

**ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΑ**  
**ΤΜΗΜΑ ΨΗΦΙΑΚΩΝ ΣΥΣΤΗΜΑΤΩΝ**



**Σχεδιασμός, Βελτιστοποίηση και Εφαρμογή αλγορίθμων  
Machine Learning και Deep Learning, με δυνατότητα αυτόματης  
εκτέλεσης διασταυρούμενης επικύρωσης και ανανέωσης βάσει  
των διατηρούμενων συνόλων πολυδιάστατων και σύνθετων  
δεδομένων**

**(Design, Optimization and Implementation of Machine Learning  
and Deep Learning algorithms, with the ability to automatically  
perform cross-validation and update based on the held sets of  
multidimensional and complex data)**

**ΔΙΔΑΚΤΟΡΙΚΗ ΔΙΑΤΡΙΒΗ**

**Γεώργιος Γ. Μακρίδης**

Μάρτιος 2023

ΠΕΙΡΑΙΑΣ







**ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΑ**  
**ΤΜΗΜΑ ΨΗΦΙΑΚΩΝ ΣΥΣΤΗΜΑΤΩΝ**

**Σχεδιασμός, Βελτιστοποίηση και Εφαρμογή αλγορίθμων  
Machine Learning και Deep Learning, με δυνατότητα αυτόματης  
εκτέλεσης διασταυρούμενης επικύρωσης και ανανέωσης βάσει  
των διατηρούμενων συνόλων πολυδιάστατων και σύνθετων  
δεδομένων**

**ΔΙΔΑΚΤΟΡΙΚΗ ΔΙΑΤΡΙΒΗ**

**Γεώργιος Γ. Μακρίδης**

**Συμβουλευτική επιτροπή:** Αν. Καθ. Δημοσθένης Κυριαζής Καθ. Ηλίας Μαγκλογιάννης  
Καθ. Μιχαήλ Φιλιππάκης

Εγκρίθηκε από την επταμελή εξεταστική επιτροπή την Πέμπτη 30 Μαρτίου 2023

.....  
Δ. Κυριαζής  
Αν. Καθ. Πανεπ. Πειραιά

.....  
Η. Μαγκλογιάννης  
Καθ. Πανεπ. Πειραιά

.....  
Μ. Φιλιππάκης  
Καθ. Πανεπ. Πειραιά

.....  
Μαρία Χαλκίδη  
Αν. Καθ. Πανεπ. Πειραιά

.....  
Ανδριάνα Πρέντζα  
Καθ. Πανεπ. Πειραιά

.....  
Χρήστος Δουλκερίδης  
Αν. Καθ. Πανεπ. Πειραιά

.....  
Π. Τσανάκας  
Καθ. Εθνικού Μετσόβιου Πολυτεχνείου

ΠΕΙΡΑΙΑΣ  
30 Μαρτίου 2023

.....  
**Γεώργιος Μακρίδης**

Copyright © Γεώργιος Μακρίδης, 2023

Με επιφύλαξη παντός δικαιώματος. All rights reserved.

Απαγορεύεται η αντιγραφή, αποθήκευση και διανομή της παρούσας εργασίας, εξ' ολοκλήρου ή τμήματος αυτής, για εμπορικό σκοπό. Επιτρέπεται η ανατύπωση, αποθήκευση και διανομή για σκοπό μη κερδοσκοπικό, εκπαιδευτικής ή ερευνητικής φύσης, υπό την προϋπόθεση να αναφέρεται η πηγή προέλευσης και να διατηρείται το παρόν μήνυμα. Ερωτήματα που αφορούν τη χρήση της εργασίας για κερδοσκοπικό σκοπό πρέπει να απευθύνονται προς τον συγγραφέα.

Οι απόψεις και τα συμπεράσματα που περιέχονται σε αυτό το έγγραφο εκφράζουν τον συγγραφέα και δεν πρέπει να ερμηνευθεί ότι αντιπροσωπεύουν τις επίσημες θέσεις του Πανεπιστημίου Πειραιά.



## ΕΥΧΑΡΙΣΤΙΕΣ

Οφείλω να ομολογήσω ότι έχοντας διαβάσει αρκετές διατριβές πάντα έδινα ιδιαίτερη σημασία στην ενότητα των ευχαριστιών. Και μετά από 3,5 χρόνια πολύ σκληρής δουλειάς, όπου ξεκίνησα την εκπόνηση της Διδακτορικής Διατριβής μου στο τμήμα Ψηφιακών Συστημάτων της σχολής Τεχνολογιών Πληροφορικής και Επικοινωνιών του Πανεπιστημίου Πειραιά (ΠΑΠΕΙ), ήρθε η ώρα να εκφράσω και εγώ τις δικές μου. Καταρχάς, θα ήθελα να εκφράσω τις ειλικρινείς μου ευχαριστίες στον επιβλέποντα καθηγητή μου, Αναπληρωτή Καθηγητή κ. Δημοσθένη Κυριαζή και Κοσμήτορα του Τμήματος Ψηφιακών Συστημάτων του ΠΑΠΕΙ τόσο για την ευκαιρία που μου προσέφερε όσο για όλη την υποστήριξη και την καθοδήγηση που μου πρόσφερε σε όλη τη διάρκεια της διατριβής. Η βοήθειά του ήταν ανεκτίμητη σε όλους αυτά τα χρόνια. Μου πρόσφερε ένα υπέροχο και απαιτητικό ακαδημαϊκό και επαγγελματικό περιβάλλον, περιτριγυρισμένος από ικανούς συναδέλφους. Θα ήθελα επίσης να τον ευχαριστήσω για την εμπιστοσύνη του και την ακαδημαϊκή ελευθερία στα πρώτα μου βήματα στην έρευνα, ενώ θα ήθελα να εκφράσω την επιθυμία μου να συνεργαστώ περαιτέρω μαζί του στο εγγύς μέλλον.

Επιπλέον, θα ήθελα να ευχαριστήσω ιδιαίτερος τους άλλους δύο συμβούλους μου και μέλη της Τριμελούς Συμβουλευτικής Επιτροπής, Καθηγητή κ. Ηλία Μαγκλογιάννη και Αν. Καθηγητή κ. Μιχαήλ Φιλιππάκη, για τις συμβουλές τους και τις για τις επιστημονικές τους γνώσεις, την ειλικρινή ανατροφοδότηση και τις επαγγελματικές συμβουλές που παρείχαν. Εν συνεχεία θα ήθελα να εκφράσω την ευγνωμοσύνη μου στο πρόσωπο του κυρίου Ιωάννη Σολδάτου για την αμέριστη εμπιστοσύνη και την αγαστή ερευνητική συνεργασία τα τελευταία 2 χρόνια. Επιπλέον θα ήθελα επίσης να εκφράσω τις ευχαριστίες μου στους συν-συγγραφείς μου, Γεώργιο Φατούρο, Δημήτριο Κότιο, Φίλιππο Μαυρεπή και την Αργυρώ Μαυρογιώργου για την εξαιρετική συνεργασία, τη συνεισφορά και την υποστήριξη τους .

Η έρευνα ήταν μια από τις φιλοδοξίες μου που έρχεται να εκπληρώσει αυτή η διατριβή, αλλά δεν θα μπορούσα ποτέ να τα καταφέρω χωρίς την υποστήριξη και την αγάπη των γονιών μου Γρηγορίου και Θεογνωσίας, του αδερφού μου Ιωάννη και της αδερφής μου Πηνελόπης.

Σε όσους ήταν δίπλα μου  
όλον αυτόν τον καιρό...



## ΠΕΡΙΛΗΨΗ

Η παρούσα διατριβή καλύπτει πρωτοποριακή έρευνα στην τομή των ερευνητικών πεδίων της μηχανικής μάθησης και της πρόβλεψης χρονοσειρών, με εφαρμογή σε πραγματικά δεδομένα βιομηχανικών διαδικασιών. Ο όρος "χρονοσειρά" αναφέρεται σε μια ακολουθία ή ροή δεδομένων που είναι χρονικά εξαρτημένη, με τη χρονική συσχέτιση να είναι θεμελιώδης. Τέτοιου είδους χρονοσειρές είναι πανταχού παρούσες στο περιβάλλον και στην καθημερινότητά μας. Η σημασία αυτών των δεδομένων απορρέει ακριβώς από την χρονική τους συσχέτιση. Ιδιαίτερο ενδιαφέρον υπάρχει στην περίπτωση της συσχέτισης μεταξύ διαφορετικών χρονοσειρών πολλών διαστάσεων. Η μηχανική μάθηση αποτελεί σημαντικό παράγοντα στις πιο πρόσφατες εξελίξεις για την αντιμετώπιση των προκλήσεων που συνεπάγεται η εξαγωγή χρήσιμων πληροφοριών από δεδομένα χρονοσειρών. Ταυτόχρονα, η ερευνητική επίδραση αυξάνεται όταν υπάρχει παράλληλο οικονομικό όφελος για τους βιομηχανικούς φορείς.

Η προσέγγισή μας για την επεξεργασία και ανάλυση δεδομένων περιλαμβάνει την προεπεξεργασία των ακατέργαστων δεδομένων, την εκπαίδευση ενός μοντέλου μηχανικής μάθησης στα προεπεξεργασμένα δεδομένα για την πρόβλεψη ή κατηγοριοποίηση επιθυμητών μεταβλητών, την απεικόνιση των αποτελεσμάτων σε χρονοσειρές για χρονική ανάλυση, και την εφαρμογή της επεξηγήσιμης τεχνητής νοημοσύνης για καλύτερη κατανόηση των μοντέλων και των προβλέψεών τους.

Η διατριβή αφορά την ανάλυση προβλημάτων που εμφανίζονται σε διάφορα στάδια της παραγωγικής διαδικασίας στον βιομηχανικό τομέα και για τα οποία δεν έχουν ακόμη εφαρμοστεί μέθοδοι Μηχανικής Μάθησης. Σε αυτές τις περιπτώσεις υλοποιούμε το προτεινόμενο ολοκληρωμένο πλαίσιο τεχνητής νοημοσύνης το οποίο καλύπτει τόσο προβλήματα κατηγοριοποίησης όσο και παλινδρόμησης. Ένας ειδικός στόχος της εργασίας είναι ο σχεδιασμός και η βελτιστοποίηση μαθηματικών μοντέλων Μηχανικής Μάθησης, με ιδιαίτερη έμφαση σε μοντέλα νευρωνικών δικτύων πολλαπλών επιπέδων που εφαρμόζονται σε πολυδιάστατα και πολύπλοκα πραγματικά δεδομένα σε διάφορους βιομηχανικούς τομείς, όπως την αγροτική, την ναυτιλιακή και τη χρηματοπιστωτική βιομηχανία. Στόχος μας είναι να βελτιώσουμε αυτές τις μεθόδους και να αναπτύξουμε νέες με βάση τις προκλήσεις και τους περιορισμούς που προκύπτουν από τη φύση των δεδομένων. Η διατριβή επικεντρώνεται στην πρόβλεψη μελλοντικών γεγονότων και συμπεριφορών, επιτρέποντας στις επιχειρήσεις να διεξάγουν αναλύσεις για την πρόβλεψη των επιπτώσεων των πιθανών αλλαγών στις επιχειρηματικές στρατηγικές τους.

Συγκεκριμένα η διατριβή στοχεύει στην δημιουργία και εφαρμογή προηγμένων αλγόριθμων ανάλυσης δεδομένων γενικότερα, παρέχοντας τη δυνατότητα οπτικής απεικόνισης αυτών με σκοπό την καλύτερη και βαθύτερη κατανόηση των προβλημάτων, ενώ παράλληλα μοντέλα προβλέψεων βασισμένα σε τεχνολογίες Μηχανικής Μάθησης προτείνονται με σκοπό την βέλτιστη λήψη αποφάσεων. Οι τεχνικές Μηχανικής Μάθησης καθώς και τεχνικές πρότυπης μοντελοποίησης θα

είναι σε θέση να διαχειριστούν τον κύκλο ζωής της προετοιμασίας, της εξερεύνησης και της ανάλυσης δεδομένων, για την επίτευξη καλύτερης απόδοσης και παρακολούθησης του επιλυόμενου προβλήματος. Αναλυτικότερα, εστιάζουμε τόσο σε πρακτικές όσο και σε θεωρητικές πτυχές των μεθόδων Μηχανικής Μάθησης για την εξόρυξη και διαχείριση δεδομένων χρονοσειρών. Οι ερευνητικοί μας στόχοι καθοδηγούνται από τρεις προκλήσεις στην εφαρμογή μεθόδων Μηχανικής Μάθησης (και κυρίως των σχετικών με τις χρονοσειρές) σε πραγματικές συνθήκες όπως η συνεχής βελτιστοποίηση των μοντέλων Μηχανικής Μάθησης (ακρίβεια, χρόνος εκτέλεσης κ.α.) την βελτίωση των αποτελεσμάτων του μοντέλου λόγω της χρήσης υποκατάστατων δεδομένων που εμπλουτίζουν το σύνολο δεδομένων και συμβάλλουν στην καλύτερη γενίκευση του προτεινόμενου μοντέλου και τη βέλτιστη επιλογή διαφορετικών μοντέλων.

Τα δεδομένα που χρησιμοποιήθηκαν σε αυτή τη διατριβή προέρχονται από 3 τομείς της πραγματικής οικονομίας τη ναυτιλία, την ασφάλεια τροφίμων και τον τραπεζικό τομέα. Όσον αφορά τον πρώτο τομέα, η έρευνα επικεντρώθηκε στο πρόβλημα της προβλεπτικής συντήρησης των κύριων μηχανών εμπορικών πλοίων. Από τον τομέα της ασφάλειας τροφίμων και ανάλυσης κειμένου επικεντρώθηκε στη πρόβλεψη συμβάντων ανάκληση τροφίμων από δεδομένα ροής μικρών κειμένων καθώς και την εξαγωγή συμπερασμάτων για την συσχέτιση περιβαλλοντικών συνθηκών με την ποιότητα του κρέατος των χοίρων. Τέλος στον χρηματοοικονομικό/τραπεζικό τομέα μελετήσαμε την πρόβλεψη ροών χρήματος σε Μικρομεσαίες Επιχειρήσεις (ΜΜΕ) βασισμένες σε ιστορικά δεδομένα συναλλαγών. Μία από τις πρώτες συνεισφορές μας είναι μια ολοκληρωμένη διερεύνηση των προαναφερθέντων πραγματικών προβλημάτων, όπου εντοπίσαμε ερευνητικές ευκαιρίες.

Τα πλοία, παρόμοια με άλλους τύπους “εξοπλισμού” είναι εφοδιασμένα με αισθητήρες ώστε να μπορούν να συλλέγουν πληροφορίες για την συνολική λειτουργία των μηχανικών και ηλεκτρικών μερών τους. Οι αισθητήρες αυτοί παρέχουν ροές δεδομένων που μπορούν να αναλυθούν σε πραγματικό χρόνο μέσω τεχνικών Μηχανικής Μάθησης και να συλλέγουν πληροφορίες για πιθανά σφάλματα στις κύριες μηχανές ή σε άλλα μέρη των πλοίων, αυτό ακριβώς αποτελεί και τον πρώτο πυλώνα των επιστημονικών μας συνεισφορών. Οι πληροφορίες αυτές μπορούν να χρησιμοποιηθούν ώστε να διευκολύνουν ή να βελτιώσουν αποφάσεις, όπως την παραγγελία ανταλλακτικών ή την αλλαγή προορισμού σε λιμάνια για την αντικατάσταση εξαρτημάτων που αναμένεται να χαλάσουν. Όλα τα παραπάνω μπορούμε να τα συνοψίσουμε ως μια συνολική προσέγγιση προβλεπτικής συντήρησης που οδηγείται από τα δεδομένα. Στην εργασία αυτή, παρουσιάζουμε μια προσέγγιση για τον εντοπισμό ανωμαλιών σε δεδομένα χρονοσειρών, εφαρμόζοντας τεχνικές Μηχανικής Μάθησης στα δεδομένα των αισθητήρων του πλοίου, με σκοπό την έγκαιρη εκτίμηση της κατάστασης συγκεκριμένων μερών της κύριας μηχανής του πλοίου έτσι ώστε η προβλεπτική συντήρηση να καταστεί δυνατή. Η προτεινόμενη προσέγγιση ενσωματώνει αρκετά μοντέλα που έχουν αναλυθεί και εφαρμοστεί στο πλαίσιο της προβλεπτικής συντήρησης στον



ναυτιλιακό κλάδο, σε μια συλλογή μοντέλων για να συλληχθούν σαφέστερες πληροφορίες για τα αποτελέσματα και την δυνατότητα εφαρμογής προβλεπτικής συντήρησης.

Η δεύτερη σημαντική συμβολή αυτής της έρευνας αφορά την ανάλυση και πρόβλεψη πληροφοριών σχετικά με πιθανά μη ασφαλή διατροφικά αγαθά και προϊόντα. Αυτές οι πληροφορίες αξιοποιούνται για να οδηγήσουν στη λήψη αποφάσεων αναφορικά με το ποια προϊόντα είναι πιο πιθανό να είναι επιβλαβή στο εγγύς μέλλον και στη συνέχεια να βελτιστοποιήσουν την αλυσίδα εφοδιασμού τροφίμων. Στο πλαίσιο της διατριβής, εισάγουμε μια προσέγγιση Βαθιάς Μάθησης αξιοποιώντας τεχνικές επεξεργασίας φυσικής γλώσσας και πρόβλεψης χρονοσειρών, για την παρακολούθηση και την ανάλυση του κινδύνου που σχετίζεται με κάθε κατηγορία προϊόντων διατροφής και τις αντίστοιχες πιθανές ανακλήσεις. Επιπλέον, προτείνουμε μια τεχνική που εκμεταλλεύεται την μεθοδολογία της Ενισχυτικής Μάθησης αξιοποιώντας ιστορικά δεδομένων ανακοινώσεων ανάκλησης προϊόντων τροφίμων για την πρόβλεψη των μελλοντικών ανακλήσεων τους. Με τον τρόπο αυτό παρέχονται πληροφορίες στις εταιρείες τροφίμων σχετικά με τις επερχόμενες τάσεις στις ανακλήσεις τροφίμων που μπορούν να οδηγήσουν σε αυξημένο ποιοτικό έλεγχο και επιτήρηση για έγκαιρες ανακλήσεις. Αξιολογούμε επίσης και αποδεικνύουμε την αποτελεσματικότητα και την προστιθέμενη αξία των προτεινόμενων προσεγγίσεων μέσω ενός πραγματικού σεναρίου με υποσχόμενα αποτελέσματα. Επιπροσθέτως, εξετάζεται στο υπόψη πραγματικό σενάριο η χρήση ανάλογων/υποκατάστατων δεδομένων ως μέθοδος επαύξησης του συνόλου δεδομένων για πιο ακριβή αποτελέσματα με θετικά αποτελέσματα.

Παράλληλα σχετικά με την ασφάλεια τροφίμων, ένα δεύτερο πραγματικό σενάριο το οποίο εξετάστηκε στο πλαίσιο της διατριβής και δείχνει την προστιθέμενη αξία της χρήσης των διαφόρων δεδομένων που συλλέγονται είναι αυτό της ανεπιθύμητης γεύσης και οσμής που μπορεί να υπάρχει στο κρέας των κάπρων, όταν μαγειρεύεται το χοιρινό, προκαλώντας το δυσάρεστο άρωμα, τεχνικά γνωστό ως κηλίδωση κάπρου. Αξίζει να σημειωθεί ότι φυσικές δυσάρεστες οσμές δεν δημιουργούν πρόβλημα ασφάλειας των τροφίμων, αλλά αποτελούν απειλή για την εμπειρία φαγητού/μαγειρικής και απαγορεύεται από τους κανονισμούς για την ποιότητα των τροφίμων στις περισσότερες χώρες. Χρησιμοποιώντας τα συμπεράσματα της εν λόγω μελέτης, που συγκαταλέγονται στα αποτελέσματα της διατριβής, οι χοιροτρόφοι μπορούν να αποκτήσουν γνώσεις σχετικά με τον τρόπο με τον οποίο πρέπει να προσαρμόσουν τη διαχείρισή των κάπρων για να μειώσουν τη κηλίδωση τους. Συγκεκριμένα, αυτή η μελέτη εξετάζει πολλαπλές προγνωστικές προσεγγίσεις που βασίζονται σε δεδομένα σε συνδυασμό με μεθόδους εξηγήσιμης τεχνητής νοημοσύνης, αξιολογώντας τις σε σχέση με διάφορες εξηγήσιμες μετρήσεις, καταλήγοντας σε χρήσιμα συμπεράσματα, πληροφορίες και συγκεκριμένες προτάσεις. Αναλυτικότερα, σε αυτήν την προσέγγιση, η εξεταζόμενη περίπτωση χρήσης μοντελοποιήθηκε ως δυαδική εργασία ταξινόμησης με αποτέλεσμα ένα εξαιρετικά ανισόρροπο σύνολο δεδομένων. Με αυτήν την προσέγγιση, έχουν προκύψει ορισμένα λειτουργικά χαρακτηριστικά τα οποία

φαίνεται να συσχετίζονται με την κηλίδωση των κάπρων και αφορούν τις συνθήκες του αγροκτήματος/στάβλου και του σφαγείου, όπως ο τύπος της τροφής, ο τύπος του συστήματος εξαερισμού, η φαρμακευτική αγωγή, ο τύπος δαπέδου και η διάρκεια της αναμονής στο χώρο αποθήκευσης.

Η τρίτη επιστημονική μας προσθήκη αφορά τον τραπεζικό κλάδο και πιο συγκεκριμένα τις Μικρομεσαίες Επιχειρήσεις (ΜΜΕ) οι οποίες αντιμετωπίζουν ένα περίπλοκο και προκλητικό περιβάλλον, καθώς στους περισσότερους τομείς υστερούν στον ψηφιακό μετασχηματισμό τους. Οι τράπεζες, διατηρώντας μια ποικιλία δεδομένων των ΜΜΕ-πελατών τους για την εκτέλεση των βασικών τους δραστηριοτήτων, θα μπορούσαν να προσφέρουν επιπλέον μια λύση χρησιμοποιώντας όλα τα διαθέσιμα δεδομένα για να παρέχουν στους πελάτες τους μια εργαλειοθήκη Οικονομικής Διοίκησης Επιχειρήσεων (ΟΔΕ), παρέχοντας ψηφιακές υπηρεσίες προστιθέμενης αξίας βασισμένες στην μηχανική μάθηση. Παρά την επιτυχία της Βαθιάς Μάθησης σε πολλούς τομείς, ο σχεδιασμός τέτοιων μοντέλων βασίζεται στη διαδικασία της δοκιμής και του λάθους. Μια αυστηρή μαθηματική θεωρία υπερπαραμετροποιημένων μοντέλων η οποία να καλύπτει όλα τις περιπτώσεις εφαρμογής εξακολουθεί να λείπει. Προς αυτή την κατεύθυνση, η παρούσα εργασία περιστρέφεται γύρω από την ανάπτυξη ενός έξυπνου, εξατομικευμένου υβριδικού μοντέλου κατηγοριοποίησης συναλλαγών, διασυνδεδεμένου με ένα μοντέλο πρόβλεψης ταμειακών ροών που βασίζεται σε επαναλαμβανόμενα νευρωνικά δίκτυα Βαθιάς Μάθησης. Καθώς η ταξινόμηση των συναλλαγών είναι μεγάλης σημασίας, αυτή η έρευνα επεκτείνεται προς επεξηγήσιμη τεχνητή νοημοσύνη, όπου τα εργαλεία /μέθοδοι LIME και SHAP χρησιμοποιούνται για την ερμηνεία και την απεικόνιση των αποτελεσμάτων ταξινόμησης Μηχανικής Μάθησης. Η προσέγγισή μας δείχνει πολλά υποσχόμενα αποτελέσματα σε μια περίπτωση τραπεζικής χρήσης σε πραγματικό κόσμο και λειτουργεί ως το θεμέλιο για την ανάπτυξη περαιτέρω τραπεζικών μικροϋπηρεσιών ΟΔΕ, όπως η πρόβλεψη ταμιακών ροών, ο εντοπισμός απάτης και η παρακολούθηση του προϋπολογισμού.

Αυτή η Διατριβή είναι ένα βήμα προς πιο αξιόπιστες και αποτελεσματικές μεθόδους Μηχανικής Μάθησης για πραγματικά προβλήματα βιομηχανίας κυρίως χρησιμοποιώντας δεδομένα χρονοσειρών. Ως εκ τούτου στο προτεινόμενο πλαίσιο της διατριβής έχει ενσωματωθεί ένα καινοτόμο μοντέλο ΧΑΙ για χρονοσειρές.

Ελπίζουμε ότι τα ευρήματά μας θα αποτελέσουν κίνητρο για τους μελλοντικούς ερευνητές και θα χρησιμεύσουν ως εργαλεία για μηχανικούς σε βιομηχανικές εφαρμογές υψηλού αντίκτυπου.



## ABSTRACT

This thesis covers pioneering research at the intersection of the research fields of machine learning and time series forecasting, with application to real industrial process data. The term "time series" refers to a sequence or stream of data that is time-dependent, with temporal correlation being fundamental. Such time series are ubiquitous in our environment and everyday life. The importance of these data derives precisely from their temporal correlation. Of particular interest is the case of correlation between different multidimensional time series. Machine learning is an important factor in the latest developments to address the challenges of extracting useful information from time series data. At the same time, research impact increases when there is a parallel financial benefit for industrial actors.

Our approach to data processing and analysis includes preprocessing the raw data, training a machine learning model on the preprocessed data to predict or categorize desired variables, visualizing the results in time series for temporal analysis, and applying explainable artificial intelligence to better understand the models and their predictions.

The thesis concerns the analysis of problems that appear at various stages of the production process in the industrial sector and for which Machine Learning methods have not yet been applied. In these cases we implement the proposed integrated artificial intelligence framework which covers both classification and regression problems. A specific aim of the work is the design and optimization of mathematical Machine Learning models, with a special emphasis on multi-layer neural network models applied to multi-dimensional and complex real data in various industrial sectors, such as agriculture, shipping and financial industries. Our goal is to improve these methods and develop new ones based on the challenges and limitations arising from the nature of the data. The thesis focuses on predicting future events and behaviors, enabling businesses to conduct analysis to predict the impact of potential changes to their business strategies.

Specifically, we focus on both practical and theoretical aspects of machine learning methods when dealing with time series data mining and management. Our research goals are driven by three common "gaps" in the application of Machine Learning methods and time series, such as the continuous optimization of Machine Learning models (and as a result, their results) the improvement of model results due to the use of surrogate data that they enrich the data set and contribute to a better generalization of the proposed model and an optimal selection of different models. The data areas covered in this project cover shipping, food safety and banking. Regarding the first area, the research focused on the problem of predictive maintenance of the main engines of merchant ships. From the field of food safety and text analysis focused on predicting food recall events from small text stream data as well as inferring the correlation of environmental conditions with meat quality. Finally, in the financial sector, we studied the forecast of cash flows in Small

and Medium Enterprises (SMEs) based on historical transaction data. One of our first is a comprehensive investigation of the actuals, where we identified problem investigations.

Ships, like other types of "equipment", in their factory settings, are equipped with sensors so that they can collect information about the overall operation of the particular sector - in this case, the marine industry and the condition of its equipment - and more specifically in the case of ships/vessels. These sensors provide data streams that can be analyzed in real time through Artificial Intelligence techniques and collect information about possible errors in the machines, this is precisely the first pillar of our scientific contributions. This information is exploited to drive decisions such as ordering spare parts or changing ports of destination to replace parts that will not fail. The second trade-off in a comprehensive, data-driven predictive maintenance approach. In this work, we present an approach for detecting anomalies in time series data by leveraging Machine Learning technique on sensory data. of the ship, to predict the condition of specific parts of the ship's main engine so that predictive maintenance is possible. The presented approach incorporates both several models that have been analyzed and applied to address the challenge of predictive maintenance in the shipping industry, and a collection of these models to gather clearer information on predictive maintenance outcomes.

The second important contribution of this research concerns the analysis and prediction of information about potentially unsafe goods and products. This information is leveraged to drive decision-making, such as which products are most likely to be harmful soon, and then optimize the food supply chain. To address this, we introduce a deep learning approach leveraging Natural Language Processing and Time Series Prediction Techniques to monitor and analyze the risk associated with each food product category and the corresponding potential recalls. Furthermore, we propose a technique that exploits reinforcement learning to use historical recall announcements of food products to predict their future recalls, thus providing information to food companies about upcoming trends in food recalls that can lead to timely recalls. We also evaluate and demonstrate the effectiveness and added value of the proposed approaches through a real scenario that gives promising results. While several techniques/models have been analyzed and applied to address the challenge of food recall predictions, the use of analog/surrogate data has also been studied and evaluated for more accurate results.

Parallel to food safety, a real scenario that shows the added value of using the various data collected is that of the unwanted taste and smell that can be present in boar meat, also known as "boar taint". Using this information, pig farmers can gain insights into how they need to adjust their management to reduce boar infestation. This study examines multiple data-driven predictive approaches combined with explainable artificial intelligence (XAI) methods, evaluating them against various explainable metrics while trying to generate useful insights and

suggestions. Specifically, in this approach, the considered use case was modeled as a binary classification task resulting in a highly unbalanced data set. With this approach, some functional characteristics related to farm/stall and abattoir conditions have been derived, such as type of feed, type of ventilation system, medication, type of floor and length of time in storage.

Our third scientific addition touches on the banking industry and more specifically Small and Medium Enterprises (SMEs) which face a complex and challenging environment, as in most areas they are lagging behind in their digital transformation. Banks, maintaining a variety of data of their SME customers to perform their core activities, could offer a solution using all available data to provide their customers with a Business Financial Management (BFM) toolkit, providing value-added services. core activity. Despite the success of deep learning in many areas, the design of such models is based on the process of trial and error. A rigorous mathematical theory of overparameterized models is still lacking. Towards this end, the present work revolves around the development of an intelligent, highly personalized hybrid transaction categorization model, interfaced with a cash flow forecasting model based on recurrent neural networks (RNNs). As transaction classification is of great importance, this research is extended to explainable artificial intelligence, where the LIME and SHAP frameworks are used to interpret and visualize ML classification results. Our approach shows promising results in a real-world banking use case and serves as the foundation for the development of further BFM banking microservices such as transaction fraud detection and budget monitoring.

This scientific work is a step towards more reliable and effective machine learning methods for real industry problems mainly using time series data. Therefore, an innovative XAI model for time series has been integrated into the proposed framework of the thesis.

We hope that our findings will motivate future researchers and use serve as tools for engineers in high-impact industrial applications.



## ΠΙΝΑΚΑΣ ΠΕΡΙΕΧΟΜΕΝΩΝ

ΕΥΧΑΡΙΣΤΙΕΣ .....	5
ΠΕΡΙΛΗΨΗ.....	7
ABSTRACT .....	11
ΠΙΝΑΚΑΣ ΠΕΡΙΕΧΟΜΕΝΩΝ .....	14
ΚΑΤΑΛΟΓΟΣ ΣΥΝΤΜΗΣΕΩΝ .....	18
ΚΑΤΑΛΟΓΟΣ ΕΙΚΟΝΩΝ .....	24
ΚΑΤΑΛΟΓΟΣ ΠΙΝΑΚΩΝ .....	27
1 ΚΕΦΑΛΑΙΟ - ΕΙΣΑΓΩΓΗ .....	29
Περιεχόμενα .....	29
Περίληψη.....	29
1.1 Μηχανική Μάθηση.....	29
1.1.1 Αρχιτεκτονικές βαθιάς μάθησης.....	30
1.1.1.1 Νευρωνικά Δίκτυα Feedforward .....	37
1.1.1.2 Επαναλαμβανόμενα νευρωνικά δίκτυα .....	38
1.1.1.3 Δένδρα Απόφασης (Decision Trees).....	41
1.1.1.4 Random Forest Classifier .....	44
1.1.1.5 Αλγόριθμος Gradient Boosting .....	46
1.2 Εξόρυξη δεδομένων χρονοσειράς.....	48
1.3 Εφαρμογές Μηχανικής Μάθησης σε βιομηχανικά περιβάλλοντα με πραγματικά δεδομένα .....	50
1.3.1 Μηχανική Μάθηση & προβλεπτική συντήρηση κύριων μηχανών εμπορικών πλοίων. ....	50
1.3.2 Μηχανική Μάθηση & ασφάλεια των τροφίμων.....	52
1.3.3 Μηχανική Μάθηση & ποιότητα χοιρινού κρέατος.....	54
1.3.4 Μηχανική Μάθηση & τραπεζικές ψηφιακές υπηρεσίες για MME.....	55
1.4 Προκλήσεις κατά την εφαρμογή μεθόδων Μηχανικής Μάθησης σε πραγματικές συνθήκες. ....	57
1.4.1 Επεξηγήσιμη Τεχνητή Νοημοσύνη - eXplainable Artificial Intelligence (XAI).....	58
1.4.2 Μη ισορροπημένο σύνολο δεδομένων.....	59
1.4.3 Named Entity Recognition (NER) – Επισήμανση ακολουθίας.....	59
1.4.4 Ενισχυτική μάθηση για χρονοσειρές .....	60
1.4.5 Υποκατάστατα-Αναλογικά Δεδομένα χρονοσειράς.....	61

1.5	Συνεισφορά και καινοτομία Διατριβής.....	62
1.5.1	Μηχανική Μάθηση & προβλεπτική συντήρηση κύριων μηχανών εμπορικών πλοίων. ....	62
1.5.2	Μηχανική Μάθηση & ασφάλεια των τροφίμων.....	63
1.5.3	Μηχανική Μάθηση & ποιότητα χοιρινού κρέατος.....	64
1.5.4	Μηχανική Μάθηση & τραπεζικές ψηφιακές υπηρεσίες για ΜΜΕ.....	65
1.6	Αντικείμενο και Δομή Διατριβής .....	66
2	ΚΕΦΑΛΑΙΟ - ΒΙΒΛΙΟΓΡΑΦΙΚΗ ΑΝΑΣΚΟΠΗΣΗ.....	69
	Περιεχόμενα.....	69
	Περίληψη.....	69
2.1	Πρόβλεψη χρονοσειρών .....	69
2.2	Επεξηγήσιμη Τεχνητή Νοημοσύνη - explainable AI (XAI) .....	71
2.3	Μη ισορροπημένα δεδομένα .....	79
2.3.1	Υπερδειγματοληψία .....	80
2.3.2	Υποδειγματοληψία.....	80
2.4	Εφαρμογές Μηχανικής Μάθησης.....	81
2.4.1	Μηχανική Μάθηση & προβλεπτική συντήρηση κύριων μηχανών εμπορικών πλοίων. ....	81
2.4.2	Μηχανική Μάθηση & ασφάλεια των τροφίμων.....	82
2.4.3	Μηχανική Μάθηση & ποιότητα χοιρινού κρέατος.....	83
2.4.4	Μηχανική Μάθηση & τραπεζικός βοηθός για μικρομεσαίες επιχειρήσεις.....	83
3	ΚΕΦΑΛΑΙΟ - ΑΡΧΙΤΕΚΤΟΝΙΚΗ ΣΧΕΔΙΑΣΗ .....	87
	Περιεχόμενα.....	87
	Περίληψη.....	87
3.1	Μηχανική Μάθηση & προβλεπτική συντήρηση κύριων μηχανών εμπορικών πλοίων. 90	
3.1.1	Σύνολο Δεδομένων .....	90
3.1.2	Αρχιτεκτονική .....	92
3.1.3	Προ-επεξεργασία Δεδομένων.....	93
3.1.4	Μοντέλα Μηχανικής Μάθησης – Στατιστικά Μοντέλα .....	95
3.2	Μηχανική Μάθηση & ασφάλεια τροφίμων. ....	102
3.2.1	Σύνολο δεδομένων .....	102
3.2.2	Αρχιτεκτονική .....	105
3.2.3	Προ-επεξεργασία .....	108

3.2.4	Υποκατάστατα Δεδομένα.....	110
3.2.5	Μοντέλα Μηχανικής Μάθησης – Στατιστικά Μοντέλα.....	111
3.2.5.1	Εργασία ταξινόμησης – Named Entity Recognition .....	111
3.2.5.2	Εργασία πρόβλεψης χρονοσειράς .....	112
3.3	Μηχανική Μάθηση & ποιότητα χοιρινού κρέατος. ....	115
3.3.1	Σύνολο δεδομένων .....	115
3.3.2	Αρχιτεκτονική .....	118
3.3.3	Προ-επεξεργασία .....	118
3.3.4	Μοντέλα Μηχανικής Μάθησης – Στατιστικά Μοντέλα.....	120
3.3.4.1	Προκαταρκτικά ευρήματα .....	123
3.3.4.2	Περαιτέρω έρευνα.....	124
3.3.4.3	Καινοτόμα προσέγγιση υποδειγματοληψίας βασισμένη σε γενετικό αλγόριθμο 127	
3.4	Μηχανική Μάθηση & τραπεζικός βοηθός για μικρομεσαίες επιχειρήσεις.....	130
3.4.1	Σύνολο δεδομένων .....	131
3.4.2	Υποκατάστατα Δεδομένα.....	133
3.4.3	Προ-επεξεργασία .....	134
3.4.4	Μοντέλα Μηχανικής Μάθησης – Στατιστικά Μοντέλα.....	135
3.5	Επεξηγήσιμη Τεχνητή Νοημοσύνη - eXplainable AI (XAI) models.....	141
3.5.1	XAI για χρονοσειρές.....	147
3.5.2	XAI για δεδομένα πίνακα .....	149
3.5.3	XAI για δεδομένα κειμένου (NLP).....	149
3.5.4	XAI για εικόνες .....	150
4	ΚΕΦΑΛΑΙΟ - ΑΞΙΟΛΟΓΗΣΗ.....	154
	Περιεχόμενα .....	154
	Περίληψη.....	154
4.1	Μηχανική Μάθηση & προβλεπτική συντήρηση κύριων μηχανών εμπορικών πλοίων. 154	
4.2	Μηχανική Μάθηση & ασφάλεια των τροφίμων.....	166
4.2.1	Εργασία ταξινόμησης – Named Entity Recognition .....	167
4.2.2	Πρόβλεψη χρονοσειρών .....	169
4.3	Μηχανική Μάθηση & ποιότητα χοιρινού κρέατος. ....	178
4.4	Μηχανική Μάθηση & τραπεζικός βοηθός για μικρομεσαίες επιχειρήσεις.....	182
4.4.1	Ταξινόμηση υβριδικών συναλλαγών .....	184



4.4.2	Πρόβλεψη χρονοσειρών ταμειακών ροών .....	191
5	ΚΕΦΑΛΑΙΟ - ΣΥΜΠΕΡΑΣΜΑΤΑ.....	202
	Περιεχόμενα .....	202
	Περίληψη.....	202
5.1	Σύνοψη – Συμπεράσματα .....	202
5.2	Μηχανική Μάθηση & προβλεπτική συντήρηση κύριων μηχανών εμπορικών πλοίων. 203	
5.3	Μηχανική Μάθηση & ασφάλεια των τροφίμων.....	204
5.4	Μηχανική Μάθηση & ποιότητα χοιρινού κρέατος. ....	205
5.5	Μηχανική Μάθηση & τραπεζικός βοηθός για μικρομεσαίες επιχειρήσεις.....	206
6	ΠΑΡΑΠΟΜΠΕΣ.....	208

**ΚΑΤΑΛΟΓΟΣ ΣΥΝΤΜΗΣΕΩΝ**

<u>ΣΥΝΤΜΗΣΗ</u>	<u>ΕΠΕΞΗΓΗΣΗ</u>
<b>AAFT</b>	Amplitude-Adjusted Fourier Transform
<b>ACM SIGKDD</b>	Association for Computing Machinery Special Interest Group for Knowledge Discovery and Data Mining
<b>ACT</b>	Actual
<b>ADASYN</b>	ADaptive SYNthetic sampling approach for imbalanced learning
<b>ADF</b>	Augmented Dickey-Fuller
<b>ADF</b>	Augmented Dickey-Fuller
<b>AI</b>	Artificial Intelligence
<b>ANN</b>	Artificial Neural Network
<b>API</b>	Application Programming Interface
<b>AUC</b>	Area Under the Curve
<b>BERT</b>	Bidirectional Encoder Representations from Transformers
<b>BFM</b>	Business Financial Management
<b>BOA</b>	Bayes Optimization Algorithm
<b>BOW</b>	Bag of Words
<b>BPTT</b>	Back-Propagation Through Time
<b>CAM</b>	Class Activation Map
<b>CART</b>	Classification And Regression Trees
<b>CEP</b>	Complex Event Processing
<b>CNN</b>	Convolution Neural Network
<b>COR</b>	Correct
<b>CPS</b>	Cyber Physical Systems
<b>CPU</b>	Central Processing Unit
<b>CY</b>	Cyprus
<b>DAN</b>	Deep Artificial Network
<b>DAQ</b>	Data Acquisition
<b>DARPA</b>	Defense Advanced Research Projects Agency
<b>DDQN</b>	Double Deep Q-Network
<b>DEC</b>	Dynamain Error Correction
<b>DEEPLIFT</b>	Deep Learning Important FeaTures
<b>DL</b>	Deep Learning
<b>DNN</b>	Deep Neural Network

<b>DQN</b>	Deep Q-Network
<b>DT</b>	Decision Tree
<b>EDA</b>	Explanatory Data Analysis
<b>EDNN</b>	Evolutional Deep Neural Network
<b>EHR</b>	Electronic Health Record
<b>ERP</b>	Enterprise Resource Planning
<b>ESN</b>	Echo State Network
<b>ET</b>	Extra Trees
<b>ETS</b>	Exponential Smoothing
<b>EWMC</b>	Equally Weighted Model Combination
<b>FDA</b>	Food and Drug Administration
<b>FEU</b>	Forty-foot Equivalent Unit
<b>FIT</b>	Framework for Intergated Tests
<b>FTP</b>	File Transfer Protocol
<b>G</b>	gram
<b>GA</b>	Genetic Algorithm
<b>GBC</b>	Gradient Boosting Classifier
<b>GBM</b>	Gradient Boosting Machine
<b>GDPR</b>	General Data Protection Regulation
<b>GIRP</b>	Global Interpretation by Retrospective Partitioning
<b>GNRH</b>	Gonadotropin-releasing hormone
<b>GPU</b>	Graphics Processing Unit
<b>GR</b>	Greece
<b>GRU</b>	Gated Recurrent Unit
<b>HTAP</b>	Hybrid transaction/analytical processing
<b>HTTPS</b>	Hypertext Transfer Protocol Secure
<b>IAAFT</b>	Iterated Amplitude-Adjusted Fourier Transform
<b>ICA</b>	Independent Component Analysis
<b>ID</b>	Identity
<b>IEC</b>	International Electrotechnical Commission
<b>INC</b>	Incorrect
<b>IOT</b>	Internet of Things
<b>IRS</b>	Iterative Random Sampling
<b>IT</b>	Italy
<b>IT</b>	Information Technology

<b>ITVPNNW</b>	in-sample training–validation pair-based neural network weighting
<b>KG</b>	kilogram
<b>KL</b>	Kullback–Leibler
<b>KN</b>	Knots
<b>KNN</b>	K Neighbors Classifier
<b>KNN-TSPI</b>	K-Nearest Neighbors Time Series Prediction with Invariances
<b>KRR</b>	Kernel Ridge Regression
<b>KS</b>	Kennard-Stone
<b>KULEUVEN</b>	Katholieke Universiteit Leuven
<b>KW</b>	kilo Watt
<b>LDA</b>	Linear Discriminant Analysis
<b>LIGHTGBM</b>	Light Gradient Boosting Machine
<b>LIME</b>	Local Interpretable Model-Agnostic Explanation
<b>LM</b>	Language Model
<b>LOC</b>	Location
<b>LORE</b>	Local Rule-Based Explanations
<b>LR</b>	Logistic Regression
<b>LRP</b>	Layer Wise Relevance Propagation
<b>LSTM</b>	Long Short-Term Memory
<b>LT</b>	litre
<b>M</b>	meter
<b>MAE</b>	Mean Absolute Error
<b>MAP</b>	Maximum A posteriori Probability
<b>MASE</b>	Mean Absolute Scaled Error
<b>MCC</b>	Merchant Code
<b>MG</b>	miligram
<b>MIS</b>	Missed
<b>MISC</b>	Miscelaneous
<b>ML</b>	Machine Learning
<b>MLP</b>	Multi-Layer Perceptrons
<b>MOPSOCD</b>	Multi-objective Particle Swarm Optimization with Crowding Distance
<b>MP</b>	Matrix Profile
<b>MPNSGA</b>	Modified mixed population NSGA-2
<b>MSE</b>	Mean Squared Error

<b>MSIS</b>	Mean Scaled Interval Score
<b>MUC</b>	Message Understanding Conference
<b>N</b>	Newton
<b>NACE</b>	Nomenclature of Economic Activities
<b>NAN</b>	Not a Number
<b>NASA</b>	National Aeronautics and Space Administration
<b>NB</b>	Naive Bayes
<b>NBEATS</b>	Neural Basis Expansion Analysis for interpretable Time Series
<b>NER</b>	Named Entity Recognition
<b>NLP</b>	Natural Language Processing
<b>NN</b>	Neural Networks
<b>NO</b>	Number
<b>OC</b>	degree Celsius
<b>OCSVM</b>	One Class Support Vector Machines
<b>OECD</b>	Organisation for Economic Co-operation and Development
<b>OLS</b>	Ordinary Least Squares
<b>ORG</b>	Organisation
<b>PA</b>	Pascal
<b>PAR</b>	Partial
<b>PCA</b>	Principal Component Analysis
<b>PDA</b>	Prediction Difference Analysis
<b>PDP</b>	Partial Dependence Plots
<b>PER</b>	Person
<b>PFM</b>	Personal Financial Management
<b>PG-GAN</b>	Progressive Growing GAN,
<b>PM</b>	Particulate Matter
<b>POS</b>	Possible
<b>POS</b>	Part Of Speech
<b>PPB</b>	parts per billion
<b>PRC</b>	Precision Recall Curve
<b>QDA</b>	Quadratic Discriminant Analysis
<b>RBF</b>	Radial Basis Function
<b>RESNET</b>	Residual Neural NETWORK
<b>REST</b>	Representational State Transfer
<b>RETAIN</b>	Reverse Time Attention

<b>RF</b>	Random Forest
<b>RL</b>	Reinforcement Learning
<b>RNA</b>	Ribonucleic acid
<b>RNN</b>	Recurrent Neural Networks
<b>ROS</b>	Random OverSampling
<b>RUS</b>	Random UnderSampling
<b>SARIMA</b>	Seasonal Autoregressive Integrated Moving Average
<b>SD</b>	Standard Deviation
<b>SEC</b>	second
<b>SGD</b>	Stochastic Gradient Descent
<b>SH</b>	Shannon
<b>SHAP</b>	SHapley Additive Explanations
<b>SMAPE</b>	Mean Absolute Percentage Error
<b>SME</b>	Small and Medium Enterprise
<b>SMOTE</b>	Synthetic Minority Oversampling Technique
<b>SOM</b>	Self-Organising Map
<b>SOTA</b>	State of The Art
<b>SPU</b>	Spurious
<b>STYLEGAN</b>	Style Generative Adversarial Network
<b>SVM</b>	Support Vector Machines
<b>SVR</b>	Support Vector Regressor
<b>TCN</b>	Temporal Convolution Network
<b>TEU</b>	Twenty-foot Equivalent Unit
<b>TS</b>	Time Series
<b>UGENT</b>	Ghent University
<b>UML</b>	Unified Modeling Language
<b>UQ</b>	Uncertainty quantification
<b>WPE</b>	Wheighted Permutation Entropy
<b>XAI</b>	eXplainable AI
<b>XGB</b>	Extreme Gradient Boosting
<b>MME</b>	Μικρομεσαίες Επιχειρήσεις
<b>ΟΔΕ</b>	Οικονομική Διοίκηση Επιχειρήσεων
<b>ΠΟΥ</b>	Παγκόσμιος Οργανισμός Υγείας
<b>ΟΟΣΑ</b>	Οργανισμός Οικονομικής Συνεργασίας και Ανάπτυξης
<b>ΕΕ</b>	Ευρωπαϊκή Ένωση

<b>ΠΑΠΕΙ</b>	Πανεπιστήμιο Πειραιώς
<b>ΗΕΓ</b>	Ηλεκτροεγκεφαλογράφημα
<b>ΗΠΑ</b>	Ηνωμένες Πολιτίες Αμερικής

## ΚΑΤΑΛΟΓΟΣ ΕΙΚΟΝΩΝ

Εικόνα 1 Σχέση μεροληψίας – διακύμανσης .....	<b>Error! Bookmark not defined.</b>
Εικόνα 2 Δέντρο απόφασης για το φύλο του πελάτη .....	42
Εικόνα 3 Εξέλιξη μηχανικής μάθησης με εφαρμογή στη βιομηχανία .....	72
Εικόνα 4 Με βάση τις βιομηχανικές εφαρμογές, θα ταξινομήσουμε τις μεθόδους XAI σύμφωνα με την απλότητά τους, το εύρος της ερμηνευσιμότητας και το ποσοστό των εξαρτήσεων από τα αναλυόμενα/χρησιμοποιημένα μοντέλα τεχνητής νοημοσύνης .....	74
Εικόνα 5 Γενικό διάγραμμα συστήματος DAQ .....	82
Εικόνα 6 Αλγόριθμοι μηχανικής μάθησης που χρησιμοποιήθηκαν ή δοκιμάστηκαν για την περίπτωση της προβλεπτικής συντήρησης κυρίων μηχανών εμπορικών πλοίων.....	92
Εικόνα 7 Οι μπλε κορυφές είναι οι χρονικές περίοδοι κατά τις οποίες εξελισσόταν ένα ελάττωμα του Crosshead Bearing. Η κόκκινη γραμμή είναι η προβλεπόμενη πιθανότητα ελαττώματος με βάση το μοντέλο XGboost.....	97
Εικόνα 8 Το διάγραμμα ροής παρεμβολής ενός συναγερμού σύμφωνα με την προτεινόμενη προσέγγιση.....	102
Εικόνα 9 α) Αναπαράσταση στο πεδίο του χρόνου της ανάκλησης συγκεκριμένων κατηγοριών φαγητού, β) Μέσες ανακλήσεις προϊόντων συγκεντρωτικά ανα μήνα, και ανα ημέρα. Στον άξονα x εμφανίζεται η μονόδα του χρόνου (μήνες ή ημέρες) και στον y άξονα μετράει τα ποσοστά.....	104
Εικόνα 10 Αρχιτεκτονική συστήματος εξαγωγής και πρόβλεψης χρονοσειρών για την ανάκληση τροφίμων .....	105
Εικόνα 11 Σχηματική επεξήγηση των δεδομένων με βάση την προσέγγιση απο άκρο σε άκρο, συμπεριλαμβανομένων των τεσσάρων βασικών βημάτων της διαδικασίας .....	107
Εικόνα 12 Κατανομή κενών τιμών στο σύνολο δεδομένων της περίπτωσης κηλίδωσης κάπρων όπου αποτελείται από 172 στήλες (δηλαδή, χαρακτηριστικά) και 16791 σειρές (δηλαδή δείγματα). .....	116
Εικόνα 13 Heatmap συσχέτισης μεταξύ των 172 μεταβλητών του συνόλου δεδομένων. ....	117
Εικόνα 14 Κατανομή της αναλογίας των μολυσμένων δειγμάτων προς την ημερομηνία σφαγής σε διαφορετικά χρονικά πλαίσια.....	118
Εικόνα 15 Σημαντικότητα χαρακτηριστικών του μοντέλου Δένδρων αποφάσεων βάσει μεθόδου SHAP για την κηλίδωση κάπρου .....	124
Εικόνα 16 Βήματα εκπαίδευσης και επαλήθευσης για την περίπτωση ταξινόμησης των δειγμάτων κρέατος με ή χωρίς κηλίδωση κάπρου, με χρήση μεθόδου SMOTE για υπερδειγματοληψία της μειονοτικής κλάσης .....	125
Εικόνα 17 Σημαντικότητα των χαρακτηριστικών για την εμφάνιση κηλίδωσης κάπρου βάσει εκπαιδευμένου μοντέλου Catboost με υπερδειγματοληψία.....	126
Εικόνα 18 Boxplot συντελεστή ανισοροπίας ως βάρος κατηγορίας. Και στις δύο γραφικές παραστάσεις απεικονίζεται η κατανομή των αποτελεσμάτων σε διάφορους παράγοντες ανισοροπίας. Στο αριστερό διάγραμμα δεν χρησιμοποιήθηκε βάρος κλάσης ενώ στη δεύτερη γραφική παράσταση το βάρος κλάσης ορίστηκε ίσο με τον 1/παράγοντα-ανισοροπίας .....	128
Εικόνα 19 Μέθοδος υποδειγματοληψίας. Κάθε τετράγωνο είναι μια παρτίδα ίσου μεγέθους. Το κόκκινο χρώμα υποδηλώνει τη κηλιδωμένη παρτίδα, το κίτρινο τη μη κηλιδωμένη. Ο αριθμός αντιπροσωπεύει το αναγνωριστικό παρτίδας. Τα «ισοροπημένα» υποσύνολα αποτελούνταν επομένως από τη μολυσμένη παρτίδα που συνενώθηκε με μία από τις μη μολυσμένες παρτίδες .....	128



Εικόνα 20 Παράδειγμα δειγματοληψίας σε χρονικό βήμα εβδομάδας .....	135
Εικόνα 21 Σχηματική εξήγηση της πτυχής δεδομένων στην προσέγγιση από άκρο σε άκρο συμπεριλαμβανομένων των τεσσάρων βασικών βημάτων στη διαδικασία.....	138
Εικόνα 22 Διάγραμμα ροής προβολής ταμειακών ροών εσωτερικών διεργασιών .....	139
Εικόνα 23 Εννοιολογική Αρχιτεκτονική της προτεινόμενης προσέγγισης .....	141
Εικόνα 24 Παροχή παραστατικών Πληροφοριών από το την βιβλιοθήκη XAI.....	142
Εικόνα 25 Παροχή κατάταξης χαρακτηριστικών από την βιβλιοθήκη XAI .....	143
Εικόνα 26 Εσωτερική ροή εργασίας XAI.....	144
Εικόνα 27 Αρχιτεκτονική του XAI μοντέλου N-BEATS .....	148
Εικόνα 28 Επεξήγηση ταξινόμηση με τη μέθοδο LIME.....	149
Εικόνα 29 Επεξηγήσεις εικόνας με χρήση LIME .....	151
Εικόνα 30 Οι μπλε κορυφές είναι οι χρονικές περίοδοι κατά τις οποίες εξελισσόταν ένα ελάττωμα του Crosshead Bearing. Η κόκκινη γραμμή είναι η προβλεπόμενη πιθανότητα ελαττώματος με βάση το μοντέλο XGboost.....	155
Εικόνα 31 Μέσο απόλυτο σφάλμα του μοντέλου LSTM με χρήση δεδομένων δοκιμής, Μπλε γραμμή - Το MAE, Κόκκινες κάθετες γραμμές - ημερομηνίες ελαττωμάτων .....	157
Εικόνα 32 Η βαθμολογία ανωμαλίας που παράγεται από το μοντέλο OCSVM χρησιμοποιώντας σύνολο δεδομένων δοκιμής, Μπλε γραμμή - η βαθμολογία ανωμαλίας, Κόκκινες κάθετες γραμμές - ημερομηνίες ελαττωμάτων .....	157
Εικόνα 33 Τιμή εντροπίας κυλιόμενης σταθμισμένης μετάθεσης (α) Πίεση εισόδου γλυκού νερού ψύξης (β) Στροβίλο-συμπιεστής #3 Πίεση εισόδου λαδιού Lude (γ) Πίεση εισόδου λαδιού καυσίμου. Με ανάγνωση κάθετης γραμμής απεικονίζονται οι ημερομηνίες ελαττωμάτων.....	159
Εικόνα 34 Μονομεταβλητά διαγράμματα LSMT της θερμοκρασίας ανίχνευσης αέρα καθαρισμού και της θερμοκρασίας εξόδου ψύξης εμβόλου (οC) – μπλε: πραγματικές τιμές, πράσινο: προβλεπόμενες τιμές, κόκκινο: το τετράγωνο σφάλμα μεταξύ προβλεπόμενων και πραγματικών τιμών.....	160
Εικόνα 35 Μονομεταβλητή LSTM θερμοκρασίας νερού ψύξης σε κάθε κύλινδρο 1-10, χρησιμοποιώντας ως ερώτημα 30 ημέρες διαδοχική της Θερμοκρασίας Νερού Ψύξης Νο 9 πριν από το πρώτο ελάττωμα .....	162
Εικόνα 36 MP Θερμοκρασίας Νερού Ψύξης σε κάθε κύλινδρο 1-10, χρησιμοποιώντας ως ερώτημα 30 ημέρες διαδοχικά της Θερμοκρασίας Νερού Ψύξης Νο 9 πριν από το αρχική βλάβη .....	164
Εικόνα 37 Διαγράμματα συνολοκλήρωσης καυσαερίων από τους κυλίνδρους 7 και 8... 164	
Εικόνα 38 Αποτέλεσμα συνδυαστικού μοντέλου (μπλε χρώμα) , πραγματικές βλάβες (κόκκινο χρώμα) Πιθανότητα μη φυσιολογικής κατάστασης του κύριου κινητήρα .....	165
Εικόνα 39 Προτεινόμενο σχέδιο επικύρωσης βημάτων προς τα εμπρός.....	170
Εικόνα 40 Ενδεικτικά παραδείγματα προβλέψεων με χρήση εκτιμητή Deep AR σε συγκεκριμένες χρονοσειρές. Σημειωμένο με σκούρο πράσινο είναι το διάστημα πρόβλεψης 0,5 (εμπιστοσύνη) και σημειωμένο με ανοιχτό πράσινο είναι το διάστημα πρόβλεψης (εμπιστοσύνη) 0,9. Ο άξονα x παρουσιάζει το χρόνο σε μήνες ενώ ο άξονας y αντιπροσωπεύει την ποσοστιαία μεταβολή των ανακλήσεων ανά μήνα. Ο οριζόντιος πρόβλεψης είναι 4 μήνες, ενώ η μπλε γραμμή αντιπροσωπεύει τις πραγματικές τιμές της χρονοσειράς.....	172
Εικόνα 41 Συσχετίσεις των μονομεταβλητών μοντέλων DeepA και RL σχετικά με τις στατιστικές μετρήσεις .....	176
Εικόνα 42 Δέντρο απόφασης για την επιλογή μοντέλου, το κόκκινο χρώμα υποδηλώνει την επιλογή του μοντέλου Deep Ar , ενώ με το μπλέ χρώμα το μοντέλο RL .....	177

Εικόνα 43 Κατανομή βαθμολογίας F1 των διαφορετικών επαναλήψεων σε διαφορετικά υποσύνολα.....	178
Εικόνα 44 Βoxplots παράγοντα ανισορροπίας ως βάρος κατηγορίας. Και στα δύο οικόπεδα η κατανομή των έχει ως αποτέλεσμα διάφορους παράγοντες ανισορροπίας. Στο αριστερό διάγραμμα δεν χρησιμοποιήθηκε κάποιος συντελεστής να αντισταθμίσει την ανισορροπία των δεδομένων ενώ στο δεύτερο οικόπεδο το βάρος κλάσης ορίστηκε ίσο με τον συντελεστή 1/ανισορροπία.....	178
Εικόνα 45 Εφαρμογή μεθόδου LIME για επεξήγηση αποτελεσμάτων ταξινομητή (βάσει ποιον χαρακτηριστικών έγινε η κάθε απόφαση).....	179
Εικόνα 46 Τα πιο σημαντικά χαρακτηριστικά όσον αφορά την αλλαγή των τιμών πρόβλεψης.....	180
Εικόνα 47 Τα πιο σημαντικά χαρακτηριστικά από την άποψη της αλλαγής λειτουργίας απώλειας.....	181
Εικόνα 48 Λογική άποψη των συνδεδεμένων στοιχείων του πλαισίου, συμπεριλαμβανομένης της κατηγοριοποίησης συναλλαγών και της πρόβλεψης ταμειακών ροών.....	183
Εικόνα 49 Σχηματική εξήγηση της πτυχής δεδομένων στην προσέγγιση από άκρο σε άκρο συμπεριλαμβανομένων των τεσσάρων βασικών βημάτων στη διαδικασία.....	183
Εικόνα 50 Διάγραμμα ροής μοντέλου υβριδικής ταξινόμησης .....	185
Εικόνα 51 Αποτελέσματα κατηγοριοποίησης βημάτων βάσει κανόνων.....	186
Εικόνα 52 Τιμές SHAP που υποδεικνύουν τη σημασία χαρακτηριστικών της συναλλαγής Cathbost μοντέλο κατηγοριοποίησης.....	190
Εικόνα 53 Παραδείγματα κατηγοριοποίησης συναλλαγών της ανάλυσης LIME, εξηγώντας τα χαρακτηριστικά και τις αξίες τους συνεισφέρουν συγκεκριμένα στα αποτελέσματα του μοντέλου με τις παρεχόμενες πιθανότητες.....	191
Εικόνα 54 Προτεινόμενο σχέδιο επικύρωσης ένα βήμα μπροστά.....	191
Εικόνα 55 Βoxplot της κάλυψης 0,8, οι περισσότερες από τις κατηγορίες έχουν υψηλή διακύμανση ενώ μόνο ο μέσος όρος της κατηγορίας φόρων φαίνεται να υπολειτουργεί 195	
Εικόνα 56 Ενδεικτικά παραδείγματα εκτιμητή σε συγκεκριμένες χρονοσειρές, σε κάθε χρόνο διάγραμμα η μπλε γραμμή αντιπροσωπεύει τις πραγματικές τιμές ενώ η πράσινη κλίση απεικονίζει τις προβλέψεις σε επίπεδα εμπιστοσύνης 50% και 95%. .....	196

## ΚΑΤΑΛΟΓΟΣ ΠΙΝΑΚΩΝ

Πίνακας 1 Χαρακτηριστικά συνόλου δεδομένων από αισθητήρες πλοίων κύριας μηχανής (main engine data), και γενικά δεδομένα του πλοίου (vessel data) που χρησιμοποιήθηκαν στην περίπτωση της προβλεπτικής συντήρησης. ....	90
Πίνακας 2 Χαρακτηρίστηκα που εκλέχθηκαν ως προς τη συνολοκλήρωση τους	101
Πίνακας 3 Παραδείγματα των δεδομένων ανακοινώσεων ανάκλησης τροφίμων	103
Πίνακας 4 Κατηγορίες χρονοσειρών συνόλου δεδομένων.....	104
Πίνακας 5 Βήματα προ-επεξεργασίας για εργασία ταξινόμησης των ανακοινώσεων τροφίμων (κάθετα) .....	109
Πίνακας 6 Κατηγορίες δεδομένων ανάκλησης τροφίμων.....	110
Πίνακας 7 Περιγραφή του συνόλου δεδομένων για την περίπτωση κηλίδωσης του κρέατος των κάπρων .....	115
Πίνακας 8 Αποτελέσματα Αμοιβαίας πληροφορίας μεταξύ των χαρακτηριστικών και την τιμής που χρησιμοποιείται ως ένδειξη κηλίδωσης του κρέατος κάπρου. ....	121
Πίνακας 9 Αποτελέσματα Chi μεταξύ των χαρακτηριστικών και την τιμής που χρησιμοποιείται ως ένδειξη κηλίδωσης του κρέατος κάπρου. ....	121
Πίνακας 10 Αποτελέσματα Logistic Regression μεταξύ των χαρακτηριστικών και την τιμής που χρησιμοποιείται ως ένδειξη κηλίδωσης του κρέατος κάπρου.....	122
Πίνακας 11 Αποτελέσματα διαφόρων μοντέλων μηχανικής μάθησης στην εργασία ταξινόμησης των δειγμάτων χοιρινού κρέατος ως προς την εκτίμηση ύπαρξης κηλίδωσης κάπρου.....	123
Πίνακας 12 Αποτελέσματα ταξινόμησης σε μορφή Confusion Matrix σε δεδομένα ελέγχου στο αρχικό σετ ελέγχου (πριν την υπερδειγματοληψία) .....	126
Πίνακας 13 Αποτελέσματα ταξινόμησης σε μορφή Confusion Matrix σε δεδομένα ελέγχου στο αρχικό σετ ελέγχου (πριν την υπερδειγματοληψία) .....	127
Πίνακας 14 Πηγές δεδομένων που διατέθηκαν από την Τράπεζα Κύπρου .....	132
Πίνακας 15 Κύριες κατηγορίες συναλλαγών για MME (έξοδος του μοντέλου ταξινόμησης συναλλαγών).....	132
Πίνακας 16 Συγκεντρωτικά Αποτελέσματα όλων των μοντέλων και ο συνδυασμός τους για το πρόβλημα προβλεπτικής συντήρησης κύριων μηχανών πλοίων. ....	166
Πίνακας 17 Αποτελέσματα συνδιαστικού μοντέλου για πρόβλημα προβλεπτικής συντήρησης κύριων μηχανών πλοίων (Confusion Matrix). ....	166
Πίνακας 18 Σχέδιο αξιολόγησης με βάση την σημασιολογική αξιολόγηση .....	169
Πίνακας 19 Αποτελέσματα του συστήματος ταξινόμησης ως προς τις μετρήσεις αξιολόγησης.....	169
Πίνακας 20 Σύγκριση μεταξύ μονομεταβλητής προσέγγισης, αύξησης δεδομένων και Ενισχυτικής Μάθησης .....	171
Πίνακας 21 Σύγκριση προσέγγισης μονομεταβλητών μοντέλων, πολυμεταβλητών μοντέλων και Ενισχυτικής Μάθησης .....	172
Πίνακας 22 Hjorth mobility, block entropy, ADF, Variance and Seasonality component of the time-series.....	175
Πίνακας 23 Κορυφαία χαρακτηριστικά όσον αφορά την αλλαγή των τιμών πρόβλεψης και την αλλαγή της λειτουργίας απώλειας.....	181

Πίνακας 24 Χαρακτηρίστηκα που χρησιμοποιήθηκαν για το μοντέλο κανόνων και εν συνεχεία στο υβριδικό μοντέλο .....	184
Πίνακας 25 Σύγκριση Διάφορων μοντέλων ταξινόμησης για εργασία κατηγοριοποίησης συναλλαγών. Αποτελέσματα 10-πλάσια Διασταυρούμενη επικύρωση (με έντονη γραφή το προτεινόμενο μοντέλο).....	188
Πίνακας 26 Confusion Matrix της κατηγοριοποίησης συναλλαγών πολύκλασης ταξινόμησης .....	189
Πίνακας 27 Μέσα αποτέλεσμα ανά Κατηγορία Συναλλαγής με βάση διάφορες μετρήσεις.....	194
Πίνακας 28 Μέσα αποτέλεσμα ανά Κατηγορία Συναλλαγής με βάση διάφορες μετρήσεις με χρήση ανάλογων/υποκατάστατων δεδομένων. ....	194



# 1 ΚΕΦΑΛΑΙΟ - ΕΙΣΑΓΩΓΗ

## ΠΕΡΙΕΧΟΜΕΝΑ

1.1 Μηχανική Μάθηση

1.2 Εξόρυξη δεδομένων χρονοσειράς

1.3 Εφαρμογές Μηχανικής Μάθησης

1.4 Προκλήσεις κατά την εφαρμογή μεθόδων Μηχανικής Μάθησης σε πραγματικές συνθήκες.

1.5 Συνεισφορά και καινοτομία

1.6 Αντικείμενο και Δομή Διατριβής

## ΠΕΡΙΛΗΨΗ

Η παρούσα διατριβή πραγματεύεται ζητήματα εφαρμογής σύγχρονων τεχνικών τεχνητής νοημοσύνης κατά την αντιμετώπιση δυναμικών πολύπλοκων συστημάτων. Στόχος της έρευνας είναι η ανάπτυξη νέων μεθοδολογιών μηχανικής μάθησης προκειμένου να ξεπεραστούν οι προκλήσεις που θέτουν η χρήση δεδομένων χρονοσειρών. Ο τομέας της έρευνας εντοπίζεται στη διασταύρωση της μηχανικής μάθησης, της ανάλυσης χρονοσειρών και την εφαρμογή αυτών σε περιπτώσεις με πραγματικά δεδομένα σε πραγματικές συνθήκες. Αυτή η ενότητα παρουσιάζει συνοπτικά αυτές τις ερευνητικές περιοχές και παρουσιάζει τις κύριες συνεισφορές της διατριβής. Ακολούθως, αναλύονται οι προκλήσεις στο πεδίο των χρονοσειρών και η εφαρμογή μοντέλων μηχανικής μάθησης σε διάφορα βιομηχανικά περιβάλλοντα (εμπορική ναυτιλία, ασφάλεια τροφίμων, τραπεζικό τομέα κατά σειρά) εκφράζοντας την ανάγκη της κάλυψης πολλών σχετικών προκλήσεων, με αυτές της προ-επεξεργασίας, και της επαύξησης των συνόλων δεδομένων να ξεχωρίζει μεταξύ άλλων. Επιπροσθέτως, αναλύεται η καινοτομία, η συνεισφορά και οι τελικοί στόχοι της παρούσας διδακτορικής διατριβής. Ενώ τέλος παρατίθεται η δομή της υπόλοιπης διατριβής.

### 1.1 Μηχανική Μάθηση

Ζούμε στην εποχή των μεγάλων δεδομένων και προτείνεται να περαιώσει σε ακόμα μεγαλύτερα επίπεδα με την αύξηση της αποδοχής και εφαρμογής του Internet of Things (IoT). Στο Διαδίκτυο υπάρχουν περισσότερες από ένα τρισεκατομμύριο ιστοσελίδες. Κάθε δευτερόλεπτο μια ώρα βίντεο μεταφέρεται στο YouTube, ενώ σε μια ημέρα μεταφέρονται δέκα χρόνια περιεχομένου. Τα γονίδια χιλιάδων ανθρώπων έχουν προσδιοριστεί σε διάφορα εργαστήρια με χρήση τεχνολογιών μεγάλων δεδομένων [240, 241].



Η μηχανική μάθηση είναι ένας σημαντικός τομέας της τεχνολογίας της πληροφορίας που στασιμεύει στην αυτόματη ανάλυση και εκτίμηση συνθηκών και προτύπων σε μεγάλα σύνολα δεδομένων. Πιο συγκεκριμένα, η μηχανική μάθηση αποτελεί ένα σύνολο αλγορίθμων και τεχνικών που χρησιμοποιούνται για να ανακαλύπτουν προτύπους και εκτιμήσεις σε δεδομένα χωρίς προσανατολιστική παραμετροποίηση από τον χειριστή. Η μηχανική μάθηση χρησιμοποιείται για να προβλέψει μελλοντική συμπεριφορά ή να λάβει αποφάσεις αυτόματα, χρησιμοποιώντας τα εντοπισμένα πρότυπα και εκτιμήσεις στα δεδομένα.

Υπάρχουν δύο βασικοί τύποι μηχανικής μάθησης: **εποπτευόμενη μάθηση** και **μη εποπτευόμενη μάθηση**. Η πρώτη περιλαμβάνει μια προγνωστική προσέγγιση που στοχεύει στην μάθηση αντιστοίχισης από δεδομένα  $X$  σε στόχους  $Y$ . Σε αυτή την περίπτωση το εκπαιδευτικό σύνολο  $D$  αποτελείται από  $N$  ζεύγη. Κάθε είσοδος εκπαίδευσης  $X_i$  μπορεί να είναι είτε μία μεταβλητή είτε ένα διάνυσμα  $D$  - διάστασης, στην περίπτωση αυτή οι διαστάσεις ονομάζονται επίσης χαρακτηριστικά. Στον πραγματικό κόσμο, οι ροές δεδομένων μπορεί να είναι πιο σύνθετες δομές όπως εικόνες, κείμενο, χρονοσειρές, χημικές ενώσεις και άλλες. Με τον ίδιο τρόπο, οι στόχοι (ανεξάρτητες μεταβλητές)  $Y_i$  μπορούν να έχουν οποιαδήποτε μορφή δεδομένων. Στις απλούστερες περιπτώσεις οι στόχοι είναι κατηγορικά δεδομένα από ένα πεπερασμένο σύνολο ή βαθμωτές από έναν συνεχή χώρο. Ανάλογα με τον τύπο του στόχου, μια εργασία μηχανικής μάθησης μπορεί να ταξινομηθεί ως ταξινόμηση όταν τα δεδομένα είναι κατηγορικά ή ως παλινδρόμηση όταν τα δεδομένα ανήκουν σε έναν συνεχή χώρο. Υπάρχουν επίσης και αρκετές παραλλαγές/υποκατηγορίες αυτών όπως η τακτική παλινδρόμηση στην οποία υπάρχει μια φυσική σειρά των τιμών. Όσον αφορά τη μη εποπτευόμενη μάθηση, είναι επίσης γνωστή ως περιγραφική μάθηση, στο μοντέλο μηχανικής μάθησης δίνονται μόνο τα δεδομένα εισόδου με στόχο την αυτόματη εύρεση μοτίβων που χαρακτηρίζουν τα δεδομένα ως πληθυσμό. Αυτή η προσέγγιση χαρακτηρίζεται επίσης ως ανακάλυψη γνώσης, καθώς ο στόχος του αλγορίθμου είναι να ανακαλύψει αυτόματα τη γνώση που βρίσκεται κάτω από ορισμένες παρατηρήσεις. Η μάθηση χωρίς επίβλεψη δεν αποτελεί ένα καλά καθορισμένο πρόβλημα και δεν γνωρίζουμε εκ των προτέρων τι είδους γνώση θα αποκαλυφθεί. Εκτός από τους δύο κύριους τύπους μηχανικής μάθησης, υπάρχει ένας τρίτος που εντοπίζεται στη διασταύρωση της μηχανικής μάθησης και της θεωρίας αποφάσεων, και ονομάζεται **ενισχυτική μάθηση**. Η ενισχυτική μάθηση είναι μια μέθοδος χρήσιμη για εργασίες όπου ο στόχος είναι ένα μοντέλο μηχανικής μάθησης να μάθει να ενεργεί ή να λαμβάνει αποφάσεις μέσα σε ένα διαδραστικό περιβάλλον. Το εκπαιδευόμενο μοντέλο (την περίπτωση της ενισχυτικής μάθησης αναφέρεται ως «πράκτορας» agent) λαμβάνει ανατροφοδότηση από το περιβάλλον με τη μορφή ανταμοιβής (θετικής ή αρνητικής) ανάλογα με την ενέργεια που εκτελεί.

### 1.1.1 Εποπτευόμενη μάθηση



Η εποπτευόμενη μάθηση είναι μία από τις πιο δημοφιλείς κατηγορίες μηχανικής μάθησης λόγω τόσο της κατανόησης της από τους χρήστες όσο και της μεγάλης επίδρασής της στον τομέα. Όπως αναφέρθηκε ανάλογα με τον τύπο του στόχου, μια εργασία εποπτευόμενης μηχανικής μάθησης μπορεί να ταξινομηθεί ως ταξινόμηση όταν τα δεδομένα είναι διακριτά και πεπερασμένα ή ως παλινδρόμηση όταν τα δεδομένα ανήκουν σε έναν συνεχή χώρο. Ξεκινώντας με την ταξινόμηση, ο στόχος είναι η αντιστοίχιση από τις εισόδους  $X$  σε στόχους  $y$ , με  $y \in \{y_0, y_1, \dots, C\}$  και  $C$  τον αριθμό των κλάσεων. Είναι πολύ συνηθισμένο να έχουμε μόνο δύο κλάσεις και τότε έχουμε το λεγόμενο πρόβλημα δυαδικής ταξινόμησης. Για περισσότερες από δύο κλάσεις το πρόβλημα ονομάζεται πολυ-ταξική ταξινόμηση. Εάν οι ετικέτες κλάσεων δεν είναι αμοιβαία αποκλειόμενες, αναφερόμαστε στο πρόβλημα ως ταξινόμηση πολλαπλών ετικετών, αλλά περιγράφεται με μεγαλύτερη ακρίβεια ως πρόβλεψη πολλαπλών σχετικών ετικετών δυαδικών κλάσεων ή μοντέλο πολλαπλών εξόδων (Mendes & Sabourin, 2019). Στη βιβλιογραφία, ο όρος ταξινόμηση αναφέρεται σε πολυ-ταξική ταξινόμηση με μία μόνο έξοδο (He & Garcia, 2009). Το πρόβλημα ταξινόμησης μπορεί να τυποποιηθεί ως προσέγγιση συνάρτησης, με υποθέσεις όπως  $y = f(x)$  για μια άγνωστη συνάρτηση  $f$ . Στη συνέχεια, ο στόχος της μηχανικής μάθησης είναι να εκτιμηθεί η συνάρτηση  $f$ , με βάση ένα επισημασμένο σύνολο δεδομένων εκπαίδευσης. Αφού η συνάρτηση  $f$  βρεθεί, μπορεί να χρησιμοποιηθεί για να γίνουν προβλέψεις. Ο πρωταρχικός στόχος είναι να γίνουν προβλέψεις για εισόδους που δεν έχουν εμφανιστεί στο παρελθόν, και αυτό είναι γνωστό στη μηχανική μάθηση ως γενίκευση.

Προκειμένου ένα μοντέλο μηχανικής μάθησης να γενικεύεται καλά σε άγνωστα δεδομένα, η πραγματοποίηση «σκληρών» προβλέψεων δεν είναι ιδανική γιατί θα πρέπει να υπάρχει τρόπος να μετρηθεί η αβεβαιότητα του μοντέλου. ο καλύτερος τρόπος επίλυσης τέτοιων προβλημάτων είναι η χρήση των εργαλείων της θεωρίας πιθανοτήτων. Δεδομένου ενός διανύσματος εισόδου  $x$  και του συνόλου εκπαίδευσης  $D$ , η κατανομή πιθανοτήτων στο χώρο των στόχων  $y$  είναι  $p(y|x, D)$ . Όταν υπάρχουν πολλά μοντέλα και θέλουμε να τα συμπεριλάβουμε στην κατανομή πιθανοτήτων συμβολίζουμε  $p(y|x, D, M)$ , με το  $M$  να αντιπροσωπεύει ένα μοντέλο. Σε αυτό το πλαίσιο, η εκτίμηση του μοντέλου είναι η μέγιστη τιμή της εκ των υστέρων πιθανότητας (Maximum A posteriori Probability-MAP), το οποίο αντιστοιχεί στην ετικέτα κλάσης με τη μέγιστη πιθανότητα και συμβολίζεται ως εξής:

$$y_b = f_b(x) = \arg \max_{c=1}^C p(y = c|x, D) \quad (1)$$

Ορισμένες δημοφιλείς εφαρμογές της εποπτευόμενης μάθησης μεταξύ άλλων είναι η ταξινόμηση εγγράφων, το φιλτράρισμα ανεπιθύμητων μηνυμάτων ηλεκτρονικού ταχυδρομείου, η ταξινόμηση εικόνων, ο εντοπισμός λέξεων-κλειδιών στην ομιλία, και η ανίχνευση προσώπου.



Όταν η εξαρτημένη μεταβλητή (έξοδος) ενός μοντέλου μηχανικής μάθησης έχει τιμές στο συνεχή χώρο, έχουμε μοντελοποίηση προβλήματος ως πρόβλημα παλινδρόμησης. Ο σκοπός σε αυτή την περίπτωση είναι να αντιστοιχίσουμε μια πραγματική τιμή  $x_i \in R$  σε μια άλλη πραγματική τιμή  $y_i \in R$ . Η απλούστερη περίπτωση παλινδρόμησης είναι η προσαρμογή μιας ευθείας γραμμής ή μιας συνάρτησης ελαχίστων τετραγώνων. Είναι δυνατές πολυάριθμες επεκτάσεις σε αυτό το θεμελιώδες πρόβλημα, συμπεριλαμβανομένων των εισόδων πολλών διαστάσεων, των ακραίων τιμών και των μη ομαλών αποκρίσεων. Μερικά παραδείγματα παλινδρόμησης περιλαμβάνουν τα ακόλουθα: (α) Πρόβλεψη θερμοκρασίας για μια συγκεκριμένη περιοχή, με δεδομένα ιστορικά στοιχεία των προηγούμενων ημερών. (β) Πρόβλεψη της τιμής μιας μετοχής για την επόμενη μέρα, βάσει των ιστορικών τιμών της εν λόγω μετοχής και της τιμής άλλων μετοχών στην αγορά. (γ) Πρόβλεψη του χρόνου που μπορεί να αφιερώσει ένας χρήστης σε έναν ιστότοπο, δεδομένων της ηλικίας, του φύλου, της εκπαίδευσης και των τύπων ιστοτόπων που επισκέφτηκε στο παρελθόν. (δ) Πρόβλεψη της αρτηριακής πίεσης ενός ανθρώπου με μια σειρά κλινικών μετρήσεων.

### 1.1.2 Μη εποπτευόμενη μάθηση

Στην μάθηση χωρίς επίβλεψη (μη εποπτευόμενη μάθηση) παρέχεται ένα μοντέλο με καθαρές παρατηρήσεις δεδομένων και δεν υπάρχει συγκεκριμένος στόχος. Ο στόχος είναι να αποκαλυφθούν κρυφές δομές μέσα στα δεδομένα, μια εργασία που περιστασιακά αναφέρεται ως ανακάλυψη γνώσης. Σε αντίθεση με την εποπτευόμενη μάθηση, δεν ενημερωνόμαστε για το επιθυμητό αποτέλεσμα για κάθε είσοδο. Η εργασία περιγράφεται τυπικά ως εκτίμηση της πυκνότητας  $p(x_i|\theta)$ . Στο πλαίσιο της θεωρίας πιθανοτήτων, η εποπτευόμενη μάθηση αντιστοιχεί στην υπό όρους εκτίμηση της πυκνότητας πιθανότητας, ενώ η μάθηση χωρίς επίβλεψη αντιστοιχεί στην εκτίμηση της άνευ όρων πυκνότητας πιθανότητας. Η μάθηση χωρίς επίβλεψη είναι, από πολλές απόψεις, πιο αντιπροσωπευτική της γνώσης του ανθρώπου και των ζώων. Δεδομένου ότι δεν υπάρχει ανάγκη για ανθρώπινη τεχνογνωσία για τη μη αυτόματη επισήμανση των δεδομένων, η δυνατότητα εφαρμογής της μάθησης χωρίς επίβλεψη καλύπτει ένα πολύ ευρύτερο φάσμα προβλημάτων. Τα δεδομένα με ετικέτα όχι μόνο είναι δύσκολο να παραχθούν, αλλά παρέχουν επίσης περιορισμένο αριθμό πληροφοριών, οι οποίες είναι ανεπαρκείς για την ορθή εκτίμηση των παραμέτρων σύνθετων μοντέλων. Είναι προφανές ότι, η μάθηση χωρίς επίβλεψη θεωρείται πιο δύσκολη και πιο πολύπλοκη από την υπό επίβλεψη μάθηση.

Ένα τυπικό πρόβλημα της μάθησης χωρίς επίβλεψη είναι η ομαδοποίηση των δεδομένων σε ομάδες. Συνήθως, δεν υπάρχουν καν στοιχεία για τον αριθμό των ομάδων στις οποίες πρέπει να γίνει ομαδοποίηση. Έστω ότι ο αριθμός των ομάδων ή των συστάδων συμβολίζεται ως  $K$ . Ο αρχικός στόχος είναι να εκτιμηθεί η





κατανομή πιθανότητας  $p(K|K)$ , η οποία συχνά προσεγγίζεται από τον τρόπο  $K^* = \operatorname{argmax}_K p(K|D)$ . Σε αντίθεση με την εποπτευόμενη περίπτωση όπου ο αριθμός των κατηγοριών είναι σταθερός και γνωστός εκ των προτέρων, στην περίπτωση χωρίς επίβλεψη ο αριθμός των συστάδων μπορεί να είναι μια υπόθεση. Στη συνέχεια, κάθε σημείο δεδομένων πρέπει να ομαδοποιηθεί στο πιο κατάλληλο σύμπλεγμα που συμβολίζεται ως  $z_i \in \{0,1,..K\}$ . Η μεταβλητή  $z$  είναι ένα παράδειγμα κρυφής ή λανθάνουσας μεταβλητής που δεν παρατηρείται άμεσα. Το σύμπλεγμα κάθε σημείου δεδομένων εκτιμάται με τον τύπο  $z^* = \operatorname{argmax}_k p(z_i = k|x_i, D)$ .

Μερικές εφαρμογές μεθόδων ταξινόμησης μη εποπτευόμενης μηχανικής μάθησης είναι οι κάτωθι:

- Σε μια εφαρμογή όπως το Netflix, οι χρήστες μπορούν να ομαδοποιηθούν σε κατηγορίες ανάλογα με τον τύπο των ταινιών που παρακολουθούν.
- Δεδομένων κάποιων κλινικών μετρήσεων μια ομάδα ανθρώπων μπορεί να συγκεντρωθεί σε διαφορετικές κατηγορίες που αντιπροσωπεύουν τον κίνδυνο για την υγεία τους όταν έχουν μολυνθεί από έναν συγκεκριμένο ιό π.χ. Covid19.
- Στη βιολογία, τα κύτταρα μπορούν να ομαδοποιηθούν σε υποομάδες με βάση την κυτταρομετρία ροής.

### 1.1.2.1 Μη εποπτευόμενη μάθηση – Μείωση Διαστάσεων

Ένα σημαντικό παράδειγμα μάθησης χωρίς επίβλεψη, το οποίο αποτελεί μια σημαντική υποκατηγορία εφαρμογών είναι η μείωση των διαστάσεων των μεταβλητών εισόδου. Ειδικά στην περίπτωση των εφαρμογών μεγάλων δεδομένων (Big Data) και υψηλών διαστάσεων, είναι συχνά ωφέλιμο να μειώνουμε τη διάσταση των δεδομένων προβάλλοντάς τα σε έναν υποχώρο χαμηλότερης διάστασης που καταγράφει τα βασικά χαρακτηριστικά των δεδομένων. Η λογική για τη χρήση αυτής της τεχνικής είναι ότι, παρά το γεγονός ότι τα δεδομένα φαίνεται να έχουν μεγάλες διαστάσεις, μπορεί να υπάρχει μόνο ένας μικρός αριθμός βαθμών μεταβλητότητας, που αντιστοιχούν σε λανθάνοντα στοιχεία, στα δεδομένα. Για παράδειγμα, για την εμφάνιση μιας εικόνας προσώπου, είναι πιθανό να υπάρχουν μόνο μερικά υποκείμενα λανθάνοντα στοιχεία που ευθύνονται για την πλειοψηφία της μεταβλητότητας. Αυτοί οι παράγοντες περιλαμβάνουν τον φωτισμό, τη θέση, την ταυτότητα κ.α.. Ως είσοδος σε άλλα στατιστικά μοντέλα, οι χαμηλών διαστάσεων αναπαραστάσεις αντικειμένων συχνά παράγουν ανώτερη προβλεπόμενη ακρίβεια από τις αναπαραστάσεις υψηλών διαστάσεων, επειδή επικεντρώνονται στα βασικά χαρακτηριστικά του αντικειμένου ενώ φιλτράρουν μη ουσιαστικές πτυχές που στα περισσότερα προβλήματα θεωρείται ως θόρυβος που μειώνει την απόδοση των μοντέλων. Ένα άλλο πλεονέκτημα της χρήσης αναπαραστάσεων χαμηλών διαστάσεων είναι ότι επιτρέπουν ταχύτερες εφαρμογές είτε στατιστικών μοντέλων



είτε μοντέλων μηχανικής μάθησης, ενώ οι δισδιάστατες προβολές είναι αρκετά ωφέλιμες για την απεικόνιση δεδομένων υψηλών διαστάσεων. Η ανάλυση PCA (Principal Component Analysis), είναι η πιο ευρέως χρησιμοποιούμενη μέθοδος μείωσης διαστάσεων στη ανάλυση δεδομένων.

### 1.1.3 Εκπαίδευση Αλγορίθμων

Η εύρεση ενός μοντέλου και των σχετικών παραμέτρων του με σκοπό την καλή απόδοση του προκύπτοντος προγνωστικού παράγοντα σε δεδομένα που δεν είχαν προηγουμένως εμφανιστεί είναι ο απώτερος στόχος της μάθησης. Όταν εξετάζονται αλγόριθμοι μηχανικής μάθησης, υπάρχουν τρεις βασικές αλγοριθμικές φάσεις που μπορούν να διακριθούν εννοιολογικά αλλά και χρονικά ως εξής:

- (α) Εκπαίδευση ή εκτίμηση παραμέτρων
- (β) Συντονισμός υπερπαραμέτρων ή επιλογή μοντέλου
- (γ) Πρόβλεψη ή συμπέρασμα

Ουσιαστικά, βάσει των δεδομένων εκπαίδευσης θέλουμε να ανακαλύψουμε καλούς προγνωστικούς παράγοντες μεταβολές των οποίων επιδρούν σημαντικά στην τιμή που λαμβάνει η (εξαρτημένη) μεταβλητή  $y$ . Υπάρχουν δύο κύριες προσεγγίσεις για να την βέλτιστη επιλογή των προγνωστικών παραγόντων: με βάση κάποιο μέτρο ποιότητας ή με χρήση του συμπεράσματος Bayes. Η εύρεση μιας σημειακής εκτίμησης μπορεί να εφαρμοστεί και στους δύο τύπους προβλέψεων, πιθανολογικούς και μη, αλλά το συμπέρασμα Bayes απαιτεί τη χρήση πιθανοτικών μοντέλων. Για να ανακαλύψουμε καλές παραμέτρους που «ταιριάζουν» στα δεδομένα, χρησιμοποιούμε αριθμητικές μεθόδους. Οι περισσότερες μέθοδοι εκπαίδευσης μπορεί να θεωρηθούν ως στρατηγικές αναρρίχησης λόφων για την εύρεση του μέγιστου ενός στόχου, όπως το μέγιστο μιας πιθανότητας. Χρησιμοποιούμε κλίσεις και προσεγγίσεις αριθμητικής βελτιστοποίησης για να εφαρμόσουμε προσεγγίσεις αναρρίχησης λόφων.

Για να εκπαιδευτεί ένα μοντέλο μηχανικής μάθησης που βασίζεται σε δεδομένα και να έχει καλή απόδοση και σε άγνωστα δεδομένα, η τοποθέτηση ενός σετ εκπαίδευσης μπορεί να μην είναι επαρκής. Η μέθοδος της διασταυρούμενης επικύρωσης χρησιμοποιείται για την επικύρωση της συμπεριφοράς του μοντέλου σε άγνωστα δεδομένα. Μια άλλη προσέγγιση για την αποφυγή υπερβολικής προσαρμογής στα δεδομένα εκπαίδευσης είναι η προσθήκη προτεραιοτήτων στο μοντέλο μας ή η χρήση τεχνικών αποφυγής υπερπροσαρμογή στα δεδομένα εκπαίδευσης.

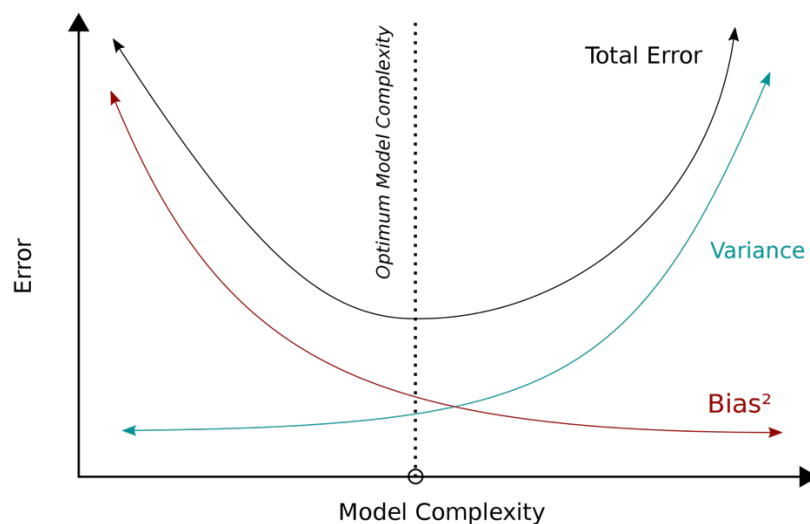
Κατά τη διάρκεια του βήματος πρόβλεψης, εφαρμόζουμε ένα εκπαιδευμένο προγνωστικό μοντέλο σε άγνωστα δεδομένα. Με άλλα λόγια, ο προγνωστικός δείκτης εφαρμόζεται σε νέα διανύσματα που αντιπροσωπεύουν νέα σημεία



δεδομένων εισόδου, ενώ οι παράμετροι και η επιλογή μοντέλου έχουν ήδη καθοριστεί. Η φάση της πρόβλεψης αναφέρεται επίσης ως συμπέρασμα. Κατασκευάζουμε το προγνωστικό μας μοντέλο ανάλογα με τα δεδομένα εκπαίδευσης κατά τη διάρκεια της φάσης εκπαίδευσης ή εκτίμησης παραμέτρων.

#### 1.1.4 Αντιστάθμιση μεροληψίας-διακύμανσης.

Μια πρόκληση που συναντάται συχνά κατά την ανάπτυξη ενός συστήματος μηχανικής μάθησης είναι η αντιστάθμιση μεροληψίας-διακύμανσης. Το σφάλμα μεροληψίας αναφέρεται στη διαφορά μεταξύ των τιμών πρόβλεψης και των πραγματικών δεδομένων. Το σφάλμα διακύμανσης αφορά τη μεταβλητότητα της πρόβλεψης όταν αλλάζουν τα δεδομένα. Η **Error! Reference source not found.** απεικονίζει την σχέση μεροληψίας-διακύμανσης όπου σημειώνεται και η βέλτιστη επιλογή πολυπλοκότητας του μοντέλου μηχανικής μάθησης.



Εικόνα 1 Σχέση μεροληψίας – διακύμανσης

Δίνεται ένα μοντέλο εποπτευόμενης μάθησης  $y = f(x) + \sigma$  και λαμβάνοντας υπόψη μια παλινδρόμηση με μέσο τετράγωνο σφάλμα (MSE), η αντιστάθμιση μεροληψίας-διακύμανσης ορίζεται ως εξής:

$$\begin{aligned} \text{Err}(x) &= E_h(Y - \hat{f}(x))^2 \\ &= E_h(Y - E(\hat{f}(x)) + E(\hat{f}(x)) - \hat{f}(x))^2 \\ &= E_h(Y - E(\hat{f}(x)))^2 + 2(Y - E(\hat{f}(x)))(E(\hat{f}(x)) - \hat{f}(x)) + (E(\hat{f}(x)) - \hat{f}(x))^2 \end{aligned}$$



$$\begin{aligned} &= E h(Y - E(\hat{f}(x)))^{2i} + E[2(Y - E(\hat{f}(x)))(E(\hat{f}(x)) - \hat{f}(x))] + E^h(E(\hat{f}(x)) - \hat{f}(x))^{2i} \\ &Y - E(\hat{f}(x)) \text{ is a constant} \\ &= E h(Y - E(\hat{f}(x)))^{2i} + 2(Y - E(\hat{f}(x)))(E(\hat{f}(x)) - E(\hat{f}(x))) + E^h(E(\hat{f}(x)) - \hat{f}(x))^{2i} \\ &= E h(Y - E(\hat{f}(x)))^{2i} + E^h(E(\hat{f}(x)) - \hat{f}(x))^{2i} \end{aligned} \quad (2)$$

Η τελευταία γραμμή της εξίσωσης (2) έχει δύο όρους. Ο πρώτος όρος αντιπροσωπεύει τη διακύμανση και ο δεύτερος το τετράγωνο της μεροληψίας. Στην πράξη αυτή η σχέση ερμηνεύεται ως εξής. Ένα απλό μοντέλο με πολύ λίγες παραμέτρους μπορεί να έχει υψηλή μεροληψία και χαμηλή διακύμανση, ένα φαινόμενο γνωστό ως υποπροσαρμογή (underfitting). Αντίθετα, ένα σύνθετο μοντέλο με μεγάλο αριθμό παραμέτρων μπορεί να έχει υψηλή διακύμανση και χαμηλή προκατάληψη, γνωστή ως υπερπροσαρμογή (overfitting). Ένα καλό μοντέλο είναι αυτό που εξισορροπεί την προκατάληψη και τη διακύμανση και ελαχιστοποιεί τη συνολική απώλεια.

Συχνά χρειάζεται να λαμβάνουμε αποφάσεις μοντελοποίησης σχετικά με τη δομή του προγνωστικού μοντέλου μηχανικής μάθησης, όπως τον αριθμό των χαρακτηριστικών που πρέπει να χρησιμοποιήσουμε ή την κατηγορία των κατανομών πιθανοτήτων που πρέπει να λάβουμε υπόψη. Ο αριθμός των παραμέτρων είναι ένα παράδειγμα υπερπαραμέτρου και μπορεί να έχει σημαντικό αντίκτυπο στην απόδοση του μοντέλου. Η επιλογή μοντέλου αναφέρεται στο πρόβλημα της επιλογής ενός μοντέλου μεταξύ πολλών επιλογών. Η επιλογή μοντέλων για μη πιθανολογικά μοντέλα πραγματοποιείται συχνά μέσω ένθετων τεχνικών διασταυρούμενης επικύρωσης.

### 1.1.5 Αρχιτεκτονικές βαθιάς μάθησης

Ο Hinton et al. (2006) [1] εισήγαγε μια πρωτοποριακή εργασία στη μάθηση χαρακτηριστικών και τη βαθιά μάθηση, η οποία υιοθετήθηκε γρήγορα από άλλους ερευνητές (Bengio et al., 2007·[2], Ranzato et al., 2007·[3] Lee et al., 2007 [4]). Ο Bengio (2009) [5] εξετάζει και συζητά εκτενώς τις ανακαλύψεις της βαθιάς μάθησης. Είναι ένας ιεραρχικός τρόπος εκμάθησης χαρακτηριστικών ανά επίπεδο χρησιμοποιώντας εκμάθηση λειτουργιών χωρίς επίβλεψη. Τα μαθησιακά χαρακτηριστικά κάθε επιπέδου χρησιμοποιούνται για την εκμάθηση των νέων χαρακτηριστικών του επόμενου. Με τη στοίβαξη αυτών των επιπέδων, μπορούν να χρησιμοποιηθούν για την εκπαίδευση ενός μοντέλου με επίβλεψη, όπως ένας ταξινομητής νευρωνικών δικτύων, ή ένα μοντέλο παραγωγής βαθιάς εμβέλειας, όπως μια μηχανή Deep Boltzmann (Salakhutdinov και Hinton, 2009) [6]. Αυτή η μεθοδολογία στοίβαξης στρωμάτων που λειτουργούν ως εξαγωγείς



χαρακτηριστικών έχει αποδειχθεί πολύ επιτυχημένη όσον αφορά το σφάλμα ταξινόμησης (Larochelle et al., 2009 [7] Erhan et al., 2010 [8]), την ποιότητα των παραγόμενων δειγμάτων στην περίπτωση των μοντέλων πιθανοτήτων (Salakhutdinov και Hinton, 2009) [6] ή από την άποψη της αμεταβλητότητας των μαθησιακών χαρακτηριστικών (Goodfellow et al., 2009) [9]. Ένα βαθύ νευρωνικό δίκτυο μπορεί να αντιπροσωπεύει συναρτήσεις αυξανόμενης πολυπλοκότητας προσθέτοντας περισσότερα επίπεδα και περισσότερους κόμβους (nodes) ανά επίπεδο. Αυτή η ενότητα περιγράφει τις βασικές έννοιες και τα βασικά δομικά στοιχεία των σύγχρονων νευρωνικών δικτύων. Δεδομένου ότι στην παρούσα διατριβή έχει γίνει εκτενής χρήση δικτύων βαθιάς μάθησης, στις επόμενες υποενότητες θα κάνουμε μια περιγραφή των βασικότερων αρχιτεκτονικών τέτοιων δικτύων.

### 1.1.5.1 Νευρωνικά Δίκτυα Feedforward

Τα νευρωνικά δίκτυα τροφοδοσίας, γνωστά και ως πολυστρωματικά perceptrons (MLPs), αποτελούν βασικό μέρος των σύγχρονων μοντέλων βαθιάς μάθησης. Τα νευρωνικά δίκτυα μπορούν να προσεγγίσουν οποιαδήποτε συνάρτηση  $f^*$ . Για παράδειγμα, δεδομένου ενός ταξινομητή  $y = f^*(x)$ , ένα δίκτυο καθορίζει μια αντιστοίχιση  $y = f(x; \theta)$  και προσεγγίζει την αρχική συνάρτηση  $f^*$  μαθαίνοντας τις σωστές παραμέτρους  $\theta$ . Το όνομα «feedforward» προέρχεται από το γεγονός ότι οι πληροφορίες ρέουν από την είσοδο  $x$  στην έξοδο μέσω των υπολογισμών που ορίζονται από την συνάρτηση  $f$ . Εάν χρησιμοποιούνται επιπλέον συνδέσεις, τότε το δίκτυο ονομάζεται επαναλαμβανόμενο.

Η συνάρτηση  $f$  τυπικά αποτελείται από πολλές συναρτήσεις και ο συνδυασμός τους περιγράφεται από ένα κατευθυνόμενο άκυκλο γράφημα. Για να ορίσουμε ένα παράδειγμα, υποθέτουμε τις συναρτήσεις  $f_1, f_2, f_3$ , τότε μια πιθανή σύνθεση είναι  $f(x) = f_1(f_2(f_3(x)))$ . Η ιεραρχική δομή των νευρωνικών δικτύων χαρακτηρίζεται από επίπεδα και σε αυτή την περίπτωση μπορούμε να πούμε ότι κάθε συνάρτηση είναι ένα στρώμα. Ο αριθμός των στρωμάτων είναι το βάθος του μοντέλου, το  $f_1$  είναι το πρώτο στρώμα, το  $f_3$  το τελευταίο και όλα τα ενδιάμεσα στρώματα ονομάζονται κρυφά. Συνήθως τα στρώματα αντιπροσωπεύονται από διανύσματα και οι διαστάσεις των διανυσμάτων αποτελούν το πλάτος κάθε στρώματος.

Οι συναρτήσεις που χρησιμοποιούνται σε δίκτυα μηχανικής μάθησης μπορούν να ομαδοποιηθούν σε δύο κύριες κατηγορίες, στις γραμμικές και τις μη γραμμικές. Οι γραμμικές μέθοδοι, όπως η λογιστική παλινδρόμηση ή η γραμμική παλινδρόμηση, είναι αποτελεσματικές αλλά το μοντέλο δεν μπορεί να μάθει περίπλοκες αλληλεπιδράσεις μεταξύ των χαρακτηριστικών εισόδου. Αυτό το πρόβλημα μπορεί να αντιμετωπιστεί εφαρμόζοντας έναν μη γραμμικό



μετασχηματισμό  $\varphi(x)$  στα δεδομένα εισόδου. Η επιλογή της σωστής συνάρτησης  $\varphi$  δεν είναι πάντα εύκολη. Στη βαθιά μάθηση η ακριβής λειτουργία υποτίθεται ότι μαθαίνεται κατά την εκπαίδευση του μοντέλου. Συγκεκριμένα, δεδομένου έστω ένα μοντέλο  $y = f(x; \theta, w) = \varphi(x; \theta)Tw$ , ο στόχος είναι το μοντέλο να «μάθει» τις παραμέτρους  $\theta$  για να προσαρμόσουμε μια γενική συνάρτηση  $\varphi$  και στη συνέχεια να χαρτογραφήσουμε τα μετασχηματισμένα χαρακτηριστικά μέσω των βαρών  $w$ . Η γενική συνάρτηση είναι επίσης γνωστή ως **συνάρτηση ενεργοποίησης** και συνήθως επιλέγεται από τους επιστήμονες δεδομένων, εκτός από τις περιπτώσεις βελτιστοποίησης υπερπαραμέτρων όπου αυτή η επιλογή σχεδιασμού είναι αυτοματοποιημένη.

Μετά το σχεδιασμό ενός deep feedforward network, ξεκινάει το στάδιο της εκπαίδευσης αυτού (training) ώστε να μάθει και να εφαρμοστεί στη κατανομή των δεδομένων. Η πιο δημοφιλής μέθοδος εκπαίδευσης είναι η backpropagation μέσω σταδιακής μάθησης. Στην περίπτωση αυτή η συνάρτηση κόστους οδηγεί τη διαδικασία της εκπαίδευσης. Το μοντέλο είναι ένα παραμετρικό μοντέλο που ορίζει μια κατανομή  $p = (y|x; \theta)$  και χρησιμοποιώντας την αρχή της μέγιστης πιθανότητας η συνάρτηση κόστους είναι συνήθως η διασταυρούμενη εντροπία (cross-entropy) μεταξύ των δεδομένων εκπαίδευσης και των προβλέψεων του μοντέλου. Προκειμένου να αποφευχθεί η υπερβολική προσαρμογή (overfitting), είναι κοινή πρακτική να συμπεριλαμβάνεται ένας όρος κανονικοποίησης στη συνάρτηση κόστους.

### 1.1.5.2 Επαναλαμβανόμενα νευρωνικά δίκτυα

Τα επαναλαμβανόμενα νευρωνικά δίκτυα (RNN) είναι εξειδικευμένα να εκπαιδεύονται από «διαδοχικά» δεδομένα. Μια θεμελιώδης ιδέα που αποτελούσε στην επιτυχή εφαρμογή των RNN είναι η κοινή χρήση παραμέτρων μεταξύ των διαφόρων στοιχείων ενός νευρωνικού δικτύου. Λόγω της κοινής χρήσης παραμέτρων, το μοντέλο μπορεί να επεκταθεί και να εφαρμοστεί σε παραδείγματα άλλων μορφών, όπως διαφορετικά μήκη διαδοχικών δεδομένων και να γενικευτεί σε αυτά. Η αλήθεια είναι ότι δεν είναι εύκολο το μοντέλο να γενικεύσει σε μήκη ακολουθίας με τις οποίες δεν έχει εκπαιδευτεί το μοντέλο, στην περίπτωση αυτή θα πρέπει να οριστούν παράμετροι για κάθε τιμή του δείκτη χρόνου με σκοπό να μοιραστούμε τη στατιστική ισχύ σε διαφορετικά μήκη ακολουθιών και θέσεις στο χρόνο. Αυτός ο τύπος κοινής χρήσης είναι ιδιαίτερα σημαντικός όταν μια συγκεκριμένη πληροφορία μπορεί να εμφανιστεί σε διάφορα χρονικά σημεία της σειράς. Για παράδειγμα, εξετάστε δύο φράσεις που παρέχουν τις ίδιες πληροφορίες αλλά με διαφορετική σύνταξη π.χ. *"Οι κορωνοϊοί είναι μια ομάδα σχετικών ιών RNA που προκαλούν ασθένειες σε θηλαστικά και πτηνά"* και *"Τα θηλαστικά και τα πτηνά μολύνονται από μια ομάδα σχετικών ιών RNA που ονομάζονται κορονοϊοί"*. Εάν η άσκηση είναι να απαντήσουμε στην ερώτηση ποιος τύπος ιού είναι ο κορονοϊός, θα



πρέπει να αναπτύξουμε ένα μοντέλο που μπορεί να παρέχει τη σωστή απάντηση που είναι "RNA" ανεξάρτητα από την θέση αυτής της λέξης στην πρόταση. Το πλεονέκτημα των RNN έναντι των κλασικών νευρωνικών δικτύων προώθησης είναι ότι μπορούν να συλλάβουν και να θυμούνται τις πληροφορίες διαδοχικά κατά την επεξεργασία της πρότασης.

Για ευκολία, θα αναφερθούμε στα RNN που δρουν σε μια ακολουθία διανυσμάτων  $x(t)$ , με τον δείκτη χρονικού βήματος  $t$  να εκτείνεται από 1 έως  $t$ . Στην πραγματικότητα, τα επαναλαμβανόμενα δίκτυα λειτουργούν συνήθως σε μικρές παρτίδες τέτοιων ακολουθιών. Τα RNN μπορούν επίσης να χρησιμοποιηθούν σε δύο διαστάσεις σε χωρικά δεδομένα, όπως εικόνες, και όταν εφαρμόζονται σε δεδομένα που σχετίζονται με το χρόνο, το δίκτυο μπορεί να περιέχει συνδέσεις που τρέχουν προς τα πίσω στο χρόνο, αρκεί να είναι γνωστή ολόκληρη η ακολουθία προτού τροφοδοτηθεί στο δίκτυο.

Υπάρχουν τρεις βασικοί τύποι επαναλαμβανόμενων αρχιτεκτονικών (Goodfellow et al., 2016) [10]. Ο πρώτος αφορά τα επαναλαμβανόμενα νευρωνικά δίκτυα που παράγουν μια έξοδο σε κάθε χρονικό βήμα και διατηρούν επαναλαμβανόμενες συνδέσεις μεταξύ κρυφών μονάδων. Ο δεύτερος τύπος περιλαμβάνει δίκτυα που παράγουν μια έξοδο σε κάθε χρονικό βήμα και έχουν επαναλαμβανόμενες συνδέσεις από την έξοδο στις κρυφές μονάδες στο επόμενο χρονικό βήμα. Τέλος, υπάρχουν επαναλαμβανόμενες αρχιτεκτονικές που έχουν επαναλαμβανόμενες συνδέσεις μεταξύ κρυφών μονάδων και διαβάζουν μια ολόκληρη ακολουθία πριν παράγουν μία έξοδο.

Στη συνέχεια, περιγράφεται ένα παράδειγμα επαναλαμβανόμενου νευρωνικού δικτύου:

Ας υποθέσουμε ότι η συνάρτηση ενεργοποίησης που χρησιμοποιείται είναι η υπερβολική συνάρτηση εφαπτομένης. Υποθέστε επίσης ότι η έξοδος είναι διακριτή, όπως η πρόβλεψη λέξεων ενός κειμένου. Ένας λογικός τρόπος για να περιγράψουμε διακριτές μεταβλητές είναι να θεωρήσουμε ότι η έξοδος ο είναι οι μη κανονικοποιημένες πιθανότητες καταγραφής των τιμών κάθε διακριτής μεταβλητής και στη συνέχεια να εφαρμόσουν συνάρτηση softmax στην έξοδο. Σε αυτό το παράδειγμα επαναλαμβανόμενου δικτύου που μεταφράζει μια ακολουθία εισόδου στην ακολουθία εξόδου του ίδιου μήκους. Η συνολική απώλεια για μια δεδομένη ακολουθία τιμών  $x$  σε σύζευξη με μια ακολουθία τιμών  $y$  είναι απλώς το άθροισμα των απωλειών σε όλα τα χρονικά βήματα. Είναι μια δαπανηρή διαδικασία για τον υπολογισμό της κλίσης αυτής της συνάρτησης κόστους σε σχέση με τις παραμέτρους. Ο υπολογισμός της κλίσης πραγματοποιείται σε δύο βασικά βήματα. (α) Αρχικά περιλαμβάνει ένα πέρασμα διάδοσης προς τα εμπρός (forward propagation pass) ακολουθώντας τη σειρά της ακολουθίας και (β) στη συνέχεια εκτελεί ένα πέρασμα διάδοσης που κινείται προς τα πίσω (backward propagation pass) στο γράφημα υπολογισμού. Λόγω του γεγονότος ότι το διάγραμμα προς τα



εμπρός είναι διαδοχικό, κάθε χρονικό βήμα μπορεί να υπολογιστεί μόνο μετά το προηγούμενο, ο χρόνος εκτέλεσης είναι  $O(t)$  και δεν μπορεί να μειωθεί εκτελώντας τη παράλληλα σε πολλές εργασίες. Οι καταστάσεις που υπολογίζονται στο «μπροστινό» πέρασμα πρέπει να διατηρούνται μέχρι να επαναχρησιμοποιηθούν στο πέρασμα προς τα πίσω, επομένως το κόστος μνήμης είναι επίσης  $O(t)$ . Αυτός ο αλγόριθμος ονομάζεται Back-Propagation Through Time (BPTT).

Μια θεμελιώδης πρόκληση στα επαναλαμβανόμενα νευρωνικά δίκτυα είναι ότι τα *gradients* που διαδίδονται σε πολλαπλά επίπεδα είτε εξαφανίζονται είτε εκρήγνυνται και το μοντέλο δεν μπορεί να συγκλίνει σε μια λύση. Ακόμα κι αν υποθέσουμε ότι οι παράμετροι είναι τέτοιες ώστε το επαναλαμβανόμενο δίκτυο να είναι σταθερό αποθηκεύοντας μνήμες, με τα *gradients* να μην εκρήγνυνται, η δυσκολία με τις μακροπρόθεσμες εξαρτήσεις προκύπτει από τα εκθετικά μικρότερα βάρη που αποδίδονται σε μακροπρόθεσμες αλληλεπιδράσεις. Αυτό συμβαίνει λόγω του πολλαπλασιασμού πολλών Jacobians, σε σύγκριση με τις βραχυπρόθεσμες αλληλεπιδράσεις. Το πρόβλημα της έκρηξης και της εξαφάνισης των *gradients* έχει μελετηθεί από πολλούς ερευνητές και υπάρχει μια πλούσια βιβλιογραφία όπου αναλύεται με μεγάλη λεπτομέρεια (Hochreiter, 1991 [11], Doya, 1993 [12], Pascanu et al., 2013 [13]). Θα μπορούσε κανείς να σκεφτεί ότι το πρόβλημα μπορεί να αποφευχθεί απλώς παραμένοντας σε μια περιοχή παραμέτρων χώρου όπου οι κλίσεις δεν εξαφανίζονται ή εκρήγνυνται. Δυστυχώς, για να αποθηκευτούν μνήμες που είναι ανθεκτικές σε μικρές διαταραχές, το RNN πρέπει να εισέλθει σε μια περιοχή του χώρου παραμέτρων όπου οι κλίσεις εξαφανίζονται Bengio et al. (1993, 1994) [14, 15]. Όταν το μοντέλο μπορεί να περιγράψει μακροπρόθεσμες εξαρτήσεις, η κλίση μιας μακροπρόθεσμης αλληλεπίδρασης έχει εκθετικά μικρότερο μέγεθος από τη διαβάθμιση μιας βραχυπρόθεσμης αλληλεπίδρασης. Αυτό δεν σημαίνει ότι η εκμάθηση μακροπρόθεσμων εξαρτήσεων είναι αδύνατη, αλλά ότι μπορεί να χρειαστεί πολύς χρόνος επειδή το σήμα σχετικά με αυτές τις εξαρτήσεις θα τείνει να κρύβεται από τις μικρότερες διακυμάνσεις που προέρχονται από βραχυπρόθεσμες εξαρτήσεις Bengio et al. (1994) [15]. Στο [15] επίσης έδειξαν ότι καθώς αυξάνεται το εύρος των εξαρτήσεων που πρέπει να αποτυπωθούν, η βελτιστοποίηση με βάση την κλίση γίνεται όλο και πιο δύσκολη. Η πιθανότητα επιτυχούς εκπαίδευσης ενός παραδοσιακού RNN μέσω SGD πλησιάζει γρήγορα το 0 για αλληλουχίες μεγέθους που δεν υπερβαίνουν το 20.

Δύο από τις πιο επιτυχημένες επαναλαμβανόμενες νευρωνικές αρχιτεκτονικές είναι η μακρο-βραχυπρόθεσμη μνήμη (Long Short-Term Memory) (LSTM) και η περιγραφόμενη επαναλαμβανόμενη μονάδα (Gated Recurrent Unit) (GRU). Αυτές οι δύο αρχιτεκτονικές ανήκουν στον ίδιο τύπο επαναλαμβανόμενων δικτύων που ονομάζονται Gated RNN. Τα Gated RNN χτίζονται στην ιδέα της δημιουργίας μονοπατιών μέσα στο χρόνο με παράγωγα που δεν εξαφανίζονται ή εκρήγνυνται. Η αρχιτεκτονική LSTM βασίζεται στη θεμελιώδη έννοια των αυτο-βρόχων (self-loops) που παράγουν μονοπάτια για την κλίση, για να ρέουν χωρίς να





εξαφανίζονται, και προτάθηκε από τους Hochreiter και Schmidhuber (1997) [16]. Η χρονική κλίμακα της ολοκλήρωσης μπορεί να τροποποιηθεί δυναμικά έχοντας το βάρος αυτού του self-loop περιφραγμένο με την έννοια ότι μπορεί να ελεγχθεί από μια άλλη κρυφή μονάδα. Σε αυτήν την περίπτωση, υπονοούμε ότι ακόμη και με ένα LSTM σταθερής παραμέτρου, η χρονική κλίμακα ολοκλήρωσης μπορεί να αλλάξει ανάλογα με την ακολουθία εισόδου, επειδή οι σταθερές χρόνου εξάγονται από το ίδιο το μοντέλο. Η αρχιτεκτονική LSTM έχει ευδοκιμήσει σε πολλές εφαρμογές που περιλαμβάνουν διαδοχικά δεδομένα όπως η αναγνώριση χειρογράφου (Graves et al., 2008) [17], η αναγνώριση ομιλίας (Graves et al., 2013, 2014) [18,19], η δημιουργία χειρογράφου (Graves et al., 2013)[18], η αυτόματη μετάφραση (Sutskever et al., 2014) [20], λεζάντες εικόνων (Kiros et al., 2014 [21]; Vinyals et al., 2015b [23]; Vinyals et al., 2015a [22]). Η αρχιτεκτονική GRU έχει προταθεί σε μια προσπάθεια να μειωθεί η πολυπλοκότητα της αρχιτεκτονικής LSTM (Cho et al., 2014a [24]; Chung et al., 2014, 2015 [25,26]; Chrupała et al., 2015 [27]; Jozefowicz et al., 2015 [28]). Η κύρια διαφορά μεταξύ GRU και LSTM είναι ότι στην αρχιτεκτονική LSTM μια ενιαία μονάδα πύλης ελέγχει τόσο το forgetting component όσο και την επιλογή ενημέρωσης της μονάδας κατάστασης.

### 1.1.5.3 Δένδρα Απόφασης (Decision Trees)

Τα δέντρα αποφάσεων θεωρούνται μία από τις πιο δημοφιλείς προσεγγίσεις για την επίλυση προβλημάτων κατηγοριοποίησης. Ερευνητές από διάφορους κλάδους όπως η στατιστική, η μηχανική μάθηση, η αναγνώριση προτύπων και η εξόρυξη δεδομένων έχουν ασχοληθεί με την ανάπτυξη και εφαρμογή μοντέλων βασισμένων στα δέντρα απόφασης από τα διαθέσιμα δεδομένα. Στην συνέχεια θα παρουσιάσουμε μια επισκόπηση κάποιων μεθόδων για την κατασκευή ταξινομητών που βασίζονται σε δέντρα αποφάσεων.

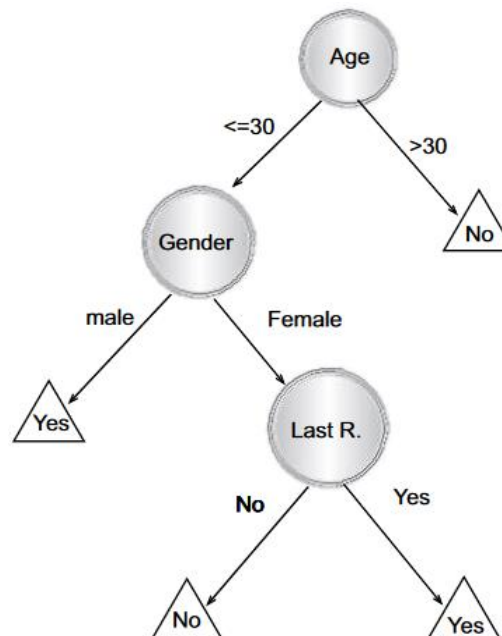
Ένα δέντρο απόφασης είναι ένας ταξινομητής που εκφράζεται ως ένα χώρισμα του συνόλου των πιθανών παραδειγμάτων ή στιγμιότυπων (instance space). Το δέντρο απόφασης αποτελείται από κόμβους που σχηματίζουν ένα ριζωμένο δέντρο, δηλαδή ένα κατευθυνόμενο δέντρο με έναν κόμβο που ονομάζεται "root" το οποίο δεν έχει εισερχόμενες ακμές (incoming edges). Όλοι οι άλλοι κόμβοι έχουν ακριβώς μια εισερχόμενη ακμή. Ένας κόμβος με εξερχόμενες ακμές ονομάζεται εσωτερικός κόμβος ή κόμβος δοκιμής (node). Όλοι οι άλλοι κόμβοι ονομάζονται φύλλα, γνωστοί και ως τερματικοί κόμβοι ή κόμβοι αποφάσεων. Σε ένα δέντρο απόφασης, κάθε εσωτερικός κόμβος χωρίζει τον χώρο πιθανών παραδειγμάτων σε δύο ή περισσότερους χώρους, σύμφωνα με μια συγκεκριμένη διακριτή συνάρτηση των τιμών των χαρακτηριστικών εισόδου. Στην απλούστερη και πιο συνηθισμένη περίπτωση, κάθε συνθήκη εξετάζει ένα μόνο χαρακτηριστικό, έτσι ώστε ο χώρος πιθανών παραδειγμάτων να είναι χωρισμένος σύμφωνα με την τιμή



του χαρακτηριστικού [29]. Στην περίπτωση των συνεχών αριθμητικών χαρακτηριστικών, η συνθήκη παραπέμπει σε ένα εύρος τιμών.

$$\begin{aligned}p(y = 1|x, \mathbf{w}) &= \mu = \sigma(\mathbf{w}^\top \mathbf{x}) = \frac{1}{1 + \exp(-\mathbf{w}^\top \mathbf{x})} = \frac{\exp(\mathbf{w}^\top \mathbf{x})}{1 + \exp(\mathbf{w}^\top \mathbf{x})} \\p(y = 0|x, \mathbf{w}) &= 1 - \mu = 1 - \sigma(\mathbf{w}^\top \mathbf{x}) = \frac{1}{1 + \exp(\mathbf{w}^\top \mathbf{x})} \\p(y|x, \mathbf{w}) &= \frac{1}{1 + \exp(-y\mathbf{w}^\top \mathbf{x})}\end{aligned}$$

Κάθε φύλλο αντιστοιχεί σε μία κατηγορία που αντιπροσωπεύει την καταλληλότερη τιμή-στόχο. Εναλλακτικά, το φύλλο μπορεί να περιέχει έναν διάνυσμα πιθανότητας που υποδεικνύει την πιθανότητα του χαρακτηριστικού-στόχου να έχει μια ορισμένη τιμή. Τα στιγμιότυπα ταξινομούνται πηγαίνοντας από τη ρίζα του δέντρου φτάνοντας σε ένα φύλλο, σύμφωνα με τα αποτελέσματα των συνθηκών κατά μήκος της διαδρομής. Η Εικόνα 2 περιγράφει ένα δέντρο απόφασης που αιτιολογεί εάν ένας πιθανός πελάτης θα απαντήσει ή όχι σε μια άμεση αποστολή αλληλογραφίας, βασισμένο σε ένα συγκεκριμένο σύνολο δεδομένων. Οι εσωτερικοί κόμβοι αντιπροσωπεύονται ως κύκλοι, ενώ τα φύλλα αναπαρίστανται ως τρίγωνα. Αυτό το δέντρο απόφασης περιέχει τόσο κατηγορικά όσο και αριθμητικά χαρακτηριστικά. Με βάση αυτόν τον ταξινομητή, ο αναλυτής μπορεί να προβλέψει την απάντηση ενός δυνητικού πελάτη και να κατανοήσει τα χαρακτηριστικά συμπεριφοράς του πληθυσμού των δυνητικών πελατών όσον αφορά την άμεση αποστολή αλληλογραφίας. Κάθε κόμβος φέρει ετικέτα ανάλογη του χαρακτηριστικού που ελέγχει και τα κλαδιά του φέρουν ετικέτες με τιμές αντίστοιχες αυτών του κόμβου.



Εικόνα 2 Δέντρο απόφασης για το φύλο του πελάτη



Σε περίπτωση αριθμητικών χαρακτηριστικών, τα δέντρα αποφάσεων μπορούν να ερμηνευθούν γεωμετρικά ως μια συλλογή υπερεπιπέδων (hyperplanes), με κάθε ορθογώνιο να αναλογεί σε έναν από τους άξονες [29]. Οι υπεύθυνοι λήψης αποφάσεων προτιμούν λιγότερο σύνθετα δέντρα αποφάσεων, καθώς μπορούν να θεωρηθούν πιο κατανοητά. Επιπλέον, σύμφωνα με τον Breiman et al. (1984) [30] η πολυπλοκότητα των δένδρων έχει κρίσιμη επίδραση στην ακρίβειά τους. Η πολυπλοκότητα των δένδρων ελέγχεται ρητά από τα κριτήρια διακοπής και τη χρησιμοποιούμενη μέθοδο «κλαδέματος». Συνήθως η πολυπλοκότητα του δέντρου μετράται με μια από τις επόμενες μετρικές: τον συνολικό αριθμό κόμβων, τον συνολικό αριθμό των φύλλων, το βάθος των δένδρων και τον αριθμό των χαρακτηριστικών που χρησιμοποιήθηκαν. Η επαγωγή του δέντρου απόφασης σχετίζεται στενά με την επαγωγή του κανόνα. Κάθε διαδρομή από τη ρίζα ενός δέντρου απόφασης σε ένα από τα φύλλα του μπορεί να μετατραπεί σε κανόνα απλώς συνδέοντας τους ελέγχους κατά μήκος της διαδρομής ώστε να σχηματίσουν το προηγούμενο μέρος, θεωρώντας την πρόβλεψη κλάσης του φύλλου ως την τιμή της κλάσης. Για παράδειγμα, ένα από τα μονοπάτια στο προηγούμενο σχήμα μπορεί να μετατραπεί στον κανόνα: «Αν η ηλικία του πελάτη είναι μικρότερη ή ίση ή ίση με 30, και το φύλο του πελάτη είναι «Ανδρας» τότε ο πελάτης θα απαντήσει στην άμεση αλληλογραφία». Το σύνολο κανόνων που προκύπτει μπορεί στη συνέχεια να απλοποιηθεί για να βελτιώσει την κατανόησή του προβλήματος από έναν χρήστη και ενδεχομένως την ακρίβειά του (Quinlan, 1987) [31].

Οι επαγωγείς των δέντρων αποφάσεων είναι αλγόριθμοι οι οποίοι υλοποιούν ένα δέντρο απόφασης από ένα σύνολο δεδομένων. Τυπικά ο στόχος είναι να βρεθεί το βέλτιστο δέντρο απόφασης ελαχιστοποιώντας το σφάλμα γενίκευσης. Ωστόσο κι άλλες συναρτήσεις στόχου μπορούν επίσης να οριστούν, όπως για παράδειγμα η ελαχιστοποίηση του αριθμού κόμβων ή η ελαχιστοποίηση του μέσου βάθους. Η δημιουργία ενός βέλτιστου δέντρου αποφάσεων από ένα σύνολο δεδομένων θεωρείται δύσκολο έργο. Η χρήση βέλτιστων αλγορίθμων δέντρων αποφάσεων είναι εφικτή μόνο σε μικρά προβλήματα. Συνεπώς, απαιτούνται ευρετικοί μέθοδοι για την επίλυση του εκάστοτε προβλήματος. Σε γενικές γραμμές, αυτές οι μέθοδοι μπορούν να χωριστούν σε δύο ομάδες: από πάνω προς τα κάτω και από κάτω προς τα πάνω με σαφή προτίμηση στην πρώτη περίπτωση. Υπάρχουν διάφοροι αλγόριθμοι δέντρων αποφάσεων που υλοποιούν την περίπτωση από πάνω προς τα κάτω όπως ο ID3 (Quinlan, 1986)[32], C4.5 (Quinlan, 1993)[33], CART (Breiman et al., 1984)[34]. Ορισμένοι αποτελούνται από δύο εννοιολογικές φάσεις: ανάπτυξη και κλάδεμα (C4.5 και CART). Αξίζει να σημειωθεί ότι αυτοί οι αλγόριθμοι είναι άπληστοι από τη φύση τους και κατασκευάζουν το δέντρο απόφασης με έναν επαναληπτικό τρόπο από την κορυφή προς τα κάτω (με το γνωστό «διαίρει και βασίλευε»). Σε κάθε επανάληψη, ο αλγόριθμος υλοποιεί το διαχωρισμό του σετ εκπαίδευσης χρησιμοποιώντας το αποτέλεσμα μιας διακριτής συνάρτησης των



χαρακτηριστικών εισόδου. Η επιλογή της καταλληλότερης συνάρτησης γίνεται σύμφωνα με μερικά μέτρα διαίρεσης. Μετά την επιλογή του κατάλληλου διαχωρισμού, κάθε κόμβος υποδιαιρεί περαιτέρω το σύνολο εκπαίδευσης σε μικρότερα υποσύνολα, μέχρις ότου δεν υπάρξει επιπλέον επαρκής διάσπαση ή ικανοποιείται ένα κριτήριο τερματισμού [29].

#### 1.1.5.4 Random Forest Classifier

Τα τυχαία δάση (Random Forests), επίσης γνωστά ως τυχαία δάση αποφάσεων, είναι μια δημοφιλής μέθοδος συνόλων που μπορεί να χρησιμοποιηθεί για την κατασκευή προγνωστικών μοντέλων τόσο για την ταξινόμηση όσο και για προβλήματα παλινδρόμησης. Οι μέθοδοι των συνόλων χρησιμοποιούν πολλαπλά μοντέλα μάθησης για να αποκτήσουν καλύτερα προγνωστικά αποτελέσματα. Στην περίπτωση του τυχαίου δάσους, το μοντέλο δημιουργεί ένα ολόκληρο δάσος τυχαίων μη συσχετισμένων δέντρων αποφάσεων για να επιτύχει την καλύτερη δυνατή απάντηση. Τα τυχαία δάση είναι ένας συνδυασμός δέντρων προβλέψεων έτσι ώστε κάθε δέντρο να εξαρτάται από τις τιμές ενός τυχαίου ανεξάρτητου δειγματοληπτημένου διανύσματος και έχοντας την ίδια κατανομή για όλα τα δέντρα στο δάσος. Το σφάλμα γενίκευσης για τα δάση συγκλίνει σε ένα όριο καθώς ο αριθμός των δέντρων στο δάσος μεγαλώνει. Το σφάλμα γενίκευσης ενός δάσους από δέντρα ταξινομητών εξαρτάται από τη δύναμη των μεμονωμένων δέντρων στο δάσος και τη συσχέτιση που υπάρχει μεταξύ τους. Αξίζει να σημειωθεί ότι ο αλγόριθμος είναι ισχυρός σε προβλήματα θορύβου.

Τα τυχαία δάση εισήχθησαν από τον Leo Breiman (Breiman, 2001) [35]. Τα τυχαία δάση είναι μια επέκταση της ιδέας του Breiman που ονομάζεται bagging και αναπτύχθηκαν ως ανταγωνιστής της μεθόδου boosting. Τα τυχαία δάση μπορούν να χρησιμοποιηθούν είτε για μια μεταβλητή κατηγορικής απόκρισης στην περίπτωση της ταξινόμησης, είτε για μια συνεχή απόκριση στην περίπτωση της παλινδρόμησης. Ομοίως, οι μεταβλητές πρόβλεψης μπορούν να είναι είτε κατηγορικές είτε συνεχείς. Από υπολογιστική άποψη, τα Τυχαία Δάση είναι ελκυστικά επειδή χειρίζονται φυσικά τόσο την παλινδρόμηση όσο και την ταξινόμηση, είναι σχετικά γρήγορα στο να εκπαιδεύουν και να προβλέπουν, εξαρτώνται μόνο από μία ή δύο παραμέτρους tuning, έχουν ως ενσωματωμένο χαρακτηριστικό την εκτίμηση του σφάλματος γενίκευσης, μπορούν να χρησιμοποιηθούν άμεσα για προβλήματα μεγάλης διάστασης και μπορούν εύκολα να υλοποιηθούν παράλληλα.

Στατιστικά, τα τυχαία δάση είναι ελκυστικά λόγω των πρόσθετων χαρακτηριστικών που παρέχουν, όπως μέτρα σημασίας μεταβλητής, διαφορετικό



βάρος κλάσης, απόδοση τιμής σε ελλείπουσες τιμές, οπτικοποίηση, ανίχνευση ακραίων τιμών και μάθηση χωρίς επίβλεψη.

Όπως υποδηλώνει το όνομα, ένα Τυχαίο Δάσος είναι σύνολο βασισμένο σε δέντρα με κάθε δέντρο να εξαρτάται από μια συλλογή τυχαίων μεταβλητών. Πιο τυπικά, για ένα  $p$ -διαστάσεων τυχαίο διάνυσμα  $X = (X_0, X_1, \dots, X_p)^T$  που αντιπροσωπεύει τις πραγματικές μεταβλητές εισόδου ή πρόβλεψης και μια τυχαία μεταβλητή  $Y$  που αντιπροσωπεύει την πραγματική απόκριση, υποθέτουμε μια άγνωστη κοινή κατανομή  $P_{XY}(X, Y)$ . Ο στόχος είναι να βρεθεί μια συνάρτηση πρόβλεψης  $f(X)$  για την πρόβλεψη του  $Y$ . Η συνάρτηση πρόβλεψης καθορίζεται από μια συνάρτηση κόστους  $L(Y, f(X))$  και ορίζεται για να ελαχιστοποιήσει την αναμενόμενη τιμή της απώλειας  $E_{XY}(L(Y, f(X)))$ , όπου οι δείκτες ( $E_{XY}$ ) υποδηλώνουν προσδοκία σε σχέση με την κοινή κατανομή των  $X$  και  $Y$  [36].

Διαισθητικά, το  $L(Y, f(X))$  είναι ένα μέτρο του πόσο κοντά στο  $Y$  είναι το  $f(X)$  και τιμωρεί τις τιμές του  $f(X)$  που είναι πολύ μακριά από το  $Y$ . Οι τυπικές επιλογές του  $L$  είναι η τετραγωνική απώλεια σφάλματος (squared error loss)  $L(Y, f(X)) = Y - f(X)$  για την παλινδρόμηση και η μηδενική απώλεια για την ταξινόμηση [36]:

Φαίνεται ότι η ελαχιστοποίηση του  $E_{XY}(L(Y, f(X)))$  για την τετραγωνική απώλεια σφάλματος καταλήγει στην προσδοκώμενη προϋπόθεση  $f(x) = E(Y|X = x)$ , γνωστή και ως συνάρτηση παλινδρόμησης. Στην ταξινόμηση, αν το σύνολο των πιθανών τιμών του  $Y$  υποδηλώνεται με  $\mathcal{Y}$ , ελαχιστοποιώντας την  $E_{XY}(L(Y, f(X)))$  για μηδενική απώλεια δίνει τον τύπο [36]

Γνωστός και ως ο κανόνας του Bayes.

Τα τυχαία δάση κατασκευάζουν την συνάρτηση  $f$  από την σκοπιά μιας συλλογής αποκαλούμενων «ταξινομητών βάσης» (base learners)  $h_0(x), h_1(x), \dots, h_j(x)$ , και αυτοί οι ταξινομητές βάσης συνδυάζονται για να δώσουν τον εκτιμητή συνόλου  $f(x)$ . Στην παλινδρόμηση, οι βασικοί μαθητές υπολογίζονται κατά μέσο όρο [36]

$$L(Y, f(X)) = I(Y \neq f(X)) = \begin{cases} 0 & \text{if } Y = f(X) \\ 1 & \text{otherwise.} \end{cases}$$

$$f(x) = \arg \max_{y \in \mathcal{Y}} P(Y = y | X = x)$$



Ενώ στην ταξινόμηση η  $f(x)$  είναι η πιο συχνά προβλεπόμενη κλάση. Στα Τυχαία Δάση ο βασικός μαθητής είναι ένα δέντρο που υποδηλώνεται ως  $h_j(X, \theta_j)$ , όπου  $\theta_j$  είναι μια συλλογή τυχαίων ανεξάρτητων μεταβλητών για  $j = 1, \dots, J$  [36].

Όπως αναφέρθηκε προηγουμένως σε αυτή την ενότητα, ένα Τυχαίο Δάσος χρησιμοποιεί τα δέντρα  $h_j(X, \theta_j)$  ως ταξινομητές βάσης. Για δεδομένα εκπαίδευσης  $D = \{(x_0, y_0), (x_1, y_1), \dots, (x_N, y_N)\}$ , όπου το  $x_i = (x_{i,0}, x_{i,1}, \dots, x_{i,p})^T$  υποδηλώνει τους προγνώστες  $p$  και το  $y_i$  δηλώνει την απόκριση και μια συγκεκριμένη υλοποίηση  $\theta_j$  του  $\theta_j$ , το προσαρμοσμένο (fitted) δένδρο συμβολίζεται με  $\hat{h}_j(x, \theta_j, D)$ . Αν και αυτή είναι η αρχική διατύπωση από τον Breiman, στην πράξη η τυχαία  $\theta_j$  χρησιμοποιείται έμμεσα για να εισάγει την τυχαιότητα (randomness) με δύο τρόπους. Πρώτον, με την μέθοδο bagging, όπου το κάθε δέντρο απόφασης (ταξινομητής βάσης) είναι προσαρμοσμένο σε ένα ανεξάρτητο δείγμα εκκίνησης (bootstrap) από τα αρχικά δεδομένα. Η τυχαιοποίηση που περιλαμβάνεται στη δειγματοληψία bootstrap αποτυπώνει ένα μέρος του  $\theta_j$ . Δεύτερον, κατά τον διαχωρισμό ενός κόμβου, ο καλύτερος διαχωρισμός βρίσκεται πάνω σε ένα τυχαία επιλεγμένο υποσύνολο μεταβλητών πρόβλεψης  $m$  αντί όλων των προγνωστών  $p$ , ανεξάρτητα σε κάθε κόμβο. Η τυχαιοποίηση που χρησιμοποιείται για τη δειγματοληψία των προγνωστών αποτυπώνει το υπόλοιπο τμήμα του  $\theta_j$  [36].

Αρχικά, ο Breiman πρότεινε την ανάπτυξη των δέντρων απόφασης έως ότου οι τερματικοί κόμβοι ήταν καθαροί (περίπτωση ταξινόμησης) ή μέχρις ότου υπήρχαν λιγότεροι κόμβοι από έναν προκαθορισμένο αριθμό σημείων δεδομένων σε κάθε τερματικό κόμβο (περίπτωση παλινδρόμησης). Τελευταία προτείνει να γίνεται έλεγχος του μέγιστου αριθμού τερματικών κόμβων. Τα δένδρα που προκύπτουν συνδυάζονται με τη μη σταθμισμένη (unweighted) ψηφοφορία στην περίπτωση όπου η απάντηση είναι κατηγορηματική (ταξινόμηση) ή με τον μη σταθμισμένο μέσο όρο εάν η απόκριση είναι συνεχής (παλινδρόμηση).

### 1.1.5.5 Αλγόριθμος Gradient Boosting

Το Gradient boosting είναι μια τεχνική μηχανικής μάθησης για προβλήματα παλινδρόμησης και ταξινόμησης, η οποία παράγει ένα μοντέλο πρόβλεψης με τη μορφή ενός συνόλου αδύναμων μοντέλων πρόβλεψης, συνήθως δέντρων αποφάσεων. Χτίζει το μοντέλο με ένα σοφό τρόπο όπως και άλλες μέθοδοι boosting, και το γενικεύει επιτρέποντας τη βελτιστοποίηση μιας αυθαίρετης διαφοροποιήσιμης συνάρτησης κόστους [37]. Η λογική του πλησιάζει αυτή των τυχαίων δασών.

Η ιδέα του Gradient boosting προέκυψε από την παρατήρηση του Leo Breiman ότι το boosting μπορεί να ερμηνευθεί ως ένας αλγόριθμος βελτιστοποίησης σε μια κατάλληλη συνάρτηση κόστους. Οι gradient boosting αλγόριθμοι



παλινδρόμησης αναπτύχθηκαν στη συνέχεια από τον Jerome H. Friedman, ταυτόχρονα με την γενικότερη λειτουργική θεώρηση του gradient boosting που ειπώθηκε από τους Llew Mason, Jonathan Baxter, Peter Bartlett και Marcus Frean. Τα τελευταία δύο έγγραφα εισήγαγαν την άποψη των αλγορίθμων ενίσχυσης ως επαναληπτικοί λειτουργικοί αλγόριθμοι gradient descent. Δηλαδή, οι αλγόριθμοι που βελτιστοποιούν μια συνάρτηση κόστους σε σχέση με το χώρο της συνάρτησης επιλέγουν με επαναληπτικό τρόπο μία αδύναμη υπόθεση που δείχνει την αρνητική gradient κατεύθυνση. Αυτή η λειτουργική όψη της gradient μεθόδου, στο πλαίσιο του boosting, έχει οδηγήσει στην ανάπτυξη boosting αλγορίθμων σε πολλούς τομείς της μηχανικής μάθησης και των στατιστικών πέρα από την παλινδρόμηση και την ταξινόμηση [37].

Οι μηχανές Gradient Boosting (Gradient Boosting Machines) είναι μια ισχυρή μέθοδος που μπορεί να καταγράψει αποτελεσματικά πολύπλοκες εξαρτήσεις μη γραμμικών λειτουργιών. Αυτή η οικογένεια μοντέλων έχει δείξει σημαντική επιτυχία σε διάφορες πρακτικές εφαρμογές. Επιπλέον, τα GBMs είναι εξαιρετικά ευέλικτα και μπορούν εύκολα να προσαρμοστούν στις διαφορετικές πρακτικές ανάγκες. Ωστόσο, όλα αυτά τα αποτελέσματα και τα οφέλη έχουν και μειονεκτήματα που αφορούν κυρίως τους πόρους που χρειάζονται [38]. Το πιο αξιοσημείωτο πρόβλημα των GBM που αναδύεται στην πράξη είναι η κατανάλωση μνήμης. Το κόστος αποθήκευσης ενός μοντέλου πρόβλεψης εξαρτάται από τον αριθμό των boosting επαναλήψεων που χρησιμοποιούνται για μάθηση. Για να μειώσουμε τις επιπτώσεις του over-fitting, ο βέλτιστος αριθμός επαναλήψεων μπορεί να είναι αρκετά μεγάλος. Σε ορισμένες εφαρμογές που απαιτούν μεγάλη ακρίβεια, όπως τα συστήματα ανίχνευσης εισβολών, ο επιθυμητός αριθμός επαναλήψεων μπορεί εύκολα να είναι της τάξης των δεκάδων χιλιάδων. Ο χειρισμός τέτοιων τεράστιων μοντέλων απαιτεί την αποθήκευση όλων των παραμέτρων καθενός από τους fitted βασικούς μαθητευόμενους. Το πρόβλημα αυτό μπορεί να παρακαμφθεί εν μέρει με την εκτεταμένη χρήση αραιών βασικών μαθητευόμενων ή με τις μεθόδους της απλοποίησης του συνόλου (Kulkarni and Sinha, 2012 [39]). Ωστόσο, αυτό το πρόβλημα με την κατανάλωση μνήμης είναι κοινό σε όλες τις μεθόδους του συνόλου και εμφανίζεται πιο έντονα στον αυξημένο αριθμό μοντέλων που κάποιος επιλέγει να αποθηκεύσει. Ένα άλλο πρόβλημα των GBM είναι η ταχύτητα αξιολόγησης. Για να χρησιμοποιήσουμε το «εκπαιδευμένο» GBM μοντέλο ώστε να λαμβάνει τις προβλέψεις, πρέπει να αξιολογήσουμε όλους τους βασικούς μαθητευόμενους αλγόριθμους στο σύνολο. Παρά την απλότητα του καθενός από τους βασικούς μαθητευόμενους, όταν το σύνολο είναι αρκετά μεγάλο, η απόκτηση προβλέψεων με γρήγορο ρυθμό μπορεί να γίνει χρονοβόρα. Επομένως, η χρήση GBM σε εντατικές online εργασίες θα απαιτούσε κατά πάσα πιθανότητα από τον επαγγελματία να δεχτεί ένα συμβιβασμό μεταξύ της πολυπλοκότητας του μοντέλου, και του επιθυμητού αριθμού αξιολογήσεων της λειτουργίας ανά χρονικό διάστημα [38].



Τα προαναφερθέντα προβλήματα είναι καθαρά υπολογιστικά και μπορούν να θεωρηθούν ως το κόστος χρήσης ενός ισχυρού μοντέλου μηχανικής μάθησης. Όπως ήδη περιγράψαμε, τα GBMs είναι εξαιρετικά εφαρμόσιμα, παρέχοντας χρήσιμες ιδιότητες στους χρήστες. Επιπλέον, επιτρέπουν τη σχετικά εύκολη ερμηνεία των αποτελεσμάτων, παρέχοντας έτσι στον χρήστη πληροφορίες για το προσαρμοσμένο μοντέλο κάτι που αυξάνει την εμπιστοσύνη των χρηστών και κατά επέκταση την αποδοχή τους.

## 1.2 Εξόρυξη δεδομένων χρονοσειράς

Σε αυτή την ενότητα θα εξεταστούμε τα χαρακτηριστικά των χρονοσειρών που αποτελούν το βασικό κομμάτι τις έρευνας μας. Όσον αφορά τα δεδομένα χρονοσειρών, οι πιο συνηθισμένες εργασίες επεξεργασίας είναι, μεταξύ άλλων, η ομαδοποίηση, η ταξινόμηση, η πρόβλεψη, η ανίχνευση ανωμαλιών, η αναγνώριση μοτίβων και η ανακάλυψη κανόνων. Η ύπαρξη αυτών των ζητημάτων έχει τεκμηριωθεί εκτενώς στα πεδία της αναγνώρισης προτύπων και της εξόρυξης δεδομένων εδώ και πολλά χρόνια. Για διάφορους λόγους, ωστόσο, οι καθιερωμένες προσεγγίσεις αναγνώρισης προτύπων και εξόρυξης δεδομένων είναι αναποτελεσματικές για την επεξεργασία δεδομένων χρονοσειρών. Στην περίπτωση αυτή, μια ιδανική αναπαράσταση χρονοσειρών που επιτυγχάνει μείωση διαστάσεων θα πρέπει επίσης να επιτρέπει στους χρήστες να πραγματοποιούν οποιαδήποτε από τις ενέργειες που αναφέρονται παραπάνω απευθείας από τις αναπαραστάσεις τους. Στη συνέχεια, συζητούνται οι πιο συνηθισμένες εργασίες χρονοσειρών, παρέχοντας αντιπροσωπευτικά παραδείγματα εφαρμογών.

Μια πολύ δημοφιλής εργασία χρονοσειρών είναι η ομαδοποίηση. Έστω  $N$  χρονοσειρές χωρίς ετικέτες που συμβολίζονται ως  $T_1 T_2 \dots T_N$ . Όταν ένα σύνολο διαιρείται σε ομάδες  $K$  χρησιμοποιώντας ένα σημαντικό μέτρο ομοιότητας, η διαδικασία είναι γνωστή ως ομαδοποίηση. Τα μέλη μιας ομάδας είναι παρόμοια μεταξύ τους και τα μέλη των άλλων ομάδων είναι σημαντικά διαφορετικά μεταξύ τους. Οι περισσότερες τεχνικές για την ομαδοποίηση δεδομένων χρονοσειρών είναι εκδόσεις αλγορίθμων που χρησιμοποιούνται αρχικά για τη ομαδοποίηση διανυσματικών δεδομένων χαμηλής διάστασης. Ένας μικρός αριθμός χαρακτηριστικών υπολογίζεται για την αναπαράσταση κάθε χρονοσειράς στις προσεγγίσεις που βασίζονται σε χαρακτηριστικά και, στη συνέχεια, αλγόριθμοι ομαδοποίησης, όπως αλγόριθμοι *k-means*, χρησιμοποιούνται για τη ομαδοποίηση των διανυσμάτων χαρακτηριστικών. Οι μέθοδοι που βασίζονται σε μοντέλα εξάγουν ένα σύνολο παραμέτρων για κάθε χρονοσειρά και στη συνέχεια βρίσκουν συστάδες με βάση τις παραμέτρους που χρησιμοποιούνται στη διαδικασία ομαδοποίησης. Είναι σπάνιο να δούμε προσεγγίσεις που βασίζονται σε ακατέργαστα δεδομένα να χρησιμοποιούνται στην ανάλυση χρονοσειρών λόγω της μεγάλης διάστασης των δεδομένων.





Ταξινόμηση είναι η διαδικασία κατανομής μιας χρονοσειράς εισόδου σε μία από τις διάφορες αναγνωρισμένες κλάσεις ή κατηγορίες με βάση τα χαρακτηριστικά της. Όλοι οι αλγόριθμοι ταξινόμησης μαθαίνουν κανόνες ταξινόμησης από δείγματα εκπαίδευσης των οποίων η ταξινόμηση είναι γνωστή πριν από την εκπαίδευση. Η χρήση της θεωρίας απόφασης Bayes για την ταξινόμηση χρονοσειρών δεν είναι πρακτική επειδή ο υπολογισμός των συναρτήσεων πυκνότητας πιθανότητας για χρονοσειρές με υψηλό βαθμό διαστάσεων δεν είναι πρακτικός. Ο αριθμός των δειγμάτων που απαιτούνται για γραμμικούς ταξινομητές όπως το perceptron, οι προσεγγίσεις του ελαχίστου μέσου τετραγώνου και οι μηχανές διανύσματος υποστήριξης είναι τουλάχιστον διπλάσιος από τη διάσταση της χρονοσειράς. Η εκπαίδευση ταξινομητών με μοτίβα τέτοιας μεγάλης διάστασης είναι υπολογιστικά απαιτητική, ακόμη και όταν είναι διαθέσιμα τα απαραίτητα δεδομένα εκπαίδευσης. Η χρήση μη γραμμικών ταξινομητών όπως τα τεχνητά νευρωνικά δίκτυα καθίσταται πολύ ακριβή λόγω του τεράστιου μεγέθους του δικτύου και του μεγάλου αριθμού δειγμάτων εκπαίδευσης που πρέπει να χρησιμοποιηθούν. Ως αποτέλεσμα, η ταξινόμηση με βάση τα χαρακτηριστικά φαίνεται να είναι η μόνη εφικτή μέθοδος.

Αν υποθέσουμε μια χρονοσειρά με  $N$  τιμές που συμβολίζονται ως  $T_1 T_2 \dots T_N$ , τότε η πρόβλεψη είναι η εκτίμηση των μελλοντικών τιμών του  $T$  για  $i > n$ . Οι μελλοντικές τιμές εκτιμούνται έχοντας ως βάση την τρέχουσα εξέλιξη που παρατηρείται στις χρονοσειρές (βάση στατιστικών δεδομένων) ή σε μαθηματικά μοντέλα όπως το Hidden Markov Model που κατασκευάστηκε από δεδομένα ιστορικών χρονοσειρών που είναι παρόμοια με τις τρέχουσες χρονοσειρές και χρησιμοποιούνται για την πρόβλεψη μελλοντικών τιμών. Συνήθως, το μοντέλο βασίζεται σε αξιοσημείωτα χαρακτηριστικά χρονοσειρών. Τέλος, δεδομένου ότι η σε πολλές περιπτώσεις η εργασία πρόβλεψης πολλαπλών βημάτων είναι το ζητούμενο, θα πρέπει να αναφερθεί ότι για την πρόβλεψη χρονοσειρών, ο σκοπός της χρήσης παρατηρούμενων χρονοσειρών στο παρελθόν για την πρόβλεψη τιμών στον μελλοντικό ορίζοντα γίνεται αναλογικά πιο δύσκολος καθώς αυτός ο ορίζοντας διευρύνεται [85]. Επί του παρόντος, οι εργασίες πρόβλεψης πολλών μελλοντικών χρονικών στιγμών επιτυγχάνονται με δύο διαφορετικούς τρόπους. Ο πρώτος, που ονομάζεται ανεξάρτητη πρόβλεψη, αποτελείται από την εκπαίδευση ενός άμεσου μοντέλου για την πρόβλεψη των τιμών αρκετών, μελλοντικών στιγμών. Η δεύτερη στρατηγική, που ονομάζεται επαναληπτική μέθοδος, αποτελείται από την επανάληψη των προβλέψεων ενός βήματος τόσες φορές έως τον επιθυμητό χρονικό ορίζοντα. Η επαναληπτική πρόβλεψη χρησιμοποιεί μόνο ένα μοντέλο για να προβλέψει όλες τις μελλοντικές τιμές. Ο στόχος είναι να αναλύσουμε μια σύντομη ακολουθία δεδομένων και να προσπαθήσουμε να προβλέψουμε την υπόλοιπη ακολουθία δεδομένων μέχρι να φτάσουμε σε ένα προκαθορισμένο χρονικό βήμα. Το κύριο μειονέκτημα αυτής της προσέγγισης είναι η συγκεκριμένη φύση του σφάλματος. Η τελευταία προσέγγιση αξιοποιήθηκε για το προτεινόμενο μοντέλο



που βασίζεται σε RL, ενώ τα χρησιμοποιούμενα πιθανοτικά μοντέλα χρονοσειρών DNN ακολούθησαν τη μέθοδο πρόβλεψης ανεξάρτητης τιμής.

Το μοτίβο σε μια χρονοσειρά ή μια ομάδα χρονοσειρών εάν υπάρχει μια περίπτωση επαναλαμβανόμενη υπό-ακολουθία. Για παράδειγμα, μοτίβα χρησιμοποιούνται στην ανάλυση των δεδομένων της χρηματιστηριακής αγοράς για την ανίχνευση των «παρόμοιων» συμπεριφορών. Τα μοτίβα έχουν χρησιμοποιηθεί ευρέως για την ανάπτυξη κανόνων, την ομαδοποίηση και την κατηγοριοποίηση δεδομένων χρονοσειρών, μεταξύ άλλων εφαρμογών. Ενώ είναι δύσκολο να γίνει διάκριση μεταξύ μοτίβων διαφορετικών χρονικών πλαισίων ακόμη και όταν μοιράζονται τις ίδιες ή πολύ παρόμοιες συνολικές ιδιότητες.

Η ανακάλυψη κανόνων είναι η διαδικασία ανακάλυψης χρονικών κανόνων που είναι κρυμμένοι ή δεν είναι ορατοί στα δεδομένα χρονοσειρών. Μια τεχνική είναι να μετατρέψετε μια χρονοσειρά σε μια ακολουθία συμβόλων και στη συνέχεια να χρησιμοποιήσετε αλγόριθμους εξόρυξης κανόνων συσχέτισης για να αποκαλύψετε τους κανόνες που διέπουν τη σειρά, η εν λόγω μέθοδος βρίσκει εφαρμογή για την αποκάλυψη αρχών συναλλαγών. Μια άλλη στρατηγική είναι να μετατρέψετε μια χρονοσειρά σε μια ακολουθία γεγονότων και στη συνέχεια να χρησιμοποιήσετε δέντρα ταξινόμησης για να μάθετε τους κανόνες που διέπουν την ακολουθία γεγονότων.

### **1.3 Εφαρμογές Μηχανικής Μάθησης σε βιομηχανικά περιβάλλοντα με πραγματικά δεδομένα**

Στην ενότητα αυτή θα περιγράψουμε ορισμένα προβλήματα τριών κλάδων της βιομηχανίας που η λύση τους μπορεί να δοθεί μέσω της ανάλυσης δεδομένων και της χρήσης μεθόδων μηχανικής μάθησης. Οι εν λόγω περιπτώσεις αποτελέσαν τα πειραματικά περιβάλλοντα για τα αποτελέσματα της ανά χειράς διατριβής. Συγκεκριμένα θα εξεταστούν συγκεκριμένα προβλήματα της εμπορικής ναυτιλίας όπου θα εξεταστεί η διαδικασία προβλεπτικής συντήρησης, της ασφάλειας τροφίμων όπου θα εξεταστούν δύο περιπτώσεις και τέλος ο τραπεζικός κλάδος όπου προτείνουμε ψηφιακές υπηρεσίες για Μικρομεσαίες επιχειρήσεις.

#### **1.3.1 Μηχανική Μάθηση & προβλεπτική συντήρηση κύριων μηχανών εμπορικών πλοίων.**

Τα τελευταία χρόνια, γινόμαστε μάρτυρες της ψηφιοποίησης των βιομηχανικών περιβαλλόντων ως μέσω για την μετάβαση στην τέταρτη βιομηχανική επανάσταση -Industry 4.0. Βασισμένη στα Φυσικά Συστήματα Κυβερνοχώρου (Cyber Physical Systems, CPS) και τις ψηφιακές τεχνολογίες όπως τα Υπολογιστικά Νέφη, το Βιομηχανικό Διαδίκτυο των Πραγμάτων και την Τεχνητή Νοημοσύνη,



υποστηρίζονται καινοτόμες λειτουργίες όπως η μαζική εξειδίκευση, η προβλεπτική συντήρηση, η εκμηδένιση των αποβλήτων και η δημιουργία ψηφιακών διδύμων. Η τεχνητή νοημοσύνη είναι αυτή που διευκολύνει περισσότερο αυτή την ψηφιακή οδό, και εισάγει καινούργιες περιπτώσεις, συμπεριλαμβάνοντας την προβλεπτική συντήρηση ποιότητας (Quality 4.0) [40]. Οι μεγάλες δυνατότητες της τεχνητής νοημοσύνης οδηγούνται από τις εξελίξεις και την πρόοδο της τεχνολογίας των συστημάτων υλισμικού και λογισμικού, που έχουν επιτρέψει την ανάπτυξη προηγούμενων σχημάτων μηχανικής μάθησης [41] και καινοτόμων αλγορίθμων [42] κατάλληλων για προβλήματα ευρείας κλίμακας σε ρεαλιστικές συνθήκες. Παρ' όλη αυτή την πρόοδο, η κατάσταση στην τεχνολογία της τεχνητής νοημοσύνης στον βιομηχανικό τομέα, απέχει από το να εφαρμοστούν οι πιο προηγμένες δυνατότητες της βαθιάς μηχανικής μάθησης. Αντ' αυτού, κυρίως εφαρμόζονται στην ενοποίηση του συνόλου των δεδομένων από ετερογενείς πηγές παρά διευκολύνουν προηγούμενες αναλύσεις (όπως η βαθιά μάθηση) για να χρησιμοποιηθούν σε περιπτώσεις όπως η προβλεπτική συντήρηση και οι βιομηχανικές προσομοιώσεις.

Στον τομέα της εμπορικής ναυτιλίας, στις μέρες μας, η κίνηση των φορητών με μικρή ταχύτητα (*slow steaming*) είναι η επικρατούσα στρατηγική λειτουργίας για ένα μεγάλο μέρος των ναυτιλιακών εταιρειών ώστε αυτές να παραμένουν προσοδοφόρες. Η κίνηση με μικρή ταχύτητα συνήθως συνεπάγεται μειώσεις της ταχύτητας των πλοίων και έχει εισέλθει ώστε να μετριάσει τις επιπτώσεις από την πρόσφατη οικονομική κρίση. Πέρα από τη σημαντική μείωση της κατανάλωσης καυσίμου, έχει ως αποτέλεσμα και την λιγότερο αποδοτική λίπανση του κινητήρα. Ως εκ τούτου ένα κοινό πρόβλημα που δημιουργείται είναι η φθορά του ρουλεμάν σταυρού κεφαλής [43]. Ένα πρόβλημα που ιδιαίτερης δυσκολίας να επιλυθεί καθώς λόγω της απουσίας κατάλληλων αισθητήρων όπως οι αισθητήρες κραδασμών της κύριας μηχανής στην περίπτωση των πλοίων. Ένα άλλο πρόβλημα ότι τα παραδοσιακά συστήματα παρακολούθησης δεν είναι ικανά να εντοπίσουν πολύπλοκα ελαττώματα, για παράδειγμα την υπό εξέταση διάβρωση που σταδιακά εξελίσσεται δεν δίνει εμφανή σημάδια μέχρι το ρουλεμάν σταυρού κεφαλής σπάσει. Ακόμα, και δεδομένου ότι το κόστος επιθεώρησης σε πολλές περιπτώσεις είναι υψηλό, αυτό το σενάριο είναι κατάλληλο ώστε να υποδείξει την προστιθέμενη αξία μίας στρατηγικής προβλεπτικής συντήρησης. Αυτή χρησιμοποιεί μια εκτίμηση της παρούσας και μελλοντικής κατάστασης ενός εξαρτήματος ώστε να προβλέπει την βέλτιστη συντήρηση, σχεδιάζοντας την αποφυγή των ζημιών, χωρίς να οδηγήσει σε υπερβολική συντήρηση [44].

Υπάρχουν δυο κυρίαρχες προσεγγίσεις που χρησιμοποιούν τη μηχανική μάθηση ώστε να εκτιμηθεί η παρούσα κατάσταση του μηχανικού εξοπλισμού: η πρώτη αναφέρεται στην πρόβλεψη του χρόνου μεταξύ σφαλμάτων του εξοπλισμού, ενώ η δεύτερη αναφέρεται στον εντοπισμό ανωμαλιών. Το σημαντικό συστατικό του εντοπισμού ανωμαλιών μπορεί να δοθεί περιληπτικά ως εξής: «ο εντοπισμός ανωμαλιών βασίζεται στο θεμελιώδη συστατικό μοντελοποίησης: τι είναι



φυσιολογικό, ώστε να ανακαλύψει ό,τι δεν είναι» [45]. Ο εντοπισμός ανωμαλιών προσφέρει πληθώρα ορισμών των τύπων ανωμαλιών αλλά σχετικά με δεδομένα χρονοσειρών είναι χρήσιμο να θεωρηθούν τρεις κατηγορίες ανωμαλιών, οι ανωμαλίες σημείου, περιεχομένου και οι συλλογικές [46]. Οι ανωμαλίες σημείου είναι μεμονωμένες τιμές που εμπίπτουν σε περιοχές τιμών χαμηλής πυκνότητας, οι συλλογικές ανωμαλίες υποδεικνύουν ότι μια ακολουθία τιμών δεν ακολουθεί την κατανομή της χρονοσειράς και οι ανωμαλίες περιεχομένου είναι τιμές που δεν εμπίπτουν σε περιοχές χαμηλής πυκνότητας αλλά είναι συγκρινόμενες με τις τοπικές τιμές.

Υπάρχουν περιπτώσεις όπου η πληροφορία που συλλέγεται από αισθητήρες (πχ. δεδομένα θερμοκρασίας, πίεσης μηχανών, άλλα και δεδομένα που υποδηλώνουν την κατάσταση του πλοίου όπως η κατανάλωση ισχύος και το φορτίο του) είναι αρκετά μεγάλη συγκριτικά με τις πληροφορίες για τον αριθμό των βλαβών, πόσο μάλλον αν εξεταστεί ως προς μία συγκεκριμένη βλάβη. Στις περιπτώσεις αυτές, η ουσιαστική πληροφορία για τις βλάβες μας οδηγεί να χειριστούμε μια τέτοια περίπτωση ως περίπτωση εντοπισμού ανωμαλίας. Συγκεκριμένα η προσέγγιση που επιστρατεύεται στην διατριβή αυτή περιλαμβάνει την περιγραφή μιας κατάλληλης τεχνικής προεπεξεργασίας για το σύνολο των δεδομένων που συλλέγονται, η ανάπτυξη ενός μοντέλου μηχανικής μάθησης που μπορεί να εκτιμήσει αν ένα σημείο δεδομένων που βασίζεται σε χρονοσειρά αντιστοιχεί σε μια κατάσταση αναφοράς (φυσιολογικής λειτουργίας) ή όχι (στην περίπτωση όχι φυσιολογικής συμπεριφοράς) που θα επιφέρει πιθανή διάβρωση του ρουλεμάν σταυρού κεφαλής σε αρχικό επίπεδο.

### **1.3.2 Μηχανική Μάθηση & ασφάλεια των τροφίμων.**

Η ασφάλεια των τροφίμων έχει συγκεντρώσει το ενδιαφέρον του κοινού και την προσοχή του Τύπου τα τελευταία χρόνια. Παγκόσμιες ενώσεις όπως ο FDA («Διαχείριση Τροφίμων και Φαρμάκων») ή ο ΠΟΥ («Παγκόσμιος Οργανισμός Υγείας») θέτουν πολύ αυστηρούς κανονισμούς σε όλη την αλυσίδα εφοδιασμού προκειμένου να παρακολουθούν και να αξιολογούν τους κίνδυνους στην ασφάλεια των τροφίμων (FAO/WHO, 2007). Αυτό είναι λογικό, δεδομένου του γεγονότος ότι πολλά προϊόντα που βρίσκονται στην αγορά τροφίμων μπορεί να είναι επιβλαβή όχι μόνο για τα άτομα αλλά και για την οικονομία και την δημόσια υγεία γενικότερα [47] [48].

Μία από τις πιο αποτελεσματικές μεθόδους για την ασφάλεια των τροφίμων είναι η απόσυρση των τροφίμων από την αγορά, όταν έχουν χαρακτηριστεί επιβλαβή, σε οποιοδήποτε στάδιο της εφοδιαστικής αλυσίδας, συμπεριλαμβανομένων και αυτών που κατέχουν οι καταναλωτές. Αυτή η διαδικασία είναι ευρέως γνωστή ως ανάκληση τροφίμων. Υπάρχουν διάφοροι λόγοι που



μπορεί να οδηγήσουν στον χαρακτηρισμό ενός προϊόντος ως μη ασφαλές ή επιβλαβές, με τον πιο διαδεδομένο από αυτούς να είναι η ύπαρξη αδήλωτων αλλεργιογόνων [49] ή η μόλυνση συγκεκριμένων κατηγοριών τροφίμων, όπως το κρέας και τα πουλερικά, όπου υπάρχουν επιβλαβή βακτήρια (*Listeria*, *Salmonella* και *Escherichia coli*) [50].

Αυτή τη στιγμή, η διαδικασία αξιολόγησης της ποιότητας των τροφίμων είναι μια κουραστική και χρονοβόρα εργασία λόγω της έλλειψης ομοιογένειας στις ανακοινώσεις ανάκλησης από τους διάφορους οργανισμούς που σχετίζονται με την ασφάλεια των τροφίμων παγκοσμίως. Φυσικά, ο χρόνος είναι ουσιαστικός σε αυτήν την περίπτωση, δεδομένου ότι τα προϊόντα αυτά μπορούν να καταναλωθούν κατά τη διάρκεια αυτής της διαδικασίας [51]. Ως εκ τούτου, είναι ζωτικής σημασίας να υπάρχει ένα αξιόπιστο σύστημα για την έγκαιρη ανίχνευση ή ακόμα και την πρόβλεψη των μη ασφαλών τροφίμων, προκειμένου να αποφευχθούν τα κρούσματα και η σοβαρή βλάβη στο κοινό. Αυτή είναι μια πρόκληση υψίστης σημασίας, στην οποία θα πρέπει να επικεντρωθεί η ερευνητική κοινότητα που σχετίζεται με την ασφάλεια των τροφίμων, καθώς κοστίζει δισεκατομμύρια στην παγκόσμια οικονομία [52] [48]. Συνοπτικά, το πρόβλημα της ασφάλειας των τροφίμων είναι διαρκώς μεταβαλλόμενο καθώς εμφανίζονται νέες απειλές ετησίως που δοκιμάζουν τους τρέχοντες τρόπους διαχείρισης των τροφίμων. Εκτός από μια τεράστια πρόκληση, επηρεάζει σε μεγάλο βαθμό τις ανθρώπινες ζωές. Για παράδειγμα, σύμφωνα με τον ΠΟΥ, τα μη ασφαλή τρόφιμα προκαλούν περισσότερες από 200 ασθένειες – που μπορεί να είναι από διάρροιες έως και καρκίνοι. Υπολογίζεται ότι σχεδόν 1 στους 10 ανθρώπους στον κόσμο αρρωσταίνουν μετά από κατανάλωση μολυσμένων τροφίμων και 420.000 πεθαίνουν κάθε χρόνο, με αποτέλεσμα την απώλεια 33 εκατομμυρίων ετών υγιών ανθρώπων. Επιπλέον, το χρηματικό κόστος είναι υψηλό. Περίπου 110 δισεκατομμύρια δολάρια χάνονται κάθε χρόνο σε παραγωγικότητα και ιατρικές δαπάνες που προκύπτουν από μη ασφαλή τρόφιμα σε χώρες χαμηλού και μεσαίου εισοδήματος. Επιπλέον, σύμφωνα με το Υπουργείο Γεωργίας των Ηνωμένων Πολιτειών, οι τροφιμογενείς ασθένειες κοστίζουν στην οικονομία των ΗΠΑ 10-83 δισεκατομμύρια δολάρια ετησίως [47].

Σε αυτήν την περίπτωση, για να αποκτήσουμε με ακρίβεια τις γνώσεις που σχετίζονται με τον τομέα που εξάγονται από ανακλήσεις τροφίμων, απαιτείται η ανάλυση κειμένου για εργασίες Επεξεργασίας Φυσικής Γλώσσας. Η μηχανική μάθηση και η βαθιά μάθηση έχει αποδείξει αποτελέσματα σε διάφορες εργασίες επεξεργασίας φυσικής γλώσσας (NLP), όπως ενσωματώσεις λέξεων [53] [54], επισήμανση μέρους του λόγου [55], ανάλυση κειμένου [56] [57], αναγνώριση ονομασίας οντότητας (Named Entity Recognition -NER) [58]. Συγκεκριμένα η πρόκληση που αντιμετωπίζουμε σε αυτή την περίπτωση των ανακλήσεων τροφίμων είναι να κατηγοριοποιήσουμε τις ανακοινώσεις ανακλήσεων βάσει του υποκειμένου προϊόντος (τρόφιμου). Λαμβάνοντας υπόψη ότι οι δεν υπάρχει κανένα



προ-εκπαιδευμένο μοντέλο για εξαγωγή οντοτήτων προϊόντων τροφίμων, κάτι που να καθιστά απαραίτητη την εκπαίδευση ενός νέου μοντέλου. Δεδομένης της φύσης του συνόλου δεδομένων που σχετίζεται με τα τρόφιμα σχετικά με την εργασία ταξινόμησης κειμένου, εφαρμόστηκε η μεθοδολογία NER, θεωρώντας τα προϊόντα ως οντότητες. Γενικά, μια οντότητα με συγκεκριμένο όνομα είναι μια λέξη ή μια φράση που προσδιορίζει ξεκάθαρα ένα στοιχείο από ένα σύνολο άλλων στοιχείων που έχουν παρόμοια χαρακτηριστικά. Παραδείγματα κατονομασμένων οντοτήτων είναι ονόματα οργανισμών, ατόμων και τοποθεσιών σε γενικό τομέα, ονόματα γονιδίων, πρωτεϊνών, φαρμάκων και ασθενειών στον βιοϊατρικό τομέα [59]. Εν συνεχεία, έχοντας ταξινομήσει της ανακλήσεις τροφίμων δημιουργούνται οι χρονοσειρές από τις οποίες θα προκύψει η πρόβλεψη πιθανών ανακλήσεων τροφίμων, με ζητούμενο τη μείωση της μελλοντικής αβεβαιότητας. Οι μέθοδοι για την πρόβλεψη χρονοσειρών βασίζονται σε ιστορικά δεδομένα, καθώς περιλαμβάνουν εσωτερικά μοτίβα που μεταφέρουν χρήσιμες πληροφορίες για τη μελλοντική περιγραφή του υπό διερεύνηση φαινομένου. Σε αυτή την περίπτωση χρησιμοποιήθηκαν τεχνικές βαθιάς μάθησης για την παροχή πληροφοριών σχετικά με τον κίνδυνο που σχετίζεται με τα τρόφιμα και τις πιθανές ανακλήσεις. Οι τεχνικές επικεντρώνονται κυρίως στην πρόβλεψη χρονοσειρών και την ενισχυτική μάθηση προκειμένου να προβλέψουν τον αριθμό των περιπτώσεων τροφίμων. Σχετικά με αυτό το έργο, έπρεπε να αντιμετωπιστούν πολλές προκλήσεις, και αυτή με τον μεγαλύτερο επιστημονικό αντίκτυπο ήταν η χρήση συνθετικά παραγόμενων υποκατάστατων δεδομένων ως ενός τρόπου εμπλουτισμού του αρχικών δεδομένων για τη βελτίωση της απόδοσης των μοντέλων βαθιάς μάθησης.

### 1.3.3 Μηχανική Μάθηση & ποιότητα χοιρινού κρέατος.

Η ταχεία τεχνολογική πρόοδος έχει οδηγήσει σε νέες προσεγγίσεις στη γεωργία, την υδατοκαλλιέργεια και την κτηνοτροφία, χρησιμοποιώντας προηγμένες μεθόδους για τη βελτιστοποίηση της συμβολής κάθε φυτού ή ζώου. Αυτές οι προσεγγίσεις στοχεύουν, για παράδειγμα, στην παράδοση προϊόντων καλύτερης ποιότητας διατηρώντας ή αυξάνοντας τον ρυθμό παραγωγής [60]. Αν και μια άνευ προηγουμένου συλλογή δεδομένων από υπάρχουσες πηγές και συνεχώς εξελισσόμενα δίκτυα αισθητήρων έχει ενισχύσει την ανάπτυξη της τεχνολογίας, αυτά τα δεδομένα μπορεί συχνά να είναι δύσκολο να αξιοποιηθούν για τη λήψη αποφάσεων λόγω του μεγέθους, της ποικιλομορφίας και της ετερογένειάς τους [61]. Ένα πραγματικό σενάριο που δείχνει την προστιθέμενη αξία της χρήσης των συλλεγόμενων δεδομένων είναι αυτό της κηλίδωσης κάπρου (δηλαδή, ανεπιθύμητη δυσσομία που μπορεί να υπάρχει σε μη ευνοχισμένους κάπρους), καθώς οι κτηνοτρόφοι χρειάζονται πληροφορίες για δράση για να μειώσουν την πιθανότητα αυτού του φαινομένου.



Το 2010, συμφωνήθηκε η «Ευρωπαϊκή Διακήρυξη για εναλλακτικές λύσεις στον χειρουργικό ευνουχισμό των χοίρων». Η Διακήρυξη ορίζει ότι από την 1η Ιανουαρίου 2012, ο χειρουργικός ευνουχισμός των χοίρων θα πραγματοποιείται μόνο με παρατεταμένη αναλγησία ή/και αναισθησία και από το 2018 ο χειρουργικός ευνουχισμός των χοίρων θα πρέπει να καταργηθεί εντελώς [62]. Αυτή η Διακήρυξη ενθάρρυνε την έρευνα από ακαδημαϊκούς και επαγγελματίες που προσπαθούσαν να βρουν εναλλακτικούς τρόπους για την εξάλειψη του φαινομένου της κηλίδωσης του κρέατος των κάπρων. Εκτός από τις μελέτες παρέμβασης, ακολουθήθηκαν διάφορες προσεγγίσεις, όπως μελέτες παρατήρησης με συλλογή δεδομένων σε φάρμες και σφαγεία. Μια προσέγγιση που παρέχει πολύτιμες γνώσεις είναι η διεξαγωγή στατιστικών δοκιμών στα συγκεντρωμένα δεδομένα, όπως δοκιμές συσχέτισης που χρησιμοποιούνται για την αναζήτηση συσχετίσεων μεταξύ μεταβλητών και δοκιμές σύγκρισης μέσων τιμών των μεταβλητών. Τα τεστ παλινδρόμησης αξιολογούν εάν μια αλλαγή σε μια μεταβλητή προκαλεί αλλαγές στην άλλη μεταβλητή, όπως διεξάγεται στο [63]. Μια άλλη προσέγγιση που εμφανίζεται τα τελευταία χρόνια είναι η εφαρμογή τεχνικών μηχανικής μάθησης. Στο πλαίσιο αυτής της διατριβής, παρουσιάζουμε μια προσέγγιση που βασίζεται σε διάφορα επεξηγηματικά μοντέλα μηχανικής μάθησης με σκοπό την εξαγωγή των χαρακτηριστικών που συσχετίζονται με την εμφάνιση του φαινομένου της κηλίδωσης του χοιρινού κρέατος.

#### **1.3.4 Μηχανική Μάθηση & τραπεζικές ψηφιακές υπηρεσίες για ΜΜΕ**

Οι Μικρομεσαίες Επιχειρήσεις (ΜΜΕ) αποτελούν τη ραχοκοκαλιά της παγκόσμιας οικονομίας καθώς, σύμφωνα με τον ΟΟΣΑ [64], οι ΜΜΕ στις χώρες μέλη της ΕΕ αντιπροσωπεύουν το 99% όλων των επιχειρήσεων και σχεδόν το 60% της προστιθέμενης αξίας. Αντίστοιχα στην Ευρώπη, οι ΜΜΕ διαδραματίζουν ζωτικό ρόλο καθώς αντιπροσώπευαν το 99,8% του συνόλου των επιχειρήσεων στον μη χρηματοπιστωτικό επιχειρηματικό τομέα της ΕΕ-28, ενώ παράγουν το 56% της προστιθέμενης αξίας και οδηγούν την απασχόληση με το 66% [65]. Οι ΜΜΕ παγκοσμίως αντιμετωπίζουν ένα περίπλοκο και προκλητικό περιβάλλον λόγω των ταχέων παγκόσμιων επιχειρηματικών αλλαγών και της συνεχιζόμενης ψηφιακής επανάστασης. Η συνεχιζόμενη πανδημία του Covid-19 διευρύνει το χάσμα ρευστότητας των ΜΜΕ, αλλά επηρεάζει επίσης σε μεγάλο βαθμό τις συμπεριφορές των πελατών και τις συναλλαγές στην αγορά μέσω της υψηλότερης ολοκλήρωσης της εφοδιαστικής αλυσίδας και της αυξημένης διαφοροποίησης των προϊόντων. Η εισαγωγή ψηφιακών τεχνολογιών και εργαλείων τελευταίας τεχνολογίας μπορεί να ενδυναμώσει τις ΜΜΕ συμβάλλοντας στη μείωση του λειτουργικού και επιχειρηματικού κόστους, εξοικονομώντας χρόνο και πολύτιμους πόρους, ειδικά για επιχειρήσεις που δείχνουν επιβράδυνση της οικονομικής δραστηριότητας και μικρότερους όγκους παραγωγής, οι οποίες επίσης τείνουν να έχει περιορισμένη



εμβέλεια στην αγορά και χαμηλότερη διαπραγματευτική ισχύ με τα ενδιαφερόμενα μέρη [66].

Οι ριζικές τεχνολογικές εξελίξεις μπορούν να επηρεάσουν σχεδόν κάθε πτυχή της λειτουργίας μιας ΜΜΕ, οδηγώντας τη σε διαδικασίες λήψης στρατηγικών αποφάσεων που βασίζονται σε δεδομένα. Ο ψηφιακός μετασχηματισμός μιας ΜΜΕ θα μπορούσε να προσφέρει μια νέα προοπτική στην επιχειρηματική και οικονομική της διαχείριση, προσφέροντας ένα ανταγωνιστικό πλεονέκτημα, εισάγοντας νέες τεχνικές μάρκετινγκ και την ικανότητα εντοπισμού νέων αγορών και πρόβλεψης επιχειρηματικών ευκαιριών.

Φυσικά, δεν αντιμετωπίζουν όλοι οι τομείς τις ίδιες προκλήσεις και εμπόδια όσον αφορά τον ψηφιακό μετασχηματισμό τους. Η παραγωγή δεδομένων και η χρήση της ανάλυσης δεδομένων φαίνεται να είναι υψηλότερη στον χρηματοπιστωτικό τομέα [67], με ορισμένες καινοτόμες νεοφυείς επιχειρήσεις και fintechs να ανοίγουν το δρόμο για τις υπόλοιπες ΜΜΕ που αγωνίζονται να συμμετάσχουν στην οικονομία που βασίζεται στα δεδομένα. Ωστόσο, ο ψηφιακός μετασχηματισμός μιας ΜΜΕ ενέχει πολλούς κινδύνους και προκλήσεις, που έχουν να κάνουν τόσο με την πρόσβαση στα δεδομένα της επιχείρησης όσο και με την ανάλυσή τους. Σύμφωνα με το 2019 OECD SME Outlook [64], οι επί του παρόντος περιορισμένες ψηφιακές δεξιότητες που εντοπίζονται στη διοίκηση των περισσότερων ΜΜΕ, η αδυναμία εντοπισμού, προσέλκυσης και διατήρησης ιδανικών υπαλλήλων, η έλλειψη απαιτούμενων πόρων ή επιλογών χρηματοδότησης και οι αυστηροί κανονισμοί σχετικά με την προστασία δεδομένων αποτελούν τα κύρια εμπόδια στην ψηφιοποίηση των ΜΜΕ.

Σημειώνεται επίσης ότι ορισμένες εμπορικές εφαρμογές ανάλυσης δεδομένων αρχίζουν να διατίθενται οριζόντια για τις ΜΜΕ μέσω υπηρεσιών υπολογιστικού νέφους που προσφέρονται από μεγάλες εταιρείες, επιτρέποντας στις ΜΜΕ να έχουν πρόσβαση σε υπηρεσίες βασισμένες με μεθόδους μηχανικής μάθησης ακόμη και όταν δεν διαθέτουν τους εσωτερικούς πόρους για την ανάπτυξή τους. Άλλες εμπορευματικές εφαρμογές περιλαμβάνουν σε πακέτα ERP ενώ εργαλεία αυτόνομης χρηματοοικονομικής διαχείρισης επιχειρήσεων (BFM) είναι επίσης διαθέσιμα για εμπορική χρήση, αλλά τα περισσότερα από αυτά είναι προσανατολισμένα στην ανάλυση μόνο προηγούμενων συναλλαγών, καθιστώντας αυτά τα εργαλεία ανεπαρκή στον σημερινό κόσμο.

Οι τράπεζες, διατηρώντας μια ποικιλία δεδομένων ΜΜΕ, όπως απαιτείται για τις βασικές τους δραστηριότητες, θα μπορούσαν να προσφέρουν μια λύση αξιοποιώντας όλα τα διαθέσιμα δεδομένα και παρέχοντας μια εργαλειοθήκη επιχειρηματικής οικονομικής διαχείρισης στους πελάτες τους, παρέχοντας υπηρεσίες προστιθέμενης αξίας πέρα από τη βασική τους δραστηριότητα. Προς αυτή την κατεύθυνση, οι τράπεζες μπορούν να αξιοποιήσουν όλα τα διαθέσιμα δεδομένα για να παρέχουν ακριβείς επιχειρηματικές πληροφορίες και αναλυτικές





υπηρεσίες σε λιανικούς ή επιχειρηματικούς πελάτες με αποτέλεσμα, την αυξημένη πελατειακή βάση και αφοσίωση.

Ωστόσο, η ανάπτυξη εξατομικευμένων υπηρεσιών δεν είναι απλή υπόθεση για τις τράπεζες, καθώς θέτει μια ποικιλία επιχειρηματικών και τεχνικών προκλήσεων. Από τεχνική άποψη, μια βασική πρόκληση που προκύπτει κατά την ανάπτυξη υπηρεσιών ανάλυσης δεδομένων και τη χρήση τεχνικών μηχανικής μάθησης για τον χρηματοπιστωτικό τομέα, είναι η έλλειψη «χαρακτηρισμένων» στοιχείων και η εμφάνιση μη αριθμητικών δεδομένων. Λαμβάνοντας υπόψη τον όγκο των χρηματοοικονομικών δεδομένων που παράγονται ή χρησιμοποιούνται από τις τράπεζες, η επισήμανση (χαρακτηρισμός) των δεδομένων γίνεται μια χρονοβόρα εργασία και η ανάπτυξη εξατομικευμένων υπηρεσιών θα μπορούσε να αποδειχθεί ιδιαίτερα απαιτητική. Επιπλέον, τα περισσότερα από τα σημερινά εργαλεία Χρηματοοικονομικής Διαχείρισης για ΜΜΕ είναι προσανατολισμένα στην ανάλυση προηγούμενων συναλλαγών, χρησιμοποιώντας κυρίως στατιστικές μεθόδους, καθιστώντας αυτά τα εργαλεία ανεπαρκή στον σημερινό κόσμο. Σήμερα, τόσο οι ΜΜΕ όσο και οι πελάτες τους, απαιτούν έγκαιρη επεξεργασία, διαφάνεια και εξατομικευμένες υπηρεσίες όχι μόνο να κατανοήσουν καλύτερα την επιχειρηματική/οικονομική κατάσταση τους, αλλά και να μπορούν να διαμορφώσουν τις επιλογές τους και να αποφασίσουν για τις βέλτιστες ενέργειες που πρέπει να κάνουν.

Στην διατριβή, εξετάζεται η αναδυόμενη πρόκληση της οικονομικής διαχείρισης των ΜΜΕ, εισάγοντας μια προσέγγιση βασισμένη στα ιστορικά δεδομένα συναλλαγών για τη διευκόλυνση της έγκαιρης λήψης αποφάσεων. Οι βασικές υπηρεσίες περιλαμβάνουν την έγκαιρη κατηγοριοποίηση συναλλαγών κάθε συναλλαγής σύμφωνα με το υποκείμενο προϊόν/υπηρεσία τόσο πολυδιάστατων δεδομένων πίνακα όσο και μη δομημένων δεδομένων κειμένου (δηλαδή περιγραφή συναλλαγής), χρησιμοποιώντας μεθόδους μηχανικής μάθησης. Εν συνεχεία έχοντας αναπαραστήσει τα δεδομένα στο πεδίο του χρόνου ως χρονοσειρές ακολουθούμε τις μεθόδους πρόβλεψης χρονοσειρών. Οι προβλέψεις αυτές αφορούν τις ταμιακές εισροές και εκροές για κάθε κατηγορία συναλλαγών, εφαρμόζοντας και συγκρίνοντας διαφορετικά πιθανοτικά μοντέλα βαθιάς μάθησης, ενώ εξετάζεται η χρησιμότητα των υποκατάστατων δεδομένων ως μια προσέγγιση εμπλουτισμού των δεδομένων χρονοσειρών.

#### **1.4 Προκλήσεις κατά την εφαρμογή μεθόδων Μηχανικής Μάθησης σε πραγματικές συνθήκες.**

Στην εν λόγω διατριβή που ως επί τω πλείστων ασχοληθήκαμε με προβλήματα χρονοσειρών, κατά τη διάρκεια των διαδικασιών εξόρυξης δεδομένων ή επεξεργασίας αυτών επιπλέον προκλήσεις βγήκαν στην επιφάνεια. Ενδεικτικά αναφέρουμε την έλλειψη εμπιστοσύνης στα μοντέλα μηχανικής μάθησης όταν αυτά



χρησιμοποιούνται ως «μαύρα κουτιά», την εφαρμογή ενισχυτικής μάθησης (reinforcement learning) για πρόβλεψη χρονοσειρών, τον εμπλουτισμό μικρών σε μέγεθος συνόλου δεδομένων με σκοπό την καλύτερη γενίκευση των μοντέλων σε άγνωστα δεδομένα.

#### 1.4.1 Επεξηγήσιμη Τεχνητή Νοημοσύνη - eXplainable Artificial Intelligence (XAI)

Αν και τόσο ο έλεγχος υποθέσεων όσο και ο συστηματικός έλεγχος της σημασίας των διακυμάνσεων είναι καλά εδραιωμένες στατιστικές μέθοδοι, δεν ισχύει το ίδιο για την επεξήγηση της μηχανικής μάθησης. Παρόλο που υπάρχουν πολλές μετρικές και μετρήσεις απόδοσης, δεν αποδίδουν το «σκεπτικό» πίσω από τις αποφάσεις οι τις προβλέψεις που ελήφθησαν [68]. Αυτή η ενεργή ερευνητική περιοχή η οποία εξετάσει και προσπαθεί να εξηγήσει τα μοντέλα μηχανικής μάθησης είναι γνωστή ως Επεξηγήσιμη Τεχνητή Νοημοσύνη (eXplainable AI (XAI)) [69]. Πιο ισχυρές μορφές ερμηνείας θα μπορούσαν να προσφέρουν πολλά πλεονεκτήματα, από την εμπιστοσύνη στις προβλέψεις μοντέλων και την ανάλυση σφαλμάτων έως τη βελτίωση του μοντέλου και την καλύτερη κατανόηση των δεδομένων εισόδου. Όπως είναι ευκολά κατανοητό το XAI είναι περισσότερο απαιτητό σε τομείς όπου εμπλέκεται η λήψη αποφάσεων με μεγάλο κόστος, όπως π.χ. η ιατρική [70].

Στο πλαίσιο της διατριβής, εφαρμόσαμε διάφορες προσεγγίσεις και μεθόδους XAI, ο οποίες συνήθως καταλήγουν στην εξαγωγή της σπουδαιότητας των χαρακτηριστικών εισόδου, καταλήγοντας σε χρήσιμη γνώση τόσο από τεχνική όσο και από επιχειρηματική άποψη. Γενικά, οι τεχνικές επεξήγησης χωρίζονται σε δύο κατηγορίες:

- Μεθόδους ενσωματωμένες στο μοντέλο Μηχανικής Μάθησης, για την εξαγωγή σημαντικότητας χαρακτηριστικών του μοντέλου (π.χ. Random Forest).
- Μεθόδους post-hoc, εξαγωγή δηλαδή της σημαντικότητας των χαρακτηριστικών με τη χρήση τρίτων μοντέλων όπως το LIME [71] και το SHAP [72].

Αξίζει να σημειωθεί ότι για πολλά υποκείμενα μοντέλα μηχανικής και κυρίως βαθιάς μάθησης και διάφορους τύπους δεδομένων η συνεπής αξιολόγηση της σημασίας των χαρακτηριστικών παραμένει πρόκληση [73]. Στην εν λόγω διατριβή ασχοληθήκαμε τόσο με επεξήγηση σε δεδομένα εικόνας αλλά και χρονοσειρών που είναι και η βάση της διατριβής.

Εν συνεχεία, αξίζει να σημειωθεί ένα επιπλέον παράγοντας που παραμένει ανοιχτή πρόκληση είναι τα εξαιρετικά μη ισορροπημένα σύνολα δεδομένων, καθώς



μπορεί να αυξήσει δραστικά την πολυπλοκότητα τόσο της διαδικασίας εκπαίδευσης ενός μοντέλου μηχανικής μάθησης όσο και της επεξήγησης αυτού.

### 1.4.2 Μη ισορροπημένο σύνολο δεδομένων

Ένα μη ισορροπημένο σύνολο δεδομένων είναι μια κοινή πρόκληση σε πολλές εφαρμογές ταξινόμησης [74] αλλά και πρόβλεψης. Το πρόβλημα προκύπτει όταν ο αριθμός των δειγμάτων σε μια κατηγορία (κατηγορία πλειοψηφίας) είναι πολύ μεγαλύτερος από την ή τις υπόλοιπες κατηγορίες (κατηγορία μειοψηφίας), επηρεάζοντας σημαντικά την ακρίβεια ταξινόμησης των μοντέλων. Τα μη ισοκατανεμημένα σύνολα δεδομένα αποτελούν πρόκληση για την προγνωστική μοντελοποίηση, δεδομένου ότι οι περισσότεροι από τους αλγόριθμους μηχανικής μάθησης έχουν σχεδιαστεί για να αποδίδουν καλά σε ένα ισορροπημένο σύνολο δεδομένων (δηλαδή, ίσος αριθμός δειγμάτων για κάθε τάξη/κλάση). Έτσι, πολλά μοντέλα έχουν μειωμένη απόδοση, ειδικά για την κατηγορία μειοψηφίας, καθώς υπάρχει ασυμφωνία μεταξύ των παρατηρήσεων που βασίζονται σε χαρακτηριστικά και σε αρχικά στατιστικά μεγέθη που έχουν χρησιμοποιηθεί κατά τη ανάπτυξη των μοντέλων. Στις περισσότερες από αυτές τις περιπτώσεις, η μειονοτική τάξη είναι αυτή που παρουσιάζει το μεγαλύτερο ενδιαφέρον (π.χ. καρκινική διάγνωση [75], επίθεση ασφάλειας [76]). Πολλές μέθοδοι ισχυρίζονται ότι μπορούν να βελτιώσουν την ακρίβεια της τάξης της μειοψηφίας χωρίς να θυσιάσουν σε μεγάλο βαθμό την ακρίβεια της τάξης της πλειοψηφίας. Μια ολοκληρωμένη ανασκόπηση των μεθόδων ανάκτησης ισορροπίας [77] παρουσιάζει ορισμένες καθιερωμένες προσεγγίσεις δειγματοληψίας, όπως το SMOTE και το ADASYN. Ενώ τα βαθιά γενετικά μοντέλα (DGM) λειτουργούν καλά σε προβλήματα συνεχών μεταβλητών, όπως εικόνες και βίντεο, όπου είναι δύσκολο να δημιουργηθούν δείγματα με τη μορφή διακριτών ακολουθιών ή μικτών κατανομών (δηλαδή, πίνακες δεδομένων που περιέχουν πολλαπλές κατηγορικές και αριθμητικές μεταβλητές) [78].

Αν και αυτοί οι τύποι προβλημάτων ανήκουν σε μια ανοιχτή ερευνητική περιοχή, στην ανα χείρας διατριβή μας, το μοντέλο Catboost [79] προτιμήθηκε με βάση τα ευρήματα του [80], σε όλες τις εφαρμογές ταξινόμησης. Αυτή η μελέτη εξέτασε τα πιο καθιερωμένα μοντέλα ενίσχυσης που βασίζονται σε μη ισορροπημένα σύνολα δεδομένων πολλών κλάσεων, καταλήγοντας στο συμπέρασμα ότι οι αλγόριθμοι CatBoost αποδίδει καλύτερα από άλλους αλγόριθμους που βασίζονται σε ενίσχυση σε σύνολα δεδομένων με ανισορροπία πολλών κλάσεων. Επιπλέον σημειώνεται ότι το Catboost μπορεί από τη φύση του να χειριστεί αποτελεσματικότερα κατηγορηματικές μεταβλητές με λιγότερη (εώς μηδενική) προ επεξεργασία αυτών.

### 1.4.3 Named Entity Recognition (NER) – Επισήμανση ακολουθίας



Η Αναγνώριση Ονομαζόμενης Οντότητας (NER-Named Entity Recognition) είναι μια υποκατηγορία του τομέα Επεξεργασίας Φυσικής Γλώσσας (NLP). Αντιστοιχεί στη δυνατότητα αναγνώρισης των ονομασμένων οντοτήτων σε έγγραφα και την επισήμανση τους με μία από τις ετικέτες οντοτήτων, όπως πρόσωπο, τοποθεσία ή οργανισμό. Ωστόσο, υπάρχει ένας μεγάλος αριθμός από ετικέτες οντοτήτων για διαφορετικούς τομείς. Η δημιουργία κλάσεων οντοτήτων που δεν ταιριάζουν στο παραδοσιακό παράδειγμα των τεσσάρων κατηγοριών (PER, LOC, ORG, MISC) είναι ένα σύνθετο πρόβλημα. Υπάρχουν πολλές μέθοδοι σχολιασμού κειμένου προκειμένου να χρησιμοποιηθούν για τις μεθόδους NER, όπου αυτές οι μέθοδοι φαίνονται ελπιδοφόρες για το NER αλλά εξακολουθούν να αφήνουν πολλά περιθώρια για βελτιώσεις.

Στην Επεξεργασία Φυσικής Γλώσσας, η εργασία της επισήμανσης/τιτλοφόρησης ακολουθίας χρησιμοποιείται για να εκχωρηθεί μια κατηγορική ετικέτα σε κάθε όρο μιας πρότασης/εγγράφου. Ένα από τα πιο κοινά σχήματα ετικετών είναι η Επισήμανση Μέρους του Λόγου (Part Of Speech tagging-POS) το οποίο αντιστοιχεί μέρη του λόγου σε κάθε λέξη. Ενώ αυτό μπορεί να αντιμετωπιστεί ως εργασία ταξινόμησης κάθε λέξης, η βιβλιογραφία υπογραμμίζει ότι οι εξαρτήσεις μεταξύ των λέξεων φέρουν χρήσιμες πληροφορίες και επομένως βελτιώνουν την ακρίβεια όταν περιλαμβάνονται. Ως εκ τούτου, ότι στη βιβλιογραφία απαντώνται πολλές προσεγγίσεις εξαγωγής των οντοτήτων, όπως η χρήση οριοθετημένων τυχαίων πεδίων [81], και προσεγγίσεις βασισμένες στην αρχιτεκτονική BERT. Εκτός από τη διαφοροποίηση στις προσεγγίσεις, υπάρχει και διαφοροποίηση στις στρατηγικές προσθήκης ετικετών θέσης της οντότητας με το "BIOE" να είναι το πιο δημοφιλές. Στη στρατηγική προσθήκης ετικετών BIOE, το "B" αντιπροσωπεύει την αρχή ενός χαρακτηριστικού, το "I" αντιπροσωπεύει το εσωτερικό ενός χαρακτηριστικού, το "O" αντιπροσωπεύει το εξωτερικό ενός χαρακτηριστικού και το "E" αντιπροσωπεύει το τέλος του. Άλλες δημοφιλείς στρατηγικές προσθήκης ετικετών περιλαμβάνουν το "UBIOE" και το "IOB". Το "UBIOE" έχει μια επιπλέον ετικέτα "U" που αντιπροσωπεύει την ετικέτα μονάδας διακριτικού που διαχωρίζει τα μονολεκτικά χαρακτηριστικά από τα πολυλεκτικά ενώ για την επισήμανση "IOB", το "E" παραλείπεται καθώς το "B" και το "I" αρκούν για να εκφράσουν το τελικό όριο ενός χαρακτηριστικού.

#### 1.4.4 Ενισχυτική μάθηση για χρονοσειρές

Η ενισχυτική μάθηση δεν είναι πολύ συνηθισμένο να χρησιμοποιείται σε προβλήματα πρόβλεψης χρονοσειρών. Παρόλα αυτά αποδείξαμε στο πλαίσιο της διατριβής ότι με κατάλληλη μοντελοποίηση του προβλήματος τα αποτελέσματα μιας τέτοιας μεθόδου είναι τουλάχιστον ανταγωνιστικά προσφέροντας ταυτόχρονα οφέλη της αυτόματης προσαρμογής στα δεδομένα.



Η Ενισχυτική Μάθηση μπορεί να αναλυθεί σε τρεις κύριες κατηγορίες: critic-only, actor-only, actor-critic [82]. Δεδομένου ότι στη διατριβή ορίσαμε και χρησιμοποιήσαμε ενισχυτική μάθηση που εμπίπτει στην κατηγορία critic-only, θα πρέπει να σημειωθεί ότι στην προσέγγιση αυτή ο πράκτορας προσπαθεί να εκπαιδευτεί σε μια συνάρτηση τιμής κατάστασης-δράσης  $Q$  ή μια προσέγγιση του  $Q$  για να δημιουργήσει μια αντιστοίχιση  $S, A \rightarrow v$  που αντιπροσωπεύει την καταλληλότητα μιας συγκεκριμένης ενέργειας δεδομένης της κατάστασης, όπου  $S$  και  $A$  είναι η κατάσταση και οι χώροι δράσης αντίστοιχα. Ενώ υπάρχουν πολλές υλοποιήσεις που εμπίπτουν σε αυτήν την κατηγορία, η πιο σημαντική είναι η  $Q$  (DQN) βαθιά μάθηση με πολλές βελτιώσεις και προκαθορισμένους  $Q$  στόχους, η διπλή DQN, οι φέροντες DQN (DDQN) και η Επανάληψη Εμπειριών Προτεραιότητας (PER) [83]. Για να λειτουργήσει αυτή η προσέγγιση είναι απαραίτητο είτε να υπάρχει ένας διακριτός χώρος δράσης είτε να ψηφιοποιηθεί ένας συνεχής χώρος τιμών με μεθόδους όπως η tile-coding [84], coarse coding, function approximation κ.λπ.

#### 1.4.5 Υποκατάστατα-Αναλογικά Δεδομένα χρονοσειράς

Το πρώτο βήμα της πρόβλεψης χρονοσειρών είναι η απόρριψη της μηδενικής υπόθεσης «λευκού θορύβου». Στη συνέχεια, εάν πρόκειται να χρησιμοποιηθούν μη γραμμικές μέθοδοι όπως μηχανική μάθηση, βαθιά μάθηση και ενισχυτική μάθηση, είναι απαραίτητο να απορριφθεί η μηδενική υπόθεση σχετικά με την ύπαρξη αποκλειστικά χρονικών γραμμικών συσχετισμών στις χρονοσειρές. Ένα στατιστικά ορθό πλαίσιο για το προαναφερθέν τεστ είναι αυτό των υποκατάστατων δεδομένων (surrogate data), μερικές φορές γνωστά ως αναλογικά δεδομένα. Αναφέρεται σε δεδομένα χρονοσειρών που αναπαράγουν διάφορες στατιστικές ιδιότητες όπως η δομή αυτοσυσχέτισης ενός μετρήσιμου συνόλου δεδομένων. Η μηδενική υπόθεση αντιπροσωπεύεται από τα υποκατάστατα δεδομένα που συγκρίνονται με τα αρχικά δεδομένα κάτω από μια μη γραμμική διακριτική στατιστική για την απόρριψη ή την έγκριση της μηδενικής υπόθεσης. Σε αυτή την εργασία η χρήση υποκατάστατων δεδομένων ήταν διπλή:

1. Να ελέγξει την υπόθεση της μη γραμμικότητας.
2. Να εμπλουτίσει το υπάρχον σύνολο δεδομένων με τα νέα δεδομένα.

Το τελευταίο προτείνεται λόγω της φύσης των περισσότερων μοντέλων βαθιάς μάθησης που μπορούν να ωφεληθούν από την αύξηση τόσο της ποικιλομορφίας όσο και του όγκου των δεδομένων εκπαίδευσης. Η μηδενική υπόθεση στην περίπτωση μας είναι ότι η χρήση ανάλογων δεδομένων θα βοηθήσει στη γενίκευση ορισμένων, αν όχι όλων των μοντέλων βαθιάς μάθησης, στην περίπτωση της πρόβλεψης χρονοσειρών. Παρόλο που υπάρχουν πολλές μέθοδοι για τη



δημιουργία τέτοιων δεδομένων, επιλέχθηκε το μοντέλο IAAFT [86] για το οποίο γίνεται εκτενέστερη αναφορά στην επόμενη ενότητα.

## 1.5 Συνεισφορά και καινοτομία Διατριβής

Αυτή η ενότητα συνοψίζει τις βασικές ερευνητικές κατευθύνσεις της διατριβής, περιγράφοντας εν συντομία τις βασικές ιδέες της εφαρμογής σε κάθε περίπτωση. Κάθε υποενότητα που ακολουθεί αντιστοιχεί σε μία από αυτές τις κατευθύνσεις, περιγράφοντας τις προκλήσεις ή/και κίνητρα που οδήγησαν στην εκάστοτε λύση, μαζί με μια συνοπτική περίληψη της μεθοδολογίας μας. Καθένα από αυτές οι ερευνητικές κατευθύνσεις συνδέεται με το αντίστοιχο κεφάλαιο και το αντίστοιχο ερευνητικό άρθρο μας.

Μπορούμε να συνοψίσουμε την κύρια συμβολή αυτής της διατριβής ως την η εξέταση και την πρόταση ενός σύνθετου πλαισίου που αξιοποιεί δεδομένα για την παροχή προβλέψεων σε βιομηχανικά περιβάλλοντα. Η μεθοδολογία διακρίνεται σε διάφορα διακριτά βήματα τα οποία κατηγοριοποιούνται ως «Εξόρυξη δεδομένων», «Προ-επεξεργασία», «Μοντελοποίηση σε επίπεδο χρονοσειρών» και «Εκπαίδευση μοντέλων Μηχανικής Μάθησης». Στα αντίστοιχες εφαρμογές επιπλέον αποκομίζοντας ιδιαίτερα οφέλη οπότε θα εξετάσουμε την κάθε περίπτωση ειδικά.

### 1.5.1 Μηχανική Μάθηση & προβλεπτική συντήρηση κύριων μηχανών εμπορικών πλοίων.

Η συνεισφορά της διατριβής στο πλαίσιο της προβλεπτικής συντήρησης συνοψίζεται στην πρόταση μεθοδολογίας, βασισμένη στη μηχανική μάθηση, για την εκτίμηση της λειτουργικής κατάστασης των μηχανικών μερών ενός εξοπλισμού. Στην περίπτωση αυτή η προτεινόμενη μεθοδολογία εφαρμόστηκε και δοκιμάστηκε σε δεδομένα εμπορικών πλοίων καλύπτοντας μια ανοιχτή πρόκληση της εμπορικής ναυτιλίας. Συγκεκριμένα, σε αυτό το πλαίσιο, προσφέρεται ένα σύνολο μοντέλων μηχανικής μάθησης, μεταξύ άλλων πολυδιάστατο και μονοπαραγοντικό μοντέλο μακράς βραχυπρόθεσμης μνήμης (Long Short Term Memory - LSTM), Support Vector Machines –(SVM), Σταθμισμένη Μετάθεση Εντροπίας (WPE), κατάλληλα για τον εντοπισμό των σημειακών και των συγκεντρωτικών ανωμαλιών σε χρονοσειρές.

Η κύρια διαφορά της προτεινόμενης εργασίας συγκρινόμενη με άλλες προσεγγίσεις για τον εντοπισμό ανωμαλιών, είναι ότι εισάγουμε μία καινοτόμα προσέγγιση συνδυάζοντας διάφορα μοντέλα μηχανικής και βαθιάς μάθησης βασισμένα σε πραγματικά δεδομένα, επιτυγχάνοντας τον έγκαιρο εντοπισμό βλάβης του ρουλεμάν σταυρού κεφαλής. Πιο συγκεκριμένα η προσέγγιση περιλαμβάνει την περιγραφή μίας κατάλληλης τεχνικής προ-επεξεργασίας του λαμβανόμενου συνόλου δεδομένων και την υλοποίηση διαφόρων μοντέλων για να



εντοπίσουν τόσο ανωμαλίες σημείου όσο και συγκεντρωτικές ανωμαλίες. Η αποτελεσματικότητα της προτεινόμενης προσέγγισης να συγκεντρώσει διαφορετικά μοντέλα, αποδεικνύεται και από τα αποτελέσματα στο σενάριο του πραγματικού κόσμου και του συνόλου δεδομένων.

### 1.5.2 Μηχανική Μάθηση & ασφάλεια των τροφίμων.

Η διατριβή αντιμετωπίζει την αναδυόμενη πρόκληση της ασφάλειας των τροφίμων εισάγοντας μια προσέγγιση προερχόμενη από την μηχανική μάθηση και την πρόβλεψη χρονοσειρών για τη διευκόλυνση της έγκαιρης λήψης αποφάσεων για ανακλήσεις τροφίμων. Εφαρμόζοντας το framework που πραγματεύεται η παρούσα διατριβή, αυτό προσαρμόζεται σε δύο κύρια στοιχεία, έναν μηχανισμό ταξινόμησης που βασίζεται σε μια τεχνική βαθιάς μάθησης για την Αναγνώριση Ονομασίας Οντότητας, που ταξινομεί τις ανακοινώσεις ανακλήσεων τροφίμων σύμφωνα με το εμπλεκόμενο προϊόν και μια μέθοδο πρόβλεψης χρονοσειρών που τροφοδοτείται από τις χρονοσειρές που ορίζονται από τις ταξινομημένες ανακλήσεις τροφίμων.

Πιο συγκεκριμένα, η προσέγγιση περιλαμβάνει:

α. Την ανάπτυξη ενός προσαρμοσμένου περιβάλλοντος ενισχυτικής μάθησης και την υλοποίηση ενός μοντέλου που μπορεί να υπολογίσει την πρόβλεψη 30 κατηγοριών για χρονικό ορίζοντα τεσσάρων μηνών. Θα πρέπει επίσης να αναφερθεί ότι αυτή η προσέγγιση δεν εφαρμόζεται ευρέως σε εργασίες πρόβλεψης χρονοσειρών, ενώ προσφέρει συνεχή βελτιστοποίηση και προσαρμοστικότητα, ως προστιθέμενη αξία του μοντέλου ενισχυτικής μάθησης.

β. Δεδομένου ότι για σύνολα δεδομένων χρονοσειρών "μικρού" μεγέθους, η εύρεση ενός μοντέλου πρόβλεψης που προσφέρει ποιοτικές προβλέψεις αποτελεί σημαντική πρόκληση στη μηχανική μάθηση όπως αναφέρεται στο [87], η διατριβή προτείνει τη χρήση υποκατάστατων δεδομένων ως μια βελτιωμένη προσέγγιση που παρέχει πιο ακριβή αποτελέσματα.

γ. Ένα προσαρμοσμένο μετα-μοντέλο (βάσει των στατιστικών χαρακτηριστικών κάθε χρονοσειράς) χρησιμοποιήθηκε ως πρόσθετο στοιχείο προκειμένου να εισαχθεί ένα υβριδικό μοντέλο για περαιτέρω βελτιστοποίηση της απόδοσης του πλαισίου.

δ. Επιπλέον, θα πρέπει να τονιστεί ότι ένα μοντέλο επεξεργασίας φυσικής γλώσσας χρησιμοποιείται για την εξαγωγή των υποκείμενων τροφίμων από την ανακοίνωση ανακλήσεων, η οποία δείχνει την προστιθέμενη αξία από την αδόμητη αναπαράσταση δεδομένων (δηλ. δεδομένα κειμένου) σε μια δομημένη μορφή όπως αυτή της αναπαράστασης χρονοσειράς.

Στο πλαίσιο αυτό, ο επιχειρησιακός στόχος είναι διπλός:



α. Έγκαιρες ανακλήσεις τροφίμων με βάση την αποτελεσματική ταξινόμηση κάθε ανάκλησης σύμφωνα με το εκάστοτε υποκείμενο προϊόν. Συγκεκριμένα προτείνεται προσέγγιση για το πώς τα μη δομημένα δεδομένα κειμένου που ανακτώνται από ετερογενείς πηγές μπορούν να μετασχηματιστούν και να ταξινομηθούν χρησιμοποιώντας μεθόδους βαθιάς μάθησης για φυσική επεξεργασία γλώσσας και να αναπαρασταθούν στον τομέα του χρόνου ως χρονοσειρές που επιτρέπουν την εκμετάλλευση της μεθόδου πρόβλεψης με χρήση πιθανοτικών μοντέλων βαθιάς μάθησης.

β. Συστηματική πρόβλεψη ανακλήσεων βάσει ιστορικών δεδομένων που εφαρμόζονται για κάθε γενική κατηγορία προϊόντων εφαρμόζοντας και συγκρίνοντας διαφορετικές πιθανολογικές μεθόδους βαθιάς μάθησης. Επιπλέον, επεκτείνεται η μελέτη τόσο προς την εισαγωγή ενός μετα-μοντέλου που επιτρέπει μια υβριδική υλοποίηση επιλογής κατάλληλου μοντέλου πρόβλεψης όσο και προς τη χρησιμότητα τεχνητών (ή ανάλογων) δεδομένων ως μέθοδο εμπλουτισμού δεδομένων.

Ενώ οι επιστημονικές συνεισφορές μπορούν να συνοψιστούν εν συντομία ως προς την προστιθέμενη αξία, καθώς η προτεινόμενη προσέγγιση επιτρέπει:

α. Συνεχή βελτιστοποίηση του μοντέλου πρόβλεψης χρονοσειρών (και των αποτελεσμάτων του) με βάση τη χρήση τεχνικής ενισχυτικής μάθησης και τον αντίστοιχο μοντέλο ταξινόμησης βάσει του υποκείμενου προϊόντος επανεκπαιδευόντας το σε συγκεκριμένα διαστήματα.

β. Βελτίωση των αποτελεσμάτων του μοντέλου πρόβλεψης των ανακλήσεων λόγω της χρήσης υποκατάστατων/ανάλογων δεδομένων που εμπλουτίζουν το σύνολο δεδομένων και συμβάλλουν στην γενίκευση του προτεινόμενου μοντέλου.

γ. Βέλτιστη χρήση διαφορετικών μοντέλων τόσο ως προς το χρόνο και την πολυπλοκότητα όσο και ως προς τα αποτελέσματα (και λήψη αποφάσεων) λόγω της εκμετάλλευσης της προτεινόμενης υβριδικής προσέγγισης για την επιλογή του καλύτερου μοντέλου για πρόβλεψη χρονοσειρών.

### 1.5.3 Μηχανική Μάθηση & ποιότητα χοιρινού κρέατος.

Σε αυτό το πλαίσιο, η διατριβή αξιολογεί διάφορες προσεγγίσεις που βασίζονται σε κυρίως σε μεθόδους μηχανικής μάθησης και σε άλλα στατιστικά μοντέλα. Εισάγει μια προσέγγιση για την αντιμετώπιση μη ισορροπημένων δεδομένων αξιοποιώντας γενετικούς αλγόριθμους για τη διευκόλυνση της εξαγωγής των σημαντικών παραγόντων και χαρακτηριστικών που σχετίζονται με τη εμφάνιση του φαινομένου της κηλίδωσης των κάπρων.

Προς αυτή την κατεύθυνση, στην διατριβή αξιοποιήσαμε ένα ιδιωτικό σύνολο δεδομένων πιθανών παραγόντων κινδύνου που σχετίζονται με τη κηλίδωση





κάπρου, όπως η εφηβεία, η στέγαση, η υγεία, οι συνθήκες πριν από τη σφαγή, η εποχή, η τροφή και η σύνθεση του σφαγίου. Με βάση αυτές τις μεταβλητές εξετάσαμε και αξιολογήσαμε στατιστικές μεθόδους και μοντέλα μηχανικής και βαθιάς μάθησης σε συνδυασμό με τεχνικές επεξηγήσιμης τεχνητής νοημοσύνης και καταλήξαμε με την πρόταση μιας νέας τεχνικής υπερ-δειγματοληψίας που βασίζεται στη μεθοδολογία των γενετικών αλγορίθμων [88].

Πέρα από τα συμπεράσματά που αφορούν στη μεθοδολογία και την αντιμετώπιση των προκλήσεων που τίθενται κατά την υλοποίησης μηχανικής μάθησης σε πραγματικά σύνολα δεδομένων, παράλληλα τα αποτελέσματα της ανάλυσής μας, σύμφωνα με την πλειονότητα των προσεγγίσεων, υποδεικνύουν χαρακτηριστικά που φαίνονται πιο σημαντικά και μπορεί να επηρεάσουν την ποιότητα του κρέατος (εμφάνιση φαινομένου κηλίδωσης). Αυτά μπορούν να τα συνοψίσουμε στα κάτωθι:

- α. ο τύπος τροφής,
- β. ο τύπος συστήματος εξαερισμού,
- γ. η φαρμακευτική αγωγή, και
- δ. η διάρκεια αναμονής στο χώρο αποθήκευσης .

Τέλος, μια υπόθεση που επιβεβαιώνεται από τα αποτελέσματα είναι ότι ο λόγος ανισορροπίας δεν είναι η μόνη αιτία της απώλειας απόδοσης ενός ταξινομητή. Άλλοι παράγοντες, όπως οι μικρές αποσυνδέσεις, ο θόρυβος και οι επικαλύψεις στα δεδομένα, παίζουν επίσης σημαντικό ρόλο και σε συνδυασμό με την αναλογία ανισορροπίας και οδηγούν σε υποαπόδοση ενός ταξινομητή.

#### **1.5.4 Μηχανική Μάθηση & τραπεζικές ψηφιακές υπηρεσίες για ΜΜΕ.**

Η διατριβή αντιμετωπίζει την ανάγκη δημιουργίας νέων ψηφιακών τραπεζικών υπηρεσιών με σκοπό την βελτίωση της οικονομικής και διοικητικής διαχείρισης των ΜΜΕ, αντιμετωπίζοντας προκλήσεις όπως η έλλειψη χαρακτηρισμένων δεδομένων, η ιδιωτικότητα των δεδομένων, και ανάγκη ανατροφοδότησης από τους χρήστες, καθώς και η γενίκευση των μοντέλων μηχανικής μάθησης σε άγνωστα δεδομένα.

Η καινοτομία της προτεινόμενης εργασίας σε σχέση με τις παλαιότερες έρευνες είναι ότι τόσο το μοντέλο κατηγοριοποίησης συναλλαγών όσο και το μοντέλο πρόβλεψης ταμειακών ροών περιστρέφονται γύρω από συγκεκριμένες ανάγκες των ΜΜΕ. Εφαρμόζοντας το framework που πραγματεύεται η παρούσα διατριβή, αυτό προσαρμόζεται σε δύο κύρια στοιχεία, έναν μηχανισμό ταξινόμησης των συναλλαγών, που τις ταξινομεί σύμφωνα με το εμπλεκόμενο προϊόν και μια μέθοδο πρόβλεψης χρονοσειρών που τροφοδοτείτε από τις χρονοσειρές που ορίζονται από τις ταξινομημένες συναλλαγές. Οι δύο προσεγγίσεις που



αναπτύχθηκαν είναι αλληλένδετες με οποιεσδήποτε αλλαγές στη διαδικασία κατηγοριοποίησης να αντικατοπτρίζονται στο μοντέλο πρόβλεψης ταμειακών ροών.

Στο πλαίσιο αυτό, ο επιχειρησιακός στόχος είναι διπλός:

α. Έγκαιρη κατηγοριοποίηση συναλλαγών σύμφωνα με την υποκείμενη κατηγορία εξόδων. Συγκεκριμένα τα δομημένα μη επισημασμένα δεδομένα μπορούν να ταξινομηθούν χρησιμοποιώντας ένα Υβριδικό Μοντέλο (σύνολο κανόνων και μηχανικής μάθησης) και να αναπαρασταθούν στον τομέα του χρόνου ως χρονοσειρές που επιτρέπει την πρόβλεψη μελλοντικών ταμειακών ροών.

β. Συστηματική πρόβλεψη δαπανών με βάση τα ιστορικά δεδομένα που εφαρμόζονται για κάθε γενική κατηγορία συναλλαγών. Στο τέλος, εξετάστηκε η χρησιμότητα των υποκατάστατων δεδομένων ως προσέγγιση εμπλουτισμού.

Η συνεισφορά μας στην εν λόγω εφαρμογή, είναι πρόταση μιας υβριδικής προσέγγισης (δηλαδή βασισμένη σε κανόνες και μέθοδο μηχανικής μάθησης) για την μοντελοποίηση του προβλήματος καταλήγοντας σε μια πρόκληση πρόβλεψης χρονοσειρών. Πιο συγκεκριμένα, η προσέγγιση περιλαμβάνει: (α) την ανάπτυξη ενός υβριδικού μοντέλου κατηγοριοποίησης συναλλαγών, που προσφέρει συνεχή βελτιστοποίηση και προσαρμοστικότητα, μέσω του βρόχου ανάδρασης, (β) δεδομένου ότι για σύνολα δεδομένων χρονοσειρών «μικρού» μεγέθους, η εύρεση ενός μοντέλου πρόβλεψης μηχανικής ή βαθιάς μάθησης που προσφέρει ποιοτικές προβλέψεις αποτελεί σημαντική πρόκληση στη μηχανική μάθηση, η παρουσιαζόμενη προσέγγιση προτείνει τη χρήση υποκατάστατων δεδομένων ως μια βελτιωμένη προσέγγιση που παρέχει πιο ακριβή αποτελέσματα.

### 1.6 Αντικείμενο και Δομή Διατριβής

Αυτή η έρευνα καθοδηγείται από εφαρμογές πραγματικού κόσμου που ασχολούνται με θεμελιώδη προβλήματα χρονοσειρών με εφαρμογές σε πραγματικά προβλήματα της βιομηχανίας.

Στο **ΚΕΦΑΛΑΙΟ 1**, πραγματοποιείται μία εισαγωγή στις έννοιες της Μηχανικής Μάθησης και των χρονοσειρών, Εν συνεχεία αναφέρονται συγκεκριμένα παραδείγματα τριών κατηγοριών προβλημάτων χρονοσειρών βασισμένα στην πραγματική βιομηχανία. Τόσο θεωρητικό όσο και συζητούνται τεχνικές πτυχές αυτών των προβλημάτων, εκφράζοντας την ανάγκη της κάλυψης πολλαπλών υπάρχουσών προκλήσεων, με αυτές της ανισόροπής κατανομής των δεδομένων και της αξιοπιστίας των μοντέλων Μηχανικής Μάθησης να ξεχωρίζουν μεταξύ άλλων. Επιπροσθέτως, αναλύεται η καινοτομία, η συνεισφορά και οι τελικοί στόχοι της παρούσας Διδακτορικής Διατριβής, λαμβάνοντας υπόψη τις προαναφερθείσες προκλήσεις. Παράλληλα, πραγματοποιείται μία σύντομη ανασκόπηση σε παλαιότερες και κομβικές προσωπικές έρευνες, οι οποίες θεωρούνται ως ορόσημα για το τελικό αποτέλεσμα της ερευνάς.



Στο **ΚΕΦΑΛΑΙΟ 2**, αρχικά αναφέρεται η σχετική βιβλιογραφική έρευνα που πραγματοποιήθηκε για διάφορους ερευνητικούς τομείς όπως αυτούς της πρόβλεψης χρονοσειρών, της επεξηγησιμότητας των μοντέλων Μηχανικής Μάθησης των μη ισοκατανεμημένων συνόλων δεδομένων αναλύοντας τη σύνδεση που υπάρχει μεταξύ αυτών και των επιμέρους αναγκών και προκλήσεων στις εφαρμογές/έρευνες που έχουν πραγματοποιηθεί μέχρι και σήμερα σε διάφορους τομείς της βιομηχανίας. Σε δεύτερο στάδιο, καθώς η παρούσα Διδακτορική Διατριβή επικεντρώθηκε σε 3 διαφορετικούς τομείς, προς αυτή την κατεύθυνση, περιγράφονται οι σχετικές προκλήσεις, καθώς και οι προσεγγίσεις που έχουν προταθεί για να συμβάλλουν στην αντιμετώπισή τους, οι οποίες επηρέασαν την παρούσα Διδακτορική Διατριβή. Βάσει της εκτεταμένης έρευνας που πραγματοποιήθηκε στα προαναφερθέντα πεδία, η παρούσα Διδακτορική Διατριβή κατέληξε πως κοινό χαρακτηριστικό των περισσότερων εφαρμογών μηχανικής μάθησης σε πραγματικά δεδομένα είναι η ύπαρξη μη ισορροπημένων δεδομένων και η έλλειψη εμπιστοσύνης των «προβλέψεων» από τους χρήστες.

Στο **ΚΕΦΑΛΑΙΟ 3**, αναλύεται η προτεινόμενη προσέγγιση εφαρμογής μοντέλων Μηχανικής Μάθησης σε βιομηχανικές διαδικασίες από την εξαγωγή των δεδομένων μέχρι την παραγωγή των αποτελεσμάτων προς βελτίωση της λήψης αποφάσεων. Ειδικότερα, γίνεται μια λεπτομερής αναφορά στις πιο κομβικές προσεγγίσεις που προτάθηκαν στο πλαίσιο της Διδακτορικής Διατριβής και συνέβαλαν στη δημιουργία της τελικής προτεινόμενης προσέγγισης, σε συνδυασμό με τα αποτελέσματά τους. Παράλληλα, σε κάθε μία από τις κομβικές αυτές προσεγγίσεις καταγράφονται οι εκάστοτε αλλαγές που πραγματοποιήθηκαν στις διαφορετικές προσεγγίσεις, επισημαίνοντας της προκλήσεις που έπρεπε να αντιμετωπίσουμε σε κάθε μία από αυτές. Τέλος, παρατίθεται μία λεπτομερής προσέγγιση των μεθόδων ΧΑΙ που δοκιμάστηκαν στο πλαίσιο της Διατριβής και σε κάθε ένα από τα 4 πεδία εφαρμογή (πειράματα) ενώ αποτυπώνεται ο κοινός τόπος μεταξύ των διαφορετικών προσεγγίσεων σε κάθε εφαρμογή.

Στο **ΚΕΦΑΛΑΙΟ 4**, λαμβάνει χώρα η αξιολόγηση της προτεινόμενης προσέγγισης της εφαρμογής μεθόδων μηχανικής μάθησης σε διαδικασίες της βιομηχανίας επιλύοντας πραγματικά προβλήματα σε πραγματικές συνθήκες μέσω τεσσάρων (4) διαφορετικών πειραμάτων. Η 1<sup>η</sup> περίπτωση που αναφέρεται στη προβλεπτική συντήρηση κύριων μηχανών εμπορικών πλοίων, αποδεικνύοντας την εγκυρότητα εφαρμογής ενός συνδυαστικού μοντέλου αποτελούμενο από διαφορετικά μοντέλα μηχανικής μάθησης και στατιστικά μοντέλα με σκοπό την έγκαιρη πρόβλεψη της διάβρωσης κεφαλής σταυρού του ρουλεμάν της μηχανής των πλοίων. Η 2<sup>η</sup> περίπτωση αποδεικνύει την εφαρμογή με ενθαρρυντικά αποτελέσματα ενός ολοκληρωμένου πλαισίου ανάλυσης δεδομένων αποτελούμενο από μοντέλα μηχανικής μάθησης που εφαρμόζεται για έγκαιρη πρόβλεψη των ανακλήσεων τροφίμων. Όπως και το 2<sup>ο</sup> πείραμα, έτσι και το 3<sup>ο</sup> πείραμα πραγματοποιείται στο πλαίσιο της ασφάλειας τροφίμων, σε αυτή την περίπτωση



εξετάστηκαν οι παράγοντες που μπορεί να οφείλονται για την εμφάνιση του φαινομένου της κηλίδωσης κάπρου. Η 4<sup>η</sup> περίπτωση εξετάζει αντίστοιχα την πρόβλεψη ταμειακών ροών των ΜΜΕ βάσει ιστορικών δεδομένων συναλλαγών. Στο τέλος του Κεφαλαίου, πραγματοποιείται σχετική συζήτηση και ανάλυση των εξαγομένων αποτελεσμάτων όλων των πειραμάτων, παραθέτοντας σχετικά συμπεράσματα για την ολική επίδοση και απόδοση του εκάστοτε μηχανισμού.

Στο **ΚΕΦΑΛΑΙΟ 5**, πραγματοποιείται η σύνοψη της παρούσας Διδακτορικής διατριβής, καταγράφοντας τις βασικές συνεισφορές της αναφορικά με την εφαρμογή μεθόδων μηχανικής μάθησης σε πραγματικά σενάρια της βιομηχανίας. Επιπλέον, περιγράφονται τα ανοιχτά ερευνητικά θέματα και οι μελλοντικοί στόχοι όπως αυτοί προέκυψαν από την τρέχουσα προσέγγιση της ερευνάς, αναφορικά με την ένταξη και την αντιμετώπιση των υπάρχουσών προκλήσεων της εφαρμογής αντίστοιχων τεχνολογιών σε όλους τους τομείς γενικότερα, αλλά και στους τομείς της ναυτιλίας, της ασφάλειας τροφίμων και του τραπεζικού κλάδου ειδικότερα.



## 2 ΚΕΦΑΛΑΙΟ - ΒΙΒΛΙΟΓΡΑΦΙΚΗ ΑΝΑΣΚΟΠΗΣΗ

### ΠΕΡΙΕΧΟΜΕΝΑ

- 2.1 Πρόβλεψη χρονοσειρών
- 2.2 Επεξηγήσιμη Τεχνητή Νοημοσύνη - eXplainable AI (XAI)
- 2.3 Μη ισορροπημένα δεδομένα
- 2.4 Εφαρμογές Μηχανικής Μάθησης

### ΠΕΡΙΛΗΨΗ

Στο ΚΕΦΑΛΑΙΟ 2, αρχικά αναφέρεται η σχετική βιβλιογραφική έρευνα που πραγματοποιήθηκε για διάφορους ερευνητικούς τομείς όπως αυτούς της πρόβλεψης χρονοσειρών, της επεξηγησιμότητας των μοντέλων Μηχανικής Μάθησης των μη ισοκατανεμημένων συνόλων δεδομένων αναλύοντας τη σύνδεση που υπάρχει μεταξύ αυτών και των επιμέρους αναγκών και προκλήσεων στις εφαρμογές/έρευνες που έχουν πραγματοποιηθεί μέχρι και σήμερα σε διάφορους τομείς της βιομηχανίας. Σε δεύτερο στάδιο, καθώς η παρούσα Διδακτορική Διατριβή επικεντρώθηκε σε 3 διαφορετικούς τομείς, προς αυτή την κατεύθυνση, περιγράφονται οι σχετικές προκλήσεις, καθώς και οι προσεγγίσεις που έχουν προταθεί για να συμβάλλουν στην αντιμετώπισή τους, οι οποίες επηρέασαν την παρούσα Διδακτορική Διατριβή. Βάσει της εκτεταμένης έρευνας που πραγματοποιήθηκε στα προαναφερθέντα πεδία, η παρούσα Διδακτορική Διατριβή κατέληξε πως κοινό χαρακτηριστικό των περισσότερων εφαρμογών μηχανικής μάθησης σε πραγματικά δεδομένα είναι η ύπαρξη μη ισορροπημένων δεδομένων και η έλλειψη εμπιστοσύνης των «προβλέψεων» από τους χρήστες.

### 2.1 Πρόβλεψη χρονοσειρών

Στη δεκαετία του 1960, η έννοια της πρόβλεψης χρονοσειρών αναπτύχθηκε αρχικά στη γλώσσα της στατιστικής. Ως τυπική μέθοδος στατιστικής πρόβλεψης, η εκθετική εξομάλυνση ερευνάται διεξοδικά και στη συνέχεια εφαρμόζεται ευρέως σε διάφορα προβλήματα όπως το οι πωλήσεις και η πρόβλεψη κίνησης. Εξάλλου, το μοντέλο Holt-winters, επίσης γνωστό ως εποχιακός αυτοπαλινδρομικός ολοκληρωμένος κινητός μέσος όρος (SARIMA), έχει κυρίαρχη χρήση στη βιομηχανία και την ακαδημαϊκή έρευνα λόγω της ευκολίας της ερμηνείας και επεξηγησιμότητας του αλλά και της αξιοπιστίας του.

Με την πρόοδο της μηχανικής μάθησης (ML), η πρόβλεψη χρονοσειρών αναδιατυπώνεται συχνά ως εργασία παλινδρόμησης και μοντελοποιείται με



μεθόδους μηχανικής μάθησης, όπως η παλινδρόμηση διανύσματος υποστήριξης (SVR) και η ενίσχυση ακραίας κλίσης (XGB). Το [89] εξετάζει την αποτελεσματικότητα και τη σκοπιμότητα της χρήσης SVR για την πρόβλεψη για συμβόλαια μελλοντικής εκπλήρωσης που συγκεντρώνονται από την αγορά εμπορευμάτων του Σικάγου σε σύγκριση με το πολυστρωματικό perceptron (MLP) και το δίκτυο συνάρτησης ακτινικής βάσης (RBF). Στο [90] πραγματοποιείται σύγκριση των μοντέλων SVR και XGB στην παγκόσμια πρόβλεψη ηλιακής ακτινοβολίας και τα αποτελέσματα επικυρώνουν την καλύτερη απόδοση του XGB.

Το βαθιά νευρωνικά δίκτυα (DNN) επιτυγχάνουν ασύγκριτη επιτυχία σε τομείς όπως η ταξινόμηση εικόνων, η ανίχνευση αντικειμένων και μοντελοποίηση ακολουθίας. Λόγω της ισχυρής γενίκευσης και επεκτασιμότητας του, τα βαθιά νευρωνικά δίκτυα γίνονται γρήγορα το κυρίαρχο ερευνητικό επίκεντρο όσον αφορά τη μεθοδολογική καινοτομία των μοντέλων χρονοσειρών. Τα επαναλαμβανόμενα νευρωνικά δίκτυα όπως η μακροπρόθεσμη βραχυπρόθεσμη μνήμη (LSTM) [16], το DeepAR [198] είναι από τη φύση τους εξαιρετικές επιλογές για τη μοντελοποίηση διαδοχικών δεδομένων. Το συνελκτικό νευρωνικό δίκτυο (CNN) επινοήθηκε αρχικά για προβλήματα που σχετίζονται με την εικόνα. Ωστόσο, λόγω της θαυμάσιας απόδοσής του στην εξαγωγή χαρακτηριστικών, κερδίζει δημοτικότητα σε προβλήματα που σχετίζονται με τις χρονοσειρές και οι παραλλαγές του όπως το Temporal Convolution Neural Network (TCNN) [91] τροποποιούνται σκόπιμα για εξαιρετικά μεγάλα δεδομένα χρονοσειρών και ως εκ τούτου πρωτοστατούν σε πολλές πολύπλοκες εφαρμογές χρονοσειρών. Στο [92] προτείνεται ένα μοντέλο βασισμένο στην αρχιτεκτονική των Transformers (μετασχηματιστών) για πρόβλεψη χρονοσειρών μακράς ακολουθίας και αξιοποιούν πλήρως τον μηχανισμό αυτό-προσοχής, ξεπερνώντας τα υπάρχοντα μοντέλα σε τέσσερα σύνολα δεδομένων μεγάλης κλίμακας.

Ένα σύνολο που βασίζεται σε συνδυασμούς θεωρείται συχνά μια σοφή στρατηγική, εκτός από τον υβριδικό τρόπο. Η υπεροχή του συνδυασμού πρόβλεψης έναντι αυτών χωρίς συνδυασμούς αποδεικνύονται στο [93]. Οι λόγοι είναι σε τρεις πτυχές. Πρώτον, η υπόθεση και ο περιορισμός των μεμονωμένων μοντέλων διαφέρουν μεταξύ τους, και ως εκ τούτου είναι δύσκολο να ισχυριστεί κανείς την επάρκεια της βάσης σε ένα μοντέλο για να δώσει ακριβείς προβλέψεις σε διάφορες περιπτώσεις. Δεύτερον, η διαδικασία παραγωγής πίσω από κάθε μοντέλο θα μπορούσε να είναι αμοιβαία συμπληρωματική για να αντισταθμίσει το μειονέκτημα ενός μόνο μοντέλου. Τρίτον, η επιλογή και η στάθμιση του μοντέλου μπορούν να μετριάσουν ακραίες καταστάσεις όπως η κατάρρευση μοντέλου που προκαλείται από δομικά εσφαλμένη προδιαγραφή. Το [94] χρησιμοποιεί μεταπιθανολογικές ασαφείς συναρτήσεις για να βρει την καλύτερη συνάρτηση βάρους και συνάθροισης για να συνδυάσει δεκατρία μοντέλα πρόβλεψης για κάθε χρονοσειρά, η οποία επιτυγχάνει μεγαλύτερη ακρίβεια σε εννέα χρονοσειρές πραγματικού κόσμου.



Η πρόβλεψη χρονοσειρών αποτελείται συνήθως από διάφορα διακριτά στάδια όπως η προεπεξεργασία, η μοντελοποίηση της ακολουθίας, η ανεξάρτητη πρόβλεψη και η συνάθροιση. Επομένως, ένας μεγάλος αριθμός υπερπαραμέτρων απαιτεί συντονισμό για την επίτευξη μεγαλύτερης ακρίβεια πρόβλεψης. Η αναζήτηση πλέγματος (grid search) χρησιμοποιείται συχνά για την προσαρμογή των διαμορφώσεων. Ωστόσο, οι επαναλαμβανόμενες αναζητήσεις χωρίς συσχέτιση μπορεί να είναι εξαιρετικά χρονοβόρες και μεροληπτικές. Όταν αντιμετωπίζουμε το έργο συντονισμού, οι εξελικτικοί και γενετικοί αλγόριθμοι γίνονται οι επιλογές για χάρη της ερμηνευσιμότητας. Το κίνητρο της εφαρμογής του αλγόριθμου βελτιστοποίησης Bayes (BOA) [95] στον συντονισμό υπερπαραμέτρων βρίσκεται σε δύο πτυχές. Πρώτον, το κόστος χρόνου είναι ελεγχόμενο και μπορεί να ρυθμιστεί σύμφωνα με τις ανάγκες μας. Δεύτερον, η ισορροπία εκμετάλλευσης και εξερεύνησης διατηρείται καλά σε σύγκριση με άλλα υπάρχοντα μοντέλα, καθιστώντας την αναζήτηση πιο αποτελεσματική. Στο [96] εφαρμόζεται το BOA για το μοντέλο LSTM. Ενώ το [97] βελτιστοποίησε τη δομή του δικτύου και τη χρονική καθυστέρηση για την πρόβλεψη των PM2.5. Όλες οι εργασίες που αναφέρονται παραπάνω επικυρώνουν τη βελτίωση της εφαρμογής BOA σε σύγκριση με τα βασικά μοντέλα.

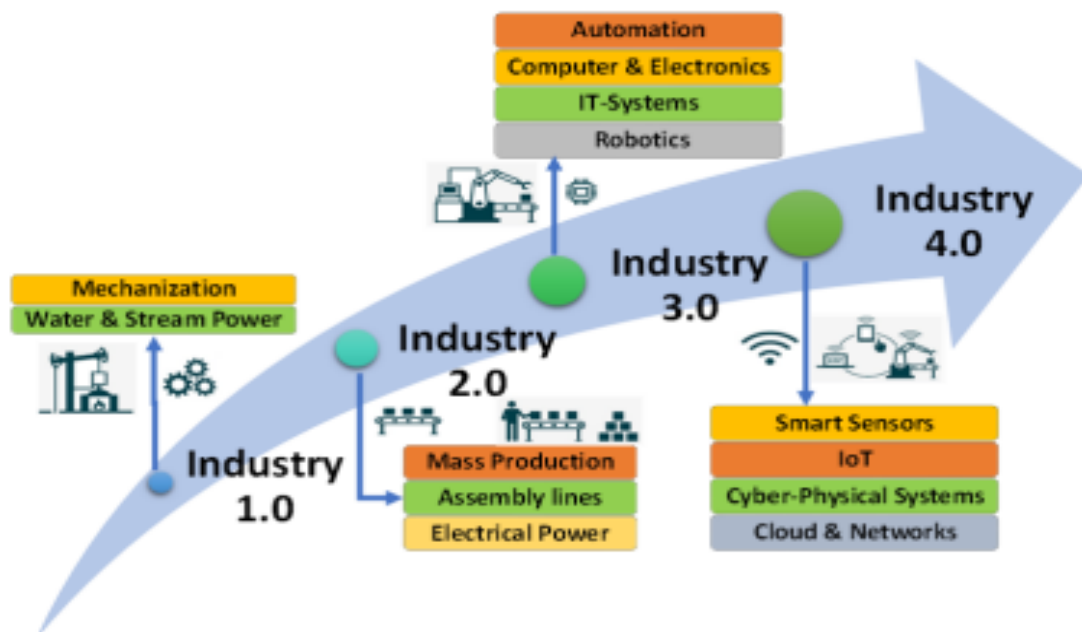
Αν και εμφάνισε μεγάλη επιτυχία, η μέθοδος συνδυασμού προβλέψεων εξακολουθεί να ενέχει προκλήσεις. Το πρώτο μειονέκτημα προκύπτει όταν τα υποψήφια μοντέλα δεν διαφέρουν πολύ. Για παράδειγμα, στο [97] οι συγγραφείς έχουν λάβει υπόψη μόνο τα υποψήφια βαθιά νευρωνικά δίκτυα (DNN) όπως το πολυστρωματικό perceptron (MLP), το συνελκτικό νευρωνικό δίκτυο (CNN) και τη μακροπρόθεσμη βραχυπρόθεσμη μνήμη (LSTM). Δεύτερον, ο συνδυασμός στατικής βέλτιστης πρόβλεψης εκτιμά το βάρος κάθε μοντέλου μέσω της αντιστοιχίας απόδοσης κατά ζεύγη εκπαίδευσης-επικύρωσης, η οποία έχει εξεταστεί θεωρητικά και εμπειρικά αναξιόπιστη. Απλές μέθοδοι στατικής στάθμισης χρησιμοποιούνται συχνά για την πρόβλεψη προβλημάτων όπως οι συνήθεις μέθοδοι ελαχίστου τετραγώνου (OLS), ο αλγόριθμος βασισμένος σε σφάλματα και ο εξίσου σταθμισμένος συνδυασμός μοντέλων (EWM) [98]. Επιπλέον, πρόσφατα οι αλγόριθμοι Ενισχυτικής Μάθησης (Reinforcement Learning (RL)), [99] που αναφέρονται σε αλγόριθμους που είναι «προσανατολισμένοι στον στόχο», άρχισαν να ευδοκίμουν σε ορισμένες εργασίες χρονοσειρών, ειδικά στον χρηματοοικονομικό τομέα. Ως εκ τούτου στο πλαίσιο της Διατριβής έχει διερευνηθεί η πιθανότητα ένα πρόβλημα χρονοσειρών να μοντελοποιηθεί σε μια προσέγγιση RL και αν ναι, να συγκριθεί αυτή η προσέγγιση με τις κυρίαρχες προβλέψεις χρονοσειρών.

## 2.2 Επεξηγήσιμη Τεχνητή Νοημοσύνη - eXplainable AI (XAI)

Η σημερινή εποχή, γνωστή και ως Ψηφιακή Εποχή ή Εποχή της Πληροφορίας, χαρακτηρίζεται από πολύπλοκα υπολογιστικά συστήματα που παράγουν τεράστιες ποσότητες δεδομένων σε καθημερινή βάση. Η συνεχής αύξηση των υπολογιστικών



πόρων και δυνατοτήτων σε σύγκριση με τη μείωση του χρηματικού τους κόστους, έχει οδηγήσει σε ένα κύμα υπολογιστικών μοντέλων που προσπαθούν να λύσουν πολύπλοκα μαθηματικά προβλήματα. Για το σκοπό αυτό, τα πεδία της μηχανικής μάθησης και της Τεχνητής Νοημοσύνης (AI), τα οποία βασίζονται σε ισχυρά στατιστικά μοντέλα, κέρδισαν μεγάλη προσοχή κατά την τελευταία δεκαετία. Η ανάγκη για μεγάλα, πυκνά και πολύπλοκα μοντέλα βαθιάς μάθησης έχει δημιουργήσει ένα ζήτημα εμπιστοσύνης έναντι των προβλέψεων αυτών των μοντέλων, ειδικά όταν χρησιμοποιούνται από ερευνητές που δεν είναι επιστήμονες δεδομένων [100]. Η έννοια της εξήγησης ενός μοντέλου μηχανικής μάθησης ονομάζεται ερμηνευσιμότητα ή επεξήγηση [101]. Αυτή η ανάγκη για ερμηνευτικότητα υπάρχει κυρίως σε μοντέλα βαθιάς μάθησης, καθώς σε πραγματικές εφαρμογές πρέπει να λειτουργούν ως μοντέλα υψηλής απόδοσης που περιέχουν τεράστιο αριθμό (έως χιλιάδες) υπερπαραμέτρων που υποδηλώνουν εξαιρετική εσωτερική πολυπλοκότητα χρησιμοποιώντας μη γραμμικούς μετασχηματισμούς. Η εσωτερική πολυπλοκότητα των Νευρωνικών Δικτύων (NN) αναφέρεται κυρίως ως «μαύρο κουτί» [102] [103]. Πολλές έρευνες αντιμετωπίζουν το πρόβλημα των «μαύρων κουτιών» της μηχανικής μάθησης, καθώς δεν έχουν σαφήνεια ή διορατικότητα, γνωστή και ως επεξηγήσιμη τεχνητή νοημοσύνη (XAI) [104].



Εικόνα 3 Εξέλιξη μηχανικής μάθησης με εφαρμογή στη βιομηχανία

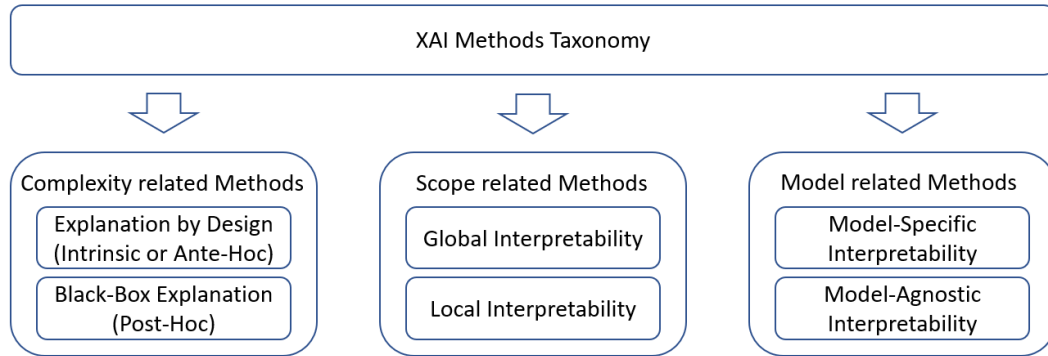
Οι ενδείξεις ότι το πεδίο της επεξηγήσιμης τεχνητής νοημοσύνης έχει γίνει ένας ταχέως αυξανόμενος τομέας είναι εμφανείς τόσο στη βιομηχανία όσο και στην έρευνα. Το 2017, η DARPA χρηματοδότησε το «επεξηγήσιμο πρόγραμμα τεχνητής νοημοσύνης» με στόχο να αυξήσει την επεξήγηση των αποφάσεων τεχνητής νοημοσύνης [105]. Την ίδια χρονιά δημοσιεύτηκε από την κινεζική κυβέρνηση «Το





Σχέδιο Ανάπτυξης για τη Νέα Γενιά Τεχνητής Νοημοσύνης» τονίζοντας την έντονη ανάγκη για αξιόπιστα και εξηγήσιμα μοντέλα τεχνητής νοημοσύνης [106]. Το 2018 ο «Γενικός Κανονισμός Προστασίας Δεδομένων» (GDPR) που δημοσιεύτηκε από την Ευρωπαϊκή Ένωση, επιτρέπει σε κάθε άτομο το δικαίωμα «εξήγησης» σε περίπτωση που τα προσωπικά δεδομένα έχουν χρησιμοποιηθεί σε αλγοριθμικούς χειρισμούς [107]. Οι βιομηχανικές εφαρμογές ενσωματώνουν μια ποικιλία λύσεων τεχνητής νοημοσύνης για να επιτρέψουν στα συστήματα, τις μηχανές και τις συσκευές να «μάθουν» από τα δικά τους δεδομένα και να ενισχύσουν τις ανθρώπινες ικανότητες [108]. Για το σκοπό αυτό, ο τομέας του ΧΑΙ κατέχει έναν πολύ ευαίσθητο, αλλά παρόλα αυτά, βασικό ρόλο στις βιομηχανικές εφαρμογές, καθώς χρησιμεύει ως γέφυρα μεταξύ πολύπλοκων μοντέλων βαθιάς μάθησης και ειδικών που δεν ανήκουν στην Τεχνολογία Πληροφορικής. Για το σκοπό αυτό, οι εξηγήσεις που παρέχονται από τις μεθόδους ΧΑΙ πρέπει να είναι ακριβείς και κατανοητές από ειδικούς διαφόρων τομέων, προκειμένου να αυξηθεί η έννοια της «εμπιστοσύνης» σε ένα βιομηχανικό περιβάλλον σε πραγματικό χρόνο.

Στο [109] παρουσιάζεται η εκτεταμένη έρευνα για τον εντοπισμό τρόπων με τους οποίους οι επαγγελματίες του κλάδου χρησιμοποιούν και αναπτύσσουν μοντέλα ΧΑΙ. Ο εντοπισμός των σφαλμάτων των μοντέλων, η παρακολούθηση, η διαφάνεια και ο έλεγχος ήταν οι πιο κοινές ανάγκες μεταξύ των ειδικών του βιομηχανικού τομέα. Όσον αφορά την αποσφαλμάτωση των μοντέλων, οι ειδικοί απαιτούν υψηλά επίπεδα επεξήγησης προκειμένου να εντοπίσουν σημεία συμφόρησης σε κακή απόδοση μοντέλου, ειδικά όταν υπάρχουν σε χώρους «στατικών» χαρακτηριστικών. Για τον σκοπό αυτό, οι ειδικοί πρέπει να μπορούν να λαμβάνουν γρήγορα και αποτελεσματικά καθοδήγηση σχετικά με τον τρόπο κατασκευής νέων λειτουργιών, απόρριψης περιττών και συλλογής περισσότερων δεδομένων για τη βελτίωση της απόδοσης του μοντέλου. Η παρακολούθηση μοντέλων, ως επί το πλείστον, αναφέρεται σε εξελισσόμενους χώρους χαρακτηριστικών στους οποίους παρατηρούνται μετατοπίσεις εννοιών [110] και οι ειδικοί πρέπει να έχουν εγγενή κατανόηση των μοντέλων μηχανικής μάθησης που χρησιμοποιούνται για την προσαρμογή στις νέες στατιστικές κατανομές που ενδέχεται να προκύψουν. Τέλος, για τον έλεγχο μοντέλων, οι ειδικοί του βιομηχανικού τομέα διεξάγουν διάφορα είδη δοκιμών, ενώ έχει διαπιστωθεί ένα χάσμα μεταξύ των πρακτικών επεξήγησης σε εφαρμογές του πραγματικού κόσμου και του στόχου για διαφάνεια, καθώς οι εξηγήσεις εξυπηρετούν κατά κύριο λόγο τους εσωτερικούς ενδιαφερόμενους.



**Εικόνα 4** Με βάση τις βιομηχανικές εφαρμογές, θα ταξινομήσουμε τις μεθόδους XAI σύμφωνα με την απλότητά τους, το εύρος της ερμηνευσιμότητας και το ποσοστό των εξαρτήσεων από τα αναλυόμενα/χρησιμοποιημένα μοντέλα τεχνητής νοημοσύνης

Στη βιβλιογραφία του πεδίου XAI, πολλές φορές, υπάρχει σύγχυση ή κακή χρήση των όρων ερμηνευσιμότητα και επεξηγησιμότητα. Πρέπει να οριστούν οι διαφορές μεταξύ των δύο όρων για να αποφευχθεί η σύγχυση. Η έννοια της ερμηνευσιμότητας είναι μια αναφορά σε ένα παθητικό χαρακτηριστικό ενός μοντέλου τεχνητής νοημοσύνης που περιγράφει την επέκταση ότι ένα μοντέλο είναι κατανοητό (έχει νόημα) σε έναν ειδικό τομέα. Αντίθετα, ο όρος επεξήγηση χρησιμοποιείται για να περιγράψει ένα ενεργό χαρακτηριστικό ενός δεδομένου μοντέλου τεχνητής νοημοσύνης, υποδηλώνοντας κάθε τύπο δράσης που λαμβάνει ένα μοντέλο για να παράγει υψηλότερα επίπεδα σαφήνειας των εσωτερικών λειτουργιών και της δομής του. Αν και τα τελευταία δύο χρόνια έχουν παρουσιαστεί αρκετές μεθοδολογίες, στρατηγικές και πλαίσια του XAI, για τους σκοπούς αυτής της έρευνας, με βάση τις βιομηχανικές εφαρμογές, θα ταξινομήσουμε τις μεθόδους XAI σύμφωνα με την απλότητά τους, το εύρος της ερμηνευσιμότητας και το ποσοστό των εξαρτήσεων από τα αναλυόμενα/χρησιμοποιημένα μοντέλα τεχνητής νοημοσύνης (Εικόνα 4).

Η έννοια της πολυπλοκότητας συνδέεται άμεσα με την ερμηνευσιμότητα ενός δεδομένου μοντέλου μηχανικής μάθησης. Η ερμηνεία και η επεξήγηση δεδομένων λύσεων μηχανικής μάθησης γίνεται μια πρόκληση καθώς αυτά τα μοντέλα αυξάνουν την πολυπλοκότητα προκειμένου να αντιμετωπίσουν πολυδιάστατα χαρακτηριστικά. Για το σκοπό αυτό, το [111] τόνισε ότι ο πιο σημαντικός τρόπος για να δημιουργηθούν ακριβείς και κατανοητές εξηγήσεις είναι η ανάπτυξη αλγορίθμων μηχανικής μάθησης που είναι εγγενώς ερμηνεύσιμοι. Επιπλέον, οι μέθοδοι που σχετίζονται με την πολυπλοκότητα μπορούν περαιτέρω να διακριθούν σε i) εγγενώς εξηγήσιμα μοντέλα (Ante-Hoc), τα οποία αναφέρονται επίσης ως προσεγγίσεις διαφανούς ή γυάλινου κουτιού και ii) πρόβλεψη μοντέλων μαύρου κουτιού (Post-hoc) που απαιτούν πληροφορίες σχετικά με την πρόβλεψη συλλογιστική διαδικασία σχετικά με την πηγή επεξήγησης.



Τα εγγενή εξηγήσιμα μοντέλα είναι συγκεκριμένα εξ ορισμού και περιλαμβάνουν την λογιστική παλινδρόμηση, τα δέντρα αποφάσεων, τα KNN, και τα μοντέλα Bayes [112]. Στο [113] οι Ustun et al. πρότειναν αραιά γραμμικά μοντέλα που χρησιμοποιούσαν μηχανισμούς βαθμολόγησης βάσει δεδομένων, οι οποίοι επικεντρώνονται στους ειδικούς τομείς ποιοτικής κατανόησης των εσωτερικών μοντέλων λόγω υψηλού επιπέδου αραιότητας. Αν και αυτές οι μέθοδοι τείνουν να είναι διαφανείς από τη φύση τους, όπως επισημαίνεται στο [112], όταν ασχολούμαστε με χώρους χαρακτηριστικών υψηλών διαστάσεων [114] στους οποίους χρειάζεται πιο εξελιγμένα μοντέλα (π.χ. βαθιά δέντρα απόφασης), το επίπεδο διαφάνειας και επεξήγησης πέφτει σημαντικά. Όπως τόνισε ο [115] υπάρχει μια λεπτή γραμμή μεταξύ ακρίβειας και επεξήγησης. Αυτή η πρόκληση μπορεί να θεωρηθεί ως αντιστάθμιση μεταξύ μοντέλων με υψηλά επίπεδα πολυπλοκότητας, και επομένως ακρίβειας, και της προσπάθειας που απαιτείται για την εξήγησή τους στους ειδικούς του τομέα. Το 2001, ο Breiman σημείωσε ότι «η ακρίβεια απαιτεί γενικά πιο σύνθετες μεθόδους πρόβλεψης καθώς οι απλές και ερμηνεύσιμες συναρτήσεις δεν κάνουν τους πιο ακριβείς προγνωστικούς παράγοντες» [116].

Οι πιο πρόσφατες εργασίες που εκτελούνται στις μεθόδους που σχετίζονται με την πολυπλοκότητα εξελίσσονται γύρω από επεξηγήσεις Black-Box, γνωστές και ως Post-Hoc εξηγήσεις. Αυτός ο τύπος επεξηγήσεων είναι ακριβώς το αντίθετο από το Explanations by Design, καθώς αναφέρεται σε πολύπλοκα και δύσκολα ερμηνευτικά μοντέλα χωρίς να είναι γνωστές τις εσωτερικές δομές τους, όπως τα SVMs, και τα νευρωνικά δίκτυα βαθιάς μάθησης Νευρωνικών Δικτύων. Οι εκ των υστέρων μέθοδοι είναι συνήθως ανεξάρτητες από το υποκείμενο μοντέλο μηχανικής μάθησης, και παρόλο που λειτουργούν ως «μαύρα κουτιά» που δεν μπορούν να δώσουν μια ολιστική εξήγηση για ολόκληρο το μοντέλο, αλλά μπορεί να παρέχουν τοπικές εξηγήσεις για συγκεκριμένες αποφάσεις/ταξινομήσεις [117]. Επιπλέον, οι Post-hoc εξηγήσεις μπορούν περαιτέρω να αναλυθούν σε μεθόδους που βασίζονται σε διαταραχές και σε μεθόδους που βασίζονται στην οπισθοδιάδοση. Οι μέθοδοι που βασίζονται σε διαταραχές, όπως η [118] Ανάλυση Διαφοράς Πρόβλεψης (PDA), βασίζονται στη σημασία συγκεκριμένων χαρακτηριστικών που δίνονται σε χώρους χαρακτηριστικών υψηλών διαστάσεων, αλλά δεν μπορούν να χειριστούν κορεσμένους ταξινομητές. Για την αντιμετώπιση του προβλήματος των κορεσμένων ταξινομητών στην επεξεργασία εικόνας [119] Ruth et al. πρότεινε μια παραλλαγή που ονομάζεται Meaningful Perturbations, η οποία αντικατέστησε περιοχές μιας εικόνας με σταθερές τιμές, θόρυβο σε μια εικόνα για να μετρήσει τις αλλαγές στις ενεργοποιήσεις χαρακτηριστικών και τις βαθμολογίες ταξινόμησης. Επέκταση του PDA [120] Zintgraf et al. αφαίρεσε πολλά χαρακτηριστικά ταυτόχρονα χρησιμοποιώντας προηγούμενες γνώσεις σχετικά με εικόνες και επιλέγοντας ενημερωμένες εκδόσεις συνδεδεμένων εικονοστοιχείων ως σύνολα χαρακτηριστικών για την ανάλυση των επιπτώσεων διαφορετικών μεγεθών



παραθύρων στις κορυφαίες κλάσεις βαθμολογίας. Το τεράστιο υπολογιστικό κόστος αυτής της μεθόδου ελαχιστοποιήθηκε αργότερα από τον [121] μέσω της Ανάλυσης Διαφορών Πρόβλεψης Συμφραζομένων, η οποία έλυσε επίσης το πρόβλημα των κορεσμένων ταξινομητών δημιουργώντας έναν χάρτη προεξοχής με επίγνωση του μοντέλου.

Ένας χάρτης προεξοχής [122] είναι μια εικόνα στην οποία η φωτεινότητα ενός εικονοστοιχείου αντιπροσωπεύει πόσο σημαντικό είναι το εικονοστοιχείο, δηλαδή, η φωτεινότητα ενός εικονοστοιχείου είναι ευθέως ανάλογη με την εξέχουσα θέση του. Γενικά είναι μια εικόνα σε κλίμακα του γκρι. Οι χάρτες προεξοχής ονομάζονται επίσης χάρτες θερμότητας (heatmaps) όπου η θερμότητα αναφέρεται σε εκείνες τις περιοχές της εικόνας που έχουν μεγάλο αντίκτυπο στην πρόβλεψη της κατηγορίας στην οποία ανήκει το αντικείμενο. Ο σκοπός του χάρτη προεξοχής είναι να βρει τις περιοχές που είναι εμφανείς σε κάθε θέση στο οπτικό πεδίο και να καθοδηγήσει την επιλογή των τοποθεσιών που παρακολουθούνται, με βάση τη χωρική κατανομή της εξέχουσας θέσης. Η διαφορά μεταξύ ενός προεξέχοντος χάρτη και ενός χάρτη θερμότητας έγκειται στο γεγονός ότι ένας χάρτης εξέχουσας θέσης εμφανίζει ιδιότητες της εικόνας (δηλαδή την πιθανότητα προσέγκυσης της προσοχής με βάση τα χαρακτηριστικά από κάτω προς τα πάνω), ενώ ένας χάρτης θερμότητας είναι μια αναπαράσταση δεδομένων συμπεριφοράς βλέμματος. Έτσι, βασίζονται σε διαφορετικά δεδομένα: η εικόνα = ανάδειξη ή η συμπεριφορά του βλέμματος = χάρτης θερμότητας.

Παρατηρώντας τους χάρτες προεξοχής:

(α) Μπορούμε να εντοπίσουμε μοτίβα που μπορούν να χρησιμοποιηθούν για να δημιουργήσουμε παραδείγματα τα οποία να ξεγελάει το μοντέλο για να προβλέψει τη λάθος ετικέτα.

(β) Μπορούμε να βρούμε το λόγο που αποτυγχάνει ένα μοντέλο μηχανικής μάθησης (λάθος προβλέψεις).

Δεδομένου ότι υπάρχουν πολλές προσεγγίσεις επεξήγησης και μπορούν να παράγουν αρκετά διαφορετικούς χάρτες εξέχουσας σημασίας, είναι σημαντικό να γνωρίζουμε τις ιδιότητές τους, όπως φαίνεται στο [122]. Οι πιο εξέχουσες ιδιότητες των χαρτών εξέχουσας σημασίας είναι το «Faithfulness» και «Stability». Δεδομένου ότι οι τεχνικές επεξήγησης χρησιμοποιούνται για να εξηγήσουν τις προβλέψεις μοντέλων για μία μόνο περίπτωση, μια βασική ιδιότητα είναι ότι είναι πιστές στις εσωτερικές λειτουργίες του μοντέλου και δεν βασίζονται σε αυθαίρετες επιλογές. Ένας καθιερωμένος τρόπος μέτρησης αυτής της ιδιότητας είναι η αντικατάσταση ορισμένων από τις πιο σημαντικές λέξεις με μία «μάσκα» και η παρατήρηση της πτώσης στην απόδοση του μοντέλου. Η έννοια της σταθερότητας μπορεί να εξηγηθεί με ένα παράδειγμα, θα μπορούσε κανείς να θεωρήσει ότι δύο περιπτώσεις έχουν παρόμοια λογική, εάν η είσοδος είναι παρόμοια και η παραγόμενη έξοδος είναι η ίδια. Μια πιο συνεπής προσέγγιση θα ήταν επίσης η μέτρηση της ομοιότητας



μεταξύ των χαρτών ενεργοποίησης στα ξεχωριστά επίπεδα, η οποία δεν ελήφθη υπόψη για υπολογιστικούς λόγους. Την εν λόγω τακτική την εξετάσαμε στο πλαίσιο της διατριβής.

Στο [124], οι Yang et al. πρότειναν το Παγκόσμιο μοντέλο Ερμηνείας μέσω Αναδρομικής Διαμέρισης (GIRP), το οποίο δημιουργεί ένα ευρύτερο δέντρο ερμηνείας για μια ποικιλία μοντέλων μηχανικής μάθησης με βάση τοπικές εξηγήσεις. Το GIRP προσδιορίζει εάν τα μοντέλα μηχανικής μάθησης τείνουν να ταιριάζουν υπερβολικά σε συγκεκριμένα μοτίβα και ειδοποιεί τους ειδικούς του τομέα σε περιπτώσεις παράλογης συμπεριφοράς. Στο [125] παρουσιάζεται η μέθοδος Deep Learning Important Features (DeepLIFT), που προτείνεται από τους Shrikumar et al. και χρησιμοποιεί μεθοδολογίες που βασίζονται σε παράγωγα για να διαδώσει τις διαφορές ενεργοποίησης αντί για τα gradients του μοντέλου. Αν και οι μερικές παράγωγοι δεν εξηγούν μια μεμονωμένη απόφαση, υποδεικνύουν ποιες τροποποιήσεις του χώρου χαρακτηριστικών θα μεταδίδονταν στις διαφορές αποτελέσματος των μοντέλων. Στο [126] συναντάμε μια προσέγγιση Ολοκληρωμένων Διαβαθμίσεων που βασίζεται στον υπολογισμό των αποδόσεων πολλαπλασιάζοντας τα χαρακτηριστικά εισόδου με τη μέση μερική παράγωγο, καθώς ο χώρος των χαρακτηριστικών εισόδου μπορεί να είναι αραιός. [127] Η χαρτογράφηση ενεργοποίησης κλάσης (CAM), η οποία παρουσιάστηκε από τους Zhou et al., βασίστηκε στην παρατήρηση ότι ορισμένα συνελκτικά στρώματα συμπεριφέρονται ως ανιχνευτές αντικειμένων χωρίς επίβλεψη και χρησιμοποιεί τη συγκέντρωση του παγκόσμιου μέσου όρου για να δημιουργήσει χάρτες θερμότητας ενός επιπέδου πριν από το επίπεδο της softmax. Οι χάρτες θερμότητας που δημιουργούνται υπογραμμίζουν τις περιοχές ενός χώρου χαρακτηριστικών που είναι πιο υπεύθυνες για μια εργασία ταξινόμησης. [128] Η χαρτογράφηση ενεργοποίησης κλάσης με στάθμιση κλίσης (GradCAM), η οποία είναι μια επέκταση του CAM, χρησιμοποιεί τις πληροφορίες κλίσης για να ταξινομήσει την ενεργοποίηση των νευρώνων στο τελευταίο στρώμα ενός Συνελκτικού Νευρωνικού Δικτύου (CNN).

Η χρήση μεθόδων επεξήγησης για την ερμηνεία των λόγων για τους οποίους ελήφθη μια συγκεκριμένη απόφαση (π.χ. εργασία ταξινόμησης ενός συστήματος ανίχνευσης ακραίων τιμών), υποδηλώνει ότι η εργασία επεξήγησης λαμβάνει χώρα τοπικά. Οι αγνωστικές λύσεις μοντέλων δεν ορίζονται από έναν συγκεκριμένο τύπο μοντέλου μηχανικής μάθησης, καθώς διαχωρίζουν τα αποτελέσματα ταξινόμησης από τις επεξηγήσεις. Φτάνουμε στο πολύ διαδεδομένο μοντέλο XAI το Local Interpretable Model-Agnostic Explanation (LIME) [129], που προτείνεται από τους Ribeiro et al. και προσεγγίζει ένα μοντέλο μαύρου κουτιού τοπικά στην «περιοχή» οποιασδήποτε πρόβλεψης που θα ήθελαν να εστιάσουν οι ειδικοί ενός τομέα. Το μοντέλο μαθαίνει μια συγκεκριμένη πρόβλεψη που σχετίζεται με την τοπική περιοχή, αντιστοιχίζοντας το δεδομένο διάνυσμα χαρακτηριστικών και τις διαταραγμένες εισόδους, με τα αποτελέσματα που λαμβάνονται από το μοντέλο



αναφοράς. Στο [130] οι δημιουργοί του LIME πρότειναν μια επέκταση του αρχικού μοντέλου, χρησιμοποιώντας κανόνες απόφασης. Δεδομένου ότι η τοπική συμπεριφορά του μοντέλου μπορεί να είναι μη γραμμική, οι συγγραφείς προτείνουν τη χρήση ενός συνόλου κανόνων if-then, οι οποίοι είναι εύχρηστοι και κατανοητοί. Παρόμοια με την επέκταση του LIME που παρουσιάστηκε από τους Ribeiro et al., το Local Rule-Based Explanations (LoRE) [131] όπου προτείνει μια μεθοδολογία δύο βημάτων χωρίς παραμέτρους, η οποία παρέχει επίσης εξηγήσεις βάσει κανόνων. Δημιουργεί ένα ισορροπημένο σύνολο γειτονικών παρουσιών χρησιμοποιώντας έναν γενικό αλγόριθμο για να διερευνήσει το όριο απόφασης των σημείων δεδομένων ενδιαφέροντος και δημιουργεί έναν ταξινομητή δέντρου αποφάσεων, ο οποίος επιτρέπει την απόκτηση κανόνων απόφασης. Μια άλλη ιδιαίτερα διαδεδομένη μέθοδος είναι η SHapley Additive Explanations (SHAP) [72] που εισήχθη από τους Lundberg και Lee, χρησιμοποιεί ένα ενιαίο μέτρο της σημασίας των χαρακτηριστικών που βασίζεται στις τιμές Shapley, μια έννοια από τη θεωρία παιγνίων. Τα πολλαπλά μοντέλα επεξήγησης που προτείνονται από το SHAP διαφέρουν ως προς τον τρόπο με τον οποίο προσεγγίζουν τον υπολογισμό των τιμών του SHAP. Λειτουργεί δημιουργώντας μια λεπτομερή αξιολόγηση για κάθε στοιχείο εισόδου σε ένα μοντέλο μηχανικής μάθησης. Αυτό γίνεται παρέχοντας μια συνάρτηση αξιολόγησης για κάθε στοιχείο εισόδου που αναπαριστά την επιρροή που έχει στην απόφαση του μοντέλου. Για παράδειγμα, μπορεί να δείξει ότι ένα στοιχείο εισόδου έχει μια ισχυρή θετική επιρροή στην απόφαση του μοντέλου, ενώ άλλα στοιχεία εισόδου έχουν μια θετική ή αρνητική επιρροή σε μικρότερο βαθμό. Έτσι, οι χρήστες μπορούν να κατανοήσουν και να ελέγξουν την απόφαση του μοντέλου, και να αξιολογήσουν την ακρίβεια και την σαφήνεια της. Στον τομέα της ταξινόμησης εικόνων μπορούν να χρησιμοποιηθούν δύο κύριες εξηγήσεις για δίκτυα βαθιάς μάθησης. [72] DEEPSHAP και [132] Gradient-SHAP: [72] Το DEEPSHAP είναι ένας αλγόριθμος προσέγγισης υψηλής ταχύτητας αφιερωμένος σε μοντέλα βαθιάς μάθησης που είναι επίσης συνδεδεμένος με το [125] DeepLift. Σε αυτό το σενάριο, η διαφορά από το αρχικό μοντέλο DeepLift [125] έγκειται στη χρήση μιας βασικής κατανομής δειγμάτων φόντου αντί για μια ενιαία τιμή και στη χρήση εξισώσεων Shapley για τη γραμμικοποίηση μη γραμμικών στοιχείων του «μαύρου κουτιού». Το Gradient-SHAP βασίζεται στους αλγόριθμους [133] IntGrad και [134] SmoothGrad. Οι τιμές IntGrad απαιτούν μια ενιαία τιμή αναφοράς για ενσωμάτωση. Ως προσαρμογή σε κατά προσέγγιση τιμές SHAP, το Gradient-SHAP επαναδιαμορφώνει το ολοκλήρωμα ως προσδοκία και συνδυάζει αυτήν την προσδοκία με δειγματοληπτικές τιμές αναφοράς από το σύνολο δεδομένων φόντου, όπως γίνεται στο SmoothGrad.

Έχουν εφαρμοστεί αρκετές τεχνικές και μέθοδοι επεξήγησης στον τομέα της κατασκευής και πιο συγκεκριμένα στον τομέα της προγνωστικής διαχείρισης ποιότητας (Ποιότητα 4.0) για την ενίσχυση της διαφάνειας των αναπτυσσόμενων μοντέλων AI. Οι Goldman et al. χρησιμοποίησε τεχνικές [135] XAI, όπως CAM και



Contrastive gradient-based χάρτες προεξοχής για να εξηγήσει τους ταξινομητές μαύρου κουτιού σε συγκολλήσεις ποιότητας σε γλωττίδες μπαταρίας συγκολλημένες με υπερήχους. Οι Lee et al. Ο [136] εφάρμοσε διάφορες μεθόδους ΧΑΙ για να παρέχει εξηγήσεις στους ειδικούς του τομέα στην ταξινόμηση ελαττωμάτων πάνελ οθόνης υγρών κρυστάλλων με τρανζίστορ λεπτής μεμβράνης. Τεχνικές όπως CAM, LRP, και SmoothGrad εφαρμόστηκαν και οπτικοποιήθηκαν σε ένα μοντέλο ταξινόμησης το VGG-16. Με βάση τα οπτικοποιημένα αποτελέσματα, επιλέχθηκαν το LRP για τη δημιουργία κατανεμημένων χαρτών θερμότητας. Προσαρμόζοντας το μοντέλο σε ένα δέντρο αποφάσεων και μετατρέποντας τα αποτελέσματα της πρόβλεψης σε ανθρώπινο ερμηνεύσιμο κείμενο, οι συγγραφείς πέτυχαν ένα αυξημένο επίπεδο επεξήγησης, όπως επιβεβαιώθηκε από μια σειρά ειδικών του τομέα που πραγματοποίησαν πολλαπλές αξιολογήσεις. Στον τομέα της εκτίμησης του κόστους κατασκευής, ο [137] περιέγραψε μια μέθοδο που βασίζεται στην απεικόνιση των χαρακτηριστικών μηχανικής επεξεργασίας ενός τρισδιάστατου μοντέλου σχεδίασης με τη βοήθεια υπολογιστή που επηρεάζουν την αύξηση του κόστους κατασκευής. Για τον προτεινόμενο σκοπό, εφαρμόστηκε μια τρισδιάστατη χαρτογράφηση ενεργοποίησης κλάσεων με στάθμιση βαθμίδας ως μέθοδος ΧΑΙ.

### 2.3 Μη ισορροπημένα δεδομένα

Οι πρώτες εργασίες για μη ισορροπημένα δεδομένα προήλθαν από προβλήματα δυαδικής ταξινόμησης. Εδώ, υποτίθεται η παρουσία τάξεων πλειοψηφίας και μειοψηφίας, με συγκεκριμένη αναλογία ανισορροπίας. Τέτοιες μη ισορροπημένες κατανομές κλάσεων αποτελούν πρόκληση για τα μοντέλα μηχανικής μάθησης, καθώς οι τυπικοί ταξινομητές οδηγούνται από μια συνάρτηση απώλειας που προϋποθέτει ομοιόμορφη ποινή και για τις δύο κατηγορίες. Επομένως, οποιαδήποτε διαδικασία μάθησης καθοδηγείται από μια τέτοια συνάρτηση θα οδηγήσει σε μια προκατάληψη προς την τάξη της πλειοψηφίας. Ταυτόχρονα, η μειονοτική τάξη είναι συνήθως πιο σημαντική και επομένως δεν μπορεί να αναγνωριστεί ελάχιστα. Ως εκ τούτου, οι μέθοδοι που είναι αφιερωμένες στην αντιμετώπιση του προβλήματος της ανισορροπίας στοχεύουν είτε στην άμβλυση της ανισορροπίας είτε στην εναλλαγή της διαδικασίας μάθησης. Οι δύο βασικές προσεγγίσεις που βασίζονται στα δεδομένα θα πρέπει να θεωρείται ως μια φάση προ-επεξεργασίας που είναι ανεξάρτητη από ταξινομητή. Εδώ, εστιάζουμε στην εξισορρόπηση του συνόλου δεδομένων πριν από την εφαρμογή οποιασδήποτε εκπαίδευσης ταξινομητή. Αυτό συνήθως επιτυγχάνεται με έναν από τους τρεις τρόπους:

- 1) μείωση του μεγέθους της πλειοψηφικής τάξης (υποδειγματοληψία)
- 2) αύξηση του μεγέθους της μειονοτικής τάξης (υπερδειγματοληψία) ή
- 3) συνδυασμό των δύο προηγούμενων λύσεων (υβριδική προσέγγιση).



Τόσο η υπο-δειγματοληψία όσο και η υπερ-δειγματοληψία μπορούν να πραγματοποιηθούν με τυχαίο τρόπο, ο οποίος έχει χαμηλή πολυπλοκότητα, αλλά οδηγεί σε δυνητικά ασταθή συμπεριφορά (π.χ. αφαίρεση σημαντικών περιπτώσεων ή ενίσχυση θορυβωδών). Ως εκ τούτου, έχουν προταθεί καθοδηγούμενες λύσεις που προσπαθούν να επιλέξουν έξυπνα στιγμιότυπα για προ-επεξεργασία.

### 2.3.1 Υπερδειγματοληψία

Υπάρχουν πολλές τεχνικές κατά της ανισορροπίας της τάξης που περιλαμβάνουν προσεγγίσεις υπερδειγματοληψίας, όπως η τυχαία υπερδειγματοληψία (ROS), η τεχνική συνθετικής μειονότητας έναντι της δειγματοληψίας (SMOTE) [138] και η τεχνική προσαρμοστικής συνθετικής δειγματοληψίας (ADASYN)[139]. Αν και το ROS είναι μια πιο απλή μέθοδος, παρέχει καλά αποτελέσματα σε πολλές περιπτώσεις, επομένως δεν πρέπει να αγνοηθεί [140]. Όπως υποστηρίζουν οι συντάκτες μιας συνολικής έρευνας [141], θα πρέπει να διερευνηθούν διαφορετικές μέθοδοι για κάθε πρόβλημα για την αντιμετώπιση της ανισορροπίας κλάσης, καθώς δεν υπάρχει καθολική καλύτερη προσέγγιση που να ταιριάζει σε όλα. Το SMOTE και το ADASYN είναι οι δύο μέθοδοι που χρησιμοποιήθηκαν στο πλαίσιο της έρευνάς μας χωρίς σημαντικές διαφορές στην απόδοση ως προς την αξιολόγηση στο σύνολο δοκιμών εκτός δείγματος. Στις τεχνικές υπερ-δειγματοληψίας, μια σημαντική ανησυχία είναι η υπερβολική προσαρμογή από τον εκπαιδευόμενο (δηλαδή, που οδηγεί σε κακή απόδοση γενίκευσης) και η αύξηση του μεγέθους των δεδομένων εκπαίδευσης. Ωστόσο, όπως προτείνεται από το [142], το SMOTE αποφεύγει αυτή την παγίδα αφού δημιουργεί συνθετικά νέες περιπτώσεις, σε σύγκριση με την επικάλυψη των υπάρχοντων παρουσιών, το ίδιο ισχύει και για την ADASYN.

### 2.3.2 Υποδειγματοληψία

Στην περίπτωση των τεχνικών υποδειγματοληψίας, ένα θέμα ανησυχίας είναι η διαγραφή πολύτιμων πληροφοριών εάν αφαιρεθεί ένας σχετικά μεγάλος αριθμός περιπτώσεων από την πλειοψηφική κατηγορία. Η ουσιαστική αφαίρεση παρουσιών από την κλάση πλειοψηφίας μπορεί επίσης να αλλάξει την κατανομή της πλειοψηφικής κλάσης (και επομένως του συνολικού συνόλου δεδομένων) που αντιπροσωπεύει τον συγκεκριμένο τομέα. Ως εκ τούτου, είναι σημαντικό να εξεταστούν διαφορετικές μέθοδοι δειγματοληψίας δεδομένων για ανισορροπία κλάσεων [141]. Παρόλο που στη βιβλιογραφία προτείνονται πολλές τεχνικές που βασίζονται σε σύνολα, όπως το SMOTEBagging, το RUSBoost και το UnderBagging, οι μέθοδοι επιλογής χαρακτηριστικών για την ανισορροπία τάξης είναι ένας σε μεγάλο βαθμό ανεξερεύνητος τομέας έρευνας. Επιπλέον, θα πρέπει να αναφερθεί ότι, όπως προτείνεται από τον [143], η Επαναληπτική Τυχαία





Δειγματοληψία (IRS) είναι πιο αξιόπιστη από τη δειγματοληψία Kennard-Stone (KS) στη δειγματοληψία αντιπροσωπευτικού εκπαιδευτικού συνόλου, επομένως πειραματιστήκαμε μόνο με μεθόδους IRS σε αυτήν τη μελέτη.

## 2.4 Εφαρμογές Μηχανικής Μάθησης

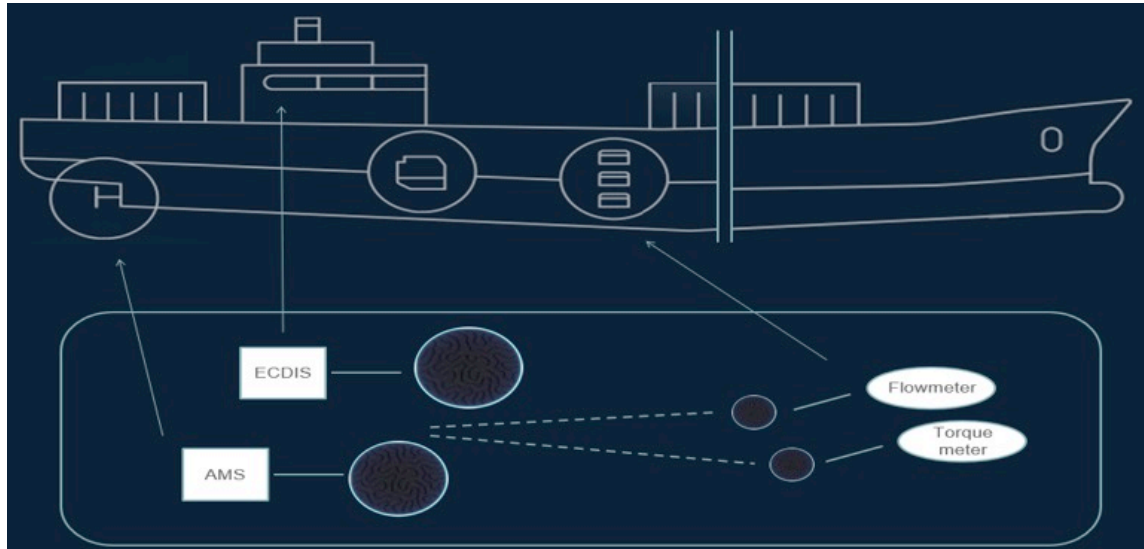
### 2.4.1 Μηχανική Μάθηση & προβλεπτική συντήρηση κύριων μηχανών εμπορικών πλοίων.

Η χρήση σε διαφορετικούς τομείς, τύπους δεδομένων και τύπους ανωμαλιών, δείχνει ότι ο Εντοπισμός Ανωμαλιών είναι ένας ενεργός χώρος έρευνας [145]. Ενώ οι προσεγγίσεις κατηγοριοποίησης είναι αρκετά δημοφιλείς, φαίνονται μη ρεαλιστικές σε πραγματικές περιπτώσεις χωρίς αρκετή ποσότητα επισημασμένων δεδομένων [144]. Τα Netflix και Twitter πρότειναν δυο υποσχόμενες μεθόδους εντοπισμού ανωμαλιών, βασισμένες στην Ισχυρή Ανάλυση Βασικών Συνιστωσών και Εποχιακή Υβριδική Υπερβολική Μαθητοποιημένη Απόκλιση αντίστοιχα [146], [147]. Φυσικά υπάρχουν πολλές άλλες προσεγγίσεις κατάλληλες για εντοπισμό ανωμαλιών [144]. Μερικές εξ αυτών έχουν εφαρμοστεί στην ναυτιλιακή βιομηχανία τα τελευταία χρόνια. Για παράδειγμα, έχει προταθεί μια μεθοδολογία όπου δεδομένα δονήσεων συνδυάζονται με δεδομένα απόδοσης (πιέσεις κυλίνδρων) για τις συνθήκες παρακολούθησης της κύριας μηχανής. Συνεπώς, σχολιάστηκαν και συγκρίθηκαν δυο διαφορετικές προσεγγίσεις που βασίζονται σε θερμοδυναμικά μοντέλα ώστε να εντοπίσουν δυο συνήθεις καταστάσεις βλάβης μεγάλων ντίζελ μηχανών δεδομένων των ιχνών των πιέσεων κυλίνδρων [149]. Επίσης οι Ραπτοδήμος και Λαζάκης ερεύνησαν την δυνατότητα του SOM (Self-Organising Map) να παρακολουθήσει τις συνθήκες της κύριας μηχανής, εντοπίζοντας συστάδες δεδομένων που αποκλίνουν από την φυσιολογική λειτουργία της μηχανής. Ακόμα οι κ.κ. Λαζάκης, Γκερέκος και Θεοτοκάτος [151] πρότειναν ένα SVM-οδηγούμενο, μιας κλάσης προσέγγιση ώστε να εκτιμηθεί η κατάσταση των συστημάτων ενός πλοίου, μια μεθοδολογία που λαμβάνει υπόψη της και η παρούσα εργασία. Οι Kowalski et.al [152] χρησιμοποίησαν το ένα εναντίον ενός σε Υπερβολική Μηχανική Μάθηση (Extreme Learning Machine) ώστε να τα ταξινομήσουν τα δεδομένα από 14 διαφορετικές βλάβες και έναν σωστό τρόπο λειτουργίας.

Λαμβάνοντας υπόψη περιπτώσεις όπου ο εντοπισμός βλάβης μπορεί να είναι χρήσιμος, πρέπει να συνυπολογιστεί και το ότι στην πραγματική ζωή υπάρχει επίσης θόρυβος και τα δεδομένα έχουν πολλές διατάσεις κάτι που κάνει τον εντοπισμό ανωμαλιών ακόμα πιο δύσκολο. Επίσης, σε διάφορα πεδία, τα δεδομένα (δηλαδή η διαφορετικές παράμετροι στα σύνολα δεδομένων) σχετίζονται μεταξύ τους σε μεγάλο βαθμό, κάτι που πρέπει να συνυπολογιστεί. Έτσι, για την εφαρμογή κάποιων μοντέλων θα πρέπει να χρησιμοποιηθεί είτε η μέση τιμή τους, είτε η



Ανάλυση Κύριων Συστατικών (Principal Component Analysis (PCA)) [153] με σκοπό τη μείωση των διαστάσεων των δεδομένων. Δεδομένου ότι τα προαναφερθέντα δεδομένα που προέρχονται κυρίως από αισθητήρες είναι συνήθως στη μορφή χρονοσειράς, θα πρέπει να ελεγχθεί η στατικότητα των δεδομένων χρησιμοποιώντας τον έλεγχο Dickey-Fuller. Αυτό είναι ουσιώδες ώστε να εφαρμοστεί οποιαδήποτε μέθοδος πρόβλεψης που βασίζεται σε ιστορικά δεδομένα, καθώς οι μέθοδοι μηχανικής μάθησης συνήθως βασίζονται στην υπόθεση ότι ο μηχανισμός παραγωγής δεδομένων δεν αλλάζει με το χρόνο [154].



Εικόνα 5 Γενικό διάγραμμα συστήματος DAQ

#### 2.4.2 Μηχανική Μάθηση & ασφάλεια των τροφίμων.

Στο θέμα της βιομηχανίας τροφίμων και ιδιαίτερα των ανακλήσεων τροφίμων, μια κινητήρια έρευνα έλαβε χώρα στην Ιατρική Σχολή του Πανεπιστημίου της Βοστώνης, όπου ένας αλγόριθμος τεχνητής νοημοσύνης βασισμένος στην αρχιτεκτονική BERT κατάφερε να ανιχνεύσει μη ασφαλή προϊόντα διατροφής με βάση τις κριτικές πελατών του Amazon με ακρίβεια σε ποσοστό 74% [155]. Εκτός από αυτό, υπάρχει μια προσέγγιση NER βασισμένη σε κανόνες για την εξαγωγή πληροφοριών τροφίμων, με το όνομα FoodIE. Αυτή αποτελείται από έναν μικρό αριθμό κανόνων υπολογιστικής γλωσσολογίας και σημασιολογικών πληροφοριών, που περιγράφουν τις οντότητες των τροφίμων [156]. Σε πρόσφατη έρευνα που πραγματοποιήθηκε και παρουσιάστηκε στο [157], περιγράφεται μια εκτενής σύγκριση μεταξύ αυτοματοποιημένων μεθόδων εξαγωγής για πληροφορίες τροφίμων. Στο [158] για τις πληροφορίες προϊόντος, όπως τίτλοι και περιγραφές, αξιοποιήθηκαν για την ανακάλυψη τιμών που λείπουν από τα χαρακτηριστικά του προϊόντος, αναπτύσσοντας ένα νέο μοντέλο βαθιάς προσθήκης ετικετών OpenTag με F1 score 83%. Επίσης, όπως παρουσιάζεται στο [159] διεξήχθη έρευνα για την



πρόβλεψη εάν η FDA θα προβλέψει ανάκληση ενός συγκεκριμένου φαρμάκου σε χρονικό ορίζοντα που κυμαίνεται από μία (1) έως σαράντα (40) ημέρες στο μέλλον, χρησιμοποιώντας χαρακτηριστικά που ποσοτικοποιούν την αλλαγή στον όγκο ερωτημάτων στο διαδίκτυο. Μια εξαιρετική επισκόπηση και σύγκριση του SoTA στον τομέα την Αναγνώρισης Ονομαστική Οντότητα τόσο όσον αφορά τις προσεγγίσεις όσο και τα σύνολα δεδομένων μπορεί να βρεθεί στο [160] όπου οι συγγραφείς ανέπτυξαν μια νέα μέθοδο εξαγωγής οντοτήτων και σχέσεων μεταξύ τους. Όλες αυτές οι δημοσιεύσεις μπορούν να δικαιολογήσουν το κίνητρο πίσω από την προσέγγισή μας, προτείνοντας ότι τα συγκεντρωτικά δεδομένα μηχανών αναζήτησης στο Διαδίκτυο μπορούν να χρησιμοποιηθούν για τη διευκόλυνση της έγκαιρης προειδοποίησης για ελαττωματικές παρτίδες τροφίμων.

### 2.4.3 Μηχανική Μάθηση & ποιότητα χοιρινού κρέατος.

Πολυάριθμες μελέτες σχετικά με τις επιδράσεις του ευνουχισμού σε κάπρους υποδηλώνουν ότι η εκτροφή των κάπρων για παραγωγή κρέατος παρουσιάζει μεγάλα οικονομικά πλεονεκτήματα (για παράδειγμα αυτές που πραγματοποιήθηκαν από τους Huber et al. [161], Backus and McGlone [162] και Gautam et al. [163]). Σε μικρό αριθμό περιπτώσεων, ωστόσο, ορισμένα σφάγια κάπρου εμφανίζουν έντονη δυσάρεστη οσμή και γεύση όταν ψήνεται το κρέας («χρώση κάπρου») [164]. Επειδή η «κηλίδωση του κάπρου» συνδέεται με τις ορμόνες των όρχεων και τη σεξουαλική ωριμότητα, ο χειρουργικός ευνουχισμός έχει εφαρμοστεί ιστορικά ως προληπτικό μέτρο, με το πρόσθετο πλεονέκτημα της πιο πειθήνιας συμπεριφοράς. Η έρευνα [165] απέδειξε ότι οι προσεγγίσεις που βασίζονται στον ευνουχισμό για την πρόληψη της αντικατάστασης της κηλίδωσης από κάπρους οδήγησαν σε γενικευμένες σε αυξημένες ανησυχίες για την καλή διαβίωση των ζώων, ενώ οι καταναλωτές δείχνουν επίσης αυξανόμενη ανησυχία για το θέμα. Εναλλακτικές προσεγγίσεις για την πρόληψη της κηλίδωσης στους κάπρους, όπως η εφαρμογή ενός εμβολίου GnRH (π.χ. Impronac) από τον [166] είναι διαθέσιμες αλλά δεν είναι ευρέως διαδεδομένες. Μια ανασκόπηση από τον [167] εστιάζει σε μελέτες σχετικά με τη χρήση της διατροφής, της αναπαραγωγής και της διαχείρισης για την εξάλειψη ή τη μείωση της κηλίδωσης του κάπρου, με μια πολλά υποσχόμενη μελέτη, από τον [168], με βάση ένα υποσύνολο των δεδομένων που χρησιμοποιήθηκαν στην διατριβή μας. Συνοπτικά, οι ακόλουθοι παράγοντες φαίνεται να συσχετίζονται με την εμφάνιση κηλίδας κάπρου: δερματικές αλλοιώσεις στη φάρμα, δερματικές αλλοιώσεις στο σφαγείο, διαφορά μεταξύ ζώντων δερματικών αλλοιώσεων και δερματικών αλλοιώσεων σφαγείου, ποσοστό άπαχου κρέατος, μέση θερμοκρασία περιβάλλοντος την ημέρα της σφαγής και διάρκεια αναμονής πριν τη σφαγή.

### 2.4.4 Μηχανική Μάθηση & τραπεζικός βοηθός για μικρομεσαίες επιχειρήσεις.



Όσον αφορά την κατηγοριοποίηση συναλλαγών στον τραπεζικό κλάδο, υπάρχουν πολλές εμπορικές εφαρμογές που βασίζονται σε προσεγγίσεις ταξινόμησης μηχανικής μάθησης, εστιάζοντας κυρίως σε λύσεις κατηγοριοποίησης προσωπικών συναλλαγών. Μια σχετική έρευνα προέρχεται από το [169], όπου χρησιμοποιήθηκαν σύντομα κείμενα προς ένα νέο σύστημα που συνδυάζει τεχνικές Επεξεργασίας Φυσικής Γλώσσας με αλγόριθμους μηχανικής μάθησης για την ταξινόμηση των περιγραφών τραπεζικών συναλλαγών για χρήση σε μια εφαρμογή Διαχείρισης Προσωπικών Οικονομικών (PFM). Χρησιμοποιήθηκε ένα επισημασμένο σύνολο δεδομένων με πραγματικές συναλλαγές πελατών, αξιοποιώντας υπάρχουσες λύσεις στον εντοπισμό ανεπιθύμητων μηνυμάτων, προτείνοντας έναν ανιχνευτή ομοιότητας σύντομου κειμένου για τη μείωση του μεγέθους του συνόλου εκπαίδευσης. Ενώ στο [170] προτείνεται μια νέα προσέγγιση που χρησιμοποιεί δεδομένα συναλλαγών πελατών με στόχο την παρακολούθηση της πρόβλεψης της συμπεριφοράς των πελατών χωρίς καμία μη αυτόματη επισήμανση των δεδομένων. Για την επίτευξη αυτού του στόχου, στοιχεία των δεδομένων τραπεζικών συναλλαγών αναπαρίστανται αυτόματα σε έναν χώρο ενσωμάτωσης υψηλών διαστάσεων ως συνεχή διανύσματα. Σε αυτόν τον νέο χώρο, οι αποστάσεις μεταξύ των διανυσματικών θέσεων είναι μικρότερες για τα στοιχεία με παρόμοια οικονομική σημασία. Ομοίως, οι αποστάσεις μεταξύ των μη σχετιζόμενων στοιχείων είναι μεγαλύτερες, κάτι που αποδεικνύεται χρήσιμο στην αυτόματη αποτύπωση των σχέσεων μεταξύ των στοιχείων χρηματοοικονομικών συναλλαγών χωρίς καμία χειροκίνητη παρέμβαση. Στην ίδια κατεύθυνση, μια άλλη έρευνα [171] παρουσιάζει μια νέα εφαρμογή εκμάθησης αναπαράστασης σε διμερή γραφήματα συναλλαγών με πιστωτικές κάρτες προκειμένου να μάθουμε ενσωματώσεις λογαριασμών και εμπορικών οντοτήτων.

Πρόσφατα, η έρευνα για την κατηγοριοποίηση συναλλαγών περιστρέφεται γύρω από την αναγνώριση κειμένου. Πρόσφατες προσεγγίσεις προσπαθούν να εξαγάγουν πληροφορία από το μη δομημένο κείμενο της περιγραφής συναλλαγών. Μια αντιπροσωπευτική έρευνα είναι [172] όπου προτείνεται ένα πλαίσιο για την κατανόηση των μεταβαλλόμενων τάσεων και τη λήψη κατάλληλων αποφάσεων με βάση την περιγραφή των συναλλαγών. Μια άλλη σχετική πρόταση προέρχεται από το [173], όπου ένα σύστημα που βασίζεται στη μηχανική μάθηση αυτοματοποιεί τη διαδικασία χαρτογράφησης οικονομικών μεταφορών στους αντίστοιχους λογαριασμούς.

Επιπλέον, περισσότερες έρευνες που βασίζονται σε πιο παραδοσιακά εποπτευόμενα μοντέλα μπορούν να βρεθούν στη σχετική βιβλιογραφία όπως [174] και [175]. Η πρώτη προσέγγιση από τις δύο διερευνά μεθόδους που βασίζονται στην εποπτευόμενη μάθηση για να συναγάγει την ταυτότητα του Χρηματοοικονομικού Ινστιτούτου αποστολής από τη συμβολοσειρά περιγραφής μιας συναλλαγής μεταφοράς χρημάτων, χρησιμοποιώντας ένα μείγμα παραδοσιακών μεθόδων επεξεργασίας φυσικής γλώσσας με βάση ένα Recurrent



Neural Network. Ενώ η τελευταία παρουσιάζει μια έρευνα που διεξήχθη στο Νορβηγικό Πανεπιστήμιο Επιστήμης και Τεχνολογίας προσπαθώντας να απαντήσει σε 3 ερευνητικά ερωτήματα, συμπεριλαμβανομένης της σύγκρισης διαφορετικών μοντέλων μηχανικής μάθησης, αξιολόγηση εξωτερικών σημασιολογικών πόρων όπως η νορβηγική βάση δεδομένων Brønnpøysundregisteret και το Google Places API ή συνδεδεμένες πηγές ανοιχτών δεδομένων όπως το DBPedia και WikiData. Τα κύρια αποτελέσματα των προσεγγίσεων υποδηλώνουν ότι οι σημασιολογικοί πόροι βελτιώνουν την ακρίβεια του συστήματος ταξινόμησης, ενώ τα Wikidata και DBpedia οδηγούν σε μείωση της ακρίβειας [175]. Εκτός από την έρευνα που αναφέρεται εδώ, υπάρχουν επίσης διάφορα διπλώματα ευρεσιτεχνίας της Google (π.χ. "Ένα σύστημα διαχείρισης συναλλαγών" και "Κατηγοριοποίηση μη συνεχόμενων συναλλαγών"), συμπεριλαμβανομένου ενός συστήματος βάσης δεδομένων που έχει διαμορφωθεί για λήψη και αποθήκευση δεδομένων για πολλές οικονομικές συναλλαγές και μια μέθοδο για βελτιωμένη διαχείριση δεδομένων συναλλαγών από δίκτυο υπολογιστών χρηματοοικονομικών υπηρεσιών. Από την άλλη πλευρά από την βιβλιογραφική έρευνα προκύπτει ότι δεν υπάρχει έρευνα ή εργαλείο κατηγοριοποίησης συναλλαγών για MME στη σχετική βιβλιογραφία.

Όσον αφορά την εργασία πρόβλεψης χρονοσειρών, που είναι ένας από τους πιο ενδιαφέροντες τομείς στην ερευνητική βιβλιογραφία, στο [176], το μοντέλο Prophet της Facebook (πλέον Meta), παρουσιάζεται ένα σπονδυλωτό μοντέλο πρόβλεψης παλινδρόμησης με ερμηνεύσιμες παραμέτρους που μπορούν να προσαρμοστούν διαισθητικά από αναλυτές με γνώση του τομέα σχετικά με τις χρονοσειρές. Στο [177] το ίδιο μοντέλο συγκρίνεται με το DeepAR στο πρόβλημα πρόβλεψης της ζήτησης τροφίμων, παρουσιάζοντας υποσχόμενα αποτελέσματα. Βέβαια πρέπει να αναφερθεί ότι αποτελέσματα δεν ήταν τόσο ικανοποιητικά στην προσπάθεια μας να προβλέψουμε τις εισροές και εκροές συναλλαγών σε ένα από τα σενάρια μας.

Εκτιμάται ότι η απόδοση του DeepAR στον τομέα αυτόν οφείλεται στις ομαδοποιήσεις ανά κατηγορία όπως εφαρμόζονται στην εργασία υβριδικής ταξινόμησης, όπου ένα παγκόσμιο μοντέλο που λαμβάνει υπόψη όλες τις ομάδες κατηγοριοποίησης πριν από την εφαρμογή της πρόβλεψης χρονοσειρών ξεπερνά τα άλλα μοντέλα που παρουσιάζονται στη σχετική βιβλιογραφία. Προς αυτή την κατεύθυνση, μια συστηματική ανασκόπηση 117 εγγράφων που σχετίζονται με χρονοσειρές, έρχεται να ενισχύσει αυτόν τον ισχυρισμό [178]. Στην ίδια ανασκόπηση, μια σύγκριση μεταξύ ορισμένων από τις πιο διαδεδομένες προσεγγίσεις για την πρόβλεψη χρονοσειρών έδειξε ότι η SARIMA είναι η μόνη στατιστική μέθοδος που μπορεί να έχει καλύτερη απόδοση (σε ορισμένες περιπτώσεις) τους ακόλουθους αλγόριθμους μηχανικής μάθησης: ANN, SVM και kNN-TSPI, αλλά χωρίς στατιστικά σημαντική διαφορά. Ωστόσο, η υψηλότερη ακρίβεια πρόβλεψης έρχεται σε βάρος ενός μεγαλύτερου αριθμού παραμέτρων [178]. Λαμβάνοντας υπόψη το τελευταίο, τις απαιτήσεις υψηλής κλίμακας και τα



πολύ διαφορετικά μεγέθη των χρονοσειρών των ΜΜΕ, το μοντέλο SARIMA δεν επιλέχθηκε. Ως εκ τούτου, καμία στατιστική μέθοδος δεν χρησιμοποιήθηκε στην έρευνά μας, αντίθετα, μοντέλα του πλαισίου GluonTS [179] χρησιμοποιήθηκαν ως πιθανές λύσεις βαθιάς μάθησης για την πρόβλεψη χρονοσειρών.

Επιπλέον, θα πρέπει να σημειωθεί ότι για την πρόβλεψη χρονοσειρών, το έργο της χρήσης παρατηρούμενων χρονοσειρών στο παρελθόν για την πρόβλεψη τιμών σε έναν ορίζοντα προοπτικής γίνεται αναλογικά δυσκολότερο καθώς αυτός ο ορίζοντας διευρύνεται [180]. Η κύρια διαφορά της προτεινόμενης εργασίας σε σχέση με τις προαναφερθείσες είναι ότι τόσο το μοντέλο κατηγοριοποίησης συναλλαγών όσο και το μοντέλο πρόβλεψης ταμειακών ροών περιστρέφονται γύρω από τις συγκεκριμένες ανάγκες των ΜΜΕ και τις συγκεκριμένες επιχειρηματικές τους ανάγκες. Επιπλέον, οι δύο προσεγγίσεις που αναπτύχθηκαν είναι αλληλένδετες με αλλαγές στη διαδικασία κατηγοριοποίησης που αντικατοπτρίζονται στο μοντέλο πρόβλεψης ταμειακών ροών. Συγκεκριμένα, αυτό το έγγραφο αντιμετωπίζει την αναδυόμενη πρόκληση της αυτοματοποιημένης επιχειρηματικής χρηματοοικονομικής διαχείρισης για τις ΜΜΕ, εισάγοντας μια προσέγγιση βασισμένη σε ιστορικά δεδομένα για τη διευκόλυνση της έγκαιρης λήψης αποφάσεων.



## 3 ΚΕΦΑΛΑΙΟ - ΑΡΧΙΤΕΚΤΟΝΙΚΗ ΣΧΕΔΙΑΣΗ

### ΠΕΡΙΕΧΟΜΕΝΑ

#### 3.1 Γενικευμένη προσέγγιση

3.1 Μηχανική Μάθηση & προβλεπτική συντήρηση κύριων μηχανών εμπορικών πλοίων.

#### 3.2 Μηχανική Μάθηση & ασφάλεια των τροφίμων.

#### 3.3 Μηχανική Μάθηση & ποιότητα χοιρινού κρέατος.

#### 3.4 Μηχανική Μάθηση & τραπεζικός βοηθός για μικρομεσαίες επιχειρήσεις.

#### 3.5 Επεξηγήσιμη Τεχνητή Νοημοσύνη - eXplainable AI (XAI)

### ΠΕΡΙΛΗΨΗ

Στο **ΚΕΦΑΛΑΙΟ 3**, αναλύεται η προτεινόμενη προσέγγιση εφαρμογής μοντέλων Μηχανικής Μάθησης σε βιομηχανικές διαδικασίες από την εξαγωγή των δεδομένων μέχρι την παραγωγή των αποτελεσμάτων προς βελτίωση της λήψη αποφάσεων. Ειδικότερα, γίνεται μια λεπτομερής αναφορά στις πιο κομβικές μεθοδολογίες που προτάθηκαν στο πλαίσιο της Διδακτορικής Διατριβής και συνέβαλαν στη δημιουργία της τελικής προτεινόμενης προσέγγισης, σε συνδυασμό με τα αποτελέσματά τους. Παράλληλα, σε κάθε μία από τις κομβικές αυτές μεθοδολογίες καταγράφονται οι εκάστοτε αλλαγές που πραγματοποιήθηκαν στις διαφορετικές προσεγγίσεις, επισημαίνοντας της προκλήσεις που έπρεπε να αντιμετωπίσουμε σε κάθε μία από αυτές. Τέλος, παρατίθεται μία λεπτομερής προσέγγιση των μεθόδων επεξηγήσιμης τεχνητής νοημοσύνης (XAI) που δοκιμάστηκαν στο πλαίσιο της Διατριβής και σε κάθε ένα από τα 4 πεδία εφαρμογή (πειράματα) ενώ αποτυπώνεται ο κοινό τόπος μεταξύ των διαφορετικών προσεγγίσεων σε κάθε εφαρμογή.

### 3.1 Γενικευμένη Προσέγγιση

Η προσέγγιση μας στην επεξεργασία και την ανάλυση δεδομένων στηρίζεται σε μια γενικευμένη διαδικασία που ξεκινά με την προεπεξεργασία των ακατέργαστων δεδομένων (raw data). Αυτή η αρχική φάση προεπεξεργασίας μπορεί να περιλαμβάνει τον καθαρισμό των δεδομένων από θόρυβο ή ατέλειες, την κανονικοποίηση ή μετασχηματισμό των δεδομένων για να καταστούν πιο εύκολα αναλύσιμα, και την εξαγωγή χαρακτηριστικών για να καταστεί πιο συνεκτική και χρήσιμη η πληροφορία που παρέχουν τα δεδομένα.

Ακολουθεί η φάση της μηχανικής μάθησης, όπου χρησιμοποιούμε τα προεπεξεργασμένα δεδομένα για να κατηγοριοποιήσουμε ή να προβλέψουμε επιθυμητές μεταβλητές. Αυτό επιτυγχάνεται μέσω ενός μοντέλου μηχανικής

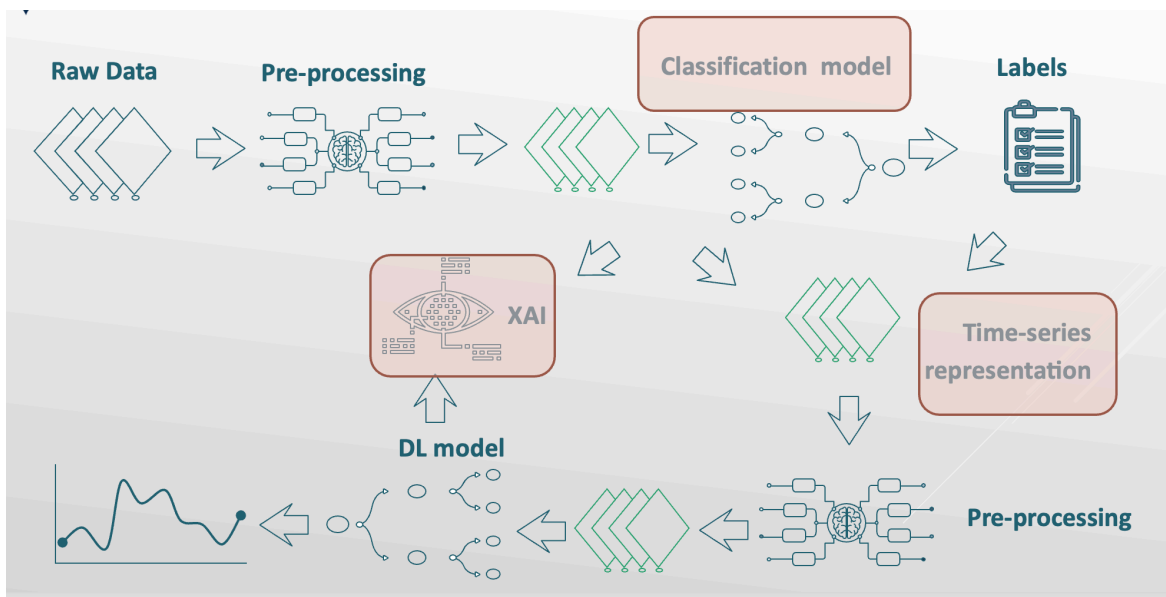


μάθησης που έχει εκπαιδευτεί στα δεδομένα μας, είτε μέσω επιβλεπόμενης μάθησης (όπου το μοντέλο μαθαίνει να αντιστοιχίζει είσοδο-έξοδο από τα δεδομένα εκπαίδευσης) ή μέσω μη επιβλεπόμενης μάθησης (όπου το μοντέλο μαθαίνει να εξαγάγει δομές ή σχέσεις από τα δεδομένα χωρίς ειδικές οδηγίες).

Τα αποτελέσματα της μηχανικής μάθησης, όπως η κατηγοριοποίηση, αναπαρίστανται στη συνέχεια σε χρονοσειρές, παρέχοντας μια δυναμική απεικόνιση των χρονικών αντίκτυπων των προβλέψεων ή των κατηγοριοποιήσεων. Αυτό είναι ιδιαίτερα χρήσιμο για την ανάλυση δεδομένων που εξελίσσονται με τον χρόνο, καθώς μας επιτρέπει να εντοπίσουμε τάσεις, κύκλους ή άλλα χρονικά μοτίβα.

Επιπρόσθετα, αυτές οι χρονοσειρές εισάγονται σε ένα μοντέλο για την πρόβλεψη χρονοσειρών, το οποίο επιτρέπει την πρόβλεψη μελλοντικών τιμών ή ενεργειών με βάση τα παρελθόντα δεδομένα. Αυτό είναι πολύ χρήσιμο για πολλές εφαρμογές, όπως η πρόβλεψη των πωλήσεων, η πρόβλεψη των τιμών των αγαθών, η πρόβλεψη των καιρικών συνθηκών και πολλά άλλα.

Τέλος, στην προσέγγισή μας εφαρμόζεται η εξηγήσιμη τεχνητή νοημοσύνη (XAI) στα μοντέλα μηχανικής μάθησης που χρησιμοποιούνται. Η XAI συμβάλλει στην κατανόηση του τρόπου λειτουργίας των μοντέλων, προσφέροντας επεξηγήσεις για τις προβλέψεις ή τις αποφάσεις τους. Αυτό μπορεί να βοηθήσει τους ανθρώπους να κατανοήσουν καλύτερα τις προβλέψεις ή τις κατηγοριοποιήσεις που παράγει το σύστημα, να εντοπίσουν πιθανά λάθη ή προβλήματα, και να εμπιστευτούν περισσότερο τις αποφάσεις που λαμβάνονται με βάση αυτές τις προβλέψεις ή κατηγοριοποιήσεις. Η εξηγήσιμη τεχνητή νοημοσύνη είναι ένα σημαντικό εργαλείο για την επίτευξη ενός πιο διαφανούς και δίκαιου συστήματος μηχανικής μάθησης.



Εικόνα 6 Αρχιτεκτονική προτεινόμενης γενικευμένης προσέγγισης





### 3.1.1 Συνεισφορά Διατριβής

Στη διατριβή μας, αποτυπώνουμε μια σειρά από καινοτόμες τεχνικές και προσεγγίσεις στην ανάλυση χρονοσειρών, την επεξεργασία δεδομένων και την ανάλυση κειμένου, επικεντρωμένες σε εφαρμογές της τεχνητής νοημοσύνης και της μηχανικής μάθησης.

Πρώτα από όλα, καταπιαστήκαμε με τον σχεδιασμό ενός νέου περιβάλλοντος ενισχυτικής μάθησης (RL) για την πρόβλεψη χρονοσειρών. Με τη χρήση αυτής της καινοτόμου προσέγγισης, καταφέραμε να επιτύχουμε βελτιωμένες προβλέψεις χρονοσειρών, οι οποίες ξεπερνούν τις τρέχουσες μεθόδους. Αυτό διευκολύνει την εξαγωγή εποικοδομητικών εισόδων για πολύπλοκες αποφάσεις και ενισχύει την αποτελεσματικότητα των πολιτικών λήψης αποφάσεων.

Εκτός αυτού, προσεγγίζουμε την πρόκληση των "μικρών" δεδομένων μέσω της χρήσης ανάλογων (συνθετικών) δεδομένων, γνωστών ως surrogate data, για τον εμπλουτισμό των δεδομένων που έχουμε στη διάθεσή μας. Με την εφαρμογή αυτής της τεχνικής, καταφέραμε να παράγουμε περισσότερα χρήσιμα και εμπλουτισμένα δεδομένα για την εκπαίδευση των μοντέλων μας, βελτιώνοντας σημαντικά την απόδοση και τη γενίκευση τους.

Επίσης, αναπτύξαμε ένα μετα-μοντέλο για την επιλογή του κατάλληλου μοντέλου μηχανικής μάθησης βάσει χαρακτηριστικών μιας χρονοσειράς. Με αυτή την προσέγγιση, επιτρέπουμε στην διαδικασία επιλογής μοντέλου να γίνεται αυτόματη, πετυχαίνοντας μεγαλύτερη ακρίβεια και αποτελεσματικότητα στην πρόβλεψη χρονοσειρών.

Στο πεδίο της εξηγήσιμης AI (XAI), εφαρμόσαμε τεχνικές εξαιρετικά αυτόματης αναγνώρισης συμπτωμάτων σε χρονοσειρές, προσφέροντας περισσότερη διαφάνεια και κατανόηση για τη λειτουργία και τις αποφάσεις των μοντέλων μας. Επιπλέον, χρησιμοποιήσαμε την XAI ως μέτρο αντιμετώπισης επιθέσεων reasoning στα δεδομένα, ενισχύοντας την ασφάλεια και την αξιοπιστία των μοντέλων μας.

Αναφορικά με την ταξινόμηση, αναπτύξαμε μια custom προσέγγιση για την υπερδειγματοληψία και την επιλογή χαρακτηριστικών βασισμένη σε Γενετικό Αλγόριθμο. Το αποτέλεσμα ήταν η παραγωγή ενός πιο ισορροπημένου και αντιπροσωπευτικού συνόλου δεδομένων, το οποίο βελτίωσε την απόδοση της ταξινόμησης.

Τέλος, στον τομέα της αναγνώρισης ονοματισμένων οντοτήτων (NER), εφαρμόσαμε τεχνικές για την αναγνώριση προϊόντων τροφίμων σε κείμενα. Αυτό άνοιξε τον δρόμο για πιο εξειδικευμένες και πληροφοριακά πλούσιες αναλύσεις



κειμένου, βοηθώντας στην ανακάλυψη νέων συσχετίσεων και παρατηρήσεων στα δεδομένα.

### 3.2 Μηχανική Μάθηση & προβλεπτική συντήρηση κύριων μηχανών εμπορικών πλοίων.

Ο σκοπός μας είναι να χρησιμοποιήσουμε μεθόδους μηχανικής μάθησης και στατιστικές προσεγγίσεις ώστε να εντοπίσουν νωρίς μη φυσιολογικές λειτουργίες σχετικά με ελαττώματα του ρουλεμάν σταυρού κεφαλής. Η αρχική υπόθεση μας είναι ότι η εφαρμογή καινοτόμων τεχνικών θα έχει βελτιωμένα αποτελέσματα σε περιπτώσεις που παραδοσιακές μέθοδοι αποτυγχάνουν.

#### 3.2.1 Σύνολο Δεδομένων

Τα δεδομένα που χρησιμοποιήθηκαν στην εν λόγω έρευνα διατέθηκαν από μια ελληνική ναυτιλιακή εταιρεία που συγκεντρώνει ροές δεδομένων από όλους τους αισθητήρες που χρησιμοποιούνται στα πλοία ιδιοκτησίας της. Η διάθεση των δεδομένων πραγματοποιήθηκε στο πλαίσιο ενός ευρωπαϊκού ερευνητικού προγράμματος HORIZAON 2020. Στα δεδομένα, κάθε ροή παριστάνει τις εγγραφές που συλλέγονται με το χρόνο από έναν συγκεκριμένο αισθητήρα στο πλοίο. Αναφέρεται ότι οι τιμές των αισθητήρων δειγματοληπτούνται σε χρονική περίοδο ενός λεπτού ή δέκα λεπτών, ανάλογα τον τύπο των μετρήσεων του αισθητήρα. Στην περίπτωση μας (εφαρμογή μεθόδων μηχανικής μάθησης) τα δεδομένα αυτά χρησιμοποιούνται σαν είσοδοι των μοντέλων αναφέρονται και ως «χαρακτηριστικά» στο και παρουσιάζονται περιληπτικά στον Πίνακας 1. Τα δεδομένα αποτελούν ιστορικά δεδομένα από 10 αδερφά πλοία συμπεριλήφθηκαν όπου τα δεδομένα για την πρόβλεψη εξάχθηκαν μέσω on-line DAQ (Data AcQuisition)(σύστημα ανάκτησης δεδομένων).

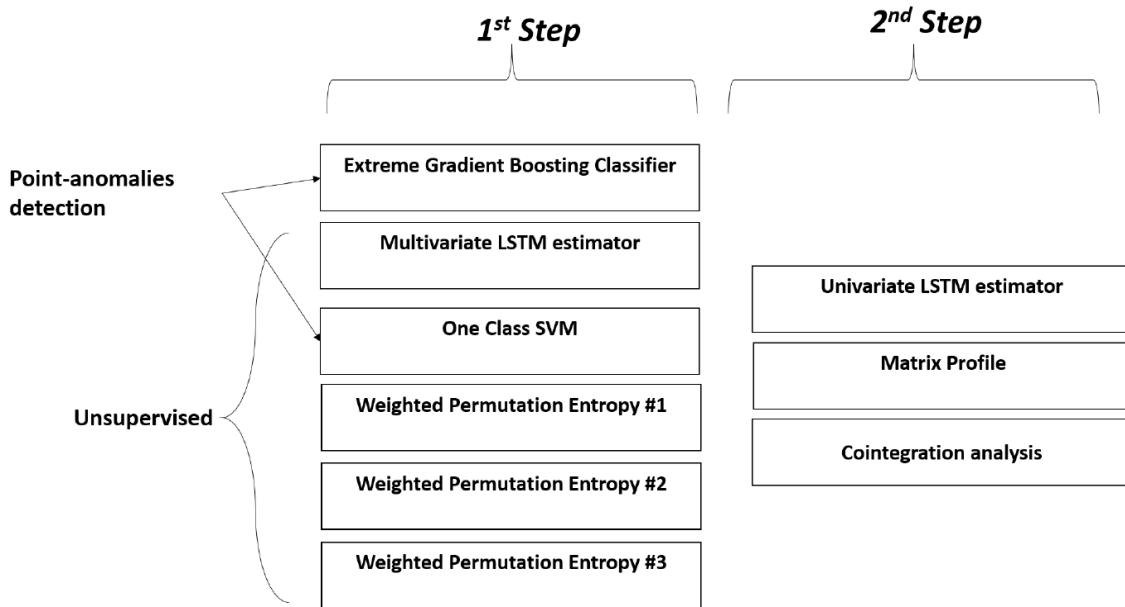
**Πίνακας 1** Χαρακτηριστικά συνόλου δεδομένων από αισθητήρες πλοίων κύριας μηχανής (main engine data), και γενικά δεδομένα του πλοίου (vessel data) που χρησιμοποιήθηκαν στην περίπτωση της προβλεπτικής συντήρησης.

VESSEL DATA table structure	MAIN ENGINE DATA table structure
Consumed power (kW)	Air Cooler Cooling Water Inlet Pressure (Pa)
Wind-speed (kn)	Cylinder #1-10 Scavenge Air Fire Detection Temperature (°C)
GPS speed (kn)	Cooling Fresh Water Inlet Pressure (Pa)
Speed through water – longitudinal (kn)	Control Air Pressure (Pa)
Speed through water – transverse (kn)	Exhaust Valve Spring Air Inlet Pressure (Pa)
Rotations per minute of the main shaft	Fuel Oil Flow Rate (lt)



### Κεφάλαιο 3 Αρχιτεκτονική Σχεδίαση

Wind angle (0-359.99 degrees)	Fuel Oil Inlet Pressure (Pa)
Vessel draft at mid-line (m)	Fuel Oil Inlet Temperature (°C)
Speed through water (kn)	Jacket Cooling Fresh Water Inlet Temperature Low (°C)
The latitude of the vessel's position	Cylinder #1-10 Exhaust Gas Out Temperature (°C)
The longitude of the vessel's position	Cylinder #1-10 Jacket Cooling Fresh Water Outlet Temper. (°C)
Total number of containers	Cylinder #1-10 Piston Cooling Outlet Temperature (°C)
Vessel draft at starboard-side (right-side looking to the fore)	Turbocharger #1-3 Exhaust Gas Inlet Temperature (°C)
Vessel draft at stern (m)	Turbocharger #1-3 Exhaust Gas Outlet Temperature (°C)
Vessel draft at fore (m)	Turbocharger #1-3 Lube Oil Inlet Pressure (Pa)
High-sulfur fuel oil consumption (tones)	Turbocharger #1-3 Lube Oil Outlet Pressure (Pa)
Low-sulfur fuel oil consumption (tones),	Turbocharger #1-3 RPMs
The trim of the vessel	Scavenge Air Inlet Pressure (Pa)
Vessel draft at port-side (m)	Scavenge Air Receiver Temperature (°C)
Total Twenty-foot Equivalent Unit (TEU) (# of containers)	Starting Air Pressure (Pa)
Total Forty-foot Equivalent Unit (FEU) (# of containers)	Thrust Pad Temperature (°C)
	Main Lube Oil Inlet Pressure (Pa)
	Main Lube Oil Inlet Temperature (°C)
	Fuel Oil Temperature (°C)
	Torque of the main shaft (N/m)
	Turbocharger #1-3 Air Cooler Cooling Water Outlet Temperature °C)



Εικόνα 7 Αλγόριθμοι μηχανικής μάθησης που χρησιμοποιήθηκαν ή δοκιμάστηκαν για την περίπτωση της προβλεπτικής συντήρησης κυρίων μηχανών εμπορικών πλοίων.

### 3.2.2 Αρχιτεκτονική

Στη συγκεκριμένη περίπτωση, προσεγγίζουμε το πρόβλημα σε δύο βήματα όπως παρουσιάζονται στην Εικόνα 7. Εκπαιδεύσαμε διάφορα μοντέλα μηχανικής και βαθιάς μάθησης, συμπεριλαμβανομένων των *LSTM*, *One Class SVM*, *XGBoost*, και υπολογισμός εντροπίας μετάθεσης, με σκοπό την εκτέλεση μεθόδων είτε ταξινόμησης είτε παλινδρόμησης προβλέποντας την αναμενόμενη τιμή για κάθε μοντέλο την επόμενη χρονική στιγμή (δηλ. το χρονικό βήμα). Δοκιμάσαμε τόσο μοντέλα μηχανικής μάθησης με επιτήρηση (π.χ., *XGBoost*) για τα οποία απαιτείτε μια πρόσθετη εργασία τεχνητής ταυτοποίησης δεδομένων όσο και μη εποπτευόμενα μοντέλα (π.χ., *LSTM*, *One Class SVM*), τα οποία χρησιμοποιήθηκαν, μαζί με κυλιόμενη εντροπία συνδυασμών/μετάθεσης. Δευτερευόντως, δημιουργήθηκαν κανόνες εντοπισμού ανωμαλιών, οι οποίοι διαφοροποιούνται ανάλογα με τη χρησιμοποιούμενη μεθοδολογία, το κοινό τους χαρακτηριστικό είναι η σύγκριση των πραγματικών με τις εκτιμώμενες τιμές. Όπως σε κάθε εφαρμογή μεθόδων μηχανικής μάθησης, ένα στάδιο προ-επεξεργασίας των δεδομένων εξετάστηκε ως προς την αποτελεσματικότητα του και εφαρμόστηκε αντίστοιχα. Η διαδικασία αυτή είναι κρίσιμη σε κάθε εφαρμογή μοντέλων μηχανικής μάθησης και συνήθως θεωρείται το πιο χρονοβόρο στάδιο.

Επιπλέον, ένα σύνολο πρόσθετων τεχνικών μελετήθηκαν και εφαρμόστηκαν, έτσι ώστε να ταυτοποιηθεί η σχετική πιθανότητα της υπό εξέταση βλάβης του ρουλεμάν σταυρού κεφαλής σε συγκεκριμένο κύλινδρο. Αυτές οι πρόσθετες τεχνικές συμπεριλαμβάνουν μονο-μεταβλητές *LSTM*, την πολύπλοκη μέθοδο *Matrix*



Profile (MP) και Cointegration Analysis (Ανάλυση Συνολοκλήρωσης). Στην προκειμένη περίπτωση, τα δεδομένα που αναλύθηκαν μέσω αυτών των προσεγγίσεων συλλέχθηκαν αποκλειστικά από αισθητήρες (θερμοκρασίας και πίεσης) που είναι εγκατεστημένες σε κάθε κύλινδρο των μηχανών των πλοίων ξεχωριστά. Στην περίπτωση αυτή η μηδενική υπόθεση που εξετάσαμε είναι ότι κάθε βλάβη είναι πιο προφανής στο συγκεκριμένο κύλινδρο που έχει διαβρωθεί. Οπότε, ούτε οι γενικές ροές δεδομένων των πλοίων, ούτε οι ροές δεδομένων της κύριας μηχανής χρησιμοποιήθηκαν στις μεθόδους αυτές, παρά μόνο τα δεδομένα από τους κυλίνδρους.

### 3.2.3 Προ-επεξεργασία Δεδομένων

Η προ-επεξεργασία των δεδομένων αποτελεί ένα από τα πιο σημαντικά στάδια στη διαδικασία εξόρυξης γνώσης από συλλογές δεδομένων. Στην υποενότητα αυτή παρουσιάζονται τα βασικά προβλήματα και οι αντίστοιχες τεχνικές αντιμετώπισης τους που σχετίζονται με την προ-επεξεργασία των δεδομένων. Γίνεται αναφορά στο καθαρισμό των δεδομένων, στην συμπλήρωση/ολοκλήρωση τους και στο μετασχηματισμό τους εάν αυτός απαιτείται. Επίσης παρουσιάζονται μεταξύ άλλων μέθοδοι μείωσης διαστάσεων και επιλογής σημαντικών χαρακτηριστικών τα οποία άπτονται στην εκμάθηση χωρίς επίβλεψη και στην επεξηγήσιμη τεχνητή νοημοσύνη αντίστοιχα.

Τα χαρακτηριστικά συλλογής δεδομένων μπορεί να περιέχουν ασυνεπή ή λανθασμένα δεδομένα (πχ. λόγω ασυνέπειας καταγραφής, ανθρώπινου λάθους ή βλάβες αισθητήρων). Συνεπώς ακολουθήσαμε συγκεκριμένα βήματα προ-επεξεργασίας των δεδομένων. Στην αρχή, έγινε δειγματοληψία των δεδομένων σε χρονικά παράθυρα ωρών και οι μέγιστες, ελάχιστες και μέσες τιμές υπολογίστηκαν για κάθε χρονικό παράθυρο. Ο σκοπός της δειγματοληψίας είναι τριπλός,

- α. παρέχει μια πρόσθετη δομή ή ματιά στο πρόβλημα της εκμάθησης,
- β. επιτρέπει την μείωση της επίδρασης θορύβου στα δεδομένα, και τέλος,
- γ. όλα τα προτεινόμενα μοντέλα και διαδικασίες είναι πιο αποτελεσματικά όσον αφορά στον χρόνο και τους πόρους.

Το χρονικό πλαίσιο των «ωρών» επιλέχθηκε μιας και αντιστοιχεί και στην ταχύτητα δεδομένων για την εξεταζόμενη περίπτωση—δηλαδή οι αισθητήρες του πλοίου δίνουν νέα δεδομένα κάθε τρεις ώρες. Επιπλέον η εξεταζόμενη διάβρωση μπορεί να διαρκέσει μέρες ως ότου επέλθει η καταστροφή του σταυρού κεφαλής, οπότε αναλογικά εκτιμάται ότι η πληροφορία (μεταβολές στις τιμές) εντοπίζεται τουλάχιστον σε χρονικό ορίζοντα ωρών.



Δεδομένων των παραπάνω, όλα τα προτεινόμενα μοντέλα και μέθοδοι είναι πιο αποδοτικά όσον αφορά τον χρόνο και τους πόρους. Η προεπεξεργασία για την προτεινόμενη προσέγγιση περιλαμβάνει τα εξής βήματα:

α. Τα περισσότερα μοντέλα δεδομένων δυσκολεύονται στη διαχείριση ελλειπόντων τιμών (missing values). Στις χρονοσειρές, αν υπάρχουν κενές τιμές, έχουμε δύο επιλογές: είτε να αγνοήσουμε ολόκληρη την εγγραφή που περιλαμβάνει την πληροφορία, είτε να συμπληρώσουμε την ελλείπουσα πληροφορία. Στην περίπτωση μας, η ελλείπουσα πληροφορία συμπληρώθηκε με την τεχνική της "προώθησης" (forward fill), και στη συνέχεια με την τεχνική της "αναδρομής" (backward fill). Αυτή η τεχνική χρησιμοποιείται ευρέως στα χρηματοοικονομικά δεδομένα και γενικά στις χρονοσειρές. Η ιδέα είναι να συμπληρώσουμε τα χαμένα δεδομένα με την τελευταία γνωστή τιμή της ίδιας σειράς.

β. Πολλές φορές, τα δεδομένα επηρεάζονται από ακραίες τιμές. Στα σταθμισμένα σύνολα δεδομένων, τα δεδομένα από ορισμένους αισθητήρες δεν είχαν υποστεί επεξεργασία και διόρθωση από ένα CEP, οπότε υπήρχαν πολλές ακραίες τιμές που δεν αντικατοπτρίζουν την πραγματικότητα. Επομένως, τα μετρητικά διορθώθηκαν με βάση τις μέγιστες και ελάχιστες τιμές που δίνει η βιβλιογραφία για τους εξεταζόμενους κινητήρες (MAN B&W DIESEL A/S 2004), όπως προτείνεται στην πηγή [181].

γ. Σε ορισμένες περιπτώσεις, οι παραπάνω διορθώσεις δεν κατάφεραν να εξαλείψουν την ύπαρξη ορισμένων ακραίων τιμών που αντιστοιχούν σε ένα σφάλμα ενός αισθητήρα. Σε αυτές τις περιπτώσεις, χρησιμοποιήσαμε μια επιπλέον τεχνική κατά την αποθήκευση των δεδομένων. Συγκεκριμένα, περιορίσαμε τις τιμές των αισθητήρων με βάση ένα κινητό όριο για τιμές που υπερβαίνουν τις 3 τυπικές αποκλίσεις (3SDs). Η τεχνική της περικοπής στην επεξεργασία σημάτων περιορίζει ένα σήμα όταν υπερβαίνει ένα καθορισμένο κατώφλι. Αυτό το κατώφλι καθορίζεται με τη χρήση των 3 τυπικών αποκλίσεων, όπως προαναφέρθηκε. Υποθέτοντας μια κανονική κατανομή των δεδομένων ενός αισθητήρα, το 99,7% των δεδομένων θα πρέπει να είναι εντός της περιοχής του μέσου +3SDs. Επομένως, αυτή η τεχνική φιλτράρει τα περισσότερα από τα μη φυσιολογικά σημεία δεδομένων από τα σύνολα των δεδομένων, χωρίς να επηρεάζει την μεγάλη πλειοψηφία των κανονικών σημείων. Επιπλέον, οι κυλινδρικοί μέσοι όροι μπορούν να χρησιμοποιηθούν για να μειώσουν ή να εξαλείψουν την μεταβλητότητα ενός σήματος αισθητήρα, βοηθώντας στην αποφυγή των ψευδών συναγερμών που βασίζονται στο θόρυβο ή άλλες μικρές αλλαγές (γνωστές και ως ψευδείς θετικές αναγνώσεις).

δ. Στα πολυμεταβλητά μοντέλα υπάρχει η πιθανότητα εμφάνισης διαφορετικών στατιστικών χαρακτηριστικών μεταξύ των μεταβλητών επιφέροντας αρνητική επίδραση στην συμπεριφορά κάποιων μοντέλων μηχανικής μάθησης καθώς μπορεί να ευνοούν χαρακτηριστικά με ορισμένα συγκεκριμένα στατιστικά μεγέθη. Αυτό συμβαίνει καθώς κάποιες αντικειμενικές συναρτήσεις είναι πιο



ευαίσθητες στην στάθμιση των χαρακτηριστικών που έχουν ευρύ φάσμα τιμών (ή έχουν μεγαλύτερες κατά απόλυτη τιμή τιμές). Επομένως, το εύρος όλων των χαρακτηριστικών πρέπει να κλιμακωθεί ούτως ώστε κάθε χαρακτηριστικό να συνεισφέρει περίπου αναλογικά στις τελικές απώλειες που δίνονται από την αντικειμενική συνάρτηση. Η κανονικοποίηση είναι ένας πιο ριζικός μετασχηματισμός με στόχο να αλλάξει τις παρατηρήσεις έτσι ώστε να μπορούν να θεωρηθούν ως μία κανονική κατανομή που βοηθάει ορισμένα μοντέλα να συγκλίνουν στις βέλτιστες τιμές κατά της διάρκεια της διαδικασίας εκπαίδευσης. Για να αντιμετωπίσουμε το φαινόμενο διαφοράς κλίμακας των μεταβλητών εισόδου χρησιμοποιήσαμε την ποσοστιαία αλλαγή της δοσμένης χρονοσειράς, για το μοντέλο *One Class SVM*. Η κλιμάκωση και η κανονικοποίηση λειτουργούν στα υπόλοιπα μοντέλα εκτός από αυτά που βασίζονται σε δέντρα αποφάσεων. Τα δέντρα κατάταξης είναι αμετάβλητα από όλους τους μονότονους μετασχηματισμούς επιμέρους διατεταγμένων μεταβλητών. Ο λόγος είναι ότι τα δέντρα κατάταξης χωρίζουν τους κόμβους με δημιουργία κατωφλίου ενώ οι μονότονοι μετασχηματισμοί δεν μπορούν, [182]. Σημειώνεται ότι η κλιμάκωση αλλάζει μόνο το εύρος κάθε χαρακτηριστικού και όχι τα στατιστικά χαρακτηριστικά του. Επιπροσθέτως, εφαρμόστηκε PCA (Ανάλυση Πρωτευόντων Συνιστωσών) ώστε να δημιουργήσει επιπλέον χαρακτηριστικά που χρησιμοποιούνται στα μοντέλα LSTM. Ο σκοπός που προτείνετε η προσέγγισης PCA είναι για να εξαλείψει την επιρροή των υψηλά σχετιζόμενων χαρακτηριστικών σε μια μοναδική ροή δεδομένων.

ε. Ένας άλλος συνηθισμένος κίνδυνος σε πολλές προσεγγίσεις μηχανικής μάθησης είναι ο λανθασμένος τρόπος (ή μέθοδος) αξιολόγησης και επικύρωσης των αποτελεσμάτων. Μια λανθασμένη επικύρωση μπορεί να οδηγήσει σε υπέρ-αισιόδοξες προσδοκίες περί του τι θα συμβεί όταν το μοντέλο εφαρμοστεί στην παραγωγή και σε καμιά περίπτωση δεν είναι αποδέκτη. Ο τρόπος που το σύνολο δεδομένων μας χωρίστηκε σε σύνολα εκμάθησης, επικύρωσης και ελέγχου, βασίστηκε στις απαιτήσεις της κάθε μεθόδου που χρησιμοποιήθηκε. Σε περιπτώσεις όπως τα *LSTMs* και *SVM* μιας κλάσης (*OneClassSVM*), όπου τα μοντέλα εκπαιδεύονται και προσομοιώνουν την κανονική λειτουργία της μηχανής, κατά την εκπαίδευση τους χρησιμοποιήθηκαν τα δεδομένα ενός πλοίου χωρίς ιστορικό ελαττωμάτων (βλαβών), ενώ ως δεδομένα επικύρωσης και αξιολόγησης χρησιμοποιήσαμε δεδομένα από πλοία που εμφάνισαν την εξεταζόμενη βλάβη. Ενώ στα μοντέλα με επίβλεψη όπως το *XGBoost* που βασίζεται σε δέντρα αποφάσεων, χρησιμοποιήσαμε δεδομένα από ελαττωματικό πλοίο, τόσο στην διαδικασία εκμάθησης όσο αυτήν της επικύρωσης.

### 3.2.4 Μοντέλα Μηχανικής Μάθησης – Στατιστικά Μοντέλα

Λαμβάνοντας υπόψη ότι ήταν αδιευκρίνιστο εάν η εντοπίσεις της προς εξέταση βλάβης απαιτεί την ανίχνευση παρατεταμένων ανωμαλιών, απλών σύντομων



καθυστερήσεων ή στιγμιαίων διαταραχών στα δεδομένα, οι προτεινόμενες μέθοδοι ανίχνευσης ανωμαλιών επικεντρώνονται τόσο σε συνεχή τμήματα λειτουργίας (γνωστά επίσης ως χρονικά τμήματα ή χρονικά παράθυρα) όσο και σε μεμονωμένα σημεία λειτουργίας (δηλαδή συγκεκριμένα χρονικά αποτυπώματα). Εφαρμόσαμε και αξιολογήσαμε πληθώρα συνδυασμών μοντέλων και προαναφερθέντων κανόνων, ώστε να προσδιοριστεί η κοινή ταυτότητα μεθόδων ανίχνευσης ανωμαλιών που αποδεικνύεται αποτελεσματική στην εντοπίσει ανωμαλιών στο ρουλεμάν σταυρού κεφαλής. Καθώς οι σημειακές ανωμαλίες μπορούν, εκτός από να είναι ενδείξεις προβλήματος ή δυσλειτουργίας του αισθητήρα, να σηματοδοτούν επίσης βλάβη στο ρουλεμάν σταυρού κεφαλής, υπήρχε ανάγκη για τη χρήση Μοντέλων Μηχανικής Μάθησης ικανών να ανιχνεύουν σημειακές ανωμαλίες. Τόσο επιβλεπόμενα όσο και ημι-επιβλεπόμενα μοντέλα χρησιμοποιήθηκαν. Τα μοντέλα που χρησιμοποιήθηκαν είναι τα εξής:

**Ταξινόμηση με δέντρο αποφάσεων gradient boosting:** Τα δέντρα ταξινόμησης και αναδρομής αποτελούν ορισμένες από τις πιο ισχυρές τεχνικές της μηχανικής μάθησης [183]. Μπορούν επιπλέον να εφαρμοστούν σε προβλήματα ανίχνευσης ανωμαλιών με διάφορους τρόπους. Πρώτον, μπορούν να χρησιμοποιηθούν για επιβλεπόμενη μάθηση, ταξινομώντας ανώμαλα και μη ανώμαλα δεδομένα. Για να επιτευχθεί αυτό, απαιτείται η επισήμανση των δεδομένων, μια διαδικασία που είναι γνωστή ως labeling [184]. Σε αυτή την εργασία, χρησιμοποιήσαμε τη βιβλιοθήκη XGBoost (ακρωνύμιο του eXtreme Gradient Boosting) [183]. Μέσω αυτής της μεθόδου, ο στόχος της ανίχνευσης ανωμαλιών μετατράπηκε σε πρόβλημα ταξινόμησης. Σε αυτό το πλαίσιο, προτείνουμε ένα στατικό χρονικό παράθυρο, υποθέτοντας ότι μέσα σε αυτό εξελίσσεται η βλάβη. Κατ' αυτόν τον τρόπο, ορίσαμε την ετικέτα '1' ή 'Αληθές' για ένα στατικό χρονικό παράθυρο 30 ημερών πριν από κάθε γνωστή ημερομηνία εμφάνισης ελαττώματος. Επιλέξαμε αυτή την προσέγγιση των 30 ημερών γιατί υπάρχει μεγάλη πιθανότητα ένα πλοίο να συνεχίσει το ταξίδι του με ένα ελαφρώς χαλασμένο ρουλεμάν σταυρού κεφαλής για περισσότερες από 30 ημέρες. Αυτή η υπόθεση πρέπει να δοκιμαστεί σε διάφορα χρονικά παράθυρα και η αξιολόγηση πρέπει να βασίζεται στα αποτελέσματα των δοκιμών σε όλα τα παράθυρα. Στην προκειμένη περίπτωση, οι δοκιμές έγιναν για 5, 10, 30 και 60 ημέρες. Κατά συνέπεια, τα δεδομένα που επισημάνθηκαν τεχνητά βάσει του προαναφερόμενου παραθύρου των 30 ημερών ήταν τα δεδομένα του εποπτευόμενου μοντέλου μας. Παρ' όλα αυτά, τα δεδομένα ήταν ανισορροπημένα καθώς η πλειοψηφία των δειγμάτων ήταν 'φυσιολογικά'. Για να ξεπεράσουμε αυτή την πρόκληση, χρησιμοποιήσαμε την τυπική τιμή

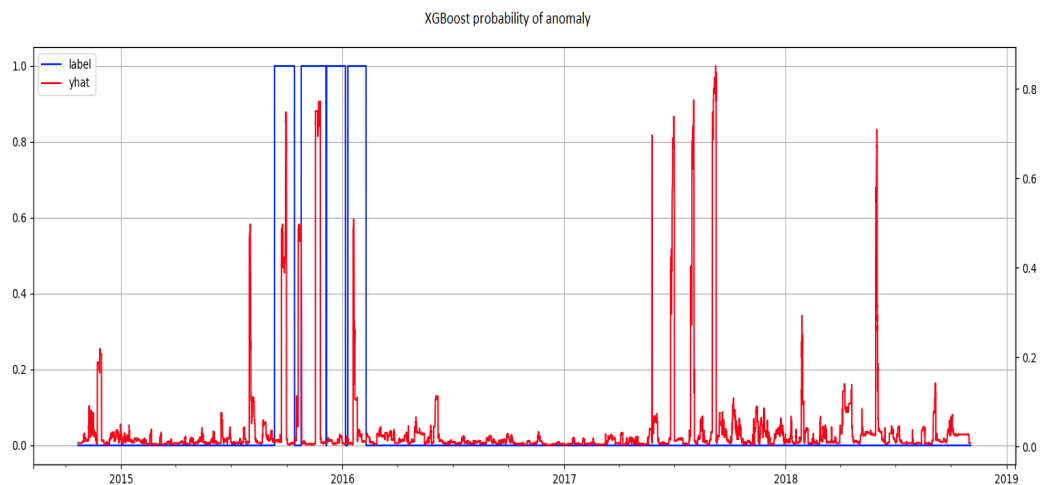
*άθροισμα (θετικών περιπτώσεων)/άθροισμα (αρνητικών περιπτώσεων)*





ως υπέρ-παράμετρο του μοντέλου *XGBoost*. Ουσιαστικά η τιμή αυτή αντιπροσωπεύει το «βάρος» που εφαρμόζεται σε θετικές περιπτώσεις, το οποίο κανονικοποιείται με το άθροισμα (θετικών περιπτώσεων) και το εξισώνεται με το άθροισμα (αρνητικών περιπτώσεων).

Ενδεικτικά τα αποτελέσματα της εν λόγω μεθόδου κατηγοριοποίησης παρουσιάζεται στην Εικόνα 8, όπου με κόκκινες γραμμές απεικονίζεται η πιθανότητα βλάβης ως προς το χρόνο (άξονας x) ενώ με μπλε οι περίοδοι όπου είχε πρόβλημα το εν λόγω πλοίο.



**Εικόνα 8** Οι μπλε κορυφές είναι οι χρονικές περιόδους κατά τις οποίες εξελισσόταν ένα ελάττωμα του Crosshead Bearing. Η κόκκινη γραμμή είναι η προβλεπόμενη πιθανότητα ελαττώματος με βάση το μοντέλο XGboost.

**One Class Support Vector Machines (OCSVM):** Τα *SVM* (*Support Vector Machines*) είναι μοντέλα εκμάθησης με επίβλεψη που αναλύουν δεδομένα και αναγνωρίζουν μοτίβα, και που μπορούν να χρησιμοποιηθούν τόσο για σκοπούς ταξινόμησης όσο και για εργασίες παλινδρόμησης. Ένα μοντέλο SVM βασίζεται στο ότι διαιρεί τα δείγματα εκμάθησης σε ξεχωριστές κατηγορίες με όσο το δυνατό μεγαλύτερο κενό μεταξύ τους «τιμωρώντας» τα δείγματα εκμάθησης που μεταπίπτουν στην λανθασμένη πλευρά του διάκενου.

Ο μηχανισμός SVM μίας κλάσης (*OCSVM*) μπορεί να χρησιμοποιηθεί για να αναγνωρίσει ανωμαλίες στο χώρο χαρακτηριστικών καθώς δημιουργεί ένα υπέρ-επίπεδο που βέλτιστα χωρίζει τα δεδομένα. Συγκρινόμενα με άλλα εργαλεία μηχανικής μάθησης, τα *SVM* προσφέρουν ανώτερες ικανότητες γενίκευσης [185]. Ο Μηχανισμός Υποστήριξης Διανυσμάτων Μίας Κλάσης, είναι ένας αλγόριθμος εκμάθησης χωρίς επιτήρηση, που εκπαιδεύεται μόνο σε “φυσιολογικά” δεδομένα, στην περίπτωση μας τα δεδομένα που αντιπροσωπεύουν την κανονική λειτουργία



(χωρίς βλάβες). Μαθαίνει τα όρια αυτών των σημείων και είναι έτσι σε θέση να ταξινομήσει όποια σημεία μεταπίπτουν εκτός των ορίων ως ακραία σημεία. Όπως επίσης προτείνεται και εφαρμόζεται στο [186], όσο μακρύτερα είναι ένα σημείο από το υπερ-επίπεδο προς την κανονική πλευρά, η κανονικότητα του συστήματος που εξετάζεται αυξάνεται και αντίστροφα. Στην περίπτωση μας, εκπαιδεύσαμε δυο μοντέλα SVM σε δεδομένα από πλοία που ποτέ δεν είχαν δυσλειτουργία του ρουλεμάν σταυρού κεφαλής. Το πρώτο εκπαιδεύτηκε σε κανονικοποιημένες ροές δεδομένων, και το δεύτερο σε ωριαία ποσοστιαία αλλαγή των ροών δεδομένων. Τελικά χρησιμοποιήσαμε το δεύτερο μοντέλο, καθώς αξιολογήθηκε ως πιο αποτελεσματικό.

Ακόμα, τα μοντέλα επικεντρώθηκαν σε δυο συνεχείς ενότητες δραστηριοτήτων που είναι οι ακόλουθες:

**Multivariate Long Short-Term Memory Model (LSTM):** Μια από τις υποσχόμενες προσεγγίσεις εντοπισμού ανωμαλιών για δεδομένα χρονοσειρών είναι τα Αναδρομικά Νευρωνικά Δίκτυα (Recurrent Neural Networks - RNNs), ειδικά αυτά της Βραχέως Χρόνου Προσωρινής Μνήμης (Long Short-Term Memory - LSTM) που φαίνεται ότι είναι ο ιδανικός υποψήφιος εντοπισμού ανωμαλιών που περιλαμβάνουν δεδομένα χρονοσειρών, καθώς το LSTM μπορεί να χειριστεί δεδομένα ροής πολυμεταβλητών χρονοσειρών, μοντελοποιώντας πολύπλοκες μη γραμμικές αλληλεπιδράσεις χαρακτηριστικών που υπάρχουν συχνά σε ροές δεδομένων πολυμεταβλητών χρονοσειρών. Αν συμβαίνει αυτό, ένα μοντέλο LSTM εκπαιδεύεται με δεδομένα θεωρούμενης φυσιολογικής λειτουργίας, έτσι ώστε να μοντελοποιήσει την φυσιολογική λειτουργία ενός συστήματος, προσομοιώνοντας τη συμπεριφορά του συστήματος υπό φυσιολογικές συνθήκες. Εν συνεχεία, το εκπαιδευμένο μοντέλο προβλέπει τις επόμενες τιμές και το σφάλμα μεταξύ των προβλεπόμενων και των πραγματικών τιμών, επιδεικνύοντας μία πιθανή ανωμαλία. Η υπόθεση που εξετάζουμε στην παρούσα μεθοδολογία είναι ότι αν η πραγματικές τιμές αποκλίνουν από αυτές της «κανονικής λειτουργίας» (προβλέψεις του LSTM) τότε έχουμε ένδειξη ανωμαλίας. Ισοδύναμες προσεγγίσεις μπορούν να βρεθούν στις προσεγγίσεις της NASA που βασίζονται στα LSTMs και μια καινοτομία μη-παραμετρικής δυναμικής τοποθέτηση κατωφλίου στην περίπτωση ανωμαλιών στην περίπτωση δεδομένων τηλεμετρίας διαστημικών σκαφών [187] και [188] παρουσιάζεται μια ισοδύναμη προσέγγιση των *LSTMs* σε υψηλά περιοδικά και θορυβώδη δεδομένα.

Επιπλέον, πολύ περισσότερες προσπάθειες μπορούν να βρεθούν στην σχετική βιβλιογραφία [189][190][191]. Στο πλαίσιο της έρευνας μας, ένα μοντέλο που βασίζεται στο *LSTM* εκπαιδεύτηκε έχοντας για είσοδο πολλαπλές ροές δεδομένων (για ένα δοσμένο χρονικό παράθυρο) και ως έξοδο την τιμή του επόμενου χρονικού βήματος (πρόβλεψη ενός βήματος μπροστά). Δεδομένου ότι δεν υπάρχει μαθηματικό μοντέλο που βασίζεται σε πραγματικά δεδομένα για τα μέρη της μηχανικής που περιγράφουν τον κύκλο της ζωής και την διάρκεια του ρουλεμάν σταυρού κεφαλής. Έτσι, τα πειράματα διεξάχθηκαν σε μια ποικιλία



χρονικών παραθύρων και σε διαφορετικά χρονικά πλαίσια ώστε να υποδείξουν το καλύτερο μοτίβο δεδομένα που οδηγούν σε πιθανή ανίχνευση ελαττώματος. Τα δεδομένα για τα πλοία στα οποία δεν εμφανίστηκε κανένα ελάττωμα στο ρουλεμάν σταυρού κεφαλής χρησιμοποιήθηκαν ως δεδομένα κανονικής λειτουργίας ώστε να εκπαιδεύσουν το μοντέλο. Στο πλαίσιο αυτής της διαδικασίας δημιουργήθηκε μια κατάσταση λειτουργίας LSTM ενός πλοίου υπό κανονική λειτουργία. Κατόπιν, χρησιμοποιώντας δεδομένα που περιείχαν περιπτώσεις των υπαρχόντων ρουλεμάν σταυρού κεφαλής ως δεδομένα ελέγχου, πιστοποιήσαμε τις εξόδους πρόβλεψης των παραπάνω αναφερόμενων εκπαιδευόμενων μοντέλων, σχετικά με τα πραγματικά δοσμένα δεδομένα εξόδου από κάθε αισθητήρα. Αφού μια τιμή πρόβλεψης παραχθεί σε κάθε βήμα, υπολογίζοντας το σφάλμα πρόβλεψης ως το μέσο απόλυτο σφάλμα του πίνακα πρόβλεψης. Το σύνολο των σφαλμάτων κατόπιν ομαλοποιείται για να μειώσει τις αιχμές των σφαλμάτων που μπορεί να προκύψουν στις προβλέψεις που βασίζονται στο LSTM, καθώς οι απότομες αλλαγές σε τιμές που συχνά δεν προβλέπονται με ακρίβεια και οδηγούν σε απότομες αιχμές στις τιμές σφαλμάτων ακόμα και στη φυσιολογική συμπεριφορά [188]. Για να εκτιμήσουμε αν οι τιμές είναι φυσιολογικές, ορίζουμε ένα κατώφλι, εμπνευσμένο από την άνω Bollinger Band (δείκτης τεχνικής ανάλυσης χρηματοοικονομικών) για τα ομαλοποιημένα σφάλματα πρόβλεψης-τιμές που αντιστοιχούν σε ομαλοποιημένα σφάλματα πάνω από το κατώφλι και που ταξινομούνται ως ανωμαλίες.

**Σταθμισμένη Εντροπία Μετάθεσης [Weighted Permutation Entropy (WPE)]:** Η εντροπία μετάθεσης είναι ένα μέτρο της πολυπλοκότητας σειριακών διαδοχικών δεδομένων. Με άλλα λόγια η μετρική WPE αναφέρει τον όγκο της νέας πληροφορίας που εμφανίζεται κατά μέσο όρο σε κάθε σημείο της ακολουθίας. Στο πλαίσιο των χρονοσειρών, αυτό μεταφράζεται σε ένα μέτρο του τρόπου με τον οποίο η πληροφορία διαδίδεται στο χρόνο. Αυτό που κρύβεται πίσω από αυτήν την προσέγγιση είναι ότι η χρήση του ως σύστημα ανίχνευσης ανωμαλίας όπου οι απότομες αλλαγές της εντροπίας μεταθέσεων υποδεικνύουν πως κάτι αναπάντεχο συμβαίνει. Αυτή η μέθοδος εφαρμόστηκε για τον εντοπισμό ανωμαλιών σε εγγραφές παλαιό-κλιματικών δεδομένων [192].

Στην περίπτωση μας, χρησιμοποιήσαμε σταθμισμένη εντροπία μετάθεσης ως μέτρο της χρονικής πολυπλοκότητας μίας χρονοσειράς και συμπεριφερόμαστε στις απότομες αλλαγές ως σήμα πιθανών ανωμαλιών να συμβαίνουν φυσικά ή που σχετίζονται με δεδομένα. Αυτός ο σκοπός δεν χρειάζεται ένα μέτρο εκπαίδευσης, μιας και αυτό είναι ένα μέτρο που μπορεί να υπολογιστεί σε πραγματικό χρόνο (online).

Όσον αφορά στην συμπληρωματική μας έρευνα στη συγκεκριμένη περίπτωση έγινε προσπάθεια να υπολογιστεί η την σχετιζόμενη πιθανότητα σε κάθε κύλινδρο του υπό εξέταση ελαττώματος του ρουλεμάν σταυρού κεφαλής κυλίνδρου, οι



επιπρόσθετες τεχνικές χρησιμοποιούν μονομεταβλητά LSTMs, τη μέθοδο Matrix Profile και ανάλυση συν-ολοκλήρωσης.

**Univariate Long Short-Term Memory (u-LSTM):** Η υπόθεση και η προοπτική πίσω από την χρήση των μονομεταβλητών LSTMs είναι ίδια με αυτή των πολύμεταβλητών, με την διαφορά ότι ένα μοναδικό μοντέλο δημιουργείται για κάθε ροή δεδομένων και κάθε μοντέλο χρησιμοποιείται ώστε να προβλέψει τιμές για τη συγκεκριμένη ροή δεδομένων. Όπως και στο πολυμεταβλητό LSTM, τα δημιουργούμενα μοντέλα, εκπαιδεύτηκαν να προσομοιώνουν τη φυσιολογική λειτουργία κάθε κυλίνδρου. Μοντελοποιώντας κάθε κανάλι χωριστά επιτρέπει την ιχνηλασιμότητα στο επίπεδο του καναλιού, δίνοντας μία ένδειξη του ποιος κύλινδρος μπορεί να είναι κατεστραμμένος. Για κάθε σημείο, μια ακολουθία μήκους  $s$  καθορίζει τον αριθμό των σημείων που θα εισαχθούν στο μοντέλο για πρόβλεψη. Η πρόβλεψη λαμβάνει χώρα την επόμενη χρονική στιγμή, κάτι που σημαίνει ότι ο αριθμός βημάτων που πρέπει να προβλεφθεί είναι μονάδα (1). Σαν αποτέλεσμα, δημιουργείται μία μόνο βαθμωτή πρόβλεψη παράγεται για την πραγματική τιμή του αισθητήρα σε κάθε βήμα. Έπειτα υπολογίσαμε το σφάλμα μεταξύ των πραγματικών και των προβλέψεων. Ο υπολογισμός, εκτελείται πάλι με βασισμένη την τιμή του σφάλματος.

**Matrix Profile (MP):** Το MP εισήχθη το 2016 [193], και είναι ένας κλιμακωτός αλγόριθμος για την διερεύνηση ομοιότητας υπο-ακολουθίας χρονοσειρών όλων των ζευγών. Για εξαιρετικά μεγάλα σύνολα δεδομένων, ο αλγόριθμος μπορεί τετριμμένα να λειτουργήσει ως ένας αλγόριθμος για κάθε χρονική στιγμή και να παράγει προσεγγιστικές λύσεις υψηλής ποιότητας σε λογικό χρόνο. Στην περίπτωση μας, για τον εντοπισμό μη ομαλών λειτουργιών μέσα σε μία χρονοσειρά μπορούμε να εφαρμόσουμε έλεγχο ομοιότητας (μικρότερη ευκλείδεια απόσταση) μεταξύ υπο-χρονοσειρών (τμημάτων χρονοσειράς σταθερού μεγέθους). Η υπόθεση εδώ υποδεικνύει ότι πριν από κάθε βλάβη, το ίδιο μοτίβο πρέπει να παρουσιάζεται σε μια ροή δεδομένων του αισθητήρα ενός κυλίνδρου. Οπότε, μία χρονική περίοδος 30 ημερών χρησιμοποιήθηκε (ως μια έγκυρη περίοδος όπως συζητήθηκε νωρίτερα στην περίπτωση ταξινόμησης αύξησης κλίσης) για κάθε πρώτο ελαττωματικό κύλινδρο και το MP υπολογίστηκε για τον άλλο κύλινδρο χρησιμοποιώντας την καθορισμένη υπο-ακολουθία, αναμένοντας η χαμηλότερη τιμή του MP να βρίσκεται πριν το ελάττωμα του υπό βλάβη κυλίνδρου. Ουσιαστικά, τμήματα της χρονοσειράς συγκρίνονται σε ζεύγη, βάση της Ευκλείδειας απόστασης τους από κάθε άλλο τμήμα. Εκμεταλλευόμενοι τους αλγόριθμους του MP είναι δυνατών να μειώσουμε δραστικά τον χρόνο υπολογισμού.

**Ανάλυση συνολοκλήρωσης (Cointegration analysis):** Η υπόθεση είναι ότι κάποια ζεύγη δεδομένων της δοσμένης χρονοσειράς συνολοκληρώνονται. Η συνολοκλήρωση είναι μια στατιστική ιδιότητα μίας συλλογής ( $X_1, X_2, X_3, \dots, X_K$ ) χρονοσειρών. Πρώτα, όλες οι σειρές πρέπει να ολοκληρωθούν σε τάξη  $d$ . Μετά, αν ένας γραμμικός συνδυασμός της συλλογής αυτής ολοκληρωθεί σε τάξη μικρότερη

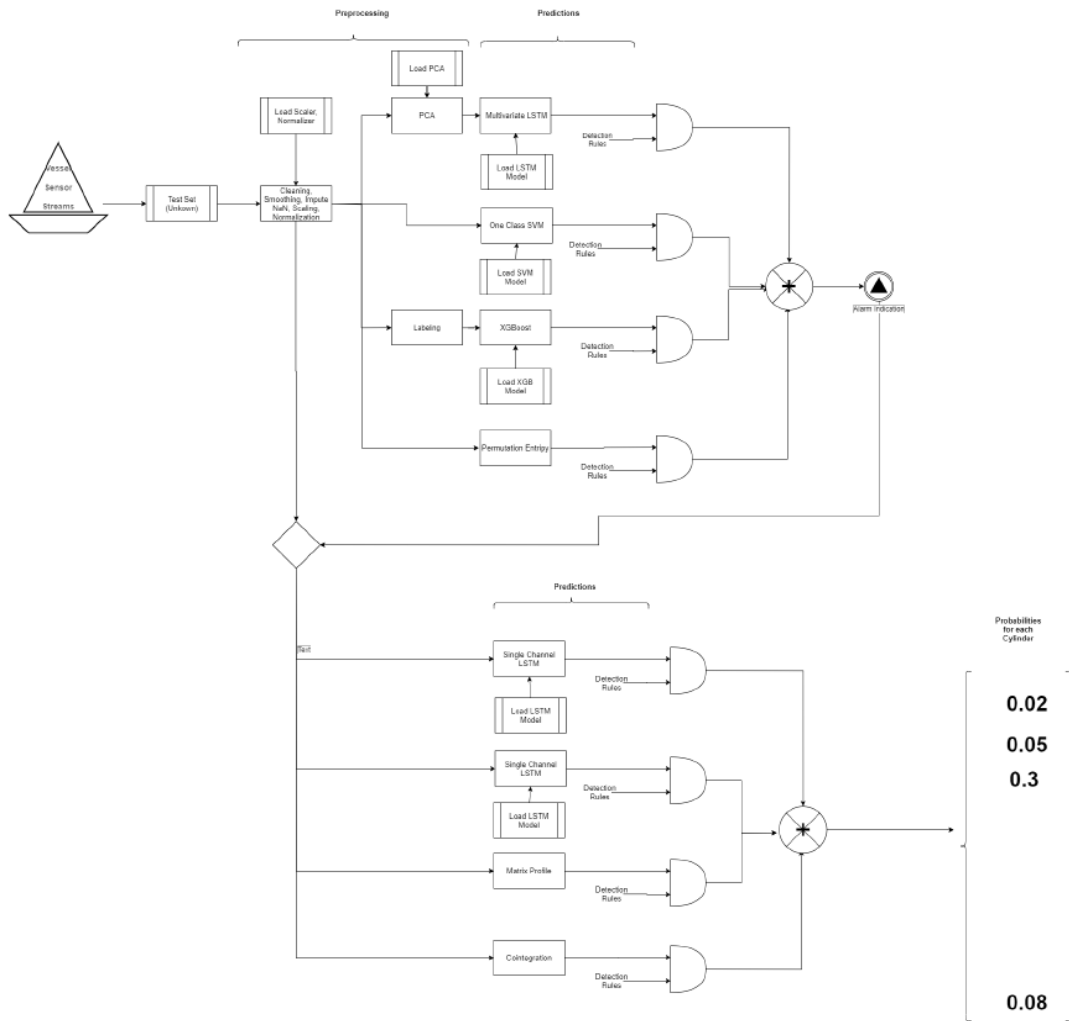


από  $d$ , λέγετε ότι η συλλογή συνολοκληρώνεται κατά τάξη  $d$ . Τυπικά, αν τα  $(X, Y, Z)$  είναι ολοκληρωμένα σε τάξη  $d$ , και υπάρχει συντελεστής  $a, b, c$  έτσι ώστε  $aX + bY + cZ$  ολοκληρώνονται σε τάξη μικρότερη του  $d$ , τότε τα  $(X, Y, Z)$  συνολοκληρώνονται.

Η συνολοκλήρωση έχει γίνει μία σημαντική ιδιότητα στην σύγχρονη ανάλυση χρονοσειρών [194]. Η συνολοκλήρωση βοηθάει να αναγνωριστεί ο βαθμός του πόσο δυο μεταβλητές είναι ευαίσθητες στην ίδια μέση τιμή για μία συγκεκριμένη χρονική περίοδο. Έτσι, η συνολοκλήρωση δεν αντικατοπτρίζει το αν δύο χρονοσειρές μπορεί να κινηθούν στην ίδια ή σε αντίθετη κατεύθυνση, αλλά μπορεί να πει αν η απόσταση μεταξύ τους παραμένει η ίδια με τον χρόνο [194]. Στα πλαίσια της περίπτωσης μας, έχοντας ελέγξει και σχεδιάσει την στατικότητα της χρονοσειράς, υποθέτουμε με βεβαιότητα ότι η πλειονότητα των στοιχείων συνολοκληρώθηκε. Βασισμένοι στην θεωρία που υποθέσαμε, αν παρατηρήσουμε μια προσωρινή έκρηξη συνολοκλήρωσης δυο χρονοσειρών, αυτό σημαίνει ότι υπάρχει μια ανωμαλία ή μια δυσλειτουργία. Έτσι, αποφασίσαμε να μελετήσουμε ειδικά την συνολοκλήρωση μεταξύ των δεδομένων των αισθητήρων ενός ορισμένου κυλίνδρου/πιστονιού μεταξύ τους και μεταξύ κάποιων κυρίων μεταβλητών θερμοκρασίας και πίεσης. Το ζεύγος των ροών που χρησιμοποιήθηκαν δίνονται στον Πίνακα 2. Θα πρέπει να αναφερθεί ότι πολλές μέθοδοι επαναδειγματοληψίας δοκιμάστηκαν και επίσης χρησιμοποιήσαμε προσπάθειες ώστε να απομακρυνθεί η παροδικότητα των δεδομένων χωρίς ουσιώδες επίπτωση στο αποτέλεσμα.

**Πίνακας 2** Χαρακτηριστικά που εκλέχθηκαν ως προς τη συνολοκλήρωση τους

Cylinder #1 - 10 Scavenge Air Fire Detection Temperature (°C)
Cylinder #1 - 10 Exhaust Gas Out Temperature (°C)
Cylinder #1 - 10 Jacket Cooling Fresh Water Outlet Temperature (°C)
Cylinder #1 - 10 Piston Cooling Outlet Temperature (°C)
Second Feature description
Mean and Max of Cylinders Scavenge Air Fire Detection Temperature
Mean and Max of Cylinders Exhaust Gas Out Temperature (°C)
Mean and Max of Cylinders Jacket Cooling Fresh Water Outlet Temperature
Mean and Max of Cylinders Piston Cooling Outlet Temperature (°C)
Fuel Oil Inlet Temperature (°C)
Jacket Cooling Fresh Water Inlet Temperature Low (°C)
Starting Air Pressure (Pa)
Thrust Pad Temperature (°C)
Main Lube Oil Inlet Temperature (°C)
Fuel Oil Temperature (°C)
GPS speed (kn)
Speed through water
Scavenge Air Inlet Pressure (Pa)
Scavenge Air Receiver Temperature (°C)



Εικόνα 9 Το διάγραμμα ροής παρεμβολής ενός συναγερμού σύμφωνα με την προτεινόμενη προσέγγιση

### 3.3 Μηχανική Μάθηση & ασφάλεια τροφίμων.

Στην παρούσα πειραματική διαδικασία, στόχος μας είναι να χρησιμοποιήσουμε προσεγγίσεις μηχανικής μάθησης για να επιτρέψουμε τον έγκαιρο εντοπισμό και την πρόβλεψη των ανακλήσεων τροφίμων, οι οποίες μπορούν να αξιοποιηθούν από εταιρείες σε ολόκληρη την αλυσίδα εφοδιασμού τροφίμων. Η αρχική υπόθεση μας είναι ότι η εφαρμογή καινοτόμων τεχνικών θα έχει βελτιωμένα αποτελέσματα σε περιπτώσεις που παραδοσιακές μέθοδοι αποτυγχάνουν.

#### 3.3.1 Σύνολο δεδομένων

Το συγκριμένο πρόβλημα, παρόλο που άπτεται του συνόλου των ανθρωπινων κοινωνιών και εμφανίζει ερευνητικό ενδιαφέρον δεν υπάρχουν διαθέσιμα ανοιχτά

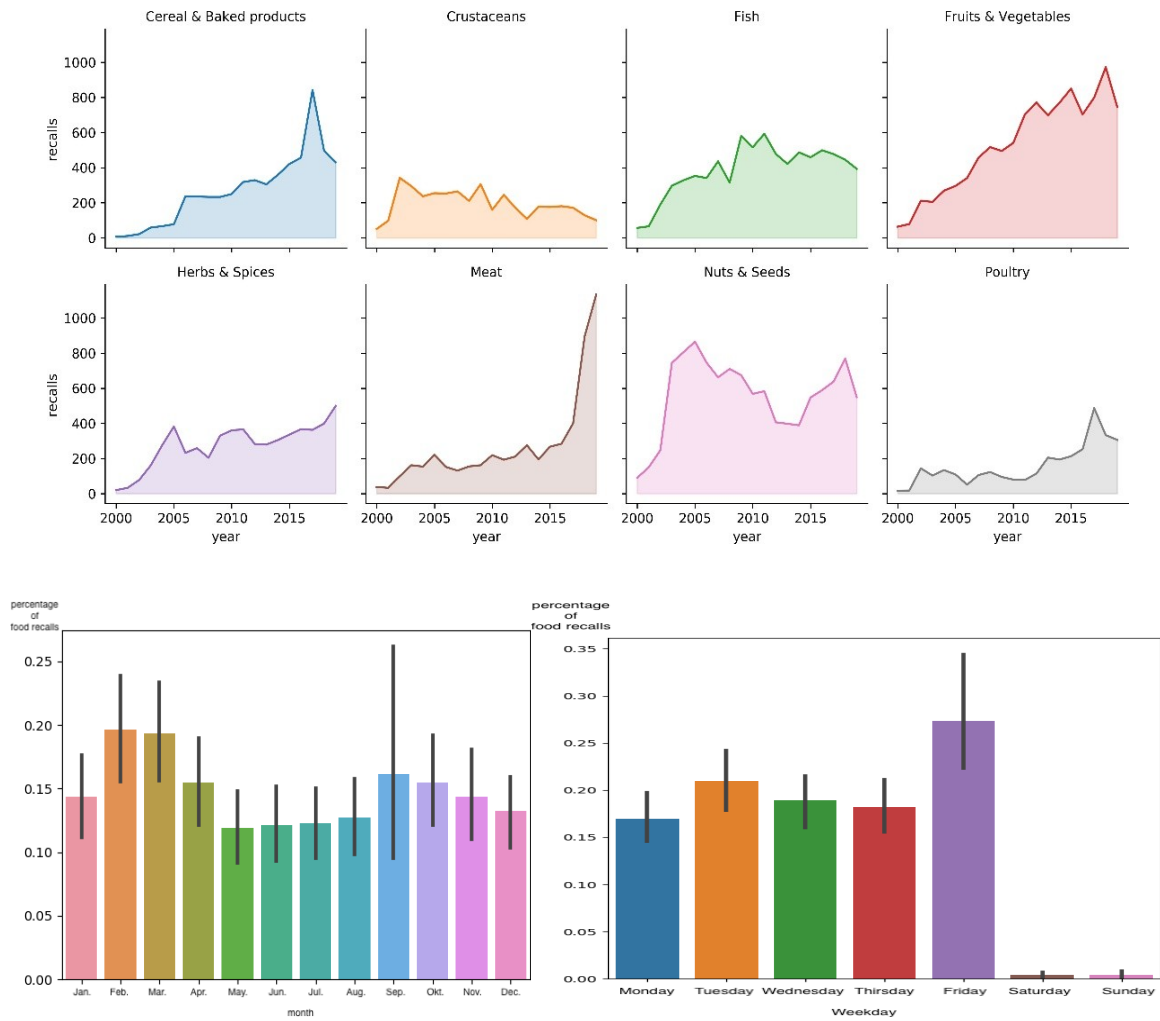


σύνολα δεδομένων ή μοντέλα ανοιχτού κώδικα σχετικά με τις ανακλήσεις τροφίμων. Το χρησιμοποιούμενο σύνολο δεδομένων παρέχεται από την Agroknow (στο πλαίσιο ενός ερευνητικού προγράμματος HORIZON 2020) ως μια πραγματική περίπτωση χρήσης, που αποτελείται από δεδομένα σύντομου κειμένου που προέρχονται από εξόρυξη κειμένου διαφορετικών διαδικτυακών πηγών, όπως ανακοινώσεις, άρθρα και αναφορές που εξετάζουν τις ανακλήσεις τροφίμων παγκοσμίως. Θα πρέπει να αναφερθεί ότι η διαδικασία ανάκτησης αυτών των δεδομένων είναι εκτός του πεδίου εφαρμογής της παρούσας εργασίας. Το σύνολο δεδομένων στην τελική του έκδοση μετασχηματίστηκε σε μορφή πίνακα, αποτελούμενο από τις ακόλουθες πληροφορίες: ID, τίτλο (σύντομο κείμενο) και το προϊόν (ετικέτα). Μερικά ενδεικτικά παραδείγματα απεικονίζονται στον Πίνακα 3.

**Πίνακας 3** Παραδείγματα των δεδομένων ανακοινώσεων ανάκλησης τροφίμων

<i>ID</i>	<i>Title</i>	<i>Product</i>
0	Aflatoxins (tot. = 54.1/mg/kg-ppb) in dried figs from turkey	Dried figs
1	Salmonella (presence /125g) in onion powder from india	Onion powder
2	Limena, llc recalls salvadorean string cheese (quesillo cheese) because of possible health risk	Semi-soft cheese
3	Fosthiazate (0.125+/-0.063 mg/kg - ppm) in fresh peppers from turkey	Fresh peppers
4	Salmonella (1 out of 5 samples /25g) in hot smoked bacon from poland	Smoked bacon

Με βάση τα δεδομένα και τα μετα-δεδομένα που περιλαμβάνουν μεταξύ άλλων και τη χρονική σήμανση κάθε ανάκλησης, δημιουργήθηκε ένα άλλο σύνολο δεδομένων 30 χρονοσειρών, που αντιπροσωπεύει τον ημερήσιο αριθμό ανακλήσεων τροφίμων 30 εμπλεκόμενων κατηγοριών (όπως 'Πρόσθετα και Αρωματικές ουσίες Τροφίμων', 'Κρέας και προϊόντα', 'Ψάρια και προϊόντα' κ.λπ.) από το έτος 2000 και έπειτα. Επιπλέον, η Εικόνα 10(α) απεικονίζει ορισμένες ενδεικτικές κατηγορίες ενδιαφέροντος, ενώ η Εικόνα 10(β) παρουσιάζει την εποχικότητα του εξεταζόμενου χρονοσειράς, τονίζοντας ότι η πλειονότητα των ανακλήσεων τροφίμων δημοσιεύονται την Παρασκευή (λαμβανομένων υπόψη των ημερών της εβδομάδας ως χρονικού πλαισίου). Επιπρόσθετα, παρατηρείται ότι κατά τους καλοκαιρινούς μήνες γίνονται λιγότερες ανακλήσεις τροφίμων συγκριτικά με τους υπόλοιπους μήνες του χρόνου.



**Εικόνα 10 α)** Αναπαράσταση στο πεδίο του χρόνου της ανάκλησης συγκεκριμένων κατηγοριών φαγητού, β) Μέσες ανακλήσεις προϊόντων συγκεντρωτικά ανα μήνα, και ανα ημέρα. Στον άξονα x εμφανίζεται η μονάδα του χρόνου (μήνες ή ημέρες) και στον y άξονα μετράται τα ποσοστά.

**Πίνακας 4** Κατηγορίες χρονοσειρών συνόλου δεδομένων

0	Food Additives and Flavorings	15	Other food products
1	Fruits and Vegetables	16	Food contact materials
2	Poultry meat and products	17	Non-alcoholic beverages
3	Prepared dishes and snacks	18	Fats and Oils
4	Dietetic foods, Food supplements, fortified foods	19	Sugars and Syrups
5	Feed materials	20	Nuts, Seeds and products
6	Honey and Royal Jelly	21	Cereals and Bakery products
7	Bivalve mollusks and products	22	Crustaceans and products
8	Meat and products	23	Eggs and products

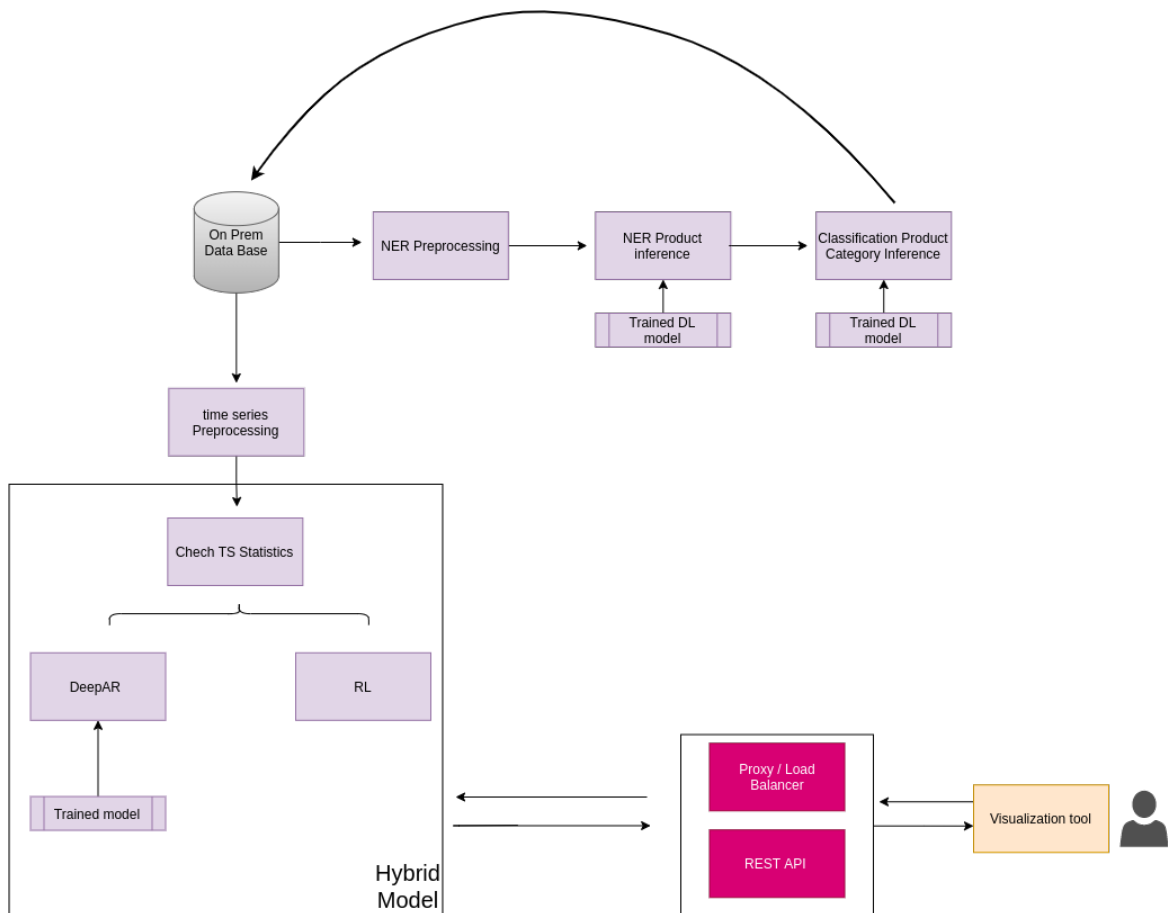




9	Feed Additives	24	Alcoholic beverages
10	Milk and Milk products	25	Cocoa, Coffee and Tea
11	Herbs and Spices	26	Gastropods
12	Confectionery	27	Cephalopods and products
13	Soups, Broths, Sauces and Condiments	28	Pet Feed
14	Fish and products	29	Ices and Desserts

### 3.3.2 Αρχιτεκτονική

Η εννοιολογική αρχιτεκτονική της προτεινόμενης υπηρεσίας απεικονίζεται στην Εικόνα 11. Αρχικά πραγματοποιείται η εξαγωγή συγκεκριμένων λέξεων-κλειδίων από σύντομα κείμενα. Η εξαγωγή αναφέρεται στην κατηγορία «ΠΡΟΪΟΝ» που έχει ανακληθεί από τον τίτλο ή την περιγραφή των ανακλήσεων.



**Εικόνα 11** Αρχιτεκτονική συστήματος εξαγωγής και πρόβλεψης χρονοσειρών για την ανάκληση τροφίμων

Δεδομένης της φύσης του συνόλου δεδομένων και του επιθυμητού αποτελέσματος, η προτεινόμενη προσέγγιση για ταξινόμηση πραγματοποιήθηκε με

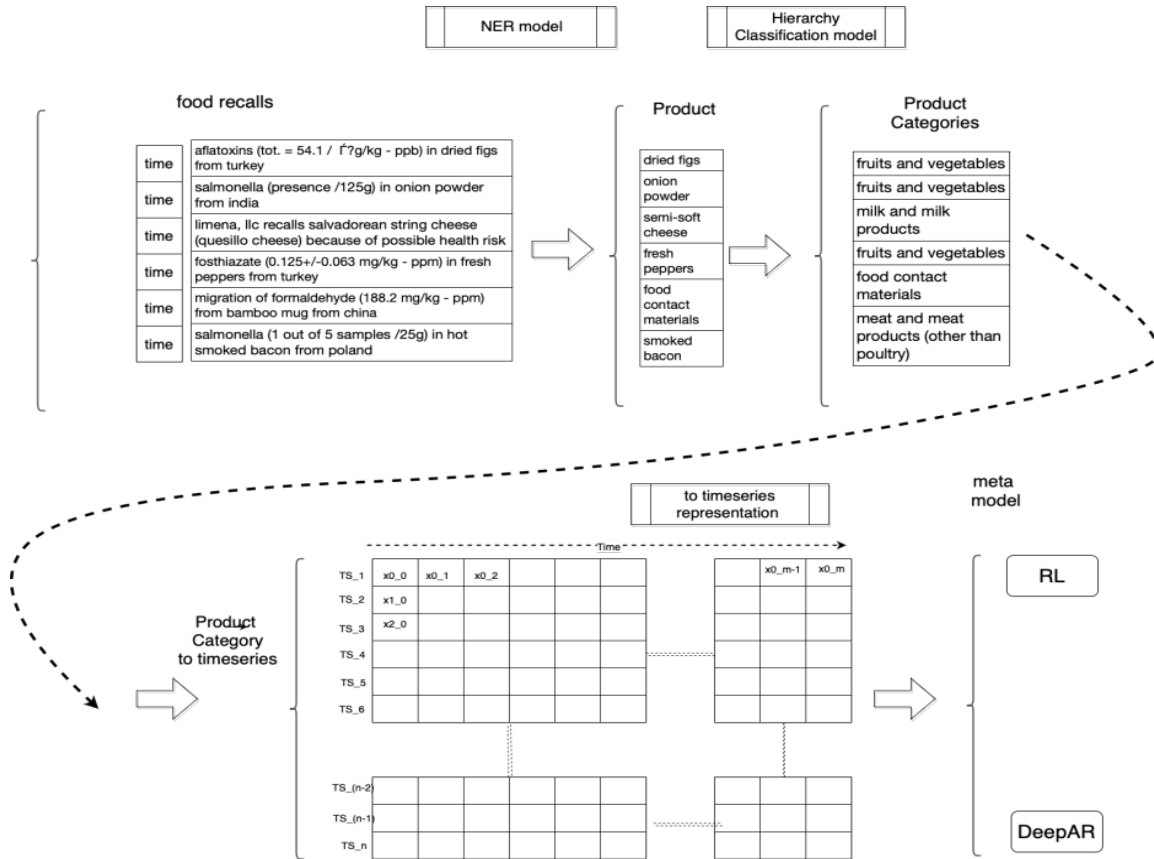


τη χρήση τεχνικών Αναγνώρισης Ονομάτων Οντότητων (Named Entity Recognition-NER). Σε αυτό το πλαίσιο, η πρόκληση του προσδιορισμού του αναφερόμενου προϊόντος που έχει ανακληθεί μπορεί να θεωρηθεί ως ένα σωστά ταξινομημένο στοιχείο. Η μεθοδολογία που εισάγεται στη διατριβή επιτρέπει την ανίχνευση του προϊόντος σε κάθε ανάκληση μέσω:

α. Μιας κατάλληλης τεχνικής προ-επεξεργασίας για το σύνολο δεδομένων.

β. Ένα μοντέλο μηχανικής μάθησης που μπορεί να εκτιμήσει ποιο προϊόν (ή η κατηγορία προϊόντος) αναφέρεται σε μια ανακοίνωση ανάκλησης.

Μια πιο λεπτομερής άποψη αυτού παρέχεται στην Εικόνα 12, η οποία απεικονίζει τη γραμμή ροής δεδομένων, σε μια σταδιακή προσέγγιση, για την εφαρμογή τόσο της διαδικασίας NER όσο και της επανεκπαίδευσης του μοντέλου ενισχυτικής μάθησης (Reinforcement Learning - RL). Στο πρώτο βήμα, τα ακατέργαστα δεδομένα υποβλήθηκαν σε κατάλληλη προ-επεξεργασία και τροφοδοτήθηκαν στο μοντέλο NER. Στη συνέχεια, τα εξαγόμενα ονόματα προϊόντων ταξινομήθηκαν σε συγκεκριμένες κατηγορίες (μια εργασία που δεν αποτελεί μέρος αυτής της έρευνας) χρησιμοποιώντας το Μοντέλο Ταξινόμησης Ιεραρχίας, χρησιμοποιώντας μια ιδιόκτητη οντολογία που είναι υπεύθυνη να συλλαμβάνει μια οντότητα που φέρει την ετικέτα «ΠΡΟΪΟΝ» και την ταξινόμησή της στην αντίστοιχη κατηγορία. Τέλος τα δεδομένα μετατράπηκαν σε αναπαραστάσεις χρονοσειρών για να χρησιμοποιηθούν για την πρόβλεψη μελλοντικών ανακλήσεων. Σε σύγκριση με άλλες σχετικές μεθοδολογίες, ένα επιπλέον νέο χαρακτηριστικό αυτής της προσέγγισης είναι η εξαγωγή του εμπλεκόμενου προϊόντος των κειμένων ανάκλησης τροφοίμων μέσω του συνδυασμού ενός συνόλου μοντέλων ανάλυσης για την αύξηση της αποτελεσματικότητάς τους.



**Εικόνα 12** Σχηματική επεξήγηση των δεδομένων με βάση την προσέγγιση απο άκρο σε άκρο, συμπεριλαμβανομένων των τεσσάρων βασικών βημάτων της διαδικασίας

Όσον αφορά το δεύτερο σκέλος της προσέγγισης, δηλαδή την πρόβλεψη ανάκλησης τροφίμων, εφαρμόστηκε μια καινοτόμα μέθοδος που βασίζεται στην ενισχυτική εκπαίδευση (Reinforcement Learning - RL) αποδίδοντας πολλά υποσχόμενα αποτελέσματα σε σύγκριση με άλλες καθιερωμένες προσεγγίσεις. Τα ταξινομημένα δεδομένα από την εργασία NER αναπαρίστανται στο πεδίο του χρόνου προκειμένου να ενσωματωθούν στο περιβάλλον του RL το οποίο κατασκευάσαμε για πρόβλεψη των ανακλήσεων τροφίμων.

Για το σκοπό αυτό, ορίσαμε ένα καινοτόμο περιβάλλον για να εκφράσουμε το «πλαίσιο» με το οποίο μπορεί να αλληλοεπιδράσει ο πράκτορας (μοντέλο μηχανικής μάθησης), σε ένα προσαρμοσμένο σύνολο ενεργειών (A), το οποίο είναι συγκεκριμένο για τα δεδομένα χρονοσειρών και μια ανταμοιβή  $Ra(s, s')$  που είναι το αποτέλεσμα της αλληλοεπίδρασης με τον περιβάλλον. Το περιβάλλον αποτελείται από ένα σύνολο καταστάσεων (S) όπου κάθε κατάσταση έχει οριστεί να είναι ένας πίνακας των τελευταίων σημείων δεδομένων της χρονοσειράς. Ένα νευρωνικό δίκτυο χρησιμοποιήθηκε ως μηχανισμός για την επιλογή της κατάλληλης δράσης για τον παράγοντα. Ο πράκτορας που προτείνουμε, εκπαιδεύεται να προβλέπει την επόμενη τιμή στη χρονοσειρά ενός προϊόντος ως προς την ποσοστιαία μεταβολή. Για να διευκολυνθεί αυτό, οι διαθέσιμες ενέργειες είναι ένα διακριτό σύνολο αριθμών (20), που προέρχονται από στατιστικά χαρακτηριστικά



χρονοσειρών (κατανομή, μέσος όρος, διάμεσος, ελάχιστο και μέγιστο) που μπορεί να προβλέψει ο πράκτορας. Ένας ομοιόμορφος μηχανισμός επανάληψης εμπειρίας χρησιμοποιείται επίσης για να μπορεί ο πράκτορας να θυμάται τις προηγούμενες εμπειρίες του και τις αντίστοιχες ανταμοιβές τους λόγω της φύσης των προβλημάτων πρόβλεψης χρονοσειρών.

### 3.3.3 Προ-επεξεργασία

Τα δεδομένα σε πραγματικές περιπτώσεις περιλαμβάνουν συχνά θόρυβο, είναι υψηλών διαστάσεων, τα περισσότερα από αυτά μπορεί να είναι διαστρεβλωμένα ή να λείπουν, επομένως πρέπει να εκτελείται κατάλληλη προ-επεξεργασία πριν από οποιαδήποτε εργασία ανάλυσης. Όσον αφορά την ταξινόμηση βάση του υποκείμενου προϊόντος, οι προσεγγίσεις προ-επεξεργασίας στοχεύουν να κάνουν τα σύντομα κείμενα (τίτλοι των ανακοινώσεων ανάκλησης), που προέρχονται από ετερογενείς πηγές, πιο ευανάγνωστα και να αφαιρέσουν λέξεις και χαρακτηριστικά που μπορεί να θεωρηθούν ως θόρυβος. Η προτεινόμενη μεθοδολογία μας αποτελείται από τα ακόλουθα:

(α) Αφαίρεση αριθμών,

(β) Αφαίρεση συγκεκριμένων λέξεων ως θόρυβο,

(γ) Αντικατάσταση των λέξεων με τα λήμματα αυτών και αντικατάσταση όλων των λέξεων σε πεζά,

(δ) Χαρακτηρισμός (labeling).

Ο Πίνακας 5 συνοψίζει τους μετασχηματισμούς που εφαρμόστηκαν σε κάθε βήμα παρέχοντας ένα αντιπροσωπευτικό παράδειγμα.

Η βιβλιοθήκη του SpaCy που χρησιμοποιήθηκε σε αυτήν την Διατριβή, για τη διαδικασία αναγνώρισης οντοτήτων προϊόντων παρέχει ένα μοντέλο για την εξαγωγή-αντιστοίχιση οντοτήτων. Σε αυτή την περίπτωση η βιβλιοθήκη, προσφέρει τη δυνατότητα χαρακτηρισμού των ζητούμενων τιμών ανεξάρτητων μεταβλητών που ακολουθεί τη μορφή (θέση αρχικού χαρακτήρα, θέση τελικού χαρακτήρα, τύπος οντότητας). Ως παράδειγμα παρέχουμε την ετικέτα (13, 22, 'ΠΡΟΙΟΝ'), η οποία αντιστοιχεί σε ένα προϊόν που υπάρχει στον 13ο έως τον 22ο χαρακτήρα του συγκεκριμένου κειμένου. Αξίζει να αναφέρουμε ότι δεν περιέχουν σε όλα τα κείμενα ακριβή αντιστοίχιση και αυτές οι περιπτώσεις σημειώνονται ως (0,0, " ") παρέχοντας τη δυνατότητα στο μοντέλο μας να αναγνωρίσει την έλλειψη μιας οντότητας. Για το λόγο αυτό, διαχωρίζουμε τα δεδομένα (κατά τη διάρκεια του διαχωρισμού δοκιμών-εκπαίδευσης) με τέτοιο τρόπο ώστε τόσο τα σύνολα δεδομένων εκπαίδευσης όσο και τα δοκιμαστικά να περιέχουν παραδείγματα με και χωρίς ύπαρξη της οντότητας του προϊόντος.



**Πίνακας 5** Βήματα προ-επεξεργασίας για εργασία ταξινόμησης των ανακοινώσεων τροφίμων (κάθετα)

<b>Initial Text</b>	Aflatoxins (b1 = 27.2; tot. = 37 g/kg - ppb) in pistachios from the united states, manufactured in turkey
<b>Remove numbers and symbols</b>	Aflatoxins b tot gkg ppb in pistachios from the united states manufactured in turkey
<b>Remove noise</b>	Aflatoxins in pistachios from the united states manufactured in turkey
<b>Lemmatization and lower case</b>	Aflatoxin in pistachio from manufacture in
<b>Annotation</b>	(13, 22, 'PRODUCT')

Για την εργασία πρόβλεψης χρονοσειρών, η σταθερότητα των δεδομένων ελέγχθηκε χρησιμοποιώντας τον έλεγχο *Augmented Dickey-Fuller* [195]. Αυτή η διαδικασία είναι μείζονος σημασίας για οποιαδήποτε μέθοδο πρόβλεψης που εκμεταλλεύεται ιστορικά δεδομένα χρονοσειρών, καθώς αυτές οι μέθοδοι βασίζονται συνήθως στην υπόθεση ότι ο μηχανισμός παραγωγής δεδομένων δεν αλλάζει με την πάροδο του χρόνου. Επιπλέον, θα πρέπει να σημειωθεί ότι όσον αφορά τις προβλέψεις, τα χρονικά πλαίσια πρόβλεψης των ανακλήσεων μπορεί να είναι 4 μήνες, 6 μήνες ή 12 μήνες. Δεδομένου ότι οι προβλέψεις θα συμβάλουν στη διασφάλιση της ποιότητας και θα επιτρέψουν στους επαγγελματίες της ασφάλειας των τροφίμων να διασφαλίσουν τη συνέχεια της αλυσίδας εφοδιασμού τους, να ελαχιστοποιήσουν τους μελλοντικούς κινδύνους και τις οικονομικές απώλειες, έτσι επιλέξαμε το μικρότερο, δηλαδή ένα παράθυρο πρόβλεψης 4 μηνών. Με βάση το τελευταίο, ξεχωρίζουν τρεις επιλογές:

α. Χρήση του συνόλου δεδομένων ως έχει, με την ημερήσια συχνότητά του να οδηγεί σε ένα παράθυρο πρόβλεψης 120 χρονικών βημάτων. Αυτό δεν συνιστάται, καθώς οι περισσότερες από τις χρονοσειρές που παράγονται είναι αραιές, χωρίς προφανές μοτίβο.

β. Δειγματοληψία των δεδομένων σε εβδομάδες, με αποτέλεσμα το παράθυρο πρόβλεψης να είναι 16 βημάτων. Ενώ ορισμένα μοτίβα αρχίζουν να εμφανίζονται, τα δεδομένα σε ορισμένες περιπτώσεις εξακολουθούν να είναι πολύ αραιά και το σωρευτικό σφάλμα από την πρόβλεψη των 16 βημάτων είναι θεωρητικά σχετικά μεγάλο.

γ. Δειγματοληψία των δεδομένων σε περιόδους μηνών, παρέχοντας σταθερές χρονοσειρές με ορατά μοτίβα και χαμηλότερο θεωρητικό συσσωρευμένο σφάλμα πρόβλεψης, καθώς το παράθυρο έχει μειωθεί σε 4-χρονικά βήματα.

Τέλος, μια άλλη υπόθεση που ελέγξαμε ήταν εάν το συγκεκριμένο πρόβλημα πρόβλεψης είναι πιο αποδοτικό να μοντελοποιηθεί ως πολλά μονομεταβλητά μοντέλα ή ως ένα πολυμεταβλητό υποθέτοντας ότι υπάρχουν σύνθετες μη



γραμμικές αλληλεπιδράσεις χαρακτηριστικών στα δεδομένα μας όταν πρόκειται για όλες τις κατηγορίες ανακλήσεων τροφίμων. Κατά τη διαδικασία των πειραμάτων διαπιστώσαμε ότι 23 από τις 30 συνολικά χρονοσειρές ήταν “στάσιμες” σύμφωνα με το τεστ Augmented Dickey-Fuller που διεξήχθη αρχικά και το γεγονός ότι διαφέρουν σημαντικά στα στατιστικά χαρακτηριστικά (ελάχιστο, μέγιστο, διακύμανση κ.λπ.) που θα είχε αρνητική επίδραση στην απόδοση κάποιου πολυμεταβλητού μοντέλου ενώ θα ευνοούσε κάποια άλλα, αντ’ αυτού χρησιμοποιήθηκε η ποσοστιαία μεταβολή των χρονοσειρών. Αυτό επιτρέπει στο μοντέλο να γενικεύει καλύτερα (αριθμητικός ρυθμός) ενώ επίσης μετατρέπει τις υπόλοιπες μη στάσιμες χρονοσειρές σε σταθερές. Συγκεκριμένα, όσον αφορά τα δεδομένα, για να εφαρμόσουμε ποσοστιαία μεταβολή έπρεπε να αντιμετωπίσουμε μηδενικές τιμές. Στις περισσότερες από τις ανάλογες εργασίες, ο χειρισμός αυτών των υποθέσεων απαιτεί εμπειρογνωμοσύνη στον τομέα, καθώς δεν υπάρχει “σωστή” μεθοδολογία. Στην περίπτωση μας προστέθηκε μία «ανάκληση» κάθε μήνα σε κάθε κατηγορία, η οποία είναι ασήμαντη και δεν μεταφέρει πραγματική αλλαγή, για να αποφευχθεί η ύπαρξη μηδενικού στον παρονομαστή της ποσοστιαίας μεταβολής. Μαθηματικά, έστω:

$$x_i \frac{(x_{t+1}) - (x_{t-1} + 1)}{x_{t-1} + 1} \text{ η ποσοστιαία μεταβολή, } x_t \text{ είναι } x_{t-1} + 1 \text{ ο αριθμός των}$$

ανακλήσεων τον μήνα t.

### 3.3.4 Υποκατάστατα Δεδομένα

Το αρχικό μέγεθος δεδομένων μετά την προ-επεξεργασία είναι 30x147 και τα παραγόμενα υποκατάστατα δεδομένα έχουν ίσο μέγεθος, καταλήγοντας σε ένα συνδυασμένο σύνολο δεδομένων με 30x247 σημεία δεδομένων. Κατά την ανάλυσή μας, αναφερόμαστε στα αντίστοιχα δεδομένα με βάση τον αριθμό ταυτότητάς τους όπως αναφέρεται στον Πίνακας 6.

**Πίνακας 6** Κατηγορίες δεδομένων ανάκλησης τροφίμων

0	Food Additives and Flavorings	15	Other food products
1	Fruits and Vegetables	16	Food contact materials
2	Poultry meat and products	17	Non-alcoholic beverages
3	Prepared dishes and snacks	18	Fats and Oils
4	Dietetic foods, Food supplements, fortified foods	19	Sugars and Syrups
5	Feed materials	20	Nuts, Seeds and products
6	Honey and Royal Jelly	21	Cereals and Bakery products
7	Bivalve mollusks and products	22	Crustaceans and products
8	Meat and products	23	Eggs and products
9	Feed Additives	24	Alcoholic beverages
10	Milk and Milk products	25	Cocoa, Coffee and Tea



11	Herbs and Spices	26	Gastropods
12	Confectionery	27	Cephalopods and products
13	Soups, Broths, Sauces and Condiments	28	Pet Feed
14	Fish and products	29	Ices and Desserts

Οι πιο συχνά χρησιμοποιούμενες τεχνικές για τη δημιουργία υποκατάστατων δεδομένων για στατιστική ανάλυση μη γραμμικών διεργασιών περιλαμβάνουν τυχαία αναδιάταξη της αρχικής χρονοσειράς, υποκατάστατα μετασχηματισμένα με μετασχηματισμό Fourier, μετασχηματισμένα Fourier με προσαρμογή πλάτους (AAFT υποκατάστατα) και επαναλαμβανόμενα υποκατάστατα AAFT (IAAFT). Στην εργασία μας χρησιμοποιήσαμε τη μέθοδο IAAFT η οποία αντιμετωπίζει το ζήτημα της λεύκανσης φάσματος ισχύος, ως το κύριο μειονέκτημα της μεθόδου AAFT, εκτελώντας μια σειρά επαναλήψεων στις οποίες το φάσμα ισχύος ενός υποκατάστατου AAFT προσαρμόζεται ξανά στα δεδομένα του αρχικού πριν η διανομή κλιμακωθεί σε εκείνη των αρχικών δεδομένων

### 3.3.5 Μοντέλα Μηχανικής Μάθησης – Στατιστικά Μοντέλα

#### 3.3.5.1 Εργασία ταξινόμησης – Named Entity Recognition

Δεδομένου ότι ο στόχος μας είναι να ταξινομήσουμε το σύντομο κείμενο της ανάκλησης φαγητού με βάση το υποκείμενο προϊόν (απορρίπτοντας πιθανά θορυβώδη μοτίβα) και δεδομένου ότι δεν υπάρχει κάποιο προ-εκπαιδευμένο μοντέλο ανοιχτού κώδικα που να έχει καθοριστεί για την αναγνώριση προϊόντων τροφίμων για την άμεση εφαρμογή της μάθησης μεταφοράς (transfer learning) για την εργασία NER, προσεγγίσαμε το πρόβλημα εκπαιδύοντας από την αρχή ένα Βαθύ Νευρωνικό Δίκτυο. Στην εκπαίδευση ελήφθησαν υπόψη οι προσαρμοσμένοι σχολιασμοί «ΠΡΟΪΟΝΤΩΝ» όπως περιεγράφηκαν παραπάνω. Οι κεντρικές δομές δεδομένων στο προτεινόμενο μοντέλο είναι το Doc και το Vocab. Το αντικείμενο Doc κατέχει την ακολουθία των ενδείξεων και όλους τους σχολιασμούς τους. Τέλος, όσον αφορά τη επίλυση του προβλήματός μας, μια αρχιτεκτονική του δικτύου Residual CNN (ResNET) [196] αποδίδει τα καλύτερα αποτελέσματα όσον αφορά στις μετρήσεις αξιολόγησης που θα αναλυθούν λεπτομερώς στην επόμενη υπο-ενότητα. Στα παραδοσιακά νευρωνικά δίκτυα, κάθε επίπεδο τροφοδοτείται στο επόμενο επίπεδο. Σε ένα δίκτυο με υπολειμματικά μπλοκ, κάθε επίπεδο τροφοδοτείται στο επόμενο επίπεδο και απευθείας στα επίπεδα που απέχουν περίπου 2-3 βήματα μπροστά.

Στον τομέα της ανάλυσης εικόνας από H/Y, τα CNN (Convolution Neural Networks) χρησιμοποιούνται για τη μείωση διαστάσεων ή τη μείωση των πινάκων σε διανύσματα με τη χρήση πολλαπλών φίλτρων που εστιάζουν στη ρύθμιση του



διασκελισμού, του μεγέθους του πυρήνα, της επένδυσης κ.λπ. Στην υλοποίηση μας τα CNN χρησιμοποιούνται με τρόπο παρόμοιο με αυτό που περιγράφεται στο [197] αλλά φυσικά αξιοποιώντας τις ενημερωμένες λεπτομέρειες εφαρμογής για τη διευκόλυνση της εκπαίδευσης και της ακρίβειας. Το βασικό μπλοκ των CNN που χρησιμοποιεί είναι το trigram-cnn που συνενώνει τις ενσωματώσεις κάθε λέξης με αυτές των δύο γειτονικών λέξεων. Για παράδειγμα, εάν η διάσταση ενσωμάτωσης κάθε λέξης ήταν 128 διαστάσεων, αυτή η συνένωση θα δημιουργούσε ένα διάνυσμα 384 διαστάσεων που αντιπροσωπεύει αυτές τις τρεις λέξεις. Μετά από αυτό, χρησιμοποιείται ένα MLP για τη μείωση της διάστασης αναπαράστασης αυτής της εισόδου πίσω στην αρχική, μαθαίνοντας εκ νέου το νόημα κάθε λέξης με βάση τα συμφραζόμενα της. Ενώνοντας αυτά τα επίπεδα μαζί, αναδύεται ένα είδος «αποτελεσματικού δεκτικού πεδίου» όσον αφορά την όραση. Αυτό συμβαίνει λόγω ένωσης, καθιστώντας έτσι τις αναπαραστάσεις των διανυσμάτων ευαίσθητες σε πληροφορίες που βρίσκονται σε λέξεις πιο μακριά από την αρχική ερμηνεία τους. Μια άλλη πτυχή που ενδιαφέρει, είναι η χρήση υπολειμματικών συνδέσεων από επίπεδο σε επίπεδο. Το τελευταίο σημαίνει ότι η έξοδος κάθε επιπέδου είναι η έξοδος που παράγεται από το επίπεδο συν την εισαγωγή που διευκολύνει την εκπαίδευση. Αυτό έχει θεμελιώδη επίδραση καθώς υπονοεί ότι το διάνυσμα χώρου εξόδου κάθε συνελκτικού επιπέδου είναι πιθανό να είναι παρόμοιο με το χώρο εξόδου του διανύσματος εισόδου εξαιτίας της χρήσης και προώθησης της δυνατότητας εισαγωγής σε κάθε επίπεδο.

### 3.3.5.2 Εργασία πρόβλεψης χρονοσειράς

Όσον αφορά την προσέγγιση πρόβλεψης χρονοσειρών, έχουν χρησιμοποιηθεί αρκετά μοντέλα, όπως συνοψίζονται παρακάτω:

**DeepAR Estimator:** εφαρμόζει ένα μοντέλο Recurrent Neural Network (Επαναλαμβανόμενου Νευρωνικού δικτύου), ως υλοποίηση του περιγραφόμενου στο [198]. Πιο συγκεκριμένα, εφαρμόζει μια μεθοδολογία για την παραγωγή πιθανοτικών προβλέψεων, βασισμένη στην εκπαίδευση ενός μοντέλου επαναλαμβανόμενου νευρωνικού δικτύου σε πολλές σχετικές χρονοσειρές. Το επαναλαμβανόμενο νευρωνικό δίκτυο είναι ένα νευρωνικό δίκτυο τροφοδοσίας με εσωτερική μνήμη. Η «επανάληψη» εξηγεί το γεγονός ότι η παραγόμενη έξοδος αντιγράφεται και αποστέλλεται πίσω στο επαναλαμβανόμενο δίκτυο ως πρόσθετη είσοδος. Για τη λήψη μιας απόφασης σχετικά με κάθε έξοδο κάθε επιπέδου, λαμβάνει υπόψη την τρέχουσα είσοδο και την έξοδο που έχει μάθει από την προηγούμενη είσοδο. Η ικανότητα αυτή την καθιστά τέλεια υποψήφιο για το χειρισμό δεδομένων ακολουθίας.

**Simple Feed Forward Estimator:** εφαρμόζει ένα απλό μοντέλο Πολυεπίπεδου Νευρώνα (Perceptrons Multi-layer (MLP)) που προβλέπει τα επόμενα χρονικά





βήματα-στόχους με βάση τα προηγούμενα. Το MLP είναι ένας εποπτευόμενος αλγόριθμος εκμάθησης που μαθαίνει μια συνάρτηση με εκπαίδευση σε ένα σύνολο δεδομένων, όπου  $n$  είναι ο αριθμός των διαστάσεων για είσοδο και  $m$  είναι ο αριθμός διαστάσεων για έξοδο. Δεδομένου ενός συνόλου χαρακτηριστικών  $X = x_1, x_2, \dots, x_n$  και ενός στόχου  $y$ , μπορεί να μάθει μια μη γραμμική συνάρτηση προσέγγισης είτε για ταξινόμηση είτε για παλινδρόμηση. Εφαρμόστηκε όπως περιγράφεται στο [199].

**Deep Factor Estimator:** μια υλοποίηση του [200]. Χρησιμοποιεί ένα καθολικό μοντέλο RNN για την εκμάθηση μοτίβων σε πολλαπλές συσχετιζόμενες χρονοσειρές και ένα αυθαίρετο τοπικό μοντέλο για τη μοντελοποίηση των χρονοσειρών ανά χρονοσειρά.

**Seasonal Naive Estimator:** βασίζεται στην εποχικότητα των δεδομένων. Αυτό το μοντέλο προβλέπει ότι  $Y(T + k) = y(T + k - h)$  όπου  $T$  είναι ο χρόνος πρόβλεψης,  $k \in [0, \text{μήκος πρόβλεψης} - 1]$  και  $h = \text{μήκος εποχής}$ . Εάν μια χρονοσειρά είναι μικρότερη από το μήκος εποχής, τότε η μέση παρατηρούμενη τιμή χρησιμοποιείται ως πρόβλεψη.

**WaveNet Estimator:** βασίζεται στην αρχιτεκτονική WaveNet [201]. Το WaveNet, είναι ένα βαθύ νευρωνικό δίκτυο που κατασκευάστηκε για τη δημιουργία ακατέργαστων κυματομορφών ήχου. Το μοντέλο είναι πλήρως πιθανοτικό και αυτοπαλινδρομικό, με την προγνωστική κατανομή για κάθε δείγμα ήχου να εξαρτάται από όλα τα προηγούμενα, αποδίδοντας αποτελέσματα τελευταίας τεχνολογίας σε εργασίες όπως μετατροπές κειμένου σε ομιλία, διάσπαση πηγών κ.λπ.

Πέρα από την εφαρμογή και τον πειραματισμό των προαναφερθέντων μοντέλων, ερευνήθηκε και αναπτύχθηκε μια προσέγγιση ενισχυτικής μάθησης, η οποία περιλαμβάνει τόσο ένα προσαρμοσμένο περιβάλλον για χρονοσειρές όσο και έναν πράκτορα-μοντέλο βασισμένο σε μοντέλα βαθιάς μάθησης. Αυτό το προσαρμοσμένο περιβάλλον θα μπορούσε επίσης να χρησιμοποιηθεί σε άλλες περιπτώσεις χρονοσειρών εκτός από την πρόβλεψη ανακλήσεων τροφίμων. Ο πράκτορας RL αλληλοεπιδρά με αυτό το περιβάλλον σε διακριτά χρονικά βήματα. Κάθε στιγμή  $t$ , ο πράκτορας λαμβάνει μια παρατήρηση  $O_t$ , η οποία τυπικά περιλαμβάνει την ανταμοιβή  $R_t$  και την κατάσταση  $S_t$ . Σε αυτό το προσαρμοσμένο περιβάλλον οι καταστάσεις από τις ιστορικές αξίες και τις ανταμοιβές παράγονται εφαρμόζοντας την κάτωθι συνάρτηση επιβράβευσης. Ο πράκτορας μετά επιλέγει μια ενέργεια από το σύνολο των διαθέσιμων ενεργειών  $A_t$ , η οποία στη συνέχεια αποστέλλεται στο περιβάλλον. Το περιβάλλον μετακινείται σε μια νέα κατάσταση και καθορίζεται η ανταμοιβή που σχετίζεται με τη μετάβαση  $(S_t, A_t, S_{t+1})$ .

Ο σχεδιασμός της συνάρτησης ανταμοιβής εξαρτάται από τις ενέργειες που θέλουμε να ευνοήσει ο πράκτορας. Σε αυτό το έργο, η ανταμοιβή  $R_t$  τη στιγμή  $t$  είναι:

$$R_t = \frac{1}{\log(\text{MSE}(Y_t, \hat{Y}_t))} + k \quad (2)$$



όπου  $Y_t$  είναι ο πραγματικός αριθμός ανακλήσεων για το δεδομένο χρονικό βήμα,  $\hat{Y}_t$  είναι το προβλεπόμενο και  $k$  είναι ένας μικρός θετικός ακέραιος.

Με την προσέγγιση αυτή, γνωστή ως «Q-Learning» ο πράκτορας επιδιώκει να μάθει μια πολιτική που μεγιστοποιεί τη συνολική ανταμοιβή για το επιλεγμένο σύνολο ενεργειών σε ένα δεδομένο περιβάλλον. Στην πράξη, ένας πίνακας  $Q$  είναι ένας πίνακας της μορφή[κατάσταση, πράξη] όπου αποθηκεύονται οι τιμές κάθε ενέργειας. Μια αρχιτεκτονική DQN με επανάληψη εμπειρίας υλοποιήθηκε όπως περιγράφεται παρακάτω:

---

**Algorithm 1: DEEP Q-LEARNING with experience replay**

---

```
1 Initialize replay memory  $D$  to capacity  $N$ 
2 Initialize action function  $Q$  with random weights  $\vartheta$ 
3 Initialize target action function  $\hat{Q}$  with weights  $\bar{\vartheta}$ 
4 for  $episode = 1, M$  do
5   Initialize the environment and get the state  $s_1$ 
6   for  $t = 1, T$  do
7     with probability  $\epsilon$  select a random action  $\alpha_t$ 
8     otherwise select  $\alpha_t = \operatorname{argmax}_\alpha Q(s_t, \alpha; \vartheta)$ 
9     execute action  $\alpha_t$  in the environment and observe the reward  $r_t$  and next
       state  $s_{t+1}$ 
10    store transition  $(s_t, \alpha_t, r_t, s_{t+1})$  in  $D$ 
11    use a random batch of transitions from  $D$ 
12    if  $episode$  terminates at step  $j + 1$  then
13      | set  $y_j = r_j$ 
14    end
15    else
16      | set  $y_j = r_j + \gamma \max_\alpha Q(s_{j+1}, \alpha; \bar{\vartheta})$ 
17    end
18    perform gradient descent on  $(y_j - Q(s_j, \alpha_j; \vartheta))^2$  with respect to parameters  $\vartheta$ 
19    every  $C$  steps reset  $\hat{Q} = Q$ 
20  end
21 end
```

---

Όπου  $\gamma$  είναι ο συντελεστής «έκπτωσης», που χρησιμοποιείται για να εξισορροπήσει τη σημασία της μελλοντικής και της άμεσης ανταμοιβής, το  $\alpha$  (ποσοστό μάθησης) ορίζει το ποσοστό στο οποίο η πρόσφατα υπολογισμένη τιμή του πίνακα  $Q$  επηρεάζει την παλιά. Σε αυτό το σημείο, θα πρέπει να τονιστεί ότι ο συγκεκριμένος αλγόριθμος αποδίδει πολύ καλά αποτελέσματα στις περιπτώσεις μικρών χώρων ενεργειών και καταστάσεων. Στις περιπτώσεις απαγορευτικού μεγέθους  $S_t, A_t$  προτείνεται ένα νευρωνικό δίκτυο για την προσέγγιση και συμπίεση του πίνακα  $Q$ , όπου η ενημέρωση των βαρών  $w$  αντιστοιχεί στην ενημέρωση των τιμών του  $Q$ . Μία βελτίωση του αλγόριθμου Deep Q-learning έχει επιτευχθεί χρησιμοποιώντας ένα πρόσθετο νευρωνικό δίκτυο με την ίδια αρχιτεκτονική αλλά με σταθερά βάρη  $\hat{w}$  που ενημερώνονται κάθε επανάληψη για να σπάσει τη συσχέτιση μεταξύ των ενημερωμένων τιμών  $w$  και  $\Delta w$ .



### 3.4 Μηχανική Μάθηση & ποιότητα χοιρινού κρέατος.

#### 3.4.1 Σύνολο δεδομένων

Στο πλαίσιο αυτής της έρευνας, το χρησιμοποιούμενο σύνολο δεδομένων παρασχέθηκε από την ILVO,<sup>1</sup> στο πλαίσιο ενός ερευνητικού προγράμματος HORIZON 2020. Τα παρεχόμενα δεδομένα παρουσιάστηκαν σε μορφή πίνακα που αποτελούνταν από μετρήσεις που υποδηλώνουν τα περιβαλλοντικά και υγειονομικά χαρακτηριστικά των χοίρων και μετρήθηκαν σε διάφορες φάρμες και σφαγεία. Η διαδικασία για την ανάκτηση αυτών των δεδομένων περιγράφεται στα [202] και [203]. Το σύνολο δεδομένων στην τελική του έκδοση αποτελείται από 172 στήλες (χαρακτηριστικά) και 16791 σειρές (δείγματα). Μια σύντομη περιγραφή του συνόλου δεδομένων παρουσιάζεται στον Πίνακα 7. Ενώ ο τρόπος με τον οποίο κατανέμονται οι missing values μέσα στο σύνολο δεδομένων απεικονίζεται στην Εικόνα 13. Η ετικέτα (δηλαδή, εξαρτημένη μεταβλητή) προήλθε από μια αισθητηριακή βαθμολόγηση από ειδικά εκπαιδευμένο προσωπικό που περιγράφεται στο [202], επίσης γνωστή ως η οσφρητική αξιολόγηση της κηλίδωσης κάπρου ή η μέθοδος της ανθρώπινης μύτης. Αξίζει να σημειωθεί ότι, αυτή παραμένει μια υποκειμενική μέθοδος με περιορισμένη ακρίβεια [204] [205].

**Πίνακας 7** Περιγραφή του συνόλου δεδομένων για την περίπτωση κηλίδωσης του κρέατος των κάπρων

Dataset Info		
Number of variables	172	
	Numeric	101
	Categorical	48
	Boolean	21
	Date	2
Number of Observations	16791	
Total Missing(%)	18%	

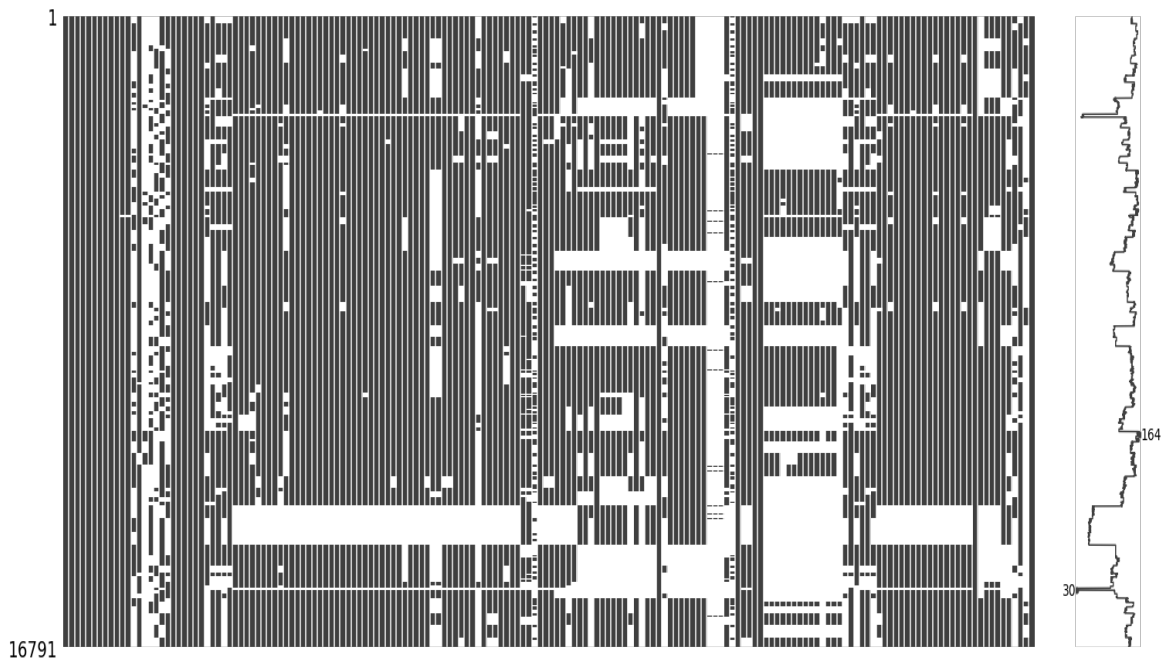
Επιπλέον, είναι σημαντικό να σημειωθεί ότι τα περισσότερα από τα δεδομένα αντιπροσωπεύουν τη μη μολυσμένη (αρνητική) κλάση, η οποία εγείρει την πρόκληση του μη ισορροπημένου συνόλου δεδομένων. Συγκεκριμένα, προσεγγίσαμε την περίπτωση χρήσης ως δυαδική ταξινόμηση (tainted=1, not tainted=0), καταλήγοντας σε ένα εξαιρετικά μη ισορροπημένο σύνολο δεδομένων

<sup>1</sup> <https://ilvo.vlaanderen.be/en>

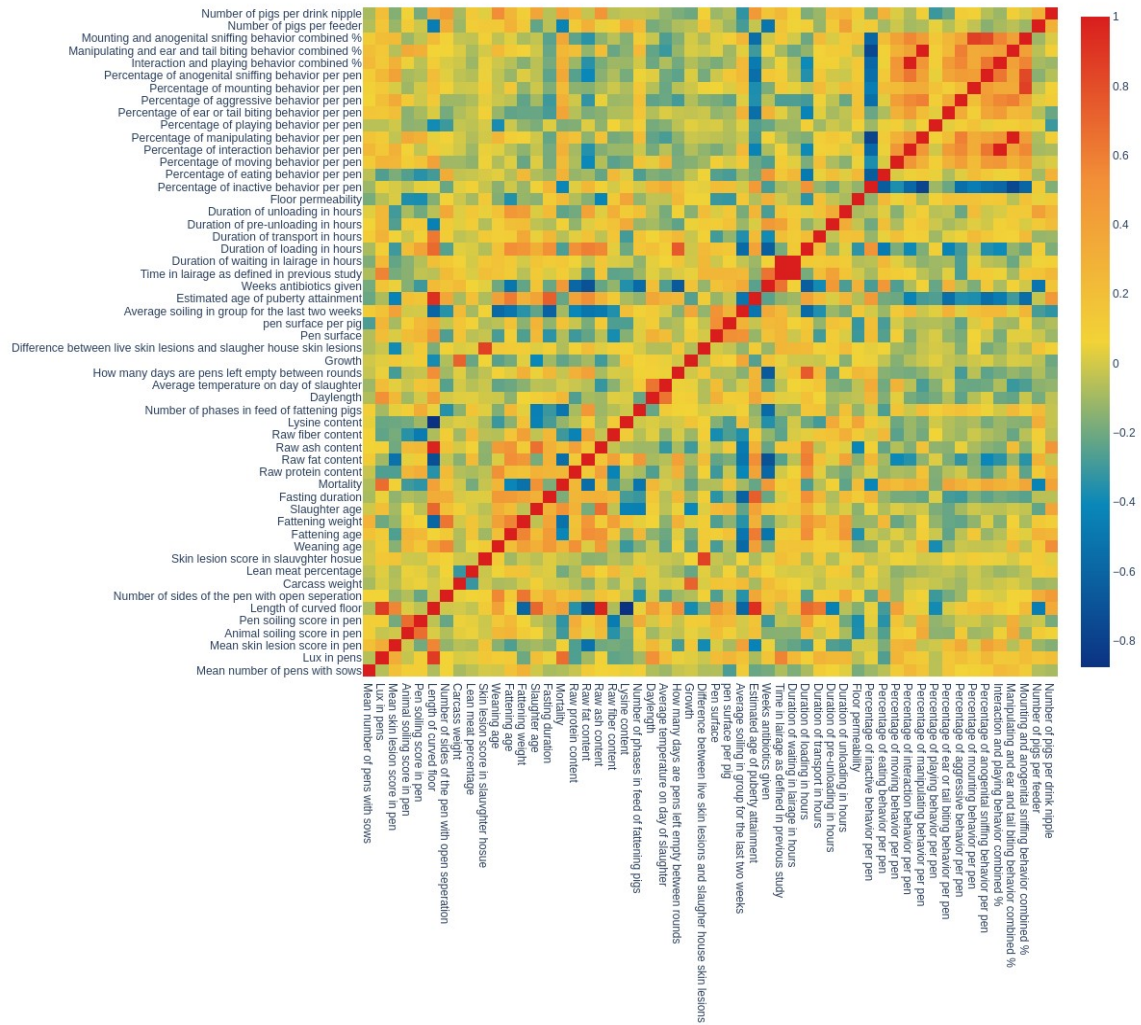


(αναλογία ανισοροπίας=1/50). Κατά συνέπεια, η ανάπτυξη ενός ισχυρού ταξινομητή για συγκεκριμένη περίπτωση χρήσης αποτελεί μια ανοιχτή ερευνητική πρόκληση.

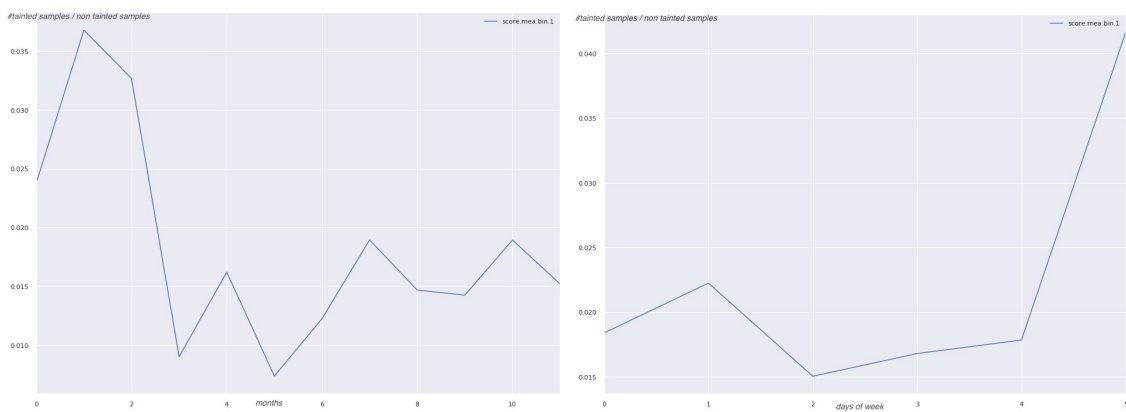
Ένα εμφανές χαρακτηριστικό, εκτός από την ύπαρξη συσχετισμένων δεδομένων, όπως απεικονίζονται στην Εικόνα 14 και τον παράγοντα ανισοροπίας των κατηγοριών, είναι η επίδραση των μεταβλητών ημερομηνίας, όπως απεικονίζεται στην Εικόνα 15. Αυτές οι ημερομηνίες αναφέρονται στην ημερομηνία που πραγματοποιήθηκε η σφαγή για κάθε εξεταζόμενο δείγμα.



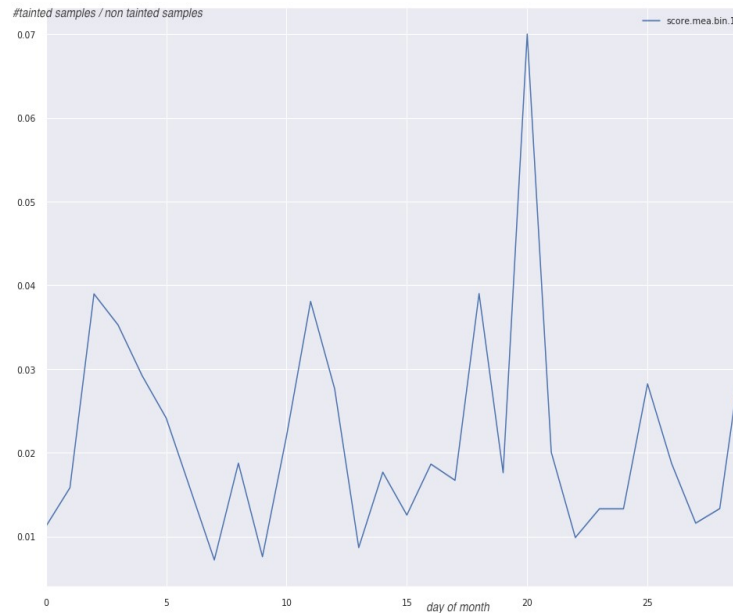
**Εικόνα 13** Κατανομή κενών τιμών στο σύνολο δεδομένων της περίπτωσης κηλίδωσης κάπρων όπου αποτελείται από 172 στήλες (δηλαδή, χαρακτηριστικά) και 16791 σειρές (δηλαδή δείγματα).



Εικόνα 14 Heatmap συσχέτισης μεταξύ των 172 μεταβλητών του συνόλου δεδομένων.



(a) Distribution of the ratio of the tainted samples per slaughter per month (b) Distribution of the ratio of the tainted samples per slaughter per day of the week



**Εικόνα 15** Κατανομή της αναλογίας των μολυσμένων δειγμάτων προς την ημερομηνία σφαγής σε διαφορετικά χρονικά πλαίσια

### 3.4.2 Αρχιτεκτονική

Στο πλαίσιο της Διατριβής στοχεύσαμε στη χρήση μεθόδων μηχανικής μάθησης και των στατιστικών μεθόδων για τον εντοπισμό των πιο σημαντικών παραγόντων που σχετίζονται με τη φάρμα και τη σφαγή για τη μείωση της «κηλίδωσης του κρέατος του κάπρου» βασισμένοι σε ιστορικά στοιχεία από 16 αγροκτήματα και τέσσερα σφαγεία.

Στην παρούσα διατριβή προσεγγίζουμε τον ερευνητικό στόχο με διάφορους τρόπους. Πρώτον, εφαρμόσαμε μια Εξηγητική Ανάλυση Δεδομένων (Explanatory Data Analysis – EDA) από την οποία απορρίψαμε πολλά χαρακτηριστικά για είτε λόγω υψηλού ποσοστού τιμών που λείπουν και την υψηλή συσχέτιση με άλλες μεταβλητές. Στη συνέχεια, εκπαιδεύτηκαν και αξιολογήθηκαν διάφορα μοντέλα μηχανικής μάθησης για ταξινόμηση για την επιλογή της καταλληλότερης οικογένειας μοντέλων. Δεδομένου ότι το σύνολο δεδομένων μας ήταν εξαιρετικά μη ισορροπημένο, αξιοποιήθηκαν και οι μέθοδοι υπερδειγματοληψίας και υποδειγματοληψίας. Τέλος προτείνεται και μια προσέγγιση που περιλαμβάνει έναν προσαρμοσμένο γενετικό αλγόριθμο για την επιλογή χαρακτηριστικών, καταλήγοντας σε διαφορετικά εκπαιδευμένα μοντέλα από τα οποία προέκυψε η σημασία των χαρακτηριστικών.

### 3.4.3 Προ-επεξεργασία



Τα χαρακτηριστικά συλλογής δεδομένων μπορεί να περιλαμβάνουν ασυνεπή και/ή ελαττωματικά δεδομένα λόγω ασυνέπειας στην καταγραφή, ανθρώπινων σφαλμάτων ή σφαλμάτων αισθητήρα. Στη συνέχεια, εφαρμόστηκαν συγκεκριμένα βήματα προ-επεξεργασίας για την αφαίρεση του θορύβου και την υποβοήθηση των ανεπτυγμένων μοντέλων στη γενίκευση σε άγνωστα δεδομένα, συγκεκριμένα:

α. Η πλειοψηφία των μοντέλων μηχανικής μάθησης δεν μπορεί να χειριστεί τις τιμές που λείπουν. Στην περίπτωση των τιμών που λείπουν, υπάρχουν πολλοί τρόποι αντιμετώπισης των ημιτελών δεδομένων, είτε παραλείποντας ολόκληρη την εγγραφή που περιέχει πληροφορίες είτε συμπληρώνοντας τις τιμές που λείπουν [206]. Στην περίπτωση μας, συμπληρώσαμε τις τιμές που λείπουν με μια αρνητική τιμή (σταθερά), δεδομένου ότι όλα τα χαρακτηριστικά περιέχουν θετικές τιμές. Αυτή η προσέγγιση μπορεί να προσφέρει τη δυνατότητα εξαγωγής/εκμετάλλευσης οποιασδήποτε λανθάνουσας πληροφορίας (γιατί έλειπε αυτή η τιμή;). Επίσης, διερευνήσαμε εάν ο αριθμός των τιμών που λείπουν ή οι τιμές Not-A-Number (NaN) σε διατηρεί πληροφορία που σχετίζεται με τη χρώση κάπρου. Για το σκοπό αυτό, ένα μονο-μεταβλητό μοντέλο λογιστικής παλινδρόμησης εφαρμόζεται μόνο σε μια τεχνητή μεταβλητή με βάση τον αριθμό των τιμών NaN. Η μετρική F-1 αυτού του μοντέλου ήταν 0,1, υποδηλώνοντας ότι υπάρχει μια ασθενής εξάρτηση.

β. Σε πολυμεταβλητά μοντέλα, υπάρχει η πιθανότητα κατά τον χειρισμό μεταβλητών με στατιστικές διαφορές μεταξύ των χαρακτηριστικών (ελάχιστο, μέγιστο, διακύμανση κ.λπ.) να βλάψει την απόδοση του μοντέλου ευνοώντας ορισμένα συγκεκριμένα χαρακτηριστικά. Αυτό συμβαίνει καθώς ορισμένες αντικειμενικές συναρτήσεις είναι επιρρεπείς στο να ζυγίζουν περισσότερα χαρακτηριστικά με ένα ευρύτερο φάσμα τιμών. Επομένως, το εύρος όλων των χαρακτηριστικών θα πρέπει να κλιμακωθεί έτσι ώστε κάθε χαρακτηριστικό να συνεισφέρει περίπου αναλογικά στην τελική απώλεια που δίνεται από την αντικειμενική συνάρτηση. Ωστόσο, τα δέντρα ταξινόμησης μονότονους μετασχηματισμούς μεμονωμένων ταξινομημένων μεταβλητών. Ο λόγος είναι ότι τα δέντρα ταξινόμησης χωρίζουν τους κόμβους με όριο. Οι μονοτονικοί μετασχηματισμοί δεν μπορούν να αλλάξουν τους πιθανούς τρόπους διαίρεσης σημείων δεδομένων με όριο [207].

γ. Όπως αναφέρθηκε παραπάνω, το σύνολο δεδομένων αποτελείται από τύπου μεταβλητών ημερομηνίας, αριθμητικές και κατηγορηματικές. Ειδικά όσον αφορά τις κατηγορικές τιμές, θα πρέπει να κωδικοποιηθούν ώστε να ενσωματωθούν σε ένα τελικό μοντέλο μεταξύ των αριθμητικών μεταβλητών. Στην έρευνά μας, χρησιμοποιήσαμε διάφορους τύπους κατηγορικών τεχνικών κωδικοποίησης δεδομένων, όπως One-hot Encoding (δημιουργεί ένα δυαδικό χαρακτηριστικό για κάθε κατηγορία), Label Encoding (μετατροπή των ετικετών σε αριθμητική μορφή με βάση την κατηγορία), και Κωδικοποίηση συχνότητας (αντικατάσταση κατηγορικών τιμών χαρακτηριστικών με τη συχνότητα της κατηγορίας στο σύνολο δεδομένων). Δεδομένου ότι δεν σημειώθηκε σημαντική



διαφορά μεταξύ αυτών των προσεγγίσεων, η προτιμώμενη προσέγγιση είναι επίσης η βασική ιδέα πίσω από την προ-επεξεργασία κατηγορικών χαρακτηριστικών CatBoost. Ονομάζεται Ordered Target Encoding και βασίζεται στην έννοια της τεχνητής χρονικής εξάρτησης. Τα στατιστικά στοιχεία στόχου για κάθε παράδειγμα υπολογίζονται χρησιμοποιώντας μόνο τον στόχο από το ιστορικό.

δ. Όπως και στην προηγούμενη εφαρμογή μοντέλων μηχανικής μάθησης, είναι ιδιαίτερης σημασίας το σχήμα αξιολόγησης και επικύρωσης της προσέγγισης. Μια λανθασμένη επικύρωση οδηγεί σε υπερβολικά αισιόδοξες προσδοκίες για το τι θα συμβεί όταν το μοντέλο εφαρμοστεί στην παραγωγική διαδικασία. Ουσιαστικό οπότε είναι η βήμα διαχωρισμού του συνόλου δεδομένων σε σύνολα εκπαίδευσης, επικύρωσης και δοκιμών με βάση τις απαιτήσεις κάθε χρησιμοποιούμενης μεθόδου.

### 3.4.4 Μοντέλα Μηχανικής Μάθησης – Στατιστικά Μοντέλα

Στο πλαίσιο της διατριβής αναφορικά με την περίπτωση της κηλίδωσης του κρέατος των κάπρων, εφαρμόστηκαν διάφορα στατιστικά μοντέλα και αλγόριθμοι μηχανικής μάθησης. Παρακάτω αναφέρονται τα βασικότερα αυτών μαζί με κάποια από τα αποτελέσματα τους:

Η **Αμοιβαία Πληροφορία (Mutual Information)** των κορυφαίων χαρακτηριστικών σχετικά με την "Ετικέτα" (δηλαδή, εξαρτημένη μεταβλητή που δηλώνει τη κηλίδωση κάπρου) απεικονίζεται στον Πίνακα 8. Η αμοιβαία πληροφορία δύο τυχαίων μεταβλητών μετρά την εξάρτηση μεταξύ των δύο μεταβλητών. Συγκεκριμένα, ποσοτικοποιεί την «ποσότητα πληροφοριών» (σε μονάδες όπως το Shannon (σύμβολο Sh), επίσης γνωστό ως bit, είναι μια μονάδα πληροφοριών και εντροπίας που ορίζονται από το IEC 80000-13) που λαμβάνεται για τη μία τυχαία μεταβλητή μέσω παρατήρησης της άλλης μεταβλητής. Η έννοια της αμοιβαίας πληροφορίας είναι στενά συνδεδεμένη με την εντροπία μιας τυχαίας μεταβλητής, μια θεμελιώδη έννοια στη θεωρία της πληροφορίας που ποσοτικοποιεί την αναμενόμενη «ποσότητα πληροφοριών» που διατηρείται σε μια τυχαία μεταβλητή. Έστω  $(X, Y)$  ένα ζεύγος τυχαίων μεταβλητών με τιμές στο διάστημα  $X \times Y$ . Εάν η κοινή κατανομή τους είναι  $P(X, Y)$  και οι οριακές κατανομές είναι  $P_X$  και  $P_Y$ , η αμοιβαία πληροφορία ορίζεται ως

$$I(X; Y) = D_{KL}(P_{(X,Y)} // P_X \otimes P_Y) \quad (3)$$

όπου  $KL$  είναι η απόκλιση Kullback–Leibler.



**Πίνακας 8** Αποτελέσματα Αμοιβαίας πληροφορίας μεταξύ των χαρακτηριστικών και την τιμής που χρησιμοποιείται ως ένδειξη κηλίδωσης του κρέατος κάπρου.

Factor name	mutual information	Factore name	mutual information
Ratio of aggressive behavior per pen	0.00566	Mortality	0.00262
Ratio of inactive behavior per pen	0.00360	Floor permeability	0.00261
Weeks antibiotics given	0.00346	Manip. & ear-tail biting behavior	0.00237
Live skin lesions/slaughter house skin lesions	0.00324	Pen surface	0.00234
Mean number of sows per pen	0.00314	Lux in pens	0.00230
Percentage of eating behavior per pen	0.00308	dietary lysine content	0.00228
Number of pigs per drink nipple	0.00292	Length of curved floor	0.00224
Meat boar	0.00287	Fasting duration	0.00219
Average temperature on day of slaughter	0.00283	Width of pen	0.00215
Number of pigs per feeder	0.00278	antibiotica	0.00211
Aggregate variable of cleaning method	0.00271	aspirine	0.00209
Mean skin lesion score in pen	0.00270	Duration of Loading	0.00207
Light aggregate	0.00267		
Deaths caused by aggression	0.00264		

Το  $\chi^2$  είναι μια στατιστική μετρική που μετρά την ανεξαρτησία ενός χαρακτηριστικού από τις ετικέτες κλάσεων. Ένα  $\chi^2$  **ΤΕΣΤ**, που αναφέρεται επίσης ως δοκιμή, είναι κάθε στατιστικό τεστ υπόθεσης στο οποίο η δειγματοληπτική κατανομή από το στατιστικό αποτέλεσμα της δοκιμής είναι μια  $\chi^2$  κατανομή, όταν η μηδενική υπόθεση είναι αληθής. Τα αποτελέσματα ενός στατιστικού τεστ που ακολουθούν μια  $\chi^2$  κατανομή προκύπτουν από την υπόθεση των ανεξάρτητων κανονικά καταμεμημένων δεδομένων, τα οποία ελέγχονται σε πολλές περιπτώσεις με το κεντρικό οριακό θεώρημα. Ένα  $\chi^2$  τεστ μπορεί στη συνέχεια να χρησιμοποιηθεί για να απορρίψουμε τη μηδενική υπόθεση, ότι τα δεδομένα είναι ανεξάρτητα. Αυτό είναι σχετικά κοινό με μη ισορροπημένα σύνολα δεδομένων. Ο Πίνακας 9 δείχνει τις τιμές με τις υψηλότερες τιμές Chi και τις τιμές p.

$$\chi^2 = \sum (O_i - E_i)^2 / E_i^2 \quad (4)$$

όπου  $O_i$  είναι η  $i$  - η παρατήρηση και  $E_i$  η αναμενόμενη τιμή.

**Πίνακας 9** Αποτελέσματα Chi μεταξύ των χαρακτηριστικών και την τιμής που χρησιμοποιείται ως ένδειξη κηλίδωσης του κρέατος κάπρου.

Factors	chi square	p value
<b>Fasting duration</b>	5.104339	0.023
<b>Number of free drink nipples</b>	3.976639	0.046
<b>Ventilation system</b>	3.434663	0.063
<b>Mean number of pens with sows</b>	3.321317	0.068
<b>Treatment 2</b>	3.262683	0.070
<b>Age group</b>	2.876586	0.089
<b>Seperation material</b>	2.732896	0.098
<b>Mean number of sows per pen</b>	2.509816	0.113



Are sows present in the compartment	2.265103	0.132
Season	2.26426	0.132
Duration of waiting in lairage	2.064702	0.150

Από τον πίνακα παρατηρούμε ότι τα κορυφαία χαρακτηριστικά που αποδίδουν υψηλή τιμή  $X^2$  (και χαμηλή τιμή  $p$ ) είναι κυρίως δείκτες (δηλαδή, τυχαίες μεταβλητές χωρίς χρήσιμες πληροφορίες). Ως εκ τούτου η συγκεκριμένη προσέγγιση δεν μας επιτρέπει να εξάγουμε συμπεράσματα.

Επιπλέον, ως μια άλλη ένδειξη της μονομεταβλητής εξάρτησης της ετικέτας με κάθε ανεξάρτητη μεταβλητή, υπολογίστηκε η βαθμολογία F-1 της εφαρμόζοντας την μέθοδο **Logistic Regression**. Τα κορυφαία 10 χαρακτηριστικά (δηλαδή, με υψηλότερες βαθμολογίες F-1) απεικονίζονται στον Πίνακα 10.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (5)$$

$$Precision = \frac{TP}{TP + FP} \quad (6)$$

$$Recall = \frac{TP}{TP + FN} \quad (7)$$

$$F1 = \frac{2 * Precision * Recall}{Precision + Recall} = \frac{2 * TP}{2 * TP + FP + FN} \quad (8)$$

όπου  $TP$  είναι τα πραγματικά θετικά αποτελέσματα,  $TN$  είναι τα πραγματικά αρνητικά,  $FP$  τα ψευδώς θετικά αποτελέσματα και  $FN$  τα ψευδώς αρνητικά αποτελέσματα.

**Πίνακας 10** Αποτελέσματα Logistic Regression μεταξύ των χαρακτηριστικών και την τιμής που χρησιμοποιείται ως ένδειξη κηλίδωσης του κρέατος κάπρου.

Factor Name	F1
Pen separation material	0.110914
Are pens present in the compartment	0.102091
Depth of drinking place	0.10102
Dietary crude protein content	0.100936
Season	0.100719
Type of feed = meal	0.100659
Frequency of anogenital sniffing behavior per pen	0.100135
Mean skin lesion score per pen	0.099887
Number of pens in slaughter batch	0.099567
Duration of pre-unloading (hours)	0.099379



Επιπλέον, διεξήχθησαν διάφορες στατιστικές δοκιμές με βάση το ίδιο σύνολο δεδομένων και παρουσιάστηκαν στο [202], από όπου αποδεικνύεται, μεταξύ άλλων, ότι σύμφωνα με διωνυμικά μονομεταβλητά μικτά μοντέλα, οι κάπροι με υψηλότερη βαθμολογία δερματικών αλλοιώσεων είχαν υψηλότερες πιθανότητες να έχουν χρώση κάπρου ( $P = 0,031$ ), όπως και οι παχύτεροι κάπροι ( $P < 0,001$ ). Στο διωνυμικό πολυμεταβλητό μικτό μοντέλο, το ποσοστό άπαχου κρέατος ( $P < 0,001$ ) και η εξωτερική μέση θερμοκρασία ( $P = 0,005$ ) παρέμειναν ως οι μόνοι σημαντικοί παράγοντες. Με βάση αυτά τα αποτελέσματα, μπορούμε να συμπεράνουμε ότι αυτοί οι στατιστικά σημαντικοί παράγοντες επηρεάζουν, τουλάχιστον εν μέρει, τον επιπολασμό της κηλίδωσης κάπρου.

### 3.4.4.1 Αρχική μεθοδολογία και ευρήματα

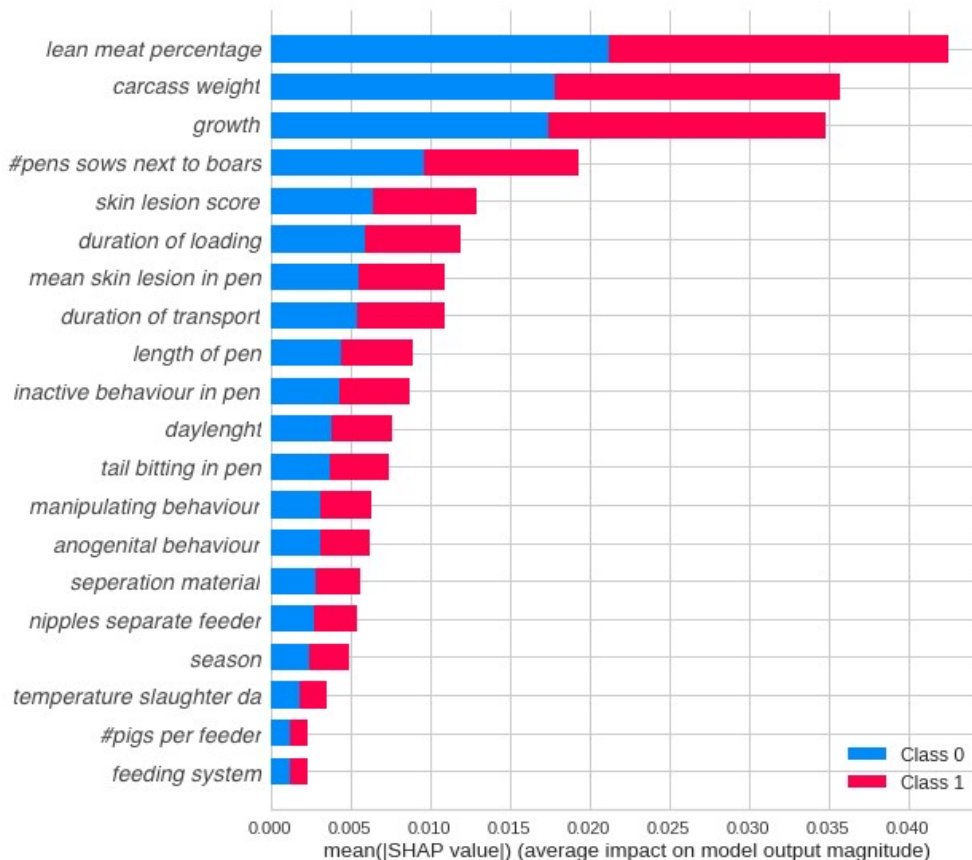
Αρχικά, προκειμένου να δοθεί μια επισκόπηση του πόσοι τύποι μοντέλων μηχανικής μάθησης μπορούν να αξιοποιηθούν στο συγκεκριμένο πρόβλημα εφαρμόσαμε τα πιο διαδομένα μοντέλα. Επιπλέον η μέθοδος SMOTE χρησιμοποιήθηκε για να αντισταθμίσει την ανισορροπία των δεδομένων. Είναι επίσης επιτακτική ανάγκη να επιλέξουμε την καταλληλότερη μετρική για την αξιολόγησή της απόδοσης των διαφορετικών μοντέλων μας. Δεδομένου ότι το σύνολο δεδομένων είναι σε μεγάλο βαθμό μη ισορροπημένο, με τη λιγότερο προεξέχουσα κατηγορία να είναι υψηλής σημασίας, η ακρίβεια δεν φαίνεται σαν μια σωστή μέτρηση αξιολόγησης [209], [210].

**Πίνακας 11** Αποτελέσματα διαφόρων μοντέλων μηχανικής μάθησης στην εργασία ταξινόμησης των δειγμάτων χοιρινού κρέατος ως προς την εκτίμηση ύπαρξης κηλίδωσης κάπρου.

	<i>Model</i>	<i>Accuracy</i>	<i>AUC</i>	<i>Recall</i>	<i>Prec.</i>	<i>F1</i>	<i>Kappa</i>	<i>MCC</i>	<i>TT (Sec)</i>
<i>lr</i>	Logistic Regression	0.9799	0.585	0	0	0	0	0	0.523
<i>knn</i>	K Neighbors Classifier	0.9799	0.5016	0	0	0	0	0	0.32
<i>ridge</i>	Ridge Classifier	0.9799	0	0	0	0	0	0	0.028
<i>ada</i>	Ada Boost Classifier	0.9797	0.5556	0	0	0	-0.0003	-0.0008	0.141
<i>catboost</i>	CatBoost Classifier	0.9796	0.5312	0	0	0	-0.0007	-0.0017	6.536
<i>lightgbm</i>	Light Gradient Boosting Machine	0.9794	0.506	0	0	0	-0.0009	-0.0018	0.093
<i>xgboost</i>	Extreme Gradient Boosting	0.9788	0.4968	0.0042	0.0333	0.0074	0.0051	0.0078	34.85
<i>svm</i>	SVM - Linear Kernel	0.9786	0	0	0	0	-0.0017	-0.002	0.072
<i>lda</i>	Linear Discriminant Analysis	0.978	0.5787	0	0	0	-0.0032	-0.0043	0.143
<i>gbc</i>	Gradient Boosting Classifier	0.9778	0.5473	0	0	0	-0.0038	-0.0064	0.551
<i>rf</i>	Random Forest Classifier	0.9775	0.5097	0	0	0	-0.0043	-0.0068	0.275
<i>et</i>	Extra Trees Classifier	0.9764	0.4982	0	0	0	-0.0057	-0.008	0.43
<i>dt</i>	Decision Tree Classifier	0.9563	0.4965	0.0174	0.0159	0.0164	-0.0053	-0.0055	0.039
<i>qda</i>	Quadratic Discriminant Analysis	0.6582	0.4932	0.3214	0.0191	0.0359	-0.0019	-0.0042	0.107
<i>nb</i>	Naive Bayes	0.1577	0.5657	0.8766	0.0205	0.0401	0.0009	0.0084	0.015



Επιπλέον, η χρήση της καμπύλης ανάκλησης ακριβείας (PRC) έχει αποθαρρυνθεί εκτενώς [211], [212] ως μετρική στις διαδικασίες ταξινόμησης σε μη ισορροπημένα σύνολα δεδομένων. Η μετρική F1, η οποία εξαρτάται από τη PRC, απορρίπτεται επίσης έμμεσα από αυτούς τους συγγραφείς. Ωστόσο, είναι η πιο διαδεδομένη μετρική στη βιβλιογραφία. Ωστόσο, κανένα από τα μοντέλα δεν απέδειξε αξιοπρεπή απόδοση στη βαθμολογία F1, όπως απεικονίζεται στον Πίνακα 11. Εν συνεχεία απεικονίσαμε το διάγραμμα τιμών SHAP του μοντέλου Decision Tree, όπως παρουσιάζεται στην Εικόνα 16 για να έχουμε μια αρχική ιδέα των καθοριστικών παραγόντων ταξινόμησης των δεδομένων, όπως π.χ. Ποσοστό άπαχου κρέατος, Βάρος του κάπρου στο σφαγείο, Ηλικία, Μέσος αριθμός με χοιρομητέρες δίπλα σε μάνδρες με κάπρους και η «βαθμολογία» δερματικών αλλοιώσεων στα σφαγεία.



Εικόνα 16 Σημαντικότητα χαρακτηριστικών του μοντέλου Δένδρων αποφάσεων βάσει μεθόδου SHAP για την κηλίδωση κάπρου

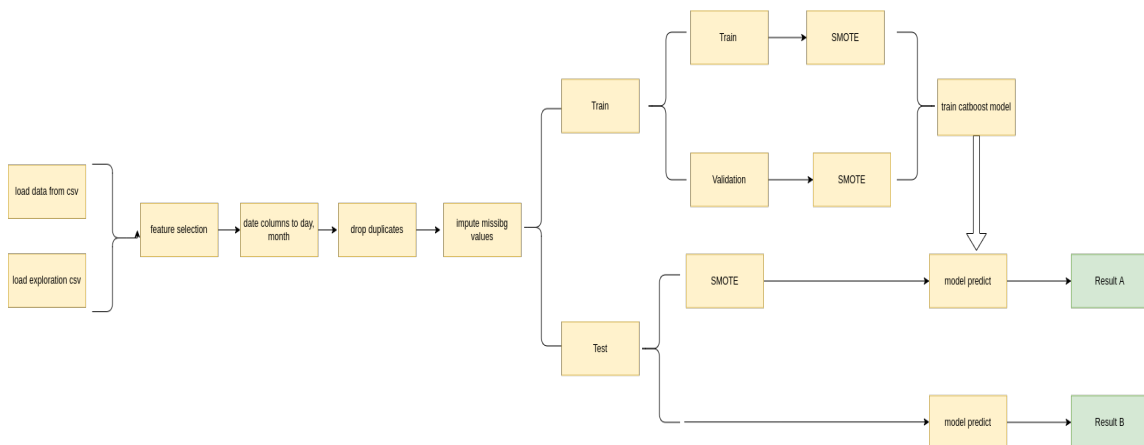
### 3.4.4.2 Περαιτέρω πειράματα και αποτελέσματα

Όπως αναλύθηκε στο εισαγωγικό κεφάλαιο, πολλοί συμβατικοί αλγόριθμοι μηχανικής μάθησης υποφέρουν από αυτήν την άνιση κατανομή των δεδομένων

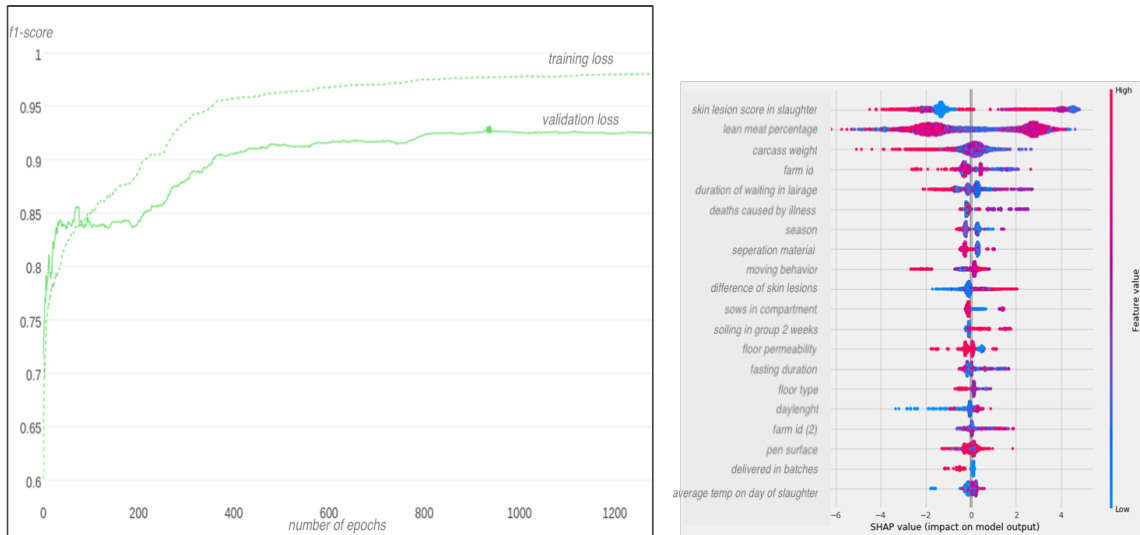


μεταξύ των τάξεων/κλάσεων, οδηγώντας σε λιγότερο αποτελεσματικές προβλέψεις. Στην εξεταζόμενη περίπτωση, προτιμήθηκε το μοντέλο Catboost [213] με βάση τα ευρήματα του [214]. Επίσης, έχοντας κατά νου ότι το Catboost μπορεί να χειριστεί αποτελεσματικά κατηγορικές μεταβλητές σε συνδυασμό με αριθμητικές, επιλέχθηκε ως ο τέλειος υποψήφιος για την συγκεκριμένη εφαρμογή.

Στη συνέχεια, εφαρμόσαμε μια αυτόνομη μέθοδο SMOTE ακολουθώντας το διάγραμμα ροής που απεικονίζεται στην Εικόνα 17, η οποία καταλήγει στην εκπαίδευση ενός μοντέλου Catboost. Τα αποτελέσματα της εν λόγω μεθοδολογίας απεικονίζονται στην Εικόνα 18. Τα αποτελέσματα αναφέρονται στη βαθμολογία F1 του σετ δεδομένων εκπαίδευσης και επικύρωσης ανά επανάληψη διαδικασίας εκπαίδευσης. Επίσης, απεικονίζονται οι τιμές SHAP του εκπαιδευμένου μοντέλου στο σύνολο επικύρωσης. Μπορεί να εξαχθεί ότι το ποσοστό άπαχου κρέατος, το βάρος σφαγίου και η βαθμολογία δερματικών αλλοιώσεων στο σφαγείο είναι οι μεταβλητές με τον υψηλότερο αντίκτυπο.



**Εικόνα 17** Βήματα εκπαίδευσης και επαλήθευσης για την περίπτωση ταξινόμησης των δειγμάτων κρέατος με ή χωρίς κηλίδωση κάρπου, με χρήση μεθόδου SMOTE για υπερδειγματοληψία της μειονοτικής κλάσης



**Εικόνα 18** Σημαντικότητα των χαρακτηριστικών για την εμφάνιση κηλίδωσης κάπρου βάσει εκπαιδευμένου μοντέλου Catboost με υπερδειγματοληψία

Επιπλέον, είναι σημαντικό να αξιολογηθούν τα αποτελέσματα τόσο στο αρχικό σύνολο δοκιμών (πριν την υπερδειγματοληψία) όσο και σε μια έκδοση αυτού με υπερδειγματοληψία, τα οποία απεικονίζονται στους Πίνακας 12 και Πίνακας 13, αντίστοιχα. Στην περίπτωση μας, τα θετικά δείγματα (δηλαδή τα μολυσμένα) είναι πιο σημαντικό να ταξινομηθούν σωστά, επομένως είναι προφανές ότι το μοντέλο φαίνεται να έχει εξαιρετικά καλή απόδοση στο σύνολο δοκιμών υπερδειγματοληψίας (καθώς τα προβλεπόμενα θετικά αποτελέσματα είναι 1398 από 1474 στην πραγματικότητα θετικά). Ωστόσο, από την άλλη πλευρά, το αρχικό σετ δοκιμών με το ίδιο μοντέλο μπορεί να ταξινομήσει σωστά μόνο 1 από τα 30 πραγματικά μολυσμένα δείγματα. Αυτό σημαίνει ότι το μοντέλο μπορεί να ταξινομήσει σωστά τις τεχνητά δημιουργημένες θετικές τιμές, όχι τις πραγματικά μολυσμένες. Από εκεί, μπορεί εύκολα να εξαχθεί ότι η προσέγγιση υπερδειγματοληψίας δεν βοήθησε πραγματικά, καθώς τα περισσότερα από τα αρχικά μολυσμένα δείγματα ταξινομήθηκαν εσφαλμένα.

**Πίνακας 12** Αποτελέσματα ταξινόμησης σε μορφή Confusion Matrix σε δεδομένα ελέγχου στο αρχικό σετ ελέγχου (πριν την υπερδειγματοληψία)

	<i>Predicted Negative</i>	<i>Predicted Positive</i>
<i>Actually Negative</i>	1465	9
<i>Actually Positive</i>	29	1



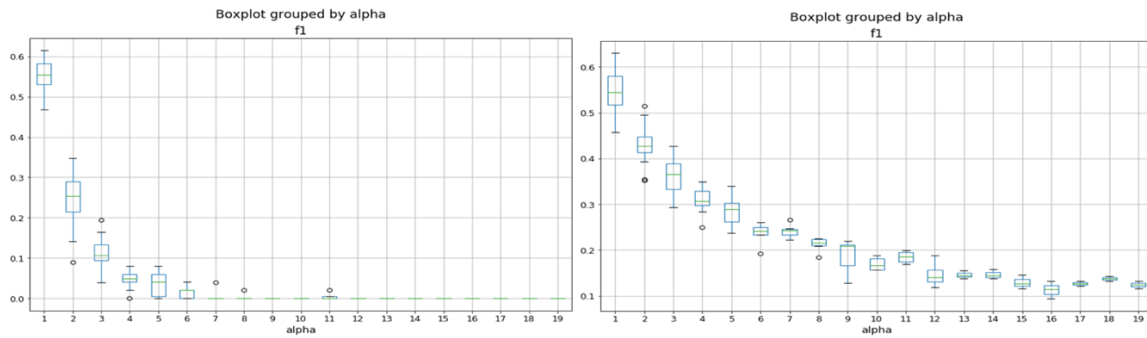
**Πίνακας 13** Αποτελέσματα ταξινόμησης σε μορφή Confusion Matrix σε δεδομένα ελέγχου στο αρχικό σετ ελέγχου (πριν την υπερδעיματοληψία)

	<i>Predicted Negative</i>	<i>Predicted Positive</i>
<i>Actually Negative</i>	1465	9
<i>Actually Positive</i>	76	1398

### 3.4.4.3 Καινοτόμα προσέγγιση υποδειγματοληψίας βασισμένη σε γενετικό αλγόριθμο

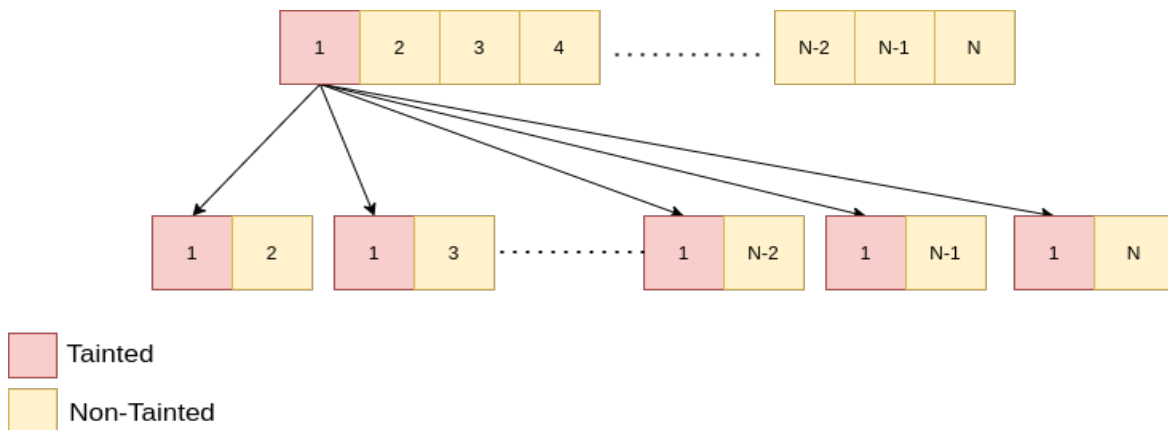
Στη σχετική βιβλιογραφία που σχετίζεται με την επίλυση του προβλήματος υπο-απόδοσης των μοντέλων μηχανικής μάθησης σε μη ισοκαταμεμημένα σύνολα δεδομένων και ακόμη και στο αρχικό έγγραφο SMOTE [26], συνιστάται η υποδειγματοληψία της πλειοψηφικής τάξης αντί της υπερδειγματοληψίας στην κατηγορία μειοψηφίας. Βέβαια, είναι εύκολα κατανοητό ότι σε οποιαδήποτε μορφή υποδειγματοληψίας αφαιρείται ένα τμήμα πληροφορίας από τα δεδομένα που απορρίπτονται. Ως εκ τούτου, στην περίπτωση μας, ακολουθήσαμε ένα υβριδικό μοντέλο δειγματοληψίας με σκοπό την διατήρηση όσο το δυνατόν της πληροφορίας από τα δεδομένα.

Συγκεκριμένα, μετρήσαμε την απόδοση ενός μοντέλου Catboost σχετικά με την αναλογία ανισορροπίας. Για αυτήν την περίπτωση, επιλέχθηκαν τυχαία 19 διαφορετικά υποσύνολα του συνόλου δεδομένων, με παράγοντες ανισορροπίας από 1 έως 1/19 αντίστοιχα. Σημειώνεται ότι, το Catboost παρέχει παραμέτρους που μπορούν να παραμετροποιηθούν, όπως τα *weight-class* και *scale-pos-weight*, οι οποίες είναι αφιερωμένες στην εξομάλυνση των επιπτώσεων της ανισορροπίας. Στη συνέχεια εφαρμόστηκε το ίδιο μοντέλο ώστε να το συγκρίνουμε με το αν χρησιμοποιήθηκαν αυτές οι υπερπαραμέτροι. Τα αποτελέσματα στη βαθμολογία F-1 στο σύνολο δεδομένων δοκιμής απεικονίζονται στα δύο διαγράμματα στην Εικόνα 19. Στη δεξιά γραφική παράσταση, όπου διαμορφώθηκαν οι υπερπαραμέτροι, οι βαθμολογίες F1 αυξάνονται όταν ο λόγος ανισορροπίας γίνεται υψηλότερος σε σύγκριση με τη χρήση μη υπερπαραμέτρων (υψηλότερος λόγος ανισορροπίας σημαίνει ότι οι περισσότερες καθιερωμένες μέθοδοι υποδειγματοληψίας παραλείπουν πολλά δείγματα της πλειοψηφικής κατηγορίας).



**Εικόνα 19** Βoxplot συντελεστή ανισορροπίας ως βάρος κατηγορίας. Και στις δύο γραφικές παραστάσεις απεικονίζεται η κατανομή των αποτελεσμάτων σε διάφορους παράγοντες ανισορροπίας. Στο αριστερό διάγραμμα δεν χρησιμοποιήθηκε βάρος κλάσης ενώ στη δεύτερη γραφική παράσταση το βάρος κλάσης ορίστηκε ίσο με τον 1/παράγοντα-ανισορροπίας

Συνεπώς μια πρόκληση που πρέπει να αντιμετωπιστεί είναι η διατήρηση των πληροφοριών από την πλειοψηφική τάξη στις περιπτώσεις της υποδειγματοληψίας. Στο πλαίσιο της Διατριβής μας, προτείνουμε μια προσαρμοσμένη μέθοδος υποδειγματοληψίας που υλοποιείται ως μια μορφή γενετικού αλγορίθμου για την επιλογή χαρακτηριστικών για τη μόχλευση όσο το δυνατόν περισσότερων πληροφοριών από την πλειοψηφική κατηγορία. Συγκεκριμένα, δημιουργήθηκαν 1/(παράγοντα ανισορροπίας) (δηλ. 50), που αποτελούνται από όλα τα μολυσμένα δείγματα και ένα διαφορετικό υποσύνολο των μη μολυσμένων δειγμάτων, όπως απεικονίζεται στην Εικόνα 20.



**Εικόνα 20** Μέθοδος υποδειγματοληψίας. Κάθε τετράγωνο είναι μια παρτίδα ίσου μεγέθους. Το κόκκινο χρώμα υποδηλώνει τη κηλιδωμένη παρτίδα, το κίτρινο τη μη κηλιδωμένη. Ο αριθμός αντιπροσωπεύει το αναγνωριστικό παρτίδας. Τα «ισορροπημένα» υποσύνολα αποτελούνται επομένως από τη μολυσμένη παρτίδα που συνενώθηκε με μία από τις μη μολυσμένες παρτίδες

Ένας γενετικός αλγόριθμος (GA) είναι μια τεχνική βελτιστοποίησης που βασίζεται στις αρχές της Γενετικής και της Φυσικής Επιλογής. Χρησιμοποιείται για την εύρεση βέλτιστων ή σχεδόν βέλτιστων λύσεων σε χρονοβόρα υπολογιστικά





προβλήματα. Οι γενετικοί αλγόριθμοι αναπτύχθηκαν από τον John Holland και έκτοτε έχουν δοκιμαστεί σε διάφορα προβλήματα βελτιστοποίησης με μεγάλη επιτυχία [215]. Επιπλέον, η ίδια προσέγγιση επαναλήφθηκε πολλές φορές, δημιουργώντας τυχαία υποσύνολα με σκοπό να ελαχιστοποιηθεί η πιθανότητα και τον αντίκτυπο ενός μεροληπτικού αρχικού διαχωρισμού (δηλαδή, ελαχιστοποίηση της πιθανότητας η αρχική τυχαία επιλογή να έχει σημαντικό αντίκτυπο στα αποτελέσματα). Με άλλα λόγια, όσο περισσότερες επαναλήψεις των πειραμάτων, τόσο μεγαλύτερη είναι η γενίκευση των αποτελεσμάτων.

Σε έναν γενετικό αλγόριθμο (GA), ένας πληθυσμός από υποψήφια λύσεις (άτομα) σε ένα πρόβλημα βελτιστοποίησης εξελίσσεται προς ένα πληθυσμό αποτελούμενο από καλύτερες λύσεις (που ελαχιστοποιούν την συνάρτηση κόστους). Κάθε πιθανή λύση έχει ένα σύνολο ιδιοτήτων (τα χρωμοσώματα ή τον γονότυπο του) που υφίσταται διεργασίες που υποκινούνται από τη φυσική γενετική, όπως η επιλογή, η διασταύρωση και η μετάλλαξη. Σε κάθε άτομο (ή υποψήφια λύση) εκχωρείται μια τιμή καταλληλότητας (με βάση την τιμή της αντικειμενικής συνάρτησης κόστους) και αυτοί που έχουν τις καλύτερες βαθμολογίες φυσικής κατάστασης είναι πιο πιθανό να συμβάλουν στην παραγωγή της επόμενης γενιάς. Με αυτόν τον τρόπο, συνεχίζει ο πληθυσμός να «εξελίσσεται» προς καλύτερα άτομα ή λύσεις με την πάροδο των γενεών μέχρι να φτάσουμε σε ένα κριτήριο αναστολής/διακοπής. Κατά την εφαρμογή του GA, είναι πρακτικό να μοντελοποιείται κάθε λύση σε μια δυαδική αναπαράσταση. Στο πλαίσιο της έρευνάς μας, ένας δυαδικός πίνακας που αντιπροσωπεύει τα χαρακτηριστικά που θα επιλεγούν θεωρείται ως λύση του προβλήματος που βελτιστοποιούμε. Για να γίνουμε πιο συγκεκριμένοι, κάθε λύση  $I$  είναι ένας δυαδικός διάνυσμα, όπου για  $i$  στην περιοχή  $(0, \text{μήκος χαρακτηριστικών})$ , υπολογίζεται ως εξής:

$$I_i = \begin{cases} 1, & \text{if feature } i \text{ is to be selected} \\ 0, & \text{otherwise} \end{cases}$$

Σε κάθε επανάληψη, για κάθε λύση, τα υποσύνολα και ο τρόπος με τον οποίο αναδιατάχθηκαν τα δεδομένα ήταν τα ίδια, καθώς, σε κάθε άλλη περίπτωση, τα αποτελέσματα δεν θα είναι συνεπή. Στη συνέχεια, ένα μοντέλο Catboost εκπαιδεύτηκε για κάθε υποσύνολο χρησιμοποιώντας μόνο τα «επιλεγμένα» χαρακτηριστικά κάθε υποσυνόλου. Ως βαθμολογία φυσικής κατάστασης, χρησιμοποιήθηκε η βαθμολογία F-1 των προβλέψεων κάθε μοντέλου έναντι των πραγματικών τιμών του συνόλου δοκιμής. Η όλη διαδικασία περιγράφεται στο περιεχόμενο του παρακάτω Αλγόριθμου. Επιπλέον, όλες οι παράμετροι του GA και οι υπερπαράμετροι του μοντέλου Catboost αντιμετωπίστηκαν ως είσοδοι του αλγορίθμου.

**Algorithm 1: GENETIC ALGORITHM** for feature selection

```
Input: A tabular dataset data, a classifier, the features.array
(i.e. discrete sequence to a sequence of ones and zeros)
Output: A vector of the indexes of the selected features.
1 for  $n \leftarrow \text{range}(\text{number\_of\_permutations})$  do
2   Split the data into balanced subsets
3   for  $\text{subset} = 1, \text{subsets}$  do
4     Initialize the Population P with size  $P_{\text{length}}$ 
5     for  $i \leftarrow \text{range}(P_{\text{length}})$  do
6       Initialize an specific selection of features as Individual I
7        $P_i = I$ 
8     end
9     Evaluate the fitness score  $P_{\text{score}}$  of the P
10    for  $i \leftarrow 0, \text{range}(L)$  do
11      Select the features of the dataset based on the I
12       $\text{data} = \text{data}[\text{features} \times I]$ 
13      Split the dataset in train, validation, test
14       $\text{model.fit}(\text{train}, \text{validation})$ 
15       $\text{prediction} = \text{model.predict}(\text{test})$ 
16       $I_{\text{score}} = F1(\text{test}, \text{prediction})$ 
17    end
18     $P_{\text{score}} = \text{average}([I_{\text{score}}])$ 
19    if  $i == 0 \vee P_{\text{score}} > P_{\text{score.best}}$  then
20      Set best score  $P_{\text{score.best}}$  and best Population  $P_{\text{best}}$ 
21       $P_{\text{score.best}} \leftarrow P_{\text{score}}$ 
22       $P_{\text{best}} \leftarrow P$ 
23       $\text{no\_improve} = 0$ 
24    end
25    Check for termination criterion
26  end
27 end
1 if  $\text{no\_improve} < \text{termination\_criterion}$  then
2   Selection
3   Select top k best Individuals I of the P with probability
    $I_{\text{score}} / \text{sum}(I_{\text{scores}})$ 
4   Crossover
5   for  $k \in P_{\text{length}}/2$  do
6      $\text{cross\_index} = \text{randint}(I_{\text{length}})$ 
7      $P[k + \text{population}_{\text{length}}/2] \leftarrow P[k]:$ 
    $\text{cross\_index}] + P[k + 1][\text{cross\_index} :]$ 
8      $P[k + \text{population}_{\text{length}}/2 + 1] \leftarrow P[k][\text{cross\_index} :]$ 
    $+ P[k + 1]:$ 
    $\text{cross\_index}]$ 
9   Mutation
10  for  $m \in P_{\text{length}}$  do
11     $\text{flag}_{\text{mutation}} = \text{rand}()$ 
12    if  $\text{mutation}_{\text{flag}} < \text{mutation}_{\text{threshold}}$  then
13       $\text{mutation}_{\text{index}} = \text{randint}(I_{\text{length}})$ 
14       $P[m][\text{mutation}_{\text{index}}] = 1 - P[m][\text{mutation}_{\text{index}}]$ 
15  else
16    return  $I_{\text{best}}, I_{\text{score.best}}$ 
17 return  $I_{\text{best}}, I_{\text{score.best}}$ 
```

### 3.5 Μηχανική Μάθηση & τραπεζικός βοηθός για μικρομεσαίες επιχειρήσεις.



Οι κύριες ερευνητικές προκλήσεις που αντιμετωπίστηκαν σε αυτή την υλοποίηση επιτρέπουν την ταξινόμηση των συναλλαγών των ΜΜΕ και τη δημιουργία προβλέψεων ταμειακών ροών ανά κατηγορία εξόδων. Οι δύο αυτές εργασίες αποτελούν ερευνητικό θέμα της ανά χείρας Διατριβής και είναι αλληλένδετες καθώς το μοντέλο πρόβλεψης βασίζεται στα αποτελέσματα της εργασίας κατηγοριοποίησης/ταξινόμησης. Ωστόσο, αυτές οι εργασίες θέτουν διαφορετικές προκλήσεις και μεθοδολογικές προσεγγίσεις, τόσο όσον αφορά τη συγκέντρωση δεδομένων και την προ-επεξεργασία, όσο και από την άποψη της χρήσης και επέκτασης των διαδοσμένων τεχνικών μηχανικής μάθησης. Προκειμένου να αντιμετωπιστούν οι διάφορες προκλήσεις και τα ερευνητικά επιτεύγματα, είναι βασικό να παρουσιαστούν τα δεδομένα που χρησιμοποιήθηκαν για αυτήν την έρευνα, τα οποία ελήφθησαν από την Τράπεζα Κύπρου στο πλαίσιο ενός ευρωπαϊκού ερευνητικού προγράμματος για την υλοποίηση ενός πραγματικού σεναρίου.

Στο πλαίσιο αυτό, η προτεινόμενη προσέγγιση επιτρέπει την ανάπτυξη εξατομικευμένων υπηρεσιών προστιθέμενης αξίας για τις ΜΜΕ. Συγκεκριμένα εφαρμόστηκε ένα μοντέλο κατηγοριοποίησης συναλλαγών που ταξινομεί τις συναλλαγές σύμφωνα με την υποκείμενη χρησιμότητα της συναλλαγής και μια προσέγγιση πρόβλεψης εισροών και εκροών με βάση ιστορικά δεδομένα για κάθε κατηγορία συναλλαγών. Ως εκ τούτου, απαιτείται η κατηγοριοποίηση όλων των συναλλαγών ΜΜΕ προκειμένου να επισημανθούν όλα τα ιστορικά δεδομένα και να ξεκλειδωθούν δυνατότητες μιας σύγχρονης εργαλειοθήκης οικονομικής διαχείρισης των ΜΜΕ.

Όσον αφορά το μοντέλο κατηγοριοποίησης συναλλαγών, προτείνεται μια υβριδική προσέγγιση, με την εφαρμογή μιας προσέγγισης σταδίων και κανόνων, και την εφαρμογή ενός αλγόριθμου μηχανικής μάθησης που βασίζεται στην αρχιτεκτονική των δέντρων αποφάσεων για τις μη κατηγοριοποιημένες συναλλαγές. Με βάση τα παραπάνω, ο αλγόριθμος μηχανικής μάθησης του μοντέλου Κατηγοριοποίησης Συναλλαγών βασίζεται σε ένα μοντέλο Catboost, με στόχο την παραγωγή ισχυρών αποτελεσμάτων. Τα δέντρα απόφασης είναι ιδανικά για περιπτώσεις όπου υπάρχουν αριθμητικά χαρακτηριστικά, όμως σε σενάρια πραγματικών περιπτώσεων, τα περισσότερα σύνολα δεδομένων περιλαμβάνουν κατηγορικά χαρακτηριστικά, τα οποία είναι επίσης σημαντικά για τις προβλέψεις. Για την αντιμετώπιση κατηγορικών χαρακτηριστικών, η ενίσχυση κλίσης είναι η μετατροπή τους σε αριθμούς. Το Catboost αξιοποιεί έναν νέο αλγόριθμο gradient boosting που χειρίζεται με επιτυχία κατηγορηματικά χαρακτηριστικά και ωφελείται από την αντιμετώπισή τους κατά τη διάρκεια της εκπαίδευσης, ως εκ τούτου κρίθηκε η καταλληλότερη λύση στην εξεταζόμενη περίπτωση.

### 3.5.1 Σύνολο δεδομένων



Δεδομένου ότι δεν υπάρχει σύνολο δεδομένων ανοιχτού κώδικα που να εκτελεί την διαδικασία ταξινόμησης συναλλαγών MME, στο πλαίσιο της έρευνάς μας, το χρησιμοποιούμενο σύνολο δεδομένων είχε παρασχεθεί από την Τράπεζα Κύπρου που αποτελείται από χαρακτηριστικές συναλλαγές πελατών της τράπεζας (MME). Τα παρεχόμενα δεδομένα παρουσιάστηκαν σε μορφή πίνακα και ήταν διαθέσιμα στο πλαίσιο του INFINITECH H2020 European Project (GA 856632). Αυτά τα δεδομένα περιλαμβάνουν δεδομένα συναλλαγών, πελατών και λογαριασμών για περισσότερες από χίλιες MME για τα έτη 2017 έως 2020, που υπερβαίνουν τις 3,5 εκατομμύρια καταχωρίσεις δεδομένων. Καθώς πολλές μεταβλητές συμπεριλήφθηκαν στα σύνολα δεδομένων που έλαβε η τράπεζα, οι κύριες κατηγορίες και οι μεταβλητές / χαρακτηριστικά που χρησιμοποιήθηκαν στην προσέγγισή μας παρουσιάζονται στον Πίνακα 14, ενώ οι κατηγορίες των ταξινομημένων συναλλαγών απεικονίζονται στον Πίνακα 15:

**Πίνακας 14** Πηγές δεδομένων που διατέθηκαν από την Τράπεζα Κύπρου

Datasets	Utilized variables / features
<b>Transaction Data</b>	Account Holder Key Identifier, Transaction Code, Transaction Date, Original Amount, Currency Code, Amount, Transaction Description, Transfer Account Key, Channel Code', Merchant Name, Merchant Code', Debit/Credit Indicator, Merchant Category Code ID
<b>Customer Data</b>	Card Holder Key Identifier, Customer NACE code (rev2), Number of Employees, District
<b>SME Accounts Data</b>	Account and Card Keys identifiers
<b>NACE Rev. 2</b>	Open-source NACE codes

**Πίνακας 15** Κύριες κατηγορίες συναλλαγών για MME (έξοδος του μοντέλου ταξινόμησης συναλλαγών)

Uncategorized Expense	Utilities	Professional Fees	Cash
Selling and distribution	Offices Expenses	Banking Expense	Saving
Taxes	Rent & Lease	Repairs & Maintenance	Transfer
Insurance	Motor Vehicles Running Cost	Sundry expenses	Supplier
Payroll & Benefits in Kind	Subscriptions Donations	Income	Assets

Αξίζει να σημειωθεί ότι όλα τα δεδομένα των MME είχαν υποστεί επεξεργασία ώστε να παραμείνουν ανώνυμα πριν από την απόκτηση τους.

Η βασική πηγή δεδομένων, τόσο για την κατηγοριοποίηση των συναλλαγών όσο και για το μοντέλο πρόβλεψης ταμειακών ροών, ήταν τα διαθέσιμα δεδομένα συναλλαγών που περιείχαν περισσότερες από 40 μεταβλητές/χαρακτηριστικά που υποδεικνύουν από βασικές πληροφορίες όπως ημερομηνία, ποσό συναλλαγής, δείκτες πίστωσης/χρέωσης και περιγραφή της συναλλαγής, έως πιο περίπλοκες πληροφορίες όπως δείκτες πληρωμής χωρίς αφής, κωδικοί κατηγορίας εμπόρου, δείκτες πάγιας εντολής κ.λπ.

Το σύνολο δεδομένων λογαριασμών MME περιείχε κωδικοποιημένες πληροφορίες των μεμονωμένων λογαριασμών των πελατών, τα διαθέσιμα υπόλοιπα και τον αντίστοιχο κωδικό NACE rev2, μια στατιστική ταξινόμηση των



οικονομικών δραστηριοτήτων που χρησιμοποιείται σε ολόκληρη την Ευρωπαϊκή Κοινότητα [216] που χρησιμοποιείται επίσης από την τράπεζα για την κατάταξη των ΜΜΕ ανάλογα με τον τομέα λειτουργείας τους. Η διαθεσιμότητα της οικονομικής δραστηριότητας των ΜΜΕ είναι σημαντική καθώς εξωτερικά εθνικά και ευρωπαϊκά δεδομένα που χρησιμοποιούν οριζόντια το σύστημα κωδικών NACE μπορούν να χρησιμοποιηθούν στα μοντέλα μας για την παροχή πιο εξατομικευμένων πληροφοριών ανά τομέα.

Τα δεδομένα πελατών περιέχουν τους διάφορους λογαριασμούς που ανήκουν σε κάθε μεμονωμένη ΜΜΕ, καθώς και τον τύπο του αντίστοιχου λογαριασμού.

Τα μετα-δεδομένα κάθε συναλλαγής περιλαμβάνουν τη χρονική σήμανση κάθε συναλλαγής, σε αυτήν την προσέγγιση δημιουργήθηκε ένα άλλο σύνολο δεδομένων μεγέθους 9494 χρονοσειρών, που αντιπροσωπεύει τον ημερήσιο αριθμό ποσού που δαπανήθηκε για κάθε λογαριασμό ανά κατηγορία για τη δεδομένη χρονική περίοδο. Εκτός από αυτές τις χρονοσειρές που αντιπροσωπεύουν τα ημερήσια έξοδα, δημιουργούμε επίσης ορισμένες συμμεταβλητές (για παράδειγμα Ημέρα της εβδομάδας, Μήνας του έτους, κ.λπ.) για μεμονωμένες χρονοσειρές.

### 3.5.2 Υποκατάστατα Δεδομένα

Η χρήση υποκατάστατων/ανάλογων δεδομένων εφαρμόστηκε σε και σε αυτή την υλοποίηση για να αυξήσει την προστιθέμενη αξία και τον αντίκτυπο των χρησιμοποιούμενων μοντέλων βαθιάς μάθησης, καθώς επωφελούνται από την αυξημένη ποικιλομορφία και όγκο των δεδομένων εκπαίδευσης. Αναμένεται ότι η χρήση ανάλογων δεδομένων θα βοηθήσει στη γενίκευση ορισμένων, αν όχι όλων, των μοντέλων βαθιάς μάθησης στην περίπτωση της πρόβλεψης χρονοσειρών.

Οι πιο συχνά χρησιμοποιούμενες τεχνικές για τη δημιουργία υποκατάστατων δεδομένων για στατιστική ανάλυση μη γραμμικών διεργασιών περιλαμβάνουν τυχαία ανακατανομή της αρχικής χρονοσειράς, υποκατάστατα μετασχηματισμένα με Fourier, μετασχηματισμένο Fourier με προσαρμογή πλάτους (AAFT υποκατάστατα) και επαναλαμβανόμενα υποκατάστατα AAFT (IAAFT). Στην εργασία μας, ενσωματώσαμε τη μέθοδο IAAFT για να αντιμετωπίσουμε το ζήτημα της λεύκανσης του φάσματος ισχύος, ως το κύριο μειονέκτημα της μεθόδου AAFT, εκτελώντας μια σειρά επαναλήψεων στις οποίες το φάσμα ισχύος ενός υποκατάστατου AAFT προσαρμόζεται ξανά σε αυτό των αρχικών δεδομένων προτού η κατανομή επανακλιμακωθεί σε αυτή των αρχικών δεδομένων. Όπως σημειώνεται από το [217], μόνο λίγες μελέτες διερευνούν την αύξηση δεδομένων από την προοπτική του τομέα συχνότητας για χρονοσειρές. Οι χρονοσειρές που δημιουργούνται από τις AAFT και IAAFT μπορούν να διατηρήσουν κατά προσέγγιση τη χρονική συσχέτιση, τα φάσματα ισχύος και την κατανομή πλάτους της αρχικής χρονοσειράς. Αυτό καθιστά αυτές τις προσεγγίσεις κατάλληλες για την



αύξηση του μεγέθους των χρονοσειρών. Στα πειράματα του [218], οι συγγραφείς διεξήγαγαν δύο τύπους αύξησης δεδομένων επεκτείνοντας τα δεδομένα κατά 10 και 100 φορές μέσω των μεθόδων AAFT και IAAFT και επέδειξαν πολλά υποσχόμενες βελτιώσεις στην ακρίβεια ταξινόμησης σε σύγκριση με την αρχική χρονοσειρά χωρίς αύξηση δεδομένων.

Μέσω αυτής της προτεινόμενης προσέγγισης, με κάθε επανάληψη η αλλαγή στην κατανομή που συμβαίνει όταν προσαρμόζονται τα πλάτη Fourier θα είναι μικρότερη από την προηγούμενη επανάληψη, και επομένως η μεταβολή του φάσματος ισχύος όταν εκτελείται η επανακλιμάκωση θα είναι επίσης μικρότερη από την προηγούμενη επανάληψη. Στην πραγματικότητα, ο Schreiber [219] έδειξε ότι, για μια μη γραμμικά μετασχηματισμένη διαδικασία αυτόματης παλινδρόμησης, η διαδικασία επανάληψης θα συγκλίνει προς το φάσμα ισχύος των αρχικών δεδομένων έως ότου επιτευχθεί ένα σημείο κορεσμού, όπου η προσαρμογή του πλάτους Fourier είναι τόσο μικρή που η επανακλιμάκωση θέτει δεδομένα με την ακριβή σειρά που είχε πριν από τη ρύθμιση του πλάτους.

### 3.5.3 Προ-επεξεργασία

Η στασιμότητα των χρονοσειρών ελέγχθηκε με τη χρήση δοκιμής ADF. Επιπλέον, θα πρέπει να σημειωθεί ότι όσον αφορά τις προβλέψεις, τα χρονικά πλαίσια πρόβλεψης των ανακλήσεων μπορεί να είναι πάνω από το 1/4 του έτους (δηλαδή 3 μήνες).

Αυτές οι προβλέψεις στοχεύουν στη βελτίωση της εσωτερικής χρηματοοικονομικής παρακολούθησης μιας επιχείρησης και των διαφόρων εισροών και εκροών της, ενώ παράλληλα επιτρέπουν στους επαγγελματίες να διασφαλίζουν την επιχειρηματική συνέχεια των ΜΜΕ πελατών τους, να ελαχιστοποιούν τους μελλοντικούς κινδύνους και τις οικονομικές απώλειες. Με βάση το τελευταίο, τρεις επιλογές ξεχωρίζουν:

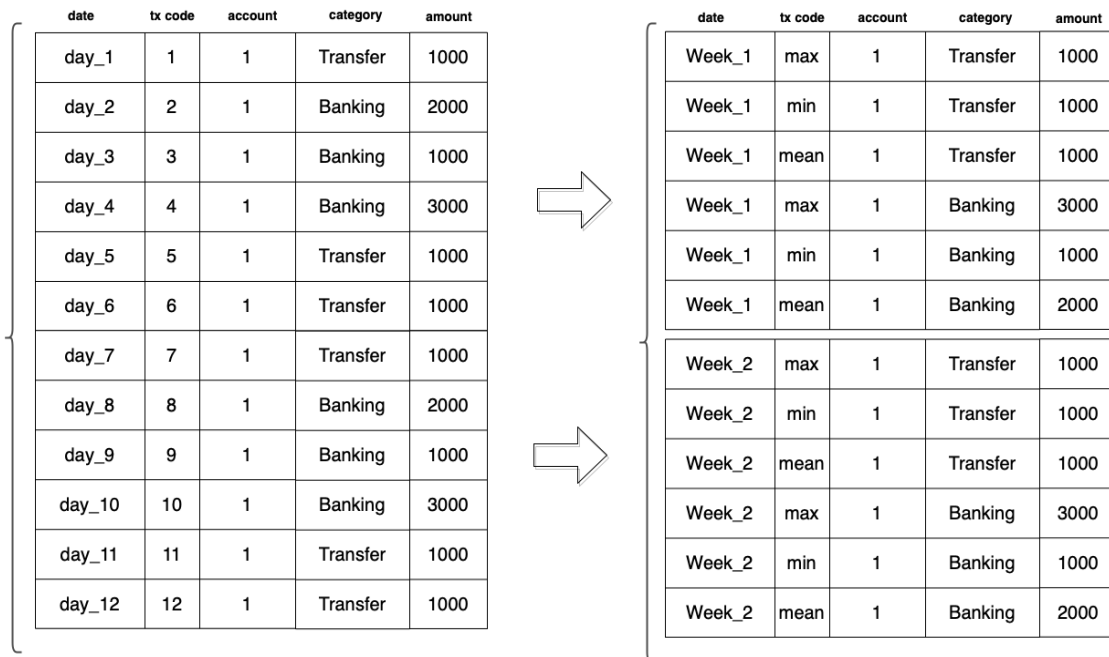
α. Το σύνολο δεδομένων ως έχει, με την ημερήσια συχνότητά του να οδηγεί σε ένα παράθυρο πρόβλεψης 90 χρονικών βημάτων. Αυτό δεν συνιστάται, καθώς οι περισσότερες από τις χρονοσειρές που παράγονται είναι αραιές, χωρίς προφανές μοτίβο να διδαχθεί το μοντέλο βαθιάς μάθησης.

β. Δειγματοληψία των δεδομένων σε χρονικό ορίζοντα εβδομάδας, με αποτέλεσμα ένα παράθυρο πρόβλεψης 16 βημάτων. Ενώ ορισμένα μοτίβα αρχίζουν να εμφανίζονται, τα δεδομένα σε ορισμένες περιπτώσεις εξακολουθούν να είναι πολύ αραιά και το σωρευτικό σφάλμα από την πρόβλεψη των 16 βημάτων είναι θεωρητικά σχετικά μεγάλο.



γ. Δειγματοληψία των δεδομένων σε περιόδους μηνών, παρέχοντας σταθερές χρονοσειρές με ορατά μοτίβα και χαμηλότερο θεωρητικό συσσωρευμένο σφάλμα πρόβλεψης, καθώς το παράθυρο έχει μειωθεί σε 3-χρονικά βήματα.

Τα δεδομένα επανα-δειγματίστηκαν σε μηνιαίο χρονικό πλαίσιο υπολογίζοντας το συνολικό ποσό των εισροών και εκροών αντίστοιχα ανά λογαριασμό, όπως απεικονίζεται στην Εικόνα 21. Το χρονικό πλαίσιο των «μηνών» επιλέχθηκε καθώς αντιστοιχεί στην ταχύτητα δεδομένων για τη συγκεκριμένη περίπτωση - δηλαδή η στρατηγική των MME μπορεί να προβληθεί σε ορίζοντα 4 μηνών.



Εικόνα 21 Παράδειγμα δειγματοληψίας σε χρονικό βήμα εβδομάδας

### 3.5.4 Μοντέλα Μηχανικής Μάθησης – Στατιστικά Μοντέλα

Η κύρια συμβολή της διατριβής αναφορικά με την συγκεκριμένη εφαρμογή είναι ένα πλήρες πλαίσιο που αξιοποιεί προσεγγίσεις μηχανικής μάθησης για τη χρηματοοικονομική διαχείριση των MME. Το πλαίσιο αποτελείται από δύο κύρια στοιχεία, έναν μηχανισμό ταξινόμησης που βασίζεται σε μια τεχνική ενίσχυσης βασισμένο σε αταξινόμητες δέντρων αποφάσεων, που ταξινομεί τις συναλλαγές σύμφωνα με το εμπλεκόμενο προϊόν/έμπορο και μια μέθοδο πρόβλεψης χρονοσειρών βασισμένο στις ιστορικές τιμές των χρονοσειρών όπως ορίζονται από τις ταξινομημένες συναλλαγές, προβλέποντας μελλοντικές ταμειακές ροές. Η κύρια διαφορά της προτεινόμενης εργασίας σε σύγκριση με τις υπάρχουσες, είναι η εισαγωγή μιας υβριδικής προσέγγισης (δηλαδή βασισμένη σε κανόνες και σε



μηχανική μάθηση) και την εν συνεχεία μοντελοποίηση του προβλήματος ως πρόβλεψη χρονοσειρών.

Πιο συγκεκριμένα, η προσέγγιση περιλαμβάνει:

α. την ανάπτυξη ενός υβριδικού μοντέλου κατηγοριοποίησης συναλλαγών, που προσφέρει συνεχή βελτιστοποίηση και προσαρμοστικότητα, μέσω του βρόχου ανάδρασης.

β. την επιλογή μοντέλου πρόβλεψης υψηλής ακρίβειας για σύνολα δεδομένων χρονοσειρών «μικρού» μεγέθους, που προσφέρει ποιοτικές προβλέψεις αποτελεί σημαντική πρόκληση στη μηχανική μάθηση,

γ. τη χρήση υποκατάστατων δεδομένων ως μια βελτιωμένη προσέγγιση που παρέχει πιο ακριβή αποτελέσματα.

Δεδομένου ότι δεν διατίθεται μοντέλο ανοιχτού κώδικα ή δεδομένα διαθέσιμα για τη συγκεκριμένη περίπτωση κατηγοριοποίησης συναλλαγών MME, τα δεδομένα που διατέθηκαν κατόπιν διαδικασίας που τα καθιστά ανώνυμα, επισημάνθηκαν από human-experts με βάση διάφορους κανόνες, ενσωματώνοντας διάφορες εσωτερικές κατηγορικές τιμές που υπάρχουν στο σύνολο δεδομένων. Κατά συνέπεια, το σύνολο δεδομένων που προέκυψε, που αποτελείται από 20 «Κύριες Κατηγορίες» ειδικά σχεδιασμένες για τις MME, ήταν εξαιρετικά ανισόρροπο όπως στα περισσότερα προβλήματα ταξινόμησης πολλαπλών κλάσεων όπου ορισμένες κλάσεις έχουν σημαντικά μικρότερες παρουσίες σε σύγκριση με άλλες κλάσεις. Πολλοί συμβατικοί αλγόριθμοι μηχανικής μάθησης υποφέρουν από αυτήν την άνιση κατανομή των παρουσιών δεδομένων μεταξύ των κλάσεων και τον καθιστούν λιγότερο αποτελεσματικό στην πρόβλεψη περιπτώσεων μειοψηφίας. Αν και αυτοί οι τύποι προβλημάτων ανήκουν σε μια ενεργό ερευνητική περιοχή, στην περίπτωση μας προτιμήθηκε το μοντέλο Catboost με βάση τα ευρήματα του [214], όπου τα πιο καθιερωμένα μοντέλα ενίσχυσης εξετάστηκαν σε σύνολα δεδομένων πολλαπλών κλάσεων, καταλήγοντας στο συμπέρασμα ότι το CatBoost αποδίδει καλύτερα από άλλους αλγόριθμους ενίσχυσης σε συμβατικά σύνολα δεδομένων πολλαπλών κλάσεων.

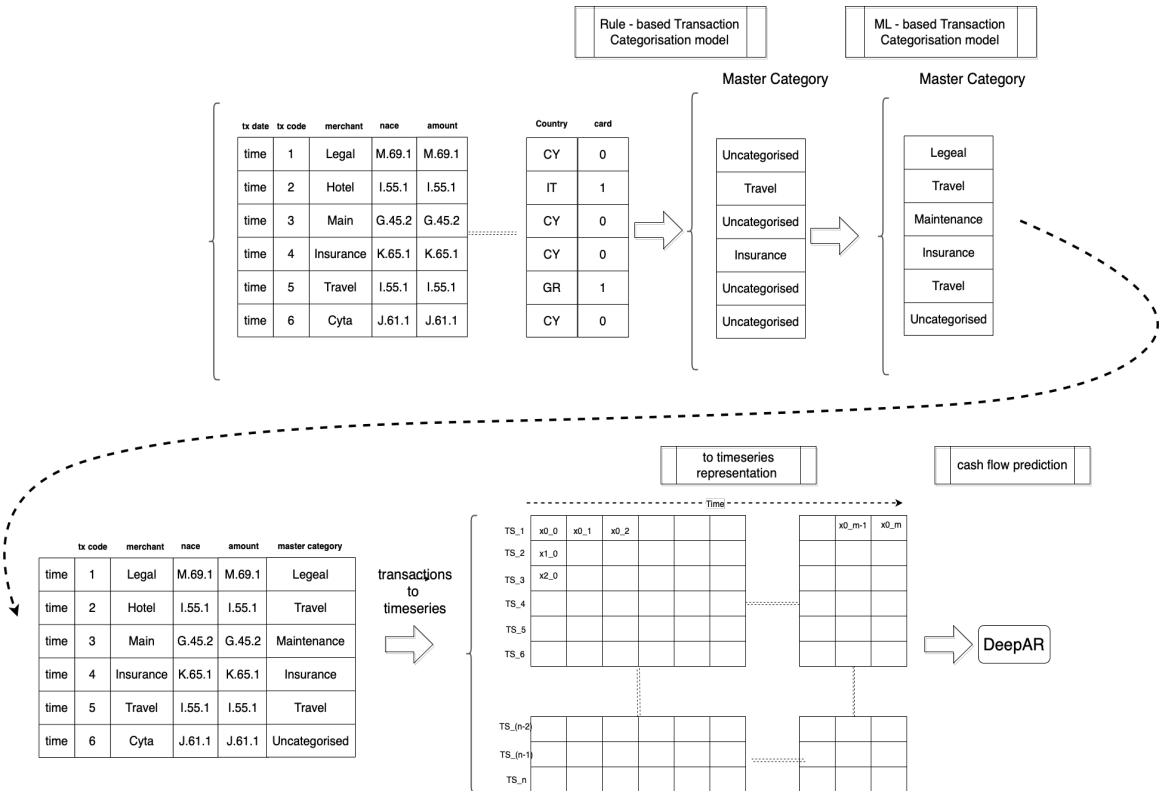
Τέλος, η εργασία πρόβλεψης ταμειακών ροών, στοχεύει στην ανάπτυξη ενός καινοτόμου και ταυτόχρονα επεκτάσιμου μοντέλου χρονοσειρών που μπορεί να προβλέψει τις οικονομικές εισροές και εκροές των MME ανά κατηγορία συναλλαγών με υπολογιζόμενη ακρίβεια. Στην περίπτωση αυτή, χρησιμοποιήθηκαν τεχνικές βαθιάς μάθησης για την παροχή πληροφοριών σχετικά με τις μελλοντικές ταμειακές ροές των MME πλαίσιο, ως μέρος ενός ολιστικού συστήματος BFM. Στα πλαίσια της Διατριβής προτείνουμε μια προσέγγιση, βασισμένη σε μεθόδους βαθιάς μάθησης, που θεωρεί ότι οι συναλλαγές των MME μετατρέπονται σε χρονοσειρές και μπορεί να προβλέψει την «συμπεριφορά» αυτών με πολύ σύντομο χρόνο εκπαίδευσης.





Οι τεχνικές επικεντρώνονται κυρίως στην πρόβλεψη κατανομής χρονοσειρών, χρησιμοποιώντας ένα μοντέλο επαναλαμβανόμενου νευρωνικού δικτύου το οποίο ακολουθείται από Monte Carlo simulation παράγοντας την κατανομή. Πιο συγκεκριμένα, για το σκοπό αυτό, αξιοποιήθηκε το μοντέλο DeepAR. Ένας εκτιμητής DeepAR υλοποιεί ένα μοντέλο που βασίζεται σε επαναλαμβανόμενο νευρωνικό δίκτυο (RNN). Πιο συγκεκριμένα, ο Εκτιμητής DeepAR εισάγει μια μεθοδολογία για την παραγωγή ακριβών πιθανοτικών προβλέψεων, βασισμένη στην εκπαίδευση ενός αυτοπαλινδρομικού μοντέλου RNN σε πολλές συσχετισμένες χρονοσειρές για την πρόβλεψη της μέσης τιμής και της τυπικής απόκλισης αυτών, ενώ μια προσομοίωση Monte Carlo παρέχει την κατανομή πιθανότητας της πρόβλεψης. Όπως περιεγράφηκε στο προηγούμενο κεφάλαιο, ένα RNN είναι ένα νευρωνικό δίκτυο τροφοδοσίας προς τα εμπρός που έχει εσωτερική μνήμη, με το "Recurrence" να εξηγεί το γεγονός ότι η παραγόμενη έξοδος χρησιμοποιείται από το RNN ως πρόσθετη είσοδος. Για τη λήψη μιας απόφασης σχετικά με κάθε έξοδο κάθε επιπέδου, λαμβάνει υπόψη την τρέχουσα είσοδο και την έξοδο που έχει μάθει από την προηγούμενη είσοδο. Σχετικά με αυτό το έργο, έπρεπε να αντιμετωπιστούν πολλές προκλήσεις, με τη μεγαλύτερη επιστημονική επίδραση αυτών, να έχει η χρήση συνθετικά παραγόμενων υποκατάστατων δεδομένων ως ένας τρόπος εμπλουτισμού του αρχικού συνόλου δεδομένων για τη βελτίωση της απόδοσης των υποκείμενων μοντέλων βαθιάς μάθησης.

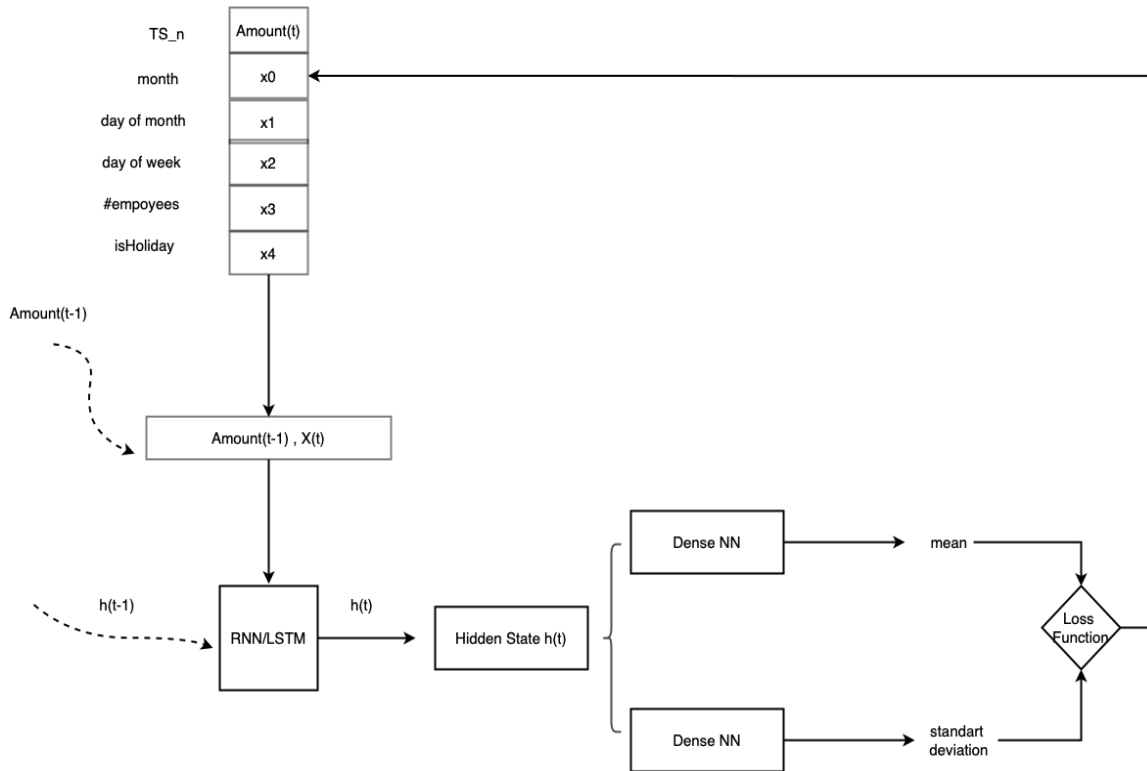
Η εργασία κατηγοριοποίησης συναλλαγών μπορεί να μεταφραστεί ως μια εργασία ταξινόμησης πολύ ανισόροπων δεδομένων πολλαπλών κατηγοριών, όπου ο ταξινομητής Catboost που χρησιμοποιείται φαίνεται να υπερτερεί σε σύγκριση με άλλες προσεγγίσεις, όπως υποστηρίζεται στο [214]. Ταυτόχρονα, δεδομένου ότι η εξεταζόμενη περίπτωση χρονοσειρών είναι μια εργασία πρόβλεψης πολλαπλών βημάτων, θα πρέπει να αναφερθεί ότι για την πρόβλεψη χρονοσειρών, το σφάλμα των προβλέψεων μεγαλώνει αναλογικά όσο αυτός ο ορίζοντας διευρύνεται [85]. Για το σκοπό αυτό, προτιμάται μια πιθανολογική προσέγγιση, η οποία παρέχει μια προβλεπόμενη κατανομή πιθανοτήτων συνοδευόμενη από διαφορετικά επίπεδα εμπιστοσύνης, με σκοπό να ποσοτικοποιηθεί και η αβεβαιότητα του μοντέλου.



Εικόνα 22 Σχηματική εξήγηση της πτυχής δεδομένων στην προσέγγιση από άκρο σε άκρο συμπεριλαμβανομένων των τεσσάρων βασικών βημάτων στη διαδικασία

Μια πιο λεπτομερής άποψη της προτεινόμενης προσέγγισης που παρέχεται στην Εικόνα 22, η οποία απεικονίζει την ροή διαχείρισης δεδομένων, σε μια προσέγγιση διακριτών «βημάτων», για την εφαρμογή τόσο της διαδικασίας κατηγοριοποίησης συναλλαγών όσο και του μοντέλου πρόβλεψης ταμειακών ροών. Στο πρώτο βήμα, τα ακατέργαστα δεδομένα υποβάλλονται σε προ-επεξεργασία και τροφοδοτούνται στο μοντέλο κατηγοριοποίησης του μοντέλου που βασίζεται σε κανόνες. Στη συνέχεια, οι συναλλαγές «μη κατηγοριοποιημένες» (που δεν επισημάνθηκαν από το μοντέλο κανόνων) ταξινομούνται σε συγκεκριμένες κατηγορίες και τελικά τα δεδομένα μετασχηματίζονται σε αναπαραστάσεις χρονοσειρών για να χρησιμοποιηθούν για την πρόβλεψη μελλοντικών εισροών και εκροών.

Επιπλέον, το μοντέλο DeepAR μαθαίνει εποχιακά μοτίβα συμπεριφοράς από δεδομένες συμμεταβλητές, γεγονός που ενισχύει τις δυνατότητες πρόβλεψής του, με την εσωτερική επεξεργασία που απεικονίζεται στην Εικόνα 23.



**Εικόνα 23** Διάγραμμα ροής προβολής ταμειακών ρών εσωτερικών διεργασιών

Επιπλέον, έχουμε συμπεριλάβει τον κάτωθι αλγόριθμο για την καλύτερη απεικόνιση της διαδικασίας κατηγοριοποίησης συναλλαγών.



---

**Algorithm 1:** HYBRID MODEL transaction categorization

---

```
1 Initialize batch of transactions  $T$  to length  $N$ 
2 Initialize the transaction-code-rule as dictionary  $txn - dict$ 
3 Initialize the merchant-code-rule as dictionary  $mcc - dict$ 
4 Initialize the nace-code-rule as dictionary  $nace - dict$  for  $i = 1, N$  do
5   Get the transaction  $T_i$  Get the Beneficiary Account  $BAcc_i$  of the
    $T_i$  Load accounts of the customer  $accounts$ 
6   if  $BAcc_i$  is in  $accounts$  then
7     set category of  $T_i =$  'Transfer between own accounts'
8     continue
9   end
10  else
11    Get the Transaction code  $Txncode_i$  of the  $T_i$ 
12    if  $Txncode_i$  is in  $txn - dict$  then
13      set category of  $T_i = txn\_dict_i$ 
14      continue
15    end
16    else
17      if Operational Use or No rule updates then
18        Load the pre-trained Catboost Model  $model$ 
19        set category of  $T_i = model.predict(T_i)$ 
20        continue
21      end
22      else
23        Get the Merchant code  $MCCid_i$  of the  $T_i$  if  $MCCid_i$  is
        in  $mcc - dict$  then
24          set category of  $T_i = mcc\_dict_i$ 
25          continue
26        end
27        else
28          Get the Beneficiary NACE code  $Nace_i$  of the  $T_i$  if
           $Nace_i$  is in  $nace - dict$  then
29            set category of  $T_i = nace\_dict_i$ 
30            continue
31          end
32        end
33      end
34    end
35  end
36 end
```

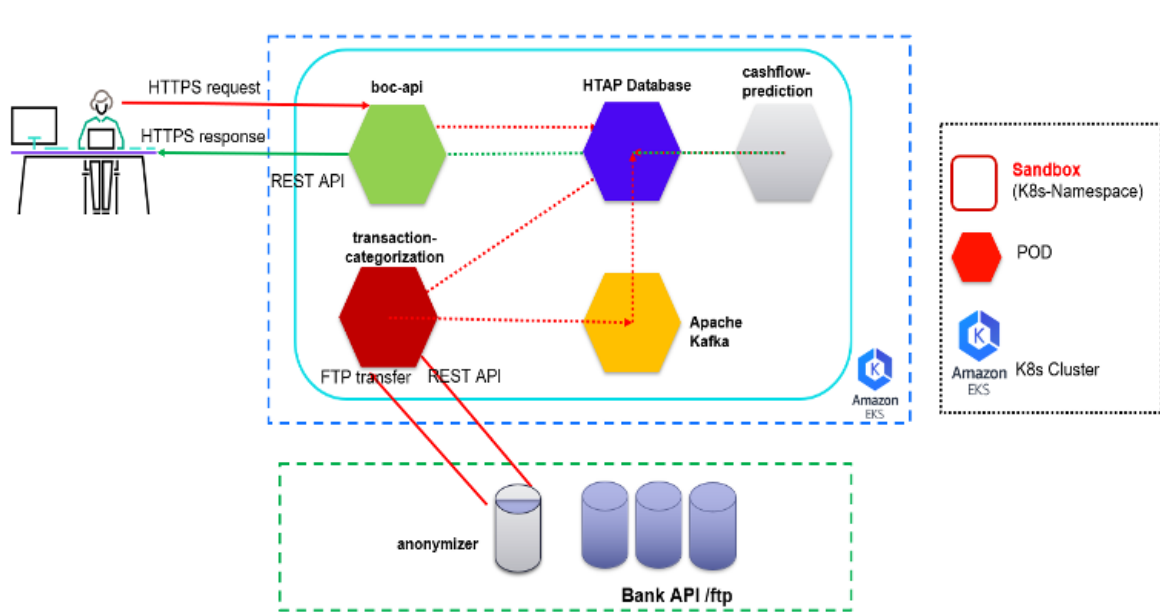
---

Ο ταξινομητής Catboost είναι ένας νέος αλγόριθμος για ενίσχυση κλίσης σε δέντρα απόφασης, ο οποίος έχει την ικανότητα να χειρίζεται τα κατηγορικά χαρακτηριστικά στη φάση της εκπαίδευσης. Αναπτύχθηκε από ερευνητές της Yandex και χρησιμοποιείται για σε διάφορες εργασίες όπως συστήματα συστάσεων, προσωπικοί βοηθοί, αυτοοδηγούμενα αυτοκίνητα, πρόβλεψη καιρού και πολλές άλλες εργασίες. Το CatBoost διαθέτει υλοποιήσεις CPU και GPU που επιταχύνουν τη διαδικασία εκπαίδευσης. Όπως όλες οι προσεγγίσεις gradient boosting, το



CatBoost αποτελείται από δύο φάσεις κατά την κατασκευή δέντρων. Η πρώτη φάση έχει ως αποτέλεσμα την επιλογή της δομής του δέντρου και η δεύτερη φάση ορίζει την τιμή των φύλλων για το σταθερό δέντρο. Μία από τις σημαντικές βελτιώσεις του CatBoost είναι η δυνατότητα της αμερόληπτης εκτίμησης κλίσης προκειμένου να ελεγχθεί η υπερπροσάρτηση. Για το σκοπό αυτό, για κάθε επανάληψη της ενίσχυσης, για την πραγματική εκτίμηση της διαβάθμισης κάθε δείγματος, αποκλείει αυτό το δείγμα από το σετ εκπαίδευσης του τρέχοντος μοντέλου συνόλου. Μια άλλη βελτίωση αναφέρεται στον αυτόματο μετασχηματισμό κατηγορικών χαρακτηριστικών σε αριθμητικά χαρακτηριστικά χωρίς να έχουν υποστεί προεπεξεργασία. Το CatBoost επιπλέον μπορεί να εφαρμοστεί τόσο σε δυαδικά όσο και σε προβλήματα πολλαπλών κλάσεων.

Στόχος μας είναι να χρησιμοποιήσουμε τα προαναφερθέντα μοντέλα μηχανικής και βαθιάς μάθησης για να παράγουμε πληροφορίες αξιοποιώντας τις διαθέσιμες χρηματοοικονομικές συναλλαγές των MME προκειμένου να παρακολουθήσουμε και να βελτιστοποιήσουμε την οικονομική τους συμπεριφορά. Η εννοιολογική αρχιτεκτονική του προτεινόμενου μηχανισμού απεικονίζεται στην Εικόνα 24.



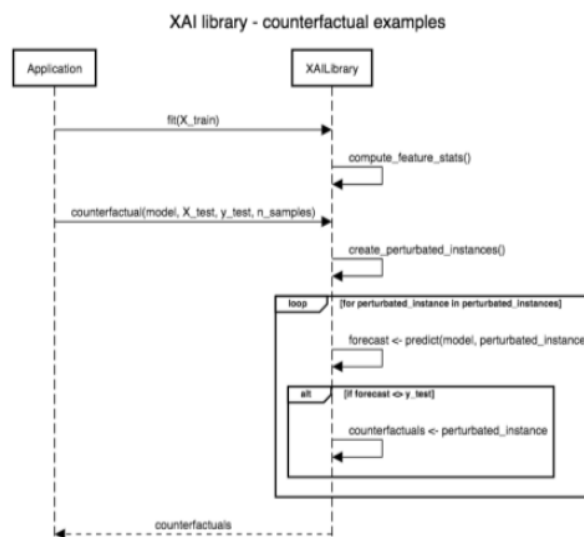
Εικόνα 24 Εννοιολογική Αρχιτεκτονική της προτεινόμενης προσέγγισης

### 3.6 Επεξηγήσιμη Τεχνητή Νοημοσύνη - eXplainable AI (XAI) models

Η εσωτερική αρχιτεκτονική ενός στοιχείου XAI που είναι μέρος ενός ολοκληρωμένου εργαλείου απεικονίζεται στα ακόλουθα διαγράμματα UML (Εικόνα 25, Εικόνα 26). Τα διαγράμματα απεικονίζουν τόσο την εκτέλεση της λογικής

αντιπαράθεσης όσο και την κατάταξη χαρακτηριστικών με βάση τη σχετική σημασία των χαρακτηριστικών του μοντέλου μηχανικής μάθησης αναφορικά με τις προβλέψεις του μοντέλου. Αυτές είναι οι δύο από τις κύριες «εσωτερικές» λειτουργίες της μονάδας XAI.

Για την παραγωγή αντιπαραδειγμάτων σε κάθε περίπτωση εφαρμογής, το εργαλείο XAI καλείται κατά την προσαρμογή του μοντέλου μηχανικής μάθησης στο πλήρες σύνολο δεδομένων εισόδου, προκειμένου να υπολογιστούν τα στατιστικά στοιχεία που είναι απαραίτητα για την ερμηνεία αυτο. Αυτό θα φέρει το XAI σε μια κατάσταση όπου μπορεί να παράγει αντιπαραδείγματα, συνήθως για όλα ή μέρος των δεδομένων δοκιμής. Για τις διαταραγμένες περιπτώσεις, ο αλγόριθμος θα ελέγξει ποιες αλλαγές επιφέρουν στις προβλέψεις μοντέλων και θα επιστρέψει τα συλλεγμένα αντι γεγονότα πίσω στον χρήστη. Φυσικά, τα παραπάνω βήματα θα διαφέρουν ανάλογα με την υποκείμενη υλοποίηση, αλλά η γενική διεπαφή για αντιπαραδείγματα αναμένεται να συμμορφώνεται με την παραπάνω ακολουθία.

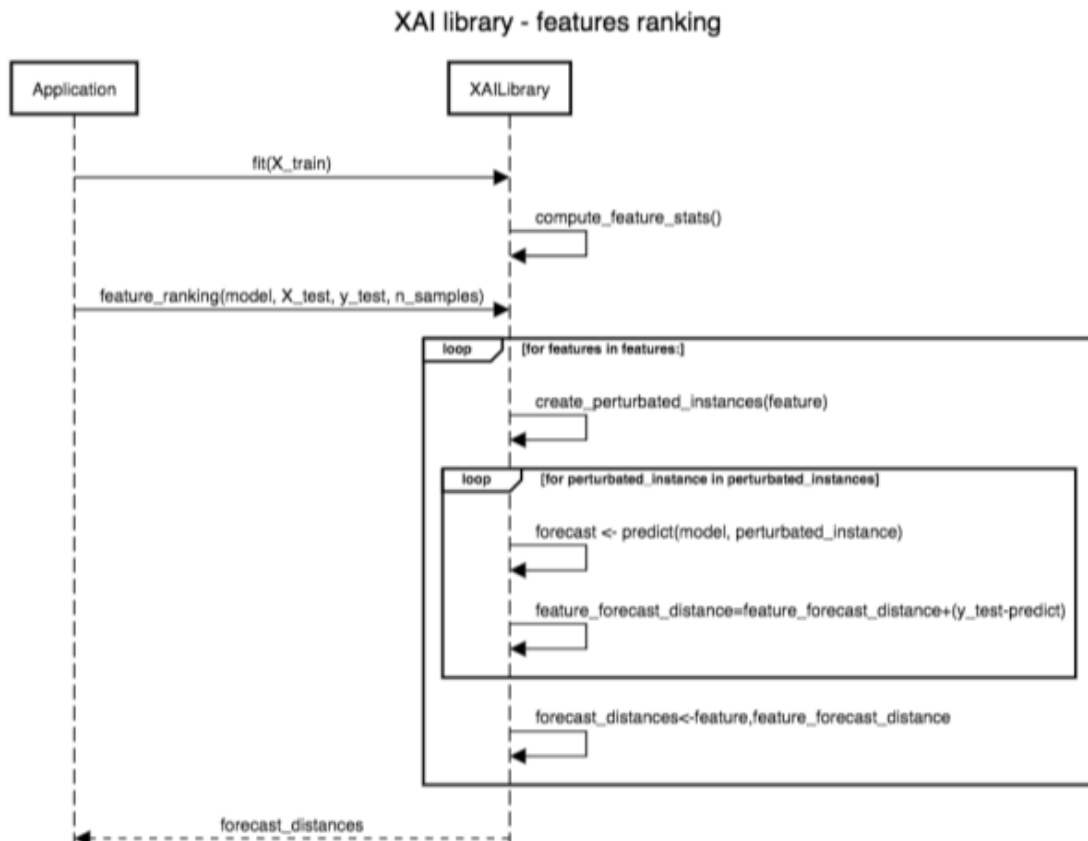


**Εικόνα 25** Παροχή παραστατικών Πληροφοριών από το την βιβλιοθήκη XAI.

Η κατάταξη χαρακτηριστικών βάση της «σημαντικότητας» τους στο μοντέλο πρόβλεψης εμφανίζεται στην Εικόνα 26. Η διαδικασία ξεκινά ξανά κατά την εκπαίδευση του μοντέλου μηχανικής μάθησης στο σετ εκπαίδευσης, το οποίο θα βοηθήσει τον αλγόριθμο XAI να προετοιμαστεί συλλέγοντας στατιστικές πληροφορίες. Στη συνέχεια, αυτές οι πληροφορίες γίνονται διαθέσιμες με εφαρμογή δυναμικών κλήσεων με δοκιμαστικά δείγματα, τα οποία θα απαντώνται με βαθμολογίες σπουδαιότητας χαρακτηριστικών για τις προβλέψεις των παρεχόμενων δειγμάτων δεδομένων. Όπως και στο προηγούμενο παράδειγμα, ο αλγόριθμος XAI θα απαιτήσει πρόσβαση στο εκπαιδευμένο μοντέλο και στη μέθοδο



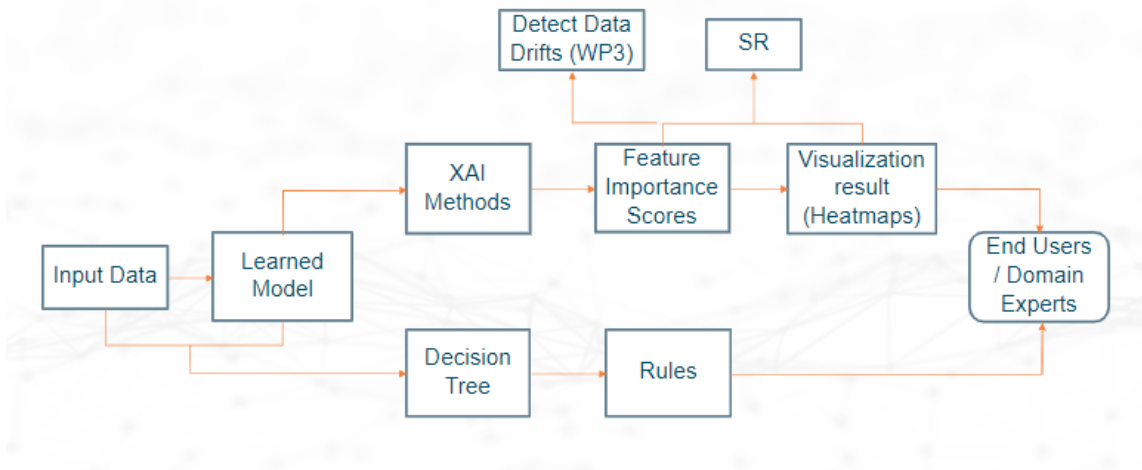
πρόβλεψης για την εκτέλεση υπολογισμών ερμηνευσιμότητας. Η διαδικασία υπολογισμού της σημασίας των χαρακτηριστικών βασίζεται, όπως και πριν, στη δημιουργία διαταραγμένων περιπτώσεων για κάθε δείγμα εισόδου. Η μορφή του αποτελέσματος μπορεί να διαφέρει ανάλογα με την υποκείμενη μέθοδο, αλλά συνήθως για αυτούς τους τύπους μεθόδων θα είναι μια αντιστοίχιση του χαρακτηριστικού με τη βαθμολογία σπουδαιότητάς του με βάση κάποια βαθμολογία διαφοράς χρησιμοποιώντας τη διαφορά δυνατότητας-πρόβλεψης.



**Εικόνα 26** Παροχή κατάταξης χαρακτηριστικών από την βιβλιοθήκη XAI

Παρακάτω στην Εικόνα 27 μπορούμε επίσης να δούμε πώς μοιάζει μια τυπική ροή εργασιών για τη λειτουργία συστημάτων XAI βαθμολόγησης χαρακτηριστικών, η οποία αποτελείται από βήματα όπως,

α. Απόκτηση και εισαγωγή δεδομένων: αυτό το αρχικό βήμα αφορά τη συλλογή των απαραίτητων δεδομένων για εκπαίδευση μοντέλων από διάφορες πηγές δεδομένων και με διάφορους τρόπους. Μπορεί επίσης να περιλαμβάνει (ανάλογα με τον επιλεγμένο αλγόριθμο XAI) στατιστικά χρήσιμα για ερμηνεία που μπορούν να εξαχθούν από τα ακατέργαστα δεδομένα εισόδου.



Εικόνα 27 Εσωτερική ροή εργασίας XAI.

β. Μοντέλα μηχανικής μάθησης: τα μοντέλα μηχανικής μάθησης χρησιμοποιούν τα δεδομένα εκπαίδευσης που παρέχονται από το προηγούμενο επίπεδο για την προσαρμογή των δεδομένων σύμφωνα με την εργασία που πρέπει να εκτελέσουν τα μοντέλα. Είναι σημαντικό να υπάρχει αντιστοιχία μεταξύ της επιλογής των μοντέλων, των τρόπων δεδομένων και των αλγορίθμων XAI που εφαρμόζονται, καθώς δεν είναι όλοι οι αλγόριθμοι XAI κατάλληλοι για όλες τις περιπτώσεις χρήσης και δεν προσφέρονται όλα τα μοντέλα για ερμηνεία. Το μοντέλο εκμάθησης θα είναι βασικό μέρος στη διαδικασία ερμηνείας, επομένως μια βιβλιοθήκη XAI θα χρειαστεί πρόσβαση στις δυνατότητες πρόβλεψης των μοντέλων μηχανικής μάθησης (σε περίπτωση μεθόδου μαύρου κουτιού), αλλά υπάρχουν και μέθοδοι που θα χρειαστούν πρόσθετη πρόσβαση στο μοντέλο, αυτές κυρίως είναι οι μέθοδοι βαθιάς μάθησης.

γ. Χρήση μεθόδων XAI: Οι μέθοδοι XAI για συγκεκριμένες περιπτώσεις εφαρμόζονται αρχικά για τη συλλογή χρήσιμων πληροφοριών επεξήγησης και στατιστικών στοιχείων κατά τη διάρκεια της εκπαίδευσης μοντέλων. Αυτό το βήμα φέρνει τους αλγόριθμους XAI σε καλύτερη θέση για να εφαρμοστούν σε νέα (δοκιμαστικά) δείγματα και να παράγουν κατανοητές εξηγήσεις από τον άνθρωπο κατά τη λειτουργία σε πραγματικό κόσμο.

δ. Εξαγωγή κανόνων (προαιρετικό): παράλληλα με την διαδικασία εξαγωγής της σπουδαιότητας των χαρακτηριστικών ενός μοντέλου μηχανικής μάθησης, αυτό το μέρος της ροής εργασίας στοχεύει στην εξαγωγή κανόνων κατανοητών από τον άνθρωπο. Είναι προαιρετικό, καθώς μπορεί να εφαρμοστεί σε περιορισμένο αριθμό τοπολογιών δεδομένων (π.χ. δεδομένα πίνακα), ενώ δεν είναι συμβατό με άλλες (π.χ. δεδομένα εικόνας που ταιριάζουν καλύτερα με χάρτες θερμότητας). Ένα παράδειγμα τέτοιου κανόνα μπορεί να είναι η εξήγηση μιας απότομης αύξησης της ζήτησης λόγω της τιμής ενός αγαθού που ξεπερνά ένα συγκεκριμένο όριο, παρόμοιο με αυτό που παράγεται από κινητήρες τεχνητής νοημοσύνης που





βασίζονται σε κανόνες. Για να επιτευχθεί αυτό, πολλές μέθοδοι προσπαθούν να μάθουν έναν υποκατάστατο ταξινομητή (π.χ. ένα δέντρο αποφάσεων) παράλληλα με έναν πιο περίπλοκο (π.χ. ένα νευρωνικό δίκτυο) και στη συνέχεια να εξαγάγουν απευθείας τους κανόνες από τον υποκατάστατο ταξινομητή (ανίχνευση μιας διαδρομής στην απόφαση δέντρο), το οποίο μπορεί να λειτουργήσει καλά και να οδηγήσει σε σαφείς κανόνες για ορισμένες περιοχές των κρατικών χώρων.

ε. Υπολογισμός βαθμολογιών σπουδαιότητας χαρακτηριστικών: τα αποτελέσματα του προηγούμενου βήματος θα αξιοποιηθούν για την εκτέλεση υπολογισμών σε πραγματικό χρόνο που σχετίζονται με την απόδοση χαρακτηριστικών μέσω των βαθμολογιών σπουδαιότητας χαρακτηριστικών και της παροχής αντιπραγματικών πληροφοριών.

στ. Οπτικοποίηση αποτελεσμάτων: Να είναι κατανοητό από τον άνθρωπο, ειδικά από μη-εμπειρογνώμονες, τα αποτελέσματα ΧΑΙ θα πρέπει να απεικονίζονται σωστά. Αυτό θα μπορούσε να είναι τόσο απλό όσο ένας πίνακας για εισόδους διανυσμάτων, σε χάρτες θερμότητας που επικαλύπτονται πάνω από την αρχική εικόνα, σε πιο εξελιγμένα γραφήματα και γραφήματα που υπογραμμίζουν τη σημασία των χαρακτηριστικών.

η. Παρουσίαση των αποτελεσμάτων σε τελικούς χρήστες και εμπειρογνώμονες τομέα: αυτό το βήμα συνήθως συνδυάζει τα αποτελέσματα διαφορετικών μεθόδων ΧΑΙ και οπτικοποίησης για να παρουσιάσει τον τρόπο λειτουργίας των αλγορίθμων και τη λογική ως συνεκτικό σύνολο, με στόχο την αύξηση της εμπιστοσύνης των μη ειδικών ενδιαφερομένων.

Τα παραπάνω διαγράμματα είναι τα προσχέδια μιας βιβλιοθήκης που στοχεύει να παρέχει ερμηνευσιμότητα σε διαφορετικά στοιχεία του STAR σε διαφορετικά WPs, που κυμαίνονται από τα μοντέλα εκμάθησης που χρησιμοποιούνται στη γραμμή παραγωγής, όπως η αυτοματοποιημένη επιθεώρηση ποιότητας στο PCL UC2 έως τις μονάδες ανίχνευσης επίθεσης του WP3 και ακόμη δυνητικά εξαρτήματα WP5, όπως ενισχυτική εκμάθηση και ανίχνευση ζωνών ασφαλείας. Οι αλληλεπιδράσεις εξαρτημάτων επεξηγούνται λεπτομερώς στην επόμενη υποενότητα.

Δεδομένου ότι στο πλαίσιο της Διατριβής το ΧΑΙ είναι μια βιβλιοθήκη αλγορίθμων, η εφαρμογή του θα ακολουθήσει διαφορετική πορεία από τα παραδοσιακά στοιχεία λογισμικού. Ο σημαντικότερος παράγοντας που παίζει ρόλο είναι η κατάλληλη επιλογή και αξιολόγηση μεθόδων, οι οποίες, μετά από αυστηρό έλεγχο και επαλήθευση της καταλληλότητάς τους για τις διάφορες περιπτώσεις χρήσης.

Τέλος, υπάρχει μια ποικιλία από βιβλιοθήκες ΧΑΙ εκτός ραφιού που αξιοποιήθηκαν ή δύναται να ενσωματωθούν στο μέλλον για την υλοποίηση αυτού του εργαλείου.



Μερικά από τα αξιολογητέα είναι τα ακόλουθα:

- **LIME**: υλοποιεί την οικογένεια μεθόδων Local Interpretable Model-agnostic Explanations για πολλαπλούς τρόπους, όπως διανύσματα, εικόνες και κείμενο. Λαμβάνει τη μέθοδο `model.predict` ως είσοδο και δεν εξαρτάται από υποκείμενες υλοποιήσεις. Δεν απαιτείται GPU εκτός εάν το μοντέλο είναι πραγματικά πολύπλοκο και πολλές περιπτώσεις χρειάζονται ερμηνεία.

- **SHAP**: Εφαρμόζει επεξηγήσεις χαρακτηριστικών και είναι από άποψη χρήσης παρόμοια με το LIME χωρίς να συνδέεται με ένα συγκεκριμένο μοντέλο.

Επεκτείνεται επίσης σε εικόνες.

- **EthicalML**: και αναπτύχθηκε από την IBM, παρέχει πολλά πρόσθετα βοηθητικά προγράμματα για φόρτωση δεδομένων, λειτουργίες και οπτικοποιήσεις αποτελεσμάτων με στόχο να είναι μια ολοκληρωμένη λύση. Μόνο η εξάρτηση είναι από το `scipy` και το `scikit`, επί του παρόντος, ίσως όχι αρκετά ώριμο για πολύπλοκα μοντέλα βαθιάς μάθησης.

- **Dalex**: έχει ενσωματώσει διαφορετικά πλαίσια ML καθώς και μεθόδους ΧΑΙ όπως SHAP και LIME και μπορεί εύκολα να ενσωματωθεί τόσο με `scikit` όσο και με `keras` (επάνω `tensorflow`) καθώς και `xgboost`, `mlr` και `mlr3`. Περιλαμβάνει επίσης εντυπωσιακά εργαλεία οπτικοποίησης και πίνακες εργαλείων.

- **InterpretML**: υλοποιήθηκε από τη Microsoft και υποστηρίζει μεταξύ άλλων SHAP, Lime, Partial Dependence Plots (PDP) και Sensitivity Analysis. Αν και είναι μια καλή εργαλειοθήκη, προς το παρόν έχει περιορισμένη υποστήριξη για δεδομένα εικόνας και κειμένου.

- **Grad-CAM**: αποτελεί μια οικογένεια μεθόδων ειδικών για συνελκτικα νευρωνικά δίκτυα. Πολλές παραλλαγές έχουν εφαρμοστεί στο πακέτο `rip grad-cam`, το οποίο εφαρμόζεται επί του παρόντος στο PyTorch και υποστηρίζει επίσης τη χρήση GPU. Σημειώστε ότι αυτές οι μέθοδοι συνήθως απαιτούν κάποια πρόσβαση στο πραγματικό μοντέλο (π.χ. βάρη, αρχιτεκτονική) για την εκτέλεση υπολογισμών `gradient` και δεν αντιμετωπίζουν το μοντέλο ως μαύρο κουτί.

Πρόθεσή μας δεν είναι μόνο να εκμεταλλευτούμε πλήρως αυτές τις μεθόδους και να εφαρμόσουμε την καλύτερη για κάθε συγκεκριμένη περίπτωση, αλλά και να αναπτύξουμε τις δικές μας επεκτάσεις και βελτιώσεις, ιδίως σε ορισμένες από τις δύσκολες περιπτώσεις χρήσης μας (π.χ. αυτοματοποιημένη αναγνώριση ελαττωμάτων) μπορεί να επεκτείνει αυτές τις πιο γενικής χρήσης μεθόδους.

Αυτή η ενότητα καταγράφει τα μοντέλα, τις τεχνικές και τα αντίστοιχα μετα-ΧΑΙ μοντέλα ΧΑΙ που ήταν υπό μελέτη στο πλαίσιο της Διατριβής. Ο κύριος σκοπός μας είναι μία διεπαφή που θα ενσωματώνει διαφορετικούς αλγόριθμους Επεξηγήσιμης Τεχνητής Νοημοσύνης (ΧΑΙ). Προς αυτή την κατεύθυνση, το προτεινόμενο πλαίσιο



θα ενσωματώσει καθιερωμένες λύσεις και μοντέλα XAI κάνοντας τις κατάλληλες τροποποιήσεις, όταν απαιτούνται, προσθέτοντας αξία σε βιομηχανικούς φορείς.

### 3.6.1 XAI για χρονοσειρές

Όλες οι μέθοδοι που αναφέρονται παρακάτω ισχύουν για χρονοσειρές. Αρχικά, αναφέρουμε μεθόδους post-hoc που προσεγγίζουν τη συμπεριφορά ενός μοντέλου εξάγοντας σχέσεις μεταξύ τιμών χαρακτηριστικών και προβλέψεων, σε αυτήν την περίπτωση, οι καθυστερήσεις χαρακτηριστικών των χρονοσειρών παίζουν επίσης ρόλο. Ενώ οι ante-hoc μέθοδοι ενσωματώνουν την εξήγηση στη δομή του μοντέλου, η οποία επομένως είναι ήδη εξηγήσιμη στο τέλος της φάσης εκπαίδευσης.

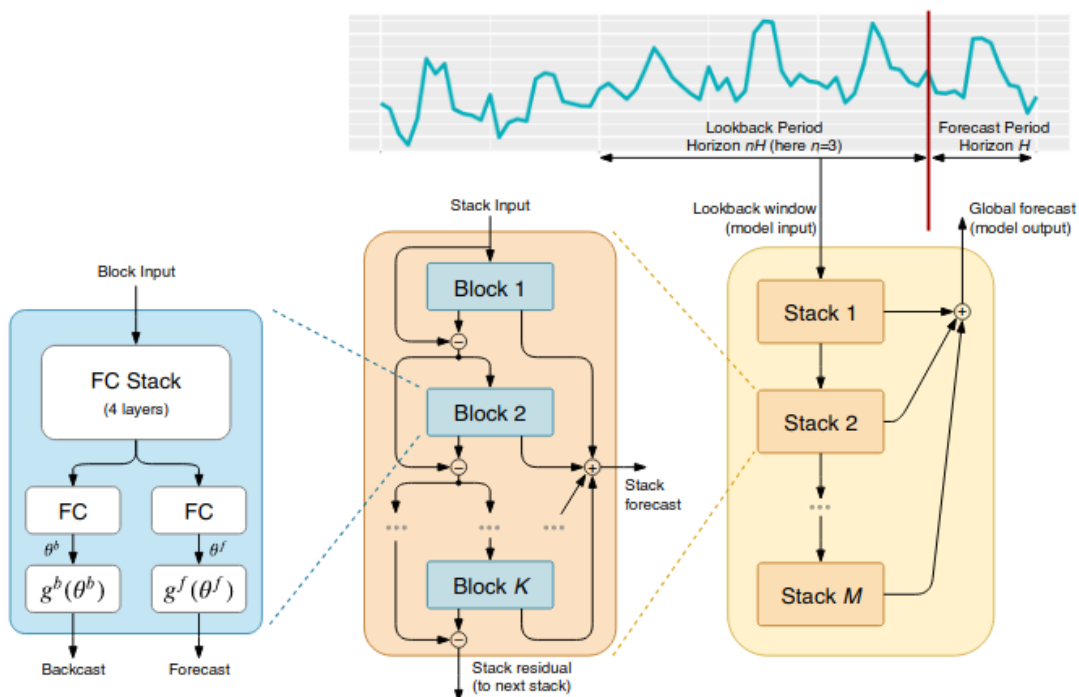
Η πρώτη κατηγορία περιλαμβάνει παραλλαγές των ευρέως χρησιμοποιούμενων [Lime] LIME, [220] k-LIME, [221] DLIME, [222] LIMETree και [Shar] SHAP (π.χ. [223] TimeSHAP) καθώς και [224] Anchors, Local Foil Trees, ή [225] LoRE. Υποκατάστατα μοντέλα κατασκευάζονται για κάθε δείγμα πρόβλεψης στις περισσότερες από αυτές τις τεχνικές, μαθαίνοντας τη συμπεριφορά του μοντέλου αναφοράς στη συγκεκριμένη περίπτωση ενδιαφέροντος προσθέτοντας διαταραχές (ή κάλυψη) στις μεταβλητές του διανύσματος χαρακτηριστικών. Οι πολυάριθμες τεχνικές παραλλαγής των χαρακτηριστικών για την αξιολόγηση της συνεισφοράς των χαρακτηριστικών στην προβαλλόμενη τιμή όταν διαγράφονται ή καλύπτονται είναι σχεδόν ένα ξεχωριστό θέμα έρευνας σε αυτό το πλαίσιο. Αυτά τα μοντέλα XAI μπορούν να χρησιμοποιηθούν και για DNN, καθώς δεν επηρεάζονται από το υποκείμενο μοντέλο μηχανικής εκμάθησης.

Από εκεί και πέρα, στα μοντέλα βαθιάς μάθησης, τα πράγματα γίνονται πιο περίπλοκα όταν πρόκειται για προσεγγίσεις ante-hoc. Αξίζει να αναφερθεί η μέθοδος Gradient-Input, η οποία υπολογίζει τις ενεργοποιήσεις νευρώνων και φίλτρων για μια συγκεκριμένη περίπτωση πολλαπλασιάζοντας την είσοδο με τη μερική παράγωγο ενός στρώματος σε σχέση με την είσοδο. Ακολούθησαν παρόμοιες προσεγγίσεις, όπως [226] Deep Learning Important Features (DeepLIFT) και Smooth-Grad. Η εισαγωγή του επιπέδου προσοχής θεωρείται επίσης ως πηγή εξήγησης, καθώς παρέχει πληροφορίες για τα χρονικά σημεία που σχετίζονται με την πρόβλεψη.

Πέρα από αυτά, αξίζει να αναφέρουμε προσεγγίσεις όπως το [227] RETAIN (Reverse Time Attention) με εφαρμογή σε δεδομένα Ηλεκτρονικών Μητρώων Υγείας (EHR). Το RETAIN επιτυγχάνει υψηλή ακρίβεια ενώ παραμένει κλινικά ερμηνεύσιμο και βασίζεται σε ένα μοντέλο νευρικής προσοχής δύο επιπέδων που ανιχνεύει προηγούμενες επισκέψεις. Τέλος, το [228] NBEATS εστιάζει σε μια αρχιτεκτονική που βασίζεται σε υπολειπόμενες συνδέσεις και σε μια στοίβα πλήρως συνδεδεμένων επιπέδων. Η αρχιτεκτονική έχει μια σειρά από επιθυμητές ιδιότητες,

καθώς είναι ερμηνεύσιμη, εφαρμόσιμη χωρίς τροποποίηση σε ένα ευρύ φάσμα περιπτώσεων.

Στο πλαίσιο της διατριβής χρησιμοποιήσαμε το και ενσωματώσαμε το μοντέλο NBEATS, καθώς πιστεύουμε ότι θα ταιριάζει με τα δεδομένα χρονοσειρών των περιπτώσεων που έχουμε πρόβλεψη χρονοσειρών. Το NBEATS είναι ένα ενδιαφέρον βήμα για την εφαρμογή βαθιάς μάθησης σε χρονοσειρές, επειδή δημιουργεί μια αρχιτεκτονική αφιερωμένη στις χρονοσειρές. Η προηγούμενη προσέγγιση συνίσταται στη μετάφραση αλληλουχιών (ακολουθίες σε ακολουθία). Τα χρονικά σημεία δίνονται στο δίκτυο το ένα μετά το άλλο και το δίκτυο ενημερώνει κάποια εσωτερική μνήμη για να ενημερώσει την εσωτερική αναπαράσταση της κατάστασης του συστήματος. Στη συνέχεια, η έξοδος υπολογίζεται χρησιμοποιώντας αναπαράσταση εσωτερικής κατάστασης και τρέχουσα έξοδο.



Εικόνα 28 Αρχιτεκτονική του XAI μοντέλου N-BEATS

Τα επαναλαμβανόμενα νευρωνικά δίκτυα κάνουν μόνο αυτό, ενώ το LSTM χρησιμοποιεί αρκετούς διαφορετικούς μηχανισμούς για να υπολογίσει ρητά ποια μέρη πρέπει να ξεχάσει και ποια μέρη να ενημερώσει δεδομένης τρέχουσας εισόδου.

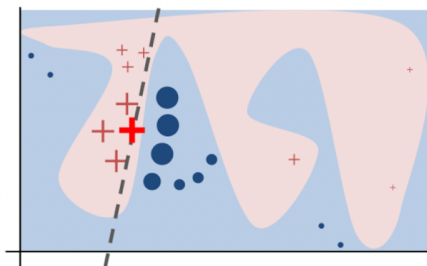
Το NBEATS χρησιμοποιεί μια εντελώς διαφορετική προσέγγιση: παίρνει ένα ολόκληρο παράθυρο προηγούμενων τιμών και υπολογίζει πολλές τιμές χρονικών σημείων πρόβλεψης σε ένα μόνο πέρασμα. Για να γίνει αυτό, χρησιμοποιεί εκτενώς πλήρως συνδεδεμένα επίπεδα. Αποτελείται από πολλά μπλοκ που συνδέονται με υπολειπόμενο τρόπο: το πρώτο προσπαθεί να μοντελοποιήσει το παρελθόν

παράθυρο (backcast) και το μέλλον (πρόβλεψη) όσο το δυνατόν ακριβέστερα, το δεύτερο προσπαθεί να μοντελοποιήσει μόνο το υπολειπόμενο σφάλμα του παρελθόντος του προηγούμενου μπλοκ (και ενημερώνει την πρόβλεψη με βάση αυτό το σφάλμα) και ούτω καθεξής. Η πρόβλεψη είναι το άθροισμα των προβλέψεων από πολλά μπλοκ, όπου το πρώτο μπλοκ πιάνει τις κύριες τάσεις, το δεύτερο ειδικεύεται σε μικρότερα σφάλματα κ.λπ. Αυτή η εναπομένουσα αρχιτεκτονική έχει επίσης τα πλεονεκτήματα της τεχνικής ενίσχυσης/συναρμολόγησης (που χρησιμοποιείται στην κλασική μηχανική μάθηση) καθώς η πρόβλεψη είναι το άθροισμα των προβλέψεων από πολλά μπλοκ.

### 3.6.2 ΧΑΙ για δεδομένα πίνακα

Μια ερμηνεύσιμη αναπαράσταση της εισόδου που είναι κατανοητή από τους ανθρώπους είναι βασική απαίτηση για το μοντέλο LIME. Ένα διάνυσμα BoW (bag of words) για NLP ή μια εικόνα είναι παραδείγματα ερμηνεύσιμων αναπαραστάσεων.

Το LIME δημιουργεί μια λίστα από επεξηγήσεις που υποδεικνύουν τη συμβολή κάθε χαρακτηριστικού στην πρόβλεψη του δείγματος δεδομένων. Αυτό επιτρέπει την τοπική ερμηνεία καθώς και τον προσδιορισμό ποιες προσαρμογές χαρακτηριστικών θα έχουν τον μεγαλύτερο αντίκτυπο στην πρόβλεψη.



Algorithm 1 Sparse Linear Explanations using LIME

```
Require: Classifier  $f$ , Number of samples  $N$   
Require: Instance  $x$ , and its interpretable version  $x'$   
Require: Similarity kernel  $\pi_x$ , Length of explanation  $K$   
 $\mathcal{Z} \leftarrow \{\}$   
for  $i \in \{1, 2, 3, \dots, N\}$  do  
   $z'_i \leftarrow \text{sample\_around}(x')$   
   $\mathcal{Z} \leftarrow \mathcal{Z} \cup \{z'_i, f(z_i), \pi_x(z_i)\}$   
end for  
 $w \leftarrow \text{K-Lasso}(\mathcal{Z}, K)$  ▷ with  $z'_i$  as features,  $f(z)$  as target  
return  $w$ 
```

Εικόνα 29 Επεξήγηση ταξινόμηση με τη μέθοδο LIME.

### 3.6.3 ΧΑΙ για δεδομένα κειμένου (NLP)

Η Επεξεργασία Φυσικής Γλώσσας είναι το θέμα του τρόπου με τον οποίο οι υπολογιστές αλληλοεπιδρούν με τις ανθρώπινες (φυσικές) γλώσσες, ιδίως πώς να εκπαιδεύουν τους υπολογιστές να επεξεργάζονται και να αναλύουν τεράστιους όγκους δεδομένων φυσικής γλώσσας.

Ωστόσο, η αναγνώριση της αξίας χαρακτηριστικών ή οντοτήτων είναι ζωτικής σημασίας στο ΧΑΙ των μοντέλων NLP, καθώς επιδιώκει να προσδιορίσει ποιο μέρος της ομιλίας οδηγεί τις πιο σημαντικές πληροφορίες. Επεξήγηση της λογικής της αλληλουχίας ερωτήσεων στο διάλογο, αντιμετώπιση προβλημάτων ενός



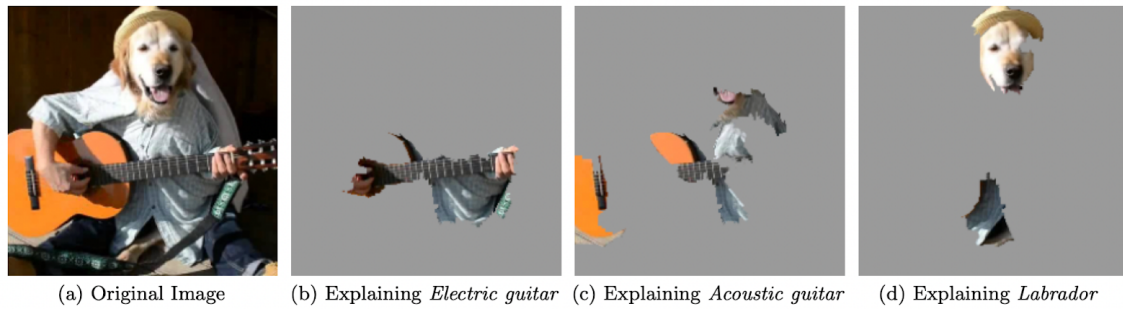
συστήματος διαλόγου που βασίζεται σε σχέδιο ή επεξήγηση των παραγόμενων εκφράσεων, για να αναφέρουμε μερικά παραδείγματα. Υπάρχουν πολλές προσεγγίσεις για τον προσδιορισμό των πιο αντιπροσωπευτικών οντοτήτων σε μια εργασία κατηγοριοποίησης κειμένου.

### 3.6.4 XAI για εικόνες

Κατά αντιστοιχία η κύρια ιδέα αυτού του στοιχείου είναι να παρέχει επεξηγήσεις όσον αφορά τα σημαντικά τμήματα της εικόνας αναφορικά με την πρόβλεψη ενός μοντέλου μηχανικής μάθησης. Δεδομένου ενός εκπαιδευμένου μοντέλου, όπως ένα νευρωνικό δίκτυο, το στοιχείο θα εξάγει τη σημασία κάθε χαρακτηριστικού εισόδου για μια συγκεκριμένη πρόβλεψη. Όταν ασχολούμαστε με δεδομένα εικόνας, η σημασία των χαρακτηριστικών μπορεί να μεταφραστεί στη σημασία κάθε pixel για την πρόβλεψη εξόδου. Το τελευταίο μπορεί επίσης να οπτικοποιηθεί σε έναν χάρτη θερμότητας όπου η σημασία κάθε χαρακτηριστικού (pixel) μπορεί να εμφανιστεί με διαφορετικά χρώματα και μπορεί να λειτουργήσει ως μια εξαιρετική εξήγηση για τον ανθρώπινο χειριστή.

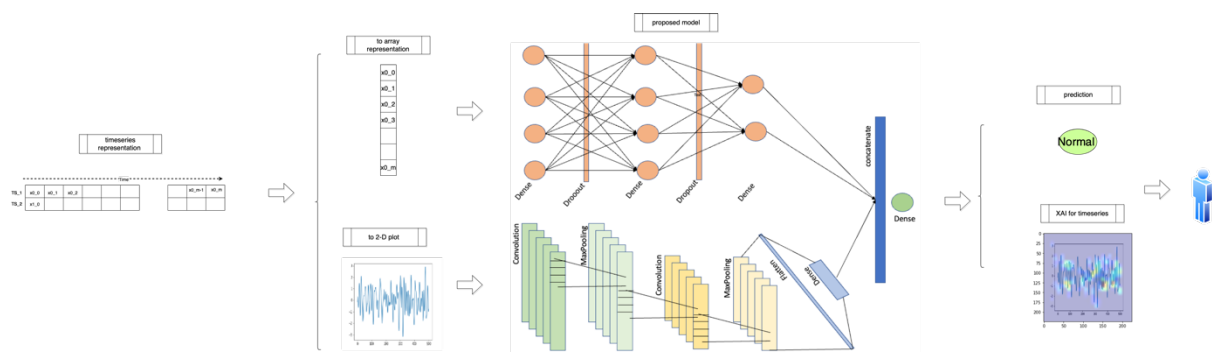
Όσον αφορά τα δεδομένα τύπου εικόνας, υπάρχουν δύο τύποι προσεγγίσεων επεξήγησης: με βάση την απεικόνιση των gradients και των βαρών των νευρώνων και με βάση τις διαταραχές. Οι μέθοδοι της πρώτης κατηγορίας απαιτούν πολλές διαδρομές προς τα πίσω στο νευρωνικό δίκτυο πριν από τον υπολογισμό των βαθμολογιών σημαντικότητας, ενώ οι μέθοδοι που βασίζονται σε διαταραχές απαιτούν απλώς περάσματα προς τα εμπρός μόλις αλλάξει η είσοδος. Το DeepLIFT (Deep Learning Important Features) και το [229] LRP (Layer Wise Relevance Propagation) είναι αλγόριθμοι επεξήγησης της πρώτης κατηγορίας, ενώ οι LIME (Local Interpretable Model Agnostic Explanations) και το PDA είναι μέθοδοι επεξήγησης βασισμένες σε διαταραχές Ανάλυση Διαφοράς). Ο αλγόριθμος SHAP για την επεξήγησης εικόνας.

Οι παραπάνω αλγόριθμοι που είναι ενσωματωμένοι στο στοιχείο έχουν αναπτυχθεί κυρίως σε βιβλιοθήκες Python και μπορούν εύκολα να εφαρμοστούν στις περιπτώσεις χρήσης που εξετάστηκαν στην Διατριβή. Για άλλη μια φορά το LIME είναι μια από τις πιο εύχρηστες τεχνικές (με ερμηνεύσιμες απεικονίσεις) για τον τύπο εικόνας των δεδομένων εισόδου.

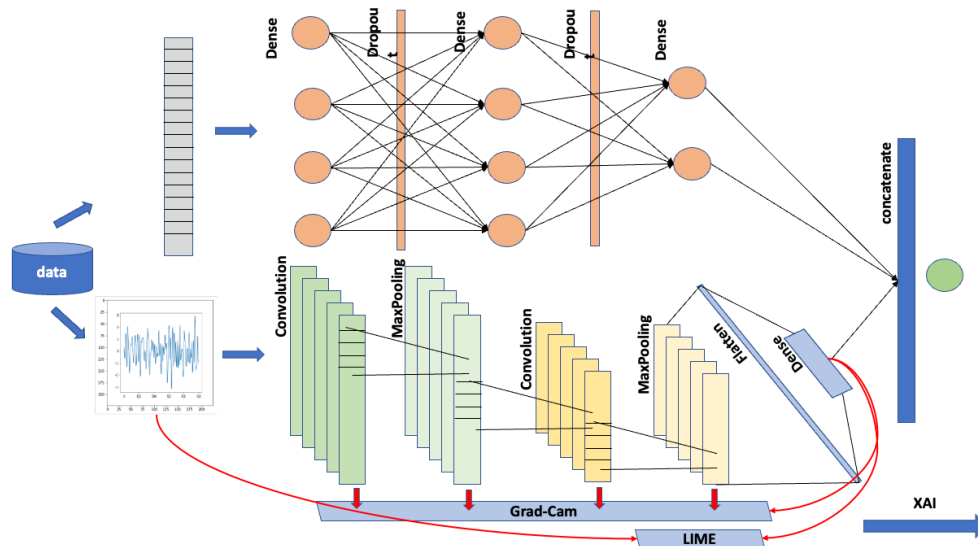


Εικόνα 30 Επεξηγήσεις εικόνας με χρήση LIME

Οι πρωταρχικές ερευνητικές προκλήσεις που έχουν αντιμετωπιστεί μέχρι τώρα στο πλαίσιο της Διατριβής περιλαμβάνουν τη διευκόλυνση της ακριβούς ταξινόμησης χρονοσειρών και τη δημιουργία ερμηνεύσιμων απεικονίσεων ΧΑΙ. Αυτές οι δύο όψεις είναι αλληλένδετες αφού το μοντέλο ΧΑΙ κατασκευάζεται με βάση τα αποτελέσματα της ταξινόμησης. Για να αυξήσουμε τη χρηστικότητα ενός μοντέλου DNN καθιστώντας το εξηγήσιμο, παρουσιάζουμε έναν τύπο μοντέλου απόσταξης (δηλαδή, μια αρχιτεκτονική δασκάλου-μαθητή). Εδώ, ο "δάσκαλος" είναι το σύνθετο μοντέλο DNN που προσφέρει κορυφαία απόδοση και ο "μαθητής" είναι ένα μοντέλο CNN που γίνεται ερμηνεύσιμο στους ανθρώπινους τελικούς χρήστες όταν χρησιμοποιούνται μέθοδοι όπως το LIME και το GradCam. Το σχήμα 1 απεικονίζει τη δομή του προτεινόμενου μοντέλου, με το Πυκνό μοντέλο να ενσαρκώνει τον «δάσκαλο» και το CNN να συμβολίζει τον «εξηγήσιμο»-«μαθητή».



Εικόνα 31 Σχηματική εξήγηση της πτυχής δεδομένων στην προσέγγιση από άκρο σε άκρο, συμπεριλαμβανομένων των κύριων βημάτων στη διαδικασία, από τα πρωτογενή δεδομένα στον χρήστη.



**Εικόνα 32** Η αρχιτεκτονική του προτεινόμενου μοντέλου, όπου το Πυκνό μοντέλο αντιπροσωπεύει τον «δάσκαλο» ενώ το CNN τον «εξηγήσιμο»-«μαθητή». Τα δεδομένα τροφοδοτούνται ως συστοιχία στο δίκτυο Dense ενώ το CNN ως 2D σχεδίαση.

Όσον αφορά το μοντέλο DNN, τα επίπεδα συστατικών του μπορούν να αναλυθούν ως εξής:

- Input layer: Αυτό το επίπεδο έχει ως αποστολή να δέχεται μια εικόνα ως είσοδο.
- Rescaling layer.
- Convolution layer: Αυτά τα στρώματα έχουν σχεδιαστεί για να αποσπάζουν χαρακτηριστικά από την εικόνα εισόδου. Η ποσότητα των φίλτρων και οι διαστάσεις τους μπορούν να καθοριστούν με βάση την πολυπλοκότητα των δεδομένων. Στη δοκιμή μας, αναπτύξαμε 32 φίλτρα μεγέθους 3x3, με βήμα 1.
- Pooling Layer: Η λειτουργία αυτών των επιπέδων είναι να μειώνουν τις χωρικές διαστάσεις των χαρτών χαρακτηριστικών και να περιορίζουν τις υπολογιστικές απαιτήσεις του δικτύου. Μια τυπική επιλογή είναι το max pooling με μέγεθος πιασίνας 2x2 και διασκελισμό 2.
- Flattening Layer: Ο ρόλος αυτού του επιπέδου είναι να μετατρέψει τους πολυδιάστατους χάρτες χαρακτηριστικών σε ένα διάνυσμα 1D, το οποίο στη συνέχεια μπορεί να εισαχθεί σε ένα πλήρως συνδεδεμένο επίπεδο.
- Dense Layer: Ένα ή πολλαπλά πυκνά (πλήρως συνδεδεμένα) στρώματα μπορούν να ενσωματωθούν για την παραγωγή προβλέψεων με βάση τα εξαγόμενα χαρακτηριστικά. Ο αριθμός κόμβων σε κάθε πυκνό στρώμα μπορεί να επιλεγεί σύμφωνα με την πολυπλοκότητα των δεδομένων και την επιθυμητή απόδοση του μοντέλου.





- Επίπεδο εξόδου: Αυτό θα μπορούσε να είναι ένα πυκνό στρώμα που διαθέτει έναν μόνο κόμβο με μια συνάρτηση ενεργοποίησης σιγμοειδούς για δυαδική ταξινόμηση ή θα μπορούσε να έχει τόσους κόμβους όσες υπάρχουν κλάσεις στα δεδομένα σας για ταξινόμηση πολλαπλών κλάσεων.



## 4 ΚΕΦΑΛΑΙΟ - ΑΞΙΟΛΟΓΗΣΗ

### ΠΕΡΙΕΧΟΜΕΝΑ

- 4.1 Μηχανική Μάθηση & προβλεπτική συντήρηση κύριων μηχανών εμπορικών πλοίων.
- 4.2 Μηχανική Μάθηση & ασφάλεια των τροφίμων.
- 4.3 Μηχανική Μάθηση & ποιότητα χοιρινού κρέατος.
- 4.4 Μηχανική Μάθηση & τραπεζικός βοηθός για μικρομεσαίες επιχειρήσεις.
- 4.5 Επεξηγήσιμη Τεχνητή Νοημοσύνη - eXplainable AI (XAI)

### ΠΕΡΙΛΗΨΗ

Στο **ΚΕΦΑΛΑΙΟ 4**, λαμβάνει χώρα η αξιολόγηση της προτεινομένης προσέγγισης της εφαρμογής μεθόδων μηχανικής μάθησης σε διαδικασίες της βιομηχανίας επιλύοντας πραγματικά προβλήματα σε πραγματικές συνθήκες μεσώ τεσσάρων (4) διαφορετικών πειραμάτων. Η 1<sup>η</sup> περίπτωση που αναφέρεται στη προβλεπτική συντήρηση κύριων μηχανών εμπορικών πλοίων, αποδεικνύει την εφικτότητα εφαρμογής ενός συνδυαστικού μοντέλου αποτελούμενο από διαφορετικά μοντέλα μηχανικής μάθησης και στατιστικά μοντέλα με σκοπό την έγκαιρη πρόβλεψη της διάβρωσης κεφαλής σταυρού του ρουλεμάν της μηχανής των πλοίων. Η 2<sup>η</sup> περίπτωση αποδεικνύει την εφαρμογή, με ενθαρρυντικά αποτελέσματα, ενός ολοκληρωμένου πλαισίου ανάλυσης δεδομένων, αποτελούμενο από μοντέλα μηχανικής μάθησης, που εφαρμόζεται για έγκαιρη πρόβλεψη των ανακλήσεων τροφίμων. Όπως και το 2<sup>ο</sup> πείραμα, έτσι και το 3<sup>ο</sup>, πραγματοποιείται στο πλαίσιο της ασφάλειας τροφίμων. Σε αυτή την περίπτωση εξετάστηκαν οι παράγοντες που μπορεί να οφείλονται για την εμφάνιση του φαινομένου της κηλίδωσης κάπρου. Η 4<sup>η</sup> περίπτωση εξετάζει αντίστοιχα την πρόβλεψη ταμειακών ροών των ΜΜΕ, βάσει ιστορικών δεδομένων των συναλλαγών. Στο τέλος του Κεφαλαίου, πραγματοποιείται σχετική συζήτηση και ανάλυση των εξαγομένων αποτελεσμάτων όλων των πειραμάτων, παραθέτοντας σχετικά συμπεράσματα για την ολική επίδοση και απόδοση του εκάστοτε μηχανισμού.

### 4.1 Μηχανική Μάθηση & προβλεπτική συντήρηση κύριων μηχανών εμπορικών πλοίων.

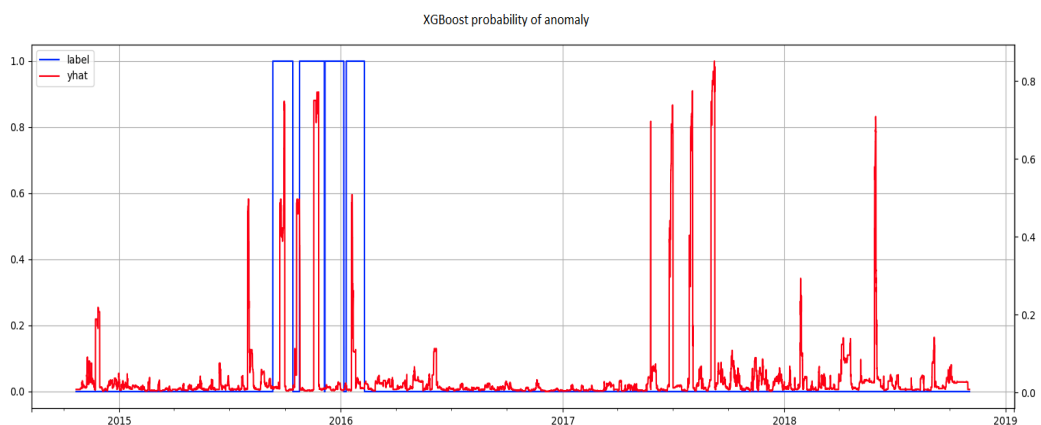


Μια κύρια πρόκληση όσον αφορά την αξιολόγηση περιπτώσεων βλαβών και δει για την πρόβλεψη αυτών, είναι το γεγονός ότι είναι άγνωστο το με ποιόν τρόπο ακριβώς το εξεταζόμενο ελάττωμα εξελίσσεται με τον χρόνο. Στην εξεταζόμενη περίπτωση προτάθηκε ένα στατικό παράθυρο που είναι κατάλληλο, και στο οποίο υποθέτουμε ότι το ελάττωμα εξελίσσεται. Το διάστημα αυτό είναι 30 ημέρες.

Ως εκ τούτου, αξιολογούμε ως Πραγματικό Θετικό αν η πρόβλεψη μας έχει εμπέσει στο στατικό παράθυρο πριν την εμφάνιση του ελαττώματος στην πραγματικότητα. Ειδικά, η αξιολόγηση κάθε μοντέλου, έλαβε υπόψη της τις υπό εξέταση τιμές precision και recall, βασισμένες στο προαναφερθέν σχήμα στατικού παραθύρου. Η αναπαράσταση από την οποία μπορούμε να εξάγουμε σημαντικά συμπεράσματα για την απόδοση των μοντέλων είναι ο πίνακας Confusion Matrix.

Επιπλέον πρέπει να αναφέρουμε ότι το τελικό αποτέλεσμα στο συγκεκριμένο πρόβλημα που αποδεικνύεται αποδοτικός ταξινομητής αποτελείται από τον συνδυασμό πολλών αδύναμων και διαφοροποιημένων ταξινομητών.

Gradient Boosting (Ταξινόμηση): Η Εικόνα 33 δείχνει την πιθανότητα που έχει προβλεφθεί, κάθε χρονική στιγμή, για ένα πλοίο που έχει αυτόν τον τύπο δυσλειτουργίας. Είναι προφανές ότι το μοντέλο προβλέπει πολλά “ψευδώς θετικά” αλλά έχει και πολλά “ψευδώς αρνητικά”. Δεδομένου του γεγονότος ότι είχαμε μικρό αριθμό επισημασμένων ανωμαλιών, αναμένεται να μην έχουμε σωστή απόκριση σε νέα-άγνωστα δεδομένα. Παρ’ ότι αυτό το μοντέλο θεωρείται ως αδύναμος ταξινομητής, είναι χρήσιμο το συνολικό τελικό μοντέλο.



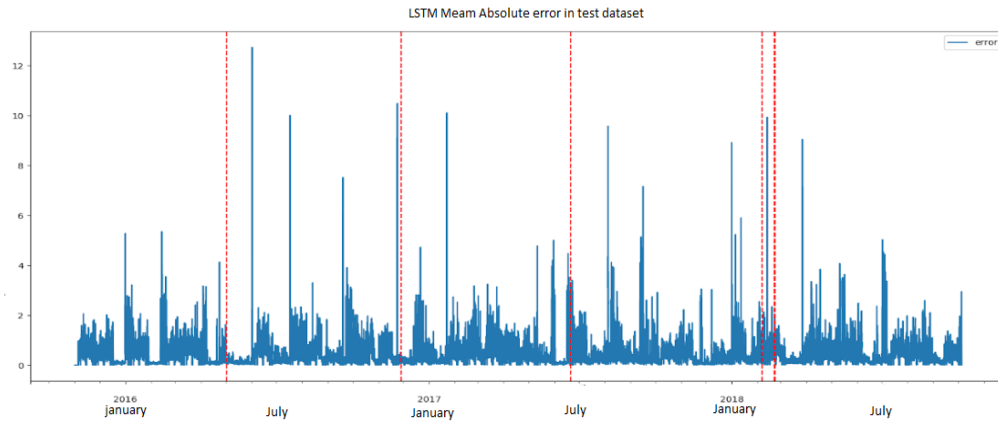
**Εικόνα 33** Οι μπλε κορυφές είναι οι χρονικές περίοδοι κατά τις οποίες εξελισσόταν ένα ελάττωμα του Crosshead Bearing. Η κόκκινη γραμμή είναι η προβλεπόμενη πιθανότητα ελαττώματος με βάση το μοντέλο XGboost.

Πολυμεταβλητό LSTM: Μετά από πολλές δοκιμές και ρυθμίσεις υπερπαραμέτρων, η αρχιτεκτονική του LSTM που χρησιμοποιήθηκε αποτελούνταν από 5 εσωτερικά (κρυφά) επίπεδα (hidden layers) [200, 500, 700, 500, 200], το παράθυρο μνήμης ήταν 100 ώρες χρησιμοποιώντας συνάρτηση ενεργοποίησης



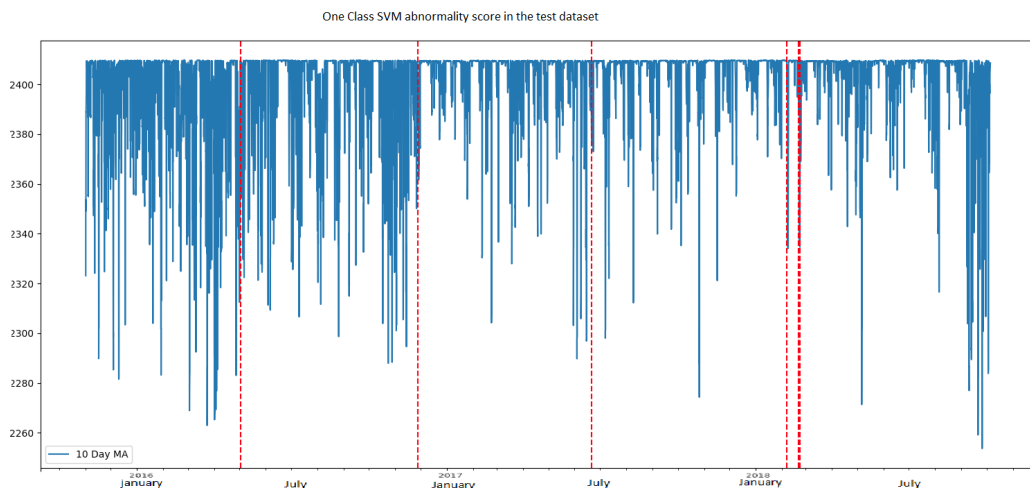
υπερβολικής εφαιπτομένης (tanh), ενώ το dropout (παράγοντας κατά του overfitting) ήταν 0.50 [230] μετά από κάθε επίπεδο. Δεν υπάρχει σαφής αιτιολόγηση γιατί η απόκρυψη ενός νευρώνα με 50% πιθανότητα είναι ο βέλτιστος στο πρόβλημά μας, αλλά είναι σύμφωνος με το [230] όπου το 0.5 ήταν η βέλτιστη τιμή για τα περισσότερα σύνολα δεδομένων, ενώ σε διάφορα πειράματα, παρόλα αυτά δοκιμάστηκαν και άλλες τιμές με χειρότερα αποτελέσματα. Στο [231] βρέθηκε ότι (τουλάχιστον για την γραμμική περίπτωση) το να χρησιμοποιήσουμε απόρριψη 50% έχει ως αποτέλεσμα το μέγιστο ποσό κανονικοποίησης. Ως κριτήριο τερματισμού της διαδικασίας εκπαίδευσης χρησιμοποιήθηκε η «πρώρη διακοπή» για να περιορίσει την εκπαίδευση σε μοντέλα και κανάλια που παρουσιάζουν μείωση στο σφάλμα αποτίμησης. Επίσης, λόγω της ύπαρξης θορύβου στα δεδομένα, παρατηρήσαμε ότι η αρχικοποίηση των βαρών, που είναι τυχαία στην περίπτωση μας, είχαν μεγάλη επίπτωση στην απόδοση. Έτσι, εκπαιδεύσαμε 10 μοντέλα, όπως προτείνει το [232], με διαφορετικό τυχαίο τρόπο και τα χρησιμοποιήσαμε ως μοντέλο Bagging. Μια πιθανή εξήγηση για την συμπεριφορά αυτή θα μπορούσε να είναι η ύπαρξη πολλών τοπικών ελαχίστων στην επιφάνεια απωλειών. Αυτό είναι ιδιαίτερα προβληματικό καθώς οι επιφάνειες απωλειών είναι γενικά μη κυρτές και μπορεί να παρουσιάζουν πολλά σημεία καμπής κάνοντας δύσκολο το να επιτευχθεί η σταθερότητα του μοντέλου.

Το σφάλμα που περιεγράφηκε στην παράγραφο 3.1 (σφάλμα της πρόβλεψης σε σύγκριση με τις πραγματικές τιμές) παρουσιάζεται στην Εικόνα 34. Υπάρχουν μερικές αιχμές στο διάγραμμα σφάλματος στην περιοχή πριν τον χρόνο βλάβης, που παρουσιάζονται με κόκκινες κατακόρυφες γραμμές, σημαίνοντας ότι ένας κατάλληλος ανιχνευτής μπορεί να εντοπίσει τα σήματα αυτά. Φυσικά υπάρχουν πολλά ψευδώς θετικά σήματα, με κυριότερο λόγο για αυτό να είναι το ότι άλλες βλάβες εκτός αυτές του ρουλεμάν σταυρού κεφαλής, μπορεί να έχουν συμβεί. Ο προτεινόμενος ανιχνευτής για αυτήν την προσέγγιση μπορεί να είναι το δυναμικό κατώφλι που βασίζεται στην τυπική απόκλιση του σφάλματος αυτού. Χρησιμοποιούμε ένα φίλτρο με 3.5 φορές την τυπική απόκλιση ως κατώφλι. Η τιμή εμπνεύστηκε από το [233], όπου 3 τυπικές αποκλίσεις από την μέση τιμή είναι μια συνηθισμένη αποκοπή για να εντοπισθούν ακραίες τιμές από μία Gaussian κατανομή ή παρόμοιά της. Για μικρότερα δείγματα δεδομένων, μια τιμή 2 SDs (95%) μπορεί να χρησιμοποιηθεί, ενώ για μεγαλύτερα δείγματα, μια τιμή 4 SDs (99.9%) μπορεί να χρησιμοποιηθεί. Διεξαγάγαμε πολλά πειράματα, αλλά τα 3.5 SDs έδωσαν τα καλύτερα αποτελέσματα. Αυτό σημαίνει ότι για κάθε ώρα προβλέψαμε τις επόμενες τιμές για κάθε αισθητήρα και μετά υπολογίσαμε το σφάλμα μεταξύ της προβλεπόμενης και της πραγματικής τιμής. Τότε, αν οι τιμές υπερβούν το συγκεκριμένο κατώφλι, το σύστημα ενεργοποιεί συναγερμό.



**Εικόνα 34** Μέσο απόλυτο σφάλμα του μοντέλου LSTM με χρήση δεδομένων δοκιμής, Μπλε γραμμή - Το MAE, Κόκκινες κάθετες γραμμές - ημερομηνίες ελαττωμάτων

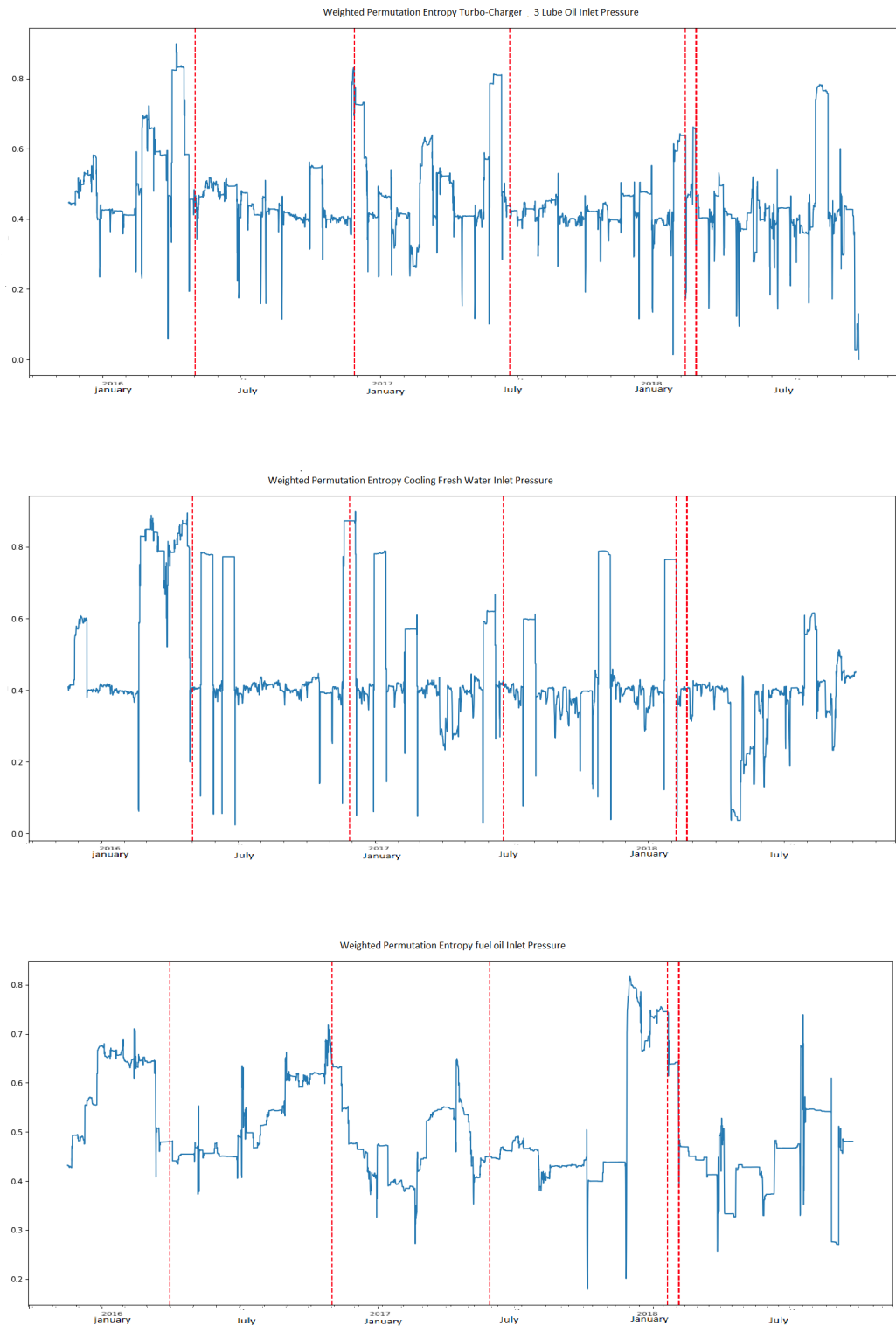
One Class SVM: Η Εικόνα 35 δείχνει τον βαθμό απόκλισης (απόστασης) μεταξύ κάθε σημείου δεδομένων και το συγκεκριμένου μοντέλου SVM (μπλε γραμμή). Στην περίπτωση μας αυτό σημαίνει ότι η χαμηλή τιμή της βαθμολογίας, συνεπάγεται σε υψηλότερη πιθανότητα ανωμαλίας. Είναι προφανές ότι υπάρχουν ορισμένες ημερομηνίες πριν από κάθε ελάττωμα (κόκκινες κάθετες γραμμές) με μικρή ομοιότητα με το εκπαιδευόμενο μοντέλο “κανονικής λειτουργίας”. Ο ανιχνευτής που χρησιμοποιείται για την προσέγγιση αυτή, χρησιμοποιεί σταθερό κατώφλι, πάνω από το οποίο υποθέτουμε ανώμαλη συμπεριφορά.



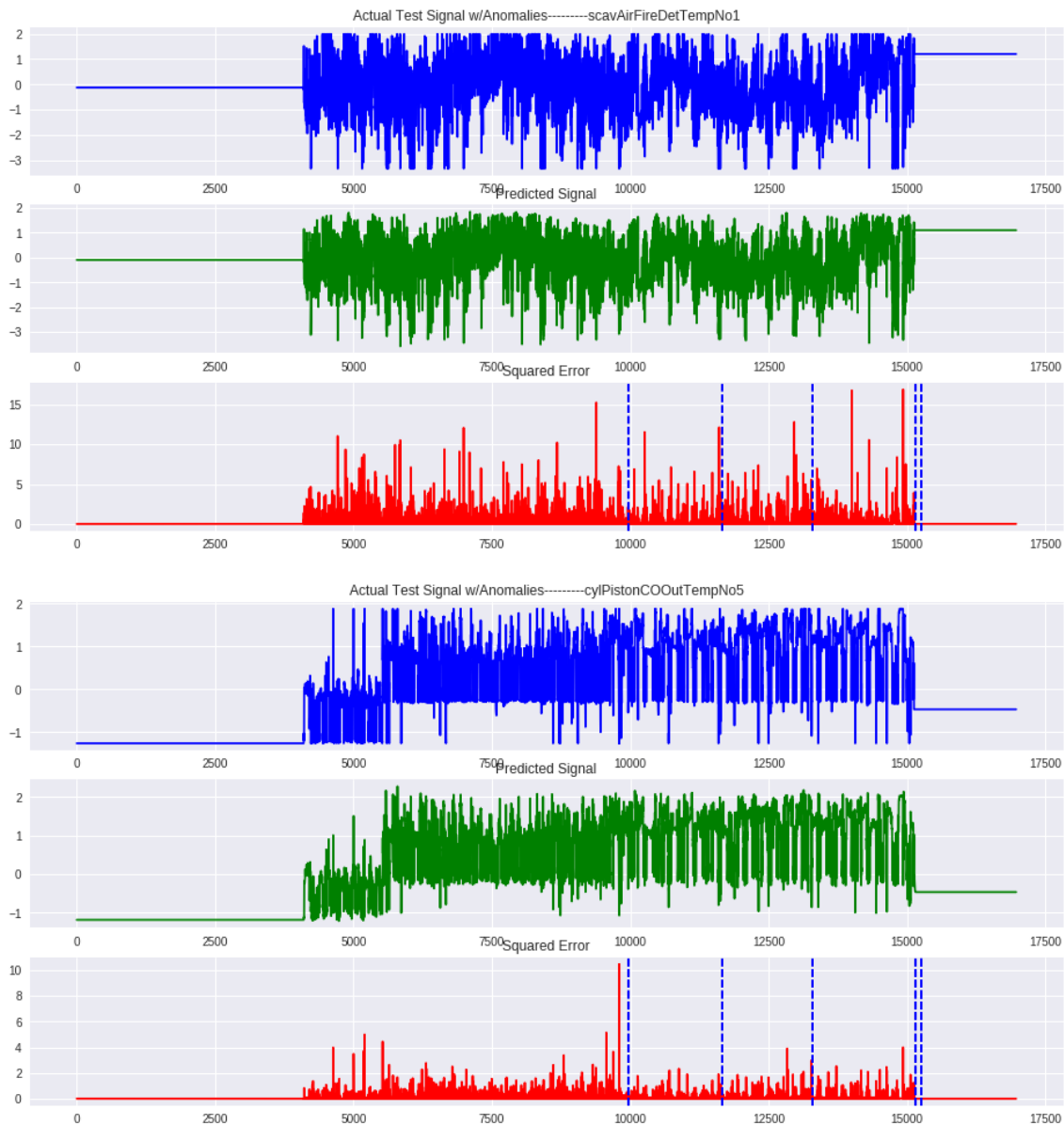
**Εικόνα 35** Η βαθμολογία ανωμαλίας που παράγεται από το μοντέλο OCSVM χρησιμοποιώντας σύνολο δεδομένων δοκιμής, Μπλε γραμμή - η βαθμολογία ανωμαλίας, Κόκκινες κάθετες γραμμές - ημερομηνίες ελαττωμάτων



Weighted Permutation Entropy (WPE): Όπως αναφέρθηκε στο προηγούμενο κεφάλαιο, η αύξηση της εντροπίας στην συγκεκριμένη περίπτωση είναι μία ένδειξη μη ομαλής λειτουργίας. Για τον υπολογισμό της WPE χρησιμοποιήσαμε ως παράμετρο  $D=3$  και κυλιόμενο παράθυρο ίσο με 600 χρονικές στιγμές. Οι τιμές αυτές ήταν αποτέλεσμα της αναζήτησης πλέγματος και χρησιμοποιήθηκαν για την ανίχνευση ανωμαλιών με την βοήθεια του WPE. Τα πειράματά μας έδειξαν ότι υπάρχουν κάποιες ροές δεδομένων που μπορούν να συλλάβουν όλα τα ελαττώματα. Οι προαναφερθείσες παράμετροι είναι οι "Cooling Fresh Water Inlet Pressure" (πίεση εισόδου ψυκτικού φρέσκου ύδατος), "Turbo-Chargerr #3 Lube Oil Inlet Pressure" και "fuel oil Inlet Pressure" ("πίεση εισόδου turbo #3 πετρελαίου φόρτισης" και "πίεση εισόδου καυσίμου πετρελαίου"). Τα διαγράμματα για τα χαρακτηριστικά αυτά βρίσκονται στην Εικόνα 36. Ο προτεινόμενος ανιχνευτής που βασίζεται στην WPE έχει σταθερό κατώφλι. Αυτό σημαίνει πρέπει να υπολογίζουμε διαρκώς την εντροπία μετάθεσης και όταν οι τιμές αυτές υπερέρβαιναν το κατώφλι, το σύστημα ενεργοποιεί την ένδειξη βλάβης.



**Εικόνα 36** Τιμή εντροπίας κυλιόμενης σταθμισμένης μετάθεσης (α) Πίεση εισόδου γλυκού νερού ψύξης (β) Στροβιλο-συμπιεστής #3 Πίεση εισόδου λαδιού Lude (γ) Πίεση εισόδου λαδιού καυσίμου. Με ανάγνωση κάθετης γραμμής απεικονίζονται οι ημερομηνίες ελαττωμάτων



**Εικόνα 37** Μονομεταβλητά διαγράμματα LSMT της θερμοκρασίας ανίχνευσης αέρα καθαρισμού και της θερμοκρασίας εξόδου ψύξης εμβόλου (οC) – μπλε: πραγματικές τιμές, πράσινο: προβλεπόμενες τιμές, κόκκινο: το τετράγωνο σφάλμα μεταξύ προβλεπόμενων και πραγματικών τιμών.

Μονομεταβλητό-LSTM : Ακολουθώντας ρυθμίσεις υπερπαραμέτρων, οι αρχιτεκτονικές του LSTMs που χρησιμοποιήθηκαν, αποτελούνταν από 5 εσωτερικά επίπεδα. Το παράθυρο μνήμης ήταν 100 ώρες, χρησιμοποιώντας tanh για την συνάρτηση ενεργοποίησης και η τιμή dropout (0.50) μετά από κάθε επίπεδο, όπως υλοποιήθηκε και στο πολυμεταβλητό μοντέλο LSTMs. Η μεθοδολογία ήταν όμοια με αυτή που ακολουθήθηκε για πολυμεταβλητά LSTM σχετικά με τον τερματισμό εκπαίδευσης, το σύστημα προειδοποίησης και την αρχικοποίηση βαρών, αφού η



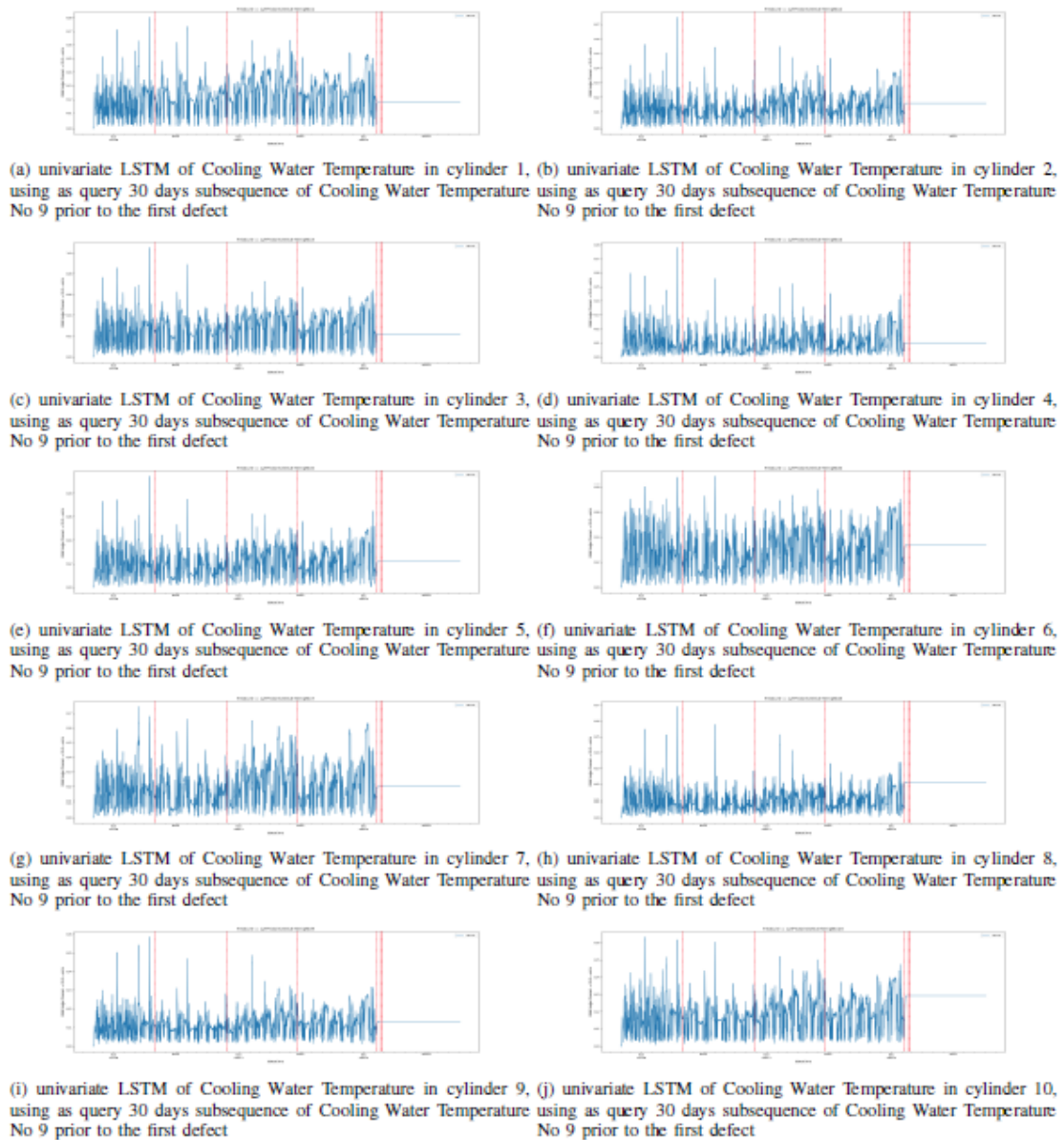


διαδικασία είναι η ίδια όσον αφορά στη ρύθμιση των παραμέτρων, με μόνη διαφορά τα δεδομένα εισόδου.

Η Εικόνα 37 απεικονίζει τα αποτελέσματα της προσέγγισης αυτής, με τις μπλε γραμμές να υποδηλώνουν τις πραγματικές τιμές κάθε καναλιού, τις πράσινες γραμμές να υποδηλώνουν τις τιμές πρόβλεψης και τις κόκκινες γραμμές να υποδεικνύουν το τετραγωνικό σφάλμα μεταξύ των πραγματικών τιμών και των τιμών πρόβλεψης. Ακόμα, με τις μπλε κατακόρυφες γραμμές στο γράφημα των σφαλμάτων, υποδεικνύεται ο χρόνος του ελαττώματος. Η υπόθεση μας δηλώνει ότι το σφάλμα πριν το χρόνο ελαττώματος πρέπει να είναι υψηλότερο από τις υπόλοιπες υπο-ακολουθίες, σε έναν συγκεκριμένο κύλινδρο, που καταστράφηκε εκείνη την στιγμή. Μετά από αρκετούς πειραματισμούς καταλήξαμε στο συμπέρασμα ότι οι μετρήσεις που παρακολουθήσαμε σε κάθε κύλινδρο έχουν σχεδόν ίδια συμπεριφορά ακόμα και πριν από μία συγκεκριμένη δυσλειτουργία ενός μοναδικού κυλίνδρου. Στην Εικόνα 38 παρουσιάζεται η πρόβλεψη του σφάλματος του κάθε κυλίνδρου/πιστονιού συναρτήσει της θερμοκρασίας ψύξεως εξόδου ( $^{\circ}\text{C}$ ). Οι κόκκινες κατακόρυφες γραμμές υποδηλώνουν ελάττωμα στο ρουλεμάν σταυρού κεφαλής του κυλίνδρου N<sup>ο</sup> 9-2-4-8-4. Είναι προφανές ότι οι κορυφές του σφάλματος πρόβλεψης είναι στις ίδιες χρονικές στιγμές στους περισσότερους κυλίνδρους, κάτι που σημαίνει ότι δεν γνωρίζουν για την υπό διερεύνηση βλάβη.



APPENDIX  
FIGURES

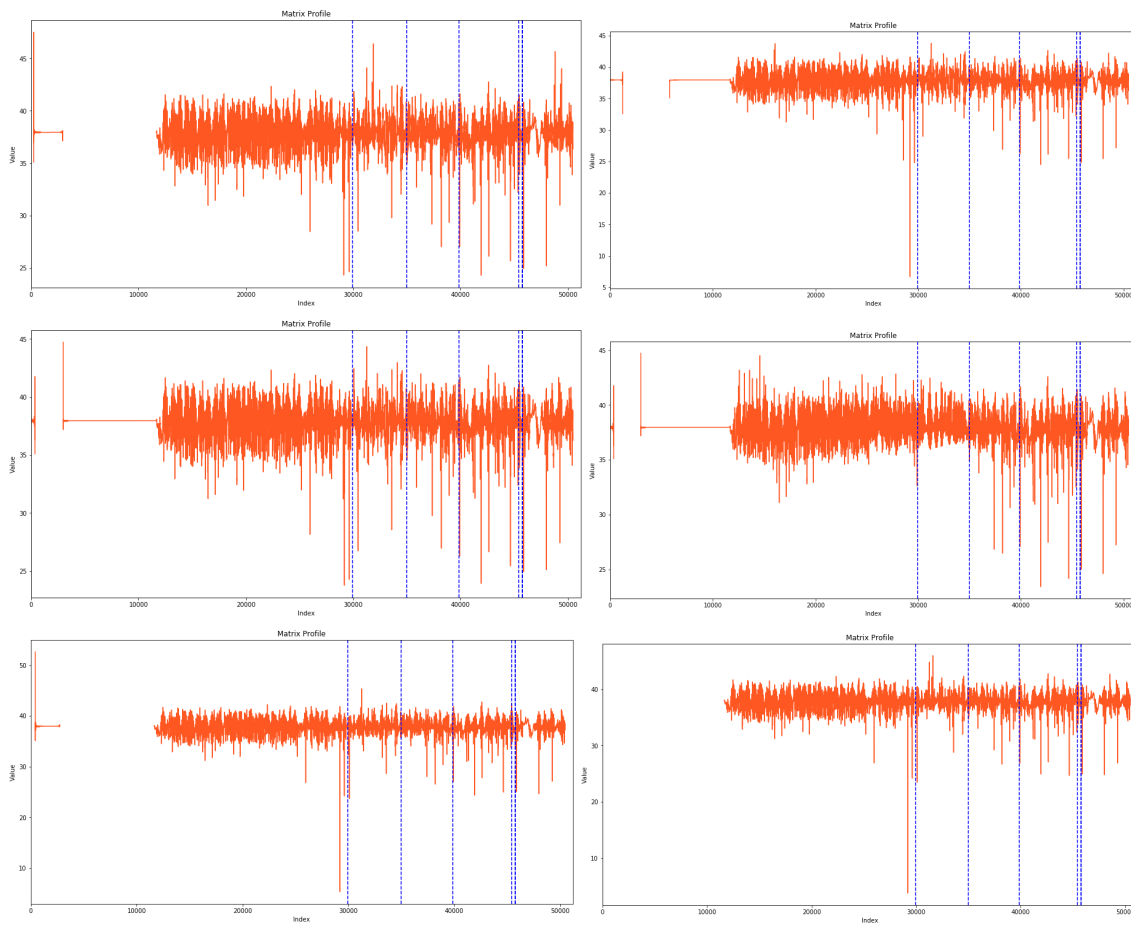


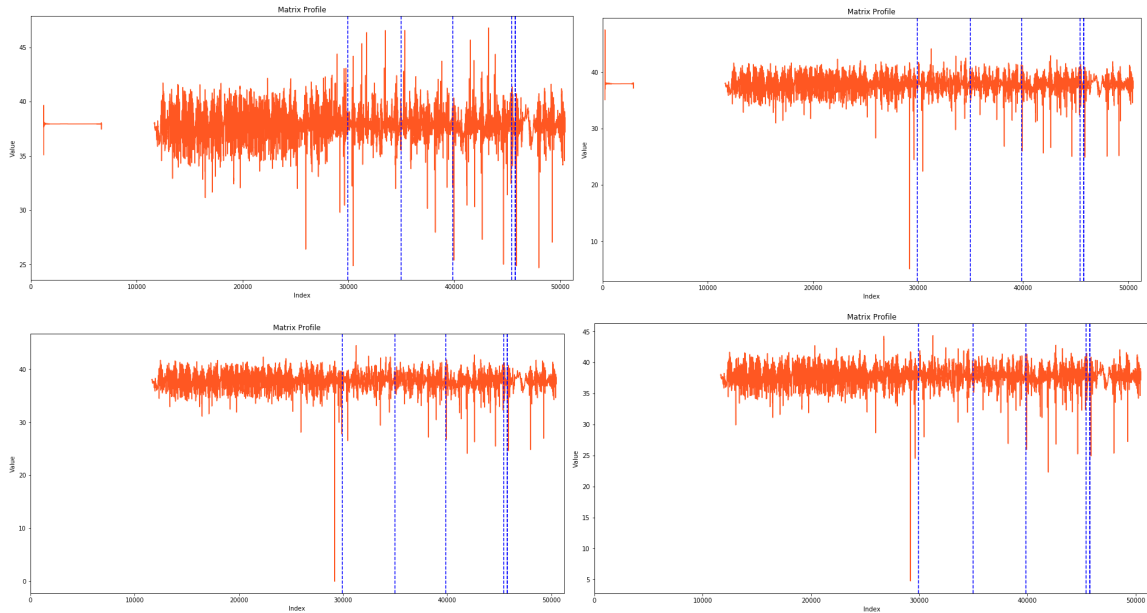
**Εικόνα 38** Μονομεταβλητή LSTM θερμοκρασίας νερού ψύξης σε κάθε κύλινδρο 1-10, χρησιμοποιώντας ως ερώτημα 30 ημέρες διαδοχική της Θερμοκρασίας Νερού Ψύξης No 9 πριν από το πρώτο ελάττωμα

**Matrix Profile (MP):** Τα διαγράμματα στην Εικόνα 39 απεικονίζουν τα διαγράμματα του MP για μια συγκεκριμένη περίοδο πριν το χρόνο της πρώτης βλάβης. Ουσιαστικά χρησιμοποιήσαμε την μεθοδολογία με σκοπό την ανίχνευσης επαναλαμβανόμενου μοτίβου. Συγκεκριμένα, τα διαγράμματα δείχνουν το MP της θερμοκρασίας ύδατος ψύξης σε κάθε κύλινδρο (1-10), για υπο-ακολουθία 30 ημερών διαδοχικά και την θερμοκρασία ύδατος ψύξης του N° 9 πριν την πρώτη βλάβη. Το σκεπτικό ήταν ακριβώς το γεγονός ότι η πρώτη βλάβη συνέβη στον

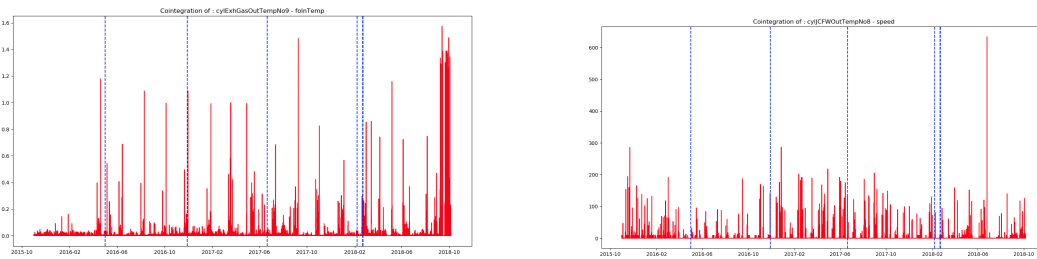


κύλινδρο 9. Οι μπλε κατακόρυφες γραμμές υποδηλώνουν ελάττωμα στο ρουλεμάν σταυρού κεφαλής του κυλίνδρου Ν° 9-2-4-8-4. Παρατηρώντας τα διαγράμματα δεν είναι δυνατό να συγκλίνουμε σε κάποιο συμπέρασμα για το ποιος κύλινδρος είναι πιο πιθανόν να δυσλειτουργεί. Είναι παραπάνω από εμφανές ότι κανένα από τα παραπάνω διαγράμματα δεν δίνει οποιαδήποτε πληροφορία για το ποιος κύλινδρος είναι πιο πιθανόν να έχει αρχίσει να φθείρεται. Όταν η εποχικότητα δεν έχει αρθεί, οι μικρότερες τιμές του MP προήλθαν από την περιοδική λειτουργία και όχι από δυσλειτουργία.



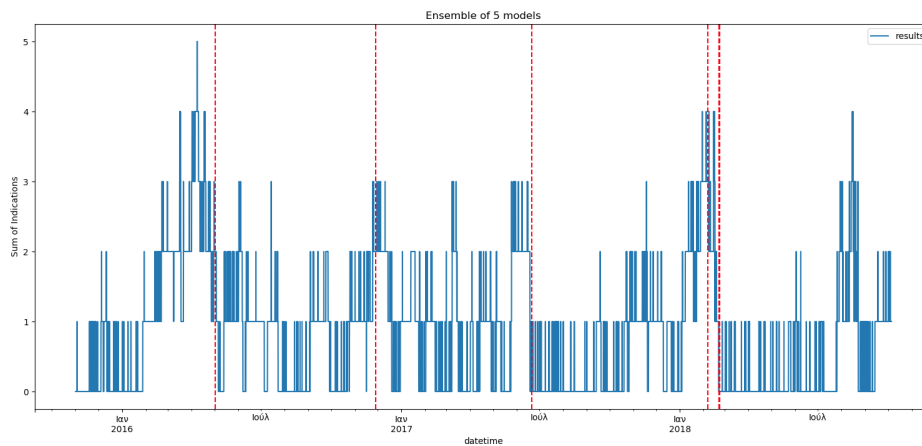


**Εικόνα 39** MP Θερμοκρασίας Νερού Ψύξης σε κάθε κύλινδρο 1-10, χρησιμοποιώντας ως ερώτημα 30 ημέρες διαδοχικά της Θερμοκρασίας Νερού Ψύξης Νο 9 πριν από το αρχική βλάβη



**Εικόνα 40** Διαγράμματα συνολοκλήρωσης καυσαερίων από τους κυλίνδρους 7 και 8

Ανάλυση συνολοκλήρωσης (Cointegration Analysis): Η Εικόνα 40 δίνει κάποια ενδεικτικά διαγράμματα συνολοκλήρωσης. Η πιο ενδιαφέρουσα πληροφορία είναι ότι καμία από τις “εκρήξεις” στα διαγράμματα δεν οφείλεται σε βλάβη του ρουλεμάν σταυρού κεφαλής. Είναι προφανές ότι τα περισσότερα είναι πανομοιότυπα, ακόμα και όταν διαφορετικοί κύλινδροι εξετάζονται. Δεδομένου του ότι η ανάλυση συνολοκλήρωσης χρησιμοποιεί τους συγκεκριμένους ειδικούς αισθητήρες κυλίνδρων, μπορούμε να απορρίψουμε την υπόθεση μας ότι η εν λόγω μέθοδος είναι ενδεικτική και δύναται να εντοπίσει πιθανές βλάβες.



**Εικόνα 41** Αποτέλεσμα συνδυαστικού μοντέλου (μπλε χρώμα) , πραγματικές βλάβες (κόκκινο χρώμα) Πιθανότητα μη φυσιολογικής κατάστασης του κύριου κινητήρα

---

**Algorithm 2: INFERENCE process**

---

```
1 models = [XgBoost, mLSTM, OCSVM, WPE]
2 Impute missing data X
3 Clip the data X (remove extreme values)
4 set (int)alert = 0
5 for model in models do
6   if model=XgBoost then
7     result = model.predict(X)
8     if resultz anomaly threshold then
9       | alert+ = 1
10    end
11  end
12  if model=mLSTM then
13    set X = PCA(X)
14    result = model.predict(X)
15    if result > anomalythreshold then
16      | set alert+ = 1
17    end
18  end
19  if model=OCSVM then
20    set X = pct(X)
21    result = model.predict(X)
22    if result > anomalythreshold then
23      | set alert+ = 1
24    end
25  end
26  if model=WPE then
27    result = calculate rolling WPE(X)
28    if result > anomalythreshold then
29      | set alert+ = 1
30    end
31  end
32  if alert > 3 then
33    | raise alert
34  end
35 end
```

---

Έχοντας μελετήσει όλες τις μεθόδους που αναφέρθηκαν παραπάνω, και εξετάσαμε εάν ο συνδυασμός μη ισχυρών ταξινομητών είναι πιο σταθερός από κάθε ένα ξεχωριστά. Χρησιμοποιήσαμε ένα μη σταθμισμένο άθροισμα αυτών, που σημαίνει ότι όταν όλα τα μοντέλα συμφωνούν ότι υπάρχει ανωμαλία (βλάβη) και το συνδυαστικό μοντέλο εμφανίζει την ένδειξη 6, που επιβεβαιώνει την προτεινόμενη



προσέγγιση. Χρησιμοποιήσαμε ως κατώφλι τον αριθμό των μισών μοντέλων (δηλ. 3) ώστε να ενεργοποιηθεί η προειδοποίηση στο σύστημα. Το όριο θα μπορούσε πιθανά να ρυθμιστεί ανάλογα με τις επιχειρηματικές απαιτήσεις και την στρατηγική της κάθε εταιρίας. Αυτή η μέθοδος μπορεί να δοθεί περιληπτικά στον ανωτέρω αλγόριθμο. Ενώ όπως φαίνεται από την Εικόνα 41 και τους Πίνακα 16 και Πίνακα 17, ο συνδυασμός είναι πιο σταθερός από ότι το κάθε μοντέλο χωριστά, ενώ τα καλύτερα αποτελέσματα σχετικά με τις τιμές F-1 επιτυγχάνονται χρησιμοποιώντας το άθροισμα των προαναφερθέντων μοντέλων, εκτός του μοντέλου LSTM.

**Πίνακας 16** Συγκεντρωτικά Αποτελέσματα όλων των μοντέλων και ο συνδυασμός τους για το πρόβλημα προβλεπτικής συντήρησης κύριων μηχανών πλοίων.

<i>Model</i>	<i>Precision</i>	<i>Recall</i>	<i>f1-score</i>
<i>Gradient Boosting</i>	0,39	0,08	0,13
<i>Multi-LSTM</i>	0,22	0,03	0,05
<i>One Class SVM</i>	0,22	0,14	0,17
<i>WPE - 1</i>	0,33	0,61	0,43
<i>WPE - 2</i>	0,47	0,53	0,5
<i>WPE - 3</i>	0,39	0,62	0,48
<i>Ensemble</i>	0,8	0,21	0,33
<i>Ensemble (no LSTM)</i>	0,5	0,7	0,58

**Πίνακας 17** Αποτελέσματα συνδιαστικού μοντέλου για πρόβλημα προβλεπτικής συντήρησης κύριων μηχανών πλοίων (Confusion Matrix).

	<i>Predicted Negative</i>	<i>Predicted Positive</i>
<i>Actually Negative</i>	31712	382
<i>Actually Positive</i>	2996	248

## 4.2 Μηχανική Μάθηση & ασφάλεια των τροφίμων.

Σε αυτή την ενότητα αρχικά παρέχουμε μια σύντομη περιγραφή της στρατηγικής αξιολόγησης σχετικά με τις προσεγγίσεις που εισάγονται σε αυτή την εργασία. Υπενθυμίζεται ότι στόχος του προτεινόμενου συστήματος είναι η ταξινόμηση και η πρόβλεψη χρονοσειρών για την ανακλήσεων τροφίμων. Στη συνέχεια παρουσιάζονται τα αποτελέσματα κάθε προσέγγισης αντίστοιχα.



### 4.2.1 Εργασία ταξινόμησης – Named Entity Recognition

Η ακρίβεια (accuracy), η ανάκληση (recall) και η βαθμολογία F-1 (F-1 score) σε συμβολικό επίπεδο είναι τα πιο συνηθισμένα μέτρα αξιολόγησης για εργασίες Αναγνώρισης Ονοματισμένων Οντοτήτων (Named Entity Recognition - NER) [234]. Ωστόσο, στην εξεταζόμενη περίπτωση χρήσης, είναι πιο χρήσιμο να γίνει αξιολόγηση σε επίπεδο πλήρους επωνυμίας της οντότητας. Υπάρχουν ορισμένα μέτρα που είναι κατάλληλα για αυτήν την προσέγγιση, καθώς μετρούν την πλήρη απόδοση του διακριτικού της ονοματισμένης οντότητας που παρουσιάζεται [235]. Ακολουθώντας αυτή τη στρατηγική, οι μετρήσεις μπορούν να καθοριστούν από την άποψη της σύγκρισης της απόκρισης ενός συστήματος έναντι της 'χρυσής ετικέτας' (gold standard) ως εξής:

- **Σωστό (COR):** και τα δύο είναι ίδια.
- **Λάθος (INC):** η έξοδος του συστήματος και η χρυσή ετικέτα δεν ταιριάζουν.
- **Μερικό (PAR):** το σύστημα και η χρυσή ετικέτα είναι κάπως «παρόμοιοι», αλλά όχι ίδια.
- **Ελλιπές (MIS):** δεν καταγράφεται χρυσή ετικέτα από το σύστημα.
- **Πλαστό (SPU):** η έξοδος του συστήματος δεν υπάρχει στη χρυσή ετικέτα αναφοράς.

Για να καλύψουμε τα σενάρια που ορίζονται στο πρόβλημά μας, πρέπει να εξετάσουμε τις διαφορές μεταξύ της εξόδου του NER και των 'χρυσών ετικετών' (gold labels) βασιζόμενοι σε δύο άξονες: τη συμβολοσειρά της οντότητας και τον τύπο της οντότητας. Μια άλλη ισοδύναμη προσέγγιση παρουσιάστηκε στο Διεθνές Εργαστήριο Σημαντικής Αξιολόγησης (International Workshop on Semantic Evaluation - SemEval 13) [236], όπου εισήγαγε τέσσερις τρόπους μέτρησης των αποτελεσμάτων της ακρίβειας/ανάκλησης/F1-βαθμολογίας με βάση τις μετρήσεις που ορίζονται από το MUC.

- **Αυστηρή:** ακριβής αντιστοίχιση συμβολοσειράς επιφάνειας ορίων και τύπος οντότητας,
- **Ακριβής:** ακριβής αντιστοίχιση ορίων πάνω από την επιφάνεια, ανεξάρτητα από τον τύπο,
- **Μερική:** μερική αντιστοίχιση ορίων πάνω από την επιφάνεια, ανεξάρτητα από τον τύπο,
- **Τύπος:** απαιτείται κάποια επικάλυψη μεταξύ της οντότητας με ετικέτα συστήματος και του χρυσού σχολιασμού;



Επιπλέον, για να οριστεί η ακρίβεια/ανάκληση/F1-score, πρέπει να υπολογιστούν δύο ακόμη μεταβλητές. Ο αριθμός των σχολιασμένων 'χρυσών ετικετών' που συμβάλλουν στην τελική βαθμολογία παρέχεται ως:

$$POSSIBLE(POS) = COR + INC + PAR + MIS = TR + FN \quad (2)$$

Και ο αριθμός των σχολίων που παράγονται από το σύστημα:

$$ACTUAL(ACT) = COR + INC + PAR + SPU = TR + FP \quad (3)$$

Συνοπτικά, η ακρίβεια είναι το ποσοστό των σωστά εντοπισμένων ονοματικών οντοτήτων από το σύστημα NER, ενώ η ανάκληση αναφέρεται στο ποσοστό των ονοματικών οντοτήτων στις 'χρυσές ετικέτες' που ανακτώνται από το σύστημα NER. Αυτά υπολογίζονται είτε για σενάριο ακριβούς αντιστοίχισης (δηλαδή αυστηρό και ακριβές), είτε για σενάριο μερικής αντιστοίχισης (δηλαδή μερικό και βάσει τύπου) με τις ακόλουθες εξισώσεις:

Ακριβής αντιστοίχιση (δηλαδή αυστηρή και ακριβής):

$$Precision = \frac{COR}{ACT} = \frac{TP}{TP+FP} \quad (4)$$

$$Recall = \frac{COR}{POS} = \frac{TP}{TP+FN} \quad (5)$$

Ακριβής αντιστοίχιση (δηλαδή μερική και τύπου):

$$Precision = \frac{COR+0.5*PAR}{ACT} = \frac{TP}{TP+FP} \quad (6)$$

$$Recall = \frac{COR+0.5*PAR}{POS} = \frac{COR}{ACT} \frac{TP}{TP+FN} \quad (7)$$

Ο Πίνακας 18 παρουσιάζει μερικά παραδείγματα από το χρησιμοποιούμενο σύνολο δεδομένων, εξηγώντας πώς μπορούν να αξιολογηθούν. Στη συνέχεια, υπολογίζεται η ακρίβεια/ανάκληση/F1-βαθμολογία για κάθε διαφορετικό σχήμα αξιολόγησης που εφαρμόζεται σε ένα τετραπλά στρωματοποιημένο σχήμα αξιολόγησης των αποτελεσμάτων μας, ενώ τα μέσα αποτελέσματα αναφορικά με τη στρατηγική αξιολόγησης παρουσιάζονται στον Πίνακας 19.





Πίνακας 18 Σχέδιο αξιολόγησης με βάση την σημασιολογική αξιολόγηση

<i>Golden Standard</i>		<i>System Prediction</i>			<i>Evaluation</i>		
Surface String	Entity Type	Surface String	Entity Type	Type	Partial	Exact	Strict
Organic Flour	PRODUCT	Organic Flour	PRODUCT	COR	COR	COR	COR
-	-	Organic Flour	PRODUCT	SPU	SPU	SPU	SPU
Organic Flour	PRODUCT	Organic Flour	PRODUCT	COR	COR	COR	COR
Organic Flour	PRODUCT	-	-	MIS	MIS	MIS	MIS
Organic Flour	PRODUCT	Tuna	PRODUCT	COR	INC	INC	INC
Organic Flour	PRODUCT	Flour	PRODUCT	COR	PAR	PAR	INC

Πίνακας 19 Αποτελέσματα του συστήματος ταξινόμησης ως προς τις μετρήσεις αξιολόγησης

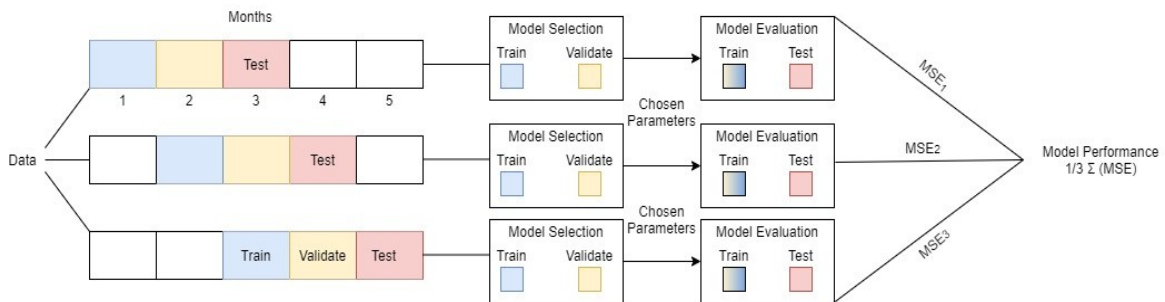
	<i>Type</i>	<i>Partial</i>	<i>Exact</i>	<i>Strict</i>
<i>Correct</i>	12715	12534	12534	12534
<i>Incorrect</i>	0	7.75	7.75	181.5
<i>Partial</i>	0	173.5	173.5	0
<i>Missed</i>	654	654.5	654.5	654.5
<i>Spurious</i>	561	561	561	561
<i>Possible</i>	12270	13370	13370	13370
<i>Actual</i>	13276	13276.5	13276.5	13276
<i>Precision</i>	0.957877	0.944245	0.9442455	0.9442455
<i>Recall</i>	0.951183	0.937578	0.9375785	0.943409

#### 4.2.2 Πρόβλεψη χρονοσειρών

Τα συνολικά δεδομένα εκπαίδευσης αποτελούνται από έναν πίνακα  $N \in \mathbb{R}^{30 \times 147}$ , όπου το «30» αντιστοιχεί στον αριθμό των χρονοσειρών προϊόντων και το «147» στα χρονικά βήματα των τελευταίων 147 μηνών. Προκειμένου να έχουμε πιο αξιόπιστα αποτελέσματα, εκπαιδεύσαμε κάθε μοντέλο 47 φορές σε ένα κυλιόμενο παράθυρο 100 χρονικών βημάτων που σχηματίζουν 47 πίνακες  $\in \mathbb{R}^{30 \times 100}$  (π.χ. [30,1-100], [30,2-101] κ.λπ.). Ενώ όλα τα μοντέλα μοιράζονται ορισμένες κοινές υπερπαραμέτρους όπως ο αριθμός των εποχών  $\leftarrow 200$ , ο ορίζοντας πρόβλεψης  $\leftarrow 4$  και το μέγεθος του ιστορικού πλαισίου  $\leftarrow 12$  μήνες, μερικές άλλες είναι συγκεκριμένες μόνο για ορισμένα μοντέλα. Αυτές οι συγκεκριμένες παράμετροι ρυθμίστηκαν με ακρίβεια χρησιμοποιώντας τεχνικές αναζήτησης πλέγματος, ως θεμελιώδες βήμα του μηχανικής μάθησης. Οι αλγόριθμοι αναζήτησης πλέγματος και τυχαίας αναζήτησης χρησιμοποιούνται ευρέως για συντονισμό υπερπαραμέτρων. Βασικά, το πεδίο των υπερπαραμέτρων τοποθετείται σε ένα διακριτό πλέγμα. Στη συνέχεια, κάθε συνδυασμός τιμών αυτού του πλέγματος (δηλ.

πλέγμα-αναζήτηση) χρησιμοποιείται ως προς τη χρήση διασταυρούμενης επικύρωσης. Το σημείο του πλέγματος που μεγιστοποιεί τη μέση τιμή στη διασταυρούμενη επικύρωση, είναι το βέλτιστο σύνολο τιμών για τις υπερπαραμέτρους.

Στην συνηθισμένη εφαρμογή μεθόδων μηχανικής μάθησης, οι μέθοδοι διασταυρούμενης επικύρωσης ενέχουν μια παγίδα που αφορά προσεγγίσεις πρόβλεψης χρονοσειρών, καθώς μπορεί να οδηγήσουν σε σημαντική επικάλυψη μεταξύ των δεδομένων εκπαίδευσης και ελέγχου. Έτσι, η βέλτιστη προσέγγιση είναι η προσομοίωση μοντέλων σε μια ακολουθία «βημάτων προς τα εμπρός», εκπαιδεύοντας περιοδικά το μοντέλο, ώστε να ενσωματώνει συγκεκριμένα κομμάτια δεδομένων διαθέσιμα σε εκείνη τη χρονική στιγμή. Αυτή η διαδικασία απεικονίζεται στην Εικόνα 39, ενώ όσον αφορά τη διαδικασία επικυρώσεων που ακολουθείται στην περίπτωση του μοντέλου RL, έχει εφαρμοστεί ο Αλγόριθμος 2.



Εικόνα 42 Προτεινόμενο σχέδιο επικύρωσης βημάτων προς τα εμπρός

---

### Algorithm 2: DEEP Q-LEARNING evaluation process

---

```
1 for validationfold = 1, Validationfolds do
2   Initialize the custom time-series environment and get the state s regarding
   validationfold
3   Load the pre-trained DQN agent as agent regarding validationfold
4   initialize list predictionsgets[]
5   for step in lengthofstepsahead do
6     prediction = agent.predict(s)
7     state = state[1:].append(prediction)
8     predictions.append(prediction)
9   end
10  evaluate(γtrue, predictions)
11 end
```

---



Όσον αφορά την αξιολόγηση των αποτελεσμάτων, παρατηρήσαμε τα ακόλουθα αποτελέσματα ως προς το Μέσο Τετραγωνικό Σφάλμα (Mean Squared Error (MSE))

α. αποκλείστηκα με τη χρήση των αρχικών δεδομένων (Πίνακας 20),

β. με το εμπλουτισμένο σύνολο δεδομένων με χρήση υποκατάστατων δεδομένων (Πίνακας 21).

Επιπλέον, τα σχετικά διαγράμματα απεικονίζονται στην Εικόνα 43, που δείχνουν τους εκτιμητές σε συγκεκριμένες χρονοσειρές.

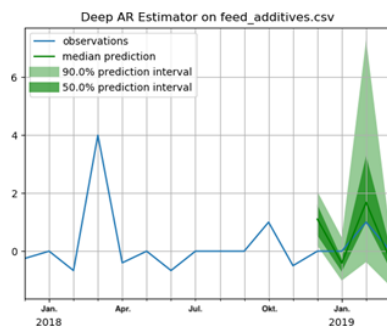
**Πίνακας 20** Σύγκριση μεταξύ μονομεταβλητής προσέγγισης, αύξησης δεδομένων και Ενισχυτικής Μάθησης

	Univariate				Univariate with Data Surrogation				RL
	Deep AR	Simple FF	Seasonal	WaveNet	Deep AR	Simple FF	Seasonal	WaveNet	
0	0.922	1.895	3.268	2.175	1.74	1.785	3.267	1.644	1.752
1	0.196	0.442	0.787	0.697	0.073	0.068	0.106	0.07	0.104
2	0.037	0.081	0.144	0.131	0.218	0.217	0.463	0.253	0.171
3	0.453	1.307	2.765	1.178	1.355	1.58	2.786	1.414	0.238
4	0.367	0.969	1.688	1.432	4.011	3.908	8.041	3.701	1.256
5	0.076	0.207	0.244	0.234	1.067	1.173	1.202	1.083	1.011
6	6.021	8.377	8.39	8.093	1.321	1.364	2.429	1.181	0.309
7	0.539	1.841	3.046	1.511	1.019	1.095	2.221	0.955	0.282
8	0.228	0.782	1.374	1.235	5.248	5.477	6.241	5.235	2.52
9	0.284	0.93	1.811	0.892	0.944	0.961	1.567	0.776	0.055
10	0.138	0.376	0.557	0.509	0.558	0.403	0.787	0.425	1.144
11	0.03	0.085	0.106	0.112	0.06	0.082	0.144	0.083	0.149
12	0.058	0.212	0.236	0.215	1.138	1.141	2.763	1.071	0.683
13	1.374	3.603	5.399	3.044	0.637	0.644	1.692	0.609	0.105
14	0.108	0.22	0.394	0.329	0.168	0.199	0.244	0.183	0.648
15	2.305	7.129	8.789	6.067	8.085	8.125	8.386	7.932	14.329
16	0.459	0.985	1.609	0.953	1.429	1.469	3.045	1.269	0.586
17	0.13	0.531	1.026	0.524	0.68	0.68	1.374	0.616	0.357
18	0.108	0.228	0.625	0.317	0.748	0.824	1.811	0.783	0.874
19	0.375	0.836	1.179	1.221	0.365	0.371	0.557	0.355	0.213
20	0.513	1.76	2.286	2.644	0.133	0.143	0.236	0.144	0.083
21	2.285	11.554	23.353	9.327	2.575	2.748	5.396	2.606	3.068
22	0.086	0.277	0.462	0.315	0.256	0.211	0.394	0.225	0.474
23	0.581	1.889	2.786	1.858	6.081	6.188	8.785	5.881	N/A
24	1.575	4.22	8.045	3.744	0.975	0.852	1.609	0.742	0.593
25	0.348	1.389	1.203	2.595	0.468	0.479	1.027	0.46	0.084
26	0.615	1.404	2.429	2.345	0.255	0.256	0.625	0.277	0.232
27	0.314	1.087	2.22	1.036	0.804	0.814	1.181	0.762	0.053
28	2.129	5.206	6.244	4.68	1.599	1.735	2.287	1.554	0.622
29	0.313	1.049	1.568	1.02	5.813	6.386	23.423	5.751	9.747

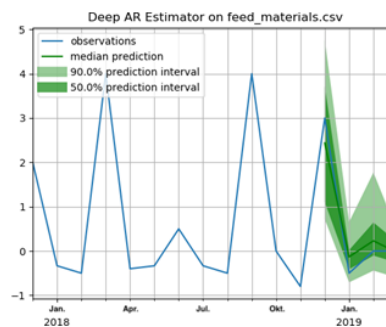


Πίνακας 21 Σύγκριση προσέγγισης μονομεταβλητών μοντέλων, πολυμεταβλητών μοντέλων και Ενιασχυτικής Μάθησης

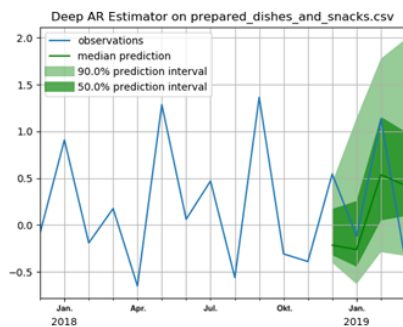
	Univariate				Multivariate					RL
	Deep AR	Simple FF	Seasonal	WaveNet	Deep AR	Simple FF	Deep Factor	Seasonal	WaveNet	
0	0.922	1.895	3.268	2.175	<b>0.154</b>	0.198	1.777	0.244	0.187	1.752
1	<b>0.196</b>	0.442	0.787	0.697	1.919	1.851	3.266	3.267	1.736	0.104
2	<b>0.037</b>	0.081	0.144	0.131	5.04	6.678	7.181	23.423	5.681	0.171
3	0.453	1.307	2.765	1.178	0.817	0.959	2.406	1.567	0.862	<b>0.238</b>
4	<b>0.367</b>	0.969	1.688	1.432	1.55	1.755	2.806	2.287	1.534	1.256
5	<b>0.076</b>	0.207	0.244	0.234	0.233	0.213	1.805	0.394	0.270	1.011
6	6.021	8.377	8.39	8.093	1.514	1.582	3.155	2.786	1.636	<b>0.309</b>
7	0.539	1.841	3.046	1.511	0.951	1.196	2.748	1.202	1.195	<b>0.282</b>
8	<b>0.228</b>	0.782	1.374	1.235	0.318	0.380	1.943	0.557	0.314	2.52
9	0.284	0.93	1.811	0.892	4.530	6.278	7.375	8.786	5.777	<b>0.055</b>
10	<b>0.138</b>	0.376	0.557	0.509	1.406	1.484	2.901	3.045	1.428	1.144
11	<b>0.03</b>	0.085	0.106	0.112	0.083	0.07	1.619	0.106	0.083	0.149
12	<b>0.058</b>	0.212	0.236	0.215	0.413	0.420	2.052	0.787	0.428	0.683
13	1.374	3.603	5.399	3.044	2.111	2.792	4.280	5.396	2.654	<b>0.105</b>
14	<b>0.108</b>	0.22	0.394	0.329	3.950	3.893	5.377	8.041	3.619	0.648
15	2.305	7.129	8.789	6.067	<b>0.643</b>	0.828	2.353	1.811	0.800	14.329
16	<b>0.459</b>	0.985	1.609	0.953	1.165	1.180	2.593	2.763	1.089	0.586
17	<b>0.13</b>	0.531	1.026	0.524	0.921	0.784	2.424	1.181	0.954	0.357
18	<b>0.108</b>	0.228	0.625	0.317	0.621	0.681	2.131	1.692	0.597	0.874
19	0.375	0.836	1.179	1.221	0.812	0.981	2.558	2.221	0.825	<b>0.213</b>
20	0.513	1.76	2.286	2.644	0.085	<b>0.079</b>	1.751	0.144	0.081	0.083
21	2.285	11.554	23.353	9.327	0.842	0.828	2.381	1.609	<b>0.790</b>	3.068
22	<b>0.086</b>	0.277	0.462	0.315	7.311	8.071	9.351	8.387	8.030	0.474
23	<b>0.581</b>	1.889	2.786	1.858	4.916	5.493	6.682	6.241	5.167	N/A
24	1.575	4.22	8.045	3.744	0.555	0.689	2.191	1.027	<b>0.492</b>	0.593
25	0.348	1.389	1.203	2.595	0.267	0.247	1.764	0.625	0.310	<b>0.084</b>
26	0.615	1.404	2.429	2.345	0.244	<b>0.22</b>	1.74	0.463	0.232	0.232
27	0.314	1.087	2.22	1.036	0.173	0.142	1.777	0.236	0.160	<b>0.053</b>
28	2.129	5.206	6.244	4.68	1.139	1.316	2.657	2.429	1.315	<b>0.622</b>
29	0.313	1.049	1.568	1.02	0.672	0.783	2.394	1.374	<b>0.647</b>	9.747



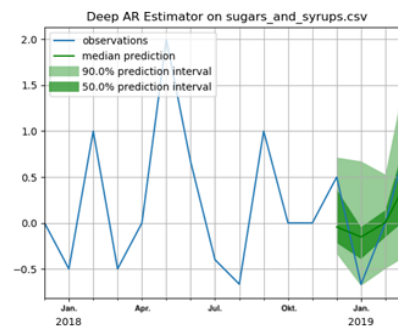
(a) Deep-AR Estimator on feed additives time-series



(b) Deep-AR Estimator on feed materials time-series



(c) Deep-AR Estimator on prepared dishes and snacks time-series



(d) Deep-AR Estimator on sugars and syrups time-series

Εικόνα 43 Ενδεικτικά παραδείγματα προβλέψεων με χρήση εκτιμητή Deep AR σε συγκεκριμένες χρονοσειρές. Σημειωμένο με σκούρο πράσινο είναι το διάστημα πρόβλεψης 0,5 (εμπιστοσύνη) και



σημειωμένο με ανοιχτό πράσινο είναι το διάστημα πρόβλεψης (εμπιστοσύνη) 0,9. Ο άξονα x παρουσιάζει το χρόνο σε μήνες ενώ ο άξονας y αντιπροσωπεύει την ποσοστιαία μεταβολή των ανακλήσεων ανά μήνα. Ο ορίζοντας πρόβλεψης είναι 4 μήνες, ενώ η μπλε γραμμή αντιπροσωπεύει τις πραγματικές τιμές της χρονοσειράς

Οι Πίνακας 21 και Πίνακας 21 χωρίζονται σε τρία μέρη όπου απεικονίζεται η σύγκριση των μοντέλων σε διαφορετικές περιπτώσεις. Το πρώτο μέρος είναι τα αποτελέσματα των μοντέλων με μονομεταβλητά δεδομένα (οι προβλέψεις βασίζονται αποκλειστικά σε μία χρονοσειρά κάθε φορά), ενώ το δεύτερο χρησιμοποιεί τα υποκατάστατα δεδομένα στον Πίνακα 21. Ενώ στον Πίνακα 21 αξιολογούνται και πολυμεταβλητά δεδομένα (και οι 30 χρονοσειρές χρησιμοποιούνται για προβλέψεις) συγκριτικά με τα μονομεταβλητά δεδομένα. Το τελευταίο μέρος και των δύο πινάκων παρέχει μια σύγκριση με την προτεινόμενη προσέγγιση ενισχυτικής μάθησης. Επίσης παρατηρούμε ότι το μοντέλο Deep Factor παραλείπεται στη μονομεταβλητή περίπτωση καθώς δεν είναι εφαρμόσιμο λόγω της ύπαρξης ενός γενικού μοντέλου RNN όπως περιγράφεται στο [237], όπου χρησιμοποιείται για την εκμάθηση μοτίβων σε πολλαπλές σχετικές χρονοσειρές.

Τα αποτελέσματα που παρουσιάζονται υπογραμμίζουν ότι το μοντέλο με την καλύτερη απόδοση στις περισσότερες περιπτώσεις είναι το μοντέλο Deep AR, ενώ το WaveNet ακολουθεί λόγω της ικανότητάς του να καταγράφει μακροπρόθεσμες εξαρτήσεις όπως LSTM αλλά με λιγότερη εκπαίδευση. Το δίκτυο Simple Feed Forward αποδίδει καλύτερα σε χρονοσειρές 4/30 στη ρύθμιση πολλαπλών μεταβλητών.

Όπως ήταν αναμενόμενο, παρόλο που το εποχιακό μοντέλο επιτυγχάνει ως επί το πλείστον χαμηλά σφάλματα λόγω του τρόπου πρόβλεψης των τιμών, δεν ξεπερνά ποτέ όλα τα βαθιά μοντέλα. Ένα άλλο ενδιαφέρον αποτέλεσμα είναι ότι η χρήση μονομεταβλητών μοντέλων φαίνεται να είναι πιο ακριβής. Αυτό έρχεται σε αντίθεση με την αρχική μας υπόθεση ότι με τη χρήση πολυμεταβλητών ροών δεδομένων, θα προκύψουν πολύπλοκες μη γραμμικές αλληλεπιδράσεις χαρακτηριστικών, διευκολύνοντας τη βελτιστοποίηση των μοντέλων. Όπως αναφέρθηκε παραπάνω, τα δεδομένα κανονικοποιήθηκαν σε περίπτωση διαφορετικής κλίμακας δεδομένων. Η χειρότερη απόδοση του πολυμεταβλητού συνόλου δεδομένων αιτιολογείται από το γεγονός ότι αξιοποιείται ένα ευρύ αλλά όχι βαθύ σύνολο δεδομένων, καθιστώντας δυσκολότερη τη διάκριση μεταξύ σήματος και θορύβου καθώς και από την έλλειψη συσχέτισης μεταξύ χρονοσειρών. Επιπρόσθετα, μπορούμε να εκφράσουμε συγκρατημένη αισιοδοξία για τη χρήση ανάλογων/υποκατάστατων δεδομένων ως τεχνικής εμπλουτισμού, η οποία μπορεί να διερευνηθεί περαιτέρω, καθώς τα αποτελέσματα που παρουσιάζονται στον Πίνακα 20, δείχνουν μια επιτυχημένη δοκιμή της προτεινόμενης προσέγγισης σε ορισμένες περιπτώσεις.

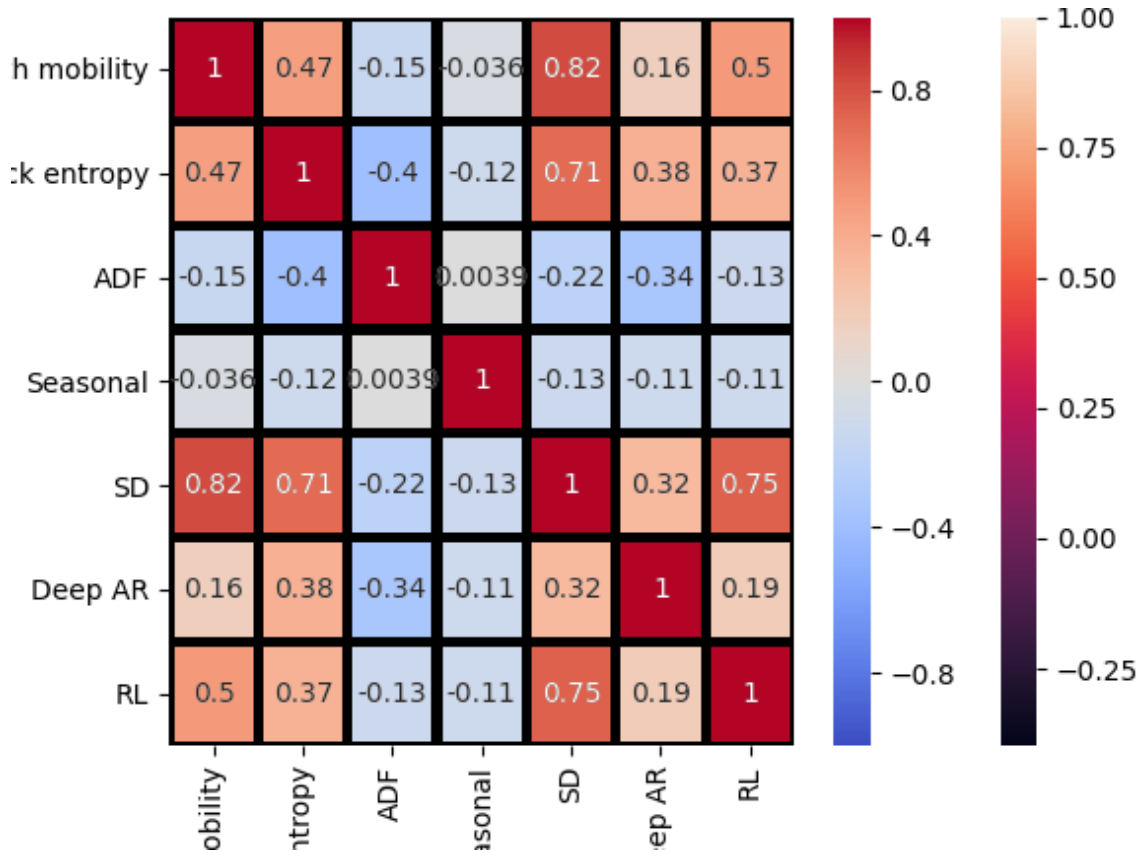


Τέλος, το μοντέλο ενισχυτικής μάθησης που χρησιμοποίησε το προσαρμοσμένο μας περιβάλλον εμφάνισε υποσχόμενα αποτελέσματα. Σύμφωνα με τον Πίνακα 20, έχει καλύτερη απόδοση από όλα τα άλλα μοντέλα σε σύνολο δεδομένων 9/30. Λαμβάνοντας υπόψη το γεγονός ότι το μοντέλο RL εκπαιδεύτηκε σε μονομεταβλητές χρονοσειρές χωρίς υποκατάστατα δεδομένα, μπορούμε να ενισχύσουμε την πεποίθηση ότι η χρήση μονομεταβλητών δεδομένων είναι κατάλληλη για τα συγκεκριμένα δεδομένα.

Σε ένα πραγματικό σενάριο, όπως η περίπτωση που παρουσιάζεται σε αυτήν την εργασία, υπάρχουν άλλα χαρακτηριστικά εκτός από την ακρίβεια των αποτελεσμάτων που μπορεί να κάνουν ένα μοντέλο καταλληλότερο. Τέτοια χαρακτηριστικά είναι για παράδειγμα οι απαιτούμενοι υπολογιστικοί πόροι και η πολυπλοκότητα των δεδομένων ή του μοντέλου. Με αυτό το σκεπτικό, εξετάσαμε επίσης μια προσέγγιση για τη βελτιστοποίηση της διαδικασίας πρόβλεψης χρονοσειρών, λαμβάνοντας υπόψη τα προαναφερθέντα χαρακτηριστικά. Το τελευταίο επιτυγχάνεται με ένα μετα-μοντέλο, το οποίο προτείνει το καλύτερο μοντέλο που θα χρησιμοποιηθεί χωρίς να δοκιμάζονται και να αξιολογούνται εξαντλητικά όλα τα πιθανά υποψήφια μοντέλα. Προς αυτή την κατεύθυνση, η επιλογή μοντέλου θα πρέπει να βασίζεται σε ορισμένες ενδεικτικές στατιστικές μετρήσεις κάθε χρονοσειράς, όπως το Augmented Dickey-Fuller (ADF) που υποδηλώνει τη σταθερότητα, την τυπική απόκλιση (SD), την βαθμό εντροπία και την h-jorth-mobility. Ως εποχιακό στοιχείο, χρησιμοποιήσαμε τα αποτελέσματα του μονομεταβλητού εποχιακού μοντέλου του GlounTS. Επιπλέον, ο Πίνακας 22 παρουσιάζει τις τιμές για τις προαναφερθείσες στατιστικές μετρήσεις για κάθε χρονοσειρά. Εστιάζοντας στις προσεγγίσεις που αποδίδουν τα καλύτερα αποτελέσματα (μονομεταβλητό DeepAR και RL), επιπλέον υπολογίσαμε τη συσχέτιση μεταξύ του σφάλματος πρόβλεψης και των διαφόρων στατιστικών μετρήσεων, παρουσιάζοντας στον θερμικό χάρτη της Εικόνα 44.

**Πίνακας 22** Hjorth mobility, block entropy, ADF, Variance and Seasonality component of the time-series.

	Hjorth mobility	block entropy	ADF	Variance	Seasonal
0	1.116	1.029	-14.606	1.462	14.0
1	1.146	0.163	-9.000	0.152	7.0
2	1.151	0.535	-9.000	0.526	63.0
3	1.168	0.970	-15.682	1.365	28.0
4	1.179	0.608	-14.140	1.466	25.0
5	1.123	0.651	-1.984	0.499	26.0
6	1.172	1.327	-15.398	2.527	38.0
7	1.175	0.878	-13.988	2.101	14.0
8	1.213	0.540	0.676	2.165	9.0
9	1.144	1.134	-6.642	1.046	5.0
10	1.152	0.846	-10.914	1.654	11.0
11	1.130	0.262	-4.956	0.307	71.0
12	1.179	0.934	-15.293	1.476	11.0
13	1.152	0.946	-16.049	0.890	37.0
14	1.081	0.181	-13.479	0.178	140.0
15	1.208	1.330	-13.788	5.791	3.0
16	1.197	0.749	-16.988	1.092	66.0
17	1.144	0.860	-11.610	0.604	9.0
18	1.147	0.961	-16.567	0.783	21.0
19	1.125	0.841	-11.958	0.383	481.0
20	1.090	0.059	-10.602	0.102	36.0
21	1.164	0.274	-14.269	0.960	31.0
22	1.177	0.571	-15.046	1.074	120.0
23	1.201	1.266	-13.241	2.577	10.0
24	1.150	1.229	-18.744	1.035	7.0
25	1.151	0.703	-5.879	0.843	50.0
26	1.118	0.791	-5.793	0.337	51.0
27	1.119	1.088	-13.468	1.047	13.0
28	1.140	1.315	-17.363	1.498	NaN
29	1.228	1.445	-12.721	7.643	7.0



**Εικόνα 44** Συσχετίσεις των μονομεταβλητών μοντέλων DeepA και RL σχετικά με τις στατιστικές μετρήσεις

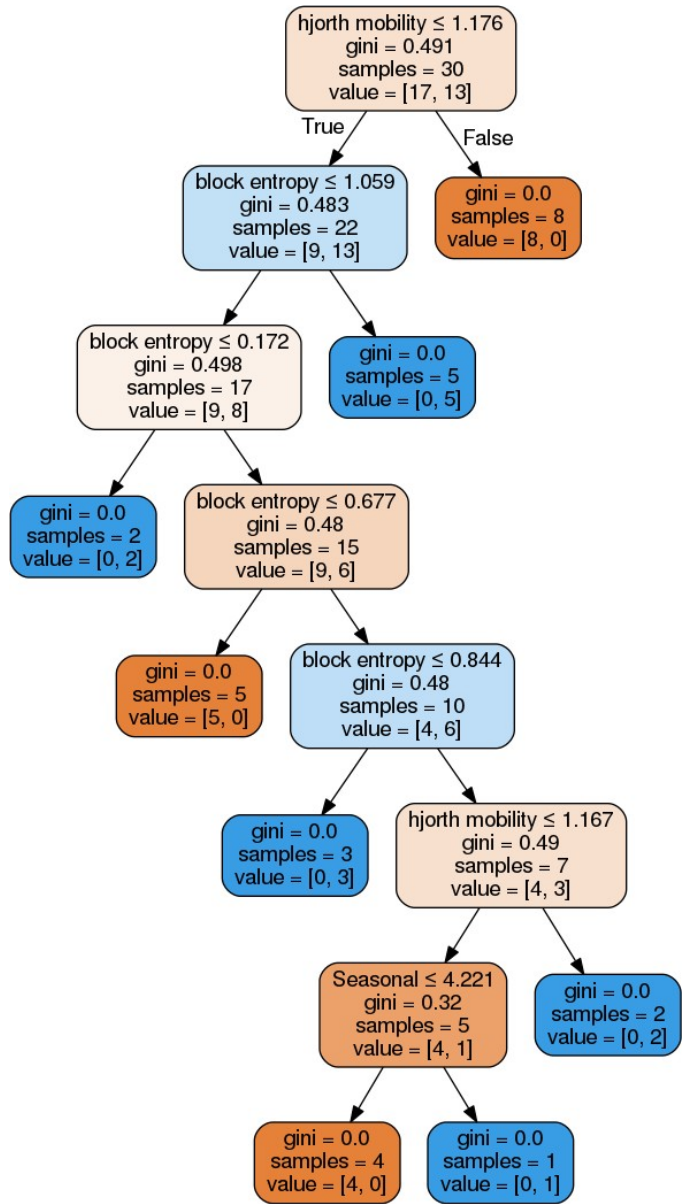
Είναι προφανές ότι το μοντέλο RL συσχετίζεται πολύ θετικά με την SD και την Hjorth mobility. Το τελευταίο σημαίνει ότι όταν μια χρονοσειρά έχει μεγαλύτερη τιμή SD, η προσδοκία του σφάλματος για το μοντέλο RL ακολουθεί την ίδια τάση. Ο λόγος για αυτό είναι ότι χρησιμοποιήσαμε έναν στατικό χώρο δράσης για το μοντέλο RL. Δεδομένων του συνόλου τιμών μας και των αποτελεσμάτων των μοντέλων, εφαρμόσαμε ένα δέντρο αποφάσεων σε μια προσπάθεια να εξηγήσουμε τις αποφάσεις που ελήφθησαν για την επιλογή του μοντέλου. Τα αποτελέσματα απεικονίζονται στην Εικόνα 45. Συγκεκριμένα, σε αυτό το σχήμα οι κόκκινοι χρωματικοί κόμβοι υποδηλώνουν την επιλογή του μοντέλου DeepAR και οι μπλε το μοντέλο RL. Τα κύρια ευρήματα της εφαρμογής του μετα-μοντέλου μπορούν να συνοψιστούν στις ακόλουθες δηλώσεις. Όταν η hjorth-mobility είναι υψηλή, μπορούμε να χρησιμοποιήσουμε με σιγουριά το μοντέλο DeepAR, πράγμα που σημαίνει ότι τα μοντέλα είναι καλύτερα στην καταγραφή διαφορετικών ειδών παραλλαγών της χρονοσειράς. Εκτός αυτού, ένας άλλος σημαντικός παράγοντας είναι η block entropy που είναι μια εκτίμηση της καμπύλης αύξησης της εντροπίας σε σχέση με ένα μέγεθος παραθύρου. Από το δέντρο βλέπουμε ότι όταν η τιμή της hjorth mobility είναι μεγάλη, χρησιμοποιούμε με σιγουριά το μοντέλο DeepAR που σημαίνει ότι αυτό το μοντέλο είναι καλύτερο στην αποτύπωση της παραλλαγής εντός του μπλοκ της χρονοσειράς. Από την άλλη πλευρά, όταν παρατηρούμε





ακραίες τιμές εντροπίας βαθμίδας, τείνουμε να επιλέγουμε το μοντέλο RL που σημαίνει ότι μπορεί να συλλάβει καλύτερα την παραλλαγή μεταξύ των μπλοκ. Η εκμετάλλευση ενός τέτοιου μοντέλου επιτρέπει τη χρησιμότητα των δύο καλύτερων μοντέλων με υβριδικό τρόπο, βελτιστοποιώντας τα αποτελέσματα πρόβλεψης όσον αφορά την ακρίβεια, επιλέγοντας το καλύτερο μοντέλο για κάθε κατηγορία χρονοσειρών.

Η εκμετάλλευση ενός τέτοιου μοντέλου επιτρέπει τη χρήση των δύο μοντέλων με την καλύτερη εφαρμογή με υβριδικό τρόπο, βελτιστοποιώντας τα αποτελέσματα πρόβλεψης όσον αφορά την ακρίβεια, επιλέγοντας το καλύτερο μοντέλο για κάθε κατηγορία χρονοσειρών.



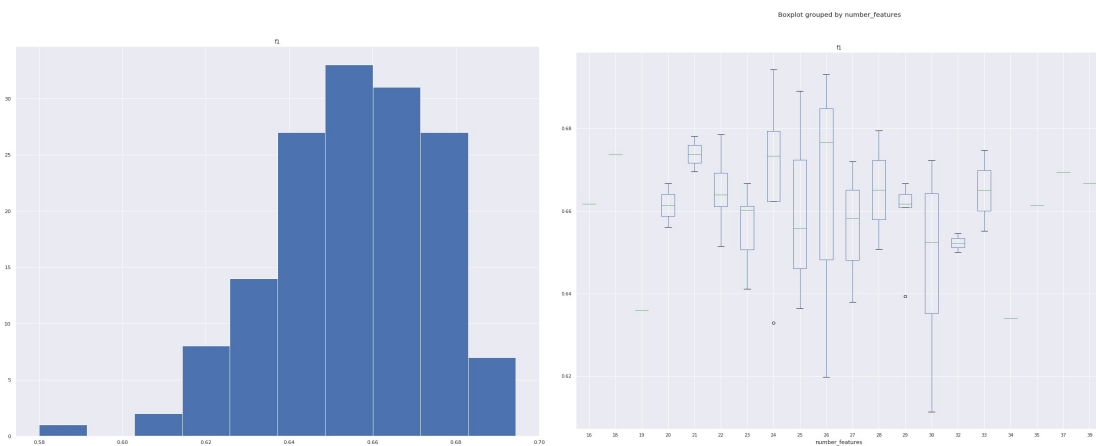
**Εικόνα 45** Δέντρο απόφασης για την επιλογή μοντέλου, το κόκκινο χρώμα υποδηλώνει την επιλογή του μοντέλου Deer Ar , ενώ με το μπλέ χρώμα το μοντέλο RL



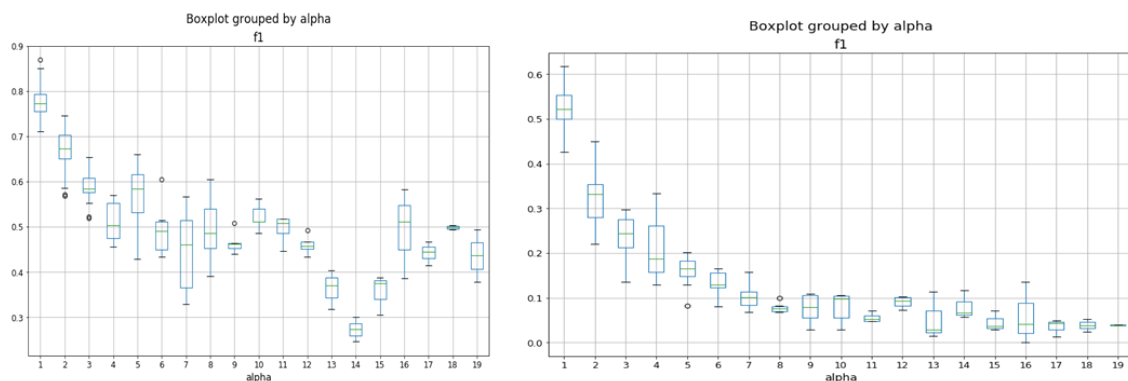
### 4.3 Μηχανική Μάθηση & ποιότητα χοιρινού κρέατος.

Σε αυτή την υπό-ενότητα παρέχεται αρχικά μια σύντομη περιγραφή της στρατηγικής αξιολόγησης σχετικά με τις προσεγγίσεις που εισάγονται σε αυτό το πείραμα, ακολουθούμενη από τα αποτελέσματα του καινοτόμου μοντέλου εκτίμησης σημαντικότητας των χαρακτηριστικών ενός μοντέλου βασισμένο σε μια μέθοδο υποδειγματοληψίας με γενετικό αλγόριθμο.

Δεδομένων των «χαρακτηριστικών» της προσέγγισής μας, καθώς πρέπει να συγκρίνουμε και να αξιολογήσουμε αποτελέσματα από διάφορα υποσύνολα, ο καλύτερος τρόπος για να τα απεικονίσουμε είναι με τη χρήση Box-plots. Με την απεικόνιση τόσο των μέσων τιμών όσο και των αποκλίσεων δίνεται η έννοια της κατανομής και τα στατιστικά στοιχεία των αποτελεσμάτων.



**Εικόνα 46** Κατανομή βαθμολογίας F1 των διαφορετικών επαναλήψεων σε διαφορετικά υποσύνολα



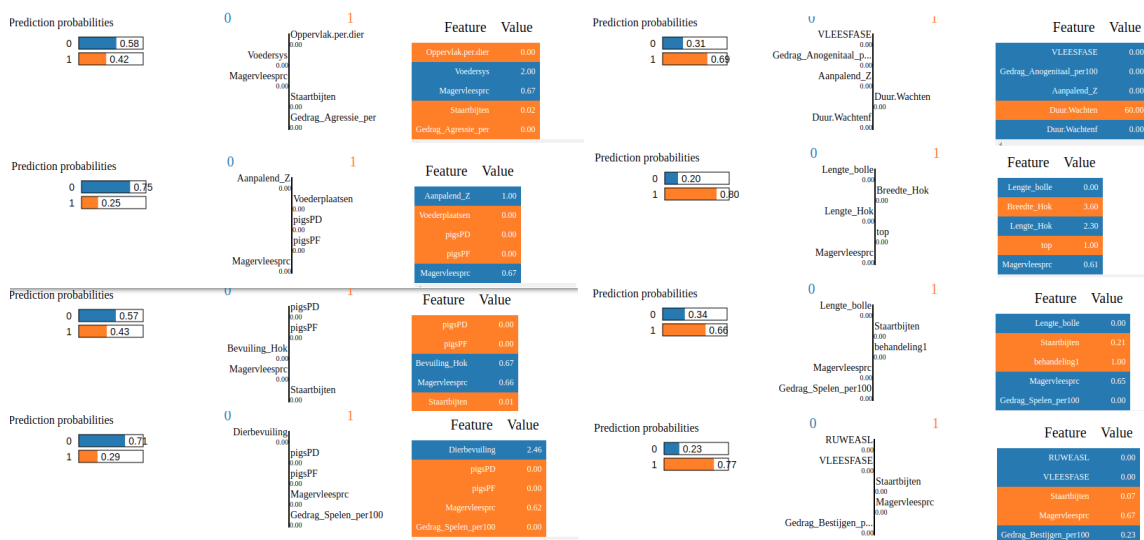
**Εικόνα 47** Boxplots παράγοντα ανισορροπίας ως βάρος κατηγορίας. Και στα δύο οικόπεδα η κατανομή των έχει ως αποτέλεσμα διάφορους παράγοντες ανισορροπίας. Στο αριστερό διάγραμμα δεν χρησιμοποιήθηκε κάποιος συντελεστής να αντισταθμίσει την ανισορροπία των δεδομένων ενώ στο δεύτερο οικόπεδο το βάρος κλάσης ορίστηκε ίσο με τον συντελεστή 1/ανισορροπία

Συγκεκριμένα στις εικόνες διακρίνονται τα αποτελέσματα του μοντέλου Catboost στο δυαδικό πρόβλημα ταξινόμησης, τόσο στο αρχικό σύνολο δεδομένων, όσο και στα υποσύνολα που δημιουργήθηκαν βάσει του προτεινόμενου τρόπου



υποδειγματοληψίας και επιλογής χαρακτηριστικών. Συγκρίνοντας τα αποτελέσματα της Εικόνα 46 και της Εικόνα 47, μπορούμε να εκτιμήσουμε ότι όσον αφορά τη μετρική F1-score, το αποτέλεσμα φαίνεται να βελτιώνεται κατά 0,1 μονάδες (από 0,55 έως 0,65) όταν εφαρμόζεται η προσαρμοσμένη προσέγγιση υποδειγματοληψίας με την επιλογή χαρακτηριστικών GA. Επιπλέον, τα αποτελέσματα που ενσωματώνουν όλα τα διαφορετικά υποσύνολα προτείνουν ότι τα κορυφαία 5 χαρακτηριστικά στην Αλλαγή Τιμών Πρόβλεψης και Αλλαγή Συνάρτησης Απώλειας φαίνονται στον Πίνακα 23, στην Εικόνα 49, και στην Εικόνα 50, αντίστοιχα.

Τέλος, σε ένα από τα υποσύνολα, εφαρμόσαμε το μεθόδους επεξηγησιμότητας τεχνητής νομοσύνης και πιο συγκεκριμένα τη μέθοδο LIME για να διερευνήσουμε ποια χαρακτηριστικά είναι αυτά που «μπερδεύουν» το μοντέλο. Παρατηρήσαμε ορισμένα κοινά χαρακτηριστικά σε ορισμένα δείγματα που δεν ταξινομήθηκαν, όπως ο αριθμός των χοίρων ανά τροφοδότη και ο αριθμός των χοίρων ανά θηλή. Επιπλέον, μπορεί να σημειωθεί ότι η εμπιστοσύνη δεν είναι πολύ υψηλή (πιθανότητες πρόβλεψης κοντά στο 0,5) στα μη ταξινομημένα δείγματα σε σύγκριση με τα σωστά ταξινομημένα.

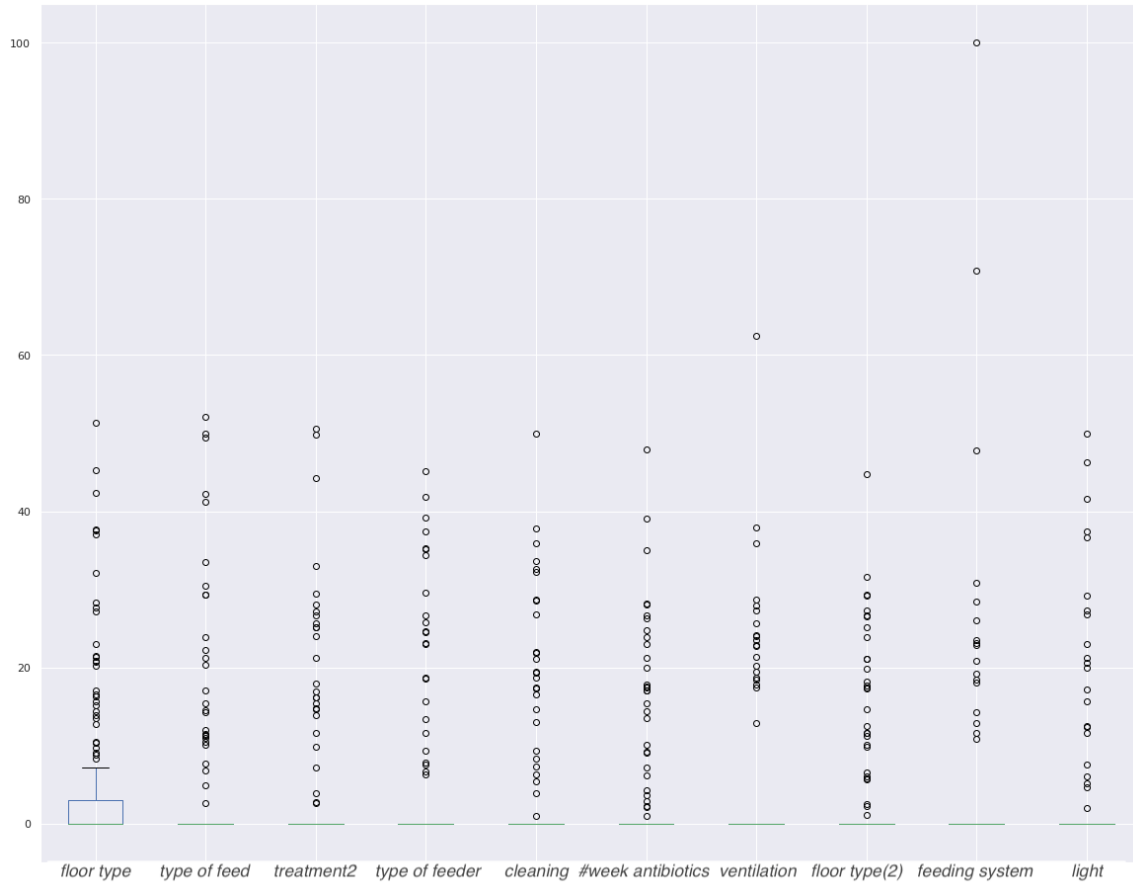


Εικόνα 48 Εφαρμογή μεθόδου LIME για επεξήγηση αποτελεσμάτων ταξινομητή (βάσει ποιων χαρακτηριστικών έγινε η κάθε απόφαση)

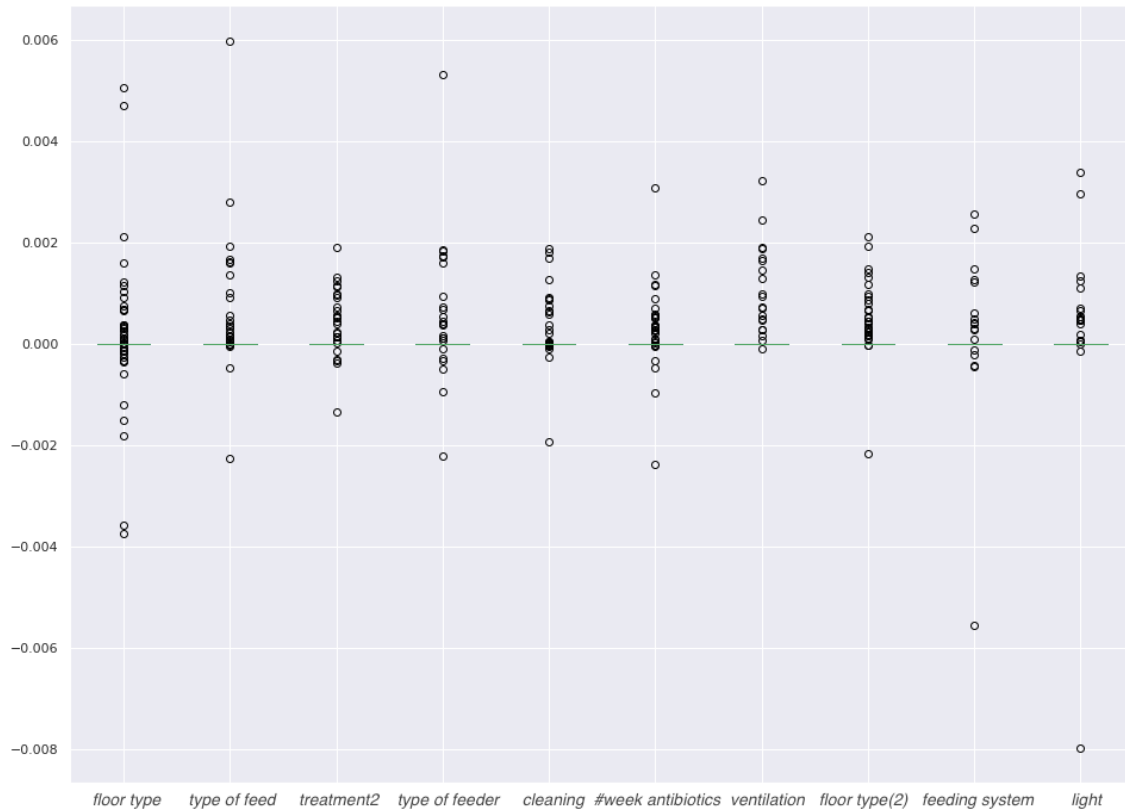
Τα αποτελέσματα του εφαρμοσμένου μοντέλου σε κάθε υποσύνολο απεικονίζονται στην Εικόνα 49 και στην Εικόνα 50 σε σχέση με τον αριθμό των χρησιμοποιούμενων χαρακτηριστικών. Θα πρέπει να σημειωθεί ότι ενδέχεται να υπάρχει ασυμφωνία μεταξύ των επιλεγμένων χαρακτηριστικών του GA και εκείνων που τελικά αξιοποιήθηκαν από το μοντέλο Catboost. Αυτό σημαίνει ότι υπάρχει η δυνατότητα να επιλεγεί ένα χαρακτηριστικό σε μια λύση του GA, αλλά από την άλλη



πλευρά, να καταλήξει σε χαμηλή βαθμολογία σημασίας χαρακτηριστικών από το μοντέλο Catboost. Έτσι, μπορούμε με βεβαιότητα να υποστηρίξουμε τη σημασία τους μόνο για τα αποτελέσματα που παρουσιάζονται στον Πίνακα 23, καθώς αυτά είναι αυτά που χρησιμοποιούνται στην πραγματικότητα από το μοντέλο.



**Εικόνα 49** Τα πιο σημαντικά χαρακτηριστικά όσον αφορά την αλλαγή των τιμών πρόβλεψης



**Εικόνα 50** Τα πιο σημαντικά χαρακτηριστικά από την άποψη της αλλαγής λειτουργίας απώλειας

**Πίνακας 23** Κορυφαία χαρακτηριστικά όσον αφορά την αλλαγή των τιμών πρόβλεψης και την αλλαγή της λειτουργίας απώλειας

Feature	Prediction Values Change	Feature	Loss Function Change
Type of feed (pellet)	4.856722	Ventilation system	1.405127e-04
Type of feed	4.039977	Type of feed	1.271893e-04
Treatment 2	3.828014	Sowline	1.140735e-04
Type of feeder	3.741619	Type of feed (meal)	1.112171e-04
Cleaning	3.662566	Duration of waiting in lairage	1.089759e-04
Weeks antibiotics given	3.570304	Type of feeder	9.738937e-05
Ventilation system	3.395850	Season	8.879115e-05
Type of feed (meal)	3.387441	Temperature on day of slaughter	8.780216e-05
Feeding system	3.330306	Illnes	8.599249e-05
Light	3.162993	Water	8.341634e-05
Distraction material	2.906545	Cleaning method	8.128506e-05
Cleaning method	2.771212	Deaths caused by illness	7.876333e-05
Water	2.337957	Treatment 2	7.857921e-05
Temperature on day of slaughter	2.319449	Boartype	7.830557e-05
Sowline	2.145860	Aggressive behavior per pen	7.734633e-05
Duration of waiting in lairage	1.984435	Cleaning	6.956905e-05
Boartype	1.947673	Lux	6.504661e-05

Όσον αφορά τη μέθοδο της ενσωματωμένης σημασίας χαρακτηριστικών ενός μοντέλου που βασίζεται σε δέντρα, υπάρχει μια ανοιχτή διαμάχη σχετικά με τα συσχετισμένα χαρακτηριστικά. Πιο αναλυτικά, εάν υπάρχουν χαρακτηριστικά υψηλής συσχέτισης σε ένα σύνολο δεδομένων και το μοντέλο χρειάζεται ένα από αυτά, θα επιλέξει ένα εξ αυτών τυχαία. Όταν πρόκειται για ένα σύνολο βασισμένο



σε δέντρα αποφάσεων μοντέλο, αυτή η τυχαία επιλογή θα γίνει για κάθε δέντρο επειδή κάθε δέντρο είναι ανεξάρτητο από τα υπόλοιπα. Ως εκ τούτου, η σημασία των πληροφοριών που περιέχονται σε αυτά τα χαρακτηριστικά μειώνεται. Επομένως, η σημασία που προκύπτει από την αναφορά σε αυτά τα χαρακτηριστικά θα είναι παραπλανητική. Αυτό δεν ισχύει για τα ενισχυτικά μοντέλα που βασίζονται σε δέντρα (δηλ. Catboost). Εκεί, όταν έχει «μαθευτεί» μια συγκεκριμένη συσχέτιση μεταξύ χαρακτηριστικού και αποτελέσματος, δεν θα αναθεωρηθεί. Επομένως, όλη η σημασία θα είναι σε ένα από τα συσχετισμένα χαρακτηριστικά. Σε κάθε περίπτωση, θα πρέπει να εξετάζονται ρητά όλα τα χαρακτηριστικά. Δεδομένου του γεγονότος ότι στις περισσότερες περιπτώσεις, ακόμη και σε λιγότερο μη ισορροπημένα υποσύνολα, το εκπαιδευμένο μοντέλο δεν είχε καλή απόδοση σε δεδομένα εκτός δείγματος (άγνωστα δεδομένα), μπορούμε να συμπεράνουμε ότι τα δεδομένα, τουλάχιστον στην παρεχόμενη μορφή, δεν περιέχουν ισχυρή προγνωστική ισχύ. Από την άλλη πλευρά, η παρουσία των ίδιων χαρακτηριστικών, όπως ο τύπος της τροφής, ο τύπος του συστήματος εξαερισμού, η φαρμακευτική αγωγή και η διάρκεια αναμονής στο στάβλο σχεδόν σε κάθε αξιοποιημένη προσέγγιση υποδηλώνει εξάρτηση από αυτά της εμφάνισης της κηλίδωσης κάπρου.

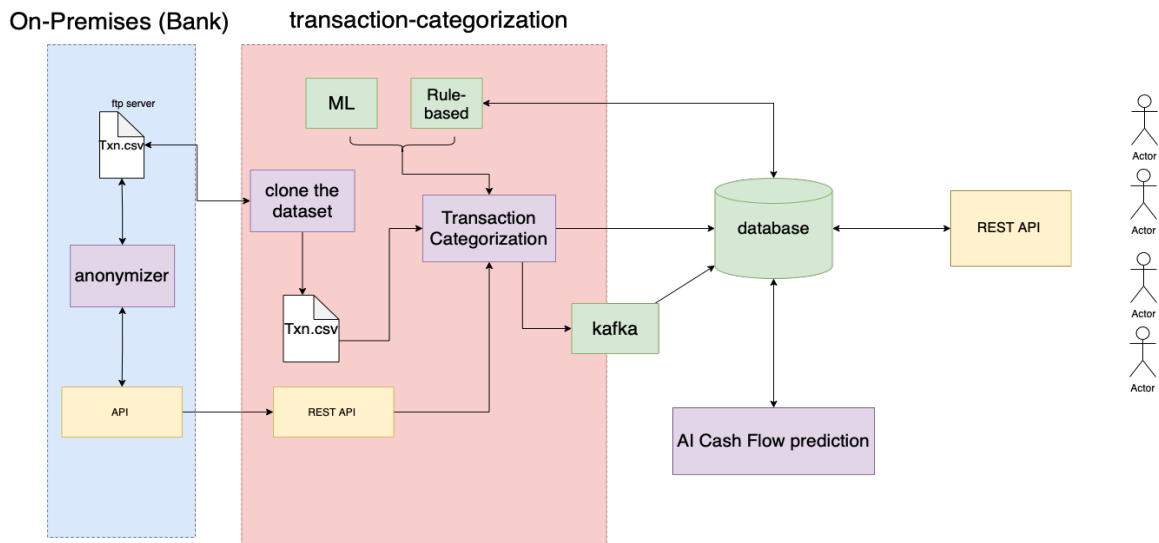
Τελικά καταλήγουμε στην κύρια ερώτηση μας, πιστεύουμε ότι σε αυτήν τη συγκεκριμένη περίπτωση χρήσης, οι περιπτώσεις που ταξινομούνται λανθασμένα ως μη μολυσμένες (ψευδώς αρνητικά) παρουσιάζουν μεγαλύτερο ενδιαφέρον και θα πρέπει να συνοδεύονται από τοπικές εξηγήσεις. Τα ψευδώς θετικά είναι επίσης ενδιαφέροντα, αλλά λίγο λιγότερο επικριτικά από τα ψευδώς αρνητικά. Τα πιο σημαντικά χαρακτηριστικά για τις λανθασμένες περιπτώσεις παρουσιάζουν ενδιαφέρον καθώς είναι παραπλανητικά.

#### **4.4 Μηχανική Μάθηση & τραπεζικός βοηθός για μικρομεσαίες επιχειρήσεις.**

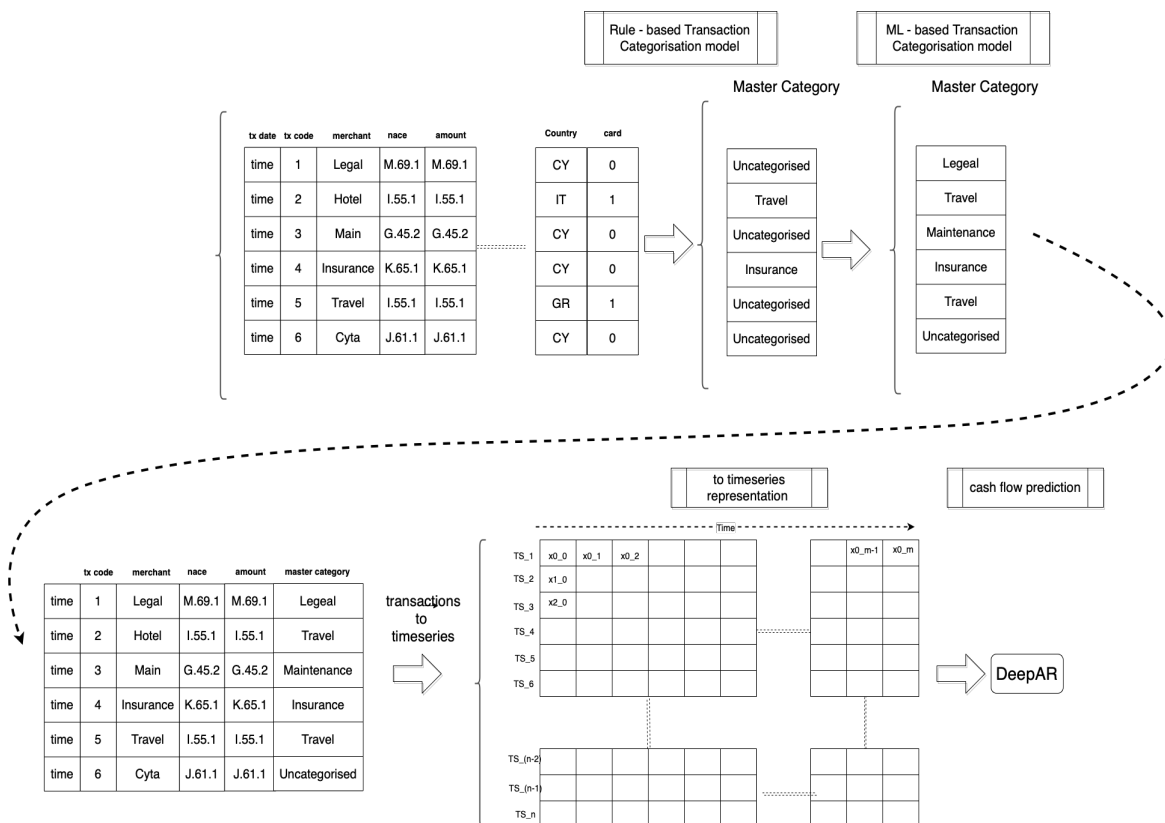
Σε αυτή την ενότητα παρέχεται αρχικά μια σύντομη περιγραφή της στρατηγικής αξιολόγησης σχετικά με τις προσεγγίσεις που εισάγονται σε αυτό το πείραμα, ακολουθούμενη από τα αποτελέσματα τόσο της κατηγοριοποίησης των συναλλαγών όσο και των μοντέλων πρόβλεψης ταμειακών ροών. Μια πιο λεπτομερής άποψη αυτού παρέχεται στην Εικόνα 51 που απεικονίζει τη λογική όψη των συνδεδεμένων εξαρτημάτων μαζί με την Εικόνα 52, που απεικονίζει τη γραμμή διαχείρισης δεδομένων, προσέγγιση με διαδοχικά και διακριτά βήματα, για την εφαρμογή τόσο της διαδικασίας κατηγοριοποίησης συναλλαγών όσο και της πρόβλεψης ταμειακών ροών μέσω πρόβλεψης χρονοσειρών. Στο πρώτο βήμα, τα ακατέργαστα δεδομένα υποβλήθηκαν σε προ-επεξεργασία και τροφοδοτήθηκαν στο μοντέλο κατηγοριοποίησης βάσει στατικών κανόνων. Στη συνέχεια, οι συναλλαγές «μη κατηγοριοποιημένες» (δηλαδή που δεν επισημάνθηκαν από το



μοντέλο βάσης κανόνων) ταξινομήθηκαν σε συγκεκριμένες κατηγορίες και τελικά τα δεδομένα μετασχηματίστηκαν σε αναπαράστασεις χρονοσειρών για να χρησιμοποιηθούν για την πρόβλεψη μελλοντικών εισροών και εκροών.



Εικόνα 51 Λογική άποψη των συνδεδεμένων στοιχείων του πλαισίου, συμπεριλαμβανομένης της κατηγοριοποίησης συναλλαγών και της πρόβλεψης ταμειακών ροών



Εικόνα 52 Σχηματική εξήγηση της πτυχής δεδομένων στην προσέγγιση από άκρο σε άκρο συμπεριλαμβανομένων των τεσσάρων βασικών βημάτων στη διαδικασία



#### 4.4.1 Ταξινόμηση υβριδικών συναλλαγών

Όπως αναφέρθηκε παραπάνω, η απουσία χαρακτηρισμένων δεδομένων (labeled data) αποτελεί μία κύρια πρόκληση κατά την ανάπτυξη ενός μοντέλου μηχανικής μάθησης για την κατηγοριοποίηση συναλλαγών. Σε αυτά τα σενάρια, προκύπτουν δύο διαδεδομένες προσεγγίσεις. Η πρώτη χρησιμοποιεί τεχνικές μηχανικής εκμάθησης χωρίς επίβλεψη για τη δημιουργία συμπλεγμάτων χωρίς προηγούμενη γνώση των αναμενόμενων αποτελεσμάτων. Ωστόσο, καθώς οι ετικέτες είναι σταθερές και αδιαπραγμάτευτες στον χρηματοοικονομικό τομέα, αυτή η προσέγγιση μπορεί να οδηγήσει σε ένα λιγότερο ισχυρό μοντέλο όσον αφορά το επίπεδο εμπιστοσύνης, με ορισμένες ομάδες να μην μπορούν να αντικατοπτρίζουν μια ξεχωριστή κατηγορία με πραγματική αξία για τις ΜΜΕ. Η δεύτερη και προτεινόμενη προσέγγιση είναι αρχικά η χειροκίνητη επισήμανση ενός αντιπροσωπευτικού υποσυνόλου που βασίζεται σε γνώσεις ειδικών, δημιουργώντας ένα μοντέλο βασισμένο σε κανόνες, το οποίο στη συνέχεια μπορεί να ενσωματωθεί με ένα μοντέλο εποπτευόμενης μηχανικής εκμάθησης, προσφέροντας υψηλό βαθμό αυτοματοποίησης για την ενημέρωση και επαναταξινόμηση των συναλλαγών. Για το σκοπό αυτό, κάθε τμήμα του Υβριδικού Μοντέλου (μοντέλο που βασίζεται σε κανόνες και μοντέλο μηχανικής μάθησης) πρέπει να αντιμετωπίσει συγκεκριμένες προκλήσεις. Συγκεκριμένα το μοντέλο βάσει κανόνων ακολούθησε μια προσέγγιση διακριτών βημάτων, αναγνωρίζοντας διάφορους κώδικες της Τράπεζας, ορισμένοι από τους οποίους ερμηνεύονται κατά περίπτωση μόνο από την ίδια την τράπεζα (δηλ. Κωδικός Τύπου Συναλλαγής) και άλλοι χρησιμοποιούνται καθολικά στον επιχειρηματικό κόσμο (δηλαδή Κωδικός κατηγορίας εμπόρου και κωδικοί NACE). Πιο αναλυτικά, οι βασικές μεταβλητές που χρησιμοποιούνται απεικονίζονται στον παρακάτω πίνακα:

**Πίνακας 24** Χαρακτηριστικά που χρησιμοποιήθηκαν για το μοντέλο κανόνων και εν συνεχεία στο υβριδικό μοντέλο

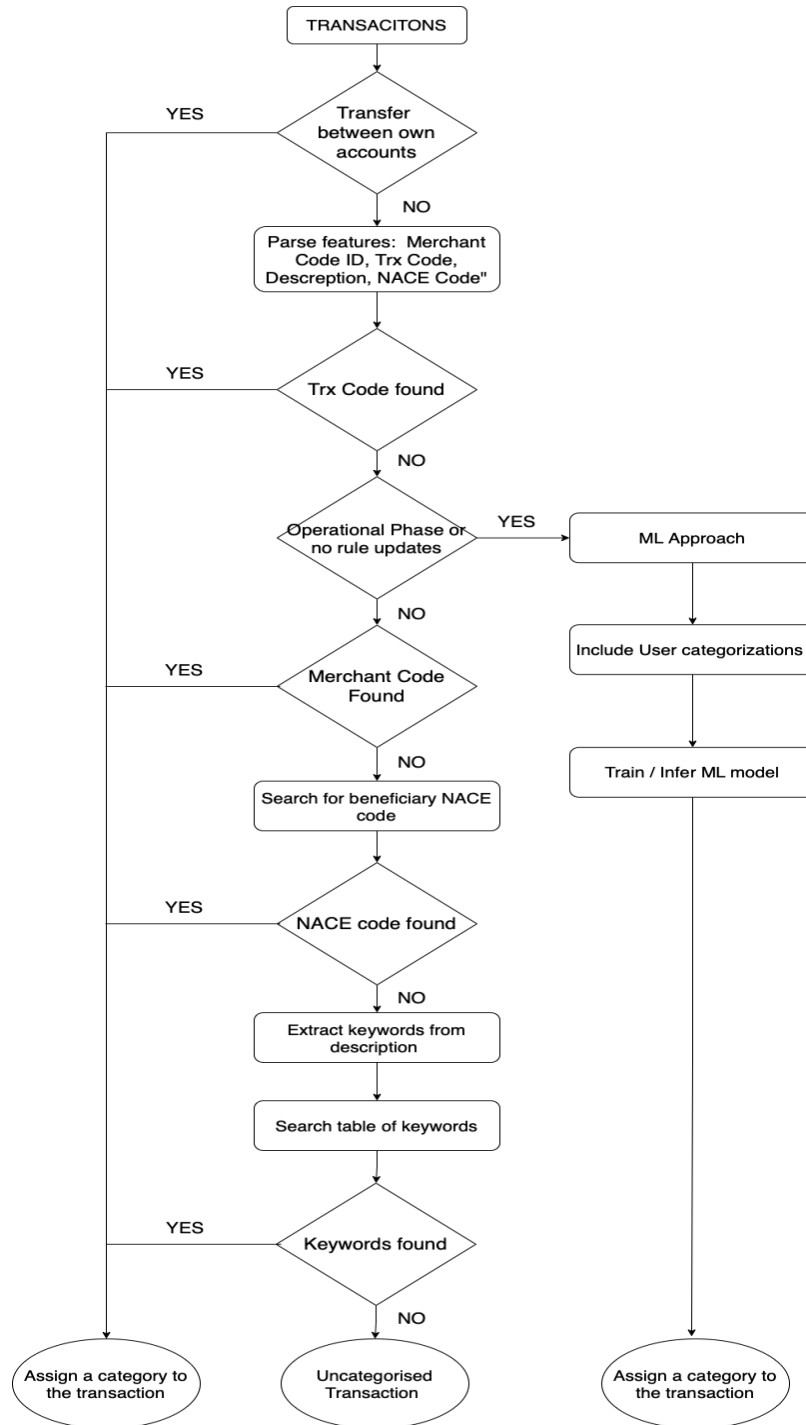
<i>Variable</i>	<i>Description</i>
<i>Transaction Type Code</i>	An internal code provided for every bank transaction, providing information regarding the transaction type and including Root Categories like cash withdrawals, deposits, banking fees etc.
<i>Merchant Category Code</i>	A four-digit number listed in ISO 18245 for financial services, mainly used to classify a business by the types of goods and services provided.
<i>NACE code</i>	A European industry standard system used across European Union for classifying business activities. Utilizing the NACE code of the transfer account found in the available SME customer data.
<i>Transaction Description</i>	Including mainly native retailers, service providers, consulting and legal firms not included in previous steps.

Πριν από την αντιστοίχιση των παραπάνω μεταβλητών σε συγκεκριμένες κατηγορίες, ήταν σημαντικό να αποτυπωθούν όλες οι συναλλαγές μεταξύ των ίδιων λογαριασμών που ανήκουν στην ίδια ΜΜΕ, καθώς αυτές θα ταξινομούσαν ως «Μεταφορές μεταξύ ιδίων λογαριασμών». Το ακριβές διάγραμμα ροής της υβριδικής





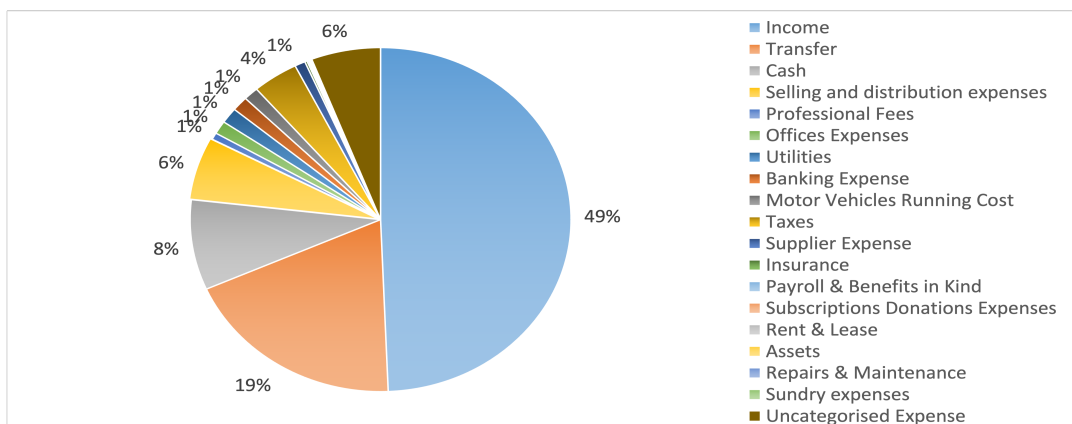
συναλλαγής απεικονίζεται στην Εικόνα 53 απεικονίζοντας τόσο την κατηγοριοποίηση βημάτων βάσει κανόνων όσο και την προσέγγιση ταξινόμησης βασισμένη σε αλγόριθμο μηχανικής μάθησης. Η αξιολόγηση κατηγοριοποίησης απεικονίζεται στην Εικόνα 54 που αναφέρεται στην κατανομή και την κάλυψη του συνόλου δεδομένων που μπορεί να επιτευχθεί με βάση την προσέγγιση βημάτων που βασίζεται σε κανόνες.



Εικόνα 53 Διάγραμμα ροής μοντέλου υβριδικής ταξινόμησης



Όσον αφορά το μοντέλο ταξινόμησης Catboost, μια βασική πρόκληση για την ανάπτυξη ενός αποτελεσματικού υβριδικού μοντέλου ταξινόμησης είναι η άμβλυση της μεροληψίας που εισάγεται από το μοντέλο που βασίζεται σε κανόνες, επιτρέποντας στον χρήστη να επανα-κατηγοριοποιήσει μια δεδομένη κατηγορία συναλλαγών. Αυτή η διαδικασία ενημέρωσης των ετικετών είναι εκτός του πεδίου εφαρμογής αυτής της εργασίας, ωστόσο, ο τρόπος ενημέρωσης των γνώσεων μας και προσαρμογής του μοντέλου σε επανακατηγοριοποιήσεις είναι θεμελιώδης. Για το σκοπό αυτό, εκπαιδεύτηκε ένα μοντέλο Catboost με τρόπο που να υιοθετεί τις αλλαγές που πραγματοποιούν οι MME. Δεδομένης της φύσης των διαφορετικών βημάτων που χρησιμοποιούνται στο μοντέλο που βασίζεται σε κανόνες και προς την υψηλή απόδοση και ακρίβεια κατηγοριοποίησης, προτιμήθηκε ένα υβριδικό μοντέλο αντί για ένα πλήρως AI, ως αποτέλεσμα της εμπιστοσύνης που παρέχει η κατηγοριοποίηση με βάση τα πρώτα δύο βήματα του μοντέλου που βασίζεται σε κανόνες.



Εικόνα 54 Αποτελέσματα κατηγοριοποίησης βημάτων βάσει κανόνων

Πιο αναλυτικά, καθώς ορισμένες βασικές κατηγορίες που παράγονται από τα δύο πρώτα βήματα (δηλαδή μεταφορά μεταξύ λογαριασμών και αντιστοίχιση κωδικού τύπου συναλλαγής) του μοντέλου που βασίζεται σε κανόνες έχουν καθοριστεί, ο στόχος αυτού του μοντέλου είναι να μάθει και να μιμηθεί τα τρία τελευταία βήματα. Ο αριθμός των υπόλοιπων ετικετών (δηλαδή των Κύριων Κατηγοριών) που μπορούν να παραχθούν σε αυτά τα βήματα είναι 16, επομένως εφαρμόστηκε η αξιολόγηση σε εργασίες πολλαπλών τάξεων ως προς τις διάφορες μετρήσεις. Προκειμένου να έχουμε καλύτερη αντίληψη του αποτελέσματος και της προγνωστικής ισχύος σε κάθε ετικέτα, παρουσιάζεται ένας πίνακας σύγχυσης (Confusion Matrix) στον Πίνακα 25. Ο πίνακας αναφέρεται στα αποτελέσματα στο σύνολο δοκιμών λαμβάνοντας υπόψη ένα σχήμα αξιολόγησης 40-30-30, όπου το 40% του συνόλου των δεδομένων είναι το σετ εκπαίδευσης 30% η επικύρωση και 30% το σετ δοκιμών. Τα αποτελέσματα που παρουσιάζονται έχουν ληφθεί από ένα μοντέλο Catboost που έχει ήδη βελτιστοποιηθεί ως προς τις υπερπαραμέτρους, έχει βελτιστοποιηθεί με τη χρήση τεχνικών αναζήτησης πλέγματος και διαίσθησης, ως θεμελιώδες βήμα ενός αγωγού ML. Έχοντας κατά νου ότι οι ετικέτες είναι εξαιρετικά



ανισορροπημένες, η εύρεση ενός κατάλληλου παράγοντα κανονικοποίησης ήταν μια άλλη πρόκληση που ξεπεράστηκε με τη βελτιστοποίηση υπερπαραμέτρων. Οι αλγόριθμοι αναζήτησης πλέγματος και τυχαίας αναζήτησης χρησιμοποιούνται ευρέως για την επιλογή των βέλτιστων παραμέτρων. Βασικά, το πεδίο των υπερπαραμέτρων σε ένα διακριτό πλέγμα. Στη συνέχεια, κάθε συνδυασμός τιμών αυτού του πλέγματος (δηλ. πλέγμα-αναζήτηση) χρησιμοποιείται σε ένα σχήμα αξιολόγησης διασταυρούμενης επικύρωσης. Το σημείο του πλέγματος που μεγιστοποιεί τη μέση τιμή στη διασταυρούμενη επικύρωση, είναι το βέλτιστο σύνολο τιμών για τις υπερπαραμέτρους. Η διαδικασία για την κατηγοριοποίηση των συναλλαγών συνοψίζεται στον κάτωθι Αλγόριθμο.

---

**Algorithm 1:** HYBRID MODEL transaction categorization

---

```
1 Initialize batch of transactions  $T$  to length  $N$ 
2 Initialize the transaction-code-rule as dictionary  $txn - dict$ 
3 Initialize the merchant-code-rule as dictionary  $mcc - dict$ 
4 Initialize the nace-code-rule as dictionary  $nace - dict$  for  $i = 1, N$  do
5   Get the transaction  $T_i$  Get the Beneficiary Account  $BAcc_i$  of the
    $T_i$  Load accounts of the customer  $accounts$ 
6   if  $BAcc_i$  is in  $accounts$  then
7     set category of  $T_i =$  'Transfer between own accounts'
8     continue
9   end
10  else
11    Get the Transaction code  $Txncode_i$  of the  $T_i$ 
12    if  $Txncode_i$  is in  $txn - dict$  then
13      set category of  $T_i = txn\_dict_i$ 
14      continue
15    end
16    else
17      if Operational Use or No rule updates then
18        Load the pre-trained Catboost Model  $model$ 
19        set category of  $T_i = model.predict(T_i)$ 
20        continue
21      end
22      else
23        Get the Merchant code  $MCCid_i$  of the  $T_i$  if  $MCCid_i$  is
        in  $mcc - dict$  then
24          set category of  $T_i = mcc\_dict_i$ 
25          continue
26        end
27        else
28          Get the Beneficiary NACE code  $Nace_i$  of the  $T_i$  if
           $Nace_i$  is in  $nace - dict$  then
29            set category of  $T_i = nace\_dict_i$ 
30            continue
31          end
32        end
33      end
34    end
35  end
36 end
```



Το κύριο αποτέλεσμα που μπορεί να προκύψει από τα αποτελέσματα είναι ότι το μοντέλο μπορεί να μάθει τους κανόνες που χρησιμοποιούνται ενσωματώνοντας τον κώδικα NACE καθώς και τον κωδικό του πωλητή και να κατηγοριοποιήσει σωστά τα αποτελέσματα με ακρίβεια 98%. Αξίζει να σημειωθεί ότι περιλαμβάνεται μια κατηγορία «Μη Κατηγοριοποιημένες Δαπάνες», καθώς οι συναλλαγές που εμπίπτουν σε αυτήν την κατηγορία θα κατηγοριοποιηθούν στη συνέχεια από τη ΜΜΕ και θα ενσωματωθούν στο μοντέλο.

Αν και η στατιστική, με τις διάφορες δοκιμές υποθέσεων και τη συστηματική μελέτη της μεταβλητής σημασίας, είναι καλά εδραιωμένη και μελετημένη, δεν ισχύει το ίδιο για την επεξήγηση διαφόρων τεχνικών στη μηχανική και τη βαθιά μάθηση. Παρόλο που έχουν εφαρμοστεί εκτενώς πολλαπλές μετρήσεις και μετρήσεις απόδοσης, ορισμένες από αυτές μπορεί να είναι μάλλον παραπλανητικές καθώς δεν μεταφέρουν την αιτιολόγηση πίσω από την απόφαση που ελήφθη. Ισχυρότερες μορφές ερμηνείας θα μπορούσαν να προσφέρουν πολλά πλεονεκτήματα, από την εμπιστοσύνη στις προβλέψεις μοντέλων και την ανάλυση σφαλμάτων έως τη βελτίωση του μοντέλου. Αυτό ισχύει καλύτερα σε τομείς όπου η λήψη αποφάσεων με βαρύ κόστος, όπως η ιατρική, ή σε περιπτώσεις όπου πρέπει να ληφθούν προληπτικά μέτρα.

**Πίνακας 25** Σύγκριση Διάφορων μοντέλων ταξινόμησης για εργασία κατηγοριοποίησης συναλλαγών. Αποτελέσματα 10-πλάσια Διασταυρούμενη επικύρωση (με έντονη γραφή το προτεινόμενο μοντέλο)

<i>Model</i>	<i>Accuracy</i>	<i>AUC</i>	<i>Recall</i>	<i>Prec.</i>	<i>F1</i>	<i>Kappa</i>	<i>MCC</i>	<i>TT (Sec)</i>
<i>Extreme Gradient Boosting</i>	0.9554	0.698	0.8012	0.9569	0.9545	0.943	0.9434	298.3
<b><i>CatBoost Classifier</i></b>	<b>0.9476</b>	<b>0.6974</b>	<b>0.7398</b>	<b>0.9503</b>	<b>0.9461</b>	<b>0.9329</b>	<b>0.9335</b>	<b>49.64</b>
<i>Extra Trees Classifier</i>	0.9457	0.6949	0.7547	0.9453	0.9447	0.9306	0.9308	15.27
<i>Logistic Regression</i>	0.9427	0.6969	0.7258	0.9458	0.9409	0.9267	0.9275	41.51
<i>Gradient Boosting Classifier</i>	0.9402	0.6944	0.7518	0.9446	0.939	0.9234	0.9243	170.4
<i>SVM - Linear Kernel</i>	0.9396	0	0.6948	0.9446	0.937	0.9226	0.9237	3.268
<i>Ridge Classifier</i>	0.937	0	0.6917	0.9437	0.9342	0.9191	0.9203	0.396
<i>Linear Discriminant Analysis</i>	0.934	0.6953	0.6867	0.9444	0.9333	0.9156	0.9175	8.38
<i>K Neighbors Classifier</i>	0.8455	0.6678	0.4984	0.8341	0.8366	0.8014	0.802	121.1
<i>Naive Bayes</i>	0.818	0.6401	0.7371	0.952	0.8615	0.7776	0.7866	1.359
<i>Ada Boost Classifier</i>	0.7951	0.5463	0.4443	0.6992	0.7353	0.7255	0.7456	3.36
<i>Light Gradient Boosting Machine</i>	0.6293	0.5559	0.3534	0.6996	0.6429	0.5442	0.5484	20.12
<i>Quadratic Discriminant Analysis</i>	0.0111	0.354	0.0496	0.2889	0.0203	0.0068	0.031	4.846

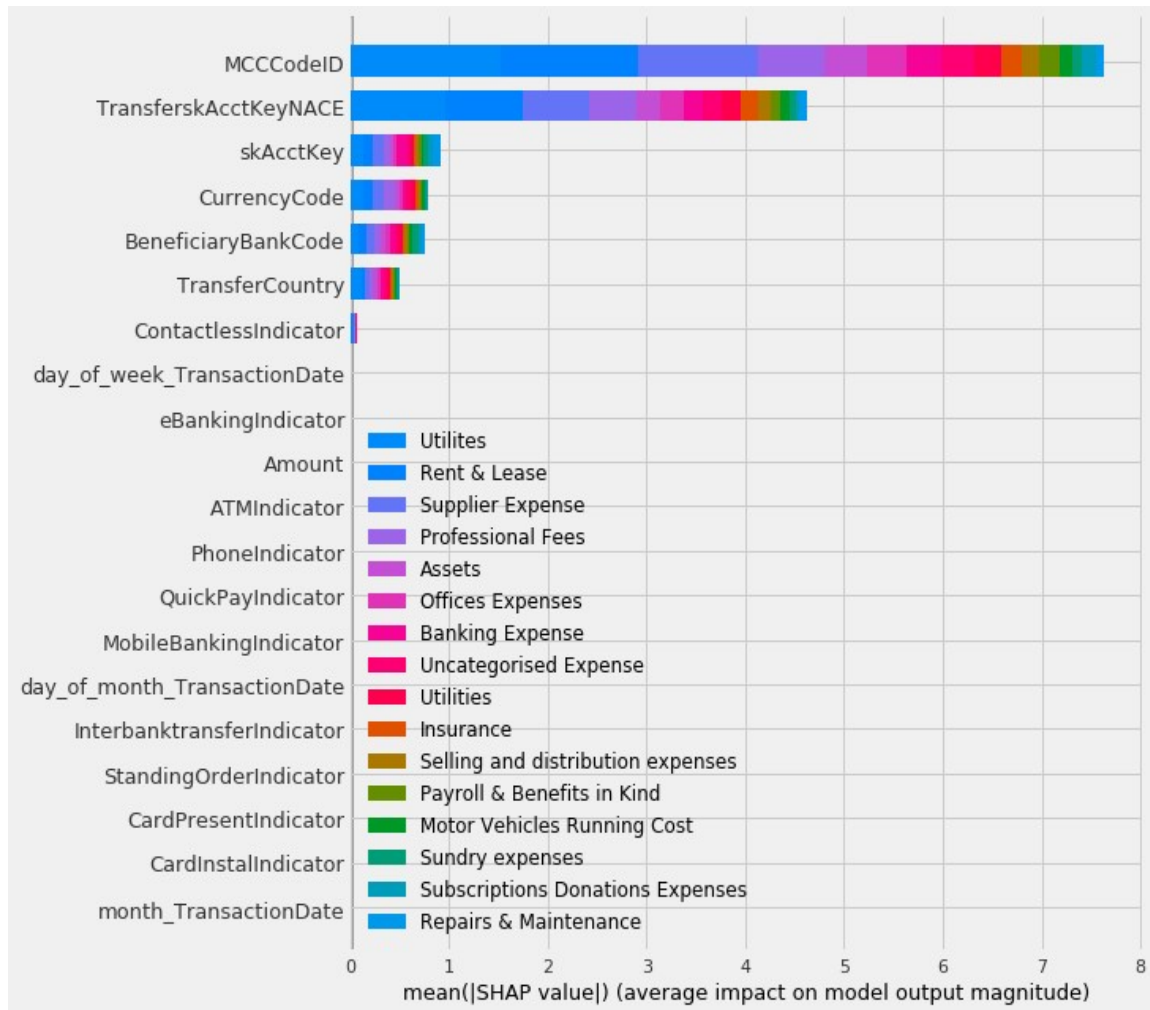


Πίνακας 26 Confusion Matrix της κατηγοριοποίησης συναλλαγών πολύκλασης ταξινόμησης

Predicted Values	Assets	566	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	Banking Expense	0	245	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	Insurance	0	0	1325	0	0	0	1	0	0	0	0	0	0	0	0	0	0
	Motor Vehicles Running Cost	0	0	0	1256	0	0	0	0	0	0	0	0	0	0	0	0	0
	Offices	0	0	0	0	15321	0	59	0	0	7	0	0	0	57	0	0	0
	Payroll Benefits in Klnd	0	0	0	0	0	958	1	0	0	0	0	0	0	0	0	0	0
	Professional Fees	0	0	0	0	0	0	31121	0	0	4	0	0	0	260	0	0	0
	Rent Lease	0	0	0	0	0	0	0	680	0	0	0	0	0	12	0	0	0
	Repairs Maintenance	0	0	0	0	0	0	0	0	53	53	0	0	0	0	0	0	0
	Selling and distribution	0	0	0	0	2	0	1	1	0	39751	0	0	0	1061	0	0	1
	Subscriptions Donations	0	0	0	0	0	0	0	0	0	0	734	0	0	23	0	0	0
	Supplier	0	0	0	0	0	0	0	0	0	0	0	1365	0	2	0	0	0
	Taxes	0	0	0	0	0	0	8	0	0	0	0	5	0	1	0	0	0
	Transfer	0	0	0	0	0	92	0	0	1	5	0	0	0	41177	0	0	34
	Utilities	0	0	0	0	0	0	0	0	0	0	0	0	0	5	0	0	0
	Uncategorized	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10124

Στο πλαίσιο της έρευνάς μας, η ερμηνεία των αποτελεσμάτων είναι πιο σημαντική από τα ίδια τα αποτελέσματα προκειμένου να συναχθούν συμπεράσματα τόσο σε τεχνικές όσο και σε επιχειρηματικές γνώσεις.

Στην περίπτωση κατηγοριοποίησης των συναλλαγών μας, τόσο οι τεχνικές LIME όσο και SHAP χρησιμοποιήθηκαν ως ποιοτική αξιολόγηση των αποτελεσμάτων. Οι τιμές SHAP που υποδηλώνουν τη σημασία κάθε χαρακτηριστικού που περιλαμβάνεται στο μοντέλο Catboost απεικονίζονται στην Εικόνα 55. Είναι εμφανές στο σχήμα ότι το μοντέλο έμαθε τους κανόνες που βάσιμα στο αναγνωριστικό κώδικα εμπόρου (δηλ. MCCCodeID) και τον κωδικό NACE δικαιούχου συναλλαγής ως αυτά τα δύο είναι τα πιο σημαντικά χαρακτηριστικά. Επιπλέον, η σημασία του κλειδιού λογαριασμού (δηλ. skAcctKey) ως το τρίτο πιο σημαντικό χαρακτηριστικό στο μοντέλο ενισχύει την προτεινόμενη μεθοδολογία μιας προσέγγισης ενημέρωσης προσανατολισμένη στον χρήστη.

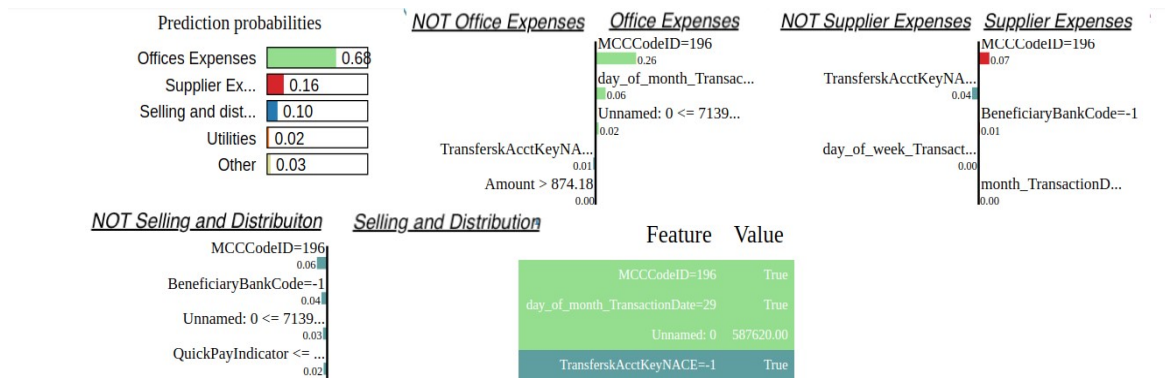


**Εικόνα 55** Τιμές SHAP που υποδεικνύουν τη σημασία χαρακτηριστικών της συναλλαγής Cathboost μοντέλο κατηγοριοποίησης

Τέλος, όσον αφορά την αξιολόγηση του ταξινομητή Catboost, εκτός από τη σημασία των χαρακτηριστικών που βασίζεται στην ανάλυση SHAP, είναι πολύ σημαντικό να ελεγχθούν ποιοτικά ορισμένα από τα αποτελέσματα που βασίζονται στην μέθοδο LIME, μια πρόσφατη τεχνική ικανή της εξήγησης των προβλέψεων οποιοδήποτε ταξινομητή ή παλινδρομητή χρησιμοποιώντας τοπικές προσεγγίσεις (βασισμένες σε δείγμα) μοντέλων με άλλα ερμηνεύσιμα μοντέλα. Στην Εικόνα 56 απεικονίζεται μία περιπτώσεις για το πώς το η βιβλιοθήκη LIME μπορεί να βοηθήσει στην ερμηνεία των προβλέψεων ενός μοντέλου ML. Για παράδειγμα, όπως φαίνεται στο σχήμα, αυτή η συναλλαγή κατηγοριοποιείται ως «Έξοδα Γραφείου» με πιθανότητα 68% και χαρακτηριστικά που συμβάλλουν σε αυτό το αποτέλεσμα είναι ο MCCCode, η ημέρα του μήνα. Ομοίως με πιθανότητα 16% μπορεί να κατηγοριοποιηθεί ως «Έξοδα Προμηθευτή» με βάση τον κωδικό MCC τον Λογαριασμό Δικαιούχου και τον Κωδικό Τράπεζας, ενώ με πιθανότητα 10% μπορεί

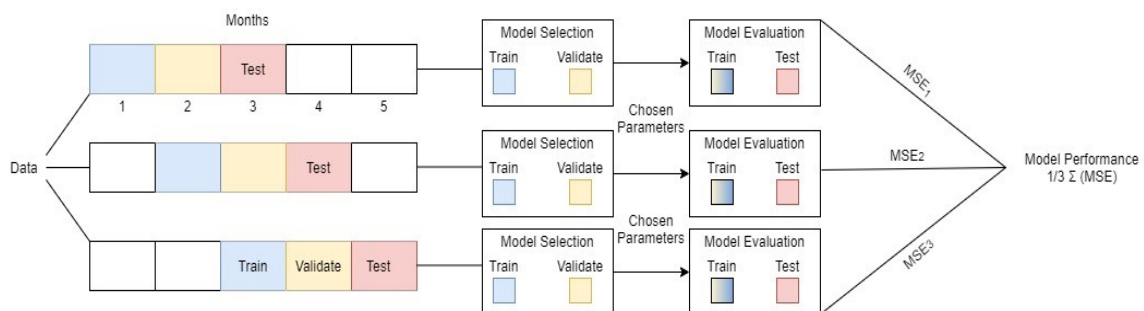


να ταξινομηθεί ως «Έξοδα Πώλησης και Διανομής», με τα χαρακτηριστικά να συμβάλλουν σε αυτές τις αποφάσεις. επίσης απεικονίζεται.



Εικόνα 56 Παραδείγματα κατηγοριοποίησης συναλλαγών της ανάλυσης LIME, εξηγώντας τα χαρακτηριστικά και τις αξίες τους συνεισφέρουν συγκεκριμένα στα αποτελέσματα του μοντέλου με τις παρεχόμενες πιθανότητες

### 4.4.2 Πρόβλεψη χρονοσειρών ταμειακών ροών



Εικόνα 57 Προτεινόμενο σχέδιο επικύρωσης ένα βήμα μπροστά

Ο στόχος αυτού του δεύτερου μέρους της έρευνάς μας είναι να προβλέψουμε με ακρίβεια τις ταμειακές εισροές και εκροές των συγκεκριμένων κατηγοριών που παράγονται από το υβριδικό μοντέλο κατηγοριοποίησης συναλλαγών μας για κάθε MME. Ο τρόπος με τον οποίο μοντελοποιήθηκε η εργασία και ο στόχος ανέδειξε την ανάγκη για μετασχηματισμό δεδομένων στην αναπαράσταση χρονοσειρών. Για το σκοπό αυτό, τόσο η επαναδειγματοληψία όσο και η συγκέντρωση του ποσού των εξόδων/εσόδων ανα κατηγορία ανα χρονικό διάστημα δημιούργησαν την αναπαράσταση των δεδομένων μας ως προς το χρόνο (χρονοσειρές).

Όπως σε κάθε εφαρμογή ερευνητικών διαδικασιών σε πραγματικά δεδομένα, προέκυψαν διάφορες προκλήσεις σχετικά με τις προβλέψεις χρονοσειρών που αντιμετωπίστηκαν στο πλαίσιο της έρευνάς μας. Αυτές οι προκλήσεις μπορούν να συνοψιστούν ως εξής:



α. Το πρόβλημα της «ψυχρής εκκίνησης», που στη συγκεκριμένη περίπτωση αναφέρεται στις χρονοσειρές που έχουν μικρό αριθμό συναλλαγών. Η «ψυχρή εκκίνηση» αναφέρεται στη δυσκολία απόκτησης ακριβών προβλέψεων για έναν νέο πελάτη. Όπως θα μπορούσε να εξεταστεί περαιτέρω σε μελλοντική έρευνα, για να δώσουμε μια ιδέα, μπορούμε να υποστηρίξουμε ότι σε τέτοιες περιπτώσεις όπου ένας νέος λογαριασμός πελάτη προστίθεται στη δεξαμενή λογαριασμών με βάση τα χαρακτηριστικά της υποκείμενης ΜΜΕ (δηλ. αριθμός εργαζομένων, κωδικός NACE κ.λπ.).

β. Η αντιστάθμιση σταθερότητας - εποχικότητας, όπου υποτίθεται ότι για να έχουμε προβλέψιμες χρονοσειρές πρέπει να είναι στάσιμες (δηλαδή χωρίς να υπάρχουν παράγοντες τάσης και εποχικότητας). Στο πλαίσιο αυτής της εργασίας, υποστηρίζεται ότι αυτή η πρόκληση έχει αντιμετωπιστεί μέσω της χρήσης του μοντέλου DeepAR. Όπως περιγράφεται, η συγκεκριμένη εφαρμογή του DeepAR που χρησιμοποιείται αντιμετωπίζει αυτούς τους τύπους προβλημάτων.

γ. Η παρουσία θορύβου στα δεδομένα ή/και πολλών παρατηρούμενων ακραίων τιμών. Η λύση σε αυτήν την πρόκληση προήλθε από την προσέγγιση επαναδειγματοληψίας που χρησιμοποιήθηκε. Το πεδίο εφαρμογής της επαναδειγματοληψίας ήταν κυρίως η μείωση της επίδρασης του θορύβου στο σύνολο δεδομένων, προσθέτοντας παράλληλα στο προτεινόμενο μοντέλο και διαδικασίες πρόσθετη αποτελεσματικότητα από άποψη χρόνου και πόρων. Το χρονικό πλαίσιο των «μηνών» επιλέχθηκε καθώς αντιστοιχεί στην ταχύτητα δεδομένων για τη δεδομένη περίπτωση - δηλαδή η στρατηγική των ΜΜΕ μπορεί να προβληθεί σε ορίζοντα 4 μηνών.

δ. Την επάρκεια του μήκους του συνόλου δεδομένων για την εφαρμογή τεχνικών μηχανικής και βαθιάς μάθησης. Η τελική πρόκληση προκαλείται ως παρενέργεια από το γεγονός ότι τα δεδομένα επαναδειγματίστηκαν σε εβδομαδιαίο χρονικό πλαίσιο λαμβάνοντας τη συγκέντρωση του Ποσού των συναλλαγών που έγιναν για μια συγκεκριμένη κατηγορία. Αυτή η διαδικασία καταλήγει στα συνολικά δεδομένα εκπαίδευσης που αποτελούνται από έναν πίνακα  $N \in R^{9494 \times 250}$ , όπου το "9494" αντιστοιχεί στον αριθμό των χρονοσειρών λογαριασμών ανά κατηγορία συναλλαγής και "250" στα χρονικά βήματα των τελευταίων 250 εβδομάδων καθεμιάς. Για να έχουμε πιο αξιόπιστα αποτελέσματα, εκπαιδεύσαμε κάθε μοντέλο 90 φορές σε ένα κυλιόμενο παράθυρο 160 χρονικών βημάτων που σχηματίζουν 90 πίνακες  $\in R^{9494 \times 250}$  (π.χ. [9494,1-160], [9494,2-161] κ.λπ.). Ενώ όλα τα μοντέλα μοιράζονται ορισμένες υπερπαραμέτρους όπως εποχές  $\leftarrow 10$ , το μήκος (πρόβλεψη)  $\leftarrow 12$  και το μήκος (πλαίσιο)  $\leftarrow 24$  μήνες. Αυτές οι συγκεκριμένες παράμετροι ρυθμίστηκαν με ακρίβεια χρησιμοποιώντας τεχνικές αναζήτησης πλέγματος και διαίσθηση, ως θεμελιώδες βήμα του αγωγού ML. Όσον αφορά τα υποκατάστατα δεδομένα, τα δεδομένα που χρησιμοποιήθηκαν για τη δημιουργία των ανάλογων δεδομένων είναι οι πρώτες 160 μονάδες, ενώ ο αριθμός των βημάτων αξιολόγησης παραμένει 90, καταλήγοντας σε 90 πίνακες  $\in R^{9494 \times 320}$ . Για





να αντιμετωπιστεί η επάρκεια διάρκειας συναλλαγής που απαιτείται για την εκτέλεση προβλέψεων, ορίστηκε ως όριο ένας ελάχιστος αριθμός συναλλαγών που εκτελέστηκαν κατά την εξεταζόμενη ιστορική περίοδο, με τη χρήση μόνο των χρονοσειρές που απεικονίζουν περισσότερες από δέκα (10) συναλλαγές.

ε. Στην κοινή χρήση ML, οι μέθοδοι διασταυρούμενης επικύρωσης αντικατοπτρίζουν μια παγίδα σε μια προσέγγιση πρόβλεψης χρονοσειρών, καθώς μπορεί να οδηγήσουν σε σημαντική επικάλυψη μεταξύ αμαξοστοιχίας και δεδομένων δοκιμής. Έτσι, η βέλτιστη προσέγγιση είναι η προσομοίωση μοντέλων σε μια ακολουθία «βήμα προς τα εμπρός», εκπαιδεύοντας περιοδικά το μοντέλο ώστε να ενσωματώνει συγκεκριμένα κομμάτια δεδομένων διαθέσιμα σε εκείνη τη χρονική στιγμή. Αυτή η διαδικασία απεικονίζεται στην Εικόνα 57.

Για την ανάπτυξη ενός αποτελεσματικού μοντέλου κατηγοριοποίησης συναλλαγών, εφαρμόστηκαν και αξιολογήθηκαν διάφορα μοντέλα τελευταίας τεχνολογίας, με τη σύγκριση που φαίνεται στον Πίνακα 25. Είναι προφανές ότι το XGboost ξεπερνά την απόδοση του μοντέλου Catboost, ωστόσο, καθώς ο χρόνος εκπαίδευσης είναι ζωτικής σημασίας για την ανάπτυξη ενός μοντέλου σε πραγματικές συνθήκες.

Στον Πίνακα 27 παρουσιάζεται η απόδοση του μοντέλου πρόβλεψης ταμειακών ροών με βάση διάφορες μετρήσεις αξιολόγησης. Αναλυτικά, εποχιακό σφάλμα, MASE, sMAPE, MSIS και Quantile Loss, ενώ παρουσιάζεται και η Κάλυψη που παρέχεται από διαφορετικά επίπεδα εμπιστοσύνης. Τα αποτελέσματα αναφέρονται σε μέσες τιμές για όλους τους ενσωματωμένους λογαριασμούς ανά κατηγορία συναλλαγής. Επιπλέον, τα σχετικά διαγράμματα απεικονίζονται στην Εικόνα 58 απεικονίζοντας τους εκτιμητές σε συγκεκριμένες χρονοσειρές. Όπως ήταν αναμενόμενο, το εποχικό σφάλμα της Κατηγορίας Συνδρομών και Δωρεών είναι το χαμηλότερο και ακολουθούν τα Ασφαλιστικά έξοδα. Από την άλλη πλευρά, τα μη κατηγοριοποιημένα έξοδα αποδίδουν το υψηλότερο εποχικό σφάλμα, υποδηλώνοντας ότι υπάρχει χαμηλός εποχικός παράγοντας σε αυτού του είδους τα δαπάνες.

Τα αποτελέσματα που παρουσιάζονται υπογραμμίζουν ότι η προτεινόμενη μεθοδολογία παράγει αποτελέσματα που δύσκολα μπορούν να επιτύχουν την απαιτούμενη προσέγγιση με βάση την Κάλυψη. Με πιθανολογικό τρόπο, αυτό μεταφράζεται ότι για ένα δεδομένο επίπεδο εμπιστοσύνης π.χ. 80%, το μοντέλο παρουσιάζει χαμηλότερες επιδόσεις δεδομένων των μέσων τιμών των κατηγοριών.

Επιπλέον, στον Πίνακα 28 παρουσιάζεται η απόδοση του μοντέλου πρόβλεψης ταμειακών ροών που εφαρμόζεται στα ενισχυμένα ανάλογα δεδομένα ως προς τον ίδιο.



Πίνακας 27 Μέσα αποτέλεσμα ανά Κατηγορία Συναλλαγής με βάση διάφορες μετρήσεις.

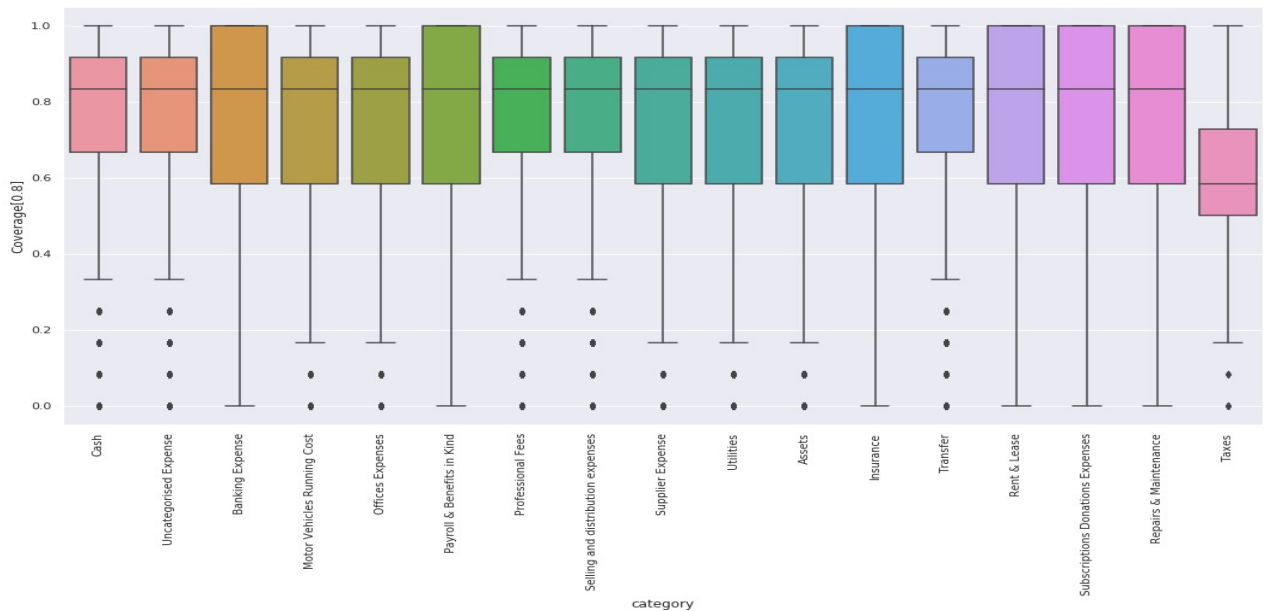
<i>Category (expenses)</i>	<i>seasonal error</i>	<i>MASE</i>	<i>sMAPE</i>	<i>MSIS</i>	<i>Quantile Loss [0.5]</i>	<i>Coverage [0.5]</i>	<i>Quantile Loss [0.8]</i>	<i>Coverage [0.8]</i>	<i>Quantile Loss [0.95]</i>	<i>Coverage [0.95]</i>
<b>Assets</b>	425.004	0.955	1.940	26.064	2401.091	0.466	2900.684	0.722	2367.565	0.871
<b>Banking</b>	694.784	72.957	1.910	2478.095	3340.891	0.477	4437.331	0.746	4268.931	0.891
<b>Cash</b>	4065.432	0.787	1.658	18.493	25428.835	0.515	31496.964	0.755	26010.145	0.882
<b>Insurance</b>	156.242	1.018	1.921	23.230	964.912	0.464	1150.166	0.724	908.232	0.872
<b>Motor Vehicles</b>	600.610	3.625	1.702	50.167	3599.007	0.442	3863.190	0.730	2646.931	0.880
<b>Running Cost</b>	249.595	3.975	1.829	57.882	2379.410	0.470	3007.474	0.736	2660.359	0.879
<b>Offices</b>	382.469	1.972	1.940	46.086	3326.830	0.468	4666.329	0.730	4582.028	0.874
<b>Payroll &amp; Benefits in Kind</b>	2117.501	7.830	1.818	265.033	16769.57	0.480	19810.913	0.743	15255.594	0.884
<b>Professional Fees</b>	803.825	0.938	1.944	15.177	4093.789	0.465	4920.855	0.709	3708.97	0.856
<b>Rent &amp; Lease</b>	210.978	7.209	1.945	54.962	1815.282	0.487	2171.715	0.740	1693.385	0.879
<b>Repairs &amp; Maintenance</b>	773.484	5.174	1.683	110.461	4951.311	0.4901	5679.797	0.751	4288.930	0.887
<b>Selling and distribution</b>	74.533	25.023	1.920	215.479	916.44	0.472	1052.954	0.7314	776.814	0.880
<b>Subscriptions</b>	2107.319	2.459	1.849	44.455	18531.24	0.463	19777.18	0.726	14281.448	0.872
<b>Donations</b>	2786.071	0.403	1.825	6.073	13210.548	0.366	12938.64	0.5712	8515.301	0.801
<b>Supplier</b>	8942.757	3.394	1.577	116.278	57575.255	0.494	63056.221	0.751	43652.938	0.888
<b>Taxes</b>	11734.934	2.673	1.634	59.273	72403.651	0.483	76042.949	0.741	49743.826	0.881
<b>Transfer</b>	291.783	4.429	1.848	76.449	1855.911	0.468	2126.565	0.731	1604.632	0.878
<b>Uncategorised</b>										
<b>Utilities</b>										

Πίνακας 28 Μέσα αποτέλεσμα ανά Κατηγορία Συναλλαγής με βάση διάφορες μετρήσεις με χρήση ανάλογων/υποκατάστατων δεδομένων.

<i>Category (expenses)</i>	<i>Seasonal error</i>	<i>MASE</i>	<i>sMAPE</i>	<i>MSIS</i>	<i>Quantile Loss [0.5]</i>	<i>Coverage [0.5]</i>	<i>Quantile Loss [0.8]</i>	<i>Coverage [0.8]</i>	<i>Quantile Loss [0.95]</i>	<i>Coverage [0.95]</i>
<b>Assets</b>	373.534	1.343	1.980	42.724	2811.85	0.466	3896.822	0.815	3899.518	0.915
<b>Banking</b>	640.346	69.007	1.949	2511.932	3172.123	0.47	4609.697	0.847	5036.039	0.924
<b>Cash</b>	3783.599	1.037	1.752	31.832	24129.639	0.452	32013.991	0.737	29854.857	0.862
<b>Insurance</b>	129.248	1.53	1.971	45.973	1165.007	0.464	1511.745	0.799	1477.439	0.908
<b>Motor Vehicles</b>	527.229	2.873	1.841	62.53	4402.294	0.428	6080.678	0.737	5910.548	0.855
<b>Running Cost</b>	204.590	2.636	1.906	68.333	2635.013	0.457	3563.313	0.792	3514.37	0.881
<b>Offices</b>	342.144	2.471	1.9811	67.818	3503.126	0.474	5015.94	0.822	5383.268	0.917
<b>Payroll &amp; Benefits in Kind</b>	1589.156	13.40	1.887	475.538	15838.086	0.479	21370.288	0.789	19950.472	0.882
<b>Professional Fees</b>	753.627	1.344	1.987	27.806	4977.029	0.474	7129.604	0.789	7523.01	0.900
<b>Rent &amp; Lease</b>	168.585	2.004	1.978	28.57	1541.248	0.468	2107.157	0.814	2150.599	0.925
<b>Repairs &amp; Maintenance</b>	637.919	5.337	1.783	163.050	5665.68	0.450	7665.938	0.757	7299.181	0.86
<b>Selling and distribution</b>	73.122	10.053	1.962	135.738	809.495	0.470	989.008	0.816	917.352	0.915
<b>Subscriptions</b>	1641.695	2.608	1.925	70.391	25259.910	0.455	31190.551	0.779	26794.230	0.882
<b>Donations</b>	3139.324	0.431	1.955	12.492	16307.595	0.415	23018.600	0.662	22284.277	0.742
<b>Supplier</b>	7650.145	6.087	1.697	224.617	66887.980	0.442	78911.379	0.715	64099.914	0.842
<b>Taxes</b>	10193.153	3.110	1.759	97.321	84007.558	0.436	95502.407	0.714	72415.80	0.845
<b>Transfer</b>	243.400	4.33	1.923	116.887	2001.960	0.465	2730.252	0.788	2622.102	0.880
<b>Uncategorised</b>										
<b>Utilities</b>										



Συγκρίνοντας τα αποτελέσματα μεταξύ των Πίνακας 27 και Πίνακας 28, τα κύρια αποτελέσματα είναι ότι το εποχικό σφάλμα είναι χαμηλότερο σε όλες τις κατηγορίες. Αυτό μπορεί να εξηγηθεί καθώς τα ανάλογα δεδομένα διατηρούν την εποχικότητα παράγοντα αφού βασίζονται στον μετασχηματισμό Fourier, τονίζοντας την εποχιακή επιρροή. Αυτό το επιχείρημα υποστηρίζεται επίσης από το γεγονός ότι, όσον αφορά το MAPE και το MSE, τα αποτελέσματα δεν είναι καλύτερα, πράγμα που σημαίνει ότι η μέση τιμή της προβλεπόμενης κατανομής δεν συγκλίνει έναντι των πραγματικών μελλοντικών τιμών. Τέλος, μπορούμε να εκφράσουμε συγκρατημένη αισιοδοξία για τη χρήση ανάλογων δεδομένων ως τεχνική εμπλουτισμού, που θα διερευνηθεί περαιτέρω, καθώς, όσον αφορά την κάλυψη, αυτή η προσέγγιση αποδίδει πολλά υποσχόμενα αποτελέσματα, βελτιστοποιώντας τα αποτελέσματα 14 από τις 17 κατηγορίες όταν λαμβάνεται υπόψη η κάλυψη του 80%. , ενώ σε κάλυψη 95% 10 στα 17 απεικονίζουν καλύτερα αποτελέσματα.



**Εικόνα 58** Βοχplot της κάλυψης 0,8, οι περισσότερες από τις κατηγορίες έχουν υψηλή διακύμανση ενώ μόνο ο μέσος όρος της κατηγορίας φόρων φαίνεται να υπολειτουργεί

Η πρόβλεψη πολλαπλών βημάτων συνίσταται στην πρόβλεψη των επόμενων τιμών της χρονοσειράς και εκτελείται με δύο διαφορετικούς τρόπους. Η πρώτη, ανεξάρτητη πρόβλεψη αξίας, αποτελείται από την εκπαίδευση ενός άμεσου μοντέλου για την πρόβλεψη των ακριβών βημάτων μπροστινής αξίας. Η δεύτερη, που ονομάζεται επαναληπτική μέθοδος, αποτελείται από την επανάληψη των προβλέψεων ενός βήματος μπροστά στον επιθυμητό ορίζοντα.

Η επαναληπτική πρόβλεψη χρησιμοποιεί μόνο ένα μοντέλο για να προβλέψει όλους τους απαιτούμενους ορίζοντες. Ο στόχος είναι να αναλυθεί μια σύντομη ακολουθία δεδομένων ενώ προσπαθεί να προβλέψει την υπόλοιπη ακολουθία δεδομένων μέχρι να επιτευχθεί ένα προκαθορισμένο χρονικό βήμα. Το κύριο



μειονέκτημα αυτής της προσέγγισης είναι η συσσωρευτική φύση του σφάλματος. Η εφαρμογή ενός μοντέλου DeepAR στην προσέγγισή μας εμπίπτει σε αυτήν την τελευταία περίπτωση, αποδίδοντας υψηλότερες τιμές σφάλματος όταν διευρύνεται ο ορίζοντας πρόβλεψης.



**Εικόνα 59** Ενδεικτικά παραδείγματα εκτιμητή σε συγκεκριμένες χρονοσειρές, σε κάθε χρόνο διάγραμμα η μπλε γραμμή αντιπροσωπεύει τις πραγματικές τιμές ενώ η πράσινη κλίση απεικονίζει τις προβλέψεις σε επίπεδα εμπιστοσύνης 50% και 95%.

### 4.5 Επεξηγήσιμη Τεχνητή Νοημοσύνη

Διεξάγουμε μια αξιολόγηση της απόδοσης του μοντέλου μας μέσω μιας εργασίας ταξινόμησης χρησιμοποιώντας ένα χωρισμένο σύνολο δεδομένων σε υποσύνολα εκπαίδευσης, επικύρωσης και δοκιμής. Η παρούσα εργασία



περιλαμβάνει την ταξινόμηση ενός συνόλου χρονοσειρών σε μία από τις δύο κατηγορίες. Το μοντέλο απόσταξης μας αποτελείται από ένα δίκτυο δασκάλων - ένα πυκνό δίκτυο - και ένα δίκτυο μαθητών - ένα CNN. Η διαδικασία απόσταξης στοχεύει στη μεταφορά της γνώσης του δικτύου εκπαιδευτικών στο δίκτυο των μαθητών, αποδίδοντας έτσι ένα ερμηνεύσιμο και πιο βελτιωμένο μοντέλο με συγκρίσιμες ή ανώτερες επιδόσεις. Ωστόσο, η απόσταξη μοντέλου έχει τους περιορισμούς της, συμπεριλαμβανομένου του ότι η απλοποίηση ενός σύνθετου μοντέλου σε ένα πιο ερμηνεύσιμο μπορεί να μην είναι πάντα εφικτή. Κατά συνέπεια, παρουσιάζουμε τα αποτελέσματα για το δίκτυο δασκάλου-μαθητή, του δάσκαλου μόνος και του μαθητή μόνος.

Το σύνολο δεδομένων εκπαίδευσης χρησιμοποιείται για την εκπαίδευση των μοντέλων και ένα σύνολο δεδομένων επικύρωσης βοηθά στον συντονισμό των υπερπαραμέτρων του δικτύου, όπως ο ρυθμός εκμάθησης και ο αριθμός των επιπέδων. Η πρώτη διακοπή ξεκινά εάν δεν υπάρξει βελτίωση στην ακρίβεια επικύρωσης σε συγκεκριμένο αριθμό εποχών. Τελικά, η απόδοση των δικτύων δασκάλων και μαθητών αξιολογείται στο σύνολο δεδομένων δοκιμών.

Αντιπαραθέτουμε τις επιδόσεις του δικτύου καθηγητών, του δικτύου των μαθητών και του «αποσταγμένου» μοντέλου μας. Αναφέρονται η ακρίβεια, η ακρίβεια, η ανάκληση και η F1-score για κάθε μοντέλο. Επιπλέον, επισημαίνουμε την ερμηνευτικότητα του μαθητικού δικτύου χρησιμοποιώντας μεθόδολογίες LIME και GradCAM που επισημαίνουν σημαντικά τμήματα των εικόνων εισόδου που επηρεάζουν την απόφαση ταξινόμησης.

Τα μοντέλα μας υλοποιούνται και εκπαιδεύονται χρησιμοποιώντας το πλαίσιο βαθιάς εκμάθησης TensorFlow σε περιβάλλον με 20 πυρήνες CPU στα 2,3 GHz και 64 GB μνήμης RAM. Ένα εύρος τιμών υπόκειται σε αναζήτηση πλέγματος για την επιλογή των υπερπαραμέτρων, με το καλύτερο σύνολο να προσδιορίζεται με βάση την ακρίβεια επικύρωσης. Τα μοντέλα εκπαιδεύονται έως και 300 εποχές κατ' ανώτατο όριο, ενώ η πρόωρη διακοπή ενεργοποιείται εάν η ακρίβεια επικύρωσης δεν δείξει βελτίωση σε διάστημα 20 εποχών. Για αναπαραγωγιμότητα, ο κώδικας και τα εκπαιδευμένα μοντέλα γίνονται δημόσια προσβάσιμα. Επιπλέον, διενεργούμε μια ποιοτική αξιολόγηση της ερμηνευσιμότητας των αποτελεσμάτων XAI της προσέγγισής μας. Τέλος, παρέχουμε μια πιο αναλυτική άποψη της απόδοσης της μεθόδου από άποψη χρόνου. Τα αποτελέσματα που παρουσιάζονται είναι ο μέσος όρος δέκα ανεξάρτητων δοκιμών για κάθε μέτρηση.

### 4.5.1 Σύνολα Δεδομένων

Ως μέρος της έρευνάς μας, χρησιμοποιήσαμε σύνολα δεδομένων που προέρχονται από το Αρχείο UCR, το οποίο αποτελείται από δεδομένα χρονοσειρών



με ετικέτα. Συγκεκριμένα, χρησιμοποιήσαμε δύο από τα διαθέσιμα σύνολα δεδομένων που έχουν περισσότερα από 1000 δείγματα για εκπαίδευση και δοκιμή:

- **Wafer:** Το σύνολο δεδομένων wafer διατυπώθηκε από τον R. Olszewski ως μέρος της διατριβής του, *Generalized Feature Extraction for Structural Pattern Recognition in Time-Series Data*, στο Πανεπιστήμιο Carnegie Mellon το 2001. Αυτό το σύνολο δεδομένων αφορά τον τομέα της κατασκευής μικροηλεκτρονικών ημιαγωγών. Αποτελείται από μετρήσεις ελέγχου διεργασίας που συλλέγονται από διάφορους αισθητήρες κατά την επεξεργασία πλακών πυριτίου για την κατασκευή ημιαγωγών. Κάθε σύνολο δεδομένων στη βάση δεδομένων πλακιδίων περιέχει τις μετρήσεις που καταγράφονται από έναν μόνο αισθητήρα κατά την επεξεργασία μιας πλακέτας από ένα εργαλείο. Τα δεδομένα μπορούν να ταξινομηθούν σε δύο κατηγορίες: φυσιολογικά και μη φυσιολογικά. Υπάρχει μια σημαντική ανισορροπία κατηγορίας μεταξύ της κανονικής και της μη φυσιολογικής κατηγορίας, με το 10,7% των δεδομένων εκπαίδευσης και το 12,1% των δεδομένων δοκιμής να εμπίπτουν στην ανώμαλη κατηγορία. Η καλύτερη ακρίβεια μέχρι σήμερα είναι 99,8%, που επιτυγχάνεται με τον αλγόριθμο ST.

- **FordA:** Αυτό το σύνολο δεδομένων χρησιμοποιήθηκε αρχικά σε έναν διαγωνισμό στο Παγκόσμιο Συνέδριο IEEE για την Υπολογιστική Νοημοσύνη το 2008. Το πρόβλημα ταξινόμησης περιλαμβάνει τη διάγνωση εάν ένα συγκεκριμένο σύμπτωμα υπάρχει σε ένα υποσύστημα αυτοκινήτου. Κάθε θήκη περιλαμβάνει 500 μετρήσεις θορύβου κινητήρα και ταξινόμηση. Το σύνολο δεδομένων περιλαμβάνει δύο ξεχωριστά προβλήματα: Για το FordA, τόσο τα σύνολα δεδομένων εκπαίδευσης όσο και δοκιμής συλλέχθηκαν υπό τυπικές συνθήκες λειτουργίας με ελάχιστη μόλυνση από θόρυβο. Η καλύτερη ακρίβεια μέχρι σήμερα είναι 96,54%, που επιτυγχάνεται με τον αλγόριθμο ST.

#### 4.5.2 Αποτελέσματα

Το δίκτυο Dense, τόσο για τα σύνολα δεδομένων FordA όσο και για το Wafer, παρήγαγε τη χαμηλότερη ακρίβεια και βαθμολογία F1 για την κλάση 1, υποδεικνύοντας τη σχετική αναποτελεσματικότητά του στον σωστό εντοπισμό περιπτώσεων της κατηγορίας μειοψηφίας σε αυτά τα μη ισορροπημένα σύνολα δεδομένων. Αυτό θα μπορούσε να οδηγήσει σε υψηλότερο ποσοστό ψευδώς θετικών για αυτήν την κατηγορία. Από την άλλη πλευρά, το CNN και τα προτεινόμενα μοντέλα πέτυχαν τέλεια ακρίβεια και βαθμολογίες F1 για την κλάση 0 και στα δύο σύνολα δεδομένων. Αυτό υποδηλώνει την ικανότητά τους να ταξινομούν με ακρίβεια παρουσίες της κλάσης πλειοψηφίας χωρίς να επισημαίνουν εσφαλμένα στιγμιότυπα από άλλες κλάσεις ότι ανήκουν στην κλάση 0. Είναι ενδιαφέρον ότι το



προτεινόμενο μοντέλο ξεπέρασε τις επιδόσεις τόσο του πυκνού δικτύου όσο και του CNN όσον αφορά την ταξινόμηση των περιπτώσεων της κλάσης 1 (την κατηγορία μειοψηφίας) και για τα δύο σύνολα δεδομένων, όπως αποδεικνύεται από την υψηλότερη ακρίβεια και τις βαθμολογίες F1. Αυτό υποδηλώνει ότι το προτεινόμενο μοντέλο είναι καλύτερα εξοπλισμένο για να χειρίζεται μη ισορροπημένα σύνολα δεδομένων, εντοπίζοντας με μεγαλύτερη ακρίβεια περιπτώσεις της μειοψηφίας και μειώνοντας τις πιθανότητες εσφαλμένης ταξινόμησης. Εξετάζοντας τη συνολική ακρίβεια, το δίκτυο Dense είχε χαμηλή απόδοση στο σύνολο δεδομένων FordA, ενώ το CNN έδωσε τη χαμηλότερη ακρίβεια στο σύνολο δεδομένων Wafer. Αντίθετα, το προτεινόμενο μοντέλο εμφάνισε την υψηλότερη ακρίβεια και στα δύο σύνολα δεδομένων, ενισχύοντας τη συνολική του αποτελεσματικότητα και την ικανότητα γενίκευσης σε διαφορετικά σύνολα δεδομένων χρονοσειρών.

Πίνακας 29 Αποτελέσματα των εργασιών ταξινόμησης με διαφορετικά μοντέλα

Model	(FordA) Dense	(FordA) CNN	(FordA) Proposed
Precision (Class 0)	0.86	0.70	0.86
Precision (Class 1)	0.82	0.75	0.84
F1-Score (Class 0)	0.84	0.73	0.86
Accuracy	0.84	0.73	0.85
Model	(Wafer) Dense	(Wafer) CNN	(Wafer) Proposed
Precision (Class 0)	1.0	1.0	1.0
Precision (Class 1)	0.98	1.0	0.99
F1-Score (Class 0)	1.0	1.0	1.0
Accuracy	1.0	1.0	1.0

Πίνακας 30 Χρόνος εκτέλεσης κάθε μοντέλου για την εκπαιδευτική εργασία καθώς και πόσες εποχές έκανε κάθε μοντέλο να συγκλίνει. Η στήλη RAM αναφέρεται στα MB που απαιτούνται για τη φόρτωση των συνόλων δεδομένων εκπαίδευσης στη μνήμη RAM

Model	Training time (sec)	Early stopping (epochs)	Train-set RAM (MB)
Dense (FordA)	31	67	1.16
CNN (FordA)	643	76	1148.44
Dense + CNN (FordA)	5106	102	1149.6
Dense (Wafer)	186	300	13.74
CNN (Wafer)	3539	300	4135.52
Dense + CNN (Wafer)	3048	257	4149.26

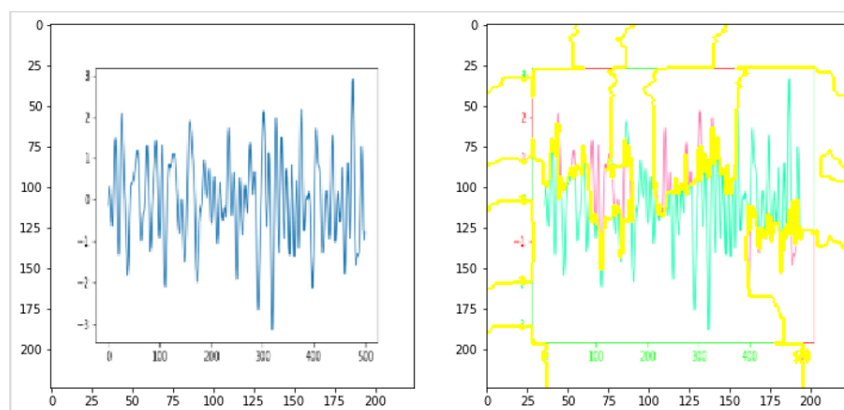
Η απόδοση του χρόνου παίζει πράγματι σημαντικό ρόλο στην αποδοτικότητα και τη σχέση κόστους-αποτελεσματικότητας των συστημάτων AI. Τα μοντέλα που χρειάζονται περισσότερο χρόνο για την εκτέλεση και την επεξεργασία δεδομένων



ενδέχεται να απαιτούν περισσότερους υπολογιστικούς πόρους, αυξάνοντας το λειτουργικό κόστος. Επομένως, η βελτιστοποίηση μοντέλων τεχνητής νοημοσύνης για χρονική απόδοση είναι μια κρίσιμη πτυχή της ανάπτυξης και ανάπτυξης της τεχνητής νοημοσύνης.

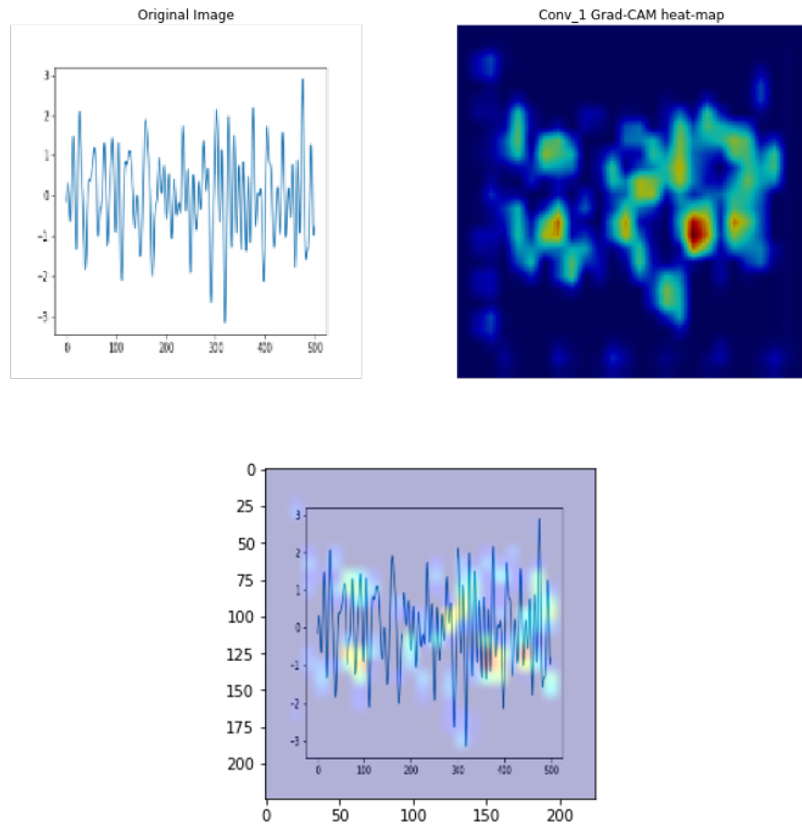
Ο Πίνακας 30 παρουσιάζει τον χρόνο εκτέλεσης κάθε μοντέλου κατά τη διάρκεια της φάσης εκπαίδευσης καθώς και τον αριθμό των εποχών που απαιτούνται για τη σύγκλιση κάθε μοντέλου. Λάβετε υπόψη ότι οι διαμορφώσεις του μοντέλου προσδιορίστηκαν όπως αναφέρθηκε στην προηγούμενη υποενότητα. Αξίζει να σημειωθεί ότι όσον αφορά τον χρόνο συμπερασμάτων (δηλαδή, τον χρόνο που χρειάζεται για το μοντέλο να δημιουργήσει μια πρόβλεψη για μία μόνο είσοδο), δεν υπάρχει σημαντική διαφορά μεταξύ των μοντέλων.

Όσον αφορά την πτυχή της «εξηγησιμότητας», το μοντέλο έχει σχεδιαστεί για να ενσωματώνει τη δημιουργία χαρτών θερμότητας LIME και Grad-Cam, διευκολύνοντας την ερμηνευτικότητα για χρήστες που μπορεί να μην έχουν εξειδίκευση στην επιστήμη δεδομένων. Το LIME προσφέρει πληροφορίες σχετικά με το γιατί ένα μοντέλο μηχανικής μάθησης έχει κάνει μια συγκεκριμένη πρόβλεψη, επισημαίνοντας τα πιο σημαντικά μέρη των δεδομένων εισόδου που το μοντέλο έλαβε υπόψη κατά την πραγματοποίηση της πρόβλεψης. Από την άλλη πλευρά, το Grad-CAM δημιουργεί έναν θερμικό χάρτη που απεικονίζει τις περιοχές μιας εικόνας που ήταν πιο κρίσιμες για ένα μοντέλο μηχανικής μάθησης να κάνει μια πρόβλεψη. Αυτή η οπτικοποίηση μας επιτρέπει να δούμε σε ποια μέρη της εικόνας το μοντέλο έδωσε προτεραιότητα όταν κάνει την πρόβλεψή του. Οι Εικόνες 57 και 59 παρέχουν μερικά αντιπροσωπευτικά παραδείγματα αυτών των απεικονίσεων XAI.



**Εικόνα 60** Παράδειγμα σημαντικών superpixel LIME όπου μπορούμε να παρατηρήσουμε τις κόκκινες περιοχές που δεν φαίνονται σχετικές με την προβλεπόμενη κατηγορία





**Εικόνα 61** Παράδειγμα του χάρτη θερμότητας Grad-Cam και της επικάλυψης του χάρτη θερμότητας στην αρχική εικόνα όπου μπορούμε να παρατηρήσουμε τις περιοχές της εικόνας που εστίασε το μοντέλο



## 5 ΚΕΦΑΛΑΙΟ - ΣΥΜΠΕΡΑΣΜΑΤΑ

### ΠΕΡΙΕΧΟΜΕΝΑ

5.1 Σύνοψη - Συμπεράσματα

5.2 Μηχανική Μάθηση & προβλεπτική συντήρηση κύριων μηχανών εμπορικών πλοίων.

5.3 Μηχανική Μάθηση & ασφάλεια των τροφίμων.

5.4 Μηχανική Μάθηση & ποιότητα χοιρινού κρέατος.

5.5 Μηχανική Μάθηση & τραπεζικός βοηθός για μικρομεσαίες επιχειρήσεις.

### ΠΕΡΙΛΗΨΗ

Στο **ΚΕΦΑΛΑΙΟ 5**, πραγματοποιείται η σύνοψη της παρούσας Διδακτορικής διατριβής, καταγράφοντας τις βασικές συνεισφορές της αναφορικά με την εφαρμογή μεθόδων μηχανικής μάθησης σε πραγματικά σενάρια της βιομηχανίας. Επιπλέον, περιγράφονται τα ανοιχτά ερευνητικά θέματα και οι μελλοντικοί στόχοι όπως αυτοί προέκυψαν από την τρέχουσα προσέγγιση της ερευνάς, αναφορικά με την ένταξη και την αντιμετώπιση των υπάρχουσών προκλήσεων της εφαρμογής αντίστοιχων τεχνολογιών σε όλους τους τομείς γενικότερα, αλλά και στους τομείς της ναυτιλίας, της ασφάλεια τροφίμων και του τραπεζικού κλάδου ειδικότερα

### 5.1 Σύνοψη – Συμπεράσματα

Στην ανα χείρας Διατριβή εξετάσαμε εκτενώς πως μπορούν καινοτόμες μέθοδοι μηχανικής και βαθιάς μάθησης να βελτιώσουν ή και να επιλύσουν προβλήματα σε διάφορους βιομηχανικούς τομείς. Η βασική ιδέα (ο κορμός) της έρευνας ξεκίνησε γύρω από την πρόβλεψη χρονοσειρών με εφαρμογή σε πραγματικά δεδομένα. Κατά την πορεία της έρευνας διάφοροι άλλοι τομείς εξερευνήθηκαν καθώς έπρεπε να αντιμετωπίσουμε τις προκλήσεις που ενέχει η χρήση πραγματικών δεδομένων. Ως εκ τούτου καταλήξαμε με μία αρχιτεκτονική που αποτελείται από 3 βασικά στοιχεία και αρκετά περιφερειακά. Πιο συγκεκριμένα σχεδόν σε κάθε πρόβλημα πρέπει να γίνει προ-επεξεργασία των δεδομένων, εν συνεχεία συνήθως απαιτείται ταξινόμηση των δεδομένων και ακολουθεί η πρόβλεψη χρονοσειρών. Περιφερειακά, εργαλεία όπως η χρήση αναλογικών δεδομένων και επεξηγησιμότητας τεχνητής νοημοσύνης.



## 5.2 Μηχανική Μάθηση & προβλεπτική συντήρηση κύριων μηχανών εμπορικών πλοίων.

Στη συγκεκριμένη περίπτωση, παρουσιάσαμε και εξετάσαμε μία σημαντική και αναπτυσσόμενη πρόκληση με την βοήθεια της οποίας οι ναυτιλιακές εταιρίες μπορούν να ωφεληθούν σημαντικά από τις σύγχρονες τεχνολογίες του εντοπισμού ανωμαλιών. Εφαρμόσαμε μερικές σύγχρονες προσεγγίσεις στον τομέα του εντοπισμού βλαβών ενώ ταυτόχρονα θέτουμε σημαντικές προκλήσεις που περιέχουν επεξήγηση, κλιμάκωση, ακρίβεια και πολυπλοκότητα που είναι σύμφυτες πολλών περιπτώσεων εντοπισμού ανωμαλιών. Συγκρίναμε και συνδυάσαμε προσεγγίσεις που βασίζονται σε βαθιά νευρωνικά δίκτυα LSTM, SVM, Ανύψωση Κλίσης και Στατιστικής. Το συμπληρωματικό μοντέλο αποτελείτο από ένα συνδυασμό των προσεγγίσεων, υποσχόμενο καλύτερο αποτέλεσμα όσον αφορά στις μετρικές αξιολόγησης, έχοντας θέσει το παράθυρο “διάβρωσης” στις 30 ημέρες πριν από κάθε ελάττωμα. Η προσέγγιση αξιολογήθηκε, δίνοντας υποσχόμενα αποτελέσματα και θα βοηθήσει τους μηχανικούς να εντοπίσουν βλάβες στο ρουλεμάν σταυρού κεφαλής στην κύρια μηχανή των πλοίων. Ακόμα, εστίασαμε στο να προσπαθήσουμε να προβλέψουμε την πιθανότητα εντοπισμού της βλάβης που ανιχνεύθηκε σε έναν συγκεκριμένο κύλινδρο. Δυστυχώς η πληροφορία που αντλήθηκε από το δοσμένο σύνολο δεδομένων ήταν ανεπαρκής, καθώς οι αλλαγές στις ροές των δεδομένων που προέρχονταν από διαφορετικούς κυλίνδρους ήταν σχεδόν πανομοιότυπες, ενώ οι ελάχιστες αποκλίσεις που βρέθηκαν δεν μπορούν να ερμηνεύσουν την υπό μελέτη βλάβη. Επίσης πρέπει να αναφερθεί ότι οι παράμετροι των μοντέλων δεν ήταν ρυθμισμένη ρητά στο βάρος, την ακρίβεια και την ανάκληση. Με άλλα λόγια τα μοντέλα έχουν την τάση να προβλέπουν τα θετικά σε μία περίπτωση που οι ψευδείς συναγερμοί δεν είναι επιβλαβείς, πράγμα που σημαίνει ότι οι περισσότερες βλάβες μπορούν να εντοπιστούν νωρίς αλλά πολλές φυσιολογικές συνθήκες να εκληφθούν ως ανωμαλίες. Η διαδικασία αυτή θα πρέπει να συζητηθεί επαρκώς με την ναυτιλιακή εταιρεία, καθώς θα πρέπει αυτές να σταθμίσουν τα λειτουργικά κόστη και τα κόστη συντήρησης. Πρόσθετη εργασία στο μέλλον θα μπορούσε να εστιάσει στο να βελτιώσει την πρόβλεψη, πρωτίστως μέσω της βελτιούμενης μηχανικής. Ακόμα, ένα συμπληρωματικό μοντέλο βασισμένο στην εναπομείνασα ζωή του κάθε μηχανικού μέρους θα ήταν χρήσιμο. Οι δύο προαναφερθείσες βελτιώσεις αποτελούν βασικούς τομείς για μελλοντική εργασία που θα είναι γενικά ωφέλιμη σχετικά με την παρακολούθηση της δυναμικής πολύπλοκων θαλάσσιων σκαφών. Σχεδιάζουμε να συνεχίσουμε να τελειοποιούμε τις προσεγγίσεις μας, δίνοντας έμφαση στη χρήση πιο πολύπλοκων κανόνων ανίχνευσης ανωμαλιών.



### 5.3 Μηχανική Μάθηση & ασφάλεια των τροφίμων.

Οι προσεγγίσεις που παρουσιάζονται σε αυτήν την εργασία έχουν εφαρμοστεί ως πλαίσιο για την αντιμετώπιση μιας από τις κύριες προκλήσεις στον τομέα της ασφάλειας των τροφίμων, που είναι η συνεχής βελτιστοποίηση της παρακολούθησης, της έγκαιρης ανίχνευσης και της πρόβλεψης των τάσεων ανάκλησης τροφίμων. Η παρούσα έρευνα πραγματοποιεί μια συνολική βελτιστοποίηση τόσο ως προς την απόδοση, καθώς το μετα-μοντέλο και η χρήση υποκατάστατων δεδομένων ενισχύουν τα προβλεπόμενα αποτελέσματα του προτεινόμενου μοντέλου πρόβλεψης, όσο και όσον αφορά την προσαρμοστικότητα και τη συνεχή βελτιστοποίηση με την εφαρμογή ενός RL μοντέλου για πρόβλεψη χρονοσειρών. Ένα σημαντικό εμπόδιο για την εφαρμογή των προτεινόμενων προσεγγίσεων σε μεγάλη κλίμακα είναι η προσαρμοστικότητα σε νέους τύπους πόρων. Αυτό αντιμετωπίζεται στο πλαίσιο της έρευνάς μας με την εισαγωγή μιας προσέγγισης πρόβλεψης που βασίζεται σε RL που προσφέρει συνεχή βελτιστοποίηση μέσω συγκεκριμένων διαστημάτων εκπαίδευσης για το μοντέλο NER.

Όσον αφορά στην πρόβλεψη χρονοσειρών, παρουσιάσαμε συγκεκριμένες προσεγγίσεις στον τομέα της πρόβλεψης χρονοσειρών, ενώ αντιμετωπίσαμε βασικές προκλήσεις που περιλαμβάνουν ερμηνεία, κλίμακα, ακρίβεια και πολυπλοκότητα (που είναι εγγενείς σε πολλές περιπτώσεις χειραγώγησης χρονοσειρών). Παρά τον πειραματισμό και την αξιολόγηση, συγκρίναμε μια ποικιλία προσεγγίσεων που βασίζονται σε βαθιά νευρωνικά δίκτυα και στατιστικούς όρους. Το συμπληρωματικό μοντέλο, το οποίο αποτελείται από ένα μοντέλο RL (DQN), παρέχει πολλά υποσχόμενα αποτελέσματα όσον αφορά την πρόβλεψη ανακλήσεων τροφίμων.

Όσον αφορά τις μελλοντικές εργασίες, οι δύο προαναφερθείσες βελτιώσεις αντιπροσωπεύουν βασικούς τομείς της μελλοντικής εργασίας που θα είναι γενικά επωφελής για την παρακολούθηση δυναμικών και πολύπλοκων ανακοινώσεων ανάκλησης που δημοσιεύονται σε διαφορετικές πηγές. Σχεδιάζουμε επίσης να συνεχίσουμε να βελτιώνουμε τις προσεγγίσεις μας εφαρμόζοντας το μοντέλο RL χώρου συνεχούς δράσης, χρησιμοποιώντας ένα περιβάλλον ενεργειών πολλαπλών μεταβλητών και πολλαπλών βημάτων και να προσπαθήσουμε να αξιοποιήσουμε μια προσέγγιση κοντά στο Signal2Vec. Μια αναπαράσταση ενσωμάτωσης χρονοσειρών που χρησιμοποιείται για τη μείωση διαστάσεων για χρονοσειρές [238]. Επιπλέον, είναι μέσα στα μελλοντικά μας σχέδια να αντιμετωπίσουμε την περίπτωση μεγάλου αριθμού «ψευδών» δεδομένων με ετικέτα. Αυτό αντιμετωπίζεται με προσεγγίσεις όπως το [239], το οποίο αποτελεί μια επέκταση του OpenTag για πολλαπλές κατηγορίες προϊόντων που ονομάζονται Textract. Συγκεκριμένα, θα εκχωρηθεί συγκεκριμένη οντότητα («σχολιασμός») για κάθε κατηγορία προϊόντος αντί για καθολική οντότητα «ΠΡΟΪΟΝ». Αυτό αναμένεται να



βελτιώσει ή τουλάχιστον να βελτιώσει την εργασία μετατρέποντας τα μη δομημένα δεδομένα κειμένου σε χρονοσειρές. Σε αυτήν τη σημείωση, σχεδιάζουμε να αναπτύξουμε ένα σενάριο εποπτευόμενης εκμάθησης που θα κωδικοποιεί τη λεξιλογική σημασιολογία της ανακοίνωσης ανάκλησης με ενσωματώσεις για την πρόβλεψη της κατηγορίας προϊόντων που αναφέρεται στην ανάκληση. Επιπλέον, αυτή τη στιγμή πειραματιζόμαστε με τη λεπτή ρύθμιση ενός μοντέλου BERT με τη χρήση της διεπαφής HuggingFace για σύγκριση ή ακόμα και βελτίωση της αρχικής μας απόδοσης NER. Τέλος, η πολυγλωσσία του προτεινόμενου μοντέλου NER είναι μια άλλη απαίτηση με υψηλή βαθμολογία βάσει της παραγωγής.

#### 5.4 Μηχανική Μάθηση & ποιότητα χοιρινού κρέατος.

Σε αυτή την εργασία αντιμετωπίσαμε το πρόβλημα της κηλίδωσης κάπρου μοντελοποιώντας το ως μια εργασία δυαδικής ταξινόμησης. Εν συνεχεία δοκιμάστηκαν διάφορες προσεγγίσεις για να αντιμετωπιστούν οι προκλήσεις ανισορροπίας των δεδομένων και να προσφερθούν πολύτιμες γνώσεις για να εξασφαλιστεί υψηλότερη ποιότητα στο χοιρινό κρέας (δηλαδή, χωρίς μόλυνση κάπρου). Συγκεκριμένα, χρησιμοποιούνται στατιστικές μέθοδοι όπως το  $X^2$  και η αμοιβαία πληροφορία για την εύρεση των περισσότερων πληροφοριακών μεταβλητών σε σχέση με την εξαρτημένη μεταβλητή (boar taint) καθώς και πολλά μοντέλα μηχανικής μάθησης μαζί με μεθόδους υπερδειγματοληψίας που δίνουν μια αρχική ιδέα των καθοριστικών παραγόντων ταξινόμησης των δεδομένων, όπως ποσοστό άπαχου κρέατος, βάρος σφαγίου, ανάπτυξη, μέση τιμή αριθμού μαντρών με χοιρομητέρες δίπλα σε μάνδρες με κάπρους και βαθμολογία δερματικής βλάβης στα σφαγεία.

Στη συνέχεια, προτάθηκε μια προσαρμοσμένη μέθοδος υποδειγματοληψίας, συνοδευόμενη από επιλογή γενετικών χαρακτηριστικών, για την επίτευξη της καλύτερης βαθμολογίας F-1 κατά μέσο όρο. Τα αποτελέσματα της ανάλυσής μας, σύμφωνα με την πλειονότητα των προσεγγίσεων, υποδεικνύουν χαρακτηριστικά που φαίνονται πιο σημαντικά και μπορεί να επηρεάσουν την ποιότητα του κρέατος: τύπος τροφής, τύπος συστήματος εξαερισμού, φαρμακευτική αγωγή και διάρκεια αναμονής στο χώρο αποθήκευσης. Μια άλλη υπόθεση που μπορεί να υποτεθεί με βάση τα αποτελέσματα είναι ότι ο λόγος ανισορροπίας δεν είναι η μόνη αιτία της απώλειας απόδοσης ενός ταξινομητή. Άλλοι παράγοντες, όπως μικρές αποσυνδέσεις, θόρυβοι και επικαλύψεις, παίζουν επίσης βασικούς ρόλους παράλληλα με την αναλογία ανισορροπίας και οδηγούν σε υποαπόδοση ενός ταξινομητή.

Η μελλοντική εργασία θα επικεντρωθεί στην εφαρμογή αλγορίθμων και προσεγγίσεων προς την κατεύθυνση της συναγωγής αιτιότητας. Για το σκοπό αυτό, πλαίσια όπως το CausalML μπορούν να χρησιμοποιηθούν και να αξιολογηθούν.



Ταυτόχρονα, θα δοθεί ιδιαίτερη έμφαση στην εφαρμογή τεχνικών GraphAI για τη διευκόλυνση της εξαγωγής λανθάνουσας σχέσης μεταξύ των δεδομένων. Οι ερευνητές θα πρέπει επίσης να επικεντρωθούν σε συστήματα που, εκτός από την επίτευξη καλύτερων αποτελεσμάτων, ταξινομούν σωστά τα δείγματα. Επιπλέον, τα διαθέσιμα δεδομένα ενδέχεται να μην είναι βέλτιστα για ανίχνευση λεκέδων κάπρου, επομένως θα πρέπει να αξιοποιηθούν νέες δυνατότητες (ή συσκευές που δίνουν νέες δυνατότητες). Ταυτόχρονα, τα αποτελέσματα σχετικά με την επίδραση της θεραπείας, τον τύπο της τροφής, τον αερισμό και τον χρόνο αναμονής στο σταυλισμό θα μπορούσαν να δοκιμαστούν στην πράξη για να επικυρωθεί το μοντέλο σχετικά με την πρόληψη της κηλίδωσης κάπρου.

### **5.5 Μηχανική Μάθηση & τραπεζικός βοηθός για μικρομεσαίες επιχειρήσεις.**

Αυτό το έγγραφο θέτει τα θεμέλια για την ανάπτυξη και την παροχή νέων τραπεζικών υπηρεσιών, προσαρμοσμένων στις ΜΜΕ, πάνω από τις βασικές δραστηριότητες της τράπεζας. Αυτό επιτυγχάνεται με την εισαγωγή ενός υβριδικού μοντέλου ταξινόμησης συναλλαγών, το οποίο με τη σειρά του χρησιμοποιείται για την ανάπτυξη ενός ακριβούς μοντέλου πρόβλεψης ταμειακών ροών που βασίζεται στο πιθανοτικό μοντέλο πρόβλεψης DeepAR. Επιπλέον, η εργασία ταξινόμησης επεκτείνεται προς την εξηγήσιμη πτυχή της τεχνητής νοημοσύνης της κατηγοριοποίησης των συναλλαγών, η οποία βοηθά στην ερμηνεία των αποτελεσμάτων του προτεινόμενου μοντέλου ενώ αυξάνει τη διαφάνεια και την αξιοπιστία.

Έχοντας αναπτύξει ένα ακριβές μοντέλο υβριδικής ταξινόμησης, βασισμένο σε 20 κύριες κατηγορίες προσαρμοσμένες στις ΜΜΕ, ανοίγουμε το δρόμο για την ανάπτυξη περαιτέρω τραπεζικών μικροϋπηρεσιών με βάση τα αποτελέσματα της ταξινόμησης. Σε συνδυασμό με το αναπτυγμένο μοντέλο πρόβλεψης ταμειακών ροών, αυτές οι υπηρεσίες θα μπορούσαν να περιλαμβάνουν:

Δυναμική παρακολούθηση προϋπολογισμού, όπου το μοντέλο DeepAR μπορεί να επεκταθεί σε προκαθορισμένους προϋπολογισμούς ανά κατηγορία.

Ανίχνευση απάτης και προστασία συναλλαγών, όπου νέες συναλλαγές πέρα από ένα δεδομένο επίπεδο εμπιστοσύνης που έχει οριστεί ανά κατηγορία θα μπορούσαν να υποδηλώνουν κακόβουλες ενέργειες και δόλιες απόπειρες.

Τομεακή ανάλυση και συγκριτική αξιολόγηση, με τη χρήση δεδομένων ανοιχτού κώδικα που βασίζονται στον διαθέσιμο κώδικα SME NACE και συγκριτική ανάλυση, βάσει αμοιβαίων πελατών ΜΜΕ του επιχειρηματικού τομέα της τράπεζας.

Εκτεταμένες υπηρεσίες αναλυτικής ανάλυσης, με επίκεντρο την παροχή επιχειρηματικών πληροφοριών και προστιθέμενης αξίας στους λειτουργικούς τομείς των ΜΜΕ ή βελτιωμένη συμφωνία ΦΠΑ.



Από την πλευρά της τράπεζας, οι προτεινόμενες υπηρεσίες θα επιτρέψουν στην τράπεζα να κατανοήσει βαθύτερα τους πελάτες της στις ΜΜΕ, επιτρέποντας την παροχή εξατομικευμένων χρηματοοικονομικών προϊόντων, προσελκύοντας παράλληλα νέους επιχειρηματικούς πελάτες, επιδιώκοντας να χρησιμοποιήσουν τις εκτεταμένες υπηρεσίες οικονομικής διαχείρισης των επιχειρήσεων της τράπεζας. ως οικονομική λύση.



## 6 ΠΑΡΑΠΟΜΠΕΣ

[1] Geoffrey E Hinton, Simon Osindero, and Yee-Whye Teh. 2006. A fast learning algorithm for deep belief nets. *Neural computation*, 18(7):1527–1554

[2] Yoshua Bengio, Pascal Lamblin, Dan Popovici, and Hugo Larochelle. 2007. Greedy layer-wise training of deep networks. In *Advances in neural information processing systems*, pages 153–160.

[3] Marc Ranzato, Christopher Poultney, Sumit Chopra, Yann LeCun, et al. 2007. Efficient learning of sparse representations with an energy-based model. *Advances in neural information processing systems*, 19:1137.

[4] Honglak Lee, Chaitanya Ekanadham, and Andrew Ng. 2007. Sparse deep belief net model for visual area v2. *Advances in neural information processing systems*, 20:873–880.

[5] Yoshua Bengio. 2009. *Learning deep architectures for AI*. Now Publishers Inc.

[6] Ruslan Salakhutdinov and Geoffrey Hinton. 2009. Deep boltzmann machines. In *Artificial intelligence and statistics*, pages 448–455. PMLR.

[7] Hugo Larochelle, Yoshua Bengio, Jérôme Louradour, and Pascal Lamblin. 2009. Exploring strategies for training deep neural networks. *Journal of machine learning research*, 10(1).

[8] Dumitru Erhan, Aaron Courville, Yoshua Bengio, and Pascal Vincent. 2010. Why does unsupervised pre-training help deep learning? In *Proceedings of the thirteenth international conference on artificial intelligence and statistics*, pages 201–208. *JMLR Workshop and Conference Proceedings*.

[9] Ian Goodfellow, Honglak Lee, Quoc Le, Andrew Saxe, and Andrew Ng. 2009. Measuring invariances in deep networks. *Advances in neural information processing systems*, 22:646–654.

[10] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. 2016. *Deep learning*. MIT press.

[11] Sepp Hochreiter. 1991. *Untersuchungen zu dynamischen neuronalen netzen*. Diploma, Technische Universität München, 91(1).

[12] Kenji Doya. 1993. Bifurcations of recurrent neural networks in gradient descent learning. *IEEE Transactions on neural networks*, 1(75):218.

[13] Razvan Pascanu, Tomas Mikolov, and Yoshua Bengio. 2013. On the difficulty of training recurrent neural networks. In *International conference on machine learning*, pages 1310–1318. PMLR.





[14]Yoshua Bengio, Paolo Frasconi, and Patrice Simard. 1993. The problem of learning long-term dependencies in recurrent networks. In IEEE international conference on neural networks, pages 1183–1188. IEEE.

[15]Yoshua Bengio, Patrice Simard, and Paolo Frasconi. 1994. Learning long-term dependencies with gradient descent is difficult. IEEE transactions on neural networks, 5(2):157–166.

[16]Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. Neural computation, 9(8):1735–1780.

[17]Alex Graves, Marcus Liwicki, Santiago Fernández, Roman Bertolami, Horst Bunke, and Jürgen Schmidhuber. 2008. A novel connectionist system for unconstrained handwriting recognition. IEEE transactions on pattern analysis and machine intelligence, 31(5):855–868.

[18]Alex Graves, Abdel-rahman Mohamed, and Geoffrey Hinton. 2013. Speech recognition with deep recurrent neural networks. In 2013 IEEE international conference on acoustics, speech and signal processing, pages 6645–6649. IEEE.

[19] Alex Graves, Greg Wayne, and Ivo Danihelka. 2014. Neural Turing machines. arXiv preprint arXiv:1410.5401.

[20]Ilya Sutskever, Oriol Vinyals, and Quoc V Le. 2014. Sequence to sequence learning with neural networks. In Advances in neural information processing systems, pages 3104–3112.

[21]Ryan Kiros, Ruslan Salakhutdinov, and Richard S Zemel. 2014. Unifying visual-semantic embeddings with multimodal neural language models. arXiv preprint arXiv:1411.2539.

[22]Oriol Vinyals, Łukasz Kaiser, Terry Koo, Slav Petrov, Ilya Sutskever, and Geoffrey Hinton. 2015a. Grammar as a foreign language. Advances in neural information processing systems, 28:2773–2781.

[23]Oriol Vinyals, Alexander Toshev, Samy Bengio, and Dumitru Erhan. 2015b. Show and tell: A neural image caption generator. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 3156–3164.

[24]Kyunghyun Cho, Bart Van Merriënboer, Dzmitry Bahdanau, and Yoshua Bengio. 2014a. On the properties of neural machine translation: Encoder-decoder approaches. arXiv preprint arXiv:1409.1259.

[25]Junyoung Chung, Caglar Gulcehre, Kyunghyun Cho, and Yoshua Bengio. 2014. Empirical evaluation of gated recurrent neural networks on sequence modeling. In NIPS 2014 Workshop on Deep Learning, December 2014.



[26] Junyoung Chung, Caglar Gulcehre, Kyunghyun Cho, and Yoshua Bengio. 2015. Gated feedback recurrent neural networks. In International conference on machine learning, pages 2067–2075. PMLR.

[27] Grzegorz Chrupala, Akos Kádár, and Afra Alishahi. 2015. Learning language through pictures. arXiv preprint arXiv:1506.03694.

[28] Rafal Jozefowicz, Wojciech Zaremba, and Ilya Sutskever. 2015. An empirical exploration of recurrent network architectures. In International conference on machine learning, pages 2342–2350. PMLR.

[29] Lior Rokach, Oded Maimon. "DECISION TREES"  
<http://www.ise.bgu.ac.il/faculty/liorr/hbchap9.pdf>

[30] Ingo Ruczinski. "Classification Algorithms and Regression Trees"  
<https://rafalab.github.io/pages/649/section-11.pdf>

[31] Quinlan, J.R. "Simplifying Decision Trees"  
<https://dspace.mit.edu/handle/1721.1/6453>

[32] Quinlan, J.R., Induction of decision trees, Machine Learning 1, 81-106, 1986.

[33] Quinlan, J. R., C4.5: Programs for Machine Learning, Morgan Kaufmann, Los Altos, 1993.

[34] Breiman L., Friedman J., Olshen R., and Stone C.. Classification and Regression Trees. Wadsworth Int. Group, 1984.

[35] Breiman, L. (2001). Random forests. Machine learning, 45(1), 5-32.

[36] Adele Cutler, David Richard Cutler, John R Stevens. "Random Forests"  
[https://www.researchgate.net/publication/236952762\\_Random\\_Forests](https://www.researchgate.net/publication/236952762_Random_Forests)

[37] Gradient boosting [https://en.wikipedia.org/wiki/Gradient\\_boosting](https://en.wikipedia.org/wiki/Gradient_boosting)

[38] Alexey Natekin<sup>1</sup>, Alois Knoll. "Gradient Boosting Machines Tutorial"  
<https://pdfs.semanticscholar.org/eec4/71897375942fd690b736c2753bb19d907273.pdf>

[39] Kulkarni, V. Y., & Sinha, P. K. (2012, July). Pruning of random forest classifiers: A survey and future directions. In 2012 International Conference on Data Science & Engineering (ICDSE) (pp. 64-68). IEEE.

[40] Y. Bai, Z. Sun, J. Deng, L. Li, J. Long, and C. Li, "Manufacturing quality prediction using intelligent learning approaches: A comparative study," Sustainability, vol. 10, no. 1, p. 85, 2018.



- [41] . Dean, G. Corrado, R. Monga, K. Chen, M. Devin, M. Mao, M. Ranzato, A. Senior, P. Tucker, K. Yang et al., “Large scale distributed deep networks,” in *Advances in neural information processing systems*, 2012, pp. 1223–1231.
- [42]D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” *arXiv preprint arXiv:1412.6980*, 2014.
- [43]A. Gard, “Slow steaming on 2-stroke engines,” *Loss Prevention Circular*, no. 03-09, 2009.
- [44]K. Verbert, B. De Schutter, and R. Babuška, “Timely condition-based maintenance planning for multi-component systems,” *Reliability Engineering & System Safety*, vol. 159, pp. 310–321, 2017.
- [45]T. Dunning and E. Friedman, “Time series databases,” *New Ways to Store and Access data*, 2015.
- [46]V. Chandola, A. Banerjee, and V. Kumar, “Anomaly detection: A survey,” *ACM computing surveys (CSUR)*, vol. 41, no. 3, pp. 1–58, 2009.
- [47]Nyachuba, D.G.: Foodborne illness: is it on the rise? *Nutrition reviews* 68(5), 257–269 (2010)
- [48]Devlin, J., Chang, M.W., Lee, K., Toutanova, K.: Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805* (2018)
- [49]Luo, F., Fang, P., Qiu, Q., Xiao, H.: Features induction for product named entity recognition with crfs. In: *Proceedings of the 2012 IEEE 16th International Conference on Computer Supported Cooperative Work in Design (CSCWD)*, pp. 491–496. IEEE (2012)
- [50]Zhao, J., Liu, F.: Product named entity recognition in chinese text. *Language Resources and Evaluation* 42(2), 197–217 (2008)
- [51]Marvin, H.J., Janssen, E.M., Bouzembrak, Y., Hendriksen, P.J., Staats, M.: Big data in food safety: An overview. *Critical Reviews in Food Science and Nutrition* 57(11), 2286– 2295 (2017)
- [52]Gendel, S.M., Zhu, J.: Analysis of us food and drug administration food allergen recalls after implementation of the food allergen labeling and consumer protection act. *Journal of food protection* 76(11), 1933–1938 (2013)
- [53]Bengio, Y., Lamblin, P., Popovici, D., Larochelle, H.: Greedy layer-wise training of deep networks. In: *Advances in neural information processing systems*, pp. 153–160 (2007)
- [54]Mikolov, T., Chen, K., Corrado, G., Dean, J.: Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781* (2013)



[55]Tsuboi, Y.: Neural networks leverage corpus-wide information for part-of-speech tagging. In: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 938–950 (2014)

[56]Socher, R., Huval, B., Manning, C.D., Ng, A.Y.: Semantic compositionality through

recursive matrix-vector spaces. In: Proceedings of the 2012 joint conference on empirical methods in natural language processing and computational natural language learning, pp. 1201–1211 (2012)

[57]Socher, R., Manning, C.D., Ng, A.Y.: Learning continuous phrase representations and syntactic parsing with recursive neural networks. In: Proceedings of the NIPS-2010 deep learning and unsupervised feature learning workshop, vol. 2010, pp. 1–9 (2010)

[58]Hammerton, J.: Named entity recognition with long short-term memory. In: Proceedings of the seventh conference on Natural language learning at HLT-NAACL 2003, pp. 172–175 (2003)

[59]Li, J., Sun, A., Han, J., Li, C.: A survey on deep learning for named entity recognition. IEEE Transactions on Knowledge and Data Engineering (2020)

[60] C. Rexroad, J. Vallet, L. K. Matukumalli, J. Reecy, D. Bickhart, H. Blackburn, M. Boggess, H. Cheng, A. Clutter, N. Cockett, et al., Genome to phenome: improving animal health, production, and well-being—a new usda blueprint for animal genome research 2018–2027 *Frontiers in Genetics* (2019) 327

[61]A. Paleyes, R.-G. Urma, N. D. Lawrence, Challenges in deploying machine learning: a survey of case studies, arXiv preprint arXiv:2011.09926589 (2020).

[62]N. De Briyne, C. Berg, T. Blaha, D. Temple, Pig castration: will the eu manage to ban pig castration by 2018?, *Porcine health management* 592 2 (1) (2016) 1–11

[63]E. Heyrman, S. Millet, F. Tuytens, B. Ampe, S. Janssens, N. Buys, J. Wauters, L. Vanhaecke, M. Aluwé, Olfactory evaluation of boar taint: effect of factors measured at slaughter and link with boar taint compounds, *animal* 11 (11) (2017) 2084–2093

[64]OECD, S., Outlook, E.: Policy Highlights. OECD [Electronic source].– Mode of access: <http://www.oecd.org/industry> . . . (2019)

[65]Muller, P., Robin, N., Jessie, W., Schroder, J., Braun, H., Becker, L.S., Farrenkopf, J., Ruiz, F., Caboz, S., Ivanova, M., et al.: Annual report on european smes 2018/2019-research & development and innovation by smes. European Commission (2019)



[66]Kergroach, S.: Giving momentum to sme digitalization. Journal of the International Council for Small Business 1(1), 28–31 (2020)

[67]Lai, G., Chang, W.-C., Yang, Y., Liu, H.: Modeling long-and short-term temporal patterns with deep neural networks. In: The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval, pp. 95–104 (2018)

[68]Islam, S.R., Eberle, W., Ghafoor, S.K., Ahmed, M., 2021. Explainable artificial intelligence approaches: A survey. arXiv preprint arXiv:2101.09429

[69]Demajo, L.M., Vella, V., Dingli, A., 2020. Explainable ai for interpretable credit scoring. arXiv preprint arXiv:2012.03749 .

[70]Antoniadi, A.M., Du, Y., Guendouz, Y., Wei, L., Mazo, C., Becker, B.A., Mooney, C., 2021. Current challenges and future opportunities for xai in machine learning-based clinical decision support systems: a systematic review. Applied Sciences 11, 5088.

[71]Lai, V., Cai, J.Z., Tan, C., 2019. Many faces of feature importance: Comparing built-in and post-hoc feature importance in text classification. arXiv preprint arXiv:1910.08534.

[72]Lundberg, S., Lee, S.I., 2017. A unified approach to interpreting model predictions. arXiv preprint arXiv:1705.07874 .

[73]Nguyen, D., 2018. Comparing automatic and human evaluation of local explanations for text classification, in: Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers), pp.1069–1078.

[74] Yang, Q., Wu, X., 2006. 10 challenging problems in 745 data mining research. International Journal of Information Technology & Decision Making 5,597–604.

[75]Rao, R.B., Krishnan, S., Niculescu, R.S., 2006. Data mining for improved cardiac care. AcmSigkdd Explorations Newsletter 8, 3–10.

[76]Azaria, A., Richardson, A., Kraus, S., Subrahmanian, V., 2014. Behavioral analysis of insider threat: A survey and bootstrapped prediction in imbalanced data. IEEE Transactions on Computational Social Systems 1,135–155.

[77]Branco, P., Torgo, L., Ribeiro, R.P., 2016. A survey of predictive modeling on imbalanced domains. ACM Computing Surveys (CSUR) 49, 1–50.

[78]Camino, R., Hammerschmidt, C., et al., 2020. Oversampling tabular data with deep generative models: Is it worth the effort?, in: "I Can't Believe It's Not Better!"NeurIPS 2020 workshop.



[79]Tanha, J., Abdi, Y., Samadi, N., Razzaghi, N., Asadpour, M., 2020. Boosting methods for multi-class imbalanced data classification: an experimental review. *Journal of Big Data* 7, 1–47.

[80]Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A.V., Gulin, A., 2017. Catboost: unbiased boosting with categorical features. *arXiv preprint arXiv:1706.09516* .

[81]Pfeiffer, J., Simpson, E., Gurevych, I.: Low resource multi-task sequence tagging–revisiting dynamic conditional random fields. *arXiv preprint arXiv:2005.00250* (2020)

[82]Zhang, Z., Zohren, S., Roberts, S.: Deep reinforcement learning for trading. *The Journal of Financial Data Science* 2(2), 25–40 (2020)

[83]Hessel, M., Modayil, J., Van Hasselt, H., Schaul, T., Ostrovski, G., Dabney, W., Horgan, D., Piot, B., Azar, M., Silver, D.: Rainbow: Combining improvements in deep reinforcement learning. In: *Thirty-Second AAAI Conference on Artificial Intelligence* (2018)

[84]Sherstov, A.A., Stone, P.: Function approximation via tile coding: Automating parameter choice. In: *International Symposium on Abstraction, Reformulation, and Approximation*, pp. 194–205. Springer (2005)

[85]Lai, G., Chang, W.C., Yang, Y., Liu, H.: Modeling long-and short-term temporal patterns with deep neural networks. In: *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*, pp. 95–104 (2018)

[86]Dolan, K.T., Spano, M.L.: Surrogate for nonlinear time series analysis. *Physical Review E* 64(4), 046128 (2001)

[87]Fong, S.J., Li, G., Dey, N., Crespo, R.G., Herrera-Viedma, E.: Finding an accurate early forecasting model from small dataset: A case of 2019-ncov novel coronavirus outbreak. *arXiv preprint arXiv:2003.10776* (2020)

[88]Whitley, D., 1994. A genetic algorithm tutorial. *Statistics and computing* 4,65–85.

[89]F.E. Tay, L. Cao, Application of support vector machines in financial time series forecasting, *Omega* 29 (2001) 309–317.

[90] J. Fan, X. Wang, L. Wu, H. Zhou, F. Zhang, X. Yu, X. Lu, Y. Xiang, Comparison of support vector machine and extreme gradient boosting for predicting daily global solar radiation using temperature and precipitation in humid subtropical climates: A case study in china, *Energy Convers. Manage.* 164 (2018) 102–111.

[91]C. Lea, M.D. Flynn, R. Vidal, A. Reiter, G.D. Hager, Temporal convolutional networks for action segmentation and detection, in: *Proceedings of the IEEE*



Conference on Computer Vision and Pattern Recognition, 2017, pp. 156–165. L. Du, R. Gao, Ponnuthurai Nagaratnam Suganthan et al. Information Sciences 591 (2022) 155–175

[92] H. Zhou, S. Zhang, J. Peng, S. Zhang, J. Li, H. Xiong, W. Zhang, Informer: Beyond efficient transformer for long sequence time-series forecasting, in: Proceedings of AAAI, 2021. L. Du, R. Gao, Ponnuthurai Nagaratnam Suganthan et al. Information Sciences 591 (2022) 155–175

[93] J. Fan, X. Wang, L. Wu, H. Zhou, F. Zhang, X. Yu, X. Lu, Y. Xiang, Comparison of support vector machine and extreme gradient boosting for predicting daily global solar radiation using temperature and precipitation in humid subtropical climates: A case study in china, Energy Convers. Manage. 164(2018) 102–111.

[94] N. Tak, Forecast combination with meta possibilistic fuzzy functions, Inf. Sci. 560 (2021) 168–182

[95] J. Snoek, H. Larochelle, R.P. Adams, Practical bayesian optimization of machine learning algorithms. arXiv preprint arXiv:1206.2944, 2012.

[96] B. Alizadeh, A.G. Bafti, H. Kamangir, Y. Zhang, D.B. Wright, K.J. Franz, A novel attention-based lstm cell post-processor coupled with Bayesian optimization for streamflow prediction, J. Hydrol. 601 (2021) 126526.

[97] J. Ma, Y. Ding, J.C. Cheng, F. Jiang, V.J. Gan, Z. Xu, A lag-flstm deep learning network based on bayesian optimization for multi-sequential-variant pm<sub>2.5</sub> prediction, Sustain. Cities Soc. 60 (2020) 102237.

[97] S. Zhang, Y. Chen, W. Zhang, R. Feng, A novel ensemble deep learning model with dynamic error correction and multi-objective ensemble pruning for time series forecasting, Inf. Sci. 544 (2021) 427–445.

[98] G. Claeskens, J.R. Magnus, A.L. Vasnev, W. Wang, The forecast combination puzzle: A simple theoretical explanation, Int. J. Forecast. 32 (2016) 754–762.

[99] D. Trigg, A. Leach, Exponential smoothing with an adaptive response rate, J. Oper. Res. Soc. 18 (1967) 53–59.

[100] G. Montavon, W. Samek, and K.-R. Muller, “Methods for interpreting and understanding deep neural networks,” Digital Signal Process., vol. 73, pp. 1–15, 2018.

[101] J. Choo and S. Liu, "Visual Analytics for Explainable Deep Learning," in IEEE Computer Graphics and Applications, vol. 38, no. 4, pp. 84-92, Jul./Aug. 2018, doi: 10.1109/MCG.2018.042731661.

[102] A. Weller, “Challenges for transparency,” in Proc. Int. Conf. Mach. Learn. Workshop Human Interpretability ML, 2017.



- [103] T. Zahavy, N. Ben-Zrihem, and S. Mannor, “Graying the blackbox: Understanding DQNs,” in Proc. Int. Conf. Mach. Learn., 2016, pp. 1899–1908.
- [104] D. Gunning. Explainable Artificial Intelligence (XAI) DARPA-BAA16-53. Technical report, Defense Advanced Research Projects Agency (DARPA), 2016.
- [105] Turek, M.: DARPA - Explainable Artificial Intelligence (XAI) Program (2017). <https://www.darpa.mil/program/explainable-artificial-intelligence>
- [106] State Council Chinese Government: Development Plan for New Generation Artificial Intelligence (2017). [http://www.gov.cn/zhengce/content/2017-07/20/content\\_5211996.htm](http://www.gov.cn/zhengce/content/2017-07/20/content_5211996.htm)
- [107] Goodman, B., Flaxman, S.: European union regulations on algorithmic decision-making and a “right to explanation”. AI Mag. 38(3), 50–57 (2017)
- [108] I. Ahmed, G. Jeon and F. Piccialli, "From Artificial Intelligence to eXplainable Artificial Intelligence in Industry 4.0: A survey on What, How, and Where," in IEEE Transactions on Industrial Informatics, doi: 10.1109/TII.2022.3146552.
- [109] Bhatt, U., Xiang, A., Sharma, S., Weller, A., Taly, A., Jia, Y., ... & Eckersley, P. (2020, January). Explainable machine learning in deployment. In Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency (pp. 648-657).
- [110] Jan Zenisek, Florian Holzinger, Michael Affenzeller, Machine learning based concept drift detection for predictive maintenance, Computers & Industrial Engineering, Volume 137, 2019, 106031, ISSN 0360-8352, <https://doi.org/10.1016/j.cie.2019.106031>.
- [111] A. Adadi and M. Berrada, "Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI)," in IEEE Access, vol. 6, pp. 52138-52160, 2018, doi: 10.1109/ACCESS.2018.2870052.
- [112] Holzinger, A.; Langs, G.; Denk, H.; Zatloukal, K.; Müller, H. Causability and explainability of artificial intelligence in medicine. Wiley Interdiscip. Rev. Data Min. Knowl. Discov. 2019, 9, e1312.
- [113] B. Ustun and C. Rudin, “Supersparse linear integer models for optimized medical scoring systems,” Mach. Learn., vol. 102, no. 3, pp. 349–391, 2015
- [114] Molnar, C.; Casalicchio, G.; Bischl, B. Interpretable Machine Learning—A Brief History, State-of-the-Art and Challenges. arXiv 2020, arXiv:2010.09337
- [115] S. Sarkar, “Accuracy and interpretability trade-offs in machine learning applied to safer gambling,” in Proc. CEUR Workshop, 2016, pp. 79–87.





[116] L. Breiman, “Statistical modeling: The two cultures (with comments and a rejoinder by the author),” *Stat. Sci.*, vol. 16, no. 3, pp. 199–231, 2001.

[117] Arrieta, A.B.; Díaz-Rodríguez, N.; Del Ser, J.; Bennetot, A.; Tabik, S.; Barbado, A.; García, S.; Gil-López, S.; Molina, D.; Benjamins, R.; et al. Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Inf. Fusion* 2020, 58, 82–115

[118] Marko Robnik-Šikonja and Igor Kononenko. Explaining classifications for individual instances. *IEEE Transactions on Knowledge and Data Engineering*, 20(5):589–600, 2008

[119] Ruth C Fong and Andrea Vedaldi. Interpretable explanations of black boxes by meaningful perturbation. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 3429–3437, 2017

[120] Luisa M Zintgraf, Taco S Cohen, Tameem Adel, and Max Welling. Visualizing deep neural network decisions: Prediction difference analysis. *arXiv preprint arXiv:1702.04595*, 2017.

[121] Jindong Gu and Volker Tresp. Contextual prediction difference analysis for explaining individual image classifications. *arXiv preprint arXiv:1910.09086*, 2019.

[122] Adebayo, Julius and Gilmer, Justin and Muelly, Michael and Goodfellow, Ian and Hardt, Moritz and Kim, Been. Sanity Checks for Saliency Maps. *Advances in Neural Information Processing Systems 31 (NeurIPS 2018)*, vol. 31, 2018.

<https://proceedings.neurips.cc/paper/2018/file/294a8ed24b1ad22ec2e7efea049b8737-Paper.pdf>

[123] Xiao-Hui Li and Yuhan Shi and Haoyang Li and Wei Bai and Yuanwei Song and Caleb Chen Cao and Lei Chen (2020). Quantitative Evaluations on Saliency Methods: An Experimental Study. *arXiv:2012.15616*

[124] C. Yang, A. Rangarajan, and S. Ranka. (2018). “Global model interpretation via recursive partitioning.” [Online]. Available: <https://arxiv.org/abs/1802.04253>

[125] Avanti Shrikumar, Peyton Greenside, Anna Shcherbina, and Anshul Kundaje. Not just a black box: Learning important features through propagating activation differences. *arXiv preprint arXiv:1605.01713*, 2016.

[126] Mukund Sundararajan, Ankur Taly, and Qiqi Yan. Axiomatic attribution for deep networks. In *International Conference on Machine Learning*, pages 3319–3328. PMLR, 2017.

[127] Bolei Zhou, Aditya Khosla, Agata Lapedriza, Aude Oliva, and Antonio Torralba. Learning deep features for discriminative localization. In *Proceedings of*



the IEEE conference on computer vision and pattern recognition, pages 2921–2929, 2016.

[128] Ramprasaath R Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. Grad-cam: Visual explanations from deep networks via gradient-based localization. In Proceedings of the IEEE international conference on computer vision, pages 618–626, 2017.

[129] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. "why should i trust you?" explaining the predictions of any classifier. In Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining, pages 1135–1144, 2016.

[130] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. Anchors: High-precision model-agnostic explanations. In Proceedings of the AAAI Conference on Artificial Intelligence, volume 32, 2018.

[131] Riccardo Guidotti, Anna Monreale, Salvatore Ruggieri, Dino Pedreschi, Franco Turini, and Fosca Giannotti. Local rule-based explanations of black box decision systems. arXiv preprint arXiv:1805.10820, 2018.

[132] Scott M Lundberg, Bala Nair, Monica S Vavilala, Mayumi Horibe, Michael J Eisses, Trevor Adams, David E Liston, Daniel King-Wai Low, Shu-Fang Newman, Jerry Kim, et al. Explainable machine-learning predictions for the prevention of hypoxaemia during surgery. Nature biomedical engineering, 2018.

[133] T. Schuchert, T. Aach and H. Scharr, "Range Flow in Varying Illumination: Algorithms and Comparisons," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 9, pp. 1646-1658, Sept. 2010, doi: 10.1109/TPAMI.2009.162.

[134] Smitha Milli, Ludwig Schmidt, Anca D. Dragan, and Moritz Hardt. 2019. Model Reconstruction from Model Explanations. In Proceedings of the Conference on Fairness, Accountability, and Transparency (FAT\* '19). Association for Computing Machinery, New York, NY, USA, 1–9. DOI: <https://doi.org/10.1145/3287560.3287562>

[135] Claudia V Goldman, Michael Baltaxe, Debejyo Chakraborty, and Jorge Arinez. Explaining learning models in manufacturing processes. *Procedia Computer Science*, 180:259–268, 2021.

[136] Minyoung Lee, Joohyoung Jeon, and Hongchul Lee. Explainable ai for domain experts: a posthoc analysis of deep learning for defect classification of tft–lcd panels. *Journal of Intelligent Manufacturing*, pages 1–13, 2021.

[137] Soyoung Yoo and Namwoo Kang. Explainable artificial intelligence for manufacturing cost estimation and machining feature visualization. arXiv preprint arXiv:2010.14824, 2020.



[138] N. V. Chawla, K. W. Bowyer, L. O. Hall, W. P. Kegelmeyer, Smote: synthetic minority over-sampling technique, *Journal of artificial intelligence research* 16 (2002) 321–357.

[139] H. He, Y. Bai, E. A. Garcia, S. Li, Adasyn: Adaptive synthetic sampling approach for imbalanced learning, in: 2008 IEEE international joint conference on neural networks (IEEE world congress on computational intelligence), IEEE, 2008, pp. 1322–1328.

[140] D. Rodriguez, I. Herraiz, R. Harrison, J. Dolado, J. C. Riquelme, Preliminary comparison of techniques for dealing with imbalance in software defect prediction, in: *Proceedings of the 18th International Conference on Evaluation and Assessment in Software Engineering*, 2014, pp. 1–10.

[141] J. L. Leevy, T. M. Khoshgoftaar, R. A. Bauder, N. Seliya, A survey on addressing high-class imbalance in big data, *Journal of Big Data* 5 (1) (2018) 1–30.

[142] A. Ferná'ndez, S. del R'io, N. V. Chawla, F. Herrera, An insight into imbalanced big data classification: outcomes and challenges, *Complex & Intelligent Systems* 3 (2) (2017) 105–120.

[143] L. C. Lee, C.-Y. Liong, A. A. Jemain, Iterative random vs. kennardstone sampling for ir spectrum-based classification task using pls2-da, in: *AIP Conference Proceedings*, Vol. 1940, AIP Publishing LLC, 2018, p. 020116.

[144] V. Chandola, A. Banerjee, and V. Kumar, “Anomaly detection: A survey,” *ACM computing surveys (CSUR)*, vol. 41, no. 3, pp. 1–58, 2009.

[145] M. Goldstein and S. Uchida, “A comparative evaluation of unsupervised anomaly detection algorithms for multivariate data,” *PLoS one*, vol. 11, no. 4, p. e0152173, 2016.

[146] J. Wong, C. Colburn, E. Meeks, and S. Vedaraman, “Rad—outlier detection on big data,” Web blog post. *The Netflix Tech Blog*. Netflix, vol. 19, 2015.

[147] A. Kejariwal, “Introducing practical and robust anomaly detection in a time series,” *Twitter Engineering Blog*. Web, vol. 15, 2015.

[148] G. Chandroth, “Condition monitoring: the case for integrating data from independent sources,” *Journal of Marine Engineering & Technology*, vol. 3, no. 1, pp. 9–16, 2004.

[149] D. Watzenig, M. Sommer, and G. Steiner, “Engine state monitoring and fault diagnosis of large marine diesel engines,” *e & i Elektrotechnik und Informationstechnik*, vol. 126, no. 5, pp. 173–179, 2009.

[150] Y. Raptodimos and I. Lazakis, “Using artificial neural network-selforganising map for data clustering of marine engine condition monitoring applications,” *Ships and Offshore Structures*, vol. 13, no. 6, pp. 649–656, 2018.



[151] I. Lazakis, C. Gkerekos, and G. Theotokatos, “Investigating an svm-driven, one-class approach to estimating ship systems condition,” *Ships and Offshore Structures*, vol. 14, no. 5, pp. 432–441, 2019.

[152] J. Kowalski, B. Krawczyk, and M. Woźniak, “Fault diagnosis of marine 4-stroke diesel engines using a one-vs-one extreme learning ensemble,” *Engineering Applications of Artificial Intelligence*, vol. 57, pp. 134–141, 2017.

[153] M. Nasution, O. Sitompul, and M. Ramli, “Pca based feature reduction to improve the accuracy of decision tree c4. 5 classification,” in *Journal of Physics: Conference Series*, vol. 978, no. 1. IOP Publishing, 2018, p. 012058.

[154] T. R. Hoens, R. Polikar, and N. V. Chawla, “Learning from streaming data with concept drift and imbalance: an overview,” *Progress in Artificial Intelligence*, vol. 1, no. 1, pp. 89–101, 2012.

[155] Maharana, A., Cai, K., Hellerstein, J., Hswen, Y., Munsell, M., Staneva, V., Verma, M., Vint, C., Wijaya, D., Nsoesie, E.O.: Detecting reports of unsafe foods in consumer product reviews. *JAMIA open* 2(3), 330–338 (2019)

[156] Popovski, G., Kochev, S., Seljak, B.K., Eftimov, T.: Foodie: a rule-based named-entity recognition method for food information extraction. In: *In Proceedings of the 8th International Conference on Pattern Recognition Applications and Methods*, pp. 915–922 (2019)

[157] Popovski, G., Seljak, B.K., Eftimov, T.: A survey of named-entity recognition methods for food information extraction. *IEEE Access* 8, 31586–31594 (2020)

[158] Zheng, G., Mukherjee, S., Dong, X.L., Li, F.: Opentag: Open attribute value extraction from product profiles. In: *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 1049–1058 (2018)

[159] Yom-Tov, E.: Predicting drug recalls from internet search engine queries. *IEEE journal of translational engineering in health and medicine* 5, 1–6 (2017)

[160] Zhong, Z., Chen, D.: A frustratingly easy approach for entity and relation extraction. *arXiv preprint arXiv:2010.12812* (2020)

[161] Huber, L., Squires, E., Mandell, I., De Lange, C., 2018. Age at castration (surgical or immunological) impacts carcass characteristics and meat quality of male pigs. *animal* 12, 648–656.

[162] Backus, B.L., McGlone, J.J., 2018. Evaluating environmental enrichment as a method to alleviate pain after castration and tail docking in pigs. *Applied animal behaviour science* 204, 37–42



- [163] Gautam, V., Lakhani, G., Jain, A., Dandotiya, A., Pradhan, S., 2021. Effect of castration on carcass traits and meat production in crossbred pigs
- [164] Bonneau, M., 1982. Compounds responsible for boar taint, with special emphasis on androstenone: a review. *Livestock Production Science* 9, 687–705.
- [165] Giersing, M., Ladewig, J., Forkman, B., 2006. Animal welfare aspects of preventing boar taint. *Acta Veterinaria Scandinavica* 48, 1–3.
- [166] Dunshea, F., Colantoni, C., Howard, K., McCauley, I., Jackson, P., Long, K., Lopaticki, S., Nugent, E., Simons, J., Walker, J., et al., 2001. Vaccination of boars with a gnrh vaccine (improvac) eliminates boar taint and increases growth performance. *Journal of animal science* 79, 2524–2535
- [167] Zamaratskaia, G., Squires, E., 2009. Biochemical, nutritional and genetic effects on boar taint in entire male pigs. *Animal* 3, 1508–1521
- [168] Heyrman, E., Millet, S., Tuytens, F., Ampe, B., Janssens, S., Buys, N., Wauters, J., Vanhaecke, L., Aluwé, M., 2017. Olfactory evaluation of boar taint: effect of factors measured at slaughter and link with boar taint compounds. *animal* 11, 2084–2093
- [169] García-Méndez, S., Fernández-Gavilanes, M., Juncal-Martínez, J., González-Castaño, F.J., Seara, Ó.B.: Identifying banking transaction descriptions via support vector machine short-text classification based on a specialized labelled corpus. *IEEE Access* 8, 61642–61655 (2020)
- [170] Dayioglugil, A.B., Akgul, Y.S.: Continuous embedding spaces for bank transaction data. In: *International Symposium on Methodologies for Intelligent Systems*, pp. 129–135 (2017). Springer
- [171] Khazane, A., Rider, J., Serpe, M., Gogoglou, A., Hines, K., Bruss, C.B., Serpe, R.: Deeptrax: Embedding graphs of financial transactions. In: *2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA)*, pp. 126–133 (2019). IEEE
- [172] Sukumaran, R.: Improved customer transaction classification using semi-supervised knowledge distillation. *arXiv preprint arXiv:2102.07635* (2021)
- [173] Jørgensen, R.K., Igel, C.: Machine learning for financial transaction classification across companies using character-level word embeddings of text fields. *Intelligent Systems in Accounting, Finance and Management* 28(3), 159–172 (2021)
- [174] Haas, N., Horesh, Y., Shahar, S., Resheff, Y.S.: Identifying financial institutions by transaction signatures. *arXiv preprint arXiv:1808.00151* (2018)
- [175] Vollset, E.E.N., Folkestad, O.E.E.: Automatic classification of bank transactions. Master's thesis, NTNU (2017)



[176] Schreiber, T., Schmitz, A.: Improved surrogate data for nonlinearity tests. *Physical review letters* 77(4), 635 (1996)

[177] Mourad, S.: Prophet Vs DeepAR: Forecasting Food Demand. <https://towardsdatascience.com/prophet-vs-deepar-forecasting-food-demand-2fdebfb8d282> Accessed 2019-10-20 Springer Nature 2021 LATEX template 30 Deep learning enhancing banking services

[178] Parmezan, A.R.S., Souza, V.M., Batista, G.E.: Evaluation of statistical and machine learning models for time series prediction: Identifying the state-of-the-art and the best conditions for the use of each model. *Information Sciences* 484, 302–337 (2019)

[179] Alexandrov, A., Benidis, K., Bohlke-Schneider, M., Flunkert, V., Gasthaus, J., Januschowski, T., Maddix, D.C., Rangapuram, S., Salinas, D., Schulz, J., et al.: Gluons: Probabilistic time series models in python. arXiv preprint arXiv:1906.05264 (2019)

[180] Lai, G., Chang, W.-C., Yang, Y., Liu, H.: Modeling long-and short-term temporal patterns with deep neural networks. In: *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*, pp. 95–104 (2018)

[181] I.Lazakis, C. Gkerekos, and G. Theotokatos, “Investigating an svm-driven, one-class approach to estimating ship systems condition,” *Ships and Offshore Structures*, vol. 14, no. 5, pp. 432–441, 2019

[182] [21] R. D’iaz-Uriarte and S. A. De Andres, “Gene selection and classification of microarray data using random forest,” *BMC bioinformatics*, vol. 7, no. 1, p. 3, 2006.

[183] T. Chen and C. Guestrin, “Xgboost: A scalable tree boosting system,” in *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, 2016, pp. 785–794.

[184] J. Frery, A. Habrard, M. Sebban, O. Caelen, and L. He-Guelton, “Efficient top rank optimization with gradient boosting for supervised anomaly detection,” in *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Springer, 2017, pp. 20–35.

[185] A. Widodo and B.-S. Yang, “Support vector machine in machine condition monitoring and fault diagnosis,” *Mechanical systems and signal processing*, vol. 21, no. 6, pp. 2560–2574, 2007.

[186] C. Gkerekos, I. Lazakis, and G. Theotokatos, “Exploiting machine learning for ship systems anomaly detection and healthiness forecasting,” in *Smart Ship Technology 2018*. Royal Institution of Naval Architects, 2018.



[187] K. Hundman, V. Constantinou, C. Laporte, I. Colwell, and T. Soderstrom, “Detecting spacecraft anomalies using lstms and nonparametric dynamic thresholding,” in Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2018, pp. 387–395.

[188] D. T. Shipmon, J. M. Gurevitch, P. M. Piselli, and S. T. Edwards, “Time series anomaly detection; detection of anomalous drops with limited features and sparse examples in noisy highly periodic data,” arXiv preprint arXiv:1708.03665, 2017.

[189] S. Chauhan and L. Vig, “Anomaly detection in ecg time signals via deep long short-term memory networks,” in 2015 IEEE International Conference on Data Science and Advanced Analytics (DSAA). IEEE, 2015, pp. 1–7.

[190] P. Malhotra, A. Ramakrishnan, G. Anand, L. Vig, P. Agarwal, and G. Shroff, “Lstm-based encoder-decoder for multi-sensor anomaly detection,” arXiv preprint arXiv:1607.00148, 2016.

[191] P. Malhotra, L. Vig, G. Shroff, and P. Agarwal, “Long short term memory networks for anomaly detection in time series,” in Proceedings, vol. 89. Presses universitaires de Louvain, 2015.

[192] J. Garland, T. R. Jones, M. Neuder, V. Morris, J. W. White, and E. Bradley, “Anomaly detection in paleoclimate records using permutation entropy,” Entropy, vol. 20, no. 12, p. 931, 2018.

[193] C.-C. M. Yeh, Y. Zhu, L. Ulanova, N. Begum, Y. Ding, H. A. Dau, D. F. Silva, A. Mueen, and E. Keogh, “Matrix profile i: all pairs similarity joins for time series: a unifying view that includes motifs, discords and shapelets,” in 2016 IEEE 16th international conference on data mining (ICDM). IEEE, 2016, pp. 1317–1322.

[194] C. W. Granger, “Time series analysis, cointegration, and applications,” American Economic Review, vol. 94, no. 3, pp. 421–425, 2004.

[195] Cheung, Y.W., Lai, K.S.: Lag order and critical values of the augmented dickey–fuller test. Journal of Business & Economic Statistics 13(3), 277–280 (1995)

[196] He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770–778 (2016)

[197] Collobert, R., Weston, J., Bottou, L., Karlen, M., Kavukcuoglu, K., Kuksa, P.: Natural language processing (almost) from scratch. Journal of machine learning research 12(ARTICLE), 2493–2537 (2011)

[198] Salinas, D., Flunkert, V., Gasthaus, J., Januschowski, T.: Deepar: Probabilistic forecasting with autoregressive recurrent networks. International Journal of Forecasting (2019)



- [199] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E.: Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research* 12, 2825–2830 (2011)
- [200] Wang, Y., Smola, A., Maddix, D.C., Gasthaus, J., Foster, D., Januschowski, T.: Deep factors for forecasting. *arXiv preprint arXiv:1905.12417* (2019)
- [201] Oord, A.v.d., Dieleman, S., Zen, H., Simonyan, K., Vinyals, O., Graves, A., Kalchbrenner, N., Senior, A., Kavukcuoglu, K.: Wavenet: A generative model for raw audio. *arXiv preprint arXiv:1609.03499* (2016)
- [202] E. Heyrman, S. Millet, F. Tuytens, B. Ampe, S. Janssens, N. Buys, J. Wauters, L. Vanhaecke, M. Aluwé, Olfactory evaluation of boar taint: effect of factors measured at slaughter and link with boar taint com-596 pounds, *animal* 11 (11) (2017) 2084–2093.
- [203] E. Heyrman, S. Millet, F. Tuytens, B. Ampe, S. Janssens, N. Buys, J. Wauters, L. Vanhaecke, M. Aluwé, On-farm prevalence of and potential risk factors for boar taint, *Animal* 15 (3) (2021) 100141
- [204] P. Mathur, J. Ten Napel, S. Bloemhof, L. Heres, E. Knol, H. Mulder, A human nose scoring system for boar taint and its relationship with androstenone and skatole, *Meat Science* 91 (4) (2012) 414–422
- [205] J. Trautmann, J. Gertheiss, M. Wicke, D. Mörlein, How olfactory acuity affects the sensory assessment of boar fat: A proposal for quantification, *Meat science* 98 (2) (2014) 255–262
- [206] A. R. T. Donders, G. J. Van Der Heijden, T. Stijnen, K. G. Moons, A gentle introduction to imputation of missing values, *Journal of clinical epidemiology* 59 (10) (2006) 1087–1091
- [207] R. Díaz-Uriarte, S. A. De Andres, Gene selection and classification of microarray data using random forest, *BMC bioinformatics* 7 (1) (2006) 3
- [208] M. Ali, PyCaret: An open source, low-code machine learning library in Python, pyCaret version 2.3 (July 2020). URL <https://www.pycaret.org>
- [209] N. V. Chawla, Data mining for imbalanced datasets: An overview, *Data mining and knowledge discovery handbook* (2009) 875–886
- [210] M. Hossin, M. Sulaiman, A review on evaluation metrics for data classification evaluations, *International Journal of Data Mining & Knowledge Management Process* 5 (2) (2015) 1





- [211] S. Daskalaki, I. Kopanas, N. Avouris, Evaluation of classifiers for an uneven class distribution problem, *Applied artificial intelligence* 20 (5)717 (2006) 381–417
- [212] Q. Gu, L. Zhu, Z. Cai, Evaluation measures of the classification performance of imbalanced data sets, in: *International symposium on intelligence computation and applications*, Springer, 2009, pp. 461–471
- [213] L. Prokhorenkova, G. Gusev, A. Vorobev, A. V. Dorogush, A. Gulin, Catboost: unbiased boosting with categorical features, *arXiv preprint arXiv:1706.09516* (2017).
- [214] J. Tanha, Y. Abdi, N. Samadi, N. Razzaghi, M. Asadpour, Boosting methods for multi-class imbalanced data classification: an experimental review, *Journal of Big Data* 7 (1) (2020) 1–47.
- [215] D. E. Goldberg, J. H. Holland, *Genetic algorithms and machine learning* (1988)
- [216] . Rev, 2. statistical classification of economic activities in the european community, (2008). Tech. rep., Methodologies and Working papers. European Communities, Luxembourg
- [217] Wen, Q., Sun, L., Yang, F., Song, X., Gao, J., Wang, X., Xu, H.: Timeseries data augmentation for deep learning: A survey. *arXiv preprint arXiv:2002.12478* (2020)
- [218] Lee, T.E.K., Kuah, Y., Leo, K.-H., Sanei, S., Chew, E., Zhao, L.: Surrogate rehabilitative time series data for image-based deep learning. In: *2019 27th European Signal Processing Conference (EUSIPCO)*, pp. 1–5 (2019). IEEE
- [219] Schreiber, T., Schmitz, A.: Surrogate time series. *Physica D: Nonlinear Phenomena* 142(3-4), 346–382 (2000)
- [220] Hall, P.; Gill, N.; Kurka, M.; Phan, W. Machine Learning Interpretability with H2O Driverless AI. H2O. AI. 2017. Available online: <http://docs.h2o.ai/driverless-ai/latest-stable/docs/booklets/MLIBooklet.pdf> (accessed on 5 August 2021).
- [221] Zafar, M.R.; Khan, N.M. DLIME: A deterministic local interpretable model-agnostic explanations approach for computer-aided diagnosis systems. *arXiv 2019*, arXiv:1906.10263.
- [222] Sokol, K.; Flach, P. LIMETree: Interactively Customisable Explanations Based on Local Surrogate Multi-output Regression Trees. *arXiv 2020*, arXiv:2005.01427.



[223] Bento, J.; Saleiro, P.; Cruz, A.F.; Figueiredo, M.A.; Bizarro, P. TimeSHAP: Explaining recurrent models through sequence perturbations. arXiv 2020, arXiv:2012.00073.

[224] Shrikumar, A., Greenside, P., & Kundaje, A. (2017, July). Learning important features through propagating activation differences. In International conference on machine learning (pp. 3145-3153). PMLR.

[225] Choi, E., Bahadori, M. T., Sun, J., Kulas, J., Schuetz, A., & Stewart, W. (2016). Retain: An interpretable predictive model for healthcare using reverse time attention mechanism. Advances in neural information processing systems, 29.

[226] Oreshkin, B. N., Carpo, D., Chapados, N., & Bengio, Y. (2019). N-BEATS: Neural basis expansion analysis for interpretable time series forecasting. arXiv preprint arXiv:1905.10437.

[227] Montavon, G., Binder, A., Lapuschkin, S., Samek, W., & Müller, K. R. (2019). Layer-wise relevance propagation: an overview. Explainable AI: interpreting, explaining and visualizing deep learning, 193-209.

[228] Xiao-Hui Li and Yuhuan Shi and Haoyang Li and Wei Bai and Yuanwei Song and Caleb Chen Cao and Lei Chen (2020). Quantitative Evaluations on Saliency Methods: An Experimental Study. arXiv:2012.15616

[229] Zavrtanik, Vitjan, Matej Kristan, and Danijel Skočaj. "DRAEM-A discriminatively trained reconstruction embedding for surface anomaly detection." Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021.

[230] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," The journal of machine learning research, vol. 15, no. 1, pp. 1929–1958, 2014.

[231] P. Baldi and P. J. Sadowski, "Understanding dropout," in Advances in neural information processing systems, 2013, pp. 2814–2822.

[232] G. Montavon, G. Orr, and K.-R. Müller, Neural networks: tricks of the trade. Springer, 2012, vol. 7700.

[233] J. Brownlee, Statistical Methods for Machine Learning: Discover how to Transform Data into Knowledge with Python. Machine Learning Mastery, 2018.

[234] Batista, D.S.: Named-entity evaluation metrics based on entity-level (2018). URL [https://www.davidsbatista.net/blog/2018/05/09/Named\\_Entity\\_Evaluation/](https://www.davidsbatista.net/blog/2018/05/09/Named_Entity_Evaluation/)

[235] Sundheim, B.M., Chinchor, N.: Survey of the message understanding conferences. In: HUMAN LANGUAGE TECHNOLOGY: Proceedings of a Workshop Held at Plainsboro, New Jersey, March 21-24, 1993 (1993)



[236] Segura Bedmar, I., Martinez, P., Herrero Zazo, M.: Semeval-2013 task 9: Extraction of drug-drug interactions from biomedical texts (ddiextraction 2013). Association for Computational Linguistics (2013)

[237] Tsuboi, Y.: Neural networks leverage corpus-wide information for part-of-speech tagging. In: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 938–950 (2014)

[238] Nalmpantis, C., Vrakas, D.: Signal2vec: Time series embedding representation. In: International Conference on Engineering Applications of Neural Networks, pp. 80–90. Springer (2019)

[239] Karamanolakis, G., Ma, J., Dong, X.L.: TXtract: Taxonomy-aware knowledge extraction for thousands of product categories. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, pp. 8489–8502. Association for Computational Linguistics, Online (2020). DOI 10.18653/v1/2020.acl-main.751

[240] McAfee, A., & Brynjolfsson, E. (2012). Big data: The management revolution. Harvard Business Review, 90(10), 60-68.

[241] IoT Analytics (2021). Internet of Things (IoT) Market Size, Share & Trends Analysis Report By Component (Hardware, Software, Services), By Application (Consumer, Enterprise, Industrial), By Region, And Segment Forecasts, 2021-2028