



Πανεπιστήμιο Πειραιώς

Σχολή Τεχνολογιών Πληροφορικής και Επικοινωνιών
Τμήμα Ψηφιακών Συστημάτων

Πρόγραμμα Μεταπτυχιακών Σπουδών
Πληροφορικά Συστήματα & Υπηρεσίες
Ειδίκευση: «Μεγάλα Δεδομένα και Αναλυτική»

Διπλωματική Εργασία

**«Επεξηγήσιμη Τεχνητή Νοημοσύνη για Ανάλυση
Χρονοσειρών και Προληπτική Συντήρηση»**

Βασίλειος Κούκος

Επιβλέπων Καθηγητής

Δημοσθένης Κυριαζής, Καθηγητής

Πανεπιστήμιο Πειραιώς

Σεπτέμβριος 2023

Πειραιάς

*Στη μνήμη του αγαπημένου μου πατέρα,
Ευάγγελο, με αγάπη και ευγνωμοσύνη για
όσα μου πρόσφερε καθ' όλη τη διάρκεια της
ζωής μου.*

Ευχαριστίες

Με μεγάλη χαρά και ευγνωμοσύνη ολοκληρώνω αυτήν τη φάση της ζωής μου και θέλω να εκφράσω τις θερμότερες ευχαριστίες μου προς όλους όσους συνέβαλαν στην επιτυχή ολοκλήρωση της παρούσας διπλωματικής μου εργασίας.

Αρχικά, στον επιβλέποντα μου, Καθηγητή και Αντιπρύτανη του Πανεπιστημίου Πειραιώς κ. Δημοσθένη Κυριαζή, για την ανεκτίμητη υποστήριξη και καθοδήγηση που μου παρείχε από την πρώτη μας γνωριμία. Η καθοδήγηση και η στήριξή σας όλα αυτά τα χρόνια ήταν καθοριστικές, και είμαι ευγνώμων για την εμπιστοσύνη που μου δείξατε.

Στον συνάδελφο και φίλο, Διδάκτωρ Γεώργιο Μακρίδη, για την επίσης πολύτιμη καθοδήγηση και τη στήριξή από την αρχή μέχρι και την ολοκλήρωση αυτής της εργασίας. Η συνεργασία μας, για μένα ήταν και είναι ανεκτίμητης αξίας.

Στην οικογένειά μου, τη μητέρα μου Ασημίνα, τα αδέρφια μου Μιχάλη και Κωνσταντίνα, τον ανιψιό μου Βαγγέλη, και τον πιστό μας σύντροφο, τον σκύλο μας Κούκι, με τον οποίο «μεγαλώσαμε μαζί», για την απεριόριστη αγάπη και υποστήριξή που με συνόδευσαν σε κάθε μου βήμα, παρά τις όποιες δυσκολίες που έχουμε περάσει.

Στους αγαπημένους κολλητούς μου φίλους-αδερφούς, Μιλτιάδη και Νικόλα, σας ευχαριστώ για τη συνολική σας υποστήριξη και τη θετική ενέργειά σας που με ενέπνευσαν.

Τέλος, σε ολόκληρο τον κύκλο μου, φίλους, συγγενείς, συνάδελφους και όλους όσους με στήριξαν κατά τη διάρκεια αυτού του ταξιδιού. Η συμπαράστασή σας ήταν ανεκτίμητη και συνέβαλε στην επίτευξη αυτού του στόχου.

Πίνακας Περιεχομένων

Ευχαριστίες.....	3
Πίνακας Περιεχομένων	4
Κατάλογος Εικόνων	5
Κατάλογος Πινάκων.....	6
Κατάλογος Ακρωνυμίων	7
Περίληψη	8
Abstract	9
1. Εισαγωγή.....	10
1.1 Πρόβλημα έρευνας και στόχοι	11
1.2 Τεχνητή Νοημοσύνη και Μηχανική Μάθηση	11
1.2.1 Εποπτευόμενη μάθηση	12
1.2.2 Μη εποπτευόμενη μάθηση	13
1.2.3 Εκπαίδευση Αλγορίθμων.....	14
1.2.4 Αντιστάθμιση μεροληψίας-διακύμανσης.....	15
1.2.5 Αρχιτεκτονικές βαθιάς μάθησης.....	16
1.2.5.1. Νευρωνικά Δίκτυα Feedforward	17
1.2.5.2. Αναδρομικά νευρωνικά δίκτυα.....	18
1.2.5.3. Δέντρα Απόφασης (Decision Trees).....	19
1.2.5.4. Τυχαία Δάση (Random Forests).....	21
1.2.5.5. Αλγόριθμος Gradient Boosting	22
1.3 Επεξήγηση μοντέλων δεδομένων χρονοσειρών	23
1.4 Σημαντικότητα και συνεισφορά της μελέτης.....	24
1.5 Οργάνωση της εργασίας.....	25
2. Βιβλιογραφική Ανασκόπηση	26
2.1 Πρόβλεψη χρονοσειρών	26
2.2 Επεξηγήσιμη Τεχνητή Νοημοσύνη – eXplainable AI (XAI)	27
2.3 XAI για Χρονοσειρές.....	29
2.4 Μη ισορροπημένα δεδομένα	30
2.5 Μηχανική μάθηση & Προβλεπτική συντήρηση.	31
3. Μεθοδολογία και Πρακτικές	32
3.1 Σχεδιασμός και Προσέγγιση της Έρευνας	32
3.2 Χρησιμοποιούμενα σύνολα δεδομένων.....	35

3.3 Προετοιμασία - Προ-επεξεργασία Δεδομένων	38
3.4 Χρησιμοποιούμενα Προβλεπτικά Μοντέλα	39
3.4.1 Μέσος Χρόνος για την Αποτυχία Υλικού / Εξοπλισμού (MTTF).....	39
3.4.2 Υπολειπόμενη χρήσιμη απόδοση κατά τη διάρκεια ζωής του υλικού (RUL)41	
3.4.3 Μοντέλα Μηχανικής Μάθησης – Στατιστικά Μοντέλα.....	43
3.5 ΧΑΙ για χρονοσειρές	46
4. Αποτελέσματα	49
4.1 Μηχανική Μάθηση & Προβλεπτική Συντήρηση κύριων μηχανών εμπορικών πλοίων. 49	
4.2 Αρχιτεκτονική Εργαλείου	53
4.3 Αποτελέσματα ΧΑΙ για Χρονοσειρές.....	60
5. Συμπεράσματα και Μελλοντικές επεκτάσεις	64
Βιβλιογραφία.....	65

Κατάλογος Εικόνων

Εικόνα 1: Σχέση μεροληψίας – διακύμανσης.....	15
Εικόνα 2: Παράδειγμα δέντρου απόφασης.....	20
Εικόνα 3: ΧΑΙ προσέγγιση για ερμηνεία προβλέψεων μοντέλου χρονοσειράς.....	24
Εικόνα 4: Ταξινόμηση ΧΑΙ μεθόδων σύμφωνα με την απλότητα τους, το εύρος της ερμηνευσιμότητας και το ποσοστό των εξαρτήσεων από τα αναλυόμενα / χρησιμοποιημένα μοντέλα τεχνητής νοημοσύνης.....	28
Εικόνα 5: Αρχιτεκτονική του ΧΑΙ μοντέλου N-BEATS.....	30
Εικόνα 6: Weibull Distribution – Probability plot & Survival function.....	40
Εικόνα 7: Weibull Distribution, Survival function x Fraction surviving.....	41
Εικόνα 8: Εξέλιξη κατάστασης μηχανής πλοίου.....	42
Εικόνα 9: Προσέγγιση εκτίμησης RUL.....	43
Εικόνα 10: 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.....	44
Εικόνα 11: Σχηματική εξήγηση της πτυχής δεδομένων στην προσέγγιση από άκρο σε άκρο, καθώς και των κύριων βημάτων στη διαδικασία, από τα πρωτογενή δεδομένα στον χρήστη.....	47
Εικόνα 12: 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.....	47
Εικόνα 13: Μέσο απόλυτο σφάλμα του μοντέλου LSTM με χρήση δεδομένων δοκιμής, Μπλε γραμμή Το MAE, Κόκκινες κάθετες γραμμές - ημερομηνίες ελαττωμάτων.....	50

Εικόνα 14: Η βαθμολογία ανωμαλίας που παράγεται από το μοντέλο OCSVM χρησιμοποιώντας σύνολο δεδομένων δοκιμής, Μπλε γραμμή - η βαθμολογία ανωμαλίας, Κόκκινες κάθετες γραμμές - ημερομηνίες ελαττωμάτων.	50
Εικόνα 15: Τιμή εντροπίας κυλιόμενης σταθμισμένης μετάθεσης για (α) Πίεση εισόδου γλυκού νερού ψύξης (β) Στρόβιλό-συμπιεστής: Πίεση εισόδου λαδιού Lude (γ) Πίεση εισόδου λαδιού καυσίμου. Με ανάγνωση κάθετης γραμμής απεικονίζονται οι ημερομηνίες ελαττωμάτων	52
Εικόνα 16: Αποτέλεσμα συνδυαστικού μοντέλου (μπλε χρώμα), πραγματικές βλάβες (κόκκινο χρώμα). Πιθανότητα μη φυσιολογικής κατάστασης του κύριου κινητήρα.....	52
Εικόνα 17: Αρχιτεκτονική εργαλείου προβλεπτικής συντήρησης – PdM tool.	54
Εικόνα 18: Η CI/CD του υλοποιημένου εργαλείου.	56
Εικόνα 19: Σελίδα σύνδεσης χρηστών στη πλατφόρμα.	56
Εικόνα 20: Κύρια σελίδα πλατφόρμας για παρακολούθηση των πλοίων και της κατάστασης τους.	57
Εικόνα 21: Σελίδα Predictive maintenance - Απεικόνιση τιμών αισθητήρων σε πραγματικό χρόνο.....	58
Εικόνα 22: Σελίδα Edit info - Εισαγωγή και επεξεργασία δεδομένων συντήρησης και βλαβών.....	59
Εικόνα 23: Παράδειγμα σημαντικών superpixel LIME όπου μπορούμε να παρατηρήσουμε τις κόκκινες περιοχές που δεν φαίνονται σχετικές με την προβλεπόμενη κατηγορία.	63
Εικόνα 24: Παράδειγμα του χάρτη θερμότητας Grad-Cam και της επικάλυψης του χάρτη θερμότητας όπου μπορούμε να παρατηρήσουμε τις περιοχές που εστίασε το μοντέλο. .63	

Κατάλογος Πινάκων

Πίνακας 1: Χαρακτηριστικά συνόλου δεδομένων από αισθητήρες πλοίων κύριας μηχανής (main engine data), και γενικά δεδομένα πλοίου (vessel data) που χρησιμοποιήθηκαν στην περίπτωση της προβλεπτικής συντήρησης.	36
Πίνακας 2: Συγκεντρωτικά αποτελέσματα όλων των μοντέλων και ο συνδυασμός τους για το πρόβλημα προβλεπτικής συντήρησης κύριων μηχανών πλοίων.	53
Πίνακας 3: Αποτελέσματα συνδυαστικού μοντέλου για πρόβλημα προβλεπτικής συντήρησης κύριων μηχανών πλοίων (Confusion Matrix).	53
Πίνακας 4: Αποτελέσματα των εργασιών ταξινόμησης με διαφορετικά μοντέλα.	61
Πίνακας 5: Χρόνος εκτέλεσης κάθε μοντέλου για την εκπαιδευτική εργασία καθώς και πόσες εποχές έκανε κάθε μοντέλο να συγκλίνει. Η τελευταία στήλη «RAM» αναφέρεται στα MB που απαιτούνται για τη φόρτωση των συνόλων δεδομένων εκπαίδευσης στη μνήμη RAM.....	62

Κατάλογος Ακρωνυμίων

Ακρωνύμιο	Επεξήγηση
ADASYN	Adaptive Synthetic Sampling
AI	Artificial Intelligence - Τεχνητή Νοημοσύνη
API	Application Programming Interface
BPTT	Back-Propagation Through Time
CAM	Class Activation Map
CNN	Convolutional Neural Network
CPS	Cyber Physical Systems
CRISP-DM	CRoss Industry Standard Process for Data Mining
DeepLIFT	Deep Learning Important FeaTures
DNN	Deep Neural Network - Βαθύ Νευρωνικό Δίκτυο
GPU	Graphics Processing Unit – Μονάδα Επεξεργασίας Γραφικών
GradCAM	Gradient weighted Class Activation Map
GRU	Gated Recurrent Units
HTTP (S)	Hypertext Transfer Protocol (Secure)
IIoT	Industrial Internet of Things
KNN	K-Nearest Neighbors
KS	Kennard-Stone
LIME	Local Interpretable Model-agnostic Explanations
LSTM	Long Short-Term Memory
ML	Machine Learning – Μηχανική Μάθηση
MLP	Multi-Layer perceptron
MTTF	Mean Time To Failure
NLP	Natural Language Processing - Επεξεργασία Φυσικής Γλώσσας
OCSVM	One Class Support Vector Machines
OLS	Ordinary Least Squares
PCA	Principal Component Analysis
PdM Tool	Podium Tool – Η ονομασία που δόθηκε στο υλοποιημένο εργαλείο
ReLU	Rectified Linear Unit – Συνάρτηση
REST	Representational State Transfer
RL	Reinforcement Learning - Ενισχυτική Μάθηση
RNN	Recurrent Neural Network – Αναδρομικό Νευρωνικό Δίκτυο
RUL	Remaining Useful Life
SHAP	SHapley Additive exPlanations
SMOTE	Synthetic Minority Over-sampling TEchnique
SVM	Support Vector Machine
SVR	Support Vector Regression
Tanh	Hyperbolic Tangent - Συνάρτηση
UI	User Interface (Web)
WPE	Weighted Permutation Entropy
XAI	eXplainable Artificial Intelligence - Επεξηγήσιμη Τεχνητή Νοημοσύνη
XGB	XGBoost (eXtreme Gradient Boosting)
NN	Neural Network - Νευρωνικό Δίκτυο
TN	Τεχνητή Νοημοσύνη

Περίληψη

Αυτή η εργασία ερευνά τη χρήση μεθόδων Τεχνητής Νοημοσύνης στην ανάλυση χρονοσειρών και την προγνωστική συντήρηση, με στόχο την παροχή εις βάθος γνώσεων και ρεαλιστικών λύσεων για προκλήσεις σε αυτούς τους τομείς καθώς και την εφαρμογή μεθόδων Επεξηγήσιμης Τεχνητής Νοημοσύνης (XAI) στην ανάλυση χρονοσειρών τονίζοντας την ανάγκη για κατανοητά και ερμηνεύσιμα αποτελέσματα. Σημαντικό μέρος αυτής της εργασίας είναι η ανάπτυξη μιας πρωτότυπης εφαρμογής για την παρακολούθηση και την προγνωστική συντήρηση των κύριων κινητήρων σε εμπορικά πλοία. Η εφαρμογή έχει σχεδιαστεί για τον έγκαιρο εντοπισμό πιθανών βλαβών, αξιοποιώντας δεδομένα χρονοσειρών από διάφορους αισθητήρες κινητήρων. Αυτή η προληπτική προσέγγιση βοηθά στην αποφυγή δαπανηρών και χρονοβόρων επισκευών, διασφαλίζοντας την αδιάλειπτη και ασφαλή λειτουργία των πλοίων.

Όπως και άλλος βιομηχανικός εξοπλισμός, τα πλοία είναι εξοπλισμένα με αισθητήρες που συλλέγουν δεδομένα σχετικά με τη συνολική τους λειτουργία και την κατάσταση των εξαρτημάτων τους. Αυτά τα δεδομένα, όταν αναλύονται μέσω τεχνικών τεχνητής νοημοσύνης, δίνουν πληροφορίες για πιθανά σφάλματα του κινητήρα. Αυτές οι πληροφορίες καθοδηγούν τις διαδικασίες λήψης αποφάσεων, όπως η παραγγελία ανταλλακτικών ή η αλλαγή δρομολόγησης πλοίων για συντήρηση. Αυτή η εργασία εισάγει μια προσέγγιση δύο επιπέδων για την προγνωστική συντήρηση, χρησιμοποιώντας τεχνικές μηχανικής και βαθιάς εκμάθησης σε δεδομένα αισθητήρων για την πρόβλεψη της κατάστασης συγκεκριμένων τμημάτων της μηχανής ενός πλοίου. Η προσέγγιση περιλαμβάνει μια σειρά μοντέλων που αναλύονται και εφαρμόζονται ειδικά για τη ναυτιλιακή βιομηχανία, καθώς και ένα σύνολο από αυτά τα μοντέλα για βελτιωμένη ακρίβεια πρόβλεψης. Η αποτελεσματικότητα αυτής της προσέγγισης αποδεικνύεται χρησιμοποιώντας δεδομένα πραγματικού κόσμου από μια ναυτιλιακή εταιρεία.

Επιπλέον, η εργασία υπερβαίνει τη συμβατική εξόρυξη δεδομένων και την προεπεξεργασία στο ML, τονίζοντας τη σημασία της εμπιστοσύνης και της ερμηνείας του μοντέλου, ειδικά σε σενάρια λήψης αποφάσεων στην πραγματικότητα. Το XAI είναι κεντρικό από αυτή την άποψη, υποστηρίζοντας σαφέστερες εξηγήσεις των αποφάσεων μοντέλων ML, πέρα από τις βασικές μετρήσεις απόδοσης. Η εργασία διερευνά διάφορες τεχνικές επεξήγησης, τόσο ενσωματωμένες όσο και εφαρμοσμένες μετά την ανάπτυξη του μοντέλου, για τη βελτίωση της κατανόησης και της εφαρμογής των αποτελεσμάτων. Παρουσιάζεται ένα Βαθύ Νευρωνικό Δίκτυο (DNN) με αρχιτεκτονική δασκάλου-μαθητή (μοντέλο απόσταξης), το οποίο προσφέρει ερμηνευτικότητα σε εργασίες ταξινόμησης χρονοσειρών. Η μέθοδος περιλαμβάνει τη μετατροπή χρονοσειρών σε 2D γραφικά και τη χρήση μεθόδων επισήμανσης εικόνων όπως το LIME και το Grad-Cam για να γίνουν κατανοητές οι προβλέψεις. Αν και αυτή η προσέγγιση προσφέρει αυξημένη ακρίβεια, έρχεται με την αντιστάθμιση του εκτεταμένου χρόνου εκπαίδευσης. Τέλος, η εργασία

αντιμετωπίζει τις προκλήσεις των μη ισορροπημένων συνόλων δεδομένων, υπογραμμίζοντας την ανάγκη για στοχευμένη εκπαίδευση μοντέλων για τη βελτίωση της ακρίβειας, ιδιαίτερα για τις κατηγορίες μειοψηφίας, με μεθόδους όπως το SMOTE και το ADASYN που χρησιμοποιούνται για τον μετριασμό αυτών των προκλήσεων.

Abstract

This thesis explores the use of Explainable Artificial Intelligence (XAI) in timeseries analysis and predictive maintenance, aiming to provide in-depth insights and pragmatic solutions for challenges in these areas. It emphasizes the need for understandable and interpretable results. A significant part of this work is the development of a prototype application for monitoring and predictive maintenance of main engines in commercial ships. The application is designed for the early detection of potential faults, leveraging timeseries data from various engine sensors. This proactive approach helps prevent costly and time-consuming repairs, ensuring uninterrupted and safe operation of ships.

Similar to other industrial equipment, ships are fitted with sensors that gather data about their overall functioning and the condition of their components. This data, when analyzed through AI techniques, yields insights into potential engine faults. These insights guide decision-making processes, such as ordering spare parts or rerouting ships for maintenance. This paper introduces a two-tiered approach to predictive maintenance, utilizing machine and deep learning techniques on sensor data to anticipate the condition of specific parts of a ship's engine. The approach includes an array of models analyzed and applied specifically for the maritime industry, as well as an ensemble of these models for enhanced predictive accuracy. The effectiveness of this approach is demonstrated using real-world data from a maritime company.

Furthermore, the thesis goes beyond conventional data mining and preprocessing in ML, stressing the importance of model trust and interpretability, especially in real-life decision-making scenarios. XAI is central in this regard, advocating for clearer explanations of ML model decisions, beyond basic performance metrics. The thesis investigates various explanation techniques, both embedded in and applied after model development, to enhance result comprehension and application. A Deep Neural Network (DNN) with a teacher-student architecture (distillation model) is presented, offering interpretability in timeseries classification tasks. The method involves transforming time series into 2D plots and using image highlighting methods like LIME and Grad-Cam to make predictions understandable. While this approach offers increased accuracy, it does come with the trade-off of extended training time. Additionally, the thesis addresses challenges of imbalanced datasets, underscoring the need for targeted model training to improve accuracy, particularly for minority classes, with methods like SMOTE and ADASYN employed to mitigate these challenges.

1. Εισαγωγή

Τα τελευταία χρόνια, γινόμαστε μάρτυρες του ψηφιακού μετασχηματισμού των βιομηχανικών περιβαλλόντων ως μέρος της μετάβασης στην τέταρτη βιομηχανική επανάσταση (Industry4.0). Ο μετασχηματισμός αυτός είναι βασισμένος σε Cyber Physical Systems (CPS) και ψηφιακές τεχνολογίες όπως το Cloud Computing, το Industrial Internet of Things (IIoT) και την Τεχνητή Νοημοσύνη (AI), και υποστηρίζει καινοτόμες λειτουργίες όπως μαζική προσαρμογή, προγνωστική συντήρηση, κατασκευή μηδενικών ελαττωμάτων και ψηφιακά δίδυμα (digital twins) [1]. Η τεχνητή νοημοσύνη είναι επί του παρόντος ο κινητήριος μοχλός της Industry 4.0 και επιτρέπει νέες περιπτώσεις χρήσης, συμπεριλαμβανομένης της προγνωστικής διαχείρισης ποιότητας [2].

Σε αυτή την εποχή, την εποχή των μεγάλων δεδομένων και της τεχνητής νοημοσύνης, η ανάλυση χρονοσειρών και η προληπτική συντήρηση έχουν γίνει πεδία με ιδιαίτερο ενδιαφέρον για τη βιομηχανία και την επιστήμη. Η χρήση τεχνητής νοημοσύνης για την πρόβλεψη και ανάλυση συμβάντων με βάση τα δεδομένα χρονοσειρών έχει το δυναμικό να ανατρέψει τον τρόπο με τον οποίο προσεγγίζουμε πολλά βασικά ζητήματα στον τομέα των επιχειρήσεων και της τεχνολογίας.

Ωστόσο, ένα από τα βασικά προβλήματα που αντιμετωπίζουμε στην εφαρμογή τεχνητής νοημοσύνης στην πραγματική ζωή είναι η έλλειψη διαφάνειας και επεξηγησιμότητας των αλγορίθμων. Η Επεξηγήσιμη Τεχνητή Νοημοσύνη (XAI) αποτελεί μια προσπάθεια να δώσει "ανθρώπινο" νόημα στις αποφάσεις που λαμβάνονται από τους αλγορίθμους, διευκολύνοντας την κατανόηση και την εμπιστοσύνη των χρηστών προς αυτούς.

Στην παρούσα διπλωματική εργασία, επικεντρωνόμαστε στην εξερεύνηση της χρήσης της Επεξηγήσιμης Τεχνητής Νοημοσύνης για την Ανάλυση Χρονοσειρών και την Προληπτική Συντήρηση, προσπαθώντας να αντλήσουμε εις βάθος γνώσεις από τα δεδομένα και να παρέχουμε πρακτικές λύσεις σε προκλήσεις της βιομηχανίας, με έμφαση στην κατανόηση και την ερμηνευτικότητα των αποτελεσμάτων.

Στο πλαίσιο της παρούσας εργασίας, επιπλέον υλοποιήσαμε μία πρότυπη εφαρμογή για την παρακολούθηση και την προβλεπτική συντήρηση κύριων μηχανών εμπορικών πλοίων. Η εφαρμογή, στοχεύει στην έγκαιρη ανίχνευση πιθανών βλαβών και την πρόληψη ακριβών και χρονοβόρων επεμβάσεων. Μέσω της ανάλυσης δεδομένων χρονοσειρών που προέρχονται από διάφορους αισθητήρες της μηχανής, η εφαρμογή μας είναι σε θέση να προβλέψει επικείμενες βλάβες (ή ανάγκη για συντήρηση), διασφαλίζοντας έτσι τη συνεχή και ασφαλή λειτουργία των πλοίων.

1.1 Πρόβλημα έρευνας και στόχοι

Κατά την εφαρμογή των μεθόδων Μηχανικής Μάθησης στην παρούσα εργασία, αντιμετωπίσαμε ένα πλήθος προκλήσεων που πηγάζουν πέρα από τα τυπικά ζητήματα εξόρυξης δεδομένων και προ-επεξεργασίας. Επιπλέον πρόκληση αποτελεί η εμπιστοσύνη και η ερμηνευτικότητα των μοντέλων καθώς υπήρχε μια συνεχής ανησυχία για το πώς τα μοντέλα μηχανικής μάθησης θα μπορούσαν να επηρεάσουν τη λήψη αποφάσεων σε πραγματικές συνθήκες.

Ειδικά η Επεξηγήσιμη Τεχνητή Νοημοσύνη (XAI) [3] έπαιξε κεντρικό ρόλο στην παρούσα εργασία. Τονίζεται η ανάγκη για έναν πιο διαφανή τρόπο εξήγησης των αποφάσεων που λαμβάνονται από τα μοντέλα μηχανικής μάθησης, πέρα από τις τυπικές μετρικές απόδοσης. Εξετάζονται ποικίλες τεχνικές επεξήγησης, τόσο ενσωματωμένες στα μοντέλα που αναπτύχθηκαν όσο και post-hoc, αναζητώντας τρόπους για την καλύτερη κατανόηση και διαμόρφωση των αποτελεσμάτων.

Τα μη ισορροπημένα σύνολα δεδομένων παρουσίασαν επιπλέον δυσκολίες, υπογραμμίζοντας την ανάγκη για προσεκτική και στοχευμένη εκπαίδευση των μοντέλων. Τα μη ισορροπημένα σύνολα δεδομένων αντιμετωπίζουν προβλήματα στην ταξινόμηση και πρόβλεψη, καθώς οι περισσότεροι αλγόριθμοι μηχανικής μάθησης είναι σχεδιασμένοι για ισορροπημένα σύνολα δεδομένων. Αυτό επηρεάζει την ακρίβεια των μοντέλων, ειδικά για την μειονοτική τάξη, που συχνά είναι η πιο σημαντική (π.χ., διαγνώσεις καρκίνου ή επιθέσεις ασφάλειας) [4] [5]. Υπάρχουν μέθοδοι που αναζητούν να αντιμετωπίσουν αυτό το πρόβλημα, όπως το SMOTE και το ADASYN [6].

Ως εκ τούτου η επεξήγηση και η ερμηνεία των αποτελεσμάτων σε πραγματικές συνθήκες ήταν εξίσου σημαντική με την ακρίβεια και την απόδοση των μοντέλων.

1.2 Τεχνητή Νοημοσύνη και Μηχανική Μάθηση

Η μηχανική μάθηση που ανήκει στις τεχνολογίες της τεχνητής νοημοσύνης είναι ένας υποτομέας της τεχνολογίας πληροφορικής που ασχολείται με την αυτοματοποίηση της ανάλυσης δεδομένων, επιτρέποντας στα συστήματα να αναγνωρίζουν πρότυπα και μοτίβα και να κάνουν προβλέψεις βασισμένες σε αυτά. Οι μέθοδοι της μηχανικής μάθησης μπορεί να κατηγοριοποιηθούν σε 2 βασικούς τύπους, την εποπτευόμενη και τη μη εποπτευόμενη μάθηση [7].

Στην εποπτευόμενη μάθηση, το σύστημα προσπαθεί να αντιστοιχίσει εισόδους (X) σε συγκεκριμένους στόχους (Y), βασισμένο σε ένα σύνολο δεδομένων που περιλαμβάνει παραδείγματα εισόδου-εξόδου. Οι είσοδοι μπορεί να περιλαμβάνουν ποικίλες μορφές

δεδομένων, όπως εικόνες, κείμενο ή χρονοσειρές, ενώ οι στόχοι μπορεί να είναι κατηγορικά δεδομένα ή αριθμητικές τιμές. Αυτός ο τύπος μηχανικής μάθησης εφαρμόζεται σε προβλήματα ταξινόμησης και παλινδρόμησης.

Η μη εποπτευόμενη μάθηση δεν χρησιμοποιεί προκαθορισμένους στόχους, αλλά αναζητά αυτόματα πρότυπα ή δομές μέσα στα δεδομένα. Αυτή η προσέγγιση χρησιμοποιείται για να ανακαλύψει γνώση που βρίσκεται κρυμμένη μέσα στα δεδομένα χωρίς να υπάρχει εκ των προτέρων ορισμένος στόχος.

Εκτός από αυτούς τους δύο βασικούς τύπους, υπάρχει και η ενισχυτική μάθηση, όπου το σύστημα μαθαίνει μέσα από την αλληλεπίδραση με ένα περιβάλλον, λαμβάνοντας ανατροφοδότηση σε μορφή ανταμοιβής για τις ενέργειές του.

Σε γενικές γραμμές, η μηχανική μάθηση είναι ένα ισχυρό εργαλείο που επιτρέπει στους υπολογιστές να ανακαλύπτουν γνώση και να λαμβάνουν αποφάσεις βασισμένες σε μεγάλα σύνολα δεδομένων.

1.2.1 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. Μια εργασία εποπτευόμενης μηχανικής μάθησης, ανάλογα με τον τύπο του στόχου, μπορεί να κατηγοριοποιηθεί ως ταξινόμηση (classification) όταν τα δεδομένα είναι διακριτά και πεπερασμένα ή ως παλινδρόμηση (regression) όταν τα δεδομένα ανήκουν σε έναν συνεχή χώρο. Ξεκινώντας με την κατηγορία της ταξινόμησης, 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. Στη βιβλιογραφία, ο όρος ταξινόμηση αναφέρεται σε πολυταξική ταξινόμηση με μία μόνο έξοδο. 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, 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:

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

Όταν η εξαρτημένη μεταβλητή (έξοδος y) 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.

1.2.2 Μη εποπτευόμενη μάθηση

Η μη εποπτευόμενη μάθηση ή αλλιώς η μάθηση χωρίς επίβλεψη είναι μια κατηγορία μηχανικής μάθησης στην οποία παρέχονται στο μοντέλο μόνο καθαρές παρατηρήσεις δεδομένων, χωρίς ετικέτες. Ο στόχος του μοντέλου είναι να μάθει μια συνάρτηση που περιγράφει τις δομές στα δεδομένα, χωρίς να γνωρίζει το επιθυμητό αποτέλεσμα για κάθε είσοδο [8]. Οι δύο κύριοι τύποι μη εποπτευόμενης μάθησης είναι η ομαδοποίηση και η ανίχνευση ανεξαρτησίας.

- Ομαδοποίηση: Ο στόχος είναι να ομαδοποιηθούν τα δεδομένα σε ομάδες με παρόμοια χαρακτηριστικά. Για παράδειγμα, ένα μοντέλο ομαδοποίησης μπορεί να χρησιμοποιηθεί για να ομαδοποιήσει τους πελάτες ενός καταστήματος σε κατηγορίες με βάση τις αγορές τους [9].
- Ανίχνευση ανεξαρτησίας: Ο στόχος είναι να εντοπιστούν οι σχέσεις μεταξύ διαφορετικών μεταβλητών. Για παράδειγμα, ένα μοντέλο ανίχνευσης ανεξαρτησίας μπορεί να χρησιμοποιηθεί για να προσδιοριστεί αν η κατανάλωση καφέ σχετίζεται με την αύξηση της αρτηριακής πίεσης.

Η μη εποπτευόμενη μάθηση έχει πολλές εφαρμογές σε διάφορους τομείς, όπως:

- Πληροφορική: Ανίχνευση ψευδών ειδήσεων, εξατομίκευση περιεχομένου, ταξινόμηση εικόνων
- Επιχειρήσεις: Ανάλυση πελατών, πρόβλεψη πωλήσεων, διαχείριση κινδύνου
- Επιστήμη: Μοριακή βιολογία, φυσική, στατιστική

Η μη εποπτευόμενη μάθηση είναι μια ισχυρή τεχνική που μπορεί να χρησιμοποιηθεί για την επίλυση μιας ευρείας γκάμας προβλημάτων. Με την κατάλληλη επιλογή αλγορίθμου και παραμέτρων, τα μοντέλα μάθησης χωρίς επίβλεψη μπορούν να παράγουν ακριβείς και χρήσιμες πληροφορίες.

Οι αλγόριθμοι μάθησης μη εποπτευόμενης μάθησης μπορούν να χωριστούν σε δύο κύριες κατηγορίες: στατιστικοί και εμπειρικοί. Οι στατιστικοί αλγόριθμοι βασίζονται σε πιθανότητες και στατιστικές θεωρίες, ενώ οι εμπειρικοί αλγόριθμοι βασίζονται σε τεχνικές μηχανικής μάθησης. Η επιλογή του κατάλληλου αλγορίθμου μάθησης χωρίς επίβλεψη εξαρτάται από διάφορους παράγοντες, όπως ο τύπος του προβλήματος που πρέπει να επιλυθεί, η ποσότητα των δεδομένων εκπαίδευσης και η πολυπλοκότητα του προβλήματος.

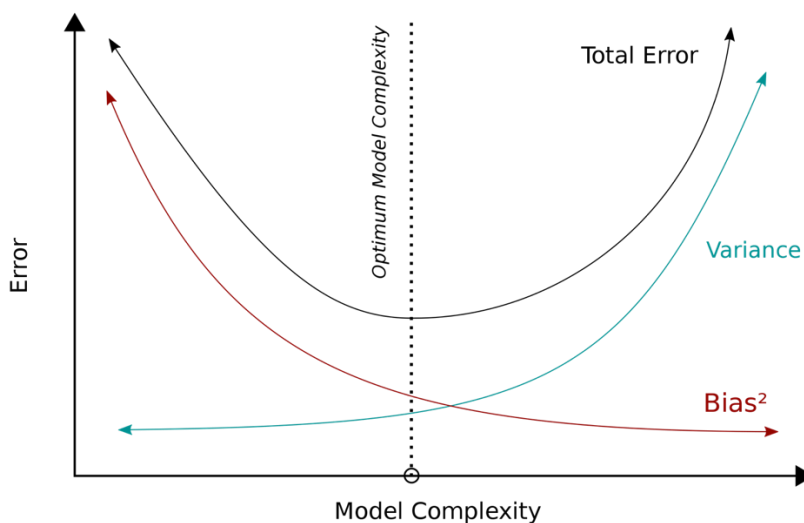
1.2.3 Εκπαίδευση Αλγορίθμων

Η ουσία της μηχανικής μάθησης είναι η εύρεση ενός μοντέλου και των συνοδευτικών παραμέτρων του, ώστε ο τελικός προγνωστικός παράγοντας να λειτουργεί αποτελεσματικά σε δεδομένα που δεν έχουν προσκομιστεί στην αρχική εκπαίδευση. Αυτό επιτυγχάνεται μέσω τριών βασικών αλγοριθμικών φάσεων: α) της εκπαίδευσης και εκτίμησης παραμέτρων, β) του συντονισμού υπερ-παραμέτρων και της επιλογής του κατάλληλου μοντέλου, και γ) της πρόβλεψης ή του συμπεράσματος. Προκειμένου να ανακαλύψουμε παραμέτρους που συμφωνούν αρμονικά με τα δεδομένα, εφαρμόζουμε σύνθετες αριθμητικές μεθόδους, οι οποίες συχνά αποτελούν στρατηγικές αναρρίχησης σε ένα τοπίο στόχων για την εύρεση του ιδανικού μέγιστου.

Παρά την προσοχή που δίνουμε στην εκπαίδευση του μοντέλου, μόνο τα αρχικά δεδομένα δεν είναι συχνά αρκετά για να εγγυηθούν την επιτυχία στην πράξη. Είναι σημαντικό να χρησιμοποιούμε μεθόδους όπως η διασταυρούμενη επικύρωση για να ελέγχουμε την ικανότητα του μοντέλου να γενικεύει σε άγνωστα δεδομένα. Πέραν αυτού, τεχνικές όπως η ρύθμιση και η προσθήκη περιορισμών στο μοντέλο μπορεί να βοηθήσουν στην αποφυγή της υπερ-προσαρμογής, διασφαλίζοντας την ικανότητά του να προβλέπει αξιόπιστα σε νέες καταστάσεις.

1.2.4 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 εικόνα απεικονίζει την σχέση μεροληψίας-διακύμανσης όπου σημειώνεται και η βέλτιστη επιλογή πολυπλοκότητας του μοντέλου μηχανικής μάθησης.



Εικόνα 1: Σχέση μεροληψίας – διακύμανσης.

Ένα πρωτογενές μοντέλο που χαρακτηρίζεται από έναν περιορισμένο χώρο παραμέτρων είναι επιρρεπές σε υψηλή μεροληψία και χαμηλή διακύμανση, αντικατοπτρίζοντας την κατάσταση της υποπροσαρμογής (underfitting). Εν αντιθέσει, ένα εξελιγμένο μοντέλο με πλουραλιστικό χώρο παραμέτρων παρουσιάζει υψηλή διακύμανση και χαμηλή μεροληψία, προσδιορίζοντας το φαινόμενο της υπερπροσαρμογής (overfitting). Το οπτικό μοντέλο προσπαθεί να εξισορροπήσει τη μεροληψία και τη διακύμανση, με στόχο την ελαχιστοποίηση του ολικού σφάλματος.

Στην πρακτική της μηχανικής μάθησης, η αναλυτική διαδικασία περιλαμβάνει συχνά στρατηγικές μοντελοποίησης, όπου επικρατούν αποφάσεις σχετικά με τις διαστάσεις των χαρακτηριστικών ή της φύσης των κατανομών πιθανοτήτων. Ο αριθμός των παραμέτρων, γνωστός ως υπερπαραμέτρος, μπορεί να προκαλέσει δραστικές διαφοροποιήσεις στην απόδοση του μοντέλου. Η οπτικοποίηση του κατάλληλου μοντέλου αποτελεί ένα πρόβλημα που αφορά την επιλογή ενός ιδανικού μοντέλου από ένα σύνολο υποψηφίων. Για μη πιθανολογικά μοντέλα, αυτή η επιλογή επιτυγχάνεται συχνά μέσω τεχνικών ενσωματωμένης διασταυρούμενης επικύρωσης.

1.2.5 Αρχιτεκτονικές βαθιάς μάθησης

Στην έρευνα του Hinton et al. (2006) [10], παρουσιάστηκε ένα επαναστατικό πλαίσιο για την εξαγωγή χαρακτηριστικών και τη Βαθιά Μάθηση (Deep Learning), που αργότερα υιοθετήθηκε από άλλες κορυφαίες ομάδες, όπως από τους Bengio et al. (2007) [11], Ranzato et al. (2007) [12], και Lee et al. (2007) [13]. Ο Bengio (2009) [14] πραγματοποίησε μια συνολική ανασκόπηση των συμβολών και των ανακαλύψεων της βαθιάς μάθησης, υπογραμμίζοντας την ιεραρχική εξαγωγή χαρακτηριστικών μέσω αφελούς μάθησης.

Στην προσέγγιση αυτή, τα χαρακτηριστικά ενός επιπέδου χρησιμεύουν ως είσοδος για την εξαγωγή πιο σύνθετων χαρακτηριστικών στο επόμενο επίπεδο. Με την προοδευτική αυξητική στοίβαξη των επιπέδων, είναι δυνατό να εκπαιδύουμε επιτυχώς επιβλεπόμενα / εποπτευόμενα μοντέλα, όπως νευρωνικά δίκτυα, ή βαθιά πιθανοτικά μοντέλο όπως οι μηχανές Deep Boltzmann, όπως παρατηρήθηκε από τους Salakhutdinov και Hinton (2009) [15]. Ευρέως αναγνώρισης εφαρμογές αυτής της προσέγγισης περιλαμβάνουν τη βελτίωση της ακρίβειας ταξινόμησης (Larochelle et al., 2009 [16], Erhan et al., 2010 [17]), την ποιότητα των παραγόμενων προσεγγίσεων σε πιθανοτικά μοντέλα (Salakhutdinov και Hinton, 2009) [15], και την αμεταβλητότητα των εκμαθημένων χαρακτηριστικών (Goodfellow et al., 2009) [18]. Τέτοια βαθιά νευρωνικά δίκτυα επιτρέπουν την αναπαράσταση συναρτήσεων αυξανόμενης πολυπλοκότητας καθώς προστίθενται περισσότερα επίπεδα και νευρώνες.

Αυτή η ενότητα παρέχει μια συνοπτική εισαγωγή στις βασικές έννοιες και τις δομικές οντότητες των σύγχρονων νευρωνικών δικτύων. Λόγω της εκτεταμένης χρήσης δικτύων βαθιάς μάθησης σε αυτήν τη εργασία, οι επόμενες υποενότητες θα περιγράψουν λεπτομερώς τις κύριες αρχιτεκτονικές αυτών των δικτύων.

1.2.5.1. Νευρωνικά Δίκτυα Feedforward

Ένα feedforward Βαθύ Νευρωνικό Δίκτυο (Deep Neural Network) είναι μια από τις θεμελιώδεις αρχιτεκτονικές στον τομέα των νευρωνικών δικτύων. Ο όρος "feedforward" προέρχεται από τη ροή πληροφοριών στο δίκτυο: οι εισροές δεδομένων κινούνται σε μια μοναδική κατεύθυνση, από το επίπεδο εισόδου, μέσω κρυφών επιπέδων, στο επίπεδο εξόδου, χωρίς επαναφορά ή παράλειψη στρωμάτων. Τα βασικά στοιχεία ενός DNN είναι τα ακόλουθα:

- Επίπεδο εισόδου: Εδώ το δίκτυο λαμβάνει είσοδο από δεδομένα. Ο αριθμός των νευρώνων (ή κόμβων) σε αυτό το επίπεδο αντιστοιχεί στον αριθμό των χαρακτηριστικών εισόδου.
- Κρυφά επίπεδα: Αυτά είναι επίπεδα μεταξύ των επιπέδων εισόδου και εξόδου. Τα DNN μπορούν να έχουν πολλαπλά κρυφά επίπεδα, κάτι που τα κάνει «βαθιά». Κάθε νευρώνας σε ένα κρυφό στρώμα μετασχηματίζει τις τιμές από το προηγούμενο επίπεδο με ένα σταθμισμένο γραμμικό άθροισμα που ακολουθείται από μια μη γραμμική ενεργοποίηση.
- Επίπεδο εξόδου: Το τελικό επίπεδο παράγει την πρόβλεψη του δικτύου. Ο αριθμός των νευρώνων εδώ εξαρτάται από την εργασία - για εργασίες παλινδρόμησης, μπορεί να είναι ένας, ενώ για εργασίες ταξινόμησης πολλών κλάσεων, ταιριάζει με τον αριθμό των κλάσεων.

Παρακάτω γίνεται μια αναπαράσταση ενός μόνο στρώματος με μαθηματικούς όρους:

Δεδομένων των εισόδων x , των βαρών W , και των προκαθορισμένων τιμών b , η έξοδος h αυτού του επιπέδου είναι: $h = \sigma(Wx + b)$. Όπου σ είναι μια συνάρτηση ενεργοποίησης όπως η σιγμοειδής, ReLU (Rectified Linear Unit) ή η tanh.

Για ένα βαθύ δίκτυο με πολλαπλά επίπεδα, αυτός ο υπολογισμός εφαρμόζεται επαναλαμβανόμενα για κάθε επίπεδο.

Η επιλογή της συνάρτησης ενεργοποίησης μπορεί να έχει σημαντική επίδραση στην απόδοση του νευρωνικού δικτύου. Οι πιο συνηθισμένες συναρτήσεις περιλαμβάνουν:

- **Σιγμοειδής (Sigmoid) συνάρτηση**: $\sigma(z) = \frac{1}{1 + e^{-z}}$
- **Συνάρτηση ReLU (Rectified Linear Unit)**: $f(z) = \max(0, z)$

- **Συνάρτηση Tanh (Hyperbolic Tangent):** $\tan h(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}}$

Η εκπαίδευση περιλαμβάνει τη ρύθμιση των βαρών και των προκαθορισμένων τιμών για την ελαχιστοποίηση της διαφοράς μεταξύ των προβλέψεων του δικτύου και των πραγματικών στόχων. Αυτό γίνεται συνήθως χρησιμοποιώντας έναν αλγόριθμο βελτιστοποίησης, πιο συχνά την καθοδική ανάβαση ή τις παραλλαγές της (π.χ. στοχαστική καθοδική ανάβαση, Adam, RMSprop).

Ο αλγόριθμος αντίστροφης διάδοσης είναι πολύ σημαντικός για την εκπαίδευση βαθιών δικτύων. Υπολογίζει την κλίση της συνάρτησης απώλειας σε σχέση με κάθε βάρος μέσω του κανόνα της αλυσίδας, κινούμενος αντίστροφα στο δίκτυο από το επίπεδο εξόδου στο επίπεδο εισόδου [19].

Οι βαθιές αρχιτεκτονικές δεν είναι χωρίς προκλήσεις:

- Εξαφανιζόμενα & Εκρηκτικά Gradients: Λόγω της εφαρμογής του κανόνα της αλυσίδας κατά τη διάρκεια της αντίστροφης διάδοσης, τα gradients μπορεί να γίνουν πολύ μικρά (εξαφανιζόμενα) ή πολύ μεγάλα (εκρηκτικά). Αυτό επηρεάζει τις ενημερώσεις βάρους και μπορεί να εμποδίσει την εκμάθηση. Τεχνικές όπως η προσεκτική αρχικοποίηση, η διαδικασία ομαδοποίησης και η χρήση συγκεκριμένων συναρτήσεων ενεργοποίησης μπορούν να μετριάσουν αυτά τα ζητήματα.
- Υπερεκπαίδευση: Όπως με κάθε μοντέλο, τα DNNs μπορούν να υπερεκπαιδευτούν στα δεδομένα εκπαίδευσης. Οι τεχνικές παλινδρόμησης όπως το dropout [20] ή η L2 παλινδρόμηση μπορεί να βοηθήσουν.
- Υπολογιστικές Απαιτήσεις: Τα βαθιά δίκτυα, ειδικά όταν είναι μεγάλα, μπορεί να είναι υπολογιστικά ακριβά για να εκπαιδευτούν. Οι υπολογιστικοί επιταχυντές (όπως τα GPUs), οι αποδοτικοί αλγόριθμοι και η περικοπή των μοντέλων είναι τρόποι για να χειριστούν αυτό.

1.2.5.2. Αναδρομικά νευρωνικά δίκτυα

Ένα αναδρομικό νευρωνικό δίκτυο (Recurrent Neural Network - RNN) είναι μια κεντρική αρχιτεκτονική για την επεξεργασία ακολουθιών δεδομένων. Σε αντίθεση με τα Feedforward DNNs, τα RNNs έχουν την ικανότητα να "θυμούνται" πληροφορίες από προηγούμενα βήματα στην ακολουθία [21], καθιστώντας τα κατάλληλα για εργασίες όπως η ανάλυση κειμένου και χρονοσειρών και η αναγνώριση φωνής.

Ο βασικός σκελετός ενός RNN περιλαμβάνει κύτταρα που δέχονται τόσο το τρέχον δείγμα της ακολουθίας όσο και μια "κρυφή" κατάσταση από το προηγούμενο βήμα.

Προκλήσεις και Βελτιώσεις

- Εξαφανιζόμενο και Εκρηκτικό Gradient: Όπως και στα DNNs, τα RNNs μπορεί να αντιμετωπίσουν προβλήματα με εξαφανιζόμενα ή εκρηκτικά gradients, καθιστώντας την εκπαίδευση δύσκολη [22].
- Long Short-Term Memory (LSTM) [23] και Gated Recurrent Units (GRU) [24]: Για να αντιμετωπίσουν τα προβλήματα των παραδοσιακών RNNs, εισήχθησαν πιο προηγμένες αρχιτεκτονικές, όπως τα LSTM και τα GRU, τα οποία εισάγουν "πύλες" για να ελέγχουν τη ροή των πληροφοριών μέσα στο δίκτυο. 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.

Η εκπαίδευση των RNNs γίνεται με τη χρήση της αντίστροφης διάδοσης στο χρόνο (Back-Propagation Through Time – BPTT) [25]. Αυτή η τεχνική απλώνει το RNN για κάθε χρονικό βήμα της ακολουθίας και εφαρμόζει την καθοδική ανάβαση.

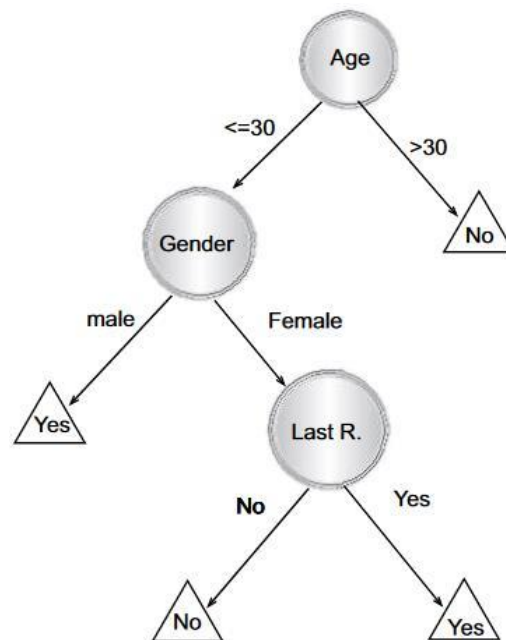
Η αρχιτεκτονική των RNNs έχει ευδοκιμήσει σε πολλές εφαρμογές που περιλαμβάνουν διαδοχικά δεδομένα όπως η αναγνώριση χειρόγραφου (Graves et al., 2008) [26], η αναγνώριση ομιλίας (Graves et al., 2013, 2014) [27][28], η δημιουργία χειρόγραφου (Graves et al., 2013) [27], η αυτόματη μετάφραση (Sutskever et al., 2014) [29], λεζάντες εικόνων (Kiros et al., 2014 [30]; Vinyals et al., 2015b [31]; Vinyals et al., 2015a [32]). Η αρχιτεκτονική GRU έχει προταθεί σε μια προσπάθεια να μειωθεί η πολυπλοκότητα της αρχιτεκτονικής LSTM (Cho et al., 2014a [33]; Chung et al., 2014, 2015 [34] [35]; Chrupala et al., 2015 [36]; Jozefowicz et al., 2015 [37]).

1.2.5.3. Δέντρα Απόφασης (Decision Trees)

Τα Δέντρα Απόφασης είναι ένα από τα πιο δημοφιλή τεχνάσματα που χρησιμοποιούνται για τους αλγόριθμους ταξινόμησης και παλινδρόμησης στη μηχανική μάθηση (Quinlan, 1986) [38]. Χρησιμοποιούν μια δομή δέντρου, όπου κάθε κόμβος αντιπροσωπεύει μια δοκιμή επάνω σε ένα χαρακτηριστικό και κάθε κλάδος είναι μια απόφαση. Είναι εύκολα κατανοητά, μπορούν να χειριστούν και αριθμητικές και κατηγορικές μεταβλητές και δεν απαιτούν προεπεξεργασία των δεδομένων (Breiman et al., 1984) [39]. Ενώ από την άλλη μεριά μπορούν να δημιουργήσουν πολύ περίπλοκα δέντρα που είναι προαιρετικά για τα δεδομένα εκπαίδευσης (overfitting) και μπορεί να μην γενικεύουν καλά σε νέα δεδομένα.

Τα Δέντρα Απόφασης εκπαιδεύονται με τη χρήση αλγορίθμων όπως το ID3, C4.5 ή CART. Η διαδικασία περιλαμβάνει την επιλογή των χαρακτηριστικών που μεγιστοποιούν

μια μετρική καθαρότητας, όπως η εντροπία ή ο δείκτης Gini (Quinlan, 1993) [40]. Τυπικά ο στόχος είναι να βρεθεί το βέλτιστο δέντρο απόφασης ελαχιστοποιώντας το σφάλμα γενίκευσης. Συνεπώς, απαιτούνται ευρετικοί μέθοδοι για την επίλυση του εκάστοτε προβλήματος. Σε γενικές γραμμές, αυτές οι μέθοδοι μπορούν να χωριστούν σε δυο 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) [38], C4.5 (Quinlan, 1993) [40], CART (Breiman et al., 1984) [39]. 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 [41].



Εικόνα 2: Παράδειγμα δέντρου απόφασης.

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.

Για να αντιμετωπιστεί το πρόβλημα του overfitting, μπορούμε να χρησιμοποιήσουμε τεχνικές όπως το pruning, ή να συνδυάσουμε πολλά δέντρα σε ensemble μεθόδους όπως το Random Forest (Breiman, 2001) [42].

1.2.5.4. Τυχαία Δάση (Random Forests)

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. Τα τυχαία δάση είναι ένας συνδυασμός δέντρων προβλέψεων έτσι ώστε κάθε δέντρο να εξαρτάται από τις τιμές ενός τυχαίου ανεξάρτητα 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) [42]. 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.

Επιπλέον, Τυχαία Δάση (Random Forests) είναι λιγότερο επιρρεπές σε overfitting σε σύγκριση με ένα μεμονωμένο δέντρο απόφασης, έχουν μεγαλύτερη ακρίβεια λόγω της συλλογικής απόφασης και της δυνατότητας χειρισμού μεγάλων συνόλων δεδομένων. Ενώ από την άλλη μεριά, μπορεί να είναι πιο χρονοβόρα στην εκπαίδευση και την πρόβλεψη τους λόγω του μεγάλου αριθμού δέντρων, και πιο δύσκολο να ερμηνευτούν σε σχέση με ένα μεμονωμένο δέντρο απόφασης.

Εκπαίδευση Random Forest:

Κατά τη διαδικασία εκπαίδευσης, το Random Forest εφαρμόζει την τεχνική του bootstrap aggregation (ή bagging). Για κάθε δέντρο, επιλέγεται ένα τυχαίο υποσύνολο των

δεδομένων εκπαίδευσης και εκπαιδεύεται το δέντρο με αυτό το υποσύνολο. Κατά τη διαδικασία αυτή, κάθε δέντρο βλέπει μόνο ένα τμήμα των δεδομένων, το οποίο οδηγεί σε διαφορετικά δέντρα με διαφορετικές προβλέψεις (Breiman, 1996) [43].

Μια σημαντική πρόκληση είναι ο υπολογιστικός χρόνος και η μνήμη που απαιτείται για την εκπαίδευση και την αποθήκευση μεγάλων δασών. Οι τεχνικές περικοπής ή η χρήση συμπαγών δέντρων μπορούν να βελτιώσουν αυτές τις προκλήσεις.

1.2.5.5. 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 [44]. Η λογική του πλησιάζει αυτή των τυχαίων δασών. Ο αλγόριθμος Gradient Boosting είναι μια τεχνική ενίσχυσης που χρησιμοποιείται για παλινδρόμηση, ταξινόμηση και άλλες εργασίες. Δημιουργεί ένα προγνωστικό μοντέλο σε μορφή συνόλου απλών μοντέλων, συνήθως δέντρων αποφάσεων (Friedman, 2001) [45].

Το Gradient Boosting λειτουργεί προσθέτοντας δέντρα σταδιακά στο μοντέλο, επιτρέποντας σε κάθε νέο δέντρο να διορθώνει και να βελτιώνει τα λάθη του προηγούμενου. Στο τέλος, το συνολικό μοντέλο είναι το αποτέλεσμα του συνδυασμού των προβλέψεων όλων των δέντρων.

Η ιδέα είναι να βελτιστοποιήσουμε μια διαφορική συνάρτηση κόστους, προσθέτοντας δέντρα που προσπαθούν να ελαχιστοποιήσουν τον κλίνοντα της συνάρτησης κόστους στο σημείο εκπαίδευσης (gradient descent στο χώρο των προβλέψεων). Στα θετικά αυτής της τεχνικής μεταξύ άλλων είναι η υψηλή ακρίβεια του, η δυνατότητα χειρισμού τόσο των κατηγορικών όσο και των πραγματικών μεταβλητών, η δυνατότητα χειρισμού των απουσιαζουσών τιμών και η ανθεκτικότητα στο overfitting. Μπορεί να είναι πιο χρονοβόρο στην εκπαίδευση λόγω της σειριακής προσθήκης των δέντρων και να έχει μεγαλύτερη πολυπλοκότητα λόγω του μεγάλου αριθμού υπερπαραμέτρων.

Η ταχύτητα και η υπολογιστική απόδοση είναι βασικά θέματα στο Gradient Boosting. Αλγόριθμοι όπως το XGBoost ή το LightGBM προσπαθούν να βελτιώσουν αυτά τα θέματα προσφέροντας πιο αποδοτικές υλοποιήσεις και πρόσθετες δυνατότητες (Chen & Guestrin, 2016 [46]; Ke et al., 2017 [47]).

1.3 Επεξήγηση μοντέλων δεδομένων χρονοσειρών

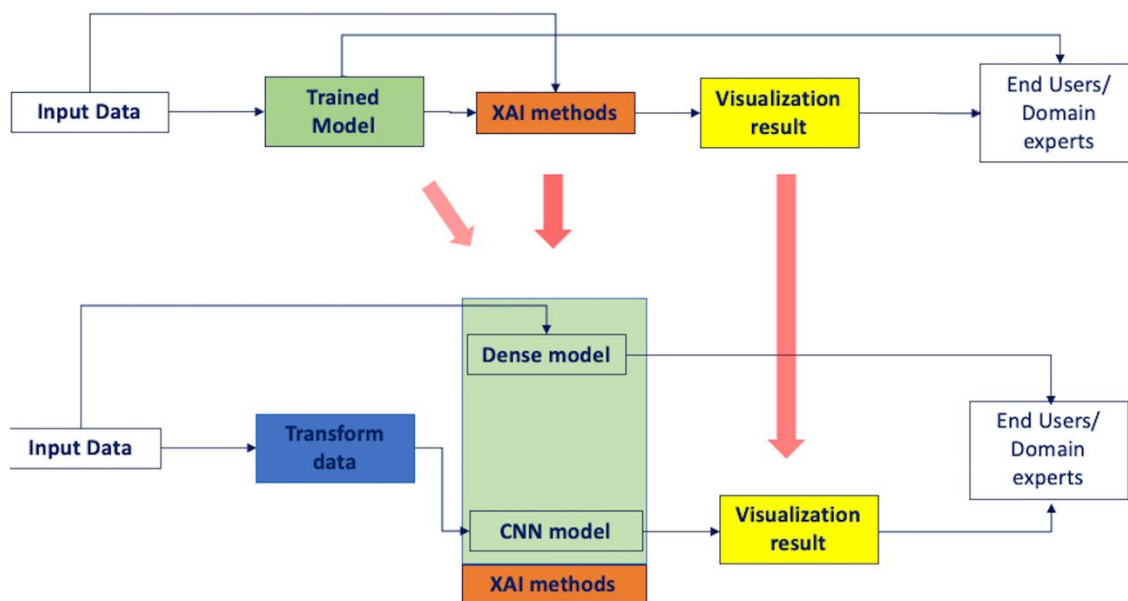
Η έννοια της επεξήγησης και της έκφρασης ενός μοντέλου Μηχανικής Μάθησης (ML) ονομάζεται ερμηνευσιμότητα ή επεξήγηση [48]. Αυτή η ανάγκη για ερμηνευτικότητα υπάρχει κυρίως στα Deep Neural Networks (DNN), τα οποία ορίζονται από μεγάλα επίπεδα πολυπλοκότητας, και έτσι εμφανίζονται ως "μαύρα κουτιά" [49]. Για την αντιμετώπιση αυτού του ζητήματος, οι ερευνητές έχουν προτείνει διάφορες μεθόδους για την εξήγηση των προβλέψεων που γίνονται από μοντέλα βαθιάς μάθησης για εργασίες ταξινόμησης εικόνων, γνωστές ως XAI (eXplainable AI) για εικόνες [50]. Αυτές οι μέθοδοι περιλαμβάνουν την οπτικοποίηση των χαρακτηριστικών που μαθαίνει το μοντέλο, τις τεχνικές απόδοσης χαρακτηριστικών και την απόσταξη μοντέλου [51].

Για παράδειγμα, η οπτικοποίηση των χαρακτηριστικών που μαθαίνει το μοντέλο μπορεί να παρέχει πληροφορίες σχετικά με τα μέρη της εικόνας εισόδου που χρησιμοποιεί το μοντέλο για να κάνει την πρόβλεψή του, ενώ οι τεχνικές απόδοσης χαρακτηριστικών μπορούν να επισημάνουν τις περιοχές της εικόνας εισόδου που είναι πιο σημαντικές για την πραγματοποίηση μιας πρόβλεψης. Η απόσταξη μοντέλων, από την άλλη πλευρά, επιχειρεί να μετατρέψει ένα σύνθετο μοντέλο σε ένα απλούστερο, πιο ερμηνεύσιμο που διατηρεί ακόμα την ακρίβεια του αρχικού μοντέλου. Ένα σύστημα απόσταξης γνώσης αποτελείται από τρία κύρια στοιχεία: τη γνώση, τον αλγόριθμο απόσταξης και την αρχιτεκτονική δασκάλου-μαθητή [52].

Η απόσταξη μοντέλου είναι μια τεχνική όπου ένα μικρότερο, απλούστερο μοντέλο εκπαιδεύεται για να μιμείται τις προβλέψεις ενός μεγαλύτερου, πιο σύνθετου μοντέλου. Αυτό επιτρέπει στο μικρότερο μοντέλο να είναι πιο εύκολα ερμηνεύσιμο και μπορεί να βοηθήσει στην παροχή εικόνας για τη λειτουργία του μεγαλύτερου μοντέλου. Για παράδειγμα, στο [53] οι συγγραφείς πρότειναν μια μέθοδο που ονομάζεται "DeepTaylor" που χρησιμοποιεί μια προσέγγιση της σειράς Taylor για να αποσπάσει ένα βαθύ νευρωνικό δίκτυο σε ένα γραμμικό μοντέλο που είναι πιο ερμηνεύσιμο.

Παρόλο που έχει γίνει πολλή δουλειά σχετικά με την επεξήγηση στα πεδία της υπολογιστικής οπτικοποίησης και της Επεξεργασίας Φυσικής Γλώσσας (NLP - Natural Language Processing), υπάρχει ακόμη πολλή δουλειά που πρέπει να γίνει για να εξηγηθούν οι μέθοδοι που εφαρμόζονται στις χρονοσειρές. Αυτό μπορεί να οφείλεται στη φύση των χρονοσειρών, την οποία δεν μπορούμε να καταλάβουμε με την πρώτη ματιά. Πράγματι, όταν ένας άνθρωπος κοιτάζει μια φωτογραφία ή διαβάζει ένα κείμενο, κατανοεί διαισθητικά και ενστικτωδώς τις υποκείμενες πληροφορίες στα δεδομένα [54]. Αν και τα χρονικά δεδομένα είναι πανταχού παρόντα στη φύση, μέσω όλων των μορφών ήχου, οι άνθρωποι δεν είναι συνηθισμένοι να αναπαριστούν αυτά τα χρονικά δεδομένα με τη μορφή ενός σήματος που ποικίλλει ανάλογα με το χρόνο. Χρειαζόμαστε ειδικές γνώσεις ή πρόσθετες μεθόδους για να αξιοποιήσουμε τις υποκείμενες πληροφορίες που υπάρχουν στα δεδομένα [55].

Σε αυτό το πλαίσιο, η προτεινόμενη προσέγγιση παρουσιάζει μια νέα προσέγγιση στο XAI για δεδομένα χρονοσειρών, στην οποία χρησιμοποιούνται τεχνικές επισήμανσης εικόνας για την απεικόνιση και την ερμηνεία των προβλέψεων ενός μοντέλου χρονοσειράς όπως απεικονίζεται στο σχήμα της Εικόνα 3.



Εικόνα 3: XAI προσέγγιση για ερμηνεία προβλέψεων μοντέλου χρονοσειράς.

1.4 Σημαντικότητα και συνεισφορά της μελέτης

Στην εργασία αυτή ερευνάται ένα πλαίσιο που συνδυάζει διάφορες μεθοδολογίες για την εξαγωγή προβλέψεων για την κατάσταση των εξαρτημάτων κύριων μηχανών πλοίων, επικεντρώνοντας την προσοχή στην αξιοποίηση διάφορων δεδομένων. Αυτό επιτυγχάνεται μέσα από μια σειρά βημάτων που περιλαμβάνουν την εξόρυξη δεδομένων, την προεπεξεργασία τους, την ανάπτυξη μοντέλων πρόβλεψης βασισμένων σε χρονοσειρές και την εκπαίδευση αλγορίθμων μηχανικής μάθησης.

Η προσφορά της εργασίας στην προβλεπτική συντήρηση είναι η πρόταση μιας νέας μεθοδολογίας για την εκτίμηση της κατάστασης των μηχανημάτων, χρησιμοποιώντας τεχνικές μηχανικής μάθησης. Αυτή η προσέγγιση έχει δοκιμαστεί σε δεδομένα πραγματικών εμπορικών πλοίων, ανταποκρινόμενη σε μια υπαρκτή ανάγκη του τομέα της ναυτιλίας. Η καινοτομία της προτεινόμενης μεθοδολογίας αναγνωρίζεται στην ικανότητα της να εντοπίζει διάφορες ανωμαλίες, συνδυάζοντας πολλαπλά μοντέλα μηχανικής και βαθιάς μάθησης. Το εύρος των εφαρμογών της, η τεχνική προεπεξεργασίας των δεδομένων και η αποδοτικότητα της στην πράξη τονίζουν τη σημασία της στον συγκεκριμένο τομέα.

Όσον αφορά την προστιθέμενη αξία της επεξήγησης για χρονοσειρές, οι επιστημονικές συνεισφορές μπορούν να συνοψιστούν στο συνδυασμό των δυνατών σημείων και των δύο δικτύων για την επίτευξη καλύτερης απόδοσης και την επεξηγησιμότητα με οπτικοποίηση των σημαντικών τμημάτων της πλοκής για χρονοσειρές, που είναι ένα ανοιχτό ερευνητικό πεδίο με περιορισμένες λύσεις τελευταίας τεχνολογίας.

1.5 Οργάνωση της εργασίας

Στην εργασία αυτή, η Ενότητα 1 παρουσιάζει συνοπτικά μια εισαγωγή στη Μηχανική Μάθηση, τις χρονοσειρές και την προβλεπτική συντήρηση, υπογραμμίζοντας προκλήσεις όπως η ανισόρροπη κατανομή των δεδομένων. Στην Ενότητα 2, αναλύεται η βιβλιογραφική έρευνα σε ερευνητικούς τομείς, ενώ η Ενότητα 3 παρουσιάζει μια προτεινόμενη προσέγγιση εφαρμογής μοντέλων Μηχανικής Μάθησης στη βιομηχανία. Η Ενότητα 4 αξιολογεί αυτή την προσέγγιση μέσω πειραμάτων, με έμφαση στη προβλεπτική συντήρηση μηχανών πλοίων. Στην Ενότητα 5, παρουσιάζονται τα συμπεράσματα, οι κύριες συνεισφορές της εργασίας και οι μελλοντικοί στόχοι.

2. Βιβλιογραφική Ανασκόπηση

Στην Ενότητα 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 στο πρόβλημα της προβλεπτικής συντήρησης κύριων μηχανών πλοίων.

2.1 Πρόβλεψη χρονοσειρών

Στη δεκαετία του 1960, η πρόβλεψη χρονοσειρών εμφανίστηκε στατιστικά. Η εκθετική εξομάλυνση και το μοντέλο Holt-winters (SARIMA) γίνονται δημοφιλή στη βιομηχανία. Με την εμφάνιση της μηχανικής μάθησης (ML), η πρόβλεψη χρονοσειρών προσεγγίζεται με μεθόδους όπως SVR και XGB. Στο [56] εξετάζεται η μεθοδολογία SVR σε συμβόλαια μελλοντικής εκπλήρωσης, ενώ το [57] συγκρίνει SVR και XGB στην πρόβλεψη ηλιακής ακτινοβολίας, υποστηρίζοντας την υπεροχή του XGB.

Τα βαθιά νευρωνικά δίκτυα (DNN) χρησιμοποιούνται επιτυχώς σε ταξινόμηση εικόνων και μοντελοποίηση ακολουθίας. Επαναλαμβανόμενα δίκτυα όπως το LSTM [58] και το DeepAR [59] είναι ιδανικά για διαδοχικά δεδομένα. Τα συνεκτικτικά δίκτυα, αρχικά σχεδιασμένα για εικόνες, έχουν προσαρμοστεί για χρονοσειρές, όπως το TCNN [60]. Στο [61] παρουσιάζεται ένα μοντέλο βασισμένο στους Transformers για προβλέψεις χρονοσειρών. Συνδυασμοί μοντέλων προτείνονται ως αποτελεσματική στρατηγική, με τα οφέλη τους επιβεβαιωμένα στο [57]. Το [62] χρησιμοποιεί μετα-πιθανολογικές μεθόδους για την βελτίωση των προβλέψεων.

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) [63] στον συντονισμό υπερπαραμέτρων βρίσκεται σε δύο πτυχές. Στο [64] εφαρμόζεται το BOA για το μοντέλο LSTM. Ενώ το [65] βελτιστοποίησε τη δομή του δικτύου και τη χρονική καθυστέρηση για την πρόβλεψη των PM2.5.

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 [66] οι συγγραφείς έχουν λάβει υπόψη μόνο τα υποψήφια βαθιά νευρωνικά δίκτυα (DNN) όπως το πολυστρωματικό perceptron (MLP - Multi-Layer perceptron), το συνελικτικό νευρωνικό δίκτυο (CNN - Convolutional Neural Network) και τη μακροπρόθεσμη βραχυπρόθεσμη μνήμη (LSTM – Long Short-Term Memory).

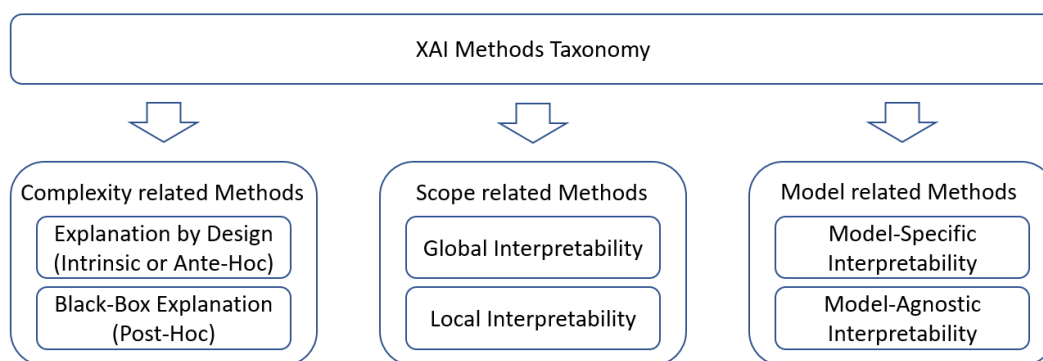
Δεύτερον, ο συνδυασμός στατικής βέλτιστης πρόβλεψης εκτιμά το βάρος κάθε μοντέλου μέσω της αντιστοιχίας απόδοσης κατά ζεύγη εκπαίδευσης-επικύρωσης, η οποία έχει εξεταστεί θεωρητικά και εμπειρικά αναξιόπιστη. Απλές μέθοδοι στατικής στάθμισης χρησιμοποιούνται συχνά για την πρόβλεψη προβλημάτων όπως οι συνήθεις μέθοδοι ελαχίστου τετραγώνου (OLS - Ordinary Least Squares), ο αλγόριθμος βασισμένος σε σφάλματα και ο εξίσου σταθμισμένος συνδυασμός μοντέλων (EVMC) [67]. Επιπλέον, πρόσφατα οι αλγόριθμοι Ενισχυτικής Μάθησης (RL - Reinforcement Learning) [68] που αναφέρονται σε αλγόριθμους που είναι «προσανατολισμένοι στον στόχο», άρχισαν να ευδοκιμούν σε ορισμένες εργασίες χρονοσειρών, ειδικά στον χρηματοοικονομικό τομέα.

2.2 Επεξηγήσιμη Τεχνητή Νοημοσύνη – eXplainable AI (XAI)

Η σημερινή εποχή χαρακτηρίζεται από υπολογιστικά συστήματα που παράγουν τεράστιες ποσότητες δεδομένων. Η αύξηση των υπολογιστικών δυνατοτήτων με μειωμένο κόστος έχει ενισχύσει τη Μηχανική Μάθηση και την Τεχνητή Νοημοσύνη (AI). Πολλά μοντέλα βαθιάς μάθησης έχουν αυξημένη πολυπλοκότητα και θεωρούνται «μαύρα κουτιά» [69] [70] [49]. Η ανάγκη για επεξηγήσιμη τεχνητή νοημοσύνη (XAI) έχει γίνει έντονη [71].

Στο [72], παρουσιάζεται η έρευνα για το πώς οι επαγγελματίες χρησιμοποιούν μοντέλα XAI. Κύριες ανάγκες είναι η ανίχνευση σφαλμάτων, η παρακολούθηση, η διαφάνεια και ο έλεγχος των μοντέλων. Για τη βελτίωση της απόδοσης των μοντέλων, χρειάζονται σαφείς εξηγήσεις, ειδικά σε “στατικούς” χώρους χαρακτηριστικών. Επίσης, υπάρχουν προκλήσεις στην παρακολούθηση μοντέλων λόγω μετατοπίσεων εννοιών [73]. Τέλος, οι ειδικοί του βιομηχανικού τομέα διεξάγουν δοκιμές και αντιμετωπίζουν διαφορές μεταξύ πρακτικής επεξήγησης και διαφάνειας.

Αν και τα τελευταία 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).



Εικόνα 4: Ταξινόμηση XAI μεθόδων σύμφωνα με την απλότητα τους, το εύρος της ερμηνευσιμότητας και το ποσοστό των εξαρτήσεων από τα αναλυόμενα / χρησιμοποιημένα μοντέλα τεχνητής νοημοσύνης.

Η πολυπλοκότητα συνδέεται με την ερμηνευσιμότητα των μοντέλων μηχανικής μάθησης. Στο [74] τονίζεται η ανάγκη για μοντέλα που είναι εγγενώς ερμηνεύσιμα. Υπάρχουν εγγενώς εξηγήσιμα μοντέλα, όπως τα δέντρα αποφάσεων και τα KNN [75], και τα μοντέλα πρόβλεψης μαύρου κουτιού (Post-hoc). Η Ustun et al. στο [76] πρότειναν μοντέλα βασισμένα στους ειδικούς τομείς ποιοτικής κατανόησης. Σε χώρους χαρακτηριστικών υψηλών διαστάσεων [77], το επίπεδο διαφάνειας μειώνεται. Υπάρχει μια ισορροπία μεταξύ ακρίβειας και επεξήγησης, όπως τονίζει ο Breiman στο [78]. Πρόσφατες εργασίες στην πολυπλοκότητα επικεντρώνονται στις Post-Hoc εξηγήσεις, 'μαύρα κουτιά' που αναφέρονται σε πολύπλοκα μοντέλα όπως SVMs και νευρωνικά δίκτυα. Παρόλο που δεν παρέχουν ολιστικές εξηγήσεις, προσφέρουν τοπικές πληροφορίες για συγκεκριμένες αποφάσεις [79]. Διαφορετικές τεχνικές όπως η PDA [80] έχουν αναπτυχθεί, με κάποιες να αντιμετωπίζουν προβλήματα στην επεξεργασία εικόνας. Επεκτάσεις και βελτιώσεις, όπως αυτής που πρότειναν οι Ruth et al. [81] και Zintgraf et al. [82], εξετάζουν τις επιπτώσεις στα χαρακτηριστικά των εικόνων και τη βαθμολόγηση ταξινόμησης.

Ένας χάρτης προεξοχής (Saliency Map) [83] 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.

Στο [84] 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. Η χαρτογράφηση ενεργοποίησης κλάσης (CAM), η οποία παρουσιάστηκε από τους Zhou et al. [85], 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. Η χαρτογράφηση ενεργοποίησης κλάσης με στάθμιση κλίσης (GradCAM) [86], η οποία είναι μια επέκταση του CAM, χρησιμοποιεί τις πληροφορίες κλίσης για να ταξινομήσει της ενεργοποίηση των νευρώνων στο τελευταίο στρώμα ενός Συνελκτικού Νευρωνικού Δικτύου (CNN).

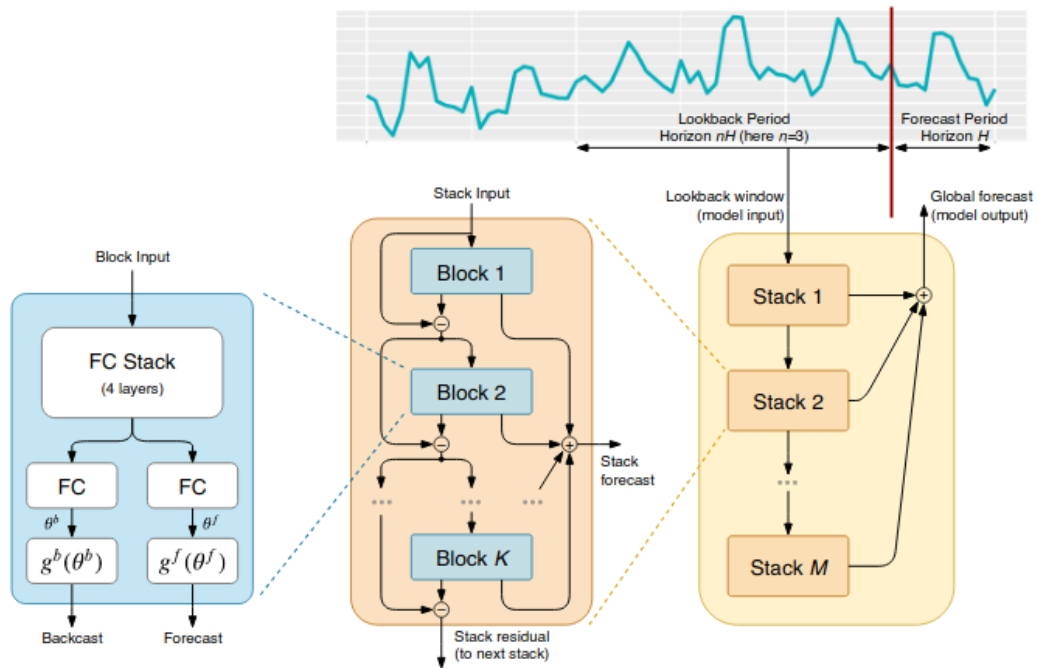
Η χρήση μεθόδων επεξήγησης επιτρέπει την τοπική ερμηνεία αποφάσεων των μοντέλων μηχανικής μάθησης. Οι αγνωστικές λύσεις μοντέλων διαχωρίζουν τα αποτελέσματα της ταξινόμησης από τις επεξηγήσεις. Προτείνονται δύο πολύ διαδεδομένα μοντέλα: το LIME [87], το οποίο προσεγγίζει μοντέλα "μαύρου κουτιού" τοπικά, και το SHAP [88], που χρησιμοποιεί τιμές Shapley για την αξιολόγηση σημασίας των χαρακτηριστικών. Στην ταξινόμηση εικόνων, δύο κύριες εξηγήσεις για δίκτυα βαθιάς μάθησης είναι το DEEPSHAP [88], που σχετίζεται με το DeepLIFT [84], και το Gradient-SHAP [89], που προέρχεται από τους αλγορίθμους IntGrad [90] και Smooth-Grad [91].

2.3 ΧΑΙ για Χρονοσειρές

Οι μέθοδοι που περιγράφονται παρακάτω αφορούν ΧΑΙ τεχνικές για χρονοσειρές. Υπάρχουν post-hoc μέθοδοι που εξετάζουν σχέσεις μεταξύ τιμών χαρακτηριστικών και προβλέψεων, όπου οι καθυστερήσεις των χρονοσειρών είναι σημαντικές, ενώ οι ante-hoc μέθοδοι έχουν ήδη ενσωματωμένη την εξήγηση στο μοντέλο κατά την φάση της εκπαίδευσης. Η πρώτη κατηγορία περιλαμβάνει παραλλαγές όπως LIME [87], k-LIME [92], DLIME [93], LIMETree [94], SHAP [88] (όπως για παράδειγμα το TimeSHAP [95]), Anchors [96], Local Foil Trees, και LoRE [97]. Σε πολλές από αυτές τις τεχνικές, δημιουργούνται υποκατάστατα μοντέλα για την αξιολόγηση προβλέψεων, με βάση τη συμπεριφορά του αρχικού μοντέλου, προσθέτοντας διαταραχές στις μεταβλητές. Τεχνικές παραλλαγής χαρακτηριστικών για την αξιολόγηση της συνεισφοράς τους στην προβλεπόμενη τιμή αποτελούν ξεχωριστό θέμα έρευνας. Τα ΧΑΙ μοντέλα μπορούν να εφαρμοστούν και σε DNN, αφού δεν εξαρτώνται από το βασικό μοντέλο μηχανικής εκμάθησης.

Στα μοντέλα βαθιάς μάθησης, οι προσεγγίσεις ante-hoc γίνονται πιο σύνθετες. Η μέθοδος Gradient-Input υπολογίζει ενεργοποιήσεις νευρώνων, ενώ υπάρχουν και

άλλες προσεγγίσεις όπως το DeepLIFT [84] και το Smooth-Grad [91]. Το επίπεδο προσοχής παρέχει πληροφορίες για χρονικά σημεία που είναι σχετικά με την πρόβλεψη. Το RETAIN [98] είναι μια προσέγγιση για δεδομένα Ηλεκτρονικών Μητρώων Υγείας που είναι ερμηνεύσιμη και κλινικά ακριβής. Το N-BEATS [99] είναι μια αρχιτεκτονική που είναι ειδικά σχεδιασμένη για χρονοσειρές και είναι εφαρμόσιμη σε διάφορες περιπτώσεις (Εικόνα 5). Η προσέγγιση αυτή διαφέρει από την παραδοσιακή μετάφραση ακολουθιών στη βαθιά μάθηση.



Εικόνα 5: Αρχιτεκτονική του ΧΑΙ μοντέλου N-BEATS.

2.4 Μη ισορροπημένα δεδομένα

Οι πρώτες εργασίες για μη ισορροπημένα δεδομένα προέρχονται από προβλήματα δυαδικής ταξινόμησης με τάξεις πλειοψηφίας και μειοψηφίας. Αυτές οι κατανομές αποτελούν πρόκληση για τα μοντέλα μηχανικής μάθησης, καθώς προκαταλαμβάνονται προς την πλειοψηφική τάξη. Οι μέθοδοι αντιμετώπισης αυτής της ανισορροπίας στοχεύουν είτε στην άμβλυνση της ανισορροπίας είτε στην εναλλαγή της διαδικασίας μάθησης. Οι βασικές προσεγγίσεις προ-επεξεργασίας των δεδομένων είναι: 1) υποδειγματοληψία της πλειοψηφικής τάξης, 2) υπερδειγματοληψία της μειονοτικής τάξης ή 3) συνδυασμός των δύο. Καθοδηγούμενες λύσεις προσπαθούν να επιλέξουν έξυπνα στιγμιότυπα για προ-επεξεργασία.

Υπάρχουν πολλές τεχνικές για την αντιμετώπιση της ανισορροπίας τάξης, μεταξύ των οποίων η τυχαία υπερδειγματοληψία (ROS), η SMOTE [100] και η ADASYN [101]. Η ROS [102], παρόλη την απλότητά της, παρέχει καλά

αποτελέσματα. Διαφορετικές μέθοδοι θα πρέπει να διερευνηθούν για κάθε πρόβλημα [103]. Η SMOTE και η ADASYN χρησιμοποιήθηκαν στην έρευνά μας χωρίς σημαντικές διαφορές στην απόδοση. Στις τεχνικές υπερ-δειγματοληψίας, υπάρχει ανησυχία για την υπερβολική προσαρμογή και την αύξηση του μεγέθους των δεδομένων. Το SMOTE και το ADASYN δημιουργούν συνθετικά νέες περιπτώσεις, αποφεύγοντας παγίδες [104].

Οι τεχνικές υποδειγματοληψίας μπορούν να οδηγήσουν στη διαγραφή πολύτιμων πληροφοριών και να αλλάξουν την κατανομή της πλειοψηφικής κλάσης. Διαφορετικές μέθοδοι δειγματοληψίας πρέπει να εξεταστούν για την ανισορροπία τάξης [103]. Παρόλο που υπάρχουν τεχνικές όπως το SMOTEBagging, το RUSBoost και το UnderBagging, οι μέθοδοι επιλογής χαρακτηριστικών για την ανισορροπία τάξης παραμένουν ένας ανεξερεύνητος τομέας. Όπως προτείνεται από τον [105], η Επαναληπτική Τυχαία Δειγματοληψία (IRS) είναι πιο αξιόπιστη από τη δειγματοληψία Kennard-Stone (KS).

2.5 Μηχανική μάθηση & Προβλεπτική συντήρηση.

Η χρήση σε διαφορετικούς τομείς, δείχνει ότι ο Εντοπισμός Ανωμαλιών είναι ένας ενεργός χώρος έρευνας [106]. Οι προσεγγίσεις κατηγοριοποίησης φαίνονται μη ρεαλιστικές χωρίς αρκετή ποσότητα επισημασμένων δεδομένων [107]. Το Netflix και το Twitter πρότειναν μεθόδους εντοπισμού ανωμαλιών, βασισμένες στην Ισχυρή Ανάλυση Βασικών Συνιστωσών και Εποχιακή Υβριδική Υπερβολική Μαθητοποιημένη Απόκλιση [108], [109]. Επιπλέον, προτάθηκε μεθοδολογία που συνδυάζει δεδομένα δονήσεων και πιέσεων κυλίνδρων για την παρακολούθηση της κύριας μηχανής. Δυο διαφορετικές προσεγγίσεις βασίζονται σε θερμοδυναμικά μοντέλα για τον εντοπισμό βλαβών μηχανών [110]. Οι Ραπτοδήμος και Λαζάκης [111] ερεύνησαν την δυνατότητα του SOM στην παρακολούθηση της κύριας μηχανής. Οι Λαζάκης, Γκερέκος και Θεοτοκάτος [112] πρότειναν μια SVM-οδηγούμενη προσέγγιση για την εκτίμηση των συστημάτων ενός πλοίου. Οι Kowalski et al. [113] χρησιμοποίησαν Υπερβολική Μηχανική Μάθηση για την ταξινόμηση δεδομένων από διαφορετικές βλάβες. Τα δεδομένα συνήθως είναι στη μορφή χρονοσειράς και θα πρέπει να χρησιμοποιηθεί η Ανάλυση Κύριων Συνιστωσών (PCA - Principal Component Analysis) [114] για τη μείωση των διαστάσεων.

3. Μεθοδολογία και Πρακτικές

3.1 Σχεδιασμός και Προσέγγιση της Έρευνας

Για τη δημιουργία μοντέλων μηχανικής μάθησης χρησιμοποιούνται επαναληπτικές διαδικασίες που αναλύουν και αξιολογούν διάφορα μοντέλα ως προς την καταλληλότητα και την απόδοσή τους. Η διαδικασία CRISP-DM είναι μια από τις πιο διαδεδομένες και περιλαμβάνει πέντε βήματα για τη στρογγυλοποίηση ενός συστήματος εξόρυξης γνώσης. Αν και αρχικά σχεδιάστηκε για συστήματα εξόρυξης γνώσης, η διαδικασία είναι ευέλικτη και μπορεί να προσαρμοστεί σε διάφορες ανάγκες ενός έργου. Στη συγκεκριμένη περίπτωση, η CRISP-DM χρησιμοποιήθηκε με ορισμένες προσαρμογές για την προγνωστική συντήρηση του εξοπλισμού ενός δικτύου.

Βήμα 1: Κατανόηση του Προβλήματος και των Επιχειρησιακών Απαιτήσεων

Το πρώτο βήμα της διαδικασίας CRISP-DM εστιάζει στην περιγραφή των στόχων με επιχειρηματικούς όρους. Επιπλέον, περιγράφει σχετικά κριτήρια επιτυχίας και δείκτες που θα σηματοδοτούσαν την επίτευξη των επιχειρηματικών στόχων. Αυτοί οι στόχοι μεταφράζονται επίσης σε πιο τεχνικούς στόχους μηχανικής μάθησης και εξόρυξης δεδομένων, δηλαδή τεχνικούς στόχους που εάν επιτευχθούν θα σηματοδοτήσουν την επιχειρηματική επιτυχία του έργου. Ως χαρακτηριστικό παράδειγμα, ένας στόχος επιχειρηματικής βελτίωσης για τη διαχείριση περιουσιακών στοιχείων μπορεί να συνοψιστεί σε δείκτες επιτυχίας όπως:

- Πρόβλεψη του RUL (Remaining Useful Life) για ένα περιουσιακό στοιχείο, ως μέσο για τη βελτιστοποίηση του προγράμματος συντήρησης του περιουσιακού στοιχείου.
- Επισήμανση ακανόνιστης ή ανώμαλης συμπεριφοράς, ως μέσο ανίχνευσης προτύπων ελαττωμάτων ή έναρξης περιόδου υποβάθμισης ενός περιουσιακού στοιχείου.
- Παραγωγή συστάσεων σχετικά με τη συντήρηση, προκειμένου να μετριάσουν προβλήματα ποιότητας μετά τον εντοπισμό ενός προτύπου ελαττώματος.

Στο πλαίσιο του βήματος κατανόησης του προβλήματος, παράγεται επίσης ένα αρχικό σχέδιο έργου, το οποίο περιλαμβάνει πληροφορίες σχετικά με τις δραστηριότητες που πρέπει να αναληφθούν, τους απαιτούμενους πόρους, τα αναμενόμενα αποτελέσματα και τις συσχετίσεις μεταξύ διαφορετικών εργασιών. Ακόμα γίνεται μια αρχική αξιολόγηση του υλικού (π.χ. υπολογιστικές πλατφόρμες), του λογισμικού (π.χ. πλατφόρμες και εργαλεία εξόρυξης δεδομένων), των δεδομένων (π.χ. σύνολα διαθέσιμων δεδομένων) και των ανθρώπινων πόρων που διατίθενται για την υλοποίηση του έργου.

Βήμα 2: Κατανόηση Δεδομένων

Δεν είναι δυνατή η ανάπτυξη αξιόπιστων συστημάτων μηχανικής μάθησης και τεχνητής νοημοσύνης χωρίς τα κατάλληλα δεδομένα. Επομένως, το δεύτερο βήμα της μεθοδολογίας CRISP-DM επικεντρώνεται στην κατανόηση των διαθέσιμων δεδομένων, ως μέσο για να αποφασιστεί εάν επαρκούν για το δεδομένο πρόβλημα. Σε αυτή την κατεύθυνση οι επιστήμονες δεδομένων της ομάδας επιθεωρούν και να οπτικοποιούν τα διαθέσιμα στοιχεία δεδομένων, προκειμένου να κατανοήσουν τον πλούτο, το εύρος και τις στατιστικές τους ιδιότητες. Για το σκοπό αυτό, αξιοποιήθηκαν εργαλεία εξόρυξης δεδομένων και πραγματοποιήθηκε οπτικοποίηση των δεδομένων σε διαφορετικά γραφήματα.

Το βήμα κατανόησης δεδομένων περιλαμβάνει την εξέταση και αναγνώριση των στατιστικών κατανομών ορισμένων βασικών χαρακτηριστικών των δεδομένων (π.χ. δεδομένα αισθητήρων), καθώς και τον εντοπισμό πιθανών αξιοσημείωτων συσχετίσεων μεταξύ διαφορετικών χαρακτηριστικών. Τέτοιες στατιστικές πληροφορίες βοηθούν τους επιστήμονες δεδομένων στον καθορισμό υποψήφιων μοντέλων μηχανικής μάθησης, όπως μοντέλα που θα μπορούσαν να αντλήσουν με επιτυχία προγνωστικές πληροφορίες από συγκεκριμένες παραμέτρους. Στο πλαίσιο αυτής της διαδικασίας, οι επιστήμονες δεδομένων ενδέχεται να χρειαστεί να συμβουλευτούν ειδικούς του πεδίου, προκειμένου να κατανοήσουν ποια χαρακτηριστικά μιας διαδικασίας παραγωγής και μιας διαδικασίας ποιοτικού ελέγχου μπορούν να χρησιμοποιηθούν ως προγνωστικοί παράγοντες της μελλοντικής κατάστασης του εξοπλισμού.

Η διαδικασία κατανόησης δεδομένων θα παρέχει επίσης χρήσιμες πληροφορίες σχετικά με τον όγκο και την ποιότητα των συνόλων δεδομένων. Για παράδειγμα, σύνολα δεδομένων από τα οποία λείπουν πολλές τιμές ενδέχεται να είναι ακατάλληλα για την εκπαίδευση μοντέλων μηχανικής μάθησης. Ομοίως, ένα σύνολο δεδομένων με πολύ λίγα περιστατικά που σχετίζονται με βλάβες του εξοπλισμού δεν θα αρκεί για την δημιουργία κατάλληλων μοντέλων μηχανικής μάθησης. Ως εκ τούτου, υπάρχει πάντα η δυνατότητα να συμπεράνουμε ότι τα διαθέσιμα στοιχεία δεν είναι επαρκή ή κατάλληλα για την αντιμετώπιση του προς επίλυση προβλήματος προγνωστικής συντήρησης. Σε τέτοιες περιπτώσεις, η ομάδα μπορεί να χρειαστεί να πάει πίσω στο βήμα της επιχειρηματικής κατανόησης προκειμένου να διατυπώσει διαφορετικά το πρόβλημα, με τρόπο που να μπορεί να επιλυθεί με βάση τα διαθέσιμα δεδομένα.

Βήμα 3: Προετοιμασία Δεδομένων

Το βήμα προετοιμασίας δεδομένων ξεκινά με την επιλογή των δεδομένων που πρέπει να αναλυθούν προκειμένου να αντιμετωπιστεί το υπάρχον προς επίλυση πρόβλημα. Μετά τη φάση κατανόησης των δεδομένων, η ομάδα θα πρέπει να έχει προσδιορίσει τα μέρη των συνόλων δεδομένων που σχετίζονται

με το έργο. Η επιλογή των δεδομένων θα επιτρέψει στους προγραμματιστές να εφαρμόσουν κατάλληλες μεθόδους για την ανάκτηση των απαιτούμενων δεδομένων από τα διαφορετικά συστήματα και τις πηγές όπου βρίσκονται. Μετά την επιλογή των δεδομένων, στο βήμα αυτό πραγματοποιείται επίσης και καθαρισμός δεδομένων προκειμένου να βελτιωθεί η ποιότητά τους.

Πέρα από τον καθαρισμό δεδομένων, το βήμα προετοιμασίας δεδομένων περιλαμβάνει την προεπεξεργασία των επιλεγμένων δεδομένων προκειμένου να μορφοποιηθούν κατάλληλα όπως απαιτούν τα μοντέλα που πρέπει να δοκιμαστούν. Ως μέρος αυτής της διαδικασίας, θα πρέπει να υπολογιστούν και να προστεθούν στο σύνολο δεδομένων ορισμένα νέα χαρακτηριστικά, ενώ άλλα χαρακτηριστικά θα πρέπει να μετατραπούν σε διαφορετικές μορφές. Για παράδειγμα, ορισμένοι αλγόριθμοι μπορεί να απαιτούν ως είσοδο τον συνολικό αριθμό βλαβών, ο οποίος μπορεί να προκύψει με βάση έναν απλό υπολογισμό πάνω στο πρωτογενές σύνολο δεδομένων. Ως άλλο παράδειγμα, παράμετροι όπως ημερομηνίες και ώρες, ενδέχεται να πρέπει να μετατραπούν σε μια κοινή μορφή δεδομένων.

Βήμα 4: Μοντελοποίηση και Εκπαίδευση Μοντέλων Μηχανικής Μάθησης

Αυτή η φάση είναι η ουσία της διαδικασίας μηχανικής μάθησης. Περιλαμβάνει την επιλογή του μοντέλου μηχανικής μάθησης που θα χρησιμοποιηθεί για την εξαγωγή γνώσης από τα υπάρχοντα σύνολα δεδομένων. Στις περισσότερες περιπτώσεις επιλέγονται, εξετάζονται και αποτιμώνται διαφορετικά μοντέλα μηχανικής μάθησης ανάλογα με τις απαιτήσεις του προβλήματος. Για παράδειγμα τα μοντέλα παλινδρόμησης μπορούν να παρέχουν έναν σχετικά απλό τρόπο για την πρόβλεψη του RUL, ενώ τεχνικές μηχανικής εκμάθησης χωρίς επίβλεψη (unsupervised learning) μπορούν να χρησιμοποιηθούν για τον εντοπισμό πιθανών μη φυσιολογικών συμπεριφορών για ένα στοιχείο.

Πολλά προβλήματα απαιτούν την ανάπτυξη πιο περίπλοκων μοντέλων, όπως τα βαθιά νευρωνικά δίκτυα. Ως χαρακτηριστικό παράδειγμα, τα δίκτυα μακροπρόθεσμης βραχυπρόθεσμης μνήμης (LSTM) είναι εξαιρετικά αποτελεσματικά στην εκμάθηση εξαγωγής προγνωστικών πληροφοριών με βάση μεγάλες ιστορικές ακολουθίες δεδομένων που περιέχουν επαρκή αριθμό βλαβών.

Η δοκιμή και η επικύρωση κάθε διαφορετικού μοντέλου περιλαμβάνει την εκπαίδευση του μοντέλου με βάση ένα υποσύνολο των διαθέσιμων δεδομένων και τη δοκιμή του σε ένα άλλο υποσύνολο δεδομένων. Τα σύνολα δεδομένων δοκιμής πρέπει να είναι διαφορετικά από τα σύνολα δεδομένων που χρησιμοποιούνται για την εκπαίδευση, διαφορετικά η διαδικασία μοντελοποίησης δεν είναι αξιόπιστη. Ως μέρος της διαδικασίας δοκιμής, η ομάδα λαμβάνει πληροφορίες σχετικά με την ακρίβεια του μοντέλου, μέσω της

αξιολόγησης παραμέτρων όπως τα ποσοστά σφαλμάτων ταξινόμησης ή η ακρίβεια του μοντέλου δηλαδή το κλάσμα των περιπτώσεων αστοχίας προς το συνολικό αριθμό περιπτώσεων κατάστασης του εξοπλισμού που υπάρχουν στα δεδομένα.

Στο πλαίσιο της διαδικασίας μοντελοποίησης, οι επιστήμονες δεδομένων συνεργάζονται με ειδικούς του τομέα προκειμένου να συζητήσουν τα ευρήματά τους σε σχέση με την ακρίβεια και την καταλληλότητα κάθε μοντέλου. Οι ειδικοί του πεδίου είναι πιθανό να συνεισφέρουν πληροφορίες του τομέα σχετικά με τα πραγματικά χαρακτηριστικά που πρέπει να ληφθούν υπόψη κατά τη διαδικασία δημιουργίας μοντέλου (π.χ. ποια χαρακτηριστικά ενός συνόλου δεδομένων είναι αντιπροσωπευτικά για την κατάσταση ενός στοιχείου). Επιπλέον, οι ειδικοί του πεδίου μπορούν να βοηθήσουν τους επιστήμονες δεδομένων στην ανάπτυξη γενικευμένων μοντέλων που είναι πιθανό να λειτουργήσουν μακροπρόθεσμα, αντί για πιο συγκεκριμένα μοντέλα που λειτουργούν μόνο για τα διαθέσιμα σύνολα δεδομένων.

Βήμα 5: Αξιολόγηση Μοντέλων Μηχανικής Μάθησης

Το τελευταίο βήμα της διαδικασίας CRISP-DM επικεντρώνεται στην αξιολόγηση των παραγόμενων μοντέλων. Κάθε μοντέλο αξιολογείται με βάση την καταλληλότητά του για το προς επίλυση πρόβλημα, πέρα από την ακρίβειά του σε σχέση με τα σύνολα δεδομένων δοκιμής. Στο πλαίσιο της αξιολόγησης, παράμετροι όπως η απλότητα, η ταχύτητα εκπαίδευσης και το πόσο κατανοητό είναι το μοντέλο μηχανικής μάθησης [115] ενδέχεται να ληφθούν υπ' όψη. Ακόμα εκτιμώνται παράμετροι όπως ο χρόνος εκπαίδευσης του μοντέλου και το συνολικό κόστος της λύσης. Το τελικό αποτέλεσμα του βήματος αυτού μπορεί να είναι η πραγματική ανάπτυξη και χρήση ενός μοντέλου μηχανικής μάθησης στην παραγωγή. Για το τελευταίο απαιτείται πρόσθετη ανάπτυξη λογισμικού, μεταξύ αυτών και κατάλληλων διεπαφών χρήστη και οπτικοποιήσεων. Ωστόσο, το βήμα αξιολόγησης μπορεί επίσης να οδηγήσει στην αναζήτηση ενός νέου μοντέλου (δηλαδή μετάβαση στη φάση μοντελοποίησης) ή ακόμα και στην αναζήτηση ενός εναλλακτικού προβλήματος (δηλαδή μετάβαση πίσω στη φάση της επιχειρηματικής κατανόησης).

3.2 Χρησιμοποιούμενα σύνολα δεδομένων.

Τα δεδομένα που χρησιμοποιήθηκαν στην δημιουργία του εργαλείου διατέθηκαν από μια ναυτιλιακή εταιρεία που συγκεντρώνει ροές δεδομένων από όλους τους αισθητήρες που χρησιμοποιούνται στα πλοία. Κάθε ροή παριστάνει τις ιστορικές εγγραφές που συλλέγονται με το χρόνο από έναν συγκεκριμένο αισθητήρα στο πλοίο. Όλα αυτά δειγματοληπτούνται βασισμένα σε χρονική περίοδο ενός λεπτού ή δέκα λεπτών. Αυτά επίσης όταν χρησιμοποιούνται σαν είσοδοι των μοντέλων αναφέρονται και ως «χαρακτηριστικά» και παρουσιάζονται περιληπτικά

στον παρακάτω Πίνακας 1. Το σύνολο των διαθέσιμων δεδομένων αφορά ιστορικά δεδομένα από 10 πλοία ίδιου τύπου.

Πλέον των ανωτέρω δεδομένων αισθητήρων, είναι απαραίτητα και τα δεδομένα (χρονικές στιγμές) εμφάνισης μιας συγκεκριμένης βλάβης, τα οποία και διετέθησαν. Η βλάβη αφορά στη φθορά του ρουλεμάν σταυρού κεφαλής που συνήθως οφείλεται στην πορεία των πλοίων με μικρότερη ταχύτητα από τα προβλεπόμενα για λόγους μείωσης του κόστους. Η μεθοδολογία του παρόντος εργαλείου έχει βασιστεί στο συγκεκριμένο σύνολο δεδομένων.

Πρέπει να σημειωθεί ότι η προτεινόμενη μεθοδολογία απαιτεί τροποποίηση για να εφαρμοστεί σε διαφορετικό σύνολο δεδομένων σε διαφορετική εταιρία.

Πίνακας 1: Χαρακτηριστικά συνόλου δεδομένων από αισθητήρες πλοίων κύριας μηχανής (main engine data), και γενικά δεδομένα 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)
Wind angle (0-359.99 degrees)	Fuel Oil Inlet Pressure (Pa)
Total Twenty-foot Equivalent Unit (TEU) (# of containers)	Fuel Oil Inlet Temperature (°C)
Total Forty-foot Equivalent Unit (FEU) (# of containers)	Jacket Cooling Fresh Water Inlet Temperature Low (°C)
Low-sulfur fuel oil consumption (metric tons),	Cylinder #1-10 Exhaust Gas Out Temperature (°C)
High-sulfur fuel oil consumption (metric tons)	Cylinder #1-10 Jacket Cooling Fresh Water Outlet Temp
Vessel draft at port-side (left-side looking to the fore) (m)	Cylinder #1-10 Piston Cooling Outlet Temperature (°C)
Vessel draft at starboard-side	Turbo-Charger #1-3 Exhaust Gas Inlet Temperature (°C)
Vessel draft at stern (m)	Turbo-Charger #1-3 Exhaust Gas Outlet Temperature (°C)
Vessel draft at fore (m)	Turbo-Charger #1-3 Lube Oil Inlet Pressure (Pa)
Speed through water (kn)	Turbo-Charger #1-3 Lube Oil Outlet Pressure (Pa)
Total number of containers	Turbo-Charger #1-3 RPMs
Vessel draft at mid-line (m)	Scavenge Air Inlet Pressure (Pa)
The trim of the vessel, calculated by draft aft and draft fore	Scavenge Air Receiver Temperature (°C)

The latitude of the vessel's position	Starting Air Pressure (Pa)
The longitude of the vessel's position	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)
	Turbo-Charger #1-3 Air Cooler Cooling Water Outlet Temp.

Για τη συνδεσιμότητα των αισθητήρων και την επικοινωνία με το κεντρικό σύστημα, στο οποίο γίνεται η συγκέντρωση και η επεξεργασία των μετρήσεων αισθητήρων, δημιουργήθηκε και χρησιμοποιείται ένα API το οποίο δύναται να δέχεται HTTP requests και με τα δεδομένα του πλοίου καθώς και από τους αισθητήρες της κύριας μηχανής.

Συγκεκριμένα, όταν λαμβάνονται μετρήσεις από τους αισθητήρες του συστήματος προβλεπτικής συντήρησης, αυτά τα δεδομένα αποθηκεύονται στη βάση δεδομένων MongoDB, η οποία έχει σχεδιαστεί για ταχεία ανάκτηση των δεδομένων. Στη συνέχεια, οι υπηρεσίες προβλεπτικής συντήρησης χρησιμοποιεί αυτά τα δεδομένα για να παράγει προβλέψεις σχετικά με την κατάσταση των μηχανών πλοίων. Επιπλέον, η βάση δεδομένων αποθηκεύει επιπλέον ιστορικά δεδομένα των πλοίων είναι διαθέσιμα για ανάλυση και σύγκριση στο μέλλον.

Ως μέρος της έρευνάς μας, χρησιμοποιήσαμε σύνολα δεδομένων που προέρχονται από το αποθετήριο 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.

3.3 Προετοιμασία - Προ-επεξεργασία Δεδομένων

Τα δεδομένα που συλλέγονται για ανάλυση μπορεί να περιέχουν ασυνεπείς ή ελαττωματικές τιμές λόγω διάφορων σφαλμάτων. Για να αντιμετωπιστούν αυτά τα ζητήματα, εφαρμόστηκαν στάδια προεπεξεργασίας στα δεδομένα χρονοσειρών. Αρχικά, τα δεδομένα δειγματίστηκαν σε χρονικά πλαίσια (π.χ. ωρών) με σκοπό την μείωση του θορύβου. Οι ελλιπείς τιμές αντικαταστάθηκαν χρησιμοποιώντας τεχνικές όπως το "Forward Fill" και το "Backward Fill". Επιπλέον, διορθώθηκαν ακραίες τιμές που ήταν αποτέλεσμα σφαλμάτων αισθητήρα ή μετάδοσης δεδομένων. Συγκεκριμένα:

- Ποιότητα των Δεδομένων: Τα δεδομένα που συλλέγονται για ανάλυση μπορεί να παρουσιάζουν ασυνεπείς ή ελαττωματικές τιμές λόγω ανθρωπίνων σφαλμάτων, σφαλμάτων αισθητήρα ή λανθασμένων καταγραφών.
- Προεπεξεργασία Δεδομένων χρονοσειρών: Για τη βελτίωση της ποιότητας των δεδομένων, εφαρμόστηκαν διάφορα στάδια προεπεξεργασίας, ειδικά για δεδομένα χρονοσειρών.
- Επαναδειγματοληψία: Τα δεδομένα δειγματίστηκαν σε χρονικά πλαίσια, όπως ωρών, υπολογίζοντας τις μέσες, μέγιστες και ελάχιστες τιμές. Αυτό επιτρέπει μια πιο καθαρή εικόνα του συνόλου δεδομένων, μειώνοντας το θόρυβο και βελτιστοποιώντας τον χρόνο και τους πόρους που απαιτούνται για την ανάλυση.
- Αντιμετώπιση Ελλιπόντων Τιμών: Ένα συνηθισμένο πρόβλημα στα δεδομένα χρονοσειρών είναι οι ελλιπείς τιμές. Αντί να αγνοηθούν τα δεδομένα που λείπουν, χρησιμοποιήθηκαν τεχνικές όπως το "Forward Fill", όπου μια ελλιπής τιμή αντικαθίσταται από την προηγούμενη γνωστή τιμή, και το "Backward Fill", όπου μια ελλιπής τιμή αντικαθίσταται από την επόμενη γνωστή τιμή.
- Διόρθωση Ακραίων Τιμών: Τα δεδομένα μπορεί να περιέχουν ακραίες τιμές λόγω σφαλμάτων αισθητήρα ή προβλημάτων μετάδοσης. Για να αντιμετωπιστεί αυτό, οι τιμές προσαρμόστηκαν βάσει των συνθηκών περιβάλλοντος και της βιβλιογραφίας του συστήματος.

3.4 Χρησιμοποιούμενα Προβλεπτικά Μοντέλα

Για την υλοποίηση της εργασίας μας σε πραγματικά δεδομένα, πέρα την χρήση των δεδομένων με μεθόδους μηχανικής μάθησης υλοποιήσαμε και μία στατιστική μέθοδο για να καλύψουμε το ενδεχόμενο έλλειψης δεδομένων αρκετών κύκλων ζωής που μπορεί να απαιτείται από τα μοντέλα μηχανικής μάθησης. Μετά την εγκατάσταση των αισθητήρων, η εν λόγω εφαρμογή λαμβάνει ως είσοδο όλα τα δεδομένα μετρήσεων από τους αισθητήρες και τις εγγραφές βλαβών και συντηρήσεων που καταχωρεί το προσωπικό. Βάσει αυτών των δεδομένων, η εφαρμογή προγνωστικής συντήρησης εξάγει αποτελέσματα με δύο μεθόδους:

- Με χρήση στατιστικών μεθόδων, υπολογίζεται η μέση διάρκεια μεταξύ βλαβών ενός εξαρτήματος (MTTF – Mean Time To Failure) και ο χρόνος μέχρι την επόμενη βλάβη.
- Με χρήση αλγορίθμων μηχανικής μάθησης, προβλέπεται ο υπολειπόμενος ωφέλιμος χρόνο ζωής ενός εξαρτήματος (RUL – Remaining Useful Life) πριν την επόμενη βλάβη.

Η λειτουργία πρόβλεψης ενεργοποιείται από το χρήστη της πλατφόρμας μέσω περιοδικών trigger από cron-jobs κάθε ημέρα.

3.4.1 Μέσος Χρόνος για την Αποτυχία Υλικού / Εξοπλισμού (MTTF)

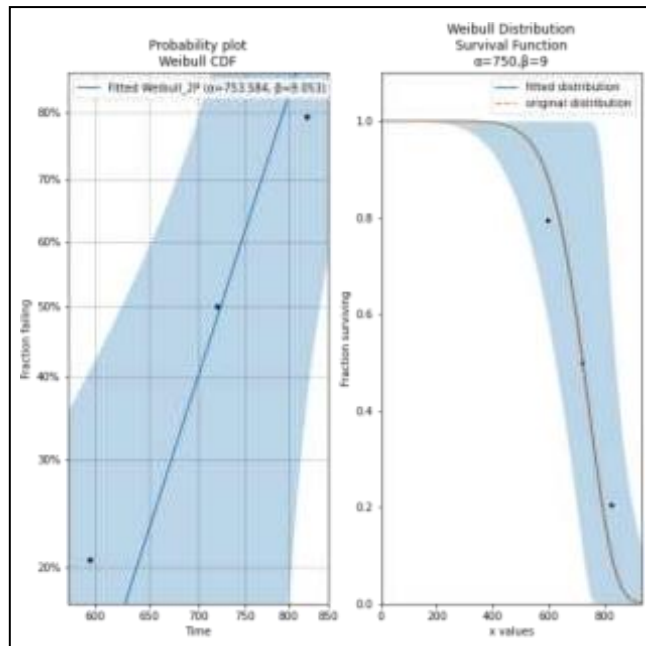
Η πρώτη λειτουργία που παρέχεται μέσω του PdM Tool αφορά την εκτίμηση του μέσου χρόνου για την αποτυχία του υλικού υπό εξέταση μέσω στατιστικής μοντελοποίησης. Για την πρόβλεψη αστοχίας ενός υλικού διάφορα επιμέρους στοιχεία πρέπει να ληφθούν υπόψη για τη στατιστική ανάλυση της διάρκειας ζωής του συστήματος.

Η κατανομή Weibull χρησιμοποιείται ευρέως για μοντελοποίηση της διάρκειας ζωής στη μηχανική αξιοπιστίας [116], ενώ έχει χρησιμοποιηθεί σε πολλές άλλες εφαρμογές, όπως η ανάλυση επιβίωσης, ηλεκτρολογική μηχανική, θεωρία ακραίων τιμών, εφαρμογές αιολικής ενέργειας, πρόγνωση καιρού και δεδομένα προσαρμογής όλων των ειδών.

Η συνάρτηση πυκνότητας πιθανότητας της τυχαίας μεταβλητής X με κατανομή Weibull με παραμέτρους α και $\beta > 0$, είναι: $f(x) = \beta/\alpha (\beta/\alpha)^{\beta-1} \exp\{-(x/\alpha)^\beta\}$

Οι ποσότητες έχουν ιδιαίτερη σημασία στην ανάλυση αξιοπιστίας και σχετίζονται με την κατανομή Weibull είναι:

- Το MTTF, $\mu = \alpha \Gamma(1 + 1/\beta)$,
- Η αξιοπιστία του υλικού (reliability), $R(x) = \exp\{-(x/\alpha)^\beta\}$
- Το ποσοστό αποτυχίας (hazard), $h(x) = \alpha^{-\beta} \beta x^{\beta-1}$



Εικόνα 6: Weibull Distribution – Probability plot & Survival function.

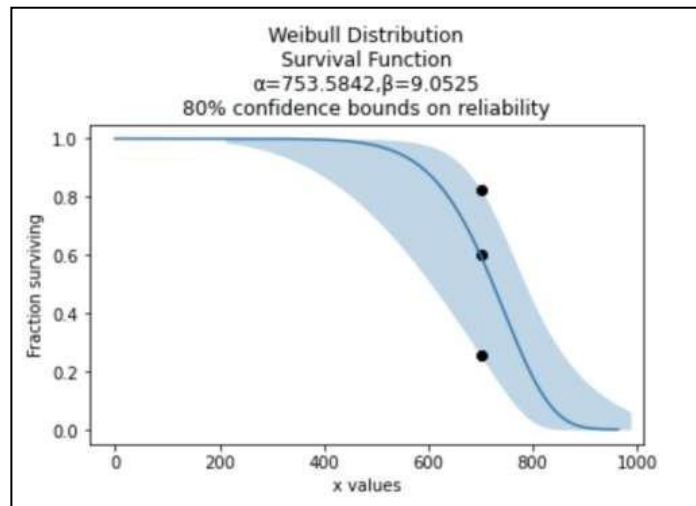
Επομένως, οι παραπάνω ποσότητες μπορούν να υπολογιστούν εκτιμώντας τις παραμέτρους β, α της κατανομής. Αυτό μπορεί να επιτευχθεί αναλυτικά μεγιστοποιώντας την μέγιστη πιθανοφάνεια (MLE estimator).

Για την υλοποίηση αυτού του παραμετρικού μοντέλου χρησιμοποιήθηκε η βιβλιοθήκη της Python *RELIABILITY* [117] που εξειδικεύεται στη μηχανική αξιοπιστίας και την ανάλυση επιβίωσης, ενώ περιλαμβάνει επίσης πολλά εξειδικευμένα εργαλεία που κατά τα άλλα είναι διαθέσιμα μόνο σε ιδιόκτητο λογισμικό.

Επισημαίνεται ότι τα δεδομένα εισόδου αφορούν τον χρόνο (σε ημέρες λειτουργίας) έως την εμφάνιση βλάβης. Στο υποκείμενο σύνολο δεδομένων υπάρχουν 4 βλάβες με αποτέλεσμα να μπορούν να δημιουργηθούν μόνο 3 data points.

Οι παραπάνω εικόνες (Εικόνα 6 και Εικόνα 7) απεικονίζουν το fit των (τεχνητών) δεδομένων στην Weibull κατανομή καθώς και τις εκτιμώμενες τιμές των παραμέτρων της κατανομής (α, β). Σύμφωνα με τις οποίες μετά από 800 ώρες λειτουργίας υπάρχει 80% πιθανότητα βλάβης. Οι χαμηλές τιμές των AIC και BIC score υποδηλώνουν ότι η εν λόγω κατανομή ταιριάζει στα δεδομένα. Το MTTF εκτιμήθηκε στις 710 ώρες λειτουργίας.

Πρακτικά, το εν λόγω στατιστικό μοντέλο προβλέπει το μέσο χρόνο για την εμφάνιση βλάβης στο εποπτευόμενο μηχάνημα / εξοπλισμό βάσει των ιστορικών βλαβών και του χρόνου που έχει μεσολαβήσει από την τελευταία βλάβη στο εν λόγω μηχάνημα.

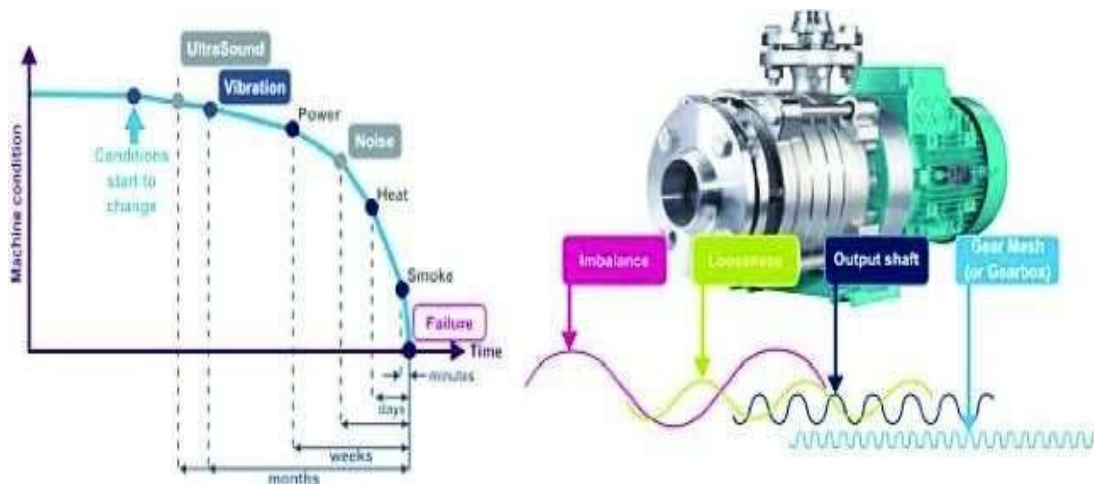


Εικόνα 7: Weibull Distribution, Survival function x Fraction surviving.

3.4.2 Υπολειπόμενη χρήσιμη απόδοση κατά τη διάρκεια ζωής του υλικού (RUL)

Η δεύτερη παρεχόμενη λειτουργία αφορά την πρόβλεψη της υπολειπόμενης χρήσιμης ζωής (RUL) του εξοπλισμού / υλικού. Στα προγνωστικά μοντέλα, η υποβάθμιση ενός εξαρτήματος ή συστήματος είναι συνήθως μια μη γραμμική συνάρτηση πολλών παραμέτρων όπως το περιβάλλον λειτουργίας και το φόρτο εργασίας. Κανονικά, ένα σύστημα ή ένα εξάρτημα βρίσκεται σε καλή κατάσταση λειτουργίας σε πρώιμο στάδιο χρήσης. Η απόδοση του συστήματος αρχίζει να υποβαθμίζεται μετά τη λειτουργία για κάποιο χρονικό διάστημα. Η διαδικασία αποικοδόμησης επιταχύνεται με την πάροδο του χρόνου έως ότου συμβεί πλήρης διάσπαση. Οι καμπύλες υποβάθμισης συνήθως αντιπροσωπεύουν μεταβάσεις από μια περίπου σταθερή τιμή σε μια γραμμικά φθίνουσα καμπύλη προς το τέλος της ζωής. Η απότομη κλίση στην καμπύλη αντανακλά το σημείο που ξεκινά η υποβάθμιση.

Η πιο εξέχουσα προσέγγιση για την εκτίμηση του RUL βασίζεται σε μεθόδους που βασίζονται σε δεδομένα, οι οποίες βασίζονται αποκλειστικά σε διαθέσιμα δεδομένα που παρατηρήθηκαν στο παρελθόν και σε στατιστικά μοντέλα. Μεταξύ των μεθόδων που βασίζονται σε δεδομένα, τα βαθιά νευρωνικά δίκτυα εφαρμόζονται ευρέως για προβλήματα ταξινόμησης και πρόβλεψης ακολουθιών. Λόγω της ικανότητάς τους να μαθαίνουν μακροπρόθεσμες εξαρτήσεις, τα δίκτυα LSTM έχουν αποδειχθεί ευρέως ότι είναι χρήσιμα για την εκμάθηση ακολουθιών με μακροπρόθεσμα μοτίβα. Πολλαπλές ερευνητικές ομάδες έχουν χρησιμοποιήσει το LSTM για την εκτίμηση RUL και πρότειναν δίκτυα LSTM ενσωματωμένα με πλήρως συνδεδεμένα νευρωνικά δίκτυα (NN), με τα αποτελέσματά να τους δείχνουν ότι το μοντέλο τους υπερέχει άλλων μεθόδων βαθιάς μάθησης στην εκτίμηση RUL σε πολλαπλά σύνολα δεδομένων.

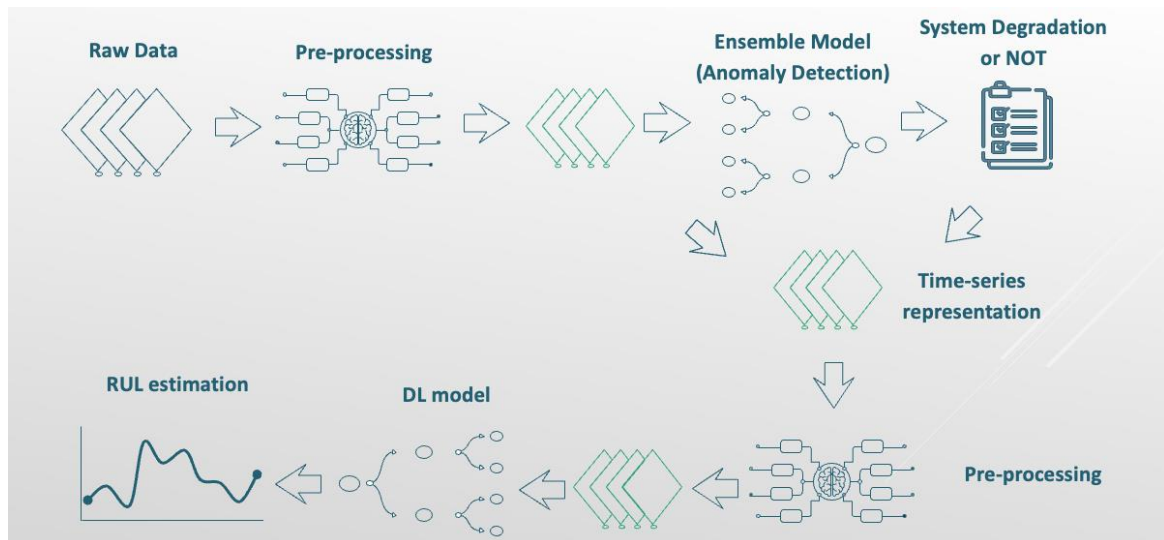


Εικόνα 8: Εξέλιξη κατάστασης μηχανής πλοίου.

Για τις ξαφνικές βλάβες, η λειτουργία υποβάθμισης ξεκινά πολύ κοντά στην αστοχία, με συνέπεια η εκτίμηση RUL να είναι πολύ δύσκολη σε τέτοιες περιπτώσεις. Στο βήμα της εκμάθησης, η μεγάλη ποσότητα κανονικών δεδομένων μπορεί να κρύψει το μοτίβο υποβάθμισης στο σύνολο δεδομένων. Στο βήμα υλοποίησης, η πραγματοποίηση εκτίμησης RUL ενώ το σύστημα είναι ακόμα σε υγιή τρόπο λειτουργίας και οι αισθητήρες δεν δείχνουν σημάδια υποβάθμισης μπορεί να οδηγήσει σε ανούσιες εκτιμήσεις και να βλάψει την αξιοπιστία του μοντέλου εκτιμητή RUL. Σε αυτή την εργασία, ακολουθήθηκε η προσέγγιση της [118] όπου χρησιμοποιείται μία μονάδα έγκαιρης ανίχνευσης σφαλμάτων για την ανίχνευση υποβαθμίσεων του συστήματος στα αρχικά στάδια. Όταν ανιχνεύονται οι τρόποι υποβάθμισης, εφαρμόζουμε μια δομή δικτύων LSTM που βασίζεται στους [119] για την εκτίμηση του συστήματος RUL.

Επομένως, το μοντέλο για την πρόβλεψη RUL αποτελείται από δύο υπομοντέλα. Η πρώτη χρησιμοποιεί έναν ταξινομητή Random για να προβλέψει εάν το εν λόγω σύστημα βρίσκεται σε κανονική λειτουργία ή σε λειτουργία υποβάθμισης. Στην περίπτωση κανονικής λειτουργίας, η εκτίμηση RUL δεν έχει έξοδο. Ωστόσο, όταν ο ταξινομητής ανιχνεύει υποβάθμιση σε ένα σύστημα, τα δεδομένα του αισθητήρα χρησιμοποιούνται ως είσοδος στον εκτιμητή LSTM RUL, ο οποίος εξάγει την υπολειπόμενη ωφέλιμη ζωή του υποβαθμισμένου εξαρτήματος / συστήματος.

Όπως φαίνεται στο παρακάτω σχήμα (Εικόνα 9), τα δεδομένα εισόδου από τους αισθητήρες υποβάλλονται σε προεπεξεργασία προκειμένου να εξαχθούν χαρακτηριστικά που σχετίζονται με τη δυσλειτουργία του εποπτευόμενου συστήματος και να πληρούν τις απαιτήσεις των μοντέλων ML, δηλαδή κανονικοποίηση, καταλογοισμός τιμών που λείπουν κτλ. Στη συνέχεια, τα μετασχηματισμένα δεδομένα εγχέονται στον ταξινομητή που εκτιμά εάν το σύστημα βρίσκεται σε κανονική λειτουργία ή σε λειτουργία υποβάθμισης. Σε περίπτωση κανονικής λειτουργίας, το RUL δεν μπορεί να εκτιμηθεί. Εάν ο ταξινομητής ανιχνεύσει υποβάθμιση, τα δεδομένα του αισθητήρα τροφοδοτούνται στον εκτιμητή LSTM RUL που εξάγει το RUL.



Εικόνα 9: Προσέγγιση εκτίμησης RUL.

Σημειώνεται ότι, για την εκπαίδευση του LSTM πρέπει για κάθε χρονική στιγμή που είναι διαθέσιμες οι τιμές των αισθητήρων, να αντιστοιχηθεί και μία τιμή του RUL. Αυτό επιτυγχάνεται υπολογίζοντας στα ιστορικά δεδομένα τις ώρες λειτουργίας του εξαρτήματος μέχρι την εμφάνιση της βλάβης.

3.4.3 Μοντέλα Μηχανικής Μάθησης – Στατιστικά Μοντέλα

Στη συγκεκριμένη εργασία, ασχοληθήκαμε με το πρόβλημα της ανίχνευσης ανωμαλιών σε δεδομένα, προκειμένου να εντοπίσουμε πιθανά προβλήματα στα ρουλεμάν σταυρού κεφαλής μηχανών πλοίων. Δεδομένης της αβεβαιότητας σχετικά με το ποιος τύπος ανωμαλίας θα έπρεπε να ανιχνεύεται - είτε παρατεταμένες, είτε στιγμιαίες - διερευνήθηκαν τόσο μέθοδοι που εστιάζουν σε συνεχή τμήματα δεδομένων (χρονικά παράθυρα) όσο και σε μεμονωμένα, συγκεκριμένα χρονικά σημεία.

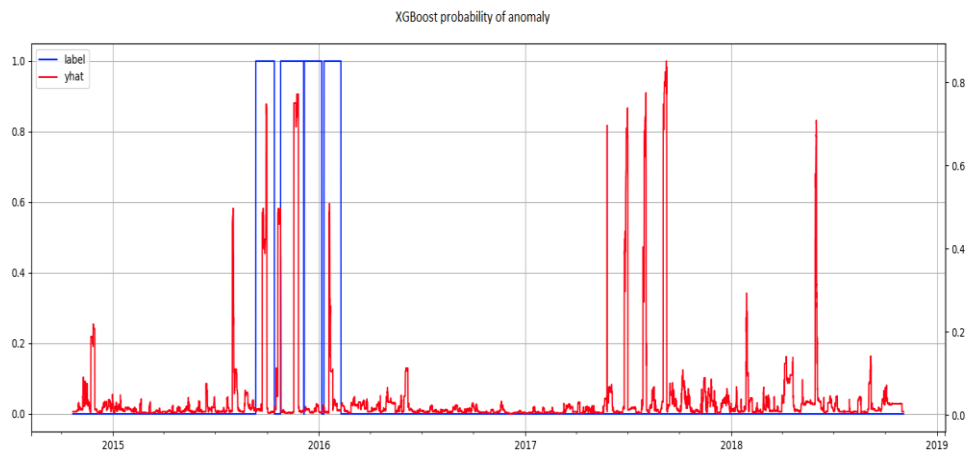
Γνωρίζοντας ότι οι σημειακές ανωμαλίες μπορεί να είναι κρίσιμες ενδείξεις βλάβης, ενσωματώθηκαν μοντέλα μηχανικής μάθησης που μπορούν να ανταποκριθούν αποτελεσματικά στον εντοπισμό τους. Τελικά, χρησιμοποιήθηκαν τόσο επιβλεπόμενα όσο και ημι-επιβλεπόμενα μοντέλα έτσι ώστε να καλυφθεί όλο το φάσμα των δυνατών ανωμαλιών:

Ταξινόμηση Gradient Boosting:

Τα δέντρα κατηγοριοποίησης και αναδρομής είναι ικανά να χρησιμοποιηθούν σε ζητήματα εντοπισμού ανωμαλιών με διαφορετικές προσεγγίσεις. Πρώτα απ' όλα, είναι δυνατόν να εφαρμοστούν σε συστήματα εποπτευόμενης μάθησης, διακρίνοντας ανώμαλες από φυσιολογικές περιπτώσεις. Για την υλοποίηση αυτού, είναι απαραίτητος ο σημασιολογικός χαρακτηρισμός των δεδομένων, διαδικασία που γνωρίζουμε ως labeling [120]. Στη συγκεκριμένη έρευνα, έγινε χρήση της

βιβλιοθήκης XGBoost, η οποία αντιπροσωπεύει το eXtreme Gradient Boosting [46]. Χάρη σ' αυτήν την τεχνική, το ζήτημα της ανίχνευσης ανωμαλιών αναγνωρίστηκε ως ένα ζήτημα κατηγοριοποίησης. Με βάση αυτό, καθορίστηκε (μετά από συμβουλές από πλοίαρχους) ένα σταθερό χρονικό διάστημα, υποθέτοντας ότι κατά τη διάρκειά του συμβαίνει η βλάβη. Έτσι, επισημάνθηκαν τα δεδομένα με ετικέτα '1' ή 'Αληθές' για ένα διάστημα 30 ημερών προτού εμφανιστεί κάποιο γνωστό σφάλμα. Επίσης, αποφασίστηκε για αυτό το διάστημα, λόγω της υπόθεσης, ότι ένα πλοίο ενδεχομένως να συνεχίσει το ταξίδι του με ένα μικρό πρόβλημα στο ρουλεμάν για πάνω από 30 μέρες. Αυτή η υπόθεση αξιολογήθηκε για διάφορες περιόδους, και τα αποτελέσματα βασίστηκαν στις δοκιμασίες τους. Για το συγκεκριμένο πείραμα, οι δοκιμές έλαβαν χώρα για 5, 10, 30 και 60 μέρες. Επομένως, τα δεδομένα που επισημάνθηκαν με βάση το προαναφερθέν 30ήμερο διάστημα αποτέλεσαν το σετ δεδομένων για το εποπτευόμενο μοντέλο μας.

Ενδεικτικά τα αποτελέσματα της εν λόγω μεθόδου κατηγοριοποίησης παρουσιάζεται στην Εικόνα 10, 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 10: 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.

One Class Support Vector Machines (OCSVM):

Τα SVM (Support Vector Machines) είναι μοντέλα εποπτευόμενης μάθησης που αξιολογούν δεδομένα για να εντοπίζουν μοτίβα και μπορούν να χρησιμεύσουν τόσο σε διαδικασίες ταξινόμησης όσο και παλινδρόμησης. Ένα SVM μοντέλο λειτουργεί έτσι ώστε να χωρίζει τα δείγματα σε διακριτές κατηγορίες δημιουργώντας το μεγαλύτερο δυνατό διάκενο μεταξύ τους, προβάλλοντας ποινές στα δείγματα που βρίσκονται στη λάθος πλευρά του διακένου.

Ο μηχανισμός OCSVM (Μηχανισμός Υποστήριξης Διανυσμάτων Μίας Κλάσης) μπορεί να αναγνωρίσει "ατακτοποιήσεις" στο χώρο των χαρακτηριστικών δημιουργώντας ένα υπέρ-επίπεδο που ορίζει ιδανικά τα δεδομένα. Σε σύγκριση με άλλες τεχνικές μηχανικής μάθησης, τα SVM παρέχουν υψηλότερες δυνατότητες γενίκευσης [121]. Ο OCSVM είναι ένας αλγόριθμος μη εποπτευόμενης μάθησης που εκπαιδεύεται βασισμένος μόνο σε "κανονικά" δεδομένα, στην περίπτωση μας δεδομένα που αντιστοιχούν σε φυσιολογική λειτουργία (χωρίς δυσλειτουργίες). Καθορίζει τα όρια των σημείων αυτών και κατατάσσει τα σημεία που βρίσκονται εκτός των ορίων ως ακραίες περιπτώσεις.

Όπως αναφέρεται και στο [122], όταν ένα σημείο αποκλίνει σημαντικά από το υπέρ-επίπεδο προς τη φυσιολογική πλευρά, η κατάσταση του συστήματος που εξετάζεται φαίνεται να είναι πιο σταθερή. Για τις ανάγκες μας, δημιουργήθηκαν δύο SVM μοντέλα με δεδομένα από πλοία χωρίς γνωστές δυσλειτουργίες στα ρουλεμάν τους. Το πρώτο μοντέλο εκπαιδεύτηκε με κανονικοποιημένες ροές δεδομένων, ενώ το δεύτερο με τις ωριαίες ποσοστιαίες αλλαγές των δεδομένων. Τελικά, επιλέχθηκε το δεύτερο μοντέλο γιατί κρίθηκε πιο αποδοτικό.

Multivariate Long Short-Term Memory Model (LSTM):

Οι χρονοσειρές αποτελούν έναν σημαντικό τομέα για την ανίχνευση ανωμαλιών. Στο πλαίσιο αυτό, τα Αναδρομικά Νευρωνικά Δίκτυα (RNNs) και συγκεκριμένα τα Long Short-Term Memory (LSTM) φαίνονται να προσφέρουν μια υποσχόμενη προσέγγιση. Τα LSTM έχουν τη δυνατότητα να χειρίζονται δεδομένα πολυμεταβλητών χρονοσειρών, κατανοώντας τις πολύπλοκες μη γραμμικές σχέσεις που μπορεί να υπάρχουν μεταξύ των χαρακτηριστικών τους.

Στη μεθοδολογία που παρουσιάζεται, ένα μοντέλο LSTM εκπαιδεύεται με δεδομένα που θεωρούνται φυσιολογικά, με σκοπό να μοντελοποιήσει τη φυσιολογική λειτουργία ενός συστήματος. Εφόσον το μοντέλο είναι εκπαιδευμένο, προβλέπει τις επόμενες τιμές σε μια χρονοσειρά. Όταν οι πραγματικές τιμές αποκλίνουν από τις προβλέψεις του μοντέλου, η απόκλιση αυτή μπορεί να επιδείξει την ύπαρξη ανωμαλίας.

Αξίζει να σημειωθεί ότι αντίστοιχες προσεγγίσεις βασισμένες στα LSTM έχουν υιοθετηθεί από τη NASA, όπως φαίνεται στις πηγές [123] και [124], όπου αναφέρεται η εφαρμογή των LSTM σε δεδομένα τηλεμετρίας διαστημικών σκαφών. Επιπλέον, στην έρευνα αναφέρεται η εκπαίδευση ενός μοντέλου LSTM για την πρόβλεψη βάσει πολλαπλών ροών δεδομένων. Η εργασία δίνει έμφαση στην ανίχνευση ελαττωμάτων, ειδικά σε μηχανικά μέρη όπως τα ρουλεμάν σταυρού κεφαλής. Στο πλαίσιο αυτό, εξετάζονται διάφορα χρονικά παράθυρα και δεδομένα για τη βέλτιστη ανίχνευση των ανωμαλιών.

Σταθμισμένη Εντροπία Μετάθεσης - Weighted Permutation Entropy (WPE):

Η εντροπία μετάθεσης αποτελεί μέτρο για την πολυπλοκότητα σειριακών δεδομένων, δηλαδή υποδεικνύει τον όγκο της νέας πληροφορίας που εμφανίζεται σε κάθε σημείο μιας ακολουθίας. Στις χρονοσειρές, αυτό αντιπροσωπεύει πώς η πληροφορία εξαπλώνεται μέσα στο χρόνο. Όταν χρησιμοποιείται για την ανίχνευση ανωμαλιών, απότομες αλλαγές στην εντροπία μετάθεσης μπορεί να δείχνουν ότι κάτι ασυνήθιστο συμβαίνει. Τέτοιες τεχνικές έχουν χρησιμοποιηθεί, για παράδειγμα, για την ανίχνευση ανωμαλιών σε δεδομένα παλαιότερου κλίματος. Στην προκειμένη περίπτωση, χρησιμοποιήθηκε η σταθμισμένη εντροπία μετάθεσης για να μετρηθεί η χρονική πολυπλοκότητα της χρονοσειράς, με απότομες αλλαγές να αντιπροσωπεύουν πιθανές ανωμαλίες. Το κύριο πλεονέκτημα είναι ότι δεν χρειάζεται εκπαίδευση και μπορεί να εφαρμοστεί σε πραγματικό χρόνο.

3.5 ΧΑΙ για χρονοσειρές

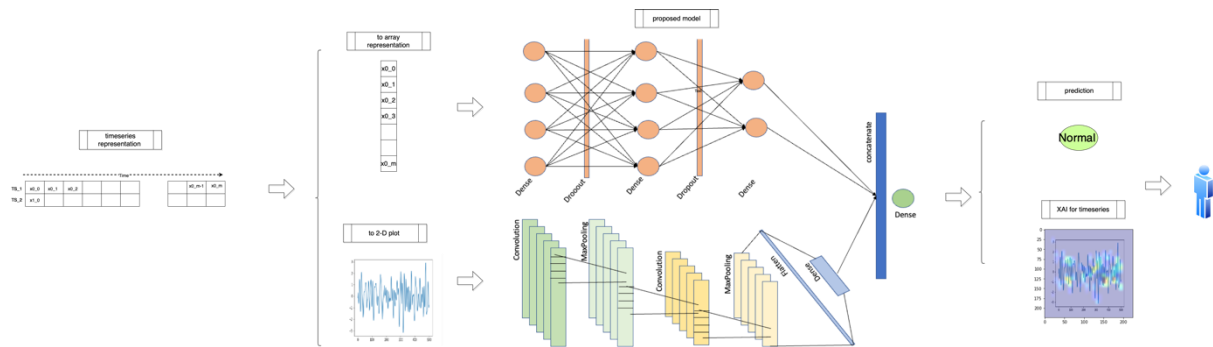
Οι πρωταρχικές ερευνητικές προκλήσεις που έχουν αντιμετωπιστεί μέχρι τώρα στο πλαίσιο της εργασίας περιλαμβάνουν τη διευκόλυνση της ακριβούς ταξινόμησης χρονοσειρών και τη δημιουργία ερμηνεύσιμων απεικονίσεων ΧΑΙ. Αυτές οι δύο όψεις είναι αλληλένδετες αφού το μοντέλο ΧΑΙ κατασκευάζεται με βάση τα αποτελέσματα της ταξινόμησης. Για να αυξήσουμε τη χρηστικότητα ενός μοντέλου DNN καθιστώντας το επεξηγήσιμο, 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. Η Εικόνα 11 και η Εικόνα 12 απεικονίζουν τη δομή του προτεινόμενου μοντέλου, με το πυκνό μοντέλο να ενσαρκώνει τον «δάσκαλο» και το CNN να συμβολίζει τον εξηγήσιμο «μαθητή».

Για να κάνουμε τη χρονοσειρά ερμηνεύσιμη, επιλέξαμε την πιο ολοκληρωμένη αναπαράσταση μιας χρονοσειράς (δηλαδή μια γραφική παράσταση 2D με τον άξονα χρόνου στον άξονα x και την τιμή της χρονοσειράς στον άξονα y). Πρώτα σχεδιάστηκε η χρονοσειρά ως εικόνα και στη συνέχεια εισάχθηκε η εικόνα στο CNN για ταξινόμηση. Σε όλα τα πειράματά μας, σχεδιάσαμε τις χρονοσειρές ως σχήματα γραμμικής γραφικής παράστασης με (224, 244, 1) διαστάσεις.

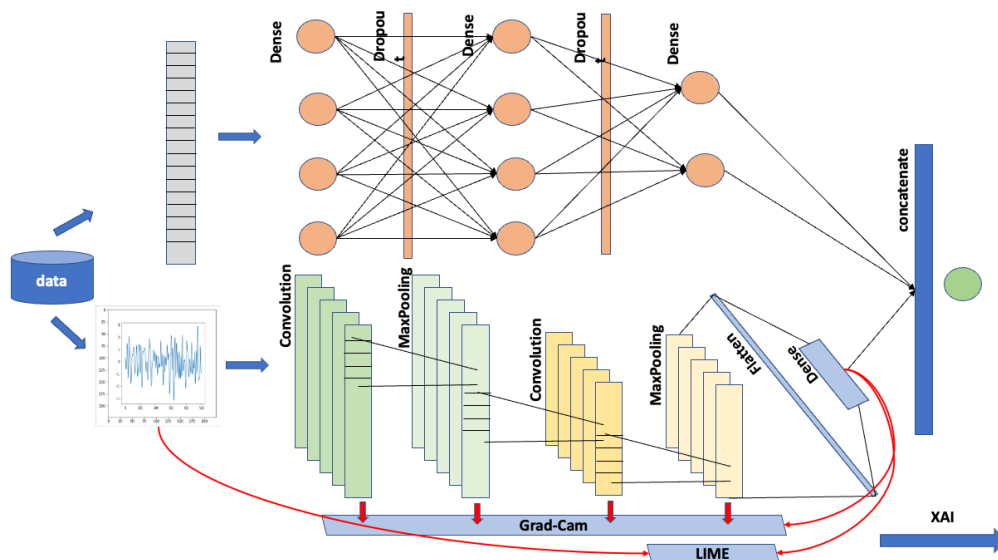
Αφού σχεδιάσουμε τις χρονοσειρές ως εικόνα, τις χρησιμοποιούμε ως είσοδο σε ένα CNN για ταξινόμηση. Το CNN θα επεξεργαστεί την εικόνα και θα χρησιμοποιήσει τα χαρακτηριστικά που έχει μάθει για να κάνει μια πρόβλεψη σχετικά με τη χρονοσειρά.

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. Στην προσέγγισή μας, επιλέξαμε 2 αξιόλογες μεθόδους XAI σε δεδομένα εικόνας ως τις καταλληλότερες για γραφικές παραστάσεις χρονοσειρών.



Εικόνα 11: Σχηματική εξήγηση της πτυχής δεδομένων στην προσέγγιση από άκρο σε άκρο, καθώς και των κύριων βημάτων στη διαδικασία, από τα πρωτογενή δεδομένα στον χρήστη.



Εικόνα 12: 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:** Το επίπεδο αυτό τροποποιεί τις διαστάσεις της εικόνας.
- **Convolution layer:** Αυτά τα στρώματα έχουν σχεδιαστεί για να αποστάζουν 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 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. Αποτελέσματα

Στην Ενότητα 4, λαμβάνει χώρα η αξιολόγηση της προτεινομένης προσέγγισης της εφαρμογής μεθόδων μηχανικής μάθησης και ΧΑΙ στη 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. Στο τέλος της ενότητας 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.

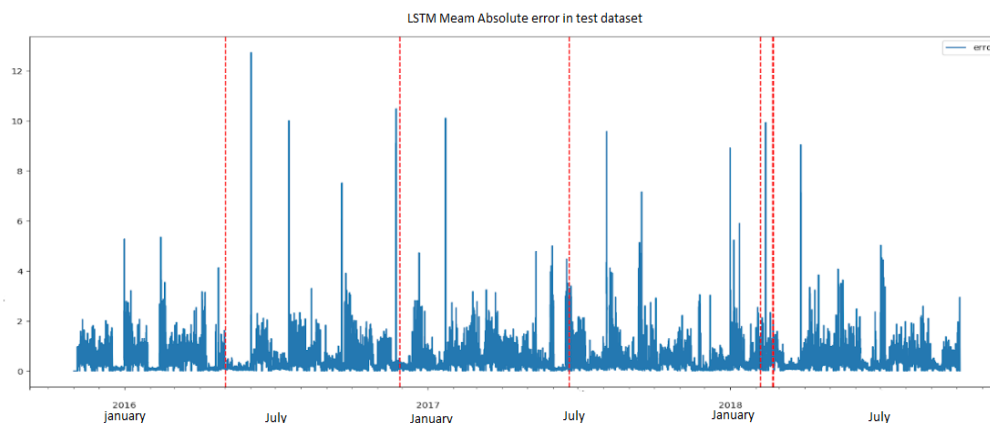
Η αξιολόγηση βλαβών και η πρόβλεψη της εξέλιξης αυτών συναντούν σημαντικές προκλήσεις, καθώς είναι δύσκολο να καθοριστεί πώς το εν λόγω ελάττωμα εξελίσσεται με την πάροδο του χρόνου. Για την εξέταση του θέματος, έχει θεσπιστεί ένα στατικό παράθυρο 30 ημερών, κατά το οποίο υποθέτουμε ότι το ελάττωμα εξελίσσεται. Με βάση αυτό το παράθυρο, αξιολογούμε την ακρίβεια των προβλέψεών μας, επικεντρωμένοι στις τιμές precision (ακρίβεια) και recall (ανάκληση). Ο πίνακας Confusion Matrix (Μήτρα Σύγχυσης) αποτελεί τον κυρίαρχο τρόπο για την ανάλυση της απόδοσης των μοντέλων.

Στην συνέχεια, το τελικό μοντέλο που παρουσιάζει αποτελεσματικότητα στο πρόβλημα αποτελείται από τον συνδυασμό διαφορετικών ταξινομητών. Παρόλο που το Gradient Boosting, ως μοντέλο, παρουσιάζει ορισμένα προβλήματα στην απόκριση, ειδικά με τα ψευδώς θετικά και τα ψευδώς αρνητικά, παραμένει χρήσιμο όταν εντάσσεται στο συνολικό σύστημα ταξινόμησης.

Στην αρχιτεκτονική του πολυμεταβλητού LSTM που αναπτύχθηκε, χρησιμοποιήθηκαν 5 εσωτερικά επίπεδα [200, 500, 700, 500, 200], με παράθυρο μνήμης 100 ώρες και υπερβολική επαπτομένη ως συνάρτηση ενεργοποίησης. Το dropout εφαρμόστηκε με ποσοστό 0.50 για την αντιμετώπιση του overfitting. Αυτό το ποσοστό βρέθηκε να είναι βέλτιστο σύμφωνα με την βιβλιογραφία [125]. Για την ορθή εκπαίδευση του μοντέλου, χρησιμοποιήθηκε "πρόωρη διακοπή" ως κριτήριο τερματισμού, ενώ δόθηκε έμφαση στην αρχικοποίηση των βαρών, λόγω του θορύβου στα δεδομένα. Προκειμένου να επιτευχθεί μια σταθερή απόδοση, δημιουργήθηκαν 10 διαφορετικά μοντέλα που χρησιμοποιούνται συλλογικά μέσω της τεχνικής Bagging, λαμβάνοντας υπόψη την πολυπλοκότητα της επιφάνειας απωλειών του μοντέλου [126].

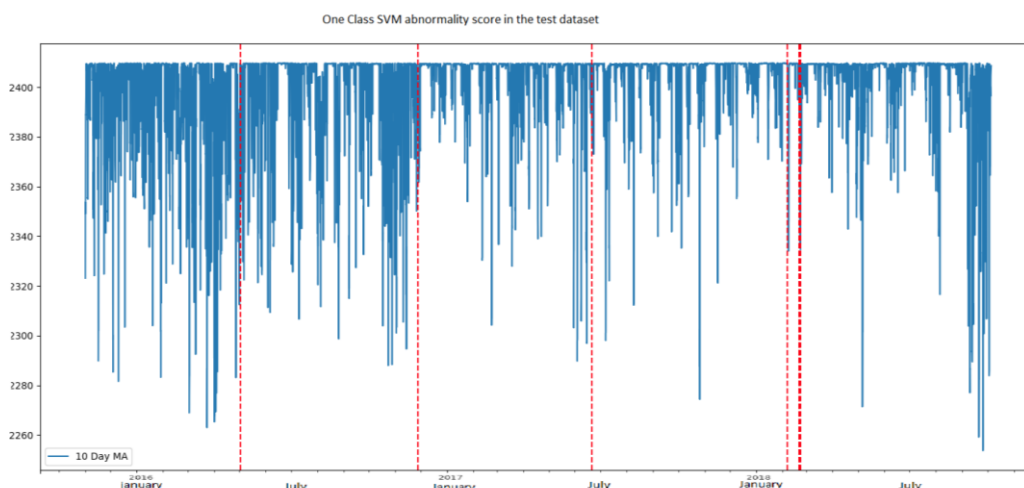
Στην Εικόνα 13 παρουσιάζεται το σφάλμα της πρόβλεψης σε σύγκριση με τις πραγματικές τιμές και συγκεκριμένα μέσο απόλυτο σφάλμα του μοντέλου LSTM. Υπάρχουν μερικές αιχμές στο διάγραμμα σφάλματος στην περιοχή πριν τον χρόνο

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.



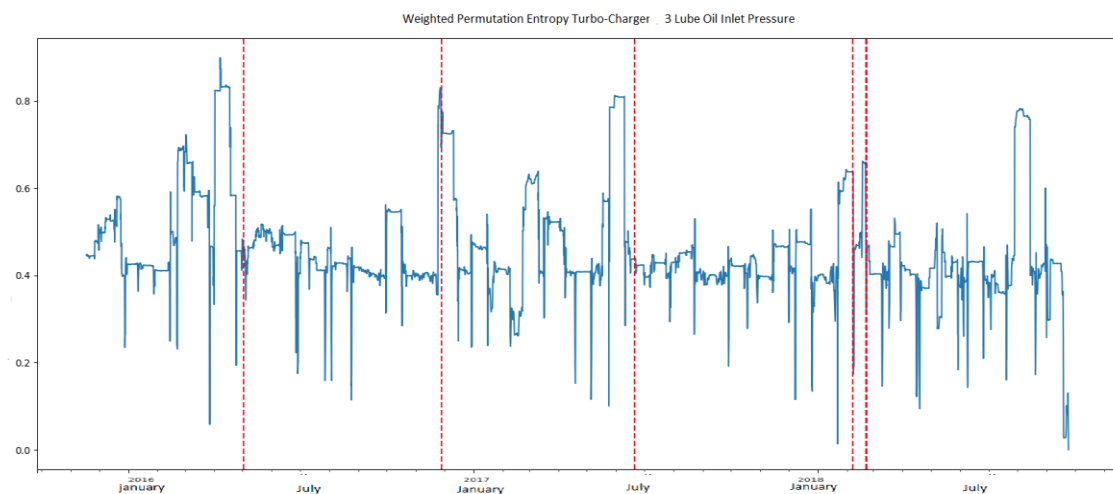
Εικόνα 13: Μέσο απόλυτο σφάλμα του μοντέλου LSTM με χρήση δεδομένων δοκιμής, Μπλε γραμμή Το MAE, Κόκκινες κάθετες γραμμές - ημερομηνίες ελαττωμάτων.

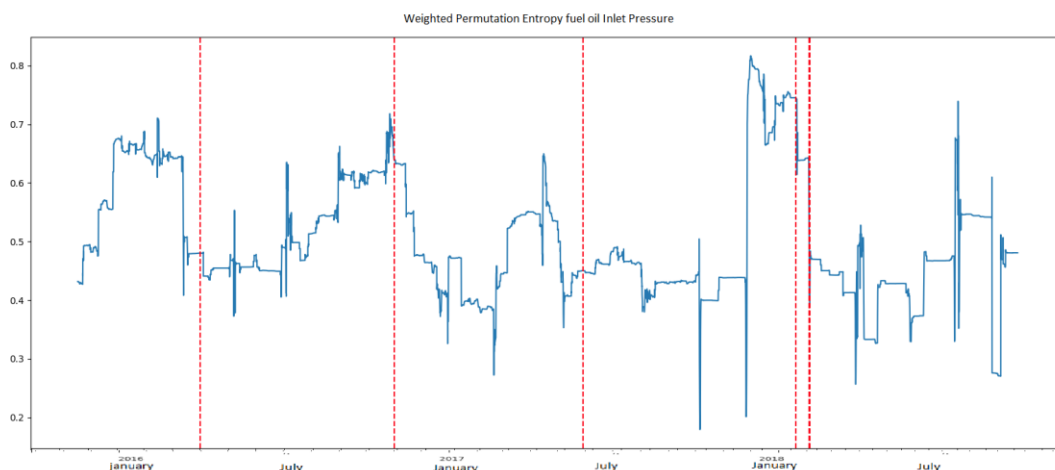
Η Εικόνα 14 παρουσιάζει την απόκλιση ανάμεσα στα σημεία δεδομένων και ένα μοντέλο SVM, με τη χαμηλή τιμή της βαθμολογίας να συμβολίζει μεγαλύτερη πιθανότητα για ανωμαλία. Προτού συμβεί κάθε ελάττωμα (αναπαριστώμενο με κόκκινες γραμμές), παρατηρούνται ημερομηνίες με σημεία που διαφέρουν σημαντικά από το "κανονικό" μοντέλο. Ο ανιχνευτής για αυτή τη μεθοδολογία χρησιμοποιεί ένα σταθερό κατώφλι για τον προσδιορισμό της ανωμαλίας.



Εικόνα 14: Η βαθμολογία ανωμαλίας που παράγεται από το μοντέλο OCSVM χρησιμοποιώντας σύνολο δεδομένων δοκιμής, Μπλε γραμμή - η βαθμολογία ανωμαλίας, Κόκκινες κάθετες γραμμές - ημερομηνίες ελαττωμάτων.

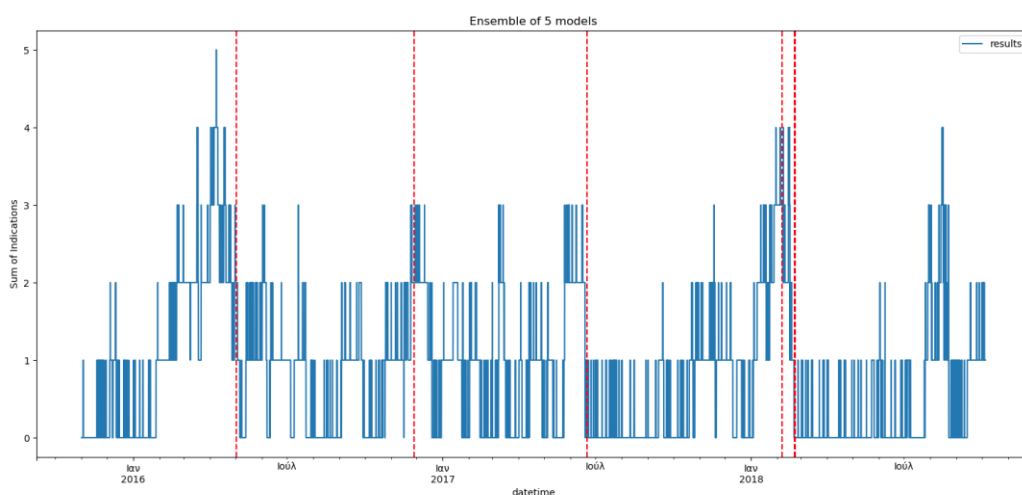
Το Weighted Permutation Entropy (WPE) χρησιμοποιείται ως ένδειξη μη ομαλής λειτουργίας, με την αύξηση της εντροπίας να σηματοδοτεί πιθανή ανωμαλία. Για τον υπολογισμό της WPE εφαρμόστηκαν παράμετροι $D=3$ και παράθυρο 600 χρονικών στιγμών, με βάση τα αποτελέσματα αναζήτησης πλέγματος. Κάποιες μετρήσεις, όπως οι πιέσεις εισόδου σε διάφορες μονάδες, κατάφεραν να ανιχνεύσουν όλα τα ελαττώματα (παρουσιάζονται στις 3 εικόνες της Εικόνα 15). Ο ανιχνευτής που χρησιμοποιείται για τη WPE λειτουργεί με σταθερό κατώφλι. Όταν η εντροπία μετάθεσης υπερβαίνει το κατώφλι, ενεργοποιείται μια ειδοποίηση για βλάβη.





Εικόνα 15: Τιμή εντροπίας κυλιόμενης σταθμισμένης μετάθεσης για (α) Πίεση εισόδου γλυκού νερού ψύξης (β) Στρόβιλο-συμπιεστής: Πίεση εισόδου λαδιού Lude (γ) Πίεση εισόδου λαδιού καυσίμου. Με ανάγνωση κάθετης γραμμής απεικονίζονται οι ημερομηνίες ελαττωμάτων

Επιπλέον, εξετάστηκε ο συνδυασμός διαφόρων ταξινομητών για να δούμε αν είναι πιο αξιόπιστος συνολικά από τα μεμονωμένα μοντέλα. Ένα συνδυαστικό μοντέλο, που βασίζεται στο άθροισμα των αποτελεσμάτων των μοντέλων, ενεργοποιεί την προειδοποίηση όταν η συνολική τιμή ξεπερνά το κατώφλι των μισών μοντέλων. Αυτό το όριο μπορεί να προσαρμοστεί βάσει των αναγκών κάθε εταιρίας. Η Εικόνα 16, ο Πίνακας 2 και ο Πίνακας 3 δείχνουν ότι η συνδυαστική προσέγγιση είναι πιο αξιόπιστη, με τα καλύτερα αποτελέσματα σε τιμές F-1 να επιτυγχάνονται με το συνδυαστικό μοντέλο, εξαιρώντας το μοντέλο LSTM.



Εικόνα 16: Αποτέλεσμα συνδυαστικού μοντέλου (μπλε χρώμα), πραγματικές βλάβες (κόκκινο χρώμα). Πιθανότητα μη φυσιολογικής κατάστασης του κύριου κινητήρα.

Πίνακας 2: Συγκεντρωτικά αποτελέσματα όλων των μοντέλων και ο συνδυασμός τους για το πρόβλημα προβλεπτικής συντήρησης κύριων μηχανών πλοίων.

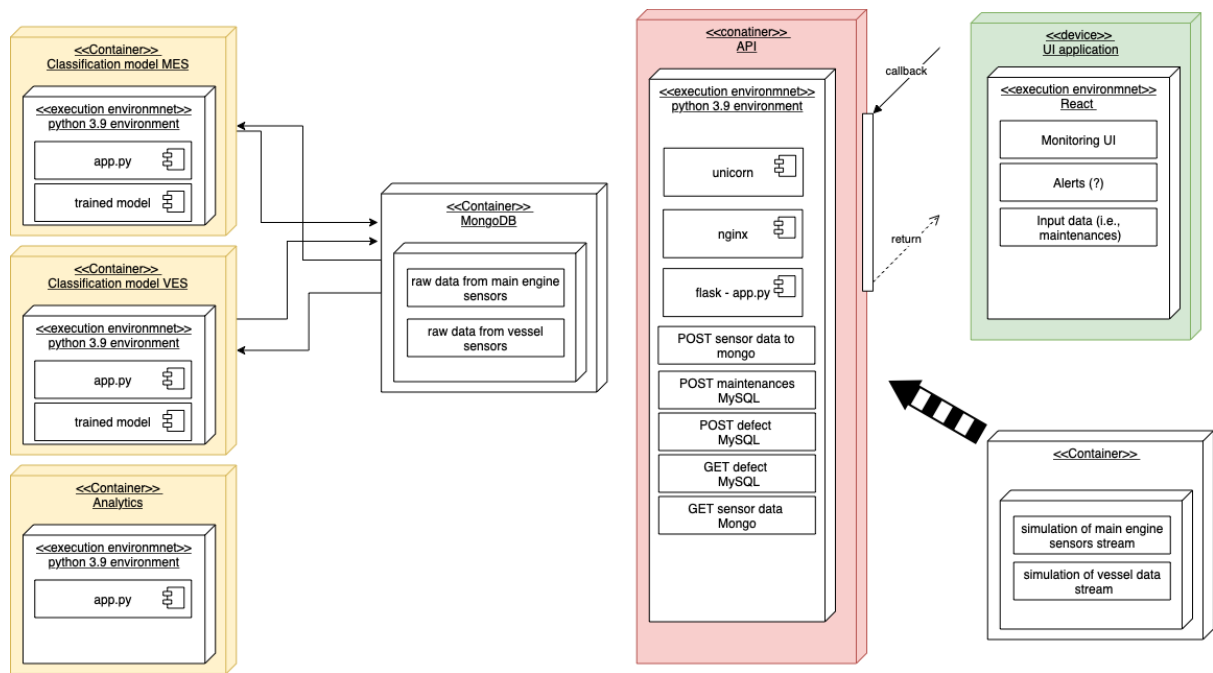
<i>Model</i>	<i>Precision</i>	<i>Recall</i>	<i>f1-score</i>
<i>Gradient Boosting</i>	0,39	0,08	0,13
<i>Multi-LSTM</i>	0,22	0,03	0,05
<i>One Class SVM</i>	0,22	0,14	0,17
<i>WPE - 1</i>	0,33	0,61	0,43
<i>WPE - 2</i>	0,47	0,53	0,5
<i>WPE - 3</i>	0,39	0,62	0,48
<i>Ensemble</i>	0,8	0,21	0,33
<i>Ensemble (no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

Πίνακας 3: Αποτελέσματα συνδυαστικού 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 Αρχιτεκτονική Εργαλείου

Πέρα των ανωτέρω ερευνητικών δραστηριοτήτων υλοποιήσαμε μία εφαρμογή με σκοπό να αναδείξουμε ένα εργαλείο προβλεπτικής συντήρησης. Το εργαλείο αφορά κατά κύριο λόγο εμπορικά-φορτηγά πλοία αλλά μπορεί να εφαρμοστεί και να υιοθετηθεί και από επιβατηγά πλοία. Εν συντομία έχει δημιουργηθεί ένα API ως back-end που βασίζεται σε microservices. Κάθε API είναι μια ξεχωριστή μικροπηρεσία που είναι υπεύθυνη για μια συγκεκριμένη λειτουργία ή σύνολο λειτουργιών και επικοινωνεί με άλλες μικροπηρεσίες όπως απαιτείται για την ολοκλήρωση των εργασιών της. Αυτή η αρχιτεκτονική επιτρέπει μεγαλύτερη ευελιξία και επεκτασιμότητα, καθώς κάθε microservice μπορεί να αναπτυχθεί, να εκτελεσθεί και να διαχειριστεί ανεξάρτητα. Αυτή η προσέγγιση επιτρέπει επίσης ταχύτερους χρόνους ανάπτυξης, καθώς οι μικροπηρεσίες μπορούν να αναπτυχθούν παράλληλα και να ενσωματωθούν σε μεταγενέστερο στάδιο. Οι υπηρεσίες ανάλυσης δεδομένων που αποσκοπούν στην πρόβλεψη λαθών έχουν σχεδιαστεί ως μικροπηρεσίες και έχουν ενσωματωθεί στην αρχιτεκτονική του εργαλείου. Πέραν του back-end προσφέρεται και ένα σχετικό front-end με φιλικό στο χρήστη περιβάλλον διεπαφής (User Interface - UI). Η αρχιτεκτονική του εργαλείου απεικονίζεται στην τοπολογία της Εικόνας 4.



Εικόνα 17: Αρχιτεκτονική εργαλείου προβλεπτικής συντήρησης – PdM tool.

Το προτεινόμενο σύστημα που επεξηγείται παρακάτω, προσφέρει μια ολοκληρωμένη διαχείριση των υπηρεσιών και εφαρμογών που αναπτύχθηκαν στα πλαίσια της εργασίας, διασφαλίζοντας τη διαλειτουργικότητα με υπάρχουσες και μελλοντικές εφαρμογές, και επιτρέποντας την εύκολη επέκταση του συστήματος μέσω της προσθήκης επιπλέον αισθητήρων (scalability). Σε αυτό το πλαίσιο, τα Docker containers μπορούν να προσφέρουν ένα πρόσθετο επίπεδο ευελιξίας και διαλειτουργικότητας.

Η υλοποίηση του PdM Tool στηρίζεται στην προσέγγιση των μικρό-υπηρεσιών όπου κάθε επιμέρους τεχνολογία προσφέρεται ως αυτόνομη εφαρμογή πακεταρισμένη σε κοντέινερ (Docker), όπως φαίνεται στο παραπάνω σχήμα. Αυτό επιτρέπει την άμεση μετάβαση των εφαρμογών από την ανάπτυξη στην παραγωγή αλλά και την εύκολη συντήρηση και αναβάθμιση τους.

Επιπλέον, μέσω της αρχιτεκτονικής αυτής, οι επιμέρους υπηρεσίες που απαιτούνται για την υλοποίηση του εργαλείου προγνωστικής συντήρησης επικοινωνούν μεταξύ τους είτε μέσω συνδέσεων στη βάση δεδομένων ασύγχρονα βάσει του πρωτοκόλλου μεταφοράς αντιπροσωπευτικής κατάστασης (REST).

Το σύστημα δομείται σε 2 επίπεδα:

- **Επίπεδο Υπηρεσιών:** Αυτό το επίπεδο περιλαμβάνει όλες τις υπηρεσίες και εφαρμογές που αναπτύχθηκαν κατά τη διάρκεια του έργου, περιλαμβάνοντας εργαλεία για προγνωστική συντήρηση, συστήματα ειδοποίησης και διαχείρισης δεδομένων, που παρουσιάζονται σε ένα ενιαίο πίνακα ελέγχου (dashboard).

- **Επίπεδο Παρουσίασης:** Αυτό το επίπεδο προσφέρει τη διεπαφή με τους χρήστες του συστήματος, παρέχοντας εύκολη πρόσβαση στα διαθέσιμα δεδομένα και υπηρεσίες μέσω μιας κεντρικής ιστοσελίδας.

Ταυτόχρονα, στο επίπεδο παρουσίασης της αρχιτεκτονικής, χρησιμοποιείται η τελευταία έκδοση του Node.js HTTP Server για την παρουσίαση των εφαρμογών στο χρήστη και την αλληλεπίδραση με αυτόν, προσφέροντας έναν αξιόπιστο και εύχρηστο μηχανισμό, συμβατό με όλα τα υποσυστήματα της αρχιτεκτονικής του συστήματος. Εδώ, η χρήση των Docker containers μπορεί να επιτρέψει μια ευέλικτη και αποτελεσματική διαχείριση των εφαρμογών και των υπηρεσιών

Για τη συνδεσιμότητα των αισθητήρων και την επικοινωνία με το κεντρικό σύστημα, στο οποίο γίνεται η συγκέντρωση και η επεξεργασία των μετρήσεων αισθητήρων, δημιουργήθηκε και χρησιμοποιείται ένα API το οποίο δύναται να δέχεται HTTP requests και με τα δεδομένα του πλοίου καθώς και από τους αισθητήρες της κύριας μηχανής. Σημειώνεται ότι το εν λόγω εργαλείο δεν αναλαμβάνει την εγκατάσταση των αισθητήριων, η εργασία αυτή παραμένει ως υποχρέωση της εταιρίας-χρήστη.

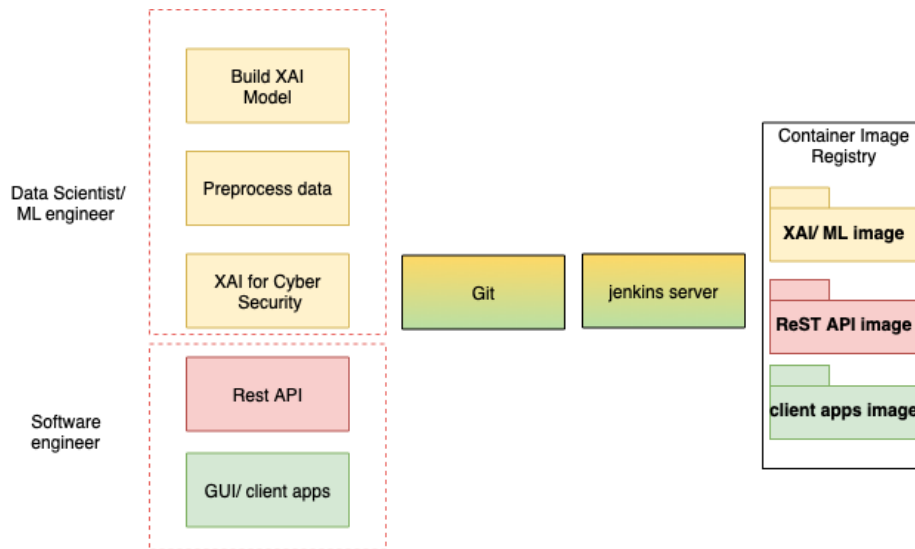
Συγκεκριμένα, όταν λαμβάνονται μετρήσεις από τους αισθητήρες του συστήματος προβλεπτικής συντήρησης, αυτά τα δεδομένα αποθηκεύονται στη βάση δεδομένων MongoDB, η οποία έχει σχεδιαστεί για ταχεία ανάκτηση των δεδομένων. Στη συνέχεια, το εργαλείο προγνωστικής συντήρησης χρησιμοποιεί αυτά τα δεδομένα για να παράγει προβλέψεις σχετικά με την κατάσταση των μηχανών πλοίων. Επιπλέον, η βάση δεδομένων αποθηκεύει επιπλέον ιστορικά δεδομένα των σταθμών που εισάγονται από το φορέα λειτουργίας του συστήματος, ώστε να είναι διαθέσιμα για ανάλυση και σύγκριση στο μέλλον. Τα δεδομένα αποθηκεύονται σε ανεξάρτητη βάση δεδομένων για να διασφαλιστεί η ακεραιότητα των δεδομένων και να αποφευχθούν συγκρούσεις και προβλήματα πρόσβασης από διαφορετικά συστήματα και υπηρεσίες.

Εκτιμητής AI RUL: Αφορά ένα μοντέλο μηχανικής μάθησης και μια στατιστική μέθοδο, όπως περιγράφεται στην προηγούμενη ενότητα.

Εκτιμητής Κατάστασης: Αφορά ένα μοντέλο μηχανικής μάθησης και μια στατιστική μέθοδος, όπως περιγράφεται στην προηγούμενη ενότητα.

Βάση δεδομένων MongoDB: Αφορά μια βάση δεδομένων MongoDB, όπου μπορούν να αποθηκευτούν γενικές πληροφορίες, όπως ανταλλακτικά ή πληροφορίες συντήρησης (δηλαδή ημερομηνία, διάρκεια, πληροφορίες τεχνικού). Σε αυτή τη βάση αποθηκεύονται και τα αποτελέσματα των προβλεπτικών μοντέλων. Επιπλέον σε αυτήν αποθηκεύονται μέσω του API τα δεδομένα από τους αισθητήρες των πλοίων.

Η αρχιτεκτονική του CI/CD απεικονίζεται στην Εικόνα 18.

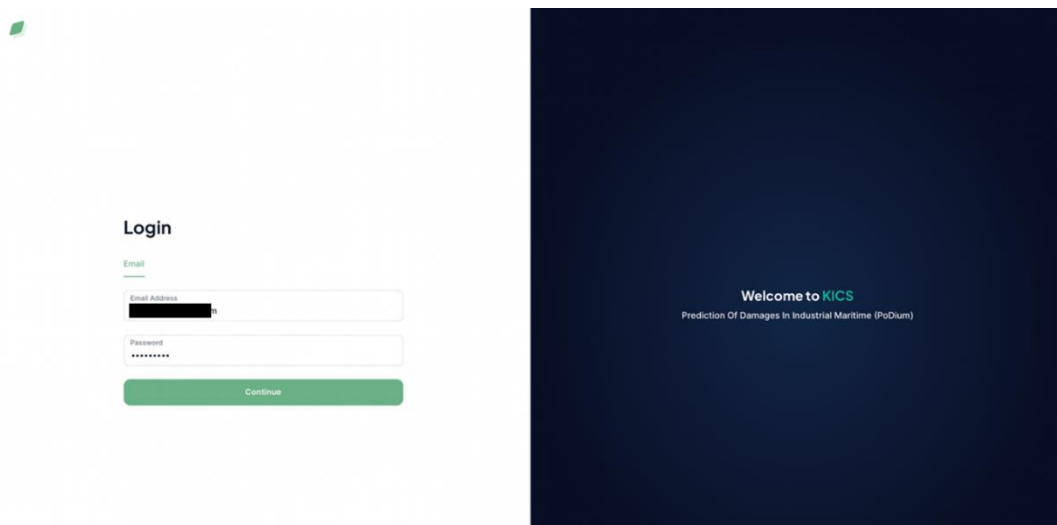


Εικόνα 18: Η CI/CD του υλοποιημένου εργαλείου.

Διεπαφή Χρήστη – User Interface: Το User Interface που αναπτύχθηκε για την οπτικοποίηση των αποτελεσμάτων της προγνωστικής συντήρησης ως πρότυπο και κάθε εταιρία δύναται να προσαρμόσει το δικό της περιβάλλον χρήστη.

Το πρότυπο σύστημα οπτικοποίησης, υπό τη μορφή πίνακα ελέγχου (dashboard), λαμβάνει όλα τα δεδομένα και τις πληροφορίες από τη βάση δεδομένων όπως αυτά έχουν επεξεργαστεί από τα προηγούμενα επίπεδα της προτεινόμενης αρχιτεκτονικής, και παράγει τις κατάλληλες απεικονίσεις προς το τελικό χρήστη, δηλαδή διαγράμματα, γραφήματα, χάρτες, line plots, πίτες, ιστογράμματα κτλ., που είναι δυναμικά και τα οποία οπτικοποιούν τα λαμβανόμενα δεδομένα.

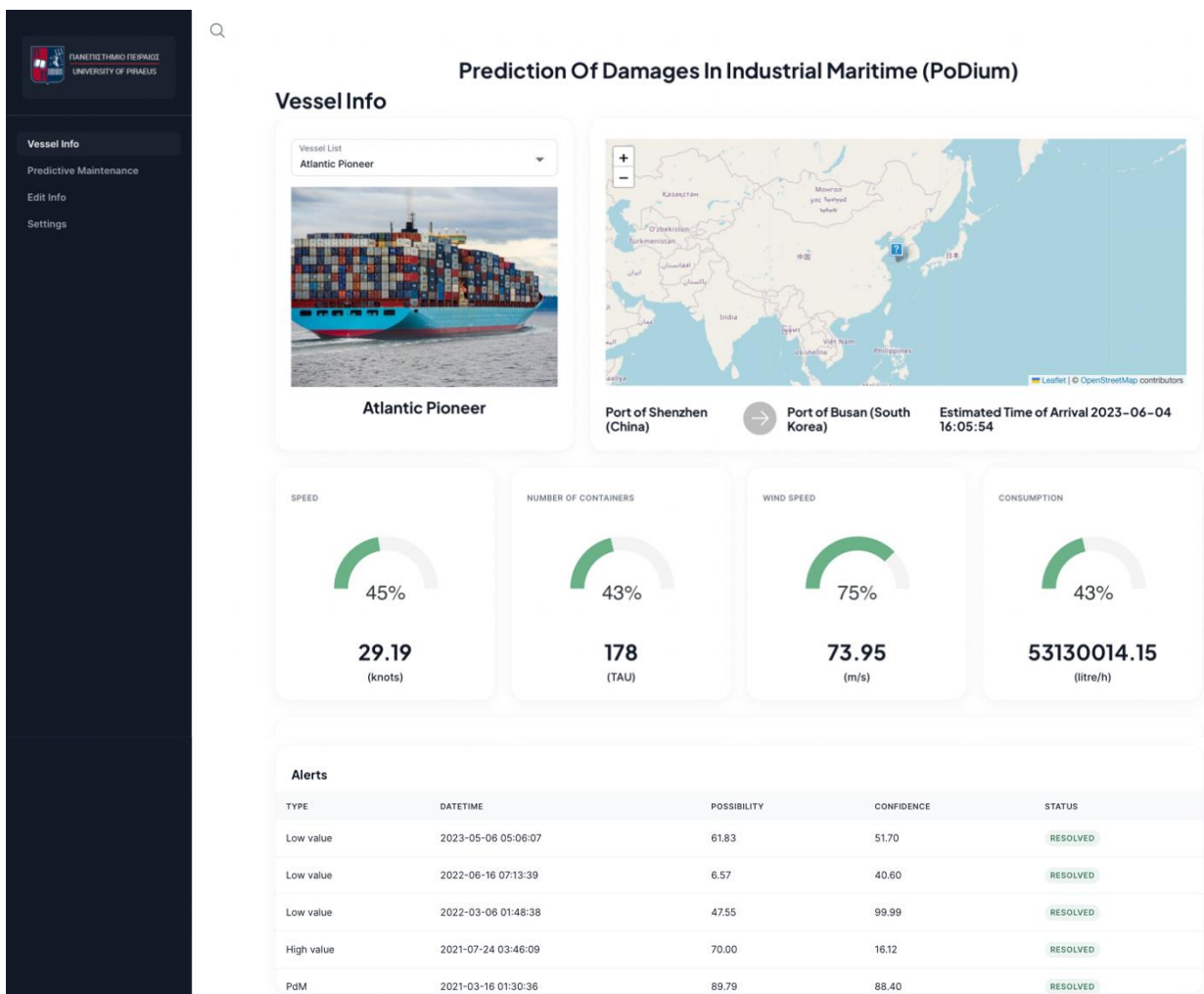
Αρχικά ο χρήστης καλείται να κάνει σύνδεση (Login) στην πλατφόρμα εισάγοντας τα «διαπιστευτήρια» του (username και password), όπως φαίνεται και στην Εικόνα 19.



Εικόνα 19: Σελίδα σύνδεσης χρηστών στη πλατφόρμα.

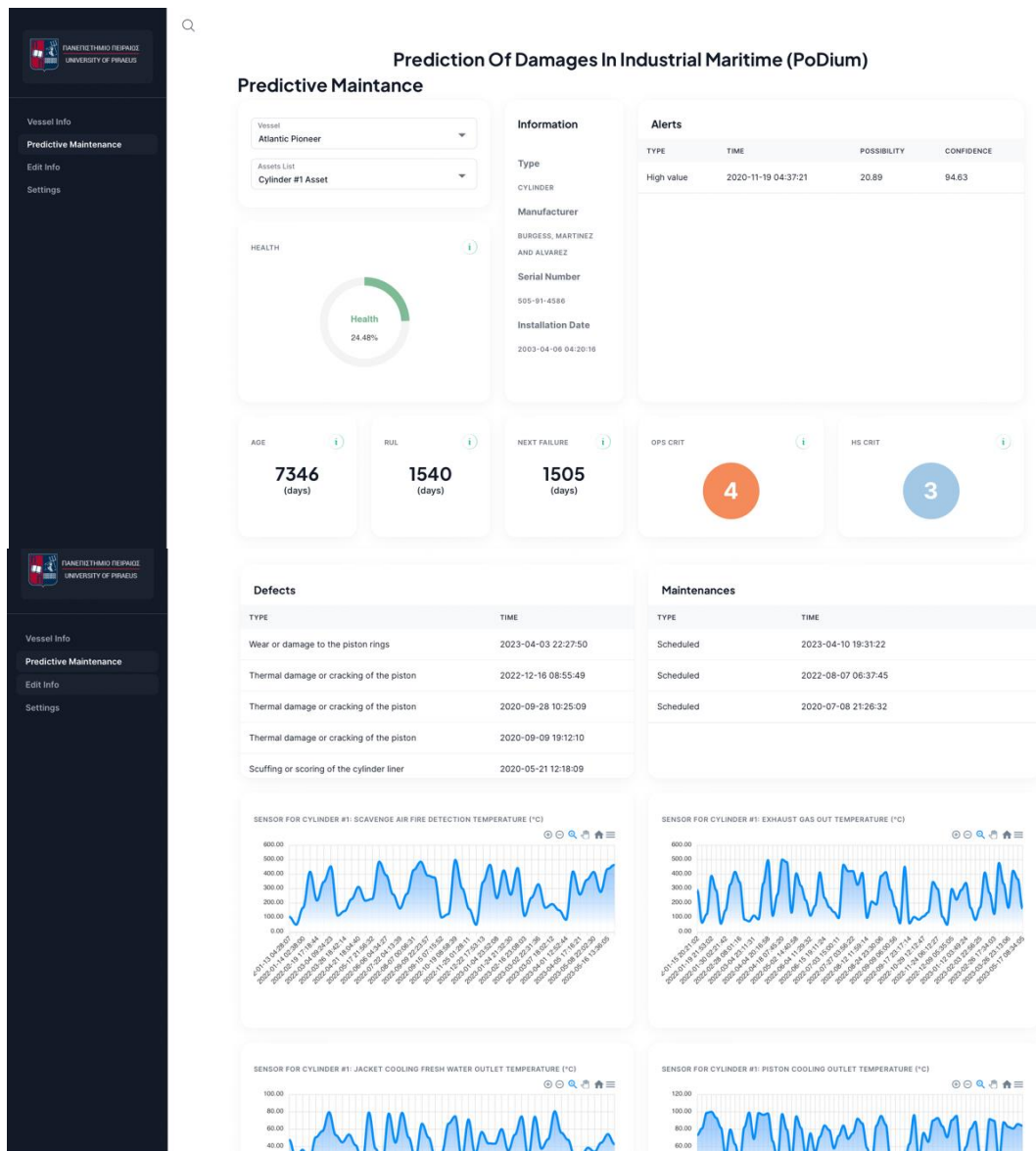
Στην κύρια οθόνη (Εικόνα 20) του πρότυπου περιβάλλοντος απεικονίζονται, από τα αριστερά προς τα δεξιά ένας χάρτης όπου αποτυπώνονται η θέση κάθε πλοίου της εταιρίας επιδεικνύοντάς την κατάσταση του πλοίου. Στη σελίδα εισόδου «μετά από τη διαδικασία σύνδεσης», βρίσκεται η «Πληροφορία Πλοίου», η οποία παρουσιάζει τις συγκεκριμένες πληροφορίες για κάθε πλοίο. Στον δυναμικό χάρτη, μπορούμε να δούμε την τελευταία γνωστή τοποθεσία του επιλεγμένου πλοίου. Κάτω από τον χάρτη, απεικονίζεται το τελευταίο λιμάνι και το λιμάνι προορισμού προς το οποίο το πλοίο κινείται. Τέλος, μπορούμε να δούμε μια εκτίμηση της ώρας άφιξης στο επόμενο λιμάνι. Στη μέση της οθόνης, μπορούμε να δούμε τα τελευταία γνωστά δεδομένα του πλοίου, τα οποία αποκαλούνται «δεδομένα πλοίου». Εδώ μπορούμε να δούμε την ταχύτητα, τον αριθμό των εμπορευματοκιβωτίων, την ταχύτητα του ανέμου και την τελευταία γνωστή κατανάλωση. Επίσης, καθώς έχουμε τις μέγιστες δυνατές τιμές, εμφανίζουμε επίσης το ποσοστό αυτών των τιμών.

Στο κάτω μέρος της πρώτης σελίδας (Εικόνα 20), μπορούμε να δούμε τις πέντε τελευταίες ειδοποιήσεις αυτού του συγκεκριμένου πλοίου. Επίσης, εκεί μπορούμε να δούμε την πιθανότητα κάθε ειδοποίησης και το επίπεδο εμπιστοσύνης των μοντέλων που δημιουργούν τις ειδοποιήσεις.



Εικόνα 20: Κύρια σελίδα πλατφόρμας για παρακολούθηση των πλοίων και της κατάστασης τους.

Στη σελίδα "Predictive maintenance" (Εικόνα 21), μπορούμε να δούμε ορισμένες πληροφορίες σχετικά με κάποιον συγκεκριμένο πόρο. Στο πεδίο "Information" υπάρχουν πληροφορίες όπως ο κατασκευαστής του πόρου, ο αριθμός σειράς του και η ημερομηνία εγκατάστασης αυτού του εξαρτήματος. Στη δεξιά πλευρά, μπορούμε να δούμε τις ειδοποιήσεις (alerts) αυτού του συγκεκριμένου πόρου, ενώ στην αριστερή πλευρά το πεδίο "health" απεικονίζει τη συνολική κατάσταση αυτού του εξαρτήματος, η οποία αποτελεί έναν βαρυσήμαντο μέσο όρο των διάφορων AI και στατιστικών δεικτών.



Εικόνα 21: Σελίδα Predictive maintenance - Απεικόνιση τιμών αισθητήρων σε πραγματικό χρόνο.

Η διεπαφή χρήστη "Edit Info" (Εικόνα 22) για την εισαγωγή δεδομένων συντηρήσεων και βλαβών είναι επίσης σημαντική για τη διαχείριση του προγράμματος συντήρησης της μηχανής του πλοίου. Η διεπαφή αυτή περιλαμβάνει drop down lists για κάθε εξάρτημα της μηχανής, έτσι ώστε ο χρήστης να μπορεί να επιλέξει το σωστό εξάρτημα για τη συντήρηση ή επιδιόρθωση.

Επιπλέον, η διεπαφή περιλαμβάνει μια ημερομηνία εισαγωγής των δεδομένων, προκειμένου να μπορεί να παρακολουθείται ο χρόνος μεταξύ των συντηρήσεων και η πρόοδος του προγράμματος συντήρησης.

The screenshot displays the 'Edit Info' page for 'Prediction Of Damages In Industrial Maritime (PoDium)'. The page is divided into two main sections: 'Maintenance' and 'Defect'.

Maintenance Section:

- Filters: Vessel List (Atlantic Pioneer), Asset List (Engine Asset), Maint List (Scheduled), Date Picker (May 2023).
- Form fields: Engineer, Description.
- Table with columns: Vessel, Asset, Type, Datetime, personnel, Description.

Vessel	Asset	Type	Datetime	personnel	Description
VSL_lj766yDEj	AST_zpNrsrvixwC_VSL_lj766y...	Routine (weekly)	2020-01-03 15:37:48	-	
VSL_xw3KctH01i	AST_0e7TCCyVrF_VSL_xw3Kct...	Routine (weekly)	2020-01-07 22:33:25	-	
VSL_lj766yDEj	AST_ex5HnMMO_VSL_lj766y...	Scheduled	2020-01-14 01:08:01	-	
VSL_C7ELkSuv8q	AST_Rav5o8yBPL_VSL_C7ELkS...	Routine (weekly)	2020-01-15 19:16:39	-	
VSL_fgWP3Udk38	AST_0GRFhUXDi6_VSL_fgWP3...	Scheduled	2020-01-27 18:49:10	-	

Defect Section:

- Filters: Vessel List, Asset List, Defect List, Date Picker.
- Form fields: Engineer, Description.
- Table with columns: Vessel, Asset, Type, Datetime, personnel, Description.

Vessel	Asset	Type	Datetime	personnel	Description
VSL_C7ELkSuv8q	AST_Rav5o8yBPL_VSL_C7ELkS...	Wear or damage to the piston rings	2023-05-31 00:00:00	Spiros	insert Defect
VSL_30fctTeg0d	AST_DW04RgAJNQ_VSL_30fct...	Worn or damaged bearings	2023-05-26 00:00:00	Petros	Test
VSL_30fctTeg0d	AST_DW04RgAJNQ_VSL_30fct...	Worn or damaged bearings	2023-05-26 00:00:00	Petros	Test
VSL_30fctTeg0d	AST_DW04RgAJNQ_VSL_30fct...	Worn or damaged bearings	2023-05-25 00:00:00	Petros	Test
VSL_DLQkD06i0	AST_7yLqRkCTD_VSL_DLQk...	Thermal damage or cracking of the piston	2023-05-17 13:08:32	-	Detected signs of 1 damage on Cylind visible cracking.

Εικόνα 22: Σελίδα Edit info - Εισαγωγή και επεξεργασία δεδομένων συντήρησης και βλαβών.

Η διεπαφή αυτή επιτρέπει επίσης στον χρήστη να καταγράφει λεπτομέρειες για τις συντηρήσεις και τις βλάβες, όπως τον τύπο της εργασίας που πραγματοποιήθηκε, τον χρόνο που πραγματοποιήθηκε και άλλες σημαντικές λεπτομέρειες. Αυτό βοηθά στην παραμετροποίηση του συστήματος για κάθε εταιρία, ενώ τα δεδομένα αυτά απαιτούνται για την εκπαίδευση των μεθόδων προβλεπτικής συντήρησης.

Συνολικά, η διεπαφή χρήστη για την εισαγωγή δεδομένων συντηρήσεων και βλαβών είναι ένα σημαντικό εργαλείο για τη διαχείριση του προγράμματος συντήρησης της μηχανής του πλοίου και για τη διασφάλιση της ασφαλούς και αποτελεσματικής λειτουργίας του πλοίου.

4.3 Αποτελέσματα ΧΑΙ για Χρονοσειρές

Σε αυτή την ενότητα, 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 μόνου του και του μαθητή μόνου του.

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 πρώιμη 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 δασκάλων, 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 μεθοδολογίες LIME και GradCAM που επισημαίνουν σημαντικά τμήματα των εισόδων 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 ΧΑΙ 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.

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 4: 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

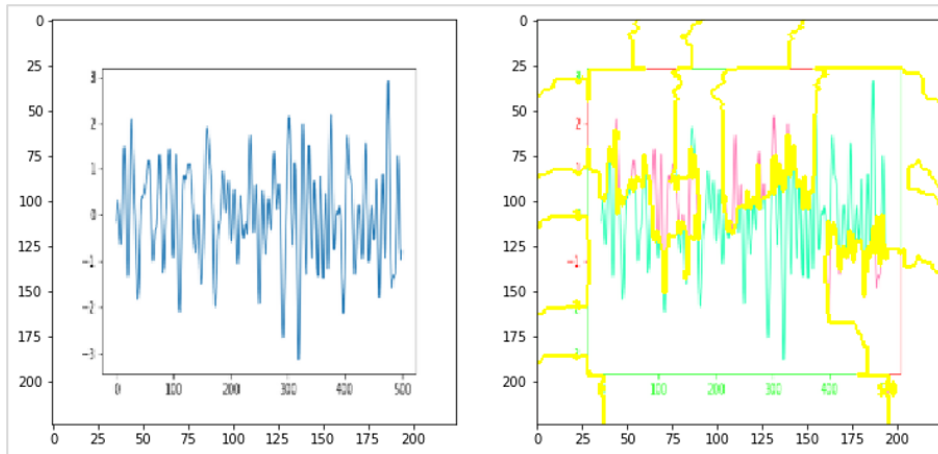
Πίνακας 5: Χρόνος εκτέλεσης κάθε μοντέλου για την εκπαιδευτική εργασία καθώς και πόσες εποχές έκανε κάθε μοντέλο να συγκλίνει. Η τελευταία στήλη «RAM» αναφέρεται στα MB που απαιτούνται για τη φόρτωση των συνόλων δεδομένων εκπαίδευσης στη μνήμη RAM.

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

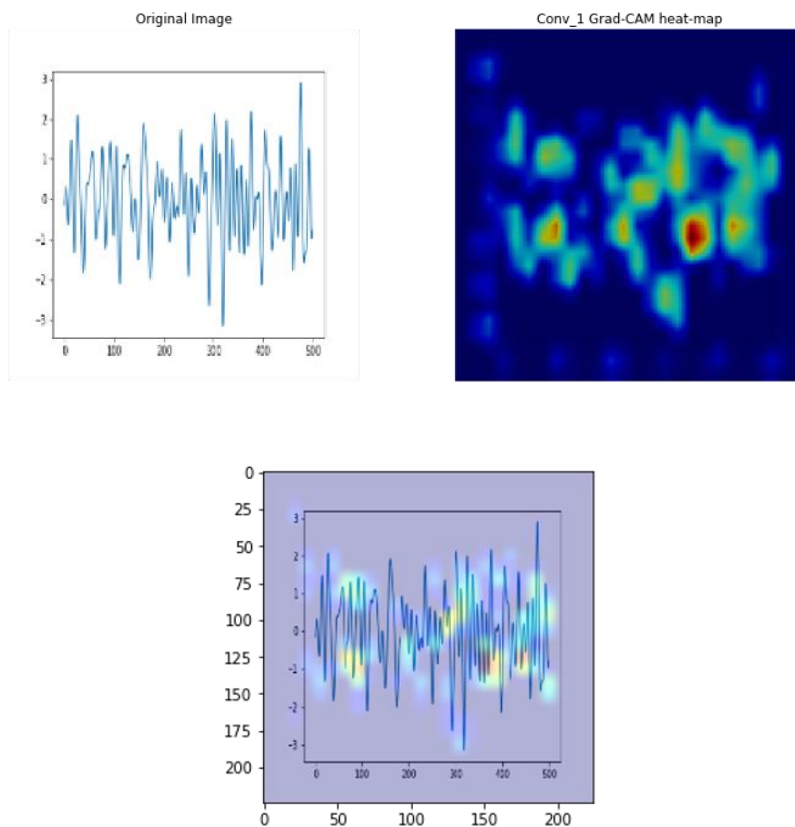
Η απόδοση του χρόνου παίζει πράγματι σημαντικό ρόλο στην αποδοτικότητα και τη σχέση κόστους-αποτελεσματικότητας των συστημάτων AI. Τα μοντέλα που χρειάζονται περισσότερο χρόνο για την εκτέλεση και την επεξεργασία δεδομένων ενδέχεται να απαιτούν περισσότερους υπολογιστικούς πόρους, αυξάνοντας το λειτουργικό κόστος. Επομένως, η βελτιστοποίηση μοντέλων τεχνητής νοημοσύνης για χρονική απόδοση είναι μια κρίσιμη πτυχή της ανάπτυξης και ανάπτυξης της τεχνητής νοημοσύνης.

Ο Πίνακας 5 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. Τονίζεται ότι οι διαμορφώσεις του μοντέλου προσδιορίστηκαν όπως αναφέρθηκε στην προηγούμενη υποενότητα. Επίσης, αξίζει να σημειωθεί ότι όσον αφορά τον χρόνο συμπερασμάτων (δηλαδή, τον χρόνο που χρειάζεται το μοντέλο για να δημιουργήσει μια πρόβλεψη για μία μόνο είσοδο), δεν υπάρχει σημαντική διαφορά μεταξύ των μοντέλων.

Όσον αφορά την πτυχή της «επεξηγησιμότητας», 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. Η Εικόνα 23 και η Εικόνα 24 παρέχουν μερικά αντιπροσωπευτικά παραδείγματα αυτών των απεικονίσεων XAI.



Εικόνα 23: Παράδειγμα σημαντικών superpixel LIME όπου μπορούμε να παρατηρήσουμε τις κόκκινες περιοχές που δεν φαίνονται σχετικές με την προβλεπόμενη κατηγορία.



Εικόνα 24: Παράδειγμα του χάρτη θερμότητας Grad-Cam και της επικάλυψης του χάρτη θερμότητας όπου μπορούμε να παρατηρήσουμε τις περιοχές που εστίασε το μοντέλο.

5. Συμπεράσματα και Μελλοντικές επεκτάσεις

Η παρούσα διπλωματική εργασία συνέβαλε στον τομέα της προβλεπτικής συντήρησης κύριων μηχανών πλοίων, προσφέροντας μια καινοτόμα μεθοδολογία που εκμεταλλεύεται τις δυνατότητες της μηχανικής μάθησης και των χρονοσειρών. Η εργασία αυτή επικεντρώθηκε στην ανάπτυξη και αξιολόγηση μοντέλων πρόβλεψης, ενισχύοντας τις διαδικασίες προεπεξεργασίας δεδομένων και εξόρυξης σημαντικών πληροφοριών. Η καινοτομία της προσέγγισης αναγνωρίζεται στην ικανότητά της να εντοπίζει ανωμαλίες, ενσωματώνοντας πολλαπλά μοντέλα μηχανικής και βαθιάς μάθησης, και να ανταποκρίνεται στις πραγματικές ανάγκες του ναυτιλιακού τομέα.

Η επιστημονική συνεισφορά της εργασίας επεκτείνεται και στην επεξήγηση των προβλέψεων για χρονοσειρές, ενσωματώνοντας τεχνικές οπτικοποίησης για την καλύτερη κατανόηση των σημαντικών τμημάτων των δεδομένων. Αυτό ανοίγει νέες δυνατότητες σε έναν τομέα με περιορισμένες υπάρχουσες λύσεις, δίνοντας έμφαση στην επεξηγησιμότητα και την αποδοτικότητα των μοντέλων.

Συνοψίζοντας, η εργασία αυτή αποτελεί έναν σημαντικό βήμα προς την προαγωγή της προβλεπτικής συντήρησης με τη χρήση προηγμένων τεχνικών μηχανικής μάθησης και ανάλυσης χρονοσειρών, ανοίγοντας τον δρόμο για μελλοντικές ερευνητικές προκλήσεις και επεκτάσεις σε αυτόν τον τομέα.

Όσον αφορά τις μελλοντικές επεκτάσεις, αυτή η εργασία ανοίγει τον δρόμο για πολλαπλές δυνατότητες στον τομέα της προβλεπτικής συντήρησης κύριων μηχανών πλοίων. Πρώτον, η βελτίωση και επέκταση των υφιστάμενων μοντέλων μηχανικής μάθησης μέσω της ενσωμάτωσης πιο προηγμένων αλγορίθμων και τεχνικών βαθιάς μάθησης μπορεί να παρέχει ακόμα πιο ακριβείς και αξιόπιστες προβλέψεις. Δεύτερον, η διερεύνηση της ενσωμάτωσης πρόσθετων τύπων δεδομένων, όπως αισθητήριες παρατηρήσεις και λειτουργικά στοιχεία, θα μπορούσε να ενισχύσει την ολοκληρωμένη εικόνα της κατάστασης των μηχανών. Τρίτον, η εφαρμογή της προτεινόμενης μεθοδολογίας σε διαφορετικά περιβάλλοντα και σενάρια λειτουργίας, καθώς και σε άλλους τύπους μηχανημάτων, θα μπορούσε να επεκτείνει το πεδίο εφαρμογής της έρευνας. Τέλος, η εξερεύνηση της ενσωμάτωσης ερμηνευτικών τεχνικών στα μοντέλα προβλέψεων θα μπορούσε να παρέχει πιο διαισθητικές και εύκολα κατανοητές εξηγήσεις για τις προβλέψεις, βελτιώνοντας την επεξήγηση και τη διαφάνεια στη λήψη αποφάσεων.

Βιβλιογραφία

- [1] J. Soldatos, S. Gusmeroli, P. Malo, and G. di Orio, “Internet of things applications in future manufacturing,” Digitizing industry-internet of things connecting the physical, digital and virtual worlds, pp. 153–183, 2016.
- [2] 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.
- [3] Demajo, L.M., Vella, V., Dingli, A., 2020. Explainable ai for interpretable credit scoring. arXiv preprint arXiv:2012.03749.
- [4] Alia, 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.
- [5] Rao, R.B., Krishnan, S., Niculescu, R.S., 2006. Data mining for improved cardiac care. ACM SIGKDD Explorations Newsletter 8, 3–10.
- [6] Branco, P., Torgo, L., Ribeiro, R.P., 2016. A survey of predictive modeling on imbalanced domains. ACM Computing Surveys (CSUR) 49, 1–50.
- [7] Russell, S. J., & Norvig, P. (2020). Artificial intelligence: A modern approach. Upper Saddle River, NJ: Pearson Education.
- [8] T. Hastie, R. Tibshirani, and J. Friedman. The Elements of Statistical Learning. Springer, 2009.
- [9] A. K. Jain, M. N. Murty, and P. J. Flynn. Data clustering: A review. ACM Computing Surveys, 31(3):264-323, 1999.
- [10] Geoffrey E. Hinton, Simon Osindero, and Yee-Whye Teh. 2006. A fast learning algorithm for deep belief nets. Neural computation, 18(7):1527–1554
- [11] 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.
- [12] 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.
- [13] 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.
- [14] Yoshua Bengio. 2009. Learning deep architectures for AI. Now Publishers Inc.

- [15] Ruslan Salakhutdinov and Geoffrey Hinton. 2009. Deep Boltzmann machines. In *Artificial intelligence and statistics*, pages 448–455. PMLR.
- [16] Hugo Larochelle, Yoshua Bengio, Jerome Louradour, and Pascal Lamblin. 2009. Exploring strategies for training deep neural networks. *Journal of machine learning research*, 10(1).
- [17] 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.
- [18] 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.
- [19] Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, 323(6088), 533-536.
- [20] Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. *The Journal of Machine Learning Research*, 15(1), 1929-1958.
- [21] Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.
- [22] Pascanu, R., Mikolov, T., & Bengio, Y. (2013). On the difficulty of training recurrent neural networks. *International conference on machine learning*, 1310-1318.
- [23] Hochreiter, Sepp, and Jürgen Schmidhuber. "Long short-term memory." *Neural computation* 9.8 (1997): 1735-1780.
- [24] Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). Learning phrase representations using RNN encoder-decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078*.
- [25] Werbos, P. J. (1990). Backpropagation through time: what it does and how to do it. *Proceedings of the IEEE*, 78(10), 1550-1560.
- [26] 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.
- [27] 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.
- [28] Alex Graves, Greg Wayne, and Ivo Danihelka. 2014. Neural Turing machines. *arXiv preprint arXiv:1410.5401*.

- [29] 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.
- [30] Ryan Kiros, Ruslan Salakhutdinov, and Richard S Zemel. 2014. Unifying visual-semantic embeddings with multimodal neural language models. arXiv preprint arXiv:1411.2539.
- [31] 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.
- [32] 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.
- [33] 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.
- [34] 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.
- [35] 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.
- [36] Grzegorz Chrupala, Akos Kádár, and Afra Alishahi. 2015. Learning language through pictures. arXiv preprint arXiv:1506.03694.
- [37] 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.
- [38] Quinlan, J.R., Induction of Decision Trees. Machine Learning, 1, 81-106, 1986.
- [39] Breiman L., Friedman J., Olshen R., and Stone C., Classification and Regression Trees. Wadsworth Int. Group, 1984.
- [40] Quinlan, J. R., C4.5: Programs for Machine Learning, Morgan Kaufmann, Los Altos, 1993.
- [41] Lior Rokach, Oded Maimon. “DECISION TREES” <http://www.ise.bgu.ac.il/faculty/liorr/hbchap9.pdf>.
- [42] Breiman, L., Random forests. Machine learning, 45(1), 5-32, 2001.
- [43] Breiman, L., Bagging Predictors. Machine Learning, 1996
- [44] Gradient boosting https://En.wikipedia.org/wiki/Gradient_boosting.

- [45] Jerome H. Friedman. "Greedy function approximation: A gradient boosting machine." *The Annals of Statistics*, 29(5) 1189-1232 October 2001.
- [46] 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.
- [47] Ke, G., Meng, Q., Finley, T., et al. (2017) LightGBM: A Highly Efficient Gradient Boosting Decision Tree. *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems*, Long Beach, 4-9 December 2017, 3146-3154.
- [48] Choo, J., Liu, S., Visual analytics for explainable deep learning. *IEEE computer graphics and applications* 38(4), 84–92 (2018).
- [49] Zahavy, T., Ben-Zrihem, N., Mannor, S., Graying the black box: Understanding DQNS. In: *International conference on machine learning*. pp. 1899–1908. PMLR (2016).
- [50] Zeiler, M.D., Fergus, R., Visualizing and understanding convolutional networks. In: *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part I* 13. pp. 818–833. Springer (2014).
- [51] Shrikumar, A., Greenside, P., Kundaje, A., Learning important features through propagating activation differences. In: *International conference on machine learning*. pp. 3145–3153. PMLR (2017).
- [52] Gou, J., Yu, B., Maybank, S.J., Tao, D., Knowledge distillation: A survey. *International Journal of Computer Vision* 129, 1789–1819 (2021).
- [53] Chiu, M.T., Xu, X., Wang, K., Hobbs, J., Hovakimyan, N., Huang, T.S., Shi, H.: The 1st agriculture-vision challenge: Methods and results. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*. pp. 48–49 (2020).
- [54] Rojat, T., Puget, R., Filliat, D., Del Ser, J., Gelin, R., Diaz-Rodriguez, N., Explainable artificial intelligence (XAI) on timeseries data: A survey. *arXiv preprint arXiv:2104.00950* (2021).
- [55] Lim, B., Zohren, S., Time-series forecasting with deep learning: a survey. *Philosophical Transactions of the Royal Society A* 379(2194), 20200209 (2021).
- [56] F.E. Tay, L. Cao, Application of support vector machines in financial time series forecasting, *Omega* 29 (2001) 309–317.
- [57] 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.

- [58] Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural computation*, 9(8): 1735–1780.
- [59] Salinas, D., Flunkert, V., Gasthaus, J., Januschowski, T., DeepAR: Probabilistic forecasting with autoregressive recurrent networks. *International Journal of Forecasting* (2019).
- [60] 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 Nagarathnam Suganthan et al. *Information Sciences* 591 (2022) 155–175.
- [61] 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 Nagarathnam Suganthan et al. *Information Sciences* 591 (2022) 155–175.
- [62] N. Tak, Forecast combination with meta possibilistic fuzzy functions, *Inf. Sci.* 560 (2021) 168–182.
- [63] J. Snoek, H. Larochelle, R.P. Adams, Practical Bayesian optimization of machine learning algorithms. *arXiv preprint arXiv:1206.2944*, 2012.
- [64] 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.
- [65] 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 pm2.5 prediction, *Sustain. Cities Soc.* 60 (2020) 102237.
- [66] 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.
- [67] 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.
- [68] D. Trigg, A. Leach, Exponential smoothing with an adaptive response rate, *J. Oper. Res. Soc.* 18 (1967) 53–59.
- [69] 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.
- [70] A. Weller, “Challenges for transparency,” in *Proc. Int. Conf. Mach. Learn. Workshop Human Interpretability ML*, 2017.
- [71] D. Gunning. *Explainable Artificial Intelligence (XAI) DARPA-BAA16-53*. Technical report, Defense Advanced Research Projects Agency (DARPA), 2016.

- [72] 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).
- [73] 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>.
- [74] 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.
- [75] 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.
- [76] B. Ustun and C. Rudin, "Supersparse linear integer models for optimized medical scoring systems," *Mach. Learn.*, vol. 102, no. 3, pp. 349–391, 2015.
- [77] Molnar, C., Casalicchio, G., Bischl, B., Interpretable Machine Learning - A Brief History, State-of-the-Art and Challenges. *arXiv* 2020, arXiv:2010.09337.
- [78] 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.
- [79] 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.
- [80] Marko Robnik-Sikonja and Igor Kononenko. Explaining classifications for individual instances. *IEEE Transactions on Knowledge and Data Engineering*, 20(5): 589–600, 2008.
- [81] Ruth C Fong and Andrea Vedaldi. Interpretable explanations of black boxes by meaningful perturbation. In *Proceedings of the IEEE International Conference on Computer Vision*, pages3429–3437, 2017.
- [82] 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.
- [83] 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>.

- [84] 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.
- [85] 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.
- [86] 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.
- [87] 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.
- [88] Lundberg, S., Lee, S.I., 2017. A unified approach to interpreting model predictions. arXiv preprint arXiv:1705.07874.
- [89] 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 hypoxemia during surgery. Nature biomedical engineering, 2018.
- [90] 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.
- [91] 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, USA, 1–9. DOI: <https://doi.org/10.1145/3287560.3287562>.
- [92] 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>.
- [93] 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.
- [94] Sokol, K.; Flach, P. LIMETree: Interactively Customisable Explanations Based on Local Surrogate Multi-output Regression Trees. arXiv 2020, arXiv:2005.01427.

- [95] Bento, J.; Saleiro, P.; Cruz, A.F.; Figueiredo, M.A.; Bizarro, P. TimeSHAP: Explaining recurrent models through sequence perturbations. arXiv 2020, arXiv:2012.00073.
- [96] 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.
- [97] 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.
- [98] 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.
- [99] Oreshkin, B. N., Carпов, D., Chapados, N., & Bengio, Y. (2019). N-BEATS: Neural basis expansion analysis for interpretable time series forecasting. arXiv preprint arXiv:1905.10437.
- [100] 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.
- [101] 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.
- [102] 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.
- [103] 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.
- [104] A. Fernandez, S. del Rio, N. V. Chawla, F. Herrera, An insight into imbalanced big data classification: outcomes and challenges, *Complex & Intelligent Systems* 3 (2) (2017) 105–120.
- [105] L. C. Lee, C.-Y. Liang, A. A. Jemail, Iterative random vs. Kennard-Stone sampling for IR spectrum-based classification task using PLS2-DA, in: *AIP Conference Proceedings*, Vol. 1940, AIP Publishing LLC, 2018, p– 020116.
- [106] 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.
- [107] V. Chandola, A. Banerjee, and V. Kumar, “Anomaly detection: A survey,” *ACM computing surveys (CSUR)*, vol. 41, no. 3, pp. 1–58, 2009.

- [108] 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.
- [109] A. Kejariwal, "Introducing practical and robust anomaly detection in a time series," Twitter Engineering Blog. Web, vol. 15, 2015.
- [110] 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.
- [111] Y. Raptodimos and I. Lazakis, "Using artificial neural network - self organising map for data clustering of marine engine condition monitoring applications," *Ships and Offshore Structures*, vol. 13, no. 6, pp. 649–656, 2018.
- [112] I. Lazakis, C. Gkerekos, and G. Theotokatos, "Investigating an SVMdriven, one-class approach to estimating ship systems condition," *Ships and Offshore Structures*, vol. 14, no. 5, pp. 432–441, 2019.
- [113] J. Kowalski, B. Krawczyk, and M. Wozniak, "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.
- [114] M. Nasution, O. Sitompul, and M. Ramli, "PCA based feature reduction to improve the accuracy of decision tree C4.5 classifications", in *Journal of Physics: Conference Series*, vol. 978, no. 1. IOP Publishing, 2018, p. 012058.
- [115] Bau, David, et al., "Network dissection: Quantifying interpretability of deep visual representations.", *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017.
- [116] Sgarbossa, Fabio, et al., "Impacts of weibull parameters estimation on preventive maintenance cost." *IFAC-PapersOnLine* 51.11 (2018): 508-513.
- [117] RELIABILITY: A Python library for reliability engineering, <https://reliability.readthedocs.io/en/latest>.
- [118] Chen, Chuang, et al., "Data-driven predictive maintenance strategy considering the uncertainty in remaining useful life prediction.", *Neurocomputing* 494 (2022): 79-88.
- [119] Zheng Liu, Norbert Meyendorf and Nezih M., "The role of data fusion in predictive maintenance using digital twin", *AIP Conference Proceedings* 1949, 020023 (2018); <https://doi.org/10.1063/1.5031520>, 20 April 2018.
- [120] 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.
- [121] 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.

- [122] 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.
- [123] 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.
- [124] 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.
- [125] P. Baldi and P. J. Sadowski, “Understanding dropout,” in Advances in neural information processing systems, 2013, pp. 2814–2822.
- [126] G. Montavon, G. Orr, and K. R. Muller, Neural networks: tricks of the trade. Springer, 2012, vol. 7700.