



ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ – ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ

Πρόγραμμα Μεταπτυχιακών Σπουδών  
«Κυβερνοασφάλεια και Επιστήμη Δεδομένων»

**Μεταπτυχιακή Διατριβή**

Τίτλος Διατριβής	<b>Εφαρμοσμένη Μηχανική Μάθηση: Μεθοδολογικό Πλαίσιο Πολυκριτήριας Επιλογής Αλγορίθμων και Σύγχρονες Εφαρμογές</b> <b>Applied Machine Learning: Methodological Framework for Multi-criteria Selection of Algorithms and Applications</b>
Όνοματεπώνυμο Φοιτητή	<b>Άγγελος Καράτσης</b>
Πατρώνυμο	<b>Αθανάσιος</b>
Αριθμός Μητρώου	<b>ΜΠΚΕΔ/21019</b>
Επιβλέπων	<b>Αποστόλου Δημήτρης, Καθηγητής</b>

Ημερομηνία Παράδοσης **Μάρτιος 2026**

**Τριμελής Εξεταστική Επιτροπή**

Δημήτριος Αποστόλου  
Καθηγητής

Κοτζανικολάου Παναγιώτης  
Καθηγητής

Γρηγόριος Κορωνάκος  
Επίκουρος Καθηγητής

## ΠΕΡΙΛΗΨΗ

Η παρούσα μεταπτυχιακή διατριβή εξετάζει σε βάθος το πεδίο της Μηχανικής Μάθησης, με πρωταρχικό σκοπό τη δημιουργία ενός δομημένου μεθοδολογικού πλαισίου για την ορθολογική επιλογή και πρακτική εφαρμογή αλγοριθμικών μοντέλων. Ξεφεύγοντας από την απλή θεωρητική βιβλιογραφική ανασκόπηση, η μελέτη υιοθετεί έναν έντονα εκπαιδευτικό χαρακτήρα. Αυτό επιτυγχάνεται αναλύοντας τη λειτουργία των βασικότερων αλγορίθμων (εποπτευόμενης και μη εποπτευόμενης μάθησης) μέσα από ρεαλιστικά σενάρια εφαρμογής.

Λαμβάνοντας υπόψη τη θεμελιώδη αρχή ότι κανένα μοντέλο δεν υπερέχει καθολικά για κάθε πιθανό πρόβλημα, ο πυρήνας της εργασίας επικεντρώνεται στον σχεδιασμό ενός συστήματος υποστήριξης αποφάσεων. Συγκεκριμένα, διαμορφώνεται ένας πίνακας πολυκριτήριας αξιολόγησης, ο οποίος τροφοδοτεί ένα δυναμικό διάγραμμα ροής (Decision Flowchart). Τα εργαλεία αυτά καθοδηγούν τον αναλυτή δεδομένων στην τεκμηριωμένη επιλογή του βέλτιστου αλγορίθμου, σταθμίζοντας μετρήσιμους παράγοντες όπως είναι η υπολογιστική ταχύτητα, η διαφάνεια των αποτελεσμάτων (ερμηνευσιμότητα) και η ανθεκτικότητα του συστήματος απέναντι σε θορυβώδη δεδομένα.

Προκειμένου να αναδειχθεί η πρακτική αξία αυτού του μεθοδολογικού πλαισίου, η διατριβή εξετάζει την εφαρμογή του σε τρεις κρίσιμους τομείς αιχμής: την ιατρική περίθαλψη, το χρηματοπιστωτικό σύστημα και τη βιομηχανία.

Συμπερασματικά, η εργασία αποδεικνύει ότι η επιλογή αλγορίθμου συνιστά έναν διαρκή συμβιβασμό μεταξύ προγνωστικής ακρίβειας, διαθέσιμων πόρων και ανάγκης για επεξηγήσιμες αποφάσεις. Κλείνοντας, προτείνονται σύγχρονες ερευνητικές κατευθύνσεις, με έμφαση στην Επεξηγήσιμη (XAI) και την Πράσινη Τεχνητή Νοημοσύνη (Green AI), την προστασία της ιδιωτικότητας μέσω της Ομόσπονδης Μάθησης (Federated Learning) και την Κβαντική Υπολογιστική.

**ΘΕΜΑΤΙΚΗ ΠΕΡΙΟΧΗ:** Επιστήμη Δεδομένων και Μηχανική Μάθηση: Ανάπτυξη Μεθοδολογικού Πλαισίου Επιλογής Αλγορίθμων

**ΛΕΞΕΙΣ ΚΛΕΙΔΙΑ:** Μηχανική Μάθηση, Επιστήμη Δεδομένων, Σύστημα Υποστήριξης Αποφάσεων, Πολυκριτήρια Ανάλυση, Ερμηνευσιμότητα Αλγορίθμων

## ABSTRACT

This master's thesis examines in depth the field of Machine Learning, with the core objective of creating a structured methodological framework for the rational selection and practical application of algorithmic models. Moving away from a simple theoretical literature review, the study adopts a strong educational character. This is achieved by analyzing the operation of the most basic algorithms (supervised and unsupervised learning) through realistic application scenarios.

Considering the fundamental principle that no model is perfect for every possible problem, the core of the work focuses on the design of a decision support system. Specifically, a multi-criteria evaluation table is formed, which feeds a Decision Flowchart. These tools guide the data analyst in the evidence-based selection of the optimal algorithm, weighing measurable factors such as computational speed, transparency of results (interpretability) and the system's resilience to noisy data.

To demonstrate the practical value of this methodological framework, the thesis examines its application in three critical cutting-edge sectors: healthcare, the financial system and industry.

In conclusion, the work demonstrates that algorithm selection constitutes a constant compromise between predictive accuracy, available resources and the need for explainable decisions.

Finally, contemporary research directions are proposed, with an emphasis on Explainable (XAI) and Green Artificial Intelligence (Green AI), privacy protection through Federated Learning and Quantum Computing.

**SUBJECT AREA:** Data Science and Machine Learning: Development of a Methodological Framework for Algorithm Selection

**KEYWORDS:** Machine Learning, Data Science, Decision Support System, Multi-Criteria Analysis, Interpretability of Algorithms

## ΚΑΤΑΛΟΓΟΣ ΣΧΗΜΑΤΩΝ

Σχήμα 1: Κατηγορίες Μηχανικής Μάθησης .....	16
Σχήμα 2: Μεθοδολογικό διάγραμμα ροής για την υποστήριξη λήψης απόφασης επιλογής αλγορίθμου.....	30

## ΚΑΤΑΛΟΓΟΣ ΕΙΚΟΝΩΝ

[Εικόνα 1](#): Δενδρόγραμμα ομαδοποίησης παρατηρήσεων από κάτω προς τα πάνω.....25

## **ΚΑΤΑΛΟΓΟΣ ΠΙΝΑΚΩΝ**

<b>Πίνακας 1:</b> Συγκριτική Ανάλυση Κατηγοριών Μηχανικής Μάθησης .....	17
<b>Πίνακας 2:</b> Συγκριτικός Πίνακας Πολυκριτήριας Αξιολόγησης .....	29

## ΠΡΟΛΟΓΟΣ

Η παρούσα διπλωματική εργασία εκπονήθηκε στο πλαίσιο του Προγράμματος Μεταπτυχιακών Σπουδών «Κυβερνοασφάλεια και Επιστήμη Δεδομένων» του Τμήματος Πληροφορικής του Πανεπιστημίου Πειραιώς. Η ενασχόλησή μου με τον συναρπαστικό χώρο της Επιστήμης Δεδομένων και ειδικότερα με τη Μηχανική Μάθηση, αποτέλεσε το έναυσμα για τη συγγραφή αυτής της μελέτης.

Βασικός μου στόχος εξ αρχής ήταν να δημιουργήσω ένα κείμενο που να μην αναλώνεται απλώς σε μια θεωρητική παράθεση αλγορίθμων, αλλά να λειτουργεί ως ένας πρακτικός και εκπαιδευτικός οδηγός. Η προσπάθεια να «μεταφραστεί» η μαθηματική πολυπλοκότητα των αλγορίθμων σε καθημερινά, κατανοητά παραδείγματα και να δομηθεί ένα μεθοδολογικό πλαίσιο επιλογής, αποτέλεσε μια απαιτητική αλλά εξαιρετικά δημιουργική διαδικασία, η οποία επισφραγίζει τον κύκλο των μεταπτυχιακών μου σπουδών.

*Άγγελος Καράτσης*

*Αθήνα, Μάρτιος 2026*

## ΠΕΡΙΕΧΟΜΕΝΑ

ΠΡΟΛΟΓΟΣ.....	8
ΚΕΦΑΛΑΙΟ 1 – ΕΙΣΑΓΩΓΗ.....	10
1.1 Στόχοι, σκοπός και ερευνητικά ερωτήματα.....	10
1.2 Σημασία της Μηχανικής Μάθησης στη σύγχρονη εποχή.....	11
1.3 Μεθοδολογία βιβλιογραφικής ανασκόπησης.....	11
1.4 Δομή της εργασίας.....	12
ΚΕΦΑΛΑΙΟ 2 – ΘΕΩΡΗΤΙΚΑ ΘΕΜΕΛΙΑ ΤΗΣ ΜΗΧΑΝΙΚΗΣ ΜΑΘΗΣΗΣ.....	13
2.1 Ορισμοί, βασικές έννοιες και ιστορική εξέλιξη.....	13
2.2 Κατηγορίες Μηχανικής Μάθησης.....	14
2.3 Συγκριτική Ανάλυση Κατηγοριών Μηχανικής Μάθησης.....	16
2.4 Διαδικασία μάθησης και αξιολόγησης.....	18
2.5 Περιορισμοί και προκλήσεις.....	19
ΚΕΦΑΛΑΙΟ 3 – ΑΝΑΛΥΣΗ ΚΑΙ ΜΕΘΟΔΟΛΟΓΙΚΟ ΠΛΑΙΣΙΟ ΕΦΑΡΜΟΓΗΣ ΑΛΓΟΡΙΘΜΩΝ ΜΗΧΑΝΙΚΗΣ ΜΑΘΗΣΗΣ.....	20
3.1 Εισαγωγή στους αλγόριθμους Μηχανικής Μάθησης.....	20
3.2 Αλγόριθμοι εποπτευόμενης μάθησης.....	20
3.2.1 Γραμμική και λογιστική παλινδρόμηση.....	20
3.2.2 k-Nearest Neighbors (k-NN).....	21
3.2.3 Δέντρα αποφάσεων και Random Forest.....	22
3.2.4 Naive Bayes.....	22
3.3 Αλγόριθμοι μη εποπτευόμενης μάθησης.....	23
3.3.1 Συσταδοποίηση K-Means (Clustering).....	23
3.3.2 Ιεραρχική Ομαδοποίηση (Hierarchical Clustering).....	24
3.3.3 DBSCAN.....	25
3.3.4 Apriori.....	26
3.3.5 Independent Component Analysis (ICA).....	27
3.4 Μεθοδολογία και Κριτήρια Επιλογής Αλγορίθμων.....	27
3.4.1 Πολυκριτήρια Αξιολόγηση Αλγορίθμων.....	28
3.4.2 Μεθοδολογικό Πλαίσιο και Διάγραμμα Ροής.....	29
ΚΕΦΑΛΑΙΟ 4 – ΣΥΓΧΡΟΝΕΣ ΕΦΑΡΜΟΓΕΣ ΤΗΣ ΜΗΧΑΝΙΚΗΣ ΜΑΘΗΣΗΣ.....	31
4.1 Εφαρμογές στον τομέα της υγείας.....	32
4.2 Εφαρμογές στη χρηματοοικονομική ανάλυση.....	33
4.3 Εφαρμογές στη βιομηχανία και τις έξυπνες τεχνολογίες.....	34
ΚΕΦΑΛΑΙΟ 5 – ΣΥΜΠΕΡΑΣΜΑΤΑ ΚΑΙ ΜΕΛΛΟΝΤΙΚΕΣ ΚΑΤΕΥΘΥΝΣΕΙΣ.....	35
5.1 Συνοπτικά συμπεράσματα.....	35
5.2 Συμβολή της εργασίας.....	35
5.3 Μελλοντικές ερευνητικές κατευθύνσεις.....	36

ΠΙΝΑΚΑΣ ΟΡΟΛΟΓΙΑΣ .....	37
ΣΥΝΤΜΗΣΕΙΣ – ΑΡΚΤΙΚΟΛΕΞΑ – ΑΚΡΩΝΥΜΙΑ .....	38
ΒΙΒΛΙΟΓΡΑΦΙΚΕΣ ΑΝΑΦΟΡΕΣ .....	40

## ΚΕΦΑΛΑΙΟ 1 – ΕΙΣΑΓΩΓΗ

### 1.1 Στόχοι, σκοπός και ερευνητικά ερωτήματα

Σε ένα περιβάλλον όπου η τεχνολογία εξελίσσεται με ραγδαίους ρυθμούς, η παραγωγή δεδομένων στη σύγχρονη κοινωνία γίνεται με όλο και περισσότερους τρόπους. Έτσι, καθίσταται απαραίτητη τόσο η πιο ακριβή και αποτελεσματική αξιοποίηση των δεδομένων αυτών, όσο και η ανάλυση νέων μεθόδων και τεχνικών που θα δώσουν απαντήσεις στις διαρκώς αυξανόμενες ανάγκες της σημερινής εποχής. Ο όγκος των πληροφοριών που παράγονται καθημερινά από πλατφόρμες δικτύωσης, χρηματοοικονομικές συναλλαγές, αισθητήρες, εφαρμογές πλοήγησης κτλ., δημιουργεί την ανάγκη να εξελιχθεί η επεξεργασία τους, εφόσον οι παραδοσιακές προσεγγίσεις φτάνουν σε περιορισμούς και δεν είναι ικανές να τις διαχειριστούν με τον πιο βέλτιστο τρόπο [1].

Απάντηση στο παραπάνω, έρχεται να δώσει η Μηχανική Μάθηση (Machine Learning). Η δυνατότητα να αξιοποιεί αλγόριθμους που μαθαίνουν από τα δεδομένα [2], η Μηχανική Μάθηση αναπτύσσει μοντέλα ικανά να αναγνωρίζουν μοτίβα, να λαμβάνουν λογικές αποφάσεις αλλά και να επεξεργάζονται λανθασμένα αποτελέσματα [3]. Η Μηχανική Μάθηση θεωρείται ένας από τους σημαντικότερους και ελπιδοφόρους τομείς της επιστήμης των δεδομένων, ώστε να εξάγει χρήσιμη γνώση ακόμα και σε πολύπλοκα περιβάλλοντα πληροφοριών [4].

Ο σκοπός της παρούσας διπλωματικής εργασίας δεν περιορίζεται σε μια απλή θεωρητική ανασκόπηση, αλλά υιοθετεί μια ξεκάθαρα εκπαιδευτική και πρακτική προσέγγιση. Πρωτίστως, επιδιώκεται η κατανόηση των θεμελιωδών αρχών και των βασικών αλγορίθμων της Μηχανικής Μάθησης, μέσα από ρεαλιστικά σενάρια εφαρμογής. Δευτερευόντως, με δεδομένο ότι κανένα μοντέλο δεν είναι ιδανικό για κάθε πρόβλημα, κύριος στόχος της μελέτης είναι η ανάπτυξη ενός δομημένου μεθοδολογικού πλαισίου (Διάγραμμα Ροής), το οποίο θα λειτουργεί ως οδηγός για την επιλογή του κατάλληλου αλγορίθμου βάσει πολυκριτήριας αξιολόγησης.

Επομένως, οι βασικοί στόχοι της εργασίας διαμορφώνονται ως εξής:

- Η παρουσίαση των θεωρητικών εννοιών της Μηχανικής Μάθησης και της λειτουργίας των αλγορίθμων μέσα από σύγχρονα εκπαιδευτικά παραδείγματα.
- Η κατασκευή ενός πίνακα πολυκριτήριας αξιολόγησης των μοντέλων, με γνώμονα την ερμηνευσιμότητα, την υπολογιστική ισχύ και την ανθεκτικότητα σε ακραίες τιμές.
- Ο σχεδιασμός ενός συστήματος υποστήριξης αποφάσεων (Διάγραμμα Ροής) που θα καθοδηγεί τον αναλυτή στην επιλογή του βέλτιστου αλγορίθμου.
- Η ανάλυση της πρακτικής εφαρμογής των αλγορίθμων μηχανικής μάθησης σε κρίσιμους τομείς, όπως η υγεία, τα χρηματοοικονομικά και η βιομηχανία.
- Η κριτική αποτίμηση της σχετικής βιβλιογραφίας και η ανάδειξη ερευνητικών κενών.

Έχοντας θέσει τους παραπάνω στόχους, τα κύρια ερευνητικά ερωτήματα διαμορφώνονται ως εξής:

- Ποιες θεμελιώδεις αρχές διέπουν τη Μηχανική Μάθηση και πώς λειτουργούν στην πράξη οι δημοφιλέστεροι αλγόριθμοι;
- Με ποια αντικειμενικά κριτήρια και με ποια μεθοδολογικά βήματα μπορεί ένας χρήστης να επιλέξει τον ιδανικό αλγόριθμο για τα δεδομένα του;
- Πώς η εφαρμογή αυτών των αλγορίθμων μεταβάλλει τη λήψη αποφάσεων στη σύγχρονη κοινωνία και βιομηχανία;

## 1.2 Σημασία της Μηχανικής Μάθησης στη σύγχρονη εποχή

Η Μηχανική Μάθηση έχει πλέον ενσωματωθεί σε πλήθος εφαρμογών της καθημερινής μας ζωής. Από τις αυτόματες προτάσεις για ταινίες, τη χρηματοοικονομική πρόβλεψη μέχρι προτάσεις για την παραγγελία προϊόντων και φαγητού που θα ταίριαζε περισσότερο με τις προτιμήσεις μας, την αναγνώριση φιλικών προσώπων στις φωτογραφίες μας, έως τη διάγνωση σοβαρών ασθενειών. Όλες αυτές οι δυνατότητες εμπεριέχουν αλγόριθμους μηχανικής μάθησης [5].

Ο κύριος λόγος που η Μηχανική μάθηση θεωρείται αδιαμφισβήτητο εργαλείο στη λήψη αποφάσεων, είναι η δυνατότητα των αλγορίθμων να μαθαίνουν και να ανταπεξέρχονται χρησιμοποιώντας δεδομένα που βελτιώνουν την απόδοσή τους χωρίς ρητό προγραμματισμό. Ανεξάρτητα από την πολυπλοκότητα και την υψηλή αβεβαιότητα που μπορεί να κυριαρχεί.

Οι καθημερινές και προφανείς εφαρμογές δεν είναι οι μοναδικές προκλήσεις που μπορεί να αντιμετωπίσει η Μηχανική Μάθηση. Η δυνατότητα να επεξεργάζεται γρήγορα και αξιόπιστα τεράστιες βάσεις δεδομένων μπορεί να δώσει καινοτόμες απαντήσεις και συσχετίσεις σε σημεία που ο ανθρώπινος παράγοντας θα αδυνατούσε ή θα ήταν εξαιρετικά χρονοβόρο να αναγνωρίσει την ελάχιστη συνοχή μεταξύ των δεδομένων αυτών [6].

Είναι σαφές πως τόσο ο ιδιωτικός όσο και ο δημόσιος τομέας επωφελείται από τη χρήση της Μηχανικής Μάθησης, διότι συμβάλλει στην ποιότητα των υπηρεσιών και την βελτίωση της αποδοτικότητας. Η αυτοματοποίηση που προσφέρει έχει ως αντίκτυπο, ένας οργανισμός να μειώσει το κόστος, να ελαχιστοποιήσει τα σφάλματα και να βελτίωση οποιαδήποτε διαδικασία που δεν αποδίδει τα μέγιστα αποτελέσματα [6].

Ωστόσο, όσο εντείνεται η χρήση των μεθόδων Μηχανικής Μάθησης, τόσο έρχονται στο προσκήνιο σημαντικές προκλήσεις και πιθανοί κίνδυνοι, όπως μεταξύ άλλων η αξιοπιστία των δεδομένων. Επομένως, η χρήση τους θα πρέπει να γίνεται με υπευθυνότητα και συνέπεια για να μην τίθενται ζητήματα αναξιόπιστων αποτελεσμάτων και να αποφεύγονται θέματα διαφάνειας των αλγορίθμων και πάνω από όλα να προστατεύονται τα ευαίσθητα προσωπικά δεδομένα. Για την αποφυγή των παραπάνω, απαιτείται ουσιαστική κατανόηση της μηχανικής μάθησης. Δεν αρκεί να αντιμετωπίζεται μόνο ως τεχνική αναγκαιότητα, αλλά πρέπει να διασφαλίζεται ότι η ανάπτυξη και η εφαρμογή της συμμορφώνονται με τα ισχύοντα νομικά και κοινωνικά πρότυπα [7].

## 1.3 Μεθοδολογία βιβλιογραφικής ανασκόπησης

Για τις ανάγκες της παρούσας βιβλιογραφικής ανασκόπησης πραγματοποιήθηκε εκτενής αναζήτηση και μελέτη της σχετικής βιβλιογραφίας, η οποία περιλάμβανε επιστημονικές δημοσιεύσεις πρωτογενών και δευτερογενών πηγών και βιβλία που άπτονται του πεδίου της επιστήμης των δεδομένων. Επιπλέον, αξιοποιήθηκε υλικό από αξιόπιστες διαδικτυακές πηγές, προερχόμενο από επαγγελματίες με αποδεδειγμένη εμπειρία στον συγκεκριμένο τομέα, με στόχο την ολοκληρωμένη κατανόηση του αντικειμένου.

Ιδιαίτερη έμφαση δόθηκε σε πηγές που παρουσιάζουν θεμελιώδεις έννοιες, καθώς και σε πιο πρόσφατες μελέτες που αποτυπώνουν τις σύγχρονες εξελίξεις στον συγκεκριμένο τομέα.

Κατά τη διαδικασία επιλογής των πηγών λήφθηκαν υπόψη κριτήρια όπως η επιστημονική εγκυρότητα, η συνάφεια με το αντικείμενο της εργασίας και η χρονολογία δημοσίευσης. Οι πηγές που επιλέχθηκαν μελετήθηκαν και συγκρίθηκαν, ώστε να εντοπιστούν κοινά σημεία, διαφοροποιήσεις και βασικές τάσεις στη βιβλιογραφία.

Η ανάλυση της βιβλιογραφίας πραγματοποιήθηκε με περιγραφικό, κριτικό και συγκριτικό τρόπο, αποτελώντας τη βάση για τη δημιουργία του μεθοδολογικού πλαισίου που προτείνεται στην εργασία. Αντί για την υλοποίηση απομονωμένων πειραματικών διαδικασιών σε ένα μόνο σύνολο δεδομένων, η προσέγγιση αυτή κρίθηκε καταλληλότερη, καθώς επέτρεψε την εξαγωγή αντικειμενικών κριτηρίων αξιολόγησης μέσα από πληθώρα μελετών. Με αυτόν τον τρόπο, η θεωρητική γνώση μετασχηματίστηκε σε ένα πρακτικά εφαρμόσιμο διάγραμμα επιλογής, προσδίδοντας άμεση χρηστικότητα στη μελέτη.

Τέλος, η βιβλιογραφική ανασκόπηση οργανώθηκε με τέτοιο τρόπο ώστε να προσφέρει μια συνολική και σφαιρική εικόνα του πεδίου, αναδεικνύοντας τόσο τα πλεονεκτήματα όσο και τους περιορισμούς των διαφορετικών προσεγγίσεων που έχουν προταθεί στη βιβλιογραφία.

## 1.4 Δομή της εργασίας

Η παρούσα εργασία οργανώνεται σε πέντε κεφάλαια, τα οποία ακολουθούν μια λογική και σταδιακή ανάπτυξη του αντικειμένου.

Στο πρώτο κεφάλαιο παρουσιάζεται το γενικό πλαίσιο της μελέτης, αναλύεται η σημασία της Μηχανικής Μάθησης στη σύγχρονη εποχή και διατυπώνονται οι βασικοί στόχοι της εργασίας. Παράλληλα, περιγράφεται η μεθοδολογία που ακολουθήθηκε για την εκπόνηση της βιβλιογραφικής ανασκόπησης.

Το δεύτερο κεφάλαιο εστιάζει στα θεωρητικά θεμέλια της Μηχανικής Μάθησης, παρουσιάζοντας τις βασικές έννοιες, κατηγορίες και αρχές που διέπουν το συγκεκριμένο πεδίο.

Στο τρίτο κεφάλαιο γίνεται αναλυτική παρουσίαση των βασικών αλγορίθμων Μηχανικής Μάθησης. Υιοθετώντας μια εκπαιδευτική προσέγγιση, η θεωρία συνοδεύεται από ρεαλιστικά παραδείγματα εφαρμογής. Στο ίδιο κεφάλαιο εισάγεται και ο βασικός σκοπός της μελέτης: η δημιουργία ενός πίνακα πολυκριτήριας αξιολόγησης και η ανάπτυξη ενός Διαγράμματος Ροής (Decision Flowchart) για την επιλογή του βέλτιστου αλγορίθμου.

Στο τέταρτο κεφάλαιο εξετάζονται οι σύγχρονες εφαρμογές της Μηχανικής Μάθησης σε κρίσιμους τομείς αιχμής, όπως η υγεία, τα χρηματοοικονομικά και η έξυπνη Βιομηχανία. Μέσα από αυτά τα πεδία, επαληθεύεται στην πράξη η χρησιμότητα του μεθοδολογικού πλαισίου που αναπτύχθηκε, αναδεικνύοντας τον καταλυτικό ρόλο των αλγορίθμων στη λήψη στρατηγικών αποφάσεων.

Τέλος, το πέμπτο κεφάλαιο συνοψίζει τα βασικά συμπεράσματα της εργασίας και προτείνει ενδεικτικές κατευθύνσεις για μελλοντική έρευνα στον τομέα της Μηχανικής Μάθησης με στόχο την ανάδειξη τάσεων και ερευνητικών κενών.

## ΚΕΦΑΛΑΙΟ 2 – ΘΕΩΡΗΤΙΚΑ ΘΕΜΕΛΙΑ ΤΗΣ ΜΗΧΑΝΙΚΗΣ ΜΑΘΗΣΗΣ

### 2.1 Ορισμοί, βασικές έννοιες και ιστορική εξέλιξη

Η Μηχανική Μάθηση αποτελεί έναν από τους σημαντικότερους κλάδους της Τεχνητής Νοημοσύνης και συνδέεται στενά με την επιστήμη και την εξόρυξη δεδομένων. Στόχος της είναι η ανάπτυξη αλγορίθμων που επιτρέπουν στα υπολογιστικά συστήματα να μαθαίνουν από δεδομένα και να βελτιώνουν σταδιακά την απόδοσή τους. Μέσα από την ανάλυση μοτίβων σε δομημένα και μη δομημένα δεδομένα, τα συστήματα μπορούν να επιλύουν πολύπλοκα προβλήματα, να παράγουν νέο περιεχόμενο και να οργανώνουν πληροφορίες σε συστάδες με βάση την ομοιότητά τους, επιδιώκοντας με αυτόν τον τρόπο την εξαγωγή χρήσιμης γνώσης [14].

Σύμφωνα με την βιβλιογραφία και τον κλασικό ορισμό του, ένα σύστημα θεωρείται ότι «μαθαίνει» όταν η επίδοσή του σε ένα συγκεκριμένο μοντέλο βελτιώνεται μέσω της εμπειρίας. Η ουσία της Μηχανικής Μάθησης βρίσκεται στην αυτόματη αναγνώριση προτύπων και σχέσεων στα δεδομένα, χωρίς την ανάγκη ρητού προγραμματισμού για κάθε πιθανή μελλοντική περίπτωση [2].

Στη διαδικασία αυτή, κομβικό ρόλο παίζουν:

- Τα δεδομένα & τα χαρακτηριστικά (features): Οι ιδιότητες που περιγράφουν την πληροφορία και λειτουργούν ως δεδομένα εισόδου [3].
- Το μοντέλο: Η μαθηματική αναπαράσταση της γνώσης που προκύπτει από τα δεδομένα.
- Ο αλγόριθμος μάθησης: Ο μηχανισμός που προσαρμόζει το μοντέλο ώστε να μειώνεται το σφάλμα πρόβλεψης.
- Η εκπαίδευση: Η διαδικασία όπου το μοντέλο μαθαίνει να γενικεύει (generalization), δηλαδή να αποδίδει σωστά σε νέα και άγνωστα δεδομένα [15].

Ανάλογα με το αν τα δεδομένα συνοδεύονται από ετικέτες ή όχι, η Μηχανική Μάθηση διακρίνεται σε εποπτευόμενη, μη εποπτευόμενη, ημι-εποπτευόμενη και ενισχυτική μάθηση ως βασικές κατηγορίες, οι οποίες θα αναλυθούν διεξοδικά στην ενότητα 2.2. [16].

Η ιστορική εξέλιξη του πεδίου της Μηχανικής Μάθησης ξεκινά από τα μέσα του 20ού αιώνα. Το 1950, ο Alan Turing εισήγαγε το γνωστό Τεστ Turing, θέτοντας το θεμελιώδες ερώτημα σχετικά με το αν μια μηχανή μπορεί να επιδείξει ανθρώπινη ευφυΐα [17]. Λίγα χρόνια αργότερα, ο Rosenblatt παρουσίασε το «Perceptron», ένα από τα πρώτα μοντέλα που μπορούσαν να μάθουν από παραδείγματα [18]. Στη συνέχεια, η έρευνα στράφηκε στην ανάπτυξη πιο σύνθετων αλγορίθμων και στη σύνδεση της μάθησης με στατιστικές μεθόδους.

Καθοριστική στιγμή αποτέλεσε η εισαγωγή της μεθόδου back-propagation του σφάλματος, η οποία επέτρεψε την αποτελεσματική εκπαίδευση πολύ-επίπεδων νευρωνικών δικτύων [19]. Κατά τη δεκαετία του 1990, η θεωρητική βάση ενισχύθηκε με τη χρήση στατιστικών μεθόδων και αλγορίθμων όπως τα δέντρα αποφάσεων και οι Support Vector Machines [16].

Από το 2010 και μετά, η διαθεσιμότητα μεγάλων ποσοτήτων δεδομένων (Big Data) και η αυξημένη υπολογιστική ισχύς οδήγησαν στην επανάσταση της Βαθιάς Μάθησης (Deep Learning). Τα βαθιά νευρωνικά δίκτυα έφεραν τεράστιες προόδους στην αναγνώριση εικόνας και λόγου, καθώς μπορούν να αναγνωρίζουν αυτόματα χαρακτηριστικά από πρωτογενείς πηγές (όπως pixels), καταργώντας την ανάγκη για χειροκίνητο σχεδιασμό χαρακτηριστικών [20] [21].

Παρόλο που η Μηχανική Μάθηση μοιάζει με την εξόρυξη δεδομένων, διαφέρουν στον στόχο τους: η εξόρυξη επικεντρώνεται στην ανάλυση υπάρχουσων βάσεων δεδομένων για την εύρεση μοτίβων, ενώ η Μηχανική Μάθηση

στοχεύει στην αυτοματοποιημένη παραγωγή γνώσης και προβλέψεων μέσω διαδικασιών μάθησης. Παρά τις διαφορές τους, τα όρια μεταξύ των δύο πεδίων παραμένουν συχνά δυσδιάκριτα [22].

Συνοψίζοντας, η Μηχανική Μάθηση δεν αποτελεί απλώς μια στατική συλλογή αλγορίθμων, αλλά ένα δυναμικό οικοσύστημα όπου η μαθηματική μοντελοποίηση συναντά την υπολογιστική ισχύ [4]. Η αποτελεσματικότητα κάθε εφαρμογής εξαρτάται από τον επιτυχή συνδυασμό τριών παραγόντων [23]:

1. Της ποιότητας των δεδομένων.
2. Της καταλληλότητας του μοντέλου.
3. Της αποδοτικότητας του αλγορίθμου εκπαίδευσης.

Καθώς η τεχνολογία εξελίσσεται, η έμφαση μετατοπίζεται πλέον από την απλή επίτευξη υψηλής ακρίβειας, στην κατανόηση της διαδικασίας με την οποία επιτυγχάνεται η μάθηση. Αυτό οδηγεί στην ανάγκη για σαφή κατηγοριοποίηση των μεθοδολογιών μάθησης, ανάλογα με το είδος της επίβλεψης που δέχεται το σύστημα κατά την εκπαιδευτική διαδικασία [12].

## 2.2 Κατηγορίες Μηχανικής Μάθησης

Η ταξινόμηση των συστημάτων Μηχανικής Μάθησης πραγματοποιείται κατά κύριο λόγο με βάση το είδος και τον βαθμό της ανθρώπινης επίβλεψης που παρέχεται κατά τη διαδικασία της εκπαίδευσης. Οι μεθοδολογίες αυτές διαφοροποιούνται ως προς τη φύση των δεδομένων εισόδου και τον τρόπο με τον οποίο το μοντέλο εξάγει συμπεράσματα.

Υπάρχουν τέσσερις βασικές κατηγορίες: εποπτευόμενη μάθηση, μη εποπτευόμενη μάθηση, ημι-εποπτευόμενη μάθηση και ενισχυτική μάθηση. Στην εποπτευόμενη μάθηση, τα δεδομένα συνοδεύονται από ετικέτες, ενώ στη μη εποπτευόμενη δεν υπάρχουν προκαθορισμένες κλάσεις. Η ημι-εποπτευόμενη μάθηση λειτουργεί ως ενδιάμεση γέφυρα των δυο πρώτων και η ενισχυτική μάθηση ακολουθεί μια τελείως διαφορετική λογική, βασισμένη στην αλληλεπίδραση με το περιβάλλον και στην έννοια της ανταμοιβής. Παρακάτω θα αναλυθούν αρχικά και οι τέσσερις αυτές κατηγορίες και στην συνέχεια θα εφαρμοστεί μια συγκριτική ανάλυση για αποτελεσματικότερη και ολοκληρωμένη κατανόηση.

- Εποπτευόμενη μάθηση (supervised learning): Στην επιβλεπόμενη μάθηση, τα δεδομένα εκπαίδευσης που τροφοδοτούνται στον αλγόριθμο περιλαμβάνουν τις επιθυμητές λύσεις που ονομάζονται ετικέτες. Επομένως, σε αυτήν τη μέθοδο δημιουργείται ένα προετοιμασμένο σύνολο δεδομένων εκπαίδευσης με τις αντίστοιχες γνωστές λύσεις (ετικέτες). Σε αυτή τη μορφή μηχανικής μάθησης, η τεχνητή νοημοσύνη είναι σε θέση να αναπαράγει μοτίβα. Τα αποτελέσματα είναι γνωστά στο σύστημα εκπαίδευσης και χρησιμοποιούνται για την εκπαίδευση του συστήματος.

Γνωστά παραδείγματα που αναφέρονται στην βιβλιογραφία είναι μεταξύ άλλων το φίλτρο της ανεπιθύμητης αλληλογραφίας, όπου ο αλγόριθμος εκπαιδεύεται με πολλά παραδείγματα ηλεκτρονικής αλληλογραφίας εάν πρόκειται για ανεπιθύμητη (spam) ή επιθυμητή (ham) αλληλογραφία, προκειμένου να μπορέσει να ταξινομήσει νέα εισερχόμενα μηνύματα σωστά [13].

Ένα επιπλέον παράδειγμα εποπτευόμενης μάθησης είναι η χρήση διαφορετικών φωτογραφιών ζώων. Το μοντέλο μαθαίνει αναλύοντας τα διάφορα δεδομένα εισόδου, με μεμονωμένα χαρακτηριστικά όπως το σχήμα, το χρώμα κ.α. Χρησιμοποιώντας αυτή τη γνώση μπορεί να αναγνωρίσει άλλα είδη ζώων είτε σε εικόνες είτε στην πραγματικότητα [14].

Κάποιοι από τους βασικότερους και ευρέως γνωστούς αλγόριθμους εποπτευόμενης μάθησης είναι οι παρακάτω [13]:

- K-nearest neighbors (KNN)

- Decision Trees
  - Random Forest
  - Naive Bayes
  - Linear & Logistic Regression
  - SVM
- Μη εποπτευόμενη μάθηση (unsupervised learning): Αντιθέτως, στην μη εποπτευόμενη μάθηση τα δεδομένα εκπαίδευσης δεν φέρουν ετικέτες. Το μοντέλο προσπαθεί να μάθει από τα δεδομένα εκπαίδευσης χωρίς καθοδήγηση, δηλαδή χωρίς να υπάρχουν προκαθορισμένες τιμές-στόχοι. Ο αλγόριθμος επιχειρεί να εντοπίσει μοτίβα και κρυμμένες δομές μέσα σε αδόμητα δεδομένα και να τα οργανώσει σε ομάδες, γνωστές ως συστάδες, χωρίς ωστόσο να υπάρχει ένα πρότυπο λύσης. Επομένως, ο στόχος αυτής της κατηγορίας μηχανικής μάθησης είναι η εξαγωγή γνώσης από τα διαθέσιμα δεδομένα, μέσω της ανάπτυξης μοντέλων που εντοπίζουν τα πρότυπα και τα μοτίβα που χαρακτηρίζουν ένα προκαθορισμένο σύνολο δεδομένων [32].

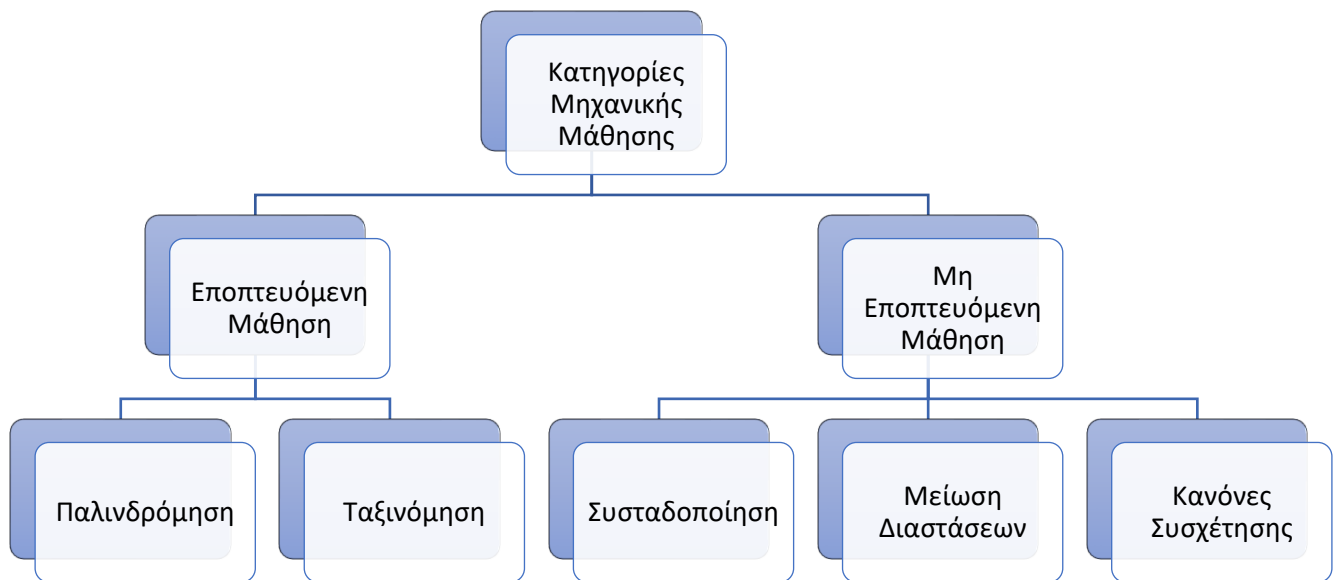
Στον τομέα της ιατρικής, η μη επιβλεπόμενη μάθηση χρησιμοποιείται συχνά για την ανίχνευση ανωμαλιών σε ιατρικές εικόνες, όπως ακτινογραφίες ή μαγνητικές τομογραφίες. Σε τέτοιες περιπτώσεις, η τεχνητή νοημοσύνη ομαδοποιεί διαφορετικού τύπου ανωμαλίες και αποτελέσματα, ενώ ένας ειδικός επαγγελματίας υγείας θα πρέπει να ερμηνεύσει το νόημά τους και να αξιολογήσει τη χρησιμότητά τους [14].

Οι πιο γνωστοί αλγόριθμοι μη εποπτευόμενης μάθησης αναφέρονται παρακάτω [35]:

- K-means
  - Ιεραρχική Συσταδοποίηση (Hierarchical Clustering)
  - Apriori
  - Principal Component Analysis (PCA)
  - Independent Component Analysis (ICA)
  - Self-Organizing Maps (SOMs)
- Ημι-εποπτευόμενη μάθηση (semi supervised): Η ημι-εποπτευόμενη μάθηση αποτελεί μια ενδιάμεση προσέγγιση ανάμεσα στην εποπτευόμενη και τη μη-εποπτευόμενη μάθηση, καθώς βασίζεται σε σύνολα δεδομένων όπου μόνο ένα μικρό μέρος δεδομένων έχει ετικέτες. Στόχος της είναι να αξιοποιήσει τον περιορισμένο αριθμό επισημασμένων δειγμάτων σε συνδυασμό με τα μη επισημασμένα, ώστε να ξεπεράσει τα μειονεκτήματα των δύο κλασικών μεθόδων. Η εποπτευόμενη μάθηση απαιτεί τεράστιο όγκο δεδομένων εκπαίδευσης για την ταξινόμηση των δεδομένων δοκιμής, η οποία είναι μια διαδικασία δαπανηρή και χρονοβόρα [36]. Αντιθέτως, η μη εποπτευόμενη μάθηση δεν απαιτεί καμία επισημάνση δεδομένων, η οποία ομαδοποιεί τα δεδομένα με βάση την ομοιότητα στα σημεία δεδομένων χρησιμοποιώντας είτε την προσέγγιση ομαδοποίησης είτε την προσέγγιση μέγιστης πιθανότητας. Το κύριο μειονέκτημα αυτής της προσέγγισης είναι ότι δεν μπορεί να ομαδοποιήσει με ακρίβεια άγνωστα δεδομένα. Για να αντιμετωπιστούν αυτά τα προβλήματα, προτάθηκε η ημι-εποπτευόμενη προσέγγιση, η οποία εκπαιδεύει ένα μοντέλο χρησιμοποιώντας λίγα χαρακτηρισμένα παραδείγματα και στη συνέχεια επιχειρεί να αποδώσει ετικέτες στα μη επισημασμένα δεδομένα [38]. Στο πλαίσιο αυτό, τα δεδομένα που έχουν ετικέτες λειτουργούν ως οδηγός, ενώ τα μη επισημασμένα συμβάλλουν στη βελτίωση της γενίκευσης του μοντέλου. Η ημι-εποπτευόμενη μάθηση διακρίνεται συνήθως σε δύο βασικές μορφές: την ημι-εποπτευόμενη ταξινόμηση και την ημι-εποπτευόμενη ομαδοποίηση [36].

- **Ενισχυτική μάθηση (reinforcement learning):** Η ενισχυτική μάθηση αποτελεί μια προσέγγιση κατά την οποία το μοντέλο μαθαίνει να επιλέγει τη βέλτιστη συμπεριφορά μέσα από μια διαδικασία αλληλεπίδρασης με το περιβάλλον του. Η διαδικασία της μάθησης πραγματοποιείται μέσω επαναλαμβανόμενων δοκιμών, όπου κάθε ενέργεια αξιολογείται βάσει της ανταμοιβής ή της ποινής που λαμβάνεται [35]. Κάθε ενέργεια που εκτελεί το σύστημα αξιολογείται μέσω μιας συνάρτησης ανταμοιβής, η οποία δείχνει αν η ενέργεια ήταν επιτυχής ή όχι. Με βάση αυτή την αξιολόγηση, η ενέργεια είτε ενσωματώνεται στη στρατηγική του συστήματος είτε απορρίπτεται. Στη συνέχεια, το σύστημα παρατηρεί στο επόμενο στάδιο ξανά το περιβάλλον και η διαδικασία επαναλαμβάνεται, έχοντας πλέον προσαρμοσμένη τη στρατηγική του. Σε πρώτη ανάγνωση, η μέθοδος αυτή ενδέχεται να θεωρηθεί συναφής με την εποπτευόμενη μάθηση. Παρ' όλα αυτά, στην ενισχυτική μάθηση η ανατροφοδότηση δεν σχετίζεται με την ορθότητα μιας ετικέτας ή μιας κατηγορίας, αλλά λειτουργεί ως δείκτης της ποιότητας της ενέργειας που εκτελέστηκε από το σύστημα [37].

Στην παρούσα διπλωματική εργασία θα επικεντρωθούμε ωστόσο στις κατηγορίες και τους αλγορίθμους εποπτευόμενης και μη-εποπτευόμενης μάθησης, όπως διακρίνονται και στο παρακάτω Σχήμα (Σχήμα 1) [37]:



Σχήμα 1: Κατηγορίες Μηχανικής Μάθησης

## 2.3 Συγκριτική Ανάλυση Κατηγοριών Μηχανικής Μάθησης

Το πεδίο της Μηχανικής Μάθησης χωρίζεται σε επιμέρους κατηγορίες, οι οποίες διαφοροποιούνται με βάση τον τρόπο εκπαίδευσης των μοντέλων, το είδος των διαθέσιμων δεδομένων και τον τρόπο με τον οποίο παρέχεται

ανατροφοδότηση στο σύστημα. Οι βασικές προσεγγίσεις, όπως αναλύθηκαν και στην προηγούμενη ενότητα, περιλαμβάνουν την εποπτευόμενη, τη μη-εποπτευόμενη, την ημι-εποπτευόμενη και την ενισχυτική μάθηση.

Η κατανόηση των διαφορών μεταξύ αυτών των τεχνικών διαδικασιών μάθησης είναι καθοριστική, καθώς επιτρέπει την επιλογή της καταλληλότερης μεθοδολογίας για την επίλυση ενός συγκεκριμένου προβλήματος.

Στον παρακάτω πίνακα συνοψίζονται τα κυριότερα χαρακτηριστικά στα οποία διαφέρουν οι τέσσερις κατηγορίες [35]:

**Πίνακας 1: Συγκριτική Ανάλυση Κατηγοριών Μηχανικής Μάθησης**

Χαρακτηριστικά	Εποπτευόμενη Μάθηση	Μη-Εποπτευόμενη Μάθηση	Ημι-Εποπτευόμενη Μάθηση	Ενισχυτική Μάθηση
Είδος Δεδομένων	Δεδομένα εμπριέχουν ετικέτες	Δεδομένα δεν εμπριέχουν ετικέτες	Λίγα δεδομένα εμπριέχουν ετικέτες, τα περισσότερα ωστόσο δεν έχουν	Δεδομένα από αλληλεπίδραση με περιβάλλον
Στόχος	Πρόβλεψη αποτελέσματος	Εύρεση μοτίβων μέσω ομαδοποίησης (Clustering) και μείωση πολυπλοκότητας	Συνδυασμός εποπτευόμενης & μη-εποπτευόμενης μάθησης για καλύτερη γενίκευση	Λήψη βέλτιστων αποφάσεων και στρατηγικής
Τύπος Μάθησης	Άμεση και συνεχή διόρθωση μέσω αντιστοίχισης δεδομένων εισόδου-εξόδου (Classification & Regression)	Καμία εξωτερική καθοδήγηση (Clustering & Association)	Μερική καθοδήγηση από υπάρχοντα παραδείγματα	Σύστημα ανταμοιβών και ποινών

Εφαρμογές	Αναγνώριση εικόνων, πρόβλεψη τιμών, Ανίχνευση απάτης	Ομαδοποίηση πελατών, συστήματα συστάσεων, ανίχνευση ανωμαλιών σε ιατρικές εικόνες	Ιατρικές διαγνώσεις, Αναγνώριση εικόνων	Αυτόνομα οχήματα χωρίς οδηγό, έξυπνα ρομπότ
-----------	--	---	---	---

## 2.4 Διαδικασία μάθησης και αξιολόγησης

Η ανάπτυξη ενός μοντέλου Μηχανικής Μάθησης αποτελεί μια σύνθετη και επαναληπτική διαδικασία, η οποία υπερβαίνει την απλή εφαρμογή ενός αλγορίθμου. Όπως έχει ήδη αναφερθεί, ο κεντρικός στόχος είναι η δημιουργία ενός συστήματος που δεν θα απομνημονεύει απλώς δεδομένα, αλλά θα είναι σε θέση να γενικεύει τη γνώση του σε νέες συνθήκες και νέα, άγνωστα δεδομένα. Η διαδικασία αυτή περιλαμβάνει τα εξής στάδια [23]:

1. Συλλογή και Προετοιμασία Δεδομένων: Η ποιότητα των δεδομένων (Data Quality) καθορίζει την επιτυχία του μοντέλου και είναι καθοριστικής σημασίας για την αποφυγή μεροληψίας (bias) στα αποτελέσματα.
2. Εκπαίδευση (Training): Ο αλγόριθμος επεξεργάζεται το σύνολο εκπαίδευσης για να εντοπίσει μοτίβα και στατιστικές σχέσεις.
3. Αξιολόγηση (Evaluation): Το μοντέλο ελέγχεται σε ένα ξεχωριστό σύνολο δεδομένων για να μετρηθεί η ακρίβειά του. Επομένως, η αξιολόγηση πραγματοποιείται συνήθως με τη χρήση ενός συνόλου δεδομένων ελέγχου (test set), το οποίο δεν έχει χρησιμοποιηθεί κατά την εκπαίδευση, ώστε να εκτιμηθεί η ικανότητα του μοντέλου να ανταπεξέρχεται σε πραγματικές συνθήκες [13]. Στην πράξη, συχνά χρησιμοποιείται και ένα ενδιάμεσο σύνολο επαλήθευσης (Validation Set) για να ρυθμίσουμε τις παραμέτρους του μοντέλου πριν την τελική δοκιμή [13] & [34].

Κομβικό σημείο σε αυτή τη διαδικασία μάθησης και αξιολόγησης είναι το σφάλμα γενίκευσης (generalization error). Στόχος δεν είναι η τέλεια απόδοση στα ήδη γνωστά δεδομένα, αλλά η πρόβλεψη σε νέα, άγνωστα δείγματα. Σύμφωνα με τον Mitchell (1997) [2], η ικανότητα ενός μοντέλου να γενικεύει από την εμπειρία είναι αυτό που διαχωρίζει τη γνήσια μάθηση από την απλή απομνημόνευση. Σε αυτό το πλαίσιο, εμφανίζονται δύο μεγάλες προκλήσεις [13]:

- Underfitting (Υποπροσαρμογή): Συμβαίνει όταν το μοντέλο είναι υπερβολικά απλό για να μάθει τη δομή των δεδομένων. Το αποτέλεσμα είναι η χαμηλή ακρίβεια τόσο στα δεδομένα εκπαίδευσης όσο και στα νέα δεδομένα [34].
- Overfitting (Υπερπροσαρμογή): Συμβαίνει όταν το μοντέλο μαθαίνει "υπερβολικά καλά" τα δεδομένα εκπαίδευσης, συμπεριλαμβανομένου ακόμα και του θορύβου και των τυχαίων διακυμάνσεων [34]. Αυτό σημαίνει ότι το μοντέλο είναι υπερβολικά ευέλικτο και προσαρμόζεται στις τυχαίες διακυμάνσεις των δεδομένων εκπαίδευσης. Σε αυτή την περίπτωση, το μοντέλο έχει εξαιρετική απόδοση στην φάση της εκπαίδευσης, αλλά αποτυγχάνει σε νέα δεδομένα, διότι δεν είναι σε θέση να γενικεύσει την γνώση του [5].

Για την αντιμετώπιση αυτών, εφαρμόζονται τεχνικές όπως η Κανονικοποίηση (Regularization) στην περίπτωση του overfitting, ή η χρήση πιο σύνθετων μοντέλων στην περίπτωση του underfitting [5].

## 2.5 Περιορισμοί και προκλήσεις

Παρά τα σημαντικά πλεονεκτήματά της, η Μηχανική Μάθηση αντιμετωπίζει προκλήσεις και περιορισμούς που σχετίζονται με την ποιότητα των δεδομένων, τη διαφάνεια των μοντέλων και τα ηθικά ζητήματα. Όπως αναλύθηκε και στην ενότητα 2.4 σχετικά με τη διαδικασία εκπαίδευσης, η ικανότητα ενός μοντέλου να ελαχιστοποιεί το σφάλμα και να γενικεύει τη γνώση του, εξαρτάται άμεσα από την εμπειρία που αποκτά μέσω των δεδομένων. Στην πράξη, αυτό σημαίνει ότι όσο περισσότερα δεδομένα εκπαίδευσης λαμβάνει ένας αλγόριθμος, τόσο πιο γρήγορα μπορεί να βελτιώσει το μοντέλο του και να μειώσει το ποσοστό σφαλμάτων. Ωστόσο, παραμένει κρίσιμο το μοντέλο να διατηρηθεί αρκετά γενικό, ώστε να μπορεί να εφαρμόζεται με επιτυχία και σε νέα δεδομένα που δεν περιλαμβάνονταν στη φάση εκπαίδευσης [38].

Ταυτόχρονα, η ποιότητα ενός μοντέλου εξαρτάται άμεσα από την ποιότητα των δεδομένων εισόδου. Εάν στον αλγόριθμο παρουσιάζονται λανθασμένα ή μη αντιπροσωπευτικά δεδομένα, η ικανότητά του να εξάγει σωστά συμπεράσματα περιορίζεται, καθιστώντας τα αποτελέσματα για νέα δεδομένα εισόδου αβέβαια [39]. Για παράδειγμα, αν ένα σύστημα πρόβλεψης πωλήσεων εκπαιδευτεί με δεδομένα που δεν περιλαμβάνουν περιόδους κρίσης, η ικανότητα γενίκευσής του σε μελλοντικές οικονομικές διακυμάνσεις θα είναι περιορισμένη.

Επιπλέον, ένα ζήτημα που αναφέρεται συχνά στη σύγχρονη έρευνα είναι η αλγοριθμική μεροληψία (algorithmic bias). Τα μοντέλα τείνουν να αναπαράγουν και να ενισχύουν προκαταλήψεις που μπορεί να υπάρχουν στα δεδομένα εκπαίδευσης [7]. Ένα παράδειγμα αποτελούν τα συστήματα αυτόματης αξιολόγησης βιογραφικών, τα οποία ενδέχεται να εμφανίσουν διακρίσεις λόγω φύλου ή καταγωγής. Αυτό μπορεί να έχει ως αποτέλεσμα στην επιλογή βιογραφικών για την κάλυψη μια θέσης με χαμηλότερες αποδοχές, να επιλέγονται κατά μέσο όρο περισσότερες γυναίκες από ότι άντρες [8]. Σύμφωνα με τους Floridi et al. [7], η έλλειψη ξεκάθαρων ηθικών κανόνων μπορεί να οδηγήσει σε άδικες κοινωνικές διακρίσεις.

Μια ιδιαίτερα σημαντική ιδιότητα στη Μηχανική Μάθηση είναι η αναγνωσιμότητα (interpretability) των μοντέλων. Συγκεκριμένα, ο όρος του «μαύρου κουτιού» (Black Box) χρησιμοποιείται στην βιβλιογραφία για να περιγράψει την ανάγκη για πιο αποτελεσματική και εις βάθος κατανόηση των αλγορίθμων [9]. Ορισμένοι αλγόριθμοι, όπως τα δέντρα αποφάσεων, έχουν τη δυνατότητα να εξηγούν τη λογική τους και να παρέχουν μια εκτίμηση για το πόσο βάσιμο είναι το αποτέλεσμα [10]. Παρόλα αυτά, πολλοί σύγχρονοι και σύνθετοι αλγόριθμοι παραμένουν τα λεγόμενα ως «μαύρα κουτιά». Αυτό σημαίνει ότι, αν και προσφέρουν υψηλή ακρίβεια, η μαθηματική λογική πίσω από τις αποφάσεις τους είναι εξαιρετικά δύσκολο να κατανοηθεί [11].

Όπως εξηγεί ο [12], σε κρίσιμους τομείς όπως μεταξύ άλλων η ιατρική και η δικαιοσύνη, το να κατανοηθεί πώς οδηγήθηκε το μοντέλο σε μια απόφαση είναι εξίσου σημαντικό με την ίδια την πρόβλεψη. Η έλλειψη αυτής της διαφάνειας μπορεί να δημιουργήσει ζητήματα εμπιστοσύνης, καθώς οι χρήστες δυσκολεύονται να βασιστούν σε ένα σύστημα που δεν μπορεί να αιτιολογήσει τα αποτελέσματά του [13]. Συνεπώς, η ισορροπία ανάμεσα στην υψηλή απόδοση και τη διαφάνεια αποτελεί μία από τις μεγαλύτερες προκλήσεις της Μηχανικής Μάθησης και της Τεχνητής Νοημοσύνης ως ευρύτερο κλάδο.

Τέλος, είναι εξίσου σημαντικό να λαμβάνεται υπόψη το οικονομικό και περιβαλλοντικό κόστος. Η εκπαίδευση σύγχρονων μοντέλων απαιτεί πολύ μεγάλη υπολογιστική ισχύ και κατανάλωση ενέργειας. Αυτό καθιστά εξαιρετικά δύσκολη την χρήση τους από μικρότερες επιχειρήσεις και οργανισμούς, ενώ παράλληλα η τεράστια κατανάλωση ηλεκτρικού ρεύματος από τους υπολογιστές που εκπαιδεύουν αυτά τα μοντέλα Μηχανικής Μάθησης στα κέντρα δεδομένων (Data Centers) έχει αρνητικές επιπτώσεις στο περιβάλλον [23].

Συνοψίζοντας, η επιτυχής εφαρμογή της Μηχανικής Μάθησης δεν βασίζεται μόνο στην τεχνική ποιότητα του αλγορίθμου, αλλά απαιτεί επίσης την τήρηση ηθικών κανόνων και τη σωστή αξιολόγηση των δεδομένων και των πόρων που χρησιμοποιούνται.

## **ΚΕΦΑΛΑΙΟ 3 – ΑΝΑΛΥΣΗ ΚΑΙ ΜΕΘΟΔΟΛΟΓΙΚΟ ΠΛΑΙΣΙΟ ΕΦΑΡΜΟΓΗΣ ΑΛΓΟΡΙΘΜΩΝ ΜΗΧΑΝΙΚΗΣ ΜΑΘΗΣΗΣ**

Στο παρόν κεφάλαιο παρουσιάζονται οι κυριότεροι αλγόριθμοι Μηχανικής Μάθησης, όχι μόνο ως θεωρητικά μοντέλα, αλλά ως εργαλεία επίλυσης σύνθετων προβλημάτων. Η προσέγγιση που ακολουθείται χωρίζεται σε δύο βασικούς πυλώνες: Ο πρώτος πυλώνας αναλύει τη λειτουργία κάθε αλγορίθμου μέσα από συγκεκριμένα παραδείγματα εφαρμογής που συνοδεύονται από αναλυτικά βήματα εκτέλεσης. Ο δεύτερος πυλώνας παρέχει μια ολοκληρωμένη μεθοδολογία επιλογής, η οποία καθοδηγεί τον χρήστη με την βοήθεια κατάλληλων κριτηρίων για να αποφασίζει ποιος αλγόριθμος ταιριάζει καλύτερα σε κάθε περίπτωση, λαμβάνοντας υπόψη τη δομή των πληροφοριών και τις ανάγκες του εκάστοτε προβλήματος

### **3.1 Εισαγωγή στους αλγορίθμους Μηχανικής Μάθησης**

Υπάρχουν αρκετά μοντέλα Μηχανικής Μάθησης, που αποτελούνται από αλγορίθμους για τη διαδικασία εκπαίδευσης δεδομένων και εξαγωγής αποτελεσμάτων. Οι υπολογιστικές πράξεις, μέσα σε αυτούς, έχουν την ικανότητα να αναγνωρίζουν μοτίβα, να προβλέπουν και να αποφασίζουν την τιμή που θα εξαχθεί. Η επιλογή του καταλληλότερου μοντέλου γίνεται από το χρήστη, αναλόγως το πρόβλημα που καλείται να επιλύσει και την ποιότητα των δεδομένων.

Στη βιβλιογραφία παρατηρείται μεγάλη ποικιλία αλγορίθμων, οι οποίοι διαφέρουν ως προς την πολυπλοκότητα, την ευκολία ερμηνείας και την απόδοσή τους. Στο παρόν κεφάλαιο παρουσιάζονται οι βασικότεροι αλγόριθμοι, με έμφαση στις θεωρητικές τους αρχές και τα πεδία εφαρμογής τους.

### **3.2 Αλγόριθμοι εποπτευόμενης μάθησης**

Η εποπτευόμενη μάθηση αποτελεί την πιο διαδεδομένη κατηγορία Μηχανικής Μάθησης, καθώς χρησιμοποιεί σύνολα δεδομένων με γνωστές ετικέτες εξόδου. Στόχος είναι η εκμάθηση μιας συνάρτησης που χαρτογραφεί τα δεδομένα εισόδου στις αντίστοιχες τιμές εξόδου.

#### **3.2.1 Γραμμική και λογιστική παλινδρόμηση**

Η γραμμική και η λογιστική Παλινδρόμηση αποτελούν τα θεμέλια της εποπτευόμενης μάθησης, λειτουργώντας ως ιδανικά εισαγωγικά μοντέλα για την κατανόηση της ευρύτερης αλγοριθμικής λογικής. Χάρη στη μαθηματική τους απλότητα, η διαδικασία εξαγωγής συμπερασμάτων παραμένει απολύτως διαφανής, ενώ τα αποτελέσματά τους είναι εύκολα επαληθεύσιμα από τον αναλυτή, καθώς οι παραγόμενες τιμές εξόδου αντιστοιχούν σε γνωστά δεδομένα εκπαίδευσης [24].

Εξετάζοντας τον πρώτο αλγόριθμο, η γραμμική Παλινδρόμηση λειτουργεί αποδίδοντας συγκεκριμένους συντελεστές βαρύτητας σε κάθε χαρακτηριστικό εισόδου, αναδεικνύοντας τον βαθμό επιρροής του στην τελική πρόβλεψη. Μέσα από διαδοχικές επαναλήψεις, το μοντέλο αναπροσαρμόζει διαρκώς αυτούς τους συντελεστές με βασικό στόχο την ελαχιστοποίηση του σφάλματος πρόβλεψης, επιδιώκοντας την ιδανική σύγκλιση. Αφού υπολογιστεί η τελική αριθμητική τιμή, η διαδικασία μπορεί να προχωρήσει στο επόμενο στάδιο, αυτό της κατηγοριοποίησης [24].

Σε αυτό ακριβώς το σημείο επεμβαίνει η λογιστική Παλινδρόμηση, η οποία εξειδικεύεται στη λήψη αποφάσεων μέσω δυαδικής ταξινόμησης. Πρακτικά, αξιοποιεί το αποτέλεσμα της γραμμικής συσχέτισης και θέτει ένα αυστηρό μαθηματικό όριο, προκειμένου να κατατάξει τη νέα παρατήρηση σε μία από δύο διακριτές κατηγορίες [24]. Αν και υπάρχει η δυνατότητα προσαρμογής του μοντέλου για ταξινόμηση σε περισσότερες κατηγορίες, η προσέγγιση αυτή ξεφεύγει από την κλασική μορφή του αλγορίθμου. Ως βασικό πλεονέκτημα, η συγκεκριμένη

μεθοδολογία απαιτεί ελάχιστους υπολογιστικούς πόρους και εκτελείται ταχύτατα. Εντούτοις, υστερεί σημαντικά όταν εφαρμόζεται σε πολύπλοκα, μη γραμμικά σύνολα δεδομένων, ενώ αποδεικνύεται ιδιαίτερα ευάλωτη στην παρουσία ακραίων τιμών, οι οποίες μπορούν να αλλοιώσουν δραματικά την τελική εκτίμηση [25].

Η πρακτική αξία αυτών των μεθόδων μπορεί να γίνει κατανοητή μέσα από την πρόβλεψη της κατανάλωσης καυσίμου ενός οχήματος βάσει της ταχύτητάς του [24] [25]:

Δεδομένα: ιστορικά δεδομένα καταγραφών οχημάτων με δύο μεταβλητές: ταχύτητα (km/h) και κατανάλωση καυσίμου (L/100km). Για παράδειγμα [80 km/h → 5L], [100 km/h → 6L].

1. Εισαγωγή Δεδομένων: Στο μοντέλο εισάγονται τα ιστορικά δεδομένα των διαδρομών και αναγνωρίζει τη μεταβλητή-στόχο (κατανάλωση).
2. Εκπαίδευση & Εύρεση Βαρών: Ο αλγόριθμος υπολογίζει τη μαθηματική συσχέτιση (ευθεία γραμμή) μεταξύ των δύο μεταβλητών, αναγνωρίζοντας ότι κάθε αύξηση της ταχύτητας κατά 20 km/h θα έχει ως αποτέλεσμα την αύξηση της κατανάλωσης κατά 1L.
3. Πρόβλεψη: Εισάγοντας μια νέα άγνωστη ταχύτητα (π.χ. 120 km/h). Το μοντέλο χρησιμοποιεί την ευθεία που δημιούργησε και εξάγει την σχετική πρόβλεψη: 7 L/100km.

### 3.2.2 k-Nearest Neighbors (k-NN)

Ο k-Nearest Neighbors (k-NN) αποτελεί ακόμα έναν από τους κυριότερους και πιο κατανοητούς αλγορίθμους της μηχανικής μάθησης. Η εφαρμογή του ενδείκνυται κυρίως σε σύνολα δεδομένων μικρού έως μεσαίου μεγέθους, καθώς σε αυτές τις περιπτώσεις οι υπολογιστικές απαιτήσεις σε μνήμη και χρόνους εκτέλεσης διατηρούνται σε χαμηλά επίπεδα. Εντάσσεται στην κατηγορία της εποπτευόμενης μάθησης, γεγονός που προϋποθέτει ότι τα δεδομένα εκπαίδευσης είναι ήδη ταξινομημένα και φέρουν γνωστές ετικέτες. Η κεντρική του φιλοσοφία βασίζεται στην έννοια της εγγύτητας: όταν εισάγεται μια νέα, άγνωστη παρατήρηση, το σύστημα εντοπίζει τα πιο κοντινά της σημεία (γείτονες) στον χώρο των δεδομένων και της αποδίδει την αντίστοιχη κλάση. Στο πλαίσιο αυτό, η παράμετρος k αντιπροσωπεύει τον ακριβή αριθμό των γειτόνων που θα ληφθούν υπόψη για την τελική συσχέτιση [26].

Η σημαντικότερη ίσως πρόκληση κατά τη χρήση του k-NN είναι η επιλογή της βέλτιστης τιμής για το k, καθώς δεν υφίσταται κάποιος αυστηρός μαθηματικός κανόνας που να την καθορίζει αυτόματα, αλλά απαιτείται πειραματισμός από τον αναλυτή. Ειδικότερα, σε περιβάλλοντα με έντονο θόρυβο και ακραίες τιμές, η επιλογή ενός μεγαλύτερου k λειτουργεί σταθεροποιητικά για την τελική πρόβλεψη. Εντούτοις, εάν η τιμή αυξηθεί υπερβολικά, ο αλγόριθμος τείνει να γενικεύει υπέρμετρα τα δεδομένα χάνοντας σημαντικά τοπικά μοτίβα, φαινόμενο που αναλύθηκε στο Κεφάλαιο 2 ως υποπροσαρμογή. Στον αντίποδα, μια πολύ μικρή τιμή καθιστά το μοντέλο εξαιρετικά ευάλωτο σε μεμονωμένους θορύβους, οδηγώντας σε υπερπροσαρμογή [5].

Αναφορικά με τον τρόπο μέτρησης της απόστασης, η πιο καθιερωμένη επιλογή είναι η Ευκλείδεια απόσταση, η οποία υπολογίζει το μήκος της ευθείας που ενώνει τη νέα παρατήρηση με τους γείτονές της. Ωστόσο, η μέθοδος αυτή παρουσιάζει αδυναμίες: εάν μια νέα, ακραία τιμή βρεθεί γεωμετρικά κοντά σε μια συμπαγή ομάδα δεδομένων, ο αλγόριθμος ενδέχεται να την ταξινομήσει λανθασμένα, αποδίδοντάς της τα χαρακτηριστικά της ομάδας. Το συγκεκριμένο πρόβλημα μπορεί να μετριαστεί εάν ο χρήστης ενσωματώσει αυστηρά όρια απόστασης ή υποσυστήματα ανίχνευσης ανωμαλιών [3]. Εναλλακτικά, προτείνεται η χρήση διαφορετικών μετρικών υπολογισμού, όπως η απόσταση Manhattan ή Minkowski, οι οποίες αποδεικνύονται σαφώς πιο ανθεκτικές και αποδοτικές όταν το σύνολο δεδομένων περιέχει ακραίες τιμές ή χαρακτηρίζεται από πολλές διαστάσεις [27].

Ένα αντιπροσωπευτικό παράδειγμα θα μπορούσε να είναι η κατάταξη μιας ταινίας με βάση τα χαρακτηριστικά της [5] [26]:

Δεδομένα: Μια βάση δεδομένων με ταινίες (ετικέτες: «Δράση» ή «Κωμωδία»). Τα χαρακτηριστικά είναι ο [Αριθμός αστειών, Αριθμός σκηνών μάχης]. Επιλέγουμε k=3 γείτονες.

1. Τοποθέτηση στον Χώρο: Όλες οι γνωστές ταινίες τοποθετούνται σε ένα γράφημα βάσει των χαρακτηριστικών τους.

2. Υπολογισμός Απόστασης: Προσθέτοντας μια νέα και άγνωστη ταινία με χαρακτηριστικά [3 αστέρια, 8 σκηνές μάχης]. Ο αλγόριθμος υπολογίζει την Ευκλείδεια απόστασή από όλες τις υπόλοιπες ταινίες και βρίσκει τις 3 πιο κοντινές ( $k=3$ ).
3. Ταξινόμηση: Αν από τους τρεις γείτονες, οι δυο ανήκουν στην κατηγορία «Δράση» και ένας στην «Κωμωδία», η νέα ταινία ταξινομείται βάσει πλειοψηφίας στην κατηγορία «Δράση».

### 3.2.3 Δέντρα αποφάσεων και Random Forest

Τα δέντρα αποφάσεων (ΔΑ) χρησιμοποιούνται σε αρκετά μοντέλα μηχανικής μάθησης. Η έννοια του δέντρου απόφασης είναι ουσιαστικά μια διαδικασία πρόβλεψης που προκύπτει από την υποδιαίρεση των δεδομένων σε διαδοχικές και ξεχωριστές ομάδες ή αλλιώς κόμβους. Κάθε κόμβος διεξάγει έναν έλεγχο και είτε θα συνεχίζει σε επόμενο κλαδί συνδεδεμένο με κόμβο όπου θα κάνει τον επόμενο έλεγχο, είτε θα τερματίσει με μια τελική δικιά του απόφαση. Απόρροια αυτής της διαδικασίας είναι να καταλήξει η απόφαση κάθε υπο-ομάδας να είναι ξεκάθαρη και συνολικά να βγει η κατηγορία ή η αριθμητική πρόβλεψη [28]. Ένα μονάχα ΔΑ είναι πιο ευανάγνωστο και βοηθάει στην κατανόηση και στη συλλογιστική διαδρομή που ακολουθήθηκε για το αποτέλεσμα. Λειτουργεί καλά σε μικρές βάσεις δεδομένων αλλά το overfitting είναι ένα συχνό πρόβλημα που προκύπτει εφόσον μεγαλώσει αρκετά το δέντρο.

Σε αυτό το πρόβλημα γίνεται η χρήση Random Forest (RF). Το RF συνδυάζει πολλά δέντρα μαζί, παράλληλα, ο συνδυασμός των δέντρων αυτών χρησιμοποιεί τη μέθοδο της τυχαιοποιημένης ομαδοποίησης. Με αυτόν τον τρόπο τα πολλά και τυχαία ομαδοποιημένα δεδομένα θα καταλήξουν σε πολλές αποφάσεις και θα υπερτερήσουν τα επικρατέστερα αποτελέσματα που επαναλαμβάνονται περισσότερο και δημιουργούν πιο αξιόπιστο συμπέρασμα. Οι λανθασμένες αποφάσεις δεν θα είναι ισχυρές, διότι όλα τα δέντρα θα “ψηφίσουν” και η επικρατέστερη τιμή θα είναι αυτή που θα εξαχθεί [29].

Ένα παράδειγμα, που τα δέντρα αποφάσεων μπορούν να αναλύσουν, είναι η έγκριση ενός τραπεζικού δανείου [28]:

Δεδομένα: Αιτήσεις πελατών για χορήγηση τραπεζικού δανείου. Χαρακτηριστικά: [Εισόδημα, Ύπαρξη ανοιχτών οφειλών].

1. 1ος Κόμβος Ελέγχου: Ο αλγόριθμος ερευνά το πρώτο χαρακτηριστικό του πιθανού δανειολήπτη: «Εισόδημα > 20.000€;». Εάν η απάντηση είναι «Όχι», το δάνειο απορρίπτεται άμεσα. Εάν είναι «Ναι», συνεχίζει στον επόμενο κόμβο.
2. 2ος Κόμβος Ελέγχου: Ο επόμενος κόμβος αναλύει: «Υπάρχει άλλη ανοιχτή οφειλή;». Έστω ότι η απάντηση είναι «Όχι».
3. Τελική Απόφαση: Εφόσον ικανοποιήθηκαν τα κριτήρια ελέγχου των διαδοχικών κόμβων, το δέντρο καταλήγει στο τελικό φύλλο, αποδίδοντας την ετικέτα «Έγκριση». Στην περίπτωση του Random Forest, αυτή η διαδικασία εκτελείται ταυτόχρονα σε εκατοντάδες δέντρα και η τελική έγκριση βγαίνει με βάσει την πλειοψηφία.

### 3.2.4 Naive Bayes

Ο αλγόριθμος Naive Bayes βασίζεται στο θεώρημα του Bayes και στην υπόθεση της ανεξαρτησίας των χαρακτηριστικών. Το θεώρημα του Bayes, σε γενικά πλαίσια, είναι η αρχική ικασία, βάση πιθανοτήτων, που γίνεται για να αποκτήσει περιγραφή μια κατάσταση. Στη συνέχεια αυτή ενημερώνεται όταν εμφανιστεί μία καινούρια παρατήρηση. Αναλόγως με την επαναληψιμότητα της, επηρεάζει την κατάσταση, η οποία θα αναπροσαρμοστεί στα νέα δεδομένα. Παρά την απλουστευτική αυτή υπόθεση, ο Naive Bayes παρουσιάζει συχνά υψηλή απόδοση [30].

Για να λειτουργήσει αποδοτικά, το dataset θα πρέπει να περιέχει κλάσεις που συνθέτουν τα χαρακτηριστικά μιας κατάστασης. Συνδυαστικά με τη βαρύτητα της πιθανότητας που παίρνει κάθε παρατήρηση κατά την εκπαίδευση, το αποτέλεσμα γίνεται όλο και πιο αξιόπιστο.

Οι μεμονωμένες ακραίες τιμές δεν το επηρεάζουν εφόσον αντιμετωπίζει τα χαρακτηριστικά ως ανεξάρτητα. Γι αυτό και δεν αποκτά λανθασμένη «άποψη» όταν υπάρχουν διαφορετικού τύπου θόρυβοι, καθώς η εκπαίδευσή του βασίζεται καθαρά σε πιθανότητες και όχι μέσο όρο, όπου μια ακραία τιμή θα είχε μεγάλη σημασία. Επί πρόσθετα, έχει την ικανότητα να λειτουργεί με μεγάλη ευστοχία τόσο σε μικρά, όσο και μεγάλα dataset [31].

Είναι σημαντικό και σε αυτό το μοντέλο τα δεδομένα που εισάγονται να έχουν ποιότητα, εξού και ανήκει στην εποπτευόμενη μάθηση. Εάν οι ακραίες τιμές ή οι θόρυβοι επαναλαμβάνονται στα δεδομένα, τότε αρχίζει να τους δίνει σημασία και να τους χρησιμοποιεί σε μια απόφαση. Επίσης, μια νέα έγκυρη παρατήρηση που δεν έχει ξαναεμφανιστεί, θα έχει την αντιμετώπιση μιας ακραίας, δηλαδή θα παραγκωνιστεί και θα αποκτήσει «δύναμη» μόνο εάν εισαχθεί πολλές φορές [31].

Για να ομαλοποιηθεί η παραπάνω περίπτωση θορύβων μπορεί να γίνει χρήση τεχνικής εξομάλυνσης Laplace ή Lidstone. Η πρώτη τεχνική λειτουργεί καλύτερα σε μεγάλα dataset και η δεύτερη σε μικρότερα. Οι τεχνικές αυτές δίνουν κατευθείαν ένα μερίδιο πιθανότητας σε κάθε νέα παρατήρηση που εισέρχεται [33].

Για παράδειγμα, η ανεπιθύμητη αλληλογραφία περιέχει πολύ συχνά λέξεις όπως "κέρδος", "δωρεάν", "κλικ". Εφόσον αυτές οι λέξεις κλειδιά έχουν χαρακτηριστεί ως Spam σε προηγούμενα mail, ο Naive Bayes εξετάζει κάθε λέξη του νέου email ξεχωριστά, υποθέτοντας ότι η μια δεν επηρεάζει την άλλη [31] [33]:

Δεδομένα: Ιστορικό email κατηγοριοποιημένων ως «Ανεπιθύμητα» (Spam) ή «Κανονικά», με καταγραφή συχνότητας λέξεων.

1. Εξαγωγή Χαρακτηριστικών: Λαμβάνεται ένα νέο email με θέμα «Επιστροφή φόρων». Ο αλγόριθμος το διαχωρίζει σε μεμονωμένες λέξεις, θεωρώντας ότι η καθεμία είναι εντελώς ανεξάρτητη.
2. Υπολογισμός Πιθανοτήτων: Βάσει του θεωρήματος Bayes, υπολογίζει συνδυαστικά την πιθανότητα οι λέξεις «επιστροφή» και «φόρων» να ανήκουν στην κατηγορία Spam (π.χ. αναγνωρίζει ότι εμφανίζονται στο 90% των γνωστών spam email).
3. Τελική Κατηγοριοποίηση: Ο αλγόριθμος υπολογίζει την πιθανότητα της ανεπιθύμητης αλληλογραφίας (π.χ. 90%). Εφόσον η πιθανότητα υπερβαίνει το καθορισμένο όριο, το μήνυμα δρομολογείται αυτόματα στον φάκελο των ανεπιθύμητων.

### 3.3 Αλγόριθμοι μη εποπτευόμενης μάθησης

Τα δεδομένα που έχουμε στη διάθεσή μας, δεν είναι πάντα ξεκάθαρα και πολλές φορές χωρίς γνωστές ετικέτες. Η μη εποπτευόμενη μάθηση μπορεί να φέρει εις πέρας αποτελέσματα εάν εφαρμοστεί σε τέτοιου είδους δεδομένα, με στόχο την ανακάλυψη κρυμμένων δομών και μοτίβων.

#### 3.3.1 Συσταδοποίηση K-Means (Clustering)

Η συσταδοποίηση επιδιώκει τη δημιουργία ομάδων δεδομένων με παρόμοια χαρακτηριστικά. Ο πιο διαδεδομένος αλγόριθμος είναι ο K-Means, ο οποίος διαχωρίζει τα δεδομένα σε προκαθορισμένο αριθμό συστάδων. Παρά τη δημοτικότητά του, ο K-Means παρουσιάζει περιορισμούς σε δεδομένα με πολύπλοκες δομές. Ο K-Means χρησιμοποιείται αποδοτικά σε μεγάλες βάσεις δεδομένων, όπου τα δεδομένα ομαδοποιούνται με βάση την ομοιότητα των χαρακτηριστικών τους. Ας θέσουμε σε βασικά βήματα τις λειτουργίες του μοντέλου [40].

Ως πρώτο βήμα ο χρήστης, μαζί με το dataset, εισάγει το K, που είναι ο αριθμός κέντρων που θα δημιουργήσουν τις ομάδες, καθώς και τις επαναλήψεις που θα γίνει η διαδικασία από την αρχή [40].

Σε δεύτερη φάση, κάθε δεδομένο θα επιλεγεί από την ομάδα όπου το κέντρο της βρίσκεται πιο κοντά σε αυτό. Ο K-Means ορίζει τυχαία το πρώτο κεντρικό σημείο κι έπειτα τα επόμενα δημιουργούνται με πιθανότητα ανάλογη της απόστασής τους από τα ήδη δημιουργημένα. Με αυτόν τον τρόπο επιτυγχάνει μια πιο απλωμένη

συσταδοποίηση. Η διαδικασία επαναλαμβάνεται με αρχικοποίηση ενός καινούριου και τυχαίου πρώτου κέντρου. Παράλληλα και σαν επόμενο βήμα, η κάθε συστάδα δεδομένων βγάζει το μέσο όρο της και συνεχίζει με το καινούριο μέσο όρο να λαμβάνετε ως νέο κέντρο. Δημιουργεί νέα ομαδοποίηση των δεδομένων, μέχρι οι τιμές των μέσων όρων να είναι σταθερές ή να μεταβάλλονται ελάχιστα [40].

Η διαδικασία θα τελειώσει με το πέρας των επαναλήψεων που έχουμε δώσει, αλλά δεν θα σταματήσει στην καταλληλότερη ομαδοποίηση, δηλαδή το σωστότερο  $K$ . Ωστόσο, μπορούμε να εισάγουμε κάποια μέθοδο, όπως το Silhouette Width, η οποία μετά από κάθε επανάληψη  $K$  κρατάει έναν αριθμό που κυμαίνεται μεταξύ του  $-1$  και  $1$ . Η μεταβλητή που έχει πιο κοντινή τιμή στο  $1$  σηματοδοτεί και το ιδανικότερο  $K$  [41].

Επίσης το μοντέλο ενώ μπορεί να αποδώσει σε αριθμητικά δεδομένα, δεν μπορεί να κάνει το ίδιο, απευθείας, σε κατηγορικά δεδομένα, εφόσον χρησιμοποιεί μέσο όρο. Ακόμη, για τον ίδιο λόγο είναι ευαίσθητος σε ακραίες τιμές. Λογίζεται ως ένας από τους πιο εύκολους αλγόριθμους, ως προς την κατανόηση, για την μη εποπτευόμενη μάθηση [41].

Για να κατανοηθεί καλύτερα ο παραπάνω αλγόριθμος θα εξετάσουμε πως μια εταιρία τηλεπικοινωνιών θα μπορούσε να ομαδοποιήσει τους πελάτες της, ώστε να στείλει προσφορές [40]:

Δεδομένα: Μια λίστα συνδρομητών τηλεφωνίας χωρίς ετικέτες, με χαρακτηριστικά: [δεδομένα ίντερνετ, λεπτά ομιλίας]. Στόχος είναι ο διαχωρισμός σε  $K=2$  ομάδες προσφορών.

1. Προσδιορισμός  $K$ : Ο χρήστης ορίζει  $K=2$ . Ο αλγόριθμος βάζει τυχαία δύο σημεία (κέντρα) στον χώρο των δεδομένων. Ένα «φοιτητικό» προφίλ και ένα «επιχειρηματικό».
2. Αντιστοίχιση & Αναπροσαρμογή: Κάθε πελάτης μπαίνει στην ομάδα του κέντρου που βρίσκεται πιο κοντά σύμφωνα με την δική του χρήση. Στη συνέχεια, υπολογίζεται ο νέος μέσος όρος κάθε ομάδας και τα κέντρα μεταβάλλονται και παίρνουν νέες συντεταγμένες.
3. Οριστικοποίηση Ομάδων: Η διαδικασία αντιστοίχισης και μετακίνησης επαναλαμβάνεται μέχρι τα κέντρα να σταθεροποιηθούν πλήρως. Πλέον, οι πελάτες έχουν ομαδοποιηθεί αποτελεσματικά για την αποστολή προσωποποιημένων προσφορών.

### 3.3.2 Ιεραρχική Ομαδοποίηση (Hierarchical Clustering)

Η Ιεραρχική Ομαδοποίηση είναι μια μεθοδολογία που ακολουθούν κάποιοι αλγόριθμοι. Βασική ιδέα και η πιο συνηθισμένη μορφή, ιεραρχικής συσταδοποίησης, είναι η αρχική διάσπαση όλων των δεδομένων σε μονάδες που αποτελούν το καθένα μια διαφορετική συστάδα. Στη συνέχεια γίνεται η εύρεση της απόστασής δύο σημείων και ξεκινούν να ομαδοποιούνται, αυτή η διαδικασία επαναλαμβάνεται μέχρι να δημιουργηθεί μια μεγάλη κλάση. Ένα δένδrogramma μπορεί εύκολα να μας απεικονίσει την ομαδοποίηση των δεδομένων, που ξεκινά από χαμηλά σαν ξεχωριστά φύλλα και καταλήγουν σε μια κύρια 'ρίζα' [42].



Ταξινομούμε τις αποστάσεις σε αύξουσα σειρά και σύμφωνα με το γράφημα, διακρίνουμε το σημείο που σπάει και αρχίζει να ανεβαίνει απότομα [45].

Ένα πρακτικό παράδειγμα θα μπορούσε να είναι η διαδικασία πρόβλεψης φωτιάς σε δασική έκταση με την βοήθεια τοποθετημένων αισθητήρων [44] [45]:

Δεδομένα: Αισθητήρες πυρκαγιάς σε δασική έκταση. Δεδομένα εισόδου: αποστάσεις μεταξύ των αισθητήρων που αναφέρουν υψηλή θερμοκρασία. Παράμετροι: Ακτίνα  $eps=10$  μέτρα, ελάχιστοι αισθητήρες  $minPts=4$ .

1. Σάρωση Πυκνότητας: Ο αλγόριθμος εντοπίζει τις περιοχές του χάρτη όπου εμφανίζονται σήματα υψηλής θερμοκρασίας.
2. Δημιουργία Συστάδας: Εάν βρει 5 αισθητήρες (σύνολο  $> minPts=4$ ) να εκπέμπουν σήμα σε απόσταση μικρότερη των 10 μέτρων ( $eps$ ) ο ένας από τον άλλον, τους ομαδοποιεί, αναγνωρίζοντας το γεγονός ως πραγματική πιθανότητα φωτιάς.
3. Απομόνωση Θορύβου: Εάν ένας μεμονωμένος αισθητήρας δώσει σήμα, ενώ δεν υπάρχουν άλλοι 3 γείτονες σε ακτίνα 10 μέτρων, ο αλγόριθμος τον χαρακτηρίζει ως "Θόρυβο", αποτρέποντας με αυτόν τον τρόπο έναν λανθασμένο συναγερμό.

### 3.3.4 Apriori

Χρησιμοποιείται σε κατηγορικά δεδομένα ή βάση σε δυαδική μορφή, διότι δεν αναγνωρίζει αριθμητικά δεδομένα, αλλά θεωρητικά απαντάει στην ερώτηση αν το  $x$  αντικείμενο ανήκει σε μία λίστα αντικειμένων. Ο αναλυτής θα πρέπει να κάνει προ επεξεργασία των δεδομένων. Ωστόσο, αυτό δε σημαίνει πως ο χρήστης δίνει την κατεύθυνση που θα πρέπει να έχει το  $x$  αντικείμενο, δηλαδή όταν  $x$  τότε η έξοδος είναι  $y$ , όπως γίνεται στην εποπτευόμενη μάθηση. Εδώ ο αναλυτής θα πρέπει να μετατρέψει τα δεδομένα σε κατηγορίες ώστε να είναι κατανοητά από τον αλγόριθμο όταν κάνει την ομαδοποίηση [46].

Επίσης σημαντικοί είναι οι δείκτες *minimum support* και *minimum confidence*. Ο πρώτος δείκτης σηματοδοτεί την ελάχιστη εμφάνιση που πρέπει να έχει το κάθε στοιχείο ή ο κάθε συνδυασμός στοιχείων σε σχέση με τις λίστες. Για παράδειγμα ας ορίσουμε τυχαία τιμή *minimum support* = 0.6, αν ένα στοιχείο εμφανίζεται από 60% και πάνω μέσα σε όλες τις λίστες, τότε παραμένει. Ενώ όσα βρίσκονται κάτω από αυτό το όριο σταματάνε να υπάρχουν στην επόμενη ομαδοποίηση. Στη συνέχεια, δημιουργεί ομάδες των δύο για την επόμενη ποσοστιαία σύγκριση, όπου θα βρει ποιες δυάδες ικανοποιούν τη συνθήκη εμφάνισης του ορίου του 60%. Οι δυάδες που δεν ικανοποιούν πάλι το όριο δεν θα εμφανιστούν μαζί, αυτό θα επαναληφθεί μέχρι να μην ικανοποιείται άλλο η συνθήκη [47].

Έπειτα, ο *minimum confidence* είναι ο δείκτης αξιοπιστίας συσχέτισης των αντικειμένων. Αν δηλαδή το αντικείμενο  $x$  βρεθεί σε μια λίστα, τότε κατά πόσο θα βρεθεί και το αντικείμενο  $w$ . Προφανώς δε σημαίνει πως αν το αντικείμενο  $x$  υπάρχει πάντα σε μία λίστα μαζί με το  $w$ , τότε και το  $w$  θα υπάρχει πάντα με το  $x$ , αφού το  $x$  μπορεί να εμφανίζεται σε παραπάνω λίστες και συνδυασμούς από ότι το  $w$  στοιχείο [75].

Για παράδειγμα, έχουμε δυο λίστες σούπερ μάρκετ, όπου παρατηρείται ότι κάποιοι κωδικοί αγοράζονται μαζί [46] [47]:

Δεδομένα: καλάθια αγορών σε σούπερ μάρκετ. π.χ. πρώτη λίστα: Γάλα, Μήλα, Δημητριακά, δεύτερη λίστα: Δημητριακά, Κοτόπουλο, Γάλα.

1. Έλεγχος Ελάχιστης Υποστήριξης - *Minimum Support*: Ο αλγόριθμος ελέγχει όλα τα καλάθια αγορών και κρατάει μόνο τους συνδυασμούς προϊόντων που εντοπίζονται σε πάνω από το 60% των πραγματοποιημένων συναλλαγών.
2. Έλεγχος Αξιοπιστίας - *Minimum Confidence*: Στη συνέχεια, προχωράει στον λεγόμενο έλεγχο αξιοπιστίας των σχετικών συνδυασμών, έτσι ώστε να διαπιστωθεί κατά πόσο διενεργούνται σταθερά συνδυαστικές αγορές μεταξύ των δυο προϊόντων (πχ. γάλα και δημητριακά).

3. Εξαγωγή Κανόνων: Ο αλγόριθμος καταλήγει στον τελικό κανόνα: «όταν ο πελάτης αγοράζει γάλα, είναι πολύ πιθανό να αγοράσει και δημητριακά». Η εκάστοτε επιχείρηση μπορεί να εφαρμόσει τον κανόνα προκειμένου να τοποθετηθούν τα δυο αυτά προϊόντα σε κοντινά ράφια για να γίνει συνδυαστική πώληση.

### 3.3.5 Independent Component Analysis (ICA)

Ένας αρκετά ενδιαφέρον αλγόριθμος είναι ο Independent Component Analysis (ICA). Βασική απαίτηση τα δεδομένα να μην ακολουθούν κανονική κατανομή, δηλαδή η μέση τιμή να μην βρίσκεται στην κορυφή μιας συμμετρικής καμπανοειδούς κατανομής. Για να συμβεί αυτό, θα πρέπει ο χρήστης να προ-επεξεργαστεί τη βάση δεδομένων και να έχει αφαιρέσει τη μέση τιμή, εφόσον ο αλγόριθμος έχει ευαισθησία και αποδίδει λανθασμένα όταν συμβαίνει αυτό. Ο ICA ψάχνει να βρει και να απομονώσει σε ξεχωριστά δεδομένα, μια αρχική πληροφορία που μοιάζει μπλεγμένη. Ωστόσο, ο ICA δεν θα μας γνωστοποιήσει αν κάποια πηγή πρόκειται για θόρυβο. Ο αναλυτής θα πρέπει να κάνει επεξεργασία και κατά την έξοδο των δεδομένων [48].

Πρόκειται για έναν χρήσιμο αλγόριθμο μηχανικής μάθησης για να αναλύει περιπτώσεις όπως είναι το εγκεφαλογράφημα, όπου υπάρχουν πολλά εγκεφαλικά κύματα που σηματοδοτούν τις ηλεκτρικές δραστηριότητες του εγκεφάλου συν τους θορύβους όπως οι μυϊκοί σπασμοί (π.χ. το βλεφάρισμα των ματιών) [48]:

Δεδομένα: Καταγραφή ηλεκτροεγκεφαλογραφήματος, το οποίο αποτελείται από ένα σύνθετο σήμα που δεν ακολουθεί κανονική κατανομή.

1. Εισαγωγή Μικτού Σήματος: Το σήμα του ασθενούς εισάγεται στο μοντέλο το οποίο εμπεριέχει ταυτόχρονα τα πραγματικά εγκεφαλικά κύματα, αλλά και κάποιες παρεμβολές (όπως μικρούς μυϊκούς σπασμούς ή βλεφάρισμα των ματιών).
2. Μαθηματικός Διαχωρισμός: Ο αλγόριθμος αναλύει την σύνθετη πληροφορία και την ξεχωρίζει στα αρχικά, ανεξάρτητα κομμάτια από τα οποία δημιουργήθηκε.
3. Απομόνωση Στοιχείων: Ως τελικό αποτέλεσμα, το μικτό σήμα έχει διαχωριστεί σε καθαρές συνιστώσες. Ο ειδικός ιατρός μπορεί με την βοήθεια του μοντέλου να απομονώσει τον θόρυβο που δημιουργούν οι μυϊκοί σπασμοί και να προχωρήσει στην διάγνωση εξετάζοντας μόνο τα καθαρά εγκεφαλικά κύματα.

### 3.4 Μεθοδολογία και Κριτήρια Επιλογής Αλγορίθμων

Η επιλογή του καταλληλότερου αλγορίθμου αποτελεί μια δυναμική πρόκληση, καθώς η ισχύς του θεωρήματος “No Free Lunch” υπενθυμίζει διαρκώς ότι κανένα μοντέλο μηχανικής μάθησης δεν υπερέχει καθολικά [49]. Για να καταλήξει ο χρήστης στην ορθή απόφαση, οφείλει να ξεκινήσει από τη θεμελιώδη διάκριση του προβλήματος σε επιβλεπόμενη ή μη επιβλεπόμενη μάθηση.

Στο πεδίο της εποπτευόμενης μάθησης, εφόσον το ζητούμενο είναι η εκτίμηση μιας συνεχούς μεταβλητής μέσα από δεδομένα απλής δομής, η Γραμμική Παλινδρόμηση ξεχωρίζει ως η πλέον ενδεδειγμένη επιλογή, κυρίως λόγω της διαφάνειας που προσφέρει. Στον αντίποδα, στα προβλήματα ταξινόμησης, ο βαθμός πολυπλοκότητας καθορίζει το κατάλληλο εργαλείο: η Λογιστική Παλινδρόμηση προκρίνεται για τη γρήγορη και σαφή επίλυση γραμμικά διαχωρίσιμων προβλημάτων δύο κατηγοριών. Αντίθετα, σε σύνολα δεδομένων με εκτενή κειμενική πληροφορία και πολλαπλές μεταβλητές, ο αλγόριθμος Naïve Bayes αποδεικνύεται ανώτερος [50]. Αυτό οφείλεται στην αρχή της στατιστικής ανεξαρτησίας που εφαρμόζει, η οποία του εξασφαλίζει υψηλή αποδοτικότητα με ελάχιστες υπολογιστικές απαιτήσεις.

Σε περιπτώσεις όπου η δομή των δεδομένων είναι πιο σύνθετη και μη γραμμική, ο χρήστης καλείται να επιλέξει ανάλογα με το αν προέχει η κατανόηση της απόφασης ή η καθαρή ακρίβεια. Ένα δέντρο αποφάσεων προσφέρει μια απόλυτα διαφανή ιεραρχική λογική, ωστόσο το Random Forest είναι αυτό που συνήθως προτιμάται στην πράξη, καθώς μετριάζει τον κίνδυνο της υπερεκπαίδευσης συνδυάζοντας τις προβλέψεις πολλών δέντρων μαζί. Παράλληλα, αν η απόφαση βασίζεται στην έννοια της γεινιάσης και τα δεδομένα είναι σχετικά «καθαρά» και μικρά σε όγκο, ο αλγόριθμος k-NN προσφέρει μια διαισθητική λύση βασισμένη στην τοπική εγγύτητα των σημείων.

Περνώντας στο πεδίο της μη επιβλεπόμενης μάθησης, τα κριτήρια μετατοπίζονται στη γεωμετρία και την πυκνότητα των δεδομένων. Ο αλγόριθμος K-Means παραμένει η κλασική επιλογή για γρήγορη ομαδοποίηση, υπό την προϋπόθεση όμως ότι ο αναλυτής γνωρίζει τον αριθμό των ομάδων και αυτές έχουν σφαιρικό σχήμα. Αν η φύση του προβλήματος απαιτεί την εύρεση μιας ιεραρχικής οργάνωσης ή αν ο αριθμός των ομάδων είναι άγνωστος, το Hierarchical Clustering παρέχει το απαραίτητο οπτικό εργαλείο του δένδρογράμματος για τη λήψη της απόφασης. Όταν όμως τα δεδομένα περιέχουν έντονο θόρυβο ή οι ομάδες έχουν ακανόνιστο σχήμα που δεν ακολουθεί την ευκλείδεια λογική, ο DBSCAN αναδεικνύεται ως η μόνη αξιόπιστη λύση λόγω της ικανότητάς του να εντοπίζει περιοχές υψηλής πυκνότητας και να απομονώνει τις ακραίες τιμές (outliers). Τέλος, σε εξειδικευμένες εφαρμογές όπου απαιτείται η αποσύνθεση ενός σύνθετου σήματος στις αρχικές του πηγές, όπως στον διαχωρισμό ήχων ή ιατρικών σημάτων, η ICA επιλέγεται λόγω της στατιστικής της προσέγγισης στην ανεξαρτησία των συνιστωσών.

Εν κατακλείδι, η διαδικασία επιλογής δεν είναι μια γραμμική πορεία αλλά ένας συμβιβασμός μεταξύ της ακρίβειας, της υπολογιστικής ισχύος και της ανάγκης για ερμηνεία των αποτελεσμάτων. Ο χρήστης πρέπει να αξιολογεί τις υποθέσεις κάθε αλγόριθμου σε σχέση με τις ιδιαιτερότητες του dataset, καθώς η επιτυχία κρίνεται από το πόσο καλά ευθυγραμμίζονται οι εγγενείς προκαταλήψεις του μοντέλου με την πραγματική δομή της πληροφορίας.

### 3.4.1 Πολυκριτήρια Αξιολόγηση Αλγόριθμων

Όπως αναλύθηκε και στην προηγούμενη ενότητα, η επιλογή του καταλληλότερου αλγορίθμου Μηχανικής Μάθησης δεν αποτελεί μια αυθαίρετη διαδικασία, αλλά εξαρτάται άμεσα από τη φύση των δεδομένων, τον επιδιωκόμενο στόχο (πρόβλεψη, ταξινόμηση ή συσταδοποίηση) και τους διαθέσιμους υπολογιστικούς πόρους. Στο σημείο αυτό, το σκεπτικό που αναπτύχθηκε προηγουμένως βασίζεται σε μια αντικειμενική διαδικασία, όπου η οργάνωση των θεωρητικών κριτηρίων αποτυπώνεται σε ένα εφαρμόσιμο πλαίσιο λήψης αποφάσεων, προκειμένου να αποφευχθεί η άσκοπη πολυπλοκότητα και να διασφαλιστεί η ερμηνευσιμότητα των αποτελεσμάτων.

Για την αξιολόγηση των αλγορίθμων που παρουσιάστηκαν στις προηγούμενες ενότητες, θεσπίζονται τα εξής τέσσερα βασικά κριτήρια:

- **Τύπος Μάθησης:** Διαχωρισμός σε επιβλεπόμενη μάθηση, όταν υπάρχουν δεδομένα με ετικέτες ή μη επιβλεπόμενη μάθηση για αδόμητα δεδομένα.
- **Ερμηνευσιμότητα (Explainability):** Η ικανότητα του μοντέλου να αιτιολογεί τη διαδικασία λήψης απόφασης, στοιχείο κρίσιμο για κλάδους υψηλού ρίσκου όπως η ιατρική και τα χρηματοοικονομικά.
- **Υπολογιστική Ισχύς:** Οι πόροι και ο χρόνος που απαιτούνται για την εκπαίδευση του μοντέλου σε σχέση με τον όγκο των δεδομένων.
- **Ανθεκτικότητα (Robustness):** Η ικανότητα και απόδοση του αλγορίθμου να διαχειρίζεται αποτελεσματικά ακραίες τιμές (outliers) και θόρυβο στα δεδομένα εισόδου.

Παρόλο που στο Κεφάλαιο 3 εξετάστηκε ένα ευρύ φάσμα τεχνικών, για τις ανάγκες του μεθοδολογικού πλαισίου επιλέχθηκαν οι πλέον αντιπροσωπευτικοί αλγόριθμοι κάθε κατηγορίας. Η επιλογή αυτή έγινε ώστε να αναδειχθεί η θεμελιώδης λογική της λήψης αποφάσεων, καθώς οι συγκεκριμένοι αλγόριθμοι παρουσιάζουν τις πιο σαφείς διαφορές μεταξύ τους:

- **Γραμμική Παλινδρόμηση:** Αντιπροσωπεύει την πρόβλεψη συνεχών αριθμητικών τιμών.
- **k-NN:** Αντιπροσωπεύει την απλή ταξινόμηση βάσει γειτνίασης.
- **Δέντρα Απόφασης:** Αντιπροσωπεύουν την ανάγκη για απόλυτη ερμηνευσιμότητα και διαφάνεια της απόφασης.
- **K-Means:** Αντιπροσωπεύει την ομαδοποίηση σε διακριτές σφαιρικές συστάδες.

- DBSCAN: Αντιπροσωπεύει τη συσταδοποίηση με βάση την πυκνότητα και την ανίχνευση θορύβου (outliers).

Ο Πίνακας 2 που ακολουθεί συνοψίζει τη βαθμολογία της απόδοσης των παραπάνω αλγορίθμων σε μια κλίμακα από το 1 έως το 5 (όπου το 5 υποδηλώνει τη βέλτιστη επίδοση), προσφέροντας μια εποπτική σύγκριση βάσει των κριτηρίων που θεσπίστηκαν:

**Πίνακας 2: Συγκριτικός Πίνακας Πολυκριτήριας Αξιολόγησης**  
 Πηγή: Ιδία επεξεργασία βάσει βιβλιογραφικής ανασκόπησης.

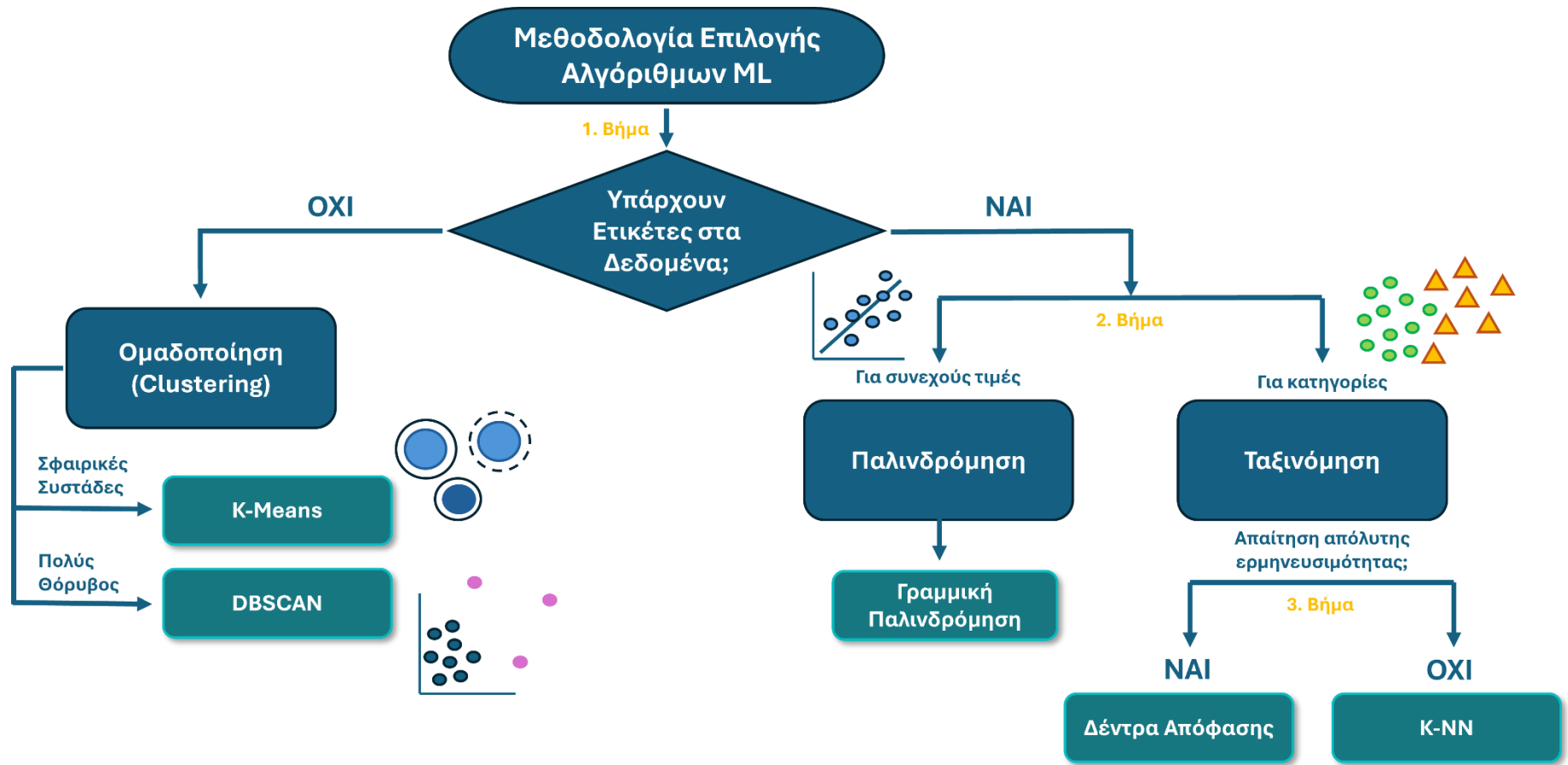
Αλγόριθμος/Κριτήριο	Τύπος Προβλήματος	Ερμηνευσιμότητα	Ταχύτητα	Ανθεκτικότητα σε θορυβο	Υπολογιστικοί Πόροι
Γραμμική Παλινδρόμηση	Παλινδρόμηση	5	5	2	1
k-NN	Ταξινόμηση	3	3	3	4
Δέντρα Απόφασης	Ταξινόμηση/ Παλινδρόμηση	5	4	4	2
K-Means	Ομαδοποίηση	2	4	2	3
DBSCAN	Ομαδοποίηση	2	2	5	5

Με βάση τα παραπάνω, γίνεται φανερό ότι κάθε αλγόριθμος έχει τα δικά του δυνατά και αδύναμα σημεία. Η τελική απόφαση δεν εξαρτάται από το ποιο μοντέλο είναι καλύτερο γενικά, αλλά από το ποιο ταιριάζει περισσότερο στις ανάγκες της εκάστοτε εφαρμογής. Για να γίνει αυτή η διαδικασία ακόμη πιο ξεκάθαρη, στην επόμενη ενότητα παρουσιάζεται το μεθοδολογικό διάγραμμα ροής, το οποίο καθοδηγεί βήμα-βήμα τον χρήστη στην επιλογή του καταλληλότερου εργαλείου.

### 3.4.2 Μεθοδολογικό Πλαίσιο και Διάγραμμα Ροής

Μετά τον καθορισμό των κριτηρίων αξιολόγησης, η παρούσα ενότητα ολοκληρώνει τη μεθοδολογική προσέγγιση, λειτουργώντας ως ένας πρακτικός οδηγός εφαρμογής. Στο Σχήμα 2 αποτυπώνεται η διαδοχή των ερωτημάτων που καλείται να απαντήσει ο αναλυτής. Η συγκεκριμένη ροή διασφαλίζει τη μετάβαση από τη θεωρία στην πράξη, οδηγώντας με ακρίβεια στην επιλογή του καταλληλότερου αλγορίθμου βάσει του εκάστοτε ερευνητικού στόχου.

Προκειμένου να αυτοματοποιηθεί η διαδικασία, προτείνεται η ακόλουθη λογική ροή, η οποία απεικονίζεται στο Σχήμα 2. Ο αναλυτής ακολουθεί μια σειρά ερωτήσεων "Ναι/Όχι" για να περιορίσει τις επιλογές του:



**Υπολογιστικοί πόροι:** Η υπολογιστική ισχύς αποτελεί ένα θεμελιώδες κριτήριο που επηρεάζει κάθε στάδιο επιλογής αλγορίθμου, από την ομαδοποίηση και την παλινδρόμηση έως την ταξινόμηση.

Σχήμα 2: Μεθοδολογικό διάγραμμα ροής για την υποστήριξη λήψης απόφασης επιλογής αλγορίθμου

Πηγή: Ιδία επεξεργασία βάσει βιβλιογραφικής ανασκόπησης

Το διάγραμμα ροής που αναπτύχθηκε οργανώνει την απόφαση σε διαδοχικά στάδια ελέγχου, μετατρέποντας τη θεωρητική ανάλυση σε τρία βασικά βήματα ερωτήσεων:

1. Έλεγχος Δεδομένων: Η ροή ξεκινά με την εξέταση της δομής του συνόλου δεδομένων. Η ύπαρξη ή η απουσία ετικετών αποτελεί τον πρώτο καθοριστικό παράγοντα που διαχωρίζει την επιβλεπόμενη από τη μη επιβλεπόμενη μάθηση. Αν δεν υπάρχουν ετικέτες στα δεδομένα, το πρόβλημα ανήκει στην ομαδοποίηση (Clustering). Εκεί προτείνεται ο K-Means για κανονικές/σφαιρικές ομάδες ή ο DBSCAN αν το σύνολο δεδομένων έχει πολύ θόρυβο.
2. Προσδιορισμός Στόχου: Στη συνέχεια, το διάγραμμα κατευθύνει τον χρήστη ανάλογα με τον τύπο της μεταβλητής που επιθυμεί να εκτιμήσει (συνεχή τιμή έναντι κατηγορίας), περιορίζοντας δραστικά το εύρος των πιθανών μοντέλων. Αν υπάρχουν ετικέτες, ελέγχουμε τι θέλουμε να προβλέψουμε. Αν ψάχνουμε μια συνεχή αριθμητική τιμή, η Γραμμική Παλινδρόμηση είναι η κατάλληλη επιλογή. Αν ψάχνουμε κατηγορία (ταξινόμηση), πάμε στο τελευταίο βήμα.
3. Ερμηνευσιμότητα: Το τελικό στάδιο επικεντρώνεται στις ειδικές απαιτήσεις του προβλήματος, όπως η ανάγκη για επεξήγηση της απόφασης ή η διαχείριση του θορύβου, επιτρέποντας στον αναλυτή να κάνει τον τελικό "συμβιβασμό" μεταξύ πολυπλοκότητας και ακρίβειας. Αν το πρόβλημα (π.χ. έγκριση δανείου ή ιατρική διάγνωση) απαιτεί απόλυτη διαφάνεια και ερμηνευσιμότητα του "γιατί" πάρθηκε η απόφαση, επιλέγονται τα Δέντρα Απόφασης. Αν αυτό δεν είναι προτεραιότητα και μας ενδιαφέρει απλώς η τοπική ανάλυση γεινίασης, επιλέγεται ο k-NN
4. Περιορισμός υπολογιστικών πόρων: Στο προτεινόμενο μοντέλο, η υπολογιστική ισχύς τοποθετείται ως κεντρικός άξονας λήψης απόφασης. Αυτό συμβαίνει διότι στην πράξη, οι διαθέσιμοι πόροι (μνήμη RAM, επεξεργαστική ισχύς) καθορίζουν τη βιωσιμότητα του μοντέλου. Για παράδειγμα, ενώ ο DBSCAN μπορεί να είναι πιο ακριβής στον εντοπισμό θορύβου, απαιτεί σημαντικά περισσότερους πόρους από τον K-Means. Συνεπώς, η μεθοδολογία που αναπτύχθηκε επιβάλλει τον έλεγχο των πόρων πριν την τελική επιλογή του αλγόριθμου.

Η χρησιμότητα του παραπάνω μεθοδολογικού διαγράμματος (Σχήμα 2) έγκειται στην ικανότητά του να μετατρέπει σύνθετες αλγοριθμικές έννοιες σε έναν ξεκάθαρο χάρτη πλοήγησης. Μέσα από αυτή την οργάνωση, η επιλογή του μοντέλου παύει να είναι μια εμπειρική διαδικασία και μετατρέπεται σε μια τεκμηριωμένη επιστημονική απόφαση. Με την ολοκλήρωση του μεθοδολογικού αυτού πλαισίου, τίθενται οι βάσεις για τη μετάβαση στο επόμενο κεφάλαιο, όπου θα εξεταστεί η πρακτική εφαρμογή της ροής αυτής σε πραγματικά σενάρια στην Υγεία, την Οικονομία και τη Βιομηχανία.

## ΚΕΦΑΛΑΙΟ 4 – ΣΥΓΧΡΟΝΕΣ ΕΦΑΡΜΟΓΕΣ ΤΗΣ ΜΗΧΑΝΙΚΗΣ ΜΑΘΗΣΗΣ

Στο παρόν κεφάλαιο εξετάζονται οι σύγχρονες εφαρμογές της Μηχανικής Μάθησης, οι οποίες έχουν επιφέρει θεμελιώδεις αλλαγές στον τρόπο λειτουργίας κρίσιμων τομέων της παγκόσμιας οικονομίας και κοινωνίας. Η μετάβαση από τον παραδοσιακό προγραμματισμό στην ικανότητα των συστημάτων να «μαθαίνουν» από τα δεδομένα, επέτρεψε την επίλυση προβλημάτων που μέχρι πρότινος θεωρούνταν υπερβολικά σύνθετα για τις ανθρώπινες δυνατότητες.

Η ανάλυση που ακολουθεί επικεντρώνεται σε τρία βασικά πεδία εφαρμογής: την Υγεία, τη Χρηματοοικονομική Ανάλυση και τη Βιομηχανία. Η επιλογή των τομέων αυτών δεν είναι τυχαία, καθώς αντιπροσωπεύουν πεδία όπου η ταχύτητα επεξεργασίας, η ακρίβεια στην πρόβλεψη και η αυτοματοποίηση των αποφάσεων έχουν άμεσο αντίκτυπο στην ανθρώπινη ζωή, τη χρηματοπιστωτική σταθερότητα και την παραγωγική αποδοτικότητα.

Μέσα από την εξέταση συγκεκριμένων παραδειγμάτων, αναδεικνύεται ο τρόπος με τον οποίο οι αλγόριθμοι που αναλύθηκαν μεθοδολογικά στο προηγούμενο κεφάλαιο, μετατρέπουν τον τεράστιο όγκο

πληροφοριών σε στρατηγική γνώση και αποτελεσματικές αποφάσεις. Ειδικότερα, εξετάζεται η πρακτική υλοποίηση των μοντέλων εποπτευόμενης και μη εποπτευόμενης μάθησης σε πραγματικά επιχειρησιακά σενάρια, επιβεβαιώνοντας τη χρησιμότητα του μεθοδολογικού πλαισίου επιλογής που προηγήθηκε.

Επομένως, το παρόν κεφάλαιο δεν αποτελεί απλώς μια παράθεση εφαρμογών, αλλά στοχεύει στην επαλήθευση του μεθοδολογικού πλαισίου που αναπτύχθηκε στο Κεφάλαιο 3. Εξετάζεται πώς τα κριτήρια επιλογής αλγορίθμων (ερμηνευσιμότητα, ανθεκτικότητα, τύπος μάθησης) εφαρμόζονται στην πράξη σε τρεις βασικούς τομείς.

#### 4.1 Εφαρμογές στον τομέα της υγείας

Στη σύγχρονη εποχή, η Μηχανική Μάθηση μπορεί να υποστηρίξει ενεργά και ουσιαστικά στο έργο του ιατρικού προσωπικού, προσφέροντας τη δυνατότητα επεξεργασίας τεράστιου όγκου σύνθετων δεδομένων που θα ήταν αδύνατο να αναλυθούν με παραδοσιακές μεθόδους. Με αυτόν τον τρόπο, η τεχνολογία εντοπίζει μοτίβα που το ανθρώπινο μάτι δύσκολα θα διέκρινε με την πρώτη ματιά. Έτσι, η ανάπτυξη της Τεχνητής Νοημοσύνης τα τελευταία χρόνια προσφέρει νέες δυνατότητες για καλύτερες θεραπείες, προσαρμοσμένες στις ανάγκες του κάθε ασθενούς. Επομένως, η εφαρμογή της Μηχανικής Μάθησης στην υγεία βασίζεται στη αξιοποίηση των διαφορετικών μεθόδων μάθησης, ανάλογα με τη φύση του εκάστοτε ιατρικού προβλήματος και τον επιδιωκόμενο στόχο.

Στον τομέα της υγείας η εφαρμογή της εποπτευόμενης μάθησης εντοπίζεται στη διαγνωστική απεικόνιση, για παράδειγμα μέσω ακτινογραφιών ή τομογραφιών. Σε αυτό το πεδίο, οι αλγόριθμοι αξιοποιούν τη γνώση από ιστορικά δεδομένα που έχουν ήδη ερμηνευθεί από ειδικούς, μαθαίνοντας να εντοπίζουν αυτόματα παθολογίες, όπως όγκους ή κατάγματα, με υψηλά ποσοστά ακρίβειας μέσα σε δευτερόλεπτα [51]. Με παρόμοιο τρόπο, η εποπτευόμενη προσέγγιση επιτρέπει τη δημιουργία «ψηφιακών αντιγράφων» οργάνων, όπως η καρδιά. Όπως επισημαίνει ο καθηγητής Olaf Dössel (2023), αναλύοντας τα βιοϊατρικά σήματα ενός συγκεκριμένου ασθενούς, η μηχανή μπορεί να προσομοιώσει τη λειτουργία του οργάνου σε ψηφιακό περιβάλλον, προσφέροντας τη δυνατότητα πρόβλεψης της εξέλιξης μιας νόσου ή της επιτυχίας μιας επέμβασης πριν από την πραγματοποίησή της. Στις κλινικές αυτές αποφάσεις, η επιλογή του αλγορίθμου καθορίζεται πρωτίστως από το κριτήριο της ερμηνευσιμότητας, όπως αυτό ορίστηκε στο μεθοδολογικό πλαίσιο της παρούσας εργασίας (Κεφάλαιο 3.4.1). Συγκεκριμένα, η χρήση των Δέντρων Απόφασης κρίνεται συχνά αναγκαία, καθώς σε κρίσιμα ζητήματα όπως είναι η υγεία, η διαφάνεια στη λήψη αποφάσεων είναι εξίσου σημαντική με την ακρίβεια

Αντίθετα, η μη εποπτευόμενη μάθηση βρίσκει εφαρμογή τόσο στην ανακάλυψη και εξέλιξη νέων φαρμάκων όσο και στην ανάλυση γενετικών δεδομένων των ασθενών. Καθώς η μέθοδος αυτή, όπως αναλύθηκε και στην παρούσα εργασία, εξειδικεύεται στον εντοπισμό κρυφών μοτίβων σε αταξινόμητα δεδομένα χωρίς ετικέτες, επιτρέπει στους ερευνητές να ομαδοποιήσουν ασθενείς με βάση άγνωστα κοινά χαρακτηριστικά. Αυτή η προσέγγιση επιταχύνει θεαματικά την εύρεση εξατομικευμένων θεραπειών για σπάνιες νόσους, μειώνοντας τόσο τον χρόνο των εργαστηριακών δοκιμών όσο και τις ενδεχόμενες παρενέργειες [52].

Πέρα από τα στάδια της διάγνωσης και της έρευνας, η Μηχανική Μάθηση επεκτείνεται και στην πρόληψη μέσω της ανάλυσης δεδομένων από συσκευές, όπως smartwatches και αισθητήρες δραστηριότητας, που παρακολουθούν συνεχώς ζωτικές λειτουργίες όπως παλμούς, οξυγόνο, ποιότητα ύπνου. Οι αλγόριθμοι μπορούν να αναλύσουν αυτά τα δεδομένα σε πραγματικό χρόνο, εντοπίζοντας έγκαιρα ενδείξεις που θα μπορούσαν να προμηνύουν ένα καρδιακό επεισόδιο. Παράλληλα παρατηρείται πρόοδος και στην ρομποτική χειρουργική, την αποκατάσταση και την ψυχοθεραπεία. Στον τομέα της ψυχοθεραπείας, η τεχνολογία βοηθά στην ανάλυση του λόγου και της συμπεριφοράς, εντοπίζοντας πρώιμα σημάδια κατάθλιψης, ενισχύοντας τον ρόλο του θεραπευτή χωρίς ωστόσο να τον αντικαθιστά [53].

Βασική πρόκληση στην εφαρμογή των τεχνολογιών που αναφέρθηκαν παραπάνω αποτελούν και εδώ τα κρίσιμα ηθικά ζητήματα σχετικά με την προστασία των προσωπικών δεδομένων. Για τη διασφάλιση του ιατρικού απορρήτου, μια σύγχρονη μέθοδος αποτελεί το λεγόμενο Federated Learning. Με την βοήθεια αυτής της μεθόδου, η εκπαίδευση των μοντέλων γίνεται τοπικά μέσα στο κάθε νοσοκομείο. Έτσι, τα νοσοκομεία μοιράζονται μεταξύ τους μόνο τα συμπεράσματα που έβγαλε ο αλγόριθμος και όχι τα ονόματα ή τα στοιχεία των ασθενών [54]. Με αυτόν τον τρόπο, η ιατρική μπορεί να εξελισσεται τεχνολογικά, διασφαλίζοντας παράλληλα το ιατρικό απόρρητο και την ιδιωτικότητα των ασθενών, όπως ακριβώς ορίζουν οι διεθνείς οδηγίες [55] & [56].

Συνοψίζοντας, η Μηχανική Μάθηση λειτουργεί ως ένα εργαλείο που βοηθάει το έργο του γιατρού και βελτιώνει συνδυαστικά τα αποτελέσματα για τον ασθενή. Με τη βοήθειά της, η διαδικασία της διάγνωσης γίνεται με μεγαλύτερη ακρίβεια, προσφέροντας λύσεις που είναι προσαρμοσμένες στις ιδιαίτερες ανάγκες που έχει ο κάθε ασθενής.

## 4.2 Εφαρμογές στη χρηματοοικονομική ανάλυση

Στον χρηματοοικονομικό τομέα, η Μηχανική Μάθηση έχει φέρει σημαντική βελτίωση των ροών και διαδικασιών, καθώς οι τράπεζες και οι επενδυτικοί οργανισμοί καλούνται να διαχειριστούν καθημερινά έναν τεράστιο όγκο δεδομένων. Σε αυτό το περιβάλλον, η ταχύτητα και η ακρίβεια στην επεξεργασία των δεδομένων αποτελούν βασικές προϋποθέσεις για τη σωστή λειτουργία της αγοράς. Η τεχνολογία επιτρέπει πλέον στα χρηματοπιστωτικά ιδρύματα να προσεγγίζουν τις ανάγκες των πελατών τους με μεγαλύτερη ασφάλεια, μειώνοντας ταυτόχρονα τα λάθη που θα μπορούσε να κάνει ο ανθρώπινος παράγοντας [57]. Οι οικονομικοί οργανισμοί αξιοποιούν πλέον την Τεχνητή Νοημοσύνη για να βελτιώσουν τις εσωτερικές τους λειτουργίες, να προσφέρουν μια πιο εξατομικευμένη εμπειρία στους πελάτες τους και να τους προστατέψουν απέναντι σε κινδύνους και ενδεχόμενες απάτες [58].

Με την συνεχή εξέλιξη της τεχνολογίας, οι τράπεζες μπορούν πλέον να ελέγχουν τις συναλλαγές με αυτοματοποιημένο τρόπο και σε πραγματικό χρόνο, προσφέροντας παράλληλα υπηρεσίες που ταιριάζουν στις ανάγκες του κάθε πελάτη. Κεντρικό ρόλο σε αυτή τη διαδικασία παίζει η εποπτευόμενη μάθηση, η οποία χρησιμοποιείται στον χρηματοοικονομικό τομέα κυρίως για τον εντοπισμό της απάτης. Οι αλγόριθμοι εποπτευόμενης μάθησης εκπαιδεύονται να αναλύουν χιλιάδες συναλλαγές σε πραγματικό χρόνο, μαθαίνοντας τις συνήθειες του κάθε χρήστη, ώστε σε περίπτωση που εντοπιστεί μια ασυνήθιστη δραστηριότητα, η οποία ξεφεύγει από τη συνηθισμένη συμπεριφορά του χρήστη, να μπορούν να την μπλοκάρουν άμεσα, χωρίς να κινδυνεύσουν τα κεφάλαιά του [58] & [59]. Με αυτόν τον τρόπο, οι τράπεζες δεν προστατεύουν μόνο τα δικά τους κεφάλαια, αλλά διασφαλίζουν και την εμπιστοσύνη των πελατών τους.

Παράλληλα, η εφαρμογή μοντέλων εποπτευόμενης μάθησης παίζει καθοριστικό ρόλο και στην αξιολόγηση του πιστωτικού κινδύνου (credit scoring). Αντικαθιστώντας τις παλιές μεθόδους με πιο ακριβή μοντέλα, όπως τα Δέντρα Απόφασης και η Γραμμική Παλινδρόμηση, οι αλγόριθμοι αυτοί έχουν πλέον τη δυνατότητα να αξιολογήσουν το ιστορικό και τη συμπεριφορά των δανειοληπτών για να προβλέψουν με μεγαλύτερη ακρίβεια την πιθανότητα αποπληρωμής [60]. Αυτή η προσέγγιση καθιστά τη διαδικασία των δανείων πιο δίκαιη και γρήγορη, προστατεύοντας ταυτόχρονα το χρηματοπιστωτικό σύστημα από υψηλό πιστωτικό κίνδυνο [60]. Επιπλέον, εφαρμόζονται συχνά οι μέθοδοι Ensemble, οι οποίες συνδυάζουν πολλαπλά μοντέλα για να μειώσουν την πιθανότητα σφάλματος και να βελτιώσουν τη συνολική απόδοση της πρόβλεψης [58].

Στον τομέα των επενδύσεων, η εφαρμογή της εποπτευόμενης μάθησης έχει επεκταθεί σημαντικά μέσω των αλγοριθμικών συναλλαγών (algorithmic trading). Χρησιμοποιώντας μοντέλα όπως η γραμμική παλινδρόμηση και τα νευρωνικά δίκτυα, τα συστήματα αυτά αναλύουν σε πραγματικό χρόνο ιστορικά δεδομένα της αγοράς και οικονομικές ειδήσεις [61]. Αυτή η ανάλυση επιτρέπει την πρόβλεψη βραχυπρόθεσμων τάσεων στις τιμές των μετοχών, υποστηρίζοντας ενεργά τη λήψη αποδοτικών αποφάσεων σε ένα εξαιρετικά μεταβαλλόμενο περιβάλλον. Παράλληλα, οι αλγόριθμοι αυτοί μπορούν να εκτελούν εντολές αγοράς ή πώλησης με αυτοματοποιημένο τρόπο, επιτυγχάνοντας ταχύτητα και ακρίβεια που ξεπερνά κατά πολύ τις ανθρώπινες δυνατότητες και τις παραδοσιακές στατιστικές μεθόδους. Με αυτόν τον τρόπο, οι επενδυτικοί οργανισμοί καταφέρνουν να μεγιστοποιούν την απόδοση των κεφαλαίων τους, αντιδρώντας άμεσα στις διακυμάνσεις της παγκόσμιας αγοράς [61].

Εκτός από την εποπτευόμενη μάθηση, ο χρηματοπιστωτικός τομέας αξιοποιεί σε μεγάλο βαθμό και την μη εποπτευόμενη μάθηση. Η βασική διαφορά εδώ είναι ότι το μοντέλο δεν χρειάζεται έτοιμα παραδείγματα για να μάθει, αλλά αναλύει τεράστιους όγκους δεδομένων για να βρει μοτίβα και συσχετίσεις. Αυτό σημαίνει ότι χωρίς να χρειάζονται προκαθορισμένες ετικέτες σε μη δομημένα δεδομένα, οι αλγόριθμοι μπορούν να ομαδοποιήσουν τους χρήστες με βάση κοινά χαρακτηριστικά στη συμπεριφορά τους, όπως οι αποταμιευτικές τους συνήθειες ή οι επενδυτικές τους προτιμήσεις [58] & [62]. Αυτό είναι επίσης πολύ χρήσιμο για τις τράπεζες, διότι τις βοηθάει να προτείνουν εξατομικευμένα προϊόντα που ταιριάζουν πραγματικά στις ανάγκες του κάθε πελάτη, βελτιώνοντας με αυτόν τον τρόπο την συνολική του εξυπηρέτηση.

Μια βασική εφαρμογή της μη εποπτευόμενης μάθησης αποτελεί ο εντοπισμός ανωμαλιών στις συναλλαγές (anomaly detection). Σύμφωνα με τον Woźniak [62], οι αλγόριθμοι ομαδοποίησης επιτρέπουν στο μοντέλο να αναγνωρίζει κινήσεις που δεν έχουν ξανασυμβεί. Αυτό είναι ιδιαίτερα χρήσιμο λόγω του ότι το σύστημα δεν ψάχνει για συγκεκριμένα είδη απάτης, αλλά ομαδοποιεί τις συναλλαγές και επισημαίνει οτιδήποτε αποκλίνει από συνηθισμένες συμπεριφορές χρηστών. Με αυτόν τον τρόπο μπορεί να βρει νέες μορφές απάτης που δεν έχουν ξανασυμβεί στο παρελθόν, καθώς δεν βασίζεται σε παλιά παραδείγματα αλλά στην αναγνώριση του «παράξενου» [62]. Στο σημείο αυτό, ο αλγόριθμος DBSCAN αποδεικνύεται ο πλέον κατάλληλος, επιβεβαιώνοντας τη θεωρητική του ανάλυση ως εργαλείο εντοπισμού θορύβου. Όπως αναδείχθηκε στη μεθοδολογία που αναπτύχθηκε στο Κεφάλαιο 3, η ικανότητα του DBSCAN να απομονώνει σημεία που αποκλίνουν από την κανονική πυκνότητα των συναλλαγών, επιτρέπει την αναγνώριση νέων μορφών απάτης που δεν βασίζονται σε ιστορικά δεδομένα.

Εν κατακλείδι, ο ρόλος της Μηχανικής Μάθησης στην χρηματοοικονομική ανάλυση είναι πλέον καθοριστικός, αποτελώντας ένα απαραίτητο εργαλείο. Δεν περιορίζεται μόνο στην αυτοματοποίηση των διαδικασιών, αλλά ενισχύει την ασφάλεια των συναλλαγών και την αξιοπιστία των οικονομικών προβλέψεων. Με αυτόν τον τρόπο, δημιουργείται ένα πιο σταθερό και προστατευμένο οικονομικό περιβάλλον, τόσο για τους οργανισμούς όσο και για τους ίδιους τους καταναλωτές.

### 4.3 Εφαρμογές στη βιομηχανία και τις έξυπνες τεχνολογίες

Στην εποχή της τέταρτης βιομηχανικής επανάστασης (Industry 4.0), η Μηχανική Μάθηση αλλάζει ριζικά τον τρόπο λειτουργίας των παραγωγικών διαδικασιών. Τα παλιά συστήματα, που εκτελούσαν απλώς τυποποιημένες εντολές, αντικαθίστανται από "έξυπνα" δίκτυα που μπορούν να παίρνουν αποφάσεις μόνα τους. Πλέον, οι αλγόριθμοι δεν ακολουθούν μόνο ένα πρόγραμμα, αλλά βελτιώνουν την παραγωγή σε πραγματικό χρόνο, την ώρα που αυτή εξελίσσεται [63]. Αυτή η αλλαγή βοηθά τις επιχειρήσεις να προσαρμόζονται αμέσως στις νέες συνθήκες, μετατρέποντας τα δεδομένα σε χρήσιμες πληροφορίες για τον έλεγχο των αποθεμάτων και την πρόβλεψη των πωλήσεων [63].

Η αποτελεσματικότητα αυτών των συστημάτων βασίζεται στη χρήση διαφορετικών μεθόδων μάθησης ανάλογα με τον στόχο. Μία από τις πιο σημαντικές εφαρμογές, η οποία βασίζεται στη μη εποπτευόμενη μάθηση, είναι η προγνωστική συντήρηση (predictive maintenance) [64]. Αντί οι εταιρείες να περιμένουν να χαλάσει μια μηχανή ή να την επισκευάζουν βάσει προκαθορισμένου χρονοδιαγράμματος, χρησιμοποιούν αισθητήρες που παρακολουθούν μεταξύ άλλων παραμέτρους όπως η θερμοκρασία και οι υπέρηχοι, εξοικονομώντας με αυτόν τον τρόπο χρόνο και πόρους [63]. Οι αλγόριθμοι ομαδοποιούν τα δεδομένα της φυσιολογικής λειτουργίας και είναι σε θέση να αξιολογήσουν πότε ένα εξάρτημα εμφανίζει ασυνήθιστη συμπεριφορά, προειδοποιώντας για μια βλάβη πριν αυτή συμβεί. Με αυτόν τον τρόπο, οι επιχειρήσεις παρεμβαίνουν μόνο όταν χρειάζεται, γλιτώνοντας χρόνο και μειώνοντας δραστικά το κόστος των επισκευών [64]. Η χρήση της μη εποπτευόμενης μάθησης σε αυτό το πεδίο υλοποιεί στην πράξη τη στρατηγική επιλογή που αναπτύχθηκε στο Σχήμα 2 της μεθοδολογίας επιλογής των κατάλληλων αλγόριθμων αναλόγως τον στόχο της ανάλυσης. Εκεί, η απουσία ετικετών στα δεδομένα πραγματικού χρόνου οδηγεί στην ανάγκη αυτόματης ομαδοποίησης, αναδεικνύοντας την ανθεκτικότητα του μοντέλου ως κρίσιμη παράμετρο για την αποφυγή βλαβών [65].

Παράλληλα, η εποπτευόμενη μάθηση βελτιώνει τον ποιοτικό έλεγχο με παραδείγματα ελλειψών και σωστών προϊόντων. Αυτό πραγματοποιείται με την ενσωμάτωση εξελιγμένων καμερών που θα καταγράφουν λεπτομερώς τα χαρακτηριστικά ενός άρτιου προϊόντος και θα αναγνωρίζουν άμεσα μια απόκλιση. Σε αντίστοιχη περίπτωση ο ανθρώπινος παράγοντας θα αδυνατούσε να ανακαλύψει με γυμνό μάτι μικρές ατέλειες δίνοντας έτσι μεγάλο προβάδισμα στη μηχανική μάθηση. Το αποτέλεσμα, εκτός της διασφάλισης των σωστών προδιαγραφών των προϊόντων, έχει αντίκτυπο και σε άλλους τομείς [66].

Η εξοικονόμηση χρόνου αλλά και η οικολογικότερη διαδικασία παραγωγής προσθέτουν ακόμα περισσότερο στην αξία των έξυπνων αλγορίθμων. Έχουν τη δυνατότητα να εξελίσσουν και να βελτιστοποιούν τις διαδικασίες, ώστε να αποφεύγουν σπατάλες πρώτων υλών και να αναγνωρίζουν με ποιον τρόπο θα καταναλώνουν λιγότερη ενέργεια. Συμβάλλοντας και προάγοντας σε ένα πιο βιώσιμο μοντέλο παραγωγής. Με αυτή την προσέγγιση, η βιομηχανία γίνεται συνολικά πιο αποδοτική και μπορεί να διαχειρίζεται καλύτερα τον

τεράστιο όγκο πληροφοριών που παράγει καθημερινά, ενισχύοντας την ανταγωνιστικότητά της στο σύγχρονο ψηφιακό περιβάλλον [66].

## **ΚΕΦΑΛΑΙΟ 5 – ΣΥΜΠΕΡΑΣΜΑΤΑ ΚΑΙ ΜΕΛΛΟΝΤΙΚΕΣ ΚΑΤΕΥΘΥΝΣΕΙΣ**

### **5.1 Συνοπτικά συμπεράσματα**

Στην παρούσα εργασία αναδείχθηκε η χρησιμότητα της Μηχανικής Μάθησης σε σύγχρονες παραγωγικές διαδικασίες, καθώς αποτελεί ένα δυναμικό μεθοδολογικό εργαλείο και όχι απλά μια συλλογή αλγορίθμων χωρίς ουσία. Σαφώς η ποιότητα των δεδομένων και η ιδανική επιλογή αλγορίθμου βάσει διαφόρων κριτηρίων, όπως η ερμηνευσιμότητα και η διαχείριση των θορύβων, καθιστούν βασικές προϋποθέσεις για την επιτυχία ενός μοντέλου, όπου διαπιστώθηκε από τη θεωρητική ανάλυση και τα εκπαιδευτικά παραδείγματα που τέθηκαν.

Εν κατακλείδι αντιλαμβανόμαστε πως η Μηχανική Μάθηση αποτελεί τον πυρήνα του σύγχρονου ψηφιακού κόσμου, στην τρέχουσα διπλωματική εργασία μελετήθηκαν αλγόριθμοι τόσο στην εποπτευόμενη όσο και στη μη εποπτευόμενη μάθηση. Η κύρια διαπίστωση είναι πως δεν υπάρχει ο τέλειος αλγόριθμος που να μπορεί να ταιριάξει παντού, κάτι το οποίο αναφέρθηκε και στο θεώρημα του “No Free Lunch”. Η πραγματική επιτυχία προέρχεται από την αντίληψη του χρήστη να καταλάβει τα δεδομένα και με ποιον τρόπο μπορεί να τα αξιοποιήσει σωστά, πάντα με την επίγνωση ότι θα πρέπει να υπάρξει και συμβιβασμός. Ο συμβιβασμός προκύπτει διότι ο αναλυτής σε κάθε πρόβλημα θα πρέπει να ερευνήσει τα εξής τρία πράγματα, το πόσο ακριβές θέλει να είναι το αποτέλεσμα, πόση είναι η υπολογιστική δύναμη αλλά και πόση ανάγκη έχει να εξηγήσει την απόφαση που πάρθηκε.

Αντιλαμβανόμαστε λοιπόν πως αυτή η μαθηματική θεωρία βρίσκει πραγματική αξία για την κοινωνία από τις σύγχρονες εφαρμογές που εξετάσαμε. Στον τομέα της Υγείας μπορούν να σωθούν ζωές από τα ευρήματα των αλγορίθμων, τα οποία είναι ικανά να αποτελέσουν αναλύσεις που θα βοηθήσουν τους γιατρούς να δουν λεπτομέρειες σε εξετάσεις που θα αδυνατούσαν να διακρίνουν με γυμνό μάτι. Στον κλάδο των χρηματοοικονομικών να εξασφαλίσουν μεγαλύτερη αξιοπιστία στον έλεγχο και να προστατεύουν τις συναλλαγές μας από απάτες σε πραγματικό χρόνο. Καθώς και στη βιομηχανία μπορεί να συντελέσει εργαλείο για να προβλεφθεί η πιθανή “ημερομηνία λήξης” ενός μηχανήματος, εξοικονομώντας χρόνο και τεράστια κόστη. Οπότε το τελικό συμπέρασμα δεν είναι απλά ότι η Μηχανική Μάθηση είναι ένα πολύτιμο εργαλείο στα χέρια μας, αλλά πως θα εφαρμοστεί με ασφάλεια και συνέπεια.

### **5.2 Συμβολή της εργασίας**

Η συμβολή της παρούσας διπλωματικής εργασίας έγκειται στην ανάπτυξη ενός συστηματικού πλαισίου επιλογής αλγορίθμων, σύμφωνα πάντα με τη φύση του προβλήματος που καλείται ο χρήστης να δώσει, ανάμεσα από τις διαφορετικές κατηγορίες μάθησης. Ο απώτερος σκοπός των ενσωματωμένων παραδειγμάτων είναι η μαθηματική θεωρία να μετατραπεί σε πρακτικά αποτελέσματα μέσα από σύγχρονα προβλήματα του πραγματικού κόσμου, προσφέροντας μια ολοκληρωμένη εικόνα για τη χρήση του εκάστοτε μοντέλου.

Το σημαντικότερο κομμάτι της εργασίας ωστόσο δεν είναι αυτό της θεωρητικής αναφοράς των αλγορίθμων, αλλά ότι πραγματοποιήσαμε ένα εργαλείο λήψης αποφάσεων. Τις περισσότερες φορές η επιλογή ενός αλγορίθμου γίνεται μέσω πειραματικών δοκιμών και κρίνοντας την καταλληλότητα από το πόσο λανθασμένο ή σωστό είναι το αποτέλεσμα κατά την έξοδο. Το διάγραμμα ροής καλείται να βάλει τάξη μέσα από μια αλληλουχία κριτηρίων που ως κατάληξη έχει τον καταλληλότερο αλγόριθμο.

Για να δημιουργήσουμε το διάγραμμα ροής χρειάστηκε ένας πίνακας αξιολόγησης για να συγκρίνουμε τα μοντέλα μας με βάση την ταχύτητα τους, την αντοχή τους σε θορύβους και την ερμηνευσιμότητα των αποφάσεων τους. Το διάγραμμα έχει ως σκοπό να καθοδηγήσει πλήρως τον χρήστη στην καλύτερη δυνατή επιλογή

αλγόριθμοι βάση λογικής, εφόσον είναι ικανός να απαντήσει σε κριτήρια όπως "έχεις ετικέτες στα δεδομένα σου;" ή "θέλεις να κάνεις πρόβλεψη τιμής ή κατηγορίας;"

Επι πρόσθετα, η εργασία δεν επικεντρώνεται σε μαθηματικούς τύπους και εξισώσεις για να εξηγήσει τους αλγόριθμους όπως άλλα κλασικά ακαδημαϊκά κείμενα. Αντί αυτού δόθηκε μια πιο θεωρητική προσέγγιση ώστε να μπορεί να εντάξει έναν αναλυτή στον κόσμο της Μηχανικής Μάθησης δίνοντας μικρά καθημερινά παραδείγματα. Αποτυπώθηκε βήμα-βήμα πώς εγκρίνεται ένα δάνειο, πώς φιλτράρεται ένα spam email ή πώς εντοπίζεται μια φωτιά στο δάσος. Η παραπάνω προσέγγιση μεταξύ θωρίας και πράξης συντελεί στη δημιουργία ενός εύχρηστου εγχειριδίου που μπορεί να βοηθήσει κάποιον στα πρώτα του βήματα στον χώρο της Επιστήμης των Δεδομένων.

### 5.3 Μελλοντικές ερευνητικές κατευθύνσεις

Η Τεχνητή Νοημοσύνη και κατά επέκταση η Μηχανική Μάθηση εξελίσσεται με ταχύτατους ρυθμούς, με τεράστιους όγκους δομημένων και αδόμητων δεδομένων να δημιουργούνται καθημερινά. Παρόλο που στην παρούσα εργασία καλύφθηκε ένα ευρύ φάσμα του πεδίου της Μηχανικής Μάθησης και αναπτύχθηκε ένα πρακτικό μεθοδολογικό πλαίσιο επιλογής αλγόριθμων, υπάρχουν ακόμα πολλές ανοιχτές προκλήσεις. Με βάση τα συμπεράσματα και τους περιορισμούς που αναδείχθηκαν, οι ερευνητές καλούνται τα επόμενα χρόνια να εστιάσουν σε τομείς που συνδυάζουν τις ήδη γνωστές ανάγκες με νέες και καινοτόμες τεχνολογικές δυνατότητες:

- **Επεξηγήσιμη Τεχνητή Νοημοσύνη (XAI) και Αυτοματοποίηση (AutoML):** Όπως τονίστηκε επανειλημμένα, οι χρήστες δυσκολεύονται να εμπιστευτούν ένα μοντέλο αν δεν κατανοούν πλήρως τον τρόπο που καταλήγει σε ένα συμπέρασμα. Η περαιτέρω έρευνα σε εργαλεία Επεξηγήσιμης Τεχνητής Νοημοσύνης είναι απολύτως απαραίτητη τη διασφάλιση της αλγοριθμικής διαφάνειας και την πλήρη κατανόηση της διαδικασίας λήψης απόφασης. Με αυτόν τον τρόπο θα μπορέσουν να αιτιολογηθούν με σαφήνεια διάφορες ιατρικές ή/και νομικές αποφάσεις [67]. Παράλληλα, το μεθοδολογικό διάγραμμα ροής που προτάθηκε στην παρούσα εργασία, θέτει τις βάσεις για την πλήρη αυτοματοποίηση στην επιλογή των καταλληλότερων αλγόριθμων. Σε αντίθεση με τα σημερινά συστήματα αυτοματοποιημένης Μηχανικής Μάθησης (AutoML), τα οποία αναζητούν σχεδόν αποκλειστικά τη μέγιστη προγνωστική ακρίβεια, η παρούσα εργασία φέρνει στο προσκήνιο την ανάγκη για μια πιο ολοκληρωμένη οπτική. Η μελλοντική επιστημονική έρευνα οφείλει να αξιοποιήσει το μεθοδολογικό πλαίσιο που αναπτύχθηκε, ενσωματώνοντάς το σε συστήματα που θα αντιλαμβάνονται τους διαθέσιμους πόρους. Δηλαδή το σύστημα δεν πρέπει απλώς να αναζητά το υψηλότερο ποσοστό επιτυχίας, αλλά να συνεκτιμά τόσο τους διαθέσιμους υπολογιστικούς πόρους (μνήμη, επεξεργαστική ισχύ και ενέργεια) όσο και την ανάγκη για διαφάνεια και ερμηνευσιμότητα αναλόγως το πεδίο εφαρμογής και τα διαθέσιμα δεδομένα. Με αυτόν τον τρόπο θα απορρίπτονται μοντέλα που, παρά την ακρίβειά τους, κρίνονται μη βιώσιμα για ένα συγκεκριμένο περιβάλλον υλοποίησης, ενώ θα είναι απόλυτα ασφαλής για ευαίσθητους τομείς [68].
- **Πράσινη Τεχνητή Νοημοσύνη (Green AI) και Κβαντική Μηχανική Μάθηση (Quantum ML):** Η επιτακτική ανάγκη για εξοικονόμηση πόρων φέρνει στο προσκήνιο και το υψηλό περιβαλλοντικό κόστος που απαιτεί η εκπαίδευση των αλγόριθμων. Για να καταστεί η Μηχανική Μάθηση βιώσιμη, το μελλοντικό στοίχημα είναι διπλό: Αφενός η δημιουργία «πράσινων» μοντέλων οι οποίοι θα προσφέρουν υψηλή απόδοση με ελάχιστη επεξεργαστική ισχύ [69]. Αφετέρου, η μετάβαση στην Κβαντική Μηχανική Μάθηση, η οποία θα εξαλείψει τους σημερινούς περιορισμούς μνήμης, καθώς τα κβαντικά συστήματα θα μπορούν να εκτελούν απαιτητικούς αλγόριθμους (όπως ο K-Means) σε τεράστια σύνολα δεδομένων, επιλύοντας εξαιρετικά πολύπλοκα προβλήματα μέσα σε λίγα δευτερόλεπτα [70].
- **Ομόσπονδη Μάθηση (Federated Learning) και Ανταγωνιστική Ασφάλεια (Adversarial ML):** Η προστασία των προσωπικών δεδομένων παραμένει ως κορυφαία πρόκληση στο πεδίο της Μηχανικής Μάθησης. Η Ομόσπονδη Μάθηση (Federated Learning) προσφέρει τη λύση μέσω της τοπικής εκπαίδευσης των μοντέλων, αποτρέποντας την αποστολή ευαίσθητων δεδομένων σε κεντρικούς διακομιστές [71]. Παράλληλα, προκύπτει η επιτακτική ανάγκη για Ανταγωνιστική Ασφάλεια (Adversarial ML). Στο μεθοδολογικό πλαίσιο που αναπτύχθηκε, τέθηκε η ανθεκτικότητα (robustness) απέναντι στον

θόρυβο ως ένα από τα πιο βασικά κριτήρια επιλογής. Έτσι λοιπόν, η μελλοντική έρευνα οφείλει να εστιάσει στις κακόβουλες παρεμβάσεις. Επομένως, η θωράκιση των αλγορίθμων απέναντι σε χάκερ που αλλοιώνουν σκόπιμα τα δεδομένα για να παραπλανήσουν κρίσιμες ιατρικές ή τραπεζικές αποφάσεις, αποτελεί πλέον αδιαπραγμάτευτη προϋπόθεση [72].

- **Μεταφορά Γνώσης (Transfer Learning) και Διαρκής Εκπαίδευση (Continuous Learning):** Σύμφωνα με τη θεμελιώδη αρχή του θεωρήματος "No Free Lunch", δεν υφίσταται ένα ιδανικό αλγοριθμικό μοντέλο για κάθε πιθανό πρόβλημα. Προκειμένου να αντιμετωπιστεί αυτό, η έρευνα επικεντρώνεται στη Μεταφορά Γνώσης, ώστε ένας αλγόριθμος να μπορεί να πάρει την υπάρχουσα γνώση από έναν τομέα και να την προσαρμόσει άμεσα σε έναν άλλο, με ελάχιστα νέα δεδομένα, βελτιστοποιώντας τη χρήση των υπολογιστικών και οικονομικών πόρων [73]. Παράλληλα, τα μελλοντικά συστήματα θα πρέπει να υποστηρίζουν και τη Διαρκή Εκπαίδευση (Continuous Learning). Όπως διαπιστώθηκε στο Κεφάλαιο 4, αναλύοντας τις σύγχρονες εφαρμογές στη Βιομηχανία και τις χρηματιστηριακές αγορές, ο πραγματικός κόσμος αλλάζει ασταμάτητα προσθέτοντας διαρκώς νέες προκλήσεις. Αυτό σημαίνει πρακτικά ότι αντί η εκπαίδευσή των μοντέλων να είναι στατική, θα πρέπει να μαθαίνουν διαρκώς από νέα δεδομένα σε πραγματικό χρόνο, χωρίς να διαγραφεί η γνώση που έχει αποκτηθεί στο παρελθόν [74].

#### ΠΙΝΑΚΑΣ ΟΡΟΛΟΓΙΑΣ

Ξενόγλωσσος όρος	Ελληνικός Όρος
Algorithmic Bias	Αλγοριθμική Μεροληψία
Algorithmic Trading	Αλγοριθμικές Συναλλαγές
Anomaly Detection	Εντοπισμός Ανωμαλιών
Black Box	Μαύρο Κουτί (Μη διαφανές μοντέλο)
Classification	Ταξινόμηση
Clustering	Συσταδοποίηση / Ομαδοποίηση
Continuous Learning	Διαρκής Εκπαίδευση
Credit Scoring	Αξιολόγηση Πιστωτικού Κινδύνου
Decision Flowchart	Διάγραμμα Ροής (Αποφάσεων)
Deep Learning	Βαθιά Μάθηση
Explainability / Interpretability	Ερμηνευσιμότητα / Αναγνωσιμότητα

Features	Χαρακτηριστικά (Δεδομένων)
Federated Learning	Ομόσπονδη Μάθηση
Generalization	Γενίκευση (Μοντέλου)
Green AI	Πράσινη Τεχνητή Νοημοσύνη
Labels	Ετικέτες (Δεδομένων)
Machine Learning	Μηχανική Μάθηση
Outliers	Ακραίες Τιμές / Θόρυβος
Overfitting	Υπερπροσαρμογή
Predictive Maintenance	Προγνωστική Συντήρηση
Quantum ML	Κβαντική Μηχανική Μάθηση
Regression	Παλινδρόμηση
Reinforcement Learning	Ενισχυτική Μάθηση
Underfitting	Υποπροσαρμογή
Unsupervised Learning	Μη Εποπτευόμενη Μάθηση
Validation Set	Σύνολο Επαλήθευσης
Robustness	Ανθεκτικότητα
Semi-supervised Learning	Ημι-εποπτευόμενη Μάθηση
Supervised Learning	Εποπτευόμενη Μάθηση
Test Set	Σύνολο Ελέγχου / Δοκιμής
Training Set	Σύνολο Εκπαίδευσης
Transfer Learning	Μεταφορά Γνώσης

## ΣΥΝΤΜΗΣΕΙΣ – ΑΡΚΤΙΚΟΛΕΞΑ – ΑΚΡΩΝΥΜΙΑ

AI	Artificial Intelligence (Τεχνητή Νοημοσύνη)
AutoML	Automated Machine Learning (Αυτοματοποιημένη Μηχανική Μάθηση)
DBSCAN	Density-Based Spatial Clustering of Applications with Noise
eps	Epsilon (Ακτίνα γειτονιάς στον αλγόριθμο DBSCAN)
ICA	Independent Component Analysis (Ανάλυση Ανεξάρτητων Συνιστωσών)
k-NN	k-Nearest Neighbors (k-Κοντινότεροι Γείτονες)
minPts	Minimum Points (Ελάχιστος αριθμός σημείων στον αλγόριθμο DBSCAN)
PCA	Principal Component Analysis (Ανάλυση Κύριων Συνιστωσών)

RF	Random Forest (Τυχαίο Δάσος)
SOMs	Self-Organizing Maps (Αυτοοργανωνόμενοι Χάρτες)
XAI	Explainable Artificial Intelligence (Επεξηγήσιμη Τεχνητή Νοημοσύνη)
ΔΑ	Δέντρα Αποφάσεων (Decision Trees)

**ΒΙΒΛΙΟΓΡΑΦΙΚΕΣ ΑΝΑΦΟΡΕΣ**

- [1] Manyika, J., et al. (2011). Big data: The next frontier for innovation, competition, and productivity. McKinsey Global Institute.
- [2] Mitchell, T. M. (1997). Machine Learning. McGraw-Hill. Ανακτήθηκε από: <https://www.cs.cmu.edu/~tom/files/MachineLearningTomMitchell.pdf>
- [3] Bishop, C. M. (2006). Pattern Recognition and Machine Learning. Springer. Ανακτήθηκε από: [crowley-coutaz.fr/jlc/Courses/2020/PRML/ENSI3.PRML.S6.Encoders.pdf](http://crowley-coutaz.fr/jlc/Courses/2020/PRML/ENSI3.PRML.S6.Encoders.pdf)
- [4] Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. Science, 349(6245), 255-260.
- [5] Müller, A. C., & Guido, S. (2016). Introduction to Machine Learning with Python: A Guide for Data Scientists. O'Reilly Media, Inc. Ανακτήθηκε από: [https://www.nrigroupindia.com/e-book/Introduction%20to%20Machine%20Learning%20with%20Python%20\(%20PDFDrive.com%20\)-min.pdf](https://www.nrigroupindia.com/e-book/Introduction%20to%20Machine%20Learning%20with%20Python%20(%20PDFDrive.com%20)-min.pdf)
- [6] Domingos, P. (2012). A few useful things to know about machine learning. Communications of the ACM, 55(10), 78–87. Ανακτήθηκε από: <https://doi.org/10.1145/2347736.2347755>  
<https://homes.cs.washington.edu/~pedrod/papers/cacm12.pdf>
- [7] Floridi, L., Cows, J., Beltrametti, M., Chatila, R., Chiriatti, G., Dignum, V., Duffy, B., Fawcett, B., Garcia, P., Giaccardi, E., Morley, J., Olsson, M., Rivard, L., Rossi, F., Schafer, B., Tsamados, A., Varsavsky, M., & Vayena, E. (2018). AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations. Minds and Machines, 28(4), 689–707. Ανακτήθηκε από: <https://doi.org/10.1007/s11023-018-9482-5>  
[https://pmc.ncbi.nlm.nih.gov/articles/PMC6404626/pdf/11023\\_2018\\_Article\\_9482.pdf](https://pmc.ncbi.nlm.nih.gov/articles/PMC6404626/pdf/11023_2018_Article_9482.pdf)
- [8] Baer, Tobias. (2020). How to Overcome Algorithmic Bias
- [9] Lei, Gao & Ling Guan. (2022). Interpretability of Machine Learning: Recent Advances and Future Prospects. Published by the IEEE Computer Society. Ανακτήθηκε από: <https://arxiv.org/pdf/2305.00537>
- [10] Lipton, Zachary. (2018). The Mythos of Model Interpretability: In machine learning, the concept of interpretability is both important and slippery.. Queue. 16. 10.1145/3236386.3241340
- [11] Adadi, A., & Berrada, M. (2018). Peeking inside the black-box: a survey on explainable artificial intelligence (XAI). IEEE access, 6, 52138-52160. Ανακτήθηκε από: <https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8466590>
- [12] Molnar, C. (2022). Interpretable Machine Learning: A Guide for Making Black Box Models Explainable.
- [13] Géron, A. (2019). Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems (2nd ed.). O'Reilly Media. Ανακτήθηκε από: <https://www.rasa-ai.com/wp-content/uploads/2022/02/Aur%C3%A9lien-G%C3%A9ron-Hands-On-Machine-Learning-with-Scikit-Learn-Keras-and-Tensorflow-Concepts-Tools-and-Techniques-to-Build-Intelligent-Systems-O%E2%80%99Reilly-Media-2019.pdf>
- [14] Pinker, Alexander. (2021). Maschinelles Lernen. Hanns-Seidel-Stiftung. Ανακτήθηκε από: [https://www.hss.de/fileadmin/user\\_upload/Factsheet\\_Maschinelles\\_Lernen.pdf](https://www.hss.de/fileadmin/user_upload/Factsheet_Maschinelles_Lernen.pdf)
- [15] Murphy, K. P. (2012). Machine Learning: A Probabilistic Perspective. MIT Press.
- [16] Hastie, T., Tibshirani, R., & Friedman, J. (2009). The Elements of Statistical Learning: Data Mining, Inference, and Prediction (2nd ed.). Springer Science & Business Media. Ανακτήθηκε από: <https://doi.org/10.1007/978-0-387-84858-7>
- [17] Turing, A. M. (1950). Computing machinery and intelligence. Mind, 59(236), 433-460.
- [18] Rosenblatt, F. (1958) The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain. Psychological Review, 65, 386. Ανακτήθηκε από: <https://doi.org/10.1037/h0042519>
- [19] Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. Nature, 323(6088), 533-536. Ανακτήθηκε από: <https://doi.org/10.1038/323533a0>
- [20] LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436-444. Ανακτήθηκε από: <https://doi.org/10.1038/nature14539>
- [21] Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press. Ανακτήθηκε από: <http://www.deeplearningbook.org>

- [22] Han, J., Pei, J., & Tong, H. (2022). Data Mining: Concepts and Techniques. Morgan Kaufmann.
- [23] Alpaydin, E. (2020). Introduction to Machine Learning. MIT Press.
- [24] Hellevik, Ottar. (2007) "Linear versus logistic regression when the dependent variable is a dichotomy" Ανακτήθηκε από: [Linear-versus-logistic-regression-when-the-dependent-variable-is-a-dichotomy.pdf](#)
- [25] Baillya,b, Alexandre, Corentin Blanc, Élie Francis, Thierry Guillotina , Fadi Jamal, Béchara Wakimd, Pascal Roy. (2022) "Effects of dataset size and interactions on the prediction performance of logistic regression and deep learning models" Ανακτήθηκε από: [Effects of dataset size and interactions on the prediction performance of logistic regression and deep learning models - ScienceDirect](#)
- [26] Smola Alex & S.V.N. Vishwanathan. (2008) "Introduction to Machine Learning"
- [27] Rajib Kumar Halder<sup>1</sup> , Mohammed Nasir Uddin<sup>1</sup> , Md. Ashraf Uddin<sup>2\*</sup> , Sunil Aryal<sup>2</sup> and Ansam Khraisat<sup>2</sup>. (2024) "Enhancing K-nearest neighbor algorithm: a comprehensive review and performance analysis of modifications" [Enhancing K-nearest neighbor algorithm: a comprehensive review and performance analysis of modifications](#)
- [28] Jijo Bahzad Taha & Abdulazeez Adnan Mohsin. (2021) "Classification Based on Decision Tree Algorithm for Machine Learning" Ανακτήθηκε από: [yaseenmustafa,+Classification+Based+on+DTA+\(1\).pdf](#).
- [29] Ren Qiong, Hui Cheng & Hai Han. (2017) "Research on machine learning framework based on random forest algorithm" Ανακτήθηκε από: [080020\\_1\\_online.pdf](#)
- [30] Zhang, Harry. (2004) "The Optimality of Naive Bayes" Ανακτήθηκε από: [The Optimality of Naive Bayes](#)
- [31] Sarker Iqbal H. & Muhammad Ashad Kabir & Alan Colman & Jun Han. (2017) "An Improved Naive Bayes Classifier-based Noise Detection Technique for Classifying User Phone Call Behavior" Ανακτήθηκε από: [1710.04461](#)
- [32] Raschka, S. (2017). Machine Learning mit Python: Das Praxis-Handbuch für Data Science und Predictive Analytics und Deep Learning. mitp Verlag. Ανακτήθηκε από: [https://www.mitp.de/out/media/9783958454224\\_Leseprobe.pdf](https://www.mitp.de/out/media/9783958454224_Leseprobe.pdf)
- [33] Raschka, S. (2014) "Naive Bayes and Text Classification I" Ανακτήθηκε από: [1410.5329](#)
- [34] Raschka, S. (2016) "Model evaluation, model selection, and algorithm selection in machine learning" <https://sebastianraschka.com/blog/2016/model-evaluation-selection-part3.html>
- [35] Dutt, S., Chandramouli, S., & Das, A. K. (2019). Machine learning. ISBN 978-93-530-6669-7, eISBN: 978-93-895-8813-2
- [36] Y C a, Padmanabha & Pulabaigari, Viswanath & B, Eswara. (2018). Semi-supervised learning: a brief review. International Journal of Engineering & Technology. 7. 81. 10.14419/ijet.v7i1.8.9977.
- [37] De Vries, H. (2025). Maschinelles Lernen. Masterstudiengänge Informatics and Business (Wirtschaftsinformatik) & Wirtschaftsingenieurwesen. Fachhochschule Südwestfalen. Ανακτήθηκε από: [https://www.fh-swf.de/media/neu\\_np/fb\\_tbw\\_1/dozentinnen\\_2/professorinnen\\_5/devries\\_1/Maschinelles-Lernen.pdf](https://www.fh-swf.de/media/neu_np/fb_tbw_1/dozentinnen_2/professorinnen_5/devries_1/Maschinelles-Lernen.pdf)
- [38] Döbel Inga et. al. (2018), MASCHINELLES LERNEN EINE ANALYSE ZU KOMPETENZEN, FORSCHUNG UND ANWENDUNG, Fraunhofer-Gesellschaft zur Förderung der angewandten Forschung e.V.
- [39] Géron, A. (2022), Hands-On Machine Learning with Scikit-Learn, Keras & TensorFlow (3rd ed.). O'Reilly.
- [40] Vankayalapati Revathi, Ghutugade Kalyani Balaso, Vannapuram Rekha, Prasanna Bejjanki Pooja Sree. (2021) "K-Means Algorithm for Clustering of Learners Performance Levels Using Machine Learning Techniques" Ανακτήθηκε από: [51932-libre.pdf](#)
- [41] Sinaga Kristina P. & Yang Miin-Shen. (2020) "Unsupervised K-Means Clustering Algorithm" Ανακτήθηκε από: [IEEE Xplore Full-Text PDF:](#)
- [42] Fionn Murtagh, Pedro Contreras. (2012) "Algorithms for hierarchical clustering: an overview" <https://www.dei.unipd.it/~capri/DM/MATERIALE/SurveyHierarchicalClustAlgs.pdf>
- [43] Fionn Murtagh, Pedro Contreras. (2012) "Algorithms for hierarchical clustering: an overview II" [https://eprints.hud.ac.uk/id/eprint/32552/1/DWD\\_Fmurtagh\\_31.pdf](https://eprints.hud.ac.uk/id/eprint/32552/1/DWD_Fmurtagh_31.pdf)
- [44] Sahu, Nagendra Kumar. (2022) "DBSCAN & HIERARCHICAL CLUSTERING ALGORITHM: ANALYSIS CYBER CRIME DATA" Ανακτήθηκε από: [Microsoft Word - 4](#)
- [45] Starczewski<sup>1</sup>, Artur, Piotr Goetzen<sup>2</sup>, Meng Joo Er. (2020) "A NEW METHOD FOR AUTOMATIC DETERMINING OF THE DBSCAN PARAMETERS" Ανακτήθηκε από: [jaiscr-2020-0014.pdf](#)
- [46] Jasdeep Singh Malik, Prachi Goyal, Mr.Akhilesh K Sharma, Rajendra Nagar Indore .(2010) "A Comprehensive Approach Towards Data Preprocessing Techniques & Association Rules" [https://bvicam.in/INDIACom/news/INDIACom%202010%20Proceedings/papers/Group3/INDIACom10\\_2\\_79\\_Paper%20\(2\).pdf](https://bvicam.in/INDIACom/news/INDIACom%202010%20Proceedings/papers/Group3/INDIACom10_2_79_Paper%20(2).pdf)

- [47] R. Uday Kiran, P. Krishna Re. (2009) “An improved multiple minimum support based approach to mine rare association rules” [https://www.researchgate.net/profile/Polepalli-Krishna-Reddy/publication/221312121\\_An\\_Improved\\_Multiple\\_Minimum\\_Support\\_Based\\_Approach\\_to\\_Mine\\_Rare\\_Association\\_Rules/links/0c9605260ddd96cb92000000/An-Improved-Multiple-Minimum-Support-Based-Approach-to-Mine-Rare-Association-Rules.pdf](https://www.researchgate.net/profile/Polepalli-Krishna-Reddy/publication/221312121_An_Improved_Multiple_Minimum_Support_Based_Approach_to_Mine_Rare_Association_Rules/links/0c9605260ddd96cb92000000/An-Improved-Multiple-Minimum-Support-Based-Approach-to-Mine-Rare-Association-Rules.pdf)
- [48] Pranali Kokate, Sidharth Pancholi, Amit M. Joshi (2021) “Classification of Upper Arm Movements from EEG signals using Machine Learning with ICA Analysis” <https://arxiv.org/pdf/2107.08514>
- [49] Tom F. Sterkenburg · Peter D. Grünwald . (2021) “The no-free-lunch theorems of supervised learning” <https://link.springer.com/content/pdf/10.1007/s11229-021-03233-1.pdf>
- [50] Lars Kotthoff, Ian P. Gent and Ian Miguel. (2012) “An Evaluation of Machine Learning in Algorithm Selection for Search Problems” [https://www.cs.uwyo.edu/~larsko/papers/kotthoff\\_evaluation\\_2012.pdf](https://www.cs.uwyo.edu/~larsko/papers/kotthoff_evaluation_2012.pdf)
- [51] MDPI Electronics. (2025). Deep Learning Architectures for Advanced Medical Imaging: A Comprehensive Review.
- [52] Nature Medicine. (2025). The Impact of Generative AI on Precision Oncology and Drug Development.
- [53] Possard, M., & Zankl, W. (2024). Artificial intelligence in medicine and psychotherapy: A concise and interdisciplinary overview. Sigmund Freud Private University Vienna.
- [54] Journal of Medical Informatics (JMIR). (2025). Federated Learning in Healthcare: Privacy-Preserving AI Implementation. Vol. 12, Issue 2.
- [55] European Medicines Agency (EMA). (2024). Reflection paper on the use of Artificial Intelligence (AI) in the medicinal product lifecycle.
- [56] World Health Organization (WHO). (2025). Ethics and governance of artificial intelligence for health: Guidance on large multi-modal models. <https://www.who.int/publications/i/item/9789240084759>
- [57] Walch Kathleen, (2024). How AI Is Transforming The Finance Industry. Forbes
- [58] Dubey, A., Khan, M. S., & Tyagi, H. (2024). From Data to Decisions: The Role of AI and Machine Learning in Modern Financial Markets. International Journal of Advanced Research in Science, Communication and Technology (IJARSCT), Volume 4, Issue 3, December 2024.
- [59] Boeddu, G., Feyen, E., Martinez Jaramillo, S., Mesquita, S., Palta, Y., Sarkar, A., Sinha, S., & Gutiérrez Traverso, A. (2025). Artificial Intelligence for Financial Sector Supervision: An Emerging Market and Developing Economies Perspective. World Bank.
- [60] Hariharan, Pappil Kothandapani. (2022). Leveraging AI for credit scoring and financial inclusion in emerging markets. World Journal of Advanced Research and Reviews. 15. 526-539. 10.30574/wjarr.2022.15.3.0904.
- [61] Bhuiyan MD Shahriar Mahmud, MD AL Rafi, Gourab Nicholas Rodrigues, MD Nazmul Hossain Mir, Adit Ishraq, M.F. Mridha, Jungpil Shin, (2025). Deep learning for algorithmic trading: A systematic review of predictive models and optimization strategies, Array, Volume 26, 2025, 100390, ISSN 2590-0056, <https://doi.org/10.1016/j.array.2025.100390>. (<https://www.sciencedirect.com/science/article/pii/S2590005625000177>)
- [62] Woźniak, R. J. (2023). Unsupervised machine learning in financial anomaly detection: clustering algorithms vs. dedicated methods. Institute of Information Systems, Faculty of Cybernetics, MUT. Warsaw, Poland.
- [63] Marco Antonio Diaz Martinez, Reina Veronica Roman Salinas, Santos Ruiz Hernandez, Gabriela Cervantes Zubirias, Mario Alberto Morales Rodriguez. (2024) [https://www.researchgate.net/profile/Gabriela-Cervantes-Zubirias/publication/386339952\\_Industry\\_40\\_Applications\\_of\\_machine\\_learning\\_in\\_the\\_field\\_of\\_industrial\\_engineering\\_Systematic\\_review\\_of\\_the\\_literature/links/674dea8ff309a268c01e513a/Industry-40-Applications-of-machine-learning-in-the-field-of-industrial-engineering-Systematic-review-of-the-literature.pdf](https://www.researchgate.net/profile/Gabriela-Cervantes-Zubirias/publication/386339952_Industry_40_Applications_of_machine_learning_in_the_field_of_industrial_engineering_Systematic_review_of_the_literature/links/674dea8ff309a268c01e513a/Industry-40-Applications-of-machine-learning-in-the-field-of-industrial-engineering-Systematic-review-of-the-literature.pdf)
- [64] Neural Concept. (2026). “Predictive Maintenance Machine Learning: A Practical Guide” <https://www.neuralconcept.com/post/how-ai-is-used-in-predictive-maintenance>
- [65] MUHAMMAD USAMA, JUNAID QADIR, AUNN RAZA, HUNAIN ARIF, KOK-LIM ALVIN YAU, YEHIA ELKHATIB, AMIR HUSSAIN<sup>5,6</sup>, AND ALA AL-FUQAHA. (2019). “Unsupervised Machine Learning for Networking: Techniques, Applications and Research Challenges” <https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=8713992>
- [66] Czimmermann, T., Ciuti, G., Milazzo, M., Chiurazzi, M., Roccella, S., Oddo, C. M., & Dario, P. (2020). Visual inspection indicators for training the next generation of manufacturing professionals. Applied Sciences, 10(7), 2443.

- [67] Alejandro Barredo Arrieta, Natalia Díaz-Rodríguez, Javier Del Ser, Adrien Bennetot, Siham Tabik, Alberto Barbado, Salvador Garcia, Sergio Gil-Lopez, Daniel Molina, Richard Benjamins, Raja Chatila, Francisco Herrera, (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI, *Information Fusion*, Volume 58, 2020, Pages 82-115, ISSN 1566-2535, <https://doi.org/10.1016/j.inffus.2019.12.012>.  
(<https://www.sciencedirect.com/science/article/pii/S1566253519308103>)
- [68] Hutter, F., Kotthoff, L., & Vanschoren, J. (Eds.). (2019). *Automated Machine Learning: Methods, Systems, Challenges*. Springer Nature. <https://doi.org/10.1007/978-3-030-05318-5> [https://www.automl.org/wp-content/uploads/2019/05/AutoML\\_Book.pdf](https://www.automl.org/wp-content/uploads/2019/05/AutoML_Book.pdf)
- [69] Schwartz, R., Dodge, J., Smith, N. A., & Etzioni, O. (2020). *Green AI*. *Communications of the ACM*, 63(12), 54–63. <https://dl.acm.org/doi/epdf/10.1145/3381831>
- [70] Biamonte, J., Wittek, P., Nicola, P., et al. (2017). Quantum Machine Learning. *Nature*, 549, 195–202.
- [71] McMahan, B., Moore, E., Ramage, D., Hampson, S., & y Arcas, B. A. (2017). Communication-efficient learning of deep networks from decentralized data. In *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (AISTATS)*. <https://proceedings.mlr.press/v54/mcmahan17a.html>
- [72] Biggio, B., & Roli, F. (2018). Wild patterns: Ten years after the rise of adversarial machine learning. *Pattern Recognition*, 84, 317-331.
- [73] Pan, S. J., & Yang, Q. (2010). A Survey on Transfer Learning. *IEEE Transactions on Knowledge and Data Engineering*, 22(10), 1345-1359. <https://ieeexplore.ieee.org/document/5288526>
- [74] Chen, Z., & Liu, B. (2016). *Lifelong Machine Learning*. Morgan & Claypool Publishers. <https://www.cs.uic.edu/~liub/lifelong-machine-learning-draft.pdf>
- [75] Radhiatul Husna, Riri Lestari, Yomei Hendra, (2019). “Inventory model of goods availability with apriori algorithm”
- [76] Dössel, O. (2023). *Computational Modeling of the Heart and Machine Learning in Medicine*. Berlin-Brandenburg Academy of Sciences and Humanities (BBAW).