



ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΙΓΑΙΟΥ
ΠΟΛΥΤΕΧΝΙΚΗ ΣΧΟΛΗ

ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ ΠΛΗΡΟΦΟΡΙΑΚΩΝ ΚΑΙ ΕΠΙΚΟΙΝΩΝΙΑΚΩΝ ΣΥΣΤΗΜΑΤΩΝ

ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ

Internet of Things, Διαδίκτυο των Πραγμάτων

Προβλεπτική Συντήρηση σε Συστήματα Βιομηχανικού
Διαδικτύου Πραγμάτων μέσω Μηχανικής Μάθησης

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

του

Παύλου Ιωάννη

Επιβλέπων : Σκούτας Δημήτριος, Επ. Καθηγητής

Μέλη εξεταστικής επιτροπής: Κορμέντζας Γεώργιος, Καθηγητής
Σκιάνης Χαράλαμπος, Καθηγητής

Σάμος, 11/2024

Πρόλογος και ευχαριστίες

Η παρούσα Διπλωματική εργασία εστιάζει στην ανάλυση των σύγχρονων τεχνολογιών του Διαδικτύου των Πραγμάτων (IoT) και της επίδρασής τους στην Βιομηχανία του Μέλλοντος. Εκπονήθηκε στα πλαίσια του μεταπτυχιακού προγράμματος σπουδών με τίτλο «Internet of Things, Διαδίκτυο των Πραγμάτων» του τμήματος Μηχανικών Πληροφοριακών και Επικοινωνιακών Συστημάτων του Πανεπιστημίου Αιγαίου. Θα ήθελα να ευχαριστήσω θερμά τον επιβλέποντα καθηγητή μου κ. Σκούτα Δημήτριο για την καθοδήγηση και την πολύτιμη βοήθειά του, καθώς και τον κ. Νομικό Νικόλαο που πρόσφερε χρήσιμες πληροφορίες όπως και ερευνητικά παραδείγματα.

Ευχαριστώ επίσης τη μητέρα μου και τη γυναίκα μου, για την αμέριστη υποστήριξη που μου παρείχανε καθ' όλη τη διάρκεια του μεταπτυχιακού, αλλά και τη κόρη μου, Χριστίνα που ήταν η πηγή της δύναμης μου.

Εύχομαι να μεγαλώσεις και να διαβάσω τη δικιά σου Διπλωματική και να αισθάνομαι το ίδιο περήφανος όπως τώρα που ολοκληρώνω εγώ τη δικιά μου.

© [2024]

του

[ΠΑΥΛΟΥ ΙΩΑΝΝΗ]

Τμήμα Μηχανικών Πληροφοριακών και Επικοινωνιακών Συστημάτων
ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΙΓΑΙΟΥ

Πίνακας περιεχομένων

1	Εισαγωγή	1
1.1	Ο Ρόλος της Τεχνολογίας IIoT και της Μηχανικής Μάθησης στη Σύγχρονη Βιομηχανία	1
1.2	Αντικείμενο διπλωματικής.....	2
1.3	Δομή της διπλωματικής	3
2	Προβλεπτική Συντήρηση Βιομηχανικού Εξοπλισμού.....	4
2.1	Η εξέλιξη της βιομηχανικής συντήρησης	4
2.2	Προ-ενεργητική συντήρηση (Proactive Maintenance)	5
2.3	Η προληπτική συντήρηση στην Industry 4.0	7
2.4	Προγνωστική και Διαχείριση Υγείας.....	8
2.5	Πρόγνωση Σφάλματος	10
3	Χαρακτηριστικά πλατφόρμας για την αξιοποίηση της Μηχανικής Μάθησης	14
3.1	Προκλήσεις για την ενσωμάτωση του ΔτΠ στην Βιομηχανία.....	14
3.2	Κύρια Χαρακτηριστικά Πλατφόρμας ΔτΠ	18
3.3	Επιλεγόμενη Πλατφόρμα: PI System.....	23
4	Μηχανισμοί και Τεχνικές για την Εκπαίδευση Αλγορίθμων Μηχανικής Μάθησης	28
4.1	Περιγραφή Dataset.....	28
4.2	Προ-επεξεργασία Δεδομένων (Data Pre-processing)	32
4.2.1	Εντοπισμός ακραίων τιμών (Outlier Detection).....	32
4.2.2	Επιλογή Χαρακτηριστικών (Feature Selection).....	34
4.2.3	Κανονικοποίηση Δεδομένων (Data Normalization).....	35
4.2.4	Ανάλυση Κυρίων Συνιστωσών (Principal Component Analysis – PCA)	35
4.2.5	Ανισορροπία κλάσης (Class Imbalance)	37
4.3	Τεχνικές Μηχανικής Μάθησης	39
4.3.1	Λογική Παλινδρόμηση (Logistic Regression – LR)	39
4.3.2	Μηχανές Διανυσμάτων Υποστήριξης (Support Vector Machines – SVM).....	41
4.3.3	Δέντρα Απόφασης (Decision Trees)	43
4.3.4	Τυχαίο Δάσος - Random Forest	44
4.3.5	Αλγόριθμος k-Nearest Neighbors (k-NN).....	46
5	Ανάπτυξη μοντέλου Μηχανικής Μάθησης για τον εντοπισμό πιθανών αστοχιών και κατηγοριοποίησης σφαλμάτων	48

5.1	Μεθοδολογία Επεξεργασίας Δεδομένων	48
5.2	Παρουσίαση Αποτελεσμάτων	50
5.3	Ροή εργασιών Έξυπνης διαχείρισης Ανεμογεννητριών	55
5.4	Συμπεράσματα	58
Βιβλιογραφία		59
Παράρτημα Ι – Αποτέλεσμα Μοντέλου Πρόβλεψης		62

Λίστα Εικόνων

Εικόνα 1. Εξέλιξη των μεθόδων συντήρησης, Πηγή: [40].	5
Εικόνα 2. Τα λειτουργικά επίπεδα της Συντήρησης βάσει κατάστασης, Πηγή: [40].	6
Εικόνα 3. Βασικά στοιχεία της Industry 4.0, Πηγή: [8].	8
Εικόνα 4. Συσχέτιση αξίας και τεχνικών συντήρησης, Πηγή: [41].	9
Εικόνα 5. Μοντέλο PHM, Πηγή: [11].	9
Εικόνα 6. Προσδιορισμός του RUL μετά τον εντοπισμό του σφάλματος, Πηγή: [42].	11
Εικόνα 7. Κατηγορίες Προγνωστικών προσεγγίσεων, Πηγή: [11].	12
Εικόνα 8. Χαρακτηριστικά Πλατφόρμας IIoT, Πηγή: [43].	19
Εικόνα 9. Αρχιτεκτονική PI System, Πηγή: [44].	23
Εικόνα 10. Οι κατηγορίες και τα στοιχεία ενός τυπικού PI System, Πηγή: [44].	24
Εικόνα 11. Χρονική Διάρκεια καταγραφής δεδομένων των υποσυνόλων Fault Data και Scada Data	29
Εικόνα 12. Στατιστικά Συνόλου Δεδομένων Scada Data	30
Εικόνα 13. Συνολικός αριθμός καταγεγραμμένων σφαλμάτων ανά κατηγορία	31
Εικόνα 14. Εντοπισμός ακραίων τιμών	32
Εικόνα 15. Συνολικός αριθμός καταγεγραμμένων σφαλμάτων ανά κατηγορία	33
Εικόνα 16. Sigmoid Συνάρτηση, Πηγή: [34].	39
Εικόνα 17. One-vs-Rest Πολυωνυμική παλινδρόμηση, Πηγή: [34].	40
Εικόνα 18. Μηχανές Διανυσμάτων Υποστήριξης, Πηγή: [45].	41
Εικόνα 19. Ταξινόμηση μη γραμμικά διαχωρίσιμων κλάσεων σε χώρο υψηλότερων διαστάσεων, Πηγή: [35].	42
Εικόνα 20. Ταξινόμηση μέσω Δέντρου Απόφασης. (Σταματάτος Ευστάθιος, σημειώσεις μαθήματος Μηχανικής Μάθησης, πανεπιστήμιο Αιγαίου [2022])	43
Εικόνα 21. Τυχαίο Δάσος, Πηγή: [46].	45
Εικόνα 22. Παράδειγμα K-NN Ταξινομητή, Πηγή: [39].	46
Εικόνα 23. Μεθοδολογία επεξεργασίας δεδομένων και εκπαίδευσης μοντέλων Μηχανικής Μάθησης	50
Εικόνα 24. Αποτελέσματα Random Forest στο test set	52
Εικόνα 25. Προβολή Heatmap	53
Εικόνα 26. Προβολή Confusion Matrix και Heatmap 2 κλάσεων	54
Εικόνα 27. Συγκεντρωτική Δομή AF	55
Εικόνα 28. Event Frames για Fault Detection	56
Εικόνα 29. Αυτοματοποιημένο μήνυμα ενημέρωσης σφάλματος εξοπλισμού	56

Εικόνα 30. Ροή Εργασίας του Συστήματος	57
---	----

Λίστα Πινάκων

Πίνακας 1. Απόδοση μοντέλων MM με την χρήση PCA και χωρίς αυτή.....	51
Πίνακας 2. Μετρικές απόδοσης των διαφορετικών μοντέλων MM.....	52

.

Ακρωνύμια

AF	Asset Framework
AI	Artificial Intelligence
API	Application Programming Interface
CBM	Condition Based Maintenance
DA	Data Archive
IIoT	Industrial Internet of Things
IT	Information Technology
KPI	Key Performance Indicators
LR	Logistic Regression
NN	Neural Networks
PCA	Principal Component Analysis
PHM	Prognostics and Health Management
RF	Random Forest
RUL	Remain Useful Life
SCADA	Supervisory Control and Data Acquisition
SMOTE	Synthetic Minority Oversampling Technique
SVM	Support Vector Machine
OT	Operation Technology

Περίληψη

Η παρούσα διπλωματική εστιάζει στην πρόβλεψη της λειτουργικής κατάστασης μιας ανεμογεννήτριας συλλέγοντας πληροφορίες από συσκευές Βιομηχανικού Διαδικτύου των Πραγμάτων (IIoT) και αξιοποιώντας εκπαιδευμένα μοντέλα Μηχανικής Μάθησης. Η σύνθεση αυτών των τεχνολογιών προσφέρει σημαντικά πλεονεκτήματα στη βιομηχανία επιτρέποντας την προληπτική συντήρηση και τη βελτιστοποίηση των πόρων με αποτέλεσμα τη μείωση του κόστους και την αύξηση της παραγωγικότητας. Ο στόχος της εργασίας είναι να αξιοποιήσει με τον βέλτιστο δυνατό τρόπο τα δεδομένα από το *Dataset: "IIoT Wind Turbine Analytics,"* το οποίο περιλαμβάνει πληροφορίες από αισθητήρες IIoT τα οποία προέρχονται μια ανεμογεννήτρια άγνωστης ταυτότητας. Ακόμα, προτείνεται μια συστηματική μεθοδολογία προ-επεξεργασίας που προετοιμάζει τα δεδομένα για την εκπαίδευση των μοντέλων, με τους αλγορίθμους LR, SVM, Decision Trees, Random Forest και K-NN να αποτελούν τη βασική εργαλειοθήκη της ανάλυσης.

Παράλληλα, η εργασία αναπτύσσει μια ολοκληρωμένη λύση αξιοποιώντας τις δυνατότητες του *PI System* για τη δημιουργία ενός ολοκληρωμένου συστήματος διαχείρισης. Το σύστημα αυτό συλλέγει τα δεδομένα από τις IIoT συσκευές και τα επεξεργάζεται μέσω του μοντέλου Μηχανικής Μάθησης για την πρόβλεψη της κατάστασης της ανεμογεννήτριας. Τα αποτελέσματα της πρόβλεψης επιστρέφουν στο *PI System*, το οποίο με την σειρά του ειδοποιεί σε πραγματικό χρόνο τους χειριστές για πιθανές δυσλειτουργίες ή ανάγκες συντήρησης, επιτρέποντας έτσι τη λήψη έγκαιρων αποφάσεων και ενισχύοντας την ασφάλεια και αποδοτικότητα της λειτουργίας.

Λέξεις Κλειδιά: IIoT, Μηχανική Μάθηση, PI System, Προληπτική Συντήρηση

Abstract

This thesis focuses on predicting the operational status of a wind turbine by collecting information from Industrial Internet of Things (*IIoT*) devices and utilizing trained Machine Learning models. The integration of these technologies provides significant advantages to the industry enabling preventive maintenance and resource optimization, which result in cost reduction and productivity improvement. The objective of this study is to make optimal use of data from the "*IIoT Wind Turbine Analytics*" dataset, which contains sensor information from an unidentified wind turbine's *IIoT* sensors. Additionally, a systematic preprocessing methodology is proposed to prepare the data for model training, with LR, SVM, Decision Trees, Random Forest, and K-NN serving as the primary analysis tools.

Furthermore, the thesis develops an integrated solution by leveraging the capabilities of the PI System to create a comprehensive management system. This system collects data from *IIoT* devices and processes it through the Machine Learning model to predict the turbine's condition. The prediction results are fed back into the PI System, which in turn notifies operators in real-time of any potential malfunctions or maintenance needs, allowing timely decision-making and enhancing the safety and efficiency of operations.

Keywords: *IIoT, Machine Learning, PI System, Predictive Maintenance*

1

Εισαγωγή

1.1 Ο Ρόλος της Τεχνολογίας IIoT και της Μηχανικής Μάθησης στη Σύγχρονη Βιομηχανία

Στη σύγχρονη εποχή, όπου ο ανταγωνισμός είναι έντονος και οι απαιτήσεις διαρκώς αυξάνονται, οι επιχειρήσεις καλούνται να μειώσουν τα λειτουργικά τους κόστη, να αυξήσουν τη διαθεσιμότητα των συστημάτων τους και να περιορίσουν τις απρόβλεπτες διακοπές λειτουργίας. Αυτές οι προκλήσεις αποτελούν κρίσιμα σημεία για τη διατήρηση ενός συγκριτικού πλεονεκτήματος και την εξασφάλιση μακροπρόθεσμης βιωσιμότητας.

Η ανάπτυξη των τεχνολογιών του Βιομηχανικού Διαδικτύου των Πραγμάτων (IIoT) έχει φέρει σημαντικές αλλαγές στον τρόπο που οι επιχειρήσεις διαχειρίζονται τα συστήματά τους. Πλέον, μέσω των συσκευών IIoT μπορούν να εγκαθιστούν αισθητήρες με μικρότερο κόστος ακόμα και στα πιο απομακρυσμένα και επικίνδυνα σημεία, καταγράφοντας με αξιοπιστία δεδομένα και στοιχεία λειτουργίας των μηχανών τους σε πραγματικό χρόνο. Αυτά τα δεδομένα προσφέρουν μια ολοκληρωμένη εικόνα της κατάστασης του εξοπλισμού, επιτρέποντας στις επιχειρήσεις να λαμβάνουν εμπεριστατωμένες αποφάσεις και να βελτιστοποιούν τις λειτουργίες τους.

Παράλληλα, τα δεδομένα αυτά μπορούν να αξιοποιηθούν για την εκπαίδευση μοντέλων μηχανικής μάθησης τα οποία έχουν τη δυνατότητα να εντοπίζουν και να προβλέπουν δυσλειτουργίες στον εξοπλισμό. Με την προγνωστική συντήρηση να καθίσταται εφικτή, οι επιχειρήσεις μπορούν να προλαμβάνουν προβλήματα προτού αυτά κλιμακωθούν, εξασφαλίζοντας μεγαλύτερη αποδοτικότητα και διάρκεια ζωής για τον εξοπλισμό τους.

Επιπλέον, η δυνατότητα προβλέψεων για την κατάσταση των συσκευών με βάση τα δεδομένα που συλλέγονται από το IIoT φέρνει νέες προοπτικές. Μέσω της εφαρμογής προηγμένων

αλγορίθμων μηχανικής μάθησης οι εταιρείες μπορούν να εξάγουν πολύτιμες πληροφορίες για την υγεία και τη φθορά των συστημάτων τους και να μειώνουν δραστικά τις ανάγκες για έκτακτες επισκευές αλλά και για προγραμματισμένες προληπτικές συντηρήσεις.

Ωστόσο, η ενσωμάτωση αυτών των τεχνολογιών θέτει και ορισμένες προκλήσεις, όπως η επεξεργασία και διαχείριση μεγάλου όγκου δεδομένων καθώς και η εξασφάλιση της ακρίβειας των προβλέψεων. Αυτές οι προκλήσεις υπογραμμίζουν την ανάγκη για μια ολοκληρωμένη προσέγγιση, η οποία θα συνδυάζει την αξιοποίηση των δεδομένων IIoT με τεχνολογίες μηχανικής μάθησης, ώστε να επιτευχθεί ο βέλτιστος έλεγχος για την παρακολούθηση των βιομηχανικών συστημάτων.

1.2 Αντικείμενο διπλωματικής

Η παρούσα διπλωματική εργασία επικεντρώνεται στην ανάπτυξη ενός μοντέλου μηχανικής μάθησης με σκοπό την πρόβλεψη της κατάστασης του εξοπλισμού και την έγκαιρη διάγνωση πιθανών προβλημάτων. Συγκεκριμένα, προτείνονται τεχνικές προ-επεξεργασίας και ανάλυσης δεδομένων, με σκοπό τον μετασχηματισμό της αρχικής πληροφορίας σε μια μορφή που θα προσφέρει βελτιωμένη ακρίβεια και αποδοτικότητα κατά την εκπαίδευση των αλγορίθμων. Παράλληλα, εξετάζονται και συγκρίνονται διαφορετικοί αλγόριθμοι Μηχανικής Μάθησης προκειμένου να επιλεγεί ο αποδοτικότερος βάσει των συγκριτικών αποτελεσμάτων. Το επιλεγόμενο μοντέλο, στη συνέχεια εντάσσεται σε ένα συνολικό σύστημα διαχείρισης και εποπτείας εξοπλισμού, όπου οι χρήστες έχουν τη δυνατότητα να παρακολουθούν την κατάσταση των μηχανημάτων σε πραγματικό χρόνο.

Αρχικώς, θα πρέπει να αντιμετωπιστεί το ζήτημα της ανισοκατανομής των δεδομένων ως προς τις κατηγορίες (κλάσεις) των δειγμάτων, η οποία αν δεν περιοριστεί μέσω κατάλληλων τεχνικών δύναται να προκαλέσει υπέρ-προσαρμογή του μοντέλου ευνοώντας την επικρατούσα κλάση, γεγονός όπου μπορεί να οδηγήσει σε περιορισμένη ικανότητα γενίκευσης και μικρότερη ακρίβεια των προβλέψεων. Για την επίλυση του προβλήματος, θα εφαρμοστούν στρατηγικές εξισορρόπησης οι οποίες θα εξασφαλίσουν ότι το μοντέλο δεν θα ευνοεί την επικρατούσα κατηγορία, επιτυγχάνοντας πιο αντιπροσωπευτική εκπαίδευση.

Επιπλέον, περιγράφεται η διαδικασία επιλογής του βέλτιστου αλγορίθμου μέσω αξιολόγησης των μακροσκοπικών μέτρων απόδοσης, όπως είναι η ευαισθησία, η εξειδίκευση και το F1-score, αποσκοπώντας στην εξασφάλιση υψηλής ακρίβειας και ισορροπίας στην απόδοση του μοντέλου.

Τέλος, η εργασία στοχεύει στη σχεδίαση και ανάπτυξη ενός έξυπνου βιομηχανικού συστήματος το οποίο θα αξιοποιεί τεχνολογίες IIoT και θα ενοποιεί επιμέρους στοιχεία, όπως τα μοντέλα Μηχανικής Μάθησης και τη πλατφόρμα PI, για τη συλλογή και διαχείριση δεδομένων σε πραγματικό χρόνο. Το σύστημα αυτό θα υποστηρίζει τη λήψη αποφάσεων και θα παρέχει αυτόματες ειδοποιήσεις που θα πραγματοποιούνται βάσει παραγόμενων προγνωστικών αναλύσεων, ενισχύοντας έτσι τη δυνατότητα προληπτικής συντήρησης και την έγκαιρη απόκριση σε ενδεχόμενες δυσλειτουργίες.

Με τον τρόπο αυτό, η εργασία αποσκοπεί στην ανάπτυξη μιας μεθοδολογίας που συνδυάζει την προηγμένη ανάλυση δεδομένων και την αξιοποίηση των δυνατοτήτων των IIoT καθώς και των σύγχρονων εργαλείων μηχανικής μάθησης, προκειμένου να υποστηρίξει τη βιομηχανική παραγωγή προς την κατεύθυνση της έξυπνης συντήρησης του εξοπλισμού.

1.3 Δομή της διπλωματικής

Στα πλαίσια της παρούσας Διπλωματικής εργασίας αρχικώς παρουσιάζεται η εξέλιξη της βιομηχανικής συντήρησης και αναλύεται η αναγκαιότητα της προβλεπτικής συντήρησης σε συστήματα Βιομηχανικού Διαδικτύου των Πραγμάτων (IIoT), καθώς και τα οφέλη που προκύπτουν από αυτήν. Στο Κεφάλαιο 3 εξετάζονται τα απαραίτητα χαρακτηριστικά μιας πλατφόρμας συλλογής δεδομένων η οποία θα υποστηρίζει τη χρήση μοντέλων Μηχανικής Μάθησης. Στη συνέχεια, το Κεφάλαιο 4 εστιάζει στους μηχανισμούς και τις τεχνικές εκπαίδευσης αλγορίθμων Μηχανικής Μάθησης για τον εντοπισμό ανωμαλιών κατά τη λειτουργία του εξοπλισμού. Τέλος το Κεφάλαιο 5 αναλύει συγκεκριμένη περίπτωση χρήσης όπου αναπτύσσεται και εκπαιδεύεται ένα μοντέλο μηχανικής μάθησης για τον εντοπισμό πιθανών ασυνεπειών και το οποίο στη συνέχεια συνδέεται με μια πλατφόρμα IIoT για την παροχή ολοκληρωμένης υποστήριξης σε πραγματικό χρόνο.

2

Προβλεπτική Συντήρηση Βιομηχανικού Εξοπλισμού

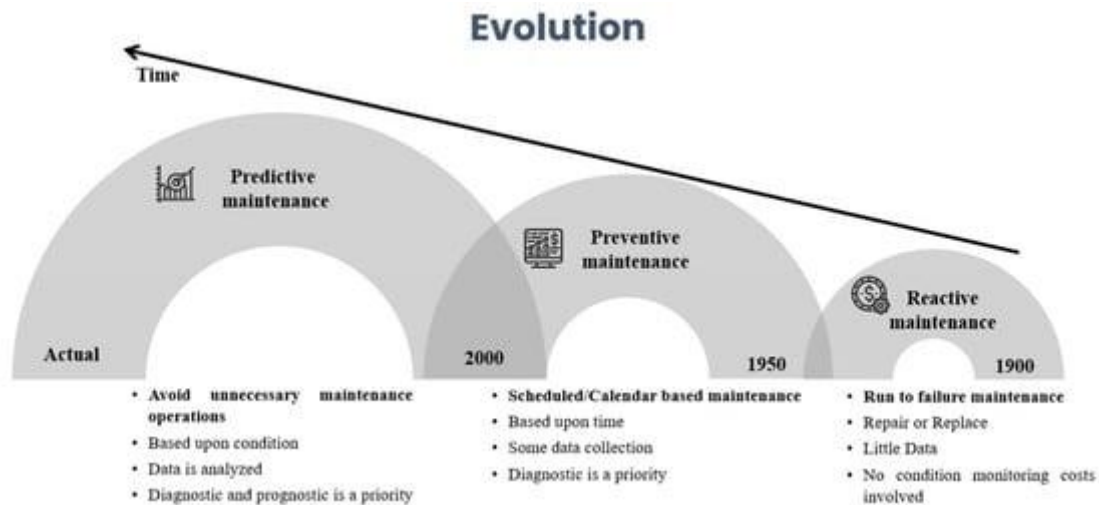
2.1 Η εξέλιξη της βιομηχανικής συντήρησης

Η Τεχνολογία πάντα αποτελούσε τον πυρήνα της καινοτομίας και της εξέλιξης του βιομηχανικού τομέα. Μέσω της συνεχούς ανάπτυξης της, δημιουργεί νέες δυνατότητες και ευκαιρίες για την βελτίωση των διαδικασιών παραγωγής, την ασφάλεια στην εργασία, την αύξηση της απόδοσης και την μείωση του κόστους. Συνεχώς διαμορφώνει το τοπίο από την πρώτη βιομηχανική επανάσταση με την χρήση ατμομηχανών μέχρι και την σημερινή ψηφιακή εποχή της τέταρτης βιομηχανικής επανάστασης όπου η παραγωγή επιτυγχάνεται με πλήρη αυτοματοποίηση και την ελαχιστοποίηση του ανθρώπινου παράγοντα. Ταυτόχρονα όμως, δημιουργείται και η ανάγκη ανάπτυξης κατάλληλων μηχανισμών για την παρακολούθηση του κρίσιμου βιομηχανικού εξοπλισμού όπως ηλεκτρικά μοτέρ (Zheng, 2020) [1].

Η εφαρμογή των μηχανισμών αυτών μπορεί να προσδώσει το απαραίτητο πλεονέκτημα στις επιχειρήσεις που πλέον ανταγωνίζονται σε διεθνές επίπεδο. Το κόστος συντήρησης αποτελεί αναπόσπαστο κομμάτι των λειτουργικών εξόδων για κάθε επιχείρηση. Επιτυγχάνοντας μια μείωση του, είτε μέσω της συρρίκνωσης της συχνότητας των επαναληπτικών ελέγχων και της επιδιόρθωσης του εξοπλισμού, είτε μέσω της εξάλειψης απρόβλεπτων συμβατών όπως της εξαναγκασμένης διακοπής της παραγωγικής διαδικασίας, οι επιχειρήσεις δύναται να αποκομίσουν σημαντικά οικονομικά και παραγωγικά οφέλη. Η αξιολόγηση της αξιοπιστίας ενός συστήματος βάσει δεδομένων σε πραγματικό χρόνο μειώνει 20-25% της εργασίες συντήρησης, ελαττώνει τις βλάβες κατά 35-45% και αυξάνει την συνολική παραγωγικότητα κατά 20-25% (Cakir, 2021) [2].

Επομένως, υπό το βάρος των συνθηκών αυτών οι βιομηχανίες κατά το πέρασμα των ετών εξέλιξαν σταδιακά τις τεχνικές τους, όπως φαίνεται και στην Εικόνα 1. Αναπτύχθηκαν δύο στρατηγικές προσεγγίσεις για την συντήρηση και τον έλεγχο του εξοπλισμού, η διορθωτική συντήρηση (Corrective Maintenance) και η προ-ενεργητική συντήρηση (Proactive Maintenance). Η πρώτη κατηγορία αναφέρεται στην απρογραμμάτιστη και απρόβλεπτη συντήρηση που πραγματοποιείται σε ένα εξοπλισμό για να τον επαναφέρει σε κατάσταση λειτουργίας, γνωστή και ως Reactive Maintenance, εφαρμόστηκε κατά την έναρξη της βιομηχανικής επανάστασης στις αρχές του προηγούμενου αιώνα και η οποία εγκαταλείφθηκε αργότερα καθώς είχε υψηλό κόστος και οδηγούσε σε παρατεταμένες διακοπές λειτουργίας των συστημάτων. Χαρακτηριζόταν από

έλλειψη συγκέντρωσης δεδομένων και προληπτικών ελέγχων, ενώ το προσωπικό επενέβαινε μόνο για να επιδιορθώσει βλάβη ή για αντικαταστήσει τον εξοπλισμό που αστοχούσε.



Εικόνα 1. Εξέλιξη των μεθόδων συντήρησης, Πηγή: [40].

Κατά την είσοδο στη δεύτερη βιομηχανική επανάσταση οι εταιρείες ξεκίνησαν να υιοθετούν ένα σύνολο δράσεων που στόχευαν στην πρόληψη των βλαβών με σκοπό την αύξηση του προσδόκιμου ζωής των συστημάτων και την μείωση των απρογραμμάτιστων διακοπών λειτουργίας. Οι στρατηγικές αυτές είναι ενεργές και εξελίσσονται διαρκώς, αφορούν την προληπτική συντήρηση (*Preventive Maintenance*), την συντήρηση βάσει συνθηκών (*Condition Based Maintenance*) και την προβλεπτική (*Predictive Maintenance*), οι συγκεκριμένες τεχνικές ανήκουν στην δεύτερη κατηγορία της Proactive Maintenance.

2.2 Προ-ενεργητική συντήρηση (*Proactive Maintenance*)

Προληπτική Συντήρηση (*Preventive Maintenance*): Αποτελεί μια σημαντική προσέγγιση στη διαχείριση του εξοπλισμού η οποία αναπτύχθηκε μετά το 1950, όπου εξελίχθηκαν η δεύτερης γενιάς στρατηγικές συντήρησης. Στοχεύει στην πρόληψη των αναπάντεχων βλαβών και την μείωση των εξαναγκασμένων κρατήσεων λόγω αστοχίας του εξοπλισμού. Οι εργασίες περιλαμβάνουν την πραγματοποίηση περιοδικών ελέγχων ανά τακτά χρονικά διαστήματα για την αξιολόγηση των συστημάτων, αλλά και προγραμματισμένες αντικαταστάσεις εξοπλισμού ανεξάρτητα της κατάσταση τους.

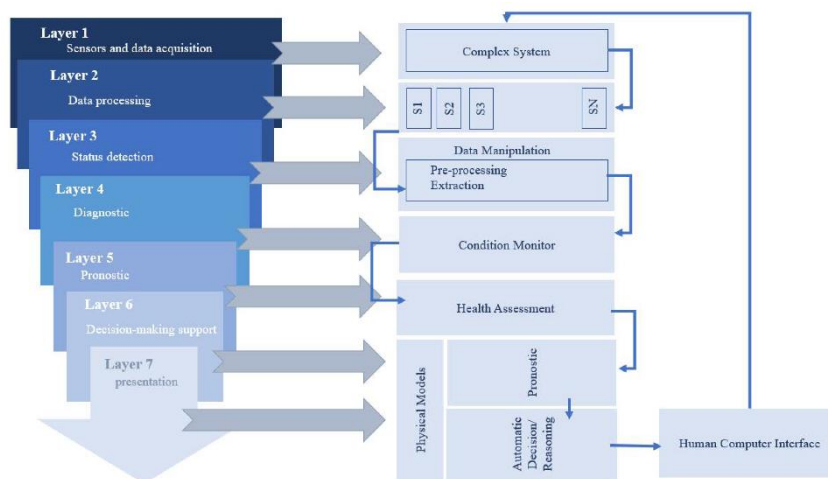
Οι τακτικές επιθεωρήσεις μπορούν να αφορούν οπτικοακουστικούς ελέγχους και λήψη μετρήσεων μέσω κατάλληλων οργάνων. Ωστόσο, οι μετρήσεις που λαμβάνονται από το πεδίο δεν αξιοποιούνται δυναμικά για την δημιουργία αναλύσεων και μοντέλων, απλά αρχειοθετούνται ως σημείο αναφοράς μέχρι την επόμενη επιθεώρηση. Επιπλέον, άλλες τακτικές ενέργειες μπορεί να περιλαμβάνουν ρύθμιση εξαρτημάτων, αλλαγή λαδιών, καθαρισμός κ.α. Παράλληλα, οι προγραμματισμένες αντικαταστάσεις του εξοπλισμού προκύπτουν είτε από οδηγίες που έχει θέσει ο ίδιος ο κατασκευαστής, είτε από τα εγχειρίδια χρήσης ορθής λειτουργίας, ενώ οι εργασίες σχεδιάζονται και εκτελούνται από το αντίστοιχο τμήμα συντήρησης.

Σε κάθε περίπτωση, οι εργασίες είναι χρονοβόρες καθώς αφαιρούν σημαντικές εργατοώρες από το προσωπικό για την εκτέλεση τους και επιπλέον επιφέρουν αύξηση του κόστους δεδομένου

ότι για την ανάγκη των εργασιών απαιτείται απομόνωση του εξοπλισμού ή ακόμα και ολόκληρου τμήματος. Επομένως, είναι αναγκαίος ο προσδιορισμός του βέλτιστου χρόνου εφαρμογής της προληπτικής συντήρησης για την μείωση των αρνητικών επιπτώσεων της. (Asadi, 2023) [3].

Συντήρηση βάσει κατάστασης (Condition Based Maintenance): Είναι μια προηγμένη στρατηγική συντήρησης που εστιάζει στην παρακολούθηση της κατάστασης και της απόδοσης του εξοπλισμού. Απαραίτητο στοιχείο είναι η συλλογή και ανάλυση δεδομένων από αισθητήρια τα οποία έχουν προκαθοριστεί ως σημαντικά αναλόγως τον εξοπλισμό. Επιτρέπει πιο δυναμικές πρακτικές σε αντίθεση με της προηγούμενες μεθόδους δίνοντας την δυνατότητα στο προσωπικό να προβλέπει και να αντιδρά σε προβλήματα συντήρησης, αξιοποιώντας τα δεδομένα που λαμβάνονται από τα όργανα όπως και από δείκτες κατάστασης των συστημάτων. Τα πλεονεκτήματα της μεθόδου αφορούν την καλύτερη διαχείριση των εξαρτημάτων στην αποθήκη, μείωση των απρόβλεπτων σφαλμάτων, αύξηση της αξιοπιστίας και της διαθεσιμότητας, μείωση του χρόνου συντήρησης και κατανάλωσης πόρων.

Η συντήρηση βάσει κατάστασης αποτελείται από δύο κατηγορίες: την εκτός σύνδεσης (offline) και την εντός σύνδεσης (online). Η παρακολούθηση της κατάστασης εκτός σύνδεσης αναφέρεται και ως περιοδική κατάσταση παρακολούθησης και πραγματοποιείται μέσω ανάλυσης των συλλεγμένων δεδομένων σε συγκεκριμένα χρονικά διαστήματα. Επομένως, το βασικό μειονέκτημα της τεχνικής αποτελεί ότι η αξιολόγηση των δεδομένων δεν πραγματοποιείται σε πραγματικό χρόνο αλλά σε κάποιο εργαστηριακό περιβάλλον εκτός των συστημάτων. Αντιθέτως, η παρακολούθηση της κατάστασης εντός σύνδεσης αξιολογεί τα δεδομένα από τους αισθητήρες σε πραγματικό χρόνο την στιγμή που συλλέγονται και συγκρίνονται άμεσα με προκαθορισμένα όρια λειτουργίας. Η δεύτερη κατηγορία καταλαμβάνει όλο και περισσότερο έδαφος τα τελευταία χρόνια με την ανάπτυξη του ΔτΠ καθώς και με την σύγχρονη τάση της αγοράς όπου όλο και περισσότερες εταιρείες τείνουν προς την ψηφιοποίηση των δεδομένων τους (Ingemarsdotter, 2021) [4]. Στην Εικόνα 2 παρουσιάζονται τα επίπεδα λειτουργιών σύμφωνα με την τεχνική συντήρησης.



Εικόνα 2. Τα λειτουργικά επίπεδα της Συντήρησης βάσει κατάστασης, Πηγή: [40].

Επομένως, μέσω της αξιοποίησης συσκευών ΔτΠ δύναται η συγκέντρωση πολυποίκιλων και μεγάλων δεδομένων, τα οποία αναλύονται και οριοθετούνται σημεία αναφοράς για την παρακολούθηση της κατάστασης του εξοπλισμού. Στη συνέχεια τα παραγόμενα νέα δεδομένα συγκρίνονται με τις αναμενόμενες τιμές και σε περίπτωση απρόβλεπτων αλλαγών αποστέλλονται

ειδοποιήσεις. Επιπλέον αξιολογείται η κατάσταση του εξοπλισμού και ο βαθμός υποβάθμισης του και υπολογίζεται η μελλοντική υγεία του. Προσφέροντας ταυτόχρονα την δυνατότητα στο προσωπικό να ελέγχει απομακρυσμένα την κατάσταση των συστημάτων σε πραγματικό χρόνο περιορίζοντας τους περιττούς ελέγχους και παρέχοντας υποστήριξη αποφάσεων και προτεινόμενες ενέργειες.

Προβλεπτική Συντήρηση (Predictive Maintenance): Αποτελεί την τελευταία στρατηγική της Proactive Maintenance και έχει στόχο τον περιορισμό των εξαναγκασμένων κρατήσεων, την αύξηση της αξιοπιστίας και την καλύτερη διαχείριση των πόρων του τμήματος συντήρησης του εργοστασίου. Με την ανάπτυξη κατάλληλων μοντέλων και μεθόδων μπορεί να προβλεφθεί ο αναμενόμενος χρόνος αστοχίας και κατ' επέκταση η διάρκεια ζωής του εξοπλισμού (Varghese, 2023) [5]. Η τεχνική αποσκοπεί στον περιορισμό των πρόωρων και δαπανηρών επισκευών που μπορούν να οδηγήσουν σε ανεπιθύμητες σπατάλες πόρων αλλά και σε αύξηση του κόστους των υλικών, καθώς και στην έγκαιρη επέμβαση του προσωπικού πριν την εμφάνιση βλάβης στον εξοπλισμό.

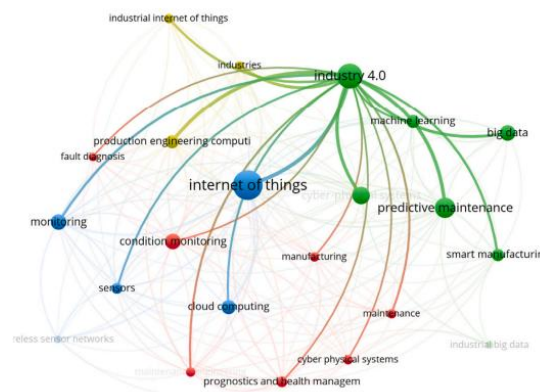
Σύμφωνα με αυτή την προσέγγιση συλλέγονται και επεξεργάζονται ιστορικά δεδομένα από αισθητήρες κατά την λειτουργία των συστημάτων ενώ επιπλέον εφαρμόζονται αναλύσεις για την αξιοποίηση δεδομένων σε πραγματικό χρόνο μέσω των οποίων ενεργοποιούνται ειδοποιήσεις όταν ξεπεραστούν κάποιες προκαθορισμένες συνθήκες. Με την βοήθεια των συλλεγμένων στοιχείων εκπαιδεύονται αλγόριθμοι μηχανικής μάθησης, όπου καλύπτουν δύο βασικά πεδία: την διάγνωση και την προγνωστική ανάλυση. Η διάγνωση αφορά στον εντοπισμό σφαλμάτων των συστημάτων και τον προσδιορισμό των αιτιών τους, ενώ η προγνωστική ανάλυση αφορά στην αξιολόγηση της υποβάθμισης του συστήματος προβλέποντας μελλοντική αστοχία, καθώς και στον υπολογισμό του εναπομένοντα χρόνου ζωής (RUL). Τέλος, δημιουργούνται αναφορές για την καλύτερη διαχείριση και προγραμματισμό των εργασιών, ενώ παρέχεται και δυνατότητα απομακρυσμένου ελέγχου της κατάστασης του εξοπλισμού.

2.3 Η προληπτική συντήρηση στην Industry 4.0

Η μετάβαση των οργανισμών στην νέα ψηφιακή εποχή είναι γνωστή και ως τέταρτη βιομηχανική επανάσταση ή και *Industry 4.0*. Ο όρος αναφέρεται στον μετασχηματισμό των συμβατικών βιομηχανιών μέσω της ενοποίησης διαφορετικών τεχνολογιών και επιχειρηματικών διαδικασιών. Μερικοί από τους κυριότερους παράγοντες ώθησης της τέταρτης βιομηχανικής επανάστασης είναι το ΔτΠ (IIoT), η ασφάλεια στον κυβερνοχώρο (Cyber security), το Cloud Computing, τα μεγάλα δεδομένα (Big Data), η αυτοματοποίηση και άλλα, όπως αυτά παρουσιάζονται στην Εικόνα 3.

Η αξιοποίηση των αλγόριθμων Μηχανικής Μάθησης και η εφαρμογή προβλεπτικής συντήρησης είναι ένας βασικός παράγοντας της τέταρτης βιομηχανικής επανάστασης. Οι οργανισμοί μπορούν να επωφεληθούν μέσω της χρήσης προβλεπτικών μοντέλων αυξάνοντας την βιωσιμότητα του εξοπλισμού, οδηγώντας στην μείωση των χρησιμοποιούμενων υλικών και στη καλύτερη διαχείριση χρόνου (Çinar, 2020) [7]. Επιπλέον, η προληπτική συντήρηση συμβάλλει στη βελτίωση της ποιότητας του παραγόμενου προϊόντος στην ασφάλεια και την παραγωγικότητα της εργασίας (Zonta, 2020)[8]. Προσφέρονται δυνατότητες ανάπτυξης ευφυών συστημάτων, όπου μέσα

από την συλλογή και ανάλυση μεγάλων δεδομένων επιτρέπουν την εξαγωγή αποφάσεων και ενεργειών, όπως σε μηχανήματα CNC (Bachim, 2020) [9].



Εικόνα 3. Βασικά στοιχεία της Industry 4.0, Πηγή: [8].

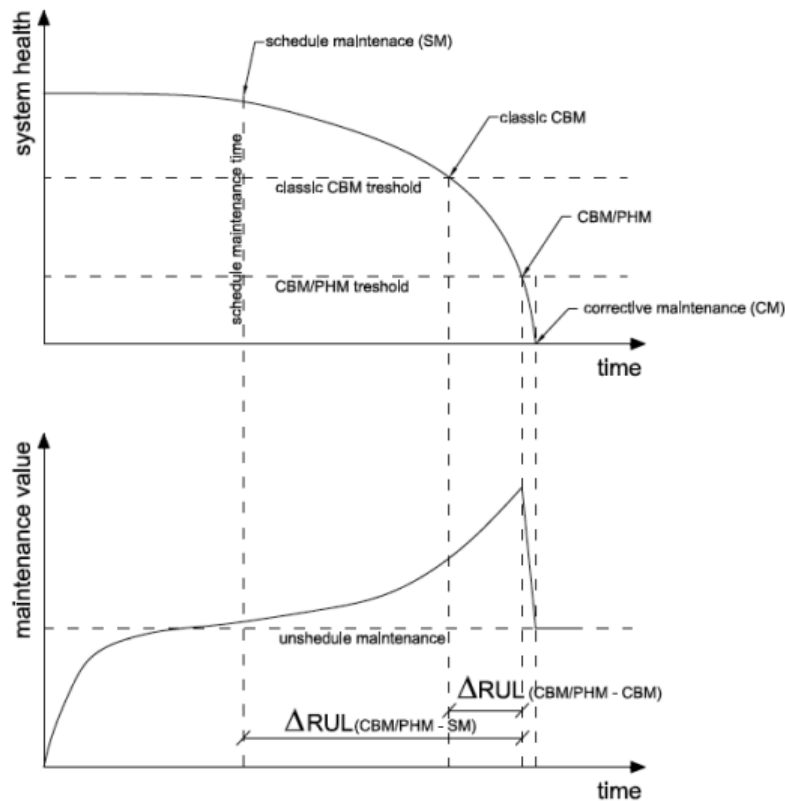
Πέρα όμως από τα πλεονεκτήματα που μπορεί να προσφέρει η ενσωμάτωση των νέων τεχνολογιών σε ένα οργανισμό οδηγεί και στην εμφάνιση νέων προκλήσεων. Η εφαρμογή των τεχνολογικών εξελίξεων θα δημιουργήσει νέες δυναμικές στο περιβάλλον εργασίας και οι οργανισμοί θα πρέπει να είναι πρόθυμοι και ικανοί να προσαρμοστούνε στις νέες απαιτήσεις (Sony M, 2020) [6].

2.4 Προγνωστική και Διαχείριση Υγείας

Η Προγνωστική και Διαχείριση Υγείας (*Prognostics and Health Management – PHM*) είναι μια μέθοδος κατά την οποία αναλύεται η τρέχουσα και μελλοντική κατάσταση ενός εξοπλισμού ή και ενός σύνθετου συστήματος, παρέχοντας υποστήριξη αποφάσεων στο τμήμα συντήρησης. Στόχος είναι να προσδιοριστεί το επίπεδο αξιοπιστίας, εντοπίζοντας αρχικά στάδια υποβάθμισης των συστημάτων ή και τμήματος αυτών, αλλά και τις αιτίες που οδηγούν σε σταδιακή φθορά. Το PHM αποτελεί κύριο στοιχείο της προβλεπτικής συντήρησης και η ορθή ανάπτυξη και εφαρμογή του προσφέρει πολλά πλεονεκτήματα στην Industry 4.0 (Zio, 2022)[10]. Ο στόχος επιτυγχάνεται μέσω της συλλογής, ανάλυσης και ερμηνείας διαφορετικών παραμέτρων, όπως της κατάστασης λειτουργίας, απρόσμενες διακοπές συστημάτων και δείκτες απόδοσης (*Key Performance Indicators – KPI*) για την μέτρηση σημαντικών μεγεθών.

Όπως έχει αναφερθεί και παραπάνω σκοπός είναι η ανάπτυξη μιας μεθόδου που θα προσφέρει υποστήριξη αποφάσεων και ευκολία στην διαχείριση ενώ επιπλέον θα περιορίζει τα κόστη και θα μειώνει τους χρόνους προληπτικών συντηρήσεων. Στην Εικόνα 4 παρουσιάζεται η συσχέτιση μεταξύ εξαγόμενης αξίας και εφαρμογής διαφορετικών τεχνικών συντήρησης. Η επιδίωξη είναι να χρησιμοποιηθεί η τεχνική αυτή που θα προσδώσει τη μεγαλύτερη αξία στην επιχείρηση, όπου σύμφωνα με το διάγραμμα οποίο προκύπτει κατά την χρήση του PHM. Ενώ αντίθετα, η διορθωτική συντήρηση, δηλαδή η αντικατάσταση ενός εξοπλισμού όταν αυτός έχει αστοχήσει, παρέχει το χαμηλότερο κέρδος για την επιχείρηση. Ακόμα, διακρίνεται η συναγόμενη αξία από την μετάβαση από τη συντήρηση βάση κατάστασης στη προβλεπτική συντήρηση μέσω πρόγνωσης και

διαχείρισης υγείας. Τέλος, παρατηρείται η χαμηλή αξία της προληπτικής συντήρησης που σταδιακά αυξάνεται όσο συγκλίνει προς τον παραδοσιακό CBM και τον PHM.



Εικόνα 4. Συσχέτιση αξίας και τεχνικών συντήρησης, Πηγή: [41].

Η οργάνωση της Προγνωστικής και Διαχείρισης Υγείας των Μηχανών προϋποθέτει την ολοκλήρωση συγκεκριμένων βημάτων για την σωστή μοντελοποίηση του συστήματος (Atamuradon, 2017) [11]. Η Εικόνα 5 μας δίνει την γενική επισκόπηση των επιμέρους στοιχείων που συναποτελούν το PHM.



Εικόνα 5. Μοντέλο PHM, Πηγή: [11].

Συλλογή Δεδομένων (Data Acquisition): Αφορά την συγκέντρωση και αποθήκευση δεδομένων μέσω αισθητηρίων για περεταίρω ανάλυση. Ξεχωριστής χρήσης όργανα και από διαφορετικούς κατασκευαστές όπως αισθητήρες, ελεγκτές και υπολογιστικές μηχανές παράγουν

διαρκώς ακατέργαστα δεδομένα ή και ειδοποιήσεις συμβάντων. Ως ειδοποίηση συμβάντων αναφέρουμε οποιαδήποτε πληροφορία σχετικά με την κατάσταση ενός συστήματος, όπως αντλία σε λειτουργία, ενεργοποίηση θερμικού, σφάλμα συστήματος και άλλα.

Η ετερογένεια των συσκευών ήταν πάντα μια ανοικτή πρόκληση για συστήματα που θα έπρεπε να συνδεθούν με όργανα διαφορετικών προμηθευτών. Στο επόμενο κεφάλαιο αναφέρεται το σύστημα επιλογής με το οποίο θα ασχοληθεί παρούσα διπλωματική (*PI System*) το οποίο πέρα από την δυνατότητα αποθήκευσης μεγάλου όγκου δεδομένων, προσφέρει και μία μεγάλη συλλογή από λειτουργικά συστήματα για την σύνδεση οργάνων με διαφορετικά πρωτόκολλα επικοινωνίας (*Modbus, OPC DA/UA/XML* κλπ.).

Προ-επεξεργασία Δεδομένων (Data Preprocessing): Αποτελεί το δεύτερο βήμα στο PHM και αφορά τον καθαρισμό και ανάλυση δεδομένων. Αρχικά περιλαμβάνει την προ-επεξεργασία των σημάτων για την αφαίρεση θορύβου από τα αρχικά δεδομένα, δηλαδή απότομες και τυχαίες διακυμάνσεις των σημάτων.

Στη συνέχεια εξάγονται και αξιολογούνται τα χαρακτηριστικά και επιλέγονται εκείνα που κρίνονται ότι έχουν την μεγαλύτερη βαρύτητα για τον έλεγχο και την προγνωστική των συστημάτων. Σκοπός του βήματος είναι η μείωση της πολυπλοκότητας του μοντέλου αφαιρώντας χαρακτηριστικά που αυξάνουν απλώς τις διαστάσεις διατηρώντας όσα είναι χρήσιμα για τον προσδιορισμό της κατάστασης της υγείας του υπό παρακολούθηση συστήματος.

Ανίχνευση, Διάγνωση και Πρόγνωση (Detection, Diagnosis and Prognosis): Είναι τα τρία κύρια στάδια που ανήκουν στο κομμάτι της Ανάλυσης του μοντέλου PHM. Στόχος είναι μέσω της αξιοποίησης των συλλεγμένων και επεξεργασμένων δεδομένων να ανιχνευθούν εγκαίρως πρώιμες μη φυσιολογικές συμπεριφορές. Ακολούθως, να εντοπιστούν οι αιτίες οι οποίες ευθύνονται για την ανωμαλία και τέλος να πραγματοποιείται μια πρόβλεψη της υγείας καθώς και της μελλοντικής κατάστασης του εξοπλισμού και να υπολογίζεται το RUL (Hu, 2017)[12].

Υποστήριξη αποφάσεων (Decision Making): Εκμεταλλεύεται τα αποτελέσματα των προηγούμενων βημάτων και προσφέρει υποστηρικτικό ρόλο κατά την λήψη αποφάσεων. Παρέχει συστάσεις για βέλτιστες ενέργειες, υποδεικνύει στους υπεύθυνους αντίστοιχες περιπτώσεις βλαβών και αποστέλλει αυτόματες αιτήσεις εργασιών για την συντήρηση του εξοπλισμού στα συστήματα επιχειρησιακής διαχείρισης.

Διεπαφή Ανθρώπου Μηχανής (HM Interface): Χρησιμεύει για την οπτικοποίηση της κατάστασης υγείας του εξοπλισμού επιτρέποντας στους χρήστες την άμεση πρόσβαση σε δεδομένα του συστήματος σε πραγματικό χρόνο μέσω γραφημάτων ή και ψηφιακών αντιγράφων. Επιπλέον, παρέχεται η δυνατότητα αλληλεπίδρασης για την εκτέλεση αναγκαίων ενεργειών, βελτιώνοντας τον χρόνο απόκρισης και απόδοσης του τμήματος συντήρησης.

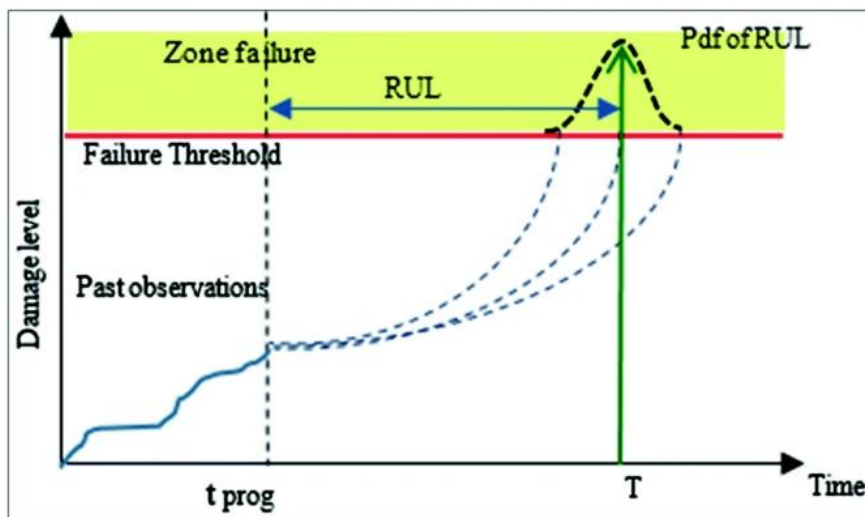
2.5 Πρόγνωση Σφάλματος

Όπως αναφέρθηκε και στη προηγούμενη υπό-ενότητα το αρχικό στάδιο της Ανάλυσης του PHM αφορά την ανίχνευση του σφάλματος. Επικεντρώνεται στον έλεγχο της κατάστασης του μηχανήματος ή του συστήματος, συγκρίνοντας τα ιστορικά δεδομένα κατά την ορθή λειτουργία του με τα δεδομένα που συλλέγονται κατά την παρούσα χρονική στιγμή. Ουσιαστικά είναι η εφαρμογή της συντήρησης βάσει κατάστασης που περιεγράφηκε παραπάνω. Περιλαμβάνει την ενεργοποίηση

συμβάντων και αποστολή ειδοποιήσεων σε περίπτωση εντοπισμού ανωμαλιών. Υπολογίζει την απόκλιση σε σχέση με την αναμενόμενη φυσιολογική τιμή καθώς και το χρονικό πλαίσιο από την στιγμή εμφάνισης του συμβάντος. Η έγκαιρη αναγνώριση ανωμαλιών κρίνεται αναγκαία καθώς μπορεί να αποδώσει σημαντικό χρονικό περιθώριο για τον προγραμματισμό των απαιτούμενων ενεργειών.

Το επόμενο στάδιο αφορά τη διάγνωση σφαλμάτων και περιλαμβάνει τον εντοπισμό και την ταυτοποίηση του παράγοντα που ευθύνεται για την αστοχία καθώς και εκτίμηση του μεγέθους της υποβάθμισης. Απαιτεί αναλυτική επεξεργασία των συλλεγμένων δεδομένων και αντιμετωπίζει προκλήσεις έλλειψης και εσφαλμένων δεδομένων, ενώ ενισχύεται από την ύπαρξη ιστορικών δεδομένων από προηγούμενες διαφορετικές καταστάσεις του συστήματος. Αποτελείται από δύο επίπεδα: την απομόνωση για τον προσδιορισμό του μέρους ή του εξαρτήματος που έχει υποστεί την φθορά ή την μείωση της απόδοσης λειτουργίας, και την ταυτοποίηση για την εκτίμηση της έκτασης της βλάβης.

Η πρόγνωση των σφαλμάτων αποτελεί το τελευταίο κύριο στάδιο της Ανάλυσης του PHM και είναι η διαδικασία η οποία ασχολείται με τον προσδιορισμό του εναπομείναντα χρόνου ζωής του μηχανήματος μετά τον εντοπισμό του σφάλματος αλλά και την διάγνωση της αιτίας που ευθύνεται για την μη φυσιολογική συμπεριφορά (Εικόνα 6).

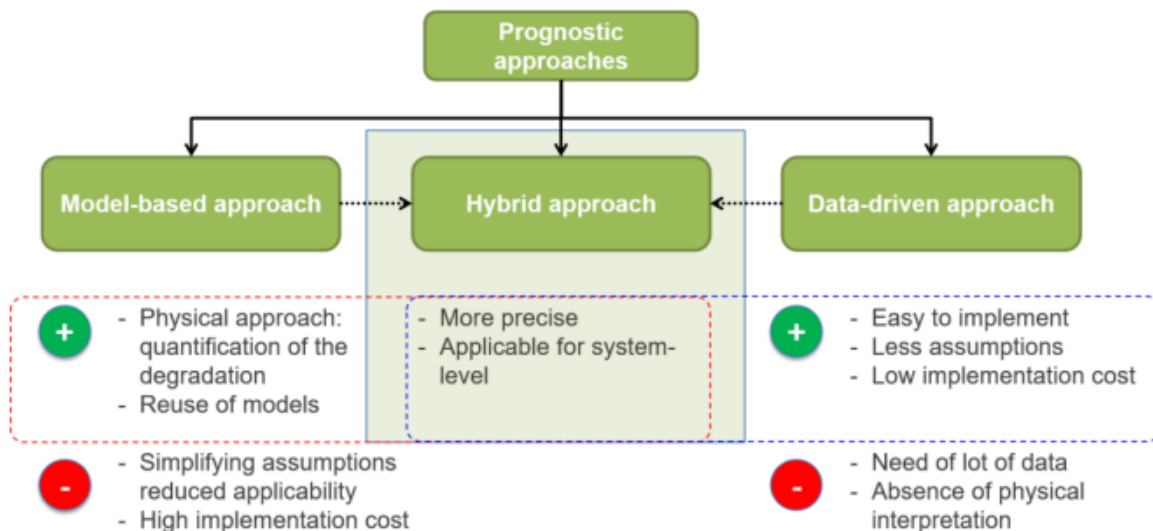


Εικόνα 6. Προσδιορισμός του RUL μετά τον εντοπισμό του σφάλματος, Πηγή: [42].

Σκοπός της Πρόγνωσης είναι να παρέχει υποστήριξη στις προ-ενεργητικές αποφάσεις προβλέποντάς την μελλοντική εξέλιξη της αποτυχίας του συστήματος ή μέρους αυτού. Χρησιμοποιεί ιστορικά δεδομένα από προηγούμενες καταστάσεις για την εκτίμηση της πιθανότητας αστοχίας καθώς και του RUL. Τα αποτελέσματα της πρόγνωσης χρησιμοποιούνται για την διαχείριση της υγείας των συστημάτων, τον χρονοπρογραμματισμό των παρεμβάσεων της συντήρησης και την καλύτερη διαχείριση των πόρων από τις επιχειρήσεις. Ωστόσο να σημειωθεί ότι ενδεχόμενες απώλειες δεδομένων από αισθητήρες ή ελλείψεις ιστορικών αναφορών από πρότερες εσφαλμένες λειτουργίες ή ανωμαλίες των μηχανημάτων καθιστούν δύσκολη την ακριβή πρόβλεψη της μελλοντικής αστοχίας. Η επιτυχία της πρόγνωσης στηρίζεται σε καταχωρημένα ιστορικά στοιχεία, σε πληροφορίες που συγκεντρώθηκαν κατά την ενεργοποίηση συμβάντων και σε διαθέσιμους βιομηχανικούς αισθητήρες που ενδεχομένως να μην είχαν εγκατασταθεί αρχικώς

στα συστήματα. Σε κάποιες περιπτώσεις υπάρχουν αρκετά δεδομένα για την ανάπτυξη εμπειρικών μοντέλων τα οποία βασίζονται σε προηγούμενες αποτυχίες, όμως υπάρχουν και περιπτώσεις που δεν υπάρχουν αρκετά δεδομένα δημιουργώντας μια αβεβαιότητα στο μοντέλο σχετικά με την ικανότητα ακριβής πρόβλεψης. Συνεπώς είναι απαραίτητο κατά την κατασκευή του κάθε μοντέλου να λαμβάνονται υπόψη οι εκάστοτε διαφοροποιήσεις που παράγουν αβεβαιότητες.

Στην Εικόνα 7 παρουσιάζονται οι τρεις κατηγορίες προσεγγίσεων για την μοντελοποίηση της προγνωστικής και διαχείρισης υγείας: i) *Model -Based*, ii) *Data driven* και iii) *Hybrid*.



Εικόνα 7. Κατηγορίες Προγνωστικών προσεγγίσεων, Πηγή: [11].

Model Based: Χρησιμοποιεί τυποποιημένα μαθηματικά μοντέλα για την περιγραφή της συμπεριφοράς των συστημάτων. Επιλέγεται όταν υπάρχει εμπειρική γνώση των διαδικασιών λειτουργίας και των φυσικών νόμων. Μέσω των μοντέλων αυτών παρακολουθείται η φυσική φθορά και προβλέπεται η πιθανή εξέλιξη της, όπως για την πρόγνωση και διαχείριση υγείας μπαταριών λιθίου (Khaleghi , 2022)[13] και βιομηχανικού εξοπλισμού με την χρήση deep learning (Xu, 2020)[14]. Τα πλεονεκτήματα της προσέγγισης είναι: η ευελιξία καθώς αλλάζουν μόνο τα μοντέλα που χρησιμοποιούν οι αλγόριθμοι, η ευρεία εφαρμογή σε μια μεγάλη ποικιλία συστημάτων, και η σαφήνεια εφόσον η προσέγγιση μπορεί να διατυπωθεί μαθηματικά προσφέροντας ξεκάθαρες διαδικασίες.

Data Driven: Αντίθετα με την πρώτη κατηγορία, η συγκεκριμένη προσέγγιση δεν προϋποθέτει τη γνώση και αντίληψη των διαδικασιών λειτουργίας των συστημάτων. Βασίζεται αποκλειστικά στην αξιοποίηση των δεδομένων χωρίς την κατασκευή υποθετικών μοντέλων για την επεξήγηση των μηχανισμών των συστημάτων. Για τον λόγο αυτό, η εφαρμογή είναι γρηγορότερη και ευκολότερη καθώς δεν απαιτείται εμπειρική γνώση ή γνώση του αντικειμένου, ωστόσο χρειάζεται μεγάλος όγκος δεδομένων από διαφορετικούς αισθητήρες και σε διαφορετικές καταστάσεις της μηχανής, προκειμένου να υπάρχει ακρίβεια πρόβλεψης. Υπάρχει πλήθος ΑΙ τεχνικών για την εκπαίδευση αλγορίθμων, όπως για παράδειγμα κατά την πρόβλεψη και διάγνωση της υγείας βαρίων βιομηχανικών μηχανημάτων με την χρήση Support Vector Machine (SVM), Neural Networks (NN), Bayesian Neural Networks and Random Forest (RF) (Cofre-Martel, 2021)[15], αλλά και για την

αξιοποίηση τους στην χημική βιομηχανία μέσω Random Forest Regressor (*RFR*) και XGBoost (Souza, 2023)[[16](#)].

Hybrid: Συνδυάζει τα πλεονεκτήματα και των δύο προσεγγίσεων, χρησιμοποιεί τυποποιημένα μαθηματικά μοντέλα για την περιγραφή της συμπεριφοράς των συστημάτων, ενώ παράλληλα εκμεταλλεύεται και τα δεδομένα για την ενίσχυση της ακρίβειας και της αξιοπιστίας των προβλέψεων. Με την συνδυαστική χρήση και των δύο τεχνικών Μηχανικής Μάθησης επιτυγχάνεται αυξημένη ακρίβεια στην ανίχνευση ανωμαλιών και στην πρόγνωση της φθοράς του εξοπλισμού.

3

Χαρακτηριστικά πλατφόρμας για την αξιοποίηση της Μηχανικής Μάθησης

3.1 Προκλήσεις για την ενσωμάτωση του ΔτΠ στην Βιομηχανία

Η ανάπτυξη του Διαδικτύου των Πραγμάτων στην σύγχρονη εποχή έχει επιφέρει σημαντικές αλλαγές στην καθημερινότητα των ανθρώπων. Νέες συσκευές κατασκευάζονται με βασικό στοιχείο την διασύνδεση τούς με επιμέρους εφαρμογές και συστήματα, επιτρέποντας την αυτοματοποίηση και την βελτίωση της αποδοτικότητας τους σε διάφορους τομείς, παράγοντας ψηφιακά δεδομένα και στατιστικά κατά την λειτουργία τους. Έξυπνες οικιακές συσκευές, αυτόνομα μεταφορικά μέσα, ενσωματωμένα φορητά συστήματα για τον έλεγχο της κατάστασης της υγείας του χρήστη, είναι μερικά από τα παραδείγματα που καταδεικνύουν την επιρροή που φέρει η νέα τεχνολογία στην καθημερινότητα του χρήστη. Μέσω αυτών των τεχνολογικών εξελίξεων προσπαθεί να επωφεληθεί και η βιομηχανία για την μετάβαση της στην Industry 4.0. Όπως αναφέρθηκε και σε προηγούμενο κεφάλαιο, το IIoT αποτελεί έναν ακρογωνιαίο λίθο για την ψηφιοποίηση των τομέων της βιομηχανίας και απαραίτητο συστατικό για την εξέλιξη των λειτουργιών και διαδικασιών ενός οργανισμού. Συνεπώς, η επιτυχής επικοινωνία και διασύνδεση των επιμέρους συσκευών IIoT με παλαιού τύπου εφαρμογές (legacy applications) καθώς και με τον υφιστάμενο εξοπλισμό ανοίγει τις πόρτες για την διερεύνηση νέων δυνατοτήτων και την αντιμετώπιση προβλημάτων που παλαιότερα θα φαινόταν αδύνατο. Ωστόσο, το ΔτΠ δεν αποτελεί τον μοναδικό ενεργοποιητή για την μετάβαση προς την έξυπνη βιομηχανία, είναι απαραίτητη η χρήση και άλλων υποστηρικτικών τεχνολογιών όπως για παράδειγμα η Μηχανική Μάθηση και τα Μεγάλα Δεδομένα. Η εξέλιξη, η διάδοση και εν τέλη η εφαρμογή των τεχνολογιών αυτών σε διαφορετικούς κλάδους της βιομηχανίας (όπως παραγωγή και διύλιση πετρελαίου, παραγωγή ηλεκτρικής ενέργειας, νερό, χημικά, φαρμακευτικά προϊόντα, τσιμέντα, μπαταρίες κλπ.) αύξησε τις απαιτήσεις για την δημιουργία ολοκληρωμένων λύσεων που αφορούν στον καλύτερο έλεγχο, διαχείριση και συντήρηση των βιομηχανικών εξοπλισμών. Στόχος ήταν να εξελιχθούν τα πρότυπα διαδικασιών συντήρησης και να απομακρυνθούν οι οργανισμοί από παλαιότερες πρακτικές και νοοτροπίες, όπως η διορθωτική συντήρηση, και να δημιουργηθούν στη θέση τους εφαρμογές και συστήματα που θα παρακολουθούν συνεχώς την κατάσταση των εξοπλισμών και θα προβλέπουν την πιθανότητα

σφαλμάτων. Με τον τρόπο αυτό οι βιομηχανίες θα είναι σε θέση να βελτιώσουν την αποδοτικότητα τους και να μειώσουν τα κόστη συντήρησης, ενώ επιπλέον αυξάνουν την αξιοπιστία και τη διάρκεια ζωής των συστημάτων τους.

Η ώθηση που προσέφεραν οι νέες τεχνολογίες σε συνδυασμό με την ακατάπαυστη προσπάθεια των εταιρειών να αποκτήσουν σημαντικό ανταγωνιστικό πλεονέκτημα, αποτέλεσαν πρόσφορο έδαφος για την ανάπτυξη μοντέλων προβλεπτικής συντήρησης και τον σχεδιασμό τους με ολοκληρωμένες λύσεις PdM. Η προγνωστική και διαχείριση της υγείας των μηχανών σχεδιάζεται για να βοηθήσει στον εντοπισμό των περιπτώσεων εκείνων κατά τον οποίο ο εξοπλισμός λειτουργεί εκτός των φυσιολογικών πλαισίων, παρέχει υποστήριξη στον έλεγχο της κατάστασης του συστήματος και προσδιορίζει τους λόγους που ευθύνονται για την υποβάθμιση του. Επιπλέον, μέσω μοντέλων εκτιμάται ο εναπομείναντας χρόνος ζωής και προγραμματίζεται μια έξυπνη και στοχευμένη προληπτική συντήρηση των μηχανημάτων που έχει ως σκοπό την αποφυγή ή συρρίκνωση των απροσδόκητων βλαβών. Κατά τον τρόπο αυτό, δημιουργείται ένα έξυπνο κεντρικοποιημένο σύστημα ελέγχου και διαχείρισης βιομηχανικού εξοπλισμού, ικανό να εντοπίζει ασυνέχειες και να ενημερώνει το προσωπικό για τις ενέργειες που πρέπει να λάβει.

Προκειμένου να αναπτυχθεί ένα σύστημα συλλογής και επεξεργασίας δεδομένων οι βιομηχανίες έπρεπε να υιοθετήσουν μεθόδους Βιομηχανικού Διαδικτύου των Πραγμάτων αξιοποιώντας κατάλληλες πλατφόρμες οι οποίες είναι διαθέσιμες στην αγορά. Στα περισσότερα συστήματα βιομηχανικού ελέγχου λαμβάνονται ψηφιακές μετρήσεις από διεργασίες, οι οποίες είναι σημαντικές για την επισκόπηση της λειτουργίας των μηχανημάτων αλλά και την ρύθμιση καίριων παραγόντων της συνολικής διαδικασίας του σταθμού (ρύθμιση ροής, στάθμη δεξαμενής, θερμοκρασία καζανιού, ταχύτητα αντλίας, έλεγχος δοσομετρικών αντλιών κ.α.). Ωστόσο, κατά το πέρας της βάρδιας συλλέγονται περιοδικά και χειροκίνητες μετρήσεις οι οποίες αποθηκεύονται σε εσωτερικά έγγραφα για μελλοντική ανασκόπηση. Ακόμα, είναι σύνηθες για την μείωση του κόστους εγκατάστασης και συντήρησης να μην υπάρχουν διαθέσιμοι όλοι οι δυνατοί ψηφιακοί μετρητές. Κατά το παρελθόν ένα μεγάλο μέρος της συλλογής δεδομένων πραγματοποιούνταν χειροκίνητα, πλέον η βιομηχανία στρέφεται προς την χρήση έξυπνων συσκευών για να αντισταθμίσει την έλλειψη δεδομένων και την συγκράτηση του κόστους. Οι συσκευές ΔτΠ δύναται να δικτυωθούν μεταξύ τους επιτρέποντας την συγκέντρωση και συλλογή των δεδομένων σε ένα τοπικό Server ή και στο Cloud, προσφέροντας ευκολία και ταχύτητα εγκατάστασης. Η αντικατάσταση των διαδικασιών και η αλλαγή νοοτροπίας είναι η αρχική πρόκληση που θα πρέπει να αντιμετωπίσει ο κάθε οργανισμός προς το ταξίδι της ψηφιακής μετάβασης. Τα Data Driven μοντέλα πρόγνωσης σφάλματος, χρειάζονται μεγάλο όγκο δεδομένων από διαφορετικούς αισθητήρες ώστε να επιτύχουν μεγαλύτερη ακρίβεια στην περιγραφή των συστημάτων. Δια το λόγο αυτό, είναι απαραίτητη η εγκατάσταση μιας πληθώρας διαφορετικών αισθητηρίων τα οποία να καλύπτουν τις ανάγκες λειτουργίας της εκάστοτε βιομηχανίας, και την αντικατάσταση παλαιότερων τοπικών οργάνων με νέες έξυπνες συσκευές.

Οι πλατφόρμες ΔτΠ θα πρέπει να είναι σε θέση να υποστηρίζουν την ετερογένεια των συσκευών και των πρωτοκόλλων επικοινωνίας τους καθώς και να έχουν την δυνατότητα συλλογής ιστορικών δεδομένων. Η έλλειψη ομοιογένειας στις βιομηχανικές συσκευές Διαδικτύου των Πραγμάτων οφείλεται σε πολλούς παράγοντες από τον σχεδιασμό μέχρι και την κατασκευή τους. Διαφορετικοί προμηθευτές κατασκευάζουν συσκευές με σκοπό την κάλυψη συγκεκριμένων αναγκών εξατομικεύοντας τις προτεινόμενες λύσεις ενώ άλλες συσκευές αναπτύσσονται βάση

ειδικών προδιαγραφών και απαιτήσεων για την αξιοποίηση τους από πολύ συγκεκριμένους τομείς όπως βιομηχανία, υγεία, φαρμακοβιομηχανία, αυτοκίνηση κλπ. Κατά αυτό τον τρόπο δημιουργείται ένα δίκτυο από συσκευές με ξεχωριστές υπολογιστικές δυνατότητες και με διαφορετικά πρωτόκολλα επικοινωνίας. Οι πλατφόρμες θα πρέπει να είναι σε θέση να διαχειριστούν την ανομοιογένεια του δικτύου προτείνοντας εφαρμογές που μπορούν να διασυνδέσουν τις συσκευές μεταξύ τους και να υποστηρίζουν την συνεχή ροή δεδομένων σε πραγματικό χρόνο. Επιπλέον, κρίνεται αναγκαία και η σύνδεση τους με βάσεις δεδομένων για την ανάκληση και αρχειοθέτηση παλαιότερων ιστορικών δεδομένων καθώς είναι απαραίτητα για την εκπαίδευση των αλγόριθμων Μηχανικής Μάθησης, επιτρέποντας τους να μάθουν από προηγούμενα παραδείγματα και να αναγνωρίσουν πρότυπα συμπεριφοράς. Μέσα από την διαδικασία αυτή τα μοντέλα μπορούν να κατανοήσουν τις σχέσεις μεταξύ των δεδομένων και να βελτιώσουν την ακρίβεια των προβλέψεων τους. Τα ιστορικά δεδομένα περιέχουν σημαντικές πληροφορίες και μετρήσεις από παρελθοντικές καταστάσεις και σφάλματα των συστημάτων, η ορθή αξιοποίηση των οποίων θα προσδώσει την ικανότητα στα μοντέλα να αναγνωρίζουν αποκλίσεις σε σχέση με την αναμενόμενη τιμή αλλά και να εντοπίζουν τυχόν υποβάθμιση του εξοπλισμού, το οποίο είναι καίριο για εφαρμογές όπως η προγνωστική συντήρηση. Χωρίς την δυνατότητα προσπέλασης των δεδομένων αυτών, τα εκπαιδευόμενα μοντέλα δεν θα έχουν την απαραίτητη βάση για να συγκρίνουν παλαιότερες και τρέχουσες καταστάσεις και να προσδιορίζουν με επιτυχία μελλοντικές εξελίξεις, καθιστώντας τα λιγότερο αξιόπιστα και αποτελεσματικά.

Επιπλέον μια από τις κύριες προκλήσεις του βιομηχανικού Διαδικτύου των Πραγμάτων είναι η διαχείριση του μεγάλου όγκου δεδομένων για τη λήψη αποφάσεων. Το βιομηχανικό πεδίο αποτελείται από εκατοντάδες ή και χιλιάδες συσκευές που παράγουν αδιάληπτα τεράστιους όγκους δεδομένων. Εξοπλισμός όπως μετρητές, αντλίες, αισθητήρες, ρυθμιστικές ή απομονωτικές βάνες είναι μόνο μερικές από τις συσκευές που βρίσκονται στο πεδίο, αναλόγως της κρισιμότητας του βιομηχανικού εξοπλισμού εντείνεται και ο ρυθμός δειγματοληψίας των δεδομένων που τα ίδια παράγουν. Επομένως, για να είναι αποτελεσματική μία πλατφόρμα ΔτΠ πρέπει να μπορεί να διαχειρίζεται τα παραγόμενα μεγάλα δεδομένα έγκαιρα και αποδοτικά, διαχειρίζοντας με ασφάλεια και συνέπεια τον μεγάλο φόρτο εργασίας. Όπως αναφέρθηκε και παραπάνω, τα μοντέλα Μηχανικής Μάθησης αξιοποιούν το μεγάλο όγκο δεδομένων για την εκπαίδευση τους αλλά και για την παραγωγή αναλύσεων σε πραγματικό χρόνο με σκοπό την εξαγωγή χρήσιμων πληροφοριών σε σχέση με την κατάσταση του εξοπλισμού αλλά και την δημιουργία ειδοποιήσεων σε περίπτωση εντοπισμού ανωμαλιών για την υποστήριξη λήψης αποφάσεων. Οι εφαρμογές Μηχανικής Μάθησης είναι αποτελεσματικό εργαλείο για την ανάλυση μεγάλων ροών δεδομένων, ωστόσο είναι απαραίτητη η ορθή διαστασιολόγηση του περιβάλλοντος που θα φιλοξενήσει την συγκεκριμένη εφαρμογή αλλά και της πλατφόρμας ΔτΠ που θα συγκεντρώνει και θα διαχειρίζεται την ροή δεδομένων. Η πληθώρα των δεδομένων που θα παράγονται από τον βιομηχανικό εξοπλισμό και από τις συσκευές βιομηχανικού ΔτΠ σε συνδυασμό με τους πόρους που θα καταλαμβάνει η επεξεργασία των δεδομένων σε πραγματικό χρόνο καθώς και η παραγωγή αναλύσεων από τα μοντέλα Μηχανικής Μάθησης, θα μπορούσαν δυνητικά να εξαντλήσουν τους επεξεργαστικούς πόρους του συστήματος. Ταυτόχρονα, η συνεχής ροή απαιτεί όχι μόνο μεγάλη αποθηκευτική ικανότητα αλλά και υψηλή διαθεσιμότητα και αξιοπιστία του δικτύου, για να διασφαλιστεί η αδιάλειπτη πρόσβαση στα δεδομένα και η συνεχής λειτουργία των εφαρμογών Μηχανικής Μάθησης. Οι πλατφόρμες θα πρέπει να είναι εξοπλισμένες με δυνατότητες επεκτασιμότητας και

ανθεκτικότητας προκειμένου να μπορούν να διαχειρίζονται τις αυξημένες απαιτήσεις των βιομηχανικών περιβαλλόντων και να υποστηρίζουν την σταθερή παροχή υπηρεσιών σε πραγματικό χρόνο.

Οι πλατφόρμες Διαδικτύου των πραγμάτων αντιμετωπίζουν σημαντικές προκλήσεις όσο αφορά την ασφάλεια, με την γρήγορη ανάπτυξη του IIoT σε διάφορους τομείς επιβάλλεται μία ολοκληρωμένη προσέγγιση που να περιλαμβάνει συγκεκριμένα πρότυπα ασφάλειας[17]. Η υιοθέτηση και εγκατάσταση διαφορετικών βιομηχανικών συσκευών ΔτΠ εισάγει ενδεχόμενες τρωτότητες στις επιχειρήσεις καθώς οι συσκευές αυτές δεν έχουν σχεδιαστεί με στόχο την ασφάλεια ως προτεραιότητα, αποτελώντας εύλογο στόχο για επιθέσεις. Ακόμα, η συνένωση των δικτύων του εργοστασίου (Operation Technology - OT) με τα δίκτυα της πληροφορικής (Information Technology – IT) επέτρεψε από την μία τα δεδομένα που παράγονται από το OT να «εκδημοκρατιστούν», μεταφέροντας τα σε ανώτερο επίπεδο στο IT για περαιτέρω ανάλυση και επεξεργασία απλουστεύοντας τη διαχείριση εξελιγμένων περιβαλλόντων, και από την άλλη να συγκεντρώνει νέους κινδύνους για την ασφάλεια. Ένας αλγόριθμος Μηχανικής Μάθησης θα πρέπει να επικοινωνήσει με το συγκεκριμένο δίκτυο προκειμένου να λάβει τα δεδομένα σε πραγματικό χρόνο για ανάλυση, δημιουργώντας με την σειρά του επιπρόσθετους κινδύνους όπως Denial of Service, Buffer Overflows, κώδικας της «πίσω πόρτας» κλπ. Επιπλέον, η διασύνδεση του OT πεδίου με τα νέα δίκτυα του IT καθώς και με νέες εφαρμογές συσκευών βιομηχανικού Διαδικτύου των Πραγμάτων ενέχει των κίνδυνο έκθεσης παλαιού τύπου συσκευών και εφαρμογών, τα οποία περιέχουν παρωχημένα μέτρα προστασίας, σε πιθανούς εισβολείς οι οποίοι θα προσπαθήσουν να εκμεταλλευτούν τις συγκεκριμένες τρωτότητες. Η βασική πρόκληση επομένως, είναι η χρησιμοποιούμενη πλατφόρμα να περιέχει συγκεκριμένες διαδικασίες και πρότυπα ασφαλείας για την απομόνωση και διασφάλιση της προστασίας του κατώτατου επιπέδου OT. Η πτυχές αυτές πρέπει να περιλαμβάνουν ένα σύνολο καλά προκαθορισμένων διαδικασιών και στόχων, όπως προστατευμένες συνδέσεις δικτύου, ανάθεση συγκεκριμένων ρόλων κατά την σύνδεση του χρήστη, ελάχιστα δικαιώματα, δημιουργία επιπλέον ζώνης δικτύου (για την προστασία του εσωτερικού δικτύου προσθέτοντας ένα επιπλέον επίπεδο ασφαλείας, Demilitarized Zone – DMZ), τοίχοι προστασίας (Firewalls) είναι κάποιες από τις προστασίες που θα πρέπει να είναι σε θέση η πλατφόρμα να εξασφαλίσει ή και να υιοθετήσει. Ωστόσο, δεν θα πρέπει να παραγνωρίζεται η σημασία του ανθρώπινου παράγοντα, οι χρήστες της πλατφόρμας θα πρέπει είναι κατάλληλα καταρτισμένοι και εκπαιδευμένοι προκειμένου να γνωρίζουν τους ενδεχόμενους κινδύνους ασφαλείας και να είναι σε θέση να αναγνωρίζουν πιθανές τρωτότητες του συστήματος που σχετίζονται με τη συλλογή, διαχείριση, ανάλυση και μεταφορά πληροφοριών.

Μία ακόμη σημαντική πρόκληση που πρέπει να αντιμετωπίσουν οι πλατφόρμες Διαδικτύου των πραγμάτων αφορά την ικανότητα τους να συνδέονται και να ανταλλάσσουν σημαντικές πληροφορίες, αποτελεσματικά με διαφορετικά συστήματα και πλατφόρμες. Ένα βιομηχανικό περιβάλλον χαρακτηρίζεται από μεγάλη ετερογένεια, ανομοιογενείς συσκευές κατασκευάζονται από διαφορετικούς προμηθευτές και χρησιμοποιούν διαφορετικές τεχνολογίες και πρωτόκολλα επικοινωνίας. Επιπλέον, οι βιομηχανίες αξιοποιούν ποικίλα λογισμικά ελέγχου, διαχείρισης και οργάνωσης της παραγωγής, όπως MES (Manufacturing Execution System), ERP (Enterprise Resource Planning) και SCADA (Supervisory Control and Data Acquisition) τα οποία επίσης αναπτύσσονται από διαφορετικούς κατασκευαστές. Συνεπώς, θα πρέπει να θεσπιστεί ένα σύνολο κανόνων και πρωτοκόλλων που θα επιτρέπει στις διαφορετικές συσκευές και λογισμικά να

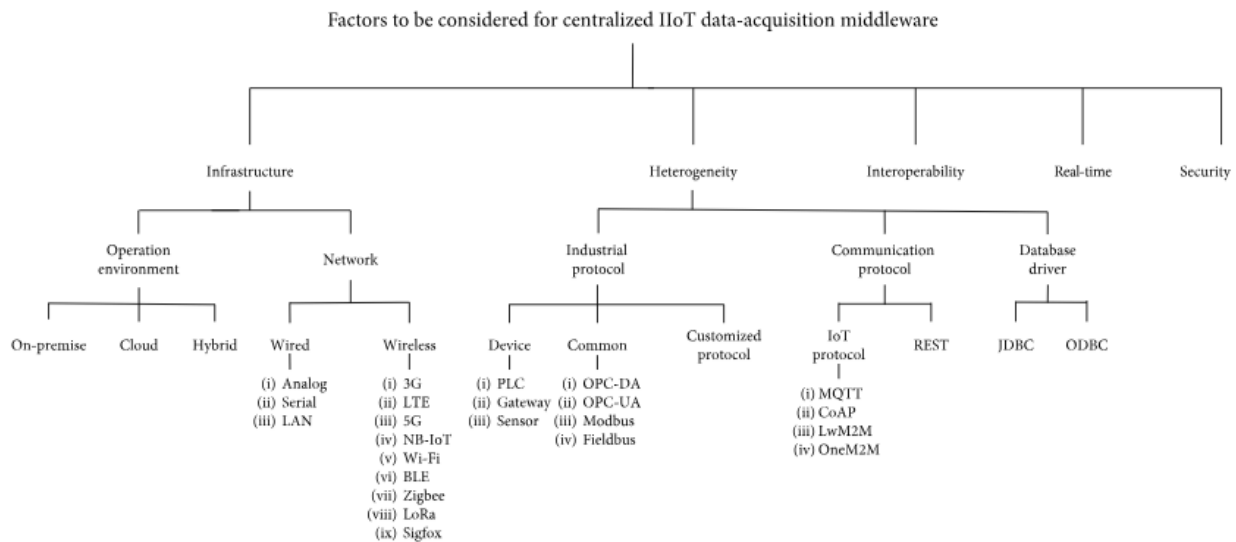
επικοινωνούν μεταξύ τους, γεφυρώνοντας τις όποιες αντιθέσεις και ενσωματώνοντας δεδομένα με οργανωμένο και αυτόματο τρόπο. Η απρόσκοπτη ανταλλαγή πληροφοριών είναι καθοριστικής σημασίας για την διαμόρφωση ενός συστήματος ελέγχου και παρακολούθησης σε πραγματικό χρόνο της κατάστασης του εξοπλισμού. Κατά συνέπεια, είναι αναγκαία η ανάπτυξη διαδικτυακών υπηρεσιών (Web Services), αυτόνομων λογισμικών, τα οποία θα επιτρέπουν την αυτόματη ανταλλαγή πληροφοριών και θα διασυνδέουν διαφορετικά συστήματα μεταξύ τους. Η ενσωμάτωση των ετερογενών στοιχείων θα παρέχει μία πολύ-επίπεδη προσέγγιση, διευκολύνοντας την ενοποίηση υφιστάμενων εφαρμογών με μελλοντικά δίκτυα, συστήματα, εξοπλισμούς και πλατφόρμες.

3.2 Κύρια Χαρακτηριστικά Πλατφόρμας ΔτΠ

Ένα βιομηχανικό περιβάλλον περιέχει εγγενής πολυπλοκότητα, η οποία οφείλεται στους διαφορετικούς τύπους εξοπλισμού, στην ποικιλία λογισμικών και συστημάτων που αναπτύσσονται για την κάλυψη εξατομικευμένων αναγκών, στις σύνθετες παραγωγικές διαδικασίες και τα πολλαπλά τους στάδια, στην ετερογένεια των προϊόντων καθώς και στις ιδιαίτερες απαιτήσεις της εκάστοτε βιομηχανίας. Οι παραπάνω παράγοντες συνδυάζονται για να δημιουργήσουν ένα πολύπλοκο περιβάλλον που απαιτεί πλατφόρμες υψηλής απόδοσης ικανές να διαχειρίζονται πολλαπλές συνδέσεις ταυτόχρονα ενοποιώντας διαφορετικά συστήματα, να αποθηκεύουν και να επεξεργάζονται μεγάλο όγκο δεδομένων με διαρκής ροή. Στη συνέχεια τα δεδομένα αυτά θα αποστέλλονται σε αλγόριθμους Μηχανικής Μάθησης οι οποίοι θα αξιοποιούν την πληροφορία και θα υπολογίζουν την κατάσταση του εξοπλισμού. Συνεπώς, η πλατφόρμα θα λειτουργεί ως ένα ενδιάμεσο λογισμικό (middleware) το οποίο όχι μόνο θα λαμβάνει δεδομένα από διάφορες πηγές, αλλά θα τα τροφοδοτεί και στο σύστημα Πρόγνωσης και Διαχείρισης Υγείας. Η πλατφόρμα θα πρέπει να διασφαλίζει την ποιότητα, την ακεραιότητα και την ακρίβεια των δεδομένων, επιτρέποντας στους αλγόριθμους να αναλύσουν τις πληροφορίες με επιτυχία παράγοντας αξιόπιστα αποτελέσματα. Η αποτελεσματική λειτουργία της πλατφόρμας καθορίζει την επιτυχία και την ακρίβεια των αλγόριθμων Μηχανικής Μάθησης, καθιστώντας την ένα κρίσιμο παράγοντα για την βελτίωση της απόδοσης, την εξασφάλιση της αδιάλειπτης λειτουργίας και την ασφάλεια της εφαρμογής. Τα κύρια χαρακτηριστικά της περιγραφόμενης πλατφόρμας παρουσιάζονται στην Εικόνα 8 και αφορούν: 1) την Συμβατότητα με ενσύρματα ή και ασύρματα Δίκτυα, 2) την υποστήριξη Ετερογενών συστημάτων και διαφορετικών πρωτοκόλλων, 3) την εξασφάλιση της διαλειτουργικότητας επιτρέποντας την ανταλλαγή πληροφοριών μεταξύ συσκευών και λογισμικών, 4) την διασφάλιση της συνεχούς ροής δεδομένων σε πραγματικό χρόνο 5) και την ενσωμάτωση προηγμένων μέτρων ασφαλείας για την προστασία των δεδομένων αλλά και του εξοπλισμού του εργοστασίου.

Υποδομή (Infrastructure): Το αρχικό χαρακτηριστικό αφορά στη προσαρμοστικότητα και διαφοροποίηση της λύσης του δικτύου βάσει των αναγκών και ιδιαίτερων απαιτήσεων της κάθε βιομηχανίας. Η προτεινόμενη, εξατομικευμένη αρχιτεκτονική λύση αφορά την σύνδεση του κατώτατου επιπέδου των διάφορων αισθητήρων, συσκευών IoT, ελεγκτών κλπ. με το ανώτερο επίπεδο των διακομιστών. Επίσης, καθορίζεται η διαχείριση της ροής πληροφορίας μεταξύ των επιπέδων, ορίζοντας μια αφιερωμένη μονάδα για την αποθήκευση και ανάλυση των συλλεγόμενων δεδομένων καθώς οι περισσότερες συσκευές δεν διαθέτουν τους απαραίτητους πόρους για την

εκτέλεση διεργασιών μεγάλης έντασης. Ο σωστός προσδιορισμός των ιδιαίτερων απαιτήσεων και χαρακτηριστικών της κάθε περίπτωσης, σε συνδυασμό με την ανάλυση της ροής των δεδομένων, αποτελούν αναγκαία βήματα για την οργάνωση της αρχιτεκτονικής του δικτύου, την διαστασιολόγηση των μονάδων και τον καθορισμό των κανόνων και προτύπων ασφάλειας της πλατφόρμας. Συνεπώς, ο σχεδιασμός και η συνδεσιμότητα των επιμέρους στοιχείων της πλατφόρμας μπορεί να επηρεάσει σημαντικά την απόδοση, την ταχύτητα εκτέλεσης και την ασφάλεια των αλγορίθμων Μηχανικής Μάθησης.



Εικόνα 8. Χαρακτηριστικά Πλατφόρμας IIoT, Πηγή: [43].

Λειτουργικό Περιβάλλον (Operation Environment): Διακρίνονται τρεις σχεδιαστικές λύσεις: 1) Κεντρικοποιημένη γύρω από μία τοπική συσκευή, 2) Συνδεδεμένη με έναν διακομιστή στο Cloud 3) και υβριδική, όπου είναι ένας συνδυασμός των δύο προηγούμενων. Για παράδειγμα, μια βιομηχανία μπορεί να επιλέξει τα συλλεγόμενα δεδομένα να επεξεργάζονται και να αποθηκεύονται σε μια πλατφόρμα που βρίσκεται σε έναν τοπικό διακομιστή, χωρίς αυτά να μεταφέρονται προς το ανώτερο επίπεδο του IT. Από την άλλη, η ίδια βιομηχανία μπορεί να αποφασίσει να τα συλλέγει τοπικά και να τα αποστέλλει στο Cloud για ανάλυση. Τέλος, μπορεί να επιλέξει το υβριδικό μοντέλο κατά το οποίο θα συλλέγει και θα επεξεργάζεται τα δεδομένα τοπικά, ενώ θα μεταφέρει ένα μέρος ή το σύνολο το συλλεγμένων πληροφοριών σε ένα διακομιστή στο Cloud για περαιτέρω ανάλυση.

Κεντρικός Διακομιστής (On-Premise): Στη παρούσα σχεδιαστική λύση τα δεδομένα παραμένουν σε ένα Κεντρικό διακομιστή, που ενδεχομένως να μην έχει και σύνδεση με το Διαδίκτυο, για τοπική επεξεργασία και αποθήκευση. Ωστόσο το κόστος αποτελεί ένα μειονέκτημα της υλοποίησης αυτής, καθώς χρειάζεται να εξαγοραστεί ένας ιδιόκτητος κόμβος προκειμένου να εγκατασταθεί η πλατφόρμα και να εκτελεί τις απαιτούμενες αναλύσεις τόσο του προγράμματος όσο και των αλγορίθμων Μηχανικής Μάθησης. Επομένως, κρίνεται απαραίτητο να διαστασιολογηθεί σωστά ο κόμβος ώστε να έχει την αναγκαία επεξεργαστική ισχύ για να διαχειρίζεται με επιτυχία τον αυξημένο φόρτο εργασίας. Να σημειωθεί, ότι σε περίπτωση παραβίασης ο επιτιθέμενος ενδέχεται να έχει πρόσβαση σε όλα τα δεδομένα, κατά συνέπεια είναι ζωτικής σημασίας η προστασία του συγκεκριμένου κόμβου, όπως και του κατώτατου επιπέδου του Δικτύου με την προσθήκη επιπλέον επιπέδων ασφαλείας όπως η DMZ.

Διακομιστής στο Cloud: Η συγκεκριμένη τοπολογία έχει υιοθετηθεί λόγω της οικονομικότητας και αποδοτικότητας του, βασίζεται στη συνδεσιμότητα στο Διαδίκτυο και στην χρησιμοποίηση διακομιστών στο Cloud. Τα βασικά πλεονεκτήματα είναι η αμεσότητα της υλοποίησης, η διαθεσιμότητα πόρων και η δυνατότητα διασύνδεσης της πλατφόρμας με συσκευές και εφαρμογές διαφορετικών κατασκευαστών, όπως αξιοποίηση ολοκληρωμένων λύσεων Μηχανικής Μάθησης (Advanced Analytics, Visplore κλπ.). Όμως, η σύνδεση με το Διαδίκτυο εισάγει πιθανές τρωτότητες που ένας εν δυνάμει επιτιθέμενος ενδεχομένως να εκμεταλλευτεί. Οι κίνδυνοι ασφάλειας και ιδιωτικότητας αυξάνονται καθώς τα δεδομένα μπορεί να υποκλαπούν, να διαγραφούν, να τροποποιηθούν ενώ παράλληλα ο επιτιθέμενος ενδέχεται να χρησιμοποιήσει τον κόμβο ως ενδιάμεσο βήμα για να επιτεθεί στο πιο κάτω επίπεδο (Denial of Service, Ping Flood κλπ.).

Υβριδική Τοπολογία (Hybrid): Η υβριδική τοπολογία συνδυάζει τα πλεονεκτήματα και από τις δύο εφαρμογές προσφέροντας μεγαλύτερη αξιοπιστία και διαθεσιμότητα σε ενδεχόμενη απουσία σύνδεσης με το Διαδίκτυο. Επιπλέον, ο φόρτος εργασίας μπορεί να διαμοιραστεί μεταξύ τοπικού κόμβου και διακομιστή Cloud. Δεδομένα που κρίνονται πιο σημαντικά ή ευαίσθητα μπορούν να παραμείνουν τοπικά, ενώ παρέχεται και η δυνατότητα σύνδεσης της πλατφόρμας με εφαρμογές τρίτων. Παρόλα αυτά και η συγκεκριμένη υλοποίηση αντιμετωπίζει προκλήσεις ασφάλειας και ιδιωτικότητας, καθώς είναι απαραίτητο να προσφέρει προστασία τόσο στα χαμηλότερα επίπεδα του OT όσο και να διασφαλίσει την αξιοπιστία και την προάσπιση των δεδομένων.

Δίκτυο (Network): Ο Δεύτερος υπό-παράγοντας που είναι καθοριστικής σημασίας για τον σχεδιασμό της υποδομής είναι το Δίκτυο. Τα δίκτυα των εγκαταστάσεων χωρίζονται σε 1) *Ενσύρματα (Wired Networks)* και 2) *Ασύρματα (Wireless Networks)*. Τα πρώτα διακρίνονται από μικρή καθυστέρηση και υψηλή αξιοπιστία, ενώ επιπλέον χαρακτηρίζονται από υψηλό κόστος εγκατάστασης και συντήρησης και περιορισμούς όσον αφορά την ευελιξία. Από την άλλη τα βιομηχανικά ασύρματα δίκτυα έχουν απλή εγκατάσταση και συντήρηση καθώς και αυξημένη ευελιξία και κινητικότητα [18]. Η βιομηχανία υιοθετεί κατά κανόνα την ενσύρματη επικοινωνία, κυρίως λόγω των πλεονεκτημάτων που έχει σε σχέση με την μικρή καθυστέρηση και τη υψηλή αξιοπιστία, διασφαλίζοντας σταθερότητα, γρήγορη απόκριση και υψηλή ακρίβεια στον έλεγχο κρίσιμων βιομηχανικών συστημάτων. Τα δεδομένα και οι εντολές από και προς το πεδίο συνήθως λαμβάνονται από διασυνδέσεις μέσω φυσικών καλωδίων, όπως είναι για παράδειγμα η αναλογική επικοινωνία (analog) όπου τα σήματα μεταδίδονται ως συνεχή κύματα μέσω τηλεφωνικών γραμμών ή σειριακή επικοινωνία (serial) όπου η πληροφορία μεταφέρεται σειριακά μέσω bit καθώς με τοπικά δίκτυα (Local Area Networks - LAN) τα οποία επιτρέπουν τη σύνδεση πολλαπλών συσκευών σε μικρές γεωγραφικές περιοχές μέσω διαφορετικών τοπολογιών (αστέρι, bus, δενδροειδής κλπ.) χρησιμοποιώντας τεχνολογίες όπως το Ethernet [19]. Ωστόσο, με την ανάπτυξη του βιομηχανικού ΔτΠ, όλο και περισσότερες επιχειρήσεις αξιοποιούν ασύρματα δίκτυα (Industrial Wireless Networks – IWN) μικρής απόστασης, για την μετάδοση μικρών δεδομένων και με χαμηλή κατανάλωση ισχύος, όπως το Bluetooth Low Energy (BLE), το Wi-Fi, το ZigBee κλπ. Τα IWN εισέρχονται σταδιακά στις βιομηχανίες και μετατρέπονται σε σημαντικά στοιχεία για την Industry 4.0 και την μετάβαση στην ψηφιακή εποχή. Τα δεδομένα συλλέγονται από ένα κεντρικό κόμβο (Gateway/Access point) το οποίο λειτουργεί και ως γέφυρα ένωσης με τα υπόλοιπα δίκτυα [20]. Επιπλέον, ο κεντρικός κόμβος είναι υπεύθυνος για την εποπτεία των υπόλοιπων συσκευών του Δικτύου, την ασφάλεια συμπεριλαμβανομένου της κρυπτογράφησης των δεδομένων, την συνδεσιμότητα με τις διαφορετικές συσκευές-κόμβους ΔτΠ υποστηρίζοντας το ανάλογο

πρωτόκολλο βάσει συσκευής, καθώς και την αξιοποίηση των δεδομένων αποθηκεύοντας τα προσωρινά σε μια τοπική βάση ή προωθώντας τα στον κεντρικό διακομιστή. Σε κάθε περίπτωση, η πλατφόρμα θα πρέπει να επιτρέπει την απόκτηση δεδομένων τόσο μέσω ενσύρματης όσο και ασύρματης επικοινωνίας.

Ετερογένεια (Heterogeneity): Ένα δεύτερο χαρακτηριστικό που πρέπει να διαθέτει μια πλατφόρμα προκειμένου να αποτελέσει μια κατάλληλη βάση για την χρήση Μηχανικής Μάθησης είναι η ετερογένεια. Η συγκεκριμένη ιδιότητα αναφέρεται στην ικανότητα της βιομηχανικής πλατφόρμας ΔτΠ να προσαρμόζεται βάσει των τεχνολογικών αναγκών και απαιτήσεων της εγκατάστασης, ενσωματώνοντας δεδομένα από ανόμοια συστήματα, διαφορετικά πρωτόκολλα επικοινωνίας και βάσεις δεδομένων.

Βιομηχανικό Πρωτόκολλο (Industrial Protocol): Αναφέρεται σε βιομηχανικά πρωτόκολλα επικοινωνίας που χρησιμοποιούνται για τη διασύνδεση και την ανταλλαγή δεδομένων μεταξύ διαφορετικών βιομηχανικών συσκευών. Διαχωρίζονται περαιτέρω σε: 1) *επίπεδο συσκευής*, όπου διαφοροποιούνται αναλόγως τις απαιτήσεις (Εύρος Ζώνης, Συχνότητα Δειγματοληψίας, Αριθμός Υποστηριζόμενων Συσκευών, Χρόνος απόκρισης κλπ.) αλλά και τους τύπους των συστημάτων όπως αισθητήρες/ενεργοποιητές, ελεγκτές, Programming Logic Controllers (PLC) κλπ. [21], 2) *κοινά πρωτόκολλα* τα οποία χρησιμοποιούνται ευρέως στη βιομηχανία για την διασύνδεση και μετάδοση σημάτων όπως OPC DA, UA, XML κλπ. τα οποία παρέχουν μια ενιαία διεπαφή για την ανταλλαγή και μετάδοση δεδομένων μεταξύ συστημάτων, λειτουργώντας ως ένα είδος μεσαίου λογισμικού (middleware) [22] και 3) *προσαρμοσμένα πρωτόκολλα*, τα οποία αναφέρονται σε ειδικές διαμορφώσεις που αναπτύσσονται για την εξυπηρέτηση συγκεκριμένων απαιτήσεων.

Πρωτόκολλα Επικοινωνίας (Communication Protocol): Περιλαμβάνει τις μεθόδους και τα πρότυπα που χρησιμοποιούνται από τις συσκευές βιομηχανικού ΔτΠ για την αξιόπιστη και αποδοτική μεταφορά δεδομένων. Η επιλογή των κατάλληλων πρωτοκόλλων επικοινωνίας είναι κρίσιμη για την εξασφάλιση της απρόσκοπτης λειτουργίας των βιομηχανικών συστημάτων, τη συνεχή ροή δεδομένων προς τον διακομιστή της πλατφόρμας και των προγραμμάτων Μηχανικής Μάθησης, καθώς και της ενσωμάτωσης συσκευών ΔτΠ στο κεντρικό Δίκτυο. Μερικά από τα πιο σημαντικά πρωτόκολλα ανταλλαγής μηνυμάτων είναι: το *Message Queuing Telemetry Transport (MQTT)*, το οποίο χρησιμοποιεί το μοντέλο «publish/subscribe» για αξιόπιστες επικοινωνίες μεταξύ μηχανών (M2M) και για την ανάπτυξη συσκευών ΔτΠ υψηλής απόδοσης με χαμηλή μνήμη και μικρή επεξεργαστική ισχύ, το *Constrain Application Protocol (CoAP)* το οποίο λειτουργεί βάσει «request/response» - το πιο δημοφιλές από τα πρότυπα «client/server»- κατάλληλο για αισθητήρες με περιορισμένη ενέργεια, το *Lightweight M2M (LwM2M)* βασισμένο στο CoAP για την διαχείριση των συσκευών ΔτΠ υποστηρίζοντας την επικοινωνία «client/server» [23].

Επιπλέον, μια ακόμη υπό-κατηγορία είναι η «Μεταφορά Καταστάσεων Αναπαράστασης» (*Representational State Transfer – REST*), το οποίο βασίζεται σε επικοινωνία «client/server» και επιτρέπει στους χρήστες να έχουν πρόσβαση στα αποθηκευμένα δεδομένα ή και να προσθέτουν εγγραφές στον διακομιστή. Χρησιμοποιεί το πρότυπο HTTP προκειμένου να εκτελεί requests μέσω σύνθεσης κατάλληλων Uniform Resource Identifiers (URL) αξιοποιώντας βασικές μεθόδους, όπως GET για την προσπέλαση των δεδομένων και πληροφοριών συστήματος, POST για την δημιουργία πόρων, PUT για την εγγραφή ενημερώσεων κλπ. καθώς και ειδικών «Controllers» που έχουν αναπτυχθεί για την διαχείριση και οργάνωση των αιτημάτων. Επομένως, το REST είναι ένα πλαίσιο

μέσα στο οποίο ο διακομιστής με την εφαρμογή που εκτελεί τα αιτήματα (requests) ανταλλάσσουν πόρους μέσω URLs χρησιμοποιώντας HTTP λειτουργίες. Το REST είναι stateless δηλαδή ο διακομιστής δεν διατηρεί καμία πληροφορία για την κατάσταση του client μεταξύ των αιτημάτων, ενώ οι εφαρμογές που εκτελούν τα αιτήματα διαφοροποιούνται αναλόγως τις απαιτήσεις της εφαρμογής καθώς κυμαίνονται από απλές εντολές σε κάποιο Web Browser μέχρι την χρήση εξατομικευμένων προγραμμάτων τα οποία λειτουργούν ως middleware (Custom Code). Η ανάπτυξη των REST προσδίδει μια εναλλακτική για την επικοινωνία των συσκευών ΔτΠ και όχι μόνο, με ένα κεντρικό διακομιστή [24].

Οδηγός Βάσης Δεδομένων (Database Driver): Παρότι οι τεχνολογικές εξελίξεις υποχρεώνουν τις βιομηχανίες να κατευθύνονται προς την ψηφιοποίηση των δεδομένων και την ενσωμάτωση νέων πρωτοποριακών συστημάτων και εφαρμογών, η ανάγκη σύνδεσης με παλαιού τύπου εφαρμογών (Legacy Applications) και των βάσεων δεδομένων τους είναι κεντρική. Συνεπώς, η επιλεγόμενη πλατφόρμα θα πρέπει να επιτρέπει την σύνδεση με οδηγούς βάσεων δεδομένων όπως το *Java Database Connectivity (JDBC)* και το *Open Database Connectivity (ODBC)* προκειμένου να επιτευχθεί η άντληση των ιστορικών δεδομένων από τα Legacy Applications προς τη πλατφόρμα και κατ' επέκταση την τροφοδότηση των αλγορίθμων Μηχανικής Μάθησης.

Δια-λειτουργικότητα (Interoperability): Ένα σημαντικό επίσης ζήτημα σε αυτά τα περιβάλλοντα είναι η διασφάλιση της δια-λειτουργικότητας, δηλαδή της ανταλλαγής δεδομένων με παλαιότερα συστήματα πληροφορικής, χρησιμοποιώντας τυποποιημένα πρωτόκολλα. Η ενσωμάτωση διαφορετικών βιομηχανικών εξοπλισμών, ανομοιογενών διατάξεων και αρχιτεκτονικών αποτελεί μία βασική πρόκληση που θα πρέπει να είναι σε θέση να αντιμετωπίσει η επιλεγόμενη πλατφόρμα, λόγω της ποικιλίας των φυσικών διαύλων, των πρωτοκόλλων επικοινωνίας και των λειτουργικών συστημάτων. Σε μία βιομηχανική εγκατάσταση είναι σύνηθες να χρησιμοποιούνται συσκευές από διαφορετικούς κατασκευαστές αυξάνοντας την πολυπλοκότητα διασύνδεσης τους, ωστόσο η δια-λειτουργικότητα διασφαλίζει την μεταξύ τους επικοινωνία και συνεργασία, διευκολύνοντας την ανταλλαγή δεδομένων. Επιπλέον, εισάγει την απαραίτητη ευελιξία και προσαρμοστικότητα στη πλατφόρμα βιομηχανικού ΔτΠ εξασφαλίζοντας τη συμβατότητα τόσο με τα Legacy Applications όσο και με νέα αναπτυσσόμενα συστήματα. Καθώς η βιομηχανία θα επεκτείνεται προκειμένου να καλύψει τις απαιτήσεις της, νέα συστήματα θα εντάσσονται ενώ παράλληλα θα ενημερώνονται τα παλιά. Κατά συνέπεια, η πλατφόρμα θα πρέπει να διαφυλάττει την απρόσκοπτη και αδιάληπτη μετάδοση σημάτων επιτρέποντας παράλληλα την προσθήκη νέων σημάτων και πληροφοριών προς τα προγράμματα Μηχανικής Μάθησης.

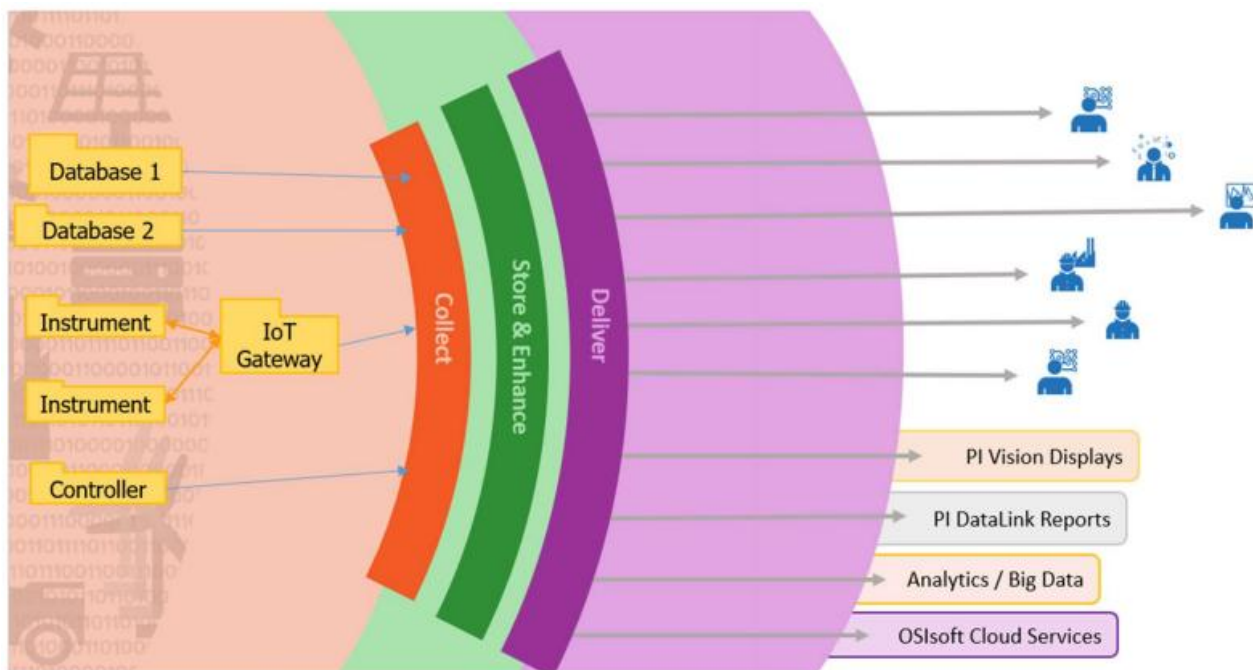
Δεδομένα σε πραγματικό χρόνο (Real time): Η ενημέρωση των δεδομένων του συστήματος σε πραγματικό χρόνο αποτελεί ένα επιπλέον σημαντικό χαρακτηριστικό καθώς αντικατοπτρίζει την λειτουργία των στοιχείων της κάθε μονάδας και παρουσιάζει την συνολική ή μεμονωμένη κατάσταση του εξοπλισμού του εργοστασίου. Ένα σύστημα παρακολούθησης και διαχείρισης της υγείας του εξοπλισμού πρέπει να είναι σε θέση να αναλύει τα δεδομένα σε πραγματικό χρόνο και να προβλέπει την μελλοντική κατάσταση της μηχανής καθώς και τον εναπομείναντα ωφέλιμο χρόνο ζωής. Η δυνατότητα παρακολούθησης και ανάλυσης των δεδομένων σε άμεσο χρόνο επιτρέπει επίσης την έγκαιρη ανίχνευση προβλημάτων και την ταχύτατη ενημέρωση, εξασφαλίζοντας την ελαχιστοποίηση των εξαναγκασμένων κρατήσεων και βελτιώνοντας την συνολική απόδοση του συστήματος. Ακόμα, η ενσωμάτωση δεδομένων από πολλαπλές πηγές σε πραγματικό χρόνο σε μία πλατφόρμα, επιτρέπει μια ολοκληρωμένη και σφαιρική προσέγγιση στη διαχείριση των πόρων και

διαδικασιών. Η κοινή χρήση εργαλείων οπτικοποίησης δεδομένων και γραφικής αναπαράστασης συστημάτων, παρέχει την δυνατότητα συντονισμού και οργάνωσης των εργασιών διευκολύνοντας τον καταμερισμό της εργασίας.

Ασφάλεια (Security): Τέλος, είναι απαραίτητο η πλατφόρμα βιομηχανικού ΔτΠ να διασφαλίζει την εμπιστευτικότητα, ακεραιότητα και διαθεσιμότητα των δεδομένων και των συστημάτων. Με την επέκταση νέων συσκευών και εφαρμογών ΔτΠ εισάγονται στο σύστημα πιθανές τρωτότητες, η πλατφόρμα θα πρέπει να εξασφαλίζει την ορθή και αποδοτική επικοινωνία μεταξύ υπηρεσιών με τρόπο που να διασφαλίζει την ακεραιότητα των δεδομένων και την αυθεντικότητα των επικοινωνιών. Επομένως, κρίνεται αναγκαία η εφαρμογή σύγχρονων τεχνικών κρυπτογράφησης και πιστοποίησης όπως κλειδιά ασφαλείας. Η πλατφόρμα πρέπει να διαθέτει μηχανισμούς για την ανίχνευση και αντιμετώπιση πιθανών κινδύνων, καθώς και να ενσωματώνει κατάλληλα πρότυπα απόδοσης ελάχιστων δικαιωμάτων προσδίδοντας στους χρήστες συγκεκριμένες δυνατότητες.

3.3 Επιλεγόμενη Πλατφόρμα: PI System

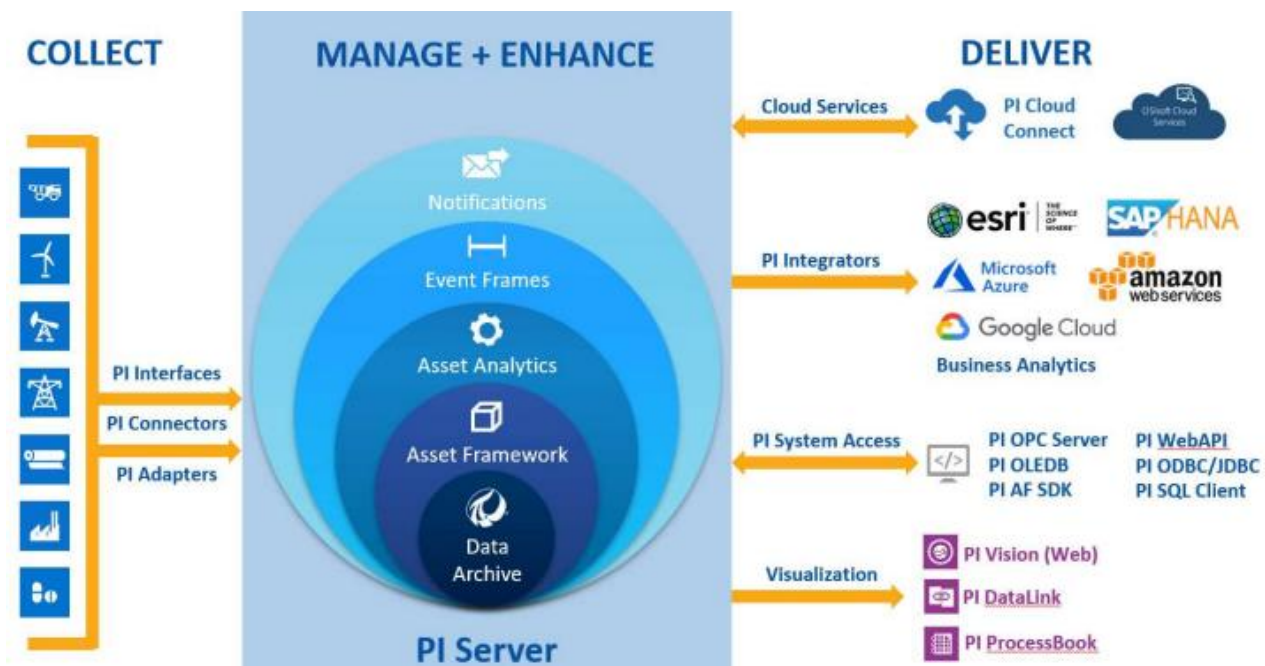
Σε ένα δυναμικά αναπτυσσόμενο τομέα η ικανότητα ενός λογισμικού να συνδυάζει και να διαχειρίζεται δεδομένα από ανομοιογενή συστήματα και διαφορετικά πρωτόκολλα επικοινωνίας προσφέροντας ασφάλεια και δια-λειτουργικότητα είναι ζωτικής σημασίας. Η επιλεγόμενη πλατφόρμα θα πρέπει να λειτουργεί ως ένα σύστημα διαχείρισης πληροφοριών, το οποίο θα παρέχει



Εικόνα 9. Αρχιτεκτονική PI System, Πηγή: [44].

κρίσιμα δεδομένα σε πραγματικό χρόνο, που θα χρησιμοποιούνται από προγράμματα Μηχανικής μάθησης για τον προσδιορισμό της κατάστασης του εξοπλισμού και θα παρέχουν υποστήριξη στη λήψη στρατηγικών και επιχειρησιακών αποφάσεων. Προκειμένου η εφαρμογή να μπορεί να ανταποκριθεί στις προκλήσεις που η σύγχρονη εποχή επιτάσσει, πρέπει να εμπεριέχει χαρακτηριστικά όπως αυτά περιεγράφηκαν στη προηγούμενη ενότητα.

Για την υλοποίηση μίας εφαρμογής προβλεπτικής συντήρησης και διαχείρισης υγείας, η παρούσα Διπλωματική εργασία θα αξιοποιήσει τις δυνατότητες που προσφέρει η πλατφόρμα *PI System (PI)*. Το *PI System (Plant Information – PI)* είναι μία πλατφόρμα λογισμικού που αναπτύχθηκε από την OSIsoft - η οποία εξαγοράστηκε από την AVEVA το 2020 - και δημιουργήθηκε με σκοπό την συλλογή, αποθήκευση, διανομή, ανάκτηση, ανάλυση και οπτικοποίηση των δεδομένων από διάφορες πηγές και συστήματα όπως παρουσιάζεται και από την αρχιτεκτονική της Εικόνας 9. Αρχικώς, το *PI System* συλλέγει δεδομένα από ανομοιογενείς πηγές, όπως βάσεις δεδομένων, ελεγκτές, ενεργοποιητές, συστήματα PLC, IIoT κλπ. μέσω της χρήσης κατάλληλων λογισμικών προγραμμάτων, τα οποία θα περιγράφουν στην συνέχεια. Τα δεδομένα που συγκεντρώνονται φέρουν χρονική σήμανση (*Timestamp*) παρελθόντος, παρόντος και μέλλοντος επιτρέποντας τη διαχρονική ανάλυση και περαιτέρω επεξεργασία των δεδομένων. Στη συνέχεια τα δεδομένα αποθηκεύονται και επεξεργάζονται στη κεντρική βάση δεδομένων, γνωστή και ως *PI Data Archive (PI DA)*, η οποία εξασφαλίζει την ασφαλή και αποτελεσματική καταχώρηση μεγάλων ποσοτήτων πληροφορίας. Για την ενίσχυση της απόδοσης και της επεξεργαστικής ισχύος, τα διάφορα στοιχεία του *PI* μπορούν να υπάρχουν χωριστά, ενώ η αρχιτεκτονική μπορεί να διαφοροποιείται από απλή έως πολύπλοκη αναλόγως τις απαιτήσεις και τις ανάγκες της κάθε βιομηχανικής μονάδας. Τέλος, τα δεδομένα είναι προσβάσιμα στον χρήστη με διαφορετικές εφαρμογές όπως μέσω *PI Vision Displays*, *PI Datalink Reports*, *OSIsoft Cloud Services*, *PI Web API* και άλλα.[25]. Στην εικόνα 10 παρουσιάζονται τα βασικά στοιχεία που αποτελούν ένα τυπικό σύστημα *PI*:



Εικόνα 10. Οι κατηγορίες και τα στοιχεία ενός τυπικού *PI System*, Πηγή: [44].

Συλλογή Δεδομένων (Data Collection): Αρχικά τα δεδομένα (ή αλλιώς *PI Tags*) συλλέγονται από υπολογιστές που έχουν εγκατεστημένο το απαραίτητο λογισμικό. Οι προσαρμογείς συνδέονται απευθείας με την πηγή και αποστέλλουν τα δεδομένα στον *PI Data Archive*. Υπάρχουν τρία είδη λειτουργικού επικοινωνίας, τα *PI Interfaces*, οι *PI Connectors* και οι *PI Adapters*. Κάθε ένα από τα τρία είδη έχει την δυνατότητα σύνδεσης με διαφορετικά συστήματα και πρωτόκολλα επικοινωνίας,

προσφέροντας προσαρμοστικότητα και ευελιξία λύσεων. Μέσω αυτών των εργαλείων, η πλατφόρμα μπορεί να απαντήσει στις προκλήσεις που θέτει ένα σύγχρονο βιομηχανικό περιβάλλον εξασφαλίζοντας την ετερογένεια και δια-λειτουργικότητα των συστημάτων. Τα PI Interfaces αποτελούν την παλιότερη οικογένεια των συγκεκριμένων εργαλείων και υποστηρίζουν σύνδεση με πολλά πρωτόκολλα επικοινωνίας όπως OPC DA, Modbus, RDBMS κλπ. Το PI Connector είναι ένα πιο εξελιγμένο εργαλείο το οποίο προσφέρει λύσεις για πολλαπλά πρωτόκολλα όπως OPC UA, MQTT κλπ. ενώ επιπλέον παρέχει ένα καλύτερο διαχειριστικό περιβάλλον για το χρήστη το οποίο βασίζεται στον ιστό. Τέλος, οι PI Adapters μπορούν να εγκατασταθούν με ευκολία και ταχύτητα τόσο σε συστήματα Windows όσο και Linux.

Διαχείριση και Ενίσχυση (Manage and Enhance): Τα δεδομένα που συλλέγονται από τις πηγές αποστέλλονται στον PI Server όπου αποθηκεύονται με ασφάλεια και είναι διαθέσιμα για επεξεργασία και ανάλυση. Ο PI Server αποτελείται από διαφορετικά εργαλεία σχεδιασμένα κατάλληλα για την αποτελεσματικότερη παρακολούθηση, ανάλυση και βελτιστοποίηση των βιομηχανικών διαδικασιών. Τα επιμέρους στοιχεία του συστήματος είναι:

PI Data Archive: Είναι μια βάση σχεδιασμένη ώστε να αποθηκεύει μεγάλα σύνολα δεδομένων σε πραγματικό χρόνο, επιτρέποντας την γρήγορη και αποδοτική ανάκτηση πληροφοριών είτε απευθείας, είτε μέσω ειδικών εργαλείων, ο χρήστης μέσω της συγκεκριμένης βάσης μπορεί να έχει πρόσβαση και σε ιστορικά δεδομένα παλαιότερων ετών. Επιπλέον, στο PI Data Archive μπορεί να οργανωθεί η Ασφάλεια (Security) για την διασφάλιση της εμπιστευτικότητας, ακεραιότητας και διαθεσιμότητας των δεδομένων, προσδίδοντας τα κατάλληλα δικαιώματα στο χρήστη κατά τον έλεγχο πρόσβασης.

Asset Framework: Το Asset Framework (AF) συμπληρώνει την αρχιτεκτονική παρέχοντας μια δομή για όλα τα δεδομένα που βρίσκονται στον οργανισμό, συνδέοντας τα PI Tags με στοιχεία των περιουσιακών στοιχείων (Assets). Διαθέτει ένα πλούσιο σύνολο λειτουργιών που επιτρέπει στους χρήστες να δημιουργούν και να οργανώνουν τα assets ιεραρχικά με τρόπο που να αντικατοπτρίζει την πραγματική δομή της βιομηχανικής μονάδας. Τα αποθηκευμένα δεδομένα που έχουν συλλεχθεί από διαφορετικές πηγές συνδέονται με τα αντίστοιχα assets και τα χαρακτηριστικά τους (Attributes) προκειμένου να αποκτήσουν νόημα και περιεχόμενο.

Asset Analytics: Αποτελεί μια κρίσιμη λειτουργία της πλατφόρμας που επιτρέπει στους χρήστες να δημιουργούν προσαρτημένες αναλύσεις βάσει αναγκών για τα Assets που εμπεριέχονται μέσα στην δομή του AF. Παρέχει την δυνατότητα ανάπτυξης προτύπων ανάλυσης (Analysis Templates) για την διευκόλυνση της διαχείρισης και επεξεργασίας πολλαπλών αναλύσεων με ένα τυποποιημένο και συνεπή τρόπο. Τυπικές εφαρμογές είναι η παρακολούθηση των βασικών δεικτών απόδοσης (KPI) αλλά και η παρακολούθηση της κατάστασης του εξοπλισμού βάσει συνθηκών (CBM).

Event Frames: Τα Event Frames αναπαριστούν σημαντικές χρονικές περιόδους κατά τις οποίες πραγματοποιούνται αλλαγές στις επιχειρηματικές δραστηριότητες, όπως αλλαγή βαρδιών, άδειασμα δεξαμενής, αντικατάσταση/επιδιόρθωση αντλίας κλπ. Μέσω των Event Frames, υπάρχει η δυνατότητα καταγραφής και ανάλυσης σημείων ενδιαφέροντος, εντός του χρονικού πλαισίου κατά το οποίο ήταν ενεργοποιημένο το συμβάν. Προσθέτει μία εύκολη εναλλακτική σε αντίθεση με την αναζήτηση βάσει χρόνου, επιτρέποντας στους χρήστες να λάβουν δεδομένα βάσει των συνθηκών που επικρατούσαν κατά την διάρκεια ενεργοποίησης του συμβάντος.

Notifications: Τα Notifications είναι μια επιπλέον δυνατότητα που προσφέρεται στο Asset Framework και επιτρέπει την άμεση ειδοποίηση των χρηστών κατά την ενεργοποίηση συγκεκριμένων Events, για το οποίο οι χρήστες θα πρέπει να λάβουν γνώση. Οι ειδοποιήσεις βοηθούν τους χρήστες να παρακολουθούν σημαντικές αλλαγές ή ανωμαλίες σε πραγματικό χρόνο και να λαμβάνουν άμεσα διορθωτικά μέτρα. Οι πληροφορίες αποστέλλονται είτε μέσω ηλεκτρονικού ταχυδρομείου είτε μέσω υπηρεσίας ιστού (*Web Services*), ενώ οι παραλήπτες έχουν την επιλογή να επιβεβαιώνουν την λήψη των ειδοποιήσεων (*Acknowledge*) σε περιπτώσεις όπου η άμεση παρέμβαση κρίνεται αναγκαία.

Παράδοση (Deliver): Το τελευταίο στάδιο αφορά την παράδοση των δεδομένων με ένα τρόπο ώστε να εξάγεται αξία για την επιχείρηση από την αξιοποίηση της συλλεχθείσας πληροφορίας. Τα στοιχεία που εμπεριέχονται σε αυτή την κατηγορία είναι:

Visualization: Αφορά τα εργαλεία που προσφέρουν δυνατότητα οπτικοποίησης των δεδομένων είτε απευθείας από τον PI Data Archive ως PI Tags είτε μέσω του Asset Framework και των Event Frames προσφέροντας πλήρη εικόνα μέσω γραφικών αναπαραστάσεων, συγκεντρωτικών πινάκων και διαγραμμάτων για την ανάλυση και παρουσίαση των πληροφοριών. Τα δύο βασικά εργαλεία οπτικοποίησης δεδομένων είναι: 1) το *PI Vision*, το οποίο είναι ένα προηγμένο εργαλείο για την δημιουργία γραφικών αναπαραστάσεων όπου ο χρήστης έχει την δυνατότητα να προσπελάσει τα δεδομένα μέσω προγράμματος περιήγησης (*browser*) και το 2) το *PI Datalink*, το οποίο είναι ένα πρόσθετο του Microsoft Excel για την ανάπτυξη προχωρημένων αναλύσεων και αναφορών.

PI Integrators: Χρησιμοποιούνται για την προετοιμασία και αποστολή των δεδομένων από το κεντρικό σύστημα του PI σε άλλο συστήματα καθιστώντας τα ιδανικά για εργαλεία επιχειρηματικής ευφυΐας (*Business Intelligence - BI*) και επιστήμης δεδομένων (*Data Science*). Τα δεδομένα μοντελοποιούνται, καθαρίζονται και παρουσιάζονται με κατάλληλα μεταδεδομένα (*metadata*), ώστε να μπορέσουν να περιηγηθούν, να αναζητηθούν και να καταναλωθούν από τα εργαλεία. Τα συστήματα BI επιτρέπουν την εκτέλεση αναδρομικών αναλύσεων σε ένα πολύ μεγάλο σύνολο των δεδομένων πραγματικού χρόνου, προσφέροντας την δυνατότητα εκμάθησης λειτουργικών συμπεριφορών και μοτίβων, εντοπίζοντας εξαρτήσεις και συσχετίσεις μεταξύ παραγόντων.

PI System Access: Αποτελεί μια συλλογή εργαλείων για την σύνδεση του *PI System* με άλλα διαφορετικά συστήματα, προσφέροντας την δυνατότητα στους χρήστες να αποκτήσουν πρόσβαση και να αλληλεπιδράσουν με τα δεδομένα του *PI*. Κάθε εργαλείο της συλλογής περιέχει διαφορετικούς τρόπους σύνδεσης, άντλησης και καταχώρησης δεδομένων. Το *PI System Access (PSA)* περιλαμβάνει *Application Programming Interfaces (APIs)*, *Software Development Kits (SDK)*, *ODBC* και άλλα εργαλεία που διευκολύνουν την ανάπτυξη προσαρμοσμένων εφαρμογών και λύσεων.

Καθώς η παρούσα διπλωματική εργασία σκοπεύει να αξιοποιήσει της δυνατότητες του *PI Web API* προκειμένου να αντλήσει δεδομένα από το *PI Data Archive*, όπως και να εγγράψει δεδομένα προς το σύστημα, τα οποία θα είναι προϊόν αποτελεσμάτων επεξεργασίας των αλγορίθμων Μηχανικής Μάθησης, είναι απαραίτητο να περιγράψουν οι λειτουργίες και τα χαρακτηριστικά του. Το *PI Web API* είναι μία *RESTful (Representational State Transfer)* υπηρεσία που επιτρέπει προγραμματιστική πρόσβαση στους πόρους του *PI System*. Μέσω της χρήσης κατάλληλων μεθόδων *HTTP* (όπως *GET, POST, PUT* και *DELETE*) και αιτημάτων (*web requests*) παρέχονται λειτουργίες ανάγνωσης, εγγραφής δημιουργίας και διαγραφής στον *PI DA* και *PI AF*. Η υπηρεσία

είναι χωρίς κατάσταση (*stateless*) συνεπώς κάθε αίτημα προς τον πελάτη πρέπει να περιέχει όλες τις απαραίτητες πληροφορίες.

Cloud Services: Τα OSIssoft Cloud Services (OCS) είναι μια ασφαλής πλατφόρμα δεδομένων που βασίζεται στο Microsoft Azure και λειτουργεί πλήρως από την OSIssoft ως SaaS/PaaS (*Software as a Service/Platform as a Service*). Το OCS επιτρέπει την συλλογή, διαχείριση, βελτίωση και κοινοποίηση ιστορικών και πραγματικών δεδομένων συμπληρώνοντας τους PI Servers που βρίσκονται τοπικά στις εγκαταστάσεις. Επιπλέον, διευκολύνει την μεταβίβαση των δεδομένων τόσο σε εσωτερικό επίπεδο όσο και σε έμπιστους συνεργάτες καθώς και σε άλλες επιχειρησιακές εφαρμογές καθιστώντας τα έτοιμα για προηγμένες αναλύσεις.

Όπως περιεγράφηκε στο παρόν κεφάλαιο, η επιλεγόμενη πλατφόρμα βιομηχανικού ΔτΠ είναι απαραίτητο να διαθέτει ορισμένα χαρακτηριστικά προκειμένου να χρησιμοποιηθεί με επιτυχία ως ενδιάμεσος μεταξύ συστημάτων *OT* και εφαρμογής πρόγνωσης και διαχείρισης υγείας. Τα χαρακτηριστικά περιλαμβάνουν την ευελιξία, στην αρχιτεκτονική των δικτύων, την δυνατότητα συλλογής δεδομένων από ετερογενείς εξοπλισμούς και πρωτόκολλα επικοινωνίας, την λειτουργικότητα μεταξύ διαφορετικών συστημάτων, την αποθήκευση πληροφορίας σε πραγματικό χρόνο και την διασφάλιση της εμπιστευτικότητας, ακεραιότητας και διαθεσιμότητας των δεδομένων. Ωστόσο, η επιτυχία της εφαρμογής δεν εξαρτάται εξ' ολοκλήρου από την επιλογή της πλατφόρμας, εξίσου σημαντική είναι και η επιλογή των κατάλληλων αλγορίθμων Μηχανικής Μάθησης καθώς και η εκπαίδευση τους. Η ακρίβεια στην πρόβλεψη του εναπομείναντος ωφέλιμου χρόνου ζωής και η εκτίμηση της κατάστασης του εξοπλισμού εξαρτάται τόσο από τη ποιότητα των δεδομένων που συλλέγονται, όσο και από την ικανότητα των προγραμμάτων να τα αναλύουν με αξιοπιστία.

4

Μηχανισμοί και Τεχνικές για την Εκπαίδευση

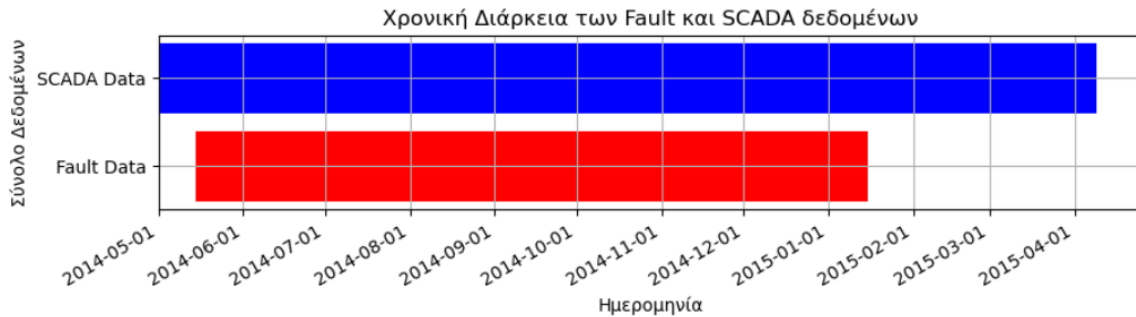
Αλγορίθμων Μηχανικής Μάθησης

4.1 Περιγραφή Dataset

Όπως έχει ήδη αναφερθεί, η προγνωστική συντήρηση είναι θεμελιώδης για τη Βιομηχανία 4.0 καθώς βελτιστοποιεί τις παραγωγικές διαδικασίες και μειώνει τα λειτουργικά κόστη. Για την ανάπτυξη ενός αξιόπιστου προγνωστικού μοντέλου είναι κρίσιμη η πρόσβαση σε ποιοτικά σύνολα δεδομένων (*Dataset*) από πραγματικές πηγές τα οποία καταγράφουν σήματα από διάφορους αισθητήρες. Αυτά τα δεδομένα πρέπει να περιλαμβάνουν πληθώρα διαφορετικών μετρήσεων που θα αντικατοπτρίζουν την κατάσταση των συστημάτων σε πραγματικό χρόνο. Επιπλέον, είναι απαραίτητο να καταγράφονται τόσο κανονικές όσο και διαφορετικές καταστάσεις σφαλμάτων. Η ποικιλομορφία αυτή προσφέρει τη δυνατότητα για την ανάπτυξη μοντέλων με μεγαλύτερη ευελιξία και εφαρμογή, ικανά να ανιχνεύουν και να προβλέπουν βλάβες σε ένα ευρύτερο φάσμα λειτουργιών. Η καταγραφή πολλαπλών τύπων βλαβών βελτιώνει σημαντικά την ακρίβεια των μοντέλων και επιτρέπει την εντόπιση προβλημάτων πριν αυτά γίνουν καταστροφικά, προσφέροντας ουσιαστικά οφέλη στην αποδοτικότητα και την επιτάχυνση των διαδικασιών συντήρησης. Ωστόσο, η διαρκής ανάγκη ανάπτυξης προγνωστικών μοντέλων απαιτεί τη διαθεσιμότητα ποιοτικών και δημόσια προσβάσιμων συνόλων δεδομένων ώστε να ενισχυθεί η συνεργασία μεταξύ των ερευνητικών κοινοτήτων και να επιτευχθεί η διάδοση αποδοτικότερων μοντέλων [26].

Η παρούσα Διπλωματική προσπαθεί να αξιοποιήσει τα δεδομένα του Dataset «*IIoT Wind Turbine Analytics*» τα οποία είναι δημόσια διαθέσιμα στην πλατφόρμα του *Kaggle* και προέρχονται από μια ανεμογεννήτρια άγνωστης ταυτότητας. Περιέχει δύο υποσύνολα δεδομένων το *Scada_Data* και το *Fault_Data*. Όπως απεικονίζεται και στην Εικόνα 11 το πρώτο υποσύνολο έχει μεγαλύτερο χρονικό διάστημα καταγραφής καλύπτοντας την περίοδο από την 1^η Μαΐου του 2014 μέχρι και την

9^η Απριλίου του 2015, ενώ το δεύτερο περιέχει πληροφορίες σχετικά με σφάλματα του συστήματος από την 14^η Απριλίου μέχρι και την 15^η Ιανουαρίου του 2015.



Εικόνα 11. Χρονική Διάρκεια καταγραφής δεδομένων των υποσυνόλων Fault Data και Scada Data

Το Πρώτο υποσύνολο αφορά δεδομένα από αισθητήρες του Βιομηχανικού Διαδικτύου των Πραγμάτων (IIoT) περιέχοντας εξήντα πέντε μεταβλητές (Features) που σχετίζονται με την γενική κατάσταση των συστημάτων αλλά και με την λειτουργία της ανεμογεννήτριας. Πιο συγκεκριμένα, τα χαρακτηριστικά περιλαμβάνουν δεδομένα σχετικά με:

- Περιβαλλοντολογικές μετρήσεις: Μέση, Μέγιστη και Ελάχιστη ταχύτητα ανέμου καθώς και Θερμοκρασία Περιβάλλοντος
- Λειτουργικότητας Ανεμογεννήτριας: Μέση, Μέγιστη και Ελάχιστη ταχύτητα περιστροφής, Γωνία Φτερού, Θέση Nacelle, Συνολικές ώρες λειτουργίας, Παραγόμενη ενέργεια, RTU Setpoint και Μέση, Μέγιστη και Ελάχιστη Reactive Ισχύς.
- Διαθεσιμότητα Συστήματος: Πληροφορίες σε σχέση με την διαθέσιμη ισχύ του συστήματος βάση Ανέμου, Τεχνικών/Εξωτερικών λόγων ή και ανωτέρας βίας.
- Θερμοκρασιακές πληροφορίες εξοπλισμού: Θερμοκρασίες των επτά καμπινών των μετατροπέων του πρώτου και δεύτερου συστήματος, Θερμοκρασίες Ρότορα και Στάτορα καθώς και Θερμοκρασίες των επιμέρους τμημάτων της ανεμογεννήτριας.

Η Εικόνα 12 συγκεντρώνει βασικές στατιστικές πληροφορίες των εν λόγω χαρακτηριστικών. Σύμφωνα με αυτές, παρατηρείται ότι δεν υπάρχει απώλεια δεδομένων για κανένα αισθητήρα καθώς το *Missing(%)* είναι μηδενικό για όλα τα χαρακτηριστικά του υποσυνόλου, ενώ για τον αισθητήρα που καταγράφει τις διακυμάνσεις της θερμοκρασίας εντός του ντουλαπιού (cabinet) ενός μετατροπέα στο δεύτερο σύστημα «Sys 2 inverter 6 cabinet temp» καταγράφεται μία μόνο μοναδική τιμή θερμοκρασίας, που υποδεικνύει ότι ο αισθητήρας δεν έχει ανταποκριθεί σε αλλαγές της θερμοκρασίας και ότι ενδεχομένως να χρίζει αντικατάστασης. Παρόμοια συμπεριφορά εντοπίζεται και για τους αισθητήρες θερμοκρασίας του δεύτερου συστήματος στον πέμπτο και έβδομο μετατροπέα, έχοντας μόνο δύο και τρεις μοναδικές τιμές αντιστοίχως. Αυτή η περιορισμένη μεταβλητότητα υποδηλώνει πιθανές ανεπάρκειες στη λειτουργία των συγκεκριμένων αισθητήρων τα οποία επηρεάζουν την ακρίβεια των δεδομένων που καταγράφονται από αυτούς. Επιπλέον, ανάμεσα στις διαθέσιμες πληροφορίες της Εικόνας 12, παρατηρείται και η τυπική απόκλιση (*Standard Deviation*) ως μέτρο μεταβλητότητας ή και σταθερότητας των δεδομένων. Όπως είναι αναμενόμενο, για τους προαναφερθέντες αισθητήρες θερμοκρασίας του δεύτερου συστήματος

υπολογίζεται μηδενική τυπική απόκλιση ή κοντά στο μηδέν, υποδεικνύοντας ότι οι τιμές τους παραμένουν σταθερές χωρίς ουσιαστική διαφοροποίηση. Συνολικά, τα δεδομένα του υποσυνόλου

Statistics						
Name	Mean	Min	Max	Standard Deviation	Distinct	Missing [%]
Ambient temp.	13.39	0	35	5.243	36	0
Blade A temp.	165.27	0	170	9.608	7	0
Blade B temp.	164.92	0	169	9.587	5	0
Blade C temp.	165.65	0	170	9.631	7	0
Control cabinet temp.	31.78	0	45	6.386	39	0
Error	0.9367	0	246	14.43	12	0
Fan inverter cabinet temp.	28.81	0	44	5.182	36	0
Front bearing temp.	22.07	0	35	4.962	29	0
Inverter averages	27.84	-14	42.55	5.6	337	0
Inverter std dev	1.855	0	23.51	1.269	1047	0
Main carrier temp.	16.56	0	29	4.794	27	0
Nacelle ambient temp. 1	12.52	0	28	4.746	29	0
Nacelle ambient temp. 2	12.44	0	28	4.743	29	0
Nacelle cabinet temp.	19.85	0	36	5.248	32	0
Nacelle temp.	16.33	0	33	4.984	31	0
Pitch cabinet blade A temp.	33.94	0	53	5.777	41	0
Pitch cabinet blade B temp.	33.91	0	52	5.681	40	0
Pitch cabinet blade C temp.	33.13	0	51	5.587	40	0
Rear bearing temp.	27.3	0	43	5.319	37	0
Rectifier cabinet temp.	30.34	0	49	5.618	43	0
Rotor temp. 1	52.74	0	125	23.03	117	0
Rotor temp. 2	52.84	0	126	22.77	118	0
RTU: ava. Setpoint 1	2988.76	0	3050	171.91	294	0
Spinner temp.	17.92	0	38	5.347	34	0
Stator temp. 1	60.89	0	119	23.81	113	0
Stator temp. 2	60.39	0	117	23.58	111	0
Sys 1 inverter 1 cabinet temp.	26.65	-14	42	6.371	41	0
Sys 1 inverter 2 cabinet temp.	27.72	-14	45	6.227	46	0
Sys 1 inverter 3 cabinet temp.	27.3	-14	43	5.818	45	0
Sys 1 inverter 4 cabinet temp.	27.38	-14	41	5.792	42	0
Sys 1 inverter 5 cabinet temp.	28.67	-14	45	6.228	44	0
Sys 1 inverter 6 cabinet temp.	27.57	-14	41	6.115	42	0
Sys 1 inverter 7 cabinet temp.	27.67	-14	41	5.779	42	0
Sys 2 inverter 1 cabinet temp.	29.15	-14	45	5.704	44	0
Sys 2 inverter 2 cabinet temp.	28.88	-14	46	5.699	44	0
Sys 2 inverter 3 cabinet temp.	27.86	-14	42	5.832	43	0
Sys 2 inverter 4 cabinet temp.	27.44	-14	41	5.87	42	0
Sys 2 inverter 5 cabinet temp.	-13.95	-14	0	0.8099	2	0
Sys 2 inverter 6 cabinet temp.	0	0	0	0	1	0
Sys 2 inverter 7 cabinet temp.	-49.83	-50	8	2.951	3	0
Tower temp.	23.13	0	36	6.362	32	0
Transformer temp.	44.03	-19	71	10.44	59	0
WEC: ava. Available P force external reasons	1040.59	0	65535	1210.66	3065	0
WEC: ava. Available P force majeure reasons	980.93	0	65535	1214.52	3063	0
WEC: ava. available P from wind	1046.32	0	65535	1217.53	3067	0
WEC: ava. available P technical reasons	1021.27	0	65535	1211.12	3065	0
WEC: ava. blade angle A	10.28	0	327.67	25	2390	0
WEC: ava. Nacel position including cable twisting	25.79	-892	1080	293.75	1495	0
WEC: ava. Power	945.15	0	3071	1010.56	3067	0
WEC: ava. reactive Power	81.89	-8	425	84.31	351	0
WEC: ava. Rotation	8.685	0	14.73	4.144	1413	0
WEC: ava. windspeed	6.889	0	32.1	3.706	254	0
WEC: max. Power	1217.8	0	3216	1170.75	3189	0
WEC: max. reactive Power	103.52	-6	1145	114.36	538	0
WEC: max. Rotation	9.559	0	18.91	4.484	1548	0
WEC: max. windspeed	9.361	0	51.1	5.175	360	0
WEC: min. Power	652.86	0	3028	819.54	2983	0
WEC: min. reactive Power	64.08	-168	390	71.38	361	0
WEC: min. Rotation	8.512	0	655.35	22.37	1361	0
WEC: min. windspeed	12.23	0	6553.5	222.94	177	0
WEC: Operating Hours	4213.6	0	7884	2111.81	7005	0
WEC: Production kWh	4021341.5	0	8579750	2362360	41753	0
WEC: Production minutes	28.83	0	59	17.89	60	0
Yaw inverter cabinet temp.	24.33	0	38	4.917	31	0

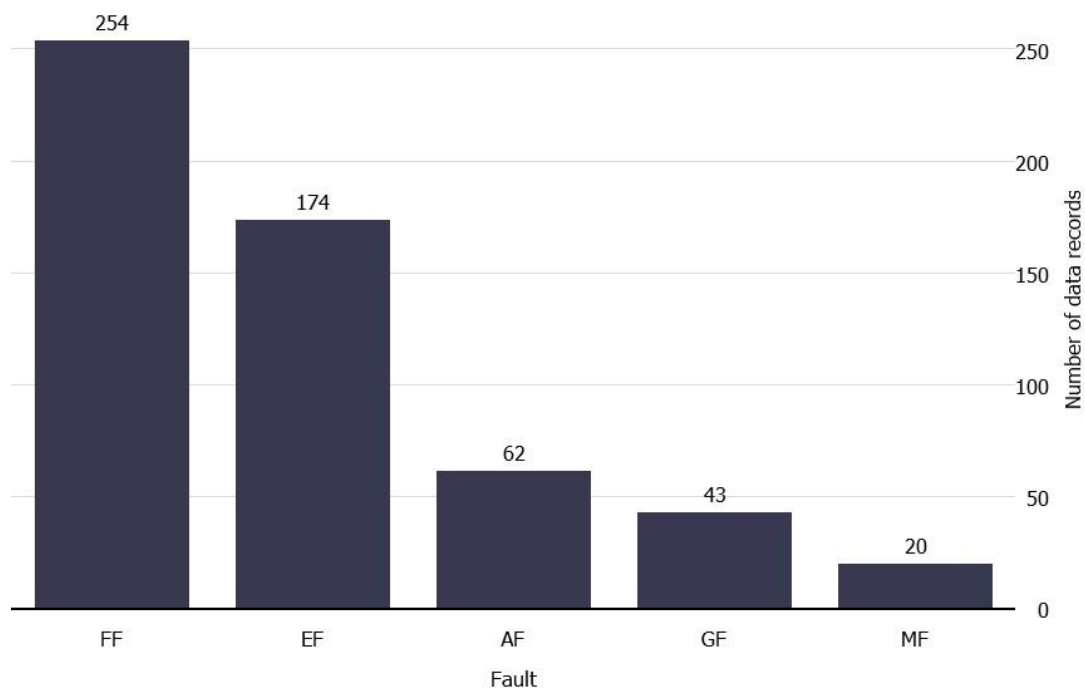
Εικόνα 12. Στατιστικά Συνόλου Δεδομένων Scada Data

προσφέρουν πολύτιμες πληροφορίες για τη λειτουργία και την απόδοση της ανεμογεννήτριας, ωστόσο η έλλειψη μεταβλητότητας σε ορισμένα χαρακτηριστικά υπογραμμίζει τη σημασία της σωστής λειτουργίας των αισθητήρων, προκειμένου να διασφαλιστεί η ακρίβεια των προγνωστικών μοντέλων που θα αναπτυχθούν στο πλαίσιο της προγνωστικής συντήρησης.

Στο Δεύτερο υποσύνολο καταγράφονται πέντε τύποι βλαβών που αντιπροσωπεύονται από 553 συμβάντα ταξινομημένα χρονολογικά, πιο συγκεκριμένα:

- *GF (Generator heating Fault)*: Βλάβη υπερθέρμανσης γεννήτριας.
- *MF (Mains failure Fault)*: Βλάβη αποτυχίας κύριου ηλεκτρικού δικτύου
- *FF (Feeding Fault)*: Βλάβη τροφοδοσίας συστήματος
- *AF (Air cooling Fault)*: Βλάβη στο σύστημα ψύξης
- *EF (Excitation Fault)*: Βλάβη στο σύστημα διέγερσης της γεννήτριας

Ωστόσο, όπως προκύπτει και από την Εικόνα 13, όπου παρουσιάζεται συγκεντρωτικά ο συνολικός αριθμός των καταγεγραμμένων σφαλμάτων ανά κατηγορία, παρατηρείται άνιση κατανομή



Εικόνα 13. Συνολικός αριθμός καταγεγραμμένων σφαλμάτων ανά κατηγορία

των παραδειγμάτων. Πιο Ειδικά, η βλάβη τροφοδοσίας συστήματος εμφανίζει τον μεγαλύτερο αριθμό περιστατικών με 254 καταγραφές, ενώ η βλάβη αποτυχίας του κύριου ηλεκτρικού δικτύου καταγράφει τα λιγότερα παραδείγματα με μόλις 20 το σύνολο.

Η άνιση κατανομή των παραδειγμάτων στις κατηγορίες βλαβών είναι ένα σύνηθες πρόβλημα σε συστήματα που βρίσκονται σε συνεχή λειτουργία και δεν αντιμετωπίζουν συχνά δυσλειτουργίες. Δεδομένου ότι οι αστοχίες εμφανίζονται σπάνια, οι βάσεις δεδομένων συνήθως περιέχουν περισσότερα παραδείγματα κανονικής λειτουργίας ή συγκεκριμένων συνηθισμένων σφαλμάτων, ενώ οι σπανιότερες βλάβες καταγράφονται λιγότερο συχνά. Η ανισοκατανομή των παραδειγμάτων

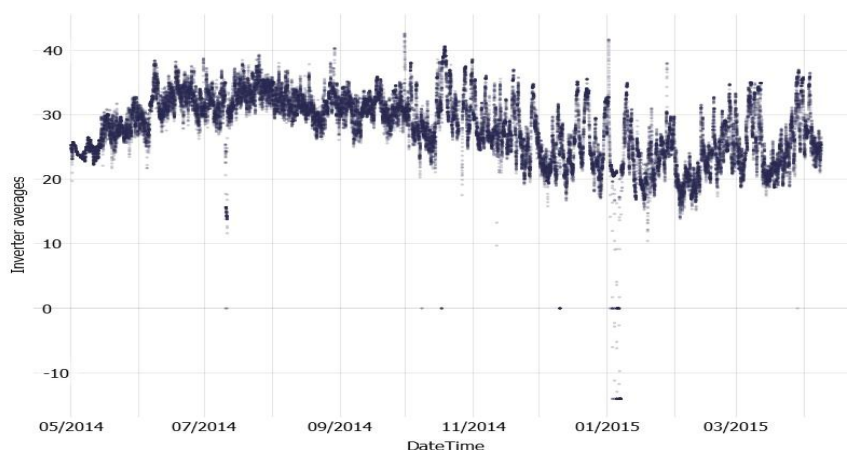
εκπαίδευσης δύναται να επηρεάσει σημαντικά την εκπαίδευση των μοντέλων προγνωστικής συντήρησης, μειώνοντας την ακρίβεια των αποτελεσμάτων και περιορίζοντας την δυνατότητα του αλγόριθμου να αναγνωρίσει ή να προβλέψει σπάνια αλλά κρίσιμα περιστατικά αστοχίας του εξοπλισμού. Προκειμένου να αντιμετωπιστούν οι περιπτώσεις της μη ισορροπημένης κλάσης (*class imbalance*), εφαρμόζονται συγκεκριμένες τεχνικές κατά την διαδικασία προ-επεξεργασίας των δεδομένων, με σκοπό την δημιουργία συνόλων με μειωμένη ανισορροπία κατηγοριών για τη βελτίωση της απόδοσης του μοντέλου Μηχανικής Μάθησης στην αναγνώριση της μειονοτικής κατηγορίας [27].

4.2 Προ-επεξεργασία Δεδομένων (Data Pre-processing)

Η προ-επεξεργασία των δεδομένων είναι ένα κρίσιμο στάδιο στη διαδικασία ανάλυσης, καθώς εξασφαλίζει την ακρίβεια, την ποιότητα και τη συνέπεια των αποτελεσμάτων. Τα ακατέργαστα δεδομένα όπως λαμβάνονται απευθείας από τους αισθητήρες, συχνά περιέχουν σφάλματα, ελλείποντα ή ανεπιθύμητα στοιχεία και θόρυβο που μπορεί να επηρεάσουν αρνητικά την ακρίβεια των προβλεπτικών μοντέλων. Το συγκεκριμένο στάδιο αποτελείται από ένα σύνολο τεχνικών που περιλαμβάνουν την ανίχνευση και δόρθωση ακραίων τιμών, την μείωση των χαρακτηριστικών όπου κρίνεται απαραίτητο, την κανονικοποίηση ή τυποποίηση των δεδομένων και την αντιμετώπιση της ανισορροπίας των κλάσεων. Στόχος της διαδικασίας αποτελεί η εξασφάλιση ενός καθαρού και ομοιογενές συνόλου δεδομένων το οποίο θα μπορεί να αξιοποιηθεί με επιτυχία από τα μοντέλα Μηχανικής Μάθησης [28].

4.2.1 Εντοπισμός ακραίων τιμών (Outlier Detection)

Η ανίχνευση ακραίων τιμών (*outliers detection*) είναι μια τεχνική που στοχεύει στον εντοπισμό καταγραφών που διαφέρουν σημαντικά από τα υπόλοιπα σε ένα σύνολο δεδομένων. Αυτές οι ακραίες τιμές μπορεί να οφείλονται σε διάφορους λόγους, όπως σφάλματα κατά τη συλλογή δεδομένων, δυσλειτουργία εξοπλισμού, σπάνια ή και ασυνήθιστα γεγονότα.



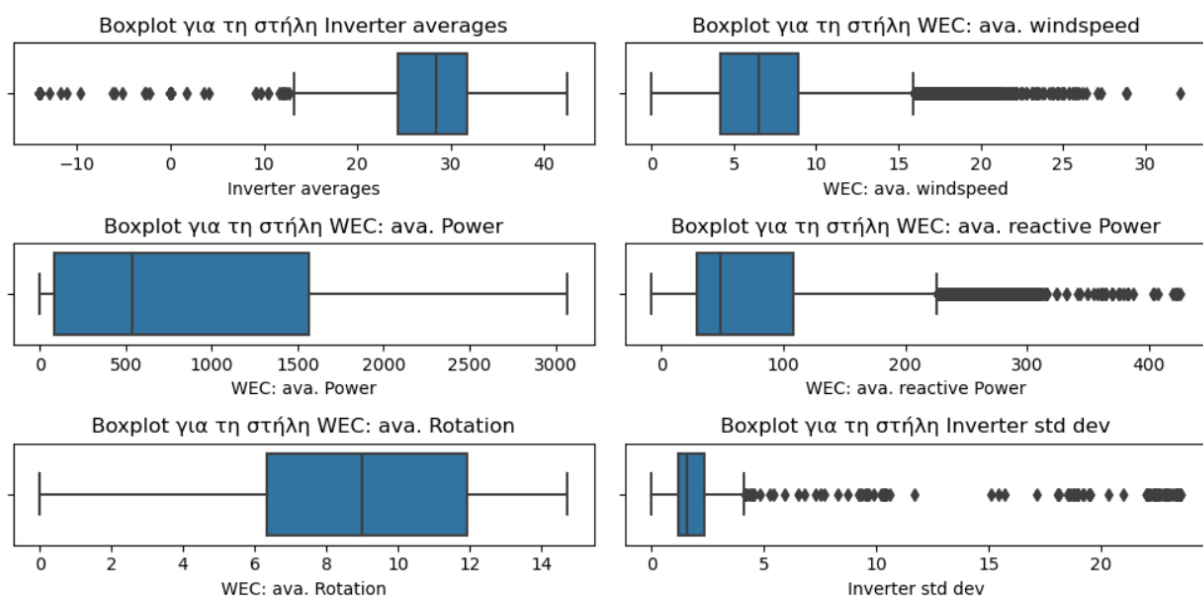
Εικόνα 14. Εντοπισμός ακραίων τιμών

Ωστόσο, είναι απαραίτητο να διαχωρίζονται οι ακραίες τιμές που προκύπτουν από σφάλματα, προσθέτοντας θόρυβο στο σύστημα μας και μειώνοντας την απόδοση του μοντέλου, από τις πραγματικά ενδιαφέρουσες πληροφορίες που υποδεικνύουν σημαντικά μοτίβα ή σπάνια φαινόμενα. Αυτές οι ακραίες παρατηρήσεις μπορεί να επηρεάσουν την ανάλυση με παραπλανητικό τρόπο,

οπότε μπορεί να χρειαστεί να αφαιρεθούν, ενώ σε άλλες περιπτώσεις μπορεί να παρέχουν πολύτιμες πληροφορίες και να είναι σημαντικό να διατηρηθούν [29]. Για τον λόγο αυτό, πριν από οποιαδήποτε απόφαση θα πρέπει να πραγματοποιείται προσεκτική ανάλυση καθώς ελλοχεύει ο κίνδυνος να απομακρυνθούν δεδομένα που αντανακλούν σε πραγματικές αλλά σπάνιες συνθήκες και όχι σε συστημικό θόρυβο. Τα Outliers μπορούν να ταξινομηθούν σε τρεις βασικές κατηγορίες:

- Τις καθολικές ακραίες τιμές (global outliers): Γνωστές και ως *point outliers*, οι οποίες αποκλίνουν έντονα από τις κανονικές τιμές, καθώς απέχουν σημαντικά από τις υπόλοιπες παρατηρήσεις ή τη σύνηθες συμπεριφορά των υπόλοιπων μετρήσεων.
- Τις ακραίες τιμές αναλόγως το πλαίσιο (contextual outliers): Οι οποίες είναι ακραίες μόνο υπό ένα συγκεκριμένο πλαίσιο.
- Τις ομαδικές ακραίες τιμές (collective outliers): Οι οποίες αναφέρονται σε ομάδες παρατηρήσεων που θεωρούνται ακραίες όταν εξετάζονται από κοινού, ενώ καμία από αυτές δεν μπορεί να χαρακτηριστεί ακραία αν αναλυθεί μεμονωμένα.

Η Εικόνα 15 παρουσιάζει την οπτική αναπαράσταση της κατανομής των δεδομένων για τα επιλεγμένα χαρακτηριστικά (Inverters Average, WEC: ava. Windspeed, κλπ.) χρησιμοποιώντας τα *Boxplot*. Κάθε *Boxplot* απεικονίζει την κεντρική τάση (διάμεσο) και την ποικιλία των δεδομένων, παρέχοντας πληροφορίες για τη παρουσία ακραίων τιμών τόσο για το χαμηλότερο όσο και για το υψηλότερο σημείο. Οι άκρες του κουτιού αντιπροσωπεύουν το 1ο και το 3ο τεταρτημόριο (Q1 και Q3) ενώ οι γραμμές *whiskers* παρουσιάζουν τα όρια μεταξύ των τιμών που δεν θεωρούνται ακραία. Οποιαδήποτε καταγραφή βρίσκεται εκτός των ορίων παρουσιάζεται ως ξεχωριστό σημείο στην γραφική απεικόνιση. Το *Boxplot* αποτελεί ένα επιπλέον εργαλείο για την λήψη αποφάσεων και την ανάλυση δεδομένων.



Εικόνα 15. Συνολικός αριθμός καταγεγραμμένων σφαλμάτων ανά κατηγορία

Στη συνέχεια και εφόσον έχει ολοκληρωθεί η διαδικασία ανίχνευσης των ακραίων τιμών, μέσω της χρήσης κατάλληλων εργαλείων, πραγματοποιείται η αφαίρεση ή αναπροσαρμογή τους. Ωστόσο, η υπερβολική αφαίρεση δεδομένων μπορεί να οδηγήσει σε μοντέλα υπερ-προσαρμογής με χαμηλές ικανότητες γενίκευσης, τα οποία θα δυσκολεύονται να προβλέψουν την κατάσταση του

συστήματος όταν τροφοδοτηθούν με νέα, άγνωστα δεδομένα. Για τον λόγο αυτό, η παρούσα διπλωματική εργασία εστιάζει στην αντικατάσταση των ακραίων τιμών με κατάλληλα όρια, διατηρώντας τη σημαντική πληροφορία που μπορεί να φέρουν, χωρίς να αλλοιώνεται η αντιπροσωπευτικότητα των δεδομένων.

4.2.2 Επιλογή Χαρακτηριστικών (Feature Selection)

Η Μείωση Χαρακτηριστικών είναι μια διαδικασία στην ανάλυση δεδομένων και τη Μηχανική Μάθηση που στοχεύει στη μείωση του αριθμού των χαρακτηριστικών που υπάρχουν σε ένα σύνολο δεδομένων. Μια εκ των κύριων τεχνικών για τη μείωση της διάστασης των χαρακτηριστικών είναι η **Επιλογή Χαρακτηριστικών (Feature Selection)**. Μέσω αυτής της προσέγγισης, επιλέγονται τα πιο αντιπροσωπευτικά χαρακτηριστικά με βάση στατιστικές πληροφορίες ή μετρικές συσχέτισης, όπως ο συντελεστής συσχέτισης *Pearson correlation*. Όσα επιλέγονται, διατηρούνται χωρίς να μετασχηματίζονται, ενώ όσα θεωρούνται περιττά ή δεν προσθέτουν αξία στο μοντέλο απομακρύνονται. Η μείωση των χαρακτηριστικών συμβάλει στην απλοποίηση του συστήματος, καθώς μέσω της μείωσης των διαστάσεων και της πολυπλοκότητας του μοντέλου διευκολύνεται η αναγνώριση μοτίβων και επιτυγχάνεται μεγαλύτερη ακρίβεια στην ταξινόμηση. Επιπλέον, η μείωση των διαστάσεων αυξάνει την ταχύτητα εκπαίδευσης του μοντέλου, μειώνοντας το απαιτούμενο υπολογιστικό φόρτο. Παράλληλα, βοηθά στην αποφυγή της υπέρ-προσαρμογής (*Overfitting*) καθώς το μοντέλο εστιάζει στις πιο σχετικές πληροφορίες και δεν προσαρμόζεται στον θόρυβο των δεδομένων, ενισχύοντας την ικανότητα του να γενικεύει σε νέα δεδομένα και να έχει καλύτερες επιδόσεις σε πραγματικές συνθήκες.

Στη παρούσα ανάλυση γίνεται χρήση του μηχανισμού μείωσης χαρακτηριστικών αξιοποιώντας στατιστικές μετρητικές παραμέτρους, απομακρύνοντας πλεονάζουσες μεταβλητές με μικρές έως μηδαμινές διακυμάνσεις. Μέσω της τεχνικής αυτής μειώνεται η μεροληψία του εκπαιδευμένου μοντέλου και ελαττώνεται ο θόρυβος που εισέρχεται στα δεδομένα. Σκοπός αποτελεί ο εντοπισμός των χαρακτηριστικών εκείνων που διαθέτουν την χαμηλότερη μεταβλητότητα και δεν προσφέρουν ουσιαστική πληροφορία στο μοντέλο. Ενδεικτικό παράδειγμα αποτελεί η μεταβλητή «*Sys 2 inverter 6 cabinet temp*» της Εικόνας 12, η οποία έχει μηδενική τυπική απόκλιση, υποδεικνύοντας ότι οι μετρήσεις του παραμένουν σταθερές και δεν προσφέρουν αξία στην εκπαίδευση του μοντέλου. Για τον προσδιορισμό των μεταβλητών εκείνων που δεν συμβάλλουν ουσιαστικά στο μοντέλο, χρησιμοποιείται ο συντελεστής μεταβλητότητας (*cv*) ο οποίος ορίζεται ως ο λόγος της τυπικής απόκλισης προς τον μέσο όρο ενός χαρακτηριστικού. Αν ο συντελεστής μεταβλητότητας είναι μικρότερος από 0.05, δηλαδή αν η τυπική απόκλιση είναι μικρότερη από το 5% του μέσου όρου τότε το χαρακτηριστικό θεωρείται ως θόρυβος και απορρίπτεται. Η συγκεκριμένη προσέγγιση παρουσιάζεται από τον τύπο της Εξίσωσης (1):

$$cv = \sigma / \mu' \quad (1)$$

Όπου:

- *cv* - Ο συντελεστής μεταβλητότητας
- σ - Η τυπική απόκλιση του χαρακτηριστικού
- μ' - Ο μέσος όρος του χαρακτηριστικού

4.2.3 Κανονικοποίηση Δεδομένων (Data Normalization)

Η Κανονικοποίηση των Δεδομένων είναι μια σημαντική τεχνική προ-επεξεργασίας των χαρακτηριστικών, καθώς προστατεύει το μοντέλο από την κυριαρχία μεταβλητών που διαθέτουν μεγαλύτερες καταγεγραμμένες αριθμητικές τιμές και μπορούν να επηρεάσουν σε μεγαλύτερο βαθμό τον αλγόριθμο. Γεγονός που αποτελεί κρίσιμης σημασίας όταν τα δεδομένα προέρχονται από αισθητήρες που καταγράφουν μετρήσεις σε διαφορετικές κλίμακες και μεγέθη. Σκοπός μέσω της διαδικασίας αυτής είναι η εξασφάλιση ότι όλα τα δεδομένα θα συμβάλλουν ισότιμα στην εκπαίδευση του μοντέλου, ανεξαρτήτως του εύρους τιμών τους. Ωστόσο, η κανονικοποίηση δεν βελτιώνει πάντα την απόδοση ταξινόμησης και εξαρτάται από τα χαρακτηριστικά των δεδομένων [30]. Παρόλο που υπάρχουν διαφορετικές τεχνικές εξισορρόπησης, στην παρούσα εργασία αξιοποιούνται οι δύο πιο διαδεδομένοι μέθοδοι: η *Min -Max Normalization* και η *Z-score Normalization*.

Min -Max Normalization: Η συγκεκριμένη μέθοδος κλιμακώνει την τιμή της μεταβλητής $x_{i,n}$ στη νέα κανονικοποιημένη τιμή της $x'_{i,n}$ σε ένα προκαθορισμένο εύρος από 0 έως 1 ή από -1 έως 1, σύμφωνα με τον τύπο της Εξίσωσης (2).

$$x'_{i,n} = \frac{(x_{i,n} - \min(x_i))}{(\max(x_i) - \min(x_i))} (nMax - nMin) + nMin \quad (2)$$

Όπου:

- $\max(x_i)$ – Ανώτατη αρχική τιμή χαρακτηριστικού i
- $\min(x_i)$ – Κατώτατη αρχική τιμή χαρακτηριστικού i
- $nMax$ – Ανώτατο προκαθορισμένο όριο για τον μετασχηματισμό του χαρακτηριστικού i
- $nMin$ – Κατώτατο προκαθορισμένο όριο για τον μετασχηματισμό του χαρακτηριστικού i

Z-score Normalization: Η τιμή κάθε μεταβλητής $x_{i,n}$ μετασχηματίζεται σε $x'_{i,n}$ με τέτοιο τρόπο ώστε τα νέα χαρακτηριστικά να έχουν μηδενικό μέσο όρο και μία μονάδα τυπική απόκλιση, σύμφωνα με τον τύπο της Εξίσωσης (3).

$$x'_{i,n} = (x_{i,n} - \mu_i) / \sigma_i \quad (3)$$

Όπου:

- μ_i – Ο μέσος όρος του χαρακτηριστικού i
- σ_i – Η τυπική απόκλιση του χαρακτηριστικού i

4.2.4 Ανάλυση Κυρίων Συνιστωσών (Principal Component Analysis – PCA)

Η Ανάλυση Κυρίων Συνιστωσών (PCA) είναι μια πολύ-μεταβλητή στατιστική μέθοδος η οποία χρησιμοποιείται κατά τη μείωση των διαστάσεων σε σύνολα δεδομένων με πολλές μεταβλητές διατηρώντας ωστόσο το μέγιστο δυνατό ποσοστό πληροφορίας. Το PCA αποτελεί ένα επιπλέον εργαλείο στη διαδικασία **Μείωσης Χαρακτηριστικών** (Feature Reduction), επιτρέποντας την απλοποίηση των δεδομένων με σκοπό την ταχύτερη και αποδοτικότερη εφαρμογή αλγορίθμων Μηχανικής Μάθησης. Η συγκεκριμένη τεχνική έχει αξιοποιηθεί σε ποικίλες εφαρμογές, όπως κατά την ανάλυση διαφορετικών συνθηκών λειτουργίας και την πρόβλεψη της διάρκειας ζωής των εδράνων [32]. Μετασχηματίζει τις πληροφορίες των αρχικών χαρακτηριστικών σε νέες και

λιγότερες τον αριθμό μεταβλητές, γνωστές ως **κύριες συνιστώσες** (Principal Components, PCs). Η "πληροφορία" που εμπεριέχεται στα δεδομένα υπολογίζεται μέσω της συνολικής διασποράς τους και οι κύριες συνιστώσες βελτιστοποιούνται ώστε να εξηγούν το μεγαλύτερο μέρος αυτής της διασποράς. Παρακάτω αναφέρεται η διαδικασία της τεχνικής:

Αρχικώς, υπολογίζεται ο πίνακας συνδιακύμανσης S , σύμφωνα με τον τύπο της Εξίσωσης (4).

$$S = \frac{1}{n-1} \sum_{i=1}^n (X_i - \mu)(X_i - \mu)^T \quad (4)$$

Όπου:

- S – ο πίνακας συνδιακύμανσης
- X_i – το διάνυσμα παρατηρήσεων για την i παρατήρηση
- μ – ο μέσος όρος των δεδομένων κάθε χαρακτηριστικού
- $(X_i - \mu)^T$ – η μεταφορά του διανύσματος $(X_i - \mu)$

Η μέση τιμή μ για κάθε χαρακτηριστικό δίνεται από τον τύπο της Εξίσωσης (5):

$$\mu = \frac{1}{n} \sum_{i=1}^n X_i \quad (5)$$

Στη συνέχεια υπολογίζονται οι ιδιοτιμές (λ_i) και τα ιδιοδιανύσματα (e_i) του πίνακα μέσω της ιδιοανάλυσης του S (6).

$$S e_i = \lambda_i e_i \quad (6)$$

Όπου:

- S – ο πίνακας συνδιακύμανσης
- λ_i – το i – οστό ιδιοδιάνυσμα
- e_i – η αντίστοιχη ιδιοτιμή

Στη συνέχεια, τα ιδιοδιανύσματα ταξινομούνται με βάση τις αντίστοιχες ιδιοτιμές τους και επιλέγονται εκείνα τα ιδιοδιανύσματα που αντιστοιχούν στις μεγαλύτερες ιδιοτιμές και εξηγούν το μεγαλύτερο μέρος της συνολικής διασποράς. Τέλος, οι νέες μεταβλητές Y_i αντιστοιχούν στις προβολές των αρχικών δεδομένων στο χώρο των κυρίων συνιστωσών, όπως δίνεται από την Εξίσωση (7).

$$Y_i = E^T (X_i - \mu) \quad (7)$$

Όπου:

- E – ο πίνακας που περιέχει τα επιλεγμένα ιδιοδιανύσματα $[e_1, e_2, \dots, e_d]$
- Y_i – οι νέες μεταβλητές για την i – η παρατήρηση
- X_i – το διάνυσμα παρατηρήσεων για την i παρατήρηση
- μ – ο μέσος όρος των δεδομένων κάθε χαρακτηριστικού

Προκειμένου να διασφαλιστεί ότι κάποιες μεταβλητές δεν θα επηρεάσουν δυσανάλογα τη διασπορά των δεδομένων, γεγονός που προκύπτει από την διαφορετικότητα της κλίμακας μεταξύ των αρχικών χαρακτηριστικών, κρίνεται αναγκαία η κανονικοποίηση των δεδομένων πριν την χρήση της Ανάλυσης Κυρίων Συνιστωσών, όπως περιεγράφηκε στη προηγούμενη υπό-ενότητα. Κατά αυτόν τον τρόπο κάθε χαρακτηριστικό συμβάλλει εξίσου στη συνολική διασπορά, αποτρέποντας μεταβλητές με μεγάλες τιμές ή διαφορετικές μονάδες μέτρησης να κυριαρχούν στην ανάλυση και να επηρεάζουν δυσανάλογα τις κύριες συνιστώσες.

Η Ανάλυση Κυρίων Συνιστωσών στοχεύει στην προβολή δεδομένων από έναν χώρο πολλών διαστάσεων σε έναν χώρο χαμηλότερων διαστάσεων, διατηρώντας τις κύριες πληροφορίες. Μέσω της τεχνικής αυτής επιτυγχάνεται ή μείωση του μεγέθους των αρχικών δεδομένων, επιτρέποντας την ευκολότερη διαχείριση και επεξεργασίας τους. Γεγονός που συμβάλλει με την σειρά του στην βελτίωση της αποδοτικότητας και ταχύτητας εκπαίδευσης των μοντέλων, καθώς μειώνεται η πολυπλοκότητα των υπολογισμών και ο υπολογιστικός φόρτος. Επιπλέον, η PCA έχει τη δυνατότητα να αποκαλύπτει τη δομή των δεδομένων, η οποία συχνά παραμένει κρυμμένη στον αρχικό πολυδιάστατο χώρο. Μέσα από αυτή τη διαδικασία δύναται να εντοπιστούν σημαντικές συσχετίσεις και μοτίβα που δεν είναι αρχικώς εμφανή. Η αποκάλυψη της υποκείμενης δομής επιτρέπει την καλύτερη κατανόηση των δεδομένων και τη βελτίωση των αποτελεσμάτων που προκύπτουν από την ανάλυσή τους. Τα παραπάνω πλεονεκτήματα καθιστούν την PCA ένα εξαιρετικά χρήσιμο εργαλείο σε διάφορους τομείς, όπως η μηχανική μάθηση και η ανάλυση μεγάλων δεδομένων. Η ικανότητα της να απλοποιεί τη διάσταση των δεδομένων χωρίς ωστόσο να παραλείπεται σημαντική πληροφορία διευκολύνει την καλύτερη αναγνώριση τάσεων και συμπεριφορών στα δεδομένα. Συνολικά, η Ανάλυση Κυρίων Συνιστωσών προσφέρει μια αποδοτική και ευέλικτη μέθοδο για την κατανόηση και επεξεργασία πολυδιάστατων συνόλων δεδομένων [31].

4.2.5 *Ανισορροπία κλάσης (Class Imbalance)*

Ένα σύνηθες πρόβλημα που αντιμετωπίζεται σε ζητήματα ταξινόμησης και πρόβλεψης της κατάστασης του εξοπλισμού αφορά την ανισορροπία κλάσης (*Class Imbalance*) και αναφέρεται στην ανισοκατανομή των παρατηρήσεων μεταξύ των διαφορετικών καταστάσεων του εξοπλισμού. Αυτό συμβαίνει καθώς σε πραγματικές συνθήκες οι βλάβες των κρίσιμων συστημάτων είναι σπάνιες, περιορίζοντας την καταγραφή των σχετικών περιπτώσεων. Με αυτό τον τρόπο, δημιουργείται υπεροχή δεδομένων προς την πλειοψηφική κλάση, που συνήθως είναι η φυσιολογική λειτουργία χωρίς σφάλμα, ενώ η μειοψηφική κλάση έχει σημαντικά λιγότερα δείγματα. Τα μοντέλα μηχανικής μάθησης τείνουν να αποδίδουν καλύτερα όταν τα δεδομένα είναι ισοκατανεμημένα. Ενώ, η ανισορροπία κλάσεων και η έλλειψη καταγραφών βλάβης δύναται να επηρεάσει αρνητικά την αξιοπιστία των αποτελεσμάτων του μοντέλου, καθώς οι μειοψηφικές κλάσεις - που ως επί το πλείστον αφορούν σφάλματα - έχουν μεγαλύτερη αξία για την πρόβλεψη, όπως για την πρόγνωση της κατάστασης τους εξοπλισμού και τον έγκαιρο εντοπισμό ανωμαλιών.

Σε προβλήματα ανισορροπίας κλάσης, το μοντέλο προσαρμόζεται κυρίως στην αναγνώριση της πλειοψηφικής κλάσης, γεγονός που το καθιστά λιγότερο ικανό στο να προσδιορίζει με επιτυχία τις περιορισμένες περιπτώσεις βλαβών. Ωστόσο, καθώς ταξινομεί με επιτυχία την κλάση που υπερτερεί σημαντικά στον αριθμό έναντι των υπολοίπων, έχει ως αποτέλεσμα να επιτυγχάνει πολύ υψηλό ποσοστό απόδοσης, οδηγώντας στο «*παράδοξο της ακρίβειας*». Όπου ένα μοντέλο φαίνεται να λαμβάνει υψηλά ποσοστά ακρίβειας χωρίς όμως να είναι πραγματικά σε θέση να εντοπίσει

αποτελεσματικά τις διαφορετικές καταστάσεις του συστήματος που είναι κρίσιμες για την πρόβλεψη, όπως για παράδειγμα τυχών ανωμαλίες. Αυτή η παραπλανητική υψηλή ακρίβεια μπορεί να οδηγήσει σε λανθασμένες αποφάσεις, καθώς η ικανότητα του μοντέλου να διακρίνει τις σημαντικές αλλά σπάνιες περιπτώσεις παραμένει περιορισμένη.

Όπως αναφέρθηκε παραπάνω η ανισορροπία των δεδομένων εμφανίζεται συχνά σε περιβάλλοντα συντήρησης, όπου οι περιπτώσεις αστοχιών είναι σχετικά λίγες αλλά ζωτικής σημασίας. Προκειμένου να αντιμετωπιστούν αποτελεσματικά, έχουν αναπτυχθεί διάφορες τεχνικές όπως η υπέρ-δειγματοληψία (*oversampling*) και η υπό-δειγματοληψία (*undersampling*) συμβάλλοντας στην εξισορρόπηση των κλάσεων[33]. Μέσω των παραπάνω τεχνικών είτε δημιουργούνται περισσότερα δείγματα της μειοψηφικής κλάσης είτε μειώνονται τα δείγματα της πλειοψηφικής, με σκοπό την ουσιαστική βελτίωση της απόδοσης των μοντέλων μηχανικής μάθησης. Οι βασικές τεχνικές που εξετάζονται στη παρούσα Διπλωματική είναι:

Random UnderSampling (RUS): Αποτελεί μια μέθοδος αντιμετώπισης της ανισορροπίας των κλάσεων κατά την οποία επιλέγεται τυχαία ένα υποσύνολο των περιπτώσεων της πλειοψηφικής κλάσης, ενώ το υπόλοιπο αγνοείται. Στόχος είναι η μείωση των παρατηρήσεων που περιγράφουν την κατάσταση που υπερτερεί αριθμητικά, ώστε να εξισορροπηθεί καλύτερα η κατανομή των δεδομένων. Χρησιμοποιείται σε προβλήματα ταξινόμησης όπου η πλειοψηφική κατάσταση κυριαρχεί σημαντικά σε σχέση με τις υπόλοιπες, με αποτέλεσμα τα μοντέλα να προσαρμόζονται στην επικρατούσα κλάση και να υστερούν στην αναγνώριση σπανιότερων καταστάσεων. Ωστόσο, η μείωση των δειγμάτων οδηγεί σε απώλεια χρήσιμων πληροφοριών, καθιστώντας την τεχνική αυτή λιγότερο αποτελεσματική.

Synthetic Minority Oversampling Technique (SMOTE): Συνιστά μια δημοφιλή τεχνική υπέρ-δειγματοληψίας που στοχεύει στην αντιμετώπιση του προβλήματος ανισορροπίας κλάσεων δημιουργώντας επιπλέον δείγματα της μειοψηφικής κλάσης. Επιπλέον, το SMOTE κατασκευάζει νέα συνθετικά δείγματα μέσω γραμμικής παρεμβολής και δεν επαναλαμβάνει απλά υπάρχοντες παρατηρήσεις. Η μέθοδος αυτή εξυπηρετεί στη βελτίωση της ισορροπίας των δεδομένων και επιτρέπει στο μοντέλο μηχανικής μάθησης να εκπαιδεύεται καλύτερα, ενισχύοντας την ικανότητά του να προβλέπει κρίσιμες περιπτώσεις από τη μειοψηφική κλάση.

Adaptive Synthetic Sampling Approach (ADASYN): Αποτελεί μια επιπλέον τεχνική υπέρ-δειγματοληψίας που στοχεύει όπως οι προηγούμενες στην αντιμετώπιση της ανισορροπίας κλάσεων σε προβλήματα ταξινόμησης. Το ADASYN όπως και το SMOTE στηρίζεται στη δημιουργία συνθετικών δεδομένων, αλλά εισάγει μια πιο δυναμική προσέγγιση προσαρμόζοντας τον αριθμό των συνθετικών δειγμάτων ανάλογα με τη δυσκολία ταξινόμησης των μειοψηφικών κλάσεων. Μέσω της μεθόδου αυτής, δημιουργούνται περισσότερα συνθετικά δείγματα για τα δεδομένα της μειοψηφικής κλάσης που είναι πιο δύσκολο να ταξινομηθούν, δηλαδή για αυτά που βρίσκονται πιο κοντά στην πλειοψηφική κλάση. Αντίθετα, για τα δεδομένα της μειοψηφικής κλάσης που είναι πιο εύκολο να ταξινομηθούν, παράγονται λιγότερα συνθετικά δείγματα. Αυτή η προσαρμοστικότητα έχει ως σκοπό την εξισορρόπηση της κατανομής των κλάσεων εστιάζοντας στην απόδοση των μοντέλων Μηχανικής Μάθησης και στην βελτίωση της μέσης ακρίβειας.

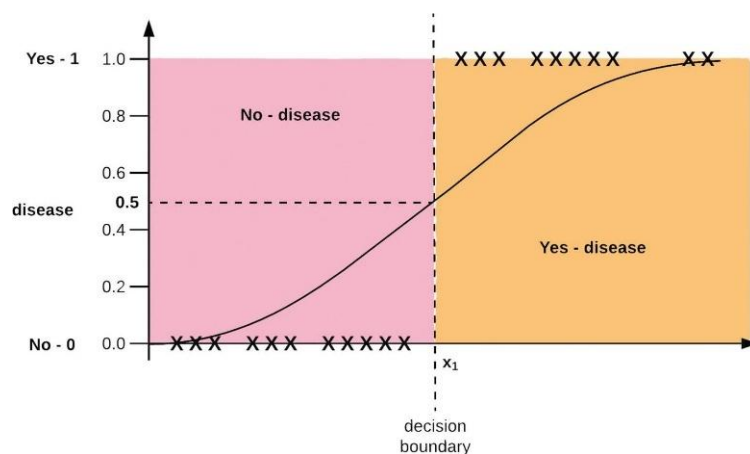
4.3 Τεχνικές Μηχανικής Μάθησης

Η Μηχανική Μάθηση είναι ένας ταχέως εξελισσόμενος τομέας της τεχνητής νοημοσύνης που επικεντρώνεται στην ανάπτυξη αλγορίθμων και μοντέλων που επιτρέπουν στους υπολογιστές να μαθαίνουν από δεδομένα και να βελτιώνουν την απόδοσή τους χωρίς την ανάγκη προγραμματισμού για κάθε πιθανό σενάριο. Μέσω της χρήσης μοντέλων που βασίζονται στην αξιοποίηση των δεδομένων (*data-driven models*), οι εκπαιδευμένοι αλγόριθμοι Μηχανικής Μάθησης έχουν τη δυνατότητα να αναλύουν τα δεδομένα σε πραγματικό χρόνο και να εντοπίζουν συσχετίσεις με παλαιότερες καταστάσεις, ανιχνεύοντας τυχόν ανωμαλίες εγκαίρως. Οι βιομηχανίες επωφελούνται σημαντικά από τις παρούσες εξελίξεις καθώς μπορούν να προετοιμάσουν μοντέλα τα οποία να προβλέπουν με ακρίβεια τη μελλοντική συμπεριφορά των συστημάτων τους, να εντοπίζουν δυσλειτουργίες, να αποστέλλουν ειδοποιήσεις, να δημιουργούν αυτόματες αιτήσεις εργασίας και να λαμβάνουν αποφάσεις με βάση τα πραγματικά δεδομένα, καθιστώντας τη λειτουργία τους περισσότερο άμεση και αποδοτική.

Τα μοντέλα Μηχανικής Μάθησης που χρησιμοποιούνται για ταξινόμηση (*classification*) στοχεύουν στην κατηγοριοποίηση δεδομένων σε προκαθορισμένες κατηγορίες, βάσει των δεδομένων εκπαίδευσης. Στην προβλεπτική συντήρηση, η ταξινόμηση εφαρμόζεται για την ανίχνευση και πρόβλεψη πιθανών αστοχιών εξοπλισμού, καθώς και για τον προσδιορισμό της φύσης και της αιτίας του σφάλματος. Μέσω αυτής της διαδικασίας είναι δυνατόν να εντοπιστεί και το είδος της βλάβης, εξασφαλίζοντας την ταχεία απόκρισή και άμεση αντιμετώπιση από το προσωπικό συντήρησης. Στο παρών κεφάλαιο θα παρουσιαστούν πέντε δημοφιλής αλγόριθμοι Μηχανικής Μάθησης για προβλήματα ταξινόμησης, οι οποίοι αξιοποιήθηκαν για τις ανάγκες της Διπλωματικής και πιο συγκεκριμένα για τον προσδιορισμό της κατάστασης της ανεμογεννήτριας.

4.3.1 Λογική Παλινδρόμηση (Logistic Regression – LR)

Το Logistic Regression (γνωστό επίσης ως logit ή sigmoid function) αποτελεί έναν αλγόριθμο Μηχανικής Μάθησης που χρησιμοποιείται κυρίως για τη ταξινόμηση δυαδικών κλάσεων, μετατρέποντας την έξοδο της συνάρτησης κόστους σε τιμές μεταξύ 0 και 1. Ο συγκεκριμένος αλγόριθμος είναι παρόμοιος με αυτόν της γραμμικής παλινδρόμησης (*Linear Regression*) με την διαφορά ότι ο πρώτος εκμεταλλεύεται τις δυνατότητες της συνάρτησης sigmoid.

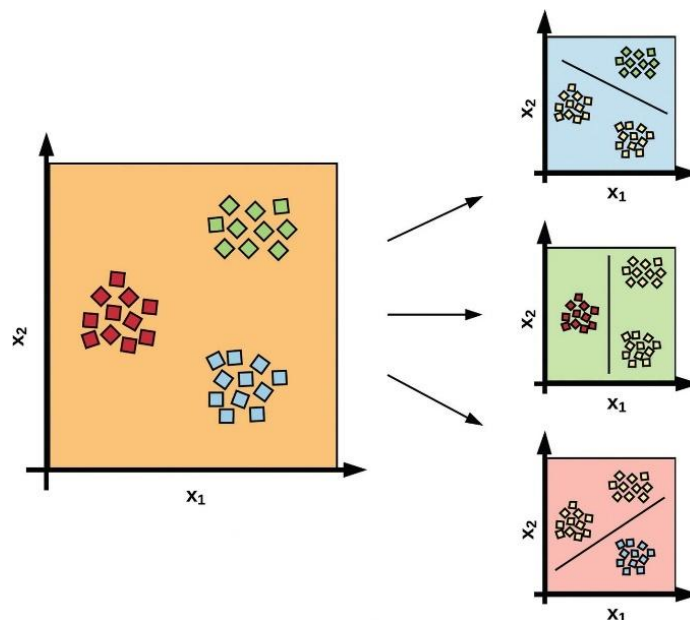


Εικόνα 16. Sigmoid Συνάρτηση, Πηγή: [34].

Η Εικόνα 16 παρουσιάζει την μορφή της sigmoid συνάρτησης, η οποία φέρει το σχήμα καμπύλης S. Παρατηρείται, ότι όσο η έξοδος έχει ως αποτέλεσμα αρνητικές τιμές που τείνουν προς το αρνητικό άπειρο, τόσο το αποτέλεσμα πλησιάζει προς το 0, ενώ αντίθετα όσο η έξοδος αυξάνεται προς τις θετικές τιμές και το θετικό άπειρο, τόσο το αποτέλεσμα πλησιάζει προς το 1. Η sigmoid συνάρτηση χρησιμοποιείται για να αποδώσει την πιθανότητα ορθής ταξινόμησης, υποδεικνύοντας την πιο πιθανή κατηγορία για κάθε παρατήρηση.

Τα συστήματα εκμάθησης λογικής παλινδρόμησης χρησιμοποιούν τη συνάρτηση κόστους (*log-loss*) ώστε να διορθώνουν τυχόν αποκλίσεις των προβλέψεων τους σε σχέση με τις πραγματικές τιμές. Η λογαριθμική απώλεια έχει ως στόχο να τιμωρεί όταν οι εκτιμήσεις είναι εκτός των πραγματικών τιμών. Τα μοντέλα βελτιώνονται με τη χρήση της κατάβασης κλίσης (*Gradient Descent*), η οποία βοηθά στον εντοπισμό των παραμέτρων εκείνων, όπου ελαχιστοποιούν τη συνάρτηση κόστους. Δια μέσου της διαδικασίας αυτής οι αλγόριθμοι προσαρμόζονται στο να προβλέπουν με μεγαλύτερη ακρίβεια τη σωστή κλάση.

Η μέθοδος αυτή έχει επεκταθεί και σε προβλήματα πολλαπλών κλάσεων (*multi-class classification*), όπου το σύστημα θα πρέπει να επιλέξει να κατηγοριοποιήσει μια παρατήρηση μεταξύ πολλών καταστάσεων. Υπάρχουν δύο βασικές προσεγγίσεις πολυωνυμικής ταξινόμησης, η One-vs-Rest και η softmax Regression [34]. Στην Εικόνα 17 απεικονίζεται ένα παράδειγμα One-vs-Rest, όπου ο αλγόριθμος προσπαθεί να ταξινομήσει κάθε δείγμα στη σωστή κλάση δημιουργώντας ξεχωριστούς ταξινομητές δυαδικής παλινδρόμησης, τα δείγματα βρίσκονται στον διδιάστατο χώρο των χαρακτηριστικών x_1 και x_2 . Η τελική ταξινόμηση γίνεται επιλέγοντας την κλάση με τη μεγαλύτερη προβλεπόμενη πιθανότητα.



Εικόνα 17. One-vs-Rest Πολυωνυμική παλινδρόμηση, Πηγή: [34].

Η δεύτερη προσέγγιση χρησιμοποιεί τη συνάρτηση **softmax** για την πρόβλεψη της πιθανότητας κάθε δείγματος να ανήκει σε κάθε μία από τις κλάσεις ταυτόχρονα. Η συνάρτηση softmax κανονικοποιεί τις πιθανότητες έτσι ώστε το άθροισμα τους να είναι ίσο με 1, επιτρέποντας στον αλγόριθμο να αποδώσει ένα δείγμα στην κλάση με την υψηλότερη πιθανότητα. Η συνάρτηση

κόστους που χρησιμοποιείται συνήθως είναι η *cross-entropy loss*, μέσω της οποίας υπολογίζεται η απόκλιση μεταξύ των προβλεπόμενων πιθανοτήτων και της πραγματικής κατανομής των κλάσεων στοχεύοντας στην ελαχιστοποίηση της κατά την εκπαίδευση. Η Εξίσωση 8 παρουσιάζει τον τύπο της συνάρτησης softmax.

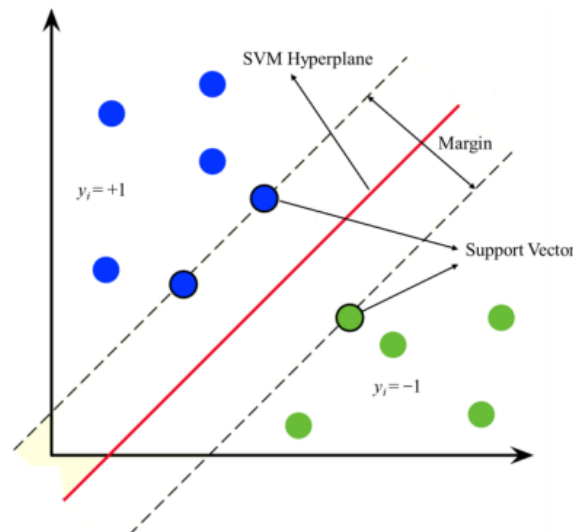
$$P(y = k|x) = (e^{\beta_k * x}) / (\sum_{j=1}^K e^{\beta_j * x}) \quad (8)$$

Όπου:

- $P(y = k|x)$ – η πιθανότητα το δείγμα x να ανήκει στην κλάση k
- K – ο συνολικός αριθμός κλάσεων
- β_k – το διάνυσμα βαρών της κλάσης k
- x – το διάνυσμα χαρακτηριστικών του δείγματος

4.3.2 Μηχανές Διανυσμάτων Υποστήριξης (Support Vector Machines – SVM)

Οι Μηχανές Διανυσμάτων Υποστήριξης είναι αλγόριθμοι Μηχανικής Μάθησης που αξιοποιούνται ευρέως για προβλήματα ταξινόμησης (*classification*) και παλινδρόμησης (*regression*). Για τα προβλήματα ταξινόμησης, οι αλγόριθμοι SVM χρησιμοποιούν ένα μοντέλο που αναπαριστά τα δεδομένα σε ένα χώρο υψηλής διαστασιμότητας και εν συνεχεία δημιουργούν ένα υπέρ-επίπεδο (*hyperplane*) που διαχωρίζει τα δεδομένα σε διαφορετικές κλάσεις, επιτυγχάνοντας τον μέγιστο διαχωρισμό μεταξύ των σημείων κάθε κλάσης. Στόχος του μοντέλου είναι ο υπολογισμός του υπέρ-επιπέδου με τέτοιο τρόπο ώστε να μεγιστοποιηθεί η απόσταση (*margin*) μεταξύ αυτού και των πλησιέστερων δειγμάτων κάθε κλάσης, γνωστά και ως *Support Vectors*. Η Εικόνα 18 παρουσιάζει την βασική λειτουργία των Μηχανών Διανυσμάτων Υποστήριξης.



Εικόνα 18. Μηχανές Διανυσμάτων Υποστήριξης, Πηγή: [45].

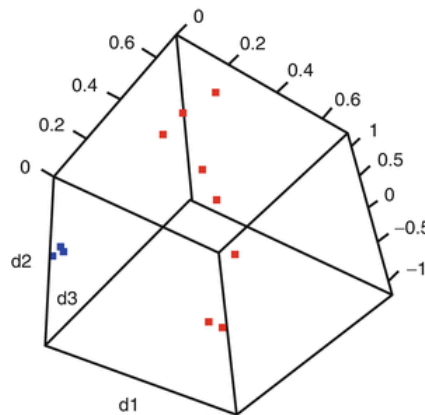
Το υπέρ-επίπεδο υπολογίζεται έτσι ώστε να εξασφαλίζει τον βέλτιστο διαχωρισμό μεταξύ κλάσεων και την ελαχιστοποίηση των σφαλμάτων, δίνεται σύμφωνα με τον τύπο της Εξίσωσης (9).

$$w * x + b = 0 \quad (9)$$

Όπου:

- w – το διάνυσμα βαρών που καθορίζει την κατεύθυνση και κλίση του υπέρ-επιπέδου
- x – το διάνυσμα των χαρακτηριστικών των σημείων δεδομένων
- b – η παράμετρος μετατόπισης (*bias*)

Οι αλγόριθμοι SVM διαχωρίζονται στα γραμμικά (*linear*) και μη-γραμμικά (*nonlinear*) μοντέλα [35]. Τα πρώτα αξιοποιούν γραμμικές συναρτήσεις – *linear kernel* – και χρησιμοποιούνται όταν το ευθύγραμμο υπέρ-επίπεδο είναι αρκετό ώστε να διαχωριστούν τα δεδομένα. Σε αυτή την περίπτωση, ο αλγόριθμος δημιουργεί ένα γραμμικό όριο αποφάσεων, επιτυγχάνοντας ταχύτητα και απλότητα στη μοντελοποίηση. Ωστόσο, όταν δεδομένα δεν είναι δυνατόν να διαχωριστούν γραμμικά εφαρμόζονται η μη γραμμικές συναρτήσεις – *non-linear kernel* – όπου προβάλλουν τα δεδομένα σε έναν χώρο υψηλότερης διάστασης (Εικόνα 18). Μέσω αυτής της μεθόδου δύναται να βρεθεί ένα διαχωριστικό υπέρ-επίπεδο όπου να χειρίζεται πιο σύνθετες δομές δεδομένων καθιστώντας τον αλγόριθμο SVM αποτελεσματικό και σε πιο περίπλοκα προβλήματα κατηγοριοποίησης.

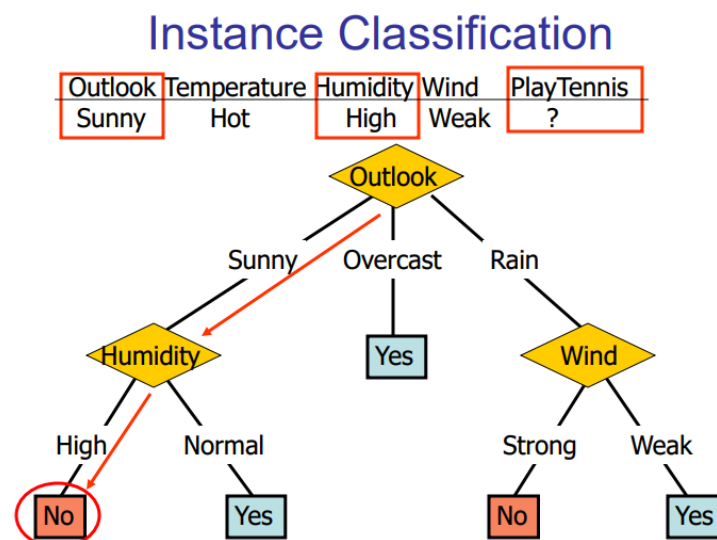


Εικόνα 19. Ταξινόμηση μη γραμμικά διαχωρίσιμων κλάσεων σε χώρο υψηλότερων διαστάσεων, Πηγή: [35].

Οι Μηχανές Διανυσμάτων Υποστήριξης αρχικά σχεδιάστηκαν για την διάκριση δυαδικών κλάσεων ($k = 2$) με χρήση κριτηρίου μέγιστης απόστασης (*maximum margin*). Ωστόσο, οι ανάγκες δημιουργίας και υλοποίησης συστημάτων αναγνώρισης πολλαπλών κλάσεων οδήγησαν στην ανάπτυξη μεθόδων που επιτρέπουν την εφαρμογή των SVM σε αντίστοιχα ζητήματα [36]. Οι δύο πιο αντιπροσωπευτικές προσεγγίσεις για αυτή τη διαδικασία είναι οι τεχνικές *one-versus-rest (1VR)* και *one-versus-one (1V1)*. Η τεχνική *One-vs-One* δημιουργεί ένα ξεχωριστό SVM για κάθε ζευγάρι κλάσεων, με αποτέλεσμα, για k κλάσεις, να προκύπτουν $k(k-1)/2$ δυαδικά SVM μοντέλα, καθένα από τα οποία αφορά ένα ζευγάρι κλάσεων. Κατά την πρόβλεψη, χρησιμοποιείται ένας πλειοψηφικός κανόνας (*majority voting*), όπου η κλάση που λαμβάνει τις περισσότερες "ψήφους" επιλέγεται. Στην προσέγγιση *One-vs-Rest (1vR)*, δημιουργούνται k ξεχωριστοί δυαδικοί ταξινομητές για ένα πρόβλημα ταξινόμησης με k κλάσεις. Κάθε ταξινομητής εκπαιδεύεται με τα δεδομένα της αντίστοιχης κλάσης ως θετικά παραδείγματα και τα δεδομένα των υπόλοιπων $k-1$ κλάσεων ως αρνητικά παραδείγματα. Κατά την πρόβλεψη, η ετικέτα της κλάσης καθορίζεται από τον δυαδικό ταξινομητή που δίνει την υψηλότερη τιμή εξόδου.

4.3.3 Δέντρα Απόφασης (Decision Trees)

Τα Δέντρα Απόφασης (Decision Trees) είναι ένας δημοφιλής αλγόριθμος μηχανικής μάθησης που χρησιμοποιείται για την επίλυση προβλημάτων ταξινόμησης και παλινδρόμησης. Λειτουργούν με τη μορφή ιεραρχικής δομής που μοιάζει με δέντρο, όπου οι κόμβοι αναπαριστούν χαρακτηριστικά (*features*) των δεδομένων. Η διαδικασία ταξινόμησης ενός δεδομένου πραγματοποιείται μέσω μιας σειράς ερωτήσεων σχετικά με τα χαρακτηριστικά του, οι ερωτήσεις επίσης μπορούν να αναπαριστώνται και ως ένα σύνολο κανόνων IF-THEN. Κάθε ερώτηση ενσωματώνεται σε έναν κόμβο, κάτω από τον οποίο προστίθενται οι κόμβοι-παιδιά που αντιστοιχούν στην κάθε πιθανή απάντηση. Μέσω της διαδικασίας των ερωτήσεων διαμορφώνεται μια ιεραρχία που κωδικοποιείται ως δέντρο, η οποία επιτρέπει την οπτική αναπαράσταση των διαδικασιών λήψης αποφάσεων. Αυτή η δομή παρέχει μια φυσική και διαισθητική κατανόηση του τρόπου με τον οποίο το μοντέλο καταλήγει σε συγκεκριμένες προβλέψεις. Η κατηγοριοποίηση ενός αντικείμενου γίνεται ακολουθώντας την πορεία από τον ανώτατο κόμβο προς τον τελικό κόμβο (φύλλο). Τέλος, το αντικείμενο κατηγοριοποιείται στην κατηγορία που σχετίζεται με το φύλλο στο οποίο φτάνει. Το βασικό πλεονέκτημα των Δέντρων Απόφασης είναι η ικανότητά τους να διαχειρίζονται προβλήματα ταξινόμησης πολλαπλών κλάσεων με φυσικό τρόπο παρέχοντας ένα σαφές και διαφανές μοντέλο ταξινόμησης. Η Εικόνα 20, παρουσιάζει ένα παράδειγμα αναπαράστασης ενός Δέντρου απόφασης για την κατηγοριοποίηση του *Target Class*: “Play Tennis”.



Εικόνα 20. Ταξινόμηση μέσω Δέντρου Απόφασης. (Σταματάτος Ευστάθιος, σημειώσεις μαθήματος Μηχανικής Μάθησης, πανεπιστήμιο Αιγαίου [2022])

Το Decision Tree αναλύει ένα σύνολο δεδομένων ακολουθώντας μια διαδικασία διαδοχικών διαχωρισμών (*splits*). Κάθε διαχωρισμός γίνεται με βάση ένα χαρακτηριστικό που επιλέγεται ώστε να βελτιστοποιεί ένα κριτήριο πληροφορίας. Τα κριτήρια αυτά μετρούν πόσο τακτοποιημένος είναι ο διαχωρισμός των κλάσεων μετά από κάθε διαίρεση. Η διαδικασία διαχωρισμού συνεχίζεται μέχρι να φτάσουμε σε φύλλα, που αναπαριστούν τις τελικές προβλέψεις για τις κλάσεις. Τα δύο πιο συχνά κριτήρια πληροφορίας που χρησιμοποιούνται για τη μέτρηση της ποιότητας κάθε διαχωρισμού είναι η *Εντροπία* και ο *Δείκτης Gini* [37]:

Εντροπία (Entropy): Μετρά το βαθμό αταξίας ή αβεβαιότητας στην κατανομή των κατηγοριών ενός συνόλου. Ουσιαστικά, καθορίζει πόσο μπερδεμένα είναι κάποια αντικείμενα που βρίσκονται σε διαφορετικές κατηγορίες μετά από κάθε διαχωρισμό (*split*). Με αυτό τον τρόπο, αν οι κλάσεις είναι ομοιόμορφα κατανεμημένες μετά τον διαχωρισμό η εντροπία λαμβάνει τιμή 1, ενώ αν όλες οι παρατηρήσεις είναι θετικές ή αρνητικές η εντροπία μηδενίζεται. Η Εντροπία δίνεται σύμφωνα με τον τύπο της Εξίσωσης (10).

$$H(S) = - \sum_{i=1}^c p_i \log_2 p_i \quad (10)$$

Όπου:

- $H(S)$ – η Εντροπία
- c – ο αριθμός των κλάσεων
- p_i – η πιθανότητα ενός δείγματος να ανήκει στη κλάση i

Δείκτης Gini (Gini Index): Εκφράζει την πιθανότητα να κατηγοριοποιηθεί λανθασμένα ένα τυχαίο δείγμα, αν επιλεγθεί μια τυχαία κλάση με βάση την κατανομή των δεδομένων στο συγκεκριμένο σύνολο. Ο Δείκτης Gini είναι 0 όταν όλα τα στοιχεία ανήκουν στην ίδια κλάση και αυξάνεται όσο αυξάνεται η ετερογένεια του συνόλου δεδομένων. Δηλαδή, όσο πιο μικρό είναι το αποτέλεσμα τόσο πιο ομοιόμορφα είναι τα δεδομένα. Ο Δείκτης Gini δίνεται από τον τύπο της Εξίσωσης (11).

$$G(S) = 1 - \sum_{i=1}^c p_i^2 \quad (11)$$

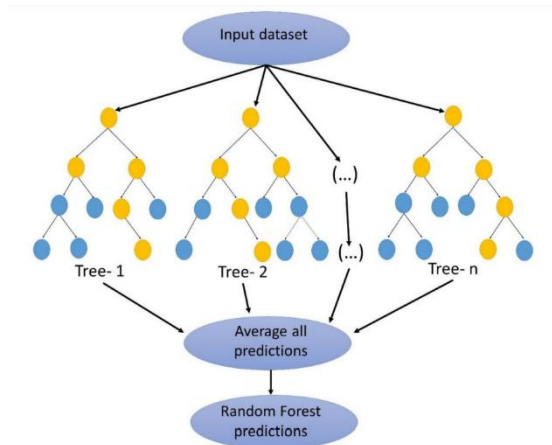
Όπου:

- $G(S)$ – ο δείκτης Gini
- c – ο αριθμός των κλάσεων
- p_i – η πιθανότητα ενός δείγματος να ανήκει στη κλάση i

Κατά τη μηχανική μάθηση, ο όρος *Overfitting* αναφέρεται στη διαδικασία κατά την οποία ένα μοντέλο προσαρμόζεται υπερβολικά στα δεδομένα εκπαίδευσης, καταλήγοντας να χάνει τη δυνατότητά του να γενικεύει σε νέα και άγνωστα δεδομένα. Στη περίπτωση αυτή, το μοντέλο αποκτά εξαιρετική απόδοση στα δεδομένα εκπαίδευσης ωστόσο, η απόδοση του μειώνεται σε δεδομένα δοκιμής ή σε πραγματικές εφαρμογές. Θα πρέπει να σημειωθεί ότι ορισμένοι αλγόριθμοι είναι επιρρεπείς στο *overfitting*, ενώ άλλοι είναι πιο ανθεκτικοί σε αυτό το φαινόμενο. Τα Δέντρα Απόφασης ανήκουν στην πρώτη κατηγορία καθώς έχουν την τάση να δημιουργούν πολύπλοκες δομές και αναπτύσσουν τόσο πολύ το δέντρο όπου καταλήγουν να προσαρμόζονται ακριβώς στις παρατηρήσεις του εκπαιδευτικού συνόλου, συμπεριλαμβανομένων των θορυβωδών δεδομένων. Για την προστασία των μοντέλων από το φαινόμενο του *overfitting* συνήθως χρησιμοποιούνται τεχνικές όπως η κοπή (*pruning*), που αφαιρεί κόμβους από το δέντρο που δεν συμβάλλουν σημαντικά στη διακριτική ικανότητα του μοντέλου, επιτρέποντας του να διατηρήσει την γενικότητα του.

4.3.4 Τυχαίο Δάσος - Random Forest

Ο Random Forest είναι ένας *ensemble* αλγόριθμος, δηλαδή βασίζεται σε ένα σύνολο απλών μοντέλων για να παράγει μια πιο ισχυρή και ακριβή τελική πρόβλεψη (Εικόνα 21). Κατά τη φάση της εκπαίδευσης, δημιουργούνται πολλά δέντρα απόφασης καθένα από τα οποία όντας ανεξάρτητο



Εικόνα 21. Τυχαίο Δάσος, Πηγή: [46].

από τα υπόλοιπα και μαθαίνοντας από ένα διαφορετικό δείγμα δεδομένων και ένα τυχαίο υποσύνολο χαρακτηριστικών (*features*) [38]. Στα προβλήματα ταξινόμησης τα δέντρα προσπαθούν να διαχωρίσουν τα δεδομένα σε διάφορες κατηγορίες, ωστόσο το αποτέλεσμα προκύπτει από τη «ψηφός» ή το μέσο όρο των αποτελεσμάτων όλων των δέντρων του δάσους. Εφόσον ολοκληρωθεί η δημιουργία και η εκπαίδευση όλων των δέντρων, ο Random Forest συνδυάζει τις προβλέψεις τους για να καθορίσει το τελικό αποτέλεσμα. Στην περίπτωση της ταξινόμησης πολλαπλών κλάσεων, το αποτέλεσμα επιλέγεται από την πλειοψηφία των δέντρων (*majority voting*) ώστε να αποφασιστεί η κατηγοριοποίηση του δείγματος.

Για την ταξινόμηση με Random Forest, η πρόβλεψη της κλάσης C για ένα δείγμα x δίνεται από την Εξίσωση (12) ως εξής:

$$C(x) = \arg_c \max \sum_{i=1}^N I(T_i(x) = c) \quad (12)$$

Όπου:

- N – ο αριθμός των δέντρων στο δάσος
- I – μια ένδειξη (*indicator function*) που παίρνει την τιμή 1 αν το δέντρο T_i ταξινομήσει το δείγμα στην κλάση c και 0 σε διαφορετική περίπτωση
- T_i – είναι η πρόβλεψη του i -ου δέντρου για το δείγμα x

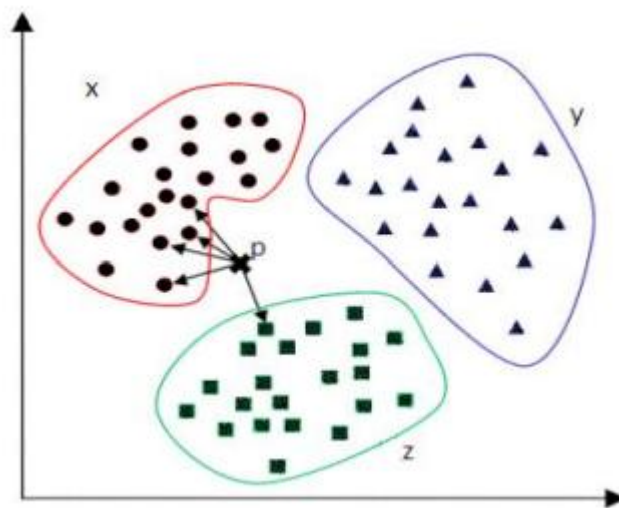
Τα πλεονεκτήματα του Random Forest είναι:

- ✓ Υψηλή ακρίβεια: Λόγω του ensemble χαρακτηριστικού του και του συνδυασμού αποφάσεων, ο Random Forest συνήθως επιτυγχάνει μεγαλύτερη ακρίβεια από μεμονωμένα δέντρα απόφασης.
- ✓ Μειωμένο Overfitting: Η χρήση πολλαπλών δέντρων και η τυχαία επιλογή χαρακτηριστικών μειώνουν το overfitting, ένα συχνό πρόβλημα των δέντρων απόφασης.
- ✓ Καλή Γενίκευση: Η τεχνική bootstrapping προσδίδει καλή απόδοση σε δεδομένα που το μοντέλο δεν έχει δει.
- ✓ Ευελιξία: Χρησιμοποιείται τόσο για προβλήματα ταξινόμησης (*classification*) όσο και για προβλήματα (*regression*).

- ✓ Ευρεία εφαρμογή: Οι Random Forest έχουν βρει εφαρμογή σε διαφορετικά πεδία, όπως η αναγνώριση προτύπων, η ανάλυση εικόνων κλπ.

4.3.5 Αλγόριθμος *k*-Nearest Neighbors (*k*-NN)

Ο αλγόριθμος *k*-Nearest Neighbors (*k*-NN) είναι ένας μη παραμετρικός, εποπτευόμενος αλγόριθμος ταξινόμησης, ο οποίος χρησιμοποιείται για την κατηγοριοποίηση δεδομένων σε διαφορετικές κατηγορίες (classes). Ανήκει στις μεθόδους μάθησης βάσει περιπτώσεων (*instance-based learning*), οι οποίες συχνά αποκαλούνται και αλγόριθμοι «τεμπέλικης μάθησης» (*lazy learners*) [39]. Θεωρείται από τους πιο απλούς και κατανοητούς αλγορίθμους μηχανικής μάθησης και είναι ιδανικός για περιπτώσεις όπου η γραμμικότητα των δεδομένων δεν μπορεί να θεωρηθεί δεδομένη. Σε αντίθεση με τους λεγόμενους αλγόριθμους "πρόθυμης μάθησης" (*eager learning*), όπως είναι τα δέντρα απόφασης, τα νευρωνικά δίκτυα και τα δίκτυα *Bayes*, ο αλγόριθμος τεμπέλικης μάθησης δεν απαιτεί πολύ υπολογιστικό χρόνο στη φάση της εκπαίδευσης. Αντίθετα, η απαιτούμενη υπολογιστική ισχύς αυξάνεται κατά την διαδικασία της ταξινόμησης, καθώς ο αλγόριθμος προσπαθεί να υπολογίσει την απόσταση του νέου δείγματος από όλα τα δείγματα εκπαίδευσης.



Εικόνα 22. Παράδειγμα *k*-NN Ταξινόμηση, Πηγή: [39].

Η βασική ιδέα του αλγορίθμου *k*-NN είναι η εύρεση των *k* πιο κοντινών δειγμάτων (*neighbors*) στο χώρο χαρακτηριστικών (*feature space*) ενός νέου δεδομένου (*unlabeled sample*). Ο αλγόριθμος ακολουθεί τα παρακάτω βήματα, σύμφωνα και με την Εικόνα (22) του παραδείγματος: 1) Αρχικώς, ορίζεται ο αριθμός των πλησιέστερων γειτόνων που θα εξεταστούν κατά τη διαδικασία ταξινόμησης ενός νέου δείγματος. Στη περίπτωση του παραδείγματος ορίζεται $k=5$. 2) Στη συνέχεια υπολογίζεται η απόσταση του νέου δείγματος από κάθε δείγμα του συνόλου της εκπαίδευσης, χρησιμοποιώντας την Ευκλείδεια απόσταση. 3) Οι υπολογιζόμενες αποστάσεις ταξινομούνται κατ' αύξουσα σειρά και 4) επιλέγονται οι k ($=5$) πρώτες τιμές που αντιστοιχούν στους κοντινότερους γείτονες. Στο παράδειγμα της Εικόνας επιλέγονται οι 5 πλησιέστεροι γείτονες, τέσσερις εκ των οποίων ανήκουν στην κλάση *x* και ένας στη *z*. 5) Τέλος χρησιμοποιείται απλά η ψήφος πλειοψηφίας (*majority voting*) για την πρόβλεψη της κλάσης του νέου δείγματος. Ουσιαστικά, η κατηγορία που εμφανίζεται πιο συχνά ανάμεσα στους γείτονες επιλέγεται ως κατηγορία του νέου δείγματος. Στη περίπτωση μας το

νέο δείγμα p κατατάσσεται στην κατηγορία x καθώς οι τέσσερις από τους πέντε πλησιέστερους γείτονες ανήκουν στη παραπάνω κλάση.

5

Ανάπτυξη μοντέλου Μηχανικής Μάθησης για τον εντοπισμό πιθανόν αστοχιών και κατηγοριοποίησης σφαλμάτων

5.1 Μεθοδολογία Επεξεργασίας Δεδομένων

Για την ανάπτυξη αξιόπιστων μοντέλων Μηχανικής Μάθησης, ικανών να προβλέπουν με επιτυχία την κατάσταση των συστημάτων βάσει των μεταβολών των σημάτων που λαμβάνουν από το πεδίο, κρίνεται απαραίτητη η κατάλληλη επεξεργασία των δεδομένων εκπαίδευσης. Όπως περιεγράφηκε στη προηγούμενη ενότητα, το εξεταζόμενο Dataset (*IIoT Wind Turbine Analytics*) περιέχει έντονη ασυμμετρία κλάσεων στο αρχικό και ακατέργαστο σύνολο δεδομένων. Αριθμεί 49.134 συνολικές παρατηρήσεις εκ των οποίων μόλις οι 553 αναφέρονται σε περιστατικά αστοχίας της ανεμογεννήτριας, διαμοιρασμένα και αυτά με την σειρά τους σε πέντε (5) διακριτές κλάσεις. Επομένως, σκοπός της περιγραφόμενης διαδικασίας είναι η αντιμετώπιση του προβλήματος της ανομοιομορφίας, εξασφαλίζοντας ποιοτικά δεδομένα εκπαίδευσης για τα μοντέλα Μηχανικής Μάθησης. Γεγονός που επιτυγχάνεται μέσω τεχνικών επεξεργασίας δεδομένων που διασφαλίζουν τη διατήρηση των κρίσιμων πληροφοριών, ενώ ταυτόχρονα αποφεύγουν την υπερβολική κυριαρχία των συχνότερων παρατηρήσεων. Στη συνέχεια παρουσιάζονται τα βήματα και οι τεχνικές που ακολουθήθηκαν για την ανάπτυξη του μοντέλου που έχει ως στόχο τον εντοπισμό πιθανών αστοχιών και την κατηγοριοποίηση σφαλμάτων.

Βήμα 1^ο

Αρχικώς θα έπρεπε να προστατευτεί το μοντέλο από την πιθανότητα *overfitting* και να εξασφαλιστεί μεγαλύτερη ισορροπία κατά την εκπαίδευση, καθώς η πλειοψηφούσα κλάση παρουσίαζε υπερβολικά μεγάλο αριθμό δειγμάτων έναντι των υπολοίπων, ήταν υπαρκτός ο κίνδυνος ο αλγόριθμος να εκπαιδευτεί ευνοώντας την σε σχέση με τις υπόλοιπες. Για τον λόγο αυτό, δεν χρησιμοποιήθηκε το σύνολο των δειγμάτων της κυρίαρχης κλάσης, αντίθετα ελήφθη ένα τυχαίο υποσύνολο παρατηρήσεων. Ο αριθμός των δειγμάτων που επιλέχθηκε ήταν αποτέλεσμα δοκιμής

διαφορετικών σεναρίων με στόχο την εύρεση της βέλτιστης ισορροπίας που θα εξασφάλιζε τη μεγαλύτερη δυνατή ακρίβεια και σταθερότητα του μοντέλου. Στη συνέχεια, η ανίχνευση και διαχείριση των ακραίων τιμών (*Outlier Detection*) αποτέλεσε ένα από τα βασικά στάδια του πρώτου βήματος για την ενίσχυση της ποιότητας των δεδομένων. Αντί της πλήρους διαγραφής συνολικά των παρατηρήσεων που περιείχαν ένα *feature* με ακραία τιμή, τα *Outliers* αντικαταστάθηκαν με τα πλησιέστερα επιτρεπτά όρια (*lower και upper limit*), ώστε να διατηρηθεί η πληροφορία κάθε δείγματος αποτρέποντας τη διαστρέβλωση των στατιστικών χαρακτηριστικών των δεδομένων που θα επηρέαζε αρνητικά την απόδοση των μοντέλων. Στη συνέχεια, εφαρμόστηκε η μείωση των χαρακτηριστικών (*Feature Reduction*) μέσω του συντελεστή μεταβλητότητας, για την αφαίρεση των *Features* που δεν συνεισφέρουν ουσιαστική και ποιοτική πληροφορία στο μοντέλο εκπαίδευσης. Αξιοποιώντας το *coefficient of variation*, εντοπίστηκαν χαρακτηριστικά των οποίων η μεταβλητότητα παρέμενε χαμηλή και επομένως αφαιρέθηκαν καθώς η προστιθέμενη αξία τους στη διαδικασία της μάθησης ήταν μικρή. Επόμενο στάδιο ήταν η εφαρμογή κανονικοποίησης στα δεδομένα για την διασφάλιση της ισότιμης συνεισφοράς όλων των *features* κατά την εκπαίδευση, αποτρέποντας με αυτή τη τεχνική την κυριαρχία συγκεκριμένων χαρακτηριστικών και βελτιώνοντας τη σύγκλιση των μοντέλων. Με την ολοκλήρωση της αναγκαίας προ-επεξεργασίας των δεδομένων, το τελικό αποτέλεσμα διαχωρίστηκε σε σύνολο εκπαίδευσης (80%) και δοκιμής (20%).

Βήμα 2^ο

Το δεύτερο βήμα στη μεθοδολογία επεξεργασίας των δεδομένων περιλαμβάνει την επιπλέον μείωση των χαρακτηριστικών. Όπως περιγράφηκε και στο προηγούμενο κεφάλαιο, η ύπαρξη πολλών *features* σε ένα σύνολο δεδομένων δύναται να επηρεάσει αρνητικά την ακρίβεια του μοντέλου κατηγοριοποίησης. Για τον λόγο αυτό κρίνεται απαραίτητος ο έλεγχος αξιοποίησης ή μη, της τεχνικής ανάλυσης κύριων συνιστωσών (*PCA*), που έχει ως σκοπό τη μετατροπή των αλληλοσχετιζόμενων χαρακτηριστικών σε μη συσχετιζόμενα χαρακτηριστικά (*κύριες συνιστώσες*) που αντικατοπτρίζουν γραμμικούς συνδυασμούς των αρχικών μεταβλητών. Ωστόσο, κατά την χρήση της τεχνικής ελλοχεύει ο κίνδυνος απώλειας κρίσιμης πληροφορίας, η οποία μπορεί να είναι σημαντική για την απόδοση του μοντέλου. Με βάση το συγκεκριμένο πλαίσιο, ο αλγόριθμος εκπαιδεύτηκε με δύο τρόπους:

- Χρησιμοποιώντας το *PCA* για τη μείωση της διαστασιμότητας
- Διατηρώντας την αρχική διάσταση των δεδομένων και παραλείποντας το *PCA*

Στόχος η σύγκριση των αποτελεσμάτων των δύο μεθόδων και η αξιολόγηση της απόδοσης ανάλυσης κύριων συνιστωσών στο υπό εξέταση σύνολο δεδομένων.

Βήμα 3^ο

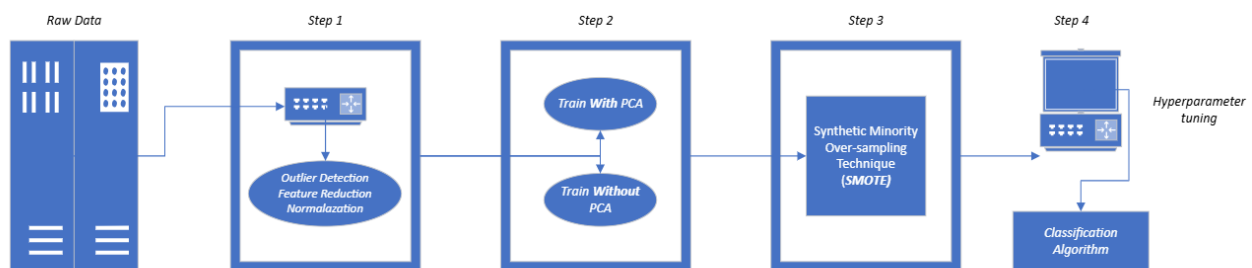
Το επόμενο βήμα της μεθοδολογίας εστιάζει στην αντιμετώπιση του προβλήματος της ανισορροπίας των κλάσεων. Όταν οι παρατηρήσεις των δειγμάτων περιέχουν ανομοιομορφία ως προς την κλάση –μία κλάση συγκεντρώνει την πλειοψηφία των παρατηρήσεων ενώ οι υπόλοιπες αποτελούνται από πολύ λιγότερα δείγματα – υπάρχει κίνδυνος μεροληψίας του μοντέλου ως προς την πλειοψηφική κατηγορία και εμφάνισης του «*παραδόξου της ακρίβειας*». Για τη βελτίωση της ισορροπίας εφαρμόζονται τεχνικές είτε υπό-δειγματοληψίας (*undersampling*) είτε υπέρ-δειγματοληψίας (*oversampling*). Κατόπιν δοκιμών και αξιολόγησης των πρώτων αποτελεσμάτων, επιλέχθηκε η χρήση της μεθόδου *SMOTE*, καθώς παρείχε τη βέλτιστη ακρίβεια έναντι των

υπολοίπων και παράλληλα διατηρούσε την κρίσιμη πληροφορία έναντι άλλων τεχνικών όπως της *undersampling* όπου κατά την αφαίρεση δειγμάτων της πλειοψηφικής κλάσης υπήρχε και απώλεια σημαντικής πληροφορίας. Η εφαρμογή της SMOTE οδήγησε σε βελτιωμένα αποτελέσματα δημιουργώντας συνθετικά δείγματα που εμπλούτιζαν τη μειονοτική κλάση χωρίς να προσθέτουν ανεπιθύμητο θόρυβο.

Βήμα 4^ο

Το τέταρτο και τελευταίο βήμα της μεθοδολογίας περιλαμβάνει τη μεγιστοποίηση της ακρίβειας του επιλεγμένου αλγόριθμου και της βελτιστοποίησης των αποτελεσμάτων του, μέσω της ρύθμισης των υπέρ-παραμέτρων (*hyperparameter tuning*). Εφόσον εκπαιδευτούν όλοι οι αλγόριθμοι και αξιολογηθούν τα αρχικά τους αποτελέσματα, θα επιλεγθεί ο επικρατέστερος αλγόριθμος βάσει την απόδοση και την ακρίβεια των προβλέψεων. Στη συνέχεια, πραγματοποιείται η διαδικασία ρύθμισης των υπέρ-παραμέτρων του συγκεκριμένου αλγόριθμου, με στόχο την αύξηση της απόδοσής του.

Η Εικόνα 23 παρουσιάζει συγκεντρωτικά τα τέσσερα βήματα της μεθοδολογίας επεξεργασίας δεδομένων και εκπαίδευσης των μοντέλων Μηχανικής Μάθησης:



Εικόνα 23. Μεθοδολογία επεξεργασίας δεδομένων και εκπαίδευσης μοντέλων Μηχανικής Μάθησης

5.2 Παρουσίαση Αποτελεσμάτων

Αρχικώς, θα πρέπει να σημειωθεί ότι τα αποτελέσματα της εκπαίδευσης των πέντε διαφορετικών μοντέλων μηχανικής μάθησης, όπως αυτά περιγράφηκαν στο παραπάνω κεφάλαιο και τα οποία ακολουθούν τη διαδικασία που απεικονίζεται στο *workflow* της Εικόνας 22, προκύπτουν μέσω της μεθόδου *10-fold cross-validation*. Η επιλογή αυτής της μεθόδου έχει ως στόχο τη διασφάλιση της αξιοπιστίας των αποτελεσμάτων, καθώς επιτρέπει την εκπαίδευση και την αξιολόγηση των μοντέλων σε πολλαπλά υποσύνολα.

Συγκεκριμένα, η *10-fold cross-validation* χωρίζει το διαθέσιμο σύνολο δεδομένων σε 10 ίσα μέρη (*folds*), όπου σε κάθε νέα επανάληψη ένα *fold* αξιοποιείται για την δοκιμή του μοντέλου (*validation*), ενώ τα υπόλοιπα εννέα χρησιμοποιούνται για την εκπαίδευση του. Η διαδικασία αυτή επαναλαμβάνεται έως ότου κάθε ένα από τα δέκα τμήματα χρησιμοποιηθούν για την δοκιμή του μοντέλου. Η μέθοδος αυτή προσφέρει μια πιο ακριβή εκτίμηση της απόδοσης του μοντέλου σε δεδομένα που δεν έχουν χρησιμοποιηθεί στην εκπαίδευση. Στόχος αποτελεί, δια μέσου της συγκεκριμένης τεχνικής τα διαφορετικά μοντέλα να αξιολογηθούν σε πολλαπλές μετρικές απόδοσης, εξάγοντας μία πιο αντικειμενική εικόνα της αποτελεσματικότητας κάθε μοντέλου.

Στο Πίνακα 1 παρουσιάζεται η ακρίβεια των πέντε διαφορετικών αλγορίθμων Μηχανικής Μάθησης με την χρήση της μεθόδου ανάλυσης κυρίων συνιστωσών και χωρίς αυτήν, προκειμένου να διερευνηθεί η επίδραση της μείωσης της διάστασης των χαρακτηριστικών στη συνολική απόδοση των μοντέλων. Σύμφωνα με τα αποτελέσματα παρατηρούμε ότι τα μοντέλα *Logistic Regression* και *SVM* παρουσιάζουν υψηλότερη ακρίβεια χωρίς τη χρήση της PCA (87.96% και 87.75%, αντίστοιχα) σε σύγκριση με την απόδοση με την PCA (83.28% και 85.90%). Το οποίο υποδεικνύει ότι οι συγκεκριμένοι αλγόριθμοι επηρεάζονται αρνητικά με την χρήση της PCA. Από την άλλη πλευρά, οι αλγόριθμοι *Decision Trees* και *Random Forest* εμφάνισαν ελαφρώς μικρές διαφορές, καθώς η απόδοσή τους ήταν αντίστοιχα 89.21% και 92.82% με την PCA, έναντι 90.58% και 92.89% χωρίς αυτήν. Οι αλγόριθμοι αυτοί φαίνεται να επηρεάζονται λιγότερο, καθώς δημιουργούν δένδροειδείς δομές για την ανάλυση δεδομένων διατηρώντας τις κρίσιμες διακρίσεις στις περισσότερες περιπτώσεις. Τέλος, αλγόριθμος *K-Nearest Neighbors (K-NN)* σημείωσε επίσης πολύ μικρή διαφορά στην απόδοσή του (89.50% με PCA έναντι 88.96% χωρίς PCA), καταδεικνύοντας ότι η χρήση της PCA μπορεί να έχει περιορισμένη επιρροή σε μοντέλα που στηρίζονται στην απόσταση μεταξύ των δεδομένων.

Πίνακας 1. Απόδοση μοντέλων MM με την χρήση PCA και χωρίς αυτή

	LR	SVM	Decision Trees	K-NN	Random Forest
Model With PCA	83.28%	85.90%	89.21%	89.50%	92.82%
Model Without PCA	87.96%	87.75%	90.58%	88.96%	92.89%

Συνολικά, παρουσιάζονται μικρές διαφορές στις αποδόσεις των μοντέλων, ωστόσο θα πρέπει να σημειωθεί ότι τα παραπάνω αποτελέσματα βασίζονται σε μία εκδοχή της PCA με διατήρηση του 99% της διασποράς των αρχικών δεδομένων. Επιπλέον, καθώς στόχος είναι η αύξηση της ακρίβειας του μοντέλου όσο το δυνατόν περισσότερο, η χρήση της PCA φαίνεται να επηρεάζει την απόδοση σε μικρό βαθμό, με πιθανή απώλεια σημαντικής πληροφορίας λόγω της μείωσης της διάστασης. Για τον λόγο αυτό, και δεδομένου ότι έχει ήδη χρησιμοποιηθεί μια μέθοδος μείωσης χαρακτηριστικών μέσω *feature selection*, δεν κρίνεται απαραίτητη η περαιτέρω εφαρμογή της PCA.

Ο Πίνακα 2 που ακολουθεί, συγκεντρώνει τις συνολικές μετρικές απόδοσης των διαφορετικών αλγορίθμων Μηχανικής Μάθησης που χρησιμοποιήθηκαν για την εκπαίδευση των μοντέλων, χωρίς την χρήση της ανάλυσης κυρίων συνιστωσών. Βάσει τα στοιχεία του πίνακα, προκύπτει ότι ο αλγόριθμος *Random Forest* παρουσιάζει την καλύτερη συνολική απόδοση, με την υψηλότερη μέση ακρίβεια (92.89%) καθώς και τους υψηλότερους σταθμισμένους μέσους όρους για precision και recall (88.00%), γεγονός που καταδεικνύει την ικανότητά του να αναγνωρίζει με ακρίβεια τόσο τις κυρίαρχες όσο και τις λιγότερο συχνές κλάσεις. Στον μακροσκοπικό μέσο όρο, το *Random Forest* υπερτερεί επίσης, με 67.00% precision και 68.00% recall, επιτυγχάνοντας καλή ισορροπία στις προβλέψεις για όλες τις κλάσεις. Οι αλγόριθμοι *Decision Trees* και *SVM* ακολουθούν με μέση ακρίβεια 90.58% και 87.75%, αντίστοιχα, ενώ το *SVM* εμφανίζει καλύτερο σταθμισμένο precision

(90.00%), γεγονός που το καθιστά κατάλληλο για συγκεκριμένα δεδομένα με υψηλότερες απαιτήσεις ακρίβειας. Ο αλγόριθμος *K-Nearest Neighbors (K-NN)* έχει καλή μέση ακρίβεια (88.96%), ωστόσο η χαμηλότερη απόδοσή του σε recall (75.00%) υποδηλώνει δυσκολίες σε ανισόρροπες κλάσεις. Τέλος, η *Logistic Regression* σημειώνει τη χαμηλότερη απόδοση στους μακροσκοπικούς μέσους όρους (55.00% precision και 69.00% recall), υποδεικνύοντας περιορισμένη ικανότητα πρόβλεψης σε λιγότερο αντιπροσωπευτικές κλάσεις.

Πίνακας 2. Μετρικές απόδοσης των διαφορετικών μοντέλων MM

	LR	SVM	Decision Trees	K-NN	Random Forest
Mean Acc.	87.96%	87.75%	90.58%	88.96%	92.89%
Weight Avg.					
Precision	87.00%	90.00%	85.00%	85.00%	88.00%
Recall	79.00%	81.00%	83.00%	75.00%	88.00%
Macro Avg.					
Precision	55.00%	61.00%	58.00%	52.00%	67.00%
Recall	69.00%	73.00%	63.00%	65.00%	68.00%

Σύμφωνα με τις μετρήσεις του Πίνακα 2, ο *Random Forest* φέρεται να προσφέρει την υψηλότερη μέση ακρίβεια και τους καλύτερους σταθμισμένους και μακροσκοπικούς μέσους όρους Precision και Recall - υποδεικνύοντας την δυνατότητα πρόβλεψης και τον λιγότερα συχνών κλάσεων - καθιστώντας τον το πιο ισορροπημένο και αξιόπιστο αλγόριθμο για το συγκεκριμένο σύνολο δεδομένων.

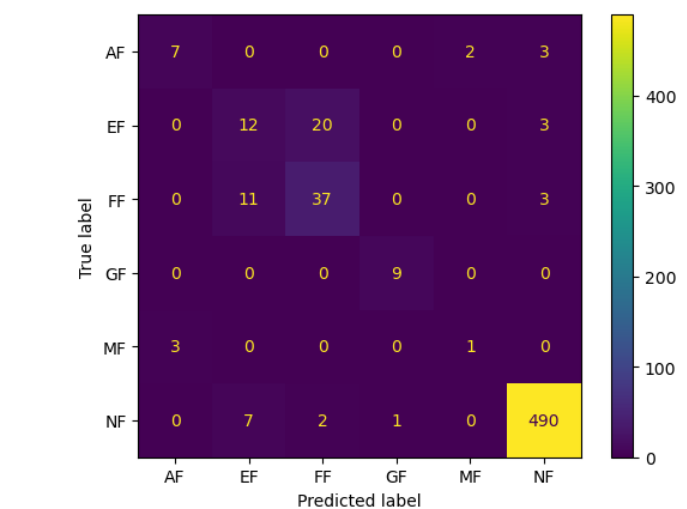
Εφόσον επιλέχθηκε η μέθοδος με την καλύτερη απόδοση, η επόμενη φάση περιλαμβάνει τη βελτιστοποίηση των υπέρ-παραμέτρων (*hyperparameter tuning*) και την αξιολόγηση του εκπαιδευόμενου μοντέλου με βάση την απόδοσή του στο test set τα αποτελέσματα του οποίου παρουσιάζονται στην Εικόνα 24.

	precision	recall	f1-score	support
AF	0.70	0.58	0.64	12
EF	0.40	0.34	0.37	35
FF	0.63	0.73	0.67	51
GF	0.90	1.00	0.95	9
MF	0.33	0.25	0.29	4
NF	0.98	0.98	0.98	500
accuracy			0.91	611
macro avg	0.66	0.65	0.65	611
weighted avg	0.91	0.91	0.91	611

Εικόνα 24. Αποτελέσματα Random Forest στο test set

Αρχικά, από το classification report της παραπάνω εικόνας παρατηρείται η άνιση κατανομή των δειγμάτων κλάση (AF, EF, FF, GF, MF, NF) καθώς και οι υψηλές τιμές *precision*, *recall* και *f1-*

score για την κατηγορία NF. Πιο αναλυτικά, διακρίνεται ότι η κλάση NF περιλαμβάνει το μεγαλύτερο πλήθος δειγμάτων (500) σε αντίθεση με τις υπόλοιπες κλάσεις, οι οποίες έχουν σημαντικά μικρότερο αριθμό δειγμάτων, όπως η AF με 12 και η MF με μόλις 4 δείγματα. Επιπλέον, η κατηγορία NF παρουσιάζει τις υψηλότερες επιδόσεις με *precision* και *recall* στο 0.98, και *f1-score* στο 0.98, γεγονός που υποδηλώνει ότι το μοντέλο μπορεί να αναγνωρίζει σχεδόν όλα τα δείγματα της συγκεκριμένης κλάσης με μεγάλη ακρίβεια, κάτι που πιθανώς οφείλεται και στην κυριαρχία της NF στον αριθμό των παρατηρήσεων. Στην κλάση GF, επίσης το μοντέλο επιτυγχάνει υψηλές επιδόσεις με *precision* 0.90, *recall* 1.00 και *f1-score* 0.95 υποδεικνύοντας καλή απόδοση στην αναγνώριση των συγκεκριμένων σφαλμάτων παρότι περιέχει λιγότερα δείγματα. Αντίθετα, για τις κλάσεις με μικρό αριθμό δειγμάτων, όπως η MF και η EF οι επιδόσεις είναι χαμηλότερες. Η κλάση MF εμφανίζει *precision* 0.33 και *recall* 0.25, με αποτέλεσμα *f1-score* 0.29, που σημαίνει ότι το μοντέλο δυσκολεύεται να αναγνωρίσει σωστά τα δείγματα αυτής της κλάσης. Το ίδιο συμβαίνει και με την EF, όπου το *precision* είναι 0.40 και το *recall* 0.34, αποδίδοντας ένα χαμηλό *f1-score* στο 0.37. Οι χαμηλότερες επιδόσεις στις κατηγορίες με μικρότερο αριθμό δειγμάτων υποδηλώνουν ότι το μοντέλο δεν έχει επιτύχει να μάθει επαρκώς τα χαρακτηριστικά τους, ενδεχομένως λόγω έλλειψης ισορροπίας των δεδομένων. Συνολικά, η ακρίβεια του μοντέλου φτάνει το 91% στο test

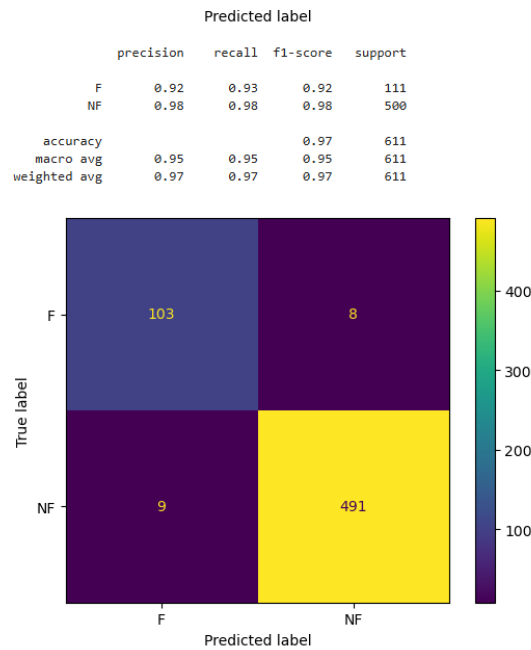


Εικόνα 25. Προβολή Heatmap

set, ενώ το macro average για *precision* και *recall* είναι στο 0.66 και 0.65 αντίστοιχα. Οι σχετικά χαμηλοί μακροσκοπικοί μέσοι όροι επηρεάζονται σημαντικά από την μη επιτυχημένη αναγνώριση των μειοψηφικών κλάσεων, όπως η MF.

Ωστόσο, η προηγούμενη ανάλυση με βάσει το *classification report* αδυνατεί να εντοπίσει ένα σημαντικό χαρακτηριστικό το οποίο είναι ορατό μέσω του *Heatmap* της Εικόνας 25. Πιο συγκεκριμένα, το *Heatmap* αποκαλύπτει τη δυνατότητα του μοντέλου να εντοπίζει τα σφάλματα λειτουργίας ακόμα κι αν δυσκολεύεται να κατηγοριοποιήσει σωστά το είδος τους. Για παράδειγμα, η κλάση MF η οποία περιέχει μόνο τέσσερα (4) δείγματα στο test set, κατηγοριοποιεί εσφαλμένα τρία (3) από τα τέσσερα δείγματα ως σφάλματα AF αλλά αντιλαμβάνεται ότι δεν πρόκειται για φυσιολογική λειτουργία του συστήματος (NF). Ομοίως, η κλάση EF που συγκεντρώνει μόνο 34% στο *recall*, έχει κατηγοριοποιήσει το μεγαλύτερο αριθμό των περιπτώσεων ως σφάλμα FF, ωστόσο το μοντέλο αντιλαμβάνεται ορθώς ότι τα συγκεκριμένα δείγματα δεν ανήκουν σε κατάσταση φυσιολογικής λειτουργίας.

Προς απόδειξη της συγκεκριμένης υπόθεσης, παραλλάσσεται ελαφρώς ο κώδικας του μοντέλου έτσι ώστε να εντοπίζει γενικώς τα σφάλματα έναντι της φυσιολογικής λειτουργίας. Με αυτό τον τρόπο περιέχονται μόνο 2 ξεχωριστές κλάσεις η NF (No Fault) και η F (Fault). Στη συνέχεια, επανεκπαιδεύεται ο αλγόριθμος του *Random Forest* κάτω από τις ίδιες συνθήκες. Το νέο confusion matrix που προκύπτει, σύμφωνα με την Εικόνα 26, αποκαλύπτει ότι το μοντέλο καταφέρνει να αναγνωρίσει τα σφάλματα με αρκετά μεγάλη ακρίβεια, εκ των οποίων 103 σωστές



Εικόνα 26. Προβολή Confusion Matrix και Heatmap 2 κλάσεων

προβλέψεις για την κατηγορία Fault και 491 σωστές προβλέψεις για την κατηγορία No Fault. Παρόλα αυτά, υπάρχουν μερικά λάθη καθώς το μοντέλο ταξινομεί οκτώ (8) δείγματα σφάλματος ως No Fault και εννέα (9) δείγματα φυσιολογικής λειτουργίας ως Fault.

Βάσει το classification report, το μοντέλο επιτυγχάνει συνολική ακρίβεια 97%, γεγονός που υποδεικνύει υψηλή ικανότητα διάκρισης μεταξύ των δύο κλάσεων. Πιο συγκεκριμένα, για την κατηγορία Fault (F), η ακρίβεια (precision) φτάνει στο 92% με ανάκληση (recall) 93%, γεγονός που υποδεικνύει ότι το μοντέλο μπορεί να ανιχνεύσει τα σφάλματα με συνέπεια και να ελαχιστοποιεί τις λανθασμένες κατηγοριοποιήσεις σφάλματος ως φυσιολογική λειτουργία. Η f1-score για την κατηγορία αυτή είναι επίσης υψηλή, στο 92%, κάτι που σημαίνει ότι το μοντέλο διατηρεί ισορροπία μεταξύ ακρίβειας και ανάκλησης για την ανίχνευση σφαλμάτων. Επιπλέον, για την κατηγορία No Fault (NF) τα αποτελέσματα είναι ακόμα καλύτερα, με ακρίβεια 98% και ανάκληση 98% το οποίο υποδεικνύει ότι το μοντέλο αναγνωρίζει τη φυσιολογική λειτουργία σχεδόν χωρίς σφάλματα, επιτυγχάνοντας υψηλό ποσοστό σωστών προβλέψεων όταν δεν υπάρχει πρόβλημα στη λειτουργία. Η f1-score για την κατηγορία αυτή είναι επίσης πολύ υψηλή, στο 98%, υποδεικνύοντας ότι το μοντέλο είναι εξαιρετικά αξιόπιστο στην αναγνώριση των φυσιολογικών περιπτώσεων.

Η παρούσα υπό-ενότητα παρουσίασε τα συνολικά αποτελέσματα των αλγόριθμων Μηχανικής Μάθησης που εξετάστηκαν, για τον εντοπισμό και την ταξινόμηση των δειγμάτων, ακολουθώντας τη μεθοδολογία προ-επεξεργασίας δεδομένων που περιγράφηκε παραπάνω. Τα αποτελέσματα υποδεικνύουν ότι η μείωση των χαρακτηριστικών μέσω της μεθόδου ελέγχου διακύμανσης του

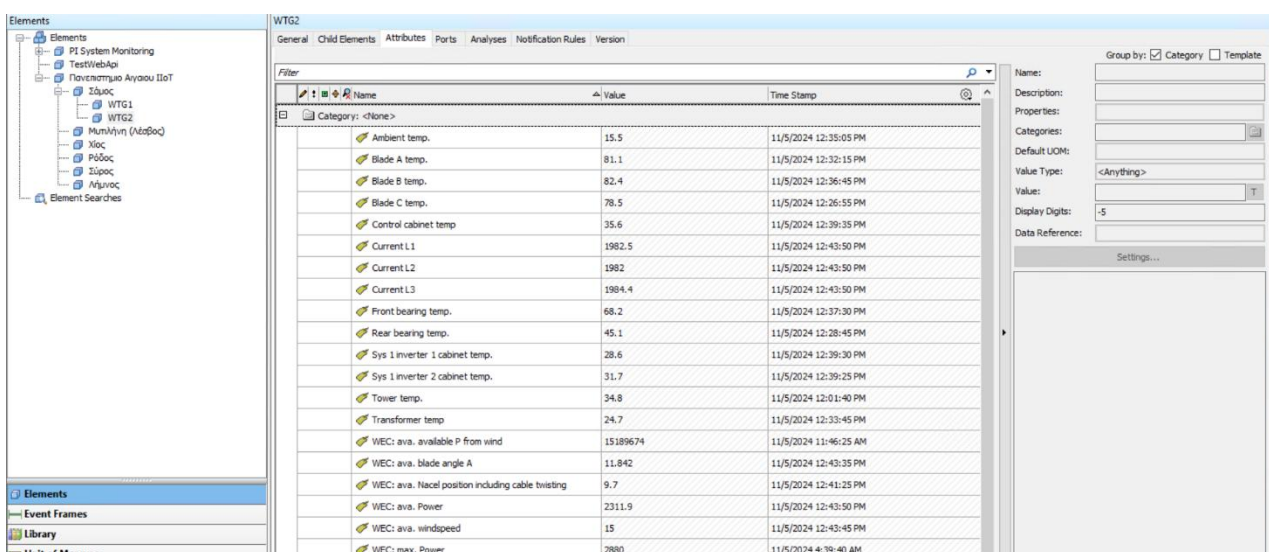
πρώτου βήματος, επαρκεί για την επίτευξη αυξημένης απόδοσης του μοντέλου, επομένως δεν κρίνεται απαραίτητη η χρήση της PCA. Παράλληλα, ο αλγόριθμος Random Forest επιλέχθηκε ως ο πλέον αποδοτικός, επιδεικνύοντας ανώτερη απόδοση τόσο σε επίπεδο ακρίβειας όσο και στις υπόλοιπες μετρικές αξιολόγησης. Ωστόσο, παρατηρήθηκε σχετική αδυναμία του μοντέλου στην ακριβή κατηγοριοποίηση των σφαλμάτων, η οποία μπορεί να αποδοθεί στην έλλειψη επαρκών δειγμάτων για τις μειοψηφικές κλάσεις. Με τη μετατροπή του προβλήματος σε πρόβλημα δύο κατηγοριών (F και NF), το μοντέλο παρουσίασε αυξημένη απόδοση, επιτυγχάνοντας αξιόπιστη διάκριση μεταξύ φυσιολογικής και μη φυσιολογικής λειτουργίας, με ιδιαίτερα χαμηλά ποσοστά σφάλματος και υψηλή γενική ακρίβεια.

5.3 Ροή εργασιών Έξυπνης διαχείρισης Ανεμογεννητριών

Στόχος της παρούσας ενότητας, είναι η παρουσίαση ενός ολοκληρωμένου σχεδίου το οποίο θα ενοποιεί τις τεχνολογίες και τα συστήματα του Βιομηχανικού Διαδικτύου των Πραγμάτων (IIoT) με εφαρμογές που θα περιέχουν τα εκπαιδευμένα μοντέλα Μηχανικής Μάθησης - για την ανίχνευση αστοχιών του εξοπλισμού – σε συνδυασμό με το PI System, το οποίο θα λειτουργεί ως πλατφόρμα για τη συλλογή, αποθήκευση και επεξεργασία δεδομένων σε πραγματικό χρόνο. Το αποτέλεσμα είναι ένα συνολικό σύστημα το οποίο θα είναι σε θέση να αξιοποιεί τις αρχές της έξυπνης βιομηχανίας και θα παρέχει ισχυρά πλεονεκτήματα, όπως ταχύτητα και άμεση αντίδραση σε πιθανά προβλήματα, μείωση κόστους και αύξηση διαθεσιμότητας εξοπλισμού, αυτοματοποίηση και βελτίωση επικοινωνίας, ενίσχυση στη λήψη αποφάσεων, αύξηση ζωής και απόδοσης συστημάτων.

Το PI System προσφέρει ισχυρές δυνατότητες για την παρακολούθηση και τη διαχείριση του εξοπλισμού, όπως οι ανεμογεννήτριες του παραδείγματος, ενώ ταυτόχρονα επιτρέπει τη σύνδεση, συλλογή και αξιοποίηση δεδομένων από ετερογενείς πηγές, όπως συσκευές Βιομηχανικού Διαδικτύου των Πραγμάτων (IIoT), αλλά και εξατομικευμένων εφαρμογών (π.χ. αλγόριθμοι Μηχανικής Μάθησης). Οι βασικές του συνιστώσες περιλαμβάνουν:

PI Data Archive: Αποτελεί το κεντρικό αποθετήριο αποθήκευσης δεδομένων, στο οποίο εισέρχονται δεδομένα από τις συσκευές IIoT σε πραγματικό χρόνο. Κατ' αυτόν τον τρόπο, διασφαλίζεται η άμεση πρόσβαση τόσο σε ιστορικά όσο και σε ζωντανά δεδομένα. Επιπλέον,

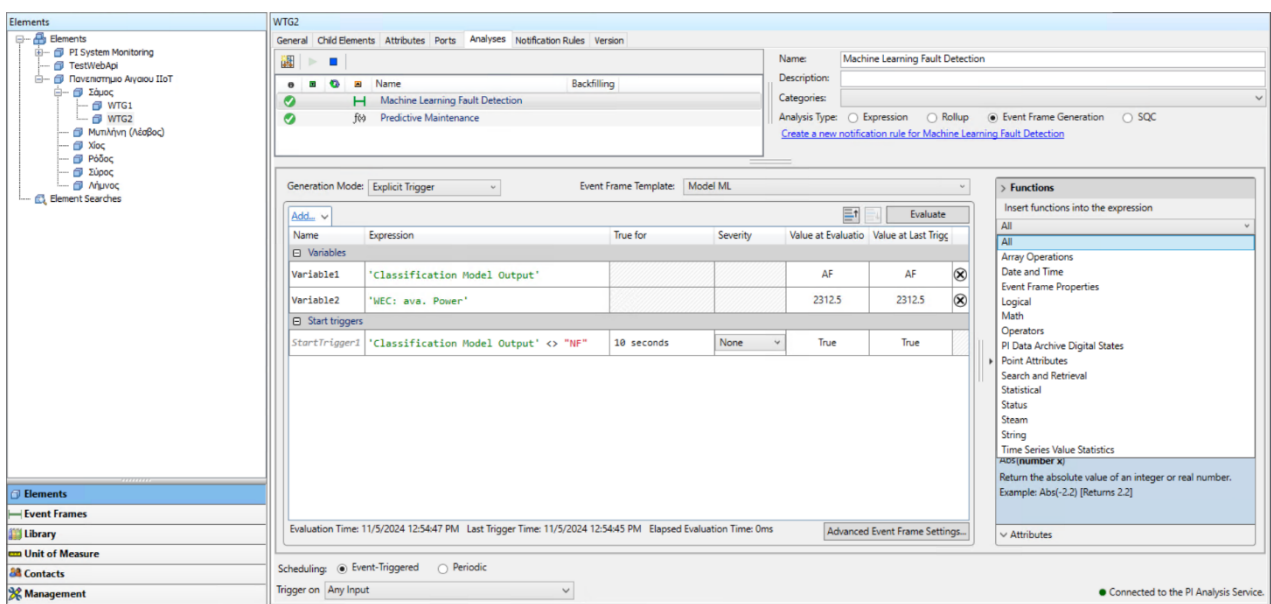


The screenshot displays the PI Data Archive interface for a specific data source labeled 'WTG2'. On the left, a tree view shows the hierarchy of elements, including 'PI System Monitoring', 'TestWebApp', and 'Πανεπιστήμιο Αιγαίου IIoT'. The main area is a table with columns for 'Name', 'Value', and 'Time Stamp'. The table lists various data points such as 'Ambient temp.', 'Blade A temp.', 'Blade B temp.', 'Blade C temp.', 'Control cabinet temp.', 'Current L1', 'Current L2', 'Current L3', 'Front bearing temp.', 'Rear bearing temp.', 'Sys 1 inverter 1 cabinet temp.', 'Sys 1 inverter 2 cabinet temp.', 'Tower temp.', 'Transformer temp.', 'WEC: ava. available P from wind', 'WEC: ava. blade angle A', 'WEC: ava. Nacel position including cable twisting', 'WEC: ava. Power', 'WEC: ava. windspeed', and 'WEC: mav. Drower'. The right-hand panel shows the 'Name' field, 'Description', 'Properties', 'Categories', 'Default UOM', 'Value Type', 'Value', 'Display Digits', and 'Data Reference' fields, along with a 'Settings...' button.

Εικόνα 27. Συγκεντρωτική Δομή ΑΡ

αποθηκεύει δεδομένα που προκύπτουν από την ανάλυση προσαρμοσμένων εφαρμογών που χρησιμοποιούν ως εισόδους τις αρχικές καταχωρήσεις των αισθητήριων, παρέχοντας δυνατότητες ανάλυσης και πρόσβαση σε συμπεράσματα ή παράγωγα δεδομένα που μπορούν να αξιοποιηθούν περαιτέρω, όπως η αξιολόγηση της κατάστασης του εξοπλισμού ή η πρόβλεψη βλαβών.

PI AF (Asset Framework): Το PI AF επιτρέπει την οργάνωση των δεδομένων σε μια δενδροειδή δομή, όπου οι χρήστες μπορούν εύκολα να περιηγηθούν στα διάφορα "elements" - στοιχεία του εξοπλισμού - όπως η ανεμογεννήτρια - και να εντοπίσουν τον συγκεκριμένο εξοπλισμό ή τη συσκευή IIoT που τους ενδιαφέρει. Στην Εικόνα 27, παρουσιάζεται μια υπόθεση εργασίας κατά την οποία το Πανεπιστήμιο Αιγαίου συγκεντρώνει στη δομή του AF όλες τις ανεμογεννήτριες που χρησιμοποιεί για ίδιο-κατανάλωση και είναι διαμοιρασμένες στα διάφορα νησιά. Ακόμα, παρατηρείται ότι κάθε element – ανεμογεννήτρια περιέχει ως *attributes* τις μετρήσεις των συσκευών IIoT με τα οποία εκπαιδεύτηκε το μοντέλο Μηχανικής Μάθησης.



Εικόνα 28. Event Frames για Fault Detection

Event Frames και Notifications: Τα Event Frames του PI AF προσφέρουν τη δυνατότητα παρακολούθησης της κατάστασης των συσκευών και του εξοπλισμού με βάση ορισμένα προκαθορισμένα κριτήρια ή γεγονότα (π.χ. κατάσταση λειτουργίας συστήματος), Εικόνα 28. Με την ενεργοποίηση ενός Event Frame, δημιουργούνται αυτόματα ειδοποιήσεις (notifications) που αποστέλλονται προς ενημέρωση στο τεχνικό προσωπικό μέσω email (Εικόνα 29), ενώ ακόμα προσφέρεται η δυνατότητα σύνδεσης με εταιρικά συστήματα μέσω *web service* όπως το SAP, επιτρέποντας την αυτόματη δημιουργία γνωστοποιήσεων.

Εντοπισμός Σφάλματος Εξοπλισμού WTG2



PI-Notification@
To: giannis.pavlou@ntec.gr

Reply Reply All Forward

Tue 11/5/2024 2:39 PM

EVENT: Machine Learning Fault Detection 2024-11-05 14:37:30.000

NAME: Predictive Maintenance

STATION: \PI-AF\PI System Monitoring\Πανεπιστήμιο Αιγαίου IIoT\Στάμος\WTG2

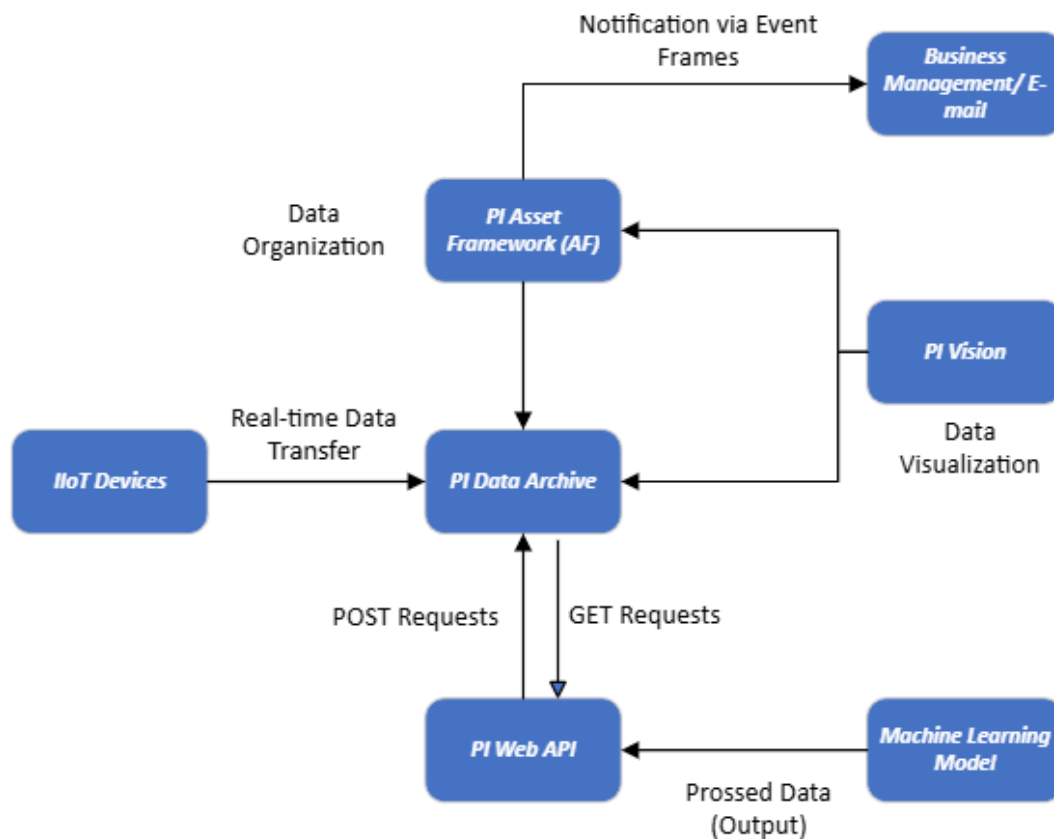
Ημερομηνία: 11/5/2024 2:37:30 PM GTB Standard Time (GMT+02:00:00)

Εντοπισμός σφάλματος στην Ανεμογεννήτρια WTG2 σύμφωνα με το Μοντέλο Μηχανικής Μάθησης. Κατηγοριοποίηση σφάλματος AF.

Επιπλέον πληροφορίες συστήματος από IIoT την στιγμή ενεργοποίησης:

Εικόνα 29. Αυτοματοποιημένο μήνυμα ενημέρωσης σφάλματος εξοπλισμού

PI Vision: Επιπλέον, μέσω του εργαλείου PI Vision παρέχεται η δυνατότητα άμεσης οπτικοποίησης της κατάστασης των ανεμογεννητριών σε όλα τα νησιά, επιτρέποντας στους χρήστες να βλέπουν γραφήματα και άλλα στοιχεία σε πραγματικό χρόνο. Οι απεικονίσεις αυτές διευκολύνουν την αναγνώριση και την κατανόηση κρίσιμων συνθηκών ή αποκλίσεων, διευκολύνοντας τη λήψη ταχύτατων αποφάσεων και συντελώντας στη βελτιστοποίηση της απόδοσης του εξοπλισμού.



Εικόνα 30. Ροή Εργασίας του Συστήματος

Η διασύνδεση των IIoT συσκευών και των μονάδων του PI System με την υποδομή της μηχανικής μάθησης επιτυγχάνεται μέσω του PI Web API, όπως περιγράφεται και από τη ροή εργασιών της Εικόνας 30. Πιο συγκεκριμένα:

GET Requests: Τα GET requests στο PI Web API χρησιμοποιούνται για την άντληση των απαραίτητων δεδομένων από τις συσκευές IIoT μέσω του PI Data Archive, τα οποία εισάγονται ως εισροές στο μοντέλο μηχανικής μάθησης για την ταξινόμηση της κατάστασης της κάθε ανεμογεννήτριας.

POST Requests: Οι αποφάσεις και τα συμπεράσματα του μοντέλου, τα οποία αντιπροσωπεύουν την εκτίμηση της κατάστασης του εξοπλισμού (πχ, *No fault (NF)*, *Air Cooling Fault (AF)*), καταγράφονται στον Data Archive μέσω POST requests. Έτσι, η τρέχουσα κατάσταση του εξοπλισμού ενημερώνεται διαρκώς επιτρέποντας την ενεργοποίηση κατάλληλων *Event Frames* που ενημερώνουν το προσωπικό για ενέργειες επισκευής ή συντήρησης.

5.4 Συμπεράσματα

Στο παρών κεφάλαιο παρουσιάστηκαν τα στάδια της μεθοδολογίας προ-επεξεργασίας δεδομένων καθώς και η διαδικασία εκπαίδευσης και επιλογής του κατάλληλου αλγορίθμου μηχανικής μάθησης για την ανίχνευση και ταξινόμηση των σφαλμάτων του εξοπλισμού. Η μεθοδολογία περιελάμβανε την ανάλυση και προετοιμασία των δεδομένων, με στόχο την εξαγωγή των πιο σημαντικών χαρακτηριστικών, καθώς και τη χρήση τεχνικών για τη μείωση της πολυπλοκότητας του μοντέλου. Επιπλέον, παρουσιάστηκε η αξιολόγηση διαφόρων αποτελεσμάτων και η τελική επιλογή του αλγορίθμου που ανταποκρίθηκε καλύτερα στις απαιτήσεις του συστήματος επιτυγχάνοντας υψηλά επίπεδα ακρίβειας και αξιοπιστίας.

Στη συνέχεια, παρουσιάστηκε ένα υποθετικό σενάριο εφαρμογής για τον έλεγχο και την παρακολούθηση των εγκατεστημένων ανεμογεννητριών, το οποίο βασίζεται στη σύνθεση επιμέρους στοιχείων και τεχνολογιών. Το σενάριο αυτό αναδεικνύει πώς τα συστήματα ΠoT, η πλατφόρμα PI και το μοντέλο μηχανικής μάθησης μπορούν να συνδυαστούν δημιουργώντας ένα ισχυρό εργαλείο που επιτρέπει την αξιόπιστη παρακολούθηση της κατάστασης του εξοπλισμού σε πραγματικό χρόνο, δημιουργώντας κατάλληλα προσαρμοζόμενες ειδοποιήσεις που έχουν ως βάση τις προβλέψεις του μοντέλου.

Ως μελλοντική προοπτική, προτείνεται η επέκταση του συστήματος για τον υπολογισμό του *Remaining Useful Life (RUL)*. Η προσθήκη της εκτίμησης του RUL θα επιτρέψει τον πιο ακριβή προσδιορισμό του χρόνου που απομένει μέχρι την επόμενη προγραμματισμένη συντήρηση ή την πιθανή αστοχία του εξοπλισμού. Με αυτόν τον τρόπο, θα είναι δυνατός ο βέλτιστος σχεδιασμός των συντηρήσεων, η αποφυγή απρόβλεπτων διακοπών λειτουργίας και η μείωση των λειτουργικών εξόδων.

Βιβλιογραφία

- [1] Zheng, H., Paiva, A. R., & Gurciullo, C. S. (2020), Advancing from predictive maintenance to intelligent maintenance with ai and iiot. *arXiv preprint arXiv:2009.00351*.
- [2] Cakir, M., Guvenc, M. A., & Mistikoglu, S. (2021). The experimental application of popular machine learning algorithms on predictive maintenance and the design of IIoT based condition monitoring system. *Computers & Industrial Engineering*, 151, 106948.
- [3] Asadi, M., Hashemi, M., & Balakrishnan, N. (2023). An overview of some classical models and discussion of the signature-based models of preventive maintenance. *Applied Stochastic Models in Business and Industry*, 39(1), 4-53.
- [4] Ingemarsdotter, E., Kambanou, M. L., Jamsin, E., Sakao, T., & Balkenende, R. (2021). Challenges and solutions in condition-based maintenance implementation-A multiple case study. *Journal of Cleaner Production*, 296, 126420.
- [5] Varghese, A., Ande, J. R. P. K., Mahadasa, R., Gutlapalli, S. S., & Surarapu, P. (2023). Investigation of fault diagnosis and prognostics techniques for predictive maintenance in industrial machinery. *Eng. Int*, 11(1), 9-26.
- [6] Sony, M., & Naik, S. (2020). Key ingredients for evaluating Industry 4.0 readiness for organizations: a literature review. *Benchmarking: An International Journal*, 27(7), 2213-2232.
- [7] Çınar, Z. M., Abdussalam Nuhu, A., Zeeshan, Q., Korhan, O., Asmael, M., & Safaei, B. (2020). Machine learning in predictive maintenance towards sustainable smart manufacturing in industry 4.0. *Sustainability*, 12(19), 8211.
- [8] Zonta, T., Da Costa, C. A., da Rosa Righi, R., de Lima, M. J., da Trindade, E. S., & Li, G. P. (2020). Predictive maintenance in the Industry 4.0: A systematic literature review. *Computers & Industrial Engineering*, 150, 106889.
- [9] Bachim, T., Martens, M. L., Digiesi, S., Trindade, D. F., & Ricci, B. (2020). Improving predictive maintenance benefits from online monitoring of spindles: case study in woodworking machine tool. *Revista Gestão & Tecnologia*, 20(4), 7-34.
- [10] Zio, E. (2022). Prognostics and Health Management (PHM): Where are we and where do we (need to) go in theory and practice. *Reliability Engineering & System Safety*, 218, 108119.
- [11] Atamuradov, V., Medjaher, K., Dersin, P., Lamoureux, B., & Zerhouni, N. (2017). Prognostics and health management for maintenance practitioners-Review, implementation and tools evaluation. *International Journal of Prognostics and Health Management*, 8(3), 1-31.
- [12] Hu, Y., Miao, X., Si, Y., Pan, E., & Zio, E. (2022). Prognostics and health management: A review from the perspectives of design, development and decision. *Reliability Engineering & System Safety*, 217, 108063.
- [13] Khaleghi, S., Hosen, M. S., Karimi, D., Behi, H., Beheshti, S. H., Van Mierlo, J., & Bercibar, M. (2022). Developing an online data-driven approach for prognostics and health management of lithium-ion batteries. *Applied Energy*, 308, 118348.

- [14] Xu, X., Tao, Z., Ming, W., An, Q., & Chen, M. (2020). Intelligent monitoring and diagnostics using a novel integrated model based on deep learning and multi-sensor feature fusion. *Measurement*, 165, 108086.
- [15] Cofre-Martel, S., Lopez Droguett, E., & Modarres, M. (2021). Big machinery data preprocessing methodology for data-driven models in prognostics and health management. *Sensors*, 21(20), 6841.
- [16] Souza, M. L. H., da Costa, C. A., & de Oliveira Ramos, G. (2023). A machine-learning based data-oriented pipeline for Prognosis and Health Management Systems. *Computers in Industry*, 148, 103903.
- [17] Bajramovic, E., Gupta, D., Guo, Y., Waedt, K., & Bajramovic, A. (2019). Security challenges and best practices for IIoT.
- [18] Xu, C., Yu, H., Zeng, P., & Li, Y. (2023). Towards critical industrial wireless control: Prototype implementation and experimental evaluation on URLLC. *IEEE Communications Magazine*.
- [19] Khan, R. A. (2018). A survey on wired and wireless network. *Lahore Garrison University Research Journal of Computer Science and Information Technology*, 2(3), 19-28.
- [20] Li, X., Li, D., Wan, J., Vasilakos, A. V., Lai, C. F., & Wang, S. (2017). A review of industrial wireless networks in the context of industry 4.0. *Wireless networks*, 23, 23-41.
- [21] Pereira, C. E., & Neumann, P. (2009). Industrial communication protocols. *Springer Handbook of Automation*, 981-999.
- [22] González, I., Calderón, A. J., Figueiredo, J., & Sousa, J. M. (2019). A literature survey on open platform communications (OPC) applied to advanced industrial environments. *Electronics*, 8(5), 510.
- [23] Madakam, S., & Uchiya, T. (2019). Industrial internet of things (IIoT): principles, processes and protocols. *The Internet of Things in the Industrial Sector: Security and Device Connectivity, Smart Environments, and Industry 4.0*, 35-53.
- [24] Maurya, R., Nambiar, K. A., Babbe, P., Kalokhe, J. P., Ingle, Y. S., & Shaikh, N. F. (2021). Application of restful apis in iot: A review. *Int. J. Res. Appl. Sci. Eng. Technol*, 9, 145-151.
- [25] AVEVA (2021, May) Brochure AVEVA PI SYSTEM
https://www.aveva.com/content/dam/aveva/documents/brochures/Brochure_AVEVA-PI_System-en-LTR-May21-080621v2.pdf
- [26] Jovicic, E., Primorac, D., Cupic, M., & Jovic, A. (2023). Publicly Available Datasets for Predictive Maintenance in the Energy Sector: A Review. *IEEE Access*.
- [27] Sridhar, S., & Sanagavarapu, S. (2021, September). Handling data imbalance in predictive maintenance for machines using SMOTE-based oversampling. In *2021 13th International Conference on Computational Intelligence and Communication Networks (CICN)* (pp. 44-49). IEEE.
- [28] Cofre-Martel, S., Lopez Droguett, E., & Modarres, M. (2021). Big machinery data preprocessing methodology for data-driven models in prognostics and health management. *Sensors*, 21(20), 6841.
- [29] Smiti, A. (2020). A critical overview of outlier detection methods. *Computer Science Review*, 38, 100306.
- [30] Singh, D., & Singh, B. (2020). Investigating the impact of data normalization on classification performance. *Applied Soft Computing*, 97, 105524.
- [31] Greenacre, M., Groenen, P. J., Hastie, T., d'Enza, A. I., Markos, A., & Tuzhilina, E. (2022). Principal component analysis. *Nature Reviews Methods Primers*, 2(1), 100.

- [32] Wang, H., Ni, G., Chen, J., & Qu, J. (2020). Research on rolling bearing state health monitoring and life prediction based on PCA and Internet of things with multi-sensor. *Measurement*, 157, 107657.
- [33] Kalafatelis, A. S., Nomikos, N., Angelopoulos, A., Trochoutsos, C., & Trakadas, P. (2023, January). An Effective Methodology for Imbalanced Data Handling in Predictive Maintenance for Offset Printing. In *International Conference on Mechatronics and Control Engineering* (pp. 89-98). Singapore: Springer Nature Singapore.
- [34] Bisong, E., & Bisong, E. (2019). Logistic regression. *Building machine learning and deep learning models on google cloud platform: A comprehensive guide for beginners*, 243-250.
- [35] Suthaharan, S., & Suthaharan, S. (2016). Support vector machine. *Machine learning models and algorithms for big data classification: thinking with examples for effective learning*, 207-235.
- [36] Wang, Z., & Xue, X. (2014). Multi-class support vector machine. *Support vector machines applications*, 23-48.
- [37] Kingsford, C., & Salzberg, S. L. (2008). What are decision trees?. *Nature biotechnology*, 26(9), 1011-1013.
- [38] Rigatti, S. J. (2017). Random forest. *Journal of Insurance Medicine*, 47(1), 31-39.
- [39] Jadhav, S. D., & Channe, H. P. (2016). Comparative study of K-NN, naive Bayes and decision tree classification techniques. *International Journal of Science and Research (IJSR)*, 5(1), 1842-1845.
- [40] Achouch, M., Dimitrova, M., Ziane, K., Sattarpanah Karganroudi, S., Dhouib, R., Ibrahim, H., & Adda, M. (2022). On predictive maintenance in industry 4.0: Overview, models, and challenges. *Applied Sciences*, 12(16), 8081.
- [41] Haddad, G., Sandborn, P. A., & Pecht, M. G. (2012). An options approach for decision support of systems with prognostic capabilities. *IEEE Transactions on reliability*, 61(4), 872-883.
- [42] Mabrouk, N., Moulahi, M. H., & Ben Hmida, F. (2020). Degradation process analysis and remaining useful life estimation in a control system. In *Proceedings of the 8th International Conference on Sciences of Electronics, Technologies of Information and Telecommunications (SETIT'18), Vol. 2* (pp. 49-65). Springer International Publishing.
- [43] Ok, J. S., Kwon, S. D., Heo, C. E., & Suh, Y. K. (2021). A Survey of Industrial Internet of Things Platforms for Establishing Centralized Data-Acquisition Middleware: Categorization, Experiment, and Challenges. *Scientific Programming*, 2021(1), 6641562.
- [44] OSIsoft Learning, (n.d.), *PI System Basics*, AVEVA Learning Academy, <https://learningacademy.aveva.com/courses/pi-system-basics-1>
- [45] Yang, I. T., & Prayogo, H. (2022). Efficient Reliability Analysis of Structures Using Symbiotic Organisms Search-Based Active Learning Support Vector Machine. *Buildings*, 12(4), 455.
- [46] Sahour, H., Gholami, V., Torkaman, J., Vazifedan, M., & Saeedi, S. (2021). Random forest and extreme gradient boosting algorithms for streamflow modeling using vessel features and tree-rings. *Environmental Earth Sciences*, 80, 1-14.

Παράρτημα Ι – Αποτέλεσμα Μοντέλου Πρόβλεψης

```
NF = [2500]
for i in NF:
    df_nf = df_combine[df_combine.Fault=='NF'].sample(i, random_state=42)
    df_f = df_combine[df_combine.Fault!='NF']
    df_combine_new = pd.concat((df_nf, df_f), axis=0).reset_index(drop=True)
    df_combine_new = df_combine_new.drop(columns=['DateTime_x', 'DateTime_y'])

    Y = df_combine_new['Fault']
    X = df_combine_new.drop(columns=['Fault'])
    # Διαχωρισμός του dataset σε training και test set
    X_train, X_test, y_train, y_test = train_test_split(X, Y, test_size=0.2, stratify=Y, random_state=42)

    # 2. Κανονικοποίηση των δεδομένων
    scaler = StandardScaler()

    X_train_scaled = scaler.fit_transform(X_train)
    X_test_scaled = scaler.transform(X_test)

    # 5. Εφαρμογή του SMOTE για υποστήριξη της Ισορροπίας Κατηγοριών
    smote = SMOTE(random_state=42)
    X_train_resampled, y_train_resampled = smote.fit_resample(X_train_scaled, y_train)

    # 6. Εκπαίδευση ενός Logistic Regression Μοντέλου στο Training Set με PCA
    #model = LogisticRegression(random_state=42, max_iter=3000)
    model = RandomForestClassifier(bootstrap=False, max_depth=30, min_samples_leaf=2, min_samples_split=2, n_estimators=200, random_state=42)
    #model = DecisionTreeClassifier(random_state=42) #ccp_alpha = 0.005
    #model = SVC(random_state=42)
    #model = KNeighborsClassifier()
    model.fit(X_train_resampled, y_train_resampled)

    # 7. Πρόβλεψη στο Test Set και Αξιολόγηση
    y_pred = model.predict(X_test_scaled)
    accuracy = accuracy_score(y_test, y_pred)

    # Δημιουργία του classification report
    report = classification_report(y_test, y_pred)
    print(report)
    ConfusionMatrixDisplay.from_estimator(model, X_test_scaled, y_test, values_format='.5g')
    plt.show()
```

