



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

**ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ
ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ**

ΤΟΜΕΑΣ ΗΛΕΚΤΡΙΚΗΣ ΙΣΧΥΟΣ

Πρόβλεψη Σφαλμάτων σε Ανεμογεννήτριες με την βοήθεια Αλγορίθμου Μηχανικής Μάθησης

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Παπαπαύλου Αλέξανδρος

Επιβλέπων : Άρης-Ευάγγελος Δημέας

Επίκουρος Καθηγητής ΕΜΠ

Αθήνα, Φεβρουάριος, 2025



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

**ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ
ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ**

ΤΟΜΕΑΣ ΗΛΕΚΤΡΙΚΗΣ ΙΣΧΥΟΣ

Πρόβλεψη Σφαλμάτων σε Ανεμογεννήτριες με την βοήθεια Αλγορίθμου Μηχανικής Μάθησης

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Παπαπαύλου Αλέξανδρος

Επιβλέπων : Άρης-Ευάγγελος Δημέας

Επίκουρος Καθηγητής ΕΜΠ

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 25^η Φεβρουαρίου 2025.

Άρης Ευάγγελος
Δημέας
Επίκουρος
Καθηγητής
Ε.Μ.Π.

Γεώργιος Κορρές

Καθηγητής Ε.Μ.Π.

Βασίλειος Νικολαΐδης
Επίκουρος
Καθηγητής
Ε.Μ.Π.

Αθήνα, Φεβρουάριος, 2025



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

**ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ
ΥΠΟΛΟΓΙΣΤΩΝ**

ΤΟΜΕΑΣ ΗΛΕΚΤΡΙΚΗΣ ΙΣΧΥΟΣ

.....
Παπαπαύλου Αλέξανδρος

Διπλωματούχος Ηλεκτρολόγος Μηχανικός και Μηχανικός Υπολογιστών Ε.Μ.Π.

Copyright © Παπαπαύλου Αλέξανδρος, 2025

Με επιφύλαξη παντός δικαιώματος. All rights reserved.

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

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

Περίληψη

Οι ανεμογεννήτριες (Α/Γ) αποτελούν μια από τις σημαντικότερες τεχνολογίες ανανεώσιμων πηγών ενέργειας, συμβάλλοντας σημαντικά στη μείωση των εκπομπών διοξειδίου του άνθρακα (CO₂) και στην ενεργειακή μετάβαση προς πιο βιώσιμες λύσεις. Η συνεχής ανάπτυξη νέων αιολικών πάρκων και η αυξανόμενη διείσδυση της αιολικής ενέργειας στο ενεργειακό μείγμα καθιστούν αναγκαία την όσο το δυνατόν αποδοτικότερη και πιο αξιόπιστη λειτουργία των ανεμογεννητριών. Ωστόσο, η πολυπλοκότητα των μηχανικών και ηλεκτρονικών τους συστημάτων και οι συνθήκες στις οποίες απαιτείται να λειτουργούν, τις καθιστούν επιρρεπείς σε σφάλματα και βλάβες, γεγονός που απαιτεί την ανάπτυξη στρατηγικών παρακολούθησης και συντήρησης για την διασφάλιση υψηλής αποδοτικότητας και αξιοπιστίας.

Μια από τις πιο προηγμένες προσεγγίσεις στη διαχείριση της συντήρησης των ανεμογεννητριών είναι η προγνωστική συντήρηση (Predictive Maintenance - PdM), η οποία βασίζεται στην ανάλυση και επεξεργασία δεδομένων πραγματικού χρόνου για την έγκαιρη ανίχνευση σφαλμάτων και την πρόβλεψη πιθανών βλαβών.

Τα συστήματα SCADA (Supervisory Control and Data Acquisition) διαδραματίζουν καθοριστικό ρόλο σε αυτή τη διαδικασία, καθώς επιτρέπουν τη συνεχή συλλογή κρίσιμων λειτουργικών δεδομένων, όπως θερμοκρασίες, ταχύτητες περιστροφής και παραγόμενη ισχύς. Αυτή η συνεχής ροή δεδομένων επιτρέπει τη δημιουργία προγνωστικών μοντέλων συντήρησης, τα οποία μπορούν να συμβάλουν στη βελτίωση της διαθεσιμότητας των ανεμογεννητριών, τη μείωση του κόστους συντήρησης και την ενίσχυση της συνολικής αξιοπιστίας του αιολικού πάρκου.

Η παρούσα διπλωματική εργασία επικεντρώνεται στην ανάπτυξη και αξιολόγηση μιας μεθοδολογίας για την πρόβλεψη σφαλμάτων σε ανεμογεννήτριες, αξιοποιώντας δεδομένα SCADA και αλγορίθμους μηχανικής μάθησης. Ιδιαίτερη έμφαση δίνεται στην εφαρμογή του XGBoost για την εκτίμηση κρίσιμων λειτουργικών παραμέτρων και στη χρήση διαγραμμάτων ελέγχου διεργασιών (SPC) για την ανίχνευση αποκλίσεων που ενδέχεται να υποδηλώνουν δυσλειτουργίες. Η συνδυαστική αυτή προσέγγιση αναμένεται να ενισχύσει την ακρίβεια των προβλέψεων και να συμβάλει στη βελτιστοποίηση της συντήρησης, βελτιώνοντας τη λειτουργική απόδοση και τη βιωσιμότητα της αιολικής ενέργειας.

Λέξεις-Κλειδιά: Αιολική Ενέργεια, Ανεμογεννήτριες, Πρόβλεψη Σφαλμάτων, συστήματα SCADA, XGBoost, Μηχανική Μάθηση, Θερμοκρασία Κιβωτίου Ταχυτήτων, Θερμοκρασία Γεννήτριας, Προγνωστική Συντήρηση, Ανίχνευση Ανωμαλιών, Διαγράμματα SPC.

Abstract

Wind turbines (WTs) are among the most significant renewable energy technologies, playing a crucial role in reducing carbon dioxide (CO₂) emissions and facilitating the transition to more sustainable energy solutions. The continuous development of new wind farms and the increasing penetration of wind energy into the energy mix necessitate the most efficient and reliable operation of wind turbines. However, the complexity of their mechanical and electronic systems, combined with the challenging conditions in which they operate, makes them susceptible to faults and failures. This highlights the need for advanced monitoring and maintenance strategies to ensure high efficiency and reliability.

One of the most advanced approaches to wind turbine maintenance management is Predictive Maintenance (PdM), which relies on real-time data analysis to detect faults early and predict potential failures. SCADA (Supervisory Control and Data Acquisition) systems play a key role in this process, enabling the continuous collection of critical operational data such as temperatures, rotational speeds, and generated power. This constant data stream facilitates the development of predictive maintenance models, which can enhance wind turbine availability, reduce maintenance costs, and improve the overall reliability of wind farms.

This thesis focuses on the development and evaluation of a methodology for fault prediction in wind turbines, leveraging SCADA data and machine learning algorithms. Special emphasis is placed on the application of XGBoost for estimating critical operational parameters and the use of Statistical Process Control (SPC) charts to detect deviations that may indicate malfunctions. This combined approach is expected to improve prediction accuracy and contribute to maintenance optimization, enhancing both operational performance and the sustainability of wind energy.

Keywords: Wind Energy, Wind Turbines, Fault Prediction, SCADA Systems, XGBoost, Machine Learning, Gearbox Temperature, Generator Temperature, Predictive Maintenance, Anomaly Detection, SPC Charts.

Ευχαριστίες

Η παρούσα διπλωματική εργασία «Πρόβλεψη Σφαλμάτων σε Ανεμογεννήτριες με την βοήθεια Αλγορίθμου Μηχανικής Μάθησης» εκπονήθηκε στα πλαίσια των προπτυχιακών μου σπουδών στη Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών του Εθνικού Μετσόβιου Πολυτεχνείου.

Θα ήθελα να ευχαριστήσω τον επιβλέποντα της διπλωματικής μου εργασίας, επίκουρο καθηγητή κ. Άρη – Ευαγγέλου Δημέα για την ευκαιρία που μου έδωσε με την ανάθεση της διπλωματικής εργασίας να ασχοληθώ με ένα τόσο ενδιαφέρον θέμα μελέτης, ιδιαίτερα επίκαιρο με την όλο αυξανόμενη διείσδυση των ανεμογεννητριών στην ηλεκτροπαραγωγή.

Επίσης, θα ήθελα να ευχαριστήσω θερμά τον κ. Δημήτρη Λαγό, για την συνεχή καθοδήγηση και παρότρυνση του σε όλη την διάρκεια προετοιμασίας της διπλωματικής εργασίας μου. Η βοήθεια του ήταν πολύτιμη και καταλυτική προκειμένου να προχωρήσει και να περατωθεί η εργασία αυτή.

Ευχαριστίες οφείλω να δώσω και στους κους Ε. Δουκάκη και Α. Κασίμη από ΔΕΗΑν, που μου παρείχαν πραγματικά μετρητικά στοιχεία απαραίτητα για να καταστεί δυνατή η ανάπτυξη των υπολογιστικών μοντέλων της διπλωματικής μου εργασίας.

Τέλος, θα ήθελα να εκφράσω την ευγνωμοσύνη μου στην οικογένεια μου, για την αμέριστη συμπαράσταση τους κατά τη διάρκεια των σπουδών μου.

Περιεχόμενα

ΚΕΦΑΛΑΙΟ 1 : Εισαγωγή	17
1.1 Αιολική Ενέργεια και Ανάγκη Συντήρησης	17
1.2 Σκοπός της Διπλωματικής.....	18
ΚΕΦΑΛΑΙΟ 2 : Ανεμογεννήτριες	19
2.1 Οι Ανεμογεννήτριες και η Λειτουργία τους	19
2.1.1 Γενικά	19
2.1.2 Βασικά εξαρτήματα ανεμογεννήτριας.....	22
2.1.3 Βασικά Οφέλη από την Λειτουργία των Ανεμογεννητριών	25
2.2 Συνηθέστερα Σφάλματα Ανεμογεννητριών	25
2.3 Συντήρηση Ανεμογεννητριών	26
Κεφαλαίο 3 : Παρακολούθηση και Βελτιστοποίηση της Λειτουργίας των Ανεμογεννητριών	28
3.1 Συστήματα SCADA	28
3.1.1 Στοιχεία ενός Συστήματος SCADA & Αρχιτεκτονική SCADA	29
3.1.2 Βασικές Λειτουργίες των SCADA.....	30
3.2 Προγνωστική Συντήρηση - Predictive Maintenance (PdM)	31
3.2.1 Ιστορική Εξέλιξη των Στρατηγικών Συντήρησης	31
3.2.2 Τρόπο Λειτουργίας της PdM.....	32
3.3 Προκλήσεις της Παρακολούθησης των Ανεμογεννητριών	33
3.4 Μελλοντικές Τεχνολογικές Τάσεις στην Παρακολούθηση και Συντήρηση των Ανεμογεννητριών	34
3.5 Οικονομικά και Περιβαλλοντικά Οφέλη της Προγνωστικής Συντήρησης	35
Κεφάλαιο 4 : Μηχανική Μάθηση και ο Αλγόριθμος XGBoost	37
4.1 Εισαγωγή στην Μηχανική Μάθηση	37
4.1.1 Προγνωστική Συντήρηση και Μηχανική Μάθηση.....	38
4.2 Ο Αλγόριθμος XGBoost	38
4.2.1 Αρχές Λειτουργίας του XGBoost	39
4.2.2 Ρύθμιση Υπερπαραμέτρων του XGBoost	40
4.2.3 Εφαρμογή του XGBoost σε δεδομένα SCADA	42
4.3 Στατιστική Αξιολόγηση Μοντέλων	43

4.3.1 Στατιστικοί Δείκτες.....	44
4.4 Διαγράμματα SPC	45
4.4.1 Θεωρητική Προσέγγιση των SPC	46
4.4.2 Κατασκευή SPC Διαγραμμάτων	46
4.4.3 Συνδυασμός των Στατιστικών Δεικτών με τα SPC Διαγράμματα	47
ΚΕΦΑΛΑΙΟ 5 : Μεθοδολογία	49
5.1 Εισαγωγή στην Μεθοδολογία	49
5.2 Απόκτηση Δεδομένων και Προεπεξεργασία	50
5.2.1 Συλλογή και Επιλογή Δεδομένων SCADA	50
5.2.2 Προεπεξεργασία και Καθαρισμός Δεδομένων	51
5.3 Επεξεργασία Μοντέλων και Εκπαίδευση	54
5.3.1 Δεδομένα Εισόδου και Εξόδου	55
5.3.2 Ανάπτυξη και Εκπαίδευση Μοντέλου XGBoost.....	58
5.4 Μεταεπεξεργασία	59
5.4.1 Αξιολόγηση Μοντέλου με Στατιστικούς Δείκτες	59
5.4.2 Ανίχνευση Ανωμαλιών (Anomaly Detection)	59
ΚΕΦΑΛΑΙΟ 6 : Αποτελέσματα και Συμπεράσματα	61
6.1 Ανάλυση Αποτελεσμάτων για Ανεμογεννήτριες χωρίς Ιστορικό Βλαβών ...	61
6.1.1 Εξέλιξη του RMSE κατά τη Διαδικασία Εκπαίδευσης του XGBoost.....	62
6.1.2 Στατιστικοί Δείκτες στο Σύνολο Εκπαίδευσης	64
6.1.3 Αξιολόγηση στο Σύνολο Δοκιμής (testing set).....	65
6.1.4 Διαγράμματα SPC για τον εντοπισμό πιθανών βλαβών	68
6.2 Ανάλυση Αποτελεσμάτων για Ανεμογεννήτριες με Ιστορικό Βλαβών.....	74
6.2.1 Βλάβη στο Κιβώτιο Ταχυτήτων	74
6.2.2 Βλάβη στην Γεννήτρια	78
6.3 Περιορισμοί και Μελλοντικές Βελτιώσεις	81
6.4 Συνολικά Συμπεράσματα.....	83
Παράρτημα : Περιγραφή των Βασικών Τμημάτων του Κώδικα	85
Βιβλιογραφικές Αναφορές – Πηγές.....	90

Ευρετήριο Εικόνων

Εικόνα 1.Εξέλιξη ανεμογεννητριών με την πάροδο του χρόνου.....	20
Εικόνα 2.Τυπική Καμπύλη Ισχύος Ανεμογεννήτριας.....	21
Εικόνα 3.Χερσαίο Αιολικό Παρκο.....	22
Εικόνα 4.Υπεράκτιο Αιολικό Πάρκο.....	22
Εικόνα 5.Κύρια Εξαρτήματα της Ανεμογεννήτριας.....	24
Εικόνα 6.Βασικά εξαρτήματα των Ανεμογεννητριών και Σύνδεση στο δίκτυο.....	25
Εικόνα 7.Αρχιτεκτονική Συστήματος SCADA.....	30
Εικόνα 8.Ροή δεδομένων SCADA στις Ανεμογεννήτριες.....	31
Εικόνα 9.Παρακολούθηση Ανεμογεννητριών.....	31
Εικόνα 10.Εξέλιξη Στρατηγικών Συντήρησης.....	32
Εικόνα 11.Ροή Δεδομένων σε Cloud Πλατφόρμα για Predictive Maintenance.....	33
Εικόνα 12.Ροή της Μεθοδολογίας.....	38
Εικόνα 13.Τρόπος Λειτουργίας του XGBoost.....	40
Εικόνα 14.Βασικές υπερπαράμετροι του XGBoost που χρησιμοποιούνται στην υλοποίηση του κώδικα.....	42
Εικόνα 15.Τυπικό Διάγραμμα SPC.....	46
Εικόνα 16.Διαδικασία Ανίχνευσης Ανωμαλιών.....	50
Εικόνα 17.Καμπύλη Ισχύος πριν τον καθαρισμό δεδομένων.....	54
Εικόνα 18.Καμπύλη Ισχύος μετά την εφαρμογή των quantiles.....	54
Εικόνα 19.Διάγραμμα Συσχέτισης του Κιβωτίου Ταχυτήτων (Gearbox).....	57
Εικόνα 20.Διάγραμμα Συσχέτισης της Γεννήτριας (Generator).....	57
Εικόνα 21.Σύγκλιση του RMSE εκπαίδευσης και επικύρωσης για την θερμοκρασία της γεννήτριας (RG01-RG04).....	63
Εικόνα 22.Σύγκλιση του RMSE εκπαίδευσης και επικύρωσης για την θερμοκρασία του κιβωτίου ταχυτήτων (RG01-RG04).....	63
Εικόνα 23.Διαγράμματα Πραγματικών-Προβλεπόμενων Τιμών για την θερμοκρασία της γεννήτριας (RG01-RG04).....	67
Εικόνα 24.Διαγράμματα Πραγματικών-Προβλεπόμενων Τιμών για την θερμοκρασία του κιβωτίου ταχυτήτων (RG01-RG04).....	68
Εικόνα 25.SPC Διαγράμματα της γεννήτριας (RG01-RG04).....	71
Εικόνα 26.SPC Διαγράμματα για του κιβώτιο ταχυτήτων(RG01-RG04).....	72
Εικόνα 27.Εξέλιξη του RMSE για την T06.....	75
Εικόνα 28.Διάγραμμα Πραγματικών και Προβλεπόμενων τιμών για την θερμοκρασία του Gearbox της T06.....	76
Εικόνα 29.Διάγραμμα SPC για την θερμοκρασία του κιβωτίου ταχυτήτων T06.....	77
Εικόνα 30.Διαγράμματα SPC για τη Θερμοκρασία Κιβωτίου Ταχυτήτων της T06 (Αύγουστος - Σεπτέμβριος 2017).....	78
Εικόνα 31.Εξέλιξη του RMSE για την T07.....	79
Εικόνα 32.Διάγραμμα Πραγματικών και Προβλεπόμενων τιμών για την θερμοκρασία της Generator της T07.....	79

Εικόνα 33.Διάγραμμα SPC για την θερμοκρασία της γεννήτριας T07.....	80
Εικόνα 34.Διαγράμματα SPC για τη θερμοκρασία της γεννήτριας της T07 (Ιούλιος - Αύγουστος 2017)	81

Ευρετήριο Πινάκων

Πίνακας 1.Μεταβλητές Εισόδου και Έξοδου για το Μοντέλο XGBoost.....	55
Πίνακας 2.Κανονικοποιημένες Τιμές RMSE και Τελευταίο Boosting Round για Κάθε Ανεμογεννήτρια	63
Πίνακας 3.Κανονικοποιημένοι Στατιστικοί Δείκτες Απόδοσης του Μοντέλου XGBoost για το Σύνολο Εκπαίδευσης	65
Πίνακας 4.Κανονικοποιημένοι Στατιστικοί Δείκτες Απόδοσης του Μοντέλου XGBoost για το Σύνολο Δοκιμής	66
Πίνακας 5.Διαδοχικές Αποκλίσεις	73
Πίνακας 6.Κανονικοποιημένοι Στατιστικοί Δείκτες για την Ανεμογεννήτρια T06 με Ιστορικό Βλαβών	76
Πίνακας 7.Κανονικοποιημένοι Στατιστικοί Δείκτες για την Ανεμογεννήτρια T07 με Ιστορικό Βλαβών	79

ΚΕΦΑΛΑΙΟ 1 : Εισαγωγή

1.1 Αιολική Ενέργεια και Ανάγκη Συντήρησης

Η αιολική ενέργεια έχει αναδειχθεί τα τελευταία χρόνια ως μία από τις κύριες πηγές ανανεώσιμης ενέργειας, κυρίως λόγω των περιβαλλοντικών οφελών της και της δυνατότητάς της να συμβάλλει στη μείωση των εκπομπών CO₂, υποστηρίζοντας έτσι την καταπολέμηση της κλιματικής αλλαγής και την επίτευξη των παγκόσμιων στόχων ενεργειακής μετάβασης. Επιπλέον, αποτελεί μια ανεξάντλητη πηγή ενέργειας, δωρεάν και διαθέσιμη σε αφθονία, καθιστώντας τα έργα αξιοποίησής της οικονομικά βιώσιμα.

Σύμφωνα με τον Global Wind Energy Council (GWEC), η παγκόσμια αιολική ισχύς αναμένεται να φτάσει τα 130 GW το 2024, ξεπερνώντας το προηγούμενο ρεκόρ. Εκτιμάται επίσης ότι τα επόμενα πέντε χρόνια θα προστεθούν 791 GW νέας ισχύος, με μέσο ετήσιο ρυθμό εγκατάστασης 158 GW μέχρι το 2028 [1]. Έτσι, λόγω της αφθονίας, της βιωσιμότητας και του ελάχιστου περιβαλλοντικού αντίκτυπου, η αιολική ενέργεια αποτελεί έναν από τους βασικούς πυλώνες της μετάβασης σε ένα σύστημα χαμηλών εκπομπών CO₂.

Ωστόσο, η διαρκής αύξηση της εγκατεστημένης αιολικής ισχύος και η λειτουργία ολοένα και πιο σύνθετων ανεμογεννητριών καθιστούν την αποδοτική λειτουργία και συντήρηση των αιολικών πάρκων, κρίσιμη για τη διατήρηση της βιωσιμότητάς τους. Οι ανεμογεννήτριες περιλαμβάνουν σύνθετα μηχανικά και ηλεκτρικά συστήματα που υποβάλλονται σε διαρκείς καταπονήσεις λόγω ακραίων και μεταβαλλόμενων περιβαλλοντικών συνθηκών, φυσικής φθοράς και μηχανικών αστοχιών [2]. Σύμφωνα με έκθεση της IRENA, το κόστος λειτουργίας και συντήρησης (O&M) των υπεράκτιων αιολικών πάρκων στις χώρες της G20 κυμαίνεται μεταξύ 16-25% του συνολικού κόστους ηλεκτρικής ενέργειας, γεγονός που υπογραμμίζει την ανάγκη για αποτελεσματικές στρατηγικές συντήρησης.

Η ανάγκη για έγκαιρη ανίχνευση βλαβών και εφαρμογή προγνωστικής συντήρησης είναι ζωτικής σημασίας, καθώς συμβάλλει στη μείωση του κόστους επισκευών και προμηθειών ανταλλακτικών, ενώ παράλληλα αυξάνει τη διάρκεια ζωής των εξαρτημάτων. Η προγνωστική συντήρηση των ανεμογεννητριών μπορεί να επιτευχθεί μέσω της ανάλυσης δεδομένων που συλλέγονται από αισθητήρες ήδη εγκατεστημένους στις ανεμογεννήτριες, οι οποίοι αποτελούν μέρος του συστήματος SCADA. Η εφαρμογή μοντέλων μηχανικής μάθησης και τεχνικών στατιστικής ανάλυσης στα δεδομένα αυτά, μπορούν να εντοπίσουν αποκλίσεις από τη φυσιολογική λειτουργία των κρίσιμων εξαρτημάτων, που πιθανόν να υποδεικνύουν βλάβη. Με τη χρήση αυτών των τεχνικών για την έγκαιρη ανίχνευση βλαβών,

επιτυγχάνεται επίσης η μείωση των απρόβλεπτων διακοπών λειτουργίας, περιορίζοντας τις οικονομικές απώλειες και τις επιπτώσεις στην ευστάθεια του ηλεκτρικού συστήματος, που μπορεί να προκληθούν από εκτεταμένες βλάβες σε ολόκληρο το αιολικό πάρκο [3,4,5].

1.2 Σκοπός της Διπλωματικής

Σκοπός της παρούσας διπλωματικής εργασίας είναι η ανάπτυξη ενός μοντέλου πρόβλεψης σφαλμάτων σε ανεμογεννήτριες, βασισμένου σε δεδομένα SCADA και αλγορίθμους μηχανικής μάθησης. Μέσω της συνεχούς παρακολούθησης κρίσιμων λειτουργικών παραμέτρων και της ανάλυσης ιστορικών δεδομένων, επιδιώκεται η έγκαιρη ανίχνευση πιθανών αστοχιών, ώστε να βελτιστοποιηθεί η λειτουργία και η συντήρηση των ανεμογεννητριών και να μειωθεί το κόστος επισκευών.

Συγκεκριμένα, η εργασία εστιάζει στην εφαρμογή του αλγορίθμου XGBoost, ενός ισχυρού αλγορίθμου μηχανικής μάθησης, για την πρόβλεψη των θερμοκρασιών σε κρίσιμα εξαρτήματα, όπως η γεννήτρια και το κιβώτιο ταχυτήτων. Η ανάλυση επικεντρώνεται στον εντοπισμό αποκλίσεων μεταξύ των πραγματικών και των προβλεπόμενων τιμών θερμοκρασίας, καθώς σημαντικές αποκλίσεις ενδέχεται να υποδεικνύουν πιθανές βλάβες ή πρώιμα σημάδια δυσλειτουργίας.

Η ανάπτυξη ενός τέτοιου προβλεπτικού μοντέλου έχει ως απώτερο στόχο να βελτιώσει τις στρατηγικές προβλεπτικής συντήρησης (Predictive Maintenance - PdM), να μειώσει τον χρόνο μη διαθεσιμότητας των ανεμογεννητριών και να συμβάλει στη βιώσιμη και αποδοτική λειτουργία των αιολικών πάρκων.

ΚΕΦΑΛΑΙΟ 2 : Ανεμογεννήτριες

Στο παρόν κεφάλαιο, γίνεται μία σύντομη αναφορά στον τρόπο λειτουργίας των ανεμογεννητριών και παρουσιάζονται τα βασικά μέρη και εξαρτήματα τους. Ακολούθως, γίνεται αναφορά στα βασικά σφάλματα που μπορεί να εμφανιστούν κατά την λειτουργία των ανεμογεννητριών.

2.1 Οι Ανεμογεννήτριες και η Λειτουργία τους

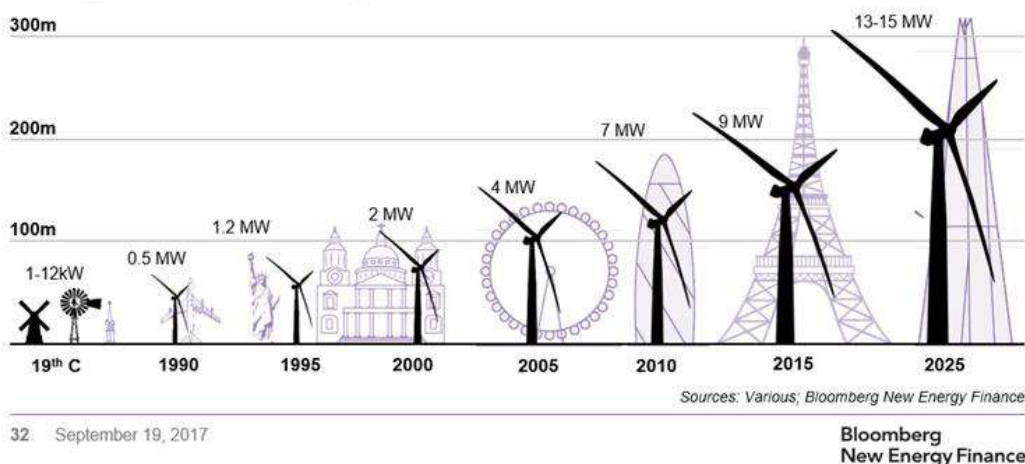
2.1.1 Γενικά

Οι ανεμογεννήτριες αποτελούν μία από τις πιο αποδοτικές τεχνολογίες αξιοποίησης ανανεώσιμης ενέργειας, μετατρέποντας την κινητική ενέργεια του ανέμου σε ηλεκτρική ενέργεια. Η παραγόμενη αιολική ενέργεια είναι συνεχώς αυξανόμενη παγκοσμίως και στην Ελλάδα στα τέλη του 2024 η εγκατεστημένη αιολική ισχύς ανέρχεται σε περίπου 5,355 MW, με την ετήσια αύξηση σε σχέση με το 2023 να σημειώνεται στο 2,4%. Η ανάπτυξη αυτή αντιστοιχεί σε επενδύσεις συνολικού ύψους 150 εκατ. Ευρώ [7].

Η λειτουργία των ανεμογεννητριών βασίζεται στην αεροδυναμική. Ο άνεμος κινεί τα πτερύγια, δημιουργώντας διαφορά πίεσης, που τα κάνει να περιστρέφονται. Αυτή η κίνηση μετατρέπει την ενέργεια του ανέμου σε μηχανική ενέργεια. Ο ρότορας της ανεμογεννήτριας μεταφέρει την ενέργεια αυτή στη γεννήτρια, είτε απευθείας είτε μέσω κιβωτίου ταχυτήτων που αυξάνει την ταχύτητα περιστροφής. Η γεννήτρια, στη συνέχεια, μετατρέπει τη μηχανική ενέργεια σε ηλεκτρική [8,11]. Αυτή η ηλεκτρική ενέργεια περνά από μετασχηματιστές για να προσαρμοστεί στην απαιτούμενη τάση πριν διοχετευτεί στο δίκτυο. Εναλλακτικά, μπορεί να αποθηκευτεί σε μπαταρίες για μελλοντική χρήση.

Η συνεχής τεχνολογικές εξελίξεις έχουν βελτιώσει την απόδοσή και την αξιοπιστία των ανεμογεννητριών, καθιστώντας τις ικανές να λειτουργούν αποτελεσματικά σε ποικίλες περιβαλλοντικές συνθήκες. Οι ανεμογεννήτριες μπορούν να παράγουν από μερικές δεκάδες ή εκατοντάδες kW έως σήμερα 15-16 MW. Στην παρακάτω εικόνα (Εικόνα 1), φαίνεται η εξέλιξη των ανεμογεννητριών με το χρόνο, τόσο όσον αφορά την παραγόμενη ηλεκτρική ισχύ από αυτές όσο και το απαιτούμενο μέγεθός τους.

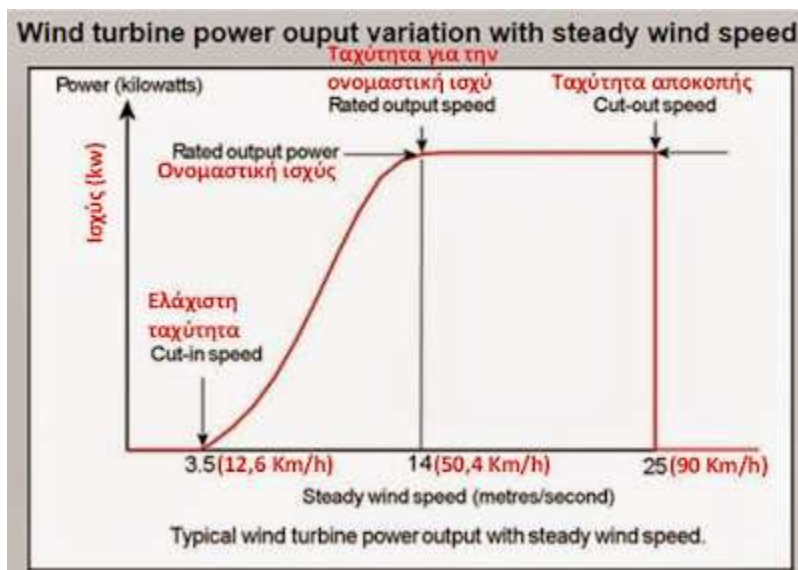
Evolution of wind turbine heights and output



Εικόνα 1. Εξέλιξη ανεμογεννητριών με την πάροδο του χρόνου.

Η τυπική καμπύλη ισχύος μιας ανεμογεννήτριας περιγράφει τη σχέση μεταξύ της ταχύτητας ανέμου (m/s) και της παραγόμενης ισχύος (kW) σε τρεις κύριες περιοχές λειτουργίας. Στην περιοχή εκκίνησης, η παραγωγή ενέργειας ξεκινά όταν η ταχύτητα ανέμου υπερβεί τα 3,5 m/s (Cut-in speed). Στην περιοχή ονομαστικής ισχύος, η ισχύς αυξάνεται γραμμικά έως το μέγιστο επίπεδο απόδοσης (Rated power) στα 14 m/s. Στην περιοχή σταθερής ισχύος και διακοπής, η ισχύς παραμένει σταθερή έως τα 25 m/s (Cut-out speed), οπότε η λειτουργία διακόπτεται για λόγους ασφαλείας [6].

Η παραγόμενη ισχύς εξαρτάται άμεσα από την ταχύτητα του ανέμου, και γι' αυτό τα πτερύγια της ανεμογεννήτριας τοποθετούνται σε μεγάλα ύψη, όπου οι ταχύτητες του ανέμου είναι υψηλότερες. Η καμπύλη αυτή αποτελεί σημείο αναφοράς για την παρακολούθηση της απόδοσης και την ανίχνευση αποκλίσεων, συμβάλλοντας στη βελτιστοποίηση της λειτουργίας της ανεμογεννήτριας.



Εικόνα 2.Τυπική Καμπύλη Ισχύος Ανεμογεννήτριας

Η επιλογή της τοποθεσίας εγκατάστασης των ανεμογεννητριών παίζει καθοριστικό ρόλο στη συνολική τους απόδοση και τη βιωσιμότητα του αιολικού πάρκου. Οι ανεμογεννήτριες μπορούν να τοποθετηθούν είτε σε χερσαίες εκτάσεις (Εικόνα 3) είτε σε θαλάσσιες περιοχές (Εικόνα 4), συνήθως σε ομάδες, σχηματίζοντας αιολικά πάρκα με εγκατεστημένη ισχύ που μπορεί να φτάνει δεκάδες ή εκατοντάδες MW.

Η επιλογή παράκτιας περιοχής προτιμάται λόγω του υψηλότερου και σταθερότερου αιολικού δυναμικού, καθώς και της μικρότερης όχλησης που προκαλείται στην τοπική κοινότητα. Στην περίπτωση των υπεράκτιων ανεμογεννητριών, η εγκατάσταση γίνεται με κατάλληλη πάκτωση στον βυθό της θάλασσας. Επιπλέον, τα τελευταία χρόνια έχουν αναπτυχθεί τεχνολογίες που επιτρέπουν τη στήριξή τους σε πλωτές εξέδρες, διευρύνοντας τις δυνατότητες αξιοποίησης της αιολικής ενέργειας σε μεγαλύτερα βάθη.



Εικόνα 3.Χερσαίο Αιολικό Πάρκο



Εικόνα 4.Υπεράκτιο Αιολικό Πάρκο

2.1.2 Βασικά εξαρτήματα ανεμογεννήτριας.

Η βασική δομή μιας ανεμογεννήτριας αποτελείται από διάφορα κρίσιμα εξαρτήματα, τα οποία συνεργάζονται για τη συλλογή και τη μετατροπή της αιολικής ενέργειας σε ηλεκτρική και τέλος την διανομής της στο δίκτυο [9,10].

Τα σημαντικότερα από αυτά είναι τα παρακάτω:

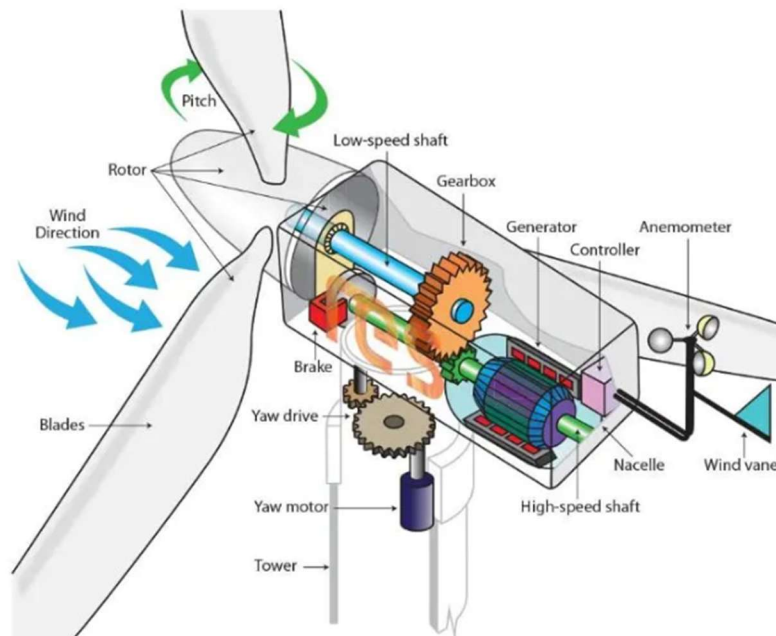
- **Ρότορας (Rotor)** : Ο ρότορας αποτελείται από δύο ή τρία περύγια, στερεωμένα στην πλήμνη (hub), τα οποία περιστρέφονται υπό την επίδραση του ανέμου, αξιοποιώντας αεροδυναμικές αρχές για τη δημιουργία

ανυψωτικής δύναμης, η οποία θέτει σε κίνηση/περιστροφή τον ρότορα [10]. Τα πτερύγια, συνήθως κατασκευασμένα από ενισχυμένο πολυεστέρα ή ανθρακονήματα, είναι ειδικού αεροδυναμικού προφίλ. Η σύνδεσή τους στην πλήμνη γίνεται μέσω ρουλεμάν στήριξης, ενώ εντός αυτής βρίσκεται το σύστημα κίνησής τους για την προσαρμογή της γωνίας λειτουργίας.

- **Σύστημα Μετάδοσης Κίνησης** : Διαφοροποιείται ανάλογα σε περίπτωση άμεσης ή έμμεσης μετάδοσης κίνησης και περιλαμβάνει κατά περίπτωση τους άξονες χαμηλής ταχύτητας περιστροφής (συνήθως 30 έως 60 rpm) και υψηλής ταχύτητας (1000 έως 1800 rpm) και το κιβώτιο ταχυτήτων, καθώς και σύστημα πέδησης (φρένο) για διακοπή περιστροφής σε περίπτωση έκτακτης ανάγκης.
- **Κιβώτιο ταχυτήτων (Gearbox)** : Χρησιμοποιείται σε ανεμογεννήτριες με έμμεση μετάδοση κίνησης και αυξάνει την ταχύτητα περιστροφής, ώστε να τροφοδοτήσει τη γεννήτρια με την κατάλληλη κινητική ενέργεια [10]. Συνοδεύτε με το απαιτούμενο σύστημα λίπανσης και σύστημα ψύξης.
- **Ηλεκτρική Γεννήτρια (Generator)** : Μετατρέπει τη μηχανική ενέργεια από την περιστροφή του ρότορα σε ηλεκτρική ενέργεια. Συνήθως είναι εναλλασσόμενου ρεύματος, σύγχρονη ή επαγωγική, με 4 ή 6 πόλους, ενώ σε ανεμογεννήτριες άμεσης μετάδοσης κίνησης μπορεί να έχει πολλούς πόλους. Είναι κυρίως αερόψυκτες, με ανεμιστήρες για βεβαιωμένη ψύξη. Τα πιο συνηθισμένα μοντέλα είναι οι γεννήτριες διπλής τροφοδοσίας (DFIG - Double Fed Induction Generator), λόγω της ικανότητάς τους να λειτουργούν σε μεταβαλλόμενες ταχύτητες ανέμου [11].
- **Περίβλημα (Nacelle)** : Το περίβλημα ή άτρακτος είναι το τμήμα της ανεμογεννήτριας που στεγάζει όλα τα κύρια μηχανικά και ηλεκτρικά εξαρτήματα, που ευρίσκονται στην κορυφή, όπως το κιβώτιο ταχυτήτων αν υπάρχει, τη γεννήτρια, τους άξονες με τα έδρανα τους και τα συστήματα ελέγχου. Παρέχει προστασία από τις καιρικές συνθήκες και διευκολύνει τη συντήρηση, επιτρέποντας την κίνηση προσωπικού εντός αυτού [10].
- **Σύστημα προσανατολισμού (Yaw System)**: Ρυθμίζει τη γωνία της ανεμογεννήτριας, διασφαλίζοντας ότι είναι προσανατολισμένη σωστά ως προς την κατεύθυνση του ανέμου για μέγιστη απόδοση [10]. Περιλαμβάνει κινητήρα εκτροπής (yaw motor) και μετρητή εκτροπής (yaw meter). Η λειτουργία του συστήματος αυτού εντέλλεται από ανεμοδείκτη (wind vane) που είναι εγκατεστημένος στην ανεμογεννήτρια και καταγράφει συνεχώς την κατεύθυνση του αέρα.
- **Πύργος / Πυλώνας (Tower)**: Στηρίζει τον παραπάνω ηλεκτρομηχανολογικό εξοπλισμό της ανεμογεννήτριας στο απαραίτητο ύψος πάνω από το έδαφος, ώστε τα πτερύγια να εκμεταλλεύονται το αιολικό δυναμικό της περιοχής με το πιο αποτελεσματικό τρόπο [10]. Είναι συνήθως χαλύβδινοι, σωληνοειδής κατασκευής, με σκάλα ή ανελκυστήρα στο εσωτερικό τους για μετακίνηση του προσωπικού στην κορυφή για επιτήρηση και εργασίες συντήρησης.

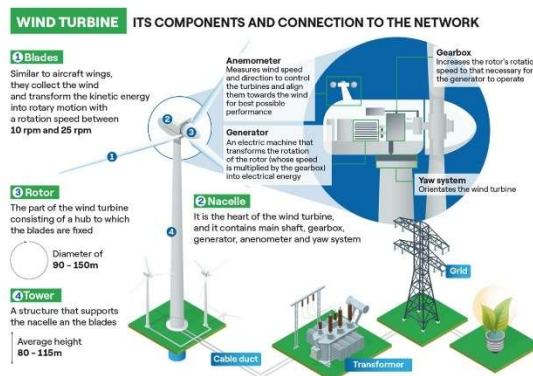
- **Σύστημα ελέγχου και ηλεκτρονικά ισχύος (Control and Power Electronics System):** Περιλαμβάνει αισθητήρες ανέμου, λογισμικό ελέγχου και μηχανισμούς ρύθμισης των πτερυγίων για τη βέλτιστη διαχείριση της παραγόμενης ενέργειας [10,11]. Βασικός «εξοπλισμός» ελέγχου είναι ο ρυθμιστής ταχύτητας περιστροφής και παραγόμενης ισχύος, που εξασφαλίζεται είτε μέσω περιστροφής των πτερυγίων γύρω από τον άξονα τους (blade pitch control) είτε μέσω αεροδυναμικού ελέγχου με κατάλληλο σχεδιασμό των πτερυγίων (passive stall control).
- **Σύστημα θεμελίωσης (Foundation):** Η βάση στήριξης από οπλισμένο σκυρόδεμα, στην οποία εδράζεται ο πύργος, εξασφαλίζει τη σταθερότητα της ανεμογεννήτριας [11].
- **Σύστημα σύνδεσης με το δίκτυο** μεταφοράς ηλεκτρικής ενέργειας (**Grid Connection**): Εξασφαλίζει τη συνεχή μεταφορά της παραγόμενης ηλεκτρικής ενέργειας στο δίκτυο κατά την λειτουργία της ανεμογεννήτριας, και συνήθως περιλαμβάνει μετασχηματιστή ισχύος και τους απαιτούμενους διακόπτες απομόνωσης [11].

Η παρακάτω εικόνα (Εικόνα 5) παρουσιάζει τη δομή και τα βασικά εξαρτήματα μιας ανεμογεννήτριας οριζόντιου άξονα, απεικονίζοντας την αλληλεπίδραση μεταξύ των μηχανικών και ηλεκτρικών συστημάτων της. Διακρίνονται τα πτερύγια, το κιβώτιο ταχυτήτων, η γεννήτρια, το περίβλημα, καθώς και τα συστήματα προσανατολισμού και ελέγχου, που συμβάλλουν στη βέλτιστη λειτουργία της ανεμογεννήτριας.



Εικόνα 5.Κύρια Εξαρτήματα της Ανεμογεννήτριας

Η παρακάτω εικόνα (Εικόνα 6) παρουσιάζει τα βασικά εξαρτήματα των ανεμογεννητριών καθώς και την σύνδεσή τους στο δίκτυο.



Εικόνα 6. Βασικά εξαρτήματα των Ανεμογεννητριών και Σύνδεση στο δίκτυο

2.1.3 Βασικά Οφέλη από την Λειτουργία των Ανεμογεννητριών

Η αιολική ενέργεια, ως βασικός πυλώνας της ενεργειακής μετάβασης, περιορίζει τις εκπομπές διοξειδίου του άνθρακα και ενισχύει την ενεργειακή ασφάλεια, παρέχοντας ένα βιώσιμο και αξιόπιστο μοντέλο παραγωγής ενέργειας για τις επόμενες γενιές. Ως ανανεώσιμη πηγή, δεν εξαντλείται και δεν επιβαρύνει το περιβάλλον με εκπομπές ρύπων, συμβάλλοντας στη μείωση της εξάρτησης από τα ορυκτά καύσιμα. Παράλληλα, η αξιοποίηση της αιολικής ενέργειας μειώνει σημαντικά την κατανάλωση ορυκτών καυσίμων, εξοικονομώντας έως και 300 τόνους ισοδύναμου πετρελαίου ανά MW εγκατεστημένης ισχύος.

Επιπλέον, η χρήση ανεμογεννητριών συμβάλλει στη σημαντική μείωση των εκπομπών CO₂, καθώς μία GWh αιολικής ενέργειας εξοικονομεί περίπου 600 τόνους διοξειδίου του άνθρακα. Αν και η κατασκευή και εγκατάσταση ανεμογεννητριών απαιτεί πόρους, οι εκπομπές που προκύπτουν αντισταθμίζονται μέσα στους πρώτους 3-6 μήνες λειτουργίας τους, καθιστώντας τες μια εξαιρετικά αποδοτική λύση από περιβαλλοντική άποψη.

Τέλος, τα οικονομικά οφέλη της αιολικής ενέργειας είναι επίσης σημαντικά, καθώς δημιουργεί θέσεις εργασίας στον τομέα της κατασκευής, συντήρησης και λειτουργίας των ανεμογεννητριών. Παράλληλα, μειώνει την εξάρτηση από εισαγόμενα καύσιμα, ενισχύοντας την ενεργειακή ανεξαρτησία των χωρών που την αξιοποιούν.

2.2 Συνηθέστερα Σφάλματα Ανεμογεννητριών

Οι ανεμογεννήτριες είναι σύνθετα ηλεκτρομηχανολογικά συστήματα που υπόκεινται σε διάφορα σφάλματα, τα οποία επηρεάζουν την απόδοση και τη λειτουργικότητά

τους. Τα συνηθέστερα σφάλματα περιλαμβάνουν ηλεκτρικές και μηχανικές βλάβες, προβλήματα στα πτερύγια, δυσλειτουργίες στα συστήματα ψύξης και ελέγχου, καθώς και ανωμαλίες στη συλλογή και επεξεργασία δεδομένων [12,13,14]. Οι ηλεκτρικές βλάβες αφορούν κυρίως τη γεννήτρια, όπου η υπερθέρμανση και η φθορά από συνεχείς μεταβολές φορτίου μπορούν να προκαλέσουν διακοπές λειτουργίας [12]. Οι μηχανικές βλάβες επηρεάζουν το κιβώτιο ταχυτήτων και τα ρουλεμάν, οδηγώντας σε αυξημένη μηχανική καταπόνηση και πιθανή αστοχία [13].

Επιπλέον, η έκθεση των πτερυγίων σε ακραίες καιρικές συνθήκες μπορεί να οδηγήσει σε διάβρωση, ρωγμές και δομικές φθορές λόγω συνεχούς έκθεσης σε βροχή, σκόνη και χαλάζι [14]. Παράλληλα, προβλήματα στα συστήματα ψύξης και λίπανσης ενδέχεται να προκαλέσουν υπερθέρμανση σε κρίσιμα εξαρτήματα, όπως το κιβώτιο ταχυτήτων και η γεννήτρια. Δυσλειτουργίες στον μηχανισμό ελέγχου γωνίας πτερυγίων (pitch control) και στο σύστημα ευθυγράμμισης (yaw control) μπορεί να μειώσουν την ενεργειακή απόδοση της ανεμογεννήτριας, προκαλώντας μηχανικές καταπονήσεις [14]. Η εμφάνιση τέτοιων σφαλμάτων αυξάνει το κόστος συντήρησης και τις μη προγραμματισμένες διακοπές λειτουργίας, καθιστώντας αναγκαία την εφαρμογή τεχνικών πρόβλεψης και συντήρησης μέσω ανάλυσης δεδομένων SCADA [12,13].

Η παρούσα διπλωματική εργασία επικεντρώνεται στη γεννήτρια και στο κιβώτιο ταχυτήτων, τα οποία αποτελούν δύο από τα πιο κρίσιμα υποσυστήματα της ανεμογεννήτριας, καθώς επηρεάζουν άμεσα την αποδοτικότητα και τη βιωσιμότητα της λειτουργίας της.

2.3 Συντήρηση Ανεμογεννητριών

Η συντήρηση των ανεμογεννητριών αποτελεί κρίσιμη διαδικασία για τη διασφάλιση της απρόσκοπτης λειτουργίας τους και τη μεγιστοποίηση της διάρκειας ζωής τους. Το κόστος και οι απαιτήσεις συντήρησης ενός αιολικού πάρκου επηρεάζουν άμεσα τόσο το αρχικό κεφάλαιο επένδυσης (CAPEX) όσο και το λειτουργικό κόστος (OPEX) του έργου.

Οι συνήθεις διαδικασίες συντήρησης καθορίζονται από τις οδηγίες του κατασκευαστή και περιλαμβάνουν προγραμματισμένες επιθεωρήσεις, ελέγχους με ειδικό εξοπλισμό, συσφίξεις, αλλαγές αναλωσίμων και αντικαταστάσεις εξαρτημάτων, βάσει των ωρών λειτουργίας και της προβλεπόμενης διάρκειας ζωής τους. Στις σύγχρονες ανεμογεννήτριες, η προγραμματισμένη συντήρηση πραγματοποιείται συνήθως μία φορά τον χρόνο, ενώ η προληπτική συντήρηση εντάσσεται επίσης σε αυτήν την κατηγορία.

Ωστόσο, η πολυπλοκότητα των ανεμογεννητριών και η έκθεσή τους σε μεταβαλλόμενες περιβαλλοντικές συνθήκες καθιστούν αναπόφευκτες τις μη προγραμματισμένες επισκευαστικές συντηρήσεις, οι οποίες προκύπτουν έπειτα από απρόοπτες βλάβες. Σε αυτές τις περιπτώσεις, η έγκαιρη διάγνωση και η διαθεσιμότητα κρίσιμων ανταλλακτικών παίζουν καθοριστικό ρόλο στη μείωση του χρόνου εκτός λειτουργίας (downtime) και στη βελτίωση της αποδοτικότητας του συστήματος.

Η συνεχώς αυξανόμενη ανάγκη για έγκαιρη συντήρηση των ανεμογεννητριών έχει οδηγήσει στην ανάπτυξη στρατηγικών που επιτρέπουν την ανίχνευση πιθανών δυσλειτουργιών πριν την εκδήλωση σοβαρών βλαβών. Αυτές οι προληπτικές προσεγγίσεις στοχεύουν στην αποτροπή αστοχιών, διασφαλίζοντας τη βέλτιστη λειτουργία του εξοπλισμού και τη μείωση του κόστους επισκευών.

Κεφαλαίο 3 : Παρακολούθηση και Βελτιστοποίηση της Λειτουργίας των Ανεμογεννητριών

Η αποδοτική λειτουργία των ανεμογεννητριών απαιτεί την συνεχή παρακολούθηση και ανάλυση δεδομένων, ώστε να ανιχνεύονται ανωμαλίες και να προλαμβάνονται πιθανές βλάβες πριν εξελιχθούν σε σοβαρά προβλήματα.

Τα συστήματα SCADA επιτρέπουν τη συλλογή και αποθήκευση δεδομένων σε πραγματικό χρόνο, και σε συνδυασμό με την χρήση αλγορίθμων μηχανικής μάθησης παρέχουν τις απαραίτητες πληροφορίες για την ανίχνευση πρώιμων ενδείξεων φθοράς. Αυτή η προσέγγιση καθιστά δυνατή την εφαρμογή προγνωστικής συντήρησης (Predictive Maintenance - PdM), μειώνοντας τον κίνδυνο σοβαρών αστοχιών.

Στην παρούσα ενότητα εξετάζεται η σημασία αυτών των τεχνολογιών, ο τρόπος λειτουργίας τους και η συμβολή τους στη βελτίωση της αποδοτικότητας και της αξιοπιστίας των ανεμογεννητριών.

3.1 Συστήματα SCADA

Τα συστήματα συλλογής δεδομένων SCADA (Supervisory Control and Data Acquisition) αποτελούν θεμελιώδη τεχνολογία για την παρακολούθηση και διαχείριση της λειτουργίας εξοπλισμού και βιομηχανικών διαδικασιών. Ειδικά στις ανεμογεννήτριες, παίζουν καθοριστικό ρόλο στη διασφάλιση της αποδοτικής λειτουργίας τους, επιτρέποντας τη συλλογή δεδομένων σε πραγματικό χρόνο. Μέσω της διεπαφής ανθρώπου-μηχανής (HMI), οι χειριστές μπορούν να παρακολουθούν κρίσιμες παραμέτρους και να αλληλεπιδρούν με τον εξοπλισμό, διασφαλίζοντας την ομαλή λειτουργία των ανεμογεννητριών.

Τα δεδομένα που συλλέγονται από τα συστήματα SCADA αποθηκεύονται, αναλύονται και χρησιμοποιούνται για την αυτοματοποίηση και την βελτιστοποίηση της λειτουργίας των ανεμογεννητριών, συμβάλλοντας στη βελτίωση της παραγωγικότητας, της αποδοτικότητας και της αξιοπιστίας τους. Επιπλέον, τα συστήματα αυτά επιτρέπουν τη διαχείριση των ανεμογεννητριών τόσο τοπικά όσο και εξ αποστάσεως, εξασφαλίζοντας την έγκαιρη λήψη αποφάσεων για την πρόληψη βλαβών και την ελαχιστοποίηση των διακοπών λειτουργίας.

Με την αξιοποίηση των δεδομένων SCADA σε μοντέλα πρόβλεψης, επιτυγχάνεται η έγκαιρη ανίχνευση ανωμαλιών που υποδηλών επικείμενες βλάβες, ενώ παράλληλα επιτρέπεται η αυτοματοποιημένη ρύθμιση των παραμέτρων λειτουργίας για τη βελτίωση της συνολικής αποδοτικότητας του συστήματος.

3.1.1 Στοιχεία ενός Συστήματος SCADA & Αρχιτεκτονική SCADA

Τα συστήματα SCADA χρησιμοποιούνται για την απομακρυσμένη διαχείριση βιομηχανικών εγκαταστάσεων, βελτιώνοντας την αποδοτικότητα και την παραγωγή. Αποτελούνται από υλικό (hardware) και λογισμικό (software), με βασικά στοιχεία [15]:

- **Συσκευές πεδίου:** Αισθητήρες που καταγράφουν δεδομένα και ενεργοποιητές που ελέγχουν τις διεργασίες.
- **Ελεγκτές πεδίου (PLCs & RTUs):** Μικροϋπολογιστές που συλλέγουν δεδομένα από τις συσκευές πεδίου και αποστέλλουν εντολές.
- **Υπολογιστές επιτήρησης SCADA:** Κεντρικά συστήματα που ελέγχουν τις διεργασίες και επεξεργάζονται δεδομένα.
- **Υποδομή επικοινωνίας:** Δίκτυα που επιτρέπουν τη σύνδεση των συσκευών πεδίου με το σύστημα επιτήρησης.

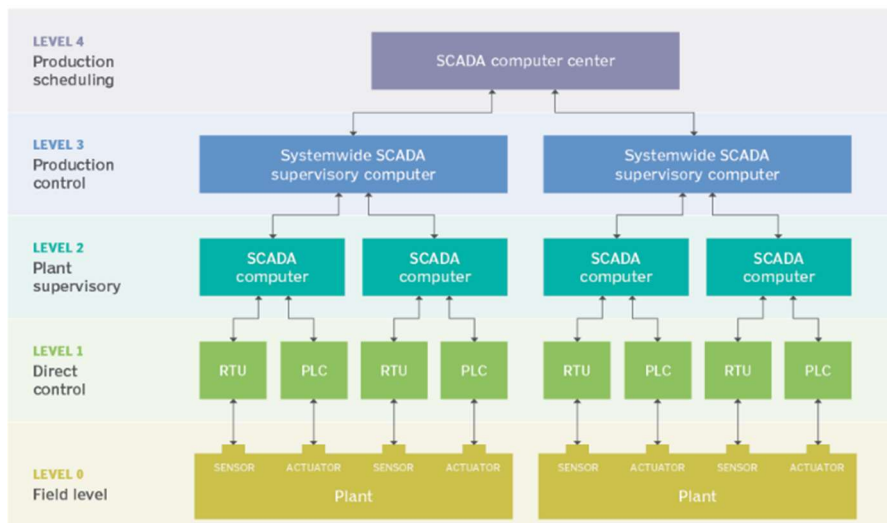
Η αρχιτεκτονική SCADA βασίζεται στο Purdue Enterprise Reference Architecture (PERA) και περιλαμβάνει πέντε επίπεδα [15]:

1. **Επίπεδο πεδίου:** Αισθητήρες και ενεργοποιητές που καταγράφουν δεδομένα και ελέγχουν διεργασίες.
2. **Επίπεδο άμεσου ελέγχου:** Τοπικοί ελεγκτές (PLCs, RTUs) που αλληλεπιδρούν με τις συσκευές πεδίου.
3. **Επίπεδο εποπτείας εγκατάστασης:** Εποπτικά συστήματα που συγκεντρώνουν δεδομένα και δίνουν εντολές.
4. **Επίπεδο εποπτείας παραγωγής:** Συνολική παρακολούθηση και αναφορές προς τα ανώτερα επίπεδα.
5. **Επίπεδο επιχειρησιακής υποστήριξης:** Επιχειρησιακά συστήματα διαχείρισης και υποδομές IT.

Αυτή η δομή επιτρέπει τη συλλογή, επεξεργασία και ανάλυση δεδομένων σε πραγματικό χρόνο, βελτιώνοντας την απόδοση και τον έλεγχο των βιομηχανικών διεργασιών.

Η παρακάτω εικόνα (Εικόνα 7) παρουσιάζει τα διάφορα επίπεδα λειτουργίας των SCADA, από το επίπεδο πεδίου μέχρι το κέντρο παραγωγής, ενώ αναδεικνύει τον ρόλο κάθε επιπέδου στη συλλογή, ανάλυση και αξιοποίηση δεδομένων.

Layers of the SCADA system architecture



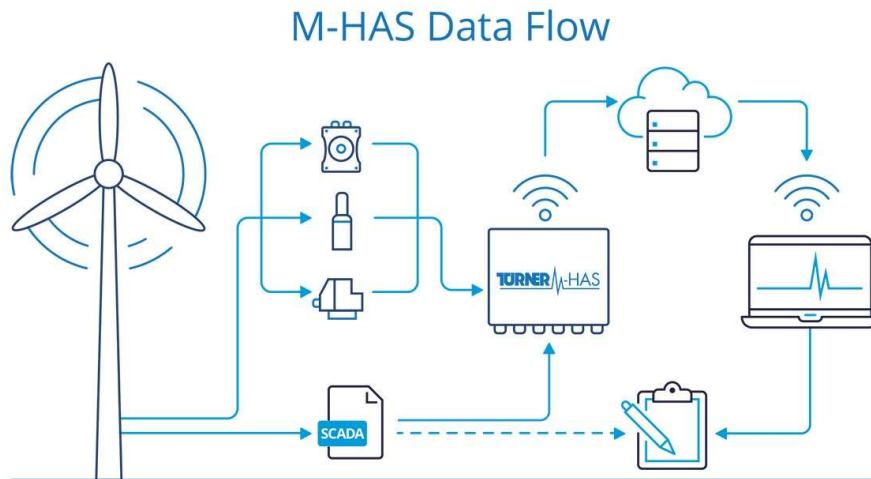
Εικόνα 7. Αρχιτεκτονική Συστήματος SCADA

3.1.2 Βασικές Λειτουργίες των SCADA

Τα συστήματα SCADA εκτελούν βασικές λειτουργίες εξασφαλίζοντας την αποδοτική και αποτελεσματική διαχείριση και εποπτεία των βιομηχανικών διαδικασιών. Βασικός τους ρόλος είναι η συλλογή δεδομένων, η οποία πραγματοποιείται μέσω αισθητήρων και ενεργοποιητών, αξιοποιώντας προγραμματιζόμενους λογικούς ελεγκτές (PLCs) ή απομακρυσμένες τερματικές μονάδες (RTUs). Τα δεδομένα αυτά περιλαμβάνουν κρίσιμες μετρούμενες παραμέτρους, προσφέροντας στους χειριστές συνεχή και ολοκληρωμένη εικόνα της λειτουργίας του συστήματος. Εκτός από την παρακολούθηση, τα SCADA επιτρέπουν τον απομακρυσμένο έλεγχο και τον αυτοματισμό των διαδικασιών, επιτρέποντας στους χειριστές ή σε αυτόματα συστήματα ελέγχου να ρυθμίζουν κρίσιμες παραμέτρους ή λειτουργίες, διασφαλίζοντας τη βέλτιστη απόδοση του εξοπλισμού.

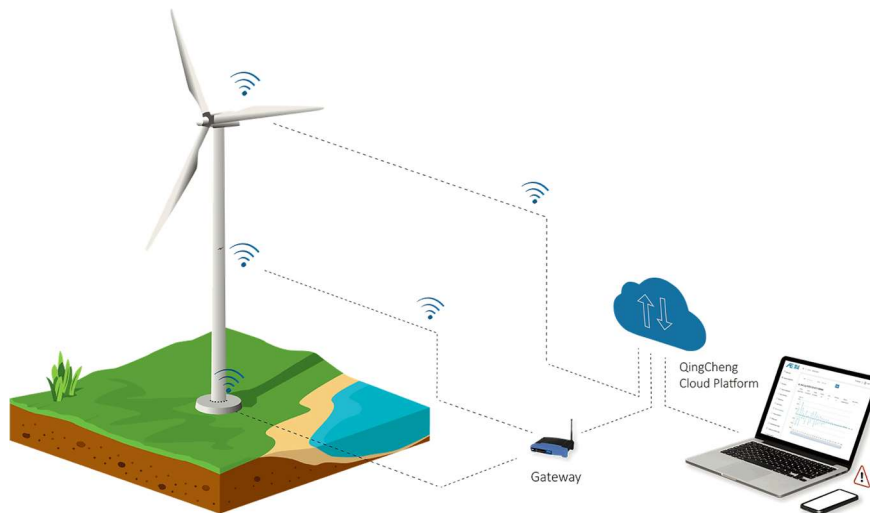
Παράλληλα, διαθέτουν μηχανισμούς συναγερμών και ειδοποιήσεων, που ενεργοποιούνται όταν κρίσιμες παράμετροι υπερβαίνουν τα προκαθορισμένα όρια, προειδοποιώντας για πιθανές δυσλειτουργίες. Επιπλέον, τα δεδομένα που συλλέγονται καταγράφονται και αποθηκεύονται για μελλοντική ανάλυση, επιτρέποντας την εφαρμογή αλγορίθμων μηχανικής μάθησης με σκοπό την ανίχνευση τάσεων και μοτίβων σφαλμάτων. Η αξιοποίηση αυτών των ιστορικών δεδομένων συμβάλλει στη βελτιστοποίηση της λειτουργίας του συστήματος και στη λήψη τεκμηριωμένων αποφάσεων για τη συντήρηση και τη βελτίωση της απόδοσης.

Στην Εικόνα 8, απεικονίζεται η ροή δεδομένων σε ένα σύστημα SCADA, περιγράφοντας τη διαδικασία συλλογής πληροφοριών από αισθητήρες, τη μεταφορά τους σε πύλες επεξεργασίας, καθώς και την ανάλυσή τους μέσω cloud πλατφορμών και εξειδικευμένων λογισμικών.



Εικόνα 8.Ροή δεδομένων SCADA στις Ανεμογεννήτριες

Η Εικόνα 9 απεικονίζει τη παρακολούθηση των ανεμογεννητριών μέσω ενός συστήματος SCADA, όπου οι αισθητήρες συλλέγουν δεδομένα και τα μεταδίδουν μέσω δικτύου σε μια κεντρική πλατφόρμα cloud για ανάλυση και επιτήρηση.



Εικόνα 9.Παρακολούθηση Ανεμογεννητριών

3.2 Προγνωστική Συντήρηση - Predictive Maintenance (PdM)

Η προγνωστική συντήρηση είναι μια στρατηγική που στοχεύει στην ανίχνευση ανωμαλιών πριν οδηγηθεί σε σφάλμα ο εξοπλισμός και επιδιώκει να μειώσει και να αποτρέψει τις βλάβες και τον χρόνο εκτός λειτουργίας του εξοπλισμού.

3.2.1 Ιστορική Εξέλιξη των Στρατηγικών Συντήρησης

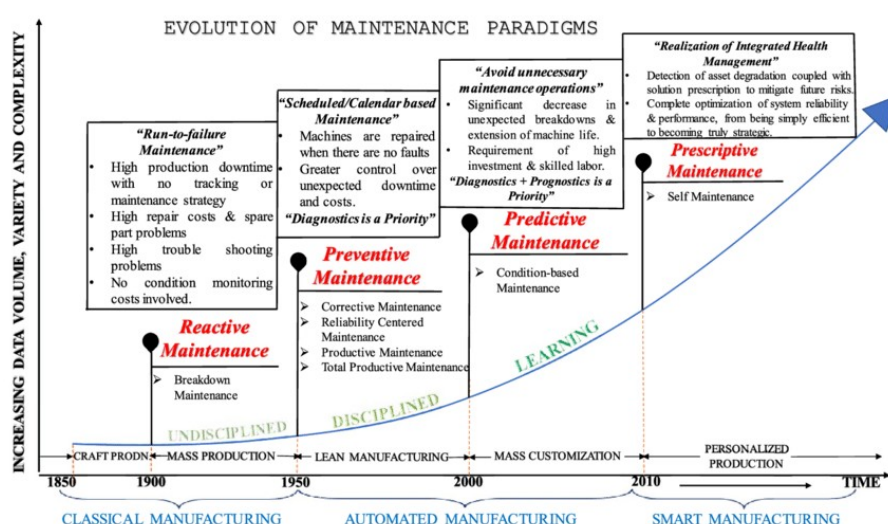
Η εξέλιξη των στρατηγικών συντήρησης ξεκίνησε με την αντιδραστική συντήρηση (Reactive Maintenance), όπου οι επισκευές πραγματοποιούνταν μόνο μετά την

εμφάνιση μιας βλάβης, οδηγώντας σε υψηλά κόστη επισκευών και απρόβλεπτες διακοπές λειτουργίας. Στη συνέχεια, υιοθετήθηκε η προληπτική συντήρηση (Preventive Maintenance), η οποία πραγματοποιούνταν σε προγραμματισμένα χρονικά διαστήματα, μειώνοντας τις βλάβες, αλλά συχνά οδηγώντας σε περιττές συντηρήσεις και αυξημένο κόστος [16].

Το επόμενο βήμα ήταν η συντήρηση βάσει κατάστασης (Condition-Based Maintenance - CBM), η οποία αξιοποιούσε δεδομένα πραγματικού χρόνου για την ανίχνευση ανωμαλιών, επιτρέποντας τη συντήρηση μόνο όταν αυτή ήταν απαραίτητη. Ωστόσο, η μεγαλύτερη πρόοδος επιτεύχθηκε με την Προγνωστική Συντήρηση (Predictive Maintenance - PdM), η οποία χρησιμοποιεί αλγορίθμους μηχανικής μάθησης και ανάλυση δεδομένων SCADA για την πρόβλεψη πιθανών βλαβών πριν αυτές εμφανιστούν. Αυτή η προσέγγιση μείωσε σημαντικά το κόστος επισκευών, αύξησε την αξιοπιστία των συστημάτων και βελτίωσε τη συνολική αποδοτικότητα [16].

Πιο πρόσφατα, έχει αναπτυχθεί η Προτρεπτική Συντήρηση (Prescriptive Maintenance - RxM), η οποία επεκτείνει τις δυνατότητες της PdM, προτείνοντας βέλτιστες ενέργειες για την αποφυγή βλαβών, βασισμένες σε ανάλυση δεδομένων και μοντέλα πρόβλεψης.

Το παρακάτω διάγραμμα (Εικόνα 10) απεικονίζει τη διαχρονική εξέλιξη των στρατηγικών συντήρησης, από την αντιδραστική συντήρηση έως τις σύγχρονες στρατηγικές όπως η Προγνωστική και η Προτρεπτική Συντήρηση, αναδεικνύοντας τη συνεχή πρόοδο προς πιο αποδοτικές και στρατηγικές προσεγγίσεις.



Εικόνα 10. Εξέλιξη Στρατηγικών Συντήρησης

3.2.2 Τρόπο Λειτουργίας της PdM

Η Predictive Maintenance (PdM) βασίζεται στην ανάλυση δεδομένων για την πρόβλεψη και πρόληψη βλαβών, επιτρέποντας τη δυναμική προσαρμογή των

εργασιών συντήρησης βάσει της πραγματικής κατάστασης του εξοπλισμού [16]. Η διαδικασία ξεκινά με τη συλλογή και επεξεργασία δεδομένων από συστήματα SCADA, όπου εφαρμόζεται φιλτράρισμα για την απομάκρυνση θορύβων και σφαλμάτων, ενώ η αποθήκευση πραγματοποιείται σε βάσεις δεδομένων ή cloud πλατφόρμες για άμεση πρόσβαση.

Στη συνέχεια, προηγμένοι αλγόριθμοι μηχανικής μάθησης, όπως το XGBoost, καθώς και στατιστικές μέθοδοι, χρησιμοποιούνται για την ανίχνευση αποκλίσεων από τη φυσιολογική λειτουργία. Οι τεχνικές αυτές επιτρέπουν την πρόβλεψη της Εναπομένουσας Ωφέλιμης Ζωής (RUL) των εξαρτημάτων και την ανίχνευση πρώιμων ενδείξεων σφαλμάτων, δίνοντας τη δυνατότητα έγκαιρων παρεμβάσεων. Το τελικό στάδιο περιλαμβάνει τη λήψη αποφάσεων βάσει των προβλέψεων, επιτρέποντας στοχευμένες εργασίες συντήρησης. [16].

Η παρακάτω εικόνα (Εικόνα 11) απεικονίζει τη ροή δεδομένων από ανεμογεννήτριες, αισθητήρες και άλλες πηγές σε μια cloud πλατφόρμα, όπου τα δεδομένα επεξεργάζονται, αναλύονται και διανέμονται σε εργαλεία και ομάδες για την υποστήριξη της Predictive Maintenance (PdM).



Εικόνα 11.Ροή Δεδομένων σε Cloud Πλατφόρμα για Predictive Maintenance

3.3 Προκλήσεις της Παρακολούθησης των Ανεμογεννητριών

Παρά τις σημαντικές εξελίξεις στην παρακολούθηση και συντήρηση των ανεμογεννητριών, η εφαρμογή συστημάτων παρακολούθησης τους εξακολουθεί να παρουσιάζει προκλήσεις. Κάποιες από τις βασικές είναι :

Ποιότητα και Πληρότητα Δεδομένων

Η απόδοση της Προγνωστικής Συμτήρησης εξαρτάται από την ποιότητα των δεδομένων SCADA. Αυτά πρέπει να είναι Σχετικά (Relevant), Ακριβή (Accurate) και

Κατανοητά (Understandable). Η παρουσία θορύβου ή λανθασμένων μετρήσεων μπορεί να οδηγήσει σε εσφαλμένες εκτιμήσεις και λανθασμένες αποφάσεις [16].

Υψηλό Κόστος Αρχικής Επένδυσης

Η εφαρμογή συστημάτων PdM απαιτεί σημαντικές επενδύσεις σε αισθητήρες, προηγμένο λογισμικό ανάλυσης δεδομένων και εξειδικευμένο προσωπικό. Το υψηλό αρχικό κόστος αποτελεί εμπόδιο για την υιοθέτηση αυτών των τεχνολογιών, καθώς περιλαμβάνει τόσο το κόστος εξοπλισμού όσο και την ανάγκη διαρκούς εκπαίδευσης προσωπικού [16].

Διαχείριση Μεγάλου Όγκου Δεδομένων

Τα συστήματα SCADA συλλέγουν τεράστιες ποσότητες δεδομένων σε πραγματικό χρόνο, οι οποίες πρέπει να υποβάλλονται σε επεξεργασία, φιλτράρισμα και ανάλυση. Η αξιοποίηση αυτών των δεδομένων για την παραγωγή αξιόπιστων προβλέψεων απαιτεί υπολογιστικούς πόρους υψηλής απόδοσης, καθώς και προηγμένους αλγορίθμους μηχανικής μάθησης, γεγονός που αυξάνει την πολυπλοκότητα και το λειτουργικό κόστος [16].

Εκπαίδευση Προσωπικού

Η χρήση αυτών των συστημάτων απαιτεί προσωπικό με εξειδικευμένες γνώσεις και δεξιότητες. Η έλλειψη εξειδικευμένου προσωπικού αποτελεί εμπόδιο στη υιοθέτηση αυτών των τεχνολογιών [16].

Μεταβαλλόμενες Περιβαλλοντικές Συνθήκες

Οι ανεμογεννήτριες λειτουργούν σε δυναμικές καιρικές συνθήκες, οι οποίες μπορούν να επηρεάσουν την ακρίβεια των προβλέψεων. Οι εξωτερικές περιβαλλοντικές μεταβολές μπορεί να οδηγήσουν σε αναξιόπιστα δεδομένα, απαιτώντας διαρκή προσαρμογή των αλγορίθμων πρόβλεψης ώστε να ανταποκρίνονται στις πραγματικές συνθήκες λειτουργίας [16].

3.4 Μελλοντικές Τεχνολογικές Τάσεις στην Παρακολούθηση και Συντήρηση των Ανεμογεννητριών

Η συνεχής εξέλιξη των τεχνολογιών παρακολούθησης και συντήρησης των ανεμογεννητριών στοχεύει στη βελτίωση της αποδοτικότητας, της ακρίβειας και της προγνωστικής ικανότητας των συστημάτων. Οι μελλοντικές εξελίξεις επικεντρώνονται στην ενίσχυση των δυνατοτήτων των SCADA, την αυτοματοποίηση διαδικασιών συντήρησης και τη βελτίωση της ανάλυσης δεδομένων. Κάποιες από αυτές είναι:

Ανάπτυξη των Συστημάτων SCADA και Ενσωμάτωση νέων Τεχνολογιών

Η εξέλιξη των συστημάτων SCADA επικεντρώνεται στην βελτίωση της συνδεσιμότητας τους και της ασφάλειας, αξιοποιώντας τεχνολογίες όπως το 5G και το Internet of Things (IoT). Επίσης η ενσωμάτωση του edge computing μειώνει την ανάγκη για κεντρική επεξεργασία δεδομένων, προσφέροντας ταχύτερη ανταπόκριση [17].

Μηχανική Μάθηση και Ψηφιακοί Δίδυμοι

Η εφαρμογή της μηχανικής μάθησης και της τεχνητής νοημοσύνης στην προγνωστική συντήρηση επιτρέπει την ακριβέστερη πρόβλεψη βλαβών και την μείωση του κόστους συντήρησης. Παράλληλα η εφαρμογή των Ψηφιακών Δίδυμων (Digital Twins), που λειτουργούν ως εικονικά αντίγραφα του εξοπλισμού, προσφέρουν προσομοιώσεις που επιτρέπουν την έγκαιρη διάγνωση και βελτιστοποίηση των διαδικασιών συντήρησης [16,18].

Αυτοματοποιημένη Συντήρηση

Στο μέλλον, η Predictive Maintenance (PdM) δεν θα περιορίζεται στην πρόβλεψη του πότε και τι πρέπει να επισκευαστεί, αλλά θα αυτοματοποιεί πλήρως τη διαδικασία συντήρησης, δημιουργώντας συστήματα αυτοσυντήρησης. Μέσω της τεχνητής νοημοσύνης (AI), τα νέα αυτά συστήματα θα μπορούν να εντοπίζουν πρότυπα σφαλμάτων και να εφαρμόζουν διορθωτικές ενέργειες σε πραγματικό χρόνο, βελτιώνοντας τη λειτουργία των ανεμογεννητριών και αποτρέποντας μη προγραμματισμένες διακοπές [16].

3.5 Οικονομικά και Περιβαλλοντικά Οφέλη της Προγνωστικής Συντήρησης

Η προγνωστική συντήρηση συμβάλλει σημαντικά στη μείωση του κόστους συντήρησης και στη βελτιστοποίηση των διαθέσιμων πόρων. Μέσω της συνεχούς παρακολούθησης των ανεμογεννητριών, περιορίζει τις μη προγραμματισμένες επισκευές και αποτρέπει τις άσκοπες παρεμβάσεις, μειώνοντας έτσι το κόστος εργασιών και την ανάγκη για ανταλλακτικά.

Η χρήση δεδομένων SCADA και αλγορίθμων μηχανικής μάθησης επιτρέπει την έγκαιρη διάγνωση σφαλμάτων, διασφαλίζοντας καλύτερο προγραμματισμό συντήρησης και αποδοτικότερη διαχείριση του ανθρώπινου δυναμικού. Παράλληλα, η στοχευμένη συντήρηση μειώνει τον χρόνο διακοπής λειτουργίας των ανεμογεννητριών, αυξάνοντας την παραγωγή ενέργειας και ενισχύοντας τη βιωσιμότητα των αιολικών πάρκων. Επιπλέον, οι έγκαιρες παρεμβάσεις περιορίζουν τη φθορά των κρίσιμων εξαρτημάτων, επιμηκύνοντας τη διάρκεια ζωής τους και μειώνοντας τις δαπάνες αντικατάστασης.

Εκτός από τα οικονομικά οφέλη, η προγνωστική συντήρηση έχει θετικό αντίκτυπο στο περιβάλλον. Η μείωση των έκτακτων ή άσκοπων επισκευών και των αντικαταστάσεων εξαρτημάτων περιορίζει τα βιομηχανικά απόβλητα και την κατανάλωση πρώτων υλών. Παράλληλα, η βελτιστοποίηση της λειτουργίας των ανεμογεννητριών μειώνει τις ενεργειακές απώλειες και την ανάγκη για εφεδρική παραγωγή ενέργειας από συμβατικές πηγές, συμβάλλοντας έτσι στη μείωση των εκπομπών CO₂ και στην περαιτέρω απεξάρτηση από ορυκτά καύσιμα.

Κεφάλαιο 4 : Μηχανική Μάθηση και ο Αλγόριθμος XGBoost

Στο παρόν κεφάλαιο γίνεται μια συνοπτική αναφορά στους αλγορίθμους μηχανικής μάθησης, με ιδιαίτερη έμφαση στον XGBoost, ο οποίος θα αναλυθεί διεξοδικά, καθώς αποτελεί τον αλγόριθμο που θα αξιοποιηθεί στη διπλωματική εργασία για την πρόβλεψη σφαλμάτων σε ανεμογεννήτριες.

Παράλληλα, παρουσιάζονται οι μέθοδοι αξιολόγησης των αποτελεσμάτων, συμπεριλαμβανομένων των στατιστικών δεικτών και των διαγραμμάτων SPC (Statistical Process Control), τα οποία συμβάλλουν στην κατανόηση της απόδοσης του μοντέλου και στην ανίχνευση σφαλμάτων. Μέσα από αυτήν την ανάλυση, αναδεικνύεται η σημασία της αξιολόγησης και βελτιστοποίησης των αλγορίθμων μηχανικής μάθησης, διασφαλίζοντας την αποτελεσματικότητα των προβλέψεων και τη βέλτιστη συντήρηση των ανεμογεννητριών

4.1 Εισαγωγή στην Μηχανική Μάθηση

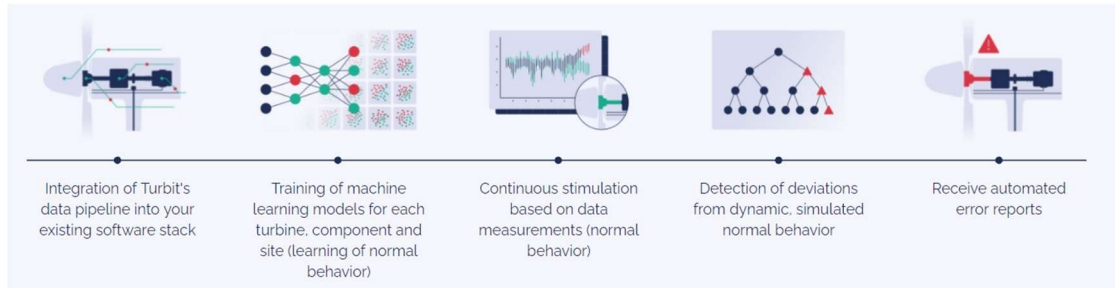
Η Μηχανική Μάθηση (Machine Learning - ML) διαδραματίζει κεντρικό ρόλο στη βελτίωση της λειτουργίας και της αξιοπιστίας πολλών βιομηχανικών διαδικασιών, επιτρέποντας την πρόβλεψη κρίσιμων παραμέτρων και τη προληπτική συντήρηση.

Αποτελεί πεδίο της Τεχνητής Νοημοσύνης (Artificial Intelligence - AI), το οποίο επιτρέπει στους υπολογιστές να μαθαίνουν και να βελτιώνονται από δεδομένα χωρίς να είναι ρητά προγραμματισμένοι. Οι αλγόριθμοι μηχανικής μάθησης εκπαιδεύονται σε μεγάλα σύνολα δεδομένων, αναγνωρίζοντας μοτίβα και συσχετίσεις, με σκοπό να κάνουν προβλέψεις και να λαμβάνουν αποφάσεις [19,20,21].

Η Μηχανική Μάθηση βασίζεται στη χρήση δεδομένων για την εκπαίδευση υπολογιστικών μοντέλων, τα οποία στη συνέχεια χρησιμοποιούνται για την ανάλυση νέων δεδομένων και τη λήψη αποφάσεων. Η διαδικασία περιλαμβάνει συγκεκριμένα στάδια που επιτρέπουν στα μοντέλα να μαθαίνουν και να βελτιώνονται με την πάροδο του χρόνου [19].

1. **Συλλογή Δεδομένων:** Τα δεδομένα μπορεί να περιλαμβάνουν μετρήσεις από αισθητήρες, ιστορικά δεδομένα λειτουργίας ή άλλες πηγές πληροφορίας, οι οποίες τροφοδοτούν το σύστημα εκπαίδευσης.
2. **Εκπαίδευση Μοντέλου:** Τα δεδομένα χρησιμοποιούνται για την εκπαίδευση αλγορίθμων, οι οποίοι αναλύουν τα διαθέσιμα στοιχεία, εντοπίζουν μοτίβα και αναπτύσσουν ικανότητα πρόβλεψης.

3. **Αξιολόγηση και Βελτιστοποίηση:** Το μοντέλο αξιολογείται με βάση την ακρίβεια των προβλέψεών του, ενώ βελτιστοποιείται συνεχώς μέσω της ανατροφοδότησης από νέα δεδομένα, προσαρμόζοντας τη συμπεριφορά του για μεγαλύτερη αξιοπιστία.



Εικόνα 12.Ροή της Μεθοδολογίας

4.1.1 Προγνωστική Συντήρηση και Μηχανική Μάθηση

Η Προγνωστική Συντήρηση αποτελεί μία από τις σημαντικότερες εφαρμογές της Μηχανικής Μάθησης στη βιομηχανία, επιτρέποντας την έγκαιρη ανίχνευση πιθανών βλαβών πριν αυτές εξελιχθούν σε σοβαρά προβλήματα. Αντί να βασίζεται σε προγραμματισμένες συντηρήσεις ή να επεμβαίνει μόνο μετά την εμφάνιση δυσλειτουργιών, αξιοποιεί ιστορικά και ζωντανά δεδομένα για την πρόβλεψη αστοχιών στα μηχανήματα και τα συστήματα [19,20].

Η χρήση αλγορίθμων μηχανικής μάθησης επιτρέπει την αυτόματη ανάλυση μεγάλων συνόλων δεδομένων, εντοπίζοντας πρότυπα και μοτίβα που σχετίζονται με αστοχίες. Αυτό οδηγεί σε μείωση των χρόνων διακοπής, βελτίωση της λειτουργικής απόδοσης και περιορισμό του κόστους συντήρησης. Επιπλέον, τα μοντέλα μηχανικής μάθησης προσαρμόζονται εύκολα σε νέα δεδομένα, βελτιώνοντας την ακρίβεια των προβλέψεων και την αποτελεσματικότητα των προληπτικών παρεμβάσεων [21,22].

4.2 Ο Αλγόριθμος XGBoost

Ο XGBoost (Extreme Gradient Boosting) είναι ένας ισχυρός αλγόριθμος μηχανικής μάθησης, γνωστός για την υψηλή αποδοτικότητα και ακρίβειά του σε προβλήματα παλινδρόμησης και ταξινόμησης. Χρησιμοποιείται ευρέως στη βιομηχανία, λόγω της ικανότητάς του να επεξεργάζεται μεγάλα και σύνθετα σύνολα δεδομένων, προσφέροντας ταχύτητα και αξιοπιστία στις προβλέψεις.

Στην παρούσα διπλωματική εργασία, ο XGBoost θα εφαρμοστεί σε δεδομένα SCADA από ανεμογεννήτριες, με σκοπό την πρόβλεψη κρίσιμων παραμέτρων, όπως οι θερμοκρασίες του κιβωτίου ταχυτήτων και της γεννήτριας, καθώς και την έγκαιρη ανίχνευση σφαλμάτων. Η ανάλυση που ακολουθεί εστιάζει στις βασικές αρχές

λειτουργίας και στα πλεονεκτήματα του XGBoost, τα οποία τον καθιστούν ιδιαίτερα κατάλληλο για την εφαρμογή αυτή.

4.2.1 Αρχές Λειτουργίας του XGBoost

Ο XGBoost είναι ένας αλγόριθμος μηχανικής μάθησης που βασίζεται στη μέθοδο Gradient Boosting Decision Trees (GBDTs) και έχει σχεδιαστεί για να προσφέρει υψηλή ακρίβεια και απόδοση. Τα δέντρα απόφασης είναι μοντέλα που λαμβάνουν αποφάσεις με βάση διαδοχικές ερωτήσεις, διαχωρίζοντας τα δεδομένα σε μικρότερες ομάδες με απλές συνθήκες. Ο XGBoost δημιουργεί διαδοχικά τέτοια δέντρα, διορθώνοντας τα λάθη των προηγούμενων για να βελτιώσει σταδιακά τις προβλέψεις του. Η λειτουργία του περιλαμβάνει τα εξής στάδια: [22,23]:

1. Εισαγωγή Δεδομένων και Αρχικοποίηση του Μοντέλου

Η διαδικασία ξεκινά με την εισαγωγή των δεδομένων και μια αρχική εκτίμηση της τιμής που θέλουμε να προβλέψουμε. Στα προβλήματα παλινδρόμησης, αυτή η αρχική εκτίμηση είναι συνήθως ο μέσος όρος των υπαρχουσών τιμών.

2. Υπολογισμός Υπολειμμάτων και Βελτιστοποίηση Προβλέψεων

Αφού γίνει η αρχική πρόβλεψη, ο XGBoost υπολογίζει τα υπολείμματα (residuals), δηλαδή τη διαφορά μεταξύ πραγματικών και προβλεπόμενων τιμών, ώστε να αξιολογήσει την απόδοσή του. Εκτός από αυτή τη διαφορά, λαμβάνει υπόψη και τη δεύτερη παράγωγο της συνάρτησης απώλειας, η οποία δείχνει πόσο «ασταθές» ή «βέβαιο» είναι το σφάλμα. Με αυτές τις πληροφορίες, προσαρμόζει τις διορθώσεις του πιο αποδοτικά, αποφεύγοντας υπερβολικές ή ανεπαρκείς προσαρμογές και βελτιώνοντας τη συνολική ακρίβεια του μοντέλου.

3. Δημιουργία Διαδοχικών Δέντρων Απόφασης

Στη συνέχεια, ο XGBoost κατασκευάζει διαδοχικά δέντρα απόφασης, όπου κάθε νέο δέντρο μαθαίνει από τα λάθη των προηγούμενων, βελτιώνοντας την ακρίβεια των προβλέψεων. Για να εξασφαλίσει η αποδοτική εκπαίδευση, χρησιμοποιεί:

- **Gradient Boosting Optimization:** κάθε νέο δέντρο μαθαίνει και διορθώνει τα λάθη των προηγούμενων δέντρων.
- **Τεχνική Pruning:** τα μη χρήσιμα κλαδιά των δέντρων απομακρύνονται για να αποφευχθεί η υπερπροσαρμογή (overfitting).
- **Ρύθμιση εκμάθησης:** : χρησιμοποιείται ο συντελεστής learning rate για την ρύθμιση της επιρροής κάθε νέου δέντρου στις τελικές προβλέψεις.

4. Συνδυασμός Προβλέψεων και Ρύθμιση του Learning Rate

Στη συνέχεια, η τελική πρόβλεψη προκύπτει από τον συνδυασμό όλων των δέντρων, όπου κάθε νέο δέντρο συνεισφέρει στη βελτίωση του αποτελέσματος. Ο XGBoost χρησιμοποιεί τον συντελεστή εκμάθησης (learning rate) για να ελέγξει τη συμβολή

κάθε νέου δέντρου, εξασφαλίζοντας σταδιακή και σταθερή προσαρμογή του μοντέλου.

Έτσι το μοντέλο αποφεύγει μεγάλες και απότομες αλλαγές στις προβλέψεις, επιτρέποντας πιο σταθερή και αποδοτική εκπαίδευση.

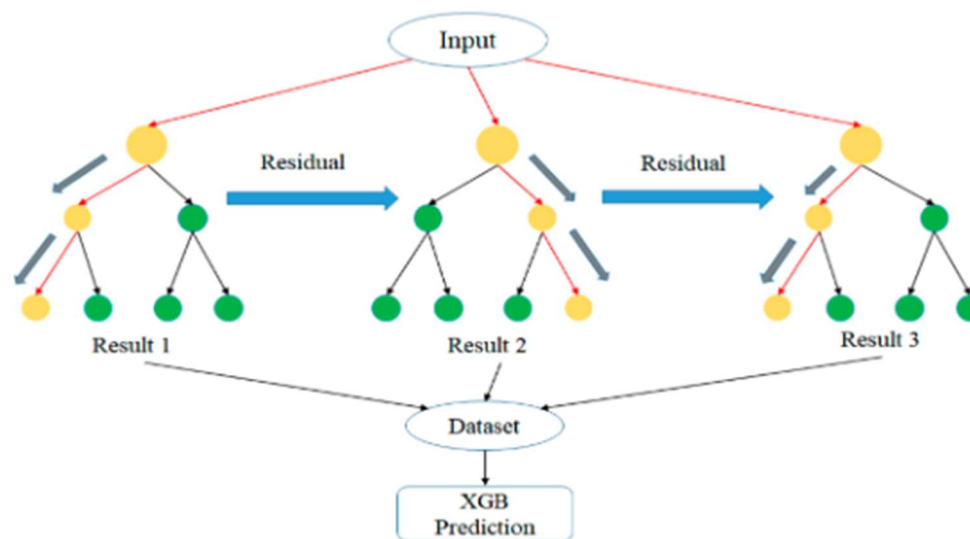
5. Επαναληπτική Διαδικασία και Early Stopping

Τέλος η διαδικασία αυτή επαναλαμβάνεται μέχρι :

- Ένα ικανοποιητικό επίπεδο ασφαλείας, ή
- Τον προκαθορισμένο αριθμός των δέντρων, ή
- Δεν παρατηρείται βελτίωση του σφάλματος στο validation test για συγκεκριμένες επαναλήψεις (early stopping).

Η τεχνική early stopping επιτρέπει στον XGBoost να διακόψει την εκπαίδευση όταν η ακρίβεια του μοντέλου δεν βελτιώνεται άλλο, αποτρέποντας την υπερπροσαρμογή.

Στην παρακάτω εικόνα (Εικόνα 13) παρουσιάζει γραφικά την διαδικασία εκπαίδευσης του XGBoost :



Εικόνα 13.Τρόπος Λειτουργίας του XGBoost

4.2.2 Ρύθμιση Υπερπαράμετρων του XGBoost

Οι υπερπαράμετροι είναι ρυθμίσεις που καθορίζονται πριν από την εκπαίδευση ενός μοντέλου και επηρεάζουν τη διαδικασία μάθησης. Σε αντίθεση με τις παραμέτρους, που προσαρμόζονται αυτόματα κατά την εκπαίδευση, οι υπερπαράμετροι πρέπει να επιλεγούν και να βελτιστοποιηθούν χειροκίνητα για να επιτευχθεί η καλύτερη δυνατή απόδοση [24].

Παρακάτω περιγράφονται οι βασικές υπερπαράμετροι που χρησιμοποιήθηκαν:

1. **Max_depth**

Καθορίζει το μέγιστο βάθος των δέντρων. Ένα μεγαλύτερο βάθος επιτρέπει στο μοντέλο να μάθει πιο περίπλοκα μοτίβα, αλλά αυξάνει τον κίνδυνο υπερπροσαρμογής (overfitting). Αντίθετα, ένα μικρότερο βάθος μπορεί να οδηγήσει σε underfitting. Στη συγκεκριμένη εφαρμογή, το βάθος έχει οριστεί στο 8, εξισορροπώντας την πολυπλοκότητα και τη γενίκευση.

2. **eta (learning_rate)**

Ελέγχει το μέγεθος του βήματος για την ενημέρωση των βαρών των δέντρων. Χαμηλότερη τιμή (π.χ., 0.05) μειώνει τον ρυθμό μάθησης, αυξάνοντας την ακρίβεια αλλά και τον χρόνο εκπαίδευσης.

3. **Subsample**

Καθορίζει το ποσοστό των δεδομένων εκπαίδευσης που χρησιμοποιούνται σε κάθε επανάληψη. Τιμή 0.9 σημαίνει ότι χρησιμοποιείται το 90% των δειγμάτων, μειώνοντας το overfitting και διατηρώντας την απόδοση.

4. **Colsample_bytree**

Ελέγχει το ποσοστό των χαρακτηριστικών που χρησιμοποιούνται τυχαία σε κάθε δέντρο. Με τιμή 0.7, το μοντέλο επιλέγει το 70% των χαρακτηριστικών, αποφεύγοντας την εξάρτηση από συγκεκριμένα.

5. **Lambda**

Αντιστοιχεί στην **L2 κανονικοποίηση**, μια τεχνική που μειώνει το overfitting, προσθέτοντας στη συνάρτηση απώλειας μια ποινή βασισμένη στο άθροισμα των τετραγώνων των βαρών, περιορίζοντας έτσι τα μεγάλα βάρη. Η τιμή 1 εξασφαλίζει σταθερότητα και γενίκευση στο μοντέλο.

6. **Alpha**

Αντιπροσωπεύει την **L1 κανονικοποίηση**, μια τεχνική που προσθέτει στη συνάρτηση απώλειας μια ποινή βασισμένη στο άθροισμα των απόλυτων τιμών των βαρών, ενισχύοντας την επιλογή χαρακτηριστικών. Με τιμή 0.5, βοηθά το μοντέλο να επικεντρωθεί στα πιο σημαντικά χαρακτηριστικά.

7. **Seed**

Ελέγχει την τυχαιότητα στο μοντέλο για αναπαραγωγή των αποτελεσμάτων. Τιμή 42 διασφαλίζει ότι τα αποτελέσματα είναι σταθερά και συγκρίσιμα.

Αυτές οι υπερπαράμετροι επιλέχθηκαν προσεκτικά, ώστε το μοντέλο να ισορροπεί μεταξύ ακρίβειας και αποδοτικότητας, παρέχοντας αξιόπιστες προβλέψεις για τα δεδομένα SCADA.

```

params = {
    "objective": "reg:squarederror",
    "eval_metric": "rmse",
    "max_depth": 8,
    "eta": 0.05,
    "subsample": 0.9,
    "colsample_bytree": 0.7,
    "lambda": 1,
    "alpha": 0.5,
    "seed": 42,
}

```

Εικόνα 14. Βασικές υπερπαραμέτροι του XGBoost που χρησιμοποιούνται στην υλοποίηση του κώδικα

4.2.2.1 Σημασία της Σωστής Ρύθμισης των Υπερπαραμέτρων

Η σωστή ρύθμιση των υπερπαραμέτρων είναι κρίσιμη για την απόδοση του XGBoost, καθώς επηρεάζει τη γενίκευση, την ακρίβεια, τη σταθερότητα και την ταχύτητα εκπαίδευσης. Οι βασικές επιπτώσεις περιλαμβάνουν [24,25,26]:

1. Γενίκευση του Μοντέλου

Η σωστή ρύθμιση εξασφαλίζει ισορροπία μεταξύ overfitting και underfitting. Υπερβολικά μεγάλα βάθη δέντρων μπορεί να οδηγήσουν σε overfitting, ενώ πολύ μικρά σε underfitting. Ρυθμίσεις όπως το subsample και το colsample_bytree μειώνουν τον κίνδυνο overfitting, βελτιώνοντας τη γενίκευση.

2. Ακρίβεια και Σταθερότητα

Χαμηλότερες τιμές στο eta (learning rate) βελτιώνουν τη σταθερότητα και τη γενίκευση. Η L1 και L2 κανονικοποίηση (από τις παραμέτρους alpha και lambda) περιορίζουν τα μεγάλα βάρη και μειώνουν τα σφάλματα από υπερπροσαρμογή.

3. Ταχύτητα Εκπαίδευσης

Το learning rate επηρεάζει την ταχύτητα εκπαίδευσης, με υψηλότερες τιμές να την επιταχύνουν αλλά να προκαλούν αστάθεια. Χρησιμοποιώντας παραμέτρους όπως το subsample και το colsample_bytree, μειώνεται ο υπολογιστικός φόρτος και βελτιώνεται η αποδοτικότητα χωρίς να μειώνεται η απόδοση.

4.2.3 Εφαρμογή του XGBoost σε δεδομένα SCADA

Η εφαρμογή του XGBoost στα δεδομένα SCADA προσφέρει ένα ισχυρό εργαλείο για την πρόβλεψη σφαλμάτων και την προγνωστική συντήρηση των ανεμογεννητριών. Μέσω της ανάλυσης μεγάλων δεδομένων και της ικανότητάς του να εντοπίζει αποκλίσεις από τη φυσιολογική λειτουργία, ο XGBoost συμβάλλει στην έγκαιρη

διάγνωση αστοχιών, μειώνοντας το κόστος συντήρησης και αυξάνοντας την αποδοτικότητα των συστημάτων.

1. Ανίχνευση Σφαλμάτων και Προγνωστική Συντήρηση

Ο XGBoost χρησιμοποιείται για την ανίχνευση δυσλειτουργιών, εκπαιδεύοντας το μοντέλο να αναγνωρίζει αποκλίσεις από τη φυσιολογική λειτουργία των ανεμογεννητριών. Στηριζόμενος σε ιστορικά δεδομένα SCADA, μπορεί να εντοπίσει πρώιμα σημάδια αστοχίας σε εξαρτήματα όπως το σύστημα μετάδοσης και η γεννήτρια [27,28,31].

Μια τυπική μεθοδολογία περιλαμβάνει :

- Εκπαίδευση του XGBoost σε δεδομένα ανεμογεννητριών υπό κανονική λειτουργία.
- Σύγκριση προβλεπόμενων και πραγματικών τιμών για την αναγνώριση αποκλίσεων.
- Αξιοποίηση της προβλεπτικής του ικανότητας για την ανίχνευση ανωμαλιών πριν εξελιχθούν σε βλάβες.

2. Πλεονεκτήματα του XGBoost στην ανάλυση δεδομένων SCADA

Η χρήση του XGBoost προσφέρει σημαντικά πλεονεκτήματα σε σύγκριση με άλλους αλγορίθμους μηχανικής μάθησης, καθιστώντας τον ιδανική επιλογή για την ανίχνευση ανωμαλιών και την προγνωστική συντήρηση των ανεμογεννητριών. Συγκεκριμένα προσφέρει :[27,28,31]

- **Υψηλή Ακρίβεια Πρόβλεψης** : Ο XGBoost αποδίδει καλύτερα σε σχέση με αλγορίθμους όπως τα νευρωνικά δίκτυα και το Random Forest.
- **Γρήγορη επεξεργασία μεγάλων δεδομένων** : Χάρη στην παράλληλη επεξεργασία, ο XGBoost μπορεί να αναλύσει μεγάλες ποσότητες δεδομένων SCADA ταχύτερα από άλλες τεχνικές, εξασφαλίζοντας ταχύτητα και αξιόπιστα αποτελέσματα.
- **Ερμηνευσιμότητα των προβλέψεων**: Προσφέρει ανάλυση της σημασίας των χαρακτηριστικών (Feature Importance Analysis), βοηθώντας στην αναγνώριση των κρίσιμων παραγόντων που επηρεάζουν την εμφάνιση σφαλμάτων.

4.3 Στατιστική Αξιολόγηση Μοντέλων

Η αξιολόγηση των μοντέλων πρόβλεψης είναι απαραίτητη για τη βελτίωση της ακρίβειας των προβλέψεων και την αξιοπιστία των αποτελεσμάτων. Η στατιστική αξιολόγηση είναι καθοριστική στη μηχανική μάθηση, καθώς επιτρέπει:

- Την ποσοτική εκτίμηση της απόδοσης του μοντέλου, διασφαλίζοντας ότι οι προβλέψεις είναι όσο το δυνατόν πιο ακριβείς.
- Την ανίχνευση και μείωση πιθανών σφαλμάτων, τα οποία μπορεί να οδηγήσουν σε λανθασμένες εκτιμήσεις.
- Τη γενίκευση των αποτελεσμάτων, ώστε το μοντέλο να έχει καλή απόδοση και σε νέα δεδομένα.

4.3.1 Στατιστικοί Δείκτες

Η αξιολόγηση ενός μοντέλου πρόβλεψης απαιτεί τη χρήση στατιστικών μετρικών, οι οποίες ποσοτικοποιούν το σφάλμα μεταξύ των προβλεπόμενων και των πραγματικών τιμών. Στην παρούσα μελέτη, η απόδοση του XGBoost αξιολογείται μέσω των ακόλουθων δεικτών [29]:

1. Root Mean Squared Error - RMSE

Το Ριζικό Μέσο Τετραγωνικό Σφάλμα (RMSE) δείχνει το μέσο σφάλμα στις προβλέψεις, λαμβάνοντας υπόψη τη ρίζα του μέσου τετραγωνικού σφάλματος. Δίνει μεγαλύτερη έμφαση στα μεγαλύτερα σφάλματα, γεγονός που το καθιστά χρήσιμο σε εφαρμογές όπου τα μεγάλα σφάλματα είναι πιο κρίσιμα από τα μικρά. Μια χαμηλή τιμή RMSE σημαίνει ότι οι προβλέψεις του μοντέλου είναι κοντά στις πραγματικές τιμές και άρα το μοντέλο έχει υψηλή ακρίβεια. Ο τύπος του είναι [29]:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

Οπού y_i είναι η πραγματική τιμή του μετρούμενου μεγέθους, $\hat{y}_i$ είναι η προβλεπόμενη τιμή του και n το πλήθος των δεδομένων.

2. Mean Squared Error - MSE

Το Μέσο Τετραγωνικό Σφάλμα (MSE) υπολογίζει τον μέσο όρο των τετραγώνων των αποκλίσεων μεταξύ των προβλεπόμενων και πραγματικών τιμών. Δίνει επίσης μεγαλύτερη βαρύτητα στα μεγαλύτερα σφάλματα, αλλά η μονάδα μέτρησής του δεν είναι η ίδια με τις πραγματικές τιμές, γεγονός που καθιστά λιγότερο διαισθητική την ερμηνεία του σε σχέση με το RMSE. Παρόλα αυτά, χρησιμοποιείται ευρέως λόγω της μαθηματικής του ευκολίας και της ικανότητάς του να τιμωρεί τις μεγάλες αποκλίσεις. Ο τύπος του είναι [29]:

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

3. Mean Absolute Error - MAE

Το Μέσο Απόλυτο Σφάλμα (MAE) εκφράζει τη μέση απόλυτη απόκλιση μεταξύ των προβλεπόμενων και πραγματικών τιμών Σε αντίθεση με το MSE και το RMSE, το MAE αντιμετωπίζει όλα τα σφάλματα με ίση βαρύτητα, χωρίς να δίνει δυσανάλογη έμφαση στις μεγαλύτερες αποκλίσεις. Το MAE είναι εύκολο στην ερμηνεία, καθώς οι μονάδες μέτρησής του είναι οι ίδιες με τις πραγματικές τιμές, γεγονός που το καθιστά κατάλληλο για εφαρμογές όπου τα σφάλματα έχουν ισότιμη σημασία ανεξαρτήτως μεγέθους [29].

$$MAE = \frac{1}{m} \sum_{i=1}^m |y_i - \hat{y}_i|$$

4. Συντελεστής Προσδιορισμού R^2

Ο συντελεστής προσδιορισμού (R^2) δείχνει πόσο καλά το μοντέλο προβλέπει τις πραγματικές τιμές. Παίρνει τιμές από 0 έως 1: όσο πιο κοντά στο 1, τόσο καλύτερα το μοντέλο εξηγεί τα δεδομένα. Αν είναι κοντά στο 0, σημαίνει ότι το μοντέλο δεν προβλέπει σωστά και χρειάζεται βελτίωση. Δίνεται από τον παρακάτω τύπο [29]:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

Ο δείκτης R^2 είναι ιδιαίτερα χρήσιμος στην ανάλυση των μοντέλων πρόβλεψης, καθώς επιτρέπει τη σύγκριση της απόδοσης διαφορετικών αλγορίθμων. Ένα υψηλό R^2 υποδεικνύει ότι το μοντέλο έχει ισχυρή προγνωστική ικανότητα, ενώ ένα χαμηλό R^2 υποδηλώνει ότι απαιτούνται βελτιώσεις, είτε μέσω επιλογής καλύτερων χαρακτηριστικών είτε μέσω διαφορετικών αλγορίθμων εκπαίδευσης.

Οι παραπάνω δείκτες επιλέχθηκαν για να διασφαλιστεί ότι το μοντέλο αξιολογείται αντικειμενικά, λαμβάνοντας υπόψη τόσο τη συνολική ακρίβεια όσο και τη σημασία των μεγάλων αποκλίσεων στις προβλέψεις.

4.4 Διαγράμματα SPC

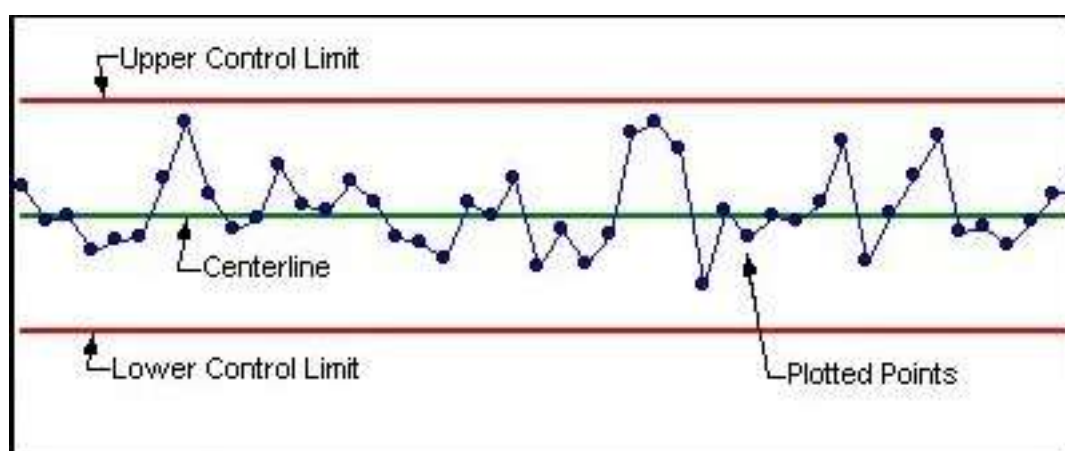
Τα διαγράμματα SPC (*Statistical Process Control*) είναι ένα ισχυρό εργαλείο για την παρακολούθηση της απόδοσης των μοντέλων πρόβλεψης και την ανίχνευση συστηματικών αποκλίσεων. Στην παρούσα μελέτη, τα SPC χρησιμοποιούνται για την ανίχνευση σφαλμάτων μέσω των αποκλίσεων, αναλύοντας πώς μεταβάλλεται η διαφορά μεταξύ των πραγματικών και των προβλεπόμενων τιμών με την πάροδο του χρόνου. Μέσω αυτής της προσέγγισης, παρέχουν πολύτιμες πληροφορίες για τη σταθερότητα και την αξιοπιστία του XGBoost, επιτρέποντας την έγκαιρη αναγνώριση πιθανών δυσλειτουργιών στο σύστημα.

4.4.1 Θεωρητική Προσέγγιση των SPC

Τα διαγράμματα SPC αποτελούν μια στατιστική μέθοδο που χρησιμοποιείται για την παρακολούθηση και τον έλεγχο της σταθερότητας σε βάθος χρόνου. Στόχος τους είναι να διακρίνουν αν οι παρατηρούμενες μεταβολές είναι μέρος της φυσιολογικής διακύμανσης ή αν υποδηλώνουν μη φυσιολογικές αποκλίσεις, οι οποίες θα σημαίνουν πιθανό σφάλμα και απαιτούν περαιτέρω διερεύνηση [30].

Τα SPC διαγράμματα είναι ισχυρά εργαλεία για την παρακολούθηση της απόδοσης μιας διαδικασίας, παρουσιάζοντας τη μεταβολή των τιμών με την πάροδο του χρόνου. Αποτελούνται από τρία βασικά στοιχεία: Τα σημεία δεδομένων, τα οποία αναπαριστούν τις μετρούμενες τιμές και βοηθούν στην ανίχνευση ακραίων τιμών ή σφαλμάτων. Η κεντρική γραμμή (CL), που αντιπροσωπεύει τον μέσο όρο της διαδικασίας, χρησιμεύει ως σημείο αναφοράς για την ανάλυση. Τα όρια ελέγχου (UCL και LCL) καθορίζουν τα αποδεκτά όρια διακύμανσης, επισημαίνοντας όταν οι τιμές ξεπερνούν τα φυσιολογικά όρια.

Στο παρακάτω σχήμα (Εικόνα 15) παρουσιάζεται ένα τυπικό παράδειγμα SPC όπου απεικονίζεται η μεταβολή τιμών σε βάθος χρόνου, η κεντρική γραμμή που αντιπροσωπεύει τον μέσο όρο της διαδικασίας, καθώς και τα άνω και κάτω όρια ελέγχου (UCL και LCL) για την ανίχνευση μη φυσιολογικών αποκλίσεων.



Εικόνα 15.Τυπικό Διάγραμμα SPC

4.4.2 Κατασκευή SPC Διαγραμμάτων

Στην παρούσα διπλωματική εργασία, τα διαγράμματα SPC χρησιμοποιούνται για την παρακολούθηση της διαφοράς (Δ) μεταξύ πραγματικών και προβλεπόμενων τιμών του μοντέλου XGBoost. Ο στόχος είναι η ανίχνευση ανωμαλιών και η έγκαιρη διάγνωση πιθανών σφαλμάτων.

Για την κατασκευή των SPC διαγραμμάτων, ακολουθούνται οι εξής υπολογισμοί [31]:

1. Διαφορά Δ (Measured - Predicted Values)

Η βασική μεταβλητή που αναλύεται στο SPC είναι η διαφορά μεταξύ των μετρούμενων τιμών (Measured Values) και των προβλεπόμενων τιμών (Predicted Values):

$$\Delta_i = Y_i - \hat{Y}_i$$

Όπου : Y_i είναι η πραγματική τιμή της θερμοκρασίας του κιβωτίου ταχυτήτων ή της γεννήτριας, $\hat{Y}_i$ η τιμή που προβλέπει το μοντέλο XGBoost και το Δ_i εκφράζει την απόκλιση του μοντέλου από την πραγματικότητα.

2. Υπολογισμός του Εύρους Μεταβολής (Moving Range - MR)

Για να εκτιμηθεί η διακύμανση των αποκλίσεων (Δ), υπολογίζεται το εύρος μεταβολής (MR) μεταξύ διαδοχικών τιμών:

$$MR_i = \Delta_i - \Delta_{i-1}$$

Το MR υπολογίζει πόσο αλλάζει η διαφορά Δ από χρονική στιγμή σε χρονική στιγμή, επιτρέποντας την ανίχνευση ακραίων διακυμάνσεων.

3. Υπολογισμός της Τυπικής Απόκλισης (σ)

Η τυπική απόκλιση (σ) υπολογίζεται με βάση το μέσο εύρος μεταβολής (MR), χρησιμοποιώντας τον στατιστικό συντελεστή 1.128 :

$$\sigma = \frac{\overline{MR}}{1.128}$$

Όπου : $\overline{MR}$ είναι ο μέσος όρος όλων των MR_i και ο συντελεστής **1.128** προέρχεται από στατιστικές μελέτες και χρησιμοποιείται για την εκτίμηση της τυπικής απόκλισης όταν το εύρος μεταβολής χρησιμοποιείται ως μέτρο διασποράς.

4. Υπολογισμός των Ορίων Ελέγχου (Control Limits - UCL & LCL)

Τα όρια ελέγχου (UCL & LCL) καθορίζονται με βάση τον μέσο όρο των διαφορών Δ και την τυπική απόκλιση σ :

$$UCL = \mu + 3\sigma$$

$$LCL = \mu - 3\sigma$$

Όπου :

- $\mu=0$, καθώς θεωρούμε ότι το μέσο σφάλμα του μοντέλου σε φυσιολογικές συνθήκες πρέπει να είναι μηδενικό.
- Τα όρια ελέγχου ($\pm 3\sigma$) καθορίζουν το εύρος των φυσιολογικών διακυμάνσεων.
- Όταν τιμές Δ βρίσκονται εκτός των ορίων ελέγχου, αυτό σημαίνει ότι το μοντέλο παρουσιάζει αποκλίσεις που απαιτούν διερεύνηση.

4.4.3 Συνδυασμός των Στατιστικών Δεικτών με τα SPC Διαγράμματα

Η ανάλυση της απόδοσης του XGBoost απαιτεί τόσο ποσοτική μέτρηση της ακρίβειας όσο και δυναμική παρακολούθηση των αποκλίσεων σε βάθος χρόνου. Οι στατιστικοί δείκτες (MSE, RMSE, MAE) αποτυπώνουν συνολικά το σφάλμα του μοντέλου, παρέχοντας αντικειμενική εκτίμηση της ακρίβειάς του. Ωστόσο, δεν αποκαλύπτουν ενδεχόμενες τάσεις ή περιόδους με συστηματικές αποκλίσεις.

Τα διαγράμματα SPC συμπληρώνουν αυτή την ανάλυση, καταγράφοντας τη διακύμανση της διαφοράς μεταξύ πραγματικών και προβλεπόμενων τιμών σε διαδοχικές χρονικές στιγμές. Η οπτική απεικόνιση επιτρέπει την ανίχνευση ακραίων τιμών, καθώς και την αξιολόγηση της σταθερότητας του μοντέλου.

Ο συνδυασμός των δύο εργαλείων παρέχει μια ολοκληρωμένη προσέγγιση, επιτρέποντας τόσο την εκτίμηση της συνολικής ακρίβειας όσο και την έγκαιρη αναγνώριση ενδεχόμενων μοτίβων απόκλισης που μπορεί να επηρεάσουν την αξιοπιστία των προβλέψεων.

ΚΕΦΑΛΑΙΟ 5 : Μεθοδολογία

Η μεθοδολογία που ακολουθείται στην παρούσα διπλωματική εργασία βασίζεται στην ανάλυση δεδομένων SCADA, τα οποία συλλέχθηκαν από ανεμογεννήτριες για χρονικό διάστημα δυο χρόνων. Οι μετρήσεις πραγματοποιήθηκαν σε διαστήματα των δέκα λεπτών, παρέχοντας έτσι ένα λεπτομερές σύνολο δεδομένων κατάλληλο για ανάλυση και επεξεργασία.

Στόχος της μεθοδολογίας είναι η ανάπτυξη ενός μοντέλου πρόβλεψης ανωμαλιών, με την αξιοποίηση του αλγορίθμου XGBoost. Μέσω της διαδικασίας αυτής επιδιώκεται η έγκαιρη ανίχνευση σφαλμάτων, η βελτίωση της αποδοτικότητας και η εφαρμογή προγνωστικής συντήρησης.

Οι μετρήσεις αυτές περιλαμβάνουν μέγιστες (max), ελάχιστες (min) και μέσες τιμές (average) για κάθε παράμετρο. Για την εκπαίδευση του μοντέλου θα χρησιμοποιηθούν μόνο οι μέσες τιμές (average). Η επιλογή αυτή βασίζεται στο γεγονός ότι αυτές οι τιμές αντιπροσωπεύουν πιο σταθερά και λιγότερο θορυβώδη δεδομένα σε σύγκριση με τις μέγιστες ή ελάχιστες τιμές, οι οποίες ενδέχεται να περιλαμβάνουν ακραίες μετρήσεις λόγω στιγμιαίων διακυμάνσεων ή σφαλμάτων. Συνεπώς η χρήση των μέσω τιμών διασφαλίζει την ακρίβεια και την αξιοπιστία των αποτελεσμάτων.

Ο κώδικας για την υλοποίηση της μεθοδολογίας αναπτύχθηκε σε Python, χρησιμοποιώντας κατάλληλες βιβλιοθήκες για την ανάλυση και επεξεργασία των δεδομένων SCADA. Οι βασικές βιβλιοθήκες και ο κώδικας που χρησιμοποιήθηκε παρουσιάζονται αναλυτικά στο Παράρτημα.

5.1 Εισαγωγή στην Μεθοδολογία

Τα κύρια βήματα της μεθοδολογίας συνοψίζονται ως εξής [31]:

1. Προεπεξεργασία:

- Συλλογή Ιστορικών Δεδομένων SCADA.
- Προεπεξεργασία και Επιλογή Χαρακτηριστικών.
- Καθαρισμός Δεδομένων.

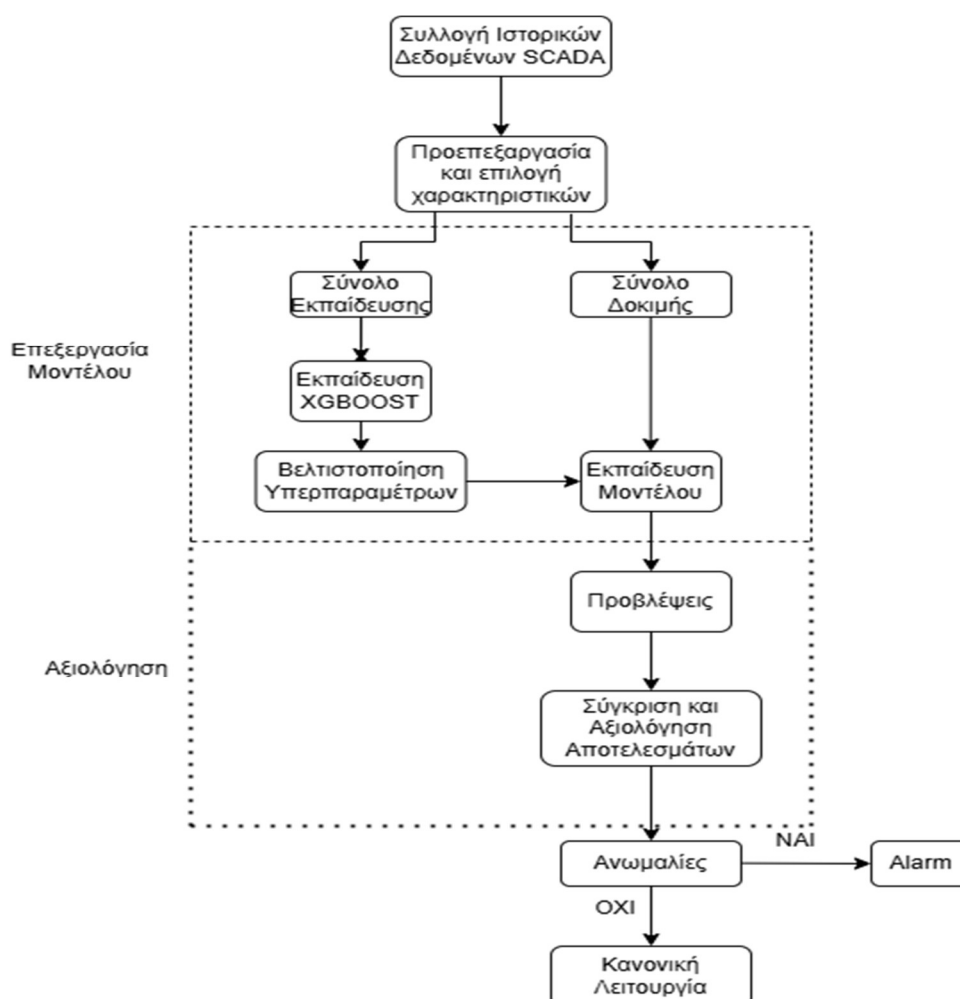
2. Εκπαίδευση Μοντέλου:

- Διαχωρισμός Δεδομένων σε σύνολο εκπαίδευσης και δοκιμής.
- Εκπαίδευση και Βελτιστοποίηση του μοντέλου XGBoost.

3. Μετα-επεξεργασία:

- Προβλέψεις.
- Αξιολόγηση και Ανάλυση Αποτελεσμάτων.
- Ανίχνευση Ανωμαλιών και Λήψη Αποφάσεων.

Το παρακάτω σχήμα (Εικόνα 16) παρουσιάζει τη ροή της μεθοδολογίας που θα ακολουθήσουμε.



Εικόνα 16. Διαδικασία Ανίχνευσης Ανωμαλιών

5.2 Απόκτηση Δεδομένων και Προεπεξεργασία

Η απόκτηση και προεπεξεργασία δεδομένων αποτελεί κρίσιμο στάδιο για την ανάπτυξη ενός αποτελεσματικού μοντέλου. Τα δεδομένα συλλέγονται μέσω συστημάτων SCADA, τα οποία παρέχουν διαδοχικές χρονικά μετρήσεις διαφόρων παραμέτρων των ανεμογεννητριών. Στη συνέχεια εφαρμόζεται καθαρισμός των δεδομένων, αφαιρώντας ακραίες τιμές και τυχόν ελλιπείς μετρήσεις. Αυτή η διαδικασία διασφαλίζει την αξιοπιστία των δεδομένων, βελτιώνοντας την ακρίβεια των προβλέψεων και ενισχύοντας την αποτελεσματικότητα του μοντέλου [31].

5.2.1 Συλλογή και Επιλογή Δεδομένων SCADA

Το πρώτο στάδιο είναι η συλλογή και η επιλογή δεδομένων. Οι επιλεγμένες παράμετροι περιλαμβάνουν:

- **Nacelle Temperature** (Nac_Temp_Avg)
- **Rotor Speed** (Rtr_RPM_Avg)
- **Active Power** (Grd_Prod_Pwr_Avg)
- **Ambient Temperature** (Amb_Temp_Avg)
- **Gearbox Oil Temperature** (Gear_Oil_Temp_Avg)
- **Gearbox Bearing Temperature** (Gear_Bear_Temp_Avg)
- **Generator Bearing** (Gen_Bear_Temp_Avg)
- **Generator Speed** (Gen_RPM_Avg)
- **Generator Stator Temperature** (Gen_Phase1_Temp_Avg, Gen_Phase2_Temp_Avg, Gen_Phase3_Temp_Avg)

Η επιλογή των χαρακτηριστικών (feature selection) αποτελεί θεμελιώδης στάδιο για τη βελτίωση της απόδοσης του μοντέλου πρόβλεψης. Η διαδικασία αυτή επικεντρώνεται στην ταυτοποίηση των πιο κρίσιμων μεταβλητών που επηρεάζουν τη λειτουργία των ανεμογεννητριών και την ακριβέστερη πρόβλεψη σφαλμάτων. Η επιλογή των χαρακτηριστικών πραγματοποιήθηκε βάσει βιβλιογραφικών αναφορών [31], διασφαλίζοντας ότι η προσέγγιση ακολουθεί επιστημονικά τεκμηριωμένες μεθόδους.

5.2.2 Προεπεξεργασία και Καθαρισμός Δεδομένων

Η προεπεξεργασία δεδομένων στοχεύει στη βελτίωση της ποιότητας και της αξιοπιστίας των δεδομένων SCADA που χρησιμοποιούνται για την ανάπτυξη μοντέλων προγνωστικής συντήρησης. Τα δεδομένα επιλέγονται από περιόδους φυσιολογικής λειτουργίας, προκειμένου να διαμορφωθεί ένα αξιόπιστο πρότυπο συμπεριφοράς των ανεμογεννητριών [31].

Η διαδικασία προεπεξεργασίας περιλαμβάνει:

- **Απομάκρυνση ανωμαλιών και θορύβου** : Εξαλείφονται σφάλματα μέτρησης από αισθητήρες, ακραίες τιμές λόγω εξωτερικών παραγόντων και κενά δεδομένων που προκύπτουν από τεχνικές δυσλειτουργίες.
- **Έλεγχος συνοχής και ακρίβειας** : Ελέγχεται η λογική συνοχή των δεδομένων και η συμβατότητά τους με τις φυσιολογικές συνθήκες λειτουργίας των ανεμογεννητριών.
- **Καθαρισμός δεδομένων καμπύλης ισχύος** : Η καμπύλη ισχύος, που χρησιμοποιείται ως δείκτης φυσιολογικής λειτουργίας, ελέγχεται για πιθανές αποκλίσεις, διασφαλίζοντας τη χαρακτηριστική μορφή της.

5.2.2.1 Καμπύλες Ισχύος : Εργαλείο Αξιολόγησης

Οι καμπύλες ισχύος αποτελούν ένα κρίσιμο εργαλείο για την αξιολόγηση της απόδοσης των ανεμογεννητριών, καθώς αποτυπώνουν τη σχέση μεταξύ της ταχύτητας ανέμου και της παραγόμενης ισχύος (Σχήμα 2). Αυτή η σχέση

χρησιμοποιείται ως σημείο αναφοράς για την ανίχνευση αποκλίσεων, διευκολύνοντας την έγκαιρη διάγνωση σφαλμάτων και την πρόβλεψη πιθανών βλαβών .

Η εφαρμογή τεχνικών καθαρισμού δεδομένων, όπως τα quantiles, διασφαλίζει ότι οι καμπύλες ισχύος βασίζονται σε αξιόπιστα δεδομένα. Ο καθαρισμός απομακρύνει ανωμαλίες και θόρυβο, δημιουργώντας καμπύλες που αντικατοπτρίζουν με ακρίβεια τη φυσιολογική λειτουργία, βελτιώνοντας έτσι την αξιοπιστία της ανάλυσης.

Η σύγκριση των δεδομένων ισχύος με την τυπική καμπύλη ισχύος επιτρέπει την ανίχνευση αποκλίσεων που μπορεί να υποδηλώνουν σφάλματα στη λειτουργία της ανεμογεννήτριας. Οι αποκλίσεις αυτές μπορεί να σχετίζονται με βλάβες σε κρίσιμα εξαρτήματα, όπως το κιβώτιο ταχυτήτων ή η γεννήτρια, και να αποτελέσουν προειδοποιητικά σημάδια για επικείμενες δυσλειτουργίες.

5.2.2.2 Καθορισμός Κριτηρίων Καθαρισμού Καμπυλών Ισχύος

Η καμπύλη ισχύος αποτελεί βασικό εργαλείο αξιολόγησης της απόδοσης των ανεμογεννητριών, συνεπώς, η απομάκρυνση δεδομένων που δεν ανταποκρίνονται στη φυσιολογική λειτουργία είναι καθοριστική.

Τα βασικά κριτήρια καθαρισμού περιλαμβάνουν:

- **Μηδενική ή αρνητική ισχύς με παρουσία ανέμου :** Αντικατοπτρίζει πιθανές δυσλειτουργίες ή λάθη αισθητήρων.
- **Ελλιπή δεδομένα :** Εγγραφές με ελλειπείς μεταβλητές, οι οποίες δεν παρέχουν αξιόπιστες πληροφορίες.
- **Τιμές εκτός φυσιολογικού εύρους :** Ενδείξεις ασυνήθιστων συνθηκών ή σφαλμάτων.
- **Παύση λειτουργίας ή απώλεια δεδομένων :** Στιγμές όπου η ανεμογεννήτρια δεν λειτουργούσε, γεγονός που μπορεί να προκαλέσει λανθασμένες εκτιμήσεις.
- **Φιλτράρισμα βάσει της μεταβλητής *AlarmActive* :** Χρησιμοποιούνται αποκλειστικά δεδομένα από περιόδους φυσιολογικής λειτουργίας των ανεμογεννητριών, όπου δεν είχε ενεργοποιηθεί συναγερμός, διασφαλίζοντας ότι η ανεμογεννήτρια λειτουργεί στην φυσιολογική λειτουργία της.

Η εφαρμογή αυτών των κριτηρίων διασφαλίζει τη δημιουργία μιας αξιόπιστης καμπύλης ισχύος, η οποία αντανακλά την κανονική λειτουργία των ανεμογεννητριών και επιτρέπει την ακριβή εκπαίδευση του αλγορίθμου. Παράλληλα, τα ίδια κριτήρια θα εφαρμοστούν και για την εκπαίδευση του μοντέλου πρόβλεψης, ώστε να διασφαλιστεί ότι το σύστημα βασίζεται αποκλειστικά σε δεδομένα φυσιολογικής λειτουργίας, αποφεύγοντας θορυβώδεις ή λανθασμένες μετρήσεις.

5.2.2.3 Χρήση Quantiles για τον Καθαρισμό των Δεδομένων

Η χρήση quantiles αποτελεί μια αποτελεσματική τεχνική για τον καθαρισμό της καμπύλης ισχύος, καθώς επιτρέπει την απομάκρυνση ακραίων τιμών και θορύβου.

Η διαδικασία καθαρισμού με quantiles περιλαμβάνει τα εξής στάδια:

1. Καθορισμός διαστημάτων ταχύτητας ανέμου

- Το εύρος των ταχυτήτων ανέμου (0–25 m/s) χωρίζεται σε μικρότερα διαστήματα (π.χ. με βήμα 0.1 m/s).
- Για κάθε διάστημα, εντοπίζονται οι αντίστοιχες τιμές ισχύος.

2. Υπολογισμός και φιλτράρισμα με quantiles

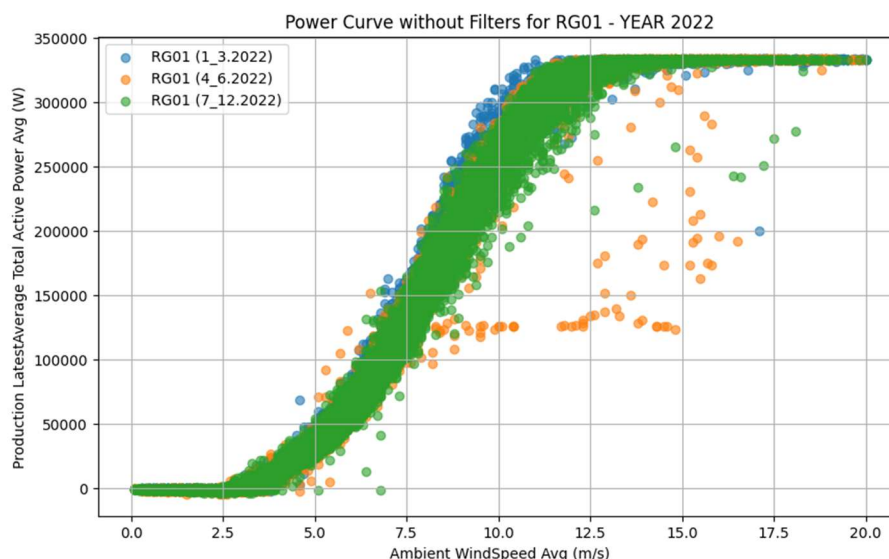
- Για κάθε διάστημα ταχύτητας ανέμου, υπολογίζονται τα 17ο και 97,5ο quantile.
- Τα δεδομένα ισχύος που βρίσκονται εκτός αυτών των ορίων απορρίπτονται ως ακραίες τιμές.

3. Δημιουργία καμπύλης ισχύος

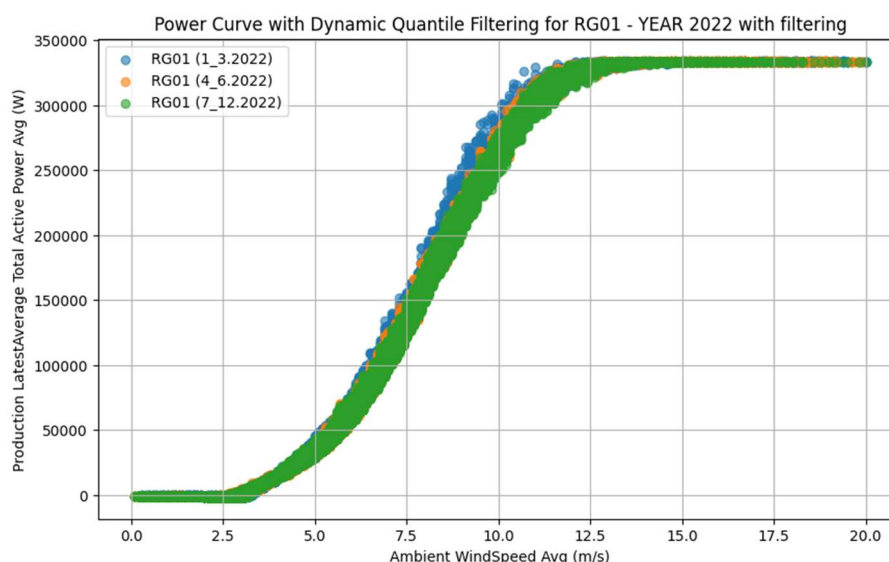
- Μετά την αφαίρεση των ακραίων τιμών, τα φιλτραρισμένα δεδομένα χρησιμοποιούνται για την κατασκευή της τελικής καμπύλης ισχύος.
- Η τελική καμπύλη ισχύος αντικατοπτρίζει με μεγαλύτερη ακρίβεια τη σχέση ταχύτητας ανέμου - παραγόμενης ισχύος.

5.2.2.4 Καμπύλες Ισχύος Πριν και Μετά τον Καθαρισμό

Στην πρώτη εικόνα (Εικόνα 17) παρουσιάζεται η καμπύλη ισχύος πριν την απομάκρυνση ακραίων τιμών. Παρατηρούνται έντονες αποκλίσεις από τη τυπική καμπύλη ισχύος (Εικόνα 2), ιδιαίτερα στις χαμηλές και στις υψηλές ταχύτητες ανέμου. Αυτές οι αποκλίσεις μπορεί να οφείλονται σε ακραίες τιμές των δεδομένων, θόρυβο και σφάλματα αισθητήρων, τα οποία επηρεάζουν την ακρίβεια της καμπύλης και υποβαθμίζουν την αξιοπιστία της ανάλυσης. Αντίθετα στη δεύτερη εικόνα (Εικόνα 18) απεικονίζεται η καμπύλη ισχύος μετά την εφαρμογή των quantiles. Ο θόρυβος και οι ανώμαλες τιμές έχουν απομακρυνθεί, με αποτέλεσμα μια σαφώς πιο ομαλή καμπύλη που αντικατοπτρίζει τη φυσιολογική λειτουργία της ανεμογεννήτριας.



Εικόνα 17. Καμπύλη Ισχύος πριν τον καθαρισμό δεδομένων



Εικόνα 18. Καμπύλη Ισχύος μετά την εφαρμογή των quantiles

5.3 Επεξεργασία Μοντέλων και Εκπαίδευση

Η φάση της επεξεργασίας των μοντέλων επικεντρώνεται στην εκπαίδευση του αλγορίθμου XGBoost, με στόχο την αναπαράσταση της φυσιολογικής λειτουργίας των ανεμογεννητριών και την πρόβλεψη κρίσιμων λειτουργικών παραμέτρων. Συγκεκριμένα, ο XGBoost χρησιμοποιείται για την πρόβλεψη της θερμοκρασίας του κιβωτίου ταχυτήτων (Gearbox Temperature) και της θερμοκρασίας της γεννήτριας (Generator Temperature), δύο παραμέτρων που σχετίζονται με τη συνολική απόδοση των ανεμογεννητριών και την πιθανή εμφάνιση σφαλμάτων.

Η διαδικασία εκπαίδευσης περιλαμβάνει:

- **Επιλογή χαρακτηριστικών (feature selection):** Οι παράμετροι που έχουν ορισθεί σε προηγούμενες ενότητες χρησιμοποιούνται ως είσοδοι/έξοδοι στο μοντέλο XGBoost.
- **Διαχωρισμός δεδομένων:** Τα δεδομένα χωρίζονται σε σύνολο εκπαίδευσης και σύνολο δοκιμής, ώστε να διασφαλιστεί η σωστή γενίκευση του μοντέλου.
- **Εκπαίδευση και βελτιστοποίηση:** Το μοντέλο εκπαιδεύεται με στόχο τη βελτιστοποίηση της ακρίβειας του.

5.3.1 Δεδομένα Εισόδου και Εξόδου

Για την εκπαίδευση και αξιολόγηση του αλγορίθμου XGBoost, επιλέχθηκαν συγκεκριμένες μεταβλητές εισόδου που επηρεάζουν τη θερμοκρασία κρίσιμων εξαρτημάτων της ανεμογεννήτριας. Στόχος είναι η πρόβλεψη της θερμοκρασίας του κιβωτίου ταχυτήτων και της γεννήτριας, όπως παρουσιάζεται στον παρακάτω πίνακα.

Component	Input	Output
Gearbox	Nacelle Temperature Rotor Speed Active Power Ambient Temperature Gearbox Oil Temperature	Gearbox Bearing Temperature
Generator	Nacelle Temperature Active Power Generator Speed Generator Stator Temperature	Generator Bearing Temperature

Πίνακας 1. Μεταβλητές Εισόδου και Έξοδου για το Μοντέλο XGBoost

5.3.1.1 Πίνακες Συσχέτισης (Correlation Matrices)

Τι είναι οι Πίνακες Συσχέτισης

Οι πίνακες συσχέτισης (Correlation Matrices) αποτελούν εργαλείο στατιστικής ανάλυσης που χρησιμοποιείται για τη μέτρηση της γραμμικής σχέσης μεταξύ δύο ή περισσότερων μεταβλητών. Κάθε κελί του πίνακα περιέχει έναν συντελεστή συσχέτισης (correlation coefficient), ο οποίος κυμαίνεται από -1 έως 1:

- **Τιμές κοντά στο +1** υποδηλώνουν ισχυρή θετική συσχέτιση (όταν αυξάνεται η μία μεταβλητή, αυξάνεται και η άλλη).
- **Τιμές κοντά στο -1** υποδηλώνουν ισχυρή αρνητική συσχέτιση (όταν αυξάνεται η μία μεταβλητή, μειώνεται η άλλη).
- **Τιμές κοντά στο 0** υποδεικνύουν απουσία γραμμικής σχέσης μεταξύ των μεταβλητών.

Στην παρούσα μελέτη, οι πίνακες συσχέτισης χρησιμοποιούνται για να εντοπιστούν οι μεταβλητές με τη μεγαλύτερη επιρροή στις εξόδους του μοντέλου XGBoost. Ο εντοπισμός αυτών των σχέσεων επιτρέπει τη βελτιστοποίηση της επιλογής χαρακτηριστικών (feature selection) και τη βελτίωση της απόδοσης του αλγορίθμου πρόβλεψης.

Συσχετίσεις Μεταβλητών Εισόδου-Εξόδου

Στα παρακάτω διαγράμματα (Εικόνες 20-21) παρουσιάζονται οι πίνακες συσχέτισης που αποτυπώνουν τις σχέσεις μεταξύ των μεταβλητών εισόδου και εξόδου για το κιβώτιο ταχυτήτων (Gearbox) και τη γεννήτρια (Generator):

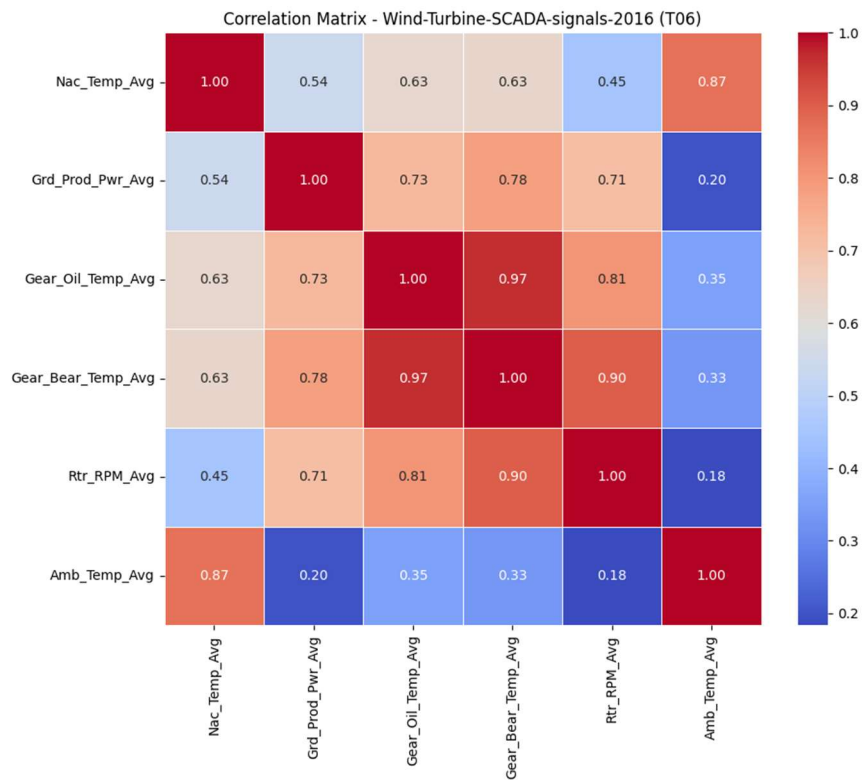
- **Κιβώτιο Ταχυτήτων (Gearbox):**

Παρατηρούνται ισχυρές συσχετίσεις μεταξύ της ενεργού ισχύος (Active Power), της ταχύτητας του ρότορα (Rotor Speed) και της θερμοκρασίας λαδιού κιβωτίου ταχυτήτων (Gearbox Oil Temperature) με τη θερμοκρασία ρουλεμάν του κιβωτίου ταχυτήτων (Gearbox Bearing Temperature).

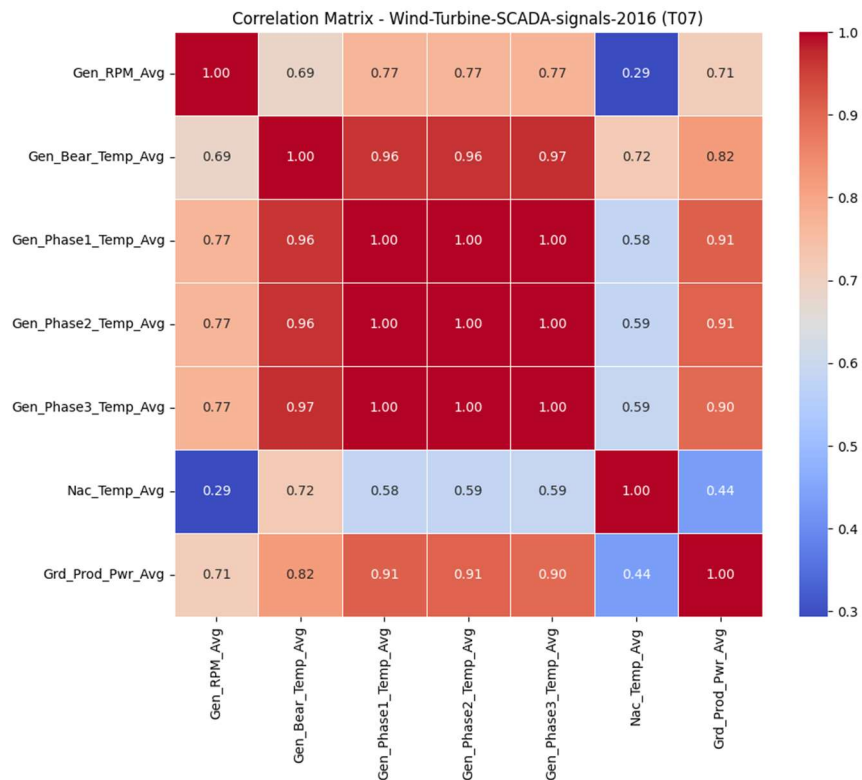
- **Γεννήτρια (Generator):**

Ισχυρές συσχετίσεις εμφανίζονται μεταξύ της ενεργού ισχύος (Active Power), της ταχύτητας της γεννήτριας (Generator RPM) και των θερμοκρασιών των τριών φάσεων της γεννήτριας (Gen_Phase1_Temp_Avg, Gen_Phase2_Temp_Avg, Gen_Phase3_Temp_Avg) με τη θερμοκρασία ρουλεμάν της γεννήτριας (Generator Bearing Temperature).

Παρόλο που η υψηλή συσχέτιση μεταξύ μεταβλητών αποτελεί ένδειξη ισχυρής σχέσης, περιλαμβάνονται και μεταβλητές με χαμηλότερη ή μέτρια συσχέτιση, καθώς μπορεί να περιέχουν κρίσιμες πληροφορίες που συμβάλλουν στη βελτίωση της ακρίβειας του μοντέλου XGBoost όταν συνδυάζονται με άλλες παραμέτρους. Η ανάλυση των συσχετίσεων επιτρέπει τη μείωση της διάστασης του προβλήματος, την εξάλειψη περιττών μεταβλητών και την ενίσχυση της γενίκευσης του μοντέλου.



Εικόνα 19. Διάγραμμα Συσχέτισης του Κιβωτίου Ταχυτήτων (Gearbox)



Εικόνα 20. Διάγραμμα Συσχέτισης της Γεννήτριας (Generator)

5.3.1.2 Κανονικοποίηση Δεδομένων

Η κανονικοποίηση των δεδομένων είναι απαραίτητη όταν τα χαρακτηριστικά έχουν διαφορετικές κλίμακες και μονάδες μέτρησης, καθώς εξασφαλίζει ότι όλα συνεισφέρουν ισότιμα στην εκπαίδευση του μοντέλου. Για παράδειγμα, η ταχύτητα ανέμου (m/s) και η θερμοκρασία (°C) έχουν διαφορετικά εύρη τιμών, γεγονός που μπορεί να προκαλέσει μεροληψία στα αποτελέσματα. Με την κανονικοποίηση, οι τιμές μετατρέπονται στην ίδια κλίμακα, βελτιώνοντας την ακρίβεια και τη σταθερότητα του μοντέλου.

Διαδικασία Κανονικοποίησης

Η κανονικοποίηση των χαρακτηριστικών γίνεται με τη μέθοδο Min-Max Scaling, όπου κάθε τιμή x διαιρείται με τη μέγιστη τιμή x_{max} του αντίστοιχου χαρακτηριστικού:

$$x' = \frac{x}{x_{max}}$$

Αυτό διασφαλίζει ότι όλα τα χαρακτηριστικά βρίσκονται στο εύρος [0,1], αποτρέποντας την υπεροχή χαρακτηριστικών με μεγάλες τιμές κατά την εκπαίδευση του μοντέλου.

5.3.2 Ανάπτυξη και Εκπαίδευση Μοντέλου XGBoost

Η διαδικασία ανάπτυξης του XGBoost περιλαμβάνει τα εξής βήματα:

- **Προετοιμασία δεδομένων:** Καθαρισμός, κανονικοποίηση και επιλογή χαρακτηριστικών, όπως περιγράφηκε αναλυτικά στις προηγούμενες ενότητες, για τη διασφάλιση της ποιότητας και αξιοπιστίας των δεδομένων εισόδου.
- **Διαίρεση δεδομένων:** Τα δεδομένα χωρίζονται σε σύνολα εκπαίδευσης και δοκιμής (training και test sets) για την εκπαίδευση του μοντέλου και την ανεξάρτητη αξιολόγηση της απόδοσής του.

5.3.2.1 Διαχωρισμός Δεδομένων

Ο διαχωρισμός των δεδομένων SCADA γίνεται με τη μέθοδο Holdout Validation, ώστε να διασφαλιστεί η αξιόπιστη αξιολόγηση του μοντέλου. Τα δεδομένα του πρώτου έτους χρησιμοποιούνται για την εκπαίδευση του XGBoost, ενώ του δεύτερου έτους για τη δοκιμή, προσομοιώνοντας ρεαλιστικές συνθήκες πρόβλεψης. Ο διαχωρισμός αποτρέπει το data leakage, διασφαλίζοντας ότι το μοντέλο αξιολογείται σε νέα δεδομένα. Με αυτόν τον τρόπο, ελέγχεται η δυνατότητα γενίκευσης, καθώς το XGBoost καλείται να προβλέψει πραγματικές μελλοντικές συνθήκες λειτουργίας, συμβάλλοντας στην αποτελεσματική προγνωστική συντήρηση των ανεμογεννητριών.

5.4 Μεταεπεξεργασία

Η μεταεπεξεργασία των αποτελεσμάτων αποτελεί το τελικό στάδιο της μεθοδολογίας και περιλαμβάνει την ανάλυση και ερμηνεία των προβλέψεων του αλγορίθμου XGBoost, την αξιολόγηση της απόδοσής του και την ανίχνευση ανωμαλιών στις λειτουργικές παραμέτρους των ανεμογεννητριών.

Οι βασικές διαδικασίες που περιλαμβάνονται στη μεταεπεξεργασία είναι:

- **Αξιολόγηση Μοντέλου με Στατιστικούς Δείκτες:** Μέτρηση της ακρίβειας του μοντέλου με δείκτες όπως MSE, MAE, RMSE, R^2 .
- **Σύγκριση Προβλέψεων και Πραγματικών Τιμών:** Γραφική απεικόνιση των προβλεπόμενων και πραγματικών τιμών για την ανίχνευση συστηματικών αποκλίσεων.
- **Ανίχνευση Ανωμαλιών (Anomaly Detection):** Ταυτοποίηση σημαντικών αποκλίσεων στις λειτουργικές παραμέτρους των ανεμογεννητριών, που ενδέχεται να υποδηλώνουν επικείμενες βλάβες.

Η μεταεπεξεργασία συμβάλλει στην κατανόηση της συμπεριφοράς του μοντέλου και επιτρέπει τη βελτιστοποίηση των προβλέψεων, με στόχο τη δημιουργία ενός αξιόπιστου συστήματος προγνωστικής συντήρησης.

5.4.1 Αξιολόγηση Μοντέλου με Στατιστικούς Δείκτες

Η απόδοση του XGBoost αξιολογείται μέσω στατιστικών δεικτών που ποσοτικοποιούν την ακρίβεια των προβλέψεων. Αυτοί οι δείκτες εφαρμόζονται τόσο στο training όσο και στο test set, εκτιμώντας τη συνολική αποτελεσματικότητα και τη σταθερότητα του μοντέλου στην πρόβλεψη των θερμοκρασιών του κιβωτίου ταχυτήτων και της γεννήτριας.

Όσο μικρότερες είναι οι τιμές αυτών των δεικτών, τόσο ακριβέστερο θεωρείται το μοντέλο.

5.4.2 Ανίχνευση Ανωμαλιών (Anomaly Detection)

Η ανίχνευση ανωμαλιών αποτελεί βασικό στάδιο της μεταεπεξεργασίας των αποτελεσμάτων, καθώς επιτρέπει την αναγνώριση σημαντικών αποκλίσεων στις λειτουργικές παραμέτρους των ανεμογεννητριών. Αυτές οι αποκλίσεις μπορεί να σηματοδοτούν επικείμενες βλάβες ή να υποδεικνύουν την ανάγκη για προληπτική συντήρηση.

Για την ανίχνευση ανωμαλιών, χρησιμοποιούνται διαγράμματα SPC, τα οποία απεικονίζουν τη διαφορά μεταξύ των προβλεπόμενων και πραγματικών τιμών. Με βάση αυτά, μια πιθανή δυσλειτουργία εντοπίζεται όταν:

- Τουλάχιστον έξι διαδοχικές τιμές υπερβαίνουν τα προκαθορισμένα όρια ελέγχου.
- Η συνεχής απόκλιση δείχνει μια συστηματική τάση και όχι μια τυχαία διακύμανση.

Αυτές οι αποκλίσεις επιτρέπουν την έγκαιρη διάγνωση και παρέμβαση, συμβάλλοντας στην αποτροπή σοβαρών δυσλειτουργιών και στη βελτιστοποίηση του συστήματος προγνωστικής συντήρησης.

ΚΕΦΑΛΑΙΟ 6 : Αποτελέσματα και Συμπεράσματα

Το παρόν κεφάλαιο εστιάζει στην παρουσίαση και ανάλυση των αποτελεσμάτων που προέκυψαν από την εφαρμογή της μεθοδολογίας που περιγράφηκε στο Κεφάλαιο 5. Στόχος είναι η αξιολόγηση της ικανότητας του μοντέλου να προβλέπει τη θερμοκρασία της γεννήτριας και του κιβωτίου ταχυτήτων, καθώς και να εντοπίζει αποκλίσεις που ενδέχεται να υποδηλώνουν πιθανά σφάλματα στη λειτουργία των ανεμογεννητριών.

Η ανάλυση πραγματοποιείται σε δύο διαφορετικές περιπτώσεις:

1. Τέσσερις ανεμογεννήτριες χωρίς ιστορικό βλαβών

Σε αυτή την περίπτωση, το μοντέλο εκπαιδεύεται και εφαρμόζεται σε δεδομένα ανεμογεννητριών που δεν έχουν καταγεγραμμένο ιστορικό σφαλμάτων. Ο στόχος είναι να αξιολογηθούν οι προβλέψεις και να διερευνηθεί εάν το μοντέλο μπορεί να ανιχνεύσει ανωμαλίες που ενδέχεται να σχετίζονται με πρώιμα σημάδια φθοράς ή δυσλειτουργίας. Τα δεδομένα προήλθαν από τη ΔΕΗ Ανανεώσιμες και αφορούν ένα αιολικό πάρκο, παρέχοντας ένα πραγματικό περιβάλλον εφαρμογής για την αξιολόγηση της μεθοδολογίας.

2. Δύο ανεμογεννήτριες με ιστορικό βλαβών

Η δεύτερη περίπτωση αφορά ανεμογεννήτριες που έχουν καταγεγραμμένες βλάβες στο παρελθόν. Η ανάλυση εξετάζει αν οι αποκλίσεις στις προβλέψεις του μοντέλου συνδέονται με τα πραγματικά σφάλματα, ώστε να εκτιμηθεί αν η προτεινόμενη μέθοδος μπορεί να χρησιμοποιηθεί για προληπτική συντήρηση. Τα δεδομένα προήλθαν από την πλατφόρμα ανοιχτών δεδομένων της EDP (EDP Open Data), παρέχοντας πραγματικές καταγραφές σφαλμάτων για την αξιολόγηση της αποτελεσματικότητας του μοντέλου.

Η παρουσίαση των αποτελεσμάτων οργανώνεται ξεχωριστά για κάθε περίπτωση, ενώ στο τέλος παρατίθεται συγκριτική ανάλυση και εξάγονται τα βασικά συμπεράσματα της μελέτης.

6.1 Ανάλυση Αποτελεσμάτων για Ανεμογεννήτριες χωρίς Ιστορικό Βλαβών

Στην παρούσα ενότητα παρουσιάζονται τα αποτελέσματα της εφαρμογής του μοντέλου XGBoost στις ανεμογεννήτριες που δεν διαθέτουν καταγεγραμμένο ιστορικό βλαβών. Η ανάλυση περιλαμβάνει την αξιολόγηση της διαδικασίας εκπαίδευσης του μοντέλου, τη σύγκριση των πραγματικών και προβλεπόμενων τιμών, καθώς και την ανίχνευση αποκλίσεων μέσω διαγραμμάτων SPC. Επιπλέον, παρατίθενται στατιστικοί δείκτες για τη συνολική αξιολόγηση της απόδοσης του μοντέλου.

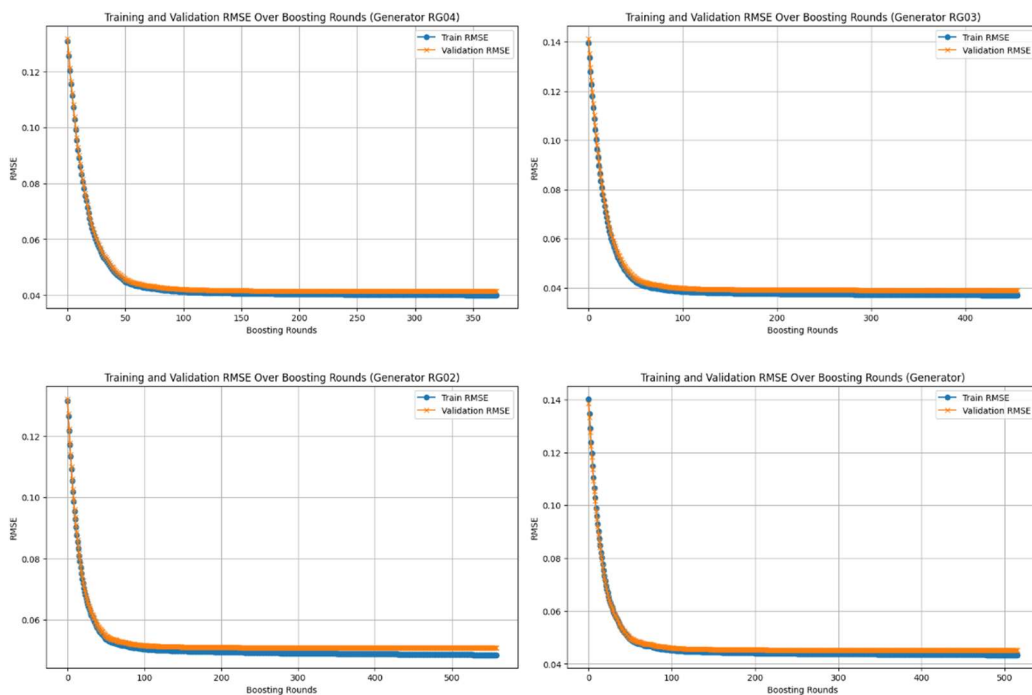
6.1.1 Εξέλιξη του RMSE κατά τη Διαδικασία Εκπαίδευσης του XGBoost

Το RMSE (Root Mean Squared Error) αποτελεί μέτρο αξιολόγησης της απόδοσης του μοντέλου εκφράζοντας μέσο τετραγωνικό σφάλμα σε σχέση με τις πραγματικές τιμές. Όσο μικρότερη είναι η τιμή του RMSE, τόσο μεγαλύτερη είναι η ακρίβεια των προβλέψεων.

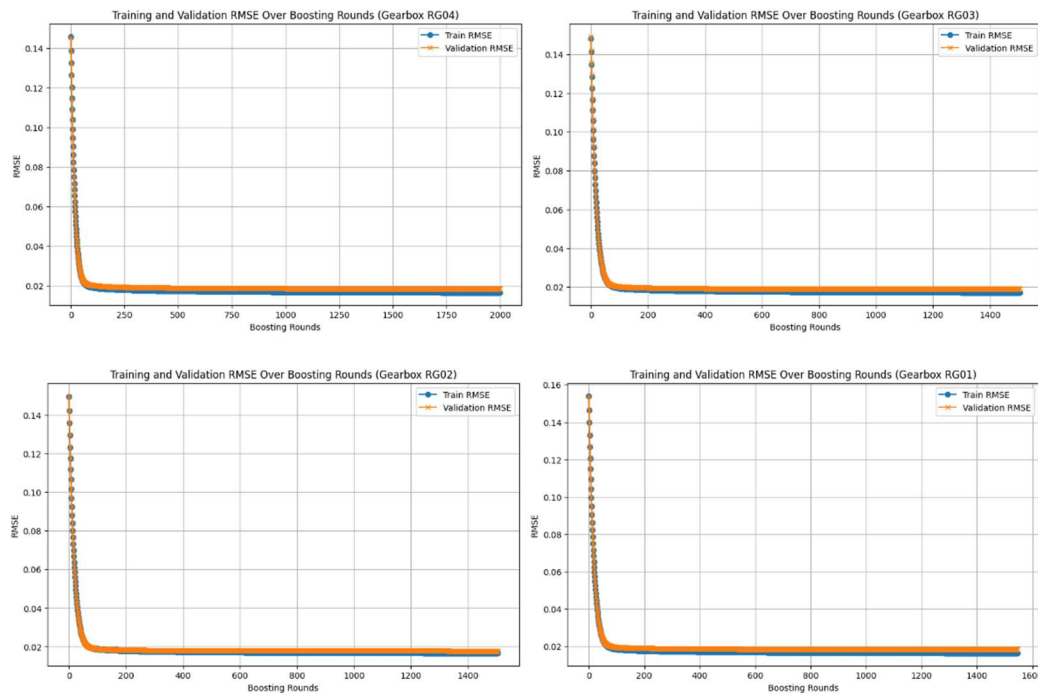
Τα παρακάτω διαγράμματα απεικονίζουν την εξέλιξη του RMSE κατά τη διαδικασία εκπαίδευσης του μοντέλου XGBoost. Τα χαρακτηριστικά του είναι :

- **Οριζόντιος άξονας (Boosting Rounds):** Δείχνει τον αριθμό των επαναλήψεων της εκπαίδευσης.
- **Κατακόρυφος άξονας (RMSE):** Παρουσιάζει το σφάλμα RMSE κατά τη διάρκεια της εκπαίδευσης.
- **Μπλε γραμμή (Train RMSE):** Αντιπροσωπεύει το σφάλμα στα δεδομένα εκπαίδευσης.
- **Πορτοκαλί γραμμή (Validation RMSE):** Αντιπροσωπεύει το σφάλμα στα δεδομένα δοκιμής.

Τα παρακάτω διαγράμματα παρουσιάζουν την σύγκλιση του RMSE των ανεμογεννητριών (RG01-RG04) για την θερμοκρασία της γεννήτρια και για την θερμοκρασία του κιβωτίου ταχυτήτων (Εικόνες 21-22).



Εικόνα 21. Σύγκλιση του RMSE εκπαίδευσης και επικύρωσης για την θερμοκρασία της γεννήτριας (RG01-RG04)



Εικόνα 22. Σύγκλιση του RMSE εκπαίδευσης και επικύρωσης για την θερμοκρασία του κιβωτίου ταχυτήτων (RG01-RG04)

Στον παρακάτω πίνακα (Πίνακας 2) συνοψίζονται οι τελικές κανονικοποιημένες τιμές RMSE για κάθε ανεμογεννήτρια, παρουσιάζοντας το τελευταίο boosting round, καθώς και τις αντίστοιχες τιμές Train RMSE και Validation RMSE. Η ανάλυση αυτών των τιμών επιτρέπει την αξιολόγηση της σύγκλισης του μοντέλου και της ικανότητάς του να γενικεύει στα δεδομένα δοκιμής.

Ανεμογεννήτρια (Generator)	Τελευταίο Boosting Round	Train RMSE	Validation RMSE
RG01	514	0.04340	0.04516
RG02	558	0.04857	0.05083
RG03	454	0.03697	0.03896
RG04	369	0.04006	0.04146
Ανεμογεννήτρια (Gearbox)	Τελευταίο Boosting Round	Train RMSE	Validation RMSE
RG01	1546	0.01659	0.01847
RG02	1503	0.01674	0.01775
RG03	1504	0.01737	0.01903
RG04	1999	0.01677	0.01857

Πίνακας 2. Κανονικοποιημένες Τιμές RMSE και Τελευταίο Boosting Round για Κάθε Ανεμογεννήτρια

Η ανάλυση των διαγραμμάτων σύγκλισης των κανονικοποιημένων τιμών RMSE αποδεικνύουν ότι το μοντέλο XGBoost συγκλίνει αποτελεσματικά και διατηρεί

σταθερή απόδοση τόσο για τη θερμοκρασία της γεννήτριας όσο και του κιβωτίου ταχυτήτων.

1. Σύγκλιση του μοντέλου και Σταθερότητα :

- Το RMSE μειώνεται απότομα στα πρώτα boosting rounds, γεγονός που δείχνει ότι το μοντέλο μαθαίνει γρήγορα τα μοτίβα των δεδομένων.
- Σταθεροποιείται μετά από έναν αριθμό επαναλήψεων, υποδεικνύοντας ότι δεν συνεχίζει να μειώνει το σφάλμα σημαντικά, άρα έχει συγκλίνει.
- Οι τιμές Train RMSE και Validation RMSE παραμένουν κοντά, κάτι που αποκλείει την υπερπροσαρμογή (overfitting) και δείχνει ότι το μοντέλο γενικεύει καλά.

2. Ακρίβεια στις Προβλέψεις και Γενίκευση :

- Οι τελικές τιμές RMSE είναι χαμηλές και σταθερές, γεγονός που δείχνει ότι το XGBoost έχει καταφέρει να μάθει τις σχέσεις στα δεδομένα και να τις εφαρμόσει σωστά στα δεδομένα δοκιμής (validation set).
- Το μοντέλο δεν επηρεάζεται σημαντικά από τη διαφορά μεταξύ εκπαίδευσης και δοκιμής, αποδεικνύοντας ότι μπορεί να προβλέπει με συνεπή ακρίβεια σε διαφορετικά σύνολα δεδομένων.

3. Υπολογιστική Αποδοτικότητα και Boosting Rounds

- Ο αριθμός των boosting rounds που απαιτούνται για τη σύγκλιση διαφέρει ανάμεσα στις δύο περιπτώσεις (γεννήτρια και κιβώτιο ταχυτήτων), με την πρόβλεψη της θερμοκρασίας του κιβωτίου ταχυτήτων να απαιτεί περισσότερα boosting rounds.
- Παρόλο που ορισμένα μοντέλα απαιτούν περισσότερη εκπαίδευση (π.χ. ~1500-2000 rounds), το RMSE δεν μειώνεται δραστικά μετά από ένα σημείο, αποδεικνύοντας ότι η αύξηση των επαναλήψεων δεν οδηγεί σε υπερβολική προσαρμογή.

Συνολικά το μοντέλο XGBoost παρουσιάζει ισχυρή απόδοση και σταθερότητα, επιτυγχάνοντας χαμηλές τιμές RMSE χωρίς υπερπροσαρμογή. Η αποτελεσματικότητά του στην πρόβλεψη τόσο της θερμοκρασίας της γεννήτριας όσο και του κιβωτίου ταχυτήτων υποδηλώνει ότι μπορεί να αξιοποιηθεί αξιόπιστα για ανάλυση ανωμαλιών και προγνωστική συντήρηση.

6.1.2 Στατιστικοί Δείκτες στο Σύνολο Εκπαίδευσης

Για την αξιολόγηση της απόδοσης του XGBoost στο σύνολο εκπαίδευσης, υπολογίστηκαν οι βασικοί στατιστικοί δείκτες σφάλματος Mean Absolute Error (MAE), Mean Squared Error (MSE), Mean Squared Error σε ποσοστό (MSE%) και ο συντελεστής προσδιορισμού R^2 . Οι δείκτες αυτοί επιτρέπουν την εκτίμηση της ακρίβειας του μοντέλου στα δεδομένα εκπαίδευσης, παρέχοντας μια πρώτη εικόνα της ικανότητάς του να μάθει τις σχέσεις μεταξύ των μεταβλητών.

Στον παρακάτω πίνακα (Πίνακας 3) παρουσιάζονται οι υπολογισμένοι στατιστικοί δείκτες για το σύνολο εκπαίδευσης, οι οποίοι έχουν υπολογιστεί για τις κανονικοποιημένες τιμές των θερμοκρασιών της γεννήτριας και του κιβωτίου ταχυτήτων. Οι δείκτες αυτοί παρέχουν μια συνολική εικόνα της απόδοσης του μοντέλου XGBoost κατά τη διαδικασία μάθησης, επιτρέποντας την αξιολόγηση της ακρίβειάς του χωρίς να επηρεάζεται από τη μονάδα μέτρησης των δεδομένων.

Ανεμογεννήτρια (Generator)	MSE	MSE(%)	MAE	R²
RG01	0.00	0.64%	0.03	0.9058
RG02	0.00	0.86%	0.04	0.8763
RG03	0.00	0.43%	0.03	0.8705
RG04	0.00	0.51%	0.03	0.8952
Ανεμογεννήτρια (Gearbox)				
RG01	0.00	0.30%	0.01	0.8862
RG02	0.00	0.27%	0.02	0.8902
RG03	0.00	0.31%	0.02	0.8630
RG04	0.00	0.22%	0.01	0.8680

Πίνακας 3.Κανονικοποιημένοι Στατιστικοί Δείκτες Απόδοσης του Μοντέλου XGBoost για το Σύνολο Εκπαίδευσης

Οι τιμές των στατιστικών δεικτών επιβεβαιώνουν τα συμπεράσματα που προέκυψαν από την ανάλυση των διαγραμμάτων σύγκλισης και των κανονικοποιημένων τιμών RMSE, αποδεικνύοντας τη σταθερότητα και την υψηλή ακρίβεια του μοντέλου XGBoost στις προβλέψεις των θερμοκρασιών της γεννήτριας και του κιβωτίου ταχυτήτων. Συγκεκριμένα:

- Οι χαμηλές τιμές MSE και MSE(%) επιβεβαιώνουν τη γρήγορη σύγκλιση του μοντέλου και τη διατήρηση της σταθερής του απόδοσης.
- Οι τιμές MAE παραμένουν χαμηλές, δείχνοντας μικρές αποκλίσεις μεταξύ των πραγματικών και των προβλεπόμενων τιμών.
- Οι υψηλές τιμές R² (πάνω από 86% σε όλες τις περιπτώσεις) επιβεβαιώνουν ότι το μοντέλο εξηγεί με ακρίβεια τη διακύμανση των πραγματικών τιμών, αποφεύγοντας το overfitting.

6.1.3 Αξιολόγηση στο Σύνολο Δοκιμής (testing set)

Για την αξιολόγηση της απόδοσης του μοντέλου XGBoost σε νέα, μη ορατά δεδομένα, πραγματοποιήθηκε ανάλυση στο σύνολο δοκιμής. Η αξιολόγηση αυτή επιτρέπει τη σύγκριση της ακρίβειας των προβλέψεων με τις πραγματικές τιμές, προσδιορίζοντας την ικανότητα του μοντέλου να γενικεύει σωστά. Στο πλαίσιο αυτό, παρουσιάζονται οι στατιστικοί δείκτες σφάλματος, καθώς και διαγράμματα που αποτυπώνουν τη σχέση μεταξύ πραγματικών και προβλεπόμενων τιμών, στο σύνολο δοκιμής.

A. Στατιστικοί Δείκτες

Στον παρακάτω πίνακα (Πίνακας 4) παρουσιάζονται οι στατιστικοί δείκτες του μοντέλου XGBoost στο σύνολο δοκιμής, σε κανονικοποιημένες τιμές. Οι δείκτες αυτοί επιτρέπουν την ποσοτική εκτίμηση της ακρίβειας των προβλέψεων και τη σύγκριση με την απόδοση στο σύνολο εκπαίδευσης.

Ανεμογεννήτρια (Generator)	MSE	MSE(%)	MAE	R²
RG01	0.0022	0.690%	0.0351	0.9058
RG02	0.0034	1.02%	0.0438	0.8763
RG03	0.0024	0.770%	0.0360	0.8705
RG04	0.0000	0.510%	0.0300	0.8952
Ανεμογεννήτρια (Gearbox)	MSE	MSE(%)	MAE	R²
RG01	0.0004	0.083%	0.0142	0.9062
RG02	0.0004	0.070%	0.143	0.8902
RG03	0.0005	0.100%	0.0168	0.8630
RG04	0.0005	0.080%	0.0151	0.8680

Πίνακας 4.Κανονικοποιημένοι Στατιστικοί Δείκτες Απόδοσης του Μοντέλου XGBoost για το Σύνολο Δοκιμής

Τα αποτελέσματα της αξιολόγησης στο σύνολο δοκιμής επιβεβαιώνουν ότι το μοντέλο XGBoost διατηρεί υψηλή ακρίβεια στις προβλέψεις του, καθώς οι στατιστικοί δείκτες απόδοσης παραμένουν σε χαμηλά επίπεδα και το R² διατηρείται στα ίδια επίπεδα με το σύνολο εκπαίδευσης. Συγκεκριμένα:

1. Απόδοση του Μοντέλου στη Θερμοκρασία της Γεννήτριας

- Το R² παραμένει σταθερό και υψηλό (≥ 0.87 σε όλες τις ανεμογεννήτριες), επιβεβαιώνοντας ότι το μοντέλο εξηγεί με ακρίβεια τη διακύμανση των πραγματικών τιμών και δεν υπερπροσαρμόζεται (overfitting).
- Οι τιμές MAE και MSE(%) είναι ελαφρώς αυξημένες σε σχέση με το σύνολο εκπαίδευσης, αλλά η διαφορά είναι μικρή, υποδεικνύοντας ότι το μοντέλο μπορεί να γενικεύσει σωστά σε νέα δεδομένα.

2. Απόδοση του Μοντέλου στη Θερμοκρασία του Gearbox

- Το R² κυμαίνεται μεταξύ 0.8630 και 0.8902, επιβεβαιώνοντας ότι το μοντέλο προβλέπει με υψηλή ακρίβεια τη θερμοκρασία του κιβωτίου ταχυτήτων.
- Οι τιμές MSE και MSE(%) είναι χαμηλότερες σε σχέση με τη γεννήτρια, γεγονός που δείχνει ότι το μοντέλο αντιμετωπίζει τη θερμοκρασία του gearbox ως μια πιο σταθερή μεταβλητή.

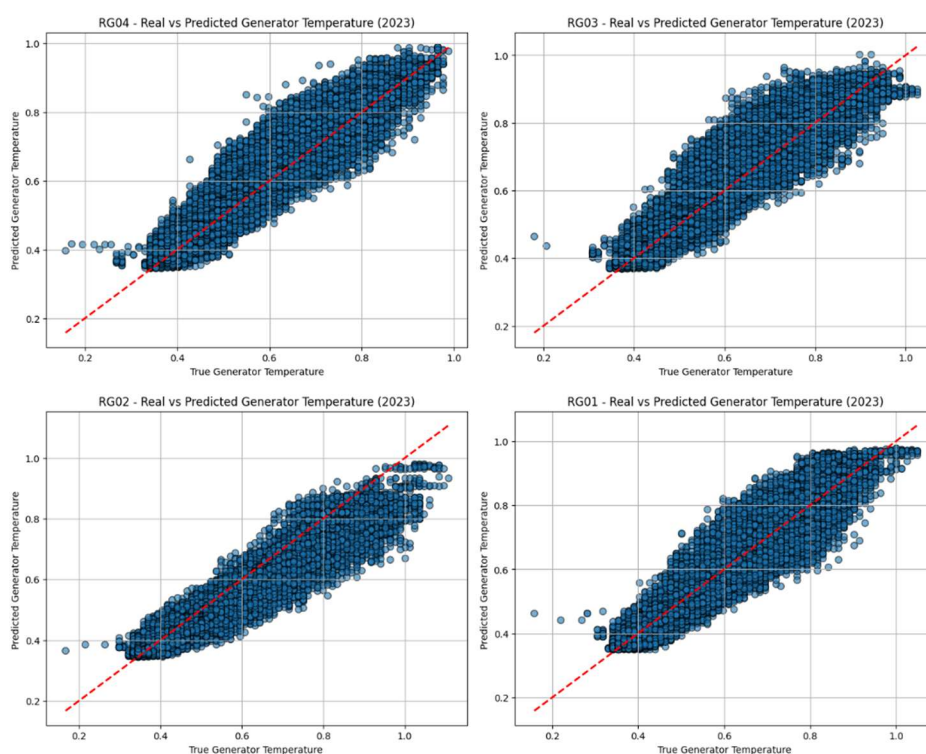
Β. Διαγράμματα Σύγκρισης Πραγματικών και Προβλεπόμενων Τιμών

Στα παρακάτω διαγράμματα (Σχήμα 23-24) παρουσιάζεται η σύγκριση μεταξύ των πραγματικών και των προβλεπόμενων κανονικοποιημένων τιμών της θερμοκρασίας της γεννήτριας και του κιβωτίου ταχυτήτων για κάθε ανεμογεννήτρια (RG01 - RG04) στο σύνολο δοκιμής (testing set).

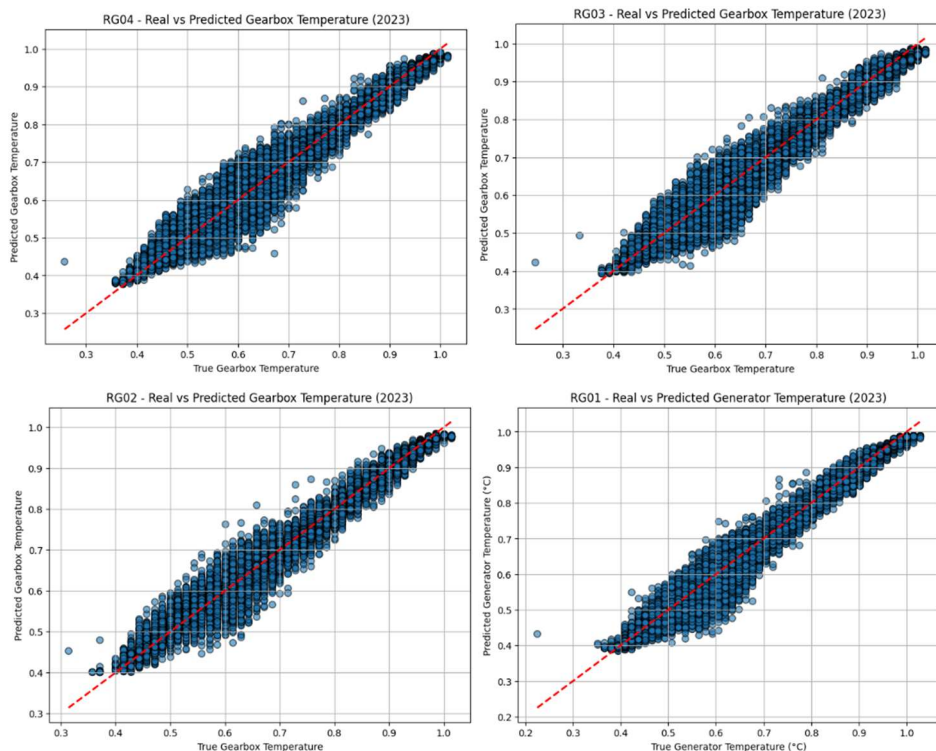
Δομικά Στοιχεία του Διαγράμματος:

- **Οριζόντιος Άξονας:** Πραγματικές τιμές της θερμοκρασίας.
- **Κατακόρυφος Άξονας:** Προβλεπόμενες τιμές της θερμοκρασίας .
- **Κόκκινη Διακεκομμένη Γραμμή:** Ιδανική ευθεία ($y=x$), που υποδηλώνει τέλεια ταύτιση μεταξύ προβλεπόμενων και πραγματικών τιμών.
- **Διάταξη Σημείων:** Τα σημεία κοντά στη διαγώνιο υποδηλώνουν υψηλή ακρίβεια στις προβλέψεις, ενώ οι αποκλίσεις μπορούν να υποδείξουν σημεία όπου το μοντέλο δεν αποδίδει ιδανικά.

Τα διαγράμματα προσφέρουν μια οπτική αναπαράσταση της ακρίβειας του μοντέλου, ενώ οι μικρές αποκλίσεις σε ορισμένα σημεία ενδέχεται να αναδεικνύουν περιοχές όπου το μοντέλο χρειάζεται περαιτέρω βελτιστοποίηση.



Εικόνα 23.Διάγράμματα Πραγματικών-Προβλεπόμενων Τιμών για την θερμοκρασία της γεννήτριας (RG01-RG04)



Εικόνα 24. Διαγράμματα Πραγματικών-Προβλεπόμενων Τιμών για την θερμοκρασία του κιβωτίου ταχυτήτων (RG01-RG04)

Από τα διαγράμματα μεταξύ πραγματικών και προβλεπόμενων τιμών προκύπτουν τα εξής συμπεράσματα :

1. Ακρίβεια Προβλέψεων και Συμμετρία

- Γενικά, τα σημεία είναι κατανεμημένα κοντά στη διαγώνιο $y = x$, γεγονός που δείχνει ότι το μοντέλο XGBoost κάνει ακριβείς προβλέψεις. Η καλή ευθυγράμμιση με τη διαγώνιο υποδηλώνει ότι το μοντέλο γενικεύει σωστά και δεν παρουσιάζει μεγάλες συστηματικές αποκλίσεις.

2. Σύγκριση Γεννήτριας και Κιβωτίου Ταχυτήτων

- Οι προβλέψεις για το κιβώτιο ταχυτήτων φαίνονται πιο ακριβείς, καθώς τα σημεία είναι πιο συγκεντρωμένα γύρω από τη διαγώνιο. Αυτό επιβεβαιώνει τα ευρήματα των στατιστικών δεικτών, που έδειξαν χαμηλότερα σφάλματα (MSE, MAE) για το gearbox.
- Οι αποκλίσεις στις προβλέψεις της γεννήτριας είναι ελαφρώς μεγαλύτερες, ειδικά στις περιοχές χαμηλών θερμοκρασιών, γεγονός που δείχνει ότι το μοντέλο αντιμετωπίζει μεγαλύτερη αβεβαιότητα στη συγκεκριμένη περίπτωση.

6.1.4 Διαγράμματα SPC για τον εντοπισμό πιθανών βλαβών

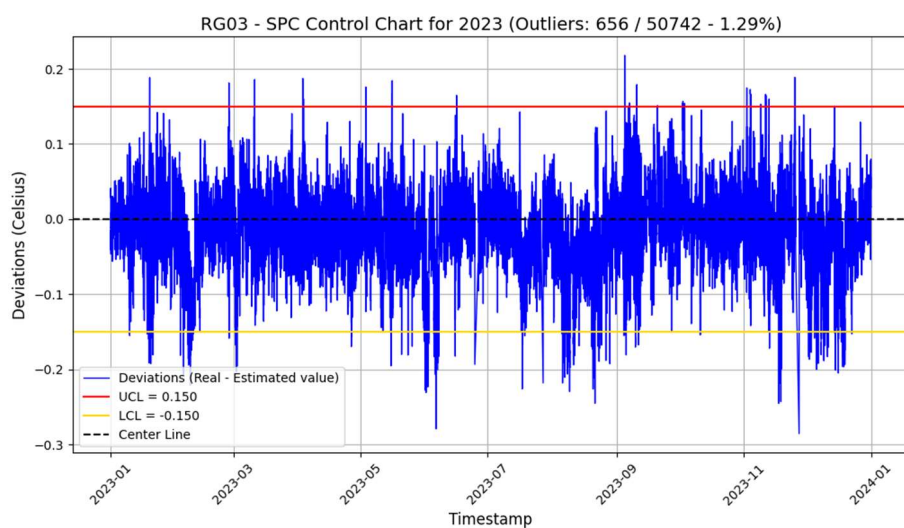
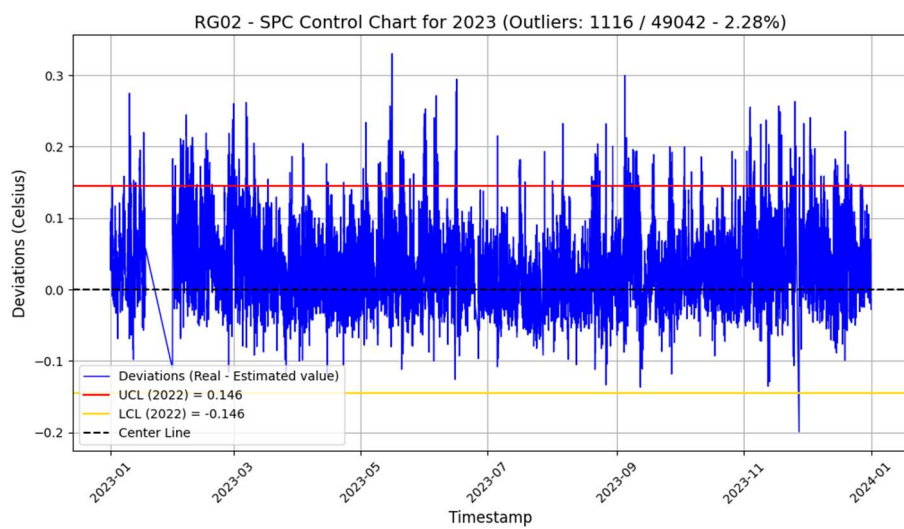
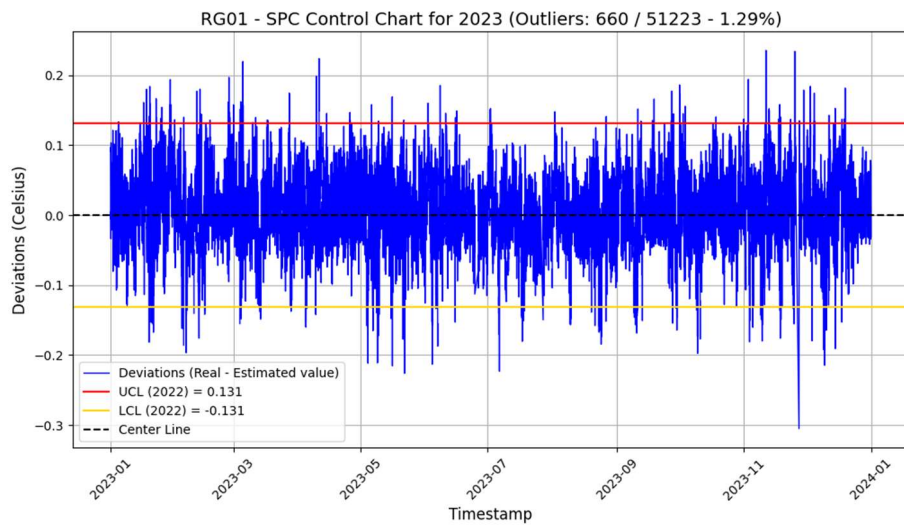
Τα διαγράμματα SPC (Statistical Process Control) χρησιμοποιούνται για την παρακολούθηση της σταθερότητας και της απόδοσης ενός συστήματος, επιτρέποντας την ανίχνευση αποκλίσεων που μπορεί να υποδηλώνουν πιθανές

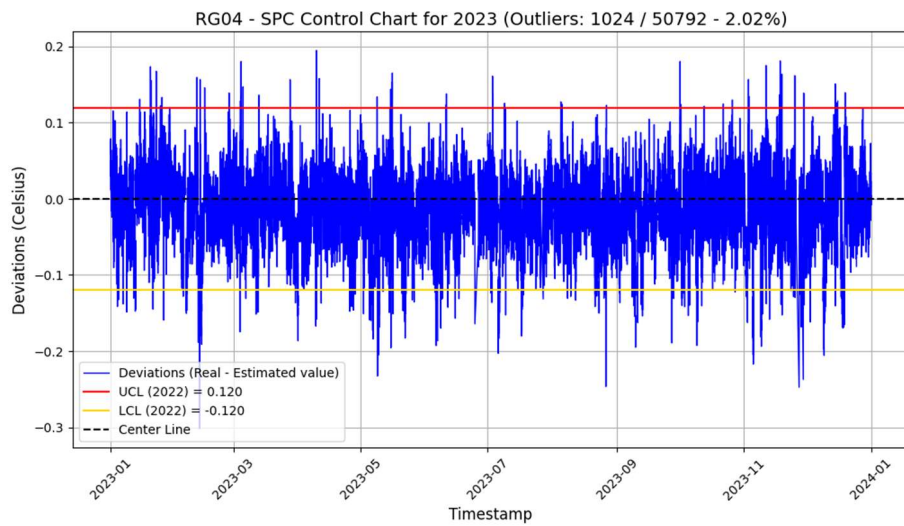
βλάβες ή αστοχίες. Στην παρούσα ενότητα, τα SPC διαγράμματα εφαρμόζονται στο σύνολο δοκιμής (testing set), καθώς αυτό επιτρέπει την αξιολόγηση της ικανότητας του μοντέλου να γενικεύει σωστά σε νέα, μη ορατά δεδομένα.

Η ανάλυση των αποκλίσεων μεταξύ πραγματικών και προβλεπόμενων τιμών συμβάλλει στην ανίχνευση ανωμαλιών και ακραίων τιμών, οι οποίες ενδέχεται να υποδηλώνουν δυσλειτουργίες στις ανεμογεννήτριες. Για τον χαρακτηρισμό μιας απόκλισης ως πιθανό σφάλμα, λαμβάνεται υπόψη η συνεχής εμφάνιση τουλάχιστον έξι διαδοχικών τιμών εκτός των ορίων ελέγχου, καθώς αυτή η συμπεριφορά μπορεί να υποδηλώνει απόκλιση από τη φυσιολογική λειτουργία.

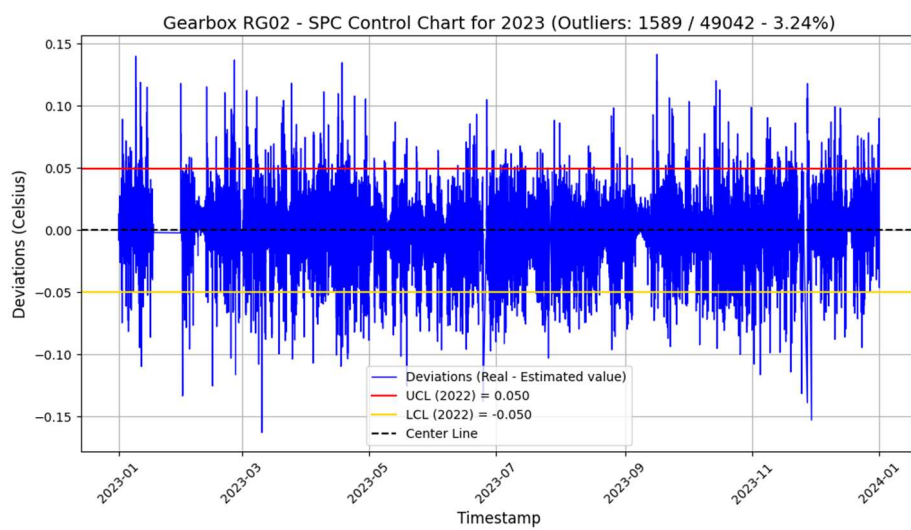
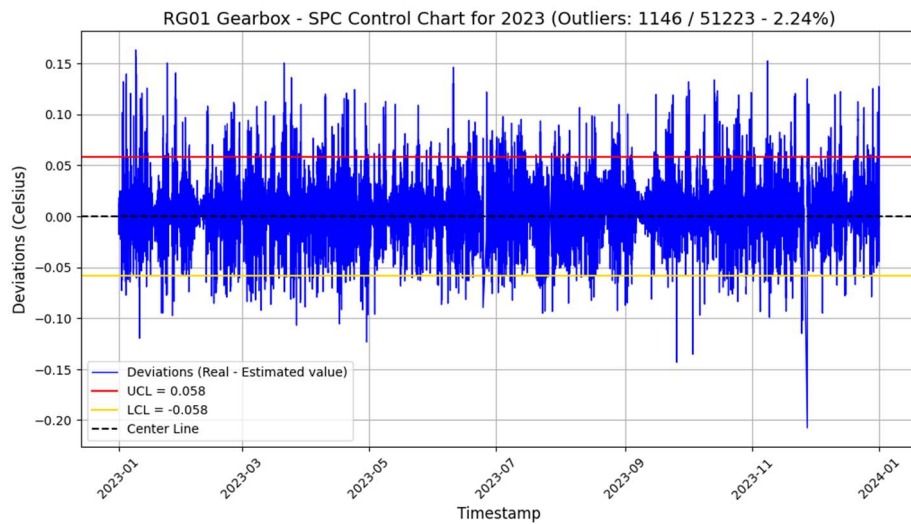
Στα παρακάτω διαγράμματα (Εικόνες 25-26) παρουσιάζονται τα διαγράμματα SPC, τα οποία απεικονίζουν τις αποκλίσεις μεταξύ των πραγματικών και των προβλεπόμενων τιμών. Συγκεκριμένα, η Εικόνα 25 αφορά τη θερμοκρασία της γεννήτριας, ενώ η Εικόνα 26 τη θερμοκρασία του κιβωτίου ταχυτήτων. Τα χαρακτηριστικά τους είναι :

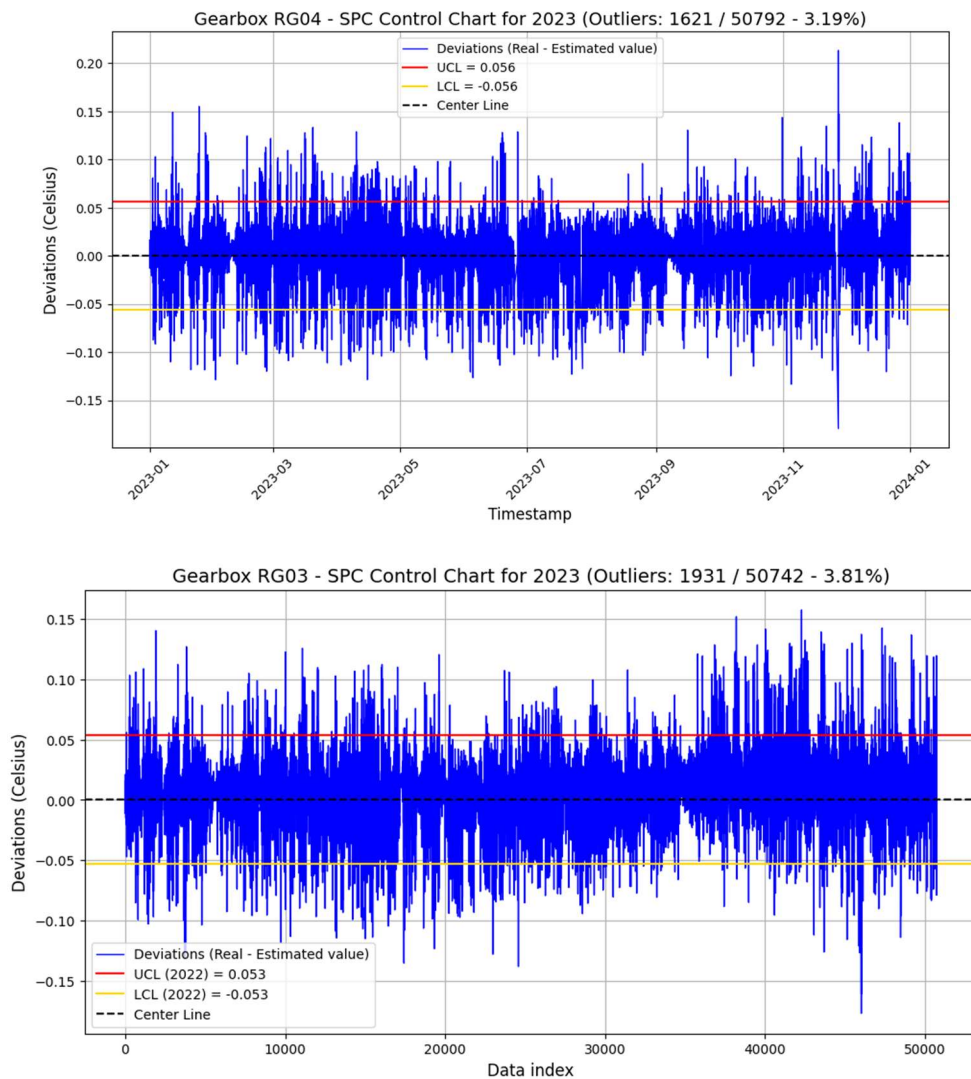
- **Οριζόντιος άξονας (Timestamp):** Δείχνει τη χρονική εξέλιξη των αποκλίσεων.
- **Κατακόρυφος άξονας (Deviations):** Παρουσιάζει τη διαφορά μεταξύ πραγματικών και προβλεπόμενων τιμών θερμοκρασίας.
- **Μπλε γραμμή (Deviations):** Απεικονίζει τις διαφορές μεταξύ των πραγματικών και των προβλεπόμενων τιμών.
- **Μαύρη διακεκομμένη γραμμή (Center Line):** Αντιπροσωπεύει τη μέση τιμή των αποκλίσεων.
- **Κόκκινη και κίτρινη γραμμή (Όρια ελέγχου - UCL & LCL):** Οριοθετούν το εύρος των αποδεκτών αποκλίσεων.
- **Σημεία εκτός ορίων:** Ενδέχεται να υποδηλώνουν πιθανές δυσλειτουργίες ή αστοχίες του μοντέλου πρόβλεψης.





Εικόνα 25.SPC Διαγράμματα της γεννήτριας (RG01-RG04)





Εικόνα 26.SPC Διαγράμματα για του κιβώτιο ταχυτήτων(RG01-RG04)

Ανάλυση των Αποκλίσεων και Ενδείξεις Πιθανών Ανωμαλιών

Τα διαγράμματα SPC για τη θερμοκρασία της γεννήτριας και του κιβωτίου ταχυτήτων, δείχνουν ότι παρά τη συνολική σταθερότητα του συστήματος, υπάρχουν περιπτώσεις όπου οι αποκλίσεις ξεπερνούν τα όρια ελέγχου, γεγονός που μπορεί να υποδηλώνει ανωμαλίες στη λειτουργία της ανεμογεννήτριας. Η ανάλυση αυτών των αποκλίσεων παρέχει χρήσιμες πληροφορίες για την ανίχνευση πιθανών βλαβών ή μη φυσιολογικών συνθηκών λειτουργίας.

1. Συγκέντρωση αποκλίσεων εκτός ορίων

- Οι αποκλίσεις που ξεπερνούν τα ανώτερα και κατώτερα όρια ελέγχου (UCL & LCL) είναι ιδιαίτερα σημαντικές, καθώς μπορεί να σηματοδοτούν καταστάσεις που αποκλίνουν από την κανονική λειτουργία

2. Εντοπισμός διαδοχικών αποκλίσεων εκτός ορίων

- Σύμφωνα με τα κριτήρια ανίχνευσης ανωμαλιών που θέσαμε, πιθανώς μια βλάβη μπορεί να υποδηλωθεί όταν εμφανίζονται τουλάχιστον 6 διαδοχικές τιμές εκτός των ορίων ελέγχου.
- Στα διαγράμματα SPC, υπάρχουν σημεία όπου παρατηρείται συνεχόμενη ακολουθία αποκλίσεων εκτός των ορίων, γεγονός που μπορεί να αποτελεί ένδειξη πιθανής αστοχίας και επικείμενης βλάβης.

3. Εποχικότητα και εξωτερικοί παράγοντες

- Και στις τέσσερις ανεμογεννήτριες, παρατηρείται ότι οι αποκλίσεις είναι πιο έντονες σε συγκεκριμένες χρονικές περιόδους, γεγονός που ενδεχομένως σχετίζεται με εξωτερικούς παράγοντες, όπως αλλαγές στις περιβαλλοντικές συνθήκες (π.χ. θερμοκρασία περιβάλλοντος, υγρασία, ταχύτητα ανέμου).

Εντοπισμός Διαδοχικών Αποκλίσεων και Ανάλυση Πιθανών Βλαβών

Ο παρακάτω πίνακας (Πίνακας 5) συνοψίζει τον αριθμό των αποκλίσεων όπου καταγράφηκαν τουλάχιστον **6 συνεχόμενες τιμές** εκτός των ορίων ελέγχου. Οι περιπτώσεις αυτές είναι ιδιαίτερα σημαντικές, καθώς μπορεί να υποδηλώνουν πιθανές ανωμαλίες που δεν οφείλονται σε τυχαία σφάλματα, αλλά σε πραγματικές δυσλειτουργίες των ανεμογεννητριών.

Component	Wind Turbine	Total Data	Outliers	Outliers (%)
Gearbox	RG01	51223	25	0.05%
Gearbox	RG02	49042	17	0.03%
Gearbox	RG03	50742	23	0.05%
Gearbox	RG04	50792	35	0.07%
Generator	RG01	51223	16	0.03%
Generator	RG02	49042	38	0.08%
Generator	RG03	50742	96	0.19%
Generator	RG04	50792	35	0.07%

Πίνακας 5. Διαδοχικές Αποκλίσεις

Δεδομένης της απουσίας ιστορικού βλαβών, οι πιθανότερες χρονικές στιγμές εμφάνισης μιας πραγματικής βλάβης είναι μετά την καταγραφή διαδοχικών αποκλίσεων (τουλάχιστον 6 στη συγκεκριμένη ανάλυση). Αυτό υποδηλώνει ότι η απόκλιση δεν είναι απλά τυχαία ή αποτέλεσμα ψευδούς συναγερμού, αλλά πιθανώς αντανακλά μια πραγματική δυσλειτουργία στο σύστημα. Ως εκ τούτου, οι συναγερμοί που σχετίζονται με συνεχόμενες αποκλίσεις θα πρέπει να λαμβάνονται σοβαρά υπόψη και να εξετάζονται προσεκτικά, καθώς μπορεί να αποτελούν πρώιμες ενδείξεις επικείμενης βλάβης.

Παρατηρήσεις

Στην RG01 έγινε συντήρηση στις 15/3 και εμφανίστηκαν ξανά διαδοχικές αποκλίσεις στις 8/5. Επιπλέον πραγματοποιήθηκε συντήρηση και στις 18/9 και οι αποκλίσεις εμφανίστηκαν δύο μήνες μετά. στις 18/11.

Στην RG02 πραγματοποιήθηκε συντήρηση στις 10/3 και μετά από αυτήν για διάστημα δύο μηνών δεν καταγράφηκαν διαδοχικές αποκλίσεις εκτός ορίων.

Στην RG03 η συντήρηση πραγματοποιήθηκε στις 20/3 και οι επόμενες διαδοχικές αποκλίσεις εμφανίστηκαν μετά από περίπου δύο μήνες (στις 11/5).

Για την RG04 η συντήρηση πραγματοποιήθηκε στις 21/3 με τις πρώτες διαδοχικές αποκλίσεις εκτός ορίων μετά την συντήρηση να εμφανίζονται μετά από 40 ημέρες, στις 29/4. Επίσης πραγματοποιήθηκε μια συντήρηση στις 7/7 και οι επόμενες διαδοχικές αποκλίσεις παρουσιάστηκαν στις 5/9, μετά από δύο μήνες περίπου.

Συνολικά και στις τέσσερις ανεμογεννήτριες (RG01-RG04) πριν την εφαρμογή των συντηρήσεων είχαμε διαδοχικές αποκλίσεις εκτός ορίων. Μετά από κάθε συντήρηση και για μεγάλο χρονικό διάστημα , από 40 ημέρες έως και 2 μήνες, δεν παρατηρήθηκαν διαδοχικές αποκλίσεις εκτός ορίων γεγονός που επιβεβαιώνει την σημασία της συντήρησης στις ανεμογεννήτριες και ενισχύει την αξιοπιστία των προβλέψεων.

6.2 Ανάλυση Αποτελεσμάτων για Ανεμογεννήτριες με Ιστορικό Βλαβών

Στην παρούσα ενότητα αξιολογείται η απόδοση του μοντέλου XGBoost σε ανεμογεννήτριες με καταγεγραμμένο ιστορικό βλαβών. Η ανάλυση επικεντρώνεται στο κατά πόσο το μοντέλο μπορεί να ανιχνεύσει και να προβλέψει σωστά τις αποκλίσεις που σχετίζονται με πραγματικά σφάλματα και δυσλειτουργίες των ανεμογεννητριών και αν αυτές εμφανίζονται την καταγραφή της βλάβης.

Συγκεκριμένα Η ανάλυση πραγματοποιείται σε δύο ανεμογεννήτριες:

- **Μία ανεμογεννήτρια με βλάβη στο κιβώτιο ταχυτήτων.**
- **Μία ανεμογεννήτρια με βλάβη στη γεννήτρια**

Στόχος είναι η σύγκριση των προβλεπόμενων τιμών του μοντέλου με τις πραγματικές μετρήσεις, ώστε να διερευνηθεί αν η προτεινόμενη μέθοδος μπορεί να ανιχνεύσει ανωμαλίες πριν από την εμφάνιση της βλάβης.

6.2.1 Βλάβη στο Κιβώτιο Ταχυτήτων

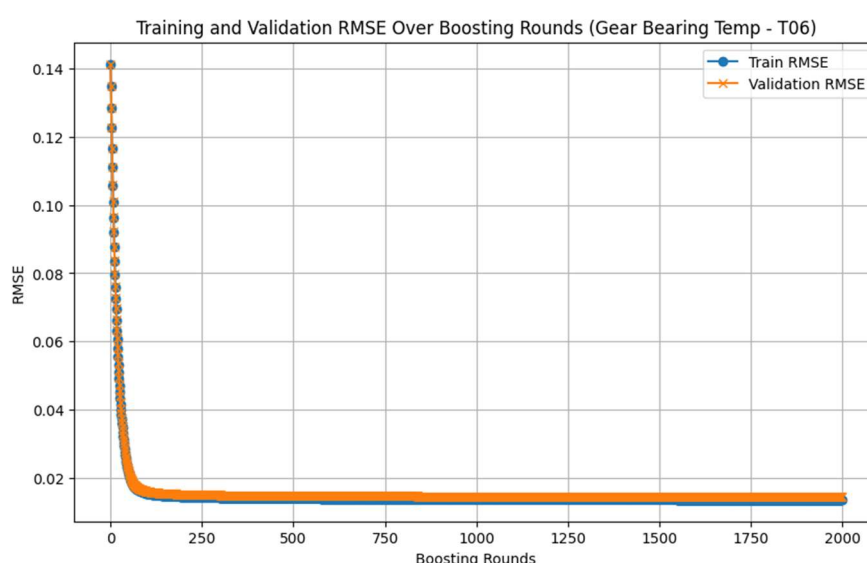
Στην παρούσα ενότητα εξετάζεται η περίπτωση ανεμογεννήτριας στην οποία καταγράφηκε βλάβη στο κιβώτιο ταχυτήτων στις 17/10/2017 στις 08:38. Η ανάλυση επικεντρώνεται στην ικανότητα του μοντέλου XGBoost να εντοπίσει ανωμαλίες στις

θερμοκρασίες του κιβωτίου ταχυτήτων πριν από την εμφάνιση της βλάβης, παρέχοντας ενδείξεις για πρόωμη διάγνωση και πιθανή εφαρμογή προγνωστικής συντήρησης (PdM).

Για την αξιολόγηση της επίδοσης του μοντέλου θα χρησιμοποιηθούν τα ίδια εργαλεία και μεθοδολογίες που παρουσιάστηκαν στην 6.1.

Εξέλιξη του RMSE κατά την εκπαίδευση

Το διάγραμμα που ακολουθεί (Εικόνα 27) παρουσιάζει την μεταβολή του RMSE κατά την διαδικασία της εκπαίδευσης, αναδεικνύοντας τη διαδικασία σύγκλισης του μοντέλου και την πιθανή εμφάνιση φαινομένων υπερπροσαρμογής (overfitting).



Εικόνα 27.Εξέλιξη του RMSE για την T06

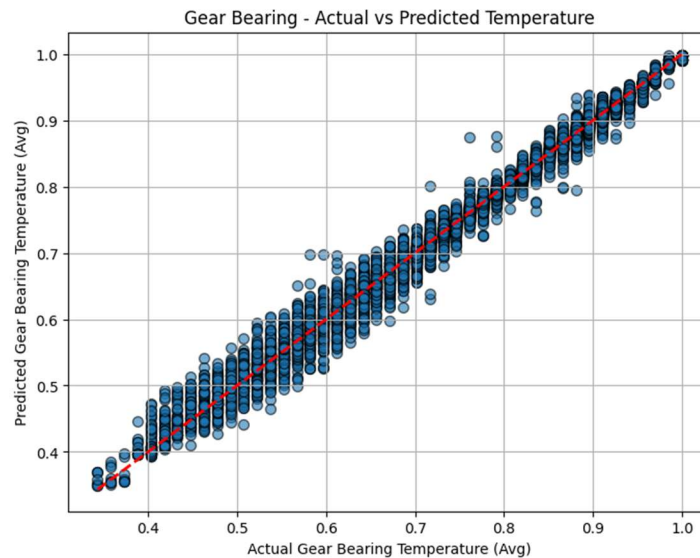
Από το παραπάνω διάγραμμα συμπεραίνουμε ότι το μοντέλο XGBoost εκπαιδεύεται καλά και είναι καλά ρυθμισμένο, καταφέροντας να μειώσει το σφάλμα του γρήγορα και να διατηρήσει χαμηλό Validation RMSE. Η έλλειψη μεγάλων αποκλίσεων μεταξύ εκπαίδευσης και δοκιμής υποδηλώνει ότι το μοντέλο μπορεί να γενικεύσει καλά σε νέα δεδομένα.

Στατιστικοί Δείκτες και Διάγραμμα Πραγματικών-Προβλεπόμενων Τιμών

Παρακάτω παρουσιάζονται (Πίνακας 6) οι στατιστικοί δείκτες και το διάγραμμα πραγματικών - προβλεπόμενων τιμών (Εικόνα 28) για το σύνολο δοκιμής, με στόχο την αξιολόγηση της ακρίβειας του μοντέλου XGBoost στην πρόβλεψη της θερμοκρασίας του κιβωτίου ταχυτήτων.

Ανεμογεννήτρια (Gearbox)	MSE	MSE%	MAE	R^2
T06	0.0003	0.05%	0.0124	0.9752

Πίνακας 6. Κανονικοποιημένοι Στατιστικοί Δείκτες για την Ανεμογεννήτρια T06 με Ιστορικό Βλαβών

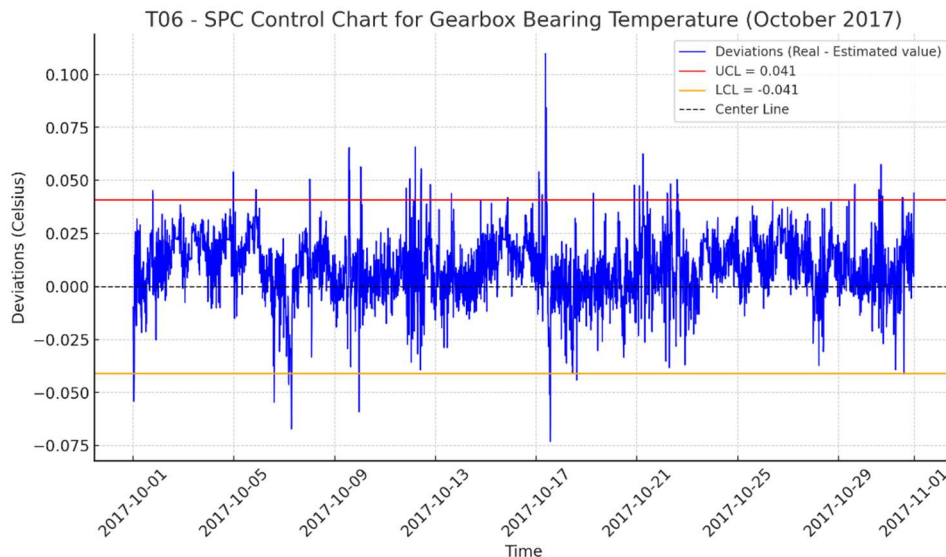


Εικόνα 28. Διάγραμμα Πραγματικών και Προβλεπόμενων τιμών για την θερμοκρασία του Gearbox της T06

Παρατηρούμε ότι ο XGBoost επιτυγχάνει υψηλή ακρίβεια στην πρόβλεψη της θερμοκρασίας του κιβωτίου ταχυτήτων, με πολύ χαμηλό MSE και MAE, ενώ το $R^2 = 0.9752$ δείχνει εξαιρετική προσαρμογή. Το διάγραμμα πραγματικών - προβλεπόμενων τιμών επιβεβαιώνει τη μικρή απόκλιση των προβλέψεων, καθώς τα σημεία συγκεντρώνονται κοντά στη διαγώνιο. Συνολικά, το μοντέλο εμφανίζει σταθερότητα και αξιοπιστία στις προβλέψεις.

Διάγραμμα SPC

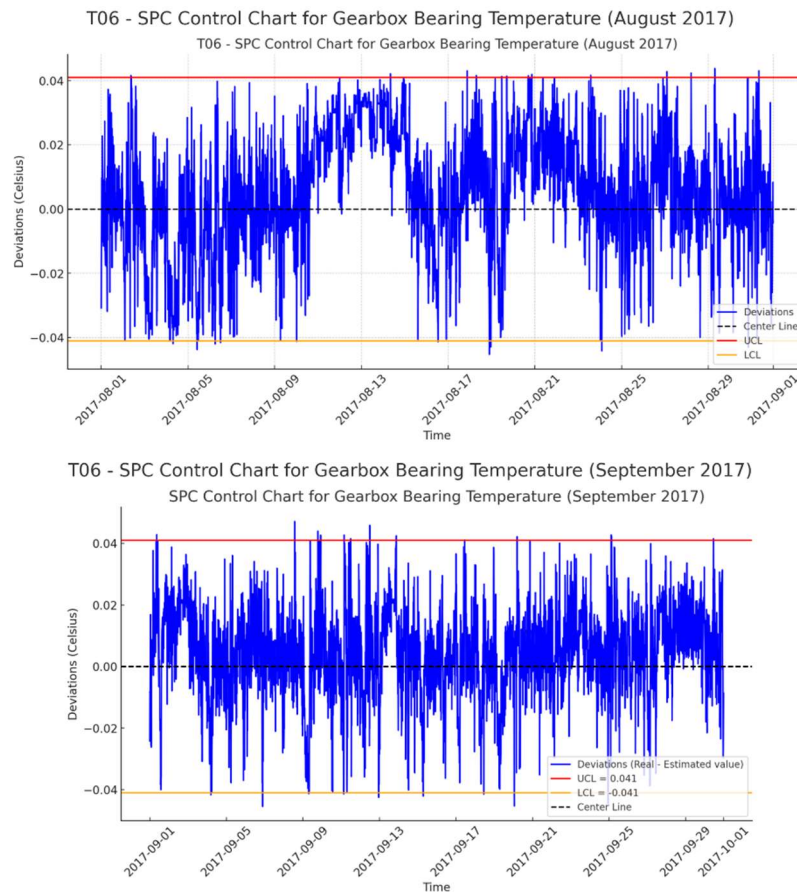
Από το ιστορικό βλαβών γνωρίζουμε ότι η βλάβη στο κιβώτιο ταχυτήτων της ανεμογεννήτριας T06 καταγράφηκε στις 17/10/2017 08:38. Το παρακάτω διάγραμμα SPC (Εικόνα 29) αφορά τον μήνα που συνέβη το σφάλμα και παρουσιάζει τις κανονικοποιημένες τιμές αποκλίσεις μεταξύ πραγματικών και προβλεπόμενων τιμών θερμοκρασίας του κιβωτίου ταχυτήτων, με κανονικοποιημένα τα όρια ελέγχου ($UCL = 0.041$, $LCL = -0.041$).



Εικόνα 29.Διάγραμμα SPC για την θερμοκρασία του κιβωτίου ταχυτήτων T06

Από το διάγραμμα SPC παρατηρούνται σημαντικές αποκλίσεις ήδη από τις 7/10/2017, υποδηλώνοντας πρώιμα σημάδια δυσλειτουργίας στο κιβώτιο ταχυτήτων. Το διάστημα 10-12/10/2017, οι αποκλίσεις αυξάνονται σημαντικά, με επτά διαδοχικές παραβιάσεις των ορίων ελέγχου (στις 12/10/2017 9:40:00 έως και τις 10:40:00), γεγονός που αποτελεί ισχυρή ένδειξη επικείμενης αστοχίας. Συνεπώς, το μοντέλο πρόβλεψης επιδεικνύει υψηλή προγνωστική ικανότητα, εντοπίζοντας τη βλάβη μία εβδομάδα νωρίτερα και επιτρέποντας την έγκαιρη διάγνωση και τη λήψη προληπτικών μέτρων συντήρησης.

Στην παρακάτω εικόνα (Εικόνα 30) παρουσιάζονται τα διαγράμματα SPC έως και δύο μήνες πριν την εμφάνιση του σφάλματος. Κατά το διάστημα αυτό, οι αποκλίσεις εμφανίζονται διάσπαρτες και περιορίζονται εντός των ορίων ελέγχου, χωρίς να παρατηρούνται συνεχόμενες ή εκτεταμένες υπερβάσεις. Αυτό υποδηλώνει ότι η λειτουργία του κιβωτίου ταχυτήτων παρέμενε σταθερή, χωρίς εμφανή σημάδια επικείμενης αστοχίας. Η ομαλή συμπεριφορά της ανεμογεννήτριας κατά την περίοδο αυτή ενισχύει την αξιοπιστία του μοντέλου πρόβλεψης, καθώς δεν εντοπίζονται ψευδείς συναγερμοί.



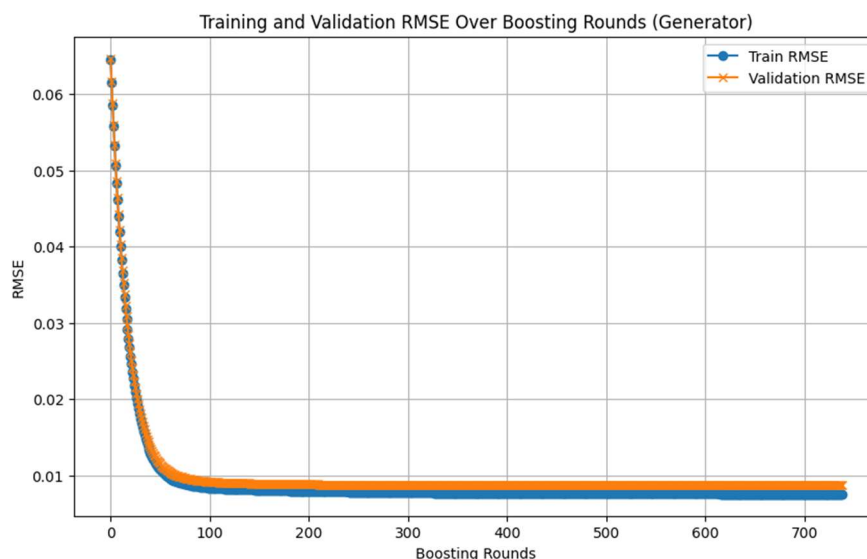
Εικόνα 30. Διαγράμματα SPC για τη Θερμοκρασία Κιβωτίου Ταχυτήτων της T06 (Αύγουστος - Σεπτέμβριος 2017)

6.2.2 Βλάβη στην Γεννήτρια

Στην παρούσα ενότητα αναλύεται η περίπτωση ανεμογεννήτριας στην οποία σημειώθηκε βλάβη στη γεννήτρια στις 20/08/2017 στις 06:08. Στόχος είναι να εξεταστεί αν το μοντέλο XGBoost μπορεί να ανιχνεύσει έγκαιρα αποκλίσεις στη θερμοκρασία της γεννήτριας πριν από τη βλάβη, παρέχοντας ενδείξεις για πιθανή αστοχία και επικείμενες βλάβες.

Εξέλιξη του RMSE κατά την εκπαίδευση

Από το παρακάτω διάγραμμα (Εικόνα 31) που απεικονίζει την εξέλιξη του RMSE κατά τη διάρκεια της εκπαίδευσης και επικύρωσης του μοντέλου XGBoost, συμπεραίνουμε ότι το μοντέλο μαθαίνει γρήγορα τα πρότυπα των δεδομένων, καθώς το σφάλμα μειώνεται απότομα στα πρώτα 100 boosting rounds. Η σταθεροποίηση του Validation RMSE σε χαμηλά επίπεδα μετά από περίπου 200 επαναλήψεις υποδηλώνει καλή γενίκευση και απουσία υπερπροσαρμογής (overfitting), γεγονός που καθιστά το μοντέλο αξιόπιστο για πρόβλεψη της θερμοκρασίας της γεννήτριας.



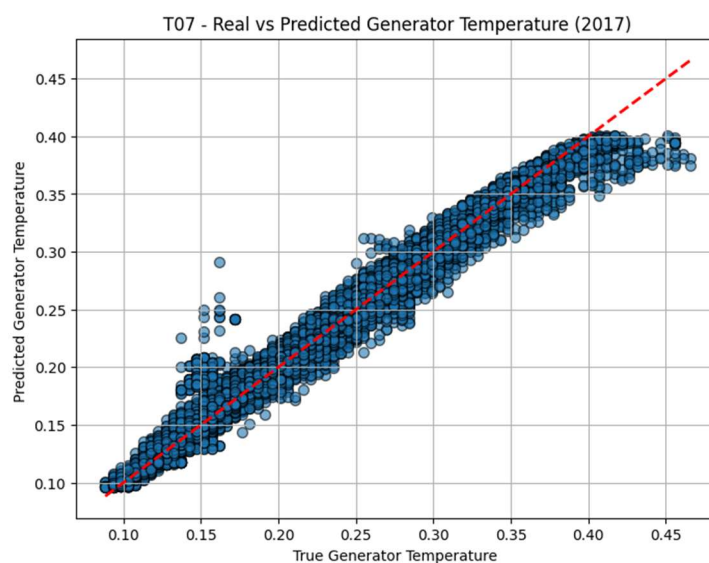
Εικόνα 31.Εξέλιξη του RMSE για την T07

Στατιστικοί Δείκτες και Διάγραμμα Πραγματικών-Προβλεπόμενων Τιμών

Παρακάτω παρουσιάζονται οι στατιστικοί δείκτες (Πίνακας 7) και το διάγραμμα πραγματικών - προβλεπόμενων τιμών για το σύνολο δοκιμής (Εικόνα 32), με στόχο την αξιολόγηση της ακρίβειας του μοντέλου XGBoost στην πρόβλεψη της θερμοκρασίας της γεννήτριας.

Ανεμογεννήτρια (Gearbox)	MSE	MSE%	MAE	R^2
T07	0.0001	0.17%	0.0065	0.9247

Πίνακας 7.Κανονικοποιημένοι Στατιστικοί Δείκτες για την Ανεμογεννήτρια T07 με Ιστορικό Βλαβών

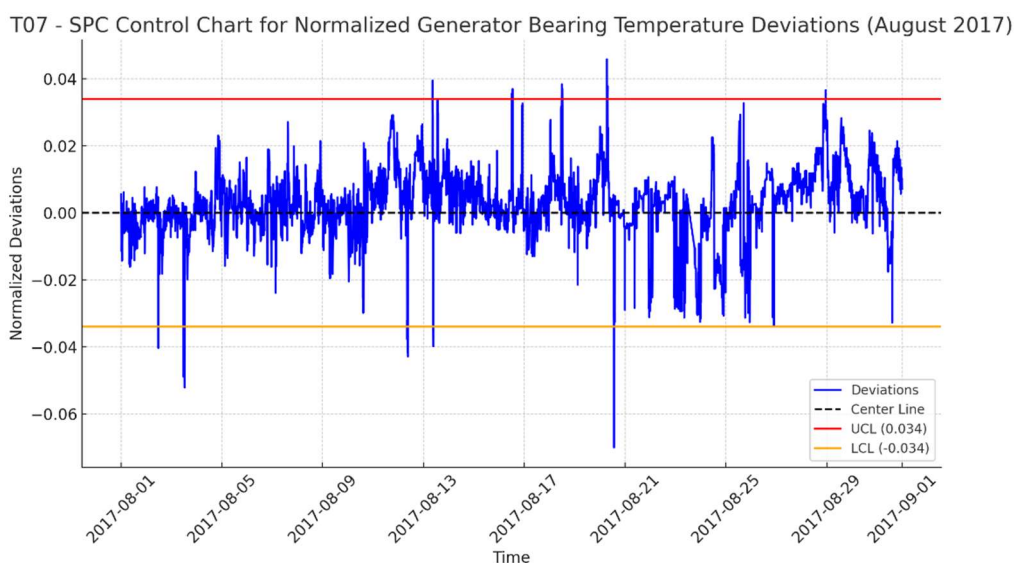


Εικόνα 32.Διάγραμμα Πραγματικών και Προβλεπόμενων τιμών για την θερμοκρασία της Generator της T07

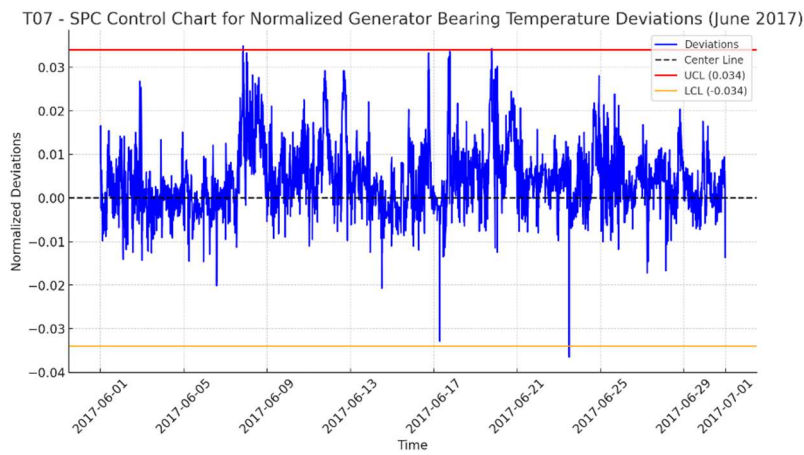
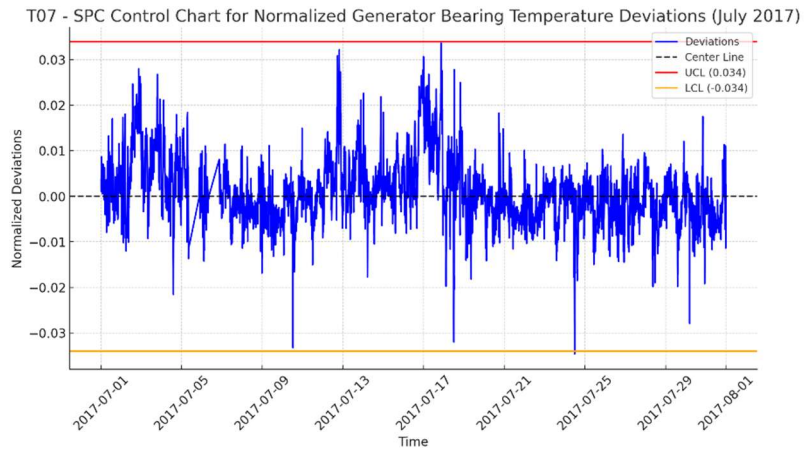
Παρατηρούμε ότι το μοντέλο επιτυγχάνει υψηλή ακρίβεια στην πρόβλεψη της θερμοκρασίας της γεννήτριας, όπως φαίνεται από το χαμηλό MSE και το MAE με κανονικοποιημένες τιμές. Το $R^2 = 0.9247$ υποδηλώνει πολύ καλή προσαρμογή, ενώ το διάγραμμα πραγματικών - προβλεπόμενων τιμών δείχνει ότι οι προβλέψεις είναι κοντά στη διαγώνιο, με ελάχιστες αποκλίσεις, επιβεβαιώνοντας τη σταθερότητα και την αξιοπιστία του μοντέλου.

Διάγραμμα SPC

Από το ιστορικό βλαβών, γνωρίζουμε ότι η βλάβη στη γεννήτρια της ανεμογεννήτριας T07 καταγράφηκε στις 20/08/2017 06:08. Το διάγραμμα SPC (Εικόνα 33) παρουσιάζει τις κανονικοποιημένες αποκλίσεις μεταξύ πραγματικών και προβλεπόμενων τιμών θερμοκρασίας της γεννήτριας κατά τη διάρκεια του Αυγούστου 2017. Από το διάγραμμα, παρατηρούνται έντονες αποκλίσεις ήδη από τις αρχές Αυγούστου, γεγονός που αποτελεί ένδειξη πρώιμων δυσλειτουργιών. Σημαντικές αποκλίσεις εκτός ορίων καταγράφονται επίσης στις 12 και 13 Αυγούστου, μια εβδομάδα πριν από τη βλάβη, επιβεβαιώνοντας τη σταδιακή επιδείνωση της κατάστασης. Η ανάλυση δείχνει ότι το μοντέλο ήταν σε θέση να προβλέψει τη βλάβη περίπου δύο εβδομάδες νωρίτερα, με διαδοχικά προειδοποιητικά σήματα τις τελευταίες ημέρες πριν από το συμβάν, επιτρέποντας την έγκαιρη διάγνωση και πιθανή εφαρμογή προγνωστικής συντήρησης.



Εικόνα 33. Διάγραμμα SPC για την θερμοκρασία της γεννήτριας T07



Εικόνα 34. Διαγράμματα SPC για τη θερμοκρασία της γεννήτριας της T07 (Ιούλιος - Αύγουστος 2017)

Τα διαγράμματα SPC των μηνών Ιουνίου και Ιουλίου 2017 (Εικόνα 34) επιβεβαιώνουν ότι η λειτουργία της γεννήτριας της ανεμογεννήτριας T07 παρέμενε σταθερή πριν την εμφάνιση του σφάλματος. Οι αποκλίσεις κατά το διάστημα αυτό ήταν περιορισμένες και διάσπαρτες, χωρίς να παρουσιάζουν διαδοχικές υπερβάσεις των ορίων ελέγχου. Η απουσία έντονων ανωμαλιών ενισχύει την αξιοπιστία του μοντέλου πρόβλεψης, καθώς δεν εντοπίστηκαν ψευδείς συναγερμοί. Η σύγκριση με το διάγραμμα του Αυγούστου καταδεικνύει ότι η σταδιακή επιδείνωση ξεκίνησε μόλις λίγες εβδομάδες πριν τη βλάβη, γεγονός που επιβεβαιώνει την προγνωστική ικανότητα της μεθόδου.

6.3 Περιορισμοί και Μελλοντικές Βελτιώσεις

Περιορισμοί

Παρά τα θετικά αποτελέσματα της ανάλυσης, υπάρχουν ορισμένοι περιορισμοί που πρέπει να ληφθούν υπόψη:

1. Περιορισμένη διαθεσιμότητα δεδομένων και ιστορικού βλαβών: Η ανάλυση βασίστηκε σε δεδομένα SCADA από συγκεκριμένο αριθμό ανεμογεννητριών και για

περιορισμένο χρονικό διάστημα, γεγονός που μπορεί να επηρεάσει τη γενίκευση των αποτελεσμάτων. Επιπλέον, στις ανεμογεννήτριες χωρίς καταγεγραμμένο ιστορικό βλαβών, δεν ήταν δυνατό να επιβεβαιωθεί αν οι ανιχνευθείσες αποκλίσεις σχετίζονται με μελλοντικές δυσλειτουργίες. Η χρήση ενός εκτενέστερου συνόλου δεδομένων, τόσο σε χρονική διάρκεια όσο και σε αριθμό ανεμογεννητριών, θα μπορούσε να βελτιώσει την ακρίβεια του μοντέλου και να προσφέρει μια πιο ολοκληρωμένη αξιολόγηση της απόδοσής του.

2. Στατικά όρια στα διαγράμματα SPC: Τα όρια ελέγχου στα SPC καθορίστηκαν με βάση τα δεδομένα της πρώτης χρονιάς και παρέμειναν σταθερά σε όλη τη διάρκεια της ανάλυσης. Αυτό μπορεί να οδηγήσει είτε σε ψευδείς συναγερμούς είτε σε καθυστερημένη ανίχνευση πραγματικών ανωμαλιών, καθώς δεν λαμβάνονται υπόψη δυναμικές αλλαγές στις συνθήκες λειτουργίας των ανεμογεννητριών.

3. Παρουσία σφαλμάτων που δεν σχετίζονται με τις εξεταζόμενες παραμέτρους: Ορισμένες αποκλίσεις στις μετρήσεις μπορεί να μην οφείλονται σε πραγματικά προβλήματα στη λειτουργία των ανεμογεννητριών, αλλά σε τυχαίους ή εξωτερικούς παράγοντες που δεν συνδέονται άμεσα με τις παραμέτρους που εξετάζονται. Σφάλματα στους αισθητήρες, διακοπές στην καταγραφή δεδομένων ή μεταβολές στις λειτουργικές ρυθμίσεις των ανεμογεννητριών μπορεί να οδηγήσουν σε εσφαλμένες ενδείξεις ανωμαλιών, αυξάνοντας τον κίνδυνο ψευδών συναγερμών.

4. Ποιότητα των δεδομένων SCADA: Παρόλο που έγινε προεπεξεργασία και καθαρισμός των δεδομένων SCADA, η ακρίβεια των προβλέψεων εξακολουθεί να επηρεάζεται από την ποιότητα των διαθέσιμων δεδομένων. Πιθανές ασυνέπειες, ελλείψεις ή μη καταγεγραμμένες κρίσιμες πληροφορίες μπορεί να περιορίσουν τη δυνατότητα του μοντέλου να ανιχνεύσει έγκαιρα τις ανωμαλίες.

Μελλοντικές Προοπτικές και Εξελίξεις

1. Εμπλουτισμός του συνόλου δεδομένων: Η συλλογή περισσότερων δεδομένων SCADA από μεγαλύτερο αριθμό ανεμογεννητριών και για μεγαλύτερο χρονικό διάστημα θα μπορούσε να βελτιώσει τη γενίκευση των αποτελεσμάτων. Επιπλέον, η ενσωμάτωση ιστορικού βλαβών θα ενισχύσει την ικανότητα πρόβλεψης του μοντέλου.

2. Ενσωμάτωση επιπλέον λειτουργικών και περιβαλλοντικών παραμέτρων: Η πρόβλεψη της θερμοκρασίας της γεννήτριας και του κιβωτίου ταχυτήτων μπορεί να βελτιωθεί με τη χρήση πρόσθετων παραμέτρων. Η ανάλυση αυτών των μεταβλητών θα μπορούσε να αποκαλύψει μοτίβα που σχετίζονται με πρώιμα σημάδια βλαβών.

3. Βελτίωση των διαγραμμάτων SPC με δυναμική προσαρμογή: Η ανάπτυξη προσαρμοστικών ορίων ελέγχου (UCL & LCL) στα διαγράμματα SPC, βασισμένων σε

στατιστικές κατανομές και σε συνθήκες λειτουργίας των ανεμογεννητριών, θα μπορούσε να μειώσει τα ψευδώς θετικά αποτελέσματα και να αυξήσει την αξιοπιστία των ανιχνεύσεων ανωμαλιών.

4. Συνεχής Βελτίωση μέσω Feedback Loop και Online Learning: Η εφαρμογή μηχανισμών συνεχούς μάθησης (online learning) θα μπορούσε να επιτρέψει στο μοντέλο να προσαρμόζεται διαρκώς σε νέα δεδομένα και να βελτιώνει την απόδοσή του με την πάροδο του χρόνου. Με την ενσωμάτωση feedback από τις προβλέψεις του μοντέλου και τις πραγματικές καταγεγραμμένες βλάβες, το σύστημα μπορεί να ανανεώνεται συνεχώς, ενισχύοντας τη μακροπρόθεσμη ακρίβεια και αποτελεσματικότητά του. Επιπλέον, η δυνατότητα συνεχούς ενημέρωσης των παραμέτρων του μοντέλου, βάσει των νέων μετρήσεων και των επιβεβαιωμένων βλαβών, θα επέτρεπε τη βελτίωση της ικανότητάς του να αναγνωρίζει ανωμαλίες σε πραγματικό χρόνο, μειώνοντας τους ψευδείς συναγερμούς.

5. Συνδυασμός μεθόδων μηχανικής μάθησης για βελτιωμένη απόδοση : Η χρήση υβριδικών μοντέλων, τα οποία συνδυάζουν διαφορετικές μεθόδους μηχανικής μάθησης, θα μπορούσε να βελτιώσει την ακρίβεια των προβλέψεων και να μειώσει τα σφάλματα. Για παράδειγμα, ο συνδυασμός XGBoost με LSTM μπορεί να εκμεταλλευτεί τόσο τη δύναμη των δέντρων απόφασης στην εξαγωγή σημαντικών χαρακτηριστικών όσο και την ικανότητα των νευρωνικών δικτύων LSTM να αναλύουν χρονοσειρές δεδομένων.

6.4 Συνολικά Συμπεράσματα

Η ανάλυση των αποτελεσμάτων έδειξε ότι το XGBoost αποδίδει αξιόπιστα και στις δύο περιπτώσεις που εξετάστηκαν, αποδεικνύοντας τη χρησιμότητά του στην προγνωστική συντήρηση των ανεμογεννητριών. Στις ανεμογεννήτριες χωρίς ιστορικό βλαβών, το μοντέλο πέτυχε χαμηλά σφάλματα πρόβλεψης (MSE, MAE, MSE%) και υψηλό συντελεστή προσδιορισμού R^2 , διατηρώντας σταθερή απόδοση στα δεδομένα δοκιμής. Αν και οι αποκλίσεις που εντοπίστηκαν μέσω των διαγραμμάτων SPC δεν συνδέθηκαν άμεσα με επιβεβαιωμένες βλάβες, καθώς υπήρχε απουσία ιστορικού βλαβών, θα μπορούσαν να αξιοποιηθούν ως δείκτες πρώιμων ενδείξεων φθοράς, συμβάλλοντας στη βελτίωση της προληπτικής συντήρησης.

Στις ανεμογεννήτριες με ιστορικό βλαβών, το μοντέλο απέδειξε την προγνωστική του ικανότητα, καθώς οι αποκλίσεις στις προβλέψεις εμφανίστηκαν αρκετές ημέρες πριν από τις καταγεγραμμένες αστοχίες. Η σταδιακή αύξηση της διαφοράς μεταξύ πραγματικών και προβλεπόμενων τιμών, όπως αποτυπώθηκε στα διαγράμματα SPC, επιβεβαίωσε τη δυνατότητα έγκαιρης διάγνωσης πιθανών δυσλειτουργιών. Το γεγονός ότι οι αποκλίσεις αυτές προέκυψαν πριν την εμφάνιση των πραγματικών βλαβών υποδηλώνει ότι η προτεινόμενη μέθοδος μπορεί να χρησιμοποιηθεί για την

έγκαιρη ανίχνευση ανωμαλιών, επιτρέποντας τη λήψη προληπτικών μέτρων πριν από την εμφάνιση σοβαρών βλαβών.

Η σταθερότητα των προβλέψεων και η απουσία φαινομένων υπερπροσαρμογής (overfitting) ενισχύουν περαιτέρω την αξιοπιστία της προσέγγισης. Η χρήση των διαγραμμάτων SPC σε συνδυασμό με το XGBoost αποδείχθηκε ιδιαίτερα αποτελεσματική, καθώς όχι μόνο κατέγραψε τις αποκλίσεις από τη φυσιολογική λειτουργία, αλλά και παρείχε μια σαφή μέθοδο εντοπισμού πιθανών προβλημάτων.

Συνολικά, η συγκριτική ανάλυση ανέδειξε τη δυνατότητα της προτεινόμενης μεθοδολογίας να λειτουργήσει τόσο ως εργαλείο παρακολούθησης κρίσιμων λειτουργικών παραμέτρων όσο και ως μηχανισμός προγνωστικής συντήρησης. Η ενσωμάτωση του XGBoost σε συνδυασμό με τα διαγράμματα SPC μπορεί να αποτελέσει μια αξιόπιστη στρατηγική για τον εντοπισμό πρώιμων ανωμαλιών, τη μείωση των αιφνίδιων αστοχιών και τη βελτιστοποίηση της λειτουργίας των ανεμογεννητριών. Η αξιοποίηση τέτοιων προγνωστικών εργαλείων μπορεί να συμβάλει στη μείωση του κόστους συντήρησης και στην αύξηση της απόδοσης των ανεμογεννητριών, διασφαλίζοντας πιο αποδοτική και αξιόπιστη λειτουργία των αιολικών πάρκων.

Η διπλωματική αυτή επιβεβαιώνει ότι η χρήση αλγορίθμων μηχανικής μάθησης, όπως το XGBoost, σε συνδυασμό με στατιστικές τεχνικές ανάλυσης, μπορεί να προσφέρει σημαντικές βελτιώσεις στην παρακολούθηση και τη συντήρηση των ανεμογεννητριών. Παρόλο που υπάρχουν ορισμένοι περιορισμοί που πρέπει να αντιμετωπιστούν, οι μελλοντικές βελτιώσεις, όπως η ενσωμάτωση περισσότερων λειτουργικών παραμέτρων και η εφαρμογή τεχνικών συνεχούς μάθησης (online learning), μπορούν να καταστήσουν τη μεθοδολογία ακόμη πιο αποτελεσματική.

Παράρτημα : Περιγραφή των Βασικών Τμημάτων του Κώδικα

Η παρούσα ενότητα περιγράφει τη βασική λειτουργία του κώδικα που χρησιμοποιήθηκε για την ανάλυση και την πρόβλεψη σφαλμάτων στις ανεμογεννήτριες, αξιοποιώντας δεδομένα SCADA και τον αλγόριθμο XGBoost. Ο κώδικας αναπτύχθηκε σε γλώσσα Python και εκτελέστηκε σε περιβάλλον Google Colab για ποια εύκολη φόρτωση, επεξεργασία και ανάλυση των δεδομένων. Η χρήση του Google Colab επιτρέπει την απευθείας σύνδεση με το Google Drive, εξασφαλίζοντας την εύκολη πρόσβαση στο αρχείων δεδομένων και την εκτέλεση των υπολογισμών χωρίς την ανάγκη τοπικής αποθήκευσης.

1) Βασικές Βιβλιοθήκες

Για την ανάλυση των δεδομένων SCADA και την εκπαίδευση του μοντέλου μηχανικής μάθησης, χρησιμοποιήθηκαν οι παρακάτω βασικές βιβλιοθήκες της Python :

- **Pandas** : χρησιμοποιείται για την επεξεργασία και την διαχείριση δεδομένων καθώς και την ανάγνωση αρχείων Excel.
- **NumPy** : παρέχει λειτουργίες για αριθμητικούς υπολογισμούς και διαχείριση πολυδιάστατων πινάκων.
- **Matplotlib** : χρησιμοποιείται για την δημιουργία γραφημάτων και την οπτικοποίηση δεδομένων, όπως οι πίνακες συσχέτισης.
- **Scikit-learn**: Περιλαμβάνει εργαλεία για την προεπεξεργασία των δεδομένων, την κανονικοποίηση χαρακτηριστικών και την αξιολόγηση του μοντέλου μέσω μετρικών όπως MSE και MAE.
- **XGBoost**: χρησιμοποιείται για την πρόβλεψη της θερμοκρασίας της γεννήτριας και του κιβωτίου ταχυτήτων, προσφέροντας υψηλή ακρίβεια και αποδοτικότητα.
- **Google Colab Drive**: χρησιμοποιείται για την πρόσβαση και τη διαχείριση αρχείων που βρίσκονται στο Google Drive.

2) Περιγραφή των Βασικών Βημάτων και Λειτουργιών του Κώδικα

Ο κώδικας οργανώνεται σε τρία κύρια τμήματα: την προεπεξεργασία δεδομένων, την εκπαίδευση μοντέλων και την οπτικοποίηση αποτελεσμάτων. Ξεκινά με τη φόρτωση και τον καθαρισμό των δεδομένων SCADA, καθώς και την επιλογή κρίσιμων χαρακτηριστικών για την ανάλυση. Στη συνέχεια, ο αλγόριθμος XGBoost εκπαιδεύεται για την πρόβλεψη κρίσιμων παραμέτρων, όπως η θερμοκρασία του κιβωτίου ταχυτήτων και της γεννήτριας, χρησιμοποιώντας φιλτραρισμένα και

κανονικοποιημένα δεδομένα. Τα αποτελέσματα αξιολογούνται μέσω δεικτών απόδοσης, όπως RMSE και MAE, και απεικονίζονται μέσω διαγραμμάτων, όπως scatter plots και SPC διαγράμματα, για την ανίχνευση αποκλίσεων και πιθανών ανωμαλιών.

2.1 Προεπεξεργασία Δεδομένων και Δημιουργία Καμπυλών Ισχύος

Σε αυτή την ενότητα πραγματοποιείται ο καθαρισμός και η φιλτράρισμα των δεδομένων SCADA, καθώς και η δημιουργία καμπυλών ισχύος για την ανάλυση της σχέσης μεταξύ ταχύτητας ανέμου και παραγόμενης ισχύος.

- **Σύνδεση με Google Drive και Φόρτωση Δεδομένων**

Για την πρόσβαση στα δεδομένα SCADA, το αρχείο Excel φορτώνεται από το Google Drive.

```
drive.mount('/content/drive')
df = pd.read_excel('/content/drive/My Drive/SCADA DATA PPCR..xlsx', sheet_name="ALARM LOG 2022")
```

- **Επιλογή χαρακτηριστικών**

Επιλέγονται οι στήλες που μας ενδιαφέρουν και αποθηκεύονται σε DataFrame για πιο εύκολη και γρήγορη επεξεργασία και ανάλυση.

```
columns_to_extract = ["NacelleTempAvg", "RPMAvg", "ActivePower", "OilTempAvg", "BearingTemp", "PCTimeStamp"]
df = df[columns_to_extract]
```

- **Δημιουργία καμπυλών ισχύος**

Η καμπύλη ισχύος δημιουργείται με τη χρήση scatter plot που απεικονίζει τη σχέση μεταξύ ταχύτητας ανέμου και παραγόμενης ισχύος, φιλτράροντας τα δεδομένα όπου το AlarmActive είναι μηδέν.

```
def plot_power_curve(df, wind_col, power_col):
    filtered_df = df[df['AlarmActive'] == 0]
    plt.scatter(filtered_df[wind_col], filtered_df[power_col], alpha=0.5, color='blue')
    plt.title('Καμπύλη Ισχύος Πριν τον Καθαρισμό')
    plt.xlabel('Ταχύτητα Ανέμου (m/s)')
    plt.ylabel('Ισχύς (kW)')
    plt.show()

# Παράδειγμα χρήσης:
plot_power_curve(data, 'Wind Speed', 'Power')
```

- **Εφαρμογή quantiles για τον Καθαρισμό της Καμπύλης Ισχύος**

Με την εφαρμογή quantiles (0.17 για το κάτω όριο και 0.95 για το πάνω όριο), απομακρύνονται οι ακραίες τιμές, βελτιώνοντας την ακρίβεια της καμπύλης ισχύος.

```
quantile_low, quantile_high = df['Power'].quantile([0.17, 0.95])
filtered_df = df[(df['Power'] >= quantile_low) & (df['Power'] <= quantile_high)]
```

2.2 Εκπαίδευση Μοντέλου Πρόβλεψης

Σε αυτή την ενότητα πραγματοποιείται το φιλτράρισμα, η προετοιμασία και η εκπαίδευση του μοντέλου πρόβλεψης, συμπεριλαμβανομένου του διαχωρισμού και της κανονικοποίησης των δεδομένων. Για την εκπαίδευση θα χρησιμοποιηθεί ο αλγόριθμος XGBoost.

- **Φιλτράρισμα Δεδομένων**

Επιλέγονται οι στήλες που μας αφορούν, ενώ απορρίπτονται σειρές όπου το AlarmActive είναι ενεργό. Επιπλέον, αφαιρούνται κενές τιμές και τα καθαρισμένα δεδομένα συνδυάζονται σε ένα νέο DataFrame για τη δημιουργία ενός ολοκληρωμένου συνόλου, διευκολύνοντας την ανάλυση και την εκπαίδευση του μοντέλου.

```
# Επιλογή κρίσιμων στηλών
columns_to_extract = ["NacelleTempAvg", "RPMAvg", "ActivePower", "OilTempAvg", "BearingTemp", "PCTimeStamp", "AlarmActive"]
df_selected = df[columns_to_extract]
# Φιλτράρισμα δεδομένων: Διατήρηση μόνο των σειρών όπου το AlarmActive είναι 0
df_filtered = df_selected[df_selected["AlarmActive"] == 0]
# Αφαίρεση σειρών με κενές τιμές
df_cleaned = df_filtered.dropna()
# Συνδυασμός δεδομένων σε νέο DataFrame
df_final = df_cleaned.reset_index(drop=True)
```

- **Διαχωρισμός Δεδομένων**

Το φιλτραρισμένο DataFrame διαχωρίζεται σε χαρακτηριστικά (features) και στόχο (target). Στη συνέχεια, τα δεδομένα χωρίζονται σε εκπαιδευτικό (80%) και δοκιμαστικό (20%) σύνολο, ενώ το εκπαιδευτικό σύνολο διαχωρίζεται περαιτέρω σε εκπαίδευσης (80%) και επικύρωσης (20%).

```
# Ορισμός χαρακτηριστικών (features) και στόχου (target)
features = ["NacelleTempAvg", "RPMAvg", "ActivePower", "OilTempAvg", "BearingTemp"]
target = "GeneratorBearingTmpt"
X = df_final[features]
y = df_final[target]
# Διαχωρισμός σε εκπαιδευτικό (80%) και δοκιμαστικό (20%) σύνολο
X_train_val, X_test, y_train_val, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
# Διαχωρισμός του εκπαιδευτικού σε εκπαίδευσης (80%) και επικύρωσης (20%)
X_train, X_val, y_train, y_val = train_test_split(X_train_val, y_train_val, test_size=0.2, random_state=42)
```

- **Κανονικοποίηση Δεδομένων**

Τα δεδομένα κλιμακώνονται στο εύρος [0,1] χρησιμοποιώντας Min-Max Scaling, ώστε όλες οι παράμετροι να έχουν ισότιμη συνεισφορά στην εκπαίδευση.

```
# Συνάρτηση για κανονικοποίηση δεδομένων
def manual_minmax_scaling(data):
    max_values = data.max(axis=0) # Υπολογισμός μέγιστων τιμών
    scaled_data = data / max_values # Κανονικοποίηση
    return scaled_data, max_values

# Κανονικοποίηση χαρακτηριστικών και στόχων
X_train_scaled, max_X = manual_minmax_scaling(X_train)
X_val_scaled = X_val / max_X
X_test_scaled = X_test / max_X
y_train_scaled, max_y = manual_minmax_scaling(y_train)
y_val_scaled = y_val / max_y
y_test_scaled = y_test / max_y
```

- **Εκπαίδευση του XGboost**

Το μοντέλο XGBoost εκπαιδεύεται με τα κανονικοποιημένα δεδομένα, ενώ ορίζονται υπερπαραμέτροι όπως το `max_depth`, το `learning_rate` (`eta`), και οι παράμετροι κανονικοποίησης `alpha` και `lambda`. Πριν την εκπαίδευση, τα δεδομένα μετατρέπονται σε `DMatrix` format, το οποίο βελτιστοποιεί τη διαχείριση της μνήμης και επιταχύνει τους υπολογισμούς. Κατά τη διαδικασία, εφαρμόζεται `early stopping`, το οποίο σταματά την εκπαίδευση αν η απόδοση στο σετ επικύρωσης δεν βελτιώνεται μετά από συγκεκριμένο αριθμό γύρων.

Ορισμός υπερπαραμέτρων

```
params = {
    "objective": "reg:squarederror",
    "eval_metric": "rmse",
    "max_depth": 8,
    "eta": 0.05, # Ρυθμός εκμάθησης
    "subsample": 0.8,
    "colsample_bytree": 0.8,
    "lambda": 1, # L2 κανονικοποίηση
    "alpha": 0.5, # L1 κανονικοποίηση
    "seed": 42,
}

# Μετατροπή των δεδομένων σε DMatrix format
train_data = xgb.DMatrix(X_train_scaled, label=y_train_scaled)
# Εκπαίδευση του μοντέλου με early stopping
model = xgb.train(
    params,
    train_data,
    num_boost_round=1000,
    evals=[(train_data, "train"), (val_data, "validation")],
    early_stopping_rounds=50,
    verbose_eval=50,
)
```

2.3 Αξιολόγηση Απόδοσης και Ανάλυση Αποτελεσμάτων με Διαγράμματα SPC

Σε αυτή την ενότητα, αξιολογείται η απόδοση του μοντέλου πρόβλεψης μέσω του RMSE και άλλων δεικτών, ενώ η ακρίβεια του αναλύεται με γραφήματα πραγματικών και προβλεπόμενων τιμών, καθώς και διαγράμματα SPC για την ανίχνευση αποκλίσεων και ανωμαλιών.

- **Υπολογισμός Μετρικών Αξιολόγησης**

Μετά την εκπαίδευση, το μοντέλο αξιολογείται χρησιμοποιώντας το σετ δοκιμής. Η απόδοσή του μετράται μέσω RMSE, ο οποίος υπολογίζει τη μέση απόκλιση των προβλέψεων από τις πραγματικές τιμές.

```
# Μετατροπή των δεδομένων του test set σε DMatrix format
test_data = xgb.DMatrix(X_test_scaled)
# Πρόβλεψη στο test set
predictions = model.predict(test_data)
# Υπολογισμός RMSE
rmse = np.sqrt(mean_squared_error(y_test_scaled, predictions))
# Εμφάνιση αποτελέσματος
print(f"RMSE: {rmse:.4f}")
```

- **Οπτικοποίηση**

Προβλέψεων

Η ακρίβεια του μοντέλου αναλύεται με scatter plot, συγκρίνοντας τις πραγματικές με τις προβλεπόμενες τιμές.

```
# Scatter plot πραγματικών vs προβλεπόμενων τιμών
plt.figure(figsize=(8, 6))
plt.scatter(y_test_scaled, predictions, alpha=0.6, edgecolors='k')
plt.plot([min(y_test_scaled), max(y_test_scaled)], [min(y_test_scaled), max(y_test_scaled)], 'r--', linewidth=2)
plt.xlabel("Πραγματικές Τιμές")
plt.ylabel("Προβλεπόμενες Τιμές")
plt.title("Πραγματικές vs Προβλεπόμενες Τιμές")
plt.grid(True)
plt.show()
```

- **Διαγράμματα SPC για Ανίχνευση Ανωμαλιών**

Η ανάλυση αποκλίσεων μεταξύ πραγματικών και προβλεπόμενων τιμών υλοποιείται με τη χρήση διαγραμμάτων SPC, τα οποία βοηθούν στην ανίχνευση ανωμαλιών. Υπολογίζονται οι ζεύγη διαδοχικών διαφορών (pairwise διαφορές), το MR (Moving Range), το σάγμα και τα όρια ελέγχου (UCL και LCL). Ο στόχος είναι η αξιολόγηση της σταθερότητας και της ακρίβειας του μοντέλου πρόβλεψης. Παρακάτω παρουσιάζεται ο σχετικός κώδικας.

```
# Υπολογισμός αποκλίσεων (Πραγματικές - Προβλεπόμενες)
deviations = y_test_scaled - predictions
# Υπολογισμός pairwise διαφορών
delta_pairwise_diff = [abs(deviations[i] - deviations[i - 1]) for i in range(1, len(deviations))]
MR_corrected_pairwise = np.mean(delta_pairwise_diff)
# Υπολογισμός των ορίων ελέγχου
sigma_pairwise = MR_corrected_pairwise / 1.128
UCL_pairwise = 3 * sigma_pairwise
LCL_pairwise = -3 * sigma_pairwise
# Σχεδίαση διαγράμματος SPC
plt.figure(figsize=(12, 6))
plt.plot(deviations, label="Αποκλίσεις (Πραγματικές - Προβλεπόμενες)", color='blue', linewidth=1)
plt.axhline(UCL_pairwise, color='red', linestyle='-', label=f'UCL = {UCL_pairwise:.3f}')
plt.axhline(LCL_pairwise, color='gold', linestyle='-', label=f'LCL = {LCL_pairwise:.3f}')
plt.axhline(0, color='black', linestyle='--', label='Κεντρική Γραμμή')
# Προσθήκη τίτλου, ετικετών και υπομνήματος
plt.title("Διάγραμμα SPC - Ανίχνευση Ανωμαλιών με Pairwise Διαφορές", fontsize=14)
plt.xlabel("Δείκτες Δεδομένων", fontsize=12)
plt.ylabel("Απόκλιση Θερμοκρασίας (°C)", fontsize=12)
plt.legend()
plt.grid(True)
plt.show()
```

Βιβλιογραφικές Αναφορές – Πηγές

- [1] Global Wind Energy Council (GWEC), Global Wind Report 2024, GWEC, Brussels, 2024.
- [2] R. J. Barthelmie and S. C. Pryor, "Potential for efficiency improvements in wind turbine operations," Renewable Energy, 2020.
- [3] International Renewable Energy Agency (IRENA), Wind Energy Operations and Maintenance Costs, 2021.
- [4] X. Jin and W. Qiao, "Feature Transfer Learning for Wind Turbine Condition Monitoring," IEEE Access, 2023.
- [5] X. Gao, L. Li, and H. Tian, "Review of SCADA data applications in wind turbine health monitoring," Journal of Renewable Energy Engineering, 2019.
- [6] Σ. Παπαθανασίου, ΑΠΕ - Ενεργειακή Απόδοση Ανανεώσιμων Πηγών Ενέργειας, Εθνικό Μετσόβιο Πολυτεχνείο (ΕΜΠ).
- [7] Hellenic Wind Energy Association (ELETAEN), Wind Energy Statistics in Greece for 2024.
- [8] U.S. Department of Energy, How Do Wind Turbines Work.
- [9] Enel Green Power: Wind Turbine Components (enelgreenpower.com).
- [10] Windmills Tech: Wind Turbine Components (windmillstech.com).
- [11] J.-Y. Son and K. Ma, "Wind Energy Systems," *Proceedings of the IEEE*, vol. 105, no. 11, pp. 2116–2131, Nov. 2017, doi: <https://doi.org/10.1109/jproc.2017.2695485>.
- [12] "Most common reasons for wind turbine failures," *www.cotes.com*. <https://www.cotes.com/blog/most-common-reasons-for-wind-turbine-failures>.
- [13] "Ten most common wind turbine failure issues identified | Sirris," *www.sirris.be*, Nov. 22, 2021.
- [14] "Wind Turbine Failures: Causes, Consequences, and Impact on Energy Output," *Enlitia.com*, May 13, 2024. <https://www.enlitia.com/resources-blog-post/wind-turbine-failures-causes-consequences-and-impact-on-energy-output>.
- [15] P. Loshin, "What is SCADA (supervisory control and data acquisition)?," *WhatIs.com*, Dec. 2021. <https://www.techtarget.com/whatis/definition/SCADA-supervisory-control-and-data-acquisition>.
- [16] H. H. Hansen, "Predictive maintenance for power plants," *Welcome to DTU Research Database*, 2024. <https://orbit.dtu.dk/en/publications/predictive-maintenance-for-power-plants>.
- [17] J. Lee, J. Ni, J. Singh, B. Jiang, M. Azamfar, and J. Feng, "Intelligent Maintenance Systems and Predictive Manufacturing," *Journal of Manufacturing Science and Engineering*, vol. 142, no. 11, Nov. 2020, doi: <https://doi.org/10.1115/1.4047856>.

- [18] N. T. Trần, H. T. Triệu, V. T. Trần, H. H. Ngô, and Q. K. Đào, "An overview of the application of machine learning in predictive maintenance," *Petrovietnam Journal*, vol. 10, pp. 47–61, Nov. 2021, doi: <https://doi.org/10.47800/pvj.2021.10-05>.
- [19] S. Brown, "Machine Learning, Explained," *MIT Sloan*, 2021. <https://mitsloan.mit.edu/ideas-made-to-matter/machine-learning-explained>.
- [20] Santosh Kumar Singh, Arun Kumar Tiwari, and H.K. Paliwal, "A state-of-the-art review on the utilization of machine learning in nanofluids, solar energy generation, and the prognosis of solar power," *Engineering Analysis with Boundary Elements*, vol. 155, pp. 62–86, Oct. 2023, doi: <https://doi.org/10.1016/j.enganabound.2023.06.003>.
- [21] A. B. Nassif, M. A. Talib, Q. Nasir, and F. M. Dakalbab, "Machine Learning for Anomaly Detection: A Systematic Review," *IEEE Access*, vol. 9, pp. 78658–78700, 2021, doi: <https://doi.org/10.1109/access.2021.3083060>.
- [22] T. Chen and C. Guestrin, "XGBoost: a Scalable Tree Boosting System," *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '16*, pp. 785–794, 2016, doi: <https://doi.org/10.1145/2939672.2939785>.
- [23] A. Hachcham, "XGBoost: Everything You Need to Know," *Neptune.ai*, 2023. <https://neptune.ai/blog/xgboost-everything-you-need-to-know>.
- [24] "XGBoost Parameters — xgboost 1.5.2 documentation," *xgboost.readthedocs.io*. <https://xgboost.readthedocs.io/en/stable/parameter.html>.
- [25] Akhmal Angga Syahputra and Rujianto Eko Saputro, "Application of the XGBoost Model with Hyperparameter Tuning for Industry Classification for Job Applicants," *Sinkron*, vol. 8, no. 3, pp. 1920–1931, Jul. 2024, doi: <https://doi.org/10.33395/sinkron.v8i3.13840>.
- [26] RITHP, "Optimizing XGBoost: A Guide to Hyperparameter Tuning," *Medium*, Jan. 16, 2023. <https://medium.com/@rithpansanga/optimizing-xgboost-a-guide-to-hyperparameter-tuning-77b6e48e289d>.
- [27] "Wind Turbine Blade Icing Fault Prediction Based on SCADA Data by XGBoost," *Academic Journal of Engineering and Technology Science*, vol. 5, no. 2, 2022, doi: <https://doi.org/10.25236/ajets.2022.050206>.
- [28] A. Wan *et al.*, "Using Transfer Learning and XGBoost for Early Detection of Fires in Offshore Wind Turbine Units," *Energies*, vol. 17, no. 10, pp. 2330–2330, May 2024, doi: <https://doi.org/10.3390/en17102330>.
- [29] A. Kumar, "MSE vs RMSE vs MAE vs MAPE vs R-Squared: When to Use?," *Analytics Yogi*, Dec. 29, 2023. https://vitalflux.com/mse-vs-rmse-vs-mae-vs-mape-vs-r-squared-when-to-use/#google_vignette.
- [30] "A guide to using SPC charts and icons, and reading demand charts :: Lincolnshire Partnership NHS Trust," *www.lpft.nhs.uk*. <https://www.lpft.nhs.uk/about-us/how-we-are-performing/guide-using-spc-charts-and-icons-and-reading-demand-charts>.

[31] W. Udo and Y. Muhammad, "Data-Driven Predictive Maintenance of Wind Turbine Based on SCADA Data," *IEEE Access*, pp. 1–1, 2021, doi: <https://doi.org/10.1109/access.2021.3132684>.