



**ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ – ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ**

**Πρόγραμμα Μεταπτυχιακών Σπουδών  
«Κυβερνοασφάλεια και Επιστήμη Δεδομένων»**

**Μεταπτυχιακή Διατριβή**

Τίτλος Διατριβής	<b>Ενίσχυση της προγνωστικής συντήρησης στο IoT: Συγκριτική μελέτη αλγορίθμων ανίχνευσης ανωμαλιών με χρήση Python</b> <b>Enhancing Predictive Maintenance in IoT: A Comparative Study of Anomaly Detection Algorithms Using Python</b>
Όνοματεπώνυμο Φοιτητή	<b>Βύρων Δημάκος</b>
Πατρώνυμο	<b>Κωσταντίνος</b>
Αριθμός Μητρώου	<b>ΜΠΚΕΔ 21011</b>
Επιβλέπων	<b>Δουληγέρης Χρήστος, Καθηγητής</b>

Ημερομηνία Παράδοσης **Νοέμβριος 2024**

---

**Τριμελής Εξεταστική Επιτροπή**

(υπογραφή)

(υπογραφή)

(υπογραφή)

Χρήστος Δουληγέρης  
Καθηγητής

Μιχαήλ Ψαράκης  
Αναπληρωτής Καθηγητής

Παναγιώτης Κοτζανικολάου  
Καθηγητής

---

## Περίληψη

Η παρούσα διατριβή, "Ενίσχυση της προγνωστικής συντήρησης στο IoT: Συγκριτική μελέτη αλγορίθμων ανίχνευσης ανωμαλιών με χρήση Python", περιηγείται στις περιπλοκές της προβλεπτικής συντήρησης στο Διαδίκτυο των Πραγμάτων (IoT). Χρησιμοποιώντας τους αλγορίθμους k-Nearest Neighbors (kNN), Support Vector Machine (SVM), Random Forest και Long Short-Term Memory (LSTM) σε Python, η μελέτη αξιολογεί την απόδοσή τους στο σύνολο δεδομένων της NASA - που περιλαμβάνει μετρήσεις κινητήρων αεριωθούμενων αεροσκαφών. Η προσαρμοστικότητα της Python αναδεικνύεται στην προεπεξεργασία δεδομένων, στην υλοποίηση αλγορίθμων και στην οπτικοποίηση. Μετρώνται βασικές μετρικές απόδοσης - R2, RMSE, χρόνος εκπαίδευσης, χρόνος πρόβλεψης και συνολικός χρόνος - παρέχοντας insights στην αποτελεσματικότητα των αλγορίθμων. Η παρούσα έρευνα συμβάλλει στη βελτιστοποίηση των στρατηγικών προβλεπτικής συντήρησης με επίκεντρο την Python, προσφέροντας στους επαγγελματίες πολύτιμη καθοδήγηση σε πραγματικές εφαρμογές IoT.

## ΚΕΦΑΛΑΙΟ 1: Εισαγωγή

Το Διαδίκτυο των Πραγμάτων (IoT) έχει εγκαινιάσει μια νέα εποχή διασυνδεδεμένων συσκευών, δημιουργώντας τεράστιες ποσότητες δεδομένων που έχουν τη δυνατότητα να μεταμορφώσουν τις βιομηχανίες. Μια κρίσιμη εφαρμογή σε αυτό το τοπίο είναι η προγνωστική συντήρηση, ακρογωνιαίος λίθος για τη διασφάλιση της αξιοπιστίας και της μακροζωίας των συστημάτων IoT. Η παρούσα διατριβή, με τίτλο "Ενίσχυση της προγνωστικής συντήρησης στο IoT: Συγκριτική μελέτη αλγορίθμων ανίχνευσης ανωμαλιών με χρήση Python", ξεκινά μια διεξοδική διερεύνηση των θεωρητικών βάσεων και των πρακτικών εφαρμογών της προγνωστικής συντήρησης. Στο πλαίσιο αυτής της έρευνας, ένα ποικίλο σύνολο αλγορίθμων μηχανικής μάθησης - συγκεκριμένα οι αλγόριθμοι k-Nearest Neighbors (kNN), Support Vector Machine (SVM), Random Forest και Long Short-Term Memory (LSTM) - υλοποιούνται με τη χρήση της ευέλικτης γλώσσας προγραμματισμού Python. Το υπόβαθρο για την παρούσα έρευνα δημιουργείται μέσω της εξέτασης των προκλήσεων που είναι εγγενείς στην ανάλυση δεδομένων IoT. Συγκεκριμένα, η εστίαση έγκειται στην αξιολόγηση αυτών των αλγορίθμων με τη χρήση του συνόλου δεδομένων της NASA, το οποίο περιλαμβάνει μετρήσεις που λαμβάνονται από έναν κινητήρα τζετ. Αυτό το σύνολο δεδομένων παρέχει ένα πλαίσιο πραγματικού κόσμου για την αξιολόγηση της αποτελεσματικότητας των διαφόρων αλγορίθμων σε σενάρια προγνωστικής συντήρησης. Η Python αναδεικνύεται ως το κεντρικό εργαλείο για την προεπεξεργασία δεδομένων, την υλοποίηση αλγορίθμων και την οπτικοποίηση, αναδεικνύοντας την ικανότητά της στο χειρισμό των περιπλοκών των δεδομένων IoT. Η επακόλουθη ανάλυση περιλαμβάνει τη μέτρηση βασικών μετρικών επιδόσεων, όπως το R2 (συντελεστής προσδιορισμού), το RMSE (Root Mean Square Error), ο χρόνος εκπαίδευσης, ο χρόνος πρόβλεψης και ο συνολικός χρόνος. Ο στόχος δεν είναι μόνο η μέτρηση της αλγοριθμικής απόδοσης στο σύνολο δεδομένων της NASA, αλλά και η παροχή διορατικών οπτικοποιήσεων που προσφέρουν μια ολιστική κατανόηση των δυνατοτήτων τους. Με την πλοήγηση μέσω αυτής της ολοκληρωμένης εξερεύνησης, η διατριβή έχει ως στόχο να εφοδιάσει τους υπεύθυνους λήψης αποφάσεων και τους επαγγελματίες με αποχρώσεις για τη διαμόρφωση αποτελεσματικών στρατηγικών προγνωστικής συντήρησης προσαρμοσμένων σε περιβάλλοντα IoT με Python.

## ΚΕΦΑΛΑΙΟ 2: Θεωρητικά Θεμέλια της Προγνωστικής Συντήρησης

Η προγνωστική συντήρηση αντιπροσωπεύει μια αλλαγή παραδείγματος στις στρατηγικές συντήρησης, μεταβαίνοντας από τις παραδοσιακές αντιδραστικές και προληπτικές προσεγγίσεις σε ένα προληπτικό μοντέλο που αξιοποιεί προηγμένες τεχνολογίες, ιδίως στο Διαδίκτυο των Πραγμάτων (IoT). Αυτό το κεφάλαιο εμβαθύνει στα θεωρητικά θεμέλια που στηρίζουν την προγνωστική συντήρηση, τονίζοντας τη σημασία της για την ενίσχυση της αξιοπιστίας του συστήματος και την ελαχιστοποίηση των λειτουργικών διαταραχών.

### 2.1 Μοντέλα Συντήρησης: Αντιδραστικά, Προληπτικά και Προγνωστικά

Τα μοντέλα συντήρησης αποτελούν τη βάση των στρατηγικών που χρησιμοποιούνται για τη διαχείριση της υγείας και της απόδοσης των βιομηχανικών περιουσιακών στοιχείων. Αυτή η ενότητα εμβαθύνει στα τρία κύρια μοντέλα συντήρησης - αντιδραστική, προληπτική και πρόβλεψη - παρέχοντας μια λεπτομερή κατανόηση του καθενός.

#### 2.1.1 Αντιδραστική Συντήρηση

Η αντιδραστική συντήρηση, που συχνά αποκαλείται "run-to-failure", είναι μια παραδοσιακή προσέγγιση όπου οι παρεμβάσεις πραγματοποιούνται μόνο ως απόκριση σε αστοχία εξοπλισμού. Σε αυτό το μοντέλο, η εστίαση είναι στην επίλυση προβλημάτων όπως αυτά προκύπτουν, χωρίς προληπτικά μέτρα. Αν και αυτή η προσέγγιση μπορεί να φαίνεται οικονομικά αποδοτική βραχυπρόθεσμα, έχει σημαντικές ελλείψεις. Ο χρόνος διακοπής λειτουργίας λόγω απροσδόκητων βλαβών μπορεί να οδηγήσει σε απώλειες παραγωγής, αυξημένο κόστος επισκευής και καταπόνηση των πόρων. Η αντιδραστική συντήρηση στερείται πρόβλεψης και μπορεί να οδηγήσει σε έναν διαρκή κύκλο βλαβών και επισκευών. [1]

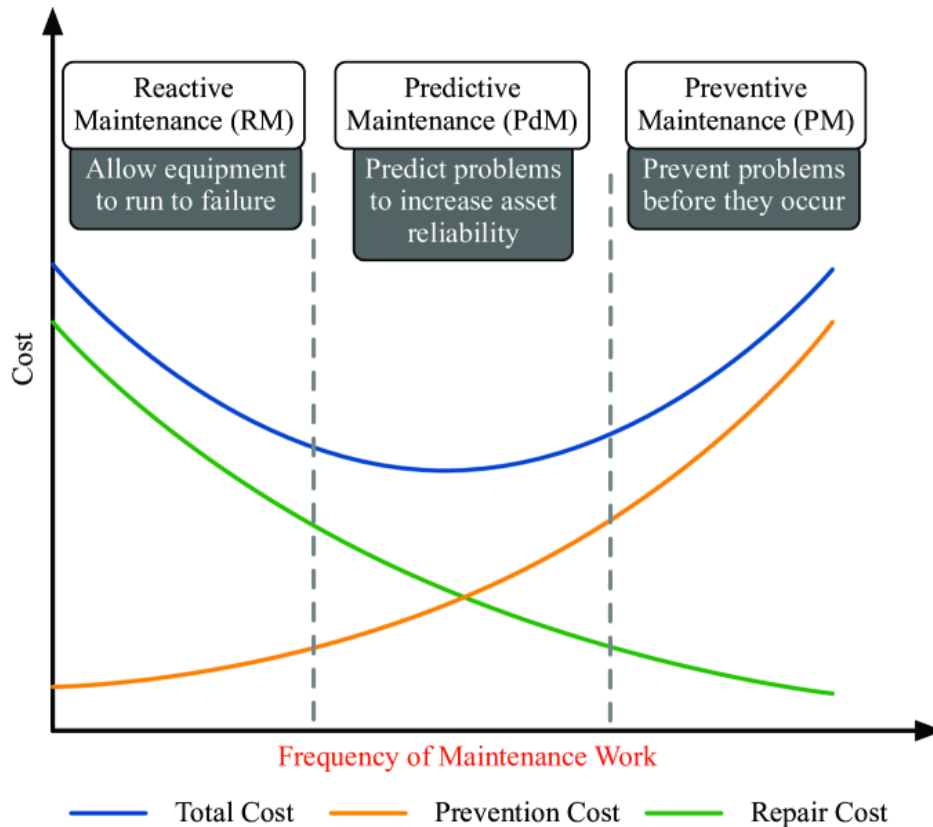
#### 2.1.2 Προληπτική Συντήρηση

Αντίθετα, η προληπτική συντήρηση υιοθετεί μια προγραμματισμένη και συστηματική προσέγγιση. Πραγματοποιούνται τακτικές επιθεωρήσεις, λίπανση και αντικαταστάσεις εξαρτημάτων σε προκαθορισμένα χρονικά διαστήματα για την αποτροπή πιθανών βλαβών. Αυτό το μοντέλο είναι μια βελτίωση σε σχέση με την αντιδραστική συντήρηση, καθώς στοχεύει στην αντιμετώπιση προβλημάτων προτού προκαλέσουν μεγάλες διακοπές. Ωστόσο, η προληπτική συντήρηση έχει τους περιορισμούς της. Συχνά βασίζεται σε σταθερά χρονικά διαστήματα ή όρια χρήσης, τα οποία μπορεί να οδηγήσουν σε περιττές παρεμβάσεις, ειδικά εάν τα εξαρτήματα δεν υποβαθμίζονται όπως αναμενόταν. Επιπλέον, ενδέχεται να παρουσιαστούν απροσδόκητες βλάβες μεταξύ των προγραμματισμένων συμβάντων συντήρησης. [1]

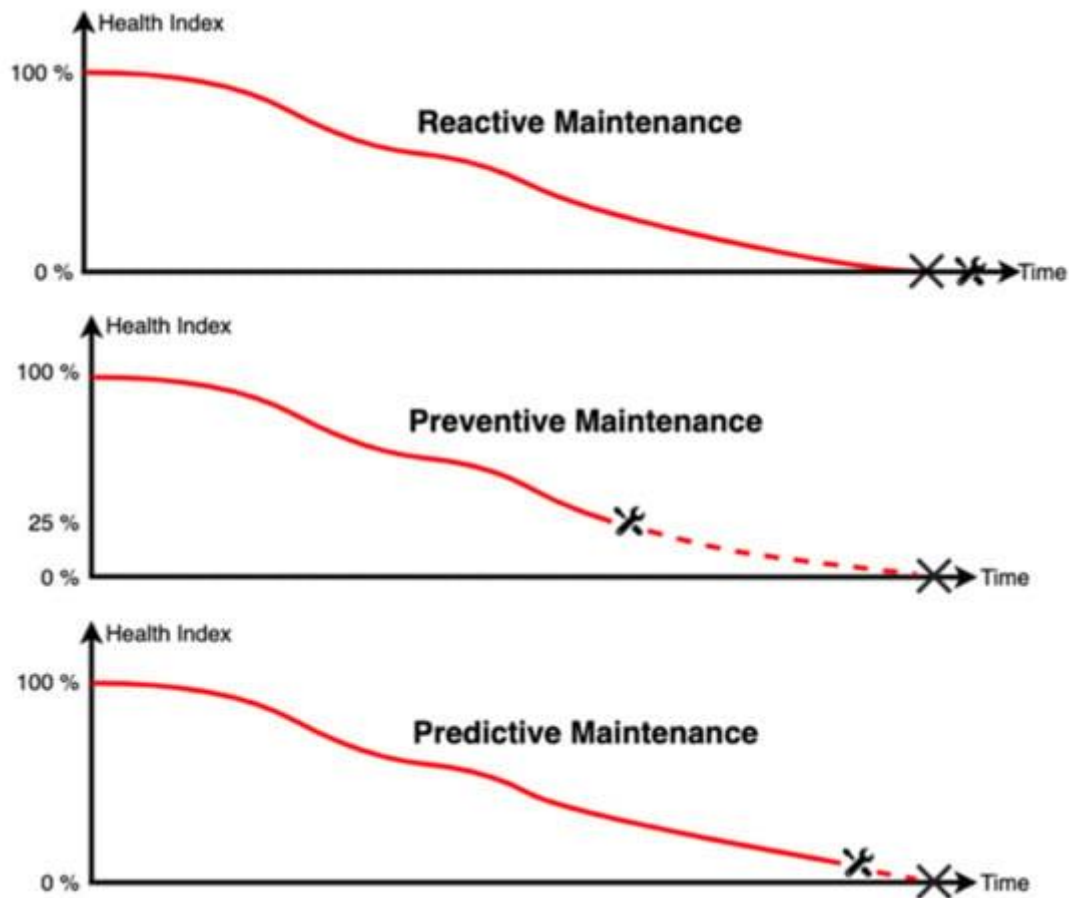
#### 2.1.3 Προγνωστική Συντήρηση

Η προγνωστική συντήρηση αντιπροσωπεύει μια αλλαγή παραδείγματος από τα αντιδραστικά και προληπτικά μοντέλα. Αυτή η προνοητική στρατηγική αξιοποιεί δεδομένα σε πραγματικό χρόνο και παρακολούθηση της διαδικασίας για να προβλέψει πότε ο εξοπλισμός είναι πιθανό να αποτύχει. Αναλύοντας δεδομένα από αισθητήρες και συσκευές IoT που είναι ενσωματωμένες σε μηχανήματα, οι αλγόριθμοι πρόβλεψης συντήρησης μπορούν να εντοπίσουν μοτίβα ενδεικτικά πιθανών προβλημάτων. Αυτή η προσέγγιση επιτρέπει επεμβάσεις ακριβώς όταν χρειάζεται, βελτιστοποιώντας τη διάρκεια ζωής των εξαρτημάτων και ελαχιστοποιώντας το χρόνο διακοπής

λειτουργίας. Ένα βασικό πλεονέκτημα της προγνωστικής συντήρησης είναι η ικανότητά της να μεταφέρει τις δραστηριότητες συντήρησης από ένα σταθερό χρονοδιάγραμμα σε ένα πιο δυναμικό σύστημα που βασίζεται σε συνθήκες. Η προσέγγιση μετριάζει τις αδυναμίες των αντιδραστικών και προληπτικών μοντέλων, επιτρέποντας την πιο αποτελεσματική κατανομή των πόρων. Επιπλέον, η προγνωστική συντήρηση διευκολύνει μια διαδικασία λήψης αποφάσεων που βασίζεται σε δεδομένα, ενισχύοντας την ακρίβεια των προβλέψεων αποτυχίας και μειώνοντας την πιθανότητα περιττών παρεμβάσεων.



Εικ. 1: Σύγκριση των 3 τύπων συντήρησης βάσει κόστους και συχνότητας των εργασιών



Εικ. 2: Σύγκριση των 3 τύπων συντήρησης βάσει του πότε πραγματοποιείται η συντήρηση

Στο πλαίσιο της διατριβής αυτής, η προγνωστική συντήρηση βρίσκεται στο επίκεντρο, ευθυγραμμισμένη με τις αρχές του Διαδικτύου των Πραγμάτων (IoT). Η ενσωμάτωση έξυπνων αισθητήρων και αναλύσεων δεδομένων σε πραγματικό χρόνο σε βιομηχανικό εξοπλισμό αποτελεί παράδειγμα της δυνατότητας προγνωστικής συντήρησης να φέρει επανάσταση στον τρόπο προσέγγισης των δραστηριοτήτων συντήρησης. Αυτό θέτει το έδαφος για μια βαθύτερη εξερεύνηση του ρόλου του IoT στην προγνωστική συντήρηση και την επακόλουθη αξιολόγηση των αλγορίθμων μηχανικής μάθησης χρησιμοποιώντας Python. [2]

## 2.2 Αρχές Προγνωστικής Συντήρησης

### 2.2.1 Προληπτική προσέγγιση

Στην καρδιά της προγνωστικής συντήρησης βρίσκεται η προληπτική φύση της. Σε αντίθεση με την αντιδραστική και προληπτική συντήρηση, οι οποίες είναι αντιδραστικές ή βάσει

χρονοδιαγράμματος, η προγνωστική συντήρηση προβλέπει βλάβες του εξοπλισμού πριν αυτές συμβούν. Αξιοποιώντας προηγμένα αναλυτικά στοιχεία και παρακολούθηση δεδομένων σε πραγματικό χρόνο, η προγνωστική συντήρηση μετατοπίζει την εστίαση από την αντίδραση σε ζητήματα στην πλήρη αποτροπή τους. Αυτή η προληπτική στάση επιτρέπει στις ομάδες συντήρησης να επέμβουν ακριβώς όταν είναι απαραίτητο, μεγιστοποιώντας το χρόνο λειτουργίας του στοιχείου και ελαχιστοποιώντας τον απρογραμματίστο χρόνο διακοπής λειτουργίας. [3]

### **2.2.2 Πληροφορίες βάσει δεδομένων**

Κεντρικό στοιχείο της προγνωστικής συντήρησης είναι η εξάρτηση από πληροφορίες που βασίζονται σε δεδομένα. Μέσω της χρήσης αισθητήρων, συσκευών IoT και άλλων τεχνολογιών παρακολούθησης, συλλέγονται τεράστιες ποσότητες δεδομένων από βιομηχανικό εξοπλισμό. Αυτά τα δεδομένα, που περιλαμβάνουν παραμέτρους όπως η θερμοκρασία, οι κραδασμοί και τα επίπεδα υγρών, παρέχουν πολύτιμες πληροφορίες για την υγεία και την απόδοση των στοιχείων. Με την ανάλυση αυτών των δεδομένων χρησιμοποιώντας προηγμένες τεχνικές ανάλυσης, οι αλγόριθμοι πρόβλεψης συντήρησης μπορούν να ανιχνεύσουν διακριτικά μοτίβα ενδεικτικά επικείμενων αστοχιών. Αυτές οι πληροφορίες που βασίζονται σε δεδομένα επιτρέπουν στις ομάδες συντήρησης να λαμβάνουν τεκμηριωμένες αποφάσεις, δίνοντας προτεραιότητα στις δραστηριότητες συντήρησης με βάση την πραγματική κατάσταση του εξοπλισμού και όχι σε προκαθορισμένα χρονοδιαγράμματα. [3]

### **2.2.3 Μετάβαση σε παρεμβάσεις βάσει συνθηκών**

Μία από τις θεμελιώδεις αρχές της προγνωστικής συντήρησης είναι η μετάβαση από τα σταθερά προγράμματα συντήρησης σε επεμβάσεις που βασίζονται σε συνθήκες. Τα παραδοσιακά μοντέλα συντήρησης βασίζονται σε προκαθορισμένα χρονικά διαστήματα ή όρια χρήσης για τον προγραμματισμό των δραστηριοτήτων συντήρησης. Ωστόσο, η προγνωστική συντήρηση ακολουθεί μια πιο δυναμική προσέγγιση, παρεμβαίνοντας μόνο όταν οι συνθήκες του εξοπλισμού το απαιτούν. Με τη συνεχή παρακολούθηση της υγείας του εξοπλισμού σε πραγματικό χρόνο, οι αλγόριθμοι πρόβλεψης συντήρησης μπορούν να ανιχνεύσουν έγκαιρα προειδοποιητικά σημάδια πιθανών αστοχιών. Αυτό επιτρέπει τον προγραμματισμό των δραστηριοτήτων συντήρησης με ακρίβεια όταν χρειάζεται, βελτιστοποιώντας την κατανομή πόρων και ελαχιστοποιώντας τις περιττές παρεμβάσεις. [3]

### **2.2.4 Βελτιστοποίηση των πόρων συντήρησης**

Η προγνωστική συντήρηση βελτιστοποιεί την κατανομή των πόρων συντήρησης εστιάζοντας τις προσπάθειες εκεί που χρειάζονται περισσότερο. Εντοπίζοντας εκ των προτέρων τις επικείμενες βλάβες, οι ομάδες συντήρησης μπορούν να ιεραρχήσουν τις δραστηριότητες με βάση την κρισιμότητα και τη σοβαρότητα. Αυτή η στοχευμένη προσέγγιση μειώνει την ανάγκη για αντιδραστικές παρεμβάσεις και ελαχιστοποιεί τον κίνδυνο περιττών εργασιών συντήρησης. Επιπλέον, η προγνωστική συντήρηση επιτρέπει στις ομάδες συντήρησης να σχεδιάζουν και να προγραμματίζουν τις δραστηριότητες πιο αποτελεσματικά, μειώνοντας το χρόνο διακοπής λειτουργίας και μεγιστοποιώντας την παραγωγικότητα του προσωπικού συντήρησης. [4]



## 2.3 ΣΧΕΤΙΚΕΣ ΜΕΛΕΤΕΣ

*Mobley, R. Keith (2002) – An Introduction to Predictive Maintenance:*

Αυτό το έργο σκιαγραφεί την εξέλιξη από την προληπτική στην προγνωστική συντήρηση και εξηγεί τις θεμελιώδεις αρχές του Predictive Maintenance.

*Jardine, A.K.S., Lin, D., & Banjevic, D. (2006) – A Review on Machinery Diagnostics and Prognostics Implementing Condition-Based Maintenance:*

Εξετάζει τον ρόλο των διαγνωστικών και προγνωστικών σε στρατηγικές συντήρησης που βασίζονται σε καταστάσεις και προγνωστικές.

## ΚΕΦΑΛΑΙΟ 3: IoT και Προγνωστική Συντήρηση

Προτού εμβαθύνουμε στην ανάλυση, θα πρέπει να διερευνηθεί το γενικότερο υπόβαθρο του Διαδικτύου των Πραγμάτων (IoT) και ο μετασχηματιστικός αντίκτυπός του σε βιομηχανίες σε όλο το φάσμα.

### 3.1 Εισαγωγή στο IoT και ο μετασχηματιστικός αντίκτυπός του

Η έλευση του IoT σηματοδοτεί μια αλλαγή παραδείγματος στο τεχνολογικό τοπίο, διεισδύοντας σε διάφορους τομείς και αλλάζοντας θεμελιωδώς τον τρόπο αλληλεπίδρασης και επικοινωνίας των συσκευών. Αυτός ο μετασχηματισμός έχει ξεπεράσει τα παραδοσιακά όρια, εγκαινιάζοντας μια εποχή όπου τα καθημερινά αντικείμενα είναι εμποτισμένα με τη δύναμη της συνδεσιμότητας, της ευφυΐας και της παραγωγής δεδομένων. Ο βαθύς αντίκτυπος του IoT εκτείνεται πολύ πέρα από την απλή τεχνολογική καινοτομία, διεισδύοντας σε βιομηχανίες που κυμαίνονται από την υγειονομική περίθαλψη έως την κατασκευή, τις μεταφορές και όχι μόνο. Στην καρδιά της επανάστασης του IoT βρίσκεται η έννοια της διασυνδεσιμότητας. Οι συσκευές, που προηγουμένως απομονώνονταν και λειτουργούσαν σε σιλό (αποκομμένες), είναι πλέον άψογα διασυνδεδεμένες, σχηματίζοντας έναν περίπλοκο ιστό επικοινωνίας. Αυτή η διασύνδεση δημιουργεί ένα οικοσύστημα όπου οι συσκευές μπορούν να μοιράζονται πληροφορίες, να ανταποκρίνονται σε δεδομένα σε πραγματικό χρόνο και να συνεργάζονται για την επίτευξη κοινών στόχων. Οι επιπτώσεις αυτού του μετασχηματισμού είναι τεράστιες, προσφέροντας άνευ προηγουμένου ευκαιρίες για αποτελεσματικότητα, αυτοματοποίηση και λήψη αποφάσεων βάσει δεδομένων. [5]

### 3.2 Εφαρμογές IoT στην Προγνωστική Συντήρηση

Οι βιομηχανίες έχουν γίνει μάρτυρες μιας μεταμόρφωσης καθώς το IoT έχει γίνει αναπόσπαστο μέρος του λειτουργικού τους ιστού. Στην κατασκευή, για παράδειγμα, αισθητήρες με δυνατότητα IoT που είναι ενσωματωμένοι σε μηχανήματα διευκολύνουν την παρακολούθηση σε πραγματικό χρόνο των διαδικασιών παραγωγής. Αυτοί οι αισθητήρες δημιουργούν μια συνεχή ροή δεδομένων, παρέχοντας πληροφορίες για την υγεία του εξοπλισμού, την απόδοση και πιθανά προβλήματα. Το παράδειγμα προγνωστικής συντήρησης, που ενεργοποιήθηκε από το IoT, αναδεικνύεται ως αλλαγή του παιχνιδιού σε αυτό το πλαίσιο. Η προγνωστική συντήρηση αξιοποιεί τη δύναμη των δεδομένων που δημιουργούνται από το IoT για την πρόβλεψη και την αντιμετώπιση πιθανών αστοχιών εξοπλισμού πριν αυτές συμβούν. Οι παραδοσιακές στρατηγικές συντήρησης συχνά ακολουθούν ένα αντιδραστικό μοντέλο, αντιμετωπίζοντας προβλήματα μόνο αφού εμφανιστούν. Ωστόσο, με το IoT, η συντήρηση μετατοπίζεται από μια αντιδραστική σε μια προληπτική στάση. Οι αισθητήρες συλλέγουν δεδομένα για διάφορες παραμέτρους όπως θερμοκρασία, πίεση, κραδασμούς και μοτίβα χρήσης. Στη συνέχεια, αυτά τα δεδομένα αναλύονται για να διακρίνουν μοτίβα, να εντοπίσουν ανωμαλίες και να προβλέψουν πότε ο εξοπλισμός είναι πιθανό να απαιτεί συντήρηση. Οι επιπτώσεις της προγνωστικής συντήρησης εκτείνονται πέρα από τον κατασκευαστικό τομέα. Στην υγειονομική περίθαλψη, για παράδειγμα, οι ιατρικές συσκευές με δυνατότητα IoT προσφέρουν συνεχή παρακολούθηση των ασθενών, επιτρέποντας στους παρόχους υγειονομικής περίθαλψης να παρεμβαίνουν με τα πρώτα σημάδια προβλημάτων υγείας. Στις μεταφορές, τα οχήματα εξοπλισμένα με IoT παρέχουν δεδομένα σε πραγματικό χρόνο για την απόδοση του κινητήρα, την απόδοση καυσίμου και πιθανές δυσλειτουργίες, επιτρέποντας την έγκαιρη συντήρηση και μειώνοντας τον κίνδυνο βλαβών. [6]

Η μετασχηματιστική δύναμη του IoT στην προγνωστική συντήρηση δεν περιορίζεται μόνο στα κέρδη αποδοτικότητας. Μεταφράζεται επίσης σε σημαντική εξοικονόμηση κόστους. Αντιμετωπίζοντας ζητήματα προτού κλιμακωθούν, οι οργανισμοί μπορούν να αποφύγουν δαπανηρές διακοπές λειτουργίας, να παρατείνουν τη διάρκεια ζωής του εξοπλισμού και να βελτιστοποιήσουν την κατανομή των πόρων. Επιπλέον, η μετάβαση από την προγραμματισμένη συντήρηση σε συντήρηση βάσει συνθηκών ή προγνωστική συντήρηση μειώνει τις περιττές παρεμβάσεις, ελαχιστοποιώντας τις λειτουργικές διακοπές και το σχετικό κόστος. Ωστόσο, η ευρεία υιοθέτηση του IoT και η προγνωστική συντήρηση δεν είναι χωρίς προκλήσεις. Οι ανησυχίες για την ασφάλεια και το απόρρητο είναι μεγάλες, δεδομένης της ευαίσθητης φύσης των δεδομένων που συλλέγονται από συσκευές IoT. Ο τεράστιος όγκος των δεδομένων που παράγονται θέτει προκλήσεις όσον αφορά την αποθήκευση, την επεξεργασία και την ανάλυση. Επιπλέον, τα ζητήματα διαλειτουργικότητας μεταξύ διαφορετικών συσκευών και πλατφορμών IoT πρέπει να αντιμετωπιστούν για να αξιοποιηθούν πλήρως οι δυνατότητες ενός διασυνδεδεμένου οικοσυστήματος. [7]

### 3.3 Σχετικές Μελέτες

*Lee, J., Bagheri, B., & Kao, H. A. (2015) – A Cyber-Physical Systems Architecture for Industry 4.0-Based Manufacturing Systems:*

Περιγράφει πώς το IoT και τα κυβερνοφυσικά συστήματα είναι κρίσιμοι παράγοντες για την προγνωστική συντήρηση, επιτρέποντας την παρακολούθηση και ανάλυση δεδομένων σε πραγματικό χρόνο.

*Zhou, K., Liu, T., & Zhou, L. (2015) – Industry 4.0: Towards Future Industrial Opportunities and Challenges:*

Αυτή η μελέτη διερευνά τον ρόλο του IoT στην έξυπνη κατασκευή και την προγνωστική συντήρηση, δίνοντας έμφαση στον τρόπο με τον οποίο οι συνδεδεμένες συσκευές οδηγούν τις στρατηγικές συλλογής δεδομένων και συντήρησης.

## ΚΕΦΑΛΑΙΟ 4: Αλγόριθμοι ανίχνευσης ανωμαλιών

Οι αλγόριθμοι ανίχνευσης ανωμαλιών διαδραματίζουν κρίσιμο ρόλο στην προγνωστική συντήρηση εντοπίζοντας αποκλίσεις από τις κανονικές συνθήκες λειτουργίας που μπορεί να σηματοδοτούν πιθανές βλάβες του εξοπλισμού. Αυτό το κεφάλαιο παρέχει μια ολοκληρωμένη επισκόπηση των βασικών αλγορίθμων ανίχνευσης ανωμαλιών, συμπεριλαμβανομένων των k-Nearest Neighbors (kNN), Support Vector Machine (SVM), Random Forest και Long Short-Term Memory (LSTM). Επεξηγούνται οι θεωρητικές αρχές πίσω από κάθε αλγόριθμο, μαζί με σκέψεις για την επιλογή κατάλληλων αλγορίθμων στο πλαίσιο της προγνωστικής συντήρησης.

### 4.1 k-Nearest Neighbors (kNN)

Ο αλγόριθμος k-Nearest Neighbors είναι μια απλή αλλά αποτελεσματική προσέγγιση για την ανίχνευση ανωμαλιών. Λειτουργεί με βάση την αρχή της εγγύτητας, όπου ένα παράδειγμα ταξινομείται με πλειοψηφία των γειτόνων του. Στο πλαίσιο της ανίχνευσης ανωμαλιών, το kNN υπολογίζει την απόσταση μεταξύ των σημείων δεδομένων και τα ταξινομεί με βάση την πλειοψηφική τάξη των k πλησιέστερων γειτόνων τους. Η θεωρητική βάση του kNN βρίσκεται στην υπόθεση ότι παρόμοιες περιπτώσεις τείνουν να ανήκουν στην ίδια κατηγορία. Οι εκτιμήσεις για την επιλογή του kNN στην προγνωστική συντήρηση περιλαμβάνουν την απλότητα, την επεκτασιμότητα και την ικανότητά του να χειρίζεται μη γραμμικές σχέσεις στα δεδομένα. [8], [9]

### 4.2 Support Vector Machine (SVM)

Ο Support Vector Machine είναι ένας ισχυρός αλγόριθμος για εργασίες ταξινόμησης και παλινδρόμησης, με εφαρμογές που επεκτείνονται στον εντοπισμό ανωμαλιών. Ο SVM λειτουργεί βρίσκοντας το υπερεπίπεδο που διαχωρίζει καλύτερα τα σημεία δεδομένων σε διαφορετικές κλάσεις, ενώ μεγιστοποιεί το περιθώριο μεταξύ των κλάσεων. Στην ανίχνευση ανωμαλιών, ο SVM προσδιορίζει ακραίες τιμές χαρτογραφώντας σημεία δεδομένων σε χώρο υψηλότερης διάστασης και βρίσκοντας το βέλτιστο υπερεπίπεδο που διαχωρίζει τις κανονικές περιπτώσεις από τις ανωμαλίες. Οι θεωρητικές αρχές πίσω από τον SVM περιλαμβάνουν τη μεγιστοποίηση του περιθωρίου μεταξύ των κλάσεων ελαχιστοποιώντας παράλληλα τα σφάλματα ταξινόμησης. Ο SVM είναι κατάλληλος για προγνωστική συντήρηση λόγω της ικανότητάς του να χειρίζεται δεδομένα υψηλών διαστάσεων και μη γραμμικές σχέσεις. [10], [11]

### 4.3 Random Forest

Ο Random Forest είναι ένας αλγόριθμος εκμάθησης συνόλου που χρησιμοποιεί μια συλλογή δέντρων αποφάσεων για να κάνει προβλέψεις. Κάθε δέντρο απόφασης στο δάσος ταξινομεί ανεξάρτητα σημεία δεδομένων και η τελική πρόβλεψη καθορίζεται με τη συγκέντρωση των αποτελεσμάτων μεμονωμένων δέντρων. Στο πλαίσιο της ανίχνευσης ανωμαλιών, ο Random Forest προσδιορίζει ανωμαλίες με βάση τη συχνότητα των σημείων δεδομένων που εμφανίζονται σε διαφορετικούς κόμβους φύλλων σε πολλά δέντρα. Η θεωρητική βάση του Random Forest βρίσκεται στην έννοια του bagging και της εκμάθησης συνόλου, που στοχεύει στη μείωση της υπερβολικής προσαρμογής και στη βελτίωση της προγνωστικής απόδοσης. Τα ζητήματα για την επιλογή του Random Forest στην προγνωστική συντήρηση περιλαμβάνουν την ανθεκτικότητά

του στο θόρυβο, την επεκτασιμότητα και την ικανότητά του να χειρίζεται μεγάλα σύνολα δεδομένων με υψηλή διαστάσεις. [12], [13]

#### 4.4 Long Short-Term Memory (LSTM)

Ο LSTM είναι ένας τύπος επαναλαμβανόμενου νευρωνικού δικτύου (RNN) που έχει σχεδιαστεί ειδικά για να καταγράφει μακροπρόθεσμες εξαρτήσεις σε διαδοχικά δεδομένα. Τα δίκτυα LSTM αποτελούνται από κελιά μνήμης που διατηρούν πληροφορίες για εκτεταμένες χρονικές περιόδους, καθιστώντας τα κατάλληλα για ανάλυση χρονοσειρών και ανίχνευση ανωμαλιών. Στην προγνωστική συντήρηση, οι αλγόριθμοι LSTM αναλύουν διαδοχικά δεδομένα αισθητήρων για να ανιχνεύσουν αποκλίσεις από τα κανονικά πρότυπα, τα οποία μπορεί να υποδεικνύουν πιθανές αστοχίες του εξοπλισμού. Οι θεωρητικές αρχές πίσω από το LSTM περιλαμβάνουν την αρχιτεκτονική των κυψελών μνήμης και τους μηχανισμούς ενημέρωσης και λήθης πληροφοριών με την πάροδο του χρόνου. Οι σκέψεις για την επιλογή του LSTM στην προγνωστική συντήρηση περιλαμβάνουν την ικανότητά του να καταγράφει χρονικές εξαρτήσεις, να χειρίζεται ακολουθίες μεταβλητού μήκους και να προσαρμόζεται σε εξελισσόμενα μοτίβα δεδομένων. [14], [15]

#### 4.5 Θεωρήσεις για την επιλογή αλγορίθμων ανίχνευσης ανωμαλιών

Κατά την επιλογή αλγορίθμων ανίχνευσης ανωμαλιών για προγνωστική συντήρηση χρησιμοποιώντας το σύνολο δεδομένων της NASA, πρέπει να ληφθούν υπόψη αρκετοί παράγοντες που αφορούν συγκεκριμένα δεδομένα. Το σύνολο δεδομένων αποτελείται από δεδομένα πολυμεταβλητών χρονοσειρών από 218 κινητήρες turbofan, με κάθε κινητήρα να περιέχει μετρήσεις από 26 αισθητήρες. Το σύνολο δεδομένων καταγράφει τη λειτουργική συμπεριφορά των κινητήρων από τη στιγμή που αρχίζουν να λειτουργούν κανονικά μέχρι να φτάσουν σε αστοχία.

##### 1) Φύση των Δεδομένων:

Η φύση των δομημένων πολυμεταβλητών χρονοσειρών του συνόλου δεδομένων PHM08 το καθιστά κατάλληλο για αλγόριθμους ανίχνευσης ανωμαλιών ικανών να χειρίζονται διαδοχικά δεδομένα. Ο LSTM είναι ιδιαίτερα ικανός στην καταγραφή χρονικών εξαρτήσεων σε δεδομένα χρονοσειρών, καθιστώντας το μια κατάλληλη επιλογή για την ανίχνευση ανωμαλιών στη λειτουργική συμπεριφορά των κινητήρων turbofan με την πάροδο του χρόνου. [16]

##### 2) Πολυπλοκότητα των σχέσεων μεταξύ μεταβλητών:

Οι αλληλεπιδράσεις μεταξύ των μετρήσεων των αισθητήρων στο σύνολο δεδομένων PHM08 μπορεί να εμφανίζουν σύνθετες σχέσεις που επηρεάζουν την υγεία και την απόδοση των κινητήρων στροβιλοανεμιστήρα. Αλγόριθμοι όπως ο SVM και ο Random Forest είναι ικανοί να καταγράφουν και να μοντελοποιούν περίπλοκες σχέσεις μεταξύ μεταβλητών, επιτρέποντάς τους να ανιχνεύουν αποτελεσματικά αποκλίσεις από την κανονική συμπεριφορά με βάση τα συνδυασμένα αποτελέσματα μετρήσεων πολλαπλών αισθητήρων. [16]

### 3) Παρουσία προσωρινών εξαρτήσεων:

Δεδομένης της διαδοχικής φύσης των δεδομένων, οι αλγόριθμοι ανίχνευσης ανωμαλιών πρέπει να λαμβάνουν υπόψη τις χρονικές εξαρτήσεις μεταξύ διαδοχικών παρατηρήσεων. Τεχνικές όπως τα επαναλαμβανόμενα νευρωνικά δίκτυα (RNN) ή τα δίκτυα μακράς βραχυπρόθεσμης μνήμης (LSTM) είναι κατάλληλες για την καταγραφή χρονικών εξαρτήσεων σε δεδομένα χρονοσειρών και την ανίχνευση ανωμαλιών που βασίζονται σε αποκλίσεις από τα αναμενόμενα μοτίβα με την πάροδο του χρόνου. [16]

### 4) Υπολογιστική αποτελεσματικότητα:

Αλγόριθμοι όπως ο k-Nearest Neighbors (kNN) και ο Random Forest είναι γνωστοί για την υπολογιστική τους απόδοση, καθιστώντας τους κατάλληλες επιλογές για την επεξεργασία και την ανάλυση δεδομένων πολυμεταβλητών χρονοσειρών από μεγάλο αριθμό κινητήρων turbofan και μετρήσεων αισθητήρων στο σύνολο δεδομένων. Αυτοί οι αλγόριθμοι μπορούν να χειριστούν αποτελεσματικά τον αυξανόμενο όγκο δεδομένων, ελαχιστοποιώντας ταυτόχρονα τους υπολογιστικούς πόρους που απαιτούνται για παρακολούθηση και ανάλυση σε πραγματικό χρόνο. [16]

### 5) Επεκτασιμότητα:

Η επεκτασιμότητα των αλγορίθμων ανίχνευσης ανωμαλιών είναι ζωτικής σημασίας για την προσαρμογή του αυξανόμενου όγκου δεδομένων από πολλαπλούς κινητήρες turbofan και μετρήσεων αισθητήρων στο σύνολο δεδομένων. Ο Random Forest, με την προσέγγιση εκμάθησης συνόλου και την παραλληλοποιήσιμη εκπαιδευτική διαδικασία, είναι κατάλληλο για κλιμάκωση για χειρισμό μεγάλων συνόλων δεδομένων με πολλαπλές μεταβλητές, καθιστώντας το μια βιώσιμη επιλογή για ανίχνευση ανωμαλιών σε εφαρμογές προγνωστικής συντήρησης. [16]

### 6) Ερμηνευσιμότητα των αποτελεσμάτων:

Στο πλαίσιο της προγνωστικής συντήρησης, η ερμηνευσιμότητα των αποτελεσμάτων ανίχνευσης ανωμαλιών είναι απαραίτητη για την κατανόηση των υποκείμενων παραγόντων που συμβάλλουν στις αστοχίες του εξοπλισμού. Αλγόριθμοι όπως ο kNN και ο Random Forest παρέχουν ερμηνεύσιμα αποτελέσματα, επιτρέποντας στους επαγγελματίες να εντοπίζουν τις πιο σημαντικές μετρήσεις αισθητήρων ή χαρακτηριστικά που συμβάλλουν στην ανίχνευση ανωμαλιών και να λαμβάνουν έγκαιρα διορθωτικά μέτρα με βάση αυτές τις πληροφορίες.

Λαμβάνοντας υπόψη αυτούς τους ειδικούς για το σύνολο δεδομένων παράγοντες και ευθυγραμμίζοντάς τους με τις δυνατότητες συγκεκριμένων αλγορίθμων ανίχνευσης ανωμαλιών όπως οι LSTM, SVM, kNN και Random Forest, μπορούν να ληφθούν τεκμηριωμένες αποφάσεις σχετικά με την επιλογή και την εφαρμογή κατάλληλων αλγορίθμων για τον εντοπισμό ανωμαλιών και τον μετριασμό πιθανής αποτυχίας εξοπλισμού χρησιμοποιώντας αποτελεσματικά το σύνολο δεδομένων. [17]

## 4.6 Σχετικές Μελέτες

Chandola, V., Banerjee, A., & Kumar, V. (2009) – *Anomaly Detection: A Survey*.

Αυτή η έρευνα είναι μια βασική αναφορά για την κατανόηση διαφόρων τεχνικών ανίχνευσης ανωμαλιών, συμπεριλαμβανομένων των μεθόδων στατιστικής, μηχανικής μάθησης και μεθόδων εγγύτητας, που χρησιμοποιούνται ευρέως στην προγνωστική συντήρηση.

*Ahmed, M., Mahmood, A. N., & Hu, J. (2016) – A Survey of Network Anomaly Detection Techniques:*

Παρέχεται μια επισκόπηση των αλγορίθμων ανίχνευσης ανωμαλιών σε δεδομένα δικτύου, πολλά από τα οποία εφαρμόζονται στην προγνωστική συντήρηση σε συστήματα IoT και δίκτυα αισθητήρων.

## **ΚΕΦΑΛΑΙΟ 5: Συλλογή δεδομένων στο IoT για Προγνωστική Συντήρηση και Προεπεξεργασία Δεδομένων**

Στο αναπτυσσόμενο τοπίο του Διαδικτύου των Πραγμάτων (IoT), η αποτελεσματική συλλογή, διαχείριση και σωστή προεπεξεργασία δεδομένων χρησιμεύουν ως θεμελιώδεις πυλώνες που στηρίζουν τις στρατηγικές πρόβλεψης συντήρησης. Αυτό το κεφάλαιο ξεκινά μια διεξοδική εξερεύνηση των πηγών δεδομένων εντός των πλαισίων IoT, με ιδιαίτερη έμφαση στον κεντρικό ρόλο που διαδραματίζουν οι αισθητήρες και οι συσκευές στην αποτύπωση ανεκτίμητων επιχειρησιακών πληροφοριών.

### **5.1 Αισθητήρες και συσκευές – Διαφορετικοί τύποι δεδομένων**

Οι αισθητήρες και οι συσκευές αποτελούν την πρωτοπορία της απόκτησης δεδομένων στα οικοσυστήματα IoT, ενεργώντας ως άγρυπνοι φρουροί που παρακολουθούν συνεχώς διάφορες πτυχές μηχανημάτων και υποδομών. Αυτοί οι αισθητήρες, που κυμαίνονται από μετρητές θερμοκρασίας και αισθητήρες πίεσης έως επιταχυνσιόμετρα και γυροσκόπια, χρησιμεύουν ως τα μάτια και τα αυτιά των συστημάτων πρόβλεψης συντήρησης, καταγράφοντας επιμελώς δεδομένα σε πραγματικό χρόνο ενδεικτικά της υγείας και της απόδοσης του εξοπλισμού. Επιπλέον, συσκευές IoT, όπως πύλες και ελεγκτές, εννορχηστρώνουν την εννορχήστρωση δικτύων αισθητήρων, διευκολύνοντας την απρόσκοπτη μετάδοση δεδομένων σε κεντρικά αποθετήρια για περαιτέρω ανάλυση και δράση. Παρόλα αυτά, η ποικιλομορφία των τύπων δεδομένων που συναντώνται σε περιβάλλοντα IoT θέτει σημαντικές προκλήσεις στις διαδικασίες συλλογής και διαχείρισης. Στο πλαίσιο της προγνωστικής συντήρησης, ο πολλαπλασιασμός δεδομένων χρονοσειρών από κρίσιμα περιουσιακά στοιχεία, όπως μετρήσεις κινητήρων τζετ ή βιομηχανικά μηχανήματα, εισάγει πολυπλοκότητες που απαιτούν εξελιγμένους μηχανισμούς χειρισμού. Τα δεδομένα χρονοσειρών, που χαρακτηρίζονται από τη διαδοχική φύση και τις χρονικές εξαρτήσεις τους, απαιτούν εξειδικευμένες προσεγγίσεις για προεπεξεργασία, ανίχνευση ανωμαλιών και προγνωστική μοντελοποίηση. Επιπλέον, η ενσωμάτωση διαφορετικών πηγών δεδομένων, συμπεριλαμβανομένων δομημένων δεδομένων αισθητήρων, μη δομημένων αρχείων καταγραφής συντήρησης και μετα-δεδομένων με βάση τα συμφραζόμενα, απαιτεί ισχυρές στρατηγικές για την εναρμόνιση και τη σύνθεση ετερογενών συνόλων δεδομένων σε συνεκτικά αναλυτικά πλαίσια. [18]

### **5.2 Στρατηγικές για αποτελεσματική συλλογή και προεπεξεργασία δεδομένων - Αντιμετώπιση περιορισμών επεκτασιμότητας και πόρων:**

Η πλοήγηση στις περιπλοκές της συλλογής και της προεπεξεργασίας δεδομένων απαιτεί την υιοθέτηση στρατηγικών προσεγγίσεων προσαρμοσμένων στις μοναδικές απαιτήσεις των εφαρμογών πρόβλεψης συντήρησης. Πρώτον, πρέπει να δοθεί σχολαστική προσοχή στην επιλογή και την ανάπτυξη αισθητήρων, διασφαλίζοντας τη βέλτιστη κάλυψη και ευαισθησία για την καταγραφή των σχετικών λειτουργικών μετρήσεων. Οι διαδικασίες βαθμονόμησης και επικύρωσης ενισχύουν περαιτέρω την αξιοπιστία και την ακρίβεια των μετρήσεων των αισθητήρων, ενσταλάσσοντας εμπιστοσύνη στην ακεραιότητα των συλλεγόμενων δεδομένων. Στη συνέχεια, τεχνικές προεπεξεργασίας όπως το φιλτράρισμα θορύβου, η ανίχνευση ακραίων τιμών και ο καταλογισμός δεδομένων που λείπουν χρησιμεύουν για τη βελτίωση της ποιότητας και της χρηστικότητας των ακατέργαστων δεδομένων αισθητήρων, θέτοντας τις βάσεις για



μεταγενέστερες προσπάθειες ανάλυσης και μοντελοποίησης. Καθώς οι αναπτύξεις του IoT κλιμακώνονται σε εύρος και πολυπλοκότητα, η επεκτασιμότητα και οι περιορισμοί πόρων εμφανίζονται ως τρομερά εμπόδια που αντιμετωπίζουν οι προσπάθειες συλλογής δεδομένων. Η εκθετική αύξηση του όγκου και της ταχύτητας των δεδομένων αισθητήρων απαιτεί επεκτάσιμες λύσεις αποθήκευσης, κατανεμημένες αρχιτεκτονικές υπολογιστών και υποδομές που βασίζονται σε cloud, ικανές να φιλοξενήσουν τον αυξανόμενο κατακλυσμό δεδομένων. Τα μοντέλα υπολογιστών αιχμής προσφέρουν πολλά υποσχόμενες οδούς για την άμβλυση των περιορισμών καθυστέρησης και εύρους ζώνης, επιτρέποντας τοπική επεξεργασία δεδομένων και ανάλυση στην περιφέρεια του δικτύου. Επιπλέον, οι συσκευές IoT με περιορισμένους πόρους απαιτούν ενεργειακά αποδοτικά πρωτόκολλα μετάδοσης δεδομένων, τεχνικές συμπίεσης δεδομένων και προσαρμοστικές στρατηγικές δειγματοληψίας για τη βελτιστοποίηση της διάρκειας ζωής της μπαταρίας και τον μετριασμό των υπολογιστικών επιβαρύνσεων. [18], [19]

### 5.3 Σημασία της προεπεξεργασίας δεδομένων

Η προεπεξεργασία δεδομένων χρησιμεύει ως το θεμέλιο πάνω στο οποίο χτίζονται ισχυρά μοντέλα πρόβλεψης συντήρησης. Με τον καθαρισμό, τον μετασχηματισμό και τον εμπλουτισμό των ακατέργαστων συνόλων δεδομένων, η προεπεξεργασία δεδομένων θέτει τις βάσεις για μεταγενέστερες προσπάθειες ανάλυσης και μοντελοποίησης. Αυτή η προπαρασκευαστική φάση διαδραματίζει κεντρικό ρόλο στον μετριασμό των επιπτώσεων των θορυβωδών, ελλιπών ή ασυνεπών δεδομένων, διασφαλίζοντας την ακεραιότητα και την αξιοπιστία των αναλυτικών αποτελεσμάτων. Επιπλέον, η προεπεξεργασία δεδομένων ενισχύει τη διαλειτουργικότητα και τη συμβατότητα μεταξύ διαφορετικών πηγών δεδομένων, εναρμονίζοντας ετερογενή σύνολα δεδομένων σε συνεκτικά αναλυτικά πλαίσια που ευνοούν τη μηχανική μάθηση και την προγνωστική ανάλυση. [19]

### 5.4 Τεχνικές χειρισμού τιμών που λείπουν

Η παρουσία τιμών που λείπουν θέτει σημαντικές προκλήσεις στις αναλύσεις πρόβλεψης συντήρησης, δυνητικά παραμορφώνοντας τα αποτελέσματα και υπονομεύοντας την ευρωστία των μοντέλων μηχανικής μάθησης. Για να αντιμετωπιστεί αυτό το ζήτημα, υπάρχουν μυριάδες τεχνικές για τον χειρισμό τιμών που λείπουν, που κυμαίνονται από απλές μεθόδους καταλογισμού έως εξελιγμένους αλγόριθμους. Οι συνήθεις προσεγγίσεις περιλαμβάνουν τον καταλογισμό μέσου όρου, τον καταλογισμό διάμεσης τιμής και τον καταλογισμό τρόπου λειτουργίας, όπου οι τιμές που λείπουν αντικαθίστανται με συνοπτικά στατιστικά στοιχεία που προέρχονται από τα παρατηρούμενα δεδομένα. Εναλλακτικά, προηγμένες τεχνικές όπως ο καταλογισμός k-πλησιέστερου γείτονα (kNN imputation), ο καταλογισμός παλινδρόμησης (regression imputation) και η ολοκλήρωση μήτρας (matrix completion) αξιοποιούν τις σχέσεις μεταξύ των μεταβλητών για να συναγάγουν τιμές που λείπουν, διατηρώντας την ακεραιότητα της υποκείμενης κατανομής δεδομένων. [20], [21]

### 5.5 Κανονικοποίηση και ανίχνευση ακραίων τιμών:

Η κανονικοποίηση και η ανίχνευση ακραίων τιμών είναι απαραίτητα βήματα προεπεξεργασίας που στοχεύουν στη βελτίωση της ερμηνείας και της ευρωστίας των μοντέλων προγνωστικής συντήρησης. Οι τεχνικές κανονικοποίησης, όπως η κλιμάκωση ελάχιστης μέγιστης κλίμακας (min-max scaling) και η κανονικοποίηση της βαθμολογίας z (z-score normalization), τυποποιούν τις τιμές των χαρακτηριστικών σε μια κοινή κλίμακα, αποτρέποντας προκαταλήψεις και διαφορές στην εκπαίδευση και την αξιολόγηση του μοντέλου. Ταυτόχρονα, οι αλγόριθμοι ανίχνευσης ακραίων τιμών, συμπεριλαμβανομένων των δασών απομόνωσης (isolation forests), της ομαδοποίησης k-means (k-means clustering) και ισχυρών στατιστικών μεθόδων, εντοπίζουν ανώμαλες παρατηρήσεις που αποκλίνουν σημαντικά από την αναμενόμενη κατανομή δεδομένων. Με την επισήμανση και την πιθανή κατάργηση των ακραίων στοιχείων, αυτές οι τεχνικές ενισχύουν την ανθεκτικότητα και τη γενίκευση των μοντέλων μηχανικής μάθησης, ενισχύοντάς τα έναντι ψευδών επιρροών και ενισχύοντας την προγνωστική τους απόδοση. [22], [23]

## 5.6 Σχετικές Μελέτες

*Wan, J., et al. (2016) – Industrial IoT: A Survey on Enabling Technologies, Protocols, and Applications:*

Το έργο αυτό περιγράφει βασικές τεχνολογίες για τη συλλογή δεδομένων IoT, απαραίτητες για την προγνωστική συντήρηση, συμπεριλαμβανομένων των αισθητήρων, των πρωτοκόλλων επικοινωνίας και της διαχείρισης δεδομένων.

*García, S., et al. (2016) – A Survey on Data Preprocessing in Data Mining:*

Παρέχει μια ολοκληρωμένη επισκόπηση των τεχνικών προεπεξεργασίας δεδομένων, όπως η κανονικοποίηση, ο καταλογισμός και η εξαγωγή χαρακτηριστικών, οι οποίες είναι ζωτικής σημασίας για το χειρισμό δεδομένων IoT πριν από την εφαρμογή προγνωστικών αλγορίθμων.

## ΚΕΦΑΛΑΙΟ 6: Η Python ως κεντρικό εργαλείο για την προγνωστική συντήρηση

Η Python αποτελεί ακρογωνιαίό λίθο στη σφαίρα της προγνωστικής συντήρησης, προσφέροντας ένα πλούσιο οικοσύστημα εργαλείων, βιβλιοθηκών και πλαισίων προσαρμοσμένων στις διαφορετικές ανάγκες προεπεξεργασίας δεδομένων, υλοποίησης αλγορίθμων και οπτικοποίησης. Αυτό το κομβικό κεφάλαιο ξεκινά μια ολοκληρωμένη εξερεύνηση των δυνατοτήτων της Python, διευκρινίζοντας τον καθοριστικό ρόλο της στην ενδυνάμωση των επαγγελματιών να εξάγουν πρακτικές ιδέες από ακατέργαστα δεδομένα και να οδηγούν στη λήψη τεκμηριωμένων αποφάσεων σε προσπάθειες πρόβλεψης συντήρησης.

### 6.1 Λεπτομερής εξερεύνηση των δυνατοτήτων της Python

Η ευελιξία και η επεκτασιμότητα της Python την καθιστούν απαραίτητο στοιχείο στο κιτ εργαλείων προγνωστικής συντήρησης, παρέχοντας στους επαγγελματίες μια ευέλικτη και διαισθητική πλατφόρμα για χειρισμό δεδομένων, ανάλυση και μοντελοποίηση. Στο επίκεντρο της απήχησης της Python βρίσκεται η απλότητα και η αναγνωσιμότητά της, ενισχύοντας την ταχεία δημιουργία πρωτοτύπων και πειραματισμό σε διάφορους τομείς. Η ισχυρή τυπική βιβλιοθήκη της Python προσφέρει μια πλούσια σειρά από ενσωματωμένες μονάδες για I/O αρχείων, σειριοποίηση δεδομένων και μαθηματικές πράξεις, θέτοντας μια γερή βάση για προηγμένη προεπεξεργασία δεδομένων και αλγοριθμικές υλοποιήσεις. Επιπλέον, η δυναμική πληκτρολόγηση και το αντικειμενοστραφή παράδειγμα της Python διευκολύνουν την απρόσκοπτη ενσωμάτωση με εξωτερικές βιβλιοθήκες και πλαίσια, επιτρέποντας στους επαγγελματίες να αξιοποιούν εξειδικευμένα εργαλεία για συγκεκριμένες εργασίες χωρίς να θυσιάζουν τη διαλειτουργικότητα ή την επεκτασιμότητα. [24]

### 6.2 Βιβλιοθήκες Python για αποτελεσματικό χειρισμό δεδομένων

Το οικοσύστημα της Python μπορεί να υπερηφανεύεται για μια πληθώρα βιβλιοθηκών προσαρμοσμένων για τον εξορθολογισμό και την επιτάχυνση της διαδικασίας χειρισμού δεδομένων, δίνοντας τη δυνατότητα στους επαγγελματίες να καθαρίζουν και να μεταμορφώνουν ακατέργαστα σύνολα δεδομένων με ευκολία και αποτελεσματικότητα. Η κυριότερη από αυτές τις βιβλιοθήκες είναι η Pandas, μια ισχυρή εργαλειοθήκη χειρισμού δεδομένων, γνωστή για τις εκφραστικές και υψηλής απόδοσης δομές δεδομένων της. Η αφαιρετικότητα του DataFrame του Pandas παρέχει στους επαγγελματίες τον λεπτό έλεγχο των δεδομένων σε πίνακα, διευκολύνοντας λειτουργίες όπως η ευρετηρίαση, ο τεμαχισμός, η συγκέντρωση και η ένωση με απaráμιλλη ευκολία. Το Complementing Pandas είναι η NumPy, μια θεμελιώδης βιβλιοθήκη για αριθμητικούς υπολογισμούς στην Python, που προσφέρει αποτελεσματικές λειτουργίες πίνακα και μαθηματικές συναρτήσεις απαραίτητες για την προεπεξεργασία δεδομένων και τις αλγοριθμικές υλοποιήσεις. Μαζί, τα Pandas και το NumPy αποτελούν τον ακρογωνιαίό λίθο των δυνατοτήτων χειρισμού δεδομένων της Python, παρέχοντας στους επαγγελματίες μια ισχυρή και ευέλικτη εργαλειοθήκη για διερευνητική ανάλυση δεδομένων και μηχανική χαρακτηριστικών. [24]

### 6.3 Εφαρμογή της Python για Αλγοριθμική Υλοποίηση

Η ικανότητα της Python εκτείνεται πέρα από το χειρισμό δεδομένων για να περιλαμβάνει αλγοριθμική υλοποίηση, προσφέροντας μια ποικιλία από βιβλιοθήκες και πλαίσια προσαρμοσμένα στις απαιτήσεις της προγνωστικής μοντελοποίησης και της μηχανικής μάθησης. Κεντρική θέση σε αυτό το οικοσύστημα είναι το scikit-learn, μια ολοκληρωμένη βιβλιοθήκη μηχανικής μάθησης, γνωστή για την απλότητα, την ευελιξία και την επεκτασιμότητα της. Το Scikit-learn προσφέρει μια πλούσια γκάμα αλγορίθμων για ταξινόμηση, παλινδρόμηση, ομαδοποίηση και μείωση διαστάσεων, συνοδευόμενα από διαισθητικά API και εκτενή τεκμηρίωση. Αξιοποιώντας το scikit-learn, οι επαγγελματίες μπορούν αβίαστα να εφαρμόσουν και να αξιολογήσουν μοντέλα πρόβλεψης συντήρησης, αξιοποιώντας αλγόριθμους αιχμής, όπως τυχαία δάση, μηχανές υποστήριξης διανυσμάτων και μηχανές ενίσχυσης κλίσης για να εξάγουν χρήσιμες πληροφορίες από ακατέργαστα δεδομένα. [25]

## 6.4 Οπτικοποίηση με χρήση Matplotlib και Seaborn

Η αποτελεσματική οπτικοποίηση χρησιμεύει ως βασικός άξονας στη ροή εργασιών προγνωστικής συντήρησης, επιτρέποντας στους επαγγελματίες να επικοινωνούν πληροφορίες, τάσεις και μοτίβα που προέρχονται από δεδομένα με συναρπαστικό και διαισθητικό τρόπο. Οι βιβλιοθήκες οπτικοποίησης της Python, μεταξύ των οποίων οι Matplotlib και Seaborn, δίνουν τη δυνατότητα στους επαγγελματίες να δημιουργούν πλοκές και γραφήματα ποιότητας έκδοσης με ελάχιστη προσπάθεια. Το Matplotlib, η θεμελιώδης βιβλιοθήκη σχεδίασης της Python, παρέχει λεπτομερή έλεγχο της αισθητικής και της σύνθεσης της πλοκής, επιτρέποντας στους επαγγελματίες να προσαρμόσουν κάθε πτυχή των απεικονίσεών τους ανάλογα με τις ανάγκες τους. Συμπληρώνοντας το Matplotlib είναι το Seaborn, μια βιβλιοθήκη οπτικοποίησης υψηλού επιπέδου που χτίστηκε πάνω από το Matplotlib, που προσφέρει μια βελτιωμένη διεπαφή για τη δημιουργία ενημερωτικών στατιστικών γραφικών με ελάχιστο κώδικα. Μαζί, η Matplotlib και η Seaborn εξοπλίζουν τους επαγγελματίες με τα απαραίτητα εργαλεία για την αποτελεσματική μετάδοση σύνθετων γνώσεων, ενισχύοντας τη βαθύτερη κατανόηση των δεδομένων και διευκολύνοντας τη λήψη αποφάσεων βάσει δεδομένων σε προσπάθειες προγνωστικής συντήρησης. [25]

## 6.5 Σχετικές Μελέτες

*Van Rossum, G., & Drake, F. L. (2009) – Python 3 Reference Manual:* Ως το επίσημο εγχειρίδιο αναφοράς για την Python, αυτή η εργασία είναι θεμελιώδης για την κατανόηση των δυνατοτήτων της Python, συμπεριλαμβανομένης της ευελιξίας της στη δημιουργία μοντέλων πρόβλεψης συντήρησης και στο χειρισμό της ανάλυσης δεδομένων.

*Raschka, S. (2015) – Python Machine Learning:* Αυτό το βιβλίο είναι απαραίτητο για την εκμάθηση του τρόπου με τον οποίο η Python μπορεί να χρησιμοποιηθεί για την εφαρμογή αλγορίθμων μηχανικής μάθησης, ιδιαίτερα αυτών που χρησιμοποιούνται στον εντοπισμό ανωμαλιών και στην προγνωστική συντήρηση.

## ΚΕΦΑΛΑΙΟ 7: Μετρήσεις απόδοσης για την αξιολόγηση αλγορίθμων

Στον τομέα της προγνωστικής συντήρησης, η αξιολόγηση της απόδοσης του αλγορίθμου αποτελεί μια κρίσιμη προσπάθεια, που χρησιμεύει ως πυξίδα για την καθοδήγηση της λήψης αποφάσεων, την αξιολόγηση της αποτελεσματικότητας του μοντέλου και τη βελτιστοποίηση των λειτουργικών αποτελεσμάτων. Αυτό το κεφάλαιο ξεκινά μια περιεκτική εξερεύνηση βασικών μετρήσεων απόδοσης, φωτίζοντας τη συνάφεια, την ερμηνεία και τις επιπτώσεις τους στο πλαίσιο των προσπαθειών πρόβλεψης συντήρησης.

### 7.1 Εισαγωγή στις βασικές μετρήσεις απόδοσης

Οι βασικές μετρήσεις απόδοσης αποτελούν το θεμέλιο πάνω στο οποίο στηρίζεται η αξιολόγηση του αλγορίθμου, παρέχοντας στους επαγγελματίες ποσοτικά μέτρα για τη μέτρηση της προγνωστικής ικανότητας και αποτελεσματικότητας των μοντέλων μηχανικής μάθησης. Ανάμεσα στο πάνθεον των μετρήσεων απόδοσης, αρκετά ξεχωρίζουν ως απαραίτητα εργαλεία για την αξιολόγηση της ακρίβειας του μοντέλου, της ικανότητας γενίκευσης και της υπολογιστικής αποτελεσματικότητας. Αυτές οι μετρήσεις περιλαμβάνουν:

- 1) Συντελεστής προσδιορισμού ( $R^2$ ): Το  $R^2$ , γνωστό και ως συντελεστής προσδιορισμού, ποσοτικοποιεί το ποσοστό διακύμανσης στη μεταβλητή στόχο που εξηγείται από το μοντέλο. Οι τιμές  $R^2$  κυμαίνονται από 0 έως 1, όπου το 1 υποδηλώνει τέλεια πρόβλεψη και το 0 υποδηλώνει ένα μοντέλο που δεν έχει καλύτερη απόδοση από την τυχαία πιθανότητα.
- 2) Root Mean Squared Error (RMSE): Το RMSE χρησιμεύει ως μέτρο της μέσης απόκλισης μεταξύ προβλεπόμενων και παρατηρούμενων τιμών, με χαμηλότερες τιμές RMSE ενδεικτικές ανώτερης ακρίβειας μοντέλου και προγνωστικής ακρίβειας. Το RMSE είναι ιδιαίτερα κατάλληλο για εργασίες παλινδρόμησης, όπου παρέχει πληροφορίες για το μέγεθος των σφαλμάτων πρόβλεψης σε σχέση με την κλίμακα της μεταβλητής στόχου.
- 3) Time to Train: Το Time to Train ενσωματώνει τα υπολογιστικά έξοδα που προέκυψαν κατά την εκπαίδευση του μοντέλου, αντικατοπτρίζοντας τη διάρκεια που απαιτείται για την προσαρμογή του αλγόριθμου στα δεδομένα εκπαίδευσης και τη βελτιστοποίηση των παραμέτρων του μοντέλου. Αυτή η μέτρηση προσφέρει πληροφορίες για την αλγοριθμική αποτελεσματικότητα και την επεκτασιμότητα, με μικρότερους χρόνους εκπαίδευσης ενδεικτικά της ταχύτερης σύγκλισης του μοντέλου και της μειωμένης κατανάλωσης πόρων.
- 4) Time to Predict: Το Time to Predict ποσοτικοποιεί τον υπολογιστικό λανθάνοντα χρόνο που σχετίζεται με τη δημιουργία προβλέψεων σε μη ορατά δεδομένα, που περιλαμβάνει τον χρόνο που απαιτείται για την επεξεργασία των χαρακτηριστικών εισόδου και την εκτέλεση αλγορίθμων συμπερασμάτων. Συντομότεροι χρόνοι πρόβλεψης είναι επιθυμητοί σε εφαρμογές σε πραγματικό χρόνο, όπου η γρήγορη απόκριση και η χαμηλή καθυστέρηση είναι πρωταρχικής σημασίας.
- 5) Total Time: Το Total Time αντιπροσωπεύει τη συνολική διάρκεια που περιλαμβάνει τόσο τις φάσεις εκπαίδευσης όσο και τις φάσεις πρόβλεψης, προσφέροντας μια ολιστική άποψη της αλγοριθμικής απόδοσης και της λειτουργικής αποτελεσματικότητας. Λαμβάνοντας υπόψη τόσο την υπολογιστική επιβάρυνση όσο και την καθυστέρηση

συμπερασμάτων, ο συνολικός χρόνος παρέχει στους επαγγελματίες μια ολοκληρωμένη μέτρηση της αλγοριθμικής αποτελεσματικότητας και της χρήσης πόρων. [26]

## 7.2 Επεξήγηση της συνάφειας και της ερμηνείας:

Η συνάφεια και η ερμηνεία κάθε μέτρησης απόδοσης εξαρτώνται από τους συγκεκριμένους στόχους, τους περιορισμούς και τις επιχειρησιακές απαιτήσεις της συγκεκριμένης εργασίας πρόβλεψης συντήρησης. Το  $R^2$ , για παράδειγμα, προσφέρει πληροφορίες για την καλή προσαρμογή μεταξύ προβλέψεων μοντέλων και παρατηρούμενων δεδομένων, παρέχοντας ένα μέτρο της επεξηγηματικότητας του μοντέλου και της προγνωστικής ισχύος. Οι υψηλότερες τιμές  $R^2$  υποδεικνύουν ισχυρότερες συσχετίσεις μεταξύ των προβλεπόμενων και των πραγματικών αποτελεσμάτων, υποδηλώνοντας ανώτερη απόδοση και αξιοπιστία μοντέλου.

Ομοίως, το RMSE χρησιμεύει ως βαρόμετρο της ακρίβειας πρόβλεψης, ποσοτικοποιώντας το μέγεθος των σφαλμάτων που ενυπάρχουν στις προβλέψεις μοντέλων. Οι χαμηλότερες τιμές RMSE υποδηλώνουν στενότερη συμφωνία μεταξύ των προβλεπόμενων και των παρατηρούμενων τιμών, υποδηλώνοντας ανώτερη προγνωστική ακρίβεια και αξιοπιστία. Αντίθετα, οι αυξημένες τιμές RMSE μπορεί να υποδεικνύουν μη βέλτιστη απόδοση του μοντέλου ή ανεπάρκειες στην αναπαράσταση χαρακτηριστικών ή στις υποθέσεις μοντελοποίησης.

Ο χρόνος εκπαίδευσης και ο χρόνος πρόβλεψης μετρήσεων προσφέρουν πληροφορίες για την αλγοριθμική αποτελεσματικότητα και την υπολογιστική κλιμάκωση, οριοθετώντας τα υπολογιστικά έξοδα που σχετίζονται με τις εργασίες εκπαίδευσης μοντέλων και πρόβλεψης, αντίστοιχα. Οι μικρότεροι χρόνοι εκπαίδευσης επιταχύνουν την ανάπτυξη μοντέλων και τους κύκλους επανάληψης, επιτρέποντας στους επαγγελματίες να εξερευνήσουν ένα ευρύτερο φάσμα αλγορίθμων και υπερπαραμέτρων μέσα σε ένα πεπερασμένο χρονικό πλαίσιο. Ομοίως, οι μειωμένοι χρόνοι πρόβλεψης διευκολύνουν τη λήψη αποφάσεων σε πραγματικό χρόνο και την ανταπόκριση, επιτρέποντας έγκαιρες παρεμβάσεις και διορθωτικές ενέργειες στις ροές εργασιών πρόβλεψης συντήρησης.

Ο συνολικός χρόνος χρησιμεύει ως μια σύνθετη μέτρηση που περιλαμβάνει τις φάσεις εκπαίδευσης και πρόβλεψης, παρέχοντας μια ολοκληρωμένη εικόνα της αλγοριθμικής απόδοσης και της λειτουργικής αποτελεσματικότητας. Υπολογίζοντας τη σωρευτική διάρκεια που απαιτείται για την εκπαίδευση του μοντέλου, τη δημιουργία προβλέψεων και την εκτέλεση αλγορίθμων συμπερασμάτων, ο συνολικός χρόνος προσφέρει στους επαγγελματίες μια ολιστική μέτρηση της αλγοριθμικής αποτελεσματικότητας, της χρήσης υπολογιστικών πόρων και της λειτουργικής σκοπιμότητας. [27]

## 7.3 Παρατηρήσεις για την επιλογή κατάλληλων μετρήσεων

Η επιλογή των κατάλληλων μετρήσεων απόδοσης εξαρτάται από μια λεπτή κατανόηση της εργασίας πρόβλεψης συντήρησης, η οποία περιλαμβάνει τους στόχους, τους περιορισμούς και το λειτουργικό της πλαίσιο. Σε σενάρια όπου η ερμηνευσιμότητα και η επεξήγηση του μοντέλου είναι πρωταρχικής σημασίας, το  $R^2$  χρησιμεύει ως κατάλληλη μέτρηση για την ποσοτικοποίηση της αξιοπιστίας των προβλέψεων του μοντέλου και της ισχύος των υποκείμενων σχέσεων μεταξύ των μεταβλητών πρόβλεψης και των αποτελεσμάτων-στόχων. Αντίθετα, σε εφαρμογές όπου η ακρίβεια και η ακρίβεια της πρόβλεψης είναι υψίστης σημασίας, το RMSE προσφέρει ένα ισχυρό

μέτρο του σφάλματος πρόβλεψης, καθοδηγώντας τις προσπάθειες βελτίωσης του μοντέλου και βελτιστοποίησης.

Επιπλέον, εκτιμήσεις όπως η υπολογιστική απόδοση, η ανταπόκριση σε πραγματικό χρόνο και οι περιορισμοί πόρων παίζουν καθοριστικό ρόλο στην επιλογή μετρήσεων, επηρεάζοντας την ιεράρχηση μετρήσεων όπως ο χρόνος εκπαίδευσης, ο χρόνος πρόβλεψης και ο συνολικός χρόνος. Σε εφαρμογές ευαίσθητες στον λανθάνοντα χρόνο που απαιτούν ταχεία λήψη αποφάσεων και ανταπόκριση, η ελαχιστοποίηση των χρόνων πρόβλεψης γίνεται πρωταρχικός στόχος, καθιστώντας απαραίτητη τη συμπερίληψη μετρήσεων με βάση το χρόνο στο πλαίσιο αξιολόγησης.

Ευθυγραμμίζοντας την επιλογή μετρήσεων με τις συγκεκριμένες απαιτήσεις και τους στόχους της προγνωστικής εργασίας συντήρησης, οι επαγγελματίες μπορούν να αποκτήσουν αξιόπιστες γνώσεις σχετικά με την απόδοση του αλγορίθμου, να βελτιστοποιήσουν την αποτελεσματικότητα του μοντέλου και να οδηγήσουν στη λήψη τεκμηριωμένων αποφάσεων σε λειτουργικά πλαίσια.

Μέσω μιας ολοκληρωμένης εξέτασης των βασικών μετρήσεων απόδοσης, της συνάφειας, της ερμηνείας και των εκτιμήσεων για την επιλογή μετρήσεων, αυτό το κεφάλαιο εξοπλίζει τους επαγγελματίες με τα εργαλεία και τις γνώσεις που είναι απαραίτητες για την αυστηρή αξιολόγηση της απόδοσης του αλγορίθμου, τη βελτιστοποίηση των ροών εργασιών πρόβλεψης συντήρησης και την απελευθέρωση ενεργών πληροφοριών από ακατέργαστα δεδομένα. Υιοθετώντας μια προσέγγιση βάσει δεδομένων για την αξιολόγηση αλγορίθμων και την αξιολόγηση της απόδοσης, οι επαγγελματίες ανοίγουν το δρόμο προς την μέγιστη απόδοση προβλεπτικής συντήρησης, αξιοποιώντας τη μεταμορφωτική δύναμη της μηχανικής μάθησης και των αναλυτικών στοιχείων για να αυξήσουν τη λειτουργική αποτελεσματικότητα, την αξιοπιστία των περιουσιακών στοιχείων και την οργανωτική ανθεκτικότητα. [27]

## 7.4 Σχετικές Μελέτες

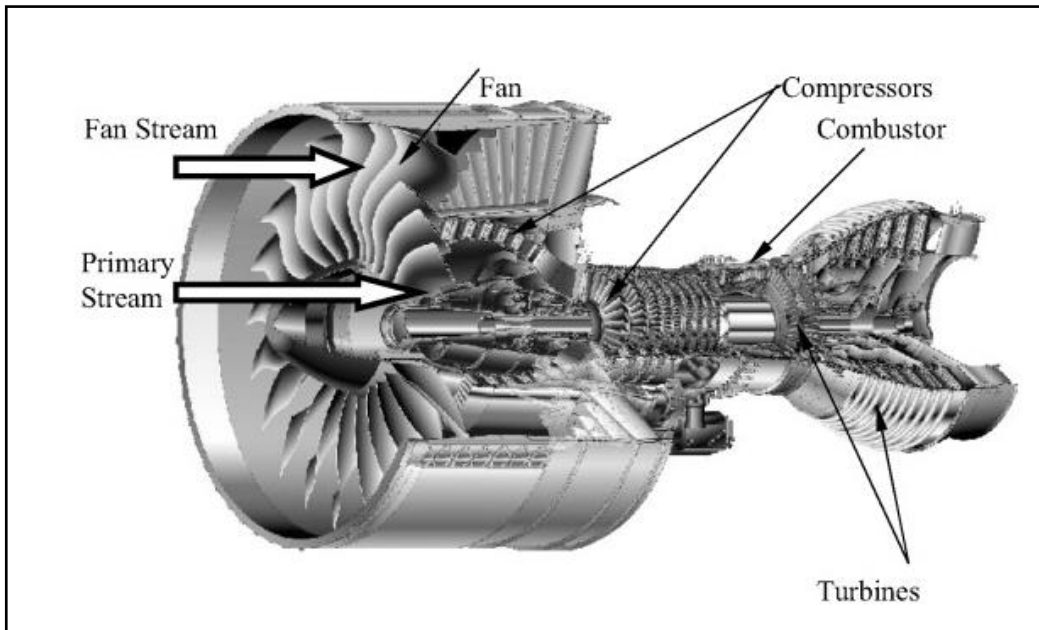
*Powers, D. M. (2011) – Evaluation: From Precision, Recall and F-measure to ROC, Informedness, Markedness & Correlation:* Αυτό το paper παρέχει μια εις βάθος επισκόπηση των βασικών μετρήσεων απόδοσης όπως η ακρίβεια, η ανάκληση, η βαθμολογία F1 και οι καμπύλες ROC, που χρησιμοποιούνται ευρέως για την αξιολόγηση της αποτελεσματικότητας των αλγορίθμων πρόβλεψης συντήρησης.

*Sokolova, M., & Lapalme, G. (2009) – A Systematic Analysis of Performance Measures for Classification Tasks:* Η μελέτη εξετάζει συστηματικά κοινές μετρήσεις απόδοσης για αλγόριθμους ταξινόμησης, όπως η μήτρα ακρίβειας και σύγχυσης, που σχετίζονται με την αξιολόγηση μοντέλων προγνωστικής συντήρησης.

## ΚΕΦΑΛΑΙΟ 8: Συγκριτική Ανάλυση Αλγορίθμων

Σε αυτό το κεφάλαιο, θα παρουσιαστεί εκτενώς η τεχνική πλευρά της παρούσας διατριβής.,

Το dataset που χρησιμοποιήθηκε ανήκει στη NASA και χρησιμοποιείται για asset degradation modeling. Περιλαμβάνει δεδομένα προσομοίωσης Run-to-Failure από κινητήρες αεριωθούμενων ανεμιστήρων turbo.



Τα σύνολα δεδομένων πειραματικού σεναρίου αποτελούνται από πολλαπλές πολυμεταβλητές χρονοσειρές. Κάθε σύνολο δεδομένων χωρίζεται περαιτέρω σε υποσύνολα εκπαίδευσης και δοκιμής. Κάθε χρονοσειρά προέρχεται από διαφορετικό κινητήρα, δηλαδή, τα δεδομένα μπορούν να θεωρηθούν ότι προέρχονται από έναν στόλο κινητήρων του ίδιου τύπου. Κάθε κινητήρας ξεκινά με διαφορετικούς βαθμούς αρχικής φθοράς και διακύμανσης κατασκευής. Αυτή η φθορά και η διακύμανση θεωρείται φυσιολογική, δηλαδή δεν θεωρείται κατάσταση σφάλματος. Υπάρχουν τρεις ρυθμίσεις λειτουργίας που έχουν ουσιαστική επίδραση στην απόδοση του κινητήρα. Αυτές οι ρυθμίσεις περιλαμβάνονται επίσης στα δεδομένα. Τα δεδομένα είναι μολυσμένα με θόρυβο αισθητήρα.

Ο κινητήρας λειτουργεί κανονικά στην αρχή κάθε χρονοσειράς και εμφανίζει σφάλμα κάποια στιγμή κατά τη διάρκεια της σειράς. Στο σετ εκπαίδευσης, το σφάλμα αυξάνεται σε μέγεθος μέχρι την αστοχία του συστήματος. Στο δοκιμαστικό σύνολο, η χρονοσειρά τελειώνει λίγο πριν από την αποτυχία του συστήματος. Τα δεδομένα παρέχονται ως αρχείο κειμένου συμπιεσμένο με zip με 26 στήλες αριθμών, διαχωρισμένες με κενά. Κάθε σειρά είναι ένα στιγμιότυπο δεδομένων που λαμβάνονται κατά τη διάρκεια ενός μόνο λειτουργικού κύκλου, κάθε στήλη είναι μια διαφορετική μεταβλητή. Οι στήλες αντιστοιχούν σε:

1) unit number



- 2) time, in cycles
- 3) operational setting 1
- 4) operational setting 2
- 5) operational setting 3
- 6) sensor measurement 1
- 7) sensor measurement 2
- ...
- 26) sensor measurement 26

Σε αυτό το σύνολο δεδομένων, ο στόχος είναι να συγκριθούν οι αλγόριθμοι που χρησιμοποιούνται για την πρόβλεψη της υπολειπόμενης ωφέλιμης ζωής (RUL) κάθε κινητήρα στο σύνολο δεδομένων δοκιμής. Το RUL ισοδυναμεί με τον αριθμό των πτήσεων που απομένουν για τον κινητήρα μετά το τελευταίο σημείο δεδομένων στο σύνολο δεδομένων δοκιμής. Ένας πρώτος στόχος είναι να καταλάβουμε τα δεδομένα μας και να δώσουμε ανάλογες ονομασίες. Αξίζει επίσης να σημειωθεί ότι το σύνολο δεδομένων εκπαίδευσης έχει δεδομένα μέχρι το σημείο αστοχίας, επομένως ο μέγιστος κύκλος για κάθε κινητήρα ήταν η διάρκεια ζωής που επιτεύχθηκε από τον κινητήρα.

```
In [3]: 1 df_train.head()
```

Out[3]:

(LPC outlet temperature) (-R)	(HPC outlet temperature) (-R)	(LPT outlet temperature) (-R)	(Fan inlet Pressure) (psia)	...	(Ratio of fuel flow to Ps30) (pps/psia)	(Corrected fan speed) (rpm)	(Corrected core speed) (rpm)	(Bypass Ratio)	(Burner fuel-air ratio)	(Bleed Enthalpy)	(Required fan speed)	(Required fan conversion speed)	(High-pressure turbines Cool air flow)	(Low-pressure turbines Cool air flow)
641.82	1589.70	1400.60	14.62	...	521.66	2388.02	8138.62	8.4195	0.03	392	2388	100.0	39.06	23.4190
642.15	1591.82	1403.14	14.62	...	522.28	2388.07	8131.49	8.4318	0.03	392	2388	100.0	39.00	23.4236
642.35	1587.99	1404.20	14.62	...	522.42	2388.03	8133.23	8.4178	0.03	390	2388	100.0	38.95	23.3442
642.35	1582.79	1401.87	14.62	...	522.86	2388.08	8133.83	8.3682	0.03	392	2388	100.0	38.88	23.3739
642.37	1582.85	1406.22	14.62	...	522.19	2388.04	8133.80	8.4294	0.03	393	2388	100.0	38.90	23.4044

```
In [3]: 1 df_train.head()
```

Out[3]:

engine	cycle	setting_1	setting_2	setting_3	(Fan inlet temperature) (-R)	(LPC outlet temperature) (-R)	(HPC outlet temperature) (-R)	(LPT outlet temperature) (-R)	(Fan inlet Pressure) (psia)	...	(Ratio of fuel flow to Ps30) (pps/psia)	(Corrected fan speed) (rpm)	(Corrected core speed) (rpm)	
0	1	1	-0.0007	-0.0004	100.0	518.67	641.82	1589.70	1400.60	14.62	...	521.66	2388.02	8138.62
1	1	2	0.0019	-0.0003	100.0	518.67	642.15	1591.82	1403.14	14.62	...	522.28	2388.07	8131.49
2	1	3	-0.0043	0.0003	100.0	518.67	642.35	1587.99	1404.20	14.62	...	522.42	2388.03	8133.23
3	1	4	0.0007	0.0000	100.0	518.67	642.35	1582.79	1401.87	14.62	...	522.86	2388.08	8133.83
4	1	5	-0.0019	-0.0002	100.0	518.67	642.37	1582.85	1406.22	14.62	...	522.19	2388.04	8133.80

5 rows x 26 columns

Σε αυτό το σημείο πρέπει να προετοιμάσουμε το σύνολο δεδομένων, σε μια αρχική προσπάθεια να διακρίνουμε ποιές μεταβλητές είναι οι κατάλληλες για τους σκοπούς της εφαρμογής και ποιές είναι δυνατό να παραλειφθούν. Θα χρησιμοποιηθεί μια τεχνική ονόματι Backward Stepwise Regression. Η Backward Stepwise Regression είναι μια προσέγγιση σταδιακής παλινδρόμησης που ξεκινά με ένα πλήρες (κορεσμένο) μοντέλο και σε κάθε βήμα εξαλείφει σταδιακά τις μεταβλητές από το μοντέλο παλινδρόμησης για να βρεθεί ένα μειωμένο μοντέλο που εξηγεί

καλύτερα τα δεδομένα. Εφαρμόζοντάς τη στο μοντέλο, προκύπτουν οι μεταβλητές “worst features” προς απόρριψη και οι “selected features” προς διατήρηση στο μοντέλο: [28]

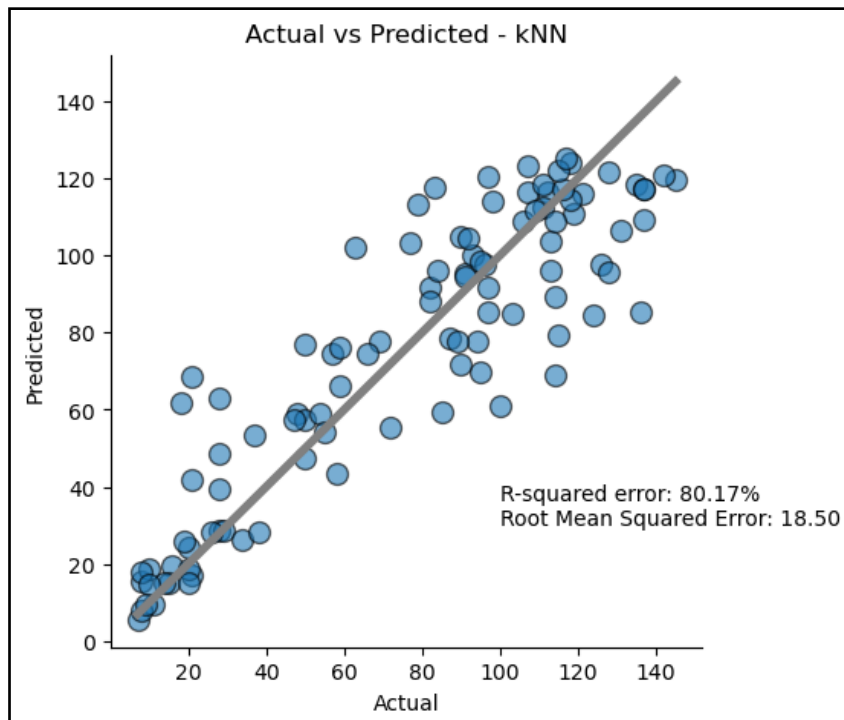
```
worst_feature : setting_1, 0.35895295901674085
worst_feature : setting_2, 0.18229853654640568

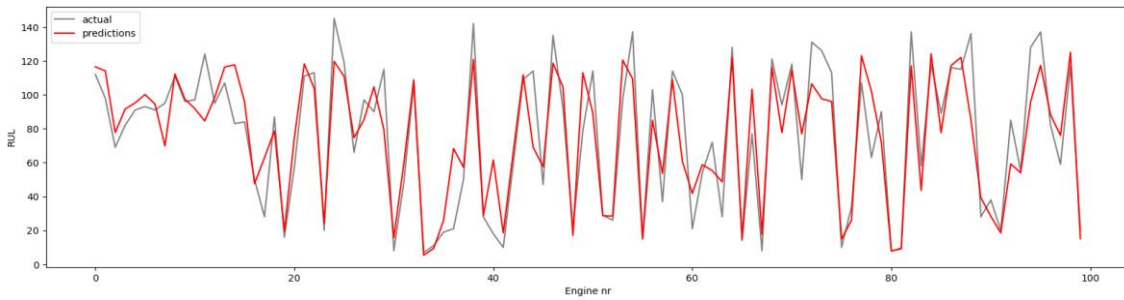
Selected Features:
['cycle', 'setting_3', '(Fan inlet temperature) (°R)', '(LPC outlet temperature) (°R)', '(HPC outlet temperature) (°R)', '(LPT outlet temperature) (°R)', '(Fan inlet Pressure) (psia)', '(bypass-duct pressure) (psia)', '(HPC outlet pressure) (psia)', '(Physical fan speed) (rpm)', '(Physical core speed) (rpm)', '(Engine pressure ratio(P50/P2))', '(HPC outlet Static pressure) (psia)', '(Ratio of fuel flow to Ps30) (pps/psia)', '(Corrected fan speed) (rpm)', '(Corrected core speed) (rpm)', '(Bypass Ratio)', '(Burner fuel-air ratio)', '(Bleed Enthalpy)', '(Required fan speed)', '(Required fan conversion speed)', '(High-pressure turbines Cool air flow)', '(Low-pressure turbines Cool air flow)']
```

Κατόπιν κανονικοποίησης των δεδομένων στο διάστημα [0,1] με τη βοήθεια του MinMaxScaler της βιβλιοθήκης Python sklearn.preprocessing, φτάνουμε στο σημείο εκπαίδευσης του μοντέλου και χρήσης των αλγορίθμων που αναφέρθηκαν στο Κεφάλαιο 4. Μιας και οι αλγόριθμοι έρχονται αρκετά κοντά στα αποτελέσματα που παράγουν και προς ευκολότερη ανάγνωση των γραφημάτων, θα χρησιμοποιηθεί ίδιος τρόπος οπτικοποίησης και στους 4 αλγορίθμους.

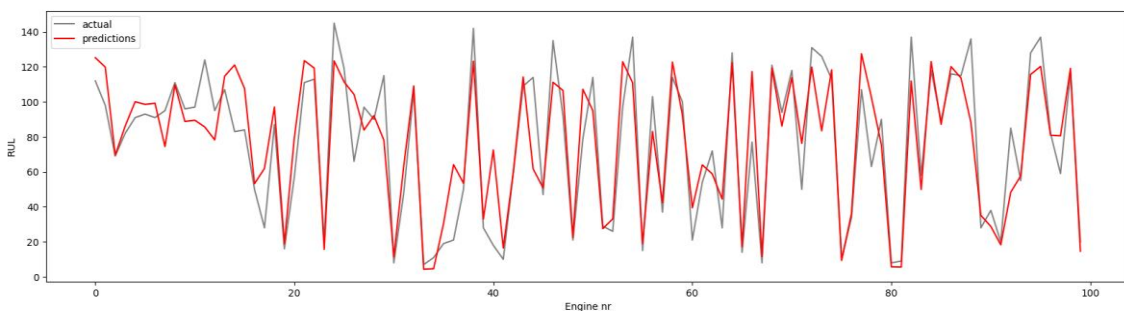
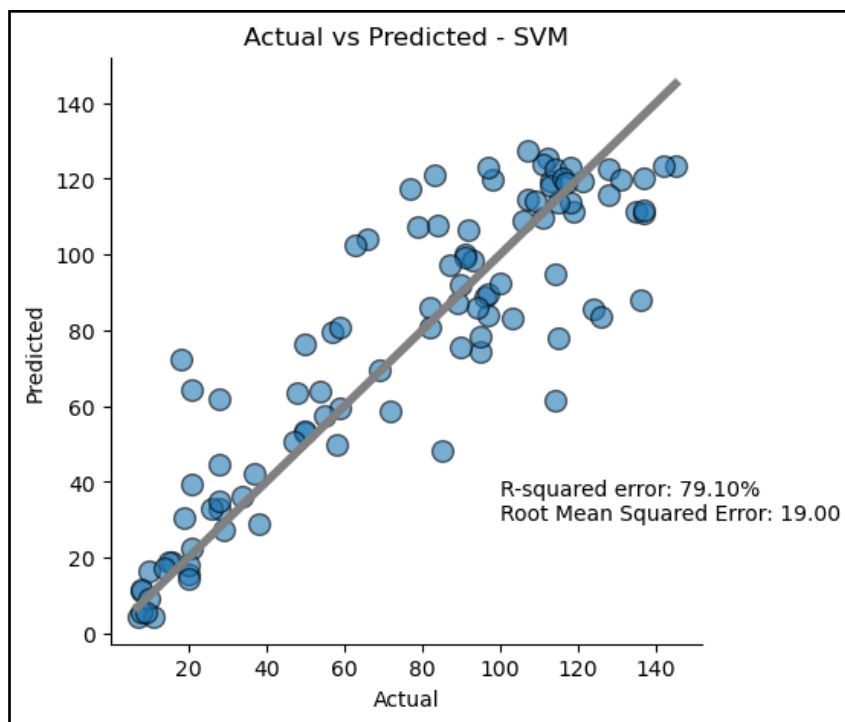
Πρέπει σε αυτό το σημείο να αναφερθεί πως στα γραφήματα συγκρίνονται οι τιμές actual vs predicted σε ότι αφορά το RUL της κάθε μηχανής, με actual να είναι οι πραγματικές τιμές των δεδομένων και predicted αυτές που παράγει ο κάθε αλγόριθμος, αναπαριστάμενες με γκρι γραμμή και γαλάζιες κουκκίδες αντίστοιχα. Τέλος, στα γραφήματα επισημαίνονται και οι τιμές error R-squared (R2) και Root Mean Square Error (RMSE). Για την R2 σε γενικές γραμμές αναζητούμε τιμές όσο το δυνατόν πιο κοντά στο 100%, ενώ για την RMSE όσο χαμηλότερα τόσο το καλύτερο.

Πρώτος αλγόριθμος που θα εξετάσουμε θα είναι ο kNN.

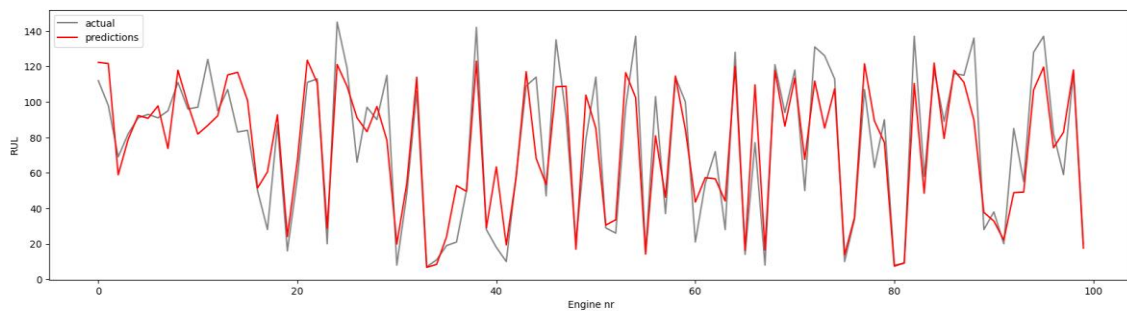
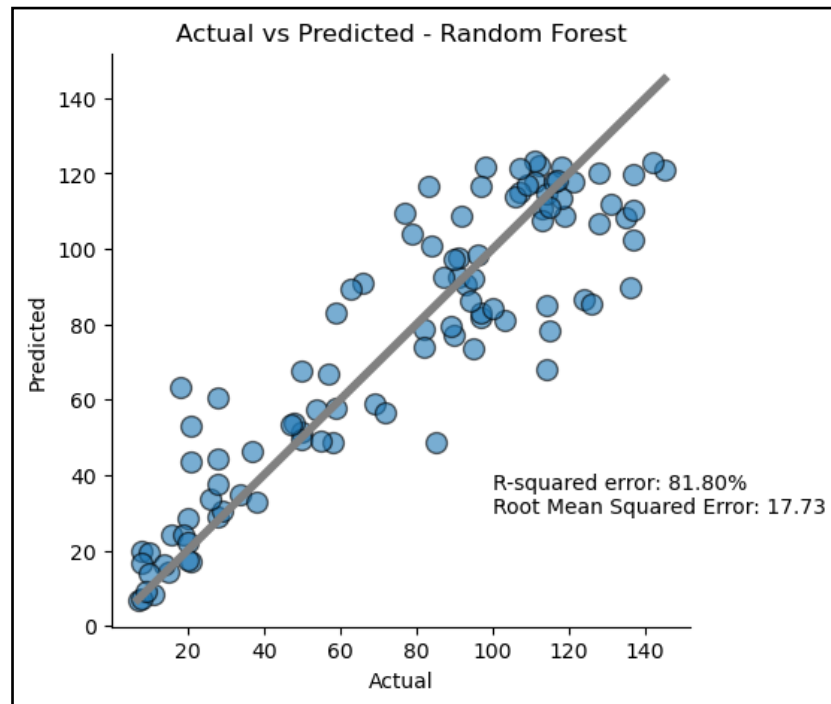




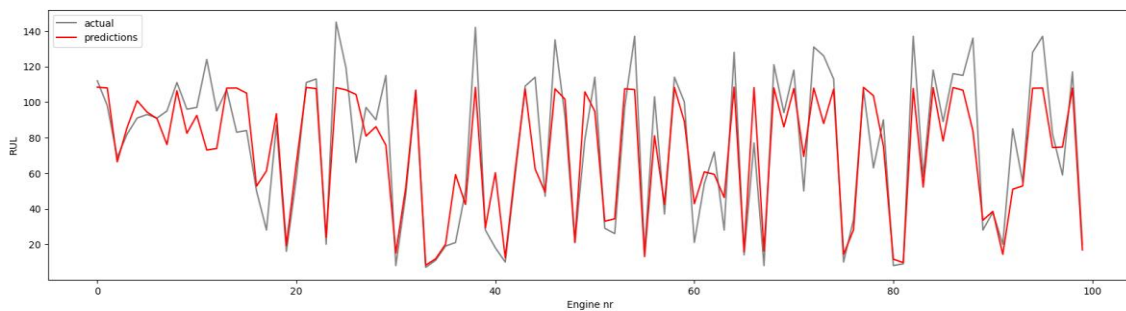
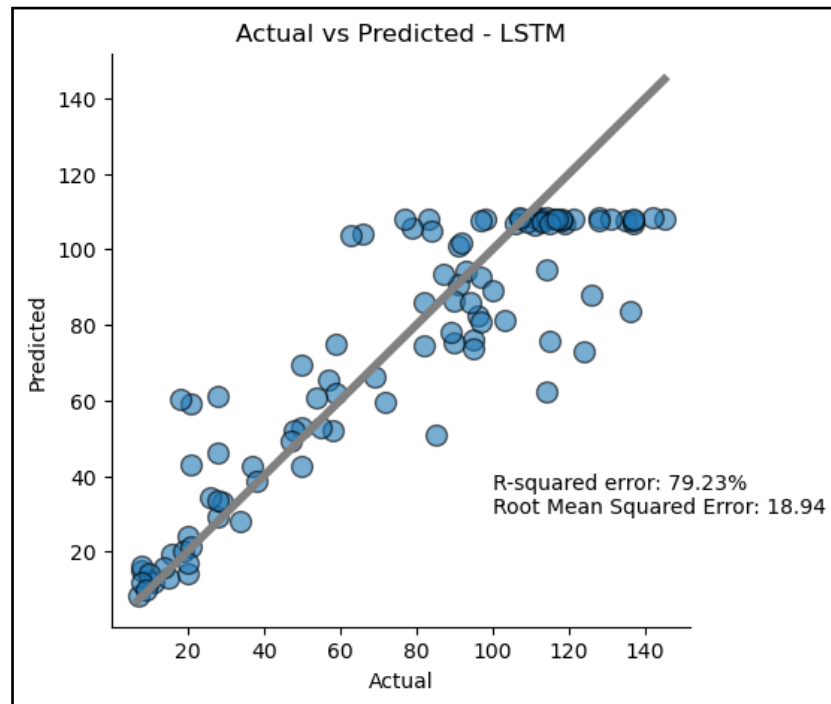
Ακολούθως έχουμε τον SVM:



Τρίτος θα είναι ο Random Forest:



Και τέλος, ο LSTM:



Εκ πρώτης όψεως, παρατηρείται πως οι διαφορές είναι αρκετά μικρές και για αυτόν τον λόγο θα συνοψίσουμε τις τιμές R2, RMSE, αλλά και τους χρόνους (σε δευτερόλεπτα) που χρειάστηκαν για την εκπαίδευση και την πρόβλεψη του κάθε αλγορίθμου ξεχωριστά:

	<b>R2</b>	<b>RMSE</b>	<b>time to train</b>	<b>time to predict</b>	<b>total time</b>
<b>kNN</b>	80.17%	18.50	0.002	0.309	0.311
<b>SVM</b>	79.10%	19.00	25.935	0.232	26.167
<b>Random Forest</b>	81.80%	17.73	7.929	0.065	7.993
<b>LSTM</b>	79.23%	18.94	16.410	0.989	17.399

Κατόπιν εξέτασης των ευρημάτων του παραπάνω πίνακα, παρατηρούμε πως για το σενάριο που εξετάζουμε και με τα δεδομένα που διαθέτουμε, ο αλγόριθμος Random Forest φαίνεται να επικρατεί του συνόλου σε ότι αφορά τους δείκτες R2 και RMSE, με αρκετά μικρές όμως διαφορές. Παρατηρείται επίσης αρκετά υψηλή απόδοση στον χρόνο πρόβλεψής του, με μόλις 0.065 δευτερόλεπτα, τη στιγμή που ο αμέσως γρηγορότερος βρίσκεται στα 0.232 δευτερόλεπτα (SVM). Αξίζει παρόλα αυτά να επισημάνουμε πως ο γρηγορότερος συνολικά αλγόριθμος ήταν ο kNN, με μόλις 0.311 δευτερόλεπτα συνολικό χρόνο εκτέλεσης (εκπαίδευσης + πρόβλεψης), με δεύτερο κατά σειρά τον Random Forest στα 7.993 δευτερόλεπτα.

Βάσει των παραπάνω παρατηρήσεων προκύπτουν διάφοροι προβληματισμοί. Δεδομένου του παρόντος dataset, στο σενάριο όπου αναζητούμε την μεγαλύτερη δυνατή ακρίβεια, ο Random Forest είναι κατά λίγες ποσοστιαίες μονάδες καλύτερος από τους υπόλοιπους 3 αλγόριθμους. Ένα σημείο που αξίζει να εστιάσουμε όμως είναι στην ταχύτητα εκπαίδευσης και πρόβλεψης. Ο kNN εδώ φαίνεται να κυριαρχεί, με πολύ μικρότερο απαιτούμενο συνολικό χρόνο και ελάχιστα χειρότερους δείκτες R2 και RMSE συγκριτικά με τον Random Forest. Σε ένα υποθετικό σενάριο με πολλαπλάσιες φορές μεγαλύτερο σετ δεδομένων, και άρα πολύ μεγαλύτερη απαίτηση υπολογιστικής ισχύος, ενδεχομένως να παρατηρούσαμε διαφορά στον παραπάνω πίνακα. Εν κατακλείδι, θα μπορούσαμε να πούμε πως και ο kNN θα μπορούσε να χρησιμοποιηθεί εξίσου καλά με τον Random Forest, ιδίως σε περιπτώσεις αρκετά μεγάλων σετ δεδομένων όπου μαζί με την ακρίβεια των προβλέψεων μας ενδιαφέρει και ο χρόνος στον οποίο παράγονται τα επιθυμητά αποτελέσματα.

## **ΚΕΦΑΛΑΙΟ 9: Προκλήσεις και προβληματισμοί στην προγνωστική συντήρηση στο IoT**

Η σύγκλιση των τεχνολογιών Διαδικτύου των πραγμάτων (IoT) και η προγνωστική συντήρηση προαναγγέλλει μια νέα εποχή προληπτικής διαχείρισης περιουσιακών στοιχείων και λειτουργικής βελτιστοποίησης. Ωστόσο, αυτή η συγχώνευση διαφορετικών τομέων φέρνει επίσης στο προσκήνιο μια μυριάδα προκλήσεων και προβληματισμών που πρέπει να αντιμετωπίζονται με ακρίβεια και προνοητικότητα. Αυτό το κεφάλαιο εμβαθύνει στο πολύπλευρο τοπίο της προγνωστικής συντήρησης που βασίζεται στο IoT, ξετυλίγοντας τις πολυπλοκότητες και διαφωτίζοντας στρατηγικές για την υπέρβαση των εμποδίων και την αξιοποίηση ευκαιριών για μεταμορφωτική αλλαγή.

### **9.1 Συζήτηση των προκλήσεων που σχετίζονται με την εφαρμογή της προγνωστικής συντήρησης σε περιβάλλοντα IoT**

Η εφαρμογή της προγνωστικής συντήρησης σε περιβάλλοντα IoT εισάγει μια πληθώρα προκλήσεων που καλύπτουν τεχνικά, λειτουργικά και οργανωτικά πεδία. Η κύρια μεταξύ αυτών των προκλήσεων είναι η ενοποίηση και η διαλειτουργικότητα ετερογενών συσκευών και συστημάτων IoT, που απαιτούν απρόσκοπτα πρωτόκολλα επικοινωνίας και ανταλλαγής δεδομένων για την εναρμόνιση διαφορετικών πηγών δεδομένων. Επιπλέον, ο τεράστιος όγκος, η ταχύτητα και η ποικιλία των δεδομένων που παράγονται από τις συσκευές IoT δημιουργούν προκλήσεις επεκτασιμότητας και επεξεργασίας, απαιτώντας επεκτάσιμη υποδομή και καταναεμημένα υπολογιστικά πρότυπα για την αντιμετώπιση του αυξανόμενου κατακλυσμού δεδομένων. Επιπλέον, η δυναμική και ετερογενής φύση των περιβαλλόντων IoT εισάγει πολυπλοκότητες στη μοντελοποίηση και την πρόβλεψη, απαιτώντας προσαρμοστικούς αλγόριθμους και τεχνικές ικανές να καταγράφουν εξελισσόμενα μοτίβα και ανωμαλίες σε πραγματικό χρόνο.

### **9.2 Θέματα για την αντιμετώπιση προβλημάτων ασφάλειας, απορρήτου και διαλειτουργικότητας**

Η ασφάλεια, το απόρρητο και η διαλειτουργικότητα αναδεικνύονται ως πρωταρχικά ζητήματα στις πρωτοβουλίες πρόβλεψης συντήρησης που βασίζονται στο IoT, υποστηρίζοντας την εμπιστοσύνη, την ακεραιότητα και την ανθεκτικότητα στα λειτουργικά οικοσυστήματα. Ισχυρά μέτρα ασφαλείας, συμπεριλαμβανομένης της κρυπτογράφησης, του ελέγχου ταυτότητας και του ελέγχου πρόσβασης, είναι απαραίτητα για την προστασία ευαίσθητων δεδομένων και τον μετριασμό των απειλών στον κυβερνοχώρο σε περιβάλλοντα IoT. Ομοίως, οι τεχνικές διατήρησης του απορρήτου, όπως το διαφορικό απόρρητο (differential privacy) και η ανωνυμοποίηση, προστατεύουν το απόρρητο των χρηστών και τηρούν τις αρχές κυριαρχίας των δεδομένων. Τα πρότυπα και τα πρωτόκολλα διαλειτουργικότητας, όπως τα MQTT, CoAP και OPC UA, διευκολύνουν την απρόσκοπτη ενοποίηση και επικοινωνία μεταξύ διαφορετικών συσκευών και πλατφορμών IoT, ενισχύοντας τη διαλειτουργικότητα και την ανταλλαγή δεδομένων μεταξύ διαφορετικών συστημάτων και προμηθευτών.

### 9.3 Διερεύνηση αναδυόμενων τάσεων και τεχνολογιών στον τομέα

Ανάμεσα στις μυριάδες προκλήσεις και προβληματισμούς, οι αναδυόμενες τάσεις και τεχνολογίες προσφέρουν ελπίδες και ευκαιρίες για καινοτομία στην προγνωστική συντήρηση που βασίζεται στο IoT. Παραδείγματα υπολογιστών αιχμής, μόχλευση κατανεμημένης νοημοσύνης και ικανότητας επεξεργασίας στην περιφέρεια του δικτύου, μετριάζουν τους περιορισμούς καθυστέρησης και εύρους ζώνης, επιτρέποντας την ανάλυση σε πραγματικό χρόνο και τη λήψη αποφάσεων στην άκρη. Ομοίως, οι εξελίξεις στην τεχνητή νοημοσύνη και τη μηχανική μάθηση, συμπεριλαμβανομένης της συνεργατικής μάθησης (collaborative/federated learning) και της εκπαίδευσης μοντέλων στη συσκευή, ενισχύουν τις συσκευές αιχμής να εκτελούν αυτόνομα προγνωστικές αναλύσεις, μειώνοντας την εξάρτηση από την κεντρική υποδομή και ενισχύοντας το απόρρητο και την επεκτασιμότητα. Επιπλέον, η τεχνολογία blockchain, με το αμετάβλητο καθολικό και την αποκεντρωμένη αρχιτεκτονική της, υπόσχεται την ενίσχυση της εμπιστοσύνης, της διαφάνειας και της προέλευσης στις προγνωστικές ροές εργασιών συντήρησης, διευκολύνοντας την ασφαλή κοινή χρήση δεδομένων και τη δυνατότητα ελέγχου στα κατανεμημένα δίκτυα.

Αντιμετωπίζοντας τις προκλήσεις κατά μέτωπο και αγκαλιάζοντας τις αναδυόμενες τάσεις και τεχνολογίες, οι οργανισμοί μπορούν να χαράξουν μια πορεία προς την καλύτερη δυνατή προβλεπτική συντήρηση σε περιβάλλοντα IoT. Μέσω στρατηγικών επενδύσεων σε μέτρα ασφάλειας, προστασίας της ιδιωτικής ζωής και διαλειτουργικότητας, σε συνδυασμό με τη δέσμευση για καινοτομία και συνεργασία, οι επαγγελματίες μπορούν να ξεκλειδώσουν το μετασχηματιστικό δυναμικό της προγνωστικής συντήρησης που βασίζεται στο IoT, την αύξηση της λειτουργικής αποτελεσματικότητας, την αξιοπιστία των περιουσιακών στοιχείων και την οργανωτική ανθεκτικότητα σε ένα ολοένα και πιο διασυνδεδεμένο περιβάλλον και έναν οδηγούμενο από δεδομένα κόσμο.

### 9.4 Σχετικές Μελέτες

*Lee, J., et al. (2014) – Service Innovation and Smart Analytics for Industry 4.0 and Big Data Environment:* Το έργο εξετάζει τις προκλήσεις της εφαρμογής προγνωστικής συντήρησης σε περιβάλλοντα IoT, συμπεριλαμβανομένης της ενοποίησης δεδομένων, της επεκτασιμότητας και της πολυπλοκότητας των αναλυτικών στοιχείων.

*Al-Turjman, F., & Malekloo, A. (2019) – Smart Environments in IoT Era: Concept, Challenges, and Future Directions:* Αντιμετωπίζει βασικά ζητήματα, όπως το απόρρητο, η ασφάλεια και η αξιοπιστία των δεδομένων στα συστήματα IoT, τα οποία επηρεάζουν άμεσα την υλοποίηση της προγνωστικής συντήρησης.



## Βιβλιογραφία

- [1] Prognostics and Health Management of Electronics, Michael G. Pecht, Myeongsu Kang, 2018
- [2] A Survey of Predictive Maintenance: Systems, Purposes and Approaches, Tianwen Zhu, Yongyi Ran, Xin Zhou, and Yonggang Wen, 2019
- [3] An Introduction to Predictive Maintenance, R. Keith Mobley, 2002
- [4] Decision Making in Predictive Maintenance: Literature Review and Research Agenda for Industry 4.0, Alexandros Bousdekis, Katerina Lepenioti, Dimitris Apostolou, Gregoris Mentzas, 2019
- [5] IoT based Predictive Maintenance in Manufacturing Sector, Shikhil Nangiaa, Sandhya Makkarb, Rohail Hassanc, 2020
- [6] Predictive Maintenance – Bridging Artificial Intelligence and IoT, Gerasimos G. Samatas, Seraphim S. Moumgiakmas, George A. Papakostas, 2021
- [7] What Is IoT Predictive Maintenance?, Anthony Moffa, 2023
- [8] Understanding K-Nearest Neighbors (KNN) Regression in Machine Learning, Nandini Verma, 2023
- [9] KNN algorithm: Introduction to K-Nearest Neighbors Algorithm for Regression, Aishwarya Singh, 2024
- [10] Guide on Support Vector Machine (SVM) Algorithm, Anshul Saini, 2024
- [11] <https://www.geeksforgeeks.org/support-vector-machine-algorithm/>, 2024
- [12] Random Forest Regression: A basic explanation and use case in 7 minutes, Nima Beheshti, 2022
- [13] A Comprehensive Guide to Random Forest Regression, Shrinath Bhat, 2023
- [14] What is LSTM? Introduction to Long Short-Term Memory, Shipra Saxena, 2021
- [15] Long Short-Term Memory (LSTM), Saba Hesarak, 2023
- [16] Algorithm selection for Anomaly Detection, Sahil Garg, 2020
- [17] Anomaly Detection in Machine Learning, Vikram Singh, 2024
- [18] Industrial Internet of Things monitoring solution for advanced predictive maintenance applications Federico Civerchia, Stefano Bocchino, Claudio Salvadori, Enrico Rossi, Luca Maggiani, Matteo Petracca, 2017
- [19] An intelligent approach for data pre-processing and analysis in predictive maintenance with an industrial case study, Ebru Turanoglu Bekar, Per Nyqvist, and Anders Skoogh, 2020
- [20] AN INTEGRATED APPROACH TO PREDICTIVE MAINTENANCE USING IOT AND MACHINE LEARNING IN MANUFACTURING, Sumanth Tatineni, 2020
- [21] IoT Data Preprocessing - A Survey, V.A. Jane, Dr. L. Arockiam, 2021
- [22] Effective Strategies for Handling Missing Values in Data Analysis, Nasima Tamboli, 2024
- [23] Here's How IoT Data Collection Works [Complete Guide], Agustin Pelaez, 2024
- [24] <https://dataheadhunters.com/academy/how-to-use-python-for-predictive-maintenance-in-manufacturing>, 2024
- [25] Building a Predictive Maintenance Application using Python with Scikit-Learn for Machine Learning, Ketan Raval, 2024
- [26] Performance Metrics in Machine Learning [Complete Guide], Aayush Bajaj, 2023

- [27] Top Performance Metrics in Machine Learning: A Comprehensive Guide, Deval Shah, 2023
- [28] House Price (step-by-step modeling) – Backwards Regression Technique, Adib Ouayjan, 2022
- [29] Machine learning and reasoning for predictive maintenance in Industry 4.0: Current status and challenges, Jovani
- [30] Dalzochio, Rafael Kunst, Edison Pignaton, Alecio Binotto, Srijnan Sanyal, Jose Favilla, Jorge Barbosa, 2020

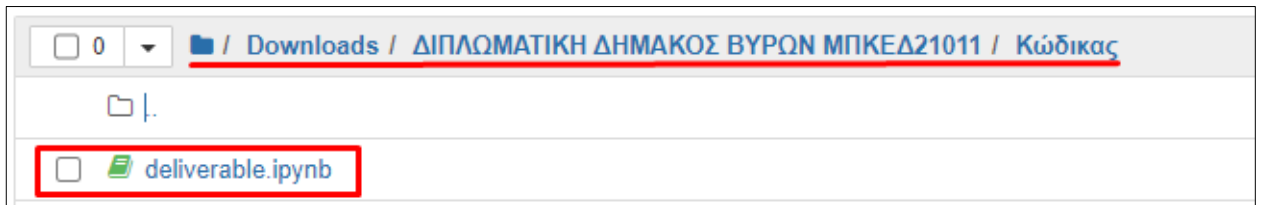
## Οδηγίες Εγκατάστασης Jupyter Notebook

<https://jupyter.org/install>

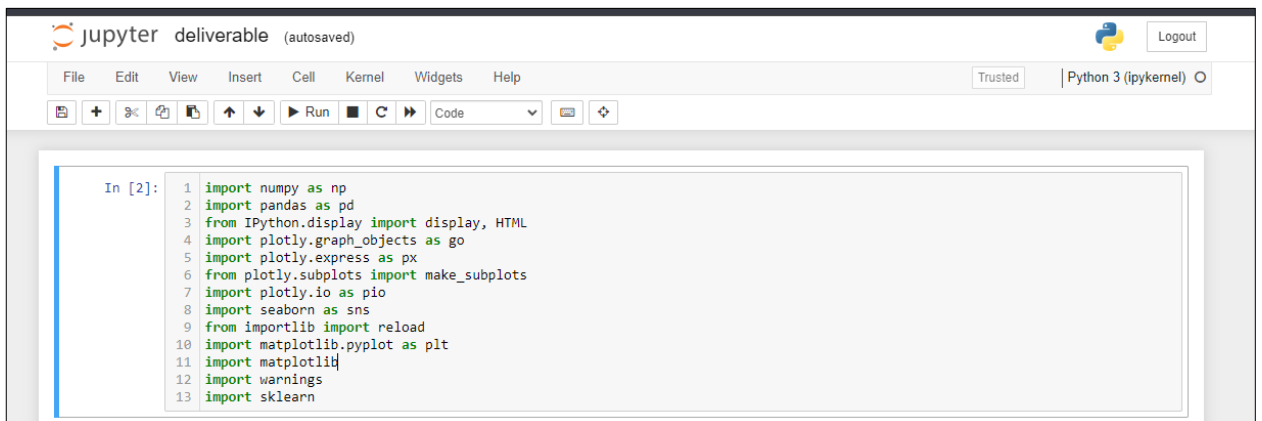
Ο παραπάνω σύνδεσμος περιέχει αναλυτικές οδηγίες για την σωστή εγκατάσταση του Jupyter Notebook σε κάθε σύστημα. Επιπροσθέτως, εμπεριέχονται οδηγίες και για την εγκατάσταση προαπαιτούμενων εργαλείων, όπως το pip.

## Οδηγίες εκτέλεσης των αρχείων της διπλωματικής

Αφού εγκατασταθεί επιτυχώς το Jupyter Notebook, μέσω της κεντρικής οθόνης, στην καρτέλα Files, πηγαίνουμε στην τοποθεσία του υπολογιστή που βρίσκεται το deliverable.ipynb, δηλαδή το κεντρικό παραδοτέο αρχείο κώδικα της διπλωματικής, και το κλικάρουμε ώστε να ανοίξει σε νέα καρτέλα.

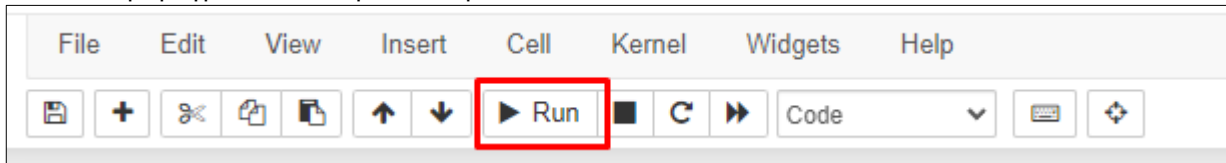


Αφού ανοίξει, παρατηρούμε blocks κώδικα, όπως το κάτωθι:

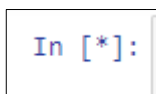


Η κάθετη μπλε γραμμή στην αριστερή πλευρά του block σημαίνει ότι το έχουμε επιλεγμένο, και καθώς ο τρόπος λειτουργίας του Jupyter είναι σειριακή εκτέλεση των blocks, πρέπει να ξεκινήσουμε από το πρώτο και να τα εκτελούμε με τη σειρά περιμένοντας την ολοκλήρωση του ενός block προτού προχωρήσουμε στο επόμενο.

Η εκτέλεση πραγματοποιείται με το κουμπί Run.



Πρέπει να σημειωθεί σε αυτό το σημείο πως η ταχύτητα εκτέλεσης του κάθε block διαφέρει από σύστημα σε σύστημα. Με το κουμπί Run, ενώ αλλάζει αμέσως το επιλεγμένο block και πηγαίνει στο επόμενο, η αλλαγή δεν ταυτίζεται με τη στιγμή περάτωσης εκτέλεσης του κώδικα στο προηγούμενο block. Η ένδειξη που μας επιβεβαιώνει αν το block τελείωσε ή όχι είναι ο κάτωθι αστερίσκος στις αγκύλες [\*].



Αν υπάρχει αστερίσκος, τότε πρέπει να περιμένουμε έως ότου εξαφανιστεί ώστε να προχωρήσουμε στην επόμενη εκτέλεση.

Τέλος, επειδή η εφαρμογή χρησιμοποιεί εξωτερικά αρχεία δεδομένων, υπάρχει ένα σημείο στον κώδικα στον οποίο υποδηλώνουμε την τοποθεσία του φακέλου των αρχείων αυτών, στο σύστημα στο οποίο εκτελούμε την εφαρμογή. Το σημείο αυτό είναι το δεύτερο block, στις τελευταίες 3 γραμμές κώδικα:

```

26 df_train = pd.read_csv('C:/Users/Byron/Downloads/unipi files/ptyxiaki/train.txt'), sep='\s+', header=None, names=col_names)
27
28 df_test = pd.read_csv('C:/Users/Byron/Downloads/unipi files/ptyxiaki/test.txt'), sep='\s+', header=None, names=col_names)
29 df_test_RUL = pd.read_csv('C:/Users/Byron/Downloads/unipi files/ptyxiaki/RUL.txt'), sep='\s+', header=None, names=['RUL'])
30

```

Είναι εμφανές πως τα 3 αυτά αρχεία είναι τα RUL.txt, test.txt και train.txt τα οποία συμπεριλαμβάνονται στον παραδοτέο φάκελο μαζί με το deliverable αρχείο κώδικα. Προκειμένου να τρέξει σωστά η εφαρμογή, θα πρέπει να αλλάξει αυτή η ονομασία με τον παρακάτω τρόπο. Έστω ότι τα αρχεία βρίσκονται στο κάτωθι path στον φάκελο «Κώδικας»:

C:\Users\Byron\Downloads\ΔΙΠΛΩΜΑΤΙΚΗ ΔΗΜΑΚΟΣ ΒΥΡΩΝ ΜΠΚΕΔ21011\Κώδικας

Θα πρέπει το path λοιπόν και στις 3 γραμμές να είναι της μορφής 'path/όνομα αρχείου', δηλαδή:

C:/Users/Byron/Downloads/ΔΙΠΛΩΜΑΤΙΚΗ ΔΗΜΑΚΟΣ ΒΥΡΩΝ ΜΠΚΕΔ21011/Κώδικας/train.txt  
 C:/Users/Byron/Downloads/ΔΙΠΛΩΜΑΤΙΚΗ ΔΗΜΑΚΟΣ ΒΥΡΩΝ ΜΠΚΕΔ21011/Κώδικας/test.txt  
 C:/Users/Byron/Downloads/ΔΙΠΛΩΜΑΤΙΚΗ ΔΗΜΑΚΟΣ ΒΥΡΩΝ ΜΠΚΕΔ21011/Κώδικας/RUL.txt

Μεγάλη προσοχή πρέπει να δοθεί στην κατεύθυνση των slashes (σύμβολο /) μιας και τα paths κατά την αντιγραφή από τον Windows File Explorer έρχονται με \ ενώ στην εφαρμογή πρέπει να γραφτούν με /.

Ακολούθως, όλη η εφαρμογή εκτελείται κανονικά με το προαναφερθέν κουμπί Run με προϋπόθεση την αναμονή εκτέλεσης του κώδικα μεταξύ blocks.

Σε περίπτωση επιθυμίας περαιτέρω οδηγιών ή/και ύπαρξης αποριών σχετικά με το Jupyter Notebook, αξίζει να μελετηθούν οι παρακάτω σύνδεσμοι:

<https://docs.jupyter.org/en/latest/> : Το επίσημο documentation site για το Jupyter Notebook.

<https://www.dataquest.io/blog/jupyter-notebook-tutorial/>: Εναλλακτικό tutorial χρήσης του Jupyter Notebook.