



**ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΙΓΑΙΟΥ**  
**ΠΟΛΥΤΕΧΝΙΚΗ ΣΧΟΛΗ**

ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ ΠΛΗΡΟΦΟΡΙΑΚΩΝ ΚΑΙ ΕΠΙΚΟΙΝΩΝΙΑΚΩΝ ΣΥΣΤΗΜΑΤΩΝ

ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ

Internet of Things, Διαδίκτυο των Πραγμάτων

ΤΙΤΛΟΣ

**Εφαρμογή Μεθόδων Μηχανικής Μάθησης για την  
Πρόβλεψη Αστοχιών σε Αεριοστρόβιλους.**

**ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ**

ΤΟΥ

**Γαϊτανάκης Γιώργος-Κωνσταντίνος**

**Επιβλέπων :** κ. Σταματάτος Ευστάθιος

**Μέλη εξεταστικής επιτροπής:** Καβαλλιεράτου Εργίνα, Κωστούλας Θεόδωρος

Σάμος, 06/2023

Η σελίδα αυτή είναι σκόπιμα λευκή

## Πρόλογος και ευχαριστίες

Η παρούσα διπλωματική εργασία εκπονήθηκε στα πλαίσια του μεταπτυχιακού προγράμματος σπουδών με τίτλο “Internet of Things, διαδίκτυο των πραγμάτων” του τμήματος Μηχανικών Πληροφοριακών και Επικοινωνιακών Συστημάτων του Πανεπιστημίου Αιγαίου. Θα ήθελα να ευχαριστήσω θερμά τον επιβλέποντα της διπλωματικής κύριο Ευστάθιο Σταματάτο, Καθηγητή και αντιπρόεδρο του Τμήματος για την σημαντική βοήθεια του, καθώς επίσης και για το χρόνο που μου αφιέρωσε. Δεν θα μπορούσα να παραλείψω την οικογένεια μου και ιδιαίτερα την Μαρία για την ενθάρρυνση, την υποστήριξη και τη βοήθεια που μου προσέφερε όλο αυτό το διάστημα.

© 2023

του

ΓΑΪΤΑΝΑΚΗ ΓΕΩΡΓΟΥ-ΚΩΝΣΤΑΝΤΙΝΟΥ

Τμήμα Μηχανικών Πληροφοριακών και Επικοινωνιακών Συστημάτων

ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΙΓΑΙΟΥ

Η σελίδα αυτή είναι σκόπιμα λευκή.

## Πίνακας Περιεχομένων

Περίληψη.....	7
ABSTRACT .....	8
1.1 Εισαγωγή .....	9
1.2 Αξιοπιστία Συστημάτων .....	11
1.3 Τύποι συντήρησης βιομηχανικού εξοπλισμού .....	16
1.3.1 Τύποι Proactive Maintenance .....	16
1.4 Προγνωστική και Διαγνωστική Μηχανών .....	18
2.1 Ο Αεριοστρόβιλος.....	26
2.2 Προσέγγιση του Predictive Maintenance.....	29
2.3 DATASET .....	32
3.1 Τί είναι η Τεχνητή νοημοσύνη;.....	35
3.1.1 Οι βασικές κατηγορίες της Τεχνητή νοημοσύνης.....	36
3.2 Τί είναι η μηχανική μάθηση;.....	39
3.3 Είδη Μηχανικής Μάθησης.....	42
3.3.1 Η διαδικασία μάθησης ενός μοντέλου βάσης.....	43
3.3.2 Ensemble learning.....	47
4.1 Το Dataset.....	63
4.2 Μεθοδολογία.....	65
4.3 Συμπεράσματα .....	81
ΑΝΑΦΟΡΕΣ .....	83

Η σελίδα αυτή είναι σκόπιμα λευκή.

## Περίληψη

Η παρούσα εργασία πραγματεύεται την εκτίμηση της εναπομένουσας ζωής (RUL-Remaining Useful Life) ενός αεριοστρόβιλου με χρήση μεθόδων μηχανικής μάθησης. Στην εποχή των Big Data και της ανάλυσης δεδομένων η σωστή εκμετάλλευση ροών δεδομένων και η επεξεργασία τους μπορεί να φέρουν μεγάλα οφέλη για την βιομηχανία. Στην εργασία αυτή σκοπός είναι να βρεθούν τα κατάλληλα δεδομένα για αεριοστρόβιλους και στη συνέχεια να γίνει επεξεργασία τους με τα κατάλληλα μαθηματικά εργαλεία για να υλοποιηθεί η όσο το δυνατόν καλύτερη εκτίμηση του RUL. Τα μαθηματικά εργαλεία τα οποία χρησιμοποιήθηκαν προέρχονται από τον χώρο της Μηχανικής Μάθησης και συγκεκριμένα χρησιμοποιήθηκαν 3 αλγόριθμοι ο SVM (SVR), ο XGBoost και ο Random Forest. Τα δεδομένα προήλθαν από το Dataset της NASA FD001 και πραγματοποιήθηκε ανάλυση των δεδομένων με σκοπό να βρεθούν τα δεδομένα αυτά τα οποία βοηθούν τους αλγορίθμους να αποδώσουν καλύτερα.

Ο σκοπός για τον οποίο χρησιμοποιήθηκαν οι συγκεκριμένοι αλγόριθμοι είναι ότι στην βιβλιογραφία πάνω στο συγκεκριμένο Dataset παρόλο που έχουν γίνει χιλιάδες εφαρμογές με χρήση νευρωνικών δικτύων δεν έχει γίνει κάποια ικανοποιητική εφαρμογή των συγκεκριμένων αλγορίθμων, η οποία να έχει καταλήξει σε αξιόλογα αποτελέσματα, ενώ γενικά είναι αλγόριθμοι οι οποίοι πολλές φορές αποδίδουν καλύτερα ή δίνουν ανταγωνιστικά αποτελέσματα ως προς τα νευρωνικά δίκτυα. Στα αποτελέσματα όπως αυτά προέκυψαν στην παρούσα εργασία, φαίνεται ότι οι αλγόριθμοι μετά από συνδυασμό διάφορων τεχνικών προσθηκών απέδωσαν καλύτερα από υπάρχουσες λύσεις Νευρωνικών δικτύων οι οποίες υπάρχουν στην βιβλιογραφία. Επίσης βάσει των αποτελεσμάτων φαίνεται να είναι οι καλύτερες λύσεις οι οποίες αφορούν μεθόδους μη χρήσης νευρωνικών δικτύων.

## **ABSTRACT**

This study focuses on the estimation of the Remaining Useful Life (RUL) of a gas turbine using machine learning methods. In the era of Big Data and data analytics, proper utilization of data streams and processing can bring significant benefits to the industry. The objective of this study is to identify suitable data for gas turbines and then process them using appropriate mathematical tools to achieve the best possible RUL estimation. The mathematical tools used are derived from the field of Machine Learning, specifically the SVM (SVR) algorithm, XGBoost, and Random Forest. The data used in this study were obtained from the NASA FD001 dataset, and data analysis was performed to identify the features that help the algorithms perform better.

The reason for using these specific algorithms is that although there have been thousands of applications using neural networks on the same dataset, no satisfactory results have been achieved. These algorithms, on the other hand, often outperform or provide competitive results compared to neural networks. The results obtained in this study show that after combining various techniques and enhancements, the algorithms performed better than existing neural network solutions found in the literature. Furthermore, based on the results, they appear to be the best solutions among non-neural network methods.



## 1.1 Εισαγωγή

Με την ανάπτυξη της τεχνολογίας και την σταδιακή ένταξή της στον καθημερινό τρόπο ζωής, ο άνθρωπος καθημερινά χρησιμοποιεί πολλούς και διαφορετικούς μηχανικούς εξοπλισμούς οι οποίοι διευκολύνουν την εργασία του, την μετακίνηση του, την ασφάλεια του, την ψυχαγωγία του και διάφορους άλλους τομείς. Νομοτελειακά όλες οι μηχανικές κατασκευές από την μηχανή ενός αεροσκάφους μέχρι το τηλεκοντρόλ της τηλεόρασης του σπιτιού μας τείνουν στην αστοχία μέχρι να αποσυρθούν μόνιμα από τον χρήστη. Δηλαδή κατά την χρήση μιας μηχανής, η φθορά η οποία συντελείται στην μηχανή, αρχικά φθίνει την απόδοση της μέχρι ένα σημείο όπου κάποιο υποσύστημα της μηχανής αστοχεί και τελικά η μηχανή τίθεται εκτός λειτουργίας. Βασικός σκοπός είναι η τακτική συντήρηση του εξοπλισμού ώστε να αποφευχθεί η αστοχία και ο εξοπλισμός να λειτουργεί όσο πιο κοντά στις αρχικές του προδιαγραφές.

Τις τελευταίες δύο δεκαετίες με την μεγάλη αύξηση της ψηφιοποίησης της πληροφορίας και την δυνατότητα αποθήκευσης της, αλλά και της μείωσης του κόστους κατασκευής των υπολογιστικών συστημάτων και ολοκληρωμένων κυκλωμάτων, συντελέστηκε μεγάλη πρόοδος στο κομμάτι της παρακολούθησης (monitoring) και της καταγραφής μετρήσεων και επιδόσεων των μηχανικών κατασκευών. Μεγάλη ανάπτυξη επήλθε σε μεγάλες μηχανολογικές κατασκευές βιομηχανικού τύπου όπως γεννήτριες, κινητήρες, αεριοστροβίλους, αντλίες. Όπου η τοποθέτηση αισθητήρων σε σημαντικά-καίρια σημεία της μηχανής πέρα από την βοήθεια στην επιτήρηση του εξοπλισμού ανά πάσα στιγμή από ένα απομακρυσμένο κέντρο ελέγχου (Control room), βοήθησε και στην επίλυση τυχών προβλημάτων τα οποία προκύπτουν. Καθώς αν μια μηχανή αστοχήσει, για κάποιο χρονικό διάστημα πριν την αστοχία της, θα επηρεαστούν μεγέθη τα οποία καταγράφονται από τους αισθητήρες οι οποίοι έχουν τοποθετηθεί στην μηχανή. Τέτοια μεγέθη είναι η καταγραφή ταλαντώσεων, πίεσης, ροής, θερμοκρασίας λαδιών, θερμοκρασίας μεταλλικών επιφανειών, θέση ανοίγματος βάνας κλπ. Έτσι στο διάβα του χρόνου με την συσσώρευση όλων αυτών των καταγραφών για την μηχανή δημιουργείται μια μεγάλη δεξαμενή δεδομένων η εκμετάλλευση της οποίας μπορεί να φανεί ιδιαίτερα χρήσιμη.

Ενοποιώντας τις δύο αυτές φαινομενικά διαφορετικές κατευθύνσεις, πρώτον της τεχνικής συντήρησης μηχανών και δεύτερον της καταγραφής δεδομένων χρησιμοποιώντας κατάλληλα μαθηματικά εργαλεία θα μπορούσε να γίνει η βέλτιστη διαχείριση της μηχανής. Δηλαδή χρησιμοποιώντας τα δεδομένα τα οποία έχουν

καταγραφεί στο παρελθόν από την λειτουργία της μηχανής και συσχετίζοντας τα με τα δεδομένα που καταγράφονται την παρούσα στιγμή να γίνει μια πρόβλεψη για το αν ελλοχεύει κάποια αστοχία στο μέλλον και μια εκτίμηση για το πότε πρόκειται να αστοχήσει η μηχανή. Ο πιο πολλά υποσχόμενος τρόπος σήμερα για να γίνει αυτή η πρόβλεψη είναι με μεθόδους μηχανικής μάθησης (Machine Learning). Οπότε σκοπός είναι η εύρεση του αποδοτικότερου μοντέλου μηχανικής μάθησης το οποίο θα είναι σε θέση να προβλέπει-εκτιμά μια μελλοντική αστοχία της μηχανής. Δίνοντας έτσι στον χρήστη την ευκαιρία να λάβει όλα τα κατάλληλα μέτρα και να προβεί στις απαραίτητες ενέργειες ώστε να μην αστοχήσει η μηχανή.

Ο Sciban [1] στο άρθρο του δηλώνει την σημαντικότητα της διάκρισης μεταξύ προβλεπτικής και προληπτικής συντήρησης καθώς πολλές φορές μπορεί να συγχέονται οι όροι τους. Όλες οι προληπτικές συντηρήσεις έχουν ως βασικό γνώμονα και αρχή τον χρόνο. Η επισκευή ή επιδιόρθωση είναι οδηγούμενη από τον χρόνο και καθορίζεται από το MTTF (Mean time to Failure). Με λίγα λόγια η προληπτική συντήρηση προσπαθεί να μειώσει την πιθανότητα αστοχίας μιας μηχανής μέσα από τακτική συντήρηση (π.χ. κάθε εξάμηνο ή όταν ορίζει το manual του κατασκευαστή). Η προβλεπτική συντήρηση από την άλλη βασίζεται σε δεδομένα (Data) για να προσδιορίσει την πιθανότητα αστοχίας ενός εξοπλισμού προτού αυτή συμβεί. Έτσι ο χρήστης ή ο κατασκευαστής μπορεί να προβλέψει και να επιδιορθώσει. Η προβλεπτική συντήρηση σήμερα σε μια αναφορά από το IoT Analytics, έναν οδηγό της αγοράς του Internet of Things [2] [3], δείχνει ότι οι στρατηγικές συντήρησης μετακινούνται από Condition-based Maintenance σε Analytics και IoT-enabled PdM. Το Report μιλά για 39% άνοδο την περίοδο 2016-2022 με τον τζίρο να υπολογίζεται να φτάσει τα 11 δις δολάρια το 2022 [2]. Στο άρθρο του ο Sciban [1] παρουσιάζει 2 κύριους λόγους για την αύξηση της προβλεπτικής συντήρησης, ο πρώτος είναι η παρουσία Embedded chip και αισθητήρων σε όλους τους εξοπλισμούς δημιουργώντας έτσι μεγάλες ροές δεδομένων και ο δεύτερος η συνεχής μείωση του κόστους των embedded αυτών συσκευών.

Η προβλεπτική συντήρηση- Predictive Maintenance (PdM) έχει διάφορους ορισμούς. Στο άρθρο τους οι Bousdekis [4] δηλώνουν ότι σκοπός της προβλεπτικής συντήρησης είναι η μείωση του κόστους συντήρησης και η αύξηση της αξιοπιστίας του εξοπλισμού. Οι Amruthnath και Gupta [5] ότι ο σκοπός της προβλεπτικής συντήρησης είναι να μειωθούν οι απρόοπτες καθυστερήσεις οι οποίες έπονται μιας ξαφνικής αστοχίας βιομηχανικού εξοπλισμού με σκοπό να αυξηθεί η παραγωγικότητα και να μειωθεί το κόστος παραγωγής. Σε γενικούς όρους η προβλεπτική συντήρηση μπορεί να θεωρηθεί

σαν μια τεχνική η οποία παρακολουθεί την λειτουργική κατάσταση του εξοπλισμού και προσπαθεί να εγγυηθεί το μέγιστο MTTR (μέσος χρόνος μεταξύ επισκευών) αλλά και την ελαχιστοποίηση του κόστους που δημιουργούν οι ξαφνικές αστοχίες μηχανών οι οποίες μπορούν να προκύψουν σε έναν βιομηχανικό εξοπλισμό.

## 1.2 Αξιοπιστία Συστημάτων

Ο όρος «αξιοπιστία» [6] στη μηχανική και στη στατιστική προσδιορίζει την πιθανότητα μια συσκευή, ένα εξάρτημα ή ένα σύστημα να ικανοποιήσει τις προσδοκίες για εύρυθμη-ομαλή λειτουργία, χωρίς αστοχίες, για ένα συγκεκριμένο χρονικό διάστημα. Η στατιστική μελέτη της «Θεωρίας της Αξιοπιστίας» καθιστά δυνατή την εκτίμηση του χρόνου ζωής και λειτουργίας χωρίς αστοχίες, μερών του συστήματος ή ολόκληρου του συστήματος. Η αξιοπιστία ενός συστήματος το οποίο αποτελείται από πολλά μέρη-υποσυστήματα, εξαρτάται από την αξιοπιστία των επιμέρους μερών του. Δηλαδή για μια μηχανή η οποία αποτελείται από 10 μέρη (σύστημα ψύξης, σύστημα λίπανσης κλπ.), η αξιοπιστία της εξαρτάται από το καθένα ξεχωριστά, αν αστοχήσει το ένα, τότε αστοχεί όλο το σύστημα-μηχανή. Με  $F(t)$  είναι η συνάρτηση κατανομής της αστοχίας (Failure) του συστήματος, με  $R(t)$  συμβολίζεται η αξιοπιστία (Reliability) του συστήματος, ενώ με  $f(s)$  είναι η συνάρτηση πυκνότητας πιθανότητας για την αστοχία του συστήματος. Οπότε παίρνοντας την συνάρτηση (2) και ολοκληρώνοντας προς το χρόνο για κάποια χρονική στιγμή  $t$ , υπολογίζεται η πιθανότητα αστοχίας του συστήματος μέχρι την χρονική στιγμή  $t$ . Στην συνέχεια τοποθετώντας το αποτέλεσμα της (2) στην (1) εξάγεται σαν αποτέλεσμα η αξιοπιστία του συστήματος στον ίδιο χρόνο  $t$ .

$$R(t) = 1 - F(t) \quad (1)$$

$$F(t) = \int_0^t f(s)ds \quad (2)$$

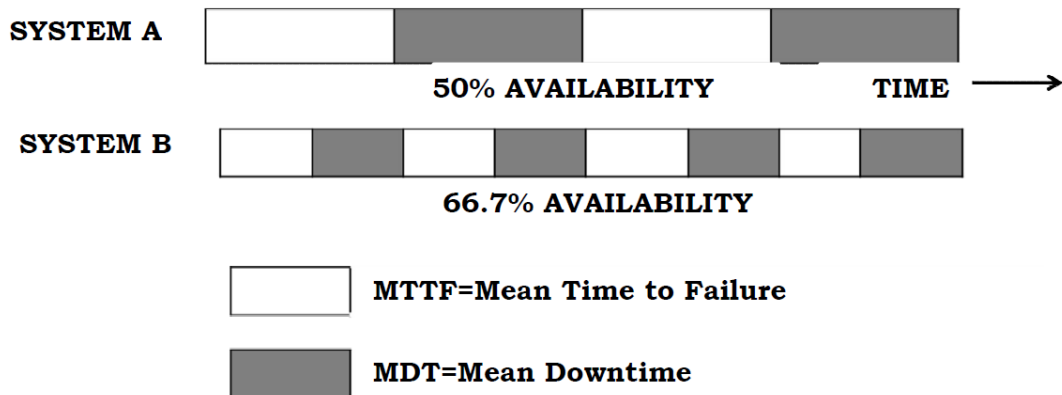
Μια από τις πιο σημαντικές παραμέτρους στην ανάλυση αξιοπιστίας είναι ο υπολογισμός του MTTF (Mean Time to Failure) ή μέσος χρόνος λειτουργίας μεταξύ βλαβών. Δηλαδή ο χρόνος ο οποίος μεσολαβεί μεταξύ δύο αστοχιών της μηχανής οι οποίες θα τη θέσουν εκτός λειτουργίας.

$$MTTF = \mu = \int_0^{+\infty} tf(t)dt \quad (3)$$

Ως εξάρτημα θεωρείται κάποιο μέλος του συστήματος όπου υπάρχει η δυνατότητα αντικατάστασης, σε περίπτωση βλάβης. Μια συσκευή αποτελείται από εξαρτήματα, ενώ ένα σύστημα είναι δυνατό να αποτελείται από πολλές συσκευές. Ο προσδιορισμός της αξιοπιστίας μπορεί να διαφέρει σε κάθε ένα από τα παραπάνω. Η διαθεσιμότητα (Availability) είναι ο πιο σημαντικός δείκτης στην επιχειρησιακή λειτουργία μιας μηχανής-συστήματος καθώς δείχνει το ποσοστό στο οποίο η μηχανή είναι διαθέσιμη για λειτουργία. Η διαθεσιμότητα είναι μια παράμετρος η οποία καθορίζεται από την αξιοπιστία και τη συντηρησιμότητα ενός συστήματος ή εξαρτήματος και προσδιορίζει τον διαθέσιμο χρόνο λειτουργίας του συστήματος. Με την λειτουργική διαθεσιμότητα ενός συστήματος εξοπλισμού να εκφράζεται ως:

$$A_o = \frac{MTBM}{MTBM + MDT} \quad (4)$$

Με MTBM είναι Mean time between Maintenance και MDT να είναι Mean Downtime. Ως MTBM θεωρείται ο καθαρός χρόνος λειτουργίας του εξοπλισμού, είναι ο χρόνος ο οποίος μεσολαβεί μεταξύ δύο διαδοχικών σταματημάτων για συντήρηση του εξοπλισμού. Ως MDT θεωρείται ο χρόνος ο οποίος απαιτείται για να γίνει συντήρηση της μηχανής. Δηλαδή η διαθεσιμότητα ενός εξοπλισμού είναι το πηλίκο του χρόνου λειτουργίας του εξοπλισμού (MTBM - ο χρόνος μεταξύ δύο επισκευών) προς το χρόνο λειτουργίας του εξοπλισμού συν το χρόνο που χρειάζεται για να γίνουν οι εργασίες συντήρησης. Στην εικόνα 1.1 φαίνεται πόσο σημαντικό ρόλο στη διαθεσιμότητα ενός συστήματος έχει το MDT, είναι πολύ σημαντικό η τιμή του να είναι όσο το δυνατόν μικρότερη καθώς αυξάνει την διαθεσιμότητα του. Βλέποντας το σύστημα A μπορεί να εξαχθεί το συμπέρασμα ότι ο χρόνος μεταξύ επισκευών είναι αρκετά μεγαλύτερος σε σχέση με αυτόν του συστήματος B, όμως το μεγάλο MDT του A μπορεί να θέτει τον εξοπλισμό εκτός λειτουργίας για αρκετό καιρό. Αυτός είναι και ο λόγος που τελικά προκύπτει το σύστημα B να έχει μεγαλύτερη διαθεσιμότητα από το σύστημα A, ενώ με την πρώτη ματιά φαίνεται το σύστημα A να συμπεριφέρεται καλύτερα.

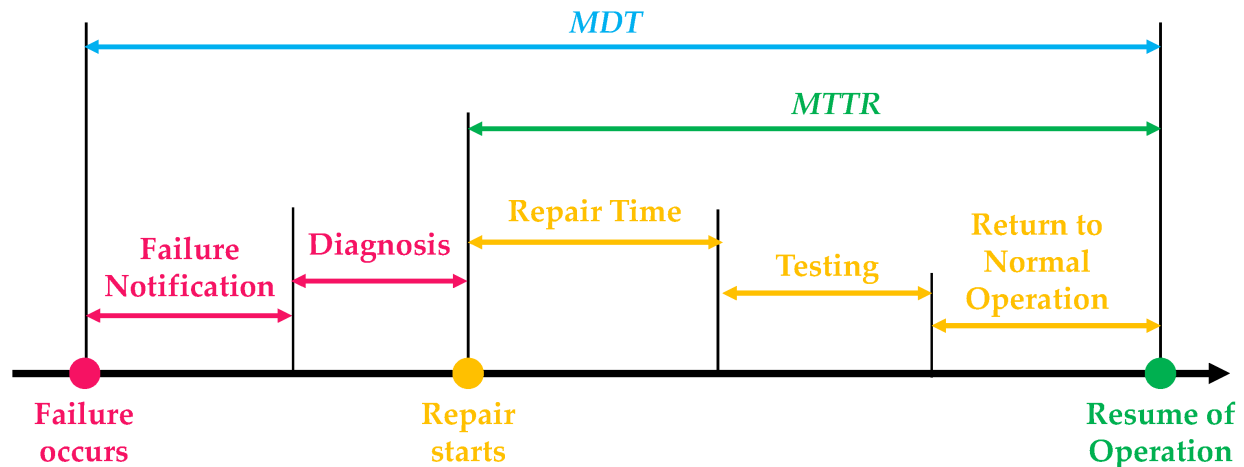


Εικόνα 1.1 Περιγραφή 2 συστημάτων και της κρίσιμότητας του MDT

(Σπυρίδων Δεσσαλερμος, Αξιοπιστία συστημάτων,(2015))

Οπότε η προσέγγιση της συντήρησης ενός εξοπλισμού έχει να κάνει με την μεγιστοποίηση της διαθεσιμότητας του. Άρα θα πρέπει να καθοριστούν πολύ προσεκτικά τα MTBM ώστε να είναι εκμεταλλεύσιμη στο μέγιστο η λειτουργική ζωή του εξοπλισμού, ενώ συγχρόνως η οργάνωση των ανταλλακτικών και της διαδικασίας συντήρησης θα πρέπει να είναι όσο το δυνατόν συντομότερη ώστε να τεθεί σε λειτουργία γρήγορα το σύστημα. Ως συντήρηση μπορούν να οριστούν οι προσπάθειες οι οποίες λαμβάνουν χώρα ώστε μια μηχανή- εξοπλισμός να διατηρήσει την κατάσταση και απόδοση της όσο πιο κοντά στις αρχικές προδιαγραφές. Η συντήρηση είναι ο βασικός κορμός-ραχοκοκαλιά μια βιομηχανίας και επιτρέπει στις παραγωγικές διαδικασίες να εκτελούνται στο προγραμματισμένο χρονικό πλαίσιο με όσο το δυνατόν μεγαλύτερη απόδοση.

Στην εικόνα 1.2 φαίνονται τα μέρη από τα οποία απαρτίζεται το MDT. Πρώτα υπάρχει ειδοποίηση αστοχίας (Failure Notification) ή ειδοποίηση για πρόβλεψη αστοχίας , ενώ ακολουθεί η διάγνωση(Diagnosis) δηλαδή οι τεχνικοί κατευθύνονται στο που οφείλεται αυτή η αστοχία π.χ. σε ποιο εξάρτημα. Στη συνέχεια και αφού έχει εντοπιστεί το πρόβλημα γίνεται εισαγωγή στο στάδιο του MTTR (Mean time to repair) το οποίο περιλαμβάνει στην σειρά τον χρόνο επισκευής (Repair Time), τη δοκιμή του εξοπλισμού ο οποίος επισκευάστηκε (Testing) και τη διαδικασία επιστροφής του εξοπλισμού στην επιχειρησιακή λειτουργία (Return to Normal Operation).



Εικόνα 1.2 Η δομή του MDT

(Reliability Analysis and Repair Activity for the Components of 350 kW Inverters in a Large Scale Grid-Connected Photovoltaic System ,(2021))

Για τις περισσότερες βιομηχανίες μια ξαφνική βλάβη-αστοχία του εξοπλισμού μπορεί να επιφέρει τεράστια κόστη , καθώς οι παράγοντες οι οποίοι συντελούν στην αποκατάσταση της βλάβης είναι πολυάριθμοι και στην εξίσωση της διαθεσιμότητας μοντελοποιούνται στο MDT της σχέσης (4):

1) Troubleshooting: Η διαδικασία εύρεσης της αστοχίας του συστήματος μπορεί να είναι μια χρονοβόρα διαδικασία καθώς η αιτία της αστοχίας μπορεί να είναι είτε ηλεκτρολογική, είτε μηχανολογική είτε από κάποιον τρίτο παράγοντα ο οποίος επηρέασε το ηλεκτρολογικό-μηχανολογικό κομμάτι. Με τη συνεχή καταγραφή δεδομένων των συστημάτων μέσω μετρήσεων, η διαδικασία του Troubleshooting εξελίχθηκε πολύ καθώς πολλές φορές όπως αναφέρθηκε και πιο πάνω, για κάποιο χρονικό διάστημα πριν την αστοχία μιας μηχανής μπορεί να καταγραφεί σε διάφορες μετρήσεις μεγεθών (δονήσεις, θερμοκρασίες κλπ.) διακύμανση σε σχέση με τις τιμές της ομαλής λειτουργίας.

2) Η συγκέντρωση εξειδικευμένου προσωπικού με σκοπό τη διερεύνηση και επίλυση του προβλήματος είναι και αυτό μια διαδικασία η οποία θέλει χρόνο και συντελεί στην παράταση της αποκατάστασης της βλάβης. Μια βλάβη μπορεί να προκύψει αργά το βράδυ ή σαββατοκύριακο αλλιώς μια βλάβη μπορεί να είναι τόσο εξειδικευμένη όπου να πρέπει η βιομηχανία να απευθυνθεί στον κατασκευαστή και να χρειαστεί να στείλει εξειδικευμένο προσωπικό ο ίδιος ο κατασκευαστής.

3) Τοποθεσία του εξοπλισμού: Πολλές φορές ένας εξοπλισμός μπορεί να βρίσκεται σε απομακρυσμένα μέρη είτε σε μέρη τα οποία οι συνθήκες δεν επιτρέπουν άμεση πρόσβαση. Για παράδειγμα η αστοχία μιας ανεμογεννήτριας η οποία βρίσκεται σε ένα Off-shore wind δίκτυο και η τοποθεσία της μπορεί να είναι στην μέση του ωκεανού ή στην κορυφή ενός βουνού. Άλλο παράδειγμα ένας βιομηχανικός αεριοστρόβιλος ο οποίος για λόγους προστασίας βρίσκεται εγκιβωτισμένος δεν μπορεί να προσεγγιστεί μετά από μια αστοχία του καθώς οι θερμοκρασίες στον χώρο είναι πολύ υψηλές και η πρόσβαση μπορεί να είναι επικίνδυνη καθώς όλες οι επιφάνειες είναι θερμές.

4) Πολλοί εξοπλισμοί είναι περίπλοκοι οπότε η αποσυναρμολόγησή τους μπορεί να χρειαστεί ώρες-ημέρες π.χ. για να αποσυνδεθούν μέρη εξοπλισμού τα οποία συνδέονται με μηχανικούς συνδέσμους, ηλώσεις, βίδες κλπ. θα πρέπει να βρίσκονται σε θερμοκρασία τέτοια ώστε να μην υπάρχουν διαστολές καθώς μπορεί να δημιουργηθούν ρωγμές και υψηλές μηχανικές τάσεις. Για παράδειγμα σε έναν βιομηχανικό εναλλάκτη θερμότητας ο οποίος προσδίδει θερμότητα μέσω ατμού σε φυσικό αέριο και ο ατμός οποίος διέρχεται τον εναλλάκτη μπορεί να είναι 300-400 °C , θα πρέπει να επέλθει ομαλή ψύξη του συστήματος ώστε να επέμβει τεχνικό προσωπικό και να αρχίσει την διαδικασία της επιδιόρθωσης της βλάβης, αποσυναρμολογώντας τον εξοπλισμό.

5) Στους βιομηχανικούς εξοπλισμούς επειδή τα ανταλλακτικά είναι ακριβά δεν υπάρχει άμεση διαθεσιμότητα τους, δηλαδή μια αποθήκη μιας βιομηχανικής μονάδας δεν έχει όλα τα ανταλλακτικά τα οποία είναι πιθανόν να χρειαστούν καθώς το κόστος αυτό θα ήταν τεράστιο. Οπότε σε περίπτωση αστοχίας εξοπλισμού και αφού έχουν προηγηθεί όλα τα παραπάνω βήματα μπορεί να προκύψει ότι δεν υπάρχει σε διαθεσιμότητα το ανταλλακτικό αυτή την στιγμή π.χ. μια βάνα , ένα μοτέρ κλπ. Χαρακτηριστικό παράδειγμα είναι τα πλοία λόγω και του περιορισμένου χώρου που έχουν για αποθήκευση. Οπότε η αποκατάσταση του εξοπλισμού θα έχει πολύ μεγάλη καθυστέρηση.

Μέχρι να λυθούν και να διευθετηθούν όλες οι παραπάνω αιτίες οι παραγωγή σταματάει και αυτό σχεδόν πάντα είναι αποτέλεσμα ξαφνικών αστοχιών του εξοπλισμού. Το σταμάτημα αυτό μπορεί να διαρκέσει ακόμα και μήνες ολόκληρους κάτι το οποίο μπορεί να σημάνει και οικονομική καταστροφή της εταιρείας π.χ. δημιουργία Arc-τόξου στη γεννήτρια θερμοηλεκτρικού σταθμού παραγωγής ρεύματος, καταστρέφοντας κομμάτι του εξοπλισμού. Σύμφωνα με το [7], το κόστος του downtime μπορεί να φτάσει τα \$540,000 για κάθε ώρα όπου εξοπλισμός, προσωπικό και σύστημα δεν είναι διαθέσιμα. Σύμφωνα με το International Society of Automation το κόστος των

αστοχιών σε βιομηχανικούς εξοπλισμούς παγκοσμίως ανέρχεται σε 864 δις δολάρια τον χρόνο [8].

### 1.3 Τύποι συντήρησης βιομηχανικού εξοπλισμού

Οι βιομηχανίες έχουν αναπτύξει 4 στρατηγικές προσεγγίσεις για την συντήρηση του εξοπλισμού στο πλαίσιο της αύξησης της διαθεσιμότητας του εξοπλισμού. Οι 3 από αυτές ανήκουν στην κατηγορία Proactive maintenance ενώ η τέταρτη η οποία δεν απασχολεί την παρούσα διπλωματική αλλά και δεν εφαρμόζεται στην βιομηχανία είναι η Reactive. Ως Proactive θεωρούνται οι προσεγγίσεις αυτές στις οποίες ο χρήστης δρα προληπτικά πάνω στην μηχανή πριν αυτή αστοχήσει, κάνοντας της τακτική συντήρηση, είτε προβλέποντας κάποια αστοχία λαμβάνοντας μέτρα για να μην αστοχήσει η μηχανή. Οι 3 Proactive προσεγγίσεις είναι η προληπτική συντήρηση (Preventive Maintenance) , η προβλεπτική συντήρηση (Predictive Maintenance) και η Condition based Maintenance. Η δεύτερη κατηγορία συντήρησης είναι η Reactive Maintenance στην οποία δεν γίνεται καμία προληπτική επέμβαση στον εξοπλισμό μέχρι αυτός να αστοχήσει(Run to Failure). Η Reactive maintenance δεν χρησιμοποιείται από καμία βιομηχανία πλέον.

#### 1.3.1 Τύποι Proactive Maintenance

**Preventive:** Σκοπός της προληπτικής συντήρησης- Preventive maintenance είναι η εγγύηση ότι ο εξοπλισμός θα είναι πάντα λειτουργικός και σε καλή κατάσταση. Μείωση του Downtime και αύξηση της διαθεσιμότητας και της απόδοσης του εξοπλισμού. Το πρόβλημα το οποίο γεννάται με την προληπτική συντήρηση είναι ότι μπορεί να είναι πολύ proactive μερικές φορές καθώς μπορεί να οδηγήσει προληπτικά στην αλλαγή μηχανικών μερών τα οποία είναι πλήρως λειτουργικά και δεν χρίζουν αντικατάστασης. Η συνεχής προληπτική αντικατάσταση μηχανικών μερών του ανεβάζει πολύ το κόστος ζωής σε όλη την υπηρεσία ενός βιομηχανικού εξοπλισμού. Πολλές φορές ένα ανταλλακτικό μπορεί να στοιχίζει χιλιάδες ευρώ, το οποίο όμως είναι προτιμότερο να αντικατασταθεί μετά το πέρας κάποιου χρόνου παρά να αστοχήσει απροειδοποίητα ο



εξοπλισμός. Έτσι γεννάται η ιδέα για μια νέα θεωρία προσέγγισης της συντήρησης εξοπλισμού η οποία θα αξιολογεί και θα προβλέπει τι χρίζει αντικατάστασης βάσει πιθανής αστοχίας ώστε να μειωθεί το κόστος.

**Condition based Maintenance:** Condition-based maintenance (CBM) [9] είναι μια στρατηγική maintenance η οποία παρακολουθεί σε πραγματικό χρόνο την κατάσταση του εξοπλισμού και δηλώνει τι χρίζει συντήρησης. Σε αντίθεση με την Preventive Maintenance η οποία είναι βασισμένη σε χρονικές περιόδους και δηλώνει πότε πρέπει να γίνει συντήρηση του εξοπλισμού, η condition-based maintenance έχει ως βάση της ότι η συντήρηση του εξοπλισμού πρέπει να γίνει όταν οι Real-time indicators δείξουν ανακολουθίες και σημάδια πτωτικής απόδοσης. Το CBM υλοποιεί real-time diagnosis με πρόγνωση για την απόδοση του συστήματος και έτσι η μη ακολούθηση της πραγματικής απόδοσης του συστήματος από την εκτιμώμενη του μοντέλου αποτελεί δείγμα πως το σύστημα χρειάζεται συντήρηση. Στην Condition Based ένα μαθηματικό μοντέλο της πραγματικής μηχανής προσομοιώνει σε ζωντανό χρόνο την λειτουργία της, κάνοντας την χρονική στιγμή  $t$  προβλέψεις τιμών για διάφορα μεγέθη της μηχανής την χρονική στιγμή  $t+1$ . Αν η πρόβλεψη του μοντέλου έχει διαφορά μεγαλύτερη από ένα προκαθορισμένο όριο-Threshold, τότε ενεργοποιείται alarm στο σύστημα. Παράδειγμα της διαφοράς Condition based με την Preventive για την αλλαγή λαδιών π.χ. σε μια μηχανή εργοστασίου ή σε ένα στόλο αυτοκινήτων με την Preventive προσέγγιση, αλλαγή για την μηχανή θα γίνει μετά από συγκεκριμένες ώρες λειτουργίας και για το αυτοκίνητο μετά από συγκεκριμένο αριθμό χιλιομέτρων π.χ. 10.000χλμ. Με την Condition based προσέγγιση συνεχώς γίνεται λήψη σήματος για την κατάσταση των σωματιδίων στο λάδι από αισθητήρα και βάσει αυτού γίνεται η αξιολόγηση πότε χρίζει αλλαγής με βάση την παραμετροποιημένη εξίσωση η οποία περιγράφει το σύστημα.

**Predictive maintenance:** Το Predictive maintenance εκμεταλλεύεται τα θετικά και των 2 προηγούμενων προσεγγίσεων (Preventive – Condition Based). Σε αυτή την προσέγγιση με την χρήση μεθόδων μηχανικής μάθησης, γίνεται εκπαίδευση των μοντέλων με δεδομένα τα οποία έχουν αποθηκευτεί από την παρακολούθηση-monitoring της μηχανής και γίνεται πρόβλεψη είτε για μελλοντική αστοχία της μηχανής(classification) είτε για τον εναπομένοντα χρόνο ζωής της -RUL- (Regression). Το μεγάλο πλεονέκτημα της Predictive maintenance ή προβλεπτικής συντήρησης είναι ότι εκ των προτέρων ακριβώς επειδή γίνεται πρόβλεψη της αστοχίας μπορεί να γίνει και προγραμματισμός

της συντήρησης του εξοπλισμού, με όλα τα θετικά τα οποία έχει ο έγκαιρος προγραμματισμός από πλευράς κόστους, όπως αναλύθηκε και πιο πάνω. Η διαφορά της Predictive Maintenance σε σχέση με την Condition Based είναι ότι η Condition based δεν προβλέπει πότε θα αστοχήσει ο εξοπλισμός, οπότε να τεθεί εκ των προτέρων ένα πρόγραμμα συντήρησής του, αλλά ειδοποιεί όταν ξεπεραστούν συγκεκριμένα Thresholds τα οποία έχουν τεθεί.

#### 1.4 Προγνωστική και Διαγνωστική Μηχανών

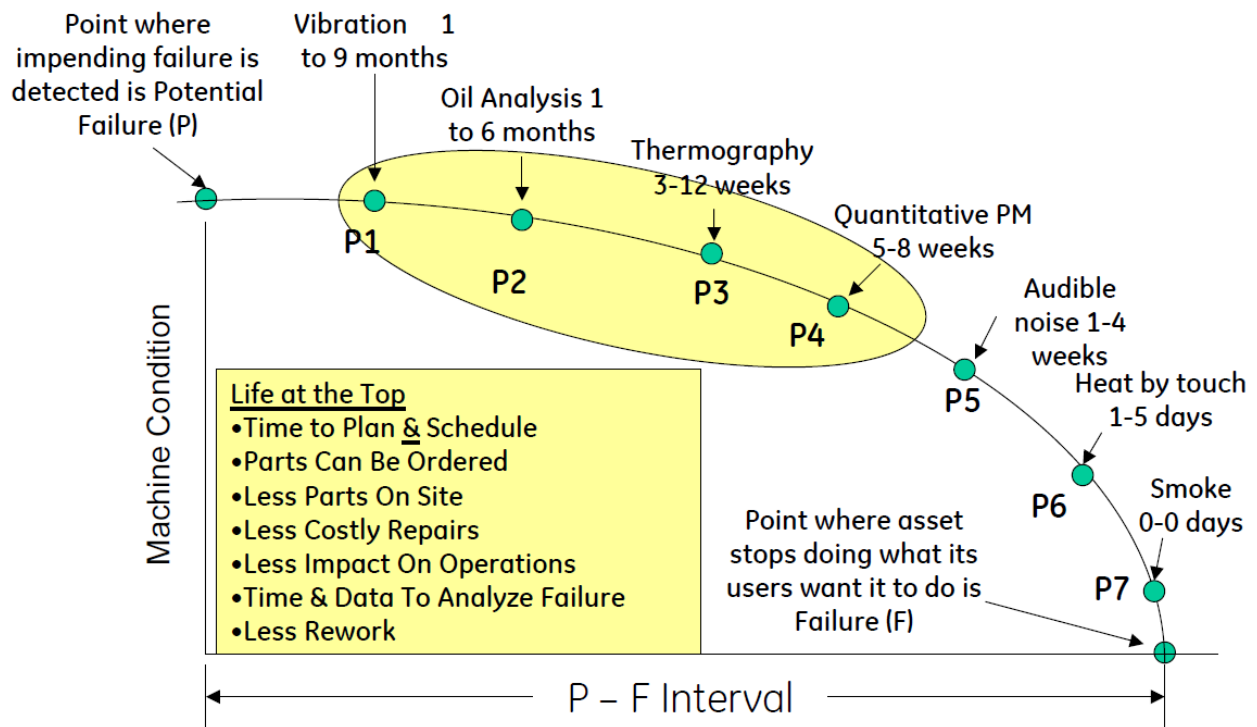
Η Προγνωστική των μηχανών εστιάζει στην πρόβλεψη των μελλοντικών καταστάσεων του συστήματος χρησιμοποιώντας προσομοιώσεις και μαθηματικές μοντελοποιήσεις του συστήματος. Συγκεκριμένα αναφέρεται στην έρευνα της πρόβλεψης πότε ένα σύστημα δεν θα είναι διαθέσιμο να αποδώσει βάσει των προδιαγραφών του, δηλαδή πότε ένα σύστημα θα αστοχήσει. Η προγνωστική είναι η προσπάθεια της εκτίμησης του εναπομένου χρόνου ζωής ενός συστήματος-Remaining Useful Life(RUL), μέχρι αυτό να αστοχήσει. Σε όλες τις βαριές βιομηχανίες ο περιστρεφόμενος εξοπλισμός είναι ευάλωτος σε αστοχίες λόγω των υψηλών δυναμικών και θερμικών καταπονήσεων που δέχεται για πολλές ώρες λειτουργίας. Η καταπόνηση αυτή μπορεί να μετρηθεί σε ώρες λειτουργίας ή σε κύκλους κόπωσης[11]. Για έναν κινητήρα αεροσκάφους ένας κύκλος κόπωσης είναι μια πλήρης πτήση, δηλαδή η έναυση της μηχανής, η πτήση του αεροσκάφους στο αεροδρόμιο προορισμού και το σταμάτημα της μηχανής. Παραδείγματα τέτοιων μηχανών είναι τα κιβώτια ταχυτήτων (π.χ. το κιβώτιο ταχυτήτων μιας ανεμογεννήτριας το οποίο μεταφέρει την κίνηση από τον άξονα των πτερυγίων που περιστρέφει ο αέρας, στην γεννήτρια) [12], τα έδρανα ολίσθησης (μέσα στα οποία περιστρέφονται και λιπαίνονται οι άξονες περιστροφής) [13], οι τουρμπίνες (Ατμοστρόβιλοι, Αεριοστρόβιλοι) [14]. Για την αποφυγή κρίσιμων και υψηλού κόστους αστοχιών στον περιστρεφόμενο εξοπλισμό οι αστοχίες πρέπει να διαγνωστούν όσο το δυνατόν γρηγορότερα.

Για την αποφυγή τέτοιων σεναρίων η συντήρηση του εξοπλισμού προγραμματίζεται νωρίτερα βάσει preventive maintenance. Όμως σε κάποιες βιομηχανίες το κόστος της συντήρησης του εξοπλισμού μπορεί να φτάσει μέχρι και 70% του ολικού κόστους[7]. Για τον λόγο αυτό η μείωση του κόστους είναι ίσως ο πιο κρίσιμος παράγοντας επιβίωσης για τις βαριές βιομηχανίες με δαπανηρό εξοπλισμό όπως π.χ. Θερμοηλεκτρικοί σταθμοί

ισχύος, διυλιστήρια , μεταλλουργίες, αεροπορική βιομηχανία. Στο διάβα του χρόνου η αύξηση της υπολογιστικής ισχύος και της ροής δεδομένων έφερε μια βιομηχανική επανάσταση και δόθηκε έμφαση στην αξιολόγηση της αξιοπιστίας ενός συστήματος βάσει των μετρήσεων των αισθητήρων. Η μέθοδος αυτή ονομάστηκε Condition Based Method (CBM)[9]. Το CBM υλοποιεί real-time diagnosis με πρόγνωση για την απόδοση του συστήματος και έτσι η μη ακολουθήση ή απόκλιση της πραγματικής απόδοσης του συστήματος από την εκτιμώμενη του μοντέλου αποτελεί ένδειξη πως το σύστημα χρειάζεται συντήρηση. Στην συνέχεια με την εξέλιξη του AI και την είσοδο του IOT στην βιομηχανία σαν IIOT(Industrial Internet of Things) αλλά και του BIG DATA η προγνωστική αναπτύχθηκε γρήγορα στο κομμάτι της μηχανικής μάθησης όπου ο εκπαιδευμένος αλγόριθμος εκτιμά αν η μηχανή πρόκειται να αστοχήσει ή πόσος χρόνος υπάρχει μέχρι την αστοχία της. Πολλοί είναι οι παράγοντες οι οποίοι συντελούν στην απόφαση για την συντήρηση μιας μηχανής. Το Προφίλ πτώσης της απόδοσης της μηχανής, το πρόγραμμα συντήρησης (πότε έκανε τελευταία φορά συντήρηση),το κόστος της συντήρησης, οι ενεργές ειδοποιήσεις που βρίσκονται στο σύστημα ελέγχου της μηχανής (Faults-diagnostics alarms).

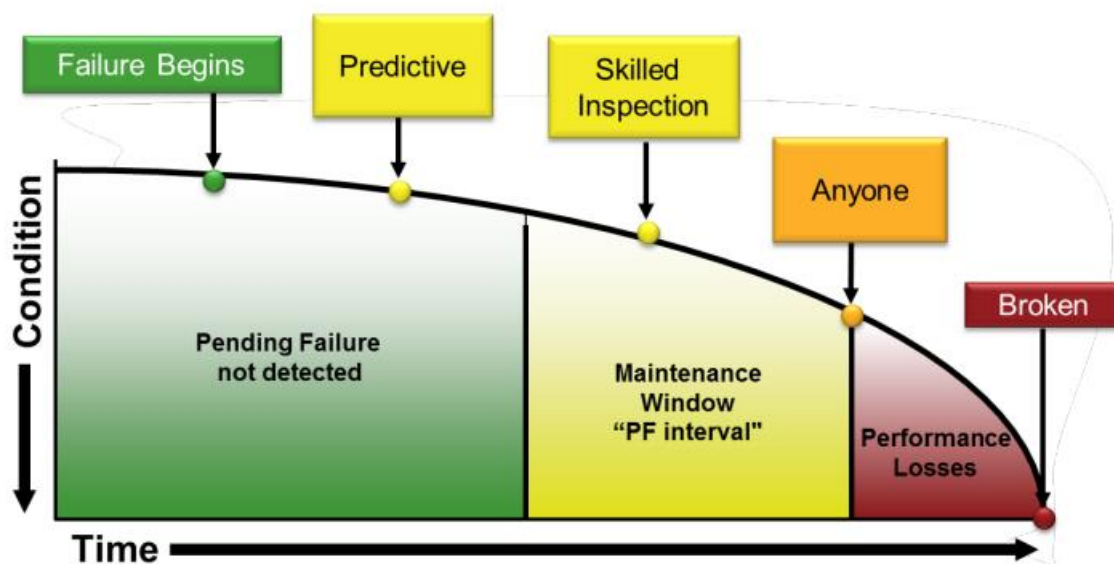
Στην παρακάτω εικόνες 1.3, 1.4 παριστάνεται η ίδια καμπύλη από διαφορετική οπτική και φαίνονται χαρακτηριστικά παραδείγματα συμπτωμάτων και ενδείξεων μιας μηχανής η οποία οδεύει προς την αστοχία , αλλά και οι δυνατότητες-μέθοδοι εύρεσης της αστοχίας για την κάθε χρονική στιγμή. Στην εικόνα 1.3 φαίνονται όλα εκείνα τα μεγέθη και οι καταστάσεις τα οποία προμηνύουν ότι έπεται μελλοντική αστοχία. Όπως φαίνεται και στην εικόνα η εξέλιξη των συμπτωμάτων τα οποία εκδηλώνονται στην μηχανή σε πρώιμο στάδιο μπορούν να ληφθούν μόνο με ειδικές καταγραφές (π.χ. αισθητήρας δονήσεων), ενώ όσο περνάει ο χρόνος γίνονται πιο έντονα και πιο αισθητά στον άνθρωπο (υψηλή θερμοκρασία μετάλλου, καπνός, θόρυβος). Τέτοιες ενδείξεις είναι οι υψηλές ταλαντώσεις , η ανάλυση του λαδιού λίπανσης του συστήματος η οποία μπορεί να δείξει την φθορά διαφόρων μερών του εξοπλισμού καθώς υπολείμματα της φθοράς θα βρίσκονται στην σύσταση του λαδιού. Επίσης στο λάδι λίπανσης μπορεί να φανεί μια διαρροή που έχει γίνει στο σύστημα (π.χ. πρόσμιξη με κύκλωμα ψύξης) ή επαφή με τον εξωτερικό αέρα δημιουργώντας οξειδωση και υποδηλώνοντας πως σε κάποιο σημείο το σύστημα είναι εκτεθειμένο. Άλλη χαρακτηριστική ένδειξη αποτελεί η θερμογραφία η οποία εκτελείται με θερμοκάμερα και στο σημείο της μελλοντικής αστοχίας (π.χ. έδρανο αντλίας) όπως στην εικόνα 1.5 παρατηρείται μεγαλύτερη θερμοκρασία στο εξωτερικό περίβλημα από ότι συνήθως. Στην συνέχεια και όσο περνάει ο καιρός το RUL μειώνεται και η μηχανή πλησιάζει προς την αστοχία, τα σημάδια γίνονται

όλο και πιο έντονα με θόρυβο της μηχανής αλλά και με αύξηση της εξωτερικής θερμοκρασίας του περιβάλλοντος της μηχανής η οποία πλέον είναι αισθητή και με το χέρι. Σκοπός του Predictive maintenance είναι η εκμετάλλευση όλων των παραπάνω ενδείξεων μέσω καταγραφών από εξειδικευμένους αισθητήρες ώστε η ανίχνευση και η συσχέτιση όλων των παραπάνω ενδείξεων σε πρώιμα στάδια να υποδηλώσουν στον χρήστη ότι έπεται αστοχία. Για παράδειγμα δονήσεις ελαφρώς μεγαλύτερες σε έναν εξοπλισμό μπορεί να είναι κάτι απολύτως φυσιολογικό λόγω αλλαγής θερμοκρασίας περιβάλλοντος ή λόγω φυσιολογικής φθοράς κάποιου μηχανισμού ο οποίος όμως έχει ακόμα αρκετή ωφέλιμη ζωή. Σκοπός του αλγορίθμου είναι να βρει το pattern π.χ. Δονήσεις στον Υ άξονα και ελαφριά αύξηση της θερμοκρασίας στο λάδι λίπανσης για τ χρόνο αυτό σημαίνει ότι έπεται αστοχία. Αν υπάρχει αρκετή πληροφορία στην βάση δεδομένων τότε ο αλγόριθμος μπορεί να κάνει και classification σχετικά με το ποιο μέρος του εξοπλισμού πρόκειται να αστοχήσει. Στο [15] αναφέρεται λεπτομερώς η διαδικασία εκπαίδευσης ενός τέτοιου αλγορίθμου για ανεμογεννήτριες και το σύστημα έχει χωριστεί σε γεννήτρια, μετασχηματιστή, κιβώτιο ταχυτήτων, έδρανα γεννήτριας και υδραυλικό σύστημα. Οπότε ο αλγόριθμος υποδηλώνει και σε ποιο σύστημα ο χρήστης θα πρέπει να επικεντρώσει την προσοχή του για να βρει την μελλοντική αστοχία.



Εικόνα 1.3 Η εξέλιξη των σταδίων μιας αστοχίας

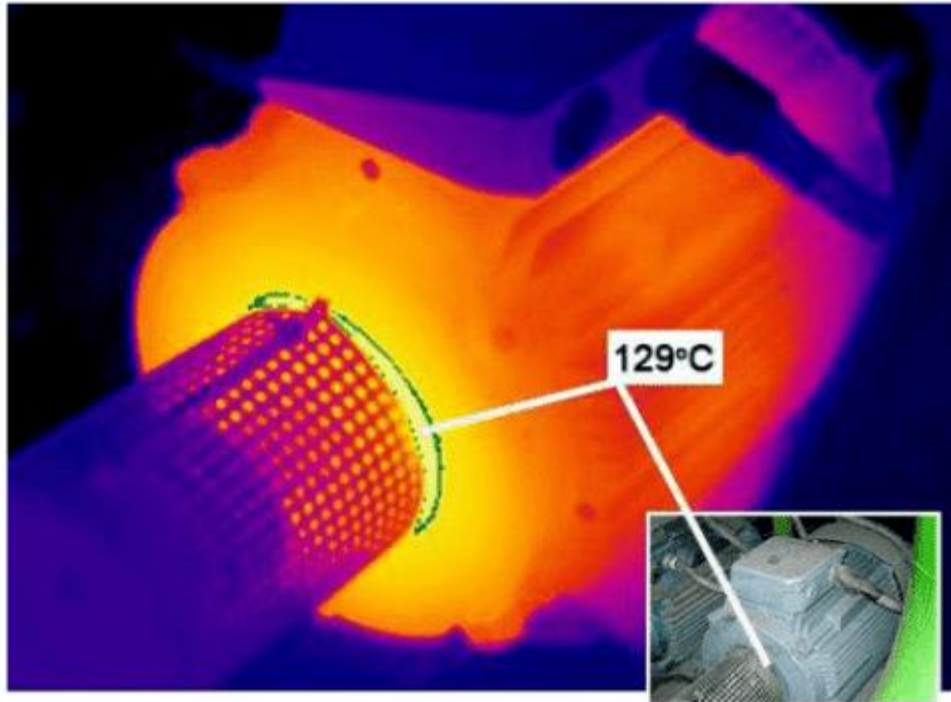
Στην Εικόνα 1. 4 φαίνεται ένα γράφημα για την εξέλιξη μιας αστοχίας στο χρόνο και όλα τα στάδια εκείνα στα οποία με κατάλληλα μέσα ο χρήστης μπορεί να την εντοπίσει πριν εκδηλωθεί και θέσει εκτός λειτουργίας τον εξοπλισμό. Η αστοχία εκδηλώνεται στο τελευταίο χρονικό σημείο του διαγράμματος (Broken). Αρχικά όπως φαίνεται και στην εικόνα η μελλοντική αστοχία (Failure Begins) δεν μπορεί να γίνει αντιληπτή από κανέναν ενώ λίγο χρόνο μετά (Predictive) μπορεί να γίνει από μεθόδους προβλεπτικής συντήρησης (Predictive Maintenance). Στη συνέχεια και ενώ η κατάσταση επιδεινώνεται, φαίνεται στην εικόνα ότι ένας εξειδικευμένος τεχνικός θα μπορούσε να αντιληφθεί την δυνητική αστοχία (Skilled Inspection), π.χ. θα μέτραγε υψηλές ταλαντώσεις ή θα κατέγραφε διαφορετικό θόρυβο στην μηχανή από ότι συνήθως. Τέλος φαίνεται ότι ένα βήμα πριν την αστοχία (Anyone) ο καθένας μπορεί να αντιληφθεί ότι υπάρχει πρόβλημα και η μηχανή οδεύει προς την αστοχία. Σημάδια θα μπορούσαν να ήταν π.χ. πολύ έντονος θόρυβος, η παρουσία καπνού. Η διάτμηση των χρονικών περιόδων του διαγράμματος είναι γενικής αναφοράς και δεν αναπαριστά απόλυτα την αναλογία με την πραγματικότητα. π.χ. ο χρόνος μεταξύ του σταδίου Anyone και Broken μπορεί να είναι πολύ μικρός.



Εικόνα 1.4 Μέθοδοι ανίχνευσης της αστοχίας στην εξέλιξη της στο χρόνο.

(Condition Monitoring & Predictive Maintenance: Machine Failure Indicators & Detection Methods(2021))

Στην εικόνα 1.5 φαίνεται μια ακόμη εφαρμογή της μηχανικής μάθησης στο κομμάτι της βιομηχανίας. Αναλυτικότερα στην εικόνα 1.5 φαίνεται η πιο διαδεδομένη μηχανή στην ιστορία της ανθρωπότητας, ο ηλεκτρικός κινητήρας ή μοτέρ ο οποίος κινεί οτιδήποτε βλέπουμε γύρω μας, από το νερό στο δίκτυο των πόλεων μέχρι τις γραμμές παραγωγής των εργοστασίων. Η θερμογραφία είναι μια βασική μέθοδος προληπτικής συντήρησης η οποία μπορεί να υποδείξει αν κάποιο εξάρτημα μιας μηχανής δυσλειτουργεί οπότε και πρόκειται να αστοχήσει στο άμεσο μέλλον. Με τακτική θερμογράφιση ενός εξοπλισμού π.χ. 1 φορά τον μήνα, διαμορφώνεται ένα θερμοκρασιακό προφίλ του εξοπλισμού αυτού σε κανονικές συνθήκες ομαλής λειτουργίας. Οπότε όταν η θερμοκρασία σε κάποιο από τα σημεία του εξοπλισμού υπερβεί τη θερμοκρασία ομαλής λειτουργίας της μηχανής-κινητήρα όπως φαίνεται στην εικόνα 1.5, τότε αυτό υποδηλώνει μελλοντική αστοχία. Η θερμογραφία σήμερα χρησιμοποιείται σαν ένα βασικό εργαλείο για την αξιολόγηση διακοπών υψηλής και μέσης τάσης όπου το ρεύμα το οποίο άγεται είναι υψηλό. Μια εργασία σταθμός πάνω στο θέμα είναι η [16] όπου και υλοποιείται εφαρμογή ενός MLP σε εξοπλισμούς σταθμών υψηλής και μέσης τάσης με εντυπωσιακά αποτελέσματα. Μέσω της εφαρμογής βαθιάς μάθησης γίνεται ανάλυση της εικόνας σχετικά με το αν έχει αυξηθεί η θερμοκρασία σε μέρη του εξοπλισμού που επιτηρείται. Αυτό το κομμάτι του Vision πρόκειται να έχει μεγάλη ανάπτυξη στο μέλλον και να επιτηρεί ολόκληρες ομάδες μηχανικών διατάξεων και συστημάτων. Επίσης η συνεργασία διαφορετικών ειδών δεδομένων από καταγραφές θα ήταν πολύ χρήσιμη για την αξιολόγηση μιας μελλοντικής αστοχίας. Δηλαδή ο συνδυασμός της εικόνας με δεδομένα τα οποία καταγράφονται από αισθητήρες π.χ. ο συνδυασμός της θερμογραφίας, με την αύξηση των ταλαντώσεων. Όλα αυτά είναι οι μεγάλες προκλήσεις του μέλλοντος όσον αφορά την βιομηχανία.



Εικόνα 1.5 Θερμογραφία σε έδρανο αντλίας

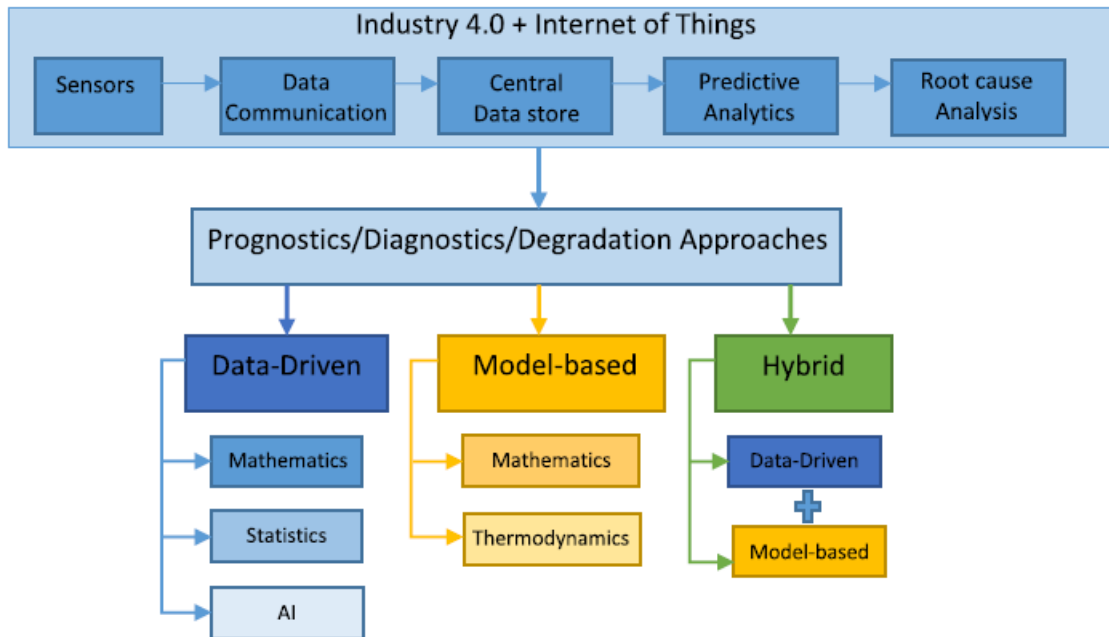
(Predictive Maintenance of Power Substation Equipment by Infrared Thermography Using a Machine-Learning Approach(2017))

Στην εικόνα 1.6 φαίνονται οι τρόποι προσέγγισης και μοντελοποίησης της διάγνωσης και πρόγνωσης σε ένα μηχανικό σύστημα. Μετρήσεις παίρνονται από αισθητήρες όπου μέσω πρωτοκόλλων επικοινωνίας (π.χ. Modbus) τα δεδομένα τους αποθηκεύονται σε έναν Server (π.χ. SQL Server). Από την στιγμή που τα δεδομένα βρίσκονται αποθηκευμένα στον Server αυτό που απομένει είναι η επιλογή των κατάλληλων μαθηματικών εργαλείων για την εκμετάλλευσή τους, με σκοπό την έγκαιρη πρόβλεψη (Predictive Analytics) της αστοχίας. Οι τρόποι-κατηγορίες είναι οι :

**Data Driven** όπου η επεξεργασία των δεδομένων γίνεται με στατιστικές μεθόδους (π.χ. διασπορές μεγεθών πάνω από ένα όριο) και με μεθόδους μηχανικής μάθησης.

**Model Based** όπου υπάρχει τυποποίηση του μοντέλου με κάποια εξίσωση η οποία αφορά την απόκριση της μηχανής ή άλλες παραμέτρους της και όταν το μοντέλο αποκλίνει από τις μετρήσεις ειδοποιείται ο χρήστης.

**Hybrid:** Όπου είναι ο συνδυασμός των παραπάνω 2 μεθόδων



Εικόνα 1.6 Μέθοδοι ανάλυσης δεδομένων

(A data-driven approach for predicting long-term degradation of a fleet of micro gas turbines (2021))

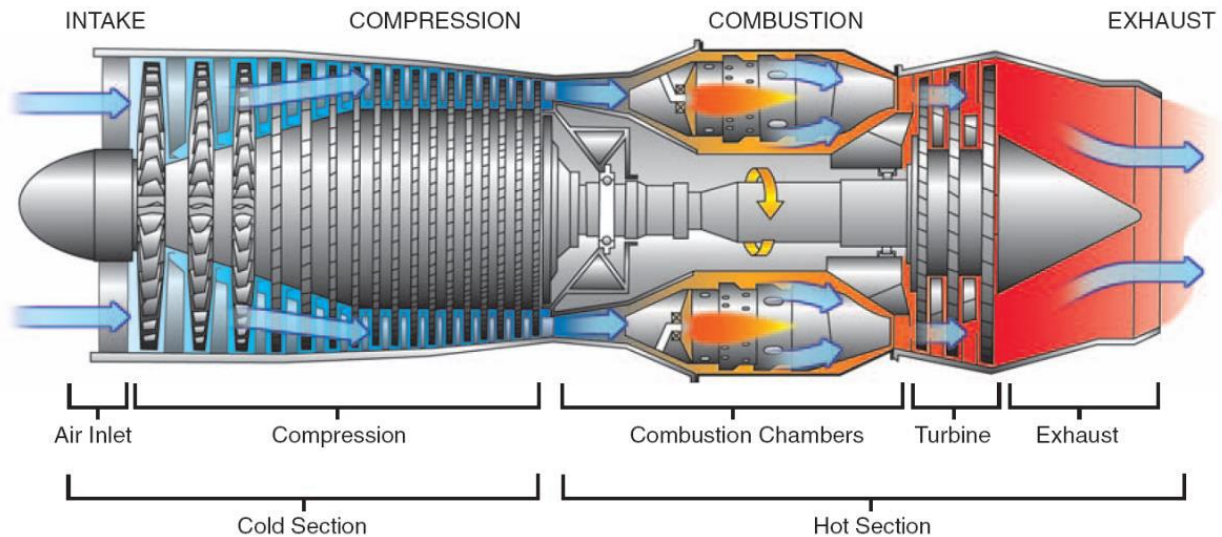
Η παρούσα διπλωματική επικεντρώνεται στο Data Driven κομμάτι και πως με μεθόδους μηχανικής μάθησης μπορούν όλα αυτά τα αποθηκευμένα δεδομένα και οι καταγραφές μιας μηχανής να αξιοποιηθούν για την πρόβλεψη μιας αστοχίας. Τα τελευταία χρόνια πολύ μεγάλη πρόοδος έχει επιτευχθεί στην παρακολούθηση της ομαλούς λειτουργίας και της απόδοσης λειτουργίας βιομηχανικών εξοπλισμών και ειδικά των αεριοστρόβιλων (Gas Turbine health and Performance monitoring). Υπάρχουν 3 κύριες χρήσεις της Μηχανικής Μάθησης στην παρακολούθηση-monitoring βιομηχανικών εξοπλισμών .

- 1) Στο πλαίσιο του Regression ένα μοντέλο μηχανής (hypothesis) έχει δημιουργηθεί κατά την εκπαίδευση του αλγορίθμου, όπου παίρνοντας τιμές από διάφορους αισθητήρες στην είσοδο προβλέπει την έξοδο .
- 2) Στο πλαίσιο του Classification όπου παίρνοντας μετρήσεις από διάφορους αισθητήρες ο αλγόριθμος προβλέπει την κατηγορία και τον τύπο του fault. π.χ. προβλέπει αστοχία στο έδρανο νούμερο 3 ή στον συμπιεστή κλπ. ανάλογα με το πως έχουν χωριστεί και οι πιθανές ομάδες αστοχίας κατά την εκπαίδευση.



- 3) Στο πλαίσιο του Regression πρόγνωση και εκτίμηση του εναπομένοντος χρόνου (RUL) ομαλής λειτουργίας της μηχανής μέχρι να εμφανιστεί βλάβη.

## 2.1 Ο Αεριοστρόβιλος



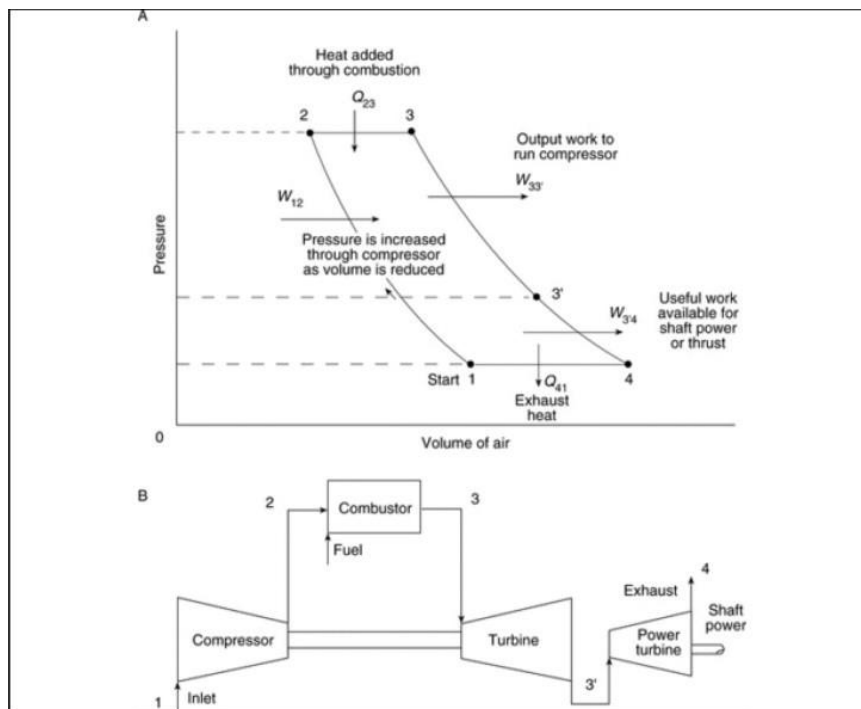
Εικόνα 2.1 Η αρχιτεκτονική ενός Turbofan κινητήρα

(General Electric Training Barcelona(Combined Cycle Power Powerplant Training (2011))

Όπως φαίνεται και στην Εικόνα 2.1, ένας αεριοστρόβιλος αποτελείται από τα εξής μέρη: τον συμπιεστή αέρα, τους θαλάμους καύσης και τον στρόβιλο. Ο συμπιεστής και ο στρόβιλος βρίσκονται σε έναν κοινό άξονα. Σε περίπτωση που ο αεριοστρόβιλος χρησιμοποιείται για παραγωγή ενέργειας τότε ο άξονας είναι σε σειρά με τον άξονα μιας επαγωγικής γεννήτριας και χρησιμοποιείται είτε σε «απλό κύκλο» (simple-cycle) είτε σε «συνδυασμένο κύκλο» (combined-cycle). Ένας αεριοστρόβιλος, που ονομάζεται επίσης στρόβιλος καύσης, είναι ένας τύπος κινητήρα συνεχούς και εσωτερικής καύσης.

Ένας αεριοστρόβιλος λειτουργεί βάσει του θερμοδυναμικού κύκλου Brayton όπως αυτός απεικονίζεται στην Εικόνα 2.1. Τα στάδια του κύκλου είναι:

Ο ατμοσφαιρικός αέρας εισέρχεται στον αεριοστρόβιλο και ρέει μέσω των σταδίων του συμπιεστή οπότε και ο αέρας συμπιέζεται φτάνοντας σε υψηλή πίεση(1→2). Στην συνέχεια και μετά την έξοδο από τον συμπιεστή ο αέρας εισέρχεται στους θαλάμους καύσης ,όπου προστίθεται καύσιμο με την μορφή ψεκασμού και γίνεται ανάφλεξη του μείγματος(2→3).Στην συνέχεια το συμπιεσμένο αέριο υψηλής θερμοκρασίας εισέρχεται σε έναν στρόβιλο, παράγοντας ένα έργο άξονα στη διαδικασία, που χρησιμοποιείται για την κίνηση του συμπιεστή(3→4). Η εναπομένουσα ενέργεια οδηγείται στην έξοδο σε μορφή καυσαερίων τα οποία μπορούν να επαναχρησιμοποιηθούν για εξωτερική εργασία, όπως η δημιουργία ώθησης σε έναν στροβιλοκινητήρα αεροσκάφους ή η περιστροφή ενός δεύτερου, ανεξάρτητου στροβίλου (τουρμπίνα ισχύος) ο οποίος μπορεί να συνδεθεί σε ανεμιστήρα, έλικα ή ηλεκτρική γεννήτρια(4→1).

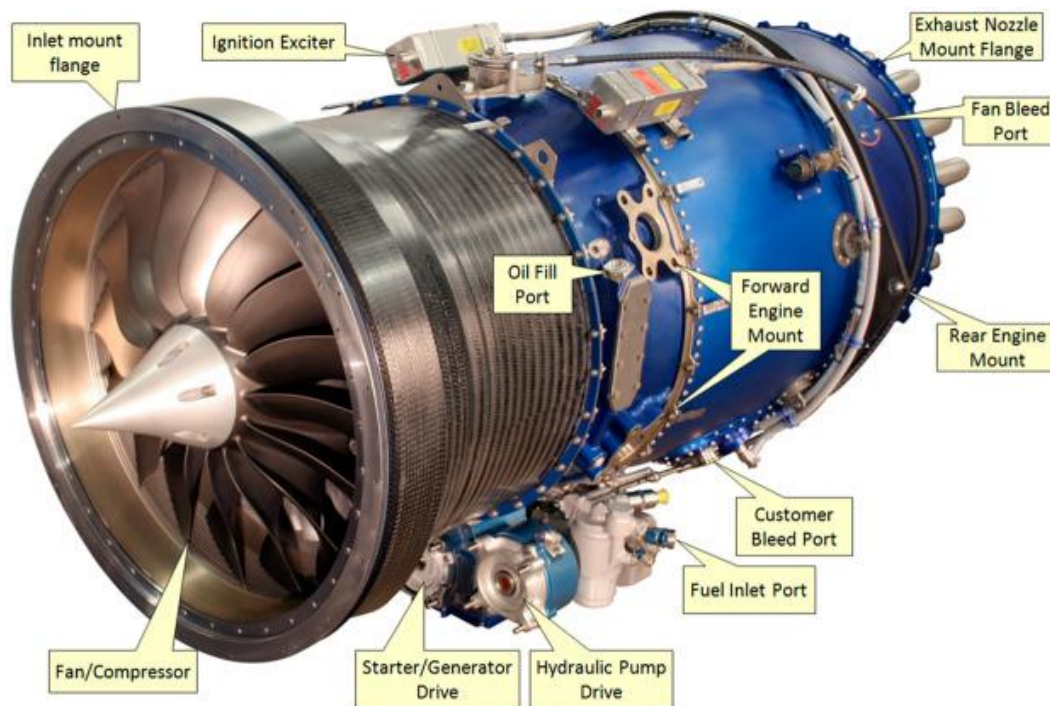


Εικόνα 2.2 Ο κύκλος του Brayton

(General Electric Training Barcelona(Combined Cycle Power Powerplant Training (2011))

Στην Εικόνα 2.3 φαίνεται ένας πραγματικός αεριοστρόβιλος με ένα κομμάτι του βοηθητικού του εξοπλισμού. Απεικονίζονται ο Ignition Exciter (μπουζί) δημιουργεί το Arc-Ηλεκτρικό τόξο για την έναυση του μείγματος αέρα-καυσίμου. Starter-Generator Drive

είναι το ανάλογο του δυναμό του αυτοκινήτου και παίρνοντας κίνηση από τον κύριο άξονα της μηχανής παράγει ρεύμα για όλες τις καταναλώσεις του αεροσκάφους. Hydraulic pump drive η υδραυλική αντλία παίρνει κίνηση από τον κύριο άξονα της μηχανής και συμπιέζει το υδραυλικό λάδι για κίνηση των επιφανειών των φτερών κλπ. Πλέον για όλα τα παραπάνω υποσυστήματα τα οποία αναφέρθηκαν και τα οποία αποτελούν ένα μικρό ποσοστό των συστημάτων του αεροσκάφους, υπάρχουν καταγραφές της απόδοσης τους και των μεγεθών τους, οπότε με τις μεθόδους του predictive maintenance γίνεται να προβλέψουμε όχι μόνο την αστοχία του κινητήρα αλλά και ποιο υποσύστημα θα αστοχήσει.



Εικόνα 2.3 Τα βοηθητικά μέρη ενός Turbofan κινητήρα

(Turbofan Engines - an overview | ScienceDirect Topics | Turbofan engine)

Για του κινητήρες αεροσκαφών και την σημαντικότητα της πρόβλεψης της αστοχίας τους, από τα στοιχεία τα οποία δίνει η Boeing [17] προκύπτουν τα εξής

- Το 20% με 30% το αστοχιών μηχανής αεροσκάφους την ώρα που βρίσκεται σε λειτουργία κοστίζει US\$500,000.
- Το 50% των καθυστερημένων πτήσεων λόγω μηχανικής βλάβης κοστίζουν US\$9,000 την ώρα.
- Το 50% των ακυρωμένων πτήσεων λόγω μηχανικής βλάβης κοστίζουν US\$66,000 για κάθε ακύρωση.

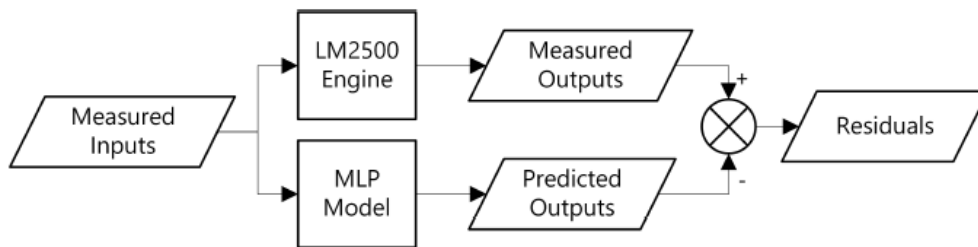
## 2.2 Προσέγγιση του Predictive Maintenance

Παρακάτω συνοψίζονται οι σημαντικότερες ως τώρα εργασίες στην βιβλιογραφία στο πλαίσιο του Predictive maintenance. Δίνεται έμφαση στους αεροστροβίλους (Turbofan) κινητήρες οι οποίοι αποτελούν και το αντικείμενο της παρούσας διπλωματικής .

Στο έργο τους οι Vimala Mathew, Mark Simpson[18] καταπιάνονται με ένα Dataset της NASA για κινητήρες αεροσκαφών turbofan. Στο Dataset λαμβάνεται ένα σετ μετρήσεων από 21 διαφορετικούς αισθητήρες οι οποίες αποθηκεύονται σε μια βάση δεδομένων για κάθε πτήση όταν μπαίνει σε λειτουργία ο κινητήρας. Το Dataset είναι Run to Failure δηλαδή για κάθε πτήση λαμβάνεται μια σειρά μετρήσεων από τους 21 αισθητήρες μέχρι το σύστημα του κινητήρα να αστοχήσει μετά από κάποιο αριθμό πτήσεων. Στο Dataset εφαρμόζεται προσέγγιση Regression οπότε γίνεται εκτίμηση (εξαρτημένη μεταβλητή) με βάση το RUL(Remaining Useful Life) όπως αυτό αναλύθηκε και πιο πάνω. Δηλαδή πόσοι κύκλοι ζωής απομένουν. Οι αλγόριθμοι οι οποίοι χρησιμοποιήθηκαν είναι Decision tree, SVM, Random Forest, K-nearest neighbors, K-means, Gradient Boosting Method, Adaboost, Deep Learning, Anova. Αποδεικνύεται ότι ο Random Forest έφερε τα καλύτερα αποτελέσματα.

Στην εργασία τους οι Daniel Mariani [19] εξετάζουν τους αεροστροβίλους LM2500 της General Electric οι οποίοι αποτελούν την ραχοκοκαλιά του στόλου των Guided Missile Destroyers (DDG) και Guided Missile Cruisers (CG) του Αμερικάνικου πολεμικού ναυτικού (US Navy). Ένα πραγματικό dataset από 60 αεροστροβίλους οι οποίοι ανήκουν σε 15 πλοία του στόλου, για χρονικό διάστημα 5 χρόνων δημιουργήθηκε για την χρονική περίοδο 2012-2017. Σκοπός είναι ο εντοπισμός ανίχνευσης μελλοντικής βλάβης εκπαιδεύοντας ένα μοντέλο με διάφορες μετρήσεις θερμοκρασιών, δονήσεων, ροών και άλλων τεχνικών μεγεθών τα οποία καταγράφονται από το σύστημα ελέγχου του αεροστροβίλου, 280.963 υγιείς μετρήσεις χρησιμοποιήθηκαν στην εργασία. Το μοντέλο

της μηχανικής μάθησης το οποίο χρησιμοποιήθηκε είναι MLP με ένα κρυμμένο επίπεδο. Η επιλογή του αριθμού των Νευρώνων και των κρυμμένων επιπέδων έγινε με χρήση γενετικού αλγορίθμου και έδειξε ότι περισσότερα επίπεδα δεν προσφέρουν καλύτερο αποτέλεσμα. Το μοντέλο που εκπαιδεύτηκε λαμβάνει τις μετρήσεις από τους αισθητήρες και προβλέπει τις τιμές μεταβλητών στην έξοδο οπότε και τις συγκρίνει με τις πραγματικές τιμές εξόδου, αν οι τιμές που καταγράφονται έχουν μεγαλύτερη απόκλιση-residual από ένα προκαθορισμένο όριο τότε υπάρχει ένδειξη για μελλοντικό fault όπως φαίνεται και στην εικόνα 2.4. Οπότε γίνεται αντιληπτό ότι με βάση τους τύπους συντήρησης όπως αυτοί αναλύθηκαν στο πρώτο κεφάλαιο, στην παρούσα εργασία ακολουθείται το Condition based Maintenance(CBM). Η CBM αρχιτεκτονική ακολουθείται από αρκετά μοντέρνα συστήματα της βαριάς βιομηχανίας. Χαρακτηριστικό παράδειγμα το σύστημα Orflex της General Electric το οποίο εφαρμόζεται σε όλα τα σύγχρονα Powerplants για αεριοστρόβιλους, ατμοστρόβιλους, γεννήτριες και τον συνδυασμό τους όταν αναφερόμαστε σε Combined Cycle Power Plant [20], με σκοπό την real time εκτίμηση δεδομένων και την απόκλιση τους από την καταγραφόμενη τιμή. Αν το Residual ξεπεράσει ένα προκαθορισμένο όριο τότε το σύστημα μπορεί να λάβει δράση είτε με την απλή εξαγωγή alarm στον operator είτε για την προστασία του ίδιου του συστήματος να δώσει Trip στον εξοπλισμό.



Εικόνα 2.4 Η CBM αρχιτεκτονική της [19]

(Development of a Data-driven Model for Marine Gas Turbine (MGT) Engine Health Monitoring 2018)

Μια ευρέως διαδεδομένη data-driven μέθοδος η οποία χρησιμοποιείται για εκτίμηση του RUL είναι η χρήση Convolutional Neural Network-CNN . Ο Babu στο [20]

είναι ο πρώτος που χρησιμοποίησε deep CNN προσέγγιση για εκτίμηση-πρόβλεψη του RUL. Τα αποτελέσματα έδειξαν ότι το CNN αποδίδει καλύτερα από τα multi-layer perceptron (MLP), τον (SVM) και τα significance vector regression (SVR) μοντέλα. Ο παραπάνω CNN εφαρμόστηκε στο C-MAPSS dataset της NASA. Οι Huang et al. [21] χρησιμοποιούν ένα multi-layer perceptron (MLP) για πρόβλεψη του RUL σε έδρανα ολίσθησης και τα αποτελέσματα τους ήταν πολύ καλύτερα από αυτά τα οποία εκτιμούσαν οι θεωρίες αξιοπιστίας για τον ωφέλιμο χρόνο ζωής του εξοπλισμού. Οι Zhao et al. [22] χρησιμοποιούν LSTM για την εκτίμηση ζωής κοπτικού εργαλείου, χρησιμοποιώντας δεδομένα από το πεδίο του χρόνου αλλά και από το πεδίο της συχνότητας. Η εικόνα 2.5 δείχνει τις πιο καινοτόμες εργασίες που έχουν γίνει βάσει του Dataset του C-MAPSS της NASA όσον αφορά το κομμάτι της βαθιάς μηχανικής μάθησης.

Authors	Year	Method Used	Benchmark Dataset	Results Achieved	Limitations/Gaps
Peng et al., [2]	2021	FCLCNN-LSTM	C-MAPSS	This model verified only with FD001 (11.17) and FD003 (9.99) subset of data	The key drawback of this model is the need to incrementally update the prognosis results.
Wen, Dong and Gao [12]	2019	ResCNN	C-MAPSS	RMSE for FD001 (12.16) RMSE for FD002 (20.85) RMSE for FD003 (12.01) RMSE for FD004 (24.79)	The limitations of the proposed method are that the imbalance of signal data is ignored, and the tuning parameter process of the ensemble ResCNN is very time-consuming.
Babu et al., [13]	2016	First attempt at a deep CNN	C-MAPSS	RMSE for FD001 (18.44) RMSE for FD002 (30.29) RMSE for FD003 (19.81) RMSE for FD004 (29.15)	The limited accuracy of the RUL estimation, means this method is not practical for real-world applications.
Li, Ding & Sun, [14]	2018	DCNN	C-MAPSS	RMSE for FD001 (18.45) RMSE for FD002 (22.36) RMSE for FD003 (12.64) RMSE for FD004 29.16	Additional architecture improvements are required, as the current training time exceeds that of the majority of shallow networks in the literature.
Zhang et al., [36]	2016	Multi-objective deep belief network ensemble	C-MAPSS	RMSE for FD001 (15.04) RMSE for FD002 (25.05) RMSE for FD003 (12.51) RMSE for FD004 (28.66)	This model suffers from slow prediction process and limited accuracy of RUL estimation, which made it not cost-effective method in industrial contexts. The main drawback can be summarised in twofold. First, the limited accuracy of the RUL prediction, which make this method is not practical for industrial contexts. Second, high computational load.
Zheng et al., [37]	2017	Deep LSTM	C-MAPSS	RMSE for FD001 (16.14) RMSE for FD002 (24.49) RMSE for FD003 (16.18) RMSE for FD004 (28.17)	Further architecture improvements are required, as the current model need more optimization.
Zhu, Chen and Peng [38]	2018	multi-scale CNN	PRONOSTIA	Tested on bearing dataset	

Εικόνα 2.5 Αποτελέσματα του C-MAPSS

(A Remaining Useful Life Prognosis of Turbofan Engine Using Temporal and Spatial Feature Fusion (2021))

## 2.3 DATASET

Το C-MAPSS είναι μια βάση δεδομένων από Dataset λειτουργίας Turbofan-αεριοστρόβιλων της NASA. Το C-MAPSS turbofan engine degradation data είναι η πιο ευρέως διαδεδομένη βάση δεδομένων προγνωστικών ελέγχων και περιέχει 4 διαφορετικά datasets FD001, FD002, FD003,FD004. Συγκεκριμένα για κάθε κύκλο λειτουργίας του κάθε αεριοστρόβιλου γίνεται μια καταγραφή δεδομένων από όλους τους αισθητήρες που φέρει ο αεριοστρόβιλος. Αυτό επαναλαμβάνεται για κάθε κύκλο λειτουργίας μέχρι τελικά ο κινητήρας να αστοχήσει. Δηλαδή για κάθε πτήση του αεροσκάφους γίνεται λήψη μιας στιγμιαίας εικόνας του κινητήρα και αυτό επαναλαμβάνεται κάθε φορά μέχρι τον κύκλο που ο κινητήρας θα αστοχήσει. Είναι ένα Run to Failure Dataset. Δηλαδή αν ο κινητήρας αστοχήσει στον 192<sup>ο</sup> κύκλο λειτουργίας του, θα έχουν ληφθεί 192 μετρήσεις (χρονοσειρές) - μία για κάθε κύκλο. Κάθε Dataset περιέχει training και test set ενώ ο αριθμός των αεριοστρόβιλων σε κάθε Dataset είναι διαφορετικός όπως φαίνεται και στην εικόνα 2.6.

Dataset	C-MAPSS			
	FD001	FD002	FD003	FD004
Engine units for training	100	260	100	249
Engine units for testing	100	259	100	248
Operating conditions	1	6	1	6
Fault modes	1	1	2	2

Εικόνα 2.6 Τα 4 διαφορετικά Dataset

(Data-Driven Deep Learning-Based Attention Mechanism for Remaining Useful Life Prediction: Case Study Application to Turbofan Engine Analysis (2021))

Η κάθε μηχανή αεριοστρόβιλου έχει παραχθεί ξεχωριστά στην γραμμή παραγωγής, οπότε η κάθε μια έχει διαφορετικές κατασκευαστικές ατέλειες και για αυτό θα αστοχήσουν σε διαφορετικό κύκλο. Επιπλέον η κάθε μηχανή επιχειρεί σε διαφορετικές συνθήκες. Αλλιώς, ίδιες ακριβώς μηχανές οι οποίες καταπονούνται κάτω από τα ίδια φορτία θα αστοχούσαν στον ίδιο κύκλο. Η κάθε μηχανή αποδίδει στην αρχή της



λειτουργίας της βάσει των εργοστασιακών της προδιαγραφών και όσο περνάει ο χρόνος η πιθανότητα σφάλματος αυξάνεται συνεχώς, μέχρι που τελικά η μηχανή αστοχεί.

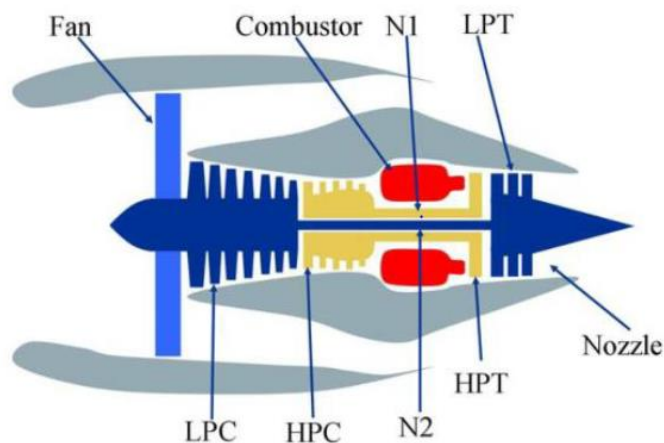
Σε κάθε χρονοσειρά γίνεται καταγραφή 21 μεγεθών από λήψεις αισθητήρων πάνω στον αεριοστρόβιλο και 3 παραμέτρους που αφορούν την λειτουργική κατάσταση της μηχανής. Στην εικόνα 2.7 φαίνονται όλα τα μεγέθη τα οποία καταγράφονται από τους αισθητήρες για κάθε πτήση ή κύκλο κόπωσης της μηχανής. Ενώ στην εικόνα 2.8 φαίνεται μια απλουστευμένη μορφή της μηχανής του αεριοστρόβιλου η οποία απεικονίζει μεγέθη-μεταβλητές τα οποία καταγράφονται στον πίνακα μετρήσεων της εικόνας 2.7.

Sensor Number	Sensor Description	Units
1	(Fan inlet temperature)	(°R)
2	(LPC outlet temperature)	(°R)
3	(HPC outlet temperature)	(°R)
4	(LPT outlet temperature)	(°R)
5	(Fan inlet Pressure)	(psia)
6	(bypass-duct pressure)	(psia)
7	(HPC outlet pressure)	(psia)
8	(Physical fan speed)	(rpm)
9	(Physical core speed)	(rpm)
10	(Engine pressure ratio (P50/P2))	—
11	(HPC outlet Static pressure)	(psia)
12	(Ratio of fuel flow to Ps30)	(pps/psia)
13	(Corrected fan speed)	(rpm)
14	(Corrected core speed)	(rpm)
15	(Bypass Ratio)	—
16	(Burner fuel-air ratio)	—
17	(Bleed Enthalpy)	—
18	(Required fan speed)	(rpm)
19	(Required fan conversion speed)	(rpm)
20	(High-pressure turbines Cool air flow)	(lb/s)
21	(Low-pressure turbines Cool air flow)	(lb/s)

Εικόνα 2.7 Τα μετρήσιμα μεγέθη του Dataset

(A Remaining Useful Life Prognosis of Turbofan Engine Using Temporal and Spatial Feature Fusion (2021))

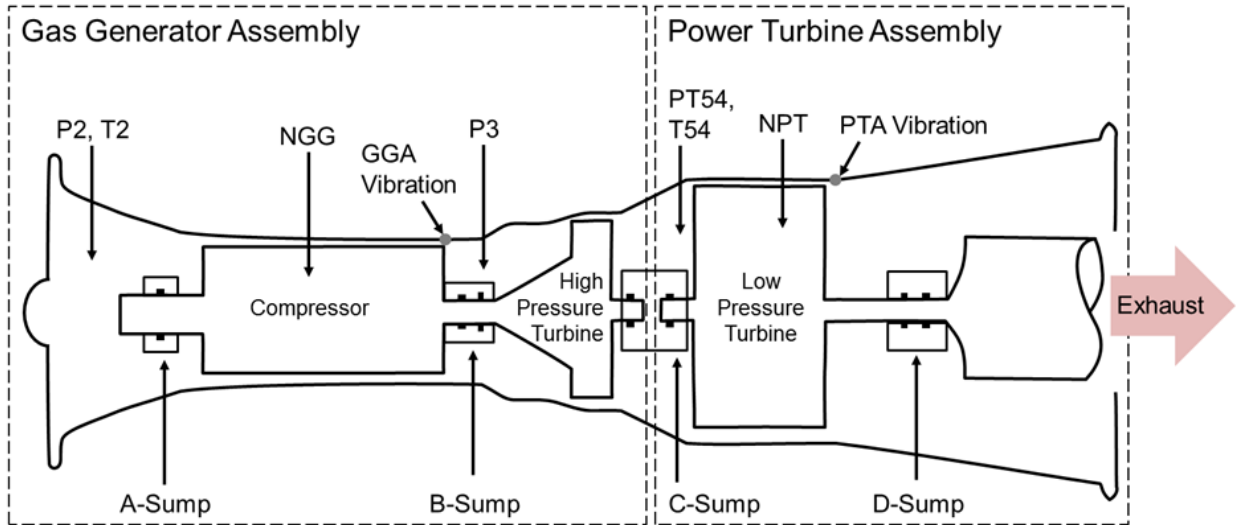
Σχετικά με τα μεγέθη της εικόνας 2.8, ο συμπιεστής χαμηλής πίεσης (LPC) και ο συμπιεστής υψηλής πίεσης (HPC) δημιουργούν αέρα υψηλής πίεσης και θερμοκρασίας ο οποίος στην συνέχεια εισέρχεται στους θαλάμους καύσης. Η Low pressure turbine (LPT) είναι η τουρμπίνα χαμηλής πίεσης, μπορεί να επιβραδύνει ή να επιταχύνει και ανάλογα να αυξηθεί ή να μειωθεί η πίεση του εισερχόμενου αέρα με σκοπό να βελτιωθεί η καύση. Η τουρμπίνα υψηλής πίεσης (HPC), δημιουργεί το κυρίως έργο. Ενώ αντίστοιχα με N1 και N2 είναι η ταχύτητα του ρότορα-άξονα της υψηλής πίεσης και με N2 της χαμηλής πίεσης.



Εικόνα 2.8 Τα μετρήσιμα μεγέθη του FD001 όπως αυτά τοποθετούνται σε έναν αεριοστρόβιλο

(A Remaining Useful Life Prognosis of Turbofan Engine Using Temporal and Spatial Feature Fusion (2021))

Για πιο λεπτομερειακή απεικόνιση της δομής του αεριοστρόβιλου ανάλυση γίνεται στην εικόνα 2.9 όπου και φαίνεται η αρχιτεκτονική των αεριοστροβίλων LM2500 όπως αυτή αναλύθηκε στην ενότητα 2.2 και αναφέρεται στην [19]. Εδώ φαίνονται με περισσότερες λεπτομέρειες τα μέρη του αεριοστρόβιλου από τα οποία λαμβάνονται μετρήσεις και μετά όπως εξηγήθηκε αναλύονται σε μια CBM αρχιτεκτονική. Η εικόνα 2.9 δεν αναφέρεται στο Dataset της παρούσας εργασίας αλλά δρα συμπληρωματικά στην εικόνα 2.8 .



Εικόνα 2.9 Γενική αρχιτεκτονική κρίσιμων μετρούμενων μεγεθών ενός αεριοστρόβιλου

(Development of a Data-driven Model for Marine Gas Turbine (MGT) Engine Health Monitoring,(2018))

### 3.1 Τί είναι η Τεχνητή νοημοσύνη;

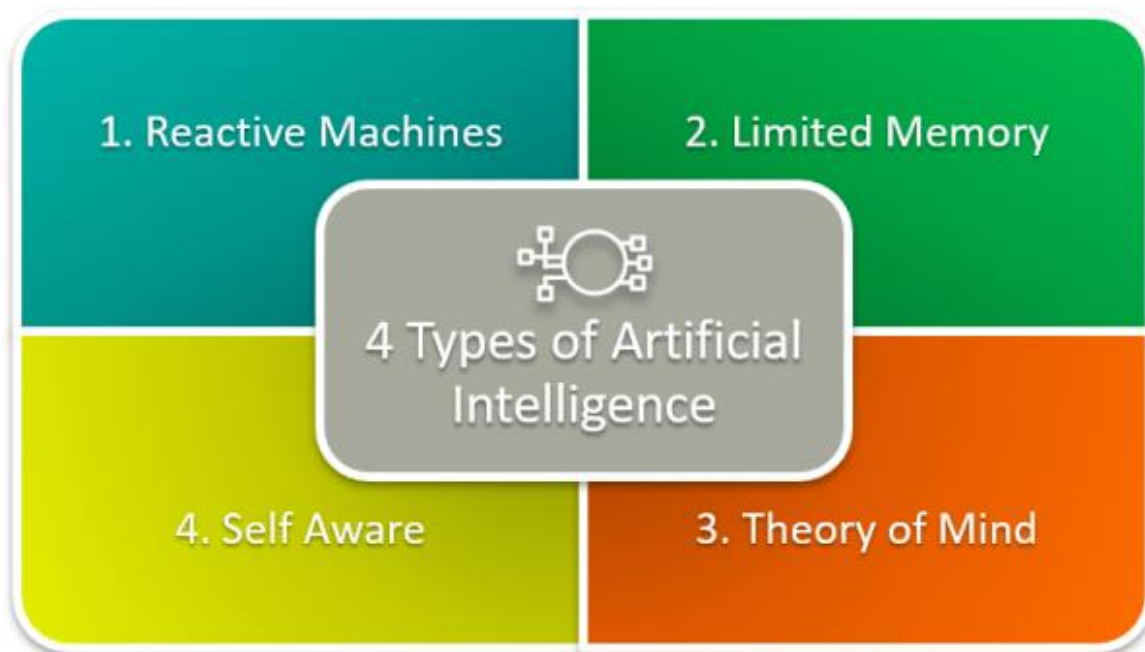
Με τον όρο τεχνητή νοημοσύνη αναφερόμαστε στον τομέα της επιστήμης των υπολογιστών ο οποίος ασχολείται με τη σχεδίαση και υλοποίηση προγραμμάτων τα οποία έχουν την δυνατότητα να μιμηθούν τις ανθρώπινες γνωστικές ικανότητες. Με τον τρόπο αυτό εμφανίζουν χαρακτηριστικά τα οποία συνήθως αποδίδουμε σε ανθρώπινη συμπεριφορά όπως η μάθηση, η επίλυση προβλημάτων, η κατανόηση της φυσικής γλώσσας και η επίλυση προβλημάτων[27].Οι ορισμοί της τεχνητής νοημοσύνης μπορεί να διαφέρουν, αλλά μπορούμε να πούμε ότι τα νευρωνικά δίκτυα είναι εμπνευσμένα από τη βιολογία.

Μπορούμε να πούμε ότι Τεχνητή Νοημοσύνη ή AI (Artificial Intelligence), είναι ο τομέας της έρευνας που επιδιώκει να κάνει τους υπολογιστές να εκτελούν διαδικασίες που όταν γίνονται από ανθρώπους ή ζώα θεωρούνται διανοητικά ευφυείς. Φυσικά, για την «AI» υπάρχουν περισσότεροι τεχνικοί ορισμοί. Η νοημοσύνη, είναι η ιδιότητα ενός παράγοντα που του επιτρέπει να βελτιστοποιεί το περιβάλλον του. Έτσι η τεχνητή νοημοσύνη βασίζεται σε αυτή την ιδιότητα.

### 3.1.1 Οι βασικές κατηγορίες της Τεχνητή νοημοσύνης

Οι 4 βασικοί τύποι της τεχνητής νοημοσύνης είναι:

1. Reactive Machines
2. Limited Memory
3. Theory of Mind
4. Self Aware



Εικόνα 3.1 Οι 4 βασικοί τύποι της τεχνητής νοημοσύνης

(AWS documentation)

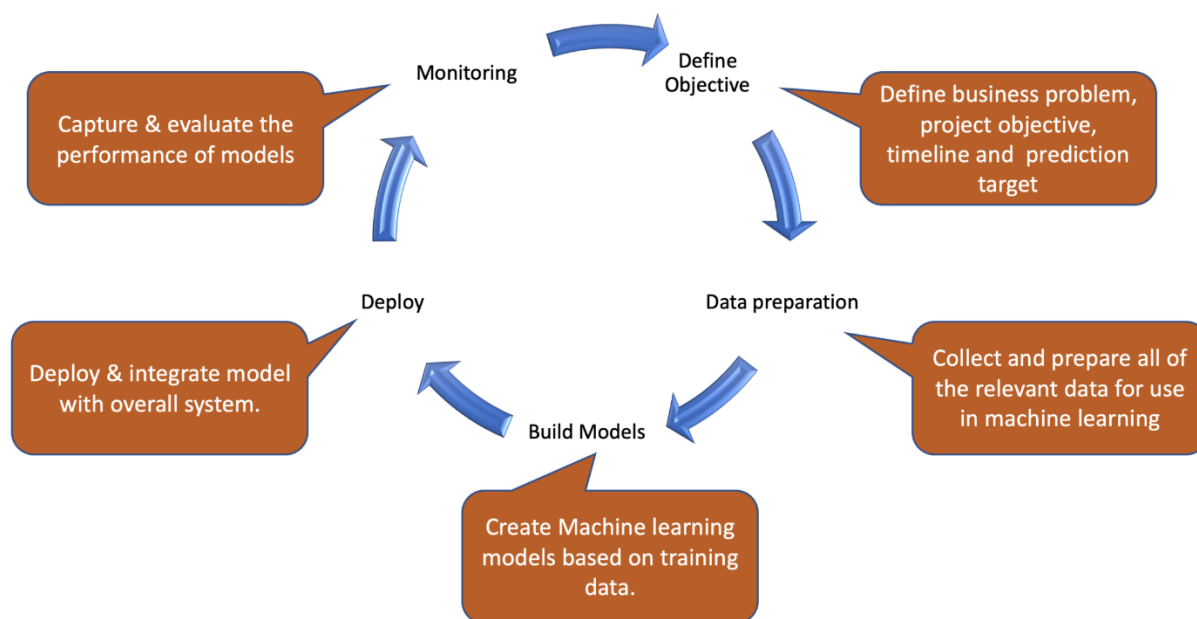
Οι **Reactive Machines** υλοποιούν βασικές λειτουργίες. Είναι η πιο απλή εφαρμογή Α.Ι. Ουσιαστικά αυτός ο τύπος Α.Ι αντιδρά σε κάποιες εισόδους με κάποιες εξόδους. Δεν υλοποιείται η διαδικασία της μάθησης. Το μοντέλο δεν αποθηκεύει εισόδους και δεν έχει διαδικασία μάθησης, π.χ. ένα μοντέλο το οποίο απλά αναγνωρίζει ένα ανθρώπινο πρόσωπο είναι ένα reactive machine.

Οι **Limited memory** αποθηκεύουν προηγούμενα δεδομένα και προβλέψεις με σκοπό να επιτύχουν καλύτερες προβλέψεις. Στα Limited Memory μοντέλα, η αρχιτεκτονική του αλγορίθμου γίνεται πιο πολύπλοκη. Κάθε μοντέλο χρειάζεται μνήμη ώστε να δημιουργηθεί και το μοντέλο αφού δημιουργηθεί μπορεί να λειτουργήσει σαν Reactive machine. Στην κατηγορία αυτή ανήκουν και τα μοντέλα του Machine learning.

Τα βασικά στάδια ενός μοντέλου μηχανικής μάθησης είναι:

1. **Training Data.** Το μοντέλο μηχανικής μάθησης πρέπει να έχει δεδομένα ώστε να εκπαιδευτεί.
2. **Build ML Model.** Το μοντέλο δημιουργείται.
3. **Model Predictions.** Το μοντέλο κάνει προβλέψεις,
4. **Feedback.** Στο μοντέλο εισέρχεται η ανάδραση βάσει των προβλέψεων του είτε από τον ανθρώπινο παράγοντα (tuning) ή από το περιβάλλον(reinforcement learning).
5. **Feedback becomes data.** Τα καινούργια δεδομένα ή hyperparameters εισέρχονται στο σύστημα.
6. **Repeat Step 1.** Επανάληψη του 1<sup>ου</sup> σταδίου.

## The Machine Learning Life cycle

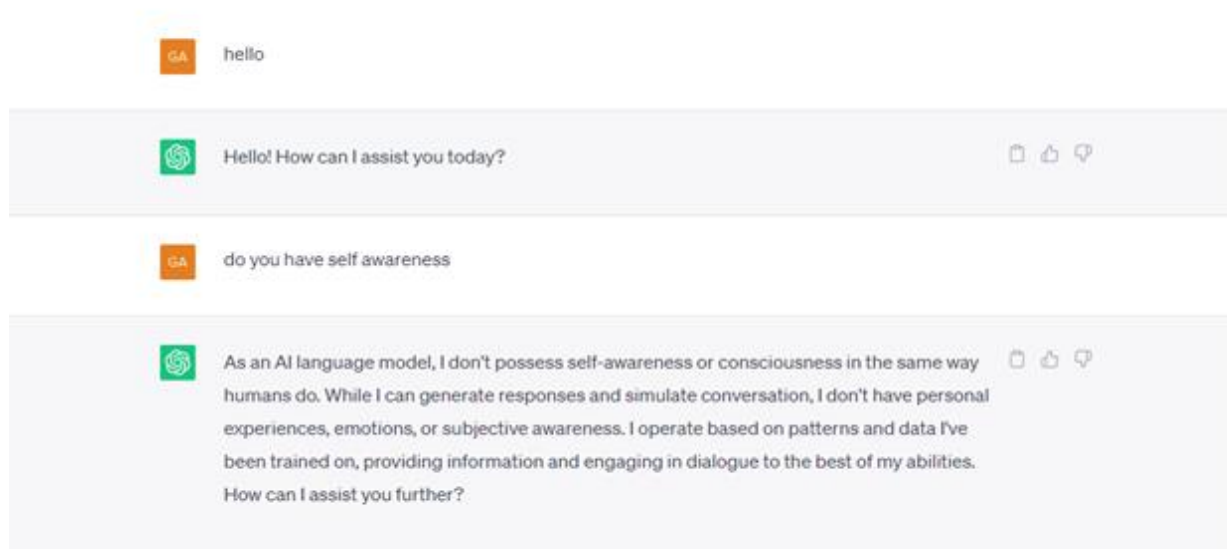


Εικόνα 3.2 Το lifecycle του machine learning

(AWS documentation)

**Theory of Mind** είναι ακόμη σε αρχικά στάδια, είναι η αλληλεπίδραση της Α.Ι με σκέψεις και συναισθήματα των ανθρώπων. Προς το παρόν το Α.Ι επιδρά με σκοπό την υλοποίηση μιας συγκεκριμένης αποστολής(task) χωρίς να αλληλοεπιδρά με τα συναισθήματα του χρήστη. Χαρακτηριστικό παράδειγμα αν απευθυνθεί κάποιος στην alexa ή στο Google maps με έντονο ύφος και τόνο δεν θα αλλάξει κάτι στα αποτελέσματα τα οποία θα δώσει στο χρήστη(π.χ. συναισθηματική υποστήριξη).Σήμερα το chat gpt φαίνεται να έχει σαφώς καλύτερα αποτελέσματα προς την κατεύθυνση αυτή.

**Self-Aware** σε ένα κοντινό μέλλον το Α.Ι ίσως δύναται να αποκτήσει επίγνωση της οντότητας του. Μια νοημοσύνη με επίγνωση της οντότητας της θα προσφέρει στον άνθρωπό την δυνατότητα να διαπραγματευτεί μαζί της βάσει συγκεκριμένων θεμάτων. Στην παρακάτω εικόνα φαίνεται διάλογος ο οποίος υλοποιήθηκε με το chatgpt τον Μάιο του 2023.



Εικόνα 3.3 Διάλογος με το Chat GPT σχετικά με το Self awareness

( OPENAI.COM (CHATGPT))

### 3.2 Τί είναι η μηχανική μάθηση;

Μάθηση είναι η διαδικασία εμπέδωσης καινούργιων δεδομένων σε υποκείμενα(ανθρώπους) και αντικείμενα(υπολογιστικά συστήματα) ικανά προς αυτό, η

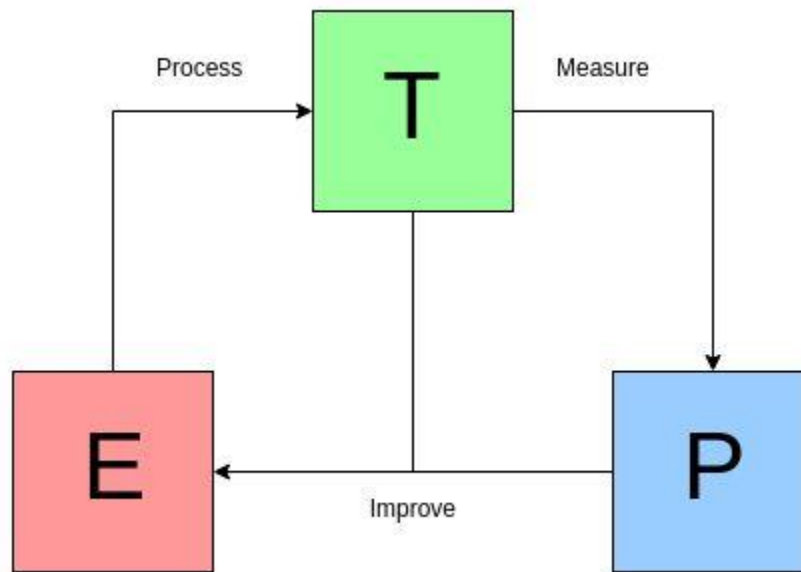
μετατροπή τους σε δεδομένα συμβατά με τις λειτουργίες του υποκειμένου ή αντικειμένου - μ' ένα κώδικα είτε ιδιοσυστατικό είτε αναγνωρίσιμο είτε και τα δύο - και η ανάκλησή τους.

Η μηχανική μάθηση θα μπορούσε να οριστεί ως:

«Η ικανότητα ενός συστήματος να βελτιώνει την απόδοσή του για την εκτέλεση μιας συγκεκριμένης εργασίας, χωρίς να χρειάζεται ο εκ νέου προγραμματισμός του συστήματος αυτού. »

Σκοπός της μηχανικής μάθησης είναι η δημιουργία μηχανών ικανών να βελτιώνουν την απόδοσή τους σε συγκεκριμένους τομείς με αξιοποίηση προηγούμενης εμπειρίας και γνώσης η οποία εισέρχεται στο σύστημα. Σύμφωνα με τον ορισμό του Mitchell:

«Ένα πρόγραμμα υπολογιστή λέμε ότι μαθαίνει από την εμπειρία E ως προς κάποια κλάση εργασιών T και μέτρο απόδοσης P, αν η απόδοσή του σε εργασίες από το T, όπως μετριέται από το P, βελτιώνεται μέσω της εμπειρίας E.»



Εικόνα 3.4 Σχεδιαγραμματικός ορισμός της ερμηνείας Mitchell

( Τεχνητή νοημοσύνη, [Βλαχαβάς 2006])



Στην **Επαγωγική Μάθηση** ή Inductive Learning ένας άνθρωπος δύναται να κατανοήσει το περιβάλλον του μέσω παρατηρήσεων δημιουργώντας μια αφαιρετική-απλοποιημένη εκδοχή του η οποία ονομάζεται νοητικό μοντέλο ή mental model. Οργανώνοντας και συσχετίζοντας εμπειρίες και παρατηρήσεις δημιουργεί νέες δομές οι οποίες ονομάζονται νοητικά πρότυπα. Σε αυτή την δημιουργία νέων προτύπων βασιζόμενη σε παλιά πρότυπα στηρίζονται οι τρόποι μάθησης οι οποίοι εξαρτώνται άλλοτε σε μεγαλύτερο και άλλοτε σε μικρότερο βαθμό από την προ υπάρχουσα γνώση για ένα συγκεκριμένο πρόβλημα, για παράδειγμα μάθηση από επεξηγήσεις και μάθηση από περιπτώσεις.

Σε φιλοσοφικό επίπεδο το ερώτημα το οποίο πρέπει να απαντηθεί είναι

«Πώς ένας επαγωγικός συλλογισμός ο οποίος οδηγεί στη μάθηση μπορεί να αξιολογηθεί για την ορθότητά του; »

Από την πλευρά της Τεχνητής Νοημοσύνης το ερώτημα το οποίο πρέπει να απαντηθεί είναι

«Πώς δύναται μια μηχανή να δημιουργήσει μοντέλα και πρότυπα μάθησης χρησιμοποιώντας συγκεκριμένα παραδείγματα και στην πράξη τα μοντέλα αυτά πόσο αξιόπιστα είναι;».

Συνοψίζοντας όλα τα παραπάνω μπορεί να δοθεί για την μηχανική μάθηση ο εξής ορισμός:

«Μηχανική Μάθηση είναι η ικανότητα δημιουργίας μοντέλων ενός υπολογιστικού συστήματος με την χρήση ενός συνόλου δεδομένων».

### 3.3 Είδη Μηχανικής Μάθησης

Η μηχανική μάθηση[27],[28] αναπτύσσεται με τρεις τρόπους, ανάλογους με τους τρόπους με τους οποίους δύναται να μάθει και ο άνθρωπος: επιβλεπόμενη μάθηση, μη επιβλεπόμενη μάθηση και ενισχυτική μάθηση.

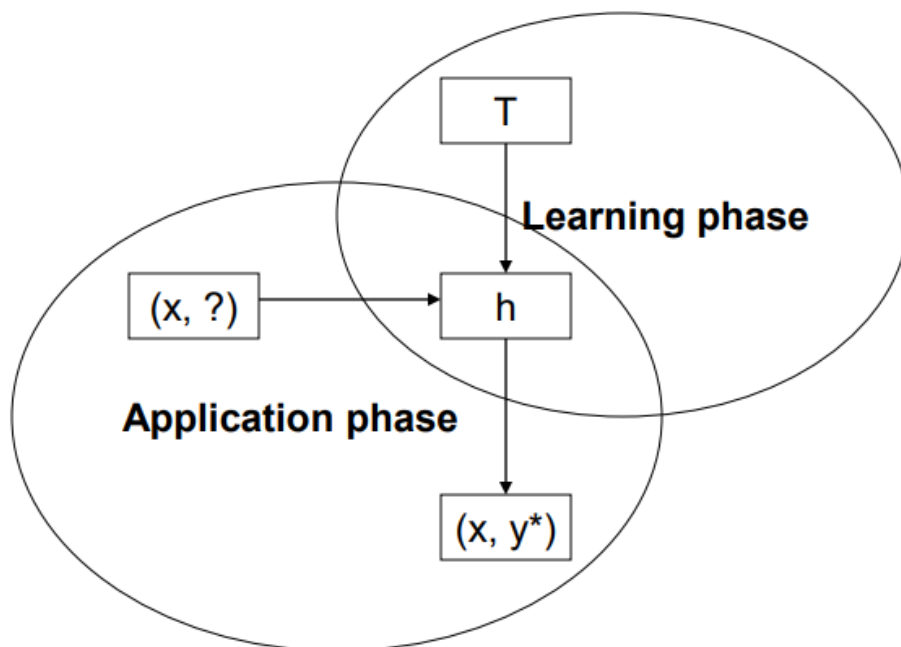
- Supervised Learning(επιβλεπόμενη μάθηση) : Ο αλγόριθμος δημιουργεί μια συνάρτηση η οποία απεικονίζει τις εισόδους του συνόλου εκπαίδευσης σε επιθυμητές εξόδους. Σκοπός είναι είναι γενίκευση της λειτουργίας αυτής για εισόδους που έχουν άγνωστη έξοδο. Τα προβλήματα στα οποία γίνεται χρήση της μεθόδου είναι:
  - 1) Προβλήματα ταξινόμησης ή Classification
  - 2) Προβλήματα πρόγνωσης ή Prediction
  - 3) Προβλήματα διερμηνείας ή Interpretation
- Unsupervised Learning (Μη επιβλεπόμενη μάθηση): Κατασκευάζεται ένα μοντέλο για ένα σύνολο εισόδων με μορφή παρατηρήσεων χωρίς όμως να είναι γνωστές οι επιθυμητές έξοδοι. Τα προβλήματα στα οποία γίνεται χρήση της μεθόδου είναι:
  - 1) Ανάλυσης συσχετισμών ή Association analysis
  - 2) Ομαδοποίησης
- Reinforcement Learning (ενισχυτική μάθηση): Μέσω της αλληλεπίδρασης με το περιβάλλον ο αλγόριθμος μαθαίνει μια στρατηγική ενεργειών.

Για όποιο πρόβλημα η μηχανική μάθηση καλείται να επιλύσει ,θα υπάρχει ένας κατάλληλος τρόπος μάθησης , ενώ για κάθε τρόπο μάθησης υπάρχει ένας τουλάχιστον κατάλληλος αλγόριθμος ο οποίος μπορεί να χρησιμοποιηθεί.

Σημαντικοί όροι στην ορολογία της μηχανικής μάθησης είναι:

- 1) Στιγμιότυπα (instances): Είναι τα παραδείγματα εκπαίδευσης, μια σειρά(row) του dataset.
- 2) Χαρακτηριστικά(Attributes or features): Είναι οι μεταβλητές εισόδου του μοντέλου. Ανεξάρτητες μεταβλητές.
- 3) Μεταβλητή στόχου: Είναι εξαρτημένη μεταβλητή και είναι ο στόχος πρόβλεψης του μοντέλου.
- 4) Συνάρτηση λάθους(loss function) : Είναι μια συνάρτηση βάσει της οποίας βελτιώνεται η μάθηση του μοντέλου, με σκοπό να κάνει λιγότερα λάθη.

### 3.3.1 Η διαδικασία μάθησης ενός μοντέλου βάσης.



Εικόνα 3.5 Η διαδικασία εκπαίδευσης και εφαρμογής ενός μοντέλου βάσης

( Σταματάτος Ευστάθιος σημειώσεις μαθήματος Μηχανικής Μάθησης, πανεπιστήμιο Αιγαίου[2021])

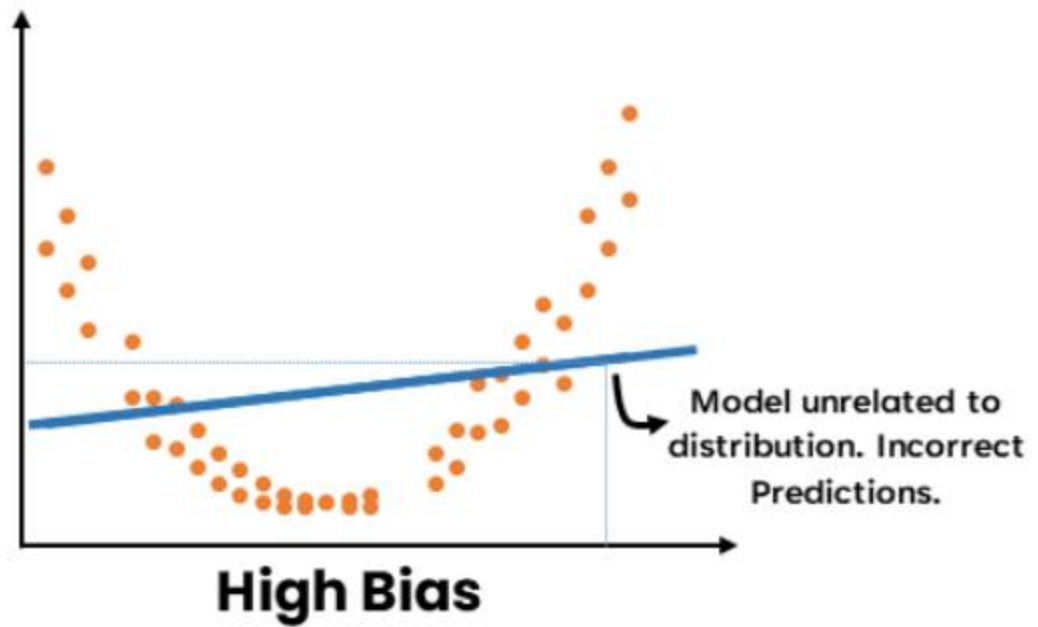
Στην εικόνα φαίνεται η διαδικασία δημιουργίας και πρόβλεψης ενός μοντέλου βάσης, δηλαδή ενός αλγορίθμου μηχανικής μάθησης. Η διαδικασία χωρίζεται σε 2 φάσεις:

Στο **Learning phase** όπου ο αλγόριθμος εκπαιδεύεται βάσει ενός συγκεκριμένου Dataset (T) και μαθαίνει ένα συγκεκριμένο hypothesis(h). Δηλαδή ένα μαθηματικό μοντέλο το οποίο μετατρέπει τα inputs σε outputs. π.χ Για έναν SVM το h του θα ήταν ένα υπερεπίπεδο.

Στο **Application phase** ο αλγόριθμος και αφού έχει δημιουργηθεί το Hypothesis εισέρχονται άγνωστα δεδομένα X (ανεξάρτητες μεταβλητές) και το μοντέλο προβλέπει το  $y^*$ (Εξαρτημένη μεταβλητή).

Ένα βασικό πρόβλημα της μηχανικής μάθησης είναι ότι τα ατομικά μοντέλα (ένα μοντέλο βάσης) δεν αποδίδουν αρκετά ικανοποιητικά. Δηλαδή τείνουν να έχουν μέτρια επίδοση ως προς την ακρίβεια στις προβλέψεις τους. Για να μετριαστεί αυτό το πρόβλημα συνδυάζουμε πολλαπλά μοντέλα με σκοπό να βελτιώσουμε αυτή την επίδοση. Τα ατομικά μοντέλα τα οποία συνδυάζονται είναι γνωστά ως weak learners. Η ονομασία τους προκύπτει λόγω ότι έχουν μεγάλο bias ή μεγάλο variance.

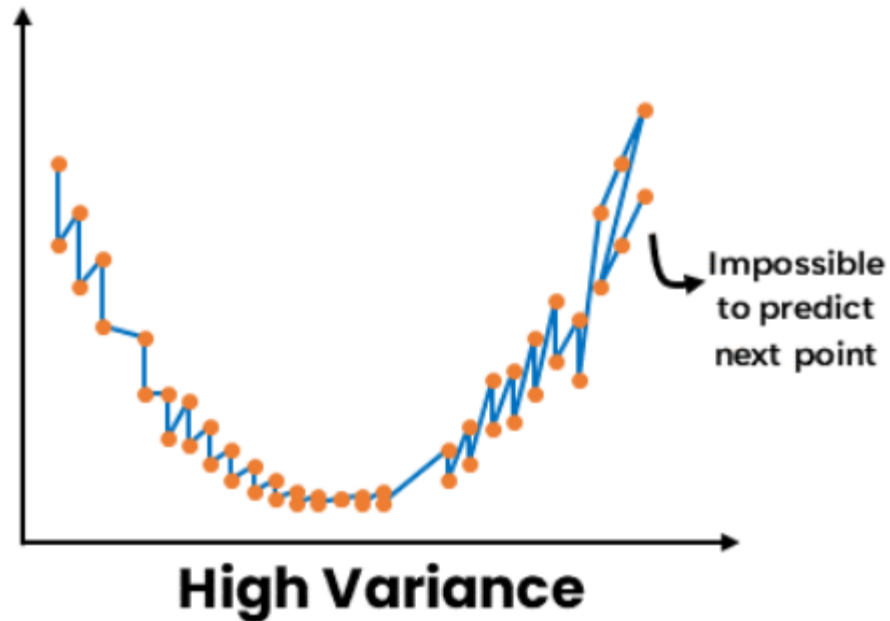
- Ένα μοντέλο με μεγάλο bias δεν μπορεί να μάθει ικανοποιητικά τα δεδομένα. Δεν μπορεί να συσχετιστεί με τα δεδομένα για αυτό, για αυτό και ως αποτέλεσμα υπάρχουν φτωχές προβλέψεις. Ένα μοντέλο με μεγάλο bias φαίνεται στην εικόνα



Εικόνα 3.6 Η αποτύπωση ενός αλγορίθμου με υψηλό Bias

(TowardsDataScience.com)

- Ένα μοντέλο με μεγάλο Variance μπορεί να μάθει εξαιρετικά το σετ εκπαίδευσης, όμως δεν δύναται να κάνει πρόβλεψη με ακρίβεια για άλλα δεδομένα, καθώς ουσιαστικά έχει πάθει Overfitting . Ένα μοντέλο με μεγάλο variance φαίνεται στην εικόνα 3.7



Εικόνα 3.7 Η αποτύπωση ενός αλγορίθμου με υψηλό Variance

(TowardsDataScience.com)

Από κοινού τα μοντέλα με μεγάλο Variance και Bias δεν μπορούν να γενικευτούν με σκοπό να κάνουν καλύτερη εκτίμηση. Η λύση είναι ο συνδυασμός μοντέλων βάσης με σκοπό τη γενίκευση και τη βελτιστοποίηση των εκτιμήσεων.

### Δημιουργία μοντέλων βάσης

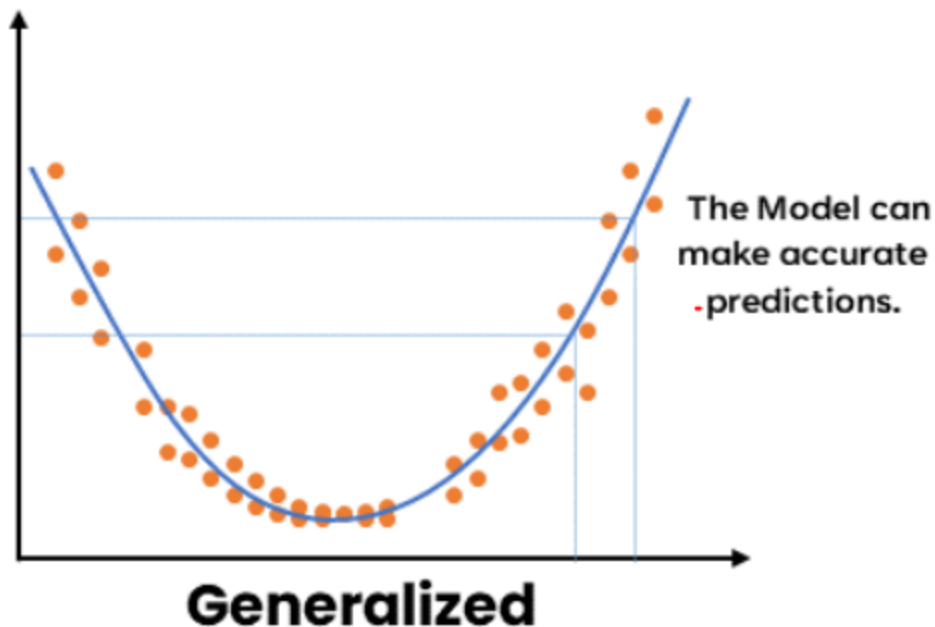
1. Ομογενή-Ετερογενή
2. Επεξεργασία dataset(training set)
3. Χειρισμός Feature set

### Ετερογενή-ομογενή μοντέλα

Δηλαδή οι Learners να βασίζονται σε όμοιους αλγορίθμους για ομογενή μοντέλα. Σε διαφορετικούς αλγορίθμους για ετερογενή μοντέλα. π.χ. ένας learner να είναι SVM, ο άλλος Random Forest, kNN κλπ. Το πλεονέκτημα της προσέγγισης των ετερογενών μοντέλων είναι ο συνδυασμός των διαφορετικών bias αλλά και των διαφορετικών υπερεπιπέδων στο σύστημα. Συνήθως τα ετερογενή μοντέλα δεν χρησιμοποιούνται καθώς η δομή πολλών dataset τα κάνει να μην είναι επεξεργάσιμα από κάποιους αλγορίθμους. Το πιο σημαντικό μειονέκτημα όμως είναι ότι οι learners δεν εξειδικεύονται, κάτι το οποίο οδηγεί σε λάθη και είναι το ζητούμενο της μεθοδολογίας ensemble.

### 3.3.2 Ensemble learning

Το ensemble learning στοχεύει να μειώσει το bias εάν έχουμε ένα αδύναμο μοντέλο με υψηλό bias και χαμηλό variance. Επίσης το ensemble learning στοχεύει να μειώσει το variance εάν έχουμε ένα αδύναμο μοντέλο με υψηλό variance και χαμηλό bias. Με αυτό τον τρόπο, το τελικό μοντέλο θα είναι πολύ πιο ισορροπημένο, με χαμηλό bias και variance. Επομένως, το τελικό μοντέλο το οποίο συντίθεται από τα μοντέλα βάσης weak learners και είναι πολύ πιο βελτιωμένο θα ονομαστεί strong learner. Αυτό το μοντέλο θα είναι πιο γενικευμένο από τα αδύναμα μοντέλα και θα είναι σε θέση να πραγματοποιεί πιο ακριβείς προβλέψεις.



Εικόνα 3.8 Η αποτύπωση ενός αλγορίθμου με ικανοποιητικό Bias και Variance

( TowardsDataScience.com)

Το ensemble learning βελτιώνει την επίδοση ενός μοντέλου με τους εξής τρεις τρόπους:

→Μείωση του variance των weak learners.

→Μείωση του bias των weak learners.

→Βελτίωση της συνολικής ακρίβειας των strong learners.

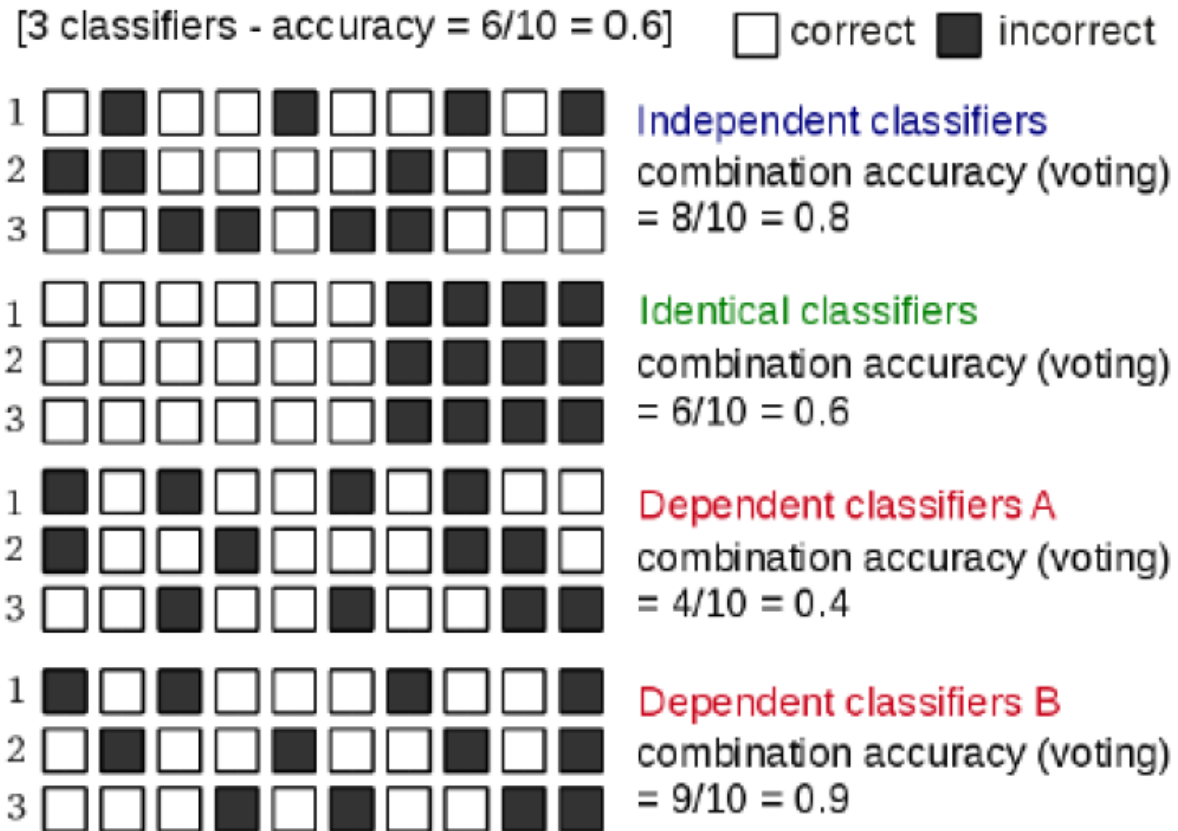
Η μέθοδος bagging χρησιμοποιείται για τη μείωση του variance των weak learner. Η μέθοδος boosting χρησιμοποιείται για τη μείωση του Bias των αδύναμων μοντέλων.

Ένα μοντέλο το οποίο είναι underfit έχει υψηλό bias και χαμηλό Variance ενώ ένα μοντέλο με overfit έχει υψηλό variance και χαμηλό bias. Σε κάθε περίπτωση για να υπάρξει ένα ισορροπημένο μοντέλο με καλές εκτιμήσεις θα πρέπει το bias και το variance να είναι χαμηλό. Το Ensemble learning είναι μια μέθοδος η οποία προσπαθεί να επιτύχει αυτή την ισορροπία(trade-off) προσπαθώντας να μειώσει εξίσου το bias και το variance.

Το Ensemble Learning είναι μια μέθοδος μάθησης η οποία στον πυρήνα της έχει τον συνδυασμό διαφορετικών Weak learners. Δηλαδή περιμένουμε από τον συνδυασμό τους να προκύψουν καλύτερα αποτελέσματα. Τι χαρακτηριστικά όμως θα πρέπει να έχουν αυτοί οι learners ώστε να βελτιώσουν τα υπάρχοντα αποτελέσματα τα οποία



δίνει ο ένας learner. Θέλουν τον συνδυασμό πολλών εξειδικευμένων Learner. Στην εικόνα 3.9 η οποία πραγματεύεται ένα πρόβλημα Classification φαίνεται τί είδους learner φέρνουν καλύτερα αποτελέσματα:



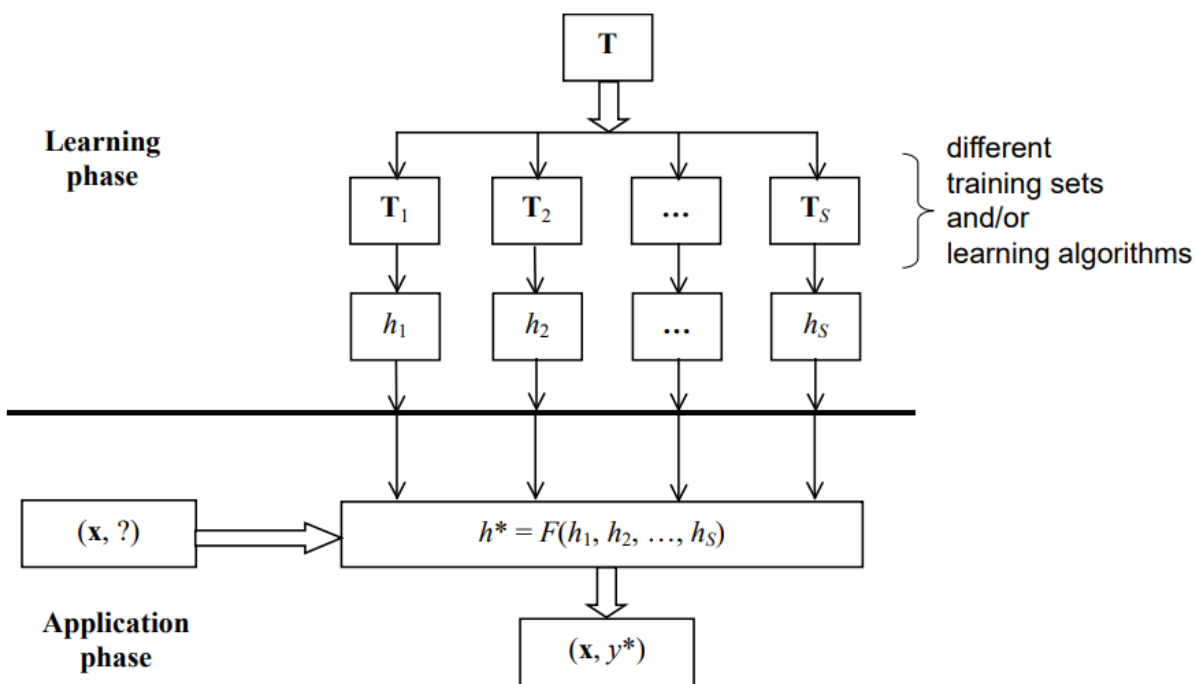
Εικόνα 3.9 Ο ρόλος των συσχετισμών μοντέλων βάσης ως προς την επίδοση του αλγορίθμου

(Σταματάτος Ευστάθιος σημειώσεις μαθήματος Μηχανικής Μάθησης, πανεπιστήμιο Αιγαίου[2021])

Τα συμπεράσματα τα οποία προκύπτουν είναι ότι οι learner είτε θα πρέπει να είναι ανεξάρτητοι μεταξύ τους, είτε θα πρέπει να είναι εξαρτημένοι και να μην κάνει λάθη ο ένας εκεί που κάνει ο άλλος .

Στην εικόνα 3.9 φαίνεται ένα παράδειγμα Ensemble learning και συγκεκριμένα bagging όπου έχει δημιουργηθεί ένα σύνολο από μοντέλα βάσης(σύνολο  $S$ ) και η εκτίμηση του μοντέλου  $h^*$  γίνεται βάσει του συνδυασμού όλων των παραπάνω μοντέλων βάσης. Τα δύο όμως ερωτήματα τα οποία γεννιούνται είναι

- Πως θα δημιουργήσουμε τους learners?
- Πως θα συνδυάσουμε τους learners?



Εικόνα 3.10 Η ένωση διαφορετικών μοντέλων βάσης με την μέθοδο του Bagging

(Σταματάτος Ευστάθιος σημειώσεις μαθήματος Μηχανικής Μάθησης, πανεπιστήμιο Αιγαίου[2021])

## BAGGing

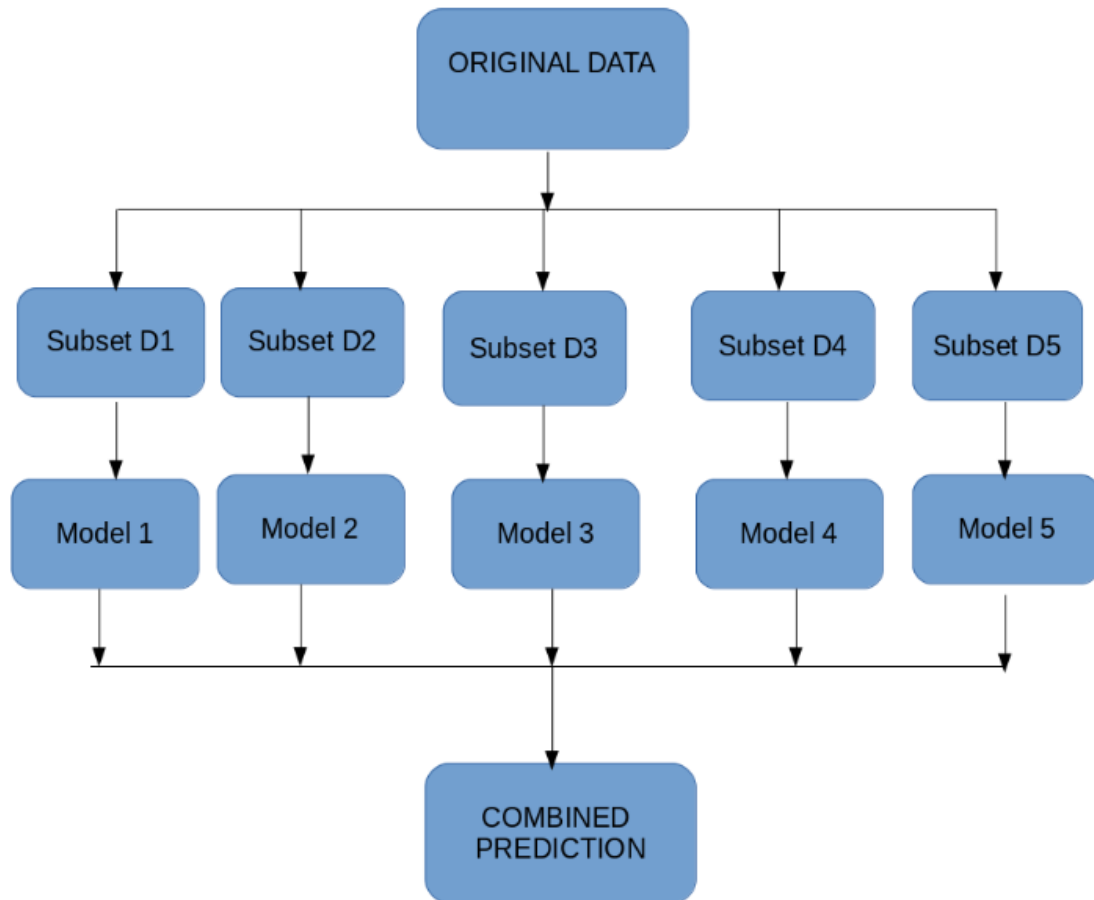
Ο Bagging ή Bootstrap aggregation[28] ανήκει στην πρώτη κατηγορία αλγορίθμων (εικόνα 3.11) .Δηλαδή βασίζεται στην επιλογή παραδειγμάτων (instances) εκπαίδευσης με επανατοποθέτηση. Κάθε φορά επιλέγεται ένα υποσύνολο παραδειγμάτων εκπαίδευσης το οποίο είναι ουσιαστικά ένα subdataset του διαθέσιμου Dataset. Το υποσύνολο αυτό χρησιμοποιείται για την εκπαίδευση του μοντέλου βάσης και επανατοποθετείται στο Dataset. Αυτό σημαίνει ότι το

Για  $i = 1, 2, \dots, K$  έχω:

$T_i \rightarrow$  Τυχαία επιλογή  $M$  παραδειγμάτων εκπαίδευσης με επανατοποθέτηση.

$h_i \rightarrow$  Εκπαίδευση με τα παραδείγματα εκπαίδευσης (kNN,RF,SVM,NN)

Όλα τα μοντέλα βάσης έχουν το ίδιο βάρος, δηλαδή συνεισφέρουν το ίδιο στο τελικό αποτέλεσμα.



Εικόνα 3.11 Η ένωση διαφορετικών μοντέλων βάσης με την μέθοδο του Bagging

( TowardsDataScience.com)

## Boosting

Το Boosting ανήκει στην δεύτερη κατηγορία αλγορίθμων δηλαδή να είναι εξαρτημένοι και να μην κάνει λάθη ο ένας εκεί που κάνει ο άλλος(εικόνα 3.12 & 3.13) . Δηλαδή σε αυτούς είναι μια διαδοχική(sequential) διαδικασία μέσα από την οποία κάθε μοντέλο βάσης προσπαθεί να βελτιώσει τα λάθη του προηγούμενου μοντέλου βάσης. Σε αυτή την τεχνική, μοντέλα βάσης εκπαιδεύονται διαδοχικά και αν ένα μοντέλο βάσης κάνει λάθος στην εκτίμηση του σε κάποια instances τότε στα Instances αυτά αυξάνεται το

βάρος τους με σκοπό το επόμενο μοντέλο βάσης να επικεντρωθεί σε αυτά ώστε να κάνει σωστή εκτίμηση.

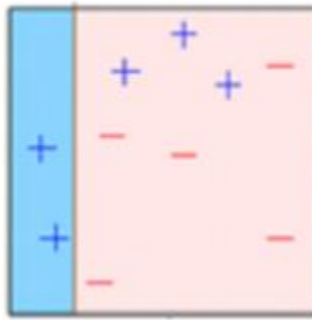
Αρχικά έχουμε ισοκατανεμημένα βάρη σε όλα τα παραδείγματα εκπαίδευσης.

Για  $t = 1, 2, \dots, T$  έχω:

- Εκπαίδευση του μοντέλου βάσης  $f_t(x) \rightarrow y$
- Αναδιανομή των βαρών και αύξηση στα βάρη όπου η πρόβλεψη δεν εκτίμησε σωστά.
- Αναδιανομή των βαρών και μείωση στα βάρη όπου η πρόβλεψη εκτίμησε σωστά.

Με την συγχώνευση όλων των αλγορίθμων από  $f_1 \dots f_t$

1. Ένα υποσύνολο(subdataset) δημιουργείται από το αρχικό Dataset
2. Αρχικά όλα τα instances έχουν το ίδιο βάρος
3. Ένα μοντέλο βάσης δημιουργείται από αυτό το dataset
4. Το μοντέλο αυτό χρησιμοποιείται για να κάνει προβλέψεις στο dataset



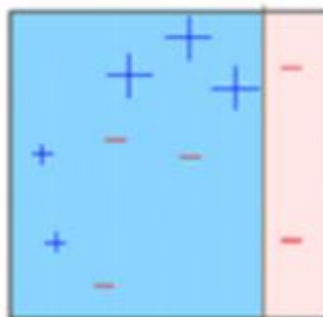
Εικόνα 3.12 Η μέθοδος Boosting 1<sup>ο</sup> βήμα

(Σταματάτος Ευστάθιος σημειώσεις μαθήματος Μηχανικής Μάθησης, πανεπιστήμιο Αιγαίου[2021])

5. Τα λάθη εντοπίζονται συγκρίνοντας πραγματικές και εκτιμώμενες τιμές

6. Στα instances τα οποία εκτιμήθηκαν λάθος δίνεται μεγαλύτερο βάρος. Στην εικόνα 3.13 φαίνονται οι 3 τιμές (μπλε σταυροί) που δεν κατηγοριοποιήθηκαν σωστά. Ενώ στην εικόνα φαίνεται ότι τα βάρος τους αυξήθηκε από το μέγεθος των σταυρών. Ενώ τα παραδείγματα τα οποία τακτοποιήθηκαν σωστά φαίνεται πως μειώθηκε το βάρος τους σε σχέση με την εικόνα 3.12.

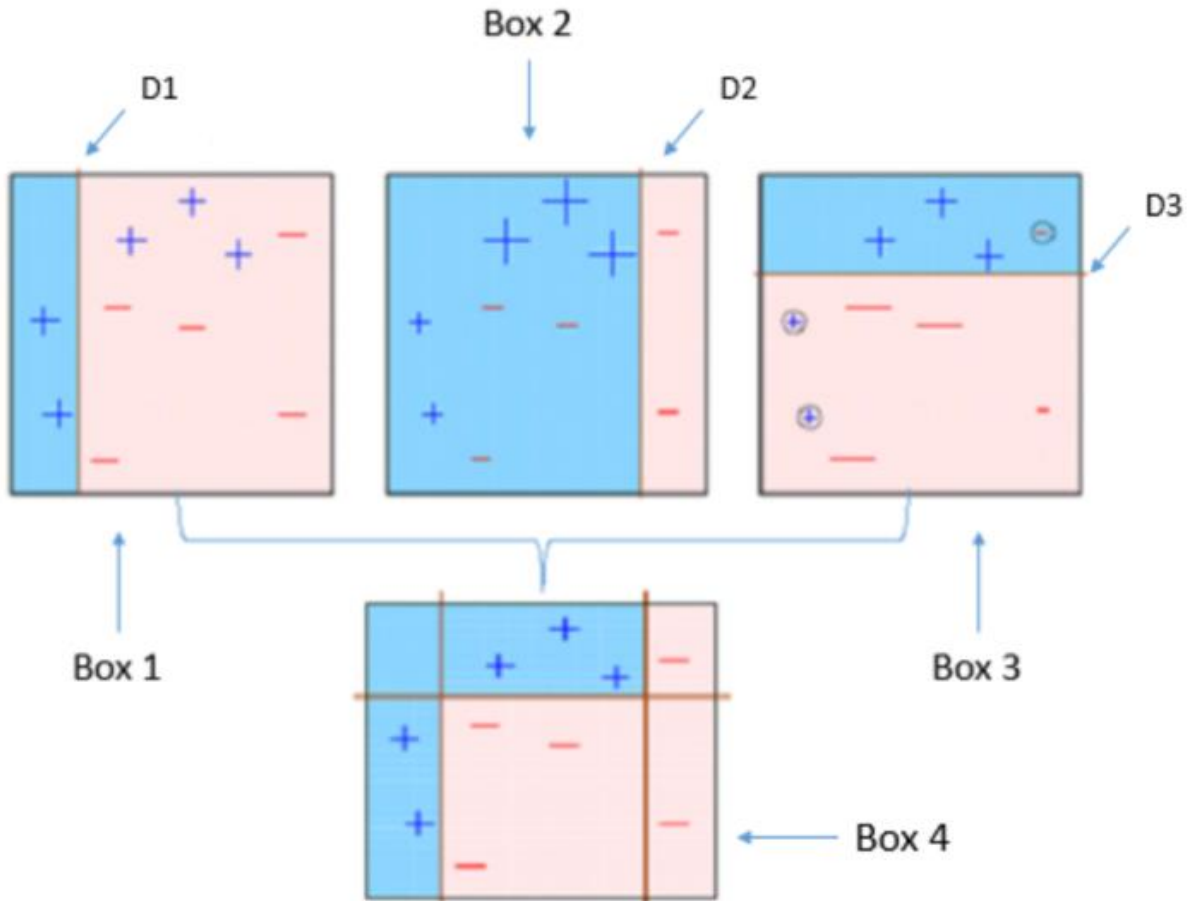
7. Ένα άλλο μοντέλο βάσης δημιουργείται και γίνονται εκτιμήσεις για το σύνολο των παραδειγμάτων εκπαίδευσης. Το μοντέλο αυτό προσπαθεί να διορθώσει τα λάθη του προηγούμενου μοντέλου βάσης.



Εικόνα 3.13 Η μέθοδος Boosting 2<sup>ο</sup> βήμα με αλλαγμένα βάρη

(Σταματάτος Ευστάθιος σημειώσεις μαθήματος Μηχανικής Μάθησης, πανεπιστήμιο Αιγαίου[2021])

Με όμοιο τρόπο δημιουργούνται πολλά μοντέλα βάσης , με το καθένα να προσπαθεί να προσπαθεί να επιλύσει τα λάθη του προηγούμενου.



Εικόνα 3.14 Η μέθοδος Boosting 3<sup>ο</sup> βήμα ένωση των μοντέλων βάσης

(Σταματάτος Ευστάθιος σημειώσεις μαθήματος Μηχανικής Μάθησης, πανεπιστήμιο Αιγαίου[2021])

Έτσι από τον συνδυασμό αυτών των μοντέλων βάσης δημιουργείται ένα πολύ δυνατό μοντέλο βάσης (strong learner). Ένα απλό μοντέλο βάσης(weak learner) θα αποδώσει καλά σε ένα συγκεκριμένο κομμάτι του dataset και όχι σε όλο. Έτσι κάθε μοντέλο από την πλευρά του βοηθάει στην καλύτερη επίδοση του αλγορίθμου.

## XGBoost

Ο αλγόριθμος XGBoost (eXtreme Gradient Boosting) είναι ένας αλγόριθμος μηχανικής μάθησης ο οποίος βασίζεται στην τεχνική του gradient boosting. Ο XGBoost κάνει συνδυασμό του gradient boosting με κάποιες βελτιστοποιήσεις για να βελτιώσει την ακρίβεια και την επίδοση του.

Η λειτουργία του είναι:

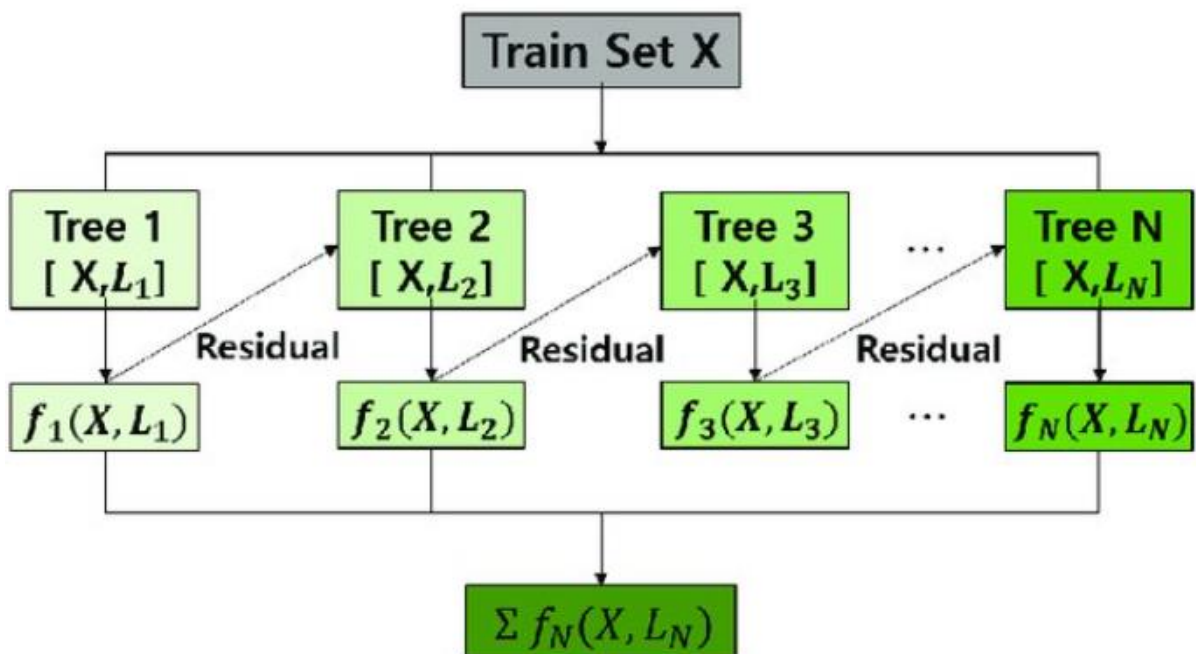
1. Ο αλγόριθμος ξεκινάει με ένα μοντέλο βάσης (base model) το οποίο είναι ένα δέντρο απόφασης.
2. Υπολογίζει την διαφορά ανάμεσα στις πραγματικές τιμές και τις τιμές τις οποίες εκτιμάει το μοντέλο βάσης. Σκοπός είναι να βρεθούν οι αδυναμίες ώστε να διορθωθούν στα επόμενα βήματα.
3. Εν συνεχεία προστίθεται ένα νέο μοντέλο βάσης (weak learner) το οποίο επικεντρώνεται στην σωστή εκτίμηση της διαφοράς.



4. Κατά την προσθήκη του νέου μοντέλου, ο XGBoost εκτελεί βελτιστοποίηση μέσω του gradient boosting του gradient descent για να βρει τις κατάλληλες τιμές των υπερπαραμέτρων του μοντέλου.

5. Η διαδικασία της προσθήκης νέων μοντέλων και της βελτιστοποίησης συνεχίζεται επαναληπτικά μέχρι να επιτευχθεί η επιθυμητή ακρίβεια ή να γίνει η εξάντληση των προκαθορισμένων επαναλήψεων.

6. Ο αλγόριθμος επαναλαμβάνει τα βήματα 3 και 4, προσθέτοντας συνεχώς νέα μοντέλα και βελτιστοποιώντας το συνολικό μοντέλο μέσω της τεχνικής του gradient boosting. Κατά τη διάρκεια αυτών των επαναλήψεων, κάθε νέο μοντέλο προσπαθεί να εστιάσει στα υπόλοιπα που παραμένουν μετά τις προηγούμενες προσθήκες.



Εικόνα 3.15 Ο XGBoost για regression

(TowardsDataScience.com)

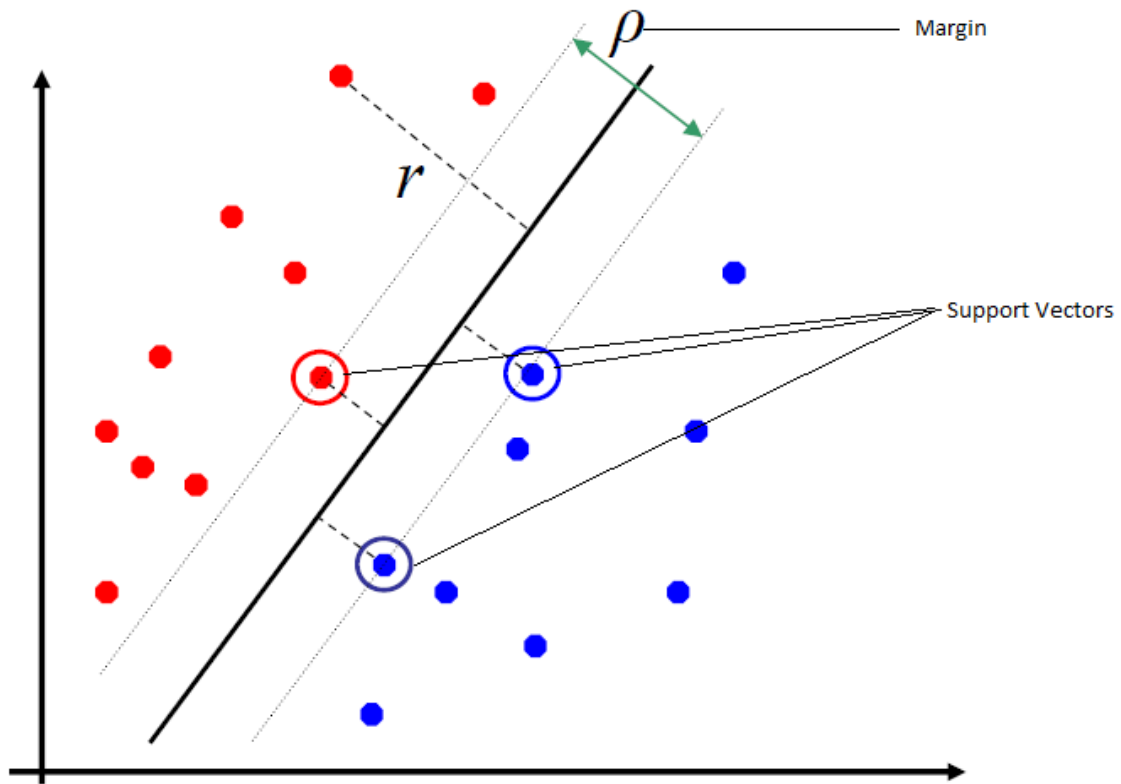
## SVM

Οι μηχανές διανυσμάτων υποστήριξης (Support Vector Machines - SVM) είναι ένας αλγόριθμος μηχανικής μάθησης που χρησιμοποιείται για Classification και Regression. Ο αλγόριθμος SVM χρησιμοποιεί ένα μοντέλο που αναπαριστά τα δεδομένα σε έναν χώρο υψηλής διαστασιμότητας και προσπαθεί να βρει ένα υπερεπίπεδο το οποίο να διαχωρίζει τις κλάσεις των δεδομένων.

Ο αλγόριθμος SVM λειτουργεί ως εξής:

1) Τα δεδομένα αρχικοποιούνται σε δύο κλάσεις με βάση τις ετικέτες τους.

2) Ο SVM επιλέγει-υπολογίζει ένα υπερεπίπεδο το οποίο διαχωρίζει όσο το δυνατόν καλύτερα τις δύο κλάσεις. Αυτό το υπερεπίπεδο υπολογίζεται με τέτοιο τρόπο ώστε η απόσταση ( $p = \text{margin}$ ) ανάμεσα στο υπερεπίπεδο και στα πλησιέστερα δείγματα κάθε κλάσης να είναι μέγιστη.



Εικόνα 3.16 Ο SVM φαίνονται κυκλωμένα τα SV και το margin

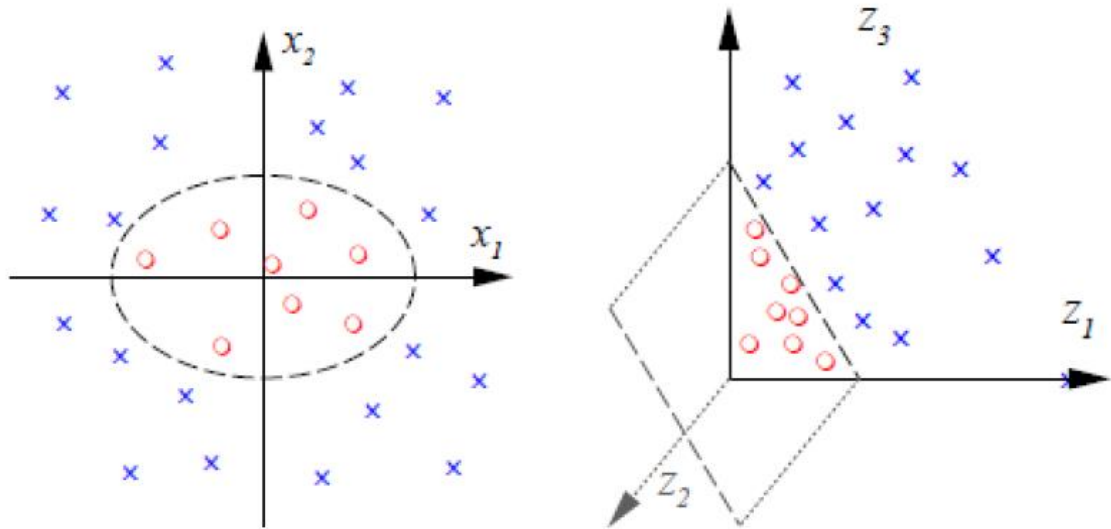
(GeekforGeeks.com)

3) Τα δείγματα που βρίσκονται στα πλησιέστερα σημεία κάθε κλάσης στο υπερεπίπεδο, είναι τα διανύσματα υποστήριξης.

Ο αλγόριθμος SVM μπορεί να χρησιμοποιήσει και μη γραμμικές συναρτήσεις (kernel) για να μπορέσει να ανταπεξέλθει σε μη γραμμικά διαχωρίσιμα δεδομένα. Τα δεδομένα αναπαρίστανται σε έναν χώρο υψηλότερης διάστασης, όπου σε αυτό τον χώρο μπορούν να είναι γραμμικά διαχωρίσιμα. Στην εικόνα φαίνεται παράδειγμα όπου τα δεδομένα δεν είναι διαχωρίσιμα στις διαστάσεις  $x_1, x_2$  αλλά στις διαστάσεις  $Z_1, Z_2, Z_3$ . Οπότε αυξάνοντας μια διάσταση τα δεδομένα είναι διαχωρίσιμα γραμμικά.

$$\Phi : \mathbb{R}^2 \rightarrow \mathbb{R}^3$$

$$(x_1, x_2) \mapsto (z_1, z_2, z_3) := (x_1^2, \sqrt{2}x_1x_2, x_2^2)$$



Εικόνα 3.17 Ο διαχωρισμός με χρήση kernel σε έναν χώρο μεγαλύτερης διάστασης

(Σταματάτος Ευστάθιος σημειώσεις μαθήματος Μηχανικής Μάθησης, πανεπιστήμιο Αιγαίου[2021])

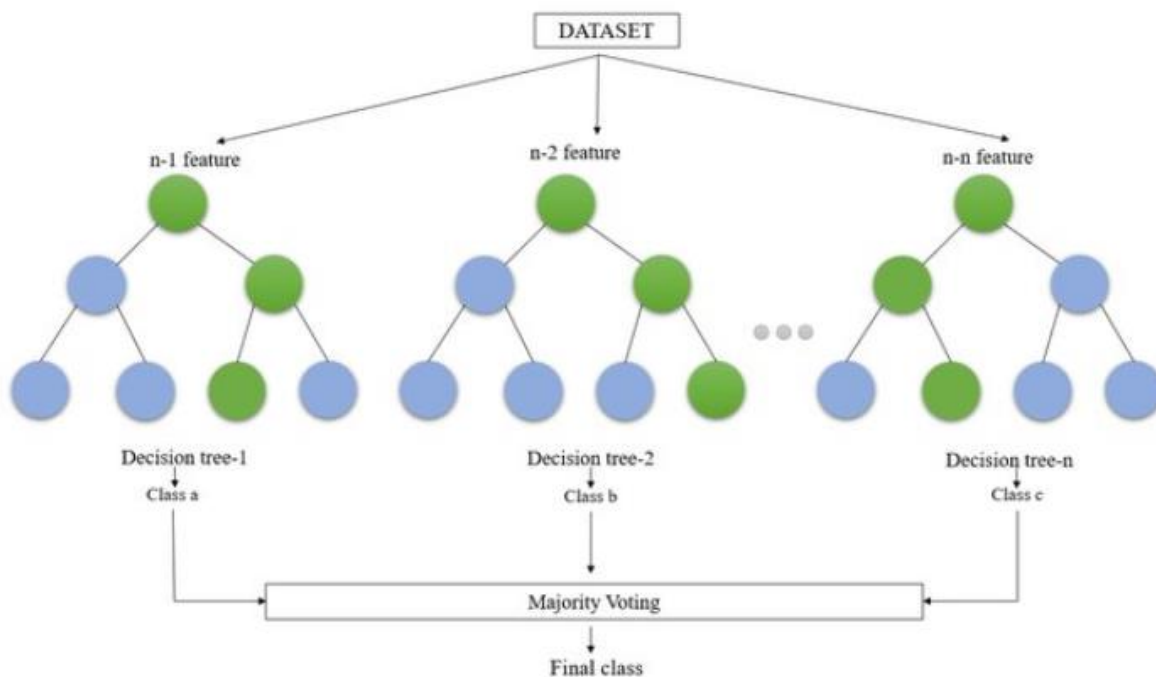
Το υπερεπίπεδο διαχωρισμού σε χώρο  $\mathbb{R}^n$  είναι  $n-1$  διαστάσεις. Οπότε στο παραπάνω παράδειγμα όπου ο χώρος είναι 3 διαστάσεων το υπερεπίπεδο διαχωρισμού θα είναι 2 διαστάσεων.

## Random forest

Ο Random Forest είναι ένα σύνολο δέντρων αποφάσεων τα οποία συνδυάζονται μεταξύ τους για να πραγματοποιήσουν Classification ή Regression σε δεδομένα.

Κατά τη φάση της εκπαίδευσης, δημιουργούνται πολλά δέντρα αποφάσεων όπου κάθε δέντρο εκπαιδεύεται ανεξάρτητα χρησιμοποιώντας ένα τυχαίο υποσύνολο των παραδειγμάτων εκπαίδευσης και ένα τυχαίο υποσύνολο των χαρακτηριστικών(Features) του Dataset. Κατά την εκπαίδευση, τα δέντρα προσπαθούν να διαχωρίσουν τα δεδομένα σε διάφορες κατηγορίες ή να προβλέψουν μια συνεχή τιμή.

Συνδυασμός: Όταν όλα τα δέντρα απόφασης εκπαιδευτούν, ο Random Forest συνδυάζει τις αποφάσεις των δέντρων για να παράξει το τελικό αποτέλεσμα. Στο classification, χρησιμοποιείται κυρίως το majority voting για να αποφασιστεί η κατηγοριοποίηση του δείγματος. Στο regression, ο μέσος όρος των προβλέψεων των δέντρων χρησιμοποιείται ως τελική πρόβλεψη.



Εικόνα 3.18 Η αρχιτεκτονική του Random Forest

(Σταματάτος Ευστάθιος σημειώσεις μαθήματος Μηχανικής Μάθησης, πανεπιστήμιο Αιγαίου[2021])

Ο αλγόριθμος Random Forest έχει τα εξής χαρακτηριστικά και πλεονεκτήματα:

→Χρησιμοποιείται για classification και regression και μπορεί να χειριστεί και μεγάλα σύνολα δεδομένων.

→Αντιμετωπίζει σε μεγάλο βαθμό το πρόβλημα του overfitting, καθώς οι αποφάσεις των δέντρων του περιορίζονται από την τυχαιότητα της επιλογής των υποσυνόλων δεδομένων και χαρακτηριστικών τα οποία χρησιμοποιούνται κατά την εκπαίδευση.

→Οι Random Forest έχουν ευρεία εφαρμογή σε πολλούς τομείς, όπως η αναγνώριση προτύπων, η ανάλυση εικόνων, η βιοπληροφορική, και η αναγνώριση φωνής.

## 4.1 To Dataset

Όπως αναφέρθηκε το Dataset αποτελείται από καταγραφές διαφόρων μετρήσεων για 100 ίδιου τύπου μηχανές αεροσκαφών. Το Dataset ονομάζεται FD001. Το Dataset αποτελείται από μετρήσεις 21 αισθητήρων και αντιστοιχούν σε μια μέτρηση για κάθε πτήση που πραγματοποιεί η μηχανή. Επίσης έχει τα 5 πρώτα features τα οποία είναι :

1. Ο αριθμός(ταυτότητα) της μηχανής από το 1 έως το 100.

2. Οι κύκλοι που αναφέρονται στο νούμερο 2 είναι οι πτήσεις ουσιαστικά που έκανε οι μηχανή μέχρι να αστοχήσει π.χ η μηχανή 1 αστόχησε στους 192 κύκλους. Οπότε στο Dataset υπάρχουν 192 instances για την μηχανή 1. Για κάθε κύκλο παίρνω και μια καταγραφή για όλα τα Features τα οποία περιγράφονται στην εικόνα 4.1 .

3. Operational settings είναι καταστάσεις πτήσεις οι οποίες δεν επηρεάζουν την διαδικασία.

Column No.	Contents
1	Engine Unit No.
2	Time in cycles
3	Operational Setting 1
4	Operational Setting 2
5	Operational Setting 3
6	Sensor Measurement 1
⋮	⋮
26	Sensor Measurement 21

Εικόνα 4.1 To Dataset

(NASA.gov)

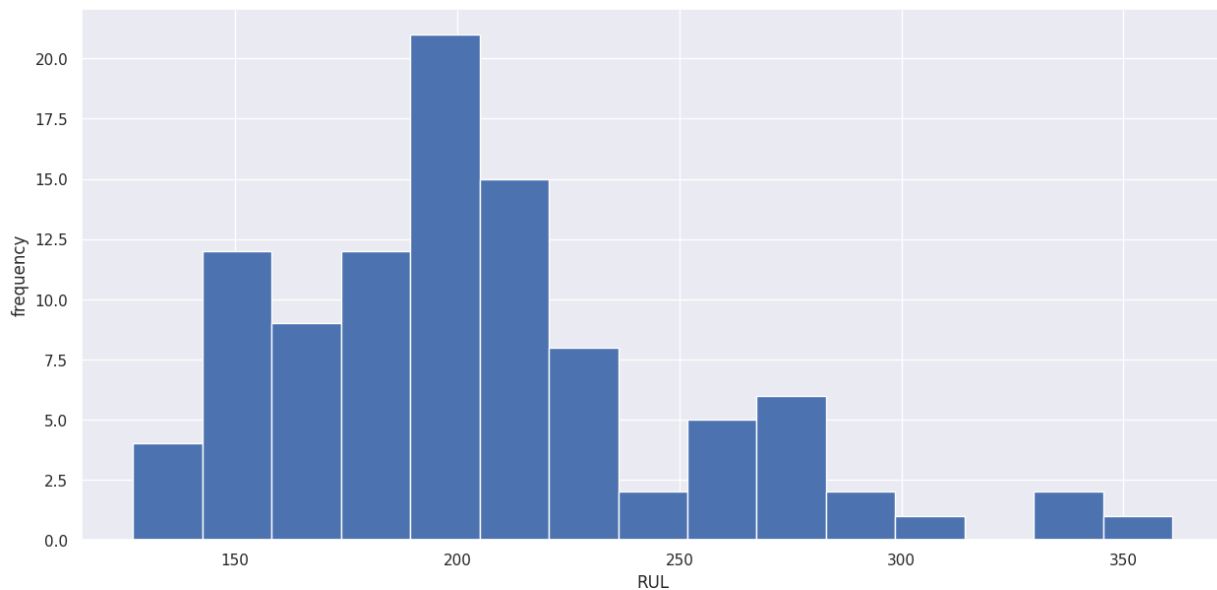
Τα υπόλοιπα 21 Features όπως φαίνεται και στην εικόνα 4.1 είναι μετρήσεις αισθητήρων. Στην εικόνα 4.2 παρουσιάζεται το training Dataset και η κατανομή των αστοχιών των μηχανών του Dataset. Το Dataset αποτελείται από 100 μηχανές ο μέσος όρος αστοχίας των οποίων είναι στους 206 κύκλους κόπωσης, ο ελάχιστος κύκλος στον

οποίο αστόχησε μια μηχανή είναι ο 128<sup>ος</sup> κύκλος κόπωσης ενώ ο μέγιστος κύκλος είναι ο 362<sup>ος</sup> κύκλος κόπωσης.

time_cycles	
count	100.000000
mean	206.310000
std	46.342749
min	128.000000
25%	177.000000
50%	199.000000
75%	229.250000
max	362.000000

Εικόνα 4.2 Στατιστικά μεγέθη για το RUL της μηχανής

Στην εικόνα 4.3 φαίνεται η κατανομή αστοχιών των μηχανών για το training Dataset. Επιβεβαιώνεται αυτό το οποίο αναφέρθηκε και στην προηγούμενη εικόνα, δηλαδή ότι ο μέσος όρος αστοχίας των μηχανών λόγω κόπωσης είναι 206<sup>ος</sup> κύκλος.



Εικόνα 4.3 Κατανομή του RUL στις 100 μηχανές του Dataset



## 4.2 Μεθοδολογία

### **Βήμα 1<sup>ο</sup>** Δημιουργία του σωστού Dataset .

Το Dataset χωρίζεται σε 3 μέρη (subdatasets). Το Training subdataset με 20631 instances το οποίο δίνεται ολόκληρο από το Site της NASA ,ενώ το Test set δίνεται σε ένα subdataset με 13095 instances για τις 100 μηχανές. Για το Test set απαιτείται να βρεθεί η εκτίμηση της εναπομένουσας ζωής της μηχανής σε κύκλους κόπωσης(RUL). Ενώ το τρίτο subdataset το οποίο δίνει τις πραγματικές τιμές του RUL οι οποίες θα πρέπει να εκτιμηθούν αποτελείται από 100 τιμές “RUL” για την εναπομένουσα ζωή της κάθε μηχανής.

RUL=Remaining Useful Life

Για το Test Set όπως αναφέρθηκε θα πρέπει να γίνει εκτίμηση του RUL για την κάθε μηχανή , όμως για την κάθε μηχανή ζητείται η πρόβλεψη του RUL σε έναν τυχαίο κύκλο του Test set. Δηλαδή για τις 100 μηχανές υπάρχουν καταγραφές μέχρι έναν τυχαίο κύκλο κόπωσης και από αυτόν τον τυχαίο κύκλο απαιτείται να γίνει η εκτίμηση του RUL. Για παράδειγμα για την μηχανή 3 ζητείται να γίνει η εκτίμηση από τον κύκλο 126 και για την μηχανή 4 από τον κύκλο 106. Οι εικόνες 4.4 και 4.5 δείχνουν αναλυτικά τα παραπάνω για το Test Set.

201 to 225 of 13096 entries Filter  ?

index	unit_nr	time_cycles	setting_1	setting_2	setting_3	s_1	s_2	s_3	s_4	s_5	s_6	s_7	s_8	s_9	s_10	s_11	s_12	s_13	s_14	s_15
200	3	121	-0.0004	-0.0001	100.0	518.67	643.05	1595.36	1412.36	14.62	21.61	553.33	2388.18	9044.93	1.3	47.64	520.87	2388.15	8125.93	8.4892
201	3	122	0.0022	-0.0005	100.0	518.67	642.65	1597.83	1412.76	14.62	21.61	552.8	2388.15	9044.98	1.3	47.75	521.25	2388.2	8129.3	8.4979
202	3	123	-0.0057	0.0002	100.0	518.67	642.54	1596.49	1409.72	14.62	21.61	552.72	2388.2	9046.83	1.3	47.67	520.53	2388.16	8125.94	8.4677
203	3	124	-0.0011	0.0003	100.0	518.67	643.06	1590.49	1412.78	14.62	21.61	552.3	2388.13	9047.84	1.3	47.61	520.87	2388.16	8133.61	8.4551
204	3	125	0.0014	0.0002	100.0	518.67	643.24	1588.64	1412.6	14.62	21.61	553.1	2388.19	9048.12	1.3	47.62	520.13	2388.21	8128.02	8.4608
205	3	126	-0.0016	0.0004	100.0	518.67	642.88	1589.75	1418.89	14.62	21.61	552.59	2388.16	9049.26	1.3	47.88	520.83	2388.14	8131.46	8.4119
206	4	1	-0.0024	0.0	100.0	518.67	642.63	1589.43	1399.93	14.62	21.61	553.1	2388.05	9048.52	1.3	47.44	521.78	2388.15	8131.61	8.4195
207	4	2	-0.0002	0.0005	100.0	518.67	642.36	1591.83	1406.26	14.62	21.61	553.61	2388.11	9047.34	1.3	47.4	521.7	2388.1	8132.51	8.4099
208	4	3	0.0021	0.0	100.0	518.67	642.58	1590.84	1409.67	14.62	21.61	554.28	2388.08	9051.81	1.3	47.41	521.89	2388.07	8137.19	8.4253
209	4	4	-0.0005	-0.0001	100.0	518.67	642.59	1589.02	1402.18	14.62	21.61	553.57	2388.11	9049.72	1.3	47.56	521.2	2388.08	8128.1	8.4205
210	4	5	0.0003	0.0004	100.0	518.67	642.29	1584.92	1410.31	14.62	21.61	553.42	2388.07	9052.16	1.3	47.35	521.42	2388.11	8135.37	8.4306
211	4	6	0.0008	0.0001	100.0	518.67	642.29	1588.37	1410.51	14.62	21.61	552.76	2388.04	9052.35	1.3	47.32	521.43	2388.11	8128.33	8.4415
212	4	7	0.0032	-0.0003	100.0	518.67	642.1	1586.93	1411.36	14.62	21.61	553.78	2388.09	9052.59	1.3	47.33	521.7	2388.09	8134.09	8.4173
213	4	8	0.0009	-0.0002	100.0	518.67	642.65	1586.81	1410.31	14.62	21.61	553.9	2388.1	9048.97	1.3	47.59	521.75	2388.09	8136.4	8.4265
214	4	9	0.0005	-0.0004	100.0	518.67	643.05	1585.09	1408.07	14.62	21.61	553.71	2388.06	9056.65	1.3	47.31	521.54	2388.09	8134.33	8.4267
215	4	10	-0.003	0.0003	100.0	518.67	642.55	1586.65	1411.38	14.62	21.61	554.13	2388.1	9052.12	1.3	47.44	522.37	2388.09	8135.74	8.4311
216	4	11	0.0004	0.0005	100.0	518.67	641.96	1585.37	1406.97	14.62	21.61	553.07	2388.08	9057.89	1.3	47.42	521.55	2388.12	8128.5	8.4367
217	4	12	-0.0018	0.0002	100.0	518.67	642.09	1594.56	1409.7	14.62	21.61	554.18	2388.08	9051.16	1.3	47.41	521.65	2388.13	8131.14	8.4731
218	4	13	-0.002	-0.0003	100.0	518.67	642.29	1588.32	1409.92	14.62	21.61	553.54	2388.09	9050.63	1.3	47.2	521.66	2388.11	8134.16	8.4663
219	4	14	-0.003	0.0004	100.0	518.67	642.22	1587.11	1400.63	14.62	21.61	553.81	2388.08	9052.24	1.3	47.54	521.43	2388.11	8130.31	8.4328
220	4	15	0.0012	-0.0001	100.0	518.67	641.84	1590.22	1404.72	14.62	21.61	553.61	2388.08	9060.56	1.3	47.7	521.75	2388.12	8134.68	8.4208
221	4	16	0.0	0.0003	100.0	518.67	642.21	1587.79	1414.51	14.62	21.61	553.92	2388.14	9053.25	1.3	47.36	521.48	2388.14	8136.03	8.4347
222	4	17	-0.0031	-0.0003	100.0	518.67	642.87	1593.3	1402.81	14.62	21.61	553.48	2388.09	9049.24	1.3	47.66	521.63	2388.12	8131.01	8.4409
223	4	18	-0.0017	-0.0	100.0	518.67	642.28	1581.83	1400.02	14.62	21.61	553.45	2388.04	9058.87	1.3	47.41	521.84	2388.08	8131.3	8.3925
224	4	19	-0.0005	0.0	100.0	518.67	642.52	1589.28	1411.29	14.62	21.61	553.19	2388.05	9053.66	1.3	47.54	521.66	2388.11	8134.19	8.4298

Εικόνα 4.4 Απεικόνιση του Dataset σχετικά με τις καταγραφές των μηχανών-μηχανή 3&4

index	unit_nr	time_cycles	setting_1	setting_2	setting_3	s_1	s_2	s_3	s_4	s_5	s_6	s_7	s_8	s_9	s_10	s_11	s_12	s_13	s_14	s_15
300	4	95	0.0017	0.0005	100.0	518.67	642.42	1595.14	1412.25	14.62	21.61	553.18	2388.18	9050.98	1.3	47.56	521.59	2388.14	8126.85	8.4347
301	4	96	0.0031	0.0001	100.0	518.67	642.67	1594.38	1412.05	14.62	21.61	552.65	2388.1	9051.65	1.3	47.64	521.67	2388.14	8133.61	8.4464
302	4	97	-0.0011	-0.0004	100.0	518.67	643.0	1591.35	1410.58	14.62	21.61	553.29	2388.17	9047.67	1.3	47.66	521.37	2388.14	8132.44	8.4303
303	4	98	-0.0017	0.0	100.0	518.67	642.5	1590.32	1415.04	14.62	21.61	553.56	2388.14	9054.42	1.3	47.56	521.01	2388.13	8137.12	8.4575
304	4	99	0.0025	0.0001	100.0	518.67	642.92	1585.88	1407.76	14.62	21.61	553.31	2388.16	9062.04	1.3	47.7	521.22	2388.09	8127.77	8.4449
305	4	100	-0.0007	0.0004	100.0	518.67	642.85	1592.05	1410.14	14.62	21.61	552.86	2388.15	9061.95	1.3	47.58	521.44	2388.12	8131.19	8.4539
306	4	101	0.0013	0.0005	100.0	518.67	643.18	1594.27	1416.02	14.62	21.61	553.2	2388.16	9055.56	1.3	47.59	521.03	2388.08	8136.59	8.4632
307	4	102	-0.0014	-0.0004	100.0	518.67	642.69	1588.87	1418.38	14.62	21.61	553.93	2388.03	9056.92	1.3	47.61	521.65	2388.07	8134.15	8.4156
308	4	103	0.0027	-0.0	100.0	518.67	642.93	1594.26	1414.05	14.62	21.61	553.07	2388.09	9049.92	1.3	47.58	521.46	2388.13	8134.72	8.4657
309	4	104	-0.0012	-0.0002	100.0	518.67	642.85	1589.92	1406.16	14.62	21.61	553.53	2388.15	9054.6	1.3	47.68	521.24	2388.14	8133.94	8.4646
310	4	105	-0.0028	-0.0003	100.0	518.67	642.68	1593.01	1408.21	14.62	21.61	554.05	2388.18	9061.26	1.3	47.69	520.94	2388.12	8135.19	8.4498
311	4	106	0.0012	0.0004	100.0	518.67	642.78	1594.53	1406.88	14.62	21.61	552.64	2388.13	9051.3	1.3	47.65	521.88	2388.11	8133.64	8.4634
312	5	1	0.0024	-0.0004	100.0	518.67	642.43	1590.79	1402.02	14.62	21.61	553.5	2388.07	9051.54	1.3	47.37	522.2	2388.07	8130.9	8.4426
313	5	2	0.0011	0.0002	100.0	518.67	642.85	1581.46	1409.7	14.62	21.61	553.44	2388.06	9051.9	1.3	47.54	521.63	2388.11	8128.71	8.4429
314	5	3	0.0056	-0.0	100.0	518.67	642.88	1586.05	1407.64	14.62	21.61	553.51	2388.06	9057.29	1.3	47.27	521.8	2388.11	8132.02	8.4657
315	5	4	-0.001	0.0003	100.0	518.67	642.79	1585.93	1403.47	14.62	21.61	553.93	2388.15	9048.64	1.3	47.43	521.8	2388.09	8130.1	8.4147
316	5	5	0.0044	0.0002	100.0	518.67	642.66	1583.4	1407.29	14.62	21.61	553.83	2388.12	9053.72	1.3	47.48	521.94	2388.01	8132.74	8.4201
317	5	6	0.0018	0.0003	100.0	518.67	642.72	1581.43	1410.88	14.62	21.61	552.84	2388.05	9048.96	1.3	47.49	521.33	2388.07	8130.32	8.4301
318	5	7	0.0061	0.0004	100.0	518.67	642.44	1592.72	1406.89	14.62	21.61	553.39	2388.1	9048.84	1.3	47.5	521.3	2388.17	8133.31	8.4494
319	5	8	-0.0009	0.0002	100.0	518.67	643.14	1590.58	1403.19	14.62	21.61	553.86	2388.09	9058.21	1.3	47.48	521.46	2388.11	8131.46	8.4504
320	5	9	0.0024	-0.0002	100.0	518.67	642.5	1586.98	1409.36	14.62	21.61	553.83	2388.12	9053.72	1.3	47.41	521.29	2388.09	8139.72	8.4558
321	5	10	-0.0009	-0.0002	100.0	518.67	643.17	1595.37	1409.32	14.62	21.61	553.16	2388.11	9057.49	1.3	47.61	521.31	2388.11	8138.55	8.4885
322	5	11	0.0015	-0.0003	100.0	518.67	643.23	1583.73	1410.89	14.62	21.61	553.24	2388.08	9046.78	1.3	47.64	521.53	2388.1	8127.61	8.3709
323	5	12	0.0003	0.0004	100.0	518.67	642.25	1592.65	1407.6	14.62	21.61	553.16	2388.06	9052.91	1.3	47.27	521.61	2388.07	8130.29	8.424
324	5	13	-0.0007	-0.0001	100.0	518.67	643.22	1586.89	1399.86	14.62	21.61	554.02	2388.1	9054.08	1.3	47.44	521.82	2388.1	8135.38	8.4294

Εικόνα 4.5 Απεικόνιση του test subdataset σχετικά με τις καταγραφές των μηχανών 4 και 5

Όπως φαίνεται και στις εικόνες 4.4 και 4.5 η μηχανή 3 τελειώνει στον 126ο κύκλο κόπωσης και ζητείται να γίνει εκτίμηση του RUL, ενώ στην κάτω σειρά ξεκινάει η μηχανή 4. Από την στιγμή που το Test Dataset περιέχει 13095 instances θα πρέπει να διαμορφωθεί κατάλληλα προγραμματιστικά ώστε από κάθε μηχανή να κρατηθεί μόνο το τελευταίο instance δηλαδή αυτό για το οποίο καλείται ο αλγόριθμος να κάνει εκτίμηση, καθώς όπως αναφέρθηκε το τρίτο subdataset με το πραγματικό RUL για το οποίο καλούμαστε να εκτιμήσουμε δίνει 100 τιμές, δηλαδή μια για κάθε μηχανή.

**Βήμα 2<sup>ο</sup>** Για να αποφευχθεί το φαινόμενο του Overfitting και να πιστοποιηθεί η εγκυρότητα του μοντέλου θα χρησιμοποιηθεί , ένα κομμάτι του training set για Validation set. Συγκεκριμένα οι μηχανές από νούμερο 85 ως και νούμερο 100 θα δημιουργήσουν το καινούργιο dataset το οποίο θα χρησιμοποιηθεί ως Validation Set. Όλες οι δοκιμές και το tuning θα υλοποιηθούν πάνω σε αυτό το dataset και τα αποτελέσματα των βέλτιστων λύσεων( για τα Hyperparameters) οι οποίες θα προκύψουν για τον καθένα από τους 3 αλγορίθμους στην συνέχεια θα χρησιμοποιηθούν στο τελικό test set. Οπότε ως training set θα χρησιμοποιηθεί το αρχικό Dataset από την μηχανή 1 ως την μηχανή 84 και από την μηχανή 85 έως την μηχανή 100 για Validation Set.

### **Βήμα 3<sup>ο</sup>** Υπολογισμός RUL

Σκοπός είναι να εκτιμηθεί η εναπομένουσα ζωή μιας μηχανής (RUL), όμως στο training dataset όπως αυτό παρουσιάζεται και στην εικόνα 4.1 δεν δίνεται για κάθε instance το αντίστοιχο RUL ώστε να είναι δυνατή η εκπαίδευση του αλγορίθμου. Οπότε αυτό θα πρέπει να υπολογιστεί-δημιουργηθεί ώστε να μπορεί να υλοποιηθεί η εκπαίδευση. Η διαδικασία έχει ως εξής:

- Επιλέγεται ο κύκλος φόρτισης στον οποίο αστόχησε η κάθε μηχανή, δηλαδή ο μέγιστος κύκλος φόρτισης.
- Από τον μέγιστο κύκλο που έφτασε η κάθε μηχανή πριν αστοχήσει αφαιρείται ο αριθμός του κάθε κύκλου κόπωσης(instance) για την μηχανή αυτή.

Δηλαδή αν η μηχανή 3 αστόχησε στον 179<sup>ο</sup> κύκλο και αυτό γίνεται αντιληπτό καθώς στο dataset η τελευταία καταγραφή είναι η 179<sup>η</sup> για την μηχανή, τότε για τον υπολογισμό του RUL στο training set θα πρέπει να δημιουργηθεί μια νέα στήλη από την οποία για κάθε instance μιας μηχανής (π.χ. για την μηχανή 3) θα αφαιρείται από το 179 ο αριθμός του instance. Οπότε π.χ. μετά την 8<sup>η</sup> πτήση της 3<sup>ης</sup> μηχανής το RUL της θα είναι:

$$179-8=171$$

Όμοια για όλα τα instances μέχρι που στο τελευταίο instance(το 179<sup>ο</sup>) θα είναι :

$$179-179=0$$

Οπότε και η μηχανή θα αστοχήσει και στο Dataset ξεκινάνε οι καταγραφές για την επόμενη μηχανή.

**Βήμα 4<sup>ο</sup>** Το degradation του εξοπλισμού δεν είναι αναλογικό.

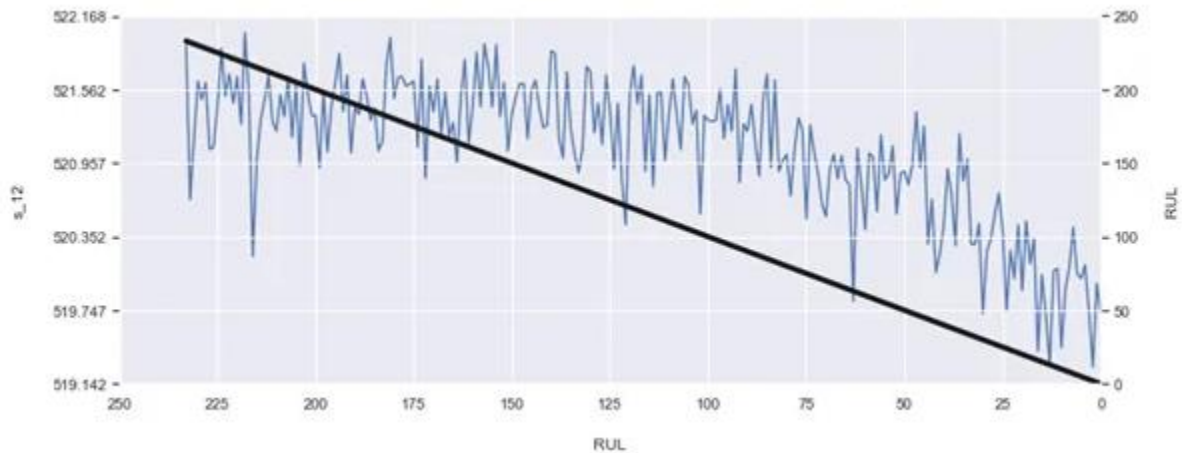
Σύμφωνα με την μεθοδολογία η οποία αναπτύχθηκε στο βήμα 1<sup>ο</sup> γίνεται αντιληπτό ότι υπάρχει μια γραμμική συσχέτιση μεταξύ αστοχίας και κύκλων κόπωσης. Δηλαδή για κάθε κύκλο κόπωσης η εναπομένουσα ζωή της εκάστοτε μηχανής μειώνεται αναλογικά, κάτι το οποίο δεν ισχύει στην πράξη αλλά δεν ισχύει και με βάση την καθημερινή μας εμπειρία. Δηλαδή για έναν εξοπλισμό δεν αρχίζει να φθίνει η μηχανική του αξιοπιστία, από την πρώτη μέρα που θέτεται σε λειτουργία. Για αρκετό καιρό στο πρώτο διάστημα χρήσης του μπορεί να λειτουργεί σαν καινούργιος και από ένα σημείο και μετά αρχίζει να φθίνει η αξιοπιστία του. Το ίδιο ισχύει και για έναν Turbofan κινητήρα. Αυτή η προσέγγιση η οποία πάρθηκε από μελέτη άλλων πρόσφατων paper[29] [30] φαίνεται ότι βελτίωσε κατά πολύ τα αποτελέσματα, των αλγορίθμων μηχανικής μάθησης οι οποίοι εφαρμόστηκαν στο Dataset. Η φιλοσοφία όπως αναφέρθηκε είναι ότι για έναν εξοπλισμό ή αξιοπιστία του δεν φθίνει γραμμικά.

Δηλαδή στην πραγματικότητα η φθορά του εξοπλισμού στις μετρήσεις οι οποίες καταγράφονται αρχίζει και έχει αντίκτυπο στον εξοπλισμό από έναν αριθμό πτήσεων (κύκλων κόπωσης) και μετά. Στην αρχή ο εξοπλισμός λειτουργεί σαν καινούργιος, οπότε στην πραγματικότητα θα πρέπει να λαμβάνονται υπόψιν οι μετρήσεις από έναν κύκλο και μετά. Αν ο αλγόριθμος λαμβάνει υπόψιν μετρήσεις για τα πρώτα instances της κάθε μηχανής, ουσιαστικά είναι σαν να εισάγει Bias στο σύστημα. Για το συγκεκριμένο Dataset μετά από δοκιμές επιλέχθηκε ο 110<sup>ος</sup> κύκλος κόπωσης. Δηλαδή αν η μηχανή αστόχησε στον 192<sup>ο</sup> κύκλο :

$$192-110=82$$

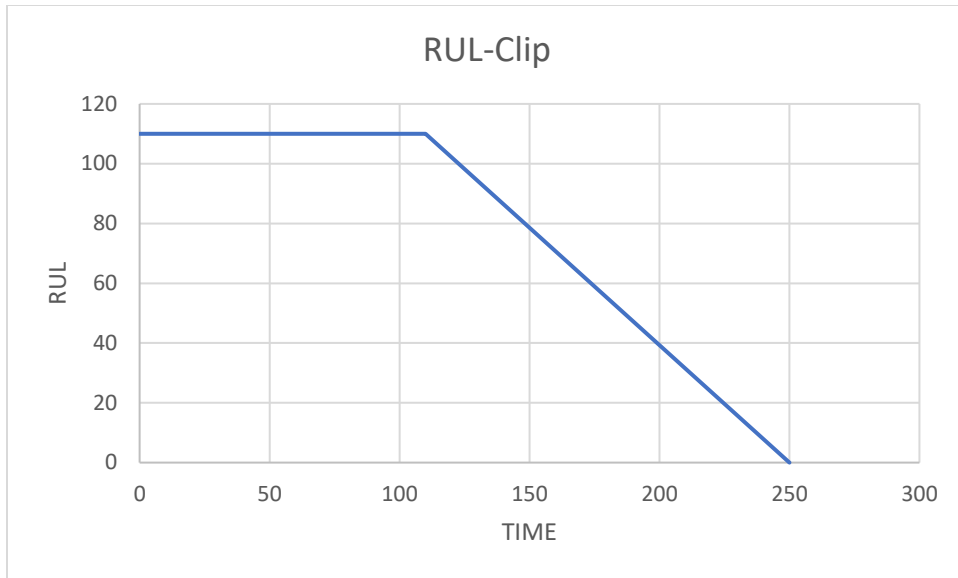
Άρα για την εκπαίδευση του αλγορίθμου κρατούνται τα δεδομένα εκπαίδευσης από τον 82<sup>ο</sup> κύκλο και μετά. Δηλαδή μέχρι το 82<sup>ο</sup> instance της συγκεκριμένης μηχανής θεωρείται ότι το RUL είναι 110. Στην εικόνα 4.6 η μαύρη γραμμή απεικονίζει το RUL μιας μηχανής μέχρι αυτή να αστοχήσει με γραμμικό μοντέλο. Στην εικόνα φαίνεται πως θεωρήθηκε το γραμμικό μοντέλο [RUL-INSTANCE→βήμα 2<sup>ο</sup>] για την φθορά του εξοπλισμού από τον πρώτο κύκλο πτήσης μιας μηχανής, ενώ με μπλε χρώμα φαίνεται η καταγραφή ενός αισθητήρα. Φαίνεται ότι η καταγραφή του αισθητήρα ξεκινάει να ακολουθεί(να συσχετίζεται) την γραμμή του RUL μετά

τον 110<sup>ο</sup> κύκλο κόπωσης. Η καταγραφή του αισθητήρα στην εικόνα 4.6 φαίνεται να ομοιάζει και με την καμπύλη αστοχίας της εικόνας 1.3 και 1.4



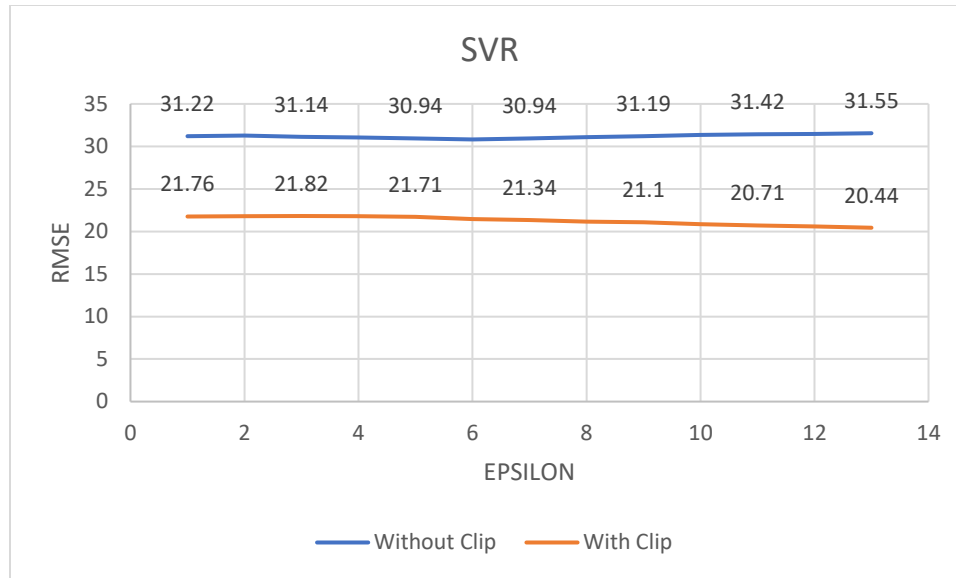
Εικόνα 4.6 Το αναλογικό μοντέλο μείωσης του RUL

Στην εικόνα 4.7 φαίνεται η στρατηγική η οποία ακολουθήθηκε. Ουσιαστικά η καταγραφή των αισθητήρων μέχρι και το  $RUL=110$  εισάγει θόρυβο στο σύστημα, καθώς δεν υπάρχει ουσιαστική μεταβολή στην καταγραφή των αισθητήρων με την προηγούμενη μέθοδο (αναλογικό RUL), ο αλγόριθμος αντιλαμβάνεται ότι το RUL μειώνεται αναλογικά οπότε είναι σαν να γίνεται εισαγωγή θορύβου στο σύστημα καθώς δεν υπάρχει πραγματική συσχέτιση των καταγραφών των αισθητήρων με την μείωση του RUL.

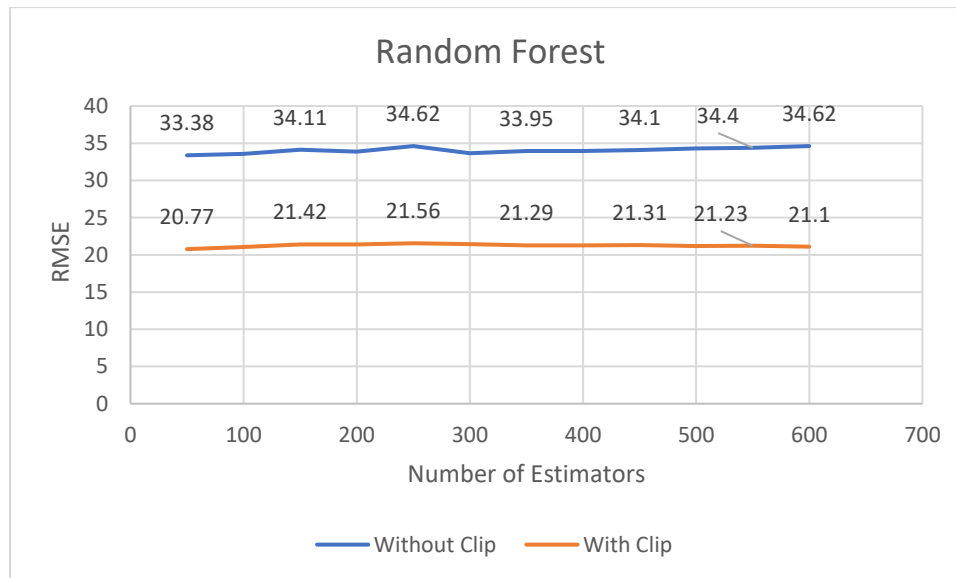


Εικόνα 4.7 Το εφαρμοσμένο μοντέλο μείωσης του RUL(για RUL=125)

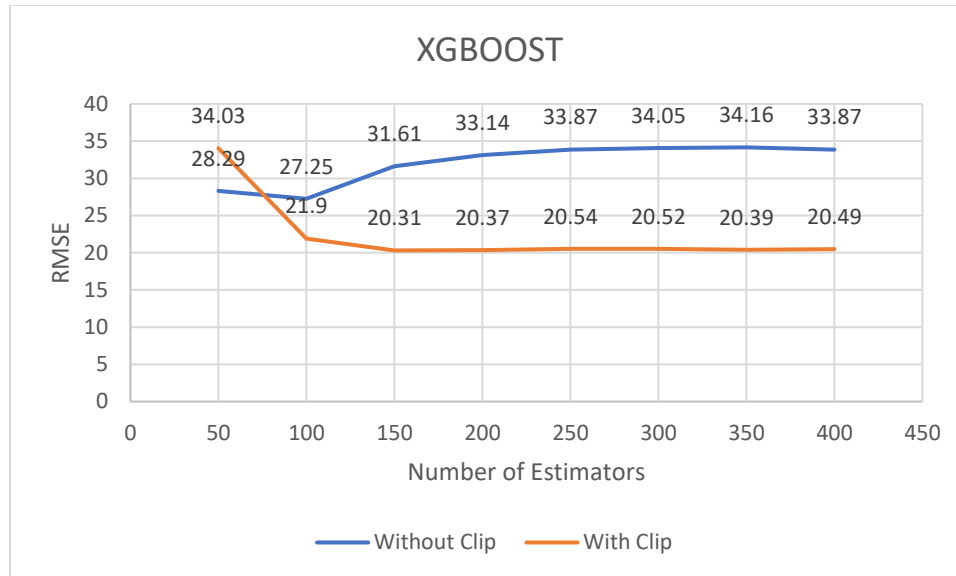
Στις εικόνες 4.8 , 4.9 και 4.10 φαίνεται η επίδοση των 3 αλγορίθμων οι οποίοι εφαρμόστηκαν στο Dataset (Validation Set) χρησιμοποιώντας την μέθοδο Clip(Πορτοκαλί γραμμή) και χωρίς χρήση της μεθόδου Clip (Μπλε γραμμή). Μέθοδος Clip ονομάζεται η διαδικασία κατά την οποία λαμβάνεται υπόψιν από το RUL= 110 μέχρι το RUL=0 όπως αυτή περιεγράφηκε πιο πάνω. Ονομάστηκε έτσι στην παρούσα εργασία από την ανάλογη εντολή η οποία χρησιμοποιείται στην Python. Γίνεται ξεκάθαρο ότι και στους 3 αλγορίθμους η μέθοδος δίνει μεγάλη βελτίωση στην εκτίμηση του RUL μειώνοντας σε πολύ μεγάλο βαθμό το RMSE. Να σημειωθεί ότι στα παρακάτω γραφήματα οι τιμές του RMSE οι οποίες παρουσιάζονται αφορούν το βασικό Dataset πριν γίνει εισαγωγή νέων Features η οποία γίνεται σε επόμενο βήμα. Γίνεται φανερό από τα γραφήματα ότι στον Random Forest η βελτίωση του RMSE έφτασε μέχρι και 13.52, στον SVR το 11.1 ενώ στον XGBoost το 13.77 . Είναι ξεκάθαρο ότι ακολουθώντας αυτή την νέα στρατηγική υπάρχει πολύ μεγάλη βελτίωση στην επίδοση των αποτελεσμάτων των αλγορίθμων.



Εικόνα 4.8 Η επίδοση του SVR με την μέθοδο clip και χωρίς clip



Εικόνα 4.9 Η επίδοση του Random Forest με την μέθοδο clip και χωρίς clip



Εικόνα 4.10 Η επίδοση του XGBoost με την μέθοδο clip και χωρίς clip

#### Βήμα 4<sup>ο</sup> Ανάλυση των features

Μια τεχνική ώστε να εξαχθούν όσο το δυνατόν περισσότερα δεδομένα από ένα Dataset είναι να βρεθεί κάποια συσχέτιση μεταξύ των Features του, η οποία να ευνοεί την επίδοση του αλγόριθμου. Μπορεί να γίνει ανάλυση των στατιστικών μεγεθών και των συσχετίσεων μεταξύ των Features του dataset για να γίνει ορατό ποιο Feature έχει μεγαλύτερες μεταβολές καθώς η εξαρτημένη μεταβλητή μεταβάλλεται. Εν προκειμένω η εξαρτημένη μεταβλητή είναι το RUL το οποίο μετά από κάθε κύκλο φόρτισης μειώνεται, οπότε θα πρέπει να διερευνηθεί η συσχέτιση των ανεξάρτητων μεταβλητών του Dataset με το RUL το οποίο είναι εξαρτημένη. Αρχικά ερευνώνται τα στατιστικά μεγέθη των ανεξάρτητων μεταβλητών του training dataset ώστε να βρεθούν οι μεταβλητές αυτές οι οποίες έχουν την μεγαλύτερη διασπορά τιμών. Αυτό υποδηλώνει ότι η μεταβολή αυτών των μεταβλητών είναι πολύ πιθανόν να συσχετιστεί με την μεταβολή είτε του RUL είτε κάποιας άλλης μεταβλητής από τα features. Έτσι αν βρεθεί συσχέτιση και η μεταβολή αυτή δεν είναι πολύ μικρή τότε μπορεί να εισέλθει ουσιαστικά πληροφορία στο σύστημα με διάφορους τρόπους:



1<sup>ος</sup> Πολλαπλασιάζοντας μεταξύ τους τις μεταβλητές.

2<sup>ος</sup> Διαιρώντας τις μεταβλητές μεταξύ τους.

3<sup>ος</sup> Κάνοντας προσθέσεις μεταξύ μεταβλητών.

4<sup>ος</sup> Κάνοντας αφαιρέσεις μεταξύ μεταβλητών.

5<sup>ος</sup> Αν μια μεταβλητή έχει μεγάλη μεταβολή της τιμής της στο χρονικό διάστημα το οποίο μελετάται προς μια κατεύθυνση, τότε θα μπορούσε να δημιουργηθεί καινούργια μεταβλητή παίρνοντας την μέγιστη τιμή στην οποία έφτασε η ανεξάρτητη μεταβλητή και αφαιρώντας από αυτήν, την τιμή της για κάθε instance. π.χ. αν μια θερμοκρασία είναι στους 500 βαθμούς στον πρώτο κύκλο κόπωσης και έχει φτάσει στους 560 μετά από 90 κύκλους τότε μπορείς να προσθέσεις νέα πληροφορία στον αλγόριθμο. Οι 60 βαθμοί διαφορά να συσχετιστούν με τους κύκλους κόπωσης.

Στην εικόνα 4.11 φαίνεται η στατιστική κατανομή όλων των Features του training dataset(μηχανή 1 έως 84), σε πρώτο επίπεδο γίνεται έρευνα των μεγεθών τα οποία έχουν μεγάλη διασπορά τιμών. Ένα δείγμα αυτών των μεγεθών είναι το std δηλαδή το standard deviation-απόκλιση μιας μεταβλητής από τον μέσο όρο της. Όταν μια μεταβλητή έχει απόκλιση δηλώνει μεταβολή της τιμής της κάτι το οποίο μπορεί να κρύβει κάποια πληροφορία, το σημαντικό όμως είναι να μελετηθεί το πως αυτή η πληροφορία συσχετίζεται με την εξαρτημένη μεταβλητή η οποία εν προκειμένω είναι το RUL.

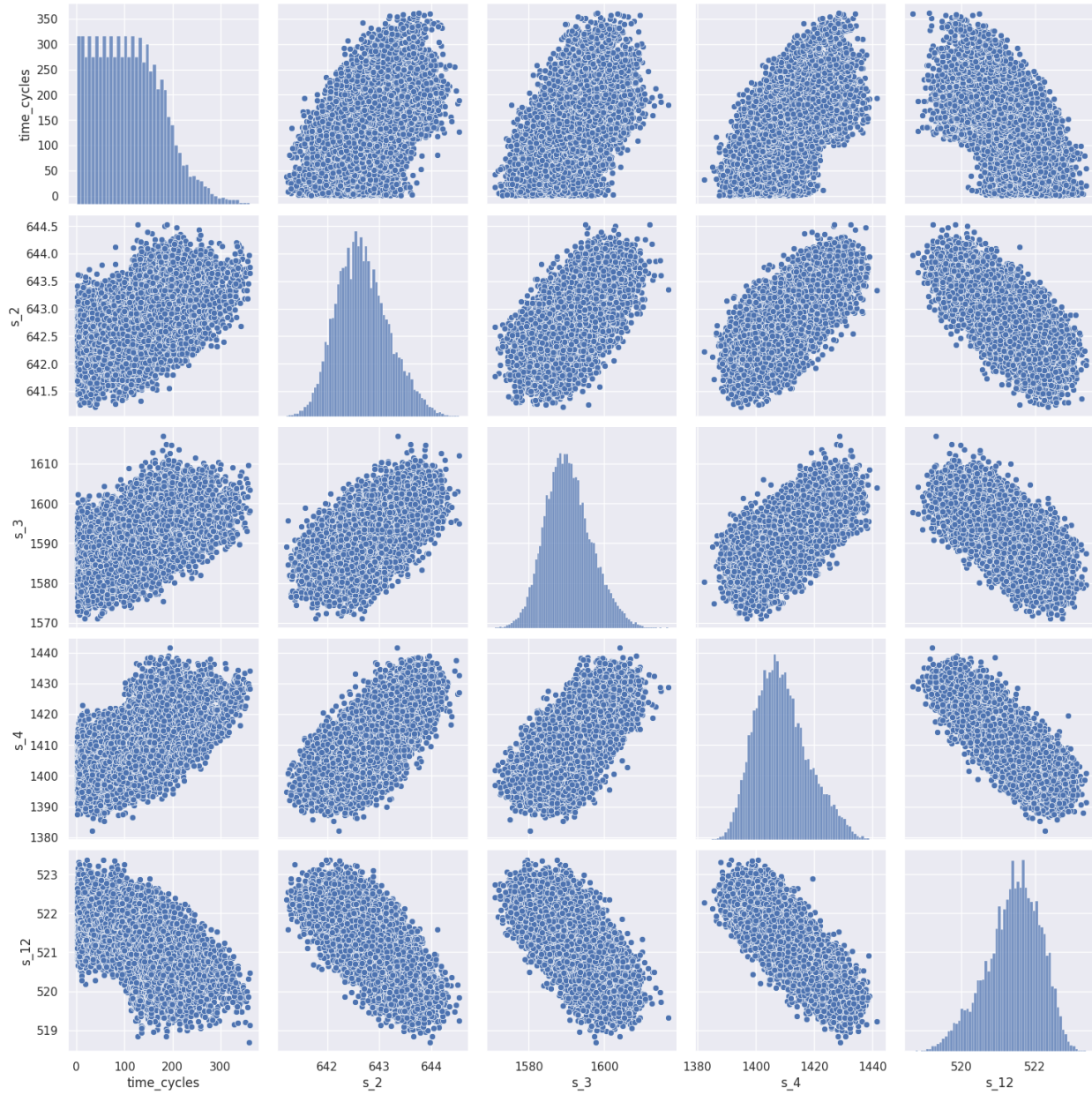
	unit_nr	time_cycles	setting_1	setting_2	setting_3	s_1	s_2	s_3	s_4	s_5	...	s_15	s_16	s_17	s_18	s_19	s_20	s_21
count	17152.000000	17152.000000	17152.000000	1.715200e+04	17152.0	17152.00	17152.000000	17152.000000	17152.000000	1.715200e+04	...	17152.000000	1.715200e+04	17152.000000	17152.0	17152.0	17152.000000	17152.000000
mean	43.183244	106.932719	0.000006	-5.247201e-08	100.0	518.67	642.684381	1590.555739	1408.975683	1.462000e+01	...	8.442349	3.000000e-02	393.218808	2388.0	100.0	38.815318	23.288984
std	24.750351	66.788082	0.002188	2.939292e-04	0.0	0.00	0.502682	6.152792	9.026965	3.552817e-15	...	0.037661	1.040864e-17	1.552754	0.0	0.0	0.181376	0.109069
min	1.000000	1.000000	-0.008700	-6.000000e-04	100.0	518.67	641.210000	1571.040000	1382.250000	1.462000e+01	...	8.324900	3.000000e-02	388.000000	2388.0	100.0	38.160000	22.907100
25%	21.000000	52.000000	-0.001400	-3.000000e-04	100.0	518.67	642.330000	1586.280000	1402.370000	1.462000e+01	...	8.415000	3.000000e-02	392.000000	2388.0	100.0	38.700000	23.220100
50%	44.000000	103.000000	-0.000000	0.000000e+00	100.0	518.67	642.650000	1590.160000	1408.180000	1.462000e+01	...	8.439600	3.000000e-02	393.000000	2388.0	100.0	38.830000	23.296300
75%	65.000000	155.000000	0.001500	3.000000e-04	100.0	518.67	643.010000	1594.440000	1414.650000	1.462000e+01	...	8.465900	3.000000e-02	394.000000	2388.0	100.0	38.940000	23.366600
max	84.000000	362.000000	0.008700	6.000000e-04	100.0	518.67	644.530000	1614.930000	1441.490000	1.462000e+01	...	8.584800	3.000000e-02	399.000000	2388.0	100.0	39.430000	23.618400

Εικόνα 4.11 Τα στατιστικά μεγέθη των Features του Dataset

Στο δεύτερο βήμα και μετά την ανάλυση της διασπορά των features πρέπει να αναλυθεί η συσχέτισή τους. Από τον πίνακα 4.11 και από τις διασπορές των features επιβεβαιώνεται ότι οι μετρήσεις των αισθητήρων S\_1, S\_5, S\_6, S\_10, S\_16, S\_18, S\_19 και τα operation setting 1,2,3 αφαιρούνται από το dataset καθώς εισάγουν θόρυβο στο σύστημα και δεν βελτιώνουν την εκτίμηση των αλγορίθμων αλλά την χειροτερεύουν. Η ίδια προσέγγιση ακολουθήθηκε και από τους [30][31][32]. Στην συνέχεια γίνεται επεξεργασία των features τα οποία παρέμειναν στο Dataset και εξετάζονται ως προς τις παραμέτρους οι οποίες αναλύθηκαν πιο πάνω. Το συμπέρασμα στο οποίο καταλήγουμε είναι ότι θα πρέπει να δοθεί ιδιαίτερη έμφαση στις καταγραφές των αισθητήρων S\_2, S\_3, S\_4, S\_12, S\_20. Στην εικόνα 4.12 φαίνεται η συσχέτιση των καταγραφών αυτών των αισθητήρων μεταξύ τους και σε σχέση με το RUL τους.

Πραγματοποιήθηκαν διάφορες δοκιμές και συνδυασμοί βάσει των στατιστικών μεγεθών που προέκυψαν, με σκοπό να βελτιωθεί η επίδοση των αλγορίθμων. Ανάλογα με την περίπτωση δημιουργήθηκαν μέχρι και 10 καινούργια features τα οποία βελτιώσαν σε μεγάλο βαθμό την επίδοση του αλγορίθμου σε σχέση με την υπάρχουσα βιβλιογραφία είτε σε επίπεδο ensemble learning είτε σε επίπεδο νευρωνικών δικτύων. Τα Features αυτά δημιουργήθηκαν από τις καταγραφές S\_2, S\_3, S\_4, S\_12, S\_20 και τους συσχετισμούς τους. Στην εικόνα 4.12 φαίνονται τα μεγέθη και οι συσχετισμοί τους.

Θα πρέπει να τονιστεί ότι τα νέα Features προέκυψαν από χρήση του Validation Set καθώς σκοπός είναι να αποφευχθεί το Overfitting. Τα νέα Features για να εισαχθούν δοκιμάστηκαν με την μέθοδο του RUL. Δηλαδή κάθε συνδυασμός Feature με τους 5 τρόπους οι οποίοι αναλύθηκαν πιο πάνω αν βελτίωνε την τελική πρόβλεψη του RUL τότε το νέο Feature έμπαινε στο Dataset. Τα Feature δοκιμάστηκαν με την σειρά ανά ζευγάρια. Δηλαδή πρώτα το S\_2 με το S\_3, μετά το S\_2 με το S\_4, μετά το S\_2 με το S\_7...κλπ. Όποιο νέο Feature μείωνε το RMSE τότε έμπαινε στα Dataset. Μετά το S\_2, το ίδιο έγινε για το S\_3 κλπ, έχοντας όμως κρατήσει τις ήδη υπάρχουσες λύσεις από τα προηγούμενα Features. Δηλαδή πχ. ο συνδυασμός S\_3 με S\_4 βελτίωνε το RUL του Dataset στο οποίο είχαν ήδη προστεθεί τα νέα Features του S\_2.



Εικόνα 4.12 Η συσχέτιση μεταξύ των Features με την μεγαλύτερη διασπορά

Με βάση τους τρόπους οι οποίοι αναλύθηκαν πιο πάνω θα αναφερθούν τα Features τα οποία προστέθηκαν σε κάθε αλγόριθμο που υλοποιήθηκε. Μετά από δοκιμές για την βέλτιστη επίδοση των αλγορίθμων προστέθηκαν τα εξής Features στους αλγορίθμους.

**XGBoost:** Δημιουργήθηκαν 3 νέα Features με τον 5<sup>ο</sup> τρόπο από τα S\_4, S\_3, S\_2 .

**SVM(SVR):** Δημιουργήθηκαν 3 νέα Features με τον 5<sup>ο</sup> τρόπο από τα S\_4, S\_3, S\_2 .

Δημιουργήθηκε 1 με τον 1<sup>ο</sup> τρόπο μεταξύ των S\_2 και S\_3.

**Random Forest:** Δημιουργήθηκαν 4 νέα Features με τον 5<sup>ο</sup> τρόπο από τα S\_20, S\_4, S\_3, S\_2.

Δημιουργήθηκαν 4 νέα Features με τον 1<sup>ο</sup> τρόπο μεταξύ των S\_4 και S\_3, μεταξύ των S\_2 και S\_4, μεταξύ των S\_4 και S\_12, μεταξύ των S\_3 και S\_12.

Δημιουργήθηκε 1 νέο Feature με τον 3<sup>ο</sup> τρόπο μεταξύ των S\_4 και S\_3

Δημιουργήθηκε 1 νέο Feature με τον 4<sup>ο</sup> τρόπο μεταξύ των S\_4 και S\_3

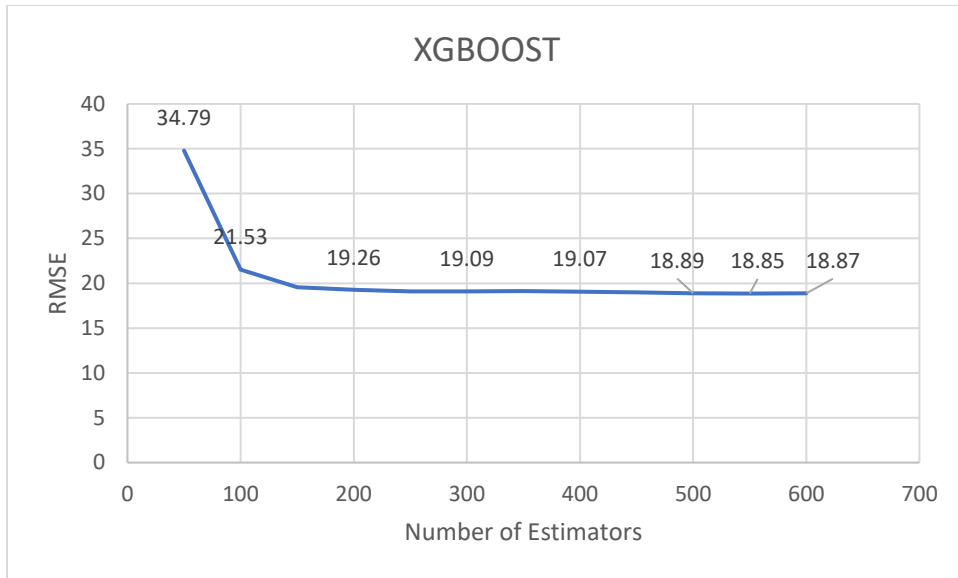
### **Βήμα 5<sup>ο</sup> Αλγόριθμοι και αποτελέσματα.**

Ένα βασικό κριτήριο της επιλογής των αλγορίθμων βασίστηκε στην έλλειψη η οποία υπάρχει στην βιβλιογραφία πάνω στο συγκεκριμένο Dataset σχετικά με μεθόδους Ensemble και SVR. Επίσης σκοπός ήταν να γίνουν πιο μοντέρνες προσεγγίσεις οι οποίες ενσωματώνουν τις τελευταίες τεχνικές των εργασιών οι οποίες έχουν δημοσιευτεί. Στην εργασία χρησιμοποιήθηκαν οι αλγόριθμοι XGboost, SVR, και Random Forest. Όπως αναφέρθηκε και πιο πάνω η βελτιστοποίηση των τιμών έγινε στο Validation Set και τα Hyperparameters τα οποία έδωσαν τα καλύτερα αποτελέσματα στο Validation Set μετά εφαρμόστηκαν στο πραγματικό Test Set. Οπότε

Για τον **XGBoost:**

Στην εικόνα 4.13 φαίνεται η επίδοση του XGBoost ως προς την αλλαγή στον αριθμό των εκτιμητών στο Validation Set. Οι υπερπαραμέτροι βάσει των οποίων ο αλγόριθμος απέδωσε τα βέλτιστα είναι:

```
XGBOOST→ n_estimators=562, max_depth=9, eta=0.1, subsample=0.7,
colsample_bytree=0.7, learning_rate=0.02
```

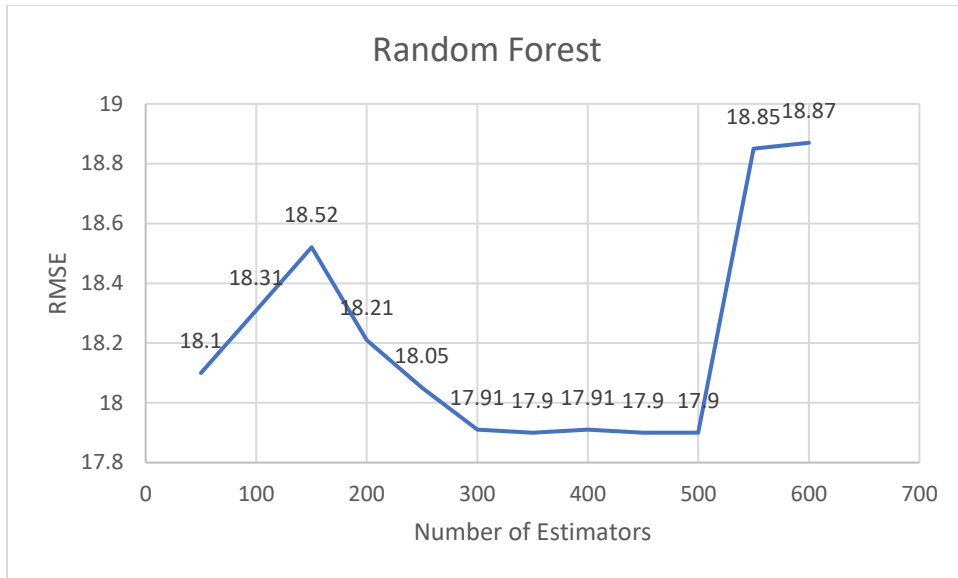


Εικόνα 4.13 Η επίδοση του XGBoost με την εισαγωγή των νέων Features

#### Για τον **Random Forest**:

Στην εικόνα 4.14 φαίνεται η επίδοση του Random Forest μετρημένη ως προς την αλλαγή στον αριθμό των εκτιμητών στο Validation Set. Οι υπερπαράμετροι βάσει των οποίων ο αλγόριθμος απέδωσε τα βέλτιστα είναι:

```
Random Forest → n_estimators = 512, random_state=42, n_jobs=-1,
max_depth=15, min_samples_leaf=1
```

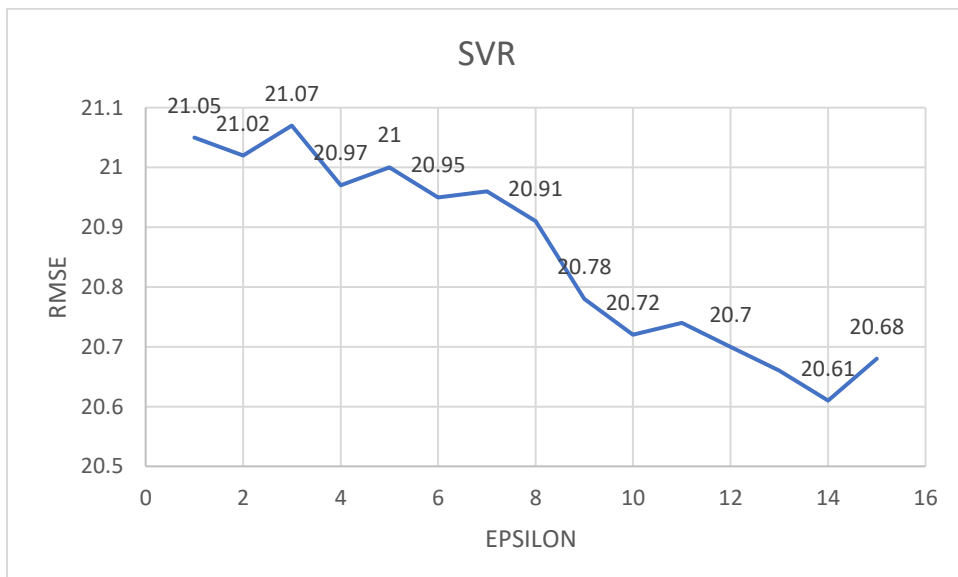


Εικόνα 4.14 Η επίδοση του Random Forest με την εισαγωγή των νέων Features

Για τον **SVR**:

Στην εικόνα 4.15 φαίνεται η επίδοση του SVR μετρημένη ως προς την αλλαγή στο  $\epsilon$  (epsilon) όπου  $2 \cdot \epsilon = \text{margin}$  στο Validation Set. Οι υπερπαράμετροι βάσει των οποίων ο αλγόριθμος απέδωσε τα βέλτιστα είναι:

```
SVR → kernel='rbf', epsilon=12, C=2.2
```



Εικόνα 4.16 Η επίδοση του SVR με την εισαγωγή των νέων Features

Στον SVR χρησιμοποιήθηκε ο Standard Scaler.

### Βήμα 6<sup>ο</sup> Αποτελέσματα

Οι υπερπαραμέτροι οι οποίοι προέκυψαν από το βήμα 5<sup>ο</sup> εδώ εφαρμόζονται στο τελικό Test Set με σκοπό να αξιολογηθούν οι εκτιμήσεις των 3 αλγορίθμων. Ως μέτρο επίδοσης των αλγορίθμων ορίζουμε το RMSE (Root Mean Squared Error-RMSE) ή μέσο τετραγωνικό σφάλμα, το οποίο υπολογίζει το σφάλμα ως την τετραγωνική ρίζα του μέσου όρου του τετραγώνου της διαφοράς μεταξύ πραγματικής και εκτιμώμενης τιμής. Ενώ ουσιαστικά με το  $R^2$  αξιολογείται η αξιοπιστία του μοντέλου.

$$RMSE = \sqrt{\frac{1}{N} \sum_{I=1}^N (Y_I - \hat{Y}_I)^2}$$

	SVR	XGBOOST	RANDOM FOREST
RMSE	18.12	16.06	17.83
$R^2$	0.82	0.87	0.80

Εικόνα 4.15 Πίνακας αποτελεσμάτων

Όπως φαίνεται και στον πίνακα της εικόνας 4.15 ο XGBOOST φαίνεται να είχε την καλύτερη επίδοση από τους 3 αλγορίθμους. Όπως θα φανεί και στις συγκρίσεις με άλλα αποτελέσματα εργασιών η επίδοση αυτή είναι πολύ καλή και πολλές φορές καλύτερη από αυτή των νευρωνικών δικτύων. Θα πρέπει να σημειωθεί ότι και ο Random Forest αλλά και ο SVR είχαν ικανοποιητική επίδοση.

Σύμφωνα με το [32] το οποίο δημοσιεύτηκε από μηχανικούς ML της AMAZON και συνοψίζει τα καλύτερα αποτελέσματα σχετικά με το Dataset FD001 εικόνα 4.16, φαίνεται ότι ο XGBOOST ο οποίος προτείνεται στην παρούσα εργασία επιτυγχάνει αρκετά ανταγωνιστικά αποτελέσματα. Στις εικόνες 4.17 και 4.18 φαίνονται τα αποτελέσματα μερικών από των πιο κομβικών εργασιών πάνω στο συγκεκριμένο dataset και εδώ επιβεβαιώνεται ότι ο XGBOOST της παρούσας εργασίας δίνει αρκετά ανταγωνιστικά αποτελέσματα ως προς υφιστάμενες λύσεις νευρωνικών δικτύων οι οποίες παρουσιάζονται στα CNN+FNN[30], Amazon SpeedWise ML[32] και Time Window Based[30]. Τα αποτελέσματα αυτά επιβεβαιώνονται και από το [33] όπου

διαπιστώνεται ότι ο XGBOOST της παρούσας διπλωματικής δίνει ανταγωνιστικά αποτελέσματα έναντι πολύ καλών υφιστάμενων λύσεων από τον χώρο των νευρωνικών δικτύων.

Method	Test RMSE
<u>CNN + FNN</u>	18.45
SML Timeseries Encoding	15.22
<u>Time Window Based NN</u>	15.16
<u>Multi-objective deep belief networks ensemble</u>	15.04
<u>CNN + FNN without rectified labels</u>	13.32

Εικόνα 4.16 Τα αποτελέσματα των βέλτιστων δημοσιευμένων λύσεων σύμφωνα με το [32]

DL approach & refs.	FD001	FD002	FD003	FD004
CNN+FNN [5]	18.45	30.29	19.82	29.16
LSTM+FNN [6]	16.14	24.49	16.18	28.17
MODBNE [19]	15.04	25.05	12.51	28.66
CNN+FNN [7]	12.61	<b>22.36</b>	12.64	23.31
Proposed semi-supervised setup	<b>12.56</b>	22.73	<b>12.10</b>	<b>22.66</b>

Εικόνα 4.17 Τα αποτελέσματα των βέλτιστων δημοσιευμένων λύσεων σύμφωνα με το [33]

TW=1	Average	MODBNE	DBN	SKF	MLP	ELM	HELM	SVM	LASSO	ETR	KNR	GB	RF
FD001	Score	640.27	1001.44	762.85	959.63	740.52	835.53	852.07	894.21	1359.38	604.26	<b>575.04</b>	802.23
	RMSE	<b>17.96</b>	18.48	19.24	18.48	19.40	20.02	20.58	22.43	22.05	19.73	18.80	20.23

Εικόνα 4.18 Τα αποτελέσματα των βέλτιστων δημοσιευμένων λύσεων σύμφωνα με το [35]



### 4.3 Συμπεράσματα

Σκοπός της παρούσας εργασίας ήταν να γίνει προσέγγιση της αστοχίας αεριοστροβίλων οι οποίοι χρησιμοποιούνται στην αεροπορική βιομηχανία με σκοπό την καλύτερη προσέγγιση στην εκτίμηση του υπόλοιπου χρόνου ζωής τους(RUL). Η προσέγγιση αυτή έγινε με χρήση αλγορίθμων μηχανικής μάθησης, ενώ η επιλογή των συγκεκριμένων αλγορίθμων στόχευε να καλύψει το κενό το οποίο υπάρχει στην βιβλιογραφία ,σχετικά με τους συγκεκριμένους αλγόριθμους στο συγκεκριμένο Dataset το οποίο επιλέχθηκε. Η επιλογή του κατάλληλου Dataset έγινε βάσει της διεθνούς βιβλιογραφίας έχοντας ως γνώμονα ένα Dataset το οποίο καλύπτει τους στόχους της εργασίας αλλά και να έχει χρησιμοποιηθεί από όσο το δυνατόν περισσότερες ερευνητικές ομάδες , με σκοπό τα αποτελέσματα τα οποία θα εξαχθούν να μπορούν να είναι συγκρίσιμα με αυτά άλλων εργασιών ,ώστε να μπορεί να αξιολογηθεί ένα αποτέλεσμα αν αποτελεί μια καλή(ανταγωνιστική) προσέγγιση. Λόγω της ευρείας χρήσης του Dataset το οποίο επιλέχτηκε, δηλαδή το Dataset FD001 της NASA τα αποτελέσματα μπορούν εύκολα να συγκριθούν με πολλές εργασίες καθώς είναι το πιο διαδεδομένο Dataset, οπότε μπορεί να αξιολογηθεί εύκολα αν μια προσέγγιση φέρνει ικανοποιητικά και ανταγωνιστικά αποτελέσματα.

Στο κομμάτι της υλοποίησης στην παρούσα εργασία πάνω στο Dataset FD001 της NASA για την κόπωση αεριοστροβίλων , πραγματοποιήθηκε προσέγγιση του προβλήματος με μεθόδους Ensemble(XGBoost & Random Forest) και SVR. Η επιλογή των συγκεκριμένων αλγορίθμων είχε σκοπό να καλύψει το κενό το οποίο υπάρχει στην βιβλιογραφία σχετικά με τους αλγορίθμους αυτούς στο συγκεκριμένο Dataset. Οι υπάρχουσες προσεγγίσεις της βιβλιογραφίας είναι κυρίως με χρήση νευρωνικών δικτύων και οι ελάχιστες υλοποιήσεις με χρήση μεθόδων ensemble περιορίζονται σε απλή εφαρμογή των αλγορίθμων, χωρίς πολύ ικανοποιητικά αποτελέσματα. Στην παρούσα εργασία ακολουθήθηκαν συγκεκριμένα βήματα με σκοπό οι αλγόριθμοι οι οποίοι αναφέρθηκαν να αποδώσουν τα μέγιστα και να οδηγήσουν σε ανταγωνιστικά αποτελέσματα.

Τα βήματα τα οποία έκαναν την ειδοποιό διαφορά και οδήγησαν σε ικανοποιητικά αποτελέσματα ήταν αρχικά η αφαίρεση κάποιων features του Dataset και η δημιουργία καινούργιων τα οποία θα προσέφεραν πρόσθετη πληροφορία στο σύστημα και έτσι θα βελτίωναν την εκτίμηση των αλγορίθμων. Η επιλογή των παλιών Features αλλά και η δημιουργία των καινούργιων έγινε βάσει στατιστικής ανάλυσης

σχετικά με το πόση διασπορά έχει ένα Feature και τι συσχέτιση έχει με το RUL. Έτσι αν ένα Feature είχε μεταβολή στην τιμή του και αυτή η μεταβολή είχε κάποια συσχέτιση με την μεταβολή του RUL τότε το Feature πολύ πιθανόν να μπορεί να βοηθήσει στην καλύτερη εκτίμηση αποτελέσματος. Με αυτό τον γνώμονα έγινε η επιλογή των Features, ενώ η δημιουργία καινούργιων έγινε με υπολογισμό του RUL. Δηλαδή αν κάθε νέο Feature το οποίο δημιουργείται βοηθάει στην μείωση της εκτίμησης του RUL τότε το Feature αυτό το κρατάμε. Χρησιμοποιήθηκαν 5 μέθοδοι δημιουργίας Features. Όπως πολύ εύκολα μπορεί να γίνει αντιληπτό η δημιουργία νέων Features πάνω στο TEST set όπου θα εφαρμοστούν οι αλγόριθμοι, θα οδηγούσε σε overfitting το μοντέλο μας, το οποίο και δεν θέλουμε για αυτό και έγινε χρήση Validation Set.

Το Validation Set αποτελείται από τις μηχανές με αριθμό 85 έως 100 του Training SET του Dataset FD001. Με βάση το Validation Set έγινε όλη η παραμετροποίηση και tuning των αλγορίθμων με σκοπό οι βελτιστοποιημένοι αλγόριθμοι έπειτα να εφαρμοστούν στο τελικό Test Set. Συγκεκριμένα από το Validation set προέκυψαν 3 νέα Features για τον XGBoost, 4 νέα Features για τον SVR και 10 για τον Random Forest. Αντίστοιχα η εφαρμογή των βελτιστοποιημένων αλγορίθμων στο Test Set έφερε πολύ ανταγωνιστικά αποτελέσματα σε σχέση με τις υπάρχουσες δημοσιευμένες εργασίες και συγκεκριμένα το RMSE για τον SVR ήταν 18.12 για τον XGBoost 16.06 και για τον Random Forest 17.83. Τα αποτελέσματα των αλγορίθμων είναι πολύ ανταγωνιστικά σε σχέση με υπάρχουσες λύσεις στο πεδίο των Νευρωνικών δικτύων και με βάση την βιβλιογραφία η οποία αναφέρεται είναι οι καλύτερες λύσεις χρησιμοποιώντας τους συγκεκριμένους αλγόριθμους. Η χρήση των νέων Features κατά μέσο όρο μείωσε 4 με 5 μονάδες το RMSE για τον κάθε αλγόριθμο και συγκεκριμένα ο XGBoost ο οποίος είχε την καλύτερη επίδοση από όλους μείωσε το RMSE του κατά 5 μονάδες.

Επιπλέον θα πρέπει να τονιστεί ότι στην βελτιστοποίηση της προσέγγισης βοήθησε πάρα πολύ η χρήση της μεθόδου Clip η οποία αναλύθηκε πιο πάνω βάζοντας ως όριο το 110, ο XGBoost μείωσε το RMSE κατά 14 μονάδες, ο SVR και αυτός μείωσε το RMSE του κατά 14 μονάδες, ενώ ο Random Forest μείωσε το RMSE του κατά 13 μονάδες. Η χρήση όλων των παραπάνω τεχνικών συνδυαστικά έφερε ως βέλτιστη λύση τον XGBoost με RMSE 16.06 η οποία είναι αρκετά ανταγωνιστική με τις υπάρχουσες στην βιβλιογραφία οι οποίες εμφανίζονται στις εικόνες 4.16, 4.17, 4.18.

Σαν μελλοντική εργασία θα είχε ενδιαφέρον ένας συνδυασμός (Fusion) των 3 αλγορίθμων οι οποίοι χρησιμοποιήθηκαν στην παρούσα εργασία, με σκοπό να βελτιωθούν ακόμη περισσότερο οι προβλέψεις του RUL των μηχανών.

## ΑΝΑΦΟΡΕΣ

- [1] Sciban, R. 6 Tools for a Successful Predictive Maintenance Program. 2017 [cited 2018 06.06]; Available from: <https://us.hitachi-solutions.com/blog/6-tools-for-a-successful-predictivemaintenance-program/>.
- [2] <https://iot-analytics.com/top-20-companies-enabling-predictive-maintenance/>
- [3]. Scully, P. New Report Indicates US\$11 Billion Predictive Maintenance Market By 2022, Driven by IoT Technology And New Services. 2017 [cited 2018 06.06]; Available from: <https://iotanalytics.com/report-us11-billion-predictive-maintenance-market-by-2022/>.
- [4] Bousdekis, A., Lepenioti, K., Apostolou, D., & Mentzas, G. (2019). Decision making in predictive maintenance: Literature review and research agenda for industry 4.0. IFAC-PapersOnLine, 53(12), 607-612. <https://doi.org/10.1016/j.ifacol.2019.11.226>
- [5] Amruthnath, N., and Gupta, T. (2018). A research study on unsupervised machine learning algorithms for early fault detection in predictive maintenance. In 2018 5th International Conference on Industrial Engineering and Applications (ICIEA) (pp. 355-361). IEEE
- [6] Σπυρίδων Δεσσαλερμος, Αξιοπιστία συστημάτων, Στρατιωτική Σχολή Ευελπίδων (2015).
- [7] Steven Aliano. The cost of downtime. In [https://blog.isa.org/worlds-largest-manufacturers-lose-1-trillion/year-to-machine-failure?\\_ga=2.44211556.1739773052.1644485415-913152564.1644485415, 2014](https://blog.isa.org/worlds-largest-manufacturers-lose-1-trillion/year-to-machine-failure?_ga=2.44211556.1739773052.1644485415-913152564.1644485415, 2014).
- [8] Andrew Lerner. World's largest manufacturers lose \$1 trillion/year to machine failure. In <https://blogs.gartner.com/andrew-lerner/2014/07/16/the-cost-of-downtime/>, 2021.
- [9] M. Mulders, M.H. Predictive Maintenance 4.0 - Predict the unpredictable. 2017.
- [10] Nowitz, A. The Economics of the Smart Factory How Does Machine Learning Lower the Cost of Asset Maintenance Part 1. 2017 [cited 2018 05.06]; Available from: <http://www.presenso.com/single-post/2017/05/24/the-economics-of-the-smart-factoryhow-does-machine-learning-lower-the-cost-of-asset-maintenance-part-1/>.
- [11] Chen, X.; Wang, S.; Qiao, B.; Chen, Q. Basic research on machinery fault diagnostics: Past, present, and future trends. *Front. Mech. Eng.* 2018, 13, 264–291. [CrossRef] 3. Wei, Y.; Li, Y.; [12] Xu, M.; Huang, W. A review of early fault diagnosis approaches and their applications in rotating machinery. *Entropy* 2019, 21, 409. [Crosse] [PubMed] 4. Lu, B.L.; Liu, Z.H.; Wei, [13] H.L.; Chen, L.; Zhang, H.; Li, X.H. A Deep Adversarial Learning Prognostics Model for Remaining Useful Life Prediction of Rolling Bearing. *IEEE Trans. Arti. In tell.* 2021, 1.

- [14]. Zeng, D.; Zhou, D.; Tan, C.; Jiang, B. Research on model-based fault diagnosis for a gas turbine based on transient performance. Appl. Sci. 2018, 8, 148. [crosser]
- [15] Andrea Devenish Data analytics for predictive maintenance of wind turbines, Stellenbosch University 2021.
- [16] Irfan Ullah, Fan Yang, Predictive Maintenance of Power Substation Equipment by Infrared Thermography Using a Machine Learning Approach, Energies 2017, 10, 1987 (2017)
- [17][https://www.boeing.com/commercial/aeromagazine/articles/qtr\\_2\\_07/AERO\\_Q207\\_article3.pdf](https://www.boeing.com/commercial/aeromagazine/articles/qtr_2_07/AERO_Q207_article3.pdf)
- [18] Vimala Mathew, Tom Toby, Vikram Singh, B Maheswara Rao, M Goutham Kumar. Prediction of Remaining Useful Lifetime (RUL) of Turbofan Engine using Machine Learning , IEEE International Conference on Circuits and Systems (ICCS 2017). (2017)
- [19] Daniel Maraini, Mark Simpson, Ronald Brown, and Michael Poporad Development of a Data-driven Model for Marine Gas Turbine (MGT) Engine Health Monitoring, Annual Conference of the Prognostics and health management society. (2018)
- [20] OpFlex advanced controlled solutions GE Gas Power, <https://www.ge.com/gas-power/products/digital-and-controls/opflex>
- [21] Babu, G.S.; Zhao, P.; Li, X.L. Deep convolutional neural network-based regression approach for estimation of remaining useful life. In International Conference on Database Systems for Advanced Applications; Springer: Cham, Switzerland, 2016; pp. 214–228.
- [22] Huang, R.; Xi, L.; Li, X.; Liu, C.R.; Qiu, H.; Lee, J. Residual life predictions for ball bearings based on self-organizing map and back propagation neural network methods. Mech. Syst. Signal Process. 2007, 21, 193–207.
- [23] Zhao, R.; Wang, J.; Yan, R.; Mao, K. Machine Health Monitoring with LSTM Networks. In Proceedings of the 2016 10th International Conference on Sensing Technology (ICST), Nanjing, China, 11–13 November 2016; pp. 1–6.
- [27] Τεχνητή νοημοσύνη, [Βλαχαβάς 2006]
- [28] Σταματάτος Ευστάθιος σημειώσεις μαθήματος Μηχανικής Μάθησης, πανεπιστήμιο Αιγαίου[2021]
- [29]. Giduthuri Sateesh Babu, Peilin Zhao Deep Convolutional Neural Network Based Regression Approach for Estimation of Remaining Useful Life(2016)

[30]. Pin Lim, C. Goth, K. Tan A time window neutral network based framework for Remaining useful life estimation. International Joint Conference on Neural Networks(IJCNN)(2016)

[31]. Chong Zhang, Pin Lim, A K Qin, Kay Chen Tan, Multiobjective Deep Belief Networks Ensemble for Remaining Useful Life Estimation in Prognostics, IEEE Trans Neural Netw Learn Syst 2017 Oct;28(10):2306-2318

[32]. Rohan Thavarajah, Yannick Agbor, and Dmitriy Tishechkin, Predicting the failure of turbofan engines using SpeedWise Machine Learning(2022)  
<https://aws.amazon.com/blogs/industries/predicting-the-failure-of-turbofan-engines-using-speedwise-machine-learning/>

[33]. Andre Listoy Ellefsen, Emil Bjørlykhaug, Remaining useful life predictions for turbofan engine degradation using semi-supervised deep architecture(2016).

[34]. Vimala Mathew, Tom Toby, Vikram Singh, B Maheswara Rao, M Goutham Kumar, Prediction of Remaining Useful Lifetime (RUL) of Turbofan Engine using Machine Learning, IEEE International Conference on Circuits and Systems (ICCS 2017)

[35]. Chong Zhang, Pin Lim, A.K. Qin, Kay Chean Tan, Deep belief Networks Ensemble for Remaining Useful life estimation in Prognostics. IEEE Transactions on Neural Networks and Learning Systems ( Volume: 28, Issue: 10, October 2017)