



**ΠΑΝΕΠΙΣΤΗΜΙΟ
ΔΥΤΙΚΗΣ ΑΤΤΙΚΗΣ**
UNIVERSITY OF WEST ATTICA

Σχολή Μηχανικών

Τμήμα Μηχανικών Βιομηχανικής Σχεδίασης και Παραγωγής

ΤΙΤΛΟΣ ΔΙΠΛΩΜΑΤΙΚΗΣ ΕΡΓΑΣΙΑΣ:

«Προβλεπτική συντήρηση στη Βιομηχανία 5.0»

Γεώργιος Ασλανίδης

8069713

Επιβλέπουσα καθηγήτρια:

Ελένη Αικατερίνη Δελίγκου

Αθήνα 2021



**ΠΑΝΕΠΙΣΤΗΜΙΟ
ΔΥΤΙΚΗΣ ΑΤΤΙΚΗΣ**
UNIVERSITY OF WEST ATTICA

School of Engineering

Department of Industrial Design and Production Engineering

ΤΙΤΛΟΣ ΔΙΠΛΩΜΑΤΙΚΗΣ ΕΡΓΑΣΙΑΣ:

«Predictive maintenance in Industry 5.0»

George Aslanidis

8069713

Supervisor:

Eleni Aikaterini Leligkou

Αθήνα 2021



ΠΑΝΕΠΙΣΤΗΜΙΟ ΔΥΤΙΚΗΣ ΑΤΤΙΚΗΣ UNIVERSITY OF WEST ATTICA

Σχολή Μηχανικών

Τμήμα Μηχανικών Βιομηχανικής Σχεδίασης και Παραγωγής

ΤΙΤΛΟΣ ΔΙΠΛΩΜΑΤΙΚΗΣ ΕΡΓΑΣΙΑΣ:

«Προβλεπτική συντήρηση στη Βιομηχανία 5.0»

Μέλη Εξεταστικής Επιτροπής συμπεριλαμβανομένου και του Εισηγητή

Η μεταπτυχιακή διπλωματική εργασία εξετάστηκε επιτυχώς από την κάτωθι Εξεταστική Επιτροπή:

A/a	ΟΝΟΜΑ ΕΠΩΝΥΜΟ	ΒΑΘΜΙΑΔΑ/ΙΔΙΟΤΗΤΑ	ΨΗΦΙΑΚΗ ΥΠΟΓΡΑΦΗ
1	Ελένη Αικατερίνη Λελίγκου	Αν. Καθηγήτρια	
2	Θεόδωρος Γκανέτσος	Καθηγητής	
3	Δημήτριος Τσελές	Καθηγητής	

ΔΗΛΩΣΗ ΣΥΓΓΡΑΦΕΑ ΜΕΤΑΠΤΥΧΙΑΚΗΣ ΕΡΓΑΣΙΑΣ

Ο/Η κάτωθι υπογεγραμμένος/η Γεώργιος Α. Αδακίδης.....
του Πανεπιστημίου με αριθμό μητρώου 8009713 φοιτητής/τρια του Προγράμματος Μεταπτυχιακών
Σπουδών Αυτοματισμός, Παραγωγή και Υπηρεσιών του Τμήματος Μηχανικής Βιομηχανικής της Σχολής
Μηχανικών του Πανεπιστημίου Δυτικής Αττικής, δηλώνω ότι:

«Είμαι συγγραφέας αυτής της μεταπτυχιακής εργασίας και ότι κάθε βοήθεια την οποία είχα για την προετοιμασία της, είναι πλήρως αναγνωρισμένη και αναφέρεται στην εργασία. Επίσης, οι όποιες πηγές από τις οποίες έκανα χρήση δεδομένων, ιδεών ή λέξεων, είτε ακριβώς είτε παραφρασμένες, αναφέρονται στο σύνολό τους, με πλήρη αναφορά στους συγγραφείς, τον εκδοτικό οίκο ή το περιοδικό, συμπεριλαμβανομένων και των πηγών που ενδεχομένως χρησιμοποιήθηκαν από το διαδίκτυο. Επίσης, βεβαιώνω ότι αυτή η εργασία έχει συγγραφεί από μένα αποκλειστικά και αποτελεί προϊόν πνευματικής ιδιοκτησίας τόσο δικής μου, όσο και του Ιδρύματος.

Παράβαση της ανωτέρω ακαδημαϊκής μου ευθύνης αποτελεί ουσιώδη λόγο για την ανάκληση του πτυχίου μου».

*Επιθυμώ την απαγόρευση πρόσβασης στο πλήρες κείμενο της εργασίας μου μέχρι και έπειτα από αίτηση μου στη Βιβλιοθήκη και έγκριση του επιβλέποντα καθηγητή.

Ο/Η Δηλών/ούσα

* Ονοματεπώνυμο /Ιδιότητα

Ψηφιακή Υπογραφή Επιβλέποντα

(Υπογραφή)

* Εάν κάποιος επιθυμεί απαγόρευση πρόσβασης στην εργασία για χρονικό διάστημα 6-12 μηνών (embargo), θα πρέπει να υπογράψει ψηφιακά ο/η επιβλέπων/ουσα καθηγητής/τρια, για να γνωστοποιεί ότι είναι ενημερωμένος/η και συναινεί. Οι λόγοι χρονικού αποκλεισμού πρόσβασης περιγράφονται αναλυτικά στις πολιτικές του Ι.Α. (σελ. 6):

https://www.uniwa.gr/uploads/2021/01/%CE%A0%CE%BF%CE%BB%CE%B9%CF%84%CE%B9%CE%BA%CE%B5%CC%81%CF%82_%CE%99%CE%B4%CF%81%CF%85%CE%BC%CE%B1%CF%84%CE%B9%CE%BA%CE%BF%CF%83%CC%81_%CE%91%CF%80%CE%BF%CE%B8%CE%B5%CF%84%CE%B7%CF%81%CE%B9%CC%81%CE%BF%CF%83_final.pdf

Περίληψη

Στη βιομηχανία γίνεται όλο και πιο επιτακτική η ανάγκη εισαγωγής νέων τεχνολογικών μεθόδων σε μια διαρκή προσπάθεια του ανθρώπου να αυξήσει την παραγωγή και να μειώσει το οικονομικό και χρονικό κόστος. Στην παρούσα εργασία περιγράφεται πως η χρήση μηχανικής μάθησης θα μπορούσε να αποτελέσει τη λύση για την έγκαιρη ανίχνευση ενδεχόμενων αστοχιών και την αποφυγή τους. Επίσης, περιγράφονται οι αλγόριθμοι που χρησιμοποιούνται και οι δυνατότητες ορισμένων στρατηγικών συντήρησης που προκύπτουν από τη μελλοντική χρήση τους στη Βιομηχανία 5.0 με τη βοήθεια της τεχνητής νοημοσύνης. Η εργασία ολοκληρώνεται με την παρουσίαση και την ανάλυση των στατιστικών που απέρρευσαν από το πειραματικό μέρος, το οποίο διεξάχθηκε πάνω σε ένα πραγματικό σύνολο δεδομένων. Τέλος, γίνεται καταγραφή συμπερασμάτων και πλεονεκτημάτων που έχει η χρήση της μηχανικής μάθησης, αλλά και μια πρόταση για περαιτέρω μελέτη.

Abstract

The need for the introduction of new technological methods in an ongoing human effort to increase production and reduce financial and time costs is becoming increasingly urgent in industry. This thesis statement describes how the use of machine learning could be the solution for the early detection of possible failures and their prevention. It also describes the algorithms used and the possibilities of some maintenance strategies arising from their future use in Industry 5.0 with the help of artificial intelligence. The master's thesis is completed by presenting and analyzing the statistics that emerged from the experimental part, which was conducted on a real data set. Finally, conclusions and advantages of the use of machine learning are recorded, as well as a proposal for further study.

Πίνακας Περιεχομένων

Περίληψη	4
Abstract	5
Εισαγωγή	9
Σκοπός της διπλωματικής εργασίας	10
Κεφάλαιο 1: Μέθοδοι συντήρησης.....	11
1.1 Τι είναι η συντήρηση.....	11
1.2 Τύποι συντήρησης.....	11
1.2.1 Προκαθορισμένη συντήρηση (Predetermined maintenance)	12
1.2.2 Διορθωτική συντήρηση (Reactive maintenance).....	12
1.2.3 Προληπτική συντήρηση (Preventive maintenance).....	13
1.2.4 Συντήρηση βάσει συνθηκών (Condition-based maintenance).....	13
1.2.5 Προβλεπτική συντήρηση (Predictive maintenance)	14
Κεφάλαιο 2: Σύγκριση στρατηγικών συντήρησης	15
2.1 Πλεονεκτήματα - Μειονεκτήματα	15
2.2 Κόστος συντήρησης	18
Κεφάλαιο 3: Βιομηχανική επανάσταση.....	20
3.1 Πρώτη βιομηχανική επανάσταση.....	20
3.2 Δεύτερη βιομηχανική επανάσταση	20
3.3 Τρίτη βιομηχανική επανάσταση.....	21
3.4 Τέταρτη βιομηχανική επανάσταση	21
3.5 Πέμπτη βιομηχανική επανάσταση	22
Κεφάλαιο 4: Συλλογή και προεπεξεργασία δεδομένων.....	24
4.1 Συλλογή δεδομένων	24
4.2 Προεπεξεργασία δεδομένων	26
Κεφάλαιο 5: Μηχανική μάθηση	28
5.1 Τύποι μηχανικής μάθησης.....	29

5.1.1 Επιβλεπόμενη μάθηση	30
5.1.2 Μη επιβλεπόμενη μάθηση	30
5.1.3 Ενισχυτική μάθηση	30
5.1.4 Ημι-επιβλεπόμενη μάθηση	31
5.2 Αλγόριθμοι	31
5.2.1 Δίκτυα Bayes (Naive Bayes)	31
5.2.2 Γενικευμένο γραμμικό μοντέλο (Generalized Linear Model)	33
5.2.3 Λογιστική παλινδρόμηση (Logistic Regression)	35
5.2.4 Δέντρο απόφασης (Decision Tree)	36
5.2.5 Τυχαίο δάσος (Random Forest)	38
5.2.6 Ενίσχυση της κλίσης (Gradient Boosting)	39
5.2.7 Μηχανή διανυσμάτων υποστήριξης (Support Vector Machine)	40
Κεφάλαιο 6: Πειραματικό μέρος	43
Συμπεράσματα	52
Βιβλιογραφία	54

Εικόνα 1 Πλαίσιο ανάλυσης (Andrade, 2021).....	10
Εικόνα 2 Η εξέλιξη των στρατηγικών συντήρησης.....	14
Εικόνα 3 Σχέση κόστους - αριθμού βλαβών.....	19
Εικόνα 4 Χρονοδιάγραμμα βιομηχανικών επαναστάσεων (Schwab, 2017)	22
Εικόνα 5 Κύκλος διεργασιών ανθρώπου–μηχανής (Francesco Longo, 2020)	23
Εικόνα 6 Οι τέσσερις τύποι των ερευνητικών δεδομένων.....	25
Εικόνα 7 Τύποι μηχανικής μάθησης (Karthik Seetharam, 2020)	29
Εικόνα 8 Δείγμα γραμμικής παλινδρόμησης (Tam, 2020)	34
Εικόνα 9 Προσέγγιση λογιστικής παλινδρόμησης (Equiskill, 2018)	35
Εικόνα 10 Παράδειγμα δέντρου απόφασης (Niculaescu, 2018).....	37
Εικόνα 11 Ταξινόμηση τυχαίου δάσους (Ampadu, 2021).....	39
Εικόνα 12 Ταξινόμηση με όριο απόφασης (Javatpoint, 2021)	41
Εικόνα 13 Τυχαίο δείγμα των δεδομένων όπως εμφανίζονται στο RapidMiner.....	45
Εικόνα 14 Στατιστική ανάλυση των δεδομένων.....	46
Εικόνα 15 Επιλογή για πρόβλεψη της συνθήκης Machine failure=1	47
Εικόνα 16 Αποτελέσματα δείκτη απόδοσης ακρίβειας	48
Εικόνα 17 Αποτελέσματα δείκτη απόδοσης σφάλματος ταξινόμησης.....	49
Εικόνα 18 Γραφική απεικόνιση για την ακρίβεια.....	49
Εικόνα 19 Γραφική απεικόνιση σφάλματος ταξινόμησης.....	50
Εικόνα 20 Γραφική παράσταση συνολικού χρόνου διεργασίας.....	51

Εισαγωγή

Το κόστος συντήρησης αποτελεί ένα σημαντικό ποσοστό στο συνολικό κόστος οποιασδήποτε παραγωγικής μονάδας. Ανάλογα με τη βιομηχανία, μπορεί να αντιπροσωπεύει το 15% έως το 60% του κόστους των παραγόμενων αγαθών. Οπότε γίνεται εύκολα κατανοητό ότι το κόστος συντήρησης έχει μεγάλη σημασία για κάθε παραγωγική μονάδα και μια βελτίωσή του μπορεί να επηρεάσει άμεσα την κερδοφορία.

Έρευνα σχετικά με τη διαχείριση συντήρησης δείχνει ότι το ένα τρίτο όλων των δαπανών συντήρησης σπαταλάται λόγω περιττής ή ακατάλληλης συντήρησης. Όταν θεωρείται ότι η βιομηχανία των ΗΠΑ δαπανά περισσότερα από 200 δισεκατομμύρια δολάρια κάθε χρόνο για τη συντήρηση εγκαταστάσεων και εξοπλισμού, ο αντίκτυπος της συντήρησης στην παραγωγικότητα και τα κέρδη καθίσταται σαφής. Το αποτέλεσμα της αναποτελεσματικής διαχείρισης συντήρησης αντιπροσωπεύει απώλεια μεγαλύτερη από 60 δισεκατομμύρια δολάρια ετησίως. Το πιο σημαντικό είναι ότι η αναποτελεσματική διαχείριση συντήρησης επηρεάζει την ικανότητα παραγωγής ποιοτικών προϊόντων που είναι ανταγωνιστικά στην παγκόσμια αγορά. Οι απώλειες στο χρόνο παραγωγής και στην ποιότητα του προϊόντος που οφείλονται σε κακή ή ανεπαρκή διαχείριση συντήρησης έχουν δραματικό αντίκτυπο στην ικανότητα του αμερικανικού κλάδου να ανταγωνιστεί με την Ιαπωνία και άλλες χώρες που έχουν υιοθετήσει πιο προηγμένες φιλοσοφίες διαχείρισης κατασκευής και συντήρησης.

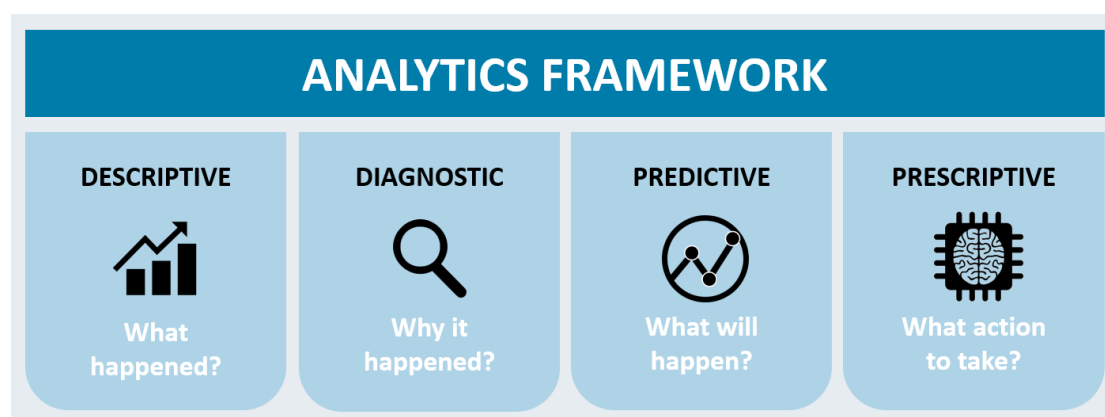
Ο πρωταρχικός λόγος για αυτήν την αναποτελεσματική διαχείριση είναι η έλλειψη πραγματικών δεδομένων για τον ποσοτικό προσδιορισμό των πραγματικών αναγκών επισκευής ή συντήρησης μηχανημάτων, εξοπλισμού και συστημάτων. Ο σχεδιασμός συντήρησης βασίστηκε, και σε πολλές περιπτώσεις, βασίζεται σε στατιστικά δεδομένα ή σε πραγματική αστοχία εξοπλισμού. Η ανάπτυξη μικροεπεξεργαστών ή εργαλείων που βασίζονται στην υπολογιστική δύναμη και μπορούν να χρησιμοποιηθούν για την παρακολούθηση της κατάστασης λειτουργίας των εγκαταστάσεων, των μηχανημάτων και των συστημάτων έχει παράσχει τα μέσα για τη διαχείριση των δραστηριοτήτων συντήρησης με στόχο τη μείωση ή την εξάλειψη περιττών επισκευών, την πρόληψη καταστροφικών βλαβών και τη μείωση των αρνητικών επιπτώσεων της συντήρησης στην κερδοφορία των βιομηχανιών.

Σκοπός της διπλωματικής εργασίας

Ο σκοπός της παρούσας διπλωματικής εργασίας είναι η ανάδειξη της σημαντικότητας, αλλά και της αναγκαιότητας από τη χρήση μεθόδων προβλεπτικής συντήρησης (predictive maintenance). Ο στόχος της προβλεπτικής συντήρησης είναι η δυνατότητα πρόβλεψης, βάσει ορισμένων παραγόντων, ακολουθούμενη από την πρόληψη της βλάβης μέσω τακτικής προγραμματισμένης και διορθωτικής συντήρησης. Έτσι, επιτυγχάνεται μείωση του κόστους συντήρησης και μεγαλύτερη αποδοτικότητα παραγωγής, καθώς περιορίζεται ο χρόνος που η μηχανή παραγωγής δε λειτουργεί.

Για να επιτευχθεί αυτό, χρειάζονται τα κατάλληλα δεδομένα, τα οποία συλλέγονται από τους αντίστοιχους αισθητήρες. Με τη βοήθεια της μηχανικής μάθησης αναλύεται όλος ο όγκος δεδομένων που είναι διαθέσιμος και με τους κατάλληλους αλγορίθμους δημιουργείται ένα προγνωστικό μοντέλο, το οποίο εμφανίζει ως αποτέλεσμα τη χρονική στιγμή κατά την οποία σε κάποιο μέρος του εξοπλισμού θα διαπιστωθεί βλάβη και κατ' επέκταση θα σταματήσει η παραγωγική λειτουργία της μηχανής ή του συστήματος.

Η χρήση τεχνητής νοημοσύνης πραγματοποιείται ήδη σε διάφορες εφαρμογές και όπως όλα δείχνουν στο άμεσο μέλλον θα χρησιμοποιηθεί και στον τομέα της βιομηχανίας. Με τη συνέργεια της τεχνητής νοημοσύνης πάνω στην προβλεπτική συντήρηση, δίνεται η δυνατότητα όχι μόνο της ένδειξης για μια πιθανή βλάβη, αλλά και προτεινόμενες ενέργειες οι οποίες θα βοηθήσουν τόσο στην αποκατάσταση της βλάβης τη σωστή χρονική στιγμή, όσο και στη μεγιστοποίηση της παραγωγής.



Εικόνα 1 Πλαίσιο ανάλυσης (Andrade, 2021)

Κεφάλαιο 1: Μέθοδοι συντήρησης

1.1 Τι είναι η συντήρηση

Ως συντήρηση ενός μηχανήματος ή ενός συστήματος εννοείται η προσπάθεια διατήρησης της κατάστασης και της απόδοσής του, όπως όταν ήταν καινούργιο. Οι δραστηριότητες συντήρησης μπορούν να χωριστούν σε δύο κατηγορίες. Στις προγραμματισμένες δραστηριότητες συντήρησης και στις μη προγραμματισμένες δραστηριότητες συντήρησης.

Η προγραμματισμένη συντήρηση είναι η συντήρηση που οργανώνεται και εκτελείται με μελλοντικές σκέψεις, έλεγχο και τήρηση των αρχείων σύμφωνα με τα προκαθορισμένα σχέδια. Η μη προγραμματισμένη δραστηριότητα συντήρησης πραγματοποιείται κατ' ανάγκη και αφού έχει εμφανιστεί κάποια αστοχία στον εξοπλισμό, όπου και χρειάζεται επιδιόρθωση ή αντικατάστασή του, με σκοπό την επαναλειτουργία.

1.2 Τύποι συντήρησης

Οι διαδικασίες και τα συστήματα συντήρησης στις βιομηχανίες είναι κρίσιμα για την αποτελεσματικότητα και την αξιοπιστία των παραγόμενων προϊόντων τους. Ως αποτέλεσμα, πολλές διαφορετικές μορφές συντήρησης συνυπάρχουν για να επιτρέπουν στους υπεύθυνους συντήρησης να μεγιστοποιούν τις γραμμές παραγωγής τους. Κάθε επιχείρηση καθορίζει τις απαιτήσεις για συντήρηση και αναπτύσσει μια στρατηγική για την κάλυψή τους. Οι ακόλουθες είναι οι πιο κοινές μορφές βιομηχανικής συντήρησης:

- Προκαθορισμένη συντήρηση (Predetermined maintenance)
- Διορθωτική συντήρηση (Reactive maintenance)
- Προληπτική συντήρηση (Preventive maintenance)
- Συντήρηση βάσει συνθηκών (Condition based)
- Προβλεπτική συντήρηση (Predictive maintenance)

1.2.1 Προκαθορισμένη συντήρηση (Predetermined maintenance)

Η προκαθορισμένη συντήρηση, σε αντίθεση με άλλους τύπους, πραγματοποιείται χρησιμοποιώντας κανόνες και συστάσεις που έχουν επινοηθεί από τον αρχικό κατασκευαστή και όχι από την ομάδα συντήρησης. Αυτές οι προτάσεις βασίζονται σε δεδομένα που λαμβάνονται από πειράματα.

Συνήθως, όταν αγοράζεται ο εξοπλισμός, ο κατασκευαστής προσφέρει στατιστικά στοιχεία και οδηγίες, οι οποίες θα παρέχουν λεπτομέρειες σχετικά με την αναμενόμενη διάρκεια ζωής ολόκληρου του μηχανήματος, αλλά και των διαφορετικών εξαρτημάτων του. Ο κατασκευαστής συμβουλεύει για το πόσο συχνά πρέπει να επιθεωρείται, να συντηρείται και να αντικαθίσταται το κάθε εξάρτημα.

Επειδή οι τεχνικοί ενδέχεται να μην είναι σε θέση να προβλέψουν αστοχίες που μπορεί να προκύψουν, η συντήρηση που γίνεται βρίσκεται σε αποκλειστική εξάρτηση από ένα προκαθορισμένο πρόγραμμα, το οποίο μπορεί να οδηγήσει στη μη σωστή λειτουργία της συσκευής. Μπορεί επίσης να οδηγηθούν σε πολλαπλές επισκευές και αντικαταστάσεις ανταλλακτικών πολύ σύντομα, με αποτέλεσμα το επιπλέον κόστος. Δεδομένου ότι το πρόγραμμα συντήρησης του κατασκευαστή επικεντρώνεται σε μέσους όρους και όχι στην πραγματική κατάσταση του εξοπλισμού, η προκαθορισμένη συντήρηση δεν εγγυάται ότι μια συσκευή δε θα χαλάσει.

1.2.2 Διορθωτική συντήρηση (Reactive maintenance)

Η πιο βασική τεχνική συντήρησης αναφέρεται ως διορθωτική συντήρηση ή συντήρηση βλάβης. Αυτή συμβαίνει όταν το μηχάνημα ή το σύστημα παραγωγής δουλεύει χωρίς κάποια δυσλειτουργία και ένα μέρος του αποτυγχάνει. Για το λόγο αυτό, γίνεται η διορθωτική συντήρηση με σκοπό την επισκευή του ή την αντικατάστασή του με στόχο την επαναφορά της πλήρους λειτουργικότητας.

Εάν άλλες μορφές συντήρησης δεν κατάφεραν να προβλέψουν τη δυσλειτουργία που διαταράσσει τη διαδικασία παραγωγής, οι τεχνικοί βιομηχανικής συντήρησης αντιμετωπίζουν το ζήτημα μόλις συμβεί. Πρόκειται για συντήρηση που δεν είναι αναμενόμενη ή προγραμματισμένη.

1.2.3 Προληπτική συντήρηση (Preventive maintenance)

Η ψηφιοποίηση των βιομηχανικών επιχειρήσεων είχε ως αποτέλεσμα μια πληθώρα μηχανογραφικών και τεχνολογικών λύσεων που επιτρέπουν στους τεχνικούς να διεξάγουν, να παρακολουθούν και να σχεδιάζουν αποτελεσματικά την προληπτική συντήρηση. Μέχρι να προκύψει οποιαδήποτε βλάβη ή δυσλειτουργία, οι τεχνικοί που είναι υπεύθυνοι για τη βιομηχανική συντήρηση ασχολούνται με την παραγγελία ανταλλακτικών, υλικών και εξοπλισμού που σκοπό έχουν τη μείωση του κινδύνου αστοχίας.

Ο βασικός στόχος της προληπτικής συντήρησης είναι ο μηδενικός χρόνος διακοπής της λειτουργίας. Μέσω της συλλογής δεδομένων και τον τακτικό έλεγχο του εξοπλισμού εντοπίζονται δευτερεύοντα ζητήματα και διορθώνονται προτού γίνουν σημαντικά. Τελικοί σκοποί της προληπτικής συντήρησης είναι η αύξηση της παραγωγικότητας του εξοπλισμού, η μείωση των κρίσιμων βλαβών και η ελαχιστοποίηση των απωλειών από την έξοδο της παραγωγικής διαδικασίας λόγω βλάβης.

1.2.4 Συντήρηση βάσει συνθηκών (Condition-based maintenance)

Η συντήρηση βάσει συνθηκών είναι μια τεχνική συντήρησης που αξιολογεί την τρέχουσα κατάσταση ενός μηχανήματος για να αποφασίσει ποιο είδος συντήρησης απαιτείται. Με αυτήν την τεχνική, υποχρεώνεται η όποια εργασία συντήρησης να πραγματοποιηθεί μόνο όταν οι δείκτες δεδομένων σε πραγματικό χρόνο υποδεικνύουν ανωμαλίες ή σημάδια φθίνουσων αποτελεσμάτων.

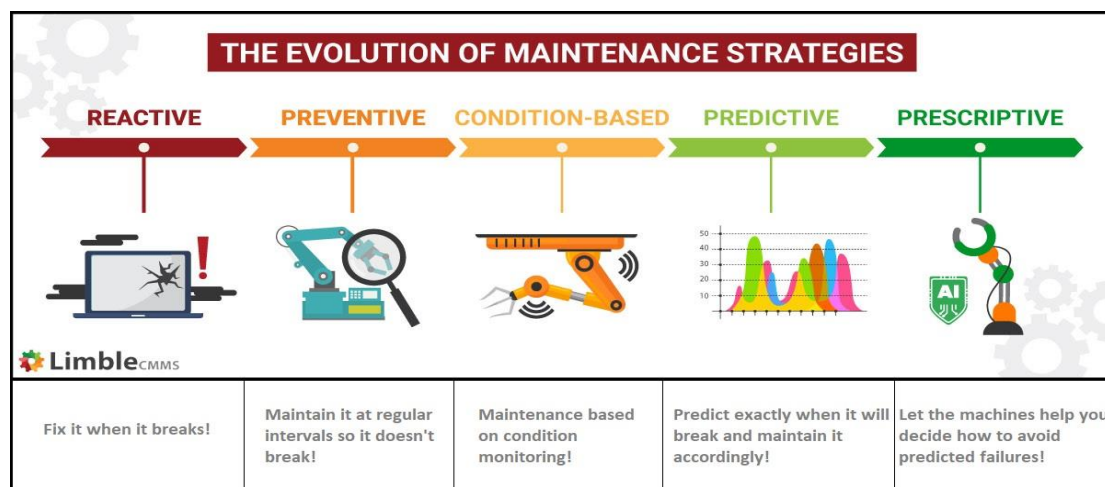
Στόχος αυτής της τεχνικής είναι η συνεχής παρακολούθηση των δεδομένων, έτσι ώστε η διορθωτική ενέργεια συντήρησης να μπορεί να προγραμματιστεί εκ των προτέρων. Η ιδέα είναι ότι μέσω αυτής της παρακολούθησης σε πραγματικό χρόνο, επιτρέπεται στις ομάδες συντήρησης να έχουν αρκετό χρόνο αντίδρασης πριν από μια δυσλειτουργία. Η πιο δημοφιλής μέθοδος για τη συλλογή δεδομένων για ανάλυση είναι μέσω αισθητήρων.

1.2.5 Προβλεπτική συντήρηση (Predictive maintenance)

Αυτός ο τύπος βιομηχανικής συντήρησης επιτρέπει στις επιχειρήσεις να προβλέπουν ζητήματα που προκύπτουν και να προγραμματίζουν τις απαραίτητες παρεμβάσεις και διαδικασίες συντήρησης. Χάρη στην εμφάνιση λύσεων επεξεργασίας και ανάλυσης δεδομένων, καθώς και τεχνητής νοημοσύνης, οι τεχνικοί συντήρησης μπορούν να προγραμματίσουν την προβλεπτική συντήρηση με βάση την πρόβλεψη σφαλμάτων και δυσλειτουργιών.

Η προβλεπτική συντήρηση παρακολουθεί την απόδοση και την κατάσταση του εξοπλισμού όταν χρησιμοποιείται. Ο σκοπός της είναι να είναι σε θέση να προβλέψει πότε ο εξοπλισμός ενδέχεται να δυσλειτουργήσει (με βάση διάφορους παράγοντες) και στη συνέχεια να μειώνει τον κίνδυνο βλάβης αποφεύγοντας την αστοχία. Η παρακολούθηση της κατάστασης, η οποία χαρακτηρίζεται ως η συνεχής παρακολούθηση των μηχανημάτων κατά τη διάρκεια των συνθηκών διεργασίας για τη διασφάλιση τη βέλτιστης χρήσης του μηχανήματος, είναι προαπαιτούμενο για τη σωστή πρόβλεψη.

Με τη χρήση της τεχνητής νοημοσύνης γεννήθηκαν και νέες δυνατότητες στην προβλεπτική συντήρηση, καθώς με τη βοήθεια της ανάλυσης των δεδομένων, όχι μόνο πραγματοποιείται η πρόβλεψη μιας ενδεχόμενης αστοχίας, αλλά δίνονται και συστάσεις για την προσαρμογή των συνθηκών λειτουργίας έτσι ώστε να επιτευχθούν τα επιθυμητά αποτελέσματα παραγωγής. Αυτή η στρατηγική συντήρησης ορίζεται ως prescriptive maintenance.



Εικόνα 2 Η εξέλιξη των στρατηγικών συντήρησης

Κεφάλαιο 2: Σύγκριση στρατηγικών συντήρησης

Στο συγκεκριμένο κεφάλαιο παρουσιάζονται τα πλεονεκτήματα και τα μειονεκτήματα της κάθε στρατηγικής συντήρησης που αναλύθηκε νωρίτερα. Σκοπός αυτού του κεφαλαίου είναι η σύγκριση των στρατηγικών συντήρησης, έτσι ώστε ο αναγνώστης να αποκτήσει μία πιο σφαιρική άποψη.

2.1 Πλεονεκτήματα - Μειονεκτήματα

Στην **προκαθορισμένη συντήρηση** η αποτελεσματικότητα βασίζεται στην προϋπόθεση ότι μπορεί να αναμένεται εμπλοκή ενός μηχανήματος ή βλάβη. Η διάρκεια του χρόνου υπολογίζεται εξετάζοντας τα ιστορικά δεδομένα ενός μηχανήματος και ακολουθώντας τις οδηγίες του κατασκευαστή που ενδέχεται να ισχύουν.

Το πλεονέκτημα της προκαθορισμένης συντήρησης είναι ότι εξαλείφει την ανάγκη εξωτερικού εξοπλισμού που δείχνει όταν ένα μηχάνημα ή μέρος του απαιτεί προσοχή για πιθανή βλάβη. Επίσης, η καθιέρωση τακτικού προγραμματισμού συντήρησης είναι σχετικά φθηνή.

Στα μειονεκτήματα αυτής της στρατηγικής συντήρησης είναι ότι υπάρχει αυξημένη πιθανότητα η εργασία που εκτελείται στο μηχάνημα να είναι περιττή. Υπολογίζεται ότι το 30% των προκαθορισμένων συντηρήσεων ήταν μόνο προληπτικό και επομένως περιττό. Ένα άλλο μειονέκτημα είναι η πιθανότητα εμφάνισης αστοχίας κάποιου στοιχείου ανά πάσα στιγμή. Ορισμένες δυσλειτουργίες είναι δύσκολο να προβλεφθούν και ορισμένα κομμάτια δε φθείρονται πάντα σύμφωνα με τις προσδοκίες.

Η **διορθωτική συντήρηση** βασίζεται στην υπόθεση πως όταν ένα κομμάτι του μηχανήματος αστοχήσει, πρέπει να φτιαχτεί ή να αντικατασταθεί. Με άλλα λόγια, είναι η τεχνική εργασία που πραγματοποιείται μετά από μια αποτυχία με στόχο την επαναφορά του στοιχείου σε κατάσταση όπου μπορεί να λειτουργεί κανονικά, είτε με επισκευή, είτε με αντικατάσταση.

Το πλεονέκτημα αυτής της μεθόδου είναι ότι θεωρείται η κατάλληλη για εξοπλισμό χαμηλής προτεραιότητας, δηλαδή εξαρτήματα που η αστοχία τους δε θα διαταράξουν σημαντικά την παραγωγική λειτουργία του μηχανήματος. Ο χρόνος και η προσπάθεια που απαιτείται για τη συντήρηση ή τη συνεχή παρακολούθηση ενδέχεται να είναι πιο δαπανηρά από την επισκευή ή την αντικατάστασή τους όταν καταρρέουν. Δεδομένου ότι αυτή η μέθοδος απαιτεί σχετικά μικρή προετοιμασία, έχει πολύ χαμηλό κόστος εφαρμογής σε σύγκριση με τις εναλλακτικές τεχνικές συντήρησης.

Το μειονέκτημα αυτής της μεθόδου προκύπτει αν οι υπεύθυνοι συντήρησης στηριχθούν στη διορθωτική συντήρηση για στοιχεία μεσαίας ή υψηλής προτεραιότητας. Δεδομένου ότι δε λαμβάνονται προληπτικά μέτρα, η διάρκεια ζωής του εξοπλισμού θα μειωθεί ταχύτερα από το αναμενόμενο. Ειδικότερα, εάν χρησιμοποιείται σε εξοπλισμό υψηλής προτεραιότητας ή υψηλής αξίας, κάτι το οποίο είναι απαγορευτικό, θα μπορούσε να οδηγήσει σε μη προγραμματισμένο χρόνο διακοπής λειτουργίας και ενδεχομένως υψηλό κόστος επισκευής.

Η συντήρηση βάσει συνθηκών βασίζεται στην παρακολούθηση της κατάστασης του εξοπλισμού και στην αξιολόγησή του. Είναι σημαντική για την ελαχιστοποίηση κάποιας κρίσιμης δυσλειτουργίας του εξοπλισμού ή και διακοπή της λειτουργίας του. Η διαδικασία εύρεσης πιθανών προβλημάτων του εξοπλισμού διασφαλίζει τη λειτουργία του. Είναι γνωστή και ως συντήρηση με επίκεντρο την αξιοπιστία.

Το βασικό πλεονέκτημα αυτής της μεθόδου είναι ότι τα εξαρτήματα που εμφανίζουν στοιχεία πιθανής βλάβης μπορούν να επισκευαστούν ή να αντικατασταθούν προτού συμβεί η αστοχία τους και επομένως το μακροπρόθεσμο κόστος είναι πολύ χαμηλό καθώς ο εξοπλισμός θα συνεχίσει να λειτουργεί με το λιγότερο δυνατό χρόνο διακοπής. Αυτό οδηγεί στην ελαχιστοποίηση της διακοπής παραγωγής, αλλά και στον καλύτερο προγραμματισμό του χρόνου συντήρησης.

Το βασικότερο μειονέκτημα για την επίτευξη της συντήρησης βάσει συνθηκών έχει να κάνει τόσο με το υψηλό αρχικό κόστος που αφορά είτε την εγκατάσταση του σωστού εξοπλισμού αισθητήρων, είτε την εκπαίδευση του ανθρώπινου δυναμικού που θα αναλάβει την πραγματοποίησή της. Επίσης, μπορεί να χρειαστούν διάφορες τροποποιήσεις στον εξοπλισμό.

Ο στόχος της **προληπτική συντήρησης** είναι να αποτρέψει την καταστροφή διαφόρων στοιχείων του εξοπλισμού πριν αυτά αποτύχουν και θέσουν το μηχάνημα εκτός παραγωγικής διαδικασίας. Για να συμβεί αυτό, πραγματοποιείται ένα σύνολο διαδικασιών με τακτικό έλεγχο του εξοπλισμού για τον εντοπισμό δευτερευόντων ζητημάτων και τη διόρθωσή τους.

Στα πλεονεκτήματα αυτής της μεθόδου συγκαταλέγεται η διατήρηση του εξοπλισμού για μεγαλύτερο χρονικό διάστημα σε σχέση με τις περισσότερες στρατηγικές συντήρησης. Το μακροπρόθεσμο κόστος επισκευής είναι συνήθως πολύ χαμηλότερο, καθώς η πιθανότητα μιας καταστροφικής αποτυχίας μειώνεται σημαντικά.

Το μειονέκτημα της προληπτικής συντήρησης είναι η δυσκολία πραγματοποίησης της από ότι με άλλους τύπους συντήρησης, αφού απαιτείται η σωστή εκπαίδευση των τεχνικών συντήρησης, αλλά και χρόνος για την εφαρμογή των συγκεκριμένων διαδικασιών συντήρησης.

Στην **προβλεπτική συντήρηση**, οι μηχανικοί που είναι υπεύθυνοι για τη συντήρηση του βιομηχανικού εξοπλισμού, χρησιμοποιούν μια προγνωστική στρατηγική συντήρησης για να προβλέψουν πότε ακριβώς θα αποτύχει κάποιο μέρος του εξοπλισμού και τότε να εκτελέσουν τη συντήρησή του με σκοπό τη μέγιστη διατήρηση της λειτουργίας των μηχανών παραγωγής. Αυτό διασφαλίζει ότι ένα κομμάτι του εξοπλισμού που χρειάζεται συντήρηση απενεργοποιείται μόνο λίγο πριν αποτύχει, επιτρέποντας έτσι ο εξοπλισμός να παραμείνει λειτουργικός για το μέγιστο εφικτό διάστημα.

Το μεγάλο πλεονέκτημα αυτής της στρατηγικής συντήρησης είναι η εξοικονόμηση του κόστους που επιτυγχάνεται, καθώς μειώνεται ο μη προγραμματισμένος χρόνος διακοπής λειτουργίας και αυξάνεται ο χρόνος παραγωγής. Παρέχονται στατιστικά δεδομένα σε πραγματικό χρόνο σχετικά με την τρέχουσα κατάσταση των στοιχείων του εξοπλισμού και έτσι διασφαλίζονται ελάχιστες διαταραχές παραγωγικότητας. Επίσης, βελτιστοποιείται όχι μόνο η χρήση ανταλλακτικών, αλλά και ο χρόνος που αφιερώνεται κατά την εκτέλεση εργασιών συντήρησης.

Η προβλεπτική συντήρηση θεωρείται ως η πιο απαιτητική στρατηγική συντήρησης διότι απαιτείται η αγορά υψηλής ακρίβειας εξοπλισμού, αλλά και το κατάλληλο λογισμικό που μπορεί να υποστηρίξει τα δεδομένα που προκύπτουν κατά τη λειτουργία του εξοπλισμού. Τέλος, ένα ακόμη μειονέκτημα είναι η απαίτηση για εξειδικευμένη εκπαίδευση που σκοπό έχει την ερμηνεία και την ανάλυση των δεδομένων παρακολούθησης, κάτι το οποίο έχει υψηλό κόστος.

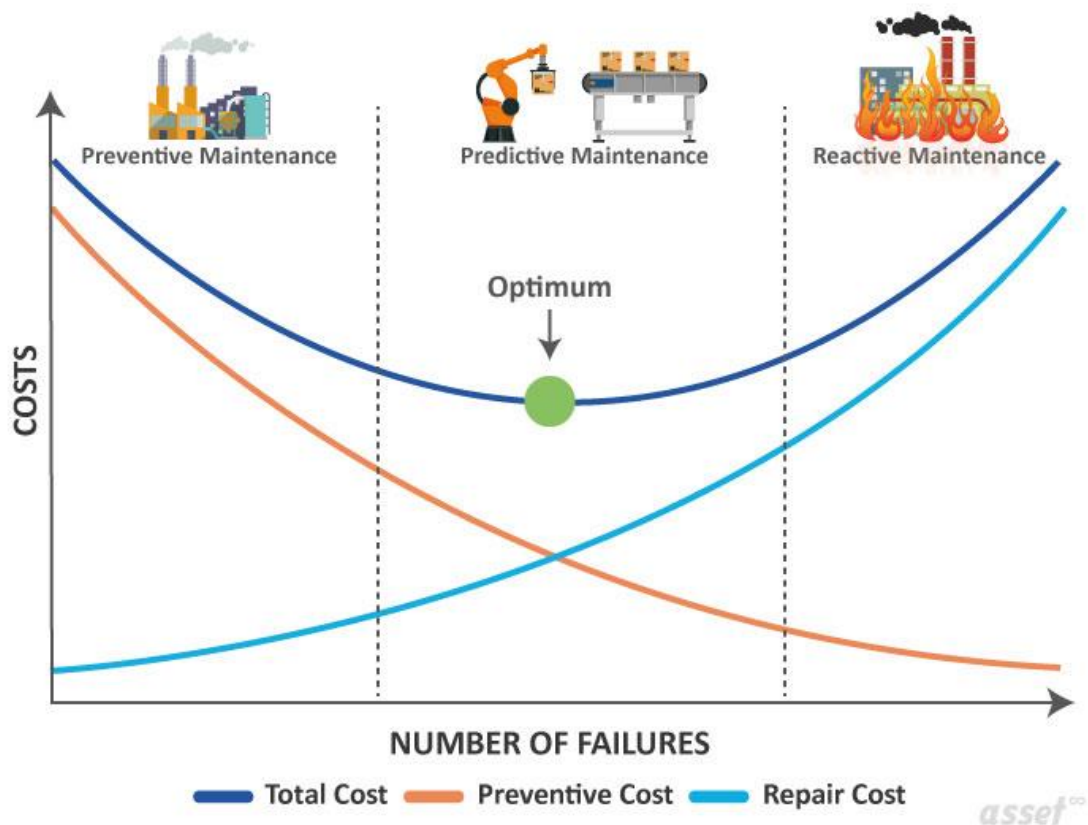
2.2 Κόστος συντήρησης

Το κόστος συντήρησης περιλαμβάνει το χρηματικό ποσό που δαπανάται από μια επιχείρηση για τη συντήρηση και την επισκευή του εξοπλισμού της που προκύπτει από τυχόν βλάβες σε όλη τη διάρκεια της λειτουργίας του. Τα μηχανήματα και τα εργαλεία παραγωγής είναι η ζωτική δύναμη κάθε εταιρείας παραγωγής και πρέπει να συντηρούνται σε τακτική βάση για να παραμένουν σε άριστη κατάσταση. Παρά τα οφέλη του, ο εξοπλισμός που απαιτεί υπερβολική συντήρηση μπορεί να είναι επιβλαβής για την επιχείρηση.

Η κατανόηση της συντήρησης, της φθοράς και των δαπανών επισκευής που προκύπτουν βοηθάει στη λήψη αποφάσεων σχετικά με το πραγματικό κόστος λειτουργίας του κάθε μηχανήματος. Τα άμεσα κόστη μιας παραγωγικής μονάδας αποτελούνται από το ανθρώπινο δυναμικό που δραστηριοποιείται στην παραγωγική διαδικασία και από τις πρώτες ύλες που χρησιμοποιούνται, ενώ το έμμεσο κόστος αποτελείται από το ανθρώπινο δυναμικό που εργάζεται στο τμήμα συντήρησης και τις όποιες ανάγκες προκύπτουν για αυτό το τμήμα, επομένως το κόστος συντήρησης καθορίζεται ως το έμμεσο κόστος της επιχείρησης.

Ειδικότερα, για την αξιολόγηση του πραγματικού κόστους ενός προγράμματος προβλεπτικής συντήρησης, είναι δύσκολο να υπολογιστεί το πραγματικό κέρδος αφού η μείωση του κόστους συντήρησης προκύπτει από τις πιθανές αστοχίες του εξοπλισμού που αποφεύχθηκαν. Βασίζεται δηλαδή σε γεγονότα που δε συνέβησαν ποτέ, άρα δεν υπάρχουν και πραγματικά κέρδη, αλλά αυτά αντιπροσωπεύονται από έναν άπειρο αριθμό πιθανών σεναρίων βλαβών ή αστοχιών του εξοπλισμού.

Συγκρίνοντας το κόστος της προβλεπτικής συντήρησης σε σχέση με δύο άλλες βασικές στρατηγικές συντήρησης, όπως της προληπτικής συντήρησης και της διορθωτικής συντήρησης, φαίνεται ξεκάθαρα ότι το κόστος που αντιστοιχεί στην προβλεπτική συντήρηση είναι σαφώς μικρότερο αφού οι δραστηριότητες συντήρησης πραγματοποιούνται μόνο όταν είναι δικαιολογημένες και έχουν προηγουμένως προβλεφθεί. Επομένως, αποφεύγονται και τα έξοδα τα οποία θα είχαν γίνει σε προληπτικό επίπεδο και κατά ένα ποσοστό δε θα είχαν οικονομικό αντίκρισμα κέρδους, αλλά και τα έξοδα που θα προκύπταν από τη μη λειτουργία της παραγωγικής διαδικασίας λόγω κάποιας αστοχίας στον εξοπλισμό.



Εικόνα 3 Σχέση κόστους - αριθμού βλαβών

Κεφάλαιο 3: Βιομηχανική επανάσταση

Η ιστορία έχει δείξει ότι οι άνθρωποι εξαρτιόντουσαν πάντα από την τεχνολογία της κάθε εποχής. Όπως είναι φυσικό, η τεχνολογία της κάθε εποχής μπορεί να μην έχει το ίδιο σχήμα και μέγεθος όπως σήμερα, αλλά για την εποχή τους ήταν σίγουρα κάτι σημαντικό. Οι άνθρωποι πάντα χρησιμοποιούν την τεχνολογία που έχουν στη διάθεσή τους για να κάνουν τη ζωή τους ευκολότερη και ταυτόχρονα να προσπαθούν να την τελειοποιήσουν και τελικά να την φτάσουν στο επόμενο στάδιο. Έτσι ξεκίνησε η έννοια της βιομηχανικής επανάστασης.

3.1 Πρώτη βιομηχανική επανάσταση

Η πρώτη βιομηχανική επανάσταση ακολούθησε την πρωτο-βιομηχανική περίοδο με αφετηρία της το 1784. Η μηχανοποίηση ήταν η αιτία που η γεωργία άρχισε να δίνει τη θέση της στη βιομηχανία ως η ραχοκοκαλιά της οικονομίας. Την εποχή εκείνη, υπήρξε η μαζική εξόρυξη άνθρακα και η πολύ σημαντική εφεύρεση της ατμομηχανής. Οι εγκαταστάσεις παραγωγής ήταν εξοπλισμένες με μηχανικά συστήματα παραγωγής ατμού ή νερού και παρατηρήθηκε σημαντική αύξηση της παραγωγικότητας με τη χρήση αυτών των βιομηχανικών μηχανημάτων και των νέων τεχνολογιών ισχύος.

3.2 Δεύτερη βιομηχανική επανάσταση

Σχεδόν έναν αιώνα μετά την πρώτη βιομηχανική επανάσταση, και πιο συγκεκριμένα το 1870, έκανε την εμφάνισή της η δεύτερη βιομηχανική επανάσταση. Χαρακτηριστικές ήταν οι μεγάλες τεχνολογικές εξελίξεις στον τομέα των βιομηχανιών και οι οποίες βοήθησαν στην εμφάνιση μιας νέας πηγής ενέργειας, τον ηλεκτρισμό. Η ηλεκτρική ενέργεια, που παραγόταν με τη χρήση φυσικού αερίου και πετρελαίου, υπήρξε η κύρια πηγή ισχύος κατά τη διάρκεια αυτής της τεχνολογικής επανάστασης. Αποτέλεσμα της δεύτερης βιομηχανικής επανάστασης ήταν η εφεύρεση του κινητήρα

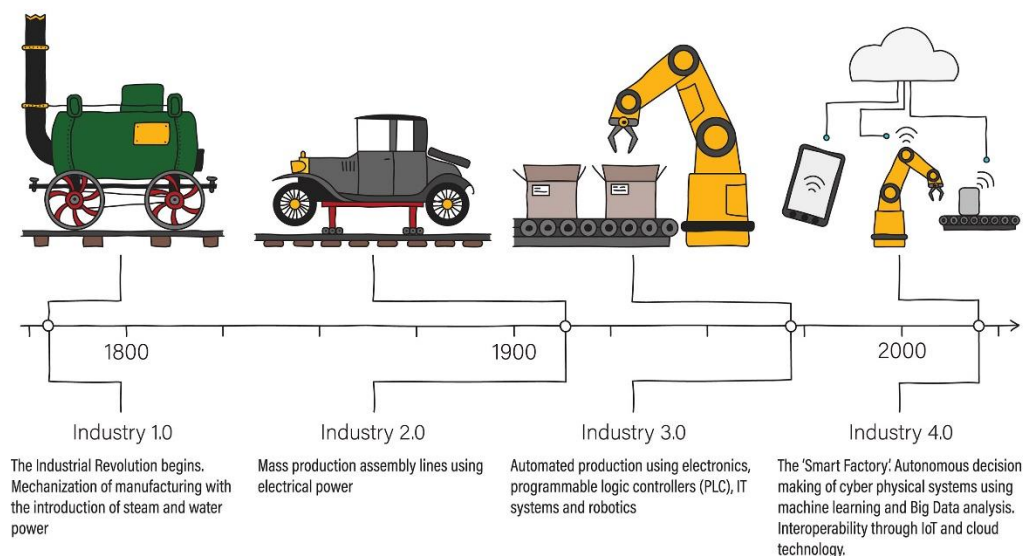
εσωτερικής καύσης και η πλήρης αξιοποίηση των δυνατοτήτων του. Άλλες σημαντικές εφευρέσεις είχαν να κάνουν με τις μεθόδους επικοινωνίας, όπως ο τηλεγράφος και το τηλέφωνο. Στις αρχές του 20ού αιώνα εφευρέθηκαν το αυτοκίνητο και το αεροπλάνο, τα οποία είναι και ο λόγος που η δεύτερη βιομηχανική επανάσταση θεωρείται μέχρι σήμερα ως η πιο σημαντική.

3.3 Τρίτη βιομηχανική επανάσταση

Στο δεύτερο μισό του 20ού αιώνα, το 1969, ξεκίνησε η τρίτη βιομηχανική επανάσταση, όπου έκανε την εμφάνισή της μία ακόμα πηγή ανεκμετάλλευτης τότε ενέργειας, η πυρηνική ενέργεια. Σημειώθηκε μεγάλη ανάπτυξη στον τομέα των ηλεκτρονικών, των τηλεπικοινωνιών και της πληροφορικής. Η υπολογιστική δύναμη θεωρείται το κλειδί αυτής της βιομηχανικής επανάστασης, καθώς άνοιξε τις πόρτες για τις διαστημικές αποστολές, την έρευνα και τη βιοτεχνολογία. Σε βιομηχανικό επίπεδο υπήρξαν δύο αξιοσημείωτες εφευρέσεις, οι προγραμματιζόμενοι ελεγκτές και τα ρομπότ που συνείσφεραν στη δημιουργία μιας εποχής υψηλού επιπέδου αυτοματισμού.

3.4 Τέταρτη βιομηχανική επανάσταση

Η τέταρτη βιομηχανική επανάσταση ξεκίνησε στις αρχές της τρίτης χιλιετίας και χρησιμοποιεί ως βασική της τεχνολογία, μία τεχνολογία που είναι ευρέως γνωστή σε όλο τον κόσμο, το διαδίκτυο. Είναι η βιομηχανική επανάσταση που ζούμε σήμερα και η παγκόσμια οικονομία βασίζεται πάνω της. Οι διαθέσιμες τεχνολογίες της κάθε εποχής αποτελούν το καύσιμο για τις βιομηχανικές επαναστάσεις. Έτσι, και η τέταρτη βιομηχανική επανάσταση επωφελείται από τις τεχνολογικές τάσεις της εποχής, που στην πραγματικότητα είναι τεχνολογίες πληροφοριών, όπως το διαδίκτυο των πραγμάτων (Internet of Things), το υπολογιστικό νέφος (Cloud Computing), τα μεγάλα δεδομένα (Big Data), η ρομποτική, η τεχνητή νοημοσύνη (Artificial Intelligence), αλλά και η τρισδιάστατη εκτύπωση (3D Printing). Με τη βοήθεια αυτών των τεχνολογιών, πραγματοποιείται η τεχνολογική εξέλιξη από τα ενσωματωμένα συστήματα σε κυβερνο-φυσικά συστήματα, με τελικό στόχο τη μαζική παραγωγή.



Εικόνα 4 Χρονοδιάγραμμα βιομηχανικών επαναστάσεων (Schwab, 2017)

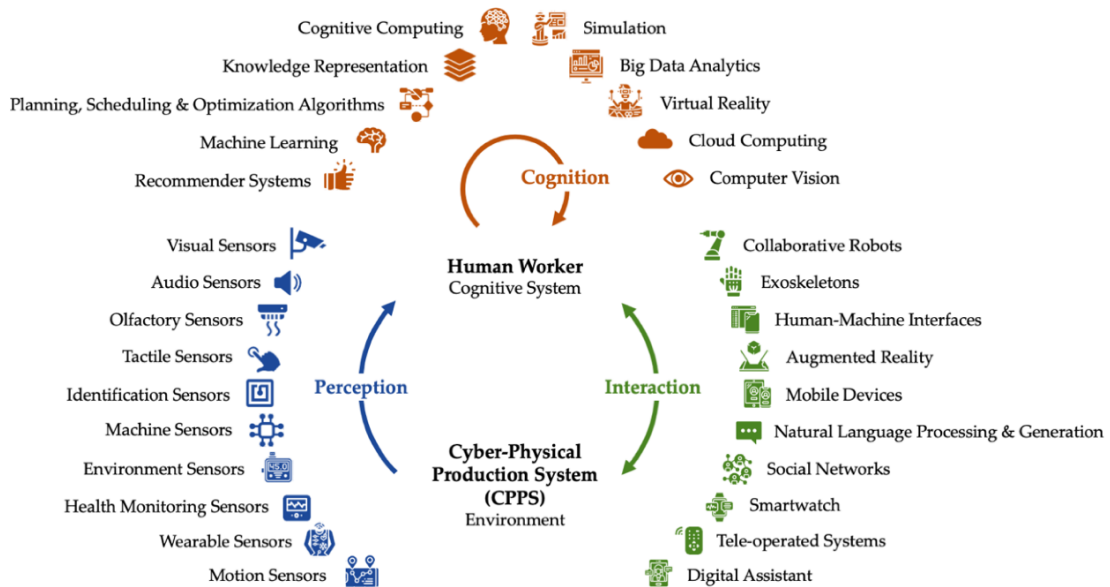
3.5 Πέμπτη βιομηχανική επανάσταση

Η τρίτη και η τέταρτη βιομηχανική επανάσταση ήταν δύσκολες τόσο για τους ανθρώπους, όσο και για το περιβάλλον. Ο άνθρωπος έπρεπε να προσαρμοστεί με βάση τις δυνατότητες των μηχανών. Τη διαφορά σε αυτό κάνει η πέμπτη βιομηχανική επανάσταση (Βιομηχανία 5.0) φέρνοντας το ανθρώπινο είδος όχι μόνο μπροστά, αλλά και στο επίκεντρο της παραγωγικής διαδικασίας, καθώς εστιάζει σε έννοιες όπως η ανθρωποκεντρική προσέγγιση, η βιωσιμότητα και η ανθεκτικότητα. Όσον αφορά την ανθρωποκεντρική προσέγγιση, σημαίνει ότι η τεχνολογία πρέπει να χρησιμοποιείται όχι μόνο στην προσαρμογή της παραγωγικής διαδικασίας πάνω στις ανάγκες του εργαζόμενου, αλλά και στη διασφάλιση ότι η χρήση της δεν παρεμβαίνει στα θεμελιώδη δικαιώματα των εργαζομένων.

Στις μέρες μας χρειάζεται η ανθρώπινη πινελιά της μαζικής εξατομίκευσης, επομένως η πέμπτη βιομηχανική επανάσταση βοηθά στην αλλαγή από τη μαζική παραγωγή στη μαζική εξατομίκευση, καθώς κάνει πραγματικότητα τη μαζική προσαρμογή.

Τα ρομπότ έχουν μια τεράστια συμβολή σε βαριές ή και επαναλαμβανόμενες εργασίες, καθώς επίσης μπορούν να χρησιμοποιηθούν σε τοποθεσίες όπου οι εργαζόμενοι είναι περιορισμένοι να λειτουργούν ή υπάρχει κίνδυνος για την υγεία τους. Αυτή η βιομηχανική επανάσταση έχει να κάνει με την αλληλεπίδραση ανθρώπου και μηχανής με σκοπό την διευκόλυνση, αλλά και την αύξηση της ταχύτητας των εργασιών και σε αυτό συμβάλλει το συνεργατικό ρομπότ, ή το λεγόμενο “COBOT”, το οποίο έχει στόχο να λειτουργεί με τους ανθρώπους και να τους βοηθάει σε ένα πλήθος καθηκόντων.

Σημαντικό όφελος της πέμπτης βιομηχανικής επανάστασης υπάρχει στη βελτιστοποίηση του κόστους, αφού άνθρωπος και μηχανή συνεργάζονται για να λάβουν τις καλύτερες οικονομικές αποφάσεις. Οι αποφάσεις αυτές δεν έχουν μόνο στόχο το οικονομικό όφελος, αλλά επικεντρώνονται στην προστασία του περιβάλλοντος ως προτεραιότητα, παρέχοντας πιο πράσινες λύσεις, σύμφωνες με αυτές που απαιτούν όχι μόνο οι διεθνείς και κυβερνητικοί οργανισμοί, αλλά και οι καταναλωτές. Τέλος, με την αυτοματοποίηση που έχει επιτευχθεί, ο εργαζόμενος απελευθερώνεται από ορισμένες εργασίες και έτσι εστιάζει στη δημιουργία πιο ισχυρών στρατηγικών ή στην εφαρμογή της δημιουργικότητάς τους.



Εικόνα 5 Κύκλος διεργασιών ανθρώπου-μηχανής (Francesco Longo, 2020)

Κεφάλαιο 4: Συλλογή και προεπεξεργασία δεδομένων

Τα ερευνητικά δεδομένα συλλέγονται με τη χρήση διάφορων τεχνικών και για αυτό το λόγο καταλήγουν να είναι σε διάφορες μορφές. Σκοπός αυτού του κεφαλαίου είναι η κατανόηση των ερευνητικών δεδομένων ανάλογα με την κατηγορία και τον τύπο που τα προσδιορίζει και ποια η σημαντικότητά τους για το τελικό επιθυμητό αποτέλεσμα.

4.1 Συλλογή δεδομένων

Τα ερευνητικά δεδομένα που προκύπτουν μπορούν να τοποθετηθούν σε δύο κατηγορίες, τα ποσοτικά και τα ποιοτικά. Οι ποσοτικές και ποιοτικές ερευνητικές μεθοδολογίες διαφέρουν μεταξύ τους με διάφορους τρόπους, συμπεριλαμβανομένου του τρόπου απόκτησης και αξιολόγησης των δεδομένων, καθώς και του είδους των πληροφοριών που έχει τη δυνατότητα να προσφέρει κάθε μέθοδος.

Όταν ένας ερευνητής προσπαθεί να ποσοτικοποιήσει ένα ζήτημα ενός ερευνητικού θέματος, χρησιμοποιούνται ποσοτικά δεδομένα. Είναι πληροφορίες που μπορούν να μετρηθούν ή να συγκριθούν αριθμητικά. Τα ποσοτικά δεδομένα αναλύονται τις περισσότερες φορές μέσω της χρήσης λογισμικών στατιστικής.

Τα ποιοτικά δεδομένα είναι ένας τύπος πληροφοριών που περιγράφει τα χαρακτηριστικά. Συνήθως τέτοιου είδους δεδομένα συγκεντρώνονται από έρευνες, συνεντεύξεις ή παρατηρήσεις. Αυτός είναι και ο λόγος που τα ποιοτικά δεδομένα μπορεί να τύχουν δυσκολίας στη μέτρηση και τη σωστή αξιολόγησή τους. Η κωδικοποίησή τους βοηθά τον ερευνητή να κατηγοριοποιήσει τα ποιοτικά δεδομένα και να κάνει την ποσοτική ανάλυσή τους, έτσι ώστε να μπορεί να τα εκμεταλλευτεί κατάλληλα για τους ερευνητικούς στόχους που έχει ορίσει.

Με βάση τις τεχνικές συλλογής, τα δεδομένα μπορούν να ταξινομηθούν σε τέσσερις βασικές κατηγορίες:

- Δεδομένα παρατήρησης
- Πειραματικά δεδομένα
- Δεδομένα προσομοίωσης
- Παράγωγα / συμμορφωμένα δεδομένα



Εικόνα 6 Οι τέσσερις τύποι των ερευνητικών δεδομένων

Τα δεδομένα παρατήρησης συλλέγονται παρακολουθώντας μία διαδικασία κατά τη λειτουργία της. Συλλέγονται με μεθόδους όπως η χρήση αισθητήρων ή άλλων οργάνων για την καταγραφή των δεδομένων, με την ανθρώπινη παρατήρηση ή και μέσω κάποιας έρευνας. Η καταγραφή τους γίνεται σε πραγματικό χρόνο και αυτός είναι ο λόγος που η εκ νέου δημιουργία τους θα ήταν εξαιρετικά δύσκολη, αν όχι αδύνατη.

Τα πειραματικά δεδομένα συγκεντρώνονται χάρη στην ενεργή συμμετοχή του ερευνητή, ο οποίος θέλει να προκαλέσει και να ποσοτικοποιήσει την όποια αλλαγή προκύψει στα δεδομένα ή να δημιουργήσει διαφορά ανάμεσα στις μεταβλητές. Αν και μπορεί να φανεί μια δαπανηρή διαδικασία, αυτός ο τύπος δεδομένων είναι συχνά αναπαραγωγίμος.

Χρησιμοποιώντας διάφορα μοντέλα δοκιμών με τη χρήση της υπολογιστικής ισχύς για την προσομοίωση της λειτουργίας μιας πραγματικής διαδικασίας σε σχέση με το χρόνο, δημιουργούνται τα δεδομένα προσομοίωσης ενός συστήματος. Αυτή η προσέγγιση συνήθως χρησιμοποιείται για την πρόβλεψη ενός συμβάντος σε μια δεδομένη κατάσταση. Το μοντέλο δοκιμής που προκύπτει είναι συχνά ακόμη και περισσότερο σημαντικό από τα δεδομένα που παράγονται από την προσομοίωση.

Μέσω της μετατροπής των υπάρχοντων σημείων δεδομένων που προκύπτουν από διάφορες πηγές, δημιουργούνται τα παράγωγα δεδομένα. Δηλαδή είναι νέα δεδομένα τα οποία αποκτούνται με τη χρήση μαθηματικών τύπων. Αυτό το είδος δεδομένων μπορεί να αποκατασταθεί, εάν συμβεί κάτι και χαθεί, αλλά μια τέτοια αναδημιουργία δεδομένων μπορεί να αποβεί χρονοβόρα και ίσως να είναι και μια δαπανηρή διαδικασία.

4.2 Προεπεξεργασία δεδομένων

Όταν υπάρχει ένα σύνολο δεδομένων μέσω της εξόρυξης δεδομένων (Data mining), είναι δυνατόν να γίνει αντιληπτή η σχέση που υπάρχει μεταξύ των μεταβλητών. Αν δηλαδή, κάποια μεταβλητή μπορεί να είναι παρόμοια με κάποια άλλη ή να είναι πιο ζωτικής σημασίας για το αποτέλεσμα που θέλουμε να επιτύχουμε. Επίσης, η εξόρυξη δεδομένων είναι πολύ χρήσιμη και αναγκαία, καθώς εξυπηρετεί τις προβλέψεις των δεδομένων. Γενικότερα, ο άνθρωπος τρομάζει με το άγνωστο και για αυτό το λόγο προσπαθεί να είναι ένα βήμα μπροστά.

Σημαντικές επίσης είναι, οι επαγωγικές σχέσεις που υπάρχουν μεταξύ των μεταβλητών. Αυτό σημαίνει ότι όταν επηρεάζεται μία μεταβλητή, επόμενο είναι να επηρεαστεί και μία άλλη. Αυτές είναι παρατηρήσεις που μπορούν να γίνουν αντιληπτές μέσω της δυνατότητας ανάλυσης των δεδομένων που μας προσφέρονται από την εξόρυξη δεδομένων. Στις προκλήσεις που προκύπτουν είναι η εύρεση των σχέσεων μεταξύ των δεδομένων, αλλά και η ταχύτητα με την οποία θα επιτευχθούν αυτές οι σχέσεις. Η εγκυρότητα αυτών των σχέσεων στηρίζεται στην ορθή εκμετάλλευση των σωστών εργαλείων και μοντέλων, για κάθε περίπτωση δεδομένων.

Η προεπεξεργασία των δεδομένων αποτελείται από ένα σύνολο διαδικασιών, οι οποίες βοηθούν στην απόκτηση μιας καλύτερης εικόνας των δεδομένων. Μια εικόνα που βοηθάει ακόμα περισσότερο στην επίτευξη του στόχου, πιο εύκολα και πιο γρήγορα. Οι διαδικασίες αυτές είναι οι εξής:

- Καθαρισμός δεδομένων.
Συμπληρώνονται κάποιες κενές καταχωρήσεις, διότι δημιουργούν θόρυβο
- Η ενσωμάτωση.
Εναρμονίζονται όλες οι πηγές των δεδομένων σε μία, έτσι ώστε να έχουν πιο εύκολη χρήση
- Η μείωση των δεδομένων.
Αφορά την ύπαρξη πολύ μεγάλου όγκου δεδομένων, ο οποίος μπορεί να είναι και μη διαχειρίσιμος
- Κανονικοποίηση δεδομένων.
Φτιάχνονται τα δεδομένα με τέτοιο τρόπο ώστε να έχουν όλα την ίδια μορφή, για παράδειγμα τις ίδιες μονάδες μέτρησης

Το πιο συχνό πρόβλημα στα δεδομένα που αντιμετωπίζεται από τους αναλυτές, είναι η ύπαρξη θορύβου. Με τον όρο θόρυβο στα δεδομένα, εννοείται η ύπαρξη λανθασμένων τιμών μέσα σε ένα σύνολο δεδομένων που προέκυψαν από την εξόρυξη δεδομένων. Τα δεδομένα που έχουν σφάλματα και ακραίες τιμές, δηλαδή μη χρήσιμες πληροφορίες, χαρακτηρίζονται ως θορυβώδη και μπορούν να προκαλέσουν σύγχυση στους αλγόριθμους εξόρυξης, γι' αυτό και πρέπει να αντιμετωπιστούν στο στάδιο της προεπεξεργασίας.

Κεφάλαιο 5: Μηχανική μάθηση

Η μηχανική μάθηση περιλαμβάνει τον προγραμματισμό ενός υπολογιστή για την αξιολόγηση των δεδομένων σε σχέση με τα δείγματα. Οι μέθοδοι μηχανικής μάθησης προσπαθούν να προσδιορίσουν τη σχέση μεταξύ του διανύσματος εισόδου και της εξόδου. Ο αλγόριθμος μηχανικής εκμάθησης πρέπει να βελτιστοποιήσει τις παραμέτρους του μοντέλου. Τα τελευταία χρόνια, η μηχανική μάθηση έχει γίνει όλο και πιο αισθητή στην επιστήμη των υπολογιστών, επειδή διευκολύνει τη συλλογή και αποθήκευση δεδομένων. Τα δεδομένα που συλλέγονται είναι συνήθως πολύ εκτεταμένα και η χειροκίνητη ανάλυση των δεδομένων δεν είναι πρακτική. Σε αυτήν την περίπτωση, η τεχνολογία της μηχανικής μάθησης παίζει σημαντικό ρόλο. Ένας ακόμη λόγος για την ταχεία άνοδο της δημοτικότητας της μηχανικής μάθησης είναι η μείωση του υπολογιστικού κόστους. Με την ανάπτυξη του υλικού υπολογιστών τα τελευταία χρόνια, η χρήση μεθόδων μηχανικής μάθησης έχει γίνει αποτελεσματική τόσο χρονικά όσο και οικονομικά, ειδικά στην ανίχνευση σφαλμάτων και στην προγνωστική συντήρηση.

Ο σκοπός κάθε επιστήμης είναι να μελετήσει φαινόμενα και να κατασκευάσει μοντέλα για να περιγράψει όσο το δυνατόν πληρέστερα πειραματικά δεδομένα και συναφή φαινόμενα. Ο όρος μηχανική μάθηση αναφέρεται σε μοντέλα προγραμματισμού (αλγόριθμοι) με ρυθμιζόμενες παραμέτρους που μπορούν να εκπαιδευτούν και να μαθευτούν αυτόματα από τα δεδομένα που τους παρέχονται για τη βελτίωση της συμπεριφοράς τους.

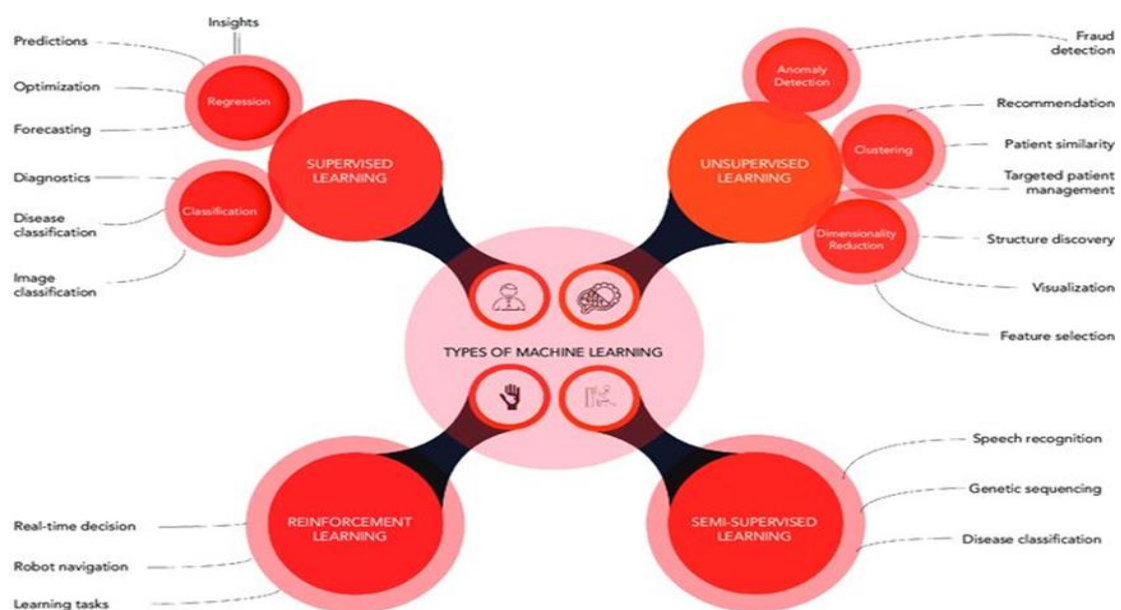
Ο στόχος είναι να δημιουργηθεί ένα μοντέλο αλγορίθμου που μαθαίνει από τα δεδομένα και ανιχνεύει και αποκαλύπτει πιθανά κρυμμένα μοτίβα που σχετίζονται με τον μηχανισμό παραγωγής δεδομένων. Αυτές οι πληροφορίες βοηθούν στη συνέχεια στην ανάλυση και κατανόηση της φύσης των δεδομένων και μπορούν να χρησιμοποιηθούν για την πρόβλεψη φαινομένων που σχετίζονται με αυτά. Εκτός από τη μοντελοποίηση κρυφών δομών δεδομένων, ένας σημαντικός κλάδος της μηχανικής μάθησης είναι η ανάπτυξη αποτελεσματικών αλγορίθμων για σχεδιασμό μοντέλου, ανάλυση και πρόβλεψη. Αντιμετωπίζουμε μεγάλο αριθμό δεδομένων, τα οποία μπορεί να περιέχουν πολύ θόρυβο και να λείπουν τιμές. Η ανάλυση δεδομένων σε παρόμοιες καταστάσεις απαιτεί υπολογιστικά αποτελεσματικούς και ισχυρούς αλγόριθμους.

Επομένως, είναι κατανοητό ότι οι μέθοδοι μηχανικής μάθησης δίνουν τη δυνατότητα σε ερευνητές, επιστήμονες, μηχανικούς και αναλυτές δεδομένων να μάθουν τις τάσεις και τις σχέσεις των δεδομένων για να αντλήσουν αξιόπιστα συμπεράσματα και να τονίσουν τη συνάφεια. Για αυτό το λόγο, αυτές οι μέθοδοι αποτελούν το κέντρο της επιστημονικής έρευνας εδώ και δεκαετίες, σε διάφορους τομείς όπως η τεχνητή νοημοσύνη, η αναγνώριση προτύπων, οι στατιστικές, η εξόρυξη δεδομένων, η επιστήμη των υπολογιστών και η επεξεργασία και ανάλυση σήματος και εικόνας.

5.1 Τύποι μηχανικής μάθησης

Τα προβλήματα μηχανικής μάθησης χωρίζονται συνήθως στις ακόλουθες κατηγορίες, ανάλογα με τη φύση των σημάτων (δείγματα δεδομένων) που παρέχουμε στον αλγόριθμο:

- επιβλεπόμενη μάθηση
- μη επιβλεπόμενη μάθηση
- ενισχυτική μάθηση
- ημι-επιβλεπόμενη μάθηση



Εικόνα 7 Τύποι μηχανικής μάθησης (Karthik Seetharam, 2020)

5.1.1 Επιβλεπόμενη μάθηση

Στην επιβλεπόμενη μάθηση, το σετ εκπαίδευσης που παρέχουμε για τη δημιουργία αλγορίθμου περιλαμβάνει τα χαρακτηριστικά εισόδων και τις αντίστοιχες εξόδους για κάθε δείγμα. Ο στόχος είναι να αφήσουμε τον αλγόριθμο να χρησιμοποιήσει αυτά τα δεδομένα για εκπαίδευση και να μάθει τους γενικούς κανόνες για την αντιστοίχιση της εισόδου με το τελικό αποτέλεσμα. Αυτό δημιουργεί ένα μαθηματικό μοντέλο που μπορεί να χρησιμοποιηθεί για την πρόβλεψη, καθώς όταν του παρέχουμε ένα νέο σύνολο εισόδων, θα είναι σε θέση να προβλέψει το αποτέλεσμα με βάση την εμπειρία που αποκτήθηκε από τα δεδομένα που χρησιμοποιήθηκαν για την εκπαίδευσή του. Υπάρχουν πολλές εφαρμογές για τα εποπτευόμενα μοντέλα μάθησης, από την αναγνώριση προτύπων και την οπτική αναγνώριση ή αναγνώριση ομιλίας έως τη βιοχημεία και το μάρκετινγκ.

5.1.2 Μη επιβλεπόμενη μάθηση

Από την άλλη πλευρά, στη μη επιβλεπόμενη μάθηση, ο αλγόριθμος δεν μπορεί να έχει πρόσβαση στις μεταβλητές εξόδου, αλλά μπορεί να έχει πρόσβαση μόνο στα δεδομένα εισόδου του κάθε δείγματος. Ο αλγόριθμος σε αυτόν τον τύπο μηχανικής μάθησης έχει σκοπό να βρει τη δομή των δεδομένων εισόδου και μπορεί να δημιουργήσει τη δική του κλάση εξόδου. Η μη επιβλεπόμενη μάθηση χρησιμοποιείται για την αποκάλυψη κρυφών μοτίβων σε δεδομένα για την επίτευξη του σκοπού της ομαδοποίησης, δηλαδή για την ταξινόμηση δεδομένων. Σημαντικός τομέας των μη επιβλεπόμενων εφαρμογών μάθησης είναι η εκτίμηση της πυκνότητας στα στατιστικά προβλήματα.

5.1.3 Ενισχυτική μάθηση

Στην ενισχυτική μάθηση, το σύστημα πρέπει να επιτύχει ένα συγκεκριμένο αποτέλεσμα όπου μέσω της συνεχούς αλληλεπίδρασης με το περιβάλλον και την

παρουσία ενός «κριτή» που δίνει θετική ή αρνητική ανάδραση καθιστά δυνατή τη μάθηση της επιθυμητής συμπεριφοράς. Επομένως, σε κάθε περίπτωση, θα επιλέξει τη συμπεριφορά που μεγιστοποιεί τα κέρδη-ανταμοιβές με βάση αυτά που έχει μάθει. Η εκμάθηση ενίσχυσης μπορεί να χρησιμοποιηθεί για την ανάπτυξη στρατηγικών παιχνιδιών, όπως το σκάκι και η αλληλεπίδραση μεταξύ ανθρώπου και μηχανής.

5.1.4 Ημι-επιβλεπόμενη μάθηση

Ο τελευταίος τύπος μηχανικής μάθησης είναι η ημι-επιβλεπόμενη μάθηση, η οποία βρίσκεται κάπου μεταξύ της επιβλεπόμενης μάθησης και της μη επιβλεπόμενης μάθησης. Στην περίπτωση αυτή, ο αλγόριθμος διαθέτει ένα σύνολο δεδομένων, το οποίο περιέχει δείγματα με είσοδο-έξοδο και συνήθως αρκετά δείγματα τα οποία στερούνται την έξοδο. Η πρακτική αξία της ημι-επιβλεπόμενης μάθησης είναι ότι σε πρακτικές εφαρμογές, είναι συνήθως πιο εύκολο να βρεθούν δεδομένα χωρίς αποτελέσματα, έτσι τα ημι-επιβλεπόμενα μοντέλα μάθησης μπορούν να παράγουν αποτελέσματα γρηγορότερα και με χαμηλότερο κόστος.

5.2 Αλγόριθμοι

5.2.1 Δίκτυα Bayes (Naive Bayes)

Τα δίκτυα Bayes είναι ένας πιθανοτικός αλγόριθμος μηχανικής μάθησης που βασίζεται στο θεώρημα του Bayes, ο οποίος μπορεί να χρησιμοποιηθεί για διάφορες εργασίες ταξινόμησης. Το θεώρημα του Bayes είναι ένας απλός μαθηματικός τύπος που χρησιμοποιείται για τον υπολογισμό της πιθανότητας. Η υπό όρους πιθανότητα είναι ένα μέτρο της πιθανότητας εμφάνισης ενός συμβάντος, υποθέτοντας ότι έχει συμβεί κάποιο άλλο συμβάν. Είναι μια τεχνική ταξινόμησης που βασίζεται στο θεώρημα του Bayes, υποθέτοντας ανεξαρτησία μεταξύ προβλέψεων. Με απλά λόγια, ο ταξινομητής Bayes υποθέτει ότι η ύπαρξη ενός συγκεκριμένου χαρακτηριστικού σε μια τάξη δεν έχει καμία σχέση με την ύπαρξη οποιωνδήποτε άλλων χαρακτηριστικών.

Το θεώρημα του Bayes παρέχει έναν τρόπο υπολογισμού της αντίστροφης πιθανότητας $P(A|B)$ από $P(A)$, $P(B)$ και $P(B|A)$. Η εξίσωση είναι η εξής:

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$$

Το θεώρημα μας ενημερώνει για το πόσο πιθανό είναι να συμβεί το A εάν εμφανιστεί το B, γνωστό ως μεταγενέστερη πιθανότητα. Με απλά λόγια, το θεώρημα του Bayes είναι μια μέθοδος υπολογισμού μιας πιθανότητας, δεδομένου ενός συνόλου άλλων πιθανοτήτων.

Πλεονεκτήματα χρήσης δικτύων Bayes:

- Είναι ένας σχετικά απλός και γρήγορος αλγόριθμος για την πρόβλεψη της τάξης του συνόλου των δεδομένων δοκιμής και αξιόπιστος στην πρόβλεψη πολλαπλών ταξινομήσεων
- Όταν συναντάται η απαίτηση της ανεξαρτησίας, τα δίκτυα Bayes απαιτούν λιγότερα δεδομένα εκπαίδευσης και έτσι ξεπερνάνε σε απόδοση τα εναλλακτικά μοντέλα, όπως η λογιστική παλινδρόμηση
- Σε σύγκριση με τις αριθμητικές μεταβλητές εισόδου, αποδίδουν καλύτερα με κατηγορίες μεταβλητών εισόδου

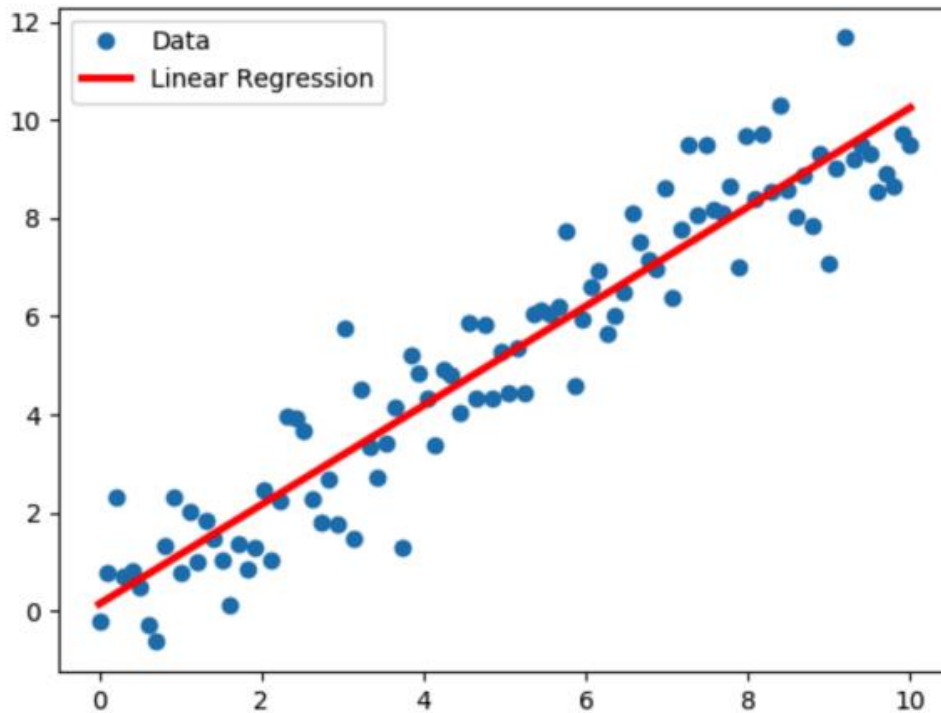
Μειονεκτήματα χρήσης δικτύων Bayes:

- Εάν μια μεταβλητή κατηγορίας στο σύνολο των δεδομένων δοκιμής έχει μια κατηγορία που δεν περιλαμβάνεται στο σύνολο των δεδομένων εκπαίδευσης, το μοντέλο αυτόματα θα εκχωρήσει πιθανότητα μηδέν και δε θα είναι σε θέση να δημιουργήσει μια πρόβλεψη. Αυτό αναφέρεται ως μηδενική συχνότητα
- Τα δίκτυα Bayes θεωρούνται κακός εκτιμητής, γι' αυτό και οι πιθανότητες που προκύπτουν από την πρόβλεψη, πρέπει να εξετάζονται με προσοχή
- Η υπόθεση των ανεξάρτητων προβλέψεων είναι ένα άλλο ελάττωμα, καθώς στην πραγματική ζωή, η απόκτηση μιας ομάδας προγνωστικών που είναι εντελώς ανεξάρτητες, είναι αδύνατη

5.2.2 Γενικευμένο γραμμικό μοντέλο (Generalized Linear Model)

Το 1972 οι John Nelder και Robert Wedderburn ανέπτυξαν το γενικευμένο γραμμικό μοντέλο, το οποίο είναι μια προηγμένη προσέγγιση στατιστικής μοντελοποίησης. Είναι μια ποικιλία μοντέλων που επιτρέπουν στη μεταβλητή απόκρισης να έχει κατανομή σφαλμάτων διαφορετική από την κανονική κατανομή. Η μεταβλητή απόκρισης «y» δηλώνεται ως γραμμική συνάρτηση ή γραμμικός συνδυασμός όλων των προβλέψεων «x» σε ένα μοντέλο γραμμικής παλινδρόμησης. Η απόκριση και οι προβλέψεις έχουν μια γραμμική σύνδεση στον πυρήνα τους. Η κατανομή σφαλμάτων της μεταβλητής απόκρισης πρέπει επίσης να κατανέμεται κανονικά, με αποτέλεσμα τη δημιουργία ενός γραμμικού μοντέλου.

Ακόμα και αν ο υποκείμενος σύνδεσμος μεταξύ της απόκρισης και των προβλέψεων δεν είναι γραμμικός, τα γενικευμένα γραμμικά μοντέλα επιτρέπουν τη δημιουργία μιας γραμμικής σχέσης μεταξύ τους. Αυτό γίνεται με τη χρήση μιας λειτουργίας σύνδεσης, η οποία συνδέει τη μεταβλητή απόκριση με ένα γραμμικό μοντέλο. Σε αντίθεση με τα μοντέλα γραμμικής παλινδρόμησης, η κατανομή σφαλμάτων της μεταβλητής απόκρισης δε χρειάζεται να διανέμεται τακτικά. Τα σφάλματα μεταβλητής απόκρισης αναμένεται να ακολουθήσουν μια εκθετική οικογένεια διανομής.



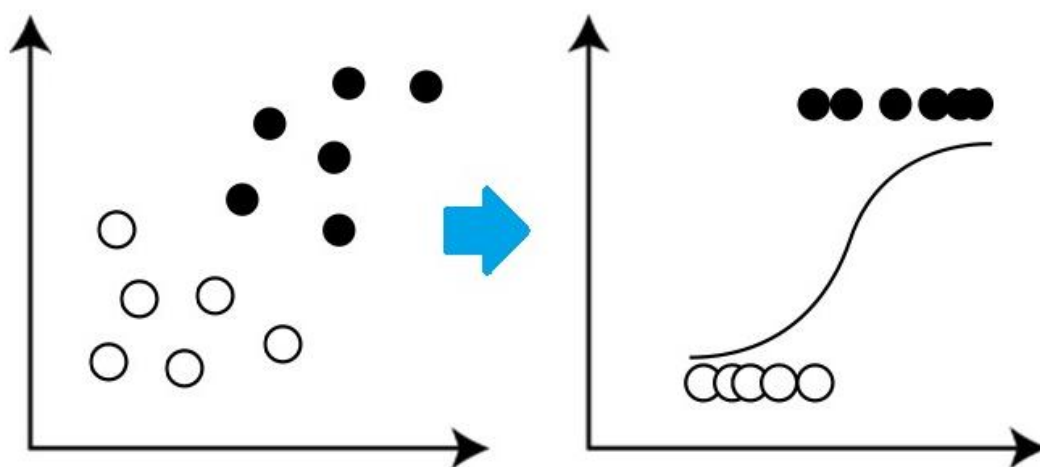
Εικόνα 8 Δείγμα γραμμικής παλινδρόμησης (Tam, 2020)

Τα δεδομένα πρέπει να είναι τυχαία και ανεξάρτητα και η μεταβλητή απόκρισης y δε χρειάζεται να διανέμεται τακτικά, αλλά πρέπει να ανήκει στην εκθετική οικογένεια. Η αρχική μεταβλητή απόκρισης δεν απαιτείται να έχει γραμμική σύνδεση με τις ανεξάρτητες μεταβλητές, αλλά η μετασχηματισμένη μεταβλητή απόκρισης. Αντί της χρήσης των αρχικών ακατέργαστων ανεξάρτητων μεταβλητών, μπορεί να χρησιμοποιηθεί μεταβλητός μετασχηματισμός και οι μετασχηματισμένες ανεξάρτητες μεταβλητές μπορούν στη συνέχεια να χρησιμοποιηθούν για την κατασκευή του γενικευμένου γραμμικού μοντέλου.

Δεν είναι απαραίτητο να ικανοποιηθεί η ομογενοποίηση, δηλαδή σταθερή διακύμανση. Μεταβλητή απόκριση με τις ανεξάρτητες μεταβλητές, η διακύμανση σφάλματος μπορεί να αυξηθεί ή να μειωθεί. Τα σφάλματα είναι ανεξάρτητα, αλλά δεν πρέπει να κατανέμονται ομοιόμορφα.

5.2.3 Λογιστική παλινδρόμηση (Logistic Regression)

Σε αντίθεση με το όνομα, η λογιστική παλινδρόμηση είναι στην πραγματικότητα ένα μοντέλο ταξινόμησης, παρά ένα μοντέλο παλινδρόμησης. Είναι μια πολύ αποτελεσματική τεχνική όσον αφορά τα δυαδικά και γραμμικά ζητήματα ταξινόμησης. Ένα μοντέλο ταξινόμησης που είναι απλό στην εφαρμογή και παρέχει εξαιρετικά αποτελέσματα με γραμμικά διαχωρίσιμες κατηγορίες. Στο βιομηχανικό κόσμο, είναι μια ευρέως χρησιμοποιούμενη μέθοδος κατηγοριοποίησης. Όπως το Adaline και το Perceptron, το μοντέλο λογιστικής παλινδρόμησης είναι μια στατιστική προσέγγιση για τη δυαδική ταξινόμηση που μπορεί να επεκταθεί σε ταξινόμηση πολλαπλών κλάσεων.



Εικόνα 9 Προσέγγιση λογιστικής παλινδρόμησης (Equiskill, 2018)

Οι συντελεστές του αλγόριθμου λογιστικής παλινδρόμησης πρέπει να υπολογίζονται χρησιμοποιώντας τα δεδομένα εκπαίδευσης. Χρησιμοποιείται η εκτίμηση μέγιστης πιθανότητας για αυτό. Παρόλο που γίνονται υποθέσεις σχετικά με τη διανομή των δεδομένων, η εκτίμηση τη μέγιστης πιθανότητας είναι μια τυπική τεχνική μάθησης που χρησιμοποιείται από έναν αριθμό αλγορίθμων μηχανικής μάθησης. Οι καλύτεροι συντελεστές θα οδηγούσαν σε ένα μοντέλο που προέβλεπε μια τιμή για την προεπιλεγμένη κλάση που ήταν πολύ κοντά στο 1 και μια τιμή για την άλλη τάξη που ήταν πολύ κοντά στο 0. Η ιδέα πίσω από τη λογιστική παλινδρόμηση μέγιστης πιθανότητας είναι ότι φαίνεται μια μέθοδος αναζήτησης για τιμές συντελεστών που ελαχιστοποιούν τη διαφορά μεταξύ των προβλεπόμενων πιθανοτήτων του μοντέλου και εκείνων στα δεδομένα.

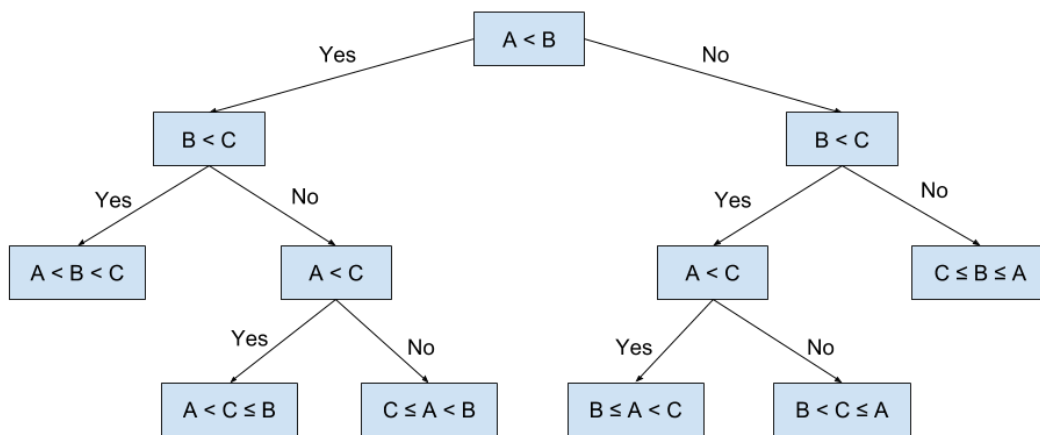
Η ονομασία λογιστική παλινδρόμηση προέκυψε από το επίκεντρο της προσέγγισης που είναι η λογιστική συνάρτηση. Η υλικότεχνική συνάρτηση, γνωστή και ως σιγμοειδής συνάρτηση, δημιουργήθηκε από στατιστικούς για να χαρακτηρίσει τα χαρακτηριστικά της αύξησης του πληθυσμού στην οικολογία, όπως το πως αυξάνεται γρήγορα και τελικά φτάνει στη φέρουσα ικανότητα του περιβάλλοντος. Είναι μια καμπύλη σχήματος S που μπορεί να μεταφράσει οποιονδήποτε ακέραιο πραγματικής αξίας σε μια τιμή μεταξύ 0 και 1, αλλά ποτέ ακριβώς μεταξύ αυτών των δύο σημείων.

5.2.4 Δέντρο απόφασης (Decision Tree)

Ο αλγόριθμος του δέντρου απόφασης είναι μέρος της οικογένειας των αλγορίθμων επιβλεπόμενης μάθησης. Η προσέγγιση του δέντρου αποφάσεων, σε αντίθεση με άλλους επιβλεπόμενους αλγόριθμους μάθησης, μπορεί να χρησιμοποιηθεί για την επίλυση προβλημάτων παλινδρόμησης και ταξινόμησης. Με την εκμάθηση βασικών κανόνων απόφασης που συνάγονται από προηγούμενα δεδομένα, ο στόχος της χρήσης ενός δέντρου αποφάσεων είναι η κατασκευή ενός μοντέλου εκπαίδευσης που μπορεί να χρησιμοποιηθεί για την πρόβλεψη της τάξης ή της τιμής της μεταβλητής στόχου.

Για την πρόβλεψη μιας ετικέτας τάξης σε μια εγγραφή όταν χρησιμοποιείται το δέντρο απόφασης, η αρχή γίνεται από τη «ρίζα» του δέντρου. Συγκρίνονται οι τιμές του χαρακτηριστικού root και του χαρακτηριστικού της εγγραφής. Ακολουθείται ο κλάδος που αντιστοιχεί σε αυτήν την τιμή και προχωράει στον επόμενο κόμβο με βάση τη σύγκριση.

Ανάλογα με τον τύπο της μεταβλητής, υπάρχουν δύο τύποι δέντρων αποφάσεων. Το δέντρο απόφασης κατηγοριοποίησης μεταβλητής, το οποίο είναι ένα δέντρο αποφάσεων μεταβλητής κατηγορίας με μια κατηγορηματική μεταβλητή στόχου και το δέντρο απόφασης συνεχούς μεταβλητής, όπου ένα δέντρο αποφάσεων συνεχούς μεταβλητής είναι αυτό που έχει μια μεταβλητή συνεχούς στόχου.



Εικόνα 10 Παράδειγμα δέντρου απόφασης (Niculaescu, 2018)

Η δυσκολία με τα δέντρα απόφασης είναι ότι επεκτείνονται πολύ, ειδικά αν υπάρχει μια βάση δεδομένων με πολλές στήλες. Για να αντιμετωπιστεί αυτή η υπερφόρτωση του δέντρου, ακολουθείται μια διαδικασία γνωστή και ως κλάδεμα του δέντρου αποφάσεων. Το κλάδεμα του δέντρου αποφάσεων είναι η διαδικασία αφαίρεσης των κόμβων απόφασης, ξεκινώντας από τον κόμβο των φύλλων, προκειμένου να διατηρηθεί η συνολική ακρίβεια. Αυτό επιτυγχάνεται διαιρώντας το πραγματικό σετ εκπαίδευσης σε δύο μέρη, για προπόνηση και για επικύρωση.

5.2.5 Τυχαίο δάσος (Random Forest)

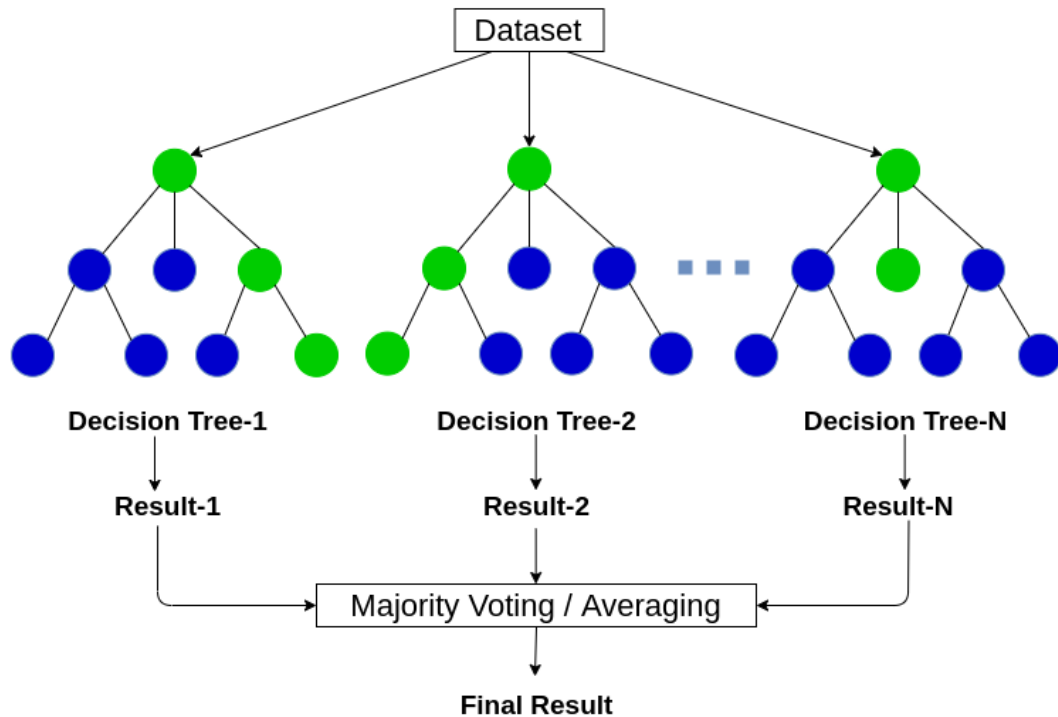
Το τυχαίο δάσος είναι μια επιβλεπόμενη τεχνική μηχανικής μάθησης που χρησιμοποιεί αλγόριθμους δέντρων αποφάσεων για την κατασκευή του. Είναι μια προσέγγιση μηχανικής μάθησης τόσο για την επίλυση θεμάτων ταξινόμησης, όσο και θεμάτων παλινδρόμησης. Χρησιμοποιεί τη μάθηση του συνόλου, η οποία είναι μια τεχνική για την επίλυση δύσκολων προβλημάτων συνδυάζοντας πολλούς ταξινομητές.

Πολλά δέντρα αποφάσεων αποτελούν έναν τυχαίο αλγόριθμο δασών. Ο αλγόριθμος αυτός καθορίζει το αποτέλεσμα βάσει των προβλέψεων του δέντρου αποφάσεων. Προβλέπει με μέσο όρο ή κατά μέσο όρο την παραγωγή διαφόρων δέντρων. Η ακρίβεια του αποτελέσματος βελτιώνεται καθώς αυξάνεται ο αριθμός των δέντρων.

Ο αλγόριθμος του τυχαίου δάσους υπερνικά σε ακρίβεια τον αλγόριθμο του δέντρου αποφάσεων και είναι ένα χρήσιμο εργαλείο για την αντιμετώπιση των δεδομένων που λείπουν. Επίσης, ξεπερνά το πρόβλημα της υπερβολικής προσαρμογής του δέντρου αποφάσεων και στο σημείο διαχωρισμού του κόμβου σε κάθε τυχαίο δασικό δέντρο, επιλέγεται τυχαία ένα υποσύνολο χαρακτηριστικών.

Η πρωταρχική διάκριση μεταξύ του δέντρου αποφάσεων και των αλγορίθμων τυχαίων δασών είναι ότι ο αλγόριθμος τυχαίων δασών δημιουργεί τυχαία ριζικούς κόμβους και διαχωρίζει τους κόμβους. Η μέθοδος της αποθήκευσης χρησιμοποιείται από το τυχαίο δάσος για τη δημιουργία της απαιτούμενης πρόβλεψης. Αντί για τη χρήση ενός μόνο δείγματος δεδομένων, η αποθήκευση απαιτεί τη χρήση πολλών δειγμάτων. Ένα σύνολο δεδομένων εκπαίδευσης είναι μια συλλογή από παρατηρήσεις και χαρακτηριστικά που χρησιμοποιούνται για την πραγματοποίηση προβλέψεων.

Ανάλογα με τα δεδομένα εκπαίδευσης που δίνονται στον τυχαίο αλγόριθμο δάσους, τα δέντρα αποφάσεων αποφέρουν ποικίλα αποτελέσματα. Οι έξοδοι θα βαθμολογηθούν και αυτή με την καλύτερη βαθμολογία θα επιλεγεί ως το τελικό αποτέλεσμα. Η ταξινόμηση τυχαίων δασών χρησιμοποιεί μια τεχνική συνόλου για να επιτευχθεί το επιθυμητό αποτέλεσμα. Διάφορα δέντρα αποφάσεων εκπαιδεύονται χρησιμοποιώντας τα δεδομένα εκπαίδευσης. Αυτό το σύνολο δεδομένων περιέχει παρατηρήσεις και χαρακτηριστικά που επιλέγονται τυχαία όταν διαιρούνται οι κόμβοι.



Εικόνα 11 Ταξινόμηση τυχαίου δάσους (Ampradu, 2021)

Στα πλεονεκτήματα από τη χρήση του αλγορίθμου τυχαίου δάσους είναι η ικανότητά του να χρησιμοποιηθεί σε θέματα και παλινδρόμησης και ταξινόμησης με προβλέψεις ακρίβειας. Είναι σε θέση να χειριστεί αποτελεσματικά μεγάλα σύνολα δεδομένων και σε σύγκριση με τη μέθοδο του δέντρου αποφάσεων, ο αλγόριθμος τυχαίων δασών είναι πιο ακριβής στην πρόβλεψη αποτελεσμάτων. Από την άλλη πλευρά, ως μειονεκτήματα μπορούν να του καταλογιστούν η απαίτηση για ενισχυμένους υπολογιστικούς πόρους και σε σύγκριση με τον αλγόριθμο δέντρων αποφάσεων, απαιτείται περισσότερος χρόνος.

5.2.6 Ενίσχυση της κλίσης (Gradient Boosting)

Μία από τις πιο ισχυρές προσεγγίσεις για τη δημιουργία προγνωστικών μοντέλων είναι η ενίσχυση της κλίσης. Η έννοια της ενίσχυσης προέκυψε από το ερώτημα εάν μία «φτωχή μάθηση» μπορεί να βελτιωθεί. Είναι μια τεχνική που μπορεί να χρησιμοποιηθεί τόσο για την παλινδρόμηση όσο και για την ταξινόμηση. Ο στόχος κάθε μεθόδου επιβλεπόμενης μάθησης είναι να ορίσει και να ελαχιστοποιήσει μια

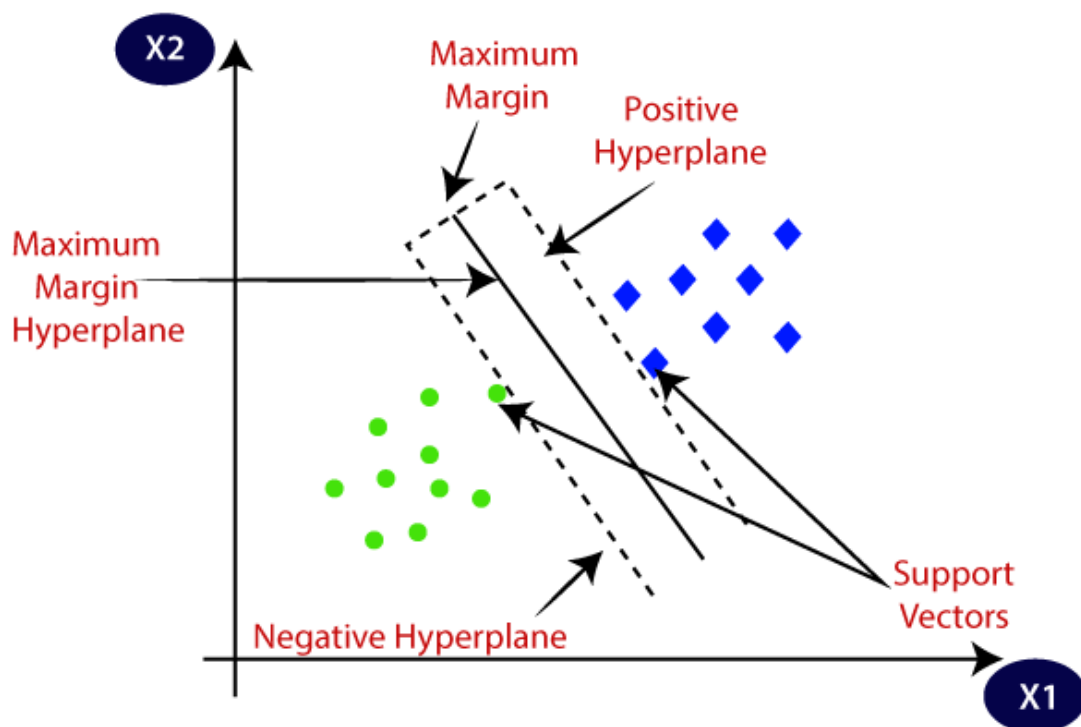
συνάρτηση απώλειας η οποία καθορίζεται στο ζήτημα που αντιμετωπίζεται. Για παράδειγμα στην παλινδρόμηση μπορεί να χρησιμοποιηθεί ένα τετραγωνικό σφάλμα, ενώ στην ταξινόμηση μπορεί να χρησιμοποιηθεί μια λογαριθμική απώλεια. Η ενίσχυση της κλίσης έχει το πλεονέκτημα ότι δεν απαιτεί την ανάπτυξη ενός νέου αλγορίθμου ενίσχυσης για κάθε συνάρτηση απώλειας που μπορεί να χρησιμοποιηθεί. Αντ' αυτού είναι ένα αρκετά γενικό πλαίσιο που μπορεί να χρησιμοποιήσει οποιαδήποτε διαφοροποιήσιμη συνάρτηση απώλειας.

Στην ενίσχυση της κλίσης, τα δέντρα αποφάσεων χρησιμοποιούνται ως ο «φτωχός μαθητής». Τα δέντρα παλινδρόμησης χρησιμοποιούνται ειδικά επειδή παρέχουν πραγματικές τιμές για διαχωρισμούς και μπορούν να συνδυαστούν μεταξύ τους, επιτρέποντας στο μελλοντικό μοντέλο να προστεθεί και να διορθώσει τα υπολείμματα στις προβλέψεις.

Η ιδέα πίσω από την τεχνική αύξησης της κλίσης είναι η χρήση μοτίβων σε υπολείμματα για την επανειλημμένη ενίσχυση και βελτίωση ενός μοντέλου με κακές προβλέψεις. Η μοντελοποίηση των υπολειμμάτων σταματάει μόλις φτάσει στο σημείο που δεν υπάρχει κανένα πρότυπο που μπορεί να μοντελοποιηθεί, διαφορετικά αυτό θα οδηγούσε σε υπερβολική προσαρμογή. Μειώνεται αλγοριθμικά η λειτουργία της απώλειας, έτσι ώστε η απώλεια της δοκιμής να φτάσει στο ελάχιστο.

5.2.7 Μηχανή διανυσμάτων υποστήριξης (Support Vector Machine)

Η μηχανή διανυσμάτων υποστήριξης είναι ίσως η πιο διαδεδομένη μέθοδος επιβλεπόμενης μάθησης που μπορεί να χρησιμοποιηθεί για την επίλυση ζητημάτων ταξινόμησης και παλινδρόμησης. Ωστόσο, χρησιμοποιείται συνήθως κυρίως στη μηχανική εκμάθηση για θέματα ταξινόμησης. Ο στόχος αυτού του αλγόριθμου είναι να βρει τη βέλτιστη γραμμή, το λεγόμενο όριο απόφασης, για την κατηγοριοποίηση του χώρου n -διαστάσεων σε κλάσεις, έτσι ώστε επιπλέον σημεία δεδομένων να μπορούν να τοποθετηθούν εύκολα στην κατάλληλη κατηγορία στο μέλλον. Το βέλτιστο όριο επιλογής ονομάζεται υπερπλάνο και τα ακραία σημεία/διανύσματα που βοηθούν στη δημιουργία του, επιλέγονται μέσω της μηχανής διανυσμάτων υποστήριξης.



Εικόνα 12 Ταξινόμηση με όριο απόφασης (Javatpoint, 2021)

Υπάρχουν δύο τύποι μηχανών διανυσμάτων υποστήριξης, ο γραμμικός και ο μη γραμμικός. Ο γραμμικός τύπος είναι ένας ταξινομητής που χρησιμοποιείται για γραμμικά διαχωρίσιμα δεδομένα, πράγμα που σημαίνει ότι εάν ένα σύνολο δεδομένων μπορεί να κατηγοριοποιηθεί σε δύο κλάσεις χρησιμοποιώντας μία μόνο ευθεία γραμμή, ονομάζεται γραμμικά διαχωρίσιμα δεδομένα και ο ταξινομητής ονομάζεται γραμμικός. Ο μη γραμμικός τύπος χρησιμοποιείται για μη γραμμικά διαχωρίσιμα δεδομένα, κάτι που σημαίνει ότι εάν ένα σύνολο δεδομένων δε μπορεί να κατηγοριοποιηθεί χρησιμοποιώντας μια ευθεία γραμμή, είναι μη γραμμικά δεδομένα και αντίστοιχα ο ταξινομητής που χρησιμοποιείται ονομάζεται μη γραμμικός.

Τα πλεονεκτήματα χρήσης των μηχανών διανυσμάτων υποστήριξης είναι ότι λειτουργούν πολύ καλά σε χώρους υψηλής διάστασης και όταν ο αριθμός των διαστάσεων υπερβαίνει τον αριθμό των δειγμάτων, η μέθοδος εξακολουθεί να είναι επιτυχής. Επίσης, η μηχανή διανυσμάτων υποστήριξης είναι αποδοτική στο θέμα της μνήμης της υπολογιστικής μονάδας, καθώς χρησιμοποιεί μόνο ένα υποσύνολο σημείων εκπαίδευσης στη συνάρτηση απόφασης και ευέλικτη γιατί μπορεί να παρέχονται διαφορετικές λειτουργίες για τη συνάρτηση απόφασης.

Στα μειονεκτήματα από τη χρήση μηχανής διανυσμάτων υποστήριξης είναι ότι εάν ο αριθμός των χαρακτηριστικών υπερβαίνει τον αριθμό των δειγμάτων, είναι κρίσιμο να αποφευχθεί η υπερβολική προσαρμογή κατά την επιλογή των λειτουργιών και των όρων κανονικοποίησης. Τέλος, οι μηχανές διανυσμάτων υποστήριξης δεν προσφέρουν απευθείας εκτιμήσεις πιθανότητας, αφού αυτές προκύπτουν από μία χρονοβόρα πενταπλή διαδικασία διασταυρούμενης επικύρωσης.

Κεφάλαιο 6: Πειραματικό μέρος

Για την πραγματοποίηση του πειραματικού μέρους της διπλωματικής εργασίας χρησιμοποιήθηκε το πρόγραμμα RapidMiner Studio. Είναι ένα λογισμικό ανοιχτού κώδικα που χρησιμοποιείται για την εξόρυξη δεδομένων, τη στατιστική αλλά και την προβλεπτική ανάλυσή τους. Το σύνολο των δεδομένων για την παρούσα διπλωματική εργασία χορηγήθηκε από το πανεπιστήμιο του Βερολίνου έπειτα από σχετική αίτηση. Τα δεδομένα που παρέχονται αφορούν μια μηχανή παραγωγής σε βιομηχανικό περιβάλλον και η χρήση τους γίνεται μόνο για ακαδημαϊκούς σκοπούς. Αποτελείται από 10.000 διαφορετικά σημεία δεδομένων τα οποία έχουν 14 χαρακτηριστικά το καθένα. Αναλυτικότερα:

UID: μοναδικό αναγνωριστικό σημείου δεδομένου που κυμαίνεται από 1 έως 10.000

Product ID: αναγνωριστικό προϊόντος

Type: ποιότητα προϊόντος που διαχωρίζεται σε L (Low – χαμηλή), M (Medium – μεσαία) και H (High – Υψηλή)

Air temperature: θερμοκρασία αέρα μετρημένη σε Kelvin με τη μέθοδο της τυχαίας βάδισης (random walk)

Process temperature: θερμοκρασία διεργασίας μετρημένη σε Kelvin και κανονικοποιημένη σε τυπική απόκλιση 1 Kelvin

Rotational speed: ταχύτητα περιστροφής (στροφές ανά λεπτό) υπολογισμένη σε ισχύ 2.860 Watt με κανονικά κατανεμημένο θόρυβο

Torque: τιμή ροπής (Nm) κατά τη διαδικασία παραγωγής

Tool wear: φθορά εξοπλισμού σε σχέση με το χρόνο. Οι ποιοτικές παραλλαγές H/M/L προσθέτουν αντίστοιχα 5/3/2 λεπτά φθοράς στον εξοπλισμό

Οι βλάβες της μηχανής παραγωγής (Machine failure) αποτελούνται από πέντε ανεξάρτητες λειτουργίες αστοχίας:

TWF: Αποτυχία λόγω φθοράς εξοπλισμού (Tool Wear Failure). Ο εξοπλισμός ή μέρος του εξοπλισμού θα αντικατασταθεί λόγω βλάβης σε τυχαία επιλεγμένο χρόνο φθοράς μεταξύ 200 και 240 λεπτών (120 φορές στο σύνολο δεδομένων).

HDF: Αποτυχία διάχυσης θερμότητας (Heat Dissipation Failure). Προκαλείται αποτυχία διαδικασίας εάν η διαφορά μεταξύ θερμοκρασίας αέρα και θερμοκρασία διεργασίας είναι κάτω από 8,6 Kelvin και η ταχύτητα περιστροφής είναι κάτω από 1.380 στροφές ανά λεπτό. Αυτό ισχύει σε 115 σημεία δεδομένων.

PWF: Αποτυχία ισχύς (Power Failure). Το γινόμενο της ροπής και της ταχύτητας περιστροφής (σε rad/s) ισούται με την απαιτούμενη ισχύ για τη διαδικασία. Εάν αυτή η ισχύς είναι κάτω από 3.500 Watt ή πάνω από 9.000 Watt, η διαδικασία αποτυγχάνει. Αυτό συμβαίνει 95 φορές στο σύνολο των δεδομένων.

OSF: Αστοχία λόγω υπερφόρτωσης (Overstrain Failure). Αν το προϊόν παραγωγής με φθορά εξοπλισμού επί την τιμή της ροπής υπερβαίνει τα 11.000 min*Nm για την παραλλαγή προϊόντος Low (αντίστοιχα για Medium 12.000 και High 13.000) η διαδικασία αποτυγχάνει. Αυτό ισχύει για 98 σημεία δεδομένων.

RNF: Τυχαίες αποτυχίες (Random Failures). Κάθε διαδικασία έχει πιθανότητα 0,1% να αποτύχει ανεξάρτητα από τις παραμέτρους της διαδικασίας αυτής. Στο συγκεκριμένο σύνολο δεδομένων αυτό συμβαίνει μόνο για 5 σημεία δεδομένων, δηλαδή λιγότερο από το αναμενόμενο.

Αν ισχύει τουλάχιστον μία από τις παραπάνω λειτουργίες αστοχίας, τότε η διαδικασία αποτυγχάνει και η ετικέτα “Machine Failure” ορίζεται σε 1. Επομένως, δεν είναι ξεκάθαρο ποια από τις λειτουργίες αστοχίας έχει προκαλέσει την αποτυχία της διαδικασίας.





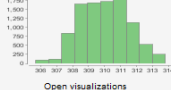
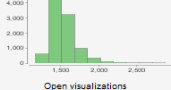
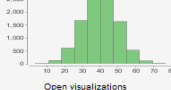
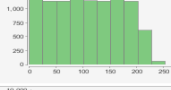
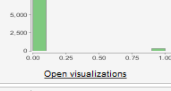
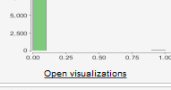
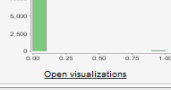
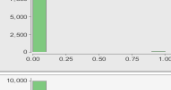
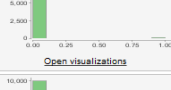
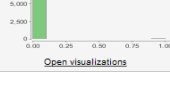
UID Number	Product ID Category	Type Category	Air temperat... Number	Process temp... Number	Rotational sp... Number	Torque [Nm] Number	Tool wear [m... Number
157	M15016	M	298.400	308.100	1495	46	203
158	H29571	H	298.400	308.200	1478	43.500	206
159	M15018	M	298.400	308.200	1499	40	211
160	L47339	L	298.400	308.200	1471	47	214
161	L47340	L	298.400	308.200	1282	60.700	216
162	L47341	L	298.300	308.100	1412	52.300	218
163	L47342	L	298.300	308.100	1586	35.500	0
164	M15023	M	298.400	308.200	1488	43.800	2

UID Number	Machine fail... Number	TWF Number	HDF Number	PWF Number	OSF Number	RNF Number
157	0	0	0	0	0	0
158	0	0	0	0	0	0
159	0	0	0	0	0	0
160	0	0	0	0	0	0
161	1	0	0	0	1	0
162	1	0	0	0	1	0
163	0	0	0	0	0	0
164	0	0	0	0	0	0

Εικόνα 13 Τυχαίο δείγμα των δεδομένων όπως εμφανίζονται στο RapidMiner

Επομένως, περάστηκαν όλα τα δεδομένα που παρείχε το αρχείο τύπου CSV (Comma Separated-Values) και πριν ξεκινήσει η διαδικασία της ανάλυσης, έγινε η επισκόπηση αυτών των αρχικών δεδομένων με σκοπό τη διόρθωση κάποιων τυπογραφικών σφαλμάτων, αλλά και της λανθασμένης ανάγνωσης κάποιων εισόδων από το πρόγραμμα.

Μετά την καταχώρηση των δεδομένων στο λογισμικό RapidMiner Studio, ακολουθεί η στατιστική ανάλυσή τους, όπως φαίνεται στην παρακάτω εικόνα:

Name	Type	Missing	Statistics	Filter (14 / 14 attributes)	<input type="text" value="Search for Attributes"/>
UID	Integer	0	 <p>Min: 1, Max: 10000, Average: 5000.500, Deviation: 2886.896</p>		
Product ID	Nominal	0	 <p>Least: M24859 (1), Most: H29424 (1)</p> <p>Values: H29424 (1), H29425 (1), H29432 (1), H29434 (1), ...[9996 more]</p>		
Type	Nominal	0	 <p>Least: H (1003), Most: L (6000)</p> <p>Values: L (6000), M (2997), H (1003)</p>		
Air temperature [K]	Real	0	 <p>Min: 295.300, Max: 304.500, Average: 300.005, Deviation: 2.000</p>		
Process temperature [K]	Real	0	 <p>Min: 305.700, Max: 313.800, Average: 310.006, Deviation: 1.484</p>		
Rotational speed [rpm]	Integer	0	 <p>Min: 1168, Max: 2886, Average: 1538.776, Deviation: 179.284</p>		
Torque [Nm]	Real	0	 <p>Min: 3.800, Max: 76.600, Average: 39.987, Deviation: 9.969</p>		
Tool wear [min]	Integer	0	 <p>Min: 0, Max: 253, Average: 107.951, Deviation: 63.654</p>		
Machine failure	Integer	0	 <p>Min: 0, Max: 1, Average: 0.034, Deviation: 0.181</p>		
TWF	Integer	0	 <p>Min: 0, Max: 1, Average: 0.005, Deviation: 0.068</p>		
HDF	Integer	0	 <p>Min: 0, Max: 1, Average: 0.011, Deviation: 0.107</p>		
PWF	Integer	0	 <p>Min: 0, Max: 1, Average: 0.009, Deviation: 0.097</p>		
OSF	Integer	0	 <p>Min: 0, Max: 1, Average: 0.010, Deviation: 0.099</p>		
RNF	Integer	0	 <p>Min: 0, Max: 1, Average: 0.002, Deviation: 0.044</p>		

Εικόνα 14 Στατιστική ανάλυση των δεδομένων

Έπειτα της στατιστικής ανάλυσης των δεδομένων που εισήχθησαν στο πρόγραμμα, ορίστηκε η επιλογή της πρόβλεψης για Machine failure = 1 και επιλέχθηκαν οι αλγόριθμοι Naive Bayes, Generalized Linear Model, Logistic Regression, Decision Tree, Random Forest, Gradient Boosting και Support Vector Machine, οι οποίοι θα χρησιμοποιηθούν για να γίνει η συγκεκριμένη πρόβλεψη.

Predict

Want to predict the values of a column?

Clusters

Want to identify groups in your data?

Outliers

Want to detect outliers in your data?

UID Number	Product ID Category	Type Category	Air temperat... Number	Process temp... Number	Rotational sp... Number	Torque [Nm] Number	Tool wear [m... Number	Machine fail... Number	TWF Number	HDF Number	PWF Number
157	M15016	M	298.400	308.100	1495	46	203	0	0	0	0
158	H29571	H	298.400	308.200	1478	43.500	206	0	0	0	0
159	M15018	M	298.400	308.200	1499	40	211	0	0	0	0
160	L47339	L	298.400	308.200	1471	47	214	0	0	0	0
161	L47340	L	298.400	308.200	1282	60.700	216	1	0	0	0
162	L47341	L	298.300	308.100	1412	52.300	218	1	0	0	0
163	L47342	L	298.300	308.100	1586	35.500	0	0	0	0	0

10,000 rows - 14 columns (2 nominal, 12 numerical)

Εικόνα 15 Επιλογή για πρόβλεψη της συνθήκης Machine failure=1

Για να γίνει η σύγκριση των αποτελεσμάτων ανάλογα με τον κάθε αλγόριθμο χρησιμοποιήθηκαν 2 βασικοί δείκτες απόδοσης (KPI – Key Performance Indicator), της ακρίβειας (Accuracy) και του σφάλματος ταξινόμησης (Classification Error).

Ο όρος της ακρίβειας αναφέρεται στην ακρίβεια της ταξινόμησης και πιο συγκεκριμένα στην ακρίβεια ενός μοντέλου μηχανικής μάθησης. Είναι μια μέτρηση που χρησιμοποιείται για τον προσδιορισμό του καλύτερου μοντέλου στην εύρεση συσχετίσεων και προτύπων μεταξύ μεταβλητών σε ένα σύνολο δεδομένων. Η ακρίβεια ενός μοντέλου μπορεί να υπολογιστεί από το λόγο του αριθμού των σωστών προβλέψεων με το συνολικό αριθμό δειγμάτων εισόδου.

$$Accuracy = \frac{\text{Number of Features classified correctly}}{\text{Number of Features}}$$

Ο άλλος δείκτης απόδοσης που χρησιμοποιήθηκε για τη σύγκριση των αποτελεσμάτων είναι αυτός του σφάλματος ταξινόμησης (Classification Error ή Misclassification Rate). Το σφάλμα ταξινόμησης εξαρτάται από τον αριθμό των δεδομένων που δεν έχουν ταξινομηθεί σωστά (False Positives + False Negatives) και αξιολογείται με τον τύπο:

$$\text{Misclassification Rate} = \frac{\text{Number of Features classified incorrectly}}{\text{Number of Features}}$$

Τα αποτελέσματα που προέκυψαν με βάση τους δύο συγκεκριμένους δείκτες απόδοσης για κάθε αλγόριθμο φαίνονται στους παρακάτω πίνακες:

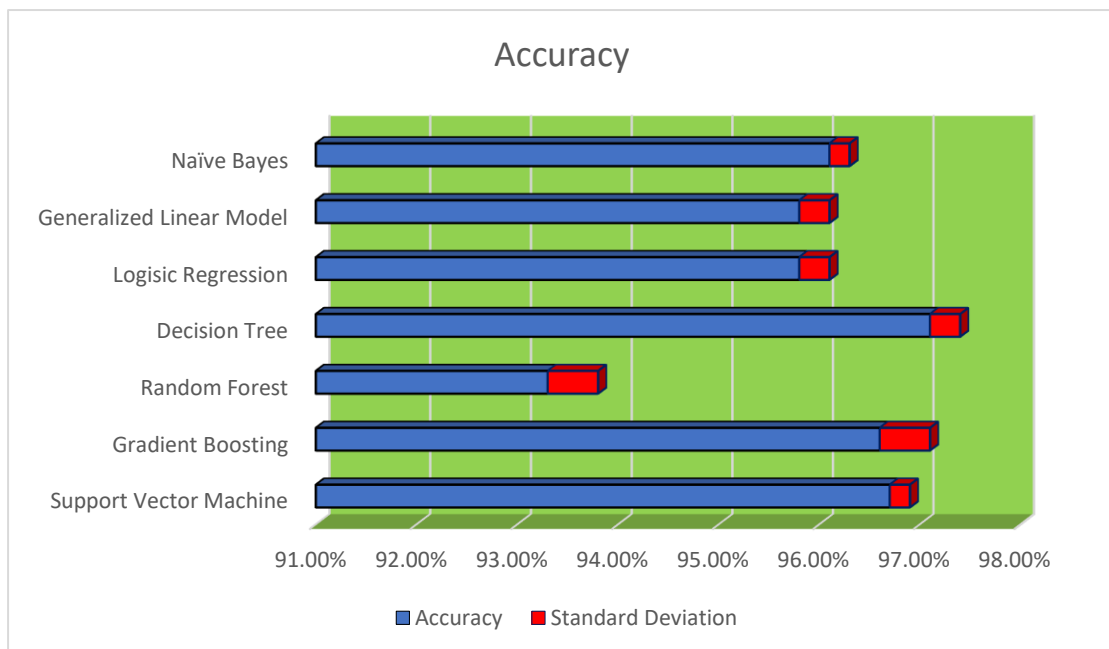
Αλγόριθμοι	Accuracy		
	Ακρίβεια	Τυπική Απόκλιση	Συνολικός Χρόνος
Naïve Bayes	96,1%	± 0,2%	1min 36sec
Generalized Linear Model	95,8%	± 0,3%	1min 54sec
Logistic Regression	95,8%	± 0,3%	1min 47sec
Decision Tree	97,1%	± 0,3%	1min 25sec
Random Forest	93,3%	± 0,5%	20min 7sec
Gradient Boosting	96,6%	± 0,5%	6min 35sec
Support Vector Machine	96,7%	± 0,2%	68min 23sec

Εικόνα 16 Αποτελέσματα δείκτη απόδοσης ακρίβειας

Αλγόριθμοι	Classification Error		
	Σφάλμα Ταξινόμησης	Τυπική Απόκλιση	Συνολικός Χρόνος
Naïve Bayes	3,9%	± 0,2%	1min 36sec
Generalized Linear Model	4,2%	± 0,3%	1min 54sec
Logistic Regression	4,2%	± 0,3%	1min 47sec
Decision Tree	2,9%	± 0,3%	1min 25sec
Random Forest	6,7%	± 0,5%	20min 7sec
Gradient Boosting	3,4%	± 0,5%	6min 35sec
Support Vector Machine	3,3%	± 0,2%	68min 23sec

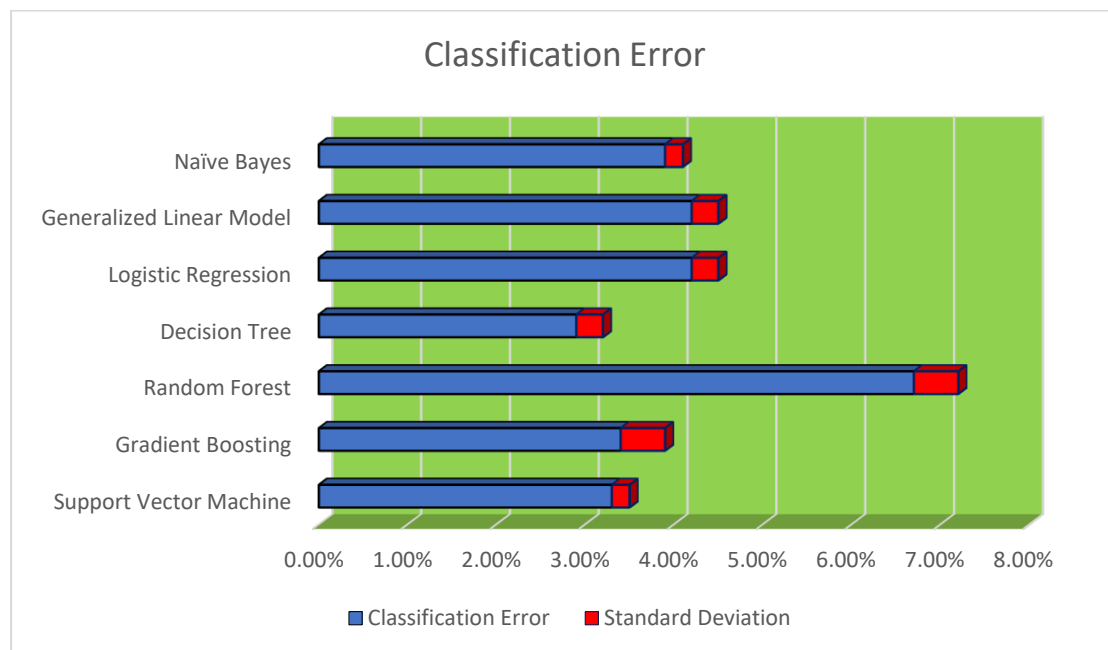
Εικόνα 17 Αποτελέσματα δείκτη απόδοσης σφάλματος ταξινόμησης

Για την καλύτερη σύγκριση και κατανόηση των αποτελεσμάτων παρατίθενται και τα αντίστοιχα γραφήματα σε σχέση με την ακρίβεια και το σφάλμα ταξινόμησης, όπως επίσης και του συνολικού χρόνου που χρειάστηκε για να ολοκληρωθεί η διεργασία για κάθε αλγόριθμο.



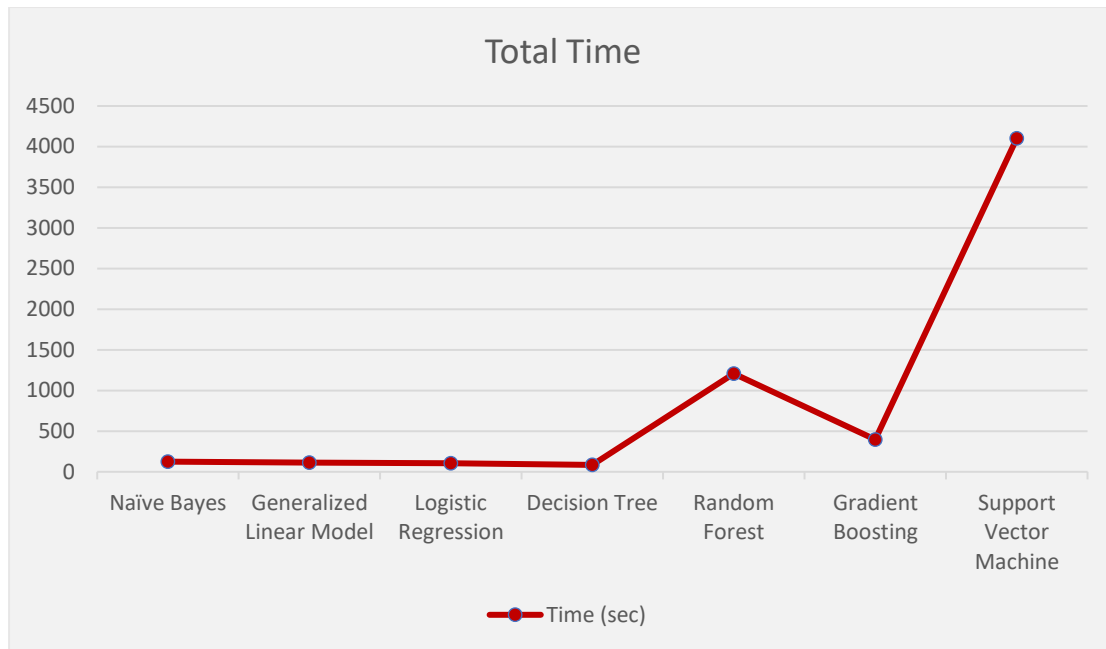
Εικόνα 18 Γραφική απεικόνιση για την ακρίβεια

Όπως φαίνεται από τις εικόνες 16 και 18, που αφορούν την ακρίβεια των αλγορίθμων, γίνεται εύκολα αντιληπτό ότι τη μεγαλύτερη ακρίβεια για την πρόβλεψη του χαρακτηριστικού Machine Failure = 1 την πέτυχε ο αλγόριθμος του Δέντρου Απόφασης (Decision Tree). Επιτεύχθηκε όχι μόνο η καλύτερη απόδοση σε σχέση με τους άλλους αλγορίθμους στο δείκτη απόδοσης της ακρίβειας, αλλά και η καλύτερη απόδοση σε σχέση με το συνολικό χρόνο που χρειάστηκε ο συγκεκριμένος αλγόριθμος για να ολοκληρωθούν οι υπολογισμοί και να προκύψουν τα αποτελέσματα.



Εικόνα 19 Γραφική απεικόνιση σφάλματος ταξινόμησης

Στις εικόνες 17 και 19 φαίνεται ξεκάθαρα ότι το μεγαλύτερο σφάλμα ταξινόμησης προέκυψε από τον αλγόριθμο του Τυχαίου Δάσους (Random Forest) παρόλο που χρονικά αυτός ο αλγόριθμος φαίνεται να είναι ο δεύτερος πιο αργός. Επομένως, η απόδοση του συγκεκριμένου αλγόριθμου για το δείκτη απόδοσης του σφάλματος ταξινόμησης είναι η χειρότερη σε σχέση με τους υπόλοιπους αλγόριθμους για το συγκεκριμένο σύνολο δεδομένων. Αξίζει να σημειωθεί ότι ο αλγόριθμος του Δέντρου Απόφασης (Decision Tree) έχει το μικρότερο σφάλμα ταξινόμησης, κάτι το οποίο ήταν σχετικά αναμενόμενο μετά την ανάλυση που προέκυψε με βάση τον δείκτη απόδοσης της ακρίβειας.



Εικόνα 20 Γραφική παράσταση συνολικού χρόνου διεργασίας

Τέλος, χρήζει ανάλυσης ο συνολικός χρόνος που χρειάστηκε από τον κάθε αλγόριθμο για τη διεκπεραίωση της διεργασίας που αφορά το σύνολο των δεδομένων που χρησιμοποιήθηκαν στη συγκεκριμένη διπλωματική εργασία. Όπως ήταν αναμενόμενο ο αλγόριθμος της Μηχανής Διανυσμάτων Υποστήριξης (Support Vector Machine) εμφάνισε μια πολύ μεγάλη χρονική διαφορά σε σχέση με όλους τους υπόλοιπους αλγόριθμους. Αυτό συνέβη διότι ο συγκεκριμένος αλγόριθμος, για να βγάλει αποτελέσματα, κάνει κάποιους εξαιρετικά πολύπλοκους μετασχηματισμούς δεδομένων και στη συνέχεια καταλαβαίνει πως να διαχωρίσει τα δεδομένα με βάση τις ετικέτες ή τις εξόδους που έχουν οριστεί. Το αντίκτυπο αυτών των πολύπλοκων διεργασιών είναι το χρονικό κόστος, κάτι το οποίο είναι αναπόφευκτο ειδικά σε μεγάλο όγκο δεδομένων. Ωστόσο, η απόδοση αυτού του αλγορίθμου, σε σχέση με την ακρίβεια των αποτελεσμάτων του, είναι πολύ καλή, αφού όπως φάνηκε είχε τη δεύτερη καλύτερη ακρίβεια για την πρόβλεψη του χαρακτηριστικού Machine failure = 1. Επομένως, αν και χρονικά αργός, είναι ένας από τους καλύτερους αλγόριθμους που μπορούν να χρησιμοποιηθούν, ειδικά αν αναλογιστεί κανείς το μεγάλο κόστος που θα προέκυπτε από μια λάθος πρόβλεψη.

Συμπεράσματα

Από την ανάλυση των δεδομένων στην παρούσα διπλωματική εργασία και την ανάδειξη των αποτελεσμάτων που καταδείχθηκαν, είναι εύκολα κατανοητό ότι η χρήση μηχανικής μάθησης είναι κατάλληλη για την πρόβλεψη σφαλμάτων στη βιομηχανία και μάλιστα με υψηλά ποσοστά επιτυχίας. Γενικότερα, η χρήση τεχνητής νοημοσύνης μπορεί να βοηθήσει στην αποπεράτωση πολλών διεργασιών με θετικό αποτέλεσμα όχι μόνο στο κόστος παραγωγής, αλλά και στο χρόνο παραγωγής. Ειδικότερα στη Βιομηχανία 5.0, όπου άνθρωπος και μηχανή συνεργάζονται άμεσα, η χρήση τεχνητής νοημοσύνης μπορεί να χρησιμοποιηθεί για εξατομίκευση των τελικών αποτελεσμάτων, αλλά και για περιβαλλοντικούς σκοπούς.

Για την επίτευξη αυτών των στόχων και για να πραγματοποιηθεί η ορθή χρήση των αλγορίθμων στη μηχανική μάθηση, θα πρέπει για τα πρωτογενή ψηφιακά δεδομένα που συλλέγονται, στο στάδιο της καταγραφής τους να λαμβάνονται όλα τα αναγκαία μέτρα ώστε να αποφεύγεται η καταγραφή θορύβου. Έτσι, επιτυγχάνεται η μείωση του τελικού σφάλματος της εκτίμησης και αποκτούν μεγαλύτερη ακρίβεια οι προγνωστικές και διαγνωστικές μέθοδοι. Αποφεύγοντας τέτοιου είδους σφάλματα και με τη χρήση τεχνητής νοημοσύνης στη Βιομηχανία 5.0 παρέχεται στον εργαζόμενο, σε συνεργασία με τη μηχανή, η δυνατότητα της καθοδήγησης για τη σωστή και έγκαιρη συντήρηση του συστήματος.

Από το πειραματικό μέρος προέκυψαν κάποια πλεονεκτήματα των μεθόδων μηχανικής μάθησης. Αρχικά αναδείχθηκε ότι για το εκάστοτε υπό εξέταση σύστημα, δεν απαιτείται από τον υπεύθυνο λειτουργίας του προγράμματος μηχανικής μάθησης να έχει γνώσεις προγραμματισμού διότι έχουν δημιουργηθεί τα κατάλληλα λογισμικά προγράμματα που διαθέτουν γραφικό περιβάλλον αλληλεπίδρασης φιλικό προς άτομα των ευρύτερων επιστημονικών πεδίων, όπως το RapidMiner που χρησιμοποιήθηκε στην παρούσα διπλωματική εργασία. Επίσης, δε χρειάζονται ειδικές τεχνικές γνώσεις του φυσικού μοντέλου, παρά μόνο βασικές στατιστικές γνώσεις για την πραγματοποίηση της εκτίμησης ύπαρξης δυσλειτουργίας, μέσω της χρήσης των αλγορίθμων μηχανικής μάθησης. Οι αλγόριθμοι μηχανικής μάθησης που χρησιμοποιήθηκαν παρουσίασαν πολύ καλή ακρίβεια (93,3% - 97,1%) στην παραγωγή αποτελεσμάτων και οι περισσότεροι από αυτούς και ταχεία.

Η δυνατότητα εξατομίκευσης της συντήρησης των συστημάτων που προσφέρεται με τη χρήση τεχνητής νοημοσύνης, είναι ένα χαρακτηριστικό της Βιομηχανίας 5.0 το οποίο θα οδηγήσει στην αλλαγή της φιλοσοφίας της συντήρησης των βιομηχανικών μηχανών παραγωγής από προγραμματισμένη ή αποκατάσταση σε προγνωστική και προβλεπτική. Κάτι τέτοιο φυσικά ακολουθείται από διάφορα οφέλη όπως η μείωση του χρόνου επισκευής και λειτουργίας και η επαύξηση της ασφάλειας χρήσης του.

Στόχος στην παρούσα εργασία ήταν η επίτευξη του μέγιστου ποσοστού πρόβλεψης για κάθε αλγόριθμο που χρησιμοποιήθηκε και αντίστοιχα η ελαχιστοποίηση του σφάλματος. Κάποια από τα μοντέλα που δοκιμάστηκαν θα μπορούσαν να εξεταστούν παραπάνω, έχοντας ως στόχο το χρόνο, αλλάζοντας το εύρος των παραμέτρων τους σε μια προσπάθεια βελτιστοποίησης (ελαχιστοποίησης) του χρόνου εκπαίδευσης ή του χρόνου πρόβλεψης. Ο χρόνος πρόβλεψης είναι ο χρόνος που απαιτείται από το μοντέλο για να πραγματοποιήσει τις προβλέψεις, αφού έχει προηγηθεί η εκπαίδευσή του στο σύνολο εκπαίδευσης. Σε ορισμένες περιπτώσεις η μεταβλητή του χρόνου μπορεί να είναι πολύ σημαντική, όπου δηλαδή απαιτείται η εκτέλεση γρήγορων προβλέψεων με όσο το δυνατόν καλύτερα αποτελέσματα.

Τέλος, αξίζει να σημειωθεί ότι όσο μεγαλύτερος είναι ο όγκος των αρχικών ψηφιακών δεδομένων που παρέχονται, τόσο καλύτερο είναι και το ποσοστό επιτυχίας της πρόβλεψης, αλλά με κόστος στο χρονικό διάστημα που απαιτείται για να ολοκληρωθεί η διεργασία, καθώς αυξάνεται και ο χρόνος εκπαίδευσης, αλλά και ο χρόνος πρόβλεψης του κάθε αλγορίθμου που χρησιμοποιείται στη μηχανική μάθηση.

Βιβλιογραφία

- Ampadu, H. (2021, May 01). *ai-pool*. Ανάκτηση από <https://ai-pool.com/a/s/random-forests-understanding>
- Andrade, F. (2021, May 06). *Netilion blog*. Ανάκτηση από <https://netilion.endress.com/blog/what-is-prescriptive-maintenance/>
- Arbour, G. (2019, Feb 01). *fixsoftware*. Ανάκτηση από <https://www.fixsoftware.com/blog/evaluating-maintenance-strategies-select-model-asset-management/>
- Asset. (2021, Jun 03). *assetinfinity*. Ανάκτηση από <https://www.assetinfinity.com/blog/reactive-vs-preventive-vs-predictive>
- Brownlee, J. (2020, Aug 15). *machinelearningmastery*. Ανάκτηση από <https://machinelearningmastery.com/gentle-introduction-gradient-boosting-algorithm-machine-learning/>
- Chauhan, N. S. (2020, Jun 11). *kdnuggets*. Ανάκτηση από <https://www.kdnuggets.com/2020/06/naive-bayes-algorithm-everything.html>
- Cousineau, M. (2020, Jan 28). *fixsoftware*. Ανάκτηση από <https://www.fixsoftware.com/blog/essential-guide-to-comparing-types-of-maintenance-strategies/>
- Deepanshi. (2021, May 25). *analyticsvidhya*. Ανάκτηση από <https://www.analyticsvidhya.com/blog/2021/05/all-you-need-to-know-about-your-first-machine-learning-model-linear-regression/>
- Equiskill. (2018, Jul 05). *equiskill*. Ανάκτηση από <https://www.equiskill.com/understanding-logistic-regression/>
- Francesco Longo, A. P. (2020, Jun 18). *mdpi*. Ανάκτηση από <https://www.mdpi.com/2076-3417/10/12/4182>
- Gandhi, R. (2018, May 05). *towardsdatascience*. Ανάκτηση από <https://towardsdatascience.com/naive-bayes-classifier-81d512f50a7c>
- Gavrilova, Y. (2021, Feb 17). *serokell*. Ανάκτηση από <https://serokell.io/blog/naive-bayes-classifiers>

- Goyal, K. (2021, Jul 15). *upgrad*. Ανάκτηση από <https://www.upgrad.com/blog/data-preprocessing-in-machine-learning/>
- Grover, P. (2017, Dec 09). *mlreview*. Ανάκτηση από <https://blog.mlreview.com/gradient-boosting-from-scratch-1e317ae4587d>
- Hunger, T. (2021, Jun 28). *technologyhunger*. Ανάκτηση από <https://www.technologyhunger.com/industry-5-0-what-you-need-to-know-and-how-it-will-impact-companies/>
- Hupjé, E. (2021, Jul 07). *roadtoreliability*. Ανάκτηση από <https://roadtoreliability.com/types-of-maintenance/>
- Jardine, J. (2020, Jul 30). *mastercontrol*. Ανάκτηση από <https://www.mastercontrol.com/gxp-lifeline/3-things-you-need-to-know-about-industry-5.0/>
- Javatpoint. (2021, Oct 13). *javatpoint*. Ανάκτηση από <https://www.javatpoint.com/machine-learning-support-vector-machine-algorithm>
- Karthik Seetharam, S. S. (2020, Feb 07). *uscjournal*. Ανάκτηση από <https://www.uscjournal.com/articles/artificial-intelligence-cardiac-imaging>
- Limble. (2021, Sep 09). *limblecmms*. Ανάκτηση από <https://limblecmms.com/blog/maintenance-strategy/>
- Martynova, O. (2019, Dec 05). *intellias*. Ανάκτηση από <https://intellias.com/industry-5-0-announcing-the-era-of-intelligent-automation/>
- Matheson, R. (2020, Nov 24). *nickelinstitute*. Ανάκτηση από <https://nickelinstitute.org/blog/2020/november/customising-the-future-the-next-industrial-revolution/>
- Mujtaba, H. (2020, Jun 06). *mygreatlearning*. Ανάκτηση από <https://www.mygreatlearning.com/blog/gradient-boosting/>
- Naveen. (2021, Nov 19). *indiachapter*. Ανάκτηση από <https://indiachapter.in/index.php?/user/article/2/2,4,6,10,17,45/121>

- Niculaescu, O. (2018, Jul 01). *elf11.github*. Ανάκτηση από <https://elf11.github.io/2018/07/01/python-decision-trees-acm.html>
- Oden, C. (2019, Jul 25). *projecttopics*. Ανάκτηση από <https://www.projecttopics.org/understanding-research-data-types-and-how-to-differentiate-them.html>
- Ray, S. (2017, Sep 11). *analyticsvidhya*. Ανάκτηση από <https://www.analyticsvidhya.com/blog/2017/09/naive-bayes-explained/>
- Rossi, B. (2018, Mar 07). *raconteur*. Ανάκτηση από <https://www.raconteur.net/manufacturing/manufacturing-gets-personal-industry-5-0/>
- Sambhav Khurana. (2021, Nov 10). *geeksforgeeks*. Ανάκτηση από <https://www.geeksforgeeks.org/naive-bayes-classifiers/>
- School, R. (2020, Sep 08). *reginsights*. Ανάκτηση από <https://insights.regenesys.net/the-fifth-industrial-revolution-5ir/>
- Schwab, K. (2017, Jan 10). *simio*. Ανάκτηση από <https://www.simio.com/applications/industry-40/industrial-revolution-through-the-ages.php>
- Stefanos Fafalios, P. C. (2020, Apr 02). *gnosisda*. Ανάκτηση από https://www.gnosisda.gr/wp-content/uploads/2020/07/Gradient_Boosting_Implementation.pdf
- Stevens, G. J. (2021, Jul 19). *thingsway*. Ανάκτηση από <https://thingsway.nl/en/onderhoud/voordelen-en-nadelen-periodiek-onderhoud/>
- Tam, T. (2020, Sep 10). *thomastam.medium*. Ανάκτηση από <https://thomasttam.medium.com/simple-and-multiple-linear-regression-for-beginners-c852ffed6700>
- Tarbani, N. (2021, Apr 19). *analyticsvidhya*. Ανάκτηση από <https://www.analyticsvidhya.com/blog/2021/04/how-the-gradient-boosting-algorithm-works/>

- Team, i. (2019, Jun 30). *ied*. Ανάκτηση από <https://ied.eu/project-updates/the-4-industrial-revolutions/>
- Trout, J. (2021, May 19). *reliableplant*. Ανάκτηση από <https://www.reliableplant.com/types-of-maintenance-31812>
- Vollmer, M. (2018, Sep 12). *medium*. Ανάκτηση από <https://medium.com/@marcellvollmer/what-is-industry-5-0-a363041a6f0a>
- Wu, W. (2020, Aug 19). *coastapp*. Ανάκτηση από <https://coastapp.com/blog/maintenance-types/>
- Yulia Gavrilova, O. B. (2020, Sep 23). *serokell*. Ανάκτηση από <https://serokell.io/blog/data-preprocessing>
- Παπακώστα, Β. (2021, Apr 26). *securnet*. Ανάκτηση από <http://www.securnet.gr/2021/04/industry-40-industry-50.html>