



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

ΤΟΜΕΑΣ ΤΕΧΝΟΛΟΓΙΑΣ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΥΠΟΛΟΓΙΣΤΩΝ

**Αναλυτική Δεδομένων για την Πρόβλεψη Ποιότητας
στην Βιομηχανική Παραγωγή**

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

της

ΧΡΙΣΤΙΑΝΑΣ ΠΕΡΙΣΤΕΡΑΣ ΚΟΥΔΙΓΚΕΛΗ

Επιβλέπων : Γρηγόρης Μέντζας
Καθηγητής Ε.Μ.Π.

Αθήνα, Νοέμβριος 2021



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

ΤΟΜΕΑΣ ΤΕΧΝΟΛΟΓΙΑΣ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΥΠΟΛΟΓΙΣΤΩΝ

Αναλυτική Δεδομένων για την Πρόβλεψη Ποιότητας στην Βιομηχανική Παραγωγή

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

της

ΧΡΙΣΤΙΑΝΑΣ ΠΕΡΙΣΤΕΡΑΣ ΚΟΥΔΙΓΚΕΛΗ

Επιβλέπων : Γρηγόρης Μέντζας
Καθηγητής Ε.Μ.Π.

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 5^η Νοεμβρίου 2021.

.....
Γρηγόρης Μέντζας
Καθηγητής Ε.Μ.Π.

.....
Δημήτριος Ασκούνης
Καθηγητής Ε.Μ.Π.

.....
Ιωάννης Ψαράς
Καθηγητής Ε.Μ.Π.

Αθήνα, Νοέμβριος 2021

.....
ΧΡΙΣΤΙΑΝΑ ΠΕΡΙΣΤΕΡΑ ΚΟΥΔΙΓΚΕΛΗ

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

Copyright © Χριστιάνα Περιστέρα Κουδιγκέλη, 2021.

Με επιφύλαξη παντός δικαιώματος. All rights reserved.

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

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

Περίληψη

Η Τέταρτη Βιομηχανική Επανάσταση είναι μια έννοια που αντιπροσωπεύει τη σύγχρονη τάση των βιομηχανικών εταιρειών ως προς την υιοθέτηση τεχνικών και διαδικασιών που επιτρέπονται από την ψηφιοποίηση, το cloud computing, το διαδίκτυο των πραγμάτων και τα μεγάλα δεδομένα για την απόκτηση ανταγωνιστικών πλεονεκτημάτων σε εγχώριες και παγκόσμιες αγορές. Η ανάλυση παρέχει στοιχεία ότι η αποτελεσματική εφαρμογή του ελέγχου ποιότητας από τον μεταποιητικό τομέα είναι πρόκληση, δεδομένου ότι, παρά την ύπαρξη ψηφιακής υποδομής σε συνδυασμό με τις αναλυτικές δυνατότητες αντιμετώπισης μεγάλων δεδομένων, το στάδιο του ποιοτικού ελέγχου εξακολουθεί να κυβερνάται από παραδοσιακές και χειροκίνητες προσεγγίσεις. Αναγνωρίζοντας αυτές τις δυσκολίες, η παρούσα διπλωματική εργασία προτείνει ένα μοντέλο αξιοποίησης στοιχείων που συλλέγονται κατά τη δοκιμή εργοστασιακών προϊόντων ακριβώς πριν τη διάθεσή τους στην αγορά (Zero Hour Test) στοχεύοντας στην πρόβλεψη πιθανού ελαττώματος και του αντίστοιχου ρίσκου ποιότητας που ενδέχεται να εμφανίσουν. Η πλήρης εκμετάλλευση των προβλέψεων μπορεί να επιδράσει σημαντικά στην επιχειρηματική αξία του εκάστοτε προϊόντος και τη βελτιστοποιημένη λήψη αποφάσεων του οργανισμού που την επιδιώκει, καθώς δεν αποτελεί απλό εργαλείο παρακολούθησης, αλλά κυρίως μηχανισμό πρόβλεψης των προβλημάτων, είτε αποτρέποντας την εμφάνισή τους, είτε επιταχύνοντας την άμεση επίλυσή τους. Συνολικά, η μελέτη αποβλέπει στην ενδυνάμωση των συστημάτων διαχείρισης ποιότητας, με σκοπό την ενίσχυση του δεσμού του οργανισμού με τον τελικό πελάτη, την εδραίωση της εικόνας του στην αγορά και την αύξηση των πωλήσεων.

Λέξεις Κλειδιά: Τέταρτη Βιομηχανική Επανάσταση, αναλυτική δεδομένων, πρόβλεψη, ποιοτικός έλεγχος, ρίσκο ποιότητας

Abstract

The Fourth Industrial Revolution is a concept that represents the modern trend of industrial companies in adopting techniques and processes allowed by digitization, cloud computing, the Internet of Things and big data to gain competitive advantages domestically and globally. The analysis provides evidence that the effective implementation of quality control by the manufacturing sector is a challenge, given that, despite the existence of digital infrastructure combined with analytical big data capabilities, the quality control phase is still governed by traditional and manual approaches. Recognizing these difficulties, this thesis proposes a model for utilizing data collected during the factory pre-market (Zero Hour Test) to anticipate a potential defect and the corresponding quality risk they may pose. The full utilization of forecasts can significantly affect the business value of each product and the optimized decision-making of the organization that seeks it, as it is not a simple monitoring tool, but mainly a mechanism for predicting defects, either preventing their occurrence or accelerating their immediate resolution. Overall, the study aims to strengthen quality management systems, in order to reinforce the organization's bond with the end customer, fortify its image in the market and increase sales.

Keywords: Fourth Industrial Revolution, data analytics, forecasting, quality control, quality risk

Ευχαριστίες

Αρχικά, θα ήθελα να αφιερώσω αυτή την διπλωματική εργασία στην οικογένειά μου, ως το ελάχιστο ευχαριστώ για την αμέριστη στήριξη και αγάπη τους σε κάθε βήμα της ζωής μου.

Έπειτα, θα ήθελα να ευχαριστήσω θερμά τον Καθηγητή μου, Γρηγόρη Μέντζα, και τους Δρ. Αλέξανδρο Μπουσδέκη και Κατερίνα Λεπενιώτη για την καθοδήγησή τους και την άριστη συνεργασία που είχαμε στα πλαίσια εκπόνησης της διπλωματικής μου εργασίας.

Τέλος, θα ήθελα να εκφράσω την τύχη που νιώθω για τη γνωριμία και αλληλεπίδραση με την κοινότητα του ΕΜΠ, καθώς ο καθένας τους με ώθησε, με τον τρόπο του, να εξελιχθώ ακαδημαϊκά και προσωπικά.

Πίνακας περιεχομένων

<i>Πίνακας περιεχομένων</i>	8
<i>Πίνακας Πινάκων</i>	10
<i>Πίνακας Εικόνων</i>	11
1 Εισαγωγή	12
1.1 Γενικά	12
1.2 Αντικείμενο και Στόχοι της Διπλωματικής Εργασίας.....	13
1.3 Διάρθρωση της Διπλωματικής Εργασίας.....	13
2 Βιβλιογραφική Επισκόπηση	15
2.1 Τέταρτη Βιομηχανική Επανάσταση.....	15
2.2 Έλεγχος Ποιότητας	18
2.2.1 Στάδια ανάπτυξης προϊόντος.....	20
2.2.2 Εργασίες Ποιότητας.....	25
2.3 Αναλυτική Δεδομένων στη Διαχείριση Ποιότητας	26
2.3.1 Προετοιμασία δεδομένων:	28
2.3.2 Προεπεξεργασία δεδομένων:.....	28
2.3.3 Εξόρυξη δεδομένων:.....	29
2.3.4 Αξιολόγηση και Ερμηνεία	32
2.3.5 Υλοποίηση.....	33
3 Ερευνητική Μεθοδολογία	34
3.1 Συλλογή Δεδομένων.....	34
3.2 Προ-επεξεργασία Δεδομένων.....	37
3.3 Πρόβλεψη Ελαττωματικών με Μηχανική Μάθηση.....	39
3.4 Αξιολόγηση Ρίσκου Εμφάνισης Ελαττώματος	44
4 Μελέτη Περίπτωσης	48

4.1	Υλοποίηση	48
4.2	Πειραματικά Αποτελέσματα.....	49
4.2.1	Εφαρμογή στη Βιομηχανία Οικιακών Συσκευών	49
4.2.2	Δεδομένα	51
4.2.3	Πρόβλεψη Ελαττωματικών.....	59
4.2.4	Υπολογισμός Ρίσκου Ποιότητας	66
5	Συμπεράσματα.....	74
6	Βιβλιογραφία.....	76

Πίνακας Πινάκων

Table 1 Δραστηριότητα ποιοτικού ελέγχου και βελτίωσης ανά Στάδιο Ανάπτυξης Προϊόντος	21
Table 2 Ταξινόμηση ερευνητικών εργασιών ανά στάδιο του κύκλου ζωής της παραγωγής ...	23
Table 3 Μετατροπή μορφής χαρακτηριστικού InsertDate.....	54
Table 4 Δείγμα εγγραφών για χαρακτηριστικό "diff"	54
Table 5 Δείγμα εγγραφών για χαρακτηριστικό "Frequency of DefectID / date".....	56
Table 6 Δομή dataset μετά την κωδικοποίηση χαρακτηριστικών	57
Table 7 Αποτελέσματα απόδοσης Random Forest.....	61
Table 8 Αποτελέσματα απόδοσης Decision Trees	62
Table 9 Αποτελέσματα απόδοσης KNN	63
Table 10 Αποτελέσματα απόδοσης Naive Bayes.....	64
Table 11 Αντιστοίχιση Βάρους Συχνότητας – Προτεινόμενη	66
Table 12 Αντιστοίχιση Βάρους Συχνότητας - Αναθεωρημένη	67
Table 13 Αντιστοίχιση Σοβαρότητας ανά Στάδιο Εντοπισμού Ελαττωματικού.....	68
Table 14 Αντιστοίχιση Βάρους Σοβαρότητας.....	69
Table 15 Αντιστοίχιση Βάρους Αναγνώρισης	69
Table 16 Τελικά Αποτελέσματα Ρίσκου Ποιότητας ανά Ελάττωμα.....	70

Πίνακας Εικόνων

Figure 1 Η εξέλιξη από τη Βιομηχανία 1.0 στη Βιομηχανία 4.0	15
Figure 2 Κύκλος ζωής προϊόντος	21
Figure 3 Στάδια της φάσης παραγωγής.....	23
Figure 4 Ερευνητική Μεθοδολογία.....	34
Figure 5 Δέντρα Αποφάσεων (πηγή: Tutorials and Example).....	41
Figure 6 Τυχαία Δάση (πηγή: Medium).....	42
Figure 7 Βάρος Συχνότητας	46
Figure 8 Βάρος Σοβαρότητας.....	46
Figure 9 Βάρος πηγής αναγνώρισης	47
Figure 10 Διαδικασίες Δοκιμής Μηδενικής Ώρας	50
Figure 11 Συλλογή Δεδομένων από τα επιμέρους στάδια της γραμμής παραγωγής	51
Figure 12 Μελέτη Περίπτωσης - Πηγές Δεδομένων.....	51
Figure 13 Μοναδικές τιμές ανά χαρακτηριστικό	53
Figure 14 Δομή dataset μετά την κατασκευή χαρακτηριστικών	56
Figure 15 Αποτελέσματα απόδοσης Random Forest	62
Figure 16 Αποτελέσματα απόδοσης Decision Trees.....	63
Figure 17 Αποτελέσματα απόδοσης KNN	64
Figure 18 Αποτελέσματα απόδοσης Naive Bayes	65
Figure 19 Κατανομή βάρους συχνότητας συναρτήσεως του αριθμού ελαττωμάτων – A.....	67
Figure 20 Κατανομή βάρους συχνότητας συναρτήσεως του αριθμού ελαττωμάτων - B.....	68
Figure 21 Πλήθος ελαττωμάτων ανά Priority Index.....	73

1

Εισαγωγή

1.1 Γενικά

Η διαχείριση ποιότητας είναι ένας από τους αειθαλείς ερευνητικούς τομείς του σύγχρονου αιώνα που επηρεάζει ολόκληρο τον κύκλο ζωής του προϊόντος, από το σχεδιασμό και την παραγωγή του μέχρι και την τελική διάθεση του στην αγορά. Το επίπεδο ποιότητας στις διαδικασίες παραγωγής αφορά ολοένα και περισσότερο τις κατασκευαστικές εταιρείες, καθώς καλούνται να ανταποκριθούν σε προκλήσεις όπως η αυξανόμενη πολυπλοκότητα και η ποικιλία των προϊόντων, οι πιο σύνθετες αλυσίδες αξίας¹ και ο συντομότερος χρόνος διάθεσης στην αγορά. Παράλληλα, το μέλλον της διαχείρισης ποιότητας στο πλαίσιο της Βιομηχανίας 4.0 προορίζεται να είναι προληπτικό, καθώς οι αρνητικές επιπτώσεις των ποιοτικών ελαττωμάτων πρέπει να αποφευχθούν πριν τη διάθεση των προϊόντων στην αγορά. Η αλλαγή παραδείγματος καθιστά κρίσιμη τη μέτρηση και την πρόβλεψη της ποιότητας των συστημάτων και των προϊόντων πολύ νωρίτερα από τις δυνατότητες παρέχονται από τις παραδοσιακές προληπτικές προσεγγίσεις. Οι παραδοσιακές πρακτικές διαχείρισης ποιότητας στις οποίες

¹ Διευκρινίζεται ότι ο όρος “αλυσίδα αξίας” αναφέρεται στη διαδικασία κατά την οποία οι επιχειρήσεις λαμβάνουν πρώτες ύλες, προσθέτουν αξία σε αυτές μέσω της παραγωγής, της κατασκευής και άλλων διαδικασιών για τη δημιουργία ενός τελικού προϊόντος και στη συνέχεια πωλούν το τελικό προϊόν στους καταναλωτές.

βασίζονται οι περισσότερες κατασκευαστικές εταιρείες, όπως ο σχεδιασμός απαιτήσεων υλικού (MRP), η μεθοδολογία Lean, η θεωρία των περιορισμών (TOC), το Six Sigma και ο Στατιστικός Έλεγχος Διαδικασιών (SPC), δεν έχουν προσαρμοστεί στις σύγχρονες απαιτήσεις και επομένως έχουν χάσει την αποτελεσματικότητά τους. Χάρη στην εισαγωγή πιο προηγμένων τεχνολογικών αισθητήρων και τον γενικότερο εκσυγχρονισμό των διαδικασιών παραγωγής που λαμβάνει χώρα στην Βιομηχανία 4.0 έχει παρατηρηθεί ραγδαία αύξηση στη διαθεσιμότητα των δεδομένων, η οποία επιτάσσει την προσαρμογή των οργανισμών στις τεχνολογικές καινοτομίες, του εκσυγχρονισμού της ανάλυσης δεδομένων και της ένταξης καινοτόμων μοντέλων πρόβλεψης των προβλημάτων, με σκοπό είτε να αποτρέψουν την εμφάνισή τους, είτε να επιταχύνουν την άμεση επίλυσή τους. Σε αυτό το πλαίσιο, η αναλυτική δεδομένων δύναται να δράσει ως αρωγός, μέσω των μεθόδων και οι αλγόριθμων που υποστηρίζουν την αξιοποίηση του μεγάλου όγκου δεδομένων στην πρόβλεψη ποιότητας.

1.2 Αντικείμενο και Στόχοι της Διπλωματικής Εργασίας

Η διπλωματική εργασία για πρόγνωση ποιότητας στη Βιομηχανική Παραγωγή στοχεύει στη διεξαγωγή ανάλυσης σχετικά με την ποιότητα πρόβλεψης που υποστηρίζεται από την αναλυτική δεδομένων. Συγκεκριμένα, επιδιώκει να διευκρινίσει πώς οι μέθοδοι και οι αλγόριθμοι μηχανικής μάθησης μπορούν να διευκολύνουν την προγνωστική ποιότητα σε περιβάλλοντα της Βιομηχανίας 4.0 και πώς οι βιομηχανίες μπορούν να προχωρήσουν πέρα από τις παραδοσιακές μεθοδολογίες ελέγχου και διασφάλισης ποιότητας με τη χρήση βασικών τεχνολογιών και μεθόδων αναλυτικής δεδομένων. Επεκτείνοντας τη διεξαγωγή βιβλιογραφικής ανασκόπησης σχετικά με τις μεθόδους και τους αλγόριθμους ανάλυσης δεδομένων στις διαδικασίες ποιότητας παραγωγής, η παρούσα μελέτη επιδιώκει, επιπλέον, να εξετάσει, να εφαρμόσει και να συγκρίνει αλγόριθμους μηχανικής μάθησης ως προς την πρόβλεψη ελαττωματικών προϊόντων, όπως και να προσδιορίσει το βαθμό του ρίσκου που το κάθε ελάττωμα εισάγει στις παραγωγικές διαδικασίες των κατασκευαστών.

1.3 Διάρθρωση της Διπλωματικής Εργασίας

Η διπλωματική εργασία διαρθρώνεται σε 6 κεφάλαια. Στο κεφάλαιο 2 παρουσιάζεται το θεωρητικό υπόβαθρο πάνω στο οποίο βασίζεται η προτεινόμενη προσέγγιση. Συγκεκριμένα, αναλύονται οι βασικές έννοιες για την Τέταρτη Βιομηχανική Επανάσταση και τον Έλεγχο Ποιότητας στο σύγχρονο καθεστώς, καθώς και η συμβολή της Αναλυτικής Δεδομένων στη Διαχείριση Ποιότητας. Στο κεφάλαιο 3 καταγράφεται η ερευνητική μεθοδολογία για την πρόβλεψη ποιότητας στη γραμμική παραγωγής εργοστασιακών προϊόντων, εξερευνώντας ως άξονες τον τρόπο καθορισμού, συλλογής και προ-επεξεργασίας των δεδομένων, όπως και την

μετέπειτα αξιοποίηση αυτών στην πρόβλεψη ελαττωμάτων και την αξιολόγηση του ρίσκου ποιότητας που αυτά εισάγουν. Το κεφάλαιο 4 επεκτείνεται στην υλοποίηση που έγινε στα πλαίσια της διπλωματικής εργασίας, στα αποτελέσματα των υπολογισμών και των προβλέψεων, και στη σύγκριση της αποτελεσματικότητας διαφορετικών αλγορίθμων μηχανικής μάθησης βάσει των διαθέσιμων στοιχείων. Στο κεφάλαιο 5 δίνονται τα συνολικά συμπεράσματα που προκύπτουν από την παρούσα εργασία, συνοδευόμενα από τις προοπτικές για μελλοντική επέκταση στο συγκεκριμένο αντικείμενο. Τέλος, στο κεφάλαιο 6 παρατίθεται η βιβλιογραφία που αξιοποιήθηκε στο σύνολο της μελέτης.

2

Βιβλιογραφική Επισκόπηση

2.1 Τέταρτη Βιομηχανική Επανάσταση

Η ταχεία πρόοδος στην εκβιομηχάνιση και στις μεθόδους πληροφόρησης έχει προκαλέσει τεράστια πρόοδο στην ανάπτυξη της επόμενης γενιάς τεχνολογίας παραγωγής, επανεφευρίσκοντας υπηρεσίες, προϊόντα και μεθόδους παραγωγής. Για να γίνει καλύτερα αντιληπτή η έντονη εξέλιξη και οι σημαντικές αλλαγές στις μηχανικές και κατασκευαστικές τεχνολογίες και πρακτικές των βιομηχανιών, όπως και οι σύγχρονες προκλήσεις που καλούμαστε να αντιμετωπίσουμε, δίνεται ακολούθως το ιστορικό πλαίσιο μέχρι την εδραίωση της Βιομηχανίας 4.0.

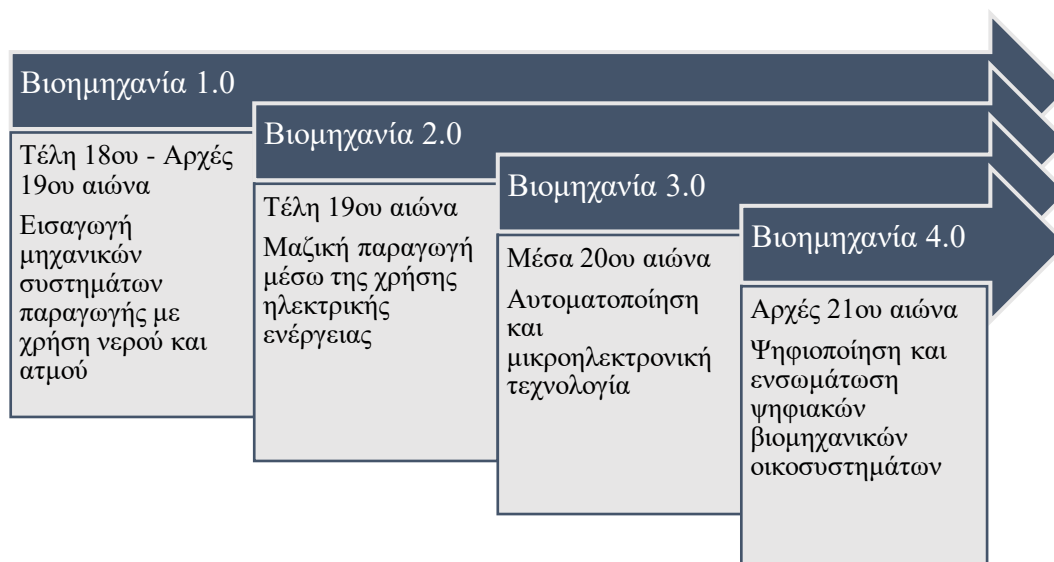


Figure 1 Η εξέλιξη από τη Βιομηχανία 1.0 στη Βιομηχανία 4.0

Η Πρώτη Βιομηχανική Επανάσταση ξεκίνησε στα τέλη του δέκατου όγδοου και στις αρχές του δέκατου ένατου αιώνα και αντιπροσωπεύτηκε από την εισαγωγή μηχανικών συστημάτων παραγωγής που χρησιμοποιούσαν νερό και ατμό. Στη συνέχεια, η Δεύτερη Βιομηχανική Επανάσταση χαρακτηρίστηκε από τη μαζική παραγωγή μέσω της χρήσης ηλεκτρικής ενέργειας. Στα μέσα του εικοστού αιώνα έγινε η μετάβαση στην Τρίτη Βιομηχανική Επανάσταση, η οποία εισήγαγε την αυτοματοποίηση και τη μικροηλεκτρονική τεχνολογία στην παραγωγή. (Xu, Xu and Li 2018). Σήμερα βρισκόμαστε στο αποκορύφωμα της Τέταρτης Βιομηχανικής Επανάστασης που ορίζεται από τη συγχώνευση των βιομηχανικών μονάδων και των ψηφιακών συστημάτων επικοινωνίας, η οποία τείνει σε μια καθολική ενοποίηση πληροφοριών, προϊόντων, διαδικασιών, μηχανών, τοποθεσιών και ανθρώπων (Chaki et. al 2019).

Η Βιομηχανία 4.0 αποτελεί ένα εν εξελίξει αποτέλεσμα της Τέταρτης Βιομηχανικής Επανάστασης, όπου η εισαγωγή δομικών αλλαγών στις παραγωγικές διαδικασίες ενίσχυσε τις ανατρεπτικές καινοτομίες και τις αλλαγές πρακτικών, οι οποίες είχαν ισχυρές επιπτώσεις στην παραγωγικότητα, δημιουργώντας τις προϋποθέσεις για νέα επιχειρηματικά μοντέλα. Πρόκειται για μια έννοια που εμφανίστηκε για πρώτη φορά το 2011 με στόχο τον χαρακτηρισμό των πολύ ψηφιοποιημένων διαδικασιών παραγωγής όπου οι πληροφορίες ρέουν μεταξύ μηχανών σε ελεγχόμενο περιβάλλον, έτσι ώστε η ανθρώπινη παρέμβαση να μειωθεί στο ελάχιστο. Η ιδέα δημιουργήθηκε και αναπτύχθηκε από τις γερμανικές βιομηχανικές και ακαδημαϊκές κοινότητες με την υποστήριξη της γερμανικής κυβέρνησης με την πρόθεση να πλαισιώσουν και να αναπτύξουν τις βιομηχανικές ικανότητες της χώρας που υποστηρίχθηκαν από την ψηφιοποίηση των διαδικασιών παραγωγής σε διάφορους βιομηχανικούς τομείς. Έκτοτε, ο όρος «Βιομηχανία 4.0» χρησιμοποιείται ως συνώνυμο των αυτοδύναμων διαδικασιών παραγωγής που καθίστανται δυνατές με την ικανότητα μηχανών και συσκευών να επικοινωνούν μεταξύ τους μέσω ψηφιακής διασυνδεσιμότητας κατά μήκος της αλυσίδας αξίας (Castelo-Branco, Cruz-Jesus, Oliveira 2019).

Στη συστηματική τους ανασκόπηση της βιβλιογραφίας για τη Βιομηχανία 4.0, οι Kamble et al. (2018) επιβεβαίωσαν ότι αυτή η τεχνολογία υπόσχεται ένα νέο κύμα επανάστασης στην παραγωγή. Αυτή η νέα προοπτική επικεντρώνεται στον τρόπο με τον οποίο η βιομηχανική παραγωγή μπορεί να συνδυάσει τις προόδους στην ψηφιοποίηση για να παράγει τη μέγιστη απόδοση, ενώ χρησιμοποιεί ελάχιστους πόρους (Kamble et al., 2018; Sony and Naik, 2019b). Η ψηφιοποίηση των οργανισμών δημιουργεί μοναδικές ευκαιρίες για τη διαχείριση της ποιότητας των προϊόντων και των υπηρεσιών που παρέχονται από τον οργανισμό, καθώς οι παραδοσιακές ποιοτικές μέθοδοι διαχείρισης ποιότητας θα πρέπει να αντικατασταθούν ή να

ενισχυθούν από τις τεχνολογικές εξελίξεις για να καταλήξουν σε νέα βέλτιστα στη λειτουργική αριστεία, την απόδοση και την καινοτομία (Gunasekaran et al., 2019). Οι προκλήσεις που αντιμετωπίζουν οι παραδοσιακές πρακτικές διαχείρισης ποιότητας ήταν ότι οι ανάγκες των πελατών αλλάζουν συνεχώς και είναι δύσκολο να διατηρηθεί ένα υψηλό επίπεδο ποιότητας (Bergman and Klefsjo, 2010). Επί του παρόντος, πολλές εταιρείες προφανώς εξακολουθούν να αγωνίζονται με τη διαχείριση της ποιότητας, δεδομένων των πολυάριθμων ανακλήσεων προϊόντων από την αυτοκινητοβιομηχανία (Gunasekaran et al., 2019) και άλλους κλάδους. Το πόσο πολύ οι παραδοσιακές πρακτικές και μέθοδοι διαχείρισης ποιότητας έχουν απορροφήσει τις αλλαγές στα στάδια ανάπτυξης του προϊόντος, τη συμπίεση του χρόνου κύκλου και την προσπάθεια των εργαζομένων να ανταποκρίνονται στη ζήτηση και τις προσδοκίες των πελατών είναι μια πρόκληση (Gunasekaran et al., 2019).

Δεδομένης της έρευνας που έχει ήδη πραγματοποιηθεί σχετικά με την ποιότητα γενικά όσον αφορά τις δεξιότητες που απαιτούνται (Ershadi, Najafi and Soleimani, 2019), είναι σκόπιμο να γίνεται αποτελεσματική διαχείριση αυτής. Περαιτέρω προκλήσεις αποτελούν η κατανομή πόρων για έρευνα και καινοτομία για την ανάπτυξη νέων μεθόδων ποιότητας, η παγκόσμια τυποποίηση των προτύπων ποιότητας όπου οι εταιρείες παράγουν από διαφορετικές τοποθεσίες (Kline and Hutchins, 2017) και, τέλος, η διατήρηση της ποιότητας κατά την προσαρμογή των προϊόντων (Du et al., 2006).

Ο όρος «Ποιότητα 4.0» πλαισιώθηκε από τον Dan Jacob, διευθυντή έρευνας και κύριο αναλυτή της LNS Research, μια κορυφαία εταιρεία έρευνας και παροχής συμβουλών στον τομέα της παραγωγής (Johnson, 2019). Ενώ βασίζεται σε και βελτιώνει τις παραδοσιακές μεθόδους ποιότητας, η Ποιότητα 4.0 ευθυγραμμίζει στενά τη διαχείριση ποιότητας με τη Βιομηχανία 4.0 για να επιτρέψει την αποδοτικότητα της επιχείρησης, την απόδοση, την καινοτομία και τα βελτιωμένα επιχειρηματικά μοντέλα. Τεχνολογίες όπως τα κυβερνοφυσικά συστήματα, το Διαδίκτυο των Πραγμάτων και το Υπολογιστικό Νέφος χρησιμοποιούνται για την κάλυψη των απαιτήσεων όσον αφορά την ποιότητα του σχεδιασμού, την ποιότητα της συμμόρφωσης και την ποιότητα της απόδοσης.

Η Ποιότητα 4.0 είναι, επομένως, η ψηφιοποίηση της ποιότητας σχεδιασμού, της ποιότητας συμμόρφωσης και της ποιότητας απόδοσης χρησιμοποιώντας σύγχρονες τεχνολογίες. Η εφαρμογή ψηφιακών τεχνολογιών μπορεί να αλλάξει την ποιότητα με διάφορους τρόπους, για παράδειγμα, ένας οργανισμός μπορεί να παρακολουθεί τις διαδικασίες και να εξάγει δεδομένα από αισθητήρες σε πραγματικό χρόνο (Sony, 2018). Τα μεγάλα δεδομένα που παράγονται από αυτούς τους αισθητήρες μπορούν να αναλυθούν περαιτέρω για να προβλέψουν ζητήματα

ποιότητας και ανάγκες συντήρησης του οργανισμού (Johnson, 2019). Η ποιότητα 4.0 αυτή καθαυτή είναι κάτι πολύ περισσότερο από τεχνολογία. Είναι μια νέα μέθοδος με την οποία μπορούν να χρησιμοποιηθούν ψηφιακά εργαλεία, ώστε να μπορεί να βελτιωθεί η ικανότητα των οργανισμών να παρέχουν σταθερά προϊόντα υψηλής ποιότητας. Σε μια πρόσφατη μελέτη από την Boston Consulting Group, την American Society for Quality και άλλους, διαπιστώθηκε ότι ο κατασκευαστικός τομέας θα ωφεληθεί πάρα πολύ λόγω της εφαρμογής της Ποιότητας 4.0 (Kupper et al., 2019). Πιο συγκεκριμένα, η εφαρμογή έξυπνων διαγνωστικών εργαλείων θα επηρεάσει την ποιότητα του σχεδιασμού, καθώς τα δεδομένα που σχετίζονται με τη χρήση του προϊόντος θα μπορούν να αναμεταδοθούν στους σχεδιαστές μέσω της ενσωμάτωσης από άκρο σε άκρο της Βιομηχανίας 4.0 (Cheng et al., 2016). Η ποιότητα 4.0 δύναται, λοιπόν, να οδηγήσει σε βελτιώσεις σε όλη την αλυσίδα αξίας.

2.2 Έλεγχος Ποιότητας

Τα προβλήματα ποιότητας στη Βιομηχανία 4.0 μπορεί να περιλαμβάνουν πολλές μεταβλητές εισόδου και εξόδου που δεν είναι εύκολο να μοντελοποιηθούν ή/και να βελτιστοποιηθούν. Όπως επισημαίνουν οι Yang και Trewn (2004), η εξόρυξη δεδομένων και η ανακάλυψη γνώσης σε βάσεις δεδομένων έχουν χρησιμοποιηθεί με επιτυχία για την επίλυση προβλημάτων βελτίωσης της ποιότητας και ελέγχου που περιλαμβάνουν δεδομένα πολλών μεταβλητών σε διάφορα στάδια του κύκλου ζωής του προϊόντος ή μιας διαδικασίας. Ωστόσο, απαιτείται μια ανασκόπηση τέτοιων προσεγγίσεων επίλυσης προβλημάτων για την καθοδήγηση των επαγγελματιών. (Gülser et. al, 2011)

Καθώς η ποιότητα αποκτούσε ολοένα και μεγαλύτερη σημασία με την πάροδο του χρόνου, έχουν προκύψει πολλές πρωτοβουλίες και έννοιες ποιότητας. Η επιθεώρηση (100%), ο στατιστικός έλεγχος ποιότητας (SQC), ο συνολικός ποιοτικός έλεγχος (TQC), τα μηδενικά ελαττώματα, η διαχείριση ολικής ποιότητας (TQM), το kaizen, τα πρότυπα ποιότητας ISO 9000, τα προγράμματα βραβείων ποιότητας (Malcolm Baldrige, Ευρωπαϊκό Βραβείο Ποιότητας κ.λπ.), το 6σ, το DFSS και το lean six sigma ήταν από τα πιο αναγνωρισμένα (Fasser & Brettner, 2002; Montgomery, 2005). Μια προφανής τάση στην εξέλιξη αυτών των πρωτοβουλιών είναι η έμφαση σε πιο προληπτικές προσεγγίσεις και «προηγούμενες διαδικασίες (σχεδιασμού)» (Kolarik, 1995; Montgomery, 2005; Phadke, 1989; Taguchi, Chowdhury, & Taguchi, 2000). Επιπλέον, η έμφαση στα τελικά αποτελέσματα σε μικρότερες χρονικές περιόδους έχει επίσης αποκτήσει σημασία και οδήγησε στα προγράμματα ποιότητας 6σ. Αυτά τα προγράμματα έχουν λάβει σημαντική αναγνώριση σε όλο τον κόσμο λόγω της επιτυχίας τους, ειδικά στη βελτίωση των διαδικασιών "on-line" ή μεταγενέστερων διαδικασιών χρησιμοποιώντας τη λεγόμενη προσέγγιση DMAIC (ορισμός-μέτρηση-ανάλυση-βελτίωση-

έλεγχος) για να φτάσουν τα Six Sigma επίπεδα ποιότητας (Brady & Allen, 2006; Fasser & Brettner, 2002). Συνειδητοποιώντας ότι τόσο ψηλά επίπεδα ποιότητας δεν μπορούν να επιτευχθούν χωρίς την ανάπτυξη ισχυρών σχεδίων προϊόντων ή διαδικασιών, τα προγράμματα Design For Six Sigma (DFSS) έχουν εφαρμοστεί στην πράξη, τα οποία ακολουθούν παρόμοια προσέγγιση όπως η DMADV (ορισμός-μέτρηση-ανάλυση-ανάπτυξη-επαλήθευση) (Brady & Allen, 2006).

Η σύγχρονη βιομηχανία δημιουργεί φυσικά μοντέλα ή μοντέλα συμπεριφοράς για σχεδόν οποιοδήποτε προϊόν, σύστημα, υποσύστημα ή λειτουργική διαδικασία. Τέτοια μοντέλα βασίζονται σε ισχυρά ηλεκτρονικά εργαλεία όπως τα MATLAB, Simulink, modeFRONTIER (Ayele, Gur και Rosen 2013), LS-DYNA (Dorsselaer et al. 2019), μεταξύ άλλων. Αυτά τα μοντέλα περιγράφουν το υπό ανάπτυξη προϊόν/σύστημα για οποιοδήποτε σχετικό στάδιο του κύκλου ζωής. Τα μοντέλα παρέχουν στους μηχανικούς ποσοτικά δεδομένα σχετικά με τις επιδόσεις των προϊόντων/συστημάτων με βάση την τεχνογνωσία σε συγκεκριμένο τομέα. Τα μοντέλα χρησιμοποιούνται για την προσομοίωση της συμπεριφοράς προϊόντων/συστημάτων για διαφορετικές εσωτερικές και εξωτερικές εισόδους. Μέσω αυτών των προσομοιώσεων, είναι δυνατό να αξιολογηθούν εναλλακτικές σχεδιαστικές λύσεις για να βρεθούν καλύτερες σχεδιαστικές λύσεις και να υποστηριχθεί ο σχεδιασμός για στιβαρότητα, ποιότητα, αξιοπιστία και ασφάλεια.

Κατά τη φάση της μοντελοποίησης και της προσομοίωσης, είναι δυνατή η ταχεία προετοιμασία πρωτοτύπων με βάση το τρέχον μοντέλο. Αυτά τα πρωτότυπα μπορούν να παραχθούν με παραδοσιακή παραγωγή ή τρισδιάστατη εκτύπωση με βάση το μοντέλο. Η δοκιμή των πρωτοτύπων σε διαφορετικά σενάρια είναι το πρώτο βήμα για την επικύρωση της ποιότητας των προϊόντων/συστημάτων και των μοντέλων τους.

Σε μεταγενέστερα στάδια της διαδικασίας ανάπτυξης, υπάρχουν δοκιμές προϊόντος/συστήματος σε επιλεγμένα σημεία του σχεδιασμού και της λειτουργίας του προϊόντος/συστήματος. Τα αποτελέσματα αυτών των δοκιμών συγκρίνονται με αποτελέσματα προσομοίωσης για την επικύρωση των μοντέλων προϊόντος/συστήματος και την επαλήθευση της ποιότητας του σχεδιασμού.

Έτσι, η μοντελοποίηση και η προσομοίωση είναι πολύ κρίσιμα εργαλεία για τους μηχανικούς ποιότητας και θα πρέπει να ενσωματωθούν στη διαδικασία μηχανικής ποιότητας, καθώς

παρέχουν πρόσθετα πλεονεκτήματα και οφέλη για την ποιοτική μηχανική στη νέα εποχή. Μερικά από αυτά είναι:

- Τα μοντέλα θεωρούνται η «αλήθεια» για τις παραμέτρους και τη συμπεριφορά του προϊόντος. Το μοντέλο θα πρέπει να επικυρωθεί και να θεωρείται τεκμηρίωση του προϊόντος.
- Τα μοντέλα ενισχύουν τις δυνατότητες ανάλυσης δεδομένων δημιουργώντας πολύτιμα δεδομένα μέσω προσομοιώσεων.
- Τα μοντέλα αποτελούν επίσης πηγή για τη μηχανική διεργασιών. Συντομεύουν το στάδιο της μηχανικής παραγωγής και σώζουν πολλές παρεξηγήσεις και σφάλματα κατά μήκος της αλυσίδας παραγωγής και ολοκλήρωσης.
- Είναι επίσης δυνατή η δημιουργία κώδικα λογισμικού αυτόματης προσομοίωσης προϊόντος. Εξαλείφει τη μη αυτόματη προσπάθεια κωδικοποίησης από το σχεδιασμό, επομένως, επιταχύνοντας τη διαδικασία ανάπτυξης ενώ μειώνει την πιθανότητα σφαλμάτων σε σύγκριση με τη χειροκίνητη κωδικοποίηση από απαιτήσεις. Τα μοντέλα υποστηρίζουν και διευκολύνουν τις πιθανές αλλαγές κατά τη διάρκεια του κύκλου ζωής των προϊόντων.
- Τα μοντέλα προσομοίωσης μπορούν επίσης να χρησιμοποιηθούν για την προετοιμασία του υλικού λειτουργίας και εκπαίδευσης για την παραγωγή, τη λειτουργία και τη συντήρηση των προϊόντων.
- Με βάση την τέταρτη βιομηχανική επανάσταση και τον στόχο της μηχανικής ποιότητας με γνώμονα τα δεδομένα και την τεκμηρίωση, οι δραστηριότητες μοντελοποίησης και προσομοίωσης αποτελούν ουσιαστικό μέρος της. (Zonnenshain & Kenett, 2020)

2.2.1 Στάδια ανάπτυξης προϊόντος

Πολλές δραστηριότητες ποιοτικού ελέγχου και βελτίωσης λαμβάνουν χώρα καθ' όλη τη διάρκεια του κύκλου ζωής ενός προϊόντος, όπως συνοψίζεται στον πίνακα 1. Αυτή η περίληψη προσαρμόστηκε από τον Phadke (1989) και ταξινομεί τη διαδικασία ανάπτυξης προϊόντος σε τρεις φάσεις:

- σχεδιασμός προϊόντος και διαδικασία παραγωγής
- παραγωγή του προϊόντος
- χρήση του προϊόντος από τον πελάτη.

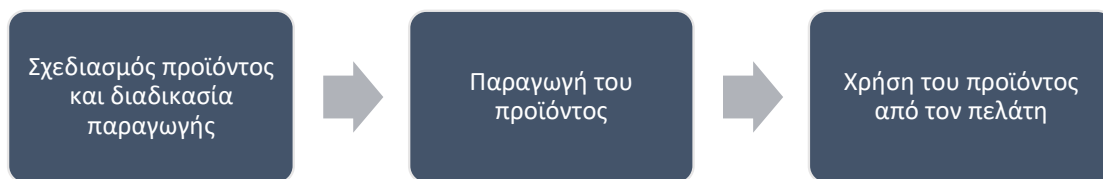


Figure 2 Κύκλος ζωής προϊόντος

Table 1 Δραστηριότητα ποιοτικού ελέγχου και βελτίωσης ανά Στάδιο Ανάπτυξης Προϊόντος

Στάδιο ανάπτυξης προϊόντος	Δραστηριότητα ποιοτικού ελέγχου και βελτίωσης
Σχεδιασμός προϊόντος και διαδικασία παραγωγής	<ul style="list-style-type: none"> - Σχεδιασμός έννοιας - Σχεδιασμός παραμέτρων - Σχεδιασμός ανοχής
Παραγωγή του προϊόντος	<ul style="list-style-type: none"> - Επιθεώρηση/Έλεγχος - Ανάλυση ποιότητας - Έλεγχος διαδικασίας - Παρακολούθηση ποιότητας
Χρήση του προϊόντος από τον πελάτη.	<ul style="list-style-type: none"> - Εγγύηση και επισκευή/αντικατάσταση

Σχεδιασμός προϊόντος και διαδικασία παραγωγής:

Τρία κρίσιμα στάδια της φάσης σχεδιασμού του προϊόντος και της διαδικασίας παραγωγής είναι:

- ο εννοιολογικός σχεδιασμός των προϊόντων/διαδικασιών,
- ο καθορισμός τιμών των παραμέτρων σχεδιασμού τους (π.χ. σχεδιασμός παραμέτρων ή βελτιστοποίηση σχεδίασης) και
- ο καθορισμός ανοχών για τις παραμέτρους σχεδιασμού (δηλαδή σχεδιασμός ανοχής).

Κατά τον εννοιολογικό σχεδιασμό ενός προϊόντος επιλέγεται η καταλληλότερη αρχιτεκτονική και τεχνολογία για την επίτευξη των επιθυμητών λειτουργιών του προϊόντος. Στη συνέχεια, καθορίζονται οι καλύτερες ρυθμίσεις για τις ελεγχόμενες παραμέτρους σχεδίασης προϊόντος/διαδικασίας. Σε αυτό το στάδιο σχεδιασμού παραμέτρων, είναι τυπικό να υποθέτουμε μεγάλες ανοχές σε μη ελεγχόμενους παράγοντες θορύβου και να αναζητούμε ρυθμίσεις παραμέτρων που δεν είναι ευαίσθητες στις δυσμενείς επιπτώσεις των παραγόντων θορύβου. Εάν δεν μπορούν να βρεθούν τέτοιες ρυθμίσεις, τότε οι ανοχές για τη μεταβλητότητα των τιμών των παραμέτρων προϊόντος/διαδικασίας μειώνονται επιλεκτικά και εξετάζεται η

χρήση υλικών υψηλότερης ποιότητας και εξελιγμένων συστημάτων ελέγχου για περαιτέρω μείωση της μεταβλητότητας.

Παραγωγή του προϊόντος:

Οι κύριες δραστηριότητες βελτίωσης της ποιότητας και ελέγχου κατά τη φάση παραγωγής του προϊόντος είναι:

- καθορισμός παραγόντων που επηρεάζουν σημαντικά την ποιότητα,
- μοντελοποίηση σχέσεων μεταξύ των χαρακτηριστικών ποιότητας εισόδου και εξόδου και
- πρόβλεψη επιπέδων ποιότητας για ένα δεδομένο σύνολο παραμέτρων εισόδου.

Τέτοιες αναλύσεις ποιότητας παρέχουν ανεκτίμητη ανατροφοδότηση για το σχεδιασμό (ή τον επανασχεδιασμό) προϊόντος/διαδικασίας και άλλες διορθωτικές ενέργειες για τη βελτίωση ποιότητας. Οι πιο συνηθισμένες εφαρμογές εξόρυξης δεδομένων αναμένεται να εμφανίζονται σε τέτοιες δραστηριότητες λόγω της ευκολίας συλλογής σχετικών δεδομένων, εκτέλεσης της ανάλυσης και ερμηνείας των αποτελεσμάτων.

Όσον αφορά τον έλεγχο της ποιότητας των προϊόντων, η παλαιότερη προσέγγιση είναι να επιθεωρούνται όλες οι μονάδες που παράγονται και να απορρίπτονται/επισκευάζονται οι ελαττωματικές. Γενικά, η αυτοματοποιημένη οπτική επιθεώρηση (AVI) έχει γίνει αναπόφευκτη στη σύγχρονη παραγωγή (Newman & Jain, 1995). Στην αυτοματοποιημένη οπτική επιθεώρηση έχουν χρησιμοποιηθεί πολλές συμβατικές λειτουργίες επεξεργασίας εικόνας, ωστόσο, πολλές πρόσφατες εργασίες επικεντρώνονται και στην ενσωμάτωση τεχνικών από τον τομέα της τεχνητής νοημοσύνης και εξόρυξης δεδομένων στα συστήματα. (Gülser et. al, 2011)

Η φάση παραγωγής, χωρίζεται με τη σειρά της σε τρία στάδια, όπως απεικονίζεται στο παρακάτω σχήμα:

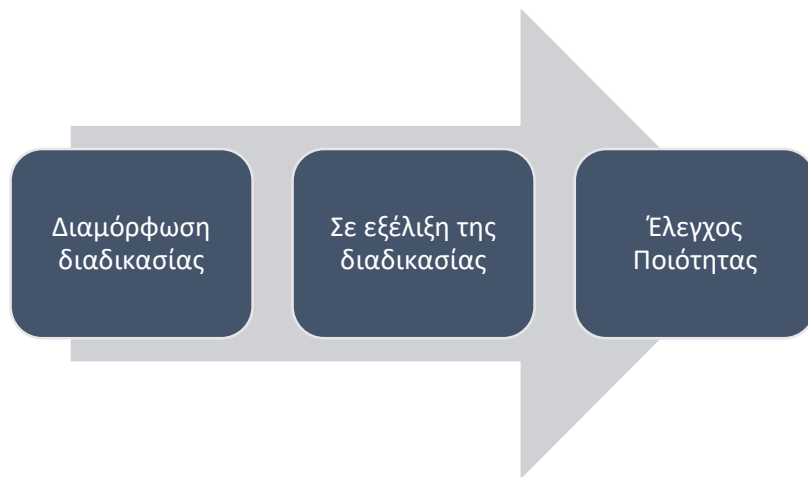


Figure 3 Στάδια της φάσης παραγωγής

Παρακάτω δίνεται η ταξινόμηση της βιβλιογραφικής ανασκόπησης σύμφωνα με το στάδιο του κύκλου ζωής της παραγωγής στο οποίο αναφέρεται η κάθε μελέτη:

Table 2 Ταξινόμηση ερευνητικών εργασιών ανά στάδιο του κύκλου ζωής της παραγωγής

Στάδιο του κύκλου ζωής της παραγωγής	Σχετική Βιβλιογραφία
Διαμόρφωση διαδικασίας	Psarommatis et al., 2020a; Zonnenshain, and Kenett, 2020; Eger et al., 2018b; Lee et al., 2014; Konrad et al., 2013; Chamkalani et al., 2014; Kim, and Ryu, 2020; Schnell et al., 2019; Chongwatpol, 2015
Σε εξέλιξη της διαδικασίας	Psarommatis, and Kiritsis, 2018; Psarommatis et al., 2020a; Zonnenshain, and Kenett, 2020; Jun et al., 2020; Schmitt et al., 2020; Liu et al., 2019; Chatterjee et al., 2019; Bai et al., 2018; Peres et al., 2018; Bustillo, and Correa, 2012; Lokrantz et al., 2018; Oliff, and Liu, 2017; He et al., 2017; Bai et al., 2017; Lee et al., 2016; Angione et al., 2019; Escobar, and Morales-Menendez, 2018; Lee et al., 2018; Frumosu et al., 2018; Jin et al., 2020; Bai et al., 2019; Teucke et al., 2020; Lindström et al., 2020; Wang, 2013; Lieber et al., 2013; Lieber et al., 2012; Hao et al., 2016; Paul, 2016; Liu et al., 2019b; Sun et al., 2017; Wuest et al., 2014; Li et al., 2012; Luckow et al., 2018; Franciosa et al., 2020; Escobar et al., 2020; Schreiber et

	al., 2019; Eger et al., 2020; Hamzeh et al., 2020; Wang et al., 2019; Kim et al., 2012; Teti, 2015
Έλεγχος ποιότητας	Psarommatis et al., 2020a; Zonnenshain, and Kenett, 2020; Nalbach et al., 2018

Είναι εμφανές ότι συντριπτική πλειοψηφία των ερευνητικών εργασιών ασχολείται με την ποιότητα της διαδικασίας, δηλαδή κατά τη λειτουργία της γραμμής παραγωγής. Η εμφάνιση του Διαδικτύου των Πραγμάτων και η αυξανόμενη χρήση αισθητήρων στο κατάστημα για την παρακολούθηση της παραγωγικής διαδικασίας και των μηχανών έχει παράσχει τεράστιο όγκο δεδομένων που χρησιμοποιούνται, μεταξύ άλλων, στο πλαίσιο της διαχείρισης ποιότητας.

Από την άλλη πλευρά, η διαμόρφωση της διαδικασίας βασίζεται κυρίως σε μη αυτόματες διεργασίες που εκτελούνται με βάση τις ειδικές γνώσεις λόγω της υψηλής πολυπλοκότητας, ποικιλίας και κρισιμότητας των αποφάσεων που πρέπει να ληφθούν. Επομένως, οι αλγόριθμοι ανάλυσης δεδομένων στη διαμόρφωση διεργασιών δεν έχουν χρησιμοποιηθεί σε μεγάλο βαθμό.

Τέλος, στη βιβλιογραφική μας ανασκόπηση, βρήκαμε μόνο τρεις εργασίες που ασχολούνται με το στάδιο του ποιοτικού ελέγχου, και όλες προτείνουν ένα εννοιολογικό πλαίσιο αντί για έναν νέο αλγόριθμο ανάλυσης δεδομένων. Η βιβλιογραφική ανασκόπηση αποκάλυψε ότι, παρά την επέκταση των αλγορίθμων ανάλυσης δεδομένων, η οποία διευκολύνεται από τις αυξανόμενες ποσότητες δεδομένων, το στάδιο ποιοτικού ελέγχου εξακολουθεί να διέπεται από παραδοσιακές και μη αυτόματες προσεγγίσεις, όπως το Six Sigma και το Lean Manufacturing. Ακόμη και ορισμένες ποσοτικές μεθοδολογίες, όπως η Θεωρία των Περιορισμών και ο Στατιστικός Έλεγχος Διαδικασιών, έχουν περιορισμούς στην ενσωμάτωση και την ανάλυση δεδομένων από μια ποικιλία πηγών δεδομένων στο δυναμικό και πολύπλοκο περιβάλλον παραγωγής.

Χρήση του προϊόντος από τον πελάτη:

Η τελική φάση της διαδικασίας ανάπτυξης προϊόντος εξετάζει τη χρήση του προϊόντος από τους πελάτες. Δεδομένου ότι το προϊόν βρίσκεται στα χέρια του πελάτη σε αυτή τη φάση, η αντίληψη της ποιότητας μπορεί να βελτιωθεί μέσω επισκευής ή αντικατάστασης του προϊόντος υπό εγγύηση και άλλων υπηρεσιών πέρα από την εγγύηση. Υπάρχουν εφαρμογές εξόρυξης δεδομένων που περιλαμβάνουν βελτίωση ποιότητας και έλεγχο στις υπηρεσίες μετά την

πώληση, οι οποίες συνήθως χρησιμοποιούν προσεγγίσεις εξόρυξης κειμένου λόγω της φύσης κειμένου των σχετικών δεδομένων. (Gülser et. al, 2011)

2.2.2 Εργασίες Ποιότητας

Καθορίζεται, ακολούθως, ένα σύνολο ποιοτικών εργασιών που διέπει τη διαχείριση ποιότητας στη Βιομηχανία 4.0. (Gülser et. al, 2011) Πιο συγκεκριμένα:

Η Περιγραφή (ή χαρακτηρισμός) ποιότητας προϊόντος και διαδικασίας:

Η ποιότητα των προϊόντων ή των διαδικασιών μπορεί να οριστεί ή να χαρακτηριστεί εκτελώντας τις ακόλουθες εργασίες:

- Προσδιορισμός ιδιοτήτων/μεταβλητών, που επηρεάζουν σημαντικά την ποιότητα.
- Κατάταξη των χαρακτηριστικών/μεταβλητών με βάση τη σημασία τους.
- Προσδιορισμός του τρόπου με τον οποίο τα προϊόντα χαμηλής, μέσης και υψηλής απόδοσης ομαδοποιούνται φυσικά σε δεδομένα και εύρεση του πιο πιθανού αιτιολογικού(ών) παράγοντα(ών) που κάνει διάκριση μεταξύ χαμηλής και υψηλής απόδοσης προϊόντων.

Αυτές οι δραστηριότητες, συνήθως, εκτελούνται στα πρώτα στάδια του DMAIC και του DMADV για ανάλυση ποιότητας και σχεδιασμό προϊόντος/διαδικασίας, αυτά του Ορισμού-Μέτρησης-Ανάλυσης.

Πρόβλεψη ποιότητας:

Όταν η ποιοτική έξοδος είναι μια μεταβλητή πραγματικής αξίας, αναπτύσσονται μοντέλα που συσχετίζουν τα χαρακτηριστικά ποιότητας εισόδου με την έξοδο και χρησιμοποιούν τέτοια μοντέλα για να προβλέψουν ποια θα είναι η προσδοκώμενη χαρακτηριστική τιμή ποιότητας για ένα δεδομένο σύνολο τιμών παραμέτρων εισόδου. Αυτή η εργασία ποιότητας εκτελείται συνήθως στο στάδιο της ανάλυσης και των δύο κύκλων DMAIC και DMADV που ασχολούνται με τα πραγματικά αποτιμημένα χαρακτηριστικά ποιότητας εξόδου. Τα μοντέλα πρόβλεψης που αναπτύσσονται σε αυτό το στάδιο μπορούν να χρησιμοποιηθούν αργότερα στο Βελτίωση ή Σχεδίαση (Βελτιστοποίηση) – Επαλήθευση σταδίων ή απευθείας στο στάδιο Ελέγχου.

Ταξινόμηση ποιότητας:

Ταξινόμηση ενός ποιοτικού χαρακτηριστικού ενδιαφέροντος για ονομαστικές, δυαδικές ή τακτικές εξόδους (όπως ελαττώματα). Για ένα δεδομένο σύνολο παραμέτρων εισόδου, αυτή η εργασία αφορά στην πρόβλεψη της κλάσης του ποιοτικού αποτελέσματος.

Αντίστοιχα με την πρόβλεψη της ποιότητας, αυτή η εργασία ποιότητας εντοπίζεται κυρίως στο στάδιο Ανάλυσης και των δύο κύκλων DMAIC και DMADV.

Βελτιστοποίηση παραμέτρων:

Με βάση τα μαθησιακά χαρακτηριστικά των περιπτώσεων που αποδίδουν υψηλή ποιότητα, αυτή η εργασία στοχεύει στην εύρεση βέλτιστων επιπέδων παραμέτρων διεργασίας/προϊόντος που αποδίδουν σταθερά την απόδοση ποιότητας στόχου.

Η συγκεκριμένη η εργασία ποιότητας εκτελείται στα στάδια Βελτίωσης ή Ανάπτυξης και Επαλήθευσης των δύο κύκλων DMAIC και DMADV.

2.3 Αναλυτική Δεδομένων στη Διαχείριση Ποιότητας

Τα μεγάλα δεδομένα στην 4η βιομηχανική επανάσταση προσφέρουν στη μηχανική ποιότητας υψηλές δυνατότητες για σχεδιασμό, εκμάθηση και μοντελοποίηση συστημάτων και προϊόντων.

Οι κύριες προκλήσεις οφείλονται:

- (i) στην ανάγκη μετάβασης από την εξέταση του μέσου όρου, στην ανάλυση της διασποράς και της συσχέτισης,
- (ii) στην εξέταση στατικών συστημάτων, στην εξέταση δυναμικών και μη σταθερών συστημάτων,
- (iii) για μετάβαση από τα δεδομένα αισθητήρων σε προφίλ υψηλότερης τάξης και
- (iv) μετάβαση από την παρακολούθηση, στη διαγνωστική, προγνωστική και προδιαγραφική ανάλυση (Reis and Kenett 2018).

Η Ποιότητα Πληροφοριών είναι ένα ουσιαστικό πλαίσιο για τη Μηχανική Ποιότητας που βασίζεται σε δεδομένα. Αναφέρεται σε ένα ολοκληρωμένο πλαίσιο που έχει σχεδιαστεί για να σχεδιάζει και να αξιολογεί το επίπεδο της ποιότητας των πληροφοριών που παρέχεται από εργαλεία και μεθόδους ανάλυσης σε μια εταιρεία ή ένα έργο. Η Ποιότητα Πληροφοριών ορίζεται στους Kenett και Shmueli (2016) ως η δυνατότητα ενός συνόλου δεδομένων να επιτύχει έναν συγκεκριμένο (επιστημονικό ή πρακτικό) στόχο, χρησιμοποιώντας μια δεδομένη μέθοδο εμπειρικής ανάλυσης. Η Ποιότητα Πληροφοριών διαφέρει από την ποιότητα των δεδομένων και την ποιότητα ανάλυσης, αλλά εξαρτάται από αυτά τα στοιχεία και από τη μεταξύ τους σχέση. Η επισημοποίηση της έννοιας της ποιότητας πληροφοριών αυξάνει την

αξία της επιστήμης δεδομένων, της στατιστικής ανάλυσης και της μηχανικής ποιότητας που βασίζεται σε δεδομένα, τόσο μεθοδολογικά όσο και πρακτικά. (Zonnenshain & Kenett, 2020)

Η βελτίωση της ποιότητας των βιομηχανικών προϊόντων και διαδικασιών απαιτεί συλλογή και αναλύσεις δεδομένων για την επίλυση προβλημάτων παραγωγής που σχετίζονται με την ποιότητα. Ενώ τα προγράμματα για τη βελτίωση της ποιότητας όπως το six sigma (6σ), το design for six sigma (DFSS) και το kaizen συνεχίζουν να ενθαρρύνουν τη συλλογή δεδομένων για να αντιμετωπίσουν προβλήματα ποιότητας, με την πρόοδο στον αυτοματισμό και τα συστήματα υπολογιστών τα δεδομένα από τις διαδικασίες παραγωγής γίνονται όλο και περισσότερο διαθέσιμα. Παρόλο που τα παραδοσιακά εργαλεία ανάλυσης δεδομένων έχουν χρησιμοποιηθεί με επιτυχία για τη βελτίωση της ποιότητας των προϊόντων και των διαδικασιών, τώρα υπάρχουν καλύτερα εργαλεία για την εξόρυξη τεράστιων συνόλων δεδομένων που συλλέγονται μέσω ηλεκτρονικών συστημάτων στη βιομηχανία.

Οι Fayyad, Piatetsky-Shapiro και Smyth (1996) ορίζουν την ανακάλυψη γνώσης ως «τη μη τετριμμένη διαδικασία προσδιορισμού έγκυρων, καινοτόμων, δυνητικά χρήσιμων και τελικά κατανοητών προτύπων στα δεδομένα». Αποτελείται από τα ακόλουθα κύρια βήματα:

- (i) προετοιμασία δεδομένων
- (ii) προεπεξεργασία δεδομένων
- (iii) εξόρυξη δεδομένων
- (iv) αξιολόγηση και ερμηνεία
- (v) υλοποίηση

Σημειώνεται ότι η εξόρυξη δεδομένων αποτελείται από την εφαρμογή αλγορίθμων ανάλυσης δεδομένων και ανακάλυψης. Οι εργασίες εξόρυξης δεδομένων μπορούν να ταξινομηθούν σε δύο ομάδες (Han & Kamber, 2006): περιγραφικές και προγνωστικές. Αυτές οι εργασίες μπορούν να επιτευχθούν χρησιμοποιώντας διάφορες μεθόδους που βασίζονται σε συναρτήσεις εξόρυξης δεδομένων, που χρησιμοποιούνται για τον καθορισμό των τύπων των προτύπων που θα εξορυχθούν. Αυτές οι λειτουργίες περιλαμβάνουν σύνοψη (χαρακτηρισμός), ομαδοποίηση, συσχέτιση, ταξινόμηση, πρόβλεψη και ούτω καθεξής. Παρακάτω, περιγράφονται συνοπτικά τα βήματα της διαδικασίας ανακάλυψης γνώσης στα πλαίσια της Διαχείρισης Ποιότητας στη Βιομηχανία 4.0. (Gülser et. al, 2011)

2.3.1 Προετοιμασία δεδομένων:

Οι λειτουργίες εξόρυξης δεδομένων χρησιμοποιούν κυρίως διαθέσιμες πηγές δεδομένων, όπως αποθήκες δεδομένων, μάρκες, βάσεις δεδομένων ή αρχεία για τη συλλογή δεδομένων (Pyle, 1999). Στις εφαρμογές, οι πηγές δεδομένων αρχικά εντοπίζονται, έχουν πρόσβαση και ενσωματώνονται. Στη συνέχεια, τα επιλεγμένα δεδομένα τοποθετούνται σε μορφή πίνακα στην οποία οι παρουσίες και οι μεταβλητές λαμβάνουν χώρα σε γραμμές και στήλες, αντίστοιχα (Giudici, 2003). Εάν το σύνολο δεδομένων που έχει δημιουργηθεί είναι πολύ μεγάλο, μπορεί να ληφθεί ένα αντιπροσωπευτικό μειωμένο σύνολο δεδομένων με δειγματοληψία. Σε ορισμένες περιπτώσεις, όλα τα δεδομένα μπορεί να μην είναι άμεσα διαθέσιμα για εξόρυξη, όπου η διαδικασία συλλογής δεδομένων μπορεί να μας βοηθήσει «να ορίσουμε χαρακτηριστικά που είναι τα πιο κατάλληλα για την εξόρυξη δεδομένων» (Kusiak, 2006). Στη συνέχεια, μπορεί κανείς να συλλέξει τα απαραίτητα δεδομένα καταγράφοντας τις τιμές των χαρακτηριστικών απευθείας από πραγματικούς χρόνους ή πειραματικές παρατηρήσεις ή έμμεσα από αποτελέσματα προσομοίωσης.

2.3.2 Προεπεξεργασία δεδομένων:

Τα δεδομένα που προκύπτουν από πραγματικές συνθήκες είναι «βρώμικα», ελλιπή και ασυνεπή. Ενδέχεται επίσης να προκύψουν επικαλύψεις λόγω ενσωμάτωσης δεδομένων από διάφορες πηγές. Ο κύριος σκοπός αυτού του βήματος είναι να χειριστεί τέτοιου είδους προβλήματα για τη βελτίωση της ποιότητας των δεδομένων. Επιπλέον, ο μετασχηματισμός και η μείωση δεδομένων μπορεί να συμβάλει στη βελτίωση της ακρίβειας και της αποτελεσματικότητας των λειτουργιών εξόρυξης δεδομένων. Οι βασικές τεχνικές προεπεξεργασίας δεδομένων είναι οι εξής (Giudici, 2003; Pyle, 1999; Witten, 2005):

- Ο καθαρισμός δεδομένων, ο οποίος περιλαμβάνει τεχνικές συμπλήρωσης τιμών που λείπουν, εξομάλυνσης του θορύβου, χειρισμού ακραίων στοιχείων, ανίχνευσης και αφαίρεσης περιττών δεδομένων.
- Ο μετασχηματισμός δεδομένων, ο οποίος τοποθετεί τα δεδομένα σε κατάλληλες μορφές για εξόρυξη όταν είναι απαραίτητο.
- Η μείωση δεδομένων, η οποία εφαρμόζεται για τη μείωση του συνόλου δεδομένων προς εξόρυξη. Ενώ η τεχνική «μείωσης διαστάσεων» εξαλείφει τα περιττά χαρακτηριστικά, οι τεχνικές «συμπίεσης δεδομένων» και «μείωσης πληθώρας» παρέχουν άλλες μορφές μειωμένων αναπαραστάσεων δεδομένων.

- Η διακριτοποίηση, η οποία αποτελεί μια μορφή μείωσης δεδομένων, μειώνοντας τον αριθμό των επιπέδων ενός χαρακτηριστικού συλλέγοντας και αντικαθιστώντας έννοιες χαμηλού επιπέδου με έννοιες υψηλού επιπέδου.

2.3.3 Εξόρυξη δεδομένων:

Οι μέθοδοι εξόρυξης δεδομένων μπορούν να κατηγοριοποιηθούν με βάση διάφορα κριτήρια. Η παρακάτω ανάλυση επικεντρώνεται στην ταξινόμηση σύμφωνα με τους «τύπους εξόρυξης γνώσης» για την επίτευξη περιγραφικών/προγνωστικών εργασιών εξόρυξης δεδομένων (Dunham, 2003).

2.3.3.1 Περιγραφική Εξόρυξη δεδομένων:

Η περιγραφική εξόρυξη δεδομένων περιλαμβάνει την εξερεύνηση προτύπων και σχέσεων που μπορεί να υπάρχουν στα δεδομένα. Οι βασικές περιγραφικές συναρτήσεις είναι η Σύνοψη (Summarization), η Ομαδοποίηση (Clustering), η δημιουργία κανόνων συσχέτισης και η ανακάλυψη ακολουθίας (Association) (Dasu & Johnson, 2003; Giudici, 2003). Αυτά χρησιμοποιούνται επίσης για την εξερεύνηση δεδομένων πριν από την εφαρμογή μιας συνάρτησης εξόρυξης δεδομένων ταξινόμησης/πρόβλεψης ή/και βελτιστοποίησης.

- Η Σύνοψη (Summarization) είναι η παρουσίαση των γενικών χαρακτηριστικών ενός συνόλου δεδομένων. Οι βασικές προσεγγίσεις είναι το OLAP και η επαγωγή προσανατολισμένη στις ιδιότητες. Υπάρχουν πολλές διαθέσιμες στατιστικές μέθοδοι για τη σύνοψη δεδομένων (Giudici, 2003). Τα περιγραφικά στατιστικά στοιχεία και οι γραφικές οθόνες μπορούν να περιγράψουν αποτελεσματικά δεδομένα μίας μεταβλητής. Για δεδομένα δύο μεταβλητών, μπορούν να χρησιμοποιηθούν πρόσθετες μέθοδοι (π.χ. ανάλυση συσχέτισης, διαγράμματα διασποράς) για τον προσδιορισμό της σχέσης που μπορεί να υπάρχει μεταξύ των μεταβλητών. Για την περιγραφή δεδομένων περισσότερων μεταβλητών, ωστόσο, χρειάζονται μέτρα εξάρτησης και συσχέτισης καθώς και πολυδιάστατα γραφήματα όπως ο πίνακας διασποράς και οι καμπύλες Andrews (Martinez & Martinez, 2002).
- Η ομαδοποίηση είναι η διαδικασία ομαδοποίησης δεδομένων σε κατηγορίες παρόμοιων αντικειμένων. Η ομοιότητα μεταξύ των αντικειμένων συνήθως μετρείται με μέτρα απόστασης. Οι κύριες μέθοδοι ομαδοποίησης που βασίζονται στην απόσταση μπορούν να οργανωθούν σε δύο κατηγορίες (Han & Kamber, 2006). Οι μέθοδοι καταμερισμού ταξινομούν τα δεδομένα σε k μέρη με τέτοιο τρόπο ώστε οι παρατηρήσεις σε κάθε μέρος να σχετίζονται στενά μεταξύ τους. Οι ιεραρχικές μέθοδοι

ομαδοποιούν τα δεδομένα σε ένα δέντρο συστάδων χρησιμοποιώντας είτε προσεγγίσεις από κάτω προς τα πάνω (συγκεντρωτικές) είτε από πάνω προς τα κάτω (διαιρετικές). Επιπλέον, υπάρχουν και άλλες που ταξινομούνται ως μέθοδοι που βασίζονται στην πυκνότητα, σε πλέγμα και σε μοντέλα.

- Η δημιουργία κανόνων συσχέτισης και η ανακάλυψη ακολουθίας (Association) προσπαθεί να προσδιορίσει ομάδες αντικειμένων που εμφανίζονται μαζί. Υποθέτοντας ότι μια βάση δεδομένων αποτελείται από ένα σύνολο εγγραφών που περιέχει ένα σύνολο στοιχείων, οι περισσότεροι αλγόριθμοι ολοκληρώνουν την εργασία συσχέτισης σε δύο βήματα: βρίσκοντας συχνά σύνολα στοιχείων και, στη συνέχεια, δημιουργούν ενδιαφέροντες κανόνες εάν-τότε (Hand, Mannila, & Smyth, 2001).

2.3.3.2 Προγνωστική Εξόρυξη δεδομένων:

Η προγνωστική μοντελοποίηση μπορεί να επιτευχθεί εκτελώντας συναρτήσεις ταξινόμησης ή πρόβλεψης για την πρόβλεψη μελλοντικών τιμών κατηγορικών ή συνεχών δεδομένων τύπου, αντίστοιχα. Οι κύριες μέθοδοι ταξινόμησης/πρόβλεψης μπορούν να κατηγοριοποιηθούν στις ακόλουθες ομάδες (Dunham, 2003):

- Οι στατιστικές μέθοδοι (βασισμένες στο S) χρησιμοποιούν κλασικές τεχνικές που εξαρτώνται από τη στατιστική θεωρία και επομένως παρέχουν στατιστικά συμπεράσματα.
- Οι αλγόριθμοι που βασίζονται σε δέντρα αποφάσεων (decision trees) κατασκευάζουν δέντρα αποφάσεων που μοιάζουν με διάγραμμα ροής χρησιμοποιώντας μια αναδρομική προσέγγιση από πάνω προς τα κάτω. Οι αλγόριθμοι που βασίζονται σε δέντρα αποφάσεων δημιουργούν αυτόματα κανόνες με δομές τύπου «αν-τότε».
- Οι αλγόριθμοι που βασίζονται σε νευρωνικά δίκτυα αποτελούνται από ένα σύνολο συνδεδεμένων μονάδων εισόδου-εξόδου που η καθεμία έχει ένα βάρος, το οποίο ενημερώνεται από έναν αλγόριθμο εκμάθησης όπως η backpropagation (BP) που χρησιμοποιείται με την τεχνική βελτιστοποίησης gradient descent (GD) ή το συνάρτηση ακτινικής βάσης (RBF).

Επιπλέον, μπορούν να συνδυαστούν διαφορετικές μέθοδοι ταξινόμησης για τη βελτίωση της ακρίβειας του μοντέλου. Μετά από μια επιτυχημένη εφαρμογή μιας συνάρτησης ταξινόμησης/πρόβλεψης ή ως αυτόνομης συνάρτησης εξόρυξης δεδομένων, μπορεί να πραγματοποιηθεί βελτιστοποίηση για τον προσδιορισμό των ρυθμίσεων των παραγόντων (ή των παραμέτρων σχεδιασμού) που δίνουν τις επιθυμητές αποκρίσεις. (Gülser et. al, 2011)

Εφαρμογές αλγορίθμων πρόβλεψης σε προηγούμενες έρευνες:

Έχουν δημοσιευτεί αρκετές εφαρμογές της βιομηχανίας υπολογιστών και ηλεκτρονικών προϊόντων που εξυπηρετούν τον στόχο της πρόβλεψης της ποιότητας. Ενδεικτικά:

- Οι Chen, Lee, Deng και Liu (2007) χρησιμοποίησαν νευρωνικά δίκτυα σε μια διαδικασία εναπόθεσης χημικών ατμών για να καθορίσουν έναν προγνωστικό παράγοντα της ποιότητας του προϊόντος.
- Οι Li, Su και Chiang (2003a) χρησιμοποίησαν νευρωνικό δίκτυο για να καθορίσουν τις ρυθμίσεις παραμέτρων σε μια παραγωγή ενώσεων πυριτίου.
- Για έναν ισχυρό σχεδιασμό κυκλωμάτων ολοκλήρωσης πολύ μεγάλης κλίμακας (VLSI), ο Ilumoka (1998) εξέτασε επίσης νευρωνικό δίκτυο.
- Οι Yang, Tsai και Yeh (2005) χρησιμοποίησαν νευρωνικό δίκτυο για να προτείνουν ένα σύστημα ποιοτικού ελέγχου για τη διαδικασία εκτύπωσης με στένσιλ συγκόλλησης.
- Οι Shi, Schillings και Boyd (2004) χρησιμοποίησαν το νευρωνικό δίκτυο για τη μοντελοποίηση μη γραμμικών σχέσεων αιτίου-αποτελέσματος στην κατασκευή PCB.
- Οι Ip, Kwong και Wong (2003) χρησιμοποίησαν ασαφή παλινδρόμηση (FR) για να μοντελοποιήσουν τη χύτευση μεταφοράς για την ενθυλάκωση με μικροτσιπ ηλεκτρονικών πακέτων.
- Νευρωνικό δίκτυο και ο συλλογισμός βάσει περιπτώσεων (CBR) χρησιμοποιήθηκαν για την πρόβλεψη της διαδικασίας παραγωγής στην παραγωγή πλαστικών οπτικών ινών (Kim & Lee, 1997).
- Για τον προσδιορισμό της αιτία της αστοχίας και της βελτίωσης της απόδοσης, οι Mieno et al. (1999) αξιοποίησαν δέντρο αποφάσεων.
- Οι Skinner et al. (2002) εξέτασαν το MLR, το γενικευμένο γραμμικό μοντέλο (GLZ) και το δέντρο αποφάσεων για τον προσδιορισμό της ποιότητας ή της απόδοσης των πλακιδίων.
- Το δέντρο αποφάσεων χρησιμοποιήθηκε επίσης για την αποσαφήνιση της συσχέτισης μεταξύ της απόδοσης και των διαφόρων παραμετρικών δεδομένων πλακιδίων (Tsuda, Shirai, Takagi, & Take, 2000) και για την πρόβλεψη στην κατασκευή ολοκληρωμένων κυκλωμάτων (Maimon & Rokach, 2001).
- Οι Last και Kandel (2001) εξέτασαν το info-fuzzy network (IFN) για την πρόβλεψη της απόδοσης γραμμής και των χρόνων ροής.
- Οι Li, Huang και Wu (2006) εξέτασαν τον γενετικό προγραμματισμό (GP) για τη βελτίωση της απόδοσης του προϊόντος σε κατασκευή ημιαγωγών.

- Οι Sanchez-Marquez et al. (2020) πρότειναν μια μέθοδο για τη μελέτη των συστημάτων διαχείρισης ποιότητας και την πρόβλεψη βασικών δεικτών απόδοσης ισορροπημένων καρτών βαθμολογίας.
- Οι Carbery et al. (2018) χρησιμοποίησαν δίκτυα Bayes για τον εντοπισμό ελαττωμάτων.
- Ο Maurya (2016) πραγματοποίησε ανίχνευση ανωμαλιών και δυαδική ταξινόμηση χρησιμοποιώντας Gradient Boosting Machine και Bayesian βελτιστοποίηση προκειμένου να εντοπίσει σπάνια ελαττώματα.
- Οι Ge et al. (2021) συνέκρινε τους αλγόριθμους Μηχανής Διανυσμάτων Υποστήριξης και Τυχαίου Δάσους με τεχνικές κεντρικής εκμάθησης για προβλέψεις αποτυχίας προϊόντων.
- Ο Pavlyshenko (2016) συνδύασε τον ταξινομητή που βασίζεται σε δέντρο XGBoost, το γενικευμένο γραμμικό μοντέλο και μια Bayesian προσέγγιση για λογιστική παλινδρόμηση.
- Οι Gashi et al. (2021) πρότειναν μια προσέγγιση για τη βελτίωση της δοκιμής End of Line όταν λείπουν δεδομένα παρακολούθησης κατάστασης προκειμένου να προβλεφθούν προϊόντα χαμηλής ποιότητας ή προϊόντα με μεγάλη πιθανότητα αποτυχίας με την πάροδο του χρόνου. Για να γίνει αυτό, χρησιμοποίησαν ένα μοντέλο ταξινόμησης συγκρίνοντας τη γραμμική διακριτική ανάλυση με τη συρρίκνωση, το Τυχαίο δάσος και το lightGBM και χρησιμοποίησαν μια προβολή συστήματος πολλαπλών συστατικών για την εξήγηση των αποτελεσμάτων πρόβλεψης.

2.3.4 Αξιολόγηση και Ερμηνεία

Η διαδικασία ανακάλυψης γνώσης που περιγράφεται παραπάνω προσπαθεί να αποκαλύψει προηγουμένως άγνωστες δομές που μπορεί να βρίσκονται σε δεδομένα. Ανάλογα με τα σύνολα δεδομένων και τους ερευνητικούς στόχους, μπορεί κανείς να ξεκινήσει σε οποιοδήποτε βήμα και να συνεχίσει με τα άλλα, εφόσον υπάρχουν ερευνητικά ερωτήματα που πρέπει να απαντηθούν. Επιπλέον, μπορεί κανείς να δοκιμάσει διάφορες μεθόδους για την περιγραφή ή τη μοντελοποίηση της ποιότητας της διαδικασίας/προϊόντος. Επομένως, η αξιολόγηση της χρησιμότητας και της αξιοπιστίας και, στη συνέχεια, η ερμηνεία των πληροφοριών που ανακαλύφθηκαν στη μοντελοποίηση θα πρέπει να είναι το τελικό στάδιο σε μια διαδικασία εξόρυξης δεδομένων (Giudici, 2003). Η αξιολόγηση των μεθόδων εξόρυξης δεδομένων για την επίτευξη τελικής απόφασης απαιτεί σύγκριση των αποτελεσμάτων που λαμβάνονται από διάφορες μεθόδους εξόρυξης δεδομένων χρησιμοποιώντας διάφορα μέτρα, όπως η ακρίβεια, ο

χρόνος και οι απαιτήσεις πόρων. Είναι προφανές ότι για να ληφθούν αξιόπιστα αποτελέσματα, η εξαγόμενη γνώση θα πρέπει να αξιολογηθεί και να ερμηνευθεί σωστά (Dunham, 2003).

2.3.5 Υλοποίηση

Η διαδικασία ανακάλυψης γνώσης που περιγράφεται παραπάνω παρέχει εργαλεία για την καλύτερη κατανόηση των σχέσεων στα ποιοτικά δεδομένα. Το τελικό βήμα ανακάλυψης γνώσης περιλαμβάνει την εφαρμογή των αποτελεσμάτων που λαμβάνονται στις αποφάσεις που σχετίζονται με τη βελτίωση ποιότητας του κλάδου (Giudici, 2003). Αυτό συνήθως εκτελείται ανεπίσημα. Ο Kusiak (2006) προτείνει ένα πλαίσιο για πιο δομημένη και διαφανή λήψη αποφάσεων που βασίζεται σε δομές λήψης αποφάσεων που ονομάζονται πίνακες αποφάσεων, χάρτες αποφάσεων, άτλαντες και βιβλιοθήκη.

3

Ερευνητική Μεθοδολογία

Συνολικά, η μεθοδολογία που προτείνεται από την συλλογή και ανάλυση των διαθέσιμων δεδομένων μέχρι και τη σύγκριση των προβλέψεων με χρήση διαφορετικών αλγορίθμων μηχανικής μάθησης δίνεται γραφικά στο παρακάτω σχήμα και εξετάζεται αναλυτικά στις ακόλουθες ενότητες.



Figure 4 Ερευνητική Μεθοδολογία

3.1 Συλλογή Δεδομένων

Ένας οργανισμός έχει πολλά φυσικά και πληροφοριακά υποσυστήματα. Τα φυσικά υποσυστήματα περιλαμβάνουν αισθητήρες και σημεία ελέγχου. Τα πληροφοριακά υποσυστήματα περιλαμβάνουν τον προγραμματισμό πόρων της επιχείρησης (ERP) και τις Εφαρμογές και τα Προϊόντα Συστημάτων (SAP). Τα φυσικά και πληροφοριακά υποσυστήματα ενσωματώνονται σε διάφορα επίπεδα μέσα στο σύστημα παραγωγής για να δημιουργήσουν ένα ευέλικτο και αναδιαμορφώσιμο σύστημα παραγωγής. Αυτό θα βοηθήσει επίσης τις έξυπνες μηχανές να σχηματίσουν ένα αυτο-οργανωμένο σύστημα που μπορεί να αναδιαμορφωθεί δυναμικά ώστε να προσαρμόζεται σε διαφορετικούς τύπους προϊόντων και οι μαζικές πληροφορίες συλλέγονται και υποβάλλονται σε επεξεργασία για να γίνει διαφανής η διαδικασία παραγωγής (Wang et al., 2016). Για την εφαρμογή της ερευνητικής μεθοδολογίας που ακολουθεί προτείνεται η πλήρης αξιοποίηση των δεδομένων που διατίθενται από τα

υποσυστήματα ενός οργανισμού / ενός εργοστασίου παραγωγής. Γίνεται, ωστόσο, παρακάτω ειδική αναφορά σε επιμέρους στοιχεία του συνόλου δεδομένων που είναι απαραίτητο να συλλεχθούν ή να οριστούν στα πλαίσια της προτεινόμενης προσέγγισης.

Η παραγωγή έτοιμων προϊόντων για χρήση ή μεταπώληση με τη χρήση εργατικού δυναμικού και μηχανών, εργαλείων, χημική και βιολογική επεξεργασία, συναρμολόγηση ή μεταποίηση, είναι η ουσία της δευτερογενούς βιομηχανίας. Ο όρος μπορεί να αναφέρεται σε ένα εύρος από ανθρώπινες δραστηριότητες, από τον τομέα της βιοτεχνίας έως τα υπερσύγχρονα εργοστάσια, αλλά εφαρμόζεται συνήθως στον βιομηχανικό σχεδιασμό, όταν οι πρώτες ύλες που λαμβάνονται από την πρωτογενή βιομηχανία μετατρέπονται μαζικά σε έτοιμα προϊόντα. Αυτά τα έτοιμα προϊόντα με τη σειρά τους μπορούν να πωληθούν σε άλλους κατασκευαστές για την περαιτέρω παραγωγή πιο σύνθετων προϊόντων ή να διανεμηθούν μέσω της τριτογενούς βιομηχανίας στους τελικούς χρήστες και τους καταναλωτές. Αυτό γίνεται συνήθως μέσω χονδρεμπόρων, που προμηθεύουν τους λιανοπωλητές, οι οποίοι στη συνέχεια τα πωλούν στους πελάτες τους (EUR-Lex, 2019).

Η διαδικασία παραγωγής περιλαμβάνει τα στάδια με τα οποία οι πρώτες ύλες μετατρέπονται στο τελικό προϊόν, όπως είναι ο σχεδιασμός του προϊόντος και ο προσδιορισμός των προδιαγραφών για τα υλικά που θα χρησιμοποιηθούν. Στα πλαίσια μιας γραμμής παραγωγής, οι παραγωγοί χρησιμοποιούν συστήματα κωδικοποίησης για να παρακολουθούν τα ενδιάμεσα βήματα της επεξεργασίας ή της συναρμολόγησης που απαιτούνται για την ολοκλήρωση του προϊόντος και εν τέλει για την ιχνηλάτιση και τον έλεγχο του ίδιου του τελικού προϊόντος. Η χρήση αυτών των κωδικών μπορεί να κάνει τη διαφορά ως προς τον έλεγχο ποιότητας στη Βιομηχανία 4.0, καθώς δίνει πληροφορίες για τη σύσταση ενός προϊόντος από τα επιμέρους εξαρτήματά του, διευκολύνει την ομαδοποίηση αυτών των εξαρτημάτων αυτών και μπορεί να συσχετίσει ενδεχόμενα εμφανιζόμενα ελαττώματα με τα εξαρτήματα και τα τελικά προϊόντα. Τον πιο χαρακτηριστικό κωδικό αποτελεί η μονάδα τήρησης αποθεμάτων (ή SKU), ο οποίος είναι ένας αλφαριθμητικός συνδυασμός οκτώ, συνήθως, χαρακτήρων που αντιστοιχεί σε ένα προϊόν και καθιστά δυνατή την παρακολούθηση της κίνησης αυτού στη διαδικασία παραγωγής. Σε αντίθεση με τους παγκόσμιους κωδικούς προϊόντων (UPC), οι SKU δεν είναι καθολικές, δηλαδή κάθε παραγωγός διαθέτει το δικό του σύνολο SKU για τα προϊόντα παραγωγής του. Επισημαίνεται, επιπλέον, ότι ένας παραγωγός μπορεί να δημιουργήσει εσωτερικά SKU που δείχνουν τις λεπτομέρειες ενός προϊόντος, όπως χρώμα, μέγεθος, κατασκευαστή και επωνυμία. Οι SKU δεν πρέπει να συγχέονται με τους αριθμούς μοντέλων, αν και οι επιχειρήσεις μπορούν να ενσωματώσουν αριθμούς μοντέλων σε SKU.

Στα πλαίσια της παρούσας μελέτης, θα χρειαστεί να υπάρχει καθορισμένη τιμή Μονάδας Τήρησης Αποθεμάτων για κάθε τελικό προϊόν και κάθε επιμέρους εξάρτημά του, ώστε να είναι δυνατή η συσχέτιση ενός εξαρτήματος με το τελικό προϊόν, όπως και ενός ελαττώματος με ένα εξάρτημα και με το τελικό προϊόν. Τα εξάρτηματα θα χρειαστεί, επιπλέον, να ταξινομηθούν σε ευρύτερες κλάσεις εξαρτημάτων (πχ. Ηλεκτρολογικά εξαρτήματα, μηχανικά εξαρτήματα) και για την κάθε κλάση να οριστεί ένα μοναδικό αναγνωριστικό. Με αυτόν τον τρόπο θα είναι δυνατή η συσχέτιση ενός εξαρτήματος με μια κλάση, όπως και ενός ελαττώματος με μια κλάση, συμβάλλοντας στην εξαγωγή χρήσιμης πληροφορίας κατά το στάδιο ανάλυσης των δεδομένων. Τα ελαττώματα, με τη σειρά τους, θα πρέπει να οριστούν και να καταγραφούν με μοναδικά χαρακτηριστικά, ώστε να γίνεται η διακριτή αναφορά σε αυτά. Επίσης, κρίνεται χρήσιμη η ευρύτερη ταξινόμησή τους σε ομάδες ελαττωμάτων, καθώς αυτή η ομαδοποίηση μπορεί να αξιοποιηθεί σε εξαγωγή επιπλέον πληροφορίας κατά το στάδιο ανάλυσης των δεδομένων. Τέλος, χρειάζεται να οριστούν και να καταγραφούν οι πηγές εμφάνισης του κάθε ελαττωματικού, καθώς ενδέχεται να συνδέονται με τις αιτίες πρόκλησης ή επίλυσης του ελαττώματος κατά τα προηγούμενα στάδια της παραγωγικής διαδικασίας.

Συνοπτικά:

Σύμφωνα με τα παραπάνω, για να είναι δυνατή η εφαρμογή της μεθοδολογίας πρόβλεψης ελαττωμάτων και της αξιολόγησης του ρίσκου ποιότητας που αυτό εισάγει στη βιομηχανική διαδικασία, ενδείκνυται ο καθορισμός των παρακάτω δεδομένων:

- Ο ορισμός μοναδικού αναγνωριστικού για το τελικό προϊόν.
- Ο ορισμός μοναδικού αναγνωριστικού για το κάθε εξάρτημα που συμμετέχει στον σχεδιασμό των τελικών προϊόντων.
- Ο ορισμός μοναδικού αναγνωριστικού για ευρύτερες ομάδες / οικογένειες εξαρτημάτων.
- Η συσχέτιση του τελικού προϊόντος με τα επιμέρους εξαρτήματά του.
- Η ομαδοποίηση του κάθε εξαρτήματος σε ομάδα / οικογένεια εξαρτημάτων.
- Ο ορισμός μοναδικού αναγνωριστικού για την παραγγελία με την οποία σχετίζεται ένα παραγόμενο προϊόν.
- Ο ορισμός μοναδικού αναγνωριστικού για την κάθε πηγή εμφάνισης ελαττωμάτων.
- Ο ορισμός μοναδικού αναγνωριστικού για το κάθε ελάττωμα που μπορεί να εμφανιστεί.
- Ο ορισμός μοναδικού αναγνωριστικού για ευρύτερες ομάδες / οικογένειες ελαττωμάτων.
- Η ομαδοποίηση του κάθε ελαττώματος σε ομάδα / οικογένεια ελαττωμάτων.
- Ο ορισμός μοναδικού αναγνωριστικού για τις γραμμές παραγωγής προϊόντος.

Εφόσον γίνει ο καθορισμός των δεδομένων που κρίνονται απαραίτητα για τα επόμενα στάδια της μεθοδολογίας που προτείνουμε, χρειάζεται να γίνει η συλλογή δεδομένων για ελαττώματα που εντοπίστηκαν κατά τον έλεγχο ποιότητας προϊόντων στο τέλος της γραμμής παραγωγής και πριν τη διάθεσή τους στην αγορά. Αυτά τα δεδομένα, με τη σειρά τους, θα μπορούν να χρησιμοποιηθούν ως ιστορικά δεδομένα για την εκπαίδευση των μοντέλων μηχανικής μάθησης και να συμβάλλουν στην πρόβλεψη ελαττωμάτων μελλοντικών προϊόντων που παράγονται. Πιο συγκεκριμένα, είτε μέσω ανθρώπινης ανατροφοδότησης, είτε με τη χρήση αυτοματοποιημένων διαδικασιών και συστημάτων, θα πρέπει να δημιουργηθεί μια βάση δεδομένων, η οποία θα περιλαμβάνει:

- Το χρονικό στίγμα
- Τη γραμμή παραγωγής
- Τον αριθμό παραγγελίας
- Τη μονάδα τήρησης αποθεμάτων
- Το εξάρτημα
- Την ευρύτερη οικογένεια εξαρτήματος
- Την πηγή ελαττώματος
- Το ελάττωμα
- Την ευρύτερη κλάση στην οποία ανήκει το ελάττωμα

για τα οποία εντοπίστηκε κάποιο ελαττωματικό.

Με αυτά τα στοιχεία δημιουργείται, εν τέλει, ένα σύνολο ιστορικών δεδομένων, τα οποία, μετά την περαιτέρω ανάλυση και επεξεργασία τους, θα αξιοποιηθούν στην εκμάθηση μοντέλων πρόβλεψης, όπως αναφέρεται εκτενέστερα στα επόμενα κεφάλαια.

3.2 Προ-επεξεργασία Δεδομένων

Επί της αρχής, η προ-επεξεργασία δεδομένων είναι ένα προαπαιτούμενο βήμα προκειμένου να μετατραπούν τα ακατέργαστα δεδομένα σε μορφή ικανή να υποστεί περαιτέρω επεξεργασία από τους αλγόριθμους ανάλυσης δεδομένων. Ανάλογα με τη μορφή και τη δομή των δεδομένων καθώς και τους εφαρμοζόμενους αλγόριθμους, η πολυπλοκότητα της προ-επεξεργασίας δεδομένων ποικίλλει. Στα πλαίσια της μηχανικής μάθησης συγκεκριμένα, η ανάλυση και επεξεργασία των διαθέσιμων δεδομένων στοχεύει, μεταξύ άλλων, στη βέλτιστη επιλογή χαρακτηριστικών (feature selection). Είναι γνωστή και ως επιλογή μεταβλητής (variable selection), μείωση χαρακτηριστικών (feature reduction), επιλογή χαρακτηριστικών (attribute selection) ή επιλογή υποσυνόλου μεταβλητών (variable subset selection) και αποτελεί μια

πρακτική επιλογής ενός υποσυνόλου μεταβλητών που στοχεύουν στη δημιουργία ισχυρών μοντέλων εκπαίδευσης. Η επιλογή χαρακτηριστικών είναι ένα ιδιαίτερα σημαντικό βήμα στην ανάλυση δεδομένων, καθώς αρκετά συχνά ο όγκος των δεδομένων είναι πολύ μικρός αλλά ο αριθμός των χαρακτηριστικών που προέκυψαν κατά την ανάλυση είναι πολύ μεγάλος. Με την αφαίρεση μη σχετικών ή/και περιττών χαρακτηριστικών από τα δεδομένα, καθίσταται δυνατή η βελτίωση της απόδοσης των μοντέλων μάθησης, η ενίσχυση της γενίκευσης του μοντέλου και η επιτάχυνση της διαδικασίας εκμάθησης. Η επιλογή των χαρακτηριστικών βοηθά στην καλύτερη κατανόηση των δεδομένων, αποκαλύπτοντας τα πιο σημαντικά από αυτά και πώς σχετίζονται μεταξύ τους. Σύμφωνα με τη συλλογή δεδομένων που αναφέρθηκε νωρίτερα, τα υποψήφια χαρακτηριστικά που θα μπορούσαν να δοθούν ως είσοδοι στους αλγόριθμους μηχανικής μάθησης είναι τα παρακάτω:

- Το χρονικό στίγμα
- Η γραμμή παραγωγής
- Ο αριθμός παραγγελίας
- Η μονάδα τήρησης αποθεμάτων
- Το εξάρτημα
- Η ευρύτερη οικογένεια εξαρτήματος
- Η πηγή ελαττώματος

Σημειώνεται, επίσης, ότι το ελάττωμα θα είναι, σε αυτή την περίπτωση, η πρόβλεψη / έξοδος του αλγόριθμου βάσει των εισαχθέντων πληροφοριών και του παραγόμενου μοντέλου.

Προκειμένου να βελτιωθούν τα αποτελέσματα για τους τέσσερις αλγόριθμους που εξετάζουμε σε αυτή τη μελέτη, είναι απαραίτητο να γίνουν κάποια βήματα προεπεξεργασίας που θα καταστήσουν δυνατή τη μείωση της διάστασης δεδομένων χωρίς να επηρεαστεί η εργασία ταξινόμησης. Το πρώτο βήμα είναι η μετατροπή όλων των στιγμιοτύπων του συνόλου δεδομένων σε ετικέτες. Η τεχνική μετατροπής δεδομένων σε ετικέτες χρησιμοποιείται για να κάνει τα αντικείμενα αναγνωρίσιμα και κατανοητά για μοντέλα μηχανικής μάθησης. Αυτά τα δεδομένα με ετικέτα χρησιμοποιούνται στη συνέχεια για την εκπαίδευση μοντέλων μηχανικής μάθησης ώστε να βρίσκουν νόημα σε νέα, σχετικά παρόμοια δεδομένα. Σε όλη αυτή τη διαδικασία, οι επαγγελματίες μηχανικής μάθησης προσπαθούν τόσο για την ποιότητα όσο και για την ποσότητα. Ο ακριβέστερος καθορισμός σε συνδυασμό με μεγαλύτερη ποσότητα επισημασμένων δεδομένων δημιουργεί πιο χρήσιμα μοντέλα βαθιάς εκμάθησης, καθώς το μοντέλο μηχανικής εκμάθησης που προκύπτει βασίζεται στις αποφάσεις του σε όλα τα δεδομένα με ετικέτα.

Τελικό βήμα αποτελεί ο διαχωρισμός των δεδομένων σε εκπαίδευση και δοκιμή. Το ποσοστό που χρησιμοποιείται για προπόνηση είναι συνήθως το 65-85% του αρχικού συνόλου και το υπόλοιπο ποσοστό των δεδομένων χρησιμοποιείται για δοκιμή. Το πρώτο σύνολο είναι απαραίτητο για την εύρεση μοτίβων στα δεδομένα εκπαίδευσης έτσι ώστε οι παράμετροι εισόδου να αντιστοιχούν στον στόχο. Το αποτέλεσμα της εκπαιδευτικής διαδικασίας είναι ένα μοντέλο μηχανικής μάθησης το οποίο στη συνέχεια χρησιμοποιείται για την παραγωγή προβλέψεων. Με την αξιοποίηση του δεύτερου συνόλου καθίσταται, αργότερα, δυνατή η αξιολόγηση του εφαρμοζόμενου μοντέλου.

3.3 Πρόβλεψη Ελαττωματικών με Μηχανική Μάθηση

Με την εμφάνιση της Βιομηχανίας 4.0 και του κινήματος Μεγάλων Δεδομένων, η βιομηχανία παρουσιάζεται με μοναδικές ευκαιρίες για να μεταφέρει την απόδοση σε ένα νέο και υψηλότερο επίπεδο με την πολύτιμη συμβολή της Τεχνητής Νοημοσύνης. Η μηχανική μάθηση (machine learning) αποτελεί έναν από τους κυριότερους τομείς έρευνας της τεχνητής νοημοσύνης. Στόχος της είναι η δημιουργία μοντέλων που εκπαιδεύονται από ιστορικά και εμπειρικά δεδομένα, ώστε να εκτελούν την εργασία για την οποία προορίζονται αποτελεσματικότερα. Η διαδικασία εκμάθησης μπορεί να αναλυθεί στα παρακάτω στάδια:

- Συλλογή εμπειρικών ή/και ιστορικών δεδομένων κατάλληλων να αξιοποιηθούν για εκπαίδευση του μοντέλου.
- Επεξεργασία των δεδομένων, ώστε να βρεθούν πιθανές γενικεύσεις ή εξειδικεύσεις (διαδικασία μάθησης).
- Χρησιμοποίηση των αποτελεσμάτων της επεξεργασίας για την εκτέλεση της εργασίας στόχου.

Η μηχανική μάθηση μπορεί να διακριθεί στην επιβλεπόμενη μάθηση (supervised learning) και στη μάθηση χωρίς επίβλεψη (unsupervised learning). Η επιβλεπόμενη μάθηση είναι μία κατηγορία μηχανικής μάθησης, στόχος της οποίας είναι ο χαρακτηρισμός δεδομένων με βάση κάποια δεδομένα εκπαίδευσης. Τα δεδομένα εκπαίδευσης αποτελούνται από ένα σύνολο παραδειγμάτων τα οποία χρησιμοποιούνται για εκπαίδευση μοντέλων. Στην επιβλεπόμενη μάθηση, κάθε παράδειγμα αποτελείται από ένα σύνολο εισόδου (συνήθως ένα διάνυσμα από χαρακτηριστικά) και μια επιθυμητή τιμή εξόδου. Οι αλγόριθμοι επιβλεπόμενης μάθησης αναλύουν τα δεδομένα εκπαίδευσης και παράγουν ένα μοντέλο το οποίο μπορεί να χρησιμοποιηθεί για να χαρακτηρίσει νέα παραδείγματα. Το βέλτιστο σενάριο επιτρέπει στον αλγόριθμο να καθορίσει σωστά την ετικέτα της κατηγορίας για άγνωστα μέχρι τώρα

παραδείγματα. Για να επιτευχθεί αυτό, απαιτείται ο αλγόριθμος μάθησης να γενικεύει από τα δεδομένα εκπαίδευσης σε αθέατες καταστάσεις με ένα "λογικό" τρόπο (Mohri, 2012).

Τυπικό παράδειγμα επιβλεπόμενης μάθησης αποτελούν τα προβλήματα ταξινόμησης (classification), όπως αυτά που χρησιμοποιούνται στην παρούσα εργασία. Αφορούν στην εκμάθηση μιας συνάρτησης στόχου ικανής να αντιστοιχεί άγνωστα αντικείμενα σε προκαθορισμένο σύνολο κατηγοριών. Στόχος της διαδικασίας μάθησης είναι η δημιουργία ενός μοντέλου πρόβλεψης της μη παρατηρούμενης ιδιότητας που είναι το διακριτικό της κατηγορίας ενός αγνώστου αντικειμένου, στην προκειμένη περίπτωση αυτής του ελαττώματος, βάσει των τιμών των χαρακτηριστικών που το προσδιορίζουν. Για να είναι δυνατή η προσέγγιση του συγκεκριμένου προβλήματος είναι απαραίτητη η συγκέντρωση αντιπροσωπευτικών παραδειγμάτων για κάθε κατηγορία, τα οποία έχουν ταξινομηθεί από τον άνθρωπο-εκπαιδευτή. Στην συνέχεια, αυτά τα παραδείγματα-στιγμιότυπα, τα οποία αποτελούν το σύνολο δεδομένων της εκπαίδευσης, δίνονται σε έναν αλγόριθμο μηχανικής μάθησης, ο οποίος, εφόσον παρατηρήσει τα χαρακτηριστικά τους, παράγει ως έξοδο ένα εκτιμώμενο μοντέλο, ικανό να αντιστοιχίζει άγνωστα στιγμιότυπα στις προδιαγεγραμμένες κατηγορίες. Το τελικό αποτέλεσμα της εκπαίδευσης ονομάζεται ταξινομητής (classifier).

Δέντρα Αποφάσεων (Decision Trees):

Το δέντρο αποφάσεων είναι μέλος του εποπτευόμενου αλγόριθμου μάθησης που χρησιμοποιείται τόσο για προβλήματα ταξινόμησης όσο και για προβλήματα παλινδρόμησης. Πιο συχνά, το δέντρο αποφάσεων χρησιμοποιείται για προβλήματα ταξινόμησης. Ένα δέντρο απόφασης παλινδρόμησης είναι ένα δέντρο που δημιουργείται σε δεδομένα των οποίων η μεταβλητή στόχος είναι συνεχής, ενώ ένα δέντρο αποφάσεων ταξινόμησης υλοποιείται σε δεδομένα των οποίων η μεταβλητή στόχος είναι κατηγορική. Ο αλγόριθμος δέντρου αποκαλείται έτσι λόγω της δομής του που μοιάζει με δέντρο στην παρουσίαση των αποφάσεων και των διαδικασιών λήψης αποφάσεων. Μας επιτρέπει να κάνουμε τη βέλτιστη επιλογή σε μια περίπλοκη κατάσταση με διάφορες εναλλακτικές λύσεις και να ακολουθήσουμε την καλύτερη στρατηγική για να επιτύχουμε το βέλτιστο αποτέλεσμα.

Τα δέντρα αποφάσεων αποτελούνται από κόμβους και μονοπάτια. Οι κόμβοι στο δέντρο αποφάσεων είναι δύο ευρειών κατηγοριών: κόμβος απόφασης και κόμβος φύλλου. Ο κόμβος απόφασης είναι όπου περνάμε έναν κανόνα και χωρίζουμε τα δεδομένα με βάση έναν τέτοιο κανόνα. Όλοι οι κόμβοι απόφασης έχουν μονοπάτια που βγαίνουν από αυτούς. Από την άλλη πλευρά, ένας κόμβος φύλλου είναι ένας κόμβος που εμφανίζεται μόνο στο τερματικό ενός δέντρου αποφάσεων. Οι κόμβοι φύλλων δεν έχουν μονοπάτια που βγαίνουν από αυτούς και

επομένως δεν έχουν περαιτέρω διαχωρισμούς. Σε αυτόν τον κόμβο, παραθέτουμε τις τιμές της μεταβλητής στόχου. Ο κόμβος φύλλων είναι επίσης γνωστός ως τερματικός κόμβος. Ο κόμβος στην κορυφή ενός δέντρου αποφάσεων ονομάζεται κόμβος ρίζας. Η κύρια ιδιότητά του είναι ότι δεν έχει εισερχόμενες διαδρομές. Επομένως, η πλειοψηφία των κόμβων απόφασης εμφανίζεται μεταξύ του κόμβου ρίζας και του κόμβου φύλλου. Αυτοί οι κόμβοι ονομάζονται εσωτερικοί κόμβοι ενός δέντρου αποφάσεων. Ένας εσωτερικός κόμβος έχει εισερχόμενο και εξερχόμενο μονοπάτι(α) και μαζί με τους τερματικούς κόμβους σχηματίζουν τον κλάδο ενός δέντρου αποφάσεων.

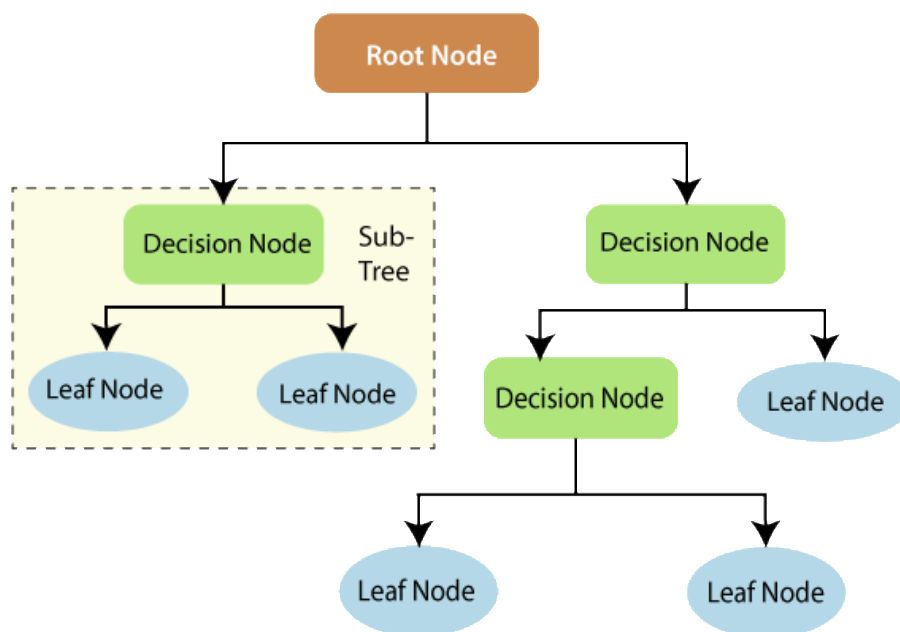


Figure 5 Δέντρα Αποφάσεων (πηγή: Tutorials and Example)

Τυχαία Δάση (Random Forests):

Ένα τυχαίο δάσος είναι ένας εποπτευόμενος αλγόριθμος μηχανικής μάθησης που κατασκευάζεται από αλγόριθμους δέντρων αποφάσεων. Χρησιμοποιείται για την επίλυση προβλημάτων παλινδρόμησης και ταξινόμησης εφαρμόζοντας την εκμάθηση συνόλου, η οποία είναι μια τεχνική που συνδυάζει πολλούς ταξινομητές για να δώσει λύσεις σε πολύπλοκα προβλήματα. Το «δάσος» που δημιουργείται από τον αλγόριθμο τυχαίων δασών εκπαιδεύεται μέσω bagging ή bootstrap. Το Bagging είναι ένας μετα-αλγόριθμος συνόλου που βελτιώνει την ακρίβεια των αλγορίθμων μηχανικής εκμάθησης. Ο αλγόριθμος αυτός καθορίζει το αποτέλεσμα βασίζοντας στις προβλέψεις των δέντρων απόφασης. Προβλέπει λαμβάνοντας τον μέσο όρο ή τον μέσο όρο της παραγωγής από διάφορα δέντρα. Έτσι, η αύξηση του αριθμού των δέντρων αυξάνει την ακρίβεια του αποτελέσματος. Ως εκ τούτου, ένα τυχαίο δάσος εξαλείφει τους περιορισμούς ενός αλγορίθμου δέντρου αποφάσεων, καθώς μειώνει την υπερπροσαρμογή των συνόλων δεδομένων και αυξάνει την ακρίβεια.

Η ταξινόμηση σε τυχαία δάση χρησιμοποιεί μια μεθοδολογία συνόλου για την επίτευξη του αποτελέσματος. Τα δεδομένα εκπαίδευσης τροφοδοτούνται για την εκπαίδευση διαφόρων δέντρων αποφάσεων. Αυτό το σύνολο δεδομένων αποτελείται από παρατηρήσεις και χαρακτηριστικά που θα επιλεγούν τυχαία κατά τη διαίρεση των κόμβων. Ένα σύστημα τυχαίων δασών βασίζεται σε διάφορα δέντρα απόφασης. Κάθε δέντρο απόφασης αποτελείται από κόμβους απόφασης, κόμβους φύλλων και έναν κόμβο ρίζας. Ο κόμβος φύλλων κάθε δέντρου είναι η τελική έξοδος που παράγεται από το συγκεκριμένο δέντρο απόφασης. Η επιλογή της τελικής παραγωγής ακολουθεί το σύστημα της πλειοψηφίας. Σε αυτήν την περίπτωση, η έξοδος που επιλέγεται από την πλειοψηφία των δέντρων απόφασης γίνεται η τελική έξοδος του συστήματος τροπικών δασών. Το παρακάτω διάγραμμα δείχνει έναν απλό τυχαίο ταξινομητή δασών:

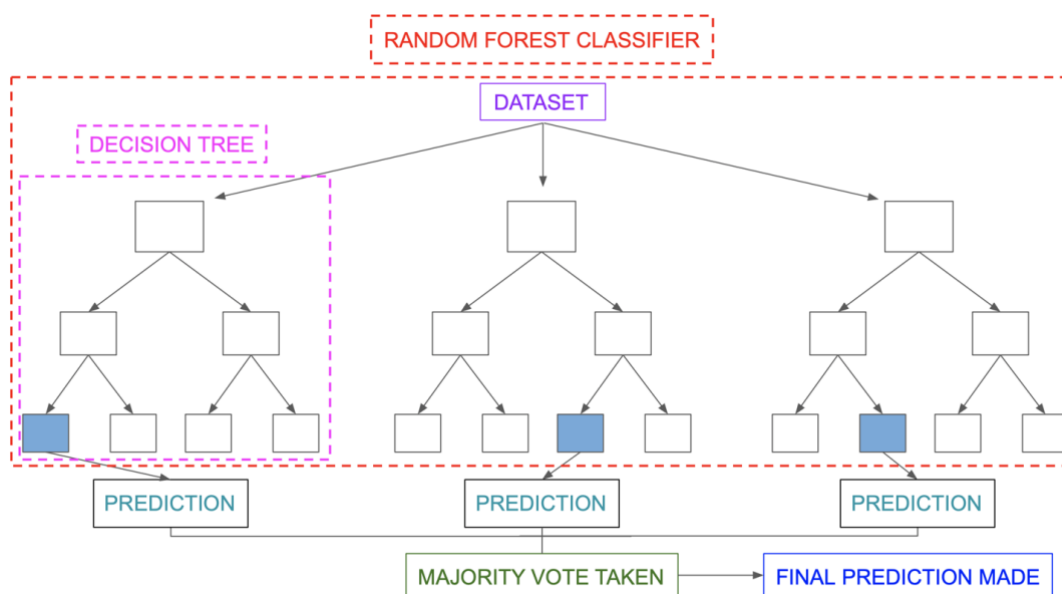


Figure 6 Τυχαία Δάση (πηγή: Medium)

K-πλησιέστεροι γείτονες (K-Nearest Neighbors):

Ο αλγόριθμος K-πλησιέστερων γειτόνων αποτελεί έναν από τους βασικότερους αλλά ουσιαστικούς αλγόριθμους ταξινόμησης στη μηχανική μάθηση, ενώ ανήκει στον τομέα της εποπτευόμενης μάθησης. Είναι ένας ευρέως αναλώσιμος αλγόριθμος σε σενάρια πραγματικής ζωής, καθώς είναι μη παραμετρικός, δηλαδή δεν κάνει υποκείμενες υποθέσεις σχετικά με τη διανομή των δεδομένων, σε αντίθεση με άλλους αλγόριθμους που υποθέτουν μια Γκαουσιανή

κατανομή των δεδομένων. Στην ταξινόμηση K -πλησιέστερων γειτόνων, η έξοδος είναι μια ιδιότητα μέλους κλάσης. Ένα αντικείμενο ταξινομείται με πλήθος ψηφοφοριών των γειτόνων του, με το αντικείμενο να εκχωρείται στην κατηγορία που είναι πιο κοινή μεταξύ των k πλησιέστερων γειτόνων του (k είναι ένας θετικός ακέραιος, συνήθως μικρός). Αν $k = 1$, τότε το αντικείμενο απλώς εκχωρείται στην κλάση αυτού του απλού πλησιέστερου γείτονα.

Ο αλγόριθμος υποθέτει ότι παρόμοια πράγματα υπάρχουν σε κοντινή απόσταση. Με άλλα λόγια, παρόμοια πράγματα είναι κοντά το ένα στο άλλο, με αποτέλεσμα η συνάρτηση να προσεγγίζεται μόνο τοπικά και όλοι οι υπολογισμοί να αναβάλλονται μέχρι την αξιολόγηση της συνάρτησης. Δεδομένου ότι αυτός ο αλγόριθμος βασίζεται στην απόσταση για ταξινόμηση, εάν τα χαρακτηριστικά αντιπροσωπεύουν διαφορετικές φυσικές μονάδες ή βρίσκονται σε πολύ διαφορετικές κλίμακες, τότε η κανονικοποίηση των δεδομένων εκπαίδευσης μπορεί να βελτιώσει δραματικά την ακρίβειά τους.

Μια χρήσιμη τεχνική στην εφαρμογή του αλγορίθμου αποτελεί η ανάθεση βαρών στις συνεισφορές των γειτόνων, έτσι ώστε οι πιο κοντινοί γείτονες να συνεισφέρουν περισσότερο στον μέσο όρο από τους πιο απομακρυσμένους. Οι γείτονες λαμβάνονται από ένα σύνολο αντικειμένων για τα οποία είναι γνωστή η κλάση. Αυτό μπορεί να θεωρηθεί ως το σύνολο εκπαίδευσης για τον αλγόριθμο, αν και δεν απαιτείται ρητό βήμα εκπαίδευσης.

Naïve Bayes:

Οι ταξινομητές Naïve Bayes είναι μια οικογένεια απλών «πιθανολογικών ταξινομητών» που βασίζονται στην εφαρμογή του θεωρήματος του Bayes με ισχυρές (αφελείς) υποθέσεις ανεξαρτησίας μεταξύ των χαρακτηριστικών. Είναι από τα απλούστερα μοντέλα δικτύων Bayes, αλλά σε συνδυασμό με την εκτίμηση της πυκνότητας του πυρήνα, μπορούν να επιτύχουν υψηλότερα επίπεδα ακρίβειας. Οι ταξινομητές Naïve Bayes είναι εξαιρετικά επεκτάσιμοι, απαιτώντας έναν αριθμό παραμέτρων γραμμικών ως προς τον αριθμό των μεταβλητών (χαρακτηριστικά/προγνωστικά) σε ένα μαθησιακό πρόβλημα. Η εκπαίδευση μέγιστης πιθανότητας μπορεί να γίνει με την αξιολόγηση μιας έκφρασης κλειστής μορφής, η οποία απαιτεί γραμμικό χρόνο, αντί με ακριβή επαναληπτική προσέγγιση όπως χρησιμοποιείται για πολλούς άλλους τύπους ταξινομητών.

Ο Naïve Bayes είναι μια απλή τεχνική για την κατασκευή ταξινομητών: μοντέλα που εκχωρούν ετικέτες κλάσεων σε στιγμιότυπα προβλημάτων, που αναπαρίστανται ως διανύσματα τιμών χαρακτηριστικών, όπου οι ετικέτες κλάσεων προέρχονται από κάποιο πεπερασμένο σύνολο.

Δεν υπάρχει ένας μόνο αλγόριθμος για την εκπαίδευση τέτοιων ταξινομητών, αλλά μια οικογένεια αλγορίθμων που βασίζεται σε μια κοινή αρχή: όλοι οι απλοί ταξινομητές Bayes υποθέτουν ότι η τιμή ενός συγκεκριμένου χαρακτηριστικού είναι ανεξάρτητη από την τιμή οποιουδήποτε άλλου χαρακτηριστικού, δεδομένης της μεταβλητής κλάσης.

Παρά τον αφελή σχεδιασμό τους και τις φαινομενικά υπεραπλουστευμένες υποθέσεις, οι αφελείς ταξινομητές Bayes έχουν λειτουργήσει αρκετά καλά σε πολλές περίπλοκες καταστάσεις του πραγματικού κόσμου. Ένα πλεονέκτημα του αφελούς Bayes είναι ότι απαιτεί μόνο μικρό αριθμό δεδομένων εκπαίδευσης για την εκτίμηση των παραμέτρων που είναι απαραίτητες για την ταξινόμηση. Επίσης, παρά το γεγονός ότι οι εκτεταμένες υποθέσεις ανεξαρτησίας είναι συχνά ανακριβείς, ο ταξινομητής Naïve Bayes έχει αρκετές ιδιότητες που τον καθιστούν ιδιαίτερα χρήσιμο στην πράξη. Συγκεκριμένα, η αποσύνδεση των κατανομών χαρακτηριστικών υπό όρους κλάσης σημαίνει ότι κάθε κατανομή μπορεί να εκτιμηθεί ανεξάρτητα ως μονοδιάστατη κατανομή. Αυτό βοηθά στην άμβλυνση προβλημάτων που προκύπτουν από την ανάγκη για σύνολα δεδομένων που κλιμακώνονται εκθετικά ανάλογα με τον αριθμό των χαρακτηριστικών. Ενώ ο Naïve Bayes συχνά αποτυγχάνει να παράγει μια καλή εκτίμηση για τις σωστές πιθανότητες κλάσης, αυτό μπορεί να μην είναι απαίτηση για πολλές εφαρμογές.

3.4 Αξιολόγηση Ρίσκου Εμφάνισης Ελαττώματος

Η διαχείριση κινδύνου είναι κεντρικός πυρήνας της στρατηγικής διαχείρισης κάθε οργανισμού. Είναι η διεργασία με την οποία οι οργανισμοί προσεγγίζουν μεθοδικά τους κινδύνους που σχετίζονται με τις δραστηριότητές τους, με σκοπό την επίτευξη κέρδους σε κάθε δραστηριότητα. Το επίκεντρο της καλής διαχείρισης κινδύνου είναι η αναγνώριση και ο χειρισμός αυτών των κινδύνων, η οποία εντέλει αυξάνει την πιθανότητα επιτυχίας των συνολικών στόχων του οργανισμού. Η αξιολόγηση των κινδύνων μπορεί να αποκαλύψει ήδη από τα αρχικά στάδια αν αξίζει ή όχι να ασχοληθεί ο οργανισμός με μια στρατηγική προσέγγιση, ενώ πολλά προβλήματα μπορούν να αντιμετωπιστούν εκ των προτέρων και να δώσουν την ευκαιρία στον διαχειριστή έργου να προσδιορίσει συγκεκριμένη πορεία.

Στα πλαίσια της παρούσας μελέτης, οι κίνδυνοι που εξετάζονται είναι τα πιθανά ελαττωματικά κατά τη διάρκεια παραγωγής των προϊόντων και ο έγκαιρος εντοπισμός τους πριν βγουν αυτά βγουν στην αγορά. Μέσω των σταδίων που προηγήθηκαν έχει επιτευχθεί η πρόβλεψη του αναμενόμενου ελαττωματικού. Ως επόμενο βήμα τίθεται η αξιολόγηση του ρίσκου που αυτό

το ελαττωματικό εισάγει στην παραγωγική διαδικασία. Σκοπός του υπολογισμού του ποιοτικού ρίσκου του κάθε ελαττώματος είναι η ένδειξη των μοναδικών αναγνωριστικών ελαττωμάτων για τα οποία ο έγκαιρος εντοπισμός και η άμεση επίλυση είναι κριτικής σημασίας, με απώτερο στόχο τη γενικότερη ενίσχυση της διαχείρισης ρίσκου του εκάστοτε παραγωγού σχετικά με την ποιότητα. Πιο συγκεκριμένα, με την διαχείριση κινδύνου παρέχεται ένα σημαντικό όπλο για την βελτίωση των σχεδίων της παραγωγικής διαδικασίας και ταυτόχρονα εξασφαλίζεται μια καλύτερη αποτύπωση της πραγματικότητας. Ήδη από το στάδιο του σχεδιασμού επισημαίνονται οι τομείς που χρειάζονται περισσότερη προσοχή. Έτσι είναι δυνατή η πιο άμεση και γρήγορη αντιμετώπιση των αιτιών που οδηγούν σε ελάττωμα. Ακόμη, δίνεται η δυνατότητα ποσοτικοποίησης του κινδύνου και ταυτόχρονα η δυνατότητα παρακολούθησης της πορείας αυτού του παράγοντα στο παρελθόν με σκοπό να αξιοποιηθεί η γνώση στο μέλλον σε άλλα έργα και διαδικασίες.

Στο συγκεκριμένο στάδιο της μεθοδολογίας που προτείνουμε στόχος είναι να απαντηθεί το ερώτημα «πόσο επικίνδυνο είναι ένα ελάττωμα;». Για την απάντηση αυτού του ερωτήματος και τον καθορισμό της κρισιμότητας του κάθε ελαττώματος, όπως αυτό έχει προκύψει μέσω των αλγορίθμων μηχανικής μάθησης, αξιοποιείται ο όρος του «Priority Index». Αυτός υπολογίζεται από τη φόρμουλα που δίνεται παρακάτω. Επισημαίνεται ότι όσο μεγαλύτερη είναι η τιμή του «Priority Index», τόσο μεγαλύτερο είναι και το ρίσκο του ελαττώματος.

Priority Index = $FG(\sum D)*B$, όπου:

- F = Frequency Points
- G = Severity Weight
- $\sum D$ = Detection Points
- B = Brand Points

Είναι εμφανές ότι αυτός ο δείκτης επηρεάζεται από την εκτίμηση της συχνότητας εμφάνισης του ελαττώματος, τον προσδιορισμό της έκτασης της επίπτωσης του ενδεχομένου (για παράδειγμα: κρίσιμη, σημαντική, δευτερεύουσας σημασίας), από τις πηγές αναγνώρισής τους, όπως και την επωνυμία των προϊόντων ή/και εξαρτημάτων με τα οποία σχετίζονται. Ακολουθεί εκτενέστερη ανάλυση ανά παράγοντα που συνυπολογίζεται για την εύρεση του Priority Index του κάθε ελαττώματος.

Βάρος συχνότητας (F: Frequency Points):

Η συχνότητα εμφάνισης του κάθε ελαττώματος επί του συνόλου των εγγραφών αποτελεί αναμφισβήτητα καθοριστικό παράγοντα για την κρισιμότητα αυτού. Για τον υπολογισμό αυτού του παράγοντα είναι αναγκαία η συλλογή όλων των διαθέσιμων εγγραφών καταγραφής ελαττωματικού, όπως και ο καθορισμός μιας αντιστοίχισης μεταξύ του ποσοστού εμφάνισης του κάθε ελαττώματος επί του συνόλου των διαθέσιμων δεδομένων και του βάρους συχνότητας που θα συνυπολογίζεται στην παραπάνω φόρμουλα.



Figure 7 Βάρος Συχνότητας

Βάρος σοβαρότητας (G: Severity Weight):

Για τον υπολογισμό αυτής της παραμέτρου κρίνονται αναγκαία τα παρακάτω βήματα:

- ο καθορισμός συγκεκριμένων σταδίων ποιοτικού ελέγχου στα οποία μπορεί να γίνει κάποια καταγραφή ενός ελαττωματικού.
- ο ορισμός κατηγοριών ως προς τη σοβαρότητα εντοπισμού ενός ελαττώματος ανά στάδιο.
- η αντιστοίχιση της κάθε κατηγορίας σοβαρότητας με το αντίστοιχο βάρος που πρόκειται να ληφθεί υπόψη στον υπολογισμό του Priority Index.

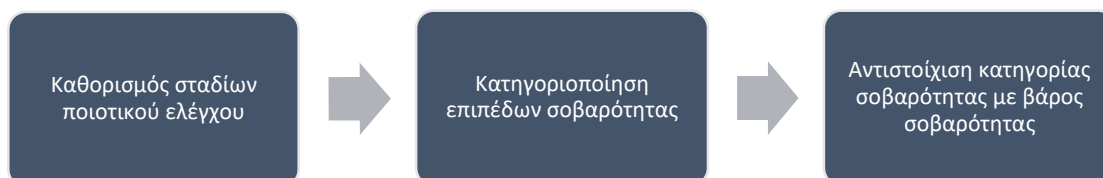


Figure 8 Βάρος Σοβαρότητας

Βάρος πηγής αναγνώρισης (ΣD: Detection Points):

Όπως προαναφέρθηκε στην ενότητα για το Βάρος Σοβαρότητας, χρειάζεται να υπάρχουν καθορισμένα στάδια ποιοτικού ελέγχου στα οποία μπορεί να γίνει κάποια καταγραφή ενός ελαττωματικού. Ο παραγωγός οφείλει, στη συνέχεια, να ορίσει τις πηγές αναγνώρισης ανά στάδιο και το βάρος που αντιστοιχεί σε αυτές. Σε κάθε περίπτωση, ο αντίστοιχος βαθμός θα συνυπολογιστεί στο τελικό αποτέλεσμα.

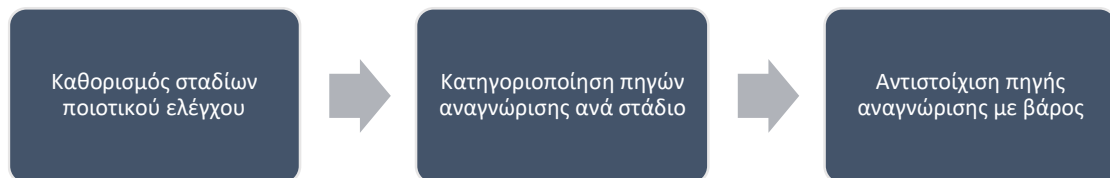
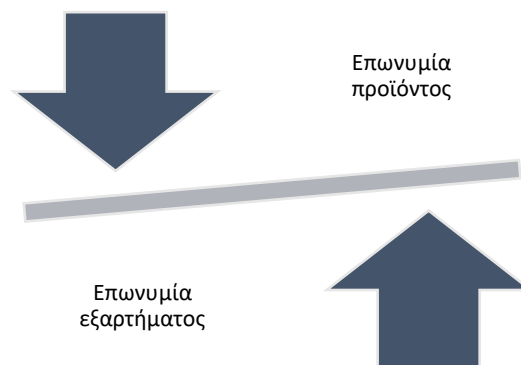


Figure 9 Βάρος πηγής αναγνώρισης

Βάρος Επωνυμίας (B: Brand Points):

Τελευταίο παράγοντα για τον υπολογισμό του Priority Index αποτελεί το Βάρος Επωνυμίας. Αυτό μπορεί να αναφέρεται είτε στην επωνυμία του τελικού προϊόντος που εμφανίζει το ελάττωμα, είτε στην επωνυμία του εξαρτήματος που εμφανίζει το ελάττωμα. Σε κάθε περίπτωση, ο διαχωρισμός και η βαθμονόμηση θα γίνεται από τον εκάστοτε παραγωγό και θα συνυπολογίζεται αντίστοιχα για τον καθορισμό της κρισιμότητας του εκάστοτε ελαττώματος.



4

Μελέτη Περίπτωσης

4.1 Υλοποίηση

Η γλώσσα προγραμματισμού που επιλέχθηκε για την κατασκευή του αλγορίθμου και την επίτευξη των στόχων της διπλωματικής είναι η Python, καθώς βιβλιογραφικά ενδείκνυται για την επίλυση προβλημάτων που κατηγοριοποιούνται ομοίως με το δικό μας πρόβλημα. Η Python είναι μια διερμηνευόμενη γλώσσα προγραμματισμού, γενικού σκοπού και υψηλού επιπέδου. Ανήκει στις γλώσσες προστακτικού προγραμματισμού και υποστηρίζει τόσο το διαδικαστικό όσο και το αντικειμενοστραφές υπόδειγμα. Είναι δυναμική γλώσσα προγραμματισμού και υποστηρίζει συλλογή απορριμμάτων. Επισημαίνεται ότι λόγω της αναγνωσιμότητας του κώδικά της, της ευκολία χρήσης της, όπως και της αφθονίας των διαθέσιμων βιβλιοθηκών της, η διαδικασία υλοποίησης του ζητούμενου διευκολύνθηκε σε σημαντικό βαθμό.

Η πλατφόρμα στην οποία αναπτύχθηκε ο κώδικας μας είναι το Jupyter Notebook. Πρόκειται για μια διαδικτυακή εφαρμογή ανοιχτού κώδικα που επιτρέπει στους επιστήμονες δεδομένων να δημιουργούν και να μοιράζονται έγγραφα που ενσωματώνουν ζωντανό κώδικα, εξισώσεις, υπολογιστικές εξόδους, απεικονίσεις και άλλους πόρους πολυμέσων, μαζί με επεξηγηματικό κείμενο σε ένα μόνο έγγραφο. Πράγματι αποδείχθηκε ένα χρήσιμο εργαλείο που επέτρεψε τον διαμοιρασμό του κώδικα και τον άμεσο έλεγχο του καθ' όλη τη διάρκεια της υλοποίησης.

Τέλος, έγινε εκτεταμένη χρήση δύο βιβλιοθηκών της Python, η Pandas και η Scikit-learn, οι οποίες προσέφεραν ισχυρές και αποτελεσματικές βάσεις για την τελική υλοποίηση. Η Pandas χρησιμοποιείται ευρέως για ανάλυση δεδομένων και είναι μια από τις πιο συχνά χρησιμοποιούμενες βιβλιοθήκες Python. Παρέχει μερικά από τα πιο χρήσιμα σετ εργαλείων

για την εξερεύνηση, την επεξεργασία και την ανάλυση των διαθέσιμων δεδομένων. Με την Pandas κατέστη, επίσης, δυνατός ο χειρισμός των δεδομένων μας μέσω της δομής των DataFrames. Σε δεύτερο στάδιο, Scikit-learn αποτέλεσε την πιο σημαντική βιβλιοθήκη για τη δημιουργία μοντέλων μηχανικής εκμάθησης χάρη στα χιλιάδες εργαλεία που διαθέτει για προγνωστική μοντελοποίηση και ανάλυση, όπως και για την τελική επικύρωση και σύγκριση της ακρίβειας αυτών.

4.2 Πειραματικά Αποτελέσματα

4.2.1 Εφαρμογή στη Βιομηχανία Οικιακών Συσκευών

Με σκοπό τον πειραματικό έλεγχο και την πρακτική υποστήριξη της προσέγγισης που προτείνουμε ως προς τον προσδιορισμό και την αξιολόγηση του ρίσκου ποιότητας που ενδέχεται να προκύψει από το αναμενόμενο ελάττωμα ενός δοσμένου προϊόντος, αξιοποιήθηκαν δεδομένα που προέρχονται από πολυεθνικό παραγωγό και έμπορο οικιακών συσκευών.

Στο μοντέλο παραγωγής του παραγωγού, ολόκληρη η παραγωγή λευκών ειδών ελέγχεται από άποψη ποιότητας και ασφάλειας για να διασφαλιστεί ένα υψηλό επίπεδο ποιότητας προϊόντος στους τελικούς πελάτες. Για τις ενέργειες δοκιμής, το σύστημα παραγωγής προσθέτει επίσης ορισμένες ενέργειες στατιστικού ελέγχου ποιότητας που εφαρμόζονται τόσο σε εσωτερικά εξαρτήματα παραγωγής σε κρίσιμες διαδικασίες ποιότητας όσο και σε τελικά προϊόντα, μετά τη διαδικασία συσκευασίας. Αυτό ονομάζεται δοκιμή μηδενικής ώρας (ZHT), και αναφέρεται στον Στατιστικό Ποιοτικό Έλεγχο που εφαρμόζεται σε εξειδικευμένο εργαστήριο ροής εκτός παραγωγής σε ορισμένα τελικά προϊόντα. Οι κύριοι στόχοι του είναι να μετρήσει το επίπεδο ποιότητας του εξερχόμενου προϊόντος από αισθητικό, λειτουργικό και κανονιστικό σημείο και να μετρήσει την αποτελεσματικότητα του ελέγχου της διαδικασίας. Αυτές οι δοκιμές εκτελούνται σε ένα αποκλειστικό εργαστηριακό περιβάλλον, που δημιουργείται σε κάθε εργοστάσιο παραγωγής και ακολουθούν μια συγκεκριμένη διαδικασία λειτουργίας STD που ορίζεται κατά τη φάση Σχεδιασμού Προϊόντος. Οι Διευθυντές Ποιότητας Εργοστασίου είναι πλήρως υπεύθυνοι για την εκτέλεση της διαδικασίας, ενώ το Κεντρικό Τμήμα Ποιότητας είναι υπεύθυνο για το σχεδιασμό της διαδικασίας με το τμήμα Μηχανικής Προϊόντων. Αυτή η μέθοδος δοκιμής έχει σχεδιαστεί για να αναπαράγει την προσέγγιση του πελάτη στο προϊόν, προσομοιώνοντας τις κανονικές συνθήκες χρήσης του προϊόντος κατά την πρώτη χρήση του τελικού πελάτη που ακολουθεί τις οδηγίες στο φυλλάδιο χρήσης και συντήρησης, εκτελώντας

την προτεινόμενη εγκατάσταση προϊόντος και δοκιμή της πρώτης χρήσης σε συνθήκες σπιτιού. Εάν κατά την εκτέλεση ενός βήματος ποιοτικού ελέγχου εντοπιστεί ένα ελάττωμα, ο χειριστής εισαγάγει το αναγνωριστικό ελάττωμα και το αναγνωριστικό εμπλεκόμενων μερών στο σύστημα DCS, όπου όλα τα ελαττώματα που εντοπίζονται στο εργοστασιακό δίκτυο ποιότητας καταγράφονται για σκοπούς ανάλυσης και βελτίωσης.

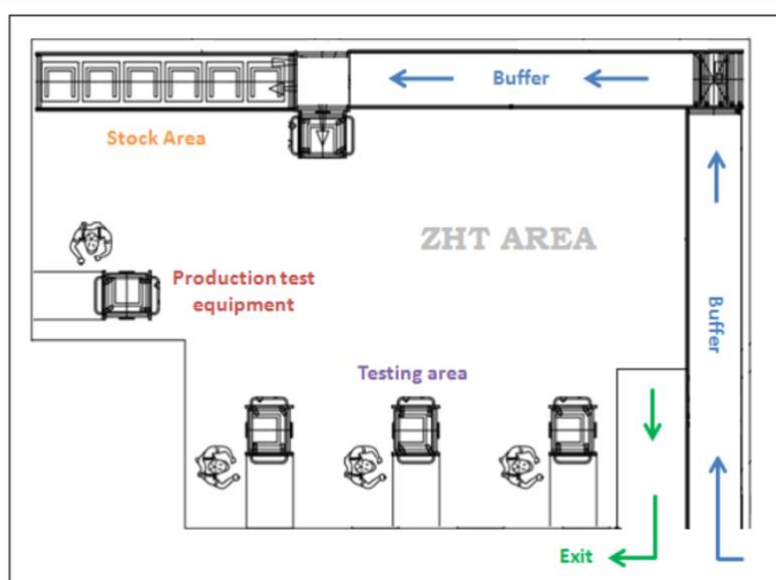


Figure 10 Διαδικασίες Δοκιμής Μηδενικής Ωρας

Σε κάθε περίπτωση, αυτή η καταγραφή δεν οδηγεί σε επανάληψη της διαδικασίας σχεδιασμού του προϊόντος, περιορίζει το εύρος αντίδρασης σε δράσεις μετριασμού σε επίπεδο παραγωγής και μειώνει την πιθανότητα πρόβλεψης παρόμοιων ζητημάτων ποιότητας για άλλα προϊόντα υπό ανάπτυξη. Ταυτόχρονα, η διαδικασία που αντιμετωπίζει το ζήτημα ποιότητας ενεργοποιείται μόνο μετά την ανακάλυψη του ζητήματος σε επίπεδο αγοράς με αντίκτυπο στην εκτίμηση των πελατών, χάνοντας τη δυνατότητα πρόβλεψης του ζητήματος για τους πελάτες και σωστής διαχείρισης του κινδύνου. Αυτή η καθυστερημένη αναγνώριση κατά τη χρήση οδηγεί σε αυξανόμενο κόστος των πολιτικών εγγύησης και στη δυσαρέσκεια των πελατών. Επιπλέον, ο ποιοτικός έλεγχος στο τελευταίο στάδιο της παραγωγικής διαδικασίας αποτελεί κυρίως μια αντιδραστική προσέγγιση με στόχο την αποτροπή παράδοσης ελαττωματικών προϊόντων στους πελάτες και τη διόρθωση των αναποτελεσματικών στη διαδικασία παραγωγής για τις επόμενες παρτίδες.

Στόχο της παρούσας προσέγγισης αποτελεί η υλοποίηση μιας προγνωστικής τεχνικής για την πρόβλεψη προϊόντων και διαδικασιών χαμηλής ποιότητας, δίνοντας τη δυνατότητα στον

παραγωγό να σχεδιάσει δράσεις μετριασμού και να δράσει προληπτικά στις βαθύτερες αιτίες των ελαττωμάτων.

4.2.2 Δεδομένα

Οι πηγές δεδομένων που είχαμε στη διάθεσή μας αντιστοιχούν σε 4 κατηγορίες:

- (i) δεδομένα ποιοτικού ελέγχου από το κατάστημα (DCS, Previdi, EOL, ZHT Lab)
- (ii) δεδομένα εξυπηρέτησης πελατών (CSBI)
- (iii) δεδομένα παραγωγής (SAP- HANA-EEP, SAP MII)
- (iv) μη δομημένα δεδομένα που σχετίζονται με γνώσεις σχετικά με θέματα ποιότητας και διαδικασίες (GDrive).

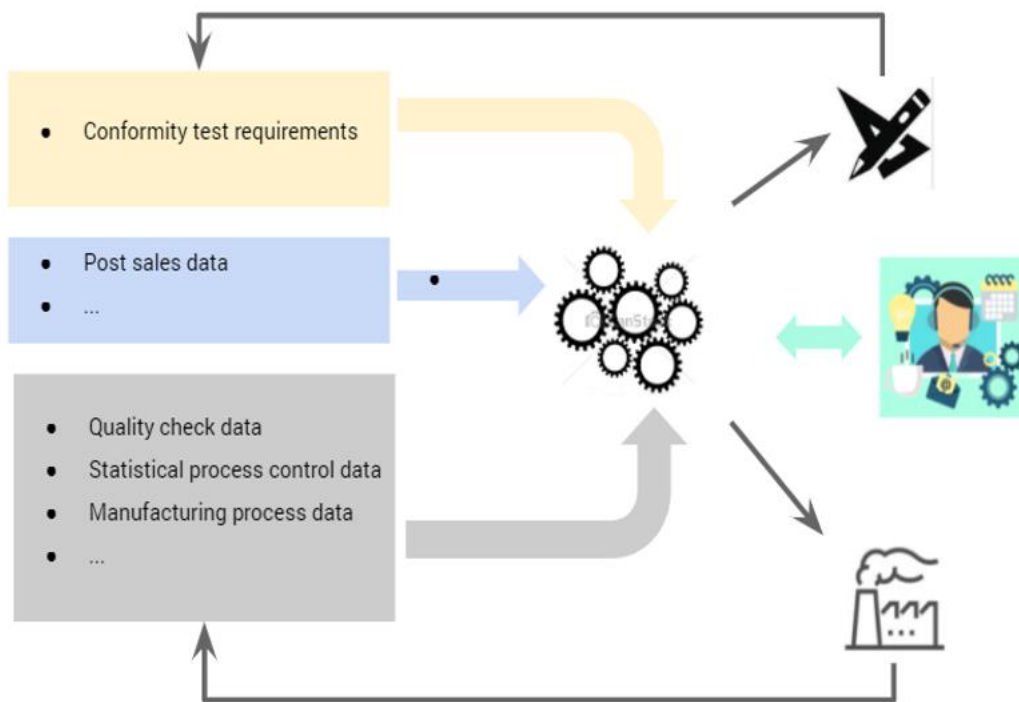


Figure 11 Συλλογή Δεδομένων από τα επιμέρους στάδια της γραμμής παραγωγής

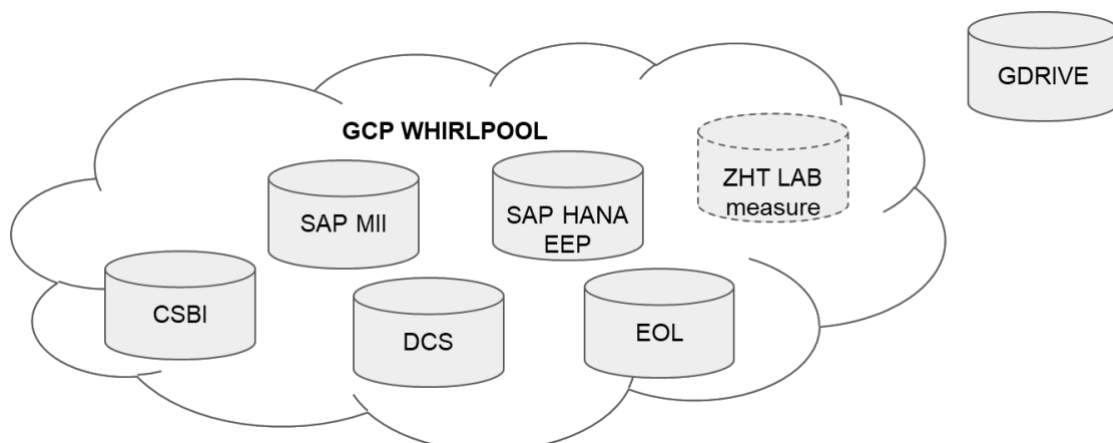


Figure 12 Μελέτη Περίπτωσης - Πηγές Δεδομένων

Μία από τις προκλήσεις που αντιμετωπίσαμε ήταν η επιλογή του κατάλληλου dataset για την εξαγωγή προβλέψεων. Λόγω της συσχέτισης των αναγνωριστικών ελαττωμάτων με τα τελικά προϊόντα αλλά και τα εμπλεκόμενα μέρη στο σύστημα DCS, όπως αυτά εντοπίζονται και καταγράφονται από τους χειριστές κατά τον έλεγχο ποιότητας στο στάδιο ZHT, αξιοποιήσαμε το συγκεκριμένο σύνολο δεδομένων ως ιστορική βάση για τα μοντέλα μηχανικής μάθησης που χρησιμοποιήσαμε στη συνέχεια.

Εφόσον εντοπίσαμε το κατάλληλο dataset, τέθηκε ως πρώτος στόχος της υλοποίησης η ανάλυση αυτού, ώστε να καθοριστούν τα χαρακτηριστικά που θα έπρεπε να αξιοποιηθούν ή/και να κατασκευαστούν για τις προβλέψεις που θα ακολουθήσουν. Έτσι, επιχειρήθηκε μια αρχική ανάλυση του ιστορικού dataset απ' όπου προέκυψαν τα εξής ευρήματα:

- Είχαμε στη διάθεσή μας 25.654 εγγραφές, για τις οποίες δεν υπήρχαν κενές τιμές σε συγκεκριμένα επιλεγμένα χαρακτηριστικά που είχαμε ήδη διαθέσιμα, όπως:
 - Το FGNum, δηλαδή τη μονάδα τήρησης αποθεμάτων (SKU -Stock keeping unit) / το μοναδικό αναγνωριστικό ενός προϊόντος.
 - Το OrdNum, δηλαδή το μοναδικό αναγνωριστικό μιας παραγγελίας.
 - DefectID, δηλαδή το μοναδικό αναγνωριστικό ενός ελαττώματος.
 - DefectGrp, δηλαδή την ευρύτερη ομάδα στην οποία ανήκει κάποιο ελάττωμα.
 - DefectSrcID, δηλαδή τις πηγές εμφάνισης ελαττωμάτων.
 - MatNum, δηλαδή το μοναδικό αναγνωριστικό ενός εξαρτήματος.
 - FamilyID, δηλαδή την ευρύτερη οικογένεια στην οποία ανήκει κάποιο εξάρτημα.
 - InsertDate, δηλαδή την ημερομηνία που είχε γίνει η κάθε εγγραφή
 - LinSeq, δηλαδή τη γραμμή παραγωγής από την οποία προέκυψε το ελαττωματικό προϊόν.
- Με υπολογισμό των μοναδικών τιμών ανά χαρακτηριστικό βρέθηκε ότι για το διάστημα 23/04/2020 έως 16/04/2021:
 - Τα μοναδικά προϊόντα (FGNum) για τα οποία υπήρχαν εγγραφές ήταν 94.
 - Οι μοναδικές παραγγελίες στις οποίες εντοπίστηκαν ελαττωματικά προϊόντα ήταν 4295.
 - Οι μοναδικές τιμές ελαττωμάτων (DefectID) ήταν 422.
 - Οι μοναδικές τιμές για τις ομάδες ελαττωμάτων (DefectGrp) ήταν 16.
 - Οι μοναδικές πηγές εμφάνισης ελαττωμάτων ήταν 7.
 - Τα μοναδικά εξαρτήματα για τα οποία είχε βρεθεί ελάττωμα ήταν 494.

- Οι οικογένειες στις οποίες ανήκαν τα καταγεγραμμένα εξαρτήματα που παρουσίαζαν ελάττωμα ήταν 62.
- Είχαμε μετρήσεις 270 μοναδικών ημερών.
- Ελαττωματικά προϊόντα προέκυψαν από 1987 διαφορετικές γραμμές παραγωγής.

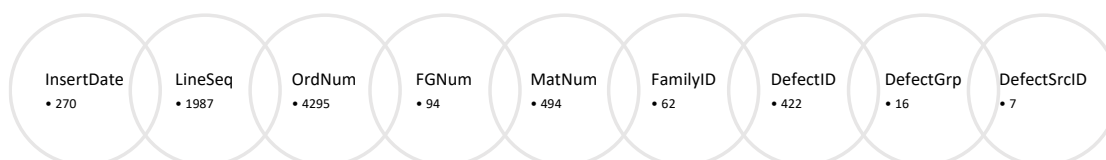


Figure 13 Μοναδικές τιμές ανά χαρακτηριστικό

Συνδυάζοντας τις παραπάνω πληροφορίες καταλήξαμε, ακολούθως, στα παρακάτω συμπεράσματα:

- Για τα περισσότερα προϊόντα υπάρχει καταγραφή περισσότερων του ενός ελαττώματων.
 - Όντως, μετά από περαιτέρω ανάλυση βρέθηκε ότι για ένα προϊόν υπήρχαν 3964 εγγραφές, ενώ μόνο δύο προϊόντα είχαν μοναδική εγγραφή.
- Κάθε ομάδα ελαττωμάτων αποτελείται από πολλά μοναδικά ελαττώματα / το DefectGrp αποτελεί υπερσύνολο του DefectID.
- Κάθε οικογένεια εξαρτημάτων αποτελείται από πολλά μοναδικά εξαρτήματα / το FamilyID αποτελεί υπερσύνολο του MatNum.
- Υπάρχουν μέρες κατά τις οποίες είτε δεν έγινε κάποια έλεγχος για ελαττώματα είτε έγινε έλεγχος και δεν εντοπίστηκαν ελαττώματα.

Μετά την πρώτη ανάλυση δεδομένων και αφού καταλήξαμε σε κάποια αρχικά συμπεράσματα, προχωρήσαμε στην περαιτέρω επεξεργασία του συνόλου δεδομένων. Πιο συγκεκριμένα:

Μετατροπή μορφής χαρακτηριστικού InsertDate:

Η διαθέσιμη μορφή καταγραφής της ημερομηνίας της κάθε εγγραφής ήταν σε λεπτομέρεια δευτερολέπτου (yyyy-mm-dd hh:mm:ss.s), με αποτέλεσμα να μην μπορούν να εξαχθούν εύκολα πληροφορίες για την ημερήσια, εβδομαδιαία ή μηνιαία συχνότητα εμφάνισης του κάθε ελαττώματος, να μην μπορούν να κατανοηθούν καλύτερα οι αιτίες πίσω από αυτήν και, τελικά,

να εμποδίζεται επιτυχώς η πρόβλεψη μελλοντικών τάσεων. Προχωρήσαμε, επομένως, σε μετατροπή της μορφής της ημερομηνίας σε λεπτομέρεια μέρας (yyyy-mmm-dd).

Table 3 Μετατροπή μορφής χαρακτηριστικού InsertDate

Δείγμα Ημερομηνίας Πριν τη Μετατροπή	Δείγμα Ημερομηνίας Μετά τη Μετατροπή
2020-04-23 17:01:49.007	2020-04-23 00:00:00

Κατασκευή του χαρακτηριστικού «diff»:

Με την μετατροπή της ημερομηνίας της κάθε εγγραφής από datetime σε date κατέστη δυνατή η δημιουργία ενός νέου χαρακτηριστικού ανά εγγραφή, αυτό του «diff», το οποίο υπολογίστηκε ως η διαφορά της εκάστοτε ημερομηνίας της εγγραφής που ελέγχουμε και για συγκεκριμένο εξάρτημα με την ημερομηνία της τελευταίας εγγραφής όπου εμφανίστηκε το ελάττωμα (DefectID). Επισημαίνεται σε αυτό το σημείο ότι ο υπολογισμός της προαναφερθείσας διαφοράς ημερομηνιών έγινε με τη χρήση ενός επιπλέον χαρακτηριστικού που κατασκευάστηκε για τις ανάγκες της υλοποίησης, το «unique». Αυτό προέκυψε από συνδυασμό του αναγνωριστικού εξαρτήματος (MatNum), της οικογένειας του εξαρτήματος (FamilyID) και του αναγνωριστικού του ελαττώματος (DefectID) (παράδειγμα: 400010914457_419_1240) και κρίθηκε αναγκαίο για τους εξής λόγους:

- Χρειαζόταν ένα μοναδικό αναγνωριστικό επί του οποίου θα υπολογιζόταν η καθυστέρηση εμφάνισης του κάθε ελαττώματος.
- Το αναγνωριστικό του ελαττώματος δεν θα μπορούσε να χρησιμοποιηθεί αυτούσιο ως κλειδί, καθώς η πληροφορία που μας ενδιέφερε ήταν η καθυστέρηση εμφάνισης ενός ελαττώματος που αφορά σε συγκεκριμένο εξάρτημα.
- Βάσει της αρχικής ανάλυσης γνωρίζουμε ότι ένα εξάρτημα μπορεί να ανήκει σε παραπάνω από μία οικογένειες εξαρτημάτων, επομένως μόνο ο συνδυασμός αυτών των δύο πεδίων και όχι η μεμονωμένη αξιοποίησή τους θα μπορούσε να δώσει σωστά στοιχεία για τα μετέπειτα μοντέλα προβλέψεων.

Αποτελέσματα από τη δημιουργία του χαρακτηριστικού «diff» παρουσιάζονται ενδεικτικά στον παρακάτω πίνακα:

Table 4 Δείγμα εγγραφών για χαρακτηριστικό "diff"

date	DefectID	diff
2020-04-23	1240	0

2020-04-23	1240	0
2020-04-23	1240	0
2020-04-23	1240	0
2020-04-23	1285	0
2020-04-23	1285	0
2020-04-23	1697	0
2020-04-23	1697	0
2020-04-23	1697	0
2020-04-23	1697	0
2020-04-23	1697	0
2020-04-23	1697	0
2020-04-23	1697	0
2020-04-27	1697	4

Πράγματι, η τιμή 4 δόθηκε για την εγγραφή της 2020-04-27, καθώς το ελάττωμα με αναγνωριστικό 1697 είχε, προηγουμένως, εμφανιστεί 4 μέρες νωρίτερα, στις 2020-04-23. Παρατηρούμε ότι τιμή 0 έχει δοθεί στις εγγραφές όπου καταγράφουν ελάττωμα που έχει καταγραφεί ξανά μέσα στην ίδια μέρα στην οποία αναφέρεται η εγγραφή. Αξίζει, επίσης, να σημειωθεί ότι το συγκεκριμένο τμήμα των δεδομένων έχει ληφθεί από την πρώτη εγγραφή και έπειτα, οι μηδενικές τιμές είναι, λοιπόν, λογικές λόγω έλλειψης προηγούμενης ιστορικότητας δεδομένων.

Κατασκευή του χαρακτηριστικού «Frequency of DefectID / date»:

Με τη μετατροπή της ημερομηνίας της κάθε εγγραφής από datetime σε date μπορούσε να δημιουργηθεί ακόμα ένα χαρακτηριστικό, το «Frequency of DefectID / date», το οποίο υποδηλώνει τη συχνότητα εμφάνισης ενός ελαττώματος σε μια συγκεκριμένη ημερομηνία. Δανειζόμενοι τις εγγραφές που χρησιμοποιήθηκαν και παραπάνω, παρατηρούμε πώς εμπλουτίστηκαν τα δεδομένα με την επιπλέον πληροφορία:

Table 5 Δείγμα εγγραφών για χαρακτηριστικό "Frequency of DefectID / date"

date	DefectID	Frequency of DefectID / date
2020-04-23	1240	4
2020-04-23	1240	4
2020-04-23	1240	4
2020-04-23	1240	4
2020-04-23	1697	6
2020-04-23	1697	6
2020-04-23	1697	6
2020-04-23	1697	6
2020-04-23	1697	6
2020-04-23	1697	6
2020-04-27	1697	1

Μετά κι από τις τελευταίες κατασκευές χαρακτηριστικών προέκυψε ένα σύνολο δεδομένων με την παρακάτω δομή:

	date	DefectID	Frequency of DefectID / date	LitSeq	OrdNum	FDNum	MatNum	FamilyID	DefectGrp	DefectDetID	dB	unique
0	2020-04-23 00:00:00	1240	4	20229	309883898	8,599820+11	4,00011E+11	419	1291	0	0	400010914457_419_1240
1	2020-04-23 00:00:00	1240	4	20281	309810389	8,599820+11	4,00011E+11	420	1291	3	0	400011045755_420_1240
2	2020-04-23 00:00:00	1240	4	10620	309644391	8,587840+11	4,00011E+11	419	1291	6	0	400010794742_419_1240
3	2020-04-23 00:00:00	1240	4	10636	309644581	8,599820+11	4,00011E+11	420	1291	3	0	400011316478_420_1240
4	2020-04-23 00:00:00	1285	2	20258	309644579	8,599820+11	4,00011E+11	420	1291	5	0	400010801501_420_1285
5	2020-04-23 00:00:00	1285	2	20259	309644492	8,599820+11	4,00011E+11	420	1291	3	0	400010687018_420_1285
6	2020-04-23 00:00:00	1697	6	10607	308728887	8,587840+11	4,00011E+11	417	1591	6	0	400010807966_417_1697
7	2020-04-23 00:00:00	1697	6	10616	308718641	8,599820+11	4,00011E+11	417	1591	2	0	400010878686_417_1697
8	2020-04-23 00:00:00	1697	6	10617	309644832	8,599820+11	4,00011E+11	417	1591	2	0	400010908913_417_1697
9	2020-04-23 00:00:00	1697	6	10627	308644392	8,599820+11	4,00011E+11	418	1591	3	0	400011131073_418_1697
10	2020-04-23 00:00:00	1697	6	10587	308718640	8,599820+11	4,00011E+11	417	1591	2	0	400010813294_417_1697
11	2020-04-23 00:00:00	1697	6	20258	309644579	8,599820+11	4,00011E+11	418	1591	3	0	400010852325_418_1697
12	2020-04-24 00:00:00	1697	2	10629	309644738	8,599820+11	4,619690+11	1047	1691	5	0	461968975611_1047_1697
13	2020-04-24 00:00:00	1697	2	10635	308726889	8,599820+11	4,619690+11	1047	1691	5	0	461968677761_1047_1697
14	2020-04-24 00:00:00	1181	1	20226	308883893	8,599820+11	4,619690+11	1047	1691	4	0	461968977761_1047_1181
15	2020-04-24 00:00:00	1238	1	10636	308718641	8,599820+11	4,00011E+11	9000	1291	1	0	400010778046_9000_1238
16	2020-04-24 00:00:00	1285	1	10642	309644578	8,591840+11	4,619690+11	9999	1291	4	0	461969361751_9999_1285
17	2020-04-24 00:00:00	1285	1	10681	303387977	8,599820+11	4,00011E+11	419	1291	6	0	400011336671_419_1285
18	2020-04-24 00:00:00	1777	1	10617	309644832	8,599820+11	4,00011E+11	0	1591	4	0	400011381755_0_1777
19	2020-04-24 00:00:00	1802	1	10641	309644578	8,591840+11	4,00011E+11	1034	1691	5	0	400010716272_1034_1802
20	2020-04-24 00:00:00	2045	1	20220	308537551	8,599820+11	4,619690+11	1007	2091	2	0	461967846483_1007_2045
21	2020-04-24 00:00:00	2017	2	10627	308644392	8,599820+11	4,00011E+11	9000	2091	5	0	400010624600_9000_2017
22	2020-04-24 00:00:00	3017	2	10641	309644578	8,591840+11	4,00011E+11	9999	2091	1	0	400010473108_9999_3017
23	2020-04-24 00:00:00	3035	6	10627	308644392	8,599820+11	4,00011E+11	9000	2091	5	0	400010624600_9000_3035
24	2020-04-24 00:00:00	3035	6	10596	308724382	8,599820+11	4,00011E+11	9999	2091	1	0	400010778046_9999_3035
25	2020-04-24 00:00:00	3035	6	10617	309644832	8,599820+11	4,00011E+11	501	2091	1	0	400011164464_501_3035
26	2020-04-24 00:00:00	3035	6	10617	309644832	8,599820+11	4,00011E+11	1016	2091	1	0	400010614553_1016_3035
27	2020-04-24 00:00:00	3035	6	10641	309644578	8,591840+11	4,00011E+11	9000	2091	1	0	400010778046_9000_3035
28	2020-04-24 00:00:00	3035	8	10636	308718641	8,599820+11	4,00011E+11	1016	2091	1	0	400010624553_1016_3035
29	2020-04-27 00:00:00	1001	1	20454	91334277	8,599820+11	4,619690+11	1047	1091	1	0	461968677761_1047_1001
30	2020-04-27 00:00:00	1019	1	10486	307587980	8,587840+11	4,619690+11	1047	1091	3	0	461968975611_1047_1019
31	2020-04-27 00:00:00	1115	1	20454	91334277	8,599820+11	4,619690+11	1047	1091	3	0	461968677761_1047_1115
32	2020-04-27 00:00:00	1181	1	20454	91334277	8,599820+11	4,00011E+11	1040	1091	3	0	400010367896_1040_1181
33	2020-04-27 00:00:00	1181	1	20454	91334277	8,599820+11	4,00011E+11	1040	1091	3	0	400010733803_1040_1181
34	2020-04-27 00:00:00	1181	1	20454	91334277	8,599820+11	4,00011E+11	1040	1091	3	0	400010733403_1040_1181
35	2020-04-27 00:00:00	1182	1	20454	91334277	8,599820+11	4,00011E+11	1040	1091	3	0	400010865553_1040_1182
36	2020-04-27 00:00:00	1240	1	10595	308134281	8,599820+11	4,00011E+11	420	1291	3	0	400010851853_420_1240

Figure 14 Δομή dataset μετά την κατασκευή χαρακτηριστικών

Πιο συγκεκριμένα, ο μορφότυπος ανά πεδίο / χαρακτηριστικό φαίνεται παρακάτω:

#	Column	Non-Null Count	Dtype
0	date	25654 non-null	datetime64[ns]
1	DefectID	25654 non-null	int64
2	Frequency of DefectID / date	25654 non-null	int64
3	LinSeq	25654 non-null	int64
4	OrdNum	25654 non-null	int64
5	FGNum	25654 non-null	int64
6	MatNum	25654 non-null	int64
7	FamilyID	25654 non-null	int64
8	DefectGrp	25654 non-null	int64
9	DefectSrcID	25654 non-null	int64
10	diff	25654 non-null	timedelta64[ns]
11	unique	25654 non-null	object

Μετατροπή όλων των χαρακτηριστικών σε ετικέτες:

Οι αλγόριθμοι που χρησιμοποιούμε στη συνέχεια δεν λειτουργούν με χαρακτηριστικά κειμένου, γι' αυτό λόγο, χρειάστηκε να μετατρέψουμε το κείμενο σε αριθμητικά χαρακτηριστικά. Για να μετατρέψουμε αυτού του είδους τα δεδομένα κατηγοριών κειμένου σε αριθμητικά δεδομένα κατανοητά από το μοντέλο, αξιοποιήσαμε την κλάση Label Encoder. Αντίστοιχα, κωδικοποιήσαμε κάθε στήλη του συνόλου δεδομένα με ετικέτα εισάγοντας την κλάση LabelEncoder από τη βιβλιοθήκη sklearn, με αποτέλεσμα να καταλήξουμε στην αντικατάσταση των υπάρχοντων δεδομένων με τα νέα κωδικοποιημένα δεδομένα.

Ως εκ τούτου, μετά από την μετατροπή των χαρακτηριστικών σε ετικέτες, το σύνολο δεδομένων έλαβε την τελική μορφή που φαίνεται παρακάτω:

Table 6 Δομή dataset μετά την κωδικοποίηση χαρακτηριστικών

date	DefectID	Frequency of DefectID / date	LinSeq	OrdNum	FGNum	MatNum	FamilyID	DefectGrp	DefectSrcID	diff	unique
0	70	3	1224	49	47	250	8	3	6	0	2175
0	70	3	1273	115	47	263	9	3	3	0	2219
0	70	3	619	71	14	92	8	3	6	0	1204
0	70	3	635	81	70	372	9	3	3	0	2903
0	92	1	1250	79	33	169	9	4	3	0	1837
0	92	1	1251	75	46	66	9	4	3	0	834
0	209	5	606	65	6	140	6	6	6	0	1705
0	209	5	615	55	43	208	6	6	2	0	2014
0	209	5	616	85	59	246	6	6	2	0	2147

0	209	5	626	72	37	292	7	6	3	0	2551
0	209	5	586	54	36	142	6	6	2	0	1714
0	209	5	1250	79	33	171	7	6	3	0	1860
1	3	1	628	89	91	476	57	0	5	0	3481
1	3	1	614	66	47	488	57	0	5	0	3594
1	49	0	1221	50	68	488	57	0	4	0	3624
1	69	0	615	55	43	122	61	3	1	0	1450
1	78	0	640	78	22	491	61	3	4	0	3691
1	92	0	680	142	65	326	8	4	6	0	2698
1	229	0	616	85	59	409	0	6	4	0	3225
2	122	3	552	46	35	34	53	4	4	0	338
2	122	3	660	112	36	34	53	4	4	0	338
2	136	2	643	91	22	147	38	4	2	0	1735
2	136	2	1326	226	88	101	38	4	2	0	1257
2	136	2	1328	228	88	101	38	4	2	0	1257

Πράγματι, οι μετατροπές φαίνονται και από τις ακόλουθες λεπτομέρειες ανά πεδίο/χαρακτηριστικό:

```
#      Column      Non-Null Count  Dtype
---  -
0      date          25654 non-null  int64
1      DefectID        25654 non-null  int64
2      Frequency of DefectID / date 25654 non-null  int64
3      LinSeq           25654 non-null  int64
4      OrdNum           25654 non-null  int64
5      FGNum            25654 non-null  int64
6      MatNum           25654 non-null  int64
7      FamilyID         25654 non-null  int64
8      DefectGrp        25654 non-null  int64
9      DefectSrcID      25654 non-null  int64
10     diff              25654 non-null  int64
11     unique            25654 non-null  int64
dtypes: int64(12)
```

Έχοντας την τελική μορφή του συνόλου δεδομένων, μπορούσαμε πλέον να προχωρήσουμε στην εφαρμογή των αλγορίθμων μηχανικής μάθησης για την πρόβλεψη του ελαττωματικού ανά εισαγόμενες πληροφορίες. Το πειραματικό σκέλος της διαδικασίας πρόβλεψης παρατίθεται λεπτομερέστερα στην επόμενη ενότητα.

4.2.3 Πρόβλεψη Ελαττωματικών

Έχοντας στη διάθεσή μας ένα σύνολο από χαρακτηριστικά, προχωρήσαμε με την εφαρμογή διάφορων αλγορίθμων μηχανικής μάθησης, και συγκεκριμένα με αλγορίθμους ταξινόμησης, με σκοπό την πρόβλεψη του μοναδικού αναγνωριστικού ελαττωματικού (DefectID). Στα πλαίσια της πειραματικής διαδικασίας ο κάθε αλγόριθμος εφαρμόστηκε πολλαπλές φορές, με τροποποίηση παραμέτρων. Πιο συγκεκριμένα, οι προσαρμοζόμενες παράμετροι ανά εφαρμογή ήταν οι εξής:

- Τα πεδία του dataset που δίνονταν ως χαρακτηριστικά – features στον αλγόριθμο για την πρόβλεψη του DefectID.

Συνδυασμός Χαρακτηριστικών	Πεδία dataset ως Features
A	"date" "Frequency of DefectID / date" "diff" "unique" "MatNum" "FamilyID" "LinSeq" "OrdNum" "FGNum"
B	"date" "Frequency of DefectID / date " "diff" "unique" "MatNum" "FamilyID" "LinSeq" "OrdNum"
C	"date" "unique" "MatNum" "FamilyID"

	"LinSeq" "OrdNum"
D	"date" "MatNum" "FamilyID" "OrdNum" "FGNum"
E	"MatNum" "FamilyID" "OrdNum" "FGNum"
F	"date" " Frequency of DefectID / date" "diff" "unique"
G	"MatNum" "FamilyID" "FGNum"
H	"date" "unique" "FGNum"
I	"MatNum" "FGNum"
J	"date" "unique"

- Το ποσοστό του συνόλου εκπαίδευσης του μοντέλου:
 - Αξιοποίηση του 65% των διαθέσιμων δεδομένων για εκπαίδευση, και του υπόλοιπου 35% για επικύρωση της πρόβλεψης.
 - Αξιοποίηση του 75% των διαθέσιμων δεδομένων για εκπαίδευση, και του υπόλοιπου 25% για επικύρωση της πρόβλεψης.
 - Αξιοποίηση του 85% των διαθέσιμων δεδομένων για εκπαίδευση, και του υπόλοιπου 15% για επικύρωση της πρόβλεψης.

Παρακάτω δίνονται τα πειραματικά αποτελέσματα με έμφαση στο ποσοστό ακρίβειας του κάθε αλγορίθμου για διαφορετικούς συνδυασμούς πεδίων ως χαρακτηριστικά (features) για την ταξινόμηση και για διαφορετικά ποσοστά του συνόλου δεδομένων που αξιοποιήθηκαν για την εκπαίδευση του εκάστοτε αλγορίθμου:

Table 7 Αποτελέσματα απόδοσης Random Forest

Συνδυασμός Χαρακτηριστικών	Ακρίβεια Πρόβλεψης 65% εκπαίδευση	Ακρίβεια Πρόβλεψης 75% εκπαίδευση	Ακρίβεια Πρόβλεψης 85% εκπαίδευση
A	74,43%	74,99%	76,05%
B	71,32%	72,39%	72,56%
C	71,07%	72,28%	73,53%
D	48,84%	48,47%	48,53%
E	47,10%	47,10%	47,10%
F	65,70%	66,96%	69,45%
G	42,59%	42,84%	43,31%
H	57,49%	58,15%	60,82%
I	42,20%	42,61%	42,61%
J	53,74%	54,85%	57,99%

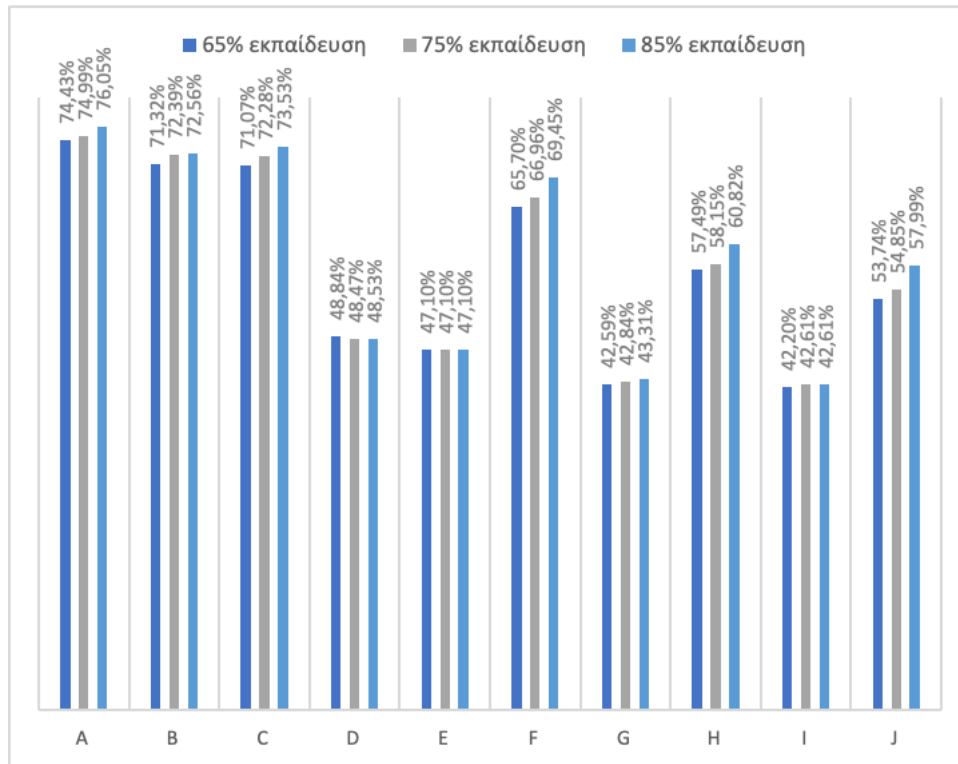


Figure 15 Αποτελέσματα απόδοσης Random Forest

Table 8 Αποτελέσματα απόδοσης Decision Trees

Συνδυασμός Χαρακτηριστικών	Ακρίβεια Πρόβλεψης 65% εκπαίδευση	Ακρίβεια Πρόβλεψης 75% εκπαίδευση	Ακρίβεια Πρόβλεψης 85% εκπαίδευση
A	74,50%	75,57%	77,45%
B	42,25%	75,94%	78,15%
C	42,30%	83,15%	84,39%
D	74,81%	45,79%	45,96%
E	44,93%	45,12%	45,26%
F	83,45%	83,80%	84,70%
G	82,65%	42,59%	42,84%
H	82,93%	84,71%	85,48%
I	45,45%	42,50%	42,79%
J	85,35%	85,91%	87,27%

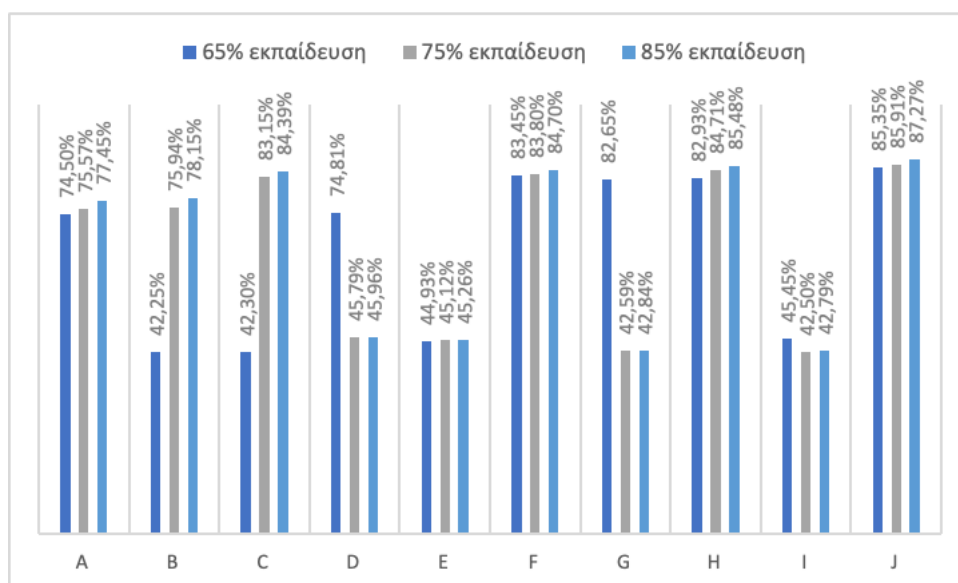


Figure 16 Αποτελέσματα απόδοσης Decision Trees

Table 9 Αποτελέσματα απόδοσης KNN

Συνδυασμός Χαρακτηριστικών	Ακρίβεια Πρόβλεψης 65% εκπαίδευση	Ακρίβεια Πρόβλεψης 75% εκπαίδευση	Ακρίβεια Πρόβλεψης 85% εκπαίδευση
A	34,19%	35,36%	36,14%
B	39,18%	35,38%	36,32%
C	40,08%	35,13%	35,70%
D	34,07%	31,81%	32,24%
E	30,64%	32,35%	32,35%
F	48,40%	52,96%	54,35%
G	51,65%	40,16%	40,45%
H	33,80%	50,36%	52,09%
I	30,81%	39,46%	38,50%
J	51,73%	53,34%	55,21%

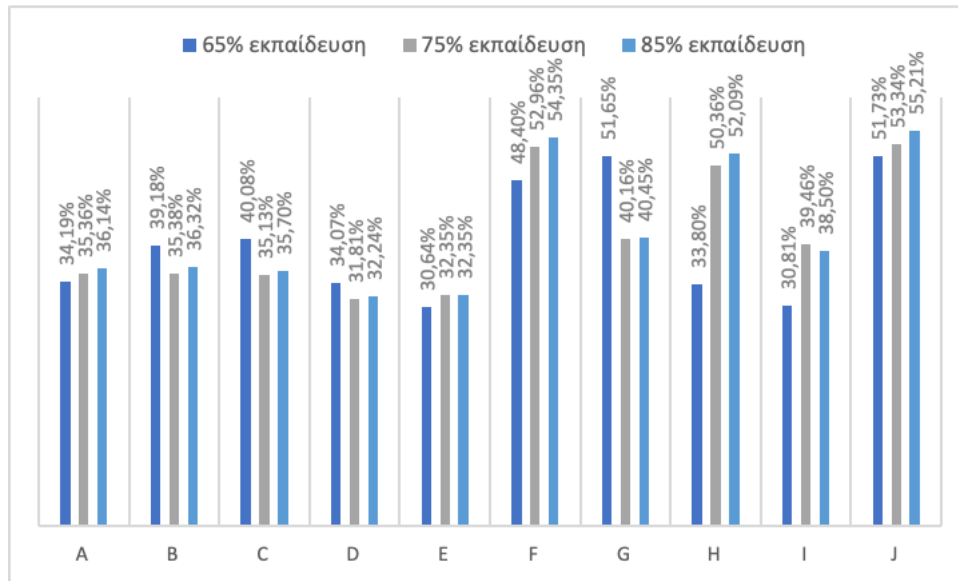


Figure 17 Αποτελέσματα απόδοσης KNN

Table 10 Αποτελέσματα απόδοσης Naive Bayes

Συνδυασμός Χαρακτηριστικών	Ακρίβεια Πρόβλεψης 65% εκπαίδευση	Ακρίβεια Πρόβλεψης 75% εκπαίδευση	Ακρίβεια Πρόβλεψης 85% εκπαίδευση
A	24,98%	25,66%	26,32%
B	7,26%	24,65%	25,28%
C	10,31%	14,28%	13,54%
D	24,51%	15,62%	15,33%
E	16,43%	16,35%	15,85%
F	15,14%	17,12%	16,97%
G	16,39%	9,14%	7,48%
H	13,78%	14,86%	15,12%
I	15,69%	6,61%	6,44%
J	14,52%	13,75%	14,24%

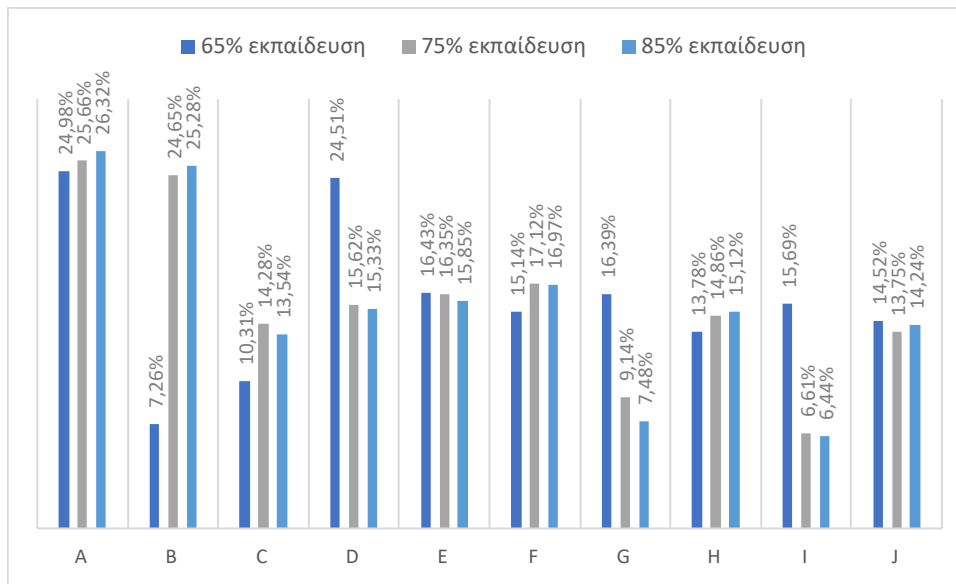


Figure 18 Αποτελέσματα απόδοσης Naive Bayes

Σχολιασμός Αποτελεσμάτων:

Σύμφωνα με τα εξαγόμενα αποτελέσματα, μπορούμε να παρατηρήσουμε τα εξής σημεία:

1. Η μέγιστη απόδοση παρατηρείται με την εφαρμογή των Δέντρων Αποφάσεων με είσοδο τα χαρακτηριστικά "date", "unique" και "FGNum" και για 85% συνόλου εκπαίδευσης.
2. Οι αλγόριθμοι των Δέντρων Αποφάσεων και των Τυχαίων Δασών αποδείχθηκαν εν γένει πιο αξιόπιστοι στα πλαίσια του στόχου πρόβλεψης.
3. Η αύξηση του ποσοστού των δεδομένων που αξιοποιείται για την εκπαίδευση των αλγορίθμων δεν επιδρά πάντα θετικά κατά τις πειραματικές δοκιμές. Πιο συγκεκριμένα, σε περίπτωση που δίνονται ως είσοδοι το "FGNum" και το "MatNum" το μικρότερο training set οδηγεί σε πιο ακριβή πρόβλεψη στα Δέντρα Αποφάσεων, στο k-nearest neighbor και στη Naive Bayes. Αυτό, θεωρούμε, ότι οφείλεται στο γεγονός ότι με την αύξηση των δεδομένων εκπαίδευσης επέρχεται αύξηση της τυχαιότητας σε περίπτωση των Δέντρων Αποφάσεων και της Naive Bayes, και μεγαλύτερο υπολογιστικό κόστος στην περίπτωση του k-nearest neighbor, οδηγώντας σε πιο αβέβαια αποτελέσματα.
4. Παρατηρούμε ότι η απόδοση του k-nearest neighbor είναι σχετικά χαμηλή, γεγονός που οφείλεται στη γενικότερη φύση του αλγορίθμου που τον καθιστά ακατάλληλο για μεγάλα σύνολα δεδομένων. Σε ένα σύνολο δεδομένων όπως αυτό που εφαρμόστηκε, το κόστος υπολογισμού της απόστασης μεταξύ του νέου σημείου και κάθε υπάρχοντος σημείου είναι μεγάλο, γεγονός που υποβαθμίζει την απόδοση του αλγορίθμου.

5. Ο k-nearest neighbor δεν λειτουργεί καλά με πολλά χαρακτηριστικά, καθώς καθίσταται δύσκολος ο υπολογισμός της απόστασης της κάθε διάστασης. Παρατηρούμε ότι όσο μειώνονται τα χαρακτηριστικά που δίνονται ως είσοδοι στον αλγόριθμο η απόδοσή του βελτιώνεται.

4.2.4 Υπολογισμός Ρίσκου Ποιότητας

Για τον υπολογισμό του Ποιοτικού Ρίσκου, όπως αυτό ορίστηκε στο Κεφάλαιο 3.4, αξιοποιήθηκε μέρος της διαθέσιμης πληροφορίας για τα δεδομένα. Πιο συγκεκριμένα, τα στοιχεία που συμμετείχαν στον καθορισμό και υπολογισμό του ρίσκου που συνοδεύει την εμφάνιση του κάθε ελαττώματος είναι τα εξής:

- Το FGNum, δηλαδή η μονάδα τήρησης αποθεμάτων (SKU -Stock keeping unit) / το μοναδικό αναγνωριστικό ενός προϊόντος.
- DefectID, δηλαδή το μοναδικό αναγνωριστικό ενός ελαττώματος.
- DefectSrcID, δηλαδή οι πηγές εμφάνισης ελαττωμάτων.
- MatNum, δηλαδή το μοναδικό αναγνωριστικό ενός εξαρτήματος.
- FamilyID, δηλαδή η ευρύτερη οικογένεια στην οποία ανήκει κάποιο εξάρτημα.

Υπολογισμός βάρους συχνότητας (F: Frequency Points):

Η συχνότητα εμφάνισης του κάθε ελαττώματος επί του συνόλου των εγγραφών αποτελεί αναμφισβήτητα καθοριστικό παράγοντα για την κρισιμότητα αυτού. Από τον παραγωγό για τον οποίο έγινε η μελέτη περίπτωσης είχε προταθεί η παρακάτω αντιστοίχιση μεταξύ του ποσοστού εμφάνισης του κάθε ελαττώματος επί του συνόλου των διαθέσιμων δεδομένων και του βάρους συχνότητας που θα συνυπολογιζόταν στην προαναφερθείσα φόρμουλα:

Table 11 Αντιστοίχιση Βάρους Συχνότητας – Προτεινόμενη

Συχνότητα %	Βάρος Συχνότητας
>80	5
>=61 και < 80	4
>=41 και < 60	3
>=21 και < 40	2
<= 20	1

Ωστόσο, τα πειραματικά αποτελέσματα έδειξαν πως, λόγω, του περιορισμένου αριθμού δεδομένων που είχαμε στη διάθεσή μας, το ποσοστό εμφάνισης του κάθε ελαττώματος ήταν κάτω από το 0.8%, με αποτέλεσμα το βάρος συχνότητας να ορίζεται σταθερά στο 1 και να μην προκύπτει σημαντικό αποτέλεσμα ως προς την αξιολόγηση του ρίσκου του εκάστοτε ελαττώματος. Ενδεικτικά:

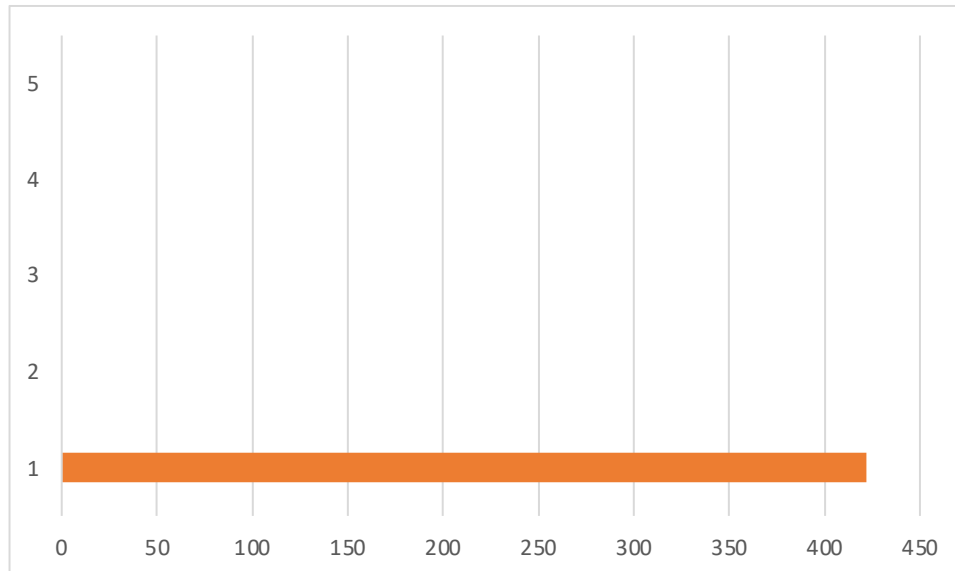


Figure 19 Κατανομή βάρους συχνότητας συναρτήσει του αριθμού ελαττωμάτων – Α

Ως εκ τούτου, λαμβάνοντας υπόψη τις τιμές που προέκυψαν από την πειραματική διαδικασία, προσαρμόσαμε την αντιστοίχιση του ποσοστού συχνότητας εμφάνισης του κάθε ελαττώματος με το αντίστοιχο βάρος συχνότητας όπως εμφανίζεται στον παρακάτω πίνακα:

Table 12 Αντιστοίχιση Βάρους Συχνότητας - Αναθεωρημένη

Συχνότητα %	Βάρος Συχνότητας
>0.80	5
>=0.61 και < 0.80	4
>=0.41 και < 0.60	3
>=0.21 και < 0.40	2
<= 0.20	1

Πράγματι, η αναθεωρημένη κλίμακα βελτίωσε την εξαγόμενη πληροφορία, καθώς εμφάνιζε την ακόλουθη κατανομή ανά τιμή βάρους, αποδεικνύοντας την αξία της αναλυτικής

διαδικασίας κατά τα αρχικά στάδια της μεθοδολογίας, καθώς κατέστησε δυνατή τη διόρθωση και βελτιστοποίηση του προτεινόμενου προτύπου.

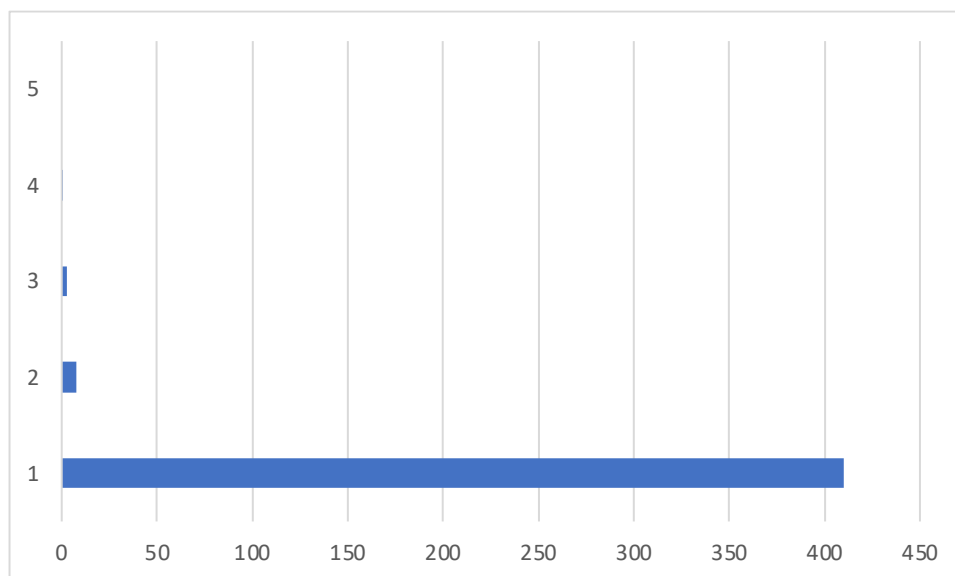


Figure 20 Κατανομή βάρους συχνότητας συναρτήσει του αριθμού ελαττωμάτων - B

Υπολογισμός βάρους σοβαρότητας (G: Severity Weight):

Σύμφωνα με τις διαδικασίες που έχουν οριστεί από τον παραγωγό για τον οποίο έγινε η μελέτη περίπτωσης, υπάρχουν καθορισμένα στάδια ποιοτικού ελέγχου στα οποία μπορεί να γίνει κάποια καταγραφή ενός ελαττωματικού. Η σοβαρότητα ποικίλει ανά στάδιο και, πιο συγκεκριμένα, ορίζεται από την παρακάτω αντιστοίχιση:

Table 13 Αντιστοίχιση Σοβαρότητας ανά Στάδιο Εντοπισμού Ελαττωματικού

Στάδιο Εντοπισμού Ελαττωματικού	Σοβαρότητα
Market	A, B, AC
Short Reliability	A
FDC Audit	A, AC
Zero Hour Test	A
Warehouse Audit	A, C
GPD Lab	A
Process Audit	A, B
Mfg Process	A, B, AC
Consumer Lab	A, B, C
Field Test	A

Δεδομένου ότι όλα τα ελαττώματα έχουν ταυτοποιηθεί κατά το Zero Hour Test, βάσει της κατηγοριοποίησης που ορίζεται παραπάνω έχουν Severity A, το οποίο αντιστοιχεί σε βάρος (weight) 4. Συνυπολογίσαμε, λοιπόν, αυτή την τιμή στα τελικά αποτελέσματα.

Table 14 Αντιστοίχιση Βάρους Σοβαρότητας

Σοβαρότητα	Βάρος Σοβαρότητας
AC - Safety	5
A Defect	4
B Defect	3
C Defect	2
Remark	1

Υπολογισμός βάρους πηγής αναγνώρισης (ΣD: Detection Points):

Σύμφωνα με τις διαδικασίες που έχουν οριστεί από τον παραγωγό για τον οποίο έγινε η μελέτη περίπτωσης, υπάρχουν καθορισμένα στάδια ποιοτικού ελέγχου στα οποία μπορεί να γίνει κάποια καταγραφή ενός ελαττωματικού. Το βάρος πηγής αναγνώρισης ποικίλει ανά στάδιο και, πιο συγκεκριμένα, ορίζεται από την παρακάτω αντιστοίχιση. Είναι προφανές ότι το βάρος πηγής αναγνώρισης αυξάνεται σε στάδια πιο κοντά στον τελικό καταναλωτή:

Table 15 Αντιστοίχιση Βάρους Αναγνώρισης

Στάδιο Εντοπισμού Ελαττωματικού	Βάρος Αναγνώρισης
Workstation	1
Primary Process	1
Aesthetical Control	3
Functional Test	3
Zero Hour Test	5
Quarantine	5
Q-Lab/Short Rel,	7
FDC.RDC Audit	10
1MIS – 12MIS	15
5 star	18
Trade Partner	20

Δεδομένου ότι όλα τα defects έχουν ταυτοποιηθεί κατά το Zero Hour Test, το βάρος της πηγής αναγνώρισής τους είναι το 5. Βάσει αυτού υπολογίζεται η στήλη Detection Points.

Επισημαίνεται ότι η στήλη Detection Points δεν συνυπολογίζεται στα τελικά αποτελέσματα, καθώς, λόγω της οικουμενικότητας της τιμής σε όλες τις εγγραφές δεν δίνει σημαντικά συμπεράσματα για το ρίσκο ποιότητας. Στα πλαίσια της μελέτης της συγκεκριμένης περίπτωσης, εισαγάγαμε ένα ακόμα κριτήριο ανάδειξης της κρισιμότητας ενός ελαττώματος, λαμβάνοντας υπόψη τον αριθμό των πηγών/αιτιών των ελαττωμάτων. Αξιοποιήσαμε, δηλαδή, το μέγεθος της λίστας στην στήλη DefectSrcID και βάει αυτού υπολογίσαμε τη στήλη Source Points, η οποία τελικά αντικαθιστά την πληροφορία των Detection Points.

Υπολογισμός Βάρους Επωνυμίας (B: Brand Points):

Σύμφωνα με τα δεδομένα που ήταν διαθέσιμα από τον παραγωγό για τον οποίο έγινε η μελέτη περίπτωσης δεν υπήρχε κάποιος διαχωρισμός ως προς την αξιολόγηση του κάθε ελαττωματικού βάσει τις επωνυμίας κάποιου προϊόντος ή επιμέρους εξαρτήματος. Ως εκ τούτου τέθηκε η τιμή 1 για όλους τους υπολογισμούς, ώστε να μην επηρεάζει ως κριτήριο το τελικό αποτέλεσμα.

Τελικά Αποτελέσματα:

Συνολικά τα αποτελέσματα που προέκυψαν από τους παραπάνω υπολογισμούς δίνονται παρακάτω στη στήλη του Priority Index, το οποίο καθορίζει σε φθίνουσα λογική το περισσότερο προς το λιγότερο κρισιμότερο ελάττωμα. Χάριν συντομίας, παρατίθενται οι 60 από τα 422 εγγραφές, όλες ανά μοναδικό αναγνωριστικό ελαττώματος:

Table 16 Τελικά Αποτελέσματα Ρίσκου Ποιότητας ανά Ελάττωμα

DefectID	Severity Weight	Frequency Percentage	Frequency Points	Detection Points	Source Points	Priority Index
1240	4	0,036953302	2	5	7	56
3035	4	0,075309893	4	5	3	48
1181	4	0,036173696	2	5	4	32
2045	4	0,005730101	1	5	6	24
1403	4	0,042644422	3	5	2	24

2217	4	0,000155921	1	5	6	24
3065	4	0,001598191	1	5	6	24
1280	4	0,000467763	1	5	6	24
1817	4	0,000467763	1	5	6	24
1287	4	0,000740625	1	5	6	24
1665	4	7,79606E-05	1	5	6	24
1355	4	0,000896546	1	5	6	24
1888	4	0,002923521	1	5	6	24
4027	4	0,001208389	1	5	6	24
1705	4	0,002572698	1	5	5	20
1627	4	0,002689639	1	5	5	20
1555	4	0,000974507	1	5	5	20
1867	4	0,003586185	1	5	5	20
1813	4	0,000155921	1	5	5	20
1873	4	0,001754112	1	5	5	20
1899	4	0,001013487	1	5	5	20
1393	4	3,89803E-05	1	5	5	20
2049	4	0,000506744	1	5	5	20
2533	4	0,001208389	1	5	5	20
3016	4	0,002026974	1	5	5	20
1663	4	7,79606E-05	1	5	5	20
1247	4	3,89803E-05	1	5	5	20
1575	4	3,89803E-05	1	5	5	20
1051	4	3,89803E-05	1	5	5	20
1777	4	3,89803E-05	1	5	4	16
1718	4	0,00144227	1	5	4	16
1615	4	0,000428783	1	5	4	16
1618	4	0,003235363	1	5	4	16
1617	4	0,000233882	1	5	4	16
3063	4	0,002455757	1	5	4	16
1446	4	0,000623684	1	5	4	16

1394	4	0,003703126	1	5	4	16
3015	4	0,006041943	1	5	4	16
1501	4	7,79606E-05	1	5	4	16
1095	4	0,000311842	1	5	4	16
3060	4	0,009121385	1	5	4	16
1295	4	0,000233882	1	5	4	16
1397	4	0,000116941	1	5	4	16
1009	4	0,00136431	1	5	4	16
1645	4	0,000389803	1	5	4	16
2028	4	0,000194901	1	5	4	16
1907	4	0,000428783	1	5	4	16
2124	4	0,000935527	1	5	4	16
1114	4	0,001793093	1	5	4	16
3085	4	0,001559211	1	5	4	16
2239	4	0,000155921	1	5	4	16
1913	4	3,89803E-05	1	5	4	16
5005	4	7,79606E-05	1	5	4	16
2607	4	3,89803E-05	1	5	4	16
3013	4	0,000545724	1	5	4	16
2229	4	3,89803E-05	1	5	4	16
2379	4	3,89803E-05	1	5	4	16
2113	4	3,89803E-05	1	5	4	16
1005	4	0,000116941	1	5	4	16
1889	4	0,000155921	1	5	4	16

Σχολιασμός Αποτελεσμάτων:

Από τα εξαγόμενα αποτελέσματα προκύπτουν τα παρακάτω ευρήματα:

- Η μικρότερη εμφανιζόμενη τιμή του Priority Index είναι η τιμή 4, ενώ η μεγαλύτερη η τιμή 56.
- Το 93,13% των ελαττωμάτων έχουν Priority Index κάτω του 20. Πιο συγκεκριμένα, το πλήθος των ελαττωμάτων ανά Priority Index δίνεται στο παρακάτω διάγραμμα:

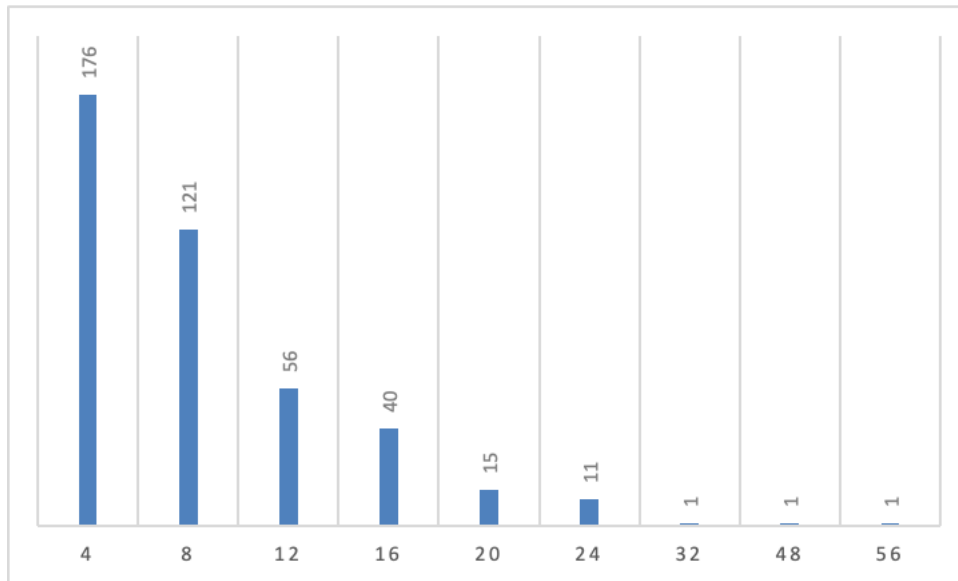


Figure 21 Πλήθος ελαττωμάτων ανά Priority Index

- Στο 6,87% των ελαττωμάτων με αυξημένο ρίσκο ποιότητας (με τιμή Priority Index από 20 έως 56) αντιστοιχούν συνολικά 29 ελαττώματα, εκ των οποίων ενδεικτικά αναφέρουμε ότι:
 - Τα 6 αντιστοιχούν σε προβλήματα που εμφανίζει η πόρτα του τελικού προϊόντος.
 - Τα 7 αντιστοιχούν στον πίνακα ελέγχου του τελικού προϊόντος.
 - Τα 2 αντιστοιχούν στα ηλεκτρολογικά εξαρτήματα του τελικού προϊόντος.
 - Τα 2 αντιστοιχούν στο καλούπι του τελικού προϊόντος.
- Η εύρεση διαδικασιών και μηχανισμών εκ μέρους του παραγωγού για την πρόληψη των 29 κρίσιμων ελαττωμάτων θα είχε οδηγήσει σε μείωση κατά 22,21% στις καταγραφές ελαττωμάτων. Συγκεκριμένα, τα 29 ελαττώματα σχετίζονται με 5.697 καταγραφές ελαττωματικών περιπτώσεων επί του συνόλου των 25.655 εγγραφών που είχαμε στη διάθεσή μας.

Γίνεται, επομένως, αντιληπτό πως ο προσδιορισμός του ρίσκου ποιότητας ανά ελάττωμα μπορεί να αξιοποιηθεί στις στρατηγικές αποφάσεις ενός παραγωγού για την επένδυση σε μεθόδους και διαδικασίες που είτε θα αποτρέπουν την εμφάνιση συγκεκριμένων προβλημάτων ποιότητας είτε θα ενισχύουν και θα διευκολύνουν την έγκυρη αντιμετώπισή τους.

5

Συμπεράσματα

Κατά την τελευταία δεκαετία παρατηρείται έντονη εξέλιξη και σημαντικές αλλαγές στις μηχανικές και κατασκευαστικές πρακτικές, τεχνολογίες και διαδικασίες των βιομηχανιών προηγμένων οικονομιών. Στα πλαίσια της Βιομηχανίας 4.0, όπως αναφέρεται συχνά η σύγχρονη αυτή τάση, η διαχείριση ποιότητας καλείται να ανταποκριθεί στις όλο και αυξανόμενες προκλήσεις. Η ποιότητα αποτελεί μια κρίσιμη διάσταση της γραμμής παραγωγής και μπορεί να δράσει ως ανταγωνιστικό πλεονέκτημα στον τομέα της κατασκευαστικής. Αποτελεί για τους παραγωγούς αδήριτη ανάγκη να προσαρμοστούν στις τεχνολογικές καινοτομίες, να εκσυγχρονίσουν την ανάλυση δεδομένων και να εντάξουν καινοτόμα μοντέλα ποιότητας ικανών να ανταποκριθούν στην αύξηση της πολυπλοκότητας και της ποικιλίας των προϊόντων που διαθέτουν στην αγορά. Με πολύτιμο αρωγό τις μεθοδολογίες και τους αλγόριθμους που προσφέρει η Μηχανική Μάθηση καθίσταται πλέον δυνατή η συλλογή δεδομένων, η δημιουργία δομών και μοντέλων προβλέψεων των ελαττωματικών προϊόντων και του προσδιορισμού το ρίσκου ποιότητας που αυτά εισάγουν στη παραγωγική διαδικασία. Ο εντοπισμός αυτός μπορεί να διευκολύνει τη λήψη αποφάσεων ως προς τους μηχανισμούς και τις κινήσεις αποτροπής της εμφάνισης των πιο επιδραστικών προβλημάτων, ή της ταχείας επίλυσής τους, ενισχύοντας το συνολικό σύστημα διαχείρισης ποιότητας του εκάστοτε παραγωγού και συντελώντας στην αύξηση της επιχειρηματικής αξίας της δραστηριοποίησής του στην αγορά.

Σε αυτή την κατεύθυνση κινήθηκε και η παρούσα εργασία, εξετάζοντας την ήδη υπάρχουσα βιβλιογραφία και επαληθεύοντας την εφαρμογή της αναλυτικής δεδομένων και μοντέλων μηχανικής μάθησης σε δεδομένα διαθέσιμα από πολυεθνικό παραγωγό και έμπορο οικιακών συσκευών με σκοπό την πρόβλεψη ελαττωμάτων και του καθορισμού του ρίσκου ποιότητας που αυτά εισάγουν στις διαδικασίες του. Πρωταρχική πρόκληση υπήρξε η συλλογή και επεξεργασία των κατάλληλων δεδομένων ικανών να οδηγήσουν στην εξαγωγή χρήσιμης

πληροφορίας που θα αξιοποιούσαμε ως ιστορικά δεδομένα για την εκπαίδευση των μοντέλων και την αύξηση της ακρίβειας της πρόβλεψης. Μετά την αρχική επεξεργασία των δεδομένων και τον καθορισμό των εισόδων των αλγορίθμων, έγινε δοκιμή και εφαρμογή τεσσάρων μοντέλων, των Τυχαίων Δασών, του Δέντρου Απόφασης, των K-πλησιέστερων γειτόνων και της Naïve Bayes, εκ των οποίων επικράτησε σε απόδοση το Δέντρο Απόφασης. Το επόμενο στάδιο της μεθοδολογίας όρισε την έννοια του “Priority Index”, το οποίο λειτουργεί ως δείκτης προσδιορισμού του ρίσκου ποιότητας ανά χαρακτηριστικό ελάττωμα. Κύριοι παράγοντες για τον υπολογισμό αυτού του δείκτη αποτελούν η συχνότητα εμφάνισης ενός ελαττώματος, το στάδιο και η πηγή εντοπισμού του. Τα πειραματικά αποτελέσματα έδειξαν πως με τον εντοπισμό των κρισιμότερων ελαττωμάτων (περί το 7%) και την προληπτική αντιμετώπιση αυτών μπορεί να μειωθεί σημαντικά η καταγραφή ελαττωματικών προϊόντων (περί το 22%), δρώντας θετικά στις διαδικασίες ποιότητας σε ένα οικοσύστημα παραγωγής.

Σημαντική αναγκαιότητα για βιομηχανικές εφαρμογές είναι, επομένως, η ταχύτητα στην εκμάθηση και ανάπτυξη των μοντέλων μηχανικής μάθησης και των συνολικών δυνατοτήτων της αναλυτικής δεδομένων, λόγω των συνεχώς μεταβαλλόμενων απαιτήσεων πελατών και τεχνικών απαιτήσεων στις μεταποιητικές βιομηχανίες. Τα παραδοσιακά εργαλεία διαχείρισης ποιότητας ήταν σε μεγάλο βαθμό επιτυχημένα κυρίως λόγω της επιτυχημένης εκπαίδευσης των ανθρώπων του κλάδου για τη χρήση των μεθόδων με τη βοήθεια προσιτού και φιλικού προς τον χρήστη λογισμικού. Για πιο αποτελεσματική και ευρεία χρήση των προσεγγίσεων μηχανικής μάθησης, πρέπει να αναπτυχθεί σχετικό λογισμικό με συγκρίσιμες ιδιότητες. Αυτό, με τη σειρά του, απαιτεί τη διαθεσιμότητα πιο ισχυρών, εύκολων στην εκμάθηση και υλοποίηση προσεγγίσεων εξόρυξης και διαχείρισης δεδομένων για προβλήματα βελτίωσης ποιότητας. Τέτοιο λογισμικό θα πρέπει επίσης να έχει τη δυνατότητα να βοηθά τους χρήστες να επιλέξουν τις καταλληλότερες μεθόδους για το πρόβλημα και να ερμηνεύουν τα αποτελέσματα που λαμβάνονται από τις εφαρμογές. Αυτή η μελέτη μπορεί να καθοδηγήσει ερευνητές και παραγωγούς λογισμικού στην προσπάθειά τους να αναπτύξουν ή/και να βελτιώσουν περαιτέρω τις μεθόδους και τα εργαλεία τους παρέχοντάς τους πληροφορίες σχετικά με τυπικά χαρακτηριστικά των δεδομένων βελτίωσης ποιότητας που συλλέγονται και τις απαραίτητες/πιο προτιμώμενες λειτουργίες και μεθόδους που μπορούν να εφαρμοστούν για την πρόβλεψη ποιότητας.

6

Βιβλιογραφία

Abhinav Maurya. Bayesian optimization for predicting rare internal failures in manufacturing processes. In 2016 IEEE International Conference on Big Data (Big Data), pages 2036–2045. IEEE, 2016.

Ahmad, M., Aftab, S. and Muhammad, S. S. (2017) ‘Machine Learning Techniques for Sentiment Analysis: A Review’, *International Journal of Multidisciplinary Sciences and Engineering*, 8(3), p. 27.

Angione, G., Cristalli, C., Barbosa, J., & Leitão, P. (2019, July). Integration Challenges for the Deployment of a Multi-Stage Zero-Defect Manufacturing Architecture. In 2019 IEEE 17th International Conference on Industrial Informatics (INDIN) (Vol. 1, pp. 1615-1620). IEEE.

Avigdor Zonnenshain & Ron S. Kenett (2020) Quality 4.0—the challenging future of quality engineering, *Quality Engineering*, 32:4, 614-626, DOI: 10.1080/08982112.2019.1706744

Bai, Y., Li, C., Sun, Z., & Chen, H. (2017). Deep neural network for manufacturing quality prediction. In 2017 Prognostics and System Health Management Conference (PHM-Harbin) (pp. 1-5). IEEE.

Bai, Y., Sun, Z., Deng, J., Li, L., Long, J., & Li, C. (2017). Manufacturing quality prediction using intelligent learning approaches: A comparative study. *Sustainability*, 10(1), 85.

Bai, Y., Sun, Z., Deng, J., Li, L., Long, J., & Li, C. (2018). Manufacturing quality prediction using intelligent learning approaches: A comparative study. *Sustainability*, 10(1), 85.

Bai, Y., Sun, Z., Zeng, B., Long, J., Li, L., de Oliveira, J. V., & Li, C. (2019). A comparison of dimension reduction techniques for support vector machine modeling of multi-parameter manufacturing quality prediction. *Journal of Intelligent Manufacturing*, 30(5), 2245-2256.

Berger, D., Zaiß, M., Lanza, G., Summa, J., Schwarz, M., Herrmann, H.G., Pohl, M., Günther, F. and Stommel, M., 2018. Predictive quality control of hybrid metal-CFRP components using information fusion. *Production Engineering*, 12(2), 161-172.

- Brady, J. E., & Allen, T. T. (2006). Six sigma literature: A review and agenda for future research. *Quality and Reliability Engineering International*, 22, 335–367.
- Bustillo, A., & Correa, M. (2012). Using artificial intelligence to predict surface roughness in deep drilling of steel components. *Journal of Intelligent Manufacturing*, 23(5), 1893-1902.
- Caoimhe M Carbery, Roger Woods, and Adele H Marshall. A bayesian network based learning system for modelling faults in large-scale manufacturing. In 2018 IEEE international conference on industrial technology (ICIT), pages 1357–1362. IEEE, 2018.
- Castelo-Branco, Isabel; Cruz-Jesus, Frederico; Oliveira, Tiago (2019). “Assessing Industry 4.0 readiness in manufacturing: Evidence for the European Union”. *Computers in Industry*, 107(), 22–32.
- Chaki, Nabendu; Devarakonda, Nagaraju; Sarkar, Anirban; Debnath, Narayan C. (2019). “Quality Control in the Context of Industry 4.0.” International Joint conference on Industrial Engineering and Operations Management XXIVIJCIEOM 2018: Industrial Engineering and Operations Management II.
- Chamkalani, A., Chamkalani, R., & Mohammadi, A. H. (2014). Hybrid of two heuristic optimizations with LSSVM to predict refractive index as asphaltene stability identifier. *Journal of dispersion science and technology*, 35(8), 1041-1050.
- Chatterjee, S., Mahapatra, S. S., Bharadwaj, V., Upadhyay, B. N., & Bindra, K. S. (2019). Prediction of quality characteristics of laser drilled holes using artificial intelligence techniques. *Engineering with Computers*, 1-24.
- Chen, W. C., Lee, A. H. I., Deng, W. J., & Liu, K. Y. (2007). The implementation of neural network for semiconductor PECVD process. *Expert Systems with Applications*, 32(4), 1148–1153.
- Chongwatpol, J. (2015). Prognostic analysis of defects in manufacturing. *Industrial Management & Data Systems*.
- Choudhary, A. K., Harding, J. A., & Tiwari, M. K. (2008). Data mining in manufacturing: A review based on the kind of knowledge. *Journal of Intelligent Manufacturing*, published online, ISSN:0956-5515.
- Dunham, M. H. (2003). *Data mining introductory and advanced topics*. New Jersey: Prentice Hall/Pearson Education.
- Eger, F., Reiff, C., Tempel, P., Magnanini, M. C., Caputo, D., Lechler, A., & Verl, A. (2020). Reaching Zero-Defect Manufacturing by Compensation of Dimensional Deviations in the Manufacturing of Rotating Hollow Parts. *Procedia Manufacturing*, 51, 388-393.

- Escobar, C. A., & Morales-Menendez, R. (2018). Machine learning techniques for quality control in high conformance manufacturing environment. *Advances in Mechanical Engineering*, 10(2), 1687814018755519.
- Escobar, C. A., Morales-Menendez, R., & Macias, D. (2020). Process-monitoring-for-quality—A machine learning-based modeling for rare event detection. *Array*, 7, 100034.
- «EUR-Lex - 52008DC0397:βιώσιμη κατανάλωση και παραγωγή και βιώσιμη βιομηχανική πολιτική». eur-lex.europa.eu. Ανακτήθηκε στις 21 Σεπτεμβρίου 2019.
- Fasser, Y., & Brettner, D. (2002). *Management for quality in high-technology enterprises*. New York: Wiley Interscience.
- Franciosa, P., Sokolov, M., Sinha, S., Sun, T., & Ceglarek, D. (2020). Deep learning enhanced digital twin for closed-loop in-process quality improvement. *CIRP Annals*, 69(1), 369-372.
- Frumosu, F. D., & Kulahci, M. (2018). Big data analytics using semi- supervised learning methods. *Quality and Reliability Engineering International*, 34(7), 1413-1423.
- Gashi, M., Ofner, P., Ennsbrunner, H., & Thalmann, S. (2021). Dealing with missing usage data in defect prediction: A case study of a welding supplier. *Computers in Industry*, 132, 103505. doi:10.1016/j.compind.2021.103 .
- Giudici, P. (2003). *Applied data mining: Statistical methods for business and industry*. New York: J. Wiley.
- Gülser Köksal; İnci Batmaz; Murat Caner Testik (2011). A review of data mining applications for quality improvement in manufacturing industry. , 38(10), 13448–13467. doi:10.1016/j.eswa.2011.04.063
- Hamzeh, R., Thomas, L., Polzer, J., Xu, X. W., & Heinzl, H. (2020). A Sensor Based Monitoring System for Real-Time Quality Control: Semi-Automatic Arc Welding Case Study. *Procedia Manufacturing*, 51, 201-206.
- Han, J., & Kamber, M. (2006). *Data mining: Concepts and techniques*. San Francisco: Morgan Kaufmann Publishers.
- Hao, L., Bian, L., Gebraeel, N., & Shi, J. (2016). Residual life prediction of multistage manufacturing processes with interaction between tool wear and product quality degradation. *IEEE Transactions on Automation Science and Engineering*, 14(2), 1211-1224.
- He, Y., Gu, C., Chen, Z., & Han, X. (2017). Integrated predictive maintenance strategy for manufacturing systems by combining quality control and mission reliability analysis. *International Journal of Production Research*, 55(19), 5841-5862.
- Hong Ge, Xinli Li, Yijiao Li, Gang Lu & Yong Yan (2019): Biomass fuel identification using flame spectroscopy and tree model algorithms, *Combustion Science and Technology*, DOI: 10.1080/00102202.2019.1680654 .

- Ilumoka, A. A. (1998). A modular neural network approach to microelectronic circuit yield optimization. *Microelectronics Reliability*, 38(4), 571–580.
- Ip, K. W., Kwong, C. K., & Wong, Y. W. (2003). Fuzzy regression approach to modeling transfer moulding for microchip encapsulation. *Journal of Materials Processing Technology*, 140(1–3), 147–151.
- Jin, Z., Zhang, Z., & Gu, G. X. (2020). Automated real-time detection and prediction of interlayer imperfections in additive manufacturing processes using artificial intelligence. *Advanced Intelligent Systems*, 2(1), 1900130.
- Jun, J. H., Chang, T. W., & Jun, S. (2020). Quality Prediction and Yield Improvement in Process Manufacturing Based on Data Analytics. *Processes*, 8(9), 1068.
- Kenett, R. S., and G. Shmueli. 2016. *Information quality: The potential of data and analytics to generate knowledge*. Chichester, UK: John Wiley and Sons.
- Kenett, R. S., Swarz, R. S., & Zonnenshain, A. (Eds.). (2019). *Systems engineering in the fourth industrial revolution: Big data, novel technologies, and modern systems engineering*. John Wiley & Sons.
- Kim, D., Kang, P., Cho, S., Lee, H. J., & Doh, S. (2012). Machine learning-based novelty detection for faulty wafer detection in semiconductor manufacturing. *Expert Systems with Applications*, 39(4), 4075-4083.
- Kim, S. H., & Lee, C. M. (1997). Nonlinear prediction of manufacturing systems through explicit and implicit data mining. *Computer and Industrial Engineering*, 33(3-4), 461–464.
- Kim, S., & Ryu, K. (2020). Intelligent Process Quality Management for Supporting Collaboration of Mold Manufacturing SMEs. *Procedia Manufacturing*, 51, 381-387.
- Kolarik, W. J. (1995). *Creating quality: Concepts, systems, strategies, and tools*. New York: McGraw-Hill.
- Konrad, B., Lieber, D., & Deuse, J. (2013). Striving for zero defect production: intelligent manufacturing control through data mining in continuous rolling mill processes. In *Robust Manufacturing Control* (pp. 215-229). Springer, Berlin, Heidelberg.
- Last, M., & Kandel, A. (2001). Data mining for process and quality control in the semiconductor industry. In D. Braha (Ed.), *Data mining for design and manufacturing* (pp. 207–234). Dordrecht: Kluwer Academic Publishers.
- Lee, C. K. H., Choy, K. L., Ho, G. T., & Lam, C. H. (2016). A slippery genetic algorithm-based process mining system for achieving better quality assurance in the garment industry. *Expert systems with applications*, 46, 236-248.

- Lee, J. H., Shin, J., & Realff, M. J. (2018b). Machine learning: Overview of the recent progresses and implications for the process systems engineering field. *Computers & Chemical Engineering*, 114, 111-121.
- Lee, J., Noh, S. D., Kim, H. J., & Kang, Y. S. (2018). Implementation of cyber-physical production systems for quality prediction and operation control in metal casting. *Sensors*, 18(5), 1428.
- Li Da Xu, Eric L. Xu & Ling Li (2018) Industry 4.0: state of the art and future trends, *International Journal of Production Research*, 56:8, 2941-2962.
- Li, D. C., Chen, W. C., Liu, C. W., & Lin, Y. S. (2012). A non-linear quality improvement model using SVR for manufacturing TFT-LCDs. *Journal of Intelligent Manufacturing*, 23(3), 835-844.
- Li, T. S., Huang, C. L., & Wu, Z. Y. (2006). Data mining using genetic programming for construction of a semiconductor manufacturing yield rate prediction system. *Journal of Intelligent Manufacturing*, 17, 355–361.
- Li, T. S., Su, C. T., & Chiang, T. L. (2003a). Applying robust multi-response quality engineering for parameter selection using a novel neural-genetic algorithm. *Computers in Industry*, 50(1), 113–122.
- Lieber, D., Stolpe, M., Konrad, B., Deuse, J., & Morik, K. (2013). Quality prediction in interlinked manufacturing processes based on supervised & unsupervised machine learning. *Procedia Cirp*, 7, 193-198.
- Lindström, J., Kyösti, P., Birk, W., & Lejon, E. (2020). An Initial Model for Zero Defect Manufacturing. *Applied Sciences*, 10(13), 4570.
- Liu, G., Gao, X., You, D., & Zhang, N. (2019b). Prediction of high power laser welding status based on PCA and SVM classification of multiple sensors. *Journal of Intelligent Manufacturing*, 30(2), 821-832.
- Liu, Y., Zhou, H., Tsung, F., & Zhang, S. (2019). Real-time quality monitoring and diagnosis for manufacturing process profiles based on deep belief networks. *Computers & Industrial Engineering*, 136, 494-503.
- Lokrantz, A., Gustavsson, E., & Jirstrand, M. (2018). Root cause analysis of failures and quality deviations in manufacturing using machine learning. *Procedia CIRP*, 72, 1057-1062.
- Luckow, A., Kennedy, K., Ziolkowski, M., Djerekarov, E., Cook, M., Duffy, E., ... & Smith, M. C. (2018). Artificial intelligence and deep learning applications for automotive manufacturing. In *2018 IEEE International Conference on Big Data (Big Data)* (pp. 3144-3152). IEEE.

- Martinez, W. L., & Martinez, A. R. (2002). *Computational statistics handbook with MATLAB*. Boca Raton: Chapman and Hall.
- Mehryar Mohri, Afshin Rostamizadeh, Ameet Talwalkar (2012) *Foundations of Machine Learning*, The MIT Press ISBN 9780262018258.
- Mieno, F., Sato, T., Slubnya, Y., Odagiri, K., Tsuda, H., & Take, K. (1999). Yield improvement using data mining system. In *Semiconductor manufacturing conference proceedings, 1999 IEEE international symposium on*, 11–13 October Santa Clara CA (pp. 391–394). IEEE.
- Montgomery, D. (2005). *Introduction to statistical quality control*. Hoboken, NJ: John Wiley.
- Nalbach, O., Linn, C., Derouet, M., & Werth, D. (2018). Predictive quality: Towards a new understanding of quality assurance using machine learning tools. In *International Conference on Business Information Systems* (pp. 30-42). Springer, Cham.
- Oliff, H., & Liu, Y. (2017). Towards industry 4.0 utilizing data-mining techniques: a case study on quality improvement. *Procedia CIRP*, 63, 167-172.
- Paul, S. K. (2016). Prediction of complete forming limit diagram from tensile properties of various steel sheets by a nonlinear regression based approach. *Journal of Manufacturing Processes*, 23, 192-200.
- Pavlyshenko, B. (2016). Machine learning, linear and bayesian models for logistic regression in failure detection problems. In *2016 IEEE international conference on Big Data (Big Data)*, 2016 (pp. 2046– 2050). IEEE. <https://doi.org/10.1109/bigdata.2016.7840828>.
- Peres, R., Rocha, A. D., Matos, J. P., & Barata, J. (2018). Go0dman data model- interoperability in multistage zero defect manufacturing. In *2018 IEEE 16th international conference on industrial informatics (indin)* (pp. 815-821). IEEE.
- Phadke, M. S. (1989). *Quality engineering using robust design*. Englewood Cliffs, NJ: Prentice-Hall.
- Psarommatis Giannakopoulos, F., Prouvost, S., May, G., & Kyritsis, D. (2020b). Product quality improvement policies in Industry 4.0: characteristics, enabling factors, barriers, and evolution toward Zero Defect Manufacturing. *Frontiers in Computer Science*, 2(ARTICLE).
- Psarommatis, F., & Kiritsis, D. (2018). A scheduling tool for achieving zero defect manufacturing (ZDM): a conceptual framework. In *IFIP International Conference on Advances in Production Management Systems* (pp. 271-278). Springer, Cham.
- Psarommatis, F., May, G., Dreyfus, P. A., & Kiritsis, D. (2020a). Zero defect manufacturing: state-of-the-art review, shortcomings and future directions in research. *International Journal of Production Research*, 58(1), 1-17.
- Pyle, D. (1999). *Data preparation for data mining*. San Francisco, CA: Morgan Kaufmann Publishers.

- Quinlan, J. and Quinlan J. R. (1996) 'Learning decision tree classifiers', *ACM Computing Surveys (CSUR)*, 28(1), pp. 2–3. Available at: <http://dl.acm.org/citation.cfm?id=234346>.
- Reis, M., and R. S. Kenett. 2017. A structured overview on the use of computational simulators for teaching statistical methods. *Quality Engineering* 29 (4):730–44. doi:10.1080/08982112.2016.1272122.
- Roblek, V., M. Mesko, and A. Krapez. 2016. "A Complex View of Industry 4.0." *SAGE Open* April-June 2016: 1–11.
- Sanchez-Marquez, Rafael; Albarracín Guillem, José Miguel; Vicens-Salort, Eduardo; Jabaloyes Vivas, José (2020). Diagnosis of quality management systems using data analytics – A case study in the manufacturing sector. *Computers in Industry*, 115(), 103183. doi:10.1016/j.compind.2019.103183 .
- Schmitt, J., Böning, J., Borggräfe, T., Beiting, G., & Deuse, J. (2020). Predictive model-based quality inspection using Machine Learning and Edge Cloud Computing. *Advanced Engineering Informatics*, 45, 101101.
- Schnell, J., Nentwich, C., Endres, F., Kollenda, A., Distel, F., Knoche, T., & Reinhart, G. (2019). Data mining in lithium-ion battery cell production. *Journal of Power Sources*, 413, 360-366.
- Schreiber, M., Klöber-Koch, J., Bömelburg-Zacharias, J., Braunreuther, S., & Reinhart, G. (2019). Automated quality assurance as an intelligent cloud service using machine learning. *Procedia CIRP*, 86, 185-191.
- Shi, X., Schillings, P., & Boyd, D. (2004). Applying artificial neural networks and virtual experimental design to quality improvement of two industrial processes. *International Journal of Production Research*, 42(1), 101–118.
- Skinner, K. R., Montgomery, D. C., Runger, G. C., Fowler, J. W., McCarville, D. R., Rhoads, T. R., et al. (2002). Multivariate statistical methods for modeling and analysis of wafer probe test data. *IEEE Transactions on Semiconductor Manufacturing*, 15(4), 523–530.
- Sony, M. (2018). Industry 4.0 and lean management: a proposed integration model and research propositions. *Production & Manufacturing Research*, 6(1), 416-432.
- Sun, H., Yang, J., & Wang, L. (2017). Resistance spot welding quality identification with particle swarm optimization and a kernel extreme learning machine model. *The International Journal of Advanced Manufacturing Technology*, 91(5), 1879-1887.
- Taguchi, G., Chowdhury, S., & Taguchi, S. (2000). *Robust engineering*. New York: McGraw-Hill.
- Teti, R. (2015). Advanced IT methods of signal processing and decision making for zero defect manufacturing in machining. *Procedia CIRP*, 28, 3-15.

- Teucke, M., Broda, E., Boerold, A., & Freitag, M. (2018). Using sensor-based quality data in automotive supply chains. *Machines*, 6(4), 53.
- Tsuda, H., Shirai, H., Takagi, O., Take, R. (2000). Yield analysis and improvement by reducing manufacturing fluctuation noise. In *Proceedings of International Symposium on Semiconductor Manufacturing (ISSM 2000)*, 26-28 September Tokyo. 249-252.
- Wang, G., Ledwoch, A., Hasani, R. M., Grosu, R., & Brintrup, A. (2019). A generative neural network model for the quality prediction of work in progress products. *Applied Soft Computing*, 85, 105683.
- Wang, K. S. (2013). Towards zero-defect manufacturing (ZDM)—a data mining approach. *Advances in Manufacturing*, 1(1), 62-74.
- Witten, I. H. (2005). *Data mining: Practical machine learning tools and techniques*. Boston, MA: Morgan Kaufman.
- Wuest, T., Irgens, C., & Thoben, K. D. (2014). An approach to monitoring quality in manufacturing using supervised machine learning on product state data. *Journal of Intelligent Manufacturing*, 25(5), 1167-1180.
- Yang, T., Tsai, T., & Yeh, J. (2005). A neural network-based prediction model for fine pitch stencil printing quality in surface mount assembly. *Engineering Applications of Artificial Intelligence*, 18(3), 335–341.
- Zonnenshain, A., & Kenett, R. S. (2020). Quality 4.0—the challenging future of quality engineering. *Quality Engineering*, 32(4), 614-626.