



**ΠΑΝΕΠΙΣΤΗΜΙΟ
ΔΥΤΙΚΗΣ ΑΤΤΙΚΗΣ**

Τμήμα Μηχανικών Βιομηχανικής
Σχεδίασης και Παραγωγής

&

**ΠΑΝΕΠΙΣΤΗΜΙΟ
ΑΙΓΑΙΟΥ**

Τμήμα Ναυτιλίας και
Επιχειρηματικών Υπηρεσιών



**ΔΙΔΡΥΜΑΤΙΚΟ
ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ
«ΝΕΕΣ ΤΕΧΝΟΛΟΓΙΕΣ ΣΤΗ ΝΑΥΤΙΛΙΑ ΚΑΙ ΤΙΣ ΜΕΤΑΦΟΡΕΣ»**

ΤΙΤΛΟΣ

*Πρόβλεψη λειτουργικών παραμέτρων πλοίου με χρήση τεχνικών
τεχνητής νοημοσύνης.*

ΤΙΤΛΟΣ ΑΓΓΛΙΚΑ

*Prediction of a ship's operational parameters using artificial
intelligence technics.*

Όνοματεπώνυμο Σπουδαστή:

Αλεξίου Κυριάκος

Όνοματεπώνυμο Υπεύθυνου Καθηγητή:

Ελένη Αικατερίνη Δελίγκου

Ευθύμιος Παριώτης

ΔΙΑΤΡΙΒΗ

Φεβρουάριος 2021



**ΠΑΝΕΠΙΣΤΗΜΙΟ
ΔΥΤΙΚΗΣ ΑΤΤΙΚΗΣ**

Τμήμα Μηχανικών Βιομηχανικής
Σχεδίασης και Παραγωγής

&

**ΠΑΝΕΠΙΣΤΗΜΙΟ
ΑΙΓΑΙΟΥ**

Τμήμα Ναυτιλίας και
Επιχειρηματικών Υπηρεσιών



Μέλη Εξεταστικής Επιτροπής

Λελίγκου Ελένη - Αικατερίνη

Παπουτσιδάκης Μιχαήλ

Δρόσος Χρήστος

ΔΗΛΩΣΗ ΣΥΓΓΡΑΦΕΑ ΜΕΤΑΠΤΥΧΙΑΚΗΣ ΕΡΓΑΣΙΑΣ

Ο κάτωθι υπογεγραμμένος Αλεξίου Κυριάκος του Δημητρίου, με αριθμό μητρώου 8056121 φοιτητής του Διδρυματικού Προγράμματος Μεταπτυχιακών Σπουδών «Νέες Τεχνολογίες στη Ναυτιλία και τις Μεταφορές» του Τμήματος Μηχανικών Βιομηχανικής Σχεδίασης και Παραγωγής της Σχολής Μηχανικών Πανεπιστημίου Δυτικής Αττικής, δηλώνω υπεύθυνα ότι: «Είμαι συγγραφέας αυτής της μεταπτυχιακής εργασίας και ότι κάθε βοήθεια την οποία είχα για την προετοιμασία της είναι πλήρως αναγνωρισμένη και αναφέρεται στην εργασία. Επίσης, οι όποιες πηγές από τις οποίες έκανα χρήση δεδομένων, ιδεών ή λέξεων, είτε ακριβώς είτε παραφρασμένες, αναφέρονται στο σύνολό τους, με πλήρη αναφορά στους συγγραφείς, τον εκδοτικό οίκο ή το περιοδικό, συμπεριλαμβανομένων και των πηγών που ενδεχομένως χρησιμοποιήθηκαν από το διαδίκτυο. Επίσης, βεβαιώνω ότι αυτή η εργασία έχει συγγραφεί από μένα αποκλειστικά και αποτελεί προϊόν πνευματικής ιδιοκτησίας τόσο δικής μου, όσο και του Ιδρύματος. Παράβαση της ανωτέρω ακαδημαϊκής μου ευθύνης αποτελεί ουσιώδη λόγο για την ανάκληση του διπλώματός μου».

Ο δηλών

Ημερομηνία

Alexiou

14/02/2021

Kiriakos

Digitally signed
by Alexiou Kiriakos
date 2021.02.14



ΤΙΤΛΟΣ

*Πρόβλεψη λειτουργικών παραμέτρων πλοίου με χρήση τεχνικών
τεχνητής νοημοσύνης*

ΟΝΟΜΑ ΦΟΙΤΗΤΗ

Αλεξίου Κυριάκος

**Μεταπτυχιακή Διατριβή που υποβάλλεται στο καθηγητικό σώμα για την μερική
εκπλήρωση των υποχρεώσεων απόκτησης του μεταπτυχιακού τίτλου του
Διδρυματικού Προγράμματος Μεταπτυχιακών Σπουδών «Νέες Τεχνολογίες
στη Ναυτιλία και τις Μεταφορές» του Τμήματος Ναυτιλίας και
Επιχειρηματικών Υπηρεσιών του Πανεπιστημίου Αιγαίου και του Τμήματος
Μηχανικών Βιομηχανικής Σχεδίασης και Παραγωγής του Πανεπιστημίου
Δυτικής Αττικής.**

Ευχαριστίες

Σε αυτό το σημείο θα ήθελα να εκφράσω τις ειλικρινείς μου ευχαριστίες στους καθηγητές μου, Ελένη-Αικατερίνη Δελίγκου και Ευθύμιο Παριώτη που συνέβαλαν καθοριστικά στην επιτυχή ολοκλήρωση της διπλωματικής διατριβής μου. Η εμπιστοσύνη που επέδειξαν στο πρόσωπό μου, η συνεχής παρότρυνση και καθοδήγησή τους καθώς και οι χρήσιμες συμβουλές τους υπήρξαν τα πολυτιμότερα βοηθήματα σε αυτή μου την προσπάθεια. Πάνω από όλους όμως θα ήθελα να εκφράσω τις ευχαριστίες μου στην οικογένεια μου και κυρίως στην σύζυγό μου, αρχικά για την προτροπή και ενθάρρυνση στην απόφαση μου να συνεχίσω τις σπουδές μου και μετέπειτα για την υπομονή που επέδειξε τα χρόνια που απαιτήθηκαν για την ολοκλήρωσή τους.

Περίληψη

Σε αυτή τη διπλωματική διατριβή ο συγγραφέας αξιολογεί και συγκρίνει τη χρήση αλγορίθμων πολλαπλής παλινδρόμησης όπως τα Τεχνητά Νευρωνικά Δίκτυα, τα Δέντρα Απόφασης, τα Δάση Τυχαίας Απόφασης, τους k-Πλησιέστερους Γείτονες (kNN), τη Γραμμική Παλινδρόμηση και του AdaBoost στην πρόβλεψη της απαιτούμενης ισχύος της κύριας μηχανής πλοίου. Αυτή η προσέγγιση βασίζεται στην εξόρυξη γνώσης με την χρήση δεδομένων που αντιστοιχούν σε πραγματικές τιμές μεγεθών που συλλέχθηκαν από το σύστημα αυτόματης καταγραφής και παρακολούθησης δεδομένων (ADLM) ενός πλοίου σε περίοδο έξι μηνών. Συγκρίνονται τα αποτελέσματα πρόβλεψης των ανωτέρω αλγορίθμων με χρήση δύο ξεχωριστών συνόλων δεδομένων: το αρχικό πλήρες σύνολο δεδομένων καθώς και ένα δεύτερο το οποίο προκύπτει από την επεξεργασία του πρωτοτύπου βάση συγκεκριμένης μεθοδολογίας, με τη χρήση ενός συνδυασμού στατιστικών μεθόδων καθώς και τη γνώση των φυσικών αρχών που διέπουν το υπό εξέταση φυσικό μέγεθος. Η μελέτη αποδεικνύει ότι με τη χρήση της Τεχνητής Νοημοσύνης (μέσω μηχανικής μάθησης) και της κατάλληλης προ επεξεργασίας των δεδομένων μπορούμε να επιτύχουμε αποτελέσματα πρόβλεψης με ποσοστό σφάλματος κάτω του 3% στο 99,31% των υπό πρόβλεψη τιμών.

Λέξεις κλειδιά: Τεχνητή Νοημοσύνη, Κατανάλωση Καυσίμου, Πρόβλεψη, Αλγόριθμοι παλινδρόμησης, Εξόρυξη Δεδομένων, Μηχανική Μάθηση

Abstract

In this paper the author tests and compares the use of multiple regression algorithms like Artificial Neural Network (ANN), Decision Tree Regressor (DTRs), Random Forest Regressor (RFR), K-Nearest Neighbor (kNN), Linear Regression and AdaBoost, on the prediction of ships main engine required power. This data driven approach is based on real values that are collected from a ship's onboard Automated Data Logging & Monitoring (ADLM) system in a period of six months. Two separated set of Algorithm benchmark tests are carried out and compared: one with the use of the above complete data set and the second with the use of a preprocessed one that is constructed from the original, with the use of a combination of statistical inference methods as well as the knowledge of the physical principles. The study shows that with the use of Artificial intelligence (through Machine Learning) and proper data preprocessing we can achieve prediction results that are below an 3% error percentage in 99.31% of the respective values.

Keywords: Artificial intelligence (AI), Fuel Consumption (FOC), Prediction, Regression Algorithms, Data Mining, Machine learning (ML).



Περιεχόμενα

Ευχαριστίες.....	5
Περίληψη.....	6
Abstract.....	7
1. ΕΙΣΑΓΩΓΗ.....	10
1.1 Σημαντικότητα Αντικείμενου και διατύπωση ερευνητικού προβλήματος.....	10
1.2 Τεχνητή Νοημοσύνη / Μηχανική Μάθηση.....	11
1.3 Σκοπός και Στόχοι Εργασίας.....	12
2. ΥΠΟΒΑΘΡΟ.....	13
2.1 Ιστορική Αναδρομή.....	13
2.2 Ανάπτυξη Θεωρητικού Πλαισίου.....	15
2.2.1 Λογικοί ή ευφυείς παράγοντες.....	15
2.2.2 Τύποι Μηχανικής Μάθησης.....	20
2.2.3 Διάγραμμα Ροής / Στάδια Εκτέλεσης	22
2.3 Ανασκόπηση βιβλιογραφίας	30
2.4 Υπόθεση / Ερευνητικά Ερωτήματα	32
3. ΑΝΑΛΥΣΗ.....	35
3.1 Αλγόριθμοι Παλινδρόμησης.....	35



3.2 Μέτρηση Επιδόσεων.....	49
4. ΜΕΘΟΔΟΛΟΓΙΑ / ΥΛΟΠΟΙΗΣΗ.....	52
4.1 Μεταβλητές Έρευνας / Πλαίσιο Δεδομένων.....	52
4.2 Προετοιμασία Δεδομένων / προ-επεξεργασία.....	58
4.3 Εξόρυξη Γνώσης.....	69
4.4 Αποτελέσματα προβλέψεων / μέτρηση επιδόσεων.....	72
5.ΣΥΜΠΕΡΑΣΜΑ.....	79
5.1 Συμπεράσματα.....	79
5.2 Περιορισμοί.....	80
5.3 Μελλοντική Έρευνα.....	81
6. ΒΙΒΛΙΟΓΡΑΦΙΑ.....	82

1. ΕΙΣΑΓΩΓΗ

1.1 Σημαντικότητα Αντικείμενου και διατύπωση ερευνητικού προβλήματος

Η πρόβλεψη μιας κρίσιμης λειτουργικής παραμέτρου ενός πλοίου όπως η κατανάλωση καυσίμου ή η αποδιδόμενη ισχύς της κύριας μηχανής μπορεί να οδηγήσει σε σημαντική αύξηση του προσδόκιμου ζωής καθώς και τη διαθεσιμότητα του. Η σύγκριση των προγνωστικών τιμών με τις τιμές αναφοράς (π.χ. θαλάσσιες δοκιμές) οδηγεί σε ασφαλή συμπεράσματα της λειτουργικής του κατάστασης. Οι φθορές των κρίσιμων μηχανικών μερών συνήθως επέρχονται προοδευτικά. Η έγκαιρη ανίχνευση δυσλειτουργίας αποτρέπει περαιτέρω επιδείνωση. Εάν μπορούμε να προβλέψουμε με ακρίβεια τη μείωση της απόδοσης ενός συγκεκριμένου στοιχείου, τότε θα είμαστε σε θέση να προγραμματίσουμε τις επισκευές με τον πιο παραγωγικό τρόπο. Από την άλλη πλευρά, η πρόβλεψη του ποσοστού αύξησης της κατανάλωσης καυσίμου λόγω π.χ. της αυξανόμενης βιολογικής ρύπανσης του κύτους των πλοίων μπορεί να βοηθήσει τους πλοιοκτήτες να επιλέξουν τον κατάλληλο χρόνο και μέρος για την εκτέλεση ενός δεξαμενισμού επιτυγχάνοντας τη μέγιστη κερδοφορία τους.

Ένα σημαντικό τμήμα του κύκλου εργασιών της σύγχρονης ναυτιλιακής βιομηχανίας αποτελεί η έρευνα και ανάπτυξη νέων εξελιγμένων πλοίων. Εκατομμύρια δολάρια δαπανούνται για τον σχεδιασμό νέων σκαφών με σκοπό την μεγιστοποίηση του κέρδους (μέσω μείωσης λειτουργικών εξόδων) και την ελαχιστοποίηση της περιβαλλοντικής ρύπανσης (Andrea Coraddua, 220). Η εύρεση μιας σχετικά φθηνής μεθόδου σχεδιασμού και υλοποίησης αποτελεί διακαή πόθο της βιομηχανίας.

Η ναυτιλία είναι ένας τομέας που βασίζεται σε παραδοσιακές προσεγγίσεις. Οι επιστημονικές γνώσεις και προπαντός η εμπειρία των στελεχών είναι ο πρωταρχικός παράγοντας διαμόρφωσης των επιλογών. Η εισχώρηση της τεχνολογίας των υπολογιστών αν και αναπόφευκτη έχει περιοριστεί κυρίως στην βελτίωση της ποιότητας καθώς και το πλήθος των δεδομένων που παρέχονται στον ανθρώπινο

παράγοντα και όχι τόσο στην αξιολόγηση και παροχή γνώσης από τα ίδια τα δεδομένα. Ο "παραδοσιακός" λοιπόν αυτός τρόπος χρήσης της τεχνολογίας των υπολογιστών λειτουργεί συνεπικουρικά με άλλα επιστημονικά πεδία (ναυπηγεία, μηχανική) επιχειρώντας τη συνεχή βελτίωση είτε της χρήσης είτε του σχεδιασμού των πλοίων. Η προσέγγιση αυτή οδηγεί σε πολύπλοκες και δαπανηρές λύσεις. Η δημιουργία πρωτοτύπων και η εξέλιξη τους στο τμήμα του Research and Development (RND) απαιτεί τον έλεγχο και κατασκευή πρωτοτύπων ή την χρήση πολύπλοκου και ακριβού λογισμικού που χειρίζεται προσωπικό με εξειδικευμένες γνώσεις. Αντίστοιχη είναι και η κατάσταση όσο αναφορά την παρακολούθηση των λειτουργικών παραμέτρων ενός πλοίου κατά την διάρκεια της επιχειρησιακής του ζωής. Η μετάπτωση της πολιτικής συντήρησης σε μοντέλα Condition Based Maintenance (CBM) απαιτεί εξειδικευμένα εργαλεία και γνώση της εγκατάστασης σε επίπεδο κατασκευαστή.

Στην σύγχρονη ναυτιλία, την εποχή της τέταρτης βιομηχανικής επανάστασης με την ευρεία χρήση της διασυνδεσιμότητας ίσως είναι εφικτό να χρησιμοποιηθεί η επιστήμη των υπολογιστών με μια διαφορετική προσέγγιση. Η εξόρυξη γνώσης που μπορεί να παραχθεί με την χρήση μεθόδων τεχνητής νοημοσύνης από τον τεράστιο όγκο δεδομένων που συλλέγονται από τα συστήματα παραμετρικής παρακολούθησης της πλατφόρμας των πλοίων και, ίσως να είναι μια φθηνότερη και αποδοτικότερη εναλλακτική στις ανωτέρω προκλήσεις.

1.2 Τεχνητή Νοημοσύνη / Μηχανική Μάθηση

Η τεχνητή νοημοσύνη, αποτελεί από μόνη της μία επανάσταση, ίσως όχι τόσο τεχνολογικά όσο εννοιολογικά. Αποτελεί την σύμπραξη της επιστήμης (Ron Kohavi, 1998) και της μηχανικής με σκοπό να κάνουν τους υπολογιστές να συμπεριφέρονται με τρόπους που, μέχρι πρόσφατα, πιστεύαμε ότι μόνο η ανθρώπινη νοημοσύνη μπορούσε. Πιο συγκεκριμένα μέσω της τεχνητής νοημοσύνης μπορούμε να καταστήσουμε τις μηχανές ικανές να 'αντιλαμβάνονται' το περιβάλλον τους, να προσομοιάζουν τον τρόπο ανθρώπινης επίλυσης προβλημάτων και να δρουν προς την

επίτευξη ενός συγκεκριμένου στόχου. Ο υπολογιστής λαμβάνει δεδομένα (έτοιμα ή συλλεγμένα μέσω αισθητήρων), τα επεξεργάζεται και αντιδρά βάσει αυτών. Τα συστήματα τεχνητής νοημοσύνης είναι ικανά να προσαρμόζουν τη συμπεριφορά τους, σε ένα ορισμένα βαθμό, αναλύοντας τις συνέπειες προηγούμενων δράσεων και επιλύοντας προβλήματα με αυτονομία. (Τι είναι η τεχνητή νοημοσύνη και πώς χρησιμοποιείται, 2020).

Η μηχανική μάθηση είναι κλάδος της τεχνητής νοημοσύνης και βασίζεται στην επεξεργασία μικρών ή μεγάλων συνόλων δεδομένων, αναλύει τους συσχετισμούς μεταξύ τους, με απώτερο σκοπό την εξαγωγή γνώσης. Διερευνά τη μελέτη και την κατασκευή αλγορίθμων που μπορούν να μαθαίνουν από τα δεδομένα και να κάνουν προβλέψεις σχετικά με αυτά. Τέτοιοι αλγόριθμοι λειτουργούν κατασκευάζοντας μοντέλα από πειραματικά δεδομένα, προκειμένου να κάνουν προβλέψεις βασισμένες στα δεδομένα αυτά. (Ron Kohavi, 1998). Μέσω της μηχανικής μάθησης μπορούν να αναλυθούν δεδομένα προκειμένου να βρεθούν αλληλεξαρτήσεις μεταξύ διαφορετικών στοιχείων που είναι απίθανο να ανακαλύψουν οι άνθρωποι αναλυτές. Χρησιμοποιώντας κατάλληλα τους αλγόριθμους πρόβλεψης, είμαστε σε θέση όχι μόνο να παρακολουθούμε και να προβλέπουμε την λειτουργική κατάσταση του προς εξέταση συστήματος αλλά και να προσομοιάζουμε τεχνικές βελτίωσης ή νέες καινοτομίες σχεδίασης.

1.3 Σκοπός και Στόχοι Εργασίας

Η εφαρμογή της τεχνητής νοημοσύνης έχει αρχίσει να γίνεται κοινός τόπος στους περισσότερους τομείς της ανθρώπινης δραστηριότητας. Η ιατρική, οι οικονομικές επιστήμες, η κοινωνιολογία και σιγά σιγά η βιομηχανία έχουν αρχίσει να την ενσωματώνουν. Στη ναυτιλία και γενικά οι στις θαλάσσιες μεταφορές η διείσδυση της τεχνητής νοημοσύνης βρίσκεται ακόμη σε εμβρυακό στάδιο άλλα στο μέλλον αυτό πρόκειται να αλλάξει. Η προσαρμογή εργαλείων μηχανικής μάθησης ίσως μπορεί να

προσφέρει το μεγάλο όφελος στην αποτελεσματικότητα, τη βιωσιμότητα και τη μείωση του λειτουργικού κόστους στις ναυτιλιακές εταιρείες.

Σκοπός της παρούσας εργασίας είναι να εξετάσει το κατά πόσο η μηχανική μάθηση μπορεί να αντικαταστήσει σήμερα πιο παραδοσιακές προσεγγίσεις στο πρόβλημα της πρόβλεψης λειτουργικών παραμέτρων όπως η αποδιδόμενη ισχύς ή η κατανάλωση καυσίμου ενός πλοίου. Είναι λοιπόν εφικτή η χρησιμοποίηση του μεγάλου όγκου δεδομένων που παράγονται από ένα αυτοποιημένο σύστημα παρακολούθησης και καταγραφής ενός πλοίου ώστε με τη χρήση σύγχρονων εξελιγμένων αλγορίθμων μηχανικής μάθησης να επιτύχουμε προβλέψεις που μπορούν να αξιοποιηθούν;

2. ΥΠΟΒΑΘΡΟ

2.1 Ιστορική Αναδρομή

Η πρώτη αναφορά σε νευρωνικά δίκτυα έγινε το 1943, όταν ο νευροφυσιολόγος Warren McCulloch και ο μαθηματικός Walter Pitts έγραψαν μια εργασία για τους νευρώνες και τον τρόπο λειτουργίας τους. Αποφάσισαν να δημιουργήσουν ένα μοντέλο αυτού χρησιμοποιώντας ένα ηλεκτρικό κύκλωμα και ως εκ τούτου γεννήθηκε το πρώτο νευρικό δίκτυο. Η Μηχανική Μάθηση λειτουργεί αντίστοιχα με τον τρόπο αλληλεπίδρασης των εγκεφαλικών κυττάρων. Το μοντέλο αυτό δημιουργήθηκε το 1949 από τον Donald Hebb σε ένα βιβλίο με τίτλο The Organization of Behavior. Το βιβλίο παρουσιάζει τις θεωρίες του Hebb σχετικά με την ενεργοποίηση των νευρώνων και την επικοινωνία μεταξύ τους. Το 1950, ο Alan Turing δημιούργησε το παγκοσμίου φήμης Turing Test. Αυτό το test είναι αρκετά απλό, για να θεωρηθεί επιτυχές, πρέπει ένας υπολογιστής να είναι σε θέση να πείσει έναν άνθρωπο ότι είναι άνθρωπος και όχι υπολογιστής.

Ο Arthur Samuel της IBM ανέπτυξε ένα πρόγραμμα υπολογιστή για να παίζει πούλια στη δεκαετία του 1950. Δεδομένου ότι το πρόγραμμα έπρεπε να λειτουργήσει

με πολύ μικρή διαθέσιμη μνήμη RAM, ο Samuel ξεκίνησε αυτό που λέγεται alpha-beta pruning. Το σχέδιό του περιλάμβανε μια ρουτίνα βαθμολογίας χρησιμοποιώντας τις θέσεις των πούλιων στο ταμπλό. Η ρουτίνα αυτή προσπαθούσε να μετρήσει τις πιθανότητες έχει κάθε αντίπαλος να νικήσει. Το πρόγραμμα προτείνει την επόμενη κίνηση μέσω μιας στρατηγική ελαχιστοποίησης σφάλματος, η οποία τελικά εξελίχθηκε στον αλγόριθμο ελάχιστων σφαλμάτων. Ο Arthur Samuel ήταν αυτός που πρωτοεισήγαγε με τη φράση «Machine Learning» το 1952. Ο Frank Rosenblatt σχεδίασε το πρώτο τεχνητό νευρικό δίκτυο το 1958, που ονομάζεται Perceptron. Ο κύριος στόχος αυτού ήταν η αναγνώριση προτύπων σε σχήματα. Ένα άλλο παράδειγμα νευρικών δικτύων έγινε 1959, όταν ο Bernard Widrow και ο Marcian Hoff δημιούργησαν δύο μοντέλα στο Πανεπιστήμιο του Στάνφορντ. Το πρώτο ονομάστηκε ADELIN και μπορούσε να εντοπίσει δυαδικά μοτίβα. Η επόμενη γενιά ονομάστηκε MADELINE, και χρησιμοποιήθηκε στην εξάλειψη της ηχώ σε τηλεφωνικές γραμμές. Ήταν η πρώτη εμπορική εφαρμογή νευρωνικού δικτύου στον κόσμο και χρησιμοποιείται μέχρι σήμερα.

Η τεχνητή νοημοσύνη υπεισέρχεται σε μία περίοδο στασιμότητας τις δεκαετίες 60, 70 όπου δεν επιτελέστηκε καμία ουσιαστική πρόοδος. Η δυναμική των ανωτέρω εντυπωσιακών ανακαλύψεων ουσιαστικά παγώνει. Από την μία η έλλειψη επεξεργαστικής ισχύος στα υπολογιστικά συστήματα της εποχής, και από την άλλη οι επιδώσεις των εφαρμογών της εποχής, που αδυνατούσαν να εκπληρώσουν τις υπερβολικά υψηλές προσδοκίες που είχαν καλλιεργηθεί για την τεχνητή νοημοσύνη συντελούν σε αυτό.

Η επάνοδος στο προσκήνιο συντελείται το 1981, όταν η κυβέρνηση της Ιαπωνίας διαθέτει 850 εκατομμύρια δολάρια στο πρόγραμμα υπολογιστών πέμπτης γενιάς. Ο στόχος του προγράμματος ήταν η δημιουργία λογισμικού που θα ήταν σε θέση να πραγματοποιούν συνομιλίες, να μεταφράσουν γλώσσες, να ερμηνεύσουν εικόνες και να προσομοιάσουν την λογική που χαρακτηρίζει τα ανθρώπινα όντα. (McCorduck, 2019)

Παρόλο που τα ανωτέρω παραδείγματα ήταν επαναστατικά στην εποχή τους και ουσιαστικά συνετέλεσαν στην αλλαγή της χρήσης των ηλεκτρονικών υπολογιστών, η τεχνητή νοημοσύνη άρχισε να γίνεται ευρέως γνωστή στους ανθρώπους πολύ αργότερα, το 1997 όταν ο υπερυπολογιστής Deep Blue της IBM χρησιμοποιώντας τεχνητή νοημοσύνη, κατάφερε να νικήσει τον παγκόσμιο πρωταθλητή σκακιού Garry Kasparov.

2.2 Ανάπτυξη Θεωρητικού Πλαισίου

2.2.1 Λογικοί ή ευφυείς παράγοντες

Θεμελιώδης λίθος της τεχνητής νοημοσύνης (TN) αποτελεί ένας λογικός (ή ευφυής) παράγοντας. Τέτοιοι λογικοί παράγοντες “κρύβονται” πίσω από την παρούσα εργασία αφού αποτελούν τη βάση λειτουργίας όλων των εφαρμογών (TN) καθώς και Μηχανικής Μάθησης (MM). Οφείλουμε λοιπόν να κατανοήσουμε την δομή λειτουργίας του. Ως ευφυή παράγοντα ορίζουμε οποιονδήποτε είναι σε θέση να λαμβάνει αποφάσεις, μπορεί να είναι ένα άτομο, ένας οργανισμός ή μια επιχείρηση, μια μηχανή ή ακόμα και ένα λογισμικό. Κάθε τέτοιος πράκτορας πραγματοποιεί μια ενέργεια βασισμένη στο καλύτερο αποτέλεσμα στο οποίο καταλήγει αφού εξετάσει τις προηγούμενες και τις τρέχουσες αντιλήψεις (ένας λογικός πράκτορας στην περίπτωση εφαρμογής τεχνητής νοημοσύνης εξετάσει τις εισόδους δεδομένων του) Το ελάχιστο δυνατό σύστημα TN αποτελείται από έναν πράκτορα και το περιβάλλον του. Οι πράκτορες αλληλοεπιδρούν με το περιβάλλον τους το οποίο δύναται να περιλαμβάνει και άλλους πράκτορες. Εάν θέλαμε λοιπόν να δώσουμε τον ορισμό ενός λογικού πράκτορα θα λέγαμε ότι είναι μία μονάδα που:

«Αντιλαμβάνεται το περιβάλλον της μέσω κατάλληλων αισθητήρων και μέσω ενεργοποιητών αλληλοεπιδρά με αυτό.»

Για να κατανοήσουμε τη δομή των ευφύων πρακτόρων, θα πρέπει να αναλύσουμε την αρχιτεκτονική δομή τους. Η αρχιτεκτονική είναι ο μηχανισμός

λειτουργίας κάθε λογικού πράκτορα. Έστω μια συσκευή που αποτελείται από αισθητήρια και ενεργοποιητές, όπως για παράδειγμα: ένα ρομποτικό αυτοκίνητο, ένας υπολογιστής ή ένα αυτόνομο UAV. Το “λογισμικό” του πράκτορα είναι η εφαρμογή μιας συνάρτησης. Η συνάρτηση είναι ένας λογικός χάρτης της ακολουθίας αντίληψης των δεδομένων (ιστορικό όλων αυτών που ένας πράκτορας έχει αντιληφθεί μέχρι σήμερα) που καταλήγει σε μια δεδομένη ενέργεια.

Παραδείγματα πράκτορα:

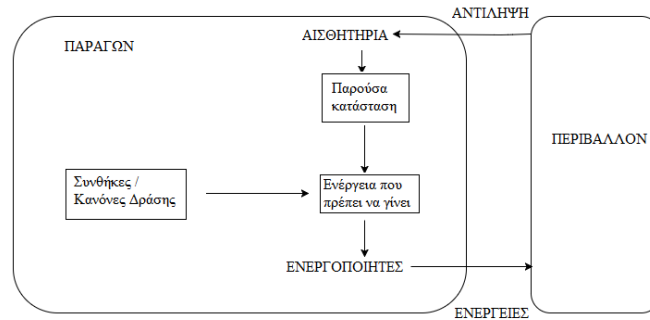
Ένας λογικός πράκτορας με την μορφή λογισμικού έχει ως αισθητήρες: οτιδήποτε εισέρχεται από τις συσκευές εισόδου ενός υπολογιστή (πληκτρολόγιο, ποντίκι), περιεχόμενο αρχείων, πακέτα δεδομένων που λαμβάνονται από το δίκτυο και ως ενεργοποιητές οτιδήποτε εμφανίζει στην οθόνη.

Ο άνθρωπος ως λογικός παράγοντας έχει μάτια, αυτιά και άλλα όργανα που λειτουργούν ως αισθητήρες και τα χέρια, τα πόδια, το στόμα και άλλα μέρη του σώματος ενεργούν ως ενεργοποιητές.

Οι βασικές μορφές των λογικών πρακτόρων είναι οι ακόλουθες:

Απλοί Αντανακλαστικοί Πράκτορες.

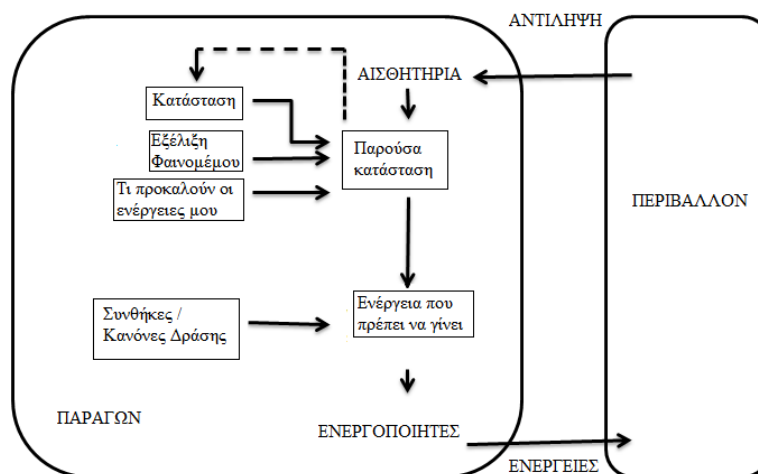
Οι απλοί αντανακλαστικοί παράγοντες αγνοούν το υπόλοιπο ιστορικό αντίληψης και ενεργούν μόνο με βάση την τρέχουσα. (Ως ιστορικό αντίληψης θεωρούμε οτιδήποτε έχει αντιληφθεί ένας πράκτορας μέχρι σήμερα.) Η συνάρτηση λειτουργίας βασίζεται στον κανόνα συνθήκης-δράσης (εάν η συνθήκη είναι αληθής τότε εκτελείται μια προσχεδιασμένη ενέργεια).



Σχ.1 Δομή απλού αντανακλαστικού πράκτορα

Αντανακλαστικοί Πράκτορες Βασιζόμενοι σε Κάποιο Μοντέλο

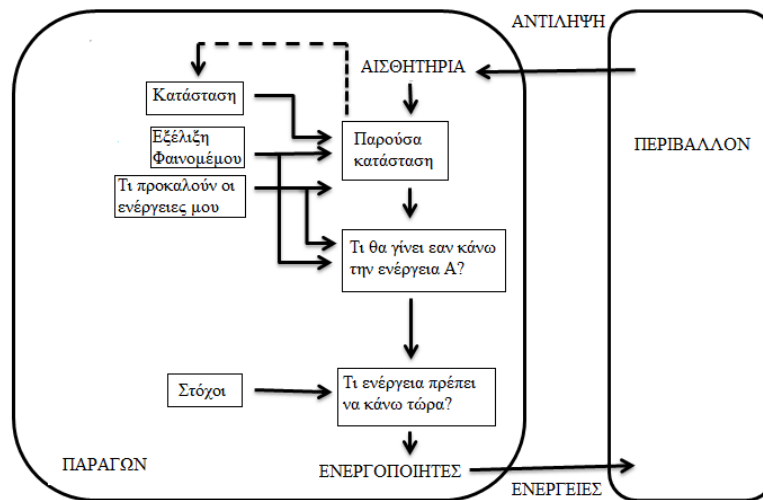
Λειτουργεί βρίσκοντας έναν κανόνα του οποίου οι συνθήκες ταιριάζουν με το περιβάλλον του. Ένας αντανακλαστικός πράκτορες βασιζόμενος σε μοντέλο μπορεί να χειριστεί μερικώς παρατηρήσιμα περιβάλλοντα με τη χρήση του αποθηκευμένου μοντέλου που έχει. Ο πράκτορας κρατάει ιστορικό της εσωτερικής κατάστασης του και την προσαρμόζει κατάλληλα μετά από κάθε νέα αντίληψη (είσοδο) με βάση το ιστορικό αντίληψης. Η τρέχουσα κατάσταση του περιβάλλοντος αποθηκεύεται μέσα στον πράκτορα ο οποίος διαμορφώνει κάποιο είδος δομής που περιγράφει το περιβάλλον του.



Σχ.2 Δομή αντανακλαστικού πράκτορα βασιζόμενου σε μοντέλο

Πράκτορες βασισμένοι στον στόχο

Αυτοί οι πράκτορες λαμβάνουν αποφάσεις με βάση την απόκλιση που έχουν εκείνη την στιγμή από τον στόχο τους (προδιαγεγραμμένη σαφής περιγραφή της επιθυμητής κατάστασης τους). Κάθε δράση τους αποσκοπεί στη μείωση της απόκλισης από τον στόχο. Αυτό τους δίνει την δυνατότητα να επιλέγουν μεταξύ διαφορετικών αποφάσεων, επιλέγοντας αυτήν που επιτρέπει την καλύτερη προσέγγιση του. Η γνώση που καθορίζει τις αποφάσεις ορίζεται ρητά αλλά μπορεί να τροποποιηθεί, γεγονός που τους καθιστά πιο ευέλικτους.

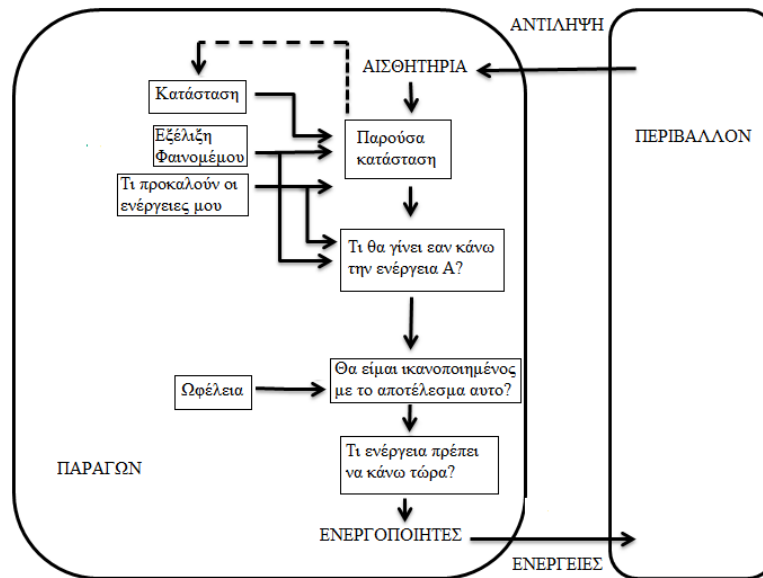


Σχ.3 Δομή πράκτορα βασιζόμενου σε στόχο

Πράκτορες βασισμένοι στη χρησιμότητα

Οι παράγοντες που αναπτύσσονται έχοντας την τελική χρήση τους ως δομικά στοιχεία και επομένως δεν μπορούν να χρησιμοποιηθούν οικουμενικά ονομάζονται πράκτορες βασισμένοι στη χρησιμότητα. Όταν υπάρχουν πολλές πιθανές

εναλλακτικές λύσεις, η καταλληλότητα των οποίων βασίζεται στο είδος του στόχου που καλούνται να επιτύχουν. Επιλέγουν ενέργειες με βάση μια προτίμηση (χρησιμότητα) για κάθε κατάσταση. Χρησιμοποιούνται εναλλακτικά σε περιπτώσεις που θέλουμε ταχύτητα στην επίτευξη ενός στόχου με έναν “φθηνό” τρόπο.



Σχ.4 Δομή πράκτορα βασισμένου στη χρησιμότητα

Εκπαιδευόμενος πράκτορας

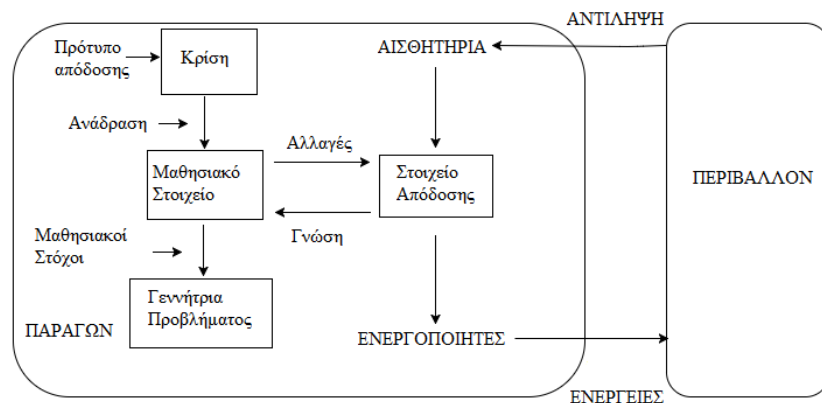
Ένας εκπαιδευόμενος πράκτορας στην ΤΝ είναι ο τύπος του πράκτορα που μπορεί να μάθει από τις προηγούμενες εμπειρίες του, ή έχει μαθησιακές δυνατότητες. Αποτελεί ουσιαστικά την πιο ολοκληρωμένη δομή και είναι ουσιαστικά αυτός που χρησιμοποιείται ευρέως στις σύγχρονες εφαρμογές υψηλής απόδοσης. Αρχίζει να ενεργεί με βασικές τις γνώσεις που του έχουν εισαχθεί ή έχει διαμορφώσει την δεδομένη χρονική στιγμή και μετά μπορεί να ενεργεί και να προσαρμόζεται αυτόματα μέσω της μάθησης. (Sahil_Bansall, 2019) Ένας εκπαιδευόμενος πράκτορας μάθησης τέσσερα βασικά εννοιολογικά στοιχεία, τα οποία είναι:

Στοιχείο εκμάθησης: Είναι υπεύθυνο για την εκτέλεση βελτιώσεων μαθαίνοντας από το περιβάλλον

Κριτική: Το στοιχείο εκμάθησης λαμβάνει ανατροφοδότηση από αυτό, το οποίο συγκρίνει τις ενέργειες του πράκτορα με ένα προδιαγεγραμμένο σταθερό πρότυπο απόδοσης.

Στοιχείο απόδοσης: Είναι υπεύθυνο για την επιλογή της ενέργειας που θα εκτελεστεί

Πρόγραμμα δημιουργίας προβλημάτων: Αυτό το στοιχείο είναι υπεύθυνο για την πρόταση ενεργειών με σκοπό την δημιουργία νέας εμπειρίας (γνώσης) του πράκτορα.



Σχ.5 Δομή εκπαιδευόμενου πράκτορα

2.2.2 Τύποι Μηχανικής Μάθησης

Με χρήση των ανωτέρω θεμελιωδών λογικών παραγόντων, ο τομέας της Μηχανικής Μάθησης αναπτύσσει τρεις τρόπους μάθησης, οι οποίοι βρίσκονται σε αντιστοιχία με τους τρόπους που χρησιμοποιεί και ο άνθρωπος για να μαθαίνει: την επιβλεπόμενη μάθηση, την μη επιβλεπόμενη μάθηση και τέλος την ενισχυτική μάθηση, αναλυτικότερα:

Η **μη επιβλεπόμενη μάθηση** χρησιμοποιείται στις περιπτώσεις εκείνες που για τα δεδομένα δεν υπάρχει προηγούμενη γνώση. Το σύστημα δεν έχει τη σωστή απάντηση. Ο αλγόριθμος που χρησιμοποιείται στη μη επιβλεπόμενη μάθηση οφείλει να καταλάβει τι εμφανίζεται. Ο στόχος του είναι να εξερευνήσει τα δεδομένα και να βρει συσχετισμούς και την δομή που κρύβεται μέσα τους. Η μη επιβλεπόμενη μάθηση λειτουργεί καλά με δεδομένα συναλλαγών. Για παράδειγμα, μπορεί να προσδιορίσει κατηγορίες πελατών με παρόμοια χαρακτηριστικά με σκοπό να μπορούμε μετά να εκτελέσουμε στενευμένες διαφημιστικές καμπάνιες.

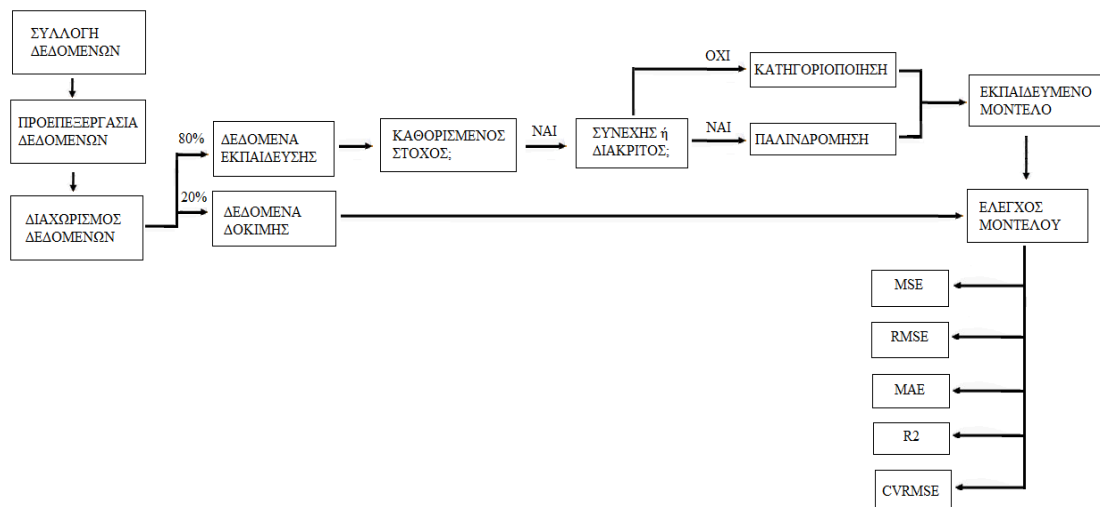
Η **επιβλεπόμενη μάθηση** χρησιμοποιεί αλγόριθμους οι οποίοι εκπαιδεύονται χρησιμοποιώντας γνωστά παραδείγματα, όπως μια είσοδο για την οποία όπου είναι γνωστή η επιθυμητή έξοδος. Για παράδειγμα, τροφοδοτούμε έναν υπολογιστή που εκτελεί επιβλεπόμενη μάθηση με πακέτα δεδομένων τα οποία έχουν χωριστεί σε δύο κατηγορίες A και B. Η κατηγορία στην οποία ανήκει κάθε πακέτο είναι γνωστή στον αλγόριθμο. Ο αλγόριθμος εκμάθησης λαμβάνει τα πακέτα δεδομένων μαζί με τις αντίστοιχες σωστές εξόδους (την κατηγορία στην οποία ανήκουν) και ο αλγόριθμος μαθαίνει συγκρίνοντας την πραγματική του έξοδο (πρόβλεψη) με τις σωστές εξόδους για να βρει λάθη. Στη συνέχεια τροποποιεί ανάλογα το μοντέλο με σκοπό την εξάλειψη των λαθών. Μέσω μεθόδων όπως η ταξινόμηση, η παλινδρόμηση, και η πρόβλεψη, η επιβλεπόμενη μάθηση χρησιμοποιεί μοτίβα για να προβλέψει την κατηγορία που ανήκουν τα πακέτα δεδομένων σε πρόσθετα δεδομένα για τα οποία δεν έχει αυτή τη γνώση. Η εποπτευόμενη μάθηση χρησιμοποιείται συνήθως σε εφαρμογές η προ υπάρχουσα γνώση μπορεί να προβλέψει πιθανά μελλοντικά γεγονότα. Για παράδειγμα, μπορεί να προβλέψει δόλιες συναλλαγές με πιστωτική κάρτα ή να βοηθήσει στην αξιολόγηση της πιστοληπτικής ικανότητας ενός πελάτη μιας τράπεζας.

Η **ενισχυτική μάθηση** χρησιμοποιείται συνήθως στην ρομποτική, σε παιχνίδια και σε εφαρμογές πλοήγησης. Οι αλγόριθμοι εκτελούν μια διαδικασία δοκιμής και σφάλματος μέσω τις οποίας προσπαθούν να ανακαλύψουν ποιες ενέργειες θα αποφέρουν την μεγαλύτερη προσέγγιση στον στόχο που έχει δοθεί. Η ενισχυτική μάθηση είναι η κατεξοχήν εφαρμογή της χρήσης των εκπαιδευόμενων πρακτόρων. Ο

στόχος είναι ο πράκτορας να επιλέξει ενέργειες που μεγιστοποιούν την προσέγγιση στο επιθυμητό αποτέλεσμα για ένα δεδομένο χρονικό διάστημα. Ο πράκτορας θα επιτύχει τον ανωτέρω στόχο πιο γρήγορα ακολουθώντας καλύτερη στρατηγική ενεργειών. Ο στόχος λοιπόν στην ενίσχυση της μάθησης είναι να μάθουμε την καλύτερη δυνατή στρατηγική.

2.2.3 Διάγραμμα Ροής / Στάδια Εκτέλεσης

Η πρόβλεψη λειτουργικών στοιχείων ενός πλοίου που αποτελεί το αντικείμενο αυτής της εργασίας και με βάση τα ανωτέρω ανήκει στην κατηγορία της επιβλεπόμενης μάθησης. Το βασικό διάγραμμα ροής εκτέλεσης ενός τέτοιου προβλήματος απεικονίζεται στο Σχ6.



Σχ.6 Διάγραμμα Ροής Μηχανικής Μάθησης

Αρχικό στάδιο εκτέλεσης της ανωτέρω διαδικασίας είναι η **Συλλογή ή Απόκτηση Δεδομένων (Data Acquisition)**. Όπως προ είπαμε, ο απώτερος σκοπός της μηχανικής μάθησης είναι η εξαγωγή γνώσης από σύνολα δεδομένων η οποία δεν είναι εμφανής. Η γνώση που θέλουμε να "εξάγουμε" βρίσκεται στα δεδομένα και συνεχώς αυξάνεται / τροποποιείται μαζί με αυτά. Οι αλγόριθμοι που χρησιμοποιούνται

“εκπαιδεύονται” από τα δεδομένα ώστε να είναι σε θέση να αντιληφθούν τους συσχετισμούς μεταξύ τους και να τους χρησιμοποιήσουν. Υπάρχουν υλοποιήσεις της μηχανικής μάθησης σε τομείς για τους οποίους υπάρχει πληθώρα ιστορικών στοιχείων, που μπορούν να αποτελέσουν ένα αξιόπιστο σύνολο δεδομένων, όμως η ευρεία χρήση της μηχανικής μάθησης μας έχει φέρει αντιμέτωπους με εφαρμογές στις οποίες η συγκέντρωση στοιχείων δεν θεωρούνταν απαραίτητη. Τέλος σε περιπτώσεις που ένα πρόβλημα πρωτοεμφανίζεται (π.χ. μία πρωτόγνωρη συμπεριφορά ενός συστήματος) για το οποίο υπάρχουν λίγα ή καθόλου δεδομένα κατάλληλα για την εκπαίδευση στη μηχανική μάθηση.

Υπάρχουν τρεις βασικές προσεγγίσεις στη βιβλιογραφία όσο αναφορά την απόκτηση δεδομένων: η ανακάλυψη δεδομένων, η επαύξηση δεδομένων και τέλος η δημιουργία δεδομένων. Η ανακάλυψη δεδομένων είναι η διαδικασία αναζήτησης συνόλων δεδομένων από διάφορες πηγές όπως το διαδίκτυο ή συγκεκριμένες εταιρικές πηγές που είναι προσβάσιμες. Ο χρήστης αναζητάει δεδομένα κατάλληλα για το πρόβλημα που έχει να αντιμετωπίσει και τα χρησιμοποιεί στην εκπαίδευση της εφαρμογής μηχανικής μάθησης που εφαρμόζει. Η διαδικασία αυτή συναντάται συνήθως σε ερευνητικά προγράμματα όπου η εκπαίδευση επί των διαδικασιών είναι ο αυτοσκοπός και λιγότερο η εξαγόμενη γνώση. Η αύξηση δεδομένων συμπληρώνει την ανακάλυψη δεδομένων. Τα όποια υπάρχοντα σύνολα δεδομένων εμπλουτίζονται με την προσθήκη περισσότερων από τον χρήστη. Χρησιμοποιώντας ένα βασικό σύνολο και εμπλουτίζοντας το με πακέτα δεδομένων που συγκεκριμενοποιούν το πρόβλημα, τα κάνουμε κατάλληλα για δική μας χρήση. Η τεχνική αυτή συνήθως εφαρμόζεται από ανεξάρτητους ερευνητές ή νεοσύστατους οργανισμούς. Η δημιουργία δεδομένων τέλος χρησιμοποιείται όταν δεν υπάρχει διαθέσιμο εξωτερικό σύνολο δεδομένων, αλλά είναι δυνατή η δημιουργία συνόλων δεδομένων. Μια βιομηχανία με ολοκληρωμένο σύστημα καταγραφής λειτουργικών παραμέτρων ή ένα πλοίο εν προκειμένω είναι δύο παραδείγματα. (Yuji Roh, 2019)

Το δεύτερο μεγάλο στάδιο της διαδικασίας εκτελέσεως είναι η **Επιλογή Δεδομένων και Μετασχηματισμός τους (Preprocessing / Transformation)**. Η

διαδικασία αυτή αποτελεί το σημαντικότερο τμήμα και είναι εκείνη που δύναται να επηρεάσει το μέγιστο το συνολικό αποτέλεσμα. Για να είναι δυνατή η χρησιμοποίηση ενός συνόλου δεδομένων, με αποδοτικό τρόπο, από έναν αλγόριθμο πρέπει να πληρούνται κάποιες ελάχιστες προϋποθέσεις.

Τα βασικά κριτήρια που καθορίζουν την ποιότητα των δεδομένων είναι η **εγκυρότητα**, η **ακρίβεια**, η **πληρότητα**, η **συνοχή** και η **ομοιομορφία**. (Refaat, 2007). Για να θεωρηθεί έγκυρο ένα σύνολο δεδομένων πρέπει να πληροί τις ακόλουθες προϋποθέσεις:

- Πρέπει να έχει σαφείς περιορισμούς στο εύρος των τιμών. Για παράδειγμα δεν μπορεί να υπάρχει καταχώρηση αρνητικής πίεσεως.
- Οι στήλες που περιγράφουν τα διάφορα δεδομένα πρέπει να είναι μοναδικές. Δεν μπορεί να υπάρχουν στήλες που περιγράφουν το ίδιο αντικείμενο.
- Τα δεδομένα κατηγοριοποίησης πρέπει να υπακούουν σε συγκεκριμένους κανόνες. π.χ. το χρώμα του δέρματος δεν μπορεί να είναι πράσινο.
- Τα δεδομένα κάθε στήλης οφείλουν να έχουν καταχωρηθεί με την ίδια μορφοποίηση.
- Εάν υπάρχουν συσχετισμοί μεταξύ διαφορετικών στηλών στα δεδομένα πρέπει να πληρούνται ορθά. Π.χ. η χρονική στιγμή ενεργοποίησης μιας μηχανής πρέπει να προηγείται της απενεργοποίησης της.

Κριτήριο ακρίβειας αποτελεί ο βαθμός στον οποίο τα δεδομένα του συνόλου προσεγγίζουν τις πραγματικές τιμές. Παρόλο λοιπόν που μπορεί να έχουμε ένα απολύτως έγκυρο σύνολο δεδομένων αυτό δεν σημαίνει πως είναι και ακριβές. Για παράδειγμα δεν αρκεί να εξαλείψουμε τις αρνητικές τιμές που έχουμε για την ταχύτητα ενός αντικειμένου, οφείλουμε να περιορίσουμε και την απόκλιση από την πραγματική τιμή. Το κριτήριο της πληρότητας είναι προφανές και είναι ο βαθμός στον

οποίο είναι γνωστά όλα τα απαιτούμενα δεδομένα. Το κριτήριο της συνοχής περιγράφει το κατά πόσο τα δεδομένα δεν δημιουργούν αντικρουόμενες πληροφορίες. Είναι ένα κριτήριο το οποίο τις περισσότερες φορές είναι δύσκολο να αξιολογηθεί εάν πληρείται διότι αποτελεί από μόνο του μια διαδικασία μηχανικής μάθησης. Ένα παράδειγμα που θα μπορούσε να δοθεί είναι να υπάρχει μια καταχώρηση σε κάποια χρονική στιγμή του βυθίσματος ενός πλοίου με τιμή που αντιστοιχεί στο μέγιστο βύθισμα και σε άλλη καταχώρηση και για την ίδια χρονική στιγμή το εκτόπισμα να είναι ελάχιστο. Τέλος η ομοιομορφία. Τα δεδομένα δεν πρέπει να έχουν διαφοροποιήσεις ως προς τον τύπο τους. Οφείλει δηλαδή κάθε μεμονωμένη στήλη να έχει αριθμητικές τιμές ή οτιδήποτε άλλο που να παραμένει όμως ίδιο σε όλο το σύνολο.

Το πρόβλημα είναι πως στην συντριπτική πλειοψηφία των περιπτώσεων μεγάλο μέρος των κριτηρίων ποιότητας των δεδομένων δεν πληρείται. Χαμένες τιμές από συστήματα καταγραφής, λάθη κατά την καταχώρηση των δεδομένων, προβλήματα κατά την αποθήκευσή τους ή ανθρώπινα λάθη είναι κοινώς τόπος σχεδόν σε όλες τις εφαρμογές. Σε μία διαδικασία ανεύρεσης γνώσης όλα αυτά τα σφάλματα εάν παραμείνουν δυσχεραίνουν ή μπορεί να καταστήσουν ακόμα και αδύνατη την εκτέλεσή της. Είναι τόσο σημαντική η "ποιότητα" των δεδομένων στην μηχανική μάθηση, που είναι δυνατόν να καταλήξουμε σε τελείως αντικρουόμενα αποτελέσματα με φαινομενικά ασήμαντες διαφοροποιήσεις στο αρχικό σύνολο δεδομένων. Καθίστανται λοιπόν επιτακτική η ανάγκη χρησιμοποίησης ενός μηχανισμού τροποποίησης του συνόλου των δεδομένων με σκοπό την καλύτερη αξιοποίηση του. Βασική προϋπόθεση στην όποια τροποποίηση είναι η μη διαφοροποίηση τους σε τέτοιο βαθμό ώστε να χαθεί ή να αλλοιωθεί η "γνώση" που εμπεριέχεται σε αυτά. Τα βασικά στάδια προ επεξεργασίας δεδομένων είναι τα ακόλουθα:

α) **Καθαρισμός των δεδομένων.** Ο καθαρισμός δεδομένων συνιστά την διαδικασία κατά την οποία εφαρμόζονται διάφορες τεχνικές αντισταθμίσεως των βασικότερων και πιο χονδροειδών σφαλμάτων τους. Εκτελούνται τα ακόλουθα:

- Απόρριψη άσχετων δεδομένων. Ως άσχετα ορίζουμε τα δεδομένα αυτά που δεν χρειάζονται πραγματικά και δεν ταιριάζουν στο πλαίσιο του προβλήματος που προσπαθούμε να λύσουμε.
- Αφαίρεση διπλότυπων.
- Μετασχηματισμός τύπου.
- Διόρθωση συντακτικών λαθών. Χρήση ομοιογενούς σύνταξης κάθε πεδίου δεδομένων.

β) **Διαχείριση χαμένων τιμών.** Κατά την διαδικασία αυτή λαμβάνεται μέριμνα να εκτελεστούν οι απαραίτητες ενέργειες ώστε να προκύψει ένα σύνολο δεδομένων που δεν θα εμπεριέχει πεδία με κενές τιμές. Η υλοποίηση του ανωτέρω μπορεί να πραγματοποιηθεί με διάφορες μεθόδους με τις σημαντικότερες να είναι:

- Διαγραφή ολόκληρης της γραμμής.
- Αναζήτηση και καταχώρηση της πραγματικής τιμής
- Χρήση μιας σταθερής τιμής για όλες τις χαμένες τιμές
- Αντικατάσταση της χαμένης τιμής με τη μέση τιμή της στήλης
- Αντικατάσταση της χαμένης τιμής με τη μέση τιμή της κλάσης
- Αντικατάσταση της χαμένης τιμής με κάθε δυνατή τιμή
- Αντικατάσταση της χαμένης τιμής με κάθε δυνατή τιμή για τις παρατηρήσεις της κλάσης
- Πρόβλεψη της χαμένης τιμής (με αλγόριθμο 1-NN)

γ) **Εξάλειψη "θορύβου".** Ως θόρυβο στα δεδομένα ορίζουμε όλες τις μη πραγματικές τιμές που εισάγονται στη συλλογή δεδομένων λόγω τυχαίων παραγόντων. Τα σφάλματα αυτά συνήθως προκύπτουν λόγω λάθους κατά την εισαγωγή δεδομένων ή λόγω σφαλμάτων μετάδοσης τους. Εξορισμού οι τιμές των σφαλμάτων αυτών παρουσιάζουν μεγάλη διαφοροποίηση σε σχέση με τα πραγματικά δεδομένα, πράγμα που συντελεί στον ευκολότερο εντοπισμό τους. Οι βασικότερες τεχνικές αντιμετώπισης του θορύβου των δεδομένων είναι:

- Κατακερματισμός σε διαστήματα και αντικατάσταση τιμών με αντικατάσταση μέσων όρων ή με αντικατάσταση οριακών τιμών
- Στατιστικός Εντοπισμός Εξαιρέσεων
- Χρήση Ανάλυσης Συστάδων
- Προσαρμογή των δεδομένων με χρήση μοντέλου

δ) **Προσαρμογή των δεδομένων.** Τα προβλήματα στα οποία προσπαθούμε να δώσουμε λύσεις μέσω μηχανικής μάθησης είναι ποικίλα και για αυτό οφείλουμε να τροποποιούμε τα δεδομένα συμφώνως αυτών. Διαφορετικές απαιτήσεις για το σύνολο δεδομένων προκύπτουν όταν θέλουμε να επιλύσουμε ένα πρόβλημα κατηγοριοποίησης και διαφορετικές όταν προσπαθούμε να προβλέψουμε μια μεταβλητή με συνεχείς τιμές. Σε κάθε περίπτωση η τμηματοποίηση του αρχικού συνόλου δεδομένων σε διάφορα μικρότερα τα οποία παρουσιάζουν κοινά γνωρίσματα βοηθάει την λειτουργία εκμάθησης των αλγορίθμων. Βασικές τεχνικές προσαρμογής των δεδομένων είναι η διακριτοποίηση και η κανονικοποίηση.

Η διακριτοποίηση (discretization) εκτελείται με:

- Διαστήματα ίσου πλάτους (equal width discretization)
- Διαστήματα ίσης συχνότητας (equal frequency discretization)
- Διακριτοποίηση βασισμένη στην Εντροπία (entropy based discretization)
- Διακριτοποίηση βασισμένη στην Ανάλυση Συστάδων (cluster based discretization)
- Τμηματοποίηση με φυσική κατάτμηση (segmentation by natural partitioning)

Ενώ η Κανονικοποίηση (normalization) με χρήση:

- Ελάχιστου-μέγιστου
- Z-score
- Δεκαδικής κλιμάκωσης

ε) **Μείωση των διαστάσεων (dimensionality reduction) / επιλογή σημαντικών χαρακτηριστικών (feature selection)**. Ο αριθμός των δεδομένων των μεταβλητών εισαγωγής ή των χαρακτηριστικών για ένα σύνολο δεδομένων αναφέρεται ως διαστάσεις του. Γενικά η συνηθέστερη μορφή των συνόλων δεδομένων είναι αυτή του πίνακα όπου στις στήλες έχουμε τα διαφορετικά χαρακτηριστικά εισόδων (Attributes) και στις γραμμές έχουμε τα πεδία τιμών (Instances). Ως μείωση διαστάσεων αναφερόμαστε σε τεχνικές που έχουν ως σκοπό να περιορίσουν τον αριθμό των χαρακτηριστικών. Υπάρχουν περιπτώσεις που η ύπαρξη υπερβολικά μεγάλου αριθμού χαρακτηριστικών, παρόλο που δεν είναι άσχετα και ταιριάζουν στο πλαίσιο του προβλήματος που προσπαθούμε να λύσουμε, δυσχεραίνουν την λειτουργία των αλγόριθμων μηχανικής μάθησης και επιφέρουν αντίθετο αποτέλεσμα. Τέτοιες τεχνικές λοιπόν μπορούν να χρησιμοποιηθούν σε εφαρμοσμένα προβλήματα κατηγοριοποίησης ή παλινδρόμησης απλοποιώντας το σύνολο δεδομένων, ώστε να ταιριάζει καλύτερα σε ένα μοντέλο πρόβλεψης. Οι βασικές τεχνικές περιορισμού των δεδομένων είναι οι ακόλουθες:

- Αναλογία ελλিপών τιμών
- Χαμηλή διακύμανση στις τιμές κάθε στήλης
- Υψηλή συσχέτιση μεταξύ δύο στηλών
- Ανάλυση βασικών συστατικών (PCA)
- Υποψήφιοι και διαχωρισμένες στήλες σε ένα τυχαίο δάσος
- Εξάλειψη χαρακτηριστικών προς τα πίσω
- Εμπρόσθια προσθήκη χαρακτηριστικών

Το επόμενο βήμα μετά την τελική διαμόρφωση του συνόλου των δεδομένων που θα χρησιμοποιηθούν, είναι ο διαχωρισμός του σε σύνολο δεδομένων εκπαίδευσης και σύνολο δεδομένων δοκιμής. Το πρώτο σύνολο χρησιμοποιείται για την εκπαίδευση των αλγορίθμων της μηχανικής μάθησης. Και στις δύο περιπτώσεις τα προς εξόρυξη δεδομένα είναι γνωστά. Ο διαχωρισμός γίνεται ως εξής: 80% του αρχικού συνόλου χρησιμοποιείται για εκπαίδευση και το υπόλοιπο 20% για δοκιμή,

αν και τα ανωτέρω νούμερα δεν είναι δεσμευτικά και υπάρχουν εφαρμογές που εφαρμόζουν διαφορετικά.

Ανάλογα με τον τύπο του προβλήματος που καλούμαστε να επιλύσουμε και πιο συγκεκριμένα ανάλογα με το είδος της πληροφορίας που αναμένουμε να αποκομίσουμε από τα δεδομένα, χρησιμοποιούμε διαφορετικές κατηγορίες μάθησης. Οι δύο κύριες κατηγορίες στην μηχανική μάθηση είναι:

- **Κατηγοριοποίηση (classification).** Κατά την οποία προσπαθούμε να δημιουργήσουμε έναν μηχανισμό υπολογισμού της κατηγορίας του κάθε αντικειμένου από τα υπόλοιπα γνωρίσματα του. Στην κατηγοριοποίηση, το αποτέλεσμα που θέλουμε να προβλέψουμε είναι η κλάση που ανήκουν τα άγνωστα δεδομένα. Η κλάση μπορεί να πάρει διακριτές τιμές από ένα πεπερασμένο σύνολο το οποίο έχουμε ορίσει από την αρχή.
- **Παλινδρόμηση (regression) :** Στόχος εδώ είναι η πρόβλεψη τιμών μιας εξαρτημένης μεταβλητής (dependent variable) με βάση κάποιες ανεξάρτητες μεταβλητές (independent variables). Χρησιμοποιούμε λοιπόν το σύνολο τιμών εισόδου για την πρόβλεψη ενός άλλου νέου μεγέθους που συσχετίζεται με αυτά. Η παλινδρόμηση προσομοιάζει την κατηγοριοποίηση αλλά εδώ η μεταβλητή-στόχος μπορεί να είναι οποιοσδήποτε πραγματικός αριθμός.

Άλλες κατηγορίες μάθησης που χρησιμοποιούνται είναι η **Ανάλυση Συστάδων (clustering)** όπου επιχειρείται ο επιμερισμός ενός συνόλου δεδομένων σε ομάδες. Έχοντας ένα σύνολο δεδομένων, στόχος της ανάλυσης συστάδων είναι η δημιουργία ομάδων (clusters), οι οποίες θα περιέχουν όμοια ή παρεμφερή δείγματα. Η **Ανάλυση Κανόνων Συσχέτισης (Mining Association Rules)** όπου προσπαθούμε να ανακαλύψουμε σχέσεις μεταξύ τιμών των γνωρισμάτων, οι οποίες εμφανίζονται συχνά μαζί. Οι κανόνες συσχέτισης ανακαλύπτουν τις κρυμμένες «αλληλεπιδράσεις» μεταξύ των γνωρισμάτων ενός συνόλου δεδομένων. Η **Ανάλυση Εξαιρέσεων** που εστιάζει

στην ανακάλυψη αποκλίσεων στα δεδομένα σε σχέση με αντίστοιχα δεδομένα, τα οποία έχουν συλλεχθεί στο παρελθόν ή με τυπικές τιμές των δεδομένων αυτών που μας είναι γνωστές. Επιχειρείται ο εντοπισμός εξαιρέσεων, χρησιμοποιώντας στατιστικές κατανομές πιθανοτήτων. Η **Ανάλυση Χρονοσειρών** αποτελεί μία άλλη ενδιαφέρουσα κατηγορία στην οποία με τη ανάλυση δεδομένων διαφορετικών χρονικών περιόδων προσπαθούμε να εξάγουμε χρήσιμα συμπεράσματα για το φαινόμενο και την εξέλιξή του στο μέλλον.

Ο αλγόριθμος κάνοντας χρήση των ανωτέρω τεχνικών "μαθαίνει" από τα δεδομένα εκπαιδεύσεως και ανακαλύπτει τους συσχετισμούς μεταξύ τους. Το δεύτερο σύνολο δεδομένων χρησιμοποιείται για την δοκιμή των συμπερασμάτων που έχουν προκύψει από την παραπάνω ανάλυση. Η μηχανική μάθηση είναι ουσιαστικά η διαδικασία «trial and error» που συντελείται σε αυτό το σημείο. Ο αλγόριθμος αποκτά γνώση συσχετισμού των δεδομένων την οποία εφαρμόζει στα δεδομένα δοκιμής. Τα αποτελέσματα πρόβλεψης συγκρίνονται με τις γνωστές τιμές του συνόλου δοκιμής και έτσι αξιολογούνται οι επιδόσεις. Μόλις φτάσουμε σε ένα σημείο στο οποίο οι επιδόσεις μας καλύπτουν, τότε είναι δυνατή η εφαρμογή του αλγορίθμου σε νέα αντίστοιχα σύνολα στα οποία είναι πλέον άγνωστες οι τιμές ενδιαφέροντος. Η διαδικασία αυτή είναι γνωστή σαν **Εξόρυξη Δεδομένων (data mining)**.

Το τελικό στάδιο της διαδικασίας ονομάζεται **Αξιολόγηση Προτύπων και Ανακάλυψη Γνώσης (Interpretation/Evaluation)**. Στο στάδιο αυτό εκτελούμε την μέτρηση των επιδόσεων με βάση εργαλείων μέτρησης (MSE, RMSE, R2 κ.α.) και αξιολογούμε τα αποτελέσματα. Η διαδικασία της μηχανικής μάθησης πλέον συγκλίνει στον πραγματικό κόσμο και τα αποτελέσματα αποκτούν φυσική υπόσταση.

2.3 Ανασκόπηση βιβλιογραφίας

Η τεχνητή νοημοσύνη και πιο συγκεκριμένα η μηχανική μάθηση κάνουν ακόμα τα πρώτα τους βήματα στις ναυτιλιακές μεταφορές. Παρόλο λοιπόν που η

επιστημονική κοινότητα έχει παρουσιάσει πληθώρα μελετών σχετικά με την χρησιμότητα της υιοθέτησης της σε πάρα πολλούς διαφορετικούς τομείς της ανθρώπινης δραστηριότητας, η έρευνα στην χρήση της μηχανικής μάθησης στην ναυτιλία είναι σχετικά πολύ περιορισμένη. Εντούτοις, έχουν παρουσιαστεί έρευνες που ελέγχουν τις δυνατότητες αξιοποίησής τους.

Οι Andrea Coraddua , Luca Onetob, Francesco Baldic και Davide Anguitad στο άρθρο “Vessels fuel consumption forecast and trim optimisation: A data analytics perspective ” (Andrea Coraddu, 2016) ερευνούν την δυνατότητα πρόβλεψης της κατανάλωσης καυσίμου και την βέλτιστη τιμή εκκρεμούς ενός σκάφους σε πραγματικές συνθήκες, και με βάση δεδομένα που μετρούνται από τα συστήματα αυτοματισμού του πλοίου. Συγκρίνουν τρεις διαφορετικές προσεγγίσεις για την πρόβλεψη της κατανάλωσης καυσίμου: Τα λευκό Μοντέλο (White Box, WBM) που βασίζεται στην γνώση των φυσικών διεργασιών που διέπουν τους παράγοντες που επηρεάζουν το φαινόμενο. Το Μαύρο μοντέλο (Black Box BBM) βασίζεται αποκλειστικά σε στατιστικές διαδικασίες συμπερασμάτων βάσει της ιστορικής συλλογής δεδομένων. Και τέλος οι συγγραφείς προτείνουν δύο Γκρι μοντέλα (Gray Box GBM) τα οποία είναι σε θέση να εκμεταλλευτούν και τις βασικές φυσικές αρχές και τις διαθέσιμες μετρήσεις.

Οι Christos Gkerekos, Iraklis Lazakis και Gerasimos Theotokatos στην μελέτη “Machine learning models for predicting shipmain engine Fuel Oil Consumption: A comparative study” (Christos Gkerekos, 2019) παρουσιάζουν μία σύγκριση επιδόσεων διαφόρων αλγόριθμων παλινδρόμησης στην πρόβλεψη της κατανάλωσης καυσίμου του κύριου κινητήρα ενός πλοίου. Εξετάζονται δύο διαφορετικές προσεγγίσεις στην απόκτηση των δεδομένων του πλοίου. Οι μεσημεριανές αναφορές του πληρώματος και το σύστημα αυτόματης καταγραφής και παρακολούθησης δεδομένων (ADLM).

Οι Fredrik Ahlgrena , Maria E. Mondejarb και ο Marcus Thernc στη μελέτη “ Predicting Dynamic Fuel Oil Consumption on Ships with Automated Machine Learning” (Fredrik Ahlgrena, 2019) παρουσιάζουν μια μέθοδο πρόβλεψης της

δυναμικής κατανάλωσης καυσίμου σε πλοία με χρήση αυτοματοποιημένου αλγόριθμου μηχανικής μάθησης, ο οποίος τροφοδοτείται μόνο με δεδομένα από μεγάλα χρονικά διαστήματα (από 12 ώρες έως 96 ώρες). Ο αλγόριθμος μηχανικής εκμάθησης εκπαιδεύτηκε σε δυναμικά δεδομένα των χαρακτηριστικών του κινητήρα σε μικρότερα χρονικά διαστήματα μαζί με δεδομένα για την κατανάλωση καυσίμου σε μεγαλύτερα χρονικά διαστήματα. Σε αυτήν τη μελέτη, προτείνεται μια προσέγγιση μηχανικής εκμάθησης χρησιμοποιώντας για τη βελτιστοποίηση της αυτόματης μηχανικής μάθησης, τα ήδη διαθέσιμα δεδομένα από το σύστημα καταγραφής μηχανημάτων.

2.4 Υπόθεση / Ερευνητικά Ερωτήματα

Είναι αδιαμφισβήτητο ότι το συντριπτικά μεγαλύτερο μέρος του παγκόσμιου εμπορίου εκτελείται δια της θαλάσσης. Καθημερινά ένας τεράστιος όγκος αγαθών και ανθρώπων διακινείται μέσω της ναυτιλίας εξασφαλίζοντας ουσιαστικά τον σύγχρονο τρόπο ζωής στον πλανήτη. Η θάλασσα ως μέσο εμπορίου δεν διαφαίνεται να σταματήσει για πολλά χρόνια να αποτελεί το πεδίο εκτέλεσης της ανωτέρω διαδικασίας, “αναγκάζοντας” ένα μεγάλο εύρος επιστημών καθώς και τεχνολογίας να εξετάζουν τρόπους διαρκούς βελτιστοποίησης των θαλάσσιων μεταφορών.

Οι βασικοί πυλώνες βελτιστοποίησης των θαλάσσιων μεταφορών αποτελούνται ουσιαστικά από δύο αλληλένδετους τομείς :

α) ΜΕΙΩΣΗ ΤΟΥ ΚΟΣΤΟΥΣ ΕΚΤΕΛΕΣΗΣ

β) ΜΕΙΩΣΗ ΤΟΥ ΕΝΕΡΓΕΙΑΚΟΥ ΑΠΟΤΥΠΩΜΑΤΟΣ

Συγκεκριμένα, γίνονται τεράστιες προσπάθειες ελαχιστοποίησης του κόστους διακίνησης αγαθών και ανθρώπων, με σκοπό να επιβιώσουν οι ναυτιλιακές εταιρείες στον αμείλικτο ανταγωνισμό του κλάδου. Ταυτόχρονα, η επιβάρυνση του περιβάλλοντος λόγω χρήσεως των κυρίαρχων πηγών ενέργειας αναγκάζει

παγκόσμιους οργανισμούς και κυβερνήσεις να πιέζουν για την θέσπιση ολοένα και πιο αυστηρών κανόνων εκπομπής μολυσματικών αερίων.

Με βάση την πρότερη ανάλυση των τεχνικών της μηχανικής μάθησης και του πως αυτές αξιοποιούνται σε διάφορους επιστημονικούς και επιχειρηματικούς κλάδους είμαστε σε θέση να διατυπώσουμε την παρακάτω υπόθεση:

Στην ικανοποίηση των πάγιων επιδιώξεων των θαλάσσιων μεταφορών ΜΠΟΡΕΙ να συνδράμει καθοριστικά η επιστράτευση της επιστήμης / τεχνολογίας των υπολογιστών

Με βάση την παραπάνω υπόθεση μπορούμε να διατυπώσουμε τα ερευνητικά ερωτήματα που αυτή η εργασία καλείται να αξιολογήσει / απαντήσει.

RQ1: Ποια είναι τα μετρήσιμα φυσικά μεγέθη που επιδρούν στην λειτουργική συμπεριφορά ενός πλοίου; Τα σύγχρονα συστήματα καταγραφής λειτουργικών παραμέτρων στα πλοία είναι σε θέση να καταγράφουν πληθώρα μεγεθών. Τα δεδομένα που συλλέγονται αυξάνονται με ταχύτατους ρυθμούς και ως προς τις σειρές, προσφέροντας συνεχώς αυξανόμενη στατιστική ισχύ, και ως προς τις στήλες (εισαγωγή περισσότερων μετρήσιμων χαρακτηριστικών) με υψηλότερη πολυπλοκότητα. Για την περιγραφή της ολοένα και αυξανόμενη διαθεσιμότητας συνόλων δεδομένων έχει αποδοθεί ο όρος Big Data. Οι προκλήσεις που υπεισέρχονται λόγω του Big Data είναι ποικίλες. Αρχικά η εισαγωγή υπερπληθώρας δεδομένων μπορεί να οδηγήσει σε υψηλότερο ποσοστό ψευδούς ανακάλυψης. Πιο σημαντικά και πρακτικά προβλήματα όμως είναι οι μεγάλες προκλήσεις που αφορούν τη λήψη των δεδομένων, την αποθήκευσή τους, την αναζήτηση, την κοινή χρήση, τη μεταφορά, την οπτικοποίηση, την αποστολή σε κέντρα λήψεως απόφασης, και την διασφάλιση του απόρρητου των πληροφοριών. Τα δεδομένα που συλλέγονται σήμερα ακόμα και από σχετικά απλά συστήματα αυτόματης καταγραφής λειτουργικών παραμέτρων έχουν τέτοιο μέγεθος που συχνά υπερβαίνουν την ικανότητα επεξεργασίας παραδοσιακού λογισμικού εντός ρεαλιστικού χρόνου και τιμής. Αντιλαμβανόμαστε λοιπόν την σημασία διερεύνησης και εντοπισμού των μεγεθών που επιδρούν σημαντικά στην

λειτουργική συμπεριφορά ενός πλοίου ώστε να διαμορφώσουν αυτά το υπο εξέταση σύνολο δεδομένων.

RQ2: Ποιες τεχνικές και αλγόριθμοι τεχνητής νοημοσύνης είναι δυνατόν αξιοποιηθούν στην ανάλυση δεδομένων λειτουργίας ενός πλοίου; Η μηχανική μάθηση εν γένει, είναι σε θέση να καλύψει ένα ευρύτατο πεδίο επιστημών διότι υπάρχουν πολλές διαφορετικές υλοποιήσεις της. Προβλήματα αντιστοίχισης σε μία κλάση αντιμετωπίζονται ως κατηγοριοποίηση, εάν αναζητάται η τιμή συγκεκριμένων μεγεθών ως παλινδρόμηση, η ανάλυση συστάδων, η ανάλυση κανόνων συσχέτισης, η ανάλυση εξαιρέσεων και τέλος η ανάλυση χρονοσειρών είναι μέθοδοι μάθησης που χρησιμοποιούνται ευρέως. Ο εντοπισμός της / των κατάλληλων μεθόδων μηχανικής μάθησης για την πρόβλεψη λειτουργικών παραμέτρων ενός πλοίου θα βοηθήσει στο να επικεντρωθούμε στην βελτιστοποίηση του. Τέλος, και με δεδομένο πως έχουμε καταλήξει σε μία από τις παραπάνω μεθόδους πρέπει να εμβαθύνουμε στα εργαλεία υλοποίησης της που δεν είναι άλλα από τους αλγόριθμους. Τα τεχνητά νευρωνικά δίκτυα, τα δέντρα απόφασης, τα δάση τυχαίας απόφασης, οι k-πλησιέστεροι γείτονες (kNN), η γραμμική παλινδρόμηση, η λογιστική παλινδρόμηση κ.α. είναι πιθανοί αλγόριθμοι που θα μπορούσαν να βοηθήσουν στην διερεύνηση. Καθένας από αυτούς λειτουργεί διαφορετικά επομένως ο έλεγχος των επιδόσεων τους παρουσιάζει ιδιαίτερο ενδιαφέρον.

RQ3: Είναι δυνατόν να τυποποιηθεί μια διαδικασία πρόβλεψης συμπεριφοράς πλοίου που να μπορεί να αξιοποιηθεί ευρέως; Η διαδικασία της μηχανικής μάθησης κατά την υλοποίηση της διέρχεται από συγκεκριμένα στάδια. Το αρχικό στάδιο είναι η συλλογή των δεδομένων της οποίας η βέλτιστη διαμόρφωση θα εξεταστεί απαντώντας στο πρώτο ερευνητικό ερώτημα. Το επόμενο στάδιο της διαμόρφωσης / προ επεξεργασίας του συνόλου αυτού που αποτελεί ίσως και το σημαντικότερο σκέλος της διαδικασίας αποτελεί ουσιαστικά το πρώτο σκέλος του

τελευταίου ερευνητικού ερωτήματος. Η διαδικασία της διαμόρφωσης του συνόλου των δεδομένων οφείλει να είναι *απαλλαγμένη* από οποιοδήποτε μέγεθος χαρακτηρίζει τον τύπο του εξεταζόμενου πλοίου. Το δεύτερο σκέλος του τρίτου ερευνητικού ερωτήματος εξετάζει την δυνατότητα αποφυγής του μεγαλύτερου κινδύνου σε μια εφαρμογή μηχανικής μάθησης. Το φαινόμενο *overfitting*. Ο όρος *overfitting* υιοθετήθηκε για να περιγράψει την κατάσταση κατά την οποία ένα μοντέλο μάθησης μαθαίνει με τέτοια λεπτομέρεια να λειτουργεί στο σύνολο δεδομένων εκπαίδευσης που αρχίζει να επηρεάζεται αρνητικά η απόδοση του μοντέλου σε νέα δεδομένα. Αυτό συμβαίνει διότι ο θόρυβος ή οι τυχαίες διακυμάνσεις στα δεδομένα εκπαίδευσης συλλέγονται και μαθαίνονται ως έννοιες από το μοντέλο. Η τυποποίηση λοιπόν μιας διαδικασίας μηχανικής μάθησης που μπορεί να χρησιμοποιηθεί σε κάθε τύπο πλοίου και για διάφορα σύνολα δεδομένων είναι το κλειδί για την ευρεία χρησιμοποίηση της.

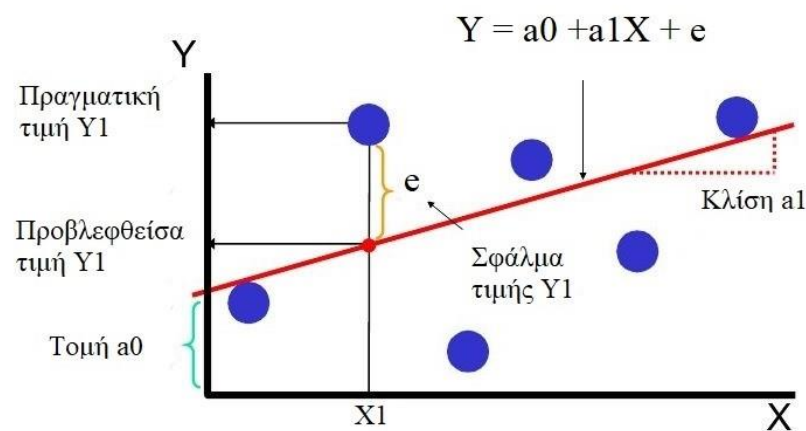
3. ΑΝΑΛΥΣΗ

3.1 Αλγόριθμοι Παλινδρόμησης

Το πρόβλημα της πρόβλεψης λειτουργικών παραμέτρων ενός πλοίου εμπίπτει στην κατηγορία της **παλινδρόμησης** δεδομένου πως προσπαθούμε να προβλέψουμε την τιμή ενός μεγέθους από την πληροφορία που "κρύβεται" σε άλλα χαρακτηριστικά του μεγέθου. Οι τιμές που δύναται να πάρει το προς αναζήτηση μέγεθος μπορεί να είναι οποιοσδήποτε φυσικός αριθμός. Οι αλγόριθμοι επιλύσεως προβλημάτων παλινδρόμησης που χρησιμοποιήθηκαν στην παρούσα εργασία είναι:

- Η Γραμμική Παλινδρόμηση (Linear Regression)
- Τα Δέντρα Απόφασης (Decision Tree)
- Οι k-πλησιέστεροι γείτονες (k-Nearest Neighbors)
- Τα Δάση Τυχαίας Απόφασης (Random Forest)
- Αλγόριθμος Ώθησης (AdaBoost)
- Τα Τεχνητά Νευρωνικά Δίκτυα (ANN)

Η **γραμμική παλινδρόμηση** είναι ένας από τους ευκολότερους και πιο δημοφιλείς αλγόριθμους Μηχανικής Μάθησης. Πρόκειται ουσιαστικά για μία στατιστική μέθοδος που χρησιμοποιείται στην προγνωστική ανάλυση. Με την γραμμική παλινδρόμηση μπορούμε να ξάνουμε προβλέψεις σε συνεχείς αριθμητικές μεταβλητές. Ο αλγόριθμος γραμμικής παλινδρόμησης εντοπίζει την γραμμική σχέση μεταξύ της εξαρτώμενης μεταβλητής που αναζητούμε (y) και μίας ή περισσότερων ανεξάρτητων (x) μεταβλητών. Δεδομένου ότι η σχέση αυτή είναι γραμμική, μπορούμε να εξακριβώσουμε εύκολα τον τρόπο που η εξαρτημένη μεταβλητή αλλάζει ανάλογα με την τιμή της ανεξάρτητης μεταβλητής. (Daksh, 2018) Παριστάνοντας γραφικά το μοντέλο λειτουργίας της γραμμικής παλινδρόμησης θα σχεδιάζαμε μια κεκλιμένη και ευθεία γραμμή που αντιπροσωπεύει το συσχετισμό μεταξύ ανεξάρτητων και εξαρτημένων μεταβλητών. Σχ7.



Σχ.7 Γραφική απεικόνιση γραμμικής παλινδρόμησης

Η μαθηματική διατύπωση του αλγορίθμου γραμμικής παλινδρόμησης είναι η εξής: $Y = a_0 + a_1X + e$

Όπου: Y = Εξαρτημένη μεταβλητή (ζητούμενο)

X = Ανεξάρτητη μεταβλητή (δεδομένα)

a_0 = Τομή γραμμής

a_1 = Συντελεστής γραμμικής παλινδρόμησης

e = Τυχαίο σφάλμα

Η γραμμική παλινδρόμηση χωρίζεται σε δύο μεγάλες κατηγορίες ανάλογα με τον αριθμό των ανεξάρτητων μεταβλητών. Απλή γραμμική παλινδρόμηση: Εάν ο αλγόριθμος μία ανεξάρτητη μεταβλητή για την πρόβλεψη της τιμής μιας εξαρτώμενης μεταβλητής τότε ονομάζεται απλή γραμμική παλινδρόμηση (Simple Linear Regression). Πολλαπλή γραμμική παλινδρόμηση: Εάν ο αλγόριθμος γραμμικής παλινδρόμησης χρησιμοποιεί περισσότερες από μία ανεξάρτητες μεταβλητές για την πρόβλεψη της τιμής μιας εξαρτώμενης αριθμητικής μεταβλητής, τότε ονομάζεται πολλαπλή γραμμική παλινδρόμηση (Multiple Linear Regression).

Κύριος στόχος όταν δουλεύουμε με γραμμική παλινδρόμηση, είναι να εντοπίσουμε την καλύτερη γραμμή προσαρμογής. Αυτό σημαίνει ότι πρέπει να βρεθεί η ευθεία που ελαχιστοποιηθεί το σφάλμα μεταξύ των προβλεπόμενων τιμών και των πραγματικών τιμών. Η καλύτερη γραμμή προσαρμογής είναι αυτή που συνολικά έχει το μικρότερο σφάλμα. Οι διαφορετικές τιμές για τους συντελεστές γραμμών (a_0 , a_1) δίνουν και μια διαφορετική γραμμή προσαρμογής, επομένως η βελτιστοποίηση του αλγορίθμου είναι ουσιαστικά ο υπολογισμός των τιμών για το a_0 και το a_1 ώστε να χαραχθεί η καλύτερη γραμμή προσαρμογής. Ο υπολογισμός αυτός εκτελείται με χρήση της συνάρτησης κόστους. Στον αλγόριθμο γραμμικής παλινδρόμησης χρησιμοποιούμε την συνάρτηση κόστους Mean Squared Error (MSE) η οποία θα αναλυθεί σε επόμενο κεφάλαιο.

Το **Δένδρο Απόφασης** είναι μια τεχνική εποπτευόμενης μάθησης η οποία χρησιμοποιείται γενικά είτε σε προβλήματα ταξινόμησης είτε σε προβλήματα παλινδρόμησης. Λόγω του τρόπου λειτουργίας του αποδίδει τα μέγιστα στην επίλυση προβλημάτων κατηγοριοποίησης αλλά υπάρχουν αρκετές περιπτώσεις που έχουν επιτευχθεί αξιοσημείωτα αποτελέσματα και σε πρόβλεψη αριθμητικών τιμών. Είναι

ουσιαστικά μίας δενδροειδούς μορφής ταξινομητής στον οποίο οι εσωτερικοί κόμβοι (nodes) αντιπροσωπεύουν τα χαρακτηριστικά ενός συνόλου δεδομένων, οι κόμβοι απόφασης αντιπροσωπεύουν τους κανόνες με τους οποίους λαμβάνονται αποφάσεις και έχουν πολλούς κλάδους ενώ κάθε φύλο αντιπροσωπεύει κάποιο αποτέλεσμα και δεν έχει περαιτέρω κλάδους. Οι αποφάσεις λαμβάνονται με γνώμονα τα χαρακτηριστικά του συνόλου δεδομένων. Ουσιαστικά είναι μια γραφική παράσταση που σκοπό έχει να διερευνήσει όλες τις πιθανές λύσεις σε ένα πρόβλημα που έχουμε θέσει βάσει συγκεκριμένων συνθηκών. Πήρε το όνομα Δένδρο Απόφασης διότι παρομοιάζει ένα ανάποδο πραγματικό δένδρο που ξεκινάει από τον ριζικό κόμβο, ο οποίος επεκτείνεται σε περαιτέρω κλαδιά και κατασκευάζει μια δομή που μοιάζει με δέντρο. Τα δέντρα απόφασης μιμούνται την λογική αλληλουχία της ανθρώπινης σκέψης κατά τη λήψη μιας απόφασης, και επομένως είναι οικείος και εύκολα κατανοητός ο τρόπος λειτουργίας τους.

Οι βασικές ορολογίες σε ένα δένδρο απόφασης είναι οι εξής:

Κόμβος ρίζας (Root Node): Ο κόμβος ρίζας είναι αφετηρία του δέντρο αποφάσεων. Αντιπροσωπεύει ολόκληρο το σύνολο δεδομένων, το οποίο εν συνεχεία θα διαιρεθεί σε δύο ή περισσότερα ομοιογενή σύνολα.

Κόμβοι Φύλλων (Leaf Node): Οι κόμβοι φύλλων είναι οι τελικοί κόμβοι εξόδου τους οποίους το δέντρο δεν μπορεί να διαχωριστεί περαιτέρω.

Διαχωρισμός (splitting) : Ονομάζεται η διαδικασία διαχωρισμού του κόμβου απόφασης / κόμβου ρίζας σε υπο-κόμβους σύμφωνα η οποία εκτελείται με βάση δεδομένες συνθήκες.

Υπό-δένδρο (Branch / Sub Tree): Ένα νέο δέντρο που σχηματίζεται μέσα στο αρχικό διαχωρίζοντας το δέντρο.

Κλάδεμα (Pruning): Το κλάδεμα είναι η διαδικασία αφαίρεσης των ανεπιθύμητων κλαδιών από το δέντρο με σκοπό την βελτίωση της λειτουργίας του. Γενικά ένα πολύ μεγάλο δέντρο αυξάνει τον κίνδυνο overfitting ενώ αντίθετα ένα

πολύ μικρό δέντρο ενδέχεται να μην καταγράψει όλα τα σημαντικά χαρακτηριστικά του συνόλου δεδομένων. Ως κλάδεμα ονομάζουμε την τεχνική μείωσης του μεγέθους του δέντρου εκμάθησης χωρίς να μειωθεί η ακρίβειά του.

Για να δημιουργήσουμε ένα δένδρο απόφασης κάνουμε χρήση του αλγόριθμου CART, ο οποίος σημαίνει Classification and Regression Tree algorithm. Το δέντρο αποφάσεων θέτει ερωτήσεις οι οποίες απαντώνται με Ναι ή Όχι και με βάση την απάντηση, το δέντρο επεκτείνεται σε υποδέντρα. Κατά την λειτουργία του δένδρου απόφασης ο αλγόριθμος ξεκινά από τον ριζικό κόμβο. Συγκρίνει τις τιμές του χαρακτηριστικού που του έχει δοθεί με το χαρακτηριστικό αναφοράς (πραγματικό σύνολο δεδομένων) και, με βάση τη σύγκριση, ακολουθεί τον κλάδο και μεταβαίνει στον επόμενο κόμβο. Για τον επόμενο κόμβο, ο αλγόριθμος συγκρίνει ξανά την τιμή χαρακτηριστικού με τους άλλους υπο-κόμβους και προχωρά στην επόμενη σύγκριση. Συνεχίζει τη διαδικασία μέχρι να φτάσει στον κόμβο των φύλλων του δέντρου όπου είναι αδύνατος ο περαιτέρω διαχωρισμός.

Τα βήματα που εκτελούνται από τον αλγόριθμο δένδρου απόφασης είναι τα παρακάτω:

Βήμα-1: Ξεκινάει το δέντρο με τον ριζικό κόμβο οποίος περιέχει το πλήρες σύνολο δεδομένων.

Βήμα-2: Βρίσκει το καλύτερο χαρακτηριστικό στο σύνολο δεδομένων χρησιμοποιώντας κάποιο μέτρο επιλογής χαρακτηριστικών Attribution Selection Measure (ASM).

Βήμα-3: Χωρίζει το αρχικό σύνολο σε υποσύνολα που περιέχουν πιθανές τιμές για τα καλύτερα χαρακτηριστικά.

Βήμα-4: Δημιουργεί έναν κόμβο δέντρου αποφάσεων, ο οποίος περιέχει το καλύτερο χαρακτηριστικό.

Βήμα-5: Δημιουργεί αναδρομικά νέα δέντρα αποφάσεων χρησιμοποιώντας τα υποσύνολα του συνόλου δεδομένων που δημιουργήθηκαν στο βήμα 3. Η διαδικασία αυτή συνεχίζεται έως ότου επιτευχθεί ένα στάδιο όπου είναι αδύνατη η περαιτέρω ταξινόμηση και διαχωρισμός των κόμβων.

Κατά την εφαρμογή ενός δέντρου απόφασης, η βασικότερη λειτουργία είναι η επιλογή του καλύτερου χαρακτηριστικού. Για την επίλυση τέτοιων προβλημάτων υπάρχει μια τεχνική που ονομάζεται μέτρο επιλογής χαρακτηριστικών ή ASM. Με αυτήν την διαδικασία, μπορούμε εύκολα να επιλέξουμε το καλύτερο χαρακτηριστικό για τους κόμβους του δέντρου. Υπάρχουν δύο δημοφιλείς τεχνικές για ASM, οι οποίες είναι:

- Το κέρδος πληροφοριών

Το κέρδος πληροφοριών είναι η τιμή της αλλαγής στην εντροπία μετά την τμηματοποίηση ενός συνόλου δεδομένων βάσει κάποιου χαρακτηριστικού. Η εντροπία αποτελεί την μέτρηση της "ακαθαρσίας" σε ένα δεδομένο χαρακτηριστικό. Καθορίζει την τυχαιότητα στα δεδομένα. Αποτελεί κριτήριο του πόσης πληροφορίας μας παρέχει ένα χαρακτηριστικό για μία κλάση. Σύμφωνα με την τιμή του κέρδους πληροφοριών, χωρίζουμε τον κόμβο και χτίζουμε το δέντρο αποφάσεων.

- Δείκτης Gini

Ο δείκτης Gini είναι ένα μέτρο ακαθαρσίας ή καθαρότητας που χρησιμοποιείται κατά τη δημιουργία ενός δέντρου αποφάσεων με τον αλγόριθμο CART (Classification and Regression Tree)

Στο Σχ8 φαίνεται μια γραφική αναπαράσταση ενός δένδρου απόφασης.



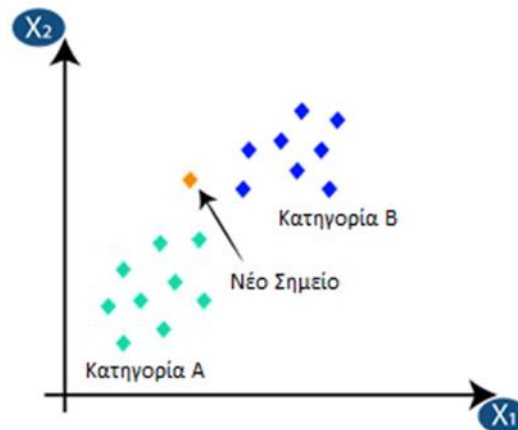
Σχ8. Γραφική αναπαράσταση δένδρου απόφασης.

Στο ανωτέρω παράδειγμα έστω ότι θέλουμε να αποφασίσουμε για το εάν είναι καλή ημέρα για να γυμναστούμε στην ύπαιθρο. Το δένδρο απόφασης με βάση τα στοιχεία που του δίδονται για τις καιρικές συνθήκες και την γνώση των συνθηκών που επιτρέπουν την άσκηση σε εξωτερικό χώρο, καλείται να λάβει την απόφαση αυτή. Απαντώντας με Ναι ή Όχι σε κάθε ερώτηση κλάδου είναι σε θέση να καταλήξει στο εάν οι εξωτερικές συνθήκες επιτρέπουν την άσκηση.

Ο αλγόριθμος **k πλησιέστερων γειτόνων** είναι ένας από τους πιο απλούς αλγόριθμους μηχανικής εκμάθησης και βασίζεται στην τεχνική της επιβλεπόμενης μάθησης. Ο αλγόριθμος K-NN αποθηκεύει τα διαθέσιμα δεδομένα και τα κατηγοριοποιεί. Σχηματίζει κατηγορίες δεδομένων με βάση τα κοινά τους χαρακτηριστικά. Τα νέα δεδομένα που του δίδονται και με βάση την ομοιότητα τους με τα βασικά χαρακτηριστικά κάθε κατηγορίας που έχει αποθηκεύσει, τα ταξινομεί. Η διαδικασία αυτή εκτελείται δυναμικά. Αυτό σημαίνει ότι όταν εμφανίζονται νέα δεδομένα τότε μπορούν εύκολα να ταξινομηθούν σε μια κατηγορία από τις διαθέσιμες. Ο αλγόριθμος K-NN μπορεί να χρησιμοποιηθεί τόσο για την παλινδρόμηση όσο και για την ταξινόμηση, αλλά ως επί το πλείστον χρησιμοποιείται για τα προβλήματα ταξινόμησης. Ο αλγόριθμος KNN στη φάση της εκπαίδευσης αποθηκεύει μόνο το

σύνολο δεδομένων και όταν λαμβάνει νέα δεδομένα, τότε ταξινομεί αυτά τα δεδομένα σε μια κατηγορία που μοιάζει πολύ με τα νέα δεδομένα.

Ας υποθέσουμε ότι έχουμε δύο κατηγορίες συνόλων δεδομένων οι οποίες παρουσιάζουν κάποια κοινά χαρακτηριστικά, την κατηγορία A και την κατηγορία B. Έστω και ένα νέο σημείο δεδομένων x_1 . Για να μπορέσουμε να εξακριβώσουμε σε ποια από τις δύο πιο πάνω κατηγορίες ανήκει, χρειαζόμαστε έναν αλγόριθμο k-NN. Σχ 9.

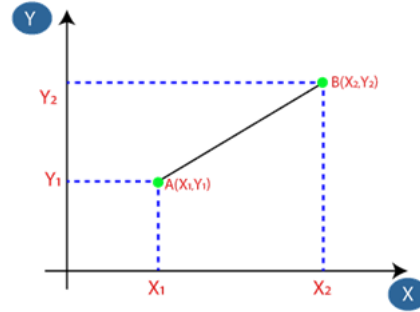


Σχ9. Νέο σημείο προς αντιστοίχιση

Τα βήματα που θα εκτελεστούν από τον αλγόριθμο K-NN για την επίλυση του ανωτέρω προβλήματος είναι τα εξής:

Βήμα-1: Επιλογή του αριθμού των K των γειτόνων

Βήμα-2: Υπολογισμός της Ευκλείδειας απόστασης μεταξύ των ανωτέρω K αριθμού γειτόνων Σχ10.

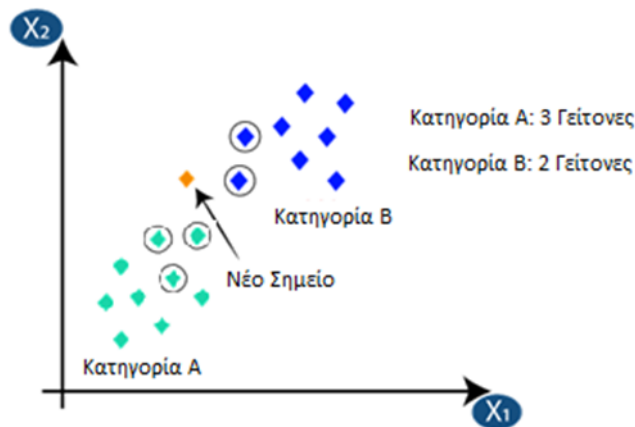


Ευκλείδεια απόσταση μεταξύ A1 και B2 = $\sqrt{(X_2 - X_1)^2 + (Y_2 - Y_1)^2}$

Σχ10 Ευκλείδεια Απόσταση Σημείων

Βήμα-3: Επιλογή των πλησιέστερων γειτόνων K σύμφωνα με την Ευκλείδεια απόσταση που υπολογίστηκε παραπάνω.

Βήμα-4: Μέτρηση του αριθμού των σημείων δεδομένων κάθε κατηγορίας από τους ανωτέρω γείτονες. Σχ 11



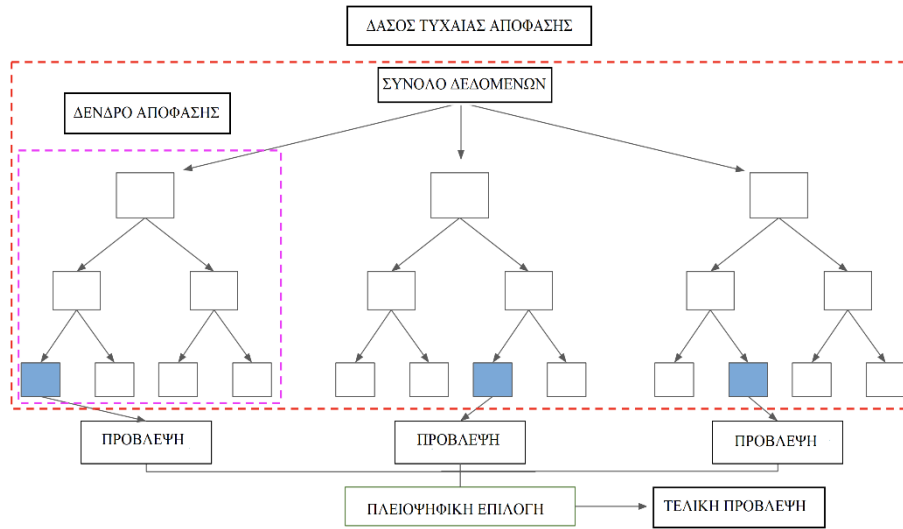
Σχ11 Αντιστοίχιση αριθμού k πλησιέστερων γειτόνων.

Βήμα-5: Αντιστοίχιση των νέων σημείων δεδομένων σε αυτήν την κατηγορία για την οποία ο αριθμός των γειτόνων είναι μέγιστος.

Η επιλογή του αριθμού k είναι θεμελιώδους σημασίας για το αποτέλεσμα του αλγόριθμου k-NN. Δυστυχώς παρόλο που στις περισσότερες περιπτώσεις η τιμή 5 για

το k είναι προτιμότερη αυτό δεν είναι πανάκεια. Δεν υπάρχει συγκεκριμένος τρόπος προσδιορισμού της καλύτερης και πρέπει να δοκιμαστούν μερικές τιμές για να βρεθεί η κατάλληλότερη ανάλογα με τα δεδομένα και τη φύση του προβλήματος. Μια πολύ χαμηλή τιμή για το K , όπως $K = 1$ ή $K = 2$, μπορεί να εισάγει "θόρυβο" και να οδηγήσει το μοντέλο σε χαμηλές επιδόσεις. Πολύ μεγάλες τιμές για το K είναι γενικά καλές, αλλά μπορεί να δημιουργήσουν δυσανάλογα αυξημένες απαιτήσεις υπολογιστικής ισχύος.

Το **Δάσος Τυχαίας Απόφασης** είναι ένας δημοφιλής αλγόριθμος μηχανικής μάθησης που ανήκει και αυτός στις τεχνικές επιβλεπόμενης μάθησης. Μπορεί να χρησιμοποιηθεί τόσο για προβλήματα κατηγοριοποίησης όσο και παλινδρόμησης. Βασίζεται στην έννοια του ensemble learning, κατά την οποία συνδυάζονται πολλαπλοί ταξινομητές ώστε να επιτύχουν τη βελτίωση της απόδοσης του μοντέλου, με σκοπό την επίλυση ενός πολύπλοκου και σύνθετου προβλήματος. Πήρε το όνομα δάσος διότι ουσιαστικά περιέχει έναν αριθμό δένδρων απόφασης τα οποία κάνοντας χρήση υποσυνόλων του αρχικού δεδομένων παίρνουν αποφάσεις και εκτελούν προβλέψεις παράλληλα. Αφού έχουν εκτελεστεί οι προβλέψεις από τα δένδρα απόφασης, παίρνει τον μέσο όρο από αυτά για να βελτιώσει την προγνωστική ακρίβεια του συνολικού και αρχικού συνόλου δεδομένων. Αντί να βασίζεται λοιπόν σε ένα μόνο δέντρο αποφάσεων, το τυχαίο δάσος λαμβάνει τις προβλέψεις από κάθε δέντρο και βασίζεται στην πλειοψηφία των προβλέψεων. Όσο μεγαλύτερος είναι ο αριθμός των δέντρων στο δάσος τόσο μεγαλύτερη είναι η ακρίβεια και αποτρέπεται το πρόβλημα του overfitting. Στο Σχ12 παρουσιάζεται η δομή του δάσους απόφασης.



Σχ12 Σχηματική αναπαράσταση δάσους απόφασης

Το δάσος απόφασης λειτουργεί σε δύο φάσεις. Αρχικά δημιουργεί το τυχαίο δάσος συνδυάζοντας έναν αριθμό από δένδρα απόφασης και σε δεύτερο χρόνο εκτελεί τις προβλέψεις κάθε δένδρου που δημιουργήθηκε στην πρώτη φάση.

Η ανωτέρω διαδικασία περιγράφεται αναλυτικά στα παρακάτω βήματα:

Βήμα-1: Επιλογή τυχαίων σημείων δεδομένων Σ από το εκπαιδευτικό σύνολο.

Βήμα-2: Δημιουργία των δένδρων αποφάσεων που σχετίζονται με τα επιλεγμένα σημεία δεδομένων (υποσύνολα).

Βήμα-3: Επιλογή του αριθμού N των δέντρων αποφάσεων που θα δημιουργηθούν.

Βήμα-4: Επανάληψη των βημάτων 1 & 2.

Βήμα-5: Για νέα σημεία δεδομένων, και με βάση τις προβλέψεις κάθε δέντρου αποφάσεων, αντιστοίχιση τους στην κατηγορία που κερδίζει την πλειοψηφία.

Ο αλγόριθμος **AdaBoost** (Adaptive Boosting) αποτελεί μια τεχνική ενίσχυσης που στοχεύει στο συνδυασμό πολλαπλών αδύναμων ταξινομητών για τη δημιουργία

ενός ισχυρού ταξινομητή. Ένας μεμονωμένος ταξινομητής μπορεί να αδυνατεί να προβλέψει με ακρίβεια την κλάση ενός αντικειμένου, αλλά όταν ομαδοποιούμε πολλούς αδύναμους ταξινομητές με τον καθένα να μαθαίνει σταδιακά από τις λανθασμένες ταξινομήσεις των άλλων, μπορούμε να δημιουργήσουμε ένα πολύ ισχυρό μοντέλο. Ο AdaBoost δεν αποτελεί ουσιαστικά ένα αυτόνομο μοντέλο αλλά μπορεί να εφαρμοστεί πάνω από οποιονδήποτε ταξινομητή για να μάθει από τις ελλείψεις του και να προτείνει ένα πιο ακριβές μοντέλο. (Kurama, 2020) Αποτελεί άλλο ένα παράδειγμα ενισχυτικής μάθησης. Λειτουργεί με βάση την αρχή της διαδοχικής βελτίωσης της μαθησιακής διαδικασίας. Εκτός από τον πρώτο, κάθε επόμενος ταξινομητής συνεχίζει το έργο από το σημείο που σταμάτησε ο προηγούμενος.

Τα Τεχνητά Νευρωνικά Δίκτυα είναι το μαθηματικό ισοδύναμο της λειτουργίας των βιολογικών νευρώνων που απαρτίζουν τον ανθρώπινο εγκέφαλο. Αποτελούνται από έναν αριθμό απλών και εσωτερικά διασυνδεδεμένων επεξεργαστικών μονάδων, οι οποίες οργανώνονται σε στρώματα. Τα ΤΝΔ είναι οργανωμένα σε στρώματα (layers). Είναι δυνατόν εκτός από το στρώμα εισόδου (input layer) και το στρώμα εξόδου (output layer) να υπάρχουν και ενδιάμεσα κρυφά στρώματα (hidden layers). Τα παραπάνω στρώματα αποτελούνται από έναν αριθμό μονάδων (units) ή κόμβων (nodes) που είναι έτσι συνδεδεμένες μεταξύ τους, ώστε μία μονάδα να έχει συνδέσμους με πολλές άλλες μονάδες του ίδιου ή άλλου επιπέδου. Οι μονάδες επιδρούν σε άλλες μονάδες. Κάθε μία επεξεργαστική μονάδα επεξεργάζεται πληροφορίες ανταποκρινόμενη δυναμικά σε εξωτερικά ερεθίσματα (εισόδους). Κάθε τεχνητός νευρώνας αποτελείται από πολλές εισόδους x_i και μία μόνο έξοδο y . Σε κάθε είσοδο x_i τοποθετείται ένας συντελεστής “ βάρους” w_i και τα αποτελέσματα αθροίζονται μέσω της συνάρτησης αθροίσματος (summation function) F

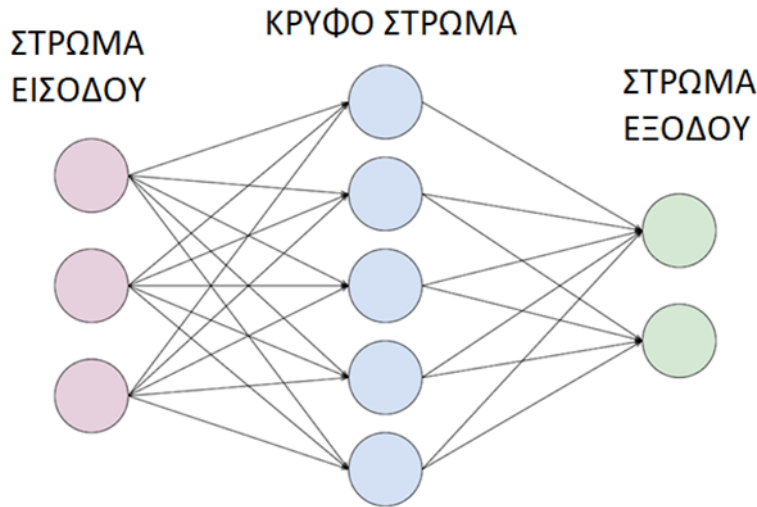
$$F = \sum_i^n x_i w_i$$

Η έξοδος y ενός τεχνητού νευρώνα προκύπτει με χρήση της συνάρτησης μετάβασης (transfer function), αλλά μόνο όταν το άθροισμα των εισόδων πολλαπλασιασμένα με τα βάρη τους είναι μεγαλύτερο μιας ορισμένης τιμής κατωφλίου (threshold value) θ , δηλαδή όταν:

$$\sum_i^n x_i w_i - \theta > 0$$

Βασικά στοιχεία της αρχιτεκτονικής των ΤΝΔ που καθορίζονται κατά τη δημιουργία τους και επιδρούν στην απόδοσή τους είναι ο αριθμός των ενδιάμεσων κρυφών επιπέδων, ο αριθμός των μονάδων (ή κόμβων) ανά επίπεδο, ο τρόπος διασύνδεσης των μονάδων μεταξύ τους, η μορφή της συνάρτησης μετάβασης, οι τιμές των αρχικών βαρών μεταξύ των μονάδων, η τιμή ενεργοποίησης (τιμή κατωφλίου) και ο αλγόριθμος εκπαίδευσης και τροποποίησης των βαρών.

Οι μονάδες μεταξύ τους μπορεί να είναι διασυνδεδεμένες με δύο τρόπους διαμορφώνοντας αντίστοιχα και δύο βασικές αρχιτεκτονικές. Πρόσθιας τροφοδότησης όπου οι μονάδες ενός επιπέδου τροφοδοτούν τις μονάδες του επόμενου επιπέδου και όχι τις άλλες μονάδες του ίδιου επιπέδου (Σχ13) και οπίσθιας τροφοδότησης όπου οι μονάδες ενός επιπέδου είναι δυνατόν να τροφοδοτούν μονάδες του ίδιου ή ακόμα και προηγούμενου επιπέδου. Η πρώτη κατηγορία αποτελεί και την συντριπτική πλειοψηφία των ΤΝΔ που χρησιμοποιούνται σήμερα.



Σχ13 Σχηματική αναπαράσταση TNN

Η εκπαίδευση του ΤΝΔ είναι ουσιαστικά η διαδικασία ευρέσεως των κατάλληλων βαρών που θα τοποθετηθούν στις εισόδους των επεξεργαστικών μονάδων. Τροποποιήσεις στα βάρη έχουν ως αποτέλεσμα την αλλαγή της γενικής συμπεριφοράς ΤΝΔ ως προς την ικανότητα του να παρέχει στο μέλλον την επιθυμητή έξοδο από μία δεδομένη είσοδο. Όταν η επιθυμητή αυτή έξοδος είναι εκ των προτέρων γνωστή, αναφερόμαστε σε επιβλεπόμενη διαδικασία εκπαίδευσης (supervised learning), αντίθετα για περιπτώσεις που η έξοδος δεν είναι γνωστή έχουμε μη επιβλεπόμενη εκπαίδευση (unsupervised learning). Η διαδικασία αυτή της αλλαγής των βαρών των εισόδων εκτελείται εσωτερικά στα ΤΝΔ και υλοποιείται από τις παρακάτω τρεις κατηγορίες εκπαιδευτικών αλγορίθμων:

- Αλγόριθμος οπισθοδιάδοσης λάθους (error-backpropagation)
- Ανταγωνιστική μάθηση (competitive learning)
- Τυχαία μάθηση (random learning)

Ο συνηθέστερη διαδικασία εκπαίδευσής ενός ΤΝΔ υλοποιείται με χρήση αλγορίθμου οπισθοδιάδοσης λάθους. Κατά την διαδικασία αυτή και για κάθε μονάδα ενός επιπέδου ξεκινώντας από το τέλος, λαμβάνονται υπόψη οι διαφορές μεταξύ του

υπολογιζόμενου και του επιθυμητού αποτελέσματος και διαδίδονται προς τα πίσω στις μονάδες του κρυφών επιπέδων, έτσι ώστε να καθορίσουν τις απαραίτητες αλλαγές που απαιτείται να γίνουν στα βάρη σύνδεσης μεταξύ των μονάδων.

3.2 Μέτρηση Επιδόσεων

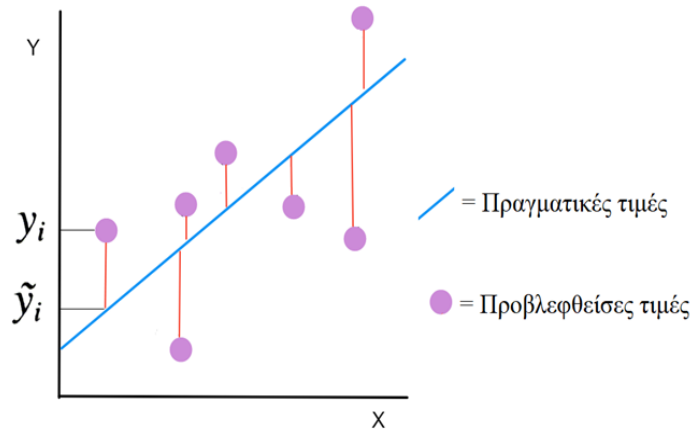
Κατά την διαδικασία πρόβλεψης τιμών σε ένα πρόβλημα με χρήση μηχανικής μάθησης ακολουθούμε τα εξής βήματα: συλλογή δεδομένων, προ επεξεργασία δεδομένων, διαχωρισμός του συνόλου δεδομένων σε σύνολο εκπαίδευσης και σύνολο δοκιμής, επιλογή κατάλληλου αλγορίθμου ανάλογα με την φύση του προβλήματος, εκπαίδευση του αλγορίθμου, έλεγχο ακρίβειας των παραγόμενων αποτελεσμάτων, βελτιστοποίηση παραμέτρων του αλγορίθμου και τέλος εξαγωγή γνώσης και ανάλυση. Η διαδικασία αυτή είναι συγκεκριμένη και μέχρι στιγμής έχει αναλυθεί, αλλά προκύπτει ένα ερώτημα. Πως ελέγχουμε την ακρίβεια των παραγόμενων από ένα αλγόριθμο αποτελεσμάτων; Ποιο ή ποια είναι τα κριτήρια σύγκρισης που θα μας βοηθήσουν να επιλέξουμε την καταλληλότερη διαμόρφωση και θα μας κάνουν να αποδεχτούμε τα αποτελέσματα;

Η ακρίβεια μίας προβλέψεως ορίζεται ως ο συνολικός αριθμός σωστών προβλέψεων διαιρούμενος με τον συνολικό αριθμό προβλέψεων που έγιναν για ένα σύνολο δεδομένων. Σε ένα πρόβλημα παλινδρόμησης όμως που αναζητείται η πρόβλεψη συγκεκριμένης τιμής ενός χαρακτηριστικού, ο ορισμός αυτός δεν μας καλύπτει. Οφείλουμε να ορίσουμε μεθόδους που ουσιαστικά θα θέτουν όρια αποκλίσεως προβλεπόμενων τιμών και μέσω αυτών θα είμαστε σε θέση να συγκρίνουμε τις επιδόσεις πρόβλεψης. Τα σημαντικότερα εργαλεία μετρήσεως επιδόσεων είναι τα ακόλουθα:

- Mean Absolute Error (MAE)
- Mean Squared Error (MSE)

- Root Mean Squared Error (RMSE)
- R Square (R²)
- Coefficient of Variation of Root-Mean Squared Error (CVRMSE)

Ας θεωρήσουμε την γραφική αναπαράσταση του Σχ4 που παριστάνει τα αποτελέσματα προβλέψεων ενός αλγορίθμου παλινδρόμησης. Η μπλε γραμμή αναπαριστά τις πραγματικές τιμές και οι μωβ κουκίδες τις προβλεφθείσες.



Σχ4 Γραφική αναπαράσταση αποτελεσμάτων πρόβλεψης

Το **Mean Absolute Error (MAE)** αποτελεί τον μέσο όρο των απόλυτων διαφορών μεταξύ των προβλέψεων και των πραγματικών τιμών και μας δίνει ένα μέτρο για το πόσο απέχουν οι προβλέψεις από την πραγματικότητα. Ορίζεται μαθηματικά από την σχέση:

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

Όπου N = αριθμός προβλέψεων

Το **Mean Squared Error (MSE)** προσομοιάζει αρκετά το MAE, με την μόνη διαφορά πως το MSE παίρνει το μέσο όρο του τετραγώνου της διαφοράς μεταξύ των

πραγματικών τιμών και των προβλεπόμενων τιμών. Το πλεονέκτημα του MSE είναι ότι είναι πιο εύκολο να υπολογιστεί. Επίσης είναι σημαντικό να τονίσουμε πως λόγω της χρησιμοποίησης των τετραγώνων του σφάλματος, η επίδραση των μεγαλύτερων σφαλμάτων γίνεται πιο έντονη, επομένως το μοντέλο επικεντρώνεται στα μεγαλύτερα λάθη. (Binieli, 2018) Η μαθηματική του διατύπωση είναι η εξής:

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

Το **Root Mean Squared Error (RMSE)** είναι η τετραγωνική ρίζα του MSE αποτελεί ουσιαστικά ίδιο εργαλείο μέτρησης με αυτό αλλά έχει επικρατήσει και χρησιμοποιείται αρκετά στη βιβλιογραφία. Η μαθηματική του διατύπωση είναι η εξής:

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

Το **Coefficient of Variation of Root-Mean Squared Error (CVRMSE)** αποτελεί το επόμενο βήμα από το RMSE ομαλοποιώντας το από τη μέση εξαρτώμενη μεταβλητή τιμή. Η μαθηματική του διατύπωση είναι:

$$CV(RMSE) = \frac{1}{\bar{Y}} \sqrt{\frac{\sum_{i=1}^N (Y_i - \hat{Y}_i)^2}{N}}$$

Τέλος το **Rsquare (R²)** επικεντρώνεται όχι τόσο στα αποτελέσματα αλλά στην ίδια την λειτουργία του αλγορίθμου. Καθορίζει τον βαθμό στον οποίο οι διακυμάνσεις των τιμών της εξαρτημένης μεταβλητής (στόχος πρόβλεψης) μπορούν να εξηγηθούν από τις αλλαγές των τιμών των ανεξάρτητων μεταβλητών (σύνολο δεδομένων). Ιδανικά θα θέλαμε οι αλλαγές των τιμών του στόχου πρόβλεψης να είναι σε θέση να

εξηγηθούν πλήρως από τις ανεξάρτητες μεταβλητές. (Yeresime Suresh, 2014) Σε αυτή την περίπτωση το R^2 θα είχε τιμή 1. Η μαθηματική του διατύπωση είναι η εξής:

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

4. ΜΕΘΟΔΟΛΟΓΙΑ / ΥΛΟΠΟΙΗΣΗ

4.1 Μεταβλητές Έρευνας / Πλαίσιο Δεδομένων

Τα δεδομένα για την πραγματοποίηση της συγκεκριμένης εργασίας είναι πραγματικά και προέρχονται από το σύστημα παραμετρικής παρακολούθησης της πλατφόρμας ενός Crude Oil Tanker. Πρόκειται για ένα πλοίο εκτοπίσματος περί τους 165.000 τόνους του οποίου το μήκος μεταξύ καθέτων είναι της τάξης των 265 μέτρων, έχει μέγιστο πλάτος 50 μέτρα και μέγιστο βύθισμα 23 μέτρα. Το προωστήριο σκεύος του συγκεκριμένου πλοίου αποτελείται από μία μεγάλη αργόστροφη δίχρονη μηχανή diesel 6S70MC-C7 κατασκευής MAN. Η μέγιστη αποδιδόμενη ισχύς της μηχανής είναι 18.660 kW που αντιστοιχεί στις μέγιστες στροφές της που είναι 90rpm. Η μηχανή μεταφέρει την ισχύ στην μοναδική προπέλα του πλοίου που έχει διάμετρο 8200 χιλιοστά. Για την ηλεκτρική ισχύ φροντίζουν τρία ηλεκτροπαραγωγά ζεύγη, καθένα εκ των οποίων παράγει 850 kW σε τάση εξόδου γεννήτριας 450 Volt. Τα βασικά χαρακτηριστικά του πλοίου που χρησιμοποιήθηκε φαίνονται στον πίνακα 1.

Χαρακτηριστικό Πλοίου	Τιμή Μεγέθους	Μονάδα Μέτρησης
Εκτόπισμα	165.000	Tons
Κύρια Μηχανή	18.000	kW

Τύπος	Δίχρονη	-
Αριθμός Κυλίνδρων	6	EA
Μέγιστες Στροφές	90	rpm
Ηλεκτρομηχανές	850 / 450	kW / Volt
Προπέλα	FPP 4EA	-
Διάμετρος	8200	mm
LBL	265	m
B (mld)	50	m
D (mld)	23	m

Πίν. 1 Βασικά Χαρακτηριστικά Πλοίου

Το εν λόγω πλοίο έχει αυτοματοποιημένο σύστημα καταγραφής και επιτήρησής Automated data logging & Monitoring system (ADLM), το οποίο είναι σε θέση να καταγράφει τιμές σε 178 διαφορετικές παραμέτρους λειτουργίας του. Η συχνότητα δειγματοληψίας του συστήματος επιτρέπει την πλήρη καταγραφή όλων των ανωτέρω τιμών κάθε ένα λεπτό της ώρας. Τα δεδομένα εκτός από την απεικόνιση τους, αποθηκεύονται σε μία βάση δεδομένων που επιτρέπει την μετέπειτα επεξεργασία τους. Στην αρχή κάθε κύκλου μετρήσεων δημιουργείται μία χρονοσφραγίδα (timestamp) μέσω της οποίας είναι δυνατή η ανασκόπηση καθώς και ο συγχρονισμός των τιμών. Στον πίνακα 2 φαίνεται το πλήθος και η περιγραφή των μεγεθών που συλλέγονται.

local_ID	Attribute	Measurement Unit	local_ID	Attribute	Measurement Unit	local_ID	Attribute	Measurement Unit	local_ID	Attribute	Measurement Unit
0	Longitudinal_Water_Speed	Knots	46	ME_CYL1_SCAV_AIR_TEMP_AMS	°C	92	DG1_FO_INLET_TEMP_AMS	°C	138	DG3_TC_EXH_GAS_OUT_TEMP_AMS	°C
1	Transverse_Water_Speed	Knots	47	ME_CYL2_SCAV_AIR_TEMP_AMS	°C	93	DG1_LO_INLET_PRESS_AMS	Bar	139	DG1_TC_A_EXH_GAS_IN_TEMP_AMS	°C
2	Longitudinal_Ground_Speed	Knots	48	ME_CYL3_SCAV_AIR_TEMP_AMS	°C	94	DG1_LO_INLET_TEMP_AMS	°C	140	DG1_TC_B_EXH_GAS_IN_TEMP_AMS	°C
3	Transverse_Ground_Speed	Knots	49	ME_CYL4_SCAV_AIR_TEMP_AMS	°C	95	DG2_FO_INLET_PRESS_AMS	Bar	141	DG2_TC_A_EXH_GAS_IN_TEMP_AMS	°C
4	Stern_Transverse_Water_Speed	Knots	50	ME_CYL5_SCAV_AIR_TEMP_AMS	°C	96	DG2_FO_INLET_TEMP_AMS	°C	142	DG2_TC_B_EXH_GAS_IN_TEMP_AMS	°C
5	Stern_Transverse_Ground_Speed	Knots	51	ME_CYL6_SCAV_AIR_TEMP_AMS	°C	97	DG2_LO_INLET_PRESS_AMS	Bar	143	DG3_TC_A_EXH_GAS_IN_TEMP_AMS	°C
6	Total_Cumulative_Water_Distance	Miles	52	ME_PC01_OUTLET_TEMP_AMS	°C	98	DG2_LO_INLET_TEMP_AMS	°C	144	DG3_TC_B_EXH_GAS_IN_TEMP_AMS	°C
7	Water_Distance_Since_Reset	Miles	53	ME_PC02_OUTLET_TEMP_AMS	°C	99	DG3_FO_INLET_PRESS_AMS	Bar	145	Aux_Boiler_Status	I/O
8	Water_Depth_Offset_From_Transducer	Meters	54	ME_PC03_OUTLET_TEMP_AMS	°C	100	DG3_FO_INLET_TEMP_AMS	°C	146	Aux_Boiler_Running_Time	Seconds
9	Water_Depth_Maximum_Range_Scale_In_U	Meters	55	ME_PC04_OUTLET_TEMP_AMS	°C	101	DG3_LO_INLET_PRESS_AMS	Bar	147	AUX_BLR1_FO_INLET_TEMP_AMS	Celcius
10	Wind_Direction	Degrees	56	ME_PC05_OUTLET_TEMP_AMS	°C	102	DG3_LO_INLET_TEMP_AMS	°C	148	AUX_BLR1_FO_INLET_PRESS_AMS	Bar
11	Wind_Speed	Knots	57	ME_PC06_OUTLET_TEMP_AMS	°C	103	DG1_CFW_OUT_TEMP_AMS	°C	149	AUX_BLR1_DRUM_LEVEL_AMS	Number
12	Magnetic_Variation	Number	58	ME_CYL1_CFW_OUTLET_TEMP_AMS	°C	104	DG2_CFW_OUT_TEMP_AMS	°C	150	AUX_BLR1_STEAM_DRUM_PRESS_AMS	Bar
13	Speed_Over_Water	Knots	59	ME_CYL2_CFW_OUTLET_TEMP_AMS	°C	105	DG3_CFW_OUT_TEMP_AMS	°C	151	AUX_BLR1_EXH_GAS_OUTLET_TEMP_AMS	Celcius
14	Latitude	0,°, "	60	ME_CYL3_CFW_OUTLET_TEMP_AMS	°C	106	DG1_CFW_IN_PRESS_AMS	Bar	152	AUX_BLR2_FO_INLET_TEMP_AMS	Celcius
15	Longitude	0,°, "	61	ME_CYL4_CFW_OUTLET_TEMP_AMS	°C	107	DG2_CFW_IN_PRESS_AMS	Bar	153	AUX_BLR2_FO_INLET_PRESS_AMS	Bar
16	UTC	Time	62	ME_CYL5_CFW_OUTLET_TEMP_AMS	°C	108	DG3_CFW_IN_PRESS_AMS	Bar	154	AUX_BLR2_DRUM_LEVEL_AMS	Number
17	True_Course_Over_Ground	Degrees	63	ME_CYL6_CFW_OUTLET_TEMP_AMS	°C	109	DG1_WINDING_TEMP_R_PHASE_AMS	°C	155	AUX_BLR2_STEAM_DRUM_PRESS_AMS	Bar
18	Magnetic_Course_Over_Ground	Degrees	64	ME_CYL1_EXH_GAS_OULET_TEMP_AMS	°C	110	DG1_WINDING_TEMP_S_PHASE_AMS	°C	156	AUX_BLR2_EXH_GAS_OUTLET_TEMP_AMS	°C
19	Speed_Over_Ground	Knots	65	ME_CYL2_EXH_GAS_OULET_TEMP_AMS	°C	111	DG1_WINDING_TEMP_T_PHASE_AMS	°C	157	Aux_Boiler_2_Status	I/O
20	Water_Depth	Meters	66	ME_CYL3_EXH_GAS_OULET_TEMP_AMS	°C	112	DG2_WINDING_TEMP_R_PHASE_AMS	°C	158	Aux_Boiler_2_Running_Time	Seconds
21	True_Heading	Degrees	67	ME_CYL4_EXH_GAS_OULET_TEMP_AMS	°C	113	DG2_WINDING_TEMP_S_PHASE_AMS	°C	159	ME_FO_Counter_Mass	Kgr
22	Rate_Of_Turn	Degrees per Minute	68	ME_CYL5_EXH_GAS_OULET_TEMP_AMS	°C	114	DG2_WINDING_TEMP_T_PHASE_AMS	°C	160	ME_FO_Flow_Mass	Kgr/min
23	Time	Time	69	ME_CYL6_EXH_GAS_OULET_TEMP_AMS	°C	115	DG3_WINDING_TEMP_R_PHASE_AMS	°C	161	DG_FO_Counter_Volumetric	Lt
24	Day	Day	70	ME_TC_RPM_AMS	Rpm	116	DG3_WINDING_TEMP_S_PHASE_AMS	°C	162	DG_FO_Flow_Volumetric	Lt/min
25	Month	Month	71	Fuel_Rack_Position_AMS	Number	117	DG3_WINDING_TEMP_T_PHASE_AMS	°C	163	DG_FO_Flowmeter_Temp	Celcius
26	Year	Year	72	ME_START_AIR_PRESS_AMS	Bar	118	DG1_START_AIR_IN_PRESS_AMS	Bar	164	FIRE_MAIN_PRESS_AMS	Bar
27	Inclinometer_X_max	Degrees	73	ME_CONTROL_AIR_PRESS_AMS	Bar	119	DG2_START_AIR_IN_PRESS_AMS	Bar	165	SG_Significant_Wave_Height	Meters
28	Inclinometer_X_min	Degrees	74	ME_FO_INLET_PRESS_AMS	Bar	120	DG3_START_AIR_IN_PRESS_AMS	Bar	166	SG_Mean_Wave_Period	Time
29	Inclinometer_X_zero_cross	Degrees	75	ME_FO_INLET_TEMP_AMS	°C	121	DG1_CYL1_EXH_GAS_OUT_TEMP_AMS	°C	167	SG_Wind_Speed_at_10m	Knots
30	Inclinometer_Y_max	Degrees	76	ME_LO_INLET_PRESS_AMS	Bar	122	DG1_CYL2_EXH_GAS_OUT_TEMP_AMS	°C	168	SG_Wind_direction	Degrees
31	Inclinometer_Y_min	Degrees	77	ME_LO_INLET_TEMP_AMS	°C	123	DG1_CYL3_EXH_GAS_OUT_TEMP_AMS	°C	169	SG_Mean_Wave_Direction	Degrees
32	Inclinometer_Y_zero_cross	Degrees	78	ME_JCFW_INLET_PRESS_AMS	Bar	124	DG1_CYL4_EXH_GAS_OUT_TEMP_AMS	°C	170	SG_Wind_Sea_Wave_Height	Meters
33	Inclinometer_DX	Degrees	79	ME_SCAV_AIR_PRESS_AMS	Bar	125	DG1_CYL5_EXH_GAS_OUT_TEMP_AMS	°C	171	SG_Wind_Sea_Wave_Period	Time
34	Inclinometer_DY	Degrees	80	ME_Rpm_AMS	Rpm	126	DG2_CYL1_EXH_GAS_OUT_TEMP_AMS	°C	172	SG_Wind_Sea_Direction	Degrees
35	Draft_Fore	Meters	81	ME_EXH_GAS_MEAN_TEMP_AMS	°C	127	DG2_CYL2_EXH_GAS_OUT_TEMP_AMS	°C	173	SG_Swell_Wave_Height	Meters
36	Draft_Aft	Meters	82	CAMSHAFT_CHAIN_W_H_BRG_TEMP_AMS	°C	128	DG2_CYL3_EXH_GAS_OUT_TEMP_AMS	°C	174	SG_Swell_Wave_Period	Time
37	Draft_Mid_S	Meters	83	ME_THRUST_BEARING_SEG_TEMP_AMS	°C	129	DG2_CYL4_EXH_GAS_OUT_TEMP_AMS	°C	175	SG_Wind_Direction	Degrees
38	Draft_Mid_P	Meters	84	ME_AIR_COOLER_CFW_OUTLET_TEMP_AMS	°C	130	DG2_CYL5_EXH_GAS_OUT_TEMP_AMS	°C	176	SG_Swell_Wave_Direction	Degrees
39	ME_Torque	kNm	85	ME_AIR_COOLER_CW_INLET_PRESS_AMS	Bar	131	DG3_CYL1_EXH_GAS_OUT_TEMP_AMS	°C	177	SG_Current_Velocity	Time
40	ME_Rpm	Rpm	86	INTERM_SHAFT_BEARING_TEMP_AMS	°C	132	DG3_CYL2_EXH_GAS_OUT_TEMP_AMS	°C			
41	ME_Power	kW	87	STERN_TUBE_AFT_BRG_TEMP_AMS	°C	133	DG3_CYL3_EXH_GAS_OUT_TEMP_AMS	°C			
42	DIGITAL_ALARMS_AMS	Number	88	DG1_Power_kW	kW	134	DG3_CYL4_EXH_GAS_OUT_TEMP_AMS	°C			
43	ME_TC_LO_INLET_TEMP_AMS	°C	89	DG2_Power_kW	kW	135	DG3_CYL5_EXH_GAS_OUT_TEMP_AMS	°C			
44	ME_TC_LO_OUTLET_TEMP_AMS	°C	90	DG3_Power_kW	kW	136	DG1_TC_EXH_GAS_OUT_TEMP_AMS	°C			
45	ME_TC_EXH_GAS_OUTLET_TEMP_AMS	°C	91	DG1_FO_INLET_PRESS_AMS	Bar	137	DG2_TC_EXH_GAS_OUT_TEMP_AMS	°C			

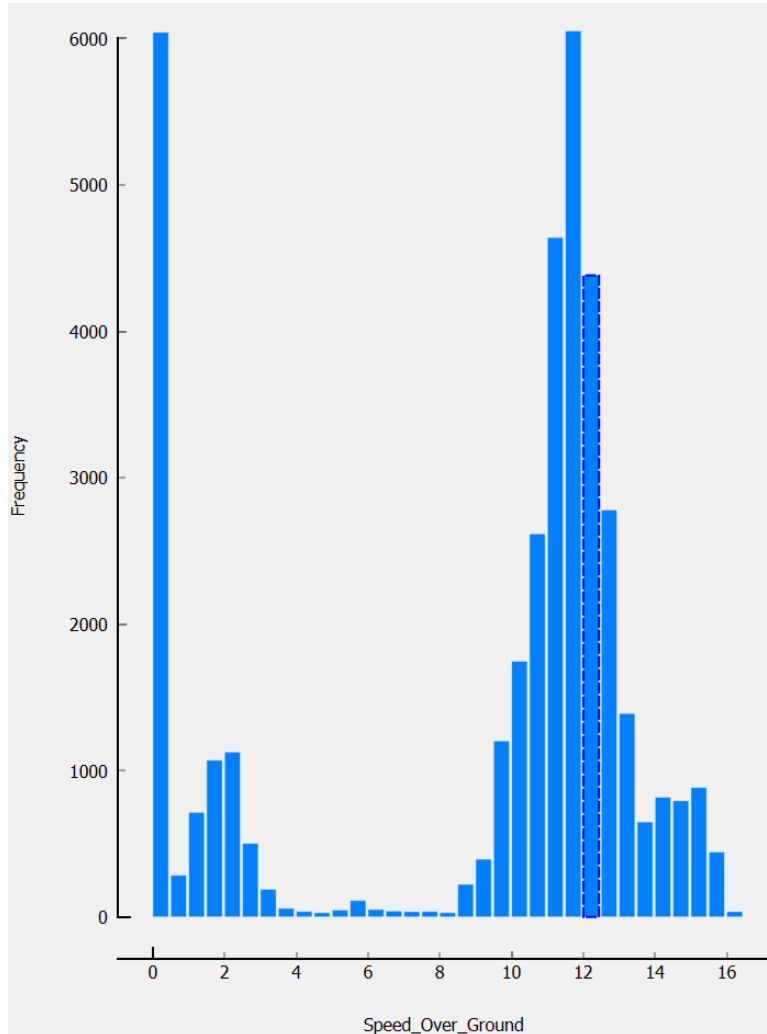
Πίν. 2 Δεδομένα συλλογής συστήματος

Τα δεδομένα που συλλέχθηκαν αντιστοιχούν σε μία περίοδο 6 μηνών και αφορούν την περίοδο από 13 Φεβρουαρίου 2020 στις 00:00 έως 26 Ιουλίου 2020 στις 00:00. Σε αυτή την χρονική περίοδο το πλοίο εκτέλεσε πολλά και διαφορετικά δρομολόγια. Όπως επίσης υπήρχαν και αρκετές περιόδους κατά τις οποίες το πλοίο βρισκόταν σε λιμάνια. Οι περιοχές πλεύσης του περιλαμβάνουν την περίμετρο της Αφρικανικής Ηπείρου, την Μεσόγειο, τον Ατλαντικό, τον Ινδικό και τον Ειρηνικό Ωκεανό. Η διασπορά στην υψήλιο είναι μεγάλη όπως και η χρονική περίοδος συλλογής τιμών, πράγμα που συνέσπει στην όσο το δυνατόν μεγαλύτερη ανεξάρτηση των δεδομένων από κλιματολογικά φαινόμενα. Στην εικόνα 1 παρουσιάζεται ο χάρτης με τα στίγματα των διαδρομών του πλοίου κατά την περίοδο ενδιαφέροντος.



Εικ1. Δρομολόγια πλοίου

Τα δρομολόγια που εκτελέστηκαν στο ανωτέρω χρονικό διάστημα παρουσιάζουν μια σύγκλιση ως προς τις τιμές ταχύτητας με την οποία εκτελέστηκαν. Γενικά σε πλοία αυτού του τύπου προτιμάται η παραμονή σε σχετικά σταθερές ταχύτητες και αποφεύγονται οι πολλές διακυμάνσεις. Από την στιγμή της ολοκλήρωσης των χειρισμών προσέγγισης ή απομάκρυνσης από ένα λιμένα, παρατηρείται η επιλογή μιας σταθερής ταχύτητας. Οι καιρικές συνθήκες, κάποια επικίνδυνη κατάσταση ή συγκεκριμένες απαιτήσεις του συμβολαίου μεταφοράς είναι οι συνηθέστερες αιτίες τροποποίησης της ταχύτητας του πλοίου αλλά και πάλι για αναλογικά σύντομο χρονικό διάστημα. Στην εικόνα 2 φαίνεται η κατανομή των ταχυτήτων οι οποίες κυριάρχησαν το χρονικό διάστημα καταγραφής τιμών. Διαφαίνεται η μεγάλη συγκέντρωση τιμών μεταξύ των ταχυτήτων 10 και 14 κόμβων.



Εικ2 Κατανομή ταχυτήτων πλοίου

Το σύνολο δεδομένων το οποίο διαμορφώθηκε από τις τιμές που καταχωρήθηκαν στο σύστημα καταγραφής και επιτήρησής του πλοίου έχει 178 διαφορετικές παραμέτρους. Οι παράμετροι αυτοί εάν διαμορφώσουμε το σύνολο δεδομένων σε πίνακα αντιστοιχούν στις στήλες του και ονομάζονται attributes. Οι διαφορετικές τιμές που έχουν οι παραπάνω στήλες αντιστοιχούν στις γραμμές του πίνακα και ονομάζονται instances. Ο συνολικός αριθμός των γραμμών του πίνακα που αναπαριστά το σύνολο δεδομένων είναι 236161.

Στο πλαίσιο πραγματοποίησης της παρούσας εργασίας τέθηκε ως ζητούμενο πρόβλεψης η απαιτούμενη ισχύς της μηχανής, η οποία αποτέλεσε και την εξαρτημένη μεταβλητή του προβλήματος μηχανικής μάθησης, και το υπόλοιπο σύνολο δεδομένων διαμόρφωσε τις ανεξάρτητες μεταβλητές που τροφοδοτούν τους αλγορίθμους. Η ισχύς που παράγεται από την κύρια μηχανή σε ένα πλοίο είναι ίση με την ισχύ που απαιτείται την δεδομένη χρονική στιγμή ώστε το πλοίο να εκτελεί την κίνηση του. Οι μηχανές εσωτερικής καύσης έχουν ως "στόχο" την διατήρηση της ισορροπίας και δεν μπορεί να ποτέ υπάρξει πλεόνασμα ισχύος. Αύξηση της παραγόμενης ισχύος π.χ. από επιλογή του χειριστή, συνεπάγεται και αύξηση των κινηματικών χαρακτηριστικών του πλοίου εφόσον όλες οι άλλες παράμετροι παραμένουν σταθερές. Αντίθετα αύξηση των ενεργειακών απαιτήσεων του πλοίου π.χ. αύξηση φορτίου ή αύξηση αντίθετου ανέμου, σημαίνει αντίστοιχη αύξηση της παραγόμενης ισχύος ώστε να επέλθει πάλι η ισορροπία.

Παραγόμενη ισχύς μηχανής = απαιτούμενη ισχύς πλοίου.

Με βάση αυτό επιλέχθηκε ως ζητούμενο πρόβλεψης στο πρόβλημα μηχανικής μάθησης που εξετάστηκε, η απαιτούμενη (ή παραγόμενη) ισχύς της μηχανής διότι αντικατοπτρίζει την συνολική κατάσταση του πλοίου.

Οι αλγόριθμοι παλινδρόμησης που παρουσιάστηκαν στο κεφάλαιο 3.1 επιστρατεύθηκαν για την πρόβλεψη της παραγόμενης ισχύς της κύριας μηχανής χρησιμοποιώντας την διαδικασία της μηχανικής μάθησης που περιγράφηκε στο κεφάλαιο 2.2.3 αλλά σε δύο στάδια. Αρχικά όλοι οι παραπάνω αλγόριθμοι κλήθηκαν να κάνουν πρόβλεψη τροφοδοτούμενοι με το πλήρες αρχικό σύνολο δεδομένων και εν συνεχεία εκτελέστηκε η ίδια διαδικασία με ένα τροποποιημένο σύνολο δεδομένων, το οποίο προκύπτει από το αρχικό μέσω μιας μεθοδολογίας που προτάσσεται από την παρούσα εργασία.

4.2 Προετοιμασία Δεδομένων / προ-επεξεργασία

Δειγματοληψία (Sampling). Αρχικά και μετά την συλλογή των δεδομένων, συγχρονίστηκαν ώστε να αφορούν συγκεκριμένες χρονικές περιόδους. Πέντε από τα πεδία τιμών του συνόλου δεδομένων που αντιστοιχούν στην ώρα (UTC), το έτος, τον μήνα, την ημέρα και την τοπική ώρα, αντικαταστάθηκαν με ένα πεδίο που αναπαριστά την ώρα ως Unix Timestamp (δευτερόλεπτα από την πρώτη Ιανουαρίου 1970). Με τον τρόπο αυτό διευκολύνεται ο χρονικός διαχωρισμός, εξαλείφθηκε οποιαδήποτε χρονική υστέρηση λήψεως ή καταγραφής των τιμών και πλέον όλες οι τιμές του συνόλου δεδομένων είναι φυσικοί αριθμοί. Η χρονική διάρκεια που απαιτείται για την καταγραφή και των 178 τιμών των παραμέτρων θεωρείται αμελητέα σε σχέση με την συχνότητα δειγματοληψίας επομένως οι τιμές που αναγράφονται σε κάθε γραμμή του συνόλου δεδομένων θεωρούνται σύγχρονες. Εξάλλου οι τιμές των φυσικών μεγεθών που ελέγχονται σε ένα πλοίο δεν παρουσιάζουν αξιοσημείωτες μεταβολές σε χρονικά διαστήματα μικρότερα του ενός λεπτού ώστε να υφίστανται κίνδυνος απώλειας πληροφορίας ή αλλοιώσεων αποτελεσμάτων.

Οποιαδήποτε δυσλειτουργία της διαδικασίας καταγραφής έχει ως αποτέλεσμα την απώλεια τιμής στο συγκεκριμένο πεδίο το οποίο εμφανίζεται κενό. Για τον εντοπισμό όλων των σημείων στα οποία δεν υπήρχαν τιμές εκτελέστηκαν τα εξής: μέσω ρουτίνας σε γλώσσα python προστέθηκε σε κάθε κελί ο αριθμός 0,00001. Ο αριθμός αυτός επιλέχθηκε επειδή είναι θετικός και πολύ μικρός σε σχέση με τις τιμές των μεγεθών που περιγράφονται στο σύνολο δεδομένων. Αυτό είχε σαν αποτέλεσμα κάθε κελί του πίνακα να αναπαριστά πλέον ένα φυσικό αριθμό. Σε δεύτερο χρόνο σε οποιοδήποτε τιμή μεγέθους ήταν μικρότερη από την τιμή 0,00001 αποδόθηκε το αναγνωριστικό NaN (non applicable).

Επιλογή Χαρακτηριστικών (Attribute selection). Οι παράμετροι που καταγράφονται από το σύστημα ADLM αναγνωρίστηκαν ως προς τί φυσικό μέγεθος περιγράφουν και τους αποδόθηκε η κατάλληλη μονάδα μέτρησης. Αφού

αναγνωρίστηκαν όλες οι μονάδες μέτρησης, τα δεδομένα χωρίστηκαν σε τέσσερις αναγνωρίστηκαν όλα τα μεγέθη χωρίστηκαν σε τέσσερις μεγάλες κατηγορίες. Οι κατηγορίες αυτές είναι: ΤΟΠΟΣ και ΧΡΟΝΟΣ, ΚΑΙΡΟΣ, ΠΡΟΩΣΗ και ΕΝΕΡΓΕΙΑ, ΚΙΝΗΜΑΤΙΚΑ ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ. Η κατηγοριοποίηση αυτή όλων των μεγεθών διευκόλυνε ιδιαίτερα την επιλογή των χαρακτηριστικών που θα διαμορφώσουν το σύνολο δεδομένων.

Στα προβλήματα μηχανικής μάθησης πολλοί ερευνητές αντιμετωπίζουν τα δεδομένα χωρίς να λαμβάνεται υπόψη η φυσική τους σημασία. Η θεώρηση αυτή βασίζεται στο γεγονός πως η πραγματική φυσική σημασία των μεγεθών "κρύβεται" μέσα τους και δεν αφορά την διαδικασία. Η γνώση που θα προκύψει εκ του αποτελέσματος της μάθησης θα την εμπεριέχει αλλά θα εκτελεστεί με αφανή για εμάς τρόπο. Τέτοια μοντέλα ονομάζονται Black Box Models (BBM) διότι βασίζονται σε στατιστικές διαδικασίες συμπερασμάτων βάσει της ιστορικής συλλογής δεδομένων και στα οποία ο εξωτερικός παρατηρητής δεν έχει ουσιαστικά καμία δυνατότητα επέμβασης επί των εσωτερικών διαδικασιών. Στην αντίπερα όχθη υπάρχουν τα White Box Models (WBM) τα οποία βασίζονται στη γνώση των φυσικών διεργασιών. Στα μοντέλα αυτά ο εξωτερικός παρατηρητής έχει πλήρη επίγνωση των εσωτερικών διαδικασιών και επηρεάζει τον τρόπο λειτουργίας τους. Ένα τέτοιο παράδειγμα θα ήταν η επίλυση του προβλήματος πρόβλεψης με χρήση μεθόδων παραδοσιακής μάθησης.

Στόχος της εργασίας είναι η διερεύνηση των χαρακτηριστικών αυτών που διαμορφώνουν την συμπεριφορά του πλοίου και χρήση αυτών στην πρόβλεψη αλλά με τρόπο όμως ανεξάρτητο του τύπου του πλοίου. Με χρήση Black Box Model στη διαδικασία επιλογής χαρακτηριστικών θα έπρεπε να επιλεγθούν προς χρήση τα χαρακτηριστικά εκείνα που μέσω στατιστικών εργαλείων θα έδειχναν ότι τείνουν να επηρεάσουν πιο σημαντικά την πρόβλεψη. Σε αυτή την περίπτωση το πρόβλημα αντιμετωπίζεται όπως ένα οποιοδήποτε άλλο πρόβλημα μάθησης. Αντίθετα εάν επιλεγεί ένα White Box Model στην επιλογή ή/και τροποποίηση των χαρακτηριστικών που θα εισαχθούν στο πρόβλημα μάθησης υπάρχει κίνδυνος να καταλήξουμε σε μία

πιο παραδοσιακή αντιμετώπιση η οποία δεν θα χαίρει ευρείας εφαρμογής. Το μοντέλο το οποίο επιλέχθηκε για την επιλογή των χαρακτηριστικών που θα διαμορφώσουν το σύνολο δεδομένων μπορεί να περιγράψει ως Gray Box Model (GBM). Στην επιλογή των χαρακτηριστικών έγινε η επιλογή εκείνων που, με βάση το φυσικό μέγεθος που αφορούν, θεωρήθηκε ότι είναι δυνατόν να επηρεάζουν το αντικείμενο της μελέτης (π.χ. η ταχύτητα του πλοίου). Τα attributes που κρίθηκαν πως δεν διαμορφώνουν καθόλου το αντικείμενο της μελέτης (π.χ. πίεση ελαίου ηλεκτρομηχανών) αφαιρέθηκαν, με σκοπό να μειωθεί ο συνολικός όγκος των δεδομένων και να διαμορφωθεί κατάλληλα το σύνολο δεδομένων (data set). Από τις 178 καταχωρίσεις επιλέχθηκαν 38.

Μετασχηματισμός Δεδομένων (Feature Engineering/date transformation). Στο πλαίσιο του περιορισμού των δεδομένων και με βάση την φυσική υπόσταση του υπό εξέταση φαινομένου είναι θεμιτός ο μετασχηματισμός των χαρακτηριστικών (Attributes) με σκοπό είτε τον περιορισμό του συνόλου δεδομένων είτε την αντικατάσταση κάποιων χαρακτηριστικών με έτερα που περιγράφουν καλύτερα την επίδραση στο φαινόμενο. Ένας από τους παράγοντες που επηρεάζουν την αντίσταση του πλοίου είναι η αντίσταση στον άνεμο. Η αντίσταση αυτή εξαρτάται από την μετωπική επιφάνεια του πλοίου καθώς και από την φαινόμενη ένταση του ανέμου που προσπίπτει σε αυτή. Για δεδομένη μετωπική επιφάνεια λοιπόν εξαρτάται από το διανυσματικό άθροισμα της ταχύτητας πλοίου και της διεύθυνσης του ανέμου. Με βάση αυτό οι στήλες (Wind Direction, Wind Speed) αντικαταστάθηκαν από την στήλη Head Wind Speed (ταχύτητα μετωπικού ανέμου). (Πολίτης, 2018). Ο μετασχηματισμός αυτός είναι ανεξάρτητος από την μορφολογία του πλοίου επομένως κρίθηκε πως μπορεί να εφαρμοστεί χωρίς να συγκεκριμενοποιήσει το πρόβλημα για τον υπό εξέταση τύπο πλοίου. Η αντικατάσταση των τιμών έγινε με χρήση των τύπων:

$$X = \text{Wind Direction} - \text{True Heading}$$

$$\text{If } X < -180 \text{ then } X' = 360 + X$$

$$\text{If } X > 180 \text{ then } X' = 360 - X$$

$$X'' = \text{ABS}(X')$$

$$\text{New instance} = \text{COS}(X'') * \text{Wind Speed}$$

Επιλογή Δεδομένων. Η διαχείριση των χαμένων τιμών είναι ένα από τα βασικά στάδια της προ επεξεργασίας των δεδομένων στην μηχανικά μάθηση. Αρχικά ως μέθοδος αντιμετώπισης επιλέχθηκε η πρόβλεψη της χαμένης τιμής (με αλγόριθμο 1-NN). Κατά την εφαρμογή της μεθόδου όμως και λόγω του πολύ μεγάλου όγκου των δεδομένων η διαδικασία πρόβλεψης χαμένων τιμών χρειάστηκε 16 ώρες για να ολοκληρωθεί με χρήση του ηλεκτρονικού υπολογιστή που χρησιμοποιήθηκε. (AMD Ryzen 5 4500U, 32GB RAM, Windows 10 Pro) Ο χρόνος αυτός θεωρήθηκε μη ρεαλιστικός για χρήση της μεθόδου σε πραγματικές συνθήκες και εγκαταλείφθηκε. Από τα δεδομένα της προηγούμενης κατηγορίας επιλέχθηκε η μέθοδος της διαγράψης γραμμής, επομένως έγινε διαγραφή όλων των γραμμών (Instances) οι οποίες περιείχαν έστω και μία χαμένη τιμή (NaN). Από τις 236161 οι γραμμές περιορίστηκαν στις 39475.

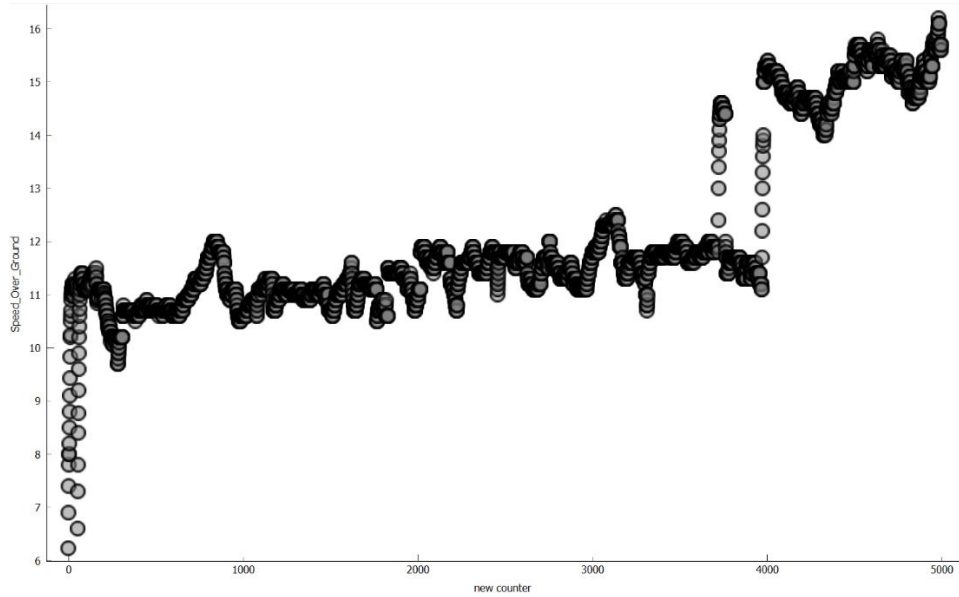
Από την βιβλιογραφία προκύπτει πως για τις ναυτικές μηχανές DIESEL υπάρχει ένα όριο ισχύος που δίδεται από τον κατασκευαστή, κάτω από το οποίο η μηχανή δεν πρέπει να λειτουργεί για μεγάλο χρονικό διάστημα. Η τιμή αυτή της ισχύος αντιστοιχεί στο 15-20% της μέγιστης συνεχούς (L1) για ηλεκτρονικά ελεγχόμενες μηχανές (electronically controlled) ή 20-25% για μηχανικά ελεγχόμενες (camshaft controlled). (Diesel, 2013) Όλες οι τιμές ισχύος λοιπόν κάτω από αυτό το όριο αντιστοιχούν σε καταστάσεις που σύμφωνα με τον κατασκευαστή καλό θα ήταν να αποφεύγονται για μεγάλα χρονικά διαστήματα.

Με βάση τα αποτελέσματα των δοκιμών αποδοχής του πλοίου (sea trials) προκύπτει ότι με το 25% της ισχύος της κύριας μηχανής επιτυγχάνονται περί τις 57 στροφές ανά λεπτό. Ο συσχετισμός αυτών ισχύος και στροφών άξονα αναφέρεται σε καινούργιο προωστήριο σκεύος. Λόγω της παλαιότητας του πλοίου και του γεγονότος πως κατά το διάστημα λήψεως των μετρήσεων δεν ήταν γνωστή η κατάσταση του, ο αριθμός αυτός περιορίστηκε στο 50 rpm. Αφαιρώντας τις γραμμές στις οποίες η τιμή

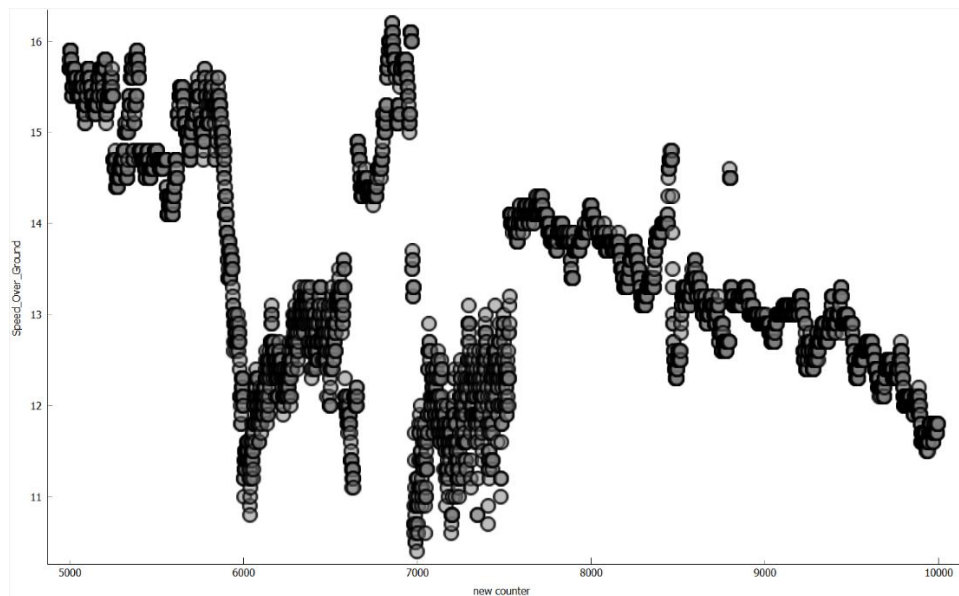
main engine RPM ήταν μικρότερη ίση των 50, προκύπτει το τελικό σύνολο δεδομένων με 29067 γραμμές. Τέλος αφαιρέθηκαν 3 γραμμές (counter:5830, 171377, 112996) στις οποίες η τιμή ισχύος μηχανής ήταν φανερά λανθασμένη σε σχέση με την τιμή στροφών της μηχανής (απόκλιση περίπου 50% σε σχέση με λοιπές τιμές συνόλου.

Αφαίρεση θορύβου / καθαρισμός δεδομένων (data cleansing). Όπως προαναφέρθηκε οι ταχύτητες πραγματοποίησης των πλου είναι συγκεκριμένες. Σε αυτές τις καταστάσεις με σταθερή ταχύτητα επέρχεται μια ισορροπία και το παραγόμενο έργο από την κύρια μηχανή πρόωσης αντιστοιχεί ακριβώς στις απαιτήσεις του πλοίου. Κατά την διάρκεια επιταχύνσεων ή επιβραδύνσεων, όπου δηλαδή έχουμε δραστηκές αλλαγές ταχυτήτων η ισορροπία αυτή διαταράσσεται. Σε καταστάσεις όπου το πλοίο βρίσκεται σε διαδικασίες χειρισμών είναι πιθανό και λόγω της συχνής αλλαγής στροφών μηχανής να μην προλαβαίνει το σύστημα προώσεως να ισορροπεί. Κατά την επιτάχυνση μέρος της παραγόμενης ισχύος απαιτείται για την αύξηση της ταχύτητας και όχι για τα κινηματικά χαρακτηριστικά που αποτυπώνονται εκείνη τη στιγμή στις μετρήσεις. Αντίστοιχα κατά την επιβράδυνση η παραγόμενη ισχύς μπορεί να είναι ακόμα και μηδενική ενώ το πλοίο κινείται ακόμη.

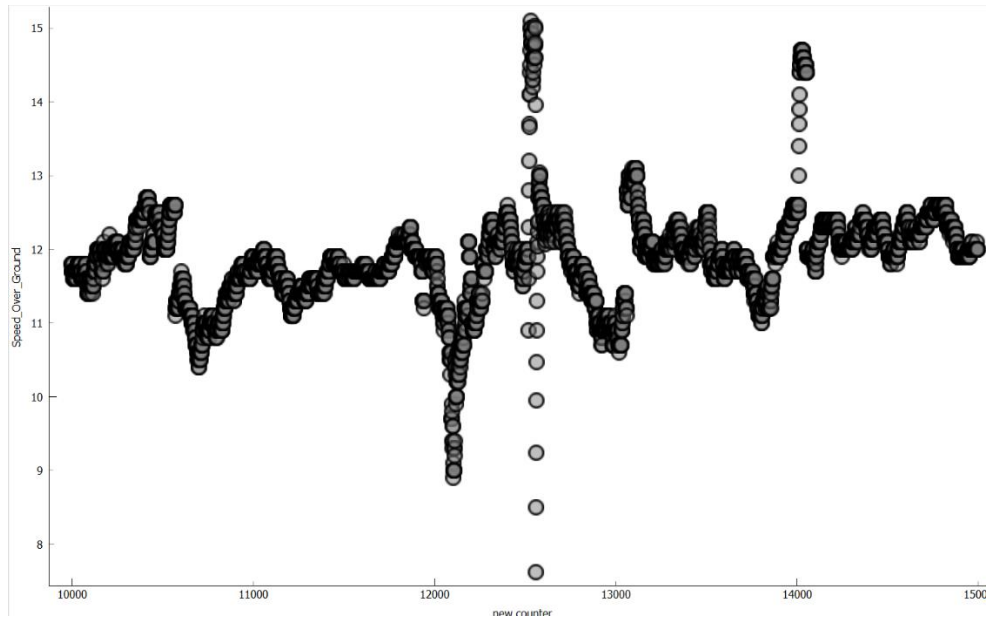
Στο πλαίσιο της αφαίρεσης των περιόδων τιμών που κρίθηκαν ως μεταβατικές με αποτέλεσμα να προσθέτουν θόρυβο στο σύστημα και δεν περιγράφουν μία σταθερή κατάσταση (π.χ. διαδικασία επιτάχυνσης) τα δεδομένα επαναριθμήθηκαν (διατηρώντας την χρονολογική τους σειρά) και χωρίστηκαν σε διαστήματα ίσης συχνότητας (τα διαστήματα αυτά έχουν ίδιο πλήθος τιμών, 5000 για την προκειμένη περίπτωση) για διευκόλυνση του υπολογιστικού βάρους και με βάση τους περιορισμούς του εν χρήση λογισμικού (Orange data minning). Στα διαστήματα αυτά και με την μέθοδο K-means αναγνωρίστηκαν τα συμπλέγματα τιμών (πάντα ως προς την πραγματική ταχύτητα του πλοίου, μέγεθος που θεωρήθηκε θεμελιώδες ως προς την μεταβατικότητα του υπό εξέταση φαινομένου). Στις εικόνες 3 έως 8 φαίνεται η διασπορά των τιμών των ταχυτήτων σε σχέση με την καινούργια αρίθμηση των γραμμών δεδομένων.



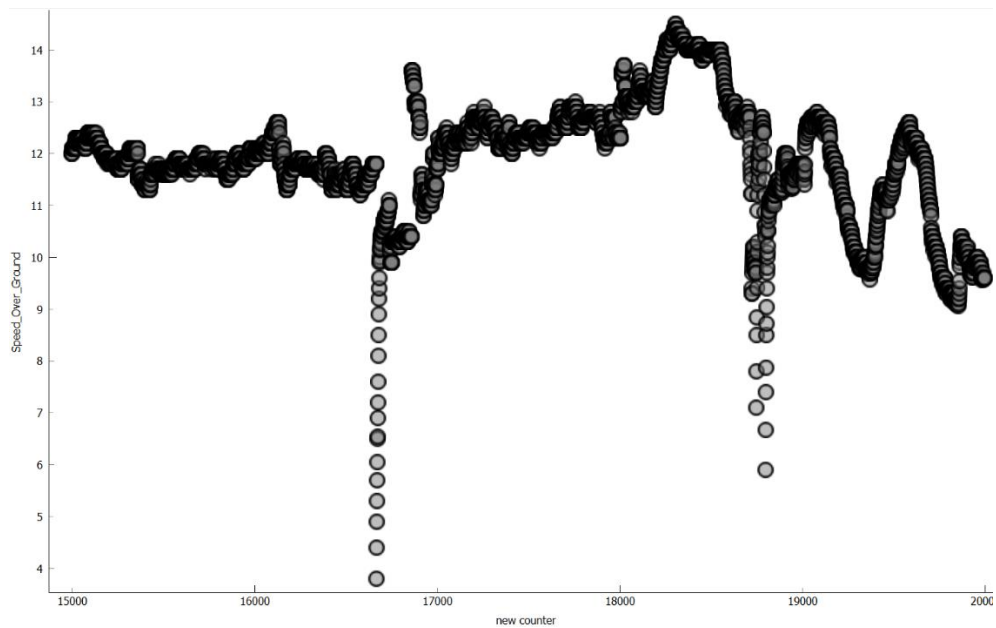
Εικ3 Διασπορά ταχυτήτων για γραμμές 1-4999



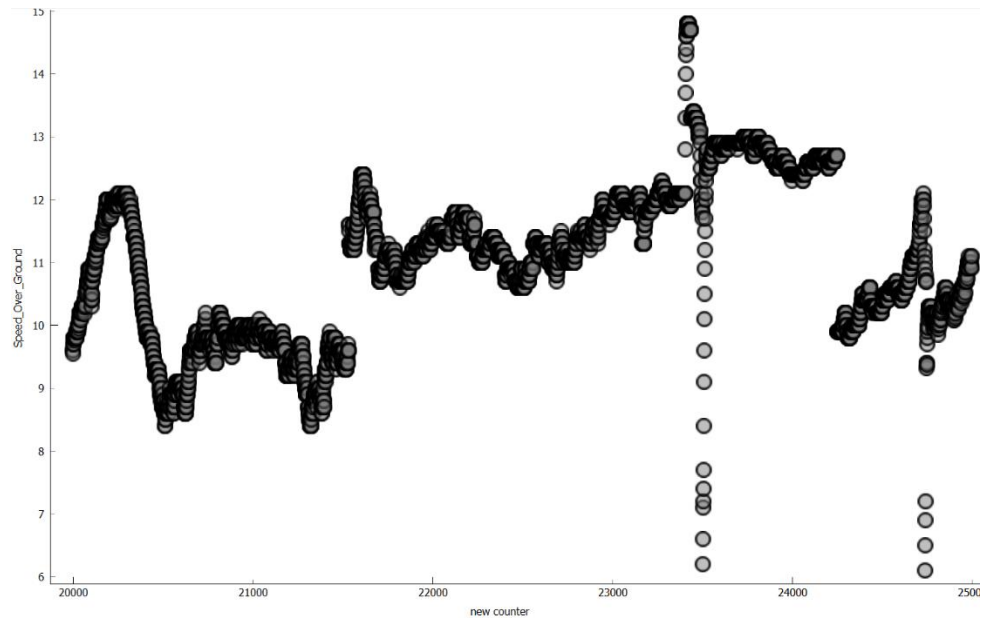
Εικ4 Διασπορά ταχυτήτων για γραμμές 5000-9999



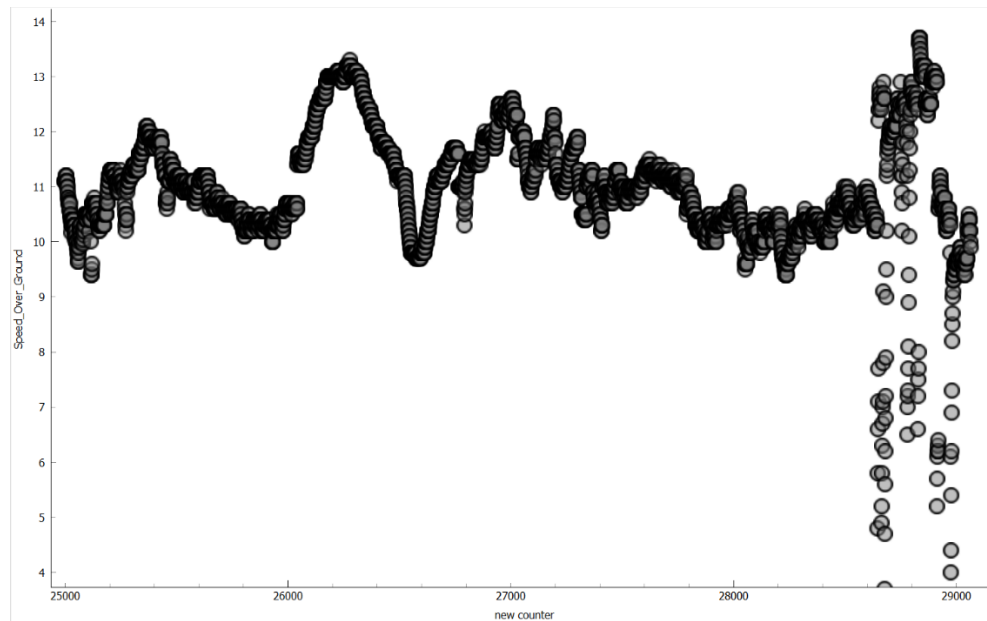
Εικ5 Διασπορά ταχυτήτων για γραμμές 10000-14999



Εικ6 Διασπορά ταχυτήτων για γραμμές 15000-19999



Εικ7 Διασπορά ταχυτήτων για γραμμές 20000-24999

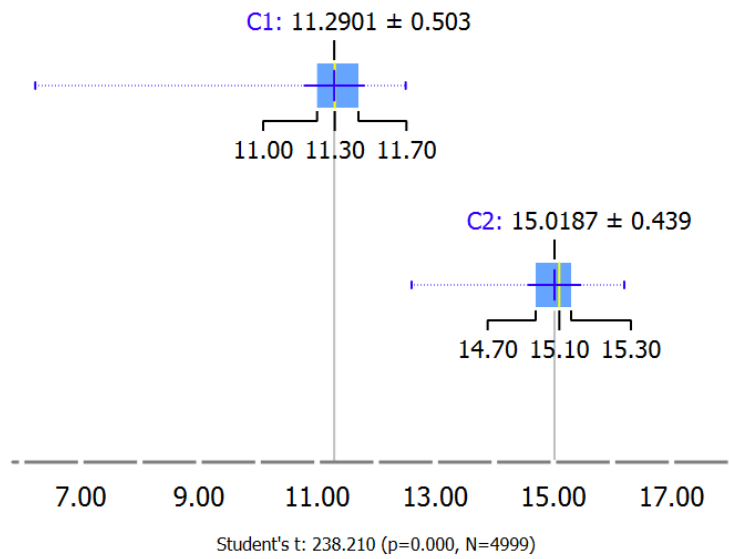


Εικ8 Διασπορά ταχυτήτων για γραμμές 25000-29067

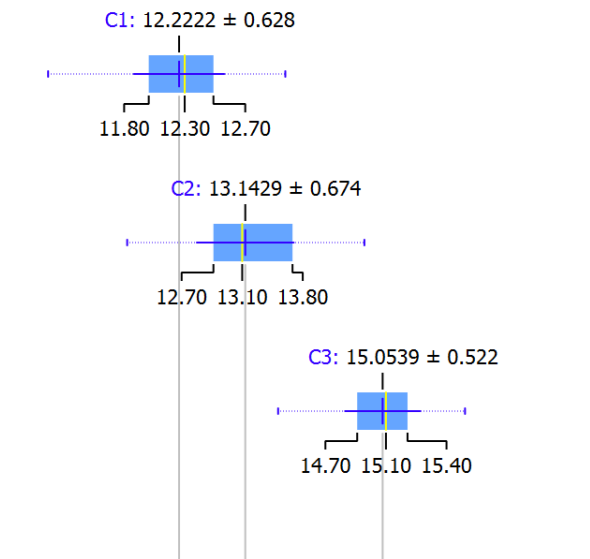
Ως ταχύτητα χρησιμοποιήθηκε η τιμή του πεδίου speed over ground που αντιστοιχεί στην ταχύτητα πλοίου ως προς το έδαφος (και γενικά ως προς οποιονδήποτε εξωτερικό ακίνητο παρατηρητή) και όχι η ταχύτητα through water που

αντιστοιχεί στην ταχύτητα του πλοίου σε σχέση με το οριακό στρώμα νερού δίπλα του.

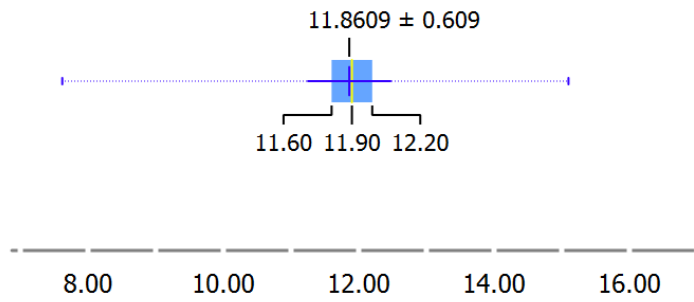
Στις εικόνες 9 έως 14 φαίνονται οι μέσες τιμές των ταχυτήτων σε κάθε ένα από τα συμπλέγματα τιμών καθώς και η διασπορά τους.



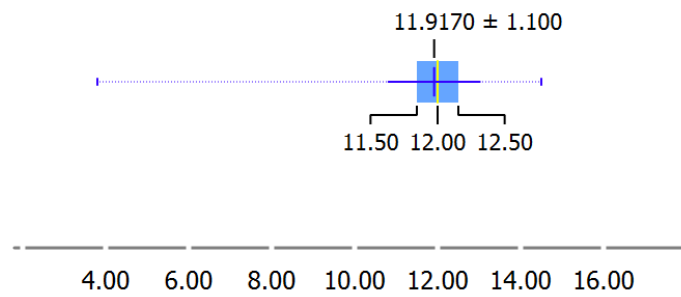
Εικ 9 Μέση τιμή ταχύτητας και διασπορά 1-4999



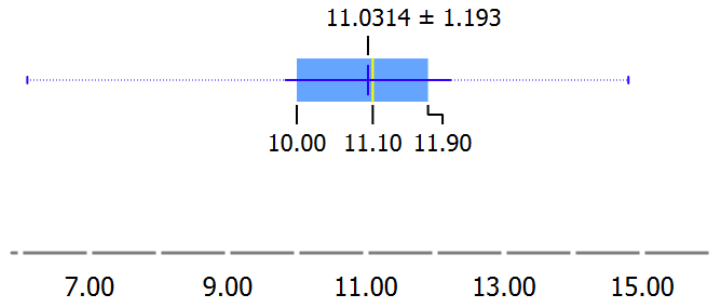
Εικ 10 Μέση τιμή ταχύτητας και διασπορά 5000-9999



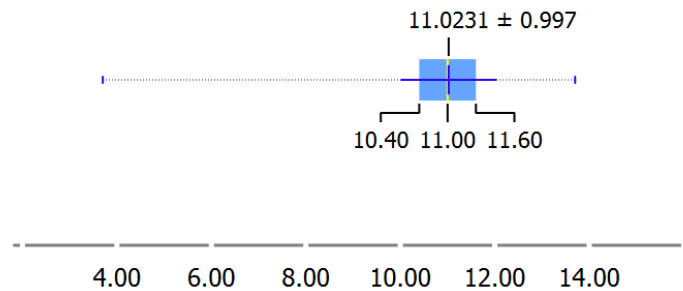
Εικ 11 Μέση τιμή ταχύτητας και διασπορά 10000-14999



Εικ 12 Μέση τιμή ταχύτητας και διασπορά 15000-19999



Εικ 13 Μέση τιμή ταχύτητας και διασπορά 20000-29067



Εικ 14 Μέση τιμή ταχύτητας και διασπορά 25000-24999

Στο σύνολο των συμπλεγμάτων αυτών οι τιμές ταχύτητας (μέση τιμή +- διασπορά) βρίσκονταν μεταξύ 9.8384 η ελάχιστη και 15.5759 η μέγιστη. Οι γραμμές με τιμές ταχύτητας έξω από αυτό το πεδίο αναγνωρίστηκαν ως θόρυβος διότι θεωρήθηκε ότι αντιστοιχούν σε λάθος καταχωρήσεις είτε περιγράφουν μεταβατικά στάδια και αφαιρέθηκαν από το σύνολο δεδομένων. Το τελικό σύνολο δεδομένων αποτελείται από 37 στήλες (attributes) και 27155 γραμμές(instances).

Κανονικοποίηση (normalization). Τα Τεχνητά Νευρωνικά Δίκτυα επωφελούνται ιδιαίτερος εάν οι τιμές εισόδου βρίσκονται στην κλίμακα 0 έως 1.

(Maynard, 2020) Για την επίτευξη αυτού του σκοπού το σύνολο δεδομένων κανικοποιήθηκε σε τιμές 0 έως 1 με χρήση του παρακάτω τύπου.

$$x_{\text{norm}} = \frac{x - \min(x)}{\max(x) - \min(x)} (1 - 0)$$

Όπου x_{norm} = η νέα τιμή μετά την κανονικοποίηση

x = η αρχική τιμή

$\min(x)$ = η ελάχιστη τιμή προ κανονικοποίησης

$\max(x)$ = η μέγιστη τιμή προ κανονικοποίησης

4.3 Εξόρυξη Γνώσης

Με έτοιμο το προ επεξεργασμένο σύνολο δεδομένων, το επόμενο στάδιο ήταν ο διαχωρισμός τους. Τα δεδομένα χωρίστηκαν στο σύνολο εκπαίδευσης (training data set) που αποτελείται από το 80% των γραμμών και το σύνολο επικύρωσης ή δοκιμής (validation data set) που αποτελείται από το υπόλοιπο 20% των γραμμών. Ο διαχωρισμός αυτός εκτελέστηκε με τυχαίο τρόπο από το λογισμικό που χρησιμοποιήθηκε (Rapid Miner Studio). Αντίστοιχη διαδικασία εκτελέστηκε και στο αρχικό σύνολο δεδομένων.

Κάνοντας χρήση του αρχικού μη προ επεξεργασμένου συνόλου δεδομένων "τροφοδοτήθηκαν" οι εξής αλγόριθμοι:

Γραμμική Παλινδρόμηση (Linear Regression)

Τα Δέντρα Απόφασης (Decision Tree)

Οι k-πλησιέστεροι γείτονες (k-Nearest Neighbors)

Τα Δάση Τυχαίας Απόφασης (Random Forest)

Αλγόριθμος Ωθησης (AdaBoost)

Τα Τεχνητά Νευρωνικά Δίκτυα (ANN)

Ως δεδομένο στόχος ορίστηκε η ισχύς της κύριας μηχανής και ξεκίνησε η διαδικασία εκπαίδευσης καθώς και αξιολόγησης των μοντέλων. Οι βασικές παράμετροι της λειτουργίας κάθε αλγορίθμου τροποποιούνταν και γινόταν έλεγχος κάθε φορά της επίδρασης που αυτό είχε στις προβλέψεις της τιμής της ανεξάρτητης μεταβλητής (στόχο που είχε δοθεί). Με την εκτέλεση αρκετών πειραμάτων και με διαδικασία trial and error βρέθηκαν οι τιμές παραμέτρων σε κάθε αλγόριθμο που έδιναν το καλύτερο αποτέλεσμα. Η διαδικασία εκπαίδευσης εκτελέστηκε κάθε φορά στον ίδιο ακριβώς ηλεκτρονικό υπολογιστή και δόθηκε ιδιαίτερη μέριμνα ώστε οι συνθήκες κάτω από τις οποίες πραγματοποιούνταν να παραμένουν όσο το δυνατόν όμοιες. (ίδιες θερμοκρασιακές συνθήκες για αποφυγή Thermal throttling cpu, εκκίνηση διαδικασίας με ακριβώς τις ίδιες διεργασίες να εκτελούνται στο παρασκήνιο)

Τα αποτελέσματα αξιολογούνταν κάθε φορά μέσω των εργαλείων μέτρησης αποτελεσμάτων που αναλύθηκαν στο κεφ. 3.2. (MAE, MSE, RMSE, R^2 , CVRMSE). Ταυτόχρονα υπήρχε καταγραφή του χρόνου που ήταν απαραίτητος για να ολοκληρωθεί η διαδικασία εκπαίδευσης διότι λόγω της χρήσεως ίδιου ηλεκτρονικού υπολογιστή κρίθηκε πως αποτελεί συγκριτικό εργαλείο της απαιτούμενης επεξεργαστικής ισχύος κάθε αλγορίθμου.

Οι ρυθμίσεις οι οποίες απέδωσαν τα καλύτερα αποτελέσματα για κάθε περίπτωση ήταν οι ακόλουθες:

k-πλησιέστεροι γείτονες (k-Nearest Neighbors):

Βέλτιστος αριθμός clusters 5 ($k=5$)

Δάσος Τυχαίας Απόφασης (Random Forest):

Αριθμός δένδρων που απαρτίζουν το δάσος = 60

Αριθμός μη περαιτέρω διαχωρισμού υποσυνόλων = 2

Τα Δέντρα Απόφασης (Decision Tree):

Αριθμός διαχωρισμού υποσυνόλων = 2

Αριθμός χαρακτηριστικών σε κλαδιά = 7

Αλγόριθμος Ώθησης (AdaBoost):

Αριθμός ταξινομητών = 50

Ρυθμός μάθησης = 1

Αλγόριθμος = SAMME.R

Συνάρτηση παλινδρόμησης = Γραμμική

Τεχνητό Νευρωνικό Δίκτυο (ANN):

Νευρώνες κρυφού επιπέδου = 100

Συνάρτηση ενεργοποίησης = ReLu

Επίλυση = Adam

Επαναλήψεις = 1000

Με αυτές τις διαμορφώσεις σε κάθε περίπτωση εκτελέστηκε πρόβλεψη της απαιτούμενης ισχύος της κύριας μηχανής πλοίου στο σύνολο δεδομένων ελέγχου. Μόλις ολοκληρώθηκε η διαδικασία εκτελέστηκαν ακριβώς τα ίδια βήματα χρησιμοποιώντας πλέον το προ επεξεργασμένο σύνολο δεδομένων. Οι συνθήκες εκτελέσεων της εκπαίδευσης και του ελέγχου όλων των αλγορίθμων παρέμειναν ίδιες με την προηγούμενη περίπτωση ώστε να εξασφαλιστεί η εγκυρότητα του συγκριτικού αποτελέσματος.

4.4 Αποτελέσματα προβλέψεων / μέτρηση επιδόσεων

Τα αποτελέσματα των επιδόσεων όλων των αλγορίθμων με χρήση του αρχικού συνόλου ελέγχου (Raw Data) φαίνονται στην εικόνα 15.

Raw Data						
Model	MSE	RMSE	MAE	R2	CVRMSE	Remarks
Random Forest	298.902	17.289	2.421	1.000	0.293	number of trees 60 do not split subnets 2
kNN	31590.858	177.738	72.709	0.998	3.010	k=5
Linear Regression	240008597722430000.000	489906723.492	2329117.484	-16376545662.923	8296814.438	
Neural Network	111588.711	334.049	233.929	0.992	5.657	neurons hidden 100 activation tanh solver SGD regularization 2 iterations 160
Decision Tree	639.464	25.288	4.957	1.000	0.428	instances in leaves 7 split subnets 2
Adaboost	356.969	18.894	4.105	1.000	0.320	

Εικ 15 Αποτελέσματα με χρήση αρχικού συνόλου δεδομένων ελέγχου

Αντίστοιχα τα αποτελέσματα με την χρήση του προ επεξεργασμένου συνόλου ελέγχου (selected data) φαίνονται στην εικόνα 16.

Selected Data						
Model	MSE	RMSE	MAE	R2	CVRMSE	Remarks
Random Forest	4542.351	67.397	41.162	0.999	0.966	number of trees :60 do not split subnets :2
kNN	6126.236	78.27	43.311	0.999	1.258	k=5
Linear Regression	12666.868	112.547	84.931	0.998	1.614	
Neural Network	6096.33	78.079	84.931	0.999	1.12	neurons hidden :100,200 activation :ReLu solver :Adam regularization :2 iterations :1000
Decision Tree	7702.947	87.766	53.181	0.999	1.122	instances in leaves :7 split subnets :2
Adaboost	4116.474	64.16	39.196	0.999	0.92	Number of estimators :50 Learning Rate :1 Classification algorithm :SAMME.R Regression loss function :Linear

Εικ 16 Αποτελέσματα με χρήση προ επεξεργασμένου συνόλου δεδομένων ελέγχου

Τα αποτελέσματα δείχνουν πως η συντριπτική πλειοψηφία των αλγορίθμων πέτυχε αποτελέσματα τα οποία είναι πολύ ενθαρρυντικά για τις δυνατότητες

χρήσης τους σε προβλήματα πρόβλεψης λειτουργικών παραμέτρων πλοίου. Ο δείκτης R^2 παρέμεινε σταθερά κοντά στην μονάδα σε όλες τις περιπτώσεις δεικνύοντας πως οι αλγόριθμοι ήταν σε θέση να "δικαιολογήσουν" σε πολύ μεγάλο βαθμό τις διακυμάνσεις τιμών της αποδιδόμενης ισχύος του πλοίου. Ο αλγόριθμος kNN καθώς και η γραμμική παλινδρόμηση ήταν οι δύο περιπτώσεις που ουσιαστικά απέτυχαν να δώσουν ποιοτικά αποτελέσματα όταν έγινε χρήση του αρχικού συνόλου γεγονός που αναστράφη όταν οι ίδιοι αλγόριθμοι έκαναν χρήση του προ επεξεργασμένου συνόλου.

Οι απόλυτα καλύτερες επιδόσεις με κριτήριο την ακρίβεια προβλέψεων επιτεύχθηκαν από το Δάσος Τυχαίας Απόφασης και με χρήση του αρχικού συνόλου. Η τιμή του MAE της τάξης του 2.421 είναι ενδεικτική. Το τεχνητό νευρωνικό δίκτυο επωφελήθηκε ιδιαίτερα από την διαδικασία της προ επεξεργασίας ενώ το δέντρο απόφασης καθώς και ο αλγόριθμος ώθησης μείωσαν ελάχιστα τις επιδόσεις τους.

Ένα δεύτερο πολύ σημαντικό κριτήριο αξιολόγησης των επιδόσεων αλγορίθμων μηχανικής μάθησης είναι η απαιτούμενη επεξεργαστική ισχύς. Ένας αλγόριθμος μπορεί να επιτυγχάνει πολύ ακριβή αποτελέσματα πρόβλεψης αλλά οι απαιτήσεις σε ισχύ να είναι τέτοιες που να καθιστούν την χρήση του ασύμφορη. Αντίστοιχα είναι πιθανό να απαιτείται τόσο υψηλός χρόνος ολοκλήρωσης των διεργασιών του που να μην μπορεί εμπράκτως να χρησιμοποιηθεί.

Ένα άλλο πολύ σημαντικό στοιχείο αξιολόγησης των αποτελεσμάτων είναι η μέγιστη απόκλιση των προβλεπόμενων τιμών. Σε πραγματικές συνθήκες και με γνώμονα την πρόθεση αξιοποίησης τους, τα εργαλεία μέτρησης ακρίβειας των αλγορίθμων δεν μπορούν να μας καλύψουν πλήρως. Όλα τα εργαλεία μέτρησης αντιμετωπίζουν τις αποκλίσεις συνολικά. Σε ένα μεγάλο εύρος τιμών πρόβλεψης μας παρέχουν τη δυνατότητα σύγκρισης της μέσης τιμής. Στην πραγματικότητα όμως σε ένα πρόβλημα αναζήτησης π.χ. δέκα τιμών, ίσως είναι χειρότερο να έχουμε μία πολύ λανθασμένη τιμή και τις υπόλοιπες εννέα πολύ κοντά στις πραγματικές, από το να έχουμε και τις δέκα τιμές εσφαλμένες μεν αλλά με μικρή απόκλιση. Στον εικόνα 16 παρουσιάζονται τα αποτελέσματα των αλγορίθμων με βάση τα δύο τελευταία αυτά κριτήρια.

Selected Data							
Model	True Value At Max Fault	Calculated Value At Max Fault	Max Difference	Max Percentage Difference	Percentage Of Values With Below 3% Error	Train Time (sec)	Test Time (sec)
Random Forest	6945	6183	762	10.97%	99.20%	2.70	0.03
kNN	6723	4973	1750	26.03%	99.00%	0.27	0.53
Linear Regression	5750	6853	1103	19.18%	95.54%	0.09	0.01
Neural Network	3955	6635	2680	67.76%	99.18%	96.02	0.08
Decision Tree	8286	9128	842	10.16%	98.40%	45.35	0.01
Adaboost	6945	6202	743	10.69%	99.31%	23.79	0.18
Raw Data							
Model	True Value At Max Fault	Calculated Value At Max Fault	Max Difference	Max Percentage Difference	Percentage Of Values With Below 3% Error	Train Time (sec)	Test Time (sec)
Random Forest	6290	3803	2487	39.54%	99.76%	1168.78	1.02
kNN	9463	2781	6682	70.61%	95.90%	19.49	25.13
Linear Regression	11245	NA	NA	NA	97.40%	5.65	0.43
Neural Network	4716	-1036	5752	121.97%	69.82%	83.97	0.68
Decision Tree	6290	4066	2224	35.36%	99.59%	684.48	0.02
Adaboost	6290	3473	2817	44.79%	99.71%	1253.63	4.12

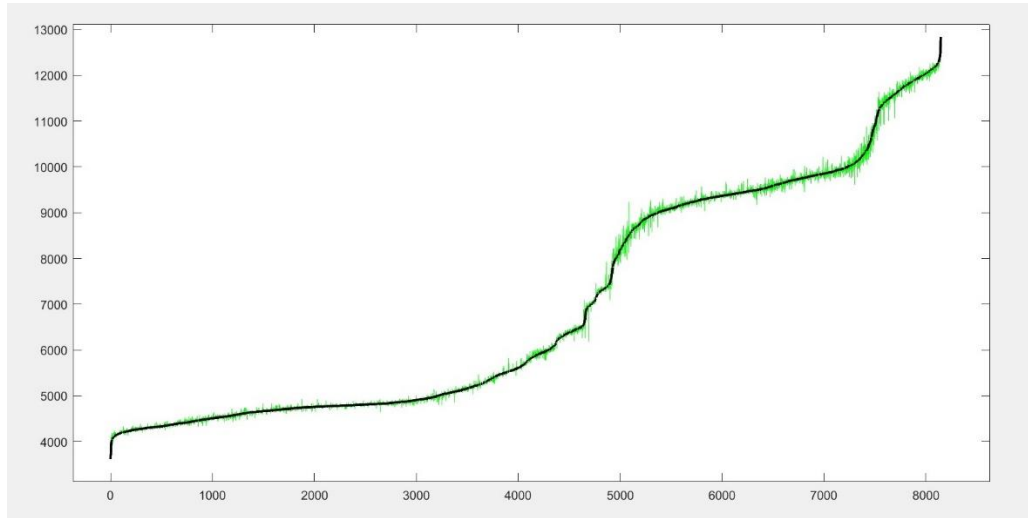
Εικ 16 Αποτελέσματα επίτευξης χρόνου ολοκλήρωσης / λανθασμένων τιμών

Ως κατώφλι ποσοστού απόκλισης λανθασμένης πρόβλεψης στις τιμές ισχύος της κύριας μηχανής τέθηκε η τιμή του 3%. Η επιλογή αυτή έγινε διότι αποτελεί ένα γενικά αποδεκτό όριο σφάλματος στην ναυτιλία καθώς και σε μη εξειδικευμένα (και συναφώς πολύ ακριβά) συστήματα μετρήσεων.

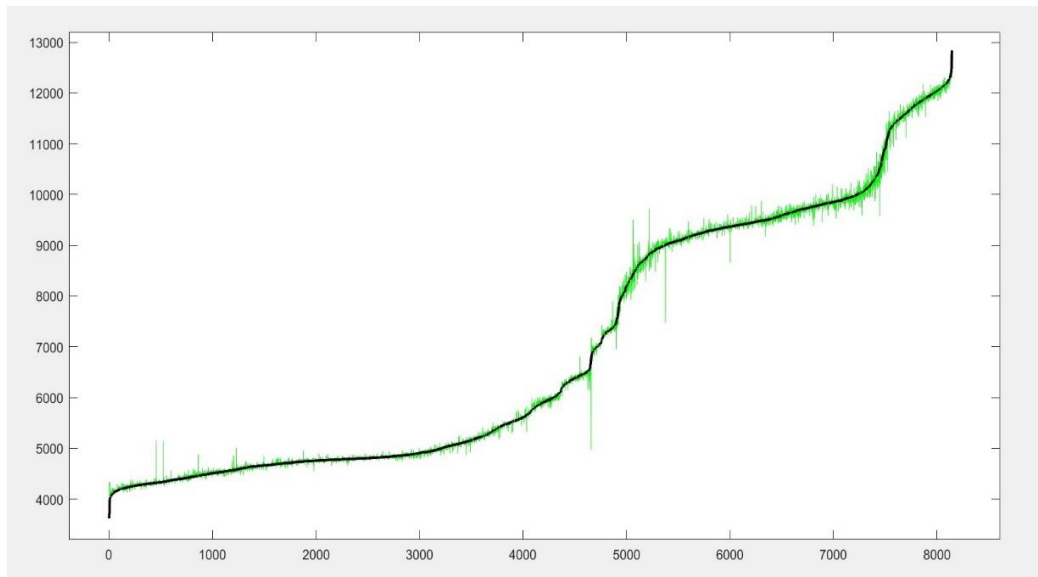
Από τα αποτελέσματα προκύπτει ότι παρόλο που η χρήση του επεξεργασμένου συνόλου δεδομένων δεν είχε σαφή ωφέλεια στην ακρίβεια των δεδομένων εάν αυτά αξιολογηθούν με εργαλεία μετρήσεων, εδώ η κατάσταση που παρουσιάζεται είναι πολύ διαφορετική. Οι χρόνοι εκπαίδευσης καθώς και ελέγχου σχεδόν σε όλες τις περιπτώσεις είναι δραματική. Με εξαίρεση το τεχνητό νευρωνικό δίκτυο όλοι οι άλλοι χρόνοι περιορίστηκαν σε πολύ μεγάλα ποσοστά. Οι χρόνοι εκπαίδευσης είναι της τάξης των δευτερολέπτων καθιστώντας την διαδικασία εύκολα εκτελέσιμη σε πραγματικούς χρόνους. Ενδεικτικά να αναφέρουμε πως στην περίπτωση του Δάσους Τυχαίας Απόφασης το οποίο επέτυχε και το καλύτερο αποτέλεσμα σύμφωνα με τα εργαλεία μετρήσεως επιδόσεων, ο χρόνος εκπαίδευσης με χρήση αρχικού συνόλου δεδομένων είναι 43159.3% πιο αυξημένος σε σχέση με την χρήση του προ επεξεργασμένου συνόλου.

Όσο αναφορά την μέγιστη απόκλιση των τιμών ισχύος που προβλέφθηκαν σε απόλυτα νούμερα βλέπουμε επίσης την σημαντική επίδραση της προ επεξεργασίας δεδομένων. Η καλύτερη επίδοση επιτυγχάνεται από τον αλγόριθμο AdaBoost του οποίου η απόκλιση της προβλεφθείσας τιμής στην χειρότερη των περιπτώσεων δεν υπερέβηκε τα 743 kW κατά απόλυτη τιμή ήτοι 10.16%. Οι αντίστοιχες καλύτερες επιδώσεις με χρήση αρχικού συνόλου δεδομένων προήλθαν από το δένδρο απόφασης με τιμές 2224 kW ήτοι 35,36%.

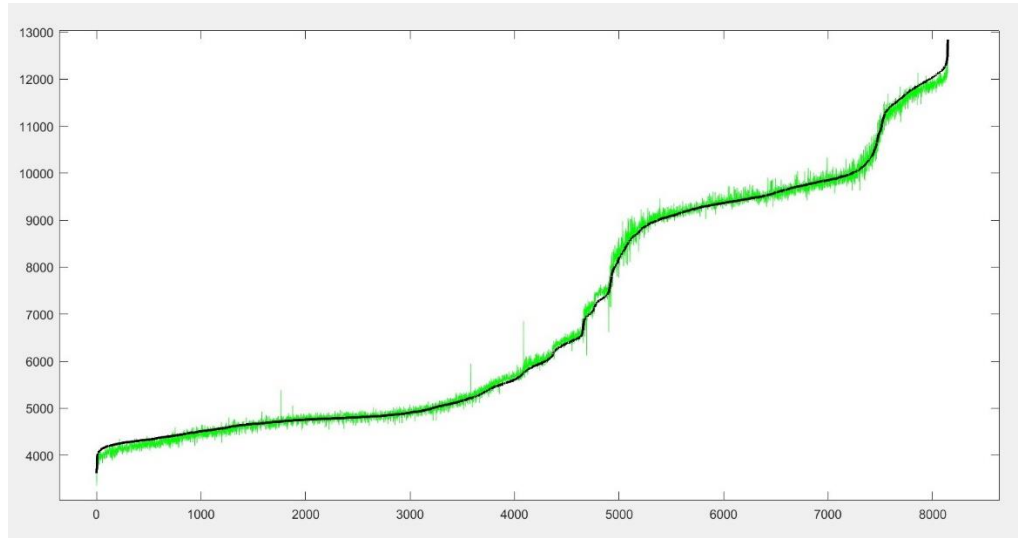
Το μέγιστο ποσοστό των τιμών που στο σύνολό τους δεν υπερέβησαν το προκαθορισμένο όριο του 3% σφάλματος ήταν 99.76% με χρήση αλγορίθμου Δάσους Τυχαίας Απόφασης στο αρχικό σύνολο δεδομένων, ποσοστό που μειώθηκε στο 99,31% όταν χρησιμοποιήθηκε το προ επεξεργασμένο σύνολο και επιτεύχθηκε από τον αλγόριθμο ώθησης (AdaBoost). Στις εικόνες 17 έως 22 παρουσιάζονται οι προβλεφθείσες τιμές ισχύος όλων των αλγορίθμων και με χρήση του προ επεξεργασμένου συνόλου δεδομένων ελέγχου (πράσινο χρώμα) σε σχέση με τις πραγματικές ισχύος της κύριας μηχανής (μαύρο χρώμα).



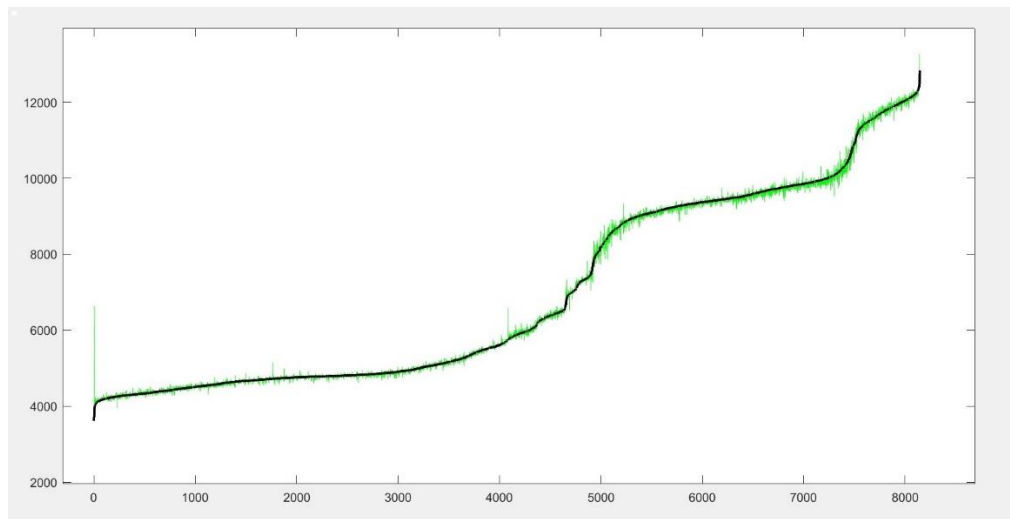
Εικ 17 προβλεφθείσες τιμές Random Forest



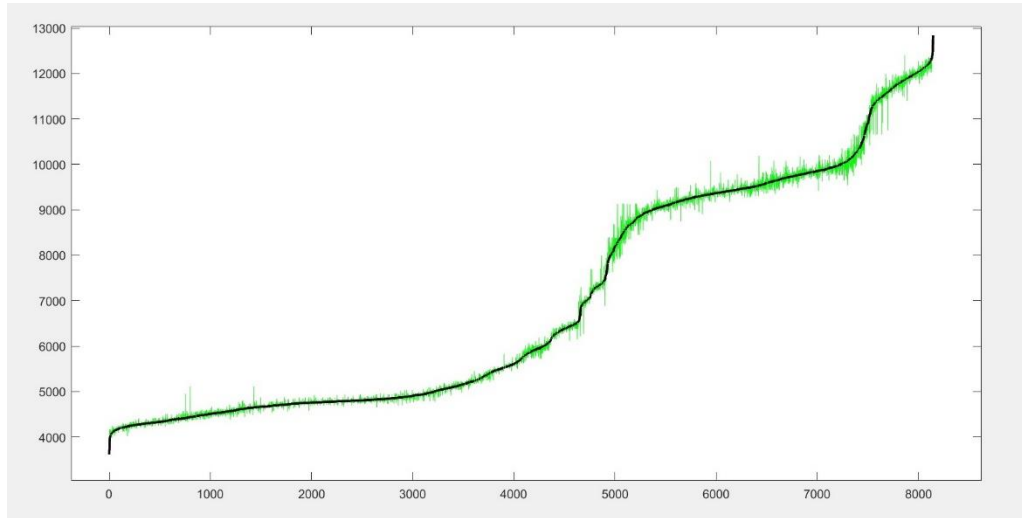
Εικ 18 προβλεφθείσες τιμές kNN



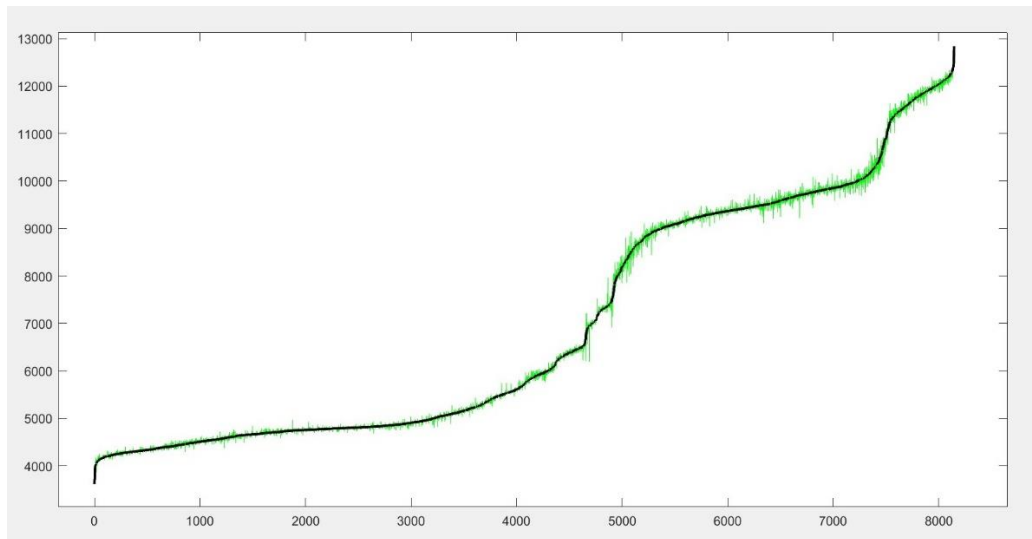
Εικ 19 προβλεφθείσες τιμές Linear Regression



Εικ 20 προβλεφθείσες τιμές Neural Network



Εικ 21 προβλεφθείσες τιμές Decision Tree



Εικ 22 προβλεφθείσες τιμές Adaboost

5.ΣΥΜΠΕΡΑΣΜΑ

5.1 Συμπεράσματα

Τα αποτελέσματα των επιδόσεων στην πρόβλεψη των τιμών της αποδιδόμενης ισχύος ενός πλοίου κρίνονται ως πολύ ικανοποιητικά. Η χρήση των πιο διαδεδομένων αλγορίθμων παλινδρόμησης μέσω μίας διαδικασίας μηχανικής μάθησης μπορεί να προβλέψει με αρκετή ακρίβεια τις τιμές ενός τόσο καθοριστικού μεγέθους όπως η ισχύς. Χρησιμοποιώντας τιμές που αποθηκεύονται στα συστήματα ελέγχου και καταγραφής, που είναι κοινός τύπος σε όλα τα νέα πλοία, μπορούμε να κάνουμε ρεαλιστικές προβλέψεις για την συμπεριφορά τους. Η αξιοποίηση των αποτελεσμάτων είναι εφικτό να συνδράμει καθοριστικά στις βασικές και διαρκείς επιδιώξεις της ναυτιλίας όπως αυτές ορίστηκαν:

α) Στη μείωση του κόστους εκτέλεσης των μεταφορών μέσω π.χ. την πρόβλεψη απαιτούμενης ισχύος, και επομένως την κατανάλωση καυσίμου, για διάφορες καταστάσεις που μπορεί να αντιμετωπίσει ένα πλοίο και την επιλογή της καλύτερης (route optimization)

β) Στην μείωση του ενεργειακού αποτυπώματος μέσω π.χ. τον διαρκή έλεγχο της κατάστασης των μηχανών / ηλεκτρομηχανών και την έγκαιρη διάγνωση οποιασδήποτε δυσλειτουργίας που μπορεί να την επηρεάσει (Condition Based Maintenance)

Το σύνολο δεδομένων που επιδρούν καθοριστικά στην λειτουργική συμπεριφορά ενός πλοίου προκύπτει από την γνώση των φυσικών διεργασιών που εκτελούνται, αλλά μπορούν να αποτελέσουν τμήμα μιας αντιμετώπισης Gray Box Model στο οποίο οι τεχνικές παραδοσιακής μάθησης (κλασικές επιστήμες) δεν συνδράμουν επιπλέον. Ο εντοπισμός τους και η χρήση σε μοντέλα πρόβλεψης επιτυγχάνει την δραματική μείωση της απαιτούμενης επεξεργαστικής ισχύος χωρίς να επηρεάζει την συνολική ακρίβεια των προβλέψεων. Το γεγονός αυτό επιτρέπει την

αξιοποίηση τέτοιων μεθόδων και σε μικρότερα ή παλαιότερα σκάφη που έχουν πρόσβαση σε περιορισμένους οικονομικούς πόρους. Η δυνατότητα δημιουργίας ενός μοντέλου που με κατάλληλη επιλογή των παραμέτρων, και μέσω μιας συγκεκριμένης μεθοδολογίας προ επεξεργασίας τους, είναι σε θέση να επιτυγχάνει αντίστοιχες επιδώσεις πρόβλεψης με τα ολοκληρωμένα συστήματα που χρησιμοποιούν μεγάλα πλοία είναι το κλειδί για την ευρεία υιοθέτηση της μηχανικής μάθησης στην ναυτιλία.

Το τελευταίο αλλά εξίσου σημαντικό συμπέρασμα είναι πως η όλη διαδικασία μηχανικής μάθησης που εκτελέστηκε είναι ανεξάρτητη του τύπου ή του μεγέθους του πλοίου στο οποίο εφαρμόστηκε. Το γεγονός αυτό την καθιστά ικανή να χρησιμοποιηθεί σε κάθε τύπο πλοίου με αντίστοιχα αποτελέσματα.

5.2 Περιορισμοί

Η έρευνα εκτελέστηκε στο πλαίσιο εκπόνησης διπλωματικής διατριβής του γράφοντα και είχε σαφή χρονικά περιθώρια ολοκλήρωσης. Στο χρονικό διάστημα αυτό έπρεπε να γίνει χρήση δύο εξειδικευμένων προγραμμάτων (Orange Data Mining και Rapid Miner Studio) καθώς και εκτέλεση όσο των δυνατών περισσότερων δοκιμών κατά την διαδικασία βελτιστοποίησης των παραμέτρων. Οι δυνατότητες επέμβασης στις παραμέτρους κατά την διαδικασία βελτιστοποίησης που επέτρεπε το λογισμικό ήταν περιορισμένες καθώς και ο διαθέσιμος χρόνος για εκτέλεση περαιτέρω δοκιμών.

Αν και δεν χρησιμοποιήθηκε κανένα μέγεθος που περιγράφει τον τύπο του πλοίου, που μας οδηγεί στο συμπέρασμα της χρησιμότητας της μεθόδου σε οποιονδήποτε τύπο, εντούτοις θα ήταν θεμιτή η επιβεβαίωση του παραπάνω ισχυρισμού με ένα σύνολο δεδομένων ενός τελείως διαφορετικού τύπου πλοίου.

5.3 Μελλοντική Έρευνα

Τα δεδομένα που χρησιμοποιούνται κατά την εκπαίδευση των αλγορίθμων αναφέρονται γενικά σε αρκετά μεγάλες χρονικές περιόδους. Προσπαθώντας να συλλέξουμε όσο τον δυνατόν περισσότερες τιμές για να έχουμε πιο ασφαλείς προβλέψεις εισαγάγουμε μέσα στην πληροφορία που κρύβουν τα δεδομένα και την φθορά του πλοίου. (π.χ. ρύπανση γάστρας). Η πληροφορία αυτή δεν μπορεί να συγκεκριμενοποιηθεί και κατανέμεται σε όλο το εύρος των τιμών. Παρατηρείται λοιπόν το οξύμωρο φαινόμενο από την μία να βελτιώνεται η πρόβλεψη λόγω πληθώρας πληροφορίας που περιγράφει το φαινόμενο και από την άλλη να αλλοιώνεται το αποτέλεσμα. Απαιτείται περαιτέρω έρευνα όσο αναφορά τον βέλτιστο αριθμό δεδομένων για εφαρμογές στην ναυτιλία.

Τα αποτελέσματα έδειξαν πως με κατάλληλη επιλογή και προ επεξεργασία των δεδομένων είμαστε σε θέση να εκτελέσουμε προβλέψεις λειτουργικών παραμέτρων με αρκετή ακρίβεια. Ένα άλλο χαρακτηριστικό ήταν πως αριθμητικά ήταν λίγες οι προβλέψεις που είχαν μεγάλη απόκλιση και επηρέασαν αρνητικά τις επιδόσεις. Οι λειτουργικές παράμετροι ενός πλοίου αντιστοιχούν σε μεγέθη των οποίων οι τιμές δεν αναμένεται να διαφοροποιηθούν δραματικά σε μικρά χρονικά διαστήματα. Είναι επομένως θεμιτό να βρεθεί μια διαδικασία μέσω της οποίας θα μπορούν να προλαμβάνονται τέτοιες στιγμιαίες αποκλίσεις στις τιμές πρόβλεψης.

6.ΒΙΒΛΙΟΓΡΑΦΙΑ

- {Daksh}, D. K. (2018). *mathematics-for-machine-learning-linear-regression-least-square-regression*. Retrieved from towardsdatascience: <https://towardsdatascience.com/mathematics-for-machine-learning-linear-regression-least-square-regression-de09cf53757c>
- Alpaydin, E. (2020). *Introduction to Machine Learning*. The MIT Press.
- Andrea Coraddu, L. O. (2016). Vessels fuel consumption forecast and trim optimisation: A data analytics perspective. *Ocean Engineering*.
- Binieli, M. (2018). Machine learning: an introduction to mean squared error and regression lines. *Freecodecamp*.
- Christos Gkerekos, I. L. (2019). Machine learning models for predicting shipmain engine Fuel Oil. *Ocean Engineering*.
- Corchado, F. J. (2018). *Hybrid Artificial Intelligent Systems*. Springer.
- Diesel, M. (2013). Basic Principles of Ships Propulsion.
- Fredrik Ahlgrena, M. E. (2019). Predicting Dynamic Fuel Oil Consumption on Ships with Automated Machine Learning. *Energy Procedia*.
- Kurama, V. (2020). A Guide to AdaBoost: Boosting To Save The Day. *Paperspace*.
- Maynard, M. (2020). Neural Networks_ Introduction to Artificial Neurons, Backpropagation and Multilayer Feedforward Neural Networks with Real-World Applications.
- McCorduck, P. (2019). *Machines Who Think: A Personal Inquiry into the History and Prospects of Artificial Intelligence 2nd Edition*. A K Peters/CRC Press.
- Michail Cheliotis, C. G. (2019, June 18). A novel data condition and performance hybrid imputation method for energy efficient operations of marine systems. *Journal of LATEX Templates*.
- Politis, G. K. (2018). *SHIP RESISTANCE AND PROMOTION*. Athens: E.M.P.
- Pyle, D. (1999). *Data Preparation for Data Mining (The Morgan Kaufmann Series in Data Management Systems)*. Morgan Kaufmann.
- Refaat, M. (2007). *Data Preparation for Data Mining Using SAS*. Elsevier.

Ron Kohavi, F. P. (1998). Special Issue on Applications of Machine Learning and the Knowledge Discovery Process. Netherland: Kluwer Academic Publishers.

Sahil_Bansall. (2019, August 06). Retrieved from geeksforgeeks:
<https://www.geeksforgeeks.org/agents-artificial-intelligence/>

Yeresime Suresh, L. K. (2014, March 4). Statistical and Machine Learning Methods for Software Fault Prediction Using CK Metric Suite: A Comparative Analysis. Hindawi Publishing Corporatio.

Yuji Roh, G. H. (2019). A Survey on Data Collection for Machine Learning. 20.

Πολίτης, Γ. Κ. (2018). *ΑΝΤΙΣΤΑΣΗ ΚΑΙ ΠΡΟΩΣΗ ΠΛΟΙΟΥ*. Αθηνά: ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ.

Τι είναι η τεχνητή νοημοσύνη και πώς χρησιμοποιείται. (2020, 9 9). Retrieved from Europarl:
<https://www.europarl.europa.eu/news/el/headlines/society/20200827STO85804>



**ΠΑΝΕΠΙΣΤΗΜΙΟ
ΔΥΤΙΚΗΣ ΑΤΤΙΚΗΣ**

Τμήμα Μηχανικών Βιομηχανικής
Σχεδίασης και Παραγωγής

&

**ΠΑΝΕΠΙΣΤΗΜΙΟ
ΑΙΓΑΙΟΥ**

Τμήμα Ναυτιλίας και
Επιχειρηματικών Υπηρεσιών

