



Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Ηλεκτρολόγων Μηχανικών
και Μηχανικών Υπολογιστών
Τομέας Τεχνολογίας, Πληροφορικής &
Υπολογιστών

**Τεχνικές Μηχανικής Μάθησης σε Προβλήματα
Δρομολόγησης και Ανάθεσης Πόρων**

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

ΑΣΗΜΙΝΑ ΜΕΡΤΖΑΝΗ

Επιβλέπων : Δημήτριος Φωτάκης
Αν. Καθηγητής Ε.Μ.Π.

Αθήνα, Σεπτέμβριος 2019



Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Ηλεκτρολόγων Μηχανικών
και Μηχανικών Υπολογιστών
Τομέας Τεχνολογίας, Πληροφορικής &
Υπολογιστών

Τεχνικές Μηχανικής Μάθησης σε Προβλήματα Δρομολόγησης και Ανάθεσης Πόρων

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

ΑΣΗΜΙΝΑ ΜΕΡΤΖΑΝΗ

Επιβλέπων : Δημήτριος Φωτάκης
Αν. Καθηγητής Ε.Μ.Π.

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 25η Σεπτεμβρίου 2019.

.....
Δημήτριος Φωτάκης
Αν. Καθηγητής Ε.Μ.Π.

.....
Νικόλαος Παπασπύρου
Καθηγητής Ε.Μ.Π.

.....
Γεώργιος Στάμου
Αν. Καθηγητής Ε.Μ.Π.

Αθήνα, Σεπτέμβριος 2019

.....
Ασημίνα Μερτζάνη

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

Copyright © Ασημίνα Μερτζάνη, 2019.

Με επιφύλαξη παντός δικαιώματος. All rights reserved.

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

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

Περίληψη

Η αποδοτική διαχείριση της εφοδιαστικής αλυσίδας θεωρείται σήμερα ιδιαίτερα σημαντική για τις επιχειρήσεις ενώ συγχρόνως αποτελεί ένα ευρύ πεδίο μελέτης για τους ερευνητές. Στο πλαίσιο αυτής της διπλωματικής εργασίας μελετάται η διαχείριση της εφοδιαστικής αλυσίδας, και ειδικότερα έμφαση δίνεται σε ένα επιμέρους πεδίο της, αυτό της δρομολόγησης οχημάτων. Με τον όρο αυτό αναφερόμαστε στο πρόβλημα της διανομής αγαθών σε καθορισμένα σημεία λαμβάνοντας υπόψιν συγκεκριμένους περιορισμούς. Αυτοί οι περιορισμοί μπορεί να είναι οι διαθέσιμοι πόροι για τη μεταφορά των αγαθών, ο χρόνος μέσα στον οποίο μπορούν να μεταφερθούν αυτά τα αγαθά και άλλα.

Μέχρι σήμερα, το πρόβλημα δρομολόγησης οχημάτων αποτελεί κυρίως πεδίο μελέτης της επιχειρησιακής έρευνας και της συνδυαστικής βελτιστοποίησης. Οι αλγόριθμοι που χρησιμοποιούνται για την επίλυση του επικεντρώνονται στην επαναληπτική εφαρμογή μιας αλληλουχίας κινήσεων έως ότου ικανοποιηθεί η συνθήκη τερματισμού. Εκτελούν, δηλαδή, διαδοχικές δοκιμές στο χώρο λύσεων με σκοπό τον εντοπισμό σημείων ελαχίστου κόστους.

Σε αντίθεση με τη συνήθη αντιμετώπιση, στην παρούσα εργασία, δοκιμάζεται η συνδυαστική εφαρμογή μεθόδων μηχανικής μάθησης και μαθηματικής βελτιστοποίησης για την επίλυση του παραπάνω προβλήματος. Καταγράφονται τα αποτελέσματα δύο διαφορετικών πρακτικών που δοκιμάστηκαν και στη συνέχεια συγκρίνονται με αυτά ενός αλγορίθμου δρομολόγησης οχημάτων.

Αναλυτικότερα, μετά την περιγραφή του γενικού πλαισίου και της ιστορίας, μεταβαίνουμε στην παρουσίαση της περίπτωσης προβλήματος υπό εξέταση. Ακολουθεί η περιγραφή των δεδομένων, η ανάλυση των βασικών χαρακτηριστικών τους και η επεξεργασία που απαιτήθηκε για να είναι δυνατή η κατανόηση τους και η αναγνώριση συσχετίσεων. Χάρη στα πρότυπα που εντοπίστηκαν στα δεδομένα, στάθηκε εφικτή η αξιοποίηση τους για παραγωγή προβλέψεων.

Έτσι, σχεδιάστηκε ένα μοντέλο μηχανικής μάθησης, με τη μέθοδο της λογιστικής παλινδρόμησης, που εκπαιδεύεται με βάση τα δεδομένα αυτά και έπειτα μπορεί να γενικεύσει αυτή τη γνώση για τον υπολογισμό της αναμενόμενης τιμής δεδομένων αγνώστων σε αυτό. Το μοντέλο αυτό παραμετροποιείται κατάλληλα ώστε να μπορεί να διαχωρίσει τα δεδομένα σε δύο κλάσεις. Στην συνέχεια της εργασίας αποτυπώνονται οι μεταβολές που έγιναν στα χαρακτηριστικά του μοντέλου για την βελτιστοποίηση της ικανότητας του για ταξινόμηση του συνόλου. Αφού ολοκληρωθεί η περιγραφή της τελικής μορφής του, παρουσιάζονται και αξιολογούνται τα αποτελέσματα της εφαρμογής τόσο στα δεδομένα εκπαίδευσης όσο και ένα άγνωστο σύνολο.

Για την ολική αξιολόγηση της παραπάνω μεθόδου, μετράται η συμβολή των εκτιμήσεων που παράγει στην κατασκευή της αρχικής λύσης ενός αλγορίθμου βελτιστοποίησης. Παράλληλα, δοκιμάζεται και η χρήση του στην επίλυση μέρους του προβλήματος. Τέλος, υπολογίζεται η αποτελεσματικότητα του αλγορίθμου βελτιστοποίησης στο ίδιο πρόβλημα και συγκρίνονται τα αποτελέσματα των παραπάνω δοκιμών με τα δεδομένα του συνόλου εκπαίδευσης.

Λέξεις κλειδιά

δρομολόγηση οχημάτων, περιορισμοί, χρονικά παράθυρα, διαθέσιμοι πόροι, βελτιστοποίηση, εφοδιαστική αλυσίδα, μηχανική μάθηση, λογιστική παλινδρόμηση, δυαδική ταξινόμηση, σύνολο δεδομένων, ανάλυση δεδομένων

Abstract

Nowadays, the efficient supply chain management is considered particularly important for companies while it is a broad field of study for researchers. This thesis examines the problem of logistics management and more specifically, it invests the vehicle routing problem. The term Vehicle Routing Problem (VRP) refers to the problem of distribution of goods at specific points, taking into account a range of constraints. These restrictions may include time, availability of resources etc.

Till now, Vehicle Routing Problem is mainly studied by operational research and combinatorial optimization. The existing algorithms apply iterative moves until they break a terminal condition. Testing different combinations, they are investigating new areas of the solution space continuously, trying to find points of minimal cost.

In contrast to the usual approach, the present work is focusing on the application of machine learning methods to solve the described problem. Then, follows the comparison of the results of the suggested method with the classic approach. Finally, the results of the combination of approaches are examined as well.

In more detail, after describing the general problem and its history, we move on the presentation of the case that we take into consideration in this thesis. Then, we describe the available set of data, we analyze its key features while in parallel we present the processing required for identifying the correlations between them. . Thanks to the patterns identified in the data, it was possible to use them to generate forecasts.

Thus, a machine learning model, using the logistic regression method, is designed based on the dataset. The goal of this model is to generalize the knowledge of the past in order to predict the outcome for a set of unknown data. Based on the correlations of data set, the scope of the Logistic Regression model is to separate optimally the data into two classes. This dissertation, then, outlines the changes made to the model's features to optimize its ability to classify the data set. After describing its final form, the results of application of the method to an unknown set are presented and evaluated.

For the overall evaluation of the model, the contribution of the predictions to the construction of the initial solution of an optimization algorithm is measured. Moreover, the efficiency of the model in solving a part of the problem is examined as well. Finally, we take into account the results of the optimization algorithm on the same problem and we compare the three different applications to the actual data (train/test set of the model).

Key words

vehicle routing, constraints, time windows, resources, optimization, supply chain, machine learning, logistic regression, binary classification, dataset, data analysis

Ευχαριστίες

Θα ήθελα να ευχαριστήσω ιδιαίτερα τον επιβλέποντα καθηγητή μου Δρ. Δημήτριο Φωτάκη ο οποίος με καθοδηγούσε καθ' όλη τη διάρκεια διεκπόνησης της διπλωματικής μου εργασίας με εποικοδομητικές συμβουλές και προτάσεις. Θα ήθελα να επισημάνω ότι δεν ήταν μόνο ένας σπουδαίος σύμβουλος με απαντήσεις σε κάθε μου απορία ή αδιέξοδο αλλά και ένας πραγματικός δάσκαλος που με έκανε να αγαπήσω την επιστήμη που μελετώ και να έχω θέληση να μαθαίνω.

Ακόμα, θα ήθελα να ευχαριστήσω θερμά τους ανθρώπους που στελεχώνουν την εταιρεία ORTEC για την παραχώρηση ανώνυμων δεδομένων που μελέτησα και χρησιμοποίησα καθώς και δοκιμαστικού περιβάλλοντος για την εκτέλεση των πειραμάτων. Ιδιαίτερα θα ήθελα να εκφράσω τις ευχαριστίες μου στους Δρ. Θεόδωρο Αθανασόπουλο, Δρ. Γεώργιο Νινίκα και MSc. Παναγιώτη Κυρίτση που με υπομονή και κατανόηση με βοηθούσαν, παρέχοντας μου επεξηγήσεις που αφορούν τόσο στα προβλήματα εφοδιαστικής αλυσίδας όσο και στη χρήση του προγράμματος της εταιρείας.

Τέλος, θα ήθελα να ευχαριστήσω ιδιαίτερα την οικογένεια μου, που έδειξε και δείχνει αμέριστη στήριξη και αγάπη και επικροτεί τις επιλογές μου.

Ασημίνα Μερτζάνη,

Αθήνα, 25η Σεπτεμβρίου 2019

Περιεχόμενα

Περίληψη	5
Abstract	7
Ευχαριστίες	9
Περιεχόμενα	11
Κατάλογος πινάκων	13
Κατάλογος σχημάτων	15
1. Εισαγωγή	17
Εισαγωγή	17
1.1 Το Πρόβλημα της Δρομολόγησης Οχημάτων	17
1.2 Κίνητρο	18
1.3 Συνεισφορά Διπλωματικής Εργασίας	18
1.4 Γενική Προσέγγιση Προβλήματος	19
1.5 Δομή Διπλωματικής Εργασίας	20
2. Θεωρητικό Υπόβαθρο	21
Θεωρητικό Υπόβαθρο	21
2.1 Πρόβλημα Δρομολόγησης Οχημάτων, Παραλλαγές και Τρόποι Επίλυσης	21
2.1.1 Περιγραφή Προβλήματος Δρομολόγησης Οχημάτων και Ιστορική Αναδρομή	21
2.1.2 Παραλλαγές Προβλήματος	22
2.1.3 Τρόποι Επίλυσης	22
2.1.4 Επίλυση ΠΔΟ με Χρήση Τεχνικών Μηχανικής Μάθησης	23
2.1.5 ORTEC	24
2.2 Η Επιστήμη των Δεδομένων και οι Βασικές Μεθοδολογίες της	24
2.2.1 Η Επιστήμη των Δεδομένων	24
2.2.2 Μεθοδολογία	25
2.2.3 Ανάλυση Χρονοσειρών	26
2.3 Μηχανική Μάθηση	26
2.3.1 Ορισμός Μηχανικής Μάθησης	26
2.3.2 Βασικά είδη μάθησης και Χαρακτηριστικοί Αλγόριθμοι	27
2.3.3 Μοντέλο Λογιστικής Παλινδρόμησης	28
2.3.4 Μετρικές Αξιολόγησης Μεθόδων Μηχανικής Μάθησης	29
2.3.5 Τεχνικές Βελτίωσης Μοντέλων Μηχανικής Μάθησης	30
2.3.6 Προκλήσεις και Περιορισμοί	31

3. Σχεδιασμός και Υλοποίηση	33
Σχεδιασμός και Υλοποίηση	33
3.1 Επεξεργασία Δεδομένων	33
3.1.1 Δεδομένα	33
3.1.2 Παρατηρήσεις επί των Δεδομένων	33
3.2 Σχεδίαση Μοντέλου Μηχανικής Μάθησης	36
3.2.1 Κατασκευή Συνόλου Δεδομένων	36
3.2.2 Το Πρώτο Μοντέλο	36
3.2.3 Εξέλιξη Μοντέλου	37
3.3 Παραμετροποίηση Περιβάλλοντος Δοκιμών	46
3.3.1 Δεδομένα Εισόδου	46
3.3.2 Αλγόριθμος Βελτιστοποίησης	46
3.3.3 Διαφορετικές Δοκιμές στο Δοκιμαστικό Περιβάλλον	49
4. Αξιολόγηση αποτελεσμάτων και συγκρίσεις	51
Αξιολόγηση Αποτελεσμάτων Συγκρίσεις	51
4.1 Αποτελέσματα μοντέλου μηχανικής μάθησης	51
4.1.1 Αξιολόγηση Αποτελεσμάτων Μετρικών	51
4.1.2 Αξιολόγηση Αποτελεσμάτων με Προσωπική Παρατήρηση	54
4.2 Αποτελέσματα στο Δοκιμαστικό Περιβάλλον	58
4.2.1 Παρουσίαση Διαφορετικών Δοκιμών	58
4.2.2 Σύγκριση Αποτελεσμάτων	59
4.3 Συνολικά Συμπεράσματα	62
5. Επίλογος	65
Επίλογος	65
5.1 Σύνοψη	65
5.2 Προοπτικές	66
5.3 Μελλοντικές Κατευθύνσεις Επιστημονικής Μελέτης	67

Κατάλογος πινάκων

3.1	Αποτελέσματα Πρώτου Μοντέλου	37
3.2	Αποτελέσματα Διορθωμένου Μοντέλου	38
3.3	Αποτελέσματα Μοντέλου Λιγότερων Χαρακτηριστικών	40
4.1	Συνολικός Πίνακας Αποτελεσμάτων	51
4.2	Συνολικός Πίνακας Αποτελεσμάτων Δοκικών στο Περιβάλλον	63

Κατάλογος σχημάτων

2.1	Συσχέτιση Γνωστών Παραλλαγών των Υποπροβλημάτων του ΠΔΟ	22
2.2	Θεμελιώδης Μεθοδολογία Επιστήμης των Δεδομένων	25
3.1	Μέσο Πλήθος Δρομολογίων ανά Ημέρα της Εβδομάδας	34
3.2	Πλήθος Δρομολογίων ανά Αριθμό Στάσεων	34
3.3	Πλήθος Δρομολογίων ανά Αριθμό Στάσεων Παραλαβής	35
3.4	Πλήθος Δρομολογίων ανά Αριθμό Στάσεων Παράδοσης	35
3.5	Πίνακας Συντελεστών Αυτοσυσχέτισης	40
3.6	Βέλτιστα Βάρη Χαρακτηριστικών	45
3.7	Συνάρτηση Κόστους - Χώρου Λύσεων	48
3.8	Μέθοδοι της Στρατηγικής Ruin and Recreate	49
4.1	Πίνακας Συντελεστών Αυτοσυσχέτισης	52
4.2	Ποσοστό Ορθών και Λανθασμένων Εκτιμήσεων	54
4.3	Ποσοστό ανά Κατηγορία Παραγγελίας	55
4.4	Διάγραμμα Ορθών Εκτιμήσεων	55
4.5	Διάγραμμα Εσφαλμένων Εκτιμήσεων	55
4.6	Διάγραμμα Παραγγελιών της Κλάσης 1 (P2P)	55
4.7	Διάγραμμα Παραγγελιών της Κλάσης 1	55
4.8	Ποσοστά Σφαλμάτων ανά Μέσο Όγκο Παραγγελίας	56
4.9	Ποσοστά Σφαλμάτων ανά Απόσταση Σημείων Παράδοσης-Παραλαβής	56
4.10	Ποσοστά Σφαλμάτων ανά Σημείο Παραλαβής	57
4.11	Ποσοστά Σφαλμάτων ανά Μέρα της Εβδομάδας	57
4.12	Πλήθος Μη Δρομολογημένων Παραγγελιών	60
4.13	Πλήθος Δρομολογημένων Παραγγελιών	60
4.14	Πλήθος Δρομολογίων 2 Στάσεων	61
4.15	Σύνολο Δρομολογίων	61
4.16	Διανυόμενη Απόσταση	62
4.17	Χρόνος Εκτέλεσης	62

Κεφάλαιο 1

Εισαγωγή

1.1 Το Πρόβλημα της Δρομολόγησης Οχημάτων

Το πρόβλημα της δρομολόγησης οχημάτων είναι ένα πρόβλημα συνδυαστικής βελτιστοποίησης και ακέραιου προγραμματισμού το οποίο εκφράζει το ακόλουθο ερώτημα "Ποιος είναι ο βέλτιστος συνδυασμός δρομολογίων για ένα στόλο οχημάτων για την παράδοση σε ένα δεδομένο σετ από πελάτες;". Αποτελεί μια γενίκευση του γνωστού προβλήματος του πλανόδιου πωλητή. Πρώτα μελετήθηκε σε εργασία των [G.B.Dantzig and J.H.Ramser \(1959\)](#), στην οποία γράφτηκε μια πρώτη αλγοριθμική προσέγγιση και εφαρμόστηκε σε παραδόσεις πετρελαίου. Συχνά, το πλαίσιο είναι η παράδοση αγαθών που βρίσκονται σε κεντρική αποθήκη σε πελάτες που έχουν παραγγείλει τέτοια αγαθά. Ο στόχος του προβλήματος δρομολόγησης οχημάτων είναι να ελαχιστοποιήσει το συνολικό κόστος της διαδρομής. Μερικά χρόνια αργότερα, οι [G.Clarke and J.W.Wright \(1964\)](#) βελτίωσαν την προσέγγιση των [G.B.Dantzig and J.H.Ramser \(1959\)](#) χρησιμοποιώντας μια αποτελεσματική άπληστη προσέγγιση που ονομάζεται αλγόριθμος εξοικονόμησης. (From Wikipedia, The free encyclopedia)

Αυτή η διπλωματική εργασία μελετά μια μορφή του παραπάνω προβλήματος και προτείνει διαφορετικές μεθόδους για την επίλυση του. Αποσκοπεί εκμετάλλευση ιστορικών δεδομένων μέσα από μεθοδολογίες που υπαγορεύει το επιστημονικό πεδίο της μηχανικής μάθησης, για την εκτίμηση του κατάλληλου τρόπου δρομολόγησης οχημάτων για την μεταφορά των αγαθών από αποθήκες σε άλλες αποθήκες και καταστήματα. Ο αλγόριθμος μηχανικής μάθησης εκπαιδεύεται με βάση το διαθέσιμο σύνολο δεδομένων στη συνέχεια αξιολογείται τόσο στα δεδομένα αυτά (train set) όσο και σε άγνωστα σε αυτόν δεδομένα (test set). Η έξοδος του αλγορίθμου χρησιμοποιείται είτε ως αρχική λύση για αλγόριθμο βελτιστοποίησης είτε για δρομολόγηση μέρους των παραγγελιών και συνδυαστική επίλυση με αλγόριθμο βελτιστοποίησης για επίλυση υπολοιπόμενων. Τέλος, τα αποτελέσματα των δύο παραπάνω δοκιμών συγκρίνονται με τα αποτελέσματα του αλγορίθμου βελτιστοποίησης χωρίς τη συμβολή του μοντέλου αλλά και με τα δεδομένα του συνόλου που χρησιμοποιήθηκε για την εκπαίδευση του μοντέλου.

Αναλυτικότερα, το μοντέλο μηχανικής μάθησης εφαρμόζει αλγόριθμο λογιστικής παλινδρόμησης για την κατηγοριοποίηση των παραγγελιών σε δύο ομάδες. Ως είσοδος του δίνεται ένας πίνακας όπου κάθε γραμμή αντιπροσωπεύει μία παραγγελία και κάθε στήλη περιγράφει κάποιο χαρακτηριστικό της όπως για παράδειγμα τον όγκο της, το σημείο παραλαβής της, η ημέρα της εβδομάδας που πρέπει αυτή να παραδοθεί και άλλα.

Ο αλγόριθμος που αναπτύχθηκε εκτιμάει αν αυτή η παραγγελία θα τοποθετηθεί μόνη της σε ένα δρομολόγιο, ή θα συνδυαστεί με άλλες. Προσπαθεί δηλαδή να λύσει ένα πρόβλημα ταξινόμησης των παραγγελιών σε δύο κλάσεις, την κλάση των παραγγελιών που τοποθετούνται μόνες τους και κατ' επέκταση συνθέτουν ένα δρομολόγιο δύο στάσεων, και την κλάση των παραγγελιών που πρέπει να συνδυαστούν με άλλες για να συνθέσουν ένα δρομολόγιο.

Εν συνεχεία, χρησιμοποιείται ο αλγόριθμος μαθηματικής βελτιστοποίησης για επίλυση του συνολικού προβλήματος. Επιχειρούνται οι δύο διαφορετικές μέθοδοι για να εντοπισθεί ποια ανταποκρίνεται καλύτερα και επιλύει με βέλτιστο τρόπο το ΠΔΟ που μελετάμε.

1.2 Κίνητρο

Η αποδοτική δρομολόγηση οχημάτων καθίσταται σήμερα ένα σημαντικό πεδίο ενδιαφέροντος για τις επιχειρήσεις. Τόσο οι απαιτήσεις των πελατών για υψηλή ποιότητα υπηρεσιών όσο και οι ανάγκες για περιορισμό των περιττών διαδρομών ασκούν ιδιαίτερες πιέσεις σε αυτές. Από τη μία, η αύξηση του ανταγωνισμού μεταξύ των διαφορετικών επιχειρήσεων και η πρόοδος της τεχνολογίας καθιστούν του πελάτες πιο αυστηρούς κριτές των υπηρεσιών που λαμβάνουν. Από την άλλη, η μεγαλύτερη ευαισθησία για το περιβάλλον και η επιθυμία για ελαχιστοποίηση των εκπομπών διοξειδίου του άνθρακα αλλά και η εκτεταμένη χρήση οχημάτων με συνέπεια τη μεγάλη κυκλοφοριακή συμφόρηση επιβάλλουν την περικοπή των άσκοπων μετακινήσεων. Έτσι, η βελτιστοποίηση σε προβλήματα δρομολόγησης οχημάτων δεν είναι μόνο ένα ενδιαφέρον πεδίο μελέτης για την επιστήμη των υπολογιστών, αλλά και μία απαίτηση της κοινωνίας του σήμερα.

Μελετώντας τις μελέτες που έχουν διεξαχθεί για την επίλυση του παραπάνω προβλήματος παρατηρούμε ένα ευρύ φάσμα λύσεων που κάνουν χρήση μεθοδολογιών επιχειρησιακής έρευνας. Μεγάλη πρόοδος έχει σημειωθεί στην μελέτη ευρετικών μεθόδων και μετα-ευρετικών μεθόδων με πολύ αξιόλογα αποτελέσματα. Παράλληλα, όμως, ιδιαίτερη ανάπτυξη σημειώνεται σήμερα στην κατεύθυνση της εφαρμογής τεχνικών μηχανικής μάθησης σε διαφορετικά πεδία μελέτης. Παρατηρούνται, έτσι, ενδιαφέρουσες δοκιμές τέτοιων μεθοδολογιών σε πληθώρα επιστημονικών πεδίων, όπου και παρουσιάζουν εκπληκτικά αποτελέσματα.

Παρά την αυξημένη μελέτη γύρω από τις δύο παραπάνω επιστημονικές περιοχές, μικρή έρευνα φαίνεται να υπάρχει στην εφαρμογή τεχνικών μηχανικής μάθησης σε προβλήματα δρομολόγησης οχημάτων. Η συγκεκριμένη διπλωματική εργασία επικεντρώνεται, λοιπόν, στην δοκιμή μεθόδων μηχανικής μάθησης για επίλυση προβλημάτων δρομολόγησης και στην μελέτη των αποτελεσμάτων ενός τέτοιου συνδυασμού.

1.3 Συνεισφορά Διπλωματικής Εργασίας

Η διατριβή αυτή μελετά τη βελτιστοποίηση ενός προβλήματος που συναντάται στην καθημερινή μας ζωή, δηλαδή στην αξιοποίηση της τεχνολογίας για ελαχιστοποίηση του κόστους της δρομολόγησης οχημάτων. Επιχειρείται η αξιοποίηση των ανερχόμενων και πολυσυζημένων, σήμερα, μεθόδων μηχανικής μάθησης για επίλυση προβλημάτων μαθηματικής βελτιστοποίησης. Η διατριβή αυτή ξεπερνάει το πλαίσιο της θεωρητικής μελέτης τέτοιου είδους προβλημάτων και αλλά εφαρμόζεται σε δεδομένα από το παραγωγικό περιβάλλον μιας επιχείρησης της ελληνικής αγοράς. Έτσι, συνεισφέρει στην μελέτη του συνδυασμού μεθοδολογιών της επιστήμης των δεδομένων, της

μηχανικής μάθησης και της μαθηματικής βελτιστοποίησης ενώ παράλληλα επιλύει ένα πρακτικό πρόβλημα.

1.4 Γενική Προσέγγιση Προβλήματος

Όπως ορίστηκε παραπάνω το ΠΔΟ έχει απασχολήσει σε μεγάλο βαθμό τόσο την επιστημονική έρευνα όσο και τις επιχειρήσεις. Μια μορφή του παραπάνω προβλήματος καλείται να λύσει καθημερινά αλυσίδα καταστημάτων που δραστηριοποιείται στην ελληνική αγορά. Κάθε μέρα σχηματίζουν κατά μέσο όρο 250 δρομολόγια για τη μεταφορά αγαθών από τις αποθήκες στα καταστήματα ή σε άλλες αποθήκες. Λαμβάνοντας υπόψιν περιορισμούς όπως η ευαισθησία των προϊόντων, η χωρητικότητα των οχημάτων, τα χρονικά παράθυρα παραλαβής και παράδοσης και άλλα προσπαθούν να σχηματίσουν δρομολόγια που ελαχιστοποιούν την διανυόμενη απόσταση, τον συνολικό χρόνο και τους απασχολούμενους πόρους ενώ παραδίδουν αν είναι δυνατόν όλες τις παραγγελίες στην ώρα τους. Για την επίλυση αυτού το προβλήματος καθημερινά, η εταιρεία αυτή έχει επιλέξει να χρησιμοποιεί το λογισμικό της εταιρείας ORTEC.

Η ORTEC παρέχει στους πελάτες της ολοκληρωμένα συστήματα δρομολόγησης και σχεδιασμού κάνοντας χρήση μαθηματικής βελτιστοποίησης και αξιοποιώντας τις σύγχρονες τεχνολογίες. Παρά την σχεδίαση πολύ αποδοτικών διαδρομών, παρατηρήθηκαν πολλές ανθρώπινες παρεμβάσεις στο σύστημα, αλλοιώνοντας σε μεγάλο βαθμό τις προτεινόμενες λύσεις και αυξάνοντας το κόστος για τη μεταφορά των αγαθών.

Με στόχο την βελτίωση του αλγορίθμου για τον περιορισμό των ανθρώπινων παρεμβάσεων, χρειάστηκε η συλλογή, η ανάλυση και η επεξεργασία του συνόλου των δεδομένων δρομολόγησης ενός έτους χρήσης του λογισμικού αυτού. Μετά από πολύπλευρη ανάλυση του παρατηρήθηκε ότι σχεδόν τα μισά δρομολόγια πραγματοποιούσαν μόλις δύο στάσεις (P2P δρομολόγια), ενώ σχεδόν το 95% έκανε το πολύ 4 στάσεις. Έτσι, φάνηκε κομβικός ο διαχωρισμός των παραγγελιών σε δύο ομάδες, αυτές που πρόκειται να σχηματίσουν ένα ανεξάρτητο δρομολόγιο και αυτές που θα χρειαστεί να συνδυαστούν με άλλες.

Για τον βέλτιστο διαχωρισμό των δεδομένων σε δύο επιμέρους σύνολα, επιλέχθηκε η χρήση μεθόδων μηχανικής μάθησης. Δεδομένης της αποτελεσματικότητας τους στη διαχείριση μεγάλου όγκου διαθέσιμων πληροφοριών για την παραγωγή αξιόπιστων εκτιμήσεων, θεωρήθηκε ότι αποτελούν την καλύτερη προσέγγιση του συγκεκριμένου προβλήματος. Έτσι, σχηματίστηκε ένα μοντέλο λογιστικής παλινδρόμησης, που με βάση τις πληροφορίες που διαθέτει, εντοπίζει το βέλτιστο επίπεδο διαχωρισμού του χώρου σε δύο ημι-χώρους για την ελαχιστοποίηση του σφάλματος των εκτιμήσεων. Έτσι, μετά την εκπαίδευση του δύναται να κατατάξει την κάθε παραγγελία σε μία από τις δύο ομάδες με πολύ μεγάλη ακρίβεια.

Μετά τον εντοπισμό των P2P παραγγελιών, σειρά έχει η διαχείριση των υπόλοιπων παραδόσεων που πρέπει να εκτελεστούν για την ίδια μέρα. Για την επίτευξη αυτού του στόχου, δοκιμάστηκαν δύο διαφορετικές μεθοδολογίες. Στην πρώτη, επιλέγονται οι παραγγελίες που έχουν επισημανθεί ως P2P και σχεδιάζουν τα αρχικά δρομολόγια. Αυτά δίνονται στον αλγόριθμο βελτιστοποίησης της εταιρείας ORTEC ως προτεινόμενη αρχική λύση. Έπειτα δίνεται η ελευθερία στον αλγόριθμο να τροποποιήσει όπως επιθυμεί αυτά τα δρομολόγια, καθώς και να σχηματίσει νέα με τις διαθέσιμες μονάδες μεταφοράς για να οργανώσει την παράδοση των παραγγελιών. Στην δεύτερη δοκιμή, ομοίως σχηματίζονται τα P2P δρομολόγια με βάση τις συστάσεις του μοντέλου. Οι

παραγγελίες που ανήκουν στην άλλη κατηγορία καθώς και ο υπόλοιπος στόλος οχημάτων δίνεται στον αλγόριθμο βελτιστοποίησης δρομολόγησης. Τα αποτελέσματα τόσο του μαθηματικού αλγορίθμου όσο και του μοντέλου συνδυάζονται για την επίλυση του συνολικού προβλήματος.

1.5 Δομή Διπλωματικής Εργασίας

Στο Κεφάλαιο 2, πρόκειται να περιγραφεί το τεχνικό υπόβαθρο, με τις αντίστοιχες παραπομπές στα πρωτότυπα κείμενα, που απαιτείται για την πλήρη κατανόηση της συγκεκριμένης διατριβής. Στο κεφάλαιο αυτό, λοιπόν, δίνεται μια εκτενέστερη αναφορά στο πρόβλημα της δρομολόγησης οχημάτων και ανάθεσης πόρων, στην ιστορία του και στις διάφορες παραλλαγές του. Ακόμα, καταγράφονται οι βασικές μέθοδοι με τις οποίες προσεγγίζονται μέχρι σήμερα τέτοιου είδους προβλήματα. Εν συνεχεία, παρουσιάζονται οι βασικές αρχές και μεθοδολογίες της επιστήμης των δεδομένων. Έπειτα, ακολουθεί η εισαγωγή στην επιστήμη της μηχανικής μάθησης και η καταγραφή των βασικών ειδών και αλγορίθμων. Ακολουθεί η παρουσίαση της τεχνικής μηχανικής μάθησης που επιλέχθηκε και η μαθηματική περιγραφή της. Τέλος, δίνεται μια συνοπτική παρουσίαση των συνηθισμένων σφαλμάτων στα οποία πρέπει να δοθεί προσοχή καθώς και σε μεθόδους αξιολόγησης τους.

Στο κεφάλαιο 3, αποτυπώνεται η περιγραφή της μεθοδολογίας που επιλέχθηκε στα πλαίσια αυτής της εργασίας. Παρουσιάζονται οι βασικοί άξονες μελέτης που αναπτύχθηκαν παράλληλα για την παροχή μιας εμπειριστατωμένης έρευνας. Δίνεται αρχικά, η περιγραφή του τρόπου απόκτησης των δεδομένων που χρησιμοποιήθηκαν για την εκπαίδευση και αξιολόγηση της επιλεγμένης τεχνικής. Στη συνέχεια, μεταβαίνουμε στα στάδια εξέλιξης του μοντέλου, από την πρώτη πρόχειρη δοκιμαστική μορφή του μέχρι την τελική. Εκεί δίνονται αναλυτικά τα χαρακτηριστικά και οι παράμετροι που χρησιμοποιήθηκαν σε κάποιες χαρακτηριστικές φάσεις του μοντέλου καθώς και οι λόγοι που οδήγησαν στις διάφορες τροποποιήσεις. Ακολουθεί η καταγραφή των δεδομένων που χρησιμοποιήθηκαν για την παραμετροποίηση ενός δοκιμαστικού περιβάλλοντος. Αυτό χρησιμοποιήθηκε για την διεξαγωγή πειραμάτων και την καταγραφή των αποτελεσμάτων των διαφορετικών προσεγγίσεων που δοκιμάστηκαν.

Στο κεφάλαιο 4, δίνεται η παρουσίαση των τελικών αποτελεσμάτων και η αξιολόγηση τους. Από τη μία γίνεται η αξιολόγηση του μοντέλου λογιστικής παλινδρόμησης με βάση κάποιες κρίσιμες μετρικές, καθώς και με δοκιμές σε μεμονωμένα ενδιαφέροντα στιγμιότυπα του προβλήματος. Μετράται, δηλαδή, η αποτελεσματικότητα του στην αποκωδικοποίηση των συσχετίσεων των αρχικών δεδομένων, για τον ορθό διαχωρισμό τους. Από την άλλη, καταγράφεται η συμβολή του στην επίλυση του γενικότερου προβλήματος. Εδώ παρουσιάζονται τα αποτελέσματα των διαφορετικών δοκιμών στο δοκιμαστικό περιβάλλον. Τέλος, αναλύονται ποιοτικά οι παραπάνω μετρήσεις και συγκρίνονται τα αποτελέσματα των διαφορετικών δοκιμών.

Στο κεφάλαιο 5, καταλήγουμε στα συμπεράσματα καθώς και σε προτάσεις για μελλοντική εξέλιξη. Παρατηρούμε σε τι υπερτερεί και τι υστερεί η κάθε μέθοδος και παρέχονται προτάσεις για εξέλιξη της παρούσας μελέτης και ιδέες για παρεμφερείς εφαρμογές που μπορεί να παρουσιάζουν ενδιαφέρον.

Κεφάλαιο 2

Θεωρητικό Υπόβαθρο

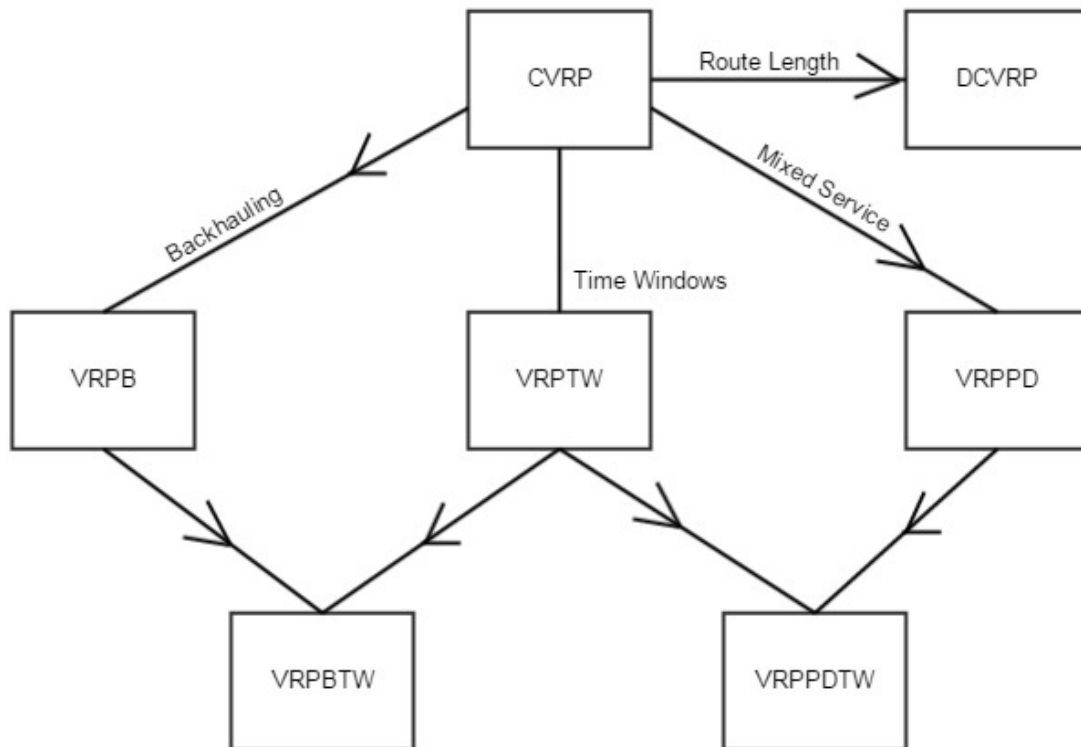
2.1 Πρόβλημα Δρομολόγησης Οχημάτων, Παραλλαγές και Τρόποι Επίλυσης

2.1.1 Περιγραφή Προβλήματος Δρομολόγησης Οχημάτων και Ιστορική Αναδρομή

Το Πρόβλημα Δρομολόγησης Οχημάτων (ΠΔΟ) μπορεί να περιγραφεί ως το πρόβλημα του σχεδιασμού των δρομολογίων ελάχιστου κόστους προς πελάτες που βρίσκονται σε διαφορετικές γεωγραφικές περιοχές, λαμβάνοντας υπόψιν ένα σύνολο περιορισμών. Η πληθώρα περιορισμών που μπορεί να υπάρξουν στην πράξη οδηγεί στην ύπαρξη πολλών παραλλαγών του ΠΔΟ. Το ΠΔΟ με χωρητικότητες ορίσθηκε αρχικά από τους [G.B.Dantzig and J.H.Ramser \(1959\)](#) όπου επιλέχθηκε μια προσέγγιση με χρήση απλής ευρετικής μεθόδου (heuristics). Τα επόμενα χρόνια δημιουργήθηκαν πιο εξελιγμένες ευρετικές μέθοδοι που ακολουθούσαν διαφορετικές αρχές όπως γεωγραφική εγγύτητα, εξοικονόμηση, αντιστοιχίες πελατών κλπ. Στο έργο των [G.Clarke and J.W.Wright \(1964\)](#) δίνεται μία προσέγγιση βασισμένη στην αρχή της εξοικονόμησης, η οποία καθίσταται μία από τις πιο γνωστές μέχρι και σήμερα. Στη συνέχεια, ακολούθησαν οι μελέτες των [N.Christofides et al. \(1981b\)](#) και [N.Christofides et al. \(1981a\)](#) που καθιέρωσαν μια νέα μέθοδο, αυτή των ακριβών αλγορίθμων (exact algorithms). Στην πρώτη δίνεται ένας αλγόριθμος δυναμικού προγραμματισμού ενώ στη δεύτερη χρησιμοποιούνται δύο μαθηματικοί φορμαλισμοί βασισμένοι σε αλγορίθμους συντομότερων μονοπατιών και ελάχιστων συνεκτικών δέντρων. Ακολούθησε το 1984 η εργασία των [G.Laporte et al. \(1984\)](#) που καθιέρωσε τις μεθόδους τομής, οι οποίες αποτέλεσαν τη βάση για τους αλγορίθμους που αναπτύσσονται μέχρι και σήμερα. Τα επόμενα χρόνια ακολούθησαν πολλοί ακριβείς αλγόριθμοι βασισμένοι σε διατυπώσεις μαθηματικού προγραμματισμού, όπως οι [R.Fukasawa et al. \(1984\)](#) και [R.Baldacci et al. \(2008\)](#). Η μελέτη γύρω από το ΠΔΟ έθεσε τα θεμέλια για τις μεθόδους που είναι γνωστές ως μετα-ευρετικές (meta-heuristics). Οι πρώτες έρευνες κινήθηκαν γύρω από τη μέθοδο της tabu-search όπως οι [E.Taillard \(1993\)](#), [M.Gendreau et al. \(1994\)](#) και [P.Toth \(2003\)](#). Παρατηρήθηκε ότι οι πιο αποδοτικές μετα-ευρετικές μέθοδοι είναι αυτές που κάνουν ευρεία και βαθιά αναζήτηση στο χώρο λύσης και έτσι μπορούν να αντεπεξέλθουν σε διάφορες παραλλαγές του προβλήματος. Τέτοιοι μέθοδοι είτε χρησιμοποιούν πολλούς τελεστές, όπως στην μελέτη προσαρμωστικής αναζήτησης μεγάλης γειτονιάς του [D.Pisinger \(2007\)](#), είτε συνδυάζουν γενετική και τοπική αναζήτηση, όπως παρουσιάζεται στο υβριδικό γενετικό αλγόριθμο (hybrid genetic algorithm) του [T.Vidal \(2012\)](#).

2.1.2 Παραλλαγές Προβλήματος

Στην εικόνα που ακολουθεί μπορούμε να παρατηρήσουμε τις διάφορες παραλλαγές του ΠΔΟ.



Σχήμα 2.1: Συσχέτιση Γνωστών Παραλλαγών των Υποπροβλημάτων του ΠΔΟ

Όπως παρουσιάζονται παραπάνω, οι πιο γνωστές παραλλαγές του ΠΔΟ είναι οι εξής:

- ΠΔΟ με παραλαβές και παραδόσεις [VRPPD]: Στόχος είναι η εύρεση βέλτιστων δρομολογίων για μεταφορά αγαθών από κάποια σημεία παραλαβής σε κάποια σημεία παράδοσης, δεδομένου ενός στόλου οχημάτων. Στην μελέτη του [G.Righini \(2000\)](#) μπορούμε να δούμε μια λύση του VRPPD με προσεγγιστικούς αλγορίθμους.
- ΠΔΟ με χρονιά παράθυρα [VRPTW]: Εδώ, τα σημεία παράδοσης έχουν χρονικά παράθυρα εντός των οποίων πρέπει να γίνουν οι παραδόσεις. Μια καλή λύση του προβλήματος δόθηκε στο [G.Righini \(1995\)](#).
- ΠΔΟ με χωρητικότητες [CVRP]: Τα οχήματα μπορούν να μεταφέρουν συγκεκριμένους όγκους αγαθών. Μια καλή προσέγγιση αυτής της παραλλαγής δίνεται στην εργασία των [T.K.Ralphs and L.E.Trotter Jr. \(2003\)](#).
- ΠΔΟ με πολλά δρομολόγια [VRPMT]: Τα οχήματα μπορούν να κάνουν περισσότερες από μία διαδρομές.
- Ανοιχτό ΠΔΟ [OVRP]: Τα οχήματα δεν χρειάζεται να επιστρέψουν στην αποθήκη.

2.1.3 Τρόποι Επίλυσης

Στην ενότητα της ιστορικής αναδρομής παρατηρήσαμε ότι πληθώρα επιστημόνων επικεντρώθηκαν στην μελέτη του ΠΔΟ. Οι διάφορες επιστημονικές ομάδες προκάλεσαν την ανάπτυξη ενός

ευρύου φάσματος μεθοδολογιών. Στο άρθρο [Solution Methods for VRP](#) δίνεται η παρακάτω κατηγοριοποίηση των μεθόδων επίλυσης του ΠΔΟ:

1. Ακριβείς προσεγγίσεις (exact approaches)
2. Ευρετικές μέθοδοι (Heuristics)
3. Κατασκευαστικές μέθοδοι (Constructive Methods)
4. Αλγόριθμοι 2-φάσεων (2-Phase Algorithm)
5. Μετα-ευρετικές μέθοδοι (2-Phase Algorithm)

Λίγο πιο αναλυτικά, στην πρώτη κατηγορία ανήκουν μέθοδοι που δοκιμάζουν όλες τις διαφορετικές λύσεις έως ότου βρεθεί η καλύτερη. Ένα τέτοιο παράδειγμα αποτελεί ο αλγόριθμος Branch and Bound του [Fisher \(1994\)](#). Στην ίδια κατεύθυνση εξελίσσονται οι ευρετικές λύσεις, όπου διακρίνονται από τις παραπάνω λόγω του ότι δοκιμάζουν περιορισμένο εύρος λύσεων προς ορισμένες κατευθύνσεις του χώρου αναζήτησης και φαίνονται συνήθως να είναι πολύ αποτελεσματικές σε μικρό υπολογιστικό χρόνο. Εν συνεχεία, μπορούμε να εντοπίσουμε αποτελεσματικούς αλγορίθμους βασισμένους σε μια άλλη φιλοσοφία. Τέτοιοι είναι οι αλγόριθμοι των [Desrochers and Verhoog \(1989\)](#) και [Thompson and Psaraftis \(1993\)](#) που αναπτύχθηκαν με τη λογική της κατασκευής δυνατών λύσεων λαμβάνοντας υπόψιν το κόστος. Περνώντας στην τέταρτη κατηγορία, βλέπουμε αλγορίθμους που αναπτύσσονται σε δύο διακριτά στάδια τα οποία μπορούν να αλληλεπιδράσουν, το στάδιο του διαχωρισμού των σημείων σε εφικτά δρομολόγια και το στάδιο της κατασκευής των κανονικής λύσης. Συχνό είναι είτε να προηγείται ο διαχωρισμός και να έπεται η δημιουργία δρομολογίων είτε και το αντίστροφο. Τέλος, στην πέμπτη κατηγορία ανήκουν οι γνωστοί αλγόριθμοι Simulated Annealing (όπως η εργασία των [O. Arbelaitz and Zamakola \(2001\)](#)), Tabu Search(για παράδειγμα [P. Badeau and Taillard](#)) και οι γενετικοί αλγόριθμοι που ενέπνευσαν πληθώρα μελετητών και σε διαφορετικά επιστημονικά πεδία. Για παράδειγμα, στο χώρο της τεχνητής νοημοσύνης αναπτύσσονται αλγόριθμοι βασισμένοι στη λογική των γενετικών αλγορίθμων και Simulated Annealing.

2.1.4 Επίλυση ΠΔΟ με Χρήση Τεχνικών Μηχανικής Μάθησης

Κατόπιν αναζήτησης στη βιβλιογραφία, καταγράφηκαν ορισμένες προσεγγίσεις του ΠΔΟ από τη σκοπιά της μηχανικής μάθησης. Χαρακτηριστικό παράδειγμα μια τέτοιας αντιμετώπισης αποτελεί η εργασία των [Mohammadreza Nazari \(2003\)](#) όπου κάνει χρήση μεθόδων ενισχυτικής μάθησης για τον εντοπισμό λύσεων. Σε αυτή τη μελέτη, το μοντέλο εκπαιδεύεται ώστε να βρίσκει σχεδόν βέλτιστες λύσεις σε στιγμιότυπα του προβλήματος, μέσα από την παρατήρηση σημάτων επιβράβευσης και ακολουθώντας τους επιτρεπτούς κανόνες. Ακόμα, στη δουλειά των [Cooray and Rupasinghe \(2017\)](#), δοκιμάζεται η βελτιστοποίηση των υπερπαραμέτρων γενετικού αλγορίθμου με τεχνικές βασισμένες σε μηχανική μάθηση με σκοπό την ελαχιστοποίηση της ενέργειας σε ΠΔΟ.

Μια ακόμα δουλειά που μας ενέπνευσε είναι αυτή των [Rousseau \(1990\)](#). Σε αυτή την εργασία, ενσωματώνονται εργαλεία της τεχνητής νοημοσύνης σε ένα υπολογιστικό σύστημα με σκοπό την διευκόλυνση της ανάπτυξης αλγορίθμων δρομολόγησης για τη μετακίνηση οχημάτων. Έτσι υποστηρίζεται η δουλειά αυτών που σχεδιάζουν αλγορίθμους και μπορούν να επικεντρώνονται στην

επίλυση νέων προβλημάτων. Στην ίδια κατεύθυνση εξελίσσεται και η μελέτη των [Nagesh Kadaba \(1991\)](#) όπου συνδυάζονται μαθηματικά μοντέλα, knowledge-based συστήματα, τεχνητά νευρωνικά δίκτυα και προσαρμοτικές γενετικές αναζητήσεις για την βελτιστοποίηση υπερπαραμέτρων μοντέλων δρομολόγησης.

Από μια άλλη σκοπιά, στο άρθρο των [Laura Calvet and Juan \(2017\)](#) δοκιμάζεται ο συνδυασμός τεχνικών επίλυσης του ΠΔΟ με μεθόδους μηχανικής μάθησης. Σε αυτή τη μελέτη επιχειρείται η δημιουργία νέου τύπου υβριδικών αλγορίθμων, μέσα από το συνδυασμό μετα-ευρετικών μεθόδων και μεθόδων τεχνητής νοημοσύνης, για την επίλυση προβλημάτων συνδυαστικής βελτιστοποίησης με δυναμικές εισόδους. Η διαφοροποίηση στις εισόδους πολλές φορές συνδέεται με την τρέχουσα μορφή της λύσης και για την αντιμετώπιση αυτού χρειάζεται επαναληπτικά να ενημερώνονται, με τεχνικές μάθησης, οι νέες εισοδοί που πρόκειται να δοθούν στις μετα-ευρετικές μεθόδους.

2.1.5 ORTEC

Στον χώρο της δρομολόγησης οχημάτων και της παροχής προηγμένων λύσεων σχεδίασης και βελτιστοποίησης δραστηριοποιείται και η εταιρεία ORTEC. Τα προϊόντα και οι υπηρεσίες της ORTEC επικεντρώνονται στην βελτιστοποίηση του στόλου δρομολόγησης και διανομής, των οχημάτων, των παλετών που φορτώνονται, του προγραμματισμού του εργασιακού δυναμικού, της πρόβλεψης ζήτησης, της σχεδίασης του δικτύου διανομής και του ελέγχου των αποθηκών. Παγκοσμίως, η ORTEC διαθέτει περισσότερους από 1200 πελάτες και απασχολεί περίπου 1000 εργαζομένους σε γραφεία σε 13 διαφορετικές χώρες σε όλο τον κόσμο. Συνδυάζοντας την Επιχειρησιακή Έρευνα, το Business και το IT, είναι αποστολή της ORTEC η υποστήριξη των πελατών της σε αποφάσεις στρατηγικές και λειτουργικές, παρέχοντας τους εκλεπτυσμένα πληροφοριακά συστήματα σχεδίασης και βελτιστοποίησης και επαγγελματικές συμβουλές. Η ORTEC Routing and Dispatch (ORD) επικεντρώνεται στην μαθηματική βελτιστοποίηση προβλημάτων δρομολόγησης και διανομής. Το αντικείμενο αυτής της διπλωματικής είναι η παροχή μιας διαφορετικής προσέγγισης επίλυσης ενός προβλήματος της πραγματικής ζωής, ενός πελάτη της ORTEC Routing and Dispatch.

2.2 Η Επιστήμη των Δεδομένων και οι Βασικές Μεθοδολογίες της

2.2.1 Η Επιστήμη των Δεδομένων

Η επιστήμη των δεδομένων είναι το πεδίο μελέτης που συνδυάζει την εξειδίκευση στον τομέα, τις προγραμματιστικές ικανότητες και τη γνώση μαθηματικών και στατιστικής προκειμένου να εξάγει ουσιαστικές γνώσεις από δεδομένα. Οι επιστήμονες που ασχολούνται με αυτή την επιστήμη εφαρμόζουν τεχνικές μηχανικές μάθησης σε νούμερα, κείμενα ή οποιαδήποτε άλλη πληροφορία μπορεί να φανεί χρήσιμη, για να παράγουν συστήματα τεχνητής νοημοσύνης που εκτελούν εργασίες που μέχρι τώρα απαιτούσαν την ανθρώπινη ευφυΐα. Έτσι, αυτά τα συστήματα παράγουν πληροφορίες που αναλυτές μπορούν να χρησιμοποιήσουν και να έχουν νόημα για τις επιχειρήσεις.

Η ανάλυση των δεδομένων, από την άλλη, είναι η διαδικασία του εντοπισμού, της μετατροπής και της μοντελοποίησης των δεδομένων με στόχο την εύρεση χρήσιμων πληροφοριών για την εξαγωγή συμπερασμάτων που υποστηρίζουν τη λήψη των αποφάσεων. Υπάρχουν πολλές σκοπιές

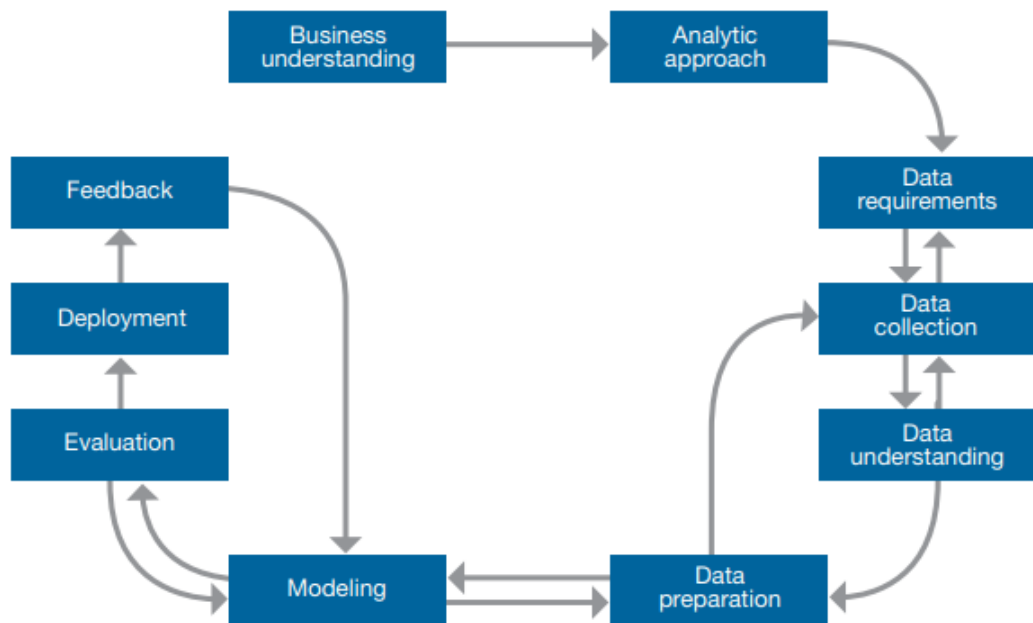
ανάλυσης των δεδομένων που χρησιμοποιούν διαφορετικές τεχνικές και εφαρμόζονται σε πληθώρα τομέων.

2.2.2 Μεθοδολογία

Η μελέτη των δεδομένων πρέπει να έχει σαν βασικό άξονα τα ακόλουθα ερωτήματα:

- Ποιό είναι το πρόβλημα για επίλυση;
- Πως μπορούν να χρησιμοποιηθούν τα δεδομένα για να δώσουν κατάλληλες απαντήσεις;
- Τι δεδομένα χρειάζονται;
- Από που μπορούν να βρεθούν αυτά τα δεδομένα και πώς;
- Είναι τα υπάρχοντα δεδομένα αντιπροσωπευτικά για το πρόβλημα;
- Τι περαιτέρω εργασία απαιτείται για να είναι τα δεδομένα έτοιμα στην μορφή που χρειάζεται;
- Πως μπορούν να οπτικοποιηθούν τα δεδομένα;
- Το μοντέλο που χρησιμοποιήθηκε απαντάει στην αρχική ερώτηση ή πρέπει να τροποποιηθεί;
- Μπορεί να εφαρμοστεί στην πράξη το μοντέλο αυτό;
- Από το μοντέλο που χρησιμοποιήθηκε μπορεί να λυθεί το αρχικό πρόβλημα;

Με βάση τα παραπάνω ερωτήματα μπορεί να δομηθεί και η μεθοδολογία εργασίας:



Σχήμα 2.2: Θεμελιώδης Μεθοδολογία Επιστήμης των Δεδομένων

[From J.B.Rollins (2015)]

Στην παραπάνω εικόνα παρουσιάζονται τα ακόλουθα στάδια:

1. Κατανόηση επιχειρησιακού αντικειμένου
2. Αναλυτική προσέγγιση
3. Απαιτούμενα δεδομένα
4. Συλλογή δεδομένων
5. Κατανόηση δεδομένων
6. Προετοιμασία δεδομένων
7. Μοντελοποίηση
8. Αξιολόγηση
9. Ανάπτυξη μοντέλου
10. Ανατροφοδότηση

2.2.3 Ανάλυση Χρονοσειρών

Σημαντικό μέρος της επιστήμης της ανάλυσης των δεδομένων επικεντρώνεται στην ανάλυση χρονοσειρών. Με τον όρο αυτό περιγράφουμε την ανάλυση πειραματικών δεδομένων που παρατηρήθηκαν σε διαφορετικές χρονικές στιγμές και τη συστηματική προσέγγιση αυτών με μαθηματικές και στατιστικές ερωτήσεις λαμβάνοντας υπόψιν τις χρονικές εξαρτήσεις, σύμφωνα με το βιβλίο [R.H.Shumway \(2017\)](#).

Σύμφωνα με την συνήθη πρακτική, μια χρονοσειρά αρχικά πρέπει να εξεταστεί ως προς την τάση, την εποχικότητα, την τυχαιότητα και την κυκλικότητα όπως καταγράφεται στο άρθρο [Components of Time Series](#). Στη συνέχεια, μια χρονοσειρά πρέπει να αποσυντεθεί στις συνιστώσες της, να επανασυνδεθούν αυτές που είναι απαραίτητες για την παραγωγή προβλέψεων (κυκλικότητα, εποχικότητα, τάση) ενώ να απομακρυνθούν αυτές που οδηγούν σε αστοχίες (τυχαιότητα).

2.3 Μηχανική Μάθηση

2.3.1 Ορισμός Μηχανικής Μάθησης

Μηχανική μάθηση είναι η επιστήμη που βοηθάει τους υπολογιστές να μαθαίνουν και να δρουν όπως οι άνθρωποι, και να βελτιώνονται με την πάροδο του χρόνου με αυτόνομο τρόπο, δίνοντας τους δεδομένα και πληροφορίες σε μορφή παρατηρήσεων και πραγματικών αλληλεπιδράσεων σύμφωνα με την περιγραφή στο [What is Machine Learning?](#).

Πολλές φορές οι έννοιες "Μηχανική Μάθηση" και "Τεχνητή Νοημοσύνη" συγχέονται. Ωστόσο, όπως μπορεί να φανεί και από τον παραπάνω ορισμό, Μηχανική Μάθηση είναι το επιμέρους πεδίο της τεχνητής νοημοσύνης, το οποίο παρέχει γνώση μέσα από δεδομένα και αναπαραστάσεις και χρησιμοποιεί αυτή τη γνώση για να γενικεύει. Η επιστήμη δηλαδή της μηχανικής μάθησης ασχολείται με την επεξεργασία πολύ μεγάλου όγκου δεδομένων για να παράγει συμπεράσματα και έχει τη δυνατότητα να προσαρμοστεί και σε δεδομένα που δεν είναι γνωστά εκ των προτέρων. Αντίθετα τα μοντέλα τεχνητής νοημοσύνης δεν διαχειρίζονται μαζικά δεδομένα και δεν είναι σε θέση

να δώσουν απάντηση σε άγνωστες εισόδους. Συνεπώς, σήμερα που υπάρχει τεράστιος όγκος δεδομένων, τα μοντέλα μηχανικής μάθησης καθίστανται ένα σημαντικό εργαλείο για την αξιοποίηση των αυτών και την εξαγωγή χρήσιμων συμπερασμάτων.

Στις παρακάτω δύο υποενότητες θα παρουσιαστούν συνοπτικά τα διαφορετικά είδη μάθησης καθώς και κάποιες χαρακτηριστικές συναρτήσεις. Ωστόσο, ανεξαρτήτως είδους ή συνάρτησης η μεθοδολογία έχει ως εξής:

1. Αναπαράσταση - ένα σύνολο από ταξινομητές ή μια γλώσσα που καταλαβαίνει ο υπολογιστής
2. Αξιολόγηση - η μετρική που χρησιμοποιείται για να υπολογιστεί η αποτελεσματικότητα του μοντέλου
3. Βελτιστοποίηση - η μέθοδος αναζήτησης, συνήθως επιλέγεται ο ταξινομητής που φέρνει το καλύτερο αποτέλεσμα

2.3.2 Βασικά είδη μάθησης και Χαρακτηριστικοί Αλγόριθμοι

- Επιβλεπόμενη μάθηση
- Μη επιβλεπόμενη μάθηση
- Ενισχυτική μάθηση

Στην πρώτη περίπτωση, συντίθεται ένα μαθηματικό μοντέλο από ένα σύνολο δεδομένων, τα δεδομένα εκπαίδευσης, που διαθέτει τόσο τις εισόδους όσο και τις επιθυμητές εξόδους. Ο αλγόριθμος εκπαιδεύεται με βάση τα δεδομένα και προσπαθεί να προσεγγίσει τη βέλτιστη συνάρτηση που μπορεί να τα περιγράψει, ώστε στη συνέχεια, δεδομένης μιας εισόδου να είναι σε θέση να προβλέψει την επιθυμητή έξοδο. Ο αλγόριθμος δηλαδή προσπαθεί να μοντελοποιήσει τις σχέσεις ανάμεσα στα χαρακτηριστικά της εισόδου και τις επιθυμητές εξόδους ώστε μπορέσει να βρει αντίστοιχα μοτίβα σε άγνωστα δεδομένα και να προβλέψει την αναμενόμενη έκβασή τους. Αυτή η μέθοδος επιλέγεται κυρίως σε προβλήματα ταξινόμησης και αναδρομής. Ορισμένοι αλγόριθμοι επιβλεπόμενης μάθησης είναι οι:

- Naive-Bayes
- Κοντινότεροι Γείτονες
- Δέντρα Αποφάσεων
- Γραμμική Παλινδρόμηση
- Νευρωνικά Δίκτυα
- Μηχανές Διανυσμάτων Υποστήριξης (Support Vector Machines)

Στην δεύτερη περίπτωση, τα δεδομένα δεν διαθέτουν επιθυμητή έξοδο. Αντίθετα, ο αλγόριθμος προσπαθεί να εντοπίσει μοτίβα, να βρει κανόνες, να ομαδοποιήσει και να εξάγει συμπεράσματα από τα δεδομένα. Είναι η μέθοδος που επιδιώκει να περιγράψει καλύτερα τα δεδομένα στο χρήστη. Συνήθως επιλέγεται για ομαδοποίηση και για συσχέτιση. Δύο δημοφιλείς αλγόριθμοι αυτής της κατηγορίας είναι οι:

- Ομαδοποίηση κ-μέσων (k-Means)
- Association Rule Learning

Στην τρίτη περίπτωση, το μοντέλο παρατηρεί τις αλληλεπιδράσεις του με το περιβάλλον, λαμβάνει δράση με βάση αυτές και προσπαθεί να αυξήσει την ανταμοιβή ή να ελαχιστοποιήσει τον κίνδυνο. Εδώ, δηλαδή, ο αλγόριθμος επαναληπτικά επιλέγει μια δράση, λαμβάνει μια ανατροφοδότηση από το περιβάλλον και μεταβαίνει σε μία κατάσταση. Από το ερέθισμα που του δίνεται κάθε φορά προσπαθεί να καταλάβει αν η επιλογή του ήταν καλή. Στην πάροδο του χρόνου μαθαίνει ποια είναι η επιθυμητή συμπεριφορά για να επιτύχει τη μέγιστη ανταμοιβή. Οι αλγόριθμοι που ακολουθούν αποτελούν χαρακτηριστικά παραδείγματα αυτής της μεθόδου:

- Q-Learning
- Monte Carlo
- Temporal difference
- Deep Competitive Networks

2.3.3 Μοντέλο Λογιστικής Παλινδρόμησης

Στο πλαίσιο αυτής της εργασίας, έγινε χρήση τεχνικών μηχανικής μάθησης για επίλυση ενός προβλήματος ταξινόμησης. Δεδομένης της φύσης του προβλήματος χρειάστηκε ένα μοντέλο της πρώτης κατηγορίας, δηλαδή ένα μοντέλο επιβλεπόμενης μάθησης. Από την πληθώρα γνωστών αλγορίθμων που υπάρχουν επιλέχθηκε αυτό της λογιστικής παλινδρόμησης. Πρόκειται για έναν αλγόριθμο που υπολογίζει την σχέση ανάμεσα στην εξαρτημένη μεταβλητή και τις ανεξάρτητες μεταβλητές εκτιμώντας πιθανότητες. Με χρήση της σιγμοειδούς συνάρτησης, αυτές οι πιθανότητες απεικονίζονται σε τιμές μεταξύ του 0 και του 1. Ανάλογα με τη σχέση αυτής της τιμής με ένα κατώφλι αποφασίζεται η τιμή της εξαρτημένης μεταβλητής για τις δοσμένες τιμές των ανεξάρτητων.

Στόχος είναι η μεγιστοποίηση της πιθανότητας ορθής κατηγοριοποίησης ενός τυχαίου σημείου του χώρου, γι' αυτό χρησιμοποιείται κάποιος αλγόριθμος που στοχεύει στην προσέγγιση της βέλτιστης ευθείας διαχωρισμού.

Πιο αναλυτικά, αν θεωρήσουμε ότι έχουμε n μεταβλητές πρόβλεψης x_1, x_2, x_n και μια δυαδική μεταβλητή Bernoulli Y , με $p = P(Y = 1)$. Θεωρούμε γραμμική σχέση ανάμεσα στις μεταβλητές πρόβλεψης και log-odds του $Y = 1$, όπου log-odds είναι ο λογάριθμος των αποδόσεων $p/(1 - p)$. Το παραπάνω μπορεί να εκφραστεί μαθηματικά ως:

$$l = \log_b \frac{p}{1-p} = b_0 + b_1 x_1 + \dots + b_n x_n$$

,όπου l η log-odds και b η βάση του λογαρίθμου.

Έτσι έχουμε:

$$\frac{p}{1-p} = b^{b_0 + b_1 x_1 + \dots + b_n x_n}$$

Δηλαδή για $Y = 1$ έχουμε:

$$p = \frac{b^{b_0 + b_1 x_1 + \dots + b_n x_n}}{b^{b_0 + b_1 x_1 + \dots + b_n x_n} + 1} = \frac{1}{1 + b^{-(b_0 + b_1 x_1 + \dots + b_n x_n)}}$$

Έτσι, έχοντας βρει με βάση κάποιον αλγόριθμο τα βέλτιστα b_i μπορεί να υπολογιστεί το p και εν συνεχεία να προβλεφθεί η αναμενόμενη τιμή της εξόδου.

Το μοντέλο που μόλις παρουσιάστηκε κάνει χρήση της συνάρτησης λογιστικής παλινδρόμησης. Για να έχει καλυφθεί πλήρως το θεωρητικό υπόβαθρο, σε αυτό το σημείο γίνεται μια σύντομη αναφορά και στην αντίστροφη της συνάρτησης αυτής. Η αντίστροφη της συνάρτησης λογιστικής παλινδρόμησης, γνωστή και ως *logit* ή *log-odds* είναι ο λογάριθμος της πιθανότητας $\frac{p}{1-p}$, όπου p η πιθανότητα. Δηλαδή, η συνάρτηση *logit* μπορεί να εκφραστεί ως:

$$\text{logit}(p) = \frac{p}{1-p} = \log(p) - \log(1-p) = -\log\left(\frac{1}{p} - 1\right)$$

Από την παραπάνω έκφραση μπορούμε να δούμε πράγματι πως αν αντιστρέψουμε αυτή τη συνάρτηση προκύπτει η παραπάνω. Φαίνεται, δηλαδή, ότι:

$$\text{logit}^{-1}(a) = \text{logistic}(a) = \frac{1}{1+e^{-a}} = \frac{\exp(a)}{\exp(a)+1}$$

2.3.4 Μετρικές Αξιολόγησης Μεθόδων Μηχανικής Μάθησης

Για τη μέτρηση της αποδοτικότητας του μοντέλων μηχανικής μάθησης χρησιμοποιούνται συνήθως οι ακόλουθες μετρικές:

- R^2 : Εκφράζει την ποιότητα προσαρμογής του μοντέλου και μπορεί να υπολογιστεί ως το ποσοστό της μεταβλητότητας της εξαρτημένης μεταβλητής που εξηγείται από το μοντέλο. Μαθηματικά μπορεί να εκφραστεί ως:

$$R^2 = 1 - \frac{SS_R}{SS_T}$$

,όπου SS_R η απόκλιση των δεδομένων από το "χειρότερο μοντέλο" (μέση τιμή) και SS_T η απόκλιση των δεδομένων από το "καλύτερο μοντέλο" (ευθεία)

- Ακρίβεια (*precision*) : Είναι ο λόγος των σωστών προβλέψεων για τη θετική κλάση ως προς όλες τις προβλέψεις που έγιναν για τη θετική κλάση, δηλαδή ισούται με:

$$\text{Precision} = \frac{TP}{TP+FP}$$

,όπου TP οι σωστές προβλέψεις για την θετική κλάση και FP οι λανθασμένες προβλέψεις για τη θετική κλάση

- Ανάκλιση (*recall*) : Είναι ο λόγος των σωστών προβλέψεων για τη θετική κλάση ως προς το άθροισμα των σωστών προβλέψεων για τη θετική κλάση με τις λάθος προβλέψεις για την αρνητική κλάση και έχει ως εξής:

$$\text{Recall} = \frac{TP}{TP+FN}$$

,όπου FN οι λανθασμένες προβλέψεις για τη αρνητική κλάση

- f1-score : Είναι μια μετρική ακρίβειας που ισούται με τον αρμονικό μέσο της ακρίβειας και της ανάκλισης, προκύπτει δηλαδή από την παρακάτω μαθηματική έκφραση:

$$F1score = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

Ακόμα, μια σημαντική πληροφορία για την περιγραφή ενός συνόλου χαρακτηριστικών είναι ο συντελεστής αυτοσυσχέτισης τους. Ως αυτοσυσχέτιση ορίζεται η μέτρηση της γραμμικής σχέσης μεταξύ δύο ποσοτικών μεγεθών. Όπως περιγράφεται στο άρθρο [Why Feature Correlation Matters...A Lot!](#) θετική αυτοσυσχέτιση σημαίνει ότι όταν το χαρακτηριστικό A αυξάνεται, τότε και το B αυξάνεται και αντίστροφα. Κατ' αναλογία, αρνητική αυτοσυσχέτιση δείχνει ότι η αύξηση του ενός χαρακτηριστικού οδηγεί στη μείωση του άλλου.

Τέλος, μια καθοριστική μετρική για την αξιολόγηση μιας μεθόδου ταξινόμησης είναι η μετρική AUC_score. Η μετρική αυτή, σύμφωνα με το [Useful properties of ROC curves and AUC scoring](#) λειτουργεί ως εξής: Για κάθε παρατήρηση υπολογίζεται μια πιθανότητα, στην συνέχεια ταξινομούνται τα δεδομένα με βάση την τιμή της προβλεπόμενης πιθανότητας κι αξιολογείται το πόσο κοντά είναι αυτή η διάταξη στην βέλτιστη σειρά που μπορούν να διαταχθούν τα δεδομένα. Η περιγραφή αυτή δείχνει ότι η μετρική αυτή είναι ανεξάρτητη της κλίμακας ή οποιασδήποτε μεταμόρφωσης που διατηρεί σχετική διάταξη.

2.3.5 Τεχνικές Βελτίωσης Μοντέλων Μηχανικής Μάθησης

Αρχικά, μια πρακτική πρακτική που επιλέγεται συχνά για την προετοιμασία ενός συνόλου δεδομένων είναι η κανονικοποίηση των χαρακτηριστικών. Αυτή η τεχνική στοχεύει στην μετατροπή των αριθμητικών τιμών του συνόλου στην ίδια κλίμακα χωρίς όμως να εξαλείφει τις διαφορές στο εύρος τιμών σύμφωνα με το [Why Data Normalization is necessary for Machine Learning models](#). Διαφορετικά είδη κανονικοποίησης καταγράφονται, και ανάλογα με τα χαρακτηριστικά, ο σχεδιασμός του μοντέλου καλείται να επιλέξει ανάμεσα στα εξής:

- Απλή κανονικοποίηση (data normalization)
- Κανονικοποίηση γύρω από το μηδέν (data standardization)
- Κανονικοποίηση ανάμεσα σε μία ελάχιστη και μία μέγιστη τιμή (MinMax Scaling)
- Κανονικοποίηση ανάμεσα στο μηδέν και μία μέγιστη τιμή (Max Abs Scaling)

Κατά την προεπεξεργασία ενός συνόλου δεδομένων για ταξινόμηση, πρέπει να καταγραφεί των πλήθος των παρατηρήσεων των διαφορετικών κλάσεων. Όταν υπάρχει μεγάλη διαφορά ανάμεσα στις τιμές αυτές απαραίτητο είναι να γίνει oversampling ή undersampling. Αυτές οι τεχνικές χρησιμοποιούνται για την προσαρμογή της κατανομής κλάσεων του συνόλου και τελικά δίνουν ένα ισορροπημένο data set.

Στη βιβλιογραφία επιλέγεται συχνά η χρήση Grid Search για την εύρεση των βέλτιστων παραμέτρων ενός μοντέλου. Με την τεχνική αυτή γίνεται ένα σκανάρισμα των δεδομένων για την εύρεση της βέλτιστης παραμετροποίησης. Η παραμετροποίηση αυτή δεν περιορίζεται την εφαρμογή ενός μόνο μοντέλου, αλλά μπορούν να δοκιμαστούν οι βέλτιστες παράμετροι για διαφορετικά μοντέλα

μηχανικής μάθησης. Ωστόσο, πρέπει να λαμβάνεται υπόψιν της υπολογιστικής πολυπλοκότητας της μεθόδου αυτής, και οι δοκιμές να μην είναι άσκοπες.

Επίσης, για την ορθή εκπαίδευση ενός μοντέλου είναι χρήσιμο να μην δίνονται περιττές πληροφορίες. Για τον εντοπισμό των προαιρετικών χαρακτηριστικών χρησιμοποιούνται μέθοδοι επιλογής χαρακτηριστικών. Στο πλαίσιο αυτής της εργασίας δοκιμάστηκε η μέθοδος επιλογής RFE. Η λειτουργία της μπορεί να περιγραφεί ως εξής: Χρησιμοποιείται ένας εξωτερικός estimator για την ανάθεση βαρών στα χαρακτηριστικά με σκοπό την επιλογή χαρακτηριστικών μέσα από αναδρομικούς περιορισμούς του συνόλου τους μέχρι να φτάσει στο επιθυμητό πλήθος.

2.3.6 Προκλήσεις και Περιορισμοί

Τα πιο συνήθη προβλήματα που μπορεί να παρουσιάσει ένα μοντέλο είναι η υπερεκπαίδευση και η διαστατικότητα. Στην περίπτωση του πρώτου, το μοντέλο τροφοδοτείται με υπερβολική πληροφορία για τα δεδομένα με αποτέλεσμα να μην έχει την ικανότητα να γενικεύσει. Στην περίπτωση του δεύτερου, εισάγονται πολλά χαρακτηριστικά στον αλγόριθμο με αποτέλεσμα να αυξάνεται η πολυπλοκότητα του και να καθίσταται δύσκολη η κατανόηση των δεδομένων.

Παρ' ότι πολλές φορές τα αποτελέσματα των μετρικών φαίνονται να βελτιώνονται όσο δίνονται περισσότερες πληροφορίες, πρέπει να αποφεύγονται οι περιττές πληροφορίες. Αυτό προκειμένου τόσο να αποφευχθούν τα παραπάνω προβλήματα όσο και να περιοριστεί ο χρόνος που χρειάζεται το μοντέλο για να εκπαιδευτεί και έπειτα να εξάγει συμπεράσματα.

Από την άλλη, δεν πρέπει να δίνονται λιγότερα από τα απαραίτητα χαρακτηριστικά για τα δεδομένα καθώς τότε το μοντέλο δεν θα έχει αρκετή πληροφορία για να μάθει και έτσι θα υποεκπαιδευτεί. Συνεπώς, πρέπει να διατηρείται πάντα η σωστή αναλογία μεταξύ του όγκου των δεδομένων και της πολυπλοκότητας του μοντέλου, αν και σε πολλές φορές δεν είναι εύκολο να βρεθεί η σωστή ισορροπία.

Όταν κάποιος ερευνητής ακολουθήσει τα παραπάνω και ερμηνεύσει ορθά τα δεδομένα, μπορεί να εκμεταλλευτεί τη δύναμη αυτών των μοντέλων για την αξιοποίηση της διαθέσιμης πληροφορίας και την εξαγωγή συμπερασμάτων που στα οποία δεν θα μπορούσε να καταλήξει με άλλα εργαλεία.

Κεφάλαιο 3

Σχεδιασμός και Υλοποίηση

Η μεθοδολογία που ακολουθήθηκε για την εκτέλεση αυτής της διπλωματικής εργασίας εξελίχθηκε σε τρεις βασικούς άξονες: α) την ανάλυση και την επεξεργασία δεδομένων, β) την ανάπτυξη ενός μοντέλου μηχανικής μάθησης και γ) την αρχικοποίηση ενός περιβάλλοντος για διεξαγωγή πειραμάτων και συγκρίσεις με αλγόριθμο βελτιστοποίησης. Στη συνέχεια περιγράφεται η σταδιακή εξέλιξη και στις τρεις αυτές κατευθύνσεις.

3.1 Επεξεργασία Δεδομένων

3.1.1 Δεδομένα

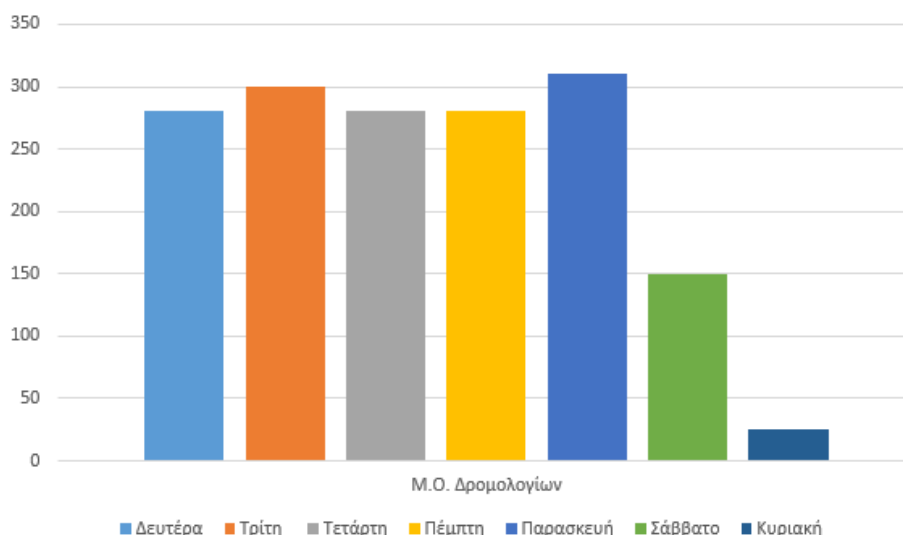
Στην εργασία αυτή παρουσιάζεται και επιλύεται ένα πρόβλημα δρομολόγησης οχημάτων. Τα δεδομένα που χρησιμοποιούνται αποτελούν τα δεδομένα από το παραγωγικό περιβάλλον μιας γνωστής εταιρείας που δραστηριοποιείται στην Ελλάδα στο χώρο του λιανικού εμπορίου τροφίμων. Στην περίπτωση που μελετάται, πρόκειται για δεδομένα που περιγράφουν τη μεταφορά αγαθών από αποθήκες σε καταστήματα ή σε άλλες αποθήκες, στη διάρκεια ενός έτους.

Αναλυτικότερα, η επιχείρηση αυτή χρησιμοποίησε το λογισμικό της εταιρείας ORTEC προκειμένου να βελτιστοποιήσει τα δρομολόγια της. Συνεπώς, οι διαδρομές που καταγράφονται είναι οι έξοδοι του αλγορίθμου βελτιστοποίησης της εταιρείας ORTEC που όμως έχουν υποστεί κάποιες τροποποιήσεις από τους χειριστές του συστήματος (εργαζόμενοι της εταιρείας σουπερ μάρκετ). Ως είσοδοι στο σύστημα δίνονται οι παραγγελίες των καταστημάτων και των αποθηκών, οι διευθύνσεις τους, τα διαθέσιμα οχήματα καθώς και όσες ακόμα πληροφορίες σχετίζονται με τα παραπάνω και είναι απαραίτητες για το σχηματισμό των δρομολογίων. Τα δεδομένα αυτά μετά την εξαγωγή τους από τη βάση δεδομένων του περιβάλλοντος εργασίας υπέστησαν επεξεργασία και ανάλυση για να είναι δυνατή η κατανόηση τους.

3.1.2 Παρατηρήσεις επί των Δεδομένων

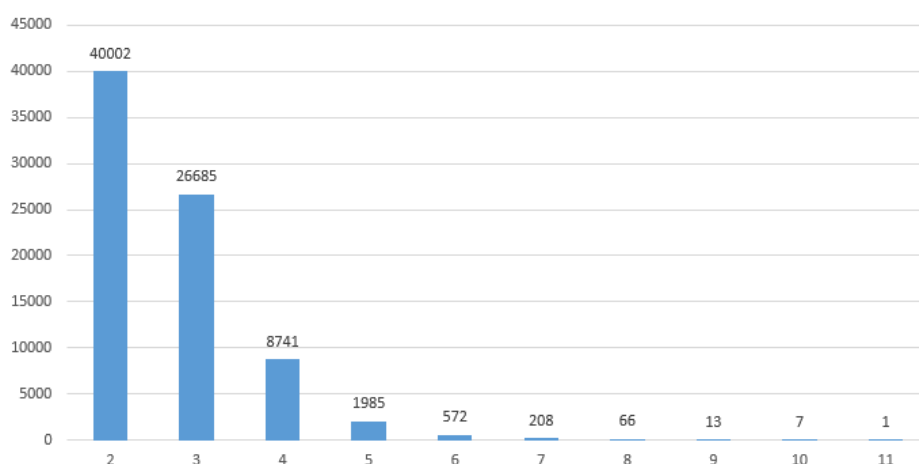
Για αρχή μελετήθηκαν τα δεδομένα ενός έτους χωρίς να υποστούν κάποια τροποποίηση. Σε αυτά τα δεδομένα πέρα από ροές αγαθών μεταξύ αποθηκών και καταστημάτων, καταγράφονται και ροές αγαθών μεταξύ προμηθευτών και αποθηκών. Παρατηρείται ότι το κάθε κατάστημα τροφοδοτείται από πολλές αποθήκες καθώς η καθεμία από αυτές διατηρεί συγκεκριμένα είδη προϊόντων.

Εν συνεχεία, για την κατανόηση του προβλήματος μελετήθηκαν οι βασικές ροές αγαθών, το πλήθος των δρομολογίων, ο όγκος των παραγγελιών και οι διαφοροποιήσεις ανά ημέρα. Στην παρακάτω εικόνα φαίνεται ο μέσος αριθμός δρομολογίων που εκτελούνταν ανά ημέρα.



Σχήμα 3.1: Μέσο Πλήθος Δρομολογίων ανά Ημέρα της Εβδομάδας

Στη συνέχεια καταγράφηκε το πλήθος των στάσεων που πραγματοποιούσε κάθε δρομολόγιο. Αυτές υπολογίστηκαν να κυμαίνονται από 2 (ζεύγος αποθήκης-καταστήματος), έως και 11 σε μόλις ένα δρομολόγιο (συνδυασμοί μίας παραπάνω αποθηκών και πολλών καταστημάτων).

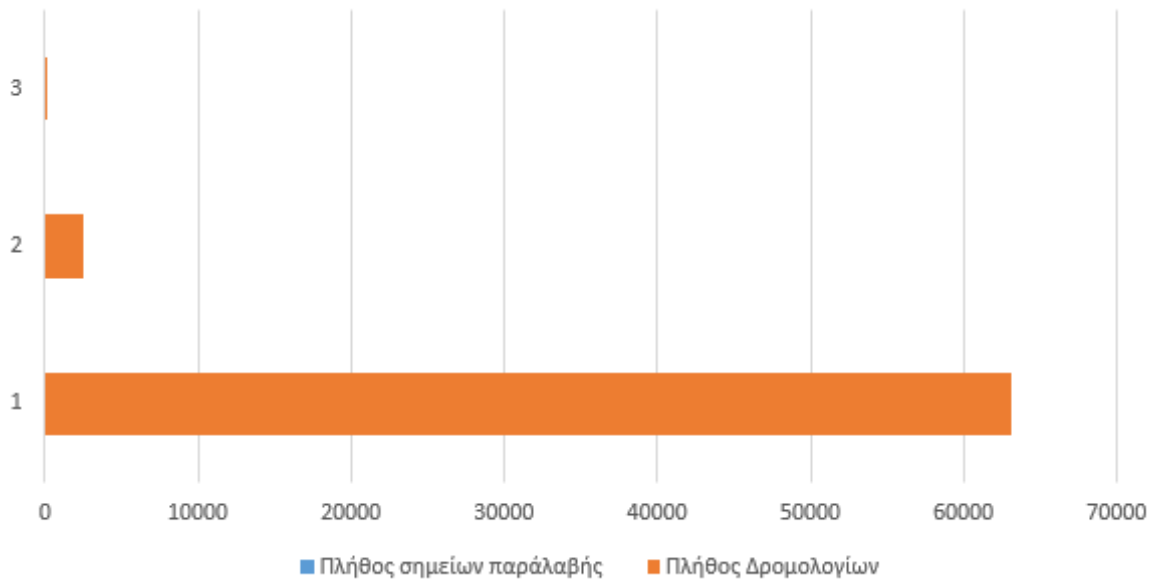


Σχήμα 3.2: Πλήθος Δρομολογίων ανά Αριθμό Στάσεων

Δεδομένου ότι τα δεδομένα υπό εξέταση εξελίσσονταν στον άξονα του χρόνου, το πρόβλημα της ανάλυσης των δεδομένων αντιμετωπίστηκε ως ένα πρόβλημα μελέτης χρονοσειρών. Το δείγμα των δεδομένων που παρατηρήθηκαν χρειάστηκε να εξεταστεί ως προς την τάση, την εποχικότητα, την τυχαιότητα και την κυκλικότητα όπως καταγράφεται στο άρθρο. Μετά την αφαίρεση των συνιστωσών της τάσης και της εποχικότητας, παρατηρήθηκε ότι υπάρχουν ορισμένα στοιχεία ακόμα που πρέπει να διαγραφούν καθώς αλλοίωναν την εικόνα του δείγματος. Τέτοια στοιχεία ήταν καταστήματα που δεν υπήρχαν από την αρχή, καταστήματα με ελάχιστες καταχωρήσεις, καθώς και ροές αγαθών από προμηθευτές σε καταστήματα που ήταν ιδιαίτερα σπάνιες και σε μη τυχαίες χρονικές στιγμές.

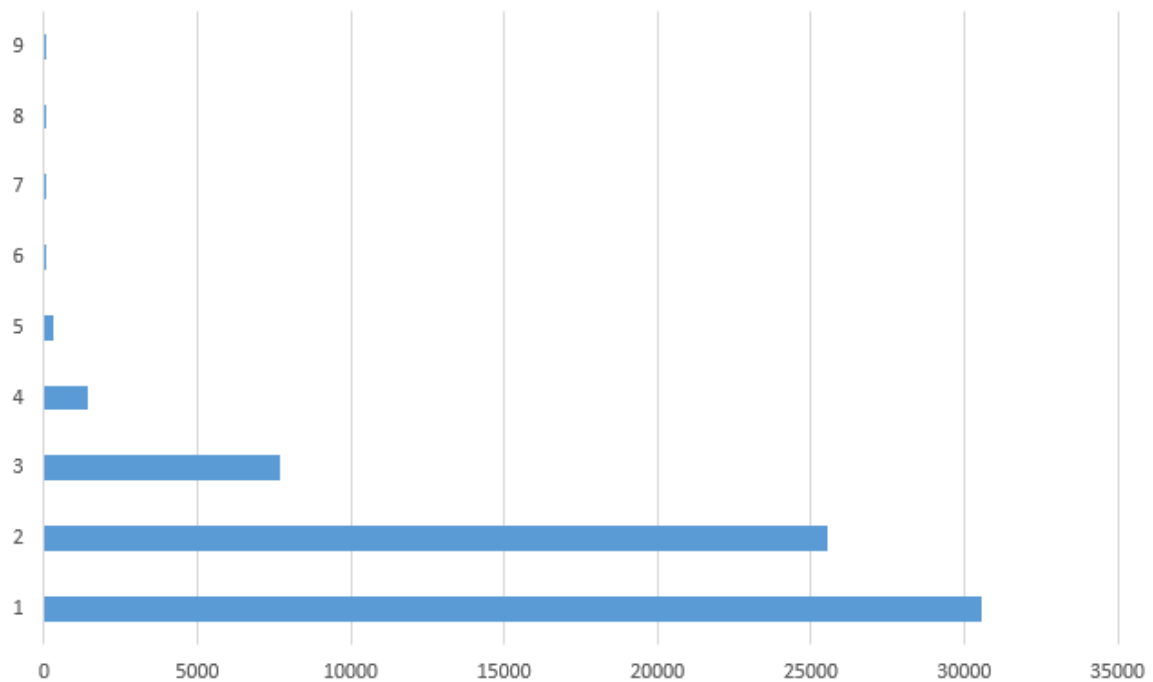
Μετά την επεξεργασία των δεδομένων, μελετήθηκε το πλήθος των δρομολογίων ανάλογα με τον αριθμό των στάσεων για παραλαβή και για παράδοση, για τις ροές που καταλήγουν σε κάποιο

κατάστημα. Στη συνέχεια ακολουθεί ένα διάγραμμα που απεικονίζει το πλήθος δρομολογίων που καταγράφεται για κάθε πλήθος των σημείων παραλαβής του:



Σχήμα 3.3: Πλήθος Δρομολογίων ανά Αριθμό Στάσεων Παραλαβής

Ακολουθεί το διάγραμμα του πλήθους των δρομολογίων ως συνάρτηση των σημείων παράδοσης του:



Σχήμα 3.4: Πλήθος Δρομολογίων ανά Αριθμό Στάσεων Παράδοσης

Μελετήθηκε ακόμα η συσχέτιση μεταξύ της συνολικής διανύομενης απόστασης και του πλήθους των στάσεων, η συχνότητα παρουσίασης ζευγών ή ομάδων καταστημάτων στο ίδιο δρομολόγιο, η διαφοροποίηση αυτών ανά ημέρα της εβδομάδας ή ανά συνολικό όγκο προς μεταφορά και άλλα. Για τους παραπάνω υπολογισμούς, την παραγωγή ποσοτικών μεγεθών για την μαθηματική περιγραφή τους και την ποιοτική απεικόνιση χρησιμοποιήθηκαν το πρόγραμμα Excel και βιβλιοθήκες της γλώσσας προγραμματισμού Python.

Μετά τη συστηματική παρατήρηση των δεδομένων, φάνηκε να υπάρχει ομοιότητα ανάμεσα στα διαφορετικά στιγμιότυπα του προβλήματος. Ακόμα, παρατηρήθηκε ότι παράγοντες όπως το είδος του καταστήματος, το μέγεθος της παραγγελίας, το μέγεθος της διανυόμενης απόστασης για την παράδοση της είναι καθοριστικοί για το αν η παραγγελία θα δρομολογηθεί μόνη της ή θα συνδυαστεί με άλλες. Αυτό, σε συνδυασμό με το σημαντικό πλήθος δρομολογίων μιας παραλαβής - μιας παράδοσης, οδήγησαν στην αντιμετώπιση του προβλήματος μεταφοράς αγαθών ως ένα πρόβλημα χαρακτηρισμού μιας παραγγελίας ως παραγγελία που θα δρομολογηθεί μόνη της ή παραγγελία που χρειάζεται να συνδυαστεί με άλλες για να συνθέσει ένα δρομολόγιο.

Η περιγραφή αυτή καθόρισε ως στόχο της διπλωματικής αυτής την ανάπτυξη ενός μοντέλου που αξιοποιεί την γνώση από προηγούμενα στιγμιότυπα του προβλήματος για να διαχωρίσει βέλτιστα τις παραγγελίες σε δύο κατηγορίες, αυτές που μπορούν να δρομολογηθούν αυτόνομα και αυτές που απαιτείται ο συνδυασμός τους με άλλες.

3.2 Σχεδίαση Μοντέλου Μηχανικής Μάθησης

3.2.1 Κατασκευή Συνόλου Δεδομένων

Με σκοπό την ταξινόμηση των παραγγελιών σε δύο κατηγορίες έπρεπε να κατασκευαστεί ένα σύνολο από δεδομένα για την αποτελεσματική περιγραφή τους. Μετά από την παρατήρηση των δεδομένων κρίθηκε σκόπιμο να ομαδοποιηθούν ορισμένες παραγγελίες. Αναλυτικότερα, οι παραγγελίες που είχαν ίδια μέρα παράδοσης, ίδιο σημείο παραλαβής, ίδιο σημείο παράδοσης και είχαν μπει στο ίδιο δρομολόγιο για το πρόβλημα μας είναι μια οντότητα και έτσι κρίθηκε σκόπιμος ο συνδυασμός του σε μία ενιαία παραγγελία. Συντέθηκε, έτσι, ένας πίνακας που κάθε γραμμή αντιστοιχεί σε μια σύνθετη παραγγελία και οι διάφορες στήλες περιγράφουν την χαρακτηριστικά της. Τα διάφορα στοιχεία που χρησιμοποιήθηκαν για να περιγράψουν τις παραγγελίες υπέστησαν πολλές τροποποιήσεις κατά την εξέλιξη του μοντέλου προκειμένου να παραχθούν οι βέλτιστες δυνατές προβλέψεις. Στην πορεία των πειραμάτων προστέθηκαν και αφαιρέθηκαν χαρακτηριστικά, δημιουργήθηκαν νέα από περαιτέρω ανάλυση και επεξεργασία των δεδομένων, τροποποιήθηκε η μορφή τους (τύπος, πλήθος διαφορετικών τιμών κλπ.) καθώς και δοκιμάστηκε η σύνθεση νέων χαρακτηριστικών από συνδυασμούς άλλων.

3.2.2 Το Πρώτο Μοντέλο

Για τις πρώτες δοκιμές δημιουργήθηκε ένα απλό μοντέλο λογιστικής παλινδρόμησης του οποίου είσοδος είναι το σύνολο χαρακτηριστικών που περιγράφεται παρακάτω, και παράγει ως έξοδο την αναμενόμενη τιμή της εξαρτημένης μεταβλητής. Επιπρόσθετα, δίνεται ένα διάνυσμα που αποτελεί την ετικέτα της κάθε παραγγελίας, δηλαδή περιγράφει σε ποια κατηγορία πρέπει να ταξινομηθεί. Πρόκειται, δηλαδή, για ένα μοντέλο που λαμβάνει έναν πίνακα που περιγράφει τα χαρακτηριστικά των παραγγελιών και με βάση αυτά μαθαίνει να προβλέπει το αν μία νέα παραγγελία θα σχηματίσει ένα ανεξάρτητο δρομολόγιο ή θα συνδυαστεί με άλλες.

Το πρώτο σύνολο δεδομένων αποτελείται από τα ακόλουθα χαρακτηριστικά:

1. Μέγεθος παραγγελίας (όγκος) : ακέραιες τιμές
2. Μέγεθος παραγγελίας (κιλά) : δεκαδικές τιμές

3. Μέρα τις εβδομάδας : ακέραιες τιμές
4. Κωδικός σημείου παραλαβής : ακέραιες τιμές
5. Κωδικός σημείου παράδοσης : ακέραιες τιμές
6. Συντεταγμένες σημείου παραλαβής : δεκαδικές τιμές
7. Συντεταγμένες σημείου παράδοσης : δεκαδικές τιμές
8. Απόσταση ανάμεσα στο σημείο παραλαβής και στο σημείο παράδοσης : δεκαδικές τιμές
9. Ετικέτα παραγγελίας : δυαδικές τιμές (ψευδή τιμή για τις παραγγελίες που πρέπει να συνδυαστούν με άλλες και αληθή για αυτές που συνθέτουν μόνες τους ένα δρομολόγιο)

Το μοντέλο αναπτύχθηκε σε γλώσσα Python. Εκεί, χρησιμοποιήθηκε η συνάρτηση LogisticRegression από τη βιβλιοθήκη scikit-learn με solver = liblinear . Χρησιμοποιήθηκε επίσης η συνάρτηση train_test_split της ίδιας βιβλιοθήκης για το διαχωρισμό των δεδομένων σε δύο ομάδες, αυτή που θα χρησιμοποιηθεί για την εκπαίδευση του μοντέλου (σύνολο εκπαίδευσης) και αυτή που θα χρησιμοποιηθεί για την δοκιμή του μοντέλου (σύνολο δοκιμών). Τα δεδομένα μοιράστηκαν στα δύο σύνολα με αναλογία 66% και 33% αντίστοιχα. Για την αξιολόγηση του μοντέλου υπολογίστηκαν οι μετρικές R^2 , Precision, Recall και F1_score και οι τιμές τους διαμορφώνονται ως εξής:

R^2	0.91213
-------	---------

Πίνακας 3.1: Αποτελέσματα Πρώτου Μοντέλου

Class	Precision	Recall	F1-score
0.0	0.93	0.95	0.94
0.1	0.86	0.80	0.83

3.2.3 Εξέλιξη Μοντέλου

3.2.3.1 Αρχικές Αλλαγές - Καλύτερη Περιγραφή των Δεδομένων

Το πρώτο μοντέλο χρειάστηκε να υποστεί αρκετή επεξεργασία προκειμένου να βελτιωθεί η απόδοσή του. Έμφαση δόθηκε στην αλλαγή των χαρακτηριστικών που του δίνονταν τόσο ως προς το περιεχόμενο όσο και ως προς τη μορφή. Επίσης, χρήσιμη φάνηκε η κανονικοποίηση των χαρακτηριστικών προκειμένου διαφορές στην τάξη μεγέθους των διαφορετικών χαρακτηριστικών να μην επηρεάζουν την σημαντικότητα τους άλλα η έμφαση να δίνεται στη βαρύτητα που τους έχει δοθεί. Ακόμα, δοκιμάστηκαν άλλοι solvers στην συνάρτηση λογιστικής παλινδρόμησης, καθώς και αναζητήθηκαν νέες μετρικές και διαγράμματα για την αξιολόγηση του.

Σε πρώτη φάση, έγινε μια προσπάθεια καλύτερης περιγραφής των σημείων παραλαβής και παράδοσης κάθε παραγγελίας. Έτσι, ο κωδικός του σημείου παράδοσης αντικαταστάθηκε από χαρακτηριστικά όπως το ποσοστό των δρομολογίων που το περιλαμβάνουν και που πηγαίνουν μόνο σε αυτό ως προς όλα τα δρομολόγια που το περιλαμβάνουν, το πλήθος των οχημάτων που μπορεί να εξυπηρετήσει το κατάστημα, το πλήθος των καταστημάτων που βρίσκονται σε μικρή

απόσταση από αυτό και το μέσο πλήθος παραλαβών που έχει ανά τύπο προϊόντος. Από την άλλη, με σκοπό την καλύτερη περιγραφή του σημείου παραλαβής, προστέθηκε στα δεδομένα εισόδου ο μέσος αριθμός που έχει το σημείο αυτό ανά ημέρα.

Για την προεπεξεργασία του συνόλου δεδομένων δοκιμάστηκε απλή κανονικοποίηση, κανονικοποίηση γύρω από το μηδέν, Min-Max κανονικοποίηση και Max Abs. Πιο αποδοτική μέθοδος φάνηκε αυτή της κανονικοποίησης γύρω από το μηδέν.

Εν συνεχεία, μετά την παρατήρηση ότι η κλάση που απεικονίζεται στο 1 αποτελεί περίπου το 30% δοκιμάστηκε oversampling χωρίς να φαίνεται όμως να έχει σημαντική επίδραση στην αποδοτικότητα του μοντέλου. Έπειτα, δοκιμάστηκε αναζήτηση πλέγματος (Grid Search) για την εύρεση των βέλτιστων παραμέτρων του μοντέλου.

Οι παραπάνω αλλαγές είχαν τις επιδράσεις που ακολουθούν στα αποτελέσματα του μοντέλου:

R^2	0.9462
-------	--------

Πίνακας 3.2: Αποτελέσματα Διορθωμένου Μοντέλου

Class	Precision	Recall	F1score
0.0	0.97	0.96	0.96
0.1	0.88	0.92	0.90

3.2.3.2 Περιορισμός Διαφορετικών Τιμών των Χαρακτηριστικών

Οι πολλές διαφορετικές τιμές των χαρακτηριστικών φάνηκαν να αυξάνουν την πολυπλοκότητα του μοντέλου χωρίς να προσφέρουν σημαντικές πληροφορίες. Γι' αυτό περιορίστηκαν οι διαφορετικές τιμές τους, ομαδοποιήθηκαν κάποιες άλλες ή τροποποιήθηκε η μορφή τους.

Ακόμα, παρατηρήθηκε ότι ένα από τα προηγούμενα χαρακτηριστικά δεν είχε την αναμενόμενη επίδραση στη συμπεριφορά του μοντέλου. Πρόκειται για τα χαρακτηριστικό που εκφράζει το ποσοστό των παραγγελιών που περιέχουν το σημείο παράδοσης της παραγγελίας και έχουν σχηματίσει ανεξάρτητα δρομολόγια ως προς το συνολικό πλήθος των παραγγελιών που παραδίδονται σε αυτό, δηλαδή:

$$\frac{\text{nof p2p orders to the delivery address}}{\text{total nof orders to the delivery address}}$$

Αυτό το χαρακτηριστικό δίνει στην πραγματικότητα την πιθανότητα μια παραγγελία να κριθεί ως ανεξάρτητη παραγγελία με βάση το σημείο παράδοσης. Ωστόσο, πρόκειται μια πληροφορία που έχει προκύψει από στατιστική ανάλυση των δεδομένων και απαντάει στο ερώτημα που επιθυμεί να απαντήσει συνολικά το μοντέλο μας. Για την σωστή αξιοποίηση αυτής της πληροφορίας, που ουσιαστικά έχει προκύψει από μια συνάρτηση λογιστικής παλινδρόμησης, χρειάζεται η αντιστροφή αυτής της τιμής. Έτσι, το ποσοστό που είχε υπολογιστεί δόθηκε ως είσοδος στην συνάρτηση logit και η έξοδος αυτού έγινε το νέο χαρακτηριστικό του μοντέλου. Έχουμε, δηλαδή:

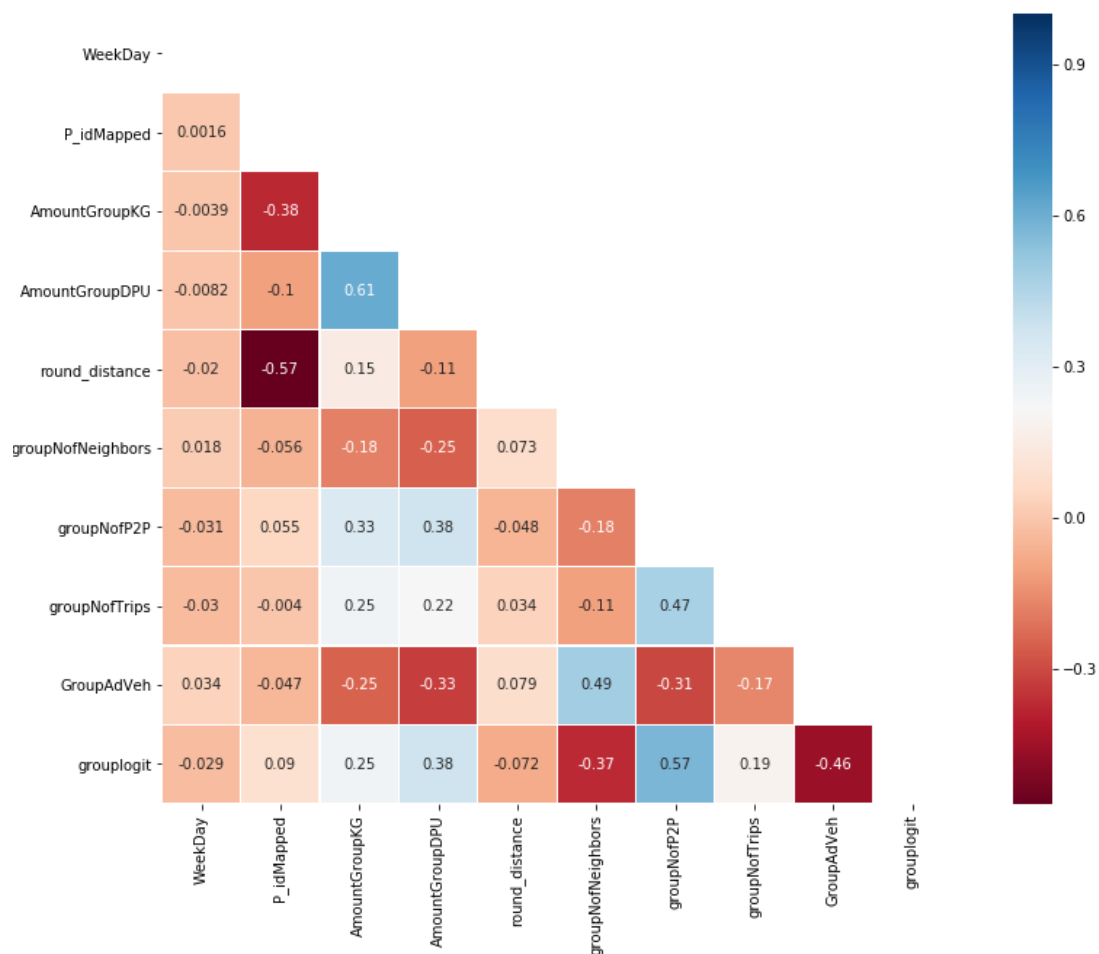
$$\text{logit}\left(\frac{\text{nof p2p orders to the delivery address}}{\text{total nof orders to the delivery address}}\right)$$

Έτσι, οι τιμές ορισμένων χαρακτηριστικών ομαδοποιήθηκαν ως εξής:

1. Μέγεθος παραγγελίας (όγκος) : απεικόνιση σε 4 ομάδες (4 διαφορετικές ακέραιες τιμές)
2. Μέγεθος παραγγελίας (κιλά) : ομοίως με από πάνω, απεικόνιση σε 6 ομάδες
3. Μέρα τις εβδομάδας : 7 ακέραιες τιμές
4. Κωδικός σημείου παραλαβής : 4 ακέραιες τιμές
5. Απόσταση ανάμεσα στο σημείο παραλαβής και στο σημείο παράδοσης : απεικόνιση σε 3 ομάδες
6. Πλήθος καταστημάτων κοντά στο σημείο παράδοσης : απεικόνιση σε 3 ομάδες
7. Πλήθος δρομολογίων δύο στάσεων (στάση για παραλαβή και παράδοση) που περιέχουν το σημείο παράδοσης : απεικόνιση σε 3 ομάδες
8. Πλήθος δρομολογίων που περιέχουν το σημείο παράδοσης : απεικόνιση σε 3 ομάδες
9. Πλήθος οχημάτων που μπορούν να εξυπηρετήσουν το σημείο παράδοσης : δυαδικό χαρακτηριστικό
10. Αντίστροφη συνάρτησης λογιστικής παλινδρόμησης για ποσοστό δρομολογίων δύο στάσεων ως προς τα συνολικά δρομολόγια του σημείου παράδοσης της παραγγελίας : απεικόνιση σε 3 ομάδες

Ιδιαίτερη σημασία δόθηκε στο τελευταίο χαρακτηριστικό. Το ποσοστό που είχε υπολογιστεί, αρχικά, δόθηκε σαν είσοδος στην αντίστροφη της συνάρτησης λογιστικής παλινδρόμησης και έτσι υπολογίστηκε η νέα του τιμή. Αυτό μετά κατατάχθηκε σε τρεις ομάδες ανάλογα με το αν είναι μικρή, μεσαία ή μεγάλη η τιμή του.

Για την καλύτερη αξιολόγηση των χαρακτηριστικών του μοντέλου, υπολογίστηκε ο συντελεστής αυτοσυσχέτισης τους. Ακολουθεί ο πίνακας που απεικονίζει το συντελεστή αυτοσυσχέτισης ανάμεσα στα χαρακτηριστικά του μοντέλου:



Σχήμα 3.5: Πίνακας Συντελεστών Αυτοσυσχέτισης

Σε αυτό το σημείο, για την περαιτέρω εξέλιξη του μοντέλου και τον εντοπισμό αστοχιών ταξινόμησης, υπολογίστηκε και η τιμή του AUC_score.

Με τις παραπάνω αλλαγές φάνηκαν τα αποτελέσματα του μοντέλου εκτός του δείγματος εκπαίδευσης να διαμορφώνονται ως εξής:

$$R^2 \parallel 0.8972 \quad AUC_score \parallel 0.8723$$

Πίνακας 3.3: Αποτελέσματα Μοντέλου Λιγότερων Χαρακτηριστικών

Class	Precision	Recall	F1-score
0.0	0.93	0.93	0.93
0.1	0.80	0.82	0.81

Τα ποσοστά που παρατηρούνται οδήγησαν στην ανάγκη για μελέτη μεμονωμένων παραδειγμάτων για αναγνώριση των αιτιών που οδηγούν σε λάθος πρόβλεψη. Παρατηρήθηκε, έτσι, ότι πολλές φορές η λανθασμένη πρόβλεψη οφειλόταν στην έλλειψη επαρκούς πληροφορίας για τις παραγγελίες. Δηλαδή, εντοπίστηκαν περιπτώσεις που τα χαρακτηριστικά των δύο παραγγελιών ταυτίζονταν και κατ' επέκταση το μοντέλο τα ταξινομούσε στην ίδια κατηγορία ενώ έπρεπε να κατηγοριοποιηθούν σε διαφορετικές κλάσεις.

Έτσι, δόθηκε έμφαση στην προσθήκη των απαραίτητων πληροφοριών που φάνηκαν να απουσιάζουν καθώς και στην μορφοποίηση των χαρακτηριστικών ώστε οι τιμές να διευκολύνουν το μοντέλο να αναγνωρίσει μοτίβα. Η επίδραση των διαφόρων χαρακτηριστικών που δοκιμάζονταν αξιολογήθηκε με βάση το διάνυσμα βαρών. Ακόμα, για τον εντοπισμό των πιο σημαντικών χρησιμοποιήθηκαν εργαλεία επιλογής χαρακτηριστικών (όπως επιλογέας RFE) της βιβλιοθήκης scikit-learn.

3.2.3.3 Το τελικό μοντέλο

Για την τελειοποίηση του μοντέλου, χρειάστηκε πληθώρα δοκιμών και παρατηρήσεων τόσο στα δεδομένα όσο και στα χαρακτηριστικά που είχαν δοθεί. Ακολουθήθηκε η εξής διαδικασία επαναληπτικά:

- * Τροποποιήσεις στα χαρακτηριστικά
- * Παρατήρηση μετρικών για πρώτη αξιολόγηση αλλαγών
- * Manual inspection των αποτελεσμάτων για βαθύτερη κατανόηση επίδρασης
- * Οπτικοποίηση σφαλμάτων για εντοπισμό μοτίβων
- * Παρατήρηση διανύσματος βαρών για αναγνώριση βαθμού επίδρασης των χαρακτηριστικών
- * Σύγκριση με προηγούμενα αποτελέσματα
- * Εκ νέου διορθώσεις στα χαρακτηριστικά

Κατά τις διαφορετικές δοκιμές χρησιμοποιήθηκαν πολλά νέα χαρακτηριστικά, διαφορετικές αναπαραστάσεις της ίδιας πληροφορίας, χωρισμός των δεδομένων σε ομάδες, δυαδική αναπαράσταση κάποιων πληροφοριών, στρογγυλοποιήσεις μεγεθών ανάλογα με την ακρίβεια που χρειάζεται, εισαγωγή και προσαρμογή δυαδικών χαρακτηριστικών για μεγέθη που λειτουργούν ως κτώφλι για τη συμπεριφορά του μοντέλου. Ακόμα, χρησιμοποιήθηκε η τεχνική με την αντίστροφη της συνάρτησης λογιστικής παλινδρόμησης που παρουσιάστηκε παραπάνω, για την απεικόνιση των χαρακτηριστικών που εκφράζουν την πιθανότητα η κάθε παραγγελία να καταταχθεί στη μία ή την άλλη κατηγορία με βάση κάποιο στοιχείο. Σε αυτό το σημείο, χρησιμοποιήθηκε μία ακόμα τεχνική για την παραγωγή συνδυαστικών χαρακτηριστικών η οποία και αναλύεται στην επόμενη παράγραφο.

Η ανάλυση των δεδομένων έδειξε πως κάποιες πληροφορίες δεν ήταν ανεξάρτητες μεταξύ τους, ενώ το μοντέλο υποθέτει ότι τα δεδομένα του διανύσματος εισόδου είναι ανεξάρτητες μεταβλητές. Σε τέτοιες περιπτώσεις έγινε χρήση της εξής τεχνικής: τα δύο δυαδικά χαρακτηριστικά που αλληλεξαρτώνται αντιστοιχίζονται σε ένα συνδυαστικό. Αυτό μπορεί να πάρει τέσσερις τιμές (μία τιμή για κάθε έναν από τους διαφορετικούς συνδυασμούς που μπορούν να προκύψουν από τα αρχικά). Το νέο χαρακτηριστικό απεικονίζεται σε τέσσερις στήλες όπου αληθή τιμή μπορεί να πάρει μόνο μία από αυτές κάθε φορά, που δείχνει ποιο συνδυασμό έχουν πάρει οι δύο αρχικές μεταβλητές.

Με τις παραπάνω μεθόδους εντοπίστηκαν σημαντικές μεταβολές στη συμπεριφορά της τεχνικής μηχανικής μάθησης που σχεδιάστηκε. Παρατηρήθηκε, δηλαδή, ότι κάποιες προσθήκες ή

τροποποιήσεις βελτίωναν συνολικά το μοντέλο (οι οποίες και διατηρήθηκαν), ενώ κάποιες άλλες διόρθωναν μια μετρική αλλά χειροτέρευαν μια άλλη. Έτσι, από κάποιο σημείο και μετά έγινε μια προσπάθεια για διατήρηση του πιο ισορροπημένου μοντέλου. Με τον τελευταίο όρο γίνεται αναφορά στην διατήρηση καλών τιμών συνολικά σε όλες τις τιμές αξιολόγησης καθώς και στην συνολική εικόνα των αποτελεσμάτων όπως προκύπτει από την παρατήρηση της μορφής των σφαλμάτων.

Επιπροσθέτως έγινε η εξής διαπίστωση αναφορικά με διάφορα σφάλματα που καταγράφηκαν και παρουσιάζονται στην ενότητα των αποτελεσμάτων. Φάνηκε ότι από κάποιο σημείο και μετά δεν ήταν δυνατή η περαιτέρω διόρθωση των συνολικών σφαλμάτων με την αλλαγή των χαρακτηριστικών. Νέες περιγραφές μπορεί να ελάττωναν, για παράδειγμα, τα σφάλματα στην εκτίμηση των τιμών της κλάσης 0 αλλά συνεπάγονταν σε αύξηση των λανθασμένων εκτιμήσεων της άλλης κλάσης. Το μόνο που φάνηκε να διαφοροποιείται, δηλαδή, ήταν η αναλογία μεταξύ των σφαλμάτων των προβλέψεων για τις δύο κλάσεις αλλά όχι ο συνολικός αριθμός τους. Από αυτό το στάδιο και μετά, θεωρήθηκε ότι οι λάθος εκτιμήσεις οφείλονταν κυρίως σε τυχαιότητα στα δεδομένα εισόδου, γεγονός που επιβεβαιώθηκε και με την μελέτη της συστηματικότητας των σφαλμάτων.

Πιο αναλυτικά, έγιναν οι εξής δοκιμές στα χαρακτηριστικά:

- Για την απεικόνιση του μεγέθους της παραγγελίας δοκιμάστηκαν διαφορετικές περιγραφές. Τέτοιες ήταν: αυτούσια η τιμή του όγκου και των κιλών, στρογγυλοποιήσεις αυτών, ομαδοποιήσεις, δυαδική αναπαράσταση και συνδυασμοί αυτών.
- Για τα διάφορα χαρακτηριστικά που περιγράφουν το σημείο παράδοσης (όπως το πλήθος οχημάτων που μπορούν να το εξυπηρετήσουν, το πλήθος δρομολογίων που το περιέχουν κλπ.) δοκιμάστηκαν αντίστοιχες τροποποιήσεις με παραπάνω.
- Για την περιγραφή της απόστασης μεταξύ των σημείων παραλαβής και παράδοσης χρησιμοποιήθηκε τόσο αυτούσια η απόσταση όσο και η εκθετική τιμή της, στρογγυλοποιήσεις αυτών σε διάφορα δεκαδικά, ομαδοποιήσεις και δυαδικές αναπαραστάσεις.
- Δοκιμάστηκε το ποσοστό των παραγγελιών που σχημάτισαν δρομολόγιο δύο στάσεων ως προς τα συνολικά δρομολόγια ανάλογα με χαρακτηριστικά όπως η μέρα της εβδομάδας, το σημείο παράδοσης, το σημείο παραλαβής και το πλήθος των διαφορετικών ειδών προϊόντος από τα οποία συντίθεται η παραγγελία. Έπειτα, υπολογίστηκε η τιμή της αντίστροφης της συνάρτησης λογιστικής για κάθε ένα από αυτά τα ποσοστά.
- Για την καλύτερη περιγραφή της ημέρας της εβδομάδας και του σημείου παραλαβής προϊόντος, τα αριθμητικά χαρακτηριστικά που χρησιμοποιούνταν αντικαταστάθηκαν από δυαδικά (π.χ. Η μέρα της παραγγελίας είναι δευτέρα;)
- Παρατηρήθηκε εξάρτηση ανάμεσα στην πληροφορία αν η παραγγελία παραδίδεται σε αποθήκη και αν η παραγγελία έχει συντεθεί από πολλές άλλες. Έτσι τα δύο αυτά δυαδικά χαρακτηριστικά αντικαταστάθηκαν από τέσσερα δυαδικά που αντιστοιχούν στο καρτεσιανό παράγωγο τους, σύμφωνα με την τεχνική που αναλύθηκε παραπάνω.
- Για την περιγραφή του τρόπου σύνθεσης της κάθε παραγγελίας, προστέθηκαν πληροφορίες που περιγράφουν πλήθος των διαφορετικών ειδών προϊόντος που περιέχονται σε αυτήν και

δυναδικά χαρακτηριστικά για την απεικόνιση αυτής της πληροφορίας (π.χ. αν αποτελείται μόνο από ένα είδος, από λίγα ή από πολλά). Ακόμα χρησιμοποιήθηκαν πληροφορίες για το βαθμό συσχέτισης των ειδών προϊόντος που συνθέτουν την παραγγελία καθώς και το πλήθος των διαφορετικών παραγγελιών που συνδυάστηκαν σε αυτήν.

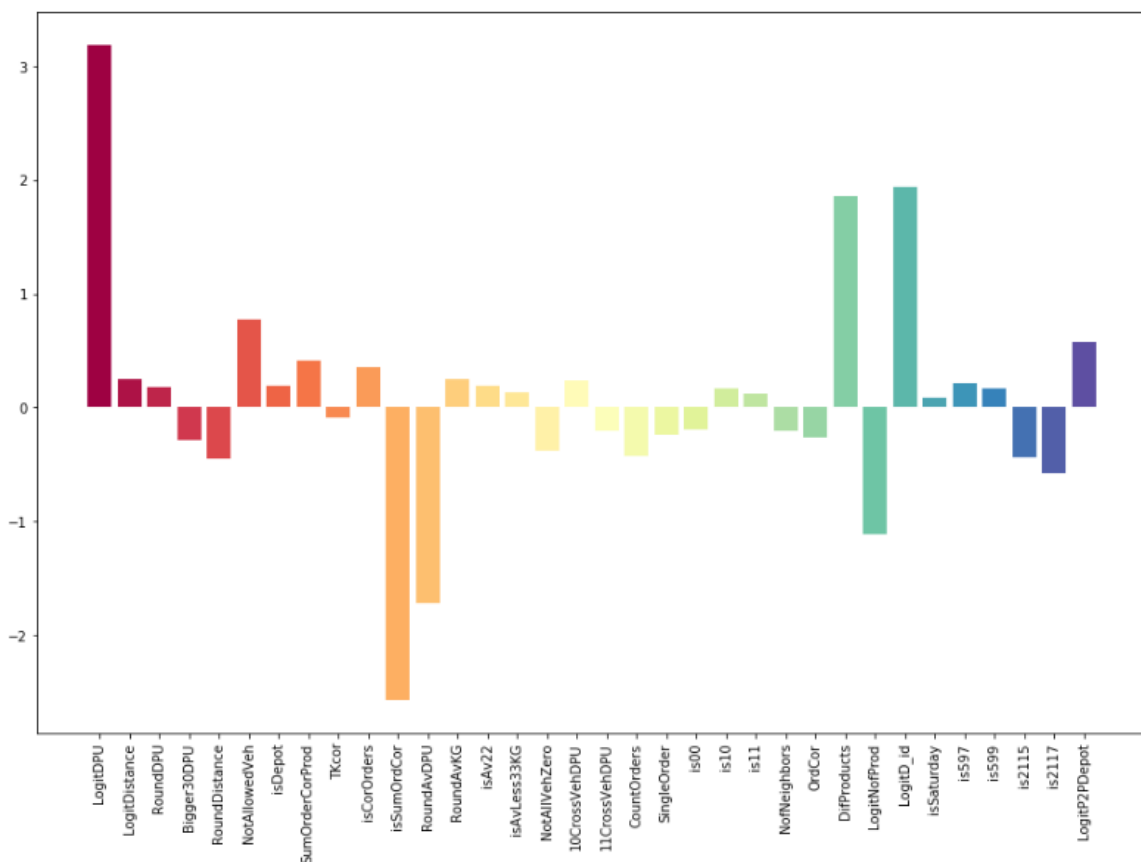
Μετά, λοιπόν, από επαναλαμβανόμενες δοκιμές χαρακτηριστικών, μελέτη μεμονωμένων παραδειγμάτων για εντοπισμό συστηματικών σφαλμάτων και αλλαγών στις παραμέτρους του μοντέλου και στον αλγόριθμο που χρησιμοποιείται για τον εντοπισμό των βέλτιστων βαρών της ευθείας διαχωρισμού, καταλήξαμε στο βέλτιστο δυνατό μοντέλο. Σε αυτό, για κάθε μία από τις (σύνθετες) παραγγελίες δίνονται τα χαρακτηριστικά που φαίνονται παρακάτω:

1. Αντίστροφη συνάρτησης λογιστικής παλινδρόμησης (logit) για ποσοστό P2P παραγγελιών επί των συνολικών παραγγελιών για την συγκεκριμένη (στρογγυλοποιημένη) ποσότητα TU (*logitDPU*)
2. Αντίστροφη συνάρτησης λογιστικής παλινδρόμησης (logit) για ποσοστό P2P παραγγελιών επί των συνολικών παραγγελιών για την συγκεκριμένη (στρογγυλοποιημένη) απόσταση μεταξύ σημείου παραλαβής και σημείου παράδοσης (*logitP2Pdistance*)
3. Στρογγυλοποιημένη (χωρίς τα δεκαδικά) ποσότητα παραγγελίας σε TU (*round amount dpu*)
4. Δυναδικό χαρακτηριστικό που εκφράζει το εάν η παραγγελία είναι μεγαλύτερη από 30 TU ή όχι (τιμή από την οποία και μετά παρατηρήθηκε ότι η πλειοψηφία των παραγγελιών δρομολογούταν ως P2P) (*biggerThan30DPU*)
5. Στρογγυλοποιημένη απόσταση μεταξύ σημείου παράδοσης-σημείου παραλαβής (*roundDistance*)
6. Πλήθος οχημάτων που δεν μπορούν να παραδώσουν στο σημείο παράδοσης (*D_Ad_Vehicles*)
7. Είδος σημείου παράδοσης (αποθήκη ή κατάστημα) – Bool χαρακτηριστικό (*isDepot*)
8. Τιμή που εκφράζει το πόσο πιθανό είναι η παραγγελία μας να συνδυαστεί με άλλες με βάση τα προϊόντα που αποτελείται (η τιμή αυτού του δείκτη ισούται με το άθροισμα των προϊόντων των παραγγελιών με τις οποίες έχουν συνδυαστεί τα είδη που συνθέτουν την παραγγελία που μελετάμε) (*sumaOrderCorrelationProd*)
9. Τιμή που εκφράζει το πόσο πιθανό είναι η παραγγελία που μελετάμε να συνδυαστεί με άλλες με βάση τον ταχυδρομικό της κώδικα (η τιμή αυτού του δείκτη ισούται με το άθροισμα των TK των παραγγελιών με τις οποίες έχει συνδυαστεί ο TK της παραγγελίας μας) (*Tkcor*)
10. Δυναδική μεταβλητή που δηλώνει αν είναι πιθανό η παραγγελία μας να συνδυαστεί με άλλες με βάση τα προϊόντα που τη συνθέτουν (*CountOrdersGrouped*)
11. Δυναδική μεταβλητή που δηλώνει το αν το *sumaOrderCorrelationProd* έχει μεγάλη ή μικρή τιμή. (*sumaOrdCorGroup*)
12. Στρογγυλοποιημένη μέση ποσότητα παραγγελιών σε TU του σημείου παράδοσης (*roundAvDPU*)

13. Στρογγυλοποιημένη μέση ποσότητα παραγγελιών σε KG του σημείου παράδοσης (*roundAvKG*)
14. Δυναμική μεταβλητή που εκφράζει αν η μέση παραγγελία στο σημείο παράδοσης είναι μεγαλύτερη από 22 TU – τιμή που φάνηκε ως *threshold* (*Avis22*)
15. Δυναμική μεταβλητή που εκφράζει αν η μέση παραγγελία στο σημείο παράδοσης είναι μικρότερη από 33 Kg – τιμή που φάνηκε ως *threshold* (*AvislessThan33_KGTrreshold*)
16. Δυναμική μεταβλητή που δηλώνει αν όλα τα οχήματα μπορούν να προσεγγίσουν το σημείο παράδοσης ή όχι (*AdVehZero*)
17. Cross Feature του αν μπορούν όλα τα οχήματα να προσεγγίσουν το σημείο παράδοσης και αν η μέση χωρητικότητα του σημείου παράδοσης είναι μεγαλύτερη από 9 TU (για να διακρίνει τα δρομολόγια που το σημείο παράδοσης είναι κατάστημα και μπορούν να παραδώσουν όλα τα οχήματα ή είναι αποθήκη) – δεν κράτησα και τα 4 διαφορετικά χαρακτηριστικά καθώς μόνο δύο φάνηκαν να είναι καθοριστικά για τη συμπεριφορά του μοντέλου (*10CrossVehDPU*, *11CrossVehDPU*)
18. Πλήθος παραγγελιών από τις οποίες αποτελείται η σύνθετη παραγγελία (*CountOrders*)
19. Δυναμική μεταβλητή που εκφράζει αν η σύνθετη παραγγελία αποτελείται από 1 ή περισσότερες απλές παραγγελίες (*SingleOrder*)
20. Cross Feature του αν το σημείο παράδοσης είναι αποθήκη και αν η σύνθετη παραγγελία αποτελείται από πολλές απλές – αφάιρεσα έναν από τους 4 δυνατούς γιατί καμία παρατήρηση που να συνδυάζει αυτές τις τιμές (*is00*, *is10*, *is11*)
21. Πλήθος σημείων με τα οποία έχει συνδυαστεί το σημείο παράδοσης (*NofNeighbors*)
22. Αυτό το χαρακτηριστικό περιγράφει πόσο πιθανό είναι η παραγγελία που εξετάζεται να συνδυαστεί με άλλες με βάση τα διαφορετικά προϊόντα που τη συνθέτουν (η τιμή αυτού του δείκτη ισούται με το άθροισμα των παραγγελιών με τις οποίες έχουν συνδυαστεί τα είδη της) (*OrdCor*)
23. Πλήθος διαφορετικών προϊόντων από τα οποία αποτελείται η παραγγελία – γραμμικό χαρακτηριστικό (*DifProducts*)
24. Αντίστροφη συνάρτησης λογιστικής παλινδρόμησης (*logit*) για ποσοστό P2P παραγγελιών επί των συνολικών παραγγελιών για το συγκεκριμένο πλήθος διαφορετικών προϊόντων που συνθέτουν αυτή την παραγγελία (*logitNofProd*)
25. Αντίστροφη συνάρτησης λογιστικής παλινδρόμησης (*logit*) για ποσοστό P2P παραγγελιών επί των συνολικών παραγγελιών για το συγκεκριμένο σημείο παράδοσης (*logitD_id*)
26. Δυναμικό χαρακτηριστικό που εκφράζει αν είναι Σάββατο ή όχι (η μόνη μέρα της εβδομάδας που παρατηρήθηκε να είναι καθοριστική για το αν μια παραγγελία θα είναι P2P ή όχι) (*isSaturday*)

27. Δυαδικό χαρακτηριστικά που αντιπροσωπεύουν τα id των τα id των 4 από τις 5 αποθήκες (η πέμπτη δεν φάνηκε να είναι καθοριστική για τη συμπεριφορά του μοντέλου) (*is597, is599, is2115, is2117*)
28. Αντίστροφη συνάρτησης λογιστικής παλινδρόμησης (logit) για ποσοστό P2P παραγγελιών επί των συνολικών παραγγελιών για το συγκεκριμένο σημείο παραλαβής (*logitP2PDepot*)

Όπως έχει περιγραφεί και παραπάνω, ο αλγόριθμος εντοπίζει την κατάλληλη βαρύτητα για το κάθε χαρακτηριστικό του διανύσματος εισόδου. Ανάλογα με το μέγεθος του κάθε χαρακτηριστικού φαίνεται και ο βαθμός επιρροής του. Στο ακόλουθο διάγραμμα φαίνεται η σημαντικότητα των χαρακτηριστικών όπως πρόκυψε από το βέλτιστο διάνυσμα βαρών:



Σχήμα 3.6: Βέλτιστα Βάρη Χαρακτηριστικών

Με βάση την παραπάνω απεικόνιση μπορούμε να διαπιστώσουμε ότι τα δύο χαρακτηριστικά με τη μεγαλύτερη επιρροή είναι η πιθανότητα της παραγγελίας να ανήκει στην κλάση των ανεξάρτητων με βάση τη χωρητικότητα της και το αν η παραγγελία είναι πιθανό να συνδυαστεί με άλλες με βάση το αν τα είδη που τη συνθέτουν φαίνονται να συνδυάζονται με άλλα ή όχι. Ακόμα, σημαντική επίδραση φαίνονται να έχουν και χαρακτηριστικά όπως η στρογγυλοποιημένη τιμή του μέσου μεγέθους παραγγελίας του σημείου παραλαβής, το πλήθος των διαφορετικών ειδών από τα οποία συντίθεται η παραγγελία και οι πιθανότητες κατηγοριοποίησης στην κλάση 1 με κριτήριο το πλήθος των διαφορετικών προϊόντων και το κατάστημα παράδοσης. Η προσαρμογή του μοντέλου στις διάφορες πληροφορίες που του δίνονται είναι αναμενόμενη καθώς και από την εμπειρία

μας από την καθημερινή μας ζωή φαίνεται λογικό τα βασικά στοιχεία μιας παραγγελίας να είναι ο όγκος της, το σημείο παράδοσης και τα διαφορετικά είδη που περιλαμβάνει.

3.3 Παραμετροποίηση Περιβάλλοντος Δοκιμών

3.3.1 Δεδομένα Εισόδου

Για την πλήρη αξιολόγηση του μοντέλου χρησιμοποιήθηκε ένα δοκιμαστικό περιβάλλον στο οποίο δόθηκαν ως είσοδοι οι διευθύνσεις, τα φορτηγά, οι παραγγελίες, δυνατοί συνδυασμοί οχημάτων - οδηγών και οι βάρδιες τους.

Στη συνέχεια, χρησιμοποιήθηκαν οι προβλέψεις του μοντέλου για την εύρεση των παραγγελιών που συνθέτουν ένα δρομολόγιο. Τα δρομολόγια αυτά δόθηκαν ως είσοδοι στον αλγόριθμο βελτιστοποίησης δρομολόγησης, συνθέτοντας έτσι την αρχική λύση του προβλήματος. Αναλυτικότερα, τα δεδομένα που δόθηκαν ήταν τα ακόλουθα:

1. Addresses: Πληροφορία για διευθύνσεις αποθηκών και καταστημάτων (εμπεριέχουν και την πληροφορία για είδη οχημάτων που δεν μπορούν να προσεγγίσουν μια διεύθυνση)
2. Resources: Πληροφορία για κάθε είδος οχήματος (resourceKind) για κάθε πινακίδα (resource). Μπορεί δύο οχήματα ίδιου είδους να έχουν διαφορετική χωρητικότητα σε TU. Για να φτιάξω resources μελέτησα στα δεδομένα:
 - αν κάθε όχημα ξεκινάει από συγκεκριμένη αποθήκη (δεν φάνηκε να υπάρχει κάποιο standard pattern)
 - αν ένα όχημα είναι ψυχόμενο (φάνηκε συγκεκριμένα οχήματα να μεταφέρουν μόνο Dry, μόνο Frozen ή και τα δύο). Έτσι έκανα τη θεώρηση ότι υπάρχουν οχήματα με ή χωρίς ψυγείο. Στα οχήματα με ψυγείο μπορούν να τοποθετηθούν όλα τα είδη παραγγελιών, ενώ σε αυτά δεν μπορούν να τοποθετηθούν παραγγελίες με capability Fresh (δηλαδή παραγγελίες με τύπο προϊόντος Fresh, Frozen, Ambient)
3. Orders: Πληροφορία για κάθε «σύνθετη» παραγγελία (ημερομηνία παράδοσης, είδος παραγγελίας, είδος προϊόντος, ποσότητα σε KG και TU, σημείο παραλαβής – αποθήκη, σημείο παράδοσης, χρονικό παράθυρο παραλαβής και χρονικό παράθυρο παράδοσης)
 - Χρησιμοποιώ σαν χρονικό παράθυρο παραλαβής 24 ώρες (με χρονική στιγμή τέλους την χρονική στιγμή λήξης του χρονικού παραθύρου παράδοσης) και σαν χρονικό παράθυρο παράδοσης το υπάρχον
4. Shift Templates: Πρότυπο δρομολογίου. Για κάθε όχημα δημιουργώ ένα shift template που το περιγράφει και σε αυτό μπορώ να βάλω ημερολογιακές διαθεσιμότητες

3.3.2 Αλγόριθμος Βελτιστοποίησης

Μετά την καταχώρηση στη βάση δεδομένων του συστήματος των απαραίτητων δεδομένων για την δρομολόγηση, στόχος είναι η διεξαγωγή πειραματικών δοκιμών. Ο αλγόριθμος που χρησιμοποιείται στο περιβάλλον για την βελτιστοποίηση των δρομολογίων διέρχεται από τρία στάδια.

Στο κάθε στάδιο δοκιμάζονται διαφορετικές μέθοδοι και στρατηγικές και στη συνέχεια ο αλγόριθμος μεταβαίνει στο επόμενο στάδιο. Ασφαλώς, η συγκεκριμένη δομή αποτελεί μια απλουστευμένη υλοποίηση στα πλαίσια της διπλωματικής.

1. Καρασκευή λύσης (Εναλλακτική επιλογή μεθόδων)

- sweep
- παράλληλη εισαγωγή
- ακολουθιακή εισαγωγή
- πολλαπλά σημεία εκκίνησης

2. Τοπική αναζήτηση

- ανταλλαγές ακμών
- ανταλλαγές κόμβων
- LNS
- ejection chains
- VLNS και άλλες

3. Ruin and Recreate

- τυχαία αφαίρεση
- αφαίρεση ομάδας
- αφαίρεση χειρότερου

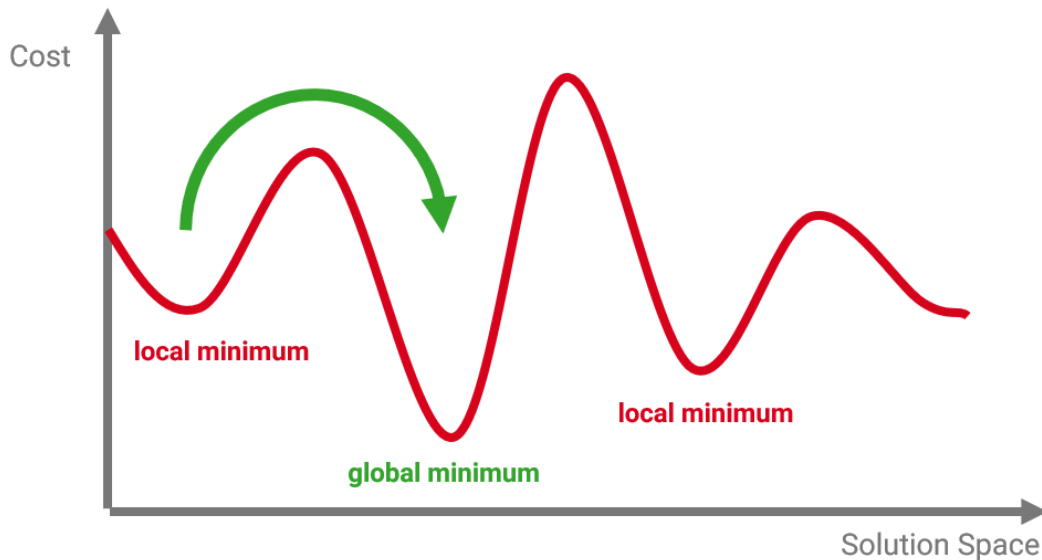
Η πρώτη φάση αποτελεί την αρχικοποίηση της λύσης. Εδώ ο αλγόριθμος προσπαθεί να συνδέσει τα σημεία προκειμένου να κατασκευάσει μία πρώτη λύση για το πρόβλημα. Ξεκινώντας από μία κοινή αφετηρία δοκιμάζει είτε την εισαγωγή ανεξάρτητων διαδρομών, είτε την παράλληλη εισαγωγή κόμβων, είτε την ακολουθιακή προσθήκη κόμβων σε ένα δρομολόγιο και μετά την ολοκλήρωση του αρχίζει να εισάγει κόμβους στο επόμενο ή δοκιμάζει ακόμα την εισαγωγή δρομολογίων με διαφορετικές αφετηρίες.

Αφού ολοκληρωθεί το στάδιο της κατασκευής, μπορούν να εφαρμοστούν διάφορες μέθοδοι επαναληπτικής βελτιστοποίησης για να βελτιώσουν την αρχική λύση. Αυτές οι μέθοδοι μπορούν να εκτελεστούν με όποια σειρά καθορίσει αυτός που παραμετροποιεί τον αλγόριθμο, ανάλογα με το επιθυμητό αποτέλεσμα. Δεδομένου ότι μετά τις αλλαγές μίας μεθόδου το τοπίο λύσης μεταβάλλεται, μπορούν οι διάφορες μέθοδοι να εκτελεστούν επαναληπτικά για εντοπισμό όλο και καλύτερων λύσεων. Αυτές τερματίζονται μόνο όταν δεν μπορεί να γίνει περαιτέρω βελτίωση της λύσης.

Λίγο πιο αναλυτικά, στη φάση της τοπικής αναζήτησης ο αλγόριθμος δοκιμάζει διαφορετικές μεταβολές στα δρομολόγια (όπως του υπαγορεύεται από την επιλεγμένη στρατηγική) με στόχο την ελάττωση του κόστους δρομολόγησης, των πόρων που αξιοποιούνται, του χρόνου και άλλα.

Το παραπάνω εμπεριέχει τον κίνδυνο του περιορισμού σε ένα τοπικό ελάχιστο του προβλήματος. Όπως μπορεί να φανεί και στην παρακάτω εικόνα, η συνάρτηση του κόστους που περιγράφει το συγκεκριμένο στιγμιότυπο του προβλήματος μπορεί να περιέχει παραπάνω από ένα ελάχιστο,

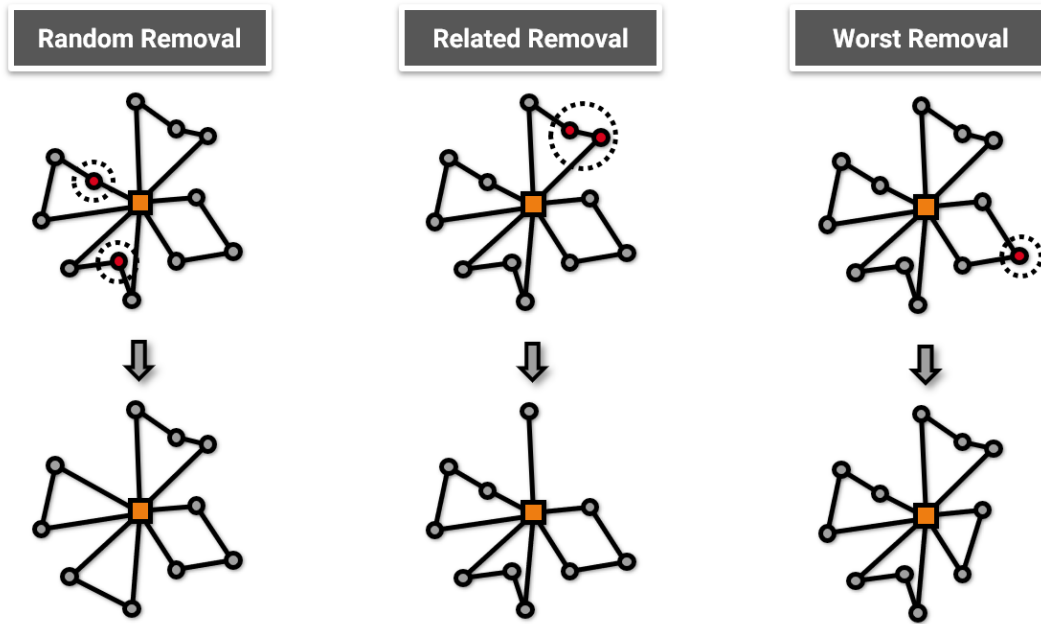
και δεν σημαίνει ότι πάντα το πρώτο που εντοπίζεται είναι και το ολικό. Η στρατηγική της τοπικής αναζήτησης, μπορεί να βρεθεί στην κατεύθυνση λύσης τοπικού ελαχίστου και να αδυνατεί να ξεφύγει από αυτή με αποτέλεσμα έπειτα από ορισμένων μεταβολών να τερματίζονται οι δυνατές επιλογές για βελτίωση. Ωστόσο, η μέθοδος της τοπικής αναζήτησης μπορεί σε λίγο χρόνο να εντοπίσει καλές περιοχές του χώρου λύσης. Πιο αναλυτικές πληροφορίες για την πολυπλοκότητα τις μεθόδου αλλά και τα σημεία που υστερεί μπορούν να αναζητηθούν στο βιβλίο των [E.Aarts \(2003\)](#).



Σχήμα 3.7: Συνάρτηση Κόστους - Χώρου Λύσεων

Προκειμένου να αποφευχθεί ο εγκλωβισμός σε τοπικό ελάχιστο, υπάρχει η τρίτη φάση του αλγορίθμου. Σε αυτό το στάδιο, δοκιμάζεται η καταστροφή ενός μέρους της λύσης ώστε να μεταβληθεί ο χώρος λύσεων του προβλήματος. Εδώ, μπορεί να δοκιμαστεί η τυχαία αφαίρεση ενός κόμβου από ένα δρομολόγιο, η αφαίρεση δύο γειτονικών ή και η αφαίρεση του πιο απομακρυσμένου. Ο αλγόριθμος Ruin and Recreate σε φυσική γλώσσα καθώς και ο μαθηματικός φορμαλισμός του περιγράφεται αναλυτική στο έργο του [Misevicius](#)

(2003). Στην παρακάτω εικόνα μπορεί να γίνει πιο κατανοητή η παραπάνω περιγραφή.



Σχήμα 3.8: Μέθοδοι της Στρατηγικής Ruin and Recreate

3.3.3 Διαφορετικές Δοκιμές στο Δοκιμαστικό Περιβάλλον

Μετά την καταχώρηση των απαραίτητων στοιχείων στο δοκιμαστικό περιβάλλον καθίσταται δυνατή η διεξαγωγή δοκιμών για τη σύγκριση αποτελεσμάτων. Για την απόκτηση μιας ολιστικής εικόνας των δυνατοτήτων του μοντέλου λογιστικής παλινδρόμησης δοκιμάστηκαν δύο διαφορετικές προσεγγίσεις και συγκρίθηκαν τόσο με τα πραγματικά δεδομένα δρομολόγησης όσο και με τα αποτελέσματα του αλγορίθμου βελτιστοποίησης για το ίδιο στιγμιότυπο του προβλήματος.

Η πρώτη δοκιμή έχει ως στόχο την χρήση των προβλέψεων του μοντέλου για κατασκευή της πρώτης λύσης του αλγορίθμου βελτιστοποίησης. Οι παραγγελίες που το μοντέλο έκρινε ικανές να συνθέσουν ανεξάρτητο δρομολόγιο τοποθετήθηκαν στα οχήματα με την ελάχιστη δυνατή χωρητικότητα για τη μεταφορά τους. Οι συνδυασμοί αυτοί οχήματος - παραγγελίας καταχωρήθηκαν στα δρομολόγια στο περιβάλλον και συνέθεσαν ανεξάρτητα δρομολόγια. Παράλληλα, δημιουργήθηκαν τα υπόλοιπα διαθέσιμα δρομολόγια της συγκεκριμένης ημέρας για την δρομολόγηση των υπόλοιπων παραγγελιών. Έπειτα, ο αλγόριθμος προσέθεσε τις υπόλοιπες παραγγελίες σε όλα τα διαθέσιμα δρομολόγια και δοκίμασε τις διαφορετικές μεθόδους των στρατηγικών τοπικής αναζήτησης και Ruin and Recreate που αναλύθηκαν παραπάνω. Επισημαίνεται, εδώ, ότι ο αλγόριθμος έχει τη δυνατότητα να αφαιρέσει μία παραγγελία από ένα δρομολόγιο της προτεινόμενης λύσης, να προσθέσει νέες με αποτέλεσμα να διαφοροποιηθεί το πλήθος των στάσεων του. Αφού ολοκληρώθηκε η εκτέλεση του αλγορίθμου βελτιστοποίησης, καταγράφηκε η έξοδος του.

Ακολούθησε η δεύτερη δοκιμή, όπου μετράει την αποτελεσματικότητα του συνδυασμού των λύσεων του μοντέλου μηχανικής μάθησης και του αλγορίθμου βελτιστοποίησης. Σε αυτό το πείραμα, συλλέχθηκε η έξοδος της συνάρτησης λογιστικής παλινδρόμησης και παράχθηκαν τα δρομολόγια δύο στάσεων όπως στο πρώτο βήμα της προηγούμενης μεθόδου. Στη συνέχεια, με τους συνδυασμούς οδηγών και οχημάτων που δεν είχαν χρησιμοποιηθεί, συντέθηκαν τα δρομολόγια που μπορούν να πραγματοποιηθούν. Ακόμα, καταχωρήθηκαν οι υπόλοιπες παραγγελίες στη βάση

δεδομένων του περιβάλλοντος. Εφαρμόστηκε ο αλγόριθμος για τη δρομολόγηση μόνο των παραγωγικών που δεν είχαν κριθεί από το μοντέλο ως ανεξάρτητες στα οχήματα που δεν είχαν απασχοληθεί. Μετά τον τερματισμό της διαδικασίας αθροίστηκαν οι αποστάσεις, οι χρόνοι εκτέλεσης, οι δρομολογημένες παραγγελίες και οι πόροι που αξιοποιήθηκαν από τις δύο μεθόδους και χρησιμοποιήθηκαν για την αξιολόγηση της δοκιμής αυτής.

Στη συνέχεια, καταχωρήθηκαν πάλι όλες οι παραγγελίες και τα διαθέσιμα δρομολόγια για το ίδιο στιγμιότυπο του προβλήματος. Εδώ, χωρίς καμία αρχικοποίηση εκτελέστηκε ο αλγόριθμος και καταγράφηκε η λύση και οι τιμές των αντίστοιχων μετρικών.

Τέλος, για την πλήρη αξιολόγηση των παραπάνω υπολογίστηκαν οι αντίστοιχες τιμές των πραγματικών δεδομένων δρομολόγησης από το παραγωγικό περιβάλλον για την υπό εξέταση ημέρα του προβλήματος. Στο επόμενο κεφάλαιο παραθέτονται τα αποτελέσματα των τεσσάρων εφαρμογών που δοκιμάστηκαν και γίνεται η σύγκριση τους.

Κεφάλαιο 4

Αξιολόγηση αποτελεσμάτων και συγκρίσεις

Αυτή η ενότητα επικεντρώνεται στην καταγραφή των τελικών αποτελεσμάτων των προσεγγίσεων που αναλύθηκαν παραπάνω. Ακολουθεί η αξιολόγηση αυτών των τιμών και η ποιοτική τους ερμηνεία ενώ στη συνέχεια παρουσιάζονται οι συγκρίσεις μεταξύ αυτών ενώ τελικά αποτυπώνονται τα καταλληλότερα πεδία εφαρμογής της κάθε δοκιμής.

4.1 Αποτελέσματα μοντέλου μηχανικής μάθησης

4.1.1 Αξιολόγηση Αποτελεσμάτων Μετρικών

Σε αυτή την υποενότητα παρουσιάζονται οι τιμές των μετρικών της ακρίβειας, της ανάκλισης, του R2 και του AUC score. Στη συνέχεια δίνεται η ποιοτική ερμηνεία των τιμών τους καθώς και η αξιολόγηση του μοντέλου με βάση αυτά.

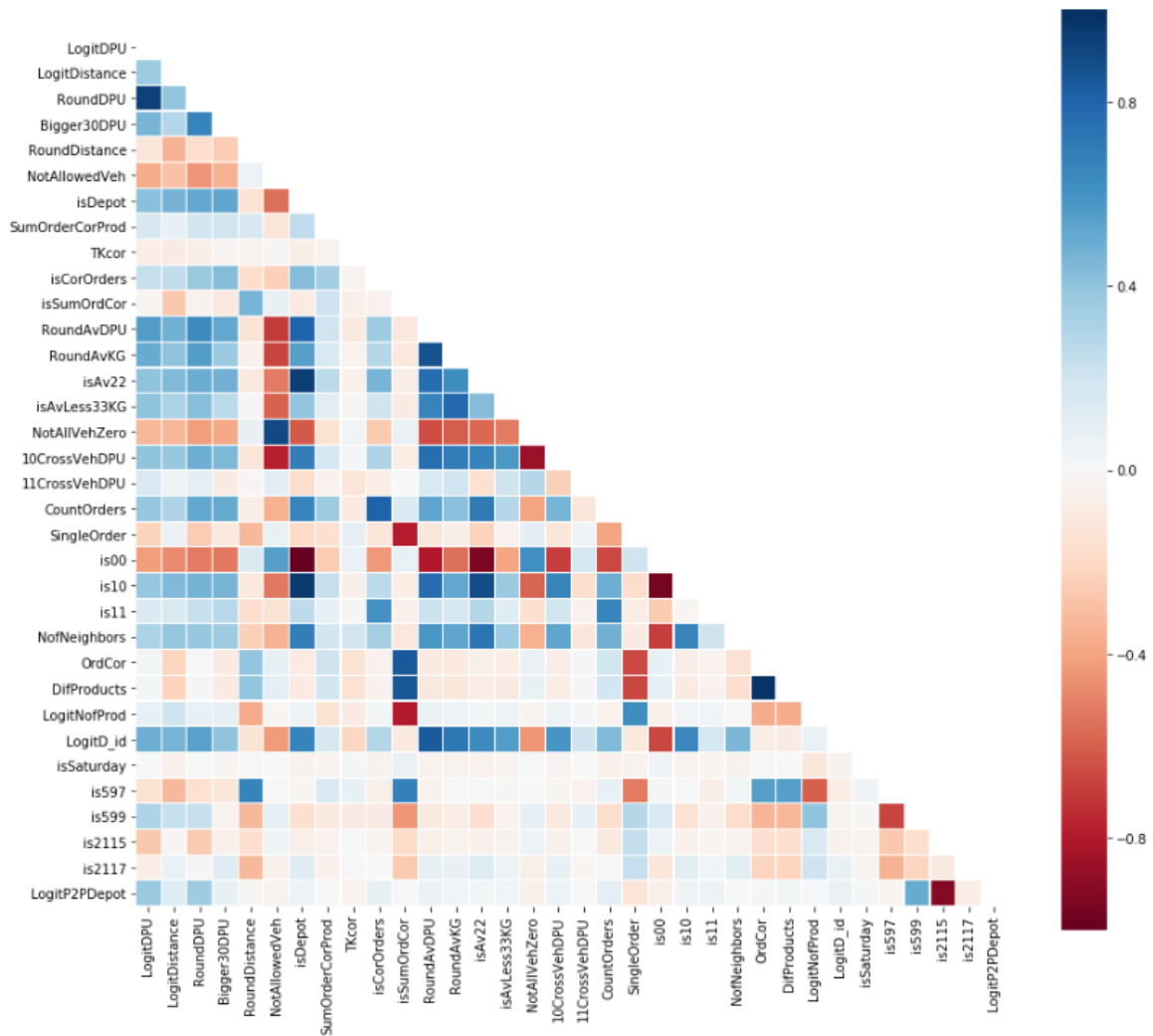
4.1.1.1 Παρουσίαση Αποτελεσμάτων Μετρικών

Οι τιμές των μετρικών που μελετάμε στο τελικό μοντέλο διαμορφώνονται όπως φαίνεται παρακάτω:

Πίνακας 4.1: Συνολικός Πίνακας Αποτελεσμάτων

Metric	Train Set	Test Set	Data Set
R2_score	0.9375	0.9365	0.9372
AUC_score	0.9152	0.9145	0.9150
precision 0	0.95	0.95	0.95
recall 0	0.96	0.96	0.96
F1_score 0	0.96	0.96	0.96
precision 1	0.89	0.89	0.89
recall 1	0.87	0.87	0.87
F1_score 1	0.88	0.88	0.88

Ο πίνακας αυτοσυσχέτισης των χαρακτηριστικών του παρακάτω μοντέλου φαίνεται παρακάτω:



Σχήμα 4.1: Πίνακας Συντελεστών Αυτοσυσχέτισης

Από τον πίνακα με τις συσχετίσεις των χαρακτηριστικών λαμβάνουμε κάποιες ενδιαφέρουσες πληροφορίες. Αναλυτικότερα, παρατηρούμε στοιχεία των δεδομένων που είχαν καταγραφεί και επιβεβαιώνονται ή αναμενόμενες συσχετίσεις. Κάποια ενδιαφέροντα παραδείγματα είναι τα παρακάτω:

- Το χαρακτηριστικό που δίνει μια εκτίμηση για το αν η παραγγελία θα είναι P2P με βάση το ποσοστό P2P δρομολογίων για το συγκεκριμένο σημείο παράδοσης (LogitD_id) έχει υψηλό συντελεστή αυτοσυσχέτισης με χαρακτηριστικά που σχετίζονται με το μέγεθος (μέσο όγκο/βάρος παραγγελίας του σημείου και όγκο/βάρος παραγγελίας) (π.χ. RoundAvDPU, RoundAvKG, isAv22). Αυτό μας δείχνει ότι αν ένα σημείο σε μεγάλο μέρος των δρομολογίων δεν συνδυάζεται με άλλα, συνήθως αυτό το σημείο λαμβάνει μεγάλες παραγγελίες.
- Παρατηρούμε σημαντική συσχέτιση ανάμεσα στο μέσο μέγεθος παραγγελίας ενός σημείου και το αν έχει πολλά γειτονικά σημεία παράδοσης. Συνεπώς, αν για παράδειγμα ένα σημείο μπορεί να συνδυαστεί με πολλά άλλα πιθανώς να του μεταφέρονται μικρότεροι όγκοι και αντίστροφα.

- Ακόμα, φαίνεται να υπάρχει αλληλεξάρτηση μεταξύ του πλήθους των οχημάτων που μπορούν να παραδώσουν στη διεύθυνση αυτή και του μέσου μεγέθους παραγγελίας. Πράγματι αν σε ένα σημείο μπορούν να παραδώσουν όλα τα οχήματα και δεν υπάρχουν μη επιτρεπτά, σημαίνει ότι ακόμα και τα μεγαλύτερα φορτηγά έχουν την δυνατότητα να την προσεγγίσουν. Έτσι, υπάρχει δυνατότητα να μεταφερθούν μεγαλύτεροι όγκοι/βάρη σε ένα μόλις δρομολόγιο, αυξάνοντας κατ'αντιστοιχεία τη μέση ποσότητα παραγγελίας του σημείου.
- Συσχέτιση εντοπίζεται ανάμεσα στο μέσο μέγεθος παραγγελίας του σημείου, και το αν αυτό το σημείο είναι αποθήκη. Όπως είναι αναμενόμενο, συνήθως οι αποθήκες παραλαμβάνουν μεγαλύτερες παραγγελίες από ότι τα καταστήματα κατά μέσο όρο.
- Η ισχυρή συσχέτιση μεταξύ του όγκου και του βάρους μία παραγγελία επιβεβαιώνει και την κοινή λογική που δείχνει ότι συνήθως αν μια παραγγελία καταλαμβάνει μεγάλο χώρο κατά κανόνα έχει και μεγάλο βάρος και αντίστροφα (γιατί για παράδειγμα αποτελείται από πολλά αντικείμενα).

4.1.1.2 Αξιολόγηση Τιμών

Παραπάνω είδαμε τα αποτελέσματα εντός του δείγματος εκπαίδευσης, εκτός του δείγματος αλλά και στο σύνολο των δεδομένων (εντός και εκτός δείγματος εκπαίδευσης). Παρατηρούμε ότι δεν υπάρχει μεγάλη διαφοροποίηση των αποτελεσμάτων μεταξύ των διαφορετικών συνόλων. Αυτό σημαίνει ότι το μοντέλο όχι μόνο αποκωδικοποιεί τα μοτίβα του συνόλου εκπαίδευσης αλλά και αποκτά την ικανότητα γενίκευσης της γνώσης.

Η συμπεριφορά του στο άγνωστο σύνολο δεδομένων το καθιστά αξιόπιστο. Ειδικότερα, τα ιδιαίτερα υψηλά ποσοστά στο σύνολο δοκιμών αποδεικνύουν ότι ενδείκνυται η χρήση του καθώς μπορεί να ανταποκριθεί και να παράγει ακριβείς εκτιμήσεις. Επίσης, δείχνει ότι στο μοντέλο δόθηκε επαρκής και ισορροπημένος όγκος πληροφοριών κατά την εκπαίδευση, και αποφεύχθηκε ο κίνδυνος της υπερεκπαίδευσης ή της υποεκπαίδευσης. Αποτελέσματα που διαφοροποιούνται πολύ ανάμεσα στο σύνολο εκπαίδευσης και το σύνολο δοκιμών θα σηματοδοτούσαν ότι η εκπαίδευση του μοντέλου δεν ήταν ομαλή.

Πιο συγκεκριμένα, η τόσο υψηλή τιμή της μετρικής AUC_score δηλώνει ότι το μοντέλο κάνει πολύ ικανοποιητικό διαχωρισμό των δύο κλάσεων. Παράλληλα, το γεγονός ότι το R2_score ανέρχεται στο 93% πιστοποιεί ότι είναι πολύ περιορισμένες οι αστοχίες στις προβλέψεις.

Εστιάζοντας στην διαφοροποίηση των αποτελεσμάτων ανάμεσα στις δύο κλάσεις, διαπιστώνουμε ότι το μοντέλο είναι πιο ακριβές στην εκτίμηση της κλάσης 0 έναντι της 1. Λίγο πιο αναλυτικά, αν ανατρέξουμε στους ορισμούς της ακρίβειας και της ανάκλισης συμπεραίνουμε ότι το πλήθος των παρατηρήσεων στο οποίο το μοντέλο εκτιμά 1 ενώ έπρεπε να πει 0 (λανθασμένη πρόβλεψη) είναι πολύ μικρό σε σχέση με το πλήθος των ορθών προβλέψεων για την κλάση 0. Ομοίως σκεπτόμενοι, βλέπουμε ότι οι αστοχίες αυτές έχουν μεγαλύτερη επίδραση στην πρόβλεψη της κλάσης 1 καθώς είναι και λιγότερες οι παρατηρήσεις που ανήκουν στην κλάση 1 και εκτιμώνται ορθά. Το παραπάνω, ωστόσο, μπορεί να οφείλεται και στην ανισορροπία του δείγματος, καθώς το 70% των παρατηρήσεων ανήκει στην κλάση 0 και μόλις το 30% στην κλάση 1. Για τους ίδιους λόγους, διαμορφώνονται όπως φαίνεται τα F1_score που αποτελούν τον αρμονικό μέσο της ακρίβειας και της ανάκλισης.

Συνολικά μπορούμε να πούμε ότι το σύνολο δεδομένων ήταν αρκετά κατάλληλο για να παραχθούν εκτιμήσεις για αυτό. Πέρα όμως από αυτό, ακολούθησε η σωστή ανάλυση του διαθέσιμου συνόλου και ο εντοπισμός των μοτίβων στις συμπεριφορές του. Επίσης, δόθηκαν οι απαραίτητες πληροφορίες και στη σωστή μορφή ώστε να προσαρμοστεί ορθά το μοντέλο στις πληροφορίες που του παρέχονταν στην είσοδο. Ακόμα, χρησιμοποιήθηκε η κατάλληλη συνάρτηση και ένας αποτελεσματικός αλγόριθμος για την εύρεση των βέλτιστων βαρών για τα χαρακτηριστικά, με αποτέλεσμα την πλήρη εκπαίδευση του. Τελικά, όλα τα παραπάνω συνετέλεσαν στην σχεδίαση ενός πολύ αποτελεσματικού μοντέλου μηχανικής μάθησης για την εκτίμηση της κατάλληλης μεθόδου δρομολόγησης της κάθε παραγγελίας.

4.1.2 Αξιολόγηση Αποτελεσμάτων με Προσωπική Παρατήρηση

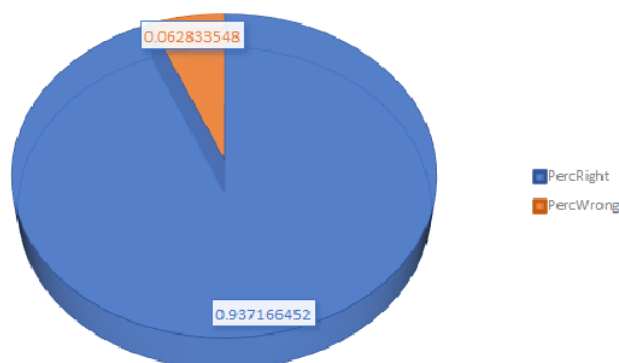
Για την επιβεβαίωση της ορθότητας των παραπάνω τιμών, ακολούθησε η ανάλυση των εξόδων του μοντέλου. Στο σημείο αυτό αναλύονται και οπτικοποιούνται ορισμένες παρατηρήσεις που έγιναν μελετώντας τις εξόδους του, με την τεχνική του *manual inspection*, και σχολιάζονται οι παραπάνω διαπιστώσεις.

4.1.2.1 Διαγραμματική Παρουσίαση Παρατηρήσεων

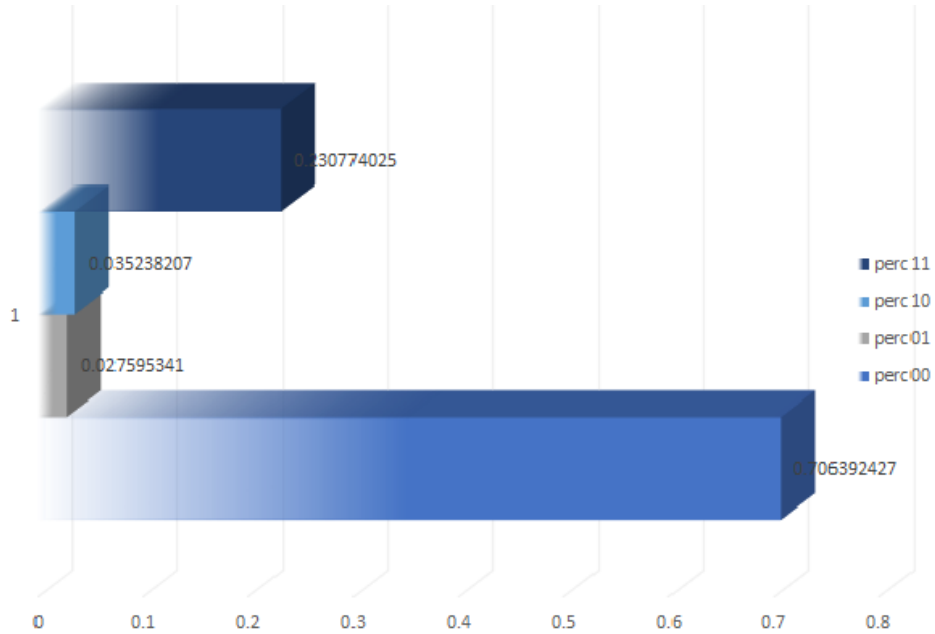
Μελετήθηκε η έξοδος του μοντέλου και δόθηκε έμφαση στην αναγνώριση συστηματικών σφαλμάτων και στη διόρθωση αυτών. Έπειτα από διάφορες διορθώσεις στο μοντέλο και επανέλεγχους των αποτελεσμάτων καταλήξαμε στο τελικό μοντέλο. Οι δείκτες που φαίνονται παρακάτω ερμηνεύονται ως εξής:

- **T1P1:** Ανήκει στην κλάση 1 και η τιμή που εκτιμήθηκε ήταν 1 (ορθή πρόβλεψη)
- **T1P0:** Ανήκει στην κλάση 1 και η τιμή που εκτιμήθηκε ήταν 0 (λανθασμένη πρόβλεψη)
- **T0P1:** Ανήκει στην κλάση 0 και η τιμή που εκτιμήθηκε ήταν 1 (λανθασμένη πρόβλεψη)
- **T0P0:** Ανήκει στην κλάση 0 και η τιμή που εκτιμήθηκε ήταν 0 (ορθή πρόβλεψη)

Στα διαγράμματα που θα ακολουθήσουν μπορούμε να δούμε αρχικά το μέρος των εκτιμήσεων που είναι ορθές έναντι των λανθασμένων ενώ έπειτα φαίνονται οι ποσοστιαίες τιμές της κάθε ομάδας παραγγελιών. Δηλαδή, διαχωρίσαμε τις παραγγελίες σε 4 κατηγορίες ανάλογα με την πραγματική την προβλεπόμενη τιμή τους (T0P0, T1P1, T1P0, T0P1), υπολογίσαμε το πλήθος των παραγγελιών σε κάθε κατηγορία και έπειτα διαιρέσαμε τις τιμές αυτές με το συνολικό πλήθος παραγγελιών.

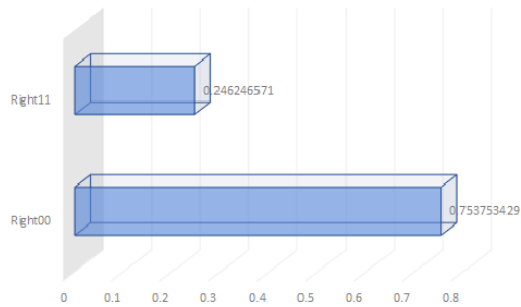


Σχήμα 4.2: Ποσοστό Ορθών και Λανθασμένων Εκτιμήσεων

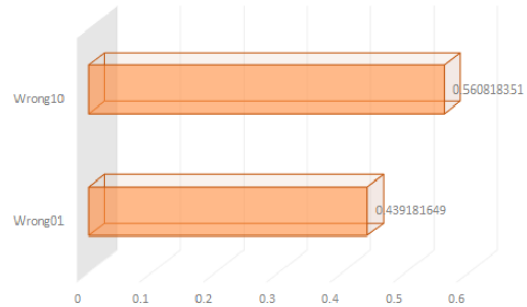


Σχήμα 4.3: Ποσοστό ανά Κατηγορία Παραγγελίας

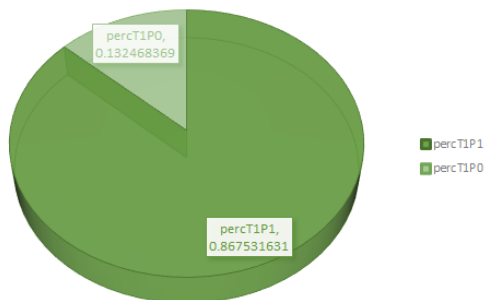
Στη συνέχεια παρουσιάζονται τα διαγράμματα που περιγράφουν την κατανομή των ορθών και λανθασμένων εκτιμήσεων, αλλά και το πως διαμορφώνονται οι εκτιμήσεις για τις 2 κλάσεις.



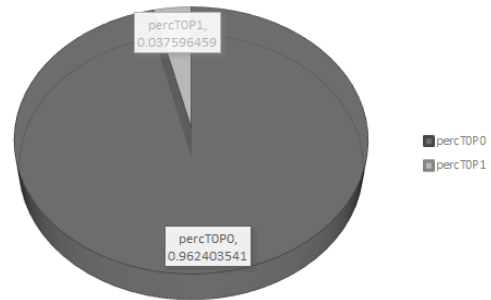
Σχήμα 4.4: Διάγραμμα Ορθών Εκτιμήσεων



Σχήμα 4.5: Διάγραμμα Εσφαλμένων Εκτιμήσεων



Σχήμα 4.6: Διάγραμμα Παραγγελιών της Κλάσης 1 (P2P)



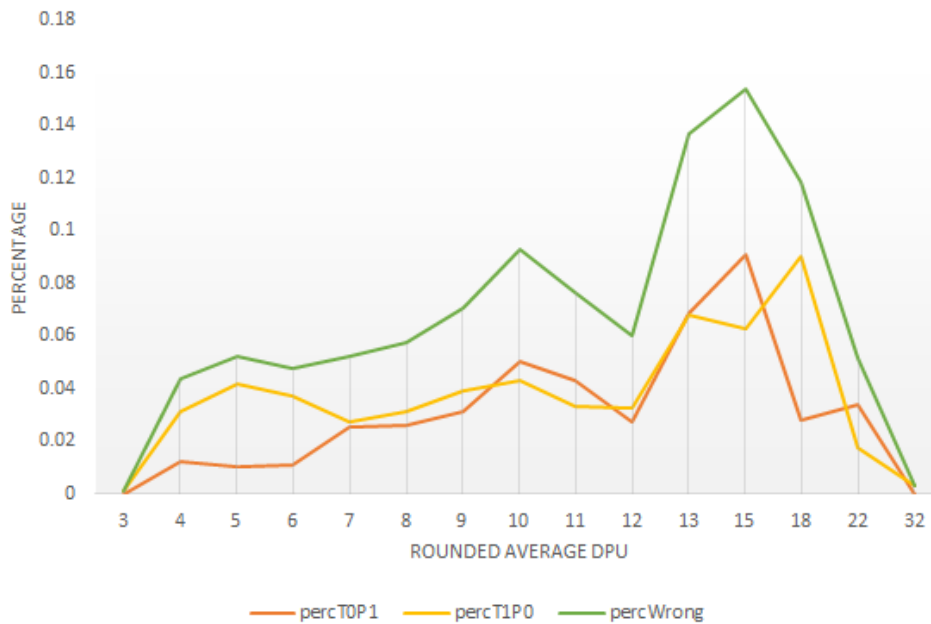
Σχήμα 4.7: Διάγραμμα Παραγγελιών της Κλάσης 1

- Ποσοστά T0P0, T1P1 ως προς το σύνολο των ορθών παραγγελιών
- Ποσοστά T0P1, T1P0 ως προς το σύνολο των λανθασμένων παραγγελιών
- Ποσοστά ορθών έναντι λανθασμένων προβλέψεων της κλάσης 1

- Ποσοστά ορθών έναντι λανθασμένων προβλέψεων της κλάσης 0

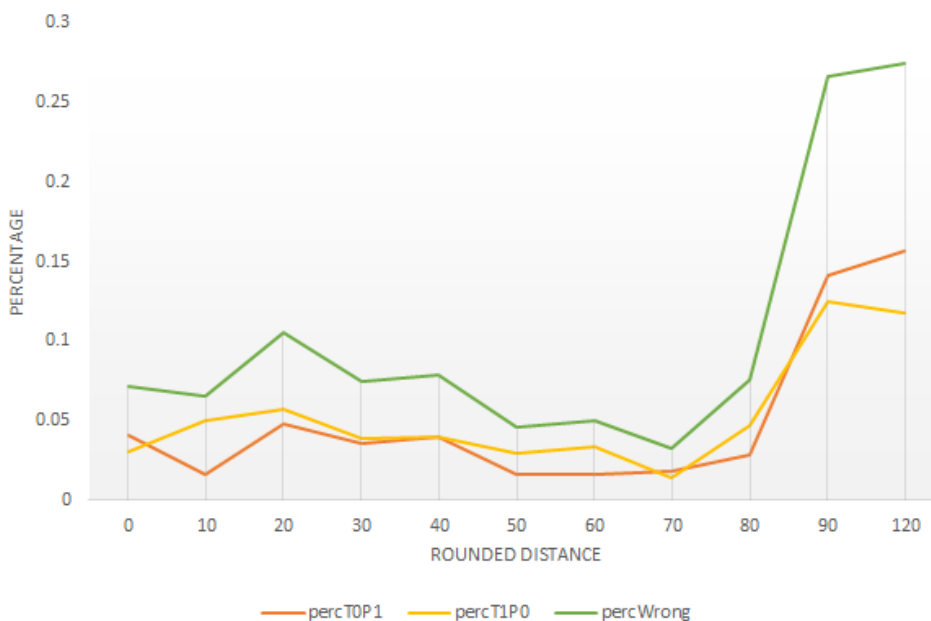
Στη συνέχεια δόθηκε έμφαση στα σφάλματα, για να ελεγχθεί εάν αυτό είναι συστηματικά και συνεπώς μπορούν να διορθωθούν, ή αν είναι κυρίως τυχαία και δεν μπορούν να προβλεφθούν. Μελετήθηκε, λοιπόν, η συμπεριφορά του μοντέλου ως προς διάφορα στοιχεία που περιγράφουν μια παραγγελία. Παρακάτω δίνονται τα διαγράμματα που παρουσίασαν ενδιαφέρον.

1. Διάγραμμα για την απεικόνιση του είδους των σφαλμάτων ανάλογα με τον όγκο της παραγγελίας.



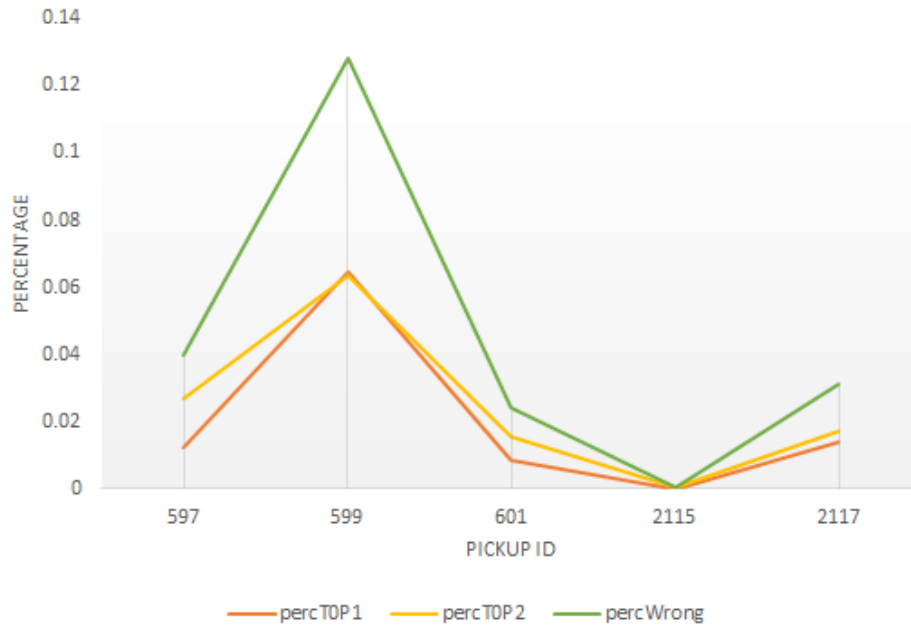
Σχήμα 4.8: Ποσοστά Σφαλμάτων ανά Μέσο Όγκο Παραγγελίας

2. Διάγραμμα που απεικονίζει το πλήθος των σφαλμάτων της κάθε κατηγορίας



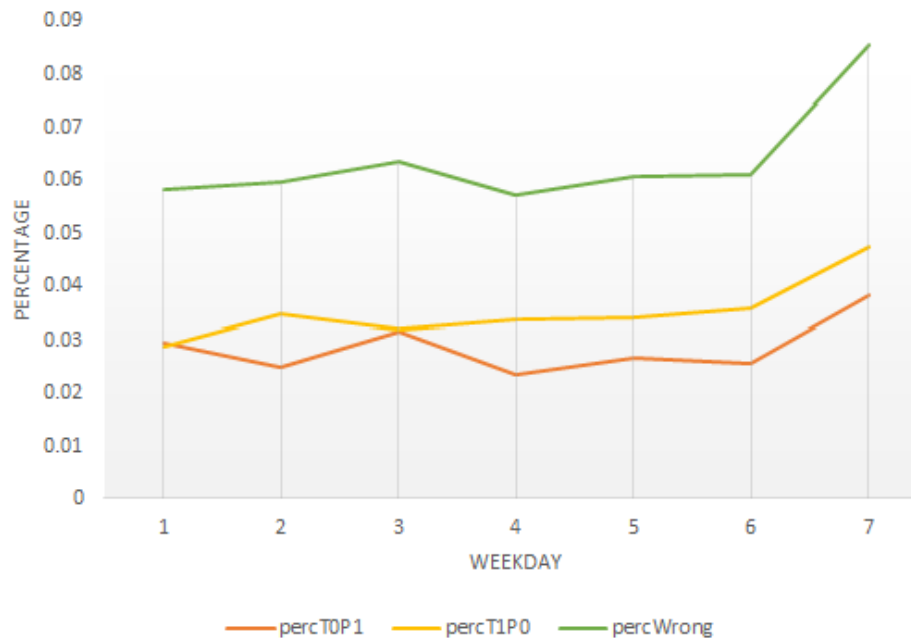
Σχήμα 4.9: Ποσοστά Σφαλμάτων ανά Απόσταση Σημείων Παράδοσης-Παραλαβής

3. Διάγραμμα που απεικονίζει το πλήθος των σφαλμάτων της κάθε κατηγορίας ανάλογα με το σημείο παραλαβής



Σχήμα 4.10: Ποσοστά Σφαλμάτων ανά Σημείο Παραλαβής

4. Διάγραμμα που απεικονίζει το πλήθος των σφαλμάτων της κάθε κατηγορίας ανάλογα με την ημέρα παράδοσης



Σχήμα 4.11: Ποσοστά Σφαλμάτων ανά Μέρα της Εβδομάδας

Αντίστοιχοι έλεγχοι έγιναν για τον εντοπισμό της σχέσης ανάμεσα στον τύπο του σφάλματος και το πλήθος των γειτονικών σημείων στο σημείο παράδοσης, το πλήθος των διαφορετικών ειδών προϊόντος που περιέχονται στην παραγγελία και άλλα. Στις περιπτώσεις αυτές δεν παρατηρήθηκε κάποιο μοτίβο στα σφάλματα και γι' αυτό δεν παρουσιάστηκαν τα αντίστοιχα διαγράμματα.

4.1.2.2 Αξιολόγηση Παρατηρήσεων

Με βάση τα παραπάνω διαπιστώνουμε ότι το πλήθος των λανθασμένων εκτιμήσεων είναι πολύ μικρό σε σχέση με το πλήθος των ορθών εκτιμήσεων. Συνεπώς μπορούμε να επιβεβαιώσουμε ότι το μοντέλο προσαρμόστηκε ορθά στα δεδομένα και έτσι είναι ικανό να παράγει αξιόπιστες εκτιμήσεις. Παρατηρούμε, δηλαδή, ότι επαληθεύονται και πειραματικά οι τιμές των παραπάνω μετρικών.

Επίσης, φαίνεται ότι δεν υπάρχει συστηματικότητα στα σφάλματα. Αυτό σημαίνει ότι τα δεδομένα δεν επιτρέπουν πολύ καλύτερη προσαρμογή καθώς εμπεριέχουν τυχαιότητα. Σε αυτό το σημείο αξίζει να αναφερθεί ότι περαιτέρω διορθώσεις του μοντέλου δεν οδηγούσαν σε περιορισμό των συνολικών σφαλμάτων αλλά σε αλλαγή του είδους τους. Αυτό μπορεί να περιγραφεί ως εξής:

Έστω ότι η μέθοδος της λογιστικής παλινδρόμησης αρχικά αδυνατούσε να κατηγοριοποιήσει ορθά τα αντικείμενα της κλάσης 1. Αυτό συνέβαινε καθώς έκανε περισσότερες προβλέψεις υπέρ της κλάσης 0. Με την τροποποίηση των χαρακτηριστικών, φαινόταν να γίνεται περισσότερο εύστοχο στην εκτίμηση της κλάσης 1. Με μία καλύτερη παρατήρηση όμως των αποτελεσμάτων, φανερωνόταν ότι αυτή η αλλαγή δεν βελτιώνει συνολικά το μοντέλο, καθώς έκανε πλέον περισσότερες λανθασμένες εκτιμήσεις της κλάσης 0.

4.2 Αποτελέσματα στο Δοκιμαστικό Περιβάλλον

4.2.1 Παρουσίαση Διαφορετικών Δοκιμών

Όπως προαναφέρθηκε, έγιναν δύο διαφορετικές προτάσεις αξιοποίησης του μοντέλου για επίλυση του προβλήματος δρομολόγησης οχημάτων. Οι δοκιμές διαμορφώθηκαν ως εξής:

1. Στο δοκιμαστικό περιβάλλον εισήχθησαν οι παραγγελίες και οι διαθέσιμοι πόροι για τη μεταφορά των αγαθών μιας μέρας μέσω φορτίου εργασίας. Έπειτα, χρησιμοποιήθηκαν τα αποτελέσματα του αλγορίθμου για αυτή τη μέρα. Δηλαδή, από τις παραγγελίες επιλέχθηκαν αυτές που δρομολογούνται μόνες τους. Η κάθε μία από αυτές σχημάτισε ένα δρομολόγιο και τα δημιουργημένα δρομολόγια δόθηκαν σαν αρχική λύση στον αλγόριθμο βελτιστοποίησης. Αυτός είχε τη δυνατότητα να τροποποιήσει τα δημιουργημένα δρομολόγια και να συνθέσει καινούργια, εφόσον το επιτρέπουν οι διαθέσιμοι πόροι, για να οργανώσει κατά βέλτιστο τρόπο τη μεταφορά των παραγγελιών από τις αποθήκες στα καταστήματα ή σε άλλες αποθήκες. Όταν ολοκληρώθηκε η διαδικασία της βελτιστοποίησης, τα δημιουργημένα δρομολόγια εξήχθησαν από τον αλγόριθμο και μελετήθηκαν οδηγώντας στις παρατηρήσεις που θα ακολουθήσουν.
2. Στη συνέχεια, τα δεδομένα που αφορούν την δρομολόγηση (παραγγελίες, δρομολόγια και διαθέσιμοι πόροι συγκεκριμένης ημέρας) αφαιρέθηκαν από το δοκιμαστικό περιβάλλον. Όλες οι παραγγελίες της ίδιας ημέρας δόθηκαν όπως πριν στον αλγόριθμο μηχανικής μάθησης και υπολογίστηκε ποιες πρόκειται να δρομολογηθούν ανεξάρτητα. Σε αυτές, αντίστοιχα με την παραπάνω δοκιμή, ανατέθηκαν τα κατάλληλα οχήματα. Έπειτα, οι υπόλοιπες παραγγελίες καθώς και τα εναπομείναντα διαθέσιμα οχήματα για την ημέρα αυτή καταχωρήθηκαν στο περιβάλλον των δοκιμών. Εκεί, αφέθηκε ο αλγόριθμος να εκτελεστεί για να σχεδιάσει τα

υπόλοιπα δρομολόγια για τη μεταφορά των παραγγελιών που δεν είχαν κριθεί ως ανεξάρτητες. Τα αποτελέσματα τόσο του αλγορίθμου μηχανικής μάθησης όσο και του αλγορίθμου βελτιστοποίησης δρομολόγησης συγκεντρώθηκαν και αθροίστηκαν για να υπολογιστεί το συνολικό κόστος δρομολόγησης της συνδυαστικής λύσης.

Για χάριν πληρότητας, εκτελέστηκε μία ακόμα δοκιμή στο δοκιμαστικό περιβάλλον. Σε αυτή την περίπτωση, για την ίδια μέρα ο αλγόριθμος βελτιστοποίησης κλήθηκε να δρομολογήσει τις παραγγελίες αξιοποιώντας τους διαθέσιμους πόρους χωρίς να αρχικοποιηθεί. Τα αποτελέσματα επίσης συλλέχθηκαν και αξιοποιήθηκαν για σύγκριση.

Πριν την σύγκριση των αποτελεσμάτων, υπολογίστηκαν οι μετρικές που αξιολογούνται στο πλαίσιο αυτής της εργασίας και για τα πραγματικά δεδομένα δρομολόγησης από το παραγωγικό περιβάλλον του πελάτη. Έτσι, στη συνέχεια παρουσιάζονται συγκριτικά διαγράμματα ανάμεσα στα αποτελέσματα των διαφορετικών μεθόδων.

Στα διαγράμματα που θα ακολουθήσουν οι διαφορετικές μεθοδολογίες επίλυσης του προβλήματος περιγράφονται ως εξής:

- Real: Αποτελέσματα από δρομολόγηση στα πραγματικά δεδομένα από το παραγωγικό περιβάλλον
- Model: Αποτελέσματα μοντέλου μηχανικής μάθησης
- Algo-Init: Αποτελέσματα αλγορίθμου βελτιστοποίησης αρχικοποιημένου με βάση την λύση του μοντέλου μηχανικής μάθησης
- Algo-NotInit: Αποτελέσματα αλγορίθμου βελτιστοποίησης χωρίς αρχικοποίηση
- Algo-NotP2P: Αποτελέσματα αλγορίθμου για τις παραγγελίες που το μοντέλο μηχανικής μάθησης έχει κρίνει ότι πρόκειται να συνδυαστούν
- Algo+Model: Άθροισμα αποτελεσμάτων συνδυασμού μοντέλου μηχανικής μάθησης για ανεξάρτητες παραγγελίες (P2P) και αλγορίθμου βελτιστοποίησης για υπόλοιπες

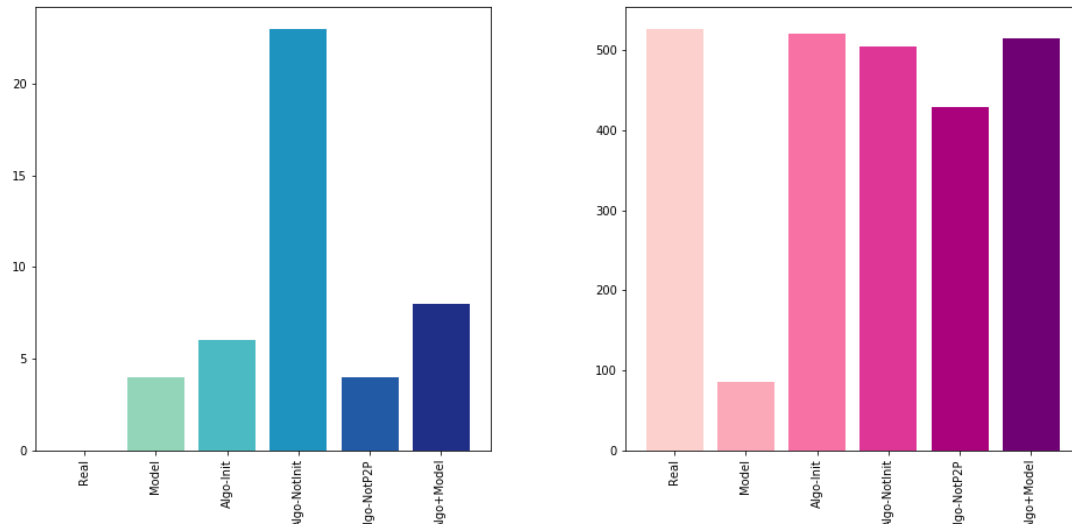
4.2.2 Σύγκριση Αποτελεσμάτων

Σε αυτό το σημείο θα δοθούν βασικά χαρακτηριστικά περιγραφής των δεδομένων με τα οποία αρχικοποιήθηκαν τα διαφορετικά περιβάλλοντα:

1. 527 παραγγελίες
2. 90 παραγγελίες κρίθηκαν ως P2P από το μοντέλο μηχανικής μάθησης
3. 437 παραγγελίες απέμειναν για τον αλγόριθμο βελτιστοποίησης στη συνδυαστική λύση
4. 267 διαθέσιμα δρομολόγια
5. 90 δρομολόγια αξιοποιήθηκαν από το μοντέλο μηχανικής μάθησης
6. 177 δρομολόγια απέμειναν για τον αλγόριθμο βελτιστοποίησης στη συνδυαστική λύση
7. 4 από τις προγραμματισμένες ως P2P παραγγελίες δεν χωρούσαν σε κανένα από τα διαθέσιμα οχήματα

4.2.2.1 Αποτελέσματα Διαχείρισης Παραγγελιών

Στα διαγράμματα που ακολουθούν φαίνεται στα αριστερά το πλήθος των παραγγελιών που δεν κατάφεραν να τοποθετηθούν σε δρομολόγια και στα δεξιά το πλήθος των παραγγελιών που δρομολογήθηκαν.



Σχήμα 4.12: Πλήθος Μη Δρομολογημένων Πα-ραγγελιών Σχήμα 4.13: Πλήθος Δρομολογημένων Παραγγελιών

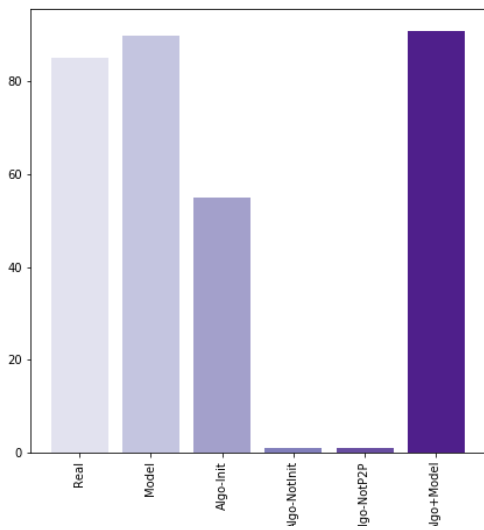
Επειδή το μοντέλο δεν χρησιμοποιήθηκε ανεξάρτητο για πλήρη δρομολόγηση, αλλά είτε για αρχικοποίηση του αλγορίθμου είτε για συνδυαστική λύση με τον αλγόριθμο βελτιστοποίησης θα εστιάσουμε στα υπόλοιπα αποτελέσματα.

Σχετικά με το αριστερό διάγραμμα, παρατηρούμε ότι η λύση του αρχικοποιημένου αλγορίθμου αφήνει μόλις 6 αδρομολόγητες παραγγελίες, ο συνδυασμός των αποτελεσμάτων του μοντέλου και του αλγορίθμου 8, ενώ ο μη αρχικοποιημένος αλγόριθμος αφήνει 23 παραγγελίες εκτός δρομολογίων. Η διαφοροποίηση του τελευταίου πιθανώς οφείλεται στο ότι χρησιμοποιήθηκαν τα μεγαλύτερα οχήματα για την συνδυαστική δρομολόγηση πολλών παραγγελιών και αυτό είχε ως αποτέλεσμα κάποιες μεγάλες παραγγελίες να μην μπορούν να τοποθετηθούν στα υπολειπόμενα. Ακόμα, αξίζει να σχολιαστεί ότι το μηδενικό πλήθος στα πραγματικά δεδομένα οφείλεται σε παραβίαση του όγκου. Φάνηκε δηλαδή, ότι οι 6 παραγγελίες που δεν κατάφεραν να δρομολογηθούν από τον αρχικοποιημένο αλγόριθμο δεν ήταν δυνατό να χωρέσουν σε κανένα από τα διαθέσιμα οχήματα.

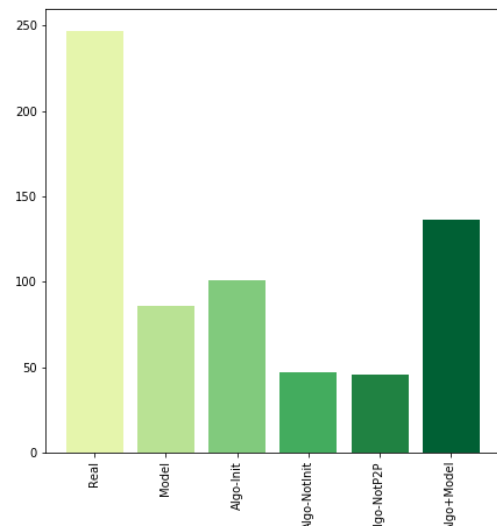
Στο διάγραμμα που βρίσκεται στα δεξιά, παρουσιάζεται το πλήθος των παραγγελιών που δρομολογήθηκαν από τις διαφορετικές μεθόδους. Το άθροισμα των δρομολογημένων και των αδρομολόγητων ταυτίζεται με το συνολικό πλήθος των παραγγελιών προς δρομολόγηση για τη συγκεκριμένη ημέρα και είναι ίσο με 527 παραγγελίες.

4.2.2.2 Αποτελέσματα Διαχείρισης Δρομολογίων

Στην υποενότητα αυτή παρουσιάζονται τα διαγράμματα της διαχείρισης των πόρων και των ανεξάρτητων παραγγελιών.



Σχήμα 4.14: Πλήθος Δρομολογίων 2 Στάσεων



Σχήμα 4.15: Σύνολο Δρομολογίων

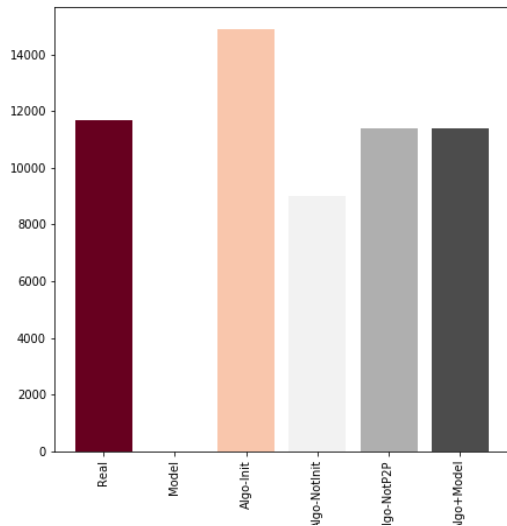
Αναφορικά με την απεικόνιση του πλήθους των P2P δρομολογίων παρατηρείται ότι ο συνδυασμός του μοντέλου με τον αλγόριθμο δίνουν την μεγαλύτερη τιμή, η οποία ωστόσο σχεδόν ταυτίζεται με τη λύση που επέλεξε ο πελάτης. Αντίθετα, ο μη αρχικοποιημένος αλγόριθμος τείνει να συνδυάζει σχεδόν σε όλα τα δρομολόγια τουλάχιστον δύο παραγγελίες, σχηματίζοντας μόλις ένα δρομολόγιο P2P. Από την άλλη, σημαντικό είναι να σημειωθεί ότι το μοντέλο ήταν εύστοχο στο 96.6% των παραγγελιών. Τέλος, ο αρχικοποιημένος αλγόριθμος φαίνεται να σχηματίζει μια λύση κάπου στο ενδιάμεσο των λύσεων των πραγματικών δεδομένων και του μη αρχικοποιημένου μοντέλου. Έτσι, είναι σαν να ομοιάζει στο τι έγινε στην πραγματικότητα, προσπαθώντας να το βελτιστοποιήσει όσο αυτό είναι δυνατό. Τέλος, αξίζει να σημειωθεί ότι ο αλγόριθμος δεν προσέθεσε άλλα P2P δρομολόγια πέραν αυτών που είχαν δημιουργηθεί στην αρχική λύση, μόνο σε ορισμένα προσέθεσε παραγγελίες αλλάζοντας των χαρακτηρισμό τους σε μη-P2P.

Αναλογικά, παρατηρούμε ότι η λύση που χαρακτηρίζεται ως Algo-NotInit δημιουργεί τα λιγότερα συνολικά δρομολόγια ενώ η Algo+Model αν και απασχολεί μεγαλύτερο πλήθος πόρων, βρίσκεται πιο κοντά στην επιλογή του πελάτη. Σημαντικό είναι να αναφερθεί ότι σε σύγκριση που έγινε με την πραγματική δρομολόγηση της αντίστοιχης ημέρας, η λύση Algo+Model όχι μόνο εντοπίζει με πολύ μεγάλη ακρίβεια τα Peer-to-Peer δρομολόγια, αλλά και συνολικά τοποθετεί τις παραγγελίες σε αντίστοιχα δρομολόγια. Ειδικότερα, το 49% των παραγγελιών τοποθετούνται σε δρομολόγια σε δρομολόγια με τις ίδιες στάσεις παραλαβής ενώ το 84.5% τοποθετείται σε δρομολόγια με απόκλιση μέχρι μία στάση. Σχετικά με τις στάσεις παράδοσης, βλέπουμε ότι η συνδυαστική λύση προσπαθεί να ελαχιστοποιήσει τα χιλιόμετρα και τα οχήματα με αποτέλεσμα να αυξάνει το πλήθος των στάσεων για παράδοση. Έτσι, με τα ίδια σημεία παραλαβής προσπαθεί να μεγιστοποιήσει το πλήθος των σημείων παράδοσης, γεγονός που δημιουργεί απόκλιση στο πλήθος τους σε σχέση με αυτό που καταγράφεται για τα πραγματικά δεδομένα. Τέλος, ο αρχικοποιημένος αλγόριθμος είναι ουσιαστικά η βελτιωμένη εικόνα των πραγματικών δεδομένων, απασχολώντας μεγαλύτερο πλήθος πόρων από τον μη αρχικοποιημένο αλγόριθμο αλλά σημαντικά λιγότερα από τα πραγματικά.

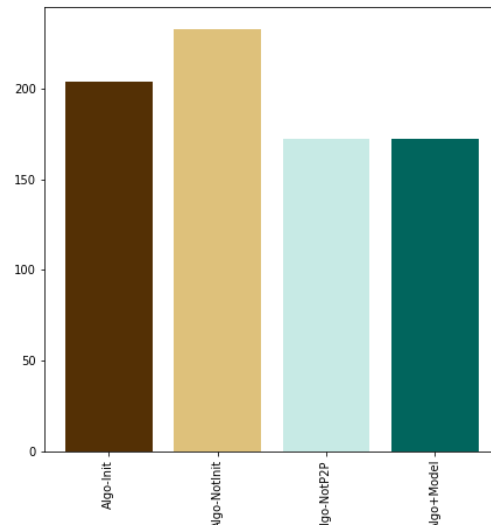
4.2.2.3 Αποτελέσματα στον Χρόνο Εκτέλεσης και στην Απόσταση

Στο πρώτο διάγραμμα παρατηρούμε την συνολική διανυόμενη απόσταση των δρομολογίων ενώ στο δεύτερο βλέπουμε το χρόνο εκτέλεσης των διαφορετικών δοκιμών.

Ο χρόνος εκτέλεσης που χρειάστηκε για τα πραγματικά δεδομένα ήταν δυνατό να ανακτηθεί ενώ ο χρόνος εκτέλεσης του αλγορίθμου μηχανικής μάθησης για την παραγωγή εκτιμήσεων, μετά την εκπαίδευσή του, ήταν αμελητέος.



Σχήμα 4.16: Διανυόμενη Απόσταση



Σχήμα 4.17: Χρόνος Εκτέλεσης

Παρατηρούμε ότι η χειρότερη λύση από άποψη διανυόμενης απόστασης είναι αυτή του αρχικοποιημένου αλγορίθμου, που διανύει περίπου 20% μεγαλύτερη απόσταση από την λύση στο παραγωγικό περιβάλλον. Από την άλλη, η δοκιμή με το χαρακτηρισμό Algo_NotInit δίνει την καλύτερη λύση. Ανάμεσα σε αυτήν την λύση και τη λύση του Real, βρίσκεται η συνδυαστική λύση του αλγορίθμου βελτιστοποίησης με το μοντέλο μηχανικής μάθησης.

Σχετικά με τους χρόνους εκτέλεσης, η συνδυαστική λύση παρέχει την πιο γρήγορη δρομολόγηση καθώς ο αλγόριθμος χρησιμοποιείται μόνο για ένα μέρος των παραγγελιών. Έπειτα, ο αρχικοποιημένος αλγόριθμος δείχνει να εκτελείται αρκετά πιο γρήγορα από το μη αρχικοποιημένο. Η πρόταση μιας αρχικής λύσης φαίνεται δηλαδή να επιταχύνει τον αλγόριθμο.

4.2.2.4 Συνολική Αξιολόγηση Διαφορετικών Μεθοδολογιών

Μελετώντας συνολικά τα αποτελέσματα των παραπάνω διαγραμμάτων καταλήγουμε στις εξής παρατηρήσεις:

4.3 Συνολικά Συμπεράσματα

Παραπάνω παρουσιάστηκαν τα αποτελέσματα τόσο καθαρά από τη σκοπιά της μηχανικής μάθησης όσο και από τη σκοπιά της δρομολόγησης οχημάτων και ανάθεσης πόρων.

Από τη μία μεριά, οι ιδιαίτερα ικανοποιητικές τιμές των μετρικών αξιολόγησης του μοντέλου μηχανικής έδειξαν ότι επιλέχθηκαν κατάλληλα η μέθοδος εκτίμησης και ο αλγόριθμος βελτιστο-

Πίνακας 4.2: Συνολικός Πίνακας Αποτελεσμάτων Δοκικών στο Περιβάλλον

Diagram	Algo_Init	Algo_NotInit	AlgoModel
Αδρομολόγητες παραγγελίες	Min	Max	Κοντά στο Min
Πλήθος P2P δρομολίων	Ενδιάμεσο	Min	Max αλλά κοντά στα Real
Πλήθος δρομολογίων	Ενδιάμεσο	Min	Max αλλά κοντά στα Real
Διανύομενη Απόσταση	Max	Min	Κοντά Min
Χρόνος Εκτέλεσης	Ενδιάμεσο	Max	Min

ποίησης. Παράλληλα, έγινε η επαλήθευση των παραπάνω αποτελεσμάτων με τον εποπτικό έλεγχο με το χέρι. Από αυτά μπορούμε να συμπεράνουμε ότι αποκωδικοποιήθηκαν σωστά τα μοτίβα στα δεδομένα και έγινε πλήρης περιγραφή του διαθέσιμου συνόλου στο μοντέλο. Έτσι, συνολικά σχεδιάστηκε ένα πολύ ακριβές μοντέλο ταξινόμησης των παραγγελιών σε δύο κλάσεις με όσο το δυνατόν καλύτερη αξιοποίηση των διαθέσιμων εργαλείων μηχανικής μάθησης.

Από την άλλη μεριά, οι δοκιμές στο δοκιμαστικό περιβάλλον μας έδωσαν την δυνατότητα παροχής μιας παλέτας λύσεων στο πρόβλημα της δρομολόγησης οχημάτων. Εάν σκοπός του χρήστη είναι να υπολογίσει σε ικανοποιητικό χρόνο μια λύση που ελαχιστοποιεί τις μη δρομολογημένες παραγγελίες και ομοιάζει με τον τρόπο που είχε μάθει να λειτουργεί ο άνθρωπος, χωρίς να τον απασχολεί ιδιαίτερα το κόστος μεταφοράς τότε η καταλληλότερη μέθοδος είναι αυτή της αρχικοποίησης του αλγορίθμου με το μοντέλο μηχανικής μάθησης. Αν, αντίθετα, προτεραιότητα είναι η μείωση του κόστους και του χρόνου μεταφοράς και ας μην μπορέσει να μεταφέρει όλες τις παραγγελίες, ενώ συγχρόνως έχει αρκετό διαθέσιμο χρόνο και υπολογιστικούς πόρους για την εύρεση λύσης, θα του προτεινόταν η επιλογή του αλγορίθμου βελτιστοποίησης χωρίς αρχικοποίηση. Τέλος, στην περίπτωση που θέλει να ακολουθήσει όσο τον δυνατόν περισσότερο τον τρόπο που έχει συνηθίσει ώστε να ελαχιστοποιηθούν οι παρεμβάσεις που χρειάζεται να κάνει ο άνθρωπος στη συνέχεια, μειώνοντας συγχρόνως την απόσταση, τους αξιοποιούμενους πόρους και τον χρόνο για εύρεση λύσης, αναμφίβολα η πιο αποτελεσματική μέθοδος είναι αυτή που συνδυάζει τα αποτελέσματα της μηχανικής μάθησης και του αλγορίθμου βελτιστοποίησης.

Κεφάλαιο 5

Επίλογος

5.1 Σύνοψη

Στην εργασία που παρουσιάστηκε παραπάνω, χρησιμοποιήθηκαν τα δεδομένα από το παραγωγικό περιβάλλον μιας εταιρείας που δραστηριοποιείται στο χώρο του λιανικού εμπορίου αγαθών.

Το σύνολο δεδομένων αποτελείται από τις παραγγελίες, τις αποθήκες, τα καταστήματα και τους διαθέσιμους πόρους της εν λόγω επιχείρησης. Η περιγραφή αυτή εκφράζει ένα πρόβλημα δρομολόγησης οχημάτων με περιορισμούς. Για την επίλυση αυτού του καθημερινού προβλήματος η αλυσίδα καταστημάτων επέλεξε το λογισμικό της εταιρείας ORTEC. Το σύνολο δεδομένων εισόδου δόθηκε στον αλγόριθμο βελτιστοποίησης με σκοπό την εύρεση του βέλτιστου συνδυασμού πόρων και δρομολογίων για τη μεταφορά αγαθών. Στη συνέχεια, οι χρήστες του συστήματος έκαναν τις παρεμβάσεις που θεωρούσαν σκόπιμες και έπειτα εκτελέστηκαν τα δρομολόγια.

Μετά από ένα χρόνο χρήσης του συστήματος, τα δεδομένα αυτά συλλέχθηκαν και μελετήθηκαν στα πλαίσια της παρούσας εργασίας. Η ομοιότητα μεταξύ των διαφορετικών ημερών παρουσίαζε ιδιαίτερο ενδιαφέρον, καθώς τόσο τα σημεία παρέμεναν σταθερά μέσα στο χρόνο (με ορισμένες προσθήκες καταστημάτων), όσο και οι όγκοι των αγαθών δεν διέφεραν πολύ μεταξύ αντίστοιχων ημερών ενός έτους παρατηρήσεων. Αυτό οδήγησε στην σκέψη ότι δεν χρειάζεται κάθε μέρα να τρέχει αλγόριθμος δρομολόγησης, μια διαδικασία που κοστίζει σε χρόνο και υπολογίζεται κάθε φορά από το μηδέν, αλλά μπορεί να αξιοποιηθεί η παρελθοντική γνώση για σύνθεση μελλοντικών λύσεων. Φάνηκε έτσι να έχει νόημα η μελέτη του προβλήματος από τη σκοπιά της τεχνητής νοημοσύνης.

Με παραπάνω ανάλυση των δεδομένων και παρατήρηση από διαφορετικές σκοπιές, φάνηκε στην συγκεκριμένη περίπτωση ένα μεγάλο ποσοστό των δρομολογίων να αποτελείται από ζεύγη δύο σημείων (μιας αποθήκης και ενός καταστήματος ή δύο αποθηκών). Υπολογίστηκε, δηλαδή, ότι το περίπου 50 τις εκατό συντίθεται από δρομολόγια μιας στάσης για παραλαβή και μιας για παράδοση, το 35 τις εκατό από δρομολόγια τριών στάσεων ενώ λιγότερο από 5 τις εκατό περιέχει πάνω από 3 στάσεις. Ακόμα, παρατηρήθηκε ομοιότητα των χαρακτηριστικών που καθόριζαν το από πόσες στάσεις θα αποτελείται ένα δρομολόγιο (όγκοι, σημείο παράδοσης, σημείο παραλαβής, απόσταση μεταξύ τους κλπ). Για την εξέλιξη της εργασίας, χρειάστηκε ο καθαρισμός των δεδομένων από λανθασμένες τιμές, σημεία που δεν υπήρχαν σε όλο το μέγεθος του δείγματος ή ροές που κάναν σύνθετο το πρόβλημα χωρίς να προσθέτουν μεγάλη γνώση σε αυτό.

Με βάση τα παραπάνω, φάνηκε ότι η κάθε παραγγελία μπορεί να αντιμετωπισθεί ως ένα αντικείμενο που δύναται να ανήκει σε δύο κλάσεις, την κλάση των παραγγελιών που δεν χρειάζεται

να συνδυαστούν με άλλες για να συνθέσουν ένα δρομολόγιο, και την κλάση αυτών που χρειάζεται ο συνδυασμός τους με άλλες. Ένα πρόβλημα δρομολόγησης οχημάτων μετασχηματίστηκε έτσι σε ένα πρόβλημα ταξινόμησης. Για την επίλυση του νέου προβλήματος, δομήθηκε ένα μοντέλο μηχανικής μάθησης που εντοπίζει τη βέλτιστη ευθεία διαχωρισμού των δεδομένων σε δύο κλάσεις κάνοντας χρήση της συνάρτησης λογιστικής παλινδρόμησης. Το μοντέλο αυτό εκπαιδεύεται σε ένα σύνολο δεδομένων του προβλήματος και έπειτα δοκιμάζεται η αποτελεσματικότητα του τόσο στα δεδομένα εκπαίδευσης όσο και σε ένα σύνολο δεδομένων άγνωστο σε αυτό.

Μετά από βελτιστοποίηση των παραμέτρων του μοντέλου ταξινόμησης, έπρεπε να μετρηθεί και η συνεισφορά του στην επίλυση του αρχικού προβλήματος. Αρχικά, δοκιμάστηκε η εφαρμογή του μοντέλου για αρχικοποίηση της λύσης του αλγορίθμου βελτιστοποίησης. Έπειτα, χρησιμοποιήθηκε το μοντέλο για να βρει τα δρομολόγια που μεταφέρουν μία παραγγελιά και ο αλγόριθμος για να συνθέσει δρομολόγια για τη μεταφορά των υπόλοιπων παραγγελιών. Παράλληλα, δοκιμάστηκε η επίλυση του προβλήματος αποκλειστικά από τον αλγόριθμο βελτιστοποίησης. Ακολούθησε η σύγκριση μεταξύ των τριών παραπάνω, καθώς και η σύγκριση αυτών με τα δεδομένα από το παραγωγικό περιβάλλον του πελάτη.

Φάνηκε ο αλγόριθμος βελτιστοποίησης να εντοπίζει λύση μικρότερου κόστους από άποψη χιλιομέτρων και απασχόλησης πόρων. Από την άλλη, η επίλυση του προβλήματος με αρχικοποίηση του αλγορίθμου από μοντέλο λογιστικής παλινδρόμησης φάνηκε να καταφέρνει να δρομολογεί μεγαλύτερο πλήθος παραγγελιών καθώς και να παρέχει μια λύση πολύ κοντινή σε αυτό που είχε επιλέξει να πραγματοποιήσει ο χρήστης του συστήματος. Τέλος, η εφαρμογή του μοντέλου λογιστικής παλινδρόμησης για εύρεση δρομολογίων δύο στάσεων και η χρήση του αλγορίθμου για την δημιουργία των υπόλοιπων δρομολογίων οδήγησε σε μια μορφή λύσης πολύ κοντά στην πραγματικά δεδομένα, χρησιμοποιώντας ωστόσο λιγότερα οχήματα, διανύοντας ελαφρώς λιγότερα χιλιόμετρα, που υπολογίστηκε στον μικρότερο υπολογιστικό χρόνο και ελαχιστοποιεί τις παρεμβάσεις που χρειάζεται να κάνει ο άνθρωπος στη συνέχεια.

Συμπερασματικά, αυτή η διπλωματική εργασία προτείνει μια παλέτα αποτελεσματικών λύσεων σε ένα πρόβλημα βέλτιστης μεταφοράς αγαθών και αξιοποίησης πόρων. Δοκιμάζονται, δηλαδή, τεχνικές μηχανικής μάθησης για την επίλυση ενός προβλήματος δρομολόγησης, καθώς και εντοπίζεται σε ποιο σημείο της επίλυσης έχουν θέση. Ακόμα, δοκιμάζεται η αντιμετώπιση του προβλήματος από τη σκοπιά της επιχειρησιακής έρευνας - μέθοδος που κυριαρχεί στη βιβλιογραφία - και τέλος ακολουθεί η σύγκριση των παραπάνω και η εξαγωγή συμπερασμάτων σχετικά με τα πλεονεκτήματα και τα μειονεκτήματα της κάθε μεθόδου.

5.2 Προοπτικές

Στο πλαίσιο της διπλωματικής εργασίας αυτής μελετήθηκε η δρομολόγηση των παραγγελιών 2 στάσεων με χρήση μεθόδου μηχανικής μάθησης. Στη συνέχεια, θα είχε νόημα η σύνθεση ενός μοντέλου για την δημιουργία παραγγελιών που συνδυάζουν δύο παραγγελίες ή και τρεις με αποτέλεσμα το σχηματισμό δρομολογίων περισσότερων στάσεων. Έτσι, θα είχε καλυφθεί το μεγαλύτερο μέρος των παραγγελιών και θα απέμεναν μόνο λίγες για τις οποίες δεν θα μπορούσε να πάρει απόφαση το μοντέλο. Ακόμα, θα μπορούσε να βρεθεί κάποια μέθοδος μηχανικής μάθησης που θα μπορούσε να δρομολογήσει και τις παραγγελίες που έχουν απομείνει ώστε συνολικά να μπορεί όλο το πρόβλημα να αντιμετωπισθεί χωρίς την χρήση αλγορίθμου βελτιστοποίησης.

Επιπλέον, θα παρουσίαζε ιδιαίτερο ενδιαφέρον η δοκιμή μεθόδων μηχανικής μάθησης σε άλλα προβλήματα βελτιστοποίησης μεταφορών. Σε ένα δίκτυο που τα σημεία δεν είναι σταθερά στο χρόνο ή σε ένα δίκτυο που οι όγκοι που μεταφέρονται είναι πολύ μικρότεροι αλλά υπάρχουν πολλά σημεία στα οποία πρέπει να γίνουν παραδόσεις θα αναζητούνταν άλλα είδη ομοιοτήτων στο χρόνο. Για παράδειγμα, για την ίδια αλυσίδα μεταπρατών λιανικού εμπορίου, θα μπορούσε να μελετηθεί η βελτιστοποίηση των μεταφορών αγαθών από τα καταστήματα ή τις αποθήκες απευθείας στους καταναλωτές. Εκεί θα έπρεπε να αναζητηθούν ομοιότητες στις ομάδες των περιοχών που συνδυάζονται, στα είδη των προϊόντων που τοποθετούνται μαζί, στη διαφοροποίηση ανάμεσα στις μέρες, στη διακύμανση του όγκου των παραγγελιών κλπ.

Μια ακόμα εφαρμογή που απαιτεί μια πολύ διαφορετική προσέγγιση είναι η περίπτωση εταιρείας ταχυδρομείων. Εδώ, κάθε πακέτο που μεταφέρεται περνάει από πολλές στάσεις μέχρι να φτάσει στον τελικό αποδέκτη. Σε ένα τέτοιο πρόβλημα, όπου κάποιος από την Κίνα θέλει να στείλει ένα δέμα στην Ελλάδα μπορεί να υπάρξουν πολλές ενδιάμεσες στάσεις και αλλαγές μέσων. Ακόμα και στα πλαίσια εντός μιας χώρας ένα γράμμα περνάει από πολλά σημεία που συλλέγεται η αλληλογραφία και αναδιανέμεται. Σε μια τέτοια μορφή μεταφορών θα αναζητούνταν κοινά στοιχεία μεταξύ του όγκου των παραγγελιών που συλλέγονται για να μεταφερθούν μαζικά στο επόμενο σημεία, στο χρόνο που παραμένουν στην κάθε στάση, στη συσχέτιση μεταξύ του πλήθους των στάσεων της κάθε παραγγελίας με την συνολική απόσταση που πρέπει να διανύσει και άλλα.

5.3 Μελλοντικές Κατευθύνσεις Επιστημονικής Μελέτης

Στα πλαίσια της εργασίας αυτής μελετήθηκαν διαφορετικές προσεγγίσεις περιγραφών του προβλήματος δρομολόγησης οχημάτων, μεθοδολογίες επεξεργασίας και ανάλυσης δεδομένων καθώς και τεχνικές μηχανικής μάθησης. Μετά την παρατήρηση αυτών των διαφορετικών πεδίων διαπιστώθηκε ότι θα μπορούσε να δοκιμαστεί ο συνδυασμός τους για διαφορετικούς σκοπούς.

Αρχικά, οι τεράστιοι όγκοι που απαιτείται να μεταφερθούν σήμερα από και προς όλες τις μεριές του πλανήτη συ συνδυασμό με πληθώρα περιορισμών που υπαγορεύονται σε κάθε περίπτωση, καθιστούν την εύρεση λύσεων ελαχιστοποίησης κόστους πολύ δαπανηρές υπολογιστικά. Ωστόσο, με προσεκτική παρατήρηση των διαφορετικών στιγμιότυπων της κάθε περίπτωσης μπορούν να εντοπισθούν ομοιότητες στο χρόνο. Ακόμα και αν οι όγκοι ή τα σημεία παραλαβής και παράδοσης διαφοροποιούνται στο χρόνο, στο μεγαλύτερο μέρος των προβλημάτων ακολουθούν κάποια πιθανοτική κατανομή. Συνεπώς, φαίνεται να μην είναι απαραίτητος ο επανυπολογισμός ολόκληρης της λύσης κάθε μέρα, αλλά να μπορεί να αξιοποιηθεί ένα μέρος της προηγούμενης γνώσης.

Για την ανάλυση της διαθέσιμης πληροφορίας υπάρχει πληθώρα διαθέσιμων μεθοδολογιών που έχουν αναπτυχθεί στα πλαίσια της επιστήμης των δεδομένων. Έτσι καθίσταται δυνατή η αναγνώριση προτύπων που επαναλαμβάνονται στο χρόνο. Αυτά στη συνέχεια μπορούν να αξιοποιηθούν από μεθόδους μηχανικής μάθησης. Συνεπώς, στο μέλλον θα φαινόταν ενδιαφέρον να μελετηθούν τρόποι για ελαχιστοποίηση του υπολογιστικού κόστους, των ανθρώπινων παρεμβάσεων ή την γρήγορη εύρεση πολύ αποδοτικών λύσεων, χάρη στον συνδυασμό των τριών αυτών πεδίων. Ενδιαφέρουσες προτάσεις για μελέτη φαίνονται να είναι:

- η αξιοποίηση της διαθέσιμης γνώσης για παραγωγή οικείων λύσεων στον άνθρωπο, ελαχιστοποιώντας έτσι τις παρεμβάσεις του αλλά και αυξάνοντας την αποδοτικότητα του (αφού από τη φύση του ο άνθρωπος όταν επαναλαμβάνει μια διαδικασία μαθαίνει να την κάνει όλο και καλύτερα)
- η σχεδίαση συνόλων μέσα από πιθανοτικές κατανομές που αποτυπώνουν τις ομοιότητες ανάμεσα στα διαφορετικά στιγμιότυπα για εκπαίδευση μοντέλων μηχανικής μάθησης που επιλύουν το ΠΔΟ σε περιπτώσεις που δεν υπάρχει καταγεγραμμένη η προηγούμενη γνώση
- η βελτιστοποίηση των υπερπαραμέτρων αλγορίθμων βελτιστοποίησης με χρήση τεχνικών τεχνητής νοημοσύνης
- η μερική λύση του προβλήματος για επιτάχυνση αλγορίθμων βελτιστοποίησης ή ο συνδυασμός μερικών λύσεων για επίλυση του συνόλου

Βιβλιογραφία

- Components of time series. <https://www.toppr.com/guides/business-mathematics-and-statistics/time-series-analysis/components-of-time-series/>, 2017.
- Will Badr. Why feature correlation matters...a lot! <https://towardsdatascience.com/why-feature-correlation-matters-a-lot-847e8ba439c4>, 2018.
- P. Cooray and Thashika Rupasinghe. Machine learning-based parameter tuned genetic algorithm for energy minimizing vehicle routing problem. *Journal of Industrial Engineering*, 2017:1–13, 01 2017. doi: 10.1155/2017/3019523.
- M. Desrochers and T. W. Verhoog. A matching based savings algorithm for the vehicle routing problem. 1989.
- S.Ropke D.Pisinger. A general heuristic for vehicle routing problems. *Computers and Operations Research*, 34(8):2403–2435, 2007.
- J.K.Lenstra E.Aarts. Local search in combinatorial optimization. 2003.
- E.Taillard. Parallel iterative search methods for vehicle routing problems. *Networks*, 23(8):661–673, 1993.
- M. L. Fisher. Optimal solution of vehicle routing problems using minimum k-trees. *Operations Research*, (42):626–642, 1994.
- G.B.Dantzig and J.H.Ramser. The truck dispatching problem. *Management Science*, 6(1):1–140, 1959.
- G.Clarke and J.W.Wright. Scheduling of vehicles from a central depot to a number of delivery points. *Operations Research*, 12(4):568–581, 1964.
- G.Laporte, M.Desrochers, and Y.Nobert. Two exact algorithms for the distance-constrained vehicle routing problem. *Networks*, 14(1):161–172, 1984.
- G.Laporte, P.Toth, and D.Vigo. Vehicle routing: historical perspective and recent contributions. *EURO Journal on Transportation and Logistics*, 2(1-2), 2013.
- G.Righini. Algorithms for the vehicle routing problem with time windows. *Transportation Science*, 29(2):156–166, 1995.
- G.Righini. Approximation algorithms for the vehicle routing problem with pick-up and delivery. *Università degli Studi di Milano*, 2000.

- <http://neo.lcc.uma.es/vrp/solution-methods/>. Solution methods vrp. <http://neo.lcc.uma.es/vrp/solution-methods/>, 2013.
- <https://emerj.com/ai-glossary-terms/what-is-machine-learning/>. What is machine learning? <https://emerj.com/ai-glossary-terms/what-is-machine-learning/>, 2019.
- Urvashi Jaitley. Why data normalization is necessary for machine learning models. <https://medium.com/@urvashilluniya/why-data-normalization-is-necessary-for-machine-learning-models-681b65a05029>, 2018.
- J.B.Rollins. Foundational methodology for data science. *IBM Analytics*, 2015.
- David Masip Laura Calvet, J sica de Armas and Angel A. Juan. Learnheuristics: hybridizing metaheuristics with machine learning for optimization with dynamic inputs. *Open Mathematics*, (15):261–280, 2017.
- Bai Lo. Useful properties of roc curves and auc scoring. <https://www.kaggle.com/learn-forum/53782>, 2018.
- M.Gendreau, A.Hertz, and G.Laporte. A tabu search heuristic for the vehicle routing problem. *Management Science*, 40(10):1276–1290, 1994.
- E.G.Alexandrakis M.G.Lioudakis. Emotional analysis in greek text using machine learning algorithms. 2017.
- Alfonso Misevicius. Ruin and recreate principle based approach for the quadratic assignment problem. pages 598–609, 07 2003.
- Lawrence V. Snyder-Martin Tak c Mohammadreza Nazari, Afshin Oroojlooy. Reinforcement learning for solving the vehicle routing problem. *arXiv:1802.04240*, 2003.
- P.L.Juell Nagesh Kadaba, K.E.Nygaard. Integration of adaptive machine learning and knowledge-based systems for routing and scheduling applications. *Expert Systems with Applications*, 2(1): 15–27, 1991.
- N.Christofides, A.Mingozzi, and P.Toth. Exact algorithms for the vehicle routing problem, based on spanning tree and shortest path relaxations. *Mathematical Programming*, 20(1):255–282, 1981a.
- N.Christofides, A.Mingozzi, and P.Toth. State-space relaxation procedures for the computation of bounds to routing problems. *Networks*, 11(2):145–164, 1981b.
- C. Rodriguez O. Arbelaitz and I. Zamakola. Low cost parallel solutions for the vrptw optimization problem. *2001 International Conference on Parallel Processing Workshops*, pages 176–181, 2001.
- F. Guertin J.-Y. Potvin P. Badeau, M. Gendreau and E. Taillard. A parallel tabu search heuristic for the vehicle routing problem with time windows. *Transportation Research*.
- D.Vigo P.Toth. The granular tabu search and its application to the vehicle-routing problem. *INFORMS Journal on Computing*, 15(4):333–346, 2003.

- R.Baldacci, N.Christofides, and A.Mingozzi. An exact algorithm for the vehicle routing problem based on the set partitioning formulation with additional cuts. *Mathematical Programming*, 115(2):351–385, 2008.
- R.Fukasawa, H.Longo, J.Lysgaard, M.Reis, E.Uchoa, and R.F.Werneck. Robust branch-and-cut-and-price for the capacitated vehicle routing problem. *Mathematical Programming*, 106(3):491–511, 1984.
- D.S Stoffer R.H.Shumway. Time series analysis and its applications. 2017.
- Jean-Yves PotvinGuy LapalmeJean-Marc Rousseau. Integration of ai and or techniques for computer-aided algorithmic design in the vehicle routing domain. *Journal of the Operational Research Society*, 41(6):517–525, 1990.
- P. M. Thompson and H. N. Psaraftis. Cyclic transfer algorithms for the multivehicle routing and scheduling problems. *Operations Research*, (41):935–946, 1993.
- W.R.Pulleyblank T.K.Ralphs, L.Kopman and L.E.TrotterJr. On the capacitated vehicle routing problem. *Mathematical Programming*, 94:343–359, 2003.
- M.Gendreau C.Prins T.Vidal, T.G.Crainic. A unified solution framework for multi-attribute vehicle routing problems. *European Journal of Operational Research*, 234(3):658–673, 2012.