



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

Σχολή Πολιτικών Μηχανικών

Τομέας Μεταφορών και Συγκοινωνιακής Υποδομής

Εκτίμηση Κυκλοφοριακού Φόρτου Αστικής Αρτηρίας Βάσει του Χρόνου Διαδρομής με Χρήση Μοντέλων Οδηγούμενων από Δεδομένα



Διπλωματική Εργασία

Χρήστος Κωνσταντινίδης

Επιβλέπουσα Καθηγήτρια : Ελένη Βλαχογιάννη,
Καθηγήτρια Σχολής Πολιτικών Μηχανικών ΕΜΠ

Αθήνα, Οκτώβριος 2023

Ευχαριστίες

Ολοκληρώνοντας τη παρούσα διπλωματική εργασία, θα ήθελα να ευχαριστήσω την κ. Ελένη Βλαχογιάννη, Καθηγήτρια της Σχολής Πολιτικών Μηχανικών του Εθνικού Μετσόβιου Πολυτεχνείου, για την ευκαιρία που μου προσέφερε να ασχοληθώ με ένα τόσο ενδιαφέρον θέμα και για την καθοδήγησή της σε όλα τα στάδια υλοποίησης του.

Στην συνέχεια θα ήθελα να ευχαριστήσω τον υποψήφιο Διδάκτορα του Εθνικού Μετσόβιου Πολυτεχνείου, Παναγιώτη Φαφουτέλλη για τον χρόνο, την προθυμία και τις καθοριστικές συμβουλές του σε όλο το διάστημα εκπόνησης της εργασίας.

Αθήνα, Οκτώβριος 2023

Τίτλος : Εκτίμηση Κυκλοφοριακού Φόρτου Αστικής Αρτηρίας Βάσει του Χρόνου Διαδρομής με Χρήση Μοντέλων Οδηγούμενων από Δεδομένα

Χρήστος Κωνσταντινίδης

Επιβλέπουσα Καθηγήτρια : Ελένη Ι. Βλαχογιάννη

Σύνοψη

Τα τελευταία χρόνια, η εξέλιξη της τεχνολογίας και η διαθεσιμότητα μεγάλου όγκου δεδομένων έχουν προσφέρει μια νέα διάσταση στον τρόπο διαχείρισης συστημάτων μεταφορών, ειδικά μέσω της πρόγνωσης κυκλοφοριακών συνθηκών. Σκοπό της παρούσας διπλωματικής εργασίας αποτελεί η ανάπτυξη προτύπων πρόβλεψης κυκλοφοριακών μεγεθών ενός οδικού τμήματος, όπως ο κυκλοφοριακός φόρτος και ο χρόνος διαδρομής. Τα πρότυπα αυτά βασίζονται σε αλγόριθμους μηχανικής μάθησης, οι οποίοι βρίσκουν συχνή εφαρμογή στη βιβλιογραφία σε προβλήματα πρόβλεψης. Τα δεδομένα χρόνου διαδρομής και φόρτων οχημάτων που χρησιμοποιήθηκαν στα μοντέλα, ελήφθησαν από οδικό τμήμα της Λ. Αλεξάνδρας στην Αθήνα. Σύμφωνα με τα αποτελέσματα, τα μοντέλα είναι ικανά για ακριβείς προβλέψεις με ικανοποιητική προσαρμογή στα δεδομένα. Τέλος, έγινε μία σύγκριση μεταξύ των μοντέλων μηχανικής μάθησης και κλασικών συναρτήσεων φόρτου-καθυστέρησης, όπως η συνάρτηση BPR. Ως συμπέρασμα, οι μέθοδοι μηχανικής μάθησης επέδειξαν μεγαλύτερη ικανότητα πρόβλεψης.

Λέξεις κλειδιά: μηχανική μάθηση, πρόβλεψη κυκλοφορίας, κυκλοφοριακός φόρτος, χρόνος διαδρομής, δέντρα απόφασης, GBDT, συνάρτηση BPR

Title: Urban Arterial Traffic Volume Estimation Based on Travel Time with Use of Data Driven Models

Chris Konstantinidis

Supervising Professor : Eleni I. Vlahogianni

Abstract

In recent years, the development of technology and the availability of large amounts of data have provided a new dimension to the way transport systems are managed, especially through the forecasting of traffic conditions. The purpose of this thesis is to develop models for predicting traffic variables in a road section, such as traffic volume and travel time. These models are based on machine learning algorithms, which find frequent application in the literature in forecasting problems. The data of travel time and traffic that were used in the models, were sourced from a road section of Alexandras Av. in Athens. According to the results, the models are capable of accurate predictions with an acceptable fit to the data. Finally, a comparison is made between machine learning models and volume-delay functions, such as the BPR function. In conclusion, machine learning methods showed greater forecasting capability.

Keywords: machine learning, traffic forecasting, traffic volume, travel time, decision trees, GBDT, BPR function

Περίληψη

Η ραγδαία εξέλιξη της τεχνολογίας σε συνδυασμό με την όλο και αυξανόμενη αστικοποίηση των τελευταίων ετών φέρουν την έννοια της βέλτιστης διαχείρισης οδικών δικτύων στο επίκεντρο της έρευνας των μεταφορικών συστημάτων. Οι σχέσεις μεταξύ των μεταβλητών ενός οδικού δικτύου, είναι δυνατό να αποκωδικοποιηθούν με την αξιοποίηση πληθώρας δεδομένων μέσω συστημάτων πληροφορικής. Βασικό στόχο της παρούσας διπλωματικής εργασίας αποτελεί η πρόβλεψη κυκλοφοριακών μεγεθών (κυκλοφοριακός φόρτος, χρόνος διαδρομής) ενός οδικού τμήματος με μοντέλα μηχανικής μάθησης. Παράλληλα, επιπρόσθετο στόχο αποτελεί η διερεύνηση της ικανότητας πρόβλεψης κυκλοφοριακών μεγεθών μέσω της συνάρτησης Φόρτου-Καθυστερήσης BPR, καθώς και η σύγκριση της με τα πρότυπα μηχανικής μάθησης που αναπτύχθηκαν.

Από τη διερεύνηση της σχετικής βιβλιογραφίας προέκυψε ότι οι μέθοδοι της μηχανικής μάθησης αποτελούν συχνά το προτιμώμενο εργαλείο πρόγνωσης κυκλοφοριακών συνθηκών. Πληθώρα μοντέλων μηχανικής μάθησης έχουν χρησιμοποιηθεί για αυτόν τον σκοπό, επιδεικνύοντας υψηλή ακρίβεια πρόβλεψης. Η πρόβλεψη χρόνου διαδρομής και κυκλοφοριακού φόρτου με απλά μοντέλα μηχανικής μάθησης (shallow learning) παρατηρείται συχνά στην βιβλιογραφία. Το ίδιο μπορεί να ειπωθεί αναφορικά με την εκτίμηση του χρόνου διαδρομής μέσω των συναρτήσεων φόρτου-καθυστερήσης.

Τα μοντέλα μηχανικής μάθησης μπορούν να ταξινομηθούν σε διάφορες κατηγορίες ανάλογα τον τύπο προβλήματος και την αρχιτεκτονική τους. Τα πρότυπα της παρούσας διπλωματικής εργασίας βασίστηκαν στους εξής αλγορίθμους:

- Δέντρα Απόφασης (Decision Trees)
- Ενισχυμένα Δέντρα Απόφασης (Gradient Boosting Decision Trees)

Ενώ η προτιμώμενη συνάρτηση φόρτου-καθυστερήσης ήταν η παρακάτω:

- Συνάρτηση Bureau of Public Roads (BPR function)

Η Λεωφόρος Αλεξάνδρας στην Αθήνα αποτέλεσε το οδικό τμήμα ενδιαφέροντος για την παρούσα διπλωματική. Από δύο σημεία της ανακτήθηκαν, μέσω επαγωγικών βρόχων (MS407, MS423), μετρήσεις κυκλοφοριακών φόρτων ανά μία ώρα για ογδόντα μη συνεχόμενες ημέρες. Οι αντίστοιχοι χρόνοι διαδρομής του παραπάνω οδικού τμήματος αντλήθηκαν από βάσεις δεδομένων της εφαρμογής Google Maps για τα ίδια χρονικά διαστήματα.

Συγκεκριμένα, για την ανάπτυξη των μοντέλων χρησιμοποιήθηκε η προγραμματιστική γλώσσα Python σε περιβάλλον Jupyter Notebooks. Αρχικά υπολογίστηκε ο συντελεστής συσχέτισης Pearson των μεταβλητών κυκλοφοριακού φόρτου και χρόνου διαδρομής, ο οποίος κρίθηκε ικανοποιητικός. Μετά την κατάλληλη επεξεργασία των δεδομένων και την προσθήκη χρήσιμων χαρακτηριστικών (features), διαμορφώθηκε η τελική βάση δεδομένων για την εκπαίδευση τους. Κατά τη διαδικασία προετοιμασίας έγινε τυχαίος διαχωρισμός των δειγμάτων με ποσοστό 80% για την εκπαίδευση των μοντέλων (train set) και 20% για τη δοκιμή τους (test set). Για την πρόβλεψη του χρόνου διαδρομής αξιοποιήθηκαν ως μεταβλητές εισόδου τα δεδομένα κυκλοφοριακού φόρτου των φωρατών MS407, MS423 σε συνδυασμό με άλλα χαρακτηριστικά. Τα μοντέλα μηχανικής μάθησης παρουσίασαν καλή προσαρμογή στα δεδομένα και χαμηλό σφάλμα πρόβλεψης ($R^2_{DT} = 0.82$, $MAPE_{DT} = 9.75\%$ και $R^2_{GBDT} = 0.76$, $MAPE_{GBDT} = 9.75\%$). Όσον αφορά την πρόβλεψη κυκλοφοριακού φόρτου χρησιμοποιήθηκαν παρόμοια χαρακτηριστικά σε συνδυασμό με τα δεδομένα χρόνου διαδρομής. Για την βελτιστοποίηση των μοντέλων μηχανικής μάθησης σε αυτή την περίπτωση προηγήθηκε η διαδικασία ομαλοποίησης της κλίμακας των δεδομένων φόρτου υπολογίζοντας τον λογάριθμο τους. Επιπρόσθετα, δεν λήφθηκαν υπόψη καταγραφές μεταμεσονύχτιων ωρών (00:00-3:00) για τον περιορισμό των ακραίων τιμών της βάσης δεδομένων. Για τον υπολογισμό του σφάλματος πρόβλεψης των κυκλοφοριακών φόρτων στα δύο σημεία τους οδικού τμήματος, πραγματοποιήθηκε η μετατροπή των τιμών στην αρχική τους κλίμακα. Τα μοντέλα μηχανικής παρουσίασαν ικανοποιητική ικανότητα πρόβλεψης και σε αυτή την περίπτωση ($R^2_{DT} = 0.79$, $MAPE_{DT} = 14.55\%$ και $R^2_{GBDT} = 0.80$, $MAPE_{GBDT} = 14.55\%$).

Για την χρήση της συνάρτησης BPR ως μέσο πρόβλεψης εφαρμόστηκε μια τεχνική βελτιστοποίησης της με την μέθοδο ελαχίστων τετραγώνων. Υπολογίστηκε ο μέσος όρος των φόρτων των δύο φωρατών της Λ. Αλεξάνδρας και καθορίστηκαν αρχικές τιμές των άγνωστων όρων της συνάρτησης ($\alpha = 0.15$, $\beta = 4$, $C = 1050$). Ο αλγόριθμος βελτιστοποίησης υπολόγισε την τελική τιμή των παραμέτρων α και β , ώστε να ελαχιστοποιείται το σφάλμα μεταξύ προβλεπόμενων τιμών της συνάρτησης και πραγματικών τιμών χρόνου διαδρομής. Οι τιμές των μέτρων αξιολόγησης της συνάρτησης ήταν οι εξής: $R^2_{BPR} = 0.48$, $MAPE_{BPR} = 16.78\%$.

Με την ανάπτυξη προτύπων πρόβλεψης ικανών για ακριβείς προβλέψεις επιτεύχθηκε ο βασικό σκοπός της διπλωματικής εργασίας. Παράλληλα, διαπιστώθηκε ότι οι σύγχρονες μέθοδοι μηχανικής μάθησης εμφανίζουν καλύτερη απόδοση σε σχέση με την συνάρτηση BPR ως εργαλείο πρόβλεψης. Επίσης, η ακριβής πρόβλεψη τιμών κυκλοφοριακού φόρτου, του οποίου η άντληση πραγματικών δεδομένων παρουσιάζει δυσκολίες, μπορεί να αποτελέσει σημαντικό εργαλείο έρευνας των μεταφορικών συστημάτων. Με την

ενσωμάτωση περισσότερων τύπων δεδομένων και τη περαιτέρω εξερεύνηση των δυνατοτήτων της μηχανικής μάθησης, θα μπορούσε να αποδοθεί μία ακριβέστερη εικόνα των συνθηκών ενός οδικού δικτύου ανά πάσα στιγμή. Ως γενικό συμπέρασμα εξάγεται ότι, τα οφέλη για τον χρήστη του οδικού δικτύου, και κατ' επέκταση για το κοινωνικό σύνολο πολλαπλασιάζονται συνεχώς με την ανάπτυξη καινοτόμων τεχνολογικά εφαρμογών διαχείρισης μεταφορικών συστημάτων.

Περιεχόμενα

Κεφάλαιο 1	1
Εισαγωγή.....	1
1.1 Το Πρόβλημα της Διαχείρισης Μεταφορικών Συστημάτων.....	1
1.2 Σκοπός της Διπλωματικής Εργασίας.....	2
1.3 Διάρθρωση της Διπλωματικής Εργασίας	4
Κεφάλαιο 2	6
Βιβλιογραφική Ανασκόπηση	6
2.1 Εισαγωγή.....	6
2.2 Μεθοδολογίες Πρόβλεψης Κυκλοφοριακών Μεγεθών.....	6
2.3 Έρευνες Πρόβλεψης Κυκλοφοριακού Φόρτου και Χρόνου Διαδρομής.....	11
2.4 Εφαρμογές της Συνάρτησης BPR	15
2.5 Συμπεράσματα Βιβλιογραφίας	18
Κεφάλαιο 3	21
Μεθοδολογική Προσέγγιση και Θεωρητικό Υπόβαθρο	21
3.1 Εισαγωγή.....	21
3.2 Μηχανική μάθηση.....	21
3.2.1 Περιγραφή Αλγορίθμων Μηχανικής Μάθησης και Συμπληρωματικές Έννοιες.	23
3.2.2 Δέντρα Απόφασης.....	26
3.2.3 Ενισχυμένα Δέντρα Απόφασης.....	28
3.3 Συναρτήσεις Φόρτου-Καθυστέρησης και BPR.....	30
3.3.1 Μέθοδος ελαχίστων τετραγώνων	31
Κεφάλαιο 4	33
Συλλογή και επεξεργασία στοιχείων	33
4.1 Εισαγωγή.....	33
4.2 Συλλογή στοιχείων	33
4.3 Διαμόρφωση δεδομένων.....	35
4.4 Επεξεργασία δεδομένων.....	36
4.5 Αποτελέσματα	42
Κεφάλαιο 5	53
Συμπεράσματα και Προτάσεις.....	53

5.1 Σύνοψη Αποτελεσμάτων	53
5.2 Βασικά Συμπεράσματα	54
5.3 Περιορισμοί Διπλωματικής Εργασίας και Προτάσεις Έρευνας	55
Βιβλιογραφία	57

Ευρετήριο Εικόνων

Εικόνα 1: Στάδια και ροή ανάπτυξης μοντέλων στα I.T.S. (Πηγή: Laña et al., 2021)	4
Εικόνα 2: Λογισμικά μεθόδων καθοδηγούμενες από μοντέλα.....	7
Εικόνα 3: Παραδείγματα μοντέλων μηχανικής μάθησης ανά κατηγορία (Πηγή: Manibardo et al., 2022).....	9
Εικόνα 4: Τα βασικά στοιχεία της πρόβλεψης της κυκλοφορίας (Πηγή: Z. Liu et al., 2018).....	12
Εικόνα 5: Κατηγοριοποίηση των μεθόδων μηχανικής μάθησης (Πηγή: Yuan et al., 2022)	23
Εικόνα 6: Δομή Δέντρου Απόφασης (Πηγή: https://www.smartdraw.com/decision-tree/)	27
Εικόνα 7: Δομή Ενισχυμένων Δέντρων Απόφασης (GBDT) (Πηγή: Deng et al., 2021)	29
Εικόνα 8: Οδικό τμήμα υπό μελέτη της Λ. Αλεξάνδρας	33

Ευρετήριο Εικόνων

Διάγραμμα 1: Θεμελιώδες διάγραμμα ταχύτητας-φόρτου (Πηγή: ΚΥΚΛΟΦΟΡΙΑΚΗ ΤΕΧΝΙΚΗ I. Φραντζεσκάκης - I. Γκόλιας - Μ. Πιτσιάβα-Λατινοπούλου)	17
Διάγραμμα 2: Διάγραμμα ταχύτητας-φόρτου της συνάρτησης BPR.....	18
Διάγραμμα 3: Διάγραμμα φόρτου-χρονικής καθυστέρησης της συνάρτησης BPR (Πηγή: Maerivoet & De Moor, 2005)	31
Διάγραμμα 4: Θερμικός χάρτης συσχέτισης φόρτου-χρόνου διαδρομής	37
Διάγραμμα 5: Διάγραμμα διασποράς φόρτου-χρόνου διαδρομής.....	38
Διάγραμμα 6: Πλήθος παρατηρήσεων διαφορετικών τιμών χρόνου διαδρομής για κάθε ώρα της ημέρας.....	42
Διάγραμμα 7: Διάγραμμα διασποράς πραγματικών και προβλεπόμενων τιμών κυκλοφοριακού φόρτου (Dec.Trees)	44
Διάγραμμα 8: Διάγραμμα κατανομής εναπομένουστος σφάλματος τιμών κυκλοφοριακού φόρτου (Dec.Trees)	44
Διάγραμμα 9: Διάγραμμα διασποράς πραγματικών και προβλεπόμενων τιμών κυκλοφοριακού φόρτου (GBDT).....	45
Διάγραμμα 10: Διάγραμμα κατανομής εναπομένουστος σφάλματος τιμών κυκλοφοριακού φόρτου (GBDT).....	46
Διάγραμμα 11: Διάγραμμα διασποράς πραγματικών και προβλεπόμενων τιμών χρόνου διαδρομής (Dec.Trees).....	47
Διάγραμμα 12: Διάγραμμα κατανομής εναπομένουστος σφάλματος τιμών χρόνου διαδρομής (Dec.Trees).....	47
Διάγραμμα 13: Διάγραμμα διασποράς πραγματικών και προβλεπόμενων τιμών χρόνου διαδρομής (GBDT).....	48
Διάγραμμα 14: Διάγραμμα κατανομής εναπομένουστος σφάλματος τιμών χρόνου διαδρομής (GBDT)	48
Διάγραμμα 15: Διάγραμμα διασποράς πραγματικών και προβλεπόμενων τιμών χρόνου διαδρομής (BPR).....	51
Διάγραμμα 16: Διάγραμμα κατανομής εναπομένουστος σφάλματος τιμών χρόνου διαδρομής (BPR)	52

Ευρετήριο Πινάκων

Πίνακας 1 : Παράγοντες που επηρεάζουν την κυκλοφοριακή ικανότητα (Πηγή: ENNOIES ΒΑΣΙΚΩΝ ΠΑΡΑΜΕΤΡΩΝ ΚΥΚΛΟΦΟΡΙΑΣ Γ. Γιαννής - Ι. Γκόλιας - Γ. Κανελλαΐδης)	16
Πίνακας 2 : Ημερομηνίες καταγραφής δεδομένων	34
Πίνακας 3 : Στατιστικά μέτρα δεδομένων κυκλοφοριακού φόρτου	34
Πίνακας 4 : Στατιστικά μέτρα δεδομένων χρόνου διαδρομής.....	35
Πίνακας 5 : Βάση δεδομένων	35
Πίνακας 6 : Χαρακτηριστικά (features) μοντέλων μηχανικής μάθησης	39
Πίνακας 7 : Βάση δεδομένων μετά την ομαλοποίηση	40
Πίνακας 8 : Στατιστικά μέτρα μέσου κυκλοφοριακού φόρτου ανά τιμή χρόνου διαδρομής.....	41
Πίνακας 9 : Z-score μέσου κυκλοφοριακού φόρτου ανά τιμή χρόνου διαδρομής.....	41
Πίνακας 10 : Υπερ-παράμετροι (hyper-parameters) μοντέλων μηχανικής μάθησης	43
Πίνακας 11 : Αποτελέσματα πρόβλεψης κυκλοφοριακού φόρτου με Δέντρα Απόφασης.....	43
Πίνακας 12 : Αποτελέσματα πρόβλεψης κυκλοφοριακού φόρτου με GBDT.....	45
Πίνακας 13 : Αποτελέσματα πρόβλεψης χρόνου διαδρομής με Δέντρα Απόφασης.....	46
Πίνακας 14 : Αποτελέσματα πρόβλεψης χρόνου διαδρομής με GBDT	48
Πίνακας 15 : Βάση δεδομένων μέσου κυκλοφοριακού φόρτου	49
Πίνακας 16 : Αποτελέσματα πρόβλεψης χρόνου διαδρομής με την συνάρτηση BPR.....	51
Πίνακας 17 : Συγκεντρωτικός πίνακας αποτελεσμάτων	53

Κεφάλαιο 1

Εισαγωγή

1.1 Το Πρόβλημα της Διαχείρισης Μεταφορικών Συστημάτων

Στη σύγχρονη εποχή, ο συνωστισμός του πληθυσμού σε αστικά κέντρα σε συνδυασμό με την όλο και αυξανόμενη κινητικότητά του με Ι.Χ. έχει οδηγήσει σε δυσεπίλυτα προβλήματα για τα δίκτυα μεταφορών. Η απαιτούμενη ζήτηση είναι δύσκολο να ικανοποιηθεί σε βραχεία χρονική κλίμακα, καθώς μόνιμες παρεμβάσεις σε συστήματα μεταφορών είναι χρονοβόρες και συνεπάγονται μεγάλο κόστος. Επομένως, για την αντιμετώπιση του προβλήματος της αυξανόμενης ζήτησης κρίνεται αναγκαία η βέλτιστη διαχείριση και αξιοποίηση των υπάρχοντων δικτύων μεταφορών.

Η εξέλιξη της τεχνολογίας και των συστημάτων πληροφορικής έχουν επιφέρει την ανάπτυξη των Ευφυών Συστημάτων Μεταφορών. (Intelligent Transport Systems – I.T.S.). Το Υπουργείο Μεταφορών των Η.Π.Α ορίζει τα I.T.S. ως την εφαρμογή τεχνολογιών ανίχνευσης, επικοινωνίας και πληροφορίας σε παραδοσιακά συστήματα με σκοπό την επίτευξη μεγαλύτερης κινητικότητας και ασφάλειας. Τα I.T.S. παρέχουν νέα πληθώρα δεδομένων, τα οποία αποτελούν χρήσιμη γνώση για τα κυκλοφοριακά χαρακτηριστικά των οδικών δικτύων ανά πάσα στιγμή. Τα συνήθη κυκλοφοριακά μεγέθη: ταχύτητα, πυκνότητα και φόρτος κυκλοφορίας, σε συνδυασμό με παράγοντες όπως: η γεωμετρία του οδικού δικτύου, η σήμανση, τα καιρικά φαινόμενα, η οδηγική συμπεριφορά, πιθανά ατυχήματα καθώς και η εκάστοτε ώρα και ημερομηνία παρατήρησης της κυκλοφορίας δημιουργούν ένα πολυσύνθετο περιβάλλον μελέτης. Η αξιοποίηση των αναδυόμενων δυνατοτήτων των I.T.S. για την ανάπτυξη εφαρμογών κυκλοφοριακής τεχνικής, συντελεί στην αντιμετώπιση του παραπάνω προβλήματος. Είναι προφανές λοιπόν, ότι η δυνατότητα πρόγνωσης συνθηκών κυκλοφορίας μπορεί να αποτελέσει εργαλείο ύψιστης σημασίας για την βέλτιστη διαχείριση των μεταφορικών συστημάτων, και κατά συνέπεια την αύξηση της οδικής ασφάλειας και της κινητικότητας ενός οδικού δικτύου.

Η πρόγνωση μιας μελλοντικής κατάστασης με βάση την ανάλυση υπάρχοντων δεδομένων αποτελεί βασική μέθοδο για την λήψη μέτρων και αποφάσεων σε όλο το φάσμα της ανθρώπινης δραστηριότητας. Όσον αφορά τον τομέα των μεταφορών, σημαντικό μέρος της επιστημονικής αναζήτησης επικεντρώνεται στην εξεύρεση της βέλτιστης μεθοδολογίας πρόβλεψης μέσω δοκιμών και συγκρίσεων. Ο κλάδος της Μηχανικής Μάθησης (Machine Learning) έχει γνωρίσει μεγάλη άνθιση με εφαρμογή σε αρκετά επιστημονικά πεδία. Η δημιουργία μοντέλων μηχανικής μάθησης μπορούν να παράγουν ακριβείς προβλέψεις μέσω του εντοπισμού μοτίβων μεταξύ μεγάλων όγκων δεδομένων. Ένα ακόμα εργαλείο πρόβλεψης κυκλοφοριακών μεγεθών αποτελούν οι

Συναρτήσεις Φόρτου-Καθυστέρησης (Volume-Delay Functions). Πρόκειται για μαθηματικές συναρτήσεις οι οποίες εκφράζουν το χρόνο διαδρομής ενός οδικού τμήματος συναρτήσει του κυκλοφοριακού φόρτου. Περισσότερα σχετικά με την Μηχανική Μάθηση και τις Συναρτήσεις Φόρτου-Καθυστέρησης θα ακολουθήσουν σε επόμενο κεφάλαιο.

1.2 Σκοπός της Διπλωματικής Εργασίας

Βασικό στόχο της παρούσας διπλωματικής εργασίας αποτελεί η ανάπτυξη μοντέλων ικανών να παράγουν ακριβείς προβλέψεις κυκλοφοριακών συνθηκών σε ένα οδικό τμήμα. Τα κυκλοφοριακά μεγέθη για το οποία έγινε η απόπειρα πρόγνωσης ήταν αυτά του κυκλοφοριακού φόρτου και του χρόνου διαδρομής.

- *Κυκλοφοριακός φόρτος είναι ο συνολικός αριθμός οχημάτων που περνούν από τη διατομή μιας λωρίδας ή οδού κατά τη διάρκεια ενός δεδομένου χρονικού διαστήματος (συνήθως οχήματα ανά ώρα).*
- *Χρόνος διαδρομής ορίστηκε ως το χρονικό διάστημα (λεπτά) που απαιτείται για την διάνυση του υπό μελέτη οδικού τμήματος.*

Συγκεκριμένα, αναπτύχθηκαν αλγόριθμοι μηχανικής μάθησης, οι οποίοι αξιοποιώντας ιστορικά δεδομένα χρόνου διαδρομής ενός οδικού τμήματος, επιχείρησαν την πρόβλεψη του κυκλοφοριακού φόρτου για το αντίστοιχο χρονικό διάστημα. Η αντίστροφη διαδικασία πραγματοποιήθηκε για την πρόβλεψη του χρόνου διαδρομής στο οδικό τμήμα. Για την ανάπτυξη των μοντέλων πρόβλεψης επιλέχθηκαν δύο αλγόριθμοι εποπτευόμενης μηχανικής μάθησης (supervised machine learning):

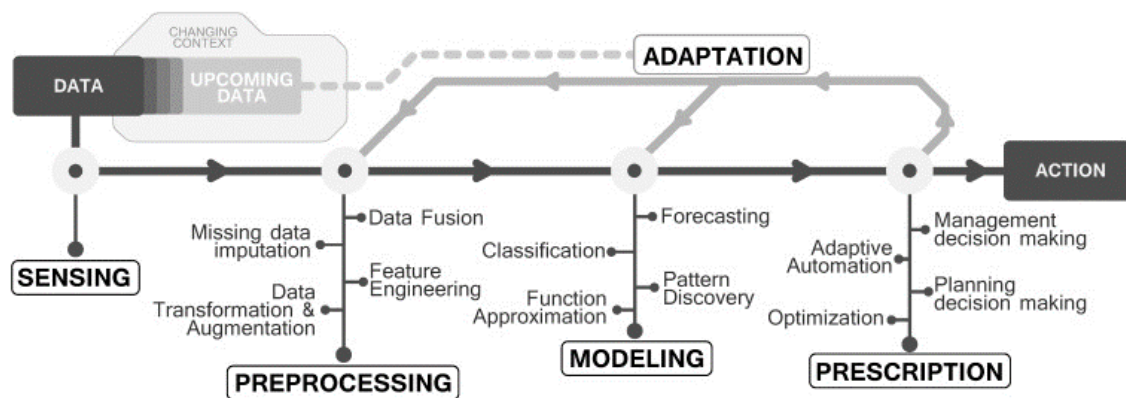
- Δέντρα Απόφασης (Decision Trees)
- Ενισχυμένα Δέντρα Απόφασης (Gradient Boosting Decision Trees - GBDT).

Τα μοντέλα αυτά τροφοδοτήθηκαν με πραγματικά δεδομένα χρόνων διαδρομής από την εφαρμογή Google maps και κυκλοφοριακών φόρτων από επαγγελματίες βρόχους επί της Λ. Αλεξάνδρας στην Αθήνα και αξιολογήθηκε τελικά η ικανότητα πρόβλεψής τους.

Επιπρόσθετο στόχο της διπλωματικής εργασίας αποτελεί η σύγκριση των καινοτόμων μεθόδων μηχανικής μάθησης με συμβατικές μεθόδους συσχέτισης κυκλοφοριακών μεγεθών, όπως οι μαθηματικές συναρτήσεις φόρτου-καθυστέρησης. Η επιστημονική βιβλιογραφία της πρόβλεψης κυκλοφοριακών συνθηκών έχει επικεντρωθεί τα τελευταία χρόνια στην χρήση καινοτόμων τεχνολογικά μεθόδων, όπως αυτές της μηχανικής μάθησης. Αυτή η σύγκριση των μεθόδων του παρελθόντος και του μέλλοντος, διασαφηνίζει το αν, και κατά πόσο έχει βελτιωθεί η ικανότητα πρόβλεψης στην εποχή της ψηφιακής επανάστασης.

Η συνάρτηση Bureau of Public Roads (BPR function) αποτελεί μία από τις πιο διαδεδομένες συναρτήσεις φόρτου-καθυστερήσης και επιλέχθηκε ως εργαλείο πρόβλεψης του χρόνου διαδρομής. Η παραπάνω συνάρτηση εφαρμόζεται αποκλειστικά σε οδικά τμήματα και απαιτεί ανάθεση τιμών στις παραμέτρους της για την παραγωγή προβλέψεων. Επομένως, αναπτύχθηκε αλγόριθμος ο οποίος με την μέθοδο ελαχίστων τετραγώνων ορίζει της παραμέτρους της συνάρτησης, ώστε να ελαχιστοποιείται το σφάλμα των προβλέψεων. Ως ορίσματα εισόδου της συνάρτησης χρησιμοποιήθηκαν τα ίδια δεδομένα κυκλοφοριακού φόρτου με τα μοντέλα μηχανικής μάθησης.

Ένα τελευταίο ερώτημα που καλούνται να απαντήσουν οι παραπάνω μέθοδοι είναι το κατά πόσο θεωρείται δυνατή η επίτευξη ικανοποιητικής πρόβλεψης για τη μία μεταβλητή έχοντας ως μοναδική γνώση δεδομένα της δεύτερης. Το όφελος αυτής της αναζήτησης δικαιολογείται μετά από την παρατήρηση των δεδομένων των δύο κυκλοφοριακών μεγεθών υπό εξέταση. Όπως αναφέρθηκε και προηγουμένως, η τεχνολογική εξέλιξη οδήγησε στην πληθώρα διαθέσιμων κυκλοφοριακών δεδομένων αναγκαίων για την βέλτιστη διαχείριση μεταφορικών συστημάτων. Το παραπάνω γεγονός είναι ιδιαίτερα εμφανές στην περίπτωση της μεταβλητής του χρόνου διαδρομής. Ιστορικά δεδομένα χρόνων διαδρομής είναι άμεσα διαθέσιμα, ευκόλως προσβάσιμα με σχεδόν ολοκληρωτική κάλυψη του οδικού δικτύου. Η συλλογή τους γίνεται συνήθως μέσω συσκευών GPS, εφαρμογών κινητών τηλεφώνων ή άλλων υπηρεσιών εντοπισμού τοποθεσίας με πολύ μικρό κόστος. Αυτές οι πηγές είναι συχνά διαθέσιμες ανά πάσα στιγμή από εμπορικές εταιρίες ή εφαρμογές πλοήγησης. Από την άλλη πλευρά, η συλλογή δεδομένων κυκλοφοριακού φόρτου παρουσιάζει δυσκολίες. Για την άντληση δεδομένων είναι απαραίτητη η εγκατάσταση υποδομών όπως οι κάμερες κυκλοφορίας και οι επαγωγικοί βρόχοι οδοστρώματος. Η κάλυψη μεγάλου μέρους του οδικού δικτύου με τον απαραίτητο εξοπλισμό καταγραφής συνεπάγεται μεγαλύτερο κόστος. Επίσης, η πρόσβαση στα αποτελέσματα καταγραφών του εξοπλισμού αυτού είναι σαφώς δυσκολότερη, καθώς στις περισσότερες περιπτώσεις η διαχείριση τους γίνεται αποκλειστικά από δημόσιες αρχές. Για τους παραπάνω λόγους, η αμφίδρομη εξαγωγή αποτελεσμάτων πρόβλεψης μεταξύ των δύο αυτών κυκλοφοριακών μεταβλητών μπορεί να προσφέρει χρήσιμη πληροφορία στα μεταφορικά συστήματα.



Εικόνα 1: Στάδια και ροή ανάπτυξης μοντέλων στα I.T.S. (Πηγή: Laña et al., 2021)

1.3 Διάρθρωση της Διπλωματικής Εργασίας

Η δομή της διπλωματικής εργασία ανά κεφάλαιο έχει ως εξής:

Στο Κεφάλαιο 1, παρουσιάστηκε το πρόβλημα διαχείρισης που αντιμετωπίζουν τα σύγχρονα συστήματα μεταφορών και τονίστηκε η αναγκαιότητα της εφαρμογής μεθόδων πρόβλεψης των κυκλοφοριακών χαρακτηριστικών. Επίσης έγινε μια επιγραμματική αναφορά στην μηχανική μάθηση και τις συναρτήσεις φόρτου-καθυστερήσης, η οποίες αποτελούν τις μεθόδους πρόβλεψης της παρούσας διπλωματικής εργασίας.

Στο Κεφάλαιο 2, πραγματοποιείται βιβλιογραφική ανασκόπηση σχετικά με το αντικείμενο της πρόβλεψης κυκλοφοριακών μεγεθών. Παρουσιάζονται σχετικές δημοσιευμένες μελέτες, γίνεται μια σύντομη αναφορά στις μεθόδους, τα δεδομένα που χρησιμοποιήθηκαν, καθώς και τα συμπεράσματα από το σύνολο της βιβλιογραφίας. Επιπλέον, παρουσιάζονται παραδείγματα μελέτης και εφαρμογής της συνάρτησης BPR στην επιστημονική βιβλιογραφία.

Στο Κεφάλαιο 3, παρουσιάζεται το θεωρητικό υπόβαθρο της συνάρτησης BPR και των αλγορίθμων μηχανικής μάθησης που επιλέχθηκαν. Περιγράφονται η μεθοδολογική προσέγγιση για την επίτευξη του στόχου της διπλωματικής εργασίας καθώς και άλλες χρήσιμες συμπληρωματικές έννοιες.

Στο Κεφάλαιο 4, γίνεται περιγραφή της διαδικασίας συλλογής δεδομένων και η επεξεργασία τους ώστε να γίνει δυνατή η χρήση τους από τα μοντέλα. Περιγράφεται λεπτομερώς η μεθοδολογία που ακολουθήθηκε και παρουσιάζονται τα αποτελέσματα της.

Στο Κεφάλαιο 5, παρουσιάζονται τα βασικά συμπεράσματα της ανάλυσης που προηγήθηκε και γίνονται προτάσεις για περαιτέρω έρευνα.

Στο Κεφάλαιο 6, παρατίθεται η σχετική βιβλιογραφία που χρησιμοποιήθηκε για την εκπόνηση της διπλωματικής εργασίας.

Κεφάλαιο 2

Βιβλιογραφική Ανασκόπηση

2.1 Εισαγωγή

Με ιστορία τεσσάρων δεκαετιών, το αντικείμενο της πρόβλεψης κυκλοφοριακών μεγεθών καταλαμβάνει πλέον σημαντικό ποσοστό της επιστημονικής βιβλιογραφίας του τομέα μεταφορών. Πληθώρα μεθοδολογιών έχουν αναπτυχθεί και παρουσιαστεί ως λύση του προβλήματος της πρόβλεψης. Χάρη στην συμβολή των πρόσφατων τεχνολογικών εξελίξεων στον τομέα της πληροφορικής, μεθοδολογίες πρόγνωσης βασισμένες στην μηχανική μάθηση έχουν προσελκύσει σε μεγάλο βαθμό το ενδιαφέρον της επιστημονικής κοινότητας τα τελευταία χρόνια. Στο κεφάλαιο αυτό αναφέρονται οι διάφορες μεθοδολογίες που συναντώνται στην βιβλιογραφία καθώς και οι ευρύτερες κατηγορίες στις οποίες ανήκουν. Κατόπιν, παρουσιάζονται έρευνες σχετικές με την πρόβλεψη χρόνου διαδρομής και κυκλοφοριακού φόρτου μέσω μεθόδων μηχανικής μάθησης, καθώς και την εφαρμογή της συνάρτησης BPR.

2.2 Μεθοδολογίες Πρόβλεψης Κυκλοφοριακών Μεγεθών

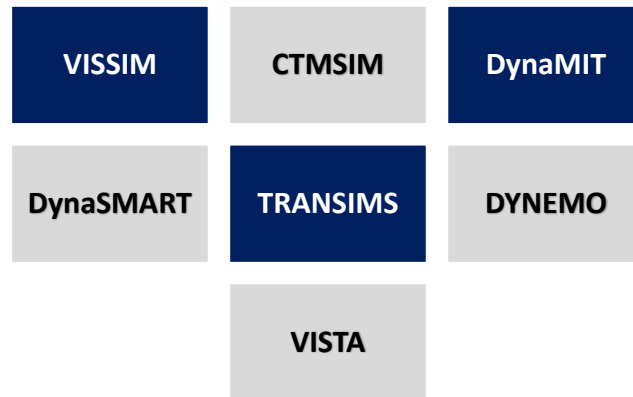
Η επιλογή κατάλληλης μεθοδολογίας αποτελεί σημαντική πρόκληση ως προς την υλοποίηση μιας τεχνικής πρόβλεψης. Η δυσκολία σύγκρισης των διαφόρων προτύπων μηχανικής μάθησης που έχουν εφαρμοστεί τα τελευταία χρόνια δικαιολογεί τον παραπάνω ισχυρισμό (Vlahogianni et al., 2014). Παρότι αντικείμενο της εργασίας αποτελεί η ανάπτυξη προτύπων πρόβλεψης με κλασσικούς αλγόριθμους μηχανικής μάθησης (shallow learning), μια επιγραμματική αναφορά γίνεται και σε άλλα γνωστά μοντέλα πρόβλεψης. Επίσης αναφορά γίνεται και στις ευρύτερες κατηγορίες στις οποίες θα μπορούσαν να ενταχθούν.

Υπάρχουν διάφοροι τρόποι με τους οποίους θα μπορούσαν να ταξινομηθούν οι υπάρχουσες μέθοδοι πρόβλεψης, όμως ένα εύχρηστο σύστημα ταξινόμησης είναι το εξής (IEEE Staff, 2018):

1. Μέθοδοι καθοδηγούμενες από μοντέλα ή δεδομένα
2. Παραμετρικές ή μη-παραμετρικές μέθοδοι
3. Άμεσες ή έμμεσες μέθοδοι πρόβλεψης

Μέθοδοι καθοδηγούμενες από μοντέλα

Οι μέθοδοι καθοδηγούμενες από μοντέλα αποσκοπούν στη δημιουργία λεπτομερών προσομοιώσεων του οδικού δικτύου με τη βοήθεια ηλεκτρονικών υπολογιστών. Μερικά στοιχεία των προσομοιώσεων αποτελούν οι διασταυρώσεις, οι λωρίδες κυκλοφορίας, οι φωτεινοί σηματοδότες και τα όρια ταχύτητας μεταξύ άλλων. Αφού έχουν οριστεί οι αρχικές συνθήκες κυκλοφορίας, κατά την προσομοίωση του συστήματος πεζοί και οχήματα αλληλοεπιδρούν μεταξύ τους σύμφωνα με τα χαρακτηριστικά του κατασκευασμένου δικτύου. Αναλύοντας τις κυκλοφοριακές συνθήκες σε μια συγκεκριμένη περιοχή του συστήματος, μπορεί να γίνει πρόβλεψη για μελλοντική κατάσταση του οδικού δικτύου. Ως εκ τούτου, μπορεί να εξαχθεί πληροφορία σχετικά με την μεταβολή των κυκλοφοριακών μεγεθών.



Εικόνα 2: Λογισμικά μεθόδων καθοδηγούμενες από μοντέλα

Ένα μειονέκτημα αυτών των μεθόδων είναι η κατασκευή μοντέλων που δεν περιγράφουν πλήρως την συσχέτιση των κυκλοφοριακών μεγεθών. Αυτό πιθανότατα καταλογίζεται στην αβεβαιότητα και την πολυπλοκότητα του φαινομένου της κυκλοφορίας, καθώς και την απουσία συνεχούς τροφοδότησης δεδομένων πραγματικού χρόνου του μοντέλου σε μακροπρόθεσμες προβλέψεις (Barros et al., 2015). Επίσης, σύμφωνα πάλι με τους (Barros et al., 2015), σε αντίθεση με τις μεθόδους καθοδηγούμενες από δεδομένα, τα μοντέλα προσομοιώσεων δέχονται συγκεκριμένους τύπους πληροφορίας. Αυτοί συνήθως περιορίζονται σε παραμέτρους κυκλοφορίας όπως η ροή, ταχύτητα, κατάληψη κτλ. Χρήσιμη πληροφορία όμως μπορεί να αντληθεί και από καιρικά δεδομένα, στατιστικά ατυχημάτων αλλά και καταχωρημένους χρόνους διαδρομής. Οι μέθοδοι καθοδηγούμενες από μοντέλα συχνά αγνοούν την πολύ-παραγοντική φύση του φαινομένου της κυκλοφοριακής ροής (Behrooz & Hayeri, 2022).

Μέθοδοι καθοδηγούμενες από δεδομένα

Ως μέτρο καταπολέμησης των παραπάνω αδυναμιών, η ερευνητική κοινότητα έστρεψε το ενδιαφέρον της στα μοντέλα καθοδηγούμενα από δεδομένα. Χάρη στην αυξανόμενη διαθεσιμότητα δεδομένων, οι εν λόγω μέθοδοι βρήκαν πρόσφορο έδαφος να εδραιωθούν ως την αιχμή του δόρατος στην βιβλιογραφία των τεχνικών προβλέψεων. Σημαντικό ρόλο σε αυτό συνετέλεσαν οι νέες τεχνολογίες άντλησης κυκλοφοριακών δεδομένων. Η εξάπλωση έξυπνων συσκευών, όπως τα σύγχρονα κινητά τηλέφωνα (GPS) και η εκσυγχρονισμός των οχημάτων με συστήματα τηλεματικής (in-vehicle telematics), ήρθαν να προστεθούν στις υπάρχουσες υποδομές ανίχνευσης δεδομένων, όπως οι ηλεκτρομαγνητικοί βρόχοι, οι κάμερες παρακολούθησης οδικού δικτύου και οι σταθμοί διοδίων. Παράλληλα, η τεχνολογική ανάπτυξη συνέβαλε στην αύξηση του όγκου και της ακρίβειας καιρικών δεδομένων, την καταγραφή και ανίχνευση οδικών ατυχημάτων κ.α. Σε αυτή τη κατηγορία μεθόδων ανήκουν και τα μοντέλα μηχανικής μάθησης. Εκτενέστερη αναφορά στο θεωρητικό υπόβαθρο του κλάδου της μηχανικής μάθησης θα γίνει στο Κεφάλαιο 3. Ωστόσο, αξίζει να γίνει διάκριση των κυριότερων υποκατηγοριών αλγορίθμων μηχανικής μάθησης που παρατηρούνται στην βιβλιογραφία.

Βασικό κριτήριο διάκρισης των εν λόγω υποκατηγοριών αποτελεί η αρχιτεκτονική των αλγορίθμων. Οι αλγόριθμοι βαθιάς μάθησης (deep learning) αποτελούνται από πολλαπλά επίπεδα επεξεργασίας (processing layers), σε αντίθεση με τους τυπικούς αλγόριθμους μηχανικής μάθησης (shallow learning). Ενώ, με την έννοια της συλλογικής μάθησης (ensemble learning) χαρακτηρίζονται αλγόριθμοι, οι οποίοι συνδυάζουν πολλαπλά μοντέλα των δύο προηγούμενων υποκατηγοριών.

Shallow Learning	Ensemble Learning	Deep Learning
<ul style="list-style-type: none"> •Least-squares Linear Regression [LR] •k Nearest Neighbors [KNN], Decision Tree [DTR] •Extreme Learning Machine [ELM] •Support Vector Machine [SVR]. 	<ul style="list-style-type: none"> •Adaboost [ADA] •Random Forest [RFR] •Extremely Randomized Trees [ETR] •Gradient Boosting [GBR] •Extreme Gradient Boosting [XGBR]. 	<ul style="list-style-type: none"> •Feed Forward Neural Network [FNN] •Convolutional Neural Network [CNN] •Recurrent Neural Network based on LSTM units [LSTM] •Convolutional-Recurrent Neural Network [CLSTM] •Attention mechanismbased Auto-encoder with Convolutional input layers [ATT]

Εικόνα 3: Παραδείγματα μοντέλων μηχανικής μάθησης ανά κατηγορία (Πηγή: Manibardo et al., 2022)

Σημαντική διαφορά μεταξύ των υποκατηγοριών deep και shallow learning δεν βρίσκεται μόνο στην πολυπλοκότητα της αρχιτεκτονικής τους, αλλά και στον τρόπο μοντελοποίησης των δεδομένων που αυτή παρέχει. Οι αλγόριθμοι βαθιάς μάθησης είναι ικανοί να εξάγουν τα υποκείμενα χαρακτηριστικά (feature extraction) των κυκλοφοριακών συνθηκών από πλήθος δεδομένων και, σε συνδυασμό με τα ορίσματα εισόδου, να παράγουν ακριβείς προβλέψεις. Στην περίπτωση των αλγορίθμων (shallow learning) αυτή η διαδικασία γίνεται χειροκίνητα (feature selection), το οποίο σε αρκετές περιπτώσεις επιδρά στην ακρίβεια πρόβλεψης (Z. Liu et al., 2018). Για αυτό τον λόγο, η έξυπνη επιλογή (feature selection) ή ακόμα και η κατασκευή (feature engineering) χρήσιμων χαρακτηριστικών για έναν αλγόριθμο shallow learning μπορεί να αποτελέσει σημαντικό μέτρο βελτίωσης της απόδοσής τους (Behrooz & Hayeri, 2022).

Παρότι οι αλγόριθμοι βαθιάς μάθησης συχνά παρουσιάζουν καλύτερες επιδόσεις σε προβλήματα πρόβλεψης, δεν παύουν να έχουν μειονεκτήματα (Wang et al., 2019). Οι υψηλές απαιτήσεις σε όγκο δεδομένων και υπολογιστική ισχύ αποτελούν εμπόδιο στην χρήση τέτοιων μεθόδων, ιδιαίτερα αν ληφθεί υπόψιν ο παράγοντας του χρόνου για την εφαρμογή τους. Ιδιαίτερα δύσκολη κρίνεται και η ρύθμιση των υπερ-παραμέτρων (hyper-parameters) τους. Απαιτείται υψηλή τεχνογνωσία για την ρύθμιση τους με στόχο την μείωση του χρόνου ανάπτυξής τους, ο οποίος μπορεί να διαρκέσει μέρες ή εβδομάδες ανάλογα τον όγκο δεδομένων.

Τέλος, η ικανότητα τους να εξάγουν αυτόματα τα υποκείμενα χαρακτηριστικά των δεδομένων (feature extraction) έρχεται σε αντίθεση τελικά με την ευκολία ερμηνείας

των αποτελεσμάτων τους. Η λογική «μαύρου κουτιού» (black box nature) που τους χαρακτηρίζει, οφείλεται στην ύπαρξη πολλών κρυφών επεξεργαστικών επιπέδων (hidden layers) τα οποία στερούν από τον χρήστη πλήρη κατανόηση των αποτελεσμάτων τους. Επομένως, σύμφωνα με τα παραπάνω μειονεκτήματα αλλά και την ανάλυση των (Manibardo et al., 2022), η χρήση ενός αλγορίθμου shallow learning ως μέσο πρόβλεψης μπορεί να λογίζεται ως η βέλτιστη λύση σε αρκετές περιπτώσεις. Πολλές φορές, όσον αφορά το πρόβλημα της διαχείρισης της κυκλοφορίας, η ευκολία ερμηνείας των αποτελεσμάτων σε συνδυασμό με την απλότητα του αλγορίθμου και το μικρό υπολογιστικό κόστος υπερτερούν έναντι ενός μικρού κέρδους απόδοσης.

Η σύγκριση διαφορετικών μοντέλων αποτελεί από μόνη της ένα σημαντικό εμπόδιο στην προσπάθεια αξιολόγησής τους (Barros et al., 2015; Vlahogianni et al., 2014). Σε κάθε περίπτωση όμως, είναι σημαντικό να λαμβάνονται υπόψη όλα τα διαθέσιμα εργαλεία κατά τη διαδικασία υλοποίησης μιας τεχνικής πρόβλεψης. Μέσω της κατανόησης των χαρακτηριστικών τους αλλά και την αναγνώριση άλλων τυχόν περιορισμών, μπορεί να αιτιολογηθεί η τελική επιλογή μεθοδολογίας.

Παραμετρικές Μέθοδοι

Έναν επιπλέον διαχωρισμό των μεθόδων καθοδηγούμενων από δεδομένα αποτελούν τα παραμετρικά και μη παραμετρικά μοντέλα. Σημαντικό μέρος των πρωταρχικών σταδίων της βιβλιογραφίας, όσον αφορά το πρόβλημα της πρόβλεψης κυκλοφοριακών συνθηκών, αφιερώθηκε στη χρήση παραμετρικών μοντέλων. Αυτά τα μοντέλα απαιτούν τον προσδιορισμό κάποιων παραμέτρων μιας μαθηματικής σχέσης μέσω των διαθέσιμων δεδομένων, ώστε να προκύψει μια πρόβλεψη. Μερικά από τα πιο γνωστά είναι τα μοντέλα αυτόματης παλινδρόμησης ενσωματωμένου κινούμενου μέσου όρου (οικογένεια ARIMA), τα μοντέλα γραμμικής παλινδρόμησης, το μοντέλο φίλτρου Kalman, τα Μπεϋζιανά δίκτυα και το μοντέλο αλυσίδας Markov. Στατιστικά μοντέλα όπως τα παραπάνω διακρίνονται για την υπολογιστική τους απλότητα και την ευκολία ερμηνείας των αποτελεσμάτων τους. Ωστόσο, η ερευνητική κοινότητα αποστασιοποιείται από τις παραδοσιακές στατιστικές μεθόδους εξαιτίας της αδυναμίας τους να αντιμετωπίσουν το ασταθές και μη γραμμικό φαινόμενο της κυκλοφοριακής ροής. Κεντρικό χαρακτηριστικό των μοντέλων αυτών είναι μελέτη μέσων όρων μεγεθών, επομένως, ραγδαίες μεταβολές και ακραίες τιμές στα υπό εξέταση δεδομένα δεν περιγράφονται σε ικανοποιητικό βαθμό (Manibardo et al., 2022).

Μη-παραμετρικές Μέθοδοι

Λύση στα παραπάνω μειονεκτήματα των στατιστικών μεθόδων παρέχουν τα μη παραμετρικά μοντέλα. Σε αυτή την περίπτωση, η δομή και οι παράμετροι των μοντέλων καθορίζονται από τα ίδια τα δεδομένα μέσω μιας διαδικασίας αυτόματης εκμάθησης.

Ως συνέπεια, παρατηρείται μεγαλύτερη ακρίβεια κατά την μοντελοποίηση πολύπλοκων και μη γραμμικών φαινομένων, όπως αυτό της κυκλοφοριακής ροής. Για τη λεπτομερή αποκωδικοποίηση μοτίβων μεταξύ των κυκλοφοριακών μεγεθών, μέσω της διαδικασίας εκμάθησης, απαιτείται μεγαλύτερος όγκος δεδομένων σε σχέση με τα στατιστικά μοντέλα. Σε αυτή την κατηγορία ανήκουν και τα μοντέλα μηχανικής μάθησης της παρούσας διπλωματικής εργασίας. Μερικά γνωστά μοντέλα της κατηγορίας είναι τα Νευρωνικά Δίκτυα (Neural Networks), η Παλινδρόμηση των k Κοντινότερων Γειτόνων (K-Nearest Neighbors Regression), τα Δέντρα Απόφασης (Decision Trees).

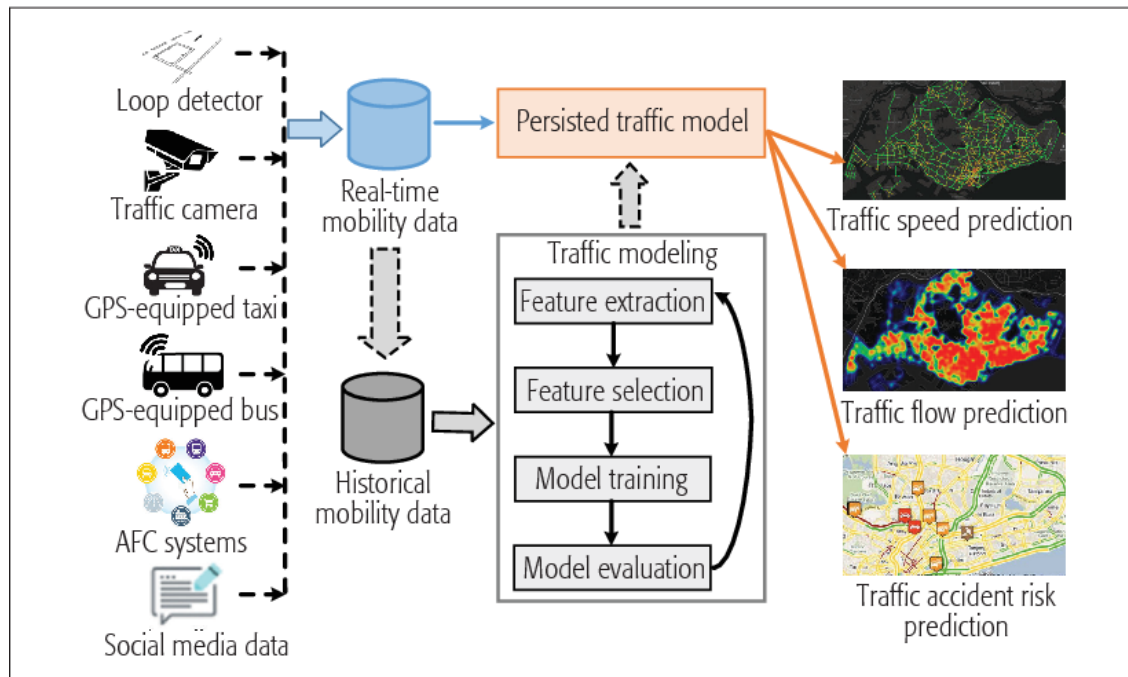
Άμεσες ή έμμεσες μέθοδοι πρόβλεψης

Το τελευταίο κριτήριο ταξινόμησης δεν αναφέρεται τόσο σε συγκεκριμένες ομάδες μοντέλων όσο στον τρόπο με τον οποίο υπολογίζεται η μεταβλητή πρόβλεψης. Στις έμμεσες μεθόδους αξιοποιούνται οι προβλεπόμενες τιμές των διάφορων κυκλοφοριακών μεταβλητών σε συνδυασμό με τη μαθηματική τους συσχέτιση με την μεταβλητή πρόβλεψης. Ενώ, οι άμεσες μέθοδοι συνήθως υλοποιούν τη διαδικασία πρόβλεψης χρησιμοποιώντας ιστορικά δεδομένα κυκλοφοριακών συνθηκών και παράγουν ως όρισμα εξόδου την πρόγνωση για τα ζητούμενα μεγέθη.

2.3 Έρευνες Πρόβλεψης Κυκλοφοριακού Φόρτου και Χρόνου Διαδρομής

Πρότυπα Πρόβλεψης Κυκλοφοριακού Φόρτου

Όπως έγινε κατανοητό η βιβλιογραφία πρόβλεψης κυκλοφοριακών μεγεθών περιλαμβάνει ποικιλία μεθόδων πρόγνωσης. Συγκεκριμένα, οι αλγόριθμοι μηχανικής μάθησης αποτελούν συχνά το προτιμώμενο εργαλείο, το οποίο ισχύει και για την παρούσα διπλωματική εργασία. Για αυτό τον λόγο παρατίθενται έρευνες πρόβλεψης κυκλοφοριακών μεγεθών (χρόνος διαδρομής, κυκλοφοριακός φόρτος) με χρήση κλασικών αλγορίθμων μηχανικής μάθησης (shallow learning).



Εικόνα 4: Τα βασικά στοιχεία της πρόβλεψης της κυκλοφορίας (Πηγή: Z. Liu et al., 2018)

Οι (Janconi et al., 2021) εφάρμοσαν πληθώρα μοντέλων μηχανικής μάθησης για την πρόβλεψη της κυκλοφορίας. Συγκεκριμένα, διεξήγαγαν δύο πειράματα:

1. την πρόβλεψη του ημερήσιου κυκλοφοριακού φόρτου σε διάφορα αστικά οδικά τμήματα,
2. την πρόβλεψη του ωριαίου κυκλοφοριακού φόρτου σε επιλεγμένα σημεία μετρήσεων για τις δύο κατευθύνσεις κίνησης.

Τα κυκλοφοριακά δεδομένα και για τις δύο περιπτώσεις προήλθαν από εγκατεστημένους επαγωγικούς βρόχους στο οδόστρωμα για την χρονική περίοδο 2011-2018. Τα αποτελέσματα υπέδειξαν ότι οι αλγόριθμοι μηχανικής μάθησης είναι ικανοί για ακριβείς προβλέψεις. Στην πρώτη περίπτωση το μοντέλο Δέντρου Απόφασης (Decision Tree) εμφάνισε την καλύτερη απόδοση, ενώ στη δεύτερη το μοντέλο που βασίστηκε στον αλγόριθμο Lazy IBk. Ως τελικό συμπέρασμα σημειώθηκε ότι όλα τα μοντέλα βασισμένα στον αλγόριθμο των Δέντρων Απόφασης συνδύαζαν καλή απόδοση και ευκολία χρήσης και στα δύο πειράματα. Επίσης, αναφέρθηκε ότι η τεχνική της συσταδοποίησης (clustering) θα μπορούσε να εφαρμοστεί για την αντιμετώπιση ασυνεπειών που παρατηρήθηκαν μεταξύ πραγματικών τιμών και τιμών πρόβλεψης.

Σε μία άλλη περίπτωση, οι (Y. Liu et al., 2021), επιχείρησαν να προβλέψουν τον κυκλοφοριακό φόρτο οδικών τμημάτων για χρονικά διαστήματα 15 λεπτών και 1 ώρας.

Εφαρμόστηκε ο αλγόριθμος των Ενισχυμένων Δέντρων Απόφασης (GBDT) που ανήκει στην κατηγορία της συλλογικής μάθησης (ensemble learning), καθώς και το μοντέλο Wavelet Neural Network (shallow learning), ένα είδος νευρωνικού δικτύου. Κυκλοφοριακά δεδομένα αντλήθηκαν για τα εξεταζόμενα οδικά τμήματα μέσω τεχνολογίας GPS, ενώ αξιοποιήθηκαν και καιρικά δεδομένα. Σύμφωνα με τα αποτελέσματα, η πρόβλεψη για μικρότερο χρονικό ορίζοντα ήταν πιο αξιόπιστη. Ενώ, το μοντέλο συλλογικής μάθησης GBDT (ensemble learning) που χρησιμοποιήθηκε υπερίσχυσε έναντι του νευρωνικού δικτύου (shallow learning).

Οι (Yang et al., 2017) κατασκεύασαν ποικιλία μοντέλων βασισμένα στον αλγόριθμο των Ενισχυμένων Δέντρων Απόφασης (GBDT). Τα δεδομένα εννέα επαγωγικών βρόχων για το οδικό τμήμα State Route 22, Garden Grove (Η.Π.Α) συλλέχθηκαν από τη βάση δεδομένων ανοιχτής πρόσβασης Caltrans Performance Measurement System (PeMS). Τα μοντέλα αναπτύχθηκαν για διάφορους συνδυασμούς μεταβλητών εισόδου, όπως διαφορετικούς συνδυασμούς των επαγωγικών βρόχων και δεδομένα για την γειτονική κυκλοφορία και για χρονικά διαστήματα πρόβλεψης πέντε, δέκα και δεκαπέντε λεπτών. Σκοπός της έρευνας αποτέλεσε η αναγνώριση της σχετικής σημαντικότητας των διάφορων μεταβλητών εισόδου (feature importance), καθώς και η βελτίωση της ακρίβειας πρόβλεψης της κυκλοφορίας. Ως συμπέρασμα, η κατάλληλη ρύθμιση των υπερ-παραμέτρων κρίθηκε απαραίτητη για την βέλτιστη απόδοση του μοντέλου GBDT. Συγκρίνοντας το με άλλα γνωστά μοντέλα μηχανικής μάθησης, όπως τα Support Vector Machine (SVM) και Back Propagation Neural Network (BPNN), το GBDT εμφάνισε μεγαλύτερη ακρίβεια πρόβλεψης. Το σφάλμα ήταν ελάχιστο στην περίπτωση του χρονικού διαστήματος πρόβλεψης 5 λεπτών. Τέλος, παρατηρήθηκε ότι η ενσωμάτωση δεδομένων σχετικά με την γειτονική κυκλοφορία, βελτίωσε σημαντικά την απόδοση του μοντέλου GBDT.

Πρότυπα Πρόβλεψης Χρόνου Διαδρομής

Ακόμα περισσότερες είναι οι έρευνες που σχετίζονται με την πρόβλεψη του χρόνου διαδρομής. Οι (Zhang & Haghani, 2015) ανέπτυξαν μεθόδους συλλογικής μάθησης (GBDT, Random Forrest) καθώς και στατιστικές μεθόδους (ARIMA) για την πρόβλεψη του χρόνου διαδρομής σε δύο τμήματα αυτοκινητοδρόμου. Ιστορικά δεδομένα χρόνων διαδρομής των οδικών τμημάτων χρησιμοποιήθηκαν για τη διεξαγωγή του πειράματος. Ο χρόνος μετακίνησης υπολογίστηκε για χρονικό ορίζοντα πέντε, δεκαπέντε και τριάντα λεπτών. Αναφορικά με το μοντέλο GBDT, παρατηρήθηκε ότι οι μεταβλητές με την μεγαλύτερη σημαντικότητα για κάθε διάστημα πρόβλεψης ήταν η αμέσως προηγούμενη τιμή του χρόνου διαδρομής και ο ρυθμός μεταβολής του χρόνου διαδρομής των δύο τελευταίων τιμών. Το παραπάνω μοντέλο κατάφερε να αποτυπώσει με μεγαλύτερη ακρίβεια την πολυπλοκότητα και τις απότομες αλλαγές της κυκλοφορίας, παρέχοντας

αξιόπιστες προβλέψεις. Για την επιβεβαίωση του παραπάνω ισχυρισμού, διεξάχθηκε ένα δευτερεύον πείραμα σε μεγαλύτερα τμήματα αυτοκινητοδρόμων με στόχο την πρόβλεψη χρόνου διαδρομής σε καταστάσεις κυκλοφοριακής συμφόρησης και ομαλής ροής. Το μοντέλο GBDT και σε αυτή την περίπτωση υπερίσχυσε έναντι των άλλων. Αξίζει να σημειωθεί ότι η ρύθμιση των υπερ-παραμέτρων του μοντέλου αποτέλεσε κομμάτι της έρευνας και κρίσιμο στάδιο για την επίτευξη βέλτιστης απόδοσης. Παρομοίως με προηγούμενες έρευνες μικρότερα χρονικά διαστήματα πρόβλεψης χαρακτηρίζονταν από μικρότερο σφάλμα.

Στην περίπτωση των (Qiu & Fan, 2021) αναπτύχθηκαν τέσσερις μη παραμετρικές μέθοδοι μηχανικής μάθησης για την πρόγνωση του χρόνου διαδρομής σε οδικά τμήματα του αυτοκινητόδρομου I-485 Charlotte (Η.Π.Α.). Χρησιμοποιήθηκαν οι αλγόριθμοι συλλογικής μάθησης (ensemble learning) Random Forest, Decision Tree, XGBoost και το νευρωνικό δίκτυο LSTM (Long short-term memory). Το μοντέλο Random Forest εμφάνισε την μεγαλύτερη ακρίβεια πρόβλεψης. Παράλληλα, έγινε διερεύνηση της σημαντικότητας των μεταβλητών εισόδου. Από τα 35 αρχικά επιλεγμένα χαρακτηριστικά (features) του μοντέλου διατηρήθηκαν τα 23 με βάση την τιμή σημαντικότητας τους. Μεταβλητές σχετικές με το χρόνο διαδρομής του αμέσως προηγούμενου χρονικού διαστήματος (προηγούμενη τιμή) από εκείνο της πρόβλεψης κρίθηκαν ως οι σημαντικότερες και σε αυτή την περίπτωση. Εκτός από χρονικές μεταβλητές (ημερομηνία, ώρα), υψηλή σημαντικότητα επέδειξαν μεταβλητές σχετικές με την ταχύτητα και τις καιρικές συνθήκες. Ως γενικό συμπέρασμα από το πείραμα προέκυψαν η υπολογιστική οικονομία και η υψηλή ακρίβεια που χαρακτηρίζουν τα μοντέλα συλλογικής μάθησης.

Σε παρόμοια συμπεράσματα κατέληξε και η έρευνα των (Cheng et al., 2019). Χρησιμοποιώντας διάφορες μεθόδους μηχανικής μάθησης επιχείρησαν την πρόβλεψη χρόνου διαδρομής σε τμήμα αυτοκινητοδρόμου. Λόγω έλλειψης επαρκούς όγκου πραγματικών δεδομένων χρόνου διαδρομής αξιοποιήθηκε το πρόγραμμα προσομοίωσης της κυκλοφορίας VISSIM. Εισάγοντας τα υπάρχοντα δεδομένα χρόνου διαδρομής και ρυθμίζοντας τις απαραίτητες κυκλοφοριακές παραμέτρους στο πρόγραμμα, είναι δυνατή η δημιουργία πολλαπλών προσομοιώσεων. Η προσαύξηση της τελικής βάσης δεδομένων πραγματοποιήθηκε με την συλλογή των δεδομένων που παρήγαγαν οι προσομοιώσεις. Τα μοντέλα που εφαρμόστηκαν ήταν τα εξής: Back Propagation Neural Network (BPNN), Support Vector Machine (SVM), Gradient Boosting Decision Trees (GBDT). Η μέθοδος συλλογικής μάθησης που εφάρμοσαν (GBDT) υπερίσχυσε έναντι των πιο απλών μοντέλων. Όπως και σε άλλες περιπτώσεις αντίστοιχων ερευνών, το σχετικό σφάλμα αυξανόταν όσο αυξανόταν και ο χρονικός

ορίζοντας πρόβλεψη. Ενώ η μεταβλητή με την μεγαλύτερη σημαντικότητα αποδείχτηκε η τελευταία τιμή του χρόνου διαδρομής πριν το διάστημα πρόβλεψης.

2.4 Εφαρμογές της Συνάρτησης BPR

Σύμφωνα με τους (Vlahogianni et al., 2014), τίθεται το ερώτημα αν οι σύγχρονες μέθοδοι που παρουσιάστηκαν προηγουμένως προσφέρουν μεγαλύτερη ακρίβεια πρόβλεψης σε σχέση με άλλες παραδοσιακές μεθόδους. Ένα εργαλείο ικανό να περιγράψει κυκλοφοριακές συνθήκες αποτελούν οι συναρτήσεις Φόρτου-Καθυστέρησης (Volume-Delay Functions), με γνωστότερη την συνάρτηση BPR (Bureau of Public Roads Function). Ακολουθεί μια σύνοψη της βιβλιογραφίας αναφορικά με την χρήση και παραμετροποίηση της συνάρτησης BPR για την πρόβλεψη κυκλοφοριακών μεγεθών.

Η συνάρτηση BPR έχει την ακόλουθη μορφή:

$$T = T_{ff} \left(1 + \alpha \left(\frac{V}{C} \right)^\beta \right) \quad (1)$$

όπου:

T = χρόνος διάνυσης οδικού τμήματος

T_{ff} = χρόνος διάνυσης οδικού τμήματος σε συνθήκες ελεύθερης ροής

V = κυκλοφοριακός φόρτος οδικού τμήματος

C = κυκλοφοριακή ικανότητα οδικού τμήματος

α, β = παράμετροι

Συνήθεις τιμές των παραμέτρων α και β είναι 0.15 και 4 αντίστοιχα. Στην πράξη όμως, έχει χρησιμοποιηθεί ένα μεγάλο εύρος τιμών και για τις δύο παραμέτρους κατά την εφαρμογή της συνάρτησης (Mahdi et al., 2022). Η παραμετροποίηση της συνάρτησης BPR αποτελεί σημαντικό εργαλείο για την αντιμετώπιση των περιορισμών που την χαρακτηρίζουν στην βασική της μορφή. Η συνάρτηση αναπτύχθηκε με σκοπό την χρήση της σε τμήματα αυτοκινητοδρόμων ανεπτυγμένων χωρών. Επομένως, οδικές υποδομές με διαφορετικά χαρακτηριστικά (φωτεινοί σηματοδότες, διασταυρώσεις, πινακίδες ρύθμισης κυκλοφορίας, παράνομη στάθμευση, γεωμετρικά χαρακτηριστικά οδού, μέγεθος οχημάτων, ποσοστό βαρέων οχημάτων κτλ.) απέχουν αρκετά από το αρχικό πλαίσιο για το οποίο σχεδιάστηκε η συνάρτηση (Mehbub Anwar et al., 2011; Pan et al., 2022; Saberi & Figliozzi, 2011). Η ρύθμιση των παραμέτρων λαμβάνοντας υπόψιν τα συγκεκριμένα χαρακτηριστικά ενός οδικού τμήματος μπορεί να βελτιώσει την συμπεριφορά της συνάρτησης. Έναν περιορισμό σε αυτή την αντιμετώπιση αποτελεί η

εκ νέου παραμετροποίηση της συνάρτησης που απαιτείται για κάθε πιθανή περίπτωση οδικού τμήματος (Pan et al., 2022).

Τα διάφορα χαρακτηριστικά των οδικών υποδομών που αναφέρθηκαν επηρεάζουν σημαντικά και τη διαδικασία εκτίμησης των κυκλοφοριακών μεγεθών που απαιτούνται για την χρήση της συνάρτησης (Mehbub Anwar et al., 2011). Συγκεκριμένα, η ακριβής εκτίμηση της κυκλοφοριακής ικανότητας (**C**) ενός οδικού τμήματος αποτελεί πολύπλοκη διαδικασία, ειδικά στην περίπτωση έλλειψης των κατάλληλων δεδομένων (Petrik et al., 2014; Saberi & Figliozzi, 2011).

Πίνακας 1 : Παράγοντες που επηρεάζουν την κυκλοφοριακή ικανότητα (Πηγή: ENNOIES ΒΑΣΙΚΩΝ ΠΑΡΑΜΕΤΡΩΝ ΚΥΚΛΟΦΟΡΙΑΣ Γ. Γιαννής - Ι. Γκόλιας - Γ. Κανελλαΐδης)

<p>Οδικές Συνθήκες:</p> <ul style="list-style-type: none"> ▪ Πλάτος λωρίδας κυκλοφορίας και οδοστρώματος ▪ Απόσταση κατακόρυφων εμποδίων ▪ Ερείσματα ▪ Βοηθητικές λωρίδες (στάθμευσης, αλλαγής ταχύτητας στροφής και αναμονής, πλέξης, ανωφέρειας) ▪ Κατάσταση επιφάνειας κύλισης ▪ Χάραξη οδού ▪ Κατά μήκος κλίση ▪ Ορατότητα 	<p>Κυκλοφοριακές Συνθήκες:</p> <ul style="list-style-type: none"> ▪ Ποσοστό φορτηγών αυτοκινήτων ▪ Ποσοστό διερχομένων λεωφορείων ▪ Ποσοστό λεωφορείων που σταματούν για επιβίβαση/αποβίβαση επιβατών ▪ Κατανομή κυκλοφορίας ανά λωρίδα ▪ Ποσοστό οχημάτων που στρέφουν δεξιά ή αριστερά ▪ Διακυμάνσεις της κυκλοφορίας, ιδιαίτερα ο ΣΩΑ ▪ Στάθμευση παρά το κράσπεδο, στάσεις λεωφορείων στην οδό ▪ Ροές πεζών ▪ Είδος οδηγών (τακτικοί και μη)
<p>Συνθήκες Λειτουργίας:</p> <ul style="list-style-type: none"> ▪ Λειτουργία της οδού ως μονόδρομου ή ως οδού δύο κατευθύνσεων ▪ Ύπαρξη φωτεινής σηματοδότησης ▪ Ύπαρξη προτεραιότητας με οδική σήμανση 	

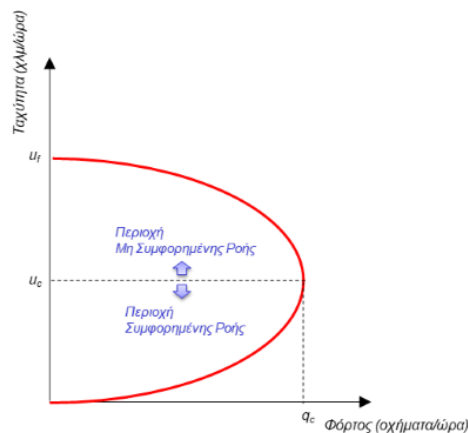
Η έννοια της κυκλοφοριακής ικανότητας έχει υποστεί αρκετές αλλαγές ιστορικά με τις διάφορες εκδόσεις του Εγχειριδίου Κυκλοφοριακής Ικανότητας (HCM), καθώς και με τις μεταβαλλόμενες ανάγκες της έρευνας των κυκλοφοριακών εφαρμογών (Petrik et al., 2014; Suh & John Kim, 1990; Wu, 2020). Σύμφωνα με το Εγχειρίδιο Κυκλοφοριακής Ικανότητας 2000 (Highway Capacity Manual 2000), η κυκλοφοριακή ικανότητα εκφράζει τον μέγιστο αριθμό οχημάτων ή πεζών που μπορεί να περάσουν από μία διατομή ή ομοιόμορφο τμήμα λωρίδας κυκλοφορίας ή οδού, κατά μία κατεύθυνση ή και κατά τις δύο κατευθύνσεις κατά τη διάρκεια μιας δεδομένης χρονικής περιόδου, με τις οδικές και κυκλοφοριακές συνθήκες καθώς και τις συνθήκες ελέγχου της κυκλοφορίας που επικρατούν.

Σε παλιότερη έκδοση του Εγχειριδίου Κυκλοφοριακής Ικανότητας (Highway Capacity Manual), παρουσιάστηκε ένα εύρος έξι επίπεδων εξυπηρέτησης (level-of-service), το οποίο ποσοτικοποιεί τις συνθήκες λειτουργίας των στοιχείων ενός οδικού συστήματος.

Το επίπεδο εξυπηρέτησης A εκφράζει τις συνθήκες ελεύθερης ροής, ενώ το επίπεδο εξυπηρέτησης F τον κορεσμό του συστήματος.

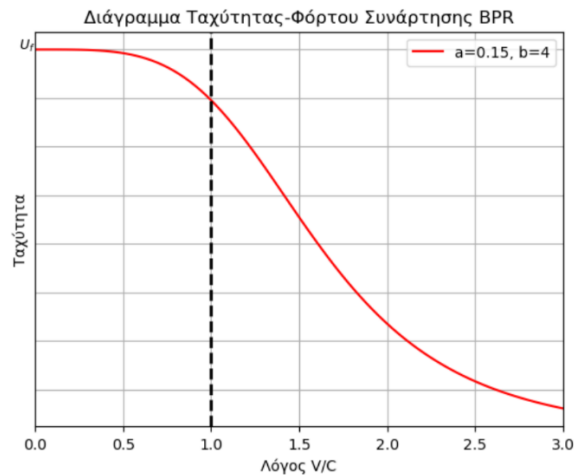
Η χρήση της έννοιας της κυκλοφοριακής ικανότητας χαρακτηρίζεται από ασυνέπεια σε αρκετές περιπτώσεις στην βιβλιογραφία. Κοινώς αποδεκτή πρακτική αποτελεί πλέον η χρήση του όρου της «μέγιστης ικανότητας», δηλαδή, του μέγιστου φόρτου οχημάτων του οδικού τμήματος που αντιστοιχεί σε επίπεδο εξυπηρέτησης E (όπου $V/C = 1$) (Marshall, 2018; Wu, 2020). Με βάση αυτό, συμπεραίνεται ότι η περίπτωση του κυκλοφοριακού φόρτου να ξεπεράσει την ικανότητα της οδού ($V/C > 1$) δεν είναι εφικτή στην πράξη, παρά μόνο για μερικά λεπτά πριν το σύστημα επιστρέψει σε χαμηλότερους ρυθμούς ροής ($V/C < 1$).

Όσον αφορά την συνάρτηση BPR, η τιμή της ικανότητας συνήθως αντιστοιχεί σε φόρτους επιπέδων εξυπηρέτησης C ή D (level-of-service) του Εγχειριδίου Κυκλοφοριακής Ικανότητας (TRB, 2011). Με βάση αυτό τον ορισμό, η κυκλοφοριακή ικανότητα δεν δεσμεύεται από την έννοια μέγιστου ρυθμού ροής που επιβάλλουν τα χαρακτηριστικά της οδού. Σε αυτή την περίπτωση ο κυκλοφοριακός φόρτος φέρει την έννοια της θεωρητικής ζήτησης και όχι της πραγματικής κυκλοφοριακής ροής. Επομένως, η έκφραση $V/C > 1$ δηλώνει ότι η ζήτηση έχει ξεπεράσει την προσφορά και το σύστημα βρίσκεται σε κατάσταση συμφόρησης. Αυτή η θεώρηση έρχεται σε αντίθεση με τη θεμελιώδη θεώρηση της κυκλοφοριακής ροής (TRB, 2011). Συνεπώς, ο λόγος φόρτου-ικανότητας (V/C) της συνάρτησης δεν μπορεί να παρέχει κάποιο πρακτικό νόημα ως προς την ανάλυση της κυκλοφορίας και την συσχέτιση του με πραγματικά δεδομένα κυκλοφοριακών μεγεθών. Η αντίθεση των δύο θεωρήσεων του λόγου φόρτου-ικανότητας γίνεται ακόμα πιο κατανοητή από την παρατήρηση των παρακάτω διαγραμμάτων.



Διάγραμμα 1: Θεμελιώδες διάγραμμα ταχύτητας-φόρτου (Πηγή: ΚΥΚΛΟΦΟΡΙΑΚΗ ΤΕΧΝΙΚΗ I. Φραντζεσκάκης - Ι. Γκόλιας - Μ. Πιτσιάβα-Λατινοπούλου)

Σύμφωνα με το θεμελιώδες παραβολικό διάγραμμα της σχέσης ταχύτητας-φόρτου (Διάγραμμα 1), η αύξηση του κυκλοφοριακού φόρτου μέχρι το σημείο κυκλοφοριακής ικανότητας (**C**) επιφέρει την μείωση της ταχύτητας (άνω κλάδος). Μετά από αυτό το σημείο τόσο ο φόρτος όσο και ταχύτητα μειώνονται μέχρι να μηδενιστούν, κάτι το οποίο συνάδει και με πραγματικές παρατηρήσεις των εν λόγω κυκλοφοριακών μεγεθών.



Διάγραμμα 2: Διάγραμμα ταχύτητας-φόρτου της συνάρτησης BPR

Ένα αντίστοιχο διάγραμμα ταχύτητας-φόρτου της συνάρτησης BPR (Διάγραμμα 2) έχει την μορφή μιας φθίνουσας συνάρτησης, ανεξάρτητα των τιμών των παραμέτρων α , β . Παρατηρείται ότι μετά το σημείο $V/C = 1$ ο φόρτος συνεχίζει να αυξάνεται έως ότου μηδενιστεί η ταχύτητα.

Τέλος, έναν ακόμα βασικό περιορισμό της συνάρτησης αποτελεί το γεγονός ότι τείνει να υποτιμά τον χρόνο διαδρομής σε συνθήκες κυκλοφοριακής συμφόρησης ($V/C > 1$). Ενώ σε αρκετές περιπτώσεις παρατηρήθηκε ότι και στην αντίστροφη περίπτωση ($V/C < 1$) συχνά η συνάρτηση μπορεί να υπερεκτιμήσει τον χρόνο διαδρομής (Barka & Politis, 2021; Wu, 2020).

2.5 Συμπεράσματα Βιβλιογραφίας

Σύμφωνα με την βιβλιογραφία, η πρόγνωση κυκλοφοριακών μεγεθών σε πολλές περιπτώσεις γίνεται με την χρήση ιστορικών δεδομένων της μεταβλητής πρόβλεψης ως όρισμα εισόδου. Στην παρούσα διπλωματική εργασία έγινε προσπάθεια ανάπτυξης μοντέλων που δεν απαιτούν ιστορικά δεδομένα για την μεταβλητή πρόβλεψης. Η εκτίμηση των κυκλοφοριακών φόρτων στην αρχή και το πέρας ενός οδικού τμήματος έγινε μόνο με γνώση των αντίστοιχων χρόνων διαδρομής. Ενώ, για την πρόβλεψη του χρόνου διαδρομής χρησιμοποιήθηκαν αποκλειστικά δεδομένα κυκλοφοριακών φόρτων. Ακόμα, αρκετές έρευνες κάνουν χρήση των χωρικών συσχετίσεων μεταξύ των

κυκλοφοριακών μεταβλητών για την επίτευξη μιας πρόβλεψης. Ένα παράδειγμα είναι η αξιοποίηση κυκλοφοριακών δεδομένων γειτονικών οδικών τμημάτων. Η χρήση του κυκλοφοριακού φόρτου εισόδου στο οδικό τμήμα θα συνέβαλλε σημαντικά στην πρόβλεψη του κυκλοφοριακού φόρτου εξόδου. Ωστόσο, αποφασίστηκε η διερεύνηση της ικανότητας πρόβλεψης των μοντέλων χωρίς αυτή την πληροφορία χωρικής συσχέτισης.

Η πλειοψηφία των ερευνών που συναντώνται στην βιβλιογραφία πρόβλεψης κυκλοφορίας αφορά οδικά τμήματα και όχι συνολικά ένα ευρύτερο οδικό δίκτυο. Μία από τις αιτίες για την οποία παρατηρείται αυτό το φαινόμενο είναι η απουσία δεδομένων μιας κυκλοφοριακής μεταβλητής για πολλά τμήματα του δίκτυο. Την επίλυση αυτού του προβλήματος διερευνά κατά ένα μέρος και η παρούσα διπλωματική εργασία. Όπως έχει ήδη αναφερθεί η καταγραφή συγκεκριμένα του κυκλοφοριακού φόρτου σε όλα τα σημεία ενός δικτύου δεν είναι πάντα εφικτή. Η πρόβλεψη του με μοντέλα που απαιτούν ως όρισμα εισόδου μόνο τον χρόνο διαδρομής ίσως μπορεί να αποτελέσει μια χρήσιμη πρακτική. Αυτό δεν σημαίνει ότι τα μοντέλα που αξιοποιούν την λιγότερη δυνατή πληροφορία αποτελούν την βέλτιστη στρατηγική. Έχει αναφερθεί ήδη αρκετές φορές ότι η πρόβλεψη της κυκλοφορίας αποτελεί πολύ-παραγοντικό πρόβλημα. Αυτό επιβεβαιώνεται και από την βιβλιογραφία όπου ποικιλία δεδομένων εκτός των βασικών κυκλοφοριακών παραμέτρων ενός οδικού τμήματος μπορούν να συμβάλλουν σημαντικά στην καλύτερη πρόβλεψη ενός κυκλοφοριακού μεγέθους. Συνήθη παραδείγματα αποτελούν τα καιρικά δεδομένα, κυκλοφοριακές συνθήκες γειτονικών οδικών τμημάτων και στατιστικά δεδομένα ατυχημάτων.

Χρήσιμα συμπεράσματα προέκυψαν και όσον αφορά τα μοντέλα μηχανικής μάθησης για την πρόβλεψη κυκλοφοριακών μεγεθών. Μοντέλα που αξιοποιούν μεγάλες βάσεις πραγματικών δεδομένων για την επίλυση πολύπλοκων προβλημάτων, όπως αυτά της μηχανικής μάθησης οδηγούν συνήθως σε καλύτερα αποτελέσματα σε σχέση με παραδοσιακά μοντέλα βασισμένα σε μαθηματικές σχέσεις κυκλοφοριακή ροής. Ειδικότερα, μοντέλα βασισμένα στα Δέντρα Απόφασης και τις παραλλαγές τους συχνά οδηγούσαν σε καλύτερες προβλέψεις συγκριτικά με άλλα μοντέλα ρηχής μάθησης (shallow learning) στις έρευνες που παρουσιάστηκαν. Τέλος, μοντέλα βαθιάς μάθησης (deep learning) συχνά παρουσιάζουν στην βιβλιογραφία μεγαλύτερη ακρίβεια σε προβλήματα πρόβλεψης, όμως παράγοντες όπως η πολυπλοκότητα ενός μοντέλου, η καταλληλότητα για το είδος του προβλήματος, η ευκολία ερμηνείας των αποτελεσμάτων και η δυνατότητα μεταφοράς των αποτελεσμάτων σε άλλες συνθήκες θα πρέπει να λαμβάνονται υπόψη για την επιλογή της κατάλληλης μεθοδολογίας.

Όσον αφορά τις συναρτήσεις φόρτου-καθυστερήσης η απουσία ενός γενικού πλαισίου παραμετροποίησης τους είναι εμφανής. Κάθε μαθηματική συνάρτηση φόρτου-

καθυστέρησης περιλαμβάνει ενδεικτικές τιμές παραμέτρων οι οποίες δεν είναι συμβατές με όλα τα πιθανά οδικά τμήματα και τα χαρακτηριστικά τους. Επομένως, για την χρήση τους απαιτείται μια διαδικασία ρύθμισης των παραμέτρων τους. Από την μέθοδο παραμετροποίησης της συνάρτησης BPR της παρούσας διπλωματικής εργασίας προκύπτουν και οι τιμές των παραμέτρων για τις οποίες θα μπορούσε να εφαρμοστεί στο υπό εξέταση οδικό τμήμα της Λ. Αλεξάνδρας. Ακόμα, η απουσία στέρεων και αυστηρών μαθηματικών ορισμών για τις έννοιες της κυκλοφοριακής συμφόρησης (V/C) και ικανότητας που συνάδουν και με εμπειρικές παρατηρήσεις, τονίζουν περαιτέρω την αδυναμία των συναρτήσεων φόρτου-καθυστέρησης ως εργαλεία πρόβλεψης και ανάλυσης της κυκλοφορίας. Επιπρόσθετα, η εκτίμηση του όρου της κυκλοφοριακής ικανότητας της συνάρτησης BPR απαιτεί κυκλοφοριακά δεδομένα που πιθανόν να μην είναι πάντα διαθέσιμα. Με την μέθοδο που αναπτύχθηκε στην παρούσα διπλωματική εργασία για την εφαρμογή της συνάρτησης, η ακριβής εκτίμηση της κυκλοφοριακής ικανότητας δεν κρίνεται αναγκαία όσον αφορά την πρόβλεψη του χρόνου διαδρομής. Είναι σημαντικό να αναφερθεί ότι οι συναρτήσεις φόρτου-καθυστέρησης αναφέρονται αποκλειστικά σε οδικά τμήματα, βασιζόμενες στη θεώρηση ότι ο κυκλοφοριακός φόρτος εισόδου είναι ίσος με τον κυκλοφοριακό φόρτο εξόδου μιας οδού. Συνεπώς αγνοούν εντελώς την συμπεριφορά δικτύου που παρουσιάζουν τα συνδεδεμένα οδικά τμήματα. Με βάση το παραπάνω, οι παραδοχές στις οποίες βασίζεται η συνάρτηση BPR αποτελούν αφορμή για την σύγκριση της με μη-παραμετρικές μεθόδους καθοδηγούμενες από δεδομένα, όπως αυτές της μηχανικής μάθησης. Με αυτό τον τρόπο είναι δυνατό να διαπιστωθεί αν οι παραδοχές ευσταθούν για την ανάλυση του φαινομένου της κυκλοφοριακής ροής.

Κεφάλαιο 3

Μεθοδολογική Προσέγγιση και Θεωρητικό Υπόβαθρο

3.1 Εισαγωγή

Η πρόγνωση χρόνου διαδρομής και κυκλοφοριακού φόρτου σε οδικό τμήμα αποτελεί στόχο της παρούσας διπλωματικής. Το Κεφάλαιο 3 πραγματεύεται την μεθοδολογία που χρησιμοποιήθηκε για την υλοποίηση των μοντέλων πρόβλεψης, καθώς και το θεωρητικό υπόβαθρο στο οποίο βασίζεται. Αλγόριθμοι μηχανικής μάθησης και η συνάρτηση BPR εφαρμόστηκαν για την πρόγνωση των ζητούμενων μεταβλητών. Ακολουθούν γενικές πληροφορίες σχετικά με τον κλάδο της μηχανικής μάθησης και ειδική αναφορά στους επιλεγμένους αλγόριθμους των μοντέλων Decision trees και GBDT. Τέλος, θα γίνει σύντομη παρουσίαση της συνάρτησης BPR και της μεθόδου ελαχίστων τετραγώνων που χρησιμοποιήθηκε για την παραμετροποίησή της.

3.2 Μηχανική μάθηση

Η Μηχανική Μάθηση (Machine Learning) αποτελεί υπο-πεδίο των κλάδων τεχνητής νοημοσύνης και επιστήμης υπολογιστών που επικεντρώνεται στην ανάπτυξη αλγορίθμων και μοντέλων ικανών να βελτιώνουν την απόδοσή τους σε ένα έργο, «μαθαίνοντας» από δεδομένα και προηγούμενη εμπειρία. Οι αλγόριθμοι αυτοί βρίσκουν εφαρμογή πλέον σε πληθώρα επιστημονικών πεδίων. Η έννοια της μηχανικής μάθησης γεννήθηκε δεκαετίες πριν, όταν από το 1950 Άλαν Τούρινγκ (Alan Turing) έθεσε την ιδέα μηχανών ικανών να μιμηθούν την ανθρώπινη νοημοσύνη. Λίγα χρόνια αργότερα, το 1959, ο Άρθουρ Σάμιουελ (Arthur Samuel), πρωτοπόρος στον κλάδο της τεχνητής νοημοσύνης επινόησε και επίσημα τον όρο.

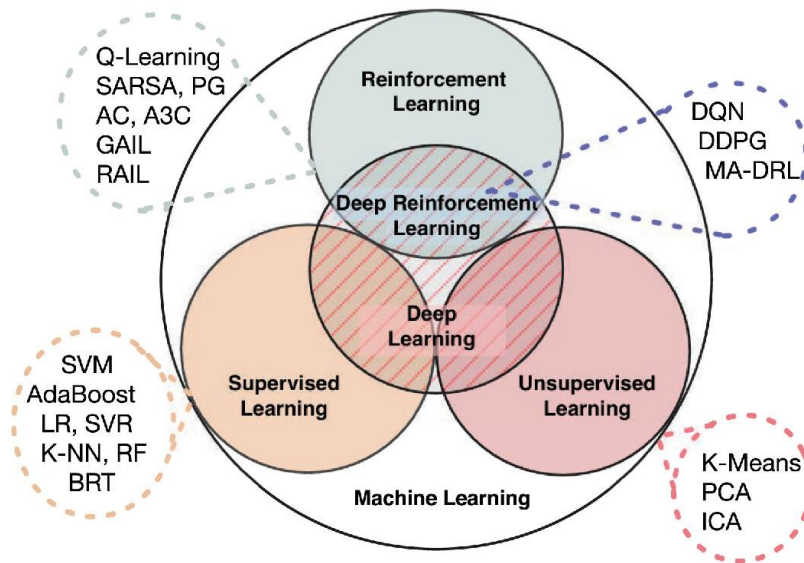
Οι μέθοδοι μηχανικής μάθησης διακρίνονται σε τέσσερις ευρείες κατηγορίες, με βάση την φύση του εκπαιδευτικού σήματος.

1. Επιτηρούμενη Μάθηση (Supervised Learning): Αποτελεί μία από τις πιο διαδεδομένες μεθόδους μηχανικής μάθησης. Περιλαμβάνει την εκπαίδευση του μοντέλου μέσω δεδομένων όπου υπάρχει προκαθορισμένη αντιστοιχία μεταξύ των ορισμάτων εισόδου και εξόδου. Συνεπώς, ο αλγόριθμος «μαθαίνει» να υπολογίζει το σωστό αποτέλεσμα με βάση την παραπάνω αντιστοιχία, καθιστώντας τον ικανό να κάνει προβλέψεις σε νέα δεδομένα. Οι αλγόριθμοι αυτού του είδους χρησιμοποιούνται κυρίως σε προβλήματα ταξινόμησης (classification) και παλινδρόμησης (regression).
 - Παλινδρόμηση (Regression): Η ταξινόμηση αποτελεί τύπο αλγορίθμου εποπτευόμενης μηχανικής μάθησης, στο οποίο το μοντέλο παράγει συνεχείς αριθμητικές τιμές ως έξοδο.

- Ταξινόμηση (Classification): Σε αυτή την περίπτωση αλγορίθμων, το μοντέλο εκπαιδεύεται να προβλέπει διακριτές κατηγορίες ή κλάσεις ως έξοδο. Στόχος της ταξινόμησης είναι να οριστεί ένα όριο απόφασης που διαχωρίζει διαφορετικές κλάσεις με βάση τα χαρακτηριστικά εισόδου, επιτρέποντας στο μοντέλο να ταξινομήσει νέα δεδομένα σε μία από τις προκαθορισμένες κατηγορίες.
2. Μη-επιτηρούμενη Μάθηση (Unsupervised Learning): Σε αντίθεση με την επιτηρούμενη μάθηση, η αντιστοιχία μεταξύ των ορισμάτων εισόδου και εξόδου των δεδομένων απουσιάζει. Αυτό συνεπάγεται ότι ένα από τα καθήκοντα του αλγορίθμου είναι να εντοπίσει κάποια δομή ή μοτίβα στα δεδομένα χωρίς ρητή καθοδήγηση. Η χρήση τους εντοπίζεται σε προβλήματα συσταδοποίησης (clustering) και μείωσης διαστασιμότητας (dimensionality reduction).
 3. Ημι-επιτηρούμενη μάθηση (Semi-supervised Learning): Αποτελεί συνδυασμό των δύο παραπάνω μεθόδων. Στην πράξη, τυχαίνει σε μεγάλους όγκους δεδομένων να μην υπάρχει η προαναφερθείσα πλήρης αντιστοιχία μεταξύ τους. Ωστόσο, έχει παρατηρηθεί σε αρκετές εφαρμογές ότι η χρήση αντιστοιχισμένων και μη δεδομένων βελτιώνει την απόδοση των αλγορίθμων.
 4. Ενισχυτική μάθηση (Reinforcement Learning): Στην προκειμένη περίπτωση, ο αλγόριθμος αλληλοεπιδρά με ένα δυναμικό περιβάλλον και, μέσω μίας διαδικασίας ανατροφοδότησης ανταμοιβών και ποινών, λαμβάνει αποφάσεις. Στόχος είναι να βρεθεί η βέλτιστη στρατηγική μεγιστοποίησης των ανταμοιβών, ώστε να επιτευχθεί ο σκοπός του μοντέλου.

Ένας δευτερεύον διαχωρισμός των μεθόδων μηχανικής μάθησης έχει προηγηθεί στο Κεφάλαιο 2 της βιβλιογραφικής ανάλυσης. Ως υπενθύμιση, οι αλγόριθμοι μηχανικής μάθησης μπορούν επίσης να ταξινομηθούν με βάση την αρχιτεκτονική τους σε:

1. Ρηχή Μάθηση (Shallow Learning)
2. Συλλογική Μάθηση (Ensemble Learning)
3. Βαθιά Μάθηση (Deep Learning)



Εικόνα 5: Κατηγοριοποίηση των μεθόδων μηχανικής μάθησης (Πηγή: Yuan et al., 2022)

3.2.1 Περιγραφή Αλγορίθμων Μηχανικής Μάθησης και Συμπληρωματικές Έννοιες

Σε αυτή την ενότητα περιγράφεται μια γενική πρακτική ανάπτυξης αλγορίθμων μηχανικής μάθησης για την επίτευξη τεχνικών προβλέψεων. Έχει γίνει αντιληπτό ότι βασική προϋπόθεση των αλγορίθμων μηχανικής μάθησης αποτελεί η συλλογή δεδομένων υψηλής ποιότητας από αξιόπιστες πηγές. Κατά κανόνα, απαιτείται ένα στάδιο προ-επεξεργασίας των δεδομένων με σκοπό αυτά να αποκτήσουν κατάλληλη μορφή να τροφοδοτήσουν τον αλγόριθμο.

Συνήθως, ο υπολογισμός της συσχέτισης μεταξύ των δεδομένων αποτελεί ένα χρήσιμο πρώτο βήμα. Ο συντελεστής συσχέτισης είναι ένα στατιστικό μέτρο που ποσοτικοποιεί το βαθμό και την κατεύθυνση της σχέσης μεταξύ δύο μεταβλητών. Παρέχει μια αριθμητική τιμή που υποδεικνύει πόσο έντονα και προς ποια κατεύθυνση αλλάζουν δύο μεταβλητές. Οι πιο γνωστοί συντελεστές συσχέτισης είναι: 1) Pearson, 2) Spearman, 3) Kendall. Οι τιμές τους κυμαίνονται από -1 έως 1. Είναι σημαντικό να τονιστεί ότι η υψηλή συσχέτιση μεταξύ των μεταβλητών δεν συνεπάγεται σχέση αιτιότητας.

Στην συνέχεια, πραγματοποιείται διόρθωση λανθασμένων ή κενών τιμών και δημιουργείται η αντιστοιχία κάθε σημείου στην βάση δεδομένων με την τιμή στόχου (target value), εφόσον αυτό χρειάζεται.

Χαρακτηριστικό στάδιο των κλασικών αλγορίθμων μηχανικής μάθησης αποτελεί η επιλογή ή και κατασκευή των κατάλληλων χαρακτηριστικών (feature selection & engineering) που περιγράφουν τη δομή των δεδομένων με ακρίβεια. Τα παραπάνω

χαρακτηριστικά παίζουν καθοριστικό ρόλο στην βελτιστοποίηση της απόδοσης και την κατανόηση των αποτελεσμάτων του αλγορίθμου. Ένα παράδειγμα κατασκευής χαρακτηριστικού αποτελεί η ομαλοποίηση της κλίμακας των δεδομένων. Η διαδικασία της ομαλοποίησης υλοποιήθηκε και στα δεδομένα της διπλωματικής εργασίας. Υπάρχουν διάφορες τεχνικές ομαλοποίησης, όμως στο πλαίσιο της εργασίας εφαρμόστηκε ο μετασχηματισμός λογαρίθμου στα δεδομένα. Συγκεκριμένα μοντέλα όπως τα Decision trees και GBDT παρέχουν βαθμολογίες σχετικά με την σημαντικότητα των εν λόγω χαρακτηριστικών (feature importance). Αυτές παρέχουν μια ποσοτική αξιολόγηση για την συνεισφορά κάθε χαρακτηριστικού στο τελικό αποτέλεσμα.

Μετά την επιλογή μοντέλου με βάση το είδος και το πλήθος των δεδομένων, την φύση του προβλήματος και λαμβάνοντας υπόψη τους περιορισμούς κάθε αλγορίθμου, ακολουθεί το στάδιο της εκπαίδευσης. Στόχος μιας μηχανής μάθησης είναι η γενίκευση μιας διαδικασίας, δηλαδή η ικανότητά της να αποδίδει σε παρόμοια προβλήματα υπό νέα δεδομένα. Αυτό επιτυγχάνεται χωρίζοντας τυχαία την βάση δεδομένων σε δεδομένα εκπαίδευσης (training data) και δεδομένα ελέγχου (test data). Τα πρώτα επιτρέπουν στον αλγόριθμο να εντοπίσει μοτίβα και σχέσεις μεταξύ των μεταβλητών, ενώ τα δεύτερα χρησιμοποιούνται ως τα δεδομένα με βάση τα οποία αξιολογείται το μοντέλο. Κατά την προσαρμογή (fitting) του μοντέλου, ο αλγόριθμος επιχειρεί μέσω διάφορων τεχνικών βελτιστοποίησης (optimization techniques) να ελαχιστοποιήσει την συνάρτηση σφάλματος (loss function). Με άλλα λόγια, σκοπός είναι η προσαρμογή των διάφορων παραμέτρων του μοντέλου ώστε να μειωθεί η απόκλιση μεταξύ των ορισμάτων εξόδου (output values) και των τιμών στόχου (target values).

Για την επίτευξη του παραπάνω πολλές φορές κρίνεται αναγκαία και η ρύθμιση των υπερ-παραμέτρων (hyper-parameters) του μοντέλου. Αυτές δεν έχουν συσχετίζονται με τα δεδομένα κάθε προβλήματος και καθορίζονται ακριβώς πριν το στάδιο της εκπαίδευσης. Η εμπειρία σε αλγόριθμους μηχανικής μάθησης συχνά επιτρέπει τον ταχύτερο καθορισμό των βέλτιστων τιμών των υπερ-παραμέτρων, χωρίς την απαίτηση πολλών δοκιμών.

Τελευταίο βήμα για την ανάπτυξη ενός μοντέλου μηχανικής μάθησης αποτελεί η αξιολόγηση των αποτελεσμάτων του. Όπως έχει αναφερθεί προηγουμένως στόχος του αλγορίθμου είναι η ελαχιστοποίηση απόκλισης μεταξύ των παραγόμενων τιμών και των δεδομένων ελέγχου (test data), μέσω της προσαρμογή των διάφορων παραμέτρων του μοντέλου. Οι μετρικές σφάλματος (error metrics) παρέχουν μια ποσοτική αξιολόγηση του παραπάνω στόχου. Μερικά από τα πιο γνωστά είναι το Μέσο Απόλυτο Σφάλμα (Mean Absolute Error – MAE), το Μέσο Απόλυτο Σχετικό Σφάλμα (Mean Absolute Percentage Error - MAPE), Μέσο Τετραγωνικό Σφάλμα (Mean Squared Error – MSE), Ρίζα

Μέσου Τετραγωνικού σφάλματος (Root Mean Squared Error – RMSE). Έχουν την ακόλουθη μαθηματική μορφή:

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (2)$$

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^n \frac{|y_i - \hat{y}_i|}{y_i} \cdot 100\% \quad (3)$$

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (4)$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (5)$$

όπου \hat{y}_i ορίζεται ως η τιμή πρόβλεψης ενώ y_i η πραγματική τιμή

Εκτός από τον έλεγχο των αποτελεσμάτων του μοντέλου, ιδιαίτερα σημαντικός είναι και ο έλεγχος προσαρμογής τους στα δεδομένα. Συχνά, ένα εκπαιδευμένο μοντέλο δεν έχει την απαιτούμενη πολυπλοκότητα ώστε να αποτυπώσει σχέσεις τόσο σε υπάρχοντα όσο και σε νέα δεδομένα (underfit), εμφανίζοντας, ως συνέπεια, περιορισμένη χρησιμότητα. Σε άλλες περιπτώσεις, παρατηρείται εξαιρετική απόδοση σε υπάρχοντα δεδομένα και μεγάλο σφάλμα σε νέα (overfit). Δηλαδή, το μοντέλο έχει καταφέρει να αποδώσει όχι μόνο τις σχέσεις μεταξύ των μεταβλητών αλλά και τυχαία μοτίβα που εμφανίζονται στα δεδομένα. Η υπερβολικά λεπτομερής προσαρμογή του στα δεδομένα ελέγχου (training data) το καθιστούν αναποτελεσματικό σε νέα δεδομένα. Για τον έλεγχο της προσαρμογής των μοντέλων στα δεδομένα χρησιμοποιούνται διάφορα στατιστικά μέτρα (goodness of fit metrics), όπως το R^2 . Εφαρμόζεται σε μοντέλα παλινδρόμησης και

παρέχει πληροφορία σχετικά με την ικανότητα πρόβλεψης τους. Έχει την ακόλουθη μαθηματική μορφή:

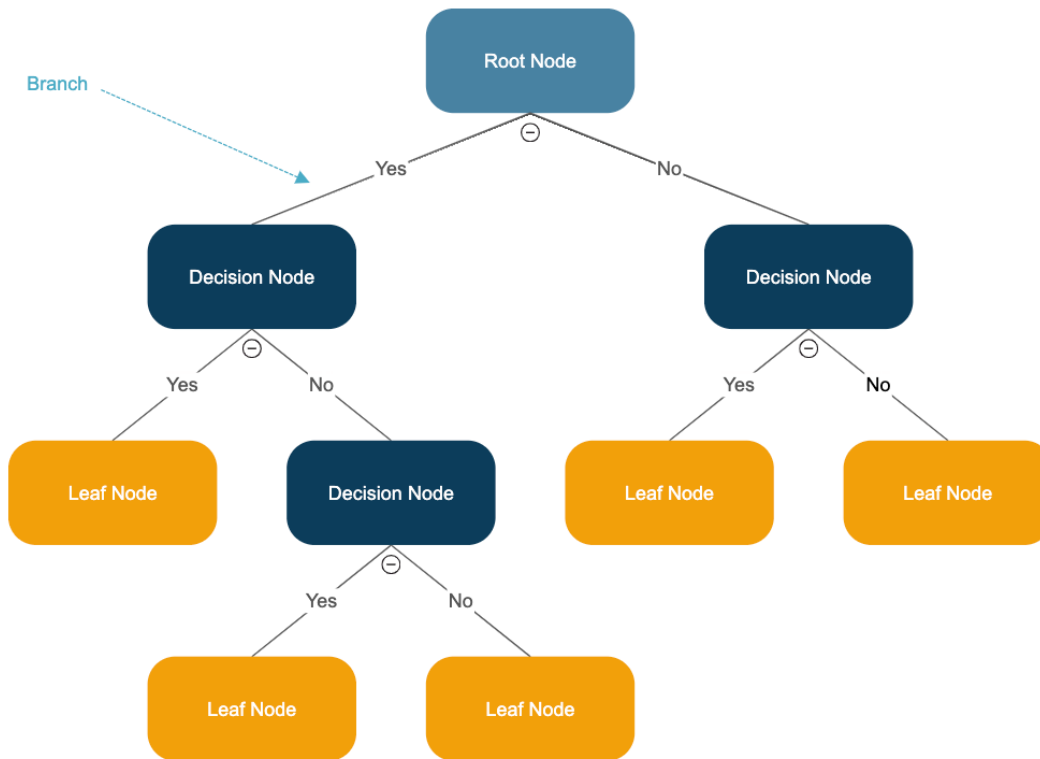
$$R^2 = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2} \quad (6)$$

όπου \bar{y} ορίζεται ως ο μέσος όρος των πραγματικών τιμών

Εκτιμάει το ποσοστό διακύμανσης για μια εξαρτημένη μεταβλητή που μπορεί να εξηγηθεί από το μοντέλο. Αξιολογεί την προσαρμογή στα δεδομένα, συγκρίνοντας τις παραγόμενες τιμές του μοντέλου με τον αντίστοιχο μέσο όρο της μεταβλητής. Λαμβάνει τιμές από 0 έως 1, όπου τιμή 1 σημαίνει τέλεια προσαρμογή στα δεδομένα και ότι το μοντέλο εξηγεί όλη τη διακύμανση που εμφανίζεται στα δεδομένα.

3.2.2 Δέντρα Απόφασης

Τα Δέντρα Απόφασης (Decision Trees) αποτελούν έναν κλασσικό αλγόριθμο εποπτευόμενης μηχανικής μάθησης. Βρίσκουν εφαρμογή τόσο σε προβλήματα ταξινόμησης (classification), όσο και σε προβλήματα παλινδρόμησης (regression), όπως στην περίπτωση της παρούσας διπλωματικής εργασίας. Χρησιμοποιώντας μια ιεραρχική δομή λαμβάνει αποφάσεις με βάση τα χαρακτηριστικά των δεδομένων εισόδου (input features). Στην κορυφή βρίσκεται ο κόμβος ρίζα (root node) ο οποίος λαμβάνει υπόψη το σύνολο των δεδομένων. Ο αλγόριθμος διαχωρίζει τα δεδομένα επιλέγοντας την τιμή του χαρακτηριστικού, η οποία αποφέρει το μικρότερο μέσο τετραγωνικό σφάλμα (MSE) στα υποσύνολα. Με αυτό τον τρόπο δημιουργούνται οι εσωτερικοί κόμβοι (internal node), τα κλαδιά (branches) των οποίων αποτελούν τις πιθανές τιμές που μπορεί να λάβει ένα χαρακτηριστικό (feature).



Εικόνα 6: Δομή Δέντρου Απόφασης (Πηγή: <https://www.smartdraw.com/decision-tree/>)

Η παραπάνω διαδικασία επαναλαμβάνεται μέχρι να ικανοποιηθεί το τερματικό κριτήριο που έχει οριστεί από τις υπερ-παραμέτρους του μοντέλου. Σε προβλήματα παλινδρόμησης ο αλγόριθμος καταλήγει στους τερματικούς κόμβους (leaf nodes), η τιμή των οποίων αποτελεί την τιμή πρόβλεψης, η οποία ορίζεται ως ο μέσος όρος των τιμών στόχου υπό του τερματικού κόμβου. Για την εκπαίδευση ενός δέντρου απόφασης απαιτείται ο καθορισμός κάποιων υπερ-παραμέτρων όπως:

- Μέγιστο βάθος (max_depth): Το μέγιστο βάθος του δέντρου. Ορίζει τον συνολικό αριθμό των κόμβων από την ρίζα (root node) έως τον τερματικό κόμβο (leaf node).
- Ελάχιστος αριθμός δειγμάτων για διαχωρισμό (min_samples_split): Με βάση την τιμή του γίνεται ο διαχωρισμός των κόμβων. Αν ένας κόμβος περιέχει λιγότερα δείγματα από την τιμή της παραμέτρου μετατρέπεται σε τερματικό κόμβο (leaf node).
- Ελάχιστος αριθμός δειγμάτων ανά τερματικό κόμβο (min_samples_leaf): Κόμβοι με λιγότερα δείγματα από την τιμή της παραμέτρου συγχωνεύονται με άλλους.
- Μέγιστος αριθμός χαρακτηριστικών (max_features): Ο μέγιστος αριθμός χαρακτηριστικών (features) τον οποίο λαμβάνει υπόψη ο αλγόριθμος σε κάθε κόμβο ώστε να πετύχει τον βέλτιστο διαχωρισμό.

- Τυχαία κατάσταση (random_state): Η διαμόρφωση ενός δέντρου απόφασης χαρακτηρίζεται από τυχαιότητα, ειδικά όταν συνυπάρχουν ισοδύναμες κατά τη δημιουργία του. Ορίζοντας την παραπάνω παράμετρο με μία σταθερή τιμή εξασφαλίζεται η επαναληψιμότητα της δομής και της απόδοσης του μοντέλου.

Πλεονεκτήματα των δέντρων απόφασης:

- Ευκολία απεικόνισης και κατανόησης του αλγορίθμου τους.
- Ιδιαίτερα αποτελεσματικά σε σύνθετα, μη γραμμικά προβλήματα.
- Δέχονται διακριτές και συνέχεις μεταβλητές.
- Παρέχουν εργαλεία για την αξιολόγηση της σημαντικότητας των χαρακτηριστικών (feature importance).
- Εύκολη ανάπτυξη αλγορίθμου και υψηλή ταχύτητα πρόβλεψης.

Μειονεκτήματα των δέντρων απόφασης:

- Παρουσιάζουν συχνά υπερ-προσαρμογή (overfitting) στα δεδομένα.
- Η από πάνω προς τα κάτω (top-down) προσέγγιση που χρησιμοποιείται από το αλγόριθμο μπορεί να μην οδηγεί πάντα στο βέλτιστο δέντρο απόφασης.
- Παρουσιάζουν αστάθεια. Διαφορετικές μορφές δέντρων σχηματίζονται από μικρές αλλαγές στα δεδομένα.
- Δυσκολία αποτύπωσης σχέσεων μεταξύ των μεταβλητών σε προβλήματα που εμφανίζουν γραμμικότητα.

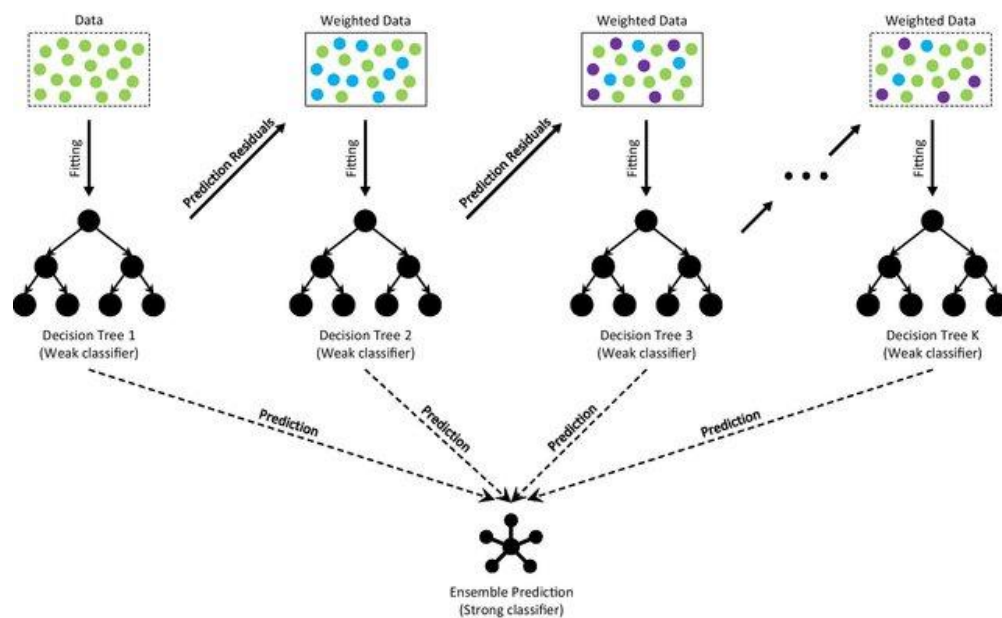
3.2.3 Ενισχυμένα Δέντρα Απόφασης

Τα δέντρα απόφασης αποτελούν ένα χρήσιμο και ισχυρό αλγόριθμο μηχανικής μάθησης, κατάλληλο για την επίλυση πληθώρας προβλημάτων. Μερικοί από τους παραπάνω περιορισμούς μπορούν να αντιμετωπιστούν συνδυάζοντας τον αλγόριθμο των δέντρων απόφασης με μεθόδους συλλογικής μάθησης (ensemble learning), όπως η τεχνική της ενίσχυσης (boosting). Μια εφαρμογή αυτού είναι ο αλγόριθμος Gradient Boosting Decision Trees (GBDT). Βασικό δομικό στοιχείο του αποτελεί το δέντρο απόφασης. Με την τεχνική Gradient Boosting βελτιώνεται η ικανότητα πρόβλεψης με τη διαδοχική προσθήκη Δέντρων Απόφασης στο μοντέλο. Ο αλγόριθμος αποτελείται αρχικά από ένα μοναδικό Δέντρο Απόφασης. Στην συνέχεια εισάγονται διαδοχικά νέα δέντρα απόφασης που έχουν ως στόχο την όλο και καλύτερη προσαρμογή στα δεδομένα. Τα νέα Δέντρα Απόφασης αποσκοπούν στην αποτύπωση μοτίβων που δεν εντοπίστηκαν από τα προηγούμενα και την ελαχιστοποίηση της συνάρτησης σφάλματος (loss function), με υπολογισμό της μερικής παραγώγου της. Για την συνάρτηση σφάλματος σε προβλήματα παλινδρόμησης συνήθως χρησιμοποιείται το μέσο τετραγωνικό σφάλμα (MSE). Οι τιμές πρόβλεψης των νέων δέντρων ενημερώνουν συνεχώς τα τελικά αποτελέσματα με έναν

συντελεστή βαρύτητας, ο οποίος εξαρτάται από τον ρυθμό εκμάθησης του αλγορίθμου (learning rate).

Η διαδικασία επαναλαμβάνεται μέχρι να ικανοποιηθεί το τερματικό κριτήριο που έχει οριστεί από τις υπερ-παραμέτρους του μοντέλου. Το τελικό αποτέλεσμα ενός μοντέλου GBDT παράγεται αθροίζοντας τις προβλέψεις όλων των μεμονωμένων δέντρων απόφασης. Η συνεισφορά κάθε νέου δέντρου ρυθμίζεται με την εφαρμογή του ρυθμού εκμάθησης (learning rate), όπως αναφέρθηκε προηγουμένως. Για την ανάπτυξη ενός αλγορίθμου GBDT απαιτείται ο καθορισμός κάποιων υπερ-παραμέτρων μερικές από αυτές είναι:

- Αριθμός εκτιμητριών (n_estimators): Ο αριθμός των μεμονωμένων δέντρων απόφασης που θα συμπεριληφθούν στο σύνολο (ensemble).
- Ρυθμός εκμάθησης (learning_rate): Ρυθμίζει την συνεισφορά κάθε δέντρου απόφασης στο τελικό αποτέλεσμα.
- Μέγιστο βάθος (max_depth): Το μέγιστο βάθος του δέντρου. Ορίζει τον συνολικό αριθμό των κόμβων από την ρίζα (root node) έως τον τερματικό κόμβο (leaf node).



Εικόνα 7: Δομή Ενισχυμένων Δέντρων Απόφασης (GBDT) (Πηγή: Deng et al., 2021)

Πλεονεκτήματα GBDT:

- Μεγαλύτερη απόδοση και ακρίβεια αποτελεσμάτων σε σχέση με άλλους απλούς αλγορίθμους μηχανικής μάθησης.
- Μετριάζει τον θόρυβο (noise) των δεδομένων, δίνοντας βαρύτητα στα σωστά ταξινομημένα δεδομένα.
- Ιδιαίτερα αποτελεσματικά σε σύνθετα, μη γραμμικά προβλήματα.
- Κατάλληλα για προβλήματα ταξινόμησης και παλινδρόμησης.
- Παρέχουν εργαλεία για την αξιολόγηση της σημαντικότητας των χαρακτηριστικών (feature importance).
- Σχετική ευκολία κατανόησης του μοντέλου, όπως στην περίπτωση των δέντρων απόφασης.

Μειονεκτήματα GBDT:

- Παρουσιάζουν συχνά υπερ-προσαρμογή (overfitting) στα δεδομένα, ειδικά στην περίπτωση που χρησιμοποιούνται βαθιές δομές δέντρων απόφασης και πολλές επαναλήψεις.
- Η κατάλληλη ρύθμιση των υπερ-παραμέτρων αποτελεί κρίσιμο βήμα για την επίτευξη ικανοποιητικής απόδοσης.
- Χαρακτηρίζονται από υψηλό υπολογιστικό κόστος σε περιπτώσεις μεγάλων βάσεων δεδομένων και βαθιών δέντρων απόφασης.
- Παρότι παρέχουν πληροφορία που συμβάλλει στην κατανόηση τους ως μοντέλα, παραμένουν πιο δυσνόητα σε σχέση με τα δέντρα απόφασης καθώς αποτελούν σύνολο (ensemble) μοντέλων.

3.3 Συναρτήσεις Φόρτου-Καθυστέρησης και BPR

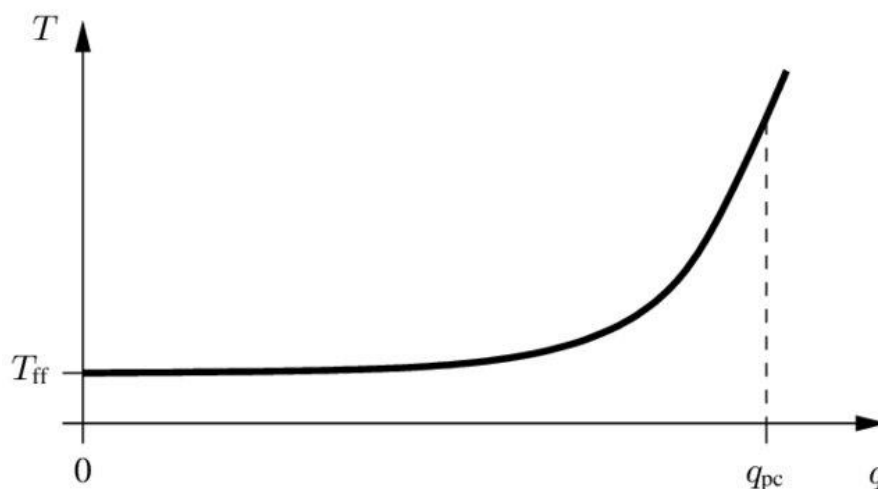
Ο χρόνος διαδρομής αποτελεί βασικό κριτήριο για την απόφαση πραγματοποίησης μιας μετακίνησης σε ένα οδικό σύστημα. Ταυτόχρονα, εξαρτάται άμεσα από την κυκλοφορία ενός οδικού τμήματος. Για την περιγραφή αυτού του φαινομένου έχουν αναπτυχθεί οι συναρτήσεις Φόρτου-Καθυστέρησης (Volume Delay Functions – VDFs). Ευρέως γνωστή είναι η συνάρτηση BPR (Bureau of Public Roads Function), η οποία αναπτύχθηκε από την Υπηρεσία Δημοσίων Οδών (Bureau of Public Roads), νυν Ομοσπονδιακή Διοίκηση Αυτοκινητοδρόμων (Federal Highway Administration) τη δεκαετία του 1960 ως μέρος του Εγχειριδίου Κυκλοφοριακής Ικανότητας (Highway Capacity Manual). Πρόκειται για μαθηματική συνάρτηση οι οποία περιγράφει το επίπεδο συμφόρησης οδικών τμημάτων εκφρασμένο ως χρονική καθυστέρηση, με βάση την μεταβολή του κυκλοφοριακού φόρτου.

Στην Σχέση (1), η παράμετρος α της συνάρτησης BPR εκφράζει τον λόγο του χρόνου μετακίνησης σε συνθήκες ελεύθερης ροής προς τον αντίστοιχο χρόνο σε συνθήκες

μέγιστης κυκλοφοριακής ικανότητας. Ενώ, η παράμετρος β εκφράζει την ταχύτητα μεταβολής από συνθήκες ελεύθερης ροής σε συνθήκες κυκλοφοριακής συμφόρησης. Μεγαλύτερες τιμές της παραμέτρου β συνεπάγονται μικρότερη χρονική καθυστέρηση σε συνθήκες χαμηλών κυκλοφοριακών φόρτων (Mahdi et al., 2022; Pan et al., 2022). Αυτό έχει ως αποτέλεσμα την μη έγκαιρη αναγνώριση κυκλοφοριακής συμφόρησης.

Η μεταβλητή T_{ff} εκφράζει τον χρόνο διάνυσης του οδικού τμήματος σε συνθήκες ελεύθερης ροής. Οι συνθήκες ελεύθερης ροής χαρακτηρίζονται από χαμηλούς κυκλοφοριακούς φόρτους και ταχύτητες, οι οποίες καθορίζονται μόνο από τις επιθυμίες των οδηγών, τα καθορισμένα όρια ταχυτήτων και τις επικρατούσες οδικές συνθήκες. Οι έννοιες της κυκλοφοριακού φόρτου (\mathbf{V} ή \mathbf{Q}) και της κυκλοφοριακής ικανότητας (\mathbf{C}) έχουν αναπτυχθεί σε προηγούμενα κεφάλαια.

Στο Διάγραμμα 3 απεικονίζεται η σχέση χρόνου διαδρομής και κυκλοφοριακού φόρτου όπως περιγράφεται από την μαθηματική συνάρτηση BPR.



Διάγραμμα 3: Διάγραμμα φόρτου-χρονικής καθυστέρησης της συνάρτησης BPR (Πηγή: Maerivoet & De Moor, 2005)

3.3.1 Μέθοδος ελαχίστων τετραγώνων

Η μέθοδος Ελαχίστων Τετραγώνων αποτελεί ευρέως γνωστή τεχνική για την βελτιστοποίηση συναρτήσεων. Συγκεκριμένα, η βιβλιοθήκη SciPy της γλώσσας προγραμματισμού Python, κάνει χρήση του αλγορίθμου Levenberg-Marquardt για την κατάλληλη τροποποίηση των παραμέτρων μιας συνάρτησης, με σκοπό την ελαχιστοποίηση του αθροίσματος του τετραγωνισμένου εναπομένου σφάλματος. Αφού οριστεί η αντικειμενική συνάρτηση \mathbf{S} , δίνονται αρχικές τιμές στις παραμέτρους. Στην συνέχεια ο αλγόριθμος υπολογίζει επαναληπτικά την τιμή της αντικειμενικής

συνάρτησης S για το σύνολο των δεδομένων, μεταβάλλοντας κατάλληλα τις τιμές των παραμέτρων. Η διαδικασία σταματά όταν μέγεθος των μεταβολών στην τιμή της αντικειμενικής συνάρτησης και των ζητούμενων παραμέτρων δεν είναι σημαντικό. Ως παράδειγμα, στην περίπτωση της συνάρτησης BPR, βασικός στόχος είναι η εύρεση κατάλληλων τιμών των παραμέτρων α και β . Η αντικειμενική συνάρτηση S θα έχει την εξής μορφή:

$$S(\alpha, \beta) = \sum_{i=1}^n (T_{obs,i} - T) \quad (7)$$

όπου:

n = πλήθος δεδομένων

$T_{obs,i}$ = πραγματικές τιμές χρόνου διαδρομής

T = τιμές χρόνου διαδρομής που υπολογίστηκαν με την συνάρτηση BPR (1)

Κεφάλαιο 4

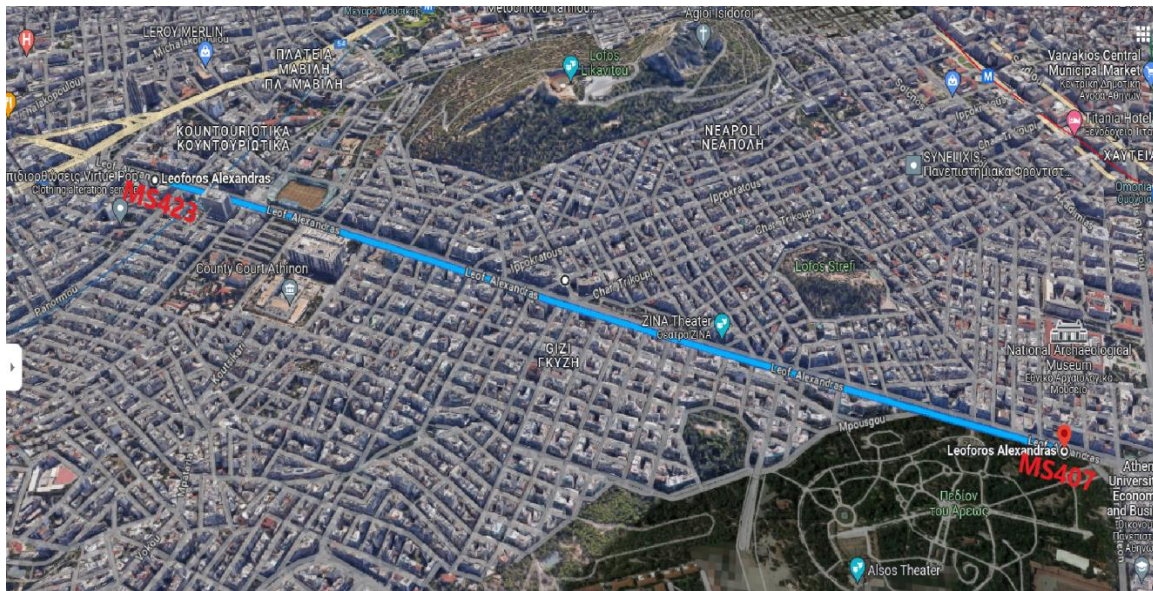
Συλλογή και επεξεργασία στοιχείων

4.1 Εισαγωγή

Σε αυτό το κεφάλαιο γίνεται επεξήγηση της συλλογής και επεξεργασίας στοιχείων που ήταν απαραίτητα για την επίτευξη του στόχου της εργασίας. Ακόμα, παρουσιάζεται η διαδικασία εφαρμογής των αλγόριθμων μηχανικής μάθησης και της συνάρτησης BPR στα δεδομένα. Για την ανάπτυξη των αλγορίθμων χρησιμοποιήθηκε η γλώσσα προγραμματισμού Python σε περιβάλλον Jupyter Notebooks.

4.2 Συλλογή στοιχείων

Για την εκπαίδευση των μοντέλων χρησιμοποιήθηκαν δεδομένα χρόνων διαδρομής και φόρτων οχημάτων. Συγκεκριμένα, επιλέχθηκαν δύο σημεία επί της λεωφόρου Αλεξάνδρας στο κέντρο της Αθήνας, όπου βρίσκονται εγκατεστημένοι φωρατές υπό του οδοστρώματος. Η απόσταση μεταξύ των φωρατών είναι περίπου 2.3km (2282m). Ο φωρατής με ονομασία **MS423** έχει τις ακόλουθες γεωγραφικές συντεταγμένες: (37.98706952, 23.75875395), ενώ ο φωρατής **MS407**: (37.9914926, 23.73333264).



Εικόνα 8: Οδικό τμήμα υπό μελέτη της Λ. Αλεξάνδρας

Από βάσεις δεδομένων του Ε.Μ.Π. αντλήθηκαν, για τους παραπάνω φωρατές, τα δεδομένα φόρτων οχημάτων ανά μία ώρα για 80 ημέρες. Οι 80 ημέρες μετρήσεων δεν ήταν συνεχείς. Επιλέχθηκαν διαστήματα 20 ημερών για κάθε μήνα από το Δεκέμβριο του 2022 έως τον Μάρτιο του 2023, όπου οι φωρατές κατέγραφαν ομαλά δεδομένα, χωρίς κάποια βλάβη. Στον Πίνακα 2 παρουσιάζονται οι ακριβείς ημερομηνίες των καταγραφών

που χρησιμοποιήθηκαν, ενώ ο Πίνακα 3 περιέχει χρήσιμα στατιστικά στοιχεία της βάσης δεδομένων των φωρατών με τις καταγραφές οχημάτων:

Πίνακας 2: Ημερομηνίες καταγραφής δεδομένων

1 Δεκεμβρίου 2022 έως 21 Δεκεμβρίου 2022

6 Ιανουαρίου 2023 έως 26 Ιανουαρίου 2023

1 Φεβρουαρίου 2023 έως 21 Φεβρουαρίου 2023

6 Μαρτίου 2023 έως 26 Μαρτίου 2023

Πίνακας 3: Στατιστικά μέτρα δεδομένων κυκλοφοριακού φόρτου

	MS407	MS423
πλήθος	1848	1848
μέσος όρος	769.34	838.55
διάμεσος	924	1000
τ. απόκλιση	319.23	380.96
ελάχιστο	10	7
μέγιστο	1365	1520
25%	513	520
50%	924	1000
75%	1013	1144

Για το υπό εξέταση οδικό τμήμα της Λ. Αλεξάνδρας μεταξύ των δύο φωρατών, αξιοποιήθηκαν δεδομένα χρόνων διαδρομής από την εφαρμογή Google Maps. Τα δεδομένα αντλήθηκαν από την ιστοσελίδα <https://outscraper.com> για τα ίδια χρονικά διαστήματα με παραπάνω ανά μία ώρα. Εκτός από τους πραγματικούς χρόνους διαδρομής (**duration**) ανά ώρα τα δεδομένα περιείχαν και το εύρος εκτίμησης του χρόνου διαδρομής ως δύο τιμές διάρκειας, μια ελάχιστη εκτίμηση (**duration_min**) και μία μέγιστη (**duration_max**). Όλα τα δεδομένα χρόνων διαδρομής ήταν σε λεπτά. Στον Πίνακα 4 παρουσιάζονται χρήσιμα στατιστικά στοιχεία της βάσης δεδομένων των χρόνων διαδρομής.

Πίνακας 4: Στατιστικά μέτρα δεδομένων χρόνου διαδρομής

	duration(minutes)
πλήθος	1848
μέσος όρος	6.29
διάμεσος	6
τ. απόκλιση	2.04
ελάχιστο	4
μέγιστο	12
25%	4
50%	6
75%	8

4.3 Διαμόρφωση δεδομένων

Οι παραπάνω βάσεις δεδομένων για τους χρόνους διαδρομής και τους φόρτους οχημάτων έπρεπε να απαλλαχθούν από λάθη και κενές τιμές. Οι γραμμές όλων των βάσεων δεδομένων όπου απουσίαζαν τιμές, αγνοήθηκαν εντελώς. Με τον ίδιο τρόπο αντιμετωπίστηκαν οι ελάχιστες γραμμές όπου εμφανίζονταν λανθασμένες τιμές. Στην συνέχεια οι δύο βάσεις συγχωνεύτηκαν σε μια με όλα τα απαραίτητα δεδομένα χρόνων διαδρομής και φόρτων, στοιχισμένα κατάλληλα ανά ώρα και ημερομηνία, τα οποία απεικονίζονται στον Πίνακα 5.

Πίνακας 5: Βάση δεδομένων

datetime_utc	duration(minutes)	MS407	MS423
12/1/2022 2:00	4	193	146
12/1/2022 3:00	4	126	97
12/1/2022 4:00	4	136	101
12/1/2022 5:00	4	280	248
12/1/2022 6:00	4	657	735
12/1/2022 7:00	5	1012	1216
12/1/2022 8:00	6	1099	1289
12/1/2022 9:00	7	1028	1122
12/1/2022 10:00	8	985	1316
12/1/2022 11:00	8	1009	1145
...
...
...
3/26/2023 12:00	6	925	966
3/26/2023 13:00	6	994	1051

3/26/2023 14:00	7	1010	1128
3/26/2023 15:00	6	900	1051
3/26/2023 16:00	5	870	959
3/26/2023 17:00	5	910	982
3/26/2023 18:00	5	937	1160
3/26/2023 19:00	6	939	1119
3/26/2023 20:00	6	934	1113
3/26/2023 21:00	6	881	994

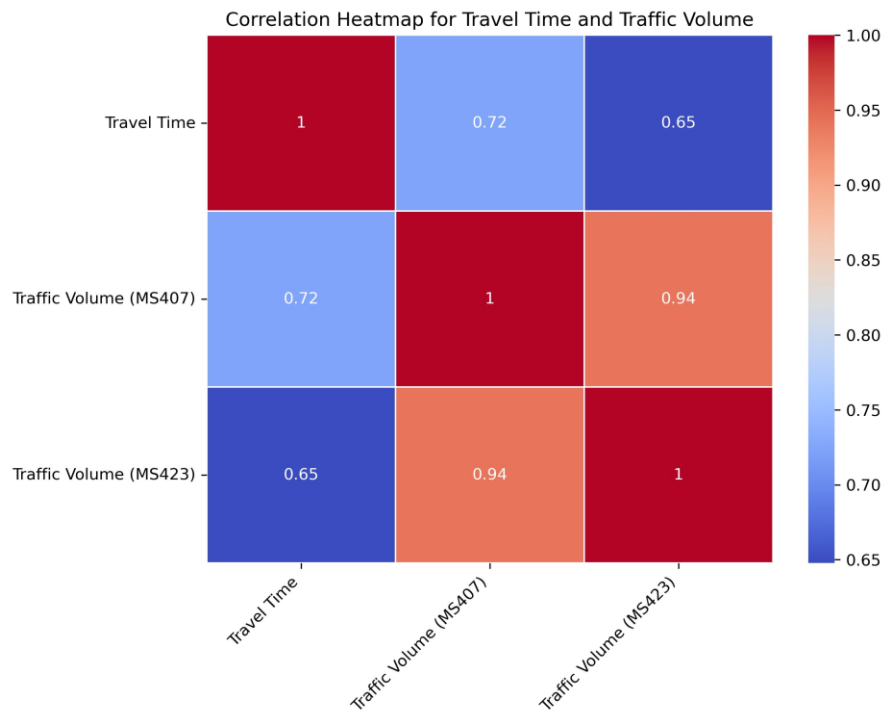
4.4 Επεξεργασία δεδομένων

Πρώτο βήμα για την επεξεργασία δεδομένων αποτέλεσε η εγκατάσταση βιβλιοθηκών στο περιβάλλον Jupyter Notebook, οι οποίες είναι απαραίτητες για όλες τις διαδικασίες ανάγνωσης και τροποποίησης των δεδομένων, την ανάπτυξη των μοντέλων, καθώς και την αξιολόγησή τους. Ορισμένες από αυτές τις βιβλιοθήκες συνοψίζονται παρακάτω:

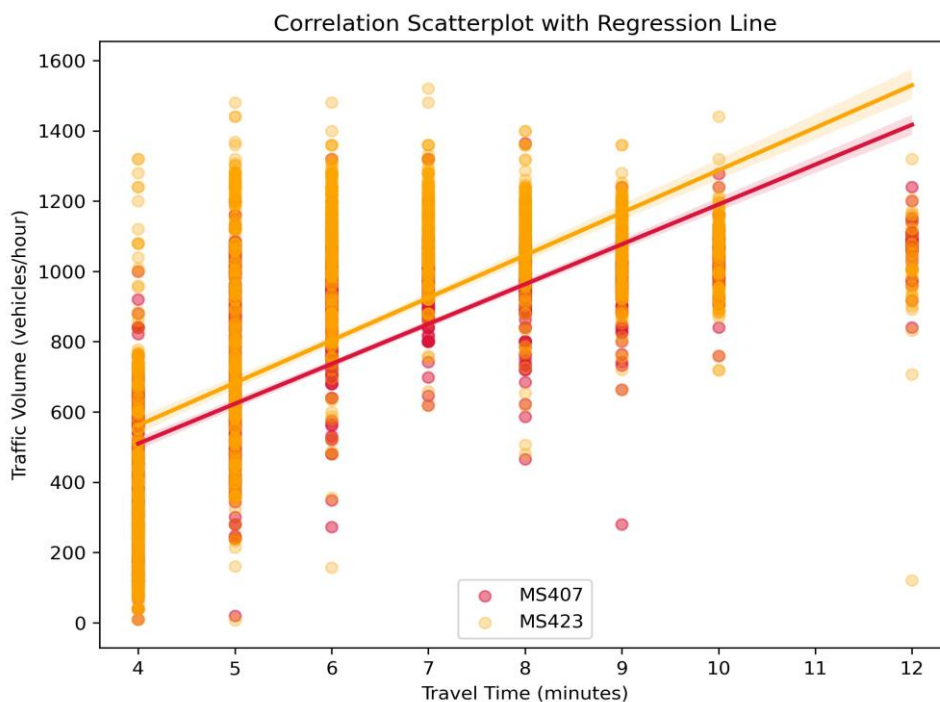
- NumPy: Μία βιβλιοθήκη Python ανοιχτού κώδικα για αριθμητικές και βασισμένες σε πίνακες λειτουργίες, παρέχοντας αποτελεσματικά εργαλεία για τη δημιουργία, το χειρισμό και την εκτέλεση μαθηματικών πράξεων σε πολυδιάστατους πίνακες και πίνακες.
- SciPy: Βιβλιοθήκη της Python η οποία επεκτείνει τις δυνατότητες της NumPy για επιστημονικές εργασίες, παρέχοντας εργαλεία στατιστικής, γραμμικής άλγεβρας κ.α.
- Pandas: Μία βιβλιοθήκη Python ανοιχτού κώδικα για χειρισμό και ανάλυση δεδομένων, που προσφέρει δομές και λειτουργίες δεδομένων για αποτελεσματικό χειρισμό και ανάλυση δομημένων δεδομένων, όπως πίνακες, με φιλικό προς τον χρήστη τρόπο.
- Scikit-learn: Μία βιβλιοθήκη μηχανικής μάθησης ανοιχτού κώδικα για την Python που παρέχει ένα ευρύ φάσμα εργαλείων και αλγορίθμων για εργασίες όπως ταξινόμηση, παλινδρόμηση, ομαδοποίηση και αξιολόγηση μοντέλων, καθιστώντας την πολύτιμη πηγή για τη δημιουργία και την ανάλυση μοντέλων μηχανικής μάθησης
- Matplotlib: Μια δισδιάστατη βιβλιοθήκη σχεδίασης που χρησιμοποιείται για τη δημιουργία γραφημάτων στη γλώσσα προγραμματισμού Python.

Για τις διαδικασίες πρόβλεψης που επιχείρησαν τα μοντέλα μηχανικής μάθησης στην συνέχεια, ιδιαίτερα σημαντική κρίνεται η γνώση του βαθμού συσχέτισης των δεδομένων χρόνων διαδρομής και φόρτων οχημάτων. Υψηλή συσχέτιση μεταξύ δύο μεταβλητών συνήθως αποτελεί ένδειξη ότι η μία πιθανόν να μπορεί να χρησιμοποιηθεί ως μέσο πρόβλεψης για την άλλη. Με τη βιβλιοθήκη Pandas που αναφέρθηκε προηγουμένως,

υπολογίστηκε εύκολα ο συντελεστής συσχέτισης Pearson μεταξύ των μεταβλητών της βάσης δεδομένων. Στο Διάγραμμα 4 απεικονίζεται ο βαθμός συσχέτισης του χρόνου διαδρομής με τον κυκλοφοριακό φόρτο για κάθε φωρατή. Ενώ στο Διάγραμμα 5 απεικονίζονται με μεγαλύτερη λεπτομέρεια οι τιμές τους στην βάση δεδομένων.



Διάγραμμα 4: Θερμικός χάρτης συσχέτισης φόρτου-χρόνου διαδρομής



Διάγραμμα 5: Διάγραμμα διασποράς φόρτου-χρόνου διαδρομής

Από τα παραπάνω διαγράμματα προέκυψε ότι η συσχέτιση των μεταβλητών είναι αρκετά καλή, συνεπώς η διαδικασία ήταν δυνατό να προχωρήσει στο στάδιο της πρόβλεψης τιμών.

Επόμενο βήμα πριν την κατασκευή των μοντέλων μηχανικής μάθησης είναι ο εμπλουτισμός και η τελική διαμόρφωση της βάσης δεδομένων. Για να καταστεί δυνατή η εύρεση μοτίβων στα δεδομένα γίνεται η επιλογή ενός υποσυνόλου σχετικών χαρακτηριστικών (feature selection) με βάση τα οποία ο αλγόριθμος θα κατηγοριοποιεί τα δεδομένα. Ο στόχος της επιλογής χαρακτηριστικών στην μηχανική μάθηση είναι η καλύτερη προσαρμογή των δεδομένων στο πρόβλημα, εντοπίζοντας και απορρίπτοντας εκείνα τα χαρακτηριστικά τα οποία παρέχουν περιττή πληροφορία. Επομένως, δημιουργήθηκαν τα απαραίτητα χαρακτηριστικά, τα οποία κρίθηκαν ικανά να περιγράψουν καλύτερα την σχέση μεταξύ χρόνων διαδρομής και φόρτων οχημάτων, επιφέροντας ένα αποδεκτό σφάλμα πρόβλεψης σε κάθε περίπτωση. Τέλος, τα χαρακτηριστικά αυτά προστέθηκαν στην τελική βάση δεδομένων. Ακολουθεί ο Πίνακας 6 με τα χαρακτηριστικά που χρησιμοποιήθηκαν.

Πίνακας 6: Χαρακτηριστικά (features) μοντέλων μηχανικής μάθησης

Weekday	Λαμβάνει τιμή 1 για τις καθημερινές και 0 για το Σαββατοκύριακο
Rush_hour	Λαμβάνει τιμή 1 για την πρωινή ώρα αιχμής (8:00 - 11:00), 2 για την απογευματινή ώρα αιχμής (15:00 - 19:00) και 0 για όλες τις άλλες ώρες της ημέρας.
W+R2	Συνδυασμός των 2 παραπάνω χαρακτηριστικών. Για τις καθημερινές λαμβάνει τιμές 1 ή 2 ανάλογα την ώρα αιχμής. Για το Σαββατοκύριακο λαμβάνει την τιμή 0 ανεξαρτήτως της ώρα παρατήρησης.
min_max	Η διαφορά της μέγιστης και της ελάχιστης τιμής του χρόνου διαδρομής που παρείχε ως εκτίμηση η εφαρμογή Google Maps (duration_max – duration_min).
diff	Η διαφορά δύο συναπτών τιμών χρόνου διαδρομής.

Όπως έχει αναφερθεί προηγουμένως, στόχος της εργασίας είναι η πρόβλεψη τιμών χρόνων διαδρομής και φόρτων οχημάτων με αλγορίθμους μηχανικής μάθησης (Decision Tree Regressor, GBDT). Η κατασκευή των αλγορίθμων σε κάθε περίπτωση πρόβλεψης ήταν παρόμοια. Προφανώς για την περίπτωση της πρόβλεψης φόρτων οχημάτων τα δεδομένα των φωρατών MS407 και MS423 αποτέλεσαν τον στόχο πρόβλεψης (target) ενώ τα δεδομένα των χρόνων διαδρομής το όρισμα εισόδου (input) στους αλγόριθμους. Στην περίπτωση της πρόβλεψης χρόνων διαδρομής πραγματοποιήθηκε το αντίστροφο. Τα υπόλοιπα χαρακτηριστικά (features) που χρησιμοποιήθηκαν για την ανάπτυξη των αλγορίθμων σε όλα τα σενάρια ήταν παρόμοια και αναγράφονται παραπάνω.

Αξίζει να σημειωθεί ότι κατά την ανάπτυξη των μοντέλων πρόβλεψης δοκιμάστηκαν διάφορα χαρακτηριστικά (features) και συνδυασμοί τους, ώστε να προκύψει η βέλτιστη ικανότητα πρόβλεψης. Στην περίπτωση της πρόβλεψης χρόνων διαδρομής ο αλγόριθμος λειτούργησε επιτυχώς, χωρίς να χρειαστούν πολλές δοκιμές. Τα μοντέλα παρήγαγαν αποδεκτό σφάλμα πρόβλεψης και καλή προσαρμογή στα δεδομένα, με κάποια από τα χαρακτηριστικά που αναφέρονται παραπάνω. Στην περίπτωση της πρόβλεψης φόρτων οχημάτων όλοι οι συνδυασμοί χαρακτηριστικών που δοκιμάστηκαν απέτυχαν να επιφέρουν το επιθυμητό αποτέλεσμα. Στην μείωση του σφάλματος πρόβλεψης συνέβαλε σημαντικά η ομαλοποίηση των δεδομένων των φόρτων οχημάτων στην βάση δεδομένων. Υπολογίζοντας τον λογάριθμο των φόρτων οχημάτων, είχε ως αποτέλεσμα την μείωση της διακύμανσης των τιμών. Αυτή η τεχνική χρησιμοποιείται συχνά σε μοντέλα μηχανικής μάθησης, καθώς ορισμένοι αλγόριθμοι είναι ιδιαίτερα ευαίσθητοι στην κλίμακα των τιμών που καλούνται να επεξεργαστούν. Στον Πίνακα 7 εμφανίζονται οι τιμές του φόρτου μετά την ομαλοποίηση.

Πίνακας 7: Βάση δεδομένων μετά την ομαλοποίηση

datetime_utc	duration(minutes)	logMS407	logMS423
12/1/2022 2:00	4	2.286	2.164
12/1/2022 3:00	4	2.100	1.987
12/1/2022 4:00	4	2.134	2.004
12/1/2022 5:00	4	2.447	2.394
12/1/2022 6:00	4	2.818	2.866
12/1/2022 7:00	5	3.005	3.085
12/1/2022 8:00	6	3.041	3.110
12/1/2022 9:00	7	3.012	3.050
12/1/2022 10:00	8	2.993	3.119
12/1/2022 11:00	8	3.004	3.059
...
...
...
3/26/2023 12:00	6	2.966	2.985
3/26/2023 13:00	6	2.997	3.022
3/26/2023 14:00	7	3.004	3.052
3/26/2023 15:00	6	2.954	3.022
3/26/2023 16:00	5	2.940	2.982
3/26/2023 17:00	5	2.959	2.992
3/26/2023 18:00	5	2.972	3.064
3/26/2023 19:00	6	2.973	3.049
3/26/2023 20:00	6	2.970	3.046
3/26/2023 21:00	6	2.945	2.997

Στην συνέχεια για τον υπολογισμό των μετρικών σφάλματος των μοντέλων οι τιμές φόρτου πρόβλεψης μετατράπηκαν πάλι στην αρχική κλίμακα. Το σφάλμα πρόβλεψης του μοντέλου δεν έφτασε σε επιθυμητά επίπεδα ακόμα και με την παραπάνω διαδικασία. Για την περαιτέρω μείωση του σφάλματος πραγματοποιήθηκε αρχικά μια στατιστική ανάλυση της βάσης δεδομένων. Υπολογίστηκε ο μέσος όρος των φόρτων των φωρατών MS407 και MS423, καθώς όπως έγινε αντιληπτό και από το διάγραμμα διασποράς (βλ. Διάγραμμα 5) οι τιμές τους στην βάση δεδομένων είναι πανομοιότυπες. Έπειτα, κατηγοριοποιήθηκε για όλες τις διαφορετικές τιμές του χρόνου διαδρομής που εμφανίζονται στην βάση δεδομένων και υπολογίστηκαν τα στατιστικά μέτρα του Πίνακα 8.

Πίνακας 8: Στατιστικά μέτρα μέσου κυκλοφοριακού φόρτου ανά τιμή χρόνου διαδρομής

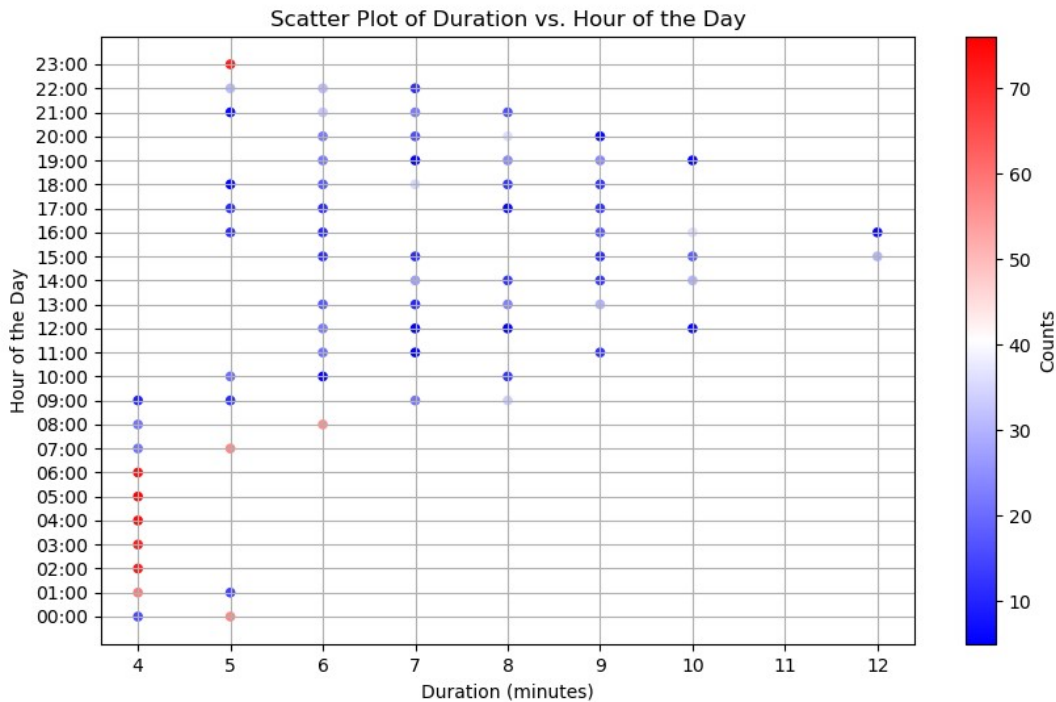
duration (minutes)	Πλήθος	Διάμεσος	Μέσος όρος	Επικρατούσα τιμή	Τυπική απόκλιση	Ελάχιστη τιμή	Μέγιστη τιμή
4	496	280.75	334.1724	120	213.0493	0	9
5	295	737.5	780.1237	560	263.8538	0	23
6	279	1031	1000.729	1080	153.0015	8	22
7	249	1063	1054.404	1080	82.15209	9	22
8	220	1057.75	1036.739	1000	96.47023	9	21
9	177	1041.5	1032.034	1080	83.23523	11	20
10	96	1040	1042.26	1040	69.13252	12	19
12	36	1045.25	1034.417	1136.5	119.5761	15	16

Στην συνέχεια υπολογίστηκε το στατιστικό μέτρο z-score για το σύνολο των δεδομένων, το οποίο εκφράζει πόσες τυπικές αποκλίσεις απέχει μια τιμή από τον μέσο όρο. Για κάθε τιμή του χρόνου διαδρομής καταγράφηκε το πλήθος των παρατηρήσεων που απείχαν περισσότερο από 1, 2 και 3 τυπικές αποκλίσεις και έπειτα υπολογίστηκε το ποσοστό τους έναντι του συνόλου των παρατηρήσεων της κατηγορίας.

Πίνακας 9: Z-score μέσου κυκλοφοριακού φόρτου ανά τιμή χρόνου διαδρομής

duration (minutes)	$ z >1$	$ z >1$ (%)	$ z >2$	$ z >2$ (%)	$ z >3$	$ z >3$ (%)
4	157	0.32	18	0.036	6	0.012
5	120	0.41	5	0.017	0	0.000
6	72	0.26	10	0.036	5	0.018
7	45	0.18	10	0.040	4	0.016
8	34	0.15	13	0.059	5	0.023
9	32	0.18	7	0.040	3	0.017
10	30	0.31	6	0.063	1	0.010
12	3	0.08	2	0.056	1	0.028

Τέλος, στο Διάγραμμα 6 απεικονίζονται οι ώρες της ημέρας που παρατηρείται κάθε διαφορετική τιμή χρόνου διαδρομής καθώς και η συνολικός αριθμός παρατήρησής τους της στην βάση δεδομένων.



Διάγραμμα 6: Πλήθος παρατηρήσεων διαφορετικών τιμών χρόνου διαδρομής για κάθε ώρα της ημέρας

Από τα παραπάνω συμπεραίνεται ότι χαμηλές τιμές χρόνου διαδρομής δημιουργούν προβλήματα στην πρόβλεψη των μοντέλων. Συγκεκριμένα, οι τιμές κυκλοφοριακού φόρτου που αντιστοιχούν σε χρόνο διαδρομής 4 και 5 λεπτών εμφανίζουν την μεγαλύτερη διακύμανση. Ενώ από τον υπολογισμό του z-score γίνεται αντιληπτό ότι κατέχουν σημαντικά μεγαλύτερο ποσοστό τιμών που απέχουν περισσότερο από μία τυπική απόκλιση ($|z| > 1$) από τον μέσο όρο, παρόλο που η τυπική απόκλιση σε αυτές τις τιμές είναι και η μέγιστη. Για την αντιμετώπιση του προβλήματος αφαιρέθηκαν από την βάση δεδομένων όλες οι καταγραφές για τις μεταμεσονύχτιες ώρες από τις 00:00-3:00, όπου παρατηρούνται αποκλειστικά χρόνοι διαδρομής 4 και 5 λεπτών. Επίσης, αφαιρέθηκε μόνο το απολύτως απαραίτητο πλήθος τιμών για την εξασφάλιση αποδεκτού σφάλματος πρόβλεψης ($MAPE < 20\%$) και καλής προσαρμογής στα δεδομένα ($R^2 > 0.6$), ώστε να αποφευχθεί η υπερβολική συρρίκνωση της αρχικής βάσης δεδομένων.

4.5 Αποτελέσματα

Κατά τη διαδικασία προετοιμασίας έγινε τυχαίος διαχωρισμός των δειγμάτων με ποσοστό 80% για την εκπαίδευση των μοντέλων (train set) και 20% για τη δοκιμή τους

(test set). Οι υπερ-παράμετροι για κάθε μοντέλο και κάθε σενάριο πρόβλεψης παρουσιάζονται στον Πίνακα 10.

Πίνακας 10: Υπερ-παράμετροι (hyper-parameters) μοντέλων μηχανικής μάθησης

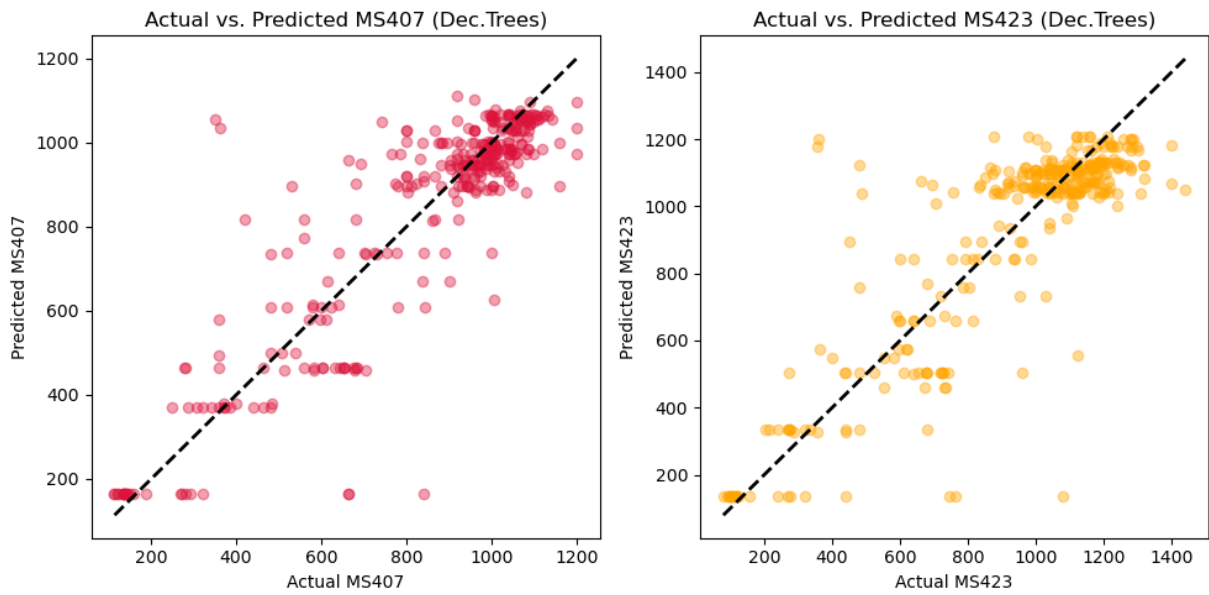
Υπερ-παράμετροι	Δέντρα Απόφασης	
	Πρόβλεψη Χρόνου Διαδρομής	Πρόβλεψη Κυκλοφοριακού φόρτου
max_depth	8	10
min_samples_split	6	2
min_samples_leaf	3	2
	GBDT	
	Πρόβλεψη Χρόνου Διαδρομής	Πρόβλεψη Κυκλοφοριακού φόρτου
n_estimators	100	100
max_depth	8	8

Μετά την εκπαίδευση οποιουδήποτε μοντέλου πρόβλεψης, απαιτείται ένα μέτρο αξιολόγησης για τις τιμές πρόβλεψης που αυτό παράγει. Στην προκειμένη περίπτωση για την αξιολόγηση χρησιμοποιήθηκαν οι μετρικές MAPE, MAE για τον έλεγχο του σφάλματος και η R^2 για τον έλεγχο προσαρμογής των μοντέλων στα δεδομένα. Ακολουθεί πίνακας με τα αποτελέσματα των παραπάνω μετρικών για κάθε περίπτωση πρόβλεψης και κάθε μοντέλο.

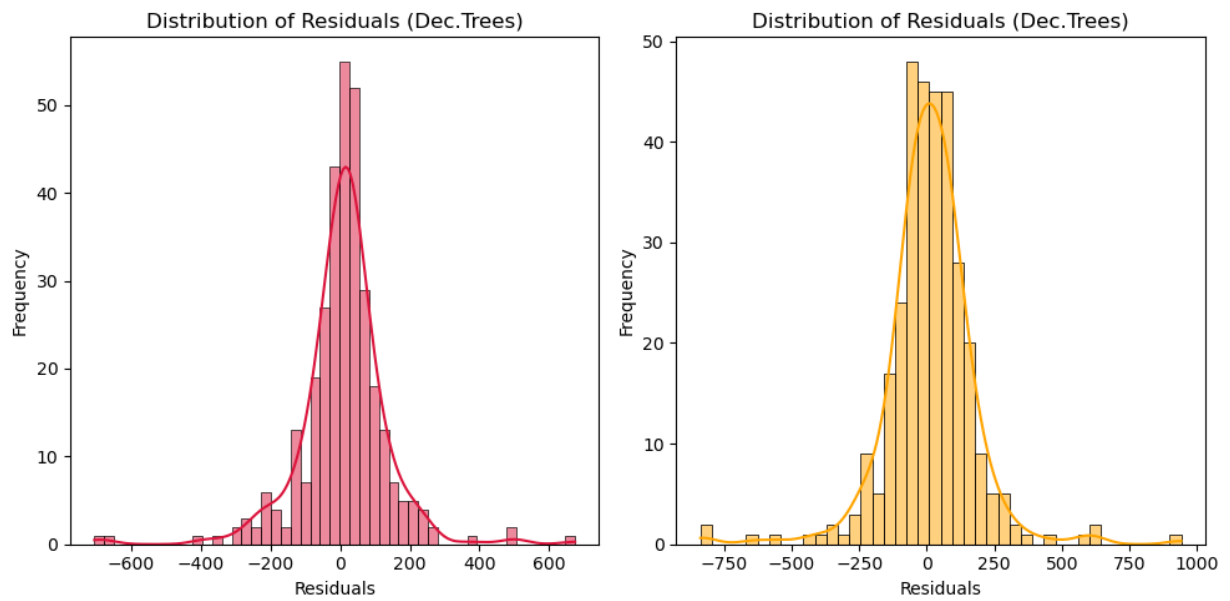
Στον Πίνακες 11 και 12 παρουσιάζονται οι μετρικές σφάλματος και ο βαθμός σημαντικότητας κάθε χαρακτηριστικού των μοντέλων μηχανικής μάθησης για την πρόβλεψη του κυκλοφοριακού φόρτου. Στην συνέχεια ακολουθούν τα διαγράμματα προβλεπόμενων-πραγματικών τιμών (Διάγραμμα 7, Διάγραμμα 9) και η κατανομή του σφάλματος (Διάγραμμα 8, Διάγραμμα 10) για κάθε φωρατή, πρώτα με το μοντέλο Δέντρων Απόφασης και στην συνέχεια με το μοντέλο GBDT.

Πίνακας 11: Αποτελέσματα πρόβλεψης κυκλοφοριακού φόρτου με Δέντρα Απόφασης

Πρόβλεψη Κυκλοφοριακού Φόρτου	Δέντρα Απόφασης
R^2	0.79
MAE	95.88
MAPE	14.55%
Βαθμός Σημαντικότητας Χαρακτηριστικών	
duration (minutes)	0.8056
min_max	0.0960
Weekday	0.0733
Rush_hour	0.0129
W+R2	0.0017
diff	0.0105



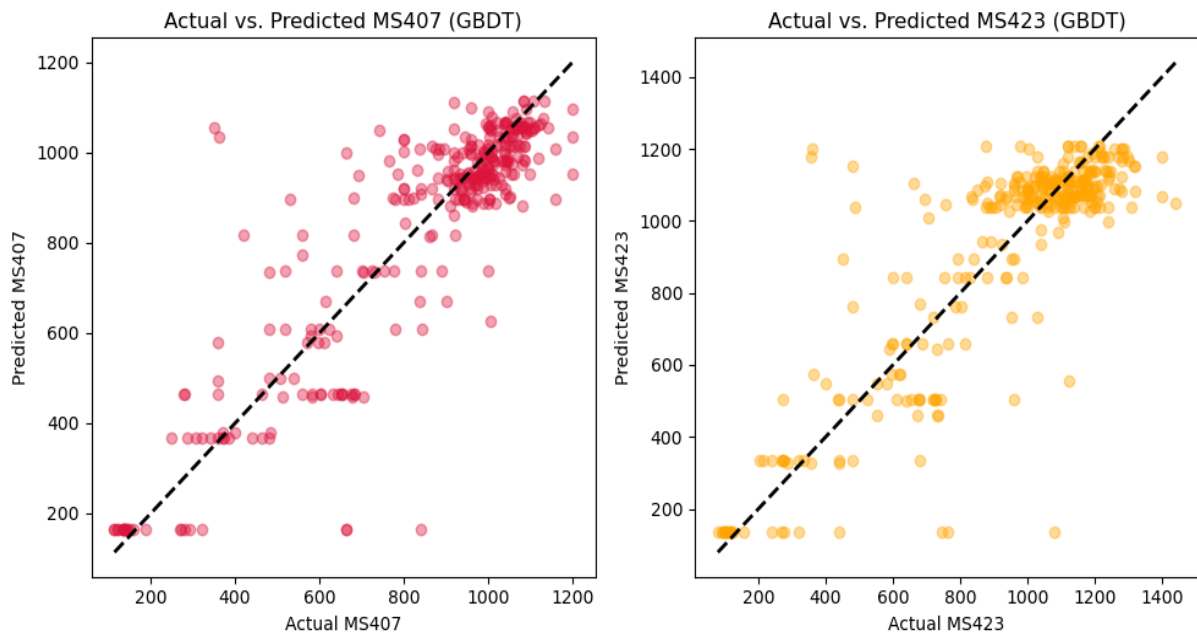
Διάγραμμα 7: Διάγραμμα διασποράς πραγματικών και προβλεπόμενων τιμών κυκλοφοριακού φόρτου (Dec.Trees)



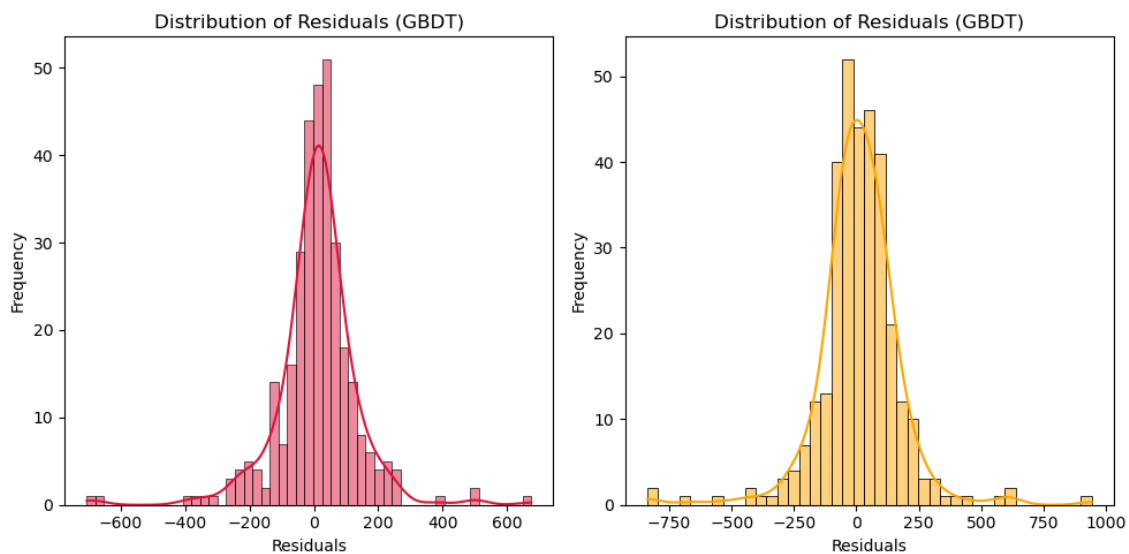
Διάγραμμα 8: Διάγραμμα κατανομής εναπομένοντος σφάλματος τιμών κυκλοφοριακού φόρτου (Dec.Trees)

Πίνακας 12: Αποτελέσματα πρόβλεψης κυκλοφοριακού φόρτου με GBDT

Πρόβλεψη Κυκλοφοριακού Φόρτου	GBDT
R ²	0.78
MAE	95.88
MAPE	14.55%
Βαθμός Σημαντικότητας Χαρακτηριστικών	
duration (minutes)	0.8092
min_max	0.0889
Weekday	0.0796
Rush_hour	0.0059
W+R2	0.0037
diff	0.0127



Διάγραμμα 9: Διάγραμμα διασποράς πραγματικών και προβλεπόμενων τιμών κυκλοφοριακού φόρτου (GBDT)

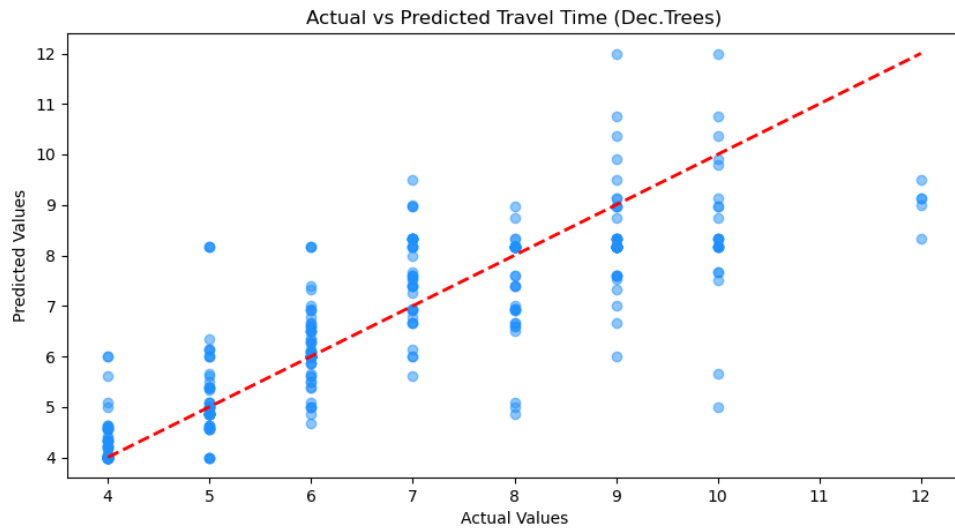


Διάγραμμα 10: Διάγραμμα κατανομής εναπομένοντος σφάλματος τιμών κυκλοφοριακού φόρτου (GBDT)

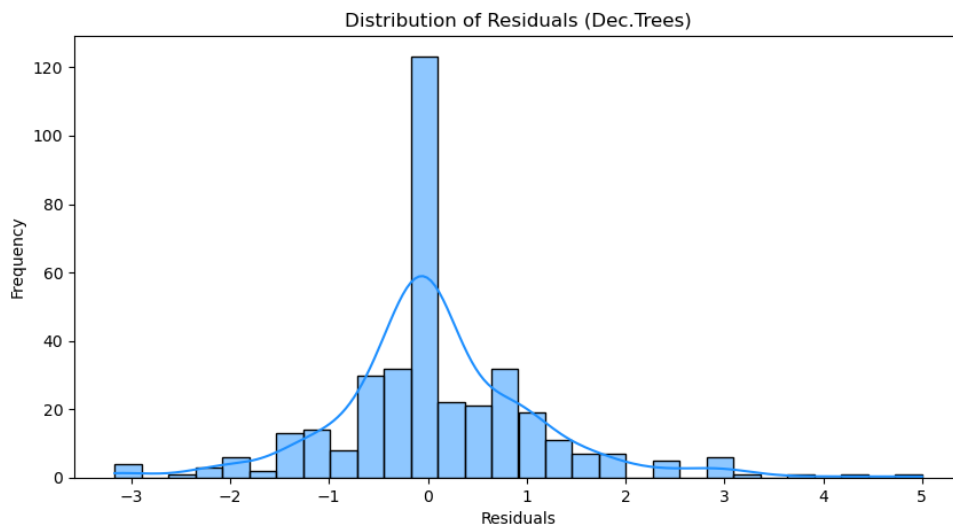
Ακολουθούν οι Πίνακες 13 και 14 με τα αποτελέσματα της πρόβλεψης του χρόνου διαδρομής με τα μοντέλα Δέντρων Απόφασης και GBDT. Το Διάγραμμα 11 και το Διάγραμμα 12 αποτελούν μια απεικόνιση των αποτελεσμάτων των πινάκων όπως και στην περίπτωση του κυκλοφοριακού φόρτου.

Πίνακας 13: Αποτελέσματα πρόβλεψης χρόνου διαδρομής με Δέντρα Απόφασης

Πρόβλεψη Χρόνου Διαδρομής	Δέντρα Απόφασης
R²	0.82
MAE	0.69
MAPE	9.75%
Βαθμός Σημαντικότητας Χαρακτηριστικών	
MS423	0.1021
MS407	0.6768
Weekday	0.1299
W+R2	0.0912



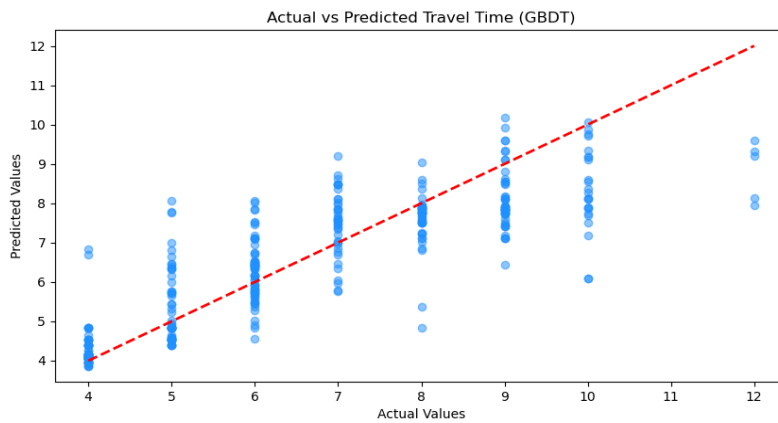
Διάγραμμα 11: Διάγραμμα διασποράς πραγματικών και προβλεπόμενων τιμών χρόνου διαδρομής (Dec.Trees)



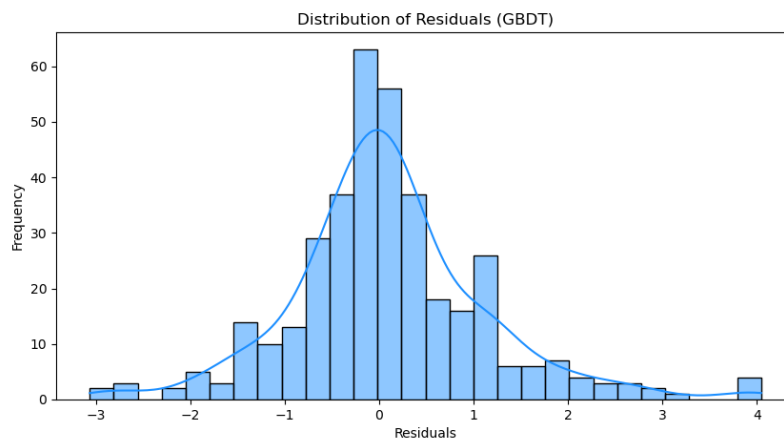
Διάγραμμα 12: Διάγραμμα κατανομής εναπομένοντος σφάλματος τιμών χρόνου διαδρομής (Dec.Trees)

Πίνακας 14: Αποτελέσματα πρόβλεψης χρόνου διαδρομής με GBDT

Πρόβλεψη Χρόνου Διαδρομής	GBDT
R²	0.76
MAE	0.69
MAPE	9.75%
Βαθμός Σημαντικότητας Χαρακτηριστικών	
MS423	0.0942
MS407	0.6865
Weekday	0.1018
W+R2	0.1175



Διάγραμμα 13: Διάγραμμα διασποράς πραγματικών και προβλεπόμενων τιμών χρόνου διαδρομής (GBDT)



Διάγραμμα 14: Διάγραμμα κατανομής εναπομένοντος σφάλματος τιμών χρόνου διαδρομής (GBDT)

Στην συνέχεια με την ίδια βάση δεδομένων πραγματοποιήθηκε η πρόγνωση των τιμών χρόνου διαδρομής στο ίδιο οδικό τμήμα με χρήση της συνάρτησης BPR. Ο μαθηματικός τύπος της συνάρτησης (1) παρουσιάστηκε σε προηγούμενα κεφάλαια της διπλωματικής εργασίας.

Για την χρήση της απαιτήθηκε ο καθορισμός των τιμών των όρων της συνάρτησης. Από τη βάση δεδομένων ήταν διαθέσιμοι οι φόρτοι οχημάτων των δύο φωρατών καθώς και οι αντίστοιχοι χρόνοι διαδρομής στο ίδιο χρονικό διάστημα. Καθώς η συνάρτηση BPR αναφέρεται σε οδικό τμήμα, χρησιμοποιήθηκε ο μέσος όρος των κυκλοφοριακών φόρτων των φωρατών που βρίσκονται στην αρχή και στο τέλος της Λ. Αλεξάνδρας.

Πίνακας 15: Βάση δεδομένων μέσου κυκλοφοριακού φόρτου

datetime_utc	duration(minutes)	MS407	MS423	avg_traffic
12/1/2022 2:00	4	193	146	169.5
12/1/2022 3:00	4	126	97	111.5
12/1/2022 4:00	4	136	101	118.5
12/1/2022 5:00	4	280	248	264
12/1/2022 6:00	4	657	735	696
12/1/2022 7:00	5	1012	1216	1114
12/1/2022 8:00	6	1099	1289	1194
12/1/2022 9:00	7	1028	1122	1075
12/1/2022 10:00	8	985	1316	1150.5
12/1/2022 11:00	8	1009	1145	1077
...
...
...
3/26/2023 12:00	6	925	966	945.5
3/26/2023 13:00	6	994	1051	1022.5
3/26/2023 14:00	7	1010	1128	1069
3/26/2023 15:00	6	900	1051	975.5
3/26/2023 16:00	5	870	959	914.5
3/26/2023 17:00	5	910	982	946
3/26/2023 18:00	5	937	1160	1048.5
3/26/2023 19:00	6	939	1119	1029
3/26/2023 20:00	6	934	1113	1023.5
3/26/2023 21:00	6	881	994	937.5

Όσον αφορά τις παραμέτρους α και β της συνάρτησης, αναφέρθηκε στη βιβλιογραφία η ύπαρξη ενδεικτικών τιμών ανάλογα τον τύπο του οδικού τμήματος. Ωστόσο, για την βέλτιστη εφαρμογή της συνάρτησης στο υπό εξέταση οδικό τμήμα αναπτύχθηκε αλγόριθμος, ο οποίος με την μέθοδο ελαχίστων τετραγώνων πραγματοποίησε την

παραμετροποίηση της συνάρτησης. Ως αρχικές τιμές των παραμέτρων α και β επιλέχθηκαν οι συνήθεις τιμές ανεξαρτήτως τύπου οδού 0.15 και 4 αντίστοιχα.

Στην συνέχεια απαιτήθηκε η εκτίμηση του χρόνου διαδρομής σε ελεύθερη ροή (T_{ff}) και η χωρητικότητα (C) του εν λόγω οδικού τμήματος. Ως χρόνος διαδρομής ελεύθερης ροής ορίστηκε η ελάχιστη τιμή χρόνου διαδρομής που παρατηρήθηκε στα δεδομένα ($T_{ff} = 4$ λεπτά).

Η θεωρητική τιμή της κυκλοφοριακής ικανότητας του εν λόγω οδικού τμήματος υπερβαίνει την μέγιστη τιμή κυκλοφοριακού φόρτου που παρατηρείται στα δεδομένα. Όμως στην πράξη, οι φωτεινοί σηματοδότες που διακόπτουν την ροή της κυκλοφορίας, η ύπαρξη αποκλειστικής λωρίδας λεωφορείων σε συνδυασμό με την στάθμευση οχημάτων στη δεξιά λωρίδα της οδού επηρεάζουν σε σημαντικό βαθμό την πραγματική τιμή ικανότητας της οδού. Ο ακριβής υπολογισμός αυτής της μεταβλητής δεν είναι εφικτός μόνο με τα διαθέσιμα δεδομένα. Γίνεται επίσης η παραδοχή ότι για κάποια χρονικά διαστήματα η ζήτηση ξεπερνάει την πραγματική κυκλοφοριακή ικανότητα της οδού ($V/C > 1$). Για την χρήση της συνάρτησης δοκιμάστηκαν αρκετές τιμές κυκλοφοριακής ικανότητας μεταξύ της μέσης και της μέγιστης τιμής φόρτου (V) που παρατηρήθηκε στη βάση δεδομένων. Ως πραγματική τιμή κυκλοφοριακής ικανότητας της οδού επιλέχθηκε η τιμή των 1050 οχημάτων, η οποία αντιστοιχεί, κατά μέσο όρο, σε χρόνο διαδρομής 10 λεπτών στη βάση δεδομένων (βλ. Πίνακα 9). Σε κάθε περίπτωση, η ακριβής τιμή της δεν φάνηκε να έχει ιδιαίτερη σημασία ως προς τα αποτελέσματα πρόβλεψης, καθώς η λογική του αλγορίθμου βελτιστοποίησης ελαχίστων τετραγώνων είναι να παράγει το μικρότερο δυνατό σφάλμα, ανεξάρτητα αν οι τιμές των όρων βρίσκονται εντός «λογικών» ορίων. Με κάθε νέα τιμή κυκλοφοριακής ικανότητας αλγόριθμος έδινε διαφορετικές τιμές στις παραμέτρους α , β και το τελικό αποτέλεσμα παρέμενε το ίδιο. Η αξία της εκτίμησης μιας όσο το δυνατόν ρεαλιστικής τιμής κυκλοφοριακής ικανότητας, έγκειται στην ανεύρεση κατάλληλων τιμών των παραμέτρων α και β για την εφαρμογή της συνάρτησης στην Λ. Αλεξάνδρας.

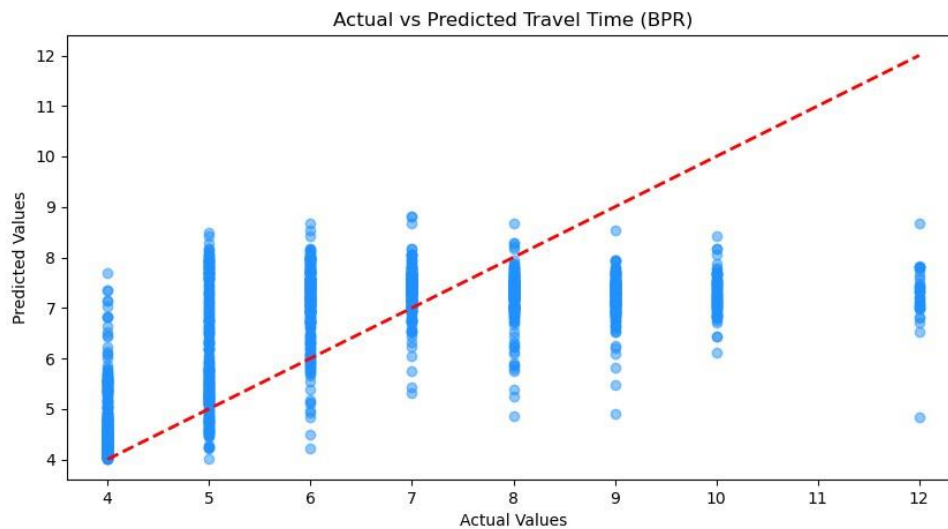
Μετά τον καθορισμό όλων των παραπάνω όρων της συνάρτησης BPR, ο αλγόριθμος, χρησιμοποιώντας τη μέθοδο ελαχίστων τετραγώνων, υπολόγισε την τελική τιμή των παραμέτρων α και β , ώστε να ελαχιστοποιείται το σφάλμα μεταξύ προβλεπόμενων τιμών της συνάρτησης και πραγματικών τιμών χρόνου διαδρομής.

Για την αξιολόγηση του αλγορίθμου χρησιμοποιήθηκαν οι ίδιες μετρικές σφάλματος και προσαρμογής στα δεδομένα. Ακολουθεί ο Πίνακας 16 με τα αποτελέσματα.

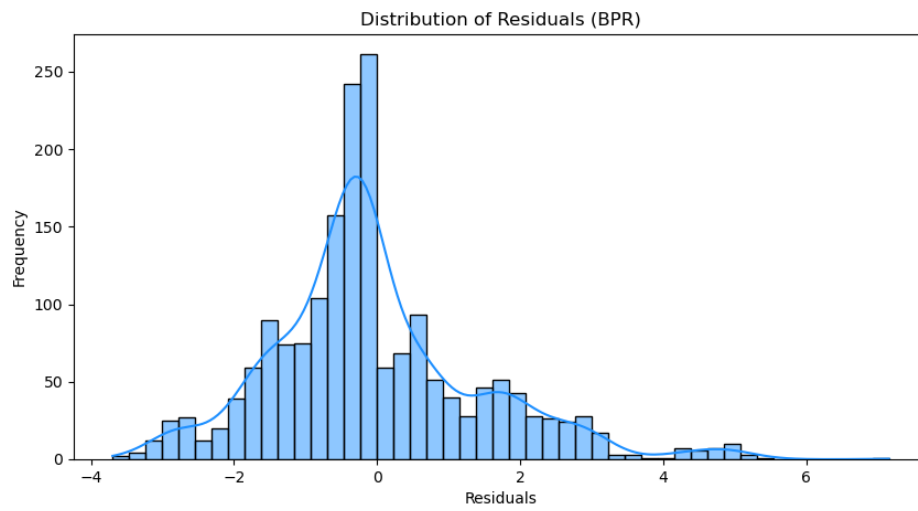
Πίνακας 16: Αποτελέσματα πρόβλεψης χρόνου διαδρομής με την συνάρτηση BPR

Πρόβλεψη Χρόνου Διαδρομής	BPR
R^2	0.48
MAE	1.08
MAPE	16.78%
Τελικές Τιμές Παραμέτρων	
α	0.90
β	1.75

Στο Διάγραμμα 15 παρουσιάζεται η διασπορά των πραγματικών και των προβλεπόμενων τιμών χρόνου διαδρομής του μοντέλου της συνάρτησης BPR. Ενώ στο Διάγραμμα 16 απεικονίζεται η κατανομή του εναπομένοντος σφάλματος.



Διάγραμμα 15: Διάγραμμα διασποράς πραγματικών και προβλεπόμενων τιμών χρόνου διαδρομής (BPR)



Διάγραμμα 16: Διάγραμμα κατανομής εναπομένοντος σφάλματος τιμών χρόνου διαδρομής (BPR)

Κεφάλαιο 5

Συμπεράσματα και Προτάσεις

5.1 Σύνοψη Αποτελεσμάτων

Σύμφωνα με Πίνακα 17 συμπεραίνεται ότι τα μοντέλα μηχανικής μάθησης παρουσιάζουν σε όλες τις περιπτώσεις ικανοποιητικό μέσο απόλυτο ποσοστιαίο σφάλμα (MAPE). Στο πλαίσιο της διπλωματικής εργασίας, βάσει του πλήθους των διαθέσιμων δεδομένων αλλά και παρόμοιων εφαρμογών πρόβλεψης της βιβλιογραφίας, θεωρήθηκε μέσο απόλυτο ποσοστιαίο σφάλμα της τάξης του 20% ως αποδεκτό όριο. Το σφάλμα πρόβλεψης του κυκλοφοριακού φόρτου έφτασε σε ικανοποιητικό επίπεδο μόνο μετά την ομαλοποίηση της βάσης δεδομένων και την αφαίρεση ακραίων τιμών. Το χαμηλότερο σφάλμα παρουσιάζεται στην πρόβλεψη του χρόνου διαδρομής με μεθόδους μηχανικής μάθησης. Όπως αναφέρθηκε και προηγουμένως, σε αυτό το σενάριο πρόβλεψης η σχέση των μεταβλητών αποτυπώθηκε εύκολα από τα μοντέλα χωρίς πολλές τροποποιήσεις στην βάση δεδομένων και δοκιμές χαρακτηριστικών (features). Ακόμα, διαπιστώνεται ότι οι υψηλές τιμές του R^2 (> 0.60) σε κάθε σενάριο πρόβλεψης επιβεβαιώνουν την χρησιμότητα των μοντέλων μηχανικής μάθησης ως εργαλεία πρόβλεψης.

Πίνακας 17: Συγκεντρωτικός πίνακας αποτελεσμάτων

Πρόβλεψη Κυκλοφοριακού Φόρτου			Πρόβλεψη Χρόνου Διαδρομής			
Μοντέλο	Δέντρα Απόφασης	GBDT	Μοντέλο	Δέντρα Απόφασης	GBDT	BPR
R^2	0.79	0.80	R^2	0.82	0.76	0.48
MAE	95.88	95.88	MAE	0.69	0.69	1.08
MAPE	14.55%	14.55%	MAPE	9.75%	9.75%	16.78%
Βαθμός Σημαντικότητας Χαρακτηριστικών			Βαθμός Σημαντικότητας Χαρακτηριστικών			
duration (min)	0.8056	0.8092	MS423	0.1021	0.0942	
min_max	0.0960	0.0889	MS407	0.6768	0.6865	
Weekday	0.0733	0.0796	Weekday	0.1299	0.1018	
Rush hour	0.0129	0.0059	W+R2	0.0912	0.1175	
W+R2	0.0017	0.0037				
diff	0.0105	0.0127				

Τα δύο μοντέλα μηχανικής μάθησης δεν φαίνεται να εμφανίζουν ιδιαίτερες διαφορές στην μεταξύ τους απόδοση, τόσο στην ακρίβεια πρόβλεψης όσο και στην προσαρμογή τους στα δεδομένα, ανεξαρτήτως της μεταβλητής πρόβλεψης. Τέλος, όπως αναμενόταν ο χρόνος διαδρομής αποτέλεσε το χαρακτηριστικό με τη μεγαλύτερη επιρροή στο αποτέλεσμα πρόβλεψης του κυκλοφοριακού φόρτου. Το αντίστροφο συνέβη στην

περίπτωση του χρόνου διαδρομής, όπου τα δεδομένα φόρτου του φωρατή 407 αποτέλεσαν την σημαντικότερη μεταβλητή.

Η συνάρτηση BPR παρουσίασε επίσης αποδεκτό μέσο απόλυτο ποσοστιαίο σφάλμα κατά την πρόβλεψη του χρόνου διαδρομής, όμως υψηλότερο σε σχέση με τις μεθόδους μηχανικής μάθησης. Από την τιμή της μετρικής R^2 διαπιστώνεται η συνάρτηση BPR δεν είναι ικανή να αποτυπώσει με ακρίβεια τα μοτίβα που παρατηρούνται στα δεδομένα. Αυτό επιβεβαιώνεται και από τα σχετικά διαγράμματα σφάλματος που παρουσιάστηκαν προηγουμένως. Ιδιαίτερα εμφανής είναι η αδυναμία της συνάρτησης να προβλέψει τον χρόνο διαδρομής στις μέγιστες τιμές της βάσης δεδομένων (9-12 λεπτά).

5.2 Βασικά Συμπεράσματα

Μετά και την σύνοψη των αποτελεσμάτων της παρούσας διπλωματικής εργασίας μπορούν να εξαχθούν τα παρακάτω συμπεράσματα.

Τα πρότυπα μηχανικής μάθησης που αναπτύχθηκαν προσέφεραν υψηλή ικανότητα πρόβλεψης των ζητούμενων κυκλοφοριακών μεταβλητών σε κάθε περίπτωση, παρά την απλότητα τους. Η μέθοδος πρόβλεψης μέσω της συνάρτησης BPR αποδείχτηκε υποδεέστερη επιλογή σε σχέση με τις μεθόδους μηχανικής πρόβλεψης όσον αφορά την ακρίβεια πρόβλεψης αλλά και την προσαρμογή στα δεδομένα. Χωρίς επαρκή δεδομένα για τον υπολογισμό της κυκλοφοριακής ικανότητας, η εφαρμογή της συνάρτησης BPR παρουσιάζει δυσκολίες. Ακόμα και με την μέθοδο βελτιστοποίησης ελαχίστων τετραγώνων που προτάθηκε στην παρούσα διπλωματική εργασία για την παράκαμψη αυτού του προβλήματος, οι προβλέψεις χρόνου διαδρομής παρουσίασαν μεγαλύτερο σφάλμα σε σχέση με τις μεθόδους μηχανικής μάθησης.

Επομένως κυκλοφοριακά δεδομένα τα οποία συλλέγονται δυσκολότερα θα μπορούσαν να εκτιμηθούν μέσω παρόμοιων μεθόδων. Στην περίπτωση της διπλωματικής εργασίας, τιμές κυκλοφοριακού φόρτου ενός οδικού τμήματος προέκυψαν μόνο με γνώση εύκολα προσβάσιμων δεδομένων χρόνου διαδρομής. Επίσης, τα μοντέλα επέδειξαν υψηλή ακρίβεια πρόβλεψης των φόρτων στα σημεία εισόδου και εξόδου του οδικού τμήματος ταυτόχρονα, χωρίς να χρειαστεί να χρησιμοποιηθούν τα δεδομένα κάποιου φωρατή ως χαρακτηριστικό (feature) για τα μοντέλα. Γίνεται κατανοητό ότι απλά μοντέλα, λίγων μεταβλητών μπορούν να αποτελέσουν χρήσιμο εργαλείο χωρίς αυτό να σημαίνει ότι αποτελούν την βέλτιστη στρατηγική για κάθε πρόβλημα πρόγνωσης κυκλοφοριακών συνθηκών.

Όσον αφορά τη διαδικασία ανάπτυξης των μοντέλων μηχανικής μάθησης, η επιλογή και διαμόρφωση των κατάλληλων χαρακτηριστικών (feature selection & engineering) ήταν σημαντική για την εξασφάλιση υψηλής απόδοσης. Ταυτόχρονα, μέσω της σημαντικότητας των χαρακτηριστικών (feature importance) έγινε κατανοητό ποι

παράγοντες επηρεάζουν σημαντικά την συσχέτιση των δύο κυκλοφορικών μεταβλητών. Ωστόσο, έγινε αντιληπτό ότι βασικό παράγοντα για την επιτυχία των μοντέλων μηχανικής μάθησης αποτελεί η ποιότητα, το πλήθος και η μορφή των δεδομένων. Στην περίπτωση πρόβλεψης του κυκλοφοριακού φόρτου, η μικρή διακύμανση των τιμών χρόνου διαδρομής δημιούργησε πρόβλημα στην προσαρμογή των μοντέλων, το οποίο επιλύθηκε με τη διαδικασία της ομαλοποίησης (normalization) και την αφαίρεση ακραίων τιμών (outliers).

Τέλος, διαπιστώνεται ότι η πρόοδος της τεχνολογίας έχει θέσει τα απαραίτητα θεμέλια για την ανάπτυξη των ευφυών συστημάτων διαχείρισης της κυκλοφορίας στα αστικά κέντρα. Η πληροφορία που αντλείται από τους ίδιους χρήστες ενός οδικού δικτύου, όπως στην περίπτωση της διπλωματικής εργασίας, συμβάλλει στην πρόγνωση κυκλοφοριακών συνθηκών και τελικά, την προσφορά καλύτερου επιπέδου εξυπηρέτησης. Για το λόγο αυτό τα αποτελέσματα της παρούσας διπλωματικής εργασίας αποκτούν ιδιαίτερη σημασία.

5.3 Περιορισμοί Διπλωματικής Εργασίας και Προτάσεις Έρευνας

Με βάση όλα τα παραπάνω θα μπορούσαν να επισημανθούν οι παρακάτω περιορισμοί των αποτελεσμάτων και κάποιες προτάσεις για έρευνα.

Δεν είναι γνωστό αν τα μοντέλα που χρησιμοποιήθηκαν στην διπλωματική εργασία μπορούν να επαναλάβουν παρόμοια ικανότητα πρόβλεψης σε άλλα οδικά τμήματα. Με εφαρμογή παρόμοιων μεθοδολογιών και μοντέλων σε διαφορετικά οδικά τμήματα μπορεί να εξακριβωθεί αυτό. Επίσης, η γεωμετρία και οι χωρικές συσχετίσεις μεταξύ των οδικών τμημάτων αυξάνουν τον βαθμό πολυπλοκότητας όσον αφορά την πρόβλεψη κυκλοφοριακών συνθηκών σε επίπεδο δικτύου. Όπως αναφέρθηκε και στο στάδιο ανάλυσης της βιβλιογραφίας, οι περισσότερες έρευνες επικεντρώνονται στο λιγότερο σύνθετο πρόβλημα της πρόβλεψης κυκλοφοριακών μεγεθών ενός οδικού τμήματος. Η έμφαση στη δημιουργία μοντέλων ικανών να κατανοήσουν την χωρική συσχέτιση οδικών αξόνων, ίσως επιτρέψει ακριβέστερες προβλέψεις σε επίπεδο δικτύου. Παρόμοια, το ζήτημα της πρόβλεψης κυκλοφοριακών συνθηκών διευρύνεται με την προσθήκη και άλλων μεταβλητών. Καιρικά δεδομένα, στατιστικά στοιχεία ατυχημάτων, η σήμανση της οδού μεταξύ άλλων, θα μπορούσαν να εισαχθούν σε μοντέλα μηχανικής μάθησης και να συνδυαστούν με τις γνωστές κυκλοφοριακές παραμέτρους.

Τα μοντέλα που αναπτύχθηκαν στην παρούσα διπλωματική εργασία θα μπορούσαν να βελτιωθούν ως προς τις παρακάτω κατευθύνσεις: εφαρμογή σε περισσότερα οδικά τμήματα, αξιοποίηση μεγαλύτερου όγκου και περισσότερων ειδών δεδομένων, διαμόρφωση περισσότερων συνδυασμών χαρακτηριστικών (features), πειραματισμός με άλλα μοντέλα μηχανικής μάθησης και σύγκρισή τους, συνδυασμός με μοντέλα βαθιάς

μηχανικής μάθησης (deep learning) με σκοπό την παραγωγή ενός υβριδικού μοντέλου. Κατά την ίδια λογική, ένα γενικό πλαίσιο μεθόδων παραμετροποίησης και αυστηρού προσδιορισμού των όρων και των παραδοχών των συναρτήσεων φόρτου-καθυστερήσης θα αύξανε σημαντικά την χρησιμότητα τους ως εργαλεία ανάλυσης της κυκλοφορίας.

Με την ραγδαία ανάπτυξη της τεχνολογίας, είναι προφανές ότι η εκμετάλλευση της όλο και αυξανόμενης πληροφορίας και η ανάπτυξη καινοτόμων εφαρμογών κυκλοφοριακής τεχνικής θα πρέπει να βρίσκεται στο επίκεντρο της έρευνας και της διαχείρισης των μεταφορικών συστημάτων.

Βιβλιογραφία

- Barka, R. E., & Politis, I. (2021). Fitting Volume Delay Functions under interrupted and uninterrupted flow conditions at Greek urban roads. *European Transport - Trasporti Europei*, 83. <https://doi.org/10.48295/ET.2021.83.7>
- Barros, J., Araujo, M., & Rossetti, R. J. F. (2015). Short-term real-time traffic prediction methods: A survey. *2015 International Conference on Models and Technologies for Intelligent Transportation Systems, MT-ITS 2015*, 132–139. <https://doi.org/10.1109/MTITS.2015.7223248>
- Behrooz, H., & Hayeri, Y. M. (2022). Machine Learning Applications in Surface Transportation Systems: A Literature Review. *Applied Sciences (Switzerland)*, 12(18). <https://doi.org/10.3390/app12189156>
- Cheng, J., Li, G., & Chen, X. (2019). Research on travel time prediction model of freeway based on gradient boosting decision tree. *IEEE Access*, 7, 7466–7480. <https://doi.org/10.1109/ACCESS.2018.2886549>
- IEEE Staff. (2018). *2018 International Conference on Information, Communication, Engineering and Technology (ICICET)*. IEEE.
- Janković, S., Uzelac, A., Zdravković, S., Mladenović, D., Mladenović, S., & Andrijanić, I. (n.d.). TRAFFIC VOLUMES PREDICTION USING BIG DATA ANALYTICS METHODS. *International Journal for Traffic and Transport Engineering*, 2021(2), 184–198. [https://doi.org/10.7708/ijtte2021.11\(2\).01](https://doi.org/10.7708/ijtte2021.11(2).01)
- Laña, I., Sanchez-Medina, J. J., Vlahogianni, E. I., & Ser, J. Del. (2021). From data to actions in intelligent transportation systems: A prescription of functional requirements for model actionability. *Sensors (Switzerland)*, 21(4), 1–34. <https://doi.org/10.3390/s21041121>
- Liu, Y., Zhang, N., Luo, X., & Yang, M. (2021). Traffic Flow Forecasting Analysis based on Two Methods. *Journal of Physics: Conference Series*, 1861(1). <https://doi.org/10.1088/1742-6596/1861/1/012042>
- Liu, Z., Li, Z., Wu, K., & Li, M. (2018). Urban Traffic Prediction from Mobility Data Using Deep Learning. *IEEE Network*, 32(4), 40–46. <https://doi.org/10.1109/MNET.2018.1700411>
- Maerivoet, S., & De Moor, B. (2005). *Transportation Planning and Traffic Flow Models*.
- Mahdi, M. B., Alrawi, A. K., & Leong, L. V. (2022). Compatibility between delay functions and highway capacity manual on Iraqi highways. *Open Engineering*, 12(1), 359–372. <https://doi.org/10.1515/eng-2022-0022>
- Manibardo, E. L., Lana, I., & Ser, J. Del. (2022). Deep Learning for Road Traffic Forecasting: Does it Make a Difference? *IEEE Transactions on Intelligent Transportation Systems*, 23(7), 6164–6188. <https://doi.org/10.1109/TITS.2021.3083957>

- Marshall, N. L. (2018). Forecasting the impossible: The status quo of estimating traffic flows with static traffic assignment and the future of dynamic traffic assignment. *Research in Transportation Business and Management*, 29, 85–92.
<https://doi.org/10.1016/j.rtbm.2018.06.002>
- Mehbub Anwar, A. H. M., Fujiwara, A., & Zhang, J. (n.d.). *Estimation of BPR Function for Dhaka City by Incorporating the Influence of Illegal On-Street Parking and Street-Occupancy*.
- Pan, Y., Guo, J., Chen, Y., Abbasi, M., List, G., & Zhou, X. (n.d.). *A review of volume-delay functions: Connecting theoretical fundamental, practical deployment and emerging applications A review of volume-delay functions: Connecting theoretical*.
<https://ssrn.com/abstract=4143355>
- Petrik, O., Moura, F., & De Abreu E Silva, J. (2014). The influence of the volume-delay function on uncertainty assessment for a four-step model. *Advances in Intelligent Systems and Computing*, 262, 293–306. https://doi.org/10.1007/978-3-319-04630-3_22
- Qiu, B., & Fan, W. (2021). Machine learning based short-term travel time prediction: Numerical results and comparative analyses. *Sustainability (Switzerland)*, 13(13).
<https://doi.org/10.3390/su13137454>
- Saberi, M., & Student, G. (n.d.). *A Study of Freeway Volume-to-Capacity Ratio Based Travel Time Approximations Using Archived 1 Loop Detector Data 2 3*.
- Suh, S., & John Kim, T. (1990). A HIGHWAY CAPACITY FUNCTION IN KOREA: MEASUREMENT AND CALIBRATION. In *Tmnspn. Res.-A* (Vol. 3).
- TRB. (2011). *Dynamic Traffic Assignment: A Primer*. www.TRB.org
- Vlahogianni, E. I., Karlaftis, M. G., & Golias, J. C. (2014). Short-term traffic forecasting: Where we are and where we're going. *Transportation Research Part C: Emerging Technologies*, 43, 3–19. <https://doi.org/10.1016/j.trc.2014.01.005>
- Wang, Y., Zhang, D., Liu, Y., Dai, B., & Lee, L. H. (2019). Enhancing transportation systems via deep learning: A survey. In *Transportation Research Part C: Emerging Technologies* (Vol. 99, pp. 144–163). Elsevier Ltd. <https://doi.org/10.1016/j.trc.2018.12.004>
- Wu, X. B. (n.d.). *Characterization and calibration of volume-to-capacity ratio in volume-delay functions on freeways based on a queue analysis approach Food supply chain management View project DTALite 2.0 strategic and semi-dynamic transportation flow assignment and simulation View project*. <https://www.researchgate.net/publication/347967287>
- Yang, S., Wu, J., Du, Y., He, Y., & Chen, X. (2017). Ensemble Learning for Short-Term Traffic Prediction Based on Gradient Boosting Machine. *Journal of Sensors*, 2017.
<https://doi.org/10.1155/2017/7074143>

Yuan, T., Da Rocha Neto, W., Rothenberg, C. E., Obraczka, K., Barakat, C., & Turletti, T. (2022). Machine learning for next-generation intelligent transportation systems: A survey. *Transactions on Emerging Telecommunications Technologies*, 33(4).
<https://doi.org/10.1002/ett.4427>

Zhang, Y., & Haghani, A. (2015). A gradient boosting method to improve travel time prediction. *Transportation Research Part C: Emerging Technologies*, 58, 308–324.
<https://doi.org/10.1016/j.trc.2015.02.019>