

ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
Σχολή Πολιτικών Μηχανικών
Τομέας Μεταφορών και Συγκοινωνιακής Υποδομής



Εκτίμηση φόρτου σε αστικό οδικό δίκτυο μέσω μοντέλων μηχανικής μάθησης



Διπλωματική Εργασία
Κωνσταντίνα Τζαβάρα
Επιβλέπουσα Καθηγήτρια: Ελένη Βλαχογιάννη,
Καθηγήτρια Σχολής Πολιτικών Μηχανικών ΕΜΠ

Αθήνα, Οκτώβριος 2024

Ευχαριστίες

Η παρούσα διπλωματική εργασία σηματοδοτεί το πέρας των προπτυχιακών μου σπουδών στην σχολή Πολιτικών Μηχανικών του Εθνικού Μετσόβιου Πολυτεχνείου.

Αρχικά, θα ήθελα να ευχαριστήσω την κα Ελένη Βλαχογιάννη που μου προσέφερε την ευκαιρία να ασχοληθώ με αυτό το θέμα έρευνας καθώς και για τις καίριες συμβουλές και την καθοδήγησή της στα στάδια υλοποίησής της.

Επίσης, θα ήθελα να ευχαριστήσω τον Υποψήφιο Διδάκτορα Χάρη Χαλκιαδάκη για την πολύτιμη βοήθεια και τον χρόνο που μου προσέφερε, καθώς και το εξαιρετικό κλίμα συνεργασίας, συμβάλλοντας ουσιαστικά στην υλοποίηση της παρούσας εργασίας.

Τέλος θα ήθελα να ευχαριστήσω την οικογένειά μου και τους φίλους μου που με στήριξαν σε όλη την διάρκεια των σπουδών μου.

Κωνσταντίνα Τζαβάρια
Αθήνα, Οκτώβριος 2024

Τίτλος : Εκτίμηση φόρτου σε αστικό οδικό δίκτυο μέσω μοντέλων μηχανικής μάθησης

Κωνσταντίνα Τζαβάρα

Επιβλέπουσα Καθηγήτρια: Ελένη Ι. Βλαχογιάννη

Σύνοψη

Τα μοντέλα μηχανικής μάθησης αποτελούν σημαντική καινοτομία στην διαχείριση της κυκλοφορίας και τη βελτίωση των υποδομών. Με την χρήση των μοντέλων αυτών είναι δυνατή η ανάλυση μεγάλου όγκου δεδομένων σε πραγματικό χρόνο και με αποτέλεσμα την πρόβλεψη των κυκλοφοριακών ροών και βελτίωση των κυκλοφοριακών ρυθμίσεων. Σκοπός της συγκεκριμένης έρευνας είναι να δημιουργηθούν μοντέλα πρόβλεψης του κυκλοφοριακού φόρτου και του χρόνου διαδρομής με τη χρήση αλγορίθμων μηχανικής μάθησης σε οδικά δίκτυα της Αθήνας. Για τον σκοπό αυτό, αντλήθηκαν δεδομένα από κόμβους της Αθήνας που χαρακτηρίζονται για τους υψηλούς κυκλοφοριακούς φόρτους μέσω αισθητήρων που βρίσκονται σε διάφορα σημεία του οδικού δικτύου (φωρατές). Στη συνέχεια, έπειτα από επεξεργασία τους αποτέλεσαν δεδομένα εισόδου σε μοντέλα μηχανικής μάθησης. Από τους αλγορίθμους που χρησιμοποιήθηκαν φάνηκε ότι πιο αποτελεσματικός στη δημιουργία ενός σταθερού και αποδοτικού μοντέλου που περιγράφει τις πραγματικές συνθήκες είναι τα 'Ενισχυμένα Δένδρα Απόφασης' (Gradient Boosting).

Λέξεις Κλειδιά: Κυκλοφοριακός φόρτος, Χρόνος Διαδρομής, Μηχανική Μάθηση, Μοντέλα Πρόβλεψης, Τυχαία Δάση, Ενισχυμένα Δέντρα Απόφασης

Title : Estimation of Traffic Load in Urban Road Networks Using Machine Learning Models

Konstantina Tzavara

Supervisor: Professor Eleni I. Vlachogianni

ABSTRACT

Machine learning models represent a significant innovation in traffic management and infrastructure improvement. These models enable the real-time analysis of large volumes of data, leading to more accurate traffic flow predictions and enhanced traffic regulation. The aim of this diploma thesis is to develop utilize machine learning algorithms to create effective and robust models that reflect real-world traffic conditions. To achieve this, data were collected from high-traffic nodes in Athens through sensors located at various points of the road network (detectors). After processing, these data served as input for machine learning models. Among the algorithms tested, Gradient Boosting Decision Trees demonstrated the highest effectiveness in producing stable and efficient predictive model.

Keywords: traffic load travel time, machine learning, prediction models, Random Forests, Gradient Boosting Trees

Περίληψη

Η πρόβλεψη της κυκλοφοριακής ροής είναι κρίσιμη για την αποτελεσματική διαχείριση των σύγχρονων συστημάτων μεταφορών. Παραδοσιακές μέθοδοι εκτίμησης όπως είναι οι συναρτήσεις φόρτου-καθυστερήσης (Volume Delay Functions) και χρόνου-διαδρομής (Travel Time Functions) έχουν εφαρμοστεί, αλλά συχνά παράγουν ανακριβείς εκτιμήσεις λόγω της πολύπλοκης και μη γραμμικής φύσης των παραμέτρων που επηρεάζουν την κυκλοφορία. Με την εξέλιξη της τεχνολογίας και τη χρήση αλγορίθμων μηχανικής μάθησης είναι δυνατή η ανάπτυξη ακριβέστερων προγνωστικών μοντέλων κυκλοφοριακής ροής στο οδικό δίκτυο. Η παρούσα διπλωματική εργασία επικεντρώνεται στην πρόβλεψη του κυκλοφοριακού φόρτου σε οδικά δίκτυα της Αθήνας με την χρήση μοντέλων μηχανικής μάθησης. Πιο συγκεκριμένα, χρησιμοποιήθηκαν ιστορικά δεδομένα χρόνου διαδρομής από την πλατφόρμα Google Maps και πληροφορίες φόρτου κυκλοφορίας από αισθητήρες που βρίσκονται σε διάφορα σημεία του οδικού δικτύου (φωρατές) κατά την περίοδο του Μαρτίου για διάφορα χρονικά διαστήματα για κάθε ημέρα λήψης δεδομένων, με σκοπό τη δημιουργία μοντέλων πρόβλεψης, ύστερα από εκπαίδευση αλγορίθμων επιβλεπόμενης μηχανικής μάθησης. Αναλυτικότερα, αξιολογήθηκε η απόδοση διάφορων αλγορίθμων για την πρόβλεψη των τιμών του κυκλοφοριακού φόρτου. Βρέθηκε ότι ο πιο αποτελεσματικός αλγόριθμος που έχει την ικανότητα να δημιουργεί ένα δυναμικό και σταθερό μοντέλο πρόβλεψης είναι τα "Ενισχυμένα Δένδρα Αποφάσεων" (Gradient Boosting). Επιπλέον, στην παρούσα εργασία αποδεικνύεται η ικανότητα ακριβέστερης πρόβλεψης εξαρτημένων κυκλοφοριακών μεταβλητών με τη χρήση των κατάλληλων ανεξάρτητων μεταβλητών, οι οποίες περιγράφουν σε μεγάλο βαθμό τις πραγματικές οδικές συνθήκες, σύμφωνα με τα αποτελέσματα σχετικού διαγράμματος SHAP. Ακόμη, τονίζονται τα οφέλη της χρήσης μοντέλων μηχανικής μάθησης, τα οποία προσαρμόζονται εύκολα σε διάφορες κυκλοφοριακές καταστάσεις και μπορούν να βελτιώσουν συνολικά τη διαχείριση και την ασφάλεια των μεταφορών. Τα αποτελέσματα δείχνουν πως μεταβλητές που δηλώνουν το είδος του οδικού δικτύου επιδρούν σε μικρότερο βαθμό στην πρόβλεψη της κυκλοφοριακής ροής ενώ ο αριθμός των λωρίδων που έχει ένα δίκτυο επιδρά σε μεγαλύτερο βαθμό. Συμπερασματικά, από αυτήν την εργασία υπογραμμίζεται η σημασία της ενσωμάτωσης προηγμένων τεχνικών μηχανικής μάθησης στο πλαίσιο πρόβλεψης της κυκλοφορίας για τη

βελτίωση της διαχείρισης των μεταφορών. Σε μελλοντικές έρευνες, η ενσωμάτωση περισσότερων παραμέτρων για τον προσδιορισμό της κυκλοφοριακής ροής, όπως οδικές σημάνσεις, στατιστικά ατυχημάτων και καιρικές συνθήκες, θα μπορέσει να οδηγήσει στην βελτίωση της ακρίβειας των αποτελεσμάτων των μοντέλων πρόβλεψης. Σε αυτό το πλαίσιο η βελτίωση των τεχνικών επεξεργασίας δεδομένων όπως ο καθορισμός και η κανονικοποίηση, θα βοηθήσει στην αντιμετώπιση του «θορύβου» στα δεδομένα, ενώ η εφαρμογή μεθόδων ελέγχου υπερπροσαρμογής θα μπορούσε να ενισχύσει την γενικευσιμότητα των μοντέλων.

Πίνακας Περιεχομένων

Εισαγωγή	1
1.1. Το Κυκλοφοριακό πρόβλημα	1
1.2. Σκοπός.....	2
1.3. Περιεχόμενα Ενοτήτων	3
Κεφάλαιο 2	4
Βιβλιογραφική Ανασκόπηση.....	4
2.1. Πρόβλεψη Κυκλοφορίας	4
2.1.1. Οι καιρικές συνθήκες	5
2.1.2. Η χωρητικότητα του οδικού δικτύου.....	6
2.1.3. Περίοδοι αιχμής	8
2.2. Μέθοδοι Πρόβλεψης Οδικής Κυκλοφορίας	9
2.3. Συναρτήσεις Φόρτου-Καθυστερήσης	10
2.4 Τεχνητή Νοημοσύνη και Μηχανική Μάθηση.....	133
2.5. Συμπεράσματα της Βιβλιογραφίας	177
Κεφάλαιο 3	199
3.1 Ροή εργασιών	199
3.2. Μηχανική Μάθηση	20
3.3. Μετρικές Αξιολόγησης.....	2121
3.4. Αλγόριθμοι Μηχανικής Μάθησης.....	244
3.4.1. Γραμμική Παλινδρόμηση.....	304
3.4.2. Παλινδρόμηση Κορυφογραμμών	255
3.4.3. Παλινδρόμηση Lasso.....	255
3.4.4. Ελαστικά δίκτυα.....	266
3.4.5. Τυχαία Δάση	277
3.4.6. Μηχανές Υποστήριξης Παλινδρόμησης.....	288
3.4.7. Ενισχυμένα Δένδρα Απόφασης.....	309
3.5. Προσδιορισμός Υπερπαραμέτρων	30
Κεφάλαιο 4	344
4.1. Εισαγωγή.....	344

4.2. Λήψη και Διαχείριση δεδομένων	344
4.3. Ανάλυση δεδομένων	355
4.4. Επεξεργασία και δόμηση δεδομένων	376
4.5. Αποτελέσματα	3939
Κεφάλαιο 5	433
Συμπεράσματα.....	433
5.1. Βασικά Συμπεράσματα.....	433
5.2 Περιορισμοί Έρευνας.....	444
5.3 Προτάσεις για Περαιτέρω Έρευνα.....	444
Βιβλιογραφία	466

Ευρετήριο Πινάκων

Πίνακας 1: Τιμές παραμέτρων α και β της συνάρτησης BPR για συγκεκριμένες ταχύτητες οχημάτων σε αυτοκινητόδρομο και δρόμο με πολλαπλές λωρίδες που δεν είναι όμως αυτοκινητόδρομος.....	10
Πίνακας 2: Περιγραφή των υπερπαραμέτρων που χρησιμοποιήθηκαν σε όλα τα μοντέλα μηχανικής μάθησης.....	33
Πίνακας 3: Δομή δεδομένων για δημιουργία μοντέλων μηχανικής μάθησης	35
Πίνακας 4: Απόδοση του καλύτερου μοντέλου που χρησιμοποιήθηκε στην παρούσα έρευνα σύμφωνα με τη μέθοδο βελτιστοποίησης Random Search ($y = \text{flow_per_lane}$).....	39
Πίνακας 5: Οι βέλτιστες τιμές των υπερπαραμέτρων του καλύτερου αλγορίθμου, που χρησιμοποιήθηκαν κατά τη διαδικασία εκπαίδευσης για τη δημιουργία μοντέλου πρόβλεψης κυκλοφοριακού φόρτου με την μέθοδο βελτιστοποίησης Random Search CV.....	40

Ευρετήριο Εικόνων

Εικόνα 1: Σχέση χρόνου διαδρομής –αδιάλειπτης ροής.....	6
Εικόνα 2: Σχέση χρόνου διαδρομής – διακοπτόμενης ροής.....	7
Εικόνα 3: Γραφική παράσταση συνάρτησης Davidson.....	8
Εικόνα 4: Γραφική αναπαράσταση της συνάρτησης BPR.....	11
Εικόνα 5: ταξινόμηση των κύριων προσεγγίσεων της μηχανικής μάθησης	14

Εικόνα 6: Οι δύο κύριες κατηγορίες της μηχανικής μάθησης, η επιβλεπόμενη και μη επιβλεπόμενη μάθηση.....	21
Εικόνα 7: Ο αλγόριθμος επιβλεπόμενης Ταξινόμησης “Τυχαία Δάση”	27
Εικόνα 8: Οπτική απεικόνιση της λειτουργίας του αλγορίθμου SVR.....	29
Εικόνα 9: Διάγραμμα SHAP για έλεγχο συνεισφοράς μεταβλητών εισόδου στην πρόβλεψη της μεταβλητής εξόδου “flow_per_lane”	41

Εισαγωγή

1.1. Το Κυκλοφοριακό πρόβλημα

Οι συνθήκες οδικής κυκλοφορίας επηρεάζονται και καθορίζονται από πολλές φυσικές και ανθρώπινες παραμέτρους. Αυτό δημιουργεί δυσκολία στην προσομοίωση των πραγματικών συνθηκών, μέσω χρήσης απλών συναρτήσεων που περιγράφουν τις οδικές συνθήκες κυκλοφορίας. Πρόκειται δηλαδή για ένα πρόβλημα που δεν μεταβάλλεται γραμμικά με το χρόνο (Boukerche & Wang, 2020).

Οι κυκλοφοριακές παράμετροι μεταβάλλονται με τον χρόνο και τον χώρο (Boukerche & Wang 2020). Σε ότι αφορά στον χρόνο, εμφανίζουν “τάση (trend)” που περιγράφει την αύξηση ή μείωση τους σε σχέση με το χρόνο και την “εποχικότητα” (seasonal change), η οποία χαρακτηρίζει τη ροή κυκλοφορίας ως μια κατάσταση που μεταβάλλεται από ορισμένες συνθήκες των ανθρώπων. Για παράδειγμα, το πρωί παρατηρείται αυξημένη κίνηση στο οδικό δίκτυο, διότι οι περισσότεροι άνθρωποι μετακινούνται στις δουλειές τους. Η τελευταία συνιστώσα που επηρεάζει τη ροή κυκλοφορίας είναι η “τυχαία διακύμανση (random variation)”, η οποία εκφράζει τις αλλαγές της κίνησης στο οδικό δίκτυο σε σχέση με κάποια μη συνηθισμένη συνθήκη, όπως για παράδειγμα είναι ένα ακραίο καιρικό φαινόμενο ή κάποιο ατύχημα (Barnett, 2005).

Με λίγα λόγια, θα μπορούσε κανείς να χαρακτηρίσει την πρόβλεψη της κυκλοφορίας των οχημάτων ως μια συνάρτηση που αποτελείται από κατάλληλες μεταβλητές που χαρακτηρίζουν την ροή κυκλοφορίας στο οδικό δίκτυο, καθώς και από χωρικά δεδομένα που αφορούν την τοποθεσία για την οποία δημιουργήθηκε το μοντέλο πρόβλεψης. Με τη δημιουργία τέτοιων μοντέλων πρόγνωσης θα παρέχεται η πληροφορία σχετικά με τις μελλοντικές συνθήκες κίνησης στους δρόμους, και αυτό θα έχει ως αποτέλεσμα την αύξηση της οδικής ασφάλειας.

Σήμερα, με την ανάπτυξη της τεχνολογίας, τα εργαλεία που χρησιμοποιούνται για τη δημιουργία μοντέλων που εκφράζουν τις συνθήκες κυκλοφορίας βασίζονται σε αλγορίθμους μηχανικής μάθησης. Έτσι έχουν αναπτυχθεί τα Ευφυή Συστήματα Μεταφορών – Ε.Σ.Μ. (Intelligent Transport Systems – ITS). Παλαιότερα, η πρόβλεψη της ροής κυκλοφορίας πραγματοποιούνταν με τη χρήση συναρτήσεων όγκου-καθυστερήσης (volume delay functions) και χρόνου-

διαδρομής (travel time functions), όπως είναι η συνάρτηση BPR (Bureau of Public Roads). Παρόλα αυτά, τέτοιου είδους συναρτήσεις πιθανότατα να δίνουν ανακριβή αποτελέσματα στην πρόβλεψη της οδικής κυκλοφορίας, με συνέπεια να ληφθούν μη ρεαλιστικές αποφάσεις που μπορεί να επιφέρουν κίνδυνο στο δίκτυο μεταφορών και κατ' επέκταση και στον ίδιο τον άνθρωπο (Hou et al., 2022).

1.2. Σκοπός

Ο κύριος στόχος της παρούσας διπλωματικής εργασίας είναι η δημιουργία μοντέλων πρόβλεψης κυκλοφοριακού φόρτου σε οδικά τμήματα από τους χρόνους διαδρομής και τους κυκλοφοριακούς φόρτους και διάφορες χωρικές παραμέτρους που έχουν συλλεγεί σε διατομές αυτών. Υπενθυμίζεται ότι ο κυκλοφοριακός φόρτος αφορά στον αριθμό των οχημάτων που περνούν από μια διατομή οδού συγκεκριμένης χωρητικότητας σε κάποιο συγκεκριμένο χρονικό διάστημα, ενώ ο χρόνος διαδρομής είναι ο χρόνος που απαιτείται για να διανύσουν τα οχήματα ένα συγκεκριμένο οδικό τμήμα. Τα μοντέλα που αναπτύσσονται έχουν χρησιμότητα σε διάφορους τομείς, όπως η μείωση της συμφόρησης, η βελτίωση της κυκλοφοριακής ροής και η παροχή αξιόπιστων πληροφοριών στους οδηγούς σχετικά με τους εκτιμώμενους χρόνους διαδρομής.

Επιπλέον, αυτά τα μοντέλα μπορούν να χρησιμοποιηθούν για τη βελτίωση της αποδοτικότητας των συστημάτων δημόσιας συγκοινωνίας και την υποστήριξη των αρχών στη λήψη αποφάσεων για την κατασκευή και συντήρηση των οδικών υποδομών. Η ακριβής πρόβλεψη των χρόνων διαδρομής είναι κρίσιμη για την ενίσχυση της οικονομικής δραστηριότητας, την οργάνωση της αστικής κινητικότητας και τη βελτίωση της ποιότητας ζωής των πολιτών.

Για την επίτευξη αυτού του στόχου, στην παρούσα διπλωματική εργασία αξιοποιούνται σύγχρονες τεχνολογίες ανάλυσης δεδομένων και μηχανικής μάθησης, οι οποίες επιτρέπουν την αναγνώριση προτύπων στα δεδομένα κυκλοφορίας και τη δημιουργία ακριβών προβλέψεων. Τα δεδομένα που συλλέγονται περιλαμβάνουν όχι μόνο τον κυκλοφοριακό φόρτο αλλά και άλλους παράγοντες όπως οι ώρες αιχμής, οι καιρικές συνθήκες, και οι ειδικές εκδηλώσεις ή οι πορείες που επηρεάζουν την κυκλοφορία.

Η ανάπτυξη και η εφαρμογή αυτών των μοντέλων προσφέρει επίσης τη δυνατότητα για συνεχή βελτίωση και προσαρμογή στις μεταβαλλόμενες συνθήκες κυκλοφορίας, καθιστώντας τα ένα πολύτιμο εργαλείο για τη διαχείριση της κυκλοφορίας και την ανάπτυξη βιώσιμων αστικών συστημάτων μεταφορών.

1.3. Περιεχόμενα Ενοτήτων

Τα δομικά στοιχεία της συγκεκριμένης διπλωματικής εργασίας περιγράφονται παρακάτω:

Στο δεύτερο κεφάλαιο παρουσιάζεται μια ανασκόπηση της βιβλιογραφίας σε σχέση με το θέμα της πρόβλεψης κυκλοφοριακών μεγεθών. Αναλυτικότερα, αναφέρονται σχετικές δημοσιευμένες έρευνες, γίνεται μια σύντομη αναφορά στις μεθόδους και τα δεδομένα που χρησιμοποιήθηκαν, καθώς και τα συμπεράσματα που προέκυψαν από τον συνολικό όγκο της βιβλιογραφίας. Επίσης, αναλύεται η θεωρητική βάση της συνάρτησης BPR και παραδείγματα εφαρμογών της στην επιστημονική βιβλιογραφία αλλά παρουσιάζονται και τα πρότυπα μηχανικής μάθησης και η εφαρμογή τους στην πρόβλεψη της οδικής κυκλοφορίας.

Στο τρίτο κεφάλαιο, αναλύεται η μεθοδολογία που ακολουθήθηκε για την επίτευξη του στόχου της διπλωματικής εργασίας. Επίσης, παρουσιάζονται και αναλύονται τα μοντέλα μηχανικής μάθησης καθώς και οι υπερπαραμέτροι που χρησιμοποιήθηκαν. Τέλος, επισημαίνονται οι μετρικές αξιολόγησης των μοντέλων μηχανικής μάθησης.

Στο τέταρτο κεφάλαιο, περιγράφεται η διαδικασία συλλογής δεδομένων και η επεξεργασία τους προκειμένου να χρησιμοποιηθούν από τα μοντέλα. Περιλαμβάνεται ακόμη λεπτομερής περιγραφή της μεθοδολογίας που εφαρμόστηκε και παρουσιάζονται τα αποτελέσματα της διαδικασίας.

Στο πέμπτο κεφάλαιο, ανακεφαλαιώνονται τα βασικά συμπεράσματα που προέκυψαν από τις αναλύσεις και διατυπώνονται προτάσεις για μελλοντικές έρευνες.

Στο έκτο κεφάλαιο, παρουσιάζεται η σχετική βιβλιογραφία που αξιοποιήθηκε κατά την εκπόνηση της διπλωματικής εργασίας.

Κεφάλαιο 2

Βιβλιογραφική Ανασκόπηση

2.1. Πρόβλεψη Κυκλοφορίας

Η πρόβλεψη των κυκλοφοριακών μεγεθών αποτελεί σημαντικό κομμάτι της επιστημονικής βιβλιογραφίας στον τομέα των μεταφορών. Αν και έχουν αναπτυχθεί πολλές μεθοδολογίες πρόβλεψης κατά καιρούς, η πρόσφατη πρόοδος στην τεχνολογία της πληροφορικής έχει επιτρέψει την ανάπτυξη μεθόδων πρόβλεψης που βασίζονται σε πολύπλοκα μοντέλα μηχανικής μάθησης. Αυτό το κεφάλαιο αναλύει τις διάφορες μεθοδολογίες που έχουν παρουσιαστεί στη βιβλιογραφία, καθώς και τις ευρύτερες κατηγορίες στις οποίες ανήκουν. Κατόπιν, αναλύονται οι παράμετροι που επηρεάζουν την οδική κυκλοφορία και είναι χρήσιμη για τη δημιουργία μοντέλων πρόβλεψης. Στη συνέχεια, παρουσιάζονται έρευνες που αφορούν την πρόβλεψη του χρόνου διαδρομής και του κυκλοφοριακού φόρτου με χρήση μεθόδων μηχανικής μάθησης, καθώς και παλαιότερων μεθόδων όπως οι συναρτήσεις όγκου-καθυστερήσης με γνωστότερη τη συνάρτηση BPR.

Σε μια κοινωνία που η έννοια του χρόνου αποτελεί μια από τους σημαντικότερους παράγοντες που καθορίζουν τις μεταφορές, η εκτίμησή του είναι αναγκαία. Ο χρόνος διαδρομής χρησιμοποιείται κατά τον σχεδιασμό έργων μεταφορών καθώς και στην επίλυση λειτουργικών προβλημάτων στις μεταφορές. Αναλυτικότερα, ο χρόνος διαδρομής αποτελεί βασική μεταβλητή τόσο για την δυναμική και στατική ανάθεση της κυκλοφορίας όσο και για τον έλεγχο ανατροφοδότησης του οδικού δικτύου. Επίσης, χρησιμοποιείται για την ανάπτυξη μέτρων για την βελτίωση των επιδόσεων του συστήματος μεταφορών κατά την διάρκεια μιας χρονικής περιόδου καθώς και για την βελτίωση της αξιοπιστίας της των μεταφορών τόσο ως προς την ακρίβεια του χρόνου, την διαθεσιμότητα, την ασφάλεια όσο και την ποιότητα υπηρεσιών που προσφέρουν. Συνοψίζοντας, στο δίκτυο μεταφορών το κόστος αφορά και τον χρόνο διαδρομής για αυτό και είναι αναγκαία η εκτίμησή του στο οδικό δίκτυο.

Η εκτίμηση του χρόνου διαδρομής είναι απαραίτητη και σε μακροσκοπικές προσομοιώσεις για την πρόβλεψη του ποσοστού θνησιμότητας στο οδικό δίκτυο καθώς και τη μελλοντική θνησιμότητα στο δίκτυο αυτό. Σε συνέχεια των παραπάνω ο Orpe (1989), ερεύνησε τη σχέση του όγκου κυκλοφορίας και της οδικής ασφάλειας με την χρήση κατάλληλων μοντέλων στους δρόμους της Ολλανδίας, της Αγγλίας και της Γερμανίας. Η έννοια του χρόνου είναι άμεσα συνδεδεμένη με τον κυκλοφοριακό φόρτο σε ένα οδικό δίκτυο, καθώς ο προσδιορισμός του καθορίζει σε μεγάλο βαθμό τις καθυστερήσεις που παρατηρούνται σε αυτό.

2.1.1. Οι καιρικές συνθήκες

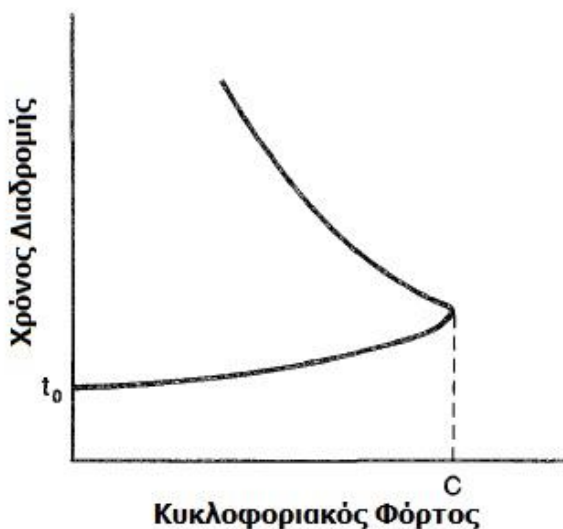
Οι καιρικές συνθήκες επηρεάζουν σημαντικά τη ροή των οχημάτων στο οδικό δίκτυο. Πολλές φορές αποτελούν την αιτία για πρόκληση διάφορων αυτοκινητιστικών ατυχημάτων. Το καιρικό φαινόμενο που έχει αναλυθεί περισσότερο στη βιβλιογραφία είναι η βροχόπτωση. Μάλιστα, σε μελέτη των Andrey & Yagar (1993) αποδείχτηκε ότι κάθε φορά που βρέχει στον Καναδά τα ατυχήματα αυξάνονται κατά 70%. Αυτό κυρίως προκαλείται από την μείωση ορατότητας και από το φαινόμενο της υδρολίσθησης. Για το λόγο αυτό, η μεταβλητή αυτή πρέπει οπωσδήποτε να εμπεριέχεται στα μοντέλα πρόβλεψης της κυκλοφοριακής ροής, ώστε σε περίπτωση ύπαρξης βροχής, χιονιού ή ομίχλης να λαμβάνονται οι κατάλληλες αποφάσεις για την αποσυμφόρηση του οδικού δικτύου.

Στη συνέχεια, οι αέριοι ρύποι που εκλύονται από τα οχήματα, συνδέονται άμεσα με την ποιότητα του αέρα στις πόλεις, και επίσης εντείνουν το πρόβλημα της κλιματικής αλλαγής (ή κρίσης), που σχετίζεται με την αύξηση της θερμοκρασίας. Αναφέρεται, πως τα οχήματα αποτελούν την κύρια πηγή έκλυσης των αερίων του θερμοκηπίου, εκτός από το υποξείδιο του αζώτου. Παρόλα αυτά το υποξείδιο του αζώτου εκλύεται σε έντονο βαθμό από τα βαρεία και ογκώδη οχήματα (Progiou & Ziomas, 2012). Η δημιουργία μοντέλων πρόβλεψης που λαμβάνουν υπόψιν την παράμετρο των αερίων του θερμοκηπίου που εκλύονται από τα οχήματα στο οδικό δίκτυο, θα βοηθήσει σε μια καλύτερη στρατηγική μείωσης ή διατήρησης αυτών των αερίων, πριν το πρόβλημα αυτό γίνει επιβλαβές για την ανθρώπινη υγεία και βιωσιμότητα.

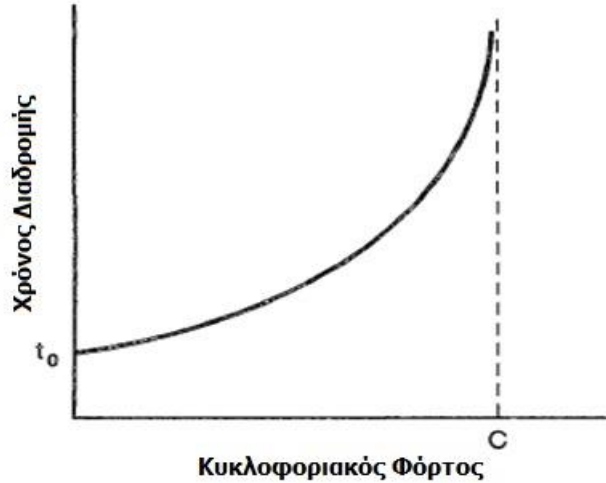
2.1.2. Η χωρητικότητα του οδικού δικτύου

Το μέγεθος του οδικού δικτύου επηρεάζει σημαντικά την κίνηση, καθώς όσο πιο μεγάλο είναι, τόσο περισσότερα οχήματα μπορούν να κυκλοφορήσουν και άρα να μειωθεί το φαινόμενο της κυκλοφοριακής συμφόρησης. Είναι γνωστό όμως ότι η χωρητικότητα των οδικών δικτύων της Ελλάδας είναι μικρή, και ιδιαίτερα εντός κατοικημένων περιοχών. Για το λόγο αυτό έχουν δημιουργηθεί μοντέλα προσομοίωσης τα οποία λαμβάνουν υπόψη το μήκος του οδικού δικτύου, καθώς και πόσο όγκο οχημάτων μπορεί αυτό να χωρέσει, ώστε να μελετηθεί ο κυκλοφοριακός φόρτος (Tsekeris & Geroliminis, 2013).

Στη συνέχεια, καθοριστική παράμετρος για την σχέση χρόνου-χωρητικότητας αποτελεί το είδος ροής της κυκλοφορίας, δηλαδή η ύπαρξη διακοπτόμενης ή μη διακοπτόμενης ροής. Διακοπτόμενη ροή ορίζεται η μορφή της ροής κυκλοφορίας όταν αυτή διακόπτεται είτε από κάποια ουρά εξαιτίας φωτεινού σηματοδότη, είτε κατάλληλης σήμανσης, ενώ συνθήκες αδιάλειπτης ροής υπάρχουν όταν τα οχήματα που διανύουν ένα οδόστρωμα δεν παρεμποδίζονται από αιτίες εξωτερικές προς τη ροή της κυκλοφορίας, όπως πινακίδες ή σήματα. Επομένως, οι αυτοκινητόδρομοι παρέχουν συνθήκες αδιάλειπτης ροής, ενώ οι αρτηριακές οδοί που περιέχουν σηματοδοτούμενες διασταυρώσεις παρέχουν συνθήκες διακοπτόμενης ροής. Η σχέση χρόνου-χωρητικότητας απεικονίζεται στις παρακάτω Εικόνες 1 και 2.



Εικόνα 1 :Σχέση χρόνου διαδρομής –αδιάλειπτης ροής



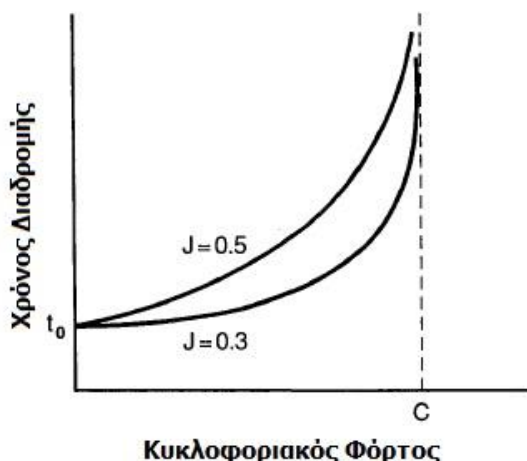
Εικόνα 2 :Σχέση χρόνου διαδρομής – διακοπτόμενης ροής

Σε ερευνά του ο Davidson (1966), προτείνει μια συνάρτηση χρόνου διαδρομής συναρτήσει της χωρητικότητας της κυκλοφορίας της οδού και του όγκου (Εικόνα 3). Για να αντιμετωπιστεί η ασύμπτωτη της συνάρτησης αυτής πολλοί ερευνητές πρότειναν τροποποιήσεις ώστε να μην δημιουργούνται υπολογιστικά προβλήματα όταν χρησιμοποιούνται σε μοντέλα ανάθεσης κυκλοφορίας. Ο τύπος του χρόνου διαδρομής κατά τον Davinson παρουσιάζεται παρακάτω:

$$T = t_0 \left(\frac{1+J \cdot V}{C-V} \right) \quad (1)$$

όπου :

- t_0 είναι ο χρόνος ελεύθερης ροής
- C η χωρητικότητα του συνδέσμου
- J η παράμετρος η οποία καθορίζει το βαθμό καμπυλότητας της συνάρτησης
- V ο όγκος.



Εικόνα 3 : Γραφική παράσταση συνάρτησης Davidson

Σημειώνεται επίσης πως στην πραγματικότητα, ένα οδικό δίκτυο δεν είναι ομοιογενές. Έχοντας ως δεδομένο την ιδιαιτερότητα αυτή του οδικού δικτύου οι Saedi et al. (2020), προχώρησαν σε κατάτμηση και μελέτη του δικτύου σε ομογενείς ζώνες χρησιμοποιώντας τις μεθόδους τροχιάς και υποτροχιάς καταλήγοντας στην εύρεση μιας βελτιωμένης συνάρτησης χρόνου με παραμέτρους την χωρητικότητα, τον μέσο χρόνο διαδρομής και την τυπική απόκλιση.

2.1.3. Περίοδοι αιχμής

Η πρόβλεψη του κυκλοφοριακού φόρτου να παρέχουν στους οδηγούς χρήσιμες και ακριβείς πληροφορίες διαδρομής και χρόνου διαδρομής καθώς τους δίνει την δυνατότητα να εξετάσουν τα πιθανά οφέλη των ευέλικτων ωραρίων εργασίας και να αξιολογήσουν τις περιβαλλοντικές επιπτώσεις της κυκλοφοριακής συμφόρησης. Από την έρευνα των Stathopoulos & Karlaftis (2001), αποδείχθηκε ότι η ροή κίνησης δεν ακολουθεί την περίπτωση της κανονικής κατανομής των δεδομένων, αλλά η κατανομή Weibull περιγράφει καλύτερα την περίπτωση των δεδομένων οδικού δικτύου της μελέτης αυτής. Παρατηρήθηκε επίσης ότι η διακύμανση στην κυκλοφορία των οχημάτων στο οδικό δίκτυο είναι έντονη. Είναι σημαντικό να ληφθεί υπόψη το επίπεδο μελέτης της ροής κίνησης των οχημάτων, αν δηλαδή περιγράφονται ετήσια, μηνιαία ή ημερήσια δεδομένα κυκλοφορίας και αν πρόκειται για δρόμους που έχουν συχνά κίνηση, ή είναι πιο απομονωμένοι. Παρόλα αυτά, στην Αθήνα, φαίνεται να μειώνεται η ροή της κίνησης τους μήνες Ιούλιο και Αύγουστο, λόγω περιόδου καλοκαιρινών διακοπών. Ταυτόχρονα φάνηκε ότι η κίνηση

στους δρόμους διαφέρει για τις ημέρες από Δευτέρα έως Παρασκευή σε σχέση με τα Σαββατοκύριακα (Stathopoulos & Karlaftis, 2001).

2.2. Μέθοδοι Πρόβλεψης Οδικής Κυκλοφορίας

Η χρήση της κατάλληλης μεθόδου πρόβλεψης αποτελεί το κλειδί για τη δημιουργία της επιθυμητής πληροφορίας προβλέψεων που αφορούν την οδική κυκλοφορία. Αναφέρονται στη συνέχεια επιγραμματικά κάποια μοντέλα πρόβλεψης που έχουν αναφερθεί στη βιβλιογραφία.

Μέθοδοι που καθοδηγούνται από μοντέλα

Πρόκειται για μοντέλα προσομοίωσης της κυκλοφορίας των οχημάτων σε ένα οδικό δίκτυο σε περιβάλλον κυκλοφοριακής προσομοίωσης. Τα μοντέλα προσομοίωσης κατηγοριοποιούνται ανάλογα με το πόσο λεπτομερές είναι η ανάλυση της ροής κυκλοφορίας, δηλαδή σε μικροσκοπικό ή μακροσκοπικό επίπεδο (Olstam & Tarani, 2004). Αν η προσομοίωση γίνεται σε μικροσκοπικό επίπεδο, τότε δίνονται πιο πολλές πληροφορίες και λεπτομέρειες σχετικά με την κυκλοφορία στο οδικό δίκτυο, τους σηματοδότες, καθώς και τις λωρίδες κυκλοφορίας. Με αυτόν τον τρόπο καθίσταται εφικτή η πρόβλεψη των συνθηκών στο οδικό δίκτυο για μελλοντικά χρονικά διαστήματα.

Μέθοδοι που καθοδηγούνται από δεδομένα

Οι μέθοδοι αυτοί διακρίνονται περεταίρω σε παραμετρικές και μη παραμετρικές μεθόδους. Οι παραμετρικές μέθοδοι, όπως είναι τα μοντέλα παλινδρόμησης, βασίζονται στη θεωρία ότι τα δεδομένα ακολουθούν το μοτίβο μιας υπάρχουσας κατανομής τιμών, όπως είναι η κανονική κατανομή (Boukerche & Wang, 2020), ενώ για τις μη παραμετρικές μεθόδους δεν χρειάζεται τα δεδομένα να στηρίζονται σε κάποια κατανομή τιμών. Τα πιο γνωστά παραμετρικά μοντέλα είναι τα αυτοπαλιδρομικά μοντέλα κινητού μέσου όρου (ARIMA), που χρησιμοποιούνται πολύ συχνά για να περιγράψουν δεδομένα χρονοσειρών (Yu & Zhang, 2004). Επίσης, ένα παράδειγμα μη παραμετρικού αλγορίθμου για τη δημιουργία μοντέλου πρόβλεψης οδικής κυκλοφορίας αποτελεί ο αλγόριθμος kNN (Peterson, 2009). Πιο συγκεκριμένα, βασίζεται στην θεωρία της ευκλείδειας απόστασης και για την πρόβλεψη των νέων αταξινόμητων δεδομένων, απαραίτητο είναι να έχει

προηγηθεί εκπαίδευση με ιστορικά δεδομένα που διαθέτουν ετικέτες (Boukerche & Wang, 2020). Πρόκειται δηλαδή για έναν αλγόριθμο επιβλεπόμενης ταξινόμησης. Σε επόμενο κεφάλαιο θα αναλυθεί η έννοια της επιβλεπόμενης ταξινόμησης.

2.3. Συναρτήσεις Φόρτου – Καθυστέρησης

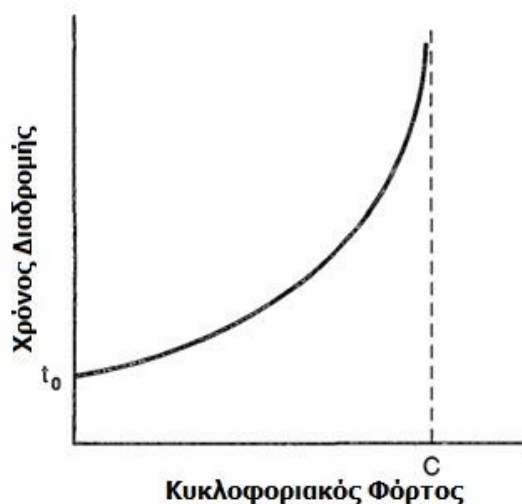
Οι συναρτήσεις φόρτου - καθυστέρησης (Volume Delay Functions) χρησιμοποιούνται συνήθως στον σχεδιασμό των μεταφορών για να μοντελοποιήσουν τις καθυστερήσεις χρόνου διαδρομής στα δίκτυα κυκλοφορίας. Διαφορετικοί τύποι συναρτήσεων φόρτου – καθυστέρησης, όπως η συνάρτηση BPR, η κωνική συνάρτηση, και η συνάρτηση Akcelik, απαιτούν εισαγωγικές παραμέτρους όπως είναι η ελεύθερη ροή του χρόνου διαδρομής και της χωρητικότητας, οι οποίες προσδιορίζονται μέσω της ανάλυσης παλινδρόμησης. Η συνάρτηση BPR, για παράδειγμα, χρησιμοποιεί εμπειρικούς συντελεστές για να εκτιμήσει τις καθυστερήσεις του όγκου κυκλοφορίας με βάση την αναλογία όγκου προς την χωρητικότητα, ενώ η κωνική συνάρτηση περιλαμβάνει μια εκτίμηση ενός μοναδικού συντελεστή α (Neuhold & Fellendorf, 2014).

Πιο συγκεκριμένα, οι συναρτήσεις φόρτου – καθυστέρησης, με πιο γνωστή τη συνάρτηση Bureau of Public Roads Function (BPR), προσφέρουν τη δυνατότητα περιγραφής των συνθηκών κυκλοφορίας (Branston, 1976). Η συνάρτηση όγκου – καθυστέρησης (BPR) περιγράφεται σύμφωνα με την παρακάτω συνάρτηση (2). Αναλυτικότερα, στη συνάρτηση, το C χαρακτηρίζει τη μέγιστη χωρητικότητα του οδικού δικτύου, η παράμετρος α συνήθως λαμβάνει την τιμή 1 και η παράμετρος β κυμαίνεται μεταξύ του εύρους των τιμών 2 έως 12 (Kachroo & Sastry, 2016). Στον Πίνακα 1 δίνονται κάποιες πιθανές τιμές για τις παραμέτρους α και β της συνάρτησης BPR (Neuhold & Fellendorf, 2014). Ακολούθως, παρουσιάζεται στην Εικόνα 4 η γραφική αναπαράσταση της συνάρτησης BPR (Rose et al., 1989).

$$\Phi_{\frac{f}{C}} = \left(1 + \beta \frac{f}{C} \right)^{\alpha} \quad (2)$$

Πίνακας 1: Τιμές παραμέτρων α και β της συνάρτησης BPR για συγκεκριμένες ταχύτητες οχημάτων σε αυτοκινητόδρομο και δρόμο με πολλαπλές λωρίδες που δεν είναι όμως αυτοκινητόδρομος (Πηγή: Neuhold & Fellendorf, 2014)

Ταχύτητα (mph)	α	β
Αυτοκινητόδρομος		
70	0.88	9.8
60	0.83	5.5
50	0.56	3.6
Δρόμος με πολλαπλές λωρίδες		
70	1.00	5.4
60	0.83	2.7
50	0.71	2.1



Εικόνα 4: Γραφική αναπαράσταση της συνάρτησης BPR

Μια πιο συγκεκριμένη μορφή της συνάρτησης BPR που περιγράφει το επίπεδο συμφόρησης των οδικών τμημάτων, εκφρασμένο ως χρονική καθυστέρηση, βάσει της μεταβολής του κυκλοφοριακού φόρτου παρουσιάζεται από την σχέση (3) (Saadullah et al., 2022).

$$TT = TT_{FF} \left(1 + \alpha \frac{V_{FR}}{C} \right)^n \quad (3)$$

Στη συνέχεια αναλύονται οι ορισμοί των μεταβλητών της συνάρτησης BPR:

- Η παράμετρος α εκφράζει την αναλογία του χρόνου διαδρομής σε συνθήκες ελεύθερης ροής προς τον χρόνο διαδρομής σε συνθήκες μέγιστης κυκλοφοριακής ικανότητας.

- Η παράμετρος β εκφράζει την ταχύτητα μετάβασης από συνθήκες ελεύθερης ροής σε συνθήκες κυκλοφοριακής συμφόρησης. Μεγαλύτερες τιμές της παραμέτρου β συνεπάγονται μικρότερη χρονική καθυστέρηση σε συνθήκες χαμηλής κυκλοφορίας, με αποτέλεσμα τη μη έγκαιρη αναγνώριση της κυκλοφοριακής συμφόρησης.
- Η μεταβλητή TT_{FF} εκφράζει τον χρόνο διαδρομής ενός οδικού τμήματος σε συνθήκες ελεύθερης ροής. Οι συνθήκες ελεύθερης ροής χαρακτηρίζονται από χαμηλή κυκλοφοριακή πυκνότητα και ταχύτητες, που καθορίζονται μόνο από τις προτιμήσεις των οδηγών, τα καθορισμένα όρια ταχύτητας και τις επικρατούσες οδικές συνθήκες.
- Η κυκλοφοριακή ικανότητα (C) εκφράζει τον μέγιστο αριθμό οχημάτων ή πεζών που μπορεί να διέλθουν από μία διατομή ή ομοιόμορφο τμήμα λωρίδας κυκλοφορίας ή οδού, προς μία ή και τις δύο κατευθύνσεις, κατά τη διάρκεια μιας δεδομένης χρονικής περιόδου, λαμβάνοντας υπόψη τις οδικές και κυκλοφοριακές συνθήκες καθώς και τις συνθήκες ελέγχου της κυκλοφορίας.
- Ο κυκλοφοριακός φόρτος (V_{FR}) είναι ο συνολικός αριθμός οχημάτων που περνούν από τη διατομή μιας λωρίδας ή οδού κατά τη διάρκεια ενός συγκεκριμένου χρονικού διαστήματος (συνήθως οχήματα ανά ώρα).

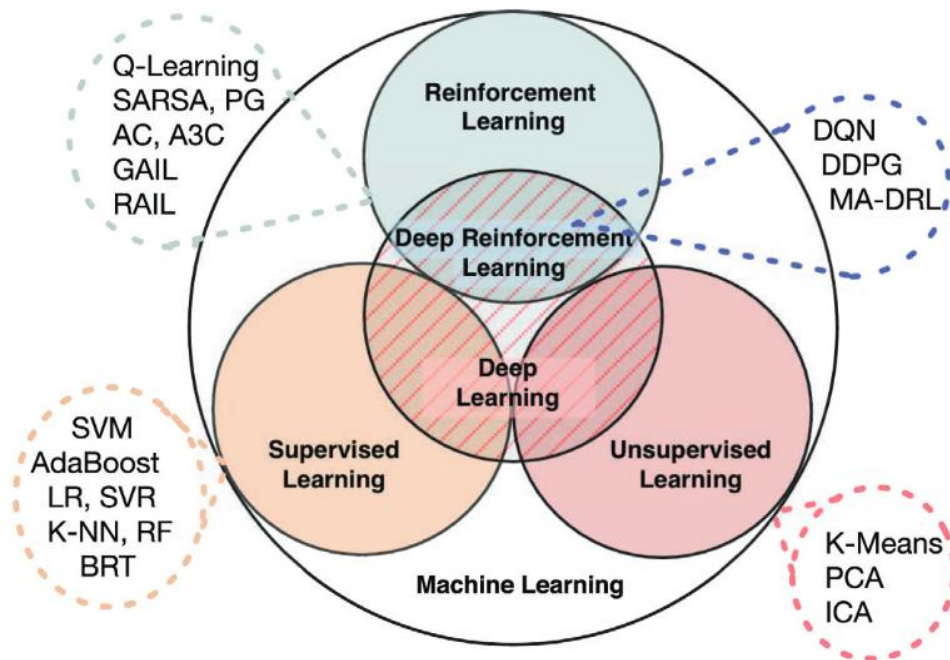
Ωστόσο, τα παραδοσιακά μοντέλα φόρτου – καθυστέρησης περιγράφουν τον κυκλοφοριακό φόρτο στους δρόμους ως διαχωρίσιμο και ως μια μονοτονικά αυξανόμενη συνάρτηση, όπου ο κυκλοφοριακός φόρτος παρουσιάζει αύξηση με την ροή της κίνησης. Αυτή η προσέγγιση θεωρεί την κυκλοφορία ως μια απλή, μονότονη διαδικασία, χωρίς πολυπλοκότητα ή δυνατότητα αλλαγής (Tsekeris & Geroliminis, 2013). Αρκετές μελέτες έχουν αμφισβητήσει τη χρήση αυτών των μεθόδων λόγω της ανάγκης να ληφθεί υπόψη η μη μονοτονικότητα του χρόνου διαδρομής σε σχέση με τον κυκλοφοριακό φόρτο (Mc Donald et al., 1999), από όπου αποδείχθηκε η εγγενής ασυνέπεια, αδυναμία και αστάθεια των στατικών μοντέλων κυκλοφοριακής συμφόρησης (Verhoef, 1999, 2001, 2005).

Η συνάρτηση BPR, αν και αναπτύχθηκε μέσω ερευνών κυκλοφορίας σε αυτοκινητόδρομους, παρουσιάζει σημαντικούς περιορισμούς όταν εφαρμόζεται σε αστικά οδικά δίκτυα, τα οποία χαρακτηρίζονται από ιδιαίτερα στοιχεία, όπως διασταυρώσεις, λωρίδες αποκλειστικής χρήσης για δημόσια συγκοινωνία και συχνή σηματοδότηση (Pan Yuyan et al.). Πιο συγκεκριμένα, η βαθμονόμηση της συνάρτησης BPR για αυτοκινητόδρομους οδηγεί σε μη αμελητέα σφάλματα

στις αστικές εφαρμογές, καθώς οι διακοπές ροής και οι σύνθετες συνθήκες κυκλοφορίας δεν είναι προβλέψιμες από το μοντέλο. Επιπλέον, η συνάρτηση BPR αποτυγχάνει να καταγράψει την πολυπλοκότητα των δυναμικών συμφόρησης, όπως το σχηματισμό, τη διάδοση και τη διάλυση των ουρών οχημάτων, κρίσιμα στοιχεία για την κατανόηση της αστικής κυκλοφοριακής συμπεριφοράς. Παράλληλα, η ακρίβεια της συνάρτησης μειώνεται σημαντικά σε υψηλές τιμές της παραμέτρου β , οι οποίες συχνά απαντώνται σε συνθήκες υψηλής πυκνότητας. Η μεταφερσιμότητα των παραμέτρων της BPR συνιστά επίσης πρόκληση, καθώς η προσαρμογή σε ποικίλα οδικά δίκτυα και περιοχές δεν επιτυγχάνεται εύκολα, ενώ η χρήση συνολικών δεδομένων συνδέσεων μειώνει την ακρίβεια, προσθέτοντας επιπλέον πολυπλοκότητα στην εξατομικευμένη βαθμονόμηση. Τα παραπάνω αναδεικνύουν τους περιορισμούς της συνάρτησης BPR στα αστικά δίκτυα και υποδεικνύουν την ανάγκη ανάπτυξης προσαρμοσμένων ή εναλλακτικών μοντέλων που να ανταποκρίνονται στις απαιτήσεις των αστικών κυκλοφοριακών συστημάτων.

2.4. Τεχνητή Νοημοσύνη και Μηχανική Μάθηση

Αρχικά, η μηχανική μάθηση αποτελεί τον κλάδο που επικεντρώνεται στο να δημιουργεί μηχανές που λειτουργούν και παίρνουν αποφάσεις σε διάφορα προβλήματα, όπως κάνει και η ανθρώπινη νοημοσύνη (Εικόνα 5). Είναι πλέον γνωστό ότι καθημερινά παράγεται ένας τεράστιος όγκος κυκλοφοριακών δεδομένων. Σύμφωνα με την βιβλιογραφία, οι τεχνικές μηχανικής μάθησης διαδραματίζουν σημαντικό ρόλο στην ανάπτυξη προηγμένων μοντέλων που υποστηρίζουν τις λειτουργίες των Ε.Σ.Μ. (Έξυπνα Συστήματα Μεταφορών). Αυτά τα μοντέλα έχουν την ικανότητα να εκμεταλλεύονται μεγάλο όγκο δεδομένων που συλλέγονται από διάφορες πηγές, όπως αισθητήρες, κάμερες και συσκευές GPS, για να δημιουργήσουν πρότυπα ώστε να προβλέπουν τον αριθμό των οχημάτων σε ένα οδικό δίκτυο και τελικά να βελτιστοποιηθούν οι λειτουργίες μεταφορών. Μέσω διάφορων αλγορίθμων, όπως είναι τα νευρωνικά δίκτυα (neural networks), τα δένδρα αποφάσεων (decision trees) και οι μηχανές διανυσματικής υποστήριξης (support vector machines), επιτρέπεται η δημιουργία σταθερών και αποδοτικών μοντέλων τα οποία είναι ικανά να εκπαιδεύονται σε παλαιότερα δεδομένα και να βελτιώνουν την αποδοτικότητα, την ασφάλεια και τη βιωσιμότητα των οδικών δικτύων μεταφοράς (Yuan et al., 2022).



Εικόνα 5: Γενική ταξινόμηση των κύριων προσεγγίσεων της μηχανικής μάθησης (Πηγή: Yuan et al., 2022)
 Στην συνέχεια αναφέρονται κάποια επιτυχημένα δημιουργούμενα μοντέλα, που οι αρχιτεκτονικές τους βασίζονται σε τεχνικές μηχανικής μάθησης.

Μοντέλα που προσδιορίζουν τον Κυκλοφοριακό Φόρτο

Αρχικά, η έρευνα των Okawa et al. (2017) επικεντρώθηκε στην πρόβλεψη του κυκλοφοριακού φόρτου για διαφορετικά τμήματα των δρόμων χρησιμοποιώντας ιστορικά δεδομένα GPS. Στη μελέτη εισάχθηκε μια καινοτόμα μέθοδος που ονομάζεται διπλή διμερής παλινδρόμηση Poisson (bilinear Poisson regression) που ενσωματώνει κρυφούς παράγοντες για την πρόβλεψη του κυκλοφοριακού φόρτου. Με την ανάλυση πραγματικών δεδομένων κυκλοφορίας από την περιοχή του Τόκιο, η μελέτη αποδεικνύει την αποτελεσματικότητα της προτεινόμενης μεθόδου στην καταγραφή περιοδικών μοτίβων και αλληλεπιδράσεων μεταξύ των οδικών τμημάτων για τη βελτίωση της πρόβλεψης της ροής της κυκλοφορίας. Πιο συγκεκριμένα, η αναφερόμενη μέθοδος λειτουργεί καλύτερα από άλλες δύο συγκρινόμενες μεθόδους πρόβλεψης της ίδιας φιλοσοφίας για βραχυπρόθεσμες προβλέψεις, αλλά όχι τόσο καλά για μακροπρόθεσμες. Αναφέρεται επίσης ότι η διμερής παλινδρόμηση Poisson συνδράμει στην καλύτερη κατανόηση

των μοτίβων κυκλοφορίας, όπως το πώς η κυκλοφορία αλλάζει με την πάροδο του χρόνου και σε διαφορετικά μέρη της πόλης.

Στην συνέχεια, η ομάδα των Lee & Fambro (1999), πραγματοποίησε έρευνα σχετικά με τη χρήση μοντέλων ARIMA στην πρόβλεψη του κυκλοφοριακού φόρτου. Η τοποθεσία μελέτης για τη συλλογή δεδομένων όγκου κυκλοφορίας ήταν στο Σαν Αντόνιο, στο Τέξας. Αναλυτικότερα, συλλέχθηκαν δεδομένα από δύο τοποθεσίες, χρησιμοποιώντας ανιχνευτές βρόχων για την παρακολούθηση των συνθηκών κυκλοφορίας και των μοτίβων ροής σε διαστήματα 5 λεπτών. Από τα πέντε μοντέλα πρόβλεψης που δημιουργήθηκαν (ESM – Exponential Smoothing Method, FAR – Full Autoregressive, SAR – Subset Autoregressive, FARIMA – Full ARIMA, SARIMA – Subset ARIMA) αποδείχθηκε ότι το μοντέλο SARIMA απέδωσε καλύτερα όσον αφορά την ακρίβεια πρόβλεψης, με χαμηλότερες τιμές σφάλματος σε σύγκριση με τα άλλα μοντέλα, υποδεικνύοντας την αποτελεσματικότητά του στην πρόβλεψη τάσεων όγκου κυκλοφορίας.

Από την μελέτη των Li et al. (2015), προτείνεται μια μέθοδος για την κατασκευή μοντέλων ακριβείας στην πρόβλεψη του κυκλοφοριακού φόρτου χρησιμοποιώντας μεγάλο όγκο δεδομένων που συλλέγονται από διάφορους τύπους αισθητήρων. Πιο συγκεκριμένα, περιγράφεται μια στρατηγική πολλαπλών βημάτων που περιλαμβάνει την ανάλυση αιτιότητας του Granger (Granger causality analysis), αλγόριθμους αποσύνθεσης (decomposition algorithms), και την παλινδρόμηση Lasso για την αποτελεσματική επεξεργασία των δεδομένων. Αυτή η προσέγγιση συμβάλλει στη μείωση του κόστους ή θορύβου και αποκαλύπτει τις σχέσεις μεταξύ των οδικών δικτύων και των μοτίβων κυκλοφορίας, βελτιώνοντας την ακρίβεια των προβλέψεων ροής κυκλοφορίας.

Αναφέρεται επίσης από τους Castro-Neto et al. (2009), η σημασία της πρόβλεψης των βραχυπρόθεσμων ροών κυκλοφορίας σχετικά με ατυχήματα οχημάτων ή ακραίες καιρικές συνθήκες για δημιουργία αποτελεσματικών συστημάτων διαχείρισης της κυκλοφορίας. Σε αυτή τη μελέτη χρησιμοποιήθηκε ένα μοντέλο που βασίζεται στην αρχιτεκτονική Μηχανών Υποστήριξης Παλινδρόμησης (OL-SVR), ως μια στατιστική τεχνική εκμάθησης που χρησιμοποιείται για την πρόβλεψη της κυκλοφορίας των αυτοκινητοδρόμων τόσο υπό συνθήκες εποχικότητας, όσο και συνθήκες τυχαίας διακύμανσης, αποδίδοντας καλύτερα από άλλα μοντέλα, όπως το μοντέλο της μέγιστης πιθανότητας Gaussian (GML) σε συνθήκες τυχαίας

διακύμανσης (random variation). Η έρευνα αυτή στόχευσε στην ενίσχυση των ευφρών συστημάτων μεταφορών παρέχοντας καλύτερα εργαλεία για τη διαχείριση πραγματικών τροχαίων φαινομένων και συνθηκών.

Μοντέλα που προσδιορίζουν τον χρόνο διαδρομής

Από την έρευνα των Wang et al. (2018) παρουσιάστηκε η σημασία της εκτίμησης του χρόνου διαδρομής των οχημάτων ή του εκτιμώμενου χρόνου άφιξης ως κρίσιμης μεταβλητής βάσει τοποθεσίας στα συστήματα πλοήγησης και μεταφοράς. Αναλυτικότερα, εισήχθη μια καινοτόμα προσέγγιση μηχανικής μάθησης για την πρόβλεψη του χρόνου διαδρομής χρησιμοποιώντας δεδομένα οχημάτων, ενσωματώνοντας χωρο-χρονική παλινδρόμηση και ένα μοντέλο αρχιτεκτονικής βαθιάς εκμάθησης Wide-Deep-Recurrent (WDR). Από τη μελέτη αποδείχθηκε ότι ο αλγόριθμος βαθιάς μάθησης ξεπερνά σε απόδοση τις υπάρχουσες μεθόδους, ωφελώντας πλατφόρμες όπως η Didi Chuxing και βελτιώνοντας την ακρίβεια των προβλέψεων του εκτιμώμενου χρόνου άφιξης για εκατομμύρια πελάτες. Σημειώνεται πως μοντέλα που βασίζονται σε αρχιτεκτονικές βαθιάς μάθησης είναι ικανά να εκπαιδευτούν και να δώσουν πολύ μικρή τιμή σφάλματος στις τιμές πρόβλεψης, καθώς έχουν την ικανότητα να αναλύουν και να επεξεργάζονται πολύπλοκα δεδομένα, πρότυπα και μοτίβα, που οι συμβατικές συναρτήσεις μελέτης χρόνου διαδρομής και άλλα απλούστερα μοντέλα αδυνατούν.

Στη συνέχεια, από την ομάδα των Servos et al. (2019) εξετάστηκε η σημασία της ακριβούς πρόβλεψης των χρόνων διαδρομής στις μεταφορές εμπορευμάτων για τη βελτίωση της ποιότητας και της αποδοτικότητας της εφοδιαστικής αλυσίδας. Στη μελέτη αυτή χρησιμοποιήθηκαν αλγόριθμοι μηχανικής μάθησης όπως είναι οι αλγόριθμοι Extra Trees, Ada Boost και SVR για την ανάλυση δεδομένων παρακολούθησης και την κατασκευή μοντέλων για την πρόβλεψη του χρόνου διαδρομής. Από την μελέτη αποδείχθηκε ότι το μοντέλο SVR έδωσε την καλύτερη ακρίβεια πρόβλεψης, αποδίδοντας καλύτερα από άλλες παραδοσιακές μεθόδους (average-based methods), με μέσο απόλυτο σφάλμα 17 ωρών για χρόνο μεταφοράς έως και 30 ημερών σύμφωνα με πραγματικά δεδομένα από μια διαδρομή μεταφορών εμπορευματοκιβωτίων μεταξύ Γερμανίας και ΗΠΑ.

2.5. Συμπεράσματα της Βιβλιογραφίας

Σύμφωνα με τη βιβλιογραφία, η πρόβλεψη των κυκλοφοριακών παραμέτρων αποδεικνύεται κομβικής σημασίας για την ορθή διαχείριση και βελτίωση του οδικού δικτύου. Οι εξελιγμένες τεχνικές πρόβλεψης που βασίζονται σε σύγχρονα μοντέλα μηχανικής μάθησης έχουν αυξήσει την ακρίβεια εκτίμησης κυκλοφοριακών μεγεθών, όπως ο χρόνος διαδρομής και ο κυκλοφοριακός φόρτος, επιτρέποντας στο σύστημα μεταφορών να προσαρμόζεται στις ανάγκες του νέου ψηφιακού περιβάλλοντος και να υποστηρίζει τη λήψη αποφάσεων. Καθώς η ακρίβεια της εκτίμησης του χρόνου διαδρομής επηρεάζει τη συνολική αποτελεσματικότητα του δικτύου, η ενσωμάτωσή του στον κυκλοφοριακό σχεδιασμό και τη λειτουργική ανάλυση είναι απαραίτητη για τη βελτίωση της αξιοπιστίας, της διαθεσιμότητας και της ασφάλειας του δικτύου.

Παράλληλα, οι καιρικές συνθήκες και η ποικιλομορφία του οδικού δικτύου καθιστούν απαραίτητη τη χρήση πιο σύνθετων μοντέλων, τα οποία να λαμβάνουν υπόψη τα καιρικά φαινόμενα και τη διακοπτόμενη ή μη διακοπτόμενη ροή, τη χωρητικότητα και τον κυκλοφοριακό φόρτο, όπως προτείνουν πρόσφατες έρευνες που αναπτύσσουν διαφοροποιημένες συναρτήσεις χρόνου διαδρομής για καλύτερη αποτύπωση των κυκλοφοριακών χαρακτηριστικών.

Οι συναρτήσεις φόρτου-καθυστέρησης και οι μέθοδοι μηχανικής μάθησης αποτελούν βασικές προσεγγίσεις για την πρόβλεψη και ανάλυση της κυκλοφοριακής ροής. Οι συναρτήσεις φόρτου-καθυστέρησης, όπως η συνάρτηση του BPR, βασίζονται σε παραδοσιακές, εμπειρικές σχέσεις που συνδέουν τον κυκλοφοριακό όγκο με την καθυστέρηση, με στόχο την εκτίμηση του χρόνου διαδρομής και του κυκλοφοριακού φόρτου. Αυτές οι συναρτήσεις είναι συνήθως απλές και εύκολες στην εφαρμογή τους, όμως τείνουν να μην αποδίδουν ικανοποιητικά σε πολύπλοκα αστικά περιβάλλοντα, όπου η κυκλοφορία είναι έντονα μεταβαλλόμενη και επηρεάζεται από εξωτερικούς παράγοντες όπως οι καιρικές συνθήκες, οι σηματοδότες και οι πεζοί.

Αντίθετα, οι μέθοδοι μηχανικής μάθησης χρησιμοποιούν προηγμένα υπολογιστικά εργαλεία και μεγάλα σύνολα δεδομένων για να «μάθουν» από ιστορικά πρότυπα κυκλοφορίας και να ενσωματώσουν μεταβλητές που επηρεάζουν άμεσα την κυκλοφορία, όπως τις καιρικές συνθήκες και τον κυκλοφοριακό φόρτο σε πραγματικό χρόνο. Έτσι, τα μοντέλα μηχανικής μάθησης είναι

περισσότερο δυναμικά και προσαρμόσιμα σε διαφορετικές κυκλοφοριακές συνθήκες, επιτυγχάνοντας συνήθως υψηλότερη ακρίβεια στις προβλέψεις, ιδιαίτερα σε μη-γραμμικά και μεταβαλλόμενα περιβάλλοντα.

Κεφάλαιο 3

Μεθοδολογική Προσέγγιση

3.1 Ροή εργασιών

Για την πρόβλεψη χρόνων διαδρομής σε οδικά τμήματα, αναπτύχθηκαν αλγόριθμοι βασισμένοι σε μεθόδους μηχανικής μάθησης, με δεδομένα εισόδου ιστορικά δεδομένα. Οι αλγόριθμοι που χρησιμοποιήθηκαν για την επίτευξη του αναφερόμενου στόχου που βασίζονται στην επιβλεπόμενη μηχανική μάθηση είναι:

- Γραμμική Παλινδρόμηση (Linear Regression)
- Παλινδρόμηση Κορυφογραμμών (Ridge Regression)
- Παλινδρόμηση Lasso (Lasso Regression)
- Ελαστικά δίκτυα (Elastic nets)
- Τυχαία Δάση (Random Forests)
- Μηχανές Υποστήριξης Παλινδρόμησης (SVR)
- Ενισχυμένα Δένδρα Απόφασης (Gradient Boosting Decision Trees - GBDT)

Με τη χρήση των παραπάνω αλγορίθμων αναπτύχθηκαν μοντέλα που εκπαιδεύτηκαν με δεδομένα χρόνων διαδρομής από το Google maps καθώς και με πληροφορίες που αφορούν στον κυκλοφοριακό φόρτο επαγωγικών βρόχων στα παρακάτω οδικά δίκτυα στην Αθήνα, από όπου αξιολογήθηκε η δυνατότητα πρόβλεψής τους.

Τα τμήματα του οδικού δικτύου που διερευνήθηκαν στην παρούσα εργασία είναι:

- Ποσειδώνος – Λεωφόρος Κηφισού
- Λεωφόρος Κηφισού – Ιερά Οδός
- Λεωφόρος Αθηνών – Λεωφόρος Κηφισού
- Λεωφόρος Αθανάσιου Διάκου – Ανδρέα Συγγρού
- Λ. Βασιλίσσης Αμαλίας – Σοφίας
- Λ. Βασιλέως Κων/νου – Λ. Βασιλίσσης Σοφίας

- Λεωφόρος Αλεξάνδρας
- Λ. Βασιλίσσης Σοφίας – Αλεξάνδρας
- Λεωφόρος Κατεχάκη

Σημειώνεται πως οι περισσότερες έρευνες σύμφωνα με τη βιβλιογραφία την τελευταία περίοδο βασίζονται στη δημιουργία μοντέλων μηχανικής μάθησης για την δημιουργία προσομοιώσεων ενός οδικού δικτύου. Τέτοια μοντέλα φαίνεται ότι αποδίδουν καλύτερα στην πρόβλεψη κυκλοφοριακών συνθηκών σε ένα οδικό τμήμα σε σχέση με τις συμβατικές μεθόδους.

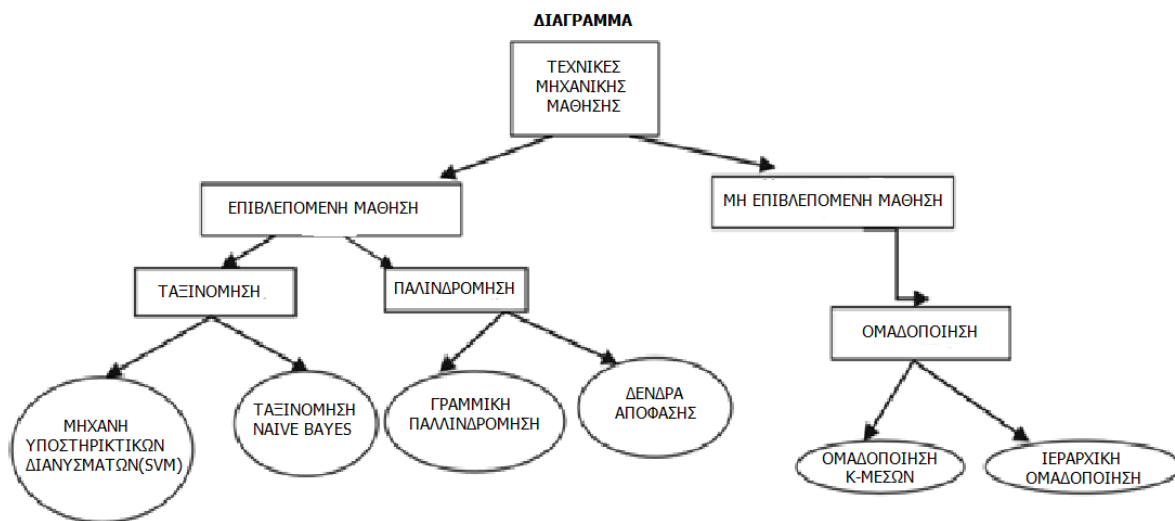
3.2. Μηχανική Μάθηση

Ο όρος Τεχνητή Νοημοσύνη χρησιμοποιήθηκε για πρώτη φορά το 1955 από τον John Mc Carthy. Η Μηχανική Μάθηση αποτελεί έναν κλάδο της Τεχνητής Νοημοσύνης. Είναι γνωστό ότι οι άνθρωποι είναι επιρρεπείς σε λάθη και ότι η έρευνα τους βασίζεται αρκετά συχνά σε υποκειμενικά κριτήρια. Για το λόγο αυτό, η ένταξη της μηχανικής μάθησης στην επίλυση αυτού του προβλήματος αποτελεί καταλυτικό ρόλο. Οι μέθοδοι της μηχανικής μάθησης έχουν δημιουργηθεί με σκοπό να μιμηθούν την ανθρώπινη νοημοσύνη, χωρίς να λαμβάνουν υπόψιν υποκειμενικές παραμέτρους που λαμβάνει ο άνθρωπος.

Οι δύο βασικές μέθοδοι της μηχανικής μάθησης που χρησιμοποιούνται ευρέως για τη δημιουργία μοντέλων είναι οι παρακάτω (Εικόνα 6):

- Μη επιβλεπόμενη μάθηση (Unsupervised learning): Πρόκειται για μοντέλα που εκπαιδεύονται χωρίς να είναι γνωστή η πληροφορία του αποτελέσματος του αθροίσματος των τιμών των επιμέρους ανεξάρτητων μεταβλητών. Σε αυτή την περίπτωση, οι αλγόριθμοι που στηρίζονται στη μη επιβλεπόμενη ταξινόμηση, εκπαιδεύονται και λαμβάνουν υπόψη πρότυπα και μοτίβα από τα υπάρχοντα δεδομένα για την επίλυση του εκάστοτε προβλήματος.
- Επιβλεπόμενη μάθηση (Supervised learning): Στην συγκεκριμένη περίπτωση, μια σειρά δεδομένων εξόδου (ή τιμές της εξαρτημένης μεταβλητής) είναι γνωστή, κατά τη διαδικασία της εκπαίδευσης. Στην περίπτωση Ταξινόμησης, το αποτέλεσμα είναι μια ετικέτα, όπως για

παράδειγμα στην περίπτωση της Δυαδικής Ταξινόμησης που υπάρχουν δύο περιπτώσεις αποτελεσμάτων, οι οποίες συνήθως εκφράζονται με 0 ή 1. Σε περιπτώσεις προβλημάτων πρόβλεψης το αποτέλεσμα της συνάρτησης είναι ένας πραγματικός αριθμός, συνήθως (συνεχή δεδομένα).



Εικόνα 6: Οι δύο κύριες κατηγορίες της μηχανικής μάθησης, η επιβλεπόμενη και μη επιβλεπόμενη μάθηση (Πηγή: Sindhu & Suriya, 2020)

3.3. Μετρικές Αξιολόγησης

Στην κλασική μηχανική εκμάθηση, μια σημαντική παράμετρος είναι η επιλογή των σωστών χαρακτηριστικών (ή ανεξάρτητων μεταβλητών) που περιγράφουν καλά τα δεδομένα. Αυτά τα χαρακτηριστικά βοηθούν στη βελτίωση της απόδοσης και στην κατανόηση των αποτελεσμάτων. Για παράδειγμα, μπορεί να χρειαστεί να προσαρμοστεί η κλίμακα των δεδομένων, χρησιμοποιώντας κάποιες τεχνικές κανονικοποίησης των τιμών. Αφού επιλεγθούν τα κατάλληλα χαρακτηριστικά, ακολουθεί η εκπαίδευση του μοντέλου με σκοπό να γενικεύει και να έχει καλή απόδοση σε νέα δεδομένα. Αναφέρεται επίσης πως γίνεται διαχωρισμός των δεδομένων σε συλλογή εκπαίδευσης και συλλογή δοκιμής/ελέγχου ώστε το μοντέλο να μάθει μοτίβα από τα δεδομένα εκπαίδευσης και να αξιολογήσει την απόδοσή του στα δεδομένα ελέγχου. Ο

γενικότερος σκοπός κάθε μοντέλου είναι να ελαχιστοποιηθούν τα σφάλματα, προσαρμόζοντας τις παραμέτρους του. Επιπλέον, καθορίζονται από τον χειριστή οι κατάλληλοι υπερπαραμέτροι για τη βελτίωση της απόδοσης του μοντέλου. Τέλος, πραγματοποιείται η αξιολόγηση του μοντέλου χρησιμοποιώντας μετρήσεις σφάλματος ή αλλιώς μετρικά αξιολόγησης για να μελετηθεί πόσο καλά ταιριάζουν τα πραγματικά δεδομένα με τα δεδομένα πρόβλεψης.

Τα μετρικά που χρησιμοποιούνται συνήθως για την αξιολόγηση των μοντέλων πρόβλεψης που χαρακτηρίζουν τον κυκλοφοριακό φόρτο παρουσιάζονται παρακάτω (Medina-Salgado et al., 2022):

- $$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i| \quad (4)$$

Το MAE (Mean Absolute Error - Μέσο Απόλυτο Σφάλμα) είναι ένα στατιστικό μετρικό αξιολόγησης που χρησιμοποιείται για την εκτίμηση της απόδοσης ενός μοντέλου πρόβλεψης και στην σχέση (4) παρουσιάζεται ο μαθηματικός του τύπος. Συγκεκριμένα, με τη χρήση του MAE υπολογίζεται η μέση απόλυτη διαφορά μεταξύ των πραγματικών και των προβλεπόμενων τιμών που παράγει το μοντέλο. Το MAE είναι χρήσιμο διότι παρέχει μια σαφή ένδειξη του μέσου μεγέθους των σφαλμάτων πρόβλεψης σε όρους των μονάδων των δεδομένων. Μικρότερες τιμές MAE υποδεικνύουν καλύτερη απόδοση του μοντέλου, δηλαδή μικρότερη απόκλιση των προβλέψεων από τις πραγματικές τιμές.

- $$MAPE = \frac{1}{N} \sum_{i=1}^N \frac{|y_i - \hat{y}_i|}{y_i} \cdot 100\% \quad (5)$$

Η χρήση του στατιστικού μέτρου MAPE (Mean Absolute Percentage Error - Μέσο Απόλυτο Ποσοστιαίο Σφάλμα) συνδράμει στον υπολογισμό της μέσης ποσοστιαίας διαφοράς μεταξύ των πραγματικών και των προβλεπόμενων τιμών που παράγει ένα μοντέλο και ο μαθηματικός του τύπος παρατίθεται στην σχέση (5). Το MAPE είναι χρήσιμο διότι εκφράζει το σφάλμα πρόβλεψης ως ποσοστό, καθιστώντας το εύκολο να ερμηνευτεί και να συγκριθεί μεταξύ διαφορετικών συνόλων δεδομένων ή μοντέλων. Μικρότερες τιμές MAPE περιγράφουν καλύτερη απόδοση του μοντέλου, δηλαδή μικρότερη απόκλιση των προβλέψεων από τις πραγματικές τιμές σε ποσοστιαία βάση.

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

Το MSE (Mean Squared Error - Μέσο Τετραγωνικό Σφάλμα) μετρά τη μέση τιμή των τετραγώνων των διαφορών μεταξύ των πραγματικών τιμών και των προβλεπόμενων τιμών που παράγει το μοντέλο και παραπάνω παρουσιάζεται στην σχέση (6) ο τύπος του. Το MSE είναι χρήσιμο γιατί δίνει έμφαση σε μεγαλύτερα σφάλματα λόγω της τετραγωνικής φύσης του. Με άλλα λόγια, μεγάλες αποκλίσεις μεταξύ των πραγματικών και προβλεπόμενων τιμών θα έχουν μεγαλύτερη επίδραση στο MSE. Μικρότερες τιμές MSE υποδεικνύουν καλύτερη απόδοση του μοντέλου, δηλαδή μικρότερη απόκλιση των προβλέψεων από τις πραγματικές τιμές.

- $$\text{RMSE} = \frac{1}{N} \sqrt{\sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad (7)$$

Με το RMSE (Root Mean Squared Error - Ριζική Μέση Τετραγωνική Απόκλιση) υπολογίζεται η ρίζα του μέσου όρου των τετραγώνων των διαφορών μεταξύ των πραγματικών τιμών και των προβλεπόμενων τιμών. Ο μαθηματικός τύπος της Ριζικής Μέσης Τετραγωνικής απόκλισης μπορεί να περιγραφεί με τον τύπο της σχέσης (7). Το RMSE είναι χρήσιμο επειδή παρέχει μια σαφή ένδειξη της μέσης απόκλισης των προβλέψεων του μοντέλου από τις πραγματικές τιμές, εκφρασμένη στις ίδιες μονάδες με τα δεδομένα. Μικρότερες τιμές RMSE υποδεικνύουν καλύτερη απόδοση του μοντέλου, δηλαδή μικρότερη απόκλιση των προβλέψεων από τις πραγματικές τιμές.

Σημειώνεται ότι το y_i υποδηλώνει της πραγματικές τιμές εξόδου ενώ \hat{y}_i τις τιμές πρόβλεψης.

Εκτός από τον έλεγχο των αποτελεσμάτων του μοντέλου, είναι εξίσου σημαντικός ο έλεγχος της προσαρμογής των προβλέψεων στα πραγματικά δεδομένα. Συχνά, ένα μοντέλο που έχει εκπαιδευτεί ενδέχεται να μην έχει αρκετή πολυπλοκότητα για να αντικατοπτρίσει πλήρως τις σχέσεις στα υπάρχοντα και νέα δεδομένα (underfitting), και να παρουσιάζει περιορισμένη χρησιμότητα. Αντίστοιχα, μπορεί να παρουσιάζει εξαιρετική απόδοση στα υπάρχοντα δεδομένα και υπερβολικό σφάλμα στα νέα (overfitting). Άρα, το μοντέλο μπορεί να ανακαλύπτει μη τυχαία μοτίβα στα δεδομένα, καθιστώντας το αναποτελεσματικό στη γενίκευση σε νέα δεδομένα. Για να ελεγχθεί η προσαρμογή των μοντέλων στα δεδομένα, χρησιμοποιούνται διάφορα στατιστικά

μέτρα, όπως το R^2 (συντελεστής προσδιορισμού), το οποίο εφαρμόζεται σε μοντέλα παλινδρόμησης και παρέχει πληροφορίες σχετικά με την ικανότητά τους να προβλέπουν. Παρακάτω, στην σχέση (8) παρουσιάζεται ο τύπος του συντελεστή προσδιορισμού (Alexander et al., 2015):

$$R^2 = 1 - \frac{\sum (y - \hat{y})^2}{\sum (y - \bar{y})^2} \quad (8)$$

Με τη συνδρομή του συντελεστή προσδιορισμού προσδιορίζεται το ποσοστό διακύμανσης για μια εξαρτημένη μεταβλητή που μπορεί να εξηγηθεί από το μοντέλο. Αξιολογείται η προσαρμογή στα δεδομένα, συγκρίνοντας τις παραγόμενες τιμές του μοντέλου με τον αντίστοιχο μέσο όρο της μεταβλητής. Το R^2 λαμβάνει τιμές από 0 έως 1, όπου η τιμή 1 δηλώνει άριστη προσαρμογή στα δεδομένα και ότι το μοντέλο εξηγεί όλη τη διακύμανση που εμφανίζεται στα δεδομένα, ενώ η τιμή 0 σηματοδοτεί ότι το μοντέλο δεν εξηγεί τη διακύμανση των δεδομένων και το αποτέλεσμα επηρεάζεται από μια άγνωστη πηγή παραλλακτικότητας. Στην παρούσα διπλωματική εργασία ως μετρικές αξιολόγησης των αποτελεσμάτων χρησιμοποιούνται ο συντελεστής προσδιορισμού R^2 και το μέσο απόλυτο ποσοστιαίο σφάλμα (MAPE).

3.4. Αλγόριθμοι Μηχανικής Μάθησης

Στη συνέχεια, παρουσιάζονται οι αλγόριθμοι μηχανικής μάθησης που χρησιμοποιήθηκαν για την ανάπτυξη των μοντέλων πρόβλεψης της ροής κυκλοφορίας στα αναφερόμενα οδικά δίκτυα της Αθήνας.

3.4.1 Γραμμική Παλινδρόμηση

Το μοντέλο μηχανικής μάθησης Linear Regression (Γραμμική Παλινδρόμηση) είναι μία από τις βασικότερες αλλά και ευρέως χρησιμοποιούμενες τεχνικές στην ανάλυση δεδομένων και στη μηχανική μάθηση. Βασίζεται στην υπόθεση ότι υπάρχει μια γραμμική σχέση μεταξύ της ανεξάρτητης μεταβλητής (ή μεταβλητών) και της εξαρτημένης μεταβλητής. Στόχος της είναι να προσδιορίσει την καλύτερη ευθεία (γραμμικό μοντέλο) που περιγράφει τα δεδομένα, με τρόπο που να ελαχιστοποιείται το σφάλμα πρόβλεψης. Η μαθηματική μορφή της παρουσιάζεται παρακάτω στην σχέση (9). Αναλυτικότερα ως β_0 ορίζεται η σταθερά (το σημείο τομής), ως β_1

είναι ο συντελεστής της ανεξάρτητης μεταβλητής x , και ε είναι το σφάλμα. Σύμφωνα με τον James et al. (2013), η γραμμική παλινδρόμηση αποτελεί μια απλή αλλά ισχυρή μέθοδο για την κατανόηση και πρόβλεψη δεδομένων, και έχει αποδειχτεί ιδιαίτερα αποτελεσματική σε περιπτώσεις όπου οι σχέσεις μεταξύ των μεταβλητών είναι, πράγματι, γραμμικές.

$$y = \beta_0 + \beta_1 x + \varepsilon \quad (9)$$

3.4.2. Παλινδρόμηση Κορυφογραμμών

Η Παλινδρόμηση Κορυφογραμμών (Ridge Regression) είναι μια τεχνική που χρησιμοποιείται στην πολλαπλή γραμμική παλινδρόμηση όταν υπάρχουν προβλήματα συσχέτισης, δηλαδή όταν οι ανεξάρτητες μεταβλητές είναι υψηλά συσχετισμένες μεταξύ τους. Για να αντιμετωπιστεί αυτό το πρόβλημα, η συγκεκριμένη μέθοδος χρησιμοποιεί μια παράμετρο στο μοντέλο. Από αυτήν την παράμετρο ελέγχεται η ισορροπία ώστε να προσαρμοστούν ικανοποιητικά τα δεδομένα και να αποφευχθεί το φαινόμενο της υπερπροσαρμογής, το οποίο παρατηρείται όταν το μοντέλο γίνεται υπερβολικά πολύπλοκο και προσαρμόζεται στο θόρυβο των δεδομένων αντί για το πραγματικό μοτίβο που περιγράφουν τα δεδομένα (McDonald, 2009).

Οι συντελεστές στην Παλινδρόμηση Κορυφογραμμών, που είναι οι τιμές που ανατίθενται σε κάθε ανεξάρτητη μεταβλητή στην συνάρτηση παλινδρόμησης, συμπεριφέρονται διαφορετικά, ανάλογα με την παράμετρο ridge. Μέσω της παρατήρησης στο πώς αλλάζουν αυτοί οι συντελεστές για διάφορες τιμές της παραμέτρου ridge, παρέχονται πληροφορίες για το πώς λειτουργεί η παλινδρόμηση ώστε να βελτιωθεί η απόδοση του μοντέλου όταν εμφανίζονται ανεξάρτητες μεταβλητές που συσχετίζονται μεταξύ τους. Κατά βάση, η συγκεκριμένη μέθοδος παλινδρόμησης συνδράμει στο να βρεθεί μια καλή συμβιβαστική λύση στην διατήρηση της καλής προσαρμογής των δεδομένων και στη δημιουργία απλών και αξιόπιστων μοντέλων πρόβλεψης (McDonald, 2009).

3.4.3. Παλινδρόμηση Lasso

Η μέθοδος της Παλινδρόμησης Lasso ομοιάζει με την περίπτωση της Παλινδρόμησης Κορυφογραμμών. Τα κλασικά μοντέλα παλινδρόμησης χρησιμοποιούνται συνήθως στις

στατιστικές αναλύσεις για την πρόβλεψη αποτελεσμάτων, αλλά οι τυποποιημένες μέθοδοι μπορούν να οδηγήσουν στην υπερπροσαρμογή των δεδομένων και σε προκαταλήψεις κατά τη διαδικασία της εκπαίδευσης. Η παλινδρόμηση Lasso αντιμετωπίζει αυτά τα προβλήματα συρρικνώνοντας τους σταθερούς όρους της συνάρτησης παλινδρόμησης προς το μηδέν για μεταβλητές που δεν είναι αντιπροσωπευτικές στο να καθορίσουν το ορθό αποτέλεσμα, επιλέγοντας τις κατάλληλες μεταβλητές που χαρακτηρίζουν το πρόβλημα, κανονικοποιώντας τα δεδομένα και ελαχιστοποιώντας το σφάλμα πρόβλεψης, με αποτέλεσμα να μην υπάρχει άγνωστη πηγή παραλλακτικότητας και τα δεδομένα πρόβλεψης να προσαρμόζονται πολύ καλά με τα πραγματικά δεδομένα. Πιο συγκεκριμένα, η Lasso παλινδρόμηση χρησιμοποιεί μια παράμετρο περιορισμού για τον έλεγχο της πολυπλοκότητας του μοντέλου με τη χρήση μιας ειδικής μεθόδου διασταυρωμένης επικύρωσης (cross validation), προσφέροντας μια ισορροπία μεταξύ της μείωσης της πολυπλοκότητας και τη διατήρηση της προγνωστικής ακρίβειας (Ranstam & Cook, 2018). Παρακάτω διακρίνεται η συνάρτηση που χρησιμοποιείται για την περίπτωση της παλινδρόμησης Lasso (10). Πιο συγκεκριμένα, το διάνυσμα β περιγράφει μια αντικειμενική συνάρτηση ελαχιστοποίησης σφάλματος λαμβάνοντας υπόψιν τον συντελεστή f_{β} ο οποίος είναι μια θετική παράμετρος κανονικοποίησης, σε μια συλλογή ανεξάρτητων μεταβλητών (DeMol et al., 2009).

$$\beta_n = \arg \min_{\beta = (\beta_{\gamma})_{\gamma \in \Gamma}} \left(\frac{1}{n} \sum_{i=1}^n |Y_i - f_{\beta}(X_i)|^2 + \lambda \sum_{\gamma \in \Gamma} |\beta_{\gamma}| \right) \quad (10)$$

3.4.4. Ελαστικά Δίκτυα

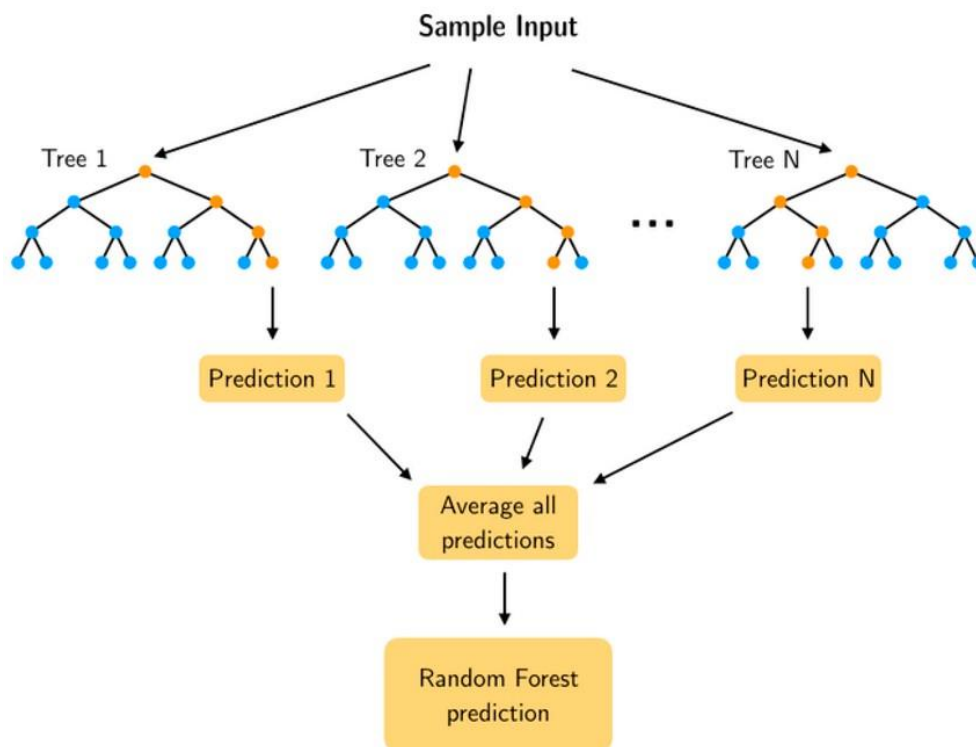
Η μέθοδος των ελαστικών δικτύων (elastic nets) είναι παρόμοια σε σχέση με τις δύο αναφερόμενες μεθόδους παλινδρόμησης για την πρόβλεψη των δεδομένων. Η μέθοδος αυτή εστιάζει στην επιλογή κατάλληλων μεταβλητών για την καλύτερη ακρίβεια πρόβλεψης. Αναλυτικότερα, συνδυάζει δύο ποινές (penalties) για την εξισορρόπηση της σπανιότητας και της σταθερότητας, ιδιαίτερα παρουσία συσχετισμένων χαρακτηριστικών. Η συνάρτηση αυτή έχει ως στόχο να αποδείξει τη συνέπεια της εκτίμησης κάτω από διαφορετικές επιλογές της παραμέτρου κανονικοποίησης και παρέχει πληροφορίες σχετικά με την επαναλαμβανόμενη διαδικασία ορίου που χρησιμοποιείται για τον υπολογισμό της εκτίμησης παρουσία συσχετιζόμενων χαρακτηριστικών (DeMol et al., 2009).

3.4.5. Τυχαία Δάση

Στην επιστήμη των δεδομένων, ο αλγόριθμος “Τυχαία Δάση” χρησιμοποιείται συχνά για τη δημιουργία αποφάσεων από πολλά δένδρα αποφάσεων. Πρόκειται για μια μέθοδο στην οποία εφαρμόζεται η τεχνική bagging, συνδυάζοντας έναν μεγάλο αριθμό δένδρων αποφάσεων ώστε να μειωθεί ο κίνδυνος υπερπροσαρμογής. Ο αλγόριθμος αυτός ανήκει στην κατηγορία του ensemble learning και χρησιμοποιείται τόσο για προβλήματα παλινδρόμησης όσο και ταξινόμησης.

Για προβλήματα ταξινόμησης, κάθε δένδρο απόφασης παράγει ένα αποτέλεσμα, το οποίο ανήκει σε μια κλάση (ή κατηγορία). Η κλάση που θα επικρατήσει από την πλειοψηφία των αποτελεσμάτων των δένδρων αποφάσεων θα είναι και η τελική απόφαση.

Για προβλήματα παλινδρόμησης, υπολογίζεται ο μέσος όρος των αποτελεσμάτων όλων των δένδρων αποφάσεων, ώστε να προκύψει η τελική προβλεπόμενη τιμή (Alam et al., 2019). Στην παρακάτω Εικόνα 7 παρουσιάζεται οπτικά η λειτουργία του αλγορίθμου “Τυχαία Δάση”.



Εικόνα 7: Ο αλγόριθμος επιβλεπόμενης Ταξινόμησης "Τυχαία Δάση"

3.4.6. Μηχανές Υποστήριξης Παλινδρόμησης

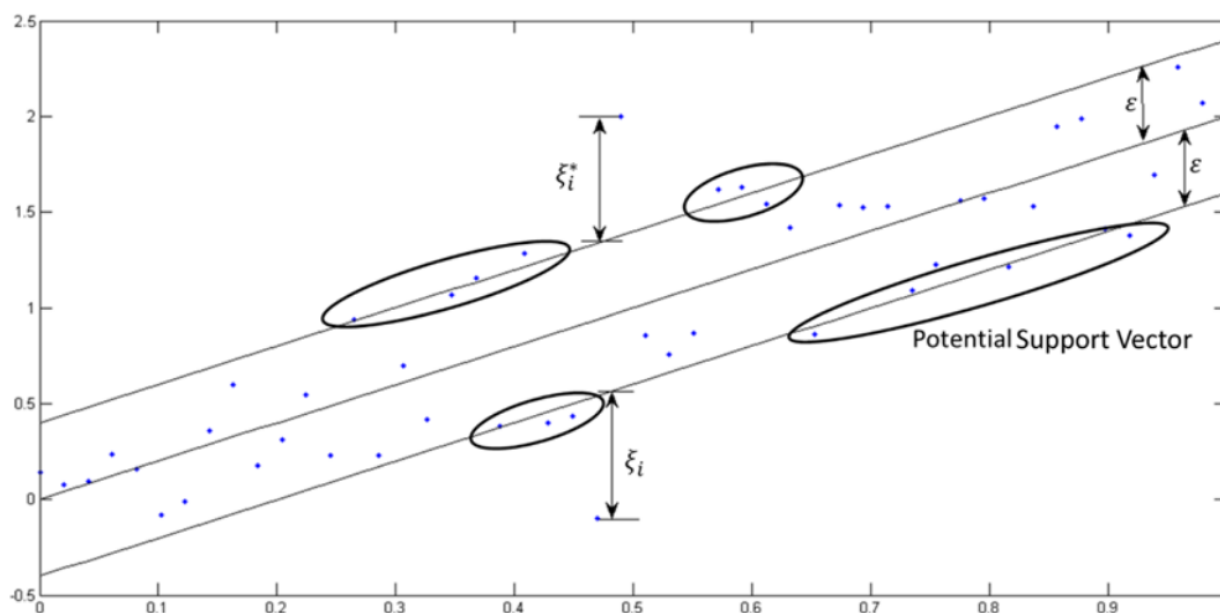
Ο στόχος των Μηχανών Υποστήριξης Παλινδρόμησης (SVR) είναι να βρεθεί μια συνάρτηση που να προσεγγίζει τη σχέση μεταξύ των δεδομένων εισόδου και των τιμών εξόδου όσο το δυνατόν ακριβέστερα. Πρόκειται για έναν αλγόριθμο που χρησιμοποιεί ως αντικειμενική συνάρτηση, την συνάρτηση βελτιστοποίησης.

Για να επιτευχθεί αυτό, η μέθοδος SVR επιδιώκει να εντοπίσει τον στενότερο δυνατό "σωλήνα" γύρω από την προβλεπόμενη λειτουργία (Εικόνα 9). Αυτός ο σωλήνας μπορεί να απεικονιστεί ως μια περιοχή που περιβάλλει την προβλεπόμενη συνάρτηση, όπου τα πραγματικά σημεία δεδομένων ιδανικά εμπίπτουν. Το πλάτος αυτού του σωλήνα καθορίζεται από μια παράμετρο, που συχνά αναφέρεται ως epsilon (ϵ), η οποία καθορίζει τη μέγιστη επιτρεπόμενη απόκλιση μεταξύ των προβλεπόμενων τιμών και των πραγματικών τιμών που θεωρείται αποδεκτή (Awad et al., 2015).

Η διαδικασία βελτιστοποίησης περιλαμβάνει δύο κύριους στόχους:

- Ελαχιστοποίηση σφάλματος πρόβλεψης: Αυτό αναφέρεται στη μείωση της απόστασης μεταξύ των προβλεπόμενων αποτελεσμάτων και των πραγματικών τιμών-στόχων. Η μέθοδος SVR έχει ως στόχο να διασφαλίσει ότι η πλειοψηφία των σημείων δεδομένων εμπίπτουν εντός του "σωλήνα", πράγμα που σημαίνει ότι το σφάλμα πρόβλεψης για αυτά τα σημεία είναι εντός της αποδεκτής περιοχής που καθορίζεται από το epsilon. Για σημεία που πέφτουν εκτός αυτού του σωλήνα, το μοντέλο περιλαμβάνει μια ποινή στη διαδικασία βελτιστοποίησης για να λογοδοτήσει για αυτά τα σφάλματα.
- Εύρεση του στενότερου σωλήνα: Ο αλγόριθμος SVR προσπαθεί επίσης να κάνει αυτόν τον σωλήνα όσο το δυνατόν στενότερο. Ένας στενότερος σωλήνας συνεπάγεται ένα πιο ακριβές μοντέλο με λιγότερη απόκλιση από τα πραγματικά σημεία δεδομένων. Ωστόσο, όσο πιο στενή είναι η περιοχή του "σωλήνα", τόσο πιο εύκολο είναι οι προβλεπόμενες τιμές να βρίσκονται έξω από αυτήν.

Εξισορροπώντας αυτούς τους δύο στόχους, το SVR διατυπώνει ένα πρόβλημα βελτιστοποίησης όπου ελαχιστοποιεί μια λειτουργία κόστους (Awad et al., 2015).



3.4.7. Ενισχυμένα Δένδρα Απόφασης

Το Gradient Boosting Decision Tree (GBDT) είναι ένας αλγόριθμος που συνδυάζει την τεχνική gradient boosting με τον αλγόριθμο decision tree. Στο GBDT, χρησιμοποιούνται δέντρα λήψης αποφάσεων ως «αδύναμοι μαθητές» και το μοντέλο δημιουργείται σταδιακά, βελτιστοποιώντας κάθε φορά τη λειτουργία απώλειας. Η ενίσχυση (boosting) είναι μια μέθοδος μάθησης του συνόλου που συνδυάζει πολλούς αδύναμους μαθητές για να δημιουργήσει έναν ισχυρό μαθητή, χρησιμοποιώντας επαναληπτικές μεθόδους (Wang et al., 2018).

Σε αντίθεση με άλλες μεθόδους boosting, ο αλγόριθμος Gradient Boosting ενημερώνει συνεχώς τη λειτουργία απώλειας και την κλίση της για να βελτιώσει το μοντέλο. Από την άλλη, ο αλγόριθμος Decision Tree είναι μια βασική μέθοδος για προβλήματα ταξινόμησης και προβλήματα πρόβλεψης. Έχει τη δομή του δέντρου, όπου κάθε εσωτερικός κόμβος αντιπροσωπεύει μια δοκιμή σε ένα χαρακτηριστικό, κάθε κόμβος φύλλου αντιπροσωπεύει μια κατηγορία, και κάθε κλάδος αντιπροσωπεύει το αποτέλεσμα μιας δοκιμής (Wang et al., 2018).

3.5. Προσδιορισμός Υπερπαραμέτρων

Για την λειτουργία των μοντέλων μηχανικής μάθησης απαραίτητος είναι ο προσδιορισμός κατάλληλων υπερπαραμέτρων. Οι υπερπαραμέτροι σε ένα μοντέλο μηχανικής μάθησης είναι σημαντικοί, καθώς καθορίζουν την συμπεριφορά και την απόδοση του αλγορίθμου. Οι τελευταίοι δεν «μαθαίνονται» από τα δεδομένα αλλά ορίζονται πριν την εκπαίδευση του μοντέλου, όπως το ποσοστό μάθησης, το μέγεθος των δειγμάτων από τα δεδομένα εκπαίδευσης (batches) και η πολυπλοκότητα του μοντέλου.

Οι υπερπαραμέτροι που ορίζονται για κάθε μοντέλο μηχανικής μάθησης είναι απαραίτητοι για την βελτίωση της ακρίβειας, της σταθερότητας, της αποτελεσματικότητας και της ικανότητας γενίκευσης του μοντέλου. Η ορθή ρύθμιση των υπερπαραμέτρων είναι κρίσιμη, διότι μπορεί να οδηγήσει στην καλύτερη δυνατή απόδοση ενός μοντέλου μηχανικής μάθησης ή αντίθετα στην δημιουργία ενός μοντέλου που «μαθαίνει» υπερβολικά τις λεπτομέρειες των δεδομένων,

οδηγώντας στην υπερεκπαίδευση. Παρακάτω στο κεφάλαιο αυτό, παρουσιάζονται οι υπερπαραμέτροι που απαιτούνται να καθοριστούν για κάθε μοντέλο μηχανικής μάθησης που χρησιμοποιήθηκε στην παρούσα έρευνα.

Αναλυτικότερα, τα μοντέλα της παλινδρόμησης Κορυφογραμμών (Ridge Regression) και της παλινδρόμησης Lasso (Lasso Regression), χρησιμοποιούν την παράμετρο α ή λ και την max_iter . Το α ή λ είναι η πιο σημαντική υπερπαραμέτρος των δύο αυτών μοντέλων και καθορίζει την δύναμη της κανονικοποίησης (regularization). Όσο μεγαλύτερη η τιμή του α τόσο λιγότερο θα προσαρμοστεί η γραμμική παλινδρόμηση στις λεπτομέρειες των δεδομένων, αποφεύγοντας την υπερπροσαρμογή των δεδομένων (overfitting). Αντίθετα, για μικρές τιμές της υπερπαραμέτρου α , για παράδειγμα $\alpha = 0$, η παλινδρόμηση κορυφογραμμών και η παλινδρόμηση Lasso παίρνει την μορφή της κλασικής γραμμικής παλινδρόμησης χωρίς κανονικοποίηση, με αποτέλεσμα την μεγαλύτερη ευαισθησία στις διακυμάνσεις των δεδομένων. Συνοψίζοντας, η αύξηση του α μειώνει τους συντελεστές και κάνει το μοντέλο πιο σταθερό αλλά μπορεί να οδηγήσει σε αντίθετα αποτελέσματα αν αυξηθεί σε μεγάλο βαθμό.

Η υπερπαραμέτρος max_iter καθορίζει τον μέγιστο αριθμό επαναλήψεων που θα πραγματοποιηθούν ώστε να έχει καλύτερη απόδοση το μοντέλο. Η γρήγορη εκπαίδευση ενός μοντέλου επιδιώκεται με χαμηλότερες τιμές των επαναλήψεων. Όμως σε μοντέλα που χρησιμοποιούν μεγάλο όγκο δεδομένων παρατηρήθηκε πως η αύξηση της συγκεκριμένης υπερπαραμέτρου οδηγεί σε βελτίωση της σύγκλισης.

Στο μοντέλο των Ελαστικών Δικτύων (Elastic Net) εκτός από την υπερπαραμέτρο α που αναλύθηκε παραπάνω απαραίτητη κρίνεται και η υπερπαραμέτρος $l1_ratio$. Δεδομένο ότι το μοντέλο των Ελαστικών Δικτύων συνδυάζει τόσο στοιχεία τόσο από την κανονικοποίηση της Παλινδρόμησης Κορυφογραμμών (L2 regularization) και της Παλινδρόμησης Lasso (L1 regularization), πετυχαίνει τον έλεγχο της ισορροπίας μεταξύ των ιδιοτήτων των επιμέρους μοντέλων μέσω της παραμέτρου $l1_ratio$.

Τα μοντέλα Τυχαία Δάση (Random Forest) και Ενισχυμένα Δέντρα απόφασης (Gradient Boosting) ορίζουν τις ακόλουθες υπερπαραμέτρους ώστε να προσαρμόσουν το μοντέλο στα δεδομένα του προβλήματος και να βελτιώσουν την απόδοσή του, την ακρίβειά του και την

ικανότητα γενίκευσης αποτελεσμάτων: `n_estimators`, `max_depth`, `min_samples_split`, `min_samples_leaf`, `max_features`, `learning_rate` και `subsample`. Συγκεκριμένα, η παράμετρος `n_estimators` (αριθμός εκτιμήσεων) περιγράφει τον αριθμό των δένδρων απόφασης που θα δημιουργήσει το κάθε μοντέλο. Η αύξηση του αριθμού των δένδρων απόφασης οδηγεί σε καλύτερη απόδοση του μοντέλου, αλλά ταυτόχρονα έχει κόστος την πιο αργή εκπαίδευσή του. Το μέγιστο βάθος κάθε δένδρου απόφασης καθορίζεται από την παράμετρο `max_depth`. Ο ρόλος της τελευταίας είναι περιορίζει το βάθος του κάθε δένδρου ώστε να επιτυγχάνεται η γενίκευση του μοντέλου και να αποφευχθούν φαινόμενα `overfitting`.

Σε συνέχεια των περιγραφών των υπερπαραμέτρων, η `min_samples_split` καθορίζει τον ελάχιστο αριθμό δειγμάτων που είναι απαραίτητοι σε έναν κόμβο να πραγματοποιηθεί διαχωρισμός του κόμβου σε επιμέρους κλάδους. Ο ορισμός ενός μεγάλου αριθμού στην υπερπάρμετρο αυτή περιορίζει την δημιουργία υπερβολικά πολλών κόμβων και οδηγούν στην δημιουργία λιγότερο ευαίσθητων δένδρων, μειώνοντας τον κίνδυνο να μην αποδίδει το μοντέλο σε νέα δεδομένα (`overfitting`). Ο ελάχιστος αριθμός δειγμάτων ανά φύλλο απεικονίζεται με την παράμετρο `min_samples_leaf`. Όσο μεγαλύτερη τιμή δέχεται ως όρισμα η `min_samples_leaf` το γενικευμένα γίνονται τα δένδρα. Ρόλος της παραμέτρου αυτής είναι να εμποδίσει την δημιουργία φύλλων με μικρό αριθμό δειγμάτων και επομένως την υπερεκπαίδευσή τους. Επιπλέον, η παράμετρος `max_features` ελέγχει την τυχαιότητα των δένδρων. Μικρότερος αριθμός χαρακτηριστικών οδηγεί σε μεγαλύτερη ποικιλία στα δένδρα και αποτρέπει το `overfitting`. Τα ορίσματα εισόδου είναι : `"sqrt"` που αφορά την τετραγωνική ρίζα των συνολικών χαρακτηριστικών και `"log2"` που αναφέρεται στον αριθμό των χαρακτηριστικών και ορίζεται ως ο λογάριθμος των συνολικών χαρακτηριστικών στη βάση του 2. Τέλος, η υπερπάρμετρος `learning_rate` ελέγχει την συνεισφορά κάθε νέου δένδρου στο σύνολο, ενώ η `subsample` καθορίζει αν θα πραγματοποιηθεί υπερδειγματοληψία από τα δεδομένα κάθε δένδρου, γεγονός που οδηγεί στην βελτιωμένη γενίκευση του μοντέλου. Συνήθως, μικρότερες τιμές της `learning_rate` όπως 0,01 έως 0,1 οδηγούν σε μεγαλύτερη απόδοση, ενώ για την `subsample` τιμές μικρότερες της μονάδας βοηθά στην γενίκευση του μοντέλου και την μείωση του `overfitting`.

Ολοκληρώνοντας την περιγραφή των υπερπαραμέτρων κάθε μοντέλου μηχανικής μάθησης, ακολουθεί η ανάλυση αυτών για τις Μηχανές Υποστήριξης Παλινδρόμησης (SVR). Η

υπερπαράμετρος C καθορίζει τον βαθμό επιρροής των σφαλμάτων στο μοντέλο. Μεγαλύτερες τιμές του C αυξάνουν την ακρίβεια των δεδομένων, μειώνοντας τα σφάλματα. Αντίθετα, μικρότερες τιμές επιτρέπουν την δημιουργία περισσότερων σφαλμάτων προωθώντας την απλότητα και επομένως προστατεύουν το μοντέλο από φαινόμενα overfitting. Ακολουθεί, η παράμετρος kernel που καθορίζει τη συνάρτηση του πυρήνα που θα χρησιμοποιηθεί για να χαρτογραφήσει τα δεδομένα σε έναν χώρο υψηλότερης διάστασης. Τιμές του πυρήνα είναι ο γραμμικός (linear), ο πολυωνυμικός (poly) και ο μη γραμμικός πυρήνας βασισμένος σε Gaussians (Radial Basis Function). Η τελευταία αποτελεί την πιο κοινή επιλογή της παραμέτρου του πυρήνα. Για το μοντέλο SVR ορίζεται και η παράμετρος gamma, η οποία υποδηλώνει την επιρροή ενός σημείου δεδομένων στο συνολικό μοντέλο και χρησιμοποιείται σε μη γραμμικούς πυρήνες. Καλύτερη προσαρμογή των δεδομένων που χρησιμοποιούνται στο μοντέλο επιτυγχάνεται με μικρότερες τιμές της παραμέτρου, ενώ αντίθετα μεγάλες τιμές του gamma ενδέχεται να σε υπερβολική προσαρμογή των δεδομένων.

Συνοψίζοντας όλα τα παραπάνω, στον Πίνακα 2 παρουσιάζονται οι υπερπαράμετροι που χρησιμοποιήθηκαν στα μοντέλα μηχανικής μάθησης στην παρούσα εργασία.

Πίνακας 2: Περιγραφή των υπερπαραμέτρων που χρησιμοποιήθηκαν σε όλα τα μοντέλα μηχανικής μάθησης

ΜΟΝΤΕΛΟ	ΥΠΕΡΠΑΡΑΜΕΤΡΟΙ	ΠΕΡΙΓΡΑΦΗ
Linear Regression	-	-
Ridge Regression	alpha max_inter	Παράμετρος κανονικοποίησης Μέγιστος αριθμός επαναλήψεων
Lasso Regressio	alpha max_inter	Παράμετρος κανονικοποίησης Μέγιστος αριθμός επαναλήψεων
Elastic Net	alpha l1_ratio	Παράμετρος κανονικοποίησης Αναλογία L1(Lasso)/L2(Ridge)
Random Forest	n-estimators max_depth min_sample_split min_sample_leaf max_features	Αριθμός εκτιμήσεων Μέγιστο βάθος δένδρου Ελάχιστος αριθμός δειγμάτων για split Ελάχιστος αριθμός δειγμάτων για φύλλο Μέγιστος αριθμός χαρακτηριστικών

Gradient Boosting	n_estimators learnig_rate max_depth min_samples_split min_sample_leaf subsample	Αριθμός εκτιμήσεων Ρυθμός μάθησης Μέγιστο βάθος δένδρου Ελάχιστος αριθμός δειγμάτων για split Ελάχιστος αριθμός δειγμάτων για φύλλο Ποσοστό υποδειγμάτων
SVR	C kernel gamma	Ελάχιστος αριθμός δειγμάτων για ένα φύλλο Πυρήνας Παράμετρος επιρροής του πυρήνα

Κεφάλαιο 4

4.1. Εισαγωγή

Σε αυτό το κεφάλαιο εξηγείται η συλλογή και επεξεργασία δεδομένων που ήταν αναγκαία για την επίτευξη του στόχου της εργασίας και η διαδικασία εφαρμογής των αλγορίθμων μηχανικής μάθησης στα δεδομένα. Για τη δημιουργία των μοντέλων πρόβλεψης με τη χρήση των αναφερόμενων αλγορίθμων, η ανάλυση των δεδομένων έγινε σε περιβάλλον Jupyter Notebooks με τη χρήση της γλώσσας προγραμματισμού Python.

4.2. Λήψη και Διαχείριση Δεδομένων

Η λήψη των δεδομένων έγινε με τη χρήση αισθητήρων οι οποίοι είναι εγκατεστημένοι σε στατικά σημεία του οδικού δικτύου, τα οποία ονομάζονται φωρατές. Με τη συλλογή αυτών των

δεδομένων δημιουργήθηκαν σημεία που περικλείουν τους φωρατές με απόσταση μεγαλύτερη ή ίση των 100 μέτρων και τα οποία βρίσκονται πριν από σηματοδοτούμενους κόμβους, σε περιβάλλον QGIS. Έπειτα, αφού καταγράφηκαν οι συντεταγμένες των σημείων αυτών, δημιουργήθηκαν μοντέλα πρόβλεψης του κυκλοφοριακού φόρτου και της διάρκειας διαδρομής, χρησιμοποιώντας δεδομένα οδικού δικτύου από το περιβάλλον της google και το Κέντρο Διαχείρισης της Κυκλοφορίας (data.gov.gr).

Αναλυτικότερα, η λήψη της πληροφορίας της μεταβλητής “χρόνος διαδρομής” πραγματοποιήθηκε σε συνεχόμενο ημερήσιο επίπεδο από τις 6 έως 19 Μαρτίου 2023 στα χρονικά διαστήματα 8:00-9:00, 11:00-12:00, 17:00-18:00, 20:00-21:00 ανά ένα τέταρτο για κάθε χρονικό διάστημα ξεχωριστά.

Ακόμη, όπως αναφέρθηκε, οι μεταβλητές που περιγράφουν τις συνθήκες κυκλοφορίας στο οδικό δίκτυο επηρεάζονται σημαντικά και από άλλες παραμέτρους, όπως είναι οι καιρικές συνθήκες, κοινωνικές και πολιτικές κινητοποιήσεις (απεργίες, διαδηλώσεις, στάσεις μετρό, ημιμαραθώνιος). Για το λόγο αυτό λήφθηκαν υπόψιν και αυτά τα στοιχεία, για τη δημιουργία σταθερών και αντιπροσωπευτικών μοντέλων με την πραγματικότητα. Τέλος, για τις αναφερόμενες ημερομηνίες, πραγματοποιήθηκε η λήψη δεδομένων σχετικά με τα οχήματα που πέρασαν από σημείο σε σημείο μεταξύ των φωρατών ανά μία ώρα για κάθε εικοσιτετράωρο.

4.3. Ανάλυση δεδομένων

Ακολούθησε η επεξεργασία των δεδομένων που λήφθηκαν. Σε πρώτο στάδιο, έγινε η διαλογή των φωρατών, καθώς από κάποιες δεν ήταν εφικτή η λήψη των δεδομένων, λόγω του ότι δεν λειτουργούσαν, ή και να λειτουργούσαν, δεν έδωσαν δεδομένα για τις ημερομηνίες και ώρες που αναφέρθηκαν παραπάνω. Σε επόμενο στάδιο, διατηρήθηκε μόνο η πληροφορία των φωρατών για τα συγκεκριμένα χρονικά διαστήματα που σημειώθηκαν. Αναλυτικότερα, χρησιμοποιήθηκε η παραδοχή για τα στοιχεία της google που λαμβάνονταν ανά τέταρτο κάθε χρονικού διαστήματος, ότι το τρίτο τέταρτο, όπως για παράδειγμα το 8:45, χαρακτηρίζει ολόκληρο το χρονικό διάστημα 8:00-9:00. Ο λόγος αυτής της παραδοχής οφείλεται στο γεγονός ότι τα στοιχεία από τη google λαμβάνονταν ανά τέταρτο, ενώ τα στοιχεία από το Κέντρο Διαχείρισης της Κυκλοφορίας data.gov.gr ανά ώρα. Στον παρακάτω Πίνακα 3, διακρίνεται η δομή

των δεδομένων που σχετίζονται με τον χρόνο (ή διάρκεια) διαδρομής, καθώς και με τον αριθμό των οχημάτων που διέρχονται από το εκάστοτε οδικό αστικό δίκτυο τις ημέρες και ώρες λήψης τους από κάθε φωρατή.

Πίνακας 3: Δομή δεδομένων για δημιουργία μοντέλων μηχανικής μάθησης

Φωρατές	Ημερομηνία	Ωρα	Αριθμός Οχημάτων	Χρόνος Διαδρομής	Πληροφορίες Οδικού Δικτύου
MS128	6/3/2023	11:00:00Z	752	4	ΚΥΡΙΟΣ ΔΡΟΜΟΣ ΜΕ ΚΑΤΕΥΘΥΝΣΗ ΛΑΜΙΑ ΠΡΙΝ ΑΠΟ ΤΗ ΡΑΜΠΑ ΕΙΣΟΔΟΥ ΣΤΗ Λ. ΚΗΦΙΣΟΥ ΑΠΟ Λ. ΠΟΣΕΙΔΩΝΟΣ ΑΠΟ ΓΛΥΦΑΔΑ
MS128	6/3/2023	17:00:00Z	1032	4	ΚΥΡΙΟΣ ΔΡΟΜΟΣ ΜΕ ΚΑΤΕΥΘΥΝΣΗ ΛΑΜΙΑ ΠΡΙΝ ΑΠΟ ΤΗ ΡΑΜΠΑ ΕΙΣΟΔΟΥ ΣΤΗ Λ. ΚΗΦΙΣΟΥ ΑΠΟ Λ. ΠΟΣΕΙΔΩΝΟΣ ΑΠΟ ΓΛΥΦΑΔΑ
MS128	6/3/2023	20:00:00Z	859	4	ΚΥΡΙΟΣ ΔΡΟΜΟΣ ΜΕ ΚΑΤΕΥΘΥΝΣΗ ΛΑΜΙΑ ΠΡΙΝ ΑΠΟ ΤΗ ΡΑΜΠΑ ΕΙΣΟΔΟΥ ΣΤΗ Λ. ΚΗΦΙΣΟΥ ΑΠΟ Λ. ΠΟΣΕΙΔΩΝΟΣ ΑΠΟ ΓΛΥΦΑΔΑ
MS129	6/3/2023	11:00:00Z	2024	6	ΡΑΜΠΑ ΕΙΣΟΔΟΥ ΑΠΟ Λ. ΠΟΣΕΙΔΩΝΟΣ ΑΠΟ ΓΛΥΦΑΔΑ ΣΤΗ Λ. ΚΗΦΙΣΟΥ ΜΕ ΚΑΤΕΥΘΥΝΣΗ ΛΑΜΙΑ
MS129	6/3/2023	17:00:00Z	2539	6	ΡΑΜΠΑ ΕΙΣΟΔΟΥ ΑΠΟ Λ. ΠΟΣΕΙΔΩΝΟΣ ΑΠΟ ΓΛΥΦΑΔΑ ΣΤΗ Λ. ΚΗΦΙΣΟΥ ΜΕ ΚΑΤΕΥΘΥΝΣΗ ΛΑΜΙΑ
MS129	6/3/2023	20:00:00Z	1803	5	ΡΑΜΠΑ ΕΙΣΟΔΟΥ ΑΠΟ Λ. ΠΟΣΕΙΔΩΝΟΣ ΑΠΟ ΓΛΥΦΑΔΑ ΣΤΗ Λ. ΚΗΦΙΣΟΥ ΜΕ ΚΑΤΕΥΘΥΝΣΗ ΛΑΜΙΑ
MS130	6/3/2023	11:00:00Z	2413	5	ΚΥΡΙΟΣ ΔΡΟΜΟΣ ΜΕ ΚΑΤΕΥΘΥΝΣΗ ΓΛΥΦΑΔΑ ΜΕΤΑ ΤΗ ΡΑΜΠΑ ΕΞΟΔΟΥ ΤΗΣ Λ. ΚΗΦΙΣΟΥ ΠΡΟΣ ΠΕΙΡΑΙΑ (ΚΟΜΒΟΣ ΠΟΣΕΙΔΩΝΟΣ)
MS130	6/3/2023	17:00:00Z	3375	6	ΚΥΡΙΟΣ ΔΡΟΜΟΣ ΜΕ ΚΑΤΕΥΘΥΝΣΗ ΓΛΥΦΑΔΑ ΜΕΤΑ ΤΗ ΡΑΜΠΑ ΕΞΟΔΟΥ ΤΗΣ Λ. ΚΗΦΙΣΟΥ ΠΡΟΣ ΠΕΙΡΑΙΑ (ΚΟΜΒΟΣ ΠΟΣΕΙΔΩΝΟΣ)
MS130	6/3/2023	20:00:00Z	2135	5	ΚΥΡΙΟΣ ΔΡΟΜΟΣ ΜΕ ΚΑΤΕΥΘΥΝΣΗ ΓΛΥΦΑΔΑ ΜΕΤΑ ΤΗ ΡΑΜΠΑ ΕΞΟΔΟΥ ΤΗΣ Λ. ΚΗΦΙΣΟΥ ΠΡΟΣ ΠΕΙΡΑΙΑ (ΚΟΜΒΟΣ ΠΟΣΕΙΔΩΝΟΣ)
MS131	6/3/2023	11:00:00Z	945	11	ΡΑΜΠΑ ΕΞΟΔΟΥ ΠΡΟΣ ΠΕΙΡΑΙΑ ΤΗΣ Λ. ΚΗΦΙΣΟΥ ΜΕ ΚΑΤΕΥΘΥΝΣΗ ΠΕΙΡΑΙΑ (ΚΟΜΒΟΣ ΠΟΣΕΙΔΩΝΟΣ)
...
...
...
MS983	18/3/2023	17:00:00Z	444	17	ΡΑΜΠΑ ΕΞΟΔΟΥ ΠΡΟΣ ΙΕΡΑ ΟΔΟ ΤΟΥ ΚΛΑΔΟΥ ΤΗΣ Λ. ΚΗΦΙΣΟΥ ΜΕ ΚΑΤΕΥΘΥΝΣΗ ΛΑΜΙΑ 40 Μ. ΠΡΙΝ ΑΠΟ ΤΗΝ ΙΕΡΑ ΟΔΟ (ΜΟΝΟ ΓΙΑ Φ/Σ)
MS983	18/3/2023	20:00:00Z	407	17	ΡΑΜΠΑ ΕΞΟΔΟΥ ΠΡΟΣ ΙΕΡΑ ΟΔΟ ΤΟΥ ΚΛΑΔΟΥ ΤΗΣ Λ. ΚΗΦΙΣΟΥ ΜΕ ΚΑΤΕΥΘΥΝΣΗ ΛΑΜΙΑ 40 Μ. ΠΡΙΝ ΑΠΟ ΤΗΝ ΙΕΡΑ ΟΔΟ (ΜΟΝΟ ΓΙΑ Φ/Σ)
MS984	18/3/2023	08:00:00Z	482	14	ΚΥΡΙΟΣ ΔΡΟΜΟΣ ΜΕ ΚΑΤΕΥΘΥΝΣΗ ΑΘΗΝΑ 40 Μ. ΠΡΙΝ ΑΠΟ ΤΗ Λ. ΚΗΦΙΣΟΥ (ΜΟΝΟ ΓΙΑ Φ/Σ)
MS984	18/3/2023	11:00:00Z	653	19	ΚΥΡΙΟΣ ΔΡΟΜΟΣ ΜΕ ΚΑΤΕΥΘΥΝΣΗ ΑΘΗΝΑ 40 Μ. ΠΡΙΝ ΑΠΟ ΤΗ Λ. ΚΗΦΙΣΟΥ (ΜΟΝΟ ΓΙΑ Φ/Σ)
MS984	18/3/2023	17:00:00Z	501	15	ΚΥΡΙΟΣ ΔΡΟΜΟΣ ΜΕ ΚΑΤΕΥΘΥΝΣΗ ΑΘΗΝΑ 40 Μ. ΠΡΙΝ ΑΠΟ ΤΗ Λ. ΚΗΦΙΣΟΥ (ΜΟΝΟ ΓΙΑ Φ/Σ)
MS984	18/3/2023	20:00:00Z	540	14	ΚΥΡΙΟΣ ΔΡΟΜΟΣ ΜΕ ΚΑΤΕΥΘΥΝΣΗ ΑΘΗΝΑ 40 Μ. ΠΡΙΝ ΑΠΟ ΤΗ Λ. ΚΗΦΙΣΟΥ (ΜΟΝΟ ΓΙΑ Φ/Σ)
MS985	18/3/2023	08:00:00Z	114	17	ΚΥΡΙΟΣ ΔΡΟΜΟΣ ΜΕ ΚΑΤΕΥΘΥΝΣΗ ΑΙΓΑΛΕΩ 25 Μ. ΠΡΙΝ ΑΠΟ ΤΗ Λ. ΚΗΦΙΣΟΥ (ΜΟΝΟ ΓΙΑ Φ/Σ)

MS985	18/3/2023	11:00:00Z	217	24	ΚΥΡΙΟΣ ΔΡΟΜΟΣ ΜΕ ΚΑΤΕΥΘΥΝΣΗ ΑΙΓΑΛΕΩ 25 Μ. ΠΡΙΝ ΑΠΟ ΤΗ Λ. ΚΗΦΙΣΟΥ (ΜΟΝΟ ΓΙΑ Φ/Σ)
MS985	18/3/2023	17:00:00Z	181	21	ΚΥΡΙΟΣ ΔΡΟΜΟΣ ΜΕ ΚΑΤΕΥΘΥΝΣΗ ΑΙΓΑΛΕΩ 25 Μ. ΠΡΙΝ ΑΠΟ ΤΗ Λ. ΚΗΦΙΣΟΥ (ΜΟΝΟ ΓΙΑ Φ/Σ)
MS985	18/3/2023	20:00:00Z	110	17	ΚΥΡΙΟΣ ΔΡΟΜΟΣ ΜΕ ΚΑΤΕΥΘΥΝΣΗ ΑΙΓΑΛΕΩ 25 Μ. ΠΡΙΝ ΑΠΟ ΤΗ Λ. ΚΗΦΙΣΟΥ (ΜΟΝΟ ΓΙΑ Φ/Σ)

4.4. Επεξεργασία και δόμηση δεδομένων

Για την ανάλυση των δεδομένων στο περιβάλλον αναλύσεων Jupyter Notebook χρειάστηκε η χρήση ορισμένων βιβλιοθηκών ώστε να καταστεί εφικτή η επεξεργασία τους. Κάποιες από τις βιβλιοθήκες που χρησιμοποιήθηκαν αναφέρονται παρακάτω:

- **Pandas:** Πρόκειται για μια βιβλιοθήκη, η οποία παρέχει τα απαραίτητα εργαλεία για τη διαχείριση και ανάλυση δεδομένων. Συγκεκριμένα, παρέχει συναρτήσεις και δομές για την παρουσίαση των δεδομένων (Data Frames, Series).
- **NumPy:** Στη βιβλιοθήκη αυτή βασίστηκε η δημιουργία της βιβλιοθήκης Pandas. Η NumPy (Numerical Python) παρέχει μια μεγάλη συλλογή μαθηματικών συναρτήσεων, όπως είναι η τριγωνομετρικές, αλγεβρικές και στατιστικές συναρτήσεις και με τη χρήση αυτής, η δόμηση των δεδομένων γίνεται σε πίνακες (arrays and matrices).
- **SciPy:** Χρησιμοποιείται για επιστημονικές και τεχνικές αναλύσεις. Χτίστηκε πάνω στις δυνατότητες της NumPy και παρέχει έναν μεγάλο αριθμό σύνθετων συναρτήσεων.
- **Matplotlib:** Είναι μια βιβλιοθήκη που συνδράμει στην οπτικοποίηση των αποτελεσμάτων, καθώς παρέχει πληθώρα γραφημάτων (scatterplots, boxplots, heatmaps) όπου είναι εφικτή η αναπαράσταση τους με διάφορους συνδυασμούς χρωμάτων.
- **Scikit-learn:** Η βιβλιοθήκη αυτή περιέχει έναν μεγάλο αριθμό αλγορίθμων μηχανικής μάθησης καθώς και το απαραίτητα μετρικά αξιολόγησης που απαιτούνται για τον έλεγχο απόδοσης ενός δημιουργούμενου μοντέλου πρόβλεψης (ή ταξινόμησης)

Στη συνέχεια, παρουσιάζονται αναλυτικά οι εξαρτημένες μεταβλητές (countedcars, flow_per_lane) και οι ανεξάρτητες μεταβλητές που χρησιμοποιήθηκαν για την εκπαίδευση των αλγορίθμων, ώστε να δημιουργηθούν τα μοντέλα πρόβλεψης της κίνησης στο οδικό δίκτυο:

- Countedcars: περιγράφει τον αριθμό των οχημάτων που διασχίζουν το κομμάτι μελέτης (τον φωρατή) σε κάθε χρονικό πλαίσιο που έγινε η λήψη δεδομένων, δηλαδή παρέχει δεδομένα σχετικά με τη ροή στο συνολικό δίκτυο
- Flow_per_lane: δίνει πληροφορίες σχετικά με τη ροή των οχημάτων σε κάθε λωρίδα που βρίσκεται ο φωρατής σε κάθε κόμβο μελέτης
- Travel_time: πρόκειται για τον χρόνο που χρειάστηκε ώστε τα οχήματα να περάσουν από κάθε φωρατή μέσα στα χρονικά πλαίσια που έγινε η λήψη των δεδομένων.
- Peak: αφορά την περίπτωση ώρας αιχμής. Λαμβάνει τιμές 0 και 1 αν δεν είναι ώρα αιχμής ή αν είναι ώρα αιχμής, αντίστοιχα.
- Normal_date: παίρνει 0 αν δεν έχει πορείες, απεργίες είτε κάποιο γεγονός που επηρεάζει την κυκλοφορία και 1 αν έχει
- Bad_weather: δέχεται την τιμή 0 αν δεν επικρατούν δυσμενείς καιρικές συνθήκες και 1 αν επικρατούν
- Lanes_occupation: αφορά τον αριθμό κατάληψης λωρίδων από τον φωρατή
- Nd_lanes: πρόκειται για τον αριθμό λωρίδων στο οδικό δίκτυο μελέτης
- Capacity: αφορά την χωρητικότητα του δικτύου
- Capacity_per_lane: πρόκειται για την χωρητικότητα κάθε λωρίδας του δικτύου
- Func_class: περιγράφει τη λειτουργική κατάταξη του οδικού τμήματος

- `Signalized`: παίρνει τιμές 0 αν η διασταύρωση στην οποία βρίσκεται ο φωρατής δεν έχει σηματοδότη και 1 αν έχει
- `is_weekend`: λαμβάνει την τιμή 0 αν πρόκειται για καθημερινή μέρα και την τιμή 1 αν είναι σαββατοκύριακο

Η έρευνα αυτή βασίστηκε στη θεωρία της επιβλεπόμενης μηχανικής μάθησης (supervised machine learning). Σε αυτόν τον τύπο μάθησης όπως αναφέρθηκε, για την εκπαίδευση είναι απαραίτητο να είναι γνωστές οι τιμές των εξαρτημένων μεταβλητών. Έτσι τα μοντέλα εκτίμησης του αριθμού των οχημάτων στους συγκεκριμένους κόμβους έρευνας έχουν ως όρισμα εξόδου την μεταβλητή `flow_per_lane`. Με την συγκεκριμένη εξαρτημένη μεταβλητή είναι δυνατή η πρόβλεψη του κυκλοφοριακού φόρτου ανά λωρίδα, δίνοντας περισσότερες πληροφορίες για την κυκλοφοριακή ικανότητα του κάθε συνδέσμου και παράλληλα βοηθώντας μετέπειτα βελτίωσή της.

4.5. Αποτελέσματα

Έπειτα από την διαμόρφωση και κατάλληλη επεξεργασία των δεδομένων για τη διαδικασία αναλύσεων, σύμφωνα με τη θεωρία της επιβλεπόμενης ταξινόμησης, υλοποιήθηκε η διαδικασία διαχωρισμού της συλλογής δεδομένων σε συλλογή εκπαίδευσης και συλλογή ελέγχου με χρήση του 80% και 20% του συνολικού αριθμού δεδομένων αντίστοιχα. Στη συνέχεια ορίστηκαν τα εύρη τιμών για τις υπερπαραμέτρους του εκάστοτε μοντέλου. Η διαδικασία εύρεσης του κατάλληλου συνδυασμού υπερπαραμέτρων πραγματοποιήθηκε σύμφωνα με τις μεθόδους βελτιστοποίησης `Randomized Search` και `Grid Search` της βιβλιοθήκη `scikitlearn`. Σημειώνεται πως τα μοντέλα που δημιουργήθηκαν έχουν ως εξαρτημένη μεταβλητή τη ροή των οχημάτων σε κάθε λωρίδα που βρίσκεται ο φωρατής σε κάθε κόμβο μελέτης (`flow_per_lane`).

Ύστερα από τον καθορισμό του κατάλληλου συνδυασμού υπερπαραμέτρων για κάθε μοντέλο που δημιουργήθηκε σύμφωνα με τις αναφερόμενες μεθόδους βελτιστοποίησης, βρέθηκε ότι ο αλγόριθμος που αποδείχθηκε αποτελεσματικότερος στο να περιγράψει τον κυκλοφοριακό φόρτο είναι αυτός του μοντέλου `Gradient Boosting` με την μέθοδο βελτιστοποίησης `Random Search`. Η μέθοδος βελτιστοποίησης αυτή κρίνεται πιο αποτελεσματική καθώς δίνει εύκολα συγκρίσιμα αποτελέσματα με λιγότερο υπολογιστικό κόστος σε σύνθετα και με μεγάλο αριθμό υπερπαραμέτρων μοντέλα. Αναλυτικότερα στον Πίνακα 4 παρουσιάζεται η απόδοση του

καλύτερου μοντέλου με την χρήση των μετρικών R^2 και MAPE που κρίθηκαν καταλυτικοί και αναλύθηκαν στην Ενότητα 3.3. Ενώ στον Πίνακα 5 απεικονίζεται για το μοντέλο αυτό, ο καλύτερος συνδυασμός υπερπαραμέτρων που το κάνει πιο αποτελεσματικό και πιο σταθερό με δεδομένα εξόδου που ανταποκρίνονται καλύτερα στην πραγματικότητα.

Πίνακας 4: Απόδοση του καλύτερου μοντέλου που χρησιμοποιήθηκε στην παρούσα έρευνα σύμφωνα με τη μέθοδο βελτιστοποίησης Random Search ($y = flow_per_lane$)

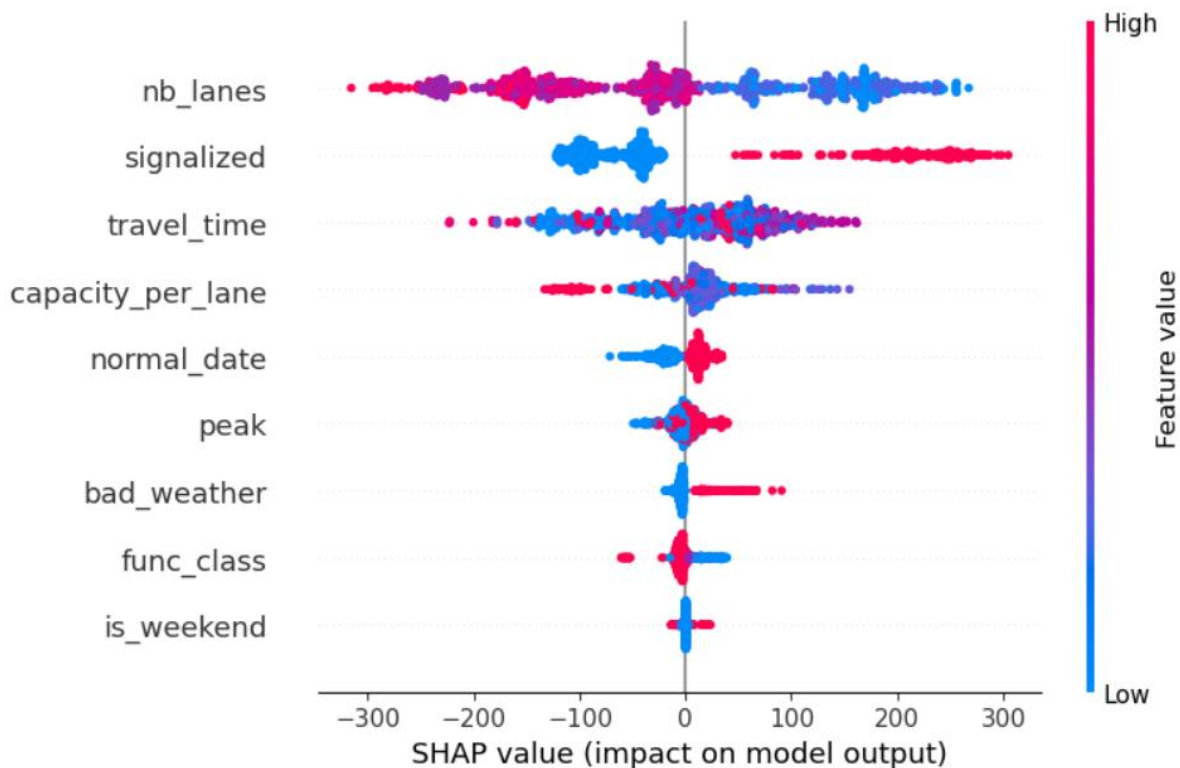
Μοντέλο Μηχανικής μάθησης	R^2	MAPE
Gradient Boosting	0,755	0,390

Πίνακας 5: Οι βέλτιστες τιμές των υπερπαραμέτρων του καλύτερου αλγορίθμου που χρησιμοποιήθηκαν κατά τη διαδικασία εκπαίδευσης για τη δημιουργία μοντέλου πρόβλεψης κυκλοφοριακού φόρτου με την μέθοδο βελτιστοποίησης Random Search

Μοντέλο Μηχανικής Μάθησης	Υπερπαραμέτροι	Τελική Τιμή
Gradient Boosting	n_estimators	15
	learning_rate	0,2
	max_depth	5
	min_samples_splits	2
	min_samples_leaf	2
	sudsample	1.0

Στη συνέχεια, αφού αποδείχτηκε ότι πιο αποτελεσματικό μοντέλο πρόβλεψης είναι αυτό που δημιουργήθηκε σύμφωνα με τον αλγόριθμο Gradient Boosting, δημιουργήθηκε και σχετικό διάγραμμα SHAP. Πιο συγκεκριμένα, ένα διάγραμμα SHAP περιγράφει τις συνεισφορές των μεμονωμένων ανεξάρτητων μεταβλητών στο να καθορίσουν την εξαρτημένη μεταβλητή. Όσο πιο υψηλή είναι η συνεισφορά μιας μεταβλητής εισόδου, σημαίνει ότι αυτή συμμετέχει σε μεγάλο ποσοστό στον καθορισμό του αποτελέσματος, δηλαδή στην πρόβλεψη της τιμής της μεταβλητής εξόδου (ή εξαρτημένης μεταβλητής). Αυτό έχει ως αποτέλεσμα να γίνεται καλύτερη

ερμηνεία της πρόβλεψης σε σχέση με την πραγματικότητα. Για το λόγο αυτό και στην παρούσα έρευνα δημιουργήθηκε ένα διάγραμμα SHAP που αφορά την πρόβλεψη των τιμών της εξαρτημένης μεταβλητής "flow_per_lane". Το διάγραμμα αυτό παρουσιάζεται στην παρακάτω Εικόνα 9.



Εικόνα 9: Διάγραμμα SHAP για έλεγχο συνεισφοράς μεταβλητών εισόδου στην πρόβλεψη της μεταβλητής εξόδου "flow_per_lane"

Ο οριζόντιος άξονας στο παραπάνω διάγραμμα (Εικόνα 9) δείχνει την τιμή SHAP, δηλαδή πόσο επηρεάζει κάθε χαρακτηριστικό την τελική πρόβλεψη του μοντέλου. Οι αρνητικές τιμές υποδηλώνουν την αρνητική επίδραση στην πρόβλεψη, ενώ οι θετικές τιμές την δείχνουν την θετική επίδραση. Όσο οι τιμές απομακρύνονται από το μηδέν, τόσο μεγαλύτερη η επιρροή των μεταβλητών στο μοντέλο. Εκτός από την ποσοτική απεικόνιση του βαθμού επίδρασης μεταβλητών του μοντέλου, παρουσιάζεται και με χρωματική κλίμακα ο ίδιος βαθμός επίδρασης. Συγκεκριμένα, η κλίμακα χρωμάτων από το μπλέ (χαμηλή τιμή του χαρακτηριστικού) έως το κόκκινο (υψηλή τιμή του χαρακτηριστικού) απεικονίζει την σημαντικότητα κάθε δεδομένου στο

μοντέλο μηχανικής μάθησης που χρησιμοποιήθηκε. Επομένως, με το διάγραμμα SHAP μπορεί να παρουσιαστεί η σημαντικότητα κάθε τιμής εξόδου στις προβλέψεις του μοντέλου καθώς και το πως επηρεάζουν οι τελευταίες, θετικά ή αρνητικά, τις προβλέψεις αυτές.

Από το δεδομένο διάγραμμα SHAP γίνεται διακριτό ότι η μεταβλητή "is_weekend" καθορίζει σε μικρότερο βαθμό την πρόβλεψη της μεταβλητής "flow_per_lane" ενώ η ανεξάρτητη μεταβλητή nb_lanes σε μεγαλύτερο. Τα χαρακτηριστικά όπως το nb_lanes, το signalized και το travel_time φαίνεται να έχουν ισχυρή επίδραση στα δεδομένα εξόδου του μοντέλου, με μεγάλες διακυμάνσεις. Ενώ τα χαρακτηριστικά όπως το bad_weather και το func_class και το normal_date έχουν μικρότερη επίδραση στο μοντέλο, καθώς η κατανομή των σημείων είναι πιο κοντά στο μηδέν.

Κεφάλαιο 5

Συμπεράσματα

5.1.Βασικά Συμπεράσματα

Σύμφωνα με την περίληψη των αποτελεσμάτων της παρούσας διπλωματικής εργασίας, μπορούν να εξαχθούν τα ακόλουθα συμπεράσματα:

Οι τεχνολογικές εξελίξεις έχουν θέσει τα θεμέλια για την ανάπτυξη ευφυών συστημάτων διαχείρισης κυκλοφορίας στα αστικά κέντρα. Η πληροφορία που λαμβάνεται από τους χρήστες ενός οδικού δικτύου, όπως στην περίπτωση της παρούσας διπλωματικής εργασίας, συμβάλλει στην πρόγνωση κυκλοφοριακών συνθηκών και τελικά προσφέρει καλύτερο επίπεδο εξυπηρέτησης. Για αυτόν τον λόγο, τα αποτελέσματα της παρούσας διπλωματικής εργασίας αποκτούν ιδιαίτερη σημασία.

Το μοντέλα μηχανικής μάθησης που αναπτύχθηκε με τη χρήση του αλγορίθμου "Ενισχυμένα Δένδρα Απόφασης"(Gradient Boosting) παρείχε υψηλή ακρίβεια πρόβλεψης για τις απαιτούμενες κυκλοφοριακές μεταβλητές δίνοντας χαμηλό ποσοστό σφάλματος και υψηλή τιμή συντελεστή προσδιορισμού.

Η χρήση συμβατικών μεθόδων στην πρόβλεψη του κυκλοφοριακού φόρτου είναι δύσκολη και δεν χαρακτηρίζει την πραγματικότητα, δεδομένου ότι η ροή κυκλοφορίας δεν μεταβάλλεται γραμμικά και σταθερά με το χρόνο.

Η επιλογή και διαχείριση των κατάλληλων χαρακτηριστικών (επιλογή και επεξεργασία χαρακτηριστικών) ήταν κρίσιμη για την εξασφάλιση υψηλής απόδοσης στα μοντέλα μηχανικής μάθησης. Επίσης, μέσω της σημαντικότητας των χαρακτηριστικών, έγινε σαφές ποιοι παράγοντες επηρεάζουν σημαντικά τη σχέση του κάθε χαρακτηριστικού με την εξαρτημένη μεταβλητή.

Ένα σημαντικό συμπέρασμα είναι ότι η επιτυχία των μοντέλων μηχανικής μάθησης εξαρτάται από την ποιότητα, την ποσότητα και τη μορφή των δεδομένων αλλά και από την επιλογή των κατάλληλων αλγορίθμων.

5.2 Περιορισμοί Έρευνας

Βάση των παραπάνω σημείων, μπορούν να επισημανθούν οι εξής περιορισμοί της έρευνας:

Είναι αβέβαιο αν τα μοντέλα που δημιουργήθηκαν και χρησιμοποιήθηκαν στην παρούσα διπλωματική διατριβή μπορούν να αναπαράγουν τις προβλεπτικές του ικανότητες σε διαφορετικά οδικά τμήματα.

Ενισχύοντας την παραπάνω αναφορά, η πολυπλοκότητα της πρόβλεψης των συνθηκών κυκλοφορίας σε επίπεδο δικτύου αυξάνεται λόγω της γεωμετρίας και των χωρικών συσχετίσεων μεταξύ των οδικών τμημάτων. Όπως αναφέρθηκε κατά την ανάλυση της βιβλιογραφίας, η έρευνα επικεντρώνεται στο απλούστερο πρόβλημα της πρόβλεψης κυκλοφορίας για ένα μόνο οδικό τμήμα. Η ανάπτυξη μοντέλων που μπορούν να λαμβάνουν υπόψη τις χωρικές συσχετίσεις μεταξύ των οδικών δικτύων μπορεί να οδηγήσει σε πιο ακριβείς προβλέψεις σε επίπεδο δικτύου.

5.3 Προτάσεις για Περαιτέρω Έρευνα

Οι προτάσεις για μελλοντική έρευνα παρουσιάζονται παρακάτω:

Τα μοντέλα που αναπτύχθηκαν σε αυτή τη διπλωματική εργασία θα μπορούσαν να βελτιωθούν με την ένταξη περισσότερων οδικών τμημάτων, την αύξηση του όγκου της ποικιλίας των δεδομένων, το πειραματισμό με διαφορετικούς συνδυασμούς ανεξάρτητων μεταβλητών και τη δοκιμή άλλων αλγορίθμων μηχανικής μάθησης. Η σύγκριση αυτών με μοντέλα βαθιάς μάθησης για τη δημιουργία ενός υβριδικού μοντέλου θα μπορούσε επίσης να είναι επωφελής.

Η πρόκληση της πρόβλεψης των συνθηκών κυκλοφορίας μπορεί να διευρυνθεί με την ενσωμάτωση επιπλέον μεταβλητών (features), όπως είναι τα καιρικά δεδομένα, οι οδικές σημάνσεις, καθώς και τα στατιστικά ατυχημάτων. Αυτό πιθανόν να βελτιώσει την ακρίβεια πρόβλεψης.

Τέλος, η χρήση οποιασδήποτε νέας τεχνολογίας, η οποία πιστοποιείται από τη βιβλιογραφία και αφορά το πρόβλημα που αναλύθηκε στην παρούσα μελέτη, θα συνδράμει στη βελτίωση των προβλέψεων της ροής κυκλοφορίας στο οδικό δίκτυο.

Βιβλιογραφία

Alam I., Farid D. M. & Rossetti R. J. (2019). The prediction of traffic flow with regression analysis. In *Emerging Technologies in Data Mining and Information Security: Proceedings of IEMIS 2018, Volume 2* (pp. 661-671). Springer Singapore.

Alexander D. L., Tropsha A., & Winkler, D. A. (2015). Beware of R²: Simple, unambiguous assessment of the prediction accuracy of QSAR and QSPR models. *Journal of Chemical Information and Modeling*, 55 (7), 1316-1322.

Andrey J., & Yagar, S. (1993). A temporal analysis of rain-related crash risk. *Accident Analysis & Prevention*, 25 (4), 465-472.

Awa M., Khanna R., Awad M., & Khanna R. (2015). Support vector regression. In *Efficient Learning Machines: Theories, Concepts, and Applications for Engineers and System Designers* (pp. 67-80). Springer.

Barnett, V. (2005). *Environmental statistics: Methods and applications*. John Wiley & Sons.

Barston D. (1976). Models of single lane time headway distributions. *Transportation Science*, 10(2), 125-148.

Boukerche, A., & Wang, J. (2020). Machine learning-based traffic prediction models for intelligent transportation systems. *Computer Networks*, 181, 107530.

Castro-Neto M., Jeong Y. S., Jeong M. K. & Han, L. D. (2009). Online-SVR for short-term traffic flow prediction under typical and atypical traffic conditions. *Expert Systems with Applications*, 36(3), 6164-6173.

Davidson K. B. (1966). A flow travel time relationship for use in transportation planning. In *Australian Road Research Board (ARRB) Conference, 3rd, 1966, Sydney* (Vol. 3, No. 1).

De Mol C., De Vito E. & Rosasco, L. (2009). Elastic-net regularization in learning theory. *Journal of Complexity*, 25(2), 201-230.

James G., Witten D., Hastie T., & Tibshirani R. (2013). *An Introduction to Statistical Learning: with Applications in R*. Springer, 59-63

Hou G., Chen S. & Bao, Y. (2022). Development of travel time functions for disrupted urban arterials with microscopic traffic simulation. *Physica A: Statistical Mechanics and its Applications*, 593, 126,961.

Kachroo, P. & Sastry, S. (2016). Traffic assignment using a density-based travel-time function for intelligent transportation systems. *IEEE Transactions on Intelligent Transportation Systems*, 17(5), 1438-1447.

Lee S., & Fambro D. B. (1999). Application of subset autoregressive integrated moving average model for short-term freeway traffic volume forecasting. *Transportation Research Record*, 1678(1), 179-188.

Li L., Su X., Wang Y., Lin Y., Li Z. & Li Y. (2015). Robust causal dependence mining in big data network and its application to traffic flow predictions. *Transportation Research Part C: Emerging Technologies*, 58, 292-307.

McDonald G. C. (2009). Ridge regression. *Wiley Interdisciplinary Reviews: Computational Statistics*, 1(1), 93-100.

McDonald J. F., d'Ouville E. L. & Liu L. N. (1999). *Economics of urban highway congestion and pricing* (Vol. 9). Springer Science & Business Media.

McCarthy J.C. (1955). Restrictive Covenants. U. Ill. LF,709

Medina-Salgado B., Sánchez-DelaCruz E., Pozos-Parra P. & Sierra J. E. (2022). Urban traffic flow prediction techniques: A review. *Sustainable Computing: Informatics and Systems*, 35, 100739.

Neuhold, R. & Fellendorf M. (2014). Volume delay functions based on stochastic capacity. *Transportation Research Record*, 2421(1), 93-102.

Okawa M., Kim H. & Toda H. (2017, May). Online traffic flow prediction using convolved bilinear poisson regression. In *2017 18th IEEE International Conference on Mobile Data Management (MDM)* (pp. 134-143). IEEE.

Olstam J. J. & Tapani A. (2004). Comparison of Car-following models (Vol.960). Linköping, Sweden: Swedish National Road and Transport Research Institute.

Oppe S. (1989). Macroscopic models for traffic and traffic safety. *Accident Analysis & Prevention*, 21(3), 225-232.

Pan Y., Guo J., Chen Y., Abbasi M., List G., Zhou X. A review of volume-delay functions: Connecting theoretical fundamental, practical deployment and emerging applications, 29-34

Peterson L. E. (2009). K-nearest neighbor. *Scholarpedia*, 4(2), 1883.

Progiou A. & Ziomas, I. (2012). Twenty-year road traffic emissions trend in Greece. *Water, Air, & Soil Pollution*, 223, 305-317.

Ranstam, J., & Cook, J. A. (2018). LASSO regression. *Journal of British Surgery*, 105(10), 1348-1348.

Rose G., Taylor M. A. & Tisato, P. (1989). Estimating travel time functions for urban roads: Options and issues. *Transportation Planning and Technology*, 14(1), 63-82.

Saadullah M., Zhang, Z. & Hu, H. (2022). Estimation of travel time variation caused by transport infrastructure development. *Smart and Resilient Transportation*, 4(2), 155-172.

Saedi R., Saeedmanesh M., Zockaie, A., Saberi M., Geroliminis N. &Mahmassani H. S. (2020). Estimating network travel time reliability with network partitioning. *Transportation Research Part C: Emerging Technologies*, 112, 46-61.

Servos N., Liu, X., Teucke, M. & Freitag, M. (2019). Travel time prediction in a multimodal freight transport relation using machine learning algorithms. *Logistics*, 4(1), 1.

Sindhu Meena K. & Suriya S. (2020). A survey on supervised and unsupervised learning techniques. In *Proceedings of International Conference on Artificial Intelligence, Smart Grid and Smart City Applications: AISGSC 2019* (pp. 627-644). Springer International Publishing.

Stathopoulos A. & Karlafti M. (2001). Temporal and spatial variations of real-time traffic data in urban areas. *Transportation Research Record*, 1768(1), 135-140.

Tsekeris T. & Geroliminis N. (2013). City size, network structure and traffic congestion. *Journal of Urban Economics*, 76, 1-14.

Verhoef E. T. (1999). Time, speeds, flows and densities in static models of road traffic congestion and congestion pricing. *Regional Science and Urban Economics*, 29(3), 341-369.

Verhoef E. T. (2001). An integrated dynamic model of road traffic congestion based on simple car-following theory: Exploring hyper congestion. *Journal of Urban Economics*, 49(3), 505-542.

Verhoef E. T. (2005). Speed-flow relations and cost functions for congested traffic: Theory and empirical analysis. *Transportation Research Part A: Policy and Practice*, 39(7-9), 792-812.

Wang J., Li P., Ran R., Che Y. & Zhou, Y. (2018). A short-term photovoltaic power prediction model based on the gradient boost decision tree. *Applied Sciences*, 8(5), 689.

Wang Z., Fu K., & Ye J. (2018, July). Learning to estimate the travel time. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (pp. 858-866).

Yu, G., & Zhang, C. (2004, May). Switching ARIMA model based forecasting for traffic flow. In *2004 IEEE International Conference on Acoustics, Speech, and Signal Processing* (Vol. 2, pp. ii-429). IEEE.

Yuan T., da Rocha Neto W. B., Rothenberg C., Obraczka K., Barakat C., &Turletti T. (2019). Harnessing machine learning for next-generation intelligent transportation systems: A survey. In *Proceedings of the Computational Intelligence, Communication Systems and Networks (CICSyN)*.

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