



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ ΣΧΟΛΗ

ΠΟΛΙΤΙΚΩΝ ΜΗΧΑΝΙΚΩΝ ΕΜΠ

ΤΟΜΕΑΣ ΜΕΤΑΦΟΡΩΝ ΚΑΙ ΣΥΓΚΟΙΝΩΝΙΑΚΗΣ ΥΠΟΔΟΜΗΣ

Διπλωματική εργασία

«Επιλογή μέσου μετακίνησης υπό την επιρροή ακραίων συμβάντων σε αστικά δίκτυα»

Γεωργίου Αλέξης



Επιβλέπουσα: Ελένη Ι. Βλαχογιάννη, Καθηγήτρια Σχολής Πολιτικών Μηχανικών
ΕΜΠ

Αθήνα, Μάρτιος 2024

Ευχαριστίες

Με την εκπόνηση της παρούσας διπλωματικής εργασίας ολοκληρώνεται ο κύκλος των προπτυχιακών μου σπουδών στη Σχολή Πολιτικών Μηχανικών του Εθνικού Μετσόβιου Πολυτεχνείου.

Πρωτίστως θα ήθελα να ευχαριστήσω θερμά την επιβλέπουσα Καθηγήτρια μου κα. Ελένη Βλαχογιάννη, τόσο για την ανάθεση ενός θέματος με ιδιαίτερο προσωπικό ενδιαφέρον, όσο και για την άψογη καθοδήγηση που μου παρείχε, κατά την προσπάθειά μου, καθώς και τους κ. Γιώργο Γιαννή, Καθηγητή ΕΜΠ και κ. Κωνσταντίνο Γκιτσαλίτη, Επίκουρο Καθηγητή ΕΜΠ για τη συμμετοχή τους στην τριμελή επιτροπή.

Θερμές ευχαριστίες οφείλω στον κ. Μάριο Γιουρουκέλη, Υποψήφιο Διδάκτορα ΕΜΠ, για τις χρήσιμες συμβουλές και υποδείξεις που μου παρείχε, καθώς και για το χρόνο που αφιέρωσε για την υποστήριξη μου σε όλα τα στάδια εκπόνησης της παρούσας διπλωματικής εργασίας.

Τέλος, θέλω να εκφράσω την ευγνωμοσύνη μου στην οικογένεια μου και σε όλους τους δικούς μου ανθρώπους για την αμέριστη συμπαράσταση και τη συνεχή στήριξη που μου παρείχαν καθ' όλη τη διάρκεια των σπουδών μου.

Τίτλος: «Επιλογή μέσου μετακίνησης υπό την επιρροή ακραίων συμβάντων σε αστικά δίκτυα»

Συγγραφέας διπλωματικής εργασίας: Αλέξης Γεωργίου

Επιβλέπουσα: Ελένη Ι. Βλαχογιάννη, Καθηγήτρια Σχολής Πολιτικών Μηχανικών ΕΜΠ

Σύνοψη

Η παρούσα έρευνα επικεντρώνεται στη διερεύνηση των παραγόντων που επηρεάζουν τις επιλογές μέσου μετακίνησης στην περίπτωση εμφάνισης απροσδόκητων συμβάντων. Τα απροσδόκητα γεγονότα που εξετάστηκαν περιλαμβάνουν δυσμενείς καιρικές συνθήκες, ακραία καιρικά φαινόμενα (π.χ. σεισμός, πλημμύρες), έντονη κυκλοφορική συμφόρηση, απρογραμμάτιστη διακοπή των μέσων μαζικής μεταφοράς (Μ.Μ.Μ) , προγραμματισμένες εκδηλώσεις (π.χ. οδικά έργα, ποδοσφαιρικός αγώνας) και επικίνδυνα γεγονότα (π.χ. COVID-19). Για τη συλλογή των απαραίτητων δεδομένων, διεξήχθη έρευνα με τη μέθοδο της δεδηλωμένης προτίμησης σε τέσσερις μεγάλες ευρωπαϊκές πόλεις: Αθήνα (Ελλάδα), Λισαβόνα (Πορτογαλία), Μάντσεστερ (Ηνωμένο Βασίλειο) και Ρεν (Γαλλία), με σκοπό να αναδειχθούν οι προτιμήσεις των πολιτών όσον αφορά στην πιθανότητα να αλλάξουν το σύνηθες μέσο μετακίνησης τους κατά την εμφάνιση ενός απροσδόκητου γεγονότος. Το τελικό δείγμα αποτελείται από 1859 έγκυρα ερωτηματολόγια. Τα δεδομένα που συλλέχθηκαν από κάθε πόλη χρησιμοποιήθηκαν για την ανάπτυξη και αξιολόγηση μοντέλων μηχανικής μάθησης. Μέσω της αξιολόγησης και της σύγκρισης των αποτελεσμάτων των μοντέλων, παρουσιάζεται εκτενής ανάλυση των παραγόντων που επηρεάζουν την επιλογή μέσου μετακίνησης υπό απροσδόκητες συνθήκες. Τα αποτελέσματα δείχνουν ότι οι περισσότεροι μετακινούμενοι τείνουν να αλλάξουν το μέσο μετακίνησης τους στην περίπτωση ακραίων συμβάντων.

Λέξεις – κλειδιά: Επιλογή μέσου μετακίνησης υπό την εμφάνιση απροσδόκητων συμβάντων, συμπεριφορά μετακινούμενου, ταξινομητές μηχανικής μάθησης

Title: «Transport mode choice under the influence of extreme events on urban networks»

Thesis Author: Alexis Georgiou

Supervisor: Eleni I. Vlahogianni, Professor at the National Technical University of Athens

Abstract

This study focuses on exploring the factors affecting travel mode choices in the event of unexpected events. The unexpected events that were examined included adverse weather, natural hazards (e.g. earthquake, flooding), intense traffic, unplanned PT disruption, planned events (e.g. roadworks, football match) and hazardous events (e.g. COVID-19). To collect the necessary data, a stated preferences survey was conducted in four major European cities: Athens (Greece), Lisbon (Portugal), Manchester (United Kingdom) and Rennes (France), aimed at the documentation of commuters' preferences when choosing modes of transport. The final sample consists 1859 valid questionnaires. Data collected from each city were used to develop and calibrate multiple mode choice models, including state-of-the-art machine learning models. Through evaluating and comparing the model results, we present an analysis of the factors influencing mode choice under unexpected events. The results show that most people tend to alter their usual mode of transport when faced with unexpected events.

Keywords: Mode choice under unexpected events, Travel behavior, Machine learning classifiers

Περίληψη

Η ανάπτυξη της κοινωνίας και η συνεχής εξέλιξη είναι στενά συνδεδεμένες με το δίκτυο μεταφορών. Η ομαλή λειτουργία των αστικών κέντρων και η διεξαγωγή των καθημερινών δραστηριοτήτων εντός των πόλεων απαιτούν αποτελεσματικές μετακινήσεις. Ο τρόπος και η φύση αυτών των μετακινήσεων ποικίλλουν και εξαρτώνται από πολλούς παράγοντες, όπως ο σκοπός, το κόστος και ο χρόνος. Κάποιες φορές, όμως, συμβαίνουν απρόβλεπτα γεγονότα που υπερβαίνουν την καθημερινότητα. Αυτά τα γεγονότα αλλάζουν τις συνήθειες των ανθρώπων και ως εκ τούτου έχουν σημαντικές επιπτώσεις στις μετακινήσεις τους, καθώς οι μετακινούμενοι υποχρεούνται πολλές φορές να επανεξετάσουν την επιλογή τους για το μέσο μεταφοράς που χρησιμοποιούν και να αναζητήσουν εναλλακτικές λύσεις.

Παρόλο που η αξιολόγηση της συμπεριφοράς μετακίνησης των χρηστών του δικτύου έχει αποτελέσει αντικείμενο εκτεταμένης έρευνας, διαπιστώθηκε κενό στις συναφείς έρευνες τόσο ως προς το αντικείμενο όσο και προς τη μεθοδολογική προσέγγιση που έχει επιλεγεί για την ανάλυση των παραγόντων που επηρεάζουν την αλλαγή μεταφορικού μέσου σε περίπτωση εμφάνισης ενός ασυνήθιστου γεγονότος.

Σκοπός της παρούσας διπλωματικής εργασίας είναι η διερεύνηση των παραγόντων που επηρεάζουν την επιλογή μέσου μετακίνησης στην περίπτωση εμφάνισης απροσδόκητων συμβάντων σε τέσσερις ευρωπαϊκές πόλεις (Αθήνα, Λισαβόνα, Μάντσεστερ, Ρεν). Τα απροσδόκητα συμβάντα που μελετήθηκαν ήταν έξι (6) και ήταν τα εξής: i) Δυσμενείς καιρικές συνθήκες, ii) ακραία καιρικά φαινόμενα (π.χ. σεισμός, πλημμύρες), iii) έντονη κυκλοφοριακή κίνηση, iv) απρογραμματίστη διακοπή των Μ.Μ.Μ., v) προγραμματισμένες εκδηλώσεις (π.χ. οδικά έργα, ποδοσφαιρικός αγώνας), vi) επικίνδυνα συμβάντα (π.χ. COVID-19). Για το λόγο αυτό, αναπτύσσονται πρότυπα μηχανικής μάθησης, όπως ο ταξινομητής δέντρων απόφασης (Decision Tree), ο ταξινομητής τυχαίου δάσους (Random Forest), ο ταξινομητής διανυσμάτων υποστήριξης (Support Vector Classifier), ο ταξινομητής K-Nearest Neighbors (KNN), ο ταξινομητής Gaussian Naive Bayes και ο πολυεπίπεδος νευρώνας perceptron (Multi-Layer perceptron).

Για τη συλλογή των δεδομένων χρησιμοποιήθηκαν ερωτηματολόγια, τα οποία συγκεντρώθηκαν τόσο από έρευνα πεδίου, που πραγματοποιήθηκε στις τέσσερις πόλεις της έρευνας, όσο και από ηλεκτρονική συμπλήρωση του ερωτηματολογίου, όπου συνολικά απάντησαν 1859 άνθρωποι. Το ερωτηματολόγιο αποτελούταν από τρεις ενότητες. Η πρώτη αφορούσε το προφίλ κινητικότητας και περιλάμβανε ερωτήσεις σχετικά με τις καθημερινές συνήθειες μετακίνησης των χρηστών, όπως ο πιο συχνός σκοπός και τρόπος μετακίνησης, αν ταξιδεύουν σε ώρες αιχμής ή όχι, η πιθανή ευελιξία που έχουν όσον αφορά την ώρα άφιξης στον προορισμό τους, το κόστος της διαδρομής κ.α.. Η δεύτερη ενότητα αφορούσε την αντίληψη των χρηστών σχετικά με διάφορες στρατηγικές διαχείρισης της κυκλοφορίας και νέες υπηρεσίες που αποσκοπούν στη βελτίωση των συνθηκών του δικτύου. Η τρίτη ενότητα

αφορούσε τα δημογραφικά χαρακτηριστικά και περιλάμβανε προσωπικές πληροφορίες σχετικά με το φύλο, την ηλικία, το ετήσιο ατομικό εισόδημα, το επάγγελμα, την ιδιοκτησία αυτοκινήτου και τον τόπο κατοικίας (εντός ή εκτός του κέντρου της πόλης).

Στη συνέχεια, ακολουθεί η διαδικασία επεξεργασίας των δεδομένων. Αρχικά χρησιμοποιήθηκε περιγραφική στατιστική και μέσω πινάκων και γραφημάτων παρουσιάζονται οι κατανομές των απαντήσεων και γίνεται λεπτομερής σχολιασμός αυτών. Ακολουθεί εκτενής ανάλυση των μοντέλων που χρησιμοποιήθηκαν. Η ανάλυση αυτή περιλαμβάνει περιγραφή της μεθοδολογίας που χρησιμοποιεί κάθε μοντέλο ώστε να καταλήξει στο αποτέλεσμα, περιγραφή των παραμέτρων του καθώς και τα πλεονεκτήματα και μειονεκτήματα του. Αναφέρονται ακόμα λεπτομερώς τα κριτήρια στα οποία βασίστηκε η αξιολόγηση των μοντέλων.

Εφαρμόζοντας τα παραπάνω παρουσιάζονται και ερμηνεύονται διεξοδικά τα μοντέλα που προέκυψαν. Τέλος, παρατίθενται τα βασικά συμπεράσματα που εξήχθησαν από το σύνολο της μελέτης. Αποδείχθηκε ότι οι μεταβλητές που αντιστοιχούν στο μέσο μετακίνησης (TRIP_MODE), στο εισόδημα (INCOME) και στο σκοπό της μετακίνησης (TRIP_PURPOSE) επηρεάζουν σημαντικά τα αποτελέσματα όλων των μοντέλων. Ως εκ τούτου, αποτελούν τους σημαντικότερους παράγοντες που επηρεάζουν την αλλαγή του συνήθους μεταφορικού μέσου σε περίπτωση εμφάνισης ενός απροσδόκητου συμβάντος.

Πίνακας περιεχομένων

1	Εισαγωγή.....	12
1.1	Γενικά στοιχεία.....	12
1.2	Σκοπός της διπλωματικής εργασίας	13
1.3	Μεθοδολογία της διπλωματικής εργασίας	13
1.4	Δομή της διπλωματικής εργασίας	14
2	Βιβλιογραφική Ανασκόπηση.....	15
2.1	Εισαγωγή.....	15
2.2	Συναφείς Έρευνες	15
3	Μεθοδολογική Προσέγγιση	22
3.1	Εισαγωγή.....	22
3.2	Δομή ερωτηματολογίου	23
3.3	Συλλογή δεδομένων.....	25
3.4	Προεπεξεργασία δεδομένων	26
3.5	Περιγραφική στατιστική	29
3.6	Ανάπτυξη μοντέλων	36
3.6.1	Αλγόριθμοι ταξινόμησης.....	36
3.6.2	Δέντρα απόφασης.....	37
3.6.3	Τυχαία Δάση	41
3.6.4	Μηχανές διανυσμάτων στήριξης.....	44
3.6.5	k κοντινότεροι γείτονες.....	49
3.6.6	Gaussian Naive Bayes Classifier	51
3.6.7	Multi-Layer perceptron (MLP)	53
3.7	Αξιολόγηση μοντέλων	56
4	Αποτελέσματα	58
4.1	Εισαγωγή.....	58
4.2	Πρότυπα επιλογής μέσου μετακίνησης για την Αθήνα.....	58

4.3	Πρότυπα επιλογής μέσου μετακίνησης για τη Λισαβόνα.....	62
4.4	Πρότυπα επιλογής μέσου μετακίνησης για το Μάντσεστερ.....	64
4.5	Πρότυπα επιλογής μέσου μετακίνησης για τη Ρεν	67
4.6	Πρότυπα επιλογής μέσου μετακίνησης για όλες τις πόλεις μαζί.....	71
4.7	Συγκριτική αξιολόγηση μοντέλων.....	76
5	Συμπεράσματα	79
5.1	Γενικά	79
5.2	Βασικά συμπεράσματα.....	79
5.3	Προτάσεις για περαιτέρω έρευνα	81
6	Βιβλιογραφία	82
	Παράρτημα	85

Ευρετήριο διαγραμμάτων

Διάγραμμα 3.1: Κατανομή των απαντήσεων της Αθήνας στο εξής ερώτημα: «Ποια η πιθανότητα να αλλάξει κάποιος μεταφορικό μέσο σε περίπτωση έντονης κυκλοφοριακής συμφόρησης»	27
Διάγραμμα 3.2: 1 ^η περίπτωση ομαδοποίησης της μεταβλητής CHA_INTTRA.....	28
Διάγραμμα 3.3: 2 ^η περίπτωση ομαδοποίησης της μεταβλητής CHA_INTTRA.....	28
Διάγραμμα 3.4: : 3 ^η περίπτωση ομαδοποίησης της μεταβλητής CHA_INTTRA.....	29
Διάγραμμα 3.5: Κατανομή συμμετεχόντων σε Αθήνα, Λισαβόνα, Μάντσεστερ, Ρεν	30
Διάγραμμα 3.6: Κατανομή συμμετεχόντων ανά ηλικία για Αθήνα.....	31
Διάγραμμα 3.7: Κατανομή συμμετεχόντων ανά ηλικία για Λισαβόνα	31
Διάγραμμα 3.8: Κατανομή συμμετεχόντων ανά ηλικία για Μάντσεστερ.....	31
Διάγραμμα 3.9: Κατανομή συμμετεχόντων ανά ηλικία για Ρεν	31
Διάγραμμα 3.10: Κατανομή των συμμετεχόντων με βάση το μέσο μετακίνησης που επιλέγουν συχνότερα για τις καθημερινές τους μετακινήσεις για Αθήνα.....	32
Διάγραμμα 3.11: Κατανομή των συμμετεχόντων με βάση το μέσο μετακίνησης που επιλέγουν συχνότερα για τις καθημερινές τους μετακινήσεις για Λισαβόνα	32
Διάγραμμα 3.12: Κατανομή των συμμετεχόντων με βάση το μέσο μετακίνησης που επιλέγουν συχνότερα για τις καθημερινές τους μετακινήσεις για Μάντσεστερ.....	33
Διάγραμμα 3.13: Κατανομή των συμμετεχόντων με βάση το μέσο μετακίνησης που επιλέγουν συχνότερα για τις καθημερινές τους μετακινήσεις για Ρεν	33
Διάγραμμα 3.14: Κατανομή των συμμετεχόντων με βάση τον κύριο σκοπό των καθημερινών τους μετακινήσεων για Αθήνα.....	34
Διάγραμμα 3.15: Κατανομή των συμμετεχόντων με βάση τον κύριο σκοπό των καθημερινών τους μετακινήσεων για Λισαβόνα	34
Διάγραμμα 3.16: Κατανομή των συμμετεχόντων με βάση τον κύριο σκοπό των καθημερινών τους μετακινήσεων για Μάντσεστερ.....	35
Διάγραμμα 3.17: Κατανομή των συμμετεχόντων με βάση το μέσο μετακίνησης που επιλέγουν συχνότερα για τις καθημερινές τους μετακινήσεις για Ρεν	35
Διάγραμμα 4.1: Κατανομή δείγματος για Αθήνα	59
Διάγραμμα 4.2: Κατανομή δείγματος για Λισαβόνα.....	62
Διάγραμμα 4.3: Κατανομή δείγματος για Μάντσεστερ	65
Διάγραμμα 4.4: Κατανομή του δείγματος για τη Ρεν.....	68

Διάγραμμα 4.5: Κατανομή δείγματος για όλες τις πόλεις μαζί.....	73
-----------------------------------------------------------------	----

Ευρετήριο πινάκων

Πίνακας 3.1: Κατανομή των απαντήσεων της Αθήνας στο εξής ερώτημα: : «Ποια η πιθανότητα να αλλάξει κάποιος μεταφορικό μέσο σε περίπτωση έντονης κυκλοφοριακής συμφόρησης;»	27
Πίνακας 3.2: Σύγκριση Random Forest και Decision Tree	44
Πίνακας 3.3: Πίνακας κατηγοριοποίησης δύο κλάσεων.....	56
Πίνακας 4.1: Κωδικοποίηση μεταβλητών του Μοντέλου της Αθήνας	60
Πίνακας 4.2: Ανεξάρτητες μεταβλητές του Μοντέλου της Αθήνας.....	61
Πίνακας 4.3: Μετρικές αξιολόγησης του Μοντέλου 1 της Αθήνας	61
Πίνακας 4.4: Κωδικοποίηση μεταβλητών των Μοντέλων της Λισαβόνας	63
Πίνακας 4.5: Ανεξάρτητες μεταβλητές των Μοντέλων της Λισαβόνας.....	63
Πίνακας 4.6: Μετρικές αξιολόγησης του Μοντέλου 2 της Λισαβόνας	64
Πίνακας 4.7: Μετρικές αξιολόγησης του Μοντέλου 3 της Λισαβόνας	64
Πίνακας 4.8: Κωδικοποίηση μεταβλητών του Μοντέλου του Μάντσεστερ	66
Πίνακας 4.9: Ανεξάρτητες μεταβλητές του Μοντέλου του Μάντσεστερ	66
Πίνακας 4.10: Μετρικές αξιολόγησης του Μοντέλου 4 του Μάντσεστερ	67
Πίνακας 4.11: Κωδικοποίηση μεταβλητών των Μοντέλων της Ρεν	68
Πίνακας 4.12: Ανεξάρτητες μεταβλητές του Μοντέλου της Ρεν	69
Πίνακας 4.13: Μετρικές αξιολόγησης του Μοντέλου 5 της Ρεν	70
Πίνακας 4.14: Μετρικές αξιολόγησης του Μοντέλου 6 της Ρεν	70
Πίνακας 4.15: Μετρικές αξιολόγησης του Μοντέλου 7 της Ρεν	71
Πίνακας 4.16: Κωδικοποίηση μεταβλητών των Μοντέλων για όλες τις πόλεις μαζί.....	72
Πίνακας 4.17: Ανεξάρτητες μεταβλητές των Μοντέλων για όλες τις πόλεις μαζί	73
Πίνακας 4.18: Μετρικές αξιολόγησης του Μοντέλου 8 για όλες τις πόλεις μαζί	74
Πίνακας 4.19: Μετρικές αξιολόγησης του Μοντέλου 9 για όλες τις πόλεις μαζί	75
Πίνακας 4.20: Μετρικές αξιολόγησης του Μοντέλου 10 για όλες τις πόλεις μαζί	75
Πίνακας 4.21: Συγκεντρωτικός πίνακας μοντέλων ανά πρόβλημα που επιλύθηκε σε κάθε πόλη	77
Πίνακας 4.22: Συγκεντρωτικός πίνακας μοντέλων σε σχέση με τις ανεξάρτητες μεταβλητές που χρησιμοποιήθηκαν	78

Ευρετήριο εικόνων

Εικόνα 3.1: Διεπαφή συνομιλίας του CONEY	25
Εικόνα 3.2: Δομή των δέντρων απόφασης.....	38
Εικόνα 3.3: Δομή του Random Forest.....	42
Εικόνα 3.4: Παράδειγμα SVC δύο κλάσεων	45
Εικόνα 3.5: Δύο μη μηδενικά διανύσματα A και B	46
Εικόνα 3.6: Παράδειγμα SVM ταξινόμησης δύο κλάσεων	47
Εικόνα 3.7: Επίπεδα L , L_1 , L_2 και η μεταξύ τους απόσταση στο SVC	48
Εικόνα 3.8: Γραφική απεικόνιση παραδείγματος KNN με $K=3$ και $K=5$	50
Εικόνα 3.9: Διαδικασία ενεργοποίησης ενός τεχνητού νευρώνα	54
Εικόνα 3.10: Δομή του Multi-layer Perceptron.....	55

1 Εισαγωγή

1.1 Γενικά στοιχεία

Η συνεχής εξέλιξη καθώς και η ανάπτυξη της κοινωνίας είναι άρρηκτα συνδεδεμένη με το δίκτυο μεταφορών. Για την πραγματοποίηση των δραστηριοτήτων των ανθρώπων εντός των πόλεων και την ορθή λειτουργία των αστικών κέντρων απαραίτητη προϋπόθεση αποτελούν οι ανεμπόδιστες αστικές μετακινήσεις, οι οποίες ποικίλουν ανάλογα με το σκοπό, το χρόνο, το κόστος και άλλους παράγοντες. Ο σχεδιασμός των μεταφορών εξελίχθηκε από μία τεχνική σε έναν εξειδικευμένο επιστημονικό κλάδο που απαιτεί, εκτός από τεχνική κατάρτιση, γνώση οικονομικών, ανθρωπιστικών και περιβαλλοντικών επιστημών. Η αύξηση της κινητικότητας στις πόλεις σε συνδυασμό με τη δομή τους καθιστούν προβληματικούς τους συνηθέστερους τρόπους μετακίνησης όπως τα Ι.Χ. και κάνουν εμφανέστερη την ανάγκη εναλλακτικών τρόπων μετακίνησης.

Οι εναλλακτικές μέθοδοι μετακίνησης αποτελούν σημαντική πτυχή στη βελτίωση της κινητικότητας στις αστικές περιοχές. Αντιπροσωπευτικά παραδείγματα αυτών των μεθόδων περιλαμβάνουν την ποδηλασία, το περπάτημα, τη χρήση των δημόσιων μέσων μαζικής μεταφοράς, την κοινή χρήση μέσων μεταφοράς (carpooling), καθώς και τη χρήση ηλεκτρικών οχημάτων και ποδηλάτων. Αυτές οι μέθοδοι προσφέρουν οφέλη όχι μόνο στο περιβάλλον λόγω της μείωσης των εκπομπών και της μείωσης της κυκλοφοριακής συμφόρησης, αλλά επίσης βελτιώνουν την υγεία των ανθρώπων και δημιουργούν πιο βιώσιμες και ενεργές κοινότητες. Η προώθηση και υποστήριξη αυτών των εναλλακτικών μεθόδων μετακίνησης αποτελεί κρίσιμο στοιχείο για τη διασφάλιση ενός βιώσιμου, αποδοτικού και υγιεινού συστήματος μεταφορών στις αστικές περιοχές.

Παρόλα αυτά, τα ιδιωτικά αυτοκίνητα αποτελούν ακόμα έναν από τους πιο δημοφιλείς τρόπους μετακίνησης λόγω των πολλών πλεονεκτημάτων που προσφέρουν, όπως η άνεση και η ευκολία που παρέχουν στους χρήστες τους. Οι επιβάτες μπορούν να απολαύσουν προσωπικό χώρο και άνετη μεταφορά, ενώ ταυτόχρονα έχουν την ελευθερία να κινούνται όποτε το επιθυμούν. Επιπλέον, τα αυτοκίνητα προσφέρουν στους μετακινούμενους άμεση πρόσβαση στον προορισμό τους χωρίς την ανάγκη για περπάτημα ή αναμονή για δημόσια μέσα μεταφοράς.

Ορισμένες ημέρες, όμως, συμβαίνουν απρόβλεπτα γεγονότα που υπερβαίνουν την καθημερινότητα. Αυτά τα γεγονότα αλλάζουν τις συνήθειες των ανθρώπων και ως εκ τούτου έχουν σημαντικές επιπτώσεις στις μετακινήσεις, υποχρεώνοντας πολλές φορές τους μετακινούμενους να επανεξετάσουν την επιλογή τους για το μέσο μεταφοράς που χρησιμοποιούν και να αναζητήσουν εναλλακτικές λύσεις.

1.2 Σκοπός της διπλωματικής εργασίας

Αντικείμενο της παρούσας διπλωματικής εργασίας είναι η διερεύνηση της αξιολόγησης των παραγόντων που επηρεάζουν την επιλογή μέσου μετακίνησης στην περίπτωση εμφάνισης απροσδόκητων συμβάντων σε τέσσερις ευρωπαϊκές πόλεις (Αθήνα, Μάντσεστερ, Ρεν, Λισαβόνα). Τα απροσδόκητα συμβάντα που μελετήθηκαν ήταν έξι (6) και αφορούσαν σε δυσμενείς καιρικές συνθήκες, ακραία καιρικά φαινόμενα (π.χ. σεισμός, πλημμύρες), έντονη κυκλοφοριακή κίνηση, απρογραμμάτιστη διακοπή των Μ.Μ.Μ., προγραμματισμένες εκδηλώσεις (π.χ. οδικά έργα, ποδοσφαιρικός αγώνας), και άλλα επικίνδυνα συμβάντα (π.χ. COVID-19).

Τα αποτελέσματα της ανάλυσης και τα συμπεράσματα της εργασίας μπορούν να παράσχουν χρήσιμες πληροφορίες, εφόσον είναι δυνατό να αξιοποιηθούν τόσο ατομικά από τους μετακινούμενους για τη βελτίωση της ταξιδιωτικής τους συμπεριφοράς, όσο και από τους αρμόδιους φορείς για τη βελτίωση της παροχής υπηρεσιών προς τους πελάτες τους.

1.3 Μεθοδολογία της διπλωματικής εργασίας

Αρχικά, μετά την οριστικοποίηση του επιδιωκόμενου στόχου πραγματοποιείται βιβλιογραφική ανασκόπηση τόσο σε ελληνικό όσο και διεθνές επίπεδο. Στη φάση αυτή πραγματοποιείται αναζήτηση παρεμφερών ερευνών, επιστημονικών άρθρων καθώς και γενικών πληροφοριών σχετικά με το εξεταζόμενο αντικείμενο που μπορούν να φανούν χρήσιμες για τη συγκεκριμένη έρευνα. Μέσω των ερευνών αυτών καταβάλλεται προσπάθεια να αποκτηθεί μια σχετική εμπειρία στην επεξεργασία τέτοιων θεμάτων, καθώς επίσης και να αποφασιστεί η μέθοδος με βάση την οποία θα πραγματοποιηθεί η επεξεργασία των στοιχείων για να επιτευχθεί ο επιδιωκόμενος στόχος.

Μετά τη διερεύνηση των στοιχείων που συγκεντρώθηκαν από τη βιβλιογραφική ανασκόπηση, σειρά έχει η εύρεση του τρόπου συλλογής στοιχείων. Για τη συλλογή των δεδομένων εφαρμόστηκε η μεθοδολογία της δεδηλωμένης προτίμησης (stated preference) χρησιμοποιώντας ερωτηματολόγια. Χίλιοι οκτακόσιοι πενήντα εννέα (1859) μετακινούμενοι συμμετείχαν στο ερωτηματολόγιο που διεξήχθη σε Αθήνα, Λισαβόνα, Μάντσεστερ και Ρεν.

Στη συνέχεια, ακολουθεί η επιλογή των μοντέλων μηχανικής μάθησης που χρησιμοποιήθηκαν για την εξαγωγή αποτελεσμάτων. Ακολουθούν τα αποτελέσματα και η διεξοδική ερμηνεία τους. Τέλος, προκύπτουν τα αντίστοιχα συμπεράσματα για το ποιοι παράγοντες επηρεάζουν την αλλαγή μεταφορικού μέσου στην περίπτωση εμφάνισης κάποιου απροσδόκητου συμβάντος.

1.4 Δομή της διπλωματικής εργασίας

Το **πρώτο κεφάλαιο** είναι εισαγωγικό και αποτελεί τη βάση για την κατανόηση της παρούσας εργασίας. Περιγράφεται επίσης, ο απώτερος σκοπός της μελέτης και γίνεται αναφορά στη μεθοδολογία που ακολουθείται για την εξαγωγή των τελικών συμπερασμάτων. Τέλος, το κεφάλαιο ολοκληρώνεται με την αναφορά στη δομή της εργασίας

Στο **δεύτερο κεφάλαιο** περιλαμβάνεται η βιβλιογραφική ανασκόπηση, όπου παρατίθενται μελέτες συναφούς αντικείμενου και μεθοδολογίας, κατόπιν αναζήτησης τους σε διεθνές επίπεδο. Αναλυτικότερα, η ανασκόπηση περιλαμβάνει μελέτες που αφορούν την καταγραφή της ταξιδιωτικής συμπεριφοράς υπό απροσδόκητα συμβάντα, καθώς και τα μοντέλα που χρησιμοποιήθηκαν για τα συμπεράσματα που προέκυψαν.

Στο **τρίτο κεφάλαιο** παρουσιάζεται η διαδικασία συλλογής και επεξεργασίας των απαραίτητων δεδομένων. Γίνεται αναλυτική περιγραφή των ενοτήτων, στις οποίες είναι χωρισμένο το ερωτηματολόγιο και ακολουθεί στατιστική περιγραφή των απαντήσεων. Μέσω πινάκων και γραφημάτων παρουσιάζονται ενδιαφέροντα μοτίβα που εντοπίζονται στις απαντήσεις και γίνεται λεπτομερής σχολιασμός αυτών. Ακολουθεί εκτενής ανάλυση των μοντέλων που χρησιμοποιήθηκαν. Η ανάλυση αυτή περιλαμβάνει περιγραφή της μεθοδολογίας που χρησιμοποιεί κάθε μοντέλο ώστε να καταλήξει στο αποτέλεσμα, περιγραφή των παραμέτρων του καθώς και τα πλεονεκτήματα και μειονεκτήματα του. Τέλος, αναφέρονται λεπτομερώς τα κριτήρια στα οποία βασίστηκε η αξιολόγηση των μοντέλων.

Στο **τέταρτο κεφάλαιο** περιγράφονται τα βήματα που ακολουθήθηκαν κατά την εφαρμογή της μεθοδολογίας και αναλύεται η διαδικασία ανάπτυξης και αξιολόγησης των τελικών μοντέλων. Παρουσιάζονται τα δεδομένα εισόδου και εξόδου, με ιδιαίτερη έμφαση στα αποτελέσματα και την ερμηνεία τους.

Στο **πέμπτο κεφάλαιο** γίνεται παράθεση των κυριότερων συμπερασμάτων που προέκυψαν ύστερα από την ανάλυση και αξιολόγηση των αποτελεσμάτων των μοντέλων. Τέλος, αναφέρονται οι δυνατότητες για περαιτέρω έρευνα στο συγκεκριμένο αντικείμενο.

2 Βιβλιογραφική Ανασκόπηση

2.1 Εισαγωγή

Η ανθρώπινη συμπεριφορά χαρακτηρίζεται από επαναλαμβανόμενα μοτίβα και τακτικότητα, με τις επιλογές μέσου μετακίνησης σε ένα αστικό περιβάλλον να μην αποτελούν εξαίρεση (Sun et al., 2021). Αυτή η συνέπεια στις μετακινήσεις μπορεί να διαταραχθεί από αναπάντεχα, ή εκτενή γεγονότα που επηρεάζουν την επιλογή μέσου μετακίνησης. Το παρόν κεφάλαιο περιέχει έρευνες που εστιάζουν στην αλλαγή της συμπεριφοράς των μετακινούμενων κατά τη διάρκεια τέτοιων συμβάντων όπως είναι οι δυσμενείς καιρικές συνθήκες, τα ακραία καιρικά φαινόμενα, η έντονη κυκλοφοριακή κίνηση, οι προγραμματισμένες εκδηλώσεις (π.χ. οδικά έργα, ποδοσφαιρικός αγώνας), τα επικίνδυνα γεγονότα (π.χ. COVID-19) και η απρογραμμάτιστη διακοπή των Μ.Μ.Μ. Παρατίθενται μελέτες που στοχεύουν στην ανάλυση τόσο των λόγων για τους οποίους η εμφάνιση τέτοιων γεγονότων επιδρά σημαντικά στην αλλαγή του μέσου μεταφοράς των ανθρώπων όσο και των μεθόδων και των μοντέλων που χρησιμοποιήθηκαν για να εξαχθούν τα αποτελέσματα αυτά. Η εξέταση των παραπάνω αποτελεί σημαντικό πεδίο έρευνας που έχει ευρείες εφαρμογές στον σχεδιασμό των αστικών μεταφορικών συστημάτων και στη βελτίωση της ποιότητας ζωής των πολιτών.

2.2 Συναφείς Έρευνες

Επικίνδυνα γεγονότα, τα οποία τυχαίνει να συμβαίνουν απροσδόκητα και υπό συγκεκριμένες συνθήκες, προκαλούν αλλαγές στην επιλογή μέσου μετακίνησης των ανθρώπων. Ένα τέτοιο γεγονός είναι η πανδημία του **COVID- 19**, κατά τη διάρκεια της οποίας η ταξιδιωτική συμπεριφορά μεταβλήθηκε πλήρως (Bhaduri et al., 2020- De Vos, 2020- Scorrano et al., 2021), όχι μόνο επειδή οι πολίτες κατακλύστηκαν από φόβο με αποτέλεσμα να περιορίσουν τις μετακινήσεις τους, αλλά και επειδή οι κυβερνήσεις έλαβαν αυστηρά μέτρα περιορισμού των μετακινήσεων. Η χρήση των Μ.Μ.Μ. μειώνεται, καθώς οι άνθρωποι αναγνωρίζουν τον αυξημένο κίνδυνο μετάδοσης, και τα μέτρα που καθιστούν υποχρεωτικές τις προστατευτικές χειρουργικές μάσκες μέσα στο Μ.Μ.Μ. το έχουν καταστήσει λιγότερο βολικό. Το μερίδιο της χρήσης τους αντικαθίσταται από ιδιωτικά αυτοκίνητα, ταξί και υπηρεσίες κοινής χρήσης (Shakibaei et al., 2020) . Ταυτόχρονα, οι άνθρωποι που έχουν την ευελιξία να αλλάξουν τρόπο μετακίνησης προτιμούν να ταξιδεύουν σε ώρες εκτός αιχμής. Σε γενικές γραμμές, η ζήτηση μετακινήσεων μειώνεται, δεδομένου ότι η τηλεργασία και η εξ αποστάσεως βοήθεια εφαρμόστηκαν για τα περισσότερα επαγγέλματα, ενώ οι άνθρωποι αλλάζουν τις καθημερινές τους συνήθειες και περνούν περισσότερο χρόνο περπατώντας και κάνοντας ποδήλατο. Ταυτόχρονα, η πανδημία δεν επηρεάζει όλες τις κατηγορίες πολιτών με τον ίδιο τρόπο.

Οι Sun et al., το 2022 προχώρησαν σε μια έρευνα στην Ταϊβάν που εξέτασε τις διαφορές στην ικανοποίηση από τις μετακινήσεις με λεωφορείο μεταξύ ατόμων με χαμηλό εισόδημα πριν και μετά την πανδημία του COVID-19. Η έρευνα επικεντρώθηκε σε άνδρες με χαμηλό εισόδημα, όπως και σε αγρότες, οι οποίοι εκπροσωπούνται σε ποσοστό 62,6% και 63,4% αντίστοιχα. Τα αποτελέσματά της έδειξαν ότι οι εργαζόμενοι και οι φοιτητές είναι λιγότερο ικανοποιημένοι από τα λεωφορεία σε σχέση με όσους μετακινούνται για λόγους αναψυχής (π.χ. συνταξιούχοι). Η μετακίνηση ή όχι με λεωφορείο πριν τον COVID-19 καθορίζονταν κυρίως από τα χαρακτηριστικά της μετακίνησης (απόσταση, χρόνος, κόστος και σκοπός μετακίνησης) ενώ μετά από τα κοινωνικοοικονομικά χαρακτηριστικά (φύλο, ηλικία, εάν το άτομο ανήκει σε ευπαθή ομάδα, επάγγελμα, εισόδημα, ιδιοκτησία Ι.Χ. κ.α.). Η διαφορά αυτή οφείλεται στο γεγονός ότι, λόγω της πανδημίας, ο παράγοντας της υγείας έχει γίνει κρίσιμος στην επιλογή του μέσου μεταφοράς. Οι άνθρωποι έχουν αρχίσει να λαμβάνουν υπόψη τους τον κίνδυνο μετάδοσης οποιοδήποτε ιού και τις συνέπειες που μπορεί να έχει αυτός ο κίνδυνος για την υγεία τους. Ως εκ τούτου, η προτίμηση μέσων μεταφοράς που προάγουν την ασφάλεια και την υγιεινή, όπως το περπάτημα, το ποδήλατο ή το ιδιωτικό όχημα, έχει αυξηθεί σημαντικά. Πριν την πανδημία οι τρεις (3) βασικοί λόγοι για τους οποίους ήταν δυσαρεστημένοι οι χρήστες των λεωφορείων ήταν οι συνθήκες οδικής κυκλοφορίας (20%), ο χρόνος αναμονής (17%), και η εξωτερική θερμοκρασία (14%) ενώ μετά έγιναν η κοινωνική απόσταση (17%), οι πιθανές προειδοποιήσεις κινδύνου εκ μέρους των Αρχών (16%), και ο χρόνος άφιξης (14%). Παρόλα αυτά οι δύο κύριοι λόγοι για τους οποίους αισθάνονται ικανοποιημένοι οι χρήστες παρέμειναν οι ίδιοι και είναι η εύκολη και οικονομική μετακίνηση, μόνο που μετά το COVID-19 αυξήθηκαν κατά λίγο τα ποσοστά.

Ένα δεύτερο γεγονός που μπορεί να επηρεάσει την επιλογή μεταφορικού μέσου είναι οι **διαταραχές στη λειτουργία των Μ.Μ.Μ.**, οι οποίες προκαλούν δυσαρέσκεια στους επιβάτες και οδηγούν επίσης σε αλλαγές στην επιλογή μεταφορικού μέσου (Rahimi et al., 2019). Όταν συμβαίνουν απροσδόκητες διαταραχές στη μεταβίβαση από το ένα μέσο στο άλλο, ένα σημαντικό ποσοστό των επιβατών που επηρεάζονται από αυτή την κατάσταση επιλέγει να αλλάξει τη διαδρομή και να χρησιμοποιήσει είτε ταξί είτε κάποιο άλλο Μ.Μ.Μ. (Drabicki et al. 2021). Επιπλέον, οι επιβάτες των μέσων μεταφοράς είναι πιο πιθανό να χρησιμοποιήσουν κάποια άλλη προτεινόμενη διαδρομή με λεωφορείο παρά να περιμένουν την αποκατάσταση της υπηρεσίας, ακόμη και όταν δεν υπάρχει διαφορά στον συνολικό χρόνο ταξιδιού (Teng & Liu, 2015). Οι Nguyen-Rhuoc et al., το 2018 ανέπτυξαν μια ποιοτική μελέτη που διερευνά τους παράγοντες που επηρεάζουν την αλλαγή τρόπου μετακίνησης των χρηστών των Μ.Μ.Μ. όταν παύει η λειτουργία τους βραχυπρόθεσμα ή μακροπρόθεσμα. Τριάντα χρήστες Μ.Μ.Μ. στη Μεμβούρνη συμμετείχαν σε semi-structured Interviews, συνεντεύξεις δηλαδή που συνδυάζουν ερωτήσεις κλειστού και ανοιχτού τύπου. Στην παρούσα έρευνα χρησιμοποιήθηκε η μέθοδος της Discourse analysis, μια μεθοδολογία ερευνητικής διερεύνησης που εστιάζει στην εξέταση του τρόπου με τον οποίο ο λόγος δομείται, παράγεται και ερμηνεύεται σε διάφορα κοινωνικά περιβάλλοντα.

Τα αποτελέσματά της έδειξαν ότι στην περίπτωση διακοπής της λειτουργίας των Μ.Μ.Μ. αναμένεται να αυξηθεί σημαντικά η κυκλοφοριακή κίνηση γεγονός που μπορεί να επηρεάσει

την επιλογή των χρηστών για να μετακινηθούν με αυτοκίνητο. Όταν η χρήση των Μ.Μ.Μ. σταματήσει βραχυπρόθεσμα, οι επιβάτες θα στραφούν σε εναλλακτικούς τρόπους μετακίνησης, όπως το αυτοκίνητο (οδηγώντας ή ως επιβάτης), το ποδήλατο, το περπάτημα ή την ακύρωση του ταξιδιού. Το μοντέλο υποδεικνύει ότι υπάρχει ένας συνδυασμός παραγόντων που επηρεάζουν τη στροφή προς το αυτοκίνητο σε περίπτωση βραχυπρόθεσμης διακοπής των Μ.Μ.Μ., οι οποίοι μπορούν να ταξινομηθούν σε τρεις κατηγορίες: παράγοντες που αφορούν τον χρήστη (π.χ. ιδιοκτησία αυτοκινήτου, κατοχή άδειας οδήγησης, αριθμός διαθέσιμων αυτοκινήτων στο νοικοκυριό, αριθμός ενηλίκων στο νοικοκυριό, εισόδημα), παράγοντες που αφορούν το πλαίσιο και τις συνθήκες κάτω από τις οποίες γίνεται η μετακίνηση (π.χ. απόσταση, χρόνος και κόστος μετακίνησης, προορισμός, καιρός ή ευελιξία) και παράγοντες που αφορούν τη διαδρομή (π.χ. προσβασιμότητα σε σταθμούς Μ.Μ.Μ., σκοπός διαδρομής). Οι χρήστες Μ.Μ.Μ. που έχουν πρόσβαση στους σταθμούς με αυτοκίνητο είναι πιθανό να στραφούν στη μετακίνηση με αυτοκίνητο όταν αυτά σταματήσουν, δεδομένου ότι είχαν ήδη χρησιμοποιήσει αυτοκίνητο για ένα μέρος της διαδρομής τους. Η αλληλεπίδραση μεταξύ της καλής πρόσβασης σε στάσεις λεωφορείων ή σταθμούς μετρό, του χρόνου μετακίνησης, του κόστους μετακίνησης, της σπουδαιότητας της μετακίνησης, των μη-CBD (Central Business District) ταξιδιών (δηλαδή μετακινήσεων που δεν έχουν ως προορισμό το επιχειρηματικό κέντρο μιας πόλης,), των καιρικών συνθηκών, της ευελιξίας και της προσβασιμότητας στους σταθμούς είναι οι σημαντικότεροι παράγοντες υπέρ της επιλογής αυτοκινήτου. Αντίθετα, η επιλογή μη μηχανοκίνητων μέσων μεταφοράς (ποδήλατο και περπάτημα) επηρεάζεται από διάφορους βασικούς παράγοντες, όπως η περιορισμένη πρόσβαση σε Ι.Χ., ο χρόνος μετακίνησης και το κόστος μετακίνησης. Η περιορισμένη πρόσβαση, ο χρόνος ταξιδιού, το κόστος ταξιδιού, η σημασία της μετακίνησης, η ανελαστικότητα και η ασφάλεια είναι οι σημαντικότεροι παράγοντες που επηρεάζουν την απόφαση ακύρωσης του ταξιδιού. Καθ' όλη τη διάρκεια των συνεντεύξεων, ήταν σαφές ότι η οδήγηση αυτοκινήτου προσφέρει πολλά πλεονεκτήματα στους χρήστες Μ.Μ.Μ. Το κύριο εμπόδιο για την οδήγηση που εντοπίστηκε από την παρούσα μελέτη είναι το κόστος μετακίνησης. Στην παρούσα μελέτη πολλοί συμμετέχοντες θα επέλεγαν να ακυρώσουν τα ταξίδια τους εάν τα Μ.Μ.Μ. σταματούσαν βραχυπρόθεσμα. Ο κύριος λόγος για τη μη πραγματοποίηση των ταξιδιών είναι ότι δεν θεωρήθηκαν πολύ σημαντικά. Άλλοι συμμετέχοντες θεώρησαν ότι θα μπορούσαν να εργάζονται από το σπίτι τους, ώστε να μην χρειάζεται να πηγαίνουν στο χώρο εργασίας τους.

Οι άνθρωποι παρουσιάζουν μια αλλαγή στην ταξιδιωτική συμπεριφορά και στο πλαίσιο ακραίων καιρικών φαινομένων. Κατά τη διάρκεια αυτών των περιπτώσεων η ταξιδιωτική συμπεριφορά μεταβάλλεται, όχι μόνο επειδή το δομημένο περιβάλλον και οι υποδομές μπορεί να υποστούν ζημιές, αλλά και επειδή οι άνθρωποι συνήθως πανικοβάλλονται στην προσπάθειά τους να παραμείνουν ασφαλείς. Πολλές μελέτες έχουν εξετάσει την προσαρμογή της ταξιδιωτικής συμπεριφοράς σε **δυσμενείς καιρικές συνθήκες** και **ακραία φυσικά φαινόμενα** (Lu et al., 2014, Zheng et al., 2015). Χαρακτηριστικό παράδειγμα φυσικού φαινομένου που επηρεάζει την ταξιδιωτική συμπεριφορά είναι οι πλημμύρες. Οι άνθρωποι που ταξιδεύουν για επαγγελματικούς λόγους έχουν μεγαλύτερη τάση να αλλάζουν τα σχέδιά

τους κατά τη διάρκεια μιας διακοπής λόγω ακραίων καιρικών συνθηκών, σε σύγκριση με αυτούς που ταξιδεύουν για προσωπικούς λόγους. Επιπλέον, κατά τη διάρκεια ακραίων φυσικών φαινομένων που προκαλούν σημαντικές διαταραχές στο δίκτυο μεταφορών, όπως μεγαλύτερες καθυστερήσεις ή δυσκολίες πρόσβασης σε σημεία προορισμού, οι ταξιδιώτες είναι πιθανό να αλλάζουν τα σχέδιά τους, όπως αναφέρεται από το προσωπικό των μέσων μεταφοράς (Lu et al., 2014). Αυτό μπορεί να περιλαμβάνει την αναχώρηση σε διαφορετική ώρα την ίδια ημέρα ή ακόμα και την αναβολή του ταξιδιού για την επόμενη ημέρα. Επιπλέον, η αλλαγή της διαδρομής του ταξιδιού μπορεί να είναι απαραίτητη για την ασφάλεια κατά τη διάρκεια ακραίων καιρικών φαινομένων. Τέλος, η επικοινωνία με την οικογένεια και τους φίλους, σε συνδυασμό με την πρόσβαση στο διαδίκτυο και το προσωπικό των μέσων μαζικής μεταφοράς, φαίνεται να επηρεάζει τις αποφάσεις των ταξιδιωτών κατά την αλλαγή σχεδίων κάτω από τέτοιες συνθήκες (Zappi & Ryley, 2015). Οι κάτοικοι παράκτιων περιοχών αντιδρούν διαφορετικά από αυτούς των εσωτερικών περιοχών κατά τις πλημμύρες και τα ακραία καιρικά φαινόμενα παρόλα αυτά οι περισσότεροι άνθρωποι τροποποιούν τα ταξίδια τους κατά την εμφάνιση πλημμυρών. Καθοριστικοί παράγοντες είναι η διατάραξη του οδικού δικτύου, η απομόνωση από τα πλημμυρικά ύδατα και η συχνότητα των πλημμυρών, οι οποίες γίνονται αντιληπτές με διαφορετικό τρόπο σε κάθε περιοχή. Οι παράκτιοι κάτοικοι έχουν περισσότερη εμπειρία πλημμυρών, αλλά οι περιορισμένες εναλλακτικές διαδρομές οδηγούν σε ακυρώσεις ταξιδιών. Συνολικά, κατά τις πλημμύρες, οι άνθρωποι προτιμούν να ακυρώνουν ή να αλλάζουν ταξίδια τους λόγω αρνητικής εμπειρίας μετακίνησης. Μια δεύτερη κατηγορία δυνητικά επικίνδυνου φυσικού φαινομένου που επηρεάζει τις επιλογές μετακίνησης είναι και οι σεισμοί (Aghababaei et al. 2020). Για τις μετακινήσεις που δημιουργούνται από περιοχές που επηρεάστηκαν λόγω του σεισμού επιλέγονται εναλλακτικές διαδρομές προκαλώντας αύξηση της ροής σε αυτούς τους δρόμους. Επιπλέον, αμέσως μετά από έναν σεισμό οι μετακινήσεις των βαρέων οχημάτων περιορίζονται σημαντικά.

Εκτός από τα φυσικά φαινόμενα, προβλήματα που επηρεάζουν τη συμπεριφορά μετακίνησης μπορούν να προκύψουν και από της ανθρώπινη δραστηριότητα. Χαρακτηριστικό παράδειγμα είναι η **κυκλοφοριακή συμφόρηση** έχει γίνει όλο και πιο οξύ πρόβλημα και ένα σημαντικό εμπόδιο που εμποδίζει τη βιώσιμη ανάπτυξη της κοινωνίας και επηρεάζει σοβαρά την ποιότητα ζωής των κατοίκων των πόλεων (Wang et al., 2015). Επί του παρόντος, σε όλο τον κόσμο, τόσο οι ανεπτυγμένες όσο και οι αναπτυσσόμενες χώρες προσπαθούν να βρουν λύσεις στο σοβαρό πρόβλημα και την πρόκληση της κυκλοφοριακής συμφόρησης. Ως εκ τούτου, η κατανόηση και η πρόβλεψη των επιλογών τρόπου μετακίνησης είναι ζωτικής σημασίας για τη διαχείριση της αστικής κυκλοφορίας. Σύμφωνα με τη μελέτη των Zhang et al., 2023 τα ηλεκτρικά ποδήλατα έχουν γίνει σταδιακά ένας κυρίαρχος τρόπος μετακίνησης στην πόλη Jinan της Κίνας με πληθυσμό 6 εκατομμύρια ανθρώπους. Ο κύριος σκοπός μετακίνησης ήταν από και προς τη δουλειά ή το σχολείο. Υπήρχαν δύο ώρες αιχμής (7:00-9:00 και 17:00-19:00).

Τα αποτελέσματα έδειξαν ότι το εισόδημα των ανθρώπων επηρέασε κατά πολύ την επιλογή τρόπου μετακίνησης και οι πλούσιοι άνθρωποι προτιμούσαν να χρησιμοποιούν Ι.Χ.

Οι Ding et al. (2017) στην μελέτη που εκπόνησαν ανέλυσαν τους λόγους που οδηγούν τους ανθρώπους στο να αποφύγουν τη χρήση Ι.Χ. και να στραφούν σε εναλλακτικούς τρόπους μεταφοράς. Επίσης ανέδειξαν και ορισμένες προτάσεις που μπορούν να υιοθετήσουν οι κυβερνήσεις ώστε να επιτύχουν τον σκοπό της μείωσης της κυκλοφοριακής συμφόρησης. Σύμφωνα με τις συνολικές επιδράσεις, παρατηρείται ότι η υψηλότερη πυκνότητα πληθυσμού και η πυκνότητα απασχόλησης αυξάνουν την καθημερινή κυκλοφοριακή κίνηση και ως εκ τούτου μειώνουν την πιθανότητα επιλογής του αυτοκινήτου. Η υψηλότερη προσβασιμότητα συνδέεται με υψηλότερη πιθανότητα επιλογής του περπατήματος/ποδηλάτου, ενώ η απόσταση από τη συγκοινωνία έχει σημαντική επίδραση στην επιλογή της συγκοινωνίας και του περπατήματος/ποδηλάτου. Τα ευρήματα της έρευνας υποστηρίζουν ότι τρόπος που ορίζονται οι περιοχές κατοικίας, εργασίας, αναψυχής και εμπορίου θα πρέπει να αποτελεί μέρος της στρατηγικής για τη μείωση της χρήσης του αυτοκινήτου. Για παράδειγμα, η δημιουργία περιοχών με μικτές χρήσης γης που συνδυάζουν κατοικία, εργασία και ψυχαγωγία σε μια περιοχή μπορεί να μειώσει την ανάγκη για μετακινήσεις με το αυτοκίνητο, καθώς οι άνθρωποι μπορούν να περπατήσουν ή να χρησιμοποιήσουν ποδήλατο για να φτάσουν σε διάφορες δραστηριότητες. Επίσης, η αύξηση των πιθανοτήτων επιλογής μέσων μεταφοράς και πεζοπορίας/ποδηλασίας θα προέλθει από τη δημιουργία πιο συμπαγών κοινοτήτων με υποδομές υψηλής διασύνδεσης που υποστηρίζουν περισσότερες μη μηχανοκίνητες μετακινήσεις. Όπως αναφέρθηκε και προηγουμένως οι αναπτυσσόμενες χώρες προσπαθούν να βρουν τους λόγους που προκαλούν την κυκλοφοριακής συμφόρηση, με αφορμή αυτό οι Le & Trinh (2016) προχώρησαν σε μια έρευνα στο Βιετνάμ η οποία απέδειξε ότι η αυξημένη κυκλοφορία έχει ως αποτέλεσμα την ραγδαία αύξηση του χρόνου μετακίνησης των χρηστών των Ι.Χ. και ταυτόχρονα την αύξηση του κόστους μετακίνησης λόγω της αυξημένης κατανάλωσης καυσίμου. Επίσης οι μετακινούμενοι λόγω της χρονικής καθυστέρησης συχνά επιλέγουν χώρους στάθμευσης ώστε να μην αργοπορήσουν με αποτέλεσμα να αυξάνεται και άλλο το κόστος. Τα παραπάνω ωθούν ένα μέρος των ανθρώπων να επιλέξουν τα Μ.Μ.Μ. (κυρίως μετρό και τρένα που δεν επηρεάζονται καθόλου από την κίνηση στους δρόμους) εφόσον έχουν να καλύψουν μεγαλύτερες αποστάσεις ενώ επιλέγουν το περπάτημα και το ποδήλατο όταν πρόκειται για σύντομες διαδρομές. Τα τελευταία δύο έχουν και το συγκριτικό πλεονέκτημα ότι είναι φιλικά προς το περιβάλλον.

Όσον αφορά στον τρόπο που προτυποποιείται η ταξιδιωτική συμπεριφορά, είναι σημαντική η ικανότητα εξήγησης των παραγόντων που επηρεάζουν τις αποφάσεις των χρηστών (Adadi and Berrada 2018). Στην έρευνα των Vlahogianni et al. (2022) χρησιμοποιήθηκαν τέσσερα (4) μοντέλα μηχανικής μάθησης και συγκεκριμένα τα Bayesian models, Random Forests, Gradient Boosting Machines και τα Ensemble Models. Για την ανάπτυξη των μοντέλων χρησιμοποιήθηκε η γλώσσα προγραμματισμού Python. Τα μοντέλα αναπτύχθηκαν δύο φορές, μια για ηλιόλουστες μέρες και μια για βροχερές. Αναπτύχθηκε επίσης το μοντέλο Mixed Logit με επιτυχία 59% γεγονός που το καθιστά ακατάλληλο για την εκτέλεση προβλέψεων αν και τα αποτελέσματα του παρέχουν ορισμένες ενδιαφέρουσες πληροφορίες. Τα αποτελέσματα έδειξαν ότι τα μοντέλα μηχανικής μάθησης με αλγόριθμους βασισμένους σε δέντρα (όπως Random Forest και Gradient Boosting) υπερτερούν των

κλασικών στατιστικών μεθόδων μιας και η ακρίβεια των μεθόδων ξεπέρασε το 80%. Σύμφωνα με άλλη έρευνα ανάμεσα στην πολυωνυμική λογιστική παλινδρόμηση (MNL) και το τεχνητό νευρωνικό δίκτυο (ANN) τα αποτελέσματα έδειξαν ότι το μοντέλο ANN ήταν καλύτερο ως προς την ακρίβεια στην προβλεψιμότητα της επιλογής μεταφορικού μέσου (Hussain et al., 2017). Στην μελέτη των Liu et al. (2019) χρησιμοποιείται το μοντέλο NL (Nested-Logit) για να αναλύσει την συνδυασμένη χρησιμότητα των διαδρομών. Επίσης χρησιμοποιείται το λογισμικό STATA, σε συνδυασμό με τα κοινωνικοοικονομικά δεδομένα και τις προσωπικές πληροφορίες των χρηστών στην έρευνα δηλωμένης προτίμησης (Stated preferences) και την επιλογή τρόπου μετακίνησης.

Στην έρευνα των Sun et al. (2022) που αναφέρθηκε και παραπάνω χρησιμοποιήθηκαν δύο (2) τύποι επεξηγηματικών μεταβλητών που κωδικοποιούν κοινωνικοοικονομικά και ψυχολογικά χαρακτηριστικά των μετακινούμενων. Χρησιμοποιήθηκε ένα μοντέλο λογιστικής παλινδρόμησης (logit). Οι μεταβλητές που χρησιμοποιούνται για την εκτίμηση της ικανοποίησης από τις μετακινήσεις με λεωφορείο είναι πολύ-κατηγορικές (multi-categorical). πολύ-κατηγορικές ονομάζονται οι μεταβλητές των οποίων οι τιμές αναφέρονται στο διαχωρισμό των στοιχείων ενός δείγματος σε επιμέρους κατηγορίες (παραπάνω από δύο). Διαχωρίζονται σε ονομαστικές (nominal) και διάταξης (ordinal). Οι πρώτες αφορούν χαρακτηριστικά των στοιχείων του δείγματος για τα οποία δεν υπάρχει κάποια σχέση ανάμεσα στις διαφορετικές κλάσεις. Για παράδειγμα σε μια ανάλυση μια τέτοια μεταβλητή θα μπορούσε να ήταν το μέσο μετακίνησης που επιλέγει κάποιος με πιθανές απαντήσεις όπως «Ι.Χ.», «Ταξί», «Περπάτημα/Ποδήλατο» και «Μ.Μ.Μ.» Οι κατηγορικές μεταβλητές διάταξης δηλώνουν μια ιεράρχηση ανάμεσα στις πιθανές κατηγορίες, για παράδειγμα τέτοια μεταβλητή μπορεί να είναι το ετήσιο εισόδημα ενός ανθρώπου με πιθανές απαντήσεις όπως «Χαμηλό», «Μεσαίο», «Υψηλό». Για τα ψυχολογικά χαρακτηριστικά χρησιμοποιήθηκε το μοντέλο δομικής εξίσωσης (Structural equation modeling), είναι μια πολύ-μεταβλητή στατιστική μέθοδος που περιλαμβάνει την εκτίμηση παραμέτρων για ένα σύστημα ταυτόχρονων εξισώσεων που χρησιμοποιούνται από επιστήμονες που διεξάγουν πειραματική έρευνα. Συνδυάζει τεχνικές ανάλυσης όπως την παλινδρόμηση (regression), την παραγοντική ανάλυση (factor analysis), την ανάλυση διαδρομής (path analysis) και τη μοντελοποίηση λανθάνουσας καμπύλης ανάπτυξης. Μια βασική διαδικασία στην εφαρμογή του SEM είναι η κατασκευή ενός δομημένου μοντέλου, το οποίο περιλαμβάνει μεταβλητές που εκφράζουν θεωρητικές υποθέσεις σχετικά με τις σχέσεις μεταξύ των μεταβλητών. Τέλος, το μοντέλο αυτό μπορεί να χρησιμοποιηθεί για να κάνει προβλέψεις ή να δοκιμάσει θεωρητικές υποθέσεις για αυτό και συνήθως χρησιμοποιείται σε επιστήμες που μελετούν την συμπεριφορά των ατόμων εξετάζοντας πειραματικά τον αντίκτυπο παραγόντων όπως σκέψεις, κίνητρα, κοινωνικές επιρροές κ.λπ. Ακόμα χρησιμοποιήθηκε η Εκτίμηση μέγιστης πιθανοφάνειας (Maximum likelihood estimate), η οποία και είναι μια από τις πιο διαδεδομένες μεθόδους εκτίμησης με πλεονεκτήματα όπως η ασυμπτωτική αμεροληψία και εγκυρότητα. Τέλος επειδή το να ταξιδεύει κάποιος μόνος ή με παρέα είναι δύο διαφορετικά σενάρια χρησιμοποιήθηκε η δυαδική λογαριθμική (BL), όπου κωδικοποιήθηκε ως "0" το να ταξιδεύει κάποιος μόνος και ως "1" το να ταξιδεύει κάποιος με παρέα.

Στη μελέτη των Zhang et al. (2023) χρησιμοποιήθηκε το μοντέλο στοίβαξης (stacking model), ένα μοντέλο μηχανικής μάθησης που χρησιμοποιείται για τον συνδυασμό των προβλέψεων πολλαπλών μοντέλων μηχανικής μάθησης. Για την επίλυση του προβλήματος ταξινόμησης το μοντέλο λειτουργεί σε δύο επίπεδα. Αρχικά, το πρώτο στρώμα του πλαισίου αποτελείται από πολλαπλούς ταξινομητές βάσης και η είσοδος κάθε ταξινομητή βάσης είναι τα δεδομένα εκπαίδευσης. Το δεύτερο στρώμα του πλαισίου είναι ο μετά-ταξινομητής. Τα δεδομένα εκπαίδευσης του μετά-ταξινομητή αποτελούνται από την έξοδο του πρώτου ταξινομητή βάσης και την ετικέτα (label) των αρχικών δεδομένων εκπαίδευσης. Στη συνέχεια, ο μετά-ταξινομητής εκπαιδεύεται για να λάβει τα τελικά αποτελέσματα. Βασικές αρχές του μοντέλου στοίβαξης αποτελούν η μη γραμμικότητα (μοντελοποιούνται πολύπλοκες μη γραμμικές σχέσεις μεταξύ των μεταβλητών) και η στοχαστικότητα (χρησιμοποιείται η τυχαιότητα κατά την εκπαίδευση αφού οι παράμετροι του μοντέλου εκτιμώνται μέσω επαναλαμβανόμενων διαδικασιών όπως η αναδρομική προσαρμογή των παραμέτρων) για την εκμάθηση και την αναπαράσταση πολύπλοκων σχέσεων στα δεδομένα. Στη συγκεκριμένη μελέτη το μοντέλο αποτελούνταν από επτά μοντέλα μηχανικής μάθησης για την πρόβλεψη της επιλογής μεταφορικού μέσου. Τα αποτελέσματα έδειξαν ότι η προτεινόμενη μέθοδος στοίβαξης πέτυχε ακρίβεια 0,83, ξεπερνώντας τον παραδοσιακό τρόπο MNL και εννέα άλλες μεθόδους μηχανικής μάθησης. Ειδικότερα, για τους τρεις κύριους τρόπους μετακίνησης - λεωφορείο, ιδιωτικό αυτοκίνητο και ηλεκτρικό ποδήλατο - το μοντέλο στοίβαξης παρουσίασε καλύτερες επιδόσεις. Παρουσιάστηκε τέλος η σημασία των χαρακτηριστικών των μοντέλων που βασίζονται σε δέντρα

3 Μεθοδολογική Προσέγγιση

3.1 Εισαγωγή

Η συλλογή των δεδομένων για την παρούσα μελέτη πραγματοποιήθηκε μέσω ενός ερωτηματολογίου που περιλαμβάνει ποικιλία ερωτήσεων. Στη βιβλιογραφία, συναντώνται δύο μέθοδοι καταγραφής των προτιμήσεων του κοινού, η μέθοδος της δεδηλωμένης προτίμησης (*stated preference*) στην οποία επιχειρείται η καταγραφή των προτιμήσεων των ερωτώμενων απέναντι σε κάποια νέα υποθετική κατάσταση και η μέθοδος της αποκαλυφθείσας προτίμησης (*revealed preference*), στην οποία καταγράφεται η συμπεριφορά και οι επιλογές του κοινού, γύρω από υπάρχοντα εναλλακτικά σενάρια και επομένως στηρίζεται σε μετρήσεις και παρατηρήσεις υπάρχοντων καταστάσεων. Προτιμήθηκε οι ερωτήσεις να βασίζονται στη μέθοδο της δεδηλωμένης προτίμησης. Τα κυριότερα πλεονεκτήματα της μεθόδου της δεδηλωμένης προτίμησης είναι τα ακόλουθα: Αρχικά είναι απλή στην κατανόηση και χρήση. Μπορεί επίσης να εφαρμοστεί και να αποφέρει αποτελέσματα σε σύντομο χρονικό διάστημα και με μικρό σχετικά κόστος. Για αυτό το λόγο αποτελεί ένα ευρέως διαδεδομένο εργαλείο για την εκπόνηση συγκοινωνιακών μελετών που επιδιώκουν τον προσδιορισμό των χαρακτηριστικών της συμπεριφοράς των οδηγών. Τέλος, αντλεί πληροφορίες για τη συμπεριφορά των ερωτώμενων βασιζόμενες σε ερωτήσεις για υποθετικά σενάρια και αξιοποιείται ως εργαλείο έρευνας προτιμήσεων σε νέες υπηρεσίες (Vittoratos et al., 2017). Σε αυτού του τύπου τις ερωτήσεις οι συμμετέχοντες καλούνται να αποφασίσουν το προτιμώμενο σενάριο μεταξύ δύο ή περισσότερων επιλογών που έχουν θέσει οι ερευνητές. Όσον αφορά στην πορεία που πρέπει να ακολουθηθεί ώστε να κατασκευαστεί ένα τέτοιο ερώτημα το πρώτο βήμα είναι η κατανόηση του συγκεκριμένου προβλήματος επιλογής για το οποίο απαιτείται ο πειραματικός σχεδιασμός. Στη συγκεκριμένη περίπτωση το ερώτημα που πρέπει να απαντηθεί είναι το εξής: «Ποια η πιθανότητα να αλλάξει κάποιος μεταφορικό μέσο σε περίπτωση εμφάνισης ενός απροσδόκητου συμβάντος;». Σε αυτό το στάδιο, ο αναλυτής πρέπει να καθορίσει τον αριθμό των εναλλακτικών λύσεων που απαιτούνται για το πείραμα, τα χαρακτηριστικά που σχετίζονται με κάθε μία από τις εναλλακτικές λύσεις και τα επίπεδα των χαρακτηριστικών που θα χρησιμοποιηθούν στη έρευνα.

Στόχος του ερωτηματολογίου είναι να προσδιοριστεί πως η εισαγωγή νέων υπηρεσιών, οι νέες στρατηγικές διαχείρισης της κυκλοφορίας ή η αλλαγή των στρατηγικών στόχων σε μια πόλη μπορεί να επηρεάσει τις επιλογές μετακίνησης και να κατανοήσει τις συνήθειες και τις αντιλήψεις των μετακινούμενων. Στις ενότητες που ακολουθούν περιγράφεται η δομή του ερωτηματολογίου που αναπτύχθηκε βάσει των αρχών των δεδηλωμένων προτιμήσεων, ενώ στη συνέχεια παρουσιάζεται αναλυτικά η διαδικασία που ακολουθήθηκε κατά τη συλλογή και

επεξεργασία των δεδομένων. Με τη χρήση της περιγραφικής στατιστικής παρατίθενται και σχολιάζονται γραφήματα που συμβάλουν στην κατανόηση και κατανομή των δεδομένων. Τέλος γίνεται εκτενής ανάλυση των μοντέλων που χρησιμοποιήθηκαν για την εξαγωγή αποτελεσμάτων.

3.2 Δομή ερωτηματολογίου

Το ερωτηματολόγιο αποτελείται από τρεις ενότητες και περίπου 25 ερωτήσεις, συμπεριλαμβανομένου ενός τμήματος με ερωτήσεις που αφορούν την πιθανότητα να αλλάξει κάποιος το μέσο που χρησιμοποιεί αν συμβεί κάποιο απροσδόκητο γεγονός. Στη συνέχεια, περιγράφονται αναλυτικά οι θεματικές ενότητες.

Ενότητα Α: Προφίλ Κινητικότητας

Αυτή η ενότητα αποσκοπεί στην ανίχνευση του τρέχοντος προφίλ κινητικότητας κάθε ταξιδιώτη. Για το σκοπό αυτό, κάθε ερωτώμενος καλείται να παράσχει πληροφορίες σχετικά με τις καθημερινές του συνήθειες μετακίνησης, όπως ο πιο συχνός σκοπός και τρόπος μετακίνησης, αν ταξιδεύει σε ώρες αιχμής ή όχι, η ευελιξία του/της όσον αφορά την ώρα άφιξης στον προορισμό του, το κόστος κ.α.. Ζητείται επίσης να δώσει μια προσέγγιση του αριθμού των εβδομαδιαίων μετακινήσεων για διάφορους σκοπούς όπως επαγγελματικοί, εκπαιδευτικοί, προσωπικοί (π.χ. ιατρικοί ή οικογενειακοί λόγοι), ή λόγοι αναψυχής (γυμναστήριο, φίλοι, κοινωνική ζωή κ.λπ.).

Ενότητα Β: Αντίληψη των χρηστών

Στην ενότητα αυτή, οι ερωτηθέντες καλούνται να παρέχουν την άποψη τους σχετικά με διάφορες στρατηγικές διαχείρισης της κυκλοφορίας και νέες υπηρεσίες που αποσκοπούν στη βελτίωση των συνθηκών του δικτύου. Χωρίζεται σε 5 υποενότητες:

B1. Τρέχουσα κατάσταση του δικτύου

Σε αυτή την υποενότητα, οι χρήστες εκφράζουν την ικανοποίησή τους από τις τρέχουσες συνθήκες του δικτύου μεταφορών κατατάσσοντας επιλεγμένα χαρακτηριστικά των δημόσιων μεταφορών (MMM) από τη χειρότερη στην καλύτερη και παρέχοντας μια ποιοτική άποψη ως προς το πόσο επηρεάζει η κυκλοφοριακή συμφόρηση τις καθημερινές τους μετακινήσεις.

B2. Συστήματα τιμολόγησης

Αυτή η υποενότητα καταγράφει την προθυμία των ερωτηθέντων να αποδεχθούν την εφαρμογή αστικών διοδίων. Επιπλέον, οι χρήστες καλούνται να δώσουν πληροφορίες σχετικά με το πώς θα αντιδράσουν σε περίπτωση που εφαρμοστεί ένα τέτοιο σύστημα.

B3. Υπηρεσίες κατά παραγγελία

Οι ερωτηθέντες δηλώνουν την προθυμία τους να χρησιμοποιήσουν μια υπηρεσία κατά παραγγελία και προχωρούν στην επιλογή του προτιμώμενου επιχειρηματικού μοντέλου

υπηρεσίας κατά παραγγελία από τρεις διαθέσιμες επιλογές: μια υπηρεσία πρώτου-τελευταίου χιλιομέτρου για το συγχρονισμό της μετεπιβίβασης μεταξύ υπηρεσιών κατά παραγγελία και τακτικών δημόσιων συγκοινωνιών (χαμηλή τιμή), μια υπηρεσία σταθερών στάσεων (μεσαία τιμή) και μια υπηρεσία πόρτα-πόρτα (υψηλή τιμή).

B4. Συγχρονισμός της κυκλοφορίας και των Μ.Μ.Μ.

Παρόμοια με τα παραπάνω, οι ερωτηθέντες εκφράζουν την προθυμία τους να χρησιμοποιήσουν μέσα μαζικής μεταφοράς για τις καθημερινές τους μετακινήσεις, εάν τους δινόταν προτεραιότητα μέσω των φωτεινών σηματοδοτών, και κατατάσσουν διάφορους στόχους που σχετίζονται με την εφαρμογή του προτεινόμενου συστήματος.

B5. Αντίδραση σε απροσδόκητα γεγονότα

Σε αυτό το τμήμα οι ερωτηθέντες απαντούν το κατά πόσο πιθανό είναι να αλλάξουν τον συνήθη τρόπο μετακίνησης τους στην εμφάνιση ενός απροσδόκητου συμβάντος. Μελετώνται 6 διαφορετικά γεγονότα:

- Δυσμενείς καιρικές συνθήκες [μεταβλητή CHA_ADVWEA]
- Ακραία καιρικά φαινόμενα (π.χ. σεισμός, πλημμύρα) [μεταβλητή CHA_ATHAZ]
- Έντονη κυκλοφοριακή συμφόρηση [μεταβλητή CHA_INTTRA]
- Απρογραμμάτιστη διακοπή των Μ.Μ.Μ. [μεταβλητή CHA_PTDIS]
- Προγραμματισμένες εκδηλώσεις (π.χ. οδικά έργα, ποδοσφαιρικός αγώνας) [μεταβλητή CHA_HUGEVE]
- Επικίνδυνα συμβάντα (όπως η πανδημία του COVID-19) [μεταβλητή CHA_UNEXEV]

Οι συμμετέχοντες έχουν να διαλέξουν μεταξύ 5 επιλογών.

- Αδύνατο
- Απίθανο
- Ούτε πιθανό ούτε απίθανο
- Πιθανό
- Δυνατό

Ακόμα δηλώνουν τον μέγιστο αποδεκτό χρόνο αναμονής για κρίσιμες και μη κρίσιμες μετακινήσεις, τη συνήθη αντίδρασή τους σε διακοπή των υπηρεσιών δημόσιων μεταφορών, καθώς και την αντίληψή τους για ενημέρωση που προσφέρει ο φορέας διαχείρισης των συγκοινωνιών σε περίπτωση διαταραχής ενός δρομολογίου.

Ενότητα Γ: Δημογραφικά χαρακτηριστικά

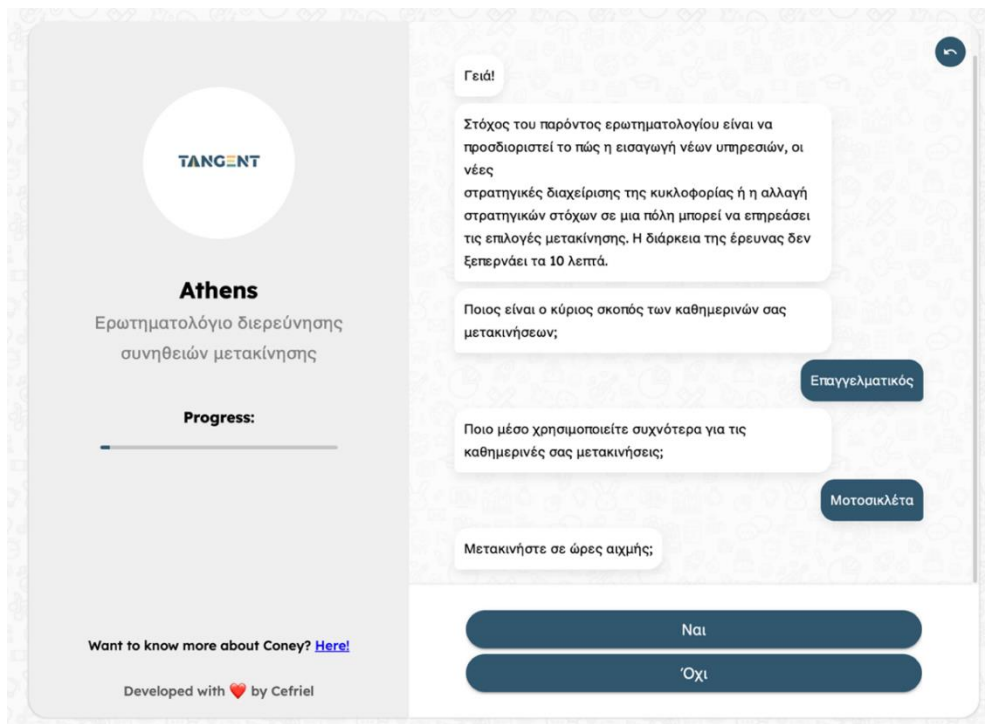
Στην τελευταία ενότητα του ερωτηματολογίου, οι χρήστες καλούνται να παράσχουν προσωπικές πληροφορίες σχετικά με το φύλο, την ηλικία, το ετήσιο ατομικό εισόδημα, το επάγγελμα, την ιδιοκτησία αυτοκινήτου και τον τόπο κατοικίας (εντός ή εκτός του κέντρου της πόλης). Επιπλέον, έπρεπε να επιλέξουν με βάση τη συνήθη επιλογή του τρόπου μεταφοράς τους ένα από τα ακόλουθα σενάρια:

- Εξαρτώμαι από το αυτοκίνητο μου,
- Είμαι οικολόγος,
- Είμαι ενθουσιώδης με την τεχνολογία,
- Εξοικονομώ χρόνο.

3.3 Συλλογή δεδομένων

Η αρχική διάρκεια της έρευνας ορίστηκε σε τέσσερις μήνες, με έναρξη στις 10 Οκτωβρίου 2022. Για την επίτευξη του προκαθορισμένου στόχου, η διάρκεια της έρευνας παρατάθηκε και συνεχίστηκε μέχρι τον Οκτώβριο του 2023. Η έρευνα διεξήχθη με τη χρήση του CONEY, μιας καινοτόμου πλατφόρμας που σχεδιάστηκε από την Cefriel για να βελτιώσει την εμπειρία του χρήστη κατά τη συμπλήρωση ερωτηματολογίων. Το CONEY αξιοποιεί μια μέθοδο που βασίζεται στη φυσική ροή της συνομιλίας με ένα γραφικό επεξεργαστή που σχεδιάστηκε με έμφαση στην ευκολία κατανόησης από τον χρήστη. Ο γραφικός επεξεργαστής προσφέρει ένα εύχρηστο περιβάλλον για τη δημιουργία και την επεξεργασία γραφικών, καθιστώντας την οπτική επεξεργασία προσβάσιμη σε όλους τους χρήστες, ανεξαρτήτως τεχνικής γνώσης.

Παρακάτω (Εικόνα 3.1) ακολουθεί στιγμιότυπο από το γραφιστικό περιβάλλον του ερωτηματολογίου.



Εικόνα 3.1: Διεπαφή συνομιλίας του CONEY

Η διάδοση του ερωτηματολογίου έγινε τόσο δια ζώσης όσο και διαδικτυακά. Οι διαδικτυακές μέθοδοι περιλάμβαναν τη χρήση πλατφορμών μέσων κοινωνικής δικτύωσης και προσωπικών επαφών. Οι πόλεις που συμμετείχαν ήταν η Αθήνα, η Λισαβόνα, το Μάντσεστερ και η Ρεν. Συγκεντρώθηκαν συνολικά 1859 έγκυρες απαντήσεις (εξαιρουμένων των ελλιπών και λανθασμένων απαντήσεων) για τις τέσσερις πόλεις. Συγκεκριμένα, συγκεντρώθηκαν 535 έγκυρες απαντήσεις για την περίπτωση της Αθήνας, 597 για τη Λισαβόνα, 333 για το Μάντσεστερ και 394 για τη Ρεν.

3.4 Προεπεξεργασία δεδομένων

Η προεπεξεργασία των δεδομένων είναι ένα σημαντικό βήμα που διασφαλίζει την ποιότητα και τη συνοχή των δεδομένων που χρησιμοποιούνται για την ανάπτυξη των μοντέλων. Το πρώτο βήμα που έγινε ήταν η ανάκτηση των δεδομένων, η οποία περιλαμβάνει την εξαγωγή των δεδομένων από το CONEY για κάθε πόλη και την αποθήκευσή τους. Μετά την ανάκληση των δεδομένων δημιουργήθηκαν τέσσερις (4) βάσεις δεδομένων, μία για κάθε πόλη. Ακόμα φτιάχτηκε μία βάση δεδομένων που περιλάμβανε τις απαντήσεις από όλες τις πόλεις.

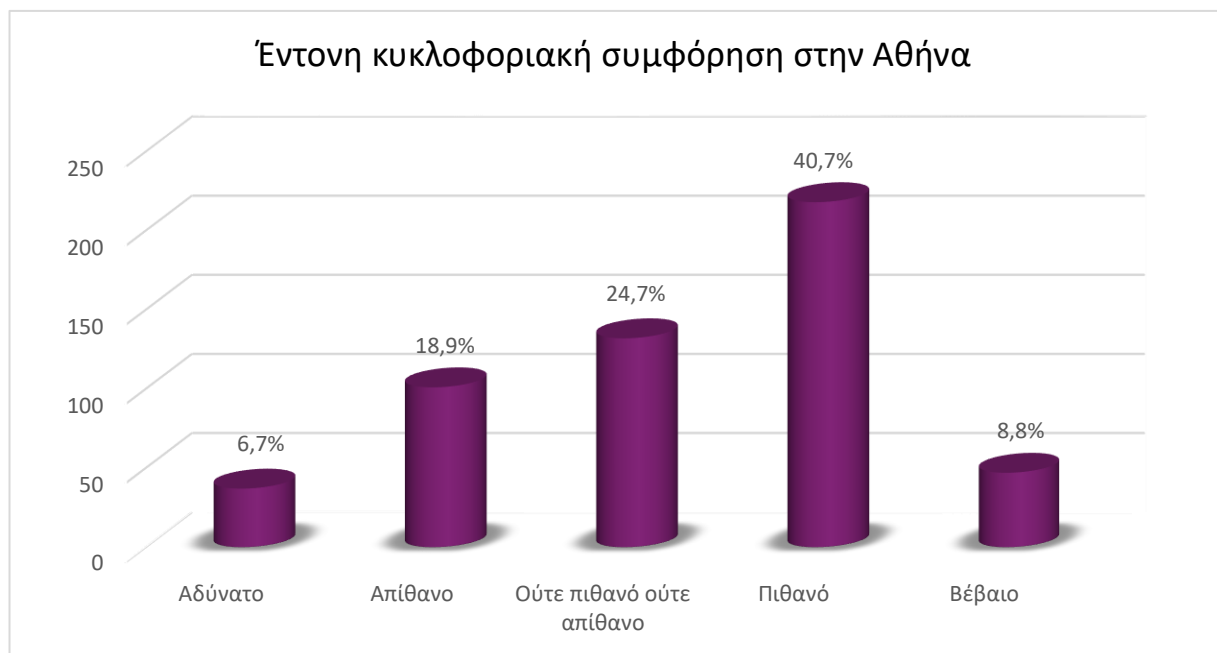
Το επόμενο βήμα ήταν ο καθαρισμός των δεδομένων. Πρόκειται για μια διαδικασία που έχει ως σκοπό τον εντοπισμό και την επίλυση των προβλημάτων που ενδέχεται να επηρεάσουν την ποιότητα των δεδομένων, όπως οι απουσιάζουσες τιμές, οι διπλές τιμές και οι ακραίες τιμές. Ακόμα περιλαμβάνει τη μετατροπή των δεδομένων σε τυποποιημένη μορφή δηλαδή κλιμακοποίηση των τιμών (feature scaling) ώστε να διασφαλιστεί ότι εμπίπτουν σε ένα συγκεκριμένο εύρος. Τέλος, έγινε κωδικοποίηση των κατηγορικών μεταβλητών δηλαδή μετατράπηκαν οι κατηγορικές μεταβλητές σε δυαδική μορφή ώστε να μπορούν να χρησιμοποιηθούν στα μοντέλα. Κάθε ένα από τα έξι απροσδόκητα συμβάντα που μελετώνται αντιστοιχεί σε μια κατηγορική μεταβλητή η οποία μπορεί να λάβει τις τιμές 0,1,2,3 και 4 όπου 0: Αδύνατο, 1: Απίθανο, 2: Ούτε πιθανό ούτε απίθανο, 3: Πιθανό και 4: Δυνατό. Επιλέχθηκαν τρεις διαφορετικοί τρόποι ομαδοποίησης των απαντήσεων.

- Στην 1^η περίπτωση ομαδοποιήθηκαν το 0, το 1 και το 2 στην πρώτη κατηγορία και το 3 και το 4 στη δεύτερη κατηγορία .
- Στην 2^η περίπτωση ομαδοποιήθηκαν το 0 και το 1 στην πρώτη κατηγορία και το 2, το 3 και το 4 στη δεύτερη κατηγορία.
- Στην 3^η και τελευταία αφαιρέθηκαν από το δείγμα οι απαντήσεις που αντιστοιχούν στις μεσαίες τιμές 1,2,3 και κρατήθηκαν το 0 στην πρώτη κατηγορία και το 4 στη δεύτερη.

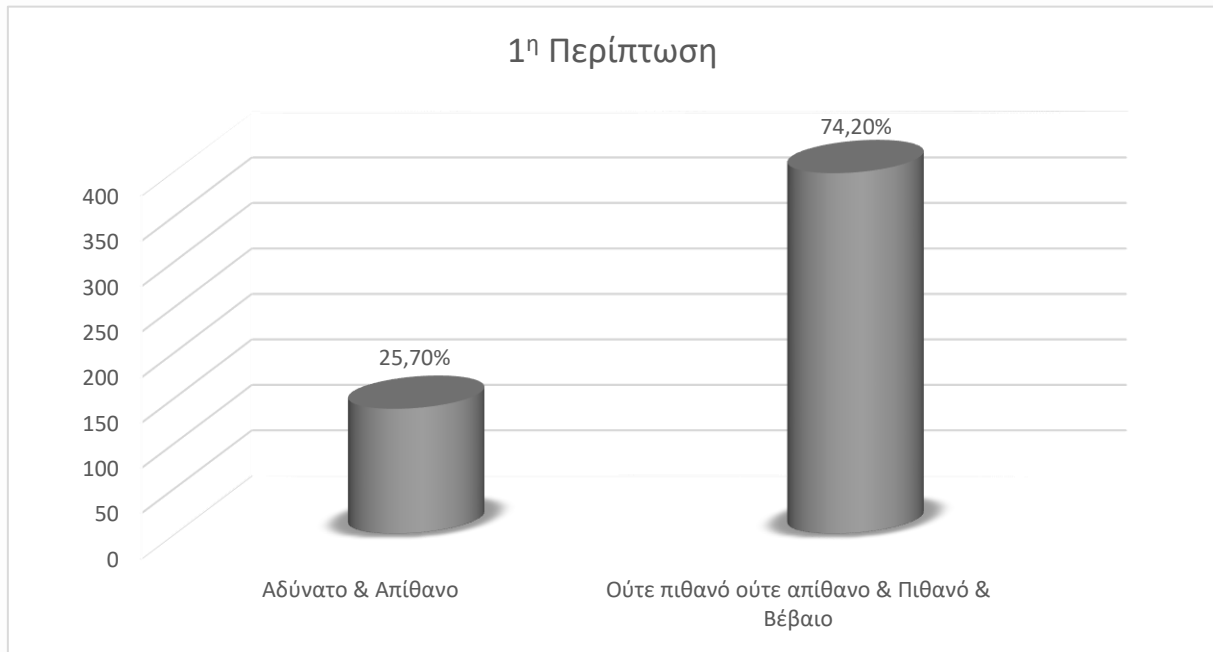
Ως επεξήγηση των παραπάνω ακολουθεί ένα παράδειγμα με την εξαρτημένη μεταβλητή CHA_INTTRA με τους τρεις (3) τρόπους (. Η μεταβλητή αυτή αντιστοιχεί στην εξής ερώτηση: «Ποια η πιθανότητα να αλλάξει κάποιος μεταφορικό μέσο σε περίπτωση έντονης κυκλοφοριακής συμφόρησης;» και οι απαντήσεις προέχονται από τη βάση δεδομένων που δημιουργήθηκε για την Αθήνα.

Πίνακας 3.1: Κατανομή των απαντήσεων της Αθήνας στο εξής ερώτημα: :
«Ποια η πιθανότητα να αλλάξει κάποιος μεταφορικό μέσο σε περίπτωση
έντονης κυκλοφοριακής συμφόρησης;»

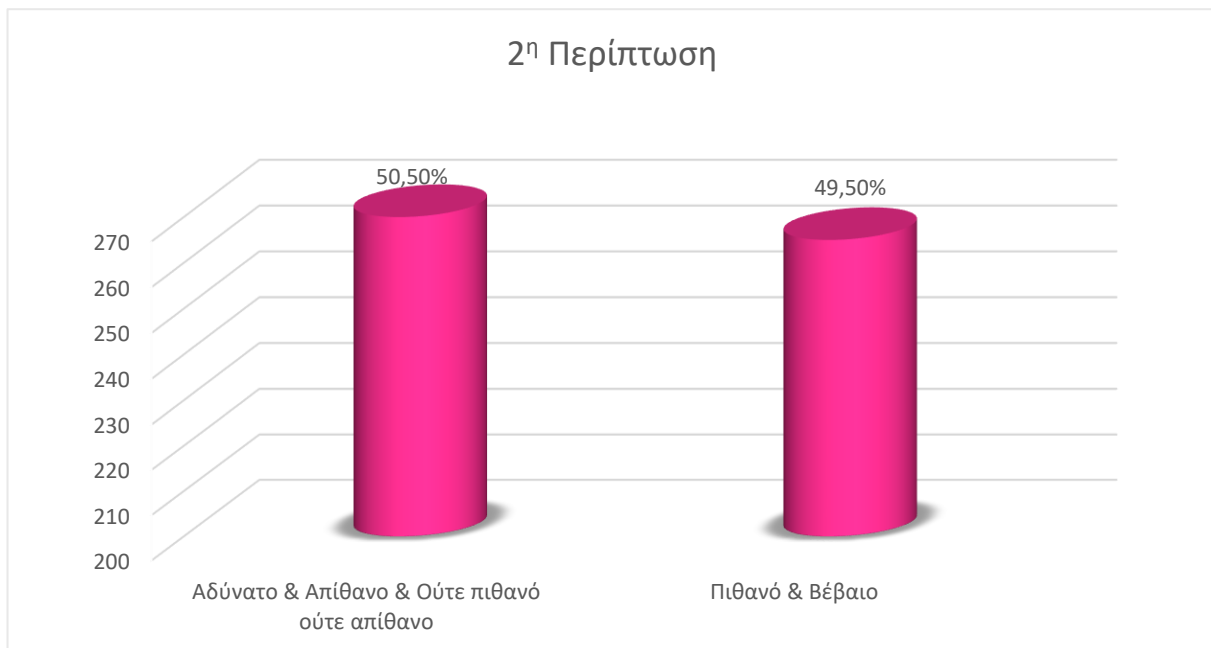
Έντονη κυκλοφοριακή συμφόρηση στην Αθήνα	
Αδύνατο	37
Απίθανο	101
Ούτε πιθανό ούτε απίθανο	132
Πιθανό	218
Βέβαιο	47



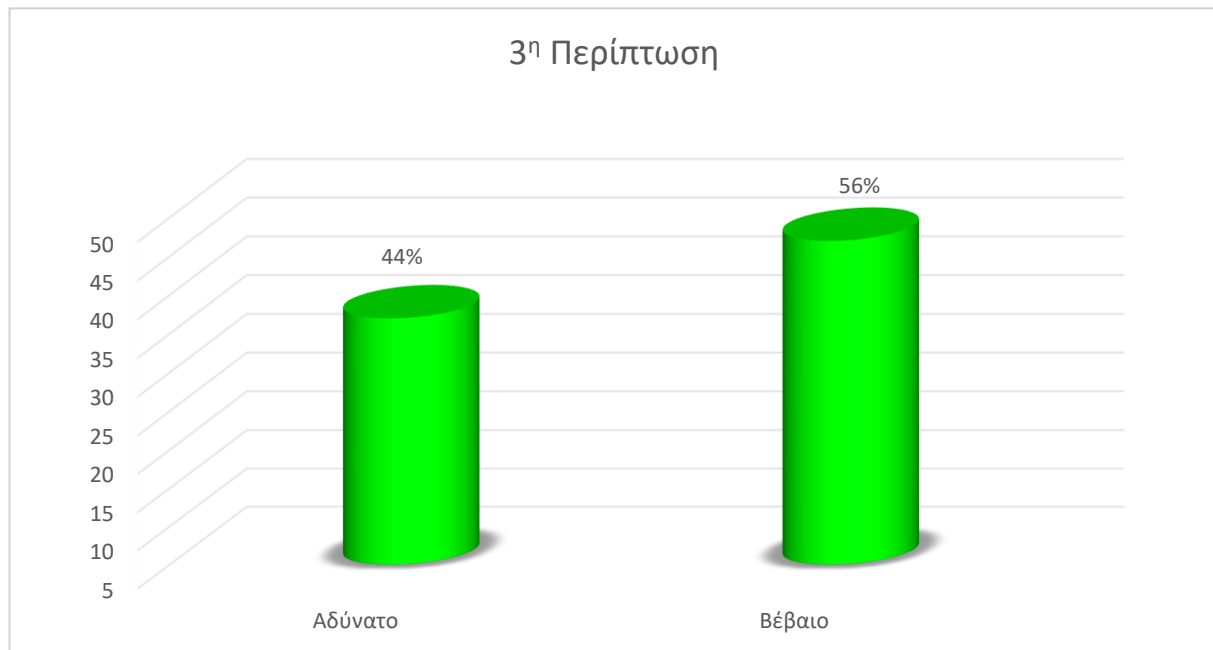
Διάγραμμα 3.1: Κατανομή των απαντήσεων της Αθήνας στο εξής ερώτημα: «Ποια η πιθανότητα να αλλάξει κάποιος μεταφορικό μέσο σε περίπτωση έντονης κυκλοφοριακής συμφόρησης»



Διάγραμμα 3.2: 1^η περίπτωση ομαδοποίησης της μεταβλητής CHA_INTTRA



Διάγραμμα 3.3: 2^η περίπτωση ομαδοποίησης της μεταβλητής CHA_INTTRA



Διάγραμμα 3.4: : 3^η περίπτωση ομαδοποίησης της μεταβλητής CHA_INTTRA

Από την 1^η περίπτωση δεν προέκυψαν αξιόπιστα μοντέλα λόγω της ανισότητας του δείγματος. Η παραπάνω διαδικασία έγινε για κάθε μία από τις έξι (6) εξαρτημένες μεταβλητές που μελετήθηκαν για όλες τις βάσεις δεδομένων που προέκυψαν από τις τέσσερις (4) πόλεις. Κρατήθηκε μόνο οι ομαδοποίηση με τον τρίτο τρόπο καθώς προέκυψαν αξιόπιστα μοντέλα. Οι λόγοι για τους οποίους ένα μοντέλο θεωρείται αξιόπιστο επεξηγούνται στο κεφάλαιο 3.7

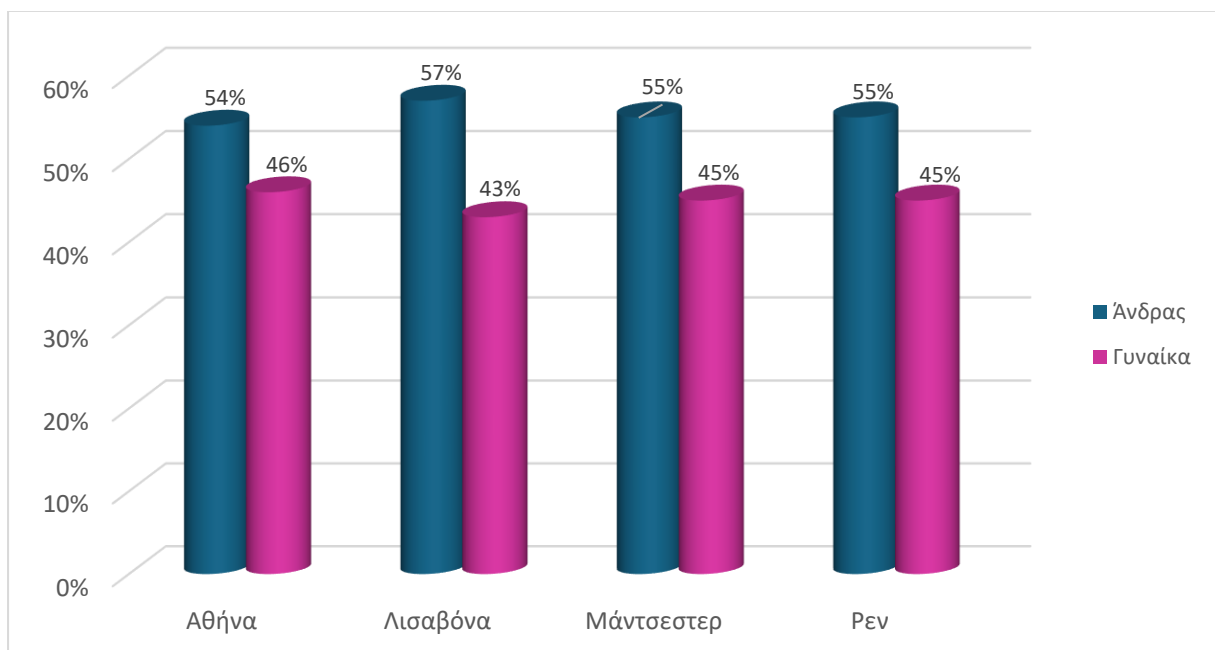
3.5 Περιγραφική στατιστική

Το επόμενο βήμα ήταν η χρήση της περιγραφικής στατιστικής η οποία είναι σημαντική για πολλούς λόγους στον τομέα της ανάλυσης δεδομένων. Ένας από τους βασικότερους λόγους είναι ό,τι συμβάλει στην κατανόηση των δεδομένων διότι παρέχει μια πλήρη εικόνα του τι αντιπροσωπεύουν αυτά τα δεδομένα. Ακόμα συνεισφέρει στην κατανομή των δεδομένων, αφού επιτρέπει να προσδιοριστούν τα σημεία που είναι πιο συχνά ή πιο ασυνήθιστα. Χρησιμοποιώντας τις δυνατότητες pivot table του Microsoft Excel κατασκευάστηκαν οπτικοποιήσεις όπως διαγράμματα πίτας, στηλών κ.α. Αμέσως μετά αναζητήθηκαν πιθανά πρότυπα ή τάσεις στα δεδομένα όπως ανίχνευση συσχετίσεων μεταξύ μεταβλητών, την εύρεση της μέσης τιμής (mean), της διαμέσου (median), της συχνότητας (frequencies) που αναφέρεται στην ποσότητα ή τον αριθμό εμφανίσεων κάθε διαφορετικής τιμής ή κατηγορίας σε ένα σύνολο δεδομένων και της διασποράς (variance) η οποία δείχνει τη διαφορά μεταξύ των προβλέψεων. Η διασπορά αποτελεί έναν τρόπο να γίνει αντιληπτό το πόσο τα δεδομένα διαφέρουν μεταξύ τους ή πόσο αποκλίνουν από τη μέση τους τιμή. Στη συνέχεια, ακολουθεί

η κατανόηση των σημείων που φαίνεται να αποτελούν εξαιρέσεις σε σχέση με την κατανομή των υπόλοιπων παρατηρήσεων. Σε αυτό το σημείο εξετάστηκαν ακραίες τιμές ώστε να γίνει αντιληπτό αν πρόκειται για φυσιολογική συμπεριφορά ή για σφάλμα στα δεδομένα. Τέλος, όλα τα παραπάνω οδηγούν στη λήψη αποφάσεων και τη διατύπωση ερωτημάτων ή υποθέσεων σχετικά με τη συνέχεια της ανάλυσης και της επεξεργασίας δεδομένων.

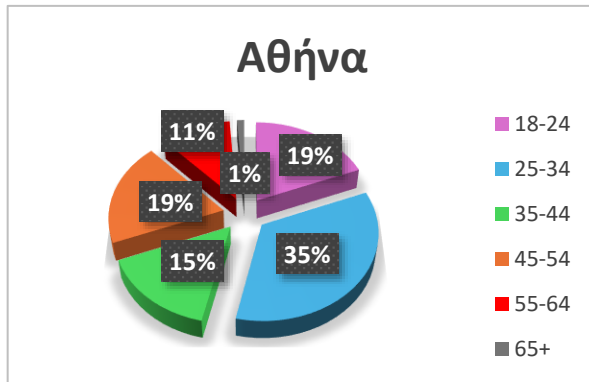
Μέσω αυτής της διαδικασίας εξήχθησαν κάποια χρήσιμα δεδομένα που αφορούν το φύλλο, την ηλικία, το σύννηθες μέσο μετακίνησης και το σκοπό μετακίνησης των ερωτηθέντων.

Στο Διάγραμμα 3.5 αποτυπώνεται η κατανομή του δείγματος σύμφωνα με την ηλικία των συμμετεχόντων. Πιο συγκεκριμένα και στις τέσσερις (4) πόλεις η πλειοψηφία του δείγματος αποτελείται από άνδρες με ποσοστά που κυμαίνονται από 54% -57%.

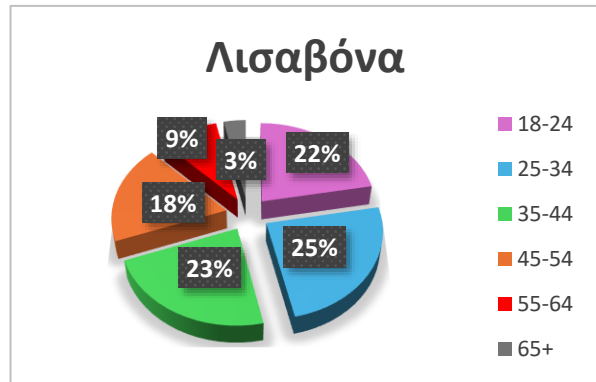


Διάγραμμα 3.5: Κατανομή συμμετεχόντων σε Αθήνα, Λισαβόνα, Μάντσεστερ, Ρεν

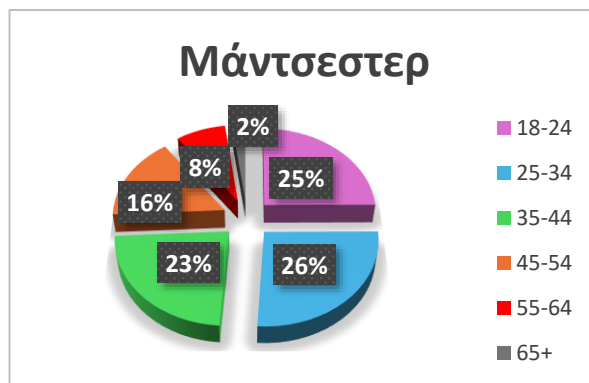
Στα διαγράμματα που ακολουθούν (Διαγράμματα 3.6, 3.7, 3.8, 3.9) εμφανίζονται τα ηλικιακά χαρακτηριστικά των συμμετεχόντων. Αν και τα ποσοστά διαφέρουν από πόλη σε πόλη παρατηρείται ότι η ηλικιακή ομάδα 65+ εκπροσωπείται με μονοψήφια ποσοστά από 1% έως 5%. Η ηλικιακή ομάδα 55-64 έχει ποσοστά από 8% στο Μάντσεστερ και φτάνει έως 11% τόσο στην Αθήνα όσο και στη Ρεν. Εκτός της Ρεν στις υπόλοιπες τρεις (3) πόλεις η ομάδα με τη μεγαλύτερη συχνότητα εμφάνισης είναι αυτή των 25-34 ετών με ποσοστά που κυμαίνονται από 25% στο Μάντσεστερ μέχρι 35% στην Αθήνα. Τα ηλικιακά διαγράμματα της Λισαβόνας και της Ρεν είναι πανομοιότυπα καθώς οι ηλικιακές ομάδες 18-24, 25-34, 35-44 και 45-54 καταλαμβάνουν παρόμοια ποσοστά. Στη Λισαβόνα από 18% (45-54) έως 25% (25-34) και στη Ρεν από 19% (35-44) έως 23% (18-24).



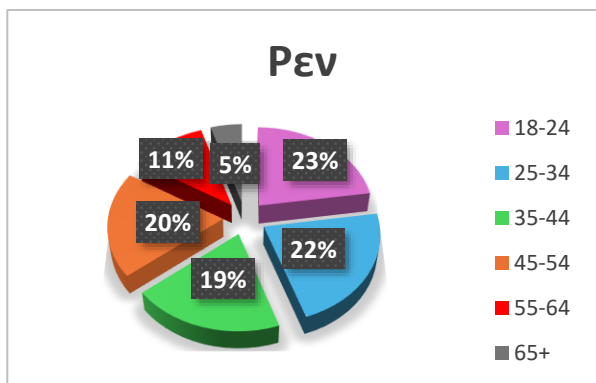
Διάγραμμα 3.6: Κατανομή συμμετεχόντων ανά ηλικία για Αθήνα



Διάγραμμα 3.7: Κατανομή συμμετεχόντων ανά ηλικία για Λισαβόνα

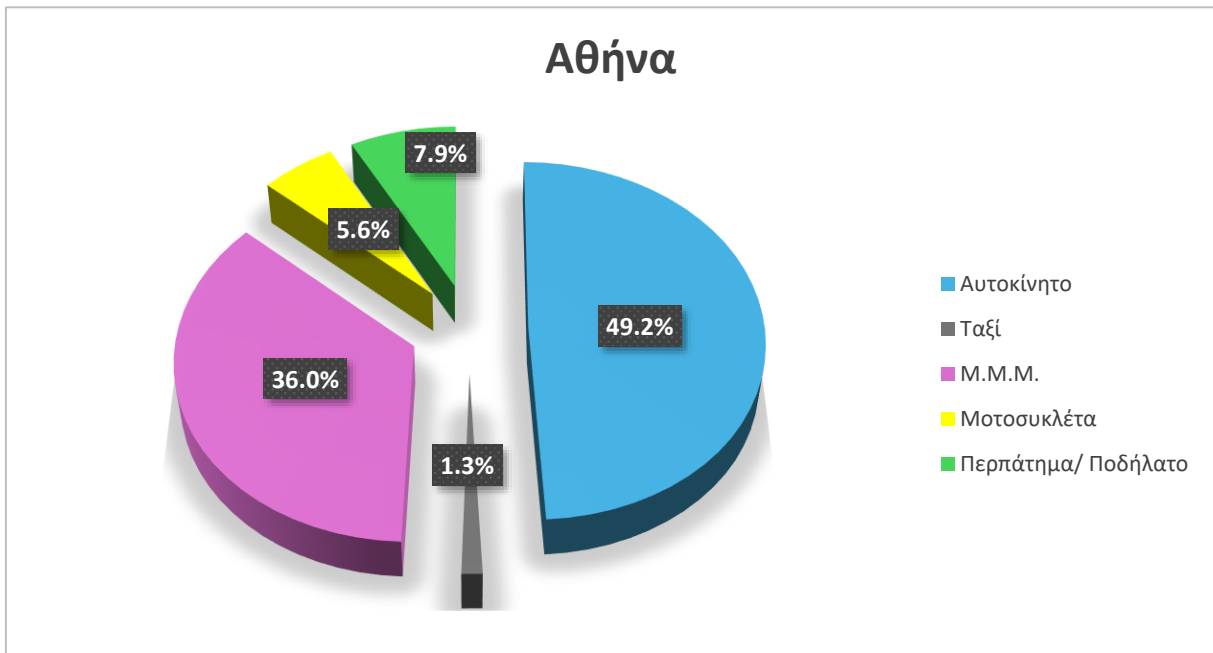


Διάγραμμα 3.8: Κατανομή συμμετεχόντων ανά ηλικία για Μάντσεστερ

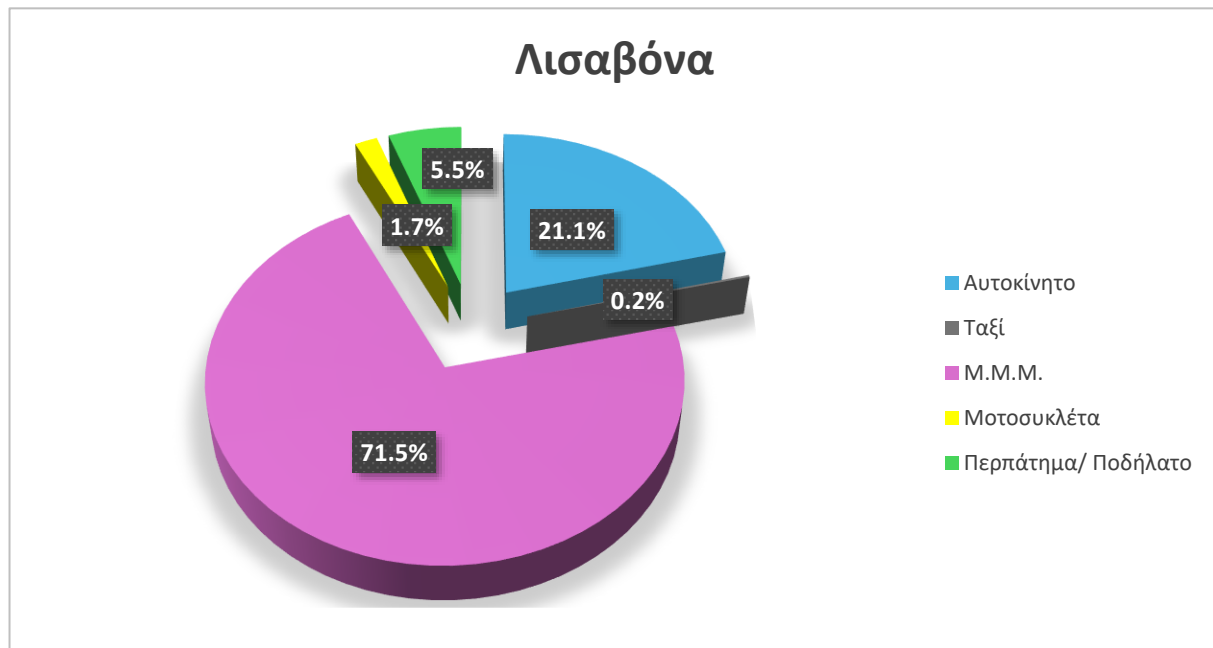


Διάγραμμα 3.9: Κατανομή συμμετεχόντων ανά ηλικία για Ρεν

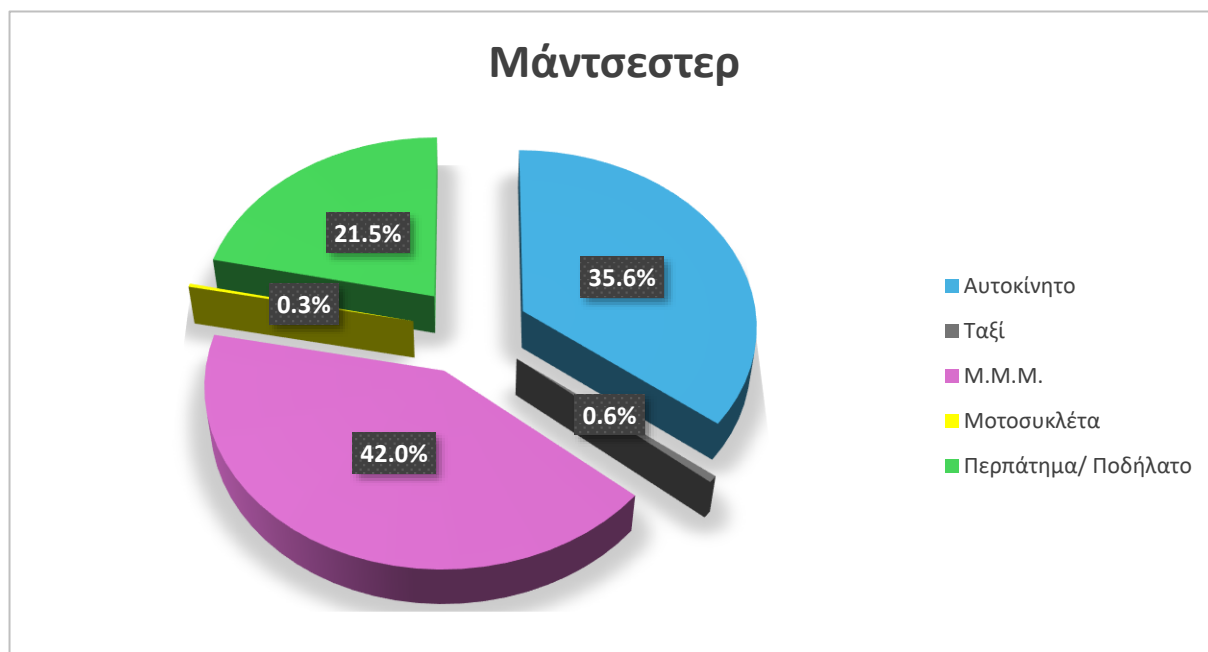
Ακολουθούν τα διαγράμματα στα οποία παρουσιάζεται η κατανομή του δείγματος με βάση το μέσο μετακίνησης που επιλέγουν οι ερωτώμενοι συχνότερα για να μετακινηθούν. Όσον αφορά τη μετακίνηση με ταξί τα ποσοστά είναι αμελητέα καθώς κυμαίνονται από 0,2% έως 1,3%. Όπως ήταν αναμενόμενο οι δύο επικρατέστερες απαντήσεις και στις τέσσερις πόλεις ήταν το αυτοκίνητο και τα Μ.Μ.Μ., με σημαντικές διαφορές όμως. Στην περίπτωση της Αθήνα (Διάγραμμα 3.6) σχεδόν ένας στους δύο (49,3%) μετακινείται με αυτοκίνητο σε αντίθεση με το Μάντσεστερ (Διάγραμμα 3.8) όπου το αντίστοιχο ποσοστό είναι 35,6%, στη Ρεν (Διάγραμμα 3.9) 34,4% ενώ στη Λισαβόνα (Διάγραμμα 3.7) μόλις το 21,1%. Τα Μ.Μ.Μ. στην Αθήνα, στο Μάντσεστερ και στη Ρεν τα επιλέγουν 36%, 42% και 38,3% αντίστοιχα ενώ στη Λισαβόνα το 71,5%. Η μοτοσυκλέτα εμφανίζεται με μεγαλύτερο ποσοστό στην Αθήνα όπου συγκεντρώνει 5,6% και μικρότερο στο Μάντσεστερ με 0,3%. Το Περπάτημα/Ποδήλατο επιλέγεται από πολύ μικρό ποσοστό στην Λισαβόνα και στην Αθήνα αφού συγκεντρώνει 5,5% και 7,9% αντίστοιχα σε αντίθεση με τους κατοίκους του Μάντσεστερ και της Ρεν που επιλέγουν τον συγκεκριμένο τρόπο μετακίνησης 21,5% και 22,2% αντίστοιχα.



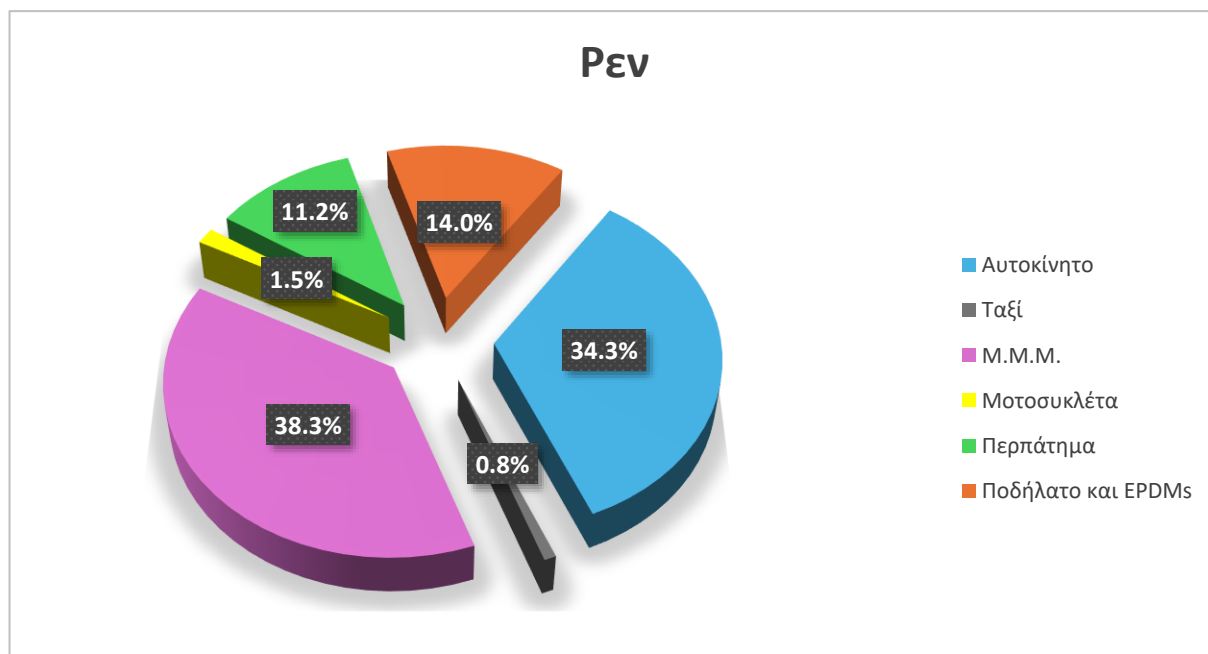
Διάγραμμα 3.10: Κατανομή των συμμετεχόντων με βάση το μέσο μετακίνησης που επιλέγουν συχνότερα για τις καθημερινές τους μετακινήσεις για Αθήνα



Διάγραμμα 3.11: Κατανομή των συμμετεχόντων με βάση το μέσο μετακίνησης που επιλέγουν συχνότερα για τις καθημερινές τους μετακινήσεις για Λισαβόνα



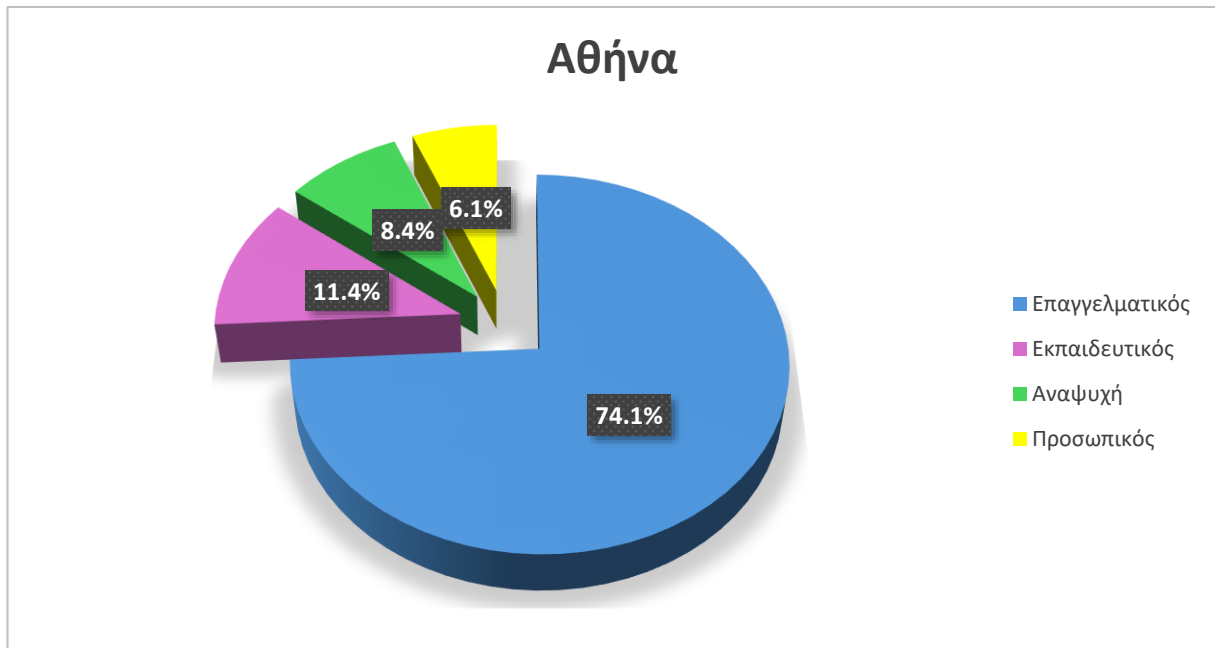
Διάγραμμα 3.12: Κατανομή των συμμετεχόντων με βάση το μέσο μετακίνησης που επιλέγουν συχνότερα για τις καθημερινές τους μετακινήσεις για Μάντσεστερ



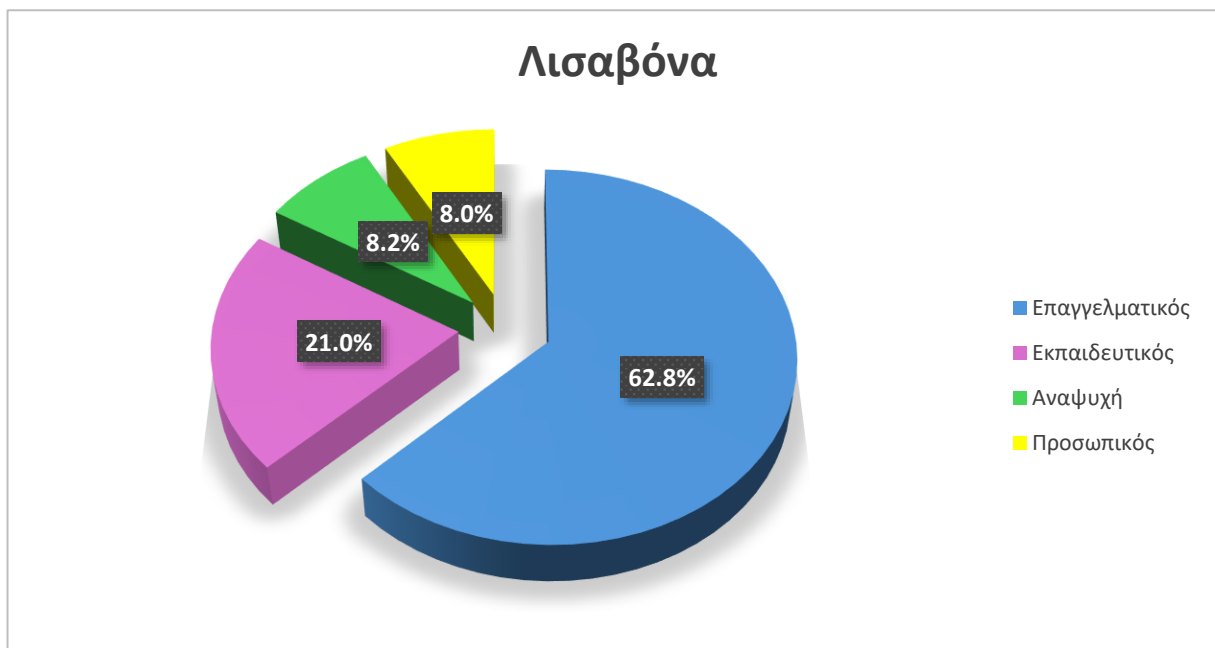
Διάγραμμα 3.13: Κατανομή των συμμετεχόντων με βάση το μέσο μετακίνησης που επιλέγουν συχνότερα για τις καθημερινές τους μετακινήσεις για Ρεν

Ακολουθεί η γραφική απεικόνιση του σκοπού της μετακίνησης των συμμετεχόντων. Ο συνηθέστερος λόγος που κάποιος μετακινείται είναι επαγγελματικός με ποσοστά που κυμαίνονται από 52,3% στη Ρεν (Διάγραμμα 3.17) έως 74,1% στην Αθήνα (Διάγραμμα 3.14). Ο εκπαιδευτικός σκοπός είναι δεύτερος και στις τέσσερις πόλεις με ποσοστά από 11,4% στην

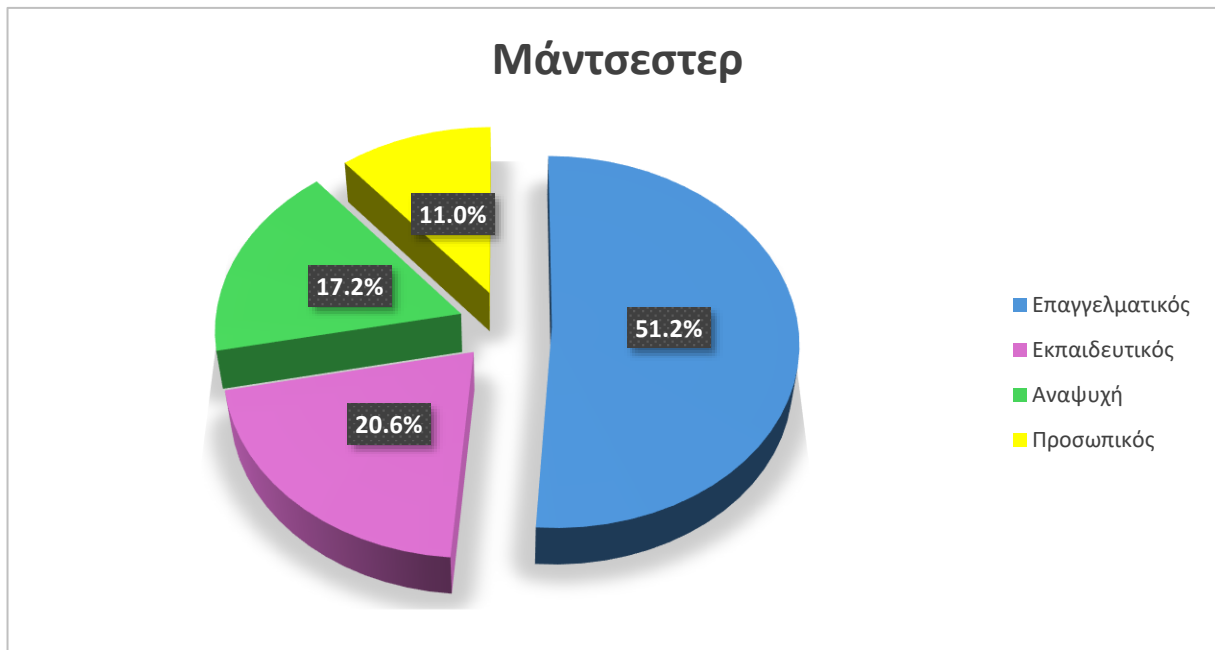
Αθήνα έως 24,4% στη Ρεν. Στην Αθήνα, στη Λισαβόνα (Διάγραμμα 3.15) και στη Ρεν το ποσοστό των ανθρώπων που μετακινείται για λόγους αναψυχής είναι παρόμοιο και κυμαίνεται από 8,2% στη Λισαβόνα έως 9,6% στη Ρεν σε αντίθεση με το Μάντσεστερ (Διάγραμμα 3.16) που το αντίστοιχο ποσοστό είναι 17,2%. Σε όλες τις πόλεις εκτός της Ρεν οι κάτοικοι μετακινούνται λιγότερο για προσωπικούς λόγους με ποσοστά που ξεκινάνε από 6,1% στην Αθήνα και φτάνουν έως 11% στο Μάντσεστερ.



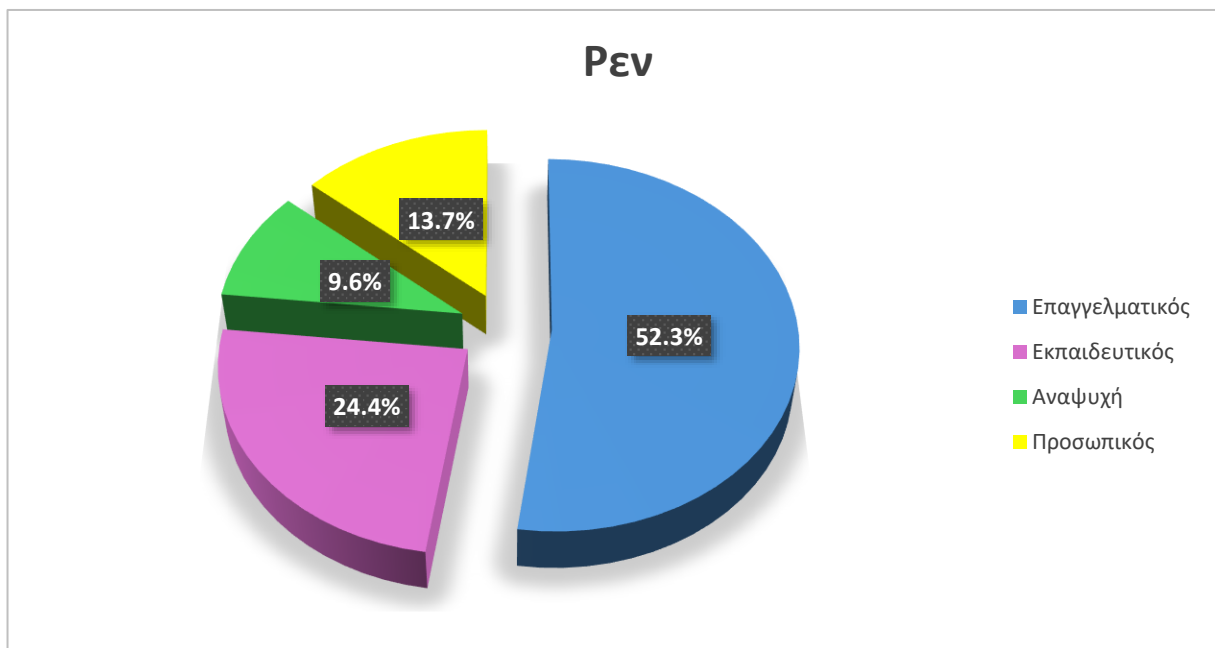
Διάγραμμα 3.14: Κατανομή των συμμετεχόντων με βάση τον κύριο σκοπό των καθημερινών τους μετακινήσεων για Αθήνα



Διάγραμμα 3.15: Κατανομή των συμμετεχόντων με βάση τον κύριο σκοπό των καθημερινών τους μετακινήσεων για Λισαβόνα



Διάγραμμα 3.15: Κατανομή των συμμετεχόντων με βάση τον κύριο σκοπό των καθημερινών τους μετακινήσεων για Μάντσεστερ



Διάγραμμα 3.16: Κατανομή των συμμετεχόντων με βάση το μέσο μετακίνησης που επιλέγουν συχνότερα για τις καθημερινές τους μετακινήσεις για Ρεν

3.6 Ανάπτυξη μοντέλων

Στο κεφάλαιο που ακολουθεί παρουσιάζονται οι βασικές αρχές λειτουργίας των αλγόριθμων ταξινόμησης μηχανικής μάθησης που εξετάστηκαν για τους σκοπούς της παρούσας διπλωματικής, όπως ο ταξινομητής δέντρων απόφασης (Decision Tree), ο ταξινομητής τυχαίου δάσους (Random Forest), ο ταξινομητής διανυσμάτων υποστήριξης (Support Vector Classifier), ο ταξινομητής K-Nearest Neighbors (KNN), ο ταξινομητής Gaussian Naive Bayes και ο πολυεπίπεδος νευρώνας perceptron (Multi-Layer perceptron).

3.6.1 Αλγόριθμοι ταξινόμησης

Οι αλγόριθμοι μηχανικής μάθησης διαχωρίζονται ανάλογα με το βαθμό επίβλεψης τους. Έτσι λοιπόν τα μοντέλα μηχανικής μάθησης διαχωρίζονται σε τρεις (3) κατηγορίες: την Επιβλεπόμενη μάθηση (Supervised learning), τη Μη επιβλεπόμενη μάθηση (Unsupervised learning) και την Ενισχυτική μάθηση (Reinforcement learning).

Στην παρούσα διπλωματική επιλέχθηκαν αλγόριθμοι ταξινόμησης οι οποίοι ανήκουν στην κατηγορία της Επιβλεπόμενης μάθησης. Ως Επιβλεπόμενη μάθηση ορίζεται το είδος μηχανικής μάθησης στο οποίο εξετάζεται μια άγνωστη συνάρτηση χρησιμοποιώντας δεδομένα τα οποία αποδίδουν τα αποτελέσματα της συνάρτησης αυτής. Σε τέτοιου είδους αλγορίθμους ορίζονται πληροφορίες εισόδου αλλά και γνωστά αποτελέσματα και ανάλογα με τους μεταξύ τους συσχετισμούς παράγεται το μοντέλο που τα περιγράφει. Η ταξινόμηση είναι μια μέθοδος, όπου το μοντέλο προσπαθεί να προβλέψει τη σωστή ετικέτα ενός δεδομένου εισόδου. Το μοντέλο εκπαιδεύεται πλήρως χρησιμοποιώντας τα δεδομένα εκπαίδευσης (train set) και στη συνέχεια αξιολογείται σε δεδομένα δοκιμής (test set) για την εκτέλεση πρόβλεψης σε νέα αχρησιμοποίητα δεδομένα. Υπάρχουν τέσσερις (4) διαφορετικοί τύποι προβλημάτων ταξινόμησης και είναι οι εξής: i) Η δυαδική ταξινόμηση (Binary classification), ii) Η ταξινόμηση πολλαπλών ετικετών (Multi-label classification), iii) Ταξινόμηση πολλαπλών κατηγοριών (Multi-Class classification), iv) Μη ισορροπημένη ταξινόμηση (Imbalanced classification).

Όπως έχει αναφερθεί και προηγουμένως η μελέτη αυτή έχει ως στόχο να απαντήσει στο εξής ερώτημα: «Ποια η πιθανότητα να αλλάξει κάποιος μεταφορικό μέσο σε περίπτωση εμφάνισης κάποιου απροσδόκητου συμβάντος;». Ενώ αρχικά οι απαντήσεις στο ερώτημα αυτό ήταν πέντε (5) με τις κατάλληλες τροποποιήσεις που περιγράφονται αναλυτικά στο κεφάλαιο 3.4 έγιναν δύο (2). Άρα το πρόβλημα εμπίπτει στην κατηγορία της δυαδικής ταξινόμησης. Η δυαδική ταξινόμηση αναφέρεται στις εργασίες ταξινόμησης που έχουν δύο ετικέτες κλάσης.

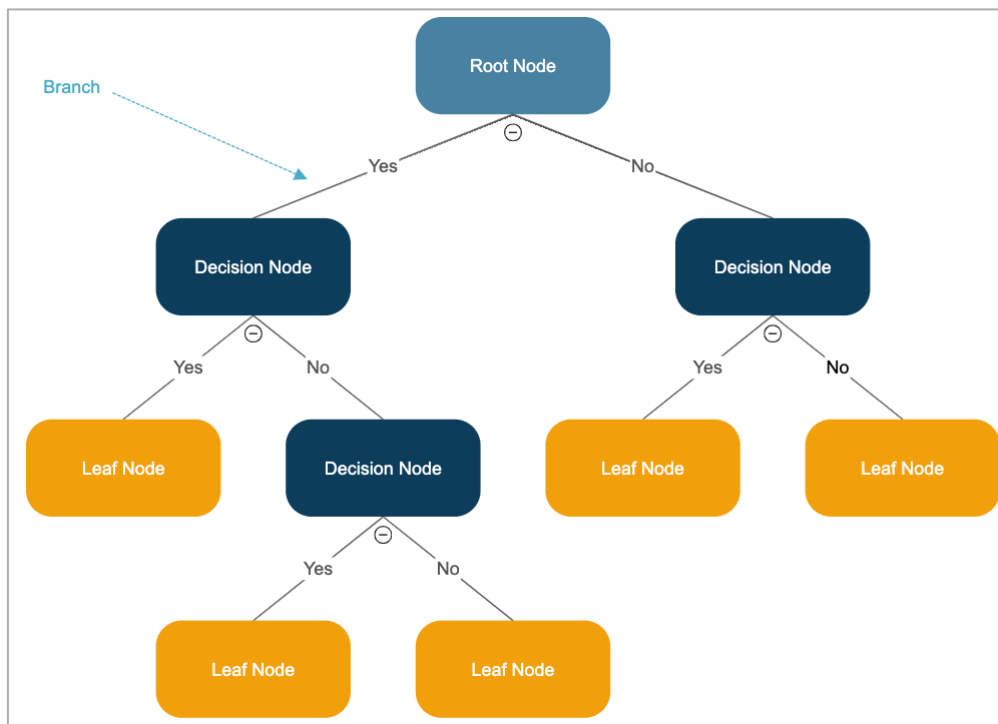
Αναπτύχθηκαν έξι (6) διαφορετικά μοντέλα μηχανικής μάθησης. Κατά τη διεξαγωγή των πειραμάτων με τους αλγόριθμους που χρησιμοποιήθηκαν χρειαζόταν η εύρεση των υπερπαραμέτρων που οδηγούν στην βέλτιστη απόδοση τους. Οι υπερπαραμέτροι ορίζονται

από τον χρήστη πριν τη διαδικασία μάθησης και καθορίζουν ιδιότητες του μοντέλου όπως η πολυπλοκότητα του, η μορφή της επιφάνειας απόφασης κ.α. Εάν κάποιος θέλει να βρει τις καλύτερες υπερπαραμέτρους μπορεί να δοκιμάσει τη μέθοδο της χειροκίνητης αναζήτησης, χρησιμοποιώντας πολλαπλές δοκιμές. Αυτό όμως θα χρειαζόταν τεράστιο χρόνο για τη δημιουργία ενός μόνο μοντέλου. Για το λόγο αυτό, εισήχθησαν μέθοδοι όπως η αναζήτηση πλέγματος (GridSearch), που χρησιμοποιήθηκε και στα έξι (6) μοντέλα που αναπτύχθηκαν. Κατά την τεχνική αυτή πραγματοποιείται εξαντλητικός έλεγχος σε συνδυασμούς υπαρπαραμέτρων με προκαθορισμένο εύρος τιμών προκειμένου να βρεθεί ο συνδυασμός που έχει την καλύτερη απόδοση. Χρησιμοποιήθηκε η βιβλιοθήκη της Python Gridsearchcn. Κάθε συνδυασμός υπαρπαραμέτρων αξιολογείται χρησιμοποιώντας κάποια μετρική αξιολόγησης, όπως η Ακρίβεια (Accuracy) ή το F1-score. Ένα παράδειγμα είναι το max_depth στον αλγόριθμο Decision Tree. Παρακάτω περιγράφονται αναλυτικά τα μοντέλα μηχανικής μάθησης που χρησιμοποιήθηκαν καθώς και ο τρόπος με τον οποίο αξιολογήθηκαν.

3.6.2 Δέντρα απόφασης

Τα Δέντρο αποφάσεων (Decision tree) είναι ένας μη παραμετρικός αλγόριθμος μηχανικής μάθησης που σημαίνει ότι δεν επιβάλλει περιοριστικές προϋποθέσεις σχετικά με τη μορφή των δεδομένων ή τη δομή του προβλήματος. Τα δέντρα αποφάσεων μπορούν να χρησιμοποιηθούν για προβλήματα ταξινόμησης και παλινδρόμησης, επίσης μπορούν να εφαρμοστούν τόσο σε κατηγορικές όσο και σε συνεχείς μεταβλητές. Έχει μια ιεραρχική δομή δέντρου που αποτελείται από τη βάση του δέντρου που ονομάζεται κόμβος ρίζας (Root node). Από εκεί προκύπτουν οι κόμβοι αποφάσεων (Decision node) που πρέπει να ληφθούν με βάση ένα κριτήριο διαχωρισμού. Οι κόμβοι φύλλων (Leaf or terminal node) είναι τα αποτελέσματα αυτών των αποφάσεων. Εκτός από τον ριζικό κόμβο, όλοι οι κόμβοι έχουν μόνο έναν εισερχόμενο κλάδο (Branch/Sub-Tree) και δύο ή περισσότερους εξερχόμενους.

Παρακάτω απεικονίζεται η δομή των δέντρων αποφάσεων (Εικόνα 3.2)



Εικόνα 3.2: Δομή των δέντρων απόφασης

Τα βήματα που ακολουθεί το Decision Tree ώστε να ταξινομήσει κάθε στοιχείο του δείγματος είναι τα εξής:

- **1^ο Βήμα :** Ο αλγόριθμος ξεκινάει από τη ρίζα (Starting at the Root), που ονομάζεται κόμβος ρίζας και αντιπροσωπεύει ολόκληρο το σύνολο των δεδομένων.
- **2^ο Βήμα :** Θέτει τις καλύτερες ερωτήσεις (Asking the Best Questions) δηλαδή αναζητά το πιο σημαντικό χαρακτηριστικό ή ερώτηση που χωρίζει τα δεδομένα στις πιο διακριτές ομάδες. Αυτό είναι σαν να θέτει μια ερώτηση σε μια διακλάδωση του δέντρου.
- **3^ο Βήμα:** Με βάση την απάντηση σε αυτή την ερώτηση, ο αλγόριθμος χωρίζει τα δεδομένα σε μικρότερα υποσύνολα, δημιουργώντας νέες διακλαδώσεις. Κάθε διακλάδωση αντιπροσωπεύει μια πιθανή διαδρομή μέσα στο δέντρο.
- **4^ο Βήμα:** Ο αλγόριθμος συνεχίζει να θέτει ερωτήσεις και να χωρίζει τα δεδομένα σε κάθε διακλάδωση μέχρι να φτάσει στους τελικούς κόμβους φύλλων που αντιπροσωπεύουν τα προβλεπόμενα αποτελέσματα ή τις ταξινομήσεις.

Παρακάτω παρατίθενται και επεξηγούνται οι παράμετροι του μοντέλου.

Split criterion (Κριτήριο διαχωρισμού): Ο αλγόριθμος αποφασίζει πώς θα διαχωρίσει τα δεδομένα σε κάθε κόμβο του δέντρου. Όταν η μεταβλητή-στόχος είναι κατηγορική όπως στις περιπτώσεις που μελετώνται στην παρούσα εργασία τότε οι μετρικές που αξιοποιεί για να αξιολογήσει τους διαχωρισμούς είναι το κέρδος πληροφορίας (Information Gain) και το ευρετήριο Gini (Gini impurity). Οι μετρικές αυτές μετράνε τη μη-καθαρότητα ενός κόμβου, δηλαδή πόσο τυχαία είναι τα δεδομένα.

Όσον αφορά την το κέρδος πληροφορίας βασίζεται στην έννοια της εντροπίας (Entropy) και δίνεται από τη σχέση:

$$\text{Information Gain} = 1 - \text{Entropy} \quad (1)$$

Η εντροπία (E) είναι μια θεμελιώδης έννοια που σχετίζεται με το δεύτερο νόμο της θερμοδυναμικής, ορίζει το μέτρο της αταξίας ενός συστήματος και χρησιμοποιείται για τον υπολογισμό της καθαρότητας ενός κόμβου. Όσο χαμηλότερη είναι η τιμή της εντροπίας, τόσο υψηλότερη είναι η καθαρότητα ενός κόμβου. Δεδομένου ότι η εντροπία αφαιρείται από τη μονάδα, το κέρδος πληροφορίας είναι υψηλότερο για τους καθαρότερους κόμβους με μέγιστη τιμή τη μονάδα (Shannon, 1948).

Η Εντροπία δίνεται από την παρακάτω σχέση:

$$E(S) = - \sum_{i=1}^n p_i \log_2 p_i \quad (2)$$

S είναι το σύνολο των δεδομένων εκπαίδευσης στον κόμβο, n οι διαφορετικές κατηγορίες και p_i το ποσοστό των δειγμάτων του S που ανήκουν στην κατηγορία i. Στην περίπτωση που υπάρχουν δύο κατηγορίες, η τιμή της είναι μηδέν (0) αν όλα τα δείγματα ανήκουν στην ίδια κατηγορία και ένα (1) αν τα μισά δείγματα ανήκουν στη μία κατηγορία και τα άλλα μισά στην άλλη.

Τα βήματα που ακολουθούνται για τη διάσπαση ενός δέντρου αποφάσεων με τη χρήση του κέρδους πληροφορίας είναι τα εξής:

- i. Για κάθε διάσπαση, υπολογίζεται ξεχωριστά η εντροπία κάθε κόμβου παιδιού (Child node).
- ii. Υπολογίζεται η εντροπία κάθε διάσπασης ως ο σταθμισμένος μέσος όρος της εντροπίας των κόμβων-παιδιών (Child nodes).
- iii. Γίνεται επιλογή της διάσπασης με τη χαμηλότερη εντροπία ή το υψηλότερο κέρδος πληροφορίας.
- iv. Τα βήματα i, ii και iii επαναλαμβάνονται μέχρι να δημιουργηθούν ομοιογενείς κόμβοι δηλαδή κόμβοι όπου τα δεδομένα που ανήκουν σε αυτούς είναι. Αυτό σημαίνει ότι τα δείγματα που βρίσκονται σε κάθε τέτοιο κόμβο είναι παρόμοια μεταξύ τους ως προς την κατηγορία ή την τιμή της μεταβλητής στόχου.

Σε ένα δέντρο αποφάσεων, ένας κόμβος που διαιρείται σε υποκόμβους είναι γνωστός ως γονικός κόμβος (Parent node) και οι υποκόμβοι που προκύπτουν από αυτόν αναφέρονται ως κόμβοι-παιδιά (Child node). Ο γονικός κόμβος αντιπροσωπεύει μια απόφαση ή συνθήκη, ενώ οι κόμβοι-παιδιά αντιπροσωπεύουν τα πιθανά αποτελέσματα ή τις περαιτέρω αποφάσεις με βάση την εν λόγω συνθήκη.

Η δεύτερη και δημοφιλέστερη μέθοδος διαχωρισμού για τη διάσπαση κόμβων όταν η μεταβλητή-στόχος είναι κατηγορική είναι το ευρετήριο Gini που δίνεται από την ακόλουθη σχέση:

$$Gini\ Impurity = 1 - Gini \quad (3)$$

Όπου το Gini είναι η πιθανότητα να χαρακτηριστεί σωστά ένα τυχαία επιλεγμένο στοιχείο, αν αυτό χαρακτηριστεί τυχαία σύμφωνα με την κατανομή των ετικετών στον κόμβο. Ο τύπος του Gini είναι:

$$Gini = \sum_{i=1}^n p_i^2 \quad (4)$$

Άρα ο τελικός τύπος του Gini Impurity είναι :

$$Gini\ Impurity = 1 - \sum_{i=1}^n p_i^2 \quad (5)$$

Όσο χαμηλότερο είναι το ευρετήριο Gini, τόσο υψηλότερη είναι η ομοιογένεια του κόμβου. Το Gini Impurity ενός καθαρού κόμβου είναι μηδέν.

Τα βήματα που ακολουθούνται για τη διάσπαση ενός δέντρου αποφάσεων με τη χρήση του ευρετηρίου Gini είναι παρόμοια με εκείνα που παρουσιάστηκαν στην περίπτωση της Εντροπίας. Γενικά, το Gini Impurity προτιμάται από το Information Gain επειδή δεν περιέχει λογαρίθμους οι οποίοι είναι υπολογιστικά απαιτητικοί.

Παραπάνω παρουσιάστηκε ο τρόπος με τον οποίο γίνεται ο διαχωρισμός ενός κόμβου. Ακολουθούν οι υπερπαράμετροι που «απαντούν» στο εξής ερώτημα: «Πότε σταματάει να μεγαλώνει ένα δέντρο απόφασης;»

Βάθος δέντρου (max_depth): Ο αριθμός των επιπέδων του δέντρου. Ένα μεγαλύτερο βάθος μπορεί να οδηγήσει σε πιο περίπλοκα μοντέλα, αλλά μπορεί να προκαλέσει και overfitting.

Ελάχιστος αριθμός δειγμάτων για διαχωρισμό (min_samples_split): Σε αυτήν την υπερπαράμετρο καθορίζεται ο ελάχιστος αριθμός δειγμάτων που απαιτείται για να γίνει ένας διαχωρισμός. Για παράδειγμα, μπορεί να τεθεί ως ελάχιστος αριθμός δειγμάτων το δέκα (10) για να καταλήξει σε μια απόφαση. Αυτό σημαίνει ότι ένας κόμβος έχει λιγότερα από 10 δείγματα, τότε χρησιμοποιώντας αυτή την υπερπαράμετρο, θα σταματήσει την περαιτέρω διάσπαση αυτού του κόμβου και θα γίνει κόμβος φύλλου.

Ελάχιστος αριθμός δειγμάτων σε φύλλο (min_samples_leaf): Ο ελάχιστος αριθμός δειγμάτων που απαιτούνται σε ένα φύλλο δέντρου. Όσο αυξάνεται η τιμή του τόσο αυξάνεται και η πιθανότητα υπερπροσαρμογής του μοντέλου.

Κάποια από τα πλεονεκτήματα του μοντέλου Decision Tree είναι τα εξής: είναι εύκολα κατανοητό διότι η έξοδος του δέντρου μπορεί να γίνει αντιληπτή ακόμη και από άτομα που δεν έχουν αναλυτικό υπόβαθρο, καθώς δεν απαιτούνται στατιστικές γνώσεις για την ανάγνωση και την ερμηνεία τους. Η γραφική του αναπαράσταση είναι πολύ διαισθητική και

οι χρήστες μπορούν εύκολα να συσχετίσουν τις υποθέσεις τους. Επίσης, είναι ένας από τους ταχύτερους τρόπους εντοπισμού των πιο σημαντικών μεταβλητών και της σχέσης μεταξύ δύο ή περισσότερων μεταβλητών. Με τη βοήθεια του Decision Tree μπορούν να δημιουργηθούν νέες μεταβλητές/χαρακτηριστικά που έχουν καλύτερη ισχύ για την πρόβλεψη της μεταβλητής-στόχου. Μπορεί επίσης να χρησιμοποιηθεί στο στάδιο διερεύνησης δεδομένων. Για παράδειγμα, σε ένα πρόβλημα όπου υπάρχουν εκατοντάδες μεταβλητές, εκεί θα βοηθήσει στον εντοπισμό της πιο σημαντικής μεταβλητής. Απαιτείται λιγότερος καθαρισμός δεδομένων αφού απαιτεί σε σύγκριση με άλλα μοντέλα επηρεάζεται ελάχιστα από τις ακραίες τιμές και τις ελλείπουσες τιμές.

Όσον αφορά τα μειονεκτήματα του, αντιμετωπίζει προβλήματα στην υπεπροσαρμογή (Overfitting). Υπερπροσαρμογή συμβαίνει όταν ένα μοντέλο μαθαίνει πολύ καλά τις λεπτομέρειες μέσα στο σύνολο δεδομένων εκπαίδευσης, με αποτέλεσμα το μοντέλο να δυσκολεύεται να κάνει αξιόπιστες προβλέψεις για εξωτερικά δεδομένα.

Το πρόβλημα αυτό επιλύεται με τον καθορισμό περιορισμών στις παραμέτρους του μοντέλου και με το κλάδεμα (Pruning). Κατά το Pruning αποκόπτονται οι κόμβοι ή οι υποκόμβοι που δεν είναι σημαντικοί. Επιπλέον, αφαιρούνται οι κλάδοι που έχουν πολύ χαμηλή σημασία. Ενώ μπορεί να χρησιμοποιηθεί για συνεχείς μεταβλητές δεν παρουσιάζει καλά αποτελέσματα. Σε προβλήματα με συνεχείς αριθμητικές μεταβλητές, το δέντρο αποφάσεων χάνει πληροφορίες όταν κατηγοριοποιεί τις μεταβλητές σε διαφορετικές κατηγορίες.

3.6.3 Τυχαία Δάση

Το τυχαία δάση (Random Forest) είναι ένας ευρέως χρησιμοποιούμενος αλγόριθμος μηχανικής μάθησης που αναπτύχθηκε από τον Leo Breiman (Breiman, 2001). Η τεχνική μάθησης συνόλου που χρησιμοποιείται ονομάζεται Ensemble και σημαίνει συνδυασμός πολλαπλών μοντέλων. Στην συγκεκριμένη περίπτωση σημαίνει τη χρήση πολλαπλών δέντρων (μεμονωμένων μοντέλων) απόφασης για να κάνει καλύτερες προβλέψεις, δημιουργώντας ένα πιο ισχυρό και ακριβές συνολικό μοντέλο. Το Ensemble χρησιμοποιεί δύο τύπους μεθόδων, το Bagging και το Boosting.

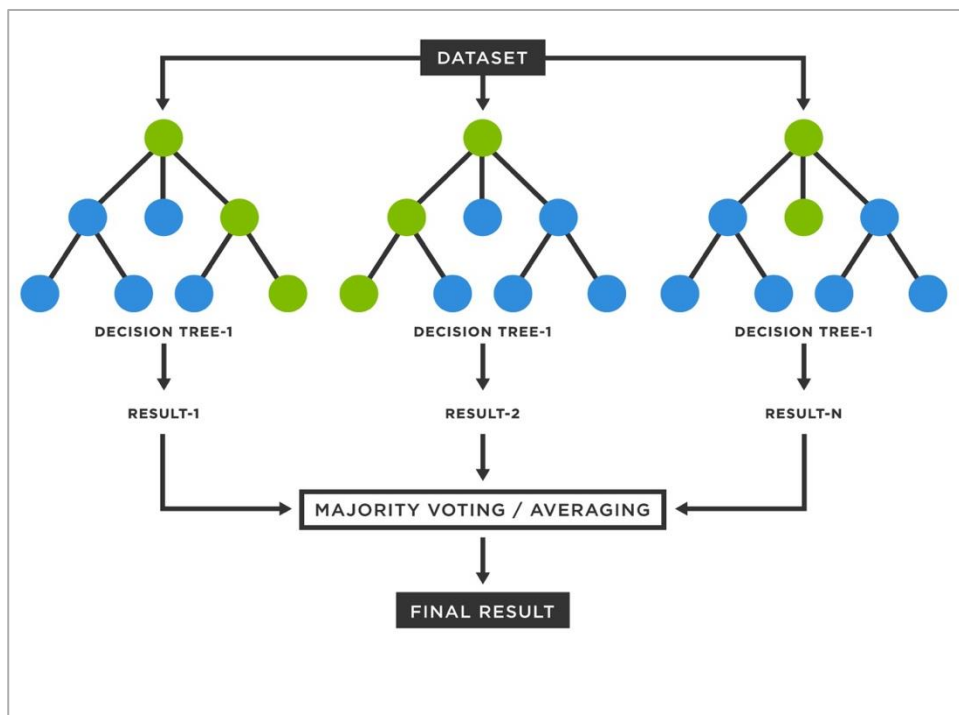
Το Bagging, γνωστό και ως Bootstrap Aggregation δημιουργεί ένα διαφορετικό υποσύνολο εκπαίδευσης από το δείγμα δεδομένων εκπαίδευσης με αντικατάσταση και η τελική έξοδος βασίζεται στην ψηφοφορία πλειοψηφίας (majority voting). Οι βασικές ενέργειες που περιλαμβάνει είναι οι εξής:

1. Επιλογή υποσυνόλου (Selection of subset): Ξεκινάει με την επιλογή ενός τυχαίου δείγματος, ή υποσυνόλου, από ολόκληρο το σύνολο δεδομένων.
2. Δειγματοληψία (Bootstrap Sampling): Κάθε μοντέλο δημιουργείται από αυτά τα δείγματα, που ονομάζονται Bootstrap Samples (είναι τυχαία επαναληπτικά δείγματα),

τα οποία λαμβάνονται από τα αρχικά δεδομένα με αντικατάσταση. Αυτή η διαδικασία είναι γνωστή ως δειγματοληψία σειρών (Bootstrapping). Κατά τη διαδικασία του Bootstrap, επιλέγεται τυχαία ένα στοιχείο από το αρχικό σύνολο δεδομένων, το οποίο προστίθεται στο νέο δείγμα και στη συνέχεια το επιστρέφει στο αρχικό σύνολο. Αυτή η διαδικασία επαναλαμβάνεται πολλές φορές μέχρι να δημιουργηθεί ένα νέο δείγμα με τον ίδιο αριθμό στοιχείων όπως και το αρχικό σύνολο δεδομένων.

3. Ανεξάρτητη εκπαίδευση μοντέλων (Independent Model Training): Κάθε μοντέλο εκπαιδεύεται ανεξάρτητα στο αντίστοιχο Bootstrap Sample. Αυτή η διαδικασία εκπαίδευσης παράγει αποτελέσματα για κάθε μοντέλο.
4. Ψηφοφορία πλειοψηφίας (Majority Voting): Το τελικό αποτέλεσμα καθορίζεται με τον συνδυασμό των αποτελεσμάτων όλων των μοντέλων μέσω ψηφοφορίας πλειοψηφίας. Επιλέγεται το πιο συνηθισμένο προβλεπόμενο αποτέλεσμα ανάμεσα στα μοντέλα.
5. Συγκέντρωση (Aggregation): Αυτό το βήμα, περιλαμβάνει τον συνδυασμό όλων των αποτελεσμάτων και τη δημιουργία του τελικού αποτελέσματος με βάση την ψηφοφορία πλειοψηφίας.

Παρακάτω (Εικόνα 3.3) απεικονίζεται η δομή ενός τυπικού Random forest.



Εικόνα 3.3: Δομή του Random Forest

[Πηγή: <https://www.spotfire.com/glossary/what-is-a-random-forest>]

Παρακάτω παρατίθενται και επεξηγούνται οι παράμετροι του μοντέλου.

Αριθμός δέντρων ($n_{estimators}$): Ο αριθμός των δέντρων που κατασκευάζει ο αλγόριθμος πριν αναλάβει τον μέσο όρο των προβλέψεων.

Μέγιστο πλήθος χαρακτηριστικών για διαχωρισμό (max_features): Το μέγιστο πλήθος των χαρακτηριστικών που το Random Forest λαμβάνει υπόψη για να διαχωρίσει έναν κόμβο. Ο όρος κόμβος αναφέρεται σε κάθε σημείο στο δέντρο όπου η απόφαση λαμβάνεται σχετικά με το πώς να διαχωριστούν τα δεδομένα εισόδου. Κάθε κόμβος έχει ένα χαρακτηριστικό και ένα κατώφλι (ή κατώτατο όριο) που χρησιμοποιείται για να διαχωριστούν τα δεδομένα σε δύο ομάδες

Ελάχιστος αριθμός δειγμάτων σε φύλλο (min_samples_leaf): Καθορίζει το ελάχιστο πλήθος φύλλων που απαιτούνται για να χωριστεί ένα εσωτερικό κόμβος. Τα «φύλλα» ενός δέντρου απόφασης είναι οι τερματικοί κόμβοι του δέντρου, δηλαδή οι κόμβοι που δεν έχουν περαιτέρω διακλαδώσεις

Κριτήριο διαχωρισμού (criterion): Πώς θα χωριστεί ο κόμβος σε κάθε δέντρο; Επιλογή ενός εκ των Information Gain και Gini impurity. Οι έννοιες αυτές επεξηγήθηκαν αναλυτικά παραπάνω.

Μέγιστος αριθμός φύλλων ανά δέντρο (max_leaf_nodes): Το μέγιστο πλήθος φύλλων σε κάθε δέντρο.

Βάθος δέντρων (max_depth): Το μέγιστο βάθος των δέντρων.

Ελάχιστος αριθμός δειγμάτων για διαχωρισμό ανά δέντρο (min_samples_split): Το ελάχιστο πλήθος των δειγμάτων που απαιτούνται για να διαχωριστεί ένας κόμβος.

Κάποια από τα πλεονεκτήματα του μοντέλου είναι τα εξής: μπορεί να χρησιμοποιηθεί σε προβλήματα ταξινόμησης και παλινδρόμησης. Είναι ανθεκτικό στο πρόβλημα της υπερπροσαρμογής (Overfitting), λόγω της διαδικασίας τυχαίας επιλογής χαρακτηριστικών και δειγματοληψίας, καθώς και επειδή η έξοδος βασίζεται στην ψηφοφορία πλειοψηφίας ή στον μέσο όρο. Επίσης, αποδίδει πολύ καλά ακόμα και αν τα δεδομένα παρουσιάζουν μηδενικές/απουσιάζουσες τιμές. Κάθε δέντρο απόφασης που δημιουργείται είναι ανεξάρτητο από το άλλο, παρουσιάζοντας έτσι την ιδιότητα του παραλληλισμού. Τέλος, ο αλγόριθμος δεν κινδυνεύει από τη διαστατικότητα (dimensionality) η οποία αναφέρεται στον αριθμό των χαρακτηριστικών ή των μεταβλητών που χρησιμοποιούνται για να περιγράψουν ένα σύνολο δεδομένων. Δεδομένου ότι κάθε δέντρο δεν εξετάζει όλα τα χαρακτηριστικά, ο χώρος των χαρακτηριστικών μειώνεται.

Όσον αφορά στα μειονεκτήματά του, είναι ιδιαίτερα πολύπλοκο σε σύγκριση με τα Decision trees, όπου οι αποφάσεις μπορούν να ληφθούν ακολουθώντας την πορεία του δέντρου. Επίσης, ο χρόνος εκπαίδευσης είναι μεγαλύτερος από άλλα μοντέλα λόγω της πολυπλοκότητάς του.

Τα τυχαία δάση (Random forest) είναι μια συλλογή από δέντρα αποφάσεων (Decision trees) ωστόσο υπάρχουν πολλές διαφορές στη συμπεριφορά τους οι οποίες συνοψίζονται στον παρακάτω πίνακα (Πίνακας 3.2).

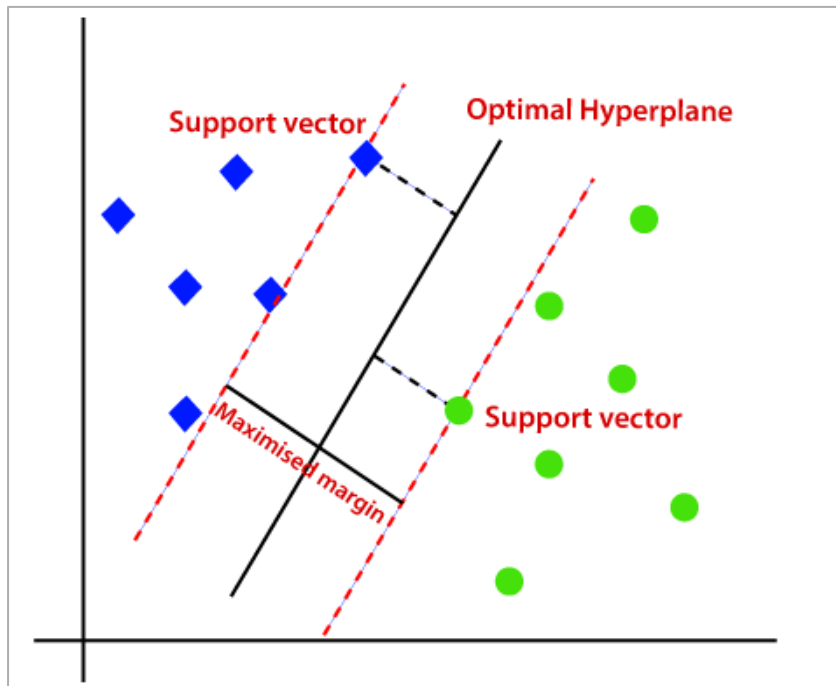
Πίνακας 3.2: Σύγκριση Random Forest και Decision Tree

	Random Forest	Decision Tree
Δομή	Σύνολο πολλαπλών δέντρων απόφασης	Ένα δέντρο απόφασης
Μεροληψία-Διασπορά	Χαμηλότερη διασπορά, δεν αντιμετωπίζει πρόβλημα υπερπροσαρμογής	Υψηλότερη διασπορά, επιρρεπές στην υπερπροσαρμογή
Ακρίβεια πρόβλεψης	Γενικά υψηλότερη λόγω συνόλου δέντρων	Επιρρεπές στο overfitting, μπορεί να ποικίλει
Ανθεκτικότητα	Πιο ανθεκτικό στις ακραίες τιμές και τον θόρυβο	Ευαίσθητο στις ακραίες τιμές και τον θόρυβο
Χρόνος εκπαίδευσης	Πιο αργό λόγω κατασκευής πολλών δέντρων	Γρηγορότερο, καθώς δημιουργεί ένα μόνο δέντρο
Ερμηνευσιμότητα	Λιγότερο ερμηνεύσιμο λόγω των πολλών δέντρων	Πιο ερμηνεύσιμο ως ένα δέντρο
Χρήση	Κατάλληλο για πολύπλοκες εργασίες, υψηλή διαστασιμότητα δεδομένων	Απλές εργασίες, εύκολη ερμηνεία

3.6.4 Μηχανές διανυσμάτων στήριξης

Πρόκειται για ένα μοντέλο μηχανικής μάθησης, το οποίο βρίσκει ένα υπερεπίπεδο που διαχωρίζει καλύτερα τις δύο κλάσεις. Ένα υπερεπίπεδο είναι μια επιφάνεια απόφασης που χρησιμοποιείται για το διαχωρισμό δεδομένων από διαφορετικές κλάσεις. Υπάρχουν δύο είδη SVM αλγορίθμων: Ο γραμμικός (Linear) και ο μη γραμμικός (Non-Linear). Στην προκειμένη περίπτωση χρησιμοποιείται ο γραμμικός. Δηλαδή τα δεδομένα είναι απόλυτα γραμμικά διαχωρίσιμα. Απόλυτα γραμμικά διαχωρίσιμο σημαίνει ότι τα σημεία δεδομένων μπορούν να ταξινομηθούν σε 2 κλάσεις με τη χρήση μιας μόνο ευθείας γραμμής (εάν πρόκειται για 2D). Στον αλγόριθμο SVM, απεικονίζεται κάθε στοιχείο δεδομένων ως ένα σημείο στον n -διάστατο χώρο (όπου n είναι ο αριθμός των χαρακτηριστικών), με την τιμή κάθε χαρακτηριστικού να είναι η τιμή μιας συγκεκριμένης συντεταγμένης. Σε ένα πρόβλημα με μόνο δύο χαρακτηριστικά, όπως το μέσο και ο σκοπός μετακίνησης ενός ατόμου, θα σχεδιασθούν πρώτα αυτές τις δύο μεταβλητές σε δισδιάστατο χώρο. Κάθε σημείο έχει δύο συντεταγμένες που ονομάζονται διανύσματα υποστήριξης (Support Vector). Εν ολίγοις, ο αλγόριθμος SVM στοχεύει στην εύρεση της καλύτερης γραμμής απόφασης ή ορίου (υπερεπίπεδο) μεταξύ των κλάσεων. Για να το επιτύχει αυτό, βρίσκει το πλησιέστερο σημείο των γραμμών και των δύο κλάσεων. Η απόσταση μεταξύ των διανυσμάτων και του υπερεπιπέδου ονομάζεται περιθώριο (margin). Και ο στόχος του SVM είναι να μεγιστοποιήσει αυτό το περιθώριο. Το υπερεπίπεδο με το μέγιστο περιθώριο (maximum margin) ονομάζεται βέλτιστο υπερεπίπεδο (optimal hyperplane).

Ακολουθεί γραφική αναπαράσταση παραδείγματος χωρισμού του δείγματος σε δύο κλάσεις (Εικόνα 3.4). Το σύνολο δεδομένων έχει δύο κλάσεις (πράσινο και μπλε). Πρέπει να ταξινομηθεί το νέο σημείο δεδομένων είτε ως μπλε είτε ως πράσινο. Δεδομένου ότι τα σημεία δεδομένων απεικονίζονται σε ένα δισδιάστατο γράφημα, αυτό το όριο απόφασης ορίζεται ως ευθεία γραμμή.



Εικόνα 3.4: Παράδειγμα SVC δύο κλάσεων

[Πηγή: <https://www.analyticsvidhya.com/blog/2021/10/support-vector-machinessvm-a-complete-guide-for-beginners/>]

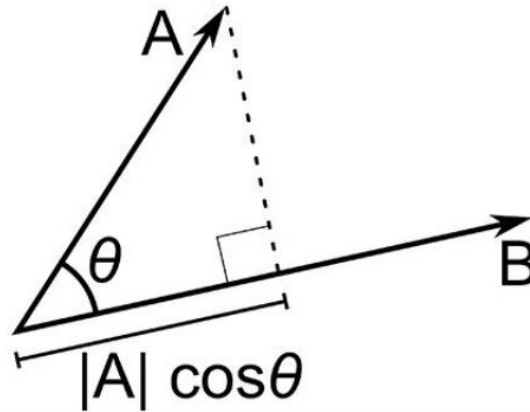
Σε περίπτωση περισσότερων διαστάσεων το όριο απόφασης ορίζεται ως υπερεπίπεδο. Το καλύτερο υπερεπίπεδο είναι εκείνο το επίπεδο που έχει τη μέγιστη απόσταση και από τις δύο κλάσεις, και αυτός είναι ο κύριος στόχος του SVM όπως αναφέρθηκε και παραπάνω. Αυτό γίνεται με την εύρεση διαφορετικών υπερεπιπέδων που ταξινομούν τις ετικέτες με τον καλύτερο τρόπο και στη συνέχεια θα επιλέξει αυτό που είναι πιο μακριά από τα σημεία δεδομένων ή αυτό που έχει το μέγιστο περιθώριο.

Ακολουθεί πλήρης επεξήγηση του μαθηματικού υπόβαθρου που βασίζεται το μοντέλο.

Ένα διάνυσμα είναι ένα μέγεθος που έχει μέτρο καθώς και κατεύθυνση και ακριβώς όπως οι αριθμοί μπορεί να χρησιμοποιηθεί σε μαθηματικές πράξεις όπως η πρόσθεση και ο πολλαπλασιασμός. Ο πολλαπλασιασμός των διανυσμάτων μπορεί να γίνει με δύο τρόπους, το εσωτερικό γινόμενο (dot product) και το εξωτερικό γινόμενο (cross product). Η ειδοποιός διαφορά τους είναι ότι το αποτέλεσμα του εσωτερικού γινομένου είναι αριθμός ενώ του εξωτερικού είναι διάνυσμα.

Το εσωτερικό γινόμενο δύο μη μηδενικών διανυσμάτων \vec{A} και \vec{B} όπως φαίνονται παρακάτω (Εικόνα 3.5) δίνεται από την ακόλουθη σχέση:

$$\vec{A} \cdot \vec{B} = |\vec{A}| |\vec{B}| \cos \theta \quad (6)$$



Εικόνα 3.5: Δύο μη μηδενικά διανύσματα A και B

Όπου,

- $|\vec{A}| \cos \theta$ είναι η προβολή του \vec{A} στο \vec{B}
- $|\vec{B}|$ είναι η απόλυτη τιμή του \vec{B}

Στο SVM αξιοποιείται μόνο η προβολή του \vec{A} . Χρησιμοποιείται το μοναδιαίο διάνυσμα του \vec{B} το οποίο έχει την κατεύθυνση του \vec{B} αλλά το μέτρο του ισούται με την μονάδα (1). Ως εκ τούτου η παραπάνω σχέση γίνεται:

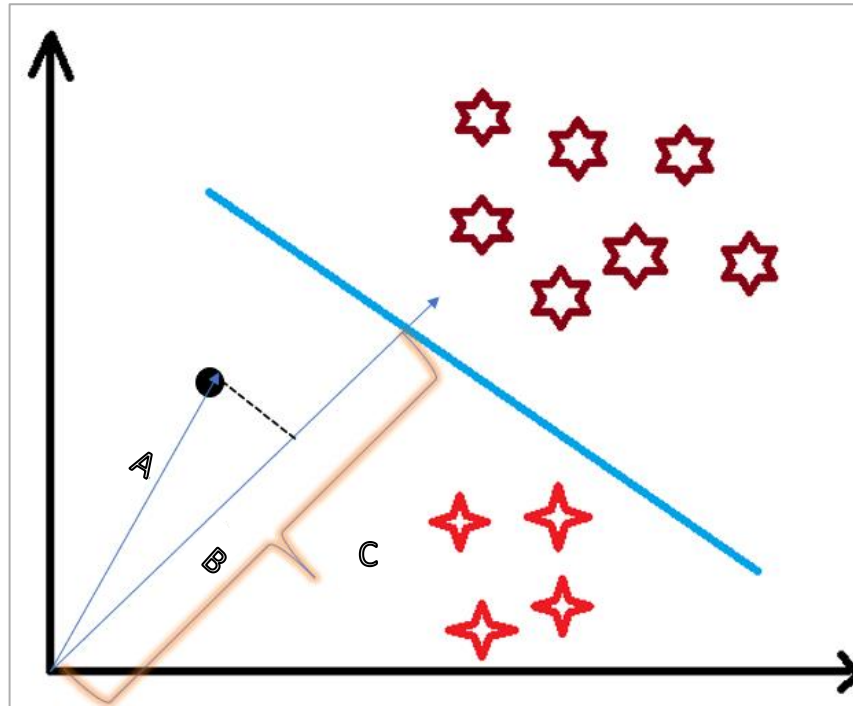
$$\vec{A} \cdot \vec{B} = |\vec{A}| \cos \theta \quad (7)$$

Αυτό αποδεικνύει ότι η προβολή οποιουδήποτε διανύσματος σε ένα άλλο διάνυσμα ισούται με το εσωτερικό τους γινόμενο.

Θεωρώντας ένα τυχαίο σημείο A και ψάχνοντας αν βρίσκεται στη δεξιά ή στην αριστερή πλευρά του επιπέδου (θετικό ή αρνητικό).

Έστω ότι το σημείο αυτό είναι ένα διάνυσμα \vec{A} , στη συνέχεια κατασκευάζεται ένα διάνυσμα \vec{B} κάθετο στο υπερεπίπεδο. Η απόσταση από την αρχή του διανύσματος \vec{B} μέχρι το όριο απόφασης (μπλε γραμμή) είναι ίση με C.

Ακολουθεί γραφική αναπαράσταση των παραπάνω (Εικόνα 3.6):



Εικόνα 3.6: Παράδειγμα SVM ταξινόμησης δύο κλάσεων

[Πηγή: <https://www.analyticsvidhya.com/blog/2021/10/support-vector-machinessvm-a-complete-guide-for-beginners/>]

- Αν ισχύει $\vec{A} \cdot \vec{B} = C$ τότε το σημείο βρίσκεται στο όριο απόφασης
- Αν ισχύει $\vec{A} \cdot \vec{B} > C$ τότε το σημείο βρίσκεται στη δεξιά πλευρά
- Αν ισχύει $\vec{A} \cdot \vec{B} < C$ τότε το σημείο βρίσκεται στην αριστερή πλευρά

Για να υπολογισθεί η απόσταση του διανύσματος \vec{A} από το όριο απόφασης χρειάζεται το διάνυσμα \vec{B} το οποίο είναι κάθετο στο όριο απόφασης και χρησιμοποιείται ως σημείο αναφοράς. Για κάθε ένα σημείο που χρειάζεται να ταξινομηθεί κατασκευάζεται η προβολή του στο κάθετο διάνυσμα και εν συνεχεία συγκρίνεται με την απόσταση C .

Για να εντοπιστεί το βέλτιστο υπερεπίπεδο διαχωρισμού πρέπει πρώτα να βρεθεί το μέγιστο περιθώριο. Αρχικά η υπερεπιφάνεια απόφασης περιγράφεται από την εξίσωση

$$w^T x + b = 0 \quad (8)$$

όπου $w = (w_1, w_2, \dots, w_i)$ το διάνυσμα βαρών που είναι κάθετο στο υπερεπίπεδο, $x = (x_1, x_2, \dots, x_i)$ το διάνυσμα εισόδου και b ο όρος μεροληψίας.

Στη συνέχεια ορίζονται δύο υπερεπίπεδα L_1 και L_2 . Όπου d είναι η απόσταση μεταξύ των δύο επιπέδων. Για να υπολογιστεί το d γίνεται η υπόθεση ότι οι εξισώσεις των L_1 και L_2 έχουν την ακόλουθη μορφή:

$$L_1 : w^T \cdot x_i + b = +1 \quad (9)$$

$$L_2 : w^T \cdot x_i + b = -1 \quad (10)$$

Άρα όπως φαίνεται και στο επόμενο σχήμα ισχύει:

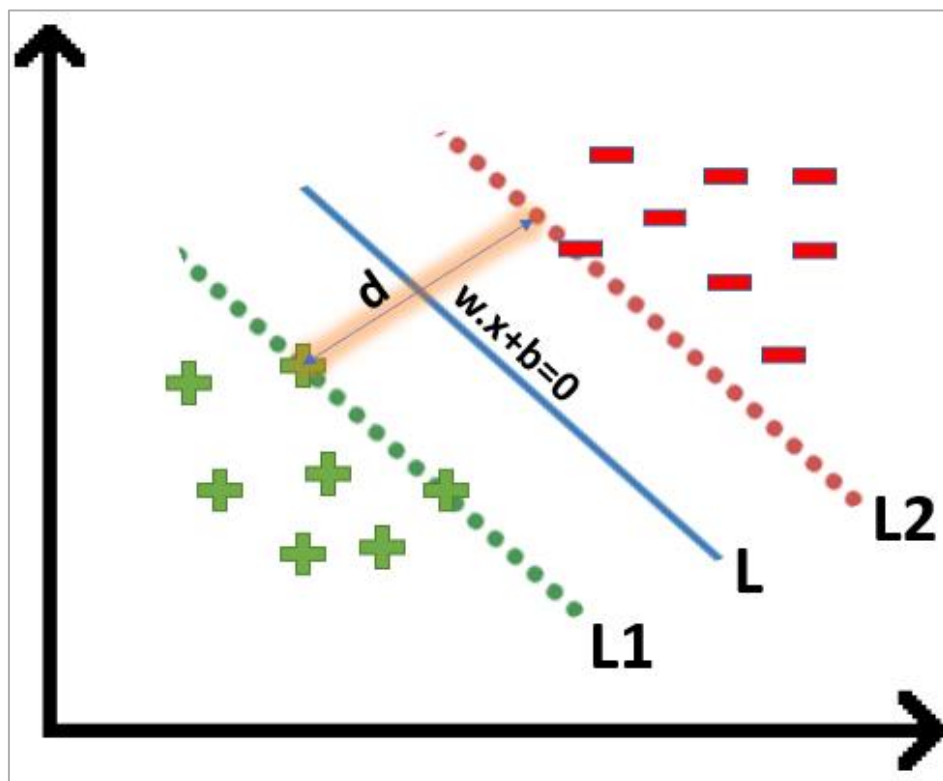
$$w^T \cdot x_i + b \geq +1 \text{ όταν } y_i = +1 \quad (11)$$

$$w^T \cdot x_i + b \leq -1 \text{ όταν } y_i = -1 \quad (12)$$

Το υπερεπίπεδο L είναι το ενδιάμεσο των L_1 και L_2 για το οποίο ισχύει:

$$L : w^T \cdot x_i + b = 0 \quad (13)$$

Το επίπεδο πρέπει να ισαπέχει από τις δύο κλάσεις δηλαδή το L πρέπει να περνάει από το κέντρο των L_1 και L_2 έτσι στηρίζεται η υπόθεση ότι τα μεγέθη ισούνται κατά απόλυτη τιμή (Εικόνα 3.7).



Εικόνα 3.7: Επίπεδα L , L_1 , L_2 και η μεταξύ τους απόσταση στο SVC

[Πηγή: <https://www.analyticsvidhya.com/blog/2021/10/support-vector-machinessvm-a-complete-guide-for-beginners/>]

Παρακάτω παρατίθενται και επεξηγούνται οι παράμετροι του μοντέλου.

c: καθορίζει τη μεροληψία (bias) του υποστηρικτικού περιθωρίου (margin) που επιτρέπεται στο μοντέλο κατά την εκπαίδευση. Μικρές τιμές του c οδηγούν σε μεγαλύτερη ευστάθεια (περισσότερη αποδοχή σφάλματος) ενώ μεγάλες τιμές του c οδηγούν σε μικρότερη μεροληψία του περιθωρίου, πιθανώς προκαλώντας overfitting στα δεδομένα εκπαίδευσης.

dual: Η παράμετρος dual καθορίζει εάν θα λύσει το πρόβλημα της εκπαίδευσης με τη διπλή (dual) ή την πρωτογενή (primal) μορφή. Η επιλογή dual=False προτιμά την πρωτογενή μορφή επίλυσης, ιδανική για περιπτώσεις όπου ο αριθμός των δειγμάτων (παρατηρήσεων) είναι μεγαλύτερος από τον αριθμό των χαρακτηριστικών (μεγάλη διαστατικότητα). Αντίθετα, η επιλογή dual=True επιλύει το πρόβλημα με τη διπλή μορφή, ιδανική για περιπτώσεις με μικρό αριθμό δειγμάτων.

Κάποια από τα πλεονεκτήματα του μοντέλου είναι τα εξής: Αποτελεί μια διαδικασία που είναι σταθερά εδραιωμένη από μαθηματικής άποψης και η λειτουργία της είναι ξεκάθαρα ερμηνεύσιμη βήμα προς βήμα. Επίσης, είναι αποτελεσματικό σε προβλήματα με υψηλή διαστατικότητα. Ακόμα, δεν είναι ευαίσθητο στις ακραίες τιμές (outliers) και ειδικεύεται στην ταξινόμηση εικόνων. Τέλος, το πλήθος των παραμέτρων που πρέπει να ρυθμιστούν στα SVMs είναι μικρότερο συγκριτικά με αρκετές αντίστοιχες μεθοδολογίες.

Όσον αφορά τα μειονεκτήματα του, δεν παρουσιάζει καλά αποτελέσματα σε ένα μεγάλο σύνολο δεδομένων. Επίσης, σε κάποια προβλήματα είναι χρονοβόρο στην εκπαίδευση. Επίσης, η παράμετρος C δεν είναι τόσο εύκολο να ρυθμιστεί λεπτομερώς και είναι δύσκολο να απεικονιστεί η επίδραση της.

3.6.5 k κοντινότεροι γείτονες

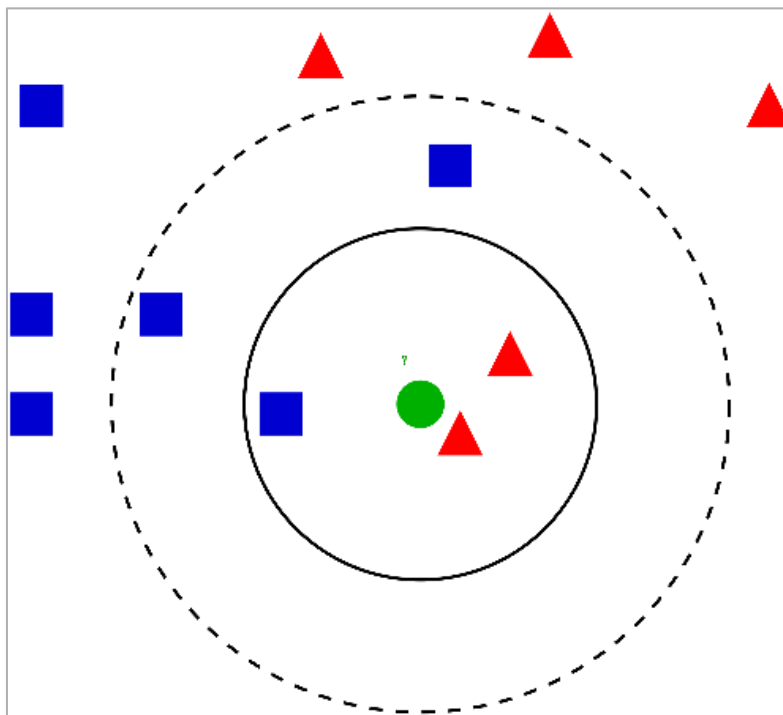
Πρόκειται για ένα δημοφιλές μοντέλο μηχανικής μάθησης που χρησιμοποιείται για εργασίες ταξινόμησης και παλινδρόμησης. Βασίζεται στην ιδέα ότι παρόμοια σημεία δεδομένων τείνουν να έχουν παρόμοιες ετικέτες ή τιμές. Δέχεται ως παραδοχή ότι τα δείγματα μπορούν να αναπαρασταθούν ως σημεία σε έναν χώρο R^n διαστάσεων, όπου n ο αριθμός των χαρακτηριστικών. Κάθε νέα περίπτωση τοποθετείται στο χώρο ως νέο σημείο και η τιμή της κλάσης προσδιορίζεται με βάση την τιμή κλάσης των k πλησιέστερων γειτονικών του σημείων. Όταν κάνει προβλέψεις, υπολογίζει την απόσταση μεταξύ του σημείου δεδομένων εισόδου και όλων των παραδειγμάτων εκπαίδευσης, χρησιμοποιώντας μια μετρική απόστασης, όπως η ευκλείδεια απόσταση. Οι πλησιέστεροι γείτονες υπολογίζονται με βάση την Ευκλείδεια απόσταση τους.

Η ευκλείδεια μετρική είναι η συνάρτηση $d: R^n \times R^n \rightarrow R$ που αντιστοιχεί σε δύο διανύσματα x, y του n -διάστατου διανυσματικού χώρου R^n , $x = \{x_1, \dots, x_n\}$, $y = \{y_1, \dots, y_n\}$ και δίνεται από τον τύπο:

$$d(x, y) = \sqrt{(y_1 - x_1)^2 + (y_2 - x_2)^2 + \dots + (y_n - x_n)^2} = \sqrt{\sum_{i=1}^n (y_i - x_i)^2} \quad (14)$$

Στη συνέχεια, ο αλγόριθμος προσδιορίζει τους K πλησιέστερους γείτονες του σημείου δεδομένων εισόδου με βάση τις αποστάσεις τους. Στην περίπτωση της ταξινόμησης, ο αλγόριθμος αναθέτει την πιο κοινή ετικέτα κλάσης μεταξύ των K γειτόνων ως την προβλεπόμενη ετικέτα για το σημείο δεδομένων εισόδου.

Στην παρακάτω εικόνα ακολουθεί ένα απλοποιημένο παράδειγμα του αλγορίθμου KNN, όπου η επιθυμητή ταξινόμηση είναι ο πράσινος κύκλος (Εικόνα 3.8). Για $K=3$ ταξινομείται με την κλάση τρίγωνο, καθώς υπάρχει μόνο ένα τετράγωνο και δύο τρίγωνα μέσα στον εσωτερικό κύκλο. Για $K=5$, αυτό ταξινομείται με την κλάση τετράγωνο, καθώς υπάρχουν δύο τρίγωνα και τρία τετράγωνα, μέσα στον εξωτερικό κύκλο.



Εικόνα 3.8: Γραφική απεικόνιση παραδείγματος KNN με $K=3$ και $K=5$.

[Πηγή: <https://www.analyticsvidhya.com/blog/2018/03/introduction-k-neighbours-algorithm-clustering/>]

Παρακάτω παρατίθενται και επεξηγούνται οι παράμετροι του μοντέλου.

Αριθμός γειτόνων ($n_neighbors$): Καθορίζει τον αριθμό των γειτόνων που θα ληφθούν υπόψη κατά την κατηγοριοποίηση ενός νέου σημείου. Επιλέγοντας μικρό αριθμό γειτόνων, το μοντέλο θα έχει τάση να αποδίδει πιο σύνθετα στοιχεία που μπορεί να οδηγήσουν σε υπερπροσαρμογή (*overfitting*), ενώ η επιλογή μεγαλύτερου αριθμού γειτόνων μπορεί να

οδηγήσει σε υποπροσαρμογή του μοντέλου που συμβαίνει όταν το μοντέλο δεν είναι αρκετά καλό για να κατανοήσει όλες τις λεπτομέρειες των δεδομένων.

Βάρη γειτόνων (weights): Καθορίζει τη μέθοδο υπολογισμού του βάρους των γειτόνων κατά τον υπολογισμό της πρόβλεψης. Στον αλγόριθμο που χρησιμοποιήσαμε, τα βάρη μπορούν να πάρουν είτε την τιμή uniform, είτε την τιμή distance. Η τιμή uniform σημαίνει ότι όλοι οι γείτονες έχουν τον ίδιο συντελεστή βαρύτητας, ενώ η τιμή distance σημαίνει ότι οι γείτονες έχουν βάρος που είναι αντιστρόφως ανάλογο με την απόστασή τους από το δείγμα που κατηγοριοποιείται.

Κάποια από τα πλεονεκτήματα του αλγόριθμου είναι τα εξής: έχει εύκολη εφαρμογή και κατανόηση, ενώ δεν απαιτείται περίοδο εκπαίδευσης. Επειδή δεν υπάρχει περίοδος εκπαίδευσης, νέα δεδομένα μπορούν να προστεθούν ανά πάσα στιγμή, αφού δεν θα επηρεάσουν το μοντέλο.

Όσον αφορά τα μειονεκτήματα του, δεν λειτουργεί καλά με μεγάλα σύνολα δεδομένων, καθώς ο υπολογισμός των αποστάσεων μεταξύ κάθε περίπτωσης δεδομένων θα ήταν πολύ δαπανηρός. Δεν αποδίδει επίσης καλά σε προβλήματα με υψηλή διαστατικότητα (dimensionality), ενώ είναι ευαίσθητος σε ελλιπή και noisy (είναι δεδομένα με μεγάλη ποσότητα πρόσθετης άσκοπης πληροφορίας που ονομάζεται θόρυβος) δεδομένα. Τέλος, ο τρόπος λήψης αποφάσεων είναι μη ευδιάκριτος.

3.6.6 Gaussian Naive Bayes Classifier

Ο ταξινομητής Gaussian Naive Bayes (NB) είναι ένα πιθανολογικό μοντέλο μηχανικής μάθησης που βασίζεται στο θεώρημα του Bayes. Υπάρχουν πέντε τύποι μοντέλων NB, στην προκειμένη χρησιμοποιείται το Gaussian NB που υιοθετεί ότι οι τιμές των χαρακτηριστικών τιμών ακολουθούν την (κανονική) κατανομή του Gauss (H. Zhang, 2004). Ο NB εξετάζει την πιθανότητα εμφάνισης κάθε κλάσης. Στη συνέχεια, συγκρίνει διαφορετικές εκ των υστέρων πιθανότητες ανάλογα με τον αριθμό των κλάσεων που υπάρχουν στα δεδομένα και η κλάση με την υψηλότερη πιθανότητα αποδίδεται στο σύνολο των χαρακτηριστικών που υπάρχουν. Τα βασικά μαθηματικά θεωρήματα που καθορίζουν τη λειτουργία του παρουσιάζονται παρακάτω.

Το Θεώρημα Bayes υπολογίζει την υπό συνθήκη πιθανότητα $P(y|x_1, \dots, x_n)$ δηλαδή την πιθανότητα να επαληθευτεί η υπόθεση y με δεδομένο ότι ισχύουν τα γεγονότα x_1, \dots, x_n . Το θεώρημα ορίστηκε μαθηματικά από την ακόλουθη εξίσωση:

$$P(y|x_1, \dots, x_n) = \frac{P(y) * P(x_1, \dots, x_n|y)}{P(x_1, \dots, x_n)} \quad (15)$$

Όπου,

- $P(x_1), \dots, P(x_n)$ είναι οι πιθανότητες των x_1, \dots, x_n αντίστοιχα. Τα x_1, \dots, x_n είναι ανεξάρτητα μεταξύ τους.
- $P(x_1, \dots, x_n | y)$ η πιθανότητα του y δεδομένου ότι το x_1, \dots, x_n είναι αληθές.

Χρησιμοποιώντας την υπόθεση της χαρακτηριστικά είναι ανεξάρτητά μεταξύ τους προκύπτει ότι:

$$P(x_i | y, x_1, \dots, x_{i-1}, \dots, x_{i+1}, \dots, x_n) = P(x_i | y) \quad (16)$$

Για ολά τα i , η σχέση αυτή απλοποιείται:

$$P(y | x_1, \dots, x_n) = \frac{P(y)}{P(X_1, \dots, X_n)} \prod_{i=1}^n P(x_i | y) \quad (17)$$

Το Gaussian NB υποθέτει ότι η πιθανότητα των χαρακτηριστικών ακολουθεί κανονική κατανομή. Η κανονική κατανομή συμβολίζεται $X \sim N(\mu, \sigma^2)$ και διαβάζεται ως τυχαία μεταβλητή X που ακολουθεί την κανονική κατανομή με συντελεστές μ και σ^2 , όπου μ ο αριθμητικός μέσος και σ η τυπική απόκλιση. Η κανονική κατανομή κατά Gauss και περιγράφεται από την ακόλουθη σχέση:

$$P(x_i | y) = \frac{1}{\sqrt{2\pi\sigma_y^2}} e^{-\frac{(x_i - \mu_y)^2}{2\sigma_y^2}} \quad (18)$$

Κάποια από τα πλεονεκτήματα του ταξινομητή είναι τα εξής: είναι εύκολο και γρήγορο να προβλέψει την κλάση του συνόλου δεδομένων δοκιμής. Αποδίδει επίσης καλά στην πρόβλεψη πολλαπλών κλάσεων. επίσης, όταν ισχύει η υπόθεση της ανεξαρτησίας, ο ταξινομητής αποδίδει καλύτερα σε σύγκριση με άλλα μοντέλα μηχανικής μάθησης, όπως η λογιστική παλινδρόμηση (logistic regression) ή το δέντρο απόφασης (decision tree), και απαιτεί λιγότερα δεδομένα εκπαίδευσης. Τέλος, αποδίδει καλύτερα στην περίπτωση κατηγορικών μεταβλητών εισόδου συγκριτικά με αριθμητικές μεταβλητές.

Όσον αφορά τα μειονεκτήματα του, υποθέτει ότι η ύπαρξη ενός συγκεκριμένου χαρακτηριστικού σε μία κλάση δεν σχετίζεται με την παρουσία οποιουδήποτε άλλου χαρακτηριστικού. Στην πραγματική ζωή, είναι σχεδόν αδύνατο να έχουμε ένα σύνολο προβλεπόμενων που να είναι εντελώς ανεξάρτητα. Ο αλγόριθμος Naive Bayes είναι επίσης γνωστός ως κακός εκτιμητής (bad estimator). Αυτό σημαίνει ότι οι πιθανότητες που προκύπτουν από τη μέθοδο predict_proba δεν πρέπει να λαμβάνονται πολύ σοβαρά. Εάν η κατηγορική μεταβλητή έχει μια κατηγορία (στο test set), η οποία δεν παρατηρήθηκε στο data set, τότε το μοντέλο θα αποδώσει πιθανότητα ίση με το 0 και δεν θα είναι σε θέση να κάνει πρόβλεψη. Αυτό είναι συχνά γνωστό ως μηδενική συχνότητα (Zero Frequency). Το πρόβλημα αυτό λύνεται με την τεχνική εξομάλυνσης. Μια από τις απλούστερες τεχνικές εξομάλυνσης ονομάζεται εκτίμηση Laplace.

3.6.7 Multi-Layer perceptron (MLP)

Ο Πολυεπίπεδος αισθητήρας (Multi-Layer perceptron) είναι ένα τροποποιημένο μοντέλο του απλού αισθητήρα (perceptron). Σε ένα τέτοιο νευρωνικό δίκτυο, μεταξύ των επιπέδων εισόδου και εξόδου, τοποθετούνται ένα ή περισσότερα επιπλέον επίπεδα, χαρακτηριζόμενα ως κρυφά επίπεδα (hidden layers). Η ροή της πληροφορίας σε ένα τέτοιο δίκτυο προχωρά πάντα προς τα εμπρός, δηλαδή, από ένα επίπεδο στο επόμενο, καθώς δεν υπάρχει αναδρομικός βρόχος. Οι νευρώνες σε κάθε επίπεδο αλληλοεπιδρούν μόνο με τους νευρώνες που ανήκουν στα άμεσα γειτονικά τους επίπεδα. Συνεπώς, το πρώτο κρυφό επίπεδο λαμβάνει τις τιμές του επιπέδου εισόδου, τα αποτελέσματα του πρώτου κρυφού επιπέδου περνάνε στο δεύτερο κρυφό, του οποίου τα αποτελέσματα στη συνέχεια περνάνε στο επίπεδο εξόδου. Το στρώμα εισόδου αποτελείται από κόμβους ή νευρώνες που λαμβάνουν τα αρχικά δεδομένα εισόδου. Στο επίπεδο εισόδου κάθε νευρώνας αντιπροσωπεύει ένα χαρακτηριστικό ή μια διάσταση των δεδομένων εισόδου. Ο αριθμός των νευρώνων στο στρώμα εισόδου καθορίζεται από τη διάσταση των δεδομένων εισόδου. Ο αριθμός των κρυφών στρωμάτων και ο αριθμός των νευρώνων σε κάθε κρυφό στρώμα είναι υπερπαραμέτροι που πρέπει να καθοριστούν κατά τη φάση σχεδιασμού του μοντέλου. Το επίπεδο εξόδου αποτελείται από νευρώνες που παράγουν την τελική έξοδο του δικτύου. Ο αριθμός των νευρώνων στο στρώμα εξόδου εξαρτάται από τη φύση της εργασίας. Στη δυαδική ταξινόμηση, μπορεί να υπάρχουν είτε ένας είτε δύο νευρώνες ανάλογα με τη συνάρτηση ενεργοποίησης (Activation Function) και να αντιπροσωπεύουν την πιθανότητα να ανήκει κανείς σε μια κλάση, ενώ σε εργασίες ταξινόμησης πολλαπλών κλάσεων, μπορεί να υπάρχουν πολλοί νευρώνες στο στρώμα εξόδου.

Τα MLP συνήθως εκπαιδεύονται με κανόνες μάθησης με επίβλεψη. Ένας αλγόριθμος που χρησιμοποιείται πολύ συχνά για το σκοπό αυτό είναι γνωστός ως αλγόριθμος Back Propagation (BP) και βασίζεται στον κανόνα μάθησης με διόρθωση σφάλματος. Η βασική ιδέα της δομής και της λειτουργίας του αλγορίθμου οπισθοδιάδοσης είναι σχετικά απλή. Το δίκτυο ξεκινά τη διαδικασία μάθησης από τυχαίες τιμές βαρών του. Εάν δώσει λάθος απάντηση τότε τα βάρη διορθώνονται έτσι ώστε το σφάλμα να γίνει μικρότερο. Η διαδικασία επαναλαμβάνεται πολλές φορές έτσι ώστε σταδιακά το σφάλμα ελαττώνεται μέχρι ότου γίνει σχετικά πολύ μικρό. Τα κρυφά στρώματα ενός MLP αποτελούνται από διασυνδεδεμένους νευρώνες που εκτελούν υπολογισμούς στα δεδομένα εισόδου. Κάθε νευρώνας σε ένα κρυφό στρώμα λαμβάνει είσοδο από όλους τους νευρώνες του προηγούμενου στρώματος. Οι είσοδοι πολλαπλασιάζονται με τα αντίστοιχα βάρη, που συμβολίζονται ως w . Τα βάρη καθορίζουν πόση επιρροή έχει η είσοδος από έναν νευρώνα στην έξοδο ενός άλλου. Εκτός από τα βάρη, κάθε νευρώνας στο κρυφό στρώμα έχει μια σχετική μεροληψία, που συμβολίζεται ως b . Για κάθε νευρώνα σε ένα κρυφό στρώμα ή στο στρώμα εξόδου, υπολογίζεται το σταθμισμένο άθροισμα (weighted sum) των εισόδων του. Αυτό περιλαμβάνει τον πολλαπλασιασμό κάθε εισόδου με το αντίστοιχο βάρος της, το άθροισμα αυτών των προϊόντων και την προσθήκη της μεροληψίας και δίνεται από την εξής σχέση:

$$\text{Weighted Sum} = \sum_{i=1}^n (w_i \cdot x_i) + b \quad (19)$$

Όπου,

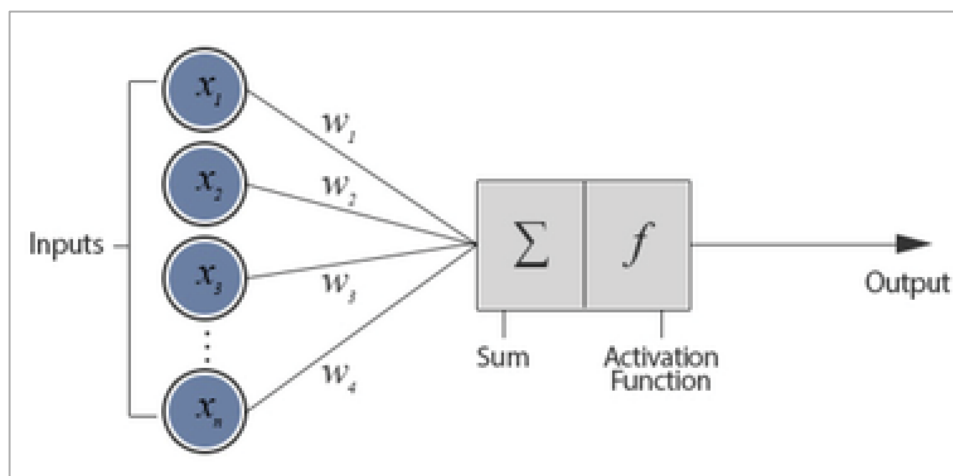
- n είναι ο συνολικός αριθμός των εισόδων
- W_i είναι το βάρος της κάθε εισόδου
- x_i είναι η τιμή της κάθε εισόδου

Το σταθμισμένο άθροισμα περνά στη συνέχεια από μια συνάρτηση που ονομάζεται συνάρτηση ενεργοποίησης. Κάθε νευρώνας στα κρυφά στρώματα και στο στρώμα εξόδου εφαρμόζει μια συνάρτηση ενεργοποίησης στο σταθμισμένο άθροισμα των εισόδων του. Αυτές οι συναρτήσεις εισάγουν μη γραμμικότητα στο δίκτυο, επιτρέποντάς του να μαθαίνει σύνθετα μοτίβα στα δεδομένα. Το μοντέλο κάθε νευρώνα στο δίκτυο περιλαμβάνει μια μη γραμμική συνάρτηση ενεργοποίησης στην έξοδο του. Αξίζει να αναφερθεί ότι η μη γραμμικότητα αυτή είναι ομαλή, δηλαδή παντού διαφορίσιμη. Μια κοινώς χρησιμοποιούμενη μορφή μη γραμμικότητας που ικανοποιεί την απαίτηση αυτή είναι η σιγμοειδής συνάρτηση ή λογιστική συνάρτηση και ορίζεται ως:

$$y_j = \frac{1}{1 + e^{-u_j}} \quad (20)$$

Όπου u_j η παράμετρος ενεργοποίησης του νευρώνα j και y_j η έξοδος του νευρώνα.

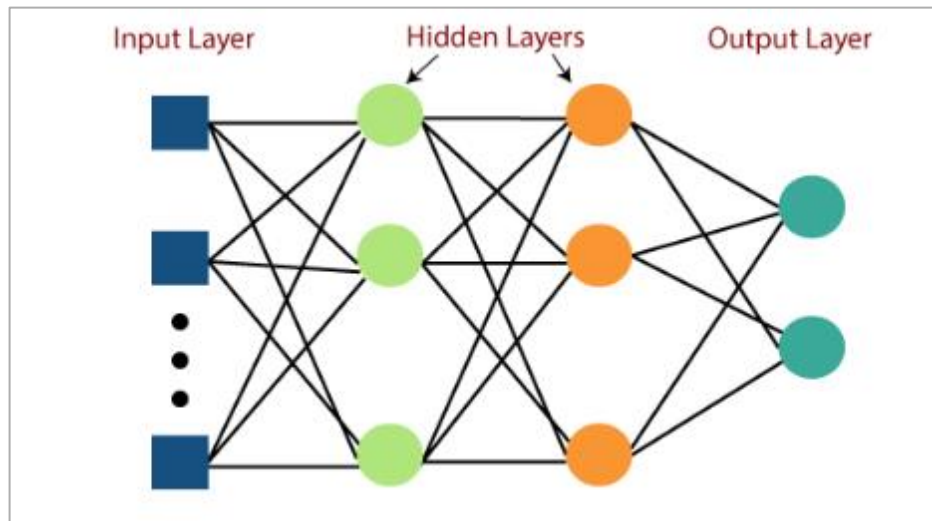
Ακολουθεί η γραφική αναπαράσταση της διαδικασίας της ενεργοποίησης ενός νευρώνα (Εικόνα 3.9).



Εικόνα 3.9: Διαδικασία ενεργοποίησης ενός τεχνητού νευρώνα

[Πηγή: <https://www.datacamp.com/tutorial/deep-learning-python>]

Η πλήρης δομή ενός MLP μοντέλου φαίνεται παρακάτω (Εικόνα 3.10)



Εικόνα 3.10: Δομή του Multi-layer Perceptron

[Πηγή: <https://www.analyticsvidhya.com/blog/2022/10/multi-layer-perceptrons-in-notations-and-trainable-parameters/>]

Παρακάτω παρατίθενται και επεξηγούνται οι παράμετροι του μοντέλου.

Αριθμός νευρώνων κρυφών επιπέδων: Καθορίζει τον αριθμό των νευρώνων σε κάθε κρυφό επίπεδο του νευρωνικού δικτύου. Για παράδειγμα, η τιμή 10 σημαίνει ένα κρυφό επίπεδο με 10 νευρώνες.

Ρυθμός μάθησης: Καθορίζει τον ρυθμό μάθησης του μοντέλου, δηλαδή πόσο γρήγορα προσαρμόζεται κατά τη διάρκεια της εκπαίδευσης. Οι επιλογές του αλγόριθμου που χρησιμοποιήθηκε περιλαμβάνουν τις τιμές constant και adaptive. Με την πρώτη ο ρυθμός μάθησης παραμένει σταθερός κατά τη διάρκεια της εκπαίδευσης ενώ με την δεύτερη ο ρυθμός μάθησης προσαρμόζεται δυναμικά κατά τη διάρκεια της εκπαίδευσης, δηλαδή ο ρυθμός μάθησης μειώνεται όσο η εκπαίδευση προχωρά, καθώς η απόδοση του μοντέλου βελτιώνεται.

Μέγιστος Αριθμός Επαναλήψεων: Καθορίζει τον μέγιστο αριθμό επαναλήψεων που θα χρησιμοποιηθούν κατά την εκπαίδευση του μοντέλου. Αν η εκπαίδευση δεν συγκλίνει πριν από την επίτευξη αυτού του αριθμού επαναλήψεων, τότε η εκπαίδευση θα σταματήσει.

Κάποια από τα πλεονεκτήματα της μεθόδου είναι ότι μπορεί να αντιμετωπίσει πολύπλοκα προβλήματα επειδή έχει πολλά επίπεδα και κόμβους και ότι μπορεί να αποδώσει σε προβλήματα ταξινόμησης πολλαπλών κλάσεων

Όσον αφορά τα μειονεκτήματά του, είναι ευαίσθητο στην υπερπροσαρμογή (overfitting) λόγω της μεγάλης του πολυπλοκότητας. Επίσης, είναι πολύ δύσκολη η εύρεση των βέλτιστων παραμέτρων για αυτό και απαιτεί πολλές δοκιμές και σφάλματα, ενώ η εκπαίδευση του μοντέλου είναι απαιτητική και χρειάζεται πολύς χρόνος και μεγάλη μνήμη.

3.7 Αξιολόγηση μοντέλων

Η αξιολόγηση ενός μοντέλου είναι μία διαδικασία δύο βημάτων. Αρχικά, το σύνολο των δεδομένων χωρίζεται σε δύο επιμέρους τμήματα. Το μεγαλύτερο ποσοστό του δείγματος χρησιμοποιείται για την εκπαίδευση/εκμάθηση του μοντέλου και ονομάζεται «Train set».

Το υπόλοιπο δείγμα χρησιμοποιείται κατά το δεύτερο βήμα που είναι η αξιολόγηση της απόδοσης του μοντέλου και ονομάζεται «Test set». Στη συγκεκριμένη περίπτωση το δείγμα χωρίστηκε σε ποσοστό 80% για την εκπαίδευση του μοντέλου και σε ποσοστό 20% για την αξιολόγηση αυτού. Αφού το μοντέλο εκπαιδευτεί στο Train set, τότε χρησιμοποιείται το Test set για να αξιολογηθεί η απόδοσή του μοντέλου, δηλαδή για να δούμε πόσο καλά προβλέπει τα νέα δεδομένα. Το Test set λειτουργεί σαν μια ανεξάρτητη επαλήθευση της απόδοσης του μοντέλου, καθώς το μοντέλο δεν έχει εκπαιδευτεί σε αυτά τα δεδομένα.

Ο πίνακας σύγχυσης (Confusion matrix) είναι ένας $N \times N$ πίνακας που χρησιμοποιείται για την αξιολόγηση της απόδοσης ενός μοντέλου ταξινόμησης, όπου N είναι ο αριθμός των κλάσεων. Υποθέτοντας ότι θέλουμε να ταξινομήσουμε ένα σύνολο δειγμάτων στις κλάσεις «θετικό» ή «αρνητικό», ο πίνακας αυτός αποτελείται από τέσσερις βασικές κατηγορίες:

- TP (True Positive): Αριθμός των στοιχείων που προβλέπονται θετικά από το μοντέλο και είναι πραγματικά θετικά.
- TN (True Negative): Αριθμός των στοιχείων που προβλέπονται αρνητικά και είναι πραγματικά αρνητικά.
- FP (False Positive): Αριθμός των στοιχείων που προβλέπονται θετικά ενώ είναι αρνητικά.
- FN (False Negative): Αριθμός των στοιχείων που προβλέπονται αρνητικά ενώ είναι θετικά.

Ακολουθεί ο πίνακας σύγχυσης για την ταξινόμηση δύο κλάσεων (Πίνακας 3.3).

Πίνακας 3.3: Πίνακας κατηγοριοποίησης δύο κλάσεων

		ΠΡΑΓΜΑΤΙΚΗ ΚΛΑΣΗ	
		ΝΑΙ	ΌΧΙ
ΠΡΟΒΛΕΠΟΜΕΝΗ ΚΛΑΣΗ	ΝΑΙ	TP	FP
	ΌΧΙ	FN	TN

Με χρήση των παραπάνω μεγεθών υπολογίζονται οι μετρικές αξιολόγησης του μοντέλου από τους παρακάτω τύπους:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (21)$$

$$Precision = \frac{TP}{TP + FP} \quad (22)$$

$$Recall = \frac{TP}{TP + FN} \quad (23)$$

$$F1 - Score = 2 * \frac{Precision * Recall}{Precision + Recall} \quad (24)$$

Η ακρίβεια (Accuracy) μπορεί να ερμηνευθεί ως μια μέτρηση για το πόσο συχνά το μοντέλο είναι σωστό στην πρόβλεψή του. Εστιάζοντας σε μια κλάση, μπορούν να εκτιμηθούν οι μετρικές Precision, Recall και F1-score.

Η μετρική ακρίβεια (Precision) υπολογίζει πόσα από τα προβλεπόμενα θετικά είναι όντως θετικά.

Η ανάκληση (Recall) υπολογίζει πόσα από τα πραγματικά θετικά εντοπίστηκαν σωστά. Δίνοντας ένα παράδειγμα αν η μετακίνηση με Μ.Μ.Μ. έχει επιλεγεί ως θετική κατηγορία, το Precision μπορεί να ερμηνευθεί ως ένα μέτρο του πόσες από τις προβλεπόμενες μετακινήσεις με δημόσια συγκοινωνία πραγματοποιήθηκαν πράγματι με δημόσια συγκοινωνία. Το Recall ερμηνεύεται ως ένα μέτρο του πόσες από τις πραγματικές μετακινήσεις με τη δημόσια συγκοινωνία προβλέφθηκαν σωστά.

Συνδυάζοντας τα δύο, το F1-score είναι ο αρμονικός μέσος όρος των Precision και Recall για τη θετική κλάση. Το ιδανικό μοντέλο θα είχε στον πίνακα όλα τα στοιχεία που δεν ανήκουν στην κύρια διαγώνιο μηδενικά. Δηλαδή θα προέβλεπε το κάθε στοιχείο στην πραγματική του κλάση.

Η καλή απόδοση ενός μοντέλου εξαρτάται από τον τύπο του προβλήματος και τις συνθήκες της εφαρμογής. Δεν υπάρχουν συγκεκριμένες τιμές που να ισχύουν γενικά για όλα τα μοντέλα και όλα τα προβλήματα. Συνήθως, όμως, η απόδοση ενός μοντέλου θεωρείται καλή όταν τα μετρικά αξιολόγησης του μοντέλου είναι πάνω από ένα κατώτατο όριο που έχει θέσει ο ερευνητής ή ο αναλυτής δεδομένων. Στην παρούσα διπλωματική ορίστηκε ίσο με 0,6.

4 Αποτελέσματα

4.1 Εισαγωγή

Στο παρόν κεφάλαιο περιγράφονται αναλυτικά τα βήματα που ακολουθήθηκαν κατά την εφαρμογή της μεθοδολογίας και παρουσιάζεται η διαδικασία ανάπτυξης των κατάλληλων μοντέλων, η οποία πραγματοποιήθηκε με τη βοήθεια του Google Colaboratory. Πρόκειται για μια υπηρεσία που παρέχεται από την Google που επιτρέπει στους χρήστες να γράφουν και να κοινοποιούν κώδικα Python μέσω του Web browser. Χρησιμοποιήθηκαν τα εξής έξι (6) μοντέλα: Random Forest, Decision Tree, K-Nearest Neighbors (KNN), Gaussian Naïve Bayes classifier (GNB), Support Vector classifier (SVC) και Multi-layer Perceptron (MLP).

Επιπρόσθετα, παρατίθενται τα αποτελέσματα που προκύπτουν από την εφαρμογή των μεθοδολογιών, η περιγραφή τους, και η ερμηνεία τους με βάση το γενικότερο πλαίσιο της έρευνας.

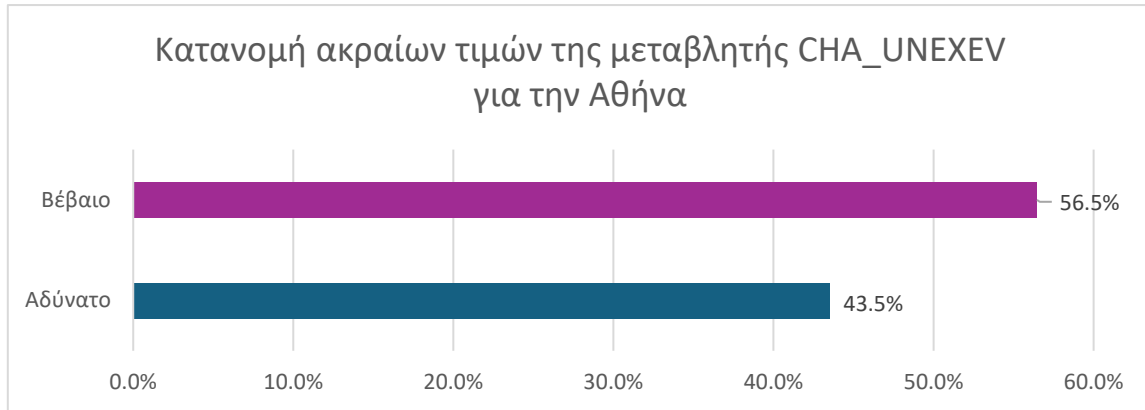
Η παρούσα διπλωματική όπως προαναφέρθηκε έχει ως σκοπό να αναδείξει τους κύριους παράγοντες που επηρεάζουν την πιθανότητα να αλλάξει κάποιος μεταφορικό μέσο υπό την επίδραση ενός απροσδόκητου συμβάντος για κάθε μία από τις τέσσερις πόλεις. Χρησιμοποιήθηκαν πέντε βάσεις δεδομένων, μία για κάθε πόλη και μία συνολική που περιείχε τις απαντήσεις και από τις τέσσερις πόλεις μαζί. Για κάθε βάση δεδομένων εκτελέστηκαν και τα έξι μοντέλα για κάθε μία από τις έξι εξαρτημένες μεταβλητές και με τους τρεις διαφορετικούς τρόπους ομαδοποίησης των απαντήσεων ώστε να μετατραπούν οι εξαρτημένες μεταβλητές σε δυαδικές. Σύνολο δηλαδή έτρεξαν 108 μοντέλα. Από την παραπάνω διαδικασία προέκυψαν δέκα (10) αξιόπιστα μοντέλα, για τα οποία ακολουθεί εκτεταμένη ανάλυση.

4.2 Πρότυπα επιλογής μέσου μετακίνησης για την Αθήνα

Στο ερωτηματολόγιο στην Αθήνα συμμετείχαν 535 μετακινούμενοι. Το πρόβλημα που διερευνήθηκε αφορά την μετακίνηση των ανθρώπων υπό την επίδραση επικίνδυνων συμβάντων και αντιστοιχεί στην εξής ερώτηση: «Ποια η πιθανότητα να αλλάξει κάποιος μεταφορικό μέσο σε περίπτωση εμφάνισης ενός επικίνδυνου συμβάντος όπως ο COVID-19;». Η εξαρτημένη μεταβλητή Y κωδικοποιείται ως CHA_UNEXEV.

Στην συγκεκριμένη περίπτωση επιλέχθηκε ο 3^{ος} τρόπος ομαδοποίησης των απαντήσεων όπου κρατήθηκαν μόνο οι ακραίες τιμές δηλαδή κωδικοποιήθηκε ως 0 η απάντηση Αδύνατο και ως 1 η απάντηση Βέβαιο.

Παρακάτω παρουσιάζεται η κατανομή των απαντήσεων αυτών. Όπως φαίνεται και παρακάτω οι συμμετέχοντες απαντούν ότι κατά 56,5% θα άλλαζαν σίγουρα το μεταφορικό τους μέσο σε περίπτωση ενός επικίνδυνου συμβάντος ενώ το 43,5% απαντά ότι αποκλείεται να το άλλαζε (Διάγραμμα 4.1).



Διάγραμμα 4.1: Κατανομή δείγματος για Αθήνα

Από τους έξι αλγορίθμους που εκτελέστηκαν αξιόπιστο θεωρήθηκε το μοντέλο K-Nearest Neighbors (μοντέλο 1). Στη συνέχεια επεξηγούνται πλήρως τα κριτήρια που θεσπίστηκαν ώστε να χαρακτηριστεί το μοντέλο αξιόπιστο.

Μοντέλο 1 - K-Nearest Neighbors

Στον παρακάτω πίνακα (Πίνακας 4.1) παρουσιάζονται οι μεταβλητές των οποίων οι απαντήσεις ομαδοποιήθηκαν. Ακόμα περιλαμβάνεται η κωδικοποίηση κάθε μεταβλητής πριν και μετά τις αλλαγές καθώς και την περιγραφή της.

Πίνακας 4.1: Κωδικοποίηση μεταβλητών του Μοντέλου της Αθήνας

Αρχικές Μεταβλητές Εισόδου	Περιγραφή	Αρχικός Τύπος	Τελικές Μεταβλητές Εισόδου	Τελικός Τύπος
TRIP_PURPOSE	Κύριος σκοπός των καθημερινών μετακινήσεων	Πολλαπλής επιλογής (4 επιλογές) [Επαγγελματικός, Εκπαιδευτικός, Προσωπικός, Αναψυχή]	TRIP_PURPOSE_1	Διαδική [Επαγγελματικός-Εκπαιδευτικός, Προσωπικός-Αναψυχή]
TIME_FLEX	Ευελιξία στην ώρα άφιξης	Πολλαπλής επιλογής (4 επιλογές) [Όχι, Έως 30 λεπτά, Έως 60 λεπτά, Πάνω από 60 λεπτά]	TIME_FLEX_1	Διαδική [Ναι-Όχι]
OCCUP	Επάγγελμα	Πολλαπλής επιλογής (6 επιλογές) [Δημόσιος υπάλληλος, Ιδιωτικός υπάλληλος, Ελεύθερος επαγγελματίας, Σε αναζήτηση εργασίας, Συνταξιούχος, Φοιτητής]	OCCUP_1	Πολλαπλής επιλογής (3 επιλογές) [Δημόσιος-Ιδιωτικός Υπάλληλος, Ελεύθερος επαγγελματίας, Φοιτητής-Συνταξιούχος-Σε αναζήτηση εργασίας]
AGE	Ηλικία	Πολλαπλής επιλογής (5 επιλογές) [18-24, 25-34, 35-44, 45-54, 65+]	AGE_1	Πολλαπλής επιλογής (3 επιλογές) [18-24, 25-64, 65+]

Στον παρακάτω πίνακα (Πίνακας 4.2) συνοψίζονται οι ανεξάρτητες μεταβλητές που χρησιμοποιήθηκαν για αυτό το μοντέλο καθώς και η περιγραφή και ο τύπος τους.

Πίνακας 4.2: Ανεξάρτητες μεταβλητές του Μοντέλου της Αθήνας

Μεταβλητές Εισόδου	Περιγραφή	Τύπος
TRIP_MODE	Συχνότερο μέσο για την καθημερινή μετακίνηση	Πολλαπλής επιλογής (5 επιλογές) [Αυτοκίνητο, Μοτοσικλέτα, Μ.Μ.Μ., Ταξί, Περπάτημα/ποδήλατο]
TRIP_COST	Κόστος μετακίνησης	Πολλαπλής επιλογής (4 επιλογές) [0, 0-5, 5-10, >10]
TRIP_PURPOSE_1	Κύριος σκοπός των καθημερινών μετακινήσεων	Διαδική [Επαγγελματικός-Εκπαιδευτικός, Προσωπικός-Αναψυχή]
INCOME	Μέσο ετήσιο εισόδημα	Πολλαπλής επιλογής (3 επιλογές) [Χαμηλό (0-10.000€), Μεσαίο (10.000€-25.000€), Υψηλό (>25.000€)]
TIME_FLEX_1	Ευελιξία στην ώρα άφιξης	Διαδική [Ναι-Όχι]
OCCUP_1	Επάγγελμα	Πολλαπλής επιλογής (3 επιλογές) [Δημόσιος-Ιδιωτικός Υπάλληλος, Ελεύθερος επαγγελματίας, Φοιτητής-Συνταξιούχος-Σε αναζήτηση εργασίας]
AGE_1	Ηλικία	Πολλαπλής επιλογής (3 επιλογές) [18-24, 25-64, 65+]

Στο κεφάλαιο 3.6.5 πραγματοποιήθηκε εκτεταμένη ανάλυση του μοντέλου KNN. Σχετικά με την επιλογή των υπερπαραμέτρων διεξήχθησαν διάφοροι πειραματισμοί οι οποίοι κατέληξαν στο ότι οι τιμές μεταξύ των οποίων πρέπει να αναζητήσει την καλύτερη το Grid search είναι οι εξής: Στην υπερπαραμέτρο που αντιστοιχεί στον αριθμό γειτόνων ($n_neighbors$) οι τρεις επιλογές που δόθηκαν ήταν το 5, το 10 και το 15 και επιλέχθηκε το 15. Όσο αυξάνεται η τιμή της παραμέτρου, τόσο το μοντέλο λαμβάνει υπόψιν περισσότερες παρατηρήσεις και μειώνει την επίδραση τυχαίων δεδομένων ή λαθών. Αν όμως αυξηθεί κατά πολύ η τιμή της τότε ελλοχεύει ο κίνδυνος να εξαχθούν αποτελέσματα επηρεασμένα κυρίως από τις γενικές τάσεις. Όσον αφορά τα βάρη γειτόνων, το Grid search είχε να επιλέξει μεταξύ της τιμής uniform και distance και επέλεξε το uniform που σημαίνει ότι όλα τα κοντινά δείγματα έχουν το ίδιο βάρος πράγμα που είναι λογικό αφού το κάθε δείγμα συνεισφέρει εξίσου στον υπολογισμό της τελικής πρόβλεψης.

Η αξιολόγησή του μοντέλου βασίστηκε στις μετρικές που επεξηγήθηκαν στο κεφάλαιο 3.7. Ακολουθεί ο πίνακας (Πίνακας 4.3) που συγκεντρώνει τις τιμές των Accuracy, Precision, Recall και F1-Score.

Πίνακας 4.3: Μετρικές αξιολόγησης του Μοντέλου 1 της Αθήνας

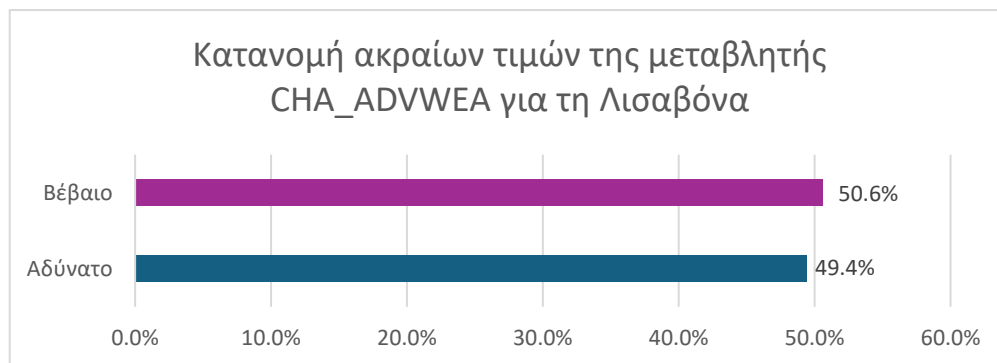
	Precision	Recall	F1-Score
Βέβαιο (1)	0,69	0,69	0,69
Αδύνατο (2)	0,64	0,64	0,64
Accuracy	0,69		

4.3 Πρότυπα επιλογής μέσου μετακίνησης για τη Λισαβόνα

Στο ερωτηματολόγιο στη Λισαβόνα συμμετείχαν 597 μετακινούμενοι. Το πρόβλημα που διερευνήθηκε αφορά την μετακίνηση των ανθρώπων κατά τη διάρκεια δυσμενών καιρικών συνθηκών και αντιστοιχεί στην εξής ερώτηση: «Ποια η πιθανότητα να αλλάξει κάποιος μεταφορικό μέσο σε περίπτωση δυσμενών καιρικών συνθηκών;». Η εξαρτημένη μεταβλητή Y είναι η CHA_ADVWEA.

Στην συγκεκριμένη περίπτωση επιλέχθηκε ο 3^{ος} τρόπος ομαδοποίησης των απαντήσεων όπου κρατήθηκαν μόνο οι ακραίες τιμές δηλαδή κωδικοποιήθηκε ως 0 η απάντηση Αδύνατο και ως 1 η απάντηση Βέβαιο.

Στο (Διάγραμμα 4.2) παρουσιάζεται η κατανομή των απαντήσεων αυτών. τους φαίνεται και παρακάτω οι συμμετέχοντες απαντούν ότι κατά 50,6% θα άλλαζαν σίγουρα το μεταφορικό τους μέσο σε περίπτωση δυσμενών καιρικών συνθηκών ενώ το 49,4% απαντά ότι αποκλείεται να το αλλάξει.



Διάγραμμα 4.2: Κατανομή δείγματος για Λισαβόνα

Από τους έξι αλγόριθμους που εκτελέστηκαν, αξιόπιστα θεωρήθηκαν τα μοντέλα K-Nearest Neighbors (μοντέλο 2) και Support Vector Classifier (μοντέλο 3).

Μοντέλο 2 - K-Nearest Neighbors

Ακολουθούν οι πίνακες 4.4 και 4.5. Στον πρώτο (Πίνακας 4.4) οι μεταβλητές των οποίων οι απαντήσεις ομαδοποιήθηκαν. Ακόμα περιλαμβάνεται η κωδικοποίηση κάθε μεταβλητής πριν και μετά τις αλλαγές καθώς και την περιγραφή της. Στον δεύτερο (Πίνακας 4.5) συνοψίζονται οι ανεξάρτητες μεταβλητές που χρησιμοποιήθηκαν για αυτό το μοντέλο καθώς και η περιγραφή και ο τύπος τους.

Πίνακας 4.4: Κωδικοποίηση μεταβλητών των Μοντέλων της Λισαβόνας

Αρχικές Μεταβλητές Εισόδου	Περιγραφή	Αρχικός Τύπος	Τελικές Μεταβλητές Εισόδου	Τελικός Τύπος
TRIP_PURPOSE	Κύριος σκοπός των καθημερινών μετακινήσεων	Πολλαπλής επιλογής (4 επιλογές) [Επαγγελματικός, Εκπαιδευτικός, Προσωπικός, Αναψυχή]	TRIP_PURPOSE_1	Διαδική [Επαγγελματικός-Εκπαιδευτικός, Προσωπικός-Αναψυχή]
TIME_FLEX	Ευελιξία στην ώρα άφιξης	Πολλαπλής επιλογής (4 επιλογές) [Όχι, Έως 30 λεπτά, Έως 60 λεπτά, Πάνω από 60 λεπτά]	TIME_FLEX_1	Διαδική [Ναι-Όχι]
OCCUP	Επάγγελμα	Πολλαπλής επιλογής (6 επιλογές) [Δημόσιος υπάλληλος, Ιδιωτικός υπάλληλος, Ελεύθερος επαγγελματίας, Σε αναζήτηση εργασίας, Συνταξιούχος, Φοιτητής]	OCCUP_1	Πολλαπλής επιλογής (3 επιλογές) [Δημόσιος-Ιδιωτικός Υπάλληλος, Ελεύθερος επαγγελματίας, Φοιτητής-Συνταξιούχος-Σε αναζήτηση εργασίας]

Πίνακας 4.5: Ανεξάρτητες μεταβλητές των Μοντέλων της Λισαβόνας

Μεταβλητές Εισόδου	Περιγραφή	Τύπος
TRIP_MODE	Συχνότερο μέσο για την καθημερινή μετακίνηση	Πολλαπλής επιλογής (5 επιλογές) [Αυτοκίνητο, Μοτοσικλέτα, Μ.Μ.Μ., Ταξί, Περπάτημα/ποδήλατο]
TRIP_COST	Κόστος μετακίνησης	Πολλαπλής επιλογής (4 επιλογές) [0, 0-5, 5-10, >10]
TRIP_PURPOSE_1	Κύριος σκοπός των καθημερινών μετακινήσεων	Διαδική [Επαγγελματικός-Εκπαιδευτικός, Προσωπικός-Αναψυχή]
INCOME	Μέσο ετήσιο εισόδημα	Πολλαπλής επιλογής (3 επιλογές) [Χαμηλό (0-10.000€), Μεσαίο (10.000€-25.000€), Υψηλό (>25.000€)]
TIME_FLEX_1	Ευελιξία στην ώρα άφιξης	Διαδική [Ναι-Όχι]
OCCUP_1	Επάγγελμα	Πολλαπλής επιλογής (3 επιλογές) [Δημόσιος-Ιδιωτικός Υπάλληλος, Ελεύθερος επαγγελματίας, Φοιτητής-Συνταξιούχος-Σε αναζήτηση εργασίας]
GENDER	Φύλλο	Διαδική [Ανδρας-Γυναίκα]

Σχετικά με την επιλογή των υπερπαραμέτρων διεξήχθησαν διάφοροι πειραματισμοί οι οποίοι κατέληξαν στο ότι οι τιμές μεταξύ των οποίων πρέπει να αναζητήσει την καλύτερη το Grid search είναι οι εξής: Για την υπερπαραμέτρο αριθμός γειτόνων οι τρεις επιλογές που δόθηκαν ήταν το 5, το 10 και το 15 και επιλέχθηκε το 15. Όσον αφορά τα βάρη γειτόνων επιλέχθηκε η τιμή uniform. Ακολουθεί ο πίνακας (Πίνακας 4.6) που συγκεντρώνει τις τιμές των Accuracy, Precision, Recall και F1-Score.

Πίνακας 4.6: Μετρικές αξιολόγησης του Μοντέλου 2 της Λισαβόνας

	Precision	Recall	F1-Score
Βέβαιο (1)	0,75	0,63	0,69
Αδύνατο (2)	0,65	0,76	0,70
Accuracy	0,69		

Μοντέλο 3 - Support Vector Classifier

Στο κεφάλαιο 3.6.4 πραγματοποιήθηκε εκτεταμένη ανάλυση του μοντέλου SVC. Για το μοντέλο αυτό χρησιμοποιήθηκαν οι ίδιες μεταβλητές με το μοντέλο 2. Σχετικά με την επιλογή των υπερπαραμέτρων διεξήχθησαν διάφοροι πειραματισμοί οι οποίοι κατέληξαν στο ότι οι τιμές μεταξύ των οποίων πρέπει να αναζητήσει την καλύτερη το Grid search είναι οι εξής: Όσον αφορά την υπερπαραμέτρο dual οι επιλογές που του δόθηκαν ήταν dual=True ή dual=False και επέλεξε το dual=True όπως συμβαίνει συνήθως σε μεγάλα σύνολα δεδομένων. Όσον αφορά την υπερπαραμέτρο C που καθορίζει την μεροληψία οι επιλογές που το δόθηκαν ήταν 0,1,2,3,5,7,10 και επιλέχθηκε η τιμή 0 όπως συμβαίνει συνήθως σε περιπτώσεις που οι κλάσεις είναι καλά διαχωρισμένες. Ακολουθεί ο πίνακας (Πίνακας 4.7) που συγκεντρώνει τις τιμές των Accuracy, Precision, Recall και F1-Score.

Πίνακας 4.7: Μετρικές αξιολόγησης του Μοντέλου 3 της Λισαβόνας

	Precision	Recall	F1-Score
Βέβαιο (1)	0,63	0,71	0,67
Αδύνατο (2)	0,68	0,60	0,64
Accuracy	0,66		

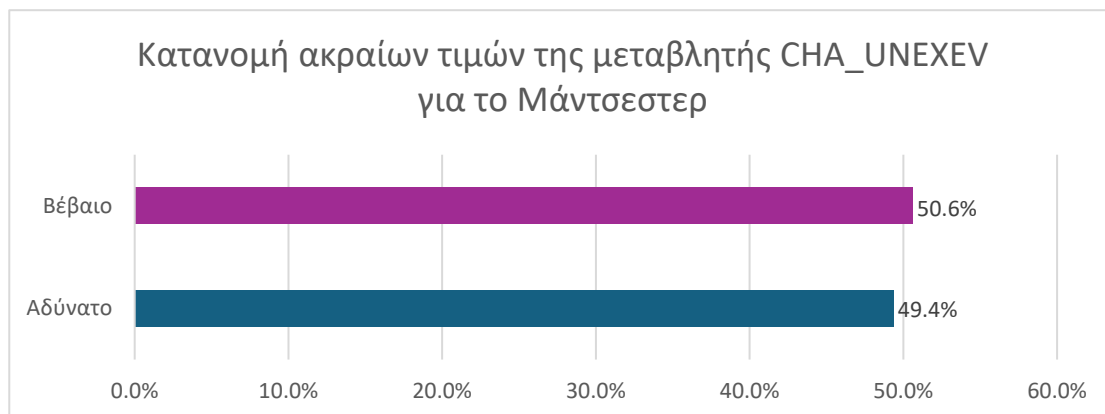
4.4 Πρότυπα επιλογής μέσου μετακίνησης για το Μάντσεστερ

Στο ερωτηματολόγιο στο Μάντσεστερ συμμετείχαν 333 μετακινούμενοι. Το πρόβλημα που διερευνήθηκε αφορά την μετακίνηση των ανθρώπων κατά τη διάρκεια επικίνδυνων

συμβάντων και αντιστοιχεί στην εξής ερώτηση: «Ποια η πιθανότητα να αλλάξει κάποιος μεταφορικό μέσο σε περίπτωση εμφάνισης ενός επικίνδυνου συμβάντος όπως ο COVID-19;». Η εξαρτημένη μεταβλητή Y είναι η CHA_UNEXEV .

Στην συγκεκριμένη περίπτωση επιλέχθηκε ο 3^{ος} τρόπος ομαδοποίησης των απαντήσεων όπου κρατήθηκαν μόνο οι ακραίες τιμές δηλαδή κωδικοποιήθηκε ως 0 η απάντηση Αδύνατο και ως 1 η απάντηση Βέβαιο.

Στο (Διάγραμμα 4.3) παρουσιάζεται η κατανομή των απαντήσεων αυτών. Όπως φαίνεται και παρακάτω οι συμμετέχοντες απαντούν ότι κατά 50,6% θα άλλαζαν σίγουρα το μεταφορικό τους μέσο σε περίπτωση επικίνδυνων συμβάντων ενώ το 49,4% απαντά ότι αποκλείεται να το αλλάξει.



Διάγραμμα 4.3: Κατανομή δείγματος για Μάντσεστερ

Από τους έξι αλγόριθμους που εκτελέστηκαν, αξιόπιστο θεωρήθηκε το μοντέλο Multi-layer Perceptron (μοντέλο 4).

Μοντέλο 4 - Multi-layer Perceptron

Ακολουθούν οι πίνακες 4.8 και 4.9. Στον πρώτο (Πίνακας 4.8) παρουσιάζονται οι μεταβλητές των οποίων οι απαντήσεις ομαδοποιήθηκαν. Ακόμα περιλαμβάνεται η κωδικοποίηση κάθε μεταβλητής πριν και μετά τις αλλαγές καθώς και την περιγραφή της. Στον δεύτερο (Πίνακας 4.9) συνοψίζονται οι ανεξάρτητες μεταβλητές που χρησιμοποιήθηκαν για αυτό το μοντέλο καθώς και η περιγραφή και ο τύπος τους.

Πίνακας 4.8: Κωδικοποίηση μεταβλητών του Μοντέλου του Μάντσεστερ

Αρχικές Μεταβλητές Εισόδου	Περιγραφή	Αρχικός Τύπος	Τελικές Μεταβλητές Εισόδου	Τελικός Τύπος
TRIP_PURPOSE	Κύριος σκοπός των καθημερινών μετακινήσεων	Πολλαπλής επιλογής (4 επιλογές) [Επαγγελματικός, Εκπαιδευτικός, Προσωπικός, Αναψυχή]	TRIP_PURPOSE_1	Διαδική [Επαγγελματικός-Εκπαιδευτικός, Προσωπικός-Αναψυχή]
TIME_FLEX	Ευελιξία στην ώρα άφιξης	Πολλαπλής επιλογής (4 επιλογές) [Όχι, Έως 30 λεπτά, Έως 60 λεπτά, Πάνω από 60 λεπτά]	TIME_FLEX_1	Διαδική [Ναι-Όχι]
OCCUP	Επάγγελμα	Πολλαπλής επιλογής (6 επιλογές) [Δημόσιος υπάλληλος, Ιδιωτικός υπάλληλος, Ελεύθερος επαγγελματίας, Σε αναζήτηση εργασίας, Συνταξιούχος, Φοιτητής]	OCCUP_1	Πολλαπλής επιλογής (3 επιλογές) [Δημόσιος-Ιδιωτικός Υπάλληλος, Ελεύθερος επαγγελματίας, Φοιτητής-Συνταξιούχος-Σε αναζήτηση εργασίας]

Πίνακας 4.9: Ανεξάρτητες μεταβλητές του Μοντέλου του Μάντσεστερ

Μεταβλητές Εισόδου	Περιγραφή	Τύπος
TRIP_MODE	Συχνότερο μέσο για την καθημερινή μετακίνηση	Πολλαπλής επιλογής (5 επιλογές) [Αυτοκίνητο, Μοτοσικλέτα, Μ.Μ.Μ., Ταξί, Περπάτημα/ποδήλατο]
TRIP_COST	Κόστος μετακίνησης	Πολλαπλής επιλογής (4 επιλογές) [0, 0-5, 5-10, >10]
TRIP_PURPOSE_1	Κύριος σκοπός των καθημερινών μετακινήσεων	Διαδική [Επαγγελματικός-Εκπαιδευτικός, Προσωπικός-Αναψυχή]
INCOME	Μέσο ετήσιο εισόδημα	Πολλαπλής επιλογής (3 επιλογές) [Χαμηλό (0-10.000€), Μεσαίο (10.000€-25.000€), Υψηλό (>25.000€)]
TIME_FLEX_1	Ευελιξία στην ώρα άφιξης	Διαδική [Ναι-Όχι]
OCCUP_1	Επάγγελμα	Πολλαπλής επιλογής (3 επιλογές) [Δημόσιος-Ιδιωτικός Υπάλληλος, Ελεύθερος επαγγελματίας, Φοιτητής-Συνταξιούχος-Σε αναζήτηση εργασίας]
GENDER	Φύλλο	Διαδική [Άνδρας-Γυναίκα]

Στο κεφάλαιο 3.6.7 πραγματοποιήθηκε εκτεταμένη ανάλυση του μοντέλου MLP. Σχετικά με την επιλογή των υπερπαραμέτρων διεξήχθησαν διάφοροι πειραματισμοί οι οποίοι κατέληξαν στο ότι οι τιμές μεταξύ των οποίων πρέπει να αναζητήσει την καλύτερη το Grid search είναι οι εξής: Για την υπερπαραμέτρο που αντιστοιχεί στον αριθμό νευρώνων κρυφών επιπέδων (`hidden_layer_sizes`) οι επιλογές που του δόθηκαν ήταν το 10, το 20 και το 50 και επέλεξε το 50. Η τιμή 50 για το μέγεθος των κρυφών επιπέδων σημαίνει ότι υπάρχει ένα κρυφό επίπεδο στο νευρωνικό δίκτυο με 50 νευρώνες οι οποίοι εκτελούν υπολογισμούς πάνω στα δεδομένα εισόδου και συμβάλλουν στη δημιουργία του μοντέλου. Μεγαλύτεροι αριθμοί νευρώνων μπορεί να οδηγήσουν σε πιο πολύπλοκα μοντέλα με μεγαλύτερη χωρητικότητα, αλλά επίσης μπορεί να αυξήσουν τον χρόνο εκπαίδευσης και την πιθανότητα υπερεκπαίδευσης (`overfitting`). Όσον αφορά την υπερπαραμέτρο ρυθμό μάθησης (`learning_rate`) οι επιλογές που του δόθηκαν ήταν `constant` και `adaptive` και επέλεξε `adaptive`. Το `adaptive` στο `learning rate` σημαίνει ότι η ταχύτητα μάθησης προσαρμόζεται κατάλληλα καθ' όλη τη διάρκεια της εκπαίδευσης, με βάση την επίδοση του μοντέλου. Για την υπερπαραμέτρο που αντιστοιχεί στο μέγιστο αριθμό επαναλήψεων (`max_iter`) οι επιλογές που του δόθηκαν ήταν το 500, το 1000 και το 2000 και επέλεξε το 500 που σημαίνει ότι θα πραγματοποιηθούν έως και 500 επαναλήψεις για την εκπαίδευση του μοντέλου. Ένας αριθμός επαναλήψεων που είναι πολύ μικρός μπορεί να μην επαρκεί για τη σύγκλιση του μοντέλου, ενώ ένας αριθμός που είναι πολύ μεγάλος μπορεί να οδηγήσει σε υπερεκπαίδευση ή προβλήματα χρόνου εκτέλεσης. Ακολουθεί ο πίνακας (Πίνακας 4.10) που συγκεντρώνει τις τιμές των Accuracy, Precision, Recall και F1-Score.

Πίνακας 4.10: Μετρικές αξιολόγησης του Μοντέλου 4 του Μάντσεστερ

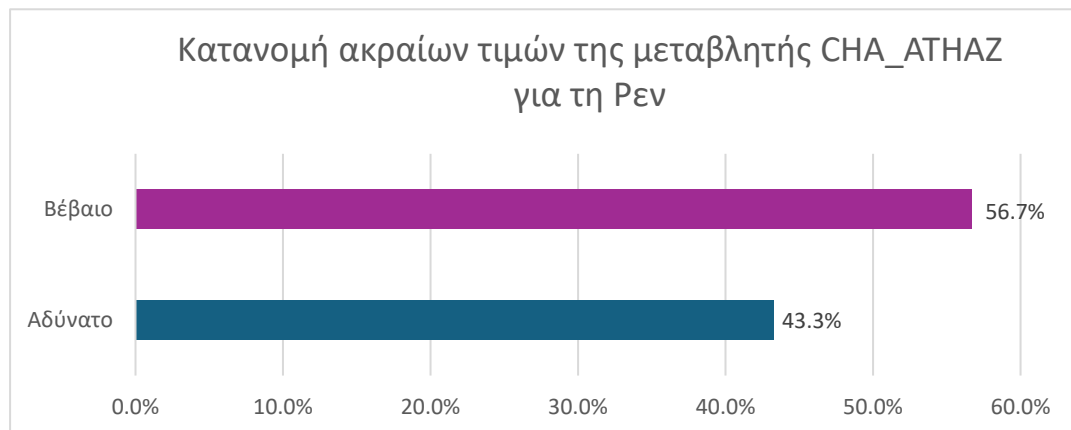
	Precision	Recall	F1-Score
Βέβαιο (1)	0,98	0,62	0,76
Αδύνατο (2)	0,73	0,98	0,83
Accuracy	0,81		

4.5 Πρότυπα επιλογής μέσου μετακίνησης για τη Ρεν

Στο ερωτηματολόγιο στη Ρεν συμμετείχαν 394 μετακινούμενοι. Το πρόβλημα που διερευνήθηκε αφορά την μετακίνηση των ανθρώπων κατά τη διάρκεια ακραίων φυσικών φαινομένων και αντιστοιχεί στην εξής ερώτηση: «Ποια η πιθανότητα να αλλάξει κάποιος μεταφορικό μέσο σε περίπτωση ακραίων καιρικών φαινομένων όπως σεισμό ή πλημμύρα;». Η εξαρτημένη μεταβλητή Y είναι η `CHA_ATHAZ`.

Στην συγκεκριμένη περίπτωση επιλέχθηκε ο 3^{ος} τρόπος ομαδοποίησης των απαντήσεων όπου κρατήθηκαν μόνο οι ακραίες τιμές δηλαδή κωδικοποιήθηκε ως 0 η απάντηση Αδύνατο και ως 1 η απάντηση Βέβαιο.

Στο (Διάγραμμα 4.4) παρουσιάζεται η κατανομή των απαντήσεων αυτών. Όπως φαίνεται και παρακάτω οι συμμετέχοντες απαντούν ότι κατά 56,7% θα άλλαζαν σίγουρα το μεταφορικό τους μέσο σε περίπτωση ακραίων φυσικών φαινομένων ενώ το 43,3% απαντά ότι αποκλείεται να το άλλαζε.



Διάγραμμα 4.4: Κατανομή του δείγματος για τη Ρεν

Από τους έξι αλγορίθμους που εκτελέστηκαν, αξιόπιστα θεωρήθηκαν τα μοντέλα K-Nearest Neighbors (μοντέλο 5), Support Vector classifier (μοντέλο 6) και Random Forest (μοντέλο 7).

Μοντέλο 5 - K-Nearest Neighbors

Στον παρακάτω πίνακα (Πίνακας 4.11) παρουσιάζονται οι μεταβλητές των οποίων οι απαντήσεις ομαδοποιήθηκαν. Ακόμα περιλαμβάνεται η κωδικοποίηση κάθε μεταβλητής πριν και μετά τις αλλαγές καθώς και την περιγραφή της.

Πίνακας 4.11: Κωδικοποίηση μεταβλητών των Μοντέλων της Ρεν

Αρχικές Μεταβλητές Εισόδου	Περιγραφή	Αρχικός Τύπος	Τελικές Μεταβλητές Εισόδου	Τελικός Τύπος
TRIP_PURPOSE	Κύριος σκοπός των καθημερινών μετακινήσεων	Πολλαπλής επιλογής (4 επιλογές) [Επαγγελματικός, Εκπαιδευτικός, Προσωπικός, Αναψυχή]	TRIP_PURPOSE_1	Διαδική [Επαγγελματικός-Εκπαιδευτικός, Προσωπικός-Αναψυχή]
TIME_FLEX	Ευελιξία στην ώρα άφιξης	Πολλαπλής επιλογής (4 επιλογές) [Όχι, Έως 30 λεπτά, Έως 60 λεπτά, Πάνω από 60 λεπτά]	TIME_FLEX_1	Διαδική [Ναι-Όχι]

Στον παρακάτω πίνακα (Πίνακας 4.12) συνοψίζονται οι ανεξάρτητες μεταβλητές που χρησιμοποιήθηκαν για αυτό το μοντέλο καθώς και η περιγραφή και ο τύπος τους.

Πίνακας 4.12: Ανεξάρτητες μεταβλητές του Μοντέλου της Ρεν

Μεταβλητές Εισόδου	Περιγραφή	Τύπος
TRIP_MODE	Συχνότερο μέσο για την καθημερινή μετακίνηση	Πολλαπλής επιλογής (5 επιλογές) [Αυτοκίνητο, Μοτοσικλέτα, Μ.Μ.Μ., Ταξί, Περπάτημα/ποδήλατο]
PEAK_HOUR	Μετακίνηση σε ώρα αιχμής	Πολλαπλής επιλογής (8 επιλογές) [Όχι, Πρωί (07:00-09:00), Μεσημέρι (12:00-14:00), Απόγευμα (17:00-19:00), Πρωί-Μεσημέρι, Πρωί-Απόγευμα, Μεσημέρι-Απόγευμα, Πρωί-Μεσημέρι-Απόγευμα]
TRIP_PURPOSE_1	Κύριος σκοπός των καθημερινών μετακινήσεων	Διαδική [Επαγγελματικός-Εκπαιδευτικός, Προσωπικός-Αναψυχή]
INCOME	Μέσο ετήσιο εισόδημα	Πολλαπλής επιλογής (3 επιλογές) [Χαμηλό (0-10.000€), Μεσαίο (10.000€-25.000€), Υψηλό (>25.000€)]
TIME_FLEX_1	Ευελιξία στην ώρα άφιξης	Διαδική [Ναι-Όχι]
OCCUP_1	Επάγγελμα	Πολλαπλής επιλογής (3 επιλογές) [Δημόσιος-Ιδιωτικός Υπάλληλος, Ελεύθερος επαγγελματίας, Φοιτητής-Συνταξιούχος-Σε αναζήτηση εργασίας]
GENDER	Φύλλο	Διαδική [Άνδρας-Γυναίκα]

Σχετικά με την επιλογή των υπερπαραμέτρων διεξήχθησαν διάφοροι πειραματισμοί οι οποίοι κατέληξαν στο ότι οι τιμές μεταξύ των οποίων πρέπει να αναζητήσει την καλύτερη το Grid search είναι οι εξής: Όσον αφορά τον αριθμό γειτόνων ($n_{\text{neighbors}}$) οι τρεις επιλογές που δόθηκαν ήταν το 5, το 10 και το 15 και επιλέχθηκε το 5 σε αντίθεση με τις δύο προηγούμενες περιπτώσεις που είχε επιλεγεί το 15. Όσο η τιμή μειώνεται τόσο το μοντέλο κινδυνεύει λιγότερο να επηρεαστεί από τιμές που είναι ακραίες και δεν αντιπροσωπεύουν τη γενική τάση. Για τα βάρη γειτόνων το Grid search είχε να επιλέξει μεταξύ της τιμής uniform και distance και επέλεξε την τιμή distance σε αντίθεση με τις άλλες δύο περιπτώσεις, αυτό σημαίνει ότι οι γείτονες συνεισφέρουν στην τελική απόφαση ανάλογα με την απόστασή τους από το νέο δείγμα.

Ακολουθεί ο πίνακας (Πίνακας 4.13) που συγκεντρώνει τις τιμές των Accuracy, Precision, Recall και F1-Score.

Πίνακας 4.13: Μετρικές αξιολόγησης του Μοντέλου 5 της Ρεν

	Precision	Recall	F1-Score
Βέβαιο (1)	0,56	0,82	0,67
Αδύνατο (2)	0,80	0,53	0,64
Accuracy	0,65		

Μοντέλο 6 - Support Vector classifier

Για το μοντέλο αυτό χρησιμοποιήθηκαν οι ίδιες μεταβλητές με το μοντέλο 5. Παρακάτω θα γίνει αναφορά στις υπερπαραμέτρους του μοντέλου: Για την υπερπαραμέτρο `dual` επιλέχθηκε η τιμή `dual=True` όπου και ερμηνεύτηκε παραπάνω. Όσον αφορά την υπερπαραμέτρο `C` μπορούσε να λάβει μία από τις ακόλουθες τιμές: 0,1,2,3,5,7,10 και επιλέχθηκε η τιμή 2. Η επιλογή μιας μεσαίας τιμής του `C`, όπως το 2, συμβάλει σε έναν καλό συμβιβασμό μεταξύ της απόδοσης και της γενίκευσης του μοντέλου.

Ακολουθεί ο πίνακας (Πίνακας 4.14) που συγκεντρώνει τις τιμές των Accuracy, Precision, Recall και F1-Score.

Πίνακας 4.14: Μετρικές αξιολόγησης του Μοντέλου 6 της Ρεν

	Precision	Recall	F1-Score
Βέβαιο (1)	0,69	0,82	0,75
Αδύνατο (2)	0,85	0,73	0,79
Accuracy	0,77		

Μοντέλο 7 – Random Forest

Στο κεφάλαιο 3.6.3 πραγματοποιήθηκε εκτεταμένη ανάλυση του μοντέλου Random Forest. Για το μοντέλο αυτό χρησιμοποιήθηκαν οι ίδιες μεταβλητές με τα μοντέλα 5 και 7. Ακολουθεί εκτενής αναφορά στις υπερπαραμέτρους του μοντέλου. Για το βάθος του δέντρου (`max_depth`) οι τρεις επιλογές που δόθηκαν στο Grid Search ήταν το 3, το 5 και το 10 και επιλέχθηκε το 3 που σημαίνει ότι το μέγιστο βάθος των δέντρων στο δάσος είναι περιορισμένο σε 3. Η τιμή αυτή επιφέρει ένα ισορροπημένο μοντέλο μεταξύ απόδοσης και γενίκευσης. Όσον αφορά την υπερπαραμέτρο που αντιστοιχεί στον ελάχιστο αριθμό δειγμάτων σε φύλλο (`min_samples_leaf`) οι τρεις επιλογές που του δόθηκαν ήταν το 1, το 2 και το 4 και επιλέχθηκε το 1 που σημαίνει ότι ο ελάχιστος αριθμός δειγμάτων που

απαιτούνται για ένα φύλλο σε ένα δέντρο του ταξινομητή Random Forest είναι. Αυτό έχει ως αποτέλεσμα το μοντέλο να έχει υψηλή διακριτική ικανότητα.

Σχετικά με την υπερπαράμετρο που αφορά τον ελάχιστο αριθμό δειγμάτων για διαχωρισμό ανά δέντρο (`min_samples_split`) οι τρεις επιλογές που του δόθηκαν ήταν το 2, το 5 και το 10 και επιλέχθηκε το 10 που σημαίνει ότι ο ελάχιστος αριθμός δειγμάτων που απαιτούνται για να διαιρεθεί ένας κόμβος σε ένα δέντρο είναι 10. Δηλαδή ένας κόμβος πριν διαχωριστεί σε περισσότερα από ένα παιδιά (`child-nodes`), θα πρέπει να υπάρχουν τουλάχιστον 10 δείγματα σε αυτόν τον κόμβο. Ο αριθμός των δέντρων (`n_neighbors`) μπορούσε να είναι 5 ή 10 ή 20 και επιλέχθηκε το 10. Συγκεκριμένα, σε αυτήν την περίπτωση, θα δημιουργηθεί ένα δάσος με 10 δέντρα. αλγόριθμος δημιουργεί τόσα δέντρα όσα και ο αριθμός που ορίζεται στην παράμετρο `n_estimators`, και καταλήγει στην τελική πρόβλεψη με βάση τον μέσο όρο των προβλέψεων των διαφόρων δέντρων. Ο αριθμός αυτός οδηγεί σε καλύτερη γενίκευση του μοντέλου, καθώς περισσότερα δέντρα εντοπίζουν διάφορες πτυχές και παραλλαγές στα δεδομένα. Ακολουθεί ο πίνακας (Πίνακας 4.15) που συγκεντρώνει τις τιμές των Accuracy, Precision, Recall και F1-Score.

Πίνακας 4.15: Μετρικές αξιολόγησης του Μοντέλου 7 της Ρεν

	Precision	Recall	F1-Score
Βέβαιο (1)	0,73	0,73	0,73
Αδύνατο (2)	0,80	0,80	0,80
Accuracy	0,77		

4.6 Πρότυπα επιλογής μέσου μετακίνησης για όλες τις πόλεις μαζί

Σε αυτήν την ενότητα δημιουργήθηκε μία βάση δεδομένων συνενώνοντας τις απαντήσεις και από τα τέσσερα (4) ερωτηματολόγια. Το ερωτηματολόγιο στην Αθήνα, στη Λισαβόνα και στο Μάντσεστερ είχε ακριβώς την ίδια δομή επομένως και οι αντίστοιχες βάσεις δεδομένων μπορούσαν να συνενωθούν χωρίς κάποια επεξεργασία. Όσον αφορά στη Ρεν σε δύο ερωτήσεις δινότουσαν διαφορετικές επιλογές στους συμμετέχοντες.

Στον παρακάτω πίνακα (Πίνακας 4.16) παρουσιάζονται οι τροποποιήσεις που απαιτούνταν να πραγματοποιηθούν ώστε να διαμορφωθεί η νέα συνολική βάση δεδομένων.

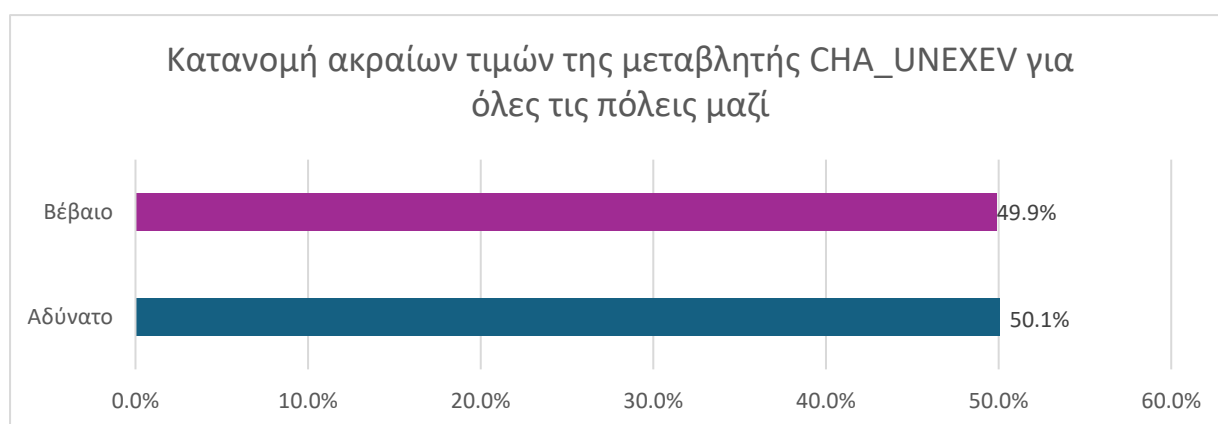
Πίνακας 4.16: Κωδικοποίηση μεταβλητών των Μοντέλων για όλες τις πόλεις μαζί

Αρχικές Μεταβλητές Εισόδου	Περιγραφή	Αρχικός Τύπος	Τελικές Μεταβλητές Εισόδου	Τελικός Τύπος
TRIP_MODE	Συχνότερο μέσο για την καθημερινή μετακίνηση	Πολλαπλής επιλογής (6 επιλογές) [Αυτοκίνητο, Μοτοσικλέτα, Μ.Μ.Μ., Ταξί, Περπάτημα, Ποδήλατο/EPDMs]	TRIP_MODE_1	Πολλαπλής επιλογής (5 επιλογές) [Αυτοκίνητο, Μοτοσικλέτα, Μ.Μ.Μ., Ταξί, Περπάτημα/ποδήλατο]
PEAK_HOUR	Μετακίνηση σε ώρα αιχμής	Πολλαπλής επιλογής (8 επιλογές) [Όχι, Πρωί (07:00-09:00), Μεσημέρι (12:00-14:00), Απόγευμα (17:00-19:00), Πρωί-Μεσημέρι, Πρωί-Απόγευμα, Μεσημέρι-Απόγευμα, Πρωί-Μεσημέρι-Απόγευμα]	PEAK_HOUR_1	Διαδική [Ναι-Όχι]

Συνολικά και στα τέσσερα (4) ερωτηματολόγια συμμετείχαν 1859 μετακινούμενοι. Το πρόβλημα που διερευνήθηκε αφορά την μετακίνηση των ανθρώπων κατά τη διάρκεια δυσμενών καιρικών συνθηκών και αντιστοιχεί στην εξής ερώτηση: «Ποια η πιθανότητα να αλλάξει κάποιος μεταφορικό μέσο σε περίπτωση εμφάνισης ενός επικίνδυνου συμβάντος όπως ο COVID-19;». Η εξαρτημένη μεταβλητή Y είναι η CHA_UNEXEV..

Στην συγκεκριμένη περίπτωση επιλέχθηκε ο 3^{ος} τρόπος ομαδοποίησης των απαντήσεων όπου κρατήθηκαν μόνο οι ακραίες τιμές δηλαδή κωδικοποιήθηκε ως 0 η απάντηση Αδύνατο και ως 1 η απάντηση Βέβαιο.

Στο (Διάγραμμα 4.5) παρουσιάζεται η κατανομή των απαντήσεων αυτών. Όπως φαίνεται και παρακάτω οι συμμετέχοντες απαντούν ότι κατά 49,9% θα άλλαζαν σίγουρα το μεταφορικό τους μέσο σε περίπτωση επικίνδυνων συμβάντων ενώ το 50,1% απαντά ότι αποκλείεται να το άλλαξε.



Διάγραμμα 4.5: Κατανομή δείγματος για όλες τις πόλεις μαζί

Από τους έξι αλγορίθμους που εκτελέστηκαν αξιόπιστα θεωρήθηκαν τα μοντέλα Support Vector classifier (μοντέλο 8), Random Forest (μοντέλο 9) και Multi-layer Perceptron (μοντέλο 10).

Μοντέλο 8 - Support Vector classifier

Ακολουθεί ο πίνακας (Πίνακας 4.17) στον οποίο παρουσιάζονται οι ανεξάρτητες μεταβλητές που χρησιμοποιήθηκαν σε αυτό το μοντέλο.

Πίνακας 4.17: Ανεξάρτητες μεταβλητές των Μοντέλων για όλες τις πόλεις μαζί

Μεταβλητές Εισόδου	Περιγραφή	Τύπος
TRIP_MODE	Συχνότερο μέσο για την καθημερινή μετακίνηση	Πολλαπλής επιλογής (5 επιλογές) [Αυτοκίνητο, Μοτοσικλέτα, Μ.Μ.Μ., Ταξί, Περπάτημα/ποδήλατο]
PEAK_HOUR_1	Μετακίνηση σε ώρα αιχμής	Διαδική [Ναι-Όχι]
TRIP_PURPOSE_1	Κύριος σκοπός των καθημερινών μετακινήσεων	Διαδική [Επαγγελματικός-Εκπαιδευτικός, Προσωπικός-Αναψυχή]
INCOME	Μέσο ετήσιο εισόδημα	Πολλαπλής επιλογής (3 επιλογές) [Χαμηλό (0-10.000€), Μεσαίο (10.000€-25.000€), Υψηλό (>25.000€)]
AGE_1	Ηλικία	Πολλαπλής επιλογής [18-24, 25-34, 35-44, 45-54, 65+]
OCCUP_1	Επάγγελμα	Πολλαπλής επιλογής (3 επιλογές) [Δημόσιος-Ιδιωτικός Υπάλληλος, Ελεύθερος επαγγελματίας, Φοιτητής-Συνταξιούχος-Σε αναζήτηση εργασίας]
TRIP_COST	Κόστος μετακίνησης	Πολλαπλής επιλογής (4 επιλογές) [0, 0-5, 5-10, >10]

Ακολουθεί αναφορά στις υπερπαραμέτρους του μοντέλου. Όσον αφορά την υπερπαραμέτρο `max_depth` επιλέχθηκε η τιμή `max_depth=True` όπου και ερμηνεύτηκε παραπάνω. Η υπερπαραμέτρος `C` που αφορά την μεροληψία μπορούσε να λάβει μία από τις ακόλουθες τιμές: 1,2,3,5,7 ή 10 και επιλέχθηκε η τιμή 5 που αντιστοιχεί σε ένα μέτριο επίπεδο κόστους. Αυτό σημαίνει ότι το μοντέλο είναι αρκετά ευαίσθητο στις εσφαλμένες ταξινομήσεις, προσπαθώντας να μειώσει τέτοιου είδους λάθη, αλλά ταυτόχρονα προσπαθεί να επιτύχει και καλή γενίκευση. Ακολουθεί ο πίνακας (Πίνακας 4.18) που συγκεντρώνει τις τιμές των Accuracy, Precision, Recall και F1-Score.

Πίνακας 4.18: Μετρικές αξιολόγησης του Μοντέλου 8 για όλες τις πόλεις μαζί

	Precision	Recall	F1-Score
Βέβαιο (1)	0,63	0,71	0,67
Αδύνατο (2)	0,67	0,60	0,63
Accuracy	0,65		

Μοντέλο 9 - Random Forest

Για το μοντέλο αυτό χρησιμοποιήθηκαν οι ίδιες μεταβλητές με το μοντέλο 8. Παρακάτω θα γίνει αναφορά στις υπερπαραμέτρους του μοντέλου. Όσον αφορά το βάθος του δέντρου (`max_depth`) οι τρεις επιλογές που δόθηκαν στο Grid Search ήταν 3 ή 5 ή 10 και επιλέχθηκε το 5 σε αντίθεση με το προηγούμενο random forest όπου ήταν 3 που σημαίνει ότι τα δέντρα αναπτύσσονται σε μεγαλύτερο βάθος, περιλαμβάνοντας περισσότερες συνθήκες και αποφάσεις για τη διαχωριστική τους ικανότητα. Για την υπερπαραμέτρο που αντιστοιχεί στον ελάχιστο αριθμό δειγμάτων σε φύλλο (`min_samples_leaf`) οι τρεις επιλογές που του δόθηκαν ήταν το 1, το 2 και το 4 και επιλέχθηκε το 2 σε αντίθεση με το προηγούμενο random forest όπου ήταν 1. Αυτό επιτρέπει πιο λεπτομερή δέντρα, θα υπήρχε κίνδυνος για υπερπροσαρμογή του μοντέλου αν το σύνολο δεδομένων δεν ήταν τόσο μεγάλο. Η υπερπαραμέτρος που αφορά τον ελάχιστο αριθμό δειγμάτων για διαχωρισμό ανά δέντρο (`min_samples_split`) μπορούσε να λάβει την τιμή 2 ή 5 ή 10 και επιλέχθηκε το 2 σε αντίθεση με το προηγούμενο random forest όπου ήταν 10. Η αλλαγή αυτή οδηγεί σε πιο σύνθετα δέντρα. Όσον αφορά την υπερπαραμέτρο που αντιστοιχεί στον αριθμό των δέντρων (`n_estimators`) οι τρεις επιλογές που δόθηκαν ήταν το 5, το 10 και το 20 και επιλέχθηκε το 20 σε αντίθεση με το προηγούμενο random forest όπου ήταν 10. Αυτή η αύξηση σημαίνει ότι το μοντέλο θα έχει μεγαλύτερη ποικιλία στις προβλέψεις του καθώς θα λαμβάνει τιμές από περισσότερα δέντρα. Επίσης με περισσότερα δέντρα, η μέση πρόβλεψη από το δάσος είναι πιο σταθερή και λιγότερο επηρεασμένη από τυχαίες διακυμάνσεις στα δεδομένα. Ακολουθεί ο πίνακας (Πίνακας 4.19) που συγκεντρώνει τις τιμές των Accuracy, Precision, Recall και F1-Score για τις δύο κλάσεις.

Πίνακας 4.19: Μετρικές αξιολόγησης του Μοντέλου 9 για όλες τις πόλεις μαζί

	Precision	Recall	F1-Score
1	0,61	0,71	0,66
2	0,68	0,60	0,64
Accuracy	0,63		

Μοντέλο 10 - Multi-layer Perceptron

Για το μοντέλο αυτό χρησιμοποιήθηκαν οι ίδιες μεταβλητές με το μοντέλο 8. Παρακάτω θα γίνει αναφορά στις παραμέτρους του μοντέλου. Όσον αφορά τον αριθμό νευρώνων κρυφών επιπέδων (*hidden_layer_sizes*) οι επιλογές που δόθηκαν στο Grid Search ήταν το 10, το 20 και το 50 και επέλεξε το 20 σε αντίθεση με το προηγούμενο Multi-Layer perceptron όπου ήταν 50. Αυτό σημαίνει ότι ενισχύεται η αποφυγή υπερεκπαίδευσης (*overfitting*). Με ένα μικρότερο αριθμό νευρώνων, το μοντέλο θα είναι λιγότερο πολύπλοκο και θα έχει λιγότερη τάση να μάθει τυχαίες λεπτομέρειες των δεδομένων. Όσον αφορά τον ρυθμό μάθησης (*learning_rate*) του δόθηκαν οι επιλογές *constant* και *adaptive* και επέλεξε την τιμή *adaptive* όπως και στο προηγούμενο MLP. Η σημασία της τιμής αυτής ερμηνεύτηκε προηγουμένως. Όσον αφορά τον μέγιστο αριθμό επαναλήψεων (*max_iter*) οι επιλογές που του δόθηκαν ήταν το 500, το 1000 και το 2000 και επέλεξε το 500 όπως και στο προηγούμενο MLP. Η σημασία της τιμής αυτής ερμηνεύτηκε προηγουμένως. Ακολουθεί ο πίνακας (Πίνακας 4.20) που συγκεντρώνει τις τιμές των Accuracy, Precision, Recall και F1-Score για τις δύο κλάσεις.

Πίνακας 4.20: Μετρικές αξιολόγησης του Μοντέλου 10 για όλες τις πόλεις μαζί

	Precision	Recall	F1-Score
1	0,62	0,69	0,66
2	0,66	0,60	0,63
Accuracy	0,64		

4.7 Συγκριτική αξιολόγηση μοντέλων

Παραπάνω παρουσιάστηκαν δέκα μοντέλα. Το πληρέστερο ήταν το μοντέλο 8 δηλαδή το Support Vector Classifier που αφορούσε τη βάση δεδομένων που απαρτίζεται και από τις τέσσερις πόλεις μαζί. Το μοντέλο αυτό έδωσε απάντηση στην εξής ερώτηση: «Ποια η πιθανότητα να αλλάξει κάποιος μεταφορικό μέσο σε περίπτωση εμφάνισης ενός επικίνδυνου συμβάντος όπως ο COVID-19;». Υπήρχαν μοντέλα με πολύ καλύτερες μετρικές όπως το 4, το 6 και το 7 αλλά θεωρήθηκε το καλύτερο κυρίως λόγω του μεγέθους της βάσης δεδομένων που ήταν περίπου τετραπλάσια από τις υπόλοιπες. Επίσης αναλύθηκαν αναλυτικά οι λόγοι για τους οποίους και οι παράμετροι του ήταν πολύ καλές.

Από τις έξι (6) ερωτήσεις που τέθηκαν στην συγκεκριμένη μελέτη προέκυψαν αξιόπιστα μοντέλα για τις τρεις. Οι τρεις (3) αυτές ερωτήσεις αντιστοιχούν στις παρακάτω εξαρτημένες μεταβλητές Y. Πρώτη είναι η CHA_ADVWEA (δυσμενείς καιρικές συνθήκες) ακολουθεί η CHA_ATHAZ (ακραία καιρικά φαινόμενα όπως σεισμός πλημμύρα) και τέλος η CHA_UNEXEV (επικίνδυνα συμβάντα όπως COVID-19).

Στα παραπάνω μοντέλα φαίνεται ότι υπάρχουν κάποιες μεταβλητές που περιέχονται σε όλα άρα είναι και οι σημαντικότερες. Σημαντικότερη θεωρείται το μέσο μετακίνησης (TRIP_MODE) και ακολουθούν το εισόδημα (INCOME), το κόστος της μετακίνησης (TRIP_COST), ο σκοπός της μετακίνησης (TRIP_PURPOSE) και το επάγγελμα (OCCUP).

Ακολουθούν δύο πίνακες, στον πρώτο (Πίνακας 4.21) παρουσιάζεται η κατανομή των μοντέλων ανά πόλη για κάθε μία ερώτηση που επιλύθηκε ενώ στον δεύτερο (Πίνακας 4.22) παρατίθενται οι μεταβλητές που χρησιμοποιήθηκαν σε κάθε ένα από τα δέκα μοντέλα.

Πίνακας 4.21: Συγκεντρωτικός πίνακας μοντέλων ανά πρόβλημα που επιλύθηκε σε κάθε πόλη

Πρόβλημα: Επικίνδυνα συμβάντα όπως ο COVID-19 (CHA_UNEXEV)				
Μοντέλα Πόλεις	KNN	SVM	MLP	RF
Αθήνα	✓			
Λίσαβονα				
Μάντσεστερ			✓	
Ρεν				
Όλες μαζί		✓	✓	✓
Πρόβλημα: Δυσμενές καιρικές συνθήκες (CHA_ADVWEA)				
Μοντέλα Πόλεις	KNN	SVM	MLP	RF
Αθήνα				
Λίσαβονα	✓	✓		
Μάντσεστερ				
Ρεν				
Όλες μαζί				
Πρόβλημα: Ακραία καιρικά φαινόμενα όπως σεισμοί και πλημμύρες (CHA_ATHAZ)				
Μοντέλα Πόλεις	KNN	SVM	MLP	RF
Αθήνα				
Λίσαβονα				
Μάντσεστερ				
Ρεν	✓	✓		✓
Όλες μαζί				

Πίνακας 4.22: Συγκεντρωτικός πίνακας μοντέλων σε σχέση με τις ανεξάρτητες μεταβλητές που χρησιμοποιήθηκαν

Μεταβλητές Μοντέλα	TRIP_MODE	TRIP_COST	INCOME	TRIP_PURPOSE	TIME_FLEX	OCCUP	AGE	GENDER	PEAK_HOUR
M1	✓	✓	✓	✓	✓	✓	✓		
M2	✓	✓	✓	✓	✓	✓		✓	
M3	✓	✓	✓	✓	✓	✓		✓	
M4	✓	✓	✓	✓	✓	✓		✓	
M5	✓		✓	✓	✓	✓		✓	✓
M6	✓		✓	✓	✓	✓		✓	✓
M7	✓		✓	✓	✓	✓		✓	✓
M8	✓	✓	✓	✓		✓	✓		✓
M9	✓	✓	✓	✓		✓	✓		✓
M10	✓	✓	✓	✓		✓	✓		✓

5 Συμπεράσματα

5.1 Γενικά

Αντικείμενο της παρούσας διπλωματικής αποτέλεσε η διερεύνηση των παραγόντων που επηρεάζουν την αλλαγή του μέσου μετακίνησης στην περίπτωση εμφάνισης απροσδόκητων συμβάντων στην Αθήνα, στο Μάντσεστερ, στη Ρεν και στη Λισαβόνα. Τα απροσδόκητα συμβάντα που μελετήθηκαν ήταν έξι (6) και ήταν τα εξής: i) Δυσμενείς καιρικές συνθήκες, ii) ακραία καιρικά φαινόμενα (π.χ. σεισμός, πλημμύρες), iii) έντονη κυκλοφοριακή κίνηση, iv) απρογραμμάτιστη διακοπή των Μ.Μ.Μ., v) προγραμματισμένες εκδηλώσεις (π.χ. οδικά έργα, ποδοσφαιρικός αγώνας), vi) επικίνδυνα συμβάντα (π.χ. COVID-19). Συγκεκριμένα το πρόβλημα που μελετήθηκε αφορά τη μετακίνηση των ανθρώπων κατά τη διάρκεια επικίνδυνων συμβάντων και αντιστοιχεί στην εξής ερώτηση: «Ποια η πιθανότητα να αλλάξει κάποιος μεταφορικό μέσο σε περίπτωση εμφάνισης κάποιου από τα έξι προαναφερθέντα απροσδόκητα συμβάντα;». Τα αποτελέσματα που εξήχθησαν κατά τη μελέτη συνδέονται άμεσα με το σκοπό της διπλωματικής εργασίας ενώ αποτέλεσαν βασικό εργαλείο για την εξαγωγή χρήσιμων συμπερασμάτων.

Στο κεφάλαιο αυτό παρουσιάζονται και σχολιάζονται διεξοδικά τα βασικά συμπεράσματα που εξήχθησαν από το σύνολο της μελέτης. Τέλος, παρατίθενται προτάσεις για περαιτέρω έρευνα στο θέμα της διπλωματικής εργασίας.

5.2 Βασικά συμπεράσματα

Από το σύνολο των μοντέλων που έτρεξαν προέκυψαν δέκα (10) αξιόπιστα μοντέλα. Κοινό τους χαρακτηριστικό είναι ότι σε όλα τα μοντέλα χρησιμοποιήθηκαν οι μεταβλητές που αντιστοιχούν στο μέσο μετακίνησης (TRIP_MODE), στο εισόδημα (INCOME) και στο σκοπό της μετακίνησης (TRIP_PURPOSE). Όσον αφορά το μέσο μετακίνησης είναι η σημαντικότερη μεταβλητή διότι ανάλογα με το μέσο που χρησιμοποιεί ένας άνθρωπος τον επηρεάζει ή όχι η εμφάνιση ενός απροσδόκητου συμβάντος (π.χ. ο χρήστης Ι.Χ. δεν επηρεάζεται από την απρογραμμάτιστη διακοπή των Μ.Μ.Μ. αλλά αντίθετα επηρεάζεται από την αυξημένη κυκλοφοριακή συμφόρηση). Η βαρύτητα του εισοδήματος στην επιλογή μεταφορικού μέσου είναι σημαντική και επηρεάζει τις αποφάσεις των ανθρώπων για πολλούς λόγους. Για παράδειγμα ένας άνθρωπος που μετακινείται συνήθως πεζός ή με ποδήλατο σε περίπτωση εμφάνισης καταιγίδας καλείται να επιλέξει κάποιο άλλο μέσο μεταφοράς. Οι επιλογές του είναι να επιλέξει τα Μ.Μ.Μ. ή ταξί με το κόστος του δεύτερου να είναι πολλαπλάσιο του πρώτου.

Όπως έχει αναφερθεί και προηγουμένως ο σκοπός της μετακίνησης μπορεί να είναι επαγγελματικός, εκπαιδευτικός, προσωπικός και αναψυχής. Ο σκοπός της μετακίνησης καθορίζεται από τις ανάγκες και τις προτεραιότητες του ατόμου. Εάν η ανάγκη για την εκτέλεση του σκοπού είναι κρίσιμη και επείγουσα τότε το άτομο ενδέχεται να αλλάξει μέσο μεταφοράς για να προσαρμοστεί στην κατάσταση (π.χ. εάν ο σκοπός της μετακίνησης είναι η επίσκεψη σε ένα γιατρό τότε σε περίπτωση ενός προγραμματισμένου γεγονότος πλησίον του ιατρείου όπως είναι ένας ποδοσφαιρικός αγώνας είναι πολύ πιο πιθανό κάποιος να αφήσει το αυτοκίνητο του και να επιλέξει το μετρό ώστε να μην χάσει το ραντεβού). Σε κάθε άλλη περίπτωση μπορεί να αναβάλει τη μετακίνηση του. Ο σκοπός της μετακίνησης συνδέεται άμεσα και με τα χρονικά περιθώρια (π.χ. σε περίπτωση καθυστέρησης ενός λεωφορείου είναι διαφορετικό αν κάποιος κατευθύνεται προς συνέντευξη για την αναζήτηση εργασίας και διαφορετικό αν πηγαίνει για να ψωνίσει).

Από τα έξι ερωτήματα που τέθηκαν προέκυψαν αξιόπιστα μοντέλα για τα τρία. Πιο συγκεκριμένα για την περίπτωση των επικίνδυνων συμβάντων (π.χ. COVID-19), των δυσμενών καιρικών συνθηκών καθώς και των ακραίων καιρικών φαινομένων (π.χ. σεισμός, πλημμύρες). Τα πέντε από τα δέκα μοντέλα αφορούν τα επικίνδυνα συμβάντα. Είναι ένα λογικό αποτέλεσμα αφού ο COVID-19 συνεχίζει να επηρεάζει σημαντικά τις μετακινήσεις των ανθρώπων σε παγκόσμιο επίπεδο και μετά το πέρας της πανδημίας. Από τα έξι απροσδόκητα συμβάντα που μελετήθηκαν αυτό των επικίνδυνων συμβάντων είναι το μόνο που έχει προεκτάσεις και στην συναισθηματική φόρτιση των ανθρώπων καθώς είναι πολύ πρόσφατες οι μνήμες των ανθρώπων από τις επιπτώσεις του COVID-19. Η ανάλυση των παραπάνω βασίστηκε σε ερωτηματολόγια δεδηλωμένης προτίμησης δηλαδή βασίστηκε στις απαντήσεις των συμμετεχόντων και όχι σε παρατηρήσεις στο πεδίο. Όπως είναι αναμενόμενο οι άνθρωποι στην ερώτηση αυτή απαντούν βασιζόμενοι στη μεγάλη τους εμπειρία λόγω της πανδημίας ενώ σε άλλες ερωτήσεις όπως αυτή που αφορά τα ακραία καιρικά φαινόμενα (π.χ. σεισμός, πλημμύρα) απαντούν περισσότερο με το ένστικτο μιας και είναι πολύ μικρή η πιθανότητα εμφάνισης ενός τέτοιου γεγονότος ώστε να στηριχθούν στο πως έπραξαν στο παρελθόν. Όσον αφορά τη μεταβλητή CHA_PTDIS (απρογραμματίστη διακοπή των Μ.Μ.Μ.) παρέδωσε πολύ καλά μοντέλα μόνο όταν περιλαμβανόταν στο μοντέλο ως ανεξάρτητη μεταβλητή η TRIP_MODE (μέσο μετακίνησης). Για τον λόγο ότι η χρησιμοποίηση της οδηγεί στο προφανές συμπέρασμα ότι π.χ. αν κάποιος είναι χρήστης των Μ.Μ.Μ. και επέλθει μία βλάβη σε αυτά τότε αναγκαστικά θα χρειαστεί να αλλάξει μέσο μετακίνησης ενώ αν μετακινείται με το Ι.Χ. ή με τη μοτοσικλέτα του δεν θα χρειαστεί να αλλάξει. Για τους παραπάνω λόγους αποφασίστηκε τα μοντέλα αυτά να μην περιληφθούν στη διπλωματική.

5.3 Προτάσεις για περαιτέρω έρευνα

Σύμφωνα με τα αποτελέσματα και τα συνολικά συμπεράσματα που εξήχθησαν κατά την εκπόνηση της εργασίας αυτής, επιχειρείται η παράθεση μιας σειράς προτάσεων, οι οποίες ενδεχομένως να συμβάλουν στην καλύτερη κατανόηση της ταξιδιωτικής συμπεριφοράς και των λόγων που επηρεάζουν την επιλογή μεταφορικού μέσου κατά την εμφάνιση ενός απροσδόκητου συμβάντος.

Οι μελλοντικές έρευνες μπορούν να εξετάσουν το θέμα αυτό βασιζόμενες και στη μέθοδο της αποκαλυφθείσας προτίμησης (revealed preference). Με αυτόν τον τρόπο θα περιοριστεί το πρόβλημα που εμφανίζεται στη μέθοδο της δεδηλωμένης προτίμησης, δηλαδή το γεγονός ότι στις έρευνες αυτές, πολλές φορές, οι ερωτώμενοι άλλο δηλώνουν- στο υποθετικό σενάριο που τους παρατίθεται- και άλλο πράττουν.

Ενδιαφέρον θα ήταν επίσης να διεξαχθεί μια έρευνα στην οποία να ερευνηθεί ο τρόπος με τον οποίο οι πολιτικές αποφάσεις και οι παρεμβάσεις επηρεάζουν την επιλογή μέσου μετακίνησης κατά την εμφάνιση απροσδόκητων συμβάντων δηλαδή πώς οι κυβερνητικές αποφάσεις, όπως οι περιορισμοί κυκλοφορίας ή οι δημόσιες ενημερωτικές καμπάνιες, επηρεάζουν τις επιλογές μετακίνησης του κοινού. Στη συνέχεια μπορεί να γίνει σύγκριση μεταξύ των τεσσάρων ευρωπαϊκών πόλεων ώστε να αναδειχθεί σε τι βαθμό επηρεάζουν οι κυβερνητικές αποφάσεις την επιλογή μέσου μετακίνησης των πολιτών. Ακόμα, θα μπορούσε να διερευνηθεί πως οι ειδικές συνθήκες που επικρατούν σε κάθε πόλη επηρεάζουν την επιλογή μέσου από τους πολίτες. Μπορούν να εξεταστούν οι υποδομές, η γεωγραφία και οι κλιματολογικές συνθήκες που επικρατούν σε κάθε πόλη. Για παράδειγμα στην Αθήνα, όπου που υπάρχει μεγάλος αριθμός γηπέδων εντός του αστικού ιστού, είναι πιθανό να παρατηρηθεί αύξηση της κυκλοφοριακής συμφόρησης κατά τη διάρκεια αθλητικών εκδηλώσεων. Επιπλέον, η γεωγραφία της Αθήνας, μπορεί να καθιστά τη μετακίνηση με ποδήλατο λιγότερο ελκυστική σε σύγκριση με άλλες πόλεις.

Τέλος, η παρακίνηση των μετακινούμενων να αλλάξουν τον τρόπο με τον οποίο μετακινούνται και να υιοθετήσουν πιο βιώσιμες συμπεριφορές είναι μία από τις μεγαλύτερες προκλήσεις στις μεταφορές. Κάποιες φορές όμως όπως στην περίπτωση εμφάνισης συγκεκριμένων απροσδόκητων συμβάντων (π.χ. δυσμενών καιρικών συνθηκών) αυτό καθίσταται εξαιρετικά δύσκολο. Θα ήταν ενδιαφέρον να εξεταστούν τα κίνητρα που θα μπορούσαν να συμβάλουν ώστε οι μετακινούμενοι να αναπροσαρμόζουν τη συμπεριφορά τους ακόμα και σε απρόβλεπτες καταστάσεις, προκειμένου να διατηρούν τη βιωσιμότητα στις μετακινήσεις τους.

6 Βιβλιογραφία

- Adadi, A., & Berrada, M. (2018). Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI). *IEEE Access*, *6*, 52138–52160.
<https://doi.org/10.1109/access.2018.2870052>
- Aghababaei, M., Costello, S. B., & Ranjitkar, P. (2020). Transportation impact assessment following a potential Alpine fault earthquake in New Zealand. *Transportation Research Part D: Transport and Environment*, *87*, 102511.
<https://doi.org/10.1016/j.trd.2020.102511>
- Bhaduri, E., Manoj, B. S., Wadud, Z., Goswami, A. K., & Choudhury, C. F. (2020). Modelling the effects of COVID-19 on travel mode choice behaviour in India. *Transportation Research Interdisciplinary Perspectives*, *8*, 100273.
<https://doi.org/10.1016/j.trip.2020.100273>
- Breiman, L. (2001). Random Forests.
<https://www.stat.berkeley.edu/~breiman/randomforest2001.pdf>
- De Vos, J. (2020). The effect of COVID-19 and subsequent social distancing on travel behavior. *Transportation Research Interdisciplinary Perspectives*, *5*, 100121.
<https://doi.org/10.1016/j.trip.2020.100121>
- Ding, C., Wang, D., Liu, C., Zhang, Y., & Yang, J. (2017). Exploring the influence of built environment on travel mode choice considering the mediating effects of car ownership and travel distance. *Transportation Research Part A: Policy and Practice*, *100*, 65–80. <https://doi.org/10.1016/j.tra.2017.04.008>
- Drabicki, A. A., Islam, M. F., & Szarata, A. (2021). Investigating the Impact of Public Transport Service Disruptions upon Passenger Travel Behaviour—Results from Krakow City. *Energies*, *14*(16), 4889. <https://doi.org/10.3390/en14164889>
- Fafoutellis, P., Mantouka, E. G., Vlahogianni, E. I., & Oprea, G.-M. (2021). Acceptability modeling of autonomous mobility on-demand services with on-board ride sharing using interpretable Machine Learning. *International Journal of Transportation Science and Technology*. <https://doi.org/10.1016/j.ijtst.2021.10.003>
- Fuster-Palà, A., Luna-Perejón, F., & Domínguez-Morales, M. (2024, February 23). *Disease screening using Artificial Intelligence*. www.researchsquare.com.
<https://www.researchsquare.com/article/rs-3969817/v1>
- Hussain, D., Hussain, A., Mohammed, A., Dawod, S., Atiq, R., Rahmat, O., & Borhan, M. (2017). ANALYSIS OF TRANSPORTATION MODE CHOICE USING A COMPARISON OF ARTIFICIAL NEURAL NETWORK AND MULTINOMIAL LOGIT MODELS. *Journal of Engineering and Applied Sciences*, *12*(5).
http://www.arnjournals.org/jeas/research_papers/rp_2017/jeas_0317_5785.pdf

- Le, T. P. L., & Trinh, T. A. (2016). Encouraging Public Transport Use to Reduce Traffic Congestion and Air Pollutant: A Case Study of Ho Chi Minh City, Vietnam. *Procedia Engineering*, 142, 236–243. <https://doi.org/10.1016/j.proeng.2016.02.037>
- Liu, Y., Chen, J., Wu, W., & Ye, J. (2019). Typical Combined Travel Mode Choice Utility Model in Multimodal Transportation Network. *Sustainability*, 11(2), 549. <https://doi.org/10.3390/su11020549>
- Lu, Q.-C., Zhang, J., Peng, Z.-R., & A.B.M. Salman Rahman. (2014). Inter-city travel behaviour adaptation to extreme weather events. *Journal of Transport Geography*, 41, 148–153. <https://doi.org/10.1016/j.jtrangeo.2014.08.016>
- Nguyen-Phuoc, D. Q., Currie, G., De Gruyter, C., & Young, W. (2018). How do public transport users adjust their travel behaviour if public transport ceases? A qualitative study. *Transportation Research Part F: Traffic Psychology and Behaviour*, 54, 1–14. <https://doi.org/10.1016/j.trf.2018.01.009>
- Rahimi, E., Shamshiripour, A., Shabanpour, R., Mohammadian, A., & Auld, J. (2019). Analysis of transit users' waiting tolerance in response to unplanned service disruptions. *Transportation Research Part D: Transport and Environment*, 77, 639–653. <https://doi.org/10.1016/j.trd.2019.10.011>
- Scorrano, M., & Danielis, R. (2021). Active mobility in an Italian city: Mode choice determinants and attitudes before and during the Covid-19 emergency. *Research in Transportation Economics*, 101031. <https://doi.org/10.1016/j.retrec.2021.101031>
- Shakibaei, S., De jong, G. C., Alpkökin, P., & Rashidi, T. H. (2020). Impact of the COVID-19 pandemic on travel behavior in Istanbul: A panel data analysis. *Sustainable Cities and Society*, 102619. <https://doi.org/10.1016/j.scs.2020.102619>
- Shannon, C. E. (1948). A Mathematical Theory of Communication. *Bell System Technical Journal*, 27(4), 623–656. <https://doi.org/10.1002/j.1538-7305.1948.tb00917.x>
- Sun, F., Jin, M., Zhang, T., & Huang, W. (2022). Satisfaction differences in bus traveling among low-income individuals before and after COVID-19. *Transportation Research Part A: Policy and Practice*, 160, 311–332. <https://doi.org/10.1016/j.tra.2022.04.015>
- Sun, L., Chen, X., He, Z., & Miranda-Moreno, L. F. (2021). Routine Pattern Discovery and Anomaly Detection in Individual Travel Behavior. *Networks and Spatial Economics*, 23(2), 407–428. <https://doi.org/10.1007/s11067-021-09542-9>
- Teng, J., & Liu, W.-R. (2015). Development of a Behavior-Based Passenger Flow Assignment Model for Urban Rail Transit in Section Interruption Circumstance. *Urban Rail Transit*, 1(1), 35–46. <https://doi.org/10.1007/s40864-015-0002-0>
- Vittoratos, N., Tselentis, D., & Yannis, G. (2017). *Συσχέτιση Δεδηλωμένης και Αποκαλυφθείσας Συμπεριφοράς του Οδηγού με Χρήση των Διαγνωστικών Στοιχείων του Οχήματος*. <https://www.nrso.ntua.gr/geyannis/wp-content/uploads/geyannis-pc264.pdf>
- Wang, B., Shao, C., Li, J., Weng, J., & Ji, X. (2015). Holiday travel behavior analysis and empirical study under integrated multimodal travel information service. *Transport Policy*, 39, 21–36. <https://doi.org/10.1016/j.tranpol.2014.12.005>

- Zanni, A. M., & Ryley, T. J. (2015). The impact of extreme weather conditions on long distance travel behaviour. *Transportation Research Part A: Policy and Practice*, 77, 305–319. <https://doi.org/10.1016/j.tra.2015.04.025>
- Zhang, H. (2004). *The Optimality of Naive Bayes*.
<https://www.cs.unb.ca/~hzhang/publications/FLAIRS04ZhangH.pdf>
- Zhang, H., Wei Min Li, Liu, Y., & Zhang, L. (2023). Understanding Travel Mode Choice Behavior: Influencing Factors Analysis and Prediction with Machine Learning Method. *Sustainability*, 15(14), 11414–11414. <https://doi.org/10.3390/su151411414>
- Zheng, Z., Lee, J. (Brian), Saifuzzaman, M., & Sun, J. (2015). Exploring association between perceived importance of travel/traffic information and travel behaviour in natural disasters: A case study of the 2011 Brisbane floods. *Transportation Research Part C: Emerging Technologies*, 51, 243–259. <https://doi.org/10.1016/j.trc.2014.12.011>

Παράρτημα

Παρακάτω παρατίθεται το ερωτηματολόγιο της διπλωματικής εργασίας για την πόλη της Αθήνας

Ερωτηματολόγιο διερεύνησης συνηθειών μετακίνησης

Ενότητα Α: Προφίλ Κινητικότητας

Αυτή η ενότητα αποσκοπεί στην ανίχνευση του τρέχοντος προφίλ κινητικότητας κάθε ταξιδιώτη και ως εκ τούτου, οι ερωτώμενοι θα πρέπει να έχουν κατά νου τι κάνουν συνήθως κατά τις καθημερινές τους μετακινήσεις όταν απαντούν στις ερωτήσεις.

A1. Ποιος είναι ο κύριος σκοπός των καθημερινών σας μετακινήσεων; (μία απάντηση)

- Επαγγελματικός
- Εκπαιδευτικός
- Προσωπικός (π.χ. ιατρικοί, οικογενειακοί ή προσωπικοί λόγοι)
- Αναψυχή (γυμναστήριο, φίλοι, κοινωνικοί ζωή κ.λπ.)

A2. Ποιο μέσο χρησιμοποιείτε συχνότερα για τις καθημερινές σας μετακινήσεις; (μία απάντηση)

- Αυτοκίνητο
- Μοτοσικλέτα
- Μ.Μ.Μ.
- Ταξί
- Περπάτημα/ποδήλατο

A3. Μετακινήστε σε ώρες αιχμής; (μία απάντηση)

- Ναι
- Όχι

A4. Έχετε την ευελιξία να αλλάξετε την ώρα άφιξης στον προορισμό σας; (μία απάντηση)

- Όχι
- Ναι, έως 30 λεπτά
- Ναι, έως 60 λεπτά
- Ναι, πάνω από 60 λεπτά

A5. Ποιο είναι το μέσο ημερήσιο κόστος των μετακινήσεών σας (συμπεριλαμβανομένης της μετακίνησης και της στάθμευσης); (μία απάντηση)

- 0€
- 0-5€
- 5-10€
- >10

A6.1. Λάβετε υπόψη σας ότι ως μετακίνηση θεωρείται η κίνηση προς μια κατεύθυνση από ένα σημείο προέλευσης σε ένα σημείο προορισμού (π.χ. Σπίτι→Εργασία, επομένως εάν μετακινείτε καθημερινά από το σπίτι στο γραφείο και από το γραφείο στο σπίτι, για πέντε ημέρες την εβδομάδα, πραγματοποιείτε 10 μετακινήσεις) Πόσες εβδομαδιαίες επαγγελματικές μετακινήσεις πραγματοποιείτε (π.χ. επίσκεψη στο γραφείο, επίσκεψη σε πελάτη κ.λπ.); _____

A6.2. Πόσες εβδομαδιαίες εκπαιδευτικές μετακινήσεις πραγματοποιείτε (π.χ. από ή προς το σχολείο ή άλλα εκπαιδευτικά ιδρύματα); _____

A6.3. Πόσες εβδομαδιαίες μετακινήσεις πραγματοποιείτε για προσωπικούς λόγους (π.χ. για ιατρικές, οικογενειακές ή οικιακές υποχρεώσεις); _____

A6.4. Πόσες εβδομαδιαίες μετακινήσεις πραγματοποιείτε με σκοπό τη διασκέδαση σας (π.χ. για επίσκεψη σε φίλους, γυμναστήριο, άλλες κοινωνικές δραστηριότητες κ.λπ.); _____

Ενότητα Β: Αντίληψη των Χρηστών

Το σημερινό δίκτυο μεταφορών αντιμετωπίζει διάφορες προκλήσεις, όπως η κυκλοφοριακή συμφόρηση και η ατμοσφαιρική ρύπανση. Σε αυτήν την ενότητα, οι συμμετέχοντες ερωτώνται σχετικά με την αντίληψη τους για διάφορες στρατηγικές διαχείρισης της κυκλοφορίας και νέες υπηρεσίες που αποσκοπούν στη βελτίωση των συνθηκών του δικτύου.

B1. ΤΡΕΧΟΥΣΑ ΚΑΤΑΣΤΑΣΗ ΤΟΥ ΔΙΚΤΥΟΥ

1. Πόσο ικανοποιημένος/η είστε από το υπάρχον δίκτυο ΜΜΜ; (μόνο για τους συμμετέχοντες που απάντησαν Μ.Μ.Μ. στο Α2)

1(Πολύ δυσαρεστημένος/η)____2____3____4____5(Πολύ ικανοποιημένος/η)

2. Παρακαλούμε κατατάξτε τα παρακάτω χαρακτηριστικά του δικτύου ΜΜΜ της πόλης σας, από το καλύτερο προς το χειρότερο. (μόνο για τους συμμετέχοντες που απάντησαν Μ.Μ.Μ. στο Α2)

- Αξιοπιστία
- Συχνότητα
- Άνεση
- Κόστος
- Διαθεσιμότητα
- Ασφάλεια
- Ενημέρωση

3. Ανεξάρτητα από το μέσο μεταφοράς που συνήθως χρησιμοποιείτε, πόσο επηρεάζει η κυκλοφοριακή συμφόρηση τις καθημερινές σας μετακινήσεις;

1(Καθόλου)____2____3____4____5(Πάρα πολύ)

B2. ΣΥΣΤΗΜΑΤΑ ΤΙΜΟΛΟΓΗΣΗΣ

Προκειμένου να ανακουφιστεί το κέντρο της πόλης από την κυκλοφοριακή συμφόρηση και να ελαχιστοποιηθεί η ατμοσφαιρική ρύπανση, εφαρμόζεται ένα σύστημα αστικής τιμολόγησης, όπου οι χρήστες πρέπει να πληρώνουν ένα διόδιο κατά την είσοδό τους στο κέντρο της πόλης.

1. Υπό ποιες προϋποθέσεις θα πληρώνατε για την είσοδό σας στο κέντρο της πόλης; (πολλαπλές απαντήσεις)

- Ύπαρξη αξιόπιστων Μ.Μ.Μ.
- Σημαντική μείωση της κυκλοφοριακής συμφόρησης
- Η χρέωση να γίνεται ανάλογα με τη διανυόμενη απόσταση
- Η χρέωση να γίνεται ανάλογα με τον τύπο του οχήματος (π.χ. οχήματα χαμηλών ή υψηλών εκπομπών)
- Άλλο, διευκρινίστε _____
- Δεν είμαι διατεθειμένος να πληρώσω για την είσοδο μου στο κέντρο της πόλης

2. Σε περίπτωση που εφαρμόζεται ένα σύστημα αστικών διοδίων στο κέντρο της Αθήνας, ποιο από τα παρακάτω θα επιλέγατε να κάνετε; (μία απάντηση)

- ο Θα συνεχίσω να χρησιμοποιώ το ιδιωτικό μου αυτοκίνητο, πληρώνοντας το αντίστοιχο τέλος
- ο Θα συνεχίσω να χρησιμοποιώ το ιδιωτικό μου αυτοκίνητο, αλλάζοντας τη διαδρομή μου ώστε να αποφύγω την περιοχή χρέωσης
- ο Θα επέλεγα τα Μ.Μ.Μ.
- ο Θα επέλεγα ένα μέσο ενεργητικής μετακίνησης (ποδήλατο, περπάτημα)
- ο Άλλο, παρακαλώ διευκρινίστε _____

B3. ΥΠΗΡΕΣΙΕΣ ΚΑΤΑ ΠΑΡΑΓΓΕΛΙΑ

Μια υπηρεσία κατά παραγγελία είναι μια ευέλικτη, κοινόχρηστη υπηρεσία δημόσιων μεταφορών, για την οποία απαιτείται προκράτηση από τον χρήστη. Οι υπηρεσίες μετακίνησης κατά παραγγελία επιτρέπουν στους επιβάτες να προγραμματίσουν τη μετακίνησή τους την ώρα που τους βολεύει και να επιβιβαστούν στο όχημα σε μια προκαθορισμένη τοποθεσία. Οι μετακινούμενοι μπορούν συνήθως να επιλέξουν τα σημεία παραλαβής και αποβίβασης μέσω μιας εφαρμογής.

1. Θα ήσασταν πρόθυμος/η να χρησιμοποιήσετε μια υπηρεσία μετακίνησης κατά παραγγελία; (μία απάντηση)

- ο Ναι
- ο Όχι

2. Ποιο από τα παρακάτω επιχειρηματικά μοντέλα υπηρεσιών κατά παραγγελία θα ήσασταν πρόθυμοι να χρησιμοποιήσετε; (μία απάντηση)

- ο Υπηρεσία **πρώτου-τελευταίου χιλιομέτρου** σε συνδυασμό με τα ΜΜΜ: συγχρονισμός της μετεπιβίβασης μεταξύ των υπηρεσιών κατά παραγγελία και των τακτικών δημόσιων συγκοινωνιών, με σκοπό την ευκολότερη πρόσβαση στις τελευταίες (χαμηλή τιμή).
- ο Υπηρεσία κατά παραγγελία με **σταθερές στάσεις**: ο χρήστης επιλέγει να επιβιβαστεί και να αποβιβαστεί σε προκαθορισμένες στάσεις (μεσαία τιμή).
- ο Υπηρεσία κατά παραγγελία **από πόρτα σε πόρτα**: τα οχήματα μπορούν να κινούνται οπουδήποτε στο δίκτυο (υψηλή τιμή).

B4. ΣΥΓΧΡΟΝΙΣΜΟΣ ΤΗΣ ΚΥΚΛΟΦΟΡΙΑΣ ΚΑΙ ΤΩΝ Μ.Μ.Μ.

Η Προτεραιοποίηση Διέλευσης Οχημάτων ΜΜΜ Μέσω της Σηματοδότησης (Transit signal priority - TSP) είναι μια γενική ονομασία που περιλαμβάνει τις διάφορες τεχνικές που διευκολύνουν την κίνηση των οχημάτων ΜΜΜ σε διασταυρώσεις που ελέγχονται από φωτεινούς σηματοδότες.

1. Είστε διατεθειμένοι να χρησιμοποιήσετε τα Μέσα Μαζικής Μεταφοράς για τις καθημερινές σας μετακινήσεις, αν τους δινόταν προτεραιότητα στις διασταυρώσεις μέσω των φωτεινών σηματοδοτών; (μία απάντηση)

- Ναι
- Όχι

2. Παρακαλούμε κατατάξτε τους ακόλουθους στόχους που σχετίζονται με την εφαρμογή ενός συστήματος προτεραιότητας διέλευσης οχημάτων ΜΜΜ σε επιλεγμένες σηματοδοτούμενες διασταυρώσεις, από τους περισσότερο προς τους λιγότερο σημαντικούς:

- Αύξηση της συνέπειας και της αξιοπιστίας των Μ.Μ.Μ.
- Μείωση του χρόνου διαδρομής και των καθυστερήσεων των Μ.Μ.Μ.
- Εύλογη επίπτωση στις συνήθεις κυκλοφοριακές συνθήκες (π.χ. συμφόρηση) του οδικού δικτύου.
- Εύλογο κόστος εφαρμογής.
- Μείωση των εκπομπών αερίων του θερμοκηπίου που εκλύουν τα Μ.Μ.Μ.
- Μείωση των συνολικών εκπομπών αερίων του θερμοκηπίου που εκλύονται από την κίνηση οχημάτων.
- Χρήση του συστήματος από οχήματα έκτακτης ανάγκης για τη μείωση του χρόνου απόκρισης.

B5. ΑΝΤΙΔΡΑΣΗ ΣΕ ΑΠΡΟΣΔΟΚΗΤΑ ΓΕΓΟΝΟΤΑ

1. Πόσο πιθανό είναι να αλλάξετε τον συνήθη τρόπο μετακίνησής σας στις ακόλουθες περιπτώσεις;

- **Δυσμενείς καιρικές συνθήκες**
1(Αδύνατο)___2___3___4___5(Βέβαιο)
- **Φυσικοί Κίνδυνοι (π.χ. σεισμός, πλημμύρες)**
1(Αδύνατο)___2___3___4___5(Βέβαιο)
- **Αυξημένη κυκλοφορία οχημάτων**
1(Αδύνατο)___2___3___4___5(Βέβαιο)
- **Απροσδόκητη διαταραχή στη λειτουργία των Μ.Μ.Μ.**
1(Αδύνατο)___2___3___4___5(Βέβαιο)

- Προγραμματισμένες δραστηριότητες (π.χ. οδικά έργα, ποδοσφαιρικός αγώνας)
1(Αδύνατο)___2___3___4___5(Βέβαιο)
- Επικίνδυνα συμβάντα (π.χ. COVID-19)
1(Αδύνατο)___2___3___4___5(Βέβαιο)

2. Πόσο συχνά μεταβάλλετε τις συνήθειες μετακίνησής σας λόγω διαταραχών στη λειτουργία των Μ.Μ.Μ. ;

Ποτέ	Σπάνια	Μερικές φορές	Συχνά	Πάντα
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

3. Ποιος είναι ο μέγιστος αποδεκτός χρόνος αναμονής για εσάς προτού εξετάσετε έναν εναλλακτικό τρόπο μετακίνησης;

Για κρίσιμες μετακινήσεις όπου είναι σημαντικό να φτάσετε εγκαίρως στον προορισμό σας (π.χ. εργασία)

5 Λεπτά	10 Λεπτά	15 Λεπτά	20 Λεπτά	30 Λεπτά	60 Λεπτά
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Για μη κρίσιμες μετακινήσεις όπου δεν υπάρχει ιδιαίτερη πίεση να φτάσετε εγκαίρως στον προορισμό σας (π.χ. ψώνια)

5 Λεπτά	10 Λεπτά	15 Λεπτά	20 Λεπτά	30 Λεπτά	60 Λεπτά
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

4. Τι είδους ενημέρωση προσφέρει στους επιβάτες ο φορέας διαχείρισης των δημόσιων συγκοινωνιών στην πόλη σας σε περίπτωση διαταραχής ενός δρομολογίου; (επιλέξτε όσα ισχύουν)

- Πληροφορίες σχετικά με το περιστατικό (π.χ. τοποθεσία, σταθμοί που επηρεάζονται κ.λπ)
- Πληροφορίες σε πραγματικό χρόνο σχετικά με την πρόοδο της αποκατάστασης της βλάβης
- Εναλλακτικές διαδρομές
- Εναλλακτικές επιλογές μετακίνησης (π.χ. υπηρεσία κατά παραγγελία)
- Τίποτα από τα παραπάνω/δεν γνωρίζω

5. Πώς θα αντιδρούσατε σε περίπτωση διακοπής της λειτουργίας του οχήματος Μ.Μ.Μ. που χρησιμοποιείτε εκείνη τη στιγμή; (μία απάντηση)

- Θα περίμενα στη στάση

- Θα χρησιμοποιούσα την εναλλακτική γραμμή Μ.Μ.Μ. που μου προτείνει ο διαχειριστής του δικτύου δημοσίων συγκοινωνιών
- Θα χρησιμοποιούσα μια εναλλακτική γραμμή Μ.Μ.Μ. που θα επέλεγα αυτόνομα
- Θα περπατούσα μέχρι τον προορισμό μου
- Θα επέλεγα ένα ιδιωτικό μέσο μετακίνησης
- Θα ματαίωνα τη μετακίνηση μου

Ενότητα Γ: Δημογραφικά Χαρακτηριστικά

D1. Φύλο:

- Άνδρας
- Γυναίκα

D2. Ηλικία:

- 18-24
- 25-34
- 35-44
- 45-54
- 55-64
- >65

D3. Ετήσιο προσωπικό εισόδημα:

- Χαμηλό (<10.000€)
- Μεσαίο (10,000 – 25.000€)
- Υψηλό (>25.000€)

D4. Επάγγελμα:

- Δημόσιος υπάλληλος
- Ιδιωτικός υπάλληλος
- Ελεύθερος επαγγελματίας
- Σε αναζήτηση εργασίας
- Συνταξιούχος
- Φοιτητής

D5.1 Διαθέτετε αυτοκίνητο:

- Ναι
- Όχι

D5.2 Αν ναι, πόσα αυτοκίνητα διαθέτετε; _____

D6. Από πόσα μέλη απαρτίζεται το νοικοκυριό σας; _____

D7. Κατοικείται στο κέντρο της Αθήνας;

- Ναι Όχι

D8. Πώς θα χαρακτηρίζατε τον εαυτό σας με βάση τις συνήθειες μετακίνησης σας;

- Είμαι οικολόγος, επιλέγω πάντα την πιο φιλική προς το περιβάλλον εναλλακτική για τη μετακίνηση μου.
- Εξοικονομώ χρόνο, επιλέγω πάντα τη συντομότερη διαδρομή/δρομολόγιο για τη μετακίνηση μου.
- Είμαι ενθουσιώδης με την τεχνολογία, χρησιμοποιώ πάντα τις πιο σύγχρονες υπηρεσίες κινητικότητας (π.χ. υπηρεσίες κινητικότητας κατά παραγγελία, ηλεκτρονικά πατίνια, υπηρεσίες διαμοιρασμού οχημάτων ή διαδρομών με άλλους χρήστες)
- Εξαρτώμαι από το αυτοκίνητο μου, πηγαίνω παντού με αυτό.