ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΠΟΛΙΤΙΚΩΝ ΜΗΧΑΝΙΚΩΝ

ΤΟΜΕΑΣ ΜΕΤΑΦΟΡΩΝ ΚΑΙ ΣΥΓΚΟΙΝΩΝΙΑΚΗΣ ΥΠΟΔΟΜΗΣ

Διπλωματική Εργασία

Ανάπτυξη προτύπων πρόβλεψης επιβατικής κίνησης ανά λεωφορειακή γραμμή

Αναστασία Υφαντή

Επιβλέπουσα Καθηγήτρια : Ελένη Βλαχογιάννη,
Αναπληρώτρια Καθηγήτρια Σχολής Πολιτικών Μηχανικών ΕΜΠ

ΑΘΗΝΑ, ΜΑΡΤΙΟΣ 2022

Ευχαριστίες

Ολοκληρώνοντας τη παρούσα διπλωματική εργασία, η οποία σηματοδοτεί το πέρας των σπουδών μου στη σχολή Πολιτικών Μηχανικών του Εθνικού Μετσόβιου Πολυτεχνείου, θα ήθελα να ευχαριστήσω όλους όσους συνέβαλαν στην περάτωση της.

Πρωτίστως ευχαριστώ θερμά την επιβλέπουσα καθηγήτρια μου κα Ελένη Βλαχογιάννη, Αναπληρώτρια Καθηγήτρια στον Τομέα Μεταφορών και Συγκοινωνιακής Υποδομής της Σχολής Πολιτικών Μηχανικών του Εθνικού Μετσόβιου Πολυτεχνείου, για την ανάθεση της συγκεκριμένης εργασίας, την καθοδήγηση και την εξαιρετική συνεργασία μας όλη αυτή την περίοδο.

Επιπλέον, θα ήθελα να ευχαριστήσω θερμά τον υποψήφιο Δρ. Μάριο Γιουρουκέλη για την εξαιρετική βοήθεια, τις πολύτιμες υποδείξεις και τον χρόνο που αφιέρωσε σε όλα τα στάδια ολοκλήρωσης της εργασίας, καθώς και για το πολύ καλό κλίμα συνεργασίας που διαμόρφωσε.

Κλείνοντας, θα ήθελα να ευχαριστήσω την οικογένειά μου και τους φίλους μου για την υποστήριξη που μου προσέφεραν καθ' όλη τη διάρκεια των σπουδών μου.

Ανάπτυξη προτύπων πρόβλεψης επιβατικής κίνησης ανά λεωφορειακή γραμμή

Αναστασία Υφαντή

Επιβλέπουσα Καθηγήτρια: Ελένη Ι. Βλαχογιάννη, Αναπλ. Καθηγήτρια ΕΜΠ.

Σύνοψη

Στόχος της παρούσας διπλωματικής εργασίας είναι η διερεύνηση της επιρροής της πανδημίας του COVID-19 στην εξέλιξη της επιβατικής ζήτησης στο δίκτυο Μέσων Μαζικής Μεταφοράς του ΟΑΣΑ. Για το σκοπό αυτό, συλλέχθηκαν στοιχεία των ακυρώσεων καθώς και στοιχεία χαρακτηριστικών των λεωφορειακών γραμμών. Για την ανάλυση των δεδομένων, αναπτύχθηκαν μοντέλα μη γραμμικής παλινδρόμησης με Δένδρα Αποφάσεων με εξαρτημένη μεταβλητή την διακύμανση των επικυρώσεων των εισιτηρίων και με ανεξάρτητες μεταβλητές χαρακτηριστικά των λεωφορειακών γραμμών. Από την εφαρμογή της μεθοδολογίας, προέκυψε ένα μοντέλο για την εκτίμηση της επιρροής των χαρακτηριστικών μιας λεωφορειακής γραμμής στο επιβατικό κοινό του ΟΑΣΑ, κατά την περίοδο του δεύτερου κύματος της πανδημίας. Χαρακτηριστικά όπως το μήκος της διαδρομής μιας γραμμής, η επικάλυψη των διάφορων λεωφορειακών γραμμών και η τυπολογία των οδών που εξυπηρετούν κάθε γραμμή, αποτελούν σημαντικούς παράγοντες επιρροής της ζήτησης.

Λέξεις κλειδιά: επικυρώσεις, μήκος, επικάλυψη, ΟΑΣΑ, Μηχανική Μάθηση, Δένδρα Αποφάσεων, Δένδρα Παλινδρόμησης, Python

Development of passenger traffic forecasting models per bus line

Anastasia Yfanti

Supervising Professor: Eleni I. Vlahogianni, Associate Professor NTUA

Abstract

The aim of this present Diploma Thesis is to investigate the impact of the COVID-19 pandemic on the evolution of passenger demand in the OASA Public Transport Network. For this purpose, data on ticket cancellations as well as data on bus lines were collected. For data analysis, models of nonlinear regression with Decision Trees were developed with the variance of ticket validation as the dependent variable and with independent variables characteristics of bus lines. From the application of the methodology, a model emerged to assess the influence of the characteristics of a bus line on the ridership of the OASA bus lines, during the period of the second wave of the pandemic. Features such as the length of the route of a line, the overlap of the various bus lines and the typology of the bus line's route, are important factors influencing the demand.

Key words: validations, length, overlap, OASA, Machine Learning, Decision Trees, Regression Trees, Python

Περίληψη

Η εκπόνηση της παρούσας διπλωματικής εργασίας έχει στόχο τη διερεύνηση της επιρροής της πανδημίας του COVID-19 στην εξέλιξη της επιβατικής ζήτησης στο δίκτυο Μέσων Μαζικής Μεταφοράς του ΟΑΣΑ. Για τον σκοπό αυτό συλλέχθηκαν δεδομένα που αφορούσαν στα ακόλουθα:

- Ημερήσιες επικυρώσεις εισιτηρίων στις λεωφορειακές γραμμές (ιστοσελίδα Κυβέρνησης https://www.data.gov.gr/) για την περίοδο Σεπτέμβριος 2020 έως Δεκέμβριος 2021 (κατά την περίοδο του δεύτερου κύματος της πανδημίας στην Ελλάδα)
- Δεδομένα για τα χαρακτηριστικά των λεωφορειακών γραμμών από τον ΟΑΣΑ και την ιστοσελίδα OpenStreetMap (https://www.openstreetmap.org/).

Σύμφωνα με το θεωρητικό υπόβαθρο, αναπτύχθηκαν μοντέλα Δένδρων Παλινδρόμησης για να εξετασθεί η σχέση μεταξύ του επιβατικού κοινού και των χαρακτηριστικών μιας γραμμής.

Το μοντέλο που προέκυψε παρουσίασε χαμηλούς δείκτες σφαλμάτων καθώς και ικανοποιητική τιμή γραμμικής συσχέτισης (δείκτης R^2), συμπεραίνοντας ότι είναι ένα σχετικά καλό μοντέλο. Στη συνέχεια, ακολουθούν τα πιο σημαντικά συμπεράσματα που προέκυψαν από την εφαρμογή των μοντέλων.

Η μεγαλύτερη διακύμανση της ζήτησης παρατηρείται σε περιπτώσεις όπου η επικάλυψη των γραμμών είναι πολύ μικρή (overlap<6.5%) και σε περιπτώσεις όπου το μήκος των λεωφορειακών γραμμών είναι μεγαλύτερο των 24 χιλιομέτρων. Αυτό οφείλεται στο γεγονός ότι κατά την περίοδο των περιοριστικών μέτρων υπήρξε μια μείωση στις άσκοπες μετακινήσεις με αποτέλεσμα οι μετακινήσεις που παραμένουν να γίνονται κοντά στον τόπο κατοικίας μέσω τοπικών λεωφορειακών γραμμών.

Επιπλέον, σχετικά μεγάλη διακύμανση της ζήτησης παρατηρήθηκε σε περιπτώσεις όπου η επικάλυψη των γραμμών και το ποσοστό της διαδρομής της που διέρχεται από κατοικημένες περιοχές είναι αυξημένα (overlap>6.5% και residential>26.7%). Πρόκειται για λεωφορειακές γραμμές που διασχίζουν πυκνά κατοικημένες περιοχές της πόλης και χρησιμοποιούν κοινές οδούς με άλλες γραμμές.

Τέλος, μικρότερη διακύμανση στη ζήτηση παρατηρήθηκε σε περιπτώσεις όπου το ποσοστό της διαδρομής των λεωφορειακών γραμμών που διέρχεται από κατοικημένη περιοχή είναι μικρότερο από 26.7%. Αυτό ίσως να οφείλεται στο ότι οι γραμμές αυτές εξυπηρετούν περιοχές μακριά από το κέντρο της πόλης ή περιοχές με άλλες χρήσεις γης. Σε αυτές τις περιπτώσεις οι μετακινούμενοι προτίμησαν μετάβαση από τη χρήση δημοσίων μέσων μεταφοράς σε ιδιωτικά και ατομικά μέσα.

Ύστερα παρουσιάστηκαν κάποιες προτάσεις για περαιτέρω έρευνα:

Η πανδημία του κορονοϊού επέφερε σημαντικές αλλαγές στις καθημερινότητες των ανθρώπων, στην κινητικότητα καθώς και στις επιλογές των μετακινούμενων. Καθώς όμως γίνεται ετοιμασία για την επιστροφή στην κανονικότητα, οι μετακινήσεις αυξάνονται ξανά σε επίπεδα προ κορονοϊού και η χρήση των μέσων μαζικής μεταφοράς είναι αναπόφευκτη. Επομένως κρίνεται απαραίτητο να επέλθουν κάποιες αλλαγές στις δημόσιες συγκοινωνίες.

Κάποιες αλλαγές που μπορούν να εφαρμοστούν από τον ΟΑΣΑ, αφορούν τις συχνότητες των δρομολογίων. Προτείνεται η αύξηση των δρομολογίων για λεωφορειακές γραμμές που διέρχονται από πυκνά κατοικημένες περιοχές με μεγάλο επιβατικό κοινό καθώς και η δημιουργία νέων γραμμών που θα έχουν ανταπόκριση με διαφορετικές υφιστάμενες λεωφορειακές γραμμές, δηλαδή με μεγάλο ποσοστό επικάλυψης. Επιπλέον, μέτρα για τη μείωση του φόβου μετακίνησης μπορούν να εφαρμοστούν από τεχνολογίες όπως η τεχνητή νοημοσύνη (AI) από τις οποίες μπορούν να γίνει ο εντοπισμός του κινδύνου μετάδοσης κατά τη διάρκεια ταξιδιών με τα δημόσια μέσα μεταφοράς.

Όσον αφορά την ανάλυση του μοντέλου υπάρχουν κάποιες προσθήκες που μπορούν να οδηγήσουν σε βελτίωση αυτού. Μπορούν να αξιοποιηθούν από το Πανεπιστήμιο της Οξφόρδης δεδομένα της κλίμακας αυστηρότητας των περιοριστικών μέτρων ώστε να γίνει λεπτομερέστερη ανάλυση για κάθε επίπεδο αυστηρότητας. Επιπλέον η ανάλυση μπορεί να επεκταθεί για άλλες χώρες καθώς επίσης να εξετασθούν οι μεταβολές κινητικότητας σε ενδεχομένως επόμενα «κύματα» ή στην περίπτωση νέας πανδημίας.

ΠΕΡΙΕΧΟΜΕΝΑ

ΕΥΡΕΤΗΡΙΟ ΕΙΚΟΝΩΝ

ΕΥΡΕΤΗΡΙΟ ΠΙΝΑΚΩΝ

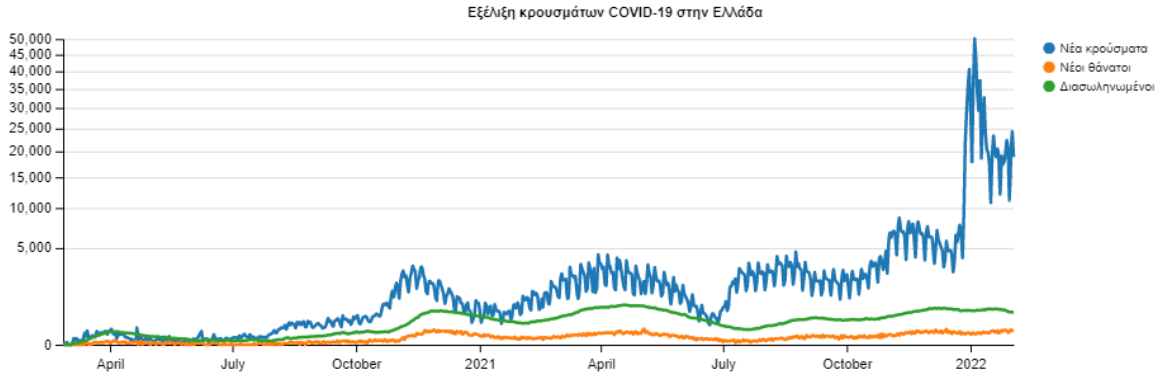
ΕΥΡΕΤΗΡΙΟ ΣΧΕΣΕΩΝ

1. ΕΙΣΑΓΩΓΗ

1.1 Γενικά

Στη Γουχάν της Κίνας, ένα ξέσπασμα πνευμονίας εντοπίστηκε τον Δεκέμβριο του 2019. Έκτοτε έχει αναγνωριστεί ως νέος και μεταδοτικός κορονοϊός, ο οποίος τώρα ονομάζεται COVID-19 (Zhu et al., 2020). Αφού εξαπλώθηκε σε όλο τον κόσμο με ανησυχητικό ρυθμό, ο Παγκόσμιος Οργανισμός Υγείας (ΠΟΥ) ανακήρυξε τον COVID-19 ως πανδημία στις 11 Μαρτίου 2020 (ΠΟΥ, 2020).

Το πρώτο κρούσμα στην Ελλάδα και ο πρώτος συσχετισμένος θάνατος καταγράφηκαν στις 26 Φεβρουαρίου και στις 19 Μαρτίου 2020, αντίστοιχα. Μέχρι τον Ιούνιο του 2021, τα επιβεβαιωμένα κρούσματα της νόσου υπερβαίνουν τις 420 χιλιάδες, συμπεριλαμβανομένων 12 χιλιάδες θανάτων (covid19.gov.gr) (Figure 1).



Εικόνα 1. Εξέλιξη κρουσμάτων COVID-19 στην Ελλάδα.

Έκτοτε η ελληνική κυβέρνηση έχει λάβει μια σειρά από μέτρα για τον περιορισμό της εξάπλωσης της πανδημίας, με κυριότερα την επιβολή lockdown σε ολόκληρη τη χώρα, δηλαδή την αυστηρή σύσταση παραμονής στην κατοικία και μετακίνησης μόνο για τα απαραίτητα, καθώς και την καθολική απαγόρευση της κυκλοφορίας, από τις αρχές Νοεμβρίου 2020, για συγκεκριμένα χρονικά διαστήματα της ημέρας, την αναστολή λειτουργίας των χώρων εστίασης και των καταστημάτων λιανικής, το κλείσιμο των εκπαιδευτικών ιδρυμάτων όλων των βαθμίδων, των χώρων εργασίας και των αθλητικών κέντρων και τη γενικότερη απαγόρευση της πλειοψηφίας των μετακινήσεων.

Η κινητικότητα και οι μεταφορές αποτελούσαν ανέκαθεν έναν από τους σημαντικότερους και κρισιμότερους τομείς για την ευημερία και την πρόοδο των

ανθρώπων, ο οποίος συνεχώς εξελίσσεται και επηρεάζει σε μεγάλο βαθμό την καθημερινότητα και τις συνθήκες ζωής τους.

Από τα παραπάνω, γίνεται αντιληπτό ότι οι μεγάλες αυτές αλλαγές στην καθημερινότητα καθώς και ο φόβος έκθεσης στον ιό, είχαν σημαντικό αντίκτυπο στις μετακινήσεις και ειδικότερα στις μετακινήσεις με τα Μέσα Μαζικής Μεταφοράς.

1.2 Σκοπός

Ο σκοπός της παρούσας Διπλωματικής Εργασίας αποτελεί **η διερεύνηση της επιρροής της πανδημίας του COVID-19 στην εξέλιξη της επιβατικής ζήτησης στο δίκτυο Μέσων Μαζικής Μεταφοράς του ΟΑΣΑ.**

Πιο συγκεκριμένα, ο στόχος είναι η εκτίμηση της επιρροής των χαρακτηριστικών μιας λεωφορειακής γραμμής στο επιβατικό κοινό του ΟΑΣΑ. Τα χαρακτηριστικά των λεωφορειακών είναι το συνολικό **μήκος** που διανύει η γραμμή, το ποσοστό που μια γραμμή **επικαλύπτεται** με άλλες γραμμές δηλαδή χρησιμοποιεί κοινούς δρόμους καθώς και την **τυπολογία** των δρόμων που χρησιμοποιεί η κάθε γραμμή.

Για την επίτευξη αυτού του στόχου, κρίθηκε αναγκαία η επιλογή κατάλληλου μοντέλου για την περιγραφή και την εκτίμηση των ιστορικών δεδομένων, πραγματοποιώντας **ανάλυση με Δένδρα Απόφασης Παλινδρόμησης**.

Τέλος, σκοπός είναι τα εξαγόμενα από αυτήν τη διπλωματική συμπεράσματα να βοηθήσουν στην καλύτερη κατανόηση του πως τα γεωμετρικά χαρακτηριστικά των γραμμών του ΟΑΣΑ επηρεάζουν την διακύμανση της ζήτησης εν μέσω της πανδημίας και να φανούν χρήσιμα στη λήψη των αποφάσεων που αφορούν στη διαχείριση της πανδημίας.

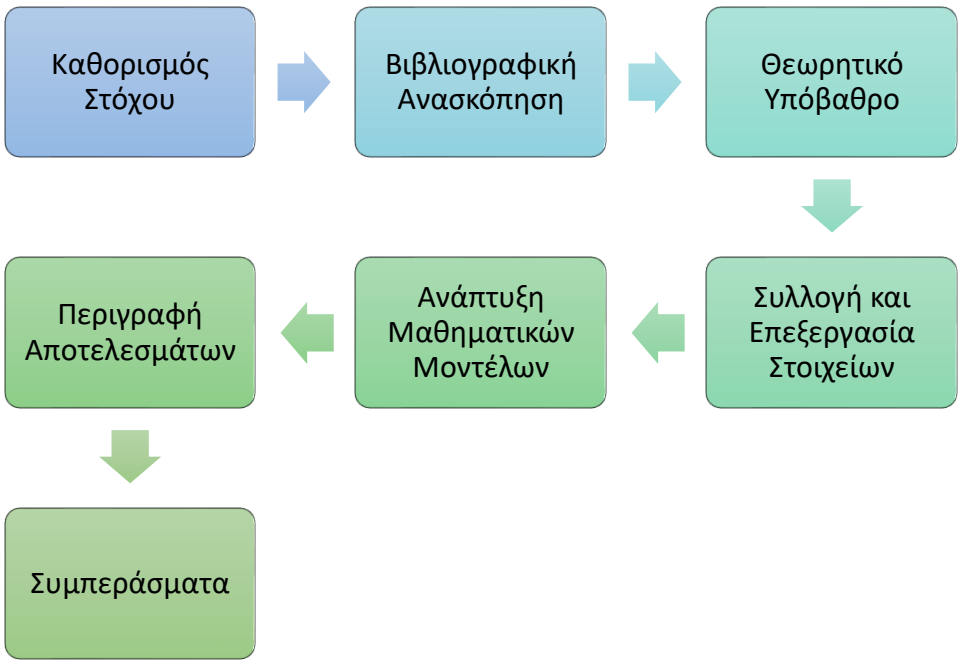
1.3 Μεθοδολογία

Αρχικό βήμα αποτελεί ο προσδιορισμός του στόχου της διπλωματικής εργασίας. Μετά την οριστικοποίηση του επιδιωκόμενου στόχου, πραγματοποιείται βιβλιογραφική ανασκόπηση τόσο σε ελληνική όσο και σε διεθνή βιβλιογραφία, με σκοπό την εύρεση παρεμφερών ερευνών και μεθοδολογιών ανάλυσης, τον εντοπισμό ζητημάτων που χρήζουν περαιτέρω έρευνας, καθώς και την αναζήτηση πιθανών τρόπων προσέγγισης και ανάλυσής τους.

Μετά την ολοκλήρωση της αναζήτησης βιβλιογραφικών αναφορών, ακολουθεί η συλλογή και επεξεργασία των στοιχείων και διαμορφώνεται η τελική, ηλεκτρονική

βάση δεδομένων.

Κατόπιν, αναπτύσσονται τα κατάλληλα μαθηματικά, στατιστικά μοντέλα που χρησιμοποιούνται για την ανάκτηση αποτελεσμάτων, τα οποία εν συνεχεία περιγράφονται και αναλύονται, για να διεξαχθούν τελικά συμπεράσματα και να διατυπωθούν προτάσεις για περαιτέρω έρευνα. Η παραπάνω διαδικασία αποτυπώνεται στην εικόνα 1.



Εικόνα 2. Μεθοδολογία διπλωματικής εργασίας

1.4 Δομή

Η υπόλοιπη εργασία δομείται στα παρακάτω κεφάλαια:

Το **κεφάλαιο 2**, αποτελεί τη βιβλιογραφική ανασκόπηση, οπού παρατίθενται χρήσιμα ευρήματα από συναφείς έρευνες και μεθοδολογίες.

Στο **κεφάλαιο 3**, γίνεται αναφορά στο θεωρητικό υπόβαθρο και τις μεθόδους που απαιτούνται για τη στατιστική ανάλυση των δεδομένων.

Στο **κεφάλαιο 4**, περιγράφεται η διαδικασία συλλογής και επεξεργασίας των στοιχείων για την δημιουργία βάσης δεδομένων και στη συνέχεια, η απαιτούμενη διαδικασία επεξεργασίας της πριν την ανάλυση της.

Στο **κεφάλαιο 5**, περιλαμβάνεται η αναλυτική περιγραφή της μεθοδολογίας που εφαρμόστηκε, τα βήματα που ακολουθήθηκαν και παρουσιάζονται τα παραγόμενα

αποτελέσματα.

Στο **κεφάλαιο 6**, παρουσιάζονται όλα τα συμπεράσματα, τα οποία προέκυψαν κατά την αξιολόγηση των μαθηματικών μοντέλων. Αναφέρονται, επίσης, προτάσεις για περαιτέρω έρευνα.

Στο τέλος, παρατίθεται σε μορφή καταλόγου η **βιβλιογραφία** που αξιοποιήθηκε κατά την εκπόνηση της Διπλωματικής Εργασίας καθώς και ο κώδικας ανάπτυξης των μοντέλων στο Παράτημα Α.

2. ΒΙΒΛΙΟΓΡΑΦΙΚΗ ΑΝΑΣΚΟΠΗΣΗ

2.1 Εισαγωγή

Στο παρόν κεφάλαιο της βιβλιογραφικής ανασκόπησης, παρουσιάζονται έρευνες της διεθνούς βιβλιογραφίας, το αντικείμενο και η μεθοδολογία των οποίων παρουσιάζουν συνάφεια με αυτά της παρούσας Διπλωματικής Εργασίας. Πιο συγκεκριμένα, παρουσιάζονται επιστημονικές εργασίες που ασχολούνται με την επίδραση της πανδημίας και των μέτρων που λήφθηκαν για την αντιμετώπισή της στην κινητικότητα, τον σκοπό και τον τρόπο μετακίνησης και κυρίως στα ΜΜΜ.

Παρακάτω περιγράφονται συνοπτικά οι σχετικές έρευνες, οι μέθοδοι ανάλυσης που ακολουθήθηκαν, καθώς και τα αποτελέσματα που προέκυψαν. Τέλος, αναδεικνύονται τα βασικά συμπεράσματα της βιβλιογραφίας και οι πιθανές ελλείψεις που παρατηρούνται.

2.2 Συναφείς Έρευνες και Μεθοδολογίες

Σύμφωνα με τις έρευνες που παρατίθενται στην συνέχεια, γίνεται κατανοητή η μεγάλη επίδραση της πανδημίας στις καθημερινότητες των ανθρώπων, στην κινητικότητα καθώς και στις επιλογές των μετακινούμενων.

2.2.1 Μείωση Κινητικότητας

Η μείωση της κινητικότητας είναι ο στόχος των περιοριστικών μέτρων, καθώς είναι **καθοριστική για τη μείωση της εξάπλωσης του κορονοϊού 2019**, όπως αποδεικνύεται από την έρευνα των Badr et al. (2020) στις Ηνωμένες Πολιτείες και έχει άμεση επίδραση στην μείωση των θυμάτων από κορονοϊό στο Ηνωμένο Βασίλειο (Hadjidemetriou et al. 2020). Από τα λογιστικά μοντέλα παλινδρόμησης της έρευνας των Maiti et al. (2021) για τις Ηνωμένες Πολιτείες, προέκυψε υψηλή συσχέτιση μεταξύ τόσο της εγχώριας όσο και της διεθνούς κινητικότητας με τον αριθμό των κρουσμάτων και των θανάτων του κορονοϊού 2019. Υψηλή συσχέτιση μεταξύ της κινητικότητας και της εξάπλωσης του ιού παρατηρήθηκε σε 52 χώρες από τους Nouvellet et al. (2021). Για παράδειγμα, στο Ηνωμένο Βασίλειο απότομη αύξηση της κινητικότητας συνδέεται με απότομη έξαρση των μολύνσεων και το αντίθετο. Ακόμη, οι Thakkar et al. (2020) παρατήρησαν πως μειώνεται ο δείκτης μεταδοτικότητας στην Washington από τις 18 Μάρτη, από όταν δηλαδή μειώθηκε και ο δείκτης της κινητικότητας.

Τον αντίκτυπο της πανδημίας στις **μετακινήσεις και την κινητικότητα** στην Ολλανδία μελετά η έρευνα των de Haas et al (2020). Τα δεδομένα συλλέχθηκαν από δείγμα 2500 ερωτηθέντων από την Ολλανδική Ομάδα Κινητικότητας (MPN) μεταξύ 27 Μαρτίου και 4 Απριλίου 2020. Χρησιμοποιήθηκαν συγκριτικές αναλύσεις για τη σύγκριση της κατάστασης κατά τη διάρκεια του κορονοϊού με αυτήν του φθινοπώρου του 2019 και συμπληρώθηκαν από μια δοκιμή X^2. Τελικά, το 85% των ερωτηθέντων μείωσε τα ψώνια εκτός σπιτιού, το 90% των ερωτηθέντων μείωσε τις επισκέψεις σε

άλλους, το ποσοστό των ατόμων που εργάζονται από το σπίτι αυξήθηκε από 6% σε 39%, οι μετακινήσεις και οι διανυόμενες αποστάσεις μειώθηκαν κατά 55% και 68% αντίστοιχα. Οι μετακινήσεις με Μ.Μ.Μ. μειώθηκαν άνω του 90% , ενώ αυξήθηκε η πεζοπορία, η ποδηλασίας και η χρήση ΙΧ.

Η έρευνα των Beria et al. (2021) επικεντρώθηκε στις εσωτερικές μετακινήσεις στην Ιταλία κατά τη διάρκεια του πρώτου «κύματος», οι οποίες μειώθηκαν έως και 80% την πρώτη εβδομάδα υποχρεωτικού περιορισμού. Η βόρεια Ιταλία, η οποία επλήγη πρώτη και σε μεγαλύτερο βαθμό, καταγράφει μείωση των μετακινήσεων από τις 24 Φεβρουαρίου, ενώ η υπόλοιπη χώρα περιορίζεται από τις 9 Μαρτίου, την ημέρα ανακοίνωσης των εθνικών περιορισμών. Αναφέρεται, επιπλέον, ότι τα ταξίδια το Σαββατοκύριακο διακόπηκαν μέχρι τον Μάιο. Στο ίδιο πλαίσιο, η έρευνα των Santamaria et al. (2020) επεσήμανε την **απότομη μείωση της κινητικότητας** κατά τις πρώτες τρεις εβδομάδες του Μαρτίου, πρώτα στην Ιταλία και έπειτα στις υπόλοιπες ευρωπαϊκές χώρες, γεγονός που μπορεί να εξηγηθεί έως και 90% λόγω των μέτρων περιορισμού.

Επιπλέον παρατηρήθηκε αύξηση της μέσης απόστασης των διαδρομών με ποδήλατο κατά 30% και των διαδρομών με τα πόδια κατά 83%. Στην έρευνα των Shamshiripour et al. (2020) φαίνεται η αλλαγή στην κινητικότητα και την ταξιδιωτική συμπεριφορά κατά τη διάρκεια του κορονοϊού στο Σικάγο. Μέσω ενός ερωτηματολογίου που διανεμήθηκε από τις 25 Απριλίου έως τις 2 Ιουνίου 2020 από τη διαδικτυακή πλατφόρμα Qualtrics και της χρήσης του Google Map API για τη συλλογή των οικιστικών τοποθεσιών των ερωτηθέντων, συλλέχθηκαν 915 δείγματα προς ανάλυση. Χρησιμοποιώντας τις μεθόδους δηλωμένης προτίμησης (SP)και αποκαλυφθήσας προτίμησης (RP) βρέθηκε αύξηση της εργασίας από το σπίτι, αύξηση του online shopping για είδη παντοπωλείου από 20% σε 33% τους μήνες της πανδημίας, αύξηση των online παραγγελιών έτοιμου φαγητού από 42% σε 55%, αύξηση της χρήσης αυτοκινήτου και ποδηλάτου και μείωση της χρήσης Μ.Μ.Μ.

Οι **επιπτώσεις των προληπτικών μέτρων** στη συμπεριφορά μετακίνησης διερευνήθηκαν επίσης από τους Muley et al. (2021) στο κράτος του Κατάρ. Διαπιστώθηκε ότι μετά την εφαρμογή κάθε απόφασης (π.χ. κλείσιμο εκπαιδευτικών ιδρυμάτων, πάρκων, όλων των εμπορικών καταστημάτων και των δημόσιων συγκοινωνιών) ο φόρτος κυκλοφορίας μειωνόταν μαζικά. Αν και επιβλήθηκαν περαιτέρω περιορισμοί, η προαναφερθείσα πτώση της κινητικότητας αναχαιτίστηκε, καθώς πλησίαζε ο Ιερός Μήνας του Ραμαζανιού. Ομοίως, η έρευνα των Parr et al. (2020) έδειξε ότι η κυκλοφορία μειώθηκε απότομα σε αντιστοιχία με την κήρυξη της πολιτείας της Φλόριντα σε κατάσταση έκτακτης ανάγκης και το κλείσιμο των σχολείων και των εστιατορίων.

Η έρευνα των Pullano et al. (2020) αναφέρει μείωση κατά 75% των μετακινήσεων σε ώρες αιχμής στη Γαλλία, ως απόρροια της εξ αποστάσεως εκπαίδευσης και της τηλεργασίας. Σύμφωνα με την έρευνα των Aloi et al. (2020) στην πόλη Santander στην Ισπανία, η πολιτική «Μένουμε σπίτι» οδήγησε στην πτώση της συνολικής κινητικότητας κατά 67% και **μείωση της οδήγησης** έως 85% σε σχέση με το 2018. Οι

πολίτες πραγματοποιούσαν, επιπρόσθετα, συντομότερες διαδρομές, κυρίως για την αγορά βασικών προϊόντων. Στη Βουδαπέστη, η κινητικότητα μειώθηκε κατά 57% σε σχέση με το 2017 και παρατηρήθηκε λιγότερη κυκλοφοριακή συμφόρηση, η οποία συνοδεύτηκε από την περιορισμένη χρήση εφαρμογών πλοήγησης (Bucksy, 2020).

Από την έρευνα των Saladié et al. (2020) στην επαρχία Tarragona στην Ισπανία, προέκυψε άμεση μείωση της κινητικότητας μετά την επιβολή των μέτρων, η οποία κατέγραψε πτώση έως 63%. Παρατηρήθηκαν, επιπλέον, **λιγότερα τροχαία ατυχήματα** έως 76% σε σύγκριση με την ίδια περίοδο τα έτη 2018-2019. Τα αποτελέσματα αυτά συμφωνούν με τα ευρήματα της μελέτης των Katrakazas et al. (2020), όπου αποδείχθηκε ότι η μείωση της οδήγησης στην Ελλάδα και στο Βασίλειο της Σαουδικής Αραβίας κατά 74% και 75%, αντίστοιχα, οδήγησε σε σημαντική μείωση των τροχαίων ατυχημάτων, αλλά και αύξηση της μέσης ταχύτητας κατά 6-11% σε σύγκριση με το 2019.

2.2.2 Αλλαγή Χαρακτηριστικών Μετακίνησης λόγω Πανδημίας

Παράγοντες που καθόριζαν την επιλογή μέσου μεταφοράς στην προ-κορονοϊού εποχή, όπως η εξοικονόμηση χρόνου, η άνεση και το κόστος, δεν αποτελούν προτεραιότητα κατά τη διάρκεια της πανδημίας (Abdullah et al., 2020). Προτεραιότητα δείχνει να είναι η **αποφυγή εστιών μετάδοσης και συνωστισμού.** Έντονη μείωση στην συνολική κινητικότητα παρατηρήθηκε στη Metro Manila των Φιλιππίνων, περισσότερο έντονη όμως, στη χρήση των μέσων μαζικής μεταφοράς που σημειώθηκε ποσοστό 74,5% (Hasselwander et al. 2021). Στη Σουηδία, στις τρεις πόλεις με τον περισσότερο πληθυσμό, (Stockholm, Skåne και Västra Götaland) παρατηρήθηκαν ποσοστά μείωσης της χρήσης των μέσων μαζικής μεταφοράς 60% για τις δύο πρώτες και 40% για την τελευταία (Jenelius et al. 2020). Στην Ινδία, με βάση το ερωτηματολόγιο των Pawar et al. (2020), το 41,65% απάντησαν πως σταμάτησαν τις μετακινήσεις, και το 51,31% συνέχισαν να χρησιμοποιούν το ίδιο μέσο. Ο σημαντικότερος λόγος αυτών των αλλαγών είναι η αίσθηση ασφάλειας. Στην Κρήτη, συγκεκριμένα στα Χανιά και στο Ρέθυμνο, παρατηρείται μείωση με ποσοστό 30% των μέσων μαζικής μεταφοράς και υποχώρηση της χρήσης ιδιωτικών οχημάτων κατά 10,7% κατά την περίοδο των περιοριστικών μέτρων (Tarasi et al. 2021). Στην Σικελία, κατά το Μάρτη του 2020 μειώθηκε η χρήση των μέσων μαζικής μεταφοράς κατά 93%. Μετά την άρση των περιορισμών, ενώ η κίνηση σχεδόν επανήλθε και ενώ υπήρχαν μέτρα προστασίας, όπως η χρήση προστατευτικής μάσκας, τα μέσα μαζικής μεταφοράς αποφεύγονταν, καθώς υπήρχε έντονη ανησυχία για την έκθεση στον ιό, οπότε προτιμούνταν η χρήση ιδιωτικού οχήματος ή εναλλακτικές όπως το περπάτημα ή το ποδήλατο. (Campisi et al., 2020).

Βέβαια, κατά τη διάρκεια των περιορισμών μειώθηκε σημαντικά **η μέση απόσταση και άλλαξε ο σκοπός των μετακινήσεων**, που έτσι δικαιολογείται η αύξηση των πεζών και των ποδηλατιστών. Σύμφωνα με το ερωτηματολόγιο των Shamshiripour et al. (2020), προκύπτει πως οι ηλεκτρονικές αγορές για τις βασικές προμήθειες αυξήθηκαν κατά 550% σε σχέση με την προ-κορονοϊού εποχή. Επίσης,

αξιοσημείωτα είναι τα ποσοστά των συμμετεχόντων που πριν την επιβολή των περιοριστικών μέτρων δεν είχαν εμπειρία με την τηλεργασία (71%) και την ηλεκτρονική αγορά των βασικών προμηθειών (55%).

Οι Padmanabhan et al. (2021) παρατήρησαν μείωση στις **μετακινήσεις με ποδήλατα** στη Νέα Υόρκη, στη Βοστόνη και στο Σικάγο κατά την περίοδο αύξησης των κρουσμάτων. Μείωση στις ενοικιάσεις ποδηλάτων παρατηρείται και στο Σικάγο από τους Hu et al. (2021), με ποσοστό 32,5% για το διάστημα Μαρτίου-Ιουνίου σε σύγκριση με την αντίστοιχη περίοδο του 2019. Η διαφορά όμως σε σχέση με τα άλλα μέσα παρατηρείται στο ρυθμό επαναφοράς κατά τον Ιούλιο, που σημειώθηκε ποσοστό 284,0%, σε σχέση με την οδήγηση και το περπάτημα που σημείωσαν 137,5% και 131,6%, αντίστοιχα. Οι Khaddar et al. (2021) υποστηρίζουν πως το ποδήλατο και το περπάτημα συμβάλουν θετικά στην ψυχολογία των ανθρώπων, που είναι ιδιαίτερα σημαντικό κατά την περίοδο περιορισμών. Οι Joseph Molloy et al. (2020) υποστηρίζουν πως σημαντικός παράγοντας της αύξησης των ποδηλατιστών είναι η σημαντική μείωση των οχημάτων στους δρόμους, που καθιστά πιο φιλική και ασφαλής την κυκλοφορία τους. Εξαιτίας της μεγάλης αυτής στροφής προς το περπάτημα και το ποδήλατο που παρατηρήθηκε σε πάνω από 500 πόλεις στο διάστημα μεταξύ Μαρτίου και Αυγούστου, που ήταν η έξαρση του πρώτου κύματος του ιού, οι Combs et al. (2020) υποστηρίζουν πως πρέπει να γίνουν αλλαγές ώστε στην κυκλοφορία να ενταχθούν και αυτοί οι τρόποι μετακίνησης, για να είναι πιο ασφαλείς όλοι οι χρήστες.

Όπως έχει ήδη αναφερθεί, καθώς άλλαξε η καθημερινότητα, άλλαξαν **οι τρόποι και οι ανάγκες για μετακίνηση**. Αξίζει να ερευνηθεί κατά πόσο αυτές οι αλλαγές έχουν διατηρηθεί στην εξέλιξη της πανδημίας.

Πολλές έρευνες, οι οποίες βασίζονται σε ερωτηματολόγια, παρουσιάζουν μια σημαντική μετάβαση από τη χρήση δημοσίων μέσων μεταφοράς σε **ιδιωτικά και ατομικά μέσα** (Abdullah et al., 2020; Tan and Ma, 2020; Shakibaei et al., 2021; Shamshiripour et al., 2020). Πιο συγκεκριμένα, το 96,5% των ερωτηθέντων στην Κωνσταντινούπολη (Shakibaei et al., 2021) και το 93% των ερωτηθέντων στο Σικάγο (Shamshiripour et al., 2020) συσχετίζουν τις δημόσιες συγκοινωνίες ως μέσο με υψηλό κίνδυνο έκθεσης στον ιό. Μία άλλη έρευνα βασισμένη σε ερωτηματολόγιο ανέφερε ότι η χρήση του αυτοκινήτου ως συνεπιβάτης μειώθηκε κατά 80% και η χρήση των δημόσιων συγκοινωνιών κατά 90% (de Haas et al., 2020).

Τα παραπάνω ευρήματα συνάδουν με τα αποτελέσματα των ερευνών των Aloi et al. (2020) και Bucksy (2020), οι οποίες μέσω ανάλυσης δεδομένων διαπίστωσαν **πτώση της χρήσης των μέσων μαζικής μεταφοράς** κατά 90%. Στη Σουηδία, σημειώθηκε μείωση των επιβατών της δημόσιας συγκοινωνίας κατά 60% στη Στοκχόλμη και κατά 40% στη Västra Götaland (Jenelius and Cebecauer, 2020). Ταυτόχρονα, στην ίδια έρευνα παρατηρήθηκε μία μετάβαση από εισιτήρια μεγάλης διάρκειας σε μονά ή πιο ευέλικτα εισιτήρια.

Η έρευνα των Kim et al. (2017) επεσήμανε ότι κατά την επιδημία του MERS το 2015 οι τουρίστες επέλεγαν τα λεωφορεία **έναντι του υπόγειου σιδηρόδρομου**, επειδή

τον συνέδεαν με αυξημένο κίνδυνο μετάδοσης της νόσου. Στο ίδιο πλαίσιο, κατά τη διάρκεια του πρώτου «κύματος» στην Κίνα, η επιλογή μέσου υπαγορευόταν από την πιθανότητα μόλυνσης. Οι ερωτηθέντες των Tan and Ma (2020) δήλωσαν ότι απέφευγαν τον υπόγειο σιδηρόδρομο και επέλεγαν τον επίγειο λόγω καλύτερου εξαερισμού.

2.3 Σύνοψη

Από την παραπάνω ανασκόπηση φαίνεται ότι η πανδημία έχει απασχολήσει έντονα την παγκόσμια επιστημονική κοινότητα. Πολλές έρευνες και μελέτες έχουν πραγματοποιηθεί προκειμένου να εξεταστεί η επίδραση του COVID-19 και των περιοριστικών μέτρων που λήφθηκαν για την αντιμετώπισή του στην κινητικότητα, την οδηγική συμπεριφορά, την οδική ασφάλεια, τον σκοπό και τον τρόπο της μετακίνησης. Παρατηρείται ότι πολλές από τις έρευνες εστιάζουν στην συγκριτική ανάλυση της κατάστασης πριν και μετά την πανδημία ενώ ένα μεγάλο μέρος των ερευνών αναλύει τις προτιμήσεις των χρηστών ως προς τα μέσα μεταφοράς. Αξίζει να σημειωθεί ότι δεν υπάρχει ικανοποιητικός αριθμός ερευνών για την Ελλάδα.

Παρακάτω παρατίθεται οι Πίνακες 1 και 2 που παρουσιάζουν συνοπτικά τα στοιχεία και αποτελέσματα των παραπάνω προαναφερθέντων μελετών.

Πίνακας 1. Σύνοψη ερευνών

Έρευνα	Χώρα μελέτης	Μέθοδοι Ανάλυσης	Αποτελέσματα
Abdullah et al (2020)	Χώρες από όλο τον κόσμο	Περιγραφική ανάλυση,Ποσοτικές συγκριτικές αναλύσεις,Μη παραμετρικές δοκιμές (όπως McNemar-Bowker, Wilcoxon signed-rank) Λογιστική παλινδρόμηση	Προτιμήσεις χρηστών ως προς τα μέσα μεταφοράς κατά τη διάρκεια της πανδημίας
Aloi (2020)	Santander, Ισπανία	Πίνακες προέλευσης-προορισμού, Ποσοστιαίες μεταβολές κινητικότητας πριν και μετά	Σύγκριση δεδομένων κινητικότητας με το 2018
Badr (2020)	Ηνωμένες Πολιτείες	Πίνακες προέλευσης-προορισμού, ομαδοποίηση των μεταβλητών, υπολογισμός συσχέτισης	Υπολογισμός συσχέτισης της κινητικότητας με την εξάπλωση των ιού (συνυπολογίζοντας κρούσματα και θανάτους)
Beria (2021)	Ιταλία	Πίνακες προέλευσης-προορισμού, Ποσοστιαίες μεταβολές εσωτερικών ταξιδιών πριν και μετά	Ανάλυση κινητικότητας ανάλογα με την περιφέρεια και σε αντιστοιχία με τα μέτρα περιορισμού
Bucsky (2020)	Βουδαπέστη, Ουγγαρία	Ποσοστιαίες μεταβολές κινητικότητας πριν και μετά	Σύγκριση δεδομένων κινητικότητας με το 2017 και το 2018
Campisi (2020)	Σικελία	Ποσοστιαίες μεταβολές κινητικότητας πριν και μετά	Προτιμήσεις χρηστών ως προς τα μέσα μεταφοράς κατά τη διάρκεια της πανδημίας
Combs (2021)	-	Ποσοστιαίες μεταβολές κινητικότητας πριν και μετά	Προτιμήσεις χρηστών ως προς τα μέσα μεταφοράς κατά τη διάρκεια της πανδημίας
DeHaas (2020)	Ολλανδία	Ερωτηματολόγιο και έλεγχος χ2	Προτιμήσεις χρηστών ως προς τα μέσα μεταφοράς κατά τη διάρκεια της πανδημίας
Hadjidemetriou (2020)	Ηνωμένο Βασίλειο	Λογιστικό μοντέλο	Υπολογισμός συσχέτισης της κινητικότητας με τον αριθμό των θανάτων
Hasselwander (2021)	Metro Manila,Φιλιππίνες	Ποσοστιαίες μεταβολές κινητικότητας πριν και μετά	Σύγκριση δεδομένων κινητικότητας πριν και μετά την εφαρμογή των περιοριστικών μέτρων
Hu (2021)	Σικάγο	Λογιστικό μοντέλο	Επιρροή του Covid-19 στις ενικοιάσεις ποδηλάτων
Jenelius (2020)	Σουηδία	Ποσοστιαίες μεταβολές κινητικότητας πριν και μετά	Σύγκριση δεδομένων κινητικότητας με το 2019
J.Molloy (2021)	Ελβετία	Ποσοστιαίες μεταβολές κινητικότητας πριν και μετά	Σύγκριση δεδομένων κινητικότητας με το 2019
Katrakazas (2020)	Ελλάδα και Σαουδική Αραβία	Διερευνητική ανάλυση μεταξύ χωρών	Σύγκριση κινητικότητας και ατυχημάτων

Πίνακας 2. Σύνοψη ερευνών

Khaddar (2021)	Kelowna,Καναδάς	LSOL model και έλεγχος χ2	Επιρροή των διαφορετικών μετακινήσεων στην ψυχολογία
Kim (2017)	Σεούλ, Βόρεια Κορέα	Ποσοστιαίες μεταβολές χρήσης καρτών MMM πριν και μετά	Σύνδεση κινητικότητας με το επίπεδο σταθερότητας
Maiti (2021)	Ηνωμένες Πολιτείες	Λογιστικό μοντέλο	Επιρροή του Covid-19 στις μετακινήσεις
Muley (2021)	Ντόχα, Κατάρ	Ποσοστιαίες μεταβολές κινητικότητας πριν και μετά	Ανάλυση επιρροής διαφόρων πανδημιών στην κινητικότητα
Nouvellet (2021)	Ηνωμένο Βασίλειο	Λογιστικό μοντέλο	Συσχέτιση της κινητικότητας με την εξάπλωση του Covid-19
Padmanabh an (2021)	Νέα Υόρκη, Βοστόνη και Σικάγο	Συσχέτιση και έλεγχος χ2	Συσχέτιση του αριθμού των κρουσμάτων με τις μετακινήσεις με ποδήλατο
Parr (2020)	Φλόριντα, ΗΠΑ	Στατιστικός έλεγχος t	Ανάλυση κινητικότητας σε αντιστοιχία με τα μέτρα περιορισμού
Pullano (2020)	Γαλλία	Γραμμική παλινδρόμηση πολλών μεταβλητών	Σύγκριση κινητικότητας Φεβρουαρίου 2020 με Απρίλιο 2020
Saladie (2020)	Tarragona, Ισπανία	Έλεγχος χ2	Σύγκριση κινητικότητας και ατυχημάτων
Santamaria (2020)	Αυστρία, Βέλγιο, Βουλγαρία, Κροατία, Δανία, Εσθονία, Φιλανδία, Γαλλία, Γερμανία, Ιταλία,Πορτογαλία,Σλοβενία, Ισπανία, Σουηδία και Νορβηγία	Πίνακες προέλευσης-προορισμού.	Σύγκριση ευρωπαϊκών χωρών
Shakibaei (2021)	Κωνσταντινούπολη, Τουρκία	Ερωτηματολόγιο	Προτιμήσεις χρηστών κατά την πανδημία
Shamshiripour (2020)	Σικάγο, ΗΠΑ	Ερωτηματολόγιο	Προτιμήσεις χρηστών κατά την πανδημία
Tan and Ma(2020)	Κίνα	Ερωτηματολόγιο και λογιστική παλινδρόμηση	Παράγοντες που καθορίζουν την επιλογή μέσου κατά την πανδημία
Tarasi (2021)	Κρήτη, Ελλάδα	Ερωτηματολόγιο	Επιρροή του Covid-19 στις μετακινήσεις
Thakkar (2020)	Washington	Συσχέτιση κρουσμάτων με τα περιοριστικά μέτρα	Μέτρα προστασίας για τον περιορισμό της εξάπλωσης του Covid-19

3. ΘΕΩΡΗΤΙΚΟ ΥΠΟΒΑΘΡΟ

3.1 Εισαγωγή

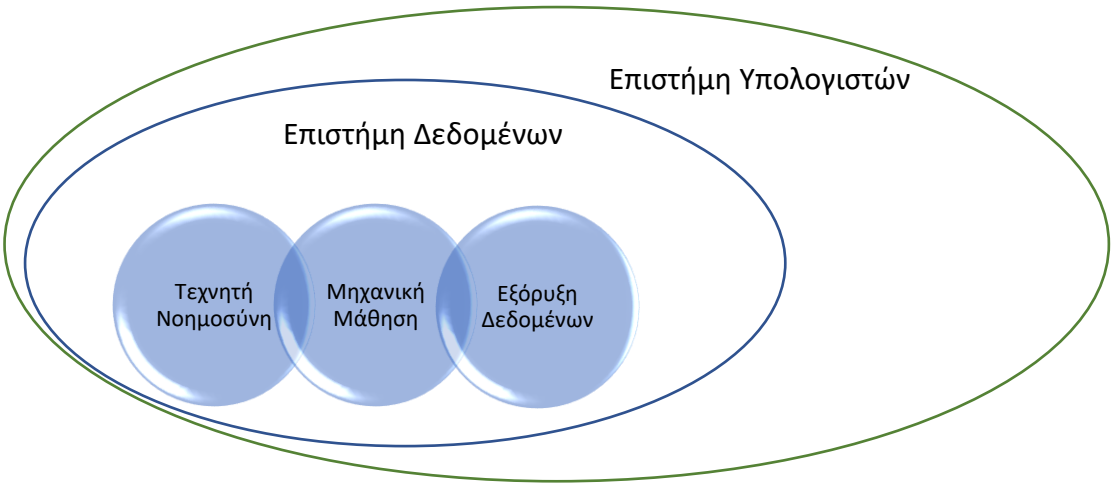
Στο παρόν κεφάλαιο αναλύεται το θεωρητικό υπόβαθρο στο οποίο στηρίχθηκε η παρούσα Διπλωματική Εργασία. Πιο συγκεκριμένα, παρουσιάζεται η ανάλυση και πρόβλεψη με χρήση δένδρων παλινδρόμησης , οι τύποι των μοντέλων, ο τρόπος αξιολόγησης των παραγόμενων προβλέψεων και τα κριτήρια αποδοχής τους.

3.2 Μηχανική Μάθηση

3.2.1 Γενικά

Μηχανική μάθηση (Machine Learning) είναι υποπεδίο της επιστήμης των υπολογιστών που αναπτύχθηκε από τη μελέτη της αναγνώρισης προτύπων και της υπολογιστικής θεωρίας μάθησης στην τεχνητή νοημοσύνη (Εικόνα 3). **Η μηχανική μάθηση διερευνά τη μελέτη και την κατασκευή αλγορίθμων που μπορούν να μαθαίνουν από τα δεδομένα και να κάνουν προβλέψεις σχετικά με αυτά.**

Ουσιαστικά, με τη μηχανική εκμάθηση ο υπολογιστής αναλύει μεγάλα δεδομένα, εξάγει αυτόματα πληροφορίες και τις χρησιμοποιεί για να κάνει προβλέψεις, να αποκρυπτογραφήσει εάν η πρόβλεψη ήταν σωστή και, αν είναι λανθασμένη, να μάθει από αυτήν για να κάνει στο μέλλον μια πιο σωστή πρόβλεψη.



Εικόνα 3. Επιστήμη Υπολογιστών

Το 1959, ο πρωτοπόρος σχεδιαστής παιχνιδιών Άρθουρ Σάμιουελ (Arthur Samuel, 1959) όρισε ως μηχανική μάθηση "Το πεδίο μελέτης όπου δίνει στους υπολογιστές την δυνατότητα να μαθαίνουν χωρίς να έχουν προγραμματιστεί".

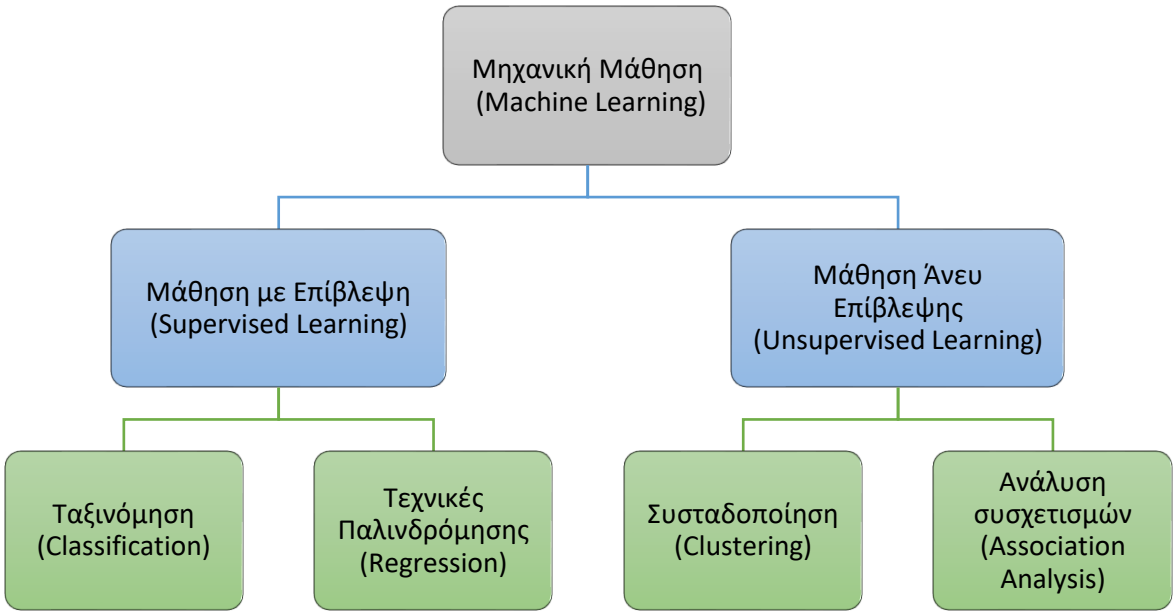
Το 1997 ο Τομ Μ. Μίτσελ (Tom M. Mitchell, 1997) έδωσε ένα πιο επίσημο ορισμό ο οποίος χρησιμοποιείται ευρέως: "Ένα πρόγραμμα υπολογιστή λέγεται ότι μαθαίνει από μια εμπειρία Ε σε σχέση μια σειρά από έργα Τ και μια μέτρηση της απόδοσης Ρ ή οποία βελτιώνεται με την εμπειρία Ε".

Εν γένει, ο τομέας της Μηχανικής Μάθησης αναπτύσσει τρεις τρόπους μάθησης, ανάλογους με τους τρόπους με τους οποίους μαθαίνει ο άνθρωπος: μάθηση με επίβλεψη, μάθηση χωρίς επίβλεψη και ενισχυτική μάθηση (Εικόνα 4). Πιο αναλυτικά:

Η Μάθηση με Επίβλεψη (Supervised Learning) είναι η διαδικασία *όπου ο αλγόριθμος κατασκευάζει μια συνάρτηση που απεικονίζει δεδομένες εισόδους (σύνολο εκπαίδευσης) σε γνωστές επιθυμητές εξόδους, με απώτερο στόχο τη γενίκευση της συνάρτησης αυτής και για εισόδους με άγνωστη έξοδο*. Χρησιμοποιείται σε προβλήματα ταξινόμησης (Classification) και παλινδρόμησης (Regression).

Στη Μάθηση χωρίς Επίβλεψη (Unsupervised Learning), *ο αλγόριθμος κατασκευάζει ένα μοντέλο για κάποιο σύνολο εισόδων υπό μορφή παρατηρήσεων χωρίς να γνωρίζει τις επιθυμητές εξόδους*. Χρησιμοποιείται σε προβλήματα ανάλυσης συσχετισμών (Association Analysis) και την ομαδοποίηση (Clustering).

Τέλος, στην Ενισχυτική Μάθηση (Reinforcement Learning), *ο αλγόριθμος μαθαίνει μια στρατηγική ενεργειών μέσα από άμεση αλληλεπίδραση με το περιβάλλον*. Χρησιμοποιείται κυρίως σε προβλήματα Σχεδιασμού (Planning), όπως για παράδειγμα ο έλεγχος κίνησης ρομπότ και η βελτιστοποίηση εργασιών σε εργοστασιακούς χώρους.



Εικόνα 4. Τύποι Μηχανικής Μάθησης

Η Μηχανική μάθηση εφαρμόζεται σε μια σειρά από υπολογιστικές εργασίες, όπου τόσο ο σχεδιασμός όσο και ο ρητός προγραμματισμός των αλγορίθμων είναι ανέφικτος. Εφαρμόζεται σε διάφορους επιστημονικούς κλάδους, όπως στις μεταφορές, την ιατρική, την οικονομία, τη ρομποτική κ.α.

Τα τελευταία χρόνια, οι τεχνικές Μηχανικής Μάθησης έχουν γίνει μέρος των έξυπνων μεταφορών. Μέσω της Βαθιάς Μάθησης (Deep Learning), η Μηχανική Μάθηση διερεύνησε τις πολύπλοκες αλληλεπιδράσεις των δρόμων, των αυτοκινητοδρόμων, της κυκλοφορίας, των περιβαλλοντικών στοιχείων, των ατυχημάτων και ούτω καθεξής. Η Μηχανική Μάθηση έχει επίσης μεγάλες δυνατότητες στην καθημερινή διαχείριση της κυκλοφορίας και στη συλλογή δεδομένων κυκλοφορίας.

Μέσω της ανάλυσης Συσταδοποίησης (Clustering), η οποία είναι μια τεχνική μηχανικής μάθησης χωρίς επίβλεψη, παρατηρείται η εφαρμογή της σε διάφορες κατηγορίες στο σχεδιασμό των μεταφορών, όπως η παραγωγή ταξιδιών, ο διαχωρισμός ζωνών κυκλοφορίας και η κατανομή ταξιδιών. Χρησιμοποιούνται επίσης μοντέλα χρονοσειρών για την πρόβλεψη της κυκλοφοριακής ροής καθώς και αλγόριθμοι για την μοντελοποίηση της επιλογής του τρόπου μετακίνησης.

Μέσω των Νευρωνικών Δικτύων (Artificial Neural Networks), πολλές μελέτες έχουν επικεντρωθεί στην μοντελοποίηση της συμπεριφοράς κατά την οδήγηση. Χαρακτηριστικές είναι οι μελέτες των Yang et al.(1992) και Dougherty και Joint (1992), οι οποίοι μοντελοποίησαν τη συμπεριφορά του οδηγού όταν λαμβάνει στρατηγικές και ενστικτώδεις αποφάσεις, καθώς και των Hunt και Lyons (1994), οι οποίοι χρησιμοποίησαν νευρωνικά δίκτυα για να μοντελοποιήσουν τη συμπεριφορά του οδηγού κατά την αλλαγή ταχύτητας και λωρίδας κυκλοφορίας σε έναν αυτοκινητόδρομο.

3.3 Δένδρα Απόφασης

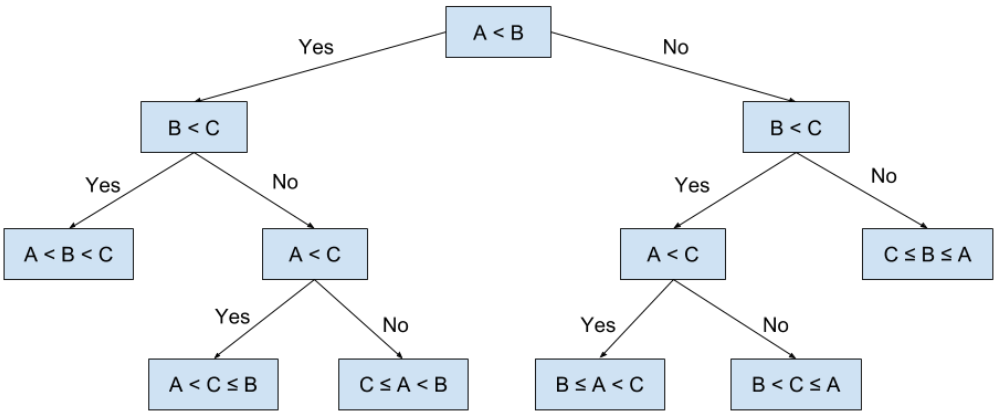
Τα **Δένδρα Απόφασης-ΔΑ** (Decision Trees) είναι ο γνωστότερος αλγόριθμος επιβλεπόμενης Επαγωγικής Μάθησης και έχει εφαρμοστεί με επιτυχία σε πολλούς τομείς όπου απαιτείται ταξινόμηση: ενδεικτικά, στην αναγνώριση προσώπων σε εικόνες, στην ιατρική για διάγνωση περιστατικών, για προβλέψεις απαραίτητες στη διαφήμιση, για προώθηση προϊόντων και, γενικότερα, για εξόρυξη γνώσης. Μπορούν να χρησιμοποιηθούν για την επίλυση προβλημάτων τόσο παλινδρόμησης όσο και ταξινόμησης.

Τα Δένδρα Αποφάσεων έχουν πολλές εφαρμογές στον τομέα μεταφορών όπως η χρήση δέντρων αποφάσεων για την πρόβλεψη αποφάσεων επιλογής τρόπου μεταφοράς και η ανάπτυξη μοντέλων ζήτησης ταξιδιού. Υπάρχουν επίσης εφαρμογές των Δένδρων Αποφάσεων για την πρόβλεψη της επιβατικής κίνησης σε υπόγειους

και μη σιδηροδρομικούς σταθμούς καθώς και πρόβλεψη της επιβατικής κίνησης για λεωφορειακές γραμμές.

Ο αλγόριθμος ΔΑ οδηγεί στη δημιουργία μιας δενδροειδούς μορφής που τα φύλλα της αποτελούν **κατηγορίες ταξινόμησης** (classes) (Εικόνα 5). Η δενδροειδής αυτή μορφή μπορεί να αναγνωστεί και ως ένα σύνολο κανόνων που καλούνται **κανόνες ταξινόμησης** (classification rules) και να δώσει μια πειστική απάντηση στο ερώτημα:

Πώς μπορεί μία μηχανή να δημιουργήσει γενικούς κανόνες από συγκεκριμένες παρατηρήσεις και πόσο αξιόπιστοι είναι αυτοί οι κανόνες στην πράξη;



Εικόνα 5. Δένδρα Αποφάσεων (Decision Trees)

Βασικές προϋποθέσεις για τη λειτουργία ενός αλγόριθμου επαγωγικής μάθησης είναι:

- Καθορισμός ενός **συνόλου χαρακτηριστικών** (features set - **FS**) ως των προϋποθέσεων του επιδιωκόμενου προς εξαγωγή κανόνα ταξινόμησης:

$$\boldsymbol{FS = \{F_1, F_2, F_3, \dots F_{|FS|}\}}$$

Σχέση 1. Σύνολο χαρακτηριστικών για τον αλγόριθμο

- Ύπαρξη προκαθορισμένων **διακριτών κατηγοριών ταξινόμησης** (classes – **C**) ως στόχου του διαχωρισμού τον οποίο θα επιδιώξει ο αλγόριθμος και, στη συνέχεια, ως συμπερασμάτων (conclusions) των κανόνων στους οποίους θα οδηγήσει η αναγνώριση της δενδροειδούς μορφής που θα αναπτύξει ο

αλγόριθμος:

$$C = \{c_1, c_2, c_3, \dots c_{|C|}\}$$

Σχέση 2. Διακριτές κατηγορίες ταξινόμησης

- ύπαρξη επαρκούς αριθμού **δειγμάτων** που θα προκύψουν από παρατηρήσεις και θα χρησιμοποιηθούν για τη δημιουργία του εκπαιδευτικού συνόλου (training set - **TS**).

Μερικά πλεονεκτήματα των δέντρων απόφασης είναι:

- Απλό στην κατανόηση και στην ερμηνεία. Τα δέντρα μπορούν να απεικονιστούν.
- Απαιτεί λίγη προετοιμασία δεδομένων. Άλλες τεχνικές συχνά απαιτούν κανονικοποίηση δεδομένων, πρέπει να δημιουργηθούν εικονικές μεταβλητές και να αφαιρεθούν κενές τιμές. Σημειώνεται ωστόσο ότι αυτή η ενότητα δεν υποστηρίζει τιμές που λείπουν.
- Ικανά να χειριστούν τόσο αριθμητικά όσο και κατηγορικά δεδομένα.
- Ικανά να χειριστούν προβλήματα πολλαπλών εξόδων.
- Παρέχουν δυνατότητα επικύρωσης ενός μοντέλου με τη χρήση στατιστικών δοκιμών. Αυτό καθιστά δυνατό τον υπολογισμό της αξιοπιστίας του μοντέλου.
- Αποδίδει καλά ακόμα κι αν οι παραδοχές του παραβιάζονται κάπως από το πραγματικό μοντέλο από το οποίο δημιουργήθηκαν τα δεδομένα.
- Τα δέντρα απόφασης δεν είναι ευαίσθητα σε ακραίες τιμές, καθώς οι ακραίες τιμές δεν προκαλούν ποτέ μεγάλη μείωση στο υπολειπόμενο άθροισμα τετραγώνων (RSS), επειδή δεν εμπλέκονται ποτέ στη διαίρεση.

Κάποια από τα μειονεκτήματα των δέντρων απόφασης περιλαμβάνουν:

- Υπερβολικά πολύπλοκα δέντρα που δεν γενικεύουν καλά τα δεδομένα μπορεί να οδηγήσουν σε υπερπροσαρμογή (Overfitting). Μηχανισμοί όπως το κλάδεμα (Pruning), ο καθορισμός του ελάχιστου αριθμού δειγμάτων που απαιτούνται σε έναν κόμβο φύλλων ή ο καθορισμός του μέγιστου βάθους του δέντρου είναι απαραίτητοι για την αποφυγή αυτού του προβλήματος.
- Τα δέντρα αποφάσεων μπορεί να είναι ασταθή επειδή μικρές παραλλαγές στα δεδομένα μπορεί να έχουν ως αποτέλεσμα τη δημιουργία ενός εντελώς διαφορετικού δέντρου.
- Σε περίπτωση που κυριαρχούν ορισμένες τάξεις, τα δένδρα αποφάσεων που προκύπτουν μπορεί να είναι προκατειλημμένα. Επομένως, συνιστάται η εξισορρόπηση του συνόλου δεδομένων πριν από την προσαρμογή με το δέντρο αποφάσεων.

- Υπάρχουν έννοιες που είναι δύσκολο να εξεταστούν επειδή τα δέντρα αποφάσεων δεν τις εκφράζουν εύκολα, όπως προβλήματα πύλης XOR, ισοτιμίας ή πολυπλέκτη (multiplexer problems).

Όπως αναφέρθηκε παραπάνω, οι αλγόριθμοι των Δένδρων Αποφάσεων είναι ιδανικοί για την επίλυση προβλημάτων ταξινόμησης και παλινδρόμησης. Τα δένδρα **παλινδρόμησης** χρησιμοποιούνται όταν η εξαρτημένη μεταβλητή είναι συνεχής ή ποσοτική και τα δένδρα **ταξινόμησης** χρησιμοποιούνται όταν η εξαρτημένη μεταβλητή είναι κατηγορική ή ποιοτική. Στη παρούσα Διπλωματική εργασία θα χρησιμοποιηθούν τα **Δένδρα Παλινδρόμησης**.

Υπάρχουν διάφοροι αλγόριθμοι για δένδρα αποφάσεων όπως:

- **ID3 (Iterative Dichotomiser 3)**: Αναπτύχθηκε το 1986 από τον Ross Quinlan. Ο αλγόριθμος δημιουργεί ένα δέντρο πολλαπλών δρόμων, βρίσκοντας για κάθε κόμβο (δηλαδή με άπληστο τρόπο) το κατηγορηματικό χαρακτηριστικό που θα αποφέρει το μεγαλύτερο κέρδος πληροφοριών για τους κατηγορικούς στόχους. Τα δέντρα αναπτύσσονται στο μέγιστο μέγεθός τους και στη συνέχεια εφαρμόζεται συνήθως ένα βήμα κλαδέματος για τη βελτίωση της ικανότητας του δέντρου να γενικεύει σε αόρατα δεδομένα.

- **C4.5** : είναι ο διάδοχος του ID3 και αφαίρεσε τον περιορισμό ότι τα χαρακτηριστικά πρέπει να είναι κατηγορηματικά ορίζοντας δυναμικά ένα διακριτό χαρακτηριστικό (με βάση αριθμητικές μεταβλητές) που χωρίζει την τιμή συνεχούς χαρακτηριστικού σε ένα διακριτό σύνολο διαστημάτων. Το C4.5 μετατρέπει τα εκπαιδευμένα δέντρα (δηλαδή την έξοδο του αλγορίθμου ID3) σε σύνολα κανόνων if-then. Αυτή η ακρίβεια κάθε κανόνα στη συνέχεια αξιολογείται για να καθοριστεί η σειρά με την οποία θα πρέπει να εφαρμοστούν. Το κλάδεμα γίνεται αφαιρώντας την προϋπόθεση ενός κανόνα εάν η ακρίβεια του κανόνα βελτιωθεί χωρίς αυτόν.

- **C5.0 :** είναι η πιο πρόσφατη έκδοση του Quinlan με άδεια αποκλειστικής χρήσης. Χρησιμοποιεί λιγότερη μνήμη και δημιουργεί μικρότερα σύνολα κανόνων από το C4.5 ενώ είναι πιο ακριβές.

- **CART (Classification and Regression Trees)** : Μοιάζει πολύ με το C4.5, αλλά διαφέρει στο ότι υποστηρίζει αριθμητικές μεταβλητές στόχου (παλινδρόμηση) και δεν υπολογίζει σύνολα κανόνων. Το CART κατασκευάζει δυαδικά δέντρα χρησιμοποιώντας το χαρακτηριστικό και το όριο που αποδίδουν το μεγαλύτερο κέρδος πληροφοριών σε κάθε κόμβο.

Για την παρούσα Διπλωματική εργασία θα χρησιμοποιηθεί η δωρεάν βιβλιοθήκη Scikit-learn, της γλώσσας προγραμματισμού Python , η οποία υλοποιεί τον αλγόριθμο **CART** για τα Δένδρα Παλινδρόμησης

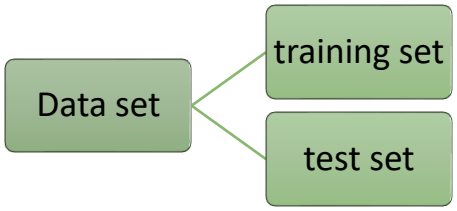
3.5 Επεξεργασία, Οπτικοποίηση και Αξιολόγηση του Αλγορίθμου

3.5.1 Επεξεργασία

Μετά την εισαγωγή του αρχείου δεδομένων ακολουθείται μια διαδικασία προετοιμασίας των δεδομένων όπως η ρύθμιση των σωστών μορφών δεδομένων, η αντιμετώπιση τιμών και ακραίων τιμών που λείπουν, εξάλειψη των διπλότυπων κλπ. Για αυτές τις διαδικασίες χρησιμοποιείται το πακέτο Pandas.

Ύστερα αυτό που απαιτείται είναι να διαιρεθούν οι στήλες του αρχείου σε δύο τύπους μεταβλητών: εξαρτημένη (ή μεταβλητή στόχου) και ανεξάρτητη μεταβλητή (ή μεταβλητές χαρακτηριστικών).

Στην συνέχεια, ακολουθεί ο διαχωρισμός του συνόλου δεδομένων σε δύο ξεχωριστά σύνολα **training set & test set** (Εικόνα 6).



Εικόνα 6. Χωρισμός δεδομένων σε training set και test set.

Training set: αυτά είναι τα δεδομένα τα οποία χρησιμοποιούνται για την κατασκευή του μοντέλου και συγκεκριμένα για την παρούσα Διπλωματική χρησιμοποιούνται για την προσαρμογή του αλγόριθμου CART.

Test set: αυτά είναι τα δεδομένα τα οποία χρησιμοποιούνται για να ελεγχθεί πώς αποδίδει το μοντέλο σε νέα δεδομένα, όπως θα έκανε σε μια πραγματική κατάσταση. Παρέχουν μια αμερόληπτη αξιολόγηση του μοντέλου που προσαρμόστηκε από τα δεδομένα του training set.

Η διαδικασία διαχωρισμού σε training set & test set χρησιμοποιείται για την εκτίμηση της απόδοσης των αλγορίθμων μηχανικής μάθησης όταν χρησιμοποιούνται για να κάνουν προβλέψεις σε δεδομένα που δεν χρησιμοποιούνται για την εκπαίδευση του μοντέλου. Η κύρια ιδέα του διαχωρισμού του συνόλου δεδομένων σε ένα σύνολο επικύρωσης (test set) είναι να αποτραπεί η υπερβολική προσαρμογή ή υποπροσαρμογή του μοντέλου μας.

Η υποπροσαρμογή είναι συνήθως η συνέπεια της αδυναμίας ενός μοντέλου να ενθυλακώσει τις σχέσεις μεταξύ των δεδομένων. Για παράδειγμα, αυτό μπορεί να συμβεί όταν γίνεται προσπάθεια να αναπαρασταθούν μη γραμμικές σχέσεις με ένα γραμμικό μοντέλο.

Η υπερπροσαρμογή εμφανίζεται όταν ένα στατιστικό μοντέλο προσαρμόζεται ακριβώς με τα δεδομένα εκπαίδευσης του. Όταν συμβαίνει αυτό, ο αλγόριθμος δεν μπορεί να αποδώσει με ακρίβεια έναντι δεδομένων που δεν έχει ξαναδεί.

Επιθυμητό είναι να δημιουργούνται δέντρα που είναι ισορροπημένα και με τα λιγότερα επίπεδα (μέγιστο βάθος). Η δημιουργία ενός δέντρου σταματά οπωσδήποτε όταν όλα τα δεδομένα του συνόλου εκπαίδευσης κατηγοριοποιούνται πλήρως. Μπορεί όμως να είναι απαραίτητο να σταματήσει νωρίτερα για να αποφευχθούν π.χ. μεγάλα δέντρα. Το πότε ή που θα σταματήσει είναι θέμα συναλλαγής μεταξύ ακρίβειας (accuracy) και απόδοσης (performance) του αλγορίθμου. Επίσης πρώιμος τερματισμός μπορεί να γίνει για την αποφυγή του φαινόμενου της προσαρμογής (overfitting). Τέλος μπορεί να προχωρήσει σε μεγαλύτερα δέντρα αν είναι γνωστό ότι υπάρχουν κατηγορίες δεδομένων που δεν αντιπροσωπεύονται στο σύνολο της εκπαίδευσης.

Τέλος, ακολουθεί το Κλάδεμα του Δέντρου - Pruning. Ύστερα από την κατασκευή ενός πλήρους δέντρου, συνήθως, προκύπτει ένα μεγάλο και περίπλοκο (με πολλά κλαδιά) δέντρο, το οποίο είναι δύσκολο να ερμηνευτεί. Έτσι θα πρέπει να αποφασισθεί σε ποιο σημείο θα σταματήσει η ανάπτυξη του δέντρου χωρίς να χάνεται σημαντική πληροφορία για τα δεδομένα. Ο στόχος είναι να αυξάνεται η ακρίβεια πρόβλεψης μειώνοντας το κόστος των λανθασμένων ταξινομήσεων. Το κλάδεμα στηρίζεται σε δυο κριτήρια, που ελέγχουν το πόσο αξιόπιστο είναι ένα δέντρο μικρότερου μεγέθους, τα οποία είναι η διασταυρωτική επικύρωση (CV- Cross Validation) με δείγμα ελέγχου και διασταυρωτική επικύρωση με δείγμα V υποδειγμάτων (V – Fold Cross Validation).

3.5.2 Οπτικοποίηση

Ένα από τα μεγαλύτερα δυνατά σημεία των Δένδρων Απόφασης είναι η ερμηνεία τους. Η οπτικοποίηση των Δένδρων Απόφασης είναι ένας ισχυρός τρόπος για να κατανοήσουμε πώς λειτουργεί το μοντέλο. Για αυτή την διαδικασία χρησιμοποιούμε τα πακέτα matplotlib και graphviz.

3.5.3 Αξιολόγηση Αλγορίθμου

Η αξιολόγηση μοντέλου είναι ένα από τα πιο σημαντικά μέρη της μηχανικής μάθησης. Μέσω της αξιολόγησης πρέπει να παρθεί η απόφαση εάν το μοντέλο που κατασκευάστηκε, θα εφαρμοστεί ή όχι. Η αξιολόγηση του μοντέλου του δένδρου παλινδρόμησης γίνεται με τους εξής δείκτες:

- **Mean Absolute Error (MAE)**: Το μέσο απόλυτο σφάλμα είναι μια μέτρηση αξιολόγησης μοντέλου που χρησιμοποιείται με μοντέλα παλινδρόμησης. Το μέσο απόλυτο σφάλμα ενός μοντέλου σε σχέση με ένα σύνολο δοκιμής είναι ο μέσος όρος των απόλυτων τιμών των μεμονωμένων σφαλμάτων πρόβλεψης σε όλες τις περιπτώσεις στο σύνολο δοκιμής. Κάθε σφάλμα πρόβλεψης είναι η διαφορά μεταξύ της πραγματικής τιμής και της προβλεπόμενης τιμής για το παράδειγμα. Προτιμάται να έχει τιμές κοντά στο 0.

$$mae = \frac{\sum_{i=1}^{n} abs\left(y_i - \lambda(x_i)\right)}{n}$$

Σχέση 3. Τύπος υπολογισμού MAE

- **Mean Square Error (MSE)**: Ο μέσος όρος της απόστασης από κάθε σημείο έως το μοντέλο προβλεπόμενης παλινδρόμησης μπορεί να υπολογιστεί και να εμφανιστεί ως το μέσο τετραγωνικό σφάλμα. Ο τετραγωνισμός είναι κρίσιμος για τη μείωση της πολυπλοκότητας με αρνητικά πρόσημα. Όταν ελαχιστοποιείται το MSE, το μοντέλο θα μπορεί να είναι πιο ακριβές, πράγμα που θα σημαίνει ότι το μοντέλο είναι πιο κοντά στα πραγματικά δεδομένα.

$$\mathrm{MSE}\left(\overline{X}\right) = \mathrm{E}\left[\left(\overline{X} - \mu\right)^2\right] = \left(\frac{\sigma}{\sqrt{n}}\right)^2 = \frac{\sigma^2}{n}$$

Σχέση 4. Τύπος υπολογισμού MSE

- **Root Mean Square Error (RMSE)**: Το Root Mean Square Error (RMSE) είναι η τυπική απόκλιση των υπολειμμάτων (σφάλματα πρόβλεψης). Τα υπολείμματα είναι ένα μέτρο του πόσο μακριά βρίσκονται τα σημεία δεδομένων της γραμμής παλινδρόμησης. Το RMSE είναι ένα μέτρο της κατανομής αυτών των υπολειμμάτων. Με άλλα λόγια, μας λέει πόσο συγκεντρωμένα είναι τα δεδομένα γύρω από τη γραμμή της καλύτερης προσαρμογής.

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

Σχέση 5. Τύπος υπολογισμού RMSE.

- **R^2:** είναι ένα στατιστικό μέτρο που αντιπροσωπεύει το ποσοστό της διακύμανσης για μια εξαρτημένη μεταβλητή που εξηγείται από μια ανεξάρτητη μεταβλητή ή μεταβλητές σε ένα μοντέλο παλινδρόμησης.

$$R^2 = \frac{\text{Variance explained by the model}}{\text{Total variance}}$$

Σχέση 6. Τύπος υπολογισμού R^2

4. ΣΥΛΛΟΓΗ ΚΑΙ ΕΠΕΞΕΡΓΑΣΙΑ ΣΤΟΙΧΕΙΩΝ

4.1 Εισαγωγή

Στόχο της παρούσας διπλωματικής εργασίας, όπως αναφέρθηκε και παραπάνω, αποτελεί η διερεύνηση της επιρροής της πανδημίας του COVID-19 στην εξέλιξη της επιβατικής ζήτησης στο δίκτυο Μέσων Μαζικής Μεταφοράς του ΟΑΣΑ.

Σε αυτό το κεφάλαιο θα αναλυθεί η **συλλογή και η επεξεργασία των στοιχείων** που χρησιμοποιήθηκαν για την εκπόνηση του στόχου της διπλωματικής εργασίας.

4.2 Συλλογή Στοιχείων

Στην παρούσα διπλωματική εργασία ήταν απαραίτητη η δημιουργία μιας **βάσης δεδομένων** η οποία αποτελείται από παραμέτρους που περιγράφουν τις λεωφορειακές γραμμές του ΟΑΣΑ καθώς και τις επικυρώσεις των εισιτηρίων που καταγράφηκαν για διάρκεια μεγαλύτερη του ενός έτους, σε καθημερινή βάση.

Το χρονικό διάστημα που μελετήθηκε είναι από τον Σεπτέμβριο του 2020 έως τον Δεκέμβριο του 2021, κατά την διάρκεια του δεύτερου «κύματος» της πανδημίας.

4.2.1 Επιβατικό κοινό ΟΑΣΑ

Για το πρώτο μέρος της βάσης δεδομένων, δηλαδή τις επικυρώσεις των εισιτηρίων, ανακτήθηκαν από την ιστοσελίδα της Κυβέρνησης https://www.data.gov.gr/ περίπου 123.000 δεδομένα σχετικά με το επιβατικό κοινό του ΟΑΣΑ κατά την παραπάνω περίοδο (Σεπτέμβρης 2020 έως Δεκέμβρης 2021).

Τα δεδομένα αυτά περιέχουν την ονομασία της κάθε γραμμής (στήλη Bus), τις ημερομηνίες (στήλη Date), την κατεύθυνση της γραμμής (στήλη Direction) δηλαδή αν πραγματοποιεί δρομολόγιο αφετηρίας-τέρματος ή το αντίθετο, τις επικυρώσεις των εισιτηρίων (dv_validations) καθώς και την συχνότητα των δρομολογίων ανά ώρα (Εικόνα 7).

Bus	Date	Direction	dv_validations	routes_per_hour
021 - ΠΛΑΤΕΙΑ ΚΑΝΙΓΓΟΣ - ΓΚΥΖΗ (ΚΥΚΛΙΚΗ)	2021-01-02	ΚΑΝΙΓΓΟΣ-ΓΚΥΖΗ	92	32
021 - ΠΛΑΤΕΙΑ ΚΑΝΙΓΓΟΣ - ΓΚΥΖΗ (ΚΥΚΛΙΚΗ)	2021-01-03	ΚΑΝΙΓΓΟΣ-ΓΚΥΖΗ	78	32
021 - ΠΛΑΤΕΙΑ ΚΑΝΙΓΓΟΣ - ΓΚΥΖΗ (ΚΥΚΛΙΚΗ)	2021-01-04	ΚΑΝΙΓΓΟΣ-ΓΚΥΖΗ	275	54
021 - ΠΛΑΤΕΙΑ ΚΑΝΙΓΓΟΣ - ΓΚΥΖΗ (ΚΥΚΛΙΚΗ)	2021-01-05	ΚΑΝΙΓΓΟΣ-ΓΚΥΖΗ	234	56
021 - ΠΛΑΤΕΙΑ ΚΑΝΙΓΓΟΣ - ΓΚΥΖΗ (ΚΥΚΛΙΚΗ)	2021-01-06	ΚΑΝΙΓΓΟΣ-ΓΚΥΖΗ	98	34
021 - ΠΛΑΤΕΙΑ ΚΑΝΙΓΓΟΣ - ΓΚΥΖΗ (ΚΥΚΛΙΚΗ)	2021-01-07	ΚΑΝΙΓΓΟΣ-ΓΚΥΖΗ	253	56
021 - ΠΛΑΤΕΙΑ ΚΑΝΙΓΓΟΣ - ΓΚΥΖΗ (ΚΥΚΛΙΚΗ)	2021-01-08	ΚΑΝΙΓΓΟΣ-ΓΚΥΖΗ	321	70
021 - ΠΛΑΤΕΙΑ ΚΑΝΙΓΓΟΣ - ΓΚΥΖΗ (ΚΥΚΛΙΚΗ)	2021-01-09	ΚΑΝΙΓΓΟΣ-ΓΚΥΖΗ	202	48
021 - ΠΛΑΤΕΙΑ ΚΑΝΙΓΓΟΣ - ΓΚΥΖΗ (ΚΥΚΛΙΚΗ)	2021-01-10	ΚΑΝΙΓΓΟΣ-ΓΚΥΖΗ	79	32
021 - ΠΛΑΤΕΙΑ ΚΑΝΙΓΓΟΣ - ΓΚΥΖΗ (ΚΥΚΛΙΚΗ)	2021-01-11	ΚΑΝΙΓΓΟΣ-ΓΚΥΖΗ	300	63
021 - ΠΛΑΤΕΙΑ ΚΑΝΙΓΓΟΣ - ΓΚΥΖΗ (ΚΥΚΛΙΚΗ)	2021-01-12	ΚΑΝΙΓΓΟΣ-ΓΚΥΖΗ	295	61
021 - ΠΛΑΤΕΙΑ ΚΑΝΙΓΓΟΣ - ΓΚΥΖΗ (ΚΥΚΛΙΚΗ)	2021-01-13	ΚΑΝΙΓΓΟΣ-ΓΚΥΖΗ	308	68
021 - ΠΛΑΤΕΙΑ ΚΑΝΙΓΓΟΣ - ΓΚΥΖΗ (ΚΥΚΛΙΚΗ)	2021-01-14	ΚΑΝΙΓΓΟΣ-ΓΚΥΖΗ	313	67
021 - ΠΛΑΤΕΙΑ ΚΑΝΙΓΓΟΣ - ΓΚΥΖΗ (ΚΥΚΛΙΚΗ)	2021-01-15	ΚΑΝΙΓΓΟΣ-ΓΚΥΖΗ	321	61
021 - ΠΛΑΤΕΙΑ ΚΑΝΙΓΓΟΣ - ΓΚΥΖΗ (ΚΥΚΛΙΚΗ)	2021-01-16	ΚΑΝΙΓΓΟΣ-ΓΚΥΖΗ	151	48
021 - ΠΛΑΤΕΙΑ ΚΑΝΙΓΓΟΣ - ΓΚΥΖΗ (ΚΥΚΛΙΚΗ)	2021-01-17	ΚΑΝΙΓΓΟΣ-ΓΚΥΖΗ	81	33
021 - ΠΛΑΤΕΙΑ ΚΑΝΙΓΓΟΣ - ΓΚΥΖΗ (ΚΥΚΛΙΚΗ)	2021-01-18	ΚΑΝΙΓΓΟΣ-ΓΚΥΖΗ	339	46
021 - ΠΛΑΤΕΙΑ ΚΑΝΙΓΓΟΣ - ΓΚΥΖΗ (ΚΥΚΛΙΚΗ)	2021-01-19	ΚΑΝΙΓΓΟΣ-ΓΚΥΖΗ	329	66
021 - ΠΛΑΤΕΙΑ ΚΑΝΙΓΓΟΣ - ΓΚΥΖΗ (ΚΥΚΛΙΚΗ)	2021-01-20	ΚΑΝΙΓΓΟΣ-ΓΚΥΖΗ	335	68
021 - ΠΛΑΤΕΙΑ ΚΑΝΙΓΓΟΣ - ΓΚΥΖΗ (ΚΥΚΛΙΚΗ)	2021-01-21	ΚΑΝΙΓΓΟΣ-ΓΚΥΖΗ	361	61

Εικόνα 7. Απόσπασμα δεδομένων επιβατικού κοινού ΟΑΣΑ

Από τα παραπάνω δεδομένα αξιοποιήθηκαν οι στήλες Bus, Date, dv_validations. Οι γραμμές που διαθέτουν δυο κατευθύνσεις (αφετηρία-τέρμα και τέρμα-αφετηρία) αντιμετωπίστηκαν ως μία συνολική γραμμή καθώς και τα δεδομένα από τις επικυρώσεις εισιτηρίων μετατράπηκαν σε εβδομαδιαίες μετρήσεις.

4.2.2 Στοιχεία λεωφορειακών γραμμών ΟΑΣΑ

Για το δεύτερο μέρος της βάσης δεδομένων, χρησιμοποιήθηκε το πρόγραμμα QGIS στο οποίο εισήχθησαν δύο αρχεία πληροφοριών σε μορφή shapefile (.shp) και πραγματοποιήθηκε μια προ-επεξεργασία των δεδομένων.

Το QGIS λειτουργεί ως λογισμικό γεωγραφικού συστήματος πληροφοριών (GIS), επιτρέποντας στους χρήστες να αναλύουν και να επεξεργάζονται χωρικές πληροφορίες καθώς και να συνθέτουν και να εξάγουν γραφικούς χάρτες. Το QGIS υποστηρίζει τόσο επίπεδα ράστερ όσο και διανυσματικά επίπεδα- τα διανυσματικά δεδομένα αποθηκεύονται είτε ως σημεία, είτε ως γραμμές, είτε ως πολυγωνικά χαρακτηριστικά. Υποστηρίζονται πολλαπλές μορφές εικόνων ράστερ και το λογισμικό μπορεί να κάνει γεωαναφορά εικόνων.

Στόχος σε αυτό το στάδιο της Διπλωματικής εργασίας είναι να αντληθούν πληροφορίες για τα γεωμετρικά χαρακτηριστικά των λεωφορειακών γραμμών. Τα χαρακτηριστικά αυτά είναι το **μήκος** των γραμμών, η **επικάλυψη** των γραμμών καθώς και η **τυπολογία** των οδών που εξυπηρετούν κάθε γραμμή.

- Το πρώτο αρχείο (active_lines), το οποίο ζητήθηκε από τον ΟΑΣΑ, περιείχε πληροφορίες για τις λεωφορειακές γραμμές (Εικόνα 8).

Εικόνα 8. Λεωφορειακές γραμμές ΟΑΣΑ

Οι πληροφορίες του παραπάνω αρχείου αποτελούνταν από στοιχεία όπως η περιγραφή της γραμμής, ο κωδικός, εάν η γραμμή είναι κυκλική και εάν η γραμμή είναι νυχτερινή (Εικόνα 9). Ο σκοπός της επεξεργασίας του αρχείου ήταν να αντλήσουμε πληροφορίες για το μήκος των γραμμών και την επικάλυψή τους.

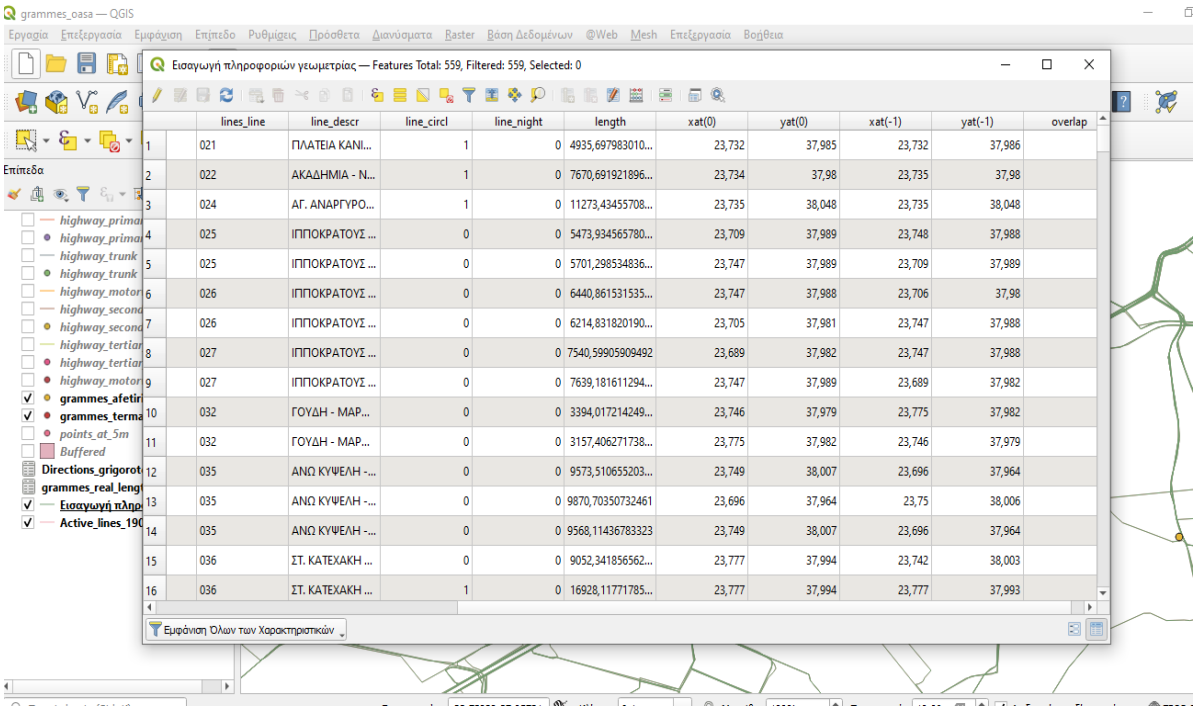
Active_lines_190321_s — Features Total: 559, Filtered: 559, Selected: 0

	route_code	route_id	route_desc	route_type	line_id	lines_line	line_descr	line_circl	line_night
1	2484	01	ΚΑΝΙΓΓΟΣ-ΓΚΥ...	1	021	021	ΠΛΑΤΕΙΑ ΚΑΝΙ...	1	0
2	3494	02	ΑΚΑΔΗΜΙΑ - Ν...	1	022	022	ΑΚΑΔΗΜΙΑ - Ν...	1	0
3	2640	01	ΑΓ.ΑΝΑΡΓΥΡΟΙ...	1	024	024	ΑΓ. ΑΝΑΡΓΥΡΟ...	1	0
4	1798	02	ΠΡ. ΔΑΝΙΗΛ - Ι...	2	025	025	ΙΠΠΟΚΡΑΤΟΥΣ ...	0	0
5	3659	03Π	ΙΠΠΟΚΡΑΤΟΥΣ ...	1	025	025	ΙΠΠΟΚΡΑΤΟΥΣ ...	0	0
6	3660	03Π	ΙΠΠΟΚΡΑΤΟΥΣ ...	1	026	026	ΙΠΠΟΚΡΑΤΟΥΣ ...	0	0
7	1800	02	ΒΟΤΑΝΙΚΟΣ - Ι...	2	026	026	ΙΠΠΟΚΡΑΤΟΥΣ ...	0	0
8	1975	02	ΟΡΦΕΩΣ - ΙΠΠ...	2	027	027	ΙΠΠΟΚΡΑΤΟΥΣ ...	0	0
9	3661	03Π	ΙΠΠΟΚΡΑΤΟΥΣ ...	1	027	027	ΙΠΠΟΚΡΑΤΟΥΣ ...	0	0
10	2781	02	ΜΑΡΑΣΛΕΙΟΣ-Γ...	2	032	032	ΓΟΥΔΗ - ΜΑΡ...	0	0
11	2780	01	ΓΟΥΔΗ-ΜΑΡΑ...	1	032	032	ΓΟΥΔΗ - ΜΑΡ...	0	0
12	2953	03	ΑΝΩ ΚΥΨΕΛΗ -...	1	035	035	ΑΝΩ ΚΥΨΕΛΗ -...	0	0
13	2949	04	ΤΑΥΡΟΣ - ΠΕΤΡ...	2	035	035	ΑΝΩ ΚΥΨΕΛΗ -...	0	0
14	2948	03Ν	ΑΝΩ ΚΥΨΕΛΗ -...	1	035	035	ΑΝΩ ΚΥΨΕΛΗ -...	0	0
15	3093	03	ΣΤ. ΚΑΤΕΧΑΚΗ ...	1	036ΕΤ	036	ΣΤ. ΚΑΤΕΧΑΚΗ ...	0	0
16	3092	03Λ	ΣΤ.ΚΑΤΕΧΑΚΗ-...	1	036	036	ΣΤ. ΚΑΤΕΧΑΚΗ ...	1	0

Εμφάνιση Όλων των Χαρακτηριστικών

Εικόνα 9. Πληροφορίες αρχείου active lines.

Το QGIS διαθέτει ενσωματωμένες συναρτήσεις και αλγόριθμους για τον υπολογισμό διαφόρων ιδιοτήτων με βάση τη γεωμετρία του χαρακτηριστικού όπως μήκος, εμβαδόν, περίμετρος κ.λπ. Επομένως για να υπολογιστεί το μήκος κάθε λεωφορειακής γραμμής χρησιμοποιήθηκε η εντολή «Προσθήκη γεωμετρικών πληροφοριών» από τη οποία προστέθηκε στο υπάρχον αρχείο μια στήλη με το **μήκος** κάθε λεωφορειακής γραμμής (Εικόνα 10).



	lines_line	line_descr	line_circl	line_night	length	xat(0)	yat(0)	xat(-1)	yat(-1)	overlap
1	021	ΠΛΑΤΕΙΑ ΚΑΝΙ...	1	0	4935,697983010...	23,732	37,985	23,732	37,986	
2	022	ΑΚΑΔΗΜΙΑ - Ν...	1	0	7670,691921896...	23,734	37,98	23,735	37,98	
3	024	ΑΓ. ΑΝΑΡΓΥΡΟ...	1	0	11273,43455708...	23,735	38,048	23,735	38,048	
4	025	ΙΠΠΟΚΡΑΤΟΥΣ ...	0	0	5473,934565780...	23,709	37,989	23,748	37,988	
5	025	ΙΠΠΟΚΡΑΤΟΥΣ ...	0	0	5701,298534836...	23,747	37,989	23,709	37,989	
6	026	ΙΠΠΟΚΡΑΤΟΥΣ ...	0	0	6440,861531535...	23,747	37,988	23,706	37,98	
7	026	ΙΠΠΟΚΡΑΤΟΥΣ ...	0	0	6214,831820190...	23,705	37,981	23,747	37,988	
8	027	ΙΠΠΟΚΡΑΤΟΥΣ ...	0	0	7540,59905909492	23,689	37,982	23,747	37,988	
9	027	ΙΠΠΟΚΡΑΤΟΥΣ ...	0	0	7639,181611294...	23,747	37,989	23,689	37,982	
10	032	ΓΟΥΔΗ - ΜΑΡ...	0	0	3394,017214249...	23,746	37,979	23,775	37,982	
11	032	ΓΟΥΔΗ - ΜΑΡ...	0	0	3157,406271738...	23,775	37,982	23,746	37,979	
12	035	ΑΝΩ ΚΥΨΕΛΗ -...	0	0	9573,510655203...	23,749	38,007	23,696	37,964	
13	035	ΑΝΩ ΚΥΨΕΛΗ -...	0	0	9870,70350732461	23,696	37,964	23,75	38,006	
14	035	ΑΝΩ ΚΥΨΕΛΗ -...	0	0	9568,11436783323	23,749	38,007	23,696	37,964	
15	036	ΣΤ. ΚΑΤΕΧΑΚΗ ...	0	0	9052,341856562...	23,777	37,994	23,742	38,003	
16	036	ΣΤ. ΚΑΤΕΧΑΚΗ ...	1	0	16928,11771785...	23,777	37,994	23,777	37,993	

Εικόνα 10. Εισαγωγή γεωμετρικών πληροφοριών στο αρχείο active lines

Στη συνέχεια για να υπολογιστεί η **επικάλυψη** των γραμμών (overlap) χρησιμοποιήθηκε το παραπάνω επίπεδο και έγινε ο χωρισμός της κάθε λεωφορειακής γραμμής σε τμήματα των 50 μέτρων (segments) με χρήση του εργαλείου διανυσματικής γεωμετρίας που ονομάζεται “Split Lines by Maximum Length” και η δημιουργία μιας ζώνης buffer 5 μέτρων γύρω από κάθε τμήμα.

Μέσω αυτής της διαδικασίας έγινε η εξαγωγή ενός αρχείου που περιέχει όλα τα τμήματα 50 μέτρων για κάθε λεωφορειακή γραμμή καθώς και έναν μοναδικό κωδικό id για το κάθε τμήμα.

Τα δεδομένα αυτά εισήχθησαν στο προγραμματιστικό περιβάλλον RStudio με σκοπό να βρεθούν ποια μοναδικά τμήματα (segments) της κάθε γραμμής εφάπτονται με τμήματα άλλων λεωφορειακών γραμμών.

Αρχικά αφαιρέθηκαν από αυτό το αρχείο τα σημεία όπου η κάθε γραμμή επικαλύπτεται με τον εαυτό της ή επικαλύπτονται οι δύο κατευθύνσεις μιας γραμμής.

Επομένως, με την βοήθεια των εντολών group by και summarize ,υπολογίστηκε για κάθε γραμμή το σύνολο των τμημάτων της το οποίο επικαλύπτεται με τμήματα άλλων γραμμών.

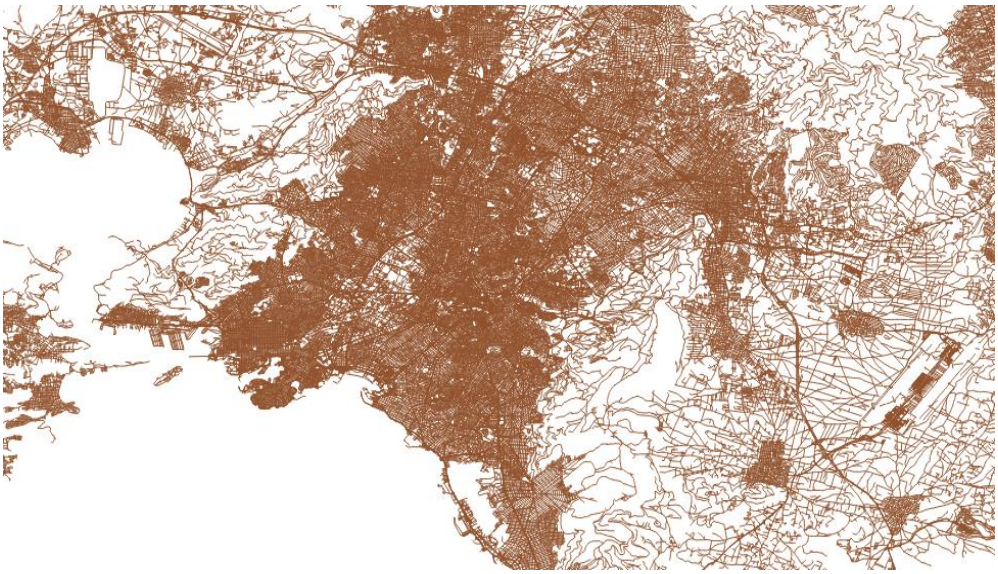
Για παράδειγμα, εάν σε μια γραμμή εντοπίστηκαν 15 μοναδικά segments των 50 μέτρων, τα οποία εφάπτονται με άλλες γραμμές, μπορεί να υπολογιστεί 15*50m= 750 μέτρα κοινής διαδρομής με άλλες γραμμές.

Το σύνολο αυτό για κάθε γραμμή διαιρεμένο με το πραγματικό μήκος της κάθε γραμμής μας δίνει το **ποσοστό επικάλυψης (overlap)**, δηλαδή το ποσοστό κοινής διαδρομής με άλλες λεωφορειακές γραμμές (Εικόνα 11).

line_id	line_descr	length	overlap
021	ΠΛΑΤΕΙΑ ΚΑΝΙΓΓΟΣ - ΓΚΥΖΗ (ΚΥΚΛΙΚΗ)	4935.70	7.215
022	ΑΚΑΔΗΜΙΑ - Ν. ΚΥΨΕΛΗ (ΚΥΚΛΙΚΗ)	7670.69	22.742
024	ΑΓ. ΑΝΑΡΓΥΡΟΙ - ΣΤ. ΚΑΤΩ ΠΑΤΗΣΙΑ (ΚΥΚΛΙΚΗ)	11273.43	3.631
025	ΙΠΠΟΚΡΑΤΟΥΣ - ΠΡΟΦΗΤΗ ΔΑΝΙΗΛ	5473.93	14.798
026	ΙΠΠΟΚΡΑΤΟΥΣ - ΒΟΤΑΝΙΚΟΣ	6214.83	15.050
027	ΙΠΠΟΚΡΑΤΟΥΣ - ΟΡΦΕΩΣ	7639.18	12.119
032	ΓΟΥΔΗ - ΜΑΡΑΣΛΕΙΟΣ (ΣΧΟΛΙΚΗ)	3394.02	13.009
035	ΑΝΩ ΚΥΨΕΛΗ - ΠΕΤΡΑΛΩΝΑ - ΤΑΥΡΟΣ	9573.51	30.562
036	ΣΤ. ΚΑΤΕΧΑΚΗ -ΣΤ. ΠΑΝΟΡΜΟΥ-ΓΑΛΑΤΣΙ-ΚΥΨΕΛΗ (ΚΥΚΛΙΚΗ)	16928.12	29.902
036ΕΑ	ΣΤ. ΚΑΤΕΧΑΚΗ -ΣΤ. ΠΑΝΟΡΜΟΥ-ΓΑΛΑΤΣΙ-ΚΥΨΕΛΗ	7903.14	34.902
036ΕΤ	ΣΤ. ΚΑΤΕΧΑΚΗ -ΣΤ. ΠΑΝΟΡΜΟΥ-ΓΑΛΑΤΣΙ-ΚΥΨΕΛΗ	9052.34	19.212
040	ΠΕΙΡΑΙΑΣ - ΣΥΝΤΑΓΜΑ	12959.65	53.076
046	ΜΟΥΣΕΙΟ - ΕΛΛΗΝΟΡΩΣΩΝ	6687.80	23.050
049	ΠΕΙΡΑΙΑΣ - ΟΜΟΝΟΙΑ	10864.33	20.783
051	ΣΤ. ΥΠΕΡ. ΛΕΩΦ. ΚΗΦΙΣΟΥ - ΟΜΟΝΟΙΑ (ΜΕΣΩ ΑΚΑΔ. ΠΛΑΤ.)(Κ	7528.70	6.100
052	ΣΤ. ΕΛΑΙΩΝΑ - ΣΤ. ΥΠΕΡ. ΛΕΩΦ. ΚΗΦΙΣΟΥ (ΚΥΚΛΙΚΗ)	3569.36	5.776
054	ΠΕΡΙΣΣΟΣ - ΛΑΜΠΡΙΝΗ - ΠΟΛΥΤΕΧΝΕΙΟ (ΚΥΚΛΙΚΗ)	15912.24	9.043
057	ΛΟΦΟΣ ΣΚΟΥΖΕ - ΟΜΟΝΟΙΑ (ΚΥΚΛΙΚΗ)	7759.95	8.546
060	ΑΚΑΔΗΜΙΑ - ΛΥΚΑΒΗΤΤΟΣ (ΚΥΚΛΙΚΗ)	8532.78	11.723
1	ΠΛ. ΑΤΤΙΚΗΣ - ΚΑΛΛΙΘΕΑ - ΜΟΣΧΑΤΟ	11614.05	22.622
10	ΤΖΙΤΖΙΦΙΕΣ - ΧΑΛΑΝΔΡΙ	14270.75	23.895

Εικόνα 11. Απόσπασμα αρχείου επικάλυψης γραμμών

- Το δεύτερο αρχείο (roads_Greece) , που ανακτήθηκε από το OpenStreetMap, περιείχε πληροφορίες για τον τύπο όλων των οδών της Ελλάδας, όμως η ανάλυση επικεντρώθηκε στις οδούς της Αττικής (Εικόνες 12 και 13).
  Οι πληροφορίες που ήταν επιθυμητό να εξάγουμε από αυτό το αρχείο ήταν η **τυπολογία** των οδών για την διαδρομή κάθε λεωφορειακής γραμμής.



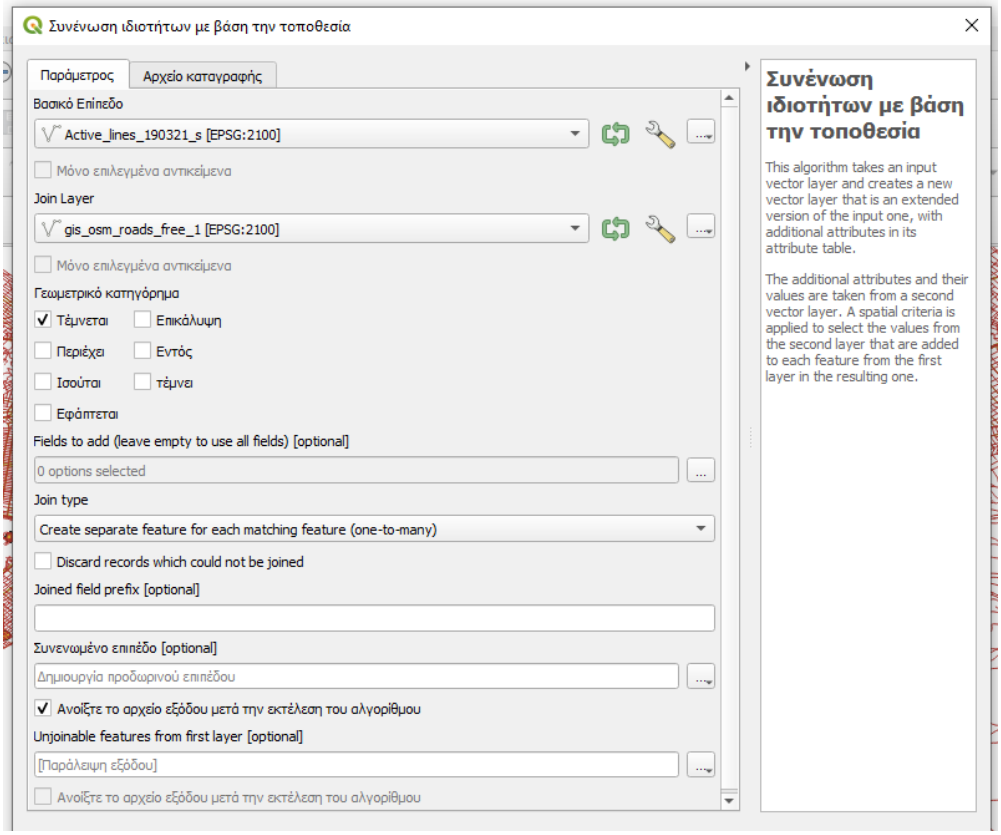
Εικόνα 12. Αρχείο shapefile με την τυπολογία οδών Αττικής

gis_osm_roads_free_1 — Features Total: 1234804, Filtered: 1234804, Selected: 0

	osm_id	code	fclass	name	ref
34	4380992	5122	residential	Νίκης	NULL
35	4380993	5122	residential	Μαλτέζου	NULL
36	4381108	5114	secondary	Αχιλλέως	NULL
37	4381112	5122	residential	Δήμητρος	NULL
38	4381113	5122	residential	Καραολή & Δη...	NULL
39	4381500	5122	residential	NULL	NULL
40	4381501	5122	residential	Γ. Μαρμαρά	NULL
41	4381503	5122	residential	Στράβωνος	NULL
42	4381504	5122	residential	Φλέμινγκ	NULL
43	4381505	5122	residential	Μ. Μητρόπουλ...	NULL
44	4381664	5141	service	NULL	NULL
45	4381672	5115	tertiary	Κ. Πλειώνη	NULL
46	4381673	5122	residential	NULL	NULL
47	4381676	5122	residential	Πλατεία Αγωνι...	NULL
48	4381678	5114	secondary	NULL	NULL
49	4381683	5115	tertiary	NULL	NULL
50	4381685	5122	residential	Αγίας Παρασκε...	NULL

Εικόνα 13. Απόσπασμα πίνακα με την τυπολογία οδών της Ελλάδας.

Επομένως για να βρεθεί η τυπολογία των οδών που χρησιμοποιεί κάθε γραμμή ήταν απαραίτητο να ενωθούν τα δύο αυτά επίπεδα, δηλαδή το επίπεδο με τις λεωφορειακές γραμμές και το επίπεδο με την τυπολογία των οδών της Ελλάδας. Για να επιτευχθεί αυτό, χρησιμοποιήθηκε η εντολή "Join Attributes by Location" (Συνένωση ιδιοτήτων με βάση την τοποθεσία) (Εικόνα 14).



Εικόνα 14. Qgis: Συνένωση ιδιοτήτων με βάση την τοποθεσία.

Αυτή η διαδικασία είχε ως αποτέλεσμα την δημιουργία ενός επιπέδου που περιέχει την **τυπολογία των οδών** που διασχίζει κάθε λεωφορειακή γραμμή. Όμως κάθε λεωφορειακή γραμμή διασχίζει, στην διαδρομή της, οδούς διαφορετικής τυπολογίας. Επομένως υπολογίστηκε για την διαδρομή της κάθε γραμμής το ποσοστό της τυπολογίας των δρόμων που διασχίζει. Για παράδειγμα, στην παρακάτω εικόνα (Εικόνα 15) φαίνεται πως η διαδρομή της λεωφορειακής γραμμής 021, διέρχεται κατά 45% από δρόμο κατοικημένης περιοχής, κατά 4,3% από πρωταρχικό δρόμο κλπ.

line_id	footway	motorway	pedestrian	primary	residential	secondary	steps	tertiary
021	7.609		9.783	4.348	45.652	4.348	2.174	22.826
022	25.301		4.819	12.651	29.518	8.434		15.663
024	10.884	1.361	2.721	6.803	46.259	2.721	1.361	15.646
025	15.833		9.167	6.667	35.833	5.000		15.000
026	15.254		7.627	4.237	36.441	5.085		17.797
027	11.940		11.194	3.731	37.313	6.716		14.925
032	16.667		5.556	5.556	45.833	6.944	1.389	16.667
035	12.598		10.236	11.023	35.959	3.149		17.061
036	8.841		2.554	9.234	37.917	11.395	0.982	22.790
036EA	9.125		2.556	8.029	37.224	12.408	2.187	24.092
036ET	10.791		2.158	9.353	33.813	10.791	0.719	25.899
040	17.051	0.922	5.991	9.217	41.935	6.912	0.922	7.834
046	8.083		2.393	9.646	49.646	4.420		21.026
049	16.049	1.235	3.704	20.370	28.395	12.346		10.494
051	11.111		1.852	10.185	36.111	14.815		16.667
051EA	11.321		1.887	5.660	49.057	9.434		16.981
052	2		2		16	16		4
054	12.121		4.762	6.061	52.814	6.494		16.883

Εικόνα 15. Απόσπασμα πίνακα που δείχνει από ποιους τύπους οδών αποτελείται η διαδρομή της κάθε λεωφορειακής γραμμής.

Παρατηρούμε ότι υπάρχουν λεωφορειακές γραμμές που παρουσιάζουν μεγάλο ποσοστό residential όπως για παράδειγμα οι γραμμές:

- 046 ΜΟΥΣΕΙΟ – ΕΛΛΗΝΟΡΩΣΩΝ με ποσοστό 49%
- 141 ΣΤ. ΔΑΦΝΗ - ΚΑΛΑΜΑΚΙ (ΚΥΚΛΙΚΗ) με ποσοστό 64%
- 752 ΑΧΑΡΝΑΙ - ΑΓ. ΙΩΑΝΝΗΣ - ΑΓ. ΠΕΤΡΟΣ (ΚΥΚΛΙΚΗ) με ποσοστό 76%

Πρόκειται για γραμμές που βρίσκονται στο κέντρο και σε πυκνά κατοικημένες περιοχές.

Επίσης παρατηρούμε γραμμές που εμφανίζουν μεγάλο ποσοστό primary (πρωταρχικός δρόμος) όπως οι γραμμές:

- 122 ΣΤ. ΑΡΓΥΡΟΥΠΟΛΗ – ΣΑΡΩΝΙΔΑ με ποσοστό 22% η οποία διασχίζει την λεωφόρο Βουλιαγμένης και την λεωφόρο Κωνσταντίνου Καραμανλή.
- 314 ΣΤ.ΔΟΥΚ.ΠΛΑΚΕΝΤΙΑΣ - ΠΑΛΛΗΝΗ – ΡΑΦΗΝΑ με ποσοστό 16.7% η οποία διασχίζει για το μεγαλύτερο τμήμα της την λεωφόρο Μαραθώνος.
- 3 Ν. ΦΙΛΑΔΕΛΦΕΙΑ - ΑΝΩ ΠΑΤΗΣΙΑ - ΝΕΟ ΨΥΧΙΚΟ με ποσοστό 15.7% η οποία διασχίζει την Πατησίων και την λεωφόρο Βασιλίσσης Σοφίας.

4.3 Επεξεργασία , Καθαρισμός και Οργάνωση Δεδομένων

Η μορφή των παραπάνω δεδομένων που συλλέχθηκαν είναι ακατέργαστη επομένως για να μπορούν να αξιοποιηθούν είναι αναγκαίο να επεξεργαστούν, να γίνει καθαρισμός και ένωση των στοιχείων ώστε να προκύψει η τελική βάση δεδομένων.

Κάποια από τα σφάλματα-προβλήματα που προέκυψαν ήταν οι διπλοεγγραφές των παρατηρήσεων και η αναντιστοιχία στα ονόματα των λεωφορειακών γραμμών μεταξύ των διαφορετικών πηγών δεδομένων.

Πλην όμως των σφαλμάτων όπως παραπάνω, έπρεπε να γίνει και καθαρισμός δεδομένων, είτε λόγω ημιτελών δρομολογίων είτε εξωγενών παραγόντων (π.χ. πορείες, κυκλοφοριακές ρυθμίσεις, τεχνικά προβλήματα). Για τον λόγο αυτό αφαιρέθηκαν οι λεωφορειακές γραμμές για τις οποίες υπήρχαν δεδομένα επικυρώσεων για λιγότερες από 300 ημερομηνίες.

Με βάση τα παραπάνω, εισήχθησαν τα δεδομένα στο προγραμματιστικό περιβάλλον της R με σκοπό να διορθωθούν τα σφάλματα που αναφέρθηκαν και να γίνει η ένωση των στοιχείων. Στο τέλος απομονώθηκαν οι στήλες που χρησιμοποιήθηκαν για την ανάπτυξη των μοντέλων και το αρχείο απέκτησε την εξής μορφή (Εικόνα 16):

line	length	overlap	residential
036EA	7903.14	34.9	37.22
036ET	9052.34	19.21	33.81
1	11718.03	18.33	39.55
10	15068.07	21.06	31.97
101	13985.40	6.88	32.92
106	17771.20	18.24	38.57
109	18696.24	16.34	31.86
11	9986.37	26.09	35.87
112	10358.21	2.68	63.23
12	8718.96	19.42	33.83
124	25566.18	22.36	39.35
126	18756.43	28.17	44.14
128	26032.22	8.85	37.99
130	22213.64	25.47	36.2
131	8471.73	5.55	50.77
136	13227.35	14.17	44.5
137	12790.66	13.24	43.3
14	9841.94	24.6	42.78
140	27534.01	11.67	33.9
15	8514.41	24.57	29.27
16	8562.54	9.84	46.32
164	8166.49	17.5	39.24
17	10473.22	21.79	42.31
18	8430.53	26.52	33.56
19	11248.48	27.6	25.42
2	9001.65	22.51	38.78

Επεξήγηση:

Line: Οι λεωφορειακές γραμμές ΟΑΣΑ.
Length: Μήκος διαδρομής λεωφορειακής γραμμής.
Overlap: Ποσοστό επικάλυψης των διαδρομών των λεωφορειακών γραμμών.
Residential: Ποσοστό κατά το οποίο η διαδρομή των λεωφορειακών γραμμών διέρχεται από περιοχή με κατοικίες.

Εικόνα 16.Γεωμετρικά χαρακτηριστικά γραμμών.

5. ΕΦΑΡΜΟΓΗ ΜΕΘΟΔΟΛΟΓΙΑΣ – ΑΠΟΤΕΛΕΣΜΑΤΑ

5.1 Εισαγωγή

Στο παρόν κεφάλαιο πραγματοποιείται αναλυτική παρουσίαση της μεθοδολογίας που εφαρμόστηκε και των αποτελεσμάτων που προέκυψαν στο πλαίσιο της μελέτης. Όπως αναφέρθηκε στο κεφάλαιο 3, θα αναπτυχθούν **μοντέλα Δένδρων Παλινδρόμησης** μέσω κατάλληλου κώδικα της προγραμματιστικής γλώσσας Python για την εκτίμηση της επίδρασης των χαρακτηριστικών των λεωφορειακών γραμμών στην διακύμανση της ζήτησης κατά την διάρκεια της πανδημίας.

5.2 Εφαρμογή Μεθοδολογίας

Τα ημερήσια δεδομένα με τις επικυρώσεις εισιτηρίων που συλλέχθηκαν, μετατράπηκαν σε εβδομαδιαία με σκοπό να υπολογιστεί ο **συντελεστής διακύμανσης CV**, ο οποίος είναι ο λόγος της τυπικής απόκλισης προς τον μέσο όρο και δείχνει την έκταση της μεταβλητότητας σε σχέση με τον μέσο όρο του πληθυσμού. Όσο υψηλότερος είναι ο δείκτης CV, τόσο μεγαλύτερη είναι η διασπορά.

Coefficient of Variation = (Standard Deviation / Mean) * 100

Σχέση 7. Συντελεστής διακύμανσης CV.

Ο συντελεστής διακύμανσης CV είναι χρήσιμος επειδή η τυπική απόκλιση δεδομένων πρέπει πάντα να γίνεται κατανοητή στο πλαίσιο του μέσου όρου των δεδομένων. Αντίθετα, η πραγματική τιμή του CV είναι ανεξάρτητη από τη μονάδα στην οποία έχει ληφθεί η μέτρηση, επομένως είναι ένας αδιάστατος αριθμός. Για σύγκριση μεταξύ συνόλων δεδομένων με διαφορετικές μονάδες ή πολύ διαφορετικά μέσα, θα πρέπει να χρησιμοποιηθεί ο συντελεστής διακύμανσης αντί της τυπικής απόκλισης.

Για τον παραπάνω λόγο, ο συντελεστής διακύμανσης CV επιλέχθηκε ως η εξαρτημένη μεταβλητή για την ανάλυση των Δένδρων Παλινδρόμησης.

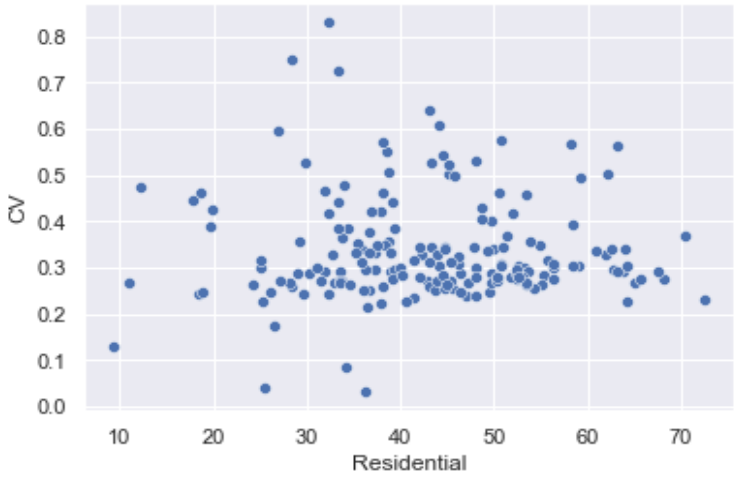
Αρχικά, δοκιμάστηκαν διάφορα μοντέλα γραμμικής παλινδρόμησης από όπου φάνηκε ότι η σχέση των μεταβλητών δεν είναι γραμμική. Τα μοντέλα γραμμικής παλινδρόμησης και λογιστικής παλινδρόμησης αποτυγχάνουν σε καταστάσεις όπου η σχέση μεταξύ των χαρακτηριστικών και του αποτελέσματος είναι μη γραμμική ή όπου τα χαρακτηριστικά αλληλοεπιδρούν μεταξύ τους. Επομένως, επιλέχθηκε και εκτελέστηκε **μη γραμμική παλινδρόμηση με δένδρα αποφάσεων.**

Πρώτο βήμα της ανάλυσης αποτελεί η εισαγωγή του πλαισίου δεδομένων στο προγραμματιστικό περιβάλλον της Python και η περιγραφή και μετονομασία των μεταβλητών καθώς και η εισαγωγή των κατάλληλων πακέτων απαραίτητα για την ανάλυση (Εικόνα 17).

line	length	overlap	residential	cv_week
036EA	7903.14	34.9	37.22	0.295205
1	11718.03	18.33	39.55	0.293418
10	15068.07	21.06	31.97	0.291226
101	13985.4	6.88	32.92	0.266851
106	17771.2	18.24	38.57	0.549818
109	18696.24	16.34	31.86	0.465774
11	9986.365	26.09	35.87	0.33971
112	10358.21	2.68	63.23	0.565545
12	8718.958	19.42	33.83	0.364316
124	25566.18	22.36	39.35	0.38682
126	18756.43	28.17	44.14	0.60885
128	26032.22	8.85	37.99	0.420159
130	22213.64	25.47	36.2	0.295319
136	13227.35	14.17	44.5	0.542054
137	12790.66	13.24	43.3	0.529068
14	9841.944	24.6	42.78	0.273098
140	27534.01	11.67	33.9	0.479079
15	8514.408	24.57	29.27	0.357724
16	8562.536	9.84	46.32	0.287476
164	8166.488	17.5	39.24	0.441328

Εικόνα 17. Απόσπασμα πλαισίου δεδομένων.

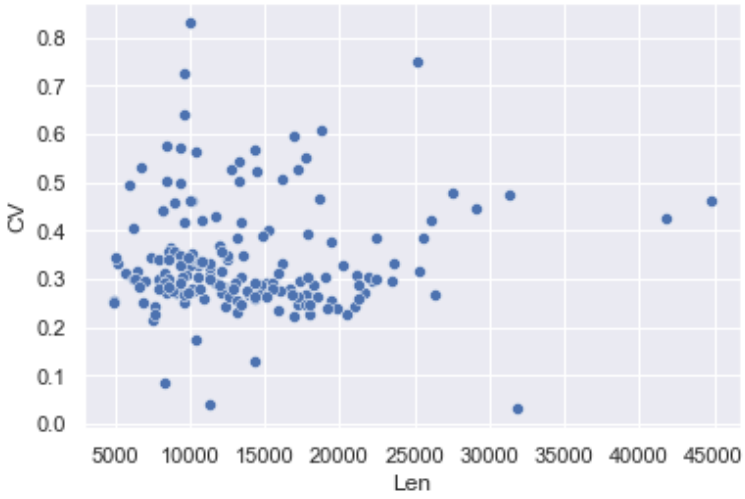
Αρχικά σχεδιάστηκαν διαγράμματα που δείχνουν την σχέση μεταξύ των ανεξάρτητων μεταβλητών και της εξαρτημένης μεταβλητής CV (Εικόνα 18).



Εικόνα 18. Σχέση του συντελεστή Διακύμανσης CV με το ποσοστό κατά το οποίο η διαδρομή των λεωφορειακών γραμμών διέρχεται από περιοχή με κατοικίες.

Στο παραπάνω διάγραμμα της σχέσης CV-Residential, παρατηρείται διασπορά ως προς το ποσοστό των γραμμών που διέρχονται από κατοικημένες περιοχές με το μεγαλύτερο αριθμό των γραμμών να παρουσιάζει την μεταβλητή Residential μεταξύ 30-55%. Παρατηρείται επίσης ότι υπάρχει περιορισμένος αριθμός ακραίων σημείων (outliers). Τα δέντρα απόφασης δεν είναι ευαίσθητα ακραίες τιμές, καθώς οι ακραίες τιμές δεν προκαλούν ποτέ μεγάλη μείωση στο υπολειπόμενο άθροισμα τετραγώνων (RSS), επειδή δεν εμπλέκονται ποτέ στη διαίρεση. Κάποιες από τις γραμμές που εμφανίζουν πολύ μικρό ποσοστό Residential (0-15%) οφείλονται σε διαδρομές που απομακρύνονται από το κέντρο της πόλης όπως για παράδειγμα η λεωφορειακή γραμμή Χ95 η οποία έχει προορισμό τον Αερολιμένα Αθηνών. Αντίθετα γραμμές με πολύ μεγάλο ποσοστό residential (60-70%) αναφέρονται σε γραμμές των οποίων η διαδρομή είναι σε πολύ κεντρικές περιοχές όπως για παράδειγμα οι γραμμές 021(ΠΛΑΤΕΙΑ ΚΑΝΙΓΓΟΣ – ΓΚΥΖΗ), 040 (ΠΕΙΡΑΙΑΣ – ΣΥΝΤΑΓΜΑ), 752 (ΑΧΑΡΝΑΙ - ΑΓ. ΙΩΑΝΝΗΣ - ΑΓ. ΠΕΤΡΟΣ) κλπ.

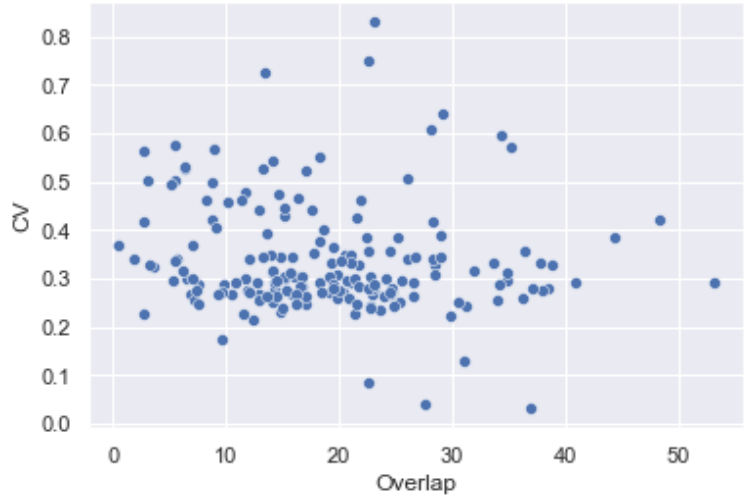
Στο επόμενο γράφημα παρουσιάζεται η σχέση μεταξύ του συντελεστή CV και του μήκους των λεωφορειακών γραμμών (Εικόνα 19).



Εικόνα 19. Σχέση του συντελεστή Διακύμανσης CV με το μήκος των λεωφορειακών γραμμών

Παρατηρείται ότι το μεγαλύτερο ποσοστό των γραμμών διαγράφει μήκος μεταξύ 5 έως 20 χιλιομέτρων (σύνολο δύο κατευθύνσεων). Υπάρχει μικρός αριθμός ακραίων τιμών οι οποίες δεν επηρεάζουν κατά πολύ τον μέσο όρο μήκους γραμμών. Από τις μεγαλύτερες σε μήκος διαδρομές διαγράφει η λεωφορειακή γραμμή Χ96 η οποία ξεκινά από τον Πειραιά και έχει προορισμό τον Αερολιμένα Αθηνών.

Στο παρακάτω διάγραμμα παρουσιάζεται η σχέση του συντελεστή CV με την επικάλυψη των λεωφορειακών γραμμών (Εικόνα 20).



Εικόνα 20. Σχέση του συντελεστή Διακύμανσης CV με την επικάλυψη των λεωφορειακών γραμμών.

Στο συγκεκριμένο διάγραμμα παρατηρείται ότι υπάρχει σχετικός διαμοιρασμός των γραμμών ως προς το ποσοστό επικάλυψης. Υπάρχει μικρός αριθμός ακραίων σημείων (>40%) , όπως για παράδειγμα η γραμμή 622 (ΓΟΥΔΗ - ΑΝΩ ΓΑΛΑΤΣΙ) η οποία διέρχεται από το κέντρο της Αθήνας χρησιμοποιώντας οδικούς άξονες που διασχίζουν επίσης πολλές κεντρικές γραμμές καθώς και για μεγάλο μέρος της διαδρομής χρησιμοποιεί κοινούς δρόμους με την λεωφορειακή γραμμή 608 (ΓΑΛΑΤΣΙ - ΑΚΑΔΗΜΙΑ - ΝΕΚΡ. ΖΩΓΡΑΦΟΥ).

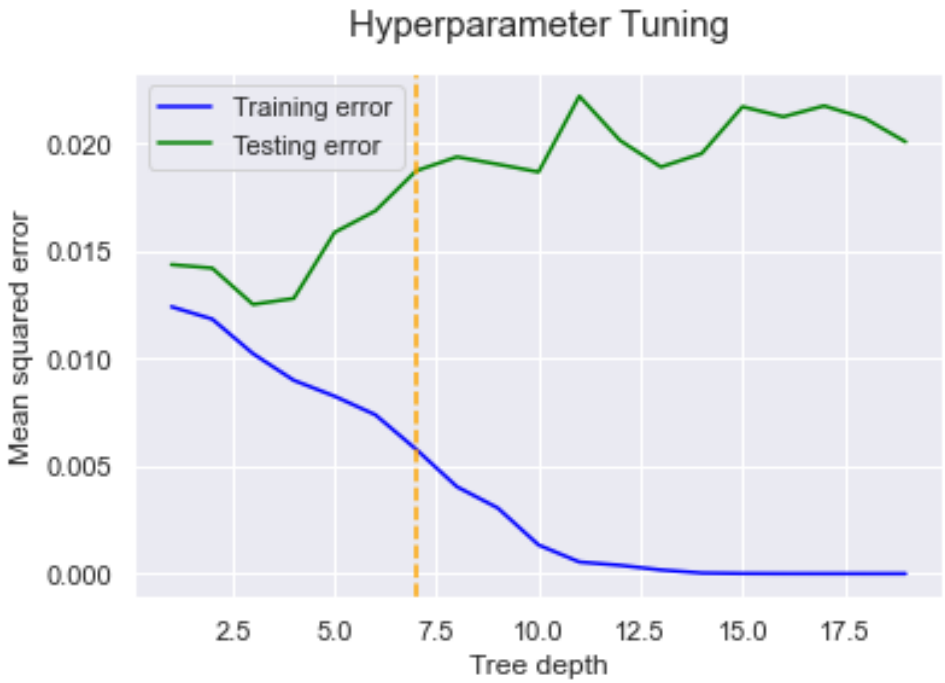
Δεύτερο βήμα της ανάλυσης αποτελεί η επιλογή της Εξαρτημένης και των Ανεξάρτητων μεταβλητών καθώς και η δημιουργία του μοντέλου.

Με βάση τα παραπάνω επιλέχθηκε ως **Εξαρτημένη** μεταβλητή ο συντελεστής διακύμανσης CV και ως **ανεξάρτητες** οι μεταβλητές “Len” ,“Overlap” και “Residential”.

Στη συνέχεια καθορίστηκε το μοντέλο με κριτήριο το μέγιστο βάθος του δένδρου (max depth). Το κριτήριο αυτό τέθηκε διότι όσο πιο βαθιά αφήνεται το δέντρο να μεγαλώσει, τόσο πιο περίπλοκο γίνεται το μοντέλο, επειδή πραγματοποιούνται περισσότερες διασπάσεις και συλλέγονται περισσότερες πληροφορίες σχετικά με τα δεδομένα. Αυτή είναι μια από τις βασικές αιτίες της υπερβολικής προσαρμογής (overfitting).

Στη συνέχεια απαιτείται ο συντονισμός υπερπαραμέτρων. Οι παράμετροι που καθορίζουν την αρχιτεκτονική του μοντέλου αναφέρονται ως υπερπαράμετροι και επομένως, η διαδικασία αναζήτησης της ιδανικής αρχιτεκτονικής μοντέλου (αυτή που μεγιστοποιεί την απόδοση του μοντέλου) αναφέρεται ως συντονισμός

υπερπαραμέτρων (Εικόνα 21). Το παρακάτω γράφημα το χρησιμοποιούμε για να διαλέξουμε το βάθος του δένδρου.



Εικόνα 21. Συντονισμός υπερπαραμέτρων.

Παρατηρείται ότι το μοντέλο κάνει υπερπροσαρμογή για μεγάλες τιμές βάθους (Πολύ μικρό training error και πολύ μεγάλο testing error). Το δέντρο προβλέπει άριστα όλα τα δεδομένα του training set, ωστόσο, αποτυγχάνει να γενικεύσει τα ευρήματα για νέα δεδομένα (test set). Παρατηρούμε επίσης ότι για βάθη μεγαλύτερα του 3 το testing error αρχίζει να αυξάνεται απότομα.

Το μοντέλο δοκιμάστηκε δύο φορές, για μέγιστο βάθος ίσο με 2 και μέγιστο βάθος ίσο με 3 και προέκυψαν οι εξής δείκτες:

Μοντέλο για μέγιστο βάθος ίσο με 3:

- Mean Absolute Error (MAE): 0.079
- Mean Squared Error (MSE): 0.012
- Root Mean Squared Error (RMSE): 0.110

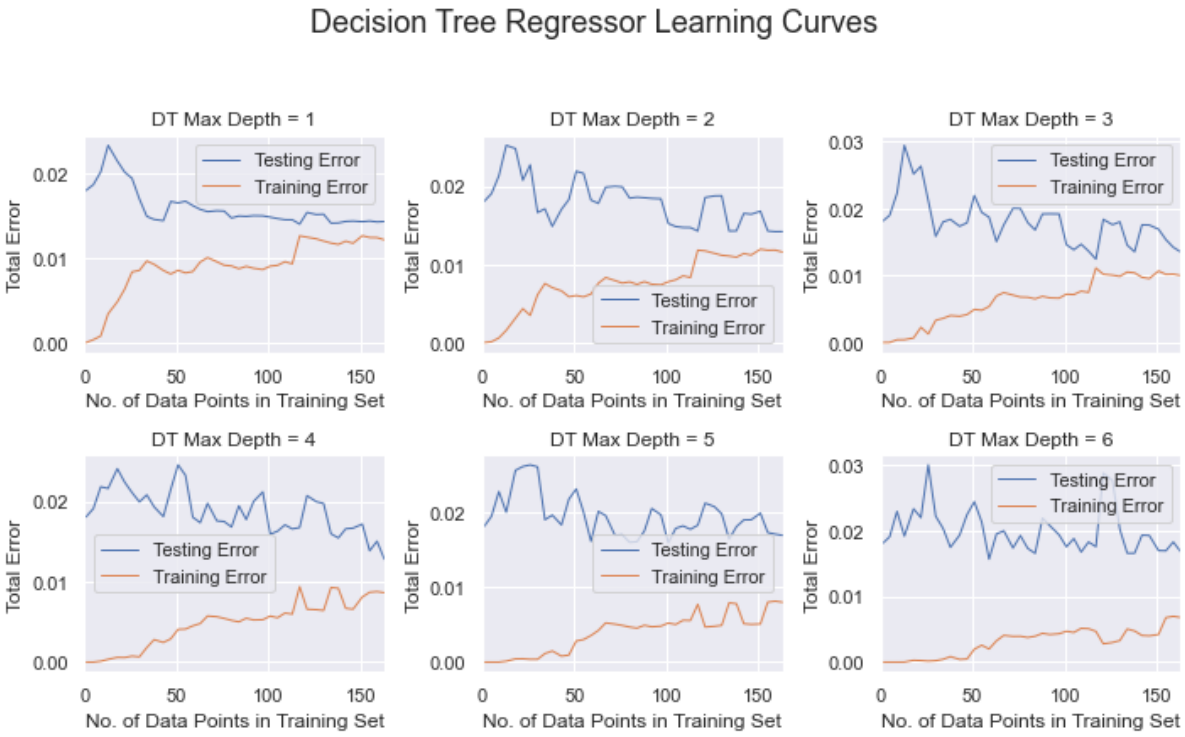
Μοντέλο για μέγιστο βάθος ίσο με 2:

- Mean Absolute Error (MAE): 0.085
- Mean Squared Error (MSE): 0.013
- Root Mean Squared Error (RMSE): 0.117

Τρίτο βήμα αποτελεί ο χωρισμός των δεδομένων σε Training set και Test set με ποσοστά 90%-10%.

Η κύρια ιδέα του διαχωρισμού του συνόλου δεδομένων σε ένα σύνολο επικύρωσης (test set) είναι να αποτραπεί η υπερβολική προσαρμογή ή υποπροσαρμογή του μοντέλου μας. Η υπερπροσαρμογή (overfitting) συμβαίνει όταν ένα μοντέλο μαθαίνει τόσο τις εξαρτήσεις μεταξύ των δεδομένων όσο και τις τυχαίες διακυμάνσεις. Με άλλα λόγια, ένα μοντέλο μαθαίνει πολύ καλά τα υπάρχοντα δεδομένα. Τα πολύπλοκα μοντέλα, τα οποία έχουν πολλά χαρακτηριστικά ή όρους, είναι συχνά επιρρεπή στην υπερπροσαρμογή. Όταν εφαρμόζονται σε γνωστά δεδομένα, τέτοια μοντέλα συνήθως αποδίδουν υψηλό R^2. Ωστόσο, συχνά δεν γενικεύονται καλά και έχουν σημαντικά χαμηλότερο R^2 όταν χρησιμοποιούνται με νέα δεδομένα. Η υποπροσαρμογή (Underfitting) συμβαίνει όταν ένα μοντέλο δεν μπορεί να συλλάβει με ακρίβεια τις εξαρτήσεις μεταξύ των δεδομένων, συνήθως ως συνέπεια της δικής του απλότητας. Συχνά αποδίδει χαμηλό R^2 με γνωστά δεδομένα και κακές δυνατότητες γενίκευσης όταν εφαρμόζεται με νέα δεδομένα.

Στο επόμενο γράφημα αναπαρίστανται οι καμπύλες εκμάθησης- Learning Curves που δείχνουν την βαθμολογία του training set και test set ενός εκτιμητή για ποικίλους αριθμούς δειγμάτων εκπαίδευσης (Εικόνα 22). Είναι ένα εργαλείο για να εκφράσει πόσο ωφελούμαστε από την προσθήκη περισσότερων δεδομένων εκπαίδευσης και εάν ο εκτιμητής υποφέρει περισσότερο από ένα σφάλμα διακύμανσης ή ένα σφάλμα μεροληψίας.

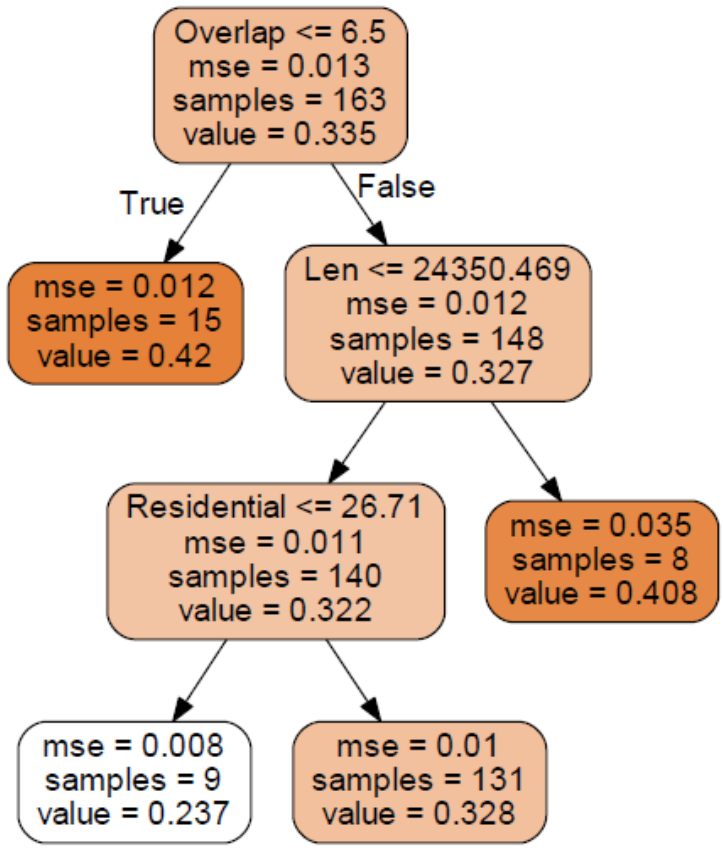


Εικόνα 22. Καμπύλες εκμάθησης δένδρου παλινδρόμησης.

Παρατηρείται από τα παραπάνω γραφήματα ότι το Testing error (μπλέ γραμμή) του Δένδρου για βάθος ίσο με 3 συνεχίζει να μειώνεται προς το τέλος του γραφήματος. Συγκρίνοντάς το με τα υπόλοιπα testing error διαφόρων βαθών, παρατηρούμε πως οπτικά φαίνεται να έχει το μικρότερο σφάλμα, κάτι που σημαίνει ότι αυτό το βάθος ίσως είναι το βέλτιστο για τα δεδομένα μας.

Όλα τα παραπάνω οδηγούν στην επιλογή του μοντέλου με μέγιστο **βάθος ίσο με 3**.

5.3 Οπτικοποίηση του Δένδρου Παλινδρόμησης



Εικόνα 23. Οπτικοποίηση Δένδρου με μέγιστο βάθος ίσο με 3.

Ο αλγόριθμος CART παίρνει ένα χαρακτηριστικό και καθορίζει ποιο σημείο διαχωρισμού **ελαχιστοποιεί τη διακύμανση** της εξαρτημένης μεταβλητής για μια εργασία παλινδρόμησης. Η διακύμανση μας λέει πόσο οι τιμές της εξαρτημένης μεταβλητής σε έναν κόμβο κατανέμονται γύρω από τη μέση τιμή τους. Ο στόχος του δέντρου ταξινόμησης είναι να χωρίσει τα δεδομένα σε μικρότερες, πιο ομοιογενείς ομάδες. Ομοιογένεια σημαίνει ότι τα περισσότερα από τα δείγματα σε κάθε κόμβο είναι από μία κατηγορία.

Επομένως, από το παραπάνω γράφημα (Εικόνα 23) παρατηρείται ότι για την μεταβλητή overlap ο διαχωρισμός έγινε για την τιμή 6.5%. Ο αλγόριθμος ξεκινά με αριθμό δείγματος 163 από τα οποία τα 148 ικανοποιούν την συνθήκη overlap> 6.5%. Τα υπόλοιπα 15 δείγματα ικανοποιούν την συνθήκη overlap<= 6.5%.

Στη συνέχεια, η μεταβλητή Len παρουσιάζει κριτήριο διαχωρισμού την τιμή 24,35 χλμ. Ο μεγαλύτερος αριθμός των δειγμάτων (140 από 148) ικανοποιεί την συνθήκη Len < 24,35 χλμ. Το τελικό κριτήριο διαχωρισμού είναι η συνθήκη Residential <= 26,71% για την οποία το μεγαλύτερο μέρος του δείγματος δεν την ικανοποιεί (131 από 140).

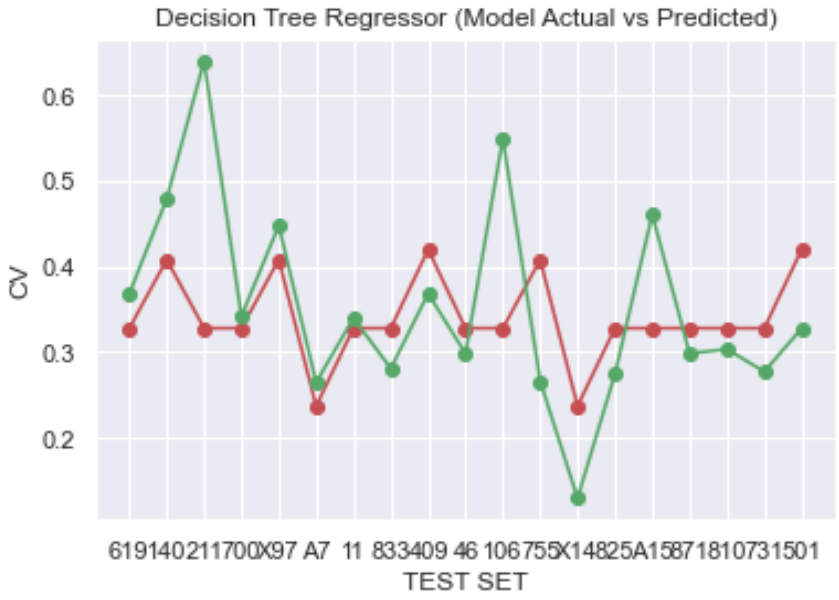
Ξεκινώντας από τον κόμβο «ρίζα» και πηγαίνοντας προς τους επόμενους κόμβους έως ότου φτάσουμε στους κόμβους «φύλλα» η συνθήκη που συνδέονται όλες οι άκρες είναι η συνθήκη AND. Επομένως παρατηρούμε ότι οι συνθήκες overlap> 6.5% **ΚΑΙ** Len < 24,35 χλμ **ΚΑΙ** Residential > 26,71% ικανοποιούνται για αριθμό δειγμάτων 131 από τα 163. Δηλαδή το μεγαλύτερο μέρος των λεωφορειακών γραμμών από το δείγμα έχει επικάλυψη μεγαλύτερη από 6.5%, μήκος μικρότερο από 24,35 χλμ και το ποσοστό κατά το οποίο διέρχεται από δρόμους κατοικημένων περιοχών ξεπερνά το 26,71%.

Με βάση τα παραπάνω:

- Αν η **επικάλυψη (Overlap)** της λεωφορειακής γραμμής είναι μεγαλύτερη από 6.5% και το **μήκος (Len)** μεγαλύτερο από 24 χλμ., παρατηρείται ότι ο συντελεστής διακύμανσης της ζήτησης CV αυξάνεται. Το ίδιο ισχύει για λεωφορειακές γραμμές με μικρό ποσοστό επικάλυψης μικρότερο του 6.5%.
- Αν η γραμμή έχει **μήκος (Len)** μικρότερο από 24 χλμ. και το ποσοστό της διαδρομής της που διέρχεται από κατοικημένη περιοχή **(Residential)** είναι μικρότερο από 26.7% , ο συντελεστής CV λαμβάνει μικρότερη τιμή επομένως δεν τόσο επιρρεπής στις αυξομειώσεις της ζήτησης.
- Για λεωφορειακές γραμμές με ποσοστό **Residential** μεγαλύτερο από 26.7% παρατηρείται αύξηση του συντελεστή CV.

5.4 Αξιολόγηση του Δένδρου Παλινδρόμησης

Στο παρακάτω γράφημα απεικονίζεται με πράσινο το πραγματικό μοντέλο και με κόκκινο το μοντέλο πρόβλεψης χρησιμοποιώντας το σύνολο επικύρωσης (test set) (Εικόνα 24).



Εικόνα 24. Μοντέλο πρόβλεψης σε σύγκριση με το πραγματικό μοντέλο

Παρατηρούμε ότι για κάποιες λεωφορειακές γραμμές οι προβλεπόμενες τιμές από το test set αποκλίνουν πολύ από τις πραγματικές, όπως οι γραμμές 211, 106 και A15. Επομένως το μοντέλο δεν παρουσιάζει καλή προσαρμογή για αυτά τα σημεία. Αντιθέτως για κάποιες γραμμές παρατηρείται ότι οι προβλεπόμενες τιμές σχεδόν ταυτίζονται με τις πραγματικές, όπως για τις γραμμές 700, A7 , 46.

Το μοντέλο παρουσίασε πολύ χαμηλά σφάλματα **MAE, MSE, RMSE** τα οποία ήταν επιθυμητό να πλησιάζουν το μηδέν.

- Mean Absolute Error (MAE): 0.079
- Mean Squared Error (MSE): 0.012
- Root Mean Squared Error (RMSE): 0.110

Το μοντέλο παρουσίασε χαμηλό $\mathbf{R^2}$ = 0.09 , κάτι που σημαίνει πως το μοντέλο δεν εξηγεί κατά τον βέλτιστο τρόπο τις διακυμάνσεις στη εξαρτημένη μεταβλητή γύρω από τον μέσο όρο της. Παρόλα αυτά, τιμές του δείκτη R2 εντός του εύρους 0.10-0.20 είναι συνηθισμένες σε αντίστοιχες εφαρμογές.

Σημειώνεται επίσης ότι ένα υψηλό ή χαμηλό R^2 δεν είναι απαραίτητα καλό ή κακό, καθώς δεν εκφράζει την αξιοπιστία του μοντέλου, ούτε αν έχει επιλεγεί η σωστή παλινδρόμηση. Μπορεί να εμφανιστεί ένα χαμηλό R^2 για ένα καλό μοντέλο ή ένα υψηλό R^2 για ένα κακώς προσαρμοσμένο μοντέλο.

6. ΣΥΜΠΕΡΑΣΜΑΤΑ

6.1 Σύνοψη Αποτελεσμάτων

Η εκπόνηση της παρούσας διπλωματικής εργασίας έχει ως στόχο τη διερεύνηση της επιρροής της πανδημίας του COVID-19 στην εξέλιξη της επιβατικής ζήτησης στο δίκτυο Μέσων Μαζικής Μεταφοράς του ΟΑΣΑ. Για τον σκοπό αυτό αναπτύχθηκαν μοντέλα Δένδρων Παλινδρόμησης για την περιγραφή και την πρόβλεψη του επιβατικού κοινού ΟΑΣΑ σε σχέση με τα χαρακτηριστικά των γραμμών.

Τα δεδομένα που αναλύθηκαν, ανακτήθηκαν από την ιστοσελίδα της Κυβέρνησης https://www.data.gov.gr/ και αφορούν τις ημερήσιες επικυρώσεις εισιτηρίων στις λεωφορειακές γραμμές. Τα στοιχεία αυτά αφορούν την περίοδο Σεπτέμβριος 2020 έως Δεκέμβριος 2021, κατά την περίοδο του δεύτερου κύματος της πανδημίας στην Ελλάδα. Επίσης αντλήθηκαν δεδομένα για τα χαρακτηριστικά των λεωφορειακών γραμμών από τον ΟΑΣΑ και την ιστοσελίδα OpenStreetMap (https://www.openstreetmap.org/).

Σύμφωνα με το θεωρητικό υπόβαθρο, αναπτύχθηκαν μοντέλα Δένδρων Παλινδρόμησης για να εξετασθεί η σχέση μεταξύ του επιβατικού κοινού και των χαρακτηριστικών μιας γραμμής.

Το μοντέλο που προέκυψε παρουσίασε χαμηλούς δείκτες σφαλμάτων καθώς και χαμηλό R^2, συμπεραίνοντας ότι είναι ένα σχετικά καλό μοντέλο.

Από το προηγούμενο κεφάλαιο εφαρμογής της μεθοδολογίας, προέκυψε μία σειρά συμπερασμάτων που δίνουν απάντηση στα αρχικά ερωτήματα της παρούσας Διπλωματικής Εργασίας.

Η μεγαλύτερη διακύμανση της ζήτησης παρατηρείται σε περιπτώσεις όπου η επικάλυψη των γραμμών είναι πολύ μικρή (overlap<6.5%) και σε περιπτώσεις όπου το μήκος των λεωφορειακών γραμμών είναι μεγαλύτερο των 24 χιλιομέτρων. Αυτό οφείλεται στο γεγονός ότι κατά την περίοδο των περιοριστικών μέτρων υπήρξε μια μείωση στις άσκοπες μετακινήσεις με αποτέλεσμα οι μετακινήσεις που παραμένουν να γίνονται κοντά στον τόπο κατοικίας μέσω τοπικών λεωφορειακών γραμμών.

Επιπλέον, σχετικά μεγάλη διακύμανση της ζήτησης παρατηρήθηκε σε περιπτώσεις όπου η επικάλυψη των γραμμών και το ποσοστό της διαδρομής της που διέρχεται από κατοικημένη περιοχή είναι αυξημένα (overlap>6.5% και residential>26.7%). Πρόκειται για λεωφορειακές γραμμές που διασχίζουν πυκνά κατοικημένες περιοχές της πόλης και χρησιμοποιούν κοινές οδούς με άλλες γραμμές. Ενώ αυτά είναι χαρακτηριστικά που βοηθούν στις μετεπιβιβάσεις και στην επιλογή περισσοτέρων συνδυασμών μετακίνησης ένας σημαντικός παράγοντας που πρέπει να ληφθεί υπόψη είναι ο φόβος των μετακινούμενων για τον κορονοϊό καθώς και η προσπάθεια τους να αποφύγουν τις εστίες μετάδοσης και τον συνωστισμό.

Τέλος, μικρότερη διακύμανση στη ζήτηση παρατηρήθηκε σε περιπτώσεις όπου το ποσοστό της διαδρομής των λεωφορειακών γραμμών που διέρχεται από κατοικημένη περιοχή είναι μικρότερο από 26.7%. Αυτό ίσως να οφείλεται στο ότι οι γραμμές αυτές εξυπηρετούν περιοχές μακριά από το κέντρο της πόλης ή περιοχές με άλλες χρήσεις γης. Σε αυτές τις περιπτώσεις οι μετακινούμενοι προτίμησαν μετάβαση από τη χρήση δημοσίων μέσων μεταφοράς σε ιδιωτικά και ατομικά μέσα.

6.3 Προτάσεις για Περαιτέρω Έρευνα

Η πανδημία του κορονοϊού επέφερε σημαντικές αλλαγές στις καθημερινότητες των ανθρώπων, στην κινητικότητα καθώς και στις επιλογές των μετακινούμενων. Καθώς όμως γίνεται ετοιμασία για την επιστροφή στην κανονικότητα, οι μετακινήσεις αυξάνονται ξανά σε επίπεδα προ κορονοϊού και η χρήση των μέσων μαζικής μεταφοράς είναι αναπόφευκτη. Επομένως κρίνεται απαραίτητο να επέλθουν κάποιες αλλαγές στις δημόσιες συγκοινωνίες.

Κάποιες αλλαγές που μπορούν να εφαρμοστούν από τον ΟΑΣΑ, αφορούν τις συχνότητες των δρομολογίων. Προτείνεται η αύξηση των δρομολογίων για λεωφορειακές γραμμές που διέρχονται από πυκνά κατοικημένες περιοχές με μεγάλο επιβατικό κοινό καθώς και η δημιουργία νέων γραμμών που θα έχουν ανταπόκριση με διαφορετικές υφιστάμενες λεωφορειακές γραμμές, δηλαδή με μεγάλο ποσοστό επικάλυψης. Επιπλέον, μέτρα για τη μείωση του φόβου μετακίνησης μπορούν να εφαρμοστούν από τεχνολογίες όπως η τεχνητή νοημοσύνη (AI) από τις οποίες μπορούν να γίνει ο εντοπισμός του κινδύνου μετάδοσης κατά τη διάρκεια ταξιδιών με τα δημόσια μέσα μεταφοράς.

Όσον αφορά την ανάλυση του μοντέλου υπάρχουν κάποιες προσθήκες που μπορούν να οδηγήσουν σε βελτίωση αυτού. Μπορούν να αξιοποιηθούν από το Πανεπιστήμιο της Οξφόρδης δεδομένα της κλίμακας αυστηρότητας των περιοριστικών μέτρων ώστε να γίνει λεπτομερέστερη ανάλυση για κάθε επίπεδο αυστηρότητας. Επιπλέον η ανάλυση μπορεί να επεκταθεί για άλλες χώρες καθώς επίσης να εξετασθούν οι μεταβολές κινητικότητας σε ενδεχομένως επόμενα «κύματα» ή στην περίπτωση νέας πανδημίας.

7. ΒΙΒΛΙΟΓΡΑΦΙΑ

1. Abdullah, M., Dias, C., Muley, D., Shahin, Md. (2020). Exploring the impacts of COVID-19 on travel behavior and mode preferences, Transp. Res. Interdiscip. Perspect. 8

2. Aloi, A., Alonso, B., Benavente, J., Cordera, R., Echániz, E., González, F., Ladisa, C., Lezama-Romanelli, R., López-Parra, Á., Mazzei, V., Perrucci, L., Prieto-Quintana, D., Rodríguez, A., Sañudo, R. (2020). Effects of the COVID-19 Lockdown on Urban Mobility: Empirical Evidence from the City of Santander (Spain), Sustainability 12

3. Badr, H.S., Du, H., Marshall, M., Dong, E., Squire, M.M., Gardner, L.M. (2020). Association between mobility patterns and COVID-19 transmission in the USA: a mathematical modelling study.

4. Beria, P., Lunkar, V. (2021). Presence and mobility of the population during the first wave of COVID-19 outbreak and lockdown in Italy, Sustainable Cities and Society 65

5. Bucsky, P. (2020). Modal share changes due to COVID-19: The case of Budapest, Transp. Res. Interdiscip. Perspect. 8

6. Campisi, T., Basbas, S., Skoufas, A., Akgün, N., Ticali, D., Tesoriere, G. (2020). The Impact of COVID-19 Pandemic on the Resilience of Sustainable Mobility in Sicily, Sustainability 12

7. Combs, T., Pardo, C. (2021). Shifting streets COVID-19 mobility data: Findings from a global dataset and a research agenda for transport planning and policy, Transp. Res. Interdiscip. Perspect. 9

8. de Haas, M., Faber, R., Hamersma, M. (2020). How COVID-19 and the Dutch 'intelligent lockdown' change activities, work and travel behaviour: Evidence from longitudinal data in the Netherlands. Transp. Res. Interdiscip. Perspect. 6

9. Hadjidemetriou, G.M., Sasidharan, M., Kouyialis, G., Parlikad, A.K. (2020). The impact of government measures and human mobility trend on COVID- 19 related deaths in the UK. Transp. Res. Interdiscip. Perspect. 6

10. Hasselwander, M., Tamagusko, T., Bigotte, J.F., Ferreira, A., Mejia, A., Ferranti, E.J.S. (2021). Building back better: The COVID-19 pandemic and transport policy implications for a developing megacity. Sustainable Cities and Society 69

11. Hu,S., Xiong, C., Liu, Z., Zhang, L. (2021). Examining spatiotemporal changing patterns of bike-sharing usage during COVID-19 pandemic. Journal of Transport Geography 91

12. Jenelius, E., Cebecauer, M. (2020). Impacts of COVID-19 on public transport ridership in Sweden: Analysis of ticket validations, sales and passenger counts. Transp. Res. Interdiscip. Perspect. 8

13. 19. Molloy, J., Tchervenkov, C., Hintermann, B., Axhausen, K.W. (2020). Tracing the Sars- CoV-2 Impact: The first month in Switzerland. Transport Findings

14. Katrakazas, C., Michelaraki, E., Sekadakis, M., Yannis, G. (2020). A descriptive analysis of the effect of the COVID-19 pandemic on driving behavior and road safety. Transp. Res. Interdiscip. Perspect. 7

15. Khaddar, S., Fatmi, M.R. (2021). COVID-19: Are you satisfied with traveling during the pandemic? Transp. Res. Interdiscip. Perspect. 9

16. Kim, C., Cheon, S.H., Choi, K., Joh, C.H., Lee, H.J. (2017). Exposure to fear: Changes in travel behavior during MERS outbreak in Seoul. KSCE J. Civ. Eng. 21

17. Maiti, A., Zhang, Q., Sannigrahi, S., Pramanik, S., Chakraborti, S., Cerda, A., Pilla, F. (2021). Exploring spatiotemporal effects of the driving factors on COVID-19 incidences in the contiguous United States. Sustainable Cities and Society 68

18. Molloy, J., Schatzmann, T., Schoeman, B., Tchervenkov, C., Hintermann, B., Axhausen, K.W. (2021). Observed impacts of the Covid-19 first wave on travel behaviour in Switzerland based on a large GPS panel. Transport Policy 104

19. Muley, D., Ghanim, M.S., Mohammad, A., Kharbeche, M., (2021). QUANTIFYING THE IMPACT OF COVID–19 PREVENTIVE MEASURES ON TRAFFIC IN THE STATE OF QATAR, Transport Policy 103

20. Nouvellet, P., Bhatia, S., Cori, A., Ainslie, K.E.C., Baguelin, M., Bhatt, S., Boonyasiri, A., Brazeau, N.F., Cattarino, L. , Cooper, L.V., Coupland, H., Cucunuba, Z.M., Cuomo-Dannenburg, G., Dighe, A., Djaafara, B.A.,Dorigatti, I., Eales, O.D., et al. (2020). Reduction in mobility and COVID-19 transmission.

21. Padmanabhan, V., Penmetsa, P., Li, X., Dhondia, F., Dhondia, S., Parrish,A. (2021). COVID-19 effects on shared-biking in New York, Boston, and Chicago. Transp. Res. Interdiscip. Perspect. 9

22. Parr, S., Wolshon, B., Renne, J., Murray-Tuite, P., Kim, K. (2020). Traffic Impacts of the COVID-19 Pandemic: Statewide Analysis of Social Separation and Activity Restriction, American Society of Civil Engineers

23. Pullano, G., Valdano, E., Scarpa, N., Rubrichi, S., Colizza, V. (2020). Evaluating the effect of demographic factors, socioeconomic factors, and risk aversion on mobility during the COVID-19 epidemic in France under lockdown: a population-based study, The Lancet Digital Health38.

24. Saladié, Ò., Bustamante, E., Gutiérrez, A. (2020). COVID-19 lockdown and reduction of traffic accidents in Tarragona province. Spain. Transp. Res. Interdiscip. Perspect. 8

25. Santamaria, C., Sermi, F., Spyratos, Sp., Iacus, S. M., Annunziato, A., Tarchi, D, Vespe,

M. (2020). Measuring the impact of COVID-19 confinement measures on human mobility using mobile positioning data. A European regional analysis, Safety Science 132

26. Shakibaei, S., de Jong, G. C., Alpkökin, P., Rashidi, T. H. (2021). Impact of the COVID-19 pandemic on travel behavior in Istanbul: A panel data analysis, Sustainable Cities and Society 65

27. Shamshiripour, A., Rahimi, E., Shabanpour, R., Mohammadian, A.K. (2020). How is COVID-19 reshaping activity-travel behavior? evidence from a comprehensive survey in chicago. Transp. Res. Interdiscip. Perspect. 7

28. Tarasi, D., Daras, T., Tournaki. S., Tsoutsos, T. (2021). Transportation in the Mediterranean during the COVID-19 pandemic era. Global Transitions 3

29. Thakkar, N., Burstein, R., Hu, H., Selvaraj, P., Klein, D. (2020). Social distancing and mobility reductions have reduced COVID-19 transmission in King County, WA.

31. 30. ΕΟΔΥ 2021. Εθνικός Οργανισμός Δημόσιας Υγείας. (2021). Covid-19 Οδηγίες - Εθνικός Οργανισμός Δημόσιας Υγείας. https://eody.gov.grhttps://el.wikipedia.org/wiki/Μηχανική-Μάθηση

32. https://towardsdatascience.com/the-complete-guide-to-decision-trees-28a4e3c7be14

33. https://towardsdatascience.com/modelling-regression-trees-b376e959d02e

34. https://scikit-learn.org/stable/modules/tree.html

35. https://www.data.gov.gr/

36. https://www.bsg.ox.ac.uk/research/research-projects/covid-19-government-response-tracker

8. ΠΑΡΑΡΤΗΜΑ Α

Κώδικας python για τα Δένδρα Παλινδρόμησης.

```
import matplotlib.pyplot as plt
import numpy as np
import pandas as pd
import seaborn as sns

#Load Dataset
sns.set()

df = pd.read_csv(r'C:/Users/Natassa/Desktop/diplomatiki/decision tree/df_weekly_cv.csv',
index_col=0, delimiter=',', encoding="utf-8-sig")
df.columns = ["Len", "Overlap", "Residential", "Cases", "Stringency", "CV", "CVPercentile"]

sns.scatterplot(x=df['Residential'], y=df['CV'])
sns.scatterplot(x=df['Len'], y=df['CV'])
sns.scatterplot(x=df['Overlap'], y=df['CV'])

plt.savefig('scatterplot.png')


#Import the class
from sklearn.tree import DecisionTreeRegressor

target = df.get('CV')
y = target

df1 = df.copy()
df1 = df1.drop('CV', axis =1)
df1 = df1.drop('Cases', axis =1)
df1 = df1.drop('Stringency', axis =1)
df1 = df1.drop('CVPercentile', axis=1)

X = df1
y = target


#Create an object (model)
dtr1 = DecisionTreeRegressor(max_depth=2,random_state=1)

dtr1.fit(X, y)

#sns.scatterplot(x=df['Residential'], y=df['CV'])
#plt.plot(X.sort_values(),dtr1.predict(X.sort_values().to_frame()), color='red', label='model',
linewidth=2)
#plt.legend()
#plt.savefig('model.png')
```

```
#We can visualize the tree diagram of this model using Graphviz.
from sklearn.tree import export_graphviz
import graphviz

dot_data = export_graphviz(dtr1, feature_names=X.columns.values, filled=True,
rounded=True)
graph = graphviz.Source(dot_data)
graph.render("tree")

#Hyperparameter tuning
#Using Scikit-learn train_test_split() function
from sklearn.model_selection import train_test_split

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.10, random_state=0,
shuffle=True)

from sklearn.metrics import mean_squared_error as mse

max_depths = range(1, 20)
training_error = []
for max_depth in max_depths:
 model_1 = DecisionTreeRegressor(max_depth=max_depth)
 model_1.fit(X, y)
 training_error.append(mse(y, model_1.predict(X)))

testing_error = []
for max_depth in max_depths:
 model_2 = DecisionTreeRegressor(max_depth=max_depth)
 model_2.fit(X_train, y_train)
 testing_error.append(mse(y_test, model_2.predict(X_test)))

plt.plot(max_depths, training_error, color='blue', label='Training error')
plt.plot(max_depths, testing_error, color='green', label='Testing error')
plt.xlabel('Tree depth')
plt.axvline(x=7, color='orange', linestyle='--')
plt.annotate('optimum = 7', xy=(7.5, 1.17), color='red')
plt.ylabel('Mean squared error')
plt.title('Hyperparameter Tuning', pad=15, size=15)
plt.legend()
plt.savefig('error.png')

#Using k-fold cross-validation
from sklearn.model_selection import GridSearchCV

model = DecisionTreeRegressor()

gs = GridSearchCV(model, param_grid = {'max_depth': range(1, 11),'min_samples_split':
range(10, 60, 10)}, cv=5,n_jobs=1,scoring='neg_mean_squared_error')

gs.fit(X_train, y_train)
```

```
print(gs.best_params_)
print(-gs.best_score_)


#best model
#sns.scatterplot(x=df['Longitude'], y=df['MedHouseVal'],label='data')

new_model = DecisionTreeRegressor(max_depth=3,
                        min_samples_split=20,
                        min_samples_leaf=2)

#or new_model = gs.best_estimator_
new_model.fit(X_train, y_train)

#plt.plot(df['Longitude'].sort_values(),
new_model.predict(df['Longitude'].sort_values().to_frame()),color='red',
label='model',linewidth=2)
#plt.legend()
#plt.title('Best Fitting', pad=15, size=15)
#plt.savefig('new_model.png')


dot_data = export_graphviz(new_model, feature_names=X.columns.values, filled=True,
rounded=True)
graph = graphviz.Source(dot_data)
graph.render("tree")

#Model Evaluation
y_pred = new_model.predict(X_test)
evaluation_df=pd.DataFrame({'Actual':y_test, 'Predicted':y_pred})

from sklearn import metrics
print('Mean Absolute Error:', metrics.mean_absolute_error(y_test, y_pred))
print('Mean Squared Error:', metrics.mean_squared_error(y_test, y_pred))
print('Root Mean Squared Error:', np.sqrt(metrics.mean_squared_error(y_test, y_pred)))
print("RSquared: " , np.round(metrics.r2_score(y_test, y_pred), 2))

#Actual vs Predicted Grapv
plt.figure()
plt.title("Decision Tree Regressor (Model Actual vs Predited)")
plt.xlabel('TEST SET')
plt.ylabel('CV')
plt.plot(y_pred, 'o-', color="r", label="Predicted")
plt.plot(y_test, 'o-', color="g", label="Actual")


def performance_metric(y_true, y_predict):
   error = metrics.mean_squared_error(y_true, y_predict)
   return error
```

```
fig = plt.figure(figsize=(10,8))
# Generate 40 evenly spaced numbers (rounded to nearest integer)
datapoints = np.rint(np.linspace(1, len(X_train), 40)).astype(int)
#initialise array of shape (40,)
train_err = np.zeros(len(datapoints))
test_err = np.zeros(len(datapoints))



# Create 6 different models based on max_depth
for k, depth in enumerate(range(1,7)):
    for i, s in enumerate(datapoints):
        reg = DecisionTreeRegressor(max_depth = depth) #increasing depth
# Iteratively increase training set size
        reg.fit(X_train[:s], y_train[:s])
# MSE for training and test sets of increasing size
        train_err[i] = performance_metric(y_train[:s], reg.predict(X_train[:s]))
        test_err[i] = performance_metric(y_test, reg.predict(X_test))

# Subplot learning curves
    sub = fig.add_subplot(3, 3, k+1)
    sub.plot(datapoints, test_err, lw = 1, label = 'Testing Error')
    sub.plot(datapoints, train_err, lw = 1, label = 'Training Error')
    sub.legend()
    sub.set_title('DT Max Depth = %s'%(depth))
    sub.set_xlabel('No. of Data Points in Training Set')
    sub.set_ylabel('Total Error')
    sub.set_xlim([0, len(X_train)])

    fig.suptitle('Decision Tree Regressor Learning Curves', fontsize=18, y=1.03)
    fig.tight_layout()
```