



Τμήμα Πληροφορικής
Πανεπιστήμιο Πειραιώς

Ανάλυση δεδομένων κινητικότητας
χρηστών κατά την διάρκεια της πανδημίας
με την χρήση τεχνικών data analytics και
machine learning

Urban Data Science and Covid-19 related
movement data

Επιβλέπων καθηγητής : κ. Ιωάννης
Θεοδωρίδης

Δημήτριος Δημητρίου Π18036

Ιούνιος 2023

Ευχαριστίες

Θα ήθελα να εκφράσω τις θερμές ευχαριστίες μου στον επιβλέποντα καθηγητή μου κ. Ιωάννη Θεοδωρίδη για την καθοδήγηση, την υπομονή και το ενδιαφέρον που επέδειξε κατά την διάρκεια εκπόνησης της πτυχιακής μου εργασίας. Επιπλέον, θα ήθελα να ευχαριστήσω τον υποψήφιο διδάκτορα κ. Ανδρέα Τριτσαρώλη για τη στήριξη και τη βοήθεια που μου προσέφερε σε δυσκολίες και απορίες που προέκυψαν κατά την συγγραφή αυτής της εργασίας. Τέλος ένα μεγάλο ευχαριστώ ανήκει στην οικογένεια μου και στους φίλους μου για την στήριξη και την κατανόηση τους.

Περίληψη

Η πανδημία του COVID-19 αποτελεί μια από τις μεγαλύτερες προκλήσεις που αντιμετώπισαν οι σύγχρονες κοινώνιες κάθως επηρέασε ριζικά κάθε πτυχή της ανθρώπινης ζωής. Εν απουσίᾳ προγενέστερης γνώσης σχετικά με τον ιό και τις επιπτώσεις του, οι κυβερνήσεις έλαβαν πρωτοφανή μέτρα περιορισμού της ανθρώπινης κινητικότητας ώστε να αναχαιτιστεί η εξάπλωση του ίου στον γενικό πληθυσμό και να δοθεί κατάλληλη υποστήριξη στα συστήματα υγειονομικής περίθαλψης. Σκοπός της εργασίας είναι η βιβλιογραφική μελέτη της επίδρασης του COVID-19 στην ανθρώπινη κινητικότητα και στους τρόπους μετακίνησης στα μεγάλα αστικά κέντρα καθώς και η ανάπτυξη μοντέλων μηχανικής μάθησης (MLP, RNN, LSTM, Random Forests, SVR, XGBoost, CNN, CNN-LSTM, GRU) για την πρόβλεψη της κινητικότητας των πολίτων στα dataset της Google και της Apple και των σημαντικών δεικτών της πανδημίας (π.χ., αριθμός των θανάτων, των κρουσμάτων) μέσα από πραγματικά σενάρια.

Λέξεις κλειδιά: Πανδημία COVID-19, Δεδομένα κινητικότητας, Αναλυτική δεδομένων, Μηχανική Μάθηση, Urban data science, Πρόβλεψη χρονοσειρών

Περιεχόμενα

1 Εισαγωγή	1
2 Βιβλιογραφική έρευνα - Σχετικές εργασίες	3
2.1 Ανθρώπινη κινητικότητα και εξάπλωση του ιού	3
2.2 Βάσεις δεδομένων COVID-19 και εργαλεία οπτικοποιήσης	9
2.3 Εφαρμογές Contact Tracing	11
2.4 Ο τομέας του Urban Data Science και οι προκλήσεις που αντιμετωπίζει	13
2.5 Επιπτώσεις της πανδημίας στους τρόπους μετακίνησης των πολιτών και στα αστικά κέντρα	15
2.6 Ευρωπαϊκά προγράμματα και πρωτοβουλίες για τον μετασχηματισμό των πόλεων	19
3 Θεωρητικό υπόβαθρο	21
3.1 Multilayer Perceptron	21
3.2 Μηχανές διανυσματικής υποστήριξης	23
3.3 Δέντρα απόφασης	25
3.4 Extreme Gradient Boosting	26
3.5 Αναδρομικά νευρωνικά δίκτυα	27
3.6 Δίκτυα μακράς βραχύχρονης μνήμης	28
3.7 Gated recurrent unit	30
3.8 Συνελικτικά νευρωνικά δίκτυα	30
4 Πηγές δεδομένων	33
5 Προεπεξεργασία	40
6 Στατιστική ανάλυση	42
6.1 Google COVID-19 Community Mobility Reports	42

6.2	Apple COVID-19 Mobility Trends Reports	45
6.3	COVID-19 Response Greece	47
6.4	Sandbird COVID-19-Greece	54
6.5	iMedD-Lab - OpenData	55
7	Πειραματική μελέτη	57
7.1	Περιγραφή των μοντέλων	57
7.2	Πειραματικά αποτελέσματα	61
8	Πειραματικά αποτελέσματα σε πραγματικά σενάρια	63
8.1	Αυστηρή καραντίνα (lockdown) στην Ελλάδα	63
8.2	Πρόβλεψη δεικτών Οδήγησης και Περπατήματος στην Αττική	66
8.3	Πρόβλεψη κινητικότητας στις κατηγορίες της Google στην Περιφέρεια της Δυτικής Ελλάδας	69
8.4	Πρόβλεψη εβδομαδιαίων θανάτων από COVID-19	71
9	Συμπεράσματα - Προτάσεις για μελλοντική έρευνα	75
9.1	Συμπεράσματα	75
9.2	Προτάσεις για μελλοντική μελέτη	76

Λίστα εικόνων

3.1	Η μορφή ενός νευρώνα ο οποίος δέχεται διανύσματα εισόδου και χρησιμοποιώντας την συνάρτηση ενεργοποιήσης παράγει την έξοδο γ	22
3.2	Η μορφή ενός Multilayer Perceptron νευρωνικού δικτύου	23
3.3	Σχηματική απεικόνιση SVM και SVR	24
3.4	Μετασχηματισμός του χώρου προβλήματος σε ανώτερη διάσταση έτσι ώστε το πρόβλημα από μη γραμμικό να μετατραπεί σε γραμμικό με την χρήση της συνάρτησης πυρήνα	25
3.5	Σχηματική αναπαράσταση του αλγορίθμου του τυχαίου δάσους .	26
3.6	Βασική δομή RNN και ξεδίπλωμα στο χρόνο όπου U, W, V τα βάρη του δικτύου, x η είσοδος στο δίκτυο, h η κρυφή κατάσταση και o η έξοδος την χρονική στιγμη t	28
3.7	Εσωτερική δομής μιας μονάδας δικτύου LSTM	29
3.8	Σχηματική αναπαράσταση του GRU δικτύου	30
3.9	Σχηματική αναπαράσταση ενός CNN	31
4.1	Στιγμιότυπο δεδομένων του αρχείου intensive_care_cases.json στο σύνολο δεδομένων COVID-19-Response Greece	36
4.2	Στιγμιότυπο δεδομένων του πίνακα apple_mobility_trends	37
4.3	Εικόνα της βάσης με τους αντίστοιχους πίνακες	39

6.1	Αναπαράσταση των εβδομαδιαίων δεικτών μέσης κινητικότητας των Ελλήνων πολιτών κατά την διάρκεια της πανδημίας στις κατηγορίες τοποθεσίας της Google στο ίδιο γράφημα. Ο δείκτης κινητικότητας απεικονίζεται με κόκκινο χρώμα στις δημόσιες συγκοινωνίες, με μπλε στα καταστήματα λιανικής και φυχαγωγίας, με πορτοκαλί στις αγορές τροφίμων και στα φαρμακεία, με γκρι στους χώρους εργασίας, με πράσινο στα πάρκα και με μωβ στις κατοικίες	42
6.2	Ιστόγραμμα για τους έξι δείκτες κινητικότητας των χρηστών της Google πριν την εφαρμογή της προεπεξεργασίας	43
6.3	Ημερήσια κινητικότητα των χρηστών της Google στην Αττική ανά τοποθεσία	43
6.4	Αναπαράσταση στο ίδιο γράφημα των ημερήσιων δεικτών κινητικότητας των πολιτών κατά την διάρκεια της πανδημίας ανά τοποθεσία/δραστηριότητα. Η κινητικότητα απεικονίζεται με κόκκινο χρώμα στην αποκεντρωμένη διοίκηση της Μακεδονίας και Θράκης, με μπλε στην αποκεντρωμένη διοίκηση του Αιγαίου, με πορτοκαλί στην αποκεντρωμένη διοίκηση της Μακεδονίας και Θράκης, με μαύρο στην αποκεντρωμένη διοίκηση της Αττικής, με πράσινο στην αποκεντρωμένη διοίκηση της Κρήτης και με λαχανί στην αποκεντρωμένη διοίκηση της Θεσσαλίας και κεντρικής Ελλάδας	44
6.5	Boxplot για τους δείκτες οδήγησης και περπατήματος μετά το preprocessing	45
6.6	Ιστόγραμμα για τους δείκτες οδήγησης και περπατήματος	45
6.7	Αναπαράσταση των ημερήσιων δεικτών οδήγησης και περπατήματος από την Apple που προσομειώνουν την κινητικότητα των πολιτών καθόλη την διάρκεια της πανδημίας	46
6.8	Αναπαράσταση των ημερήσιων χρονοσειρών οδήγησης και περπατήματος της Apple της Ελλάδας και της Αττικής. Με κόκκινο χρώμα απεικονίζεται ο δείκτης της οδήγησης και με μπλε αυτός του περπατήματος	46
6.9	Bar plot για συνολικά κρούσματα COVID-19 ανά νομό της Ελλάδας	47
6.10	Pie plot για συνολικά κρούσματα COVID-19 ανά αποκεντρωμένη διοίκηση της Ελλάδας	48
6.11	Line plot που δείχνει τον αριθμό των ανθρώπων που βρίσκονται στις ΜΕΘ των νοσοκομείων από λοίμωξη COVID-19	48

viii ΛΙΣΤΑ ΕΙΚ'ΟΝΩΝ

6.12 Bar plot για εμβολιασμούς κατά του COVID-19 με πρώτη, δεύτερη και τρίτη δόση καθώς και για το σύνολο των εμβολιασμών που πραγματοποιήθηκαν ανά νομό της Ελλάδας	49
6.13 Εβδομαδιαία κρούσματα COVID-19 ανά νομό της Ελλάδας	50
6.14 Plot της πορείας των συνολικών κρουσμάτων COVID-19 ανά νομό της Ελλάδας	51
6.15 Η πορεία των συνολικών εμβολιασμών απεικονίζεται με μπλε χρώμα, με πορτοκαλί οι εμβολιασμοί με τουλάχιστον μια δόση, με πράσινο οι ολοκληρωμένοι εμβολιασμοί και κόκκινο οι εμβολιασμοί με τρείς δόσεις	52
6.16 Ήμερησίος αριθμός εμβολιασμών κατά του COVID-19 που πραγματοποιήθηκαν ανά νομό της Ελλάδας	53
6.17 Συνολικοί θάνατοι COVID-19 και ποσοστά εκδήλωσης αυτών με βάση την ηλικία και το φύλο. Με μπλε χρώμα αναπαρίστανται τα δεδομένα που αφορούν τους άντρες και με ροζ αντίστοιχα τις γυναίκες	54
6.18 Αριθμός ασθενών COVID-19 που βρίσκονται στις ΜΕΘ με βάση την ηλικία και το φύλο	54
6.19 Εβδομαδιαίοι θάνατοι από COVID-19	55
6.20 Πορεία ημερήσιων συνολικών θανάτων από COVID-19	55
6.21 Εβδομαδιαία πανελληγία κρούσματα από COVID-19	56
6.22 Πορεία ημερήσιων πανελληγίων κρουσμάτων από COVID-19	56
 8.1 Δείκτες κινητικότητας Οδήγησης και Περπατήματος στην Αττική	67
8.2 Ήμερησιοί εμβολιασμοί στην Αττική	67
8.3 Πρόβλεψη δεικτών της Apple στην Αττική με GRU με βάση τους εμβολιασμούς και τους αντίστοιχους δείκτες	68
8.4 Δείκτες κινητικότητας στις κατηγορίες Google στην Περιφέρεια της Δυτικής Ελλάδας	70
8.5 Ήμερησια κρούσματα στην Περιφέρεια της Δυτικής Ελλάδας	70
8.6 Πρόβλεψη δεικτών κινητικότητας στις κατηγορίες της Google στην Περιφέρεια Δυτικής Ελλάδας με βάση τα κρούσματα και τους αντίστοιχους δείκτες	71
8.7 Εβδομαδιαία μέση κινητικότητα στις κατηγορίες της Google στην Ελλάδα	72
8.8 Εβδομαδιαίοι μέσοι θάνατοι στην Ελλάδα	72
8.9 Εβδομαδιαίος μέσος αριθμός ατόμων που νοσηλεύονται σε μονάδες εντατικής θεραπείας στην Ελλάδα	73

- 8.10 Πρόβλεψη μέσων εβδομαδιαίων θανάτων στην Ελλάδα χρησιμοποιώντας την μέση εβδομαδιαία κινητικότητα στις κατηγορίες μερών της Google, τον μέσο εβδομαδιαίο αριθμό ατόμων σε ΜΕΘ και τον αντίστοιχο δείκτη 73

1. Εισαγωγή

Το αντικείμενο της εν λόγω πτυχιακής εργασίας, είναι η ανάλυση δεδομένων που αντικατοπτρίζουν την κινητικότητα των χρηστών στην διάρκεια της πανδημίας που οφείλεται στον ιό SARS-CoV-2 (COVID 19) με τεχνικές Αναλυτικής Δεδομένων και Μηχανικής Μάθησης, με στόχο να προσδιοριστούν οι επιπτώσεις της στις ανθρώπινες μετακινήσεις. Η παρούσα εργασία πραγματεύεται ένα αρκετά πρωτότυπο και σύγχρονο θέμα, καθώς αποτελεί ένα από τα βασικότερα προβλήματα που ταλανίζουν τις κοινωνίες ανά την υφήλιο και τα συμπεράσματα που θα εξαχθούν ενδεχομένως να αποτελέσουν παρακαταθήκη στις επόμενες γενιές και στην επιστημονική κοινότητα σε περίπτωση ανάγκης αντιμετώπισης και διαχείρισης παρόμοιων μελλοντικών πανδημιών.

Ο COVID-19 είναι μια μεταδοτική ασθένεια που προκύπτει από τον ιό SARS-CoV-2, ένα νέο στέλεχος κορωνοϊού που μέχρι το 2019 δεν είχε απομονωθεί στον άνθρωπο, που προσβάλλει το αναπνευστικό σύστημα και μπορεί να προκαλέσει πνευμονική ίνωση, όπως και άλλοι κορώνα-ιοί, ακόμα και σε άτομα με ήπια ή καθόλου συμπτώματα. Ο ιός μεταδίδεται από άνθρωπο σε άνθρωπο κυρίως μέσω του αναπνευστικού με σταγονίδια από το φτέρνισμα, το βήχα, την ομιλία, την αναπνοή και το τραγούδι καθώς και λιγότερο συχνά με την επαφή μολυσμένων επιφανειών και αντικειμένων από σταγονίδια.

Η πρώτη καταγραφή του ιού έγινε στην πόλη Wuhan της Κίνας στα τέλη του 2019 όπου εκδηλώθηκε και παρουσιάζει ποικιλία συμπτωμάτων με πιο συχνά αυτά του πυρετού, του ξηρού βήχα, της σωματικής εξάντλησης, της απώλειας γεύσης/όσφρησης, του πονόλαιμου, των μυαλγιών και αρθραλγιών. Σε περίπτωση σοβαρής εξέλιξης της νόσου τα συμπτώματα περιλαμβάνουν δύσπνοια, ανορεξία και υψηλό πυρετό και είναι δυνατόν να προέλθει θάνατος σε περίπτωση που ο ασθενής είναι μεγάλης ηλικίας ή αντιμετωπίζει καρδιακά προβλήματα, διαβήτη, χρόνια αναπνευστική νόσο ή καρκίνο. Μεγάλο πρόβλημα αποτελούν οι συνεχόμενες μεταλλάξεις του ιού που διαιωνίζουν τη διασπορά του στην κοινότητα [World Health Organization].

2 Κεφάλαιο 1 Εισαγωγή

Η πανδημία του νέου κορωνοϊού αποτέλεσε μια από τις μεγαλύτερες παγκόσμιες προκλήσεις που αντιμετώπισαν οι σύγχρονες κοινωνίες τα τελευταία χρόνια και αναμφίβολα επηρέασε όλες τις πτυχές της ανθρώπινης δραστηριότητας. Η παγκόσμια υγειονομική κρίση επέφερε σοβαρές επιπτώσεις στο δημόσιο σύστημα υγείας της χώρας με αυξημένες οικονομικές δαπάνες για την θωράκιση του και την εξάντληση του υγειονομικού προσωπικού ενώ υπήρξαν δυστυχώς σημαντικές απώλειες ανθρώπινων ζωών. Η εφαρμογή της καραντίνας για ύποπτα μολυσμένα άτομα, η κοινωνική αποστασιοποίηση με την αποφυγή κοινωνικών επαφών και συναθροίσεων και το κλείσιμο των σχολείων, καταστημάτων, εστιατορίων και κέντρων διασκέδασης την περίοδο των lockdown εκτός από το αρνητικό αντίκτυπο στην οικονομία της χώρας έπληξαν και την φυχική υγεία των πολιτών με την εκδήλωση προβλημάτων άγχους, φόβου και κατάθλιψης δεδομένης και της παρατεταμένης ανασφάλειας για το μέλλον.

Στην παρούσα εργασία, εστιάζουμε στην γεωγραφική περιοχή της Ελλάδας χρησιμοποιώντας παράλληλα υγειονομικά δεδομένα που αφορούν τον ίο (όπως τον αριθμό των κρουσμάτων, των θανάτων, των εμβολιασμών και των ασθενών που νοσηλεύονται σε Μονάδες Εντατικής Θεραπείας), και δείκτες ανθρώπινης κινητικότητας της Google και της Apple, για την δημιουργία μοντέλων μηχανικής μάθησης που θα έχουν την δυνατότητα να προβλέπουν την μεταβολή της ανθρώπινης κινητικότητας στο μέλλον ή κάποιας μετρικής που καταγράφει την πορεία του ιού στην χώρα μας (π.χ., μελλοντικό αριθμό κρούσματων).

2. Βιβλιογραφική έρευνα - Σχετικές εργασίες

Σε αυτό το κεφάλαιο γίνεται μια παρουσίαση των θεμελιωδών γνωρισμάτων και των συμπερασμάτων των επιστημονικών άρθρων που μελετήθηκαν κατά την διάρκεια της βιβλιογραφικής έρευνας.

2.1 Ανθρώπινη κινητικότητα και εξάπλωση του ιού

Στην εποχή των Μεγάλων Δεδομένων (Big data), ο μεγάλος ρυθμός συλλογής και δημιουργίας δεδομένων επιβάλλει την ανάγκη περαιτέρω μελέτης της μορφής, του περιεχόμενου και του τρόπου αξιοποίησης τους. Τα δεδομένα που αφορούν την πανδημία διαφέρουν τόσο στην μορφή τους όσο και την προέλευση τους (από τοποθεσίες GPS, εφαρμογές κοινωνικής δικτύωσης και υγείας που είναι εγκατεστημένες σε κινητά τηλέφωνα και wearables, κάμερες και αισθητήρες IoT, εθνικά συστήματα υγείας, δημόσιες συγκοινωνίες – smartcards). Δεδομένα που περιγράφουν την κίνηση των ανθρώπων από μια τοποθεσία σε μια άλλη καθώς και τον τρόπο μετακίνησης (πόδια, ποδήλατο, αυτοκίνητο, τρένο) ονομάζονται mobility data. Ως σημεία ενδιαφέροντος (Points of Interest; PoI) ορίζονται τοποθεσίες (εκτός κατοικιών) στις οποίες οι ανθρώποι επισκέπτονται όπως εστιατόρια, ξενοδοχεία, καφετέριες και θρησκευτικοί χώροι λατρείας. Τα δεδομένα κίνησης είναι ζωτικής σημασίας καθώς με την μελέτη τους μπορεί να ερευνηθεί αν οι ανθρώποι τηρούν τα περιοριστικά μέτρα που έχουν επιβάλλει οι κυβερνήσεις, να εντοπιστούν οι γεωγραφικές περιοχές υψηλού κινδύνου και ο τρόπος εξάπλωσης του ιού συνδυαστικά με την γεωγραφική ανίχνευση της επιδημιολογικής του πορείας (geographical symptom tracking) και να συμβάλλουν στην αποδοτική λήψη αποφάσεων από κυβερνητικά στελέχη για την αντιμετώπιση της πανδημίας και την χάραξη βιώσιμου και οικολογικού (χωρίς περιβαλλοντικούς ρύπους) πλάνου μετακίνησης. Η σημασία των mobility data

4 Κεφάλαιο 2 Βιβλιογραφική έρευνα - Σχετικές εργασίες

αποτυπώνεται στο γεγονός ότι τόσο εταιρείες του δημοσίου όσο και του ιδιωτικού τομέα συλλέγουν δεδομένα. Εταιρίες του δημοσίου ενδέχεται, είτε να μισθώσουν ιδιωτικές εταιρίες για την συλλογή δεδομένων για τον λογαριασμό τους, είτε να προβούν στη συλλογή τους με δικά τους μέσα και έπειτα να προχωρήσουν σε ανταλλαγή δεδομένων με τον ιδιωτικό τομέα με κατάλληλη διαπραγμάτευση για την προσβασιμότητα που θα έχει κάθε φερόμενο μέλος σε αυτά [Kotsch 2020].

Η αναλυτική των συγκεντρωτικών δεδομένων που δόθηκαν από τις εταιρίες τηλεπικοινωνιών σύμφωνα με τους αυστηρούς κανόνες και όρια που έχει θεσπίσει το European Data Protection Board, μπορεί να εξηγήσει την δυναμική του ιού και να συμβάλλει στην πρόβλεψη του χωρίς να χρειαστεί να γίνει αποκαλύψη προσωπικών δεδομένων χρηστών αλλά ούτε και χρήση κάποιας εφαρμογής ιχνηλάτευσης επαφών (contact tracing). Χρησιμοποιώντας ανώνυμα δεδομένα κινητής τηλεφωνίας από τις χώρες της Ιταλίας, Ισπανίας και Γαλλίας και αξιοποιώντας τις τεχνικές μητρώων αφετηρίας-προορισμού (origin-destination matrices) και θεωρίας γράφων η έρευνα των [Iacus et al. 2020] κατέληξε στο συμπέρασμα ότι η ανθρώπινη κινητικότητα μπορεί να αιτιολογήσει την εξάπλωση του ιού σε ποσοστό 92 % ενώ έχει μικρότερο ποσοστό επιτυχίας μετά την άρση των περιοριστικών μέτρων.

Επιπρόσθετα, η εργασία των [Christidis et al. 2022] καταλήγει στο συμπέρασμα ότι η κινητικότητα των πολιτών είναι συσχετιζόμενη με την εξέλιξη της πανδημίας καθώς η αύξηση της κινητικότητας μπορεί να ερμηνευτεί ως αύξηση κοινωνικής επαφής με άλλους ανθρώπους και ενδεχόμενη χαλάρωση των περιοριστικών μέτρων μπορεί να οδηγήσει σε ένα νέο κύμα πανδημίας. Εντοπίστηκε μείωση της κινητικότητας τα Σαββατοκύριακα και τις αργίες στις αστικές περιοχές γεγονός που εξηγείται εν μέρει από το σημαντικό μερίδιο του πληθυσμού που εργάζεται ή μετέχει στο εκπαιδευτικό σύστημα της εκάστοτε χώρας και είναι υπεύθυνο για την αυξημένη κινητικότητα τις καθημερινές. Άλλα ιδιαίτερα γνωρίσματα της έρευνας είναι πως ο βαθμός αστικοποίησης μιας πόλης και το προφίλ της οικονομικής δραστηριότητας των πολιτών αποτελούν παράγοντες που επηρεάζουν την κινητικότητα τους και ότι ο αριθμός των θανάτων από COVID-19 σε μια συγκεκριμένη περιοχή βασίζεται πιο πολύ στα συστήματα υγείας και στα δημογραφικά χαρακτηριστικά της περιοχής πάρα στην δυναμική μετάδοση του ιού.

Αρκετά άρθρα υποδηλώνουν μια θετική σχέση ανάμεσα στα χρούσματα κορονοϊού και στην ανθρώπινη κινητικότητα, σχέση που μπορεί να είναι αρκετά χρήσιμη για την κατανόηση της κινητικότητας σε επίπεδο πληθυσμού, τον προσδιορισμό περιοχών με υψηλό κινδύνο μόλυνσης από τον ιό και την πρόβλεψη επόμενων κυμάτων επιδημίας. Στόχος της εργασίας των [Strzelecki 2022]

είναι ο προσδιορισμός των επιπτώσεων της κινητικότητας στην εξάπλωση του COVID-19. Τα 35 άρθρα που επιλέχθηκαν εστιάζουν στα θέματα της μετάδοσης του ιού, των μέσων συγκοινωνιών, της ποιότητας του ατμοσφαιρικού αέρα και της επανέναρξης χρηματοοικονομικών δραστηριοτήτων. Βασικό γνώρισμα της εργασίας είναι ότι η χρήση δεδομένων κινητικότητας μόνο από μία χώρα κατά τη διάρκεια της πανδημίας διευκολύνει την κατανόηση της πορείας της σε διαφορετικές περιοχές. Δεδομένα από διαφορετικές χώρες μπορεί να είναι ανόμοια στην φύση τους και στο περιεχόμενο τους, γεγονός που καθιστά δύσκολο να πραγματοποιηθούν συγκρίσεις και να ερμηνευθούν τα αποτελέσματα αυτών.

Στην εργασία των [Alamo et al. 2021] εφαρμόζεται η 3M ανάλυση σε σχέση με την πανδημία η οποία βασίζεται σε τρεις πυλώνες : επιτήρηση (monitoring) της εξέλιξης της και αξιολόγηση των περιοριστικών μέτρων που εφαρμόστηκαν, μοντελοποίηση και πρόβλεψη της διασποράς (modelling) με την χρήση μοντέλων που αξιοποιούν γνώσεις ιολογίας και λήψη έγκαιρων αποφάσεων για τον έλεγχο της πανδημίας (managing). Τα δεδομένα κινητικότητας είναι ιδιαίτερα χρήσιμα καθώς μπορούν να βοηθήσουν στην πρόβλεψη της χωρικής μετάδοσης και την αναγνώριση περιοχών με υψηλή επικινδυνότητα και να συνεισφέρουν στην θέσπιση κατάλληλων μέτρων περιορισμού ενάντια στην πανδημία. Επιπρόσθετα η χρήση της θεωρίας ελέγχου (control theory) είναι απαραίτητη για την λήψη αποφάσεων για την επίλυση δύσκολων και απαιτητικών προβλημάτων που αφορούν την πανδημία και που χαρακτηρίζονται από αβεβαιότητα, από απαιτήσεις σε χρόνο και υλικοτεχνική υποδομή, όπως π.χ., η βέλτιστη κατανομή περιορισμένων τεστ και εμβολίων και η εύρεση της κατάλληλης χρονικής στιγμής για την επιβολή ή την άρση των περιοριστικών μέτρων.

Η πανδημία φαίνεται να παρουσιάζει επαναλαμβανόμενα κύματα μετάδοσης τα οποία είναι πολύ σημαντικό να μπορούν να ανιχνευθούν, να ερμηνευθούν και να προβλεφθούν από συστήματα επιτήρησης τα οποία χρησιμοποιούν data driven τεχνικές. Η πρόβλεψη νέων κυμάτων πανδημίας είναι ιδιαίτερα δύσκολη, καθώς δεν υπάρχει διαθεσιμότητα ιστορικών δεδομένων εποχικότητας και ο προσδιορισμός του τρέχοντος ποσοστού του μολυσμένου πληθυσμού από τον ιό είναι απαιτητικός στην περίπτωση που υπάρχουν πολλά ασυμπτωματικά κρούσματα. Κάποιες από τις σημαντικές παρατηρήσεις της εργασίας των [Alamo et al. 2021] που αξίζουν να αναφερθούν είναι πως η υποδομή και η ετοιμότητα του συστήματος υγείας αποτελούν σημαντικούς παράγοντες της επίδρασης που θα έχει η πανδημία στον πληθυσμό και ότι η συνεισφορά των ασυμπτωματικών κρουσμάτων στον δείκτη μεταδοτικότητας του ιού είναι μεγαλύτερη από αυτή των συμπτωματικών.

6 Κεφάλαιο 2 Βιβλιογραφική έρευνα - Σχετικές εργασίες

Για την αποτελεσματικότερη διαχείριση της πανδημίας σημαντικό εργαλείο είναι η ανάπτυξη ενός μεταπληθυσμιακού μοντέλου μορφής επιρρεπή-εκτειμέμενου-μεταδοτικού-αφαιρούμενου (metapopulation susceptible exposed infectious removed - SEIR model) το οποίο ενσωματώνει δίκτυα κινητικότητας (mobility networks) [Chang et al. 2021] και χρησιμοποιήθηκε για την προσομοίωση της εξάπλωσης του ιού στις 10 μεγαλύτερες πόλεις των ΗΠΑ κάνοντας χρήση των συνόλων δεδομένων των Places Patterns, Weekly Patterns και Social Distancing Metrics από το Safegraph, το COVID-19 των New York Times και το US census. Τα δίκτυα κινητικότητας είναι δίκτυα βασισμένα σε δεδομένα κινητών τηλεφώνων και αναπαριστούν τις ωριαίες μετακίνησης από τις γειτονίες των πολιτών (census blocks) σε σημεία ενδιαφέροντος. Όπως αναφέρεται στο [Ponce-de Leon et al. 2021] η ανακατασκευή των δικτύων κινητικότητας βασίζεται σε δεδομένα εγγραφών γεγονότων, που περιλαμβάνουν ενεργά δεδομένα κινητής τηλεφωνίας (CDR Call Retail Records) και παρέχουν μια εγγραφή κάθε φορά που μια συσκευή επικοινωνεί στο δίκτυο (κλήσεις, αποστολή μηνυμάτων, σύνοδος δεδομένων) και σε παθητικά γεγονότα (περιοδική ενημέρωση της τοποθεσίας της συσκευής, αλλαγές στην κάλυψη των δεδομένων).

Ένα SEIR μοντέλο διαιρείται στις ομάδες των ατόμων που είναι επιρρεπείς στον ίο (Susceptible), των ατόμων που έχουν εκτεθεί σε αυτόν αλλά δεν μπορούν να τον μεταδώσουν ακόμα (Exposed), των μολυσμένων - μεταδοτικών (Infected) και αυτών που έχουν αναφράσει (Recovered). Το μοντέλο μπορεί να αξιολογήσει σε μεγάλο βαθμό τις επιδράσεις της άρσης των περιοριστικών μέτρων που είχαν επιβληθεί σε σημεία ενδιαφέροντος και να εκτιμήσει τον ωριαίο ρυθμό μόλυνσης του κάθε POI με βάση την περιοχή, την μέση παραμονή και το μεταβαλλόμενο ως προς το χρόνο μέγεθος των πολιτών που το επισκέπτονται. Σύμφωνα με τα αποτελέσματα το μοντέλο προβλέπει ότι μια μικρή μειονότητα υπερμεταδοτικών POIs είναι υπεύθυνη για την εξάπλωση του ιού και η εφαρμογή περιορισμών στην χωρητικότητα τους θα αποτελούσε καλύτερη στρατηγική από την ομοιόμορφη μείωση της κινητικότητας. Επίσης καταλήγει ότι υψηλότερο κίνδυνο μόλυνσης από τον ίο έχουν όσοι ανήκουν σε ασθενείς κοινωνικά και οικονομικά στρώματα επειδή ενδεχομένως επισκέπτονται πιο πολυσύχναστα POIs, μένουν πιο πολύ ώρα σε αυτά και δεν μπορούν να μεταβάλλουν την τρόπο κινητικότητας με πιο ασφαλής τρόπους (π.χ., χρήση αυτοκινήτου από ότι μετρό).

Στις μελέτες [Georgiou 2020a] και [Georgiou 2020b] ερευνάται η εξάπλωση του ιού στην Ελλάδα με την χρήση του επιδημιολογικού μοντέλου SEIQRDP και της τεχνικής Ελαχίστων Τετραγώνων (Least Squares). Η επιπλέον μεταβλητή R στο μοντέλο αναφέρεται σε ένα μέρος του πληθυσμού που όταν έρχεται

σε επαφή με τον ιό δεν μολύνεται από αυτόν όποτε δεν εισάγεται και στην μεταβλητή E . Φαίνομενα αργής αναφοράς του αριθμού των ανθρώπων που ανάρρωσαν από τον ιό R_t από τον ΕΟΔΥ και under-reporting των κρουσμάτων εντοπίστηκαν στην μελέτη. Οι καθυστερήσεις αναφοράς μπορεί να είναι σε χρονικό διάστημα έως και μιας εβδομάδας εισάγοντας μια υποβάθμιση της αντίστοιχης χρονοσείρας ενώ ο επιβεβαιωμένος αριθμός των ατόμων που μολύνθηκαν από τον ιό I_t εικάζεται ότι ήταν σε αναλογία 20:1 (δηλαδή από τα 20 κρούσματα το ένα δηλώνεται και θεωρείται επιβεβαιωμένο), γεγονός που δυσχεραίνει τον έλεγχο της πανδημίας καθώς όσο υψηλότερος είναι ο αριθμός under-reporting των κρουσμάτων, τόσο πιο αργά εκτελείται η διαδικασία αποσυμφόρησης, δηλαδή η μετάβαση από την κορυφή της καμπύλης των κρουσμάτων σε χαμηλότερα επίπεδα. Ακόμη επιβεβαιώθηκε ότι ο αριθμός των θανάτων ακολουθεί την λογιστική (σιγμοειδής) συνάρτηση και ότι η ανάλυση των κρουσμάτων πρέπει να γίνεται σε ένα μικρό χρονικό διάστημα 4 έως 6 ημερών το οποίο θεωρείται περίοδος επώασης (5.1 – 5.2 ημέρες) και σε ένα μεγάλο χρονικό διάστημα 20 ημερών, το οποίο θεωρείται κατάλληλο χρονικό διάστημα καραντίνας την περίοδο που δεν είχε ανακαλυφθεί το εμβόλιο καθώς μέσα σε αυτό έχει μεσολαβήσει η εμφάνιση των συμπτωμάτων και η ανάρρωση από τον ιό.

Αξιοσημείωτα συμπεράσματα σχετικά με την χρήση τεχνικών Machine Learning στην διαχείριση της διασποράς του ιού παρουσιάζει η δουλεία των [Dandekar and Barbastathis 2020]. Αναλυτικότερα, αξιοποιώντας δημόσια δεδομένα από την περιοχή Wuhan της Κίνας και από τις χώρες της Ιταλίας, Νότιας Κορέας και ΗΠΑ με την χρήση ενός νευρωνικού δικτύου και εξισώσεων επιδημιολογίας, γίνεται προσπάθεια σύγκρισης της επίδρασης που είχαν η καραντίνα και τα διάφορα περιοριστικά μέτρα στην διαχείριση της πανδημίας σε αυτές τις χώρες. Βασικά πορίσματα είναι η αρνητική συσχέτιση ανάμεσα στο ρυθμό αναπαραγωγής ($\text{reproduction number-}R_t$) και στην ισχύ της καραντίνας ($\text{quarantine strength-}Q_t$) και το συμπέρασμα ότι ότι χώρες που έλαβαν πιο έγκαιρα ή πιο αυστηρά υγειονομικά πρωτόκολλα και μέτρα απομόνωσης περιόρισαν με μεγαλύτερη επιτυχία την διασπορά του ιού. Ένα SIR μοντέλο δεν μπορεί να καταγράψει την συμπεριφορά του πληθυσμού στα διάφορα περιοριστικά μέτρα και η χρήση ενός νευρωνικού δικτύου σαν μια μη γραμμική συνάρτηση προσέγγισης (approximator) μπορεί να συμβάλλει σημαντικά στον καθορισμό της τιμής της μεταβλητής των μολυσμένων από τον ιό ατόμων I_t . Το νευρωνικό δίκτυο εκπαιδεύεται με δημόσια δεδομένα μόλυνσης από τον ιό και με πληθυσμιακά δεδομένα ανά περιοχή.

Στην μελέτη των [Arvanitis et al. 2021] πραγματοποιήθηκε πρόβλεψη βραχέος χρόνου του αριθμού μεταδοτικότητας R_t του ιού σε περιοχές της

8 Κεφάλαιο 2 Βιβλιογραφική έρευνα - Σχετικές εργασίες

Ελλάδας με βάση τα κρούσματα που παρέχονται από τον ΕΟΔΥ και την κινητικότητα κατά την διάρκεια της πανδημίας από το σύνολο δεδομένων της Google (COVID-19 Community Mobility Reports) [Google LLC] με την χρήση μοντέλων μηχανικής μάθησης. Η εκπαίδευση των μοντέλων πρόβλεψης ξεκινά από τα μέσα Ιουνίου 2020 έως τον Ιανουάριο του 2021, προϋποθέτει ότι η εξάρτηση του αριθμού αναπαραγωγής του ιού από την κινητικότητα δεν αλλάζει κατά τη διάρκεια αυτής της περιόδου και χρησιμοποιεί τα δεδομένα κινητικότητας της Google της προηγούμενης εβδομάδας για την εκτίμηση της τιμής R_t . Η χρήση μοντέλου τυχαίου δάσους (Random forests) απέδωσε καλύτερα αποτελέσματα σε σχέση με τα άλλα μοντέλα μηχανικής μάθησης (ARIMA, KNN, SVM, Decision Trees και τα νευρωνικά δίκτυα) στις μετρικές R^2 , Pearson, Spearman, Mean Squared Error και Mean Absolute Error. Η τιμή της μετρικής R^2 ήταν 0.5 για τα SVM και ANN και 0.75 για τα RF. Αυτό σημαίνει ότι περίπου το 75% της διακύμανσης των παρατηρούμενων δεδομένων R_t μπορεί να προβλεφθεί από το μοντέλο RF, υποδηλώνοντας ότι οι αλλαγές της προηγούμενης εβδομάδας στα δεδομένα κινητικότητας μπορούν να εξηγήσουν το 75% των ημερήσιων διακυμάνσεων του αριθμού αναπαραγωγής. Το υπόλοιπο 25% της διακύμανσης που δεν μπορεί να συλλάβει η RF ενδεχομένως να οφείλεται στα περιοριστικά μέτρα που έλαβε η κυβέρνηση, στην ανοσία της αγέλης και στις νέες παραλλαγές του ιού.

Επειτα από την πραγματοποιήση ανάλυσης σημασίας των χαρακτηριστικών προκύπτει ότι οι κατηγορίες των χώρων εργασίας (workplaces) και των πάρκων (parks) θεωρούνται σημαντικότεροι σε σχέση με τις άλλες, με την κατηγορία των πάρκων να είναι υψηλά συσχετισμένη με τις κατηγορίες της λιανικής - ψυχαγωγίας (retail & recreation) και των δημόσιων συγκοινωνιών (Transit stations). Η διαμονή στο σπίτι είναι ο πιο αποτελεσματικός τρόπος για την μείωση των κρουσμάτων με αρνητικούς συντελεστές αλλά μεγάλους σε όρους απόλυτης τιμής ενώ ο χρόνος που περνούν οι πολίτες στα πάρκα έχει ελάχιστη αινητική επίδραση στον αριθμό των κρουσμάτων σε σχέση με τις άλλες ανθρώπινες δραστηριότητες. Επίσης η κινητικότητα που αφορά τα καταστήματα τροφίμων και τα φαρμακεία δεν επηρεάζει την μετάδοση του ιού. Τέλος ο τομέας κινητικότητας που συμβάλλει λιγότερο στην R_t , η λιανική και ψυχαγωγία, έχει βρεθεί συχνά στο επίκεντρο των μέτρων ελέγχου, περιορίζοντας μερικές φορές τον αριθμό των επισκεπτών μέσα στα καταστήματα και συχνά κλείνοντας τον εντελώς, στρατηγική που φαίνεται ότι δεν αποδίδει μείωση των κρουσμάτων [Noland 2021].

Στο άρθρο των [Nouvellet et al. 2021] γίνεται παρουσίαση ενός framework που αναπτύχθηκε για να μελετήσει την σχέση ανάμεσα στον ρυθμό μετάδοσης του ιού σε πληθυσμιακό επίπεδο με την ανθρώπινη κινητικότητα. Το frame-

work περιλαμβάνει 52 χώρες μαζί με τα υγειονομικά δεδομένα των χωρών αυτών (π.χ., κρούσματα, θάνατοι) και την κινητικότητα των πολιτών τους που παρέχεται από τις πηγές δεδομένων της Google και της Apple (το συνδυασμένο dataset της Google και της Apple απέδωσε καλύτερα αποτελέσματα σε σχέση με τα άλλα streams κινητικότητας). Το framework συνέβαλλε στην εκτίμηση κατωφλίων τα οποία είναι μεταβλητά ως προς τον χώρο και χρόνο σχετικά με την κινητικότητα ανά χώρα με απότερο στόχο τον έλεγχο της μετάδοσης. Βασικό γνώρισμα της μελέτης είναι ότι αν η μείωση της κινητικότητας φτάσει ένα συγκεκριμένο επίπεδο, τότε τα κρούσματα από τον ίδιο θα μειωθούν με σταθερούς τους άλλους παράγοντες που επηρεάζουν την μεταδοτικότητα του. Η ετερογένεια στα εκτιμώμενα κατώφλια κινητικότητας ανάμεσα στις χώρες πιθανά αντανακλά τις κοινωνικοπολιτικές διαφορές ή/και τις διαφορές στα περιοριστικά μέτρα που έχει εφαρμόσει η εκάστοτε χώρα.

2.2 Βάσεις δεδομένων COVID-19 και εργαλεία οπτικοποιήσης

Στην εργασία των [Wahltez et al. 2022] ερευνάται το COVID-19 Open Dataset (COD), ένα μεγάλο “meta – dataset” σχετικά με τον ίδιο το οποίο περιλαμβάνει επιδημιολογικά δεδομένα από 22.579 μοναδικές τοποθεσίες σε 232 διαφορετικές χώρες και ανεξάρτητες περιοχές. Το dataset αποτελείται από τρείς τύπους δεδομένων: χρονοσειρές για βιολογικά δεδομένα σε διάφορες τοποθεσίες και χώρες, χρονοσειρές για πιθανά αποτελέσματα πρόβλεψης κινητικότητας και κυβερνητικών παρεμβάσεων, και στατικά δεδομένα που αφορούν τα ιδιαίτερα χαρακτηριστικά κάθε τοποθεσίας. Οι πίνακες του dataset περιλαμβάνουν δεδομένα επιστημονικής (ανά ηλικία και φύλο), υγειονομικής (εμβολιασμοί, νοσηλείες), οικονομικής, γεωγραφικής, μετεωρολογικής και αστικής φύσης. Επίσης όλα τα δεδομένα του dataset είναι χωρικά αρχειοθετημένα: αυτό σημαίνει ότι ορίζεται ένας τρόπος ιεραρχίας για την μοναδική αναγνώριση κάθε τοποθεσίας σε επίπεδο χωρών (level 0), επαρχίας ή πολιτείας (level 1), δήμου ή κομητείας (level 2) και κοινότητας (level 3). Κάθε τοποθεσία επισημαίνεται με ένα μοναδικό αναγνωριστικό έτσι ώστε διαφορετικοί τύποι πληροφοριών να μπορούν εύκολα να συνδυαστούν. Επιδημιολογικά δεδομένα μπορούν να συνδυαστούν με δημογραφικά και κινητικότητας δεδομένα, λαμβάνοντας υπόψη τις κυβερνητικές αποφάσεις ώστε να κατανοηθεί ο τρόπος εξάπλωσης του ιού.

Αξιοσημείωτη τεχνική που χρησιμοποιείται είναι η αποθήκευση τόσο των ημερησίων (daily) όσο και των αθροιστικών (cumulative) δεδομένων. Αυτό συμβαίνει καθώς τα αθροιστικά δεδομένα δεν είναι κατά ανάγκη πάντα ίσα

10 Κεφάλαιο 2 Βιβλιογραφική έρευνα - Σχετικές εργασίες

με το άθροισμα των ημερησίων δεδομένων αφού οι αρχές που δημοσιοποιούν τα δεδομένα ενδέχεται να πραγματοποιήσουν αλλαγές στα κριτήρια και στον τρόπο υπολογισμού αυτών χωρίς να τα ενημερώσουν. Επιπλέον μερικές φορές τα ημερήσια δεδομένα μπορεί να είναι αρνητικά εξαιτίας κάποιας συσχέτισης ή μεταβολής στον τρόπο μέτρησης. Για αυτό το λόγο προτιμάται η χρήση πηγών αθροιστικής φύσης δεδομένων παρά ημερήσιας μορφής καθώς όταν μια αρχή δημοσιοποιεί ιστορικά επιδημιολογικά δεδομένα υπάρχει η δυνατότητα διόρθωσης αυτών σε περίπτωση σφάλματος ή αλλαγής μεθοδολογίας.

Μια άλλη αξιόπιστη βάση είναι η “The demography of COVID-19 Deaths” [Caporali et al. 2022] η οποία ειδικεύεται στην αποθήκευση του αριθμού των θανάτων από COIVD-19. Παρέχει τον ημερήσιο και τον αθροιστικό αριθμό θανάτων από την νόσο ανάλογα το φύλο, την ηλικία, την ημερομηνία και την τοποθεσία. Η βάση επικεντρώνεται και στην ανάπτυξη εργαλείων για την αξιολόγηση της πληρότητας των δεδομένων κατά την διάρκεια του χρόνου στις διάφορες χώρες/περιοχές διευκολύνοντας την trend analysis. Τα δεδομένα που συλλέγονται υφίστανται μια σειρά από ελέγχους πριν και μετά τη δημοσίευση τους και προκειμένου να διασφαλιστεί ότι τόσο τα δεδομένα όσο και η σχετική τεκμηρίωση είναι υψηλής ποιότητας, επιστρατεύεται και ένας εμπειρογνώμονας της χώρας για την αξιολόγηση τους.

Στην μελέτη των [Ponce-de Leon et al. 2021] γίνεται παρουσίαση των COVID-19 Flow Maps, ενός GIS συστήματος που ενσωματώνει εγγραφές που αφορούν τις χρονοσειρές κινητικότητας του πληθυσμού και τα καθημερινά κρούσματα της Ισπανίας σε διαφορετική χωρική και χρονική κλίμακα ανάλυσης. Παρέχει εργαλεία για εξαγωγή, διαχείριση, ανάλυση και οπτικοποίηση των δεδομένων. Τα δεδομένα κινητικότητας παρέχονται ως μήτρες προέλευσης προορισμού σε διαφορετικό βαθμό χωρικής και χρονικής ανάλυσης (περιφέρεια, επαρχία, αυτόνομη περιοχή). Τα χαρακτηριστικά γεωαναφοράς αναπαρίστανται ως γεωγραφικές συντεταγμένες ή δείκτες γεωμετρίας που προσδιορίζουν ένα γεωγραφικό επίπεδο. Όπως αναφέρεται ο συνδυασμός των δεδομένων των κρουσμάτων και της ανθρώπινης κινητικότητας μπορεί να πραγματοποιηθεί όταν όλες οι εγγραφές προβληθούν στο ίδιο γεωγραφικό επίπεδο και για την αντιμετώπιση της επικάλυψης των γεωμετρικών σχημάτων χρησιμοποιείται γραμμική παρεμβολή στις επικαλυπτόμενες περιοχές.

Σημαντικό εργαλείο οπτικοποιήσης του ιού είναι το data dashboard. Παρέχει έναν δια δραστικό και συγκεντρωτικό τρόπο για την προβολή, ανάλυση και εξαγωγή χρήσιμων συμπερασμάτων από δεδομένα που προέρχονται από διαφορετικά dataset με έναν ταχύ και «εύπεπτο» γραφικό τρόπο. Έγιναν ιδιαίτερα δημοφιλή το 2020 όπου data scientists, κυβερνητικά στελέχη και δημοσιογράφοι τα χρησιμοποιήσαν για την αναπαράσταση των κρουσμάτων,

θανάτων, εμβολιασμών και νοσηλειών από τον COVID-19. Τα dashboards θα μπορούσαν να χρησιμοποιηθούν στο μέλλον για έρευνες που σχετίζονται με την φτώχεια, την κλιματική αλλαγή και την αστυνομική βία εις βάρος φυλετικών ομάδων αλλα πρέπει να αναφερθεί ότι η πλήρης κατανόηση των δεδομένων απαιτεί περαιτέρω ενασχόληση με πολυπλοκότερα εργαλεία και μέθοδούς [Peeples et al. 2022]. Τέλος έχουν αναπτυχθεί ερευνητικά και κατάλληλες βιβλιοθήκες οπτικοποίησης όπως της VASA στο άρθρο των [Noi et al. 2021] η οποία είναι σε γλώσσα Python και βασίζεται στις τεχνικές των stacked recency and consistency map, line-path scatter plot και categorical strip (dot) plot.

2.3 Εφαρμογές Contact Tracing

Η ραγδαία εξάπλωση του ιού ιδιαίτερα όταν τα περιοριστικά μέτρα αίρονται απαιτεί αποδοτικότερη διαχείριση και αυτό μπορεί να επιτευχθεί με την χρήση των εφαρμογών ιχνηλάτισης επαφών (contact tracing). Η έγκαιρη ιχνηλάτευση θετικών κρουσμάτων (ακόμα και ασυμπτωματικών) μπορεί να μειώσει τον δείκτη μόλυνσης R_0 κάτω από 1, ανακόπτοντας την επιδημιολογική πορεία του ιού. Η πρώτη εφαρμογή contact tracing υλοποιήθηκε από την Σιγκαπούρη με όνομα TraceTogether ενώ και άλλες χώρες της Ασίας (π.χ., Κίνα, Ταϊβάν) και της Ευρώπης (π.χ., Ηνωμένο Βασίλειο, Δανία) ακολουθήσαν την υιοθέτηση τέτοιων εφαρμογών, δεδομένου ότι η υλοποίηση τους απαιτούσε μόνο την ανάπτυξη του λογισμικού καθώς όλο το απαιτούμενο hardware παρέχονταν από τις συσκευές. Μια εφαρμογή contact tracing κάνει χρήση της τεχνολογίας Bluetooth και καταγράφει άλλες συσκευές χρηστών που είναι σε κοντινή απόσταση με αυτήν καθώς και την διάρκεια επαφής, την ημερομηνία και την χρονική στιγμή που συνέβη το γεγονός. Αναλυτικότερα η εφαρμογή αναγνωρίζει κοντινές συσκευές με την αποστολή tokens – μεταβαλλόμενων strings που έχουν χαμηλή διάρκεια ισχύος προτού ανανεωθούν εξασφαλίζοντας ένα σημαντικό επίπεδο ασφάλειας. Η εκπομπή και η λήψη σημάτων Bluetooth πρέπει να γίνεται σε τακτά χρονικά διαστήματα και να μην είναι συνεχής.

Σε περίπτωση που ένας χρήστης διαγνωστεί θετικός στον ιό η εφαρμογή επιτρέπει στο Υπουργείο Υγείας να αποκτήσει πρόσβαση στα δεδομένα του ασθενή, τα οποία περιλαμβάνουν μια λίστα με tokens άλλων συσκευών που έχει έρθει σε επαφή. Το υπουργείο καθώς διαθέτει την πλήρη λίστα αντιστοιχισμάτων tokens με τα κινητά τηλέφωνα, ειδοποιεί μέσω ενός εκπροσώπου όλους τους χρήστες που συνάντησαν το μολύσμενο άτομο το προηγούμενο χρονικό διάστημα. Η εφαρμογή αποθηκεύει τα δεδομένα που παράγονται μόνο στο χρονικό περιθώριο των είκοσι μια ημερών και αυτά δεν πρόκειται

12 Κεφάλαιο 2 Βιβλιογραφική έρευνα - Σχετικές εργασίες

να προσπελαστούν παρά μόνο στην περίπτωση που ο χρήστης νοισήσει από τον κορονοϊό. Το γεγονός ότι ο ασθενής οφείλει να διαμοιράσει τα δεδομένα του στην κυβέρνηση για να επικοινωνήσει με τα άτομα στα οποια ήρθε σε επαφή, έχει οδηγήσει την επιστημονική κοινότητα να αναζητήσει τροποποιήσεις και τεχνικές που να διατηρούν την αποτελεσματικότητα που παρουσιάζουν οι εφαρμογές contact tracing αλλά και να προσφέρουν μεγαλύτερη ασφάλεια και προστασία προσωπικού απορρήτου από κακόβολους χρήστες. Σημαντικά βήματα για την περαιτέρω διασφάλιση του απορρήτου θα ήταν η εφαρμογή να μετατραπεί σε μορφή ανοικτού κώδικα και μη υποχρεωτική για γενικούς σκοπούς, με ανάγκη χρήσης μόνο για είσοδο σε συγκεκριμένους χώρους [Cho et al. 2020].

Τέτοιες εφαρμογές contact tracing έχουν επικρατήσει και στην Ευρωπαϊκή ήπειρο με κυριότερο πρότυπο το Pan-European Privacy Preserving Proximity Tracing (PEPP-PT). Το PEPP-PT εφαρμόζει αυστηρότερα πρωτοκόλλα ασφάλειας σύμφωνα με τους Ευρωπαϊκούς νόμους προστασίας δεδομένων σε σχέση με αυτά των ασιατικών χωρών και κάνει χρήση ενός αποκεντρωμένου μοντέλου (decentralized) για την αποθήκευση των δεδομένων [Busvine 2020]. Άλλα χαρακτηριστικά μιας Ευρωπαϊκής εφαρμογής είναι η μη παρεμβατική παρακολούθηση των δεδομένων τοποθεσίας και η δυνατότητα πρόσβασης των δεδομένων μόνο από τις τοπικές υγειονομικές αρχές για την ειδοποίηση και την παρόχη συστάσεων σε ατόμα που ήρθαν σε επαφή με επιβεβαιωμένο κρούσμα. Ο τρόπος επικοινωνίας αυτών που συναντήθηκαν με επιβεβαιωμένο κρούσμα παραμένει υποχρέωση και επιλογή της κάθε κυβέρνησης που αναπτύσσει την εφαρμογή της και θα πρέπει να βρίσκονται στα πλαίσια της νοοτροπίας και των τεχνικών ανίχνευσης εγγύτητας του PEPP-PT [Cooper et al. 2020].

Παρόμοια γνωρίσματα αναφέρει και το επιστημονικό άρθρο των [Nanni et al. 2021] που εξετάζει την DP3T (Decentralized Privacy-Preserving Proximity Tracing) προσέγγιση, όπου η κεντρική αρχή απλά εκπέμπει τα ανώνυμα ID των θετικών περιστατικών σε μορφή λίστας και η κάθε επαφή χρειάζεται να την ανατρέξει για να ελέγξει αν συμπεριλαμβάνεται και το δικό του αναγνωριστικό σε αυτή. Σύμφωνα με το άρθρο μια ορθότερη λύση βασισμένη στην Personal Data Store προσέγγιση σε όρους ιδιωτικότητας και προσωπικού απορρήτου, θα αποτελούσε αν ο χρήστης έχει πλήρη έλεγχο των δεδομένων του και ένα προσωπικό περιβάλλον λογισμικού με μοναδική προσβασιμότητα όπου θα μπορεί να αποθηκεύει και να ελέγχει τα δεδομένα του και να τα διαχειρίζεται όπως επιθυμεί π.χ., αποστολή συγκεκριμένου βαθμού πληροφορίας σε άλλους χρήστες ή υγειονομικές αρχές εφόσον το επιθυμεί. Η ενίσχυση της εμπιστοσύνης των πολίτων με την παροχή ελέγχου στα δεδομένα τους και ο

σεβασμός στα ανθρώπινα δικαιώματα θα πρέπει να είναι προτεραιότητες για τις εφαρμογές contact tracing.

Για την ανάπτυξη αυτών των εφαρμογών η Google και η Apple συνεργάστηκαν αρχικά σε πρώτη φάση για την δημιουργία ενός API framework contact tracing (το οποίο αξιοποιούν τα apps που αναπτύσσονται από τους κατάλληλους φορείς) που θα μπορεί να πραγματοποιεί ανίχνευση συσκευών ακόμα και αν έχουν διαφορετικά λειτουργικά συστήματα το καθένα (Android και iOS) και μετέπειτα στην εφαρμογή αυτής της λειτουργίας κατευθείαν μέσα στο λειτουργικό σύστημα του τηλεφώνου. Τα exposure notifications τα όποια δημιουργούνται από τον κάθε φορέα υγείας με βάση το πρότυπο που καθορίζεται από τις δύο εταιρείες, αποτελούνται από παραμέτρους όπως η απόσταση μεταξύ των χρηστών και ο χρόνος επαφής με ένα θετικό κρούσμα τα οποία ρυθμίζονται αποκλειστικά από αυτόν, και θα προβάλλονται στον χρήστη σαν ειδοποιήσεις όπου θα προτείνουν την συμμετοχή στην διαδικασία ανίχνευσης ύποπτων κρουσμάτων. Η χρήση του app και η ενεργοποίηση της λειτουργίας ανίχνευσης δεν καθίσταται υποχρεωτική προς τον χρήστη και η συμμετοχή είναι εθελοντική [Brandom 2020][Robertson 2020].

Τέλος αξίζει να αναφερθεί ότι υπάρχει μια πληθώρα περιορισμών για τις εφαρμογές contact tracing. Μια εφαρμογή απαιτεί υψηλό αριθμό υιοθέτησης της από τους πολίτες για αποδοτική λειτουργία και επειδή βασίζεται στο GAEN (Google Apple Exposure Notification Service) μπορεί να αναγνωρίσει μόνο τις επαφές όπου διάρκεια μαζί τους ήταν τουλάχιστον πέντε λεπτά. Επίσης οι εφαρμογές contact tracing πραγματοποιούν σφάλματα καθώς εξασθενείται το σήμα με την επαφή του ανθρωπίνου σώματος ή αντανακλάται από το περιβάλλον. Αυτό έχει ως αποτέλεσμα επαφές με ανθρώπους που δεν έχουν προσβληθεί από τον ίο να θεωρηθούν ως επικίνδυνες – επαφές με κρούσμα (false positive) και το αντίστροφο [Kindt et al. 2021].

2.4 Ο τομέας του Urban Data Science και οι προκλήσεις που αντιμετωπίζει

Ο τομέας του Urban Data Science ασχολείται με την τεχνική υποδομή που απαιτείται για τη μελέτη και αντιμετώπιση κρίσιμων ζητημάτων που πλήττουν τις σημερινές πόλεις μέσα από μια data-driven προσέγγιση. Στην σύγχρονη εποχή όπου υπάρχει μεγάλη ποσότητα διαθέσιμων δεδομένων (Big Data) η αξιοποίηση τους μπορεί να συμβάλει καθοριστικά στον καλύτερο σχεδιασμό των πόλεων και στην βελτίωση της ποιότητας ζωής σε αυτές με την χρήση καινοτόμων μεθόδων όπως για παράδειγμα την πρόβλεψη απυχημάτων και κυκλο-

14 Κεφάλαιο 2 *Βιβλιογραφική έρευνα - Σχετικές εργασίες*

φορικής συμφόρησης στους δρόμους με προτάσεις ενναλακτικών διαδρομών στους χρήστες. Άλλα παραδείγματα ενός επιτυχημένου urban plan είναι η δημιουργία αποδοτικής δημόσιας συγκοινωνίας (όπως υπάρχει στην Ιαπωνία και Νότια Κορέα) που βοηθούν στην μείωση του περιβαλλοντικού αποτυπώματος άνθρακα και της κυκλοφορικής συμφόρησης στους δρόμους [Net4things 2021] καθώς και η ύπαρξη πάρκων και «πράσινων χώρων» στις πόλεις όπου συμβάλλουν στην φυσική υγεία των πολιτών και στην ποιότητα του αέρα [Duarte and deSouza 2020].

Νέες τεχνολογίες όπως τα αυτόνομα αυτοκίνητα, τα ηλεκτρικά αυτοκίνητα και οι υπηρεσίες κοινής κινητικότητας (shared mobility) αναπτύσσονται και εκτιμάται ότι θα παίξουν πρωταρχικό ρόλο στο μέλλον των πόλεων. Σύμφωνα με τους [Miskolczi et al. 2021] το πιο πιθανό από τα τέσσερα σενάρια που παρουσιάζονται για την κινητικότητα στις πόλεις το 2030 είναι το shared mobility να έχει κυριαρχήσει, με την δημόσια συγκοινωνία να έχει αναβαθμιστεί και να αποτελεί πλέον το βασικότερο μέσο μετακίνησης και τα ηλεκτρικά αυτοκίνητα να έχουν αντικαταστήσει σε μεγάλο βαθμό τα συμβατικά. Παράλληλα η ανάπτυξη και η υιοθέτηση αυτόνομων αυτοκινήτων είναι σχεδόν μηδενική και νέοι τρόποι μετακίνησης έχουν υιοθετηθεί όπως το ποδήλατο, περπάτημα, ηλεκτρικά ποδήλατα (e-bikes) και ηλεκτρικά σκούτερ (e-scooters). Υπάρχει μεγάλη πτωτική τάση στα ιδιωτικά αυτοκίνητα τέτοιου μεγέθους που μπορεί ενδεχομένως να προκαλέσει κλείσιμο κάποιων αυτοκινητοβιομηχανιών και το αρνητικό αποτύπωμα της κινητικότητας στην κυκλοφορική συμφόρηση και στο περιβάλλον έχει μειωθεί σημαντικά.

Ο σχεδιασμός ενός urban mobility plan πρέπει να γίνεται από κατάλληλη ομάδα η οποία απαρτίζεται και από μέλη διαφορετικού επιστημονικού υποβάθρου (Αρχιτεκτονική, Μηχανική, Τέχνες, Στατιστική, Νομική, Διοίκηση Επιχειρήσεων) και διαθέτει τις κατάλληλες γνώσεις και δεξιότητες σε θέματα τεχνητής νοημοσύνης και μηχανικής μάθησης, διαχείρισης βάσεων δεδομένων, ηλεκτρονικής διακυβέρνησης, οπτικοποιήσης δεδομένων και θεωρίας αισθητήρων [Ampountolas et al. 2018]. Για αυτό το σκοπό η ύπαρξη πρακτικής, κατάλληλων μαθημάτων urban science στα πανεπιστήμια και η κατάλληλη καθοδήγηση από τους καθηγητές μπορούν να συντελέσουν στην ανάπτυξη νέας γενιάς urban planners [Zhongwen 2018]. Είναι ιδιαίτερα σημαντικό να επισημανθεί ότι καθώς τα δεδομένα έχουν συγκεκριμένη διάρκεια ζωής και ο τρόπος εξαγωγής και χρήσης τους χαρακτηρίζονται από ιδεολογικές αξίες, η εφαρμογή urban data-driven μοντέλων επιφέρει πολιτικές, οικονομικές, κοινωνικές επιτώσεις για τις όποιες θα πρέπει να είναι ενήμεροι οι σχεδιαστές [Duarte and deSouza 2020].

Η εργασία των [Andrienko et al. 2021] καλύπτει τις κύριες πτυχές της ανάλυσης δεδομένων που αφορούν τις πόλεις, εστιάζοντας σε ζητήματα απορρήτου, αλγόριθμους, εφαρμογές και δεδομένα γεωαναφοράς από τα μέσα κοινωνικής δικτύωσης. Τα κοινωνικά δίκτυα βασισμένα στην τοποθεσία (Location based social networks) που μελετώνται είναι online κοινωνικές πλατφόρμες που επιτρέπουν στους χρήστες να διαμοιράσουν την τοποθεσία που βρίσκονται στο δημόσιο κύκλο τους μέσω check-in μαζί με την ημερομηνία (ώρα, ημέρα) και τον σκοπό της δραστηριότητας τους (ψώνια, φαγητό). Από την ανάλυση μεγάλου αριθμού αυτών των δεδομένων προκύπτει η ταυτότητα της κάθε περιοχής-γειτονιάς της πόλης. Η εύρεση της τοποθεσίας των χρηστών δεδομένου ενός συνόλου GPS στιγμάτων-συντεταγμένων των χρηστών είναι ουσιαστικά πρόβλημα ομαδοποίησης των παρατηρήσεων που αντιστοιχούν στην ίδια τοποθεσία. Για την πραγματοποίηση της ομαδοποίησης χρησιμοποιούνται διάφοροι αλγόριθμοι με πιο σημαντικό τον TOSCA, ο οποίος είναι ένας εύρωστος και χαμηλού υπολογιστικού κόστους αλγόριθμος ο οποίος συνδυάζει δύο clustering μεθόδους και μια στατιστική προσέγγιση. Ο TOSCA αντιμετωπίζει αποδοτικά τα μειονεκτήματα των DBSCAN και OPTICS καθώς καθορίζει αυτόματα ένα κατώφλι απόστασης ενώ παράλληλα προσαρμόζει την ομαδοποίηση στα μοτίβα της ατομικής κινητικότητας του κάθε χρήστη.

2.5 Επιπτώσεις της πανδημίας στους τρόπους μετακίνησης των πολιτών και στα αστικά κέντρα

Ιταλία

Οι [Cavallaro et al. 2021] αναλύουν τις επιπτώσεις των περιοριστικών μέτρων που εφαρμόστηκαν στην Ιταλία κατά το πρώτο μισό του 2020 στις ροές κινητικότητας, προτείνοντας παράλληλα μια κατάλληλη τροποποιημένη εκδοχή του αλγορίθμου *Apriori* για την ποσοτικοποίηση και αξιολόγηση των εν λόγω ροών. Επιπλέον, υποστηρίζεται ότι η χρήση clustering και συχνών ακολουθιακών προτύπων είναι οι δύο βασικές τεχνικές για την εύρεση μοτίβων – προτύπων σε τροχιές κινούμενων αντικειμένων καθώς με την συσταδοποίηση τοποθετούνται στις ίδιες ομάδες παρόμοιες τροχιές ενώ με την εξαγωγή συχνών προτύπων αναλύονται κινούμενα αντικείμενα τα οποία ταξιδεύουν με μια προκαθορισμένη αλληλουχία σε συγκεκριμένες τοποθεσίες σε ένα προκαθορισμένο χρονικό διάστημα. Παρόμοια συμπεράσματα μπορούν να εξαχθούν από το ερευνητικό project [Pepe et al. 2020] που αναπτύσσει ένα proximity network βασισμένο στις τοποθεσίες και στις ώρες κινητικότητας των χρηστών κατά την διάρκεια της ημέρας στην Ιταλία. Η εγγύτητα ανάμεσα σε δύο

16 Κεφάλαιο 2 Βιβλιογραφική έρευνα - Σχετικές εργασίες

χρήστες που ορίστηκε για το μοντέλο, συμβαίνει όταν δύο άνθρωποι βρίσκονται στην ίδια γεωγραφική περιοχή ανάμεσα σε ένα κύκλο ακτίνας 50m σε χρονικό διάστημα μιας ώρας. Σημαντικά εργαλεία που χρησιμοποιήθηκαν και μπορεί να αποδειχθούν χρήσιμα και σε μελλοντικές έρευνες αποτελούν ο μεσος ρυθμός επαφών σε ένα προκαθορισμένο διάστημα, το εβδομαδιαίο μητρώο προέλευσης – προορισμού και ο μέσος ημερήσιος βαθμός του χρήστη στο proximity network.

Cardiff, Ουαλία, Ηνωμένο Βασίλειο

Η έρευνα των [Angell and Potoglou 2022] εξετάζει τις άμεσες και τις μακροπρόθεσμες επιπτώσεις της πανδημίας στο Cardiff της Ουαλίας. Τα ερωτήματα της έρευνας εστιάστηκαν στην ανάλυση των χαρακτηριστικών ενός ταξιδιού και συγκεχριμένα στην συχνότητα του ταξιδιού, στον τρόπο μετακίνησης και στους χρόνους αναχώρησης από/προς την δουλεία των ερωτηθέντων. Τα αποτελέσματα εντόπισαν μείωση της συχνότητας μετακίνησης προς την εργασία και αύξηση των ημερών εργασίας από το σπίτι. Οι ώρες μετακίνησης προς και από την εργασία ενδέχεται να παραμείνουν ίδιες μετά την πανδημία.

Η ευελιξία των πολιτών βρέθηκε να είναι ο κύριος παράγοντας που καθορίζει τα αποτελέσματα των επιδράσεων της πανδημίας, η οποία εξαρτάται από τη γεωγραφική περιοχή στην οποία διαμένουν, την ύπαρξη της δημόσιας συγκοινωνίας, τον αστικό ιστό της πόλης, το κοινωνικό – οικονομικό προφίλ τους και τα χαρακτηριστικά της μετακίνησης που επιθυμούν να πραγματοποιήσουν κάθε φορά. Η πανδημία οδήγησε τους πολίτες να ξανασκεφτούν τον βασικό τρόπο μετακίνησης τους. Αυτοί που ταξιδεύουν με αυτοκίνητο εμφανίζουν ένα μεγαλύτερο βαθμό προσαρμοστικότητας σε σχέση με τα άλλα μέσα μεταφοράς λόγω του μεγαλύτερου εύρους επιλογών στον σχεδιασμό του ταξιδιού και βρέθηκε ότι ήταν ο προτιμητέος τρόπος μετακίνησής καθώς μειώνει την πιθανότητα μόλυνσης από τον ιό. Το γεγονός ότι οι υποδομές για την δημόσια μετακίνηση και για την ενεργή μετακίνηση (ποδηλατοδρόμοι, δρόμοι για e-scooters) χρειάζονται βελτιώσεις εξηγεί το χαμηλό αποτέλεσμα της έρευνας σε σχέση με τη χρήση του αυτοκινήτου πριν την έλευση της πανδημίας. Παρόλα αυτά η ζήτηση για χρήση ενεργών τρόπων μετακίνησης αναμένεται να αυξηθεί στο μέλλον.

Hong Kong, Λαϊκή δημοκρατία της Κίνας

Η εργασία των [Loo and Huang 2022] εστιάζει στην επίδραση της πολιτικής της εργασίας από το σπίτι (WFH), που υιοθετήθηκε από τις εταιρείες στο Hong Kong και στην κυκλοφοριακή συμφόρηση στους δρόμους. Τα αποτελέσματα

της έρευνας είναι ιδιαίτερα σημαντικά, καθώς παρά την εφαρμογή του WFH τόσο στον ιδιωτικό όσο και στον δημόσιο τομέα δεν υπήρξαν περιορισμοί στις τοπικές μετακινήσεις. Αναλυτικότερα οι πρωινές και απογευματινές ώρες αιχμής συνέχισαν να υφίστονται κατά την διάρκεια εφαρμογής του μέτρου, δεδομένου ότι δεν δυνατόν όλες οι δουλείες να πραγματοποιηθούν από το σπίτι και ότι οι εταιρείες στην Hong Kong συνέχισαν να δουλεύουν με υβριδικό τρόπο εργασίας.

Οι αλλαγές στην κίνηση στους δρόμους αντικατοπτρίζουν την αποτελεσματικότητα του μέτρου WFH. Εξαιτίας της κοινωνικής ανισορροπίας του Hong Kong κυκλοφοριακή συμφόρηση παρατηρούνταν στους δρόμους πριν την έλευση του κορονοϊού. Κατά την διάρκεια της πανδημίας το μέτρο της εργασίας από το σπίτι επηρέασε θετικά τις πρωινές και απογευματινές ώρες αιχμής και κυρίως εξομάλυνε την πρωινή καμπύλη αιχμής (κάτι που αποτύπωνται στον δείκτη CI στην εμπορική συνοικία και στα κεντρικά σημεία της πόλης). Τέλος αξίζει να αναφερθεί ότι χωρίς την εφαρμογή του μέτρου, το μέγεθος της πρωινής κυκλοφοριακής αιχμής αναμένεται να επιστρέψει σε επίπεδα προ πανδημίας ενώ μια μικρή μείωση θα εμφανιστεί στην απογευματινή ώρα αιχμής.

Κωνσταντινούπολη, Τουρκία

Σκοπός της τρέχουσας έρευνας των [Paköz and Iik 2022] είναι η μελέτη της σχέσης ανάμεσα στην ζωτικότητα της πόλης της Κωνσταντινούπολης (urban vitality), του υγειούς περιβάλλοντος και της πυκνότητας της κατά την διάρκεια της πανδημίας. Οι συνοικίες ανάλογα την πυκνότητα τους κατηγοριοποιήθηκαν σε χαμηλής, μέτριας και υψηλής πυκνότητας και αξιολογήθηκαν από τους κατοίκους σύμφωνα με τους ακόλουθους έξι τομείς: ζωντάνια, κινητικότητα, ασφάλεια, υγεία, καθαριότητα, οργάνωση. Οι έξι τομείς περιορίστηκαν στους δύο παράγοντες του urban vitality και health environment χρησιμοποιώντας Principal Component Analysis. Η έννοια του urban vitality δείχνει το πλήθος των δραστηριοτήτων που πραγματοποιούνται σε μια πόλη και σχετίζεται έντονα με την ζωντάνια και κινητικότητα των πολιτών. Ένα υγιές περιβάλλον στην πόλη σχετίζεται όχι μόνο από όρους υγείας αλλά και με ασφάλειας, καθαριότητας, προσβασιμότητας, κινητικότητας και αλληλεπίδρασης. Ο ορισμός ενός τέτοιου περιβάλλοντος διαφέρει ανάλογα την ηλικία, το εισόδημα, την οικογενειακή κατάσταση και την περιοχή στην οποία διαμένει κάποιος. Τα ευρήματα ανακάλυψαν ότι η διασπορά τού ιού επηρεάζεται από την υψηλή πυκνότητα στις πόλεις καθώς οι άνθρωποι που διαμένουν εκεί δεν είναι σε θέση να εφαρμόσουν τα μέτρα κοινωνικής αποστασιοποίησης.

18 Κεφάλαιο 2 Βιβλιογραφική έρευνα - Σχετικές εργασίες

Σκοπός της έρευνας των [Bagdatli and Ipek 2022] είναι η κατανόηση των προτιμήσεων των φοιτητών των Πανεπιστημίων της Κωνσταντινούπολης σε διάφορα διαθέσιμα μέσα μεταφοράς κατά την διενέργεια των μετακινήσεων τους την περίοδο μετά την πανδημία. Οι προτιμήσεις στα διάφορα μέσα μετακίνησης εξαρτώνται από την απόσταση της μετακίνησης, την διάρκεια, το κόστος, το εισόδημα, το μορφωτικό επίπεδο, την ηλικία, το φύλο και τα προσωπικά χαρακτηριστικά του ενδιαφερόμενου (συνήθειες, τρόπος ζωής, κουλτούρα). Ο λόγος που επιλέχθηκαν οι φοιτητές ως δείγμα έρευνας είναι ότι η κινητικότητα τους είναι αρκετά υψηλή και είναι πιο ευέλικτοι σε σχέση με τα άλλα ηλικιακά γκρουπ στο να χρησιμοποιήσουν διαφορετικούς τρόπους μετακίνησης. Οι δημόσιες συγκοινωνίες παρέχουν ένα σημαντικό πλεονέκτημα στις πόλεις όσον αφορά τις βιώσιμες μετακινήσεις αλλα η ύπαρξη μέτρων με στόχο την τήρηση κοινωνικών αποστάσεων για την μείωση εξάπλωσης του ιού μετριάζει την αποτελεσματικότητα τους. Ειδικότερα η πανδημία προκάλεσε μια κατακόρυφη μείωση της ζήτησης για δημόσια λεωφορεία, shared minibuses και υπόγειους σιδηροδρόμους ενώ μεγάλη αύξηση παρατηρήθηκε για ιδιωτικά αυτοκίνητα. Αύξηση παρατηρήθηκε και για e-scooters, hoverboard και σε ενεργούς τρόπους μετακίνησης (ποδήλατο, περπάτημα) αλλά παραμένει αβέβαιο κατά πόσο η ζήτηση τους θα παραμείνει σταθερή στο μέλλον.

Στο μέλλον έχουν αναπτυχθεί 3 πιθανά σενάρια για το μέλλον της μετακίνησης στις πόλεις: Οι άνθρωποι δεν θα χρησιμοποιούν δημόσιες συγκοινωνίες και θα συνεχίσουν να χρησιμοποιούν παραδοσιακά ιδιωτικά μέσα, μακροπρόθεσμά οι άνθρωποι θα στραφούν σε e-scooters /hoverboards και σε ενεργούς τρόπους μετακίνησης όσο και σε ιδιωτικά αυτοκίνητα και τελευταίο σενάριο είναι ότι όταν τελειώσει η πανδημία οι άνθρωποι θα επιστρέψουν στους δημόσιους τρόπους μετακίνησης. Το δεύτερο σενάριο φαίνεται να επιβεβαιώνεται από τα αποτελέσματα της έρευνας καθώς βρέθηκε μια αύξηση της ζήτησης σε ποδήλατα, e-scooters και του περπατήματος ως τρόπου μετακίνησης. Παρόλα αυτά το τρέχον σενάριο που ενδέχεται να πραγματοποιηθεί μακροπρόθεσμα στο μέλλον χωρίς την παρέμβαση του κράτους είναι το πρώτο, δηλαδή η κυριαρχία των ιδιωτικών αυτοκινήτων – σενάριο που θεωρείται εφιαλτικό για το urban mobility και το περιβάλλον. Για αυτό τον λόγο απαιτείται η δημιουργία περισσότερων υποδομών για e-scooters και ποδήλατα και πραγματοποιήση βελτιώσεων και βραχυπρόθεσμων αλλαγών στο δίκτυο των δημόσιων συγκοινωνιών όπως είναι η μείωση του ποσοστού χωρητικότητας στα οχήματα, μείωση του πλήθους που περιμένουν στους σταθμούς με περισσότερα δρομολόγια, τήρηση κοινωνικών αποστάσεων και δημιουργία συστήματος υποστήριξης ανέπαφων συναλλαγών για την έκδοση εισιτηρίων.

2.6 Ευρωπαϊκά προγράμματα και πρωτοβουλίες για τον μετασχηματισμό των πόλεων

Η Ευρωπαϊκή Ένωση έχει συμβάλει ενεργά στην θεσμοθέτηση αρχών για την δημιουργία επιτυχημένων και ορθών Sustainable Urban Mobility Plans (SUMPs) για αστικές πόλεις. Βασικοί πυλώνες τους είναι ο σχεδιασμός βιώσιμης κινητικότητας για όλη την πόλη, ο καθορισμός μακροπρόθεσμου σχεδίου και σαφής υλοποίηση του, η συνεργασία με άλλους φορείς και ινστιτούτα χωρίς περιορισμούς, η ανάπτυξη των διάφορων μέσων μετακίνησης με ενιαίο τρόπο, η ανάμειξη των πολιτών και ενδιαφερόμενων επενδυτών στην διαδικασία, η συμφωνία για έλεγχο και αξιολόγηση, η ερμηνεία της απόδοσης σε όλη την φάση του σχεδιασμού και η διασφάλιση ποιότητας [Doe 2021]. Στην Ελλάδα μόνο η Θεσσαλονίκη, η Σίφνος, η Κέα και η Νάξος έχουν υιοθετήσει μέχρι στιγμής ένα πλάνο SUMP. Βασικοί στόχοι που χαρακτηρίζουν το σχέδιο των νησιών είναι η χρήση έξυπνης ενέργειας για την ηλεκτροδότηση τους, η παροχή και δημιουργία ασφαλών διαδρομών για τρέξιμο και ποδήλατο προτρέποντας τους τουρίστες να μειώσουν την χρήση diesel αυτοκινήτου στην κινητικότητα τους και η δημιουργία σημείων φόρτισης ηλεκτρικών αυτοκινήτων [Figg 2021]. Αντίστοιχα για την Θεσσαλονίκη κάποιοι από τους βασικούς στόχους του σχεδίου που εφαρμόστηκε με επιτυχία είναι η χρήση ενιαίου εισιτηρίου στην δημόσια συγκοινωνία, η αναδιοργάνωση των ταξί, η προώθηση και η χρήση του τραμ, η δημιουργία κατάλληλου ποδηλατοδρόμου, bike – sharing συστήματος και θαλάσσιου συστήματος μεταφορών στο Θερμαϊκό και η συνέχιση των έργων κατασκευής του μετρό [Mourey 2015], [Beekman 2014].

Το EIT Urban Mobility είναι μια πρωτοβουλία του Ευρωπαϊκού Ινστιτούτου Καινοτομίας και Τεχνολογίας. Ήπαρχουν πέντε κόμβοι Καινοτομίας Urban Mobility του EIT στην Ευρώπη και ένας κεντρικός ο οποίος βρίσκεται στην Βαρκελώνη (Η Ελλάδα βρίσκεται στο νότιο κόμβο που βρίσκεται και αυτός στην Βαρκελώνη). Στόχος του προγράμματος είναι η επίτευξη βιώσιμης αστικής ανάπτυξης, η αποσυμφόρηση των μέσων μαζικής μεταφοράς, η δημιουργία οικολογικού και ασφαλούς τρόπου μετακίνησης για τους ανθρώπους και για τα αγαθά τους, η αξιοποίηση δεδομένων που αφορούν την κινητικότητα των πολιτών και την μορφολογία των πόλεων, η ενίσχυση της ανταγωνιστικότητας των μέσων μεταφορών στην Ευρώπη και διαμόρφωση ενός κατάλληλου πλαισίου για την αστική διακυβέρνηση [European Institute of Innovation and Technology 2022].

Το 2020 ο φορέας του EIT Urban Mobility διεξήγαγε μια έρευνα [EIT Urban Mobility 2021b] σε 16 ευρωπαϊκές πόλεις για να κατανοήσει και να

20 Κεφάλαιο 2 Βιβλιογραφική έρευνα - Σχετικές εργασίες

αναλύσει τις στρατηγικές κινητικότητάς των πολιτών κατά τη διάρκεια της πανδημίας COVID-19 με τις πόλεις της Ρώμης στην Ιταλία, Λουντ της Σουηδίας, Ταρτού της Εσθονίας να πρωτοπορούν. Στην Ρώμη για την προώθηση ασφαλών ταξιδιών εφαρμόστηκε η μείωση της χωρητικότητας επιβατών στα λεωφορεία και στους συρμούς του μετρό κατά τα δύο τρίτα. Η υλοποίηση του μέτρου πραγματοποιήθηκε με την αφαίρεση των θέσεων και την καταγραφή του αριθμού των επιβατών στα βαγόνια. Παράλληλα ενισχύθηκε η μικροκινητικότητα με την επιβολή εκπτώσεων σε όλη την πόλη για την ενοικίαση ηλεκτρικών σκούτερ, συμβατικών ποδήλατων και ηλεκτρικών ποδηλάτων, και την δημιουργία 1200 σταθμών φόρτισης σε όλη την Ρώμη έπειτα βέβαια από την συνεισφορά των πολίτων με τις προτάσεις τους για τις τοποθεσίες κατασκευής τους μέσω κατάλληλης πλατφόρμας.

Στην πόλη Lund της Σουηδίας οι πεζόδρομοι και οι ποδηλατόδρομοι έγιναν πιο «φιλικοί» σε περίπτωση πτώσης με την επίστρωση τους με ανάμεικτη άσφαλτο και σωματίδια καιουτσούκ προερχόμενα από ανακυκλωμένα ελαστικά, διευκολύνοντας την ποδηλασία και αποτρέποντας τους τραυματισμούς σε περίπτωση πτώσης. Στην πόλη Tartu της Εσθονίας υλοποιήθηκαν τα συστήματα κοινής χρήσης ποδηλάτων με 500 ηλεκτρονικά και 250 συμβατικά ποδήλατα και 69 σταθμούς κοινής χρήσης ποδηλάτων σε όλη την πόλη, με μακροπρόθεσμο στόχο το ποδήλατο να γίνει ο προτιμώμενος τρόπος μετακίνησης. Ο υπεύθυνος του έργου δήλωσε ότι αξιοποιήθηκαν δεδομένα της πόλης και κινητικότητας για καθοριστούν οι θέσεις των σταθμών, προσδιορίζοντας τις ζώνες με την μεγαλύτερη ροή πεζών και την χαμηλότερη πρόσβαση στα λεωφορεία.

Άλλα ευρωπαϊκά έργα EIT URBAN Mobility που ξεχώρισαν στην έκθεση [EIT Urban Mobility 2021a] είναι αυτό της κατασκευής και της εκπαίδευσης ενός συστήματος μηχανικής μάθησης από τις εταιρείες HaCon και Siemens Mobility που προβλέπει τον αριθμό των επιβατών στα μέσα μαζικής μεταφοράς, παρέχοντας στους ταξιδιώτες εκτιμήσεις για το μέγεθος του συνωστισμού που θα παρουσιαστεί στα μέσα μαζικής μεταφοράς και λαμβάνοντας υπόψη την πληρότητα ανάλογα την ώρα, την ημέρα της εβδομάδας, τις επίσημες αργίες και τις καιρικές συνθήκες. Τέλος ξεχώρισε και το Spin, τα σκούτερ κοινής ιδιοκτησίας με την ολιστική προσέγγιση που στοχεύουν στην αύξηση της ασφάλειας και της άνεσης μιας μετακίνησης με μέσο μικροκινητικότητας, δίνοντας έμφαση στη πρόληψη και στη μείωση της σοβαρότητας των τραυματισμών. Τα σκούτερ μπορούν να αναγνωρίσουν πεζοδρόμια με την χρήση κάμερας που είναι στερεωμένη σε αυτά και μπορούν να αποθαρρύνουν τον χρήστη σε περίπτωση οδήγησης ή σταύθμευσης σε αυτά, επιβραδύνοντας αυτόματα το σκούτερ και κάνοντας χρήση ηχητικών σημάτων.

3. Θεωρητικό υπόβαθρο

Η σύντομη εισαγωγή στα μοντέλα μηχανικής μάθησης και στις μεθόδους προεπεξεργασίας που υιοθετήθηκαν στην πτυχιακή βασίστηκε στα εξής βιβλία [Theodoridis and Koutroumbas 2012], [Vlahavas et al. 2020] και [Tang et al. 2020].

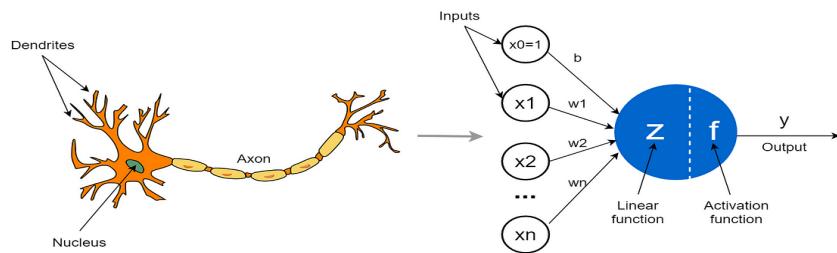
3.1 Multilayer Perceptron

Ένας απλός τεχνητός νευρώνας (neuron) είναι το υπολογιστικό ισοδύναμο ενός Perceptron και μπορεί να λύσει ρεαλιστικά προβλήματα που είναι γραμμικώς διαχωρίσιμα. Η ανάπτυξη των τεχνητών νευρωνικών δικτύων προσομειώνει την λειτουργία του ανθρώπινου εγκεφάλου. Ο ανθρώπινος εγκέφαλος αποτελείται από νευρικά κύτταρα – νευρώνες οι οποίοι συνδέονται μεταξύ τους με νευρίτες (προεκτάσεις από το σώμα των κυττάρων). Κάθε φορά που ένα ερέθισμα παράγεται, διεγείρεται ο βιολογικός νευρώνας ο οποίος μεταδίδει νευρικές διεγέρσεις μεταξύ των νευριτών σε άλλους νευρώνες μέσω των δενδριτών. Κάθε τεχνητός νευρώνας δέχεται ένα πλήθος από εισόδους οι οποίες πολλαπλασιάζονται με βάρη (weights) και στην συνέχεια αθροίζονται στον αθροιστή. Εκτός από τα εισερχόμενα σήματα, ο νευρώνας δέχεται ως είσοδο (input) και μια σταθερή τιμή $x_0 = 1$ η οποία διαμορφώνεται από το βάρος w_0 το οποίο ονομάζεται πόλωση (bias). Η συνολική τιμή που παράγει ο αθροιστής εισέρχεται στην συνέχεια στην συνάρτηση ενεργοποίησης (activation function) f του νευρώνα, η οποία παράγει την έξοδο (output) του. Η πόλωση χρησιμοποιείται για την καλύτερη προσέγγιση της επιθυμητής συνάρτησης αφού με την χρήση της σταθεράς μπορεί να γίνει κατάλληλη μετατόπιση αυτής.

Ο κυρίαρχος σκοπός της λειτουργίας ενός τεχνητού νευρωνικού δικτύου είναι να μαθαίνει μέσω εκπαίδευσης πολύπλοκα μη γραμμικά όρια αποφάσεων, αμιγώς από δεδομένα. Κάθε δίκτυο δέχεται ορισμένες εισόδους και δίνει ορισμένες εξόδους. Η εκπαίδευση γίνεται με την παρουσίαση ενός συνόλου

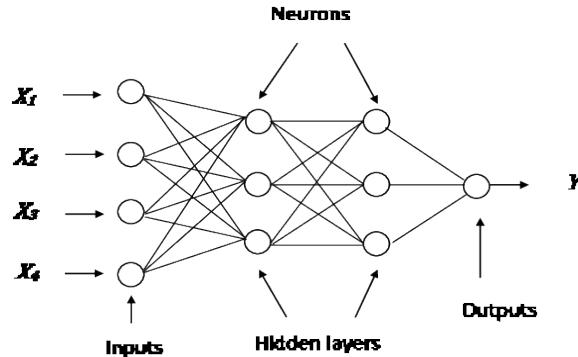
22 Κεφάλαιο 3 Θεωρητικό υπόβαθρο

δεδομένων στο δίκτυο, αντιπροσωπευτικού του προβλήματος που είναι επιθυμητό να μάθει. Ένας νευρώνας είναι το βασικό δομικό στοιχείο πολυπλοκότερων μηχανισμών που ονομάζονται πολυστρωματικά Perceptron ή πολυστρωματικά τεχνητά νευρωνικά δίκτυα και μπορούν να λύσουν προβλήματα που δεν είναι γραμμικώς διαχωρίσιμα. Η αρχιτεκτονική πολλών επιπέδων μπορεί να θεωρηθεί ως μια γενίκευση του perceptron και είναι γνωστή ως εμπροσθοτροφοδοτούμενο perceptron. Εδώ αντί για ένα και μοναδικό νευρώνα, υπάρχει επαλληλία κρυφών νευρωνικών επιπέδων στο δίκτυο. Κάθε επίπεδο μπορεί να έχει ενδεχομένως ένα διαφορετικό πλήθος από νευρώνες που βρίσκονται σε επαλληλία.



Εικόνα 3.1: Η μορφή ενός νευρώνα ο οποίος δέχεται διανύσματα εισόδου και χρησιμοποιώντας την συνάρτηση ενεργοποίησης παράγει την έξοδο y

Στα τεχνητά νευρωνικά δίκτυα πολλών επιπέδων χρησιμοποιούνται τρεις τύποι επιπέδων. Το πρώτο επίπεδο του δικτύου ονομάζεται επίπεδο εισόδου και χρησιμοποιείται για την αναπαράσταση των εισόδων από τα χαρακτηριστικά. Οι είσοδοι στην συνέχεια είναι αυτοί που τροφοδοτούν τα ενδιάμεσα επίπεδα, γνωστά και ως κρυφά επίπεδα. Οι κόμβοι στα κρυφά επίπεδα λειτουργούν με σήματα που λαμβάνουν από τους κόμβους εισόδου ή από τους κόμβους του προηγούμενου κρυφού επίπεδου και παράγουν τιμές ενεργοποίησης, οι οποίες μεταδίδονται στα επόμενα επίπεδα. Το τελευταίο επίπεδο ονομάζεται επίπεδο εξόδου και επεξεργάζεται τις τιμές ενεργοποίησης που λαμβάνει από το προηγούμενο επίπεδο για να παράγει προβλέψεις στις τιμές εξόδου.



Εικόνα 3.2: Η μορφή ενός Multilayer Perceptron νευρωνικού δικτύου

3.2 Μηχανές διανυσματικής υποστήριξης

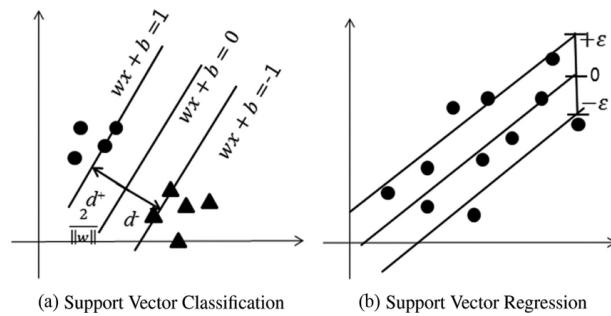
Η μέθοδος της μηχανής διανυσματικής υποστήριξης SVM (Support Vector Machine) επιδιώκει να βρει το σύνορο που απέχει όσο το δυνατόν περισσότερο από τα παραδείγματα των κλάσεων που διαχωρίζει. Σε προβλήματα παλινδρόμησης (regression) στόχος είναι να βρεθεί μια συνάρτηση που προσεγγίζει τα σημεία εκπαίδευσης ελαχιστοποιώντας το σφάλμα πρόβλεψης. Η συνάρτηση απώλειας σε SVR προβλήματα δίνει έμφαση στο να μειώσει τους συντελεστές του ζητούμενου υπερεπιπέδου μέσω L2 norm που περιορίζει την υπερπροσαρμογή (overfitting). Αξίζει να σημειωθεί ότι ο βέλτιστος ταξινομητής ενός περιθωρίου μιας μηχανής διανυσματικής στήριξης είναι μοναδικός.

Το υπερεπίπεδο χαρακτηρίζεται μέγιστου περιθωρίου σε γραμμικώς διαχωρίσιμα προβλήματα (υπάρχει υπερεπίπεδο που διαχωρίζει τέλεια τις κατηγορίες του συνόλου δεδομένων) και ορίζεται από έναν πεπερασμένο αριθμό παραδειγμάτων του συνόλου εκπαίδευσης που ονομάζονται διανύσματα υποστήριξης. Τα διανύσματα υποστήριξης είναι δηλαδή τα διανύσματα εκπαίδευσης που είναι πλησιέστερα στον ταξινομητή. Τα σύνορα απόφασης όπου χαρακτηρίζονται από μεγάλα περιθώρια εμφανίζουν μεγαλύτερη ανοχή σε φαινόμενα υπερπροσαρμογής, έχουν σχετικά χαμηλές υπολογιστικές απαιτήσεις σε σχέση με τα νευρωνικά δίκτυα και τείνουν να έχουν καλύτερη απόδοση γενίκευσης από εκείνα με μικρά περιθώρια.

Σε πραγματικά προβλήματα είναι αρκετά συχνό φαινόμενο τα δεδομένα να μην είναι γραμμικά διαχωρίσιμα ή κάποια να έχουν ταξινομηθεί λάθος. Για την αντιμετώπιση αυτού χρησιμοποιείται ένα χαλαρό περιθώριο (soft margin) που επιτρέπει στην μηχανή διανυσματικής υποστήριξης να κατασκευάσει γραμμικά

24 Κεφάλαιο 3 Θεωρητικό υπόβαθρο

υπερεπίπεδα λαμβάνοντας υπόψη τον συμβιβασμό μεταξύ του πλάτους του περιμένοντος και του πλήθους των σφαλμάτων εκπαίδευσης του γραμμικού ορίου απόφασης.



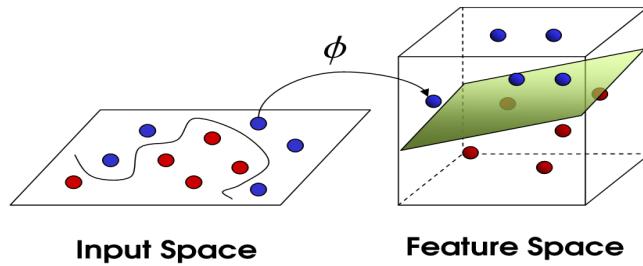
Εικόνα 3.3: Σχηματική απεικόνιση SVM και SVR

Επιπρόσθετα μέσω των μεθόδων συναρτήσεων πυρήνα (kernel function), τα SVM μπορούν να μετασχηματίσουν τον αρχικό χώρο υποθέσεων έτσι ώστε μη γραμμικώς διαχωρίσιμα προβλήματα να μετατραπούν σε γραμμικώς διαχωρίσιμα. Μετά τον μετασχηματισμό, το υπερεπίπεδο που έχει προκύψει μπορεί να προβληθεί στον αρχικό χώρο των χαρακτηριστικών παράγοντας το ζητούμενο μη γραμμικό όριο απόφασης. Η συνάρτηση πυρήνα ορίζεται ως εξής:

$$\Phi(u) * \Phi(v) = f(u * v)$$

όπου Φ είναι μετασχηματισμός στα δεδομένα ώστε το ζητούμενο μη γραμμικό σύνορο στον αρχικό χώρο να καταστεί γραμμικό στο μετασχηματισμένο.

Η συνάρτηση f ονομάζεται συνάρτηση πυρήνα και συμβολίζεται με $K(u, v)$ και υποστηρίζει ότι για δύο διανύσματα του αρχικού χώρου θα πρέπει να είναι ισοδύναμη με το εσωτερικό γινόμενο των διανυσμάτων στον μετασχηματισμένο χώρο. Ένα αξιοσημείωτο χαρακτηριστικό των SVM είναι ότι η υπολογιστική πολυπλοκότητα είναι ανεξάρτητη από την διάσταση του χώρου του πυρήνα, όπου απεικονίζεται ο χώρος χαρακτηριστικών εισόδου.



Εικόνα 3.4: Μετασχηματισμός του χώρου προβλήματος σε ανώτερη διάσταση έτσι ώστε το πρόβλημα από μη γραμμικό να μετατραπεί σε γραμμικό με την χρήση της συνάρτησης πυρήνα

3.3 Δέντρα απόφασης

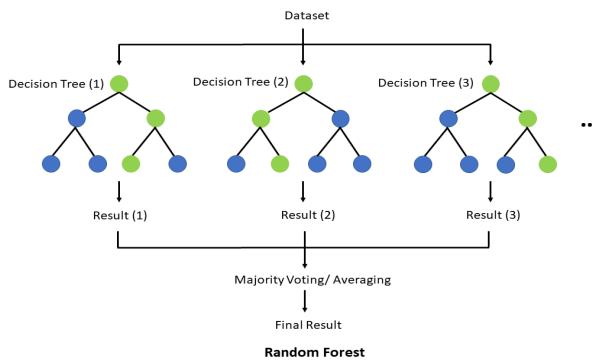
Μια μεγάλη κατηγορία μη γραμμικών ταξινομητών είναι τα δέντρα απόφασης (decision trees). Η ρίζα ενός δέντρου είναι ο κόμβος που δεν έχει εισερχόμενες ακμές αλλά μπορεί να έχει μηδέν ή πιο πολλές εξερχόμενες, ο εσωτερικός κόμβος έχει μια εισερχόμενη ακμή και δύναται να έχει δύο ή παραπάνω ακμές και τα φύλλα ή οι τερματικοί κόμβοι, όπου κάθε κόμβος έχει ακριβώς μια εισερχόμενη ακμή και καμία εξερχόμενη ακμή. Κάθε κόμβος στο δέντρο ορίζει μια συνθήκη ελέγχου της τιμής κάποιου χαρακτηριστικού (attribute) των στιγμιότυπων και κάθε κλαδί που φεύγει από τον κόμβο αυτό αντιστοιχεί σε μια διαφορετική τιμή του χαρακτηριστικού αυτού. Μια περίπτωση ταξινομείται αρχίζοντας από την ρίζα και ακολουθώντας τα κλαδιά του δέντρου προς κάποιο φύλλο, το οποίο περιέχει και μια διακριτή τιμή της κατηγορίας. Σε κάθε κόμβο ελέγχεται η τιμή της περίπτωσης για το χαρακτηριστικό του κόμβου και ακολουθείται το κλαδί.

Κάθε διαίρεση ενός κόμβου πρέπει να παράγει υποσύνολα τα οποία είναι ομοιογενής προς την κλάση τους σε σύγκριση με το υποσύνολο του προγόνου κόμβου τους (αμιγή) και πραγματοποιείται με την χρήση ενός μέτρου μη καθαρότητας του κόμβου (πχ Gini, εντροπία). Το χαρακτηριστικό ως προς το οποίο διαιρείται ένας κόμβος επιλέγεται ως το καλύτερο από ένα σύνολο F τυχαίων επιλεγμένων χαρακτηριστικών (όπου F μια παράμετρος). Αξίζει επίσης να αναφερθεί ότι οι περισσότεροι αλγόριθμοι μάθησης δέντρων ταξινόμησης είναι παραλλαγές ενός βασικού αλγορίθμου που κάνει μη πλήρη αναζήτηση στο χώρο των πιθανών δέντρων απόφασης, χτίζοντας το υποψήφιο

26 Κεφάλαιο 3 Θεωρητικό υπόβαθρο

δέντρο από την ρίζα προς τα φύλλα και άπληστα επιλέγοντας κάθε φορά ως παράμετρο διακλάδωσης το καλύτερο τοπικό χαρακτηριστικό.

Το τυχαίο δάσος (Random Forest) είναι μια μέθοδος συλλογικής μάθησης (ensemble learning) που κατασκευάζει πολλά δέντρα και λαμβάνει αποφάσεις βάσει πλειοψηφίας (σε προβλήματα ταξινόμησης) ή βάσει μέσης τιμής (στην παλινδρόμηση). Η βασική ιδέα πίσω από τον αλγόριθμο τυχαίου δάσους είναι ότι ένα μεγάλο πλήθος από μη συσχετιζόμενα δένδρα που αποφασίζουν από κοινού θα παράγουν καλύτερα αποτελέσματα από κάθε δένδρο ξεχωριστά ή από την απλή χρήση ενός ταξινομητή.



Εικόνα 3.5: Σχηματική αναπαράσταση του αλγορίθμου του τυχαίου δάσους

Βασίζεται στην μέθοδο της ενθυλάκωσης (bagging) η οποία χρησιμοποιείται σε αλγορίθμους μάθησης οι οποίοι είναι ασταθείς δηλαδή παρουσιάζουν χαμηλή μεροληφία και μεγάλη διακύμανση αλλά και διαφέρει από αυτήν καθώς σε κάθε εσωτερικό κόμβο ενός δέντρου επιλέγεται το καλύτερο κριτήριο διαχωρισμού ανάμεσα σε ένα μικρό σύνολο τυχαίων επιλεγμένων χαρακτηριστικών. Από τα N δεδομένα του συνόλου εκπαίδευσης φτιάχνονται η υποσύνολα μεγέθους N με τυχαία επιλογή και δειγματοληφία με αντικατάσταση (bootstrapping) ή και ορίζεται ένας αριθμός $m < M$ που διατηρείται σταθερός στην διάρκεια του αλγορίθμου, που καθορίζει το πλήθος των τυχαίων m χαρακτηριστικών από τα M .

3.4 Extreme Gradient Boosting

Ο Extreme Gradient Boosting - XGBoost είναι ένας αλγόριθμος ενίσχυσης (boosting) ενός δέντρου απόφασης και βασίζεται σε μια μέθοδο που ενισχύει ένα αδύναμο μοντέλο μέσω της διαδικασίας βελτιστοποίησης της βαθμωτής

κατάβασης (gradient descent - gradient boosting), παράλληλα ελαχιστοποιώντας μια κατάλληλη συνάρτηση κόστους. Η μέθοδος της ενίσχυσης είναι μια επαναληπτική διαδικασία που στηρίζεται στην ανάθεση βαρών στα δεδομένα εκπαίδευσης έτσι ώστε ο αλγόριθμος μάθησης να επικεντρωθεί σε δεδομένα που είναι δύσκολα στην κατηγοριοποίηση ή πρόβλεψη τους και να πετύχει βελτιωμένη πρόβλεψη όσον αφορά το επόμενο μοντέλο που θα παραχθεί. Αναλυτικότερα στα δεδομένα που το μοντέλο πραγματοποιεί λάθος προβλέψεις αυξάνονται τα βάρη τους ενώ σε εκείνα που προβλέπει ορθά ελαττώνονται εστιάζοντας σε αυτά. Η ενίσχυση επομένως είναι ένας αλγόριθμος που συνδυάζει μοντέλα μηχανικής μάθησης με σκοπό την μείωση της δειγματοληψίας και της διακύμανσης μέσω μιας διαδικασίας ελαχιστοποίησης μιας κυρτής συνάρτησης κόστους.

3.5 Αναδρομικά νευρωνικά δίκτυα

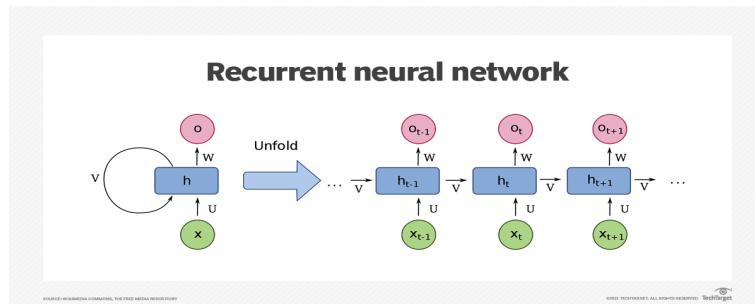
Τα αναδρομικά νευρωνικά δίκτυα RNN (recurrent neural networks) είναι δίκτυα με ανατροφοδότηση, με εσωτερική κατάσταση αναπαριστώμενη σε μορφή διανύσματος που διαμορφώνεται σε κάθε κύκλο λειτουργίας και προωθείται στον επόμενο. Η λειτουργία του RNN εμπεριέχει την έννοια του χρόνου ή της αλληλουχίας και είναι κατάλληλα για προβλήματα που εξελίσσονται στην ροή του χρόνου (όπως δεδομένα κειμένου, χρονοσειρές, οικονομικά δεδομένα, πρόβλεψη καιρού) αφού έχουν εσωτερική μνήμη και μπορούν να θυμούνται σημαντικά γνωρίσματα από την είσοδο και από τα προηγούμενα επίπεδα.

Σε κάθε κύκλο, η εσωτερική κατάσταση συνδυάζεται με το διάνυσμα εισόδου και με την χρήση μιας κατάλληλης συνάρτησης, παράγεται η έξοδος και υπολογίζεται η νέα εσωτερική κατάσταση. Βασικοί παράμετροι του RNN μοντέλου είναι ο αριθμός των κρυφών στρωμάτων, ο αριθμός των νευρώνων σε κάθε τέτοιο κρυφό στρώμα, οι συνδέσεις μεταξύ στρωμάτων και η αρχικοποίηση των βαρών (V , U , W) και οι καταστάσεις. Ο μηχανισμός εκπαίδευσης στα RNN είναι ο αλγόριθμος οπισθοδιάδοσης μέσω χρόνου (back-propagation through time - BPTT) μια παραλλαγή του back-propagation που εστιάζει στην ανάστροφη διάδοση του σφάλματος στα επίπεδα του δικτύου και στους κύκλους λειτουργίας. Αξίζει να αναφερθεί ότι τα RNN μοιράζονται τους ίδιους παραμέτρους (βάρη) επειδή εκτελούν τις ίδιες λειτουργίες σε κάθε χρονική στιγμή με διαφορετικές εισόδους.

Το πρόβλημα της εξασθενούσης παραγώγου (vanishing gradient problem) είναι ένα σημαντικό φαινόμενο που αντιμετωπίζουν τα RNN καθώς η κλίση της συνάρτησης απώλειας (loss function), εξασθενεί εκθετικά με το χρόνο και γίνε-

28 Κεφάλαιο 3 Θεωρητικό υπόβαθρο

ται στην πραγματικότητα ασήμαντη εμποδίζοντας αποτελεσματικά την ενημέρωση των παραμέτρων. Στη χειρότερη περίπτωση, υπάρχει η πιθανότητα να σταματήσει εντελώς η εκπαίδευση του δικτύου. Το αντίστροφο πρόβλημα που εμφανίζεται και αυτό στα RNN είναι το πρόβλημα του exploding gradient, που παρουσιάζεται όταν η κλίση γίνεται πολύ μεγάλη και προσθέτει αστάθεια στο μοντέλο. Σε αυτήν την περίπτωση, συσσωρεύονται μεγαλύτερες διαβαθμίσεις σφάλματος και τα βάρη του μοντέλου γίνονται πολύ μεγάλα. Αυτό το ζήτημα μπορεί να προκαλέσει μεγαλύτερους χρόνους εκπαίδευσης και κακή απόδοση μοντέλου.



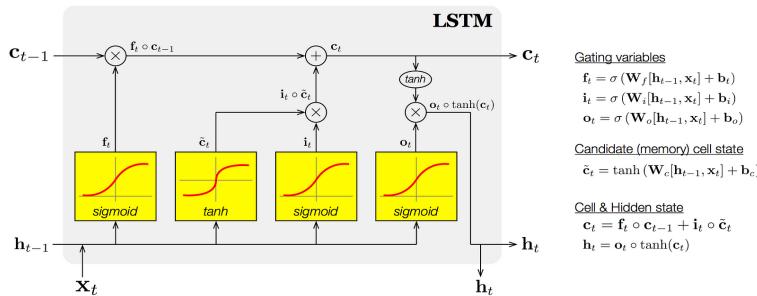
Εικόνα 3.6: Βασική δομή RNN και ξεδίπλωμα στο χρόνο όπου U, W, V τα βάρη του δικτύου, x η είσοδος στο δίκτυο, h η κρυφή κατάσταση και o η έξοδος την χρονική στιγμή t

3.6 Δίκτυα μακράς βραχύχρονης μνήμης

Τα δίκτυα μακράς βραχύχρονης μνήμης LSTM (Long short – term Memory) είναι μια παραλλαγή των RNN και είναι ένα μοντέλο που διαθέτει βραχυπρόθεσμη μνήμη, μια μορφή μνήμης που μπορεί να διαρκέσει για μεγάλο χρονικό διάστημα. Το LSTM είναι ένα ανατροφοδοτούμενο δίκτυο για την αντιμετώπιση των καίριων προβλημάτων RNN και κυρίως αυτό της εξασθενούσης παραγώγου. Ακριβώς όπως το RNN, το LSTM έχει επαναλαμβανόμενες μονάδες, αλλά η δομή του συνίσταται από τέσσερα αλληλοεπιδρώντα στρώματα που επικοινωνούν μεταξύ τους. Μια κοινή μονάδα LSTM αποτελείται από ένα κελί, μια πύλη εισόδου, μια πύλη εξόδου και μια πύλη λήθης. Η μόνη διαφορά τους από τα RNN είναι ο τρόπος που υπολογίζεται η εσωτερική τους κατάσταση.

Βασική έννοια στα LSTM είναι η εσωτερική κατάσταση της μονάδας κελιού (cell state) καθώς είναι η μνήμη του δικτύου που αποθηκεύει μακροπρόθεσμες

πληροφορίες. Το LSTM μπορεί να προσθέσει ή να αφαιρέσει πληροφορίες στην κατάσταση του κελιού (cell state) με την χρήση πυλών. Η επιλογή των πληροφοριών που θα διαγραφούν, θα γραφούν ή θα διαβαστούν ελέγχεται από τις τρεις αντίστοιχες πύλες. Οι πύλες είναι επίπεδα νευρωνικών δικτύων με λογιστική συνάρτηση ενεργοποίησης και είναι δυναμικές καθώς η τιμή τους υπολογίζεται με βάση το τρέχον γενικό πλαίσιο. Τιμές 0 στην έξοδο των πυλών καταστέλλουν (μηδενίζουν) την πληροφορία της αντίστοιχης θέσης του διανύσματος εισόδου ενώ τιμές 1 σημαίνουν την τιμή και επιτρέπουν την διέλευση της. Ενδιάμεσες τιμές επιτρέπουν την διέλευση σε κάποιο βαθμό.



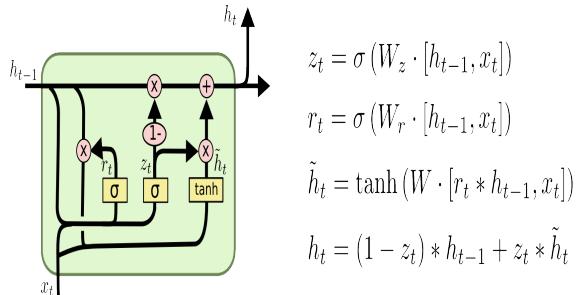
Εικόνα 3.7: Εσωτερική δομής μιας μονάδας δικτύου LSTM

Η πύλη λήθης ελέγχει το πως η προηγούμενη κατάσταση c_{t-1} θα επηρεάσει την νέα κατάσταση c_t δηλαδή τι θα πρέπει να ξεχαστεί από την προηγούμενη κατάσταση, η πύλη εισόδου επιλέγει τι πληροφορία να προστεθεί στο cell state από τον τρέχον κύκλο και τέλος η πύλη εξόδου επιλέγει τι από το cell state θα προκύψει ως πρόβλεψη του τρέχοντος κύκλου.

Σε ένα δίκτυο LSTM, η είσοδος ρέει προς μία κατεύθυνση, είτε προς τα πίσω είτε προς τα εμπρός. Ωστόσο στο Bidirectional LSTM η είσοδος ρέει και προς τις δύο κατευθύνσεις με απώτερο σκοπό την ταυτόχρονη διατήρηση των μελλοντικών και των προηγούμενων πληροφοριών. Χρησιμοποιείται συχνά στην Επεξεργασία Φυσικής Γλώσσας για τη μοντελοποίηση των διαδοχικών εξαρτήσεων μεταξύ λέξεων και φράσεων και στις δύο κατευθύνσεις της ακολουθίας. Ουσιαστικά σε προβλήματα όπου όλα τα χρονικά βήματα της ακολουθίας εισόδου είναι διαθέσιμα, τα αμφίδρομα LSTM εκπαιδεύουν δύο αντί για ένα LSTM στην ακολουθία εισόδου, όπου το πρώτο μοντέλο μαθαίνει την ακολουθία της εισόδου που παρέχεται και το δεύτερο μοντέλο μαθαίνει την αντίστροφη αυτής. Ο συνδυασμός των δύο εκπαιδευμένων μοντέλων γίνεται στο βήμα της συγχώνευσης όπου πραγματοποιείται με τις λειτουργίες του αθροίσματος, του πολλαπλασιασμού, του μέσου όρου και της συνένωσης.

3.7 Gated recurrent unit

Μια παραλλαγή των LSTM θεωρούνται τα gated recurrent unit - GRU με ελαφρώς απλοποιημένη εσωτερική δομή και γρηγορότερη λειτουργία. Σε αντίθεση με το LSTM, το GRU δεν έχει κατάσταση cell state c_t . Έχει μόνο μια κρυφή κατάσταση h_t και λόγω της απλής αρχιτεκτονικής του, η GRU έχει μικρότερο χρόνο εκπαίδευσης σε σύγκριση με τα μοντέλα LSTM. Το GRU χρησιμοποιεί μια πύλη ενημέρωσης και μια πύλη επαναφοράς για να λύσει το πρόβλημα του vanishing gradient. Αυτές οι πύλες αποφασίζουν ποιες πληροφορίες είναι σημαντικές και τις διαβιβάζουν στην έξοδο. Η πύλη ενημέρωσης βοηθά το μοντέλο να προσδιορίσει πόσες από τις πληροφορίες του παρελθόντος (από προηγούμενα χρονικά βήματα) πρέπει να διαβιβαστούν στο μέλλον και η πύλη επαναφοράς χρησιμοποιείται από το μοντέλο για να αποφασίσει πόσες από τις προηγούμενες πληροφορίες θα πρέπει να ξεχάσει. Οι πύλες μπορούν να εκπαιδευτούν ώστε να αποθηκεύουν πληροφορίες από προηγούμενες χρονικές στιγμές, χωρίς να τις απολλύουν ή να αφαιρούνται άσχετες πληροφορίες.

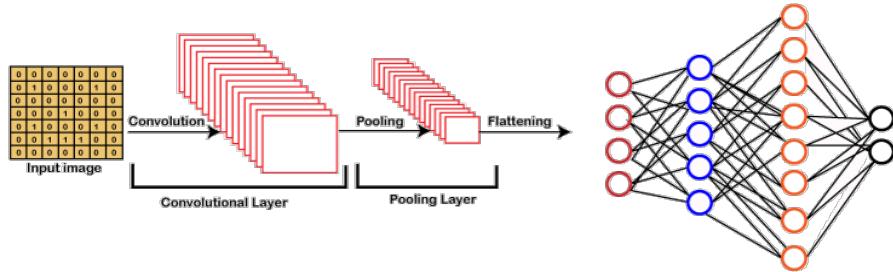


Εικόνα 3.8: Σχηματική αναπαράσταση του GRU δικτύου

3.8 Συνελικτικά νευρωνικά δίκτυα

Τα συνελικτικά νευρωνικά δίκτυα CNN (Convolutional neural network) είναι εμπνευσμένα από τον βιολογικό εγκέφαλο και ιδιαίτερα από το τμήμα του εγκεφάλου που είναι υπεύθυνο για την αντίληψη της όρασης. Το οπτικό σήμα περνά από ένα στάδιο εντοπισμού χαρακτηριστικών λεπτής υφής, τα οποία συνδυαζόμενα συνθέτουν πρότυπα (patterns) περισσότερο αδρής μορφής από τα οποία τελικά αναγνωρίζεται κάποιο αντικείμενο (ταξινόμηση). Ένα CNN περιλαμβάνει τρία είδη σταδίων επεξεργασίας: το στάδιο συνέλιξης (convolution layer), το στάδιο συγκέντρωσης/σμίκρυνσης και το πλήρως διασυνδεδεμένο

στάδιο(fully connected layer). Σε ένα στάδιο επεξεργασίας εκτελείται κάποιο είδος υπολογισμού σε μια ομάδα από στρώματα νευρώνων ή πολλαπλοί υπολογισμοί ίδιας μορφής σε ένα επίπεδο.



Εικόνα 3.9: Σχηματική αναπαράσταση ενός CNN

Το στάδιο της συνέλιξης είναι υπεύθυνο για τον εντοπισμό χαρακτηριστικών λεπτής ή αδρής υφής πάνω στην εικόνα χρησιμοποιώντας κατάλληλα φίλτρα εντοπισμού χαρακτηριστικών. Σε ένα στάδιο συνήθως εφαρμόζονται περισσότερα από ένα φίλτρα παράγοντας αντίστοιχα και περισσότερα από ένα επίπεδα με νευρώνες, που καθένα τροφοδοτείται με δεδομένα μιας υπεριοχής του προηγούμενου σταδίου. Ο πίνακας της εικόνας που δίνεται ως είσοδος πολλαπλασιάζεται με έναν πίνακα χαρακτηριστικών για την εξαγωγή των βασικών χαρακτηριστικών της εικόνας και έπειτα ο πίνακας του φίλτρου ολισθαίνει μέσω της μήτρας εισόδου και εξάγει τις πληροφορίες από τις μικρότερες υποενότητες της εικόνας. Αξίζει να επισημανθεί ότι οι παράμετροι/βάρη και οι τιμές πόλωσης που ορίζονται σε κάθε φίλτρο δεν είναι προκαθορισμένες αλλά προκύπτουν μέσω της διαδικασίας εκπαίδευσης του CNN με τον αλγόριθμο back-propagation και είναι ίδιες καθόλη την διάρκεια της εκπαίδευσης του δικτύου.

Στο στάδιο σμίκρυνσης μικραίνουν οι διαστάσεις των επιπέδων, καθώς το δίκτυο απαλλάσσεται από χαρακτηριστικά που δεν είναι τόσο σημαντικά ως προς την εκπαίδευση του δικτύου ενώ παράλληλα μειώνονται τα βάρη. Αναλυτικότερα ο προκύπτων πίνακας πολλαπλασιάζεται με τον πίνακα σμίκρυνσης που συνήθως εξάγει τις μέγιστες ή μέσες τιμές από τις μικρές υποενότητες, με αποτέλεσμα η εικόνα στην οποία έχει εφαρμοστεί σμίκρυνση να συμπιέζεται σημαντικά σε μέγεθος σε σχέση με τον πίνακα της αρχικής εικόνας. Το δίκτυο με αυτόν τον τρόπο γίνεται πιο ανθεκτικό σε αλλοιώσεις δεδομένων εισόδου, αποφεύγει την υπερπροσαρμογή και χρησιμοποιείται χαμηλή υπολογιστική ισχύ για την επεξεργασία των δεδομένων μειώνοντας τις διαστάσεις

32 Κεφάλαιο 3 Θεωρητικό υπόβαθρο

τους. Τα είδη των υπολογισμών σμίκρυνσης είναι το μέγιστο (max pooling) και ο μέσος όρος (average pooling).

Τέλος το στάδιο των πλήρως διασυνδεμένων δικτύων αποτελεί την έξοδο του δικτύου και στο τελευταίο στάδιο επεξεργασίας υλοποιείται ο εκάστοτε ταξινομητής ή παλινδρομητής (regressor). Ο όρος πλήρως συνδεδεμένος υποδηλώνει ότι όλες οι είσοδοι ή οι κόμβοι από ένα επίπεδο συνδέονται με κάθε μονάδα ενεργοποίησης ή κόμβο του επόμενου επιπέδου.

4. Πηγές δεδομένων

Στην παρούσα διπλωματική χρησιμοποιήθηκαν 2 πήγες δεδομένων σχετικά με την κινητικότητα των χρηστών κατά την διάρκεια της πανδημίας του κορωνοϊού και πιο συγκεκριμένα τα dataset της Apple [Apple Inc.], της Google [Google LLC] και οι πηγές δεδομένων iMEdD-Lab [iMEdD-Lab], Sandbird COVID19-Greece [Sandbird] και COVID-19 Response Greece [Covid-19 Response Greece] τα οποία υποστηρίζονται μέσω κατάλληλων repositories του GitHub και παρέχουν χρήσιμες πληροφορίες σχετικά με τα κρούσματα, τους εμβολιασμούς, τους θανάτους και τα άτομα που νοσηλεύονται σε μονάδες εντατικής θεραπείας στην Ελλάδα. Η βάση δεδομένων που χρησιμοποιείται στην εργασία αναπτύχθηκε σε PostgreSQL και αποτελείται από δέκα πίνακες οι οποίοι ενημερώνονται αυτόματα μέσω κατάλληλου κώδικα Python. Ιδιαίτερα σημαντικό να αναφερθεί είναι ότι τα δεδομένα που χρησιμοποιήθηκαν και επιλέχθηκαν αφορούν μόνο την γεωγραφική περιοχή της Ελλάδας.

1. Apple COVID-19 Mobility Trends Reports

Οι δείκτες κινητικότητας της Apple είναι διαθέσιμοι σε εθνικό και περιφερειακό επίπεδο για την εκάστοτε χώρα και έχουν προκύψει σύμφωνα με τις οδηγίες και τις κατευθύνσεις που λαμβάνουν οι χρήστες μέσω της αλληλεπίδρασης τους με την εφαρμογή Apple Maps. Για την διασφάλιση της ιδιωτικότητας τυχαία μεταβαλλόμενα αναγνωριστικά αποστέλλονται στην υπηρεσία έτσι ώστε να μην καταγραφούν οι αναζητήσεις των χρήστων. Η διαθεσιμότητα δεδομένων σε μια συγκεκριμένη πόλη, χώρα ή περιοχή εξαρτάται από διάφορους παράγοντες λαμβάνοντας υπόψη τα ελάχιστα όρια για τα αιτήματα λήψης οδηγιών που πραγματοποιούνται ανά ημέρα. Οι δείκτες αφορούν τις δραστηριότητες της οδήγησης και του περπατήματος κατά την διάρκεια της πανδημίας, οι οποίοι παρουσιάζουν μια κατακόρυφη πτώση παγκόσμιως σε επίπεδο 50% έως 90% σε σχέση με την περίοδο πριν την έλευση του ιού. Η Ελλάδα σύμφωνα με τα επίσημα στοιχεία που δημοσίευσε η εταιρεία βρέθηκε να είναι τέταρτη ανάμεσα

34 Κεφάλαιο 4 Πηγές δεδομένων

σε 60 χώρες ανά τον κόσμο με την ελάχιστη μεταβολή της ανθρώπινης κινητικότητας των πολιτών της καθόλη διάρκεια της πανδημίας με τα Ηνωμένα Αραβικά Εμιράτα, την Κροατία και την Πορτογαλία μόνο να προηγούνται [Apple Inc. 2020].

2. Google COVID-19 Community Mobility Reports

Οι Αναφορές κινητικότητας της Google εστιάζουν στην παροχή μιας γενικότερης εικόνας των μεταβολών της ανθρώπινης κινητικότητας των πολιτών μιας χώρας στην διάρκεια της πανδημίας σύμφωνα με τα υγειονομικά μέτρα που λαμβάνονται από την εκάστοτε κυβέρνηση για την καταπολέμηση του COVID-19. Οι αναφορές παρουσιάζουν με την μορφή γραφήματος τις τάσεις στις μετακινήσεις των πολιτών με την πάροδο του χρόνου ανά γεωγραφική περιοχή, σε διαφορετικές κατηγορίες μερών, όπως καταστήματα λιανικής πώλησης και φυχαγωγίας, παντοπωλεία και φαρμακεία, πάρκα, σταθμούς δημόσιας συγκοινωνίας, χώρους εργασίας και κατοικίες.

Οι πληροφορίες και οι τιμές των δεικτών στο σύνολο δεδομένων της Google δημιουργήθηκαν με συγκεντρωτικά, ανώνυμα δεδομένα των χρηστών που είχαν ενεργοποιήσει τη ρύθμιση του Ιστορικού τοποθεσίας στους Χάρτες της Google του λογαριασμού τους, η οποία είναι απενεργοποιημένη εξαρχής (παρόμοιος τρόπος υπολογισμού όπως η εμφάνιση των δημοφιλών ωρών για μέρη στους χάρτες Google). Τα δεδομένα συμπεριλαμβάνονται στον υπολογισμό ανάλογα με τις ρυθμίσεις χρήστη, τη συνδεσιμότητα και το κατά πόσο τα δεδομένα πληρούν τα όρια απορρήτου που έχει θεσπίσει η εταιρεία, καθώς όταν δεν υπάρχουν σημαντικά στατιστικά δεδομένα για μια περιοχή προκειμένου να διασφαλιστεί η ανωνυμία δεν εμφανίζονται ημερήσιες αλλαγές στις κατηγορίες της Google. Είναι ιδιαίτερα σημαντικό να αναφερθεί ότι δεν χρησιμοποιήθηκαν στους υπολογισμούς προσωπικά στοιχεία, όπως η τοποθεσία, οι επαφές ή οι μετακινήσεις των χρηστών.

Οι αναφορές καταγράφουν τις μεταβολές της κινητικότητας στα μέρη επίσκεψης και την διάρκεια παραμονής σε αυτά. Η τιμή ενός δείκτη κινητικότητας Google προκύπτει από την συγχρίση της κινητικότητας για την ημερομηνία της αναφοράς με την ημέρα που αποτελεί το σημείο αναφοράς και εκφράζεται με θετικό ή αρνητικό ποσοστό. Ένα σημείο αναφοράς ορίζεται ως η μέση τιμή για την αντίστοιχη ημέρα της εβδομάδας κατά τη διάρκεια μιας περιόδου 5 εβδομάδων, από τις 3 Ιανουαρίου έως 6 Φεβρουαρίου 2020, ενώ μια ημερομηνία αναφοράς ως η ποσοστιαία αλλαγή για την τελευταία ημέρα του συνόλου δεδομένων. Τέλος το γράφημα

τάσεων αναπαριστά τις ποσοστιαίες αλλαγές στις 6 εβδομάδες πριν από την ημερομηνία αναφοράς [Google LLC 2023].

Σύμφωνα με την τελευταία αναφορά κινητικότητας της Google για την Ελλάδα στις 15 Οκτωβρίου 2022, η κινητικότητα παρέμεινε σταθερή με ελάχιστες μεταβολές σε σύγκριση με το σημείο αναφοράς (πριν την πανδημία) όσον αφορά τους χώρους εργασίας (0%), τις κατοικίες (+1%) και την λιανική και ψυχαγωγία (-1%) ενώ παρουσίασε αξιοσημείωτη αύξηση στις δημόσιες συγκοινωνίες, στις αγορές τροφίμων (+39%) και στα πάρκα (+50%) [Google LLC 2022].

3. Sanbird COVID-19 Greece

Σκοπός αυτού του αποθετηρίου είναι η δημοσίευση περιφερειακών στατιστικών για τα κρούσματα COVID-19 στην Ελλάδα στην αρχή σε καθημερινή βάση (μέχρι τις 10/7/2022) και έπειτα εβδομαδιαία. Από αυτήν την πηγή δεδομένων επιλέχθηκαν τα αρχεία demography_men_details.csv και demography_women_details.csv τα οποία συνενώθηκαν σε ένα αρχείο και παρέχουν πληροφορίες για τα κρούσματα, τους θανάτους και τους ασθενείς που βρίσκονται σε μονάδες εντατικής θεραπείας σύμφωνα με την ηλικιακή κατανομή και το φύλο.

4. iMEdD-Lab

Τα σύνολα δεδομένων και τα γραφήματα αυτού του Github repository δημιουργήθηκαν από την iMEdD, ως μέρος της ανάπτυξης της διαδικτυακής εφαρμογής του iMEdD Lab που παρακολουθεί την εξάπλωση του COVID-19 στην Ελλάδα και σε όλο τον κόσμο. Από αυτήν την πηγή δεδομένων επιλέχθηκε το αρχείο greeceTimeline.csv και συγκεκριμένα οι εγγραφές που αφορούν τα κρούσματα και τους θανάτους από COVID-19 στην Ελλάδα τόσο ημερησίως όσο και αθροιστικά ανά ημερομηνία.

5. COVID-19-Response Greece

Το συγκεκριμένο repository παρέχει πηγές δεδομένων σχετικά με την έξαρση του COVID-19 στην Ελλάδα. Στόχοι της συγκεκριμένης κοινότητας με την δημιουργία αυτού του αποθετηρίου είναι η διεξαγωγή ανάλυσης δεδομένων, η δημιουργία οπτικοποιήσεων και η ανάπτυξη εργαλείων παρακολούθησης της πορείας του ιού. Από αυτήν την πηγή δεδομένων επιλέχθηκαν τα αρχεία intensive_care_cases.json, cumulative_per_area_vaccinations.json, regions_cumulative.csv, regions_history_cases.csv και vaccinations_data_history.json που αφορούν αντίστοιχα τα άτομα που βρίσκονται σε μονάδα εντατικής θεραπείας ανά ημερομηνία, τον αριθμό των εμβολιασμών που έχουν γίνει ανά περιοχή της Ελλάδας, τον

36 Κεφάλαιο 4 Πηγές δεδομένων

συνολικό αριθμό κρουσμάτων ανά νομό, τον αριθμό συνολικό κρουσμάτων ανά νομό και ανά ημερομηνία κατά την διάρκεια της πανδημίας και τέλος τον αριθμό των συνολικών εμβολιασμών ανά περιοχή της Ελλάδας και ανά ημερομηνία κατά την διάρκεια της πανδημίας. Ως γεωγραφική περιοχή στα σύνολα δεδομένων έπειτα από κατάλληλη επεξεργασία όπου απαιτείται, ορίζεται ένας νομός της Ελλάδας (σύνολο 52 νομοί συμπεριλαμβανομένου του Αγίου Όρους).

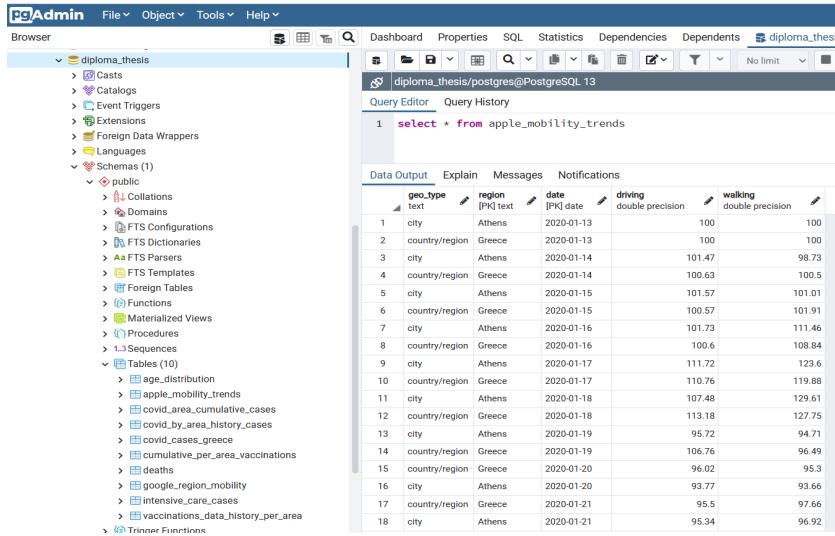
The screenshot shows a GitHub repository page for 'Covid-19-Response-Greece/covid19-data-greece'. The file 'intensive_care_cases.json' is displayed. The code content is as follows:

```
1 {
2   "cases": [
3     {
4       "date": "2020-02-26",
5       "intensive_care": 0
6     },
7     {
8       "date": "2020-02-27",
9       "intensive_care": 0
10    },
11    {
12      "date": "2020-02-28",
13      "intensive_care": 0
14    },
15    {
16      "date": "2020-02-29",
17      "intensive_care": 0
18    }
19 }
```

Εικόνα 4.1: Στιγμιότυπο δεδομένων του αρχείου intensive_care_cases.json στο σύνολο δεδομένων COVID-19-Response Greece

Η βάση αποτελείται από 10 πίνακες:

- **apple_mobility_trends** : Αποτελείται από τα πεδία γεωγραφικός τύπος π.χ. πόλη ή χωρα (geo_type - text), περιοχή (region -text), ημερομηνία (date - date), οδήγηση (driving - double) και περπατήμα (walking - double). Πρωτεύον κλειδί είναι ο συνδυασμός του region με το date.



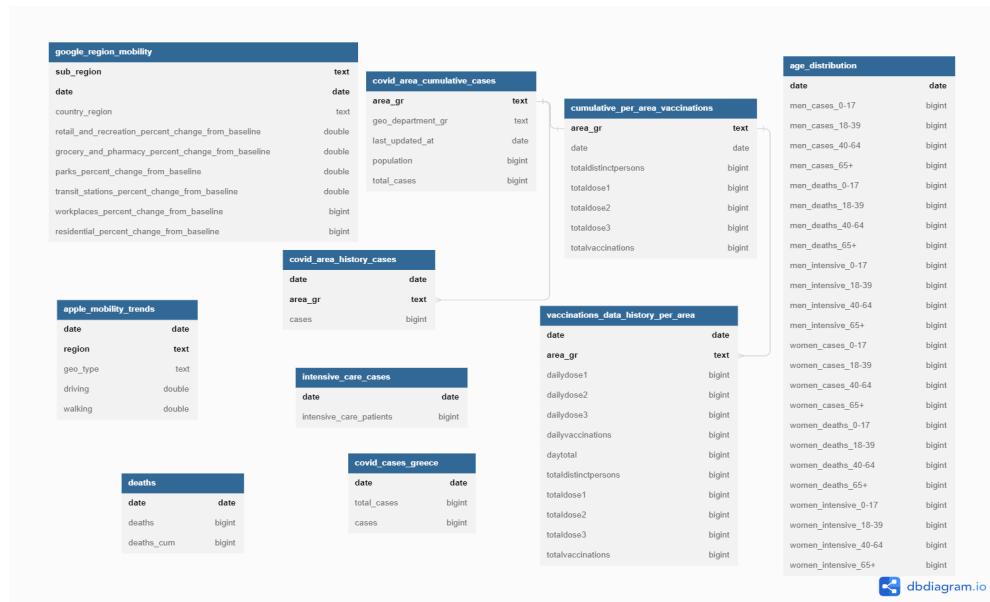
Εικόνα 4.2: Στιγμιότυπο δεδομένων του πίνακα apple_mobility_trends

- **cumulative_per_area_vaccinations** : Αποτελείται από τα πεδία νομός (area_gr - text), ημερομηνία (date - date), μοναδικός αριθμός ατόμων που έκαναν το εμβόλιο (totaldistinctpersons - bigint), συνολικοί εμβολιασμοί με πρώτη δόση (totaldose1 - bigint), συνολικοί εμβολιασμοί με δεύτερη δόση (totaldose2 - bigint), συνολικοί εμβολιασμοί με τρίτη δόση (totaldose3 - bigint) και συνολικοί εμβολιασμοί (totalvaccinations - bigint). Πρωτεύον κλειδί είναι το area_gr.
- **google_region_mobility** : Αποτελείται από τα πεδία χώρα (country_region - text), αποκεντρωμένη διοίκηση (sub_region - text), ημερομηνία (date - date) και δείκτες κινητικότητας στην λιανική και φυχαγωία (retail_and_recreation_percent_change_from_baseline - double), στις αγορές τροφίμων και φαρμακεία (grocery_and_pharmacy_percent_change_from_baseline - double), στα πάρκα (parks_percent_change_from_baseline - double), στις δημόσιες συγκοινωνίες (transit_stations_percent_change_from_baseline - double), στους χώρους εργασίας (workplaces_percent_change_from_baseline - bigint) και στις κατοικίες (residential_percent_change_from_baseline - bigint). Πρωτεύον κλειδί είναι ο συνδυασμός του sub_region με το date.
- **intensive_care_cases**: Αποτελείται από τα πεδία ημερομηνία (date - date) και αριθμός ασθενών σε ΜΕΘ (intensive_care_patients - bigint). Πρωτεύον κλειδί το date.

38 Κεφάλαιο 4 Πηγές δεδομένων

- covid_area_cumulative_cases: Αποτελείται από τα πεδία νομός (area_gr - text), γεωγραφική περιοχή στην οποία βρίσκεται ο νομός (geo_department_gr - text), τελευταία ενημέρωση της βάσης (last_updated_at - text), αριθμός πληθυσμού (population - bigint) και αριθμός συνολικών κρουσμάτων (totalcases - bigint). Πρωτεύον κλειδί είναι το area_gr.
- age_distribution: Αποτελείται από τα πεδία ημερομηνία (date - date), κρούσματα σε άνδρες σε ηλικίες 0-17 (men_cases_0-17 - bigint), 18-39 (men_cases_18-39 - bigint), 40-64 (men_cases_40-64 - bigint), 65+ (men_cases_65+ - bigint), θάνατοι ανδρών σε ηλικίες 0-17 (men_deaths_0-17 - bigint), 18-39 (men_deaths_18-39 - bigint), 40-64 (men_deaths_40-64 - bigint), 65+ (men_deaths_65+ - bigint), άνδρες ασθενείς σε ΜΕΘ σε ηλικίες 0-17 (men_intensive_0-17 - bigint), 18-39 (men_intensive_18-39 - bigint), 40-64 (men_intensive_40-64 - bigint), 65+ (men_intensive_65+ - bigint), και τα κρούσματα σε γυναίκες σε ηλικίες 0-17 (women_cases_0-17 - bigint), 18-39 (women_cases_18-39 - bigint), 40-64 (women_cases_40-64 - bigint), 65+ (women_cases_65+ - bigint), θάνατοι γυναικών σε ηλικίες 0-17 (women_deaths_0-17 - bigint), 18-39 (women_deaths_18-39 - bigint), 40-64 (women_deaths_40-64 - bigint), 65+ (women_deaths_65+ - bigint), γυναίκες ασθενείς σε ΜΕΘ σε ηλικίες 0-17 (women_intensive_0-17 - bigint), 18-39 (women_intensive_18-39 - bigint), 40-64 (women_intensive_40-64 - bigint) και 65+ (women_intensive_65+ - bigint). Πρωτεύον κλειδί είναι το date.
- vaccinations_data_history_per_area: Αποτελείται από τα πεδία νόμος (area_gr - text), ημερομηνία (date - date), ημερήσιοι εμβολιασμοί με πρώτη δόση (dailydose1 - bigint), ημερήσιοι εμβολιασμοί με δεύτερη δόση (dailydose2 - bigint), ημερήσιοι εμβολιασμοί με αναμνηστική δόση (dailydose3 - bigint), ημερήσιοι εμβολιασμοί (dailyvaccinations - bigint), ημερήσια διαφόρα εμβολιασμών που πραγματοποιήθηκαν (daydiff - bigint), αριθμός εμβολιασμένων (totaldistinctpersons - bigint), συνολικοί εμβολιασμοί με πρώτη δόση (totaldose1 - bigint), συνολικοί εμβολιασμοί με δεύτερη δόση (totaldose2 - bigint), συνολικοί εμβολιασμοί με αναμνηστική δόση (totaldose3 - bigint) και συνολικοί εμβολιασμοί (totalvaccinations - bigint). Πρωτεύον κλειδί ο συνδυασμός του (area_gr, date).
- covid_area_history_cases: Αποτελείται από τα πεδία νομός της Ελλάδας (area_gr - text), ημερομηνία (date - date) και κρούσματα (cases - bigint). Πρωτεύον κλειδί ο συνδυασμός του (area_gr, date).

- covid_cases_greece: Αποτελείται από τα πεδία ημερομηνία (date - date), αριθμός ημερήσιων κρουσμάτων (cases - bigint) και συνολικών κρουσμάτων (total_cases - bigint). Πρωτεύον κλειδί είναι το date.
- deaths: Αποτελείται από τα πεδία αριθμός καθημερινών θανάτων (deaths - bigint), ημερομηνία (date - date) και αριθμός συνολικών θανάτων (deaths_cum - bigint). Πρωτεύον κλειδί είναι το date.



Εικόνα 4.3: Εικόνα της βάσης με τους αντίστοιχους πίνακες

5. Προεπεξεργασία

Η προετοιμασία και επεξεργασία των δεδομένων αποτελεί σημαντικό βήμα οποιαδήποτε εργασίας Αναλυτικής Δεδομένων. Ειδικότερα, το πρόβλημα των ελλειπών τιμών αποτελεί φαινόμενο που αποδίδεται είτε σε σφάλματα συστήματος είτε σε ανθρώπινο παράγοντα. Αναλυτικότερα, στο dataset της Google ελλειπές τιμές στην στήλη του «sub-region1» συμπληρώνονται με την κατάλληλη υποπεριοχή της εγγραφής με βάσει το place id της, και στις κατηγορίες «retail and recreation», «grocery and pharmacy», «parks και transit stations» που εμφανίζουν κενές τιμές με την χρήση της γραμμικής παρεμβολής. Με μηδενικά συμπληρώνονται τα χαρακτηριστικά των πανελλήνιων κρούσματων και των θανάτων που εμφανίζουν NaN ή missing values που βρίσκονται χρόνια στην αρχή της πανδημίας στις διάφορες περιοχές της χώρας covid_cases_greece, covid_by_area_history_cases. Με γραμμική παρεμβολή συμπληρώνονται οι υπόλοιπες τιμές που λείπουν στα διάφορα datasets.

Στην συνέχεια αφαιρούνται αν υπάρχουν διπλότυπες εγγραφές στα σύνολα δεδομένων βάσει όλων των columns κρατώντας κάθε φόρα την πρώτη παρατήρηση. Έπειτα γίνεται προσπάθεια αντιμετώπισης του προβλήματος των ακραίων τιμών που μπορεί να έχουν προκύψει από θόρυβο, σφάλματα μετάδοσης ή και από περιορισμούς της τεχνολογίας. Για τον εντοπισμό των ακραίων τιμών γίνεται η χρήση του Inter-quartile range στα κατάλληλα numeric features (τιμές του χαρακτηριστικού που είναι μικρότερες του $Q1 - 1.5 * IQR$ και μεγαλύτερες του $Q3 + 1.5 * IQR$ θεωρούνται ακραίες τιμές) και παρουσιάζονται τρόποι αντικατάστασης τόσο των ακραίων τιμών όσο και των ελλειπών τιμών με τον median Q2, rolling window ή με Iterative και KNN imputation παρόλο που δεν χρησιμοποιούνται εν τέλει στα επόμενα βήματα της προεπεξεργασίας.

Τέλος για την κανονικοποίηση των δεδομένων χρησιμοποιήθηκε ο StandardScaler κάθως στα διάφορα πειράματα απέδωσε καλύτερα αποτελέσματα σε σχέση με άλλες τεχνικές (π.χ. MinMaxScaler, RobustScaler). Στόχος της κανονι-

κοποίησης είναι να αντιμετωπίσει το πρόβλημα της επικράτησης ενός χαρακτηριστικού με μεγάλες τιμές στους διάφορους υπολογισμούς όταν γίνεται συνδυασμός πολλαπλών χαρακτηριστικών τα οποία διαφέρουν στον τρόπο μέτρησης τους, με την μετατροπή των αριθμητικών δεδομένων σε ένα συγκεκριμένο εύρος τιμών.

Η τεχνική Standardization αποτελεί βασική απαίτηση για αρκετούς αλγορίθμους μηχανικής μάθησης καθώς μπορεί να έχουν χαμήλη απόδοση όσον αφορά τα αποτελέσματα τους εάν τα μεμονωμένα χαρακτηριστικά τους δεν έχουν την μορφή της τυπικής κατανομής. Ουσιαστικά η μέθοδος της z-score πραγματοποιεί μετασχηματισμό των αριθμητικών τιμών αφαιρώντας την μέση τιμή και κανονικοποίηση σε unit variance.

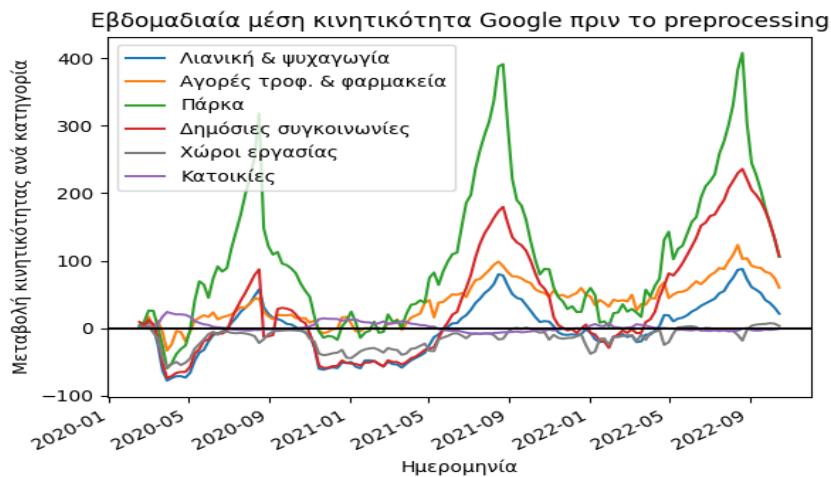
$$x' = \frac{x - \mu}{\sigma}$$

όπου x η αρχική τιμή της μεταβλητής και x' η νέα κανονικοποιημένη τιμή. Η τυπική απόκλιση υπολογίζεται

$$\sigma \cdot \sqrt{\sum_{n=1}^{\infty} \frac{(x_i - \mu)^2}{n}}$$

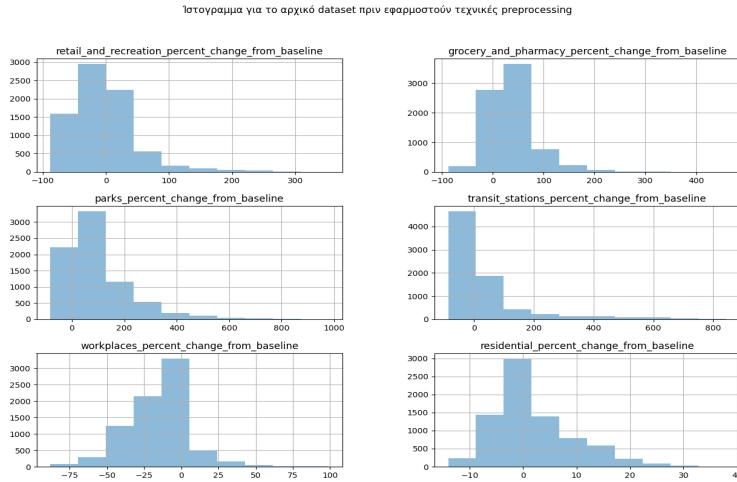
6. Στατιστική ανάλυση

6.1 Google COVID-19 Community Mobility Reports

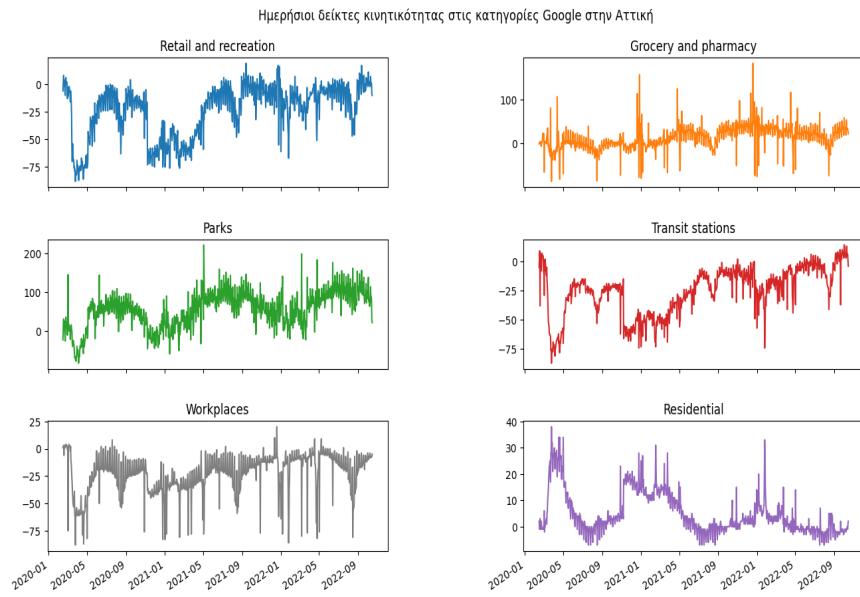


Εικόνα 6.1: Αναπαράσταση των εβδομαδιαίων δεικτών μέσης κινητικότητας των Ελλήνων πολιτών κατά την διάρκεια της πανδημίας στις κατηγορίες τοποθεσίας της Google στο ίδιο γράφημα. Ο δείκτης κινητικότητας απεικονίζεται με κόκκινο χρώμα στις δημόσιες συγκοινωνίες, με μπλε στα καταστήματα λιανικής και φυχαγωγίας, με πορτοκαλί στις αγορές τροφίμων και στα φαρμακεία, με γκρι στους χώρους εργασίας, με πράσινο στα πάρκα και με μωβ στις κατοικίες

6.1 Google COVID-19 Community Mobility Reports 43

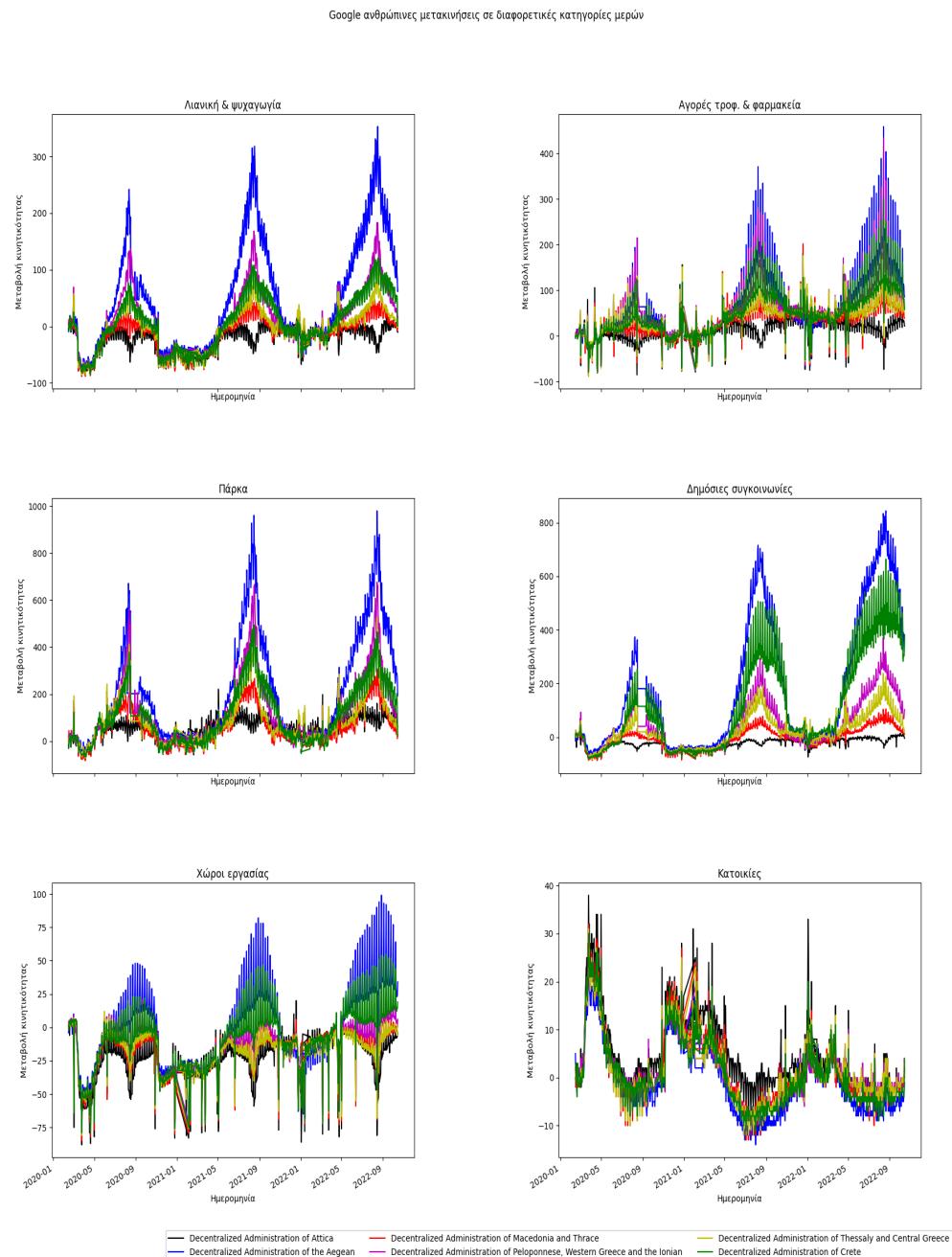


Εικόνα 6.2: Ιστόγραμμα για τους έξι δείκτες κινητικότητας των χρηστών της Google πριν την εφαρμογή της προεπεξεργασίας



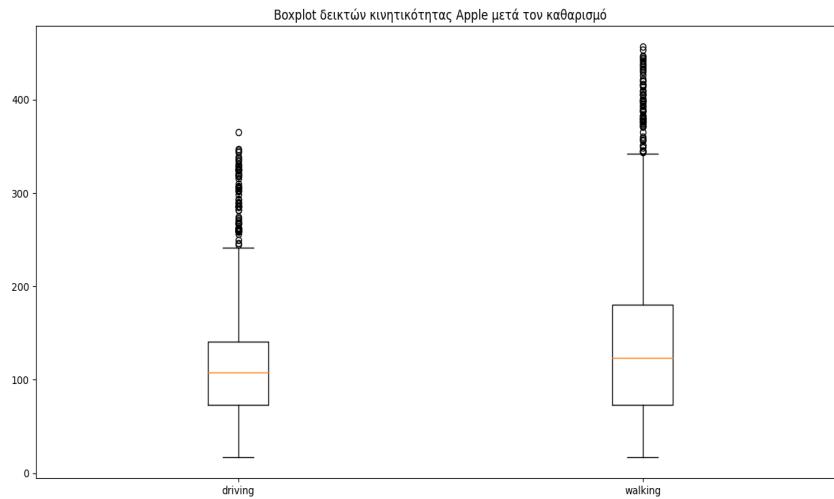
Εικόνα 6.3: Ημερήσια κινητικότητα των χρηστών της Google στην Αττική ανά τοποθεσία

44 Κεφάλαιο 6 Στατιστική ανάλυση

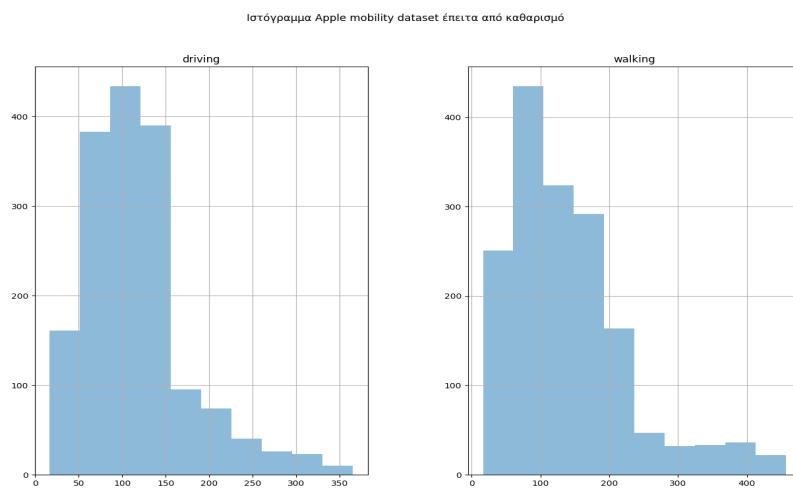


Εικόνα 6.4: Αναπαράσταση στο ίδιο γράφημα των ημερήσιων δεικτών κινητικότητας των πολιτών κατά την διάρκεια της πανδημίας ανά τοποθεσία/δραστηριότητα. Η κινητικότητα απεικονίζεται με κόκκινο χρώμα στην αποκεντρωμένη διοίκηση της Μακεδονίας και Θράκης, με μπλε στην αποκεντρωμένη διοίκηση του Αιγαίου, με πορτοκαλί στην αποκεντρωμένη διοίκηση της Μακεδονίας και Θράκης, με μαύρο στην αποκεντρωμένη διοίκηση της Αττικής, με πράσινο στην αποκεντρωμένη διοίκηση της Κρήτης και με λαχανί στην αποκεντρωμένη διοίκηση της Θεσσαλίας και κεντρικής Ελλάδας

6.2 Apple COVID-19 Mobility Trends Reports



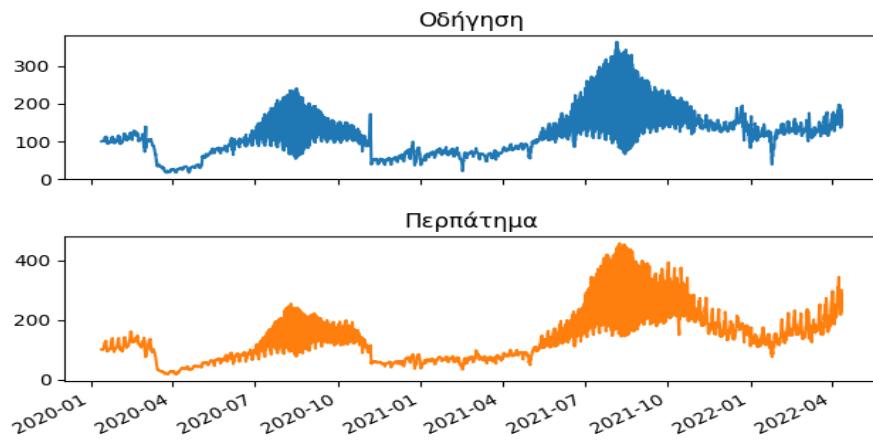
Εικόνα 6.5: Boxplot για τους δείκτες οδήγησης και περπατήματος μετά το preprocessing



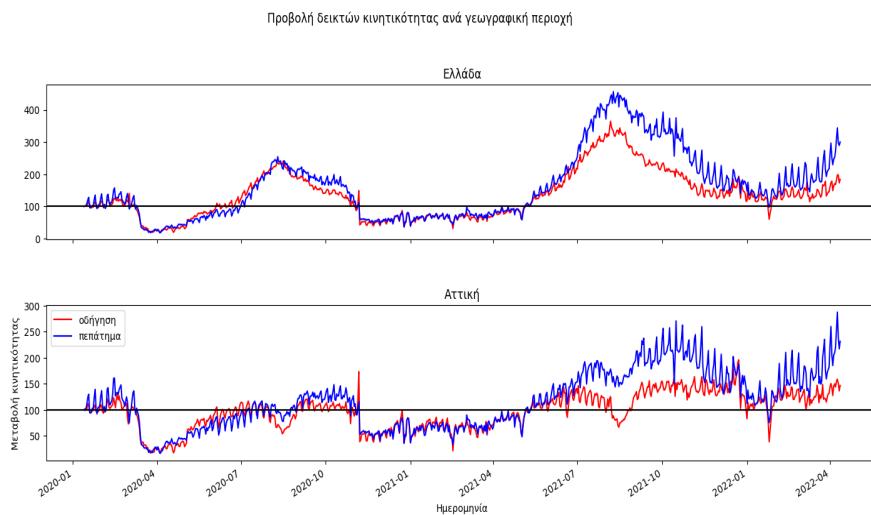
Εικόνα 6.6: Ιστόγραμμα για τους δείκτες οδήγησης και περπατήματος

46 Κεφάλαιο 6 Στατιστική ανάλυση

Χρονοσειρές κινητικότητας Apple κάτα την διάρκεια της πανδημίας



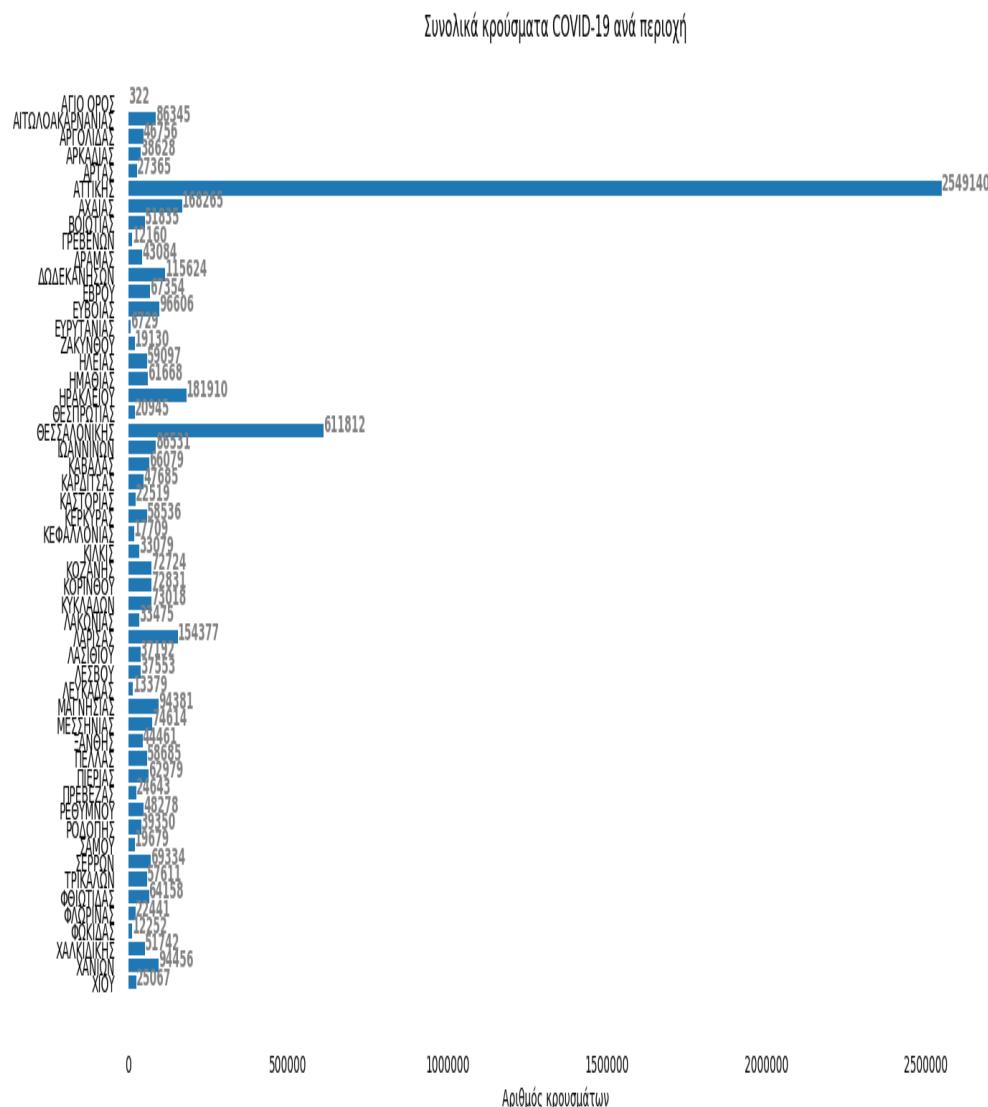
Εικόνα 6.7: Αναπαράσταση των ημερήσιων δεικτών οδήγησης και περπατήματος από την Apple που προσομειώνουν την κινητικότητα των πολιτών καθόλη την διάρκεια της πανδημίας



Εικόνα 6.8: Αναπαράσταση των ημερήσιων χρονοσειρών οδήγησης και περπατήματος της Apple της Ελλάδας και της Αττικής. Με κόκκινο χρώμα απεικονίζεται ο δείκτης της οδήγησης και με μπλε αυτός του περπατήματος

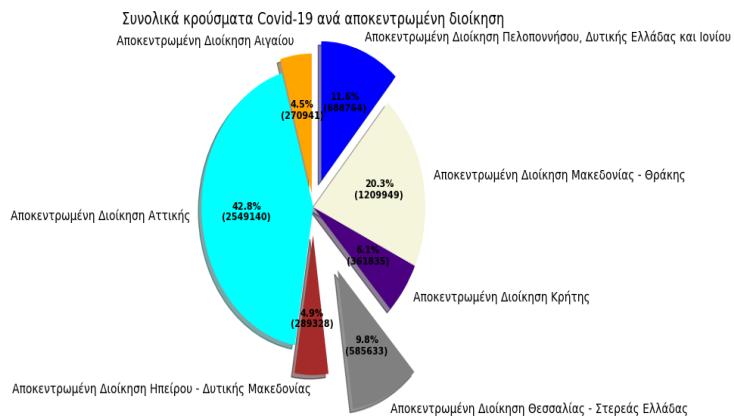
6.3 COVID-19 Response Greece

COVID-19 cumulative cases by area



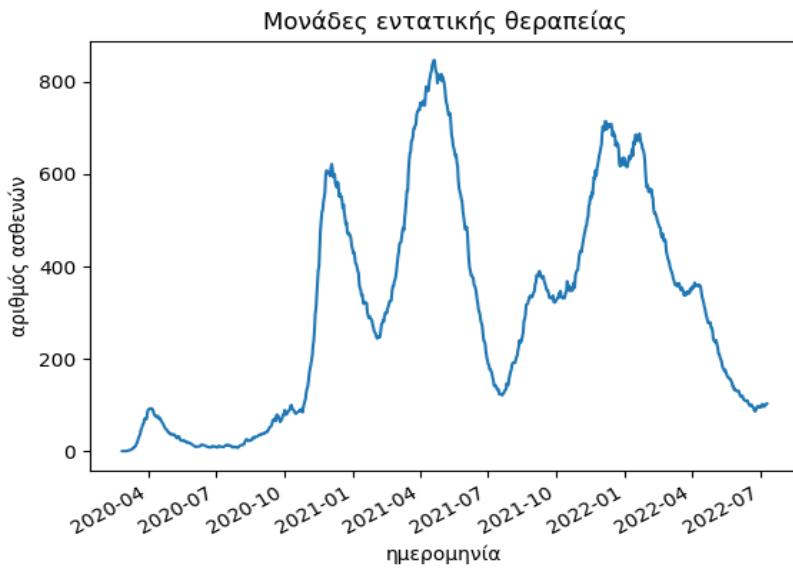
Εικόνα 6.9: Bar plot για συνολικά κρούσματα COVID-19 ανά νομό της Ελλάδας

48 Κεφάλαιο 6 Στατιστική ανάλυση



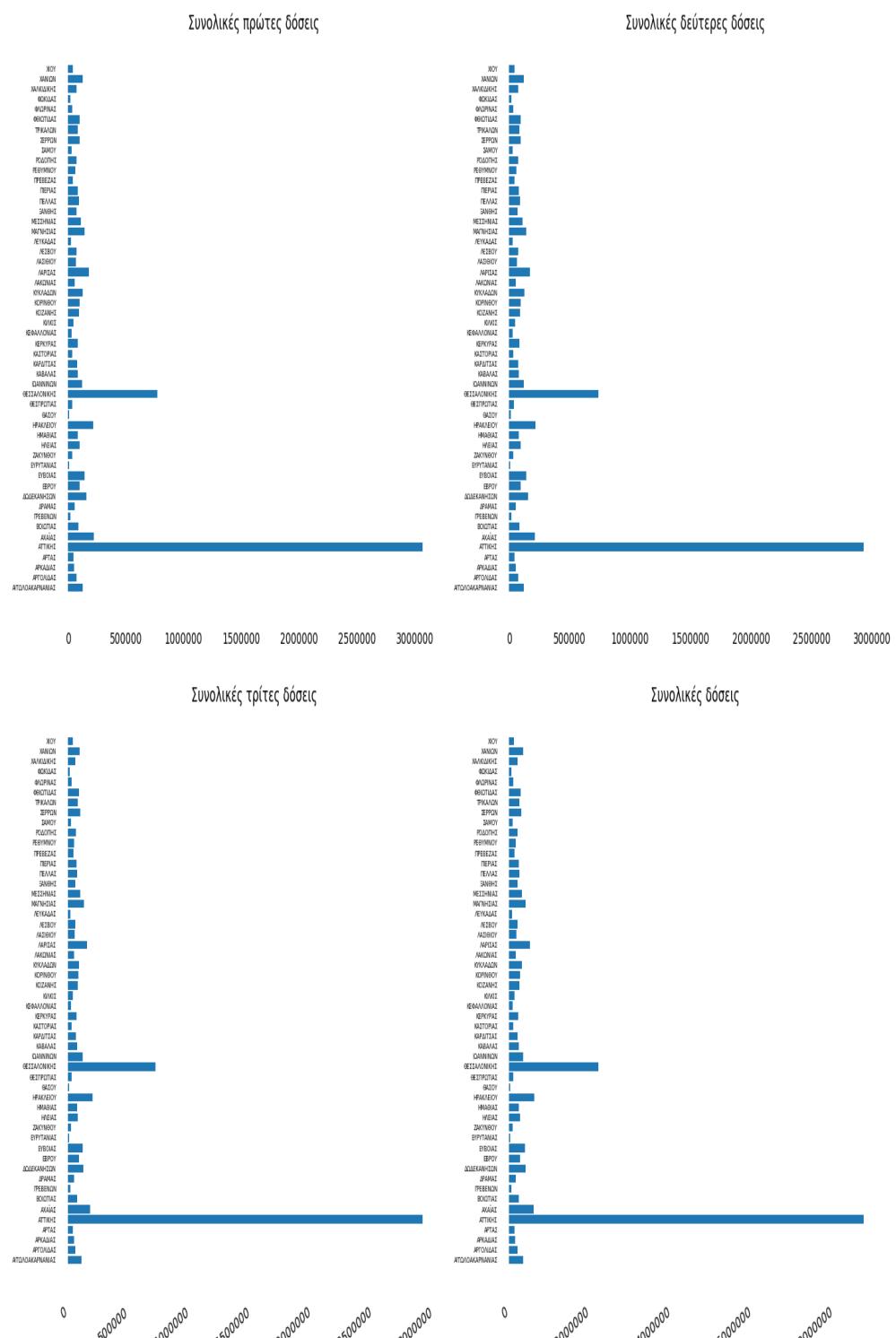
Εικόνα 6.10: Pie plot για συνολικά κρούσματα COVID-19 ανά αποκεντρωμένη διοίκηση της Ελλάδας

Intensive care cases



Εικόνα 6.11: Line plot που δείχνει τον αριθμό των ανθρώπων που βρίσκονται στις ΜΕΘ των νοσοκομείων από λοίμωξη COVID-19

Cumulative vaccinations per area



Εικόνα 6.12: Bar plot για εμβολιασμούς κατά του COVID-19 με πρώτη, δεύτερη και τρίτη δόση καθώς και για το σύνολο των εμβολιασμών που πραγματοποιήθηκαν ανά περιοχή

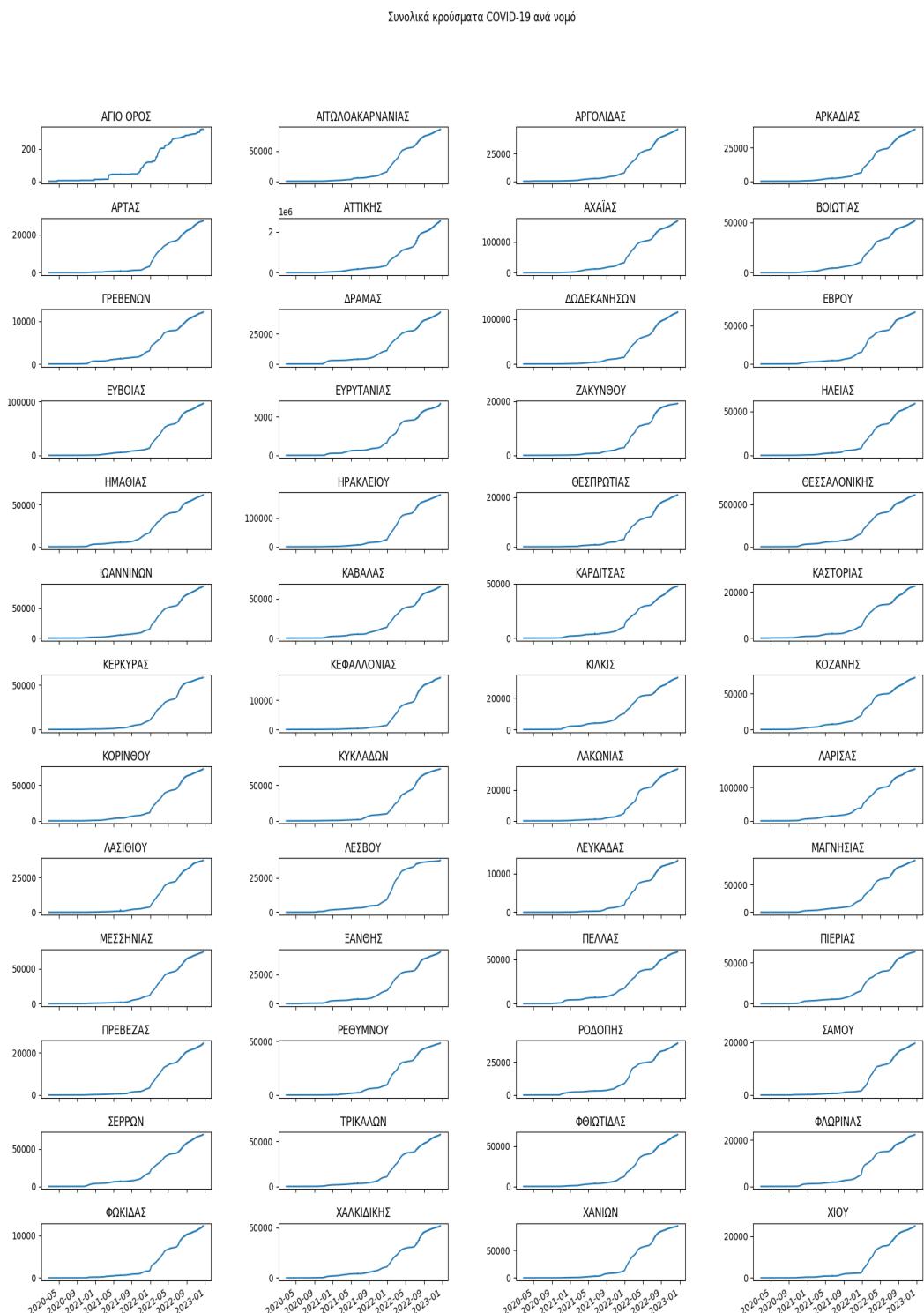
50 Κεφάλαιο 6 Στατιστική ανάλυση

COVID-19 history cases by area

Εβδομαδιαία κρούσματα COVID-19 ανά νομό



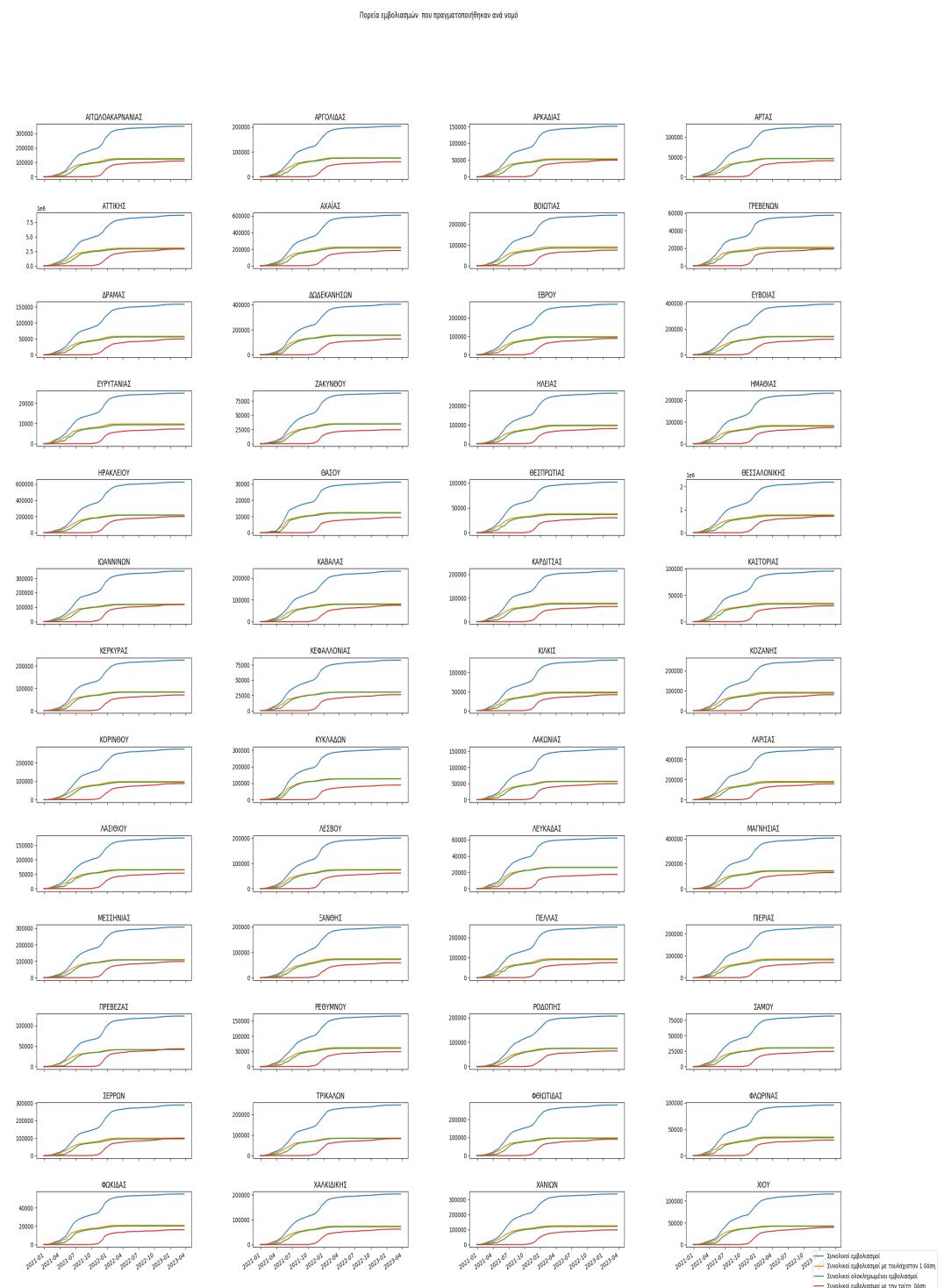
Εικόνα 6.13: Εβδομαδιαία κρούσματα COVID-19 ανά νομό της Ελλάδας



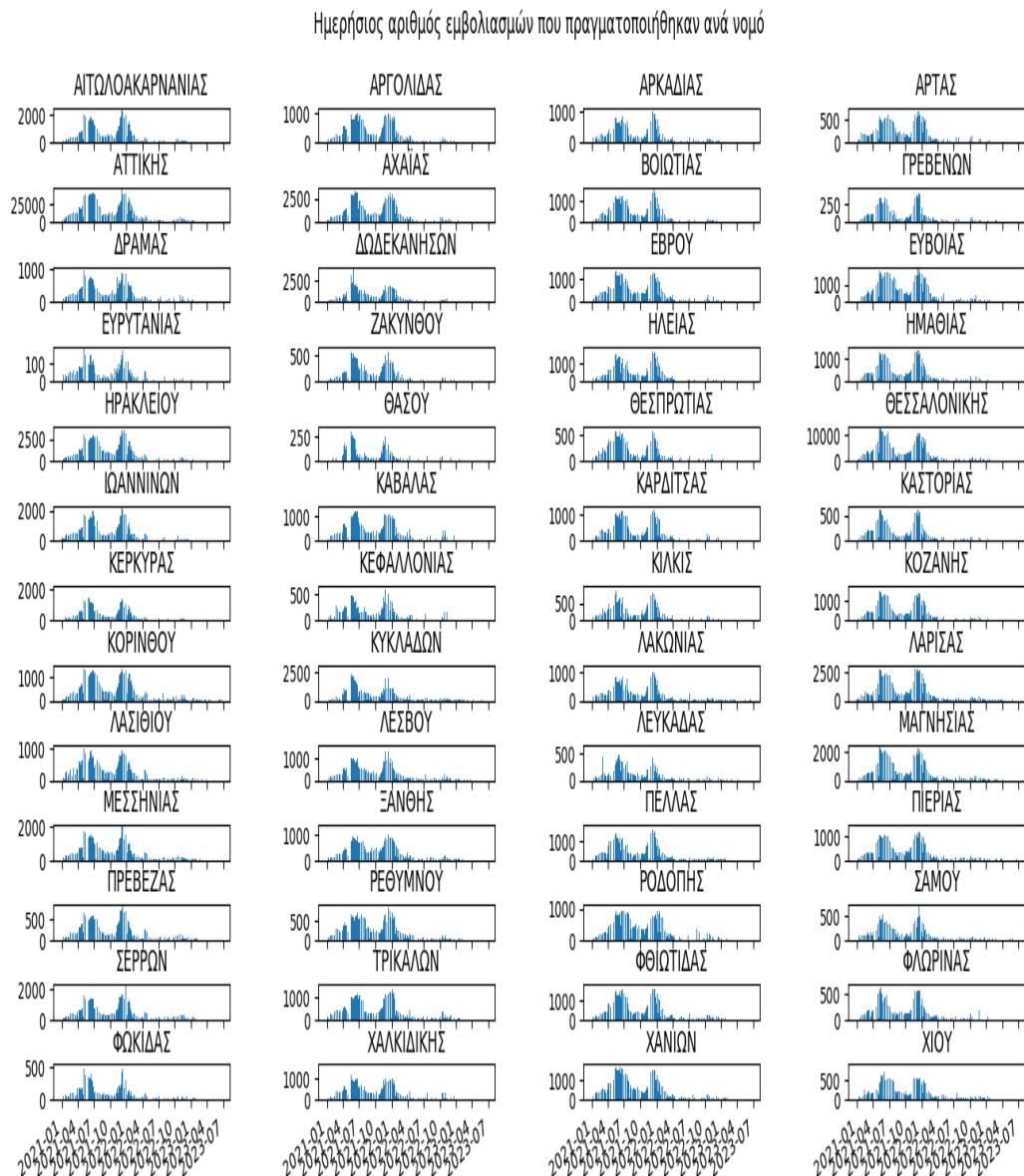
Εικόνα 6.14: Plot της πορείας των συνολικών κρουσμάτων COVID-19 ανά νομό της Ελλάδας

52 Κεφάλαιο 6 Στατιστική ανάλυση

Vaccination history data per area

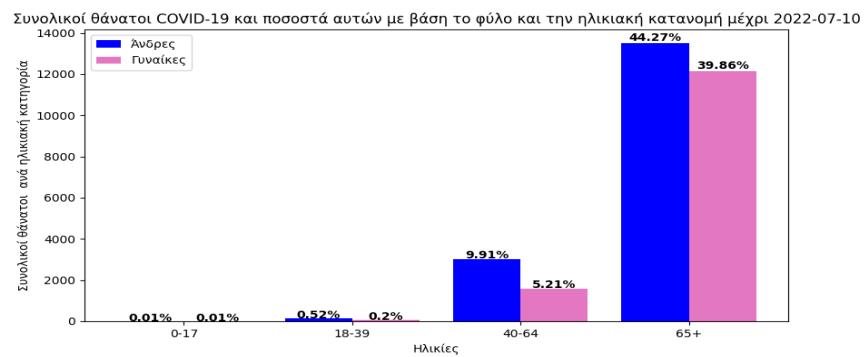


Εικόνα 6.15: Η πορεία των συνολικών εμβολιασμών απεικονίζεται με μπλε χρώμα, με πορτοκαλί οι εμβολιασμοί με τουλάχιστον μια δόση, με πράσινο οι ολοκληρωμένοι εμβολιασμοί και κόκκινο οι εμβολιασμοί με τρείς δόσεις

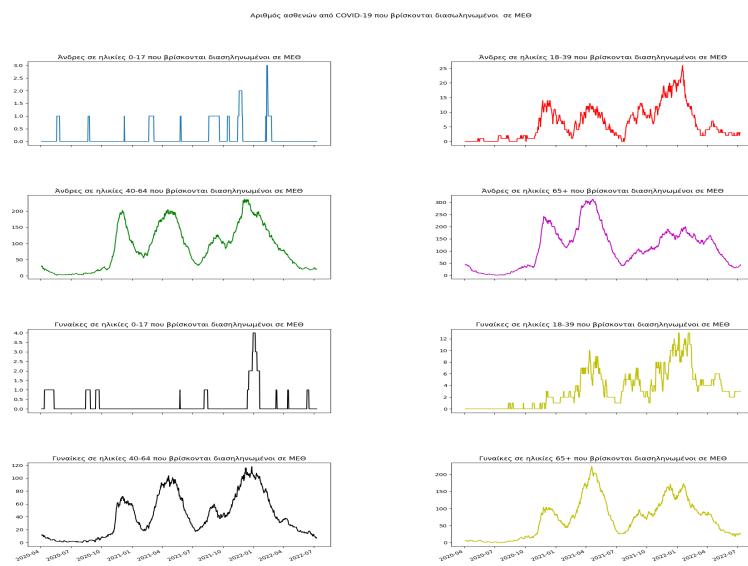


Εικόνα 6.16: Ημερήσιος αριθμός εμβολιασμών κατά του COVID-19 που πραγματοποιήθηκαν ανά νομό της Ελλάδας

6.4 Sandbird COVID-19-Greece



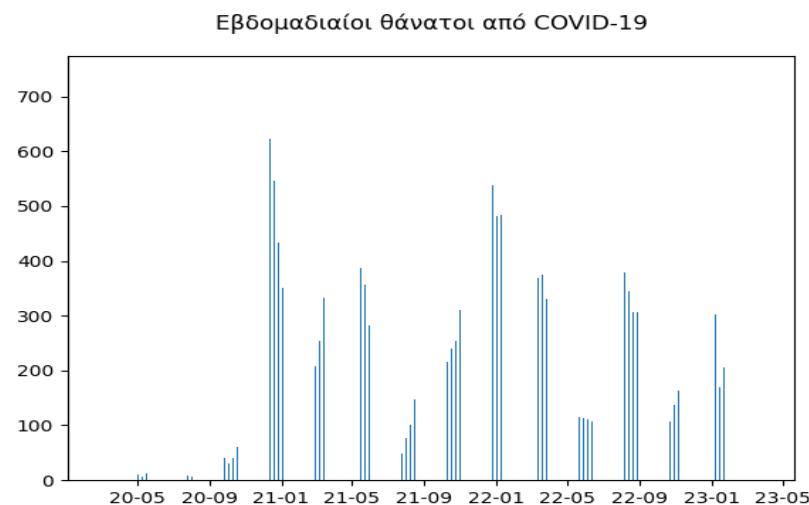
Εικόνα 6.17: Συνολικοί θάνατοι COVID-19 και ποσοστά εκδήλωσης αυτών με βάση την ηλικία και το φύλο. Με μπλε χρώμα αναπαρίστανται τα δεδομένα που αφορούν τους άντρες και με ροζ αντίστοιχα τις γυναίκες



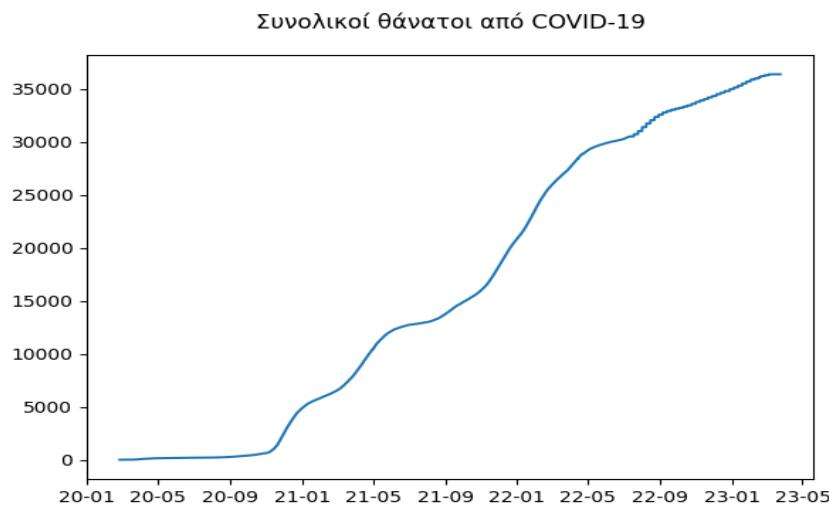
Εικόνα 6.18: Αριθμός ασθενών COVID-19 που βρίσκονται στις ΜΕΘ με βάση την ηλικία και το φύλο

6.5 iMedD-Lab - OpenData

Deaths



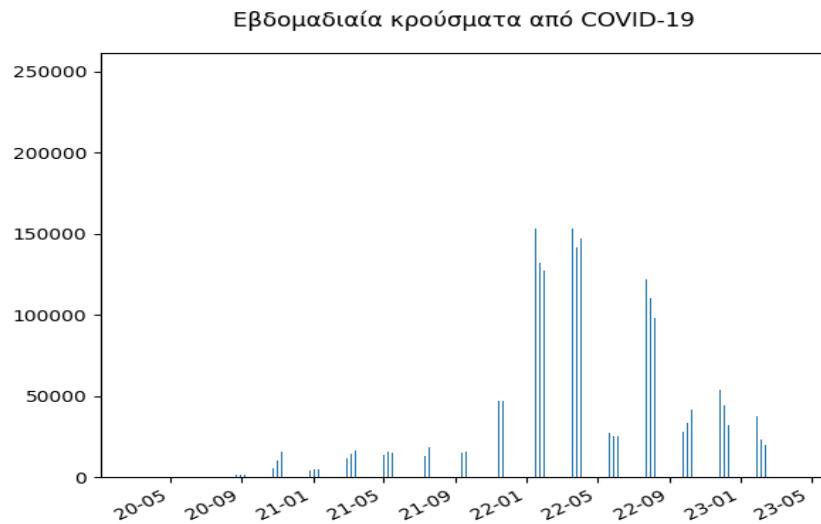
Εικόνα 6.19: Εβδομαδιαίοι θάνατοι από COVID-19



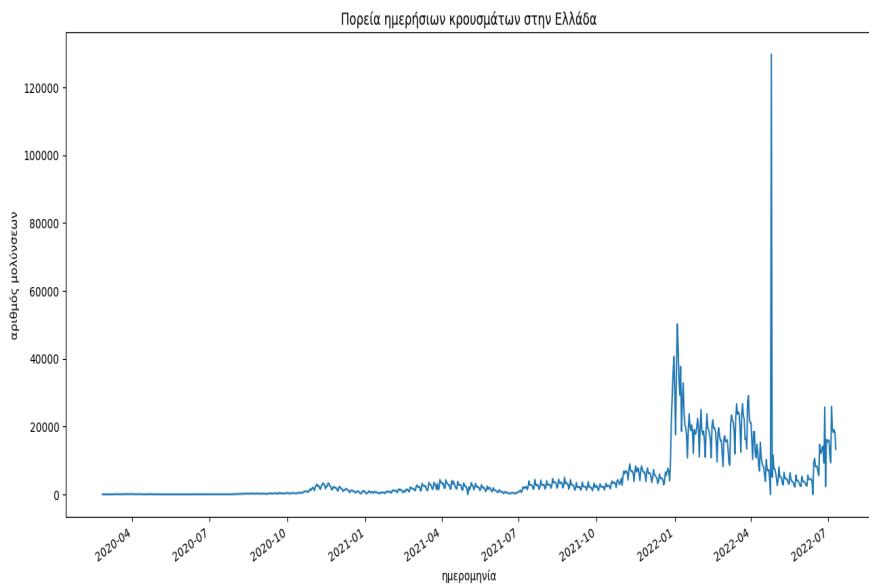
Εικόνα 6.20: Πορεία ημερήσιων συνολικών θανάτων από COVID-19

56 Κεφάλαιο 6 Στατιστική ανάλυση

COVID-19 cases greece



Εικόνα 6.21: Εβδομαδιαία πανελλήγια κρούσματα από COVID-19



Εικόνα 6.22: Πορεία ημερήσιων πανελλήγιων κρουσμάτων από COVID-19

7. Πειραματική μελέτη

7.1 Περιγραφή των μοντέλων

Σε αυτή την ενότητα πραγματοποιείται μια σύντομη περιγραφή των μοντέλων που αναπτύχθηκαν και αξιοποιήθηκαν στην εργασία. Η γλώσσα προγραμματισμού που αναπτύχθηκαν τα μοντέλα ήταν σε Python (εκδ. 3.9) και οι βιβλιοθήκες που αξιοποιήθηκαν είναι οι keras (εκδ. 2.7), tensorflow (εκδ. 2.7), xgboost (εκδ. 1.5.2), scikit-learn (εκδ. 1.0.1), matplotlib (εκδ. 3.7.1), pandas (εκδ. 1.3.5) και numpy (εκδ. 1.24.1). Η ανάπτυξη και η εκτέλεση του κώδικα πραγματοποιήθηκε σε έναν υπολογιστή με επεξεργαστή τεσσάρων πυρήνων και 8 threads και 8 GB μνήμης.

Νευρωνικό δίκτυο Multilayer Perceptron

Το νευρωνικό δίκτυο που αναπτύχθηκε αποτελείται από 3 χρυφά επίπεδα με 120, 60 και 30 νευρώνες αντίστοιχα. Το επίπεδο εισόδου αποτελείται από αριθμό νευρώνων ίσο με το γινόμενο του πλήθους των χαρακτηριστικών εισόδου X που χρησιμοποιούνται κατά την εκπαίδευση του μοντέλου επί των αριθμό των χρονικών στιγμών του παρελθόντος ενώ το επίπεδο εξόδου έχει αριθμό νευρώνων ίσο με το γινόμενο του πλήθους των χαρακτηριστικών πρόβλεψης γεγονότος την αριθμό των χρονικών βημάτων πρόβλεψης του μέλλοντος. Ο αριθμός των επιπέδων και των νευρώνων στο εκάστοτε επίπεδο προέκυψε έπειτα από πειραματισμό και σύμφωνα με την απόδοση στα σύνολα δεδομένων validation και test. Σε όλα τα χρυφά επίπεδα εφαρμόστηκε η συνάρτηση ενεργοποίησης ReLU [Agarap 2018] και στο μοντέλο ως συνάρτηση απώλειας χρησιμοποιήθηκε το RMSE μαζί με τον βελτιστοποιητή Adam [Kingma and Ba 2014]. Η επιλογή της συνάρτησης ενεργοποίησης ReLU βασίστηκε στο γεγονός ότι είναι ιδιαίτερα απλή και αρκετά αποδοτική καθώς συμπεριφέρεται σαν γραμμική για τιμές μεγαλύτερες του μηδενός και σαν μη γραμμική αφού αρνητικές τιμές αντιστοιχίζονται στο μηδέν. Ο βελτιστοποιήτης adam

58 Κεφάλαιο 7 Πειραματική μελέτη

επιλέχτηκε καθώς συνδυάζει τα πλεονεκτήματα των αλγορίθμων AdaGrad και RMSProp και απαιτεί ελάχιστη μνήμη, είναι εύκολα υλοποιήσιμος και αντιμετωπίζει αποτελεσματικά τον θόρυβο. Τα δίκτυα εκπαιδεύτηκε με βάση την τεχνική της έγκαιρης διακοπής (early stopping) [Prechelt 2002] σύμφωνα με το σφάλμα επαλήθευσης (validation loss) και με αριθμό ανοχής (patience) ίσο με 15 σε 200 εποχές.

Τυχαίο δάσος (Random Forest)

Χρησιμοποιήθηκαν 200 δέντρα στο σύνολο του τυχαίου δάσους με κάθε δέντρο να έχει μέγιστο βάθος 5, πλήθος χαρακτηριστικών ίσο με τον λογάριθμο του αριθμού των χαρακτηριστικών του συνόλου δεδομένων και κριτήριο διαχωρισμού το όποιο μετράει την ποιότητα του χωρίσματος το τετραγωνικό σφάλμα (squared_error).

Extreme Gradient Boosting (XGBoost)

Χρησιμοποιήθηκε ο XGBRegressor με αριθμό gradient boosted δέντρων ίσο με 5000, μέγιστο βάθος των δέντρων ίσο με 3, τιμή alpha που αφορά το $L1$ regularization (για την καταπολέμηση του overfitting γίνεται ανάθεση μηδενικών στα βάρη του μοντέλου και τιμωρείται το άθροισμα των απόλυτων τιμών των βαρών) με penalty ίσο με 2 και ελάχιστο άθροισμα των βαρών σε έναν κόμβο ίσο με 6. Επιπλέον ορίστηκαν ποσοστά sampling των στηλών και των training δειγμάτων σε κάθε δέντρο ίσα με 0.8, ρυθμός μάθησης ίσος με 0.05 και ως μετρική απώλειας το σφάλμα της μέσης τετραγωνικής ρίζας.

Μηχανή διανυσματικής στήριξης (SVR)

Αναπτύχθηκε μια μηχανή διανυσματικής στήριξης κατάλληλη για προβλήματα regression χρησιμοποιώντας την NuSVR από το sklearn με τιμή της παραμέτρου nu , που αποτελεί ένα κατώφλι για τα σφάλματα κατά την εκπαίδευση και τον αριθμό των διανυσμάτων, ίσο με 0.6 και τιμή του C που αφορά την $L2$ regularization (η $L2$ καταπολεμά την υπερπροσαρμογή αναγκάζοντας τα βάρη να είναι αρκετά μικρά αλλα όχι μηδενικά και τιμωρεί το άθροισμα των τετραγώνων των βαρών) ίσο με 1.5. Στην αρχιτεκτονική του μοντέλου αξιοποιείται και η MultiOutputRegressor για να υποστηρίζει το μοντέλο πολλαπλού εξόδου regression καθώς η δυνατότητα αυτή δεν παρέχεται από μόνη της από την NuSVR.

Αναδρομικά νευρωνικά Δίκτυα (RNN)

Το μοντέλο που αναπτύχθηκε αποτελεί έναν συνδυασμό αναδρομικών δικτύων τύπου SimpleRNN μαζί με Dense στρώματα. Τα αναδρομικά δίκτυα ενδέκχουνται για την χειρισμό χρονοσειρών. Τα κρυφά επίπεδα αποτελούνται από δύο επίπεδα SimpleRNN 40 και 25 μονάδων αντίστοιχα, όπου παρεμβάλλεται ένα στρώμα Dropout όπου θέτει με συχνότητα 0.2 με μηδενικές τιμές τυχαία κάποιες παραμέτρους (η ενεργοποίηση των νευρώνων δεν πραγματοποιείται στο πέρασμα προς τα εμπρός και τυχόν ενημερώσεις βάρους δεν εφαρμόζονται στον νευρώνα στο πίσω πέρασμα κατά την διάρκεια του training). Επιπλέον εφαρμόζεται στα στρώματα των SimpleRNN regularization τύπου $L2$ με τίμη 0.01 στους πύρηνες, στα bias και στις αναδρομικές σχέσεις των μονάδων. Χρησιμοποιήθηκε εξίσου και με τα υπόλοιπα μοντέλα η συνάρτηση ενεργοποίησης ReLU, ο αλγόριθμος βελτιστοποίησης Adam και early stopping στο σφάλμα επαλήθευσης (validation loss) με αριθμό ανοχής (patience) ίσο με 15.

Μακράς βραχύχρονης μνήμης δίκτυα (LSTM)

Η αρχιτεκτονική του μοντέλου αποτελείται από τρία κρυφά επίπεδα με δύο στρώματα LSTM και ένα Dense και από 80, 40, 20 νευρώνες αντίστοιχα. Ανάμεσα στα κρυφά επίπεδα υπάρχουν στρώματα Dropout με συχνότητα 0.2. Το επίπεδο εισόδου αποτελείται από αριθμό νευρώνων ίσο με το γινόμενο του πλήθους των χαρακτηριστικών X που χρησιμοποιούνται κατά την εκπαίδευση του μοντέλου επί τον αριθμό των χρονικών βημάτων του παρελθόντος, ενώ το επίπεδο εξόδου έχει αριθμό νευρώνων ίσο με το γινόμενο του πλήθους των χαρακτηριστικών πρόβλεψης y επί τον αριθμό των χρονικών βημάτων πρόβλεψης του μέλλοντος. Επιπλέον εφαρμόζεται στα στρώματα των LSTM regularization τύπου $L2$ με τίμη 0.01 στους πύρηνες, στα bias και στις αναδρομικές σχέσεις των μονάδων LSTM. Τέλος χρησιμοποιούνται αντίστοιχα οι συναρτήσεις ενεργοποίησης ReLU στα διάφορα στρώματα και early stopping για την αποφυγή της υπερπροσαρμογής με έλεγχο του σφάλματος επαλήθευσης (validation loss) και με αριθμό εποχών όπου δεν παρατηρείται βελτίωση των παραμέτρων του μοντέλου ίσο με 15.

Gated recurrent unit (GRU)

Το μοντέλο που δημιουργήθηκε αποτελείται από δύο κρυφά επίπεδα GRU με 70 και 40 νευρώνες αντίστοιχα. Ανάμεσα στα κρυφά επίπεδα υπάρχουν στρώματα Dropout με συχνότητα 0.2. Το επίπεδο εισόδου προσδιορίζεται με

βάση το διάνυσμα batch_input_shape όπου αποτελείται από τον αριθμό των batches, τον αριθμό των χρονικών βημάτων και το πλήθος των χαρακτηριστικών εισόδου ενώ το επίπεδο εξόδου έχει αριθμό νευρώνων ίσο με το γινόμενο του πλήθους των χαρακτηριστικών πρόβλεψης γεγονότος. Επιπλέον εφαρμόζεται στα στρώματα των GRU regularization τύπου L2 με τίμη 0.01 στους πύρηνες, στις αναδρομικές σχέσεις των μονάδων GRU και στα bias των στρώματων. Τέλος χρησιμοποιείται συνάρτηση απώλειας το RMSE μαζί με τον adam βελτιστοποιητή και early stopping για την αποφυγή της υπερπροσαρμογής με έλεγχο του σφάλματος επαλήθευσης (validation loss) και με αριθμό εποχών patience ίσο με 15.

Συνελικτικά νευρωνικά δίκτυα (CNN)

Αναπτύχθηκε ένα μοντέλο CNN για την επίλυση του προβλήματος της πρόβλεψης των χρονοσειρών. Αρχικά το μοντέλο περιλαμβάνει ένα επίπεδο συνέλιξης 1D που αφορά την χρονική συσχέτιση, το οποίο απαρτίζεται από αριθμό φίλτρων (προσδιορίζουν την διάσταση εξόδου έπειτα από την εφαρμογή της συνέλιξης) ίσο με 64, kernel_size (προσδιορίζει το μέγεθος του παραθύρου συνέλιξης) ίσο με 2, συνάρτηση ενεργοποίησης ReLU, input shape ίσο με τις διαστάσεις των διανυσμάτων χαρακτηριστικών εισόδου και εφαρμόζεται regularization L1 και L2 μεγέθους 0.01 στους πυρήνες και στα bias του στρώματος. Στην συνέχεια παρεμβάλλεται ένα στρώμα σμίκρυνσης MaxPooling1D με την εξαγωγή των μέγιστων τιμών των διανυσμάτων με βάση ένα χωροχρονικό παράθυρο μεγέθους 2 ενώ εφαρμόζεται padding έτσι ώστε η έξοδος να έχει ίδιες διαστάσεις με την είσοδο.

Έπειτα προστίθενται στην αρχιτεκτονική του μοντέλου ένα στρώμα Flatten για να μετατραπούν τα διανύσματα του μοντέλου προσωρινά σε 2D μορφή και ένα στρώμα RepeatVector με αριθμό επανάληψης n ίσο με 30 έτσι ώστε να ξαναμετατραπούν σε 3D, τώρα όμως με αριθμό χρονικών στιγμών ίσο με 30. Ακόμη προστίθενται δύο στρώματα Dense (το χρυφό με 30 νευρώνες και το επίπεδο εξόδου με αριθμό νευρώνων ίσο με το πλήθος των χαρακτηριστικών πρόβλεψης) μαζί με το wrapper layer TimeDistributed, καθώς είναι επιθυμητό να εφαρμοστεί ένα Dense layer σε κάθε χρονική στιγμή των διανυσμάτων εισόδου. Τέλος ορίζεται αριθμός επόχων ίσο με 150 και χρησιμοποιείται πάλι early stopping βασιμένο στην μετρική του σφάλματος επαλήθευσης (validation loss) με ανοχή ίση με 15 καθώς και η κλάση ReduceLROnPlateau του keras, η οποία συμβάλλει στην μείωση του ρυθμού μάθησης με έναν παράγοντα 0.2 όταν η μετρική του σφάλματος επαλήθευσης πάψει να βελτιώνεται έπειτα από 15 εποχές.

Νευρωνικό δίκτυο Convolutional-Long short term memory (CNN-LSTM)

Η αρχιτεκτονική του μοντέλου απαρτίζεται από ένα επίπεδο συνέλιξης 1D με αριθμό φίλτρων που προσδιορίζουν την διάσταση εξόδου έπειτα από την εφαρμογή της συνέλιξης ίσο με 64, kernel_size που προσδιορίζει το μέγεθος του παραθύρου συνέλιξης ίσο με 2 και συνάρτηση ενεργοποίησης ReLU. Στο στρώμα Conv1D ορίζονται ακόμη kernel_initializer τύπου he_normal που λαμβάνει δείγματα από μια περικομμένη κανονική κατανομή, input shape ίσο με τις διαστάσεις των διανυσμάτων χαρακτηριστικών εισόδου και εφαρμόζεται regularization L2 0.01 στους πυρήνες του στρώματος. Στην συνέχεια παρεμβάλλεται ένα στρώμα συμίχρυνσης MaxPooling1D με την εξαγωγή των μέγιστων τιμών των διανυσμάτων με βάση ένα χωροχρονικό παράθυρο μεγέθους 2. Τέλος χρησιμοποιείται ένα στρώμα Bidirectional LSTM με συνάρτηση ενεργοποίησης την ReLU, με 64 νευρώνες και regularization τύπου L2 εφαρμοσμένο στον πυρήνα. Στο μοντέλο ανάμεσα σε όλα τα επίπεδα εμφαρμόζεται Dropout με τιμή 0.2 και εκπαιδεύεται με βελτιστοποιητή adam και συνάρτηση απώλειας το RMSE.

7.2 Πειραματικά αποτελέσματα

Στους παρακάτω πίνακες παρατίθονται τα αποτελέσματα των μοντέλων στις μετρικές του συντελεστή προσδιορίσμου R^2 , του μέσου απόλυτου σφάλματος (MAE), του μέσου απόλυτου ποσοστού σφάλματος (MAPE) και του σκορ επεξήγησης διακύμανσης (EVS) σε προβλήματα πρόβλεψης των χρονοσειρών κινητικότητας από τα dataset της Apple και της Google αντίστοιχα.

Παρατηρούμε ότι τα καλύτερα αποτελέσματα όπως φαίνεται και στον πίνακα 7.1 για το πρόβλημα της πρόβλεψης των δεικτών Οδήγησης και Περπατήματος, τα έδωσε το μοντέλο του GRU με τα μοντέλα των LSTM και Vanilla RNN να ακολουθούν, γεγονός που χαρακτηρίζεται κάπως αναμενόμενο καθώς τα μοντέλα αυτά είναι πιο κατάλληλα για την πραγματοποίηση forecasting σε προβλήματα χρονοσειρών. Για αυτό το λόγο το μοντέλο GRU προτιμάται για την προβλέψη των δεικτών κινητικότητας του Apple dataset στην επόμενη ενότητα.

Στο πρόβλημα της εκτίμησης της κινητικότητας στις κατηγορίες της Google, καλύτερα αποτελέσματα όπως φαίνεται και στον πίνακα 7.2 έδωσε το XGBoost μοντέλο ακολουθούμενο από τα μοντέλα SVR και LSTM και για αυτό το

62 Κεφάλαιο 7 Πειραματική μελέτη

Model	R^2	MAE	MAPE	EVS
SVR	0.4230	19.9947	0.1447	0.4830
Random Forests	0.3431	46.1552	0.1640	0.6092
XGBoost	0.3072	45.811	0.1584	0.5673
MLP	0.5518	17	0.1826	0.6409
CNN	0.10	27.898	0.1840	0.2310
Vanilla RNN	0.5558	19.900	0.1218	0.5558
LSTM	0.5561	18.1912	0.1052	0.5675
GRU	0.6078	17.236	0.1041	0.6576
CNN-LSTM	0.0565	23.2472	0.1337	0.1983

Πίνακας 7.1: Συγκριτικός πίνακας των τιμών των μοντέλων στις μετρικές αξιολόγησης στο πρόβλημα της πρόβλεψης των χρονοσειρών Οδήγησης και Περπατήματος κατά την φάση του ελέγχου

Model	R^2	MAE	MAPE	EVS
SVR	0.576	12.8178	4355444	0.6625
Random Forests	0.4716	14.2007	5981226	0.5744
XGBoost	0.64	9.9148	3313697	0.6688
MLP	0.3348	10.2638	1935565	0.5377
CNN	-0.4334	16.9680	3546890	0.2136
Vanilla RNN	0.3362	9.4508	2307260	0.4698
LSTM	0.5421	11.0198	3025442	0.6450
GRU	0.3901	10.86	2293708	0.44
CNN-LSTM	0.4014	10.0916	1927163	0.6727

Πίνακας 7.2: Συγκριτικός πίνακας των τιμών των μοντέλων στις μετρικές αξιολόγησης κατά την φάση του ελέγχου για το πρόβλημα της πρόβλεψης των χρονοσειρών κινητικότητας σε μέρη λιανικής πώλησης και ψυχαγωγίας, παντοπωλεία και φαρμακεία, πάρκα, σταθμούς δημόσιας συγκοινωνίας, χώρους εργασίας και κατοικίες

λόγο το μοντέλο αυτό προτιμάται για την πραγματοποιήση προβλέψεων στις κατηγορίες της Google.

8. Πειραματικά αποτελέσματα σε πραγματικά σενάρια

8.1 Αυστηρή καραντίνα (lockdown) στην Ελλάδα

Για την εξαγωγή συμπερασμάτων σχετικά με την κινητικότητα των χρηστών στην διάρκεια της πανδημίας είναι αρκετά σημαντικό να παρουσιαστεί ένα χρονολόγιο με τις ημερομηνίες εφαρμογής και άρσης των περιοριστικών μέτρων που επιβλήθηκαν την περίοδο της αυστηρής καραντίνας που πέρασε η χώρα. Η Ελλάδα μέχρι στιγμής έχει εισέλθει σε δύο γενικά lockdown ενώ μέχρι και σήμερα η χρήση μάσκας στα νοσοκομεία παραμένει υποχρεωτική. Η σύνταξη αυτού του συνοπτικού χρονολογίου έγινε με βάση τα άρθρα για τα περιοριστικά μέτρα που δημοσιεύονταν κατά καιρούς στην εφημερίδα της Καθημερινής¹.

Περίοδος πρώτου αυστηρού lockdown (27 Φεβρουαρίου - 1 Ιουλίου 2020)

Περιοριστικά μέτρα άρχισαν να εφαρμόζονται από την ελληνική κυβέρνηση με την εμφάνιση του πρώτου κρούσματος του ιού στην χώρα στις 26

¹ Ενδεικτικά παρατίθενται τα άρθρα <https://www.kathimerini.gr/society/561394330/telos-i-nychterini-apagoreysi-apo-1i-ioylioy-poia-metra-airontai/>, <https://www.kathimerini.gr/society/561497539/odikos-chartis-gia-ta-christoygenna-toy-lockdown-ola-ta-metra/>, <https://www.kathimerini.gr/society/561147502/epistrofi-sto-lockdown-me-ti-chora-sto-kokkino/>, <https://www.kathimerini.gr/society/562322929/koronoios-telos-sta-metra-enanti-tis-pandimias-covid-19/>, <https://www.kathimerini.gr/society/561652528/paramoni-protouchrias-pos-tha-giotasyme-tin-allagi-toy-chronoy-ti-ischyei-gia-estiasi-kai-diaskedasi/>, <https://www.kathimerini.gr/society/561596221/nea-metra-gia-ton-koronoio-sti-dimosiotita-to-fek-oles-oi-allages/> και <https://www.kathimerini.gr/society/561568174/koronoios-ola-ta-metra-gia-toys-anemvoliastoys/>

64 Κεφάλαιο 8 Πειραματικά αποτελέσματα σε πραγματικά σενάρια

Φεβρουαρίου 2020. Αρχικά απαγορεύτηκαν οι αποκριάτικες εκδηλώσεις στις 27 Φεβρουαρίου και στις 9 Μαρτίου αποφασίστηκε η αναστολή των πολιτιστικών και καλλιτεχνικών εκδηλώσεων σε κλειστούς και ανοιχτούς χώρους που περιλαμβάνουν πάνω από 1000 άτομα, καθώς και η απαγόρευση του επισκεπτηρίου σε κέντρα αποκατάστασης και μονάδων θεραπείας στα νοσοκομεία. Εξαιτίας της μεγάλης διασποράς του ιού στις 10 Μαρτίου αποφασίστηκε το κλείσιμο όλων των εκπαιδευτικών δομών σε όλες τις βαθμίδες της χώρας και στη συνέχεια στις 13 Μαρτίου, η αναστολή λειτουργίας των καφετεριών, μπαρ, μουσείων, εμπορικών κέντρων, αθλητικών εγκαταστάσεων και εστιατορίων. Τρείς μέρες αργότερα, στις 16 Μαρτίου αποφασίστηκε και η καθολική αναστολή όλων των θρησκευτικών - δογματικών λειτουργιών.

Από τις 6 το πρώιμον έτος 2020 τέθηκε σε ισχύ το πρώτο γενικό lockdown και επιβλήθηκαν περιορισμοί στην κυκλοφορία και μετακίνηση των πολιτών σε ολόκληρη την επικράτεια με μόνες εξαιρέσεις την μετακίνηση προς και από την εργασία τους, την αγορά αγαθών πρώτης ανάγκης ή φάρμακων και την επίσκεψη σε γιατρό ή σε άτομο που χρήζει ανάγκης για βοήθεια. Οι φούρνοι, τα ζαχαροπλαστεία, τα pet shops, τα πρατήρια καυσίμων, τα περίπτερα, τα σουπερμάρκετ και τα καταστήματα τηλεπικοινωνιών παρέμειναν ανοιχτά ενώ τα μπαρ, οι καφετέριες και τα εστιατόρια λειτουργούσαν μόνο με take away. Στις υπόλοιπες ανθρώπινες δραστηριότητες ενισχύθηκε το μοντέλο της τηλεργασίας και του γλεκτρονικού εμπορίου.

Η σταδιακή άρση του lockdown ξεκίνησε από τις 4 Μαΐου 2020 με την κατάργηση της υποχρέωσης αποστολής μηνύματος στο 13033 και συνεχίστηκε με την επαναλειτουργία των εμπορικών καταστημάτων και τάξεων της Γ Λυκείου (11 Μαΐου), το άνοιγμα των εκκλησιών, μουσείων και αρχαιολογικών χώρων (17 Μαΐου) και την απελευθέρωση της δυνατότητας μετακίνησης και εκτός της οικείας περιφερειακής ενότητας (18 Μαΐου). Η άρση των περιοριστικών μέτρων συνεχίστηκε με την επαναλειτουργία των επιχειρήσεων εστίασης (25 Μαΐου), το άνοιγμα των τουριστικών μονάδων, σχολείων, κινηματογράφων και θεάτρων (1 Ιουνίου), την επαναλειτουργία των μπαρ, ίντερνετ καφέ, καταστημάτων τυχερών παιχνιδιών και των διαφόρων κέντρων διασκέδασης (15 Ιουνίου) και τέλος με την διοργάνωση παραστάσεων, συναυλιών, συνέδριων και εκδηλώσεων (1 Ιουλίου).

Περίοδος δεύτερου αυστηρού lockdown (7 Νοεμβρίου 2020 - 1 Ιουλίου 2021)

Το δεύτερο lockdown που επιβλήθηκε στην χώρα ξεκίνησε στις 7 Νοεμβρίου 2020 και επιβλήθηκαν αντίστοιχα μέτρα με το πρώτο γενικό lockdown με

την διαφορά ότι τα Νηπιαγωγεία και σχολεία Ειδικής Αγωγής παρέμειναν ανοιχτά. Ακολούθησε μια πιλοτική περίοδος επανέναρξης δραστηριοτήτων λόγω έξαρσης του ιού όπως έγινε για παράδειγμα στο λιανεμπόριο, το όποιο άνοιξε ξανά στις 18 Ιανουαρίου 2021 με click inside και κατάλληλη αποστολή SMS στο 13033 για τις χειμερινές εκπτώσεις και έκλεισε ξανά στις 11 Φεβρουαρίου, και στα λύκεια και γυμνάσια. Η τελική άρση των μέτρων ξεκίνησε στις 22 Μαρτίου 2021 με τα κομμωτήρια, κέντρα αισθητικής και τους αρχαιολογικούς χώρους, και συνεχίστηκε με το άνοιγμα των καταστημάτων στις 5 Απριλίου (εκτός πολυκαταστημάτων και εμπορικών κέντρων) με click-away και click-in-shop με SMS ενώ επιτράπηκαν οι διαδημοτικές μετακινήσεις κάθε Σαββατοκύριακο.

Η άρση των μέτρων συνεχίστηκε με το άνοιγμα των λυκείων (11 Απριλίου 2021), των μεγάλων εμπορικών καταστημάτων με click away και click inside (24 Απριλίου), της εστίασης σε εξωτερικούς χώρους (3 Μαΐου), των γυμνασίων και των δημοτικών σχολείων (10 Μαΐου) και την κατάργηση των click away και click inside στα καταστήματα. Πραγματοποιήθηκε επανέναρξη της λειτουργίας των μουσείων και καζίνων, μαζί με τον τουρισμό (15 Μαΐου), των συναυλιών σε εξωτερικούς χώρους (28 Μαΐου) και των γυμναστήριων (31 Μαΐου). Τέλος καταργήθηκε ο περιορισμός της μουσικής σε εξωτερικούς χώρους και η απαγόρευση της κυκλοφορίας το βράδυ (12 Ιουνίου) ενώ αργότερα επαναλειτούργησαν οι παιδότοποι, οι κλειστοί κινηματογράφοι και τα θέατρα (1 Ιουλίου 2021).

Λόγω της παραλλαγής όμικρον από τις 30 Δεκεμβρίου 2021 η λειτουργία των καταστημάτων εστίασης και διασκέδασης πραγματοποιούνταν έως τις 12 τα μεσάνυχτα χωρίς μουσική (επομένως δεν λειτουργούν κλαμπ και νυχτερινά κέντρα) ενώ καταργήθηκε η μουσική στα υπόλοιπα καταστήματα εστίασης. Επιπλέον η μάσκα ήταν υποχρεωτική σε εσωτερικούς και εξωτερικούς χώρους, υπήρχε απαγόρευση διοργάνωσης πάρτι σε δημόσιο ή ιδιωτικό χώρο, στα νοσηλευτικά ιδρύματα απαγορεύονταν η είσοδος σε ιατρικούς επισκέπτες και στα γήπεδα η πληρότητα εμβολιασμένων θεατών ήταν έως το 10% της χωρητικότητας τους, με ανώτερο όριο θεατών τα 1.000 άτομα. Από τις 31 Ιανουαρίου 2022 γίνεται σταδιακή άρση των εναπομείναντων μέτρων με τα καταστήματα εστίασης και τα κέντρα διασκέδασης να λειτουργούν αποκλειστικά για πλήρως εμβολιασμένους ή νοσήσαντες του τελευταίου τριμήνου και με την άρση της απαγόρευσης όρθιων στον κλάδο της διασκέδασης, την επαναφορά των σχολικών εκδρομών, την αύξηση της πληρότητας στα γήπεδα στο 50% και την εφαρμογή της τηλεργασίας στον ιδιωτικό και δημόσιο τομέα στο 20%.

Στις 15 Μαρτίου 2022 γίνεται κατάργηση της φόρμας Passenger Locator Form (PLF) για την είσοδο επισκεπτών στην Ελλάδα και των πιστοποι-

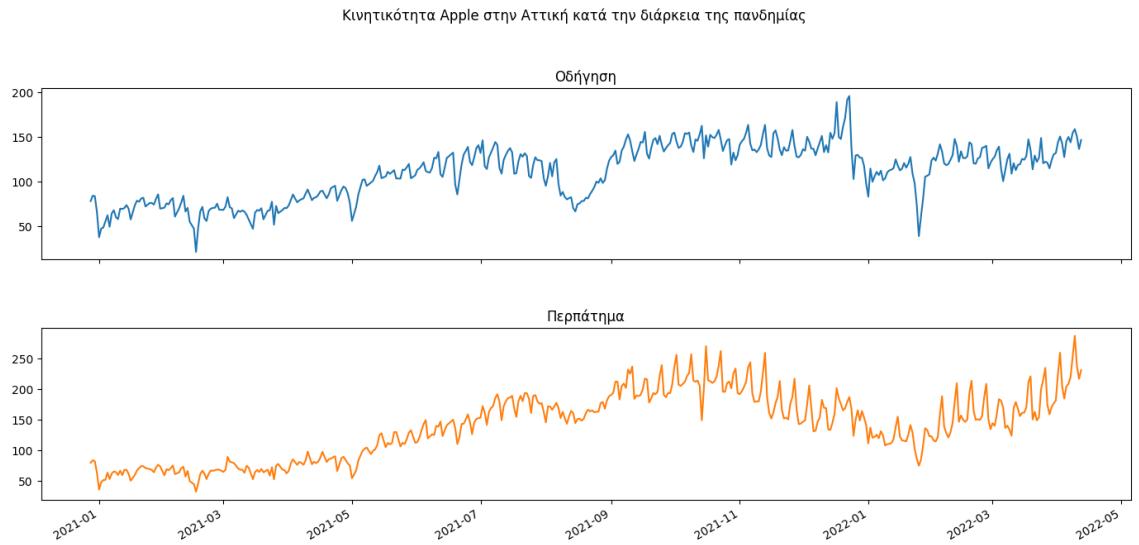
ητικών εμβολιασμού (1η Μαΐου). Επίσης από τις 15 Απριλίου αναστέλλεται το πρόστιμο των 100 ευρώ για τους ανεμβολίαστους άνω των 60 ετών και από 1/5/2022 επανέρχεται η λειτουργία όλων των επαγγελματικών δραστηριοτήτων στο 100% της χωριτικότητας τους. Η υποχρεωτική χρήση μάσκας στους εσωτερικούς χώρους καταργείται (αλλα παραμένει στα μέσα μαζικής μεταφοράς) και ορίζεται ενα rapid test την εβδομάδα για την πρόσβαση των ανεμβολίαστων στους χώρους εργασίας (1 Ιουνίου 2022). Τέλος στις 14 Μαρτίου 2023 η Επιτροπή Εμπειρογνωμόνων του υπουργείου Υγείας εισηγήθηκε την άρση των περιοριστικών μέτρων που αφορούν την υποχρεωτική χρήση μάσκας στα Μέσα Μαζικής Μεταφοράς και σε τμήματα των νοσοκομείων όπου οι εργαζόμενοι δεν έρχονται σε άμεση επαφή με τους ασθενείς και τη διενέργεια rapid test από τους ανεμβολίαστους στον ιδιωτικό τομέα. Παρόλα αυτα ένα από τα μέτρα που ακόμα παραμένει σε ισχύ είναι η διενέργεια υποχρεωτικού rapid test στους ανεμβολίαστους ασθενείς και εργαζόμενους των νοσοκομείων.

8.2 Πρόβλεψη δεικτών Οδήγησης και Περπατήματος στην Αττική

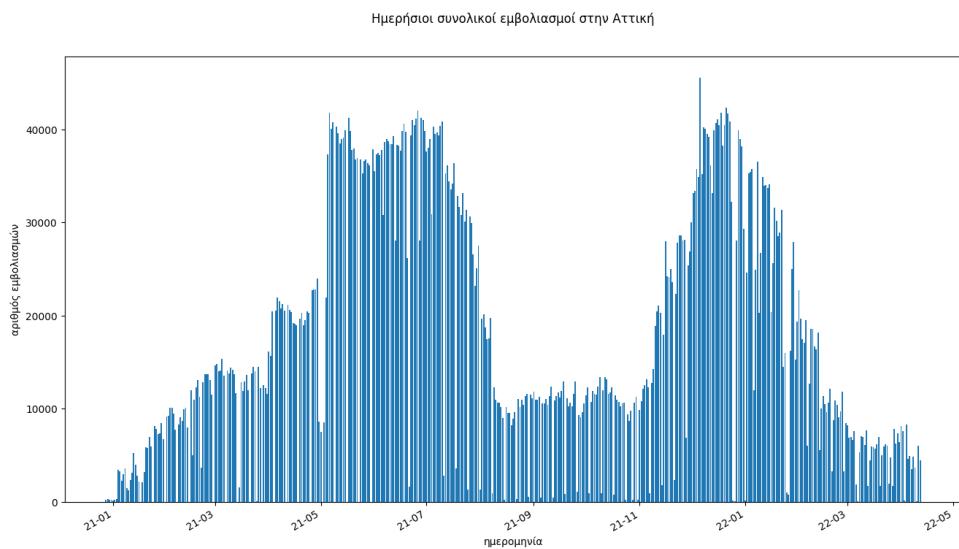
Στο πραγματικό αυτό σενάριο γίνεται μια προσπάθεια πρόβλεψης των Apple δεικτών κινητικότητας των πολιτών στην Αττική κατά την διάρκεια της πανδημίας, σε μια περιφέρεια της χώρας όπου ο ίδιος είχε έντονη επιδημιολογική πορεία. Το μοντέλο που χρησιμοποιήθηκε για να προβλεφθούν οι τιμές των δεικτών Οδήγησης (Driving) και Περπατήματος (Walking) ήταν το GRU. Ως χαρακτηριστικά εισόδου στο μοντέλο δόθηκε ο αριθμός των ημερήσιων συνολικών εμβολιασμών και οι δείκτες κινητικότητας στην Αττική, και ως χαρακτηριστικά πρόβλεψης ορίζονται οι αντίστοιχοι δείκτες.

Έχοντας επιλέξει τις κατάλληλες εγγραφές, τα δεδομένα συνενώνονται σε ένα data frame. Στην συνέχεια αφού ορίστει μέγεθος παραθύρου των προηγούμενων χρονικών στιγμών ίσο με τέσσερα και βήμα πρόβλεψης του μέλλοντος ίσο με ένα, το σύνολο δεδομένων μετατρέπεται σε χρονοσειρα. Έπειτα εφαρμόζεται μετασχηματισμός των χαρακτηριστικών, κανονικοποιώντας τα εφαρμόζοντας αφαίρεση της μέσης τιμής τους και διαιρέση με την διασπορά τους (z-scaling) και χωρίζεται το σύνολο δεδομένων σε training (70%), validation (10%) και test (20%) sets.

8.2 Πρόβλεψη δεικτών Οδήγησης και Περπατήματος στην Αττική 67

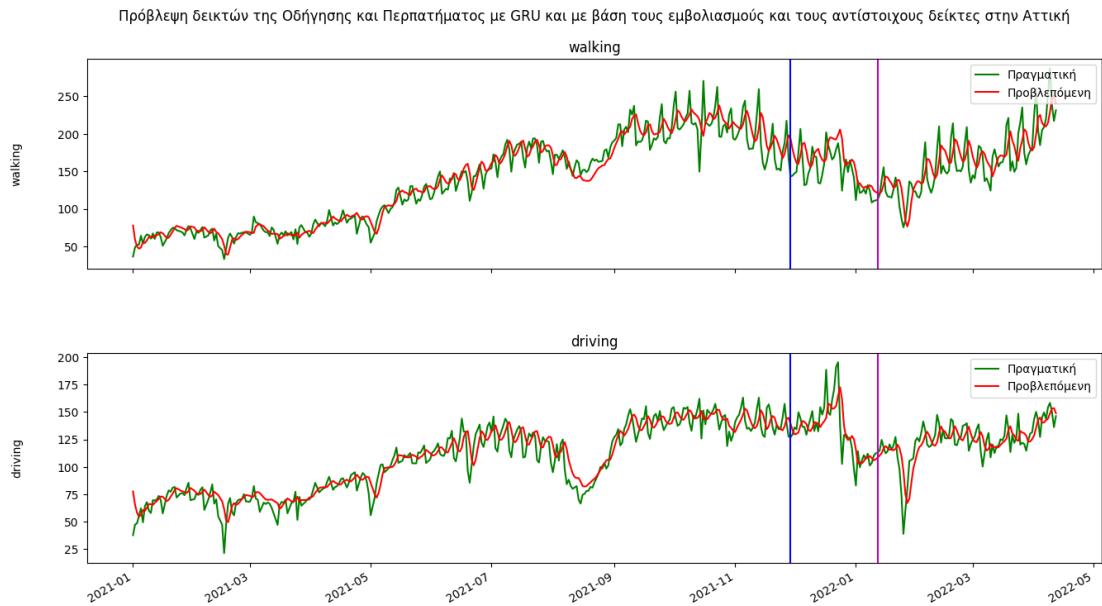


Εικόνα 8.1: Δείκτες κινητικότητας Οδήγησης και Περπατήματος στην Αττική



Εικόνα 8.2: Ημερήσιοι εμβολιασμοί στην Αττική

68 Κεφάλαιο 8 Πειραματικά αποτελέσματα σε πραγματικά σενάρια



Εικόνα 8.3: Πρόβλεψη δεικτών της Apple στην Αττική με GRU με βάση τους εμβολιασμούς και τους αντίστοιχους δείκτες

Στο παραπάνω γράφημα 8.3 αναπαρίσταται τα αποτελέσματα πρόβλεψης του GRU μοντέλου χρησιμοποιώντας τους ημερήσιους συνολικούς εμβολιασμούς και τους ημερήσιους δείκτες κινητικότητας του Apple dataset. Οι χρονοσειρές της Οδήγησης και του Περπατήματος με τα πραγματικά δεδομένα αναπαρίστανται με πράσινο χρώμα και οι εκτιμώμενες από το μοντέλο GRU με κόκκινο. Περίοδος εκπαίδευσης ορίστηκε το χρονικό διάστημα από 2020-12-28 έως 2021-11-28, περίοδος validation από 2021-11-29 έως 2022-01-11 και τέλος περίοδος test από 2022-01-12 έως το τέλος των δεδομένων. Το περπάτημα ήταν η χρονοσειρά η οποία ήταν δυνατόν να προβλεφθεί με μεγαλύτερη ακρίβεια από το μοντέλο σε σχέση με την οδήγηση σύμφωνα με τα διάφορα πειράματα τα οποία εκτελέστηκαν.

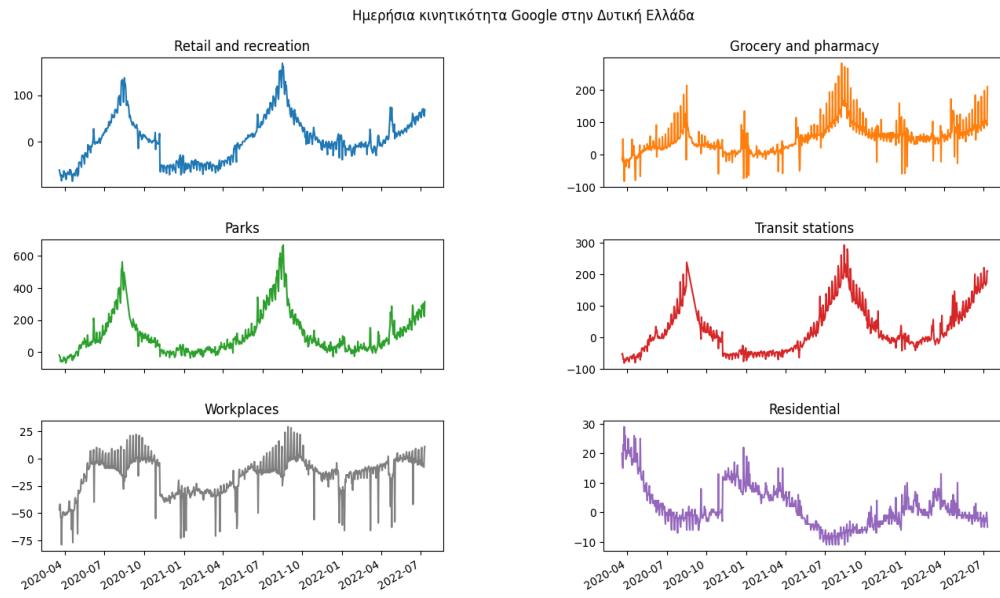
8.3 Πρόβλεψη κινητικότητας στις κατηγορίες της Google στην Περιφέρεια της Δυτικής Ελλάδας

Στο συγκεκριμένο real-life σενάριο προσπαθούμε να προβλέψουμε την κινητικότητα των πολιτών στις κατηγορίες του συνόλου δεδομένων της Google στην περιφέρεια της Δυτικής Ελλάδος (περιλαμβάνει τις περιοχές της Αιτωλοακαρανίας, Αργολίδας, Αρκαδίας, Αχαΐας, Ηλείας, Κέρκυρας, Κεφαλλονιάς, Κορίνθου, Λευκάδας, Μεσσηνίας και Ζακύνθου) χρησιμοποιώντας παράλληλα τον ημερήσιο αριθμό κρουσμάτων. Το μοντέλο που χρησιμοποιήθηκε για να προβλεφθούν οι τιμές των δεικτών κινητικότητας στις κατηγορίες «Retail and Recreation», «Grocery and pharmacy», «Parks», «Transit stations», «Workplaces» και «Residential» από το Google dataset ήταν το XGBoost. Ως χαρακτηριστικά εισόδου στο μοντέλο δόθηκαν τα ημερήσια κρούσματα και οι δείκτες κινητικότητας στις Google κατηγορίες στην Περιφέρεια της Δυτικής Ελλάδας και ως χαρακτηριστικά πρόβλεψης ορίζονται οι αντίστοιχοι δείκτες.

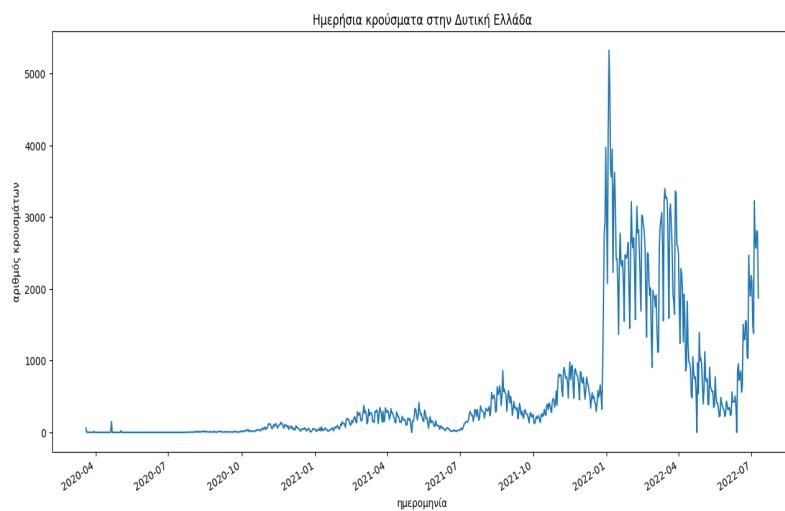
Αφού επιλεχθούν οι εγγραφές που αφορούν μόνο την περιφέρεια Δυτικής Ελλάδας πραγματοποιώντας κατάλληλη ομαδοποίηση και άθροισμα των δεδομένων των περιοχών που απαρτίζουν αυτήν την περιφέρεια, ορίζεται ο αριθμός παρελθοντικών χρονικών βημάτων ίσος με επτά και το βήμα πρόβλεψης του μέλλοντος ίσο με ένα. Στην συνέχεια το σύνολο δεδομένων μετατρέπεται σε χρονοσειρά, γίνεται χρήση της τεχνικής Standardization και χωρίζεται σε training (60%) και test (40%) sets για την πραγματοποίηση Forecasting.

Στο γράφημα της εικόνας 8.6 αναπαρίσταται τα αποτελέσματα πρόβλεψης του XGBoost μοντέλου χρησιμοποιώντας τα ημερήσια κρούσματα και τους ημερήσιους δείκτες στις Google κατηγορίες. Οι χρονοσειρές με τα πραγματικά δεδομένα αναπαρίστανται με πράσινο χρώμα και οι εκτιμώμενες από το μοντέλο με κόκκινο. Περίοδος εκπαίδευσης ορίστηκε το χρονικό διάστημα από 2020-12-28 έως 2021-08-17 και περίοδος test από 2021-08-18 έως το τέλος των δεδομένων. Οι χρονοσειρές κινητικότητας των πολιτών στις κατηγορίες Google που προβλέψθηκαν με μεγαλύτερη ακρίβεια ήταν οι κατηγορίες της λιανικής και φυχαγωγίας, των σταθμών δημόσιας συγκοινωνίας και των πάρκων ενώ οι κατηγορίες με την μικρότερη ακρίβεια πρόβλεψης ήταν οι κατοικίες, τα παντοπωλεία -φαρμακεία και οι χώροι εργασίας.

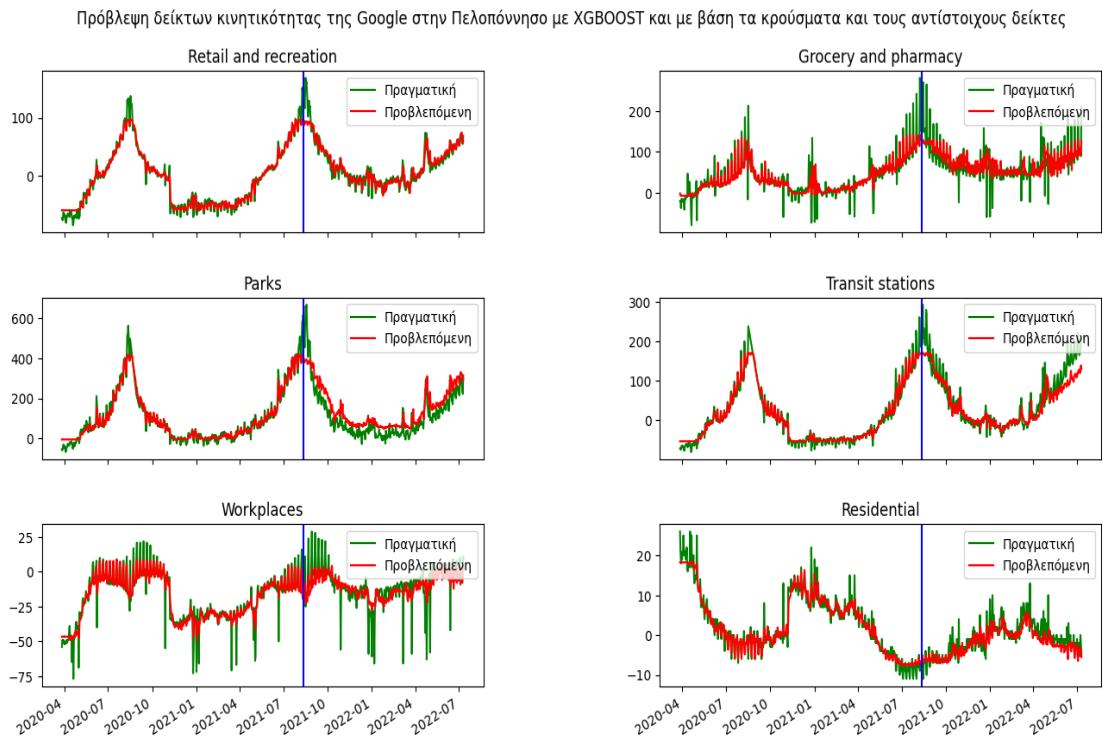
70 Κεφάλαιο 8 Πειραματικά αποτελέσματα σε πραγματικά σενάρια



Εικόνα 8.4: Δείκτες κινητικότητας στις κατηγορίες Google στην Περιφέρεια της Δυτικής Ελλάδας



Εικόνα 8.5: Ημερήσια κρούσματα στην Περιφέρεια της Δυτικής Ελλάδας

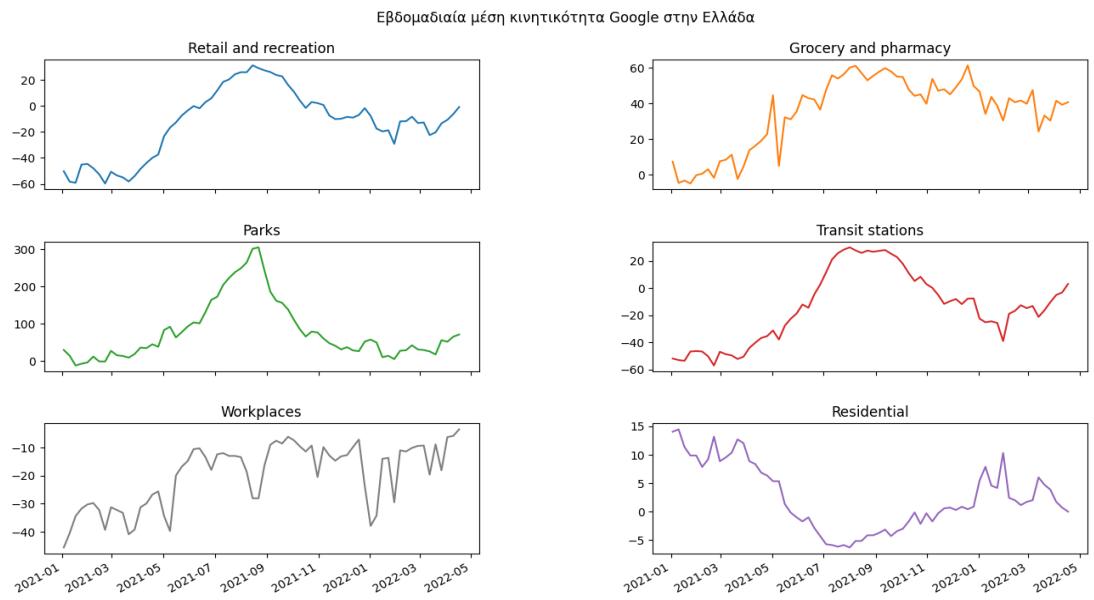


Εικόνα 8.6: Πρόβλεψη δεικτών κινητικότητας στις κατηγορίες της Google στην Περιφέρεια Δυτικής Ελλάδας με βάση τα κρούσματα και τους αντίστοιχους δείκτες

8.4 Πρόβλεψη εβδομαδιαίων θανάτων από COVID-19

Στην ενότητα αυτή περιγράφεται το πρόβλημα της πρόβλεψης των μέσων εβδομαδιαίων θανάτων από τον ιό στην Ελλάδα με την χρήση ενός LSTM μοντέλου. Για την πραγματοποίηση των προβλέψεων σχετικά με τους θανάτους χρησιμοποιήθηκαν δεδομένα που αφορούν τους ασθενείς που βρίσκονται σε μονάδες εντατικής θεραπείας και την κινητικότητα των ανθρώπων στις κατηγορίες μερών της Google καθώς, και ο ίδιος αριθμός των θανάτων. Ορίστηκε batch size ίσο με 1, μέγεθος παραθύρου παρελθοντικών δεδομένων εδβομαδιαίας φύσης ίσο με 4 και αριθμός βήματος πρόβλεψης ίσος με 1. Τέλος εφαρμόστηκε κανονικοποιήση στα χαρακτηριστικά (z-score normalization) και το σύνολο δεδομένων χωρίστηκε σε training (60%), validation (20%) και test (20%) sets.

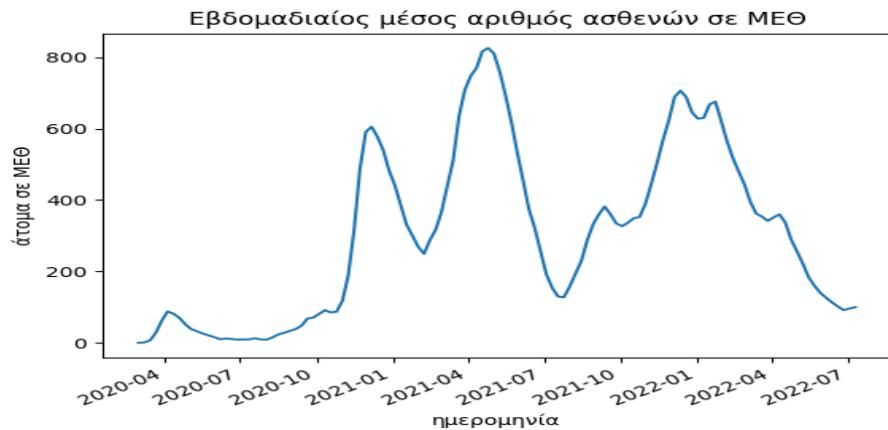
72 Κεφάλαιο 8 Πειραματικά αποτελέσματα σε πραγματικά σενάρια



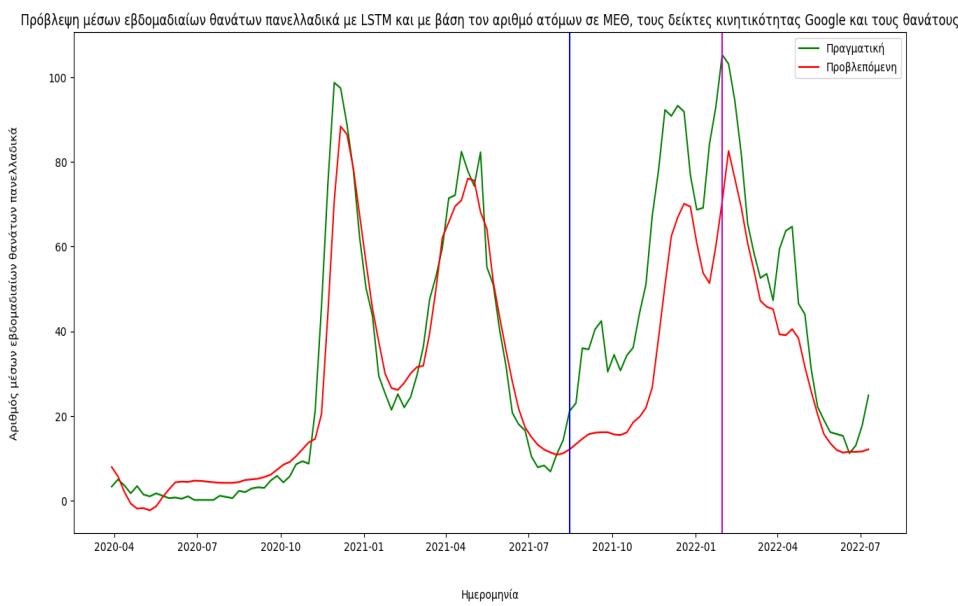
Εικόνα 8.7: Εβδομαδιαία μέση κινητικότητα στις καητηγορίες της Google στην Ελλάδα



Εικόνα 8.8: Εβδομαδιαίοι μέσοι θάνατοι στην Ελλάδα



Εικόνα 8.9: Εβδομαδιαίος μέσος αριθμός ασθενών που νοσηλεύονται σε μονάδες εντατικής θεραπείας στην Ελλάδα



Εικόνα 8.10: Πρόβλεψη μέσων εβδομαδιαίων θανάτων στην Ελλάδα χρησιμοποιώντας την μέση εβδομαδιαία κινητικότητα στις κατηγορίες μερών της Google, τον μέσο εβδομαδιαίο αριθμό ατόμων σε ΜΕΘ και τον αντίστοιχο δείκτη

74 Κεφάλαιο 8 Πειραματικά αποτελέσματα σε πραγματικά σενάρια

Στο παραπάνω γράφημα 8.10 αναπαρίσταται τα αποτελέσματα πρόβλεψης του LSTM μοντέλου στο πρόβλημα της πρόβλεψης του αριθμού των μέσων εβδομαδιαίων θανάτων. Η χρονοσειρά των θανάτων με τα πραγματικά δεδομένα αναπαρίστανται με πράσινο χρώμα και η εκτιμώμενη από το μοντέλο με κόκκινο. Περίοδος εκπαίδευσης ορίστηκε το χρονικό διάστημα από 2020-12-28 έως 2021-08-14, περίοδος validation από 2021-08-15 έως 2022-01-29 και περίοδος test από 2022-01-30 έως το τέλος των δεδομένων. Σε γενικές γραμμές, παρατηρούμε ότι η τάση της χρονοσειράς των πανελλήνιων εβδομαδιαίων μέσων θανάτων προβλέφθηκε επιτυχώς από το μοντέλο.

9. Συμπεράσματα - Προτάσεις για μελλοντική έρευνα

9.1 Συμπεράσματα

Στην παρούσα εργασία πραγματοποιήθηκε ανάλυση των δεδομένων κινητικότητας των Ελλήνων πολιτών κατά την διάρκεια της πανδημίας και αναπτύχθηκαν μοντέλα μηχανικής μάθησης που είναι σε θέση να προβλέπουν με βάσει τα δεδομένα αυτά, την μεταβολή της ανθρώπινης κινητικότητας στο μέλλον ή κάποια μετρική που καταγράφει την πορεία του ιού στην χώρα μας (π.χ., μελλοντικό αριθμό κρουσμάτων ή θανάτων). Βασικό συμπέρασμα που προέκυψε από την πειραματική μελέτη σε πραγματικά σενάρια είναι ότι η κινητικότητα των πολιτών είναι θετικά συσχετιζόμενη με την μετάδοση του ιού, καθώς η ύπαρξη του εμβολίου κατά του COVID19 και οι αριθμοί των κρουσμάτων, θανάτων και ατόμων σε Μονάδες Εντατικής θεραπείας αλληλοεπιδρούν τόσο μεταξύ τους όσο και με την ανθρώπινη κινητικότητα.

Η χρονοοσειρά τους περιπατήματος κατά την διάρκεια της πανδημίας εμφάνισε μεγαλύτερα ποσοστά επιτυχίας πρόβλεψης από τα μοντέλα, γεγονός που ίσως οφείλεται στις μικρότερες μεταβολές που παρουσιάζει το περπάτημα σε σύγκριση με την χρονοοσειρά της οδήγησης (η ύπαρξη των lockdown επέβαλε περιορισμούς στις μετακινήσεις με αυτοκίνητο κάτι που δεν συνέβη τόσο έντονα με την άσκηση σε εξωτερικούς χώρους και το περπάτημα). Επιπρόσθετα η εξέλιξη του ίου στην ελληνική κοινωνία στα διάφορα πειράματα φάνηκε να είναι πιο άρρηκτα συνδεδεμένη με την κινητικότητα των πολίτων στα καταστήματα της λιανικής και φυχαγωγίας, στους σταθμούς δημόσιας συγκοινωνίας

και στα πάρκα σε σύγκριση με τις ανθρώπινες μετακινήσεις στα φαρμακεία – παντοπωλεία, στους χώρους εργασίας και την παραμονή στο σπίτι.

9.2 Προτάσεις για μελλοντική μελέτη

Αντικείμενο μελλοντικής μελέτης θα μπορούσε να ήταν η αναγνώριση περιοχών με υψηλή επικινδυνότητα για μόλυνση από τον ιό με βάσει δεδομένα κινητικότητας, η οποία θα βοηθήσει στην θέσπιση αποτελεσματικότερων μέτρων περιορισμού ενάντια στην πανδημία. Επιπλέον ενδιαφέρον παρουσιάζει η χρήση δεδομένων αστικοποίησης μιας περιοχής και δημογραφικών δεδομένων των κατοίκων της, ως είσοδο στα μοντέλα που αναπτύχθηκαν για την πραγματοποίηση προβλέψεων κινητικότητας. Άλλωστε ο βαθμός αστικοποίησης και το προφίλ της οικονομικής δραστηριότητας των πολιτών αποτελούν παράγοντες που επηρεάζουν την κινητικότητα τους.

Μια περαιτέρω πρόταση για μελλοντική έρευνα θα αποτελούσε η διερεύνηση της συνεισφοράς των τρόπων μικροκινητικότητας (πχ. ηλεκτρικά ποδήλατα και σκούτερ) που υιοθετήθηκαν στην πανδημίας στην μείωση της μετάδοσης του ιού. Τέλος σημαντικές προσθήκες στην εργασία θα ήταν η ανάπτυξη και χρήση πολυπλοκότερων μοντέλων βαθιάς μηχανικής μάθησης (π.χ. Transformer μοντέλα) και επιδημιολογικών μοντέλων (όπως το SEIR) για τον προσδιορίσμο της κατάλληλης χρονικής στιγμής για την επιβολή ή την άρση των περιοριστικών μέτρων σε σημεία ενδιαφέροντος με βάση την κινητικότητα των πολιτών και η υλοποίηση αποδοτικότερων τεχνικών οπτικοποιήσης των δεδομένων.

Βιβλιογραφία

- A. F. Agarap. 2018. Deep learning using rectified linear units (relu). *arXiv preprint arXiv: 1803.08375*.
- T. Alamo, D. G. Reina, P. M. Gata, V. M. Preciado, and G. Giordano. 2021. Data-driven methods for present and future pandemics: Monitoring, modelling and managing. *Annual Reviews in Control*, 52: 448–464.
- K. Ampountolas, A. McHugh, and V. Thakuriah. 2018. Developing smart statistics for urban mobility: challenges and opportunities. *10442/15809*, p. 14. <https://www.statistics.gr/event/smart-statistics/docs/presentations/SS4SC-Session%20I-Presentation%20by%20Ampountolas.pdf>.
- G. Andrienko, N. Andrienko, C. Boldrini, G. Caldarelli, P. Cintia, S. Cresci, A. Facchini, F. Giannotti, A. Gionis, R. Guidotti, et al. 2021. (so) big data and the transformation of the city. *International Journal of Data Science and Analytics*, 11(4): 311–340.
- C. Angell and D. Potoglou. 2022. An insight into the impacts of covid-19 on work-related travel behaviours in the cardiff capital region and following the uk’s first national lockdown. *Cities*, 124: 103602. <https://www.sciencedirect.com/science/article/pii/S0264275122000415>.
- Apple Inc. Mobility trends reports. <https://covid19.apple.com/mobility>.
- Apple Inc., 2020. Apple covid-19 mobility trends point to weak recovery in transportation activities. <https://knoema.com/vldnawd/apple-covid-19-mobility-trends-point-to-weak-recovery-in-transportation-activities>.
- A. Arvanitis, I. Furxhi, T. Thomas, and K. Karatzas. 2021. Prediction of the effective reproduction number of covid-19 in greece. a machine learning approach using google mobility data. *medRxiv*, pp. 2021–05.
- M. E. C. Bagdatli and F. Ipek. 2022. Transport mode preferences of university students in post-covid-19 pandemic. *Transport Policy*, 118: 20–32. ISSN 0967-070X. <https://www.sciencedirect.com/science/article/pii/S0967070X22000233>.
- R. Beekman, 2014. Thessaloniki: Sustainable urban mobility planning in times of crisis (greece). <https://www.eltis.org/discover/case-studies/thesaloniki-sustainable-urban-mobility-planning-times-crisis-greece>.
- R. Brandom. 2020. Answering the 12 biggest questions about apple and google’s new coronavirus tracking project. *The Verge*. <https://www.theverge.com/2020/4/11/2121680>

78 ΒΙΒΛΙΟΓΡΑΦΙΑ

- 3/apple-google-coronavirus-tracking-app-covid-bluetooth-secure.
- D. Busvine, 2020. European experts ready smartphone technology to help stop coronavirus. <https://www.reuters.com/article/us-health-coronavirus-europe-tech/european-experts-ready-smartphone-technology-to-help-halt-coronavirus-spread-idUSKBN21J4HI>.
- A. Caporali, J. Garcia, É. Couppié, S. Poniakina, M. Barbieri, F. Bonnet, C. G. Camarda, E. Cambois, I. Hourani, D. Korotkova, et al. 2022. The demography of covid-19 deaths database, a gateway to well-documented international data. *Scientific Data*, 9(1): 1–9.
- C. Cavallaro, A. Bujari, L. Foschini, G. Di Modica, and P. Bellavista. 2021. Measuring the impact of covid-19 restrictions on mobility: A real case study from italy. *Journal of Communications and Networks*, 23(5): 340–349.
- S. Chang, E. Pierson, P. W. Koh, J. Gerardin, B. Redbird, D. Grusky, and J. Leskovec. 2021. Mobility network models of covid-19 explain inequities and inform reopening. *Nature*, 589(7840): 82–87.
- H. Cho, D. Ippolito, and Y. W. Yu. 2020. Contact tracing mobile apps for covid-19: Privacy considerations and related trade-offs. *arXiv preprint arXiv:2003.11511*.
- P. Christidis, B. Ciuffo, and M. Vespe. 2022. Regional mobility during the covid-19 pandemic: Analysis of trends and repercussions using mobile phones data across the eu. *Case Studies on Transport Policy*, 10(1): 257–268.
- D. Cooper, K. Quathem, and A. Meneses, 2020. Covid-19 apps and websites—the “pan-european privacy preserving proximity tracing initiative” and guidance by supervisory authorities. <https://www.insideprivacy.com/covid-19/covid-19-apps-and-websites-the-pan-european-privacy-preserving-proximity-tracing-initiative-and-guidance-by-supervisory-authorities/>.
- Covid-19 Response Greece. Datasets and analysis of novel coronavirus (covid-19) outbreak in greece. <https://github.com/Covid-19-Response-Greece/covid19-data-greece>.
- R. Dandekar and G. Barbastathis. 2020. Quantifying the effect of quarantine control in covid-19 infectious spread using machine learning. *MedRxiv*.
- J. Doe, 2021. The sump concept. <https://www.eltis.org/mobility-plans/sump-concept>.
- F. Duarte and P. deSouza. 2020. Data Science and Cities: A Critical Approach. *Harvard Data Science Review*, 2(3). <https://hdsr.mitpress.mit.edu/pub/lum18ajd>.
- EIT Urban Mobility, 2021a. Ai mobility landscape in the eu. https://www.eiturbanmobility.eu/wp-content/uploads/2021/09/EITUM-UrbanMobilityNext3_Final.pdf.
- EIT Urban Mobility, 2021b. Short version: urban mobility strategies during covid-19. https://www.eiturbanmobility.eu/wp-content/uploads/2021/03/Urban-mobility-strategies-during-COVID-19_short-2.pdf.
- European Institute of Innovation and Technology, 2022. Eit urban mobility. <https://www.eiturbanmobility.eu/>.
- H. Figg, 2021. Simps: Islands propose a new approach to the sump concept. <https://www.eltis.org/resources/case-studies/simps-islands-propose-new-approach-sump-concept>.

- H. Georgiou, May 2020a. Greece is past its COVID-19 epidemic peak, now entering asymptotic decay and ending its 42-day country-wide lockdown. <https://doi.org/10.5281/zenodo.3819558>.
- H. V. Georgiou. 2020b. Covid-19 outbreak in greece has passed its rising inflection point and stepping into its peak. *medRxiv*.
- Google LLC. Αναφορά κινητικότητας κοινότητας covid-19. <https://www.google.com/covid19/mobility/>.
- Google LLC, 2022. Αναφορά κινητικότητας κοινότητας για τον covid-19 (10 Οκτωβρίου). https://www.gstatic.com/covid19/mobility/2022-10-15_GR_Mobility_Report_el.pdf.
- Google LLC, 2023. Apple covid-19 mobility trends point to weak recovery in transportation activities. https://support.google.com/covid19-mobility/answer/9824897?hl=el&ref_topic=9822927.
- S. M. Iacus, C. Santamaria, F. Sermi, S. Spyros, D. Tarchi, and M. Vespe. 2020. Human mobility and covid-19 initial dynamics. *Nonlinear Dynamics*, 101(3): 1901–1919.
- iMEdD-Lab. Greece covid-19 dataset. <https://github.com/iMEdD-Lab/open-data/tree/master/COVID-19>.
- P. H. Kindt, T. Chakraborty, and S. Chakraborty. 2021. How reliable is smartphone-based electronic contact tracing for covid-19? *Communications of the ACM*, 65(1): 56–67.
- D. P. Kingma and J. Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv: 1412.6980*.
- M. Kotsch, 2020. What mobility data has been collected and published during covid-19? <https://theodi.org/article/what-mobility-data-has-been-collected-and-published-during-covid-19/>.
- B. P. Loo and Z. Huang. 2022. Spatio-temporal variations of traffic congestion under work from home (wfh) arrangements: Lessons learned from covid-19. *Cities*, 124: 103610. <https://www.sciencedirect.com/science/article/pii/S026427512200049X>.
- M. Miskolczi, D. Földes, A. Munkácsy, and M. Jászberényi. 2021. Urban mobility scenarios until the 2030s. *Sustainable Cities and Society*, 72: 103029.
- T. Mourey, 2015. Monitoring and evaluation at the very core of thessaloniki's sump (greece). <https://www.eltis.org/discover/case-studies/monitoring-and-evaluation-ver-y-core-thessalonikis-sump-greece>.
- M. Nanni, G. Andrienko, A.-L. Barabási, C. Boldrini, F. Bonchi, C. Cattuto, F. Chiaromonte, G. Comandé, M. Conti, M. Coté, et al. 2021. Give more data, awareness and control to individual citizens, and they will help covid-19 containment. *Ethics and Information Technology*, 23(1): 1–6.
- Net4things, 2021. Improving urban mobility with data analytics. <https://www.net4things.com/improving-urban-mobility-with-data-analytics/?cn-reloaded=1>.
- E. Noi, A. Rudolph, and S. Dodge. 2021. A novel method for mapping spatiotemporal structure of mobility patterns during the covid-19 pandemic. *GIScience 2021 Short Paper Proceedings*.

80 ΒΙΒΛΙΟΓΡΑΦΙΑ

- R. B. Noland. 2021. Mobility and the effective reproduction rate of covid-19. *Journal of Transport & Health*, 20: 101016.
- P. Nouvellet, S. Bhatia, A. Cori, K. E. Ainslie, M. Baguelin, S. Bhatt, A. Boonyasiri, N. F. Brazeau, L. Cattarino, L. V. Cooper, et al. 2021. Reduction in mobility and covid-19 transmission. *Nature communications*, 12(1): 1090.
- M. Z. Paköz and M. İşık. 2022. Rethinking urban density, vitality and healthy environment in the post-pandemic city: The case of istanbul. *Cities*, 124: 103598. ISSN 0264-2751. <https://www.sciencedirect.com/science/article/pii/S0264275122000373>.
- L. Peebles et al. 2022. Lessons from the covid data wizards. *Nature*, 603(7902): 564–567.
- E. Pepe, P. Bajardi, L. Gauvin, F. Privitera, B. Lake, C. Cattuto, and M. Tizzoni, 2020. Covid-19 mobility monitoring project.
- M. Ponce-de Leon, J. del Valle, J. M. Fernandez, M. Bernardo, D. Cirillo, J. Sanchez-Valle, M. Smith, S. Capella-Gutierrez, T. Gullón, and A. Valencia. 2021. Covid-19 flow-maps: An open geographic information system on covid-19 and human mobility for spain. *medRxiv*, 8(1): 1–16.
- L. Prechelt. 2002. Early stopping-but when? In *Neural Networks: Tricks of the trade*, pp. 55–69. Springer.
- A. Robertson. 2020. How you'll use apple and google's coronavirus tracking tool. *The Verge*. <https://www.theverge.com/2020/4/10/21216715/apple-google-coronavirus-covid-19-contact-tracing-app-details-use>.
- Sandbird. Daily regional statistics for covid19 cases in greece. <https://github.com/Sandbird/covid19-Greece>.
- A. Strzelecki. 2022. The apple mobility trends data in human mobility patterns during restrictions and prediction of covid-19: A systematic review and meta-analysis. In *Healthcare*, volume 10, p. 2425. MDPI.
- P. Tang, M. Steinbach, A. Karpatne, and V. Kumar. 2020. *Εισαγωγή στην εξόρυξη δεδομένων*, chapter 2.3, 2.4, 3.3, 4.7, 4.8, 4.9, 4.10. Εκδόσεις Τζιόλα, 2.
- S. Theodoridis and K. Koutroumbas. 2012. *Ανανώριση Προτύπων*, chapter Γραμμικοί ταξινομητές. Μη γραμμικοί ταξινομητές, Επιλογή χαρακτηριστικών, Γέννηση Χαρακτηριστικών : Μετασχηματισμοί Δεδομένων και Μείωση Διάστασης. Broken Hill Publishers LTD.
- I. Vlahavas, P. Kefalas, N. Bassiliades, F. Kokkoras, and Sakellariou. 2020. *Τεχνητή νοημοσύνη*, chapter Μηχανική μάθηση. Τεχνητά νευρωνικά δίκτυα, Ανακάλυψη γνώσης σε βάσεις δεδομένων. Εκδόσεις Πανεπιστημίου Μακεδονίας, Εγνατία 156, 54006 Θεσσαλονίκη, 4.
- O. Wahltinez, A. Cheung, R. Alcantara, D. Cheung, M. Daswani, A. Erlinger, M. Lee, P. Yawalkar, P. Lê, O. P. Navarro, et al. 2022. Covid-19 open-data a global-scale spatially granular meta-dataset for coronavirus disease. *Scientific data*, 9(1): 1–11.
- World Health Organization. Coronavirus disease (covid-19). https://www.who.int/health-topics/coronavirus#tab=tab_1.

- H. Zhongwen, 2018. Data analytics in urban planning: New tools for old problems.
<https://www.brinknews.com/data-analytics-in-urban-planning-new-tools-for-old-problems/>. interview.