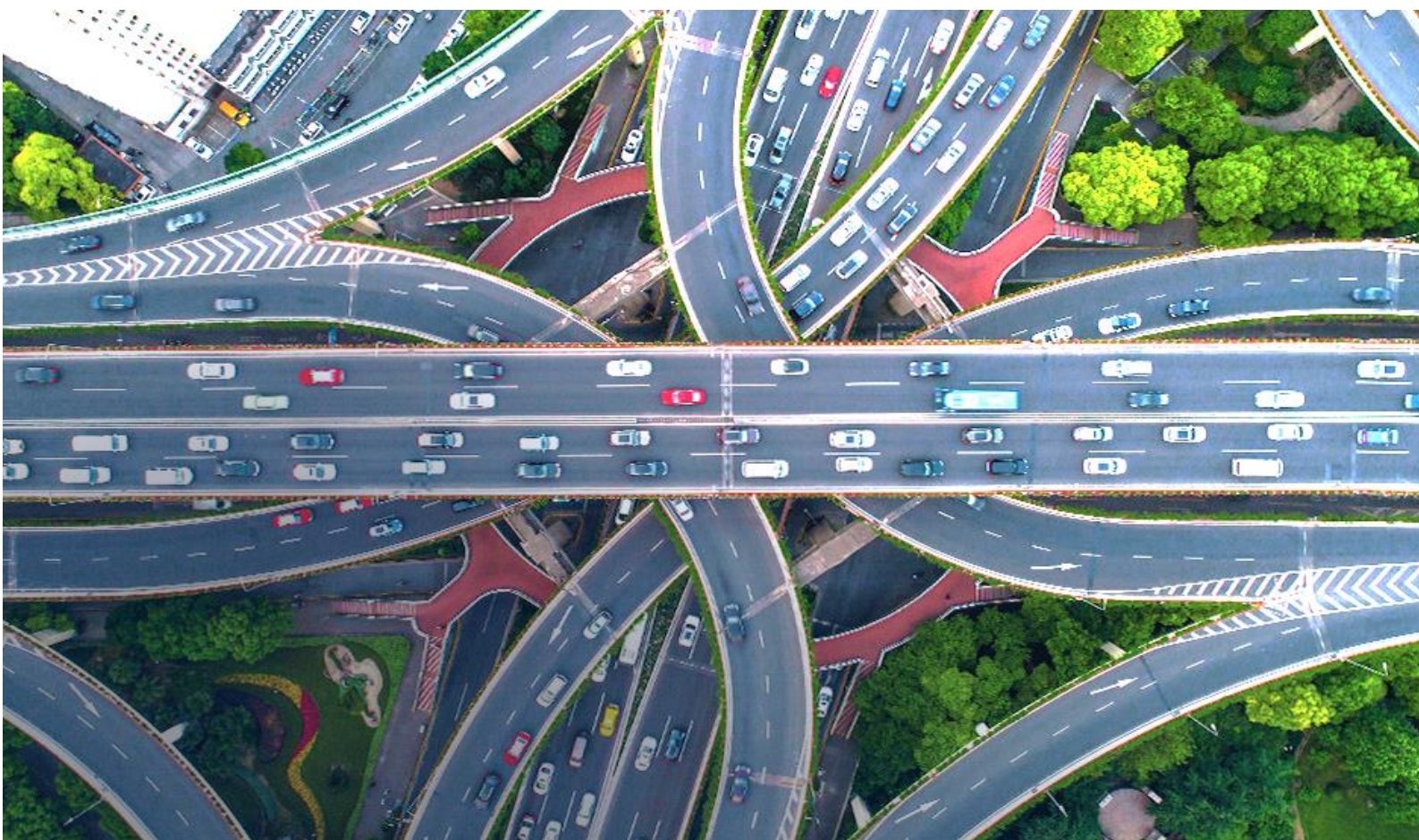




ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΠΟΛΙΤΙΚΩΝ ΜΗΧΑΝΙΚΩΝ
ΤΟΜΕΑΣ ΜΕΤΑΦΟΡΩΝ ΚΑΙ ΣΥΓΚΟΙΝΩΝΙΑΚΗΣ ΥΠΟΔΟΜΗΣ

Πρότυπα Πρόβλεψης Κυκλοφορίας με Χωρο-Χρονικές Συσχετίσεις



Διπλωματική Εργασία

Ευαγγελία Στεφανή

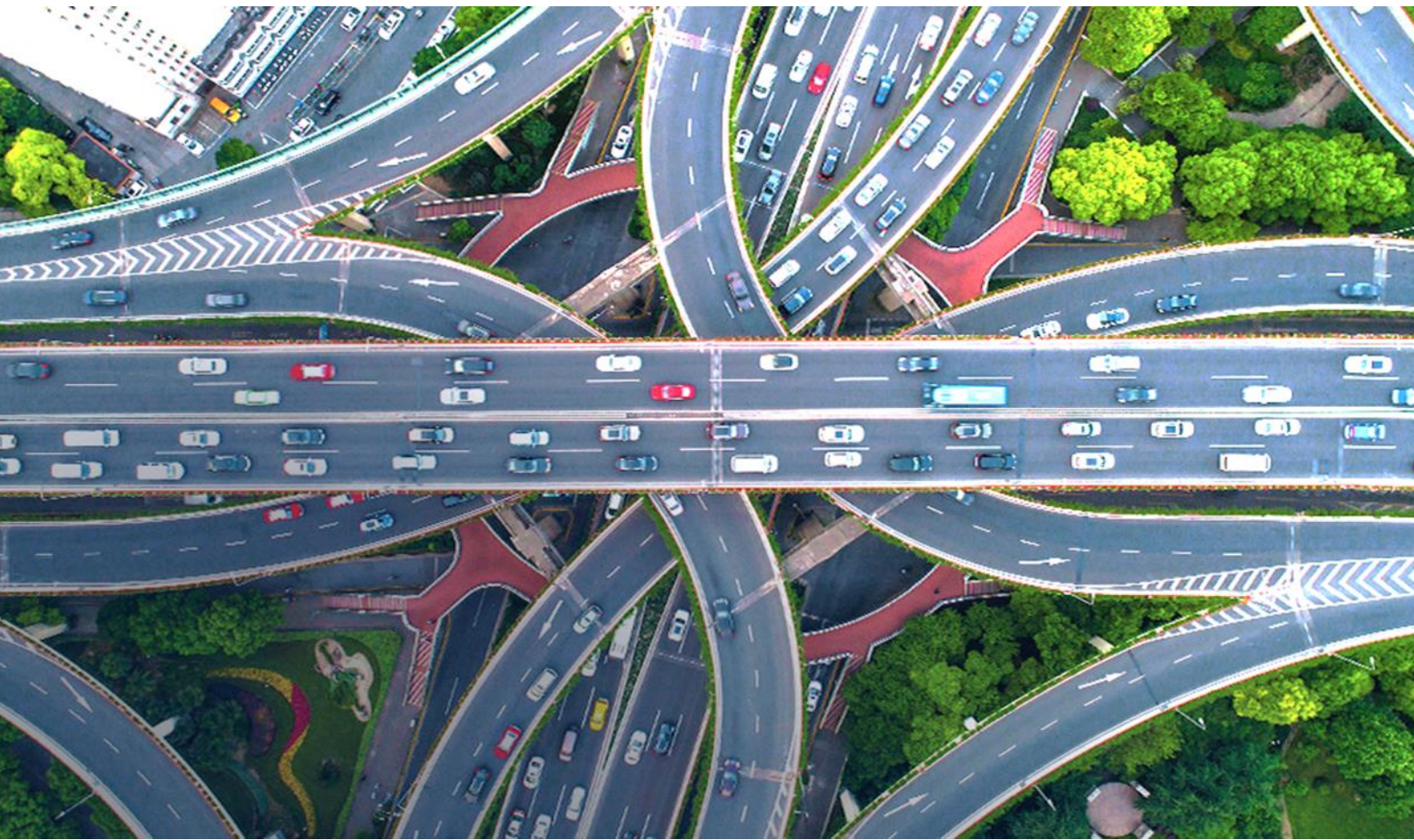
Επιβλέπουσα Καθηγήτρια : Ελένη Βλαχογιάννη,
Αναπληρώτρια Καθηγήτρια Σχολής Πολιτικών Μηχανικών ΕΜΠ

Αθήνα, Οκτώβριος 2021



National Technical University of Athens
School of Civil Engineering
Department of Transportation Planning and Engineering

Spatiotemporal Traffic Forecasting Methods



Diploma Thesis

Evangelia Stefani

Supervising Professor: Eleni Vlahogianni,
Associate Professor of the School of Civil Engineering, NTUA

Athens, October 2021

Εθνικό Μετσόβιο Πολυτεχνείο

Σχολή Πολιτικών Μηχανικών

Τομέας Μεταφορών και Συγκοινωνιακής Υποδομής

Πρότυπα Πρόβλεψης Κυκλοφορίας με Χωρο-Χρονικές Συσχετίσεις

Ευαγγελία Κ. Στεφανή

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

Αθήνα, 2021

National Technical University of Athens

School of Civil Engineering

Department of Transportation Planning and Engineering

Spatiotemporal Traffic Forecasting Methods

Thesis Author: Evangelia K. Stefani

Supervising Professor: Eleni I. Vlahogianni, Associate Professor NTUA

Athens, 2021

Copyright © Ευαγγελία Στεφανή, 2021

Με επιφύλαξη παντός δικαιώματος

Απαγορεύεται η αντιγραφή, αποθήκευση σε αρχείο πληροφοριών, διανομή, αναπαραγωγή, μετάφραση ή μετάδοση της παρούσας εργασίας, εξ ολοκλήρου ή τμήματος αυτής, για εμπορικό σκοπό, υπό οποιαδήποτε μορφή και με οποιοδήποτε μέσο επικοινωνίας, ηλεκτρονικό ή μηχανικό, χωρίς την προηγούμενη έγγραφη άδεια του συγγραφέα. Επιτρέπεται η αναπαραγωγή, αποθήκευση και διανομή για σκοπό μη κερδοσκοπικό, εκπαιδευτικής ή ερευνητικής φύσης, υπό την προϋπόθεση να αναφέρεται η πηγή προέλευσης και να διατηρείται το παρόν μήνυμα. Ερωτήματα που αφορούν στη χρήση της εργασίας για κερδοσκοπικό σκοπό πρέπει να απευθύνονται προς τον συγγραφέα.

Η έγκριση της διπλωματικής εργασίας από τη Σχολή Πολιτικών Μηχανικών του Εθνικού Μετσόβιου Πολυτεχνείου (ΕΜΠ) δεν υποδηλώνει αποδοχή των απόψεων του συγγραφέα (Ν. 5343/1932, Άρθρο 202).

Copyright © Evangelia Stefani 2021

All Rights Reserved

Neither the whole nor any part of this diploma thesis may be copied, stored in a retrieval system, distributed, reproduced, translated, or transmitted for commercial purposes, in any form or by any means now or hereafter known, electronic or mechanical, without the written permission from the author. Reproducing, storing and distributing this thesis for non-profitable, educational or research purposes is allowed, without prejudice to reference to its source and to inclusion of the present text. Any queries in relation to the use of the present thesis for commercial purposes must be addressed to its author.

Approval of this diploma thesis by the School of Civil Engineering of the National Technical University of Athens (NTUA) does not constitute in any way an acceptance of the views of the author contained herein by the said academic organization (L. 5343/1932, art. 202).

Ευχαριστίες

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

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

Επίσης, ευχαριστώ τον υποψήφιο Διδάκτορα του Εθνικού Μετσόβιου Πολυτεχνείου, Παναγιώτη Φαφουτέλλη για τις καθοριστικές συμβουλές που μου προσέφερε στην εκπόνηση της διπλωματικής μου εργασίας.

Τέλος, ένα μεγάλο ευχαριστώ στους γονείς μου, Κώστα και Ευγενία, και την αδερφή μου Κατερίνα, για τη συνεχή υποστήριξη που μου πρόσφεραν καθ' όλη τη διάρκεια των σπουδών μου, καθώς και στους φίλους μου, Μαρίνο και Ορέστη που στέκονται πάντα δίπλα μου και με ενθαρρύνουν σε κάθε μου βήμα.

Αθήνα, Οκτώβριος 2021

Τίτλος: Πρότυπα Πρόβλεψης Κυκλοφορίας με Χωρο-Χρονικές Συσχετίσεις

Ευαγγελία Κ. Στεφανή

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

Σύνοψη

Η πρόβλεψη μελλοντικών συνθηκών κυκλοφορίας έχει μεγάλη σημασία για τη βελτίωση των υπαρχόντων Ευφών Συστημάτων Μεταφορών (ITS). Η παρούσα διπλωματική εργασία διερευνά την εφαρμοσιμότητα και την ακρίβεια ενός προτύπου Μακράς Βραχυπρόθεσμης Μνήμης μοντέλου (LSTM) που εμπλουτίζεται με χωροχρονική πληροφορία συσχέτισης μεταξύ διαφορετικών σημείων του δικτύου. Η πληροφορία της συσχέτισης προέρχεται από την εφαρμογή μετρικών της Θεωρίας Αμοιβαίας Πληροφορίας. Το πρότυπο βραχυπρόθεσμης πρόβλεψης αναπτύχθηκε χρησιμοποιώντας δεδομένα κυκλοφορίας και δημόσιων συγκοινωνιών από το κέντρο της Αθήνας. Τα αποτελέσματα δείχνουν ότι το προτεινόμενο μοντέλο παρείχε υψηλότερη ακρίβεια πρόβλεψης για βραχυπρόθεσμες προβλέψεις όταν ελήφθησαν υπόψη οι μη γραμμικές χωροχρονικές συσχέτισεις. Επίσης, με πιο προηγμένη βελτιστοποίηση των υπερπαραμέτρων και μεθόδων μηχανικής χαρακτηριστικών μπορεί να βελτιωθεί περαιτέρω το μοντέλο μηχανικής μάθησης.

Λέξεις Κλειδιά: Βραχυπρόθεσμη Πρόβλεψη Κυκλοφορίας, Κυκλοφοριακός Φόρτος, Μηχανική Μάθηση, Αμοιβαία Πληροφορία, Μοντέλο Μακράς Βραχυπρόθεσμης Μνήμης.

Title: Spatiotemporal Traffic Forecasting Methods

Evangelia K. Stefani

Supervising Professor: Eleni I. Vlahogianni, Associate Professor NTUA

Abstract

Forecasting future traffic situations is of great importance for improving existing Intelligent Transportation Systems (ITS). This thesis explores the applicability and accuracy of a Long-Short Term Model (LSTM), which is enhanced with spatio-temporal correlation information based on Information Theoretic metrics. The model is developed and trained using traffic and public transport data from central Athens. Results indicate that the LSTM model provided a higher forecasting accuracy for short-term predictions when nonlinear spatio-temporal correlations were taken into account. Also, more advanced hyper parameter optimization and feature engineering methods would further improve the machine learning model.

Keywords: Short-term Traffic Forecasting, Traffic Flow, Machine Learning, Mutual Information, Long-Short Term Model.

ΠΕΡΙΛΗΨΗ

Η πρόβλεψη κυκλοφοριακών συνθηκών αποτελεί ένα συνεχώς εξελισσόμενο ερευνητικό θέμα, που ενισχύεται από τις τεράστιες πρόσφατες εξελίξεις της μηχανικής μάθησης (Machine Learning) και της βαθιάς μάθησης (Deep Learning), καθώς και από την υπολογιστική ισχύ των σύγχρονων υπολογιστών.

Η φύση της κυκλοφορίας, όπως αντικατοπτρίζεται στη μελέτη των μεταβλητών κίνησης, όπως η ροή, ο φόρτος και η ταχύτητα, είναι εξαιρετικά περίπλοκη. Τα χρονικά μοτίβα της κυκλοφορίας δεν διακρίνονται εύκολα, ειδικά όταν μελετάται η βραχυπρόθεσμη συμπεριφορά της. Επιπλέον, οι σχετικά συχνές μετατοπίσεις σε οριακές συνθήκες, όπως η έναρξη της συμφόρησης, δημιουργούν ένα δύσκολο πρόβλημα μοντελοποίησης. Βασικό ζήτημα στην πρόβλεψη κυκλοφορίας αποτελεί η αποκωδικοποίηση αυτής της σύνθετης συμπεριφοράς προκειμένου να βελτιωθεί η προβλεψιμότητα της.

Οι μέθοδοι μηχανικής μάθησης, παρά το γεγονός ότι αδυνατούν να εξηγήσουν ποιοτικά τις συσχετίσεις που μοντελοποιούν, παρουσιάζουν μεγαλύτερη ευελιξία καθώς επιτρέπουν την εισαγωγή περισσότερων παραμέτρων. Για αυτό το λόγο στην παρούσα Διπλωματική Εργασία διερευνάται η εφαρμοσιμότητα και η ακρίβεια ενός προτύπου Μακράς Βραχυπρόθεσμης Μνήμης μοντέλου (LSTM) που εμπλουτίζεται με χωροχρονική πληροφορία συσχέτισης μεταξύ διαφορετικών σημείων του οδικού δικτύου.

Τα διαθέσιμα δεδομένα δημοσιεύονται σε τακτικά ενημερωμένες δημόσιες βιβλιοθήκες (open data) και προέρχονται από τοποθετημένους αισθητήρες στους δρόμους της Αθήνας. Πιο συγκεκριμένα, χρησιμοποιώντας δεδομένα κυκλοφορίας I.X. και δεδομένα επικυρωμένων εισιτηρίων επιβατών ΟΑΣΑ από το Μάρτιο έως τον Ιούνιο του 2021, επιδιώκεται η πρόβλεψη του κυκλοφοριακού φόρτου σε ένα οδικό τμήμα στο αστικό δίκτυο μελέτης μέσω του συσχετισμού του με την επιβατική ζήτηση σε στάσεις ΣΤΑ.ΣΥ. και με άλλους αισθητήρες καταγραφής κυκλοφορίας οχημάτων.

Στην παρούσα εργασία γίνεται διερεύνηση μίας μεθόδου εντοπισμού χωροχρονικών συσχετίσεων που καθορίζουν τη λειτουργία ενός αστικού οδικού δικτύου. Πιο συγκεκριμένα, χρησιμοποιούνται μετρικές ευρύτερης χρησιμότητας, όπως η γραμμική συσχέτιση κατά Pearson αλλά και πιο εξειδικευμένες μετρικές χρονοσειρών όπως η Αμοιβαία Πληροφορία, για να προσδιοριστούν οι χωροχρονικές συσχετίσεις στο οδικό δίκτυο. Αφού εντοπιστούν και απεικονιστούν σε θερμικούς χάρτες αυτές οι συσχετίσεις, επιλέγεται ένας φωρατής τοποθετημένος σε ένα κεντρικό σημείο του αστικού οδικού δικτύου μελέτης για τον οποίο θα αναπτυχθεί ένα Μακράς Βραχυπρόθεσμης μνήμης μοντέλο πρόβλεψης (LSTM). Έπειτα επιλέγονται οι παράμετροι που θα χρησιμοποιηθούν στο δίκτυο LSTM, οι οποίες είναι στάσεις

μέσων μαζικής μεταφοράς και φωρατές τοποθετημένοι στο υπόλοιπο οδικό δίκτυο, με τους οποίους ο φωρατής – στόχος (target) εμφανίζει υψηλή συσχέτιση.

Αξιοποιώντας τις συσχετίσεις που προκύπτουν, αναπτύσσεται ένα δίκτυο LSTM με στόχο την πρόβλεψη του κυκλοφοριακού φόρτου του επιλεγμένου φωρατή με χρονικό ορίζοντα πρόβλεψης μία ώρα μπροστά και λαμβάνοντας υπόψη τις 6 προηγούμενες παρατηρήσεις. Το σφάλμα που προκύπτει είναι της τάξης του 15% και θεωρείται ικανοποιητικό λαμβάνοντας υπόψη τη βιβλιογραφία αλλά και το γεγονός πως ο όγκος των δεδομένων είναι σχετικά μικρός επειδή χρησιμοποιήθηκαν ακατέργαστα ανοιχτά δεδομένα (raw open data) από μια καινούργια ελληνική δημόσια βιβλιοθήκη δεδομένων.

Το κυριότερο συμπέρασμα που προέκυψε από την ανάπτυξη του συγκεκριμένου μοντέλου μηχανικής μάθησης ήταν πως ο εμπλουτισμός του δικτύου LSTM με χωροχρονική πληροφορία συσχέτισης μεταξύ των παραμέτρων του, μειώνει τη διαστατικότητα του προβλήματος βοηθώντας την καλή λειτουργία του αλγορίθμου και την αποδοτικότητα του υπολογιστικού συστήματος.

Για μελλοντική έρευνα προτείνεται η διερεύνηση και η ανάπτυξη μεθοδολογίας που θα περιλαμβάνει την ανάλυση μεγαλύτερου όγκου δεδομένων και την αξιοποίηση πιο εξειδικευμένων τεχνικών αναγνώρισης χωροχρονικών συσχετίσεων και πρόβλεψης κυκλοφοριακών συνθηκών που απαιτούν ισχυρότερη υπολογιστική ισχύ.

ΠΕΡΙΕΧΟΜΕΝΑ

Σύνοψη	i
Abstract.....	ii
ΠΕΡΙΛΗΨΗ.....	iii
Κεφάλαιο 1	1
Εισαγωγή.....	1
1.1. Γενική Ανασκόπηση.....	1
1.1.1. Κυκλοφοριακή Συμφόρηση σε Μεγάλα Αστικά Κέντρα	1
1.1.2. Ευφυή Συστήματα Μεταφορών.....	3
1.1.3. Βραχυπρόθεσμη Πρόβλεψη Κυκλοφορίας	5
1.2. Αντικείμενο Διπλωματικής.....	5
1.2.1. Συνεισφορά Παρούσας Εργασίας	6
1.3. Οργάνωση Εργασίας.....	8
Κεφάλαιο 2	9
Βιβλιογραφική Ανασκόπηση.....	9
2.1 Εισαγωγή.....	9
2.2 Μεθοδολογίες Πρόβλεψης Κυκλοφοριακών Μεγεθών.....	9
2.2.1 Παραμετρικές και Μη Παραμετρικές Προσεγγίσεις.....	11
2.3 Βαθιά Μηχανική Μάθηση στην Πρόβλεψη Κυκλοφορίας	12
2.4 Η Συμβολή της Αμοιβαίας Πληροφορίας.....	14
2.5 Συμπεράσματα Βιβλιογραφικής Ανασκόπησης	15
Κεφάλαιο 3	17
Μεθοδολογική Προσέγγιση.....	17
3.1 Προτεινόμενη Μεθοδολογία.....	17

3.2	Θεωρητικό Υπόβαθρο	18
3.2.1	Θεωρία Πληροφορίας	18
3.2.2	Αμοιβαία Πληροφορία	19
3.2.3	Επιλογή Χαρακτηριστικών	21
3.2.4	Δίκτυα Μακράς-Βραχείας Μνήμης.....	23
3.2.5	Αξιολόγηση Μοντέλου Πρόβλεψης.....	26
3.2.6	Multiplex Δίκτυα	29
	Κεφάλαιο 4	30
	Ανάλυση και Αποτελέσματα	30
4.1	Εισαγωγή	30
4.1.1	Εγκατάσταση πακέτων	30
4.2	Βάση Δεδομένων	31
4.2.1	Περιγραφή Αρχικής Βάσης Δεδομένων.....	31
4.1.3	Προ – Επεξεργασία Βάσης Δεδομένων	33
4.3	Επεξεργασία Δεδομένων	35
4.3.1	Εφαρμογή Μεθόδων Συσχέτισης και Αποτελέσματα Συσχετίσεων.....	35
4.3.2	Προετοιμασία Δεδομένων για το δίκτυο LSTM.....	44
4.3.3	Αρχιτεκτονική μοντέλου LSTM.....	45
4.4	Αποτελέσματα LSTM μοντέλου.....	48
	Κεφάλαιο 5	52
	Συμπεράσματα και Προτάσεις	52
5.1	Γενικά Συμπεράσματα.....	52
5.2	Βασικά Συμπεράσματα.....	53
5.3	Προτάσεις για περαιτέρω έρευνα	53
	Βιβλιογραφία	55

Παράρτημα Α – Κώδικας.....	58
Παράρτημα Β	63

ΕΥΡΕΤΗΡΙΟ ΕΙΚΟΝΩΝ

Εικόνα 1: Κυκλοφοριακή Συμφόρηση σε πόλεις παγκοσμίως.	2
Εικόνα 2: Δακτύλιος Αθηνών.	7
Εικόνα 3: Η σχέση της Θεωρίας Πληροφορίας με τις υπόλοιπες επιστήμες	18
Εικόνα 4: Αρχική Ακατέργαστη Βάση Δεδομένων που παρέχει η Περιφέρεια Αττικής....	31
Εικόνα 5: Αρχική Ακατέργαστη Βάση Δεδομένων που παρέχει ο ΟΑΣΑ.	32
Εικόνα 6: Επεξεργασμένη - Ενοποιημένη Βάση Δεδομένων.	34
Εικόνα 7: Θέσεις φωρατών και στάσεων που χρησιμοποιούνται στο μοντέλο πρόβλεψης.	43

ΕΥΡΕΤΗΡΙΟ ΣΧΗΜΑΤΩΝ

Σχήμα 1: Σχηματική απεικόνιση της διαδικασίας Επιλογής Χαρακτηριστικών.....	22
Σχήμα 2: Επαναλαμβανόμενη δομή σε ένα τυπικό RNN, πηγή [25]......	23
Σχήμα 3: Η πύλη λήθης (forget gate) του δικτύου LSTM.	24
Σχήμα 4: Η πύλη εισόδου (input gate) του δικτύου LSTM.	24
Σχήμα 5: Η πύλη εξόδου (output gate) του δικτύου LSTM.	25
Σχήμα 6: Αναπαράσταση της τυπικής δομής του δικτύου LSTM, πηγή [25]......	25
Σχήμα 7: Αναπαράσταση ενός Πολυπλεξικού Δικτύου, πηγή [29].	29

ΕΥΡΕΤΗΡΙΟ ΔΙΑΓΡΑΜΜΑΤΩΝ

Διάγραμμα 1: Ροή Κεφαλαίων Εργασίας.....	8
Διάγραμμα 2: Μεθοδολογική Προσέγγιση της κατασκευής διανύσματος κατάστασης κίνησης μέσω MI για τη βραχυπρόθεσμη πρόβλεψη κυκλοφοριακής ροής, πηγή [22].	14
Διάγραμμα 3: Διάγραμμα Ροής Εργασιών	17
Διάγραμμα 4: Διάγραμμα Venn μεταξύ αμοιβαίας πληροφορίας και εντροπίας	20
Διάγραμμα 5: Διαχωρισμός Χρονοσειρών.....	26
Διάγραμμα 6: Θερμικός Χάρτης Συσχέτισης Κατά Pearson για τα δεδομένα κυκλοφορίας Ι.Χ.36	
Διάγραμμα 7: Θερμικός Χάρτης Συσχέτισης Κατά Pearson για τα δεδομένα των στάσεων ΣΤΑ.ΣΥ. εντός του δακτυλίου Αθηνών.	37
Διάγραμμα 8: Θερμικός χάρτης κατά Pearson για την Ενοποιημένη βάση δεδομένων.	38
Διάγραμμα 9: Θερμικός χάρτης Αμοιβαίας Πληροφορίας για τα δεδομένα κυκλοφορίας.	39
Διάγραμμα 10: Θερμικός χάρτης Αμοιβαίας Πληροφορίας για τις επιλεγμένες στάσεις ΣΤΑ.ΣΥ εντός του δακτυλίου Αθηνών.....	40
Διάγραμμα 11: Θερμικός χάρτης Αμοιβαίας Πληροφορίας για την ενοποιημένη βάση δεδομένων.	41
Διάγραμμα 12: Διάγραμμα Διασποράς Πραγματικών και Προβλεπόμενων Τιμών.	49
Διάγραμμα 13: Διάγραμμα Χρονοσειράς Πραγματικών και Προβλεπόμενων.....	51
Διάγραμμα 14: Διάγραμμα χρονοσειρών φωρατή-στόχου και επιλεγμένων φωρατών και σταθμών μετρό.	63

Κεφάλαιο 1

Εισαγωγή

1.1. Γενική Ανασκόπηση

1.1.1. Κυκλοφοριακή Συμφόρηση σε Μεγάλα Αστικά Κέντρα

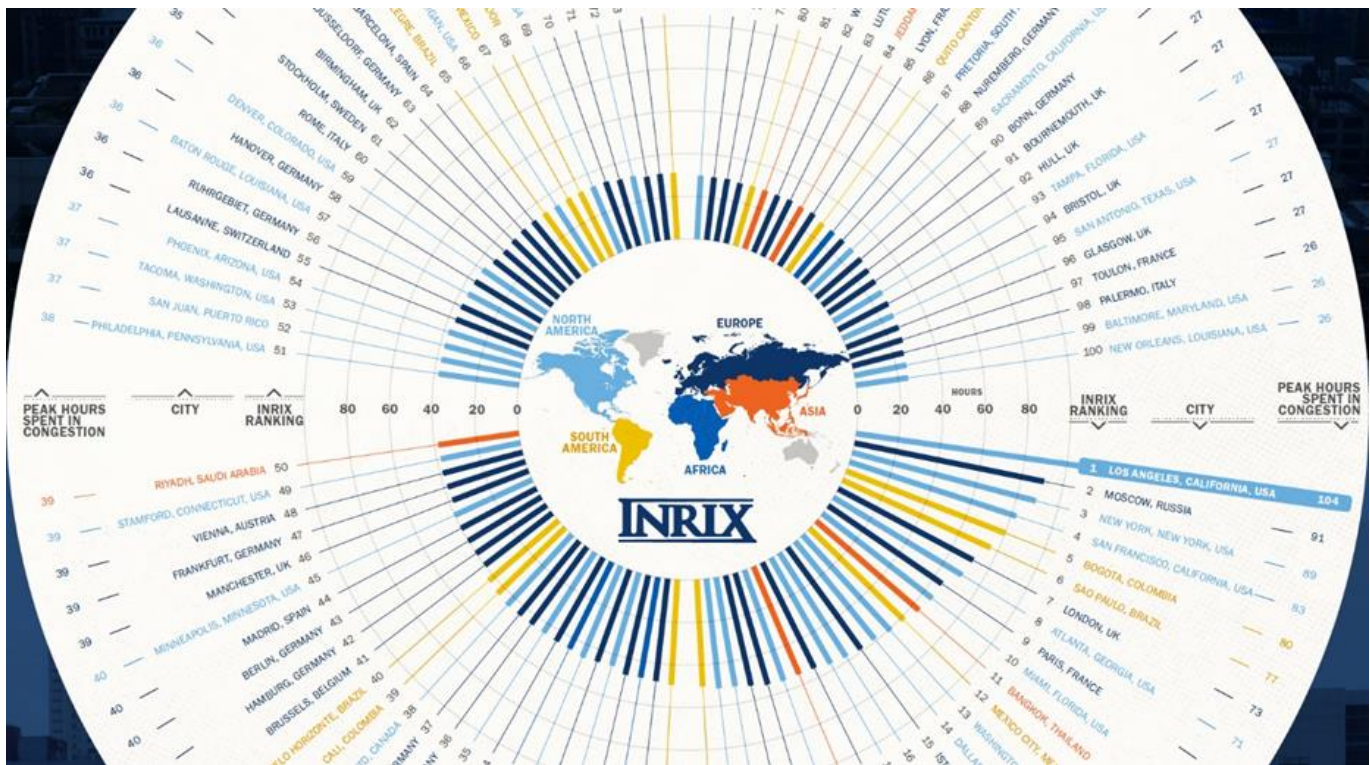
Ένα κοινό πρόβλημα σε οποιοδήποτε οδικό δίκτυο σύγχρονων μητροπολιτικών κέντρων είναι αυτό της κυκλοφοριακής συμφόρησης (traffic congestion). Συμφόρηση της κυκλοφορίας συμβαίνει όταν το πλήθος των μετακινούμενων οχημάτων υπερβαίνει την υφιστάμενη χωρητικότητα του συγκοινωνιακού συστήματος. Η υπέρβαση της κυκλοφοριακής ικανότητας ενός οδικού δικτύου μπορεί να είναι καθημερινή και σε συγκεκριμένες ώρες της ημέρας (recurrent congestion) ή μη αναμενόμενη (non-recurrent congestion) εξαιτίας απρόβλεπτων συμβάντων όπως για παράδειγμα ατυχημάτων, εργασιών στο οδικό δίκτυο ή έντονων καιρικών φαινομένων.

Τις τελευταίες δεκαετίες οι ανάγκες για μετακίνηση έχουν αυξηθεί εντείνοντας το φαινόμενο. Πιο συγκεκριμένα, το μεγάλο κύμα αστικοποίησης σε συνδυασμό με την αύξηση της ιδιοκτησίας οχημάτων παγκοσμίως, έχουν επιβαρύνει ιδιαίτερα τις αστικές περιοχές που δεν είχαν σχεδιαστεί για να αντιμετωπίσουν τον κορεσμό της υφιστάμενης οδικής υποδομής. Ο κορεσμός αυτός έχει ισχυρό αντίκτυπο στη ζωή των σύγχρονων πόλεων καθώς δεν επηρεάζει μόνο την καθημερινή ζωή των κατοίκων τους, αλλά περιορίζει και την οικονομική τους ανάπτυξη.

Στην Εικόνα 1 φαίνεται η κατάταξη χωρών παγκοσμίως με βάση τον μέσο όρο των ωρών που περνούν σε συνθήκες συμφόρησης, σύμφωνα με έρευνα που πραγματοποίησε η εταιρεία INRIX η οποία ειδικεύεται στη συγκέντρωση και ανάλυση δεδομένων κίνησης.

Ένα σωστά σχεδιασμένο δίκτυο μεταφορών και ειδικότερα το οδικό δίκτυο, συμβάλει στη μείωση της ανεργίας, στην αύξηση των επενδύσεων και στην καλύτερη πρόσβαση στις αγορές, αφού αποτρέπει την άσκοπη σπατάλη χρόνου για τους χρήστες του κάτι το οποίο μεταφράζεται σε αύξηση της κοινωνικής παραγωγικότητας [1]. Ωστόσο, οι επιπτώσεις του φαινομένου της κυκλοφοριακής συμφόρησης δεν είναι μόνο κοινωνικοοικονομικές αλλά και περιβαλλοντικές.

Σύμφωνα με έρευνες του Ευρωπαϊκού Οργανισμού Περιβάλλοντος (ΕΟΠ), μεγάλο μέρος των εκπομπών αερίων του θερμοκηπίου της ΕΕ – πάνω από το 70% – προέρχεται από τις μεταφορές και συγκεκριμένα παράγεται από τα αυτοκίνητα, τα φορτηγά, τα ημιφορτηγά και τα λεωφορεία, τα οποία συμβάλλουν σημαντικά στην κλιματική αλλαγή.



Εικόνα 1: Κυκλοφοριακή Συμφόρηση σε πόλεις παγκοσμίως.

Την τελευταία δεκαετία, παρά την προσπάθεια μείωσης των ατμοσφαιρικών ρύπων (αιωρούμενα σωματίδια (PM), διοξείδιο του άνθρακα (CO₂), διοξείδιο του αζώτου (NO₂) κ.α), οι συγκεντρώσεις αυτών εξακολουθούν να είναι πολύ υψηλές, ιδίως στις μεγάλες πόλεις, βλάπτοντας την ανθρώπινη υγεία και το περιβάλλον [2]. Μέτρα όπως η εισαγωγή προτύπων ποιότητας των καυσίμων, η εφαρμογή προτύπων Euro για μειωμένες εκπομπές ρύπων των οχημάτων και η χρήση τεχνολογιών με μειωμένο περιβαλλοντικό αποτύπωμα, αδυνατούν να αντισταθμίσουν τις υψηλές συγκεντρώσεις που οφείλονται σε περιπτώσεις κυκλοφοριακής συμφόρησης, δηλαδή σε οχήματα που κινούνται με πολύ χαμηλές ταχύτητες, κάνοντας συχνές και πολλές στάσεις, καταναλώνοντας έτσι μεγαλύτερη ποσότητα καυσίμου και απελευθερώνοντας μεγαλύτερη ποσότητα βλαβερών αερίων στην ατμόσφαιρα.

Επιπλέον, η οδική κυκλοφορία είναι η πλέον εκτεταμένη πηγή ηχορύπανσης στην Ευρώπη (“Noise in Europe – 2020”). Τα επιβλαβή επίπεδα της επηρεάζουν πάνω από 100 εκατομμύρια άτομα στις χώρες μέλη του ΕΟΠ. Τα επίπεδα θορύβου αναμένεται να αυξηθούν στις αστικές περιοχές την προσεχή δεκαετία εξαιτίας της αστικής ανάπτυξης και της αυξημένης ζήτησης για μεταφορές όπως έχει προαναφερθεί. Η μακροχρόνια έκθεση σε ηχητικά επίπεδα περιβάλλοντος που υπερβαίνουν τα συνήθη επίπεδα άνεσης έχει σημαντικές επιπτώσεις στην υγεία, προκαλώντας 12.000 πρόωρους θανάτους στην ΕΕ [3].

Ακόμη, το απρόβλεπτο των τάσεων της κυκλοφορίας και οι αυξανόμενες ώρες οδήγησης συνδέονται με ένα ευρύ φάσμα σωματικών και συναισθηματικών προβλημάτων υγείας για τα άτομα. Μελέτες έχουν δείξει ότι το επίπεδο της ορμόνης της αδρεναλίνης στο αίμα αυξάνεται σε συνθήκες κυκλοφοριακού κορεσμού με αποτέλεσμα οι μετακινούμενοι να αισθάνονται άγχος και εκνευρισμό. Όσον αφορά τη σωματική υγεία, οι συνέπειες της συμφόρησης εμφανίζονται κυρίως με πόνο στη μέση και τα άκρα. Όλα τα παραπάνω μειώνουν την ικανότητα λήψης σωστών οδικών αποφάσεων και προκαλούν αύξηση των οδικών ατυχημάτων [4].

Συνεπώς, η μείωση της κυκλοφοριακής συμφόρησης σε αστικές περιοχές ώστε το οδικό δίκτυο να λειτουργεί στο βέλτιστο επίπεδο της στάθμης εξυπηρέτησης, είναι μια από τις κύριες ανησυχίες ιδιωτικών και δημόσιων φορέων. Οι στρατηγικές αντιμετώπισης που επιλέγονται έχουν ως στόχο την αποτελεσματική διαχείριση και έλεγχο της κυκλοφορίας καθώς και της πληροφόρησης του χρήστη του μεταφορικού συστήματος. Ορισμένες από αυτές είναι η πλήρης αξιοποίηση των ωφελειών των δημόσιων συγκοινωνιών, η ρύθμιση της σηματοδότησης, η σωστή σήμανση, η διοχέτευση της κυκλοφορίας από κατάλληλα οδικά τμήματα και η εκτροπή της σε περίπτωση αναπάντεχων συμβάντων όπως καιρικών συνθηκών κ.ά.

Λαμβάνοντας τις βέλτιστες αποφάσεις οι αρμόδιες αρχές και οι διαχειριστές του δικτύου επιδιώκουν τη βελτίωση της βιώσιμης κινητικότητας και τη δημιουργία ενός αποτελεσματικού συστήματος μεταφορών, το οποίο να παρέχει τα μέσα και τις δυνατότητες για να ικανοποιούνται οι οικονομικές, κοινωνικές και περιβαλλοντικές ανάγκες με ασφάλεια και αποτελεσματικά και επιπροσθέτως να ελαχιστοποιεί ή να αποφεύγει τις περιττές οικονομικές, κοινωνικές και περιβαλλοντικές επιδράσεις και το αντίστοιχο κόστος [5].

1.1.2. Ευφυή Συστήματα Μεταφορών

Η διαχείριση της κυκλοφορίας έχει καταστεί ζήτημα ζωτικής σημασίας σε πόλεις σε όλο τον κόσμο αφού περίπου το 75% όλων των αγαθών μεταφέρονται οδικώς όμως τα οδικά δίκτυα δεν είναι απείρως επεκτάσιμα επομένως η κατασκευή νέων υποδομών για την κάλυψη των απαιτήσεων μεταφοράς είναι πολύ δαπανηρή. Η ερευνητική κοινότητα έχει στρέψει την προσοχή της στην ανάπτυξη νέων τεχνολογιών πληροφόρησης και επικοινωνίας που συμβάλλουν στην αποδοτικότερη, ασφαλέστερη και οικονομικότερη χρήση, διαχείριση και έλεγχο του οδικού δικτύου και στη βελτίωση των περιβαλλοντικών επιδόσεων των οχημάτων. Μέσω εφαρμογών των ευφυών συστημάτων μεταφορών οι χρήστες του οδικού δικτύου μπορούν να χρησιμοποιήσουν αποτελεσματικότερα τις υπάρχουσες υποδομές.

Όπως αναφέρεται στην Οδηγία 2010/40/ΕΕ του Ευρωπαϊκού Κοινοβουλίου και του Συμβουλίου, της 7ης Ιουλίου 2010 τα ευφυή συστήματα μεταφορών (Intelligent Transportation Systems - ITS) ορίζονται με τον ακόλουθο τρόπο [6]:

Ένα ευφρές σύστημα μεταφορών ITS είναι μια προηγμένη εφαρμογή που στοχεύει στην παροχή καινοτόμων υπηρεσιών σχετικά με διαφορετικούς τρόπους μεταφοράς και διαχείρισης της κυκλοφορίας ώστε να επιτρέπει στους χρήστες να είναι καλύτερα ενημερωμένοι και να κάνουν ασφαλέστερη, πιο συντονισμένη και πιο «έξυπνη» χρήση των δικτύων μεταφορών. Τα ευφρή συστήματα μεταφορών ITS ενσωματώνουν τηλεπικοινωνίες, ηλεκτρονικά συστήματα και τεχνολογίες πληροφοριών με γνώμονα τη μηχανική μεταφορών (traffic engineering) με σκοπό το σχεδιασμό, τη λειτουργία, τη συντήρηση και τη διαχείριση συστημάτων μεταφοράς.

Αυτός ο ορισμός υποδεικνύει ότι οποιαδήποτε τεχνολογία πληροφοριών βοηθά τις μεταφορές, με τον ένα ή τον άλλο τρόπο, μπορεί να συμπεριληφθεί ως μία από τις πολλές καινοτομίες του ευφρούς συστήματος μεταφορών - ITS.

Οι εφαρμογές που παρέχουν τις ώρες ταξιδιού ή την καλύτερη επιλογή διαδρομής προς έναν συγκεκριμένο προορισμό είναι παραδείγματα τέτοιων τεχνολογιών. Παραδοσιακά, αυτές οι τεχνολογίες λειτουργούν με βάση στατιστικές προσεγγίσεις και ενημερώνονται με βάση τρέχοντα συμβάντα. Ωστόσο, στη συνεχώς μεταβαλλόμενη δυναμική της αστικής κυκλοφορίας, χρειάζεται ένα ακριβές πρότυπο πρόβλεψης των θεμελιωδών μεταβλητών της κυκλοφοριακής τεχνικής (φόρτος, πυκνότητα, ταχύτητα) σε κάποιο οδικό τμήμα σε συνάρτηση με το χρόνο.

Για να επιτευχθεί αυτό απαιτείται η καταγραφή ενός επαρκούς όγκου δεδομένων προηγούμενων περιόδων. Η ανάλυση αυτών των δεδομένων με τη χρήση σύγχρονων τεχνολογιών (Data Engineering – Data Analysis) επιτρέπει την κατανόηση των κυκλοφοριακών τάσεων καθώς βοηθάει και στην αποκάλυψη διάφορων μοτίβων κυκλοφορίας. Οι μετρήσεις αυτών των μεταβλητών μπορούν να γίνουν με πολλούς τρόπους, όπως με καταγραφή με βίντεο, με χρήση επαγωγικών βρόγχων ή μαγνητικών αισθητήρων και με χρήση του στίγματος GPS από έξυπνα κινητά τηλέφωνα.

Στη συγκεκριμένη εργασία γίνεται χρήση ανοιχτών δεδομένων αξιοποιώντας την ψηφιακή πύλη data.gov.gr. Τα επιλεγμένα δεδομένα αφορούν τον κυκλοφοριακό φόρτο στο κέντρο της Αθήνας και έχουν συλλεχθεί με φωρατές (ανιχνευτές τοποθετημένοι στο οδόστρωμα). Τα μοτίβα κυκλοφορίας που προκύπτουν μετά την επεξεργασία και ανάλυση των δεδομένων, συμβάλλουν στην πρόβλεψη βραχυπρόθεσμων καταστάσεων κυκλοφορίας για τη λήψη αποφάσεων, όπως η αντιμετώπιση απρόσμενων συμβάντων, αλλά και για τον μακροπρόθεσμο σχεδιασμό της λειτουργίας ενός οδικού δικτύου.

1.1.3. Βραχυπρόθεσμη Πρόβλεψη Κυκλοφορίας

Η βραχυπρόθεσμη πρόβλεψη κυκλοφοριακής ροής σε οδικό δίκτυο μεγάλης κλίμακας, που σημαίνει πώς θα εξελιχθούν οι συνθήκες κυκλοφορίας έως και μία ώρα μπροστά, είναι δύσκολη στις σύνθετες χωροχρονικές εξαρτήσεις, την κατευθυνόμενη τοπολογία δικτύου και το υψηλό υπολογιστικό κόστος. Σε πολλές περιπτώσεις είναι επομένως δύσκολο για τα παραδοσιακά μοντέλα πρόβλεψης να παράγουν καλά αποτελέσματα, καθώς δεν είναι σε θέση να προσεγγίσουν τη μη γραμμικότητα στα δεδομένα.

Λόγω των πρόσφατων εξελίξεων στον τομέα της Τεχνητής Νοημοσύνης (Artificial Intelligent - AI) και της εκθετικής αύξησης των ιστορικών δεδομένων, η πρόβλεψη έχει σημειώσει μεγάλες βελτιώσεις. Πιο συγκεκριμένα, ο υποτομέας της Τεχνικής Νοημοσύνης που ονομάζεται Μηχανική Μάθηση (Machine Learning - ML) έχει αναπτύξει μοντέλα οδηγούμενα από δεδομένα (data-driven) που σε συνδυασμό με αλγορίθμους εξόρυξης δεδομένων (Data Mining) και αξιοποίησης δεδομένων ευρείας κλίμακας (Big Data), έχουν αποδειχθεί ικανά να προσεγγίζουν σχεδόν οποιαδήποτε συνάρτηση μεταξύ δεδομένων εισόδου και εξόδου, ανεξάρτητα από το βαθμό γραμμικότητάς της και χωρίς προηγούμενη γνώση της λειτουργικής της μορφής [7].

Επίσης, πλέον οι σύγχρονοι υπολογιστές μπορούν να φέρουν εις πέρας πολύπλοκους υπολογισμούς και να χειριστούν μεγάλες ποσότητες δεδομένων αποτελεσματικά μέσα σε λίγα δευτερόλεπτα. Για αυτό το λόγο όλο και περισσότερο χρησιμοποιούνται τεχνικές μηχανικής μάθησης για την πρόβλεψη των επιθυμητών κυκλοφοριακών μεγεθών. Ωστόσο, συχνά τα αποτελέσματα που προκύπτουν είναι ασαφή ως προς τις σχέσεις αιτίας-αποτελέσματος που καθορίζουν το μηχανισμό λειτουργίας του οδικού δικτύου μελέτης και χρειάζεται περαιτέρω επεξηγήσεις αξιοποιώντας χωροχρονικές συσχετίσεις.

1.2. Αντικείμενο Διπλωματικής

Λαμβάνοντας υπόψη τη σημαντικότητα της ανάλυσης του φαινομένου της κυκλοφοριακής συμφόρησης, η παρούσα διπλωματική εργασία αποσκοπεί στην υλοποίηση μιας μεθοδολογίας για την εύρεση και την ανάλυση των χωρικών και χρονικών εξαρτήσεων σε ένα σύνθετο αστικό οδικό δίκτυο (multiplex urban road network). Ειδικότερα, αναζητούνται χωροχρονικές συσχετίσεις μεταξύ του κυκλοφοριακού φόρτου που επικρατεί και της επιβατικής ζήτησης των στάσεων σταθερών συγκοινωνιών ενός αστικού οδικού δικτύου με γραμμικές και μη γραμμικές μεθόδους κι έπειτα επιδιώκεται η πρόβλεψη των μελλοντικών κυκλοφοριακών συνθηκών με αξιοποίηση τεχνικών μηχανικής μάθησης.

Για το σκοπό αυτό αξιοποιούνται δύο βάσεις δεδομένων, μία πρώτη που αφορά την κυκλοφορία σε επιβατικά Ι.Χ. σε κεντρικούς άξονες στο κέντρο της Αθήνας και μια δεύτερη που αφορά επικυρώσεις εισιτηρίων επιβατών σε κεντρικές στάσεις του μετρό Αθηνών.

Κύριος στόχος είναι η εξέταση των χωροχρονικών εξαρτήσεων και η σύνδεση τους με την πρόβλεψη των κυκλοφοριακών συνθηκών.

Επισημαίνεται πως οι προβλέψεις εκπονούνται για περιορισμένο αριθμό οδικών τμημάτων και στάσεων ΣΤΑ.ΣΥ. στο κέντρο της Αθήνας. Συνεπώς η παρούσα εργασία περιορίζεται στην επεξεργασία δεδομένων κίνησης ενός στατικού οδικού δικτύου. Ως αποτέλεσμα, η απόδοση της τεχνικής πρόβλεψης που παρουσιάζεται ενδέχεται να μην ισχύει απαραίτητα για άλλα αστικά οδικά δίκτυα σε άλλες πόλεις. Επιπλέον, η εφαρμογή του ίδιου μοντέλου σε εντελώς διαφορετικούς τύπους οδικών δικτύων, όπως αυτοκινητόδρομους, είναι επίσης πιθανό να αποτύχει. Ο λόγος για αυτό είναι ότι τα μοτίβα κυκλοφορίας σε διαφορετικά μέρη του κόσμου ενδέχεται να διαφέρουν.

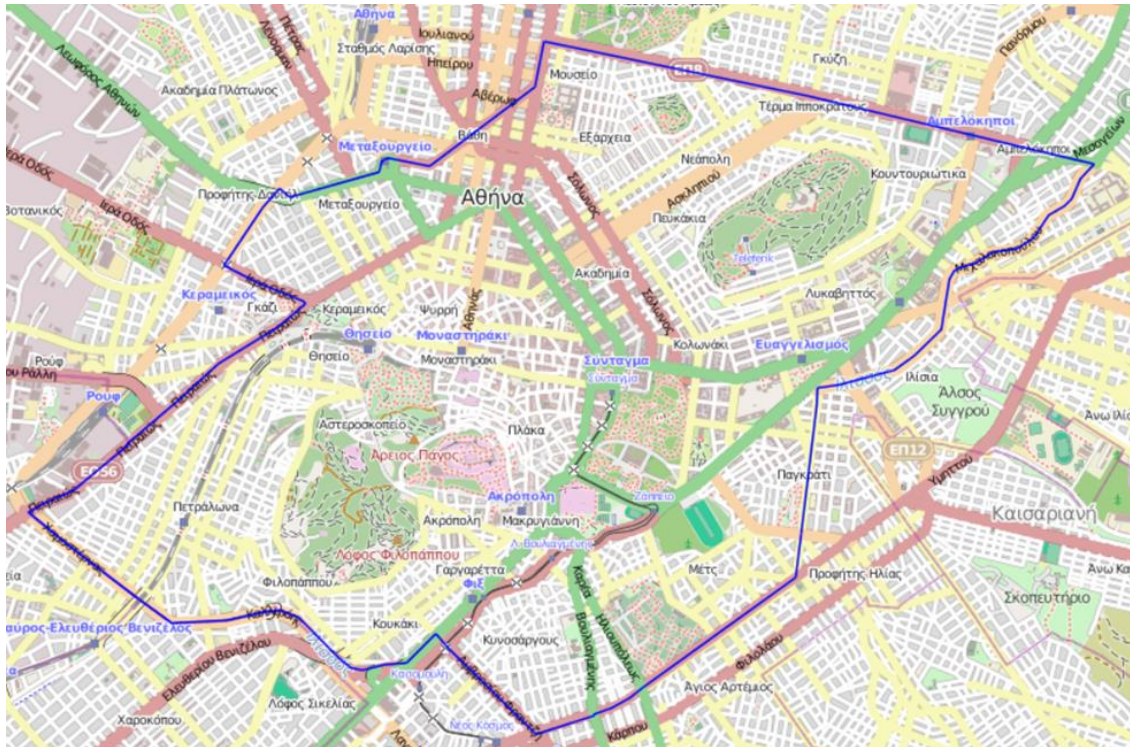
1.2.1. Συνεισφορά Παρούσας Εργασίας

Για την επίτευξη του κύριου στόχου που διατυπώθηκε, εφαρμόζονται μεθοδολογίες που επεκτείνονται από την κλασική γραμμική συσχέτιση του Pearson (Pearson's Correlation), τη Θεωρία Πληροφορίας (Information Theory), την ανάλυση χρονοσειρών έως τα Νευρωνικά Δίκτυα τύπου Μακράς και Βραχείας Μνήμης (Long-Short Term Memory Networks).

Η εφαρμογή της προτεινόμενης μεθοδολογίας εφαρμόζεται στο αστικό οδικό δίκτυο του κέντρου της Αθήνας, συγκεκριμένα εντός του δακτυλίου της ο οποίος φαίνεται στην Εικόνα 2. Να σημειωθεί πως η επιλογή του δακτυλίου Αθηνών δεν είναι τυχαία, καθώς εξυπηρετεί τον μεγαλύτερο όγκο κυκλοφοριακής και επιβατικής κίνησης του δήμου Αθηναίων.

Τα δεδομένα που χρησιμοποιήθηκαν χαρακτηρίζονται και ως ανοικτά δημόσια δεδομένα επειδή διαθέτονται από την ιστοσελίδα www.data.gov.gr η οποία αποτελεί τον κεντρικό κατάλογο των δημόσιων δεδομένων που παρέχει πρόσβαση σε βάσεις δεδομένων όλων των δημόσιων φορέων της ελληνικής κυβέρνησης.

Πιο συγκεκριμένα, τα δεδομένα που αφορούν μετρήσεις αριθμού διελεύσεων ανά σταθμό μέτρησης του δικτύου παρακολούθησης της κυκλοφορίας στην Αττική, παρέχει η περιφέρεια Αττικής ενώ τα δεδομένα που αφορούν επικυρώσεις μετακίνησης/μετεπιβίβασης σε σταθερές συγκοινωνίες, παρέχει ο Οργανισμός Αστικών Συγκοινωνιών Αθηνών (Ο.Α.Σ.Α.).



Εικόνα 2: Δακτύλιος Αθηνών.

Στην παρούσα εργασία:

- Γίνεται ανάκτηση δεδομένων σε μορφή csv.
- Τα δεδομένα επεξεργάζονται και τροποποιούνται στην επιθυμητή μορφή.
- Οι βάσεις δεδομένων αναλύονται μεμονωμένες αλλά και ενοποιημένες.
- Υπολογίζονται στατιστικά μεγέθη συσχέτισης των χρονοσειρών ανά δύο, ειδικότερα υπολογίζονται η συσχέτιση κατά Pearson και η Αμοιβαία Πληροφορία (Mutual Information).
- Με βάση τα παραπάνω μεγέθη εντοπίζονται τα οδικά τμήματα και οι στάσεις ΣΤΑ.ΣΥ. που σχετίζονται περισσότερο μεταξύ τους.
- Αναπτύσσονται Long-Short Term Memory (LSTM) Δίκτυα που αξιοποιούν τα συμπεράσματα της παραπάνω ανάλυσης στην πρόβλεψη του κυκλοφοριακού φόρτου και της ζήτησης στις στάσεις ΣΤΑ.ΣΥ. σε μελλοντικές χρονικές στιγμές.

1.3. Οργάνωση Εργασίας

Συνοπτικά, η οργάνωση της διπλωματικής εργασίας, η οποία φαίνεται και στο Διάγραμμα 1, έχει ως ακολούθως:

Στο παρόν Κεφάλαιο 1 δίνεται μια σύντομη εισαγωγή και περιγραφή του προβλήματος της κυκλοφοριακής συμφόρησης και έπειτα ορίζονται οι στόχοι της παρούσας διπλωματικής εργασίας, που είναι η ανάπτυξη ενός μοντέλου βραχυπρόθεσμης πρόβλεψης που θα λαμβάνει υπόψη χωροχρονικές συσχετίσεις.

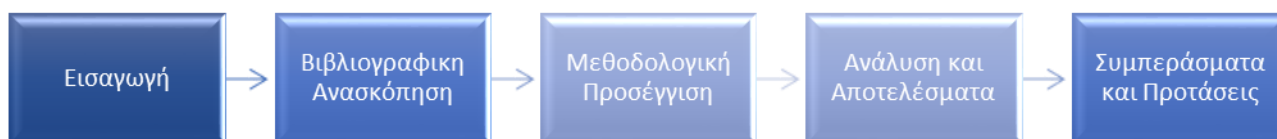
Στο κεφάλαιο 2 παρατίθενται η βιβλιογραφική ανασκόπηση, δηλαδή συναφείς έρευνες και μεθοδολογίες που έχουν προηγηθεί σχετικά με την βραχυπρόθεσμη πρόβλεψη κυκλοφορίας και των μεγεθών της.

Στο κεφάλαιο 3 συνίσταται η περιγραφή της διαδικασίας της μεθοδολογικής προσέγγισης που ακολουθείται ώστε να επιτευχθεί ο στόχος της διπλωματικής εργασίας και επεξηγείται αναλυτικά το θεωρητικό υπόβαθρο.

Στο κεφάλαιο 4 περιγράφεται η συλλογή και επεξεργασία των δεδομένων καθώς και η εφαρμογή της προτεινόμενης μεθοδολογίας και στο τέλος παρατίθενται τα αποτελέσματα της.

Στο κεφάλαιο 5 παρουσιάζονται τα συμπεράσματα που προκύπτουν από την ερμηνεία των αποτελεσμάτων της ανάλυσης. Επίσης γίνονται προτάσεις για περαιτέρω μελλοντική έρευνα.

Τέλος, συγκεντρώνεται η βιβλιογραφία που συνέβαλε στη συγγραφή της διπλωματικής εργασίας και αποτελείται από εργασίες σε διεθνή περιοδικά και από επιστημονικά βιβλία.



Διάγραμμα 1: Ροή Κεφαλαίων Εργασίας

Κεφάλαιο 2

Βιβλιογραφική Ανασκόπηση

2.1 Εισαγωγή

Η συνεχής εξέλιξη του τομέα των τεχνικών προβλέψεων συγκεντρώνει την αυξανόμενη προσοχή της επιστημονικής κοινότητας σε πολλούς και διαφορετικούς κλάδους όπως οι οικονομικές και κοινωνικές επιστήμες, η ιατρική κ.α. Ειδικά ο τομέας των μεταφορών, σχεδόν πέντε δεκαετίες – από τις αρχές της δεκαετίας του '80 – ασχολείται με το πρόβλημα της βραχυπρόθεσμης πρόβλεψης κυκλοφοριακών μεγεθών. Με τον όρο «βραχυπρόθεσμη» εννοείται χρονικός ορίζοντας έως 2 ώρες περίπου.

Σήμερα, χάρη στη συμβολή και ανάπτυξη της υπολογιστικής ισχύος των σύγχρονων υπολογιστών, η ερευνητική κοινότητα έχει καταφέρει να παρουσιάσει πολλές εξελιγμένες μεθοδολογίες που αφορούν την πρόβλεψη μελλοντικών κυκλοφοριακών μεγεθών.

Ανατρέχοντας στην πιο πρόσφατη βιβλιογραφία, εντοπίζονται οι πτυχές του προβλήματος της βραχυπρόθεσμης πρόβλεψης κυκλοφοριακών συνθηκών και οι διάφορες μεθοδολογίες με τις οποίες έχει αναλυθεί μέχρι τώρα. Στο παρόν κεφάλαιο παρουσιάζεται αναλυτικότερα το σύνολο των προσεγγίσεων και τα κριτήρια με τα οποία ταξινομούνται.

Πιο συγκεκριμένα, επισημαίνονται οι μεθοδολογίες πρόβλεψης που έχουν αναπτυχθεί, καθώς και οι σχετικές έρευνες που αφορούν την εφαρμογή της βαθιάς μηχανικής μάθησης στην πρόβλεψη κυκλοφοριακών μεγεθών. Επίσης, εφόσον το επίκεντρο της διπλωματικής εργασίας είναι η ανάπτυξη ενός μοντέλου πρόβλεψης χρονοσειρών χρησιμοποιώντας ως μέθοδο επιλογής χαρακτηριστικών τη θεωρία αμοιβαίας πληροφορίας, αναφέρονται έρευνες που αφορούν την εφαρμογή της στον τομέα των μεταφορών.

2.2 Μεθοδολογίες Πρόβλεψης Κυκλοφοριακών Μεγεθών

Η επιλογή της κατάλληλης μεθοδολογίας για την πρόβλεψη κυκλοφοριακών μεγεθών έχει ιδιαίτερη σημασία και πρέπει να λαμβάνει υπόψη τις ιδιαιτερότητες που παρουσιάζουν τα συγκεκριμένα μεγέθη.

Οι υπάρχουσες μέθοδοι πρόβλεψης μπορούν να κατηγοριοποιηθούν με πολλούς τρόπους, ωστόσο το πιο σημαντικό σύστημα ταξινόμησης είναι η ταξινόμησή τους σε τρία είδη [8]:

- i. Μέθοδοι που καθοδηγούνται από μοντέλα (model-driven methods)
- ii. Στατιστικές μέθοδοι (statistical methods)
- iii. Μέθοδοι που καθοδηγούνται από δεδομένα (data – driven methods)

Στην πρώτη μεθοδολογία χρησιμοποιούνται προγράμματα προσομοίωσης του κυκλοφοριακού δικτύου και ανάλυσης της οδηγικής απόδοσης και συμπεριφοράς σε μικροσκοπικό επίπεδο. Μέσω της λεπτομερής αναπαράστασης της τοπολογίας του δικτύου, συμπεριλαμβάνοντας κόμβους, διασταυρώσεις, αριθμό λωρίδων, όρια ταχύτητας, φωτεινούς σηματοδότες κ.α., επιδιώκεται η καλύτερη δυνατή πρόβλεψη. Πιο συγκεκριμένα, η αναπαράσταση αυτών των πληροφοριών γίνεται υπό μορφή γράφου (graph) με τις ακμές (edges) να αναπαριστούν τους δρόμους και τους κόμβους (nodes) τις διασταυρώσεις ή τους - κυκλικούς ή μη – κόμβους. Κατά την εφαρμογή της προσομοίωσης, στο υπολογιστικό σύστημα ενεργοποιούνται όλα τα στοιχεία σύμφωνα με τη δοσμένη εικονική αναπαράσταση του οδικού δικτύου και τις αλληλεπιδράσεις μεταξύ των εικονικών υποδομών και των οδηγών καθώς και τις αλληλεπιδράσεις οδηγών με οδηγούς και πεζούς σύμφωνα με εξελιγμένα μοντέλα θεωρίας κυκλοφορίας. Η πρόβλεψη για τη μελλοντική κατάσταση του δικτύου γίνεται μελετώντας την κίνηση ενός οχήματος και αναλύοντας τις κυκλοφοριακές συνθήκες σε μια συγκεκριμένη περιοχή.

Η καθοδηγούμενη από μοντέλα μέθοδος επιλέγεται πιο συχνά όταν ζητείται η μακροπρόθεσμη πρόβλεψη κυκλοφορίας, δηλαδή για αλλαγές στο οδικό δίκτυο με ορίζοντα πρόβλεψης μήνες ή και χρόνια, χωρίς να λαμβάνεται υπόψη στο ενδιάμεσο χρονικό διάστημα οποιασδήποτε μορφής πραγματικού χρόνου δεδομένα [9]. Επομένως, λόγω της πολυπλοκότητας και της αβεβαιότητας του φαινομένου της κυκλοφορίας, είναι δύσκολο να καθοριστούν μοντέλα που να εκφράζουν πλήρως τη σύνθετη σχέση μεταξύ των κυκλοφοριακών παραμέτρων.

Ως εκ τούτου, πολλοί ερευνητές έχουν χρησιμοποιήσει γενικά τις στατιστικές μεθόδους. Πολλά στατιστικά μοντέλα έχουν εφαρμοστεί εκτενώς στη βραχυπρόθεσμη πρόβλεψη κυκλοφορίας, συμπεριλαμβανομένων των μοντέλων παλινδρόμησης και προσέγγισης/εκτίμησης συναρτήσεων. Συχνή είναι η χρήση μεθοδολογιών ανάλυσης χρονοσειρών μιας και τα εξεταζόμενα κυκλοφοριακά μεγέθη μεταβάλλονται με το χρόνο. Μια πολύ δημοφιλής στατιστική προσέγγιση για την βραχυπρόθεσμη πρόβλεψη κίνησης ήταν πάντα η χρήση στατιστικών μοντέλων παλινδρόμησης και μοντέλων αυτόματης παλινδρόμησης ενσωματωμένου κινούμενου μέσου όρου (οικογένεια ARIMA), λόγω της ισχυρής μαθηματικής τους βάσης και των δυνατοτήτων ερμηνείας. Ορισμένες ακόμα τυπικές μέθοδοι είναι το τοπικό μοντέλο γραμμικής παλινδρόμησης (LLR), η αυτόματη παλινδρόμηση κινητού μέσου (ARMA), η εποχιακή αυτόματη παλινδρόμηση ενσωματωμένου κινητού μέσου (SARIMA). Οι δύο τελευταίες μέθοδοι αποτελούν επεκτάσεις των μοντέλων Autoregressive Integrated Moving Average (ARIMA), που θεωρούνται ως τα πιο δημοφιλή μοντέλα χρονοσειρών και χρησιμοποιούνται ευρέως στον τομέα των μεταφορών.

Οι μεθοδολογίες πρόβλεψης χρονοσειρών όπως το πρότυπο ARIMA και το μοντέλο Φίλτρου Kalman (Kalman Filter), βασίζονται στις παρελθοντικές τιμές του μεγέθους που παριστάνει η χρονοσειρά για την πρόβλεψη της μελλοντικής εξέλιξης. Αυτές οι μέθοδοι έχουν υψηλή υπολογιστική απόδοση και καλή ακρίβεια, όμως έρευνες έχουν δείξει πως είναι οι πλέον κατάλληλες για στάσιμες ή γραμμικές χρονοσειρές και δεν είναι ικανές να αντιμετωπίσουν τις μη γραμμικές εξαρτήσεις της κυκλοφοριακής ροής [10] [11].

Σήμερα, η ερευνητική προσοχή έχει στραφεί από τις αναλυτικές μεθόδους στις καθοδηγούμενες από τα δεδομένα. Οι προσεγγίσεις που βασίζονται σε δεδομένα έγιναν δημοφιλείς λόγω της αυξανόμενης διαθεσιμότητας δεδομένων αστικής κυκλοφορίας και της ραγδαίας εξέλιξης του κλάδου της μηχανικής μάθησης τα τελευταία χρόνια. Η συλλογή δεδομένων γίνεται με διάφορους αισθητήρες όπως με ανιχνευτές επαγωγικού βρόχου, που είναι τοποθετημένοι στους δρόμους, με κάμερες τοποθετημένες σε σχετικά υψηλές θέσεις πάνω από τις οδούς, με χρήση του στίγματος GPS από τα έξυπνα κινητά τηλέφωνα οδηγών αλλά και από τα οχήματα που είναι εξοπλισμένα με συστήματα τηλεματικής (in-vehicle telematics). Η εκτεταμένη χρήση έξυπνων συσκευών και συστημάτων, που είναι σε θέση να παρακολουθούν και να συλλέγουν τεράστιο όγκο δεδομένων κινητικότητας σε πραγματικό χρόνο, δίνει στους ερευνητές την ευκαιρία να αναπτύξουν μοντέλα πρόβλεψης που είναι πιο ακριβή και συνεχώς ενημερωμένα, καθώς και υψηλής χρονικής ανάλυσης [12]. Αυτό ήταν το σημείο καμπής που έστρεψε την προσοχή των ερευνητών από την κλασσική στατιστική προσέγγιση στη μηχανική μάθηση και τη βαθιά μηχανική μάθηση μοντέλων που βασίζονται σε δεδομένα με τη βοήθεια αλγορίθμων εξόρυξης δεδομένων (Data Mining) και αξιοποίησης δεδομένων ευρείας κλίμακας (Big Data).

2.2.1 Παραμετρικές και Μη Παραμετρικές Προσεγγίσεις

Η πρόβλεψη κυκλοφοριακών συνθηκών με μεθόδους καθοδηγούμενες από δεδομένα μπορεί να κατηγοριοποιηθεί περαιτέρω σε παραμετρικές και μη παραμετρικές προσεγγίσεις. Με τον όρο «παραμετρικά» χαρακτηρίζονται μοντέλα των οποίων η δομή είναι προκαθορισμένη και για τα οποία χρειάζεται να εκτιμηθούν μόνο οι παράμετροι τους μέσω των διαθέσιμων δεδομένων. Από τα πιο γνωστά παραμετρικά μοντέλα είναι τα αυτοπαλινδρομούμενα μοντέλα κινητών μέσων όρων (ARIMA models) και οι ποικίλες παραλλαγές τους. Πλεονέκτημα τέτοιων προσεγγίσεων είναι ότι για την "εκμάθηση" των παραμέτρων τους απαιτούν λιγότερους υπολογισμούς και δεδομένα σε σχέση με τα μη παραμετρικά μοντέλα [13]. Άλλα παραδείγματα περιλαμβάνουν το φιλτράρισμα Kalman, τα δίκτυα Μπεϋζιανών, το μοντέλο της αλυσίδας Markov. Ωστόσο, αυτές οι μέθοδοι χρησιμοποιήθηκαν για βραχυπρόθεσμη πρόβλεψη της κυκλοφορίας σε μικρά αστικά αρτηριακά δίκτυα ή αυτοκινητόδρομους και όλες υστερούν σε υψηλή υπολογιστική πολυπλοκότητα.

Ο όρος «μη παραμετρικά» δεν υποδηλώνει ότι δεν απαιτείται η εκμάθηση παραμέτρων σε αυτή την κατηγορία μοντέλων αλλά ότι ο αριθμός και η φύση αυτών των παραμέτρων είναι ευέλικτα και δεν καθορίζονται εκ των προτέρων. Η δομή του μοντέλου και των παραμέτρων του προσδιορίζονται από τα δεδομένα κι επομένως συνήθως απαιτούνται περισσότερα δεδομένα για την ‘εκμάθηση’ των μη παραμετρικών μοντέλων.

Ένα από τα πλεονεκτήματά τους είναι ότι οι πολύπλοκες, δυναμικές και μη γραμμικές διαδικασίες που χαρακτηρίζουν το φαινόμενο της ροής κυκλοφορίας μπορούν να μοντελοποιηθούν. Επιπλέον δεν απαιτείται εκ των προτέρων γνώση για τις υποκείμενες διαδικασίες αφού μαθαίνονται από τα δεδομένα. Από τα πιο γνωστά μη παραμετρικά μοντέλα είναι η παλινδρόμηση των k κοντινότερων γειτόνων (K-Nearest Neighbors Regression – KNN) [14] και τα μοντέλα Νευρωνικών Δικτύων (Neural Networks) και οι παραλλαγές τους. Τα μη παραμετρικά μοντέλα υπερισχύουν από τις στατιστικές παραμετρικές μεθόδους όσο αφορά την ακρίβεια πρόβλεψης και έχουν περιθώριο βελτίωσης της απόδοσης τους [15].

2.3 Βαθιά Μηχανική Μάθηση στην Πρόβλεψη Κυκλοφορίας

Δεδομένης της πολυπλοκότητας του προβλήματος της βραχυπρόθεσμης πρόβλεψης κυκλοφορίας οι αλγόριθμοι βαθιάς μηχανικής μάθησης προσελκύουν όλο και περισσότερο το ερευνητικό ενδιαφέρον. Η βαθιά μηχανική μάθηση αποτελεί ένα συγκεκριμένο είδος μηχανικής μάθησης που επιτρέπει σε ένα υπολογιστικό σύστημα να βελτιώνεται με την εμπειρία και τα δεδομένα. Τα Βαθιά Νευρωνικά Δίκτυα (Deep Neural Networks), είναι δημοφιλή επειδή μπορούν να εκπαιδευτούν ώστε να αποτυπώσουν τις πιο σύνθετες εξαρτήσεις που επικρατούν στα αστικά οδικά δίκτυα.

Οι αλγόριθμοι βαθιάς μηχανικής μάθησης αξιοποιούν αποτελεσματικά μεγάλες ποσότητες δεδομένων, από πολλά οδικά τμήματα του δικτύου, προσεγγίζοντας σχεδόν κάθε συνάρτηση είτε είναι γραμμική είτε όχι [16]. Χωρίς ο χρήστης να χρειάζεται να παραμετροποιήσει τα δεδομένα, αυτοί οι αλγόριθμοι έχουν την ικανότητα να συμπεριλαμβάνουν εξωγενείς παράγοντες που επηρεάζουν τις κυκλοφοριακές συνθήκες και να προσαρμόζονται σε απρόσμενα συμβάντα.

Αν και αυτά τα μοντέλα πετυχαίνουν ικανοποιητική ακρίβεια πρόβλεψης, τα περισσότερα από αυτά μπορούν να εφαρμοστούν μόνο σε απλά δίκτυα κυκλοφορίας όπως μια ακολουθία οδικών τμημάτων ενώ οι τοπολογίες αστικών δικτύων είναι γενικά πιο περίπλοκες. Εκτός από αυτό το μειονέκτημα, τα υπάρχοντα μοντέλα τείνουν να εξάγουν στατικές χωροχρονικές εξαρτήσεις στην κίνηση, ενώ αυτές οι εξαρτήσεις εξελίσσονται με την πάροδο του χρόνου.

Η πιο πρόσφατη βιβλιογραφία περιλαμβάνει πολυάριθμες έρευνες που αξιοποιούν διαφορετικές αρχιτεκτονικές των Νευρωνικών Δικτύων για να παράγουν ακριβείς κυκλοφοριακές προβλέψεις. Το Multi-layer Perceptron (MLP) αποτελεί μια κλασική αρχιτεκτονική νευρωνικών δικτύων που επιτρέπει την εκμετάλλευση δεδομένων που προέρχονται από διαφορετικά τμήματα του οδικού δικτύου, ενώ δεν απαιτείται προηγούμενη βαθμονόμηση ή γνώση [17], όμως παρά το γεγονός ότι χειρίζεται αποτελεσματικά τεράστιο όγκο δεδομένων, παραλείπει τις χωρικές εξαρτήσεις μεταξύ των οδικών τμημάτων.

Για να ληφθούν υπόψη τα χωροχρονικά χαρακτηριστικά της κυκλοφοριακής ροής προτάθηκαν πιο εξελιγμένες προσεγγίσεις που περιλαμβάνουν πιο περίπλοκες αρχιτεκτονικές όπως τα Συνελικτικά Δίκτυα (Convolutional Neural Networks – CNN) τα οποία αντιμετωπίζουν το οδικό δίκτυο ή μόνο κάποιο τμήμα του δικτύου και τις επικρατούσες κυκλοφοριακές συνθήκες ως εικόνες, οι οποίες στη συνέχεια, μετατρέπονται σε τρισδιάστατο πίνακα. Ομοίως, ένα οδικό δίκτυο μπορεί να αναπαρασταθεί ως γράφημα, χρησιμοποιώντας έναν πίνακα γειτνίασης, ο οποίος μπορεί να θεωρηθεί ως μονοδιάστατη αναπαράσταση μιας εικόνας (κλίμακα του γκρι), με την τιμή κάθε κελιού του πίνακα να ισούται με μια παράμετρο κυκλοφορίας (π.χ. ταχύτητα) του αντίστοιχου οδικού τμήματος [18].

Επίσης, έχουν εφαρμοστεί και υβριδικές προσεγγίσεις συνδυάζοντας τα Συνελικτικά Νευρωνικά Δίκτυα με άλλες μεθοδολογίες όπως τα Επαναλαμβανόμενα Νευρωνικά Δίκτυα (Recurrent Neural Networks – RNN) και τα Δίκτυα Μακράς και Βραχείας Μνήμης (Long Short-Term Memory Networks – LSTM). Αυτά τα τελευταία δίκτυα είναι εξίσου δημοφιλή και χρησιμοποιούνται ευρέως καθώς είναι πιο κατάλληλα για δεδομένα που έχουν και χρονική διάσταση, παραδείγματος χάρη χρονοσειρές, επειδή μεταφέρουν την επιρροή των προηγούμενων παρατηρήσεων σε αυτές που ακολουθούν [19]. Τα υπάρχοντα μοντέλα πετυχαίνουν ικανοποιητική ακρίβεια πρόβλεψης κυκλοφορίας όμως δεν παρέχουν την ερμηνεία του τρόπου που προκύπτουν οι συσχετίσεις και οι προβλέψεις [20]. Εξαιτίας της φύσης των αλγορίθμων της τεχνητής νοημοσύνης ο εντοπισμός των συσχετίσεων γίνεται στα κρυφά επίπεδα (layers) των νευρωνικών δικτύων και δεν είναι διαθέσιμα στο χρήστη, για αυτό συχνά αναφέρεται πως οι αλγόριθμοι μπορούν να ξεπεράσουν πολλά προβλήματα (π.χ. αποτυχία δεδομένων) και να δώσουν λύσεις «μ αύρου κουτιού».

Ο προσδιορισμός των χωρικών και χρονικών συσχετίσεων αποτελεί ένα σημαντικό ζήτημα έρευνας στη βραχυπρόθεσμη πρόβλεψη της κυκλοφοριακής ροής όπως αναφέρει και η έρευνα των Vlahogianni et al. [7]. Αρκετές δημοσιευμένες έρευνες αποδεικνύουν τη βελτίωση των προβλέψεων κυκλοφοριακών συνθηκών λόγω της εισαγωγής χωροχρονικών χαρακτηριστικών στη φάση της μοντελοποίησης μέσω χωρικών και χρονικών συσχετίσεων [21]. Δηλαδή πληροφορίες για το αν η κυκλοφορία είναι ανάντη ή κατάντη, αν ο τύπος του οδικού δικτύου που μελετάται είναι υπεραστικές οδοί ή δίκτυο, αυτοκινητόδρομοι ή οδοί σε αστικό δίκτυο, βοηθούν στην αναπαράσταση των δεδομένων και της ερμηνείας των μοντέλων.

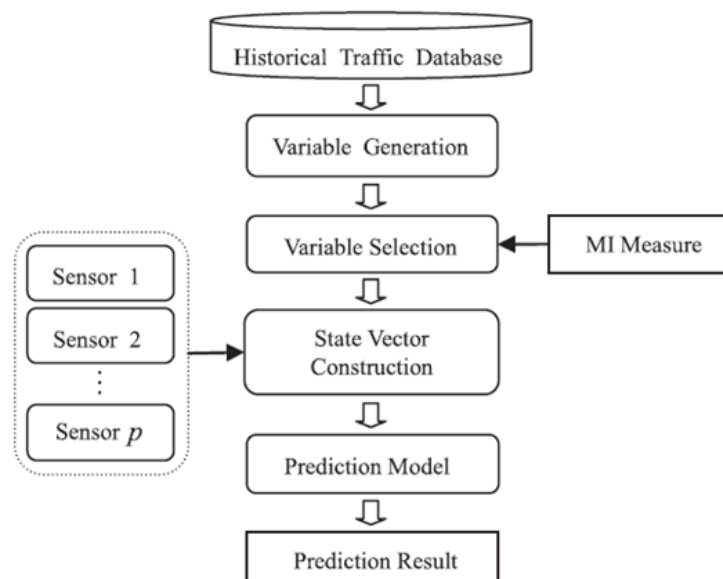
2.4 Η Συμβολή της Αμοιβαίας Πληροφορίας

Για την αξιολόγηση των χωροχρονικών συσχετισμών μεταξύ των τμημάτων ενός οδικού δικτύου, απαιτείται ένα μέτρο που να αποτυπώνει την εξάρτηση μεταξύ των μεταβλητών. Το πιο ευρέως χρησιμοποιούμενο μέτρο εξάρτησης είναι ο συντελεστής συσχέτισης, αλλά αντιπροσωπεύει μόνο τη γραμμική σχέση μεταξύ των μεταβλητών και μπορεί να μην λειτουργεί καλά όταν υπάρχουν μη γραμμικές σχέσεις.

Σε αντίθεση με τον συντελεστή συσχέτισης, ο συντελεστής της αμοιβαίας πληροφορίας, που εν συντομία αναφέρεται MI (Mutual Information), μπορεί να εκτιμήσει κάθε είδους σχέση μεταξύ τυχαίων μεταβλητών, συμπεριλαμβανομένων των μη γραμμικών σχέσεων [22].

Αρκετές προσεγγίσεις που έχουν ήδη χρησιμοποιήσει έννοιες εντροπίας και αμοιβαίας πληροφόρησης για την ανάλυση και πρόβλεψη κυκλοφορίας. Οι Lan et al. [23] χρησιμοποίησαν μια μη γραμμική συνάρτηση αυτοσυσχέτισης που ονομάζεται μέση αμοιβαία πληροφόρηση (average MI) στην ανακατασκευή των σειρών κυκλοφορίας (traffic series) για να αποκτήσουν περισσότερες πληροφορίες για τη δυναμική εξέλιξη της κυκλοφορίας. Η αμοιβαία πληροφορία χρησιμοποιήθηκε επίσης για τη διερεύνηση της επίδρασης των μεταβατικών συνθηκών κυκλοφοριακής ροής στη χρονική εξέλιξη της ροής κυκλοφορίας. Επίσης άλλες έρευνες εισήγαγαν διάφορα μέτρα που βασίζονται στην εντροπία για τον εντοπισμό ομοιότητας στα μοτίβα χρόνου ταξιδιού και να ενισχύσουν τις προβλέψεις.

Ένα ακόμη παράδειγμα αποτελεί η έρευνα των Ryu et al. [22], στην οποία προτείνεται μία μέθοδος κατασκευής ενός διανύσματος κατάστασης κίνησης, όπως παρουσιάζεται στο ακόλουθο Διάγραμμα 2.



Διάγραμμα 2: Μεθοδολογική Προσέγγιση της κατασκευής διανύσματος κατάστασης κίνησης μέσω MI για τη βραχυπρόθεσμη πρόβλεψη κυκλοφοριακής ροής, πηγή [22].

Ειδικότερα, κάνοντας χρήση του κριτηρίου της αμοιβαίας πληροφορίας επιλέγονται οι καταλληλότερες μεταβλητές για την περιγραφή της κυκλοφοριακής ροής. Για την έρευνα των χωροχρονικών συσχετισμών των γειτονικών οδικών τμημάτων με το εξεταζόμενο κομμάτι δρόμου έγινε χρήση της σταθμισμένης απόστασης στο μοντέλο του k κοντινότερου γείτονα. Τα αποτελέσματα των πειραμάτων της μεθόδου αυτής, την ανέδειξαν ως καλύτερη των ήδη υπαρχόντων.

2.5 Συμπεράσματα Βιβλιογραφικής Ανασκόπησης

Από την παραπάνω βιβλιογραφική ανασκόπηση προκύπτουν πολλά ενδιαφέροντα και σημαντικά συμπεράσματα σχετικά με το ζήτημα της βραχυπρόθεσμης πρόβλεψης κυκλοφοριακών συνθηκών σε διαφορετικές τύπους οδών. Ο συνεχώς αυξανόμενος ρυθμός εκπόνησης μελετών τα τελευταία χρόνια γύρω από το ζήτημα αυτό, υποδεικνύει μια ανοδική τάση του ενδιαφέροντος των ερευνητών για την πρόγνωση της κυκλοφορίας. Οι λόγοι που έχουν οδηγήσει προς αυτή την κατεύθυνση είναι η εξέλιξη του μαθηματικού υποβάθρου, των συστημάτων συλλογής κυκλοφοριακών μετρήσεων αλλά και η αύξηση της υπολογιστικής ισχύος των σύγχρονων υπολογιστών.

Στις παραπάνω έρευνες παρατηρείται ότι η συντριπτική πλειοψηφία έχει ως περιοχή μελέτης έναν αυτοκινητόδρομο, μία ελεύθερη οδό ή τμήματα αυτών λόγω της τοπολογικής απλότητας που τους διακρίνει και παρέχουν μια ευκολία στη διερεύνησή τους. Ωστόσο τελευταία παρουσιάζονται στη βιβλιογραφία ολοένα και περισσότερες έρευνες που αφορούν πρότυπα πρόβλεψης της κυκλοφορίας σε αστικές αρτηρίες.

Κατά την παρουσίαση όλων των διαφορετικών μεθοδολογιών πρόβλεψης έγινε κατανοητή η ικανότητα της κάθε μιας να αποτυπώνει τις ιδιαίτερες κυκλοφοριακές συνθήκες. Η κάθε προσέγγιση έχει τα δικά της πλεονεκτήματα και μειονεκτήματα και η επιλογή του κατάλληλου κάθε φορά προτύπου είναι ίσως αυτό που δυσκολεύει περισσότερο τον μελετητή. Για παράδειγμα, οι μη παραμετρικές μέθοδοι απαιτούν μεγάλες βάσεις δεδομένων αλλά καταφέρνουν να περιγράψουν επαρκώς σύνθετα φαινόμενα της κυκλοφορίας. Από την άλλη, οι παραμετρικές μέθοδοι δίνουν αξιόπιστα αποτελέσματα ακόμη και με μικρές βάσεις δεδομένων όμως αδυνατούν να ακολουθήσουν τις απότομες κυκλοφοριακές μεταβολές. Δεν είναι τυχαίο λοιπόν ότι οι δύο αυτές κατηγορίες μεθόδων έχουν χρησιμοποιηθεί εξίσου από τους ερευνητές.

Στην βιβλιογραφία φαίνεται ακόμα η ανάπτυξη μιας μεγάλης ποικιλίας αλγορίθμων βαθιάς μηχανικής μάθησης που έχουν χρησιμοποιηθεί για την πρόβλεψη των κυκλοφοριακών μεγεθών. Παρά την υψηλή ακρίβεια προβλέψεων που προσφέρουν αυτού του είδους οι μεθοδολογίες, δεν παρουσιάζουν τη χωρική και χρονική συσχέτιση των δεδομένων που λαμβάνεται υπόψη εντός των

μοντέλων, για αυτό συχνά τα αποτελέσματα που δίνουν χαρακτηρίζονται ως λύσεις «μαύρου κουτιού».

Συνοψίζοντας υπάρχουν αρκετά ζητήματα που δεν εξετάζονται στην τρέχουσα βιβλιογραφία. Πρώτον, πρέπει να διερευνηθεί η ακριβέστερη αναπαράσταση ενός οδικού δικτύου καθώς η σύνθετη τοπολογία των μελετούμενων οδικών δικτύων δεν έχει ληφθεί επαρκώς υπόψη στα μοντέλα βαθιάς μηχανικής μάθησης. Δεύτερον, οι διάφορες χωροχρονικές εξαρτήσεις δεν έχουν αξιοποιηθεί πλήρως. Για παράδειγμα, η κυκλοφοριακή ροή παρουσιάζει διαφορετικές χρονικές περιοδικότητες και η χωρική εξάρτηση μπορεί να διαφέρει σε όλο το δίκτυο.

Ακόμη και στην απλούστερη περίπτωση των λειτουργιών του αυτοκινητόδρομου, η ακριβής καταγραφή των χαρακτηριστικών της χωρικής κίνησης εξακολουθεί να είναι ένα ανοιχτό ζήτημα, καθώς δεν έχει εισαχθεί γενικευμένη προσέγγιση.

Επίσης, η καλύτερη κατανόηση των χωροχρονικών συσχετίσεων σε ένα δίκτυο κίνησης που αξιοποιείται από τα μοντέλα μηχανικής μάθησης θα ήταν χρήσιμη για την πλήρη ενσωμάτωση εφαρμογών ευφυών συστημάτων μεταφορών ITS και κατ' επέκταση για τη διαχείριση της κυκλοφορίας. Συνεπώς, η παρούσα διπλωματική εργασία, επιδιώκει την ανάδειξη της σημαντικότητας των αναφερόμενων εξαρτήσεων μέσω της παρουσίασης μιας απλής μεθοδολογίας υπολογισμού τους και την ενσωμάτωσή τους στη διαδικασία πρόβλεψης.

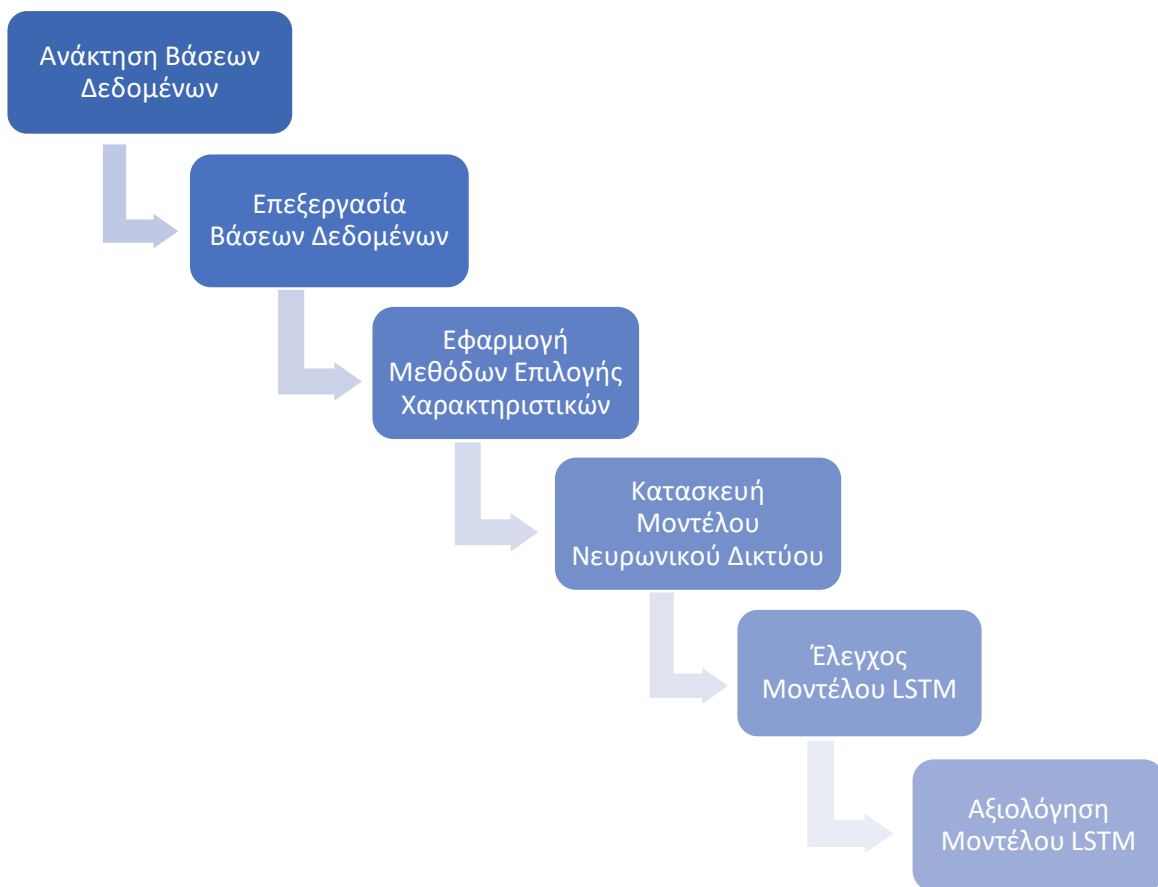
Κεφάλαιο 3

Μεθοδολογική Προσέγγιση

3.1 Προτεινόμενη Μεθοδολογία

Στην παρούσα διπλωματική εργασία επιδιώκεται η ανάπτυξη ενός μοντέλου βαθιάς μηχανικής μάθησης με την επιλογή των κατάλληλων παραμέτρων και δεδομένων για τη διενέργεια των υπολογισμών και την εξαγωγή συμπερασμάτων σχετικά με τη βραχυπρόθεσμη πρόβλεψη του κυκλοφοριακού φόρτου και της επιβατικής ζήτησης στο κέντρο της Αθήνας.

Η προτεινόμενη μεθοδολογία για την ανάπτυξη ενός μοντέλου νευρωνικού δικτύου βαθιάς μάθησης με χρήση καταγεγραμμένων δεδομένων από μια βάση δεδομένων ευρείας κλίμακας, περιγράφεται σε βήματα στο παρακάτω Διάγραμμα 3. Έπειτα, ακολουθεί μια εκτενής ανάλυση και επεξήγηση του θεωρητικού υποβάθρου στο οποίο βασίστηκε η ανάλυση των δεδομένων και η εφαρμογή της μεθόδου ανάπτυξης του μοντέλου πρόβλεψης.

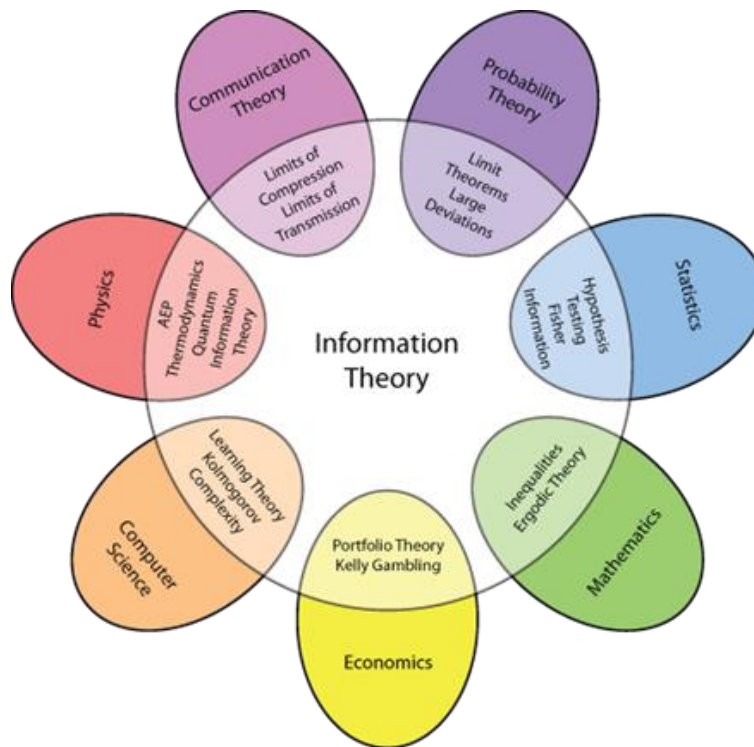


Διάγραμμα 3: Διάγραμμα Ροής Εργασιών

3.2 Θεωρητικό Υπόβαθρο

3.2.1 Θεωρία Πληροφορίας

Η Θεωρία της Πληροφορίας (Information Theory) αποτελεί τομέα των Εφαρμοσμένων Μαθηματικών που μελετά τρόπους ποσοτικοποίησης, διαχείρισης και απεικόνισης της αβεβαιότητας της πληροφορίας. Εφευρέθηκε από τον C.Shannon το 1948 με στόχο τη θεμελίωση εννοιών και θεωρημάτων που επιτρέπουν την μαθηματική περιγραφή της διαδικασίας της επικοινωνίας. Αν και αρχικά η θεωρία της πληροφορίας αποτέλεσε τμήμα της επιστήμης των επικοινωνιών, από πολύ νωρίς οι αρχές, οι μεθοδολογίες και τα συμπεράσματά της βρήκαν εφαρμογή σε πολλές επιστημονικές περιοχές, όπως η επιστήμη των υπολογιστών, οι πιθανότητες, η στατιστική, τα οικονομικά κλπ όπως φαίνεται στην Εικόνα 3.



Εικόνα 3: Η σχέση της Θεωρίας Πληροφορίας με τις υπόλοιπες επιστήμες

Οι τεχνικές που χρησιμοποιούνται στη Θεωρία της Πληροφορίας είναι πιθανολογικής φύσης και συνήθως αφορούν δύο συγκεκριμένες ποσότητες, την Εντροπία και την Αμοιβαία Πληροφορία. Η εντροπία κατά Shannon (Shannon's Entropy) είναι ένα μέτρο το οποίο φανερώνει το ποσό της αβεβαιότητας που περιέχεται στην έκβαση της τιμής μιας τυχαίας μεταβλητής ή ενός πειράματος. Έστω X μια τυχαία μεταβλητή με συνάρτηση κατανομής πιθανότητας $P(x)$ τότε η εντροπία Shannon της τυχαίας μεταβλητής X ορίζεται ως η ακόλουθη εξίσωση (1):

$$H(X) = - \sum_{i=1}^n P(x_i) \log_b P(x_i) \quad (1)$$

Όπου b η βάση του αλγορίθμου.

Συνήθεις τιμές για το b είναι το 2, ο αριθμός του Euler e , και το 10. Οι μονάδες μέτρησης της πληροφόρησης αντίστοιχα είναι bits (2), nats (e), bans (10). Όπως προκύπτει από τον ορισμό της εντροπίας, αν σε ένα πείραμα είναι γνωστό κάθε φορά το αποτέλεσμα, που θα προκύψει, τότε η εντροπία ισούται με μηδέν.

Στο πλαίσιο της μηχανικής μάθησης (Machine Learning) η εφαρμογή της ενδείκνυται για συνεχείς μεταβλητές.

3.2.2 Αμοιβαία Πληροφορία

Εφαρμόζοντας τη μεθοδολογία υπολογισμού της Αμοιβαίας Πληροφορίας (Mutual Information – MI) και συγκρίνοντας τα αποτελέσματα με την κλασσική μεθοδολογία ανάλυσης συσχετίσεων κατά Pearson (Pearson’s Correlation Analysis) εντοπίζονται τα οδικά τμήματα και οι στάσεις του μετρό που παρουσιάζουν υψηλή συσχέτιση όσον αφορά το φόρτο των οχημάτων που διέρχονται στο καθένα καθώς και της ζήτησης του επιβατικού κοινού. Αυτές οι δύο μεθοδολογίες χρησιμοποιούνται συχνά, ακόμα και σε δεδομένα που δεν είναι απαραίτητα χρονοσειρές. Για τη συγκεκριμένη εργασία μέσω αυτής της σύγκρισης γίνεται η επιλογή των χαρακτηριστικών που θα χρησιμοποιηθούν στην πρόβλεψη (Feature Selection), δηλαδή ποιι φωρατές και ποιες στάσεις θα χρησιμοποιηθούν στην πρόβλεψη.

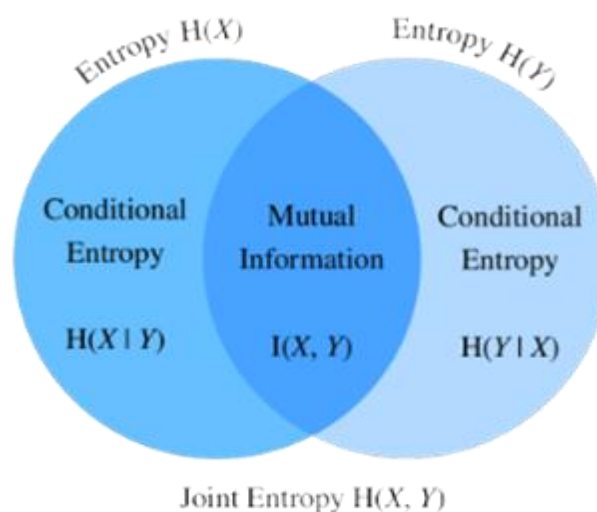
Η Αμοιβαία Πληροφορία (MI) δύο τυχαίων μεταβλητών, σύμφωνα με τη Θεωρία της Πληροφορίας (Information Theory), είναι μια μετρική που ποσοτικοποιεί το μέγεθος της πληροφορίας που λαμβάνεται για την πρώτη μεταβλητή, παρατηρώντας τη δεύτερη. Σε αντίθεση με την κλασσική ανάλυση συσχετίσεων, η αμοιβαία πληροφορία λαμβάνει υπόψη και μη γραμμικές συσχετίσεις, καθώς το μέγεθος που υπολογίζεται δε σχετίζεται με τους γραμμικούς ή μη γραμμικούς κανόνες εξέλιξης των δύο ποσοτήτων που εμπλέκονται, αλλά με την εντροπία κατά Shannon (Shannon’s Entropy). Άρα η αμοιβαία πληροφορία παίρνει τιμές ανάλογες της ανεξαρτησίας των μεταβλητών και δεν χρειάζεται τη συνδιασπορά που χρησιμοποιούν άλλες μέθοδοι.

Έστω x_n και y_n δύο τυχαίες μεταβλητές που ορίζονται στο ίδιο διάστημα με συναρτήσεις πυκνότητας πιθανότητας $p(x_n)$ και $p(y_n)$ και από κοινού συνάρτηση πυκνότητας πιθανότητας $p(x_n, y_n)$. Η Αμοιβαία Πληροφορία $I(x_n, y_n)$, η οποία ποσοτικοποιεί την αναμενόμενη πληροφορία που λαμβάνεται για τη x_n κατά την παρατήρηση της y_n δίνεται από την εξίσωση (2):

$$I(x_n, y_n) = -\sum x_n y_n P(x_n y_n) \log_2 \frac{P(x_n, y_n)}{P(x_n)P(y_n)} \quad (2)$$

Η αμοιβαία πληροφορία $I(X, Y)$ είναι ένα μέγεθος που μετράει την ποσότητα της πληροφορίας που μια τυχαία μεταβλητή περιέχει για μια άλλη τυχαία μεταβλητή. Με άλλα λόγια, υπολογίζει σε ποιον βαθμό μπορεί η γνώση που σημειώνεται για την δεύτερη μεταβλητή να μειώσει την αβεβαιότητα που υπάρχει για την πρώτη μεταβλητή. Το μέτρο αυτό ουσιαστικά, βασίζεται στην αμοιβαία εξάρτηση που υπάρχει μεταξύ των δύο μεταβλητών.

Η θεωρία Αμοιβαίας Πληροφορίας μπορεί να χρησιμοποιηθεί για την αξιολόγηση των εξαρτήσεων μεταξύ τυχαίων διακριτών και συνεχών μεταβλητών. Στο Διάγραμμα 4 απεικονίζεται το διάγραμμα Venn της αμοιβαίας πληροφορίας $I(X, Y)$ των συσχετιζόμενων μεταβλητών X και Y . Η περιοχή που καταλαμβάνεται και από τους δύο κύκλους ονομάζεται κοινή εντροπία (joint entropy) και συμβολίζεται ως $H(X, Y)$. Ο ολόκληρος αριστερός κύκλος είναι η μεμονωμένη εντροπία Shannon της μεταβλητής X , ενώ ο δεξιός κύκλος είναι η εντροπία της μεταβλητής Y και συμβολίζονται αντίστοιχα ως $H(X)$ και $H(Y)$. Τέλος, η υπό όρους εντροπία μεταξύ δύο τυχαίων μεταβλητών με κατανομή πιθανότητας υπό όρους, που ορίζεται ως $H(X|Y)$ αφορά τη δεσμευμένη πληροφόρηση μιας τυχαίας μεταβλητής X δεδομένου μιας άλλης τυχαίας μεταβλητής Y και είναι η ποσότητα της πληροφορίας που απαιτείται για την περιγραφή της έκβασης της μεταβλητής X δεδομένου ότι είναι γνωστή η τιμή της μεταβλητής Y . Αντίστοιχα, ως $H(Y|X)$ αναφέρεται στην εντροπία της μεταβλητής Y δεδομένου της X .



Διάγραμμα 4: Διάγραμμα Venn μεταξύ αμοιβαίας πληροφορίας και εντροπίας.

Έτσι προκύπτουν οι παρακάτω σχέσεις:

$$H(X) = H(X, Y) - H(Y|X) \quad (3)$$

$$I(X, Y) = H(X, Y) - H(X|Y) - H(Y|X) \quad (4)$$

Επιπλέον, η αμοιβαία πληροφορία έχει δύο βασικά χαρακτηριστικά:

1. Είναι συμμετρική, δηλαδή:

$$I(X, Y) = I(Y, X) \quad (5)$$

2. Είναι θετική, δηλαδή:

$$I(X, Y) \geq 0 \quad (6)$$

Σημειώνεται ότι δύο τυχαίες μεταβλητές X, Y με αμοιβαία πληροφόρηση ίση με μηδέν είναι μεταξύ τους ανεξάρτητες.

Ένας από τους περιορισμούς αυτής της προσέγγισης στην ανάλυση χρονοσειρών είναι ότι δε λαμβάνει υπόψη τη χρονική διαδοχή των διαφορετικών τιμών του κυκλοφοριακού φόρτου των οδικών τμημάτων και της ζήτησης των στάσεων αντίστοιχα αλλά τη στατιστική τους κατανομή μόνο. Ωστόσο, η εφαρμογή της αμοιβαίας πληροφορίας έχει αποδειχθεί πολύ αποτελεσματική για την επιλογή χαρακτηριστικών (Feature Selection) σε εργασίες ταξινόμησης και πρόβλεψης.

3.2.3 Επιλογή Χαρακτηριστικών

Κατά την κατασκευή ενός μοντέλου μηχανικής μάθησης, το πρώτο βήμα είναι συνήθως η επιλογή ενός υποσυνόλου σχετικών χαρακτηριστικών (Feature Selection). Ο στόχος της επιλογής χαρακτηριστικών στην μηχανική μάθηση είναι να προσαρμόσει καλύτερα τα δεδομένα στο πρόβλημα. Εντοπίζοντας και απορρίπτοντας εκείνα τα χαρακτηριστικά τα οποία παρέχουν περιττή πληροφορία, μειώνεται η διάσταση των δεδομένων αφού τα περιττά διανύσματα δεν συμμετέχουν στις περαιτέρω διαδικασίες.

Οι τεχνικές επιλογής χαρακτηριστικών αποσκοπούν στην βελτίωση της προγνωστικής απόδοσης ενός μοντέλου, στη μείωση των υπολογιστικών αναγκών και στη διευκόλυνση της ερμηνείας των αποτελεσμάτων.

Οι αλγόριθμοι επιλογής χαρακτηριστικών χωρίζονται στις εξής τρεις κατηγορίες:

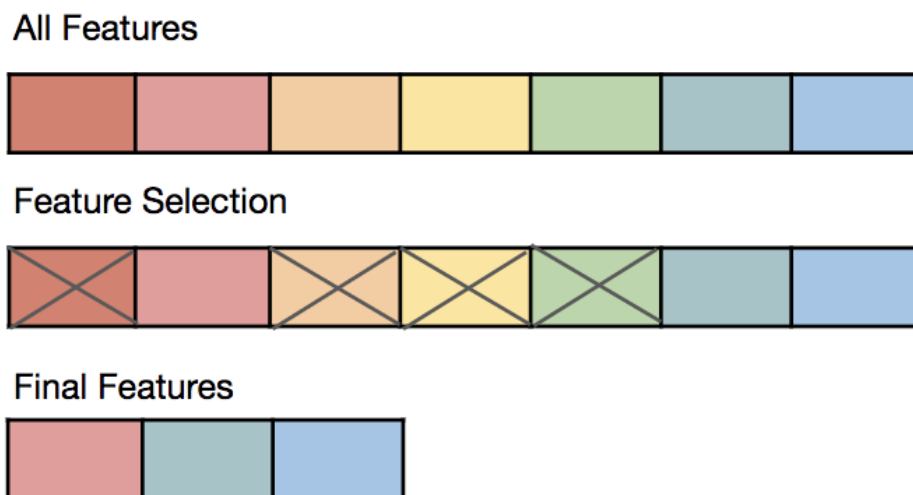
1. Οι μέθοδοι περιτύλιξης (Wrapper Methods), οι οποίες χρησιμοποιούν αλγόριθμους μάθησης στα αρχικά δεδομένα και επιλέγουν σχετικά χαρακτηριστικά με βάση την (εκτός δείγματος) απόδοση του μαθησιακού αλγορίθμου. Παράδειγμα μοντέλου περιτύλιξης αποτελεί η εκπαίδευση

ενός γραμμικού μοντέλου με διαφορετικούς συνδυασμούς χαρακτηριστικών και η επιλογή των χαρακτηριστικών που έχουν ως αποτέλεσμα την καλύτερη απόδοση εκτός δείγματος.

2. Οι μέθοδοι φίλτραρίσματος (Filter Methods) που δεν χρησιμοποιούν αλγόριθμο εκμάθησης στα αρχικά δεδομένα, αλλά λαμβάνουν υπόψη μόνο τα στατιστικά χαρακτηριστικά των δεδομένων εισόδου. Για παράδειγμα, μπορεί να γίνει επιλογή των χαρακτηριστικών για τα οποία ο συσχετισμός μεταξύ των χαρακτηριστικών και της μεταβλητής στόχου υπερβαίνει ένα όριο συσχέτισης. Επίσης, η συσχέτιση κατά Pearson και η Αμοιβαία Πληροφορία αποτελούν παραδείγματα μεθόδων φίλτραρίσματος [24].

3. Οι ενσωματωμένες μέθοδοι (Embedded Methods) οι οποίες είναι μια ομάδα τεχνικών που καλύπτουν όλες τις λειτουργίες που εκτελούν την επιλογή χαρακτηριστικών ως μέρος της διαδικασίας εκπαίδευσης του μοντέλου εκμάθησης. Η εκπαίδευση του μοντέλου παλινδρόμησης LASSO είναι ένα κλασικό παράδειγμα ενσωματωμένης μεθόδου για την επιλογή χαρακτηριστικών.

Το παρακάτω Σχήμα 1 απεικονίζει τη διαδικασία επιλογής χαρακτηριστικών, δηλαδή δείχνει τι συμβαίνει στα αρχικά δεδομένα μετά τη χρήση αλγορίθμου επιλογής χαρακτηριστικών.

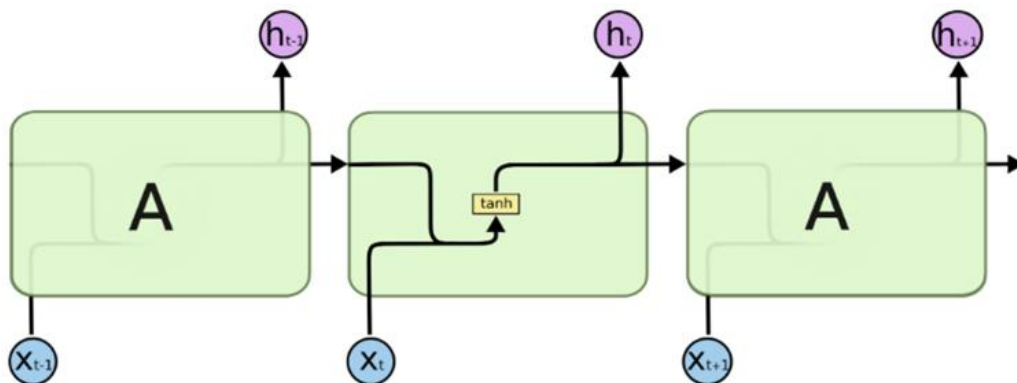


Σχήμα 1: Σχηματική απεικόνιση της διαδικασίας Επιλογής Χαρακτηριστικών.

3.2.4 Δίκτυα Μακράς-Βραχείας Μνήμης

Τα Δίκτυα Μακράς και Βραχείας Μνήμης (Long Short-Term Memory Networks) που συνήθως αναφέρονται ως "LSTM" για συντομία, αποτελούν μια τροποποιημένη εκδοχή των Επαναλαμβανόμενων Νευρωνικών Δικτύων (Recurrent Neural Networks – RNNs) και έχουν αποδειχθεί ιδιαίτερα αποτελεσματικά στην ταξινόμηση σε κλάσεις, την επεξεργασία και την πρόβλεψη συνεχών τιμών και δεδομένων σε μορφή χρονοσειρών.

Με δομή πιο περίπλοκη από τα Vanilla RNN, είναι σχεδιασμένα με τρόπο τέτοιο ώστε να έχουν την ικανότητα να αποθηκεύουν πληροφορίες από μεγάλες χρονικές περιόδους δηλαδή επιτρέπουν την αλληλεπίδραση μεταξύ των τρεχουσών και όλων των προηγούμενων εισόδων, και όχι μόνο της τελευταίας όπως τα RNNs. Όλα τα RNNs δίκτυα έχουν την μορφή μιας αλυσίδας επαναλαμβανόμενων δομών (modules) του νευρωνικού δικτύου. Στα απλά RNN αυτή η επαναλαμβανόμενη δομή (Module) έχει απλή αρχιτεκτονική, η οποία φαίνεται στο Σχήμα 2, μόνο ένα στρώμα υπερεφαπτομένης (tanh), δηλαδή έναν απλό μηχανισμό ανατροφοδότησης που μεταφέρει πληροφορία σε ένα νευρώνα από τα δεδομένα της προηγούμενης χρονικής στιγμής και μόνο [25].



Σχήμα 2: Επαναλαμβανόμενη δομή σε ένα τυπικό RNN, πηγή [25].

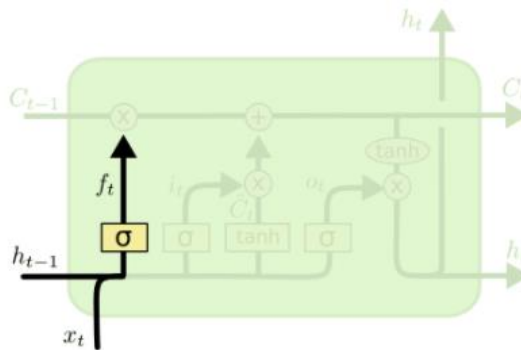
Η ιδέα των επαναληπτικών νευρωνικών δικτύων εμφανίζεται στη βιβλιογραφία από την δεκαετία του 1990 [26]. Στις πρώτες εφαρμογές και αρχιτεκτονικές, γίνεται άμεσα εμφανές το πρόβλημα της εκλειπόμενης κλίσης (vanishing gradient), καθώς οι αρχιτεκτονικές αδυνατούν να συγκρατήσουν και να χρησιμοποιήσουν με αποδοτικό τρόπο την παρελθοντική γνώση.

Αντιθέτως στα δίκτυα μακράς και βραχείας μνήμης επιτρέπεται η μεταφορά παρελθοντικών πληροφοριών από τους προηγούμενους στους επόμενους νευρώνες, κάτι που είναι ιδιαίτερα χρήσιμο στην πρόβλεψη χρονοσειρών. Το διακριτικό στοιχείο των LSTM δικτύων είναι η ικανότητά τους να αναγνωρίζουν συσχετίσεις σε δεδομένα που απέχουν αυθαίρετα χρονικά διαστήματα μεταξύ τους.

Επιπλέον, όπως μαρτυρά και η ονομασία τους, συνδυάζουν πολύ αποτελεσματικά τόσο την βραχυπρόθεσμη (short-term) πληροφορία που δέχονται όσο και την ιστορική τους μακροπρόθεσμη (long-term) γνώση. Τα δύο αυτά χαρακτηριστικά τα ξεχωρίζουν από άλλες μορφές επαναλαμβανόμενων νευρωνικών δικτύων και μεθόδων ανάλυσης χρονοσειρών.

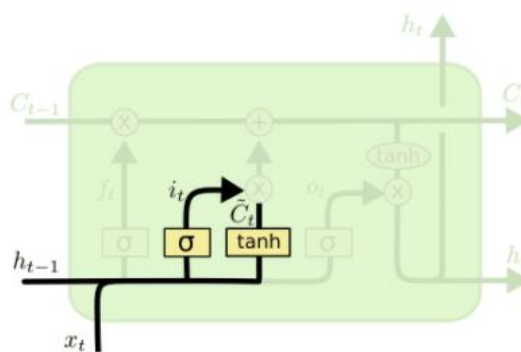
Η βασική δομή μιας μονάδας LSTM βασίζεται στο κύτταρο μνήμης c_t (memory cell) το οποίο είναι υπεύθυνο για την διατήρηση της μνήμης της ακολουθίας δεδομένων. Σε κάθε βήμα το LSTM δέχεται ως είσοδο μία παρατήρηση x_t (διάνυσμα), ενημερώνει το κύτταρο μνήμης c_t και παράγει ως έξοδο την κατάσταση ή κρυφή κατάσταση h_t (hidden state). Η διαδικασία ενημέρωσης χρησιμοποιεί έναν μηχανισμό που βασίζεται σε τρεις πύλες (gates), Σχήματα 3 έως 5.

- Μία πύλη λήθης f_t (forget gate) που ελέγχει πόση πληροφορία από το παρελθόν c_{t-1} διατηρείται. Ελέγχοντας την προηγούμενη κατάσταση h_{t-1} και το περιεχόμενο της τωρινής εισόδου x_t , χρησιμοποιεί τη σιγμοειδή συνάρτηση με τιμές στο εύρος $[0,1]$. Το 0 σημαίνει ‘απέριψε την πληροφορία’ ενώ το 1 ‘κράτα την πληροφορία’.



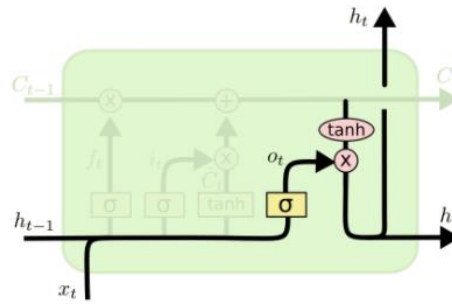
Σχήμα 3: Η πύλη λήθης (forget gate) του δικτύου LSTM.

- Μία πύλη εισόδου i_t (input gate) που ελέγχει το πόση πληροφορία από την είσοδο x_t θα ενημερώσει την κατάσταση του κυττάρου μνήμης c_t ώστε να διατηρηθεί για τα επόμενα βήματα.



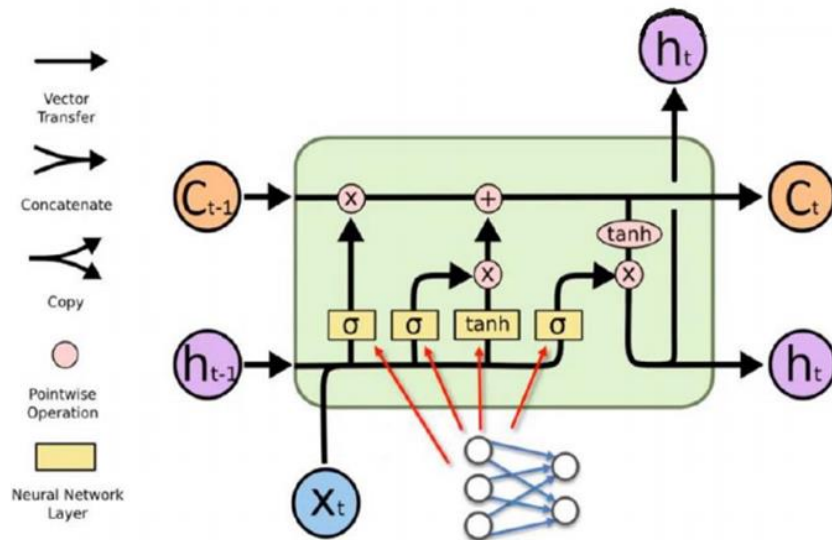
Σχήμα 4: Η πύλη εισόδου (input gate) του δικτύου LSTM.

- Μία πύλη εξόδου o_t (output gate) η οποία ελέγχει την ποσότητα πληροφορίας που της τροφοδοτείται από την μνήμη στην έξοδο ως κρυφή κατάσταση h_t .



Σχήμα 5: Η πύλη εξόδου (output gate) του δικτύου LSTM.

Στο παρακάτω Σχήμα 6 φαίνεται η τυπική αναπαράσταση μιας δομής δικτύου LSTM.



Σχήμα 6: Αναπαράσταση της τυπικής δομής του δικτύου LSTM, πηγή [25].

Οι σχέσεις που ακολουθούνται για τον ακριβή υπολογισμό της διαδικασίας είναι οι εξής:

$$i_t = \sigma(W_{x_i}x_t + W_{h_i}h_{t-1} + b_i) \quad (7)$$

$$f_t = \sigma(W_{x_f}x_t + W_{h_f}h_{t-1} + b_f) \quad (8)$$

$$o_t = \sigma(W_{x_o}x_t + W_{h_o}h_{t-1} + b_o) \quad (9)$$

$$C_t = f_t c_{t-1} + i_t \tan h(W_{x_c}x_t + W_{h_c}h_{t-1} + b_c) \quad (10)$$

$$h_t = o_t \tan h(c_t) \quad (11)$$

όπου i, f, o, c είναι οι πύλες εισόδου, λήθης, εξόδου και μνήμης αντίστοιχα. Τα βάρη W και τα διανύσματα των τιμών κατωφλίων b είναι παράμετροι που προσαρμόζονται κατά την εκπαίδευση του δικτύου [27].

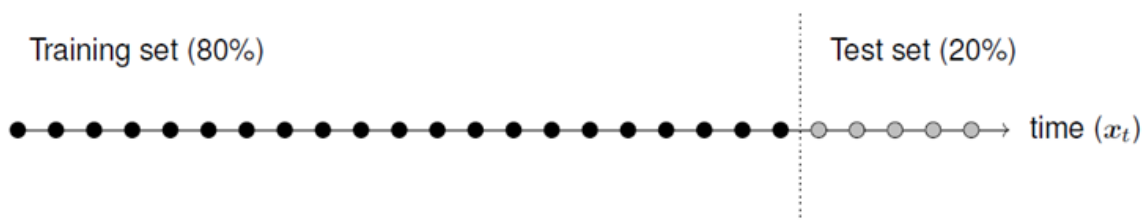
Τα LSTM μπορούν να διαχειρίζονται τόσο απλά δεδομένα (π.χ. στατικές εικόνες) όσο και ακολουθίες αυτών (π.χ. ροή ήχου, ροή εικόνων και βίντεο, χρονοσειρές), καθιστώντας τα ικανά να χρησιμοποιηθούν σε προβλήματα αναγνώρισης γραφής, αναγνώρισης φωνής, ερμηνείας εικόνας και βίντεο, ανάλυσης και πρόβλεψης χρονοσειρών, παρατήρησης ανωμαλιών σε δικτυακή κίνηση και πολλά άλλα.

3.2.5 Αξιολόγηση Μοντέλου Πρόβλεψης

Μετά την επιλογή του κατάλληλου μοντέλου και των παραμέτρων αυτού, ακολουθεί η αξιολόγηση της ακρίβειας της πρόβλεψης που προκύπτει. Αυτό επιτυγχάνεται με το διαχωρισμό των δεδομένων σε ένα σύνολο εκπαίδευσης (training set) και ένα σύνολο ελέγχου (test set). Το training set χρησιμοποιείται για την εκπαίδευση του μοντέλου και τον υπολογισμό των προβλεπόμενων τιμών. Στη συνέχεια το test set για να μετρήσει πόσο καλά το εκπαιδευμένο μοντέλο μπορεί να παράγει προβλέψεις σε δεδομένα που δεν είχαν παρατηρηθεί προηγουμένως. Η επίδοση πρόβλεψης που επιτυγχάνεται στο σύνολο ελέγχου είναι αυτή που μπορεί να αναμένεται όταν θα γίνουν πραγματικές προβλέψεις για μελλοντικά χρονικά βήματα.

Είναι σύνηθες να χωρίζεται το 80% των αρχικών χρονοσειρών για το training set και το υπόλοιπο 20% για το test set. Αυτό απεικονίζεται στο Διάγραμμα 5 όπου κάθε τελεία αντιστοιχεί σε ένα χρονικό βήμα σε μια χρονική σειρά.

Είναι σημαντικό να μην ανακατεύονται οι χρονοσειρές, καθώς αυτό θα αφαιρέσει τη χρονική του δομή. Επίσης, το test set πρέπει να φτάσει μετά το train set για να αποφευχθεί πρόβλεψη του παρελθόντος.



Διάγραμμα 5: Διαχωρισμός Χρονοσειρών.

Σημειώνεται ότι οι τιμές αυτές εξαρτώνται και από το μέγεθος των δεδομένων. Ιδανικά το test set θα πρέπει να είναι τουλάχιστον, τόσο μεγάλο, όσο και η πρόβλεψη που απαιτείται.

Για να προσδιοριστεί η ακρίβεια μιας πρόβλεψης χρησιμοποιούνται διάφορα μέτρα αξιολόγησης. Το σφάλμα της πρόβλεψης ορίζεται ως η διαφορά μεταξύ μίας παρατηρούμενης τιμής και της πρόβλεψης της. Με τον όρο σφάλμα δεν εννοείται κάποιου είδους λάθος αλλά το μέρος της παρατηρούμενης τιμής που δεν μπορεί να προβλεφθεί και εκφράζεται ως:

$$e_t = Y_t - \hat{Y}_t \quad (12)$$

όπου Y_t η πραγματική παρατηρούμενη τιμή

\hat{Y}_t η προβλεπόμενη τιμή για τη χρονική στιγμή t .

Επομένως για να μελετηθεί με αξιοπιστία μία μέθοδος πρόβλεψης, αναλύεται η διαχρονική εξέλιξη των σφαλμάτων πρόβλεψης με χρήση των παρακάτω βασικότερων κριτηρίων αξιολόγησης:

Μέτρα Αξιολόγησης - Accuracy Metrics:

- Μέσο Απόλυτο Σφάλμα (Mean Absolute Error - MAE)

Το μέσο απόλυτο σφάλμα μετράει τη μέση απόλυτη απόκλιση των προβλεπόμενων τιμών από τις πραγματικές για όλο το σύνολο της πρόβλεψης. Λόγω της απόλυτης τιμής του οι θετικές τιμές δεν αναιρούνται από τις αρνητικές. Επίσης το σφάλμα είναι στην ίδια κλίμακα με τα δεδομένα, σχέση (13):

$$MAE = \frac{1}{N} \sum_{i=1}^N |Y_t - \hat{Y}_t| \quad (13)$$

- Μέσο Τετραγωνικό Σφάλμα (Mean Squared Error - MSE)

Το μέσο τετραγωνικό σφάλμα MSE ορίζεται ως το άθροισμα των τετραγώνων των σφαλμάτων πρόβλεψης διαιρούμενο με τον αριθμό των περιόδων n , στις οποίες έγιναν οι προβλέψεις και υπολογίζεται σύμφωνα με την σχέση (14):

$$MSE = \frac{1}{N} \sum_{i=1}^N (Y_t - \hat{Y}_t)^2 \quad (14)$$

Με το MSE παρουσιάζεται το πρόβλημα ότι οι μονάδες μέτρησης του είναι υψωμένες στο τετράγωνο ενώ της εξαρτημένης μεταβλητής όχι. Για τον λόγο αυτό χρησιμοποιείται η τετραγωνική ρίζα του μεγέθους αυτού.

- Ρίζα Μέσου Τετραγωνικού Σφάλματος (Root Mean Squared Error - RMSE)

Η Ρίζα Μέσου Τετραγωνικού Σφάλματος RMSE ορίζεται ως η ρίζα του αθροίσματος των τετραγώνων των σφαλμάτων πρόβλεψης διαιρούμενο με τον αριθμό των περιόδων n , στις οποίες έγιναν οι προβλέψεις και υπολογίζεται σύμφωνα με την σχέση (15):

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (Y_t - \hat{Y}_t)^2} \quad (15)$$

Με το RMSE οι μονάδες μέτρησης είναι κοινές με αυτές της εξεταζόμενης μεταβλητής και ερμηνεύει καλύτερα την ύπαρξη μεγάλων σφαλμάτων. Συνεπώς το RMSE είναι στατιστικά πιο έγκυρο και ως αυτού χρησιμοποιείται συχνότερα.

- Μέσο Απόλυτο Ποσοστιαίο Σφάλμα (Mean Absolute Percentage Error - MAPE)

Το Μέσο Απόλυτο Ποσοστιαίο Σφάλμα MAPE ορίζεται ως η απόλυτη απόκλιση της πραγματικής από την αντίστοιχη προβλεπόμενη τιμή προς την πραγματική τιμή διαιρούμενη με τον αριθμό των περιόδων n , στις οποίες έγιναν οι προβλέψεις και υπολογίζεται σύμφωνα με την σχέση (16):

$$MAPE = \frac{1}{N} \sum_{i=1}^N \frac{|Y_t - \hat{Y}_t|}{Y_t} * 100\% \quad (16)$$

Το MAPE είναι απαλλαγμένο από μονάδες μέτρησης που το καθιστά αξιόπιστο κριτήριο στη σύγκριση μεθόδων πρόβλεψης χρονοσειρών αλλά και μεταξύ διαφορών χρονοσειρών. Όσο το MAPE τείνει στο 0 τόσο πιο αξιόπιστη χαρακτηρίζεται η μέθοδος πρόβλεψης.

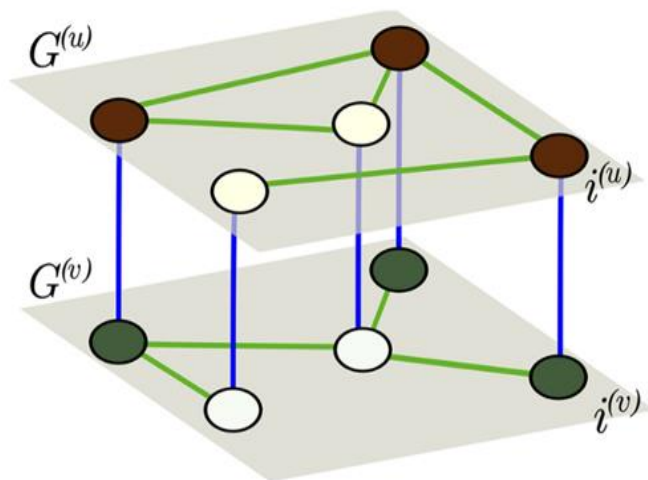
Γενικά για την επιλογή της καταλληλότερης μεθόδου πρόβλεψης με την χρήση κριτηρίων αξιολόγησης, η μέθοδος που εμφανίζει τις μικρότερες τιμές στα επιμέρους κριτήρια αποτελεί και την πιο αξιόπιστη. Συχνά όμως, υπάρχουν περιπτώσεις όπου μία μέθοδος πρόβλεψης μπορεί να εμφανίσει μικρότερη τιμή σε ένα κριτήριο και μεγαλύτερη σε ένα άλλο σε σχέση με μία εναλλακτική μέθοδο. Σε τέτοιες περιπτώσεις η επιλογή της μεθόδου εναπόκειται στην κρίση και την εμπειρία του ειδικού που θα λάβει την απόφαση [28].

3.2.6 Multiplex Δίκτυα

Στην πρόσφατη βιβλιογραφία, παρατηρείται μια αξιόλογη προσπάθεια για τον χαρακτηρισμό και τη μοντελοποίηση των λεγόμενων πολυπλεξικών δικτύων, γνωστών ως Multiplex Networks.

Τα δίκτυα πολυπλεξίας είναι απλώς ένας τύπος αυτού που συνήθως ονομάζεται δίκτυο πολλαπλών επιπέδων, στο οποίο υπάρχει ένα σύνολο επιπέδων με ακριβώς ίδιο σύνολο κόμβων σε κάθε διαφορετικό επίπεδο. Μεταξύ των επιπέδων μπορεί να υπάρχουν διαφορετικές συνδέσεις ωστόσο υπάρχει μια σχέση ένας προς έναν μεταξύ του κόμβου σε ένα επίπεδο και των κόμβων στα άλλα επίπεδα [29]. Αυτή η απλή σύνδεση ένας προς έναν είναι πολύ σημαντική για αυτήν την έρευνα.

Με άλλα λόγια, τα δίκτυα πολυπλεξίας (multiplex networks) θεωρούνται ως σύνολο κόμβων σε διαφορετικά επίπεδα. Σε κάθε επίπεδο το σύνολο των κόμβων είναι το ίδιο, αλλά οι συνδέσεις μεταξύ των κόμβων μπορεί να είναι διαφορετικές στα επίπεδα. Οι συνδέσεις μεταξύ των επιπέδων περιγράφονται από ένα «δίκτυο επιπέδων». Στο Σχήμα 7 δίνεται η αναπαράσταση ενός πολυπλεξικού δικτύου.



Σχήμα 7: Αναπαράσταση ενός Πολυπλεξικού Δικτύου, πηγή [29].

Ένα τυπικό παράδειγμα multiplex δικτύων είναι τα συστήματα μεταφορών, όπου οι κόμβοι είναι τοποθεσίες και τα επίπεδα αντιπροσωπεύουν διαφορετικούς τρόπους μεταφοράς, π.χ. λεωφορείο, μετρό, τρένα κ.λπ. Ένα δίκτυο μεταφορών μπορεί να έχει ένα πρώτο επίπεδο που αναπαριστά τις οδούς του οδικού δικτύου κι ένα δεύτερο επίπεδο για τις στάσεις των μέσων μαζικής μεταφοράς (M.M.M.).

Κεφάλαιο 4

Ανάλυση και Αποτελέσματα

4.1 Εισαγωγή

Η επεξεργασία και ανάλυση των δεδομένων καθώς και η εφαρμογή της μεθόδου του δικτύου βαθιάς μάθησης LSTM πραγματοποιήθηκαν με τη γλώσσα προγραμματισμού Python 3 [29]. Το κύριο περιβάλλον ανάπτυξης του αλγορίθμου που χρησιμοποιήθηκε ήταν το Jupyter Notebook.

4.1.1 Εγκατάσταση πακέτων

Αρχικά, πραγματοποιήθηκε η εγκατάσταση αρκετών βιβλιοθηκών στο περιβάλλον Jupyter Notebook της Python, που χρησιμοποιήθηκαν για την ανάκτηση, ανάγνωση των δεδομένων και την τροποποίησή τους, τη σύνταξη του μοντέλου νευρωνικού δικτύου και την εκπαίδευση του, την αξιολόγηση του αλγορίθμου και των μετρικών τιμών του, καθώς και για την κατασκευή διαγραμμάτων και άλλων σημαντικών διαδικασιών. Ορισμένες από αυτές τις βιβλιοθήκες συνοψίζονται παρακάτω:

- NumPy: Μια βιβλιοθήκη που επιτρέπει τον εύκολο χειρισμό μεγάλων πολυδιάστατων λιστών και πινάκων. Χρησιμοποιώντας την μπορούν να εκτελεστούν μαθηματικές και λογικές πράξεις σε πίνακες.
- Pandas: Μια βιβλιοθήκη που χρησιμοποιείται για επεξεργασία και ανάλυση δεδομένων από διάφορες μορφές αρχείων όπως csv, JSON, SQL. Κατασκευάστηκε με σκοπό την εφαρμογή και βελτίωση των υπαρχόντων εργαλείων χειρισμού δεδομένων που βρίσκονται σε γλώσσες στατιστικού προγραμματισμού όπως η R. Επιπλέον, προσφέρει καλή υποστήριξη για χειρισμό δεδομένων χρονοσειρών που είναι η μορφή δεδομένων που χρησιμοποιείται σε αυτή την εργασία.
- Scikit-learn: Μια βιβλιοθήκη για χειρισμό συνόλων δεδομένων και εφαρμογή διαφορετικών αλγορίθμων μηχανικής μάθησης.
- Matplotlib: Μια δισδιάστατη βιβλιοθήκη σχεδίασης που χρησιμοποιείται για τη δημιουργία γραφημάτων στη γλώσσα προγραμματισμού Python.
- TensorFlow: Μια βιβλιοθήκη που αποτελείται από χαμηλού επιπέδου υπολογισμούς (π.χ. πράξεις μεταξύ tensors) διαφόρων αλγορίθμων μηχανικής μάθησης, όπως νευρωνικά δίκτυα. Έχει σχεδιαστεί ώστε να συνδυάζει την υπολογιστική άλγεβρα με τεχνικές βελτιστοποίησης υπολογίζοντας γρήγορα και αποδοτικά μαθηματικές

εκφράσεις που απαιτούν μεγάλο χρόνο εκτέλεσης λόγω της υψηλής υπολογιστικής τους πολυπλοκότητας.

- Keras: Μια φιλική προς το χρήστη βιβλιοθήκη λογισμικού ανοιχτού κώδικα που χρησιμοποιείται για την εύκολη εφαρμογή διαφόρων αρχιτεκτονικών βαθέων νευρωνικών δικτύων (Deep Neural Networks) παρέχοντας τη διεπαφή με την Python. Τρέχει πάνω από άλλες βιβλιοθήκες τις οποίες χρησιμοποιεί ως Backend (TensorFlow, Theano κ.α.) επειδή από μόνη της δεν πραγματοποιεί χαμηλού επιπέδου υπολογισμούς.

4.2 Βάση Δεδομένων

4.2.1 Περιγραφή Αρχικής Βάσης Δεδομένων

Στην παρούσα διπλωματική εργασία ανακτήθηκαν από την ιστοσελίδα <https://www.data.gov.gr/> περίπου 512.000 δεδομένα σχετικά με την κυκλοφορία δρόμων στην Αττική και περίπου 1.553.000 σχετικά με το επιβατικό κοινό του ΟΑΣΑ κατά την περίοδο από 2 Μαρτίου 2021 έως 30 Ιουνίου 2021. Να σημειωθεί ότι ο όγκος των δεδομένων είναι μεγάλος επομένως απαιτείται ισχυρό υπολογιστικό σύστημα καθώς και πολύς χρόνος για την ανάκτηση τους από το δοθέν API αλλά και την επεξεργασία τους.

Η αρχική μορφή της βάσης δεδομένων για την κυκλοφορία στους δρόμους της Αττικής φαίνεται στην Εικόνα 4:

Unnamed: 0	deviceid	countedcars	appprocesstime	road_name	road_info	average_speed
0	0	MS116	1917.0	2021-03-01T19:00:00Z	Α. ΚΗΦΙΣΟΥ ΚΥΡΙΟΣ ΔΡΟΜΟΣ ΜΕ ΚΑΤΕΥΘΥΝΣΗ ΠΕΙΡΑΙΑ ΜΕΤΑ ΤΗ ΡΑ...	106.972353
1	1	MS116	797.0	2021-03-01T21:00:00Z	Α. ΚΗΦΙΣΟΥ ΚΥΡΙΟΣ ΔΡΟΜΟΣ ΜΕ ΚΑΤΕΥΘΥΝΣΗ ΠΕΙΡΑΙΑ ΜΕΤΑ ΤΗ ΡΑ...	102.736512
2	2	MS125	370.0	2021-03-01T19:00:00Z	Α. ΚΗΦΙΣΟΥ ΡΑΜΠΑ ΕΞΟΔΟΥ ΠΡΟΣ ΟΔΟ ΠΕΙΡΑΙΩΣ ΚΑΙ ΓΕΦΥΡΑ ΑΝΑΣ...	60.000000
3	3	MS125	142.0	2021-03-01T21:00:00Z	Α. ΚΗΦΙΣΟΥ ΡΑΜΠΑ ΕΞΟΔΟΥ ΠΡΟΣ ΟΔΟ ΠΕΙΡΑΙΩΣ ΚΑΙ ΓΕΦΥΡΑ ΑΝΑΣ...	61.577465
4	4	MS175	2185.0	2021-03-01T19:00:00Z	Α. ΣΥΓΓΡΟΥ ΚΥΡΙΟΣ ΔΡΟΜΟΣ ΜΕ ΚΑΤΕΥΘΥΝΣΗ ΦΑΛΗΡΟ ΜΕΤΑ ΤΗ Λ...	84.282380
5	5	MS175	889.0	2021-03-01T21:00:00Z	Α. ΣΥΓΓΡΟΥ ΚΥΡΙΟΣ ΔΡΟΜΟΣ ΜΕ ΚΑΤΕΥΘΥΝΣΗ ΦΑΛΗΡΟ ΜΕΤΑ ΤΗ Λ...	84.894263
6	6	MS207	1667.0	2021-03-01T19:00:00Z	Β. ΣΟΦΙΑΣ ΚΥΡΙΟΣ ΔΡΟΜΟΣ ΜΕ ΚΑΤΕΥΘΥΝΣΗ ΠΡΟΣ ΚΗΦΙΣΙΑ (Β.Α....	47.283143
7	7	MS207	865.0	2021-03-01T21:00:00Z	Β. ΣΟΦΙΑΣ ΚΥΡΙΟΣ ΔΡΟΜΟΣ ΜΕ ΚΑΤΕΥΘΥΝΣΗ ΠΡΟΣ ΚΗΦΙΣΙΑ (Β.Α....	52.153757
8	8	MS219	1247.0	2021-03-01T19:00:00Z	ΒΑΣ. ΚΩΝΣΤΑΝΤΙΝΟΥ ΚΥΡΙΟΣ ΔΡΟΜΟΣ ΜΕ ΚΑΤΕΥΘΥΝΣΗ ΠΡΟΣ ΧΙΛΤΟΝ (Β.Α.)...	22.595028
9	9	MS219	482.0	2021-03-01T21:00:00Z	ΒΑΣ. ΚΩΝΣΤΑΝΤΙΝΟΥ ΚΥΡΙΟΣ ΔΡΟΜΟΣ ΜΕ ΚΑΤΕΥΘΥΝΣΗ ΠΡΟΣ ΧΙΛΤΟΝ (Β.Α.)...	46.975104
10	10	MS224	1402.0	2021-03-01T19:00:00Z	ΚΑΛΛΙΠΡΩΗΣ ΚΥΡΙΟΣ ΔΡΟΜΟΣ ΜΕ ΚΑΤΕΥΘΥΝΣΗ ΑΝΑΤΟΛΙΚΑ ΜΕΤΑ ΤΗΝ...	64.568474

Εικόνα 4: Αρχική Ακατέργαστη Βάση Δεδομένων που παρέχει η Περιφέρεια Αττικής.

Κάθε γραμμή αντιστοιχεί σε έναν συγκεκριμένο φωρατή που για μία συγκεκριμένη ώρα καταγράφει τον φόρτο και το μέσο όρο της ταχύτητας των οχημάτων που διέρχονται από το σημείο που είναι τοποθετημένος.

Αναλυτικότερα, τα χαρακτηριστικά των δεδομένων είναι:

- deviceid: μοναδικό όνομα (ID) για κάθε φωρατή
- countedcars: φόρτος οχημάτων που καταγράφει κάθε φωρατής μία συγκεκριμένη ώρα
- appprocesstime: ημερομηνία και ώρα καταγραφής δεδομένων
- road_name: όνομα οδού όπου είναι τοποθετημένος κάθε φωρατής
- road_info: αναλυτική περιγραφή της οδού
- average_speed: μέσος όρος ταχύτητας του συνόλου των οχημάτων που καταγράφει κάθε φωρατής

Αντίστοιχα, η αρχική μορφή της βάσης δεδομένων του επιβατικού κοινού του ΟΑΣΑ φαίνεται στην ακόλουθη Εικόνα 5:

Unnamed: 0	dv_validations	dv_agency	dv_platenum_station	dv_route	routes_per_hour	load_dt	date_hour
0	0	7	ZKZ 9873	833 - ΕΥΓΕΝΕΙΑ-ΧΑΡΑΥΓΗ-ΠΕΙΡΑΙΑΣ Β (ΚΥΚΛΙΚΗ) :....	1.0	2021-03-02T06:14:25Z	2021-03-01T13:00:00Z
1	1	8	ZKZ 9873	832 - ΕΥΓΕΝΕΙΑ-ΧΑΡΑΥΓΗ-ΠΕΙΡΑΙΑΣ Α (ΚΥΚΛΙΚΗ) :....	1.0	2021-03-02T06:14:25Z	2021-03-01T14:00:00Z
2	2	12	ZKZ 9873	833 - ΕΥΓΕΝΕΙΑ-ΧΑΡΑΥΓΗ-ΠΕΙΡΑΙΑΣ Β (ΚΥΚΛΙΚΗ) :....	1.0	2021-03-02T06:14:25Z	2021-03-01T15:00:00Z
3	3	8	ZKZ 9873	832 - ΕΥΓΕΝΕΙΑ-ΧΑΡΑΥΓΗ-ΠΕΙΡΑΙΑΣ Α (ΚΥΚΛΙΚΗ) :....	1.0	2021-03-02T06:14:25Z	2021-03-01T16:00:00Z
4	4	9	ZKZ 9873	833 - ΕΥΓΕΝΕΙΑ-ΧΑΡΑΥΓΗ-ΠΕΙΡΑΙΑΣ Β (ΚΥΚΛΙΚΗ) :....	1.0	2021-03-02T06:14:25Z	2021-03-01T17:00:00Z
5	5	20	ZKZ 9873	833 - ΕΥΓΕΝΕΙΑ-ΧΑΡΑΥΓΗ-ΠΕΙΡΑΙΑΣ Β (ΚΥΚΛΙΚΗ) :....	1.0	2021-03-02T06:14:25Z	2021-03-01T18:00:00Z
6	6	3	ZKZ 9873	833 - ΕΥΓΕΝΕΙΑ-ΧΑΡΑΥΓΗ-ΠΕΙΡΑΙΑΣ Β (ΚΥΚΛΙΚΗ) :....	1.0	2021-03-02T06:14:25Z	2021-03-01T19:00:00Z
7	7	16	ZKZ 9873	NaN	NaN	2021-03-02T06:14:25Z	2021-03-01T20:00:00Z
8	8	5	ZKZ 9873	833 - ΕΥΓΕΝΕΙΑ-ΧΑΡΑΥΓΗ-ΠΕΙΡΑΙΑΣ Β (ΚΥΚΛΙΚΗ) :....	1.0	2021-03-02T06:14:25Z	2021-03-01T21:00:00Z
9	9	5	ZKZ 9873	832 - ΕΥΓΕΝΕΙΑ-ΧΑΡΑΥΓΗ-ΠΕΙΡΑΙΑΣ Α (ΚΥΚΛΙΚΗ) :....	1.0	2021-03-02T06:14:25Z	2021-03-01T23:00:00Z
10	10	3	ZKZ 9873	832 - ΕΥΓΕΝΕΙΑ-ΧΑΡΑΥΓΗ-ΠΕΙΡΑΙΑΣ Α (ΚΥΚΛΙΚΗ) :....	2.0	2021-03-02T06:14:25Z	2021-03-02T00:00:00Z

Εικόνα 5: Αρχική Ακατέργαστη Βάση Δεδομένων που παρέχει ο ΟΑΣΑ.

Κάθε γραμμή αντιστοιχεί σε συγκεκριμένη λεωφορειακή γραμμή ή σταθερή συγκοινωνία που για μία συγκεκριμένη ώρα καταγράφει τον αριθμό επικυρωμένων εισιτηρίων μετακίνησης/μετεπιβίβασης.

Από τη συγκεκριμένη βάση δεδομένων χρησιμοποιήθηκαν μόνο οι ακόλουθες στήλες:

- `dv_validations`: επικυρώσεις μετακίνησης/μετεπιβίβασης
- `dv_platenum_station`: όνομα σταθερής συγκοινωνίας ή αριθμός πινακίδας λεωφορείων που καλύπτουν συγκεκριμένη λεωφορειακή γραμμή
- `date_hour`: ημερομηνία και ώρα καταγραφής δεδομένων

Είναι εμφανές ότι τα ακατέργαστα δεδομένα του πραγματικού κόσμου (real world raw data) των παραπάνω δύο βάσεων χρειάζονται επεξεργασία ώστε να μπορέσουν να χρησιμοποιηθούν για την εκπαίδευση του νευρωνικού δικτύου στη συνέχεια.

4.1.3 Προ – Επεξεργασία Βάσης Δεδομένων

Για να μπορούν να αξιοποιηθούν οι παραπάνω βάσεις δεδομένων χρησιμοποιήθηκε για κάθε μια από αυτές η βιβλιοθήκη Pandas. Η χρήση της συγκεκριμένης βιβλιοθήκης βοήθησε στη δημιουργία νέων πινάκων όπου οι ονομασίες των φωρατών, των στάσεων ΣΤΑ.ΣΥ. και των λεωφορειακών γραμμών τοποθετήθηκαν ως στήλες, ενώ στις γραμμές τοποθετήθηκαν οι αντίστοιχες μετρήσεις, δηλαδή ο φόρτος οχημάτων ανά ώρα και αντίστοιχα οι επικυρώσεις εισιτηρίων μετακίνησης/μετεπιβίβασης ανά ώρα.

```
df_traffic = pd.read_csv("Traffic.csv", parse_dates=True)
df_traffic["appprocesstime"] = pd.to_datetime(df_traffic.appprocesstime)
new_traffic = df_traffic.sort_values(by='appprocesstime')
data_traffic = new_traffic.pivot(index="appprocesstime", columns="deviceid", values = "countedcars")
```

```
df_oasa = pd.read_csv("OASA.csv", parse_dates=True)
df_oasa["date_hour"] = pd.to_datetime(df_oasa.date_hour)
new_oasa = df_oasa.sort_values(by='date_hour')
data_oasa = new_oasa.pivot_table(index="date_hour", columns="dv_platenum_station", values = "dv_validations")
```

Κατά την ανάλυση των δεδομένων επιλέχθηκε το χρονικό πλαίσιο 07:00 έως 23:00 κάθε ημέρας. Οι ώρες που αφαιρέθηκαν πέραν του ότι δεν αποτελούν ώρες αιχμής, παρουσιάζουν σφάλματα στις καταγραφές και η αφαίρεση τους εξυπηρετεί και τη μείωση του όγκου των δεδομένων χωρίς αυτά να επηρεάζονται ποιοτικά.

Επιπλέον, όσες κενές τιμές υπήρχαν στον επιλεγμένο χρονικό ορίζοντα συμπληρώθηκαν με την επόμενη έγκυρη τιμή μέσω της εντολής `DataFrame.fillna(method='bfill')`. Έπειτα, έγινε επιλογή όσων φωρατών και στάσεων ΣΤΑ.ΣΥ. βρίσκονται εντός του δακτυλίου της Αθήνας όπως φαίνεται στον κώδικα παρακάτω. Έτσι τα αρχικά δεδομένα μειώθηκαν κατά πολύ, βοηθώντας στην καλύτερη ερμηνεία τους.

```
data_traffic=data_traffic.between_time('07:00:00' , "23:00:00")
data_traffic_new=data_traffic.fillna(method='ffill',axis=1)
```

```
traffic = data_traffic_new.filter(['MS207','MS219','MS224','MS225','MS228','MS230','MS231','MS233','MS234','MS238',
'MS239','MS242','MS243','MS244','MS245','MS246','MS248','MS249','MS252','MS255',
'MS256','MS258','MS261','MS310','MS317','MS342','MS350','MS356','MS364','MS365',
'MS367','MS374','MS376','MS377','MS381','MS390','MS392','MS393','MS415','MS416',
'MS418','MS426','MS427','MS432','MS434','MS435','MS440','MS446','MS447','MS474',
'MS482','MS483','MS484','MS488','MS500','MS515','MS516','MS572','MS574','MS575',
'MS581','MS631','MS634','MS636','MS638','MS639','MS658','MS669','MS670','MS672',
'MS673','MS700','MS704','MS710','MS712','MS770','MS772','MS933'], axis=1)
```

```
data_oasa=data_oasa.between_time('07:00:00' , "23:00:00")
data_oasa_new = data_oasa.fillna(method='bfill',axis=1)
```

```
oasa_data=data_oasa_new.filter(['ΑΚΡΟΠΟΛΗ','ΑΜΠΕΛΟΚΗΠΟΙ','ΒΙΚΤΩΡΙΑ','ΕΥΑΓΓΕΛΙΣΜΟΣ','ΘΗΣΕΙΟ','ΚΕΡΑΜΕΙΚΟΣ',
'ΜΕΤΑΞΟΥΡΓΕΙΟ','ΜΕΓΑΡΟ ΜΟΥΣΙΚΗΣ','ΜΟΝΑΣΤΗΡΑΚΙ','ΟΜΟΝΟΙΑ','ΠΑΝΟΡΜΟΥ',
'ΠΑΝΕΠΙΣΤΗΜΙΟ','ΠΕΤΡΑΛΩΝΑ','ΣΥΝΤΑΓΜΑ','ΣΥΓΓΡΟΥ ΦΙΞ','ΤΑΥΡΟΣ'], axis=1)
```

Μέσω αυτής της διαδικασίας δημιουργήθηκαν δύο ομοιόμορφες βάσεις δεδομένων, απαλλαγμένες από λάθη και κενές τιμές, οι οποίες ενώθηκαν δημιουργώντας την ακόλουθη μορφή της τελικής βάσης δεδομένων, Εικόνα 6:

	ΑΚΡΟΠΟΛΗ	ΑΜΠΕΛΟΚΗΠΟΙ	ΒΙΚΤΩΡΙΑ	ΕΥΑΓΓΕΛΙΣΜΟΣ	ΘΗΣΕΙΟ	ΚΕΡΑΜΕΙΚΟΣ	ΜΕΤΑΞΟΥΡΓΕΙΟ	ΜΕΓΑΡΟ ΜΟΥΣΙΚΗΣ	ΜΟΝΑΣΤΗΡΑΚΙ	ΟΜΟΝΟΙΑ	...	MS670	MS672	MS673	MS700	MS704	MS710	MS712	MS770	MS772	MS933
2021-03-02 07:00:00+00:00	23	219	263	226	24	96	119	107	84	322	...	861	1065	1344	428	1030	520	641	1049	2269	998
2021-03-02 08:00:00+00:00	41	782	642	728	61	250	329	470	168	824	...	1069	1190	1469	444	1382	563	837	1226	2574	1265
2021-03-02 09:00:00+00:00	146	1205	917	1100	115	374	528	722	417	1518	...	1316	1390	1431	557	1290	719	859	1444	2507	1568
2021-03-02 10:00:00+00:00	216	1549	976	1351	123	509	491	985	633	2110	...	1251	1355	1380	458	1303	713	814	1373	2060	1657
2021-03-02 11:00:00+00:00	189	1228	769	1070	134	411	410	857	735	2062	...	1081	1377	1258	472	1186	614	953	1261	1852	1485
2021-03-02 12:00:00+00:00	163	957	634	884	142	307	365	706	729	1956	...	1101	1566	1318	445	1176	592	920	1286	1090	1684
2021-03-02 13:00:00+00:00	139	899	684	824	143	309	410	680	721	2018	...	1261	1785	1404	515	1218	684	1050	1533	1599	2014
2021-03-02 14:00:00+00:00	160	979	758	905	158	329	434	737	756	1952	...	1284	1818	1362	504	1180	698	997	1698	939	2409
2021-03-02 15:00:00+00:00	178	1044	774	1011	198	332	391	866	701	1876	...	1236	1922	1328	571	1227	631	945	1688	705	2763
2021-03-02 16:00:00+00:00	207	1314	840	1277	194	444	500	973	785	2060	...	1254	1908	1344	569	1149	693	1058	1931	969	2405
2021-03-02 17:00:00+00:00	262	1421	1060	1377	204	442	518	880	884	2435	...	1140	1710	1109	522	1136	623	995	1733	480	2293
2021-03-02 18:00:00+00:00	320	1135	1063	1152	183	400	532	800	1055	2206	...	975	1364	979	491	1202	562	873	1860	631	2350
2021-03-02 19:00:00+00:00	573	1109	978	1380	207	506	489	846	1766	2156	...	758	1041	745	305	984	385	608	1308	744	1994
2021-03-02 20:00:00+00:00	518	926	980	1129	158	387	364	733	1358	1505	...	561	860	575	289	643	322	489	1035	921	1545
2021-03-02 21:00:00+00:00	200	706	757	727	138	252	308	594	652	1152	...	335	578	396	162	364	203	257	542	986	868
2021-03-02 22:00:00+00:00	152	587	454	613	171	191	229	479	401	606	...	190	268	174	99	218	100	147	259	766	434
2021-03-02 23:00:00+00:00	47	304	230	348	49	167	134	216	123	226	...	77	212	153	60	129	54	83	185	760	227
2021-03-03 07:00:00+00:00	33	211	229	189	19	82	129	94	69	296	...	878	1190	1378	472	1084	514	695	1188	1381	1106

Εικόνα 6: Επεξεργασμένη - Ενοποιημένη Βάση Δεδομένων.

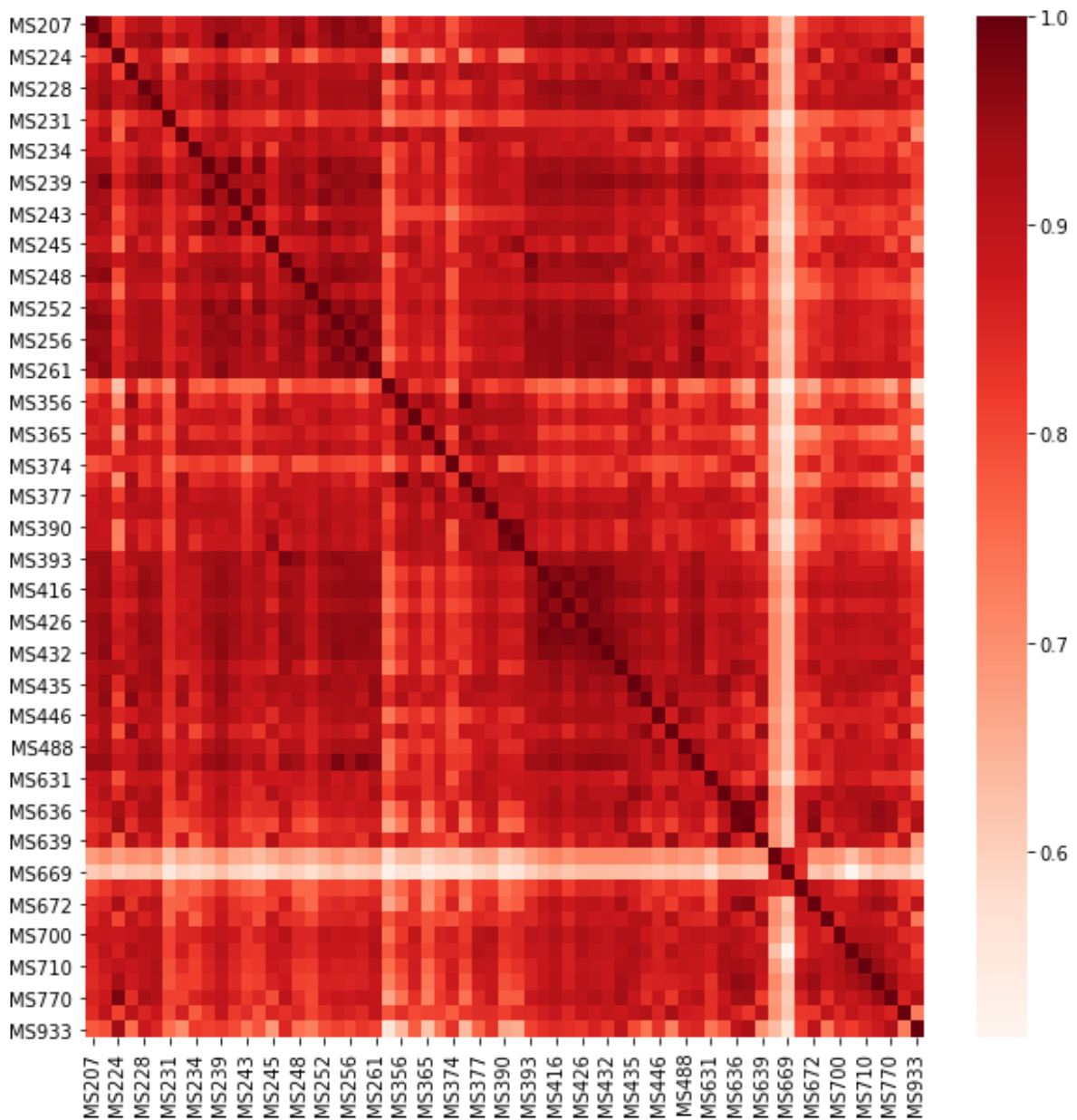
4.3 Επεξεργασία Δεδομένων

4.3.1 Εφαρμογή Μεθόδων Συσχέτισης και Αποτελέσματα Συσχετίσεων

Οι χωροχρονικές συσχετίσεις στο οδικό δίκτυο προσδιορίστηκαν μέσω της συσχέτισης κατά Pearson και της Αμοιβαίας Πληροφορίας. Με τη θεώρηση ότι υπάρχουν τμήματα δρόμων και στάσεις με κοινή συμπεριφορά αν προκύψουν υψηλές τιμές των δύο μετρικών. Ο υπολογισμός των δύο μεγεθών έγινε για όλα τα ζεύγη χρονοσειρών φωρατών και στάσεων δηλαδή τις δύο αρχικές βάσεις χωριστά καθώς και για την ενοποιημένη βάση.

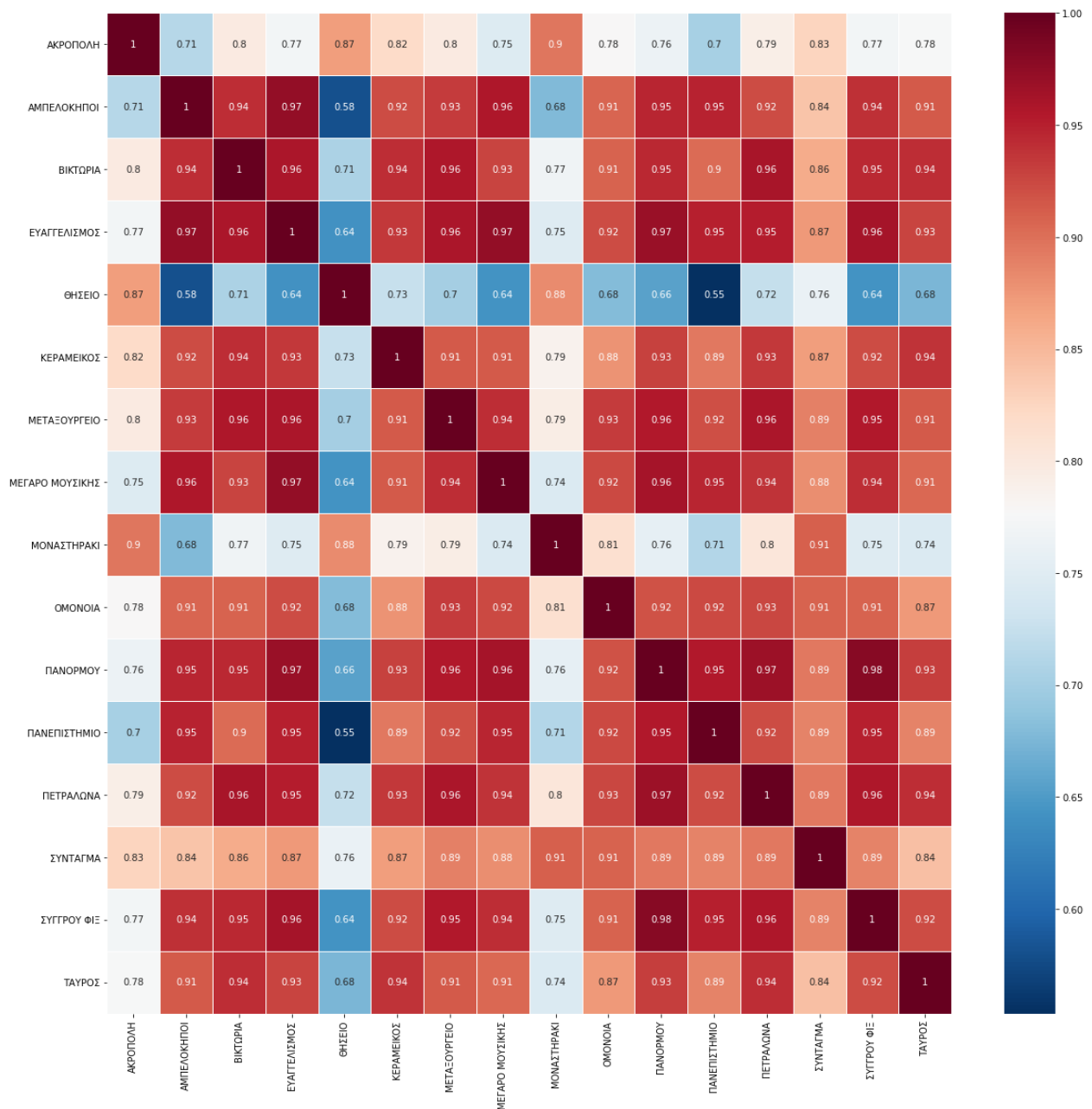
Υπενθυμίζεται πως με αυτές τις δύο μεθοδολογίες λαμβάνονται υπόψη κατά κύριο λόγο οι πιθανοτικές κατανομές των κυκλοφοριακών φόρτων των τμημάτων και της επιβατικής ζήτησης των στάσεων και δε λαμβάνεται υπόψη η χρονική διάσταση του προβλήματος. Για αυτό το λόγο χρησιμοποιούνται χρονοσειρές ίδιου χρονικού βήματος. Τα δύο τμήματα που σχετίζονται με βάση αυτές τις μεθοδολογίες θα έχουν παραπλήσιες τιμές φόρτων κατά τη διάρκεια της ημέρας όχι όμως απαραίτητα πολύ κοντινές.

Οι παρακάτω θερμικοί χάρτες που παρουσιάζονται, Διάγραμμα 6 και Διάγραμμα 7, δημιουργήθηκαν μέσω της βιβλιοθήκης Pandas με την εντολή `pandas.DataFrame.corr()` και αναπαρίστανται με τη χρήση της `seaborn` που είναι μια βιβλιοθήκη οπτικοποίησης δεδομένων βασισμένη στη βιβλιοθήκη `matplotlib`. Αυτό που δείχνουν είναι η συσχέτιση κατά Pearson για τα δεδομένα κυκλοφορίας I.X. και αντίστοιχα για τα δεδομένα των στάσεων ΣΤΑ.ΣΥ.



Διάγραμμα 6: Θερμικός Χάρτης Συσχέτισης Κατά Pearson για τα δεδομένα κυκλοφορίας Ι.Χ.

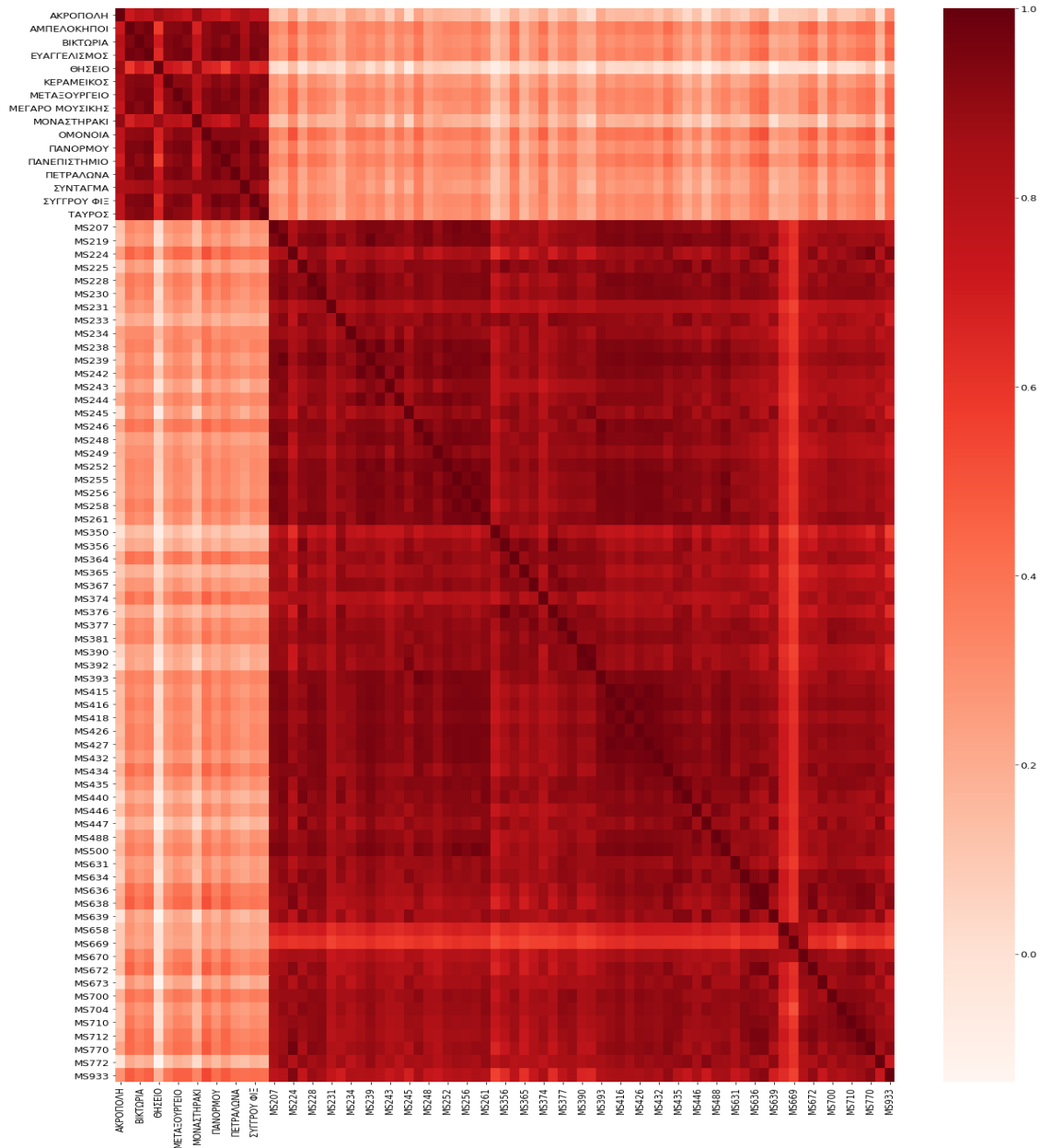
Από τον παραπάνω θερμικό χάρτη παρατηρείται πως οι φωρατές που είναι τοποθετημένοι κατά μήκος μιας οδού παρουσιάζουν υψηλή συσχέτιση μεταξύ τους, Για παράδειγμα οι MS245, MS248, MS252, MS255, MS256 οι οποίοι είναι τοποθετημένοι σε διαφορετικά σημεία, είτε ανάντη είτε κατόντη, επί της οδού Βασιλίσσης Σοφίας παρουσιάζουν υψηλότερες συσχετίσεις μεταξύ τους σε σχέση με τους φωρατές που είναι τοποθετημένοι σε πιο μακρινά σημεία, όπως οι MS639, MS669 που βρίσκονται κοντά στη Λεωφόρο Αθηνών.



Διάγραμμα 7: Θερμικός Χάρτης Συσχέτισης Κατά Pearson για τα δεδομένα των στάσεων ΣΤΑ.ΣΥ. εντός του δακτυλίου Αθηνών.

Όσο αφορά το θερμικό χάρτη των στάσεων παρατηρούνται χαμηλές συσχετίσεις μεταξύ στάσεων που δε βρίσκονται στην ίδια γραμμή του μετρό της Αθήνας, πιο συγκεκριμένα η χαμηλότερη συσχέτιση φαίνεται με το σκούρο μπλε χρώμα, μεταξύ των στάσεων Θησείο με Πανεπιστήμιο και Θησείο με Αμπελόκηποι, δηλαδή ανάμεσα σε μια στάση της γραμμής 1 με στάσεις που ανήκουν στις γραμμές 2 και 3 αντίστοιχα. Ενώ στάσεις που βρίσκονται στην ίδια γραμμή, όπως για παράδειγμα η στάση Ευαγγελισμός με τη στάση Μέγαρο Μουσικής που η μία διαδέχεται την άλλη στη γραμμή 3, εμφανίζουν συσχέτιση πολύ υψηλή, κοντά στο 1.

Στο ακόλουθο Διάγραμμα 8, φαίνεται ο θερμικός χάρτης κατά Pearson για την ενοποιημένη βάση δεδομένων:

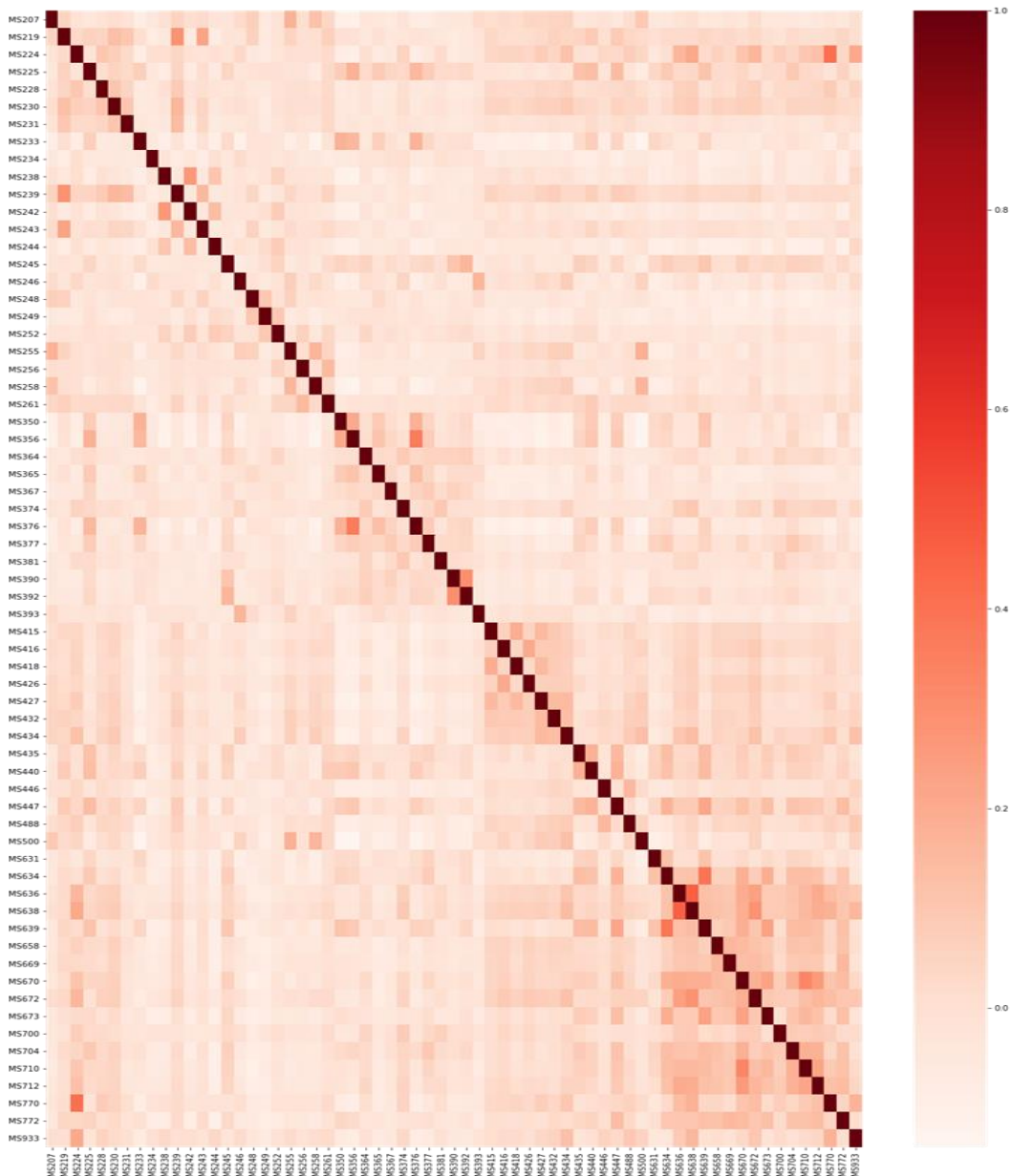


Διάγραμμα 8: Θερμικός χάρτης κατά Pearson για την Ενοποιημένη βάση δεδομένων.

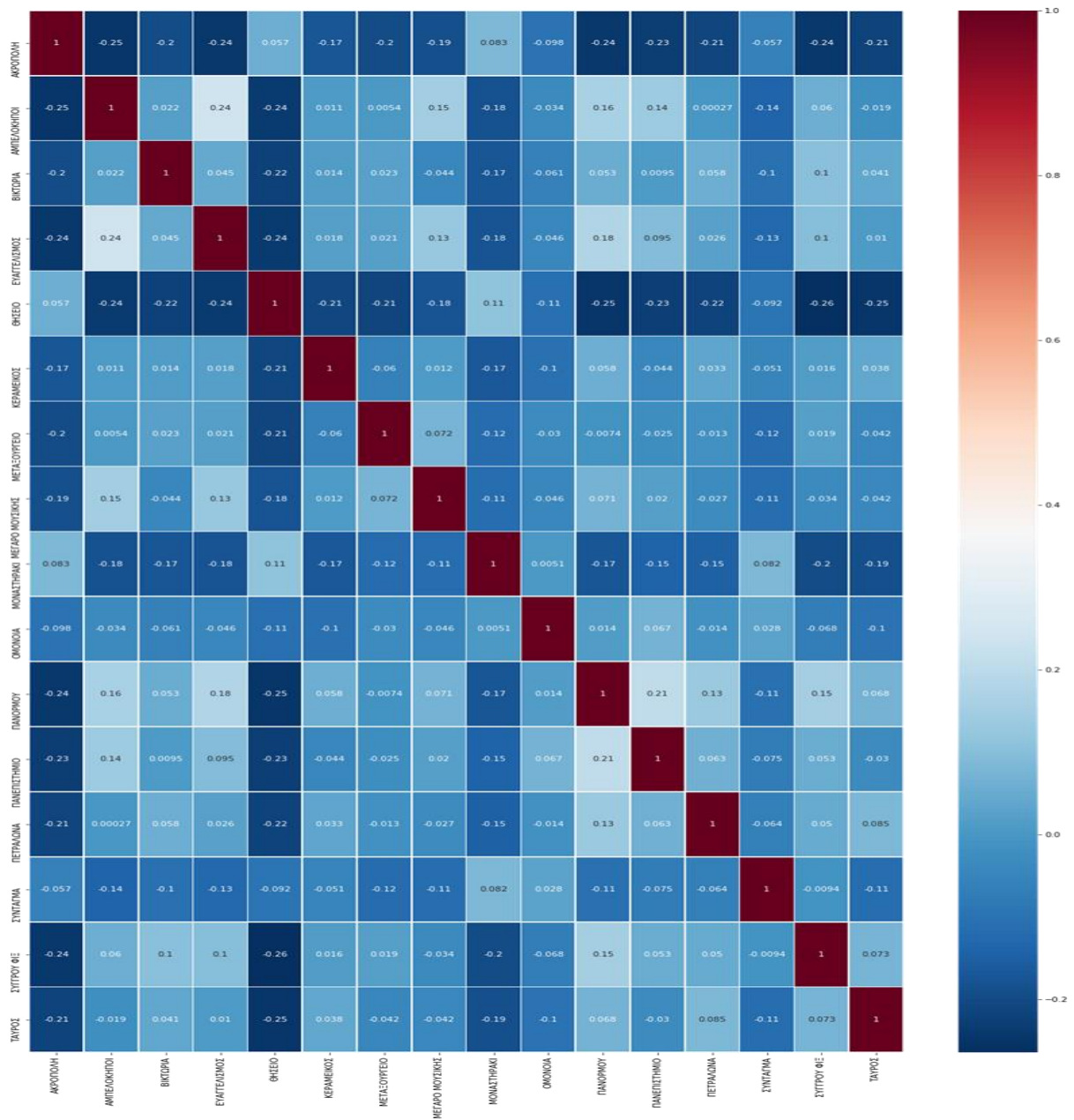
Από το θερμικό χάρτη συσχέτισης μεταξύ φωρατών και στάσεων παρατηρείται πως στάσεις μετρό που βρίσκονται κοντά στα οδικά τμήματα που είναι τοποθετημένοι οι φωρατές εμφανίζουν σχετικά μεγαλύτερες συσχετίσεις, για παράδειγμα ο MS224 επί της οδού Καλλιρόης έχει μεγαλύτερη συσχέτιση με τη στάση Συγγρού-Φιξ και μικρότερη με τη στάση Θησείο.

Ο υπολογισμός της Αμοιβαίας Πληροφορίας οδικών τμημάτων βάσει κυκλοφοριακών φόρτων και αντίστοιχα στάσεων βάσει επιβατικής ζήτησης, έγινε χρησιμοποιώντας τη βιβλιοθήκη scikit-learn [30], μία από τις πιο διαδεδομένες βιβλιοθήκες μηχανικής μάθησης, κι εν συνεχεία τα αποτελέσματα αναπαραστάθηκαν σε θερμικούς χάρτες με χρήση των ίδιων βιβλιοθηκών οπτικοποίησης που αναφέρθηκαννωρίτερα.

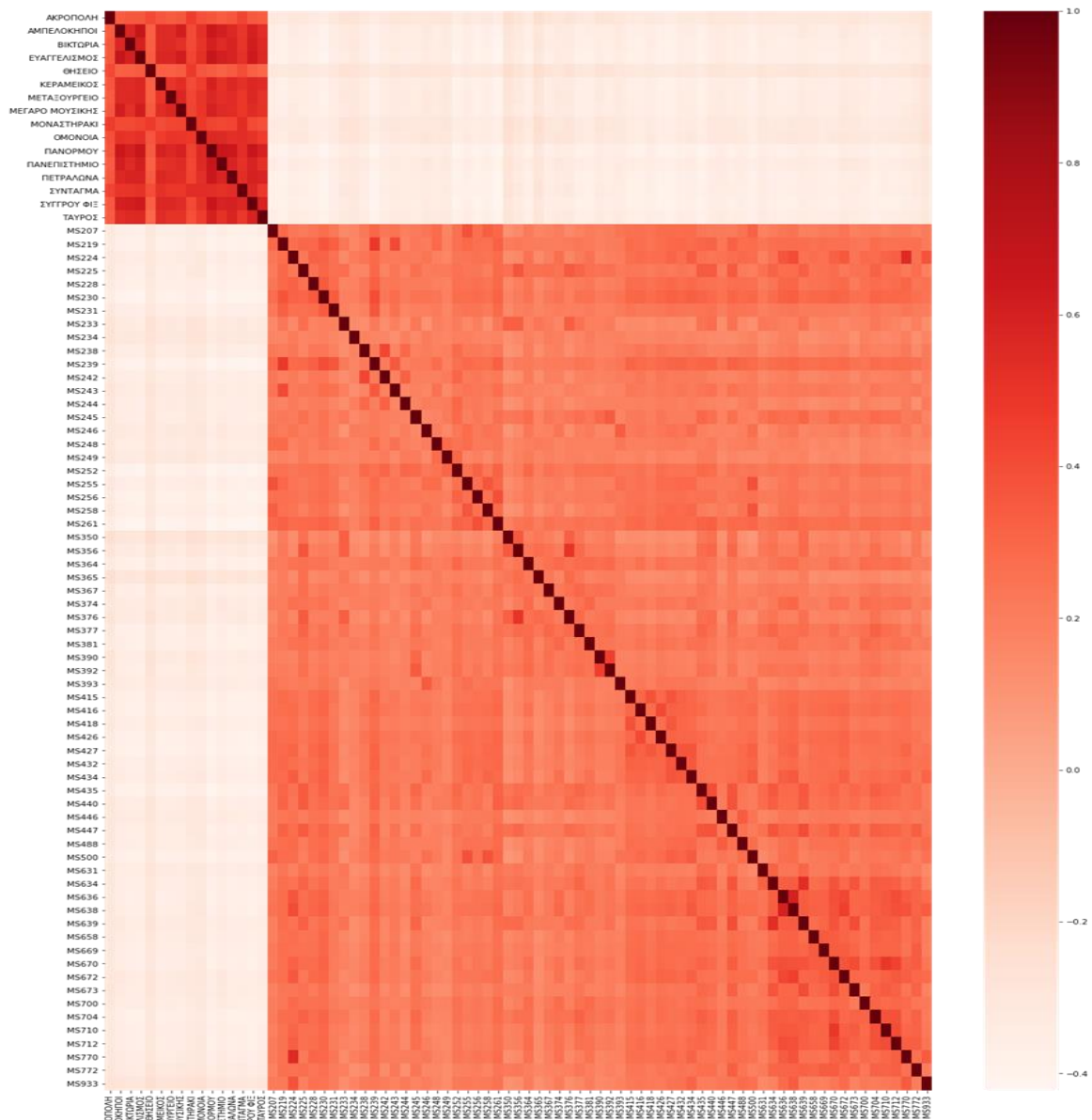
Στα Διάγραμμα 9 έως 11 παρατίθενται ο θερμικοί χάρτες για όλες τις βάσεις δεδομένων:



Διάγραμμα 9: Θερμικός χάρτης Αμοιβαίας Πληροφορίας για τα δεδομένα κυκλοφορίας.



Διάγραμμα 10: Θερμικός χάρτης Αμοιβαίας Πληροφορίας για τις επιλεγμένες στάσεις ΣΤΑ.ΣΥ εντός του δακτυλίου Αθηνών.



Διάγραμμα 11: Θερμικός χάρτης Αμοιβαίας Πληροφορίας για την ενοποιημένη βάση δεδομένων.

Στους χάρτες που προκύπτουν από τις δύο μεθοδολογίες παρατηρούνται κάποια κοινά μοτίβα όμως παρουσιάζουν και πολλές διαφορές, παρ'ότι τα δύο μεγέθη εκφράζουν σχεδόν το ίδιο. Ο θερμικός χάρτης της Αμοιβαίας Πληροφορίας έχει περισσότερο ανοιχτό χρώμα συνολικά, υποδεικνύοντας λιγότερο ισχυρές συσχετίσεις. Επειδή με τη συσχέτιση κατά Pearson υπολογίζεται η γραμμική συσχέτιση δύο μεταβλητών, η Αμοιβαία Πληροφορία μπορεί να θεωρηθεί πιο γενική μέθοδος αφού λαμβάνει υπόψη και μη γραμμικές συσχετίσεις. Συνεπώς, η χρήση διαφορετικών μετρικών στον εντοπισμό των εξαρτήσεων σε ένα οδικό δίκτυο οδηγεί σε διαφορετικά συμπεράσματα, τα οποία πρέπει να αξιολογούνται και να εξετάζεται στη συνέχεια αν βελτιώνουν την πρόβλεψη των κυκλοφοριακών συνθηκών.

Οι θερμικοί χάρτες που αφορούν τα δεδομένα των στάσεων είναι εύκολο να διαβαστούν μιας και δεν περιλαμβάνουν μεγάλο όγκο δεδομένων. Αντιθέτως ο χάρτης για την ενοποιημένη βάση είναι δύσκολο να διαβαστεί εξαιτίας του μεγάλου όγκου πληροφορίας που περιλαμβάνει και του μεγέθους του. Έτσι, ο πίνακας ΜΙ που παρουσιάζεται, Πίνακας 1, και έχει δημιουργηθεί για την ενοποιημένη βάση, βοηθάει περισσότερο από την οπτική απεικόνισή του.

	ΑΚΡΟΠΟΛΗ	ΑΜΠΕΛΟΚΗΦΟΙ	ΒΙΚΤΩΡΙΑ	ΕΥΑΓΓΕΛΙΣΜΟΣ	ΘΗΣΕΙΟ	ΚΕΡΑΜΕΙΚΟΣ	ΜΕΤΑΞΟΥΡΓΕΙΟ	ΜΕΓΑΡΟ ΜΟΥΣΙΚΗΣ	ΜΟΝΑΣΤΗΡΑΚΙ	ΟΜΟΝΟΙΑ	...
ΑΚΡΟΠΟΛΗ	6.284228	0.756583	0.847309	0.812131	1.186585	0.888233	0.840789	0.829877	1.258855	0.912657	...
ΑΜΠΕΛΟΚΗΦΟΙ	0.802827	6.036994	1.505891	2.011574	0.790874	1.431262	1.463110	1.701903	0.888821	1.240705	...
ΒΙΚΤΩΡΙΑ	0.886758	1.421956	6.293637	1.535369	0.800809	1.417210	1.483726	1.304170	0.893534	1.194679	...
ΕΥΑΓΓΕΛΙΣΜΟΣ	0.882237	1.933357	1.555924	6.107672	0.835809	1.449942	1.505742	1.672016	0.944413	1.237344	...
ΘΗΣΕΙΟ	1.195147	0.755861	0.768636	0.801351	6.175403	0.795775	0.803869	0.822509	1.296583	0.907923	...
ΚΕΡΑΜΕΙΚΟΣ	0.911808	1.414176	1.455235	1.505139	0.771854	6.387806	1.281049	1.399457	0.852816	1.089327	...
ΜΕΤΑΞΟΥΡΓΕΙΟ	0.831443	1.375245	1.486331	1.468155	0.750392	1.236833	6.321497	1.517425	0.910270	1.258228	...
ΜΕΓΑΡΟ ΜΟΥΣΙΚΗΣ	0.913796	1.778755	1.289774	1.778650	0.894481	1.460603	1.656697	6.264924	1.036400	1.221398	...
ΜΟΝΑΣΤΗΡΑΚΙ	1.392105	1.088234	1.079737	1.144429	1.431772	1.057138	1.252481	1.175572	5.928343	1.298468	...
ΟΜΟΝΟΙΑ	1.117352	1.379091	1.327805	1.398546	1.009431	1.185885	1.380283	1.312203	1.276193	5.813230	...
ΠΑΝΟΡΜΟΥ	0.895546	1.812936	1.607352	1.944671	0.819520	1.603561	1.491640	1.652777	1.012957	1.427061	...
ΠΑΝΕΠΙΣΤΗΜΙΟ	0.820959	1.746855	1.461302	1.646559	0.776515	1.299617	1.415509	1.447504	0.969379	1.468896	...
ΠΕΤΡΑΛΩΝΑ	0.832848	1.414943	1.603625	1.560007	0.788523	1.518017	1.437624	1.353485	0.942460	1.277622	...
ΣΥΝΤΑΓΜΑ	1.092446	1.087805	1.207126	1.155231	0.997003	1.299358	1.193124	1.120488	1.380749	1.424940	...
ΣΥΓΓΡΟΥ ΦΙΕ	0.814517	1.501393	1.666485	1.686744	0.759109	1.419483	1.474844	1.299543	0.877830	1.174929	...
ΤΑΥΡΟΣ	0.753396	1.317577	1.504557	1.430012	0.635881	1.453691	1.321129	1.231834	0.790735	1.044533	...
MS207	0.068754	0.161015	0.118340	0.130120	0.122915	0.068915	0.128797	0.129894	0.133310	0.230189	...
MS215	0.111961	0.195156	0.127748	0.175530	0.113241	0.132073	0.178002	0.155254	0.158964	0.256991	...
MS224	0.155873	0.317074	0.228603	0.248711	0.173020	0.174605	0.254010	0.287171	0.173036	0.364398	...
MS225	0.157446	0.208962	0.155434	0.160270	0.175300	0.132277	0.178653	0.181109	0.186967	0.338845	...
MS228	0.070635	0.195212	0.111377	0.165856	0.089061	0.136514	0.144597	0.178877	0.078365	0.238435	...
MS230	0.122218	0.164204	0.142047	0.174752	0.110532	0.152000	0.117311	0.188097	0.122743	0.210816	...
MS231	0.093689	0.218767	0.196814	0.193660	0.077335	0.170940	0.184459	0.214537	0.099692	0.230400	...
MS233	0.140436	0.113334	0.118011	0.099102	0.146616	0.118239	0.169986	0.101333	0.193532	0.224032	...
MS234	0.105629	0.158089	0.162788	0.165797	0.101998	0.124922	0.169186	0.148125	0.132682	0.243656	...
MS236	0.074268	0.181212	0.138734	0.150900	0.116720	0.103258	0.169350	0.172452	0.140875	0.232185	...
MS239	0.108620	0.145887	0.117375	0.165051	0.100877	0.063078	0.144555	0.172124	0.135563	0.182454	...
MS242	0.064566	0.165422	0.153820	0.171688	0.085803	0.133571	0.148379	0.155290	0.093359	0.159954	...
MS243	0.175364	0.247120	0.191340	0.200298	0.139261	0.171714	0.223195	0.192923	0.149584	0.237836	...
MS244	0.094031	0.114936	0.107179	0.083425	0.064846	0.073073	0.143940	0.091656	0.074586	0.132217	...
MS245	0.097078	0.204918	0.164027	0.194916	0.123200	0.155931	0.173608	0.188625	0.158239	0.225838	...
MS246	0.086049	0.218449	0.193343	0.204883	0.111203	0.166892	0.217264	0.209107	0.088145	0.288457	...
MS248	0.136704	0.183433	0.153054	0.164000	0.110388	0.162376	0.166110	0.149641	0.113354	0.200845	...
MS249	0.097104	0.155632	0.133240	0.155770	0.057380	0.137375	0.183693	0.152547	0.119617	0.223263	...

Πίνακας 1: Πίνακας Αμοιβαίας Πληροφορίας για την Ενοποιημένη Βάση Δεδομένων.

Αξιοποιώντας τη δομή του Πίνακα 1 όπου στις γραμμές πρώτα είναι οι στάσεις και μετά οι φωρατές και το ίδιο και στις στήλες, επιλέγεται ο φωρατής – στόχος (target) για τον οποίο θα αναπτυχθεί στη συνέχεια το μοντέλο πρόβλεψης καθώς και οι φωρατές και στάσεις με τις οποίες παρουσιάζει μεγαλύτερη συσχέτιση Αμοιβαίας Πληροφορίας (MI).

Επομένως μέσω αυτής της μεθοδολογίας που χρησιμοποιεί τη λογική των Multiplex Networks, επιλέγονται φωρατές και στάσεις με χωρική συσχέτιση που θα χρησιμοποιηθούν στο μοντέλο. Με αυτόν τον τρόπο λύνονται πολλά προβλήματα με κυριότερο το γεγονός ότι το μοντέλο θα έχει λιγότερα δεδομένα και θα τρέχει πιο γρήγορα, αφού αντί για δεδομένα από 200 φωρατές θα επιλεγούν δεδομένα μόνο από 5.

Ως φωρατής - στόχος (target) πρόβλεψης επιλέχθηκε ο MS219 ο οποίος βρίσκεται επί της οδού Βασιλέως Κωνσταντίνου στον κύριο δρόμο με κατεύθυνση προς Χίλτον (Β.Α.), 30μ. μετά την οδό Ριζάρη. Το συγκεκριμένο οδικό τμήμα είναι ένα από τα πολυσύχναστα και βρίσκεται στο κέντρο της πόλης Αθήνας.

Έπειτα με βάση την υψηλότερη συσχέτιση ΜΙ επιλέχθηκαν δύο σταθμοί μετρό, το Σύνταγμα και ο Ευαγγελισμός και 3 φωρατές ο MS230, ο MS239, ο MS243. Όσο αφορά τους φωρατές:

- ο MS230: βρίσκεται επί της οδού Καλλιρρόης στην κύρια οδό με κατεύθυνση προς το κέντρο πριν από τη Λεωφόρο Βουλιαγμένης.
- ο MS239: βρίσκεται επί της οδού Βασιλέως Κωνσταντίνου στον κύριο δρόμο με κατεύθυνση προς την οδό Βασιλίσσης Σοφίας πριν από την οδό Βασιλέως Γεωργίου.
- ο MS243: βρίσκεται επί της οδού Βασιλέως Κωνσταντίνου στον κύριο δρόμο με κατεύθυνση προς τη Βασιλίσσης Σοφίας πριν από τη Βασιλίσσης Σοφίας.

Οι ακριβείς θέσεις τους φαίνονται στην παρακάτω Εικόνα 7:



Εικόνα 7: Θέσεις φωρατών και στάσεων που χρησιμοποιούνται στο μοντέλο πρόβλεψης.

4.3.2 Προετοιμασία Δεδομένων για το δίκτυο LSTM

Για να χρησιμοποιηθούν τα δεδομένα για την ανάπτυξη του μοντέλου νευρωνικού δικτύου LSTM πραγματοποιήθηκε μία διαδικασία προετοιμασίας. Αρχικά, έγινε τυχαίος διαχωρισμός δειγμάτων με ποσοστό 85% για την εκπαίδευση του μοντέλου (train) και 15% για τις δοκιμές και την αξιολόγησή του (test). Από τον διαχωρισμό προέκυψαν τα μητρώα `training_set`, `test_set` για τα δεδομένα των χρονοσειρών.

```
split_ratio=0.15
split=int(len(data)*(1-split_ratio))
training_set=data[:split]
test_set=data[split:]
```

Στη συνέχεια, έγινε ομαλοποίηση χαρακτηριστικών που αλλιώς συνηθίζεται να λέγεται Feature Engineering. Επισημαίνεται πως η κλιμάκωση χαρακτηριστικών πραγματοποιείται λόγω της ευαισθησίας ορισμένων αλγορίθμων μηχανικής μάθησης στην κλίμακα των τιμών εισόδου. Αυτή η τεχνική κλιμάκωσης χαρακτηριστικών αναφέρεται μερικές φορές ως ομαλοποίηση χαρακτηριστικών. Οι διαδικασίες κλιμάκωσης που χρησιμοποιούνται συνήθως περιλαμβάνουν:

- Min-Max Scaling (Κλιμάκωση ελάχιστου μεγέθους): Αυτή η διαδικασία περιλαμβάνει την επανατοποθέτηση όλων των τιμών σε ένα χαρακτηριστικό της περιοχής 0 έως 1. Με άλλα λόγια, η ελάχιστη τιμή στο αρχικό εύρος θα λάβει την τιμή 0, η μέγιστη τιμή θα πάρει 1 και οι υπόλοιπες τιμές μεταξύ των δύο άκρων θα κλιμακωθούν κατάλληλα.
- Standardization/Variance scaling (Κλιμάκωση τυποποίησης/διακύμανσης): Όλα τα σημεία δεδομένων αφαιρούνται με τη μέση τιμή τους και το αποτέλεσμα διαιρείται με τη διακύμανση της κατανομής για να φτάσει σε μια κατανομή με 0 μέσο όρο και διακύμανση 1.

Ένας κίνδυνος που κρύβεται κατά την κλιμάκωση αραιών δεδομένων χρησιμοποιώντας τις δύο παραπάνω τεχνικές, είναι ότι σε ορισμένες περιπτώσεις μπορεί να οδηγήσει σε επιπλέον υπολογιστικό φόρτο [32].

Στη συγκεκριμένη περίπτωση επιλέχθηκε η κλιμάκωση ελάχιστου μεγέθους (min/max scaling).

```
scaler = MinMaxScaler(feature_range=(0, 1))
```

Έπειτα επιλέχθηκε ως φωρατής – στόχος (target) πρόβλεψης ο MS219 για τον οποίο διαμορφώθηκαν τα `x_train`, `y_train`, `x_test`, `y_test`.

Επειδή ζητείται η πρόβλεψη των κυκλοφοριακών συνθηκών στο φωρατή-στόχο μια ώρα μετά τη χρονική στιγμή στην οποία αντιστοιχούν τα δεδομένα, ως χρονικός ορίζοντας πρόβλεψης επιλέχθηκε η 1 ώρα μπροστά και για την πρόβλεψη λαμβάνονται υπόψη οι 6 προηγούμενες παρατηρήσεις. Για την πρόβλεψη αξιοποιήθηκαν 2 στάσεις μετρό και 3 φωρατές που σχετίζονται χωρικά με τον προβλεπόμενο.

4.3.3 Αρχιτεκτονική μοντέλου LSTM

Σε αυτή την υποενότητα περιγράφεται η αρχιτεκτονική του μοντέλου, η δυσκολότερη ίσως παράμετρος που πρέπει να αποφασιστεί για την υλοποίηση ενός Νευρωνικού Δικτύου, δηλαδή ο αριθμός και ο τύπος των επιπέδων που θα χρησιμοποιηθούν και ο τρόπος με τον οποίο αυτά συνδέονται μεταξύ τους.

Σχετικά με το πλήθος και τον τύπο των επιπέδων που χρησιμοποιούνται σε μία αρχιτεκτονική ενός Νευρωνικού Δικτύου, δεν υπάρχουν συγκεκριμένοι κανόνες. Κατά βάση μία κατάλληλη αρχιτεκτονική μπορεί να προκύψει λαμβάνοντας υπόψη το πλήθος και τον τύπο των δεδομένων και παρακολουθώντας την απόδοση του μοντέλου στο πέρασμα του χρόνου. Αρκετές φορές η παρατήρηση *overfitting* ή *underfitting* αποτελεί μία ένδειξη για την αφαίρεση ή την προσθήκη επιπλέον επιπέδων ή/και τη χρήση περισσότερων δειγμάτων στο μοντέλο.

Για την πρόβλεψη χρονοσειράς αποφασίστηκε να υλοποιηθεί ένα Μακράς και Βραχείας Μνήμης Νευρωνικό Δίκτυο, λόγω της καλής απόδοσης των LSTM σε προβλήματα όπου τα δεδομένα είναι χρονοσειρές. Η αρχιτεκτονική που επιλέχθηκε ως βέλτιστη μετά από διερεύνηση, επεξηγείται παρακάτω.

Συνάρτηση Ενεργοποίησης (Activation Function)

Οι συναρτήσεις ενεργοποίησης είναι ιδιαίτερα σημαντικές υπερπαραμέτροι για το σχεδιασμό ενός νευρωνικού δικτύου. Ο στόχος τους είναι ο μετασχηματισμός της εισόδου ενός επιπέδου σε μία έξοδο, τέτοια ώστε το επόμενο επίπεδο να μπορεί να τη δεχτεί ως είσοδο. Υπάρχουν διάφορα είδη συναρτήσεων ενεργοποίησης, με το κάθε ένα από αυτά να ενδείκνυται για κάποιον συγκεκριμένο τύπο προβλήματος.

Στο παρόν μοντέλο χρησιμοποιήθηκε για την βελτίωσή της εκπαίδευσης και της απόδοσης του μοντέλου η συνάρτηση ενεργοποίησης διορθωμένης γραμμικής μονάδας ReLU. Το πλεονέκτημά της βρίσκεται στο γεγονός πως όταν η είσοδος της είναι αρνητική, μετατρέπεται σε 0 και έτσι ο αντίστοιχος νευρώνας δεν ενεργοποιείται. Συνεπώς, οι υπολογισμοί γίνονται γρηγορότερα και πιο αποτελεσματικά.

Αλγόριθμος Βελτιστοποίησης (Optimizer)

Σε αυτό το μοντέλο χρησιμοποιήθηκε ο αλγόριθμος βελτιστοποίησης Adam, ο οποίος είναι μία στοχαστική μέθοδος (παραλλαγή του κλασικού Stochastic Gradient Descent) που βασίζεται στην προσαρμοστική εκτίμηση, με ποσοστό μάθησης ίσο με 0.001. Η επιλογή του έγινε διότι είχε καλύτερα αποτελέσματα από όλες τις εναλλακτικές. Σε γενικές γραμμές, ο Adam είναι ένας optimizer που προτιμάται σε διάφορα είδη προβλημάτων Μηχανικής Μάθησης, καθώς ελαχιστοποιεί γρηγορότερα το cost function, γι' αυτό και τα τελευταία χρόνια χρησιμοποιείται κατά κόρον.

Batch_size

Στη Μηχανική Μάθηση ο όρος batch_size χρησιμοποιείται για να περιγράψει τον συνολικό αριθμό των δειγμάτων εκπαίδευσης που χρησιμοποιούνται σε μία επανάληψη. Για παράδειγμα, αν η τιμή του batch_size είναι 100, αυτό σημαίνει ότι ο αλγόριθμος μάθησης θα χρησιμοποιήσει τα πρώτα 100 από τα δείγματα εκπαίδευσης για να εκπαιδεύσει το μοντέλο, θα συνεχίσει με τα επόμενα 100, κ.ο.κ. Στην περίπτωση που περισσέψουν δείγματα λιγότερα από τον ορισμένο αριθμό του batch_size, θα χρησιμοποιηθούν μόνο αυτά για το τελικό βήμα εκπαίδευσης του δικτύου. Οι τιμές του batch_size συνήθως κυμαίνονται από 1 ως 256, ωστόσο συχνά χρησιμοποιούνται και μεγαλύτερες τιμές, ανάλογα με τον όγκο των δεδομένων και τη μνήμη του υπολογιστή στον οποίο εκτελείται η εκπαίδευση. Συνήθως προτείνεται ο αριθμός του batch_size να είναι μικρότερος από τον συνολικό αριθμό των δειγμάτων εκπαίδευσης, καθώς αυτό επιταχύνει την εκπαίδευση ωστόσο υπάρχει ένα μειονέκτημα καθώς όσο μικρότερη είναι η τιμή του batch_size, αυξάνεται η πιθανότητα ο υπολογισμός των gradients να είναι ανακριβής λόγω της έντονης διακύμανσης. Επομένως, είναι προτιμότερο να χρησιμοποιείται η μεγαλύτερη δυνατή τιμή batch_size για την οποία τα δεδομένα χωράνε στη μνήμη.

Στο συγκεκριμένο μοντέλο το batch_size είναι 32.

Εποχές Εκπαίδευσης (Epochs)

Ο όρος εποχή εκπαίδευσης σε ένα πρόβλημα Μηχανικής Μάθησης υποδηλώνει τον αριθμό των περασμάτων ολόκληρου του συνόλου δεδομένων εκπαίδευσης που έχει ολοκληρώσει ο αλγόριθμος μηχανικής μάθησης. Δεν υπάρχει κανόνας που να ορίζει το ποιος είναι ο σωστός αριθμός εποχών για ένα πρόβλημα Μηχανικής Μάθησης, καθώς οι εποχές είναι μία υπερπαραμέτρος (hyperparameter) του μοντέλου που εξαρτάται από ένα σύνολο παραγόντων, όπως π.χ. ο αριθμός των κρυφών επιπέδων του μοντέλου, το σύνολο των δειγμάτων εκπαίδευσης κ.ά. Όταν ο αριθμός των εποχών είναι πολύ μικρός είναι πολύ πιθανό το μοντέλο να μην προλάβει να εξάγει πολλά χαρακτηριστικά από τα δεδομένα, ενώ η αντίθετη περίπτωση συνήθως οδηγεί στην εξαγωγή ασήμαντων για το πρόβλημα χαρακτηριστικών.

Γενικά προτείνεται ο ορισμός αρχικά ενός μεγάλου αριθμού εποχών εκπαίδευσης, και ο περιορισμός μετά από παρακολούθηση του πειράματος στον αριθμό των εποχών για τις οποίες το μοντέλο δεν παρουσιάζει overfitting. Για τη διευκόλυνση αυτής της διαδικασίας, το Keras παρέχει ένα callback που ονομάζεται EarlyStopping και στο οποίο ο προγραμματιστής έχει την ευκαιρία να ορίσει τον μέγιστο επιτρεπτό αριθμό εποχών μη βελτίωσης της ακρίβειας ή του κόστους. Σε περίπτωση που ο αριθμός αυτός ξεπεραστεί, η εκπαίδευση σταματά.

Το παρόν μοντέλο εκπαιδεύτηκε για 180 εποχές.

Συνάρτηση Κόστους (Cost Function)

Η επιλογή του cost function κατά βάση εξαρτάται από τον εκάστοτε τύπο του προβλήματος. Το Keras παρέχει ένα πλήθος διαφορετικών cost functions, η κάθε μία από τις οποίες εφαρμόζεται καλύτερα σε συγκεκριμένο τύπο προβλήματος (mean_squared_error, categorical_crossentropy, binary_crossentropy, cosine_proximity κ.ά.). Η συνάρτηση κόστους είναι μια τεχνική αξιολόγησης «της απόδοσης του αλγορίθμου/μοντέλου». Λαμβάνει τόσο τις προβλεπόμενες εκροές από το μοντέλο όσο και τις πραγματικές εκροές και υπολογίζει πόσο λάθος ήταν το μοντέλο στην πρόβλεψή του. Βγάζει μεγαλύτερο αριθμό εάν οι προβλέψεις διαφέρουν πολύ από τις πραγματικές τιμές.

Στο συγκεκριμένο μοντέλο πρόβλεψης χρησιμοποιήθηκε η συνάρτηση κόστους mean_absolute_error.

Μετρικές

Υπάρχουν διάφοροι τύποι μετρικών που χρησιμοποιούνται για την αξιολόγηση των αποτελεσμάτων ενός προβλήματος. Η mean_absolute_percentage_error είναι αυτή που χρησιμοποιήθηκε στο μοντέλο.

Συνοψίζοντας για την εκπαίδευση του μοντέλου πρόβλεψης οι υπερπαραμέτροι που χρησιμοποιήθηκαν είναι οι εξής:

- Για το επίπεδο του LSTM, 64 νευρώνες ως μέγεθος των Νευρωνικών Δικτύων που υλοποιούνται εντός του κελιού του LSTM.
- Η συνάρτηση ενεργοποίησης διορθωμένης γραμμικής μονάδας (ReLU), η οποία χρησιμοποιήθηκε για την βελτίωσή της εκπαίδευσης και της απόδοσης του μοντέλου.
- Ο αλγόριθμος βελτιστοποίησης Adam με ποσοστό μάθησης ίσο με 0.001.
- Το πλήθος δειγμάτων (batch size) που εκπαιδεύονται μαζί σε κάθε εποχή ορίστηκε σε 32, που αποτελεί τη βέλτιστη επιλογή που προέκυψε μετά από αναζήτηση.

- Οι εποχές εκπαίδευσης (epochs) ορίστηκαν σε 180 επαναλήψεις για τη μεταφορά ολόκληρου του συνόλου των δεδομένων μέσω του νευρωνικού δικτύου μόνο μία φορά. Με την ολοκλήρωση κάθε επανάληψης (epoch) τυπώθηκαν η τιμή απώλειας των δεδομένων εκπαίδευσης, καθώς και ο υπολογιστικός χρόνος ολοκληρωμένης επανάληψης και κάθε βήματος ξεχωριστά.
- Ως συνάρτηση κόστους (loss function) η MAE.
- Για την αξιολόγηση η μετρική MAPE.

Αυτές είναι οι διευκρινίσεις σχετικά με τις επιλογές των υπερπαραμέτρων, ακολουθεί ο κώδικας για την υλοποίηση του δικτύου.

```
import tensorflow as tf
from keras import Sequential
from keras.layers import Dense, Dropout, LSTM, Activation
from keras.optimizers import Adam

adam = Adam(learning_rate=0.001)
# Create LSTM model
model = Sequential()
model.add(LSTM(64, input_shape=(x_train.shape[1],5), return_sequences=False))
model.add(Dense(3, activation='relu'))
model.add(Dense(1))
model.compile(loss='mean_absolute_error', optimizer=adam)
history=model.fit(x_train[:, :, [3,13,21,26,28]], y_train, epochs=180, batch_size=32)
```

4.4 Αποτελέσματα LSTM μοντέλου

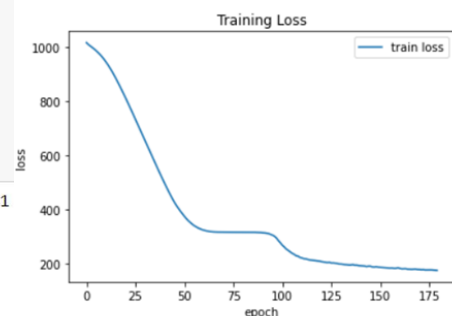
Στην ενότητα αυτή παρουσιάζονται τα αποτελέσματα που προέκυψαν από την πρόβλεψη των κυκλοφοριακών συνθηκών με την υλοποίηση του μοντέλου LSTM. Το μέσο απόλυτο ποσοστιαίο σφάλμα είναι 14,55% που αποτελεί μια ικανοποιητική τιμή, ανάλογη των εργασιών της βιβλιογραφίας που πιστοποιεί την επάρκεια της συγκεκριμένης τεχνικής στην πρόβλεψη χρονοσειρών.

```
model.evaluate(x_test[:, :, [3,13,21,26,28]], y_test)
pre = model.predict(x_test[:, :, [3,13,21,26,28]])
'''for i in range(len(y_test)):
    print(pre[i], y_test[i)'''

error_df = pd.DataFrame({'real': y_test, 'predicted': pre[:, 0]})
error_df['mae'] = abs(error_df['real'] - error_df['predicted'])
error_df['mape'] = error_df['mae'] / error_df['real']
print(error_df['mape'].mean() * 100, '%')
d_error = error_df.describe()

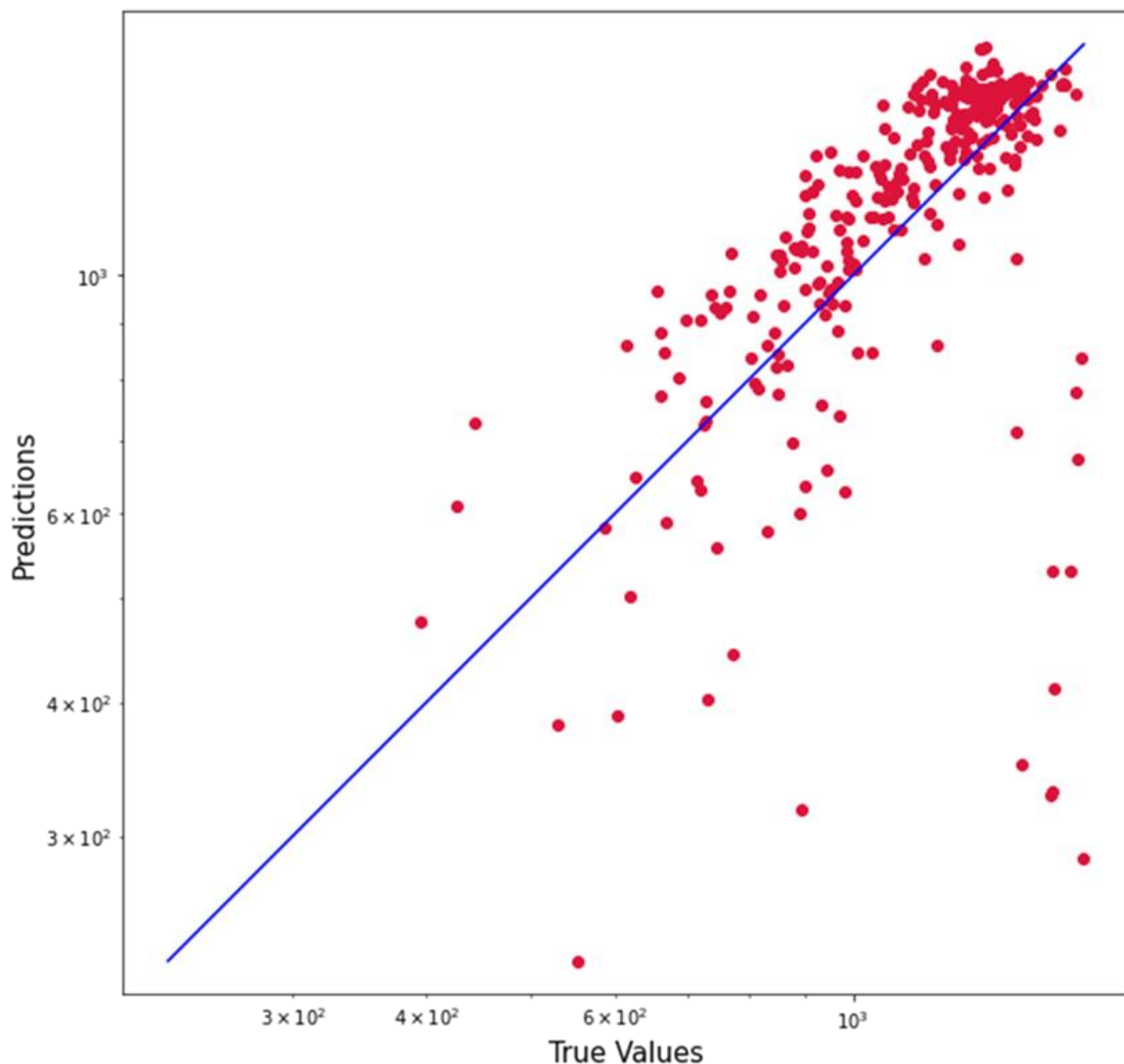
10/10 [=====] - 1s 3ms/step - loss: 159.4741
159.4740753173828

'for i in range(len(y_test)):
    print(pre[i], y_test[i])'
14.557206948348437 %
```



Επισημαίνεται πως ο όγκος δεδομένων είναι περιορισμένος επειδή χρησιμοποιήθηκαν ακατέργαστα πραγματικά δεδομένα από μια καινούργια ελληνική δημόσια βιβλιοθήκη δεδομένων σε συνδυασμό με το χρόνο εκπόνησης της διπλωματικής εργασίας. Ως αποτέλεσμα η απόδοση του μοντέλου περιορίζεται σημαντικά. Πιο συγκεκριμένα η επιπλέον επεξεργασία των διαθέσιμων δεδομένων ώστε να χρησιμοποιηθούν μόνο όσα αφορούν την περιοχή εντός του δακτυλίου Αθηνών μείωσε περαιτέρω τον όγκο τους.

Ωστόσο, με τη διαδικασία επιλογής παραμέτρων που έχουν χωροχρονική συσχέτιση στην ανάπτυξη του δικτύου LSTM, μειώθηκε η διαστατικότητα του προβλήματος βοηθώντας την καλή λειτουργία του αλγορίθμου και την αποδοτικότητα του υπολογιστικού συστήματος.

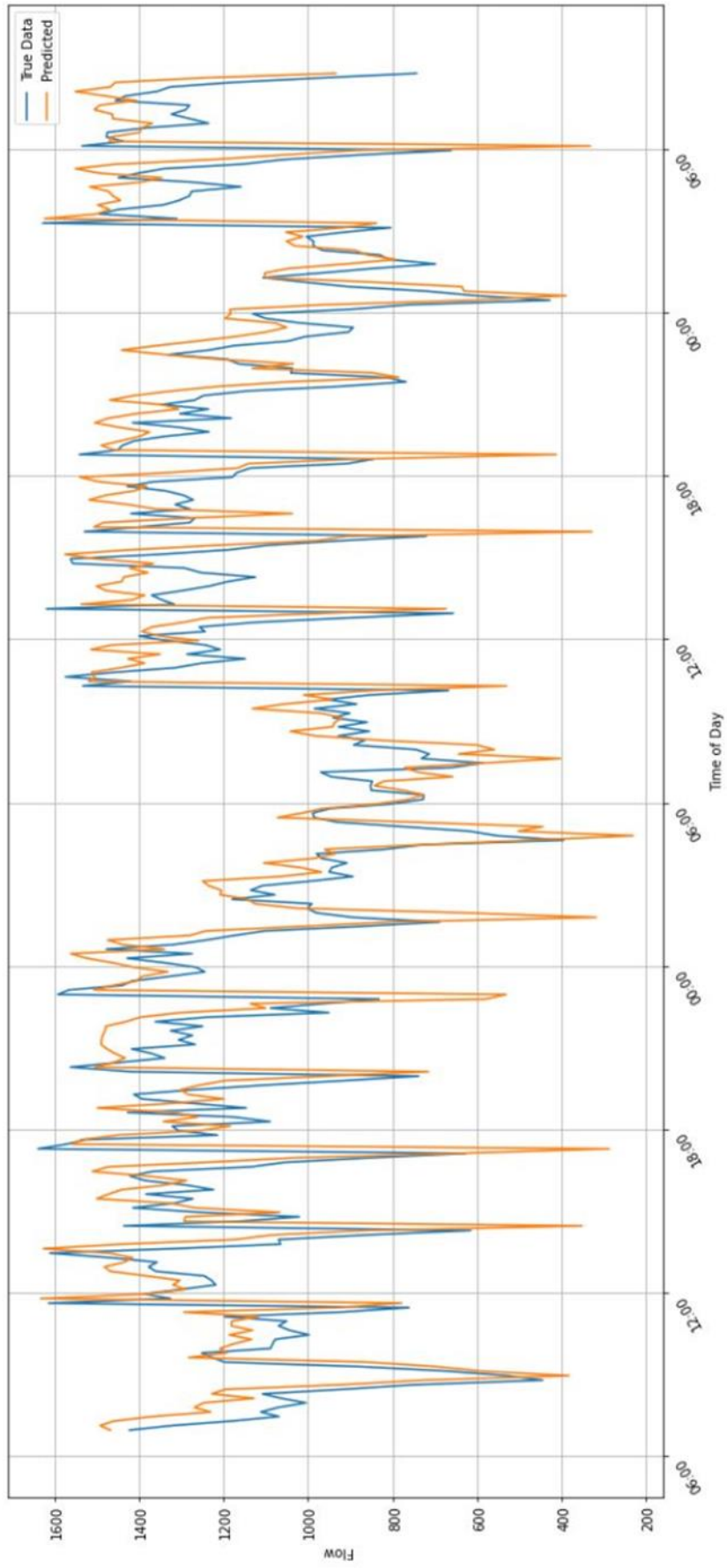


Διάγραμμα 12: Διάγραμμα Διασποράς Πραγματικών και Προβλεπόμενων Τιμών.

Στο Διάγραμμα 12 φαίνεται το διάγραμμα διασποράς, ή αλλιώς Scatter Plot, μεταξύ των πραγματικών και των προβλεπόμενων τιμών του κυκλοφοριακού φόρτου. Χρησιμοποιώντας το διάγραμμα διασποράς εξετάζεται η σχέση μεταξύ δύο μεταβλητών. Στο συγκεκριμένο διάγραμμα εξετάζεται η συσχέτιση μεταξύ των πραγματικών δεδομένων και των δεδομένων πρόβλεψης που προκύπτουν ως αποτέλεσμα του μοντέλου που αναπτύχθηκε. Το συμπέρασμα που προκύπτει είναι πως υπάρχει υψηλή γραμμική συσχέτιση μεταξύ των δύο μεταβλητών (πραγματικών και προβλεπόμενων τιμών), και αυτό φαίνεται από τη μορφή κατανομής των δεδομένων σε σχέση με τη γραμμική τάση.

Οι τιμές που εμφανίζονται σε ακραίες θέσεις με σημείο αναφοράς τη γραμμική τάση αποδίδονται είτε στο ποσοστό σφάλματος του αναπτυγμένου μοντέλου είτε σε τιμές των δεδομένων που διαφέρουν κατά πολύ από τις συνήθεις καταγεγραμμένες τιμές.

Παρακάτω ακολουθεί το Διάγραμμα 13 το οποίο απεικονίζει τη χρονοσειρά των πραγματικών τιμών και των προβλέψεων. Απο το διάγραμμα φαίνεται πως το μοντέλο LSTM ανταποκρίνεται ικανοποιητικά με βάση τα πραγματικά δεδομένα. Με αρκετά μεγάλη ακρίβεια ακολουθεί τη ροή τιμών κατά τη διάρκεια της ημέρας. Εδώ αξίζει να σημειωθούν δύο παρατηρήσεις. Η πρώτη αφορά τις απότομες πτώσεις τιμών οι οποίες προκύπτουν εξαιτίας της επιλογής τα δεδομένα να ξεκινούν από τις 7:00. Η δεύτερη αφορά τις ακραίες τιμές των δεδομένων. Πιο συγκεκριμένα μπορεί κανείς να παρατηρήσει πως για τιμές φόρτου που κυμαίνονται μεταξύ 500 και 1400 το μοντέλο ανταποκρίνεται πολύ καλά. Υστερεί ωστόσο να προβλέψει τιμές πέρα από αυτά τα όρια. Η αντίδραση αυτή είναι φυσιολογική αφού τιμές πέρα των ορίων που αναφέρθηκαν παρατηρούνται σε μεμονωμένες μόνο περιπτώσεις.



Διάγραμμα 13: Διάγραμμα Χρονοσειρές Πραγματικών και Προβλεπόμενων.

Κεφάλαιο 5

Συμπεράσματα και Προτάσεις

5.1 Γενικά Συμπεράσματα

Πράγματι όλο και περισσότεροι οδηγοί βασίζονται σε εφαρμογές και συσκευές πλοήγησης για να τους καθοδηγήσουν σε ταχύτερες διαδρομές και να αποφύγουν τη συμφόρηση. Δεν γίνεται πάντα να αποφεύγεται η κίνηση, αλλά με ακριβείς προβλέψεις για την κυκλοφορία μπορούν να γίνονται πιο έξυπνες επιλογές από τους χρήστες ώστε να μπορούν να εξοικονομήσουν χρήματα, χρόνο και να αυξήσουν τη συνολική οδική ασφάλεια.

Η βραχυπρόθεσμη πρόβλεψη κυκλοφοριακών μεγεθών σε αστικές σηματοδοτούμενες αρτηρίες αποτελεί πολύπλοκη διαδικασία που δεν επηρεάζεται μόνο από τη σηματοδότηση αλλά και από άλλους μη μετρήσιμους παράγοντες (μη αναμενόμενα συμβάντα, βραχυπρόθεσμη επιρροή ροής από μη ελεγχόμενες προσβάσεις, παράνομη στάθμευση, έργα κατά μήκος της αρτηρίας, συμπεριφορά οδηγών, κακές καιρικές συνθήκες κ.α.). Η επιρροή των παραπάνω παραγόντων αντανακλάται στην έντονη μεταβλητότητα, στις ασυνέχειες και στην μεταβατική συμπεριφορά των χρονοσειρών των κυκλοφοριακών μεγεθών που συλλέγονται σε πραγματικό χρόνο σε σημείο αστικής αρτηρίας. Τα παραπάνω χαρακτηριστικά των χρονοσειρών των κυκλοφοριακών μεγεθών επηρεάζουν καθοριστικά την ακρίβεια της βραχυπρόθεσμης πρόβλεψης κυρίως κοντά σε συνθήκες κορεσμού αλλά και σε πιθανές μεταβάσεις της κυκλοφορίας.

Οι μετρικές γενικής χρήσης, όπως η συσχέτιση κατά Pearson, καθώς και οι πιο εξειδικευμένες για χρονικές σειρές, όπως η αμοιβαία πληροφορία, παρέχουν μια σαφή εικόνα για τις χωροχρονικές σχέσεις των οδικών τμημάτων ενός αστικού οδικού δικτύου. Αυτές οι σχέσεις προκύπτουν από τη σχετική θέση των οδικών τμημάτων και τη ροή κυκλοφορίας που το κάθε τμήμα εξυπηρετεί. Ως εκ τούτου, παρέχουν εξηγήσιμα αποτελέσματα.

Είναι σαφές ότι η μηχανική μάθηση έχει μεγάλες δυνατότητες όσον αφορά την πρόβλεψη χρονοσειρών. Αυτό έχει αποδειχθεί σε αυτή την εργασία καθώς και σε άλλες βιβλιογραφικές αναφορές. Στη βιβλιογραφία φαίνεται ότι οι προσεγγίσεις με βάση τα δεδομένα αποτελούν αξιόπιστη εναλλακτική στη βραχυπρόθεσμη πρόβλεψη σε σχέση με τα συνήθη πρότυπα εκτίμησης κυκλοφορίας. Επιπλέον, η καλύτερη επεξηγηματικότητα της χωροχρονικής εξέλιξης των κυκλοφοριακών μεγεθών είναι πολύ σημαντική για την ανάπτυξη ενός μοντέλου πρόβλεψης που θα έχει υψηλή ακρίβεια.

5.2 Βασικά Συμπεράσματα

Η πρόβλεψη της κυκλοφοριακής ροής κατέχει μια εξέχουσα θέση στον κλάδο των μεταφορών με την άνοδο της βαθιάς μάθησης. Αυτό συμβαίνει ουσιαστικά λόγω του γεγονότος ότι τα μοντέλα μηχανικής μάθησης έχουν ξεπεράσει και αντικαταστήσει τα προηγούμενα στατιστικά μοντέλα, τόσο όσον αφορά την απόδοση όσο και την ακρίβεια.

Όπως αναφέρθηκε και προηγουμένως, η πρόβλεψη ροής κυκλοφορίας σχετίζεται με πολλές μεταβλητές κίνησης και ο τρόπος επιλογής κατάλληλου συνδυασμού μεταβλητών κίνησης είναι πολύ σημαντικός καθώς μπορεί να μειώσει το κόστος υπολογισμού και να βελτιώσει την ακρίβεια πρόβλεψης. Συνεπώς, η ακρίβεια των προβλέψεων εξαρτάται σε μεγάλο βαθμό από την ανάλυση των χωροχρονικών συσχετίσεων ενός οδικού δικτύου πριν από την ανάπτυξη προτύπων πρόβλεψης. Στην παρούσα εργασία προτείνεται ως τεχνική εύρεσης συσχετίσεων και επιλογής χαρακτηριστικών η συσχέτιση κατά Pearson και η Αμοιβαία Πληροφορία.

Τα αποτελέσματα έδειξαν πως η Αμοιβαία Πληροφορία είναι η πιο κατάλληλη μετρική για τη συγκεκριμένη εφαρμογή επειδή εστιάζει στην ύπαρξη γραμμικών και μη γραμμικών συσχετίσεων. Οι συσχετίσεις προέρχονται από τις σχετικές θέσεις των τμημάτων και του φόρτου που αυτά εξυπηρετούν, συνεπώς τα αποτελέσματα που προκύπτουν είναι λογικά και εξηγήσιμα.

Τα αποτελέσματα, όπως παρουσιάστηκαν και στην προηγούμενη ενότητα υποδεικνύουν ότι η προτεινόμενη προσέγγιση είναι κατάλληλη για το πρόβλημα που καλείται να λύσει η διπλωματική εργασία. Η ακρίβεια του μοντέλου είναι συγκρίσιμη με αυτή άλλων προσεγγίσεων της βιβλιογραφίας. Ωστόσο η χρήση διαφορετικών μετρικών στον εντοπισμό των εξαρτήσεων σε ένα οδικό τμήμα οδηγεί σε διαφορετικά συμπεράσματα, τα οποία αφού πρώτα αξιολογηθούν πρέπει στη συνέχεια να εξετάζεται αν βελτιώνουν την πρόβλεψη των κυκλοφοριακών συνθηκών.

5.3 Προτάσεις για περαιτέρω έρευνα

Βάσει της βιβλιογραφικής ανασκόπησης και της παρούσας διπλωματικής εργασίας έγινε κατανοητό ότι υπάρχουν περιθώρια εξέλιξης της έρευνας στη βραχυπρόθεσμη πρόβλεψη των κυκλοφοριακών συνθηκών.

Ο επόμενος στόχος είναι η αναζήτηση περισσότερο περίπλοκων αρχιτεκτονικών, ενδεχομένως αξιοποιώντας και συνελκτικά νευρωνικά δίκτυα, τα οποία συνδυάζονται αρκετά συχνά με δίκτυα LSTM στην βιβλιογραφία. Επίσης για τα σχεδιασμό της πρόβλεψης κυκλοφοριακής ροής πρέπει να λαμβάνονται υπόψη χωρικά και χρονικά χαρακτηριστικά τα οποία μπορούν να αποτυπώσουν περισσότερα χαρακτηριστικά των κυκλοφοριακών δεδομένων, για αυτό το λόγο μελλοντικά

προτείνεται ο σχεδιασμός υβριδικών δομών βαθιάς μηχανικής μάθησης που θα είναι καλύτερες στην καταγραφή χωροχρονικών πληροφοριών κίνησης.

Επιπλέον, η βελτιστοποίηση των παραμέτρων ενός δικτύου βαθιάς μάθησης είναι επίσης μια άλλη πρόκληση που μπορεί να διερευνηθεί περαιτέρω στην πρόβλεψη ροής κίνησης. Η μελλοντική έρευνα θα περιλαμβάνει σίγουρα επεξεργασία δεδομένων από περισσότερες ημέρες, προκειμένου να υπάρξουν πιο γενικευμένα συμπεράσματα.

Ακόμη προτείνεται η διερεύνηση της σχέσης μεταξύ χωρικών και χρονικών μορφών κυκλοφορίας από περισσότερα από ένα σημεία αστικής αρτηρίας, του τρόπου που επηρεάζουν τη διαμόρφωση των περιοχών κυκλοφοριακής ροής αλλά και του τρόπου που επηρεάζουν τη βελτίωση της βραχυπρόθεσμης πρόβλεψης κυκλοφορίας (ακριβέστερες προβλέψεις, πιο αξιόπιστες προβλέψεις για πολλά βήματα μπροστά κ.α.).

Τέλος, όσο αφορά το μοντέλο ο που αναπτύχθηκε στην παρούσα διπλωματική εργασία θα μπορούσε να βελτιωθεί και να επεκταθεί ως προς τις πιο κάτω κατευθύνσεις:

- Σχεδιασμός υβριδικού μοντέλου βαθιάς μάθησης.
- Αξιοποίηση μεγαλύτερου όγκου δεδομένων.
- Ενσωμάτωση πιο εξειδικευμένων και σύγχρονων τεχνικών αναγνώρισης χωροχρονικών συσχετίσεων.
- Πειραματισμός με άλλα μοντέλα μηχανικής μάθησης πέρα από το LSTM.
- Ανάπτυξη πιο βραχυπρόθεσμων προβλέψεων έως και προβλέψεις σε πραγματικό χρόνο.

Βιβλιογραφία

- [1] Cookson G.;Pishue B., “Global Traffic Scorecard,” INRIX Research, 2017.
- [2] European Environment Information and Observation Network (Eionet), March 2020. [Online].Available: <https://www.eea.europa.eu/el/pressroom/newsreleases/themes/transport/intro>.
- [3] EEA Report No 22/2019, [Online]. Available: <https://www.eea.europa.eu/publications/environmental-noise-in-europe>.
- [4] D. L. Dwight A.H., “The relationship between traffic congestion, driver stress and direct versus indirect coping behaviours,” *Ergonomics*, vol. 40, no. 3, pp. 348-361, 1997.
- [5] Gilbert Richard, “Defining Sustainable Transportation,” The Centre for Sustainable Transportation, Canada, 2005.
- [6] E. Union, “Directive 2010/40/EU of the European Parliament and of the Council of 7 July 2010 on the framework for the deployment of Intelligent Transport Systems in the field of road transport and for interfaces with other modes of transport Text with EEA relevance,” [Online]. Available: <https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX%3A32010L0040>.
- [7] E. I. Vlahogianni, M. G. Karlaftis, and J. C. Golias, “Short-term traffic forecasting: Where we are and where we’re going,” *Transp. Res. Part C Emerg. Technol.*, 2014.
- [8] Meng, M., Shao, Cf., Wong, Yd. et al., “A two-stage short-term traffic flow prediction method based on AVL and AKNN techniques.,” *J. Cent. South Univ.*, no. 22, p. 779–786, 2015.
- [9] J. Barros, M. Araujo, R. Rossetti, “Short-term real-time traffic prediction methods: A survey.,” *International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS)*, 2015.
- [10] T. Alghamdi, K. Elgazzar, M. Bayoumi, T. Sharaf, and S. Shah, “Forecasting Traffic Congestion Using ARIMA Modeling,” *15th Int. Wirel. Commun. Mob. Comput. Conf.*, pp. 1227-1232, 2019.

- [11] Y. Xie, Y. Zhang, and Z. Ye, "Short-Term Traffic Volume Forecasting Using Kalman Filter with Discrete Wavelet Decomposition.," vol. 22, pp. 326-334, 2007.
- [12] I. Lana, J. Del Ser, M. Velez, and E. I. Vlahogianni, "Road Traffic Forecasting: Recent Advances and New Challenges," *IEEE Intell. Transp. Syst. Mag.*, vol. 10, no. 2, pp. 3-19, 2018.
- [13] C.P.IJ. van Hinsbergen, C., van Lint, J., & Sanders, F., "Short term traffic prediction models," in Proc. 14th World Congress Intelligent Transportation Systems,, Beijing, 2007.
- [14] Pinlong Cai, Yunpeng Wang, Guangquan Lu, Peng Chen, Chuan Ding, Jianping Sun, "A spatiotemporal correlative k-nearest neighbor model for short-term traffic multistep forecasting," *Transportation Research Part C*, no. 62, pp. 21-34, 2016.
- [15] Brian L. Smith, Billy M. Williams, R. Keith Oswald, "Comparison of parametric and nonparametric models," pp. 303-321, 2002.
- [16] M. G. J. C. Eleni I. Vlahogianni, "Optimized and meta-optimized neural networks for short-term traffic flow prediction: A genetic approach," *Transportation Research Part C: Emerging Technologies*, vol. 13, no. 3, pp. 211-234, 2005.
- [17] L. Qu, W. Li, W. Li, D. Ma, and Y. Wang,, "Traffic flow prediction with big data: A deep learning approach.," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, pp. 865-873, 2015.
- [18] Q. Zhang, Q. Jin, J. Chang, S. Xiang, and C. Pan, "Kernel-Weighted Graph Convolutional Network: A Deep Learning Approach for Traffic Forecasting," *Proceedings - International Conference on Pattern Recognition*, pp. 1018-1023, 2018.
- [19] Zhao, Z., Chen, W., Wu, X., Chen, P. C., & Liu, J., "LSTM network: A deep learning approach for short-term traffic forecast," *Intelligent Transport Systems*, vol. 11, pp. 68-75, 2017.
- [20] Loan N.N. Doa,, Hai L. Vu, Bao Q. Vo, Zhiyuan Liu, Dinh Phung, "An effective spatial-temporal attention based neural network for traffic flow prediction," *Transportation Research Part C*, vol. 108, pp. 12-28, 2019.

- [21] A. Stathopoulos, M. Karlaftis, “A multivariate state space approach for urban traffic flow modeling and prediction,” *Transportation Research Part C: Emerging Technologies*, vol. 11, no. 2, pp. 121-135, 2003.
- [22] Unsok Ryu, Jian Wang, Thaeyong Kim, Sonil Kwak, Juhyok U., “Construction of traffic state vector using mutual information for short-term traffic flow prediction,” *Transportation Research Part C*, vol. 96, pp. 55-71, 2018.
- [23] Lawrence W. Lan, Jiuh-Biing Sheu, Yi-San Huang, “Investigation of temporal freeway traffic patterns in reconstructed state spaces,” *Transportation Research Part C*, pp. 116-136, 2008.
- [24] T.Huijskens, 2017. [Online]. Available: <https://thuijskens.github.io/2017/10/07/feature-selection/#fnref:4>.
- [25] C.Olah, “Understanding LSTM Networks,” 2015. [Online]. Available: <https://colah.github.io/posts/2015-08-Understanding-LSTMs/>.
- [26] A.W. Smith, D. Zipser, “Recurrent Neural Net Learning and Vanishing Gradient,” *International Journal of Neural Systems*, vol. 1, pp. 125-131, 1989.
- [27] S. Hochreiter, “Long Short-term Memory,” 1997.
- [28] Davydenko, A., & Fildes, R., “Forecast Error Measures: Critical Review and Practical Recommendations.,” in *In Business Forecasting: Practical Problems and Solutions.*, John Wiley & Sons Inc, 2016.
- [29] N. H. S. & G. A. Kouvaris, “Pattern formation in multiplex networks.,” *Scientific Reports*, vol. 5, 2015.
- [30] “Python Software Foundation python language reference, version 3.9.5.,” [Online]. Available: <https://www.python.org/>.
- [31] “<https://scikit-learn.org/>,” [Online].
- [32] “<https://www.kaggle.com/>,” [Online].

Παράρτημα Α – Κώδικας

```
#Εισαγωγή Βιβλιοθηκών
```

```
import requests
import pandas as pd
import json
import pandas as pd
import numpy as np
import matplotlib
import matplotlib.pyplot as plt
import seaborn as sns
import sklearn.feature_selection
```

```
#Ανάκτηση Δεδομένων από API
```

```
#TRAFFIC
```

```
url_traffic = 'https://data.gov.gr/api/v1/query/road_traffic_attica?date_from=2021-03-02&date_to=2021-06-30'
```

```
headers_traffic = {'Authorization': 'Token b15753babb3b98cac1bccd04c236e654743bb3bd'}
```

```
response_traffic = requests.get(url_traffic, headers=headers_traffic)
```

```
text_traffic = (response_traffic.text)
```

```
json_data_traffic = json.loads(text_traffic)
```

```
df_traffic = pd.read_json(text_traffic)
```

```
traffic = df_traffic.to_csv("Traffic.csv")
```

```
#OASA
```

```
url_oasa = 'https://data.gov.gr/api/v1/query/oasa_ridership?date_from=2021-03-02&date_to=2021-06-30'
```

```
headers_oasa = {'Authorization': 'Token b15753babb3b98cac1bccd04c236e553343bb3bd'}
```

```
response_oasa = requests.get(url_oasa, headers=headers_oasa)
```

```
text_oasa = (response_oasa.text)
```

```
json_data_oasa = json.loads(text_oasa)
```

```
df_oasa = pd.read_json(text_oasa)
```

```
oasa = df_oasa.to_csv("OASA.csv")
```

```
#Αρχική Επεξεργασία Βάσεων Δεδομένων
```

```
df_traffic = pd.read_csv("Traffic.csv", parse_dates=True)
```

```
df_traffic["appprocesstime"] = pd.to_datetime(df_traffic.appprocesstime)
```

```
new_traffic= df_traffic.sort_values(by='appprocesstime')
```

```
data_traffic = new_traffic.pivot(index="appprocesstime", columns="deviceid", values =  
"countedcars")
```

```
data_traffic=data_traffic.between_time('07:00:00' , "23:00:00")
```

```
data_traffic_new=data_traffic.fillna(method='bfill',axis=1)
```

```
traffic = data_traffic_new.filter(['MS207','MS219','MS224','MS225',
```

```
'MS228','MS230','MS231','MS233','MS234','MS238',  
'MS239','MS242','MS243','MS244','MS245','MS246','MS248','MS249','MS252','MS255',  
'MS256','MS258','MS261','MS310','MS317','MS342','MS350','MS356','MS364','MS365',  
'MS367','MS374','MS376','MS377','MS381','MS390','MS392','MS393','MS415','MS416',  
'MS418','MS426','MS427','MS432','MS434','MS435','MS440','MS446','MS447','MS474',  
'MS482','MS483','MS484','MS488','MS500','MS515','MS516','MS572','MS574','MS575',  
'MS581','MS631','MS634','MS636','MS638','MS639','MS658','MS669','MS670','MS672',  
'MS673','MS700','MS704','MS710','MS712','MS770','MS772','MS933'], axis=1)
```

```
df_oasa = pd.read_csv("OASA.csv", parse_dates=True)
```

```
df_oasa['date_hour'] =pd.to_datetime(df_oasa.date_hour)
```

```
new_oasa = df_oasa.sort_values(by='date_hour')
```

```
data_oasa = new_oasa.pivot_table(index="date_hour" , columns="dv_platenum_station", values  
= "dv_validations") , aggfunc='sum')
```

```
data_oasa=data_oasa.between_time('07:00:00' , "23:00:00")
```

```
data_oasa_new = data_oasa.fillna(method='bfill',axis=1)
```

```
oasa_data=data_oasa_new.filter(['ΑΚΡΟΠΟΛΗ','ΑΜΠΕΛΟΚΗΠΟΙ','ΒΙΚΤΩΡΙΑ',
```

```
'ΕΥΑΓΓΕΛΙΣΜΟΣ','ΘΗΣΕΙΟ','ΚΕΡΑΜΕΙΚΟΣ','ΜΕΤΑΞΟΥΡΓΕΙΟ',
```

```
'ΜΕΓΑΡΟ ΜΟΥΣΙΚΗΣ','ΜΟΝΑΣΤΗΡΑΚΙ','ΟΜΟΝΟΙΑ', 'ΠΑΝΟΡΜΟΥ',
```

```
'ΠΑΝΕΠΙΣΤΗΜΙΟ','ΠΙΕΤΡΑΛΩΝΑ','ΣΥΝΤΑΓΜΑ', 'ΣΥΓΓΡΟΥ ΦΙΞ','ΤΑΥΡΟΣ'], axis=1)
```

```
#Ενοποίηση Βάσεων Δεδομένων
```

```
frames = [oasa, traffic]
```

```
data = pd.concat(frames, axis=1)
```

```
data.to_csv("data_final.csv")
```

```
#Mutual Information
```

```
from sklearn.feature_selection import mutual_info_classif
```

```
devices = traffic_data.columns.tolist()
```

```
X = traffic_data.to_numpy()
```

```
mi_list_traffic = []
```

```
for i in devices:
```

```
    y = traffic_data[i]
```

```
    mi = mutual_info_classif(X,y)
```

```
    mi_list_traffic.append(mi)
```

```
mi_df_traffic = pd.DataFrame.from_records(mi_list_traffic)
```

```
mi_df_traffic_index = mi_df_traffic.set_axis(devices, axis="index")
```

```
mi_df_traffic_format = mi_df_traffic_index.set_axis(devices, axis="columns")
```

```
mi_df_traffic_format.head()
```

```
mi_df_traffic_format.to_csv("MI_traffic.csv")
```

```
stations=oasa.columns.tolist()
```

```
X = oasa.to_numpy()
```

```
mi_list_oasa = []
```

```
for i in stations:
```

```
    y = oasa[i]
```

```
    mi = mutual_info_classif(X,y)
```

```
    mi_list_oasa.append(mi)
```

```

mi_df_oasa = pd.DataFrame.from_records(mi_list_oasa)
mi_df_oasa_index = mi_df_oasa.set_axis(stations, axis="index")
mi_df_oasa_format = mi_df_oasa_index.set_axis(stations, axis="columns")
mi_df_oasa_format.head()
mi_df_oasa_format.to_csv('MI_oasa.csv')

```

```

d_all= data.columns.tolist()
X = data.to_numpy()
mi_list_all = []
for i in d_all:
    y = data[i]
    mi = mutual_info_classif(X,y)
    mi_list_all.append(mi)
mi_df_all = pd.DataFrame.from_records(mi_list_all)
mi_df_all_index = mi_df_all.set_axis(d_all, axis="index")
mi_df_all_format = mi_df_all_index.set_axis(d_all, axis="columns")
mi_df_all_format.head()
mi_df_all_format.to_csv("MI_all.csv")
#Θερμικοί Χάρτες

```

```

plt.figure(figsize=(20,30))
cor = mi_traffic_data.corr()
sns.heatmap(cor, annot=False, cmap=plt.cm.Red)
plt.show()

```

```

plt.figure(figsize=(20,30))
cor = mi_oasa_data.corr()
sns.heatmap(cor, annot=False, cmap=plt.cm.Red)
plt.show()

```

```

plt.figure(figsize=(20,30))
cor = mi_all_data.corr()
sns.heatmap(cor, annot=False, cmap=plt.cm.Red)
plt.show()

```

#Ο κώδικας παραμετροποίησης για την προετοιμασία των x_train, y_train, x_test, y_test του φωρατή – στόχου

#Βιβλιοθήκες

```

import tensorflow as tf
from keras import Sequential
from keras.layers import Dense, Dropout, LSTM, Activation
from keras.optimizers import Adam
from sklearn.preprocessing import MinMaxScaler

scaler = MinMaxScaler(feature_range=(0, 1))

```

```

t_tr = 'MS219' # choose device or station id
y_tr = training_set[t_tr]
x_tr = pd.DataFrame(data=scaler.fit_transform(training_set))
ytr = y_tr.iloc[6:].values # remove first 6 values and transform to np array
lx_tr = []
for i in range(len(x_tr)-6):
    xi = x_tr.iloc[i:i+6,:] # get all 6-timestep time series
    lx_tr.append(xi)
x_train = np.asarray(lx_tr)
y_train = np.asarray(ytr)# np array
print(x_train.shape)

t_te = 'MS219' # choose device or station id
y_te = test_set[t_te]
x_te = pd.DataFrame(data=scaler.fit_transform(test_set))
yte = y_te.iloc[6:].values # remove first 6 values and transform to np array
lx_te = []
for i in range(len(x_te)-6):
    xi = x_te.iloc[i:i+6,:] # get all 6-timestep time series
    lx_te.append(xi)

x_test = np.asarray(lx_te)
y_test = np.asarray(yte)# np array
print(x_test.shape)

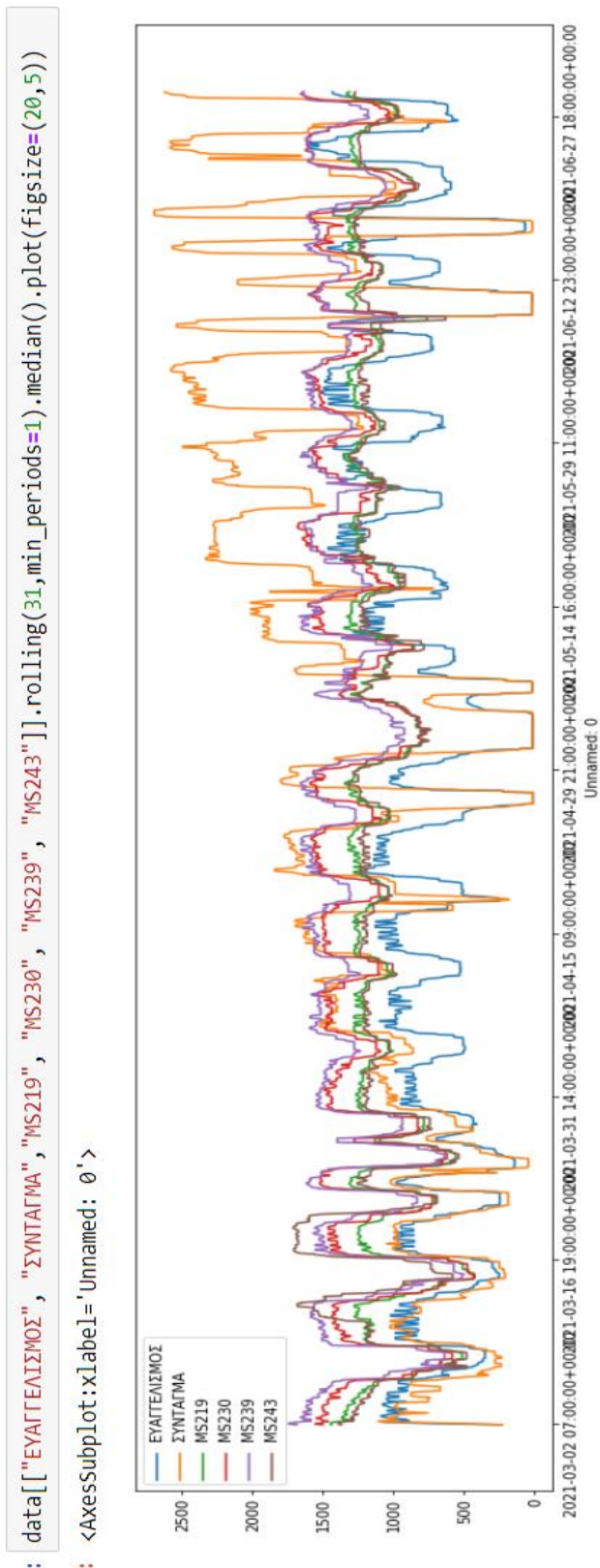
#Model

adam = Adam(learning_rate=0.001)
# Create LSTM model
model = Sequential()
model.add(LSTM(64, input_shape =(x_train.shape[1],5),return_sequences=False))
model.add(Dense(3,activation='relu'))
model.add(Dense(1))
model.compile(loss='mean_absolute_error',optimizer=adam)
history=model.fit(x_train[:,:[3,13,21,26,28]], y_train, epochs=180, batch_size=32)
model.predict(x_test[:,:[3,13,21,26,28]])

```

Παράρτημα Β

Διάγραμμα χρονοσειρών των επιλεγμένων στάσεων και φορατών και του φορατή – στόχου.



Διάγραμμα 14: Διάγραμμα χρονοσειρών φορατή-στόχου και επιλεγμένων φορατών και σταθμών μετρό.