Ακαδημαϊκό
Έτος
2020-21

Πρόβλεψη Αναμενόμενου Χρόνου Ροής Πλοίων Σήμερα Αφίξης Μηχανικής Μάθησης

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ
UNIVERSITY OF PIRAEUS

Θεοφιλόπουλος Γεώργιος

Τμήμα Πληροφορικής
Πανεπιστήμιο Πειραιώς, Πειραιά

Ακαδημαϊκό Έτος 2020/21

Πρόβλεψη Αναμενόμενου Χρόνου Άφιξης Πλοίων με Τεχνικές Μηχανικής Μάθησης

ΕΚΠΟΝΝΗΣΗ ΑΠΟ
ΘΕΟΦΙΛΟΠΟΥΛΟΣ ΓΕΩΡΓΙΟΣ

Τμήμα Πληροφορικής

ΕΤΟΣ 2020/21

Εξεταστική Επιτροπή

Πελέκης Νικόλαος

Θεοδωρίδης Ιωάννης

Ευχαριστίες

Θα ήθελα να εκφράσω τις ευχαριστίες μου προς τους καθηγητές μου Πελέκη Νικόλαο και Θεοδωρίδη Ιωάννη για το ενδιαφέρον και τη πολύτιμη βοήθειά τους στην ολοκλήρωση της πτυχιακής μου εργασίας. Ιδίως θα ήθελα να τονίσω πως η ευκαιρία που μου έδωσαν για να συνεργαστώ μαζί τους αποτέλεσε μια εκτενής επιμορφωτική εμπειρία για εμένα καθώς και αφορμή για να ενασχοληθώ με τον τόσο ενδιαφέρον τομέα των γεωγραφικών πληροφοριακών συστημάτων και της μηχανικής μάθησης.

Επιπλέον, θα ήθελα να ευχαριστήσω θερμά τον υποψήφιο Διδάκτορα του τμήματος Πληροφορικής του Πανεπιστημίου Πειραιώς, Ιωάννη Κοντούλη, για την εμπιστοσύνη που μου έδειξε καθώς και για την διαθεσιμότητα να βοηθήσει με τις όσες απορίες προέκυπταν κατά τη διάρκεια της εκπόνησης της πτυχιακής εργασίας. Θα ήταν παράλειψή μου να μην ευχαριστήσω τα υπόλοιπα μέλη του εργαστηρίου Data Science Lab του Πανεπιστημίου Πειραιώς για την υποστήριξη τους και για το φιλικό και ευχάριστο κλίμα που είχαν δημιουργήσει στα αντίστοιχα εργαστήρια.

Τέλος, θέλω από τα βάθη της καρδιάς μου να ευχαριστήσω την οικογένειά μου, για την ανεκτίμητη στήριξή τους, τόσο στην παρούσα εργασία, όσο και στην ολοκλήρωση των σπουδών μου.

Θεοφιλόπουλος Γεώργιος

Πρόβλεψη Αναμενόμενου Χρόνου Άφιξης Πλοίων με τεχνικές Μηχανικής Μάθησης

Γιώργος Θεοφιλόπουλος

7 Ιουλίου 2021

Περίληψη

Η ναυτιλία αποτελεί ένα σημαντικό κομμάτι της παγκόσμιας οικονομίας. Στο πλαίσιο της εποχής του ψηφιακού μετασχηματισμού ο τομέας αυτός δεν θα μπορούσε να μείνει ανεπηρέαστος. Εξελιγμένες εφαρμογές και περίπλοκα μοντέλα πρόβλεψης χρησιμοποιούνται καθημερινά για την διεκπεραίωση εργασιών και την στήριξη των ανθρώπινων αποφάσεων.

Το πρόβλημα του έγκυρου χρονοπρογραμματισμού και της μακροχρόνιας πρόβλεψης χρονικών καθυστερήσεων ταλανίζει ιδιαίτερα τον χώρο. Η γνώση του χρόνου άφιξης ενός πλοίου έχει μεγάλη αξία τόσο για τις ναυτιλιακές εταιρίες όσο και για τους λιμένες και τους ίδιους τους επιβάτες. Ο σκοπός της παρούσας πτυχιακής εργασίας είναι η σχεδίαση και η πρόταση ενός συστήματος μηχανικής μάθησης, το οποίο αναλύει τόσο τα υπάρχοντα δυναμικά όσο και τα στατικά δεδομένα για τον υπολογισμό του αναμενόμενου χρόνου άφιξης. Τέλος, τα υλοποιημένα μοντέλα και τους αντίστοιχους αλγορίθμους θα τους χρησιμοποιήσουμε για να πραγματοποιήσουμε δοκιμές σε αφίξεις στον λιμένα της Σαντορίνης.

Περιεχόμενα

1 Εισαγωγή **2**
1.1 Στόχος . . . 3
1.2 Ερευνητικα Ερωτήμα . . . 3

2 Βιβλιογραφική Ανασκόπηση **4**
2.1 Σχετικές Εφαρμογές . . . 4

3 Θεωρητικό Υπόβαθρο **7**
3.1 Μηχανική Μάθηση . . . 7
3.2 Νευρώνας . . . 8
3.3 Συνάρτηση Ενεργοποίησης . . . 9
3.4 Συνάρτηση Κόστους . . . 10
3.5 Αλγόριθμος Οπισθοδιάδοσης(Backpropagation) . . . 12
3.6 Δίκτυα RNN . . . 13

4 AIS Δεδομένα και Big Data **14**
4.1 AIS Δεδομένα . . . 14
4.2 Θόρυβος και Σφάλματα . . . 17
4.3 Λιμένες . . . 19
4.4 Κυκλοφοριακός Φόρτος Λιμένα . . . 21
4.5 Ημερολόγιο Αφίξεων Κρουαζιερόπλοιων Σαντορίνης . . . 22
4.6 Δεδομένα Καιρού . . . 23

5 Προεπεξεργασία Δεδομένων **25**
5.1 Καθαρισμός Δεδομένων . . . 25
5.2 Χωρική και Χρονική Διαμέριση . . . 27
5.3 Εξαγωγή Στάσιμων Σημείων και Οριοθέτηση Τροχιών . . . 29
5.4 Εύρεση Λιμένα Προέλευσης και Προορισμού . . . 30

6 Πρόβλεψη από Λιμένα σε Λιμένα **33**
6.1 Νευρωνικό Δίκτυο . . . 33
6.1.1 Αποτελέσματα . . . 34
6.2 Αναδρομικό Νευρωνικό Δίκτυο . . . 36
6.2.1 Αποτελέσματα . . . 37
6.3 Σύγκριση με το ημερολόγιο επισκέψεων κρουαζιερόπλοιων Σαντορίνης . . . 38
6.4 Πρόβλεψη σε πλοία ακτοπλοίας με προορισμό την Σαντορίνη . . . 43

7 Δοκιμές σε συνεχόμενες διαδρομές **45**

8 Βιβλιογραφία **48**

Κεφάλαιο 1

Εισαγωγή

Η συνεισφορά της ναυτιλίας σε μεταφορές ανθρώπων και αγαθών, όχι μόνο δεν παραμένει στάσιμη αλλά αυξάνεται σε παγκόσμιο επίπεδο. Προβλέπεται πως μέσα στην πενταετία 2019 - 2024 θα υπάρχει ετήσια αύξηση στις θαλάσσιες μεταφορές επί της τάξης του 3.4%. Βέβαια η σπουδαιότητα της ναυτιλίας πέρα από τους στατιστικούς δείκτες διαφαίνεται και στον ρόλο που έχει διαδραματίσει για την ανθρωπότητα από το παρελθόν μέχρι και σήμερα.

Τα τελευταία έτη παρατηρείται πως γίνεται μια επιτυχημένη προσπάθεια ψηφιακού μετασχηματισμού του τομέα. Οι αλλαγές αυτές αποσκοπούν τόσο στη δημιουργία ασφαλέστερων συνθηκών κατά την κίνηση στην θάλασσα όσο και συστημάτων που συμβάλουν στην λήψη αποφάσεων για την βέλτιστη αξιοποίηση των πόρων ενός πλοίου και της πλοήγησης του. Οι βέλτιστες αποφάσεις μεταφράζονται άμεσα σε αύξηση κέρδους. Βέβαια δεν επρόκειτο μόνο για χρηματικό κέρδος αλλά και περιβαλλοντικό και χρονικό. Στο πλαίσιο της παρούσας πτυχιακής εργασίας θα ασχοληθούμε με τον τελευταίο τύπο κέρδους, το χρονικό.

Οποιαδήποτε χρονική απόκλιση ενός πλοίου, μπορεί να προκαλέσει σύγχυση σε ανθρώπινο καθώς και σε λειτουργικό επίπεδο στο κομμάτι της εφοδιαστικής αλυσίδας. Για τα πλοία που μεταφέρουν εμπορεύματα σε λιμάνια ανά τον κόσμο, μια αρχική καθυστέρηση στην πραγματικότητα προκαλεί μια αλυσιδωτή καθυστέρηση τόσο στο επόμενο όσα και σε όλα τα επόμενα λιμάνια. Επιπρόσθετα, για ένα επιβατικό πλοίο, η καθυστέρηση δημιουργεί ένα αρνητικό κλίμα για τους επιβάτες, αυξάνοντας την πιθανότητα να μην ξανά επιλέξουν την ίδια εταιρεία για τις μεταφορές τους. Το σενάριο αυτό μεταφράζεται σε ένα υψηλό χρονικό και χρηματικό κόστος για τις εκάστοτε ναυτιλιακές εταιρείες. Επιπλέον, ο φορέας διαχείρισης του εκάστοτε λιμένα έχει προγραμματίσει το χρόνο αφίξεων και αναχωρήσεων των πλοίων και η οποιαδήποτε απόκλιση από το πρόγραμμα του, του δημιουργεί προβλήματα.

Για τον υπολογισμό του χρόνου άφιξης, παραδοσιακά χρησιμοποιούνταν συναρτήσεις που λαμβάνανε υπόψιν χωρικά και κινηματικά δεδομένα όπως η απόσταση μέχρι τον προορισμό και η ταχύτητα του πλοίου. Έπειτα σύμφωνα με την γεωγραφική θέση του πλοίου και την εμπειρία του εκάστοτε αναλυτή γινόταν μια πρόβλεψη του ενδεικτικού χρόνου άφιξης.

Τα φαινόμενα καθυστερήσεις όμως φαίνεται να επηρεάζονται άμεσα από τις καιρικές συνθήκες, την κατάσταση στα λιμάνια προορισμού, την εποχή του χρόνου καθώς και από άλλα δυναμικά δεδομένα τα οποία σχετίζονται με την θέση, τη ταχύτητα, την γενικότερη πορεία του πλοίου καθώς και τον αριθμό των προγραμματισμένων συνεχών επισκέψεων. Επειδή ακριβώς το πρόβλημα χαρακτηρίζεται από αρκετές στοχαστικές μεταβλητές με την μεταξύ τους σχέση περίπλοκη, είναι δύσκολο να αντιμετωπιστεί αποδοτικά χωρίς την συνεισφορά ενός έξυπνου συστήματος πρόβλεψης.

1.1 Στόχος

Στόχος της παρούσας πτυχιακής εργασίας είναι η μελέτη και η δημιουργία διαδικασιών για πρόβλεψη αναμενόμενων χρόνων άφιξης, μεγαλύτερης εγκυρότητας, κάνοντας χρήση των τελευταίων εξελίξεων στον τομέα της μηχανικής μάθησης.

1.2 Ερευνητικά Ερωτήμα

Στο πλαίσιο της εργασίας θα επιχειρηθεί να δοθούν απαντήσεις στις κάτωθι ερωτήσεις:

1. Πως μπορούν να αναπαρασταθούν τα διάσπαρτα γεωχωρικά σήματα σε χρήσιμες και νοηματικές τροχιές;
2. Τι παράγοντες και σε ποιόν βαθμό επηρεάζουν την χρονική συνέπεια των πλοίων;
3. Πόσα καλά μπορούν τα διάφορα νευρωνικά δίκτυα να δώσουν ικανοποιητικές προβλέψεις και για ποιόν λόγο οι προβλέψεις τους αποκλίνουν από την πραγματικότητα;
4. Πόσο διαφέρουν οι προγραμματισμένοι χρόνοι άφιξης των πλοίων από τους πραγματικούς;

Κεφάλαιο 2

Βιβλιογραφική Ανασκόπηση

Η πρόβλεψη του αναμενόμενου χρόνου άφιξης έχει απασχολήσει πολλούς επιστημονικούς κλάδους και έχει γίνει αντικείμενο έρευνας πολλάκις φορές. Οι περισσότερες εφαρμογές παρατηρούνται στον χώρο των γεωγραφικών συστημάτων. Μάλιστα, οι τεχνικές που χρησιμοποιούνται για προβλέψεις σε πλοία, εφαρμόζονται και σε αεροπλάνα, οδικά οχήματα ή και λεωφορεία. Στην πράξη, το πρόβλημα πρόβλεψης του αναμενόμενου χρόνου άφιξης ανάγεται σε πρόβλημα επεξεργασίας και ανάλυσης χρονοσειρών, εμπλουτισμένες με γεωχωρικά, στατικά και κινηματικά δεδομένα.

Το πρόβλημα έχει προσεγγιστεί με πολλούς και διαφορετικούς τρόπους. Για παράδειγμα στην μελέτη DeepTTE, λαμβάνονται υπόψη οι χωρικές συντεταγμένες των σημάτων κάθε τροχιάς και ύστερα με τη χρήση συνελικτικων μοντέλων γίνεται προσπάθεια για την κατανόηση της σχέσης μεταξύ της χωρικής διάταξης και της ανάλογης καθυστέρησης. Με τον τρόπο αυτό παράμετροι όπως διασταυρώσεις, κυκλικοί κόμβοι καθώς και φανάρια μπορούν να ληφθούν υπόψιν για τον υπολογισμό του τελικού αναμενόμενου χρόνου άφιξης. Έπειτα με τη συμμετοχή των αναδρομικών δικτύων καταγράφονται και οι χρονικές σχέσεις όπως ο ρυθμός μεταβολής της απόστασης, ο τρόπος οδήγησης κτλ.

Έρευνες αντιμετωπίζουν το πρόβλημα με την ανάλυση των τελικών τροχιών σε μικρότερα κομμάτια και έπειτα κάνοντας χρήση ιστορικών δεδομένων προβαίνουν στον υπολογισμό του τελικού χρόνου άφιξης, δηλαδή για κάθε επιμέρους κομμάτι της τροχιάς λαμβάνουν υπόψιν το συνολικό χρόνο που χρειάστηκε σε κάθε όχημα για να διανύσει το αντίστοιχο κομμάτι και με βάση αυτό πραγματοποιούν την τελική πρόβλεψη.

Άλλοι συγγραφείς κάνουν χρήση των φίλτρων Kalman τα οποία επιχειρούν να μοντελοποιήσουν τον τρόπο κίνησης ενός κινητού ώστε να προβλέψουν τον αναμενόμενο χρόνο άφιξης. Τέλος, ειδικά μοντέλα τύπου Spatio-Temporal Hidden Markov Models (STHMM) έχουν προταθεί στην βιβλιογραφία με τα οποία υπολογίζεται η διασπορά μεταξύ χρονοσειρών κίνησης και επιλέγεται κάθε φορά η κοντινότερη χρονοσειρά με αυτή του κινητού που εξετάζουμε.

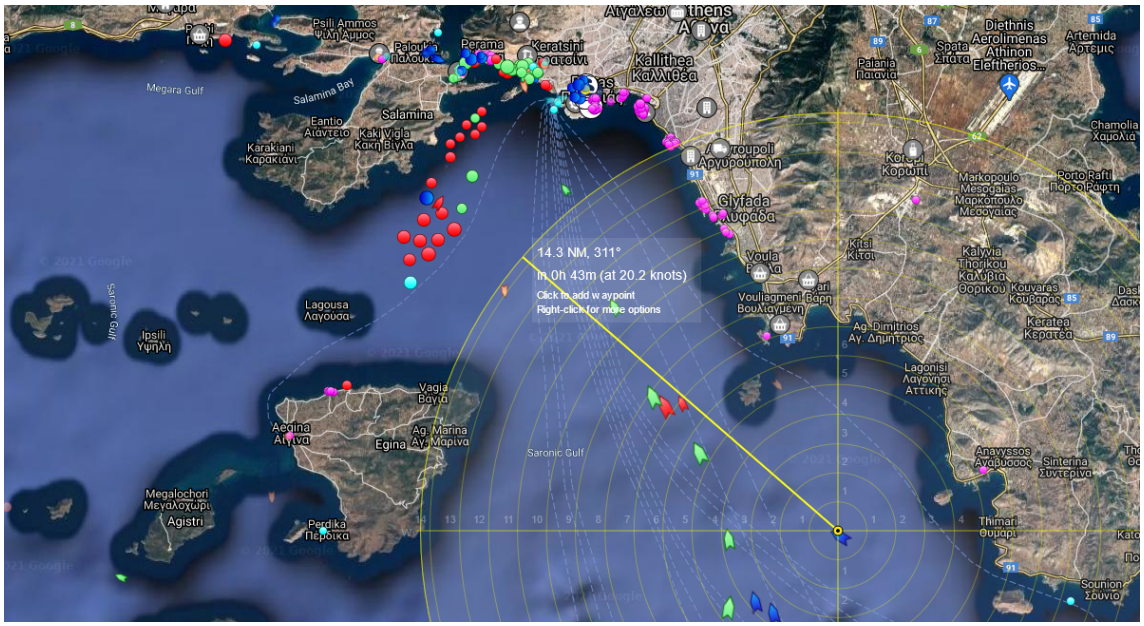
Στην παρούσα πτυχιακή γίνεται πρόβλεψη σε όλο το μήκος της τροχιάς χωρίς να γίνεται χώρισμα σε μικρότερα τμήματα και χωρίς να συμπεριλαμβάνονται στατικές πληροφορίες που περιορίζουν την χρήση των μοντέλων. Έτσι, η πρόβλεψη των νευρωνικών δικτύων δεν περιορίζεται σε μια συγκεκριμένη περιοχή ή στην πρόβλεψη συγκεκριμένων πλοίων.

2.1 Σχετικές Εφαρμογές

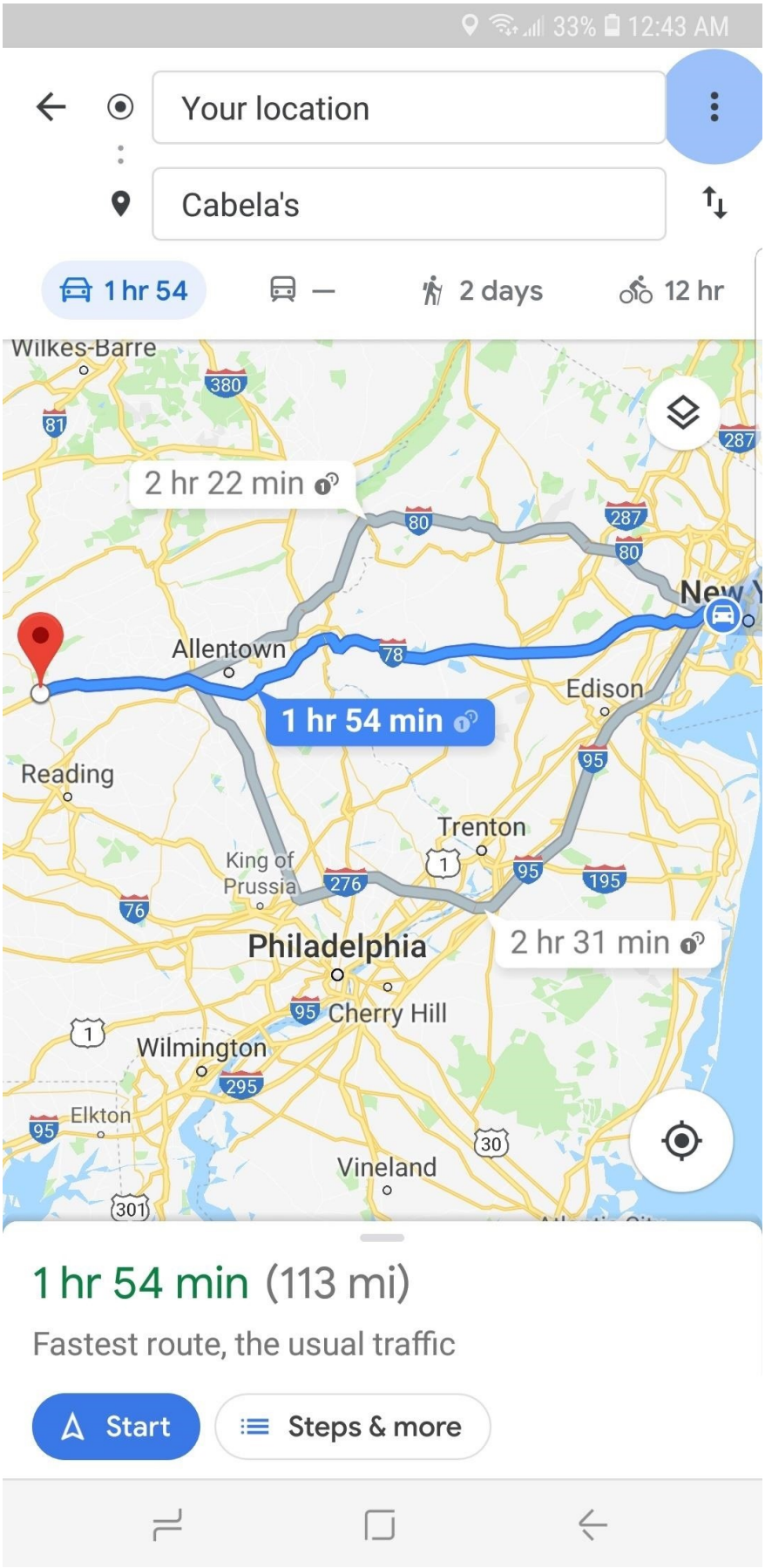
Ενδεικτικά εταιρείες που δραστηριοποιούνται στον τομέα της πρόβλεψης ETA για ναυτιλιακές εφαρμογές είναι η MarineTraffic, η Fleetmoon και η Vesselfinder. Αντίστοιχα, παρόμοιες υλοποιήσεις συναντώνται στις αεροπορικές μεταφορές όπως για παράδειγμα η FlightRadar24

και η FlightAware και στον τομέα των οδικών μεταφορών όπως η λύση που δίνει το Google Maps.

Στην παρακάτω εικόνα παρουσιάζεται στιγμιότυπο από την εφαρμογή MarineTraffic αναφορικά με την πληροφορία που δίνεται στους δωρεάν χρήστες για την πρόβλεψη της διαδρομής του πλοίου καθώς και τον αναμενόμενο χρόνο άφιξης του τόσο για τον τελικό προορισμό όσο και τα ενδιάμεσα turning points.



Σχήμα 2.1: Στιγμιότυπο από την εφαρμογή MarineTraffic



Σχήμα 2.2: Στιγμιότυπο από την εφαρμογή Google Maps ETA

Κεφάλαιο 3

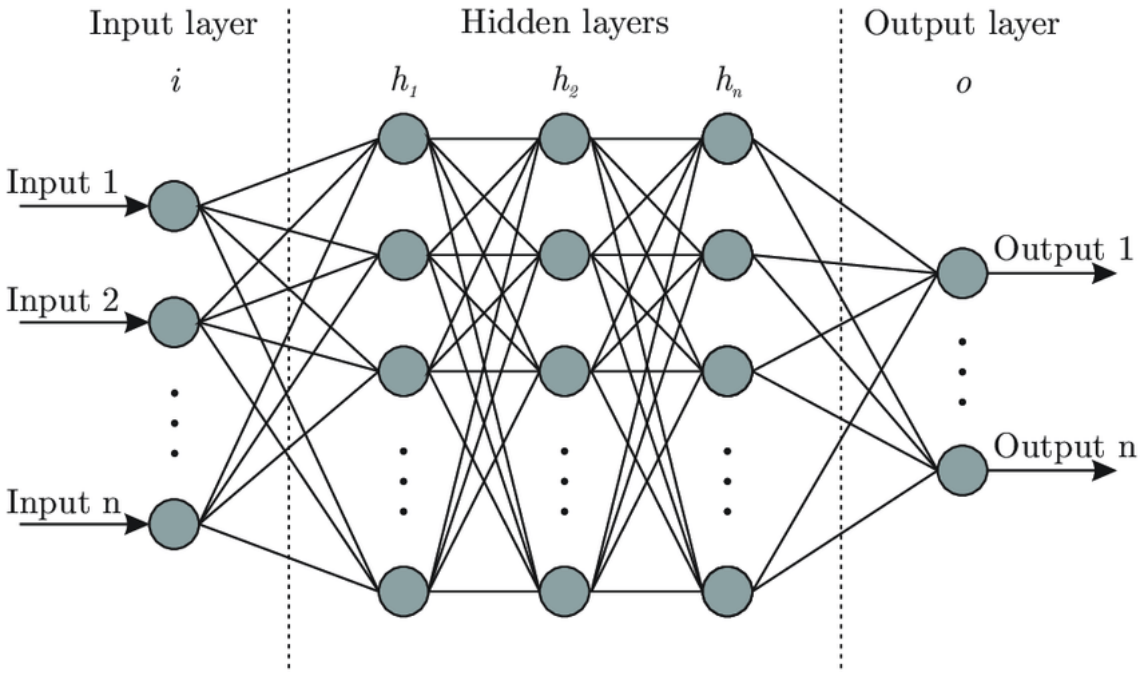
Θεωρητικό Υπόβαθρο

Η μηχανική μάθηση είναι υποπεδίο της επιστήμης των υπολογιστών που αναπτύχθηκε από τη μελέτη της αναγνώρισης προτύπων και της αριθμητικής ανάλυσης. Η μηχανική μάθηση είναι στενά συνδεδεμένη και συχνά συγχέεται με την υπολογιστική στατιστική, ένας κλάδος, που επίσης επικεντρώνεται στην πρόβλεψη μέσω της χρήσης των υπολογιστών. Έχει ισχυρούς δεσμούς με την μαθηματική βελτιστοποίηση, η οποία παρέχει μεθόδους με τις οποίες εκπαιδεύονται τα αντίστοιχα μοντέλα. Στο σημείο αυτό σημειώνονται όλοι οι αλγόριθμοι, οι τεχνικές, οι μεθοδολογίες καθώς και η ανάλογη θεωρία που επρόκειτο να χρησιμοποιηθούν στα επόμενα κεφάλαια.

3.1 Μηχανική Μάθηση

Η μηχανική μάθηση ορίζεται ως η δυνατότητα του υπολογιστή να μαθαίνει από τα δεδομένα που του δίνονται. Οι αλγόριθμοι με τους οποίους μαθαίνει ένας υπολογιστής μπορούν να ταξινομηθούν σε τρεις κατηγορίες. Επιτηρούμενη μάθηση, μη επιτηρούμενη μάθηση και ενισχυτική μάθηση. Στο πλαίσιο της παρούσας πτυχιακής θα εξεταστούν μόνο οι αλγόριθμοι επιβλεπόμενης μάθησης. Τα νευρωνικά δίκτυα(neural networks) και ο αλγόριθμος MLP - Multilayer Perceptron είναι μια κατηγορία τέτοιων αλγορίθμων.

Ο όρος επιβλεπόμενη μάθηση αναφέρετε στην προσπάθεια του υπολογιστή να αντιστοιχήσει ένα ανεξάρτητο διάνυσμα εισόδου στην εξαρτημένη τιμή, στο μέγεθος δηλαδή που θέλουμε να προβλέψουμε. Η διαδικασία μάθησης με την χρήση ενός MLP, ταυτίζεται με την μαθηματική προσέγγιση μιας συνάρτησης $y \approx f^*(x)$. Ένα σύστημα χρονικής πρόβλεψης στην πράξη αυτό που κάνει είναι να αντιστοιχεί ένα διάνυσμα εισόδου, έστω $\vec{x}$ σε μια πραγματική συνεχής τιμή y που συμβολίζει τον αναμενόμενο χρόνο. Στόχος του νευρωνικού δικτύου είναι η εύρεση της καλύτερης δυνατής συνάρτησης $y = f(x; w)$ μαθαίνοντας τις παραμέτρους w που οδηγούν στην καλύτερη δυνατή προσέγγιση. Είναι φανερό πως στην πραγματικότητα ο όρος μηχανική μάθηση αναφέρεται στον υπολογισμό/μάθηση αυτών των παραμέτρων.



Σχήμα 3.1: Τεχνητό Νευρωνικό Δίκτυο - Τα νευρωνικά δίκτυα αυτού του τύπου είναι ο θεμελιώδης λίθος της μηχανικής μάθησης μιας και οι περισσότεροι αλγόριθμοι είναι παραλλαγές αυτών.

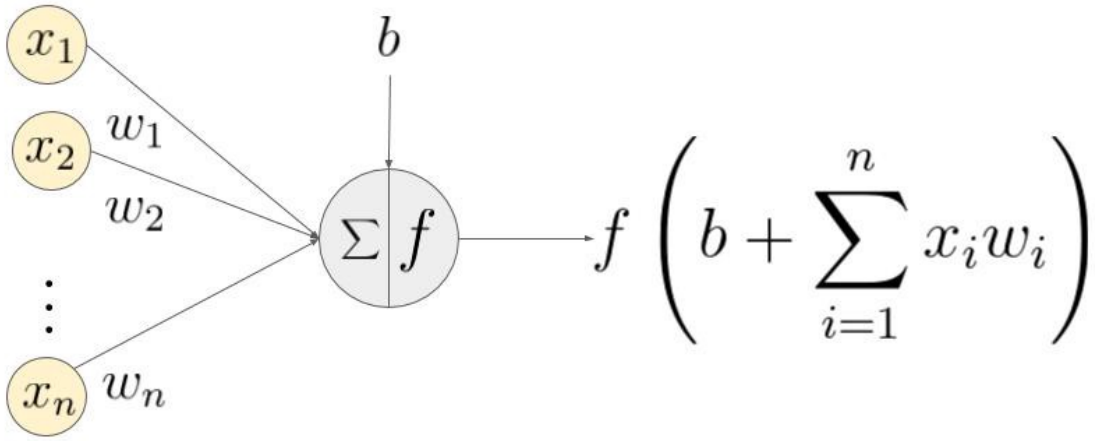
3.2 Νευρώνας

Ένα νευρωνικό δίκτυο είναι ένα σύνολο συνδεδεμένων νευρώνων. Καθένας από τους νευρώνες αυτούς ονομάζεται και perceptron. Ο νευρώνας perceptron επινοήθηκε το 1958 και ήταν αποτέλεσμα της προσπάθειας του ανθρώπου στο να μοντελοποιήσει τον εγκέφαλο και συγκεκριμένα έναν βιολογικό νευρώνα. Όπως ο βιολογικός νευρώνας είναι η κεντρική μονάδα του νευρικού μας συστήματος, έτσι και ένα νευρωνικό δίκτυο αποτελείται από ένα σύνολο νευρώνων.

Οι βιολογικοί νευρώνες δέχονται και παράγουν ερεθίσματα υπό την μορφή ηλεκτρικών σημάτων τα οποία μεταδίδονται μέσα από τις νευροαξονικές απολήξεις. Στον perceptron τα ερεθίσματα εισόδου είναι αριθμητικές τιμές που δίνονται σε μορφή διανύσματος.

Στο επόμενο σχήμα παρατηρείται ένα σύνολο μεταβλητών $(x_1, x_2, \cdots, x_n)$ οι οποίες αφού πολλαπλασιαστούν με κάποιο βάρος $(w_1, w_2, \cdots, w_n)$, δίνονται ως είσοδοι σε μια συνάρτησης ενεργοποίησης. Το αποτέλεσμα της συνάρτησης ενεργοποίησης είναι η αντίδραση του νευρώνα.

Στην πράξη, ένας νευρώνας είναι σύνθεση δύο συναρτήσεων όπου η πρώτη υπολογίζει των γραμμικό συνδυασμό του διανύσματος εισόδου $(x_1, x_2, \cdots, x_n)$ με το διάνυσμα βαρών $(w_1, w_2, \cdots, w_n)$ και η δεύτερη μετασχηματίζει το αποτέλεσμα αυτό με βάση την δοσμένη συνάρτηση ενεργοποίησης.



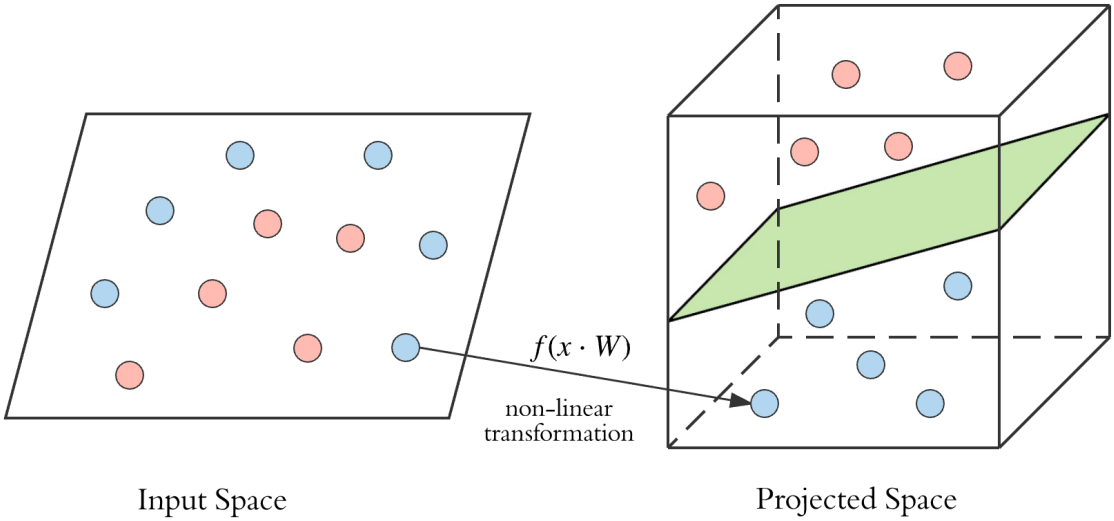
Σχήμα 3.2: Ένα παράδειγμα νευρώνα που η αντίδραση του ισούται με τον γραμμικό συνδυασμό των διανυσμάτων $(w_1, w_2, \cdots, w_n)$ και $(x_1, x_2, \cdots, x_n)$ και πρόθεσης με την σταθερά b - bias. Παρατηρείται πως στο αποτέλεσμα του αθροίσματος εφαρμόζεται η συνάρτηση ενεργοποίησης και ύστερα προκύπτει το αποτέλεσμα του νευρώνα.

Η σταθερά β - bias είναι το αποτέλεσμα του νευρώνα όταν δίνεται ένα διάνυσμα μηδενικών τιμών. Χρησιμοποιείται για την καλύτερη προσέγγιση της επιθυμητής συνάρτησης αφού με την χρήση της σταθεράς μπορεί να γίνει μετατόπιση της συνάρτησης.

3.3 Συνάρτηση Ενεργοποίησης

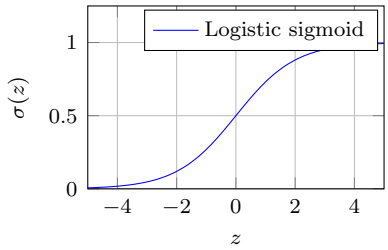
Παρατηρώντας τη δομή και τη λειτουργία των νευρικών κυττάρων, παρατηρήθηκε πως αν και μπορεί να δέχονται συχνά εξωτερικά ερεθίσματα, δεν αντιδρούν πάντα με το να παράγουν κάποιο σήμα εξόδου προς άλλα κύτταρα. Αυτό συμβαίνει στις περιπτώσεις που το σήμα εισόδου δεν είναι αρκετό(σημαντικό) ώστε να τα ενεργοποιήσει. Στα τεχνητά νευρωνικά δίκτυα, τον ρόλο αυτό κατέχει η συνάρτηση ενεργοποίησης. Η συνάρτηση ενεργοποίησης επηρεάζει την απόκριση του νευρώνα, μετασχηματίζοντας την, είτε αυξάνοντας το μέγεθος της άρα και την επιρροή της, είτε ελαττώνοντας την με την αλλαγή της σε μια πολύ μικρή ή μηδενική τιμή.

Από μαθηματικής σκοπιάς, στόχος των συναρτήσεων ενεργοποίησης είναι η εισαγωγή μη - γραμμικότητας στο νευρωνικό δίκτυο. Με τον τρόπο αυτό δίνεται η δυνατότητα να προσεγγιστούν συναρτήσεις οι οποίες δεν είναι γραμμικές, κάτι το οποίο είναι αρκετά σύνηθες στην πραγματικότητα. Αφού ένας νευρώνας στην πράξη είναι μια ευθεία γραμμή και μπορεί να διαχωρίσει έναν χώρο σε μόνο δύο υποχώρους, η συνάρτηση ενεργοποίησης μετασχηματίζει τα δεδομένα έτσι ώστε να είναι γραμμικώς διαχώρισμα.

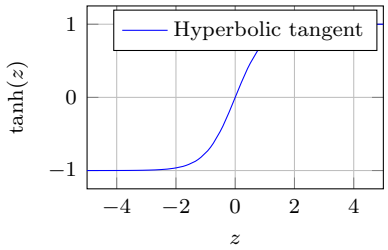


Σχήμα 3.3: Μετασχηματισμός προβλήματος σε γραμμικά διαχωρίσημο μέσω της συνάρτησης ενεργοποίησης.

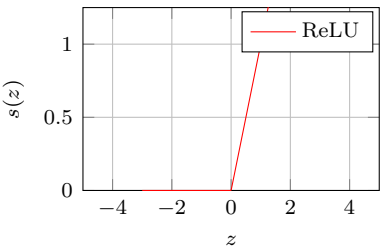
Υπάρχουν διάφορες συναρτήσεις ενεργοποίησης όπως βλέπουμε παρακάτω. Η τελική επιλογή εξαρτάται από τον στόχο του εκάστοτε δικτύου και επιλέγεται ύστερα από δοκιμές και συγκρίσεις.



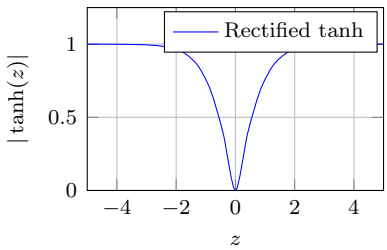
(α′) Logistic sigmoid activation function.



(β′) Hyperbolic tangent activation function.



(γ′) Relu activation function.



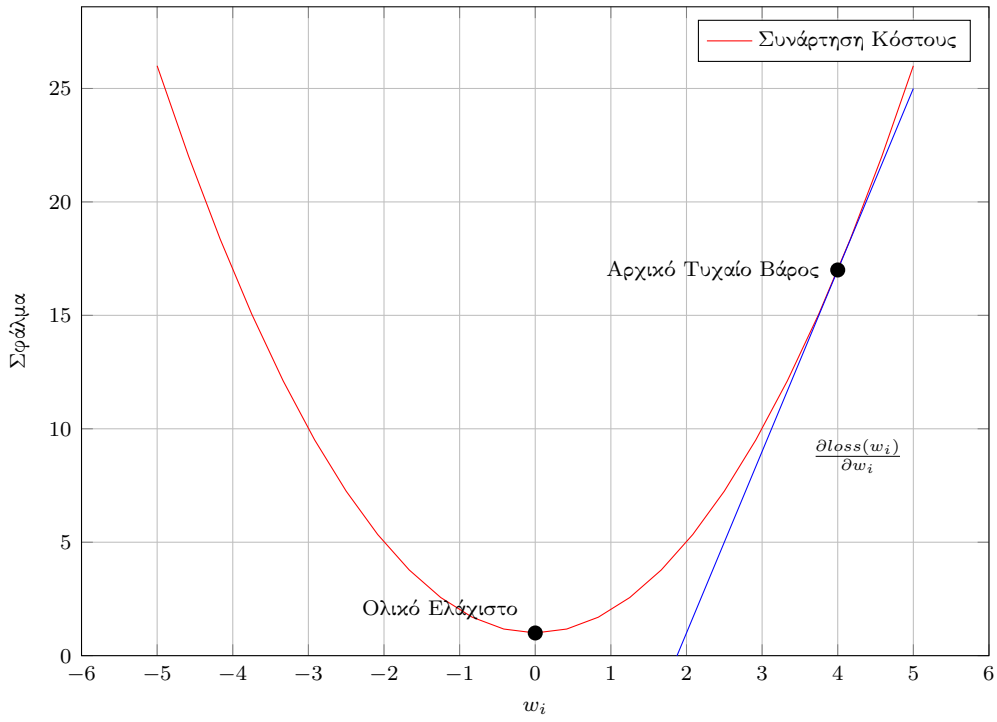
(δ′) Rectified hyperbolic tangent activation function.

3.4 Συνάρτηση Κόστους

Η συνάρτηση κόστους είναι πολύ σημαντική για τα νευρωνικά δίκτυα και είναι ο μόνο τρόπος με τον οποίο μπορεί να ποσοτικοποιηθεί το σφάλμα τους. Η εύρεση ενός ελάχιστου σημείου, ισοδυναμεί με βελτίωση της απόδοσης του μοντέλου. Για κάθε δεδομένο συνδυασμό βαρών, η συνάρτηση κόστους υπολογίζει το πραγματικό σφάλμα. Συνεπώς, είναι φανερό πως αν

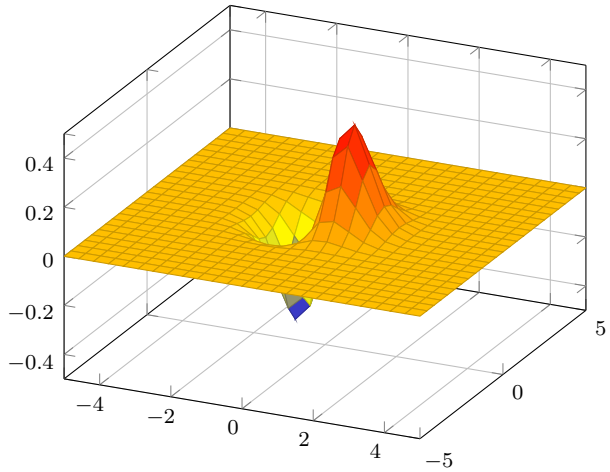
υπάρχει το ολικό της ελάχιστο δίνεται από τον καλύτερο δυνατό συνδυασμό βαρών, δηλαδή όπως επισημάνθηκε και προηγουμένως από τον συνδυασμό που προσεγγίζει καλύτερα την επιθυμητή συνάρτηση που μαθαίνει το νευρωνικό δίκτυο.

Αν θεωρητικά υπήρχε άπειρη υπολογιστική δύναμη και χρόνος θα μπορούσε να υπολογιστεί το σφάλμα για κάθε διαφορετική τιμή της μεταβλητής του βάρους. Έτσι, η συνάρτηση σφάλματος θα μπορούσε να αναπαρασταθεί όπως παρακάτω, με τον άξονα των τετμημενών αν αναπαριστά τις διαφορετικές τιμές βαρών και των άξονα των τεταγμένων το τελικό σφάλμα για την κάθε δυνατή τιμή.



Σχήμα 3.4: Συνάρτηση κόστους στον δυσδιάστατο χώρο

Σκοπός είναι ο εντοπισμός της τιμής εκείνης που δίνει το ολικό ελάχιστο όπως φαίνεται στο σχήμα. Στην πράξη επειδή οι συναρτήσεις δεν είναι πάντα δισδιάστατες γραμμικές συναρτήσεις και υπάρχουν πολλά διαφορετικά βάρη(πολυμεταβλητές συναρτήσεις), η συνάρτηση σφάλματος είναι πιθανόν να μοιάζει ως κάτωθι.



Σχήμα 3.5: Συνάρτηση κόστους στον τρισδιάστατο χώρο

Δεδομένου ότι η πολυπλοκότητα της συνάρτησης αυξάνεται καθώς αυξάνεται η πολυπλοκότητα του προβλήματος που προσπαθούμε να λύσουμε και επειδή δεν έχουμε άπειρη υπολογιστική δύναμη και χρόνο, δεν γίνεται να υπολογίσουμε την τιμή του σφάλματος για κάθε δυνατό συνδυασμό βαρών. Στην πράξη, χρησιμοποιούμε επαναληπτικούς αλγορίθμους βελτιστοποίησης που προσπαθούν να προσεγγίσουν το ολικό ελάχιστο ή κάποιο επιθυμητό τοπικό ελάχιστο που μας προσδίδει ικανοποιητική ακρίβεια.

Μέσο Τετραγωνικό Σφάλμα	$\mathrm{MSE} = \frac{1}{n}\sum_{t=1}^{n} e_t^2$
Ρίζα Μέσου Τετραγωνικού Σφάλματος	$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{t=1}^{n} e_t^2}$
Μέσο Απόλυτο Σφάλμα	$\mathrm{MAE} = \frac{1}{n}\sum_{t=1}^{n} \lvert e_t \rvert$
Μέσο Απόλυτο Ποσοστιαίο Σφάλμα	$\mathrm{MAPE} = \frac{100\%}{n}\sum_{t=1}^{n} \left\lvert \frac{e_t}{y_t} \right\rvert$

Σχήμα 3.6: Συχνές συναρτήσεις κόστους

3.5 Αλγόριθμος Οπισθοδιάδοσης(Backpropagation)

Η τελική έξοδος του νευρωνικού δίκτυο καθώς και το σφάλμα είναι αποτέλεσμα της σύνθεσης ενός συνόλου συναρτήσεων, δηλαδή ενός συνόλου νευρώνων. Από τον ορισμό του νευρώνα γίνεται κατανοητό πως το αρχικό διάνυσμα εισόδου πολλαπλασιάζεται με τα αντίστοιχα βάρη και τροφοδοτείται στο επόμενο επίπεδο. Η διαδικασία αυτή επαναλαμβάνεται από επίπεδο σε επίπεδο μέχρι να καταλήξει στο στρώμα εξόδου, όπου παράγεται και το τελικό αποτέλεσμα. Έτσι το κάθε στρώμα και συγκεκριμένα τα βάρη του κάθε νευρώνα είναι άμεσα υπεύθυνα και το καθένα επηρεάζει λιγότερο ή περισσότερο το σφάλμα.

Το ερώτημα που προκύπτει είναι κατά πόσο επηρεάζει το κάθε βάρος το τελικό αποτέλεσμα και πως θα μεταβληθούν τα βάρη καθαυτά για να επιτευχθούν επίπεδα υψηλότερης ακρίβειας. Τις απαντήσεις σε αυτά τα ερωτήματα μας της δίνει ο αλγόριθμος Backpropagation. Επί της

ουσίας ο αλγόριθμος υπολογίζει τις μερικές παραγώγους της συνάρτησης κόστους με βάση τα αντίστοιχα βάρη του δικτύου. Ο υπολογισμός των παραγώγων γίνεται με την τεχνική της αυτόματης διαφοροποίησης. Η αυτόματη διαφοροποίηση είναι μια τεχνική που παρέχει αυτόματα τις μερικές παραγώγους των συναρτήσεων στον αλγόριθμο εκπαίδευσης.

$$w_i = w_i - \eta \frac{\partial C}{\partial w_i} \tag{3.1}$$

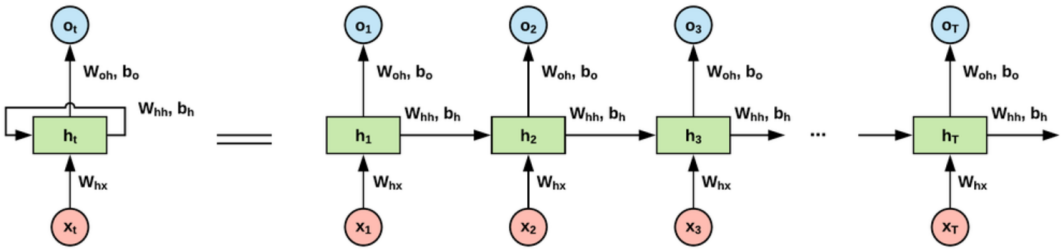
Κατά τον αλγόριθμο Gradient Descent η τιμή του νέου - ανανεωμένου βάρους ισούται με την προηγούμενη μεταβεβλημένη κατά έναν αριθμό. Ο αριθμός αυτός ισούται με την μερική παράγωγο της συνάρτησης κόστους του κάθε βάρους w_i πολλαπλασιασμένη με μια σταθερά η. Ο αριθμός η ονομάζεται learning rate και επί της ουσίας ρυθμίζει τον ρυθμό μάθησης αυξάνοντας ή μειώνοντας την επίδραση της παραγώγου.

Ο λόγος που γίνεται χρήση μερικών παραγώγων είναι διότι μας δίνουν απάντηση στο ερώτημα 'Πόσο μια μικρή αλλαγή στο βάρος w_i θα επηρεάσει το σφάλμα μας;' , με άλλα λόγια μας βοηθάνε να υπολογίσουμε την συμβολή κάθε βάρους στο τελικό σφάλμα.

Πρέπει να τονιστεί πως υπάρχουν διάφορες τεχνικές υπολογισμού παραγώγων με τη χρήσης μηχανής. Για παράδειγμα υπάρχει η αριθμητική παραγώγιση η οποία βασίζεται στον υπολογισμό της παραγώγου με την χρήση του ορισμού και των ορίων. Βέβαια αυτή η μέθοδος είναι επιρρεπής σε σφάλματα λόγο της πεπερασμένης ακρίβειας για να την αναπαράσταση αριθμών κινητής υποδιαστολής. Αντίστοιχα, η συμβολική παραγώγιση είναι αργή και επηρεάζει την ταχύτητα σύγκλισης των δικτύων. Στην πράξη χρησιμοποιείται η αυτόματη παραγώγιση η οποία προαπαιτεί την κατασκευή ενός υπολογιστικού γράφου ο οποίος περιγράφει την εκάστοτε μαθηματική έκφραση. Η εφεύρεση της αυτόματης παραγώγισης είναι ένας από τους παράγοντες που οδήγησε στην μεγάλη απόδοση των δικτύων καθώς και ο λόγος που επανήλθε το ερευνητικό ενδιαφέρον για το πεδίο της μηχανικής μάθησης.

3.6 Δίκτυα RNN

Τα ανατροφοδοτούμενα νευρωνικά δίκτυα αποτελούν εξέλιξη του βασικού νευρωνικού δικτύου. Επί της ουσίας οφείλουν την ύπαρξη τους στην αδυναμία του ANN να καταγράψει την χρονική εξάρτηση μεταξύ των δεδομένων του προβλήματος. Τα δίκτυα τύπου RNN δεν επιχειρούν μόνο να μοντελοποιήσουν την σχέση μεταξύ των χαρακτηριστικών του διανύσματος εισόδου αλλά επιπλέον μελετούν και την επίδραση των προηγούμενων στιγμιοτύπων των δεδομένων στην έξοδο του δικτύου. Για τον λόγο αυτό χρησιμοποιούνται συχνά σε προβλήματα χρονοσειρών όπως για παράδειγμα στον τομέα της φυσικής επεξεργασίας της γλώσσας και στην ανάλυση κινούμενων αντικειμένων και τροχιών.



Σχήμα 3.7: Δομή Κλασσικού PNN

Με βάση το παραπάνω σχήμα γίνεται κατανοητό πως η απόκριση του δικτύου σε μια χρονική στιγμή t ισούται με $h_t = f(h_{t-1}; x_t; w)$

Κεφάλαιο 4

AIS Δεδομένα και Big Data

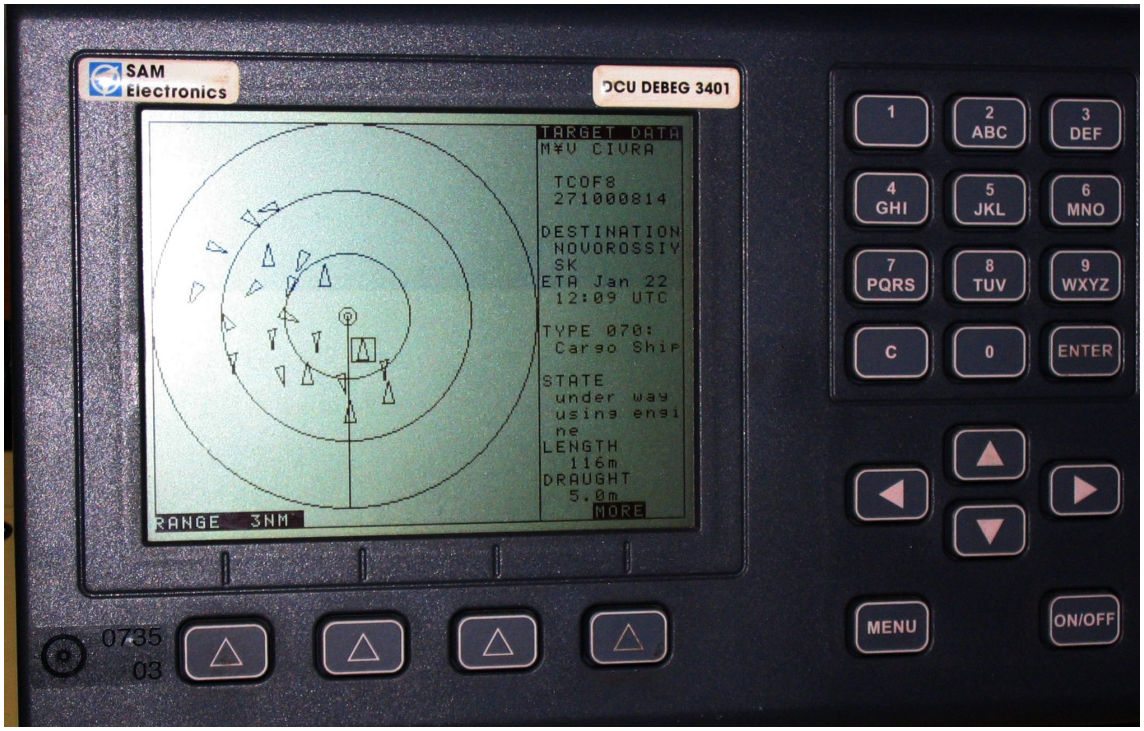
Ο όρος "Μεγάλα Δεδομένα"(Big Data) χρησιμοποιήθηκε για να περιγράψει σύνολα δεδομένων, με όγκο τέτοιο που καθιστά ανίκανα τα διαθέσιμα συστήματα επεξεργασίας να τα επεξεργαστούν σε βιώσιμο χρόνο. Προσεγγιστικά υπολογίζεται πως καθημερινά παράγονται 2.5 πεντακισεκατομμύρια bytes τέτοιων δεδομένων. Παρ΄ όλα αυτά τα μεγάλα δεδομένα δεν χαρακτηρίζονται μόνο από τον όγκο τους αλλά και από την ταχύτητα με την οποία παράγονται καθώς και την πολυπλοκότητα τους. Η ραγδαία ανάπτυξη της τεχνολογίας και συγκεκριμένα του διαδικτύου δημιούργησε τις ευνοϊκές συνθήκες οι οποίες οδήγησαν στην κατασκευή, αποθήκευση και επεξεργασία των δεδομένων αυτών.

Τεράστιοι φυσικοί χώροι αποθήκευσης(data centers), μέσα μεγάλης χωρητικότητας καθώς και συστήματα παράλληλης επεξεργασίας δώσανε την δυνατότητα σε επιστήμονες από διαφορετικούς κλάδους να αξιοποιήσουν τα περίπλοκα δεδομένα και να βγάλουμε απλά και πολύτιμα αποτελέσματα. Εφαρμογές του Big Data παρατηρούμε τόσο σε τεχνολογικές εφαρμογές, όσο και σε ιατρικές, εκπαιδευτικές, αθλητικές, ναυτιλιακές κλπ.

Όσον αφορά στον ναυτιλιακό τομέα, οι περισσότερες εταιρείες έχουν αναπτύξει τα δικά τους συστήματα ανάλυσης και αξιοποίησης των δεδομένων ή συνεργάζονται με μια αντίστοιχη εταιρεία του χώρου. Μερικές ενδεικτικές εφαρμογές είναι η αυτόματη πρόβλεψη διαδρομής, λαμβάνοντας υπόψη περίπλοκους παράγοντες όπως ο καιρός και η θαλάσσια ασφάλεια, η βέλτιστη χρήση και ο χειρισμός του πλοίου καθώς και η πρόβλεψη διάφορων προβλημάτων - ανωμαλιών όπως προβλήματα σε λειτουργικά μέρη και πιθανές καθυστερήσεις.

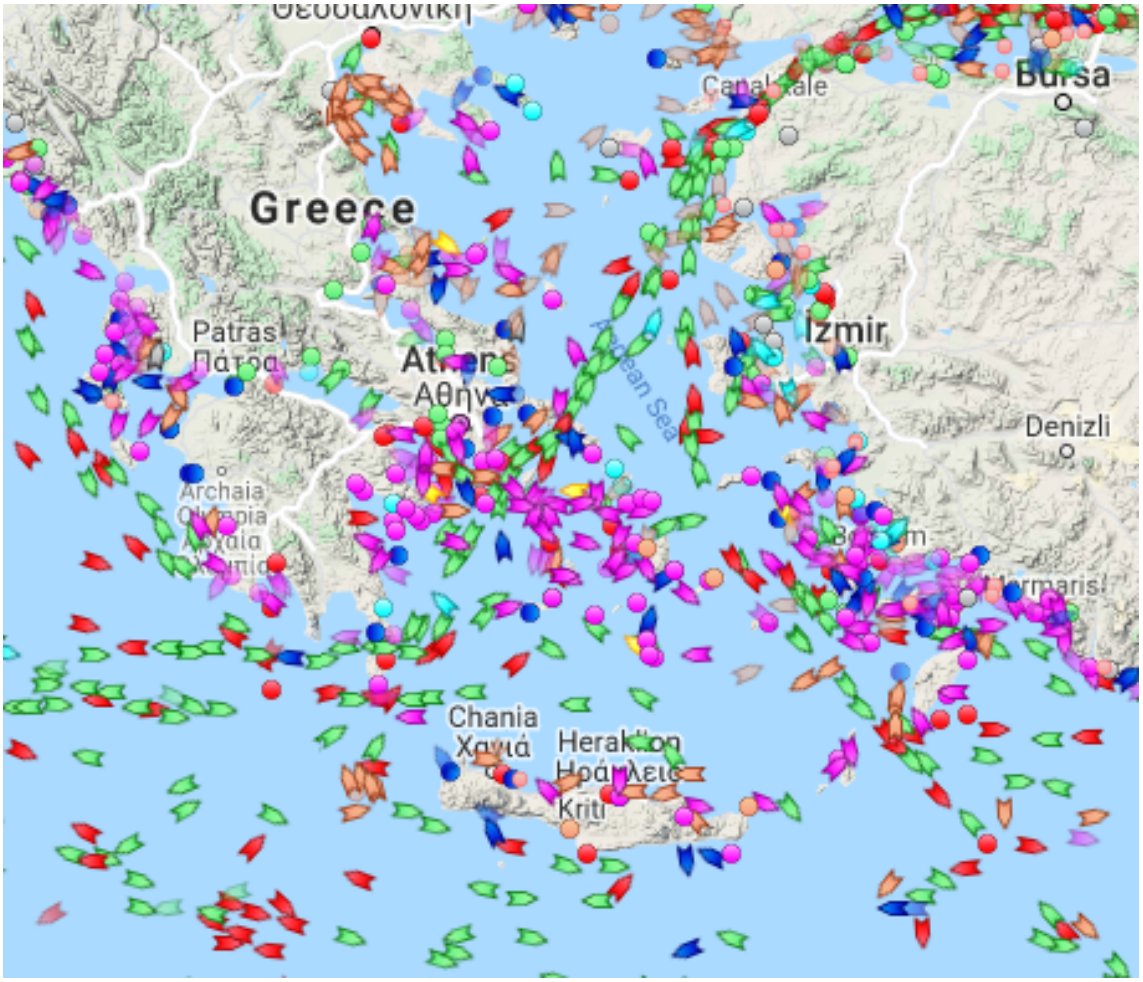
4.1 AIS Δεδομένα

Το Αυτόματο Σύστημα Αναγνώρισης - AIS χρησιμοποιείται για την ανταλλαγή ψηφιακών σημάτων μεταξύ πλοίων. Σκοπός των σημάτων τύπου AIS είναι να ενισχυθεί η ασφάλεια και η αποτελεσματικότητα της θαλάσσιας κυκλοφορίας, να βελτιωθεί η ανταπόκριση των αρμόδιων αρχών σε συμβάντα, ατυχήματα, ή επικίνδυνα περιστατικά στη θάλασσα, μεταξύ άλλων με επιχειρήσεις έρευνας και διάσωσης, και να διευκολυνθεί η πρόληψη και η ανίχνευση της ρύπανσης από πλοία. Αν και αρχικά χρησιμοποιήθηκε για λόγους ασφαλείας, σήμερα βρίσκει εφαρμογές σε προηγμένα συστήματα υποβοήθησης και αξιολόγησης πολλών ναυτιλιακών διαδικασιών όπως η αποδοτικότερη επιλογή θαλάσσιας διαδρομής ή βέλτιστη χρήση των πόρων του πλοίου. Σε αυτές εντάσσεται και η χρονική πρόβλεψη που μελετάμε στην παρούσα εργασία. Αξίζει να σημειωθεί πως τα δεδομένα που εκπέμπονται από τέτοιους πομπούς εντάσσονται στην κατηγορία του Big Data αφού χαρακτηρίζονται από συνέχεια και όγκο. Όλα τα επαγγελματικά πλοία, τα οποία πληρούν συγκεκριμένα χαρακτηριστικά όγκου, βάρους και χρήσης, είναι υποχρεωμένα να φέρουν πομπούς τύπου AIS.



Σχήμα 4.1: Εξοπλισμός AIS

Τελευταίες τεχνολογικές εξελίξεις επιτρέπουν την παρακολούθηση των σημάτων στο διαδίκτυο παρακάμπτοντας έτσι την περιορισμένη εμβέλεια των κεραιών που φέρουν τα πλοία. Ένα παράδειγμα τέτοιου συστήματος φαίνεται στην εικόνα 4.2.



Σχήμα 4.2: Ο χάρτης της MarineTraffic που απεικονίζει σε πραγματικό χρόνο τις κινήσεις των πλοίων

Ένα σήμα τύπου AIS αποτελείται από στατικά και δυναμική δεδομένα. Ως στατικά ή μόνιμα στοιχεία ορίζουμε οποιαδήποτε πληροφορία αφορά στα πλοία και δεν μπορεί να αλλάξει βραχυπρόθεσμα, όπως για παράδειγμα οι αριθμοί MMSI, IMO, το όνομα του πλοίου, τον τύπου του, το μέγεθος του κτλ.

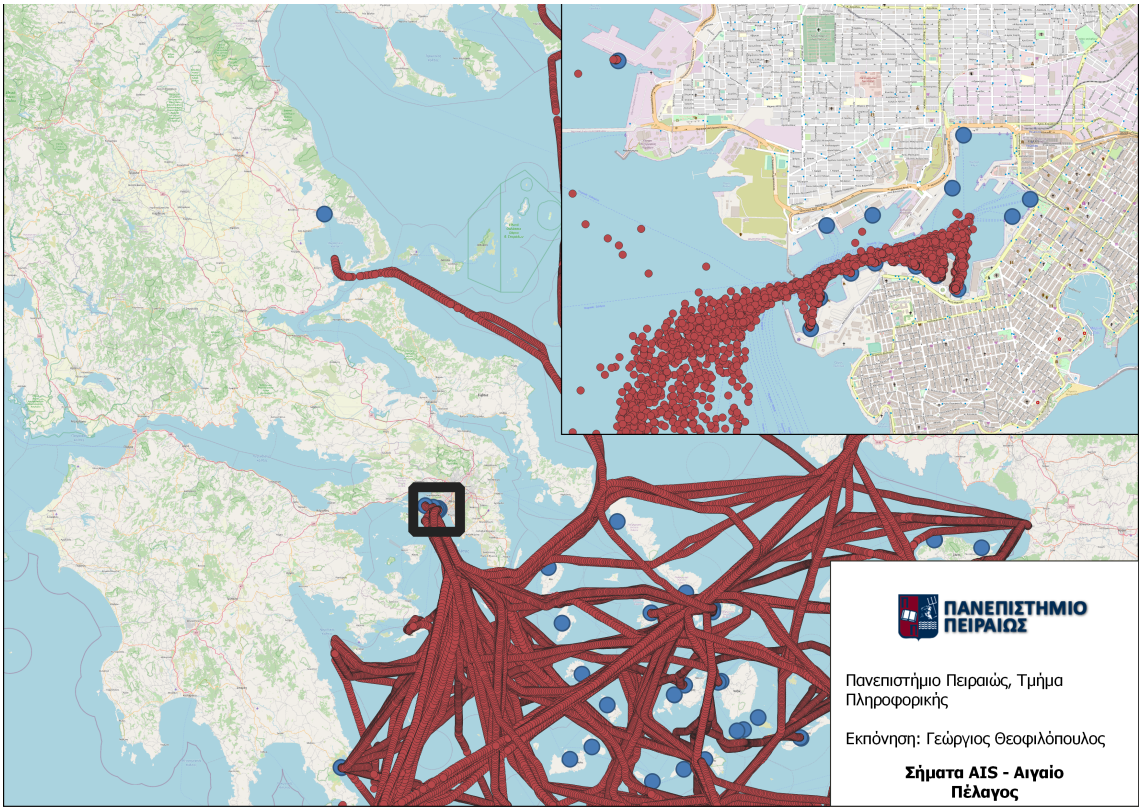
Αντιθέτως τα δυναμικά δεδομένα περιγράφουν την κατάσταση του πλοίου οποιαδήποτε χρονική στιγμή. Αυτά είναι η ταχύτητα του, η πορεία του και ο προσανατολισμός πλώρης του σε μοίρες καθώς και το γεωχωρικό στίγμα του.

Παρατίθεται αναλυτικότερα η πληροφορία που μεταδίδεται.

- Maritime Mobile Service Identity number (MMSI) - a unique identification number for each vessel station (the vessel's flag can also be deducted from it)
- AIS Navigational Status (read more on the subject)
- Rate of Turn - right or left (0 to 720 degrees per minute)
- Speed over Ground - 0 to 102 knots (0.1-knot resolution)
- Position Coordinates (latitude/longitude - up to 0.0001 minutes accuracy)
- Course over Ground - up to 0.1° relative to true north

- Heading - 0 to 359 degrees
- Bearing at own position - 0 to 359 degrees
- UTC seconds - the seconds field of the UTC time when the subject data-packet was generated.

Για την υλοποίηση της παρούσας εργασίας χρησιμοποιήθηκαν δεδομένα AIS του έτους 2018, τα οποία καταγράφουν τις θέσεις των πλοίων κατά την κίνηση τους στο Αιγαίο Πέλαγος.



Σχήμα 4.3: Δεδομένα έτους 2018

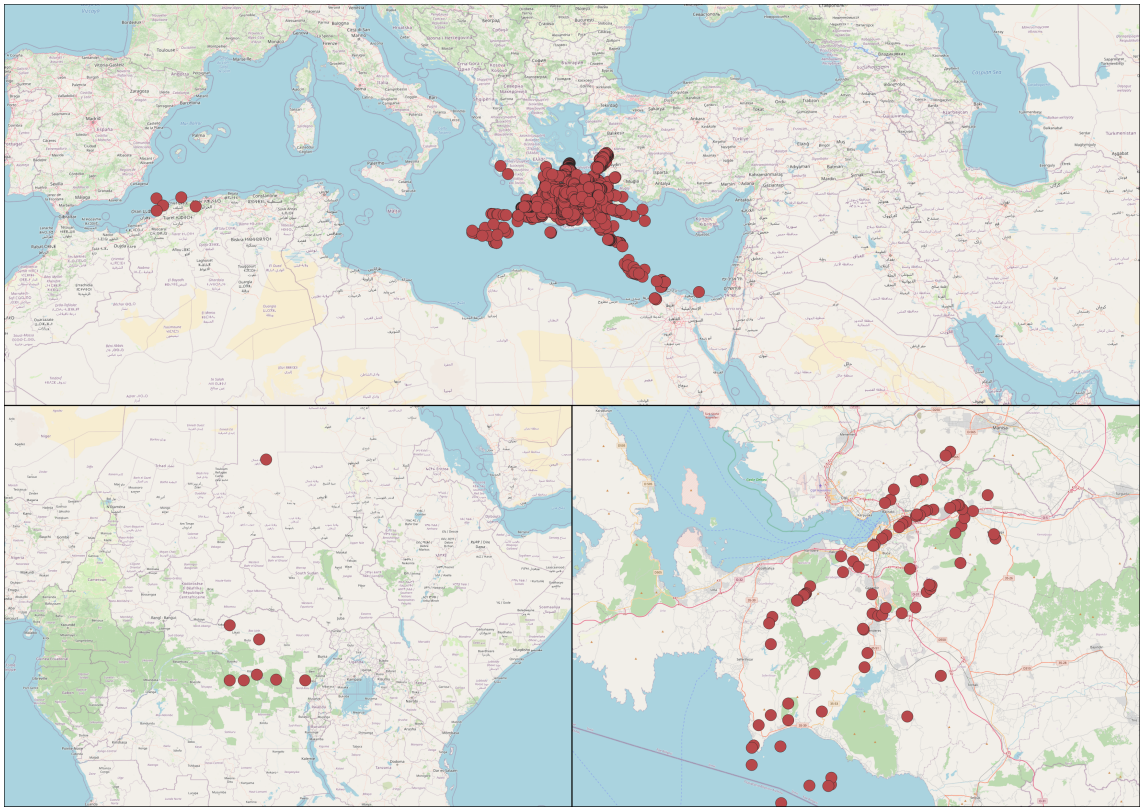
4.2 Θόρυβος και Σφάλματα

Όπως κάθε αναλογικό και ψηφιακό σήμα, έτσι και τα σήματα τύπου AIS είναι θορυβώδεις. Ο θόρυβος αυτός μπορεί να προέρχεται τόσο από τεχνικό σφάλμα των συσκευών μετάδοσης όσο και από ανθρώπινους παράγοντες. Τα τεχνικά σφάλματα οφείλονται σε μεγάλες αποστάσεις μεταξύ πομπού και δέκτη, στην λανθασμένη εγκατάσταση του πομπού καθώς και στην μετεωρολογική κατάσταση στην οποίο βρίσκεται το πλοίο. Σε αυτά συμπεριλαμβάνεται η ανακριβής απόδοση της πραγματικής γεωγραφικής θέσης, καθώς και η μετάδοση ελλιπή δεδομένων, όπως για παράδειγμα η έλλειψη ταχύτητας, πορείας κλπ.

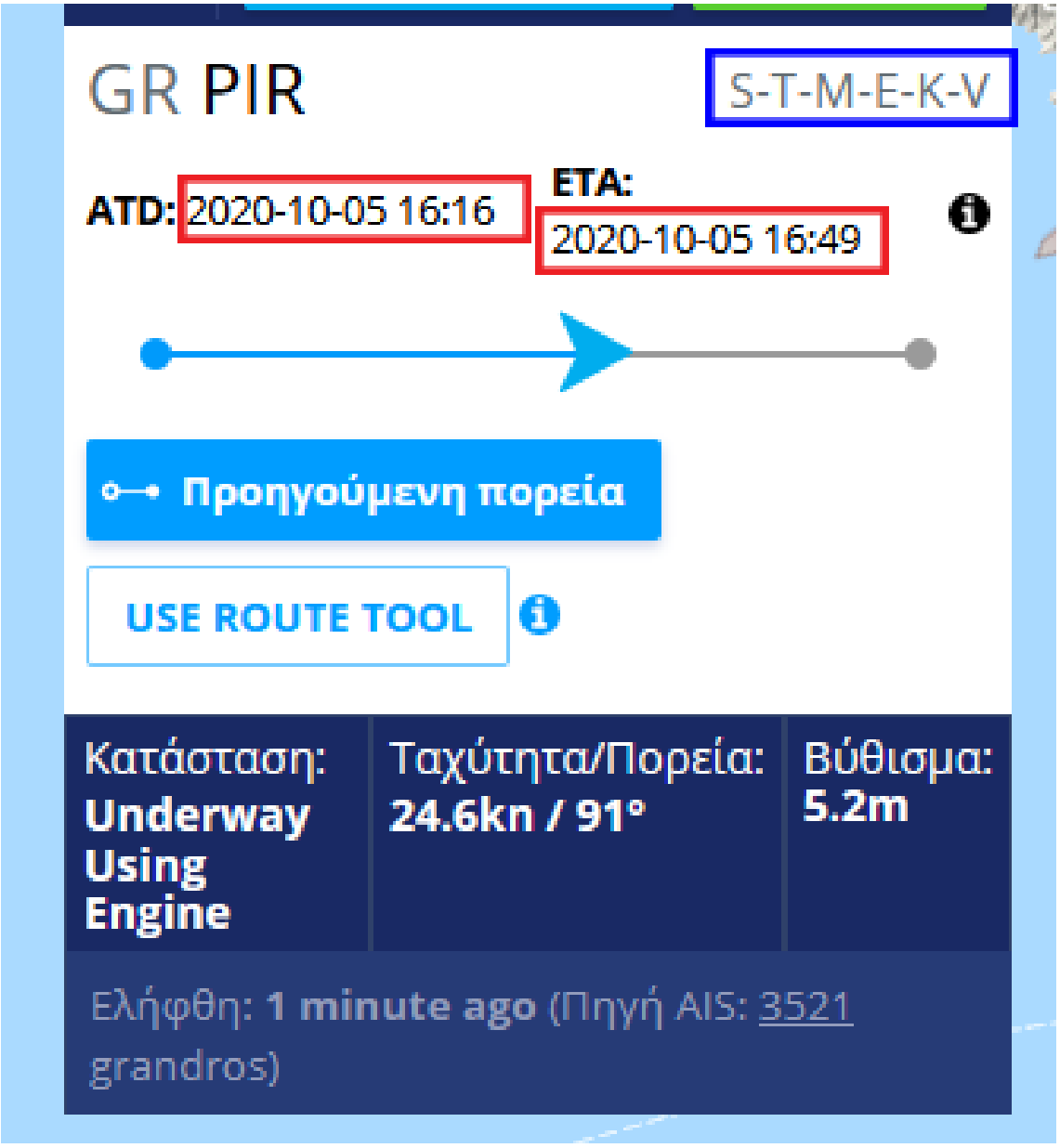
Όσον αφορά τα ανθρώπινα σφάλματα, μπορεί να δημιουργούνται είτε από μη ορθολογική χρήση του συστήματος είτε και να εισάγονται εσκεμμένα. Το 49% του συνολικού θορύβου οφείλεται στην αδυναμία διατύπωσης μιας έγκυρης χρονικής πρόβλεψης αναμενόμενου χρόνου άφιξης(ETA) καθώς και στην λανθασμένη ονομασία του προηγούμενου και επόμενου λιμένα(port calls). Επίσης, συχνά παρατηρούνται σήματα από διαφορετικά πλοία με το ίδιο αναγνωριστικό MMSI, καθώς και εσφαλμένη αναφορά του μεγέθους ή της κατάστασης του

πλοίου. Για παράδειγμα σκάφη με ταχύτητα μεγαλύτερη των 9 κόμβων να αναφέρουν στο πεδίο status ότι είναι αγκυροβολημένα - berthed.

Στις επόμενες εικόνες φαίνεται ένα παράδειγμα φυσικού και ανθρώπινου σφάλματος. Στην πρώτη εικόνα παρατηρούμαι πλοία που φαίνονται να είναι πάνω σε στεριά ενώ στην δεύτερη παρατηρούμαι πως για το προηγούμενο και επόμενο λιμάνι δεν δίνεται η κωδική ονομασία του εκάστοτε λιμένα αλλά μια συμβολοσειρά που συμβολίζει το δρομολόγιο που θα ακολουθήσει το πλοίο. Επιπροσθέτος για ένα τόσο μεγάλο δρομολόγιο δίνεται ETA, 33 λεπτά.



Σχήμα 4.4: Εσφαλμένα σήματα όπου αναφέρουν θέση πάνω σε στεριά

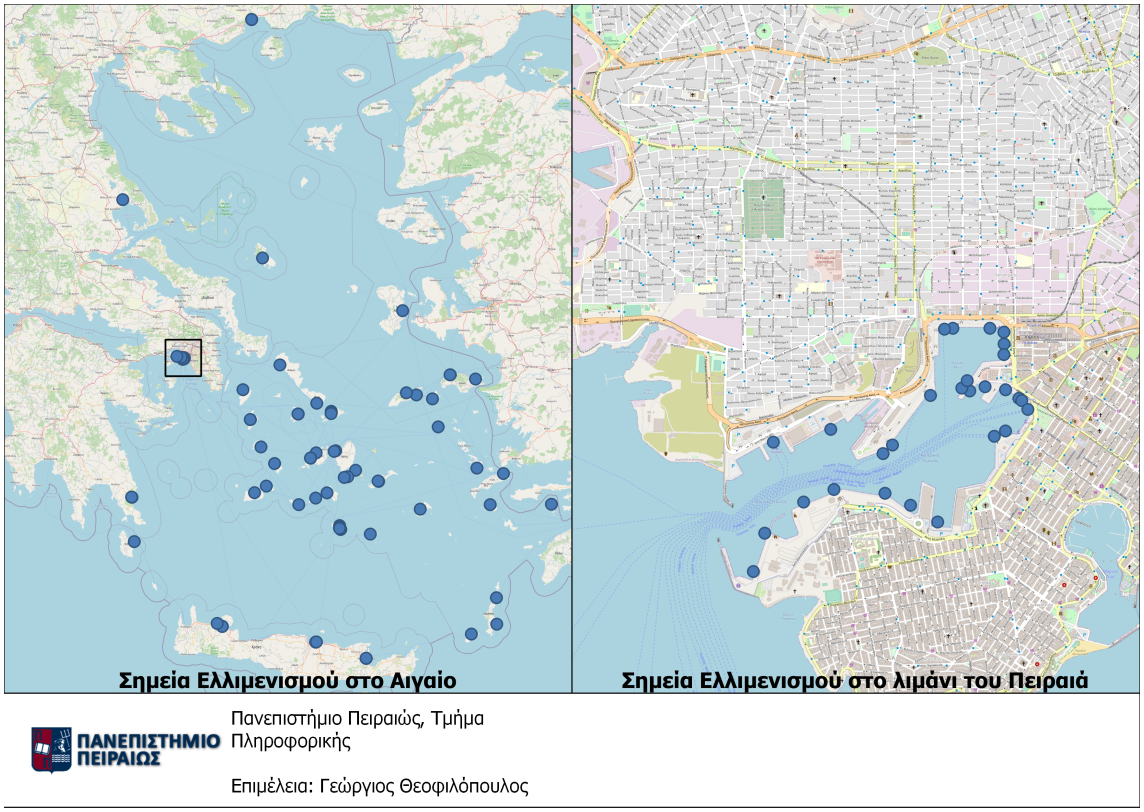


Σχήμα 4.5: Παράδειγμα κακώς αναφερόμενου επόμενου λιμένα και αναμενόμενου χρόνου άφιξης

Για τον λόγο αυτό πρέπει πρέπει να πραγματοποιηθεί κατάλληλη επεξεργασία των δεδομένων, να αφαιρεθούν τυχών outliers καθώς και δεδομένα με λάθος καταχωρημένη πληροφορία, όπως το επόμενο λιμάνι άφιξης, ο προβλεπόμενος χρόνος άφιξης, το όνομα και τα αναγνωριστικά του πλοίου.

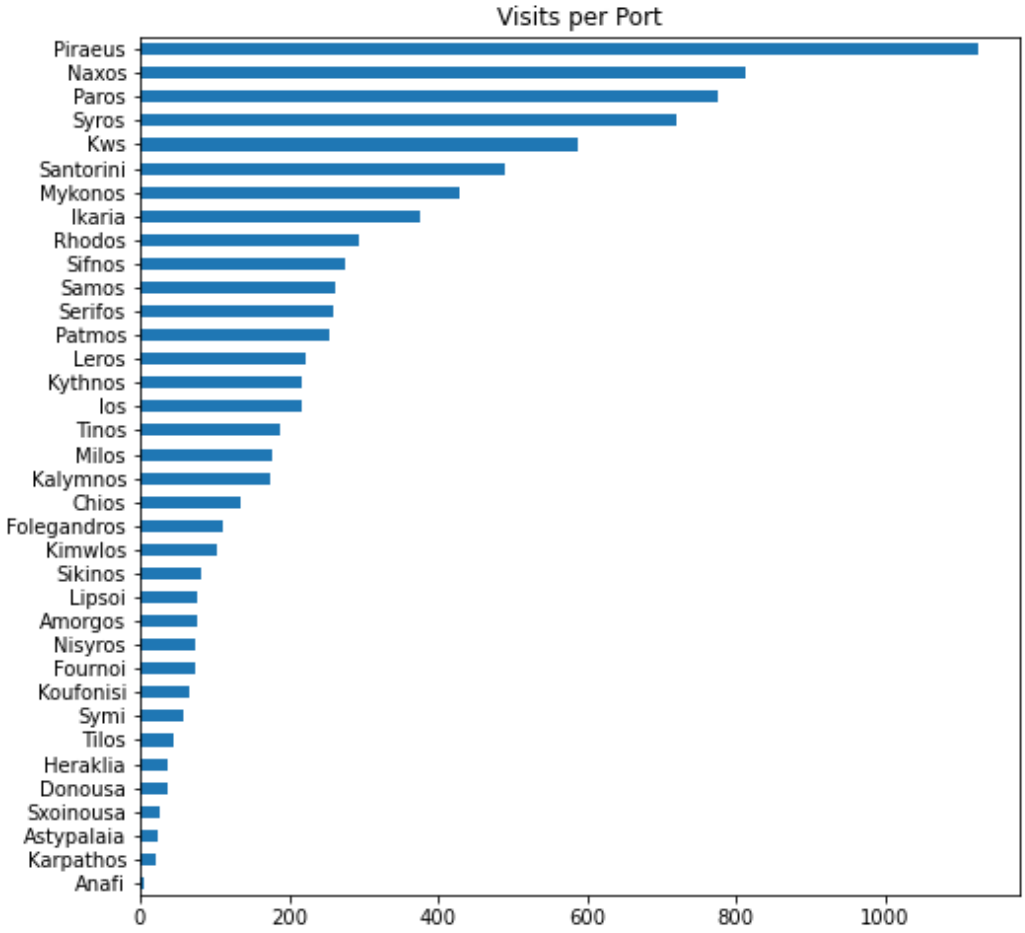
4.3 Λιμένες

Για τη δημιουργία των τροχιών όπως διατυπώθηκε προηγουμένως είναι η απαραίτητη η γνώση των λιμένων στην ευρύτερη περιοχή που κινούνται τα πλοία. Για τους σκοπούς της πτυχιακής εργασίας έγινε καταγραφή 130 λιμένων και προβλυτών. Τα λιμάνια αποθηκεύτηκαν υπό την μορφή γεωμετρικών σημείων στο παγκόσμιο σύστημα αναφοράς WGS84.



Σχήμα 4.6: Καταγεγραμμένα Σημεία Ελλιμενισμού

Σε μεγάλα λιμάνια όπως ο Πειραιάς έγινε περαιτέρω καταγραφή των ακριβών θέσεων πρόσδεσης των πλοίων. Ακολούθως φαίνεται η κατανομή των επισκέψεων για τα λιμάνια.



Σχήμα 4.7: Κατανομή λιμανιών ανά αριθμό επισκέψεων

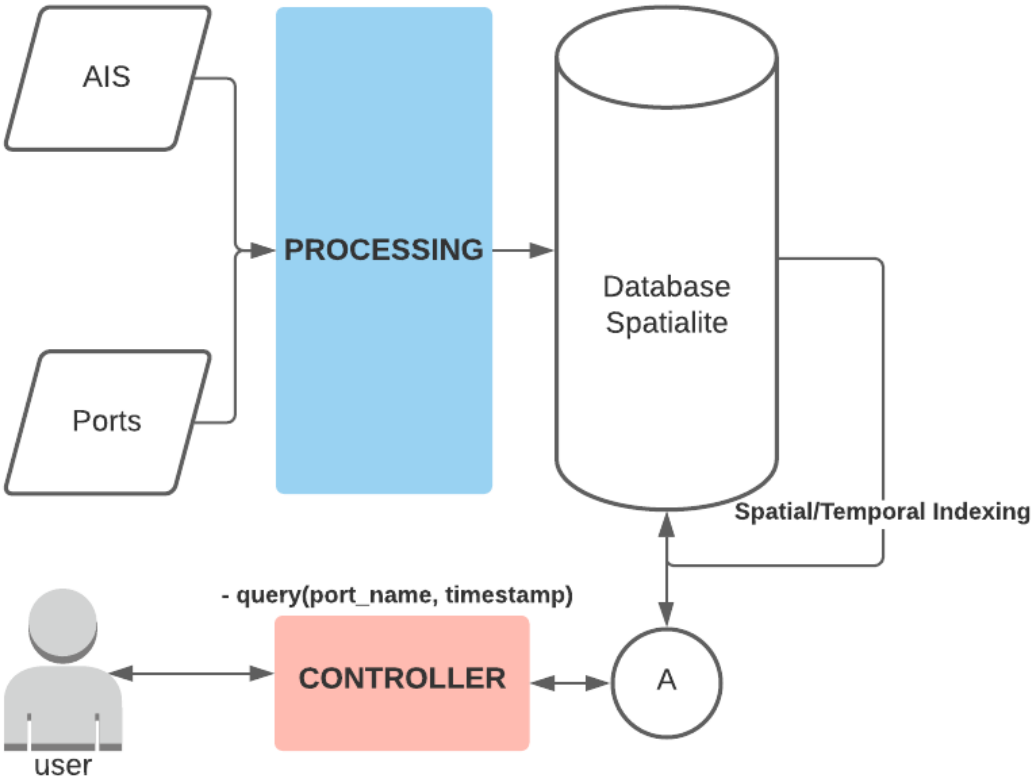
4.4 Κυκλοφοριακός Φόρτος Λιμένα

Το επίπεδο κίνησης κάθε λιμένα είναι ένας πολύ σημαντικός παράγοντας που επηρεάζει τον χρόνο άφιξης. Σε μεγάλα λιμάνια όπου υπάρχουν πολλές, ταυτόχρονες αφίξεις και αναχωρήσεις πλοίων, όπως ο Πειραιάς, τα πλοία είναι υποχρεωμένα να κινούνται σύμφωνα με ορισμένες οδηγίες. Ο λόγος που συμβαίνει αυτό είναι για την αποτροπή θαλάσσιων κίνδυνων. Είναι φανερό λοιπόν πως για να αυξηθούν αποδόσεις των μοντέλων πρόβλεψης του επόμενου κεφαλαίου πρέπει να συμπεριληφθεί το επίπεδο φόρτου του λιμένα άφιξης αφού επηρεάζει άμεσα τον τρόπο με τον οποίο κινείται το κάθε πλοίο σε κοντινές αποστάσεις από τα λιμάνια.

Για να πραγματοποιηθεί η αποτίμηση της κατάστασης του εκάστοτε λιμένα, αρχικά έγινε προεπεξεργασία του σύνολο των AIS δεδομένων και επιλέχτηκαν μόνο τα σήματα με ταχύτητα μικρότερη των 15 κόμβων. Η επιλογή αυτή έγινε διότι ταχύτητες μεγαλύτερες από τους 15 κόμβους παρατηρούνται μόνο στην ανοιχτή θάλασσα και όχι σε ένα κλειστό περιβάλλον όπως ένα λιμάνι. Έπειτα, στρογγυλοποιήσαμε τη στήλη χρόνου κάθε σήματος στο κοντινότερο

τέταρτο της ώρας. Με αυτόν τον τρόπο σήματα που για παράδειγμα εντοπίστηκαν στο 15ο λεπτό με αυτά που εντοπίστηκαν σε 5ο ή στο 10ο μπαίνουν στην ίδια κατηγορία. Ο λόγος που δεν επιλέχτηκε κάποιο μικρότερο χρονικό διάστημα, π.χ. ανά λεπτό είναι επειδή τα πλοία δεν εκπέμπουν με σταθερό ρυθμό. Κάποια πλοία εκπέμπουν ανά λεπτό άλλα ανά 3,4,5,10 λεπτά, οπότε ένα μικρότερο κατώφλι δεν θα συμπεριλάμβανε τον πραγματικό αριθμό.

Ο σχεδιασμός του συστήματος έγινε με γνώμονα την ταχύτητα των επερωτήσεων 500.000, πράγμα που μας ανάγκασε να απορρίψουμε κάποια naive λύση που θα χειριζόντανε αρχεία τύπου csv. Τα προεπεξεργασμένα δεδομένα που αναφέραμε προηγουμένως, φορτώνονται σε μια βάση δεδομένων Spatialite με το σύνολο το δεδομένων να τοποθετείται σε διαφορετικά μηνιαία partitions. Τελικά με την χρήση δύο διαφορετικών ευρετηρίων, ένα για την στήλη του χρόνου και ένα χωρικό ευρετήριο τύπου Rtree καταφέραμε να υπολογίζουμε τον αριθμό των πλοίων σε κάθε λιμάνι σε βιώσιμο χρόνο. Πρέπει να αναφέρουμε πως τα AIS δεδομένων τοποθετήθηκαν με μια επιπλέον στήλη τύπου POINT GEOMETRY όπως επίσης και τα λιμάνια με την μόνη διαφορά ότι γύρω από κάθε σημείο κατασκευάσαμε μια ζώνη(buffer) με διάμετρο 1χλμ ώστε να μπορούμε να πραγματοποιήσουμε τις αναζητήσεις μεταξύ των δύο γεωμετριών.



Σχήμα 4.8: Σύστημα Υπολογισμού Φόρτου

4.5 Ημερολόγιο Αφίξεων Κρουαζιερόπλοιων Σαντορίνης

Τα νευρωνικά δίκτυα που θα υλοποιηθούν στην επόμενη ενότητα θα χρησιμοποιηθούν για πρόβλεψη στα ιστορικά δεδομένα των κρουαζιερόπλοιων που επισκέφτηκαν το νησί της Σαντορίνης, το έτος 2018. Τα αντίστοιχα δεδομένα γίνανε εξόρυξη από τον επίσημο οργανισμό του λιμένα Σαντορίνης. Για κάθε επίσκεψη περιλαμβάνεται το όνομα του πλοίου καθώς και η

ώρα άφιξης και αναχώρισης.

	date	name	line	ship's max passenger capacity	arrival time	arrival epoch	departure time	departure epoch
0	2017-06-01	WIND STAR	WINDSTAR CRUISES	168	07:00:00	1496300400	21:00:00	1496350800
1	2017-06-01	MS WESTERDAM	HOLLAND AMERICA LINE	1916	08:00:00	1496304000	17:00:00	1496336400
2	2017-06-01	CELESTYAL OLYMPIA	CELESTYAL CRUISES	1450	15:30:00	1496331000	20:30:00	1496349000
3	2017-06-03	HORIZON	CROISIERES DE FRANCE	1828	07:00:00	1496473200	17:00:00	1496509200
4	2017-06-03	NORWEGIAN SPIRIT	NORWEGIAN CRUISE LINE	2018	08:00:00	1496476800	17:00:00	1496509200
...	...	...	...	...	...	...	...	...
433	2018-09-27	CELESTYAL CRYSTAL	CELESTYAL CRUISES	1200	09:00:00	1538038800	23:00:00	1538089200
434	2018-09-27	CELESTYAL OLYMPIA	CELESTYAL CRUISES	1450	16:45:00	1538066700	21:30:00	1538083800
435	2018-09-28	GEMINI	ETS TUR	1074	10:30:00	1538130600	20:00:00	1538164800
436	2018-09-29	HORIZON	PULLMANTUR CRUISES	1828	09:00:00	1538211600	18:00:00	1538244000
437	2018-09-30	CELESTYAL OLYMPIA	CELESTYAL CRUISES	1450	16:45:00	1538325900	21:30:00	1538343000

Σχήμα 4.9: Ημερολόγιο Επισκέψεων Σαντορίνης

4.6 Δεδομένα Καιρού

Η καιρική κατάσταση της περιοχής όπου κινείται το πλοίο μπορεί να έχει άμεση επιρροή στον χρόνο άφιξης του. Δυνατοί άνεμοι, βροχές και γενικότερα έντονα καιρικά φαινόμενα τείνουν να δυσκολεύουν την κίνηση του πλοίου. Στο διαδίκτυο υπάρχουν πολλές υπηρεσίες όπου προσφέρουν ιστορικά δεδομένα καιρού για κάθε δοσμένη γεωγραφική θέση αλλά οι πιο πολλές από αυτές είτε είναι επί πληρωμή είτε έχουν περιορισμό στα api calls που μπορούν να γίνουν.

Στο πλαίσιο της εργασίας ανακτήθηκαν δεδομένα από την ελληνική ιστοσελίδα Meteo Search που με τους 450 μετεωρολογικούς σταθμούς που μας παρείχαν μπορέσαμε να καλύψουμε ένα ικανοποιητικό εύρος τοποθεσιών. Τα ιστορικά δεδομένα του meteo.gr δεν παρέχονται σε επίπεδο ώρας αλλά σε επίπεδο ημέρας. Ενδεικτικά δίνεται η μέγιστη, ελάχιστη, μέση θερμοκρασία, τα επίπεδα του ανέμου και της βροχής.

MONTHLY CLIMATOLOGICAL SUMMARY for APR. 2020

NAME: amfikliatown CITY: STATE:
ELEV: 440 m LAT: LONG:

TEMPERATURE (°C), RAIN (mm), WIND SPEED (km/hr)

DAY	MEAN TEMP	HIGH	TIME	LOW	TIME	HEAT DEG DAYS	COOL DEG DAYS	RAIN	AVG WIND SPEED	HIGH	TIME	DOM DIR
1	9.3	11.0	16:10	7.6	7:40	9.1	0.0	6.4	4.7	33.8	15:50	ESE
2	8.3	13.3	16:50	4.2	7:10	10.0	0.0	0.0	6.3	29.0	14:20	NW
3	6.1	8.5	12:50	3.4	4:20	12.2	0.0	7.2	2.6	20.9	12:50	ESE
4	6.7	7.6	15:50	5.4	19:40	10.8	0.0	49.8	3.9	32.2	20:10	ESE
5	7.7	8.8	15:30	6.6	1:20	10.7	0.0	25.0	6.6	35.4	4:50	W
6	6.1	7.1	0:20	5.3	17:40	12.3	0.0	7.4	3.4	25.7	1:20	SE
7	7.2	10.7	17:50	4.0	8:10	11.2	0.0	2.2	3.5	27.4	14:30	SE
8	9.6	13.6	15:40	5.6	7:50	8.7	0.0	0.0	4.8	25.7	13:30	SE
9	11.9	17.2	14:30	5.4	6:30	6.4	0.0	0.0	4.3	25.7	15:40	SW
10	14.4	21.6	16:30	8.4	4:10	4.4	0.6	0.0	2.3	19.3	16:10	SW
11	16.4	22.8	14:40	10.4	6:50	3.1	1.1	0.0	3.2	19.3	11:00	SW
12	15.4	20.5	15:00	10.7	6:50	3.3	0.4	0.0	4.0	25.7	15:10	SW
13	15.5	21.4	16:50	9.4	7:10	3.4	0.6	0.0	3.9	22.5	14:00	SW
14	16.8	22.1	17:20	11.1	5:20	2.4	0.9	0.0	4.2	33.8	15:00	SW
15	10.6	13.9	1:20	7.0	21:50	7.8	0.0	1.2	4.0	33.8	10:50	SW
16	10.9	16.6	17:10	6.6	6:50	7.4	0.0	0.0	4.5	20.9	17:20	SW
17	14.7	22.6	16:20	7.2	4:40	4.5	0.9	0.0	4.0	22.5	13:40	SW
18	18.1	24.8	15:50	11.6	7:00	2.2	2.0	0.0	4.7	25.7	13:50	SW
19	19.3	25.2	16:50	13.3	7:00	1.3	2.3	0.0	2.9	17.7	14:30	SW
20	19.1	24.6	14:00	14.4	00:00	0.9	1.7	0.0	3.1	19.3	12:10	SW
21	10.2	14.6	0:50	8.0	21:00	8.2	0.0	6.6	0.6	14.5	5:20	N
22	7.5	8.4	17:30	4.9	23:30	10.8	0.0	10.4	1.8	12.9	10:10	ESE
23	7.8	11.6	13:30	4.9	5:00	10.6	0.0	3.0	2.6	25.7	13:20	SW
24	11.6	17.4	16:10	5.2	1:50	6.8	0.0	0.0	3.4	19.3	17:00	SW
25	14.9	21.8	15:30	7.7	5:30	4.2	0.8	0.0	2.7	20.9	16:50	SW
26	15.8	21.4	16:30	10.5	6:30	3.1	0.6	0.0	4.2	24.1	14:10	SW
27	16.7	22.3	14:40	12.3	4:10	2.3	0.7	0.0	3.1	25.7	14:00	SW
28	17.8	23.4	14:20	12.2	6:20	2.0	1.4	0.0	3.7	22.5	14:50	SW
29	17.8	22.7	16:30	13.2	4:00	1.7	1.2	0.0	3.7	27.4	19:00	SW
30	17.7	23.2	14:00	12.6	5:20	1.9	1.3	1.8	3.1	24.1	20:30	SW
	12.7	25.2	19	3.4	3	183.8	16.5	121.0	3.7	35.4	5	SW

Max >= 32.0: 0
Max <= 0.0: 0
Min <= 0.0: 0
Min <= -18.0: 0
Max Rain: 49.81 ON 04/04/20
Days of Rain: 11 (> .2 mm) 9 (> 2 mm) 2 (> 20 mm)
Heat Base: 18.3 Cool Base: 18.3 Method: Integration

Σχήμα 4.10: Δεδομένα από τον μετεωρολογικό σταθμό της Αμφίκλειας

Κεφάλαιο 5

Προεπεξεργασία Δεδομένων

Τα δεδομένα αποτελούνται από 36.808.679 εγγραφές για το έτος 2018 από τους μήνες Μάιο μέχρι και Αύγουστο. Στο σύνολο δεδομένων υπάρχουν εγγραφές οι οποίες δεν έχουν ενδιαφέρον στο πλαίσιο της πτυχιακής εργασίας. Σήματα από αλιευτικά, ιδιωτικά σκάφη, φορτηγά πλοία, πλοία υποβοήθησης εργασιών καθώς και σήματα από φάρους αφαιρούνται. Οι αλγόριθμοι μηχανικής μάθησης θα εφαρμοστούν σε επιβατικά πλοία που πραγματοποιούν καθημερινά δρομολόγια καθώς και σε κρουαζιερόπλοια.

	mmsi	timestamp	lon	lat	speed	course	heading
13541195	240807000	2018-05-01 00:00:01	24.646931	37.267551	10.2	323	511.0
13541192	538003876	2018-05-01 00:00:01	28.846470	40.743580	0.6	145	105.0
13541196	271040024	2018-05-01 00:00:01	28.346340	40.729568	11.8	45	43.0
13541194	229001000	2018-05-01 00:00:01	24.462740	37.511211	9.2	86	86.0
13541193	273343200	2018-05-01 00:00:01	26.088890	37.397678	6.4	313	313.0
...	...	...	...	...	...	...	...
29206582	237001000	2018-08-31 23:59:58	24.011999	38.023819	0.0	261	91.0
29206580	374171000	2018-08-31 23:59:58	28.760389	40.785770	0.9	290	511.0
29206581	271062144	2018-08-31 23:59:58	26.677130	38.211361	0.0	0	511.0
29206583	271000025	2018-08-31 23:59:59	26.916000	38.764999	0.0	29	154.0
29206584	616999019	2018-08-31 23:59:59	27.542339	40.955791	0.0	0	511.0

36808679 rows × 7 columns

Σχήμα 5.1: Σύνολο Δεδομένων

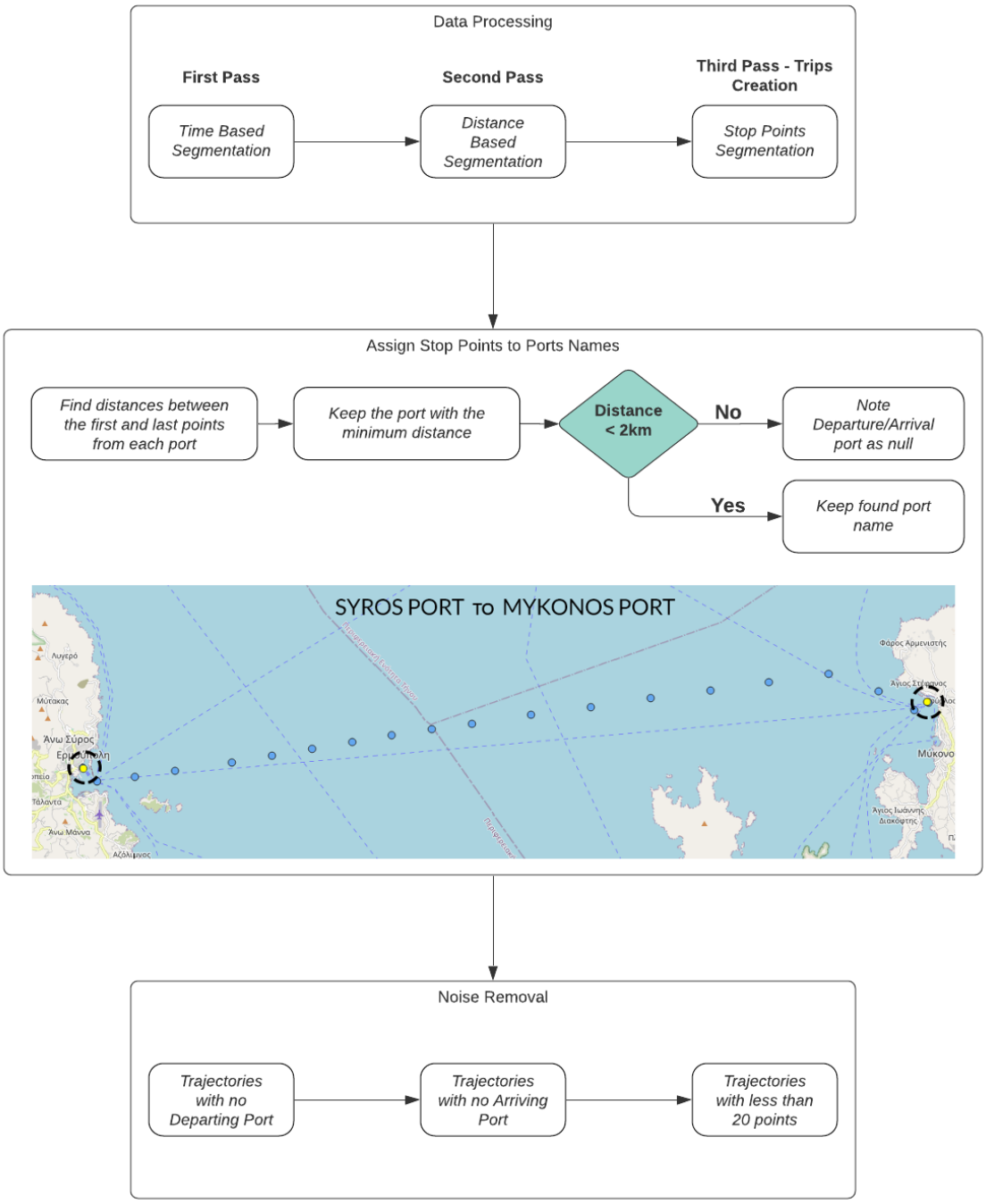
5.1 Καθαρισμός Δεδομένων

Όπως αναφέρθηκε και σε προηγούμενη ενότητα τα AIS δεδομένα εμπεριέχουν συχνά θόρυβο. Για τον λόγο αυτόν, αφαιρέθηκε οποιοσδήποτε θόρυβος οφείλεται σε ανθρώπινο παράγοντα όπως σήματα με λανθασμένο αναγνωριστικό αριθμό MMSI. Στη συνέχεια υλοποιώντας τους αντίστοιχους αλγορίθμους αφαιρέθηκαν τα σήματα τα οποία συνοδευόντουσαν από λανθασμένη

δυναμική πληροφορία.

Συγκεκριμένα, εντοπίστηκαν τα σήματα με τιμές άνω των 360 μοιρών στις στήλες heading, course, σήματα με ταχύτητα μεγαλύτερη των 100 κόμβων καθώς και σημεία τα οποία βρισκόντουσαν πάνω στην στεριά. Για τον εντοπισμό των τελευταίων, υπολογίστηκε η χωρική και χρονική απόσταση μεταξύ διαδοχικών σημείων του ίδιου πλοίου με τελικό στόχο τον υπολογισμό της πραγματικής ταχύτητας του σκάφους. Εγγραφές με ταχύτητα μεγαλύτερη των 100 κόμβων απορρίφθηκαν.

Από την στιγμή που τα δεδομένα βρίσκονται σε αξιοποιήσιμη μορφή, πρέπει να διαταχθούν σε χρονικά ταξινομημένες τροχιές με αρχή και τέλος πάντα ένα λιμάνι. Η γενική μεθοδολογία η οποία θα αναλυθεί εκτενώς στην επόμενη υποενότητα είναι τα διάσπαρτα σήματα να ομαδοποιούνται σε πρώτη φάση με βάση ένα χρονικό κατώφλι, έπειτα με ένα χωρικό και τέλος με βάση τις αποστάσεις από τα καταγεγραμμένα ελληνικά λιμάνια. Η προσέγγιση που θα αναλυθεί στις επόμενες ενότητες φαίνεται και στην κάτωθι εικόνα.



Σχήμα 5.2: Διαδικασία Παραγωγής Τροχιών

5.2 Χωρική και Χρονική Διαμέριση

Τα διάσπαρτα σημεία του συνόλου δεδομένων D μπορούν να περιγραφούν ως τριπλέτες της μορφής $t^s(k), lon^s(k), lat^s(k)$, και s το μοναδικό αναγνωριστικό του κάθε πλοίου. Ο αριθμός k δεν αφορά κάποια πραγματική χρονική στιγμή αλλά την σειρά με την οποία παρουσιάστηκαν τα αντίστοιχα σήματα, δηλαδή είναι ένας απλός αριθμητικός δείκτης. Στόχος είναι να ομαδοποιηθούν τα δεδομένα αυτά σε τροχιές, με αρχή και τέλος κάποιο λιμάνι. Έτσι θα προκύψει το σύνολο $D' = \{trajectory_1, trajectory_2, \ldots, trajectory_i\}$ με κάθε τροχιά να είναι ένα σύνολο

διανυσμάτων της μορφής

$$trajectory_i = \begin{bmatrix} mmsi & lon_1 & lat_1 & t_1 & port1 & port2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \\ mmsi & lon_j & lat_j & t_j & port1 & port2 \end{bmatrix}$$

Για την κατασκευή του συνόλου D' εμπλουτίζουμε τις τριπλέτες $\{t^s(k), lon^s(k), lat^s(k)\}$ με δύο στήλες που αναφέρονται στην χρονική και χωρική απόσταση μεταξύ των διαδοχικών σημείων για κάθε πλοίο s. Οι στήλες αυτές είναι οι

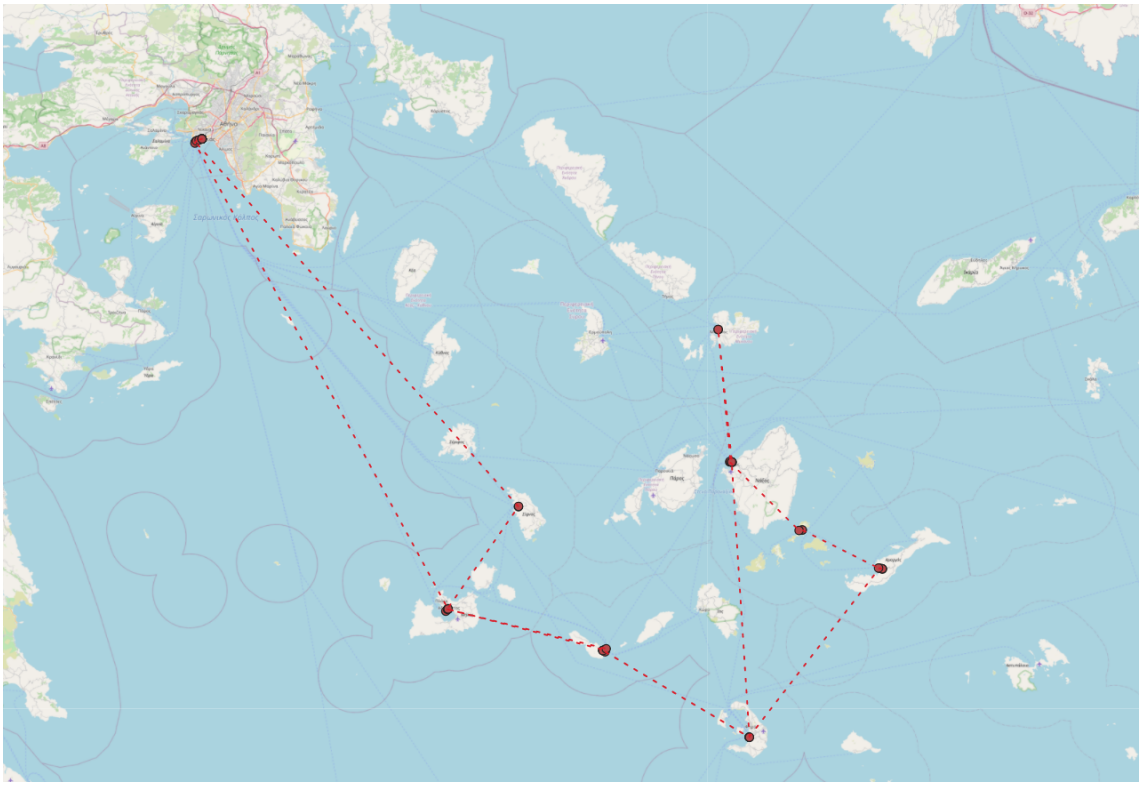
$$\Delta t^s(k) = t^s(k) - t^s(k-1)$$
$$\Delta d^s(k) = haversine(lon^s(k), lat^s(k), lon^s(k-1), lat^s(k-1))$$

Έπειτα πραγματοποιούμε το πρώτο χώρισμα των σημείων σε τροχιές με χρονικό πλάτος 24 ωρών. Η χρονική διάρκεια της τροχιάς μέχρι το σημείο k ισούται με το άθροισμα $\sum_{n=1}^{k} \Delta t^s(n)$ και αναζητούμε το τελευταίο σημείο k μέχρι το οποίο να ισχυεί $\sum_{n=1}^{k} \Delta t^s(n) < 24$. Επιλέξαμε το παράθυρο των 24 ωρών διότι μας επιτρέπεται να δημιουργήσουμε τροχιές οι οποίες εμπεριέχουν συνεχόμενες επισκέψεις σε λιμάνια.

Σε δεύτερη φάση πραγματοποιούμε ένα επιπλέον χωρικό χώρισμα στις τροχιές οι οποίες δημιουργήθηκαν από το προηγούμενο βήμα. Συγκεκριμένα, διαιρούμε τις τροχιές που έχουν κενά μεγαλύτερα των 20 μιλίων μεταξύ διαδοχικών σημείων. Η διαδικασία αυτή επιλέχθηκε ώστε οι τελικές παραγόμενες τροχιές να έχουν χωρική και χρονική ομοιομορφία. Στην συγκεκριμένη περίπτωση αναζητούμε τα σημεία k για το οποίο ισχύει $\Delta d^s(n) > 20$. Έτσι έχουμε το διαταγμένο σύνολο:

$$P = \{secondPassTrajectory_1, secondPassTrajectory_2, \ldots, secondPassTrajectory_i\}$$

που κάθε διαφορετικό στοιχείο έχει είτε χρονική απόσταση μεγαλύτερη των 24 ωρών είτε χωρική απόσταση 20 μίλια από τα γειτονικά του στοιχεία είτε απλά αναφέρεται σε διαφορετικό πλοίο. Στο συγκεκριμένο σύνολο εφαρμόζεται μια πράξη φιλτραρίσματος ώστε να αφαιρεθούν τροχιές με χρονική διάρκεια μικρότερη των 15 λεπτών καθώς και συνολικής διανυόμενης απόστασης μικρότερης των 2 μιλιών Στο επόμενο βήμα οι προσωρινές τροχιές που έχουν παραχθεί μέχρι στιγμής χωρίζονται περαιτέρω με βάση τα στάσιμα σημεία τους. Έτσι, αν για παράδειγμα είχαμε μια τροχιά των 24 ωρών και χωρίς κενά, αυτή η τροχιά θα διαχωριστεί σε n διαφορετικές τροχιές όπου n ο αριθμός των λιμανιών που επισκέφθηκε το πλοίο.



Σχήμα 5.3: Τα παραπάνω σημεία και κατ΄ επέκταση οι προβληματικές - προσωρινές υποτροχιές που παράγονται από αυτά είναι μια κλάση παραδειγμάτων που αφαιρούνται με τους προηγούμενους μετασχηματισμούς.

5.3 Εξαγωγή Στάσιμων Σημείων και Οριοθέτηση Τροχιών

Για την εξαγωγή των στάσιμων σημείων, κατασκευάσαμε τον αλγόριθμο stop points based-segmentation. Ο αλγόριθμος δέχεται για ορίσματα τρεις διαφορετικές τιμές time threshold, distance threshold, speed threshold όπου αντιπροσωπεύουν το κατώφλι του αντίστοιχου φυσικού μεγέθους, δηλαδή το χρονικό, χωρικό και το κατώφλι ταχύτητας αντίστοιχα.

Για την εύρεση των πιθανών σημείων που αφορούν την άφιξη πλοίου σε ένα λιμάνι, αρχικά ο αλγόριθμος εντοπίζει τα σημεία της κάθε τροχιάς του συνόλου P τα οποία έχουν ταχύτητα μικρότερη από το δοσμένο κατώφλι ταχύτητας. Στην περίπτωση μας όταν ισχύει $speed^s(k) < speed_threshold = 1kn$. Τα σημεία αυτά ονομάζεται πιθανά στάσιμα σημεία και τοποθετούνται στο σύνολο:

$$candidate_stop_points = \{(t^s(k), lon^s(k), lat^s(k)) : speed^s(k) < speed_threshold\}$$

Έπειτα γύρω από κάθε υποψήφιο σημείο θεωρούμε πως δημιουργείται ένας κύκλος με κέντρο το ίδιο το σημείο, $(xt^s(k), lon^s(k), lat^s(k))$ και ακτίνα η απόσταση που έχω δοθεί στην παράμετρο distance threshold. Τα σημεία που εξετάζονται είναι χρονικά διατεταγμένα, έτσι τα προηγούμενα σημεία από το κέντρο(σύμφωνα με το πρωτεύων κλειδί του αρχείου) αφορούν χρόνο που έχει προηγηθεί ενώ τα επόμενα έχουν παρατηρηθεί σε μεταγενέστερο χρόνο. Με γνώμονα το κέντρο πραγματοποιούμε μια αναζήτηση προς τα πίσω και μια προς τα μπροστά μέχρι να εντοπίσουμε δύο σημεία x_1, x_2, ένα για κάθε κατεύθυνση ώστε να ισχύει:

$$\max dist(x_1, x_c) \wedge dist(x_1, x_c) < distance\ threshold$$

$$\max dist(x_1, x_c) \wedge dist(x_c, x_2) < distance\ threshold$$

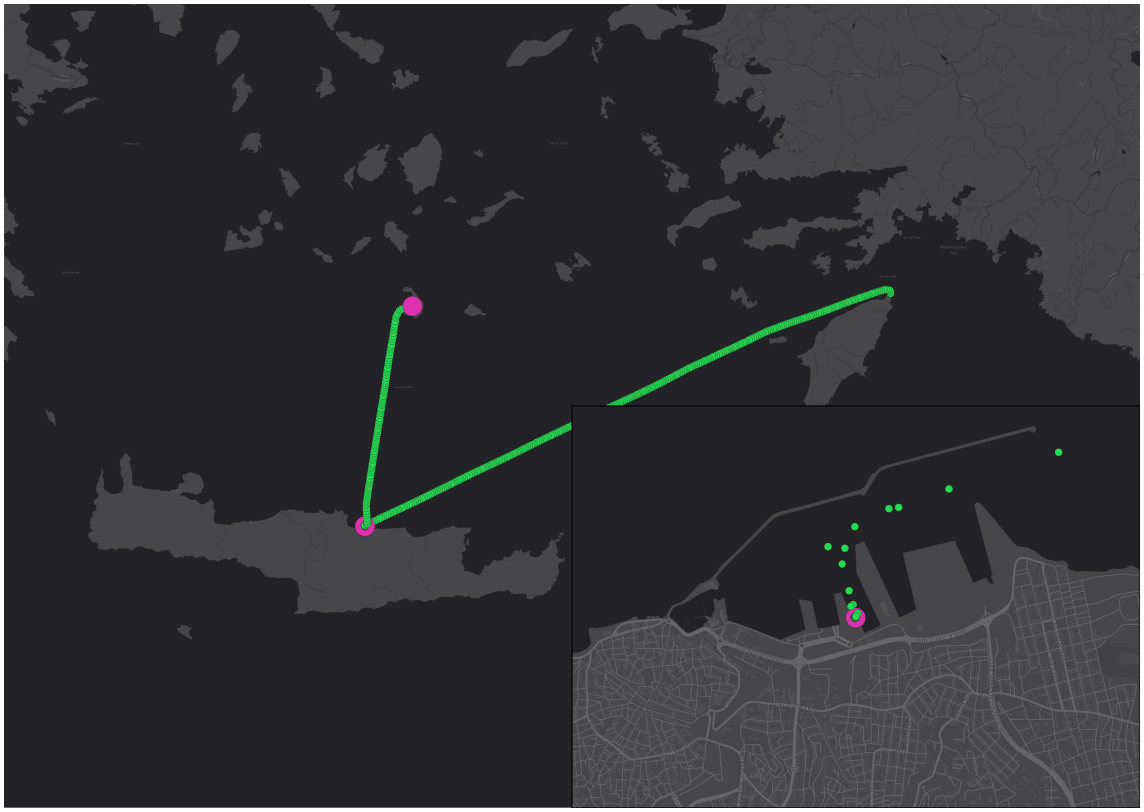
Αν το πλοίο έχει διανύσει την τροχιά $x_1, \ldots, x_c, \ldots, x_2$ σε χρόνο μεγαλύτερο από το δοσμένο χρονικό κατώφλι, αν δηλαδή $x_2(t) - x_1(t) > time_threshold$ τότε το υποψήφιο σημείο θεωρείται στάσιμο σημείο και σηματοδοτεί το τέλος κάθε επιμέρους τμήματος των τροχιών του συνόλου P.

5.4 Εύρεση Λιμένα Προέλευσης και Προορισμού

Οι τελικές τροχιές που έχουν δημιουργηθεί οριοθετούνται από ένα στάσιμο σημείο. Οι στάσεις των πλοίων μπορεί να οφείλονται τόσο στην προσωρινή στάση των πλοίων στην ανοιχτή θάλασσα στα μέσα της διαδρομής τους ή σε έναν ελλιμενισμό μακριά ή κοντά σε κάποιο λιμάνι. Στο πλαίσιο της πτυχιακής το ενδιαφέρον στρέφεται μόνο σε στάσεις που γίνονται κοντά στα καταγεγραμμένα λιμάνια διότι η πρόβλεψη που θα πραγματοποιηθεί στην επόμενη ενότητα θα γίνεται από λιμάνι σε λιμάνι και συγκεκριμένα σε κρουαζιερόπλοια και πλοία γραμμής.

Έτσι, αφού όλα τα σημεία έχουν τοποθετηθεί σε τροχιές με βάση τα στάσιμα σημεία, στόχος μας είναι ο συσχετισμός αυτών με τα ονόματα των κοντινών λιμένων. Το σημείο αναχώρησης θεωρείται το πρώτο σημείο της τροχιάς ενώ το τελευταίο σημείο το σημείο άφιξης. Εκμεταλλευόμενοι αυτή την ιδιότητα υλοποιήσαμε δυο συναρτήσεις find-arrival, find-depature όπου εμπλουτίζουν τις υπάρχουσες τροχιές με τα ονόματα των λιμένων αναχώρησης και άφιξης.

Οι διαδικασίες υπολογίζουν την απόσταση των οριακών σημείων και των διαθέσιμων λιμένων που έχουμε καταγράψει. Επί της ουσίας πραγματοποιούν την πράξη full join μεταξύ των οριακών σημείων και όλων των καταγεγραμμένων λιμανιών. Αν η υπολογισμένη απόσταση των σημείων με το κοντινότερο λιμάνι είναι μικρότερη ή ίση των 2 χλμ τότε η τροχιά εισάγεται στο τελικό σύνολο αλλιώς απορρίπτεται. Παρακάτω παρουσιάζεται ενδεικτικά μια τροχιά από λιμάνι σε λιμάνι καθώς και μια διαδρομή συνεχόμενων επισκέψεων σε διαφορετικά λιμάνια.



Σχήμα 5.4: Ενδεικτική τροχιά πλοίου με αφετηρία το λιμάνι της Ρόδου και προορισμό το λιμάνι του Ηρακλείου. Με ροζ χρώμα διακρίνονται τα στάσιμα σημεία της τροχιάς.

Σχήμα 5.5: Ολόκληρη η διαδρομή που ακολούθησε το αναφερόμενο πλοίο εντός της διάρκειας μιας ημέρας.

Κεφάλαιο 6

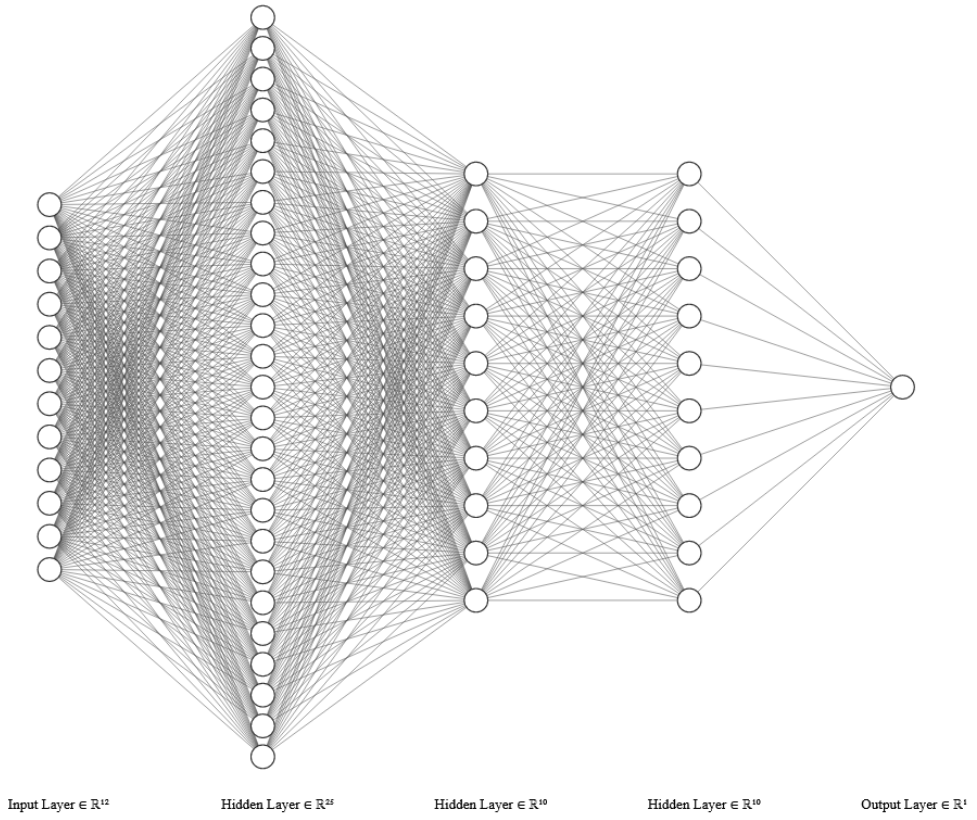
Πρόβλεψη από Λιμένα σε Λιμένα

Τα δύο πρώτα μοντέλα χρησιμοποιούνται για την πρόβλεψη του αναμενόμενου χρόνου άφιξης για μια τροχιά την φορά, δηλαδή από ένα λιμάνι σε ένα μόνο επόμενο. Για την προσέγγιση του συγκεκριμένου προβλήματος προτείνουμε δυο διαφορετικές αρχιτεκτονικές νευρωνικών δικτύων. Το πρώτο μοντέλο είναι ένα κλασικό νευρωνικό δίκτυο κρυφών επιπέδων και το δεύτερο ένα αναδρομικό μοντέλο τύπου RNN σε συνδυασμό με ένα απλό νευρωνικό δίκτυο στο επίπεδο επίπεδο.

Η μεταβλητή η οποία προβλέπεται, δηλαδή ο αναμενόμενος χρόνος άφιξης είναι κοινός και για τα δύο μοντέλα όπως και το σύνολο εκπαίδευσης που χρησιμοποιείται. Το ground truth(υπολοιπόμενος χρόνος άφιξης) προκύπτει από τον υπολογισμό της χρονικής απόστασης κάθε σημείου της τροχιάς εν συγκρίση με το τελικό σημείο άφιξης. Τα δίκτυα εκπαιδεύονται σε ένα σύνολο 3707 τροχιών(70% επί του συνολικού μεγέθους) και δοκιμάζονται σε 1589 τροχιές. Ο έλεγχος της απόδοσης και η επιλογή του καλύτερου μοντέλου γίνεται μέσω της τεχνικής k-fold validation με $k = 10$. Πριν την φάση της εκπαίδευσης, το σύνολο των δεδομένων μετασχηματίζεται στο διάστημα 0 - 1.

6.1 Νευρωνικό Δίκτυο

Το νευρωνικό δίκτυο που επιλέχθηκε αποτελείται από 3 κρυφά επίπεδα με 25, 10 10 νευρώνες αντίστοιχα. Ο αριθμός των επιπέδων και των νευρώνων επιλέχθηκε πειραματικά με βάση την απόδοση στο σύνολο επαλήθευσης(validation set). Το δίκτυο εκπαιδεύτηκε με βάση τον αλγόριθμο βελτιστοποίησης Adam και ρυθμό μάθησης ίσο με 0.01. Για την εισαγωγή μη γραμμικότητας, χρησιμοποιήθηκε η συνάρτηση ενεργοποίησης ReLU. Η επιλογή αυτής της συνάρτησης ενεργοποίης βασίζεται τόσο στην ταχύτητα σύγκλισης που πετυχαίνει όσο και στο γεγονός ότι η μεταβλητή που προβλέπεται δεν πρέπει να λαμβάνει αρνητικές τιμές. Παρακάτω παρουσιάζεται η διαγραμματική απεικόνιση του δικτύου. Τα δίκτυα εκπαιδεύτηκαν με βάση την τεχνική earling stopping σύμφωνα με το validation loss και με patience = 20 και min delta = 0.05.



Σχήμα 6.1: Δομή Νευρωνικού Δικτύου

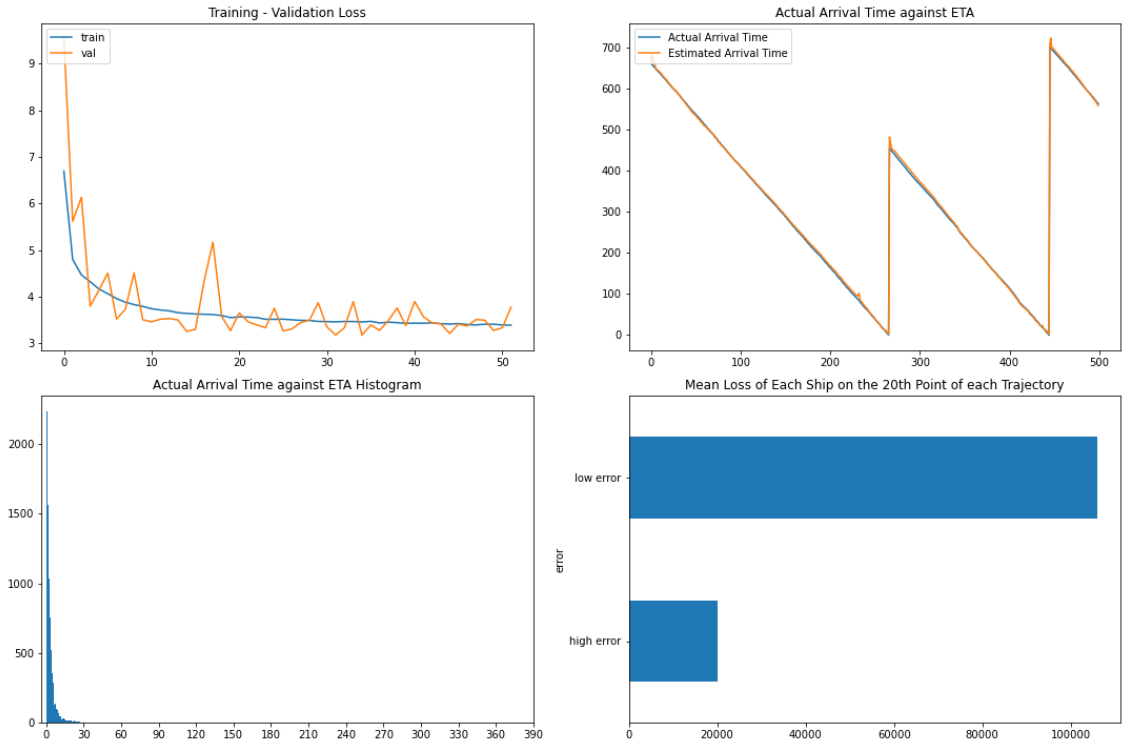
Το σύνολο εκπαίδευσης περιλαμβάνει τα εξής χαρακτηριστικά:

- μέση ταχύτητα πλοίου υπολογισμένη από το σύνολο των διαθέσιμων εγγραφών
- απόσταση μέχρι τον επόμενο λιμένα σε ναυτικά μίλια
- στιγμιαία ταχύτητα πλοίου σε κόμβους όπως αναφέρεται από το σύστημα AIS
- στιγμιαία ταχύτητα όπως υπολογίζεται με βάση την απόσταση που διανύθηκε στον αντίστοιχο χρόνο
- απολύτη διαφορά heading και course
- στροφή πλοίου ανά λεπτό
- κατάσταση ταξιδιού - έξοδος από λιμάνι, πορεία, είσοδος σε λιμάνι
- τύπος πλοίου - (γραμμής, κρουαζιερόπλοιο)
- αριθμός πλοίων στο λιμάνι άφιξης

6.1.1 Αποτελέσματα

Το δίκτυο που επιλέχθηκε έχει μέσο απόλυτο σφάλμα 2 λεπτά στο σύνολο εκπαίδευσης και 3 στο σύνολο επαλήθευσης. Η εκπαίδευση διήρκεσε 50 εποχές. Όπως φαίνεται και στο ιστόγραμμα όπου αναπαριστά την διαφορά του πραγματικού χρόνου άφιξης με την προβλεπόμενη τιμή, δηλαδή το σφάλμα, οι περισσότερες τιμές που παρατηρούνται βρίσκονται στο

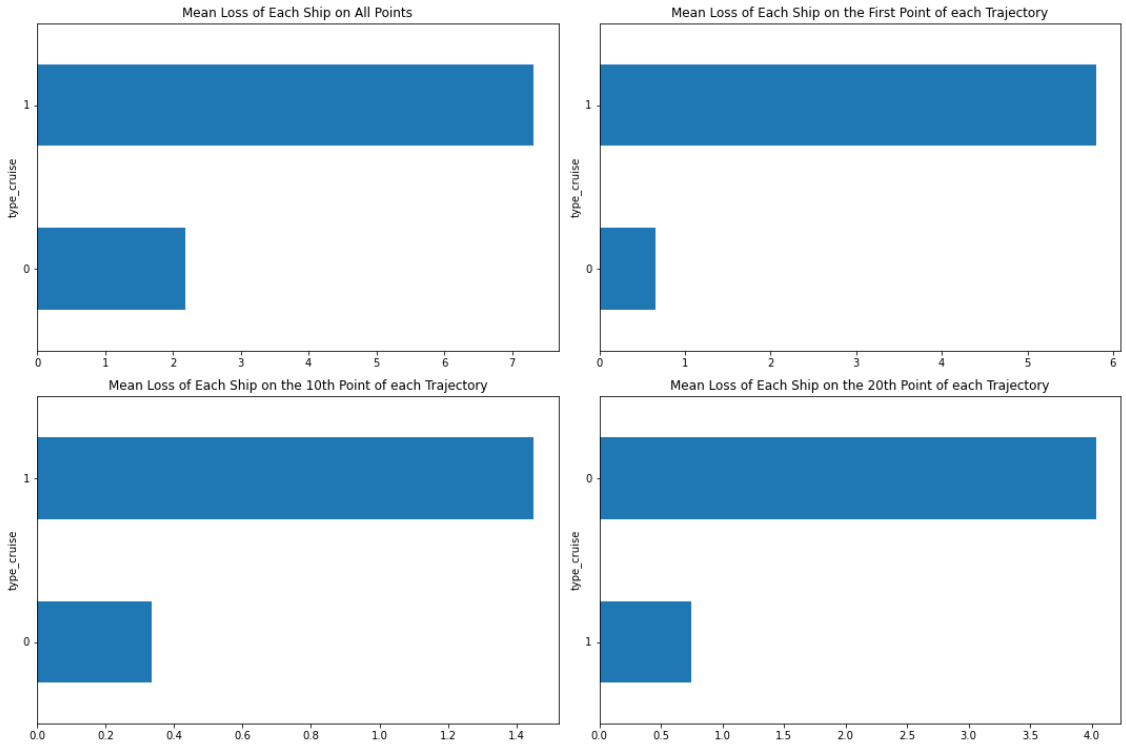
διάστημα 0 - 15(λεπτών). Αντίστοιχα από το ραβδόγραμμα φαίνεται ότι τα περισσότερα σημεία χαρακτηρίζονται ως low error, όπου low error όλα τα σημεία με σφάλμα μικρότερο των 5 λεπτών ενώ high error μεγαλύτερο των 5 λεπτών.



Σχήμα 6.2: Στατιστικά 1ου Μοντέλου

Παρατηρήθηκε πως το σφάλμα είναι πολύ μεγαλύτερο όταν το πλοίο για τα οποίο πραγματοποιείται πρόβλεψη είναι κρουαζιερόπλοιο. Πιο συγκεκριμένα το σφάλμα για τα πλοία γραμμής είναι 2 λεπτά ενώ το μέσο σφάλμα για τα κρουαζιερόπλοια είναι 7 λεπτά. Αυτό οφείλεται στο γεγονός πως τα πλοία γραμμής χαρακτηρίζονται από μεγαλύτερη συνέπεια όσον αφορά την κίνηση τους, δηλαδή έχουν λιγότερες αυξομειώσεις στην ταχύτητα τους κατά την πορεία τους καθώς και χρειάζονται διαφορετικό χρόνο για να προσεγγίσουν και να εξέρχονται από κάθε λιμάνι.

Στα παρακάτω σχήματα παρουσιάζονται τα διαφορετικά επίπεδα σφάλματος για κάθε τύπο πλοίου σε διαφορετικό σημείο της τροχιάς. Το πρώτο ραβδόγραμμα αφορά στο μέσο σφάλμα υπολογισμένο σύμφωνα με όλα τα σημεία. Το δεύτερο αφορά στο μέσο σφάλμα που προκύπτει μόνο από τα πρώτο σημείο της κάθε τροχιάς και τα επόμενα γραφήματα για το 10ο και 20ο σημείο. Οι στήλες με την τιμή 1 αφορούν κρουαζιερόπλοια ενώ για την τιμή 0 πλοία ακτοπλοίας.



Σχήμα 6.3: Ιστόγραμμα Διαφοράς Χρόνου

6.2 Αναδρομικό Νευρωνικό Δίκτυο

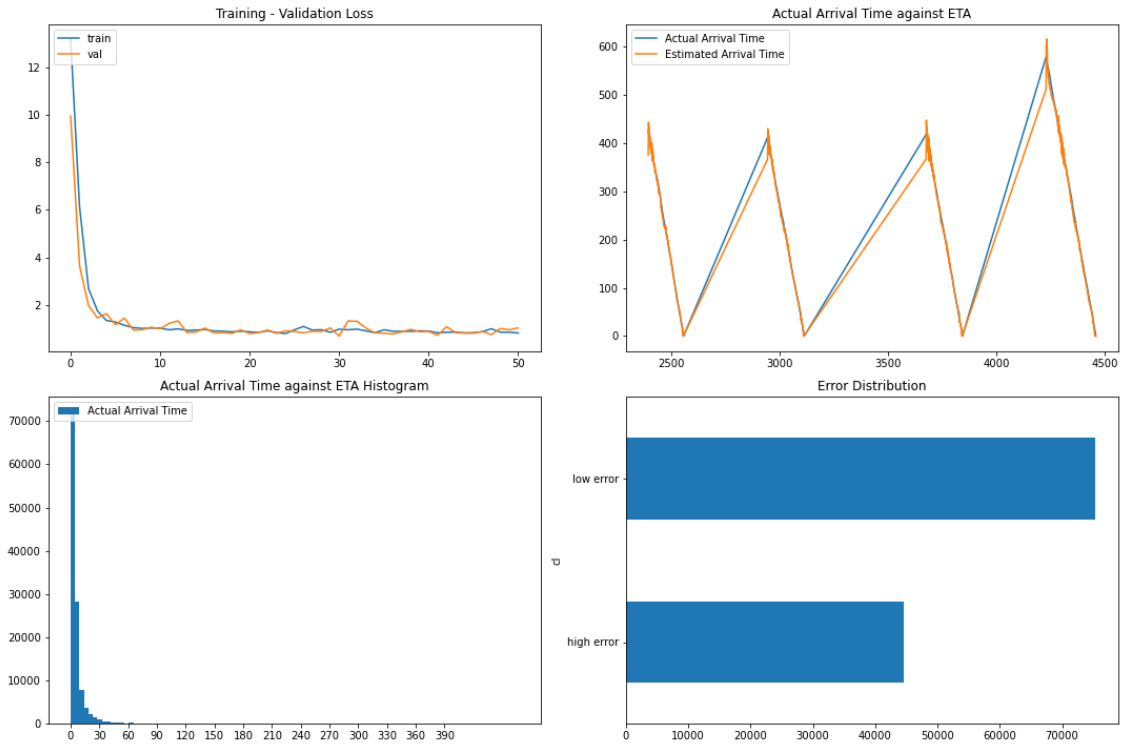
Το δεύτερο μοντέλο που επιλέξαμε αποτελεί έναν συνδυασμό αναδρομικών δικτύων τύπου RNN με ένα ANN. Ο λόγος που επιλέξαμε να χρησιμοποιήσουμε αναδρομικό δίκτυο είναι ώστε να μπορέσουμε να εκμεταλλευτούμε την δυνατότητα τους να εξετάζουν δεδομένα και προηγούμενων χρόνων για να πραγματοποιήσουν την πρόβλεψη. Η αρχιτεκτονική που επιλέξαμε αποτελείται από ένα στρώμα RNN με 25 units το οποίο προβλέπει σε όλο το μήκος της χρονοσειράς καθώς και ένα κρυφό στρώμα 100 νευρώνων. Όπως και για το προηγούμενο μοντέλο, χρησιμοποιήθηκε εξίσου η συνάρτηση ενεργοποίησης ReLU και ο αλγόριθμος βελτιστοποίησης Adam με learning rate ίσο με 0.01. Το διάνυσμα εισόδου έχει οργανωθεί σε ένα batch και το κάθε sample αποτελεί μια ολόκληρη τροχιά. Για τον λόγο ότι ο αριθμός των σημείων από τροχιά σε τροχιά δεν είναι σταθερός, εντοπίσαμε την τροχιά με το μεγαλύτερο μήκος και συμπληρώσαμε τα αντίστοιχα κενά στις μικρότερες τροχιές από αυτή. Πιο συγκεκριμένα, δουλέψαμε με την τεχνική zero padding καθώς και συμπεριλάβαμε ένα στρώμα τύπου Masking ώστε να παρακάμπτουμε τα κενά timesteps για να μην επηρεάζουν το συνολικό σφάλμα και την ταχύτητα εκπαίδευσης του δικτύου.

Οι μεταβλητές που χρησιμοποιήθηκαν για την πραγματοποίηση της πρόβλεψης είναι οι ίδιες με το αρχικό νευρωνικό δίκτυο μαζί με δύο επιπλέον, οι οποίες περιγράφουν τον χρονική απόσταση δυο διαδοχικών σημάτων καθώς και την αντίστοιχα απόσταση που το πλοίο διάνυσε στο μεσοδιάστημα.

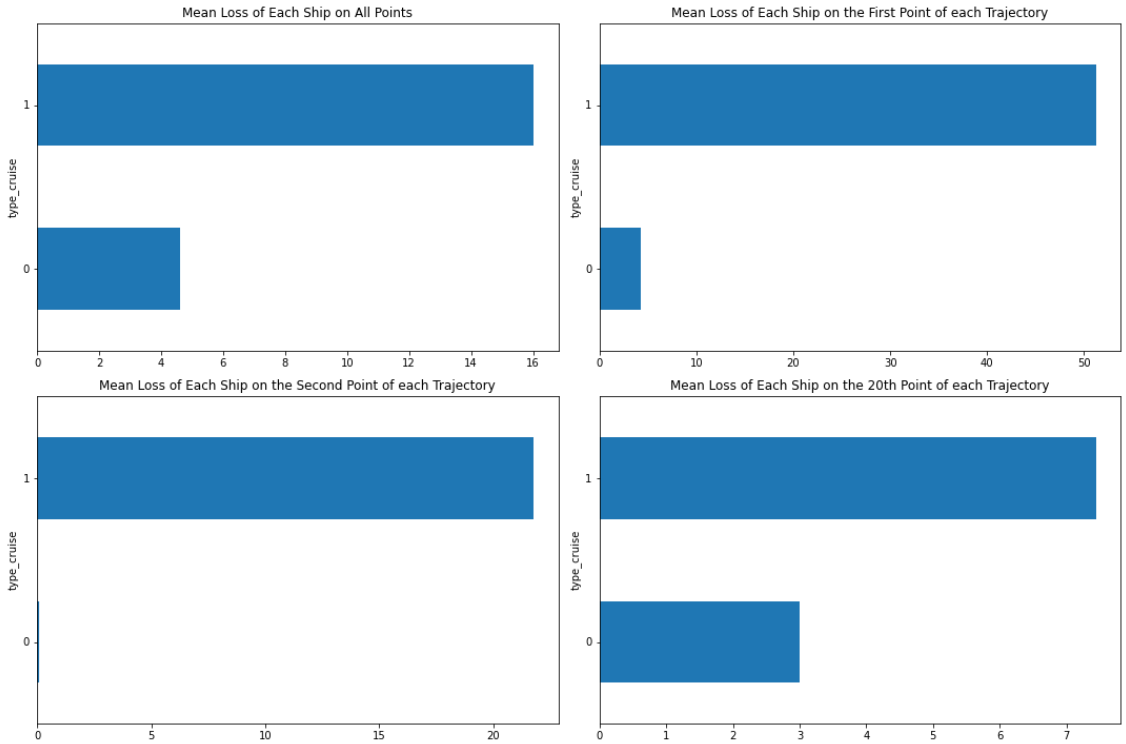
Στην προκειμένη περίπτωση υλοποιήσαμε μια παραλλαγή της συνάρτησης του μέσο απόλυτου σφάλματος η οποίο δεν συνυπολογίζει τα κενά διανύσματα που προστέθηκαν μετά την διαδικασία του padding. Αν δεν συνέβαινε αυτό τότε το δίκτυο δεν παρουσιάζει μικρότερο φαινομενικό σφάλμα.

6.2.1 Αποτελέσματα

Το δίκτυο που επιλέχτηκε έχει μέσο απόλυτο σφάλμα στο training set ίσο με 0.8 ενώ στο validation set 1.1. Αν και τα στατιστικά που παράγει η βιβλιοθήκη Keras με την οποία έχουν υλοποιηθεί τα μοντέλα, είναι πολύ χαμηλά, στην πράξη παρατηρούμε μεγαλύτερο σφάλμα. Συγκριτικά με το ANN της προηγούμενης ενότητας παρατηρούμε ότι το αναδρομικό νευρωνικό δίκτυο προβλέπει ένα μεγαλύτερο σύνολο σημείων με σφάλμα μεγαλύτερο των 5 λεπτών. Επίσης φαίνεται πως για να βελτιωθεί η ακρίβεια του δικτύου, θα πρέπει το πλοίο που εξετάζουμε να διανύσει έναν αριθμό χιλιομέτρων.



Σχήμα 6.4: Ιστόγραμμα Διαφοράς Χρόνου



Σχήμα 6.5: Ιστόγραμμα Διαφοράς Χρόνου

6.3 Σύγκριση με το ημερολόγιο επισκέψεων κρουαζιερόπλοιων Σαντορίνης

Στην παρούσα ενότητα θα παρουσιάσουμε μια εφαρμογή των δύο δικτύων και θα συγκρίνουμε τις προβλέψεις που θα πραγματοποιήσουμε με πραγματικά δεδομένα από το πρόγραμμα επισκέψεων κρουαζιερόπλοιων του λιμένα Σαντορίνης. Στην διάθεση μας έχουμε των πίνακα επισκέψεων από τον Ιούνιο του 2017 μέχρι και τον Σεπτέμβριο του 2018. Φυσικά δεν θα χρησιμοποιήσουμε όλα αυτά τα δεδομένα αλλά την τομή τους με το σύνολο των AIS δεδομένων. Έτσι, πραγματοποιώντας μια πράξη τύπου inner join στις στήλες ημερομηνία και όνομα πλοίου μεταξύ των δύο συνόλων λαμβάνουμε το επιθυμητό αποτέλεσμα.

Ο πίνακας που θα πραγματοποιήσουμε τις προβλέψεις περιλαμβάνει 51 διαφορετικά δρομολόγια. Είναι απαραίτητο να τονίσουμε πως στο αντίστοιχο χρονικό διάστημα σίγουρα υπήρχαν περισσότερες από 51 επισκέψεις. Ο λόγος που υπάρχει αυτή η απόκλιση δεν είναι άλλος πέρα από την έλλειψη μεταδεδομένων. Πιο συγκεκριμένα, στα AIS δεδομένα το μόνο αναγνωριστικό που δίνεται είναι ο αριθμός mmsi ενώ στους πίνακες της Σαντορίνης μόνο το όνομα του κάθε πλοίου. Για τον λόγο αυτό προσπαθήσαμε να αντιστοιχήσουμε όσα περισσότερα από τα ονόματα των πλοίων με τον αντίστοιχο αριθμο mmsi. Για τα πλοία με κοινό όνομα δεν καταφέραμε να βρούμε τον αντίστοιχο αριθμό διότι υπήρχαν πάρα πολλές υποψήφιες και διαφορετικές επιλογές.

Ύστερα από ανάλυση των τελικών δεδομένων προκύπτει πως, η μέσο χρονική απόκλιση από το χρονοδιάγραμμα είναι 11 λεπτά. Από τα 51 δρομολόγια, 37 από αυτά έφτασαν στο λιμάνι σε χρόνο πριν από τον αναγραφόμενο ενώ τα υπόλοιπα 14, σε μεταγενέστερο χρόνο. Στον παρακάτω πίνακα φαίνονται οι αντίστοιχοι προγραμματισμένοι χρόνοι όλων των δρομολογίων και οι αντίστοιχες προβλέψεις των δύο δικτύων στο πρώτο σημείο κάθε τροχιάς δηλαδή μόλις ξεκινήσει το πλοίο. Το μέσο σφάλμα για το ANN είναι 7.8 λεπτά ενώ για το RNN είναι 47

λεπτά. Η διάμεσος που προκύπτει από το δίκτυο ANN είναι 6.2 λεπτά και για το δίκτυο RNN 28.2 λεπτά.

Ημερομηνία	Προγραμματισμένη Άφιξη	Πραγματική Άφιξη	Πρόβλεψη ANN	Πρόβλεψη RNN
2018-05-31	16:30:00	16:55:52	16:56:11.838000	16:34:45.029999
2018-06-12	07:30:00	07:09:47	07:10:13.700400	07:34:26.547600
2018-06-21	08:00:00	07:46:26	07:47:04.296400	08:01:18.384399
2018-07-05	16:45:00	16:29:57	16:30:36.172199	15:55:12.029999
2018-06-04	07:00:00	06:41:11	06:40:10.154400	06:56:31.997999
2018-05-27	16:30:00	16:39:37	16:41:19.428999	16:16:10.674999
2018-08-10	08:00:00	07:43:46	07:45:41.374799	07:37:48.677999
2018-05-20	16:30:00	16:33:07	16:35:28.038000	15:58:45.627599
2018-06-28	08:00:00	07:25:20	07:27:47.081000	07:31:19.878799
2018-05-24	07:00:00	06:45:27	06:48:02.278800	06:47:05.680799
2018-07-15	16:45:00	16:47:10	16:49:58.940399	16:25:50.174999
2018-08-23	08:00:00	07:40:51	07:44:13.676000	07:50:23.493200
2018-08-26	10:00:00	09:57:29	10:01:05.841799	10:14:07.987999
2018-05-03	07:00:00	06:56:00	06:59:40.058199	06:52:45.417999
2018-06-07	08:00:00	07:55:59	07:59:54.789200	07:52:13.145600
2018-07-26	16:45:00	16:27:42	16:31:56.050199	16:14:00.435600
2018-08-09	16:45:00	16:03:20	16:08:04.815599	15:32:38.990400
2018-08-30	16:45:00	16:06:17	16:11:09.038800	15:44:37.851399
2018-05-03	16:30:00	15:58:06	16:03:46.383399	14:36:38.393800
2018-06-10	16:45:00	16:43:26	16:49:06.402400	17:19:28.133199
2018-06-21	16:45:00	16:44:17	16:50:07.770200	17:32:36.490400
2018-06-14	08:00:00	07:48:22	07:54:14.743200	08:01:14.046000
2018-08-16	07:00:00	07:27:24	07:33:22.199599	07:01:00.890000
2018-05-13	16:30:00	16:41:14	16:47:18.580399	17:24:30.059000
2018-07-12	08:00:00	07:15:34	07:21:42.585999	07:17:58.735599
2018-07-19	08:00:00	07:34:22	07:40:34.523999	07:25:49.344000
2018-06-07	16:45:00	16:46:19	16:52:35.237999	17:29:13.549200
2018-07-29	16:45:00	16:26:25	16:32:48.258399	15:40:07.991800
2018-08-19	16:45:00	15:58:13	16:04:39.855999	15:34:23.569000
2018-07-05	08:00:00	07:20:39	07:27:13.941200	07:21:17.279599
2018-08-23	16:45:00	15:58:18	16:05:12.096199	15:31:56.720000
2018-07-19	16:45:00	16:32:51	16:39:47.837599	17:01:20.313199
2018-06-14	16:45:00	16:46:29	16:53:30.851799	17:35:05.185600
2018-07-01	16:45:00	16:27:32	16:34:36.289999	13:38:40.706000
2018-06-03	16:45:00	16:33:08	16:40:20.616800	13:31:02.688800
2018-06-28	16:45:00	16:53:58	17:01:18.636800	13:46:04.417999
2018-08-26	09:00:00	08:39:58	08:32:08.430200	08:05:45.140000
2018-05-24	16:30:00	16:35:30	16:43:25.225199	13:33:51.619560
2018-07-22	16:45:00	16:23:51	16:31:53.905999	17:09:23.880800
2018-05-06	16:30:00	16:33:15	16:41:30.165199	17:19:51.700400
2018-06-21	07:00:00	07:33:44	07:25:04.249799	06:43:20.974800
2018-06-24	16:45:00	16:39:36	16:48:27.512200	13:41:47.632299
2018-06-17	16:45:00	16:38:59	16:49:23.351000	13:44:35.482399

Παρατηρήθηκε πως αν και το δίκτυο RNN στην αρχή έχει μεγάλη απόκλιση όταν χρησιμοποιείται μόνο μια στιγμή της κάθε χρονοσειράς - τροχιάς, με το πέρας του χρόνου η απόδοση του αυξάνεται. Για το 20ο σημείο της κάθε τροχιάς παρατηρούμαι 7.7 λεπτά απόκλιση για το αρχικό δίκτυο ενώ 8.3 λεπτά απόκλιση για το δίκτυο RNN. Η αντίστοιχη διάμεσος για το δίκτυο ANN είναι 6.5 λεπτά και 4.3 λεπτά για το RNN.

Ημερομηνία	Προγραμματισμένη Άφιξη	Πραγματική Άφιξη	Πρόβλεψη ANN	Πρόβλεψη RNN
2018-05-03	07:00:00	06:56:00	06:56:32.032000	06:43:58.019200
2018-06-04	07:00:00	06:41:11	06:41:51.766000	06:37:00.068200
2018-06-28	08:00:00	07:25:20	07:26:06.993000	07:06:30.325799
2018-08-30	08:00:00	07:44:18	07:45:33.116199	07:15:55.380399
2018-08-10	08:00:00	07:43:46	07:45:49.266200	07:36:30.098000
2018-05-31	16:30:00	16:55:52	16:57:55.746799	16:50:47.375000
2018-06-21	08:00:00	07:46:26	07:48:31.740400	07:31:50.506000
2018-06-12	07:30:00	07:09:47	07:12:01.026000	07:18:10.340399
2018-07-15	16:45:00	16:47:10	16:49:27.181200	16:41:44.341799
2018-07-05	16:45:00	16:29:57	16:33:01.309999	16:26:03.332000
2018-05-24	07:00:00	06:45:27	06:48:41.487999	06:48:45.384999
2018-07-05	08:00:00	07:20:39	07:24:14.440000	07:16:00.687999
2018-07-12	08:00:00	07:15:34	07:19:16.569999	07:17:08.743000
2018-05-27	16:30:00	16:39:37	16:43:44.234400	16:35:25.149600
2018-08-30	16:45:00	16:06:17	16:10:44.938400	16:08:52.722199
2018-08-26	10:00:00	09:57:29	10:02:05.638000	09:52:35.862999
2018-05-20	16:30:00	16:33:07	16:37:48.357599	16:34:11.399399
2018-07-26	16:45:00	16:27:42	16:32:25.245800	16:26:43.497800
2018-08-23	08:00:00	07:40:51	07:45:46.710000	07:27:45.763799
2018-06-07	08:00:00	07:55:59	08:01:00.490400	07:57:34.292000
2018-08-09	16:45:00	16:03:20	16:08:34.218400	16:07:38.121999
2018-08-26	09:00:00	08:39:58	08:33:53.628999	08:21:58.907799
2018-08-19	16:45:00	15:58:13	16:04:29.383399	16:06:19.174400
2018-05-03	16:30:00	15:58:06	16:04:22.631999	16:04:28.945199
2018-06-07	16:45:00	16:46:19	16:52:44.354199	16:45:15.329200
2018-06-14	16:45:00	16:46:29	16:52:59.510400	16:44:51.707999
2018-07-19	16:45:00	16:32:51	16:39:28.755000	16:34:28.832999
2018-05-13	16:30:00	16:41:14	16:47:57.525199	16:40:27.342200
2018-06-28	16:45:00	16:53:58	17:00:58.694999	16:52:07.944000
2018-06-03	16:45:00	16:33:08	16:40:12.092000	16:32:12.147799
2018-07-01	16:45:00	16:27:32	16:34:38.043400	16:26:51.730000
2018-07-29	16:45:00	16:26:25	16:33:32.361200	16:27:11.216000
2018-06-21	16:45:00	16:44:17	16:51:28.188800	16:44:38.994199
2018-06-24	16:45:00	16:39:36	16:46:51.015599	16:38:17.929800
2018-05-17	16:30:00	16:25:44	16:33:10.128599	16:26:34.097399
2018-08-16	07:00:00	07:27:24	07:35:11.021800	07:50:13.493799
2018-08-23	16:45:00	15:58:18	16:06:09.766000	16:03:50.588200
2018-07-22	16:45:00	16:23:51	16:31:43.568799	16:24:36.406599
2018-06-10	16:45:00	16:43:26	16:51:25.606199	16:51:27.207600
2018-06-14	08:00:00	07:48:22	07:56:53.296799	07:44:11.023200
2018-05-06	16:30:00	16:33:15	16:41:56.401799	16:36:07.766400
2018-05-24	16:30:00	16:35:30	16:45:04.157599	16:36:08.340800

6.4 Πρόβλεψη σε πλοία ακτοπλοίας με προορισμό την Σαντορίνη

Ο παρακάτω πίνακας περιλαμβάνει τον πραγματικό και τον προβλεπόμενο υπολειπόμενο χρόνο άφιξης(σε λεπτά) για πλοία γραμμής τα οποία επισκέπτονται το λιμάνι της Σαντορίνης. Όπως επισημάνθηκε το δίκτυο τύπου ANN σημειώνει υψηλότερη ακρίβεια έναντι του δικτύου RNN. Η πρόβλεψη έγινε στο πρώτο σημείο της τροχιάς, δηλαδή μόλις το πλοίο ξεκίνησε για τον προορισμό του, με το δίκτυο ANN. Πέρα από τα αριθμητικά δεδομένα που περιγράφουν την ακρίβεια του δικτύου, μπορεί να παρατηρηθεί και στον παρακάτω πίνακα πως η διαφορά του προβλεπόμενου και το πραγματικού αναμενόμενου χρόνου είναι μικρότερη έναντι αυτής των κρουαζιερόπλοιων.

Πραγματικός Υπολειπόμενος Χρόνος Άφιξης	Προβλεπόμενος Υπολειπόμενος Χρόνος Άφιξης
65.433333	68.632462
69.350000	70.294754
115.716667	111.181862
296.783333	298.863617
251.550000	252.440582
52.250000	48.122597
67.100000	68.858475
66.333333	68.470337
57.266667	59.464920
115.933333	112.907623
335.750000	343.713409
110.116667	111.253723
112.033333	112.955330
315.900000	318.145966
250.383333	250.527573
311.100000	312.532532
246.333333	247.550003
55.100000	48.301991
66.750000	70.359283
281.000000	285.666443
253.366667	256.393127
117.233333	117.551933
118.583333	117.755981
125.266667	126.745911
115.516667	113.791084
296.483333	295.900970
296.983333	300.721039
115.100000	116.326469
300.416667	303.318390
320.250000	323.733826
322.633333	325.951965
112.433333	116.220345
115.100000	115.073898
50.333333	48.083725
114.783333	117.205757
52.400000	51.699619
50.750000	49.403061
114.616667	114.761658
117.300000	115.514481
311.566667	314.289154
52.383333	51.055676

Κεφάλαιο 7

Δοκιμές σε συνεχόμενες διαδρομές

Στο 2ο κομμάτι της εργασίας επιχειρείται η πρόβλεψη του χρόνου άφιξης σε διαδοχικούς λιμένες. Η κύρια διαφορά μεταξύ αυτής και των δύο προηγούμενων υλοποιήσεων είναι πως τώρα στο σύνολο εκπαίδευσης συμπεριλαμβάνονται οι αποστάσεις για τα 9 επόμενα λιμάνια καθώς και χρονική διάρκεια των ενδιάμεσων στάσεων. Αναλυτικότερα χρησιμοποιούμε τα παρακάτω features.

- αποστάσεις μέχρι τα επόμενα εννέα λιμάνια
- ταχύτητα πλοίου σε κόμβους όπως αναφέρει το σύστημα AIS
- ταχύτητα πλοίου σε κόμβους όπως υπολογίζεται από τα γεωγραφικά σημεία
- heading πλοίου σε μοίρες
- course πλοίου σε μοίρες
- στροφή ανά λεπτό
- διαφορά course και heading σε μοίρες
- κατάσταση ταξιδιού - έξοδος από λιμάνι, πορεία, είσοδος σε λιμάνι
- μέση ταχύτητα πλοίου υπολογισμένη από το σύνολο των διαθέσιμων εγγραφών
- χρονική διάρκεια ενδιάμεσων στάσεων
- αριθμός πλοίων στο λιμάνι άφιξης

Το νευρωνικό δίκτυο που χρησιμοποιήθηκε αποτελείται από ένα στρώμα RNN και 4 στρώματα κρυφών νευρώνων. Οι υπόλοιπες παραμέτροι ρυθμίστηκαν όπως και στα προηγούμενα μοντέλα. Στην προκειμένη περίπτωση το σφάλμα στο σύνολο επαλήθευσης είναι 5 λεπτά.

1ο Λιμάνι	2ο Λιμάνι	3ο Λιμάνι	4ο Λιμάνι	Πρόβλεψη$_1$	Πρόβλεψη$_2$	Πρόβλεψη$_3$	Πρόβλεψη$_4$	Ground Truth$_1$	Ground Truth$_2$	Ground Truth$_3$	Ground Truth$_4$
Παρος	Ναξος	Σαντορινι	Ναξος	248.0	287.0	460.0	587.0	247	309	474	609
Παρος	Ναξος	Σαντορινι	Ναξος	243.0	286.0	460.0	583.0	240	303	467	603
Παρος	Ναξος	Σαντορινι	Ναξος	248.0	288.0	464.0	593.0	238	301	465	601
Παρος	Ναξος	Σαντορινι	Ναξος	251.0	325.0	473.0	597.0	235	298	462	597
Παρος	Ναξος	Σαντορινι	Ναξος	226.0	294.0	453.0	588.0	232	295	459	594
Παρος	Ναξος	Σαντορινι	Ναξος	225.0	292.0	450.0	585.0	229	291	456	591
Παρος	Ναξος	Σαντορινι	Ναξος	224.0	290.0	446.0	581.0	225	287	452	587
Παρος	Ναξος	Σαντορινι	Ναξος	222.0	287.0	445.0	580.0	223	285	450	585
Παρος	Ναξος	Σαντορινι	Ναξος	212.0	278.0	438.0	573.0	216	279	443	579
Παρος	Ναξος	Σαντορινι	Ναξος	211.0	276.0	436.0	570.0	214	277	441	577
Παρος	Ναξος	Σαντορινι	Ναξος	178.0	245.0	405.0	537.0	183	246	410	546
Παρος	Ναξος	Σαντορινι	Ναξος	161.0	225.0	385.0	519.0	165	227	392	527
Παρος	Ναξος	Σαντορινι	Ναξος	154.0	220.0	377.0	512.0	157	220	384	519
Παρος	Ναξος	Σαντορινι	Ναξος	139.0	205.0	359.0	494.0	140	203	367	503
Παρος	Ναξος	Σαντορινι	Ναξος	138.0	204.0	357.0	491.0	138	201	365	501
Παρος	Ναξος	Σαντορινι	Ναξος	124.0	188.0	341.0	479.0	124	186	351	486
Παρος	Ναξος	Σαντορινι	Ναξος	115.0	179.0	332.0	475.0	112	175	339	474
Παρος	Ναξος	Σαντορινι	Ναξος	110.0	168.0	323.0	464.0	106	168	333	468
Παρος	Ναξος	Σαντορινι	Ναξος	96.0	155.0	309.0	450.0	92	155	319	455
Παρος	Ναξος	Σαντορινι	Ναξος	90.0	151.0	306.0	446.0	88	151	315	451
Παρος	Ναξος	Σαντορινι	Ναξος	89.0	149.0	304.0	444.0	86	149	313	449
Παρος	Ναξος	Σαντορινι	Ναξος	72.0	135.0	290.0	427.0	71	133	298	433
Παρος	Ναξος	Σαντορινι	Ναξος	54.0	117.0	270.0	407.0	52	114	279	414
Παρος	Ναξος	Σαντορινι	Ναξος	50.0	113.0	267.0	404.0	48	110	275	410
Παρος	Ναξος	Σαντορινι	Ναξος	42.0	106.0	258.0	395.0	39	102	266	402
Παρος	Ναξος	Σαντορινι	Ναξος	40.0	104.0	256.0	393.0	37	100	264	400
Παρος	Ναξος	Σαντορινι	Ναξος	36.0	100.0	252.0	389.0	33	96	260	396
Παρος	Ναξος	Σαντορινι	Ναξος	34.0	98.0	250.0	387.0	31	94	258	394
Παρος	Ναξος	Σαντορινι	Ναξος	27.0	91.0	241.0	381.0	25	87	252	387
Παρος	Ναξος	Σαντορινι	Ναξος	25.0	89.0	238.0	379.0	23	85	250	385
Παρος	Ναξος	Σαντορινι	Ναξος	15.0	78.0	224.0	349.0	12	75	239	375
Παρος	Ναξος	Σαντορινι	Ναξος	20.0	70.0	249.0	360.0	10	73	237	373

Κεφάλαιο 8

Βιβλιογραφία

- Anon. 2021. Activation function. (June 2021). Retrieved July 7, 2021 from https://en.wikipedia.org/wiki/Activation_function
- Anon. 2021. Automatic differentiation. (June 2021). Retrieved July 7, 2021 from https://en.wikipedia.org/wiki/Automatic_differentiation
- Oleh Bodunov, Florian Schmidt, André Martin, Andrey Brito, and Christof Fetzer. 2018. Real-time Destination and ETA Prediction for Maritime Traffic. Proceedings of the 12th ACM International Conference on Distributed and Event-based Systems (2018). DOI:http://dx.doi.org/10.1145/3210284.3220502
- Ian J. Goodfellow, Yoshua Bengio, and Aaron Courville. 2016. Deep learning, Cambridge, MA: The MIT Press.
- Kwang Il Kim and Keon Myung Lee. 2017. Preprocessing Ship Trajectory Data for Applying Artificial Neural Network in Harbour Area. 2017 European Conference on Electrical Engineering and Computer Science (EECS) (2017). DOI:http://dx.doi.org/10.1109/eecs.2017.36
- B.t. Morris and M.m. Trivedi. 2008. A Survey of Vision-Based Trajectory Learning and Analysis for Surveillance. IEEE Transactions on Circuits and Systems for Video Technology 18, 8 (2008), 1114–1127. DOI:http://dx.doi.org/10.1109/tcsvt.2008.927109
- Charul Paliwal and Pravesh Biyani. 2019. To each route its own ETA: A generative modeling framework for ETA prediction. 2019 IEEE Intelligent Transportation Systems Conference (ITSC) (2019). DOI:http://dx.doi.org/10.1109/itsc.2019.8917465
- Anon. 2021. Perceptron. (June 2021). Retrieved July 7, 2021 from https://en.wikipedia.org/wiki/Perceptron
- Anon. 2021. Rectifier (neural networks). (July 2021). Retrieved July 7, 2021 from https://en.wikipedia.org/wiki/Rectifier_(neural_networks)
- Kaushik Roy, Benjamin Levy, and Claire Tomlin. 2006. Target Tracking and Estimated Time of Arrival (ETA) Prediction for Arrival Aircraft. AIAA Guidance, Navigation, and Control Conference and Exhibit (2006). DOI:http://dx.doi.org/10.2514/6.2006-6324

- Sergios Theodoridis and Konstantinos Koutroumbas. 2009. Pattern Recognition, 4th Edition, Academic Press.
- Fan Wu and Lixia Wu. 2019. DeepETA: A Spatial-Temporal Sequential Neural Network Model for Estimating Time of Arrival in Package Delivery System. Proceedings of the AAAI Conference on Artificial Intelligence 33 (2019), 774–781. DOI:http://dx.doi.org/10.1609/aaai.v33i01.3301774
- Feixiang Zhu. 2011. Mining ship spatial trajectory patterns from AIS database for maritime surveillance. 2011 2nd IEEE International Conference on Emergency Management and Management Sciences (2011). DOI:http://dx.doi.org/10.1109/icemms.2011.6015796