

Πανεπιστήμιο Πειραιά
Τμήμα Ψηφιακών Συστημάτων
Πρόγραμμα Μεταπτυχιακών Σπουδών
Πληροφοριακά Συστήματα και Υπηρεσίες
Ειδίκευση: Προηγμένα Πληροφοριακά Συστήματα



Πανεπιστήμιο Πειραιώς
Τμήμα Ψηφιακών Συστημάτων

Τίτλος Εργασίας

«Ενεργειακή αποδοτικότητα στις θαλάσσιες μεταφορές χρησιμοποιώντας
intelligent route planning και ανάλυση με Python»

Όνοματεπώνυμο φοιτητή

Αποστολάκης Αντώνιος

Αθήνα 2025

Πανεπιστήμιο Πειραιά
Τμήμα Ψηφιακών Συστημάτων
Πρόγραμμα Μεταπτυχιακών Σπουδών
Πληροφοριακά Συστήματα και Υπηρεσίες
Ειδίκευση: Προηγμένα Πληροφοριακά Συστήματα



Πανεπιστήμιο Πειραιώς
Τμήμα Ψηφιακών Συστημάτων

Επιτροπή Επίβλεψης Διπλωματικής Εργασίας

Επιβλέπων: Φιλιππάκης Μιχαήλ,

Καθηγητής Τμήματος Ψηφιακών Συστημάτων Πανεπιστημίου Πειραιά

Μέλος: Χαλκίδη Μαρία,

Καθηγήτρια Τμήματος Ψηφιακών Συστημάτων Πανεπιστημίου Πειραιά

Μέλος: Κυριαζής Δημοσθένης

Καθηγητής Τμήματος Ψηφιακών Συστημάτων Πανεπιστημίου Πειραιά

Ευχαριστίες

Η παρούσα διπλωματική εργασία δεν θα μπορούσε να ολοκληρωθεί χωρίς τη σημαντική συμβολή ανθρώπων που με στήριξαν επιστημονικά και προσωπικά σε όλη τη διάρκεια των μεταπτυχιακών μου σπουδών στο Πανεπιστήμιο Πειραιά. Πρώτα απ' όλα, θα ήθελα να ευχαριστήσω τον επιβλέποντα καθηγητή κ. Μιχαήλ Φιλιππάκη για την εμπιστοσύνη που μου έδειξε, τη συστηματική καθοδήγηση και τις στοχευμένες παρατηρήσεις του, οι οποίες διαμόρφωσαν ουσιαστικά την κατεύθυνση και το τελικό αποτέλεσμα της εργασίας. Η εμπειρία και η επιστημονική του προσέγγιση αποτέλεσαν σταθερό σημείο αναφοράς καθ' όλη τη διάρκεια της έρευνας.

Θα ήθελα επίσης να εκφράσω τις ευχαριστίες μου προς τα μέλη της τριμελούς επιτροπής, την καθηγήτρια κα. Μαρία Χαλκίδη και τον καθηγητή κ. Δημοσθένη Κυριαζή, για την προσεκτική μελέτη της εργασίας και τα ουσιαστικά σχόλια που παρείχαν. Οι επισημάνσεις και οι προτάσεις τους συνέβαλαν καθοριστικά στη βελτίωση της ποιότητας και της επιστημονικής τεκμηρίωσης της μελέτης.

Σημαντική υπήρξε και η συμβολή των συναδέλφων και συνεργατών μου στον τομέα της ναυτιλίας, οι οποίοι μοιράστηκαν γνώσεις, εμπειρίες και δεδομένα από την διεθνή πραγματικότητα. Μέσα από τις συζητήσεις και την πρακτική τους οπτική, κατέστη δυνατή η σύνδεση της θεωρητικής ανάλυσης με τις πραγματικές συνθήκες της ναυτιλίας και η ανάπτυξη ενός μοντέλου με εφαρμόσιμο και ρεαλιστικό χαρακτήρα.

Τέλος, αισθάνομαι την ανάγκη να ευχαριστήσω την οικογένειά μου και τους φίλους μου για τη στήριξη, την υπομονή και την κατανόηση που έδειξαν σε όλη τη διάρκεια αυτής της απαιτητικής διαδρομής. Η ενθάρρυνσή τους και η παρουσία τους αποτέλεσαν καθοριστικό παράγοντα για να φτάσει αυτή η προσπάθεια στην ολοκλήρωσή της.

Περίληψη

Η εργασία παρουσιάζει ένα ολοκληρωμένο σύστημα ανάλυσης δεδομένων για τη βελτίωση της ενεργειακής αποδοτικότητας στις θαλάσσιες μεταφορές μέσω intelligent route planning. Ο πυρήνας της μεθόδου είναι η ενοποίηση πραγματικών δεδομένων AIS με μετεωρολογικές μετρήσεις (ταχύτητα ανέμου και ύψος κύματος), ώστε να προβλεφθεί με ακρίβεια η ταχύτητα πλεύσης (SOG) και να υπολογιστεί ο απαιτούμενος χρόνος ταξιδιού σε ρεαλιστικές θαλάσσιες διαδρομές. Το σύνολο δεδομένων καλύπτει την δεκαετή περίοδο 2015 έως 2024 στη περιοχή με γεωγραφικό μήκος 35° έως 42° και γεωγραφικό πλάτος από -77° έως -70° (περιοχή ΒΑ Αμερικής). Η προεπεξεργασία περιλαμβάνει τον καθαρισμό των δεδομένων, τον έλεγχο ακραίων τιμών καιρού, τη χαρτογράφηση των τύπων των πλοίων σε κατηγορίες και την επιλογή των πλοίων που έχουν τουλάχιστον την ταχύτητα των 4 κόμβων.

Για την πρόβλεψη της ταχύτητας των πλοίων χρησιμοποιείται ένα ενιαίο μοντέλο Random Forest για όλα τα έτη, με χαρακτηριστικά εισόδου το γεωγραφικό μήκος και πλάτος (θέση), την ταχύτητα και τον τύπο των πλοίων, την ταχύτητα του ανέμου και το ύψος του κύματος στην περιοχή, αποκλείοντας τα πλοία με πολύ χαμηλή ταχύτητα και στόχο την εύρεση της ταχύτητας. Η εκπαίδευση και η αξιολόγηση του μοντέλου πρόβλεψης δείχνει μεγάλη ακρίβεια στην ταχύτητα (MAE περίπου 1 κόμβος) και σε υψηλό ποσοστό ακρίβειας ($R^2 = 85\%$). Παράγονται επίσης αναλυτικά αποτελέσματα ανά κατηγορία πλοίου, αναδεικνύοντας τη σταθερή συμπεριφορά του μοντέλου. Στη συνέχεια, υλοποιήθηκε δρομολογητής (Route Planner) για την εύρεση διαδρομής μεταξύ δύο λιμανιών με τη χρήση γράφου κόμβων/ακμών, ο οποίος αποφεύγει τη ξηρά. Το σύστημα αποδίδει συνολικά την απόσταση του δρομολογίου δια θαλάσσης, υπολογίζοντας τον εκτιμώμενο χρόνο άφιξης και τη μέση ταχύτητα για την πλήρη διαδρομή, ενώ παρέχει χαρτογραφικές απεικονίσεις με χρωματική κωδικοποίηση της προβλεπόμενης ταχύτητας ανά σκέλος.

Η συμβολή της εργασίας είναι διττή:

(α) αποδεικνύει ότι η αξιοποίηση πραγματικών δεδομένων AIS και μετεωρολογίας (σε συνδυασμό με μοντέλο μηχανικής μάθησης) επιτρέπει ακριβείς και χρηστικές εκτιμήσεις ταχύτητας πλοίων και

(β) ενσωματώνει τις εκτιμήσεις αυτές σε έναν πρακτικό δρομολογητή που αποτυπώνει ρεαλιστικές θαλάσσιες διαδρομές, προσφέροντας μετρήσιμα μεγέθη (απόσταση, χρόνος άφιξης και μέση ταχύτητα) για την τεκμηριωμένη λήψη αποφάσεων.

Η υλοποίηση είναι σε γλώσσα προγραμματισμού Python, με ξεκάθαρη δομή φακέλων και αποθήκευση του εκπαιδευμένου μοντέλου, διευκολύνοντας την αναπαραγωγή και την ενσωμάτωση σε επιχειρησιακά περιβάλλοντα.

Λέξεις κλειδιά: AIS, Μετεωρολογικά Δεδομένα, Random Forest, Θαλάσσια δρομολόγηση, Python, ETA

Abstract

The thesis presents a comprehensive data analysis system for improving energy efficiency in maritime transport through intelligent route planning. The core of the method is the integration of real AIS data with meteorological measurements (wind speed and wave height) in order to accurately predict the speed of travel (SOG) and calculate the required travel time on realistic sea routes. The dataset covers the period from 2015 to 2024 in the region with a longitude of 35° to 42° and a latitude of -77° to -70° (NE America region). The preprocessing includes data cleaning, checking for extreme weather values, mapping ship types into categories, and selecting ships that have a speed of at least 4 knots.

To predict ship speed, a single Random Forest model is used for all years, with input features including geographical longitude and latitude (location), ship speed and type, wind speed, and wave height in the area, excluding ships with very low speeds and aiming to find the speed. The training and the evaluation of the model show high accuracy (MAE = approximately 1 knot) and a high accuracy rate ($R^2 = 85\%$). Detailed results are also produced per ship category, highlighting the stable behavior of the model. Next, a router was implemented to find a route between two ports using a node/edge graph, which avoids land. Based on the date of the route, local meteorological values are obtained for each leg of the route and fed into the trained Random Forest model to predict the speed, from which the travel time for each leg is calculated. The system calculates the total distance of the sea route, estimating the arrival time and average speed for the entire route, while providing color-coded maps of the predicted speed for each leg.

The contribution of this work is twofold:

It demonstrates that the use of real AIS and meteorological data, in combination with a machine learning model, allows for accurate and useful speed estimates, and

(b) integrates these estimates into a practical route planner that maps realistic sea routes, providing measurable quantities (distance, arrival time, and average speed) for informed decision-making.

The implementation is in the Python programming language, with a clear folder structure and storage of the trained model, facilitating reproduction and integration into operational environments.

Keywords: AIS, Weather data, Random Forest, Maritime routing, Python, ETA

Πίνακας Περιεχομένων

Περίληψη	iv
Abstract	vi
Κατάλογος Εικόνων	x
Εισαγωγή	1
1 Θεωρητικό υπόβαθρο	6
1.1 Τρόπος λειτουργίας συστημάτων AIS	6
1.2 Καιρικά στοιχεία (ένταση ανέμου και ύψος κύματος)	9
1.3 Μοντέλο Παλινδρόμησης Random Forest και βασικές υπερπαράμετροι	11
1.4 Sea mask και γράφος κόμβων/ακμών με A*	13
2 Δεδομένα και Προεπεξεργασία	15
2.1 Συλλογή Δεδομένων AIS - Λήψη αρχείων, επιλογή πεδίων και χρονική κανονικοποίηση	15
2.2 Συλλογή Δεδομένων Καιρού	17
2.3 Συγχώνευση Δεδομένων	18
2.4 Προεπεξεργασία Δεδομένων	20
3 Μεθοδολογία & Υλοποίηση	22
3.1 Μοντέλο Πρόβλεψης της Ταχύτητας	22
3.2 Διαγράμματα Δεδομένων Μοντέλου Πρόβλεψης Ταχύτητας	27
3.2.1 Πολυπληθέστερες Κατηγορίες Τύπων Πλοίων	27
3.2.2 Κατανομή Ποσοστού Καταγραφών Πολυπληθέστερων Τύπων Πλοίων ...	28
3.2.3 Συσχέτιση MAE ανά Τύπο πλοίου	29
3.2.4 Συσχέτιση R ² ανά Τύπο Πλοίου	30
3.2.5 Συνολική Αποτίμηση ανά Τύπο Πλοίου (συνδυασμός MAE & R ²)	31
3.3 Sea Mask	33
3.4 Route Planner	35
4 Αποτελέσματα και Αξιολόγηση	42
4.1 Αξιολόγηση μοντέλου πρόβλεψης της ταχύτητας	42
4.2 Παραδείγματα διαδρομών	45
4.3 Ευαισθησία παραμέτρων δρομολογητή, μέγιστο μήκος ακμής, αραιοποίηση πλέγματος	51

5 Συμπεράσματα & Μελλοντικές Επεκτάσεις	53
5.1 Συμπεράσματα	53
5.2 Περιορισμοί - Προσέγγιση χωρίς χρονική μνήμη και έλλειψη ρευμάτων	55
5.3 Μελλοντικές διορθώσεις	56
Βιβλιογραφία	58

Κατάλογος Εικόνων

Εικόνα 1 : Αξιολόγηση ακρίβειας μοντέλου ανά τύπο πλοίου	25
Εικόνα 2 : Τελική απεικόνιση μέσου σφάλματος και συνολικού R^2 του μοντέλου	26
Εικόνα 3 : Horizontal Barplot με τις πολυπληθέστερες κατηγορίες πλοίων	28
Εικόνα 4 : Vertical Barplot κατανομής ποσοστού καταγραφών πολυπληθέστερων τύπων πλοίων	29
Εικόνα 5 : Horizontal Barplot Συσχέτισης MAE ανά Τύπο πλοίου	30
Εικόνα 6 : Horizontal Barplot Συσχέτισης R^2 ανά Τύπο Πλοίου	31
Εικόνα 7 : Απεικόνιση Sea Mask	33
Εικόνα 8 : Εισαγωγή αρχικού λιμανιού από το χρήστη	36
Εικόνα 9 : Εισαγωγή τελικού λιμανιού από το χρήστη	36
Εικόνα 10 : Εισαγωγή ημερομηνίας δρομολογίου από το χρήστη	36
Εικόνα 11 : Εισαγωγή κωδικού του τύπου του πλοίου από το χρήστη	36
Εικόνα 12 : Επιμέρους στάδια διαδικασίας δρομολόγησης	37
Εικόνα 13 : Καταγραφή συνολικής απόστασης, συνολικού χρόνου και μέσης ταχύτητας	41
Εικόνα 14 : Τελικό αποτέλεσμα χάρτη με σχεδιασμένο δρομολόγιο	41
Εικόνα 15 : Απεικόνιση μετρικών ορισμένων τύπων πλοίων	44
Εικόνα 16 : Δρομολόγιο Port of New York - Port of Norfolk (Ημερομηνία: 06/08/2016) ..	47
Εικόνα 17 : Δρομολόγιο Port of New York - Port of Norfolk (Ημερομηνία: 01/01/2019) - Τύπος πλοίου 30	48
Εικόνα 18 : Δρομολόγιο Port of New York - Port of Norfolk (Ημερομηνία: 01/01/2019) - Τύπος πλοίου 1001	50

Εισαγωγή

1) Στόχος, κίνητρο και ερευνητικά ερωτήματα της εργασίας

Στόχος της παρούσας εργασίας είναι η ανάπτυξη και η τεκμηρίωση ενός πλήρους συστήματος ανάλυσης δεδομένων που συνδυάζει πραγματικά δεδομένα AIS με μετεωρολογικές πληροφορίες, προκειμένου να προβλέψει με ακρίβεια την ταχύτητα πλεύσης πλοίων και να τη χρησιμοποιήσει για τον υπολογισμό του χρόνου και της απόστασης ταξιδιών σε ρεαλιστικές θαλάσσιες διαδρομές. Με άλλα λόγια, η εργασία αποσκοπεί στην πρόβλεψη της πραγματικής ταχύτητα πλεύσης (Speed Over Ground - SOG) ενός πλοίου σε συγκεκριμένο σημείο και χρονική στιγμή, λαμβάνοντας υπόψη τη θέση του, τον τύπο του και τις τοπικές συνθήκες ανέμου/κύματος. Σημαντικό είναι να αναφερθεί ότι η χαμηλή μέση ταχύτητα (SOG) χρησιμοποιείται ως ένδειξη χαμηλότερης ενεργειακής αποδοτικότητας και αυξημένων ενεργειακών απωλειών για ίδιες συνθήκες πλεύσης.

Το κίνητρο της εργασίας είναι η αξιοποίηση της ανάλυσης δεδομένων, ώστε να παραχθούν αξιόπιστες εκτιμήσεις ταχύτητας και να υποστηριχθεί ένας ρεαλιστικός Route Planner και επιστρέφει μετρήσιμα μεγέθη (απόσταση ταξιδιού, εκτιμώμενος χρόνος άφιξης - ETA (Estimated Time of Arrival), μέση ταχύτητα - SOG (Speed Over Ground)) για τη λήψη αποφάσεων. Επιπλέον, κίνητρο για την συγκεκριμένη εργασία είναι και η παρουσίαση των αποτελεσμάτων με τρόπο αξιοποιήσιμο από μη ειδικούς με απλές χαρτογραφικές απεικονίσεις (διαδρομή με ένδειξη ταχύτητας με χρωματική απεικόνιση).

Τα ερευνητικά ερωτήματα που καλείται να απαντήσει η εργασία είναι τα εξής:

- Η ακρίβεια της πρόβλεψης της ταχύτητας δεδομένης της θέσης, του τύπου του πλοίου και τον υφιστάμενο καιρό. Ουσιαστικά, διερευνάται η ικανότητα του μοντέλου να χαρτογραφεί αυτά τα στοιχεία προβλέποντας την αναμενόμενη ταχύτητα του πλοίου, εξαγοντας βασικές μετρικές (MAE σε knots και R^2). Η ανάλυση συνοδεύεται από

αποτελέσματα ανά τύπο πλοίου, ώστε να αποτυπωθεί η συμπεριφορά σε πολυπληθείς και λιγότερο πολυπληθείς κλάσεις.

- Ο τρόπος μετάφρασης μίας ακριβούς πρόβλεψης ταχύτητας σε αξιόπιστο χρόνο άφιξης. Εν ολίγοις, στην εργασία αξιολογείται με ποιο τρόπο η ταχύτητα ανά τμήμα επηρεάζει τον συνολικό χρόνο της διαδρομής. Η απάντηση δίνεται με τη χρήση της ένδειξης του μήκους του σκέλους/συνόλου διαδρομής, ταχύτητα σκέλους/μέση ταχύτητα διαδρομής και χρόνο σκέλους/συνολικό, ώστε να είναι σαφές το πρακτικό όφελος της πρόβλεψης.

Με αυτόν τον ορισμό στόχων και ερωτημάτων, η εργασία δεν περιορίζεται σε μία μεμονωμένη τεχνική, αλλά παρουσιάζει μια πλήρη αλυσίδα χρήσιμων αποτελεσμάτων, από την προετοιμασία των δεδομένων, στην πρόβλεψη ταχύτητας και από εκεί στη σχεδίαση του δρομολογίου με μετρήσιμα αποτελέσματα.

2) Γεωγραφικό πεδίο και χρονικό εύρος της μελέτης - Παραδοχές

Όσον αφορά την γεωγραφική περιοχή που μελετάται, η μελέτη εστιάζει στην παράκτια ζώνη του Βορειοδυτικού Ατλαντικού στο ορθογώνιο πλαίσιο που περιγράφεται από τις γεωγραφικές συντεταγμένες: LAT: 35°-42° και LON: -77° έως -70°. Όλες οι διεργασίες (συλλογή δεδομένων AIS, εξαγωγή μετεωρολογικών στοιχείων, συγχώνευση αυτών, κατασκευή της περιοχής της στεριάς - sea mask, δρομολόγηση μεταξύ δύο λιμανιών της περιοχής) εφαρμόζονται εντός αυτού του γεωγραφικού πλαισίου, ώστε τα αποτελέσματα να είναι συνεπή χωρικά και να ελέγχεται το υπολογιστικό κόστος.

Αναφορικά με το χρονικό εύρος, χρησιμοποιούνται δεδομένα μεταξύ των ετών 2015-2024, ενώ ορίζεται κοινό χρονικό βήμα 15 λεπτών για να είναι συγκρίσιμα όλα τα στιγμιότυπα σε δεδομένα AIS και τα στοιχεία καιρού. Η εκπαίδευση του μοντέλου για την πρόβλεψη της ταχύτητας γίνεται ενιαία πάνω στο σύνολο της δεκαετίας, με τον τυπικό διαχωρισμό train/test σε αναλογία 80/20.

Οι παραδοχές που υιοθετούνται ρητά είναι οι κάτωθι:

- Ο τύπος του πλοίου λειτουργεί ως αντιπρόσωπος κατηγορίας/μεγέθους πλοίου και παραμένει σταθερό κατά τη δρομολόγηση.
- Στη δρομολόγηση (routing), κάθε σκέλος αντιπροσωπεύει την πλεύση πάνω στη θάλασσα, άρα αποκλείονται πλοία με πολύ χαμηλή ταχύτητα (<4 kts) για να προκύπτουν προβλέψεις ταχύτητας εν πλω, χωρίς να επηρεάζεται η μελέτη από στάσιμα πλοία.

Με αυτόν τον ορισμό πεδίου και παραδοχών, η εργασία περιέχει καθαρά δεδομένα, αξιόπιστη πρόβλεψη ταχύτητας και δρομολόγηση πάνω σε θαλάσσιο γράφο που αποδίδει απόσταση, χρόνο και μέση ταχύτητα για τεκμηριωμένες αποφάσεις.

3) Αρχιτεκτονική της εργασίας

Αναφορικά με την αρχιτεκτονική της εργασίας, σε γενικό πλαίσιο σχεδιάζεται και υλοποιείται ένα πλήρες pipeline δεδομένων για την περιοχή της ΒΑ Αμερικής, με χρήση δεδομένων της περιόδου 2015-2024. Η ροή περιλαμβάνει τη συλλογή και τον καθαρισμό δεδομένων πλοίων βάσει των στοιχείων AIS των πλοίων στην περιοχή ενδιαφέροντος (MMSI, χρόνος καταγραφής των δεδομένων, γεωγραφικό πλάτος (LAT), γεωγραφικό μήκος (LON), ταχύτητα (SOG), τύπος πλοίου), ενοποίηση με μετεωρολογικά μεγέθη (ταχύτητα ανέμου και ύψος κύματος) και ευθυγράμμιση του χρόνου για καλύτερη αντιστοίχιση δεδομένων AIS με δεδομένα καιρού ανά 15 λεπτά. Επίσης, χρησιμοποιήθηκε απλό αλλά αποτελεσματικό feature engineering (ενδεικτικά, χρησιμοποιήθηκε το γινόμενο της έντασης ανέμου επί το ύψος του κύματος και ο δείκτης που υποδεικνύει αν το πλοίο θεωρείται σχεδόν ακίνητο). Το αποτέλεσμα είναι ετήσια σύνολα δεδομένων σε μορφή κατάλληλη για μοντελοποίηση και χρήση σε περαιτέρω βήματα.

Δεύτερον, εκπαιδεύτηκε ένα μοντέλο παλινδρόμησης Random Forest που χαρτογραφεί τα χαρακτηριστικά που αναφέρθηκαν παραπάνω στην ταχύτητα πλεύσης SOG. Το

μοντέλο επιτυγχάνει πολύ ικανοποιητικές μετρικές ($MAE = 1.14$ knots και $R^2 = 0.85$) σε ενιαία εκπαίδευση για όλα τα έτη, επιβεβαιώνοντας ότι η συνδυαστική πληροφορία θέσης - τύπου- καιρού επαρκεί για υψηλής ποιότητας εκτιμήσεις, έχοντας ποσοστό ακρίβειας 85% στις προβλέψεις του μοντέλου με μέσο σφάλμα περίπου ένα κόμβο. Παράλληλα, παρέχονται αναλυτικά αποτελέσματα ανά τύπο πλοίου, ώστε να γίνεται σαφές πού το μοντέλο αποδίδει καλύτερα και πώς οι διαφορετικές κατηγορίες πλοίων ανταποκρίνονται στις συνθήκες.

Τρίτον, ενσωματώθηκαν οι προβλέψεις της ταχύτητας σε έναν δρομολογητή που λειτουργεί πάνω σε γράφο θαλάσσιων κόμβων/ακμών. Η επιφάνεια της θάλασσας μοντελοποιείται με ανάλυση της επιφάνειας του χάρτη 0.01° , ώστε να αποκλείονται τα σημεία ξηράς και το σχεδιασμένο δρομολόγιο να διέρχεται μόνο από τη θάλασσα. Η εύρεση της βέλτιστης πορείας μεταξύ λιμένων γίνεται με τον αλγόριθμο A^* . Σε κάθε τμήμα της διαδρομής, το μοντέλο Random Forest προβλέπει την τοπική ταχύτητα από τις μετεωρολογικές τιμές, επιτρέποντας τον υπολογισμό του χρόνου κίνησης του πλοίου ανά τμήμα και αθροιστικά, της συνολικής απόστασης, του συνολικού χρόνου ταξιδιού και της μέσης ταχύτητας. Έτσι, γεφυρώνεται η πρόβλεψη της ταχύτητας με τη χωρική δρομολόγηση σε ένα ενιαίο σύστημα, κάτι το οποίο είναι πολύ χρήσιμο για χρήση από μία εταιρία που επιζητεί συγκεκριμένα νούμερα, απεικόνιση σε χάρτη και συγκεκριμένη διαδρομή.

Συνοψίζοντας, η εργασία προσφέρει: (α) ένα καθαρό και εφαρμόσιμο πλαίσιο επεξεργασίας AIS σε συνδυασμό με τον καιρό, (β) ένα αποδοτικό και ακριβές μοντέλο πρόβλεψης ταχύτητας, και (γ) έναν δρομολογητή που χρησιμοποιεί τις προβλέψεις για να σχεδιάσει ρεαλιστικές διαδρομές με εκτιμώμενη ταχύτητα και χρόνο. Το αποτέλεσμα είναι ένα εργαλείο έτοιμο να υποστηρίξει τεκμηριωμένες αποφάσεις στη θαλάσσια μεταφορά, με έμφαση στην ενεργειακή αποδοτικότητα και την αξιοπιστία του χρονοπρογραμματισμού.

4) Δομή εργασίας

Η εισαγωγή εισάγει το αντικείμενο και το πλαίσιο της μελέτης, αφού παρουσιάζει το στόχο, το κίνητρο και τα ερευνητικά ερωτήματα της εργασίας, το γεωγραφικό πεδίο και το χρονικό εύρος της μελέτης, τις βασικές παραδοχές, ενώ κλείνει με τη χαρτογράφηση της αρχιτεκτονικής της εργασίας από τη λήψη των δεδομένων μέχρι την τελική δρομολόγηση και με την δομή της εργασίας. Το κεφάλαιο του θεωρητικού υποβάθρου συγκροτεί το θεωρητικό υπόβαθρο, εξηγώντας συνοπτικά τον τρόπο λειτουργίας των συστημάτων AIS και τις μεταβλητές του (ιδίως ταχύτητα), αποσαφηνίζει τις μετεωρολογικές μεταβλητές που χρησιμοποιούνται (ταχύτητα ανέμου, ύψος κύματος) και περιγράφει τις βασικές έννοιες μηχανικής μάθησης για παλινδρόμηση με τον αλγόριθμο Random Forest, καθώς και τις αρχές της θαλάσσιας δρομολόγησης πάνω σε γράφο με sea mask και αλγόριθμο A*. Το δεύτερο κεφάλαιο εστιάζει στα δεδομένα και την προεπεξεργασία. Αρχικά αναλύεται ο τρόπος συλλογής και η οργάνωση των ημερήσιων αρχείων δεδομένων AIS, η εξαγωγή των μετεωρολογικών μετρήσεων για την ίδια περιοχή, η συγχώνευση των δεδομένων που οδηγεί στα τελικά ετήσια σύνολα δεδομένων για μοντελοποίηση, καθώς και η προεπεξεργασία, πριν την είσοδο των δεδομένων στο μοντέλο πρόβλεψης. Στο τρίτο κεφάλαιο αναλύεται η μεθοδολογία και η υλοποίηση της δρομολόγησης. Για να επιτευχθεί η δρομολόγηση, περιγράφηκε το μοντέλο πρόβλεψης της ταχύτητας (χαρακτηριστικά, εκπαίδευση, υπερπαραμέτροι), σχεδιάστηκαν διαγράμματα των δεδομένων του μοντέλου και αναλύθηκε η διαδικασία κατασκευής της sea mask, του γράφου κόμβων/ακμών και της εύρεσης διαδρομής με A*. Το τέταρτο κεφάλαιο παρουσιάζει τα αποτελέσματα και την αξιολόγηση του μοντέλου με τις σχετικές συνολικές επιδόσεις του μοντέλου (MAE, R^2), την ανάλυση ανά κατηγορία πλοίου και παραδείγματα διαδρομών με συνοδευτικές απεικονίσεις. Τέλος, το πέμπτο κεφάλαιο συνοψίζει τα συμπεράσματα, αποτυπώνει τη χρησιμότητα του συστήματος για εκτιμήσεις ταχύτητας και χρόνου σε θαλάσσιες διαδρομές, παραθέτει τους περιορισμούς που επηρεάζουν τη ρεαλιστική χρησιμότητα του μοντέλου και σκιαγραφεί δυνητικές κατευθύνσεις για περαιτέρω εμβάθυνση.

1 Θεωρητικό υπόβαθρο

1.1 Τρόπος λειτουργίας συστημάτων AIS

Το Automatic Identification System (AIS) είναι το κυριότερο ναυτιλιακό σύστημα ασφάλειας και επίγνωσης κατάστασης στις θαλάσσιες μεταφορές, το οποίο μεταδίδει αυτόματα πληροφορίες που αφορούν την ταυτότητα, τη θέση και την κίνηση κάθε πλοίου μέσω VHF αυτόματα. Το σύστημα αυτό επιτρέπει τη χρήση των εκπεμπόμενων πληροφοριών από άλλα πλοία και παράκτια κέντρα ελέγχου κυκλοφορίας (VTS) για αποφυγή συγκρούσεων και μεγαλύτερη ασφάλεια. Ο κανονισμός του IMO (International Maritime Organization) που επιβάλλει την υποχρεωτική εγκατάσταση του AIS σε πολλές κατηγορίες πλοίων είναι ο SOLAS V/19, καθιστώντας το AIS το βασικό ναυτιλιακό εξοπλισμό πλοήγησης (International Maritime Organization, n.d.).

Κάθε πλοίο/σταθμός AIS συνδέεται με ένα MMSI (Maritime Mobile Service Identity), δηλαδή έναν εννιαψήφιο αναγνωριστικό κωδικό που χρησιμοποιείται για την μοναδική ταυτοποίηση του πλοίου/σταθμού (Federal Communications Commission, 2022). Το MMSI εκχωρείται επίσημα από τις αρμόδιες ρυθμιστικές αρχές και προγραμματίζεται στη συσκευή AIS (United States Coast Guard Navigation Center, n.d.). Τα γεωχωρικά δυναμικά δεδομένα που μεταδίδει το σύστημα AIS προέρχονται συνήθως από τους ναυτιλιακούς αισθητήρες του πλοίου (π.χ. GNSS για θέση/ταχύτητα, γυροπυξίδα για πορεία) και συγκροτούν τις αναφορές θέσεως των μηνυμάτων κλάσης A (AIS Messages 1/2/3). Στα μηνύματα αυτά περιλαμβάνονται:

- Γεωγραφικό πλάτος/μήκος (LAT/LON) στο παγκόσμιο γεωδαιτικό σύστημα συντεταγμένων WGS-84
- Speed Over Ground (SOG), δηλαδή η ταχύτητα πλοίου σε βήματα 0.1 κόμβου (εύρος 0-102.2 kts) και
- Course Over Ground (COG), δηλαδή η γωνία πορείας ως προς το έδαφος.

Οι ίδιες αναφορές θέσεις (position reports) καταγράφουν και την πορεία του πλοίου (Heading) (όταν αυτή είναι διαθέσιμη), μαζί με άλλα πεδία (π.χ. ναυτιλιακή κατάσταση) (United States Coast Guard Navigation Center, n.d.).

Είναι κρίσιμο να διαχωριστούν οι έννοιες:

- SOG (Speed Over Ground): η πραγματική ταχύτητα του σκάφους ως προς την επιφάνεια της Γης (όχι ως προς το νερό). Είναι το μέτρο της μεταβολής θέσης στο έδαφος ανά μονάδα χρόνου, γι' αυτό και χρησιμοποιείται εκτενώς σε εκτίμηση του χρόνου άφιξης (United States Coast Guard Navigation Center, n.d.).
- COG (Course Over Ground): η πραγματική κατεύθυνση κίνησης του σκάφους πάνω στο έδαφος (σε μοίρες), που μπορεί να διαφέρει από την πορεία που διατηρεί η πλήρη του πλοίου (heading) λόγω ρευμάτων/ανέμων.

Στις οδηγίες της IALA για τη χρήση AIS σε VTS αναφέρεται ρητά ο ρόλος της πορείας και της ταχύτητας στην απεικόνιση της κίνησης και στη διαχείριση της κυκλοφορίας των πλοίων (International Association of Marine Aids to Navigation and Lighthouse Authorities [IALA], 2024).

Τα στατικά και ταξιδιωτικά δεδομένα (όνομα πλοίου, τύπος, διαστάσεις, προορισμός κ.λ.π.) αποστέλλονται με ξεχωριστά μηνύματα (π.χ. Message 5 για μηνύματα κλάσεως A) σε αραιότερα διαστήματα έναντι των αναφορών θέσεως, έτσι ώστε το φορτίο του καναλιού VHF να αφιερώνεται κατά προτεραιότητα στην ενημέρωση της θέσης και της κίνησης. Η τεχνικές προδιαγραφές και τα πρωτόκολλα της λειτουργίας του συστήματος AIS ορίζονται στο έγγραφο "ITU-R M.1371", το οποίο επικαλούνται διεθνείς φορείς, όπως οι IMO/IALA/USCG.

Οι ρυθμοί μετάδοσης εξαρτώνται από την κλάση της συσκευής που χρησιμοποιεί το πλοίο:

- Κλάση A: όταν το πλοίο κινείται («underway»), εκπέμπει κάθε δύο έως δέκα δευτερόλεπτα (ταχύτερα σε υψηλότερες ταχύτητες ή μεγάλες αλλαγές πορείας και ταχύτητας). Όταν το πλοίο είναι αγκυροβολημένο ή ακίνητο, εκπέμπει κάθε περίπου τρία λεπτά. Αυτές οι τιμές και η περιγραφή των πεδίων (όπως ο βηματισμός της ταχύτητας που ορίζεται στο 0.1 kn) τεκμηριώνονται από τον κανονισμό USCG Navigation Center στην ενότητα «Class A AIS Position Report» (United States Coast Guard Navigation Center, n.d.).
- Κλάση B (CS): σύμφωνα με το προαναφερθέν ITU-R M.1371, οι αναφορές θέσεως μεταδίδονται ανά 30 δευτερόλεπτα όταν η ταχύτητα είναι μεγαλύτερη από δύο κόμβους και ανά τρία λεπτά όταν η ταχύτητα είναι μικρότερη ή ίση από δύο κόμβους.

Το σύστημα AIS λειτουργεί σε προκαθορισμένες VHF συχνότητες (κανάλια 87B/88B) και χρησιμοποιεί TDMA σχήμα χρονοθυρίδων, ώστε χιλιάδες σταθμοί να μοιράζονται αποτελεσματικά το μέσο. Ο IMO υπογραμμίζει ότι το AIS είναι μέρος του υποχρεωτικού εξοπλισμού πλοήγησης (βάσει κανονισμών SOLAS), ενώ η IALA παρέχει κατευθυντήριες οδηγίες για την αξιοποίησή του AIS σε VTS και ευρύτερα στη ναυσιπλοΐα, όπου οι τιμές της πορείας και της ταχύτητας είναι θεμελιώδεις για την απεικόνιση των τροχιών και την αποφυγή των συγκρούσεων (IALA, n.d.).

Συνοψίζοντας, στο πλαίσιο της παρούσας μελέτης, το MMSI λειτουργεί ως σημείο διαφοροποίησης κάθε πλοίου στα δεδομένα AIS, οι γεωγραφικές συντεταγμένες (LAT/LON) παρέχουν τη χωρική αναφορά για ανάλυση/σύντηξη με άλλα πλέγματα (π.χ. καιρού), η ταχύτητα (SOG) (σε κόμβους) δίνει το φυσικό μέγεθος που το μοντέλο προβλέπει και είναι άμεσα συνδεδεμένο με την εκτίμηση διάρκειας διαδρομής, καθώς και η πορεία (COG) παρέχει το διανυσματικό πλαίσιο κίνησης στο έδαφος, συμπληρώνοντας τη ταχύτητα για την κατανόηση των δρομολογιών. Οι τυποποιημένοι ρυθμοί μετάδοσης (συχνά δευτερόλεπτα για μηνύματα A κλάση εν κινήσει, ή/και λεπτά για πλοία σε αγκυροβολία/βραδεία κίνηση ή μηνύματα κλάσης B) εξηγούν γιατί τα δεδομένα AIS διαθέτουν υψηλή χρονική ανάλυση, το οποίο αποτελεί στοιχείο στην

παρούσα εργασία αξιοποιείται για τον συγχρονισμό της θέσης/κίνησης των πλοίων με εξωτερικά πλέγματα περιβάλλοντος.

1.2 Καιρικά στοιχεία (ένταση ανέμου και ύψος κύματος)

Στη μελέτη που εκπονήθηκε χρησιμοποιήθηκαν δύο τυποποιημένες μετεωρολογικές μεταβλητές, οι οποίες είναι η ένταση/ταχύτητα του ανέμου (wind speed) και το ύψος του κύματος (wave height). Και οι δύο μεταβλητές προέρχονται από δίκτυα μετρήσεων, τα οποία διατίθενται από το NOAA NDBC (National Oceanic and Atmospheric Administration - National Data Buoy Center), το οποίο υπάγεται στην κυβέρνηση των ΗΠΑ, και της διεθνούς πρακτικής, διαθέτοντας σαφείς ορισμούς και μονάδες (National Data Buoy Center, n.d.).

- Ένταση ανέμου (Wind Speed): Η ταχύτητα του ανέμου είναι το διανυσματικό μέτρο της ροής αέρα πάνω από τη θάλασσα (m/s ή kts) αναγόμενη σε ύψος αναφοράς 10 m με καθιερωμένες μεθόδους μετατροπής. Η χρήση της σε ύψος εξασφαλίζει συγκρισιμότητα μεταξύ σταθμών, αλλά και μεταξύ των ετών και είναι η συνήθης πρακτική στη μετεωρολογία θαλάσσης (National Data Buoy Center, n.d.; Integrated Ocean Observing System, n.d.).
- Ύψος κύματος (Wave Height): Το ύψος του κύματος ορίζεται ως ο μέσος όρος του ανώτερου ενός τρίτου των κυμάτων ($H_{1/3}$), μετρημένος από κοιλάδα σε κορυφή. Η ίδια ποσότητα μπορεί να οριστεί στο πεδίο συχνοτήτων ως $H_{m0} = 4V_{m0}$, δηλαδή τέσσερις φορές τη ρίζα του μηδενικού φασματικού ροπής (διασποράς) της ανύψωσης επιφάνειας. Οι δύο ορισμοί ευθυγραμμίζονται σε τιμές που διαφέρουν λίγες μόνο ποσοστιαίες μονάδες. Το NDBC περιγράφει ρητά το ύψος κύματος με βάση αυτές τις έννοιες, ενώ άλλοι φορείς (όπως η BOM) το χρησιμοποιούν ως μέτρο της θαλάσσιας κατάστασης με μονάδα μέτρα (m) (National Data Buoy Center, n.d.; Bureau of Meteorology, n.d.).

Οι λόγοι που ο άνεμος και τα κύματα επηρεάζουν την ταχύτητα είναι κατά βάση δύο από πλευράς ναυπηγικής υδροδυναμικής:

(α) η πρόσθετη αντίσταση από τον άνεμο (aerodynamic windage/air drag), ιδίως σε headwinds/crosswinds, και

(β) η πρόσθετη αντίσταση λόγω κυματισμού (added resistance in waves), που οφείλεται στη σχετική κίνηση πλοίου-κύματος (συμπεριλαμβανομένων ταλαντώσεων και φαινομένων όπως re-radiation).

Η συνέπεια είναι ότι, για δεδομένη ισχύ πρόωσης, η πραγματική ταχύτητα ως προς το έδαφος (SOG) τείνει να μειώνεται σε δυσμενή άνεμο/κύμα. Εναλλακτικά, για τη διατήρηση της ίδιας SOG απαιτείται αυξημένη ισχύς. Η βιβλιογραφία τεκμηριώνει ότι η πρόσθετη αντίσταση λόγω κυματισμού είναι από τους κύριους λόγους αύξησης απαιτούμενης ισχύος σε πραγματικές συνθήκες (Martić et al., 2020), άρα και βασικός παράγοντας απώλειας ταχύτητας/έξτρα κατανάλωσης (Chirosca, 2023).

Η σχέση έντασης ανέμου και ύψους κύματος με την απόδοση πλου είναι διττή. Πρώτον, ο ισχυρός άνεμος αυξάνει την αεροδυναμική αντίσταση και το αποτέλεσμα είναι μικρότερη ταχύτητα για την ίδια ισχύ ή ανάγκη αύξησης ισχύος για διατήρηση του χρονοδιαγράμματος. Δεύτερον, τα υψηλότερα κύματα αυξάνουν την υδροδυναμική πρόσθετη αντίσταση (heave, pitch, slamming/green water σε ακραίες καταστάσεις), επιδεινώνοντας την απώλεια ταχύτητας, ιδίως σε όταν είναι μικρό το μήκος του κύματος. Πειραματικές μελέτες παρουσιάζουν ποσοτικοποιήσεις της απώλειας ταχύτητας και της μεταβολής απαιτούμενης ισχύος υπό διαφορετικά θαλάσσια κράτη, τεκμηριώνοντας την πρακτική σημασία των δύο μεταβλητών (Wang et al., 2019; Mittendorf, 2023).

Ο επίσημος ναυτικός οργανισμός (IMO) επεσήμανε το 1983 (IMO A.528(13)) ότι η επιλογή διαδρομών με βάση τις καιρικές συνθήκες βελτιώνει την ασφάλεια και τη λειτουργικότητα του ταξιδιού, μια πολιτική που προϋποθέτει την ύπαρξη μεγεθών όπως οι δύο αυτές μεταβλητές και τη χρήση τους σε επιχειρησιακές αποφάσεις (International Maritime Organization, 1983).

Στο μοντέλο που υλοποιεί η εργασία, οι δύο μεταβλητές (ένταση ανέμου και ύψος κύματος) εισέρχονται ως αριθμητικές είσοδοι στο μοντέλο παλινδρόμησης και λειτουργούν:

- ως δείκτες περιβαλλοντικής έντασης (με μονάδες μέτρησης m/s και m αντίστοιχα).
- ως συνδυαστικό χαρακτηριστικό μέσω του γινομένου των δύο μεταβλητών, που συμπυκνώνει την ένταση και τη συχνότητα της επίδρασης της θάλασσας στον πλου. Με αυτόν τον τρόπο, το μοντέλο εκπαιδεύεται στις μη γραμμικές σχέσεις μεταξύ ανέμου και κύματος, γεωγραφίας και τύπου πλοίου, και επιστρέφει ρεαλιστικές εκτιμήσεις ταχύτητας που συνδέονται άμεσα με χρόνο ταξιδιού (ETA) και απόσταση πάνω στη σχεδιασθείσα διαδρομή (National Data Buoy Center, n.d.).

1.3 Μοντέλο Παλινδρόμησης Random Forest και βασικές υπερπαράμετροι

Το μοντέλο παλινδρόμησης Random Forest είναι ένα σύνολο (ensemble) από δέντρα απόφασης που εκπαιδεύονται σε διαφορετικά δείγματα του ίδιου συνόλου δεδομένων και, σε κάθε διαχωρισμό (split), εξετάζουν τυχαίο υποσύνολο γνωρισμάτων. Η τελική εκτίμηση δίνεται από το μέσο όρο των επιμέρους προβλέψεων των δέντρων. Η τυχαιοποίηση στο επίπεδο των δειγμάτων (bootstrap) και των γνωρισμάτων (random subspace) μειώνει τη συσχέτιση στα δέντρα και τη διακύμανση του μοντέλου, διατηρώντας την εκφραστικότητα των δέντρων ως βασικών “μαθητών”. Η θεμελιώδης εργασία του Breiman (2001) έδειξε ότι το σφάλμα γενίκευσης του δάσους συγκλίνει καθώς αυξάνεται ο αριθμός των δέντρων και εξαρτάται από τη “ισχύ” των δέντρων και τη μεταξύ τους συσχέτιση (Breiman, 2001).

Για την εκπαίδευση χρησιμοποιούνται δέντρα CART (Classification And Regression Trees), τα οποία στην παλινδρόμηση επιλέγουν διαχωρισμούς που ελαχιστοποιούν τη μέση τετραγωνική απόκλιση (MSE) εντός κόμβων. Η μονοτονία του κριτηρίου (impurity) οδηγεί σε “καθαρότερους” κόμβους, ενώ το δέντρο επεκτείνεται μέχρι κάποιον περιορισμό (π.χ. μέγιστο βάθος) ή μέχρι οι βελτιώσεις να γίνουν αμελητέες. Το κλασικό έργο των Breiman-Friedman-Olshen-Stone (1984) τυποποίησε τη μεθοδολογία CART, ενώ τα

σύγχρονα εγχειρίδια αποτυπώνουν το αντίστοιχο κριτήριο για παλινδρόμηση (Breiman et al., 1984).

Από πλευράς bias-variance, τα δέντρα έχουν μικρό bias (μαθαίνουν περίπλοκες, μη γραμμικές σχέσεις), αλλά έχουν συχνά υψηλή διακύμανση (ευαισθησία σε μικρές αλλαγές των δεδομένων). Η ιδέα του Random Forest είναι ότι, αν εκπαιδευτούν πολλά “θορυβώδη” αλλά ασυσχέτιστα δέντρα και υπολογιστεί ο μέσος όρος, η συνολική διακύμανση μειώνεται δραστικά χωρίς να εκτινάσσεται το bias. Στην πράξη, ο αλγόριθμος Random Forest Regressor υλοποιείται με το παρακάτω σύνολο υπερπαραμέτρων, που ελέγχουν την πολυπλοκότητα και την τυχαιοποίηση:

- `n_estimators`: αριθμός δέντρων που περιέχει το δάσος. Μεγαλύτερες τιμές συνήθως μειώνουν τη διακύμανση και σταθεροποιούν την απόδοση, με κόστος χρόνου και μνήμης.
- `max_depth` (και συναφή `min_samples_split`, `min_samples_leaf`): ελέγχουν με πόση λεπτομέρεια διακλαδίζονται τα δέντρα. Ουσιαστικά, το μικρό βάθος αυξάνει το bias, ενώ το μεγάλο βάθος αυξάνει variance.
- `max_features`: μέγεθος του τυχαίου υποσυνόλου γνωρισμάτων που εξετάζεται σε κάθε split (π.χ. "sqrt"). Μικρό `max_features` δυσκολεύει περισσότερο τη συσχέτιση στα δέντρα (καλό για variance), αλλά μπορεί να αυξήσει το bias αν γίνει υπερβολικά μικρό.
- `bootstrap` και `max_samples`: αν θα εκπαιδευτεί κάθε δέντρο σε bootstrap δείγμα και με τι μέγεθος δείγματος.
- `oob_score`: εκτίμηση σφάλματος χωρίς ξεχωριστό validation, χρήσιμη για γρήγορη αξιολόγηση.
- `random_state`, `n_jobs`: αφορούν την αναπαραγωγικότητα και την παράλληλη εκτέλεση.

Τα παραπάνω (και άλλα, όπως `criterion="squared_error"`) τεκμηριώνονται αναλυτικά στη βιβλιοθήκη `scikit-learn`.

Η θεωρία του Breiman υποδεικνύει γιατί η προσθήκη δέντρων δεν οδηγεί σε `overfitting` με τον συνηθισμένο τρόπο: καθώς το πλήθος δέντρων αυξάνει, το σφάλμα γενίκευσης τείνει σε όριο που καθορίζεται από την μέση συσχέτιση των δέντρων και την ισχύ (`predictive strength`) των επιμέρους μαθητών (δέντρων). Ενστικτωδώς, όσο πιο “ισχυρό” είναι κάθε δέντρο αλλά ασυσχέτιστο από τα υπόλοιπα, τόσο καλύτερα η μέση πρόβλεψη ακυρώνει το θόρυβο. Αυτός είναι και ο λόγος που η τυχαιοποίηση στο επίπεδο των γνωρισμάτων (`max_features`) είναι κρίσιμη, ώστε να μειώνει τη συσχέτιση, άρα και τη συνολική διακύμανση του δάσους (Breiman et al., 1984).

Εξετάζοντας τον τρόπο λειτουργίας του αλγορίθμου, προκύπτει ότι ο `Random Forest` είναι μια ισχυρή και πρακτική επιλογή για παλινδρόμηση σε ετερογενή, μη γραμμικά δεδομένα, δεδομένου ότι ελέγχεται εύκολα μέσω υπερπαραμέτρων, κλιμακώνεται παράλληλα και μειώνει τη διακύμανση μέσω της τυχαιοποίησης.

1.4 Sea mask και γράφος κόμβων/ακμών με A*

Για να χαραχθούν οι διαδρομές χωρίς αυτές να διασχίζουν την ξηρά, κατασκευάστηκε μια μάσκα θάλασσας (`sea mask`), η οποία είναι ένα πλέγμα (`raster`) όπου κάθε κελί του πλέγματος χαρακτηρίζεται ως θάλασσα ή ξηρά. Πρακτικά, χρησιμοποιούνται πολυγωνικά περιγράμματα ξηράς υψηλής ποιότητας (π.χ. `Natural Earth 1:10m Land`), τα οποία θεωρούνται σημείο αναφοράς για θεματικούς χάρτες και εφαρμογές χωρικής ανάλυσης. Αυτά τα πολύγωνα καλύπτουν την επιφάνεια της ξηράς και μπορούν να κοπούν στην εκάστοτε περιοχή του δρομολογίου, πριν μετατραπούν (με δειγματοληψία) σε δυαδικό πλέγμα που σηματοδοτεί `True=θάλασσα / False=ξηρά` (Natural Earth, n.d.).

Ο έλεγχος αν ένα σημείο του πλέγματος είναι πάνω σε ξηρά γίνεται με γεωμετρικές πράξεις πάνω στα πολυγωνικά περιγράμματα. Για απόδοση, σε βιβλιοθήκες όπως το `Shapely` συνηθίζεται η χρήση των `prepared geometries` (προετοιμασμένες γεωμετρίες)

που διατηρούν εσωτερικούς ελέγχους τμημάτων και επιταχύνουν μαζικούς χωρικούς ελέγχους (αν για παράδειγμα περιέχεται ή αν τέμνεται). Αυτό είναι ιδιαίτερα χρήσιμο όταν αξιολογούνται χιλιάδες σημεία ενός πλέγματος ή πολλές ακμές του γράφου (Shapely, n.d.).

Πάνω στη μάσκα θάλασσας (sea mask) χτίζεται ένας γράφος κόμβων/ακμών. Οι κόμβοι είναι επιλεγμένες θέσεις (π.χ. υποσύνολο των θαλάσσιων κελιών ενός πλέγματος ανάλυσης 0.01°), ενώ οι ακμές συνδέουν γειτονικούς κόμβους εντός ενός μέγιστου γεωδαιτικού μήκους (π.χ. ≤ 25 ν.μ.). Για κάθε υποψήφια ακμή εφαρμόζεται έλεγχος ευθείας οράσεως (Line Of Sight - LOS), η οποία είναι η γεωμετρική ευθεία (ή δείγματα πάνω στην ευθεία) που πρέπει να μην τέμνει την ξηρά. Αν υπάρχει τομή με τη στεριά, η ακμή αυτή απορρίπτεται και αναζητείται άλλη. Με αυτόν τον τρόπο, ο γράφος τηρεί την ακτογραμμή και τις νησίδες, επιτρέποντας μόνο θαλάσσιες συνδέσεις.

Η αναζήτηση διαδρομής εκτελείται με τον αλγόριθμο A^* , ο οποίος χρησιμοποιείται συχνά για την εύρεση της συντομότερης διαδρομής σε γράφους με βάρη. Ο αλγόριθμος A^* επεκτείνει τον αλγόριθμο Dijkstra και εγγυάται τη βέλτιστη λύση, ενώ εξερευνά αισθητά λιγότερους κόμβους από τον Dijkstra. Η θεμελιώδης εργασία των Hart-Nilsson-Raphael (1968) διατυπώνει αυστηρά τις προϋποθέσεις και τις εγγυήσεις για τον βέλτιστο και αποδοτικό τερματισμό της διαδικασίας, ενώ σύγχρονες βιβλιοθήκες όπως το NetworkX προσφέρουν τεκμηριωμένες υλοποιήσεις του `astar_path` (Hart, Nilsson, & Raphael, 1968)

Συνολικά, ο συνδυασμός (1) της μάσκας θάλασσας από έγκυρα πολυγωνικά δεδομένα, (2) του γράφου που τηρεί την ακτογραμμή μέσω LOS (Line Of Sight), και (3) του αλγορίθμου A^* με χρήση της συνάρτησης `haversine` παρέχει μια αποδοτική και θεωρητικά τεκμηριωμένη μεθοδολογία δρομολόγησης σε θαλάσσιους χώρους. Οι επιλογές αυτές είναι καθιερωμένες στην ψηφιακή χαρτογραφία και στην υπολογιστική γεωμετρία και μπορούν να υλοποιηθούν με εργαλεία ανοιχτού κώδικα.

2 Δεδομένα και Προεπεξεργασία

2.1 Συλλογή Δεδομένων AIS - Λήψη αρχείων, επιλογή πεδίων και χρονική κανονικοποίηση

Η συλλογή των δεδομένων είναι το βασικότερο μέλος όλης της ροής, καθώς αποθηκεύονται ημερήσια αρχεία δεδομένων AIS και μετατρέπονται σε καθαρά και ομογενοποιημένα σύνολα. Ο στόχος της διαδικασίας συλλογής των δεδομένων είναι διπλός. Αρχικά, γίνεται αξιόπιστη εισαγωγή μεγάλου όγκου δεδομένων με έλεγχο σφαλμάτων και δευτερευόντως εκτελείται τυποποίηση της χρονικής δειγματοληψίας και ταξινόμηση σε 15 λεπτά, ώστε να είναι εφικτή η σύζευξη με τα αντίστοιχα δεδομένα καιρού και η εκπαίδευση του μοντέλου πρόβλεψης της ταχύτητας. Τα στάδια εκτέλεσης της συλλογής των δεδομένων AIS είναι ως εξής:

1) Δημιουργία URLs για αποθήκευση δεδομένων AIS

Ο κώδικας παράγει ονόματα αρχείων για κάθε ημέρα της περιόδου 2015-2024, χωρίς να προϋποθέτει ότι όντως υπάρχει αρχείο για κάθε ημερομηνία. Αν ένα αρχείο απουσιάζει, θα παρακαμφθεί αργότερα χωρίς να επηρεάσει τη ροή του κώδικα. Παράλληλα, οι έξοδοι οργανώνονται σωστά σε υποφακέλους ανά έτος και ανά μήνα. Η λήψη των αρχείων δεδομένων γίνεται συνεχόμενα (streaming) σε τμήματα (chunks) σταθερού μεγέθους, με timeout για να μη σταματάει η διαδικασία σε ασταθή δίκτυα. Αν το αρχείο υπάρχει ήδη αποθηκευμένο, η λήψη παραλείπεται, ενώ διαγράφονται πολύ μικρά αρχεία καθώς θεωρούνται κατεστραμμένα. Ομοίως, υπάρχει έλεγχος και απόρριψη κατά την αποσυμπίεση του αρχείου, σε περίπτωση που το συμπιεσμένο αρχείο δεν περιέχει αρχείο CSV. Μόλις ολοκληρωθεί η εξαγωγή, το ZIP διαγράφεται ώστε να ελευθερωθεί χώρος.

2) Επιλογή πεδίων και τύποι δεδομένων

Από το αρχείο CSV επιλέγονται μόνο οι απαραίτητες στήλες, οι οποίες είναι το MMSI (αναγνωριστικό πλοίου), η χρονοσήμανση, τα LAT, LON (γεωγραφικές συντεταγμένες), SOG (ταχύτητα πλοίου), και ο τύπος του πλοίου. Η χρήση της βιβλιοθήκης Dask επιτρέπει την εργασία σε lazy mode (δηλαδή η ανάλυση και ο συνδυασμός σχεδιάζονται πριν γίνει

η “βαριά” επεξεργασία), ενώ ορίζονται ρητά τύποι δεδομένων για να επιτευχθεί μέγιστη σταθερότητα (για παράδειγμα χρησιμοποιείται τύπος δεδομένων string για το MMSI και τον τύπο του πλοίου, ενώ χρησιμοποιείται τύπος δεδομένων float για γεωγραφικές συντεταγμένες και ταχύτητα). Σε αυτό το στάδιο επίσης απορρίπτονται γραμμές με κενές τιμές σε βασικά αριθμητικά πεδία (LAT/LON/SOG), γιατί είναι αδύνατο να αξιοποιηθούν σε χωρικές πράξεις ή σε παλινδρόμηση.

3) Κανονικοποίηση χρόνου σε 15 λεπτά και ομαδοποίηση βάσει MMSI και χρόνου

Ουσιαστικά, όλα τα σημεία τακτοποιούνται σε ομοιόμορφα bins των 15 λεπτών (για παράδειγμα 10:00, 10:15, 10:30, 10:45), κάτι το οποίο γίνεται ώστε να απλοποιεί την ευθυγράμμιση με τα μετεωρολογικά δεδομένα και ταυτόχρονα για να καταπολεμά τον “θόρυβο” μικρών αποκλίσεων στη λήψη των χρονικών στιγμών της καταγραφής των δεδομένων AIS. Μετά την κανονικοποίηση σε 15λεπτα bins, ο κώδικας κάνει ομαδοποίηση ανά MMSI σε συνάρτηση με το χρόνο και κρατά την τελευταία εγγραφή ανά ομάδα για τις γεωγραφικές συντεταγμένες, την ταχύτητα και τον τύπο του πλοίου. Αυτό λύνει πρακτικά το ζήτημα πολλαπλών καταγραφών ενός σκάφους μέσα στο ίδιο 15λεπτο. Έτσι, για κάθε 15λεπτο bin προκύπτει μία μοναδική γραμμή ανά πλοίο, κάτι που είναι βασικό προαπαιτούμενο για τη μοντελοποίηση (σε row-wise πρόβλεψη) και για να γίνει χωρίς προβλήματα η συγχώνευση με τον καιρό (1 προς 1 ανά χρονική στιγμή).

4) Αποθήκευση σε αρχεία .parquet ανά ημέρα

Τα αρχεία που δημιουργούνται είναι .parquet, με σκοπό την δραματική επιτάχυνση στα επόμενα στάδια (φόρτωση/φιλτράρισμα/συγχώνευση), την διατήρηση των τύπων των δεδομένων, και ταυτόχρονα τα αρχεία αυτά είναι φιλικά τόσο σε Pandas/Dask όσο και σε άλλα εργαλεία. Η χρήση της Dask και του partitioning βοηθούν τη διαχείριση μεγάλων ημερήσιων αρχείων χωρίς ανάγκη υπερβολικής υπερφόρτωσης της μνήμης RAM.

Συνολικά, για κάθε έγκυρη ημέρα της περιόδου, προέκυψε ένα αρχείο τύπου Parquet, που έχει για στήλες τα MMSI, χρόνο, LAT, LON, SOG, τύπο πλοίου, όπου ο χρόνος είναι κανονικοποιημένος σε 15 λεπτά και χωρίς διπλοεγγραφές εντός bin. Το dataset που προκύπτει είναι σε θέση να προσαρμοστεί κατάλληλα με την προσθήκη των τιμών καιρού (στο ίδιο 15λεπτο και με την εύρεση του πλησιέστερου σημείου αν δεν υπάρχουν

μετεωρολογικά στοιχεία σε εκείνο το σημείο). Αφού στην συνέχεια εφαρμοστούν τα φίλτρα των δεδομένων και το feature engineering, το μοντέλο πρόβλεψης της ταχύτητας θα εκπαιδευτεί για την κατασκευή του θαλάσσιου γράφου με τις προβλεπόμενες ταχύτητες. Η επιτυχία όλης της αλυσίδας εξαρτάται από την καθαρότητα, την ομοιομορφία και την ανθεκτικότητα αυτής της πρώτης φάσης εξασφαλίζοντας την αξιόπιστη λήψη, την ελεγχόμενη αποσυμπίεση, την προσεκτική επιλογή των πεδίων, την κανονικοποίηση του χρόνου και την αποδοτική αποθήκευση σε Parquet.

2.2 Συλλογή Δεδομένων Καιρού

Ο στόχος της αποθήκευσης των μετεωρολογικών στοιχείων είναι να εφαρμοστούν πάνω στα στιγμιότυπα AIS, για την ίδια γεωγραφική περιοχή (LAT 35-42, LON -77 έως -70) και για όλη την περίοδο 2015-2024. Για το σχετικό μετεωρολογικό αρχείο από το NOAA (National Oceanic and Atmospheric Administration της κυβέρνησης των ΗΠΑ), οι στήλες που επιλέγονται για αποθήκευση είναι το γεωγραφικό μήκος και το γεωγραφικό πλάτος (LON, LAT), ο χρόνος καταγραφής (ώρα UTC) και δύο φυσικά μεγέθη (ταχύτητα ανέμου και το ύψος κύματος). Η επιλογή περιορίστηκε σκόπιμα σε αυτά τα δύο, επειδή αφενός αποτυπώνουν με άμεσο τρόπο την “ένταση” του περιβάλλοντος που επηρεάζει την ταχύτητα πλεύσης και αφετέρου έχουν καθιερωμένους ορισμούς και μονάδες στη ναυτική μετεωρολογία.

Εισάγοντας τις κατάλληλες στήλες στο αρχείο CSV που αποθηκεύεται, αυτό θα διαβαστεί σε chunks, ώστε να φιλτραριστούν ανά ημέρα και ανά περιοχή οι εγγραφές που μας αφορούν κάθε φορά, και έπειτα να γίνει η χρονική ευθυγράμμιση και η χωρική αντιστοίχιση. Το αρχείο μπορεί να περιέχει κενές (NaN) τιμές είτε στην ένταση του ανέμου είτε στο ύψος του κύματος. Αυτό είναι λογικό και αντικατοπτρίζει την πραγματικότητα των μετρήσεων (διαφορετικοί αισθητήρες/συνθήκες). Με αυτόν τον σχεδιασμό, σε αυτό το βήμα δεν γίνονται καθαρισμοί/απορρίψεις/μετασχηματισμοί, αλλά δημιουργείται το ακαθάριστο σύνολο δεδομένων. Συνειδητά δεν έγινε καθαρισμός ώστε να υπάρχει αποθηκευμένο το αρχικό dataset και στη συνέχεια να μπορεί να καθαριστεί, χωρίς να απαιτείται κατέβασμα του αρχείου ξανά αν υπάρχει οποιοδήποτε πρόβλημα κατά τον καθαρισμό των δεδομένων για να τρέξει σωστά η μοντελοποίηση.

Ταυτόχρονα, ελαχιστοποιείται η εξάρτηση από εξωτερικούς servers στο υπόλοιπο pipeline. Έτσι, όλα τα μεταγενέστερα βήματα (merge, preprocess, training, routing) τρέχουν offline πάνω στο τοπικό αρχείο CSV.

Συνοψίζοντας, σε αυτό το βήμα εξασφαλίζεται ότι διατίθεται, για ολόκληρο το χρονικό και χωρικό πεδίο της μελέτης, ένα σταθερό σύνολο μετεωρολογικών μετρήσεων (ένταση ανέμου και ύψος κύματος). Με αυτόν τον τρόπο, το επόμενο στάδιο έχει ό,τι χρειάζεται για να συνδέσει με ακρίβεια τον καιρό πάνω στα AIS στοιχεία της ίδιας ημέρας και να παράγει τα ημερήσια συνδυαστικά αρχεία που τροφοδοτούν τη μοντελοποίηση της ταχύτητας και, τελικά, τον δρομολογητή.

2.3 Συγχώνευση Δεδομένων

Το επόμενο βήμα είναι το σημείο σύνδεσης ανάμεσα στα καθαρισμένα δεδομένα AIS και στο ενιαίο CSV καιρού. Ο σκοπός του είναι να παράγει, για κάθε ημέρα, ένα συγχωνευμένο σύνολο όπου κάθε γραμμή AIS εμπλουτίζεται με τις πλησιέστερες στον χώρο τιμές καιρού της ίδιας ακριβώς χρονικής στιγμής (μετά από κοινό χρονικό βήμα 15 λεπτών). Το αποτέλεσμα είναι ημερήσια αρχεία τύπου Parquet, τα οποία τροφοδοτούν άμεσα την προεπεξεργασία και αργότερα το μοντέλο πρόβλεψης ταχύτητας και τον δρομολογητή.

Δεδομένου ότι το αρχείο CSV του καιρού είναι μεγάλο, διαβάζεται τμηματικά, όπου σε κάθε τμήμα γίνονται μετονομασίες στηλών σε κοινή ορολογία με τα δεδομένα AIS (longitude σε LON, latitude σε LAT, time σε BaseDateTime) και γίνεται parsing για τον χρόνο από ISO8601 σε UTC. Οι παρτίδες (chunks) ενώνονται και δημιουργούν ένα ημερήσιο DataFrame καιρού με τις εγγραφές της κάθε ημέρας, μόνο με τις στήλες του χρόνου, της γεωγραφικής θέσης, και τα στοιχεία ανέμου και κύματος. Ομοίως, δημιουργείται ένα ημερήσιο Dataframe με δεδομένα AIS βάσει του αντίστοιχου ημερήσιου αρχείου Parquet που δημιουργήθηκε στο πρώτο βήμα. Στη συνέχεια, γίνεται κοινή χρονική κανονικοποίηση (15 λεπτά) και στα δύο σύνολα (AIS και καιρός), με χαρακτηριστικό γνώρισμα τη στήλη του χρόνου, η οποία μετατρέπεται σε κοινά bin 15 λεπτών, εντός των οποίων θα γίνει η χωρική αντιστοίχιση.

Έπειτα, για κάθε χρονική τιμή που εμφανίζεται στα αρχεία δεδομένων AIS, φιλτράρεται το ημερήσιο DataFrame του καιρού, ώστε να επιλεγούν μόνο τις γραμμές με ίδιο timestamp (μετά το floor). Αν δεν υπάρχουν μετεωρολογικά σημεία στο συγκεκριμένο στιγμιότυπο (timestamp), τότε όλες οι γραμμές AIS που πέφτουν εκεί έχουν την τιμή NaN στη στήλη του ανέμου και του κύματος. Αν υπάρχουν μετεωρολογικά σημεία, χτίζεται ένα KDTree πάνω στις γεωγραφικές συντεταγμένες (LAT, LON) των μετεωρολογικών στοιχείων. Για να επιτευχθεί αυτό, για κάθε γραμμή AIS του συγκεκριμένου timestamp, υπολογίζεται ο πλησιέστερος χωρικά μετεωρολογικός κόμβος και αντιγράφονται οι τιμές έντασης ανέμου και ύψους κύματος. Η χρήση του KDTree ανά timestamp έχει δύο κρίσιμα οφέλη. Το βασικότερο είναι ότι ο χρονικός συγχρονισμός είναι απόλυτος, χωρίς να επιλέγεται ο καιρός για άλλη χρονική περίοδο και η χωρική εύρεση εκτελείται ανά ερώτημα, άρα πολύ γρήγορη ακόμη και για δεκάδες χιλιάδες σημεία της εκάστοτε ημέρας. Στη συνέχεια, τα πεδία καιρού ενώνονται με τα δεδομένα AIS σε ένα ενιαίο DataFrame με στήλες.

Ο λόγος που το KDTree δημιουργείται ανά timestamp και όχι για όλη την ημέρα είναι ότι ο άνεμος και το κύμα μεταβάλλονται στον χρόνο. Αν δημιουργούταν ένα KDTree για όλα τα σημεία της ημέρας, θα υπήρχε το ρίσκο του ταιριάσματος λάθος χρονική μέτρηση σε ένα σημείο με στοιχεία AIS. Έτσι, εξασφαλίζεται ότι η σύζευξη είναι χρονικά ακριβής και ότι κάθε AIS στιγμιότυπο παίρνει τιμές μόνο από το ίδιο 15λεπτο. Η αρχιτεκτονική αυτή για την εύρεση του καιρού σε γειτονικά σημεία επιλέχθηκε γιατί ταιριάζει στο υπόλοιπο pipeline της εργασίας. Στο επόμενο βήμα, στην προεπεξεργασία θα εφαρμοστούν φίλτρα (φυσικά όρια και IQR) και feature engineering πριν χτιστεί το μοντέλο πρόβλεψης της ταχύτητας. Αν στη συγχώνευση των δεδομένων AIS και καιρού επιχειρούταν παρεμβολές ή σύνθετες επιλογές (π.χ. nearest στον χρόνο ± 15), θα χρησιμοποιούταν υπόθεση πριν την προεπεξεργασία. Η παρούσα σχεδίαση κρατά αυτό το βήμα ουδέτερο, αφού κάνει μόνο ακριβή χρονική σύζευξη (ίδιο 15λεπτο) και πλησιέστερο χωρικό ταιρίασμα. Έτσι, όταν στο παρακάτω γίνει λόγος για τα χαρακτηριστικά εισόδου του μοντέλου πρόβλεψης της ταχύτητας (LAT, LON, τύπος πλοίου, ένταση ανέμου, ύψος κύματος, γινόμενο μεταβλητών καιρού, μεταβλητή στασιμότητας - αν είναι σχεδόν ακίνητο το πλοίο), θα είναι σαφές ότι οι τιμές των μετεωρολογικών δεδομένων που χρησιμοποιούνται, αντιστοιχούν στο πραγματικό περιβάλλον όπου και όταν κινήθηκε το πλοίο, χωρίς

αναμίξεις με άλλες χρονικές στιγμές. Αυτό κάνει τα συγχωνευμένα ημερήσια αρχεία χρήσιμα, αφού στηρίζουν με συνέπεια τόσο την ποσοτική εκπαίδευση και αξιολόγηση του μοντέλου όσο και τις εκτιμήσεις ταχύτητας κατά μήκος πραγματικών θαλάσσιων διαδρομών.

2.4 Προεπεξεργασία Δεδομένων

Το επόμενο βήμα είναι το τελικό στάδιο πριν την είσοδο των δεδομένων στο μοντέλο πρόβλεψης της ταχύτητας, αφού ελέγχει την ποιότητα των δεδομένων. Ουσιαστικά, παίρνει τα ημερήσια συγχωνευμένα αρχεία δεδομένων (AIS με δεδομένα καιρού) και τα μετασχηματίζει σε καθαρά, συνεπή και ελαφριά σύνολα για μοντελοποίηση και δρομολόγηση. Η λογική που ακολουθεί είναι η διατήρηση μόνο έγκυρων γραμμών, η απομάκρυνση ακραίων τιμών με ενιαία κριτήρια, η τυποποίηση των πολλών τύπων πλοίων, η δημιουργία ετήσιων αρχείων για ευκολότερη αναζήτηση όποτε απαιτείται και η πρόσθεση επιπλέον χρήσιμων χαρακτηριστικών, με στόχο την μέγιστη δυνατή ακρίβεια πρόβλεψης του μοντέλου στο επόμενο βήμα.

1) Αρχικά, εφαρμόζονται ρεαλιστικοί έλεγχοι για να αποκλειστούν τιμές που δεν έχουν λογική. Έτσι, ορίστηκε ότι η μέγιστη ταχύτητα των πλοίων είναι 35 κόμβοι, η μέγιστη ταχύτητα του ανέμου είναι 50m/s και το μέγιστο ύψος των κυμάτων είναι τα 20m. Οι τιμές εκτός ορίων απομακρύνονται. Τα όρια είναι σταθερά, ώστε να μην περάσουν προφανή λάθη αισθητήρων. Έπειτα, οι κωδικοί των τύπων των πλοίων ταξινομήθηκαν όταν επρόκειτο για παρόμοια είδη πλοίων, ώστε κάθε κατηγορία να έχει επαρκή δείγματα.

2) Όσον αφορά το feature engineering, δημιουργήθηκαν οι επιπλέον μεταβλητές του γινομένου των δύο μεταβλητών καιρού και η μεταβλητή της πολύ χαμηλής ταχύτητας του πλοίου. Το γινόμενο της έντασης του ανέμου και του ύψους του κύματος είναι ένας συνδυαστικός δείκτης έντασης του περιβάλλοντος που συλλαμβάνει μη γραμμικές επιδράσεις ανέμου και κύματος στην ταχύτητα. Ο δυαδικός δείκτης που σηματοδοτεί σχεδόν στάση ($SOG < 4$) προσφέρει στο μοντέλο διακριτή ένδειξη για συνθήκες χαμηλής κινητικότητας (αναμονή, χειρισμοί, επιβράδυνση). Στο μοντέλο είναι κρίσιμη η χρήση της

χαμηλής κινητικότητας στην χάραξη του δρομολογίου, για αυτό και δεν γίνεται απλή παράλειψη των σχεδόν σταματημένων πλοίων.

Τα τελικά training sets προκύπτουν από την τελική ενοποίηση των ημερήσιων αρχείων, αρχικά σε μηνιαία και στη συνέχεια σε ετήσια αρχεία. Η σταδιακή ενοποίηση έγινε για την μικρότερη επιβάρυνση της μνήμης RAM λόγω του μεγάλου όγκου των δεδομένων. Τα ενοποιημένα αρχεία είναι τα κύρια δεδομένα εισόδου (inputs) για την εκπαίδευση και την αξιολόγηση του μοντέλου πρόβλεψης της ταχύτητας, ώστε να είναι αξιόπιστη και γρήγορη η διαδικασία της μηχανικής μάθησης.

3 Μεθοδολογία & Υλοποίηση

3.1 Μοντέλο Πρόβλεψης της Ταχύτητας

Το επόμενο βήμα αποτελεί τη σύνδεση όλης της προετοιμασίας των προηγούμενων βημάτων σε ένα πρακτικό μοντέλο παλινδρόμησης που προβλέπει την ταχύτητα της πλεύσης (SOG) σε κόμβους βάσει της γεωγραφικής θέσης, του τύπου του πλοίου και των τοπικών μετεωρολογικών συνθηκών. Ο τρόπος λειτουργίας του μοντέλου βασίζεται στη λογική ότι κάθε γραμμή δεδομένων αντιστοιχεί σε ένα στιγμιότυπο πλοίου (σε συγκεκριμένο χρόνο και θέση), και το μοντέλο μαθαίνει τη χαρτογράφηση από τα χαρακτηριστικά αυτού του στιγμιότυπου προς την πραγματική ταχύτητα. Η υλοποίηση βασίζεται στον αλγόριθμο Random Forest Regressor, που είναι αξιόπιστος σε ετερογενή και μη γραμμικά δεδομένα και απαιτεί ελάχιστη παραμετροποίηση.

Όσον αφορά την εκπαίδευση του μοντέλου, εισάγονται τα παρακάτω χαρακτηριστικά εισόδου και επιστρέφεται μία συνεχής έξοδος. Τα χαρακτηριστικά που έχουν εισαχθεί έχουν αναφερθεί και προηγουμένως, όμως σε αυτό το κεφάλαιο θα γίνει λόγος για την επίδραση αυτών των χαρακτηριστικών στην πρόβλεψη του μοντέλου:

- Γεωγραφικό πλάτος/μήκος: Επιτρέπουν στο μοντέλο να μάθει γεωγραφικές διαφοροποιήσεις ανάλογα με το αν το πλοίο κινείται πιο ανοικτά ή κοντά σε ακτές και διαύλους.
- Τύπος πλοίου: Κατηγοριοποίηση η οποία προσφέρει πληροφορία μεγέθους, χρήσης και συμπεριφοράς του πλοίου. Η πληροφορία αυτή για το μοντέλο είναι κρίσιμη για την αναμενόμενη ταχύτητα υπό παρόμοιο καιρό.
- Ταχύτητα ανέμου (m/s): Συσχετίζεται συνήθως με τη μείωση της επιτεύξιμης ταχύτητας για δεδομένη ισχύ.
- Ύψος κύματος (m): Περιλαμβάνει την κατάσταση θαλάσσης και την πρόσθετη αντίσταση λόγω κυματισμού.

- Το γινόμενο της ταχύτητας του ανέμου \times το ύψος του κύματος: Ένα απλό και μη γραμμικό χαρακτηριστικό που συνοψίζει την ισχύ του περιβάλλοντος (άνεμος και κύμα μαζί), είναι απλός δείκτης έντασης κυματικού και αεροδυναμικού φορτίου και σχετίζεται με την πρόσθετη αντίσταση πλεύσης και δεν είναι φυσικό μοντέλο.
- Δυαδικός δείκτης χαμηλής κινητικότητας: Σηματοδοτεί διαφορετικό καθεστώς συμπεριφοράς σε ειδικές περιπτώσεις όπως είναι η στάση, οι χειρισμοί και η πολύ χαμηλή ταχύτητα.

Αφού διαγράφονται οι μη πλήρεις γραμμές και γίνεται ενοποίηση σε ένα κοινό training corpus, το επόμενο βήμα είναι ο διαχωρισμός του dataset σε εκπαίδευση και έλεγχο (train / test) σε αναλογία 80/20. Ουσιαστικά, το 80% των γραμμών χρησιμοποιείται για εκπαίδευση, το 20% κρατείται αυστηρά για αξιολόγηση. Ο διαχωρισμός γίνεται τυχαία ως προς τις γραμμές (per-row), στοιχείο που ταιριάζει με τη λογική των αρχείων, όπου κάθε γραμμή είναι ένα στιγμιότυπο. Δεν γίνεται διαχωρισμός βάσει του χρόνου (time-series split), γιατί το μοντέλο δεν αξιοποιεί τη σειρά και τη μνήμη.

Στην πλήρη εκτέλεση, το τελικό ενιαίο μοντέλο (όλων των ετών) απέδωσε $MAE \approx 1.14$ kn και $R^2 \approx 0.85$ στο test set. Η επίδοση αυτή επιβεβαιώνει ότι το σύνολο των χαρακτηριστικών αποτυπώνει ικανοποιητικά τη σχέση περιβάλλοντος - τύπου πλοίου και γεωγραφίας με την πραγματική ταχύτητα πλεύσης.

Το μοντέλο που δημιουργήθηκε με τη χρήση του αλγορίθμου Random Forest Regressor περιλαμβάνει τις παρακάτω βασικές παραμέτρους:

- `n_estimators=100`: Πλήθος δέντρων στο “δάσος”. Περισσότερα δέντρα σημαίνουν συνήθως μικρότερη διακύμανση (σταθερότερη πρόβλεψη) και σχετική βελτίωση στα μετρικά (metrics), αλλά περισσότερο χρόνο εκπαίδευσης και μνήμη. Η τιμή 100 είναι μία λογική τιμή σε τέτοιου είδους μοντέλα.
- `max_depth=20`: Μέγιστο βάθος κάθε δέντρου, δηλαδή πόσο πολύ επιτρέπεται να διακλαδιστεί. Με πολύ μικρό βάθος, το δάσος υποπροσαρμόζει (υψηλό bias), ενώ

με υπερβολικά μεγάλο βάθος, τα δέντρα εκπαιδεύονται με θόρυβο (υψηλή διακύμανση). Η τιμή 20 λειτουργεί ως ισορροπία, παρέχοντας αρκετή εκφραστικότητα και ελεγχόμενη πολυπλοκότητα.

- `max_features="sqrt"`: Σε κάθε `split`, εξετάζεται μόνο ένα τυχαίο υποσύνολο γνωρισμάτων μεγέθους $\sim\sqrt{n_features}$. Αυτό μειώνει τη συσχέτιση στα δέντρα και μειώνει τη διακύμανση του δείγματος. Παρ' όλα αυτά, αν χρησιμοποιούταν αντί για "sqrt" το "auto" (δηλαδή να χρησιμοποιούνται όλα τα χαρακτηριστικά), τα δέντρα γίνονται πιο όμοια μεταξύ τους και η μέση πρόβλεψη δεν ακυρώνει εξίσου καλά τον θόρυβο, ενώ το MAE συνήθως δεν βελτιώνεται. Με τη χρήση του "sqrt" επιτυγχάνεται καλύτερη γενίκευση.
- `n_jobs=-1`: Αυτή η παράμετρος αφορά τη διαχείριση του υπολογιστή, αφού με τη συγκεκριμένη τιμή ο υπολογιστής αξιοποιεί όλους τους πυρήνες CPU για να εκπαιδεύσει και να αξιολογήσει παράλληλα. Αυτή παράμετρος είναι πολύ σημαντική όταν ο αριθμός δέντρων είναι ≥ 100 και τα δεδομένα πολυάριθμα.
- `random_state=42`: Σταθερή τυχαιότητα ώστε το αποτέλεσμα να είναι μπορεί να αναπαραχθεί (ίδιο δάσος, ίδια κατανομή splits, ίδιο αποτέλεσμα όταν τα δεδομένα είναι ίδια).
- Άλλα `hyperparams` αφήνονται στις προεπιλογές (όπως `criterion="squared_error"`, `min_samples_split=2`, `min_samples_leaf=1`). Ο συγκεκριμένος αλγόριθμος γενικά δεν χρειάζεται βαριά ρύθμιση για να επιτύχει καλές επιδόσεις. Γενικά, ο αλγόριθμος Random Forest έχει την έμφυτη ικανότητα να μαθαίνει χρήσιμες μη γραμμικότητες με περιορισμένο ρίσκο υπερπροσαρμογής (`overfitting`).

Στη διαδικασία της εκπαίδευσης, ο αλγόριθμος αποθηκεύει τα δεδομένα και τα ενώνει σε ένα `DataFrame`. Στη συνέχεια, διαχωρίζει τα δεδομένα σε τμήματα και ξεκινάει η εκπαίδευση του Random Forest δημιουργώντας τα 100 δέντρα. Για κάθε δέντρο, παίρνει ένα δείγμα από τις γραμμές της εκπαίδευσης. Στη συνέχεια, αναπτύσσει το δέντρο με διαχωρισμούς που ελαχιστοποιούν τον δείκτη MSE εντός των κόμβων, μέχρι να

ικανοποιηθεί το `max_depth`. Σε κάθε `split` χρησιμοποιεί μόνο ένα υποσύνολο των γνωρισμάτων λόγω του `max_features="sqrt"`.

Στη διαδικασία της αξιολόγησης υπολογίζονται οι εξής δείκτες:

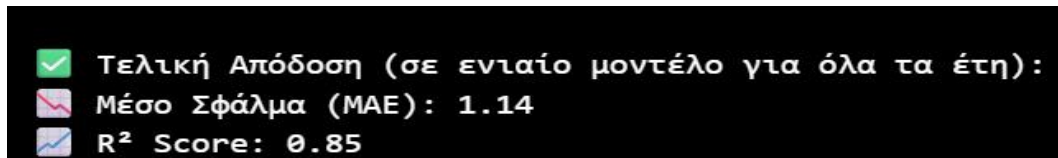
- MAE (Mean Absolute Error) (σε κόμβους): μετρική για την εύρεση του μέσου σφάλματος.
- R² (Coefficient of Determination): Ποσοστό διακύμανσης της ταχύτητας, δηλαδή το ποσοστό ακρίβειας του μοντέλου.

Επίσης παράγονται μετρικές ανά τύπο πλοίου (MAE, R², αριθμός καταγραφών) και αποθηκεύονται σε αρχείο CSV (Εικόνα 1) και το εκπαιδευμένο αντικείμενο αποθηκεύεται, ώστε να μπορεί να φορτωθεί άμεσα στη σχεδίαση του δρομολογίου. Αυτή η απεικόνιση ανά τύπο πλοίου είναι πολύ χρήσιμη ώστε να διαπιστώσει ο χρήστης του μοντέλου σε ποιο είδος πλοίων το μοντέλο αποδίδει καλύτερα και να δει ποιες κλάσεις έχουν λιγότερα δείγματα (όπου οι μετρικές είναι ασταθείς). Γενικά, το γεγονός ότι έγινε ομαδοποίηση των πλοίων παρόλο που αυξάνει τον αριθμό των πλοίων ανά κατηγορία, είναι αναμενόμενο να μην έχουν όλα τα πλοία της κάθε κατηγορίας ακριβώς ίδια στοιχεία, γεγονός που επηρεάζει την αποτελεσματικότητα του μοντέλου.

☑ Αξιολόγηση ανά VesselType:				
	VesselType	MAE	R2	Count
0	0	1.532844	0.777090	397.0
1	10	1.402697	0.808929	353.0
2	20	1.187848	0.678822	105.0
3	30	1.025412	0.787289	982540.0
4	40	1.948322	0.726735	593.0
5	50	1.259841	0.776175	44324.0
6	60	1.294795	0.818308	74188.0
7	70	1.153413	0.907792	286643.0
8	80	1.014238	0.924403	62444.0
9	90	1.325571	0.770313	77206.0
10	107	0.280200	-0.107772	4.0
11	200	1.897724	0.843550	5.0
12	207	1.793939	-8.611985	7.0
13	255	2.886980	0.547725	9.0
14	1001	1.082096	0.748582	98815.0
15	1003	1.644125	0.835678	109.0
16	1004	1.962493	0.804684	72132.0
17	1005	1.182834	0.774024	13055.0

Εικόνα 1: Αξιολόγηση ακρίβειας μοντέλου ανά τύπο πλοίου

Συνολικά τα αποτελέσματα $MAE = 1.14$ knots και $R^2 = 0.85$ (Εικόνα 2) αποτελούν ένα λογικό επίπεδο σφάλματος για εφαρμογές εκτίμησης χρόνου ταξιδιού με διαδρομές της τάξης εκατοντάδων ναυτικών μιλίων.



Εικόνα 2: Τελική απεικόνιση μέσου σφάλματος και συνολικού R^2 του μοντέλου

Για μικρές κλάσεις (π.χ. αριθμός καταγραφών < 200), τα R^2 μπορεί να εμφανίζονται ασταθή-σε τέτοιες περιπτώσεις το MAE και το ίδιο το Count (αριθμός καταγραφών) βοηθούν στην ερμηνεία. Όσον αφορά την αποθήκευση του μοντέλου, το αρχείο .pkl είναι πολύ χρήσιμο γιατί δεν χρειάζεται επανεκπαίδευση, και κατά την εκτέλεση της δρομολόγησης, φορτώνεται τον εκπαιδευμένο μοντέλο και κάνει μαζικές προβλέψεις ταχύτητας για τα τμήματα της διαδρομής.

Κάθε γραμμή στο training set είναι ένα στιγμιότυπο, εκπαιδύοντας ουσιαστικά το μοντέλο να αναγνωρίζει ότι σε μία συγκεκριμένη θέση (LAT/LON), για συγκεκριμένο τύπο πλοίου, και με συγκεκριμένο άνεμο/κύμα, ποια εκτιμάται ότι είναι η ταχύτητα. Ο λόγος που επιλέχθηκε ο αλγόριθμος Random Forest είναι ότι δεν διατηρείται ιστορική μνήμη και δεν κοιτάει τροχιές, παρά μόνο την τωρινή κατάσταση. Αυτό ακριβώς ταιριάζει με την χρήση στην κατασκευή του δρομολογίου, αφού για κάθε σκέλος της διαδρομής, χρησιμοποιείται ο τοπικός καιρός και αναζητείται η ταχύτητα από το μοντέλο.

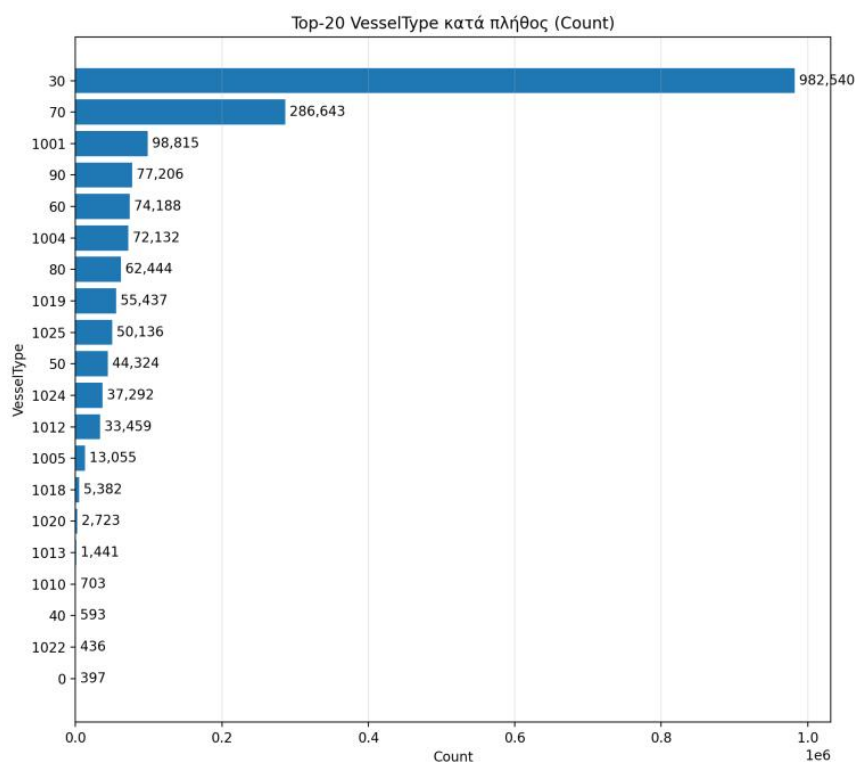
Ολοκληρώνοντας το συγκεκριμένο πολύ σημαντικό υποκεφάλαιο, πρέπει να αναφερθεί ότι το μοντέλο αυτό αποτελεί το σημείο κλειδί της προγνωστικής συνιστώσας. Συνοπτικά, αυτό λαμβάνει τα προεπεξεργασμένα δεδομένα μιας δεκαετίας, τα διαχωρίζει σε train και test, εκπαιδύει τον αλγόριθμο Random Forest με λογικές παραμέτρους και αξιολογεί την ικανότητα του μοντέλου να μεταφράζει τη γεωγραφία, τον καιρό και τον τύπο του πλοίου σε εκτιμώμενη ταχύτητα. Τα αποτελέσματα του μοντέλου είναι ενθαρρυντικά και, κυρίως, χρήσιμα, αφού επιτρέπουν στον δρομολογητή να προβλέψει ρεαλιστικά το χρόνο άφιξης, τη μέση ταχύτητα και την απόσταση, σε διαδρομές που

εκτελούνται στην πραγματική ναυσιπλοΐα. Με την αποθήκευση του μοντέλου και την παραγωγή των μετρικών και των διαγραμμάτων, η μεθοδολογία μπορεί να εκτελεστεί πολλές φορές και να είναι χρήσιμη στο υπόλοιπο του προγράμματος.

3.2 Διαγράμματα Δεδομένων Μοντέλου Πρόβλεψης Ταχύτητας

3.2.1 Πολυπληθέστερες Κατηγορίες Τύπων Πλοίων

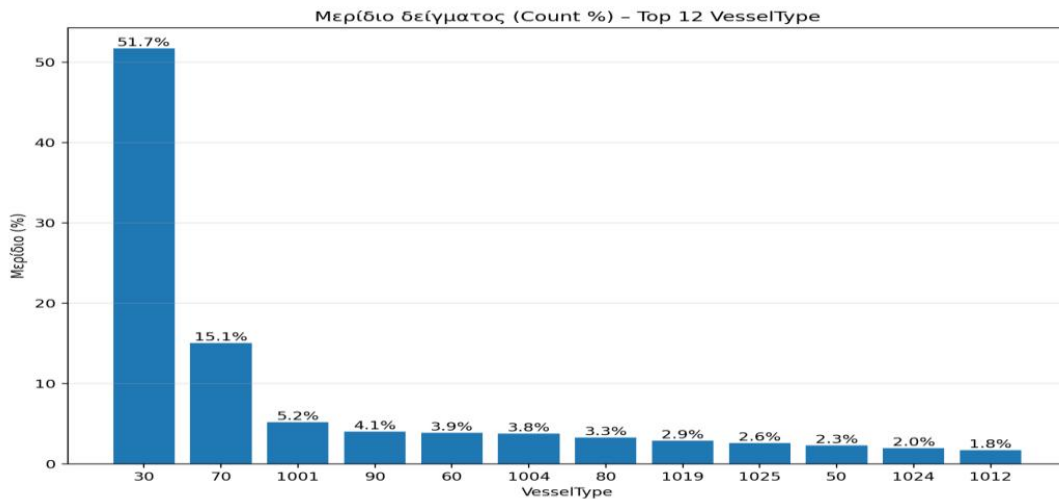
Το διάγραμμα της Εικόνας 3 αποτυπώνει τις 20 πολυπληθέστερες κατηγορίες πλοίων στην αξιολόγηση του μοντέλου, σε φθίνουσα κατάταξη ως προς τον αριθμό των καταγραφών. Η εικόνα είναι εξαιρετικά χρήσιμη για να γίνει κατανοητό σε ποιο τύπο πλοίων συγκεντρώνεται η μάζα των δεδομένων που επηρεάζει τον συνολικό μέσο όρο (MAE, R^2). Κυριαρχεί απόλυτα ο τύπος πλοίου 30 με 982 χιλιάδες γραμμές, δηλαδή πάνω από μισό εκατομμύριο παρατηρήσεις μόνο του. Δεύτερος είναι ο 70 με ~286 χιλιάδες, και ακολουθούν οι 1001 (~98 k), 90 (~77 χιλιάδες), 60 (~74 χιλιάδες), 1004 (~72 χιλιάδες), 80 (~62 χιλιάδες), 1019 (~55 χιλιάδες), 1025 (~50 k), 50 (~44 χιλιάδες), 1024 (~37 χιλιάδες), 1012 (~33 χιλιάδες), ενώ στο τέλος της εικοσάδας των τύπων πλοίων εμφανίζονται πολύ μικρότερες κλάσεις (π.χ. 1010: 703, 40: 593, 1022: 436, 0: 397). Αυτό το διάγραμμα δείχνει ότι παρότι παρουσιάζονται οι πρώτες 20 κατηγορίες, μέσα στο ίδιο γράφημα συνυπάρχουν κλάσεις με τεράστιο αριθμό καταγραφών και άλλες με πολύ μικρότερο. Για την ερμηνεία των επιδόσεων ανά τύπο πλοίου, η στήριξη (πλήθος καταγραφών) δείχνει ότι οι πολυπληθείς κλάσεις οδηγούν την συνολική επίδοση και δίνουν σταθερές, χαμηλού θορύβου μετρικές, ενώ οι υπόλοιπες κλάσεις (μερικές εκατοντάδες εγγραφές) μπορεί να εμφανίζουν μεγάλες μεταβολές (π.χ. μερικά αρνητικά R^2) χωρίς αυτό να σημαίνει ότι το μοντέλο αποτυγχάνει. Ο οριζόντιος προσανατολισμός των μπαρών βοηθά την ανάγνωση των μεγάλων αριθμών, ενώ οι αριθμητικές ετικέτες στο άκρο κάθε μπάρας δείχνουν άμεσα το μέγεθος (με διαχωριστικό χιλιάδων). Εύκολα λοιπόν γίνεται αντιληπτό ότι το 67% του δείγματος προέρχεται από τις κλάσεις 30 και 70, κάτι που προετοιμάζει τον αναγνώστη για τα αποτελέσματα MAE και R^2 που ακολουθούν, εξηγώντας γιατί ο συνολικός $MAE \approx 1.14$ και $R^2 \approx 0.85$ είναι ουσιαστικά «ζυγισμένοι» προς αυτές τις κατηγορίες, άρα και πιο αντιπροσωπευτικοί του πραγματικού φόρτου.



Εικόνα 3: Horizontal Barplot με τις πολυπληθέστερες κατηγορίες πλοίων

3.2.2 Κατανομή Ποσοστού Καταγραφών Πολυπληθέστερων Τύπων Πλοίων

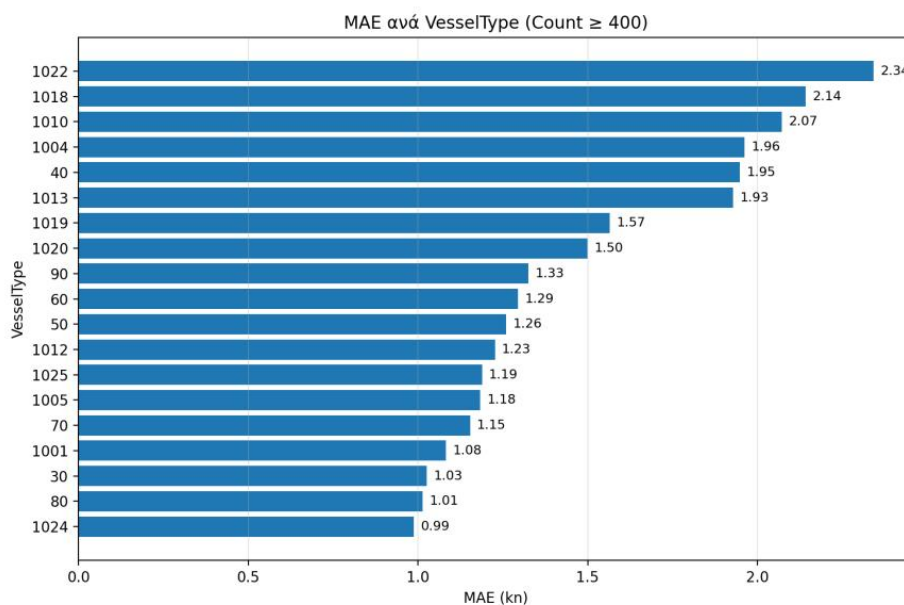
Το δεύτερο διάγραμμα (Εικόνα 4) μεταφράζει τον αριθμό των καταγραφών σε ποσοστό επί του συνόλου, για τις 12 πιο πολυπληθείς κλάσεις. Ο στόχος είναι να αναζητήσει τους τύπους των πλοίων που διαμορφώνουν τον μέσο όρο. Όπως φάνηκε προηγουμένως, τύπος πλοίων με αριθμό 30 καλύπτει το 51.7% των δεδομένων, ενώ ο τύπος πλοίων με αριθμό 70 καταλαμβάνει το 15.1%. Δηλαδή, δύο κλάσεις συνεισφέρουν τα 2/3 των παρατηρήσεων. Ακολουθούν σε πολύ χαμηλότερα ποσοστά οι 1001 (~5%), 90 (~4%), 60 (~4%), 1004 (~4%), 80 (~3%), 1019 (~3%), 1025 (~2.5%), 50 (~2%), 1024 (~2%), 1012 (~2%). Η οπτικοποίηση με ποσοστά πάνω από κάθε μπάρα διευκολύνει τον αναγνώστη να αντιληφθεί το βάρος επιρροής κάθε κλάσης στις συνολικές μετρικές. Έτσι, όταν ο Route Planner υπολογίζει το χρόνο άφιξης σε ένα τυπικό δρομολόγιο, το μεγαλύτερο μέρος των προβλέψεων ανά σκέλος θα πέσει σε κλάσεις που αποτελούν το κύριο ποσοστό της κίνησης και στις οποίες το μοντέλο είναι ιδιαίτερα ακριβές.



Εικόνα 4: Vertical Barplot κατανομής ποσοστού καταγραφών πολυπληθέστερων τύπων πλοίων

3.2.3 Συσχέτιση MAE ανά Τύπο πλοίου

Στην εικόνα 5 απεικονίζεται η μέση απόλυτη απόκλιση (MAE σε κόμβους) ανά κατηγορία πλοίου, αλλά μόνο για κλάσεις με επαρκή κάλυψη (αριθμός καταγραφών ≥ 400). Το φίλτρο είναι απαραίτητο καθώς απομακρύνει κλάσεις με ελάχιστα δείγματα, οι οποίες αποτελούν θόρυβο, ώστε ο αναγνώστης να εστιάσει σε σταθερές εκτιμήσεις του σφάλματος. Το γράφημα είναι ταξινομημένο (αύξουσα τιμή MAE), οπότε στο κάτω μέρος απεικονίζονται οι καλύτερες κλάσεις σε όρους σφάλματος, ως εξής 1024 \approx 0.99 kn, 80 \approx 1.01 kn, 30 \approx 1.03 kn, 1001 \approx 1.08 kn, 70 \approx 1.15 kn, 1005 \approx 1.18 kn, 1025 \approx 1.19 kn, 1012 \approx 1.23 kn, 50 \approx 1.26 kn, 60 \approx 1.29 kn, 90 \approx 1.33 kn. Στη μέση κλίμακα φαίνονται οι 1019 \approx 1.57 kn, 1020 \approx 1.50 kn, ενώ προς τα πάνω εμφανίζονται οι πιο κλάσεις με μεγάλο σφάλμα ως εξής: 1013 \approx 1.93 kn, 40 \approx 1.95 kn, 1004 \approx 1.96 kn, 1010 \approx 2.07 kn, 1018 \approx 2.14 kn, 1022 \approx 2.34 kn. Η εικόνα αυτή συνάδει με όσα αποτελέσματα φάνηκαν κατά την εκπαίδευση. Οι πολύ πολυπληθείς κλάσεις όπως 30, 70, 80, 1001, 1024, 1025 έχουν χαμηλό MAE, γιατί το μοντέλο έχει πλούσιο φάσμα σε καιρό και θέση και μαθαίνει σταθερές σχέσεις. Αντίθετα, κλάσεις με πιο ετερογενή συμπεριφορά ή λιγότερη ποικιλία (ακόμη κι αν περνούν το όριο των 400 εγγραφών) εμφανίζουν υψηλότερο MAE.

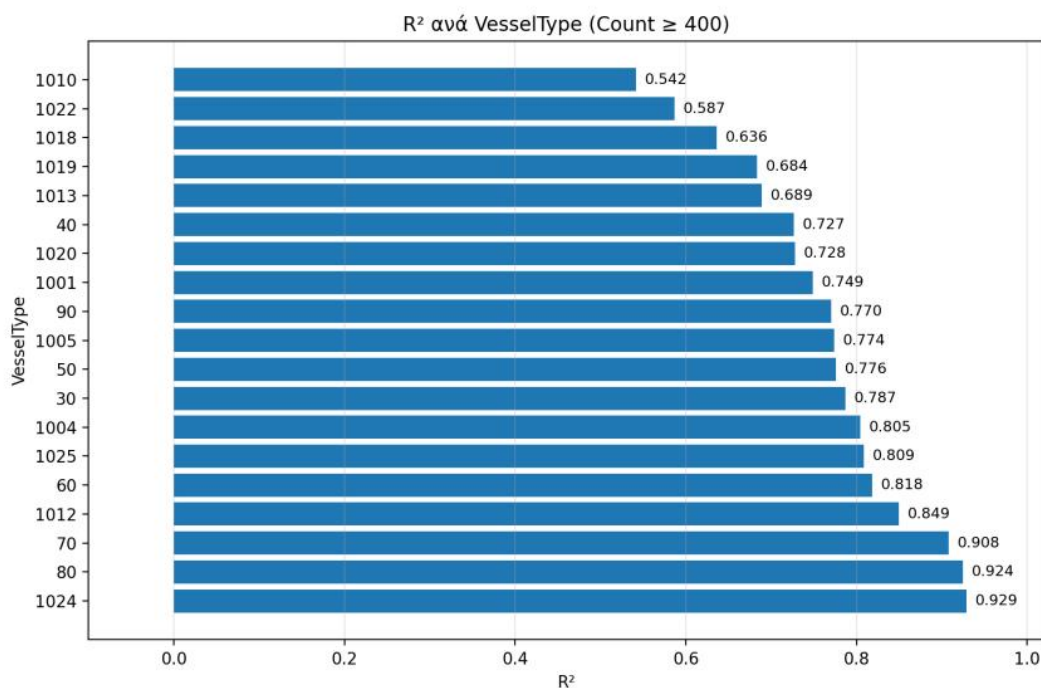


Εικόνα 5: Horizontal Barplot Συσχέτισης MAE ανά Τύπο πλοίου

3.2.4 Συσχέτιση R^2 ανά Τύπο Πλοίου

Το τρίτο γράφημα (Εικόνα 6) παρουσιάζει τον συντελεστή R^2 ανά τύπο πλοίου, και πάλι μόνο για κατηγορίες πλοίων με τουλάχιστον 400 καταγραφές. Η ταξινόμηση γίνεται φθίνουσα, ώστε στο άνω μέρος να απεικονίζονται οι κλάσεις όπου το μοντέλο εξηγεί πολύ μεγάλο μέρος της διακύμανσης. Ξεχωρίζουν οι κατηγορίες 1024 (~ 0.929), 80 (~ 0.924), 70 (~ 0.908), οι οποίες έχουν μεγάλη στήριξη και ομοιόμορφα πρότυπα συμπεριφοράς. Πολύ καλές τιμές εμφανίζουν και οι κατηγορίες 1012 (~ 0.849), 60 (~ 0.818), 1025 (~ 0.809) και 1004 (~ 0.805). Η μεγαλύτερη κατηγορία πλοίων (με τον αριθμό 30, καταλαμβάνοντας περίπου τις μισές παρατηρήσεις) βρίσκεται στο 0.78. Αυτό θεωρείται αναμενόμενο, αφού περιλαμβάνει τεράστια ετερογένεια (γεωγραφία, προφίλ αποστολών πλοίων). Πιο κάτω φαίνονται οι κατηγορίες 50 (~ 0.776), 1005 (~ 0.774), 90 (~ 0.770), 1001 (~ 0.749), 1020 (~ 0.728), 40 (~ 0.727) και, στο τέλος, κλάσεις όπως οι 1013 (~ 0.689), 1019 (~ 0.684), 1018 (~ 0.636), 1022 (~ 0.587), 1010 (~ 0.542). Η χρήσιμη ανάγνωση εδώ είναι διπλή. Αρχικά το ψηλό R^2 σημαίνει ότι τα διαθέσιμα χαρακτηριστικά (θέση, τύπος, καιρικά στοιχεία, γινόμενο μεταβλητών καιρού) επαρκούν για να αναλυθεί η ταχύτητα των πλοίων της αντίστοιχης κατηγορίας. Δευτερευόντως, το χαμηλότερο R^2 δεν σημαίνει κατ' ανάγκην κακή πρόβλεψη, αλλά συχνά συμβαδίζει με μεγαλύτερη ετερογένεια ή με λιγότερο καλές ενδείξεις των καιρικών δεδομένων (π.χ. ελλιπέστερη

διαθεσιμότητα ύψους κύματος). Σε συνδυασμό με το διάγραμμα που απεικονίζει τα αντίστοιχα MAE, ο χρήστης του προγράμματος αποκτά πλήρη εικόνα για τις κλάσεις με $R^2 > 0.80$ τείνουν να έχουν και MAE $\sim 1.0-1.3$, άρα υποδεικνύουν σταθερότητα στο σύστημα. Ο οριζόντιος άξονας οριοθετείται έως ~ 1.0 , αλλά επιτρέπει και ελαφρώς αρνητικές τιμές αν τυχόν εμφανιστούν σε άλλες εκτελέσεις (εδώ δεν απεικονίζεται αυτό λόγω του φίλτρου των 400 καταγραφών). Οι περισσότερες κλάσεις με ουσιαστικό μερίδιο του δείγματος εμφανίζουν $R^2 > 0.75$, επιβεβαιώνοντας την αξιοπιστία και την προγνωστική ισχύ του μοντέλου.



Εικόνα 6: Horizontal Barplot Συσχέτισης R^2 ανά Τύπο Πλοίου

3.2.5 Συνολική Αποτίμηση ανά Τύπο Πλοίου (συνδυασμός MAE & R^2)

Ο πιο ασφαλής τρόπος να εντοπιστούν οι κατηγορίες που είναι όντως αξιόπιστες, συσχετίστηκαν τα δύο κριτήρια (χαμηλό MAE και υψηλό R^2) και σταθμίστηκαν με τον αριθμό των καταγραφών. Υπό αυτό το πρίσμα, ξεχωρίζει μια ομάδα κλάσεων που είναι και ακριβείς και σταθερές:

- I. 80 (MAE \approx 1.01, $R^2 \approx$ 0.924, ~62k),
- II. 1024 (MAE \approx 0.99, $R^2 \approx$ 0.929, ~37k),
- III. 70 (MAE \approx 1.15, $R^2 \approx$ 0.908, ~287k),
- IV. 1012 (MAE \approx 1.23, $R^2 \approx$ 0.849, ~33k),
- V. 60 (MAE \approx 1.29, $R^2 \approx$ 0.818, ~74k),
- VI. 1025 (MAE \approx 1.19, $R^2 \approx$ 0.809, ~50k).

Οι παραπάνω συνδυάζουν χαμηλά σφάλματα με υψηλή ακρίβεια και σημαντικό όγκο παρατηρήσεων. Στη συνέχεια βρίσκεται η υπερπολυπληθής κατηγορία πλοίων με τον αριθμό 30 (MAE \approx 1.03, $R^2 \approx$ 0.787, ~983k), όπου παρότι το R^2 είναι λίγο χαμηλότερο, ο τεράστιος όγκος δείγματος και το πολύ χαμηλό MAE την καθιστούν κομβική για σταθερή πρόβλεψη του χρόνου άφιξης στο σύνολο των διαδρομών.

Μια δεύτερη ζώνη επιτυχημένων (αλλά με μεγαλύτερη διασπορά) κλάσεων περιλαμβάνει τα κάτωθι:

- I. 1001 (MAE \approx 1.08, $R^2 \approx$ 0.749, ~99k),
- II. 90 (MAE \approx 1.33, $R^2 \approx$ 0.770, ~77k),
- III. 50 (MAE \approx 1.26, $R^2 \approx$ 0.776, ~44k),
- IV. 1004 (MAE \approx 1.96, $R^2 \approx$ 0.805, ~72k).

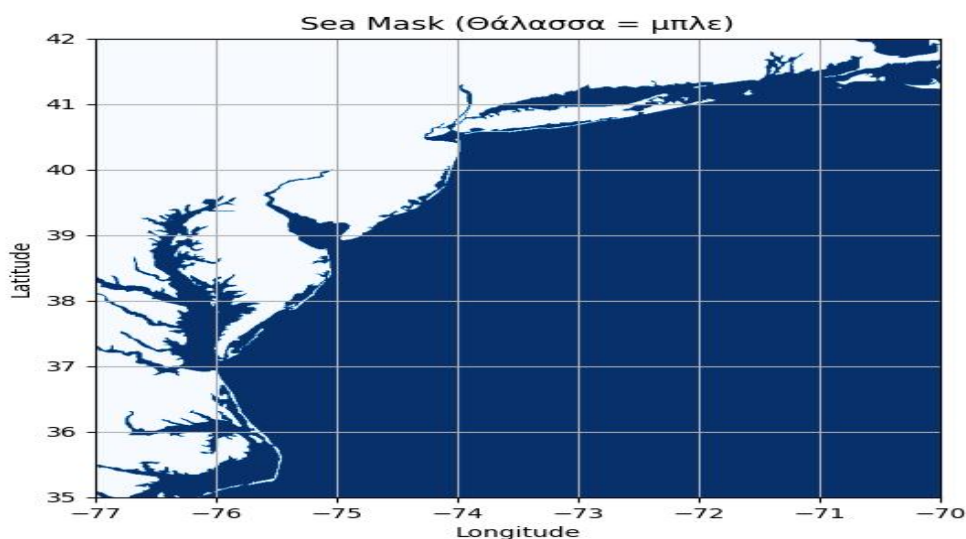
Εδώ, τα ελαφρώς υψηλότερα MAE και πιο μέτρια R^2 υποδηλώνουν πιο ετερογενή προφίλ ταχύτητας. Ωστόσο, το γεγονός ότι το R^2 παραμένει υψηλό (π.χ. 1004) σημαίνει πως η σχέση των χαρακτηριστικών με την ταχύτητα είναι συνεπής, απλά το εύρος πραγματικών ταχυτήτων είναι μεγαλύτερο.

Τέλος, οι υπόλοιπες κατηγορίες δείχνουν είτε υψηλότερα σφάλματα είτε χαμηλότερη ακρίβεια με εκτιμώμενες αιτίες τη μικρότερη ποικιλία/αριθμού καταγραφών σε κρίσιμα μοτίβα καιρού (ιδίως ύψους κύματος), μεγαλύτερη ετερογένεια ή διαφορετική χρήση των σκαφών.

Συνολικά, ο συνδυασμός MAE με R^2 και με το πλήθος των καταγραφών δείχνει ότι ο κορμός του στόλου (κατηγορίες πλοίων με κωδικούς 30, 70, 80, 60, 1001, 1024, 1025, 1012) παρέχει υψηλή ακρίβεια και σταθερότητα, άρα μεταφέρει την αξιοπιστία των χρόνων αφίξεως στις περισσότερες πραγματικές διαδρομές. Οι πιο ετερογενείς κλάσεις επηρεάζουν οριακά το συνολικό αποτέλεσμα και απαιτούν πιο εξειδικευμένη διαχείριση.

3.3 Sea Mask

Το επόμενο βήμα είναι η κατασκευή του υποστρώματος πάνω στο οποίο στηρίζεται ο Route Planner, το οποίο είναι μια μάσκα της θάλασσας (sea mask) σε μορφή πλέγματος, που σηματοδοτεί για κάθε κελί αν είναι θάλασσα ή ξηρά. Η επιλογή αυτή επιβάλλει από την αρχή τον βασικό κανόνα ασφαλείας, δηλαδή οι διαδρομές δεν επιτρέπεται να διασχίζουν την ξηρά. Η μάσκα αποθηκεύεται ως δυαδικός πίνακας και ευθυγραμμίζεται με ένα κανονικό πλέγμα γεωγραφικών συντεταγμένων στην ανάλυση (Εικόνα 7).



Εικόνα 7: Απεικόνιση Sea Mask

Όσον αφορά την ανάλυση του πλέγματος και τα όρια του, η επιλεγθείσα περιοχή είναι η ίδια για την οποία έχουν συλλεχθεί τα δεδομένα AIS και καιρού και θα γίνει και η δρομολόγηση ($LAT \in [35, 42]$, $LON \in [-77, -70]$). Η χωρική ανάλυση (Resolution) που επιλέχθηκε ώστε να μην είναι ανούσια και υπερβολικά αναλυτική, να μην επιβαρύνει πολύ τη μνήμη RAM κατά την δρομολόγηση, αλλά και ταυτόχρονα να διαχωρίζει τα στενά σημεία θαλάσσης σε διαύλους, είναι το 0.01° . Σε αυτή την επιλογή, κάθε κελί

αντιστοιχεί περίπου σε 1.11 χιλιόμετρο στον άξονα γεωγραφικού πλάτους (και ~0.8-0.9 χιλιόμετρα στον άξονα του γεωγραφικού μήκους κοντά στα γεωγραφικά πλάτη που μελετώνται). Η ανάλυση είναι πολύ ικανοποιητική για παράκτιες γραμμές και ταυτόχρονα υπολογιστικά διαχειρίσιμη. Με τη χρήση των δεδομένων του “Natural Earth 1:10m Land”, φορτώθηκαν οι περιοχές της ξηράς, έγινε χωρική περικοπή (“crop”) στις πολυγωνικές γεωμετρίες στην περιοχή ενδιαφέροντος για να μειωθεί δραστικά το κόστος των χωρικών πράξεων. Κάθε πολύγωνο ξηράς μετατρέπεται σε “prepared geometry”, οι οποίες διατηρούν εσωτερικές δομές για γρήγορους ελέγχους χωρικής σχέσης, κάτι που μειώνει κατακόρυφα το χρόνο όταν ελέγχονται δεκάδες και εκατοντάδες χιλιάδες σημεία του πλέγματος. Σημαντική λεπτομέρεια είναι ότι για κάθε σημείο (lat, lon) του πλέγματος δημιουργείται ένα Shapely Point. Στη συνέχεια, ελέγχεται αν οποιαδήποτε “prepared geometry” της ξηράς το καλύπτει (μέθοδος “covers”) για να διαπιστωθεί αν το δρομολόγιο μπορεί να περνάει από εκείνο το σημείο. Αν το καλύπτει, τότε το κελί χαρακτηρίζεται ξηρά (sea_mask = “False”), αλλιώς πρόκειται για σημείο της θάλασσας.

Πολύ σημαντικό είναι να γίνει λόγος για το περίγραμμα της περιοχής, το οποίο προφανώς είναι μέρος της ξηράς. Έγινε χρήση του “covers” (και όχι π.χ. “contains”), γιατί το “covers” επιστρέφει “True” και για σημεία πάνω στο περίγραμμα. Έτσι, σημεία ακριβώς πάνω στην ακτογραμμή λογίζονται ως ξηρά. Αυτό είναι κρίσιμο για τον δρομολογητή, καθώς αποφεύγει την χάραξη ακμής πάνω στην ακτογραμμή, κάτι που σε χάρτη μοιάζει αποδεκτό αλλά στην πράξη είναι επικίνδυνο να διέρχεται από παραλίες, προβλήτες ή πολύ ρηχά νερά.

Ο έλεγχος των σημείων πάνω στα πολύγωνα γίνεται σε δύο loops (για κάθε γεωγραφικό μήκος και για κάθε γεωγραφικό πλάτος). Η χρήση των “prepared geometry” και η περικοπή “crop” της περιοχής κάνουν τη διαδικασία γρήγορη, παρά το μεγάλο πλήθος κυψελών. Επίσης, χρησιμοποιήθηκε αντιστοίχιση από γεωγραφικές συντεταγμένες σε δείκτες πίνακα (“floor”), ώστε κάθε πραγματική θέση (lat, lon) να μπορεί να συνδεθεί με τον πλησιέστερο κόμβο του πλέγματος (κελί της μάσκας).

Η sea mask είναι ο δομικός περιορισμός που θα αξιοποιηθεί στο γράφο της δρομολόγησης, για τη δημιουργία των κόμβων του δικτύου, μόνο όπου το sea mask είναι “True”. Έτσι, δεν μπορεί να υπάρξει κόμβος πάνω σε ξηρά. Αναφορικά με τις ακμές, προτού προστεθεί ακμή, γίνεται έλεγχος ευθείας οράσεως (Line Of Sight) μεταξύ των δύο άκρων (δειγματοληπτώντας σημεία πάνω στη γραμμή). Αν οποιοδήποτε ενδιάμεσο σημείο αντιστοιχεί σε “False” (ξηρά), η ακμή απορρίπτεται. Αυτό εξαλείφει περιπτώσεις όπου μια «σχεδόν ευθεία» θα περνούσε από ακρωτήρια ή νησίδες. Με λίγα λόγια, η sea mask εγγυάται ότι δεν υπάρχουν άκυρες διαδρομές. Είναι ο απλούστερος και πιο στιβαρός τρόπος να μεταφερθεί η γεωγραφική απεικόνιση (ξηρά/θάλασσα) μέσα στον αλγόριθμο δρομολόγησης.

Επιπλέον, αξίζει να σημειωθεί ότι η επιλογή του Resolution είναι ένας συμβιβασμός ανάμεσα στη λεπτομέρεια και το κόστος. Όσο μικρότερη τιμή έχει, τόσο καλύτερα αναγνωρίζει το πλέγμα τις λεπτές ακτογραμμές και τους διαύλους, όμως αυξάνει κατακόρυφα τον αριθμό των κελιών και αναπόφευκτα το χρόνο των χωρικών πράξεων. Στην περιοχή της μελέτης της εργασίας, το 0.01° δίνει χιλιάδες κελιά, αρκετά για να αποτυπωθεί ρεαλιστικά η ακτογραμμή και να γίνουν με ασφάλεια οι έλεγχοι LOS (Line Of Sight) χωρίς υπερβολική επιβάρυνση του υπολογιστή. Αν υπάρξει ανάγκη για υψηλότερη ανάλυση σε ειδικές περιοχές (π.χ. λιμάνια), υπάρχει η δυνατότητα τοπικής υπερδειγματοληψίας μόνο σε αυτή τη μικρή περιοχή, αλλά στο πλαίσιο αυτής της εργασίας, δεν απαιτείται.

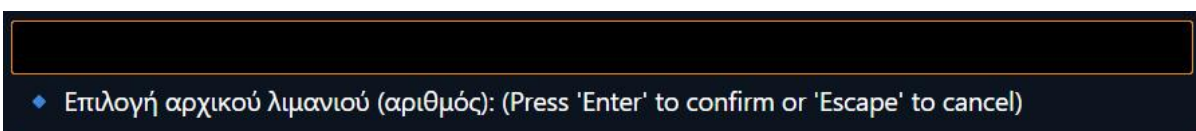
3.4 Route Planner

Το τελευταίο βήμα της εργασίας είναι αυτό που μετατρέπει όλα όσα προετοιμάστηκαν μέχρι εδώ (sea mask, συγχωνευμένα αρχεία δεδομένων AIS και καιρού και το εκπαιδευμένο μοντέλο πρόβλεψης της ταχύτητας) σε σχεδίαση δρομολογίου. Ουσιαστικά, σε αυτό το στάδιο βρίσκεται η ελάχιστη θαλάσσια διαδρομή ανάμεσα σε δύο λιμάνια (προφανώς χωρίς το δρομολόγιο να τέμνει την ξηρά). Αυτό προϋποθέτει ότι ο Route Planner είναι ικανός να εκτιμήσει την ταχύτητα σε κάθε τμήμα με βάση τον τοπικό καιρό και να υπολογίσει τη συνολική απόσταση, τον χρόνο άφιξης και τη μέση

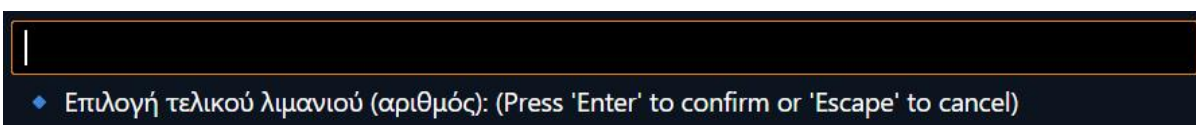
ταχύτητα. Όλα αυτά απεικονίζονται σε χάρτη με χρωματική κωδικοποίηση, ανάλογα με την προβλεπόμενη ταχύτητα.

1) Εκκίνηση διαδικασίας Route Planning

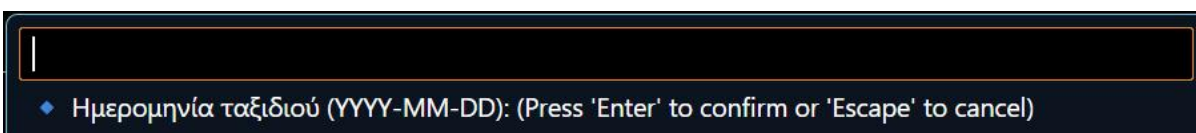
Για να ξεκινήσει η διαδικασία δρομολόγησης, πρέπει ο χρήστης να δώσει τα σημεία εκκίνησης (Εικόνα 8) και άφιξης (Εικόνα 9). Για να είναι πιο εύκολη η διαδικασία, φορτώνονται τα διαθέσιμα λιμάνια από το αντίστοιχο αρχείο με την αντίστοιχη αρίθμηση και ο χρήστης μπορεί να επιλέξει τον αύξοντα αριθμό του λιμανιού. Προφανώς, όπως ειπώθηκε παραπάνω, αφού διαβάζεται το αρχείο, με τα ονόματα και τις συντεταγμένες, γίνεται “snap” στο πλησιέστερο θαλάσσιο κελί/κόμβο. Επιπλέον, βασικές εισοδοι είναι η ημερομηνία (Εικόνα 10) για την οποία υπάρχει ημερήσιο αρχείο δεδομένων AIS και καιρού και ο κωδικός του τύπου του πλοίου (Εικόνα 11), κάτι που είναι απολύτως απαραίτητο για να συνηπολογίσει τον καιρό στην πρόβλεψη της ταχύτητας κατά μήκος των τμημάτων. Στην Εικόνα 12 φαίνονται τα επιμέρους στάδια της διαδικασίας για να διευκολύνεται ο χρήστης πριν δει το τελικό αποτέλεσμα (σε χάρτη).



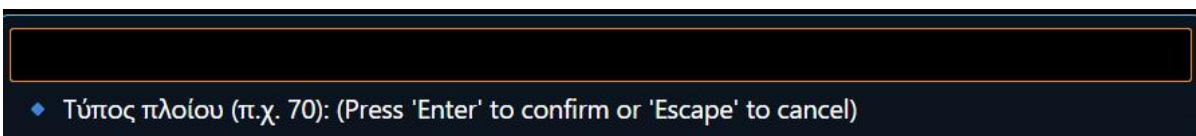
Εικόνα 8: Εισαγωγή αρχικού λιμανιού από το χρήστη



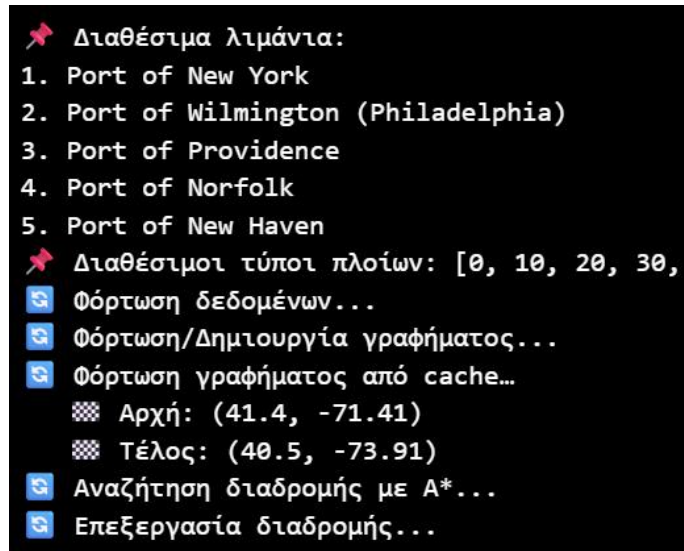
Εικόνα 9: Εισαγωγή τελικού λιμανιού από το χρήστη



Εικόνα 10: Εισαγωγή ημερομηνίας δρομολογίου από το χρήστη



Εικόνα 11: Εισαγωγή κωδικού του τύπου του πλοίου από το χρήστη



Εικόνα 12: Επιμέρους στάδια διαδικασίας δρομολόγησης

2) Σύνδεση με το Sea Mask

Με βάση τη μάσκα όπου δημιουργήθηκε προηγουμένως, το τελευταίο στάδιο της διαδικασίας είναι η κατασκευή του γράφου με τους κόμβους και τις ακμές στη θάλασσα. Για να παραμείνει ο γράφος “ελαφρύς” και να γίνει γρηγορότερη η δημιουργία του, χρησιμοποιείται αραιοποίηση. Ουσιαστικά, από όλα τα κελιά της μάσκας που είναι θάλασσα (True), επιλέγεται υποσύνολο με βήμα 2 (δηλαδή αναζητείται διαδρομή κάθε δεύτερο κελί στο γεωγραφικό πλέγμα). Έτσι επιτυγχάνεται ισορροπία, ούτως ώστε να υπάρχει αρκετή πυκνότητα για να αξιοποιούνται οι πιο περίπλοκες ακτογραμμές, αλλά λίγοι κόμβοι ώστε η αναζήτηση να είναι γρήγορη.

Για κάθε επιλεγμένο κελί της θάλασσας, δημιουργείται κόμβος με γεωμετρικές συντεταγμένες (lat, lon). Στη συνέχεια, ακολουθεί η δημιουργία ακμών, όπου οι κόμβοι αναγνωρίζουν μόνο τους κοντινούς τους γείτονες. Η εγγύτητα δεν υπολογίζεται άναρχα, αφού πρώτα χτίζεται ένας KDTree πάνω στις συντεταγμένες (lat, lon) των κόμβων και για κάθε κόμβο αναζητούνται οι γείτονες εντός μέγιστης απόστασης 25 ναυτικών μιλίων (≤ 25 ν.μ.). Η απόσταση είναι γεωδαιτική (great-circle) και υπολογίζεται με τη συνάρτηση haversine. Ο τύπος haversine υπολογίζει την απόσταση ανάμεσα σε δύο σημεία με γεωγραφικό πλάτος και μήκος πάνω στη σφαιρική Γη, από τη μεταξύ τους κεντρική γωνία. Βασίζεται στη συνάρτηση $\text{hav}(\theta) = \sin^2(\theta/2)$, που είναι αριθμητικά σταθερή για

μικρές γωνίες, και επιστρέφει απευθείας τη γεωδαιτική απόσταση. Στη συγκεκριμένη περίπτωση χρησιμοποιείται ο τύπος haversine για να μετρηθούν τα μήκη των ακμών ανά σκέλος του γράφου και να μετατραπούν σε ναυτικά μίλια. Κάθε υποψήφια ακμή περνά από αυστηρό έλεγχο μέσω του Line-of-Sight (LOS), το οποίο θα αναλυθεί παρακάτω.

3) Έλεγχος Line of Sight (LOS)

Ο έλεγχος LOS εφαρμόζεται για να εξασφαλίσει ότι καμία ακμή δεν θα περάσει πάνω από ξηρά. Η δημιουργία της ευθείας που συνδέει τα άκρα μιας υποψήφιας ακμής (lat1, lon1) και (lat2, lon2) στο επίπεδο lat-lon, απαιτεί τη δειγματοληψία σημείων (π.χ. σε μικρά κλασματικά βήματα). Για κάθε σημείο, υπολογίζεται ο δείκτης πλέγματος (i, j) με τον τύπο: $i = \text{floor}((\text{lat} - \text{lat_min})/\text{Resolution})$, $j = \text{floor}((\text{lon} - \text{lon_min})/\text{Resolution})$. Έπειτα, ελέγχεται στη sea mask το sea_mask[i, j]. Αν οποιοδήποτε δείγμα βρεθεί σε False (ξηρά), η ακμή απορρίπτεται. Μόνο αν όλα τα δείγματα είναι θάλασσα (True) η ακμή γίνεται δεκτή, με βάρος ίσο με τη θαλάσσια απόσταση (σε ναυτικά μίλια). Ο έλεγχος αυτός ενσωματώνει την κρίσιμη παρατήρηση που έγινε στο προηγούμενο υποκεφάλαιο, ότι δηλαδή τα σημεία πάνω στο περίγραμμα λογίζονται ως ξηρά. Επομένως, δεν υπάρχει τομή της ακτογραμμής από ακμές και το γράφημα είναι αμιγώς θαλάσσιο.

4) Αποθήκευση Caching για ταχύτητα

Επειδή η κατασκευή γράφου είναι δαπανηρή, χρησιμοποιείται το cache. Με άλλα λόγια, δημιουργείται ένα κλειδί (hash) που περιγράφει το χώρο και την ανάλυση βάσει των ρυθμίσεων που χρησιμοποιήθηκαν για την κατασκευή του γράφου. Αν βρεθεί ήδη έτοιμος γράφος στο φάκελο που αποθηκεύονται τα "cache" για αυτό το κλειδί, φορτώνεται απευθείας (pickle). Διαφορετικά, κατασκευάζεται από την αρχή και αποθηκεύεται για επόμενες χρήσεις. Έτσι, το κόστος σε χρόνο και επεξεργαστική ισχύ αποδίδεται μόνο την πρώτη φορά σε μια περιοχή - ανάλυση.

5) Χρήση της μετατροπής “snap” στα λιμάνια της διαδρομής

Έχει δημιουργηθεί ένα αρχείο με τις συντεταγμένες των βασικών λιμανιών της περιοχής που εξετάζεται. Τα λιμάνια θεωρούνται σημεία στο χάρτη, που όμως αυτά τα σημεία δεν είναι απαραίτητα πάνω σε κόμβους του γράφου (ούτε καν σε θαλάσσιο κελί, δεδομένου ότι μπορεί να πέφτουν σε προβλήτα). Η διαδικασία σύνδεσης των λιμανιών με τον γράφο έχει δύο βήματα. Πρώτον, γίνεται μετατροπή (“Snap”) σε θάλασσα. Ουσιαστικά, αν οι συντεταγμένες του λιμένα αντιστοιχούν σε κελί False (ξηρά), εκτελείται σπειροειδής αναζήτηση (spiral search) γύρω από το σημείο έως ότου βρεθεί το κοντινότερο θαλάσσιο κελί (True). Αυτό δίνει ένα ασφαλές θαλάσσιο σημείο εκκίνησης και τερματισμού. Δεύτερον, γίνεται ακόμα ένα “snap” σε κόμβο γράφου. Πάνω στον KDTree των κόμβων που έχει δημιουργηθεί, επιλέγεται ο πλησιέστερος κόμβος στο σημείο του πρώτου βήματος (συντεταγμένες λιμανιού). Αυτός γίνεται ο start node (κόμβος έναρξης) ή ο end node (κόμβος τερματισμού) του αλγορίθμου A*. Αυτό το διπλό “snap”, εφαρμόστηκε για να μην υπάρχει εξάρτηση από την αυθαίρετη θέση του λιμένα και σε κάθε περίπτωση η αναζήτηση αρχίζει και τελειώνει σε θαλάσσιο κόμβο.

6) Αλγόριθμος A* με ευρετική haversine

Η αναζήτηση διαδρομής γίνεται με A*, έναν αλγόριθμο που είναι ουσιαστικά συνέχεια του αλγορίθμου Dijkstra με τη διαφοροποίηση ότι σε κάθε κόμβο n , το κόστος προτεραιότητας είναι $g(n) + h(n)$ όπου $g(n)$ το κόστος από την αρχή ως εδώ (άθροισμα ν.μ. των ακμών που διανύθηκαν) και $h(n)$ μια ευρετική εκτίμηση του κόστους που απομένει μέχρι τον στόχο.

Η συνάρτηση haversine είναι ιδανική, διότι εκτιμά τη μεγάλη κυκλική απόσταση ανάμεσα στον τρέχοντα κόμβο και τον κόμβο-στόχο. Επειδή οι πραγματικές διαδρομές στο γράφο δεν μπορεί να είναι συντομότερες από την ευθεία γραμμή, η συνάρτηση δεν υπερεκτιμά ποτέ το πραγματικό κόστος και σε μεγάλο βαθμό είναι συνεπής. Άρα ο αλγόριθμος A* παραμένει βέλτιστος (επιστρέφει διαδρομή ελάχιστης απόστασης στο γράφο και ταυτόχρονα αποφεύγει να εξερευνήσει άσχετες περιοχές). Το αποτέλεσμα είναι μία ακολουθία κόμβων που, όταν ενωθούν με ευθύγραμμα τμήματα, δίνουν

διαδρομή πάντα μέσω θαλάσσης, με συνολικό μήκος το άθροισμα των αποστάσεων των ακμών σε ναυτικά μίλια.

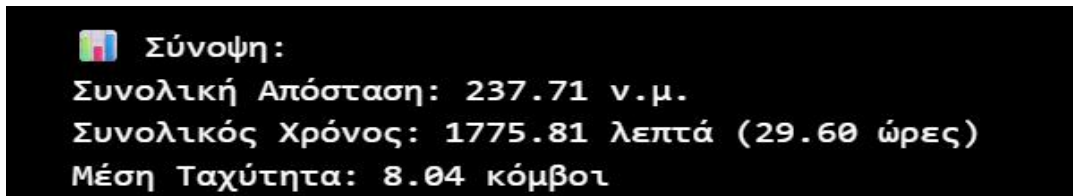
7) Διαχωρισμός διαδρομής σε τμήματα και συσχέτιση με τον καιρό της περιοχής

Δεδομένου ότι έχει σχεδιαστεί η διαδρομή, το επόμενο στάδιο για την απεικόνιση του αποτελέσματος είναι η εκτίμηση της ταχύτητας του σκέλους και, ως συνέπεια, η εκτίμηση του απαιτούμενου χρόνου. Η διαδρομή περιγράφεται από διαδοχικούς κόμβους ($P_0, P_1, P_2, \dots, P_k$), όπου για κάθε ζεύγος (P_i, P_{i+1}) ορίζεται ένα τμήμα (segment). Για κάθε τμήμα υπολογίζεται η απόσταση σε ναυτικά μίλια (υπολογισμός κόστους με χρήση της συνάρτησης haversine μεταξύ P_i και P_{i+1}) και το ενδιάμεσο σημείο (midpoint) του τμήματος (μέσο lat, lon). Για την εύρεση του καιρού στο ενδιάμεσο αυτό σημείο, διαβάζεται το ημερήσιο αρχείο με τα δεδομένα AIS και καιρού, διατηρώντας μόνο τις στήλες με τις συντεταγμένες και τις μεταβλητές του καιρού (ταχύτητα ανέμου, ύψος κύματος), φιλτράροντας την περιοχή γύρω από τη διαδρομή (ώστε να μην φορτώνεται άσκοπα όλη η μέρα). Από το Dataframe που προκύπτει εντοπίζεται για κάθε ενδιάμεσο σημείο ο πλησιέστερος καιρός (nearest neighbor σε χώρο) ανά τμήμα.

8) Πρόβλεψη ταχύτητας και χρόνου ανά σκέλος με το εκπαιδευμένο μοντέλο πρόβλεψης ταχύτητας

Δεδομένου ότι τα στοιχεία καιρού είναι διαθέσιμα, αρχικά σχηματίζεται το διάνυσμα των χαρακτηριστικών για κάθε σκέλος βάσει των συντεταγμένων του ενδιάμεσου σημείου. Έπειτα, χρησιμοποιείται ο τύπος του πλοίου που δίνεται από το χρήστη κατά την έναρξη της διαδικασίας, φορτώνεται το εκπαιδευμένο μοντέλο Random Forest και εκτελείται μαζική πρόβλεψη για όλα τα σκέλη της διαδρομής. Έτσι, για κάθε τμήμα γίνεται η πρόβλεψη της ταχύτητας. Ο χρόνος προκύπτει από το πηλίκο της απόστασης (σε ναυτικά μίλια) δια της εκτιμώμενης ταχύτητας του σκέλους. Για κάθε τμήμα υπολογίζεται ο χρόνος (σε ώρες και λεπτά) και, τελικά, προκύπτει η συνολική απόσταση (σε ν.μ.) από το άθροισμα όλων των αποστάσεων των ακμών. Ομοίως, ο συνολικός χρόνος (ώρες) προκύπτει από το άθροισμα όλων των πηλίκων των αποστάσεων δια των αντίστοιχων ταχυτήτων. Η μέση ταχύτητα της διαδρομής υπολογίζεται γρήγορα από το

σταθμισμένο μέσο όρο των ταχυτήτων με βάση τις αποστάσεις. Αυτά τα τρία μεγέθη (απόσταση, χρόνος, μέση ταχύτητα) είναι τα χρήσιμα δεδομένα που αναζητά ο χρήστης για τον προγραμματισμό ενός δρομολογίου (Εικόνα 13). Επιπλέον, επειδή όλες οι εκτιμώμενες ταχύτητες είναι διαθέσιμες ανά σκέλος, είναι εφικτή η δημιουργία πίνακα αναλυτικών αποτελεσμάτων για κάθε σκέλος.

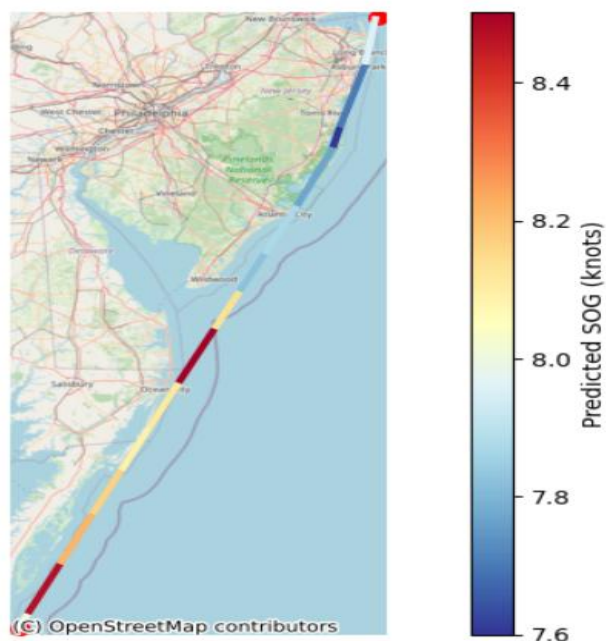


Εικόνα 13: Καταγραφή συνολικής απόστασης, συνολικού χρόνου και μέσης ταχύτητας

9) Χάρτης με χρωματική κωδικοποίηση ταχύτητας

Μετά τον υπολογισμό των κρίσιμων μεγεθών (απόσταση, χρόνος, ταχύτητα), φορτώνεται ο χάρτης που περιλαμβάνει τα σημεία της διαδρομής, στον οποίο προβάλλεται η τεθλασμένη γραμμή της διαδρομής, και κάθε τμήμα χρωματίζεται με χρώμα ανάλογο της εκτιμώμενης ταχύτητας. Στο υπόμνημα του χάρτη φαίνεται το εύρος των ταχυτήτων (Εικόνα 14).

Βέλτιστη Διαδρομή με Χρωματική Κωδικοποίηση SOG



Εικόνα 14: Τελικό αποτέλεσμα χάρτη με σχεδιασμένο δρομολόγιο

4 Αποτελέσματα και Αξιολόγηση

4.1 Αξιολόγηση μοντέλου πρόβλεψης της ταχύτητας

Όσον αφορά το μοντέλο παλινδρόμησης για την πρόβλεψη της ταχύτητας, στο σύνολο της εκπαίδευσης (80/20 με σταθερό seed), το συνολικό MAE διαμορφώνεται περίπου στον 1.14 kn, ενώ ο δείκτης R^2 φτάνει περίπου το 0.85. Η ανάγνωση αυτών των δύο μεγεθών δεν είναι απλώς αξιόπιστη, αλλά πρόκειται για σφάλμα που βρίσκεται σταθερά κάτω από 1.2 κόμβους κατά μέσο όρο, σε πρόβλημα με έντονη ετερογένεια (τύποι πλοίων, μεγάλες αποστάσεις δρομολογίων, διαφορών επιπέδων καιρικές εντάσεις) και με πολύ ασθενή γραμμικότητα. Επιπλέον, ο υψηλός R^2 υποδηλώνει ότι το μοντέλο εξηγεί ένα μεγάλο μέρος της διακύμανσης της πραγματικής ταχύτητας, γεγονός που αντανακλάται στην πρακτική ακρίβεια του εκτιμώμενου χρόνου άφιξης, όταν οι επιμέρους προβλέψεις συντίθενται κατά μήκος πραγματικής διαδρομής.

Αξίζει να ξεκαθαριστεί πρώτα το ζήτημα των μετρικών. Όταν πρόκειται για συνολικό MAE και συνολικό R^2 πάνω σε ολόκληρο το test set, ουσιαστικά υιοθετείται μικρο-μέσο όρο (micro average), όπου κάθε γραμμή/παράδειγμα έχει ίσο βάρος, άρα οι πολυπληθείς κατηγορίες επηρεάζουν περισσότερο το τελικό νούμερο. Αυτό είναι επιθυμητό στην περίπτωση της συγκεκριμένης εργασίας, γιατί το πραγματικό φορτίο χρήσης του συστήματος προέρχεται ακριβώς από αυτές τις κατηγορίες με μεγάλο αριθμό καταγραφών. Το εναλλακτικό σενάριο, με μακρο-μέσο όρο (macro average) ανά τύπο πλοίου που δίνει ίδιο βάρος σε κάθε κλάση, θα αυξήσει πάρα πολύ την επίδραση των πολύ μικρών κατηγοριών όπου οι μετρικές είναι ευμετάβλητες λόγω μικρής υποστήριξης. Με άλλα λόγια, ο αναφερόμενος συνολικός $MAE \approx 1.14$ και $R^2 \approx 0.85$ είναι ήδη, στην πράξη, ζυγισμένος με βάση τον αριθμό των δειγμάτων.

Στα στατιστικά ανά τύπο πλοίου, η εικόνα είναι διαφωτιστική. Οι πολυπληθείς κλάσεις αποδίδουν πολύ καλά, δείχνοντας ότι το δάσος μαθαίνει σταθερά μοτίβα συμπεριφοράς. Για παράδειγμα, για τον τύπο πλοίου 70 καταγράφεται $MAE \approx 1.15$ kn και $R^2 \approx 0.91$ σε δείγμα 286 χιλιάδων γραμμών, για τον τύπο πλοίου 80 καταγράφεται $MAE \approx 1.01$ kn και $R^2 \approx 0.92$ σε 62 χιλιάδες γραμμές, ενώ για την εξαιρετικά πολυπληθή κατηγορία 30 το

MAE \approx 1.03 kn με R² \approx 0.79 σε περίπου 982 χιλιάδες γραμμές. Αυτές οι τιμές δεν είναι τυχαίες και δείχνουν ότι ο συνδυασμός των γεωγραφικών συντεταγμένων (χωρικό πλαίσιο), τύπου πλοίου (προφίλ ταχύτητας ανά κλάση) και καιρικών δεδομένων (ένταση περιβάλλοντος) έχει επαρκή πληροφορία ώστε να παραχθεί προβλεπτική ισχύς με σταθερότητα. Σε δεύτερο επίπεδο, κλάσεις όπως οι τύποι 50 και 60 (με δεκάδες χιλιάδες γραμμές) κινούνται επίσης σε χαμηλά MAE (\sim 1.26-1.30 kn) με R² περίπου 0.78-0.82, ενώ οι κλάσεις 90, 1001, 1024 και 1025 εμφανίζονται με πολύ καλή ισορροπία μεταξύ σφάλματος και αριθμού καταγραφών. Για παράδειγμα, διακρίνονται οι τύποι πλοίων 1001 με MAE \approx 1.08 και R² \approx 0.75 σε 99 χιλιάδες καταγραφές, 1024 με MAE \approx 0.99 και R² \approx 0.93 σε 37 χιλιάδες καταγραφές, 1025 με MAE \approx 1.19 και R² \approx 0.81 σε 50 χιλιάδες καταγραφές. Σε όλες αυτές τις περιπτώσεις, το υψηλό πλήθος καταγραφών επιτρέπει αξιόπιστες εκτιμήσεις των μετρικών, χωρίς την επίδραση του θορύβου από μικρές διακυμάνσεις.

Στον αντίποδα, υπάρχουν κλάσεις με πολύ μικρό αριθμό καταγραφών, όπου οι μετρικές δεν έχουν στατιστική σταθερότητα. Για παράδειγμα, τιμές όπως R²<0 ή έντονα κυμαινόμενα MAE σε κλάσεις με μονοψήφιο ή διψήφιο αριθμό δειγμάτων (κάτω από 50 καταγραφές) δεν είναι ενδείξεις αποτυχίας του μοντέλου, απλώς καταγράφουν ότι το σύστημα δεν έχει αρκετή πληροφορία για να μάθει σταθερά μοτίβα σε αυτές τις σπάνιες περιπτώσεις. Στον πίνακα της εικόνας 15 (παρακάτω), φαίνονται ενδεικτικά κλάσεις πλοίων όπως 107 (Count=4), 207 (Count=7), 255 (Count=9), όπου ο R² μπορεί να είναι αρνητικός ή χαμηλός και το MAE εμφανίζει διασπορά. Αυτά τα παραδείγματα λειτουργούν κυρίως ως υπενθύμιση ότι η ερμηνεία των μετρικών πρέπει να γίνεται με αναφορά στον αριθμό των καταγραφών της κάθε κλάσης, καθώς ο μεγάλος αριθμός καταγραφών συνεπάγεται πιο σταθερές μετρικές, ενώ όταν ο αριθμός καταγραφών είναι πολύ μικρός προκύπτουν μετρικές που μεταβάλλονται έντονα από λίγα περιστατικά.

Αξιολόγηση ανά VesselType:				
	VesselType	MAE	R2	Count
0	0	1.532844	0.777090	397.0
1	10	1.402697	0.808929	353.0
2	20	1.187848	0.678822	105.0
3	30	1.025412	0.787289	982540.0
4	40	1.948322	0.726735	593.0
5	50	1.259841	0.776175	44324.0
6	60	1.294795	0.818308	74188.0
7	70	1.153413	0.907792	286643.0
8	80	1.014238	0.924403	62444.0
9	90	1.325571	0.770313	77206.0
10	107	0.280200	-0.107772	4.0
11	200	1.897724	0.843550	5.0
12	207	1.793939	-8.611985	7.0
13	255	2.886980	0.547725	9.0

Εικόνα 15: Απεικόνιση μετρικών ορισμένων τύπων πλοίων

Πρακτικά, αυτό σημαίνει ότι ο συνολικός $MAE \approx 1.14$ και $R^2 \approx 0.85$ αντικατοπτρίζει την απόδοση εκεί ακριβώς όπου υπάρχει το μεγαλύτερο όφελος, στις κλάσεις που συναντώνται συχνότερα στο πεδίο και, κατά συνέπεια, σε εκείνες που συνθέτουν το μεγαλύτερο μέρος της ζητούμενης πρόβλεψης ταχύτητας κατά μήκος μιας διαδρομής. Εφόσον ο δρομολογητής (Route Planner) χωρίζει μια διαδρομή σε πολλά σκέλη και ζητάει την ταχύτητα για καθένα τους, ο συνολικός χρόνος άφιξης εξαρτάται από τη μέση ακρίβεια των επιμέρους προβλέψεων. Οι μικρές, σποραδικές κλάσεις έχουν μικρό βάρος σε αυτή τη σύνθεση, ενώ οι μεγάλες κλάσεις (εκεί όπου ήδη υπάρχει χαμηλό MAE και υψηλό R^2) διαμορφώνουν την κύρια εικόνα χρόνου ταξιδιού.

Ένα πρακτικό ερώτημα είναι πώς μεταφράζεται ο δείκτης $MAE \approx 1.14$ kn σε όρους χρόνου. Αν κάποιος επιχειρούσε μια απλοϊκή ερμηνεία του σταθερού σφάλματος ταχύτητας σε όλη τη διαδρομή, θα έπαιρνε μια συντηρητική εκτίμηση του σφάλματος στον χρόνο. Στην πράξη όμως, ο Route Planner χωρίζει την πορεία σε πολλά σκέλη με διαφορετική τοπική ταχύτητα, και τα σφάλματα τείνουν να μην αθροίζονται γραμμικά (μερικές φορές σβήνουν μεταξύ τους). Έτσι, ο τελικός εκτιμώμενος χρόνος προκύπτει συχνά σταθερός σε σχέση με την πραγματικότητα, ειδικά όταν η διαδρομή διέρχεται από τμήματα όπου το μοντέλο (λόγω της εκπαίδευσης) είναι ούτως ή άλλως πολύ ακριβές (όπως για παράδειγμα οι κατηγορίες πλοίων 30, 70, 80, 90, 1001). Αυτό είναι εμφανές όταν συγκρίνονται διαδρομές ίδιου γεωμετρικού μήκους, αλλά με διαφορετικό προφίλ καιρού,

όπου παρατηρείται ότι οι εκτιμήσεις του χρόνου διαφέρουν εκεί που πρέπει (σε δυσμενείς καιρούς), με μέτριο σφάλμα που παραμένει στα μερικά λεπτά ανά τμήμα και όχι σε ώρες.

Τέλος, είναι σημαντικό να αναφερθεί ότι η πρόβλεψη ανά γραμμή (χωρίς χρονική μνήμη) λειτουργεί πλεονεκτικά εδώ. Η ταχύτητα που ζητά ο Route Planner αφορά τοπικές, στιγμιαίες συνθήκες σε συγκεκριμένα ενδιάμεσα σημεία των τμημάτων. Η χρήση καιρικών μεταβλητών από τα συγχωνευμένα αρχεία (δεδομένα AIS με δεδομένα καιρού) της ίδιας ημέρας, η μεταβλητή που προκύπτει από τον πολλαπλασιασμό των δυο μεταβλητών (ένταση ανέμου x ύψος κύματος) και η υπόθεση ότι το πλοίο κινείται, φτιάχνουν ένα συνεκτικό διάνυσμα εισόδου που το μοντέλο έχει αξιοποιήσει στην εκπαίδευση με πολύ παρόμοιο τρόπο. Αυτό εξηγεί γιατί, παρά την απουσία ακολουθιακής μνήμης, το σύστημα επιτυγχάνει υψηλό R^2 και χαμηλό MAE, δεδομένου ότι τα χαρακτηριστικά που τροφοδοτούν την πρόβλεψη αντιστοιχούν στενά στην ταχύτητα του πλοίου σε συγκεκριμένο σημείο και στιγμή.

Συνοψίζοντας, η απόδοση του μοντέλου ($MAE \approx 1.14$ kn και $R^2 \approx 0.85$) είναι η ένδειξη ενός μοντέλου που γενικεύει εκεί όπου υπάρχουν πολλά δεδομένα και που, παράλληλα, παραμένει χρηστικό ως προς το τελικό ζητούμενο, δηλαδή τη σταθερή ταχύτητα σε διαδρομές με ποικίλες συνθήκες. Οι κλάσεις με πολλές καταγραφές αποτελούν τη βάση της πρόβλεψης και εξηγούν γιατί η συνολική εικόνα είναι τόσο θετική. Οι υπόλοιπες κλάσεις με τις λίγες καταγραφές παραμένουν οριακές για το συνολικό σκορ και, κυρίως, δεν υπονομεύουν τη χρησιμότητα του συστήματος στην πράξη. Με αυτή την οπτική, το υποσύστημα της ταχύτητας επιτυγχάνει να προβλέψει τις ακριβείς ταχύτητες που χρειάζεται ο Route Planner για να μετατρέψει αποστάσεις σε χρόνο.

4.2 Παραδείγματα διαδρομών

Σε αυτή την ενότητα παρουσιάζονται δύο εκτελέσεις του δρομολογητή για το ίδιο δρομολόγιο και το ίδιο τύπο πλοίου (30), αλλά σε διαφορετικές ημέρες και μία εκτέλεση για το ίδιο δρομολόγιο άλλου τύπου πλοίου. Η πορεία είναι από το “Port of New York” προς το “Port of Norfolk”, ακολουθώντας την ανατολική ακτή, το οποίο είναι ένα σενάριο

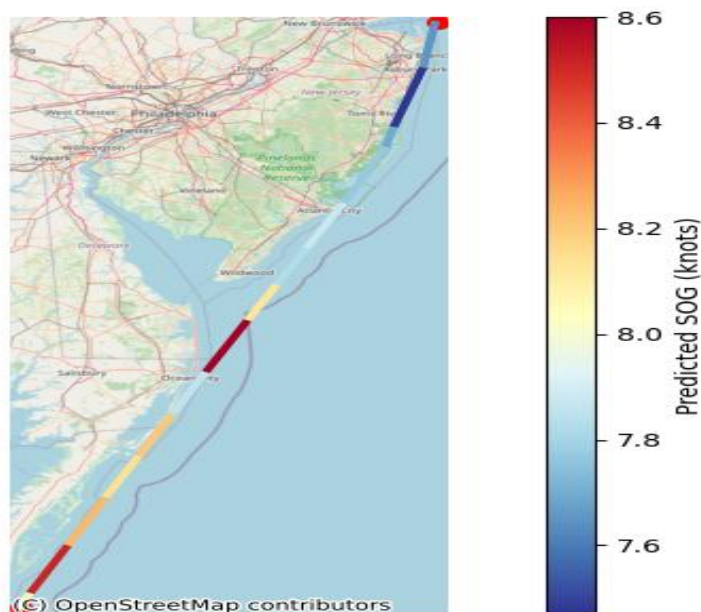
που βρίσκεται εξ ολοκλήρου μέσα στην περιοχή της μελέτης και εκθέτει τη διαδρομή σε παράκτιους και ανοικτούς τομείς με διαφορετικά θαλάσσια καθεστώτα. Σε κάθε δοκιμή του προγράμματος, ο Route Planner φορτώνει τον θαλάσσιο γράφο από τη sea mask, βρίσκει τη βέλτιστη πορεία με A* και στη συνέχεια σχεδιάζει την πορεία. Η λογική της σχεδίασης, όπως φαίνεται στις εικόνες 16 και 17, για κάθε σκέλος είναι η λήψη του τοπικού καιρού της αντίστοιχης ημερομηνίας από το αντίστοιχο αρχείο και η χρησιμοποίηση της εκτιμώμενης ταχύτητας από το εκπαιδευμένο μοντέλο Random Forest. Γνωρίζοντας την απόσταση και την ταχύτητα, προκύπτει άμεσα ο χρόνος κάθε τμήματος και κατ' επέκταση προκύπτουν η συνολική απόσταση, ο συνολικός χρόνος και η μέση ταχύτητα.

1) Πρώτη εκτέλεση (Ημερομηνία: 06-08-2016 και τύπος πλοίου 30) (Εικόνα 16)

Ο πρώτος χάρτης (06-08-2016) δείχνει την πορεία χρωματισμένη με βάση την προβλεπόμενη ταχύτητα (όπως φαίνεται από τη χρωματική κλίμακα 7.6-8.6 kn). Παρατηρείται ισχυρός χρωματικός βαθμός στα τμήματα κοντά στο βόρειο άκρο (είσοδος/έξοδος στον κόλπο της Νέας Υόρκης), όπου εμφανίζονται σε ψυχρούς τόνους (μπλε), που αντιστοιχούν σε χαμηλότερες SOG (7.6-7.8 kn), ενώ τμήματα νοτιότερα, σε πιο ανοιχτή θάλασσα, γίνονται θερμά (κίτρινο και κόκκινο), με υψηλότερες SOG (8.2-8.6 kn). Η ερμηνεία αυτής της μετάβασης είναι ότι καθώς απομακρύνεται ένα πλοίο από τις παράκτιες περιοχές και, συνδυαστικά, μεταβάλλονται οι τιμές του καιρού, το μοντέλο προβλέπει μεγαλύτερη ταχύτητα. Το γεγονός ότι η κλίμακα φτάνει έως τους 8.6 kn σημαίνει πως υπήρξαν τμήματα με σαφώς ευνοϊκότερες τοπικές συνθήκες (είτε χαμηλότερο κύμα είτε άνεμο που δεν πρόσθετε σημαντική αντίσταση). Με άλλα λόγια, στα πρώτα τμήματα (πλησίον της Νέας Υόρκης) εμφανίζονται μικρότερες τιμές εκτιμώμενης ταχύτητας και αντίστοιχα μεγαλύτεροι χρόνοι ανά ναυτικό μίλι, ενώ στα μεσαία τμήματα η SOG αυξάνεται και ο χρόνος ανά ναυτικό μίλι μειώνεται. Στο τέλος, πλησιάζοντας προς το Norfolk, το προφίλ μπορεί να μετριάζεται ή να μεταβάλλεται ανάλογα με τα τοπικά δεδομένα καιρού των τελευταίων ενδιάμεσων σημείων κάθε σκέλους. Σε επίπεδο ερμηνείας, αυτή η εκτέλεση δείχνει καθαρά τον τρόπο που λειτουργεί το μοντέλο. Δηλαδή ο τύπος του πλοίου (30) δίνει την βάση γύρω από την οποία κυμαίνεται η ταχύτητα, ο καιρός διορθώνει την ταχύτητα τοπικά, και η γεωγραφία

παρέχει το δρομολόγιο. Ο χρήστης, χωρίς να χρειαστεί να μελετήσει κάποιο σχετικό πίνακα, εντοπίζει άμεσα πάνω στον χάρτη πού το πλοίο αναμένεται να επιβραδύνει (μπλε κομμάτια) και πού θα κινηθεί γρηγορότερα (πορτοκαλί-κόκκινα). Αυτή η οπτική αντιστοιχία κάνει την παρουσίαση πιο σαφή και ευανάγνωστη.

Βέλτιστη Διαδρομή με Χρωματική Κωδικοποίηση SOG



Εικόνα 16: Δρομολόγιο Port of New York - Port of Norfolk (Ημερομηνία: 06/08/2016)

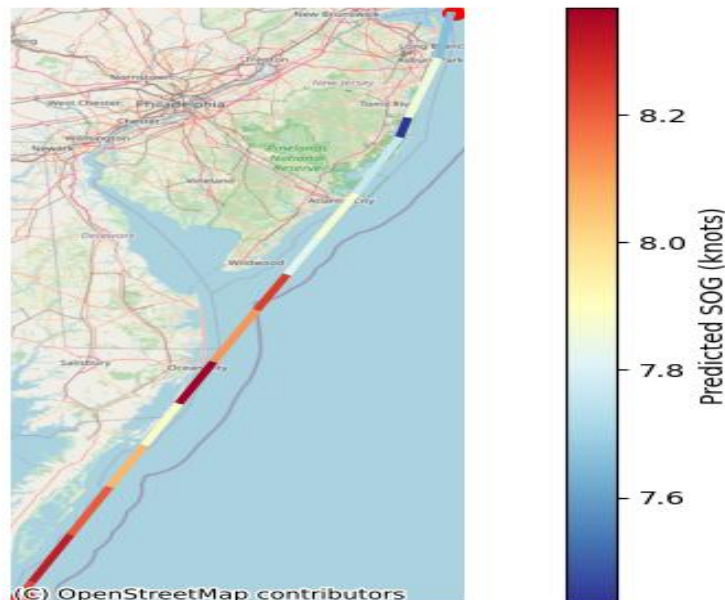
Τύπος πλοίου 30

2) Δεύτερη εκτέλεση (Ημερομηνία: 01-01-2019 και τύπος πλοίου 30) (Εικόνα 17)

Ο δεύτερος χάρτης αφορά το ίδιο ζεύγος λιμένων και τον ίδιο τύπο πλοίου, αλλά σε άλλη ημέρα (01-01-2019). Η κλίμακα της ταχύτητας εδώ κυμαίνεται περίπου μεταξύ 7.5 με 8.3 kn. Ο συνολικός χρωματισμός μοιάζει πιο ομοιογενής (με έντονες θερμές ραβδώσεις κατά μήκος της κύριας θαλάσσιας διαδρομής και περιορισμένα μπλε σκέλη), υποδηλώνοντας ότι μεγάλο μέρος της πορείας βρέθηκε σε συνθήκες που το μοντέλο εκτίμησε ως σταθερές. Παρά το γεγονός ότι το μέγιστο της κλίμακας είναι ελαφρώς χαμηλότερο (8.3 έναντι 8.6 kn στο προηγούμενο χάρτη), η μεγάλη σταθερότητα κατά μήκος της διαδρομής συχνά συνεπάγεται βέλτιστο χρόνο άφιξης και οι μικρές τοπικές διακυμάνσεις τείνουν να εξισορροπούνται καθώς αθροίζονται τα πολλά σκέλη με παρόμοιες ταχύτητες. Σε αυτή τη διαδρομή αναμένονται πολλές γραμμές με ταχύτητα

πρόβλεψης σταθερή (περίπου 7.9-8.2 kn), δηλαδή χρόνους/ν.μ. κοντά στο ιδανικό σημείο για τη συγκεκριμένη κλάση. Αν υπάρχει κάποιο σκέλος με χρωματισμό μπλε (π.χ. έξω από τη Νέα Υόρκη), θα θεωρηθεί ως outlier με ελαφρώς μεγαλύτερο χρόνο. Ωστόσο, η επίδρασή αυτού του σκέλους στον συνολικό χρόνο ταξιδιού είναι μικρή όταν το υπόλοιπο μονοπάτι παραμένει σε μεγαλύτερες ταχύτητες. Στη σύνοψη η μέση ταχύτητες υπολογίζονται ως το πηλίκο της συνολικής απόστασης δια το συνολικό χρόνο (σταθμισμένος μέσος), γι' αυτό και οι τοπικές διαφορετικές ταχύτητες συνεισφέρουν με βάση το μήκος τμήματος που αντιστοιχούν.

Βέλτιστη Διαδρομή με Χρωματική Κωδικοποίηση SOG



Εικόνα 17 : Δρομολόγιο Port of New York - Port of Norfolk (Ημερομηνία: 01/01/2019) -

Τύπος πλοίου 30

3) Σύγκριση και ανάγνωση των δύο χαρτών

Το ζεύγος χαρτών εξυπηρετεί τον βασικό στόχο της απομόνωσης της επίδρασης της ημερομηνίας (καιρού) κρατώντας σταθερά τα υπόλοιπα (γεωμετρία και τύπο πλοίου). Ακόμη και οπτικά, φαίνεται ότι στην πρώτη εκτέλεση υπάρχει πιο έντονη διακύμανση ταχύτητας μέσα στην ίδια διαδρομή (από 7.6 έως 8.6 kn), ενώ στη δεύτερη απεικονίζονται σκέλη με μικρότερη διαφορά μεταξύ ελαχίστου και μεγίστου (7.5-8.3 kn). Αυτό σημαίνει ότι στην πρώτη περίπτωση η μέση ταχύτητα μπορεί να καθοριστεί από

αυξομειώσεις ανάμεσα σε τμήματα περισσότερο ευνοϊκά και τμήματα πιο αργά, ενώ στη δεύτερη περίπτωση η μέση ταχύτητα καθορίζεται από συσσωρευμένη ομοιότητα (πολλά σκέλη σε παρόμοιο εύρος). Πρακτικά, οι δύο εικόνες παρουσιάζουν την ίδια γεωμετρία πορείας, αλλά όχι τον ίδιο χρόνο.

Σημαντικό χαρακτηριστικό είναι το μέρος που συμβαίνουν οι διαφορές. Στο βόρειο τμήμα της διαδρομής φαίνεται και στις δύο εκτελέσεις το μπλε χρώμα στο συγκεκριμένο σημείο, γεγονός που υποδεικνύει ένα συνδυασμό παράκτιου περιβάλλοντος και τοπικού κύματος/ανέμου που τείνει να μειώνει την ταχύτητα εκεί. Αντίθετα, στα εκτεθειμένα τμήματα παράλληλα της ακτής του New Jersey και προς τα νότια, τα αντίστοιχα σκέλη στις δυο εκτελέσεις της διαδρομής γίνεται θερμότερα, άρα χαρακτηρίζεται από ελεύθερη πλεύση υπό καιρό που, στις μέρες της σύγκρισης, φαίνεται να ήταν σε γενικές γραμμές ανεκτός.

4) Σύνδεση με τις μετρικές του μοντέλου SOG

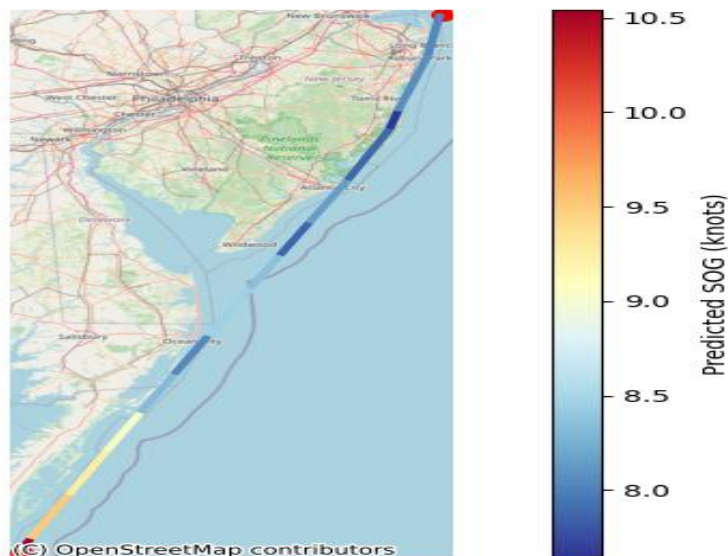
Η επιλογή του τύπου του πλοίου με αριθμό 30 έγινε καθώς πρόκειται για εξαιρετικά πολυπληθή κλάση, με χαμηλό MAE ≈ 1.03 kn και $R^2 \approx 0.79$. Ο μεγάλος αριθμός καταγραφών στα δεδομένα κάνει τη μέση συμπεριφορά πολύ προβλέψιμη. Για αυτό το λόγο, στα παραδείγματα διαδρομών, οι χρωματισμοί στους χάρτες έχουν νόημα, καθώς το μοντέλο έχει εκπαιδευτεί αρκετά σε παρόμοια στιγμιότυπα και μαθαίνει αξιόπιστα πώς ο συνδυασμός καιρού και θέσης μεταφράζεται σε ταχύτητα. Με άλλα λόγια, ο εκτιμώμενος χρόνος που υπολογίζεται πάνω στα σκέλη κληρονομεί τη σταθερότητα του χρησιμοποιούμενου μοντέλου ταχύτητας.

5) Τρίτη εκτέλεση (Ημερομηνία: 01-01-2019 και τύπος πλοίου 1001) (Εικόνα 18)

Η τρίτη εικόνα (Εικόνα 18) συνεχίζει τη σύγκριση, κρατώντας την ίδια γεωμετρία και την ίδια ημερομηνία (01-01-2019) με την δεύτερη εκτέλεση, αλλά αλλάζοντας τον τύπο του πλοίου σε 1001. Το πρώτο που ξεχωρίζει είναι η κλίμακα της ταχύτητας, καθώς το εύρος της χρωματικής μπάρας ανεβαίνει έως 10.5 kn, σημαντικά υψηλότερα από τους 8.3 kn που παρουσιάστηκε για τον τύπο με αριθμό 30 την ίδια μέρα. Αυτό αποτυπώνει ακριβώς

τον ρόλο του τύπου του πλοίου ως βάση στο μοντέλο. Δηλαδή, με πανομοιότυπες καιρικές τιμές κατά μήκος των ενδιαμέσων σημείων του κάθε σκέλους, ο τύπος πλοίου 1001 μεταφράζει τον ίδιο καιρό σε μεγαλύτερη επιτεύξιμη ταχύτητα. Παρ' όλα αυτά, η χωρική υφή των χρωμάτων παραμένει συνεπής με ό,τι περιγράφηκε προηγουμένως. Στην είσοδο/έξοδο της Νέας Υόρκης και στο New York Bight διακρίνονται σκούρες μπλε νησίδες (χαμηλότερη ταχύτητα), ενώ νοτιότερα, σε πιο ανοιχτή θάλασσα, εμφανίζονται κιτρινωπές/ανοικτές ζώνες (υψηλότερη ταχύτητα). Δηλαδή, ο γεωγραφική απεικόνιση είναι παρόμοια, αλλά η ένδειξη της ταχύτητας όλης της διαδρομής έχει αυξηθεί. Όσον αφορά την ερμηνεία, αυτό σημαίνει ότι για τον τύπο πλοίου 1001 ο εκτιμώμενος χρόνος της ίδιας πορείας τείνει να είναι μικρότερος, επειδή μεγάλα τμήματα έχουν ταχύτητες 9-10 kn αντί για 8 kn. Η αξιοπιστία της εκτίμησης στηρίζεται και στα συνολικά μετρικά του τύπου 1001 ($MAE = 1.08$, $R^2 = 0.75$ με 99 χιλιάδες δείγματα), άρα οι προβλέψεις ανά σκέλος δεν είναι μεμονωμένες περιπτώσεις, αλλά στατιστικά σταθερές εκτιμήσεις για αυτήν την κλάση. Σε συνδυασμό με τις δύο προηγούμενες εκτελέσεις, φαίνεται ότι για τον ίδιο καιρό και ίδια διαδρομή, αλλά με διαφορετικό τύπο πλοίου, θα προκύψει διαφορετική ταχύτητα και, κατ' επέκταση, διαφορετικός συνολικός χρόνος. Αυτό είναι ακριβώς το είδος διαφοροποίησης που η εργασία στόχευε να αναδείξει μέσω του δρομολογητή για συγκρίσεις μεταξύ πλοίων

Βέλτιστη Διαδρομή με Χρωματική Κωδικοποίηση SOG



Εικόνα 18: Δρομολόγιο Port of New York - Port of Norfolk (Ημερομηνία: 01/01/2019) -

Τύπος πλοίου 1001

4.3 Ευαισθησία παραμέτρων δρομολογητή, μέγιστο μήκος ακμής, αραιοποίηση πλέγματος

Ο Route Planner βασίζεται σε γράφο που χρησιμοποιεί τη γεωμετρία της θάλασσας από τη sea mask, με στόχο να οργανώσει την αναζήτηση με τον αλγόριθμο A*. Η ποιότητα και η ταχύτητα του αποτελέσματος εξαρτώνται έντονα από τρεις παραμέτρους, οι οποίες είναι το μέγιστο μήκος ακμής (σε ν.μ.), η αραιοποίηση του πλέγματος κόμβων ("Grid Step") και τα όρια της περιοχής μέσα στην οποία κατασκευάζεται ο γράφος. Κάθε μία επηρεάζει άμεσα την πληρότητα των εναλλακτικών και το υπολογιστικό κόστος.

Το μέγιστο μήκος κάθε ακμής θέτει τη μέγιστη απόσταση που επιτρέπεται να συνδέονται δύο κόμβοι της θάλασσας. Με μικρό όριο ο γράφος αναγκάζεται να χωρίζει τη διαδρομή σε περισσότερα σύντομα σκέλη. Αυτό αυξάνει τις επιλογές κοντά στην ακτογραμμή, βοηθά να σχεδιαστεί γραμμή γύρω από ακρωτήρια και μικρές νησίδες, και κάνει τον έλεγχο LOS ευκολότερο (η ευθεία δύο κοντινών κόμβων σπανιότερα τέμνει ξηρά). Το τίμημα είναι περισσότερες ακμές, μεγαλύτερος χώρος μνήμης και περισσότερη δουλειά για το A*. Με μεγάλο όριο ο γράφος γίνεται αραιότερος, καθώς μειώνονται οι ακμές και επιταχύνεται η αναζήτηση, αλλά αυξάνεται ο κίνδυνος οι ευθείες να προσπαθούν να παρακάμψουν τις λεπτές λεπτομέρειες της ακτογραμμής. Παρότι ο έλεγχος LOS θα απορρίψει τις άκυρες ακμές, ο αλγόριθμος ίσως χρειαστεί να εξετάσει περισσότερους συνδυασμούς πριν βρει την επιτρεπτή, με συνέπεια έναν μεγαλύτερο χρόνο εκτέλεσης. Στην πράξη, η επιλογή της τιμής των 25 ν.μ. ισορροπεί την ακρίβεια με την ταχύτητα για τη περιοχή μελέτης της εργασίας.

Η αραιοποίηση των κόμβων ("Grid Step") ελέγχει πόσα κελιά της sea mask γίνονται κόμβοι. Μικρό βήμα (π.χ. "Grid Step"=1) σημαίνει ότι το δίκτυο θα είναι πολύ πυκνό, άρα με πολλές εναλλακτικές πορείες, ιδίως σε στενούς διαύλους· όμως το πλήθος κόμβων/ακμών εκτοξεύεται και η αναζήτηση βαραίνει. Μεγαλύτερο βήμα (π.χ. "Grid Step"=3/4) κάνει τον γράφο πολύ ελαφρύ, αλλά μπορεί να μην αναγνωρίσει στενά ανοίγματα μεταξύ νησίδων ή να εξαναγκάσει υπερβολικά πλατιές καμπύλες γύρω από ακτές, αυξάνοντας την τελική απόσταση. Σημαντικό είναι ότι το "Grid Step" δεν μπορεί να δώσει λεπτομέρεια πιο ακριβή από την ανάλυση της sea mask (0.01°), καθώς αν η μάσκα δεν διακρίνει ένα στενό όρμο, η μείωση του "Grid Step" δεν θα τον ανασύρει. Το

ταίριασμα της αραιοποίησης (με την τιμή "Grid Step"=2) με την Resolution=0.01° της μάσκας, αποδίδουν πρακτικά δίνοντας αρκετή ευκαμψία κοντά στην ακτή, χωρίς να μεγαλώνει υπερβολικά το μέγεθος του γράφου.

Η περιοχή που έχει επιλεγεί καθορίζει το πεδίο μέσα στο οποίο χτίζεται ο γράφος. Μία πολύ περιορισμένη περιοχή γύρω από τους δύο λιμένες μπορεί να αποκλείσει χρήσιμες εναλλακτικές που περνούν λίγο εκτός των στενών ορίων και να οδηγήσει σε διαδρομές με μεγαλύτερη απόσταση ή ακόμη και σε αποτυχία εύρεσης πορείας. Αντιστρόφως, μία υπερβολικά μεγάλη περιοχή του γράφου αυξάνει τον αριθμό κόμβων/ακμών, επιβαρύνει τη μνήμη και μεγαλώνει τον χώρο που πρέπει να εξερευνήσει ο αλγόριθμος A*, χωρίς απαραίτητα να προσφέρει ουσιαστικά καλύτερη λύση. Η πρακτική στρατηγική είναι να ορίζεται η περιοχή από τα άκρα (λιμάνι αφετηρίας και άφιξης) και να προστίθεται ένα λογικό περιθώριο γύρω τους ("buffer"), αρκετό ώστε να χωράνε παρακάμψεις σε ανοιχτή θάλασσα αλλά όχι τόσο ώστε να περιλαμβάνονται άσχετες περιοχές.

Οι τρεις παράμετροι αλληλεπιδρούν, έτσι ώστε μία μικρή περιοχή με μεγάλο "μέγιστο μήκος ακμής" μπορεί να κρύψει την έλλειψη χωρικής ελευθερίας, αλλά αυξάνει τον κίνδυνο να προταθούν μεγάλες ευθείες που τελικά απορρίπτονται από τον έλεγχο LOS, επιμηκύνοντας την αναζήτηση. Ένα πυκνό δίκτυο κόμβων (μικρό "Grid Step") με μικρό όριο ακμής προσφέρει εξαιρετική προσαρμογή στην ακτογραμμή, όμως δημιουργεί γράφους με πολλές χιλιάδες ακμές και κάνει το A* να δουλεύει περισσότερο. Παρ' όλα αυτά, η ευρετική συνάρτηση haversine (μέτρηση κόστους) κρατά την επέκταση υπό έλεγχο. Συνολικά, ο εκτιμώμενος χρόνος επηρεάζεται έμμεσα, αφού συνήθως, πιο πυκνά δίκτυα και μικρότερα άλματα ακμών δίνουν λίγο συντομότερες ή ισοδύναμες αποστάσεις (άρα και εκτιμώμενο χρόνο) επειδή βρίσκουν πορείες αυστηρά γύρω από ακτές. Το κέρδος είναι μικρό σε ανοιχτή θάλασσα αλλά ουσιώδες σε περίπλοκες ακτογραμμές.

5 Συμπεράσματα & Μελλοντικές Επεκτάσεις

5.1 Συμπεράσματα

Η εργασία έδειξε με σαφήνεια ότι η ταχύτητα ως προς το έδαφος (SOG) δεν είναι μια σταθερά, αλλά το αποτέλεσμα μιας αλληλεπίδρασης ανάμεσα σε τρεις άξονες, οι οποίοι είναι τα στοιχεία καιρού, ο τύπος και η θέση του πλοίου. Στο επίπεδο του μοντέλου πρόβλεψης της ταχύτητας, αυτό αποτυπώθηκε σε μεγάλο ποσοστό ακρίβειας του μοντέλου (85%) με ελάχιστο σφάλμα (περίπου 1 κόμβο), τιμές που επιβεβαιώνουν ότι ο συνδυασμός των επιλεγμένων γνωρισμάτων περιγράφει επαρκώς την πραγματική συμπεριφορά πλου. Στο επίπεδο της εφαρμογής (δρομολόγηση), οι ίδιες προβλέψεις ταχύτητας, όταν συντεθούν κατά μήκος μιας διαδρομής, μεταφράζονται σε σταθερές και χρηστικές εκτιμήσεις χρόνου άφιξης.

Πρώτον, ως προς τον καιρό, οι μεταβλητές του ύψους κύματος και ταχύτητας ανέμου, και ιδίως ο συνδυαστικός δείκτης του γινόμενου αυτού, αποδείχθηκαν καίριες στην πρόβλεψη της ταχύτητας. Όταν το κύμα αυξάνεται ή/και ο άνεμος εντείνεται, η πρόσθετη υδροδυναμική και αεροδυναμική αντίσταση μειώνει την επιτεύξιμη ταχύτητα. Αυτή η σχέση εμφανίζεται με συνέπεια στα δεδομένα. Έτσι, η ίδια γεωμετρία διαδρομής με διαφορετικό καιρό οδηγεί σε διαφορετική ταχύτητα ανά τμήμα και, τελικά, σε διαφορετικό χρόνο άφιξης.

Δεύτερον, ο τύπος του πλοίου λειτουργεί ως γραμμή βάσης. Διαφορετικές κατηγορίες πλοίων έχουν διαφορετικά προφίλ ταχυτήτων υπηρεσίας και διαφορετική ευαισθησία στον καιρό. Η υψηλή σημαντικότητα του γνωρίσματος στο μοντέλο Random Forest υποστηρίζει ακριβώς αυτό, ότι το μοντέλο διαχωρίζει πρώτα σε ποια ζώνη ταχύτητας ανήκει ένα πλοίο και στη συνέχεια προσαρμόζει την πρόβλεψη με βάση τη θέση και τις τοπικές καιρικές συνθήκες. Το αποτέλεσμα φαίνεται και στις μετρικές ανά κλάση. Παρατηρήθηκε ότι οι πολυπληθείς κατηγορίες (π.χ. 30, 70, 80, 90, 1001, 1024, 1025) παρουσιάζουν χαμηλά MAE και υψηλό R^2 , άρα προσφέρουν σταθερότητα στον υπολογιζόμενο τελικό χρόνο ταξιδιού, γιατί αυτές είναι που κυριαρχούν στον πραγματικό φόρτο των διαδρομών.

Τρίτον, η γεωγραφία (LAT/LON) δεν καθορίζει μόνη της την ταχύτητα, όμως παρέχει κρίσιμες πληροφορίες. Σε παράκτιες ζώνες και σε διαύλους οι ταχύτητες είναι συχνά πιο συντηρητικές, ενώ στο ανοιχτό πέλαγος, όταν ο καιρός το επιτρέπει, αυξάνονται. Το μοντέλο αξιοποιεί αυτή την πληροφορία για να ρυθμίσει τον αντίκτυπο του καιρού στον χώρο, γεγονός που καθιστά τις προβλέψεις πιο ρεαλιστικές κατά μήκος σύνθετων ακτογραμμών.

Σε επίπεδο δρομολόγησης, η αλυσίδα που από τα δεδομένα AIS και καιρού έγινε η πρόβλεψη της ταχύτητας και κατ' επέκταση η δρομολόγηση, λειτούργησε ακριβώς όπως σχεδιάστηκε. Η sea mask (με κανόνα ότι σημεία πάνω στο περίγραμμα θεωρούνται ξηρά) και ο έλεγχος LOS εξασφαλίζουν ότι η γεωμετρική λύση είναι ναυτιλιακά έγκυρη, αφού κανένα τμήμα δεν τέμνει την ξηρά. Ο γράφος με τους κόμβους και τις ακμές, επιτρέπει στον αλγόριθμο A* με χρήση της ευρετικής συνάρτησης haversine να βρίσκει τη βέλτιστη πορεία. Πάνω σε αυτή τη διαδρομή, η ταχύτητα προβλέπεται ανά τμήμα από το μοντέλο, χρησιμοποιώντας τις τοπικές καιρικές συνθήκες της ίδιας ημέρας στο ενδιάμεσο σημείο του σκέλους και τα ίδια γνωρίσματα με αυτά της εκπαίδευσης. Η μετάβαση από απόσταση σε χρόνο γίνεται με πολύ απλό τρόπο ($time = distance / SOG$), οπότε η τελική χρονική διάρκεια είναι το άθροισμα καθαρών, αιτιολογημένων συνιστωσών.

Το σημαντικότερο συμπέρασμα είναι ότι η αξιοπιστία των ETA δεν προκύπτει από μια “μέση σταθερή ταχύτητα”, αλλά από πολλές μικρές, τοπικά εύστοχες προβλέψεις ταχυτήτων. Έτσι, ακόμη κι αν σε επιμέρους τμήματα υπάρχει σφάλμα της τάξης του κόμβου, το συνολικό αποτέλεσμα παραμένει σταθερό, καθώς τα σφάλματα δεν αθροίζονται γραμμικά και συχνά εξισορροπούνται καθώς η διαδρομή περνά από ζώνες με διαφορετικές συνθήκες. Η ισχυρή συνολική επίδοση ($MAE \approx 1.14$, $R^2 \approx 0.85$), σε συνδυασμό με την καλή συμπεριφορά στις πολυπληθείς κλάσεις, καθιστούν τις εκτιμήσεις χρηστικές για επιχειρησιακό προγραμματισμό και επιτρέπουν τη σύγκριση των εναλλακτικών πορειών, την ανάδειξη ευαίσθητων τμημάτων και τεκμηριωμένες αποφάσεις με θετικό αποτύπωμα στην ενεργειακή αποδοτικότητα.

Εν κατακλείδι, έγινε κατανοητό ότι η ενσωμάτωση του πραγματικού καιρού στην εκτίμηση της ταχύτητας, με σαφή διάκριση του τύπου πλοίου και της γεωγραφικής

περιοχής, οδηγεί σε ερμηνεύσιμες και επαναλήψιμες προβλέψεις του χρόνου άφιξης. Αυτό ακριβώς κάνει τη μελέτη κατάλληλη για υιοθέτηση και κλιμάκωση σε πραγματικές ροές ναυτιλιακού σχεδιασμού.

5.2 Περιορισμοί - Προσέγγιση χωρίς χρονική μνήμη και έλλειψη ρευμάτων

Παρότι το σύστημα αποδίδει αξιόπιστες εκτιμήσεις ταχύτητας και χρόνου άφιξης, υπάρχουν δύο ουσιαστικοί περιορισμοί που επηρεάζουν την ακρίβεια με την οποία μπορεί να αντιδρά σε κάθε σενάριο.

1) Πρόβλεψη χωρίς χρονική μνήμη.

Ο αλγόριθμος Random Forest εκπαιδεύεται και προβλέπει ανά γραμμή (στιγμιότυπο), χωρίς να συνδυάζει τι προηγήθηκε και χωρίς να εξετάζει το δρομολόγιο συνολικά. Αυτό είναι συνειδητή σχεδιαστική επιλογή (απλότητα, ταχύτητα, ευθυγράμμιση με τη δρομολόγηση ανά σκέλος), αλλά συνεπάγεται ορισμένες απώλειες πληροφορίας. Για παράδειγμα, δεν μοντελοποιούνται δυναμικές αδράνεις (όπως ο χρόνος για επιτάχυνση ή επιβράδυνση), λειτουργικές πολιτικές (π.χ. πλους σε traffic separation), ούτε βραχύχρονες αλληλεπιδράσεις με τοπική κυκλοφορία ή χειρισμούς. Επίσης, τα καιρικά χαρακτηριστικά είναι εντάσεις (μέτρα) και όχι διευθύνσεις, άρα θα ήταν πιο κατατοπιστικό χαρακτηριστικό αν ήταν γνωστός ο σχετικός προσανατολισμός του άνεμου και του κύματος ως προς το διάνυσμα της πορείας. Παρότι μέρος αυτών των επιδράσεων ενσωματώνεται στην ταχύτητα, το μοντέλο δεν διαθέτει ρητές μεταβλητές για να διακρίνει τη σχέση αυτή. Κατά συνέπεια, η πρόβλεψη είναι στατιστικά ισχυρή σε μέσο όρο, αλλά λιγότερο ακριβής σε εφήμερα φαινόμενα ή σε καθεστώτα όπου η συμπεριφορά του πλοίου εξαρτάται από ακολουθίες αποφάσεων.

2) Έλλειψη ρευμάτων (currents) στα χαρακτηριστικά.

Το μοντέλο δεν δέχεται ρητά ωκεάνια και παλιρροϊκά ρεύματα. Η ταχύτητα SOG είναι, εξ ορισμού, ταχύτητα ως προς το έδαφος, άρα τα ρεύματα επηρεάζουν άμεσα (το tail current αυξάνει την ταχύτητα, ενώ το head current τη μειώνει). Εφόσον όμως δεν

χρησιμοποιείται κάποιο σχετικό χαρακτηριστικό, ο αλγόριθμος δεν μπορεί να προσαρμόσει την πρόβλεψη σε τοπικές και χρονικές μεταβολές ρευμάτων που δεν συσχετίζονται με τον άνεμο και το κύμα. Σε περιοχές όπως το δυτικό ρεύμα του Βόρειου Ατλαντικού, οι προβλέψεις τείνουν προς μία μέση επίδραση ρευμάτων, όπως τη γνώρισε το μοντέλο στο στάδιο της εκπαίδευσης, με πιθανές αποκλίσεις όταν το πραγματικό ρεύμα είναι ασυνήθιστα ισχυρό ή μετατοπισμένο. Για την μελέτη του χρόνου άφιξης, το σφάλμα εμφανίζεται συσσωρευτικά σε τμήματα που διατρέχουν ισχυρά ρεύματα με σταθερό πρόσημο.

Συνολικά, προσέγγιση χωρίς χρονική μνήμη του αλγορίθμου Random Forest Regressor μειώνει την περιγραφικότητα στην επίδραση του καιρού και η απουσία ρευμάτων στερεί ένα βασικό φυσικό παράγοντα της τηρούμενης ταχύτητας. Παρά ταύτα, στο τυπικό εύρος χρήσης οι υπολογισμένες επιδόσεις ($MAE=1.14, R^2=0.85$) παραμένουν σταθερά χρήσιμες για την εκτίμηση του χρόνου του ταξιδιού. Η σωστή ανάγνωση είναι ότι οι εκτιμώμενοι χρόνοι άφιξης είναι αξιόπιστοι σε γενικές γραμμές και ιδιαίτερα επιτυχημένες στις συνηθέστερες διαδρομές, με περιθώριο βελτίωσης όταν οι τρεις παραπάνω διαστάσεις γίνουν πιο πλούσιες σε δεδομένα ή χαρακτηριστικά.

5.3 Μελλοντικές διορθώσεις

Ένα πρώτο, άμεσα εφαρμόσιμο βήμα είναι η βελτίωση της χρονικής ευθυγράμμισης δεδομένων AIS και καιρού. Πέρα από την αυστηρή σύνδεση στο ίδιο 15λεπτο, μπορούν να εξεταστεί η εναλλακτική του "nearest-time join" με ανοχή $\pm 15'$. Ουσιαστικά, θα πρέπει για κάθε στιγμιότυπο καταγραφής δεδομένων AIS στιγμιότυπου να επιλέγεται η πλησιέστερη μετεωρολογική μέτρηση εντός παραθύρου (για παράδειγμα 15 λεπτών), κάτι το οποίο εκτιμάται ότι θα αυξήσει την κάλυψη των καιρικών δεδομένων, τα οποία είναι ελλιπή.

Δεύτερη προτεραιότητα είναι η προσθήκη των ρευμάτων. Πρακτικά, θα πρέπει να ενσωματωθεί το διανυσματικό ρεύμα (u,v) και η προβολή του διανύσματος στην κατεύθυνση κάθε σκέλους της διαδρομής (along-track component). Αυτό μπορεί να μπει ως χαρακτηριστικό στο μοντέλο πρόβλεψης της ταχύτητας, μία προσέγγιση που θα ήταν

ιδιαίτερα χρήσιμη στην σχεδίαση του δρομολογίου, γιατί συνδέει απευθείας την κατεύθυνση της πλεύσης με το τοπικό ρεύμα.

Τρίτον, πολύ σημαντική είναι η λογική της δρομολόγησης βάσει του χρόνου (time-dependent routing). Ουσιαστικά, αντί για στατικό βάρος ακμής, το βάρος να εξαρτάται από τοπική πρόβλεψη καιρού ανάλογα με την ώρα της διέλευσης. Αυτή η υλοποίηση απαιτεί έναν διευρυμένο γράφο (time-expanded graph ή με A^* που ενημερώνει δυναμικά το κόστος ακμής βάσει ανέμου, κύματος και ρευμάτων) στον εκτιμώμενο χρόνο άφιξης σε κάθε κόμβο. Αυτό επιτρέπει την επιλογή της ευνοϊκής περιόδου αναφορικά με τον καιρού και τα ρεύματα στην περιοχή. Για παράδειγμα, αν είχε τόσο χρήσιμη πληροφορία το πλοίο βάσει πρόβλεψης καιρού, θα μπορούσε να αποπλεύσει νωρίτερα ή να καθυστερήσει ή να αλλάξει τη διαδρομή του για να έχει ευνοϊκότερο ρεύμα στην υπόλοιπη διαδρομή. Αυτή η επιλογή θα πολλαπλασίαζε την χρησιμότητα της εργασίας, αλλά θα απαιτούσε πολλαπλάσιους υπολογισμούς και επεξεργαστική ισχύ, για αυτό δεν επελέγη.

Τέλος, η μελέτη θα μπορούσε να γίνει ακόμα πιο ακριβής για τις ανάγκες μίας εταιρία αν κατά τη δρομολόγηση υπήρχαν διαθέσιμα δεδομένα κατανάλωσης καυσίμου. Σε αυτή την περίπτωση, η μελέτη θα μπορούσε να αποσκοπεί όχι μόνο στην ελαχιστοποίηση του χρόνου αλλά και στην μείωση του ενεργειακού κόστους ή ακόμα και στην αύξηση της ασφάλειας (αποκλείοντας επικίνδυνες περιοχές). Έτσι, ο χρήστης θα λάμβανε όχι μόνο την πιο γρήγορη διαδρομή, αλλά και την πιο οικονομική ή την πιο ασφαλή, ανάλογα με την αποστολή.

Βιβλιογραφία

Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). *Classification and regression trees*. Wadsworth.

Breiman, L. (2001). *Random forests (Technical Report No. 567)*. University of California, Berkeley, Department of Statistics.

<https://www.stat.berkeley.edu/~breiman/randomforest2001.pdf>

Bureau of Meteorology. (n.d.). *Marine weather services*.

<https://www.bom.gov.au/marine/>

Chirosca, A.-M., Medina, A., Pacuraru, F., Saettone, S., Rusu, L., & Pacuraru, S. (2023). Experimental and numerical investigation of the added resistance in regular head waves for the DTC hull. *Journal of Marine Science and Engineering*, 11(4), 852.

<https://doi.org/10.3390/jmse11040852>

Federal Communications Commission. (2022). Maritime mobile service identities (MMSI).

<https://www.fcc.gov/maritime-mobile-service-identities-mmsi>

Hart, P. E., Nilsson, N. J., & Raphael, B. (1968). A formal basis for the heuristic determination of minimum cost paths. *IEEE Transactions on Systems Science and Cybernetics*, 4(2), 100–107.

<https://doi.org/10.1109/TSSC.1968.300136>

Integrated Ocean Observing System. (n.d.). *IOOS data access*.

<https://ioos.noaa.gov/data/>

International Association of Marine Aids to Navigation and Lighthouse Authorities. (2024). *G1177 – Portrayal of VTS information (Edition 2.0)*.

<https://www.iala.int/product/g1177-portrayal-of-vts-information/>

International Association of Marine Aids to Navigation and Lighthouse Authorities. (n.d.). *AIS guidelines*.

<https://www.iala.int/product-category/publications/guidelines/>

International Maritime Organization. (1983). *Resolution A.528(13): Recommendation on weather routing*.

<https://www.imo.org/en/OurWork/Safety/Pages/Weather-Routeing.aspx>

International Maritime Organization. (n.d.). *Automatic Identification System (AIS)*.

<https://www.imo.org/en/OurWork/Safety/Pages/AIS.aspx>

Martić, I., Degiuli, N., Majetić, D., & Farkas, A. (2020). *Evaluation of the effect of container ship characteristics*. *Journal of Marine Science and Engineering*, 8(9), 696.

<https://doi.org/10.3390/jmse8090696>

Mittendorf, M. (2023). *Data-driven prediction of added resistance on ships in waves (Doctoral dissertation, Technical University of Denmark)*.

<https://orbit.dtu.dk/en/publications/data-driven-prediction-of-added-resistance-on-ships-in-waves>

National Data Buoy Center. (n.d.). *National Data Buoy Center*.

<https://www.ndbc.noaa.gov/>

Natural Earth. (n.d.). *Natural Earth data*.

<https://www.naturalearthdata.com/>

Shapely Development Team. (n.d.). *Shapely documentation*.

<https://shapely.readthedocs.io>

United States Coast Guard Navigation Center. (n.d.). *Automatic Identification System (AIS)*.

<https://www.navcen.uscg.gov/ais>

Wang, W., Wu, T., Zhao, D., Guo, C., Luo, W., & Pang, Y. (2019). Experimental–numerical analysis of added resistance to container ships under presence of wind–wave loads. *PLOS ONE*, 14(8), e0221453. <https://doi.org/10.1371/journal.pone.0221453>