Βέλτιστη Πορεία Πλοίου και Μηχανική Μάθηση

Εθνικό Μετσόβιο Πολυτεχνείο

Σχολή Εφαρμοσμένων Μαθηματικών και Φυσικών Επιστημών

Γεωργίτσης Γεώργιος

Επιβλέπων Καθηγητής: Στεφανέας Πέτρος

Αθήνα, Σεπτέμβριος 2021

ΤΡΙΜΕΛΗΣ ΕΠΙΤΡΟΠΗ

- Π. Στεφανέας, Επικ. Καθηγητής ΣΕΜΦΕ
- Ι. Κολέτσος, Αναπλ. Καθηγητής ΣΕΜΦΕ
- Π. Ψαρράκος, Καθηγητής ΣΕΜΦΕ

Περιεχόμενα

Ευχαριστίες

Πρώτον απ’ όλους, θα ήθελα να ευχαριστήσω τον επιβλέποντα καθηγητή της διπλωματικής μου εργασίας, κο Στεφανέα Πέτρο, Επικ. Καθηγητή του Ε.Μ.Π., για την άμεση ανταπόκριση και την βοήθεια που παρείχε καθ’ όλη την διάρκεια εκπόνησης της εργασίας, καθώς και για την κατανόηση και συνεννόηση που διείπε την συνεργασία μας.

Στην συνέχεια, θα ήθελα να εκφράσω τις πιο θερμές μου ευχαριστίες στον πατέρα μου, Μιχάλη, την μητέρα μου, Ελένη και την αδερφή μού, Αιμιλία, για όλα όσα μου παρείχαν και συνεχίζουν να μου προσφέρουν, συμβάλλοντας καθημερινά και ο καθένας με τον δικό του τρόπο, στην επίτευξη των στόχων μου.

Ευχαριστώ τον οικογενειακό μας φίλο Κοντούλη Γιώργο για την πολύτιμη βοήθεια του, καθώς συνέβαλε στην καλύτερη και βαθύτερη κατανόηση του κόσμου της ναυτιλίας.

Τέλος, ευχαριστώ τους φίλους και συμφοιτητές μου οι οποίοι ήταν δίπλα μου σε όλα τα στάδια της φοιτητικής μου ζωής.

Λίστα πινάκων

Λίστα εικόνων και γραφημάτων

Περίληψη

Οι θαλάσσιες μεταφορές είναι μακράν ο πιο χρησιμοποιούμενος τρόπος μεταφοράς εμπορευμάτων παγκοσμίως καθώς εκτιμάται ότι περισσότερο από το 90% των παγκόσμιων εμπορευμάτων μεταφέρονται μέσω θαλάσσης. Αυτού του είδους η ανάγκη για τη μεταφορά εμπορευμάτων έχει αυξήσει την κίνηση στις θαλάσσιες οδούς. Το αποτέλεσμα αυτής της αύξησης σε συνδυασμό με την συνεχή ανάγκη για την μείωση του κόστους μεταφοράς των προϊόντων έχει οδηγήσει στην συνεχή αναζήτηση μεθόδων που θα μπορούν να βελτιστοποιήσουν την κίνηση σε μια θαλάσσια εμπορική οδό. Τα τελευταία χρόνια έχει γίνει μια αξιοσημείωτη ανάπτυξη εφαρμογών τεχνητής νοημοσύνης σχεδόν σε όλους τους τομείς των ανθρώπινων δραστηριοτήτων και φυσικά και στην ναυσιπλοΐα. Με σκοπό την αξιολόγηση αυτών των μεθόδων πραγματοποιήθηκε εξέταση της ακρίβειας μιας σειράς μοντέρνων μεθόδων πρόβλεψης της κίνησης ενός πλοίου για τον καθορισμό της βέλτιστης πορείας του. Η εξέταση αυτή έγινε σε διάφορους τύπους πλοίων και οι μέθοδοι που εφαρμοστήκαν έδειξαν αναμενόμενα αποτελέσματα. Πιο συγκεκριμένα διαπιστώθηκε ότι πλοία προκαθορισμένης τροχιάς ή που διανύουν μικρότερες αποστάσεις αναμένεται να έχουν μεγαλύτερη ακρίβεια πρόβλεψης της πορείας σε σύγκριση με πλοία που πορείας τους εξαρτάται από περισσότερους όπως π.χ. ποντοπόρα πλοία. Επιπλέον διαπιστώθηκε ότι τα μέτρα ακρίβειας του απόλυτου τετραγωνικού σφάλματος και της ακρίβειας παραμένουν σταθερά ανεξάρτητα από τον τύπο του εξεταζόμενου , γεγονός που τα καθιστά ως πιο ασφαλή μέσα για την εκτίμηση της πορείας ενός πλοίου.

Λέξεις Κλειδιά: Δεδομένα AIS, Πορεία πλοίου, Πρόβλεψη

Abstract

Maritime transport is by far the most widely used mode of transport worldwide as it is estimated that more than 90% of global freight is transported by sea. This kind of need for freight has increased the traffic on the sea routes. The result of this increase combined with the continuing need to reduce the cost of transporting products has led to the constant search for methods that can optimize traffic on a maritime trade route. In recent years there has been a remarkable development of artificial intelligence applications in almost all areas of human activity and of course in navigation. In order to evaluate these methods, the accuracy of a series of modern methods of predicting the movement of a ship was determined to determine its optimal course. This test was performed on different types of ships and the methods applied showed expected results. More specifically, it was found that ships with a predetermined orbit or that travel shorter distances are expected to have greater accuracy in predicting the course compared to ships whose course depends on more than e.g. ocean-going ships. In addition, it was found that the measures of accuracy of absolute square error and accuracy remain constant regardless of the type of examinee, which makes them a safer means of estimating the course of a ship.

Keywords: AIS data, Ship trajectory, Prediction

Εισαγωγή

Ο σχεδιασμός διαδρομών στις θαλάσσιες μεταφορές ενέπνευσε τους επιστήμονες ήδη από την εποχή που τα ιστιοφόρα σκάφη ήταν ακόμα το μόνο μέσο μεταφοράς που επέτρεπε το ταξίδι στον ωκεανό. Οι πρώτες προσπάθειες για δημιουργία διαδρομών των ιστιοφόρων πλοίων είχαν ως βάση τα στατιστικά δεδομένα τον άνεμο και προηγούμενες διαδρομές έγιναν από τον Maury τον 19ο αιώνα. Τον 20ό αιώνα, τα πλοία που κινούνται με κινητήρες έγιναν το κύριο μέσο θαλάσσιας μεταφοράς. Παρόλη την αύξηση της δυναμικότητας των μέσων (πλοίων) που χρησιμοποιούταν στις θαλάσσιες εμπορικές μεταφορές ακόμη παραμένουν προβλήματα στην μετακίνηση τους, κάποια από τα οποία δεν παρουσιάζουν διαφορές από αυτά του προηγούμενου αιώνα όπως π.χ. η ταχύτητα και η κατεύθυνση του ανέμου. Πέραν όμως από τα προβλήματα που δημιουργούνται από τα στοιχεία της φύσης υπάρχουν και άλλα προβλήματα που μπορεί να εμφανιστούν χωρίς προειδοποίηση όπως π.χ. απότομη απόκλιση από την πορεία ενός πλοίου λόγου πειρατείας ή παροχή βοήθειας προς κάποιο άλλο πλοίο. Για τον λόγο αυτό η μοντέρνα ναυσιπλοΐα χρησιμοποιεί μοντέρνες μεθόδους πρόβλεψης της πορείας ενός πλοίου βάση παραμέτρων και προσπαθεί να υπολογίσει την βέλτιστη πορείας. Αν και βασικό κριτήριο για τον ορισμό της βέλτιστης πορείας είναι η οικονομία στο ταξίδι πάντοτε υπάρχει περίπτωση και άλλων παραμέτρων όπως η ταχύτητα του ταξιδιού ιδιαίτερα στην περίπτωση που πρέπει να καλυφθούν χρονικές απαιτήσεις λόγω συμβολαίου. Σημαντικό βοήθημα σε αυτή την προσπάθεια αποτελεί η χρήση δεδομένων AIS (Automated Information System) με την βοήθεια του οποίου οι εφαρμογές των μεθόδων πρόβλεψης της πορείας ενός πλοίου και του καθορισμού (ή επανακαθορισμού) της πορείας μπορεί να γίνει σε πραγματικό χρόνο. Σκοπός της παρούσας εργασίας είναι η αξιολόγηση της ακρίβειας ενδεικτικών μεθόδων πρόβλεψής της πορείας ενός πλοίου και η σύγκριση των αποτελεσμάτων αυτών με βάση την εφαρμοζόμενη μέθοδο και τον τύπο του πλοίου. Η εργασία δομείται σε 5 κεφάλαια ως εξής, στο πρώτο κεφάλαιο γίνεται περιγραφή του δικτύου των θαλάσσιων μεταφορών. Στο δεύτερο κεφάλαιο γίνεται η περιγραφή των μεθόδων που θα χρησιμοποιηθούν για την αξιολόγηση της ακρίβειας τους. Στο τρίτο κεφάλαιο γίνεται η βιβλιογραφική επισκόπηση προηγουμένων παρόμοιων μελετών. Στο τέταρτο κεφάλαιο παρουσιάζονται τα αριθμητικά αποτελέσματα εξέτασης της ακρίβειας των μεθόδων και στο πέμπτο κεφάλαιο παρουσιάζονται τα συμπεράσματα της έρευνας.

Κεφάλαιο 1

Δίκτυα θαλάσσιων μεταφορών

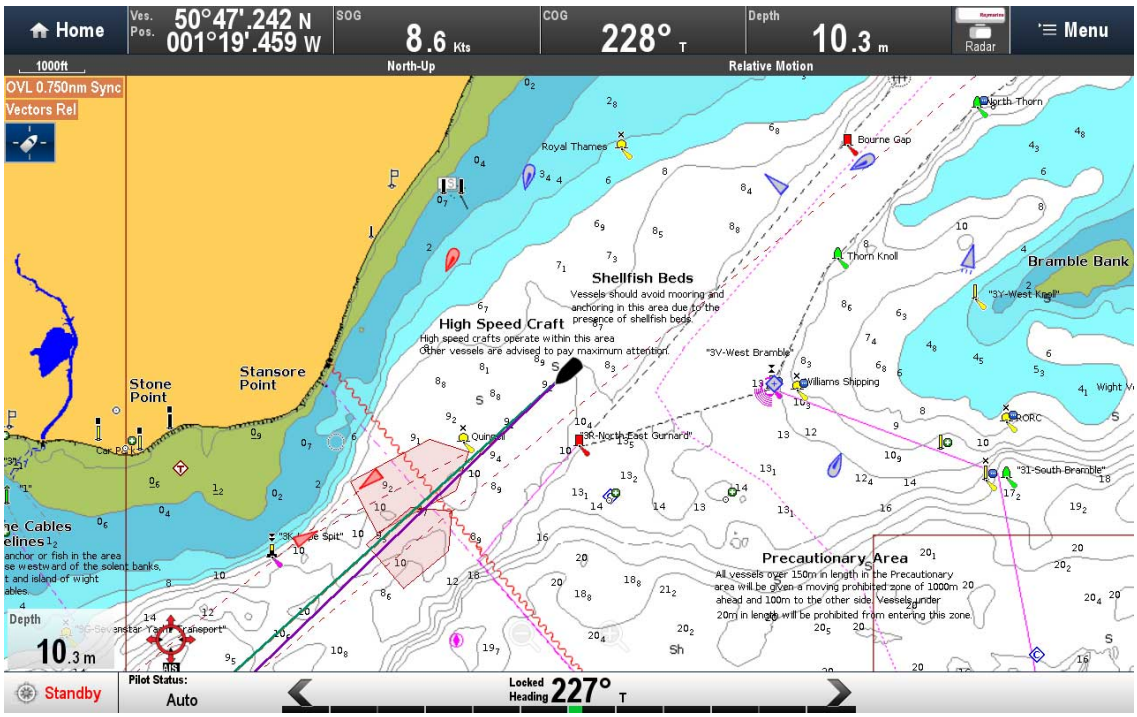
1.1. Η μεταφορά αγαθών μέσω της θαλάσσιας οδού

Ο Παγκόσμιος Οργανισμός Ναυτιλίας ή ΠΟΝ (International Maritime Organization – IMO) εκτιμά ότι πάνω από το 90 % του παγκόσμιου εμπορίου πραγματοποιείται μέσω θαλάσσης, καθώς η ναυτιλία εξακολουθεί να είναι η πιο οικονομικά αποδοτική μέθοδος για τη μεταφορά αγαθών και πρώτων υλών παγκοσμίως (Tu, et al.,2016). Κατά συνέπεια, η προστασία και η ασφάλεια των διεθνών θαλάσσιων γραμμών επικοινωνίας έχουν συνεχώς αυξανόμενη σημασία τόσο για τους εμπόρους όσο και για τους καταναλωτές. Η αυξανόμενη ζήτηση για αγαθά και υλικά σε όλο τον κόσμο αυξάνει και το εμπόριο μέσω θαλάσσης, γεγονός που αυξάνει την πιθανότητα συγκρούσεων σε περιοχές με συμφόρηση. Παρόμοια, περιοχές με θαλάσσια συμφόρηση είναι επικίνδυνες για την εμφάνισης πειρατείας ή τρομοκρατικών. Οι Harati-Mokhtari et al., (2007) εκτιμούν ότι το ανθρώπινο λάθος αντιπροσωπεύει το 80 έως 85 τοις εκατό των καταγεγραμμένων θαλάσσιων ατυχημάτων.

Επιπλέον, οι ασύμμετρες απειλές όπως αυτές που επιτέθηκαν στο USS Cole το 2000 στην Υεμένη είναι πάντοτε παρούσες και ανησυχητικές. Μια ακριβής πρόβλεψη σημείου για τη μελλοντική τοποθεσία ενός σκάφους μπορεί να είναι χρήσιμη τόσο για την παρακολούθηση της κυκλοφορίας του στόλου μιας επιχείρησης όσο και για τον εντοπισμό ανωμαλιών (παρεκκλίσεων) που θα μπορούσαν να αποτελούν πιθανή απειλή για την ασφάλεια του σκάφους. Λόγω των πιθανών σφαλμάτων σε μια πρόβλεψη, θα πρέπει οι προβλέψεις της τοποθεσίας ενός σκάφους να συνοδεύονται από περιοχές αβεβαιότητας π.χ. διαστήματα εμπιστοσύνης των προβλέψεων, που περιέχουν την πραγματική μελλοντική τοποθεσία ενός σκάφους σε ένα προκαθορισμένο επίπεδο ανοχής π.χ. p-level=0.05.

Καθώς η συμφόρηση στις πλωτές οδούς αυξάνεται συνεχώς, η συνεχής ενημέρωση για το θαλάσσιο τομέα (Maritime Domain Awareness MDA) γίνεται όλο και πιο σημαντική τόσο για εμπορικές επιχειρήσεις όσο και για κρατικούς οργανισμούς όπως π.χ. το Πολεμικό Ναυτικό των ΗΠΑ ([DON, 2007). Ένα βασικό εργαλείο για τη διατήρηση της συνεχούς ενημέρωσης για το θαλάσσιο τομέα είναι το Αυτόματο

Πληροφοριακό Σύστημα (AIS), ένα δίκτυο πομποδεκτών που παρέχει πληροφορίες σχετικά με την παγκόσμια κίνηση των πλοίων στη θάλασσα. Από το 2002, ο ΠΟΝ απαίτησε να εγκατασταθούν πομποδέκτες AIS σε πλοία άνω των 300 κόρων και σε όλα τα επιβατηγά πλοία, για να αυξηθεί η ασφάλεια της ζωής στη θάλασσα. Επειδή το AIS επιτρέπει σε όλους τους αερομεταφορείς να δουν τη θέση, την κατεύθυνση και την ταχύτητα άλλων πλοίων στη γύρω περιοχή, οι συγκρούσεις μπορούν να αποφευχθούν, αποτρέποντας έτσι τόσο τη νομισματική απώλεια όσο και την απώλεια ζωών. Άλλα οφέλη του AIS περιλαμβάνουν την παρακολούθηση της κυκλοφορίας, τις επιχειρήσεις έρευνας και διάσωσης, έρευνες ατυχημάτων, βοήθεια πλοήγησης και παρακολούθηση πλοίων (Balduzzi, Pasta, & Wilhoit, 2014). Ένα παράδειγμα για το πώς μπορεί να εμφανιστεί μια οθόνη AIS σε ένα σκάφος φαίνεται στο σχήμα 1.



Σχήμα 1.1. *Συνήθης απεικόνιση ενός συστήματος AIS (Πηγή: RayMarine, 2016)*

Ένας χειριστής AIS είναι σε θέση να λάβει χρήσιμες πληροφορίες για τα άλλα σκάφη της περιοχής επιλέγοντας ένα εικονίδιο σκάφους που απεικονίζονται ως τρίγωνα στο σχήμα 1. Πληροφορίες όπως ταχύτητα, κατεύθυνση, γεωγραφικό πλάτος και γεωγραφικό μήκος βοηθούν τον πιλότο στην πλοήγηση του σκάφους. Εκτός όμως από αυτά τα βασικά χαρακτηριστικά, άλλα πεδία ενημερώνονται από τον πιλότο όπως ο προορισμός, η χώρα προέλευσης και η τρέχουσα δραστηριότητα στο οποίο ασχολείται

το πλοίο. Ένα παράδειγμα δραστηριότητας που έχει αναληφθεί και καταχωρηθεί χειροκίνητα από τον χειριστή μπορεί να είναι το «Αλίευση » ή « Αβαρία ». Ενώ αυτά είναι χρήσιμα χαρακτηριστικά του AIS, οι πληροφορίες δεν είναι πάντα τέλειες. Τα δεδομένα εισόδου των χρηστών είναι συχνά αμφίβολα και μπορεί να μην είναι πολύ χρήσιμα για την πρόβλεψη της μελλοντικής θέσης ενός πλοίου που βρίσκεται σε κίνηση.

Εάν ένας αναλυτής διαθέτει πληροφορίες για άλλα σκάφη της περιοχής, που βοηθούν στην απόφαση αποφυγής σύγκρουσης, τότε πώς μπορεί να αναπαραστήσει καλύτερα αυτήν την απόφαση ως αλγόριθμο; Όπως ένας χειριστής σκαφών, ένας αλγόριθμος πρέπει να προβλέπει τη μελλοντική τοποθεσία ενός ή περισσότερων σκαφών για να αποτρέψει μια σύγκρουση. Ομοίως, όταν ένα σκάφος εξαφανιστεί, μια ομάδα έρευνας και διάσωσης πρέπει να αποφασίσει πού να αναζητήσει, πράγμα που περιλαμβάνει επίσης την πρόβλεψη της μελλοντικής θέσης του σκάφους. Οι σταθμοί παρακολούθησης σκαφών επωφελούνται από τέτοιου είδους αλγορίθμους καθώς οι πομποδέκτες AIS παράγουν εκπομπές μόνο κατά διαστήματα με βάση την ταχύτητα ενός σκάφους.

Οι οργανώσεις θαλάσσιας ασφάλειας μπορούν επίσης να επωφεληθούν από έναν αλγόριθμο που προβλέπει τη μελλοντική τοποθεσία του σκάφους και την αβεβαιότητα που σχετίζεται με αυτήν την πρόβλεψη για τον εντοπισμό ανώμαλης συμπεριφοράς των σκαφών. Εάν ένας αναλυτής μπορεί να υπολογίσει αυτόματα μια ακριβή πρόβλεψη σημείου για τη θέση του σκάφους και μια περιοχή πρόβλεψης γύρω από αυτήν τη θέση, μπορεί να δικαιολογήσει έρευνα εάν ένα σκάφος δεν περιέχεται σε αυτήν την περιοχή πρόβλεψης. Οι Pallotta, Vespe και Bryan (2013) περιγράφουν επιχειρήσεις καταπολέμησης της πειρατείας που εξαρτώνται από την ικανότητα πρόβλεψης του χώρου κυκλοφορίας ενός σκάφους σε συνδυασμό με την περιοχή που παρουσιάζονται πειρατικές δραστηριότητες. Επίσης, σημειώνουν ότι τα εμπορικά σκάφη συχνά απενεργοποιούν τους πομποδέκτες AIS όταν μεταβαίνουν σε περιοχές υψηλού κινδύνου για πειρατεία φοβούμενοι την αναγνώριση του σήματος (χακάρισμα) από τους πειρατές.

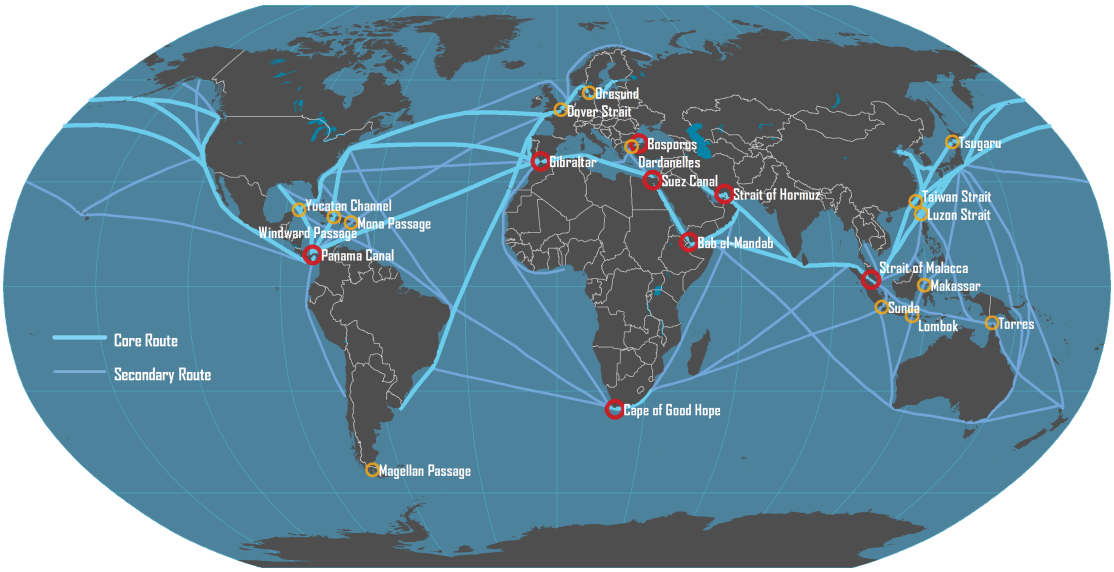
Η ανίχνευση ανωμαλιών είναι σημαντική για διάφορους λόγους. Πρώτον, είναι χρήσιμο για τον εντοπισμό πιθανών απειλών για την ασφάλεια κοντά σε πυκνοκατοικημένες παράκτιες πλωτές οδούς. Εάν η ανώμαλη συμπεριφορά εντοπιστεί αρκετά νωρίς, μπορεί να είναι δυνατή η έγκαιρη αντίδραση για την πρόληψη ή τον περιορισμό της βλάβης. Επιπλέον, η ανίχνευση ενός σκάφους που παρουσιάζει ανωμαλίες (αποκλίσεις) στην κίνηση τους μπορεί να βοηθήσει στον εντοπισμό πλοίων που έχουν χάσει τον έλεγχο ή αντιμετωπίζουν σοβαρά μηχανικά προβλήματα. Εάν ένα σκάφος εμφανίζει ανώμαλη συμπεριφορά, τότε κάνει κάτι που δεν ορίζεται από τα καθιερωμένα πρότυπα (πρωτοκόλλα) της ρότας του. Ενώ η ανώμαλη συμπεριφορά δεν συνεπάγεται κακόβουλη συμπεριφορά, η δυνατότητα αυτόματης ανίχνευσης ανώμαλης συμπεριφοράς θα μπορούσε να βοηθήσει τους αναλυτές ασφάλειας να αποφασίσουν πώς να διαθέσουν με τον καλύτερο τρόπο περιορισμένους πόρους για τη διερεύνηση πιθανών απειλών.

1.2. Τύποι θαλάσσιων διαδρομών

Η γεωγραφία πάνω στην οποία λειτουργούν βασίζονται οι θαλάσσιες μεταφορές είναι μοναδική, και περιλαμβάνει τον συνδυασμό φυσικών, στρατηγικών και εμπορικών παραμέτρων. Τα φυσικά ζητήματα είναι προφανώς σταθερά με την πάροδο του χρόνου, αλλά οι στρατηγικές, και ιδιαίτερα οι εμπορικές, εκτιμήσεις αλλάζουν συνεχώς με την εξέλιξη της παγκοσμιοποίησης. Η φυσιογραφία των θαλάσσιων μεταφορών αποτελείται από ωκεάνια και ποτάμια συστήματα κυκλοφορίας, τα οποία καθορίζονται από γεωφυσικά κριτήρια όπως το βάθος, τα ρεύματα, οι άνεμοι και η διαμόρφωση των ακτογραμμών και των διόδων. Παρόλο που οι ωκεανοί αντιπροσωπεύουν το 71% της επιφάνειας της γης, οι θαλάσσιες μεταφορές πραγματοποιούνται ως επί το πλείστων μόνο σε συγκεκριμένες διαδρομές που χρησιμοποιούνται τακτικά για ναυτιλιακά δρομολόγια. Αυτές οι διαδρομές αποτελούν συνάρτηση υποχρεωτικών σημείων διέλευσης, που είναι στρατηγικές τοποθεσίες-περάσματα, φυσικών περιορισμών (ακτές, άνεμοι, θαλάσσια ρεύματα, βάθος, ύφαλοι, πάγος) και πολιτικών συνόρων (Levinson, 2006).

Η διαμόρφωση του παγκόσμιου ναυτιλιακού δικτύου είναι σχετικά απλή και οργανώνεται κατά μήκος ενός ισημερινού διαδρόμου που συνδέει τη Βόρεια Αμερική, την Ευρώπη και την Ασία του Ειρηνικού μέσω της διώρυγας του Σουέζ, του Στενού

της Μαλάκας και της διώρυγας του Παναμά (Σχήμα 1.2). Περιορισμένη χρήση γίνεται στα βορειότερα μέρη του Ατλαντικού, καθώς και στα νοτιότερα τμήματα του Ατλαντικού, του Ινδικού και του Ειρηνικού, λόγω επικίνδυνων συνθηκών πλοήγησης (κυρίως πάγου) και της απόστασής τους από τα κέντρα της οικονομικής δραστηριότητας. Κατά τη θερινή περίοδο των μουσώνων (Απρίλιος έως Οκτώβριος), η ναυσιπλοΐα μπορεί να γίνει πιο επικίνδυνη στον Ινδικό Ωκεανό και τη Θάλασσα της Νότιας Κίνας. Η κλιματική αλλαγή μπορεί επίσης να παρέχει θαλάσσιες μεταφορές με μικρότερες διαδρομές ναυτιλίας μέσω της Αρκτικής, αλλά αυτές οι προοπτικές παραμένουν περιορισμένες σε αυτό το συγκεκριμένο σημείο (Levinson, 2006).



Σχήμα 1.2. *Θαλάσσιες διαδρομές μείζονος σημασίας και στρατηγικά περάσματα (Πηγή: Rodriguez, 2017)*

Για ευνόητους λόγους, οι θαλάσσιες διαδρομές προσπαθούν να ακολουθήσουν τη μεγάλη κυκλική διαδρομή, ένα μοτίβο που είναι εύκολα διακριτό στη διαμόρφωση των υπερατλαντικών και των διασυνοριακών διαδρομών. Οι βασικές διαδρομές είναι αυτές που χρησιμοποιούνται περισσότερο επειδή εξυπηρετούν μεγάλες αγορές, ενώ οι δευτερεύουσες διαδρομές είναι κυρίως συνδέσεις μεταξύ δευτερογενών και κύριων αγορών. Στις θαλάσσιες μεταφορές έχουν επικρατήσει οι διαμήκεις κινήσεις (ανατολής-δύσης). Το πλεονέκτημα των θαλάσσιων μεταφορών δεν είναι η ταχύτητα, αλλά η διατήρηση της κίνησης (τακτικές υπηρεσίες) και η ικανότητα να μεταφέρονται μεγάλες ποσότητες φορτίου σε αυτές τις διαδρομές. Ένας σημαντικός λόγος για την επιλογή των θαλάσσιων οδών είναι ότι τα συστήματα σιδηροδρομικών και οδικών

μεταφορών δεν μπορούν να υποστηρίξουν τέτοιας κλίμακας και έντασης εμπορικής δραστηριότητα Stopford, (2009)..

Λόγω της μορφολογίας, της γεωπολιτικής και της εμπορικής δραστηριότητας, συγκεκριμένες τοποθεσίες παίζουν στρατηγικό ρόλο στο παγκόσμιο θαλάσσιο δίκτυο. Αυτές οι τοποθεσίες χαρακτηρίζονται ως στρατηγικά περάσματα και μπορούν να ταξινομηθούν σε δύο κύριες κατηγορίες. Τα βασικά περάσματα (primary passages) είναι τα πιο σημαντικά αφού, χωρίς αυτά, θα υπήρχαν περιορισμένες οικονομικά αποδοτικές εναλλακτικές λύσεις που θα μπορούσαν να βλάψουν σοβαρά το παγκόσμιο εμπόριο. Μεταξύ αυτών είναι η Διώρυγα του Παναμά, η Διώρυγα του Σουέζ, το Στενό του Ορμούζ και το Στενό της Μαλάκας (Σχήμα 1.3), τα οποία αποτελούν βασικές τοποθεσίες στο παγκόσμιο εμπόριο αγαθών και αγαθών.

Σχήμα 1.3. *Στενά της Malaccas (Πηγή: Puigrefagut, 2021).*

Η επέκταση της Διώρυγας του Παναμά, η οποία πραγματοποιήθηκε το 2016, είναι μια προσπάθεια παροχής πρόσθετης χωρητικότητας, τόσο στο μέγεθος του πλοίου που μπορεί να εισέλθει όσο και στην συνολικό αριθμό των πλοίων που μπορεί να εξυπηρετήσει, διατηρώντας την φήμη της ως ένα από τα πιο σημαντικά στρατηγικά παγκόσμια περάσματα που λειτουργούν εδώ και 100 χρόνια. Τα δευτερεύοντα περάσματα (secondary passages) υποστηρίζουν θαλάσσιες διαδρομές για τις οποίες υπάρχουν εναλλακτικές λύσεις που θα συνεπαγόταν ακόμα μια αξιόλογη παράκαμψη. Αυτά περιλαμβάνουν το πέρασμα του Μαγγελάνου, το Στενό του Ντόβερ, το Στενό του Σούντα και το Στενό της Ταϊβάν. Ιστορικά, αυτά τα περάσματα έχουν αμφισβητηθεί και πολλές φορές εξετάστηκε και το κλείσιμο π.χ. η Διώρυγα του Σουέζ

που έκλεισε μεταξύ 1967 και 1975 και τα Στενά του Ορμούζ. Πιο πρόσφατα, η θαλάσσια πειρατεία γνώρισε μια αναζωπύρωση λόγω της αύξησης του όγκου και της αξίας του εμπορίου, ιδιαίτερα στην περιοχή του Στενού του Bab-el-Mandab, που συνδέει τον Ινδικό Ωκεανό με την Ερυθρά Θάλασσα Rodriguez, (2013).

Ωστόσο, αρκετές τεχνικές αλλαγές έχουν βελτιώσει τις θαλάσσιες μεταφορές και ως προς την ικανότητα αλλά και ως προς την αξιοπιστία τους. Το πρώτο και πιο προφανές οφείλεται στο ότι οι περισσότερες κατηγορίες πλοίων έχουν γίνει μεγαλύτερες, γεγονός που έχει βελτιώσει σημαντικά την απόδοση της κάθε διαδρομής που ήταν πάντα προς όφελος της θαλάσσιας ναυτιλίας. Οι μόνοι περιορισμοί που οφείλονται στο μέγεθος του πλοίου είναι η ικανότητα των λιμανιών και καναλιών να τα φιλοξενήσουν. Δεύτερον, η ταχύτητα των πλοίων έχει βελτιωθεί οριακά, καθώς τα πλοία μεταφοράς εμπορευματοκιβωτίων είναι ταχύτερα από τα συμβατικά πλοία που έχουν αντικαταστήσει. Αυτές οι μικρές βελτιώσεις ταχύτητας σημαίνουν ότι, σε υπερωκεάνιες αποστάσεις, οι διαδρομές μπορούν να γίνουν ταχύτερα, κερδίζοντας έως και λίγες ημέρες, κάτι που είναι σημαντικό. Τρίτον, τα πλοία εξελίσσονται όλο και περισσότερο, με πολλά να έχουν σχεδιαστεί αποκλειστικά για τη μεταφορά ενός τύπου φορτίου, όπως εμπορευματοκιβώτια, πετρέλαιο, οχήματα ή υγρό φυσικό αέριο. Αυτή η εξέλιξη περιλαμβάνει και τη υιοθέτηση νέων τύπων πλοήγησης που στηρίζονται σε αλγορίθμους τεχνητής μάθησης και του στου οποίους γίνεται εκτενής αναφορά στο επόμενο κεφάλαιο. Τέταρτον, ο σχεδιασμός των πλοίων έχει βελτιωθεί, επιτρέποντας την κατασκευή μεγαλύτερων πλοίων με μεγαλύτερη ενεργειακή απόδοση. Πέμπτον, ο αυτοματισμός επέτρεψε στα πλοία να επανδρώνονται από μικρότερα πληρώματα βελτιώνοντας παράλληλα τα πρότυπα ασφαλείας Rodriguez, (2017).

Τα δίκτυα θαλάσσιων μεταφορών έχουν σχεδιαστεί για να χρησιμοποιούν τη συντομότερη διαδρομή και παράλληλα να μπορούν να εξυπηρετούν τις κύριες αγορές. Αυτό οδηγεί σε διάφορους συμβιβασμούς μεταξύ του αριθμού των λιμένων που καλούνται να εξυπηρετήσουν και του αριθμού των πλοίων που προορίζονται για συγκεκριμένες συναλλαγές. Η μεταφορά εμπορευματοκιβωτίων είχε εκτεταμένο αντίκτυπο στη διαμόρφωση των θαλάσσιων διαδρομών, ιδιαίτερα επειδή, πριν από τη μεταφορά εμπορευματοκιβωτίων, η φόρτωση ή εκφόρτωση ενός πλοίου ήταν ένα πολύ ακριβό και χρονοβόρο έργο. Ένα φορτηγό πλοίο περνούσε συνήθως περισσότερο

χρόνο ελλιμενισμένο παρά στη θάλασσα. Η κατάσταση αυτή έχει πλέον αντιστραφεί και τα πλοία μεταφοράς εμπορευματοκιβωτίων περνούν περισσότερο χρόνο στη θάλασσα παρά στο λιμάνι, καθώς είναι σε συνεχή κίνηση μεταξύ των λιμένων που καλούνται να εξυπηρετήσουν.

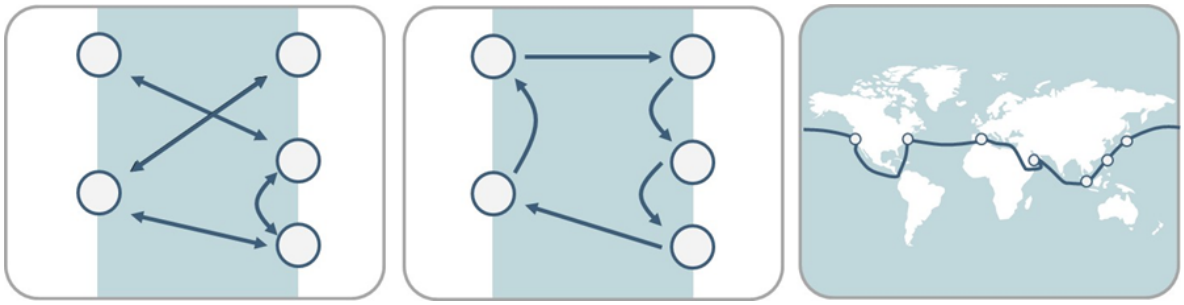
Ως εκ τούτου, τα πλοία μεταφοράς εμπορευματοκιβωτίων δημιούργησαν μια νέα δομή δικτύου: η μεταφορά από λιμάνι σε λιμάνι γνωστή και ως port to port έχει εξελιχθεί σε μια πιο σύνθετη δομή που καλύπτει πολλές αγορές. Οι θαλάσσιες διαδρομές οργανώνονται σύμφωνα με τις εμπορικές υπηρεσίες που υποστηρίζουν. Αυτές οι υπηρεσίες μπορούν να χωριστούν (Notteboom, 2012; Stopford, 2009; Rodriguez, 2013) σε τρεις κύριες κατηγορίες (Σχήμα 1.4):

Το **port-to-port** αντιπροσωπεύει τη συμβατική δομή εξυπηρέτησης που αφορά τις τακτικές κλήσεις μεταξύ δύο λιμένων. Συνήθως τα πλοία κινούνται μπρος και πίσω με πλήρες φορτίο προς τη μία κατεύθυνση και χωρίς φορτίο στην άλλη. Αυτή η δομή δικτύου είναι χαμηλής συνδεσιμότητας και αποτελεί την συνηθέστερη μορφή μεταφοράς πρώτων υλών όπως το πετρέλαιο, τα ορυκτά και τα σιτηρά. Αυτές οι αγορές συνήθως εξυπηρετούνται από ναυλωμένα πλοία που φορτώνουν σε ένα λιμάνι (δίπλα σε μια σημαντική περιοχή εξόρυξης πόρων) και εκφορτώνουν το φορτίο τους σε ένα έως τρία λιμάνια.

Η μέθοδος της **ταλάντευσης ή pendulum ή inter range**, χρησιμοποιείται συνήθως για τη μεταφορά εμπορευματοκιβωτίων με τακτικά δρομολόγια που καλύπτουν ένα σύνολο κλήσεων λιμένων που εξυπηρετούνται διαδοχικά. Η ακολουθία αυτών των λιμένων είναι προφανώς επιλεγμένη για τη μεγιστοποίηση του συντελεστή φορτίου των πλοίων. Ο όρος «ταλάντευση» αναφέρεται στην υπηρεσία αποστολής που κινείται συνεχώς και διαδοχικά μεταξύ δύο ή και τριών θαλάσσιων περιοχών. Οι σημαντικότερες διαδρομές αυτού του τύπου είναι μεταξύ Ανατολικής Ασίας, Βόρειας Αμερικής και Δυτικής Ευρώπης, των τριών κύριων πόλων της παγκόσμιας οικονομίας

Η μέθοδος **παγκόσμιας εξυπηρέτησης ή round the world ή multi ranges** σχετίζονται επίσης με τη μεταφορά εμπορευματοκιβωτίων και περιλαμβάνουν αποστολές σε μια σειρά λιμένων, συχνά και προς τις δύο κατευθύνσεις, έτσι ώστε η ακολουθία να αποτελεί ένα ταξίδι σε όλο τον κόσμο. Σε αυτή την περίπτωση, μόνο

ένας περιορισμένος αριθμός λιμένων ανά ήπειρο εξυπηρετείται, αλλά αυτοί οι λιμένες είναι είτε μεγάλες πύλες είτε κόμβοι μεταφοράς. Έτσι, η μέθοδος της παγκόσμιας εξυπηρέτησης είναι μια προσπάθεια καλύτερης σύνδεσης οριζόντιων (γεωγραφικό μήκος) και κάθετων (γεωγραφικό πλάτος) γεωγραφικών εμπορικών ροών.



Port to Port **Pendulum (Inter range)** **Round the world ή Multi Ranges**

Σχήμα 1.4. *Τύποι θαλάσσιων διαδρομών (Πηγή: Rodriguez, 2013)*

1.3. Παράμετροι που ορίζουν την πορεία ενός πλοίου

Ο σχεδιασμός της πορείας ενός πλοίου εξαρτάται από ένα πλήθος παραγόντων. Ο βασικότερος παράγοντας είναι σχεδόν πάντοτε το κόστος της διαδρομής. Πιο συγκεκριμένα, το κόστος μιας διαδρομής αποτελεί τον καθοριστικό παράγοντα στην επιλογή του και περιλαμβάνει και τη θέση του λιμανιού. Άλλοι παράγοντες που επηρεάζουν μια διαδρομή είναι (Ren, Lutzen & Rasmussen, 2018)

- **Τεχνολογία.** Οι τεχνολογικά αναπτυγμένοι λιμένες είναι πιο αξιόπιστοι και αποδοτικοί από άλλους λιμένες, καθώς επιτρέπουν την πιο γρήγορη φόρτωση ή εκφόρτωση. Επιπλέον, η τεχνολογία συμβάλλει στη βελτίωση των λειτουργιών λιμένων και ελαχιστοποιεί τα ανθρώπινα λάθη.
- **Τοποθεσία.** Οι περισσότεροι εταιρείες επιλέγουν το πλησιέστερο λιμάνι για την αποστολή των εμπορευμάτων τους, καθώς τα κοντινά λιμάνια είναι συνήθως προσβάσιμο και φθηνότερα .
- **Μέγεθος.** Το μέγεθος και ο τύπος του φορτίου επηρεάζει τον σχεδιασμό της πορείας ενός πλοίου καθώς έχει άμεση σχέση με το λιμάνι εκφόρτωσης. Υποδομή & διαθεσιμότητα εξοπλισμού
- **Κλίμα.** Στην εποχή των έντονων κλιματικών αλλαγών, το κλίμα που αναμένεται να επικρατήσει κατά την διάρκεια του δρομολογίου μπορεί να επηρεάσει τόσο την ταχύτητα του πλοίου όσο και την ασφάλεια του.

- **Κατάσταση του πλοίου.** Οι παράγοντες που έχουν σχέση με τις επιδόσεις του πλοίου όπως επίδοση κινητήρα, παλαιότητα πλοίου κ.α. μπορεί να επηρεάσουν και την τελική επιλογή της διαδρομής του.
- **Ασφάλεια διαδρομής.** Σε διαδρομές όπου παρατηρούνται επιθέσεις πειρατών ή τρομοκρατικές ενέργειες ή διαδρομές που περνούν κοντά από χώρες σε εμπόλεμη κατάσταση είναι πιθανόν να αποκλειστούν από τον σχεδιασμό της διαδρομής, ακόμη και εάν είναι πιο οικονομικές με βασικό κριτήριο την ασφάλεια του πλοίου, του πληρώματος και του φορτίου.
- **Περιβαλλοντικοί παράγοντες (Green Shipping).** Η πράσινη ναυτιλία αφορά καθαρότερες πρακτικές στον έλεγχο των εκπομπών, τη διαχείριση των λιμένων και τους κύκλους ζωής του εξοπλισμού, δηλαδή την κυκλική οικονομία.

Επιπλέον, και σύμφωνα με την Das, (2019), οι παράγοντες που επηρεάζουν τον ηλεκτρονικό σχεδιασμό της πορείας ενός πλοίου είναι:

- **Η εκτίμηση της κυκλοφορίας σε πραγματικό χρόνο :** Σύμφωνα με την Das, (2019) το 2017, οι ΗΠΑ ανέφεραν απώλειες της τάξης των 305 δισεκατομμυρίων δολαρίων λόγω συμφόρησης. Η εξέταση της κίνησης σε μια διαδρομή, σε πραγματικό χρόνο μπορεί όχι μόνο να εξοικονομήσει τα έξοδα υλικοτεχνικής υποστήριξης, αλλά μπορεί επίσης να διασφαλίσει την έγκαιρη παράδοση και την καλύτερη τήρηση των συμφωνηθέντων ενός συμβολαίου (Service Level Agreement ή SLA).
- **Ακριβής γεωκωδικοποίηση** : Σχεδόν κάθε λογισμικό βελτιστοποίησης διαδρομής που διατίθεται στην αγορά διαθέτει έναν γεωκωδικοποιητή. Ωστόσο, η μετατροπή διευθύνσεων σε ένα συγκεκριμένο σημείο του χάρτη σε ακριβείς συντεταγμένες γεωγραφικού πλάτους και γεωγραφικού μήκους είναι μια επίπονη εργασία. Η κατανόηση διφορούμενων διευθύνσεων και η κατανόηση τοπικών πλαισίων είναι το κλειδί για την ακριβή γεωκωδικοποίηση.
- **Επιθεώρηση παλαιότερων δεδομένων** : Ένα λογισμικό βελτιστοποίησης μιας διαδρομής θα πρέπει να διδαχθεί από προηγούμενες εμπειρίες και να σχεδιάσει διαδρομές ανάλογα.
- **Analytics και Report Management** : Το λογισμικό βελτιστοποίησης μιας διαδρομής θα πρέπει να σας δίνει τη δυνατότητα παρακολούθησης και

διαχείρισης του συνόλου των λειτουργιών σε πραγματικό χρόνο σε μία μόνο πλατφόρμα. Με αυτό τον τρόπο γίνεται παρακολούθηση των πραγματικών διαδρομών έναντι των προγραμματισμένων διαδρομών και παράλληλα μπορεί να γίνει σύγκριση επιδόσεων μεταξύ τους.

- **Δυναμικός Σχεδιασμός Διαδρομής** : Η εναλλαγή δρομολογίων εν κινήσει είναι ένα χαρακτηριστικό που επιτρέπει την επιτόπου επανα-χάραξη της πορείας ενός πλοίου σε περιπτώσεις που αυτό κριθεί απαραίτητο.

Σύμφωνα με τα προηγούμενα, γίνεται φανερό ότι ο σχεδιασμός της πορείας ενός πλοίου εμπεριέχει πάρα πολλές παραμέτρους, μερικές από τις οποίες είναι απρόβλεπτες. Επιπλέον ο σχεδιασμός αυτός εμπεριέχει και το εμπειρικό στοιχείο και είναι πιθανόν να υπόκειται σε συνεχείς αλλαγές. Για αυτόν τον λόγο θα πρέπει να γίνεται συνεχής εκτίμηση της και επανα-σχεδιασμός με βάση ειδικούς αλγορίθμους που περιγράφονται στα επόμενα κεφάλαια.

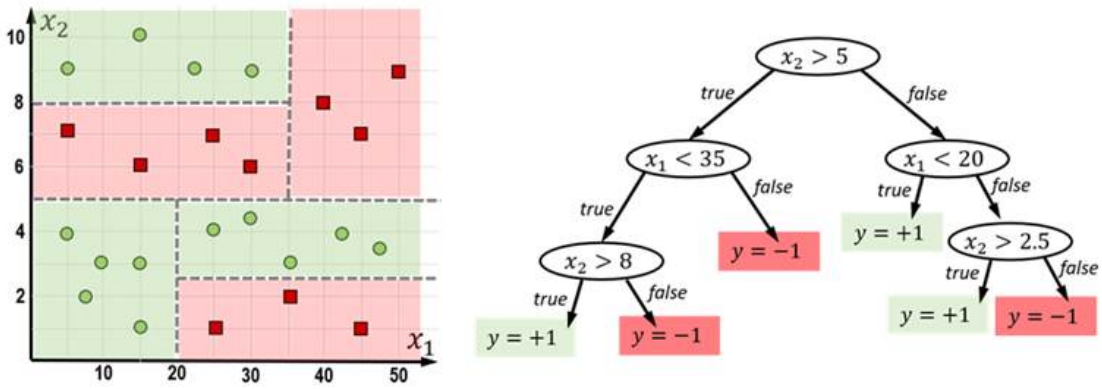
Κεφάλαιο 2

Περιγραφή μεθόδων

Σε αυτό το κεφάλαιο γίνεται η περιγραφή των μεθόδων που θα χρησιμοποιηθούν για την αξιολόγηση των αλγορίθμων πρόβλεψης της μελλοντικής θέση ενός πλοίου.

2.1. Η μέθοδος Random forest

Για την παρουσίαση της μεθόδου Random Forest, πρέπει πρώτα γίνει εισαγωγή του δέντρου διαχωρισμού (partition tree). Ένα δέντρο διαχωρισμού μπορεί να χρησιμοποιηθεί για την εκτέλεση παλινδρόμησης ή ταξινόμησης. Οι Hastie et al. (2009) εξηγούν πώς οι μέθοδοι που βασίζονται σε δενδροειδή ταξινόμηση π.χ. η Random Forest διαχωρίζουν μια ομάδα μεγεθών π.χ. μετρήσεων με βάση τις μεταβολές ενός ή και περισσότερων χαρακτηριστικών. Ο χώρος που δημιουργείται μετά τις διαμερίσεις ονομάζεται διαμοιρασμένος χώρος (Partitioned Feature Space). Οι μέθοδοι που βασίζονται σε δέντρα επιμερίζουν τον χώρο σε ένα σύνολο ορθογωνίων και στη συνέχεια μεταφέρονται σε ένα απλό μοντέλο (Hastie et al., 2009). Στην εικόνα 2.1 παρουσιάζεται ένας χώρος με δύο μεταβλητές πρόβλεψης και η αντιστοίχιση του με το μοντέλου δέντρου.



Εικόνα 2.1. *Παράδειγμα διαμοιρασμένου χώρου και του αντίστοιχου δέντρου που παράγεται (Πηγή: Kunapuli, 2020)*

Kunapuli, G, (2020). Ensemble Methods for Machine Learning. NY, Manning Publications

Σύμφωνα με το σχήμα 2.1 το δέντρο διαχωρίζεται για τις τιμές x_i i=1,2. Ο πρώτο διαχωρισμός αφορά την τιμή 5 για την μεταβλητή x_2. Έτσι οι δύο πρώτοι κλάδοι αντιστοιχούν στην αληθή ή ψευδή συνθήκη $x_2 < 5$. Ο αμέσως επόμενος διαχωρισμός

αφορά την συνθήκη $x_1 < 35$ στο αριστερό άκρο του δέντρου και $x_1 < 20$ κ.ο.κ. Οι τελικές απολήξεις του δέντρου ονομάζονται τερματικοί κόμβοι ή terminal nodes και υπολογίζονται από τους μέσους των εξαρτημένων μεταβλητών για το σύνολο των παρατηρήσεων που αντιστοιχούν σε εκείνη την περιοχή. Αυτή η διαδικασία διαμερίζει των χώρο σε 6 περιοχές οι οποίες εμφανίζονται ως αποτελέσματα της εξαρτημένης μεταβλητής στο δέντρο.

Έστω ότι έχουμε N παρατηρήσεις, και p μεταβλητές και πρέπει να κάνουμε M διαχωρισμούς. Έστω επίσης ότι $x_{i.} = \{x_{i1}, x_{i2}, .., x_{ip}\}$ αποτελεί σύνολο p εισόδων για κάθε παρατήρηση i, $y_{i.} = \{y_{i1}, y_{i2}, \ldots, y_{ip}\}$ οι αντίστοιχες μετρήσεις, $R_{i.} = \{R_{i1}, R_{i2}, \ldots, R_{ip}\}$ το σύνολο των περιοχών και έστω c_m σταθερά του μοντέλου με εξαρτημένη μεταβλητή την y_i. Τότε

$$f(x) = \sum_{m=1} c_m I\,(x \in R_m), \qquad (2.1)$$

όπου I(A) ισούται με 1 εάν η συνθήκη A είναι αληθής και 0 σε αντίθετη περίπτωση. Σκοπός είναι η ελαχιστοποίηση του τετραγωνικού αθροίσματος των καταλοίπων ή RSS (Residual Sum of Squares)

$$RSS = \sum_{i=1}^{N} (y_i - f(x_i))^2$$

(2.2)

όπου $f(x_i)$είναι η μέση τιμή των παρατηρήσεων στην περιοχή R_m. Για τον προσδιορισμό της βέλτιστης δυαδικής διαίρεσης η διαδικασία θα πρέπει να ξεκινά στο σύνολο των δεδομένων και στην συνέχεια να γίνεται επιλογή της μεταβλητής διαχωρισμού j, του σημείου s. Τότε κάθε περιοχή συμβολίζεται ως

$$R_1\,(j,s)\ \{X \mid X\,j \le s\}, \text{ και}$$

$$R_2\,(j,s)\ \{X \mid X\,j > s\} \qquad (2.3)$$

Ο χώρος διαχωρίζεται καλύτερα επιλέγοντας τη μεταβλητή διαίρεσης j και ένα σημείο διάσπασης s ελαχιστοποιεί το άθροισμα του RSS σε κάθε μία από τις δύο ξεχωριστές περιοχές. Στην συνέχεια οι δύο περιοχές χωρίζονται ξανά χρησιμοποιώντας την ίδια διαδικασία σε περισσότερα διαμερίσματα. Για λεπτομερή περιγραφή του τρόπου με

τον οποίο μπορεί να επιτευχθεί το καλύτερο μέγεθος δέντρου, δείτε (Hastie et al., 2009).

Τα δέντρα αποφάσεων από μόνα τους είναι γνωστό ότι έχουν μειονεκτήματα όπως η χρήση περισσότερων μεταβλητών από τις αναγκαίες για προσαρμογή (overfitting) όπως και η ευαισθησία σε ακραίες τιμές (outliers). Τα μεμονωμένα δέντρα χαρακτηρίζονται από υψηλή διακύμανση, αλλά έχουν μικρή μεροληψία. Η μέθοδος αυτή κάνει μεμονωμένους διαχωρισμούς χρησιμοποιώντας την τεχνική bootstrapping, δηλαδή μέσω τυχαίων δειγμάτων από το σύνολο των δεδομένων. Ο εμπειρικός κανόνας $m \approx \sqrt{p}$ είναι ένας κατάλληλος αριθμός μεταβλητών πρόβλεψης για χρήση σε κάθε διαίρεση. Τέλος, τα αποτελέσματα της πρόβλεψης από όλα τα δέντρα υπολογίζονται κατά μέσο όρο για να παράγουν ένα τελικό μοντέλο σύμφωνα με τις χαμηλότερες τιμές του RSS.

2.2.2 Η μέθοδος Gradient Boosting

Μία άλλη πολύ διαδεδομένη μέθοδος βασισμένη σε δέντρα είναι η Boosting. Η τεχνική λειτουργεί παρόμοια με την Random Forest, εκτός από το γεγονός ότι τα δέντρα δημιουργούνται διαδοχικά. Κάθε δέντρο εκπαιδεύεται χρησιμοποιώντας πληροφορία από τα προηγούμενα δέντρα, σε αντίθεση με την προηγούμενη μέθοδο που τα δέντρα είναι ασυσχέτιστα μεταξύ τους. Ο αλγόριθμος λειτουργεί έως εξής:

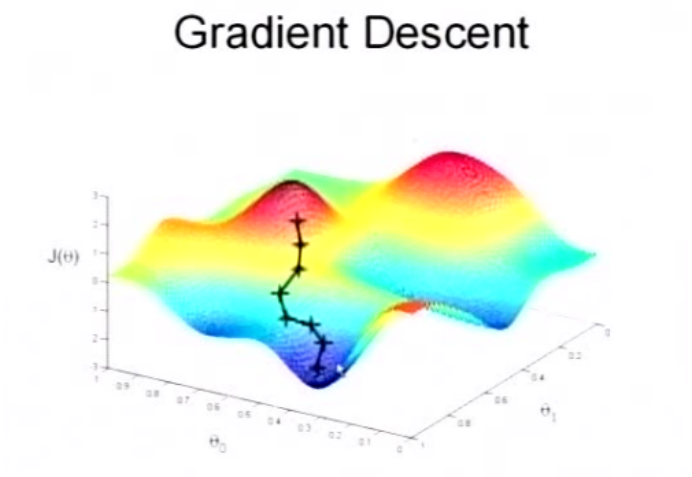
1. Πρώτα τίθεται $f(x) = 0$ και τα κατάλοιπα $\varepsilon_i = y_i$ για κάθε παρατήρηση στο σύνολο των δεδομένων εκπαίδευσης.
2. Γίνεται εκπαίδευση ενός δέντρου f^k σε κάθε επανάληψη k με d κόμβους έχοντας σαν μεταβλητή απόκρισης τα κατάλοιπα.
3. Γίνεται προσθήκη μίας περικομμένης έκδοσης του νέου δέντρου : $f(x) \leftarrow f(x) + \lambda f^k(x)$
4. Αναβαθμίζονται τα κατάλοιπα : $\varepsilon_i \leftarrow \varepsilon_i - \lambda f^k(x)$
5. Γίνεται επανάληψη της διαδικασίας από το βήμα 2 K φορές (σύμφωνα με τον χρήστη) καταλήγοντας στην τελική μορφή του μοντέλου:

$$f(x) = \lambda \sum f k(x) \tag{2.4}$$

Η gradient boosting τεχνική προτάθηκε από τον Friedman(2001) και ουσιαστικά αποτελεί μία επέκταση της τεχνικής Boosting. Το όνομα της το πήρε από δύο μεθόδους, τον αλγόριθμο Gradient Descent και την τεχνική Boosting. Η Gradient Descent αποτελεί μία μέθοδο βελτιστοποίησης πρώτης τάξης. Για τον υπολογισμό του ολικού ελαχίστου μιας συνάρτησης χρησιμοποιώντας τη συγκεκριμένη τεχνική, αρχικά γίνεται ο υπολογισμός της παραγώγου και στην συνέχεια κινείται ανάποδα από την κατεύθυνση της παραγώγου. Η παράγωγος μετράει κατά πόσο θα αλλάξει η τιμή μίας συνάρτησης $J(\theta)$ εάν μεταβληθεί ελάχιστα η παράμετρος θ. Ουσιαστικά αποτελεί την κλίση της συνάρτησης και υψηλές τιμές της συνάρτησης υποδηλώνουν μεγάλη κλίση άρα και μεγάλη μεταβολή της $J(\theta)$ για μικρές μεταβολές του θ.Ο συγκεκριμένος αλγόριθμος είναι επαναληπτικός και ορίζει μία τυχαία αρχική τιμή για το θ. Στην συνέχεια υπολογίζει την παράγωγο της συνάρτησης στο συγκεκριμένο σημείο και μεταβάλει το θ κατά :

$$\theta = \theta - \rho \, dJ/d\theta \tag{2.5}$$

όπου η παράμετρος ρ καθορίζει το πόσο γρήγορα γίνει η κίνηση στην αρνητική κατεύθυνση της παραγώγου. Η διαδικασία επαναλαμβάνεται έως ότου συγκλίνει ο αλγόριθμος. Ο αλγόριθμος παρουσιάζεται και οπτικά στην εικόνα 2.2.



Εικόνα 2.2. *Γραφική αναπαράσταση διαδικασίας Gradient Descent (Πηγή: Awasthi, 2021)* ·

Ο λόγος που η μέθοδος Gradient Boosting είναι καλύτερη σε σύγκριση με την απλή μέθοδο Boosting οφείλεται στο ότι επιτρέπει τη δυνατότητα επιλογής διαφορετικών συναρτήσεων απώλειας (Loss function). Έτσι, ανάλογα με τη δομή των δεδομένων χρησιμοποιούνται και διαφορετικές συναρτήσεις απώλεια. Για παράδειγμα αν τα δεδομένα περιέχουν αρκετές ακραίες τιμές το άθροισμα των τετραγώνων RSS επηρεάζεται σε μεγαλύτερο βαθμό από ότι το άθροισμα των απόλυτων τετραγωνικών σφαλμάτων (Absolute RSS). Οι παράμετροι που πρέπει να οριστούν από τον χρήστη είναι ίδιες με αυτές της τεχνικής Boosting με βασική προϋπόθεση ότι η συνάρτηση που επιλέγεται θα πρέπει να είναι παραγωγίσιμη.

2.3. Η μέθοδος Naive Bayes

Ο αλγόριθμος Naïve Bayes βασίζεται στο θεώρημα πιθανοτήτων του Bayes (Heckerman, 1999), που έχει στόχο να προβλέψει αποτελέσματα από μη-επισημασμένα δεδομένα. Στο θεώρημα Bayes, η μέθοδος ταξινόμησης προϋποθέτει την ανεξαρτησία μεταξύ των χαρακτηριστικών(features) και της κλάσης κατηγοριοποίησης(data). Τα μοντέλα Naïve Bayes χωρίζονται σε διάφορους τύπους ανάλογα με το χειρισμό των χαρακτηριστικών τους. Στο μοντέλο Bernoulli, οι τιμές των χαρακτηριστικών πρέπει να είναι δυαδικές(0-1, True-False κλπ.). Ο αλγόριθμος Naïve Bayes ταξινόμησης είναι χρήσιμος για να χαρακτηρίσει ακόμα και σύνολα δεδομένων με υψηλό όγκο πληροφοριών, καθώς εκτελείται αποτελεσματικά και είναι εύκολο να εφαρμοστεί. Ο κανόνας του Bayes εκφράζεται από την εξίσωση

$$P(c|X) = [P(X|c) * P(c)] / P(X) \quad (2.6)$$

όπου P(c|x) εκφράζει την εκ των υστέρων πιθανότητα, P(x|c) εκφράζει την δεσμευμένη πιθανότητα, P(c) την εκ των προτέρων πιθανότητα κλάσης και P(X) την εκ των προτέρων πιθανότητα ταξινομητή Η λογική έκφραση της 2.6 είναι, η εύρεση της πιθανότητα του γεγονότος c δεδομένου ότι έχει προηγηθεί ένα γεγονός X. Στην περίπτωση που εξετάζονται περισσότερο από ένα γεγονότα τότε ισχύει ότι

$$P(c \mid X) = P(x1 \mid c) \times P(x2 \mid c) \times \ldots \times P(xn \mid c) \times P(c) \quad (2.6B)$$

Η εκ των υστέρων πιθανότητα υπολογίζει την πιθανότητα του αποτελέσματος που προκύπτει από μια νέα πληροφορία. Στην 2.6 στο P(c|x), το c αναπαριστά την κλάση που ταξινομούνται τα δεδομένα και το x τον ταξινομητή. Η δεσμευμένη πιθανότητα

εκφράζει τη πιθανότητα να βρίσκεται ο ταξινομητής (χαρακτηριστικό) μέσα στην κλάση (Webb, 2011).

2.4. Η μέθοδος Multilayer Perceptron (MLP)

Τα Νευρωνικά Δίκτυα ή Neural Networks θεωρούνται, και αποτελούν, ένα από τα σημαντικότερα επιτεύγματα της Τεχνητής Νοημοσύνης. Με κύρια πηγή έμπνευσης το βιολογικό νευρικό σύστημα, και πιο συγκεκριμένα τον ανθρώπινο εγκέφαλο, εμφανίζουν αξιοσημείωτα χαρακτηριστικά, όπως π.χ. τη δυνατότητα να αναπαριστούν σύνθετες εξαρτήσεις και την ικανότητα να προβλέπουν την κλάση μεταξύ άγνωστων παρατηρήσεων. Χάρη στη στιβαρή (robust) θεωρητική τους θεμελίωση και στις αξιόλογες δυνατότητες τους έχουν καταξιωθεί και είναι ιδιαίτερα δημοφιλή με αμέτρητες εφαρμογές σε τομείς, όπως η ιατρική, η οικονομία, η διαφήμιση κ.α. Τα Νευρωνικά Δίκτυα είναι μια τεχνική η οποία καθοδηγείται ισχυρά από τα δεδομένα. Αυτό πρακτικά σημαίνει ότι δεν επιτρέπει την επιβολή αυθαίρετων υποθέσεων και τα μοντέλα τους εξάγονται μέσω της επεξεργασίας των δεδομένων. Τα Νευρωνικά δίκτυα περιέχουν μεθόδους τόσο της επιβλεπόμενης, όσο και της μη επιβλεπόμενης μάθησης.

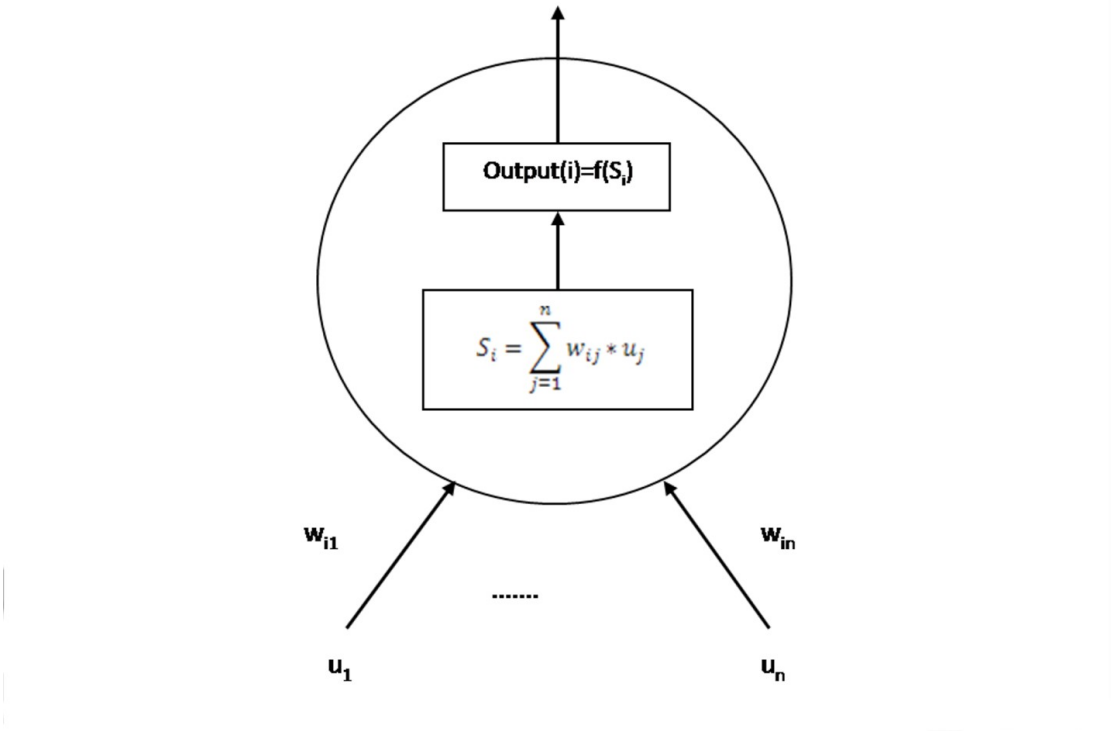
Η βασική δομική μονάδα των Νευρωνικών Δικτύων είναι οι νευρώνες. Αυτοί οι νευρώνες ονομάζονται κόμβοι ή κελιά και κάθε ένας νευρώνας αποτελεί μια στοιχειώδη υπολογιστική μονάδα. Η μονάδα αυτή δέχεται πολλές τιμές εισόδου και υπολογίζει μια τιμή εξόδου. Η γραφική αναπαράσταση των νευρώνων περιλαμβάνει την απεικόνιση των σχέσεων μεταξύ τους μεταξύ τους με κατευθυνόμενα βέλη ή συνδέσεις. Αυτές οι αναπαραστάσεις (εικόνα 2.3) απεικονίζουν την σχέση όπου ένας νευρώνας παραλαμβάνει την πληροφορία (τιμές εσόδου) από άλλους νευρώνες και την μεταβιβάζει ως τιμή εξόδου σε άλλους νευρώνες. Σε κάθε σύνδεση παρουσιάζεται και μία αριθμητική τιμή που ονομάζεται βάρος w. Σκοπός αυτού του μεγέθους είναι να επηρεάσει την επίδραση μεταξύ των συνδεδεμένων νευρώνων. Έτσι, εάν με u_j συμβολίσουμε την τιμή εξόδου του νευρώνα j, για την μεταβίβαση του στον νευρώνα i, το u_j θα πολλαπλασιαστεί με το βάρος της σύνδεσης των δύο νευρώνων w_{ij}.

Η επεξεργασία που διενεργεί ένας νευρώνας i ολοκληρώνεται σε δύο στάδια, στο πρώτο στάδιο γίνεται άθροιση των τιμών εισόδου που ισούνται με τις τιμές εξόδου των

συνδεδεμένων νευρώνων, πολλαπλασιασμένες με τα βάρη των αντίστοιχων συνδέσεων. Για τον i νευρώνα που δέχεται τιμές εισόδου u_j από n νευρώνες, το συνολικό σήμα εισόδου S_i υπολογίζεται σύμφωνα με την Εξίσωση 2.7

$$S_i = \sum_{j=1}^{n} w_{ij} \cdot u_j \qquad (2.7)$$

Σε δεύτερο στάδιο, γίνεται μετασχηματισμός των αθροισμάτων των τιμών εισόδου, με τη χρήση μιας συνάρτησης γνωστής ως συνάρτηση ενεργοποίησης (activation function) ή συνάρτησης μετασχηματισμού. Η τελική τιμή υπολογισμού είναι η τιμή εξόδου του νευρώνα. Τα παραπάνω απεικονίζονται στην εικόνα 2.3.

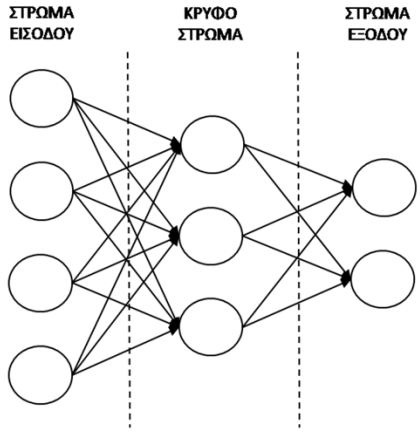


Εικόνα 2.3. *Διαδικασία ενεργοποίησης νευρών (Πηγή: Κύρκος, 2015)*

Ως συναρτήσεις ενεργοποίησης μπορούν να χρησιμοποιηθούν διάφορες μαθηματικές συναρτήσεις όπως η συνάρτηση συνημίτονου, η συνάρτηση ημίτονου κ.α. Η Σιγμοειδής συνάρτηση αποτελεί την πιο συχνά χρησιμοποιούμενη, καθώς είναι απλή και μη γραμμική αλλά και επειδή έχει παρόμοια συμπεριφορά με τη συμπεριφορά των πραγματικών νευρώνων. Η Σιγμοειδής συνάρτηση ορίζεται από την εξίσωση

$$f(x) = \frac{1}{1+e^{-x}} \quad (2.8)$$

Το μοντέλο Perceptron αποτελεί την απλούστερη μορφή νευρωνικού δικτύου, και έχει εφαρμογές στην ταξινόμηση (Classification), γραμμικά διαχωρίσιμων προτύπων. Διακρίνεται για την δυνατότητά του στην εύκολη προσαρμογή στην μηχανική μάθηση και στην ακριβή εφαρμογή του σε δεδομένα.. ο perceptron πολλαπλών επιπέδων (MLP). Αποτελείται από τρεις τύπους στρωμάτων - το στρώμα εισόδου (input layer), το επίπεδο εξόδου (output layer) και το κρυφό επίπεδο (hidden layer), όπως φαίνεται στο σχήμα 2.4. Το επίπεδο εισόδου λαμβάνει το σήμα εισόδου προς επεξεργασία. Η απαιτούμενη εργασία όπως η πρόβλεψη και η ταξινόμηση εκτελείται από το επίπεδο εξόδου . Ένας αυθαίρετος, και ορισμένος από τον χρήστη, αριθμός κρυφών στρωμάτων που τοποθετούνται μεταξύ του επιπέδου εισόδου και εξόδου είναι ο πραγματικός υπολογιστικός κινητήρας του MLP. Παρόμοια με ένα δίκτυο προώθησης τροφοδοσίας σε ένα MLP, τα δεδομένα ρέουν προς την κατεύθυνση προς τα εμπρός από την είσοδο στο επίπεδο εξόδου. Οι νευρώνες στο MLP εκπαιδεύονται με τον αλγόριθμο εκμάθησης πίσω διάδοσης (back propagation). Τα MLP έχουν σχεδιαστεί για να προσεγγίζουν κάθε συνεχή λειτουργία και μπορούν να επιλύσουν προβλήματα που δεν είναι γραμμικά διαχωρίσιμα. Οι κυριότερες περιπτώσεις χρήσης του MLP είναι η ταξινόμηση προτύπων, η αναγνώριση, η πρόβλεψη και η προσέγγιση.



Εικόνα 2.4. *Νευρωνικό δίκτυο τριών επιπέδων (Πηγή: Κύρκος, 2015)*

2.5 Η μέθοδος Logistic regression

Η λογιστική παλινδρόμηση (Logistic regression) είναι ένα μοντέλο ταξινόμησης των τιμών μιας εξαρτημένης μεταβλητής Y με βάση τη θεωρία των πιθανοτήτων. Στο μοντέλο αυτό μεταβλητή Y μπορεί να είναι δυαδική δηλαδή να παίρνει μόνο δύο τιμές π.χ. 1=άνδρας και 2=Γυναίκα. Σκοπός είναι η πρόβλεψη της εξαρτημένης μεταβλητής μέσα από ένα πλήθος προβλεπτικών (ανεξάρτητων) μεταβλητών οι οποίες μπορεί να είναι ποσοτικές ή κατηγορικές (ονομαστικές ή/και διατάξιμες). Η πιο σημαντική διαφορά με την γραμμική παλινδρόμηση είναι ότι η εξαρτημένη μεταβλητή δεν μπορεί να είναι ποσοτική.

Έτσι, ενώ κατά την κλασική γραμμική παλινδρόμηση η εκτίμηση των παραμέτρων α και bi γίνεται με τη μέθοδο των ελάχιστων τετραγώνων, κατά τη λογιστική παλινδρόμηση η εκτίμηση των παραμέτρων γίνεται με τη μέθοδο του λόγου πιθανοφάνειας δηλαδή επιλέγονται οι πιο πιθανοφανείς τιμές των παραμέτρων, προκειμένου να οδηγήσουν στα παρατηρούμενα αποτελέσματα. Ως επακόλουθο, η πρώτη παραδέχεται την ύπαρξη ομοιογένειας (ομοσκεδαστικότητας) στα κατάλοιπα ενώ στη δεύτερη αναπτύσσεται πάντα ετεροσκεδαστικότητα σε κάθε προβλεπόμενη τιμή εξαιτίας του μεταβαλλόμενου ποσοστού διακύμανσης που αναλογεί σε αυτήν.

Σύμφωνα με τον Πετρίδη, (2015) διακρίνονται τρεις τύποι λογιστικής παλινδρόμησης ανάλογα με τον τύπο της εξαρτημένης μεταβλητής η οποία μπορεί να είναι:

1. Δυαδική ή διχοτομική (binary)
2. Τακτική (ordinal) μεταβλητή. Η εξαρτημένη μεταβλητή περιέχει 3 ή περισσότερες κατηγορίες μεταξύ των οποίων ισχύει η έννοια της ανισότητας, όπως π.χ. σε μια ερώτηση της κλίμακας Likert.
3. Ονομαστική (Nominal) η οποία περιέχει τρεις ή περισσότερες κατηγορίες χωρίς κάποια φυσική διαβάθμιση, όπως π.χ. ο χαρακτηρισμός ενός χρώματος.

Η δίτιμη λογιστική παλινδρόμηση έχει τη μορφή

$$f(z) = \frac{e^z}{1+e^z} = \frac{1}{1+e^{-z}}$$

(2.9)

όπου z είναι η μεταβλητή εισόδου και f(z) το αποτέλεσμα αυτής. Στα πλεονεκτήματα της εξίσωσης συγκαταλέγεται και το γεγονός ότι η μεταβλητή εισόδου λαμβάνει θετικές και αρνητικές τιμές ενώ το αποτέλεσμα αυτής f(z) περιορίζεται σε εύρος τιμών μεταξύ 0 και 1. Αναλυτικότερα, η μεταβλητή z εκπροσωπεί τη δράση μιας ομάδας ανεξάρτητων μεταβλητών ενώ η f(z) προσδιορίζει την πιθανότητα ενός συγκεκριμένου αποτελέσματος λόγω της δράσης της ομάδας αυτής.

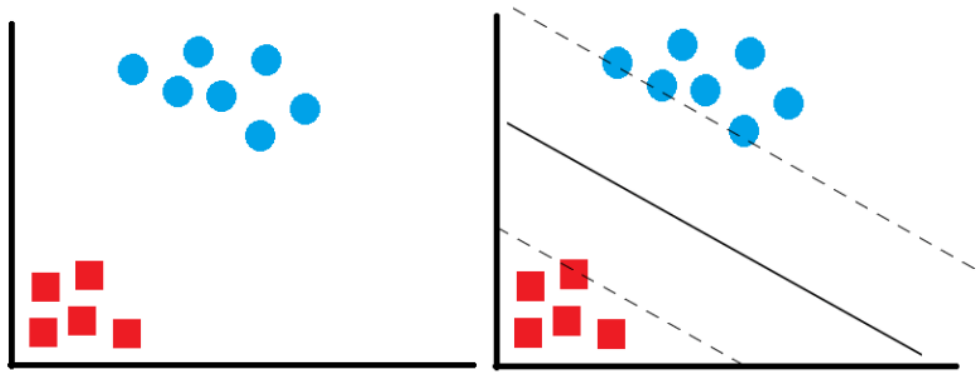
2.6. Η μέθοδος SVM

Οι Μηχανές Διανυσμάτων Υποστήριξης (Support Vector Machines - SVMs) είναι μια τεχνική η οποία ανήκει στην ομάδα των μηχανών εκμάθησης (learning machines) και ως στόχο έχει την επεξεργασία δεδομένων. Χρησιμοποιείται σε προβλήματα ταξινόμησης και στην προσέγγιση της μορφής της συνάρτησης σε προβλήματα παλινδρόμησης. Η γενική ιδέα αυτή της μεθόδου είναι ευκολονόητη και περιλαμβάνει τον διαχωρισμό των δεδομένων με την βοήθεια κατάλληλων ορίων (ή συνόρων) ανάλογα με τις διαστάσεις του χώρου R στον οποίο εφαρμόζονται. Πιο συγκεκριμένα και ανάλογα με το χώρο στον οποίο βρίσκονται οι παρατηρήσεις, διαχωρίζονται:

- από ένα σημείο στον μονοδιάστατο χώρο R^1
- από μία ευθεία γραμμή στο δισδιάστατο χώρο R^2
- από ένα επίπεδο στον τρισδιάστατο R^3
- από ένα υπερεπίπεδο (hyperplane) σε μεγαλύτερες διαστάσεις R^n

Σε κατηγοριοποιημένα δεδομένα η SVM καλείται να δημιουργήσει τα κατάλληλα διαχωριστικά υπερεπίπεδα έτσι ώστε η περιοχή των δεδομένων να διαχωριστεί σε περιοχές ή τμήματα (segments) που περιέχουν μόνο μια κατηγορία δεδομένων. Αυτή η τεχνική είναι ιδιαίτερα χρήσιμη για δεδομένα των οποίων η κατανομή είναι άγνωστη. Μια σύντομη οπτική περιγραφή της μεθόδου παρουσιάζεται στην εικόνα 2.5 όπου παρουσιάζεται μια απλή περίπτωση διαχωρισμού δεδομένων σε δυο κατηγόριες που απεικονίζονται με μπλε και κόκκινες αποχρώσεις. Σε αυτή την ιδανική περίπτωση τα δεδομένα εξάσκησης είναι ξεκάθαρα διαχωρισμένα σε δύο τμήματα και κάθε γραμμή που διαχωρίζει αυτές τις δύο κατηγορίες μπορεί να χρησιμοποιηθεί για την κατηγοριοποίηση (classification) τους. Η γραμμή που θα κατασκευαστεί θα πρέπει να έχει την βέλτιστη απόσταση μεταξύ των δύο κατηγοριών καθώς γραμμές πολύ κοντά στα δεδομένα δεν επιτρέπουν τον διαχωρισμό σημείων που παρεμβάλλονται ή

βρίσκονται πολύ κοντά σε διαφορετική κατηγορία. Η απλούστερη γραμμή περιγράφεται από την εξίσωση y = ax+b και στόχος της μεθόδου είναι να προσδιοριστεί η βέλτιστη ευθεία διαχωρισμού των κλάσεων, για την οποία θα επιτυγχάνεται η ελαχιστοποίηση του σφάλματος κατάταξης. Δηλαδή, να καταταχθεί σωστά, στην κλάση που πραγματικά ανήκει, όσο το δυνατόν μεγαλύτερος αριθμός σημείων. Απαίτηση για την εφαρμογή αυτής της μεθόδου είναι η διαχωριστική ευθεία (που καθορίζει το όριο των κλάσεων) να μην βρίσκεται κοντά στα δεδομένα σημεία των κλάσεων ή ακόμη καλύτερα να ισαπέχει από τα δεδομένα και των δύο κατηγοριών και εκφράζεται με τη συνεχή γραμμή στο δεξί μέρος της εικόνας 2.5 ενώ η περίπτωση των διακεκομμένων παράλληλων γραμμών μπορεί να έχουν την ίδια κλήση με την βέλτιστη ευθεία αλλά επιτρέπουν την σωστή ταξινόμηση τριών σημείων της μπλε ομάδας (άνω διακεκομμένη ευθεία) και ενός σημείου της κόκκινης ομάδας (κάτω διακεκομμένη ευθεία).



Εικόνα 2.5. *Παράδειγμα διαχωρισμού δεομένων με τη μέθοδο SVM (Πηγή: www.r-bloggers.com)*

Για ένα σύνολο n παρατηρήσεων οι οποίες αποτελούν τα δεδομένα εξάσκησης παριστάνουμε κάθε ζεύγος παρατηρήσεων ως (x_i, y_i) όπου i=1…n, και $x_i \in R^n$, $y_i \in \{-1,+1\}$ υπό της υπόθεσης ότι το σύνολο προέρχεται από μία άγνωστη από κοινού συνάρτηση κατανομής (distribution function) P(x,y) και τα δεδομένα είναι ανεξάρτητα και ομοιογενώς κατανεμημένα (independently and identically distributed). Έστω W ένα διάνυσμα βαρών διάσταση n όπου $W=\{w_1,\ldots,wn)$. Τότε το υπερεπίπεδο μπορεί να γραφεί και ως

$$\mathbf{W}\cdot\mathbf{X} + b = 0 \qquad (2.10)$$

όπου **X** ο πίνακας πλειάδων των σημείων x_i που στο συγκεκριμένο παράδειγμα είναι δύο διαστάσεων και περιέχει τα ζεύγη (x_1, x_2) και b ένα βαθμωτό μέγεθος που εφαρμόζεται ως επιπρόσθετο βάρος για την επίλυση της 2.10 που τώρα μετατρέπεται σε

$$w_0 + w_1 \cdot x_1 + w_2 \cdot x_2 = 0 \qquad (2.11)$$

Με αυτόν τον τρόπο για κάθε σημείο που βρίσκεται άνω του υπερεπιπέδου η 2.11 θα είναι μεγαλύτερη του μηδενός ενώ κάτω, μικρότερη του μηδενός. Με τη βοήθεια της σχέσης 2.11. οι πλευρές των περιθωρίων μπορούν να περιγράφουν από τις εξισώσεις (Kecman, 2005)

H_1: $w_0 + w_1 \cdot x_1 + w_2 \cdot x_2 \geq 1$ για $y_i = +1$ (2.12) και

H_2: $w_0 + w_1 \cdot x_1 + w_2 \cdot x_2 \leq -1$ για $y_i = -1$ (2.13)

Κάθε πλειάδα σημείων που περιέχονται στην H_1 ή την H_2 ονομάζονται υποστηρικτικά διανύσματα (support vectors) και το μέγιστο εύρος ή περιθώριο μεταξύ των δύο ομάδων υπολογίζεται από τον τύπο $2/||W||$ όπου $||W|| = \sqrt{\sum_{i=1}^{n} w_i^2}$

Η μέθοδος SVM αποτελεί μια δυνατή και ευρέως διαδομένη τεχνική με ευρύ φάσμα εφαρμογών όπως στην οικονομία, το εμπόριο και την ιατρική. Τα πλεονεκτήματά της μεθόδου SVM είναι:

- Η εξέταση δεομένων άγνωστης κατανομής
- Η ευχέρεια της ερμηνείας των αποτελεσμάτων και μέσω απεικονίσεων
- Ότι αποτελεί μια εύρωστη (robust) μέθοδο που μπορεί να δεχθεί δεδομένα με πολλές παρεμβολές μεταξύ τους (noisy test data) ή μεροληπτικά δεδομένα εξάσκησης (biased train data) αντιμετωπίζοντας με επιτυχία αυτού του είδους τις διενέξεις (incongruencies).

Επειδή η μέθοδος SVM μοιάζει της κατηγορίας των τεχνικών παλινδρόμησης περιέχει το πρόβλημα της υπερβολικής χρήσης παραμέτρων (overfit) το οποίο θεωρείται και ως

το μόνο μειονέκτημα της μεθόδου. Οι επαναλήψεις του αλγορίθμου πρέπει να περιέχουν την συνεχή ρύθμιση (tuning) των επιπέδων ακρίβειας (ε) των παραγόμενων μοντέλων για την αποφυγή σφαλμάτων στα αποτελέσματα.

2.7 Μέτρα απόδοσης

Στην παρούσα ενότητα θα επεξηγηθούν συνοπτικά βασικές έννοιες οι οποίες θα χρησιμοποιηθούν εκτενώς στη συνέχεια, σχετικά με την αξιολόγηση των αποδόσεων των διαφόρων τεχνικών μηχανικής μάθησης. Οι χρησιμοποιούμενες έννοιες ως στατιστικά μέτρα απόδοσης έχουν ως εξής:

2.7.1 Μέτρα που βασίζονται στην μέση διαφορά εκτιμώμενων και πραγματικών τιμών

Αυτή η οικογένεια μέτρων βασίζεται στην εξέταση της διαφοράς $y_i - \widehat{y_i}$ όπου με $\widehat{y_i}$ συμβολίζεται η εκτιμώμενη τιμή της yστην παρατήρηση iκαι με yη πραγματική τιμή της μεταβλητής στην παρατήρηση i. Η κάθε μέθοδος παίρνει την ονομασία της ανάλογα και με την συνάρτηση υπολογισμού της διαφοράς. Πιο συγκεκριμένα οι μέθοδοι αυτής της οικογένειας που θα εξεταστούν είναι:

2.7.1.1 Μέσο Απόλυτο Σφάλμα (Mean Absolute Error)

Το μέτρο αυτό στηρίζεται στην εκτίμηση της τιμής

$$\text{Mean Absolute Error} = \frac{1}{N} \cdot \sum_{j=1}^{N} \left| y_j - \hat{y}_j \right|$$

και επιτρέπει τον υπολογισμό των απόλυτων διαφορών μεταξύ των πραγματικών και εκτιμώμενων. Ένα σημαντικό πλεονέκτημα της μεθόδου είναι ότι μετρά με ακρίβεια τις αποκλίσεις στην πραγματική τους τιμή. Παρόλα αυτά η μέθοδος εκφυλίζεται στην περίπτωση ίσων παρατηρήσεων καθώς επιτρέπει την εμφάνιση της μηδενικής τιμής και στην ουσία μετατρέπεται σε ένα μέτρο επιτυχία/αποτυχίας.

2.7.1.2 Μέσο Τετραγωνικό Σφάλμα (Mean Squared Error)

Παρόμοια με την προηγούμενη μέθοδο, η εξέταση της ακρίβειας σε αυτή την περίπτωση γίνεται με την βοήθεια του τύπου:

$$\text{Mean Squared Error} = \frac{1}{N} \cdot \sum_{j=1}^{N} (y_j - \hat{y}_j)^2$$

Ένα σημαντικό μειονέκτημα αυτής της μεθόδου είναι ότι πολλαπλασιάζει τις περιπτώσεις μεγάλων διαφορών εμφανίζοντας μια ασυνέχεια στην σειρά εμφάνισης των αγαλμάτων, μη επιτρέποντας τις πληροφοριακές συγκρίσεις. Για αυτό τον λόγο, συνήθως χρησιμοποιείται η τετραγωνική ρίζα αυτού του σφάλματος:

$$\text{Mean Root Squared Error} = \sqrt{\frac{1}{N} \cdot \sum_{j=1}^{N} \left(y_j - \hat{y}_j\right)^2} = \sqrt{MSE}$$

Αυτή η οικογένεια μέτρων ακρίβειας εφαρμόζεται σε περιπτώσεις αποτελεσμάτων που προκύπτουν από την εφαρμογή ενός φίλτρου π.χ. γραμμικής παλινδρόμησης ή στην περίπτωση μας, στα Νευρωνικά δίκτυα, αλλά μειονεκτεί στην εξέταση της τάσης των σφαλμάτων καθώς η κλίση της, όπως εκφράζεται από την πρώτη παράγωγο, δεν μπορεί να συλλάβει μεταβαλλόμενες διαφορές και ως εκ τούτου δεν μπορεί να εξετάσει τα αίτια μεταβολής τους.

2.7.2 Μέτρα ακρίβειας που βασίζονται στην πιθανότητα

Σε αυτή την οικογένεια μέτρησης της ακρίβειας των μεθόδων ανήκει και η πλειοψηφία των εξεταζόμενων μεθόδων ακρίβειας. Σκοπός τους είναι η εξέταση της ακρίβειας της μεθόδου με την μέτρηση της πιθανότητας της εμφάνισης σωστών και λανθασμένων τοποθετήσεων του αλγορίθμου.

Πίνακας πιθανοτήτων (Confusion matrix)

Η μέθοδος αυτή χρησιμοποιεί την θεωρία πιθανοτήτων και μέσα από έναν πίνακα πιθανοτήτων εξάγει τα απαραίτητα μέτρα υπολογισμού ακρίβειας. Παρόμοια με την μέθοδο για τον υπολογισμό του odds ratio κατασκευάζεται ένας πίνακας ενδεχομένων στον οποίο η κατάταξη (κατηγορία) ενδιαφέροντος είναι η κατηγορία Β και η κατηγορία ελέγχου η Α.

Πίνακας 2.1. Πίνακας πιθανοτήτων για την μέτρηση της ακρίβεια μεθόδων ταξινόμησης

		Εκτίμηση κατηγορίας	
		A	B
Πραγματική	A	TN	FP
Κατάταξη	B	FN	TP

Ανάλογα με τη σχέση των αποτελεσμάτων και της πραγματικής κατάστασης του ατόμου μπορεί να ληφθούν τα εξής αποτελέσματα:

- **Αληθώς θετικό (true positive-TP)**: Ασθενής αναγνωρίστηκε σωστά ως ασθενής.
- **Ψευδώς θετικό (false positive-FP)**: Υγιής αναγνωρίστηκε λανθασμένα ως ασθενής.
- **Αληθώς αρνητικό (true negative-TN)**: Υγιής αναγνωρίστηκε σωστά ως υγιής.
- **Ψευδώς αρνητικό (false negative-FN)**: Ασθενής αναγνωρίστηκε λανθασμένα ως υγιής.

2.7.3 Accuracy

Η ακρίβεια ή Accuracy υπολογίζεται ως το άθροισμα των αληθών μερών του πίνακα πιθανοτήτων δηλαδή το άθροισμα των TP και TN προς το σύνολο των εξεταζόμενων περιπτώσεων. Μαθηματικά εκφράζεται από τον τύπο:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Άλλα μέτρα ενδιαφέροντος είναι το precision που εκφράζει την ακρίβεια μιας μεθόδου-ενός αλγορίθμου με την βοήθεια ψευδώς θετικών περιπτώσεων:

$$precission = \frac{TP}{TP + FP}$$

και το recall που εκφράζει το μέτρο της ακρίβειας με τον υπολογισμό των περιπτώσεων ψευδώς αρνητικών περιπτώσεων:

$$recall = \frac{TP}{TP + FN}$$

Κεφάλαιο 3

Επισκόπηση μελετών

Αυτή η βιβλιογραφική ανασκόπηση περιγράφει παρόμοιες έρευνες που εφαρμόστηκαν την τελευταία δεκαετία. Αυτή η ανασκόπηση ολοκληρώνεται με μια συζήτηση για το πώς αυτή εργασία βοηθά στην τρέχουσα γνώση όπως και η σημαντικότητα σχετικά με την πρόβλεψη της μελλοντικής θέσης ενός πλοίου και την ανίχνευση ανωμαλιών χρησιμοποιώντας δεδομένα AIS. Αρκετές μέθοδοι έχουν αναπτυχθεί την τελευταία δεκαετία για την αντιμετώπιση του προβλήματος της πρόβλεψης θέσης στον τομέα της ναυτιλίας. Οι τρέχουσες μεθοδολογίες για την πρόβλεψη των τροχιών των σκαφών χωρίζονται σε τρεις κατηγορίες ανάλογα με τον τρόπο εφαρμογής τους: μεθόδους που βασίζονται σε κλασικό μοντέλο, μεθόδους βασισμένες σε μοντέλα μηχανικής μάθησης και υβριδικά μοντέλα (Tu, Zhang, Rachmawati, Rajabally, & Huang, 2016). Τα κλασικά μοντέλα εξετάζουν όλους τους πιθανούς παράγοντες που επηρεάζουν και χρησιμοποιούν φυσικούς νόμους κίνησης για να προβλέψουν τη μελλοντική τροχιά ενός σκάφους. Ωστόσο, αυτή η μέθοδος χρησιμοποιείται κυρίως για προσομοιώσεις. Τα μοντέλα μηχανικής εκμάθησης χρησιμοποιούν ιστορικά δεδομένα AIS για να την ανάπτυξη μοντέλων κίνησης. Τέλος, τα υβριδικά μοντέλα μπορεί να περιλαμβάνουν τόσο συστατικά ενός κλασικού μοντέλου όσο και ιστορικά δεδομένα κίνησης.

Οι Khan, Cees και Kaye το (2005) χρησιμοποιήσαν ένα εκπαιδευμένο νευρωνικό δίκτυο προώθησης πολλαπλών επιπέδων (MLFF) με τη χρήση αποσύνθεσης μοναδικής αξίας και γενετικούς αλγόριθμους για να προβλέψουν τη γωνία του πλοίου (βήμα) έως και 160 δευτερόλεπτα. Οι συγγραφείς παραθέτουν ανεπάρκειες άλλων μεθόδων, όπως μοντέλα ARMA(p,q) και φίλτρο Kalman για τον υπολογισμό ακριβών βραχυπρόθεσμων εκτιμήσεων της θέσης ενός πλοίου σε ταραχώδεις θαλάσσιες διαδρομές, ώστε να καταστεί δυνατή η ασφαλέστερη προσγείωση αεροσκαφών σε πλοίο.

Οι Palacios και Doshi (2008) χρησιμοποιούν ένα νευρωνικό δίκτυο για να προβλέψουν τη μελλοντική θέση ενός αεροσκάφους χρησιμοποιώντας δύο διαφορετικές προσεγγίσεις. Στην προσέγγισή τους μέσω σημείων του χώρου R^2 (Σημ.X-Y), ο ίδιος τύπος νευρωνικού δικτύου εφαρμόζεται δύο φορές, μία για την πρόβλεψη της

μελλοντικής συντεταγμένης γεωγραφικού μήκους και μία για την πρόβλεψη της συντεταγμένης γεωγραφικού πλάτους. Για να εφαρμόσουν αυτήν τη μέθοδο, επιλέγουν να χρησιμοποιήσουν την απόσταση που διανύθηκε τα τελευταία δευτερόλεπτα για να προβλέψουν τα επόμενα τριάντα δευτερόλεπτα στο της; κίνησης. Η δεύτερη μέθοδος τους, που ονομάζεται προσέγγιση κατεύθυνσης/απόστασης (bearing/distance), βασίζεται στην εκτίμηση της κατεύθυνσης της κίνησης και της απόστασης που θα διανύσει το αεροσκάφος. Στη συνέχεια υπολογίζουν τη μελλοντική θέση χρησιμοποιώντας τριγωνομετρία. Οι συγγραφείς διαπίστωσαν ότι η προσέγγιση της κατεύθυνσης ήταν 5 % έως 10 % λιγότερο ακριβής από την προσέγγιση X-Y.

Οι Morris και Trivedi (2008) αντιπροσωπεύουν την εκμάθηση διαδρομών μεταξύ διαφορετικών σημείων ενδιαφέροντος (Points of Interest ή POI) στη καταγραφή εικόνας ως διαδικασία τριών βημάτων. Τα μέρη που δημιουργούνται από κινούμενα αντικείμενα υποβάλλονται σε προεπεξεργασία χρησιμοποιώντας μια μορφή μείωσης διαστάσεων, όπως γραμμική παρεμβολή για να τοποθετηθούν τα ίχνη σε συγκρίσιμη μορφή. Δεύτερον, τα μέρη ομαδοποιούνται για να αντιπροσωπεύουν διαφορετικές διαδρομές που αποτελούνται από παρόμοιες τροχιές. Τέλος, οι διαδρομές διαμορφώνονται είτε χρησιμοποιώντας ολόκληρη τη διαδρομή είτε διαχωρίζοντας τη διαδρομή σε τμήματα. Στο τελευταίο βήμα, συνοψίζουν τη μέθοδο απόστασης από κεντρικό σημείο αναφοράς για τον προσδιορισμό της ελάχιστης διαδρομής καθώς και των αποκλίσεων της.

Οι Ristic, La Scala, Morelande και Gordon (2008) χρησιμοποιούν ιστορικά δεδομένα AIS για να εξάγουν μοτίβα κίνησης τα οποία στη συνέχεια χρησιμοποιούνται για την κατασκευή μεθόδων ανίχνευσης ανωμαλιών κίνησης. Στη συνέχεια, χρησιμοποιούν τις μεθόδους ανίχνευσης ανωμαλιών διαδοχικά στα εισερχόμενα δεδομένα AIS για τον εντοπισμό ανωμαλιών υπό την μηδενική υπόθεση ότι δεν υπάρχει ανωμαλία. Επιπλέον, οι συγγραφείς χρησιμοποιούν ιστορικά δεδομένα κίνησης για να προβλέψουν την κίνηση των αγγείων χρησιμοποιώντας την μέθοδο CUMSUM.

Ο Zhu (2011) διακρίνει μια περιοχή ενδιαφέροντος χρησιμοποιώντας κώδικες κατακερματισμού (hash codes) και χρησιμοποιεί κανόνες συσχέτισης για να εξάγει πληροφόρηση από δίκτυα που χρησιμοποιούνται σε μεγάλη συχνότητα. Ο συγγραφέας

χρησιμοποιεί επίσης τεχνικές εξέτασης σχέσεων (εξέταση συσχετίσεων) για να αποφανθεί εάν υπάρχουν σχέσεις μεταξύ των εξεταζόμενων σημείων .

Οι Morris και Trivedi (2011) χρησιμοποιούν τη μοντελοποίηση υπό Gauss για να βρουν σημεία πορείας (σημεία εισόδου, εξόδου, μεταβολής της κίνησης και εξόδου από την κίνηση) και χρησιμοποιούν την ομαδοποίηση των δεδομένων τροχιάς (trajectory clustering) για να εξάγουν τις περιεχόμενες διαδρομές και χρησιμοποιούν Μαρκοβιανά (Markov models) μοντέλα για να προβλέψουν τη μελλοντική τοποθεσία των οχημάτων που κινούνται σε διασταυρώσεις όπως και να διαγνώσουν αποκλίσεις από την διαδρομή.

Οι Vespe, Visentini, Bryan και Braca (2012) υποδεικνύουν ότι η συμπεριφορά της κίνησης των πλοίων είναι ένα δίκτυο σημείων πορείας (σημεία εισόδου, εξόδου, μεταβολής της κίνησης και εξόδου από την κίνηση). Ορίζουν μια διαδρομή ως μια ακολουθία θαλασσίων διαδρομών που η καθεμία χαρακτηρίζεται από συγκεκριμένα μέτρα ιδιότητες, όπως Course Over Ground (COG), και Speed Over Ground (SOG). Τέλος, οι συγγραφείς καταδεικνύουν την αποτελεσματικότητα της ταυτοποίησης των σημείων τροχιάς κάνοντας σύγκριση μεταξύ δύο περιοχών, μιας με υψηλή επίγεια κάλυψη σημάτων AIS (Αδριατική θάλασσα) και μια περιοχή με χαμηλότερη επίγεια κάλυψη σημάτων AIS (Ερυθρά Θάλασσα και Κόλπος του Άντεν).

Οι Pallotta, Vespe και Bryan (2013) αναφέρουν ότι η χρήση των σημείων στροφής σε μοντέλο δικτύου δεν λειτουργεί καλά σε περιοχές όπου η κυκλοφορία δεν είναι ρυθμισμένη. Έτσι προτείνουν τον αλγόριθμο DBSCAN για την εξαγωγή στατικών περιοχών όπως και σημείων εισόδου και εξόδου και υπολογίζουν την κίνηση με βάση τα σημεία αυτά και τα διανύσματα κίνησης του σκάφους που ενσωματώνουν και τη στροφή του. Οι διαδρομές δημιουργούνται με ομαδοποίηση των ροών μεταξύ των σημείων ενδιαφέροντος. Επιπρόσθετα, προτείνουν μια μέθοδο πρόβλεψης της μελλοντικής θέσης ενός σκάφους με βάση μια ακολουθία κύκλων καθορισμένης από τον χρήστη ακτίνας με επίκεντρο τις παρατηρούμενες θέσεις. Οι συγγραφείς αναφέρουν ότι ένα μειονέκτημα στη χρήση της μεθόδου αυτής είναι ότι η επιλεγμένη ακτίνα d θα μπορούσε να είναι πολύ μικρή για τη διαδρομή με αποτέλεσμα ο χαρακτηρισμός της συμπεριφοράς της τοπικής διαδρομής να βασίζεται σε μειωμένο

αριθμό γειτονικών σημείων (neighbors). Εάν η ακτίνα d είναι πολύ μεγάλη, τότε ο χαρακτηρισμός θα έχει μεροληπτικά σφάλματα λόγω των διαδρομών που δεν είναι ευθείες. Τέλος, ισχυρίζονται ότι μια ακτίνα «της τάξης μερικών ναυτικών μιλίων» είναι αποτελεσματική για οποιαδήποτε διαδρομή.

Ο McAbee (2013) χρησιμοποιεί τον μετασχηματισμό Hough για να εξαγάγει γραμμικά πρότυπα κίνησης από τα δεδομένα AIS και για να δημιουργήσει θαλάσσιες διαδρομές τόσο σε ανοιχτούς ωκεανούς όσο και σε παράκτιες περιοχές. Μόλις καθοριστούν οι θαλάσσιες διαδρομές, χωρίζονται σε τμήματα κατά την κατεύθυνση της διαδρομής. Μια κανονική κατανομή είναι κατάλληλη για κάθε τμήμα της λωρίδας ώστε να λαμβάνεται υπόψη η ετεροσκεδαστικότητα και τα σκάφη που παρατηρούνται εκτός ενός προκαθορισμένου ορίου να ορίζονται ως αποκλίνοντα (ανώμαλα) από τον χρήστη.

Ο Tester (2013) χρησιμοποιεί ομαδοποίηση την μέθοδος k mean clustering για να ομαδοποιήσει πλοία με παρόμοια πορεία και ταχύτητα για να ταξινομήσει την κίνηση των πλοίων. Στη συνέχεια παρακολούθησε τον τρόπο δημιουργίας των ομάδων (clusters) συγκρίνοντας την απόσταση μεταξύ τους με την πάροδο του χρόνου. Ωστόσο, η μέθοδος αυτή δεν μπορεί να προβλέψει τη μελλοντική τοποθεσία του πλοίου.

Οι Stone, Streit, Corwin και Bell (2014) απεικονίζουν την παρακολούθηση ενός πλοίου χρησιμοποιώντας το φίλτρο Kalman. Καθόρισαν ένα μοντέλο κίνησης χρησιμοποιώντας την διαδικασία Integrated Ornstein- Uhlenbeck (IOU). Στη συνέχεια, οι καθόρισαν ένα μοντέλο μέτρησης υποθέτοντας ότι η σχέση μεταξύ της μέτρησης και της θέσης είναι γραμμική. Οι συγγραφείς καταλήγουν στο συμπέρασμα ότι αυτές οι υποθέσεις είναι οι βέλτιστες για ένα φίλτρο Kalman, αλλά ένα μοντέλο κίνησης που βασίζεται στη διαδικασία Ornstein-Uhlenbeck «δεν αποτελεί καλή αναπαράσταση της πραγματικής κίνησης των πλοίων» (σελ. 10).

Οι Pallota, Horn, Braca και Bryan (2014) παρουσιάζουν μια μέθοδο πρόβλεψης της μελλοντικής θέσης ενός σκάφους με βάση τη στοχαστική διαδικασία Ornstein-Uhlenbeck (OU). Οι παράμετροι του μοντέλου υπολογίζονται από επαναλαμβανόμενα μοτίβα διαδρομών που περιέχονται στα δεδομένα AIS, όπου οι διαδρομές είναι τα τόξα

μεταξύ των σημείων ενδιαφέροντος. Πρώτον, οι συγγραφείς υποθέτουν ότι ένα σκάφος έχει ταξινομηθεί σωστά σε μια συγκεκριμένη διαδρομή. Στη συνέχεια, υποθέτουν ότι η δυναμική του σκάφους αντιπροσωπεύεται από ένα σύνολο γραμμικών διαφορικών εξισώσεων και υπολογίζουν τρεις διαφορετικές παράμετροι που χαρακτηρίζουν τις ιδιότητες της διαδρομής. Το βασικό όφελος της μεθόδου OU είναι ότι η διακύμανση της θέσης του σκάφους αυξάνεται γραμμικά με το χρόνο σε αντίθεση με έναν υψηλότερο μη γραμμικό ρυθμό σε άλλα προηγούμενα μοντέλα. Τα δεδομένα μετατρέπονται από το γεωγραφικό πλάτος και το γεωγραφικό μήκος σε Μερκατοριανές συντεταγμένες (Universal Transverse Mercator - UTM). Οι συγγραφείς παρουσιάζουν αποτελέσματα για τρεις περιπτώσεις για τις οποίες η τυπική απόκλιση σφάλματος πρόβλεψης είναι της τάξης των 1000 μέτρων σε χρονικό διάστημα πρόβλεψης πέντε ωρών. Οι Millifiori, Braca, Bryan και Willett (2016) συνεχίζουν την προηγούμενη εργασία των Pallotta et al. (2013) με επίκεντρο τα σκάφη που ταξιδεύουν χωρίς ελιγμούς όπως μπορεί να συμβούν σε ανοιχτή θάλασσα. Οι συγγραφείς διαπιστώνουν ότι το μοντέλο σχεδόν σταθερής ταχύτητας (Near Constant Velocity - NCV) μπορεί να μην είναι ρεαλιστικό για τα περισσότερα σενάρια κυκλοφορίας πλοίων, καθώς οι χειριστές πλοίων μεταβάλλουν συχνά την ταχύτητα. Επιπλέον, παρουσιάζουν στοιχεία που υποδηλώνουν ότι η ταχύτητα του σκάφους χωρίς ελιγμούς ακολουθεί μια διαδικασία Ornstein-Uhlenbeck (OU), και κατά συνέπεια η θέση του σκάφους αντιπροσωπεύεται από μια ολοκληρωμένη διαδικασία OU (Integrated OU). Τα αποτελέσματά τους δείχνουν ότι η τυπική απόκλιση του σφάλματος πρόβλεψης κατά μήκος των συντεταγμένων x και y είναι της τάξης των 3 km μετά από πέντε ώρες.

Οι Mao et al. (2016) χρησιμοποιήσαν ακραία μηχανική μάθηση (Extreme Learning Machine - ELM) για την πρόβλεψη της μελλοντική τοποθεσία ενός πλοίου με βάση δεδομένα AIS στα ανοικτά των ακτών της Καλιφόρνια. Αφού επιλέξουν μια διαδρομή, χρησιμοποιούν το γεωγραφικό πλάτος, το γεωγραφικό μήκος, το SOG, το COG, το Rate of Turn (ROT), το χρόνο και την ταυτότητα του σκάφους ως δεδομένα. Οι συγγραφείς υπολόγισαν ότι η κατανομή των σφαλμάτων μέτρησης κυμαίνονταν μεταξύ 0 και 2,5 ναυτικών μιλίων για μια πρόβλεψη 20 λεπτών και μεταξύ 0 έως 6 ναυτικών μιλίων για μια πρόβλεψη 40 λεπτών.

Οι Tu et al. (2016) περιγράφουν τρεις από τις πιο συχνά χρησιμοποιούμενες μεθόδους μοντελοποίησης που χρησιμοποιούνται στην πρόβλεψη θέσης, συμπεριλαμβανομένων κλασσικών μοντέλων, μοντέλων μηχανικής μάθησης και υβριδικών μοντέλων. Οι συγγραφείς σημειώνουν ότι τα κλασσικά μοντέλα μπορεί να είναι πρακτικά για εφαρμογή σε μεμονωμένο πλοίο ή σε προσομοίωση, αν και οι λεπτομερείς πληροφορίες που απαιτούνται για την προσαρμογή αυτών των μοντέλων δεν είναι πιθανό να είναι διαθέσιμες για άλλα σκάφη, όπως συμβαίνει με το σύνολο δεδομένων AIS. Σημειώνουν ότι η προσέγγιση νευρωνικού δικτύου είναι ιδιαίτερα καλή στην προσαρμογή σύνθετων λειτουργιών, αλλά η διαδικασία μάθησης μπορεί να αργήσει να συγκλίνει και δεν υπάρχουν γενικοί κανόνες για τον τρόπο επιλογής του αριθμού των κρυφών στρωμάτων (hidden layers), του αριθμού των νευρώνων (Neurons) ή της συνάρτησης ενεργοποίησης (Activation Function). Οι συγγραφείς αναφέρουν επίσης τη χρήση διαδικασιών OU στις οποίες υποτίθεται η υπόθεση της στασιμότητας (καμία αλλαγή στη μέση τιμή ή στην διακύμανση του διανύσματος τροχιάς) και σημειώνουν ότι αυτή είναι μια περιοριστική υπόθεση για εφαρμογές σε πραγματικό κόσμο. Οι συγγραφείς αναφέρουν ότι υπάρχει δυνητικό όφελος από τον συνδυασμό κλασσικών μοντέλων με μοντέλα μηχανικής μάθησης για την επίτευξη καλύτερων αποτελεσμάτων πρόβλεψης.

Ο Bay (2017) χρησιμοποιεί δεδομένα AIS στο Port Fourchon, LA για να εξετάσει την αποτελεσματικότητα της ομαδοποίησης για τον εντοπισμό διαδρομών πλοήγησης στο βόρειο κόλπο του Μεξικού και να μετρήσει τις επιπτώσεις του καιρού και της θαλάσσιας κατάστασης στην πλοήγηση. Ο Κόλπος του Μεξικού κοντά στο Port Fourchon έχει πολλές πλατφόρμες πετρελαίου και φυσικού αερίου που εξυπηρετούνται από σκάφη που εδρεύουν στο Port Fourchon, γεγονός που καθιστά δύσκολο τον διαχωρισμό των γραμμών σε μικρό αριθμό ομάδων. Η έρευνά του στοχεύει στον εντοπισμό παραγόντων που θα μπορούσαν να είναι χρήσιμοι για τη δημιουργία καλύτερων μοντέλων πρόβλεψης. Οι μελέτες που εξετάστηκαν χρησιμοποιούν ομαδοποίηση παρόμοιων τροχιών για τον εντοπισμό θαλάσσιων διαδρομών. Χρησιμοποιούν τη διασπορά των θέσεων των σκαφών εντός αυτών των διαδρομών για την ανάπτυξη περιοχών πρόβλεψης για μελλοντική κίνηση των σκαφών. Τέλος, Μερικοί συγγραφείς χρησιμοποιούν μοντέλα νευρωνικών δικτύων σε πλαίσιο χρονικής σειράς για το σκοπό αυτό.

Κεφάλαιο 4

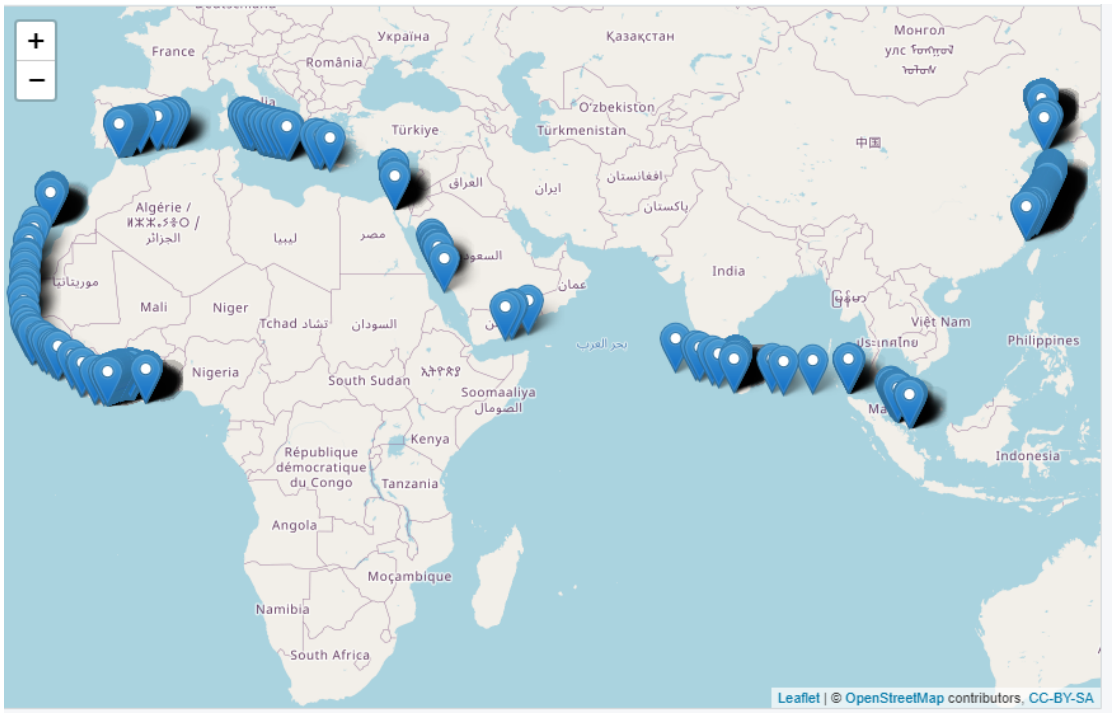
Εφαρμογή και αποτελέσματα

4.1. Εισαγωγή

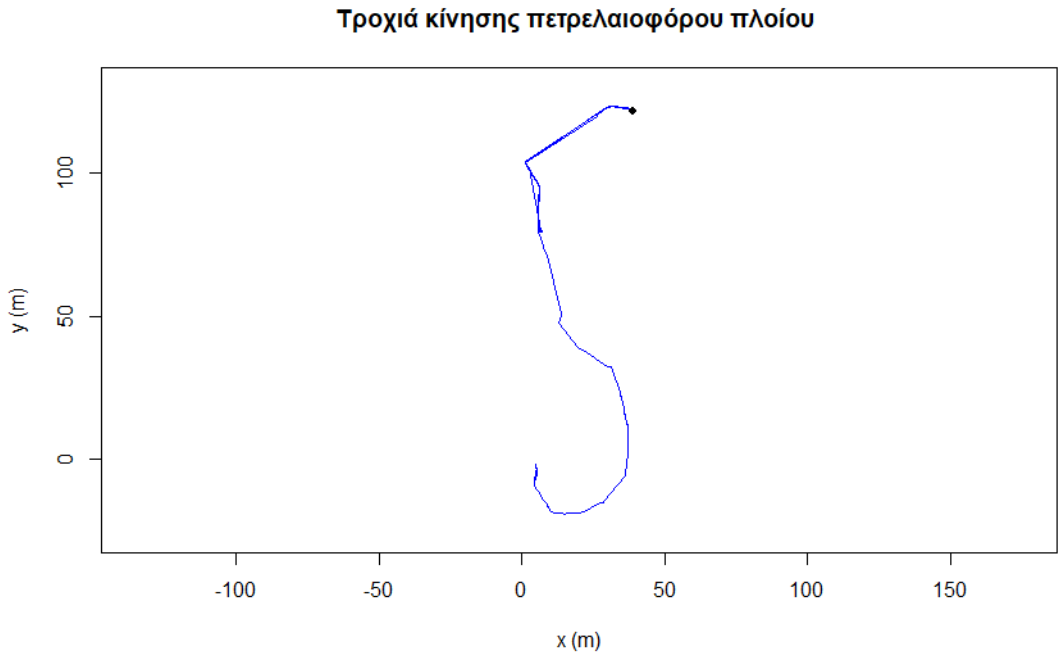
Σε αυτό το κεφάλαιο γίνεται η εφαρμογή των μεθόδων που περιεγράφηκαν στο κεφάλαιο 2. Επίσης γίνεται η παρουσίαση των δεδομένων που χρησιμοποιήθηκαν όπως και η παρουσίαση των αποτελεσμάτων ακρίβειας της εφαρμογής τους. Για την εφαρμογή των μεθόδων και την εξέταση της ακρίβειας τους επιλέχθηκαν 5 διαφορετικοί τύποι πλοίων. Αυτοί ήταν ένα Δεξαμενόπλοιο (Tanker), ένα Εμπορικό (Cargo), ένα Επιβατηγό (Transport), ένα αλιευτικό ανοικτής θαλάσσης και ένα παράκτιο αλιευτικό (πίνακας 4.1). Η επιλογή αυτή δεν έγινε τυχαία καθώς κάθε ένας τύπος αντιπροσωπεύει και μια διαφορετική κίνηση. Πιο συγκεκριμένα, το tanker ως ποντοπόρο πλοίο αναμένεται να διανύσει μεγαλύτερη απόσταση (γράφημα 4.1) η οποία χαρακτηρίζεται από μια σταθερή πορεία χωρίς αποκλίσεις (γράφημα 4.1.Α). Παρόλα αυτά η λείανση της τροχιάς του αναδεικνύει την ύπαρξη περιθωρίων βελτίωσης της πορείας του και ειδικότερα κατά την πορεία του μετά την Μαλαισία προς την Κίνα (γράφημα 4.1Β).

Πίνακας 4.1. Περιγραφή δεδομένων

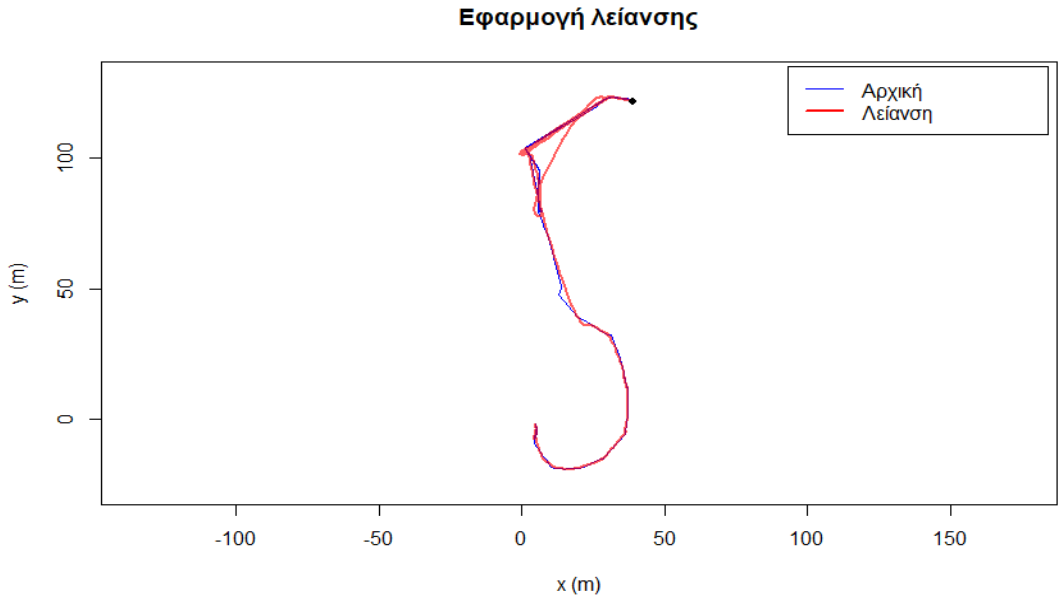
Αναγνωριστικό AIS	**Πλήθος σημείων**	**Τύπος σκάφους**	**Απόσταση σε km**
246929977513855	4.729	Tanker	13.511
44081436890597	14.904	Cargo	4.862
260142468095943	2.760	Transport	1.052
259600076146315	1.896	Αλιευτικό ανοιχτής θαλάσσης	1.503
187372144064677	10.396	Παράκτιο Αλιευτικό	48



Γράφημα 4.1. *Πορεία πετρελαιοφόρου πλοίου*

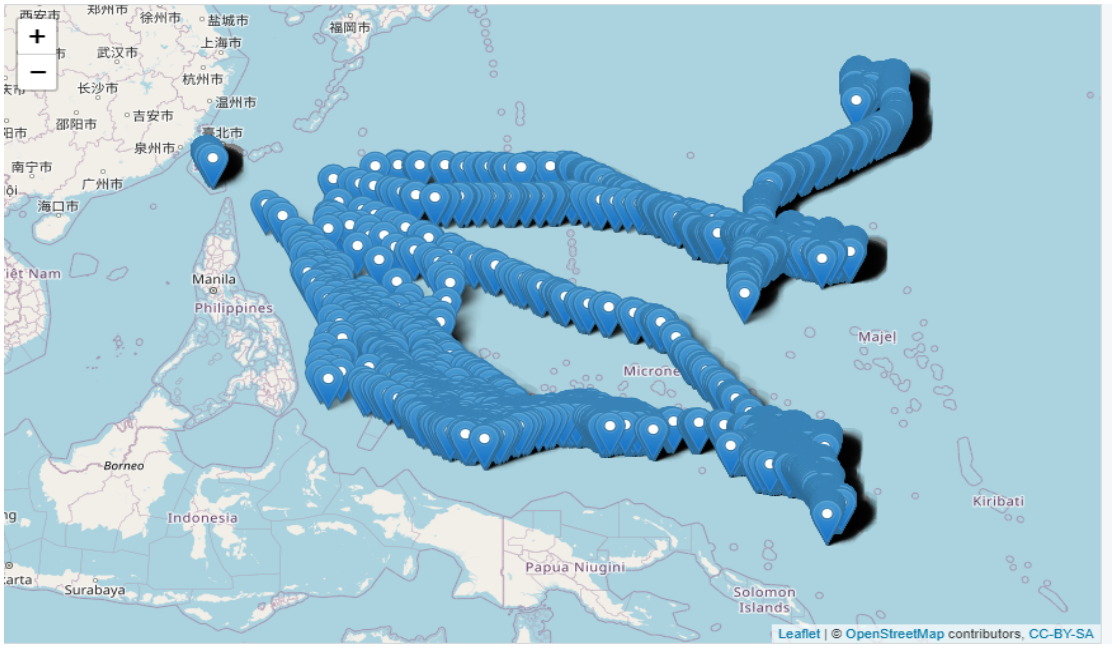


Γράφημα 4.1Α. *Δεδομένα τροχιάς πετρελαιοφόρου πλοίου*

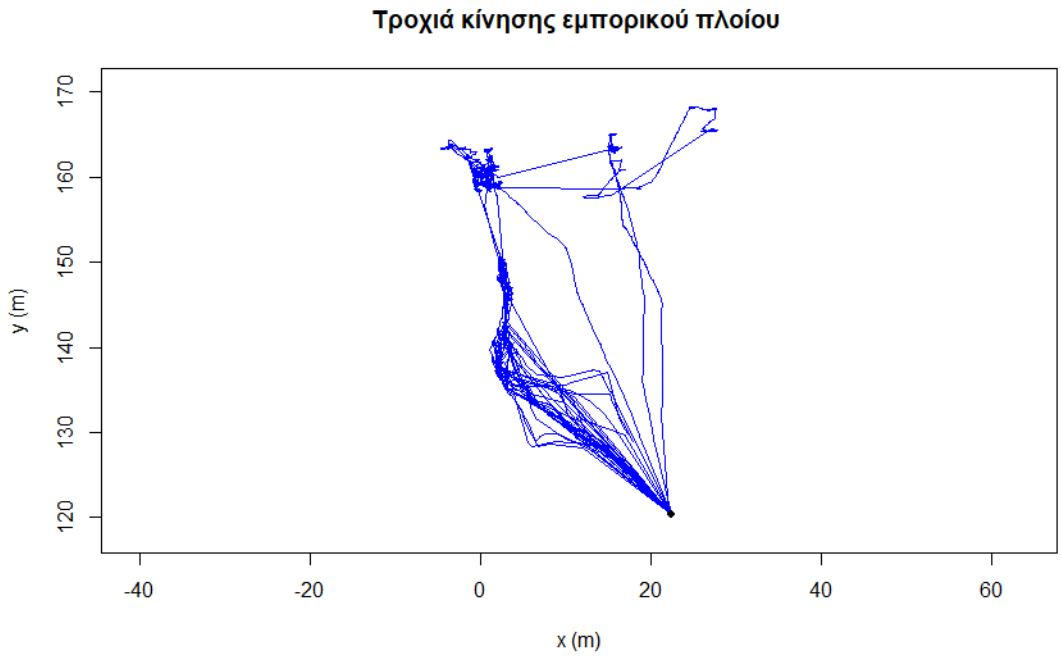


Γράφημα 4.1Β. *Λείανση δεδομένων τροχιάς πετρελαιοφόρου πλοίου*

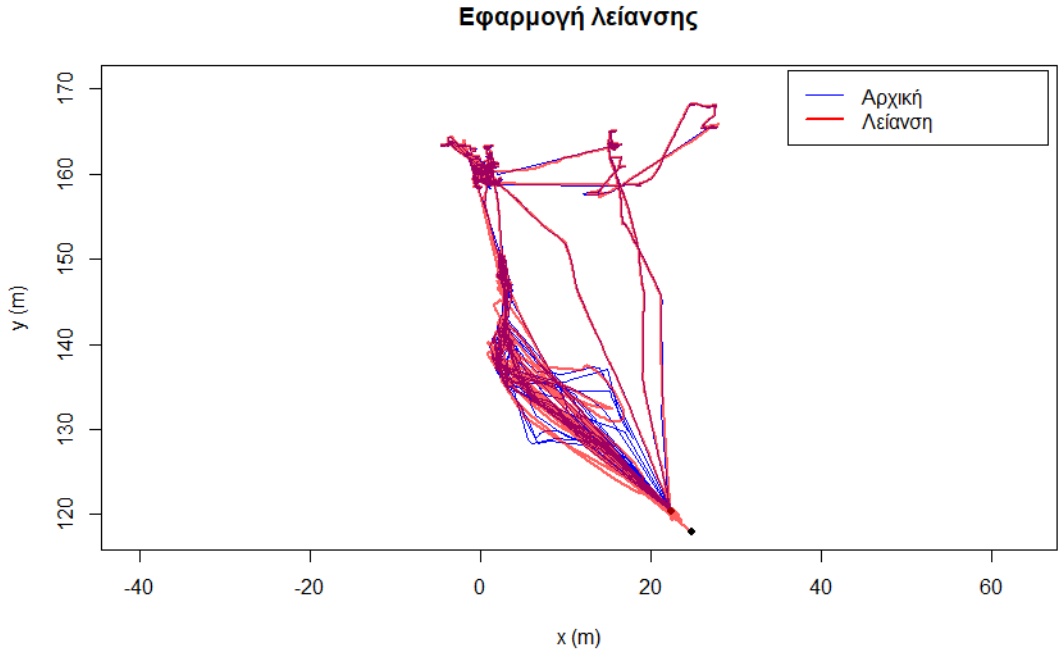
Η πορεία του εμπορικού πλοίου χαρακτηρίζεται από συνεχείς μετακινήσεις μεταξύ πολλών λιμανιών (γράφημα 4.2). Αν και η απόσταση αρχικού-τελικού σημείου αναφοράς είναι ίση με 4.862 χιλιόμετρα, η πραγματική απόσταση που διένυσε είναι αρκετά μεγαλύτερη όπως μπορεί να διαπιστωθεί και από το γράφημα 4.2Α ενώ επιπλέον από το γράφημα 4.2Β παρατηρείτε ότι η κύρια μετακίνηση του πλοίου γίνεται μεταξύ συγκεκριμένων λιμανιών της Πολυνησίας.



Γράφημα 4.2. *Πορεία εμπορικού πλοίου*



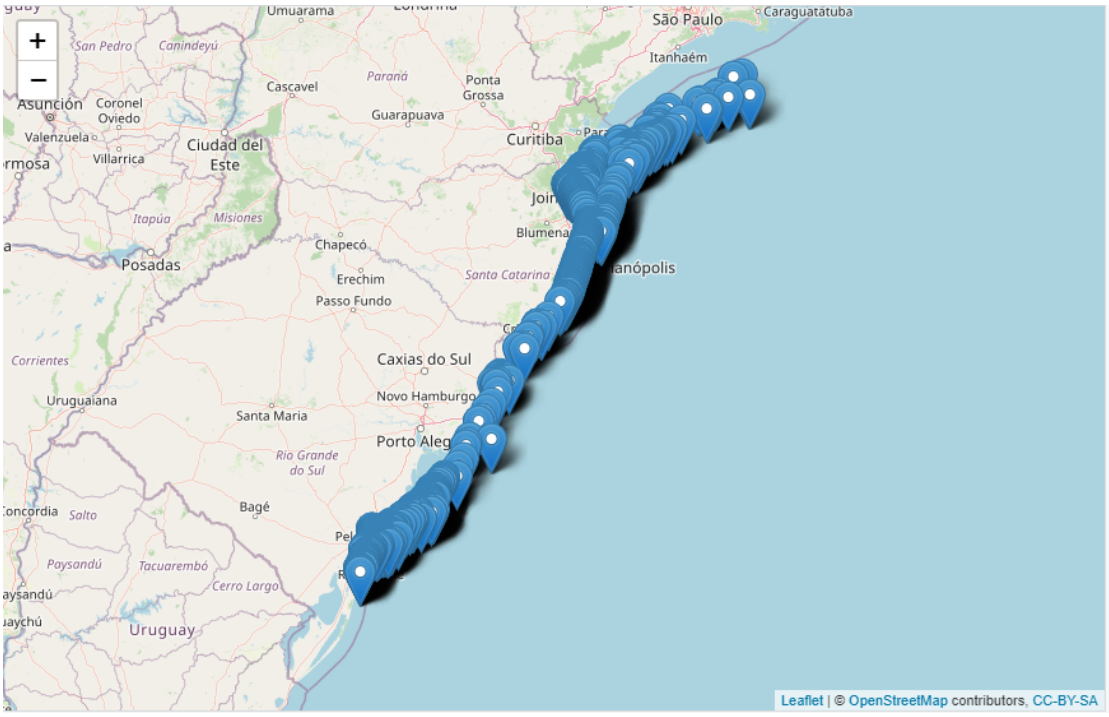
Γράφημα 4.2Α. *Δεδομένα τροχιάς εμπορικού πλοίου*



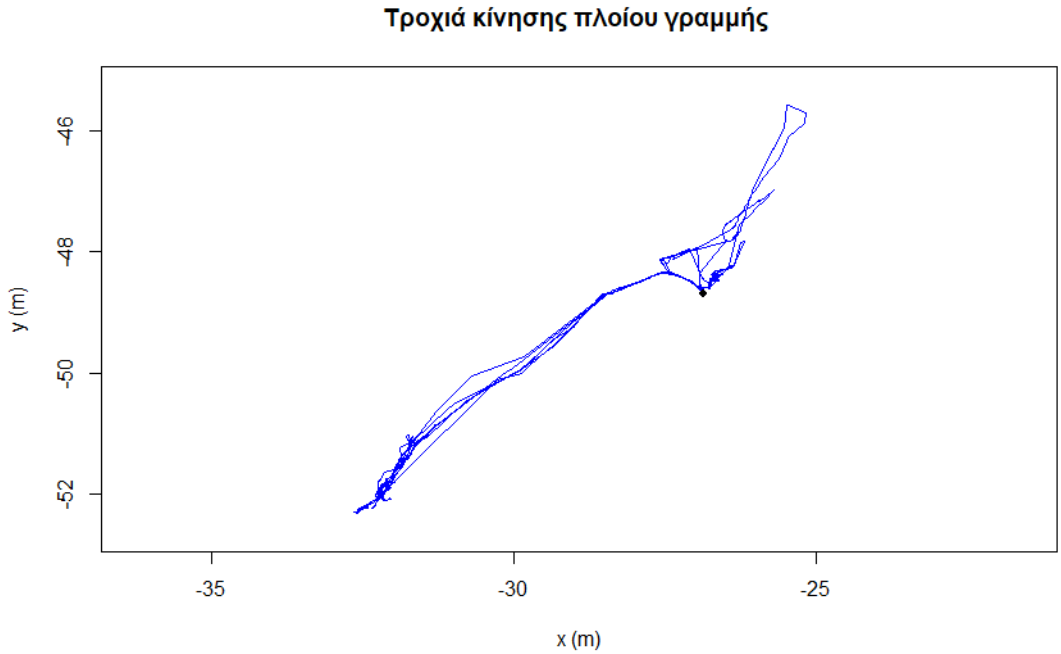
Γράφημα 4.2Β. *Λείανση δεδομένων τροχιάς εμπορικού πλοίου*

Η πορεία του επιβατηγού πλοίου (γράφημα 4.3) περιέχει αρκετά λιγότερες μεταβολές όπως είναι φανερό και από το γράφημα 4.3Α. Η πορεία του είναι τόσο σταθερή ώστε

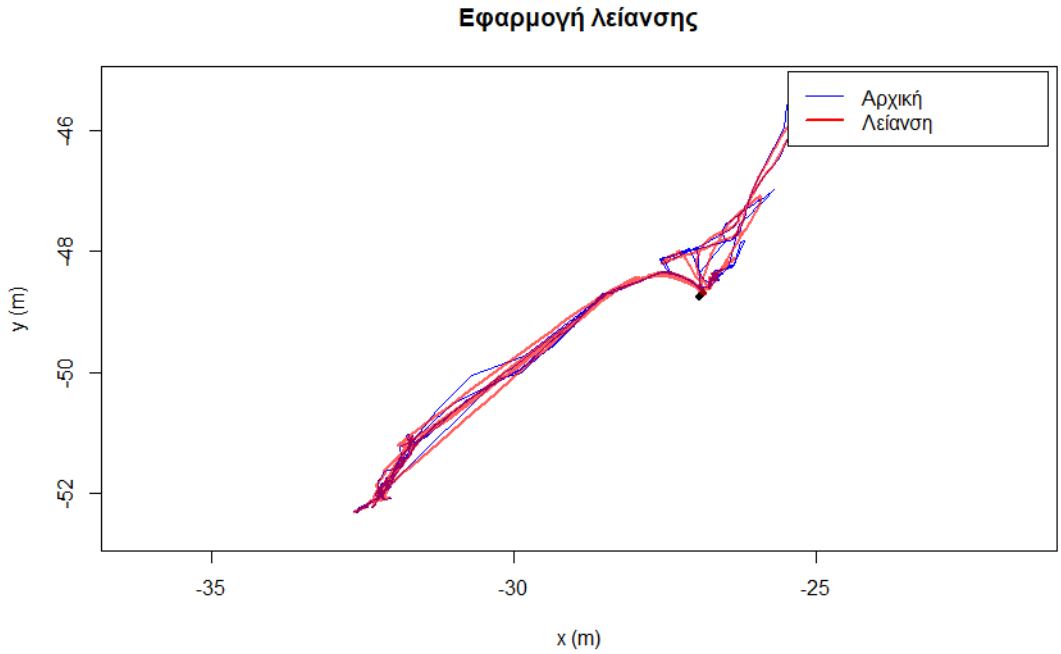
και η λείανση δεν παρουσιάζει καμία ορατή αλλαγή από την αρχική του πορεία (γράφημα 4.3Β).



Γράφημα 4.3. *Πορεία πλοίου γραμμής*



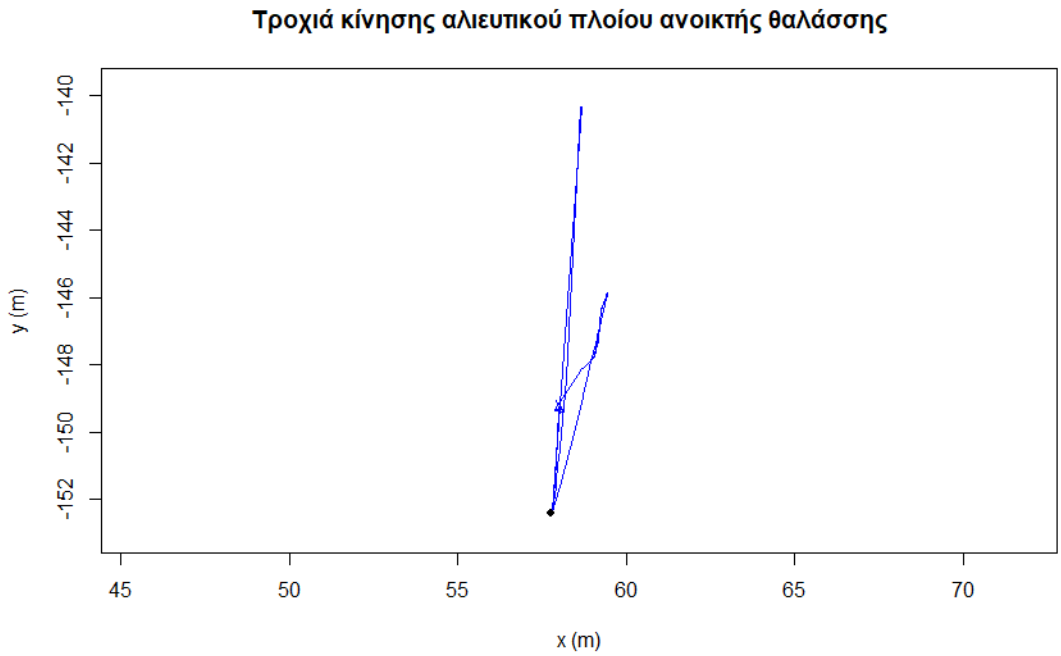
Γράφημα 4.3Α. *Δεδομένα τροχιάς πλοίου γραμμής*



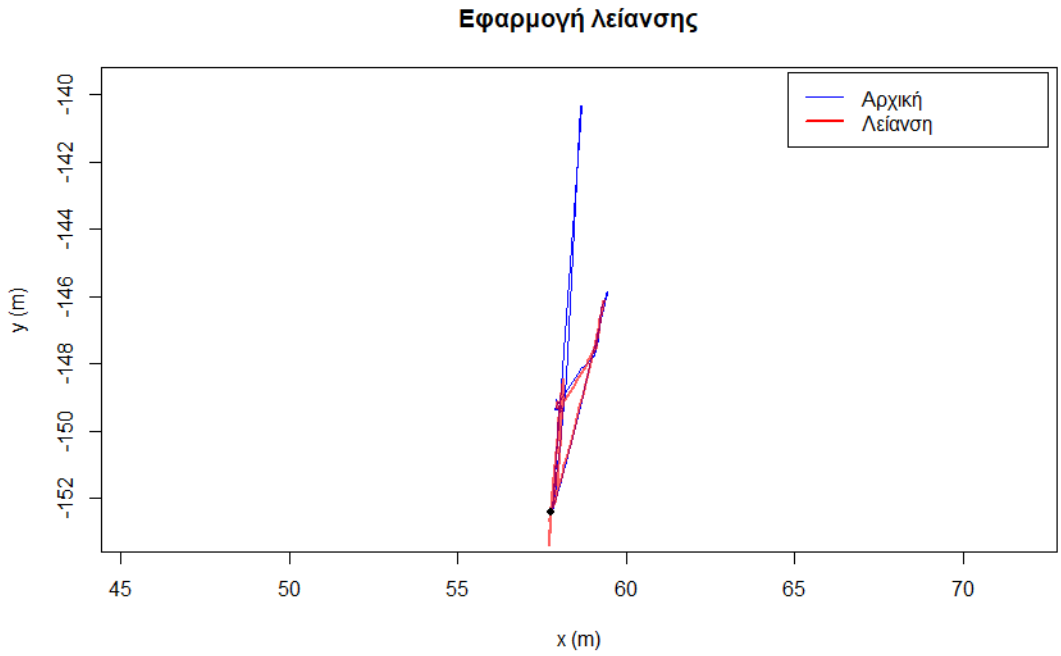
Γράφημα 4.3Β. *Λείανση δεδομένων τροχιάς πλοίου γραμμής*

Οι μεταβολές της κίνησης του αλιευτικού ανοικτής θαλάσσης ανοικτά του Άνκορατζ στην Αλάσκα (γράφημα 4.4) χαρακτηρίζεται από μηδενικές ταχύτητες και μια τριγωνική πορεία (γράφημα 4.4Α) η οποία γίνεται πιο εμφανής στο γράφημα 4.4Β. Η κίνηση αυτή δείχνει ότι το αλιευτικό ακολουθεί κάτι συγκεκριμένο π.χ. κοπάδι ψαριών και δεν έχει ορίσει μια προκαθορισμένη πορεία.

Γράφημα 4.4. *Πορεία αλιευτικού πλοίου ανοικτής θαλάσσης*



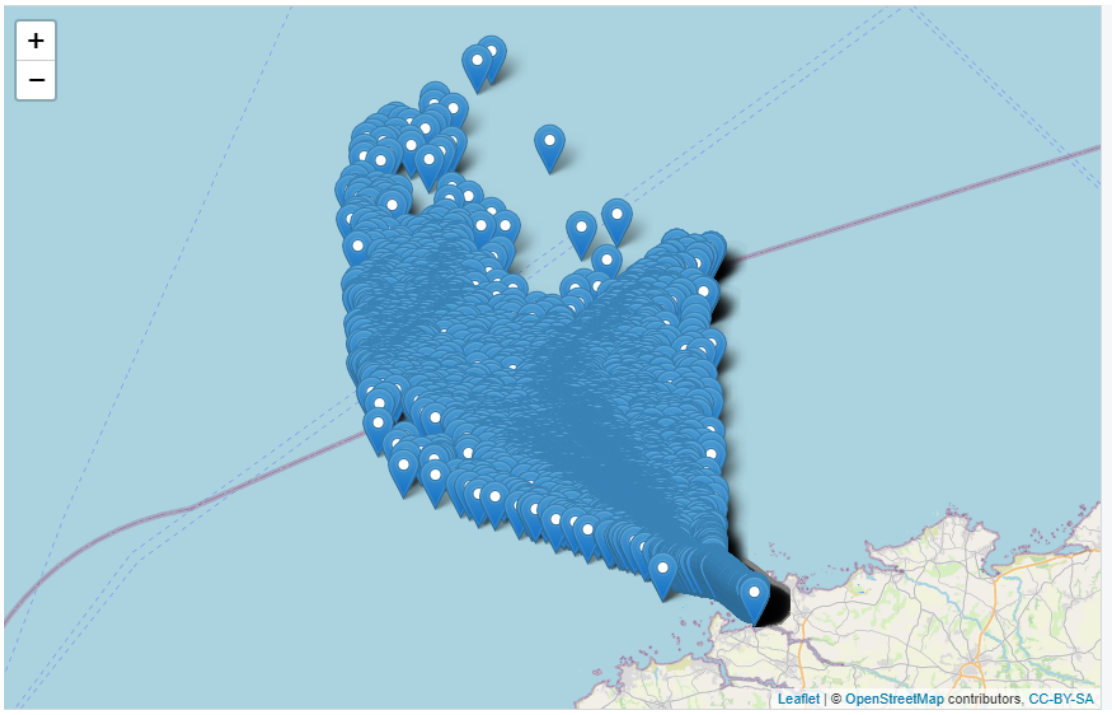
Γράφημα 4.4Α. *Δεδομένα τροχιάς αλιευτικού πλοίου ανοικτής θαλάσσης*



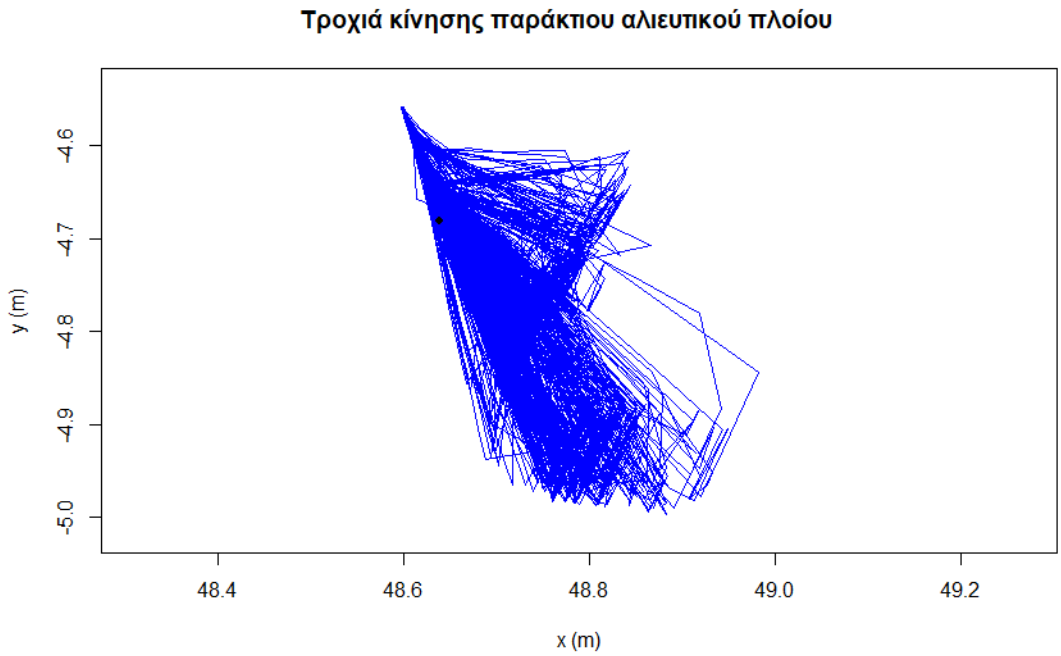
Γράφημα 4.4Β. *Λείανση δεδομένων τροχιάς αλιευτικού πλοίου ανοικτής θαλάσσης*

Το τελευταίο πλοίο του οποίου εξετάστηκε η πορεία του ήταν ένα παράκτιο αλιευτικό στην βορειοδυτική Γαλλία (γράφημα 4.5). Τα δύο μακρύτερα σημεία της πορείας του έχουν απόσταση μόλις 48 χιλιόμετρα αλλά η κίνηση του είναι συγκεκριμένη και

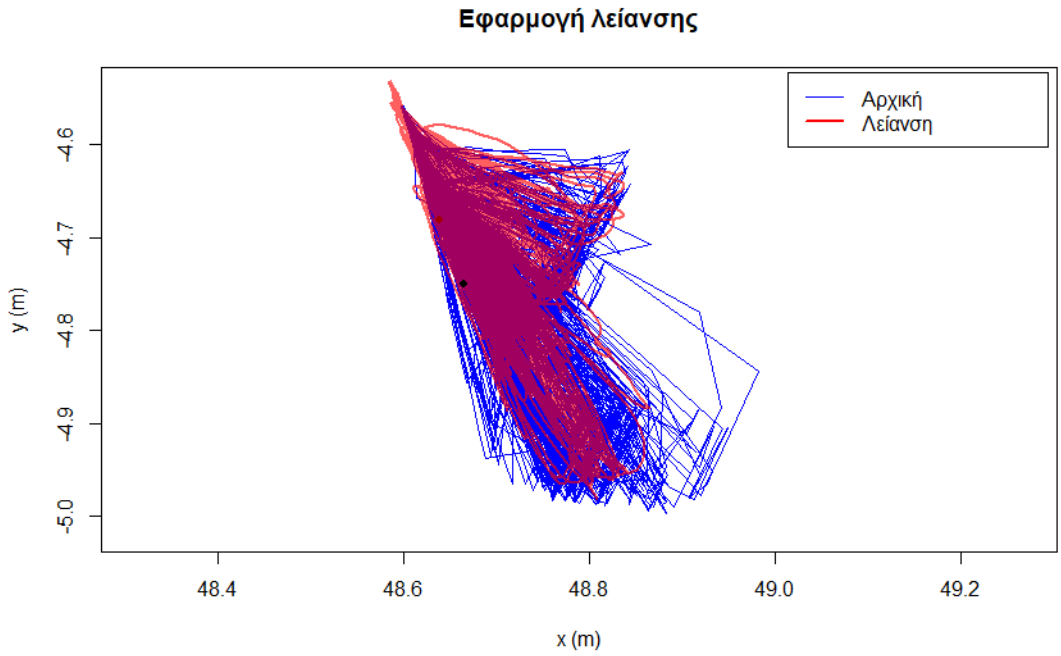
χαρακτηρίζεται από διαδοχικά τρίγωνα που έχουν το ίδιο κέντρο, το οποίο είναι το λιμάνι (γράφημα 4.5Α). Σύμφωνα με το γράφημα 4.5Β υπάρχουν δύο κύριες πορείες. Η πρώτη χαρακτηρίζεται από μια μεγαλύτερου μήκους τριγωνική πορεία και έχει κατεύθυνση βορειοδυτικά και η δεύτερη από μια πορεία μικρότερου μήκους η οποία έχει βορειοανατολική κατεύθυνση.



Γράφημα 4.5. *Πορεία παράκτιου αλιευτικού πλοίου*



Γράφημα 4.5Α. *Δεδομένα τροχιάς παράκτιου αλιευτικού σκάφους*



Γράφημα 4.5Β. *Λείανση δεδομένων τροχιάς παράκτιου αλιευτικού πλοίου*

4.2. Εφαρμογή μεθόδων

Πριν την εφαρμογή των μεθόδων έγινε υπολογισμός των αποστάσεων μεταξύ των σημείων της πορείας των πλοίων με την μέθοδο του Haversine που υπολογίζεται από τον τύπο

$$d((x_1, y_1), (x_2, y_2)) = 2r_0\sqrt{\sin^2\left(\frac{y_2 - y_1}{2}\right) + \cos(y_1)\cos(y_2)\sin^2\left(\frac{x_2 - x_1}{2}\right)}.$$

Επίσης υπολογίστηκαν, η απόσταση που καλύφθηκε ανά σημείο σε ναυτικά μίλια, ο χρόνος μετάβασης μεταξύ των σημείων και η μεταβολή της πορείας του πλοίου. Για την καλύτερη ποιότητα του περιεχόμενου των δεδομένων και για την εξαγωγή ακριβέστερων αποτελεσμάτων διαγράφηκαν τα δεδομένα με μηδενική ταχύτητα και διπλότυπες εγγραφές. Τα δεδομένα αντλήθηκαν από την βάση δεδομένων του οργανισμού διαχείρισης ακτών (https://coast.noaa.gov/htdata/CMSP/AISDataHandler/2020/index.html) και της Αυστραλιανής υπηρεσίας για την ασφάλεια στα θαλάσσια ταξίδια (https://www.operations.amsa.gov.au/Spatial/DataServices/DigitalData).

Τα αποτελέσματα της εφαρμογής των μεθόδων παρουσιάζονται στον πίνακα 4.2. Σύμφωνα με αυτά παρατηρήθηκε ότι καμία μέθοδος δεν είχε ακρίβεια πρόβλεψης μεγαλύτερη του 90%. Ο μεγαλύτερος συντελεστής ακρίβειας διαπιστώθηκε στην περίπτωση του παράκτιου αλιευτικού πλοίου (89,08%). Αυτός ο τύπος πλοίου είχε και τα μεγαλύτερα ποσοστά ακρίβειας εφαρμογής των μεθόδων και κατά πλειοψηφία τις χαμηλότερες τιμές μέσου τετραγωνικού σφάλματος.

Πίνακας 4.2. Αποτελέσματα ακρίβειας πρόβλεψης κίνησης ανά τύπο πλοίου και ανά μέθοδο

Τύπος πλοίου	**Μέθοδος**	**RMSE**	**MSE**	**MAE**	**Accuracy**
Tanker	Random Forest	0.3477798	0.1209508	39.5009979	82.73%
Tanker	Logistic Regression	0.4538818	0.2060087	0.4115274	61.70%
Tanker	Naive Bayes	0.4552986	0.2072968	0.3503022	75.26%
Tanker	Gradient Boosting	0.9560246	0.913983	0.8407316	79.14%
Tanker	SVM	0.4623169	0.2137369	0.2606308	76.60%
Tanker	MLP	0.4251054	0.1807146	0.3619503	65.68%
Cargo	Random Forest	0.3157902	0.0997234	155.9317842	85.90%
Cargo	Logistic Regression	0.4869825	0.237152	0.4742699	56.55%
Cargo	Naive Bayes	0.5549715	0.3079934	0.5316302	58.13%
Cargo	Gradient Boosting	0.9397111	0.883057	0.7687996	62.06%
Cargo	SVM	0.5873646	0.3449972	0.404228	60.81%
Cargo	MLP	0.4956881	0.2457067	0.4914134	56.55%
Transport	Random Forest	0.3257892	0.1061386	51.3904483	85.92%
Transport	Logistic Regression	0.3640905	0.1325619	0.2654623	79.73%
Transport	Naive Bayes	0.3757953	0.1412221	0.2579949	81.92%
Transport	Gradient Boosting	1.0775105	1.1610289	0.9866567	86.46%
Transport	SVM	0.4043217	0.1634761	0.2023715	81.79%
Transport	MLP	0.4267094	0.1820809	0.1820809	81.79%
Fish Open sea	Random Forest	0.4509019	0.2033125	152.0262744	67.78%
Fish Open sea	Logistic Regression	0.4914887	0.2415612	0.4838405	52.51%
Fish Open sea	Naive Bayes	0.6910218	0.4775111	0.4827468	61.30%
Fish Open sea	Gradient Boosting	0.8180891	0.6692698	0.6473216	67.99%
Fish Open sea	SVM	0.5917885	0.3502136	0.4146045	58.16%
Fish Open sea	MLP	0.4993694	0.2493698	0.4987395	47.49%
Fish Coastal	Random Forest	0.294038	0.0864583	4.872566	88.36%
Fish Coastal	Logistic Regression	0.2987109	0.0892282	0.1786963	85.61%
Fish Coastal	Naive Bayes	0.861306	0.7418481	0.8105801	87.21%
Fish Coastal	Gradient Boosting	2.8966417	8.3905332	2.6180478	88.84%
Fish Coastal	SVM	0.2982399	0.0889471	0.1410033	89.08%
Fish Coastal	MLP	0.4028809	0.162313	0.162313	83.77%

Για την ακριβή εξέταση εάν οι διαφορές των μέτρων ακρίβειας επηρεάζονται από τον τύπο του πλοίου, δηλαδή από την απόσταση που καλύπτει το κάθε ένα και από το τον τρόπο υπολογισμού, έγινε ανάλυσης της διασποράς με έναν παράγοντα ή one way ANOVA. Τα αποτελέσματα της εξέτασης των διαφορών των τιμών ακρίβειας παρουσιάζονται στον πίνακα 4.3 και έδειξαν ότι μόνο η ακρίβεια παρουσιάζει στατιστικά σημαντικές διαφορές των μέσων τιμών της ανά τύπου πλοίου ($F(4,25)=16,04$, $p<<0.001$).

Επιπλέον από τον ίδιο πίνακα διαπιστώθηκε ότι μεγαλύτερες τιμές ακρίβειας στην πρόβλεψη της πορείας ενός πλοίου παρουσιάζονται στα παράκτια αλιευτικά (Μ.Τ.=0,873, Τ.Α.=0,022) και στα επιβατηγά πλοία (Μ.Τ.=0,824, Τ.Α.=0,022) δηλαδή σε πλοία των οποίων οι διαδρομές δεν παρουσιάζουν σημαντικές αποκλίσεις από την προκαθορισμένη πορεία τους. Αντίθετα, τα αλιευτικά ανοικτής θαλάσσης παρουσίασαν την χαμηλότερη ακρίβεια (Μ.Τ.=0,579, Τ.Α.=0,0746) καθώς η πορεία τους δεν είναι προκαθορισμένη και μπορεί να μεταβληθεί ανά πάσα στιγμή καθώς εξαρτάται από την πορεία των ψαριών που προσπαθούν να αλιεύσουν.

Πίνακα 4.3. Εξέταση διαφορών των αποτελεσμάτων ακρίβειας ανά τύπο πλοίου.

		Μέση Τιμή	Τυπική Απόκλιση	F(4,25)	p
RMSE	Αλιευτικό ανοικτής θαλάσσης	.6008	.13870	0.465	0.761
	Παράκτιο Αλιευτικό	.8328	1.00761		
	Εμπορικό πλοίο	.5795	.18573		
	Δεξαμενόπλοιο	.5209	.21291		
	Επιβατηγό	.4996	.28373		
MSE	Αλιευτικό ανοικτής θαλάσσης	.3770	.18167	0.826	0.521
	Παράκτιο Αλιευτικό	1.5396	3.20852		
	Εμπορικό πλοίο	.3646	.25981		
	Δεξαμενόπλοιο	.3091	.29263		
	Επιβατηγό	.3167	.41261		
MAE	Αλιευτικό ανοικτής θαλάσσης	.6805	.42112	0.462	0.763
	Παράκτιο Αλιευτικό	.7737	1.04615		
	Εμπορικό πλοίο	26.4329	63.44021		
	Δεξαμενόπλοιο	6.9525	15.94401		
	Επιβατηγό	8.8795	20.82745		
Accurracy	Αλιευτικό ανοικτής θαλάσσης	.5791	.07464	16.04	<<0.001
	Παράκτιο Αλιευτικό	.8731	.02235		
	Εμπορικό πλοίο	.6146	.06909		
	Δεξαμενόπλοιο	.7282	.07339		
	Επιβατηγό	.8247	.02241		

Στον πίνακα 4.4 παρουσιάζονται τα αποτελέσματα της εξέτασης των διαφορών των μέσων τιμών ανά εφαρμοζόμενη μέθοδο πρόβλεψης. Σύμφωνα με τα αποτελέσματα αυτά διαπιστώθηκε ότι η εφαρμοζόμενη μέθοδος δεν επηρεάζει την ακρίβεια της μέτρησης ($F(5,24)=0.998$, $p=0.44$) αλλά διαφοροποιεί τις τιμές RMSE ($F(5,24)=4.899$, $p=0.003$) και MAE ($F(5,24)=6.754$, $p<0.001$). Τα αποτελέσματα αυτά δείχνουν ότι για την μέτρηση του συγκεκριμένου τύπου δεδομένων θα πρέπει να χρησιμοποιούνται τα μέτρα της ακρίβειας και σαν δεύτερο κριτήριο το μέσο τεραγονικο σφάλματο (MSE).

Αυτό ισχύει διότι δεν μεταβάλλονται σε σχέση με την εφαρμοζόμενη μέθοδο οπότε μπορούν να θεωρηθούν ως ακριβή μέτρα συγκρίσεων.

Πίνακας 4.4. Εξέταση διαφορών των αποτελεσμάτων ακρίβειας ανά εφαρμοζόμενη μέθοδο.

		Μέση Τιμή	Τυπική Απόκλιση	F(4,25)	p
RMSE	Gradient Boosting	1.3281	.84973	4.899	0.003
	Logistic regression	.4190	.08452		
	MLP	.4500	.04446		
	Naïve Bayes	.5877	.19303		
	Random Forest	.3868	.07504		
	SVM	.4688	.12497		
MSE	Gradient Boosting	2.3415	3.20548	2.12	0.097
	Logistic regression	.1813	.06748		
	MLP	.2040	.04049		
	Naïve Bayes	.3752	.24101		
	Random Forest	.1541	.05999		
	SVM	.2323	.11428		
MAE	Gradient Boosting	1.1682	.79446	6.754	<0.001
	Logistic regression	.3628	.13493		
	MLP	.3393	.16211		
	Naïve Bayes	.4867	.21086		
	Random Forest	62.0835	66.07967		
	SVM	.2846	.12162		
Accurracy	Gradient Boosting	.7714	.11473	0.998	0.44
	Logistic regression	.6722	.14622		
	MLP	.6706	.15744		
	Naïve Bayes	.7276	.12694		
	Random Forest	.7697	.11458		
	SVM	.7323	.13374		

Κεφάλαιο 5

Συζήτηση – Συμπεράσματα

Τα αποτελέσματα της έρευνας δεν παρουσίασαν αποκλίσεις από τα αναμενόμενα. Σύμφωνα με τα αποτελέσματα διαπιστώθηκε ότι η ακρίβεια της πρόβλεψης της πορείας ενός πλοίου εξαρτάται από τον τύπου του πλοίου και πιο συγκεκριμένα διαπιστώθηκε ότι μεγαλύτερες τιμές ακρίβειας στην πρόβλεψη της πορείας ενός πλοίου παρουσιάζονται στα παράκτια αλιευτικά και στα επιβατηγά πλοία, δηλαδή σε πλοία των οποίων οι διαδρομές δεν παρουσιάζουν σημαντικές αποκλίσεις από την προκαθορισμένη πορεία τους. Ήδη από την οπτική ανάλυση είχε διαπιστωθεί ότι η πρόβλεψη της πορείας αυτών των δύο τύπων πλοίων είναι αρκετά εύκολή καθώς οι αποστάσεις που καλύπτουν είναι σχετικά μικρές και το μοτίβο που ακολουθούν επαναλαμβανόμενο. Δεν συμβαίνει όμως το ίδιο και με τα αλιευτικά ανοιχτής θαλάσσης που δεν έχουν ένα συγκεκριμένο τόπο αλίευσης και θα πρέπει να ακολουθούν ένα κοπάδι ψαριών. Δεν είναι τυχαίο λοιπόν που αυτού του είδους πλοίου η πορεία παρουσίασε και την χαμηλότερη ακρίβεια πρόβλεψης. Χαμηλή ήταν η ακρίβεια της πρόβλεψης της εξεταζόμενης πορείας του εμπορικού πλοίου η οποία όμως για την συγκεκριμένη περίπτωση δεν αναμενόταν τόσο χαμηλή καθώς το μοτίβο της κίνησης του ήταν επαναλαμβανόμενο (Port to Port). Τέλος, η εκτίμηση της πορείας του δεξαμενόπλοιου ήταν υψηλή αν και αναμενόταν μεγαλύτερη καθώς η πορεία του δεν είχε έντονες αποκλίσεις. Με βάση τα προηγούμενα γίνεται φανερό ότι πράγματι ο σχεδιασμός βέλτιστης πορείας είναι μεγαλύτερος σε εμπορικά πλοία και δεξαμενόπλοια, ενώ δεν έχει καμία πρακτική αξία η παρόμοια προσπάθεια σε αλιευτικά ανοιχτής θαλάσσης η σε πλοία που κινούνται με παρόμοιο τρόπο όπως π.χ. σε ιστιοφόρα πλοία αναψυχής.

Για τον συγκεκριμένο τύπο δεδομένων, ανεξαρτήτως του τύπου του πλοίου διαπιστώθηκε ότι μόνο η ακρίβεια (accuracy) αποτελεί ασφαλές μέτρο για την εκτίμηση της βέλτιστης πορείας ενός πλοίου καθώς έδειξε ότι δεν επηρεάζεται ούτε εξαρτάται από την εφαρμοζόμενη μέθοδο. Παρόμοια συμπεριφορά έδειξε και το μέσο τετραγωνικό σφάλμα όμως η σημαντικότητα του συγκεκριμένου μέτρου ήταν οριακή.

5.1. Περιορισμοί της έρευνας – Προτάσεις για μελλοντική έρευνα

Βασικός περιορισμός της έρευνας ήταν η έλλειψη δωρεάν δεδομένων AIS. Λόγω της εμπορικότητας των εφαρμογών που βασίζονται σε δεδομένα AIS υπήρξε μεγάλη δυσκολία στην εύρεση δεδομένων αυτού του τύπου. Ένας ακόμη λόγος στην δυσκολία ανάκτησης δεδομένων AIS είναι και η ασφάλεια των θαλάσσιων διαδρομών καθώς για την ανάκτηση δεδομένων από το Κέντρο Πλοήγησης Λιμενικού Σώματος των ΗΠΑ θα πρέπει να γίνει εγγραφή και πλήρη αιτιολόγηση του αιτήματος. Επιπλέον, τα δωρεάν δεδομένα μπορεί να μην περιέχουν ακριβείς αποτυπώσεις της πορείας καθώς μπορεί τα δεδομένα που θα συμπεριληφθούν στο τελικό dataset να έχουν προκύψει μετά από δειγματοληψία των αρχικών συντεταγμένων. Με βάση τα παραπάνω γίνεται σαφές ότι για μια πιο ακριβή αποτύπωση των δεομένων σε μια παρόμοια μελλοντική έρευνα θα πρέπει να υπάρχει συνεισφορά ενός συνεργαζόμενου ιδρύματος για την εύκολη πρόσβαση στην πληροφορία.

Βιβλιογραφία

Awasthi, S., (2021). Complete Analysis of Gradient Descent Algorithm . Ανακτήθηκε 12/7/2021 από https://datamahadev.com/complete-analysis-of-gradient-descent-algorithm/

Balduzzi, M., Pasta, A., & Wilhoit, K. (2014). A security evaluation of AIS Automated Identification System. 30th Annual Computer Security Applications Conference (pp. 436–445). doi: 10.1145/2664243.2664257

Bay, S. (2017). Evaluation of factors on the patterns of ship movement and predictability of future ship location in the Gulf of Mexico. (Master's thesis). Retrieved from http://calhoun.nps.edu/handle/10945/53021.

Das, S., (2019). 8 Factors to Consider When Choosing Route Optimization Software for Your Logistics Business. Ανακτήθηκε 15/7/2021 από https://www.supplychain247.com/article/8_factors_to_consider_when_choosing_route_optimization_software/locus

DON, (2007). Department of the Navy. Maritime Domain Awareness Concept. Washington, DC: Chief of Naval Operations. Ανακτήθηκε 1/7/2021 από https://www.hsdl.org/?view&did=719590

Friedman, J.H. (2001). Greedy Function Approximation A Gradient Boosting Machine

Harati-Mokhtari, A., Wall, A., Brooks, P., & Wang, J. (2007). Automatic Identification System (AIS): Data reliability and human error implications. The Journal of Navigation, 60(3), 373-389.

Hastie, T., Tibshiran, R., & Friedman, J. (2009). The elements of statistical learning.New York: Springer Science + Business Media, LLC.

Heckerman, D.(1999). Learning in graphical models, chapter A tutorial on learning with Bayesian networks, pages 301–354. MIT Press

Κύρκος, Ε. (2015). Κατηγοριοποίηση. [Κεφάλαιο Συγγράμματος]. Στο Κύρκος, Ε. 2015. Επιχειρηματική ευφυΐα και εξόρυξη δεδομένων. [ηλεκτρ. βιβλ.] Αθήνα: Σύνδεσμος Ελληνικών Ακαδημαϊκών Βιβλιοθηκών. κεφ 9. Διαθέσιμο στο: http://hdl.handle.net/11419/1236

Kecman, V., (2005). Support Vector Machines – An Introduction. 10.1007/10984697_1.

Khan, A., Cees, B., & Kaye, M. (2005). Ship motion prediction for launch and recovery of air vehicles. Proceedings of MTS/IEEE OCEANS. doi: 10.1109/OCEANS.2005.1640198

Levinson, M. (2006). The Box: How the Shipping Container Made the World Smaller and the World Economy Bigger. Princeton: Princeton University Press.

Mao, S., Tu, E., Zhang, G., Rachmawati, L., Rajabally, E., & Huang, G.-B. (2016). An Automatic Identification System (AIS) database for maritime trajectory prediction and data mining. Ανακτήθηκε 1/7/2021 από https://arxiv.org/pdf/1607.03306.pdf

Maury, M.F.: The Physical Geography of the Sea and Its Meteorology, Harper Bros, New York, 1855.

McAbee, A. (2013). Traffic pattern detection using the Hough transformation for anomaly detection to improve maritime domain awareness. (Master's thesis). Retrieved from http://calhoun.nps.edu/handle/10945/38977

Millifiori, L., Braca, P., Bryan, K., & Willett, P. (2016). Modeling vessel kinematics using a stochastic mean-reverting process for long-term prediction. IEEE Transactions on Aerospace and Electronic Systems, 52(5). doi: 10.1109/TAES.2016.150596

Morris, B., & Trivedi, M. (2008). A survey of vision-based trajectory learning and analysis for surveillance. IEEE Transactions on Circuits and Systems for Video Technology, 18(8), 1114–1127. doi: 10.1109/TCSVT.2008.927109

Morris, B., & Trivedi, M. (2011). Trajectory learning for activity understanding: unsupervised, multilevel, and long-term adaptive approach. IEEE Transactions on Pattern Analysis and Machine Intelligence, 33(11), 2287–2301. doi: 10.1109/TPAMI.2011.64

Notteboom, T. (2012). “Container Shipping.” In The Blackwell Companion to Maritime Economics, edited by W.K. Talley, 230–262. Oxford: Wiley-Blackwell.

Πετρίδης, Δ., (2015.) Ανάλυση πολυμεταβλητών τεχνικών. [ηλεκτρ. βιβλ.] Αθήνα:Σύνδεσμος Ελληνικών Ακαδημαϊκών Βιβλιοθηκών. Διαθέσιμο στο: http://hdl.handle.net/11419/2126

Palacios, R., & Doshi, A. G. (2008). Computing aircraft position prediction. The Open Transportation Journal, 2, 94–97.

Pallota, G., Horn, S., Braca, P., & Bryan, K. (2014). Context-enhanced vessel prediction based on Ornstein-Uhlenbeck proceses using historical AIS traffic

patterns: Real- world experimental results. Fusion. Ανακτήθηκε 1/7/2021 από https://www.academia.edu/23838731/Context-enhanced_vessel_prediction_based_on_Ornstein-Uhlenbeck_processes_using_historical_AIS_traffic_patterns_Real-world_experimental_results

Pallotta, G., Vespe, M., & Bryan, K. (2013). Vessel pattern knowledge discovery from AIS data: A framework for anomaly detection and route prediction. Entropy, 15(6), 2218-2245.

Puigrefagut, A., (2021). China and India fight for the gates of the Strait of Malacca. Ανακτήθηκε 10/7/2021 από https://www.unav.edu/web/global-affairs/detalle/-/blogs/china-and-india-fight-for-the-gates-of-the-strait-of-malacca

RayMarine, (2016). AIS - The Ultimate Guide to Automatic Identification Systems. Ανακτήθηκε 1/7/2021 από https://www.raymarine.com/view/blog/news/details/index-ID=15032385552.html

Ren, J., Lutzen, M., Rasmussen, H.M. (2018). Identification of Success Factors for Green Shipping with Measurement of Greenness Based on ANP and ISMIn Paul Tae-Woo Lee, Zaili Yang (eds.) Multi-Criteria Decision Making in Maritime Studies and Logistics: Applications and Cases. Texas, Springer International Publishing

Ristic, Branko & Scala, B. & Morelande, M.R. & Gordon, N.. (2008). Statistical Analysis of Motion Patterns in AIS Data: Anomaly Detection and Motion Prediction. Information Fusion - INFFUS. 1-7. 10.1109/ICIF.2008.4632190.

Rodrigue, J.-P., ed. (2013). The Geography of Transport Systems, 3rd edn. London: Routledge.

Rodrigue, Jean-Paul. (2017). Maritime Transport. 10.1002/9781118786352.wbieg0155.

Stone, L., Streit, R., Corwin, T., & Bell, K. (2014). Bayesian multiple target tracking. Norwood, MA: Artech House.

Stopford, M. (2009). Maritime Economics, 3rd edn. London: Routledge.

Tester, K. A. (2013). A spatiotemporal clustering approach to maritime domain awareness. (Master's thesis). Retrieved from http://calhoun.nps.edu/handle/10945/37731

Tu, E., Zhang, G., Rachmawati, L., Rajabally, E., & Huang, G.-B. (2016). Exploiting AIS data for intelligent maritime navigation: A comprehensive survey. Ανακτήθηκε 2/7/2021 από https://arxiv.org/abs/1606.00981

Vespe, M., Visentini, I., Bryan, K., & Braca, P. (2012). Unsupervised learning of maritime traffic patterns for anomaly detection. Entropy 2013, 15, 2218–2245. doi:10.3390/e15062218

Zhu, F. (2011). Mining ship trajectory patterns from AIS database for maritime surveillance. Emergency Management and Management Sciences (ICEMMS). doi: 10.1109/ICEMMS.2011.6015796

Webb G.I. (2011) Naïve Bayes. In: Sammut C., Webb G.I. (eds) Encyclopedia of Machine Learning. Springer, Boston, MA. https://doi.org/10.1007/978-0-387-30164-8_576

Παράρτημα

Κώδικας R

```
#Εκδοση 4.1.1 (2021-08-10) -- "Kick Things"
#Εγκατάσταση πακέτων
install.packages("trajr")
install.packages("leaflet")
install.packages("geosphere")
install.packages("Metrics")
#Φόρτωση βιβλιοθηκών και επιλογών
library(trajr)
library(sp)
library(leaflet)
library(geosphere)
library(Metrics)
options(scipen = 999)



#Εισαγωγή δεδομένων
setwd("C:\\Users\\user\\Downloads")
df<-read.csv('input.csv',sep=",")
df<-df[,-c(2,3,4,6,9,10)]

#Υπολογισμός αποστάσεων αρχικού και τελικού σημείου
distm(c(31.7032, -39.01601), c(-5.608113,4.979350 )*2, fun = distHaversine)/1000
#Απόσταση 259600076146315 (df106)
distm(c(78.2377, 24.66277), c(30.4178,22.44139), fun = distHaversine)/1000
#Απόσταση 44081436890597 (df110)
distm(c(-52.29864, -32.65470), c(-45.69751,-25.17469), fun = distHaversine)/1000
#Απόσταση 260142468095943 (df111)
distm(c(59.45430, -75.9210), c(57.77395, -62.4281), fun = distHaversine)/1000
#Απόσταση 259600076146315 (df110)
```

```
distm(c(-4.559413, 48.59725), c(-4.843955,48.98191), fun = distHaversine)/1000
#Απόσταση 187372144064677 (df86)


#Πίνακας αποτελεσμάτων
results<-matrix(0,30,6)
results[1:6,1]<-"Tanker"
results[7:12,1]<-"Cargo"
results[13:18,1]<-"Trasport"
results[19:24,1]<-"Fish Open sea"
results[25:30,1]<-"Fish Coastal"
results[c(1,7,13,19,25),2]<-"Random Forest"
results[c(2,8,14,20,26),2]<-"Logistic Regression"
results[c(3,9,15,21,27),2]<-"Naive Bayes"
results[c(4,10,16,22,28),2]<-"Gradient Boosting"
results[c(5,11,17,23,29),2]<-"SVM"
results[c(6,12,18,24,30),2]<-"MLP"
##############################################################################
#Πετρελαιοφόρο
#Διαγράμματα κίνησης
df106 <- df[ which(df$mmsi=='246929977513855' ),]

trj <- TrajFromCoords(df106[,3:4])
plot(trj,main="Τροχιά κίνησης πετρελαιοφόρου πλοίου     ",col="blue")

plot(trj,main="Εφαρμογή λείανσης",col="blue")
smoothed <- TrajSmoothSG(trj, p = 3, n = 31)
lines(smoothed, col = "#FF0000A0", lwd = 2)
legend("topright", c("Αρχική", "Λείανση"), lwd = c(1, 2),
       lty = c(1, 1), col = c("blue", "red"), inset = 0.01)


cdf2<- data.frame(longitude = df106$lat,latitude = df106$lon)
```

```
coordinates(cdf2) <- ~latitude+longitude
leaflet(cdf2) %>% addMarkers() %>% addTiles()

#Υπολογισμός αποστάσεων όλων των σημείων
df106 <- df[ which(df$mmsi=='246929977513855' ),]
df106<-df106[-(which(df106$speed == 0)), ]
k<-dim(df106)[1]

for (i in 1:k)
{
  if (df106[i,1] > 90) df106[i,1]<-df106[i,1]-90
  if (df106[i,1] < -90) df106[i,1]<-df106[i,1]+90
  if (df106[i,2] > 90) df106[i,2]<-df106[i,2]-90
  if (df106[i,2] < -90) df106[i,2]<-df106[i,2]+90
  if (i > 1) df106[i,5]<-0.539956803*distm(c(df106[i,4], df106[i,3]),
                                           c(df106[i-1,4], df106[i-1,3]),
                                           fun = distHaversine)/1000
  if (i > 1) df106[i,6]<-atan((df106[i,4] - df106[i-1,4]) / (df106[i,3] - df106[i-1,3]))



}

#Υπολογισμοί
names(df106)[5]<-"dist"
df106[5]<-unlist(df106[5],use.names = FALSE)
df106[,7]<-df106[5]/df106[2]
names(df106)[7]<-"hours"
names(df106)[6]<-"direction"
df106<-df106[-(which(df106$dist == 0)), ]
df106<-df106[-1,]
df106[,8]<-ifelse(df106[,6] > 0,1,0)
names(df106)[8]<-"dirchange"
```

```
#Random forest
library(randomForest)
set.seed(123)
df106.rf <- randomForest(dirchange~lon+lat, data = df106, importance = TRUE, na.action = na.omit)
print(df106.rf)
plot(df106.rf)
pvrf<-unlist(df106.rf$predicted,use.names = FALSE)
df106[,9]<-round(pvrf,3)
names(df106)[9]<-"predvrf"
df106[,10]<-ifelse(df106$predvrf>0.6,1,0)
names(df106)[10]<-"probrf"
tblrf<-table(df106$dirchange,df106$probrf)
tblrf
acc<-sum(diag(tblrf))/sum(tblrf)
results[1,6]<-round(acc,7)
results[1,3]<-round(sqrt(df106.rf$mse[which.min(df106.rf$mse)]),7) #RMSE
results[1,4]<-round((df106.rf$mse[which.min(df106.rf$mse)]),7)    #MSE
results[1,5]<-round(mae(df106$lon, unlist(df106.rf$predicted,use.names = FALSE)),7)   #MAE

# Logistic regression
df106.lr<-glm(dirchange~lon+lat,family="binomial",data=df106)
plot(df106.lr)
pvlr<-unlist(df106.lr$fitted.values,use.names = FALSE)
df106[,11]<-pvlr
names(df106)[11]<-"predvlr"
results[2,3]<-round(rmse(df106$predvlr,df106$dirchange),7)
results[2,4]<-round(mse(df106$predvlr,df106$dirchange),7)
results[2,5]<-round(mae(df106$predvlr,df106$dirchange),7)
df106[,12]<-ifelse(df106$predvlr<0.65,0,1)
names(df106)[12]<-"problr"
tbllr<-table(df106$dirchange,df106$problr)
```

```
tbllr
acc<-sum(diag(tbllr))/sum(tbllr)
results[2,6]<-round(acc,7)

#Naive Bayes
library(naivebayes)
df106.nb <- naive_bayes( factor(dirchange)~lon+lat, data = df106, usekernel = T)
df106.nb
plot(df106.nb)
pvnb <- predict(df106.nb, type = 'prob')
k<-length(pvnb[,1])
for (i in 1:k) {
  df106[i,13]<-max(pvnb[i,])
  if (max(pvnb[i,]) == pvnb[i,1]) df106[i,14]<-0
  if (max(pvnb[i,]) == pvnb[i,2]) df106[i,14]<-1
}
names(df106)[13]<-"predvnb"
names(df106)[14]<-"probnb"
tblnb<-table(df106$dirchange,df106$probnb)
tblnb
acc<-sum(diag(tblnb))/sum(tblnb)
results[3,6]<-round(acc,7)
results[3,3]<-round(rmse(df106$predvnb,df106$dirchange),7)
results[3,4]<-round(mse(df106$predvnb,df106$dirchange),7)
results[3,5]<-round(mae(df106$predvnb,df106$dirchange),7)

#Gradient boosting
library(gbm)
df106.gb <- gbm(dirchange~lon+lat, data = df106)
plot(df106.gb$fit)
pvgb<-df106.gb$fit
df106[,15]<-pvgb
df106[,16]<-ifelse(df106.gb$fit>0,1,0)
```

```
names(df106)[15]<-"predvgb"
names(df106)[16]<-"probgb"
tblgb<-table(df106$dirchange,df106$probgb)
tblgb
acc<-sum(diag(tblgb))/sum(tblgb)
results[4,6]<-round(acc,7)
results[4,3]<-round(rmse(df106$predvgb,df106$dirchange),7)
results[4,4]<-round(mse(df106$predvgb,df106$dirchange),7)
results[4,5]<-round(mae(df106$predvgb,df106$dirchange),7)

#SVM
library(e1071)
df106.svm<-svm(dirchange~lon+lat, data = df106)
df106.svm
pvsvm<-predict(df106.svm, df106)
plot(df106.svm,pvsvm)
df106[,17]<-pvsvm
df106[,18]<-ifelse(df106.svm$fit>0.6,1,0)
names(df106)[17]<-"predvsvm"
names(df106)[18]<-"probsvm"
tblsvm<-table(df106$dirchange,df106$probsvm)
tblsvm
acc<-sum(diag(tblsvm))/sum(tblsvm)
results[5,6]<-round(acc,7)
results[5,3]<-round(rmse(df106$predvsvm,df106$dirchange),7)
results[5,4]<-round(mse(df106$predvsvm,df106$dirchange),7)
results[5,5]<-round(mae(df106$predvsvm,df106$dirchange),7)

#MLP
library(nnet)
set.seed(123)
df106.dat<-df106[,c(3,4,8)]
df106.mlp<-nnet(dirchange~lon+lat, size=1,data = df106.dat,hidden = 1 )
```

```
pvmlp<-predict(df106.mlp, df106.dat[,-3])
df106[,19]<-pvmlp
df106[,20]<-ifelse(df106[,19]>0.6,1,0)
names(df106)[19]<-"predvmlp"
names(df106)[20]<-"probmlp"
tblmlp<-table(df106$dirchange,df106$probmlp)
tblmlp
acc<-sum(diag(tblmlp))/sum(tblmlp)
results[6,6]<-round(acc,7)
results[6,3]<-round(rmse(df106$predvmlp,df106$dirchange),7)
results[6,4]<-round(mse(df106$predvmlp,df106$dirchange),7)
results[6,5]<-round(mae(df106$predvmlp,df106$dirchange),7)

# Logistic regression
   df106.lr<-glm(dirchange~lon+lat,family="binomial",data=df106)
    pvlr<-unlist(df106.lr$fitted.values,use.names = FALSE)
     df106[,11]<-pvlr
      names(df106)[11]<-"predvlr"
      results[2,3]<-round(rmse(df106$predvlr,df106$dirchange),7)
      results[2,4]<-round(mse(df106$predvlr,df106$dirchange),7)
      results[2,5]<-round(mae(df106$predvlr,df106$dirchange),7)
      df106[,12]<-ifelse(df106$predvlr<0.65,0,1)
      names(df106)[12]<-"problr"
      tbllr<-table(df106$dirchange,df106$problr)
      tbllr
       acc<-sum(diag(tbllr))/sum(tbllr)
       results[2,6]<-round(acc,7)

#Naive Bayes
       library(naivebayes)

       df106.nb <- naive_bayes( factor(dirchange)~lon+lat, data = df106, usekernel = T)
         df106.nb
```

```
  plot(df106.nb)
  pvnb <- predict(df106.nb, type = 'prob')
  k<-length(pvnb[,1])
 for (i in 1:k) {
  df106[i,13]<-max(pvnb[i,])
  if (max(pvnb[i,]) == pvnb[i,1]) df106[i,14]<-0
  if (max(pvnb[i,]) == pvnb[i,2]) df106[i,14]<-1
 }
  names(df106)[13]<-"predvnb"
  names(df106)[14]<-"probnb"
  tblnb<-table(df106$dirchange,df106$probnb)
  tblnb
  acc<-sum(diag(tblnb))/sum(tblnb)
  results[3,6]<-round(acc,7)
  results[3,3]<-round(rmse(df106$predvnb,df106$dirchange),7)
  results[3,4]<-round(mse(df106$predvnb,df106$dirchange),7)
  results[3,5]<-round(mae(df106$predvnb,df106$dirchange),7)

#Gradient boosting
  library(gbm)
  df106.gb <- gbm(dirchange~lon+lat, data = df106)
  plot(df106.gb$fit)
  pvgb<-df106.gb$fit
  df106[,15]<-pvgb
  df106[,16]<-ifelse(df106.gb$fit>0,1,0)
  names(df106)[15]<-"predvgb"
  names(df106)[16]<-"probgb"
  tblgb<-table(df106$dirchange,df106$probgb)
  tblgb
  acc<-sum(diag(tblgb))/sum(tblgb)
 results[4,6]<-round(acc,7)
 results[4,3]<-round(rmse(df106$predvgb,df106$dirchange),7)
 results[4,4]<-round(mse(df106$predvgb,df106$dirchange),7)
```

```
results[4,5]<-round(mae(df106$predvgb,df106$dirchange),7)

#SVM
    library(e1071)
    df106.svm<-svm(dirchange~lon+lat, data = df106)
    df106.svm
    pvsvm<-predict(df106.svm, df106)
    plot(df106.svm,pvsvm)
    df106[,17]<-pvsvm
    df106[,18]<-ifelse(df106.svm$fit>0.6,1,0)
    names(df106)[17]<-"predvsvm"
    names(df106)[18]<-"probsvm"
    tblsvm<-table(df106$dirchange,df106$probsvm)
    tblsvm
    acc<-sum(diag(tblsvm))/sum(tblsvm)
    results[5,6]<-round(acc,7)
    results[5,3]<-round(rmse(df106$predvsvm,df106$dirchange),7)
    results[5,4]<-round(mse(df106$predvsvm,df106$dirchange),7)
    results[5,5]<-round(mae(df106$predvsvm,df106$dirchange),7)

#MLP
    library(nnet)
    set.seed(123)
    df106.dat<-df106[,c(3,4,8)]
    df106.mlp<-nnet(dirchange~lon+lat, size=1,data = df106.dat,hidden = 1 )
    pvmlp<-predict(df106.mlp, df106.dat[,-3])
    df106[,19]<-pvmlp
    df106[,20]<-ifelse(df106[,19]>0.6,1,0)
    names(df106)[19]<-"predvmlp"
    names(df106)[20]<-"probmlp"
    tblmlp<-table(df106$dirchange,df106$probmlp)
    tblmlp
    acc<-sum(diag(tblmlp))/sum(tblmlp)
```

```
        results[6,6]<-round(acc,7)
        results[6,3]<-round(rmse(df106$predvmlp,df106$dirchange),7)
        results[6,4]<-round(mse(df106$predvmlp,df106$dirchange),7)
        results[6,5]<-round(mae(df106$predvmlp,df106$dirchange),7)

#########################################################################
#Εμπορικό πλοίο
#Διαγράμματα κίνησης
df19 <- df[ which(df$mmsi=='44081436890597' ),]
trj <- TrajFromCoords(df19[,3:4])
plot(trj,main="Τροχιά κίνησης εμπορικού πλοίου     ",col="blue")

plot(trj,main="Εφαρμογή λείανσης",col="blue")
smoothed <- TrajSmoothSG(trj, p = 3, n = 31)
lines(smoothed, col = "#FF0000A0", lwd = 2)
legend("topright", c("Αρχική", "Λείανση"), lwd = c(1, 2),
       lty = c(1, 1), col = c("blue", "red"), inset = 0.01)


cdf2<- data.frame(longitude = df19$lat,latitude = df19$lon)
coordinates(cdf2) <- ~latitude+longitude
leaflet(cdf2) %>% addMarkers() %>% addTiles()

#Υπολογισμός αποστάσεων όλων των σημείων
df19 <- df[ which(df$mmsi=='44081436890597' ),]
df19<-df19[-(which(df19$speed == 0)), ]
k<-dim(df19)[1]

for (i in 1:k)
{
  if (df19[i,1] > 90) df19[i,1]<-df19[i,1]-90
  if (df19[i,1] < -90) df19[i,1]<-df19[i,1]+90
  if (df19[i,2] > 90) df19[i,2]<-df19[i,2]-90
```

```
  if (df19[i,2] < -90) df19[i,2]<-df19[i,2]+90
  if (i > 1) df19[i,5]<-0.539956803*distm(c(df19[i,4], df19[i,3]),
                                          c(df19[i-1,4], df19[i-1,3]),
                                          fun = distHaversine)/1000
  if (i > 1) df19[i,6]<-atan((df19[i,4] - df19[i-1,4]) / (df19[i,3] - df19[i-1,3]))


}

#Υπολογισμοί
names(df19)[5]<-"dist"
df19[5]<-unlist(df19[5],use.names = FALSE)
df19[,7]<-df19[5]/df19[2]
names(df19)[7]<-"hours"
names(df19)[6]<-"direction"
df19<-df19[-(which(df19$dist == 0)), ]
df19<-df19[-1,]
df19[,8]<-ifelse(df19[,6] > 0,1,0)
names(df19)[8]<-"dirchange"

#Random forest
library(randomForest)
set.seed(123)
df19.rf <- randomForest(dirchange~lon+lat, data = df19, importance = TRUE,
na.action = na.omit)
print(df19.rf)
plot(df19.rf)
pvrf<-unlist(df19.rf$predicted,use.names = FALSE)
df19[,9]<-round(pvrf,3)
names(df19)[9]<-"predvrf"
df19[,10]<-ifelse(df19$predvrf>0.6,1,0)
names(df19)[10]<-"probrf"
tblrf<-table(df19$dirchange,df19$probrf)
```

```
tblrf
acc<-sum(diag(tblrf))/sum(tblrf)
results[7,6]<-round(acc,7)
results[7,3]<-round(sqrt(df19.rf$mse[which.min(df19.rf$mse)]),7) #RMSE
results[7,4]<-round((df19.rf$mse[which.min(df19.rf$mse)]),7)     #MSE
results[7,5]<-round(mae(df19$lon, unlist(df19.rf$predicted,use.names = FALSE)),7)

# Logistic regression
df19.lr<-glm(dirchange~lon+lat,family="binomial",data=df19)
pvlr<-unlist(df19.lr$fitted.values,use.names = FALSE)
df19[,11]<-pvlr
names(df19)[11]<-"predvlr"
results[8,3]<-round(rmse(df19$predvlr,df19$dirchange),7)
results[8,4]<-round(mse(df19$predvlr,df19$dirchange),7)
results[8,5]<-round(mae(df19$predvlr,df19$dirchange),7)
df19[,12]<-ifelse(df19$predvlr<0.65,0,1)
names(df19)[12]<-"problr"
tbllr<-table(df19$dirchange, df19$problr)
tbllr
acc<-sum(diag(tbllr))/sum(tbllr)
results[8,6]<-round(acc,7)

#Naive Bayes
library(naivebayes)
df19.nb <- naive_bayes( factor(dirchange)~lon+lat, data = df19, usekernel = T)
df19.nb
plot(df19.nb)
pvnb <- predict(df19.nb, type = 'prob')
k<-length(pvnb[,1])
for (i in 1:k) {
  df19[i,13]<-max(pvnb[i,])
  if (max(pvnb[i,]) == pvnb[i,1]) df19[i,14]<-0
  if (max(pvnb[i,]) == pvnb[i,2]) df19[i,14]<-1
```

```
}
names(df19)[13]<-"predvnb"
names(df19)[14]<-"probnb"
tblnb<-table(df19$dirchange,df19$probnb)
tblnb
acc<-sum(diag(tblnb))/sum(tblnb)
results[9,6]<-round(acc,7)
results[9,3]<-round(rmse(df19$predvnb,df19$dirchange),7)
results[9,4]<-round(mse(df19$predvnb,df19$dirchange),7)
results[9,5]<-round(mae(df19$predvnb,df19$dirchange),7)

#Gradient boosting
library(gbm)
df19.gb <- gbm(dirchange~lon+lat, data = df19)
plot(df19.gb$fit)
pvgb<-df19.gb$fit
df19[,15]<-pvgb
df19[,16]<-ifelse(df19.gb$fit>0,1,0)
names(df19)[15]<-"predvgb"
names(df19)[16]<-"probgb"
tblgb<-table(df19$dirchange,df19$probgb)
tblgb
acc<-sum(diag(tblgb))/sum(tblgb)
results[10,6]<-round(acc,7)
results[10,3]<-round(rmse(df19$predvgb,df19$dirchange),7)
results[10,4]<-round(mse(df19$predvgb,df19$dirchange),7)
results[10,5]<-round(mae(df19$predvgb,df19$dirchange),7)

#SVM
library(e1071)
df19.svm<-svm(dirchange~lon+lat, data = df19)
df19.svm
pvsvm<-predict(df19.svm, df19)
```

```
plot(df19.svm,pvsvm)
df19[,17]<-pvsvm
df19[,18]<-ifelse(df19.svm$fit>0.6,1,0)
names(df19)[17]<-"predvsvm"
names(df19)[18]<-"probsvm"
tblsvm<-table(df19$dirchange,df19$probsvm)
tblsvm
acc<-sum(diag(tblsvm))/sum(tblsvm)
results[11,6]<-round(acc,7)
results[11,3]<-round(rmse(df19$predvsvm,df19$dirchange),7)
results[11,4]<-round(mse(df19$predvsvm,df19$dirchange),7)
results[11,5]<-round(mae(df19$predvsvm,df19$dirchange),7)

#MLP
library(nnet)
set.seed(123)
df19.dat<-df19[,c(3,4,8)]
df19.mlp<-nnet(dirchange~lon+lat, size=1,data = df19.dat,hidden = 1 )
pvmlp<-predict(df19.mlp, df19.dat[,-3])
df19[,19]<-pvmlp
df19[,20]<-ifelse(df19[,19]>0.6,1,0)
names(df19)[19]<-"predvmlp"
names(df19)[20]<-"probmlp"
tblmlp<-table(df19$dirchange,df19$probmlp)
tblmlp
acc<-sum(diag(tblmlp))/sum(tblmlp)
results[12,6]<-round(acc,7)
results[12,3]<-round(rmse(df19$predvmlp,df19$dirchange),7)
results[12,4]<-round(mse(df19$predvmlp,df19$dirchange),7)
results[12,5]<-round(mae(df19$predvmlp,df19$dirchange),7)
```

```
#########################################################################
        #Πλοίο γραμμής
#Διαγράμματα κίνησης
df111 <- df[ which(df$mmsi=='260142468095943' ),]
trj <- TrajFromCoords(df111[,3:4])
plot(trj,main="Τροχιά κίνησης πλοίου γραμμής",col="blue")

plot(trj,main="Εφαρμογή λείανσης",col="blue")
smoothed <- TrajSmoothSG(trj, p = 3, n = 31)
lines(smoothed, col = "#FF0000A0", lwd = 2)
legend("topright", c("Αρχική", "Λείανση"), lwd = c(1, 2),
       lty = c(1, 1), col = c("blue", "red"), inset = 0.01)



cdf2<- data.frame(longitude = df111$lat,latitude = df111$lon)
coordinates(cdf2) <- ~latitude+longitude
leaflet(cdf2) %>% addMarkers() %>% addTiles()

#Υπολογισμός αποστάσεων όλων των σημείων
df111 <- df[ which(df$mmsi=='260142468095943' ),]
df111<-df111[-(which(df111$speed == 0)), ]
k<-dim(df111)[1]

for (i in 1:k)
{
  if (df111[i,1] > 90) df111[i,1]<-df111[i,1]-90
  if (df111[i,1] < -90) df111[i,1]<-df111[i,1]+90
  if (df111[i,2] > 90) df111[i,2]<-df111[i,2]-90
  if (df111[i,2] < -90) df111[i,2]<-df111[i,2]+90
  if (i > 1) df111[i,5]<-0.539956803*distm(c(df111[i,4], df111[i,3]),
                                           c(df111[i-1,4], df111[i-1,3]),
                                           fun = distHaversine)/1000
  if (i > 1) df111[i,6]<-atan((df111[i,4] - df111[i-1,4]) / (df111[i,3] - df111[i-1,3]))
```

```
}

#Υπολογισμοί
names(df111)[5]<-"dist"
df111[5]<-unlist(df111[5],use.names = FALSE)
df111[,7]<-df111[5]/df111[2]
names(df111)[7]<-"hours"
names(df111)[6]<-"direction"
df111<-df111[-(which(df111$dist == 0)), ]
df111<-df111[-1,]
df111[,8]<-ifelse(df111[,6] > 0,1,0)
names(df111)[8]<-"dirchange"

#Random forest
library(randomForest)
set.seed(123)
df111.rf <- randomForest(dirchange~lon+lat, data = df111, importance = TRUE,
na.action = na.omit)
print(df111.rf)
plot(df111.rf)
pvrf<-unlist(df111.rf$predicted,use.names = FALSE)
df111[,9]<-round(pvrf,3)
names(df111)[9]<-"predvrf"
df111[,10]<-ifelse(df111$predvrf>0.6,1,0)
names(df111)[10]<-"probrf"
tblrf<-table(df111$dirchange,df111$probrf)
tblrf
acc<-sum(diag(tblrf))/sum(tblrf)
results[13,6]<-round(acc,7)
results[13,3]<-round(sqrt(df111.rf$mse[which.min(df111.rf$mse)]),7) #RMSE
results[13,4]<-round((df111.rf$mse[which.min(df111.rf$mse)]),7)     #MSE
```

```
results[13,5]<-round(mae(df111$lon, unlist(df111.rf$predicted,use.names = FALSE)),7)

# Logistic regression
df111.lr<-glm(dirchange~lon+lat,family="binomial",data=df111)
pvlr<-unlist(df111.lr$fitted.values,use.names = FALSE)
df111[,11]<-pvlr
names(df111)[11]<-"predvlr"
results[14,3]<-round(rmse(df111$predvlr,df111$dirchange),7)
results[14,4]<-round(mse(df111$predvlr,df111$dirchange),7)
results[14,5]<-round(mae(df111$predvlr,df111$dirchange),7)
df111[,12]<-ifelse(df111$predvlr<0.65,0,1)
names(df111)[12]<-"problr"
tbllr<-table(df111$dirchange,df111$problr)
tbllr
acc<-sum(diag(tbllr))/sum(tbllr)
results[14,6]<-round(acc,7)

#Naive Bayes
library(naivebayes)
df111.nb <- naive_bayes( factor(dirchange)~lon+lat, data = df111, usekernel = T)
df111.nb
plot(df111.nb)
pvnb <- predict(df111.nb, type = 'prob')
k<-length(pvnb[,1])
for (i in 1:k) {
  df111[i,13]<-max(pvnb[i,])
  if (max(pvnb[i,]) == pvnb[i,1]) df111[i,14]<-0
  if (max(pvnb[i,]) == pvnb[i,2]) df111[i,14]<-1
}
names(df111)[13]<-"predvnb"
names(df111)[14]<-"probnb"
tblnb<-table(df111$dirchange,df111$probnb)
```

```
tblnb
acc<-sum(diag(tblnb))/sum(tblnb)
results[15,6]<-round(acc,7)
results[15,3]<-round(rmse(df111$predvnb,df111$dirchange),7)
results[15,4]<-round(mse(df111$predvnb,df111$dirchange),7)
results[15,5]<-round(mae(df111$predvnb,df111$dirchange),7)

#Gradient boosting
library(gbm)
df111.gb <- gbm(dirchange~lon+lat, data = df111)
plot(df111.gb$fit)
pvgb<-df111.gb$fit
df111[,15]<-pvgb
df111[,16]<-ifelse(df111.gb$fit>0,1,0)
names(df111)[15]<-"predvgb"
names(df111)[16]<-"probgb"
tblgb<-table(df111$dirchange,df111$probgb)
tblgb
acc<-sum(diag(tblgb))/sum(tblgb)
results[16,6]<-round(acc,7)
results[16,3]<-round(rmse(df111$predvgb,df111$dirchange),7)
results[16,4]<-round(mse(df111$predvgb,df111$dirchange),7)
results[16,5]<-round(mae(df111$predvgb,df111$dirchange),7)

#SVM
library(e1071)
df111.svm<-svm(dirchange~lon+lat, data = df111)
df111.svm
pvsvm<-predict(df111.svm, df111)
plot(df111.svm,pvsvm)
df111[,17]<-pvsvm
df111[,18]<-ifelse(df111.svm$fit>0.6,1,0)
names(df111)[17]<-"predvsvm"
```

```
names(df111)[18]<-"probsvm"
tblsvm<-table(df111$dirchange,df111$probsvm)
tblsvm
acc<-1981/sum(tblsvm)
results[17,6]<-round(acc,7)
results[17,3]<-round(rmse(df111$predvsvm,df111$dirchange),7)
results[17,4]<-round(mse(df111$predvsvm,df111$dirchange),7)
results[17,5]<-round(mae(df111$predvsvm,df111$dirchange),7)

#MLP
library(nnet)
set.seed(123)
df111.dat<-df111[,c(3,4,8)]
df111.mlp<-nnet(dirchange~lon+lat, size=1,data = df111.dat,hidden = 1 )
pvmlp<-predict(df111.mlp, df111.dat[,-3])
df111[,19]<-pvmlp
df111[,20]<-ifelse(df111[,19]>0.6,1,0)
names(df111)[19]<-"predvmlp"
names(df111)[20]<-"probmlp"
tblmlp<-table(df111$dirchange,df111$probmlp)
tblmlp
acc<-1981/sum(tblmlp)
results[18,6]<-round(acc,7)
results[18,3]<-round(rmse(df111$predvmlp,df111$dirchange),7)
results[18,4]<-round(mse(df111$predvmlp,df111$dirchange),7)
results[18,5]<-round(mae(df111$predvmlp,df111$dirchange),7)

#############################################################################
        #Αλιευτικό ανοιχτής θαλάσσης
#Διαγράμματα κίνησης
df110 <- df[ which(df$mmsi=='259600076146315' ),]
trj <- TrajFromCoords(df110[,3:4])
plot(trj,main="Τροχιά κίνησης αλιευτικού πλοίου ανοικτής θαλάσσης",col="blue")
```

```
plot(trj,main="Εφαρμογή λείανσης",col="blue")
smoothed <- TrajSmoothSG(trj, p = 3, n = 31)
lines(smoothed, col = "#FF0000A0", lwd = 2)
legend("topright", c("Αρχική", "Λείανση"), lwd = c(1, 2),
       lty = c(1, 1), col = c("blue", "red"), inset = 0.01)



cdf2<- data.frame(longitude = df110$lat,latitude = df110$lon)
coordinates(cdf2) <- ~latitude+longitude
leaflet(cdf2) %>% addMarkers() %>% addTiles()

#Υπολογισμός αποστάσεων όλων των σημείων
df110 <- df[ which(df$mmsi=='259600076146315' ),]
df110<-df110[-(which(df110$speed == 0)), ]
k<-dim(df110)[1]

for (i in 1:k)
{
  if (df110[i,1] > 90) df110[i,1]<-df110[i,1]-90
  if (df110[i,1] < -90) df110[i,1]<-df110[i,1]+90
  if (df110[i,2] > 90) df110[i,2]<-df110[i,2]-90
  if (df110[i,2] < -90) df110[i,2]<-df110[i,2]+90
  if (i > 1) df110[i,5]<-0.539956803*distm(c(df110[i,4], df110[i,3]),
                                           c(df110[i-1,4], df110[i-1,3]),
                                           fun = distHaversine)/1000
  if (i > 1) df110[i,6]<-atan((df110[i,4] - df110[i-1,4]) / (df110[i,3] - df110[i-1,3]))



}

#Υπολογισμοί
names(df110)[5]<-"dist"
```

```
df110[5]<-unlist(df110[5],use.names = FALSE)
df110[,7]<-df110[5]/df110[2]
names(df110)[7]<-"hours"
names(df110)[6]<-"direction"
df110<-df110[-(which(df110$dist == 0)), ]
df110<-df110[-1,]
df110[,8]<-ifelse(df110[,6] > 0,1,0)
names(df110)[8]<-"dirchange"

#Random forest
library(randomForest)
set.seed(123)
df110.rf <- randomForest(dirchange~lon+lat, data = df110, importance = TRUE,
na.action = na.omit)
print(df110.rf)
plot(df110.rf)
pvrf<-unlist(df110.rf$predicted,use.names = FALSE)
df110[,9]<-round(pvrf,3)
names(df110)[9]<-"predvrf"
df110[,10]<-ifelse(df110$predvrf>0.6,1,0)
names(df110)[10]<-"probrf"
tblrf<-table(df110$dirchange,df110$probrf)
tblrf
acc<-sum(diag(tblrf))/sum(tblrf)
results[19,6]<-round(acc,7)
results[19,3]<-round(sqrt(df110.rf$mse[which.min(df110.rf$mse)]),7) #RMSE
results[19,4]<-round((df110.rf$mse[which.min(df110.rf$mse)]),7)    #MSE
results[19,5]<-round(mae(df110$lon,        unlist(df110.rf$predicted,use.names        =
FALSE)),7)

# Logistic regression
df110.lr<-glm(dirchange~lon+lat,family="binomial",data=df110)
pvlr<-unlist(df110.lr$fitted.values,use.names = FALSE)
```

```
df110[,11]<-pvlr
names(df110)[11]<-"predvlr"
results[20,3]<-round(rmse(df110$predvlr,df110$dirchange),7)
results[20,4]<-round(mse(df110$predvlr,df110$dirchange),7)
results[20,5]<-round(mae(df110$predvlr,df110$dirchange),7)
df110[,12]<-ifelse(df110$predvlr<0.65,0,1)
names(df110)[12]<-"problr"
tbllr<-table(df110$dirchange,df110$problr)
tbllr
acc<-sum(diag(tbllr))/sum(tbllr)
results[20,6]<-round(acc,7)

#Naive Bayes
library(naivebayes)
df110.nb <- naive_bayes( factor(dirchange)~lon+lat, data = df110, usekernel = T)
df110.nb
plot(df110.nb)
pvnb <- predict(df110.nb, type = 'prob')
k<-length(pvnb[,1])
for (i in 1:k) {
  df110[i,13]<-max(pvnb[i,])
  if (max(pvnb[i,]) == pvnb[i,1]) df110[i,14]<-0
  if (max(pvnb[i,]) == pvnb[i,2]) df110[i,14]<-1
}
names(df110)[13]<-"predvnb"
names(df110)[14]<-"probnb"
tblnb<-table(df110$dirchange,df110$probnb)
tblnb
acc<-sum(diag(tblnb))/sum(tblnb)
results[21,6]<-round(acc,7)
results[21,3]<-round(rmse(df110$predvnb,df110$dirchange),7)
results[21,4]<-round(mse(df110$predvnb,df110$dirchange),7)
results[21,5]<-round(mae(df110$predvnb,df110$dirchange),7)
```

```
#Gradient boosting
library(gbm)
df110.gb <- gbm(dirchange~lon+lat, data = df110)
plot(df110.gb$fit)
pvgb<-df110.gb$fit
df110[,15]<-pvgb
df110[,16]<-ifelse(df110.gb$fit>0,1,0)
names(df110)[15]<-"predvgb"
names(df110)[16]<-"probgb"
tblgb<-table(df110$dirchange,df110$probgb)
tblgb
acc<-sum(diag(tblgb))/sum(tblgb)
results[22,6]<-round(acc,7)
results[22,3]<-round(rmse(df110$predvgb,df110$dirchange),7)
results[22,4]<-round(mse(df110$predvgb,df110$dirchange),7)
results[22,5]<-round(mae(df110$predvgb,df110$dirchange),7)

#SVM
library(e1071)
df110.svm<-svm(dirchange~lon+lat, data = df110)
df110.svm
pvsvm<-predict(df110.svm, df110)
plot(df110.svm,pvsvm)
df110[,17]<-pvsvm
df110[,18]<-ifelse(df110.svm$fit>0.6,1,0)
names(df110)[17]<-"predvsvm"
names(df110)[18]<-"probsvm"
tblsvm<-table(df110$dirchange,df110$probsvm)
tblsvm
acc<-sum(diag(tblsvm))/sum(tblsvm)
results[23,6]<-round(acc,7)
results[23,3]<-round(rmse(df110$predvsvm,df110$dirchange),7)
```

```
results[23,4]<-round(mse(df110$predvsvm,df110$dirchange),7)
results[23,5]<-round(mae(df110$predvsvm,df110$dirchange),7)

#MLP
library(nnet)
set.seed(123)
df110.dat<-df110[,c(3,4,8)]
df110.mlp<-nnet(dirchange~lon+lat, size=1,data = df110.dat,hidden = 1 )
pvmlp<-predict(df110.mlp, df110.dat[,-3])
df110[,19]<-pvmlp
df110[,20]<-ifelse(df110[,19]>0.6,1,0)
names(df110)[19]<-"predvmlp"
names(df110)[20]<-"probmlp"
tblmlp<-table(df110$dirchange,df110$probmlp)
tblmlp
acc<-sum(diag(tblmlp))/sum(tblmlp)
results[24,6]<-round(acc,7)
results[24,3]<-round(rmse(df110$predvmlp,df110$dirchange),7)
results[24,4]<-round(mse(df110$predvmlp,df110$dirchange),7)
results[24,5]<-round(mae(df110$predvmlp,df110$dirchange),7)

#########################################################################
        #Παράκτιο Αλιευτικό
#Διαγράμματα κίνησης
df86 <- df[ which(df$mmsi=='187372144064677' ),]
trj <- TrajFromCoords(df86[,3:4])
plot(trj,main="Τροχιά κίνησης παράκτιου αλιευτικού πλοίου",col="blue")

plot(trj,main="Εφαρμογή λείανσης",col="blue")
smoothed <- TrajSmoothSG(trj, p = 3, n = 31)
lines(smoothed, col = "#FF0000A0", lwd = 2)
legend("topright", c("Αρχική", "Λείανση"), lwd = c(1, 2),
       lty = c(1, 1), col = c("blue", "red"), inset = 0.01)
```

```
cdf2<- data.frame(longitude = df86$lat,latitude = df86$lon)
coordinates(cdf2) <- ~latitude+longitude
leaflet(cdf2) %>% addMarkers() %>% addTiles()

#Υπολογισμός αποστάσεων όλων των σημείων
df86 <- df[ which(df$mmsi=='187372144064677' ),]
df86<-df86[-(which(df86$speed == 0)), ]
k<-dim(df86)[1]

for (i in 1:k)
{
  if (df86[i,1] > 90) df86[i,1]<-df86[i,1]-90
  if (df86[i,1] < -90) df86[i,1]<-df86[i,1]+90
  if (df86[i,2] > 90) df86[i,2]<-df86[i,2]-90
  if (df86[i,2] < -90) df86[i,2]<-df86[i,2]+90
  if (i > 1) df86[i,5]<-0.539956803*distm(c(df86[i,4], df86[i,3]),
                                          c(df86[i-1,4], df86[i-1,3]),
                                          fun = distHaversine)/1000
  if (i > 1) df86[i,6]<-atan((df86[i,4] - df86[i-1,4]) / (df86[i,3] - df86[i-1,3]))



}

#Υπολογισμοί
names(df86)[5]<-"dist"
df86[5]<-unlist(df86[5],use.names = FALSE)
df86[,7]<-df86[5]/df86[2]
names(df86)[7]<-"hours"
names(df86)[6]<-"direction"
df86<-df86[-(which(df86$dist == 0)), ]
df86<-df86[-1,]
```

```
df86[,8]<-ifelse(df86[,6] > 0,1,0)
names(df86)[8]<-"dirchange"

#Random forest
library(randomForest)
set.seed(123)
df86.rf <- randomForest(dirchange~lon+lat, data = df86, importance = TRUE,
na.action = na.omit)
print(df86.rf)
plot(df86.rf)
pvrf<-unlist(df86.rf$predicted,use.names = FALSE)
df86[,9]<-round(pvrf,3)
names(df86)[9]<-"predvrf"
df86[,10]<-ifelse(df86$predvrf>0.6,1,0)
names(df86)[10]<-"probrf"
tblrf<-table(df86$dirchange,df86$probrf)
tblrf
acc<-sum(diag(tblrf))/sum(tblrf)
results[25,6]<-round(acc,7)
results[25,3]<-round(sqrt(df86.rf$mse[which.min(df86.rf$mse)]),7) #RMSE
results[25,4]<-round((df86.rf$mse[which.min(df86.rf$mse)]),7)     #MSE
results[25,5]<-round(mae(df86$lon, unlist(df86.rf$predicted,use.names = FALSE)),7)

# Logistic regression
df86.lr<-glm(dirchange~lon+lat,family="binomial",data=df86)
pvlr<-unlist(df86.lr$fitted.values,use.names = FALSE)
df86[,11]<-pvlr
names(df86)[11]<-"predvlr"
results[26,3]<-round(rmse(df86$predvlr,df86$dirchange),7)
results[26,4]<-round(mse(df86$predvlr,df86$dirchange),7)
results[26,5]<-round(mae(df86$predvlr,df86$dirchange),7)
df86[,12]<-ifelse(df86$predvlr<0.65,0,1)
names(df86)[12]<-"problr"
```

```
tbllr<-table(df86$dirchange,df86$problr)
tbllr
acc<-sum(diag(tbllr))/sum(tbllr)
results[26,6]<-round(acc,7)

#Naive Bayes
library(naivebayes)
df86.nb <- naive_bayes( factor(dirchange)~lon+lat, data = df86, usekernel = T)
df86.nb
plot(df86.nb)
pvnb <- predict(df86.nb, type = 'prob')
k<-length(pvnb[,1])
for (i in 1:k) {
  df86[i,13]<-max(pvnb[i,])
  if (max(pvnb[i,]) == pvnb[i,1]) df86[i,14]<-0
  if (max(pvnb[i,]) == pvnb[i,2]) df86[i,14]<-1
}
names(df86)[13]<-"predvnb"
names(df86)[14]<-"probnb"
tblnb<-table(df86$dirchange,df86$probnb)
tblnb
acc<-sum(diag(tblnb))/sum(tblnb)
results[27,6]<-round(acc,7)
results[27,3]<-round(rmse(df86$predvnb,df86$dirchange),7)
results[27,4]<-round(mse(df86$predvnb,df86$dirchange),7)
results[27,5]<-round(mae(df86$predvnb,df86$dirchange),7)

#Gradient boosting
library(gbm)
df86.gb <- gbm(dirchange~lon+lat, data = df86)
plot(df86.gb$fit)
pvgb<-df86.gb$fit
df86[,15]<-pvgb
```

```
df86[,16]<-ifelse(df86.gb$fit>0,1,0)
names(df86)[15]<-"predvgb"
names(df86)[16]<-"probgb"
tblgb<-table(df86$dirchange,df86$probgb)
tblgb
acc<-sum(diag(tblgb))/sum(tblgb)
results[28,6]<-round(acc,7)
results[28,3]<-round(rmse(df86$predvgb,df86$dirchange),7)
results[28,4]<-round(mse(df86$predvgb,df86$dirchange),7)
results[28,5]<-round(mae(df86$predvgb,df86$dirchange),7)

#SVM
library(e1071)
df86.svm<-svm(dirchange~lon+lat, data = df86)
df86.svm
pvsvm<-predict(df86.svm, df86)
plot(df86.svm,pvsvm)
df86[,17]<-pvsvm
df86[,18]<-ifelse(df86.svm$fit>0.6,1,0)
names(df86)[17]<-"predvsvm"
names(df86)[18]<-"probsvm"
tblsvm<-table(df86$dirchange,df86$probsvm)
tblsvm
acc<-sum(diag(tblsvm))/sum(tblsvm)
results[29,6]<-round(acc,7)
results[29,3]<-round(rmse(df86$predvsvm,df86$dirchange),7)
results[29,4]<-round(mse(df86$predvsvm,df86$dirchange),7)
results[29,5]<-round(mae(df86$predvsvm,df86$dirchange),7)

#MLP
library(nnet)
set.seed(123)
df86.dat<-df86[,c(3,4,8)]
```

```
df86.mlp<-nnet(dirchange~lon+lat, size=1,data = df86.dat,hidden = 1 )
pvmlp<-predict(df86.mlp, df86.dat[,-3])
df86[,19]<-pvmlp
df86[,20]<-ifelse(df86[,19]>0.6,1,0)
names(df86)[19]<-"predvmlp"
names(df86)[20]<-"probmlp"
tblmlp<-table(df86$dirchange,df86$probmlp)
tblmlp
acc<-sum(diag(tblmlp))/sum(tblmlp)
results[30,6]<-round(acc,7)
results[30,3]<-round(rmse(df86$predvmlp,df86$dirchange),7)
results[30,4]<-round(mse(df86$predvmlp,df86$dirchange),7)
results[30,5]<-round(mae(df86$predvmlp,df86$dirchange),7)

#############################################################################
 #ANOVA
install.packages("aov")
install.packages("dplyr")
install.packages("readxl")
library("readxl")
library("dplyr")
library("aov")

dfres<-data.frame(results)
type<-factor(dfres$X1)
dfres$X1<-type
names(dfres)[1]<-"Type"
method<-factor(dfres$X2)
dfres$X2<-method
names(dfres)[2]<-"Method"
names(dfres)[3]<-"RMSE"
names(dfres)[4]<-"MSE"
names(dfres)[5]<-"MAE"
```

```
names(dfres)[6]<-"Accuracy"
attach(dfres)
summary(aov(RMSE~Type))
summary(aov(MSE~Type))
summary(aov(MAE~Type))
summary(aov(Accuracy~Type))
summary(aov(RMSE~Method))
summary(aov(MSE~Method))
summary(aov(MAE~Method))
summary(aov(Accuracy~Method))

require(dplyr)
dfres %>%
  group_by(Type) %>%
  summarise(sd = sd(RMSE))
```