ESTIMATION OF SHIP ROUTE USING DATA MINING TECHNIQUES

CUMHUR KIZILKAYA

HAZİRAN, 2022

ESTIMATION OF SHIP ROUTE USING DATA MINING TECHNIQUES

BAHÇEŞEHİR ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ
YÜKSEK LİSANS TEZİ

CUMHUR KIZILKAYA

BÜYÜK VERİ ANALİTİĞİ VE YÖNETİMİ YÜKSEK LİSANS DERECESİ İÇİN GEREKLİ ÇALIŞMALAR YERİNE GETİRİLMİŞTİR

HAZİRAN, 2022

BAHCESEHIR UNIVERSITY

GRADUATE EDUCATION INSTITUTE

...../...../.....

MASTER THESIS APPROVAL FORM

Program Name:	Big Data Analytics And Management
Student's Name and Surname:	Cumhur KIZILKAYA
Name Of The Thesis:	Estimation of Ship Route Using Data Mining Techniques
Thesis Defense Date:	

This thesis has been approved by the Graduate Education Institute, which has fulfilled the necessary conditions as a Master thesis.

Prof. Dr. Ahmet ÖNCÜ

Institute Director

This thesis was read by us, quality and content as a Master's thesis has been seen and accepted as sufficient.

	Title/ Name Surname	**Signature**
Thesis Advisor's:	Dr. Selmin DANIŞ ÖNCÜL	
Member's:	Prof.Dr. Gül TEKİN TEMUR	
Member's:	Dr. Şeniz DEMİR	

All information in this thesis is obtained and presented in accordance with academic rules and ethical principles; I also declare that I have made all references not cited by this study as required by these rules and principles.

Name, Surname : Cumhur KIZILKAYA
Signature :

ABSTRACT

ESTIMATION OF SHIP ROUTE USING DATA MINING TECHNIQUES

Kızılkaya, Cumhur
Master Program in Big Data Analytics And Management
Thesis Supervisor: Dr. Selmin Danış ÖNCÜL

May 2022, 35 Pages

With the development of sea transportation, the dangers such as sea accidents and environmental pollution occur as a result of increased sea traffic. It is very important to examine the sea traffic density in the area of interest and to increase situational awareness about that area. The need for a variety of auxiliary analytical tools to increase situational awareness and maritime safety at sea is increasing in each day. The ship uses many data sources while afloat. One of them is the Automatic Identification System (AIS) device, which is widely used on ships. It is a device that publishes information such as ship type, id number, country, destination, estimated arrival time, location, speed, direction, cargo. Ship route planning is one of the most important issues to improve traffic safety and efficiency. The main purpose of this study is to develop a model to help safe navigation while afloat. In this study, a trajectory creation and estimation model is developed to assist operators on board. We recommend a route prediction method based on Automated Identification System (AIS) data. The AIS data set used in this study is unique. For the application area in this study, the Dardanelles Strait region, which has not been done before, is chosen. AIS messages is used to develop a system for safe navigation. The next location and sequential points are estimated using the DBSCAN algorithm to estimate ship orbits.

Keywords: Ship Route Analysis, DBSCAN, Vessel, AIS, Maritime

ÖZET

VERİ MADENCİLİĞİ TEKNİKLERİNİ KULLANARAK GEMİ GÜZERGAHI TAHMİNİ

Kızılkaya, Cumhur
Büyük Veri Analitiği ve Yönetimi Yüksek Lisans Programı
Tez Danışmanı: Dr. Selmin Danış ÖNCÜL

Mayıs 2022, 35 Sayfa

Denizyolu taşımacılığının gelişmesiyle beraber artan deniz trafiği sonucunda deniz kazası, çevre kirliliği gibi tehlikeler meydana gelmektedir. Bir bölgede bulunan deniz trafik yoğunluğunun incelenmesi ve o bölge hakkında durumsal farkındalığın artması oldukça önemlidir. Denizde durumsal farkındalığı ve deniz güvenliğini artırmak için çeşitli yardımcı analitik araçlara duyulan ihtiyaç her geçen gün artmaktadır. Gemi seyir halindeyken birçok veri kaynağını kullanmaktadır. Gemilerde yaygın olarak kullanılan Otomatik Tanımlama Sistemi (OTS) cihazı da bunlardan biridir. Gemi tipi, kimlik numarası, ülke, varış yeri, tahmini varış zamanı, konum, hız, yön, kargo gibi bilgileri yayınlayan bir cihazdır. Gemi rota planlaması, trafik güvenliğinin ve verimliliğinin artırılmasında en önemli konulardan biridir. Bu çalışmanın temel amacı, deniz sularında seyir yapılırken güvenli navigasyona yardımcı olacak bir model geliştirmektir. Bu çalışmada, seyir sırasında gemideki operatörlere yardımcı olmak için yörünge çıkarma ve tahmin etme modeli geliştirilmiştir. Otomatik tanımlama sistemi verilerine dayalı bir rota tahmin yöntemi öneriyoruz. Bu çalışmada kullanılan AIS veri seti özgündür. Bu çalışmada uygulama alanı olarak daha önce çalışılmayan Çanakkale Boğazı bölgesi tercih edilmiştir. AIS mesajları, güvenli navigasyon için bir sistemi geliştirmede kullanıldı. Gemi yörüngelerinin tahmini için DBSCAN algoritması kullanılarak bir sonraki konum ve sıranın tahminini yapıldı.

Anahtar Kelimeler: Gemi Rotası Analizi, DBSCAN, Gemi, AIS, Denizcilik

To my wife

ACKNOWLEDGEMENTS

I would like to thank my thesis advisor Dr. Selmin DANIŞ ÖNCÜL. I was always greeted by a smiling face and conscious ears listening patiently. She was managing the thesis professionally and gave me lots of helpful advice and tutelage.

I thank my team leader Dr. Emrah ERGÜL for his patience and research grants during my thesis.

I would like to thank my family for their support during my life time. I always feel their support in every way.

I would like to thank my wife, who is always by my side and make me feel her support until the end.

TABLE OF CONTENTS

TABLES

FIGURES

ABBREVIATIONS

AIS	Automatic Identification System
OTS	Otomatik Tanımlama Sistemi
DBSCAN	Density-Based Spatial Clustering of Applications with Noise
EMSA	European Maritime Safety Agency
IMO	International Maritime Organization
SOLAS	Safety of Life at Sea
ETA	Estimated Time of Arrival
MAE	Mean Absolute Error
VHF	Very High Frequency
GPS	Global Positioning System
COG	Course Over Ground
SOG	Speed of Ground

Chapter 1

Introduction

1.1 Overview

Sea transportation is widely used in the world. It is used not only for cruising purposes, but also for transporting cargo, transporting oil, liquids such as chemicals and oils. While maritime transportation costs less, its safety is controversial. Recently, it has been observed that the number of people who are concerned about maritime safety in both the civilian and military purposes have increased. The need for assisted analytical tools to increase situational awareness and maritime safety has been addressed by various platforms.

Increasing situational awareness and marine safety at sea is a very important issue. According to the European Maritime Safety Agency (EMSA)[1], the number of marine accidents has increased recently. According to the 2020 Marine accidents and incidents report[2], there were approximately 22000 accidents between 2014 and 2020.

There are many data sources used for the ship on while afloat. The Automated Identification System (AIS) is one of them (Turgut, 2019). The International Maritime Organization (IMO)[3] Maritime Safety at Sea (SOLAS) convention sets standards for the construction, equipment and other operations of the vessel (Mankabady, 1986). It is mandatory to equip ships with the AIS.

It is possible to monitor sea traffic and take measures on ship movements instantly, as well as to examine and analyze the movements of ships using past AIS data. With AIS data, the probabilities of collision and stranding can be calculated, sea traffic can be mapped, traffic-intensive areas can be uncovered and marine traffic analyses can be performed (Wu, Xu, Wang, Wang and Xu, 2017; Natale, Gibin, Alessandrini, Vespe and Paulrud,2015). Thus, it is possible to ensure that certain measures are taken in the region by using the information exploited for the decision support purposes.

[1] www.emsa.europa.eu Access date: November 2021
[2] http://www.emsa.europa.eu/accident-investigation-publications/annual-overview.html Access date: November 2021
[3] www.imo.org Access date: November 2021

The Dardanelles is one of the regions with high sea traffic. Many ships pass through this area. In this study, AIS data, mapping of Dardanelles Sea Traffic and analysis of ship distributions on the desired line were carried out. Then the routes followed by different types of ships were revealed and the next position of the ships was estimated.

The data-set used in this study is the data created from AIS. Essential data engineering and modeling studies have been carried out using python programming language.

The main purpose of this study is to develop a model to help safe navigation while afloat in sea waters. In this study, a model of orbital extraction and forecasting is developed to assist the operators on board during the afloat. By using past AIS messages, the model extracts ship orbits and estimates the ship's next location.

The contributions of this study to the studies in the literature are twofold. The first is that the AIS database used is unique and that the attribute selection is preprocessed on the AIS data such that each attribute is subjected to certain weighting. Thus, the accuracy analysis results in the route prediction model are improved. The second is a unique study for the Dardanelles Strait with a high accuracy rate.

The contributions of this study is a proposed AIS-based route mining method. The method includes data preprocessing, build similarity calculation, clustering, and route inference. Estimation of the next location and sequence are performed using the DBSCAN algorithm. This method is useful for understanding marine traffic patterns. It provides a reliable basis for route planning and provides a basis for ship abnormal behavior detection. With AIS data, the routes followed by the ships are revealed and the possible movements of the ships are determined and thus the operators contribute to decision support. Thanks to this working model, it can be ensured that ships trying to pass through the Dardanelles at the same time do not hit each other or that security personnel can quickly reach a ship that has an oil spill during the passage. In addition, ships are currently obliged to pass with the harbor pilot when passing through the Dardanelles. This scientific study helps ships to safely pass the Dardanelles without a harbor pilot.

In the first and second part of the study, issues such as monitoring and analyzing maritime traffic, route studies in the maritime sector, Automatic Identification System, AIS message data, trajectory forecasting were discussed.

In the third part, information about the theoretical background in the development of the model is given.

In the fourth and fifth parts, the proposed methods and results, the principle of operation of the model and the analysis samples obtained from the model are shown.

In the discussion and conclusion parts, the benefits of the developed model and its similarities with other studies are given.

1.2 Sample Route Studies in Maritime

There are many studies in the literature on finding the ship's roadmap with AIS data and examining maritime traffic. While some studies have been related to the monitoring of sea traffic in different ways or with data from AIS receivers in different locations (such as satellites) (Watagawa, Kobayashi and Wakabayashi, 2012); a number of other studies are related to mapping and analyzing regional traffic (Natale and others, 2015; Wu and others, 2017).

The data sent by the AIS device is highly valuable data for researchers and analysts, as it gives position information about the ship and gives retrospective information and the positions in the past of the ship (Seta and others, 2016:102). Thanks to this historical data showing the actual movements of the ship, you can get information about traffic (Lei, Tsai and Peng, 2016). The automatic delivery of this data to other ships and coastal stations increases marine safety and navigational effectiveness and ensures the protection of the environment (Watagawa and others, 2012).

Past AIS data can be analyzed using traffic flow simulations, as well as these data can be used for analysis in different ways.

For example: In their study, Suman, Nagarajan, Sha, Khanfir and Kobayashi (2012) uses AIS data to analyze the risk of shipwrecks in the Strait of Malakka with the help of the MATLAB program. As a result of their analysis, high-risk areas are determined in the Strait of Malakka (Suman and others, 2012).

Viran (2014) used the environmental tension model to map and analyze the risk of the southern region of the Bosphorus Strait. As a result of the his study using one day of AIS data, high risk areas are specified in the Bosphorus Strait. (Viran, 2014).

In some of the studies, AIS data is displayed and mapped (Jiacai, Qingshan, Jinxing and Zheping, 2012; Pallotta, Vespe and Bryan, 2013; Mustaffa and Ahmad, 2015; Natale and others, 2015; Shelmerdine, 2015). There is also a study in which mapping is carried out covering the whole world (Wu and others, 2017). These processes help to monitor ship movements, identify traffic-heavy areas, make decisions and gain insight into many other issues related to sea traffic. Thanks to the model developed in this study, it is possible to obtain information about the region by showing AIS data on the map. If the studies are looked at in more detail:

In study Jiacai and others (2012), an imaging model is created with AIS data and the maritime traffic status is evaluated. Marine traffic status was analyzed using AIS data in Xiamen Bay and Meizhou Wan. Their work will help with decision-making and regulating sea traffic.

Pallotta and others (2013) analyzed sea traffic with AIS data. In the study, AIS data from the past and the routes followed by the ships are removed and the possible movements of the ships are determined and thus the operator is contributed to the decision support.

The study by Mustaffa and Ahmad (2015) analyzed ship traffic in Klang Port and the Strait of Malakka using AIS data and mapped the routes taken by the ships and examined the maritime traffic density. With the 5-day-data they collected, the areas with high sea traffic were examined and this study was carried out as an example of the use of the AIS device.

In study Natale and others, (2015) analyzed fishing boats using AIS data and analyzed fishing activities by mapping the data. By mapping the movements of fishing boats, the environmental impact of fishing activities and economic studies related to fishing are contributed.

The study by Shelmerdine (2015) analyzed and mapped AIS data from 2013 around Shetland Island. With the help of ArcGIS, AIS data were analyzed to extract density map, ship tracks and afloat routes. In the study, AIS data was used to provide information to many areas in the marines.

In studyWu and others, (2017) mapped global sea density using AIS data from August 2012 to April 2015. Maps were created in three different spatial resolutions using approximately 20 billion-row data. In the study, the issues of ship density, traffic density, AIS data reception frequency were defined and equations were given to calculate. As a result of the studies on the mapping of the data and the time taken during the processing of the data, a global maritime traffic map was emerged.

1.3 Automatic Identification System

The Automated Identification System (AIS) works on AIS devices. The AIS device provides an automated and digital exchange of data/information between AIS stations to identify, locate and track ships through broadcasts using Very High Frequency (VHF) radio waves, a form of wireless communication.

The AIS is a device that shows detailed information about the ship. It includes the identification information of a ship to the ships around it, which allows the identification of that vessel, what the ship is doing, what it carries and where it is grouting to, etc. The AIS device allows a ship to be tracked by other surrounding vessels or stations. Some of the land stations have AIS devices in aircraft, helicopters, many floating and flying machines providing naval service. Thanks to the AIS device, ships at sea are monitored. Without the AIS, it would be very difficult for the ships to sail.

AIS information is updated by ships after each navigation, after each anchor departure, after docking at each port. AIS is actually the identity card of a ship. It is a kind of communication in the sea. If you want to communicate with a ship while sailing in the international waters, you must have an AIS device that allows us to access the name, size, course, speed, etc. of that ship. If it were not for the AIS device, there would be no knowledge of maritime traffic.

Historical AIS data seems to be an ideal source for reconstructing ships' motion trajectories and extracting shipping route information. However, there are a number of challenges in the analysis of marine orbital data. First, unlike the restricted movement

of vehicles in road networks, ships move relatively freely in the marine environment. There are major shipping channels recommended for ships to monitor, but it is difficult to define the normal movement of the vessel. Secondly, the refresh rate of the AIS is every few seconds or minutes, depending on the different mode of movement. The volume of AIS data to be processed is huge and complex. Traditional methods of analysis and evaluation are overloaded with dramatically increasing AIS data.

AIS publishes information such as ship type, id number, destination, estimated arrival time (ETA), location, speed, country, direction, and so on. AIS consists of three different types of messages. The first type of message includes the ship's location and speed information, the second type of message contains the ship's singular identifier information, the third type of message includes the type of ship and travel information.

The AIS device consists of three types. These are: Class A, Class B and receiving devices. Class A is mandatory for large commercial vessels. Class A transmits more information than Class B. Class B is usually used by small ships.

The signal of the AIS device has about a range of 20 miles. AIS messages are collected by AIS recipients onshore and stored in data centers. When the ships move away from the coast, AIS signals are sent to the satellites. Satellites transmit information to shore.

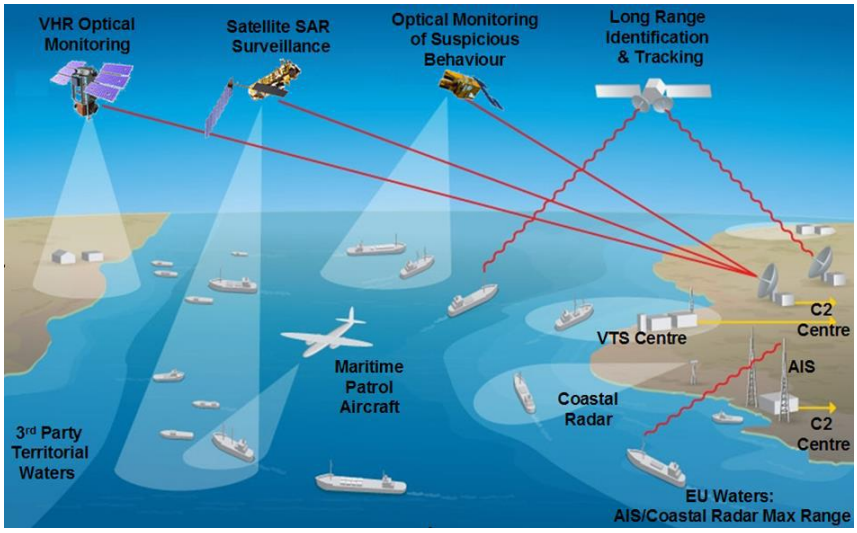

Figure 1. Marine surveillance architecture.

1.4 AIS Message Data

The data set used in this study is obtained from the data created by the Turkish Navy Command as unclassified to provide measurement data to universities.

AIS data publishes information such as ship type, id number, destination, estimated arrival time (ETA), location, speed, country, direction, and so on. There are about 570,000 lines. It contains AIS messages between 11 September and 24 September 2020.

- SourceId: Message source.
- Time: Time of receiving the message.
- TrackId: Unique identifier (system) in the system reporting the track.
- Latitude: Latitude of the ship coordinates.
- Longitude: Longitude of the ship coordinates.
- Xvelocity: The ship's velocity in the x-axis direction (m/s).
- Yvelocity: Velocity (m/s) in the y-axis direction of the ship.
- Zvelocity: Velocity (m/s) of the ship in the z-axis direction.
- HeightDepth: The depth or height (feet) of the ship.
- TrackNumber: The identifier (operator) in the system reporting the track.
- Environment: The environment in which the ship is located.
- AISTrackId: Mmsi number of the ship.
- IMONumber: The ship's IMO number.
- ShipName: The name of the ship.
- ShipCallsign: The ship's call name.
- ShipFlag: Country code of the ship.
- Destination: The ship's port of destination.
- ETA: The estimated time of arrival of the ship.

Table 1

Sample AIS Dataset

message	time	sourceId	trackId	aisTrackId	imoNumber	shipName	shipCallsign	shipFlag	destination	eta	latitude	longitude	xVelocity	yVelocity	zVelocity	heightDepth	trackNumber	environment
ais	11:58:05	1	2876976	271002673	9463176	KURTARMA 6	TCTM9	243	CANAKKALE	00/00 24:60	40.0277	26.2725	0.537839	0.148546	0	0	4345	surface
ais	14:21:16	1	2882341	215021000	9220550	SAPPHIRE	9HA4927	151	GELIBOLU	02/16 18:30	40.3202	26.5904	0.000000	0.000000	0	0	9945	surface
ais	12:04:39	1	2877255	271002673	9463176	KURTARMA 6	TCTM9	243	CANAKKALE	00/00 24:60	40.0294	26.2806	1.810030	0.922475	0	0	4754	surface
ais	15:03:34	1	2883908	215021000	9220550	SAPPHIRE	9HA4927	151	GELIBOLU	02/16 18:30	40.3202	26.5904	0.000000	0.000000	0	0	1166	surface
ais	15:20:37	1	2884623	271002673	9463176	KURTARMA 6	TCTM9	243	CANAKKALE	00/00 24:60	40.3782	26.6678	-3.890130	-5.335080	0	0	5661	surface
ais	19:47:41	1	2894726	271042963	9598593	KURTARMA 10	TCSG5	243	AMBARLI	01/01 00:00	40.4004	26.6804	-6.212850	1.537610	0	0	7667	surface
ais	19:48:58	1	2894795	271042963	9598593	KURTARMA 10	TCSG5	243	AMBARLI	01/01 00:00	40.4017	26.6751	-4.983970	2.291980	0	0	6930	surface

Chapter 2

Literature Review

In this section, I will explain similar studies on sea transportation. Sea transport has risks such as collision, fire and spillage of hazardous or pollutants. To mitigate these risks, marine surveillance data are collected in different criteria. In the literature, we can see the applications of basic statistical models and machine learning algorithms where AIS data is used to pre-identify these dangerous situations and to carry out controlled and safe journeys of ships. It seems that computer-aided systems and algorithms are becoming increasingly common to make meaningful inferences from this big data that arises with AIS technology.

The Markov model defines the transition from one state to another, with the ability to make the next state dependent only on the current state. The movements of the ships have the same characteristics. The location of a ship, the speed and the route, its next status related only to its location, current speed and current route on land. The Markov model is regularly used in the field of natural language processing. The model predicts the next letter or word based on the past training set. The implementation of the Markov model in location states and tracks is usually for GPS data. For the AIS data, only in its investigation of port management, it implements the Hidden Markov model and strengthened the port's surveillance (Deng and others, 2014).

In their study, Yitao and others (2020) proposed a satellite-AIS based route extraction method. The method includes data preprocessing, structural similarity calculation, clustering, and route inference. To verify the effectiveness of this method, real data from warrior strait collected from the seas near South Australia is used. However, the method of derailment does not perform well enough in distinguishing different routes in high-density areas, this may be because the clustering algorithm misclassifyes some sub-orbits.

Density-Based Spatial Clustering of Noisy Applications (DBSCAN) is a spatial clustering algorithm commonly used in many applications. It is capable of finding arbitrarily shaped clusters in the presence of noise data and performs well without prior knowledge of the number of clusters. To cluster spatial data, DBSCAN's properties determine its suitability. Therefore, the DBSCAN algorithm has been selected to cluster milestones (He, D. Zhang, J.F. Zhang, M.Y. Zhang and Li, 2019).

Although many scientists have studied ship route planning, the current research has the following problems. First, shipping lines to the ocean require a lot of satellite data support; also, if meteorological factors are not taken into account, the resulting ship route cannot meet safe navigation requirements; Secondly, due to the fact that coastal shipping lines are less affected by meteorological factors and have a large amount of AIS data, new requirements are required for ship speed and route from the perspective of ship traffic safety management. It is especially important to analyze and examine marine traffic and identify methods for the rapid and accurate completion of optimal road design in a complex navigation environment without relying on fine modeling of the marine environment (D. Zhang, Y. Zhang and C. Zhang, 2021).

2.1 Trajectory Prediction

This section touches on previous ship trajectory predicting studies. With the widespread implementation of the Automatic Identification System, large AIS data is generated. By analyzing the attributes of AIS data such as latitude and longitude, we can extract ship movement patterns, estimate movement status, and detect abnormal motion status.

Table 2 summarizes the studies on the ship route estimation literature. The most used algorithm model for the Ship Route estimation problem is the DBSCAN algorithm.

Table 2

Taxonomy of Researches About Vessel Route Estimation

Application areas	Algorithm	Citations
South Korea (longitude E 119:833 to E 122:667, latitude N 30:75 to N 32:417)	Markov Model	Deng and others (2014)
South Australian Nears	DBSCAN	Yitao and others (2020)
AIS Data of Cargo Ships in Port of Tianjin from 1 October 2017 to 10 October 2017.	DBSCAN	Sheng and Yin (2018)
West Coast of The United States	Extreme Learning Machine (ELM)	Mao and others (2016)
Chinese Zhoushan Islands from January To February 2015.	Spatial Logical Integrity	Zhao and others (2018)
AIS Data of The Three Gorges Dam Area	DBSCAN	He and others (2019)
Alaska-South Coast (zone1) and West Coast California (zone2)	LSTM, SRU, DBSCAN	Zhang and others (2021)
Cape Roca to the Ports of Lisbon, Setúbal and Sines.	Kernel Density Estimation Method	Rong and others (2020)
Yellow Sea and Bohai Sea Region on February 15, 2014.	DBSCAN	Li and others (2016)

Chapter 3

Material & Methods

This section provides information about the theoretical background under development of the model. AIS message clustering and evaluation are described in detail.

3.1 AIS Message Clustering

Some research suggests methods for using orbits derived from AIS data for marine traffic analysis; however, most of these studies are focused on exploring orbital models for identification of maritime traffic routes and anomaly detection. Several methods have been developed to facilitate sea traffic analysis based on trajectories collected from AIS data (Santos, Silva, Moura-Pires and Wachowicz, 2012). Clustering begins with an analysis of the position and direction of the ships to determine the routes. The authors propose a method for automatically detecting orbits and bringing them together in routes based on space-temporal information.

The AIS data processed contains more accurate longitude and latitude information at each moment of the ship and can be used to identify the ship's trajectory. Ship routes on short haul flights lack reference value and should be deleted. Also, due to the failure of the AIS equipment or the weak signal, there are very few AIS position points on the ship's course. Such routes may lack important information and also need to be deleted. The route identification method with less AIS information is to calculate the frequency of the route's AIS message. When the frequency is lower than a threshold, the route is considered to lack key information. The AIS interval varies between 2 seconds and 3 minutes for devices to send packages. This is determined by the route condition of the ship and is highly variable. Therefore, this article selects a series of ships with different speeds and counts the time interval during which the AIS equipment sends messages.

Cluster analysis is one of the primary methods used for data mining. It is used as a stand-alone tool for obtaining information about the distribution of a data set. It is used as preprocessing steps for focusing more analysis and data processing or for other algorithms running in detected clusters. Almost all well-known clustering algorithms

require input parameters that are difficult to determine but have a significant impact on the clustering result.

Our first approach to trajectory estimation is the clustering of AIS messages. The first step in this method is to cluster existing AIS messages. Next, we used classification methods to estimate the route of new AIS messages. We implemented DBSCAN (Ester, Kriegel, Sander and Xu, 1996), K-Means algorithms (MacQueen, 1967) for clustering. We used the Random Forest Algorithm (Breiman, 2001) to classify the new data.

3.1.1 Density Based Spatial Clustering of Applications with Noise

Density Based Spatial Clustering of Applications with Noise (DBSCAN) (Ester and others, 1996) is an unsupervised learning algorithm. It works based on the radius of a cluster and the minimum number of points in the clusters. This algorithm creates sets of elements based on the density of points in their neighborhood. DBSCAN is an algorithm based on a number of neighborhoods. This method clusters instances based on the density-availability relationship. DBSCAN can find outliers in the data. Distance value must be given in order to work with DBSCAN. Theoretically, the algorithm visits each point and detects its neighbors and clusters within the distance. All points are clustered by DBSCAN as core points, density-achievable points, or outliers, as shown in the following illustration. A point is inside a cluster that is a core point. Points where density can be reached are points that can be reached according to the core points within the distance. Other points are outliers. DBSCAN's basic algorithm architecture is given in Figure 2 and the pseudocode of DBSCAN algorithm is expressed in Table 3 (Ester and others, 1996).

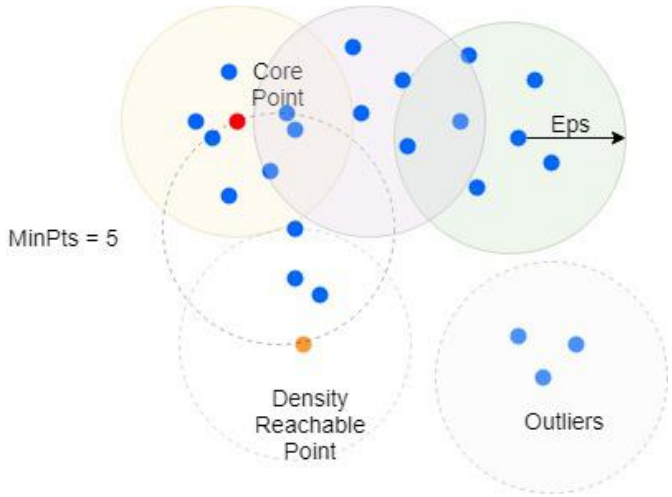

Figure 2. DBSCAN algorithm architecture

Table 3

DBSCAN Algorithm

Algorithm 1: DBSCAN (Ð, ε, MinPts)

```
C=0
for each unsinged point P_i in dataset Ð
    mark P_i as signed
    Neighbor (P_i) = RegionQuery ( P_i , ε)
    if size (Neighbor (P_i) < MinPts)
        mark  P_i as Noise
    else
        C = next cluster
        ExpandCluster( P_i , Neighbor (P_i), ε, MinPts, C)
    End if
End for
Algorithm ExpandCluster( P , Neighbor (P), ε, MinPts, C)
Add P to cluster C
For each point Ṕ in Neighbor (P)
    if Ṕ is not unsigned
        mark Ṕ as signed
        Neighbor (Ṕ) = RegionQuery ( Ṕ , ε)
        if size (Neighbor (Ṕ)) >= MinPts
            Neighbor (P)  joined with Neighbor (Ṕ)
            if Ṕ is not yet member of any cluster
                add Ṕ to cluster C
            End if
        End if
    End if
End for
Algorithm RegionQuery ( P , ε)
return all points within Ṕ s ε -neighborhood (including  P )
```

The DBSCAN algorithm does not require the specification of the number of clusters in the data a priori, however, there is no general way of choosing minPts and the radius ε. The setting of minPts requires some knowledge on the data density on study area. However, in the present case the algorithm has shown to be robust with respect to minPts, as similar clustering structures are obtained for a wide range of minPts values (between 5 and 10). The value of the radius ε is determined by a accessibility plot of the ordered points produced by the OPTICS algorithm (Ankerst and others, 1999) that is a plot showing on the x-axis the ordering of the points as processed by OPTICS and the accessibility distance on the y-axis. The accessibility plot of the ordered points enables the visualization of the clustering structure of the data set (Rong, Teixeira and Soares, 2020).

3.1.2 Linear Estimation Model

In this forecasting model, it is based on estimating and calculating the next route of ships using the SOG, COG, lat and long attributes in the motion data. This technique is one of the most widely used traditional methods. It is an approach that can be quite high in the strait crossings, especially when ship movements are linear or close to correct (Le-Tien & Phung-The, 2010). The Dardanelles strait is like that. In the tests, these qualities are evaluated only as they were and not tested by weighting them. Because lat, lng, SOG and COG are evaluated in trigonometric formulas in the accounts as given in the following so-called codes. The so-called code that calculates the next position of a ship is as in Table 4.

Table 4

The Pseudocode Calculating The Next Position of A Ship

Algorithm 2: Calculating the next position of the ship.
COG = from_degree_to_radia(COG)
new_x = (1,85 x SOG x cos(COG) x min) / 60
new_x = (1,85 x SOG x cos(COG) x min) / 60
new_lat = lat + (new_x /6378) x (180/3,14)
new_long = long+ (new_y /6378) x (180/3,14) / cos(long x 3,14 / 180)
target_coordinates = {new_lat, new_long}

The so-called code that calculates the distance between the two ships based on their lat and long information is as in Table 5.

Table 5

The Pseudocode Calculating The Distance Between Two Ships

Algorithm 3: Calculating the distance between two ships.

meter_constant = 6371

$\varphi 1$ = fromDegreeToRadia (lat1)

$\varphi 2$ = fromDegreeToRadia (lat2)

$\Delta\varphi$ =fromDegreeToRadia (lat2-lat1)

$\Delta\gamma$ = fromDegreeToRadia (lon2-lon1)

sin_cos_ multiplication = cos($\varphi 1$) x cos($\varphi 2$) x sin($\Delta\gamma$/2) x sin($\Delta\gamma$/2)

sin_cos_calue = sin($\Delta\gamma$/2) x sin($\Delta\gamma$/2) + sin_cos_ multiplication

arctan2_value = 2 x arctan2(sqrt(sin_cos_calue), sqrt(1 - sin_cos_calue))

distance difference = meter_constant x arctan2_value

3.1.3 Random Forest

Random Forest Algorithm (Breiman, 2001) is one of the supervised classification methods. It is a special form of the Decision Tree Algorithm. The Decision Tree Algorithm uses the entire data property to create a rule-based tree. However, the Random Forest Algorithm randomly selects properties to create multiple decision trees. The more trees in the forest, the stronger the model. It's important that there are more trees. The Random Forest Algorithm processes missing values. It can also work with categorical features. Another feature of the random forest model is that it gives us how important attributes are. The importance of an attribute depends on how much that attribute contributes to the explanation of variance in the dependent variable. We can give X number of attributes to the random forest algorithm and ask it to select the most useful Y and we can use this information in another model of our choice. For these reasons, the Random Forest algorithm was preferred. The Random Forest Algorithm architecture is given in Figure 3.

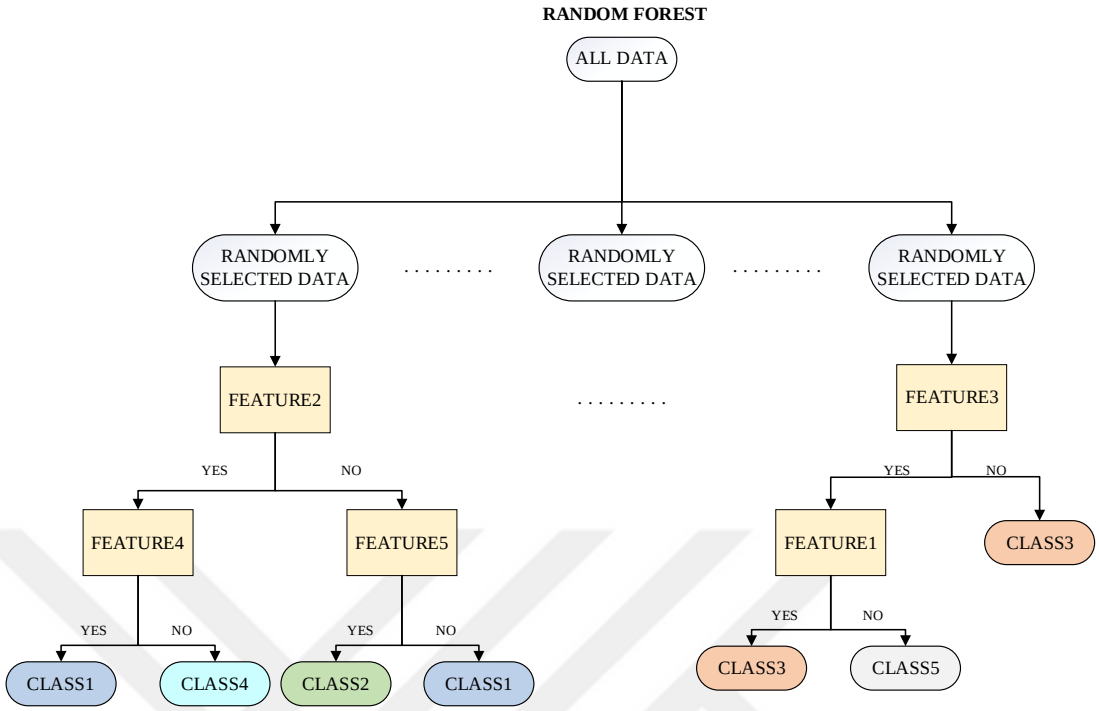

Figure 3. Random forest algorithm architecture.

3.1.4 K-Means

K-Means (MacQueen, 1967) is an unsupervised learning and clustering algorithm. The purpose of the algorithm is to find groups in the data. The K value in K-Means determines the number of clusters and should take this value as a parameter. This is actually a disadvantage. There is another algorithm called X-Means that calculates the K value by itself. The algorithm has a simple way of working.

K-Means works by analogy. It uses Euclidean Distance to measure similarity. The K-Means algorithm has two main steps. The first is to find the nearest centroid for each point. The second step is the centroid update. After the K value is determined, the algorithm randomly selects K center points. It calculates the distance between each data and randomly determined center points and assigns the data to a cluster according to the nearest center point. Then, a center point is selected again for each cluster and clustering is done according to the new center points. This situation continues until the system becomes stable.

3.1.5 Cosine Distance

The cosine distance is a measure for the similarity of the two vectors. It measures similarity by calculating the cosine of the angle between them. Cosine similarity measures the similarity between two vectors of an internal product area. It is measured by the cosine of the angle between the two vectors and determines whether the two vectors are marking in roughly the same direction. It only has positive values. The cosine distance method is used to determine the set of AIS messages.

$$1 - \cos(u, v) = 1 - \frac{\sum_{i=1}^{D}(\text{Ui x } Vi)}{\sqrt{\sum_{i=1}^{D} Ui^2}\sqrt{\sum_{i=1}^{D} Vi^2}} \quad \textbf{(3.1)}$$

3.1.6 Euclidean Distance

Euclidean distance is a measure of distance that measures the normal distance between the given two points. The euclidean distance between the two points in Euclidean is the length of a line segment between the two points. It can be calculated from the cartesian coordinates of points using pythagorean theorem, so it is sometimes called pythagorean distance. The formula for euclidean distance is below.

$$d = \sqrt{\sum_{i=1}^{n}(\text{X} - Y_i)} \quad \textbf{(3.2)}$$

3.2 Evaluation

In this study, we use evaluation criteria such as Mean Absolute Error, Accuracy, Precision, Recall and F1 score. First, metrics need true positive, true negative, false positive and false negative definitions.

It means that the true positive model predicts the correct value, which is the actual value.

The false positive is that the model estimates the correct value, but it is not the actual value.

True Negative means that the prediction is incorrect and in fact false.

False Negative prediction is incorrect. But the actual value is correct.

3.2.1 Accuracy

Accuracy shows the percentage of accurate prediction in general data. The accuracy formula equation is below.

$$Acc = \frac{TruePositive + TrueNegative}{\text{Total Data}} \quad \textbf{(3.3)}$$

3.2.2 Mean Absolute Error

Average Absolute Error shows how close our forecast results are to actual results. The negative value of the difference is positive due to the absolute value. Smaller Average Absolute Error means better guessing.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left(|Y_i - \overline{Y}_i|\right) \quad \textbf{(3.4)}$$

3.2.3 Precision

Precision shows how accurately the model predicts. Precision is the ratio of the value of true positive estimates to the value of overall positive estimates.

$$Precision = \frac{TruePositive}{\text{TruePositive+FalsePositive}} \quad \textbf{(3.5)}$$

3.2.4 F1 score

The F measure balances precision and recall results. Ranges 0 to 1. The F1 score equation is presented below.

$$F1Score = 2 * \frac{\text{Recall}*\text{Precision}}{\text{Recall+Precision}} \quad \textbf{(3.6)}$$

Chapter 4

Proposed Methods

In this section, orbital forecasting methods are explained and applications for route estimation are presented. AIS message clustering has been proposed to determine ship routes. The application language is designated Python because of its ease of use and the libraries it contains.

4.1 AIS Message Clustering

Using clustering methods, we've identified the ship's trajectory routes. So we remove the ship's trajectorys. Initially, we clustere latitude and longitude information from AIS messages with different algorithms. Therefore, density is very important to remove routes, where we use the DBSCAN (Ester and others, 1996) algorithm, an algorithm based on density. We use DBSCAN to cluster orbits. DBSCAN is an algorithm based on a number of neighborhoods. After completing the clustering process, a classification algorithm is applied. AIS messages in each cluster are used as inputs for the classification algorithm. Various classification algorithms are tested with AIS data to determine the supervised learning algorithm. The Random Forest Algorithm give the best results. The results are presented in Chapter 5. After classifying the data, routes for new locations are estimated.

Table 6

Clustering Algorithm

Algorithm 4: AIS Message Clustering Algorithm

```
data ◄── readAISMessagesFromDataset
train◄── data [Lat, Lon]
dbscan.fitPredict (train)
for cluster in cluster do
    trainX◄─ Lat, Lon
    trainY◄─ Destination
    randomForest.fit (trainX, trainY)
end for
```

As mentioned in the introduction, AIS transmits static data such as ship name, call sign and IMO (International Maritime Organization) number, as well as ship-related data including dynamic information such as speed and route. The types and content of AIS data are many, but our research only needs the most remarkable features. For example, from our point of view, the ship name, call sign, IMO and MMSI are similar in terms of identifying and recognizing the ship. Therefore, we are only able to use IMO, which is a 7 digit unique identification number, to represent the ship and edit it as a shi

Chapter 5

Experiments

Chapter 5 includes orbital prediction and the results of these orbital forecasting methods. First, data preprocessing methods are discussed. Afterwards, the prediction methods used are explained.

5.1 Trajectory Prediction

In this section, the results of various estimation algorithms are discussed. First, data preprocessing is performed as usual. Before proceeding to the estimation stage, different types of clustering and classification algorithms are applied on the data and are tested. Clustering have been applied as the estimation method.

By analyzing orbital data, we can discover motion behavior and location awareness information, and then develop many interesting applications such as motion behavior discovery, location prediction, traffic analysis, etc. However, orbital data mining is a challenging task due to the fact that orbital data is available with uncertainty. Also, due to the fact that the maritime area is the free movement area, it becomes even more difficult to discover valuable information from the sea orbit. Unlike the fact that the movement of vehicles is limited by road networks, in the maritime field there is no such sea route for ships to follow. The movement of one ship may not exactly repeat the same trajectory, even if the ship has similar movement behavior to the others. For maritime surveillance and traffic analysis, we can enable operators to better understand the movement of ships from ship trajectory data. Finally, the operators on board can be facilitated by making an orbital forecast to enable the operators on board to better understand the ship's trajectory data.

5.1.1 Data Preprocessing

Different types of classification methods have been experimented to reveal the route models used and produce an effective result. The data includes AIS messages from ships in Turkey and surrounding sea waters. The dataset contains 568015 AIS messages. Their distribution is in Figure 4. In the model study, the study area was narrowed by concentrating only on the Dardanelles Strait region. The dataset was obtained from the Naval Forces Command Surveillance Control Center as unclassified.

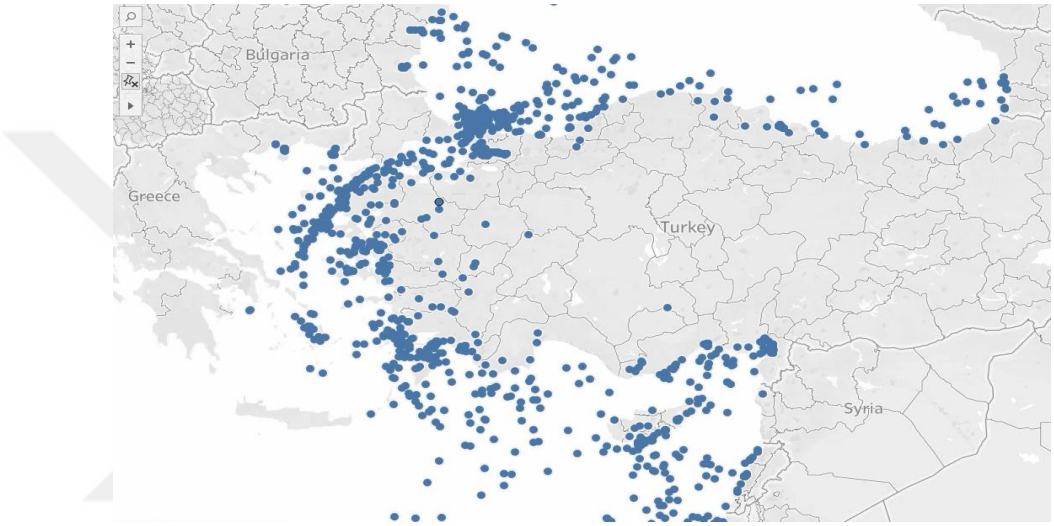

Figure 4. Presentation of all AIS messages

The raw version of the data set consists of three separate message file lines. These are: kinetic, characteristic, and ais message files. Messages in files must be processed sequentially to create an instant system state. In order for a trace to exist in the system at the time t, the kinetic and characteristic information about the trace must have been received and processed before t. traces that do not receive update information should be deleted at the end of a certain time.

Table 7

Dataset Specification

The Number of Messages	568015
Time Frame	From 11 Sept to 24 Sept 2020
The Number of Ships	2685

Initially, the dataset has noisy messages. There are empty messages and fields that the dataset generates. First, empty messages are eliminated from the dataset. Also, blank fields such as the reported draft are removed.

Currently, Automatic Identification System (AIS) is widely used in various types of ships. Outlier detection is a very important research aspect in data mining. An outlier is data that is not consistent with the overall behavior of the data or model (Hawkins, 1980). Figure 4 shows the AIS location points of all past ships between 11 September and 24 September 2020.

In this study, mainly dynamic and static AIS data are used. Dynamic data provides the latitude and longitude of the ship at any moment and can be used to map the ship's past routes. Static data is mainly used to classify ships(passenger ships, cargo ships, oil tankers, tugboats and official vessels).

Also, ship type is not required for first-stage subtraction patterns. The header indicates which direction a vehicle is facing. Therefore, it is constantly changing. It is more helpful to use the course value rather than the title value. The timestamp field is used for prediction.

5.1.2 Clustering Method

This section introduces the recommended route design method. This method obtains the ship's historical route through AIS data and then sets the optimization to design the route.

Clustering methods are applied to detect destination port clusters. After clustering, classification methods are executed to capture branches in orbital clusters. DBSCAN (Ester and others, 1996) is executed as a clustering algorithm. The Haversine distance was used as a distance measure. The epsilon metric, which is the radius of the clusters, is tested with different values. The parameters are presented in Table 8.

Table 8

DBSCAN Parameters

Eps (in meter)	50	500	5k	50k	500k
Silhouette Score	-0,36	-0,11	0,51	0,80	0,53

Initially we work with all datasets. The silhouette score for clustering is 0.13. Then we decided to reduce the data. After that, the first 10000 lines are used as input to DBSCAN. 1000 rows are randomly selected as test data. The latitude and longitude information is used in the input string. DBSCAN is clustered according to latitude and longitude densities.

After the clusters are observed, the classification algorithm Random Forest (Breiman, 2001) is applied. The purpose of the classification is to reveal the route branches in an orbital cluster. There are 10 clusters in 10000 ordered clusters. The data of each cluster is used as input to the Random Forest. Latitude, longitude, speed and route properties are classified by destination ports. Figure 5 shows the route branches in the Dardanelles.

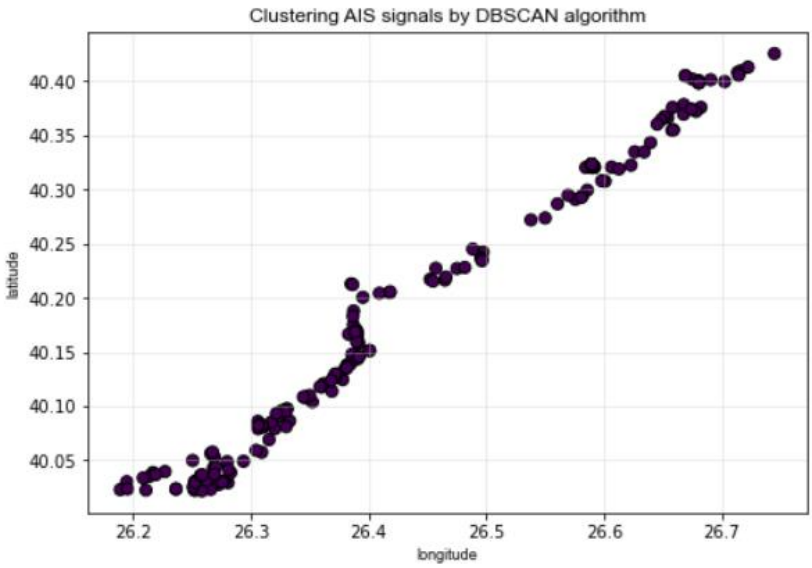

Figure 5. The route branches in dardanelles

Finally, 1000 rows of data are used to test the resulting orbital branches. Test data are not used as inputs for classification. Test data is read line by line. Next, the haversine distance between the center of gravity of the clusters and the position in each row is calculated. The cluster with the smallest distance is perceived as the cluster it belongs to. Then, the destination point is determined with the classification models of the cluster.

1000 AIS messages were used as test data. The test dataset have eight destination ports that are Gelibolu, Çanakkale, Lapseki, 1915 Çanakkale Prj, Çanakkale PS, Dardanelels, Dardanelles Northbou and Gelibolu Anchorage. For Çanakkale, the confusion matrix is presented in Table 9 and the results are presented in Table 10.

Table 9

The Confusion Matrix For Çanakkale Port

Çanakkale	**Positive**	**Negative**
True	100	600
False	150	150

Table 10

The Results For Çanakkale Port in Clustering Method

Port Name	**Accuracy**	**F1 Score**	**Precision**	**Recall**
Çanakkale	0.83	0.75	0.69	0.83

5.1.3 Prediction

In the estimation part, the clustering approach method is applied. First, we tried different supervised and unsupervised algorithms.

After the data cleanup procedures, several attempts have been made to estimate the destination port. First, various classifying algorithms are applied to the data. The arrival port is used as the forecast label, while trackId, imoNumber, shipFlag, latitude, longitude, and speed information are used for forecast input. The same data pertaining to the ship does not exist in both test and train.

As a result of our experimental study, the best estimation result is shown in figure 6, which is given by the Random Forest algorithm with an accuracy of 0,92.

```
import numpy as np
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.metrics import accuracy_score
from sklearn.ensemble import RandomForestClassifier
```

```
AISdata= AISdata.loc[(AISdata['latitude'] >=40.0200 ) & (AISdata['latitude'] <=40.4300 ) &
           (AISdata['longitude'] >=26.1800 ) & (AISdata['longitude'] <=26.7500 )]
AISdata.dropna(subset=['destination'], inplace=True)
```

```
x_train,x_test,y_train,y_test = train_test_split(x,y,test_size = 0.30, random_state= 42)
rf = RandomForestClassifier(n_estimators =100,  random_state= 42)
rf.fit(x_train, y_train)
y_pred = rf.predict(x_test)
accuracy_score(y_test,y_pred)
```

```
0.92
```

Figure 6. Modelling with random forest

Random Forest is used as the classification algorithm because the Random Forest algorithm gave the best prediction result. The next position is used as the prediction label, whereas latitude, longitude, speed and course is used for prediction input. In Table 11, the results are given.

Table 11

Next Position Prediction Accuracy

Algorithm	Accuracy
Random Forest	0.92
K Nearest Neighbors	0.83
Extra Trees Classifier	0.90
Bagging Classifier	0.88
Logistic Regression	0.44

5.2 Discussion

In this section, we compare the results of each method. Although there are several experiments on AIS messages, not all of them worked well. Existing work using the same AIS dataset is difficult to find. Therefore, some methods are evaluated with observations. The flowchart of the model is shown in Figure 7.

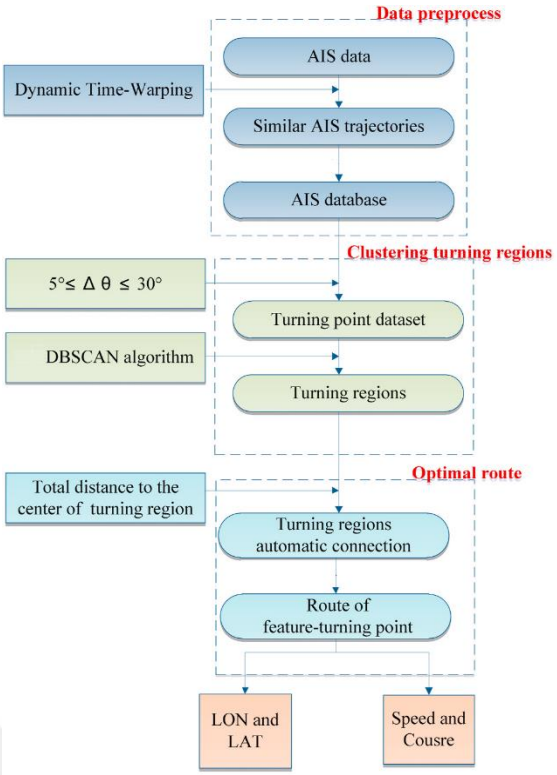

Figure 7. Flowchart of model

Studying and analyzing the navigational trajectory pattern of a particular target ship or target cluster can provide important assistance for maritime traffic management, channel clustering analysis and monitoring of key marine area. Experimental results using real AIS data show that this generated model can effectively dig the ship's typical trajectory sequence and avoid the impact of sudden situations such as interference points and transient channel change.

Basic clustering and classification methods are initially tried. The structure of the input data has been changed through methods. This has shown that it is more efficient to use AIS messages in sequences. Random AIS messages are inefficient as input for any algorithm. Also the best classification algorithm is Random Forest. The best clustering algorithm is DBSCAN. Therefore, the first port of arrival is used as a estimation method.

We created the route information based on Automatic Identification System (AIS) data. We made a prediction of the next position of the ship using the DBSCAN algorithm to make the ship trajectory estimation. Key results can be summarized as follows:

- The revised DBSCAN algorithm is well suited for exploring AIS data. Structural similarity measurement and hierarchical density estimates are created to automatically cluster AIS data on different orbital characteristics and overcome ship high-density limitations.
- The experimental results demonstrate the effectiveness of this ship orbital clustering model, which has a much lower computer time and expected result.

The results can be used to improve safety and security in the marine environment. First, the model may transform aggregate and complex data into reliable information about contingencies and help port authorities make relevant decisions in advance. This will help reduce maritime accidents. The cause of many maritime accidents is that operators are unable to anticipate possible situations in the area and react instantly to the current traffic situation. But the model in this study can decide on such issues in advance. For example, outliers as a result of clustering can inform decision makers in port of abnormal ship behaviour and give early warning to ships, which increases monitoring capability and improves the level of maritime situational awareness. It allows seafarers to be aware of the proposed route used by the majority of ships and avoid accidents caused by an unfamiliar navigational environment.

5.3 Implementation Environment and Complexity

The development environment is as follows; The system has 64 GB memory, i5 Intel processor running on Windows 10 operating system. We use Python 3.8.5 as the programming language. Jupyter is used for development in the Anaconda environment. We use the Scikit-Learn machine learning library to analyze AIS messages.

Chapter 6

Conclusion

The need for an auxiliary vehicle in the maritime field is inevitable. Streaming data is much more than operators can process at sea. Having a tool that captures and reports abnormalities overlooked by operators will help operators. Therefore, the study is organized for this purpose.

Orbital analysis and the extraction of orbital models are crucial to improving marine safety and marine situation awareness. The study estimated ship routes based on the discovery of the expected location after their current location.

It has been foreseen that algorithms will become difficult to use because AIS data is large in size. Therefore, only the Dardanelles Strait region has been studied by narrowing the field in the data set. Dardanelles AIS data from a certain time is used as a data set for use in the tests of this study. When looking at the results, it is observed that all approaches are close to each other and have high accuracy due to the fact that the ships in the Dardanelles have a linear movement structure. It has also been observed that ships travel in transit and often on a linear route in the Dardanelles.

Ship route planning is one of the key issues in increasing traffic safety and efficiency. Many route planning methods have been developed, but most of them are based on information in charts. This article proposes a method for creating the ship's course based on shipways used in the past. AIS data is processed to obtain the ship's past route information. The ship return point is evicted and clustered as nodes. The DBSCAN algorithm is used to create the optimized route. AIS data of ships in the Dardanelles Strait region were selected as case studies. The optimized trajectory of the ships is created, and also compared with the navigation trajectory of the real ship.

Unlike the fact that the movement of vehicles is restricted by road networks, ships move freely in the sea area. There is no such maritime network that ships can follow. Therefore, in this study, we developed a route prediction model by exploring the ship routes that ships often navigate from their sea orbits. The experimental results show that the Accuracy value of the model is 0.92.

The study estimates the next location and destination. The common clustering method detects message sets. Classification specifies branches in clusters. With the

first 10000 messages, there are ten clusters as a result of this method. Predictions of the next point and the next series for the coasts should be improved. However, the consequences are acceptable for the high seas.

In the model created in route prediction, the experimental results show that the accuracy value is 0,92. This is a rather high rate. In the literature, it has been observed that the accuracy value in similar studies is 0,65. With AIS data, the routes followed by the ships are revealed and the possible movements of the ships are determined and thus the operators contribute to decision support. Thanks to this working model, it can be ensured that ships trying to pass through the Dardanalles at the same time do not hit each other or that security personnel can quickly reach a ship that has an oil spill during the passage. In addition, ships are currently obliged to pass with the harbor pilot when passing through the Dardanelles. This scientific study helps ships to safely pass the Dardanelles without a harbor pilot.

The results contribute to a better understanding of maritime route patterns and assist maritime authorities and officers in stable and sustainable ship traffic management.

For further researchs deep learning model can be adapted and used in other seas and straits. In addition, it will consist of the automatic issuance of route models to increase the level of maritime surveillance and create a stable and sustainable port environment. We'll add more AIS information to the model to improve the completeness of the model. For example, in this model only the characteristics of position (latitude and longitude), direction and speed are taken into account. If we were to take into account the more contextual qualities of ship trajectory, the clustering result would be more accurate. Another goal would be to create a convenient method for pre-processing AIS data without expert knowledge in determining parameters. Route detection algorithms can be improved with streaming AIS messages. The algorithm improves when messages arrive. Moreover, other types of anomalies can be studied. For example: two or more ships meeting at the same point in the near future, not arriving at the destination port at the time of arrival, speed and location information mismatches, kinematic information incompatibility with navigational status information, location and speed, very low speed value, Unexplained high speed value anomalies may be examined in the future.

Morover further studies can improve on improving accuricy of the result. Better results can be achieved by keeping the time interval of the data in the data set longer. Other data mining and machine learning algorithms and methods will also be considered in the implementation of AIS data.

REFERENCES

Ankerst, M., Breunig, M.M., Kriegel, H.-P. & Sander, J. (1999). OPTICS: ordering points to identify the clustering structure. ACM SIGMOD Record 28 (2), 49–60.

Breiman, L. (2001). "Random forests," Machine Learning, vol. 45, pp. 5–32,

Deng, F., Guo, S., Deng, Y., Chu, H., Zhu, Q. & Sun, F. (2014). Vessel track information mining using AIS data.

Ester, M., Kriegel, H., Sander, J. & Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. In: Proceedings of Second International Conference on Knowledge Discovery and Data Mining, pp. 226–231.

European Maritime Safety Agency: www.emsa.europa.eu

Hawkins, D. M. (1980). "Identification of Outliers," 1. Monographs on Applied Probability & Statistics, 80(2):321-8.

He, Y. K., Zhang, D., Zhang, J. F., Zhang, M. Y., & Li, T. W. (t.y.). Ship Route Planning Using Historical Trajectories Derived from AIS Data. 8.

International Maritime Organization: www.imo.org

Jiacai, P., Qingshan, J., Jinxing, H., & Zheping, S. (2012). An AIS data Visualization Model for Assessing Maritime Traffic Situation and its Applications. Procedia Engineering, 29, 365-369. https://doi.org/10.1016/j.proeng.2011.12.724

Lei, P.R., Tsai, T.H. & Peng, W.C. (2016). Discovering Maritime Traffic Route from AIS Network. The 18th Asia-Pacific Network Operations and Management Symposium (APNOMS)

Le-Tien, T., & Phung-The, V. (2010). Routing and Tracking System for Mobile Vehicles in Large Area. 2010 Fifth IEEE International Symposium on Electronic Design, Test & Applications, 297-300. https://doi.org/10.1109/DELTA.2010.38

Li, Y., Zhang, Y., Zhu, F. (2016). The Method of Detecting AIS Isolated Information Based on Clustering and Distance.

MacQueen, J. B. (1967). “Some methods for classification and analysis of multivariate observations.”

Mankabady, S. (1986). THE INTERNATIONAL MARITIME ORGANIZATION, VOLUME 1: INTERNATIONAL SHIPPING RULES.

Mao, S., Tu, E., Zhang, G., Rachmawati, L., Rajabally, E., & Huang, G.-B. (2018). An Automatic Identification System (AIS) Database for Maritime Trajectory Prediction and Data Mining. In Cao, J., Cambria, E., Lendasse, A., Miche, Y. & Vong C. M. (Ed.), Proceedings of ELM-2016 (C. 9, ss. 241-257). Springer International Publishing. https://doi.org/10.1007/978-3-319-57421-9_20

Mapping Vessel Path of Marine Traffic Density of Port Klang, Malaysia using Automatic Identification System (AIS) Data. (2015). International Journal of Science and Research (IJSR), 4(11), 245-248. https://doi.org/10.21275/v4i11.NOV151099

Marıne Casualtıes And Incıdents. (2020). Prelımınary Annual Overvıew Of Marıne Casualtıes And Incıdents 2014-2020: http://www.emsa.europa.eu/accident-investigation-publications/annual-overview.html

Natale, F., Gibin, M., Alessandrini, A., Vespe, M. & Paulrud, A. (2015). Mapping Fishing Effort through AIS Data. PLOS ONE. 10(6).

Pallotta, G., Vespe, M. & Bryan, K. (2013). Traffic Knowledge Discovery from AIS Data. Proceedings of the 16th International Conference on Information Fusion.

Rong, H., Teixeira, A. P., Soares, C.G. (2020). Data mining approach to shipping route characterization and anomaly detection based on AIS data.

Santos, M.Y., Silva, J.P., Moura-Pires, J. & Wachowicz M. (2012). Automated traffic route identification through the shared nearest neighbour algorithm. In: Bridging the geographic information sciences. pp 231–248. Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-29063-3

Seta, T., Matsukura, H., Aratani, T. & Tamura, K. (2016). An estimation method of message receiving probability for a satellite automatic identification system using a binomial distribution model. 46 Scientific Journals of the Maritime University of Szczecin, 118(46), 101-107. https://doi.org/10.17402/125

Shelmerdine, R. L. (2015). Teasing out the Detail: How Our Understanding of Marine AIS Data Can Better Inform Industries, Developments, and Planning. Marine Policy. 54:17–25.

Sheng, P., & Yin, J. (2018). Extracting Shipping Route Patterns by Trajectory Clustering Model Based on Automatic Identification System Data. Sustainability, 10(7), 2327. https://doi.org/10.3390/su10072327

Suman, S., Nagarajan, V., Sha, O. P., Khanfir, S., & Kobayashi, E. (2012). Ship Collision Risk Assessment Using AIS Data. 10, 16.

Turgut, B. S. (2019). POTENTIAL BENEFITS OF SATELLITE-BASED AUTOMATIC IDENTIFICATION SYSTEM IN THE CONTEXT OF INTELLIGENT TRANSPORTATION SYSTEMS. 88.

Viran, A. (2014). İSTANBUL BOĞAZI YOĞUN TRAFİK BÖLGESİNİN (GÜNEY BÖLGESİ) OTOMATİK TANIMLAMA SİSTEMİ TABANLI RİSK HARİTASININ ÇIKARILMASI.[CREATION OF AUTOMATED IDENTİFİCATİON SYSTEM-BASED RISK MAP OF THE BOSPHORUS HEAVY TRAFFIC ZONE (SOUTHERN REGION).] 123.

Watagawa, M., Kobayashi, E. & Wakabayashi, N. (2012). Monitoring of Vessel Traffic Using AIS Data and ALOS Satellite Image. In 2012 OCEANS 2012. Yeosu. IEEE.

Wu, L., Xu, Y., Wang, Q., Wang, F., & Xu, Z. (2017). Mapping Global Shipping Density from AIS Data. Journal of Navigation, 70(1), 67-81. https://doi.org/10.1017/S0373463316000345

Yan, W., Wen, R., Zhang, A.N. & Yang D. (2016). Vessel movement analysis and pattern discovery using densitybased clustering approach. In: Proceedings of 2016 IEEE international conference on big data (Big Data), pp 3798–3806

Yitao, W., Lei, Y., & Xin, S. (2020). Route Mining from Satellite-AIS Data Using Density-based Clustering Algorithm. Journal of Physics: Conference Series, 1616(1), 012017. https://doi.org/10.1088/1742-6596/1616/1/012017

Zhang, D., Zhang, Y., Zhang, C. (2021). Data mining approach for automatic ship-route design for coastal seas using AIS trajectory clustering analysis.

Zhao, L., Shi, G. & Yang, J. (2018). Ship Trajectories Pre-processing Based on AIS Data.

APPENDICES