Data Analysis of AIS Data

Neha Devi Shakya

Internship Report

International Bachelor YGREC in Modelling and Data Science

Supervisor: Omar Eriksson

Advisor: Matthieu Cisel

Company:

International Association of Marine Aids to Navigation and Lighthouse Authorities (IALA-AISM),

Saint-Germain-En-Laye, France

Abstract

Navigational safety is one problem that frequently causes an accident, particularly a ship collision. To determine the likelihood of a collision event and comprehend its effects on human life, the shipping industry, and the environment, conducting a collision risk assessment in a waterway is essential. Due to the need for more accurate accident data, it could be challenging to assess the accident. As a result, a near-miss study can be viewed as a suitable replacement. The methodology suggested in this study is tested using a sample of a single week's worth of Automatic Identification System (AIS) data. Denmark is the target country for the near-miss analysis mentioned in this study. This paper utilises the closest point of approach (CPA) as the indicator to categorise the encounter as a safe or unsafe encounter, and it contains two parameters: the distance to CPA (DCPA) and the time to CPA (TCPA). We used the DCPA and TCPA thresholds suggested by Fukuto and Imazu.

Contents

1 Introduction

1.1 Introduction to the Maritime Navigation Industry, Aids to Navigation and Automatic Identification System (AIS)

The maritime navigation industry stands at the heart of the global economy, facilitating the movement of goods, raw materials, and people across vast oceans and waterways. It is a complex web of interconnected systems, technologies, and regulations that collectively enable safe and efficient maritime transportation. The industry's significance cannot be overstated, as approximately 80% of global trade by volume and over 70% by value are transported by sea. This intricate network encompasses shipping companies, port operators, navigation equipment manufacturers, maritime technology developers, and regulatory bodies such as the International Maritime Organization (IMO). The maritime navigation industry has undergone remarkable transformations throughout history. From the ancient use of stars for navigation to the modern use of satellite-based Global Positioning System (GPS) technology, advancements have continually improved maritime travel's accuracy, safety, and efficiency. With digitalisation, data analytics, and automation, the industry is poised for further innovation, including integrating artificial intelligence, crewless vessels, and smart port management systems. Amidst concerns about environmental sustainability, the industry is also exploring cleaner propulsion technologies and more efficient shipping practices to reduce its carbon footprint.

Aids to Navigation (AtoN) are essential tools that guide mariners through waterways and around potential hazards, ensuring safe navigation even in challenging conditions. These aids take various forms, including visual markers such as lighthouses and beacons, audible signals like foghorns, and electronic systems like radar and the Automatic Identification System (AIS). AtoN is an intricate network of reference points, enabling mariners to determine their position, chart courses, and avoid collisions. Lighthouses, dating back centuries, remain iconic symbols of AtoN. They emit distinctive light patterns that mariners recognise, guiding them away from dangerous coastlines or reefs. Buoys, floating markers anchored in strategic locations, convey information through their shape, colour, and sound signals. Modern technologies like AIS have revolutionised AtoN by providing real-time vessel information to mariners and traffic management centres, which enhances navigational safety by offering comprehensive situational awareness.

The Automatic Identification System (AIS) has emerged as a cornerstone technology in the realm of maritime navigation, revolutionising the way vessels communicate and enhancing navigational safety. AIS is a tracking and communication system that enables ships to transmit crucial data with other adjacent ships and coastal authorities, such as vessel identity, position, course, speed, and navigational status. It uses self-organising time-division multiple access (SOTDMA) technology and operates on Very High Frequency (VHF) radio bands to prevent data collisions. AIS technology relies on a network of transponders installed on ships and coastal stations. These transponders continuously broadcast information about the vessel's identity, position, speed, and course and navigational and static information such as vessel name, type, dimensions, and destination, which is received by nearby vessels, coastal stations and vessel traffic services (VTS), enabling mariners and authorities to track vessel movements, anticipate potential collisions, and facilitate effective traffic management. AIS data is collected through radio signals transmitted by the AIS transponders and received by nearby vessels and coastal base stations.

AIS messages are categorised into different types, each serving a specific purpose in enhancing maritime safety and communication. Some of the key types of AIS messages include:

- Class A Position Report: Provides real-time information about the vessel's position, speed, and course.
- Class B Position Report: Similar to Class A, but with lower transmission frequency, suitable for smaller vessels.
- Static Data Message: Contains vessel-specific information such as vessel name, type, dimensions, and destination.
- Voyage Data Message: Offers details about the vessel's voyage, including route, ETA, and draft.
- Safety-Related Message: Includes urgent safety information such as navigational warnings and distress calls.
- Aids to Navigation (AtoN) Report: Provides data about navigational aids such as buoys and lighthouses.
- Interrogation and Response: Allows authorities to query specific vessels for additional information.

The history of AIS dates back to the late 20th century when the International Maritime Organization (IMO) recognised the need for a standardised system to improve maritime situational awareness and reduce the risk of collisions at sea. The AIS concept gained momentum, and in 2000, the IMO mandated AIS as a mandatory requirement for all vessels over a certain size and certain types of vessels, such as passenger ships and commercial vessels operating in international waters, facilitating identification and location tracking on a global scale (IMO, 2003).

The AIS technology has witnessed continuous advancements since its inception, with the introduction of different classes of transponders to accommodate various types of vessels and improve compatibility. Furthermore, integrating AIS with satellite-based augmentation systems, such as the Global Navigation Satellite System (GNSS), has enhanced the accuracy and reliability of position information, contributing to improved navigational safety. In recent years, AIS data has become a valuable source of information for various maritime stakeholders beyond collision avoidance. Port authorities, maritime agencies, and researchers now use AIS data for vessel traffic analysis, route optimisation, environmental monitoring, and law enforcement. This expanded utility has led to integrating AIS data with other technologies, such as geographic information systems (GIS) and machine learning algorithms, for more comprehensive and insightful maritime analytics.

In conclusion, the Automatic Identification System (AIS) has transformed maritime navigation by enabling real-time tracking and communication, enhancing navigational safety and efficiency. Its journey from a conceptual solution to a globally adopted technology underscores its vital role in modernising the maritime industry and promoting safer and more efficient global maritime trade.

Implementing the Automatic Identification System (AIS) in Navigation and Aids to Navigation serves multiple critical objectives. Firstly, it significantly enhances navigational safety by providing real-time vessel information to mariners and traffic management centres. It enables early detection of potential collisions, especially in congested waterways and adverse weather conditions. Secondly, AIS aids in efficient maritime traffic management by allowing authorities to monitor vessel movements, identify hotspots, and optimise routes to prevent bottlenecks.

Moreover, AIS data is instrumental in post-incident analysis and investigation. In case of accidents or maritime incidents, AIS data can provide insights into vessel movements, speeds, and courses leading up to the event, which aids in determining the causes of accidents and developing preventive measures. Furthermore, AIS is a valuable tool for environmental protection, as it helps monitor vessel emissions, track the movement of hazardous cargo, and identify instances of illegal fishing or pollution. Integrating Automatic Identification System (AIS) data with Aids to Navigation presents many promising opportunities. One of the significant advantages is improved situational awareness for mariners. By accessing real-time AIS information, mariners can make informed decisions about course adjustments, avoid potential collisions, and navigate more efficiently, which leads to enhanced safety and reduced risk of accidents.

Additionally, AIS data can revolutionise traffic management in busy maritime areas. Port operators and maritime authorities can monitor vessel movements in real-time, allowing them to allocate berths, manage ship arrivals and departures, and optimise traffic flow, increasing port operations' efficiency and minimising vessel waiting times, reducing fuel consumption and emissions. Moreover, AIS data can be harnessed for predictive analytics and trend analysis. By studying historical AIS data, maritime stakeholders can identify patterns in vessel traffic, analyse seasonal variations, and predict congestion trends. This proactive approach enables better planning of navigation routes and resource allocation.

While utilising Automatic Identification System (AIS) data in Aids to Navigation offers numerous benefits, it also presents certain challenges. One primary challenge is data quality and reliability. AIS data is reliant on vessel compliance and accurate transponder information. Instances of deliberate data manipulation or malfunctioning transponders can lead to inaccurate or incomplete data sets, affecting the system's overall effectiveness. Data privacy and security are also paramount concerns. AIS data contains sensitive information about vessel movements and activities, raising questions about who can access it and how it is used. Ensuring data privacy while enabling legitimate use for navigational and safety purposes is a delicate balance that must be maintained.

Furthermore, the scalability of AIS infrastructure to accommodate the ever-increasing maritime traffic volume is a technical challenge. As the number of vessels equipped with AIS transponders grows, the data processing and management systems must be robust enough to handle the influx of information without compromising speed and accuracy.

In conclusion, integrating Automatic Identification System (AIS) data into Aids to Navigation presents a transformative opportunity for the maritime industry. By overcoming data quality, privacy, security, and scalability challenges, stakeholders can harness AIS's full potential to enhance maritime safety, traffic management, and environmental protection on the world's waterways.

1.2 Introduction of Organization

The International Association of Marine Aids to Navigation and Lighthouse Authorities (IALA) is a preeminent global non-profit organisation established in 1957 (IALA-AISM, n.d.). Founded to facilitate collaboration among entities involved in Marine Aids to Navigation (AtoN), IALA has evolved into a central hub that brings together various stakeholders, including authorities, manufacturers, consultants, scientific institutions, and training centres from diverse corners of the world. This convergence of expertise and knowledge sharing within IALA has led to the development of standardised practices that enhance maritime navigation's safety, efficiency, and environmental sustainability. IALA's core function lies in offering a platform for the exchange and comparison of experiences and achievements related to Marine Aids to Navigation. By fostering dialogue, IALA stimulates innovation and the implementation of best practices in this critical field. Over the years, IALA's efforts have contributed significantly to harmonising Aids to Navigation practices globally, ensuring consistency and compatibility across various regions.

The IALA Worldwide Academy was established in 2012 to complement IALA's strategic goals (IALA-AISM, n.d.). The Academy operates as an integral yet independently funded entity within IALA. Its mission revolves around education, training, and capacity-building to empower national authorities in providing effective Marine Aids to Navigation services. The Academy's objectives align with IALA's second strategic goal, which emphasises the development of a global network of efficient Marine Aids to Navigation services through capacity-building and knowledge-sharing. By focusing on education, training, and research, the Academy equips coastal States with the skills and expertise necessary to fulfil their obligations under international conventions. Through its activities, the Academy contributes to elevating the standards of Marine Aids to Navigation services globally, resulting in safer navigation, reduced maritime accidents, and improved environmental stewardship. The Academy's work resonates with IALA's commitment to fostering a maritime community prioritising safety, sustainability, and responsible navigation practices.

IALA's vision is deeply rooted in safety, economic viability, and environmental preservation principles. Its paramount objective is to facilitate vessels' secure and efficient movement through the standardisation and harmonisation of Marine Aids to Navigation worldwide. This vision aligns with the broader goal of safeguarding the maritime community and protecting the marine environment. To realise this vision, IALA engages in a multifaceted approach. It encourages its members to collaborate on harmonising Aids to Navigation practices, ultimately facilitating vessels' smooth and safe passage while minimising their environmental impact. By integrating the needs of mariners, technological advancements, and the requirements of AtoN authorities, IALA's technical committees work diligently to establish common best practices. These practices are documented in IALA Standards, Recommendations, Guidelines, and Model courses, which collectively provide a framework for developing and maintaining safe and effective Marine Aids to Navigation systems. Through these efforts, IALA plays a pivotal role in reducing marine accidents, enhancing the safety of lives and property at sea, and promoting the responsible use of the marine environment. They offer the following services:

- IALA Maritime Buoyage System (IALA MBS): IALA's most notable contribution is the IALA Maritime Buoyage System. This system standardises the design and application of navigation marks and signals, ensuring a consistent approach across different regions. As a result, vessels navigating international waters can rely on a unified set of navigational aids, enhancing their situational awareness and reducing the risk of accidents.

- Differential GPS System (DGPS): IALA's involvement in the Differential GPS System has led to significant advancements in navigation accuracy. By providing vessels with precise positioning information, DGPS improves navigational precision, particularly in areas where accuracy is crucial, such as narrow waterways and congested ports.

- Automatic Identification System (AIS): IALA has been instrumental in developing and promoting the Automatic Identification System. AIS enables real-time vessel tracking, enhancing collision avoidance and allowing maritime authorities to monitor vessel movements effectively. This technology has revolutionised vessel communication and safety at sea.

- VHF Data Exchange System (VDES): IALA's contributions extend to the VHF Data Exchange System, which facilitates maritime data exchange between vessels and authorities. VDES enhances communication, allowing seamless information sharing, improving situational awareness and supporting decision-making.

- Development of Vessel Traffic Services (VTS): IALA's dedication to enhancing vessel safety is evident in its support for developing Vessel Traffic Services. VTS systems provide critical information to vessel operators and traffic controllers, enabling the safe and efficient management of vessel movements in busy waterways.

- Education and Training: Recognising the importance of well-trained professionals in Marine Aids to Navigation, IALA established the IALA Worldwide Academy. This Academy focuses on education, training, and capacity-building to ensure coastal States meet their obligations under international conventions, such as SOLAS.

- Capacity Building: The Academy conducts workshops, seminars, and analytical missions to identify gaps in Aids to Navigation practices and provide guidance for improvement. It elevates the competencies of individuals and institutions responsible for delivering Marine Aids to Navigation services.

- Research and Development: The Academy fosters research and development by collaborating with maritime universities and relevant organisations. By identifying research topics and encouraging exploration, the Academy contributes to advancing the knowledge and practices in the field of Aids to Navigation.

IALA's primary aim is to foster the safe, economical, and efficient movement of vessels while upholding sustainability principles. It envisions successful voyages contributing to a sustainable planet, reflecting its commitment to maritime safety and environmental responsibility.

2 Work in the Company

My internship journey as a Data Analyst at IALA spanned five transformative months, from April to August 2023. During this time, I engaged in various tasks, projects, and immersive experiences that broadened my skill set and granted me valuable insights into the maritime industry and its technological landscape. This section outlines the various aspects of my work, including production tasks, self-directed learning, research and documentation, integration into the company, communication, and observation. The heart of my internship experience was analysing Automatic Identification System (AIS) data, a crucial aspect of maritime navigation. My colleague and I started with our initial assignment of decoding AIS messages and storing them in a Cassandra Database. This initial task not only familiarised us with the intricacies of the data but also set the foundation for subsequent advanced analyses. Table 1 gives an overview of the duration and working hours during the internship.

Start Date	3rd April 2023
End Date	31st August 2023
Total Duration	5 Months
Intern Position	Data Analyst
Mentor	Omar Eriksson
Office Hours	9:00 to 17:00

Table 1: Duration and Work week information

We were entrusted with specific tasks that posed real-world challenges as we delved deeper. My focus primarily revolved around detecting near-miss collision situations, paramount for maritime safety. The Closest Point of Approach (CPA) is an estimated point at which the distance between the own ship and another object target will reach the minimum value. To achieve the task, I used mathematical models to calculate the time to CPA (TCPA) and the distance to CPA (DCPA). On the other hand, my colleague was tasked to interpolate and predict ship trajectories, which was deemed complex as it demanded a good understanding of maritime dynamics and predictive analytics. Both tasks were executed using Python, a versatile programming language that effectively empowered us to process, analyse, and visualise the AIS data.

Given our limited knowledge of the maritime industry and AIS data, my colleague and I embarked on an intensive self-directed learning journey. Our primary resources were the Company's proprietary software documentation, IWRAP (IALA-AISM, 2023), and the AIVDM/AIVDO protocol decoding (Raymond, 2023) documentation. This concerted effort provided valuable insights into the inner workings of AIS messages, shedding light on the layers of information conveyed through this communication protocol and offering a comprehensive overview of the Company's prior undertakings. By immersing ourselves in maritime literature and industry publications, we also understood the broader context holistically. This knowledge proved instrumental in contextualising our analyses and solutions within maritime navigation challenges and trends.

To enhance our analytical arsenal, we delved into existing research on AIS data analysis, specifically academic papers focused on near-miss collision detection and ship route prediction, aligning with our designated tasks. These research papers and academic articles guided us towards established methodologies and approaches. Engaging in discussions with our Supervisor, we identified opportunities to implement research methodologies on the AIS data at our disposal. For instance, I replicated the approaches discussed in these research papers to detect near-miss collisions, which involved implementing complex algorithms to verify the results of the data they use and then adapting it to our specific context to assess the approach's viability within our data.

My internship at IALA was marked by seamless integration into the Company's dynamic professional ecosystem despite the hybrid work arrangement (combining on-site and remote work). We established a Microsoft Teams channel to facilitate continuous interaction, task updates, and collaboration. Our integration extended beyond technical collaborations. Weekly status meetings with our Supervisor, held every Friday, provided a platform where we presented our progress, highlighted upcoming tasks, and sought guidance from him. These interactions fostered a sense of accountability, enabling us to remain aligned with project goals and timelines. The Company's inclusive spirit was evident in various facets of our experience. Invitations to company lunches and dinner events showcased a welcoming atmosphere transcending professional boundaries. These social interactions facilitated informal knowledge sharing, enabling us to tap into the collective wisdom of seasoned professionals.

Throughout my internship, I gained valuable insights into the maritime industry, including risk assessment methodologies and the practical application of navigation-related software. The exposure to real-world projects, collaborative work environment, and integration into the Company's activities provided a holistic perspective on the industry's dynamics. In conclusion, my internship at IALA was marked by a dynamic and enriching experience encompassing diverse tasks, continuous learning, effective communication, and active participation within the Company's ecosystem. During my internship, I observed the Company's unwavering commitment to excellence and innovation. The Company's spirit was characterised by a relentless pursuit of innovative solutions, constantly challenging the status quo. Collaborative discussions and brainstorming sessions exemplified the Company's culture of open dialogue, where diverse perspectives converged to fuel creative problem-solving.

A relentless drive for capacity building and skill enhancement was woven into the Company's spirit. A notable and transformative highlight was our participation in the week-long IALA Risk Management Toolbox course, a learning opportunity conducted by the Company in July 2023. This course was designed to foster continuous professional development in the maritime industry. During this course, we had the privilege of interacting with a diverse spectrum of seasoned professionals in the maritime industry, many of whom boasted a wealth of experience far surpassing our own, enriching our learning journey with their insights. The course environment encouraged open discussions beyond the predefined curriculum by sharing our personal and professional experiences and perspectives.

Central to the curriculum was a comprehensive exploration of the IWARP software. Beyond technical nuances, we delved into the strategic deployment of the SIRA (Systematic Identification of Risk Assessment) and PAWSA (Probability Analysis of Waterspace Safety) methods. Although seemingly technical, these methodologies were presented in a manner that rendered them accessible even to those unfamiliar with the intricacies of the software and risk assessment procedures. The course design catered to professionals at varying levels of expertise, ensuring accessibility and inclusivity. The week-long course deepened our understanding of maritime risk management by delving into discussions, hands-on exercises, and real-world case studies. Beyond theory, it equipped us with practical skills to assess risk levels, evaluate mitigation strategies, and implement measures effectively. Armed with newfound insights and skills, we exited the course equipped to contribute to maritime safety.

Looking back at my internship at IALA, the multifaceted tasks entrusted to me and my colleague were complemented by self-directed learning, extensive research, seamless integration, and insightful observations, culminating in an experience far exceeding my initial expectations. Through my tenure at IALA, I was not merely exposed to the intersection of technology and maritime navigation; I was immersed in it. This immersion illuminated the pivotal role of data analytics in upholding maritime safety, effectively revealing how technology serves as the cornerstone of informed decision-making in this domain. While I conclude this chapter of my journey, it is with great excitement that I transition into a new phase, continuing my association with IALA through an apprenticeship contract as part of my pursuit of an MSc in Data in Finance at Albert School. I carry forward newfound technical skills and understanding and a profound appreciation for the complexities and opportunities that define the maritime industry.

3 Methodology

3.1 Data and Area of Interest

This study focuses on the Denmark waters (i.e. the Maidenhead Locator Grid JO) for decoding AIS data and Australia for detecting near-miss collision situations and predicting ship trajectories. The AIS messages provide latitude and longitude, facilitating ship identification in the designated areas. The dataset used encompasses AIS data from January 2018 to April 2018 and August 2018 to April 2019, culminating in a comprehensive ten-month database. The research paper for the ship domain theory approach implements their model on AIS data for the Port of Rio de Janeiro. On the other hand, the research paper for the CPA approach implements their model on AIS data for the Sudan Strait. So, I also use these areas to verify if my implementation of their method is correct.

3.2 Decoding AIS Data

AIS messages are commonly transmitted using the NMEA 0183 standard, developed and maintained by the National Marine Electronics Association (NMEA) to establish interface standards for marine electronic equipment. NMEA 0183 is a standard framework for exchanging marine instrument data between various onboard equipment. An illustrative example of an

AIS sentence is as follows:

”!AIVDM,1,1,,B,15NBj > PP1gG > 1PVKTDTUJOv00 < 0M,0*09”

This sentence represents a ”Position Report” message, conveying essential details about a vessel’s position, course, and speed. These digital AIS messages are structured with comma-separated fields containing distinct data types. For the given AIS sentence breakdown:

- The ”!AIVDM” denotes an AIS message in ”VDM” (VDO Message) format.
- ”1,1” signifies the total number of sentences in the message and the current sentence number. This message contains one sentence.
- The” ”field can contain the sequence number.
- ”B” indicates the communication channel (”B” channel).
- ”15NBj>PP1gG>1PVKTDTUJOv00<0M” is the payload encoded in ”Six-bit ASCII”, encompassing vessel identification, position, course, and speed.
- ”0*09” represents a checksum for sentence integrity.

The PyAIS Python library (Richter, 2022) plays a pivotal role in encoding and decoding AIS messages. Decoding AIS messages involves a three-step process: initially parsing the NMEA 0183 physical protocol layer, then parsing the inner protocol layer, and ultimately decoding the AIS payload based on its ID. The PyAIS library efficiently handles these tasks. Some AIS messages possess ”Gatehouse wrappers” containing additional information like time and checksums, which PyAIS processes using specific classes. The library offers accessibility to these wrapper messages for in-depth analysis. Internal message type 1 in Gatehouse wrappers encapsulates NMEA messages for seamless interpretation (IALA-AISM, 2023). This decoding and encapsulation capability of the PyAIS library greatly enhances the extraction and utilisation of AIS information for maritime applications. The internal message type 1 is specified as follows:

$PGHP, < msgtype > , < date format > , < country > , < region > , < pss > , < online data > , < cc > *hh < CR >< LF >
< NMEA message >

where,
msgtype is 1 for this message type (Gatehouse internal message type 1)
date format as specified above
country is the MMSI country code where the message originates from.

region is the MMSI number of the region
region is the MMSI number of the site transponder
online data buffered data from a BSC will be designated with 0. Online data with 1.
cc is the checksum value of the following NMEA sentence
**hh* checksum as described in /IEC 61162-1/
NMEA message is the NMEA message that follows immediately after the $PGHP sentence

Example:

$PGHP,1,2008,5,9,0,0,0,10,338,2,,1,09*17

”!AIVDM,1,1,,B,15NBj > PP1gG > 1PVKTDTUJOv00 < 0M,0*09”

Figure 1 shows a snippet of the provided AIS data.

```
\s:Cape Schanck A,c:1665792008,T:2022-10-15 00.00.08*78\!ABVDM,1,1,8,A,19NSadh00i:EflAawwUP6h0>0@59,0*78
\s:Cape Schanck A,c:1665792009,T:2022-10-15 00.00.09*78\!BSVDM,1,1,,A,14`Uu`001R:G7A?bAvc1?Pvl00S4,0*58
\s:Cape Schanck A,c:1665792009,T:2022-10-15 00.00.09*78\!BSVDM,1,1,,B,17P9OP0000:DnuMb<Soi>7lF0400,0*48
\s:Cape Schanck B,c:1665792009,T:2022-10-15 00.00.09*7B\!BSVDM,1,1,,B,17PCie0P00:F>bOb6jA8?gvB0400,0*05
\s:Cape Schanck A,c:1665792009,T:2022-10-15 00.00.09*78\!ABVDM,1,1,1,A,18JlSf800<:F7c9ariw3L6RD0D0V,0*53
\s:Cape Schanck A,c:1665792009,T:2022-10-15 00.00.09*78\!BSVDM,1,1,,B,19NSK4@02p:GHlGb7Sah6@4@0l0k,0*70
\s:Cape Schanck A,c:1665792009,T:2022-10-15 00.00.09*78\!ABVDM,1,1,2,A,39NSLsm000:Dsjkb=1jHJF<@0Dg:,0*73
\s:Cape Schanck A,c:1665792010,T:2022-10-15 00.00.10*78\!ABVDM,1,1,5,B,17Oj?9?000:F>8;b6i64BG4D2@6e,0*1B
\s:Cape Schanck A,c:1665792010,T:2022-10-15 00.00.10*78\!ABVDM,1,1,4,A,17P73@0036:GFuaajPR=1JF>0L0C,0*19
\s:Cape Schanck A,c:1665792011,T:2022-10-15 00.00.11*78\!ABVDM,1,1,6,A,17Of=60P00bF?LWb6fNcGwwp2d0L,0*5D
\s:Cape Schanck A,c:1665792013,T:2022-10-15 00.00.13*78\!BSVDM,1,1,,B,17P3wt0000:DnAib<@K334B`0<09,0*24
\s:Cape Schanck B,c:1665792013,T:2022-10-15 00.00.13*7B\!BSVDM,1,1,,B,18IT7B0vRQ:Fsfeb4aPT6S@J00Rf,0*7C
\s:Cape Schanck A,c:1665792014,T:2022-10-15 00.00.14*78\!ABVDM,1,1,,A,13Q;83002I:C5uSagFmHdW6H0400,0*73
\s:Cape Schanck A,c:1665792014,T:2022-10-15 00.00.14*78\!BSVDM,1,1,,A,14hDFf001t:G<b1b4ABJl8hD0400,0*02
\s:Cape Schanck A,c:1665792014,T:2022-10-15 00.00.14*78\!ABVDM,1,1,1,A,17P0?e0001:FV6ab45W:D36L088d,0*0B
\s:Cape Schanck A,c:1665792014,T:2022-10-15 00.00.14*78\!BSVDM,1,1,,A,19NSK4@02p:GHlwb7Tlh6@4L0d0l,0*4A
\s:Cape Schanck A,c:1665792016,T:2022-10-15 00.00.16*78\!ABVDM,1,1,6,A,17Oe>w5OhtbFe8ib3GulpStR0D0S,0*15
\s:Cape Schanck A,c:1665792016,T:2022-10-15 00.00.16*78\!ABVDM,1,1,7,B,17OgKC0000:F?Jqb6fQ;7lcp0W3h,0*6B
\s:Cape Schanck A,c:1665792016,T:2022-10-15 00.00.16*78\!ABVDM,1,1,8,B,17P73@0036:GFpgajQOM2:FJ0400,0*75
\s:Cape Schanck B,c:1665792016,T:2022-10-15 00.00.16*7B\!ABVDM,1,1,4,B,18LdSJ01iPbGhE1amCPrU8PP2@9e,0*0D
\s:Cape Schanck B,c:1665792016,T:2022-10-15 00.00.16*7B\!BSVDM,1,1,,B,37Ohfb?P@ObDtk?b3qpE2gvP02wQ,0*3F
\s:Cape Schanck A,c:1665792018,T:2022-10-15 00.00.18*78\!ABVDM,1,1,1,B,14`Uu`001R:G7FEbAwOi?Pw60HDm,0*1D
\s:Cape Schanck A,c:1665792018,T:2022-10-15 00.00.18*78\!ABVDM,1,1,4,B,19NSadh00h:EflWb00;07h0T0D0p,0*4E
\s:Cape Schanck B,c:1665792018,T:2022-10-15 00.00.18*7B\!BSVDM,1,1,,B,19NSduP01l:FatUatIU;a9BP00RC,0*50
\s:Cape Schanck A,c:1665792018,T:2022-10-15 00.00.18*78\!BSVDM,1,1,,A,35UNur1000:FsKAb?F;McIhR0Dvb,0*3C
\s:Cape Wickham B,c:1665792019,T:2022-10-15 00.00.19*76\!ABVDM,1,1,9,B,13Q;83002B:C5o1agF9`cW6T0400,0*5D
\s:Cape Schanck A,c:1665792019,T:2022-10-15 00.00.19*78\!ABVDM,1,1,6,B,17Of=60P00bF?Lab6fNsH?wp2l0L,0*37
```

Figure 1: Snippet of encoded AIS data

We started by looping through all messages in each file and used the checksum value of each AIS message to check if the message was valid or not. A checksum value is a mathematical function used to verify the integrity of data transmitted over a communication channel. The checksum value is used to ensure that the AIS message received by a receiver is error-free and has not been corrupted during transmission. Only valid messages were decoded and added to the Cassandra database. When performing this validation, we lost about 20% of the messages.

3.2.1 Data Filteringand cleaning

This paper aims to find situations of near-miss collisions for cargo ships. Since two ships are involved in each collision, filtering in the data must be done to eliminate situations where non-cargo ships are involved in a near-miss collision. Each instance of near-miss collision in the database stores the MMSI code, type of ship and other information for all ships. It is important to reemphasise that we only consider the situation when a cargo ship collides with another ship, i.e. at least one of the ships must be a cargo ship. Many types of cargo ships are described in AIS. However, the AIS does not give all the information for every message. Sometimes, the type of the ship is not declared (or declared as NULL), or even the type of ship described is too general for classifying.

The first step to start filtering is selecting only near-miss collisions presented in the Denmark Area. The second step is eliminating Tug and Pilot boats. Considering those ships are made for contact with another ship, they will often be in a near-miss collision situation. So, all tug and pilot ships are excluded. After those two steps, defining which ship types are included in the analysis is important. With this purpose, all the remaining ship types were observed. They are listed in Table 2.

The table separates the types into three different categories:

- The cargo ship types are well-defined.
- Cargo ship types that need to be better defined or can be cargo must be verified.
- Ship types that are not cargo.

3.3 Detecting Near Miss Collision Situations: Time to and Distance at Closest Point of Approach (CPA)

The following section discusses detecting real near-miss collision situations using AIS historical traffic data. The CPA framework was developed to evaluate the collision behaviour of two moving objects (Sang et al., 2016). Two objects will arrive at the nearest point if their speed and course remain constant. The distance between vessels at CPA (DCPA) reflects how serious the hypothetical situation is. The remaining time for the two objects to reach CPA at constant speed and course is called the time to CPA (TCPA). Negative TCPA implies that items are moving away from one another. A near-miss collision situation is when two vessels will come within a hazardous distance of each other in the near future. These are identified for each candidate situation by an unsafe DCPA – below a set threshold – and a limited positive TCPA.

First, from our AIS data, we need position messages with the message's timestamp, MMSI

Cargo Ships	Possible Cargo or not	Non-cargo Ships
Bulk Carrier	Cargo Hazard A	Fishing
Chemical Tanker	Cargo Hazard A (major)	Hopper Barge
Container ship	Inland Unknown	Hopper Dredger
Crude Oil Tanker	NULL	Military Ops
Fruit Juice Tanker	Other	Offshore Supply Ship
General Cargo	Unspecified	Passenger Ship
LPG Tanker		Research Vessel
Oil Products Tanker		Research/Survey Vessel
Oil/Chemical Tanker		Sailing Vessel
Ro-Ro Cargo		Tanker Barge
Ro-Ro/Container Carrier		Work Vessel
Tanker		
Vehicles Carrier		

Table 2: List of ship types

number, latitude, longitude, speed over ground(SOG) and course over ground(COG). We only consider the messages where the ship coordinates lie within our study area. Second, we loop through our messages and the ship coordinate is used to model the ship and the next ship is used to model the target ship; the two ships should have different MMSI numbers.

Based on past research, the calculation is performed to assess whether the ship positions are classified as safe encounters or unsafe encounters. The CPA comprises two parameters: geospatial (the distance to CPA) and temporal (the time to CPA). Numerous DCPA and TCPA threshold settings are available, but for our purposes, we refer to Fukuto and Imazu's research, which found that the safe DCPA is 1.0 nm and the safe TCPA is 5 minutes. So, if our calculated TCPA and DCPA are less than these threshold values, we set the encounter as unsafe. The program automatically takes the first ship coordinates to model its own ship and takes the next ship to model the target ship. The calculation of TCPA and DCPA is based on the following equations (Prastyasari & Shinoda, 2020, p. 6):

$$TCPA(t) = -\frac{[(y_j - y_i)(\dot{y}_j - \dot{y}_i) + (x_j - x_i)(\dot{x}_j - \dot{x}_i)]}{(\dot{y}_j - \dot{y}_i)^2 + (\dot{x}_j - \dot{x}_i)^2}$$

$$DCPA(t) = \sqrt{[(y_j - y_i) + (\dot{y}_j - \dot{y}_i) \times TCPA]^2 + [(x_j - x_i) + (\dot{x}_j - \dot{x}_i) \times TCPA]^2}$$

where,

x_i, y_i: position of own ship in x and y-axis
x_j, y_j: position of target ship in x and y-axis
$\dot{x}_i, \dot{y}_i$: speed component of own ship in x and y-axis
$\dot{x}_j, \dot{y}_j$: speed component of target ship in x and y-axis

$$\dot{x} = SOG\, sin(COG) \times \frac{1852}{3600}$$
$$\dot{y} = SOG\, cos(COG) \times \frac{1852}{3600}$$

where,
SOG: the speed over ground of the ship
COG: the course over ground of the ship

3.4 Predicting Ship trajectories

In order to predict ship routes, there had to be previous data interpolating all the data points, for which we used cluster analysis and found groups for each route. In our case, a route contains n trajectories(sub-tracks) consisting of p latitude-longitude coordinate pairs and is placed as an $n \times 2p$ matrix represented as follows:

$$\begin{array}{cc} & 2p\, variables \\ n \text{ objects} & \begin{bmatrix} (x_{11}, y_{11}) & \dots & (x_{1f}, y_{1f}) & \dots & (x_{1p}, y_{1p}) \\ \vdots & & \vdots & & \vdots \\ (x_{i1}, y_{i1}) & \dots & (x_{if}, y_{if}) & \dots & (x_{ip}, y_{ip}) \\ \vdots & & \vdots & & \vdots \\ (x_{n1}, y_{n1}) & \dots & (x_{nf}, y_{nf}) & \dots & (x_{np}, y_{np}) \end{bmatrix}. \end{array}$$

We then calculated the distance between objects to quantify the dissimilarity between each object. We use the sum of haversine distances at the interpolation points:

$$d(I, j) = \sum_{f=1}^{p} distHaversine((x_{if}, yif), (x_{if}, y_{if})).$$

These inter-object distances are placed in an $n \times n$ distance matrix **D**.

To classify the routes into sub-tracks, we picked a port, flagged any ship within a 5 km distance, and started moving. We marked it as the start of the sub-track. Later, when the ship stops, we flag it as the end of the sub-track. We then use a Random Forest to predict a more optimised route for the ship. Our clustered route data is divided into train, validation, and test sets. We set aside 10% of the sub-tracks as a test set. The remaining sub-tracks are then divided, with 80% allocated to the training set and 20% to the validation set.

4 Results

4.1 Decoding Data

We were provided 373 GB of encoded data, of which we decoded around 10 GB, resulting in almost 60 GB of decoded data. It took us almost 4 months to achieve this. Figure 2 shows a snippet of the decoded data. We analysed only the AIS messages received between 23rd February 2021 and 4^{th} March 2021, where around 34.5 million position messages from 4839 ships were found. However, only around 7.6 million messages from 437 ships lie within our study area. From this number of ships, the own ship and the target ship are modelled for detecting near-miss situations. And Figure 3 shows the a heatmap of the ships in our area of study.

	timestamp_gh	mmsi	msg_type	lat	lon	speed	course	heading	ship_type	to_bow	to_stern	to_port	to_starboard
87682	2022-01-03 09:10:33.653	244670800	1	53.660263	14.517508	0.0	275.200012	511.0	70.0	69.0	11.0	4.0	4.0
215612	2022-01-03 09:10:33.653	261004860	1	54.795746	18.415932	0.0	0.000000	511.0	30.0	16.0	8.0	3.0	3.0
251036	2022-01-03 09:10:33.653	261010050	3	53.563034	14.590933	0.0	189.399994	249.0	52.0	15.0	19.0	5.0	4.0
402024	2022-01-03 09:10:33.653	261183730	2	53.428085	14.584202	0.0	360.000000	511.0	80.0	70.0	10.0	3.0	5.0
447536	2022-01-03 09:10:33.653	261185450	2	53.403099	14.605347	0.0	0.000000	511.0	79.0	53.0	12.0	7.0	1.0

Figure 2: Preview of decoded data

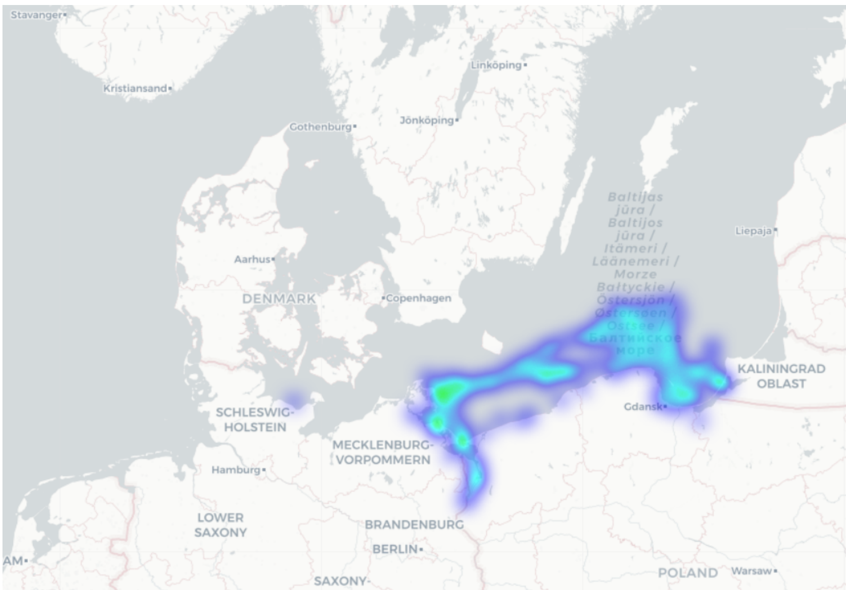

Figure 3: Heatmap in the area of study

4.2 Detecting Near-miss Collision Situations

4.2.1 For the data used in the research paper

As expected, the calculation resulted in 18,754 total ship encounters (NTE). Among these, 4,900 encounters exhibit negative CPA (NNE), implying that the ships were moving away from each other and can be classified as safe encounters. As mentioned, Fukuto and Imazu set the safe DCPA and TCPA thresholds at 1 nm and 5 minutes. As a result, this threshold's total number of near-miss encounters (NNM) was 5,870 involving 65 ships. Table 3 summarises the results of this calculation.

No.	Type of Encounter	No. of encounters for Fukuto and Imazu threshold
1	Negative encounter (NNE)	4,900
2	Safe encounter (NSE)	7,984
3	Near miss encounters (NNM)	5,870
4	Total encounter (NTE)	18,754
5	Ships involved in NNM	65

Table 3: Calculation result on research paper data

4.2.2 For our data

Implementing the same calculations to our data resulted in 7,599,938 total ship encounters. Among these, 3,175,047 encounters exhibit a negative CPA, implying that such encounters have crossed the crossing point and can be classified as safe encounters. However, we have yet to find any near-miss encounters. Table 4 summarises the results of this calculation.

No.	Type of Encounter	No. of encounters for Fukuto and Imazu threshold
1	Negative encounter (NNE)	3,175,047
2	Safe encounter (NSE)	4,424,891
3	Near miss encounters (NNM)	0
4	Total encounter (NTE)	7,599,938
5	Ships involved in NNM	0

Table 4: Calculation result on our decoded AIS data

5 Limitations

For the data decoding part, we could have made the compilation faster in several ways. Since the device we used did not have much memory space, we could only decode 1 million messages at a time so we had to upload the files in small chunks.

Also, since the code was in Python, the memory space taken by the code itself was huge. Also, we were using an inbuilt library, which was not optimised, and we did not have enough knowledge to make another library that worked the same way.

As shown in the results, we did not find any near-miss encounters. One reason could be that we only analysed one week's data. Another issue we faced was the processing power of the PC. We also had to analyse the data in chunks, so we do not have the possible ship interactions between the chunks.

6 Conclusion

There are multiple ways we could improve our code and results. One way is to vectorise our functions, as suggested by Vestre et al., 2021. We could also increase the time frame from 1 week to about 2-4 weeks. However, we could not achieve this due to the PC's processing power. As we used Python, the memory used by the code was huge, so we could process more data by switching the programming language to C++.

List of Figures

List of Tables

References

a. About iala. URL https://www.iala-aism.org/about-iala/.

b. AIVDM/AIVDO protocol decoding. URL https://gpsd.gitlab.io/gpsd/AIVDM.html.

c. Gh ais message format - iwrap. URL https://www.iala-aism.org/wiki/iwrap/index.php/GH_AIS_Message_Format.

d. Iala world-wide academy. URL https://academy.iala-aism.org/wwa/.

e. Iwrap. URL https://www.iala-aism.org/wiki/iwrap/index.php/Main_Page.

f. Traffic accidents involving trains and vessels. URL https://www.dst.dk/en/Statistik/emner/transport/trafikulykker/trafikulykker-med-tog-og-skib.

g. Welcome to Pyais's documentation! — Pyais 2.1.1 documentation. URL https://pyais.readthedocs.io/en/latest/.

h. Welcome to pymove's documentation! — PyMove 2019 documentation. URL https://pymove.readthedocs.io/en/latest/.

BHATTACHARJEE, SHILAVADRA. 2022. What is automatic identification system (Ais)- types and working(Faqs). URL https://www.marineinsight.com/marine-navigation/automatic-identification-system-ais-integrating-and-identifying-marine-communicatio

CEPEDA, MARICRUZ AURELIA FUN SANG; GABRIEL PREMOLI MONTEIRO; JOÃO VITOR MARQUES DE OLIVEIRA MOITA; und JEAN-DAVID CAPRACE. 2018. Estimating possible near miss collisions based on AIS big data for Port of Rio de Janeiro. *27º Congresso Internacional de Transporte Aquaviário, Construção Naval e Offshore*. Galoa. URL https://proceedings.science/proceedings/100040/_papers/87574.

IMO. 2003. Guidelines for the Installation of a Shipborne Automatic Identification System (AIS).

KRAETZIG, NIKITA MARWAHA. 2020. A complete guide to marine traffic tracking technologies and ais data. URL https://up42.com/blog/a-complete-guide-to-marine-traffic-tracking-tech-and-ais-data.

LEI, BAO. 2020. Apriori-based spatial pattern mining algorithm for big data. *2020 International Conference on Urban Engineering and Management Science (ICUEMS)*, 310–313. Zhuhai, China: IEEE. URL https://ieeexplore.ieee.org/document/9151537/.

PRASTYASARI, F I, und T SHINODA. 2020. Near miss detection for encountering ships in sunda strait. *IOP Conference Series: Earth and Environmental Science* 557.012039. URL https://iopscience.iop.org/article/10.1088/1755-1315/557/1/012039.

SANG, LING-ZHI; XIN-PING YAN; ALAN WALL; JIN WANG; und ZHE MAO. 2016. Cpa calculation method based on ais position prediction. *Journal of Navigation* 69.1409–1426. URL https://www.cambridge.org/core/product/identifier/S0373463316000229/type/journal_article.

SUN, FUMIN. 2016. Mining spatio-temporal ais data using prefix-span. *Proceedings of The fourth International Conference on Information Science and Cloud Computing — PoS(ISCC2015)*, 005. Guangzhou, China: Sissa Medialab. URL https://pos.sissa.it/264/005.

TRANSPORTATION RESEARCH BOARD. 2003. *Shipboard automatic identification system displays: meeting the needs of mariners – special report 273*. Washington, D.C.: Transportation Research Board. URL https://www.nap.edu/catalog/10708.

UNCTAD. 2022. Review of maritime transport 2022. URL https://unctad.org/system/files/official-document/rmt2022_en.pdf.

VESTRE, ARNSTEIN; AZZEDDINE BAKDI; ERIK VANEM; und ØYSTEIN ENGELHARDTSEN. 2021. AIS-based near-collision database generation and analysis of real collision avoidance manoeuvres. *Journal of Navigation* 74.985–1008. URL https://www.cambridge.org/core/product/identifier/S0373463321000357/type/journal_article.

YANG, DONG; LINGXIAO WU; SHUAIAN WANG; HAIYING JIA; und KEVIN X. LI. 2019. How big data enriches maritime research – a critical review of Automatic Identification System (Ais) data applications. *Transport Reviews* 39.755–773. URL https://www.tandfonline.com/doi/full/10.1080/01441647.2019.1649315.

ZHANG, WEIBIN; FLORIS GOERLANDT; PENTTI KUJALA; und YINHAI WANG. 2016. An advanced method for detecting possible near miss ship collisions from AIS data. *Ocean Engineering* 124.141–156. URL https://linkinghub.elsevier.com/retrieve/pii/S0029801816303080.