AIS Data-Based Maritime Statistics Analysis in the Strait of Malacca

Depandi Enda[1]*, Noerdin Basir[2], Boby Rahman[3] and I Made Surya Kumara[4]

[1] Informatics Department, Politeknik Negeri Bengkalis, Bengkalis, Indonesia
[2,3] Civil Engineering Department, Politeknik Negeri Bengkalis, Bengkalis, Indonesia
[4] Department of Computer Engineering, Warmadewa University, Bali, Indonesia
depandienda@polbeng.ac.id

Abstract. This study analyzes Automatic Identification System (AIS) data collected by the AIS-POLBENG Remote Base Station (RBS) in the Malacca Strait over a five-month period (February–June 2025) to support maritime traffic monitoring. The AIS data processing system was developed to encompass raw data recording, data cleaning and transformation, information extraction, data visualization and analysis. By utilizing open-source technologies such as PyAIS and PostgreSQL integrated with Python in Jupyter Notebook, the research generates strategic insights including vessel flag-state distribution, vessel types, navigational status, and traffic density. The analysis reveals that maritime traffic is dominated by foreign-flagged vessels, particularly those from Panama, Singapore, and the Marshall Islands, with cargo ships (54.49%) and tankers (24.02%) as the predominant vessel types. Most vessels were in the UnderWayUsingEngine status (84.88%), indicating high levels of engine-powered navigation activity. The average traffic density was recorded at 1.36 vessels/km² per month, totaling 6.83 vessels/km² during the observation period. These findings reaffirm the Malacca Strait as a high-intensity strategic shipping lane, underscoring the need for optimal traffic management to ensure navigational safety and efficiency.

Keywords: AIS Data, Malacca Strait, Maritime Statistic

1 Introduction

AIS (Automatic Identification System) is a technology used to track vessel movements through the transmission of ship data via Very High Frequency (VHF) radio waves (Ma'ruf et al., 2024). AIS technology is commonly used in conjunction with marine radar for maritime traffic monitoring to prevent vessel collisions. Each ship transmits information such as position, heading, and speed, which can be received by nearby vessels and by coastal monitoring stations equipped with Vessel Traffic Services (VTS). Based on the received information, a vessel can determine its position relative to other nearby ships, and coastal authorities can monitor maritime traffic within their jurisdiction (Triharjanto, 2019).

M. U. H. Al Rasyid et al. (eds.), *Proceedings of the International Conference on Applied Science and Technology on Engineering Science 2025 (iCAST-ES 2025)*, Advances in Engineering Research 283,
https://doi.org/10.2991/978-94-6463-926-1_92

The raw AIS data received by an AIS Receiver at a Remote Base Station (RBS) cannot be directly utilized by end users. Raw AIS data must first be decoded into plain text using specialized backend applications, such as AIS decoding software. Once decoded, the data can be stored in a database for future use. This process ensures that the information is accessible and usable for various maritime monitoring and analysis purposes (Enda et al., 2021). Once the historical data on vessel traffic movements has been stored, the next step is to further process or extract this data to obtain valuable information that supports effective maritime management. This stage involves analyzing patterns, identifying trends, and generating insights that can be used for strategic decision-making in maritime operations, surveillance, and spatial planning.

Several previous studies have explored the processing and utilization of AIS data in various use case scenarios. For instance, (Ma'ruf et al., 2024) identified shipping routes based on Maritime Mobile Service Identity (MMSI). Meanwhile, Spadon et al. (2024) conducted an analysis of maritime tracking data and its integration with AISdb, a system designed to address challenges in processing and analyzing Automatic Identification System (AIS) data used for maritime traffic monitoring.

Furthermore, AIS data can be utilized to estimate exhaust gas emissions from ship engines passing through specific waters, based on vessel type and cargo, to assess air pollution levels in a given region (Feng et al., 2022; Handani et al., 2018; Ikhsan et al., 2025; Ma et al., 2023; Ribeiro da Silva et al., 2024; Rong et al., 2024; Sudiantara et al., 2022; Xi et al., 2023; Zaman et al., 2015). Another use case of AIS data includes predicting coastal abrasion effects based on the number of vessels passing through an area, considering vessel type and size, as well as the estimated wave force impacting the shoreline.

AIS data can be used to visualize and identify various aspects such as maritime traffic density, shipping routes, navigation patterns, hierarchical structures of maritime routes, mapping of fishing zones and activities, spatiotemporal interactions among activities, and evaluation of accident risk indicators (Ferreira et al., 2022; Yan et al., 2024). The AIS team at the UN Global Platform also analyzes AIS data for use in diverse research fields, including migration, environmental studies, maritime and fisheries, as well as economics and trade. Additionally, AIS data can be processed and analyzed to generate relevant information that supports Maritime Spatial Planning (MSP) (Le Tixerant et al., 2018).

This study focuses on analyzing information obtained from AIS data extraction to determine the number of vessels based on unique MMSI, ship frequency by hour, day, month, and year, monthly ship arrivals, number of vessels by specific countries, number of vessels by vessel type, classification of vessels based on navigational status, and visualization of ship traffic density in the Malacca Strait region over a six-month period from January to June 2025. This research addresses the problem of how to extract maritime information in the Malacca Strait region using AIS data. Furthermore, it explores how to analyze the extracted data and visualize it through graphs that are useful for maritime purposes. The scope of this research is defined to ensure that the discussion remains focused and does not become overly broad. The limitations of this study are as follows; the recorded vessel traffic data is limited to maritime traffic within the Malacca Strait, the extracted and analyzed data consists of AIS data collected during the period from February to June 2025, and the vessel traffic data was recorded using a Remote Base Station (RBS) AIS device installed at the POLBENG campus.

2 Methodology

This study employs an experimental approach using AIS-POLBENG Remote Base Station data collected in the Strait of Malacca between January and June 2025. Raw AIS messages were extracted, cleaned, and transformed into structured datasets using a Python Jupyter Notebook-based preprocessing script. This study focuses on extracting AIS statistical data including the number of vessels by flag, navigation status, frequency, and traffic density. The data were analyzed through statistical summaries and geospatial visualization techniques to reveal maritime traffic density and movement patterns. The overall research workflow is illustrated in Figure 1.

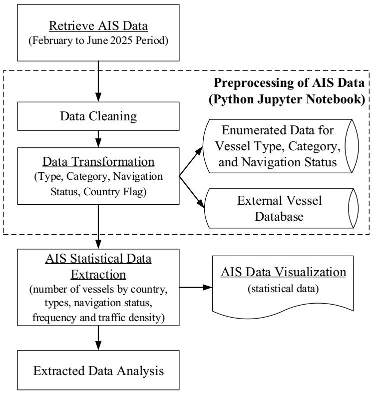

Figure 1. Methodology overview

2.1 Retrieve AIS Data

AIS data were obtained from a PostgreSQL database server via a backend Restful API, accessed through a web application. The dataset comprised records collected over a five-month period, from February to June 2025, and was stored in CSV format.

2.2 AIS Data Cleaning

Before further processing, the AIS data underwent a cleaning stage using a Python-based Jupyter Notebook environment. The valid value ranges for each AIS data field are presented in Table 1. Noise records defined as entries outside the valid range of AIS data fields (e.g., values below the minimum, above the maximum threshold, or null values) were removed to improve data extraction accuracy.

Table 1. Summarizes the valid ranges for each AIS data field

Data Field	Unit	Range	Noise Data

Longitude	[°]	-180 to 180	181
Latitude	[°]	-90 to 90	-91
Turn	$[\frac{o}{m}]$	-127 to 127	128
Speed	[knot]	0 to 100	-1
Course	[°]	0 to 359.9	3600
Heading	[°]	0 to 359	361
Accuracy	-	True or False	2

2.3 AIS Data Transformation

Data transformation was then performed to convert numeric field values linked to specific enumerations into semantically meaningful string representations. For example, the ship_type value 71 corresponds to "Cargo_HazardousCategory_A" in the ShipType enumeration and is classified under the broader "Cargo" category in the ShipCategory enumeration; thus, the *ship_type* value was transformed to the string "Cargo." Figure 2 illustrates the AIS data transformation workflow.

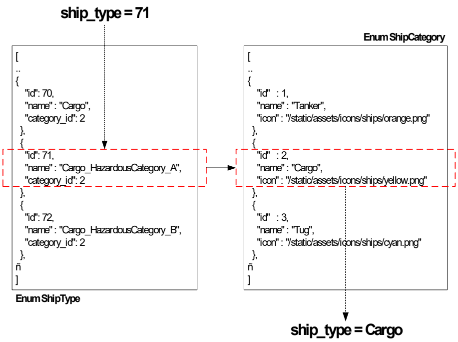

Figure 2. Ship type field transformation

The Navigation Status field, originally stored as the integer value 1, was transformed into the textual representation "AtAnchor" by referencing the NavigationStatus enumeration as shown in Figure 3.

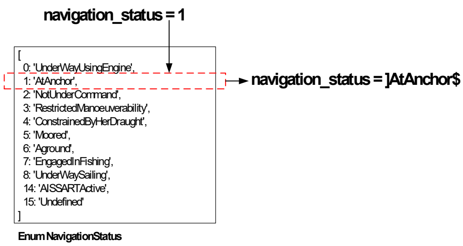

Figure 3. Navigation status field transformation

Following the transformation of the ship_type and navigation_status fields, the next step involved augmenting the AIS dataset with additional attributes flag, flagname, Gross Tonnage (GT), and DeadWeight Tonnage (DWT) by referencing an external database. Data retrieval was performed by querying the mmsi field in the AIS dataset against the external database, which contains records for 162,284 vessels. The inclusion of flagname in the AIS data is particularly important for extracting statistical information on the number of vessels registered under specific countries.

2.4 AIS Statistical Data Extraction

The extracted AIS statistics comprise: (i) the number of AIS records categorized by message type and AIS class, (ii) the number of vessels by country flag, (iii) the number of vessels by ship category, (iv) the number of vessels by navigational status, and (v) the monthly vessel counts and corresponding traffic densities over a five-month period.

The monthly vessel frequency transiting the Strait of Malacca was calculated based on unique vessel counts. Meanwhile, the vessel traffic density was calculated using the traffic density equation shown in Equation (4):

$$hip\ Traffic\ Density = \frac{Number\ of\ vessels\ in\ the\ area}{area\ (km^2)} \quad (1)$$

Variable Definitions:

- Number of vessels: calculated from the count of unique AIS records (*MMSI*) representing vessels present within or transiting the designated area.
- Area Size: computed based on the spatial extent defined by the area's coordinate boundaries (bounding box) or polygon, typically expressed in square kilometres (km^2).

2.5 AIS Data Visualization and Analysis

Following the computation of vessel frequencies and traffic density, the next stage focused on visualizing the extracted vessel trajectory data and AIS statistics in a more interactive and interpretable format. Visual outputs included bar charts, pie charts, and informative tables to facilitate clearer data interpretation. Prior to querying, the timestamp field in the dataframe was converted into the pd.to_datetime format. This standardized datetime format in Pandas enables greater flexibility in performing time-series queries when processing dataframes. The extracted data were subsequently analysed to derive insights and draw conclusions from the dataset.

3 Result and Discussion

3.1 AIS Statistical Data Result

A total of 46,251,723 raw AIS records were queried from the database. These records comprise eight AIS message types: message types 5 and 24 for static data, and message types 1, 2, 3, 18, 19, and 27 for dynamic data. Table 2 presents the distribution of records across the 8 AIS message types. Prior to analysis, the dataset was filtered to reduce computational load and improve database storage efficiency (Enda et al., 2021).

Table 2. Data distribution of 8 types of AIS messages

Message Type	Count	Percentage
1	38,694,718	83,66%
18	3,093,593	6,69%
3	2,102,301	4,55%
5	1,144,373	2,47%
24	1,030,016	2,23%
19	147,835	0,32%
2	34,300	0,07%
27	4,587	0,01%
Total	46,251,723	100%

Based on the data distribution, dynamic AIS messages account for approximately 95.3% of the dataset, with message type 1 being the most frequent, followed by types 18 and 3. From these data, several maritime information metrics were extracted, including the number of vessels by country, vessel counts by ship type, vessel counts by navigational status, monthly vessel frequencies, and vessel traffic densities within the Strait of Malacca monitoring area.

3.2 Discussion

Based on the results of the query on the AIS dataframe, covering data types as well as the counts of null and erroneous entries, the summary of findings is presented in Table 3.

Table 3. Data cleaning AIS message

#	Column	DType	Null or Error	Percentage
1	id	int64	0	0%
2	timestamp	object	0	0%
3	mmsi	int64	0	0%
4	lat	float64	1580	0,003%
5	lon	float64	15	0%
6	heading	float64	6299	0,013%
7	speed	float64	15	0%
8	turn	float64	3098120	6,698%
9	maneuver	float64	3099832	6,702%
10	course	float64	1699	0,004%
11	status	float64	3093521	6,688%
12	accuracy	object	15	0%

These results indicate that cleaning the AIS dataset is essential to reduce noise within the data. The detailed outcomes of the data cleaning process are shown in Table 4.

Table 4. Outcomes of the data cleaning AIS message

Item	Before	After	Percentage
Number of row	46,251,723	40,500,547	12,43%
Number of vessel (MMSI)	16,856	15,749	6,57%

The number of rows removed after cleaning represented approximately 12.43% of the dataset, while the reduction in invalid vessel records was 6.57%. These findings underscore the importance of initial data cleaning in minimizing noise within the AIS dataset.

3.3 Vessel Count by Country

The statistical data on vessel counts by country for traffic transiting the Strait of Malacca over a six-month monitoring period are summarized in Figure 4.

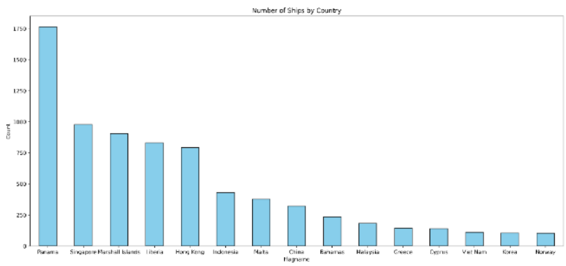

Figure 4. Number of ships by country

The top five flags with the highest number of vessels transiting the Strait of Malacca were Panama (1,764 vessels), Singapore (978 vessels), Marshall Islands (906 vessels), Liberia (830 vessels), and Hong Kong (791 vessels). In comparison, the number of domestically flagged vessels recorded in the strait was 431. These data indicate that maritime traffic in the Strait of Malacca is dominated by foreign-flagged vessels, particularly those registered under Panama.

3.4 Number of Ships by Ship Type

The distribution of vessels transiting the Strait of Malacca by ship type is illustrated in Figure 5.

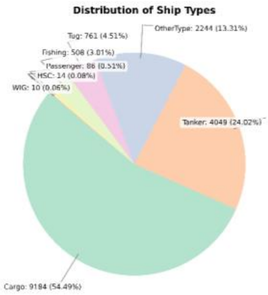

Figure 5. Distribution of ship types

Approximately 54.49% of vessels operating within the Strait of Malacca were cargo ships, followed by tankers at 24.02%. These vessels predominantly transport trade commodities and petroleum products involved in both export and import activities, either entering or leaving Indonesian waters.

3.5 Navigation Status Based on Ship Position Data

The navigational status of vessels transiting the Strait of Malacca, derived from dynamic AIS position data, is summarized in Table 5.

Table 5. Navigation status based on ship position data

Navigation Status	Count	Percentage
UnderWayUsingEngine	34,375,499	84,88%
UnderWaySailing	1,800,718	4,45%
Undefined	1,316,143	3,25%
ConstrainedByHerDraught	1,262,340	3,12%
AtAnchor	644,988	1,59%
NotUnderCommand	558,203	1,38%
Moored	412,896	1,02%
RestrictedManoeuverability	107,869	0,27%
Aground	12,890	0,03%
EngagedInFishing	8,980	0,02%
AISSARTActive	21	0,00%
Total	40,500,547	100%

Analysis of the navigational status distribution shows that the majority of vessels (84.88%) were classified as "UnderWayUsingEngine", indicating that they were operating under engine power while navigating the strait. This was followed by "UnderWaySailing" (4.45%), representing vessels propelled primarily by sails; "Undefined" (3.25%), indicating vessels with undefined navigational status; "ConstrainedByHerDraught" (3.12%), referring to vessels restricted by their draft; and "AtAnchor" (1.59%), representing vessels moored at a berth or specific location.

3.6 Monthly Vessel Frequency and Traffic Density

The estimated monitoring area was determined using Google Earth Pro, where three-dimensional polygon vertices were defined based on the farthest detectable AIS signal coverage, as illustrated in Figure 6.

Figure 6. AIS signal coverage and area 3D polygon

The measured area was approximately 13,121 km^2 with a perimeter of 586 km. This spatial extent was used as a reference for calculating the monthly vessel traffic density. The dataset covers the period from February to June 2025 and includes statistics on the monthly frequency of unique vessels and the corresponding traffic density, as presented in Table 6.

Table 6. Frequency of the number of ships each month and their density

Month	February	March	April	May	June	Total	Average
Count	17,122	18,027	18,687	19,887	16,120	89,843	17,968
Density (km^2)	1,30	1,37	1,42	1,51	1,23	6,83	1,36

The tabulated results indicate that a total of 89,843 unique vessels transited the Strait of Malacca during the observation period, with an average monthly detection of 17,968 vessels. The overall traffic density for the five-month period was calculated at 6.83 vessels/km^2, corresponding to an average monthly density of 1.36 vessels/km^2.

4 Conclusion

Based on AIS data analysis for February–June 2025, maritime traffic in the Strait of Malacca was predominantly composed of foreign-flagged vessels, particularly those registered in Panama, Singapore, and the Marshall Islands. The primary vessel categories were cargo ships (54.49%) and tankers (24.02%). The majority of vessels were classified as "UnderWayUsingEngine" (84.88%), reflecting the high prevalence of engine-powered navigation in the region. The average monthly traffic density of 1.36 vessels/km^2, and a total density of 6.83 vessels/km^2 over the observation period, highlight the Strait of Malacca's role as a strategic and high-intensity shipping lane. This underscores the need for optimal traffic management to ensure both navigational safety and operational efficiency.

References

Enda, D., Agustiawan, A., Milchan, M., & Pratiwi, E. (2021). Rancang Bangun Aplikasi AIS Backend Untuk Pemantauan Lalu Lintas Kapal di Selat Melaka. *INOVTEK Polbeng - Seri Informatika*, *6*(2), 284.

Feng, C., Fu, B., Luo, Y., & Li, H. (2022). The design and development of a ship trajectory data management and analysis system based on ais. *Sensors*, *22*(1), 1–22.

Ferreira, M. D., Campbell, J. N. A., & Matwin, S. (2022). A novel machine learning approach to analyzing geospatial vessel patterns using AIS data. *GIScience and Remote Sensing*, *59*(1), 1473–1490.

Handani, D. W., Ariana, I. M., Nugroho, T. F., & Indrayuni, F. (2018). AIS Based Spatial Distribution of Ship Emission in AIS Based Spatial Distribution of Ship Emission in. *Journal of the JIME*, *53*(3), 113–118.

Ikhsan, M. R., Pamungkasari, P. D., Purbantoro, B., & Diana, I. (2025). *Grid-Based Ship Density Analysis and Anomaly Detection for Ship Movements Monitoring at Tanjung Priok Port*. *6*(1), 107–118. https://doi.org/10.59395/ijadis.v6i1.1367

Le Tixerant, M., Le Guyader, D., Gourmelon, F., & Queffelec, B. (2018). How can Automatic Identification System (AIS) data be used for maritime spatial planning? *Ocean and Coastal Management*, *166*(734271), 18–30.

Ma'ruf, M., Ruslin, R., Hasanah, U., Efendi, E., Arifin, W. A., Handani, D. W., Suasti, G. B. D., Syamsudin, F., Sinaga, J. V., Eryawan, B. S., & Al-Kautsar, K. C. B. (2024). Information Extraction from Automatic Identification System (AIS) Data for Identifying Shipping Line based on Maritime Mobile Service Identity (MMSI). *IOP Conference Series: Earth and Environmental Science*, *1423*(1).

Ma, J., Zhou, Y., & Zhu, Z. (2023). Identification and analysis of ship waiting behavior outside the port based on AIS data. *Scientific Reports*, *13*(1), 1–13.

Ribeiro da Silva, J. N., Santos, T. A., & Teixeira, A. P. (2024). Methodology for Predicting Maritime Traffic Ship Emissions Using Automatic Identification System Data. *Journal of Marine Science and Engineering*, *12*(2). https://doi.org/10.3390/jmse12020320

Rong, H., Teixeira, A. P., & Guedes Soares, C. (2024). A framework for ship abnormal behaviour detection and classification using AIS data. *Reliability Engineering and System Safety*, *247*(December 2023), 110105.

Sudiantara, I. G., Widyantara, I. M. O., & Wiharta, D. M. (2022). Identifikasi Aktivitas Illegal Transshipment Berbasis Kepadatan Point Lintasan Pada Data Ais. *Jurnal RESISTOR (Rekayasa Sistem Komputer)*, *5*(1), 38–46.

Triharjanto, R. (2019). *Pengembangan teknologi satelit di indonesia : sistem, subsistem, dan misi operasi* (Issue October).

Xi, D., Feng, Y., Jiang, W., Yang, N., Hu, X., & Wang, C. (2023). Construction of a Real-Time Ship Trajectory Prediction Model Based on Ship Automatic Identification System Data. *ISPRS International Journal of Geo-Information*, *12*(12).

Yan, Y., Wu, E., & Lv, T. (2024). *Research on AIS Data Analysis and Its Integration into Discipline-specific Teaching*. *14*(3).

Zaman, M. B., Kobayashi, E., Wakabayashi, N., & Maimun, A. (2015). Risk of Navigation for Marine Traffic in the Malacca Strait Using AIS. *Procedia Earth and Planetary Science*, *14*(October), 33–40. https://doi.org/10.1016/j.proeps.2015.07.082