SHIP TRAJECTORY PREDICTION FOR ANOMALY DETECTION USING AIS DATA AND ARTIFICIAL INTELLIGENCE: A SYSTEMATIC LITERATURE REVIEW

ARIF BADRUDIN¹, SISWO HADI SUMANTRI², RUDY AGUS GEMILANG GULTOM³, INENGAH PUTRA APRIYANTO⁴, HENDY RISDIANTO WIJAYA⁵, INDRAJANI SUTEDJA⁶

¹,²,³,⁴Doctoral Program of Defence Science, Faculty of Defence Technology, University of Defence, Jakarta, Indonesia
³Health Informatics Research Cluster, Faculty of Public Health, University of Indonesia, Jakarta, Indonesia.
⁵Information Systems Department, School of Information Systems, Bina Nusantara University, Jakarta, Indonesia.

Email: ¹arif.badrudin11379@gmail.com, ²siswo_32@yahoo.com, ³rudygultom67@gmail.com,
⁴bintangsby@yahoo.co.id, ⁵hanry132@ui.ac.id, ⁶indrajani@binus.ac.id

ABSTRACT

According to reports from the Maritime Security Agency (Bakamla), Marine and Fisheries Ministry (KKP), and Ministry of Transportation, many illegal ship activities have occurred in 2022, such as dropping anchors at the wrong time and place, which indicates illegal fishing, not turning on AIS while in Indonesian territory, violating ALKI limits (Indonesian Archipelagic Ocean Currents); and ship activities that are of the greatest concern are the activities of transferring oil cargo (ship-to-ship) in Indonesian territory without going through ports. This is not only detrimental but can also endanger the security and safety of shipping in the Indonesian Sea area. Relevant parties need to find out about these violations, provide evidence of violations, and warn. This study aims to find the right model based on artificial intelligence so that it can detect ship violations earlier and produce evidence of violations in Indonesian waters. This study uses a systematic literature review (SLR) of written sources from Scopus, IEEE Explorer, and Google Scholar with the keywords AIS data, trajectory prediction, and anomaly detection to find solutions that have been done before. This systematic review explores the current state of research on ship trajectory prediction for anomaly detection using Automatic Identification System (AIS) data and artificial intelligence. AIS is a widely used system for vessel tracking that collects information such as ship location, speed, and direction. This method produces 559 related papers, of which only 28 are appropriate and form the basis of this research. The results show that AIS data and artificial intelligence can help find violations through ship trajectory prediction using DBSCAN. This research can be further developed by exploring AIS data and completing it with comparisons with speed, heading, and travel time which will produce activity predictions where they stop.

Keywords: Ship Trajectory Prediction, Automatic Identification System (AIS) Data, Systematic Literature Review (SLR), Anomaly Detection, Machine Learning, Artificial Intelligence (AI).

1. INTRODUCTION

Indonesia is known as the largest archipelagic country in the world. Geographically, Indonesia spans from 60°N to 110°S and 92°E to 142°E. Indonesia has a sea area of 5.8 million km² consisting of a Territorial Sea with an area of 0.8 million km², an Archipelago Sea of 2.3 million km² and an exclusive economic zone of 2.7 million km². Indonesia also has 17,480 islands and a coastline of 95,181 km² [1]. Indonesian waters have become the passage for ships from various countries in the world.

The vastness of the Indonesian Sea is a passageway for ships from various countries around the world. Apart from being the largest archipelagic country, Indonesia also has a sea area of 3,544,743.9 km² and a land area of 1,910,931.32 km², based on these conditions, Indonesia is one of the most maritime countries in the world [2]. The Exclusive Economic Zone (EEZ) covers an area of 2,981,211.00 km² and 12 miles of sea, or an area of 279,322.00 km² [3].

Indonesia’s geographical conditions can be both a potential and a challenge for Indonesia because the role of sea waters can be significant and affect the progress of a nation. It is stated that the earth, water, and the natural resources contained therein are controlled by the state and used for the prosperity of the people. Therefore, the government must try its best to take advantage of its natural potential, especially marine resources, even though many foreign companies manage our marine resources.
Indonesia also has international sea lanes that are important for international-scale transportation, namely the Three Indonesian Archipelagic Sea Lanes (ALKI), so Indonesia must be able to maximize its role and position as a country that has strength between two oceans, namely the Indian Ocean and the Pacific Ocean. However, the vast sea area is also a challenge for Indonesia so that it can be managed and secured for the benefit of the state. This situation certainly requires the government to maintain and manage the wealth and potential of the sea in Indonesia. In accordance with the government program summarized in the president's Nawacita (Nine development priorities for the next five years), Indonesia has become a world maritime axis country, so the government programs carried out are maritime-oriented.

Ship Trajectory Prediction research as a method for detecting anomalies is important. This research focuses on finding the use of machine learning algorithms that are often used for Ship Trajectory prediction for articles searched on search engine literature from 2018 to 2023 using the Systematic Literature Review (SLR) method. This SLR uses keywords in its search, searches are carried out on Scopus, IEEE Explorer, Google Scholar, and many other searches. This is done in order to produce a structured approach to generate significant and innovative value.

The keywords in this study use Ship Trajectory Prediction, Machine Learning, Artificial Intelligence, Algorithms and use AIS data.

2. THEORITICAL FOUNDATIONS

2.1. Automatic Identification System (AIS)

AIS stands for Automatic Identification System, which is an automatic tracking system installed on ships, using a tranceivers and used by Vessel Traffic Services (VTS), which is a marine traffic monitoring system established by harbor or port authorities. AIS is intended primarily to enable ships to view maritime traffic in their territory and be seen by that traffic. This requires a special AIS VHF transceiver that allows local traffic to be viewed on an AIS-enabled computer monitor while transmitting information about the ship itself to other AIS receivers.

AIS operates mainly on two dedicated frequencies (VHF channels), namely: 1) AIS type 1: Working on 161.975 MHz-Channel 87B (simplex, for ship to ship); 2) AIS type 2: 162.025 MHz-Channel 88B (duplex, for ship to shore). It uses Self Organizing Time Division Multiple Access (STDMA) technology to meet high broadcast rates. This frequency limits line of sight, which is about 40 miles. Besides that, AIS is divided into 2 classes based on its use, namely: AIS Class A (mandatory for all ships 300 GT and above) and AIS Class B (mainly used for ships such as cruise ships).

Figure 1 shows the journey of the AIS signal, starting from the ship sent to the AIS Network (AIS Base Station, AIS-VTS Control Center, VTS Radar Site), and then the data is distributed via satellite to those who need it. AIS data can be obtained by subscribing and being sent in the form of raw data that needs to be processed by the user himself. The most fundamental differences between AIS Class A and B are transmit power, data transmission methods, communication schemes, and the ships that use them.

Ships equipped with AIS devices can transmit and receive various data information about ships released automatically, either in the form of displays on the Radar screen, or electronic maps (Electronic Navigation Charts – ENC via Electronic Chart Display and Information System – ECDIS). In addition to sending and receiving data information, ships equipped with AIS can also be verified and track the movements of other ships equipped with AIS (at VHF range). Data information on these ships can also be received by ground base stations, for example VTSs (Ship Traffic Services) stations. Ship data information sent includes [6] [7] [8] [15]:

1. IMO Number;
2. Call sign;
3. MMSI number;
4. ship position (latitude and longitude);
5. ship type, heading, speed;
6. Statis Draught, ship dimension;
7. origin, destination, navigation status;
8. others information for safety.

Figure 1. Automatic Identification System [5]
2.2. AIS and VMS Comparison

The initial concept of the proposal that ships should be equipped with AIS factors is for maritime security. After being introduced to IMO membership by participating in the 69th MSC session and 22nd Assembly session, by equipping ships with AIS equipment, life safety at sea can be supported by means of safety, security, and ease of navigation, as well as increasing protection of the maritime environment from pollution. In addition, AIS is also very useful for SAR operations in cases of ship accidents at sea.

Previously, other methods besides AIS had been used, namely the Vessel Monitoring System (VMS) which was developed in the 1990s for monitoring, controlling and supervising ships. At that time, this system was designed to address the main problems facing fishing vessels and authorities. Starting in the 2000s, AIS was defined as the standard for ships subject to SOLAS (International Convention for the Safety of Life at Sea) by the International Maritime Organization. AIS was developed to show where each ship should avoid a collision; it was never intended to support fisheries monitoring. However, AIS data is openly available, and this makes sense since the goal is to avoid collisions. AIS data is sent via VHF, which can be accessed by anyone with an AIS receiver, potentially exposing the vessel's main fishing grounds to competitors, therefore AIS is not used for fisheries.

2.3. Anomaly Detection

AIS has the potential to make a substantial and cost-efficient impact in these critical applications. Anomalies, in simple terms, are events that fall outside the anticipated or normal outcome. Real-time tracking using AIS data can highlight many types of anomalies, such as intrinsic anomalies, contextual anomalies, and behavioral anomalies. Each anomaly detection use case necessitates a distinct approach. Generally, the most interesting anomalies are behavioral. Is the ship following the expected route? Is it being detected correctly? Is it headed for restricted waters?

What do we do after detecting anomalies? Some common choices are a) count and pass on to the next stage of the process; b) count, tag, and pass; c) alert and report. If we take action and report it, then action needs to be taken at the next stage.

2.4. Machine Learning

In certain situations, related to predicting data, machine learning can recognize a large set of given data, as in the case of knowledge about ship trajectories, using classification [9] [10] [11] and regression [12] illustrations. Pattern recognition or unsupervised machine learning methods can be said to be so because there are no guidelines for the data, including in certain groups. Some studies use this method as the best way to detect anomalies in a sequence of data.

After conducting several literature studies, it has been found that conducting research by detecting anomalies in AIS ship data for ship travel history is a very efficient way to support ship monitoring activities. Several methods show that data characteristics can affect calculation quality. These methods have similarities in the process to be carried out. Unsupervised The learning method is used for data that does not have examples or labels indicating that it contains certain values. Grouping is carried out based on calculations through appropriate algorithms to determine positive or negative values in the data. This research applies machine learning to detect anomalies in ship behavior. The data used for this study is unlabeled, and there are no examples for data comparison. Anomalies in ship behavior during ship monitoring activities have various categories in each organization, such as ship monitoring for oil bunkering, human smuggling, piracy, and other illegal activities.

2.5. K-Means and DBSCAN Algorithm

K-means clustering is a non-hierarchical cluster analysis method that finds ways to partition existing objects into one or more clusters or groups of objects based on their characteristics. Cluster refers to a collection of data points that are gathered together due to certain similarities. The K-Means Clustering method attempts to group existing data into several groups, where data in one group have the same characteristics as each other but have different characteristics from data in other groups.

In contrast to DBSCAN (The Density-Based Spatial Clustering of Applications with Noise) algorithm is a density-based clustering method of observed data positions with the principle of relatively close data grouping. DBSCAN is often applied to data that contains a lot of noise, this is because DBSCAN will not enter data that is considered noise into any cluster. DBSCAN is one of the algorithms in machine learning. DBSCAN has a clustering method that is almost similar to DENCLUE. Significantly, DBSCAN works
efficiently in forming arbitrary-shaped clusters. Grouping is done on points with their neighbors that are within a certain distance, which must meet the minimum number of points (minPts). In the process of creating a cluster using DBSCAN, data will be grouped with its neighbors. A pair of observations is considered to be neighboring if the distance between them is smaller than the epsilon value.

Figure 2. DBSCAN works to process and create clusters

In simple terms, the way DBSCAN works is as follows:

1. Determine the minPts and epsilon (eps) values to be used;
2. Select the initial data "p" randomly;
3. Calculate the distance between data "p" and all data using the Euclidian distance;
4. Take all density-reachable observations with "p" observations;
5. If the observations that meet the epsilon value are more than the minimum number of observations in one cluster, then the "p" observations are categorized as core points and a cluster is formed;
6. If the observation "p" is the border points and there are no density-reachable observations with the observation "p", then continue to other observations;
7. Repeat steps 3 to 6 until all observations are processed.

3. METHODOLOGY

The method used in this study is a Systematic Literature Review using Perish software (free) to search for articles, Microsoft Excel (licenced) to perform inclusion and extraction, as well as perform a comparative analysis of the articles collected on the relevance of keywords and the year of the article using VOSviewer. The adoption of a Systematic Literature Review consists of six stages [13] [14]:

1. Defining research question used;
2. Determine the required characteristics of the primary study related to inclusion, craft criteria for exclusion, and/or provide examples of research methods, focus of study, outlets, and use of language;
3. Relevant article samples related to determining the search procedure. For example, searching databases and references and specifying keywords to create an initial sample;
4. Selecting relevant articles related to the application of inclusion and/or exclusion criteria;
5. Synthesize articles related to applying coding schemes, extract relevant information from literature, and synthesize studies to summarize, integrate, or collect findings that differ across study majors;
6. Report results related to the review report and provide an overview of the literature review and thematic discussion findings.

3.1. Perish Search

Relevant articles were collected using Perish software, and article disbursement was carried out on Google Scholar, Scopus, and Crossref using the selected keywords, after which it was continued by selecting the journal output obtained. Through existing sources (Google Scholar, Scopus, and Crossref), articles are searched using several keywords to get relevant articles. Key words include "ship trajectory", "AIS Data", "Anomaly Detection", "machine learning" and "prediction". In particular, some keyword combinations are also created as follows:

1. "Ship Trajectory" AND "Algorithm" AND ("Machine Learning" OR "Artificial Intelligence" OR "Prediction")
2. “AIS Data” AND ("Machine Learning" OR "Artificial Intelligence" OR “Prediction")
3. “Ship Trajectory” AND “AIS Data”
4. “Anomaly Detection”

With the addition of keywords, it is hoped that you can find more specific articles. The 2 common words used are AIS Data and Ship Trajectory while the observation keywords are Machine learning, Artificial intelligence, and Prediction.

3.2. Article Inclusion and Extraction

After the process of searching for articles according to the keywords (inclusion), then selecting articles that are in accordance with the objectives of this research. A total of 559 articles were obtained,
then extraction was carried out, namely grouping suitable substances to be studied. Comparisons were made in terms of goals and algorithms that were often used, resulting in a total of 114 articles that could be suitable candidates for study. Then a final selection was carried out, namely selecting articles related to Ship Trajectory Prediction, anomaly detection and machine learning so that a total of 28 articles were obtained. All of these processes can be shown at Table 1.

Table 1. Paper Number Have Already Collected and Selected

<table>
<thead>
<tr>
<th>No</th>
<th>Source</th>
<th>Studies Found</th>
<th>Candidate Studies</th>
<th>Selected Studies</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Google Scholar (<a href="https://scholar.google.com/">https://scholar.google.com/</a>)</td>
<td>340</td>
<td>71</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>Scopus (<a href="https://www.scopus.com/home.uri">https://www.scopus.com/home.uri</a>)</td>
<td>10</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>ACM (<a href="https://dl.acm.org/">https://dl.acm.org/</a>)</td>
<td>81</td>
<td>41</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>IEEE Explore (<a href="https://ieeexplore.ieee.org/Xplore/home.jsp">https://ieeexplore.ieee.org/Xplore/home.jsp</a>)</td>
<td>128</td>
<td>50</td>
<td>27</td>
</tr>
<tr>
<td></td>
<td><strong>Total</strong></td>
<td><strong>559</strong></td>
<td><strong>114</strong></td>
<td><strong>28</strong></td>
</tr>
</tbody>
</table>

With a comparison of the articles obtained and extracted against the algorithm used and the resulting level of accuracy, it can be concluded that the development of machine learning models for ship trajectory prediction research has an increasing trend.

4. RESULT AND DISCUSSION

This research filters article sources from the period of 2018 to 2023 for "ship trajectory detection" using 3 search sources, namely Google Scholar, Scopus, and Crossref that 559 articles have been found. The final grouping uses additional keywords, namely machine learning, artificial intelligence, and prediction, and 28 articles are found that can be analyzed.

4.1. Analysis based on Algorithm

Table 2 shows the most dominant distribution of the use of the DBSCAN model for ship trajectory prediction [16][17][18][5][7], while the second most used is the use of K-mean. K-Means [19][7].

Table 2. Articles Count by Machine Learning Algorithm

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Total Articles</th>
</tr>
</thead>
<tbody>
<tr>
<td>DBSCAN</td>
<td>12</td>
</tr>
<tr>
<td>K-Means</td>
<td>8</td>
</tr>
<tr>
<td>LSSVM</td>
<td>4</td>
</tr>
<tr>
<td>CNN</td>
<td>3</td>
</tr>
<tr>
<td>GAN</td>
<td>2</td>
</tr>
</tbody>
</table>

Figure 3 and Figure 4 show that there is a close relationship between keyword DBSCAN, algorithms, detection, trajectory data, and
applications with Trajectory. The other relation is Algorithm like such as GRU [20], LSTM [21][22][23], LSSVM [24], CNN [25], dan GAN [26]. The next relationship is clustering, data, and AIS. The last one is connection AIS data [27][28][29] with cluster [30] [31][32], K-means, dan density.

4.2. Analysis based on Algorithm vs Year

The use of DBSCAN and K-Means algorithms for ship trajectory prediction from AIS data has been increasing since 2020. The algorithms used at that time were clustering and K-means. Over time, the DBSCAN algorithm has been increasingly utilized, often in combination with other applications. A density-based clustering approach can be used to find groups of arbitrary shapes by searching for high-density connected areas for clustering. The Density-Based Spatial Clustering with Noise Algorithm (DBSCAN) proposed in this study utilizes a different approach to clustering compared to traditional methods that rely on distance-based data similarity measures.

Instead, DBSCAN identifies areas with sufficiently high density and groups them into clusters, disregarding areas with low density. This method can be used to identify high-density regions as cluster clusters in spatial databases with "noise" data, but the DBSCAN algorithm requires the user to adjust the initial parameters, and the results are very sensitive to the values of both parameters. At the same time, it is difficult to find different density groups. In addition, clustering by fast search and finding density peaks, common nearest neighbors, and density-ratio...
estimation all enhance the DBSCAN algorithm for finding clusters with different densities.

Based on Figure 5 above, it explains that trajectory is very closely related to DBSCAN, clustering, and algorithms. In addition, there is a relationship with AIS data, trajectory data, and detection.

5. CONCLUSION

The results that have been obtained from the SLR method above show that there is still not much research on the topic of Ship Trajectory using AIS data, using machine learning algorithms. Only 28 articles can be used to find out that the most frequently used algorithms in order are DBSCAN, K-Mean, LSSVM, CNN, and GAN.

DBSCAN has advantages over other algorithms, which uses the partitioning method, DBSCAN does not require the number of clusters as input; instead, it uses epsilon and minPts as input parameters. Epsilon is the maximum allowed distance between two observations in one cluster, and minPts is the minimum amount of data within the epsilon distance to form a cluster. The advantage of this method is that it can capture clusters that have shapes and can detect noise in the data. The drawback of this method is that it is not suitable for data that has various density levels; DBSCAN can be sensitive to the choice of parameters, such as epsilon and minPts. It is also known to have difficulty with data that has large dimensions or varying densities. Therefore, it is important to carefully choose the parameters and preprocess the data appropriately before applying DBSCAN.

REFERENCES


