IAES International Journal of Artificial Intelligence (IJ-AI)
Vol. 12, No. 1, March 2023, pp. 1~11
ISSN: 2252-8938, DOI: 10.11591/ijai.v12.i1.pp1-11

Automatic identification system-based trajectory clustering framework to identify vessel movement pattern

I Made Oka Widyantara[1], I Putu Noven Hartawan[2], Anak Agung Istri Ngurah Eka Karyawati[3], Ngurah Indra Er[1], Ketut Buda Artana[4]
[1]Department of Electrical Engineering, Faculty of Engineering, Udayana University, Bali, Indonesia
[2]Postgraduate Program of Electrical Engineering, Faculty of Engineering, Udayana University, Bali, Indonesia
[3]Department of Computer Science, Faculty of Math and Natural Science, Udayana University, Bali, Indonesia
[4]Department of Marine Engineering, Faculty of Marine Technology, Sepuluh Nopember Institute of Technology, Surabaya, Indonesia

Article Info

Article history:

Received Nov 20, 2021
Revised Jul 20, 2022
Accepted Aug 18, 2022

Keywords:

Automatic identification system
Data mining
Density-based spatial clustering of applications with noise
Longest common subsequence
Trajectory
Vessel

ABSTRACT

Automatic identification system (AIS) is a vessel radio navigation equipment that has been determined by international maritime organization (IMO). Historical AIS data can be utilized for anomaly detection, trajectory prediction, and vessel trajectory planning. These benefits can be achieved by identifying the vessel's trajectory pattern through trajectory clustering. However, more effort is needed in trajectory clustering using AIS data due to their large volume and the significant number of deficiencies. In addition, trajectory clustering cannot be directly applied to trajectory data, which also applies to vessel trajectory. Therefore, we propose a trajectory clustering framework by combining douglas peucker (DP), longest common subsequence (LCSS), multi-dimensional scaling (MDS), and density-based spatial clustering of applications with noise (DBSCAN). Our experiments, carried out with AIS data for the Lombok Strait, Indonesia, showed that the trajectory compression with DP significantly accelerates the similarity measurement process. Moreover, we found that the LCSS is the optimal algorithm for similarity measurement of vessel trajectories based on AIS data. We also applied the right combination of MDS and DBSCAN in density-based clustering. The proposed framework can distinguish trajectoriess in different directions, identify the noise, and produce good quality clusters in relatively fast total processing time.

Corresponding Author:

I Made Oka Widyantara
Department of Electrical Engineering, Faculty of Engineering, Udayana University
Kampus Unud Road, Jimbaran, Bali, Indonesia
Email: oka.widyantara@unud.ac.id

1. INTRODUCTION

The automatic identification system (AIS) is a radio navigation device that uses very high frequency (VHF) to transmit vessel data automatically between vessels at sea and receivers on land. Every vessel over 300 gross tons (GT) must have an AIS signal transmitter, according to the international maritime organization (IMO) regulation [1]–[4]. Vessel location, speed, lane, direction, turn rate, destination, and expected time of arrival are among the dynamic data supplied by AIS. Static data as are vessel name, vessel maritime mobile service identity (MMSI ID), message identity (ID), vessel type, vessel size, and current time also provided. Furthermore, AIS data has the advantage of providing the highest volume of vessel position data with wide water area coverage [5] and commercially accessible or open-source ais data, which other vessel reporting systems do not have [6]. Many things may be evaluated using AIS data due to the vast number of data, including traffic analysis, transportation logistics, monitoring, collisions, pollutants, oil spills, and fishing activity [7].

Based on the history of existing AIS data, data mining techniques and artificial intelligence systems can be utilized to identify vessel route pattern at sea. Anomaly detection, trajectory prediction, and vessel trajectory planning can all be done after vessel trajectory pattern has gotten [8]. Moreover, clustering can group vessel trajectories based on the characteristics of each trajectory. The vessel trajectory pattern, whether it is already or not yet known, will appear based on the clusters resulting from the AIS data clustering process [9]–[11]. However, more effort is needed in trajectory clustering using AIS data due to its large volume and usually many deficiencies, such as low data quality, irregular AIS message time intervals, and poor data integrity [12]. The anomalies occur because AIS messages are sent from vessels with various types, delivery distances, and geographical conditions. In addition, trajectory clustering cannot be directly applied to the trajectory data, including the vessel trajectory data from AIS. Inherently, vessel trajectories are different from traditional data commonly used in clustering methods [10]. Therefore, a suitable trajectory clustering framework is needed to generate vessel trajectory patterns using AIS data. The main steps that need to be carried out in AIS data trajectory clustering are data pre-processing, similarity measurement, and the clustering process itself.

Data pre-processing is a crucial phase in data mining, and it also applies to vessel trajectory clustering [13]. The most time-consuming phase in data mining is data-preparation, which will take longer than the main data mining process itself. Incomplete data, noise, data without attributes, and repeating data are all possible to found in real-world scenarios. The length and shape of the vessel's trajectory varies greatly in the AIS data. Moreover, AIS data often has an abnormal trajectory pattern, which will mislead the algorithm used [14].

Trajectory similarity measurement is a determining factor in trajectory clustering [15]. The method used must be able to make the distance between different trajectories as far as possible and the same trajectory as close as possible. Based on previous research, several similarity measurement methods are commonly used in trajectory clustering with AIS data. Research in [16]–[18] applied hausdorff distance (HD) for trajectory clustering, where HD can identify the shape of the trajectory, calculates the maximum shortest distance value from one trajectory to another, and calculates the average value of the two maximums as distance. However, HD is inadequate in identifying the direction of the trajectory due to its sensitivity to noise [15]. HD also has shortcomings in measuring distance in dense water areas, thus giving incorrect cluster results [16]. Li *et al.* [9] applied dynamic time warping (DTW) in trajectory clustering on the bridge area waterway and Mississippi river. Li *et al.* [10] used merge distance (MD) in trajectory clustering on the bridge waterways. Furthermore, Li *et al.* [9] and Li *et al.* [10] conducted clustering with less varied trajectory data. Li *et al.* [10] showed that DTW and MD have the same accuracy, but DTW is superior in terms of processing time, because MD is a more complex algorithm. The shortcoming of DTW is that the resulting distance greatly affects the noise and sampling rate of the track. It is a potential challenge because the AIS data contains redundant vessel positions caused by the vessel sending AIS messages within the span of 3-10 seconds [19].

Based on the research in [20], partition-based methods, hierarchy-based methods, density-based methods, grid-based methods, and model-based methods are the five categories of clustering methods. The following are some previous studies in the context of trajectory clustering. Partition based clustering is a type of clustering method in which the number or center of clusters is identified before processing is applied. K-means and K-medoids are representations of partition-based methods. Li *et al.* [9] utilized K-means as a vessel trajectory clustering method using AIS data. Furthermore, principal component analysis (PCA) is used to find the value of k in the same research. Zhen *et al.* [17] used K-medoids as a clustering method which will later be classified to detect anomalies. However, both methods cannot automatically detect noise. Density based is a clustering method based on point density. Density-based spatial clustering of applications with noise (DBSCAN) is the most frequently used density-based method. Research in [10], [16], [21], used DBSCAN in the vessel trajectory clustering process, where it can automatically search for the number of clusters based on density. DBSCAN can group clusters with irregular shapes and discover noise automatically and effectively [10].

In this study, DBSCAN was chosen as the algorithm for trajectory clustering. DBSCAN is an unsupervised clustering algorithm that does not need the specification of the number of clusters at the beginning [22]. DBSCAN with epsilon (*Eps)* concept is highly dependent on spatial density. However, DBSCAN has a "curse of dimensionality" problem [23]–[25] and to overcome this, our study applies dimensional reduction with the multi-dimensional scaling (MDS) algorithm. MDS is used to reduce the dimensions of the similarity matrix into relative position data which is a low-dimensional representation of the similarity matrix. The MDS data will be utilized and injected into the DBSCAN while the distance from the MDS data will be used to find the optimal epsilon parameter. Furthermore, the similarity measurement stage uses the longest common subsequence (LCSS) algorithm. LCSS was chosen because it has less effect on noise and different trajectory lengths while can also detect the direction of the trajectory [26]. Douglas peucker (DP) algorithm is proposed at the pre-processing stage to speed up the similarity measurement process. By combining those algorithms into a framework proposed in this study, we can perform trajectory clustering with a good quality and speed from a collection of vessel trajectories based on complex and diverse AIS data, so that the cluster results can be used as a basis for anomaly detection, trajectory prediction, and vessel trajectory planning.

This study uses AIS data in the waters of the Lombok Strait which has the third highest shipping traffic density in Indonesia [27]. The firts stage of the proposed framework are the cleaning of data and the translation of the AIS coordinate data rows into trajectory data. The next stage is to remove unnecessary coordinate points from the vessel's trajectory using DP while also measuring the similarity between existing vessel trajectories using LCSS. The next stage is to change the trajectory similarity distance data from the similarity matrix into spatial points using MDS, and finally to conduct the clustering using DBSCAN. To evaluate the quality of the resulting cluster, a comparison is made between the proposed algorithm and some benchmark algorithms, based on the total time and cluster quality measurements using the silhouette coefficient (SC) method.

2. METHOD

Figure 1 shows an overview of the proposed framework. It starts with raw AIS data containing row coordinates and vessel information based on time. It follows by several processes. The first stage is preprocessing, which includes data cleaning and converting it into vessel trajectory data and then proceed with trajectory compression. Furthermore, each trajectory which is a combination of several coordinates is simplified using DP. After simplifying the trajectory, each trajectory will be measured using LCSS to find the similarity distance between the trajectories. Then, the results of the distance matrix from DTW need to go through a dimension reduction process using MDS to convert three-dimensional (3D) data into two-dimensional (2D) spatial before proceeding to the clustering stage using DBSCAN.

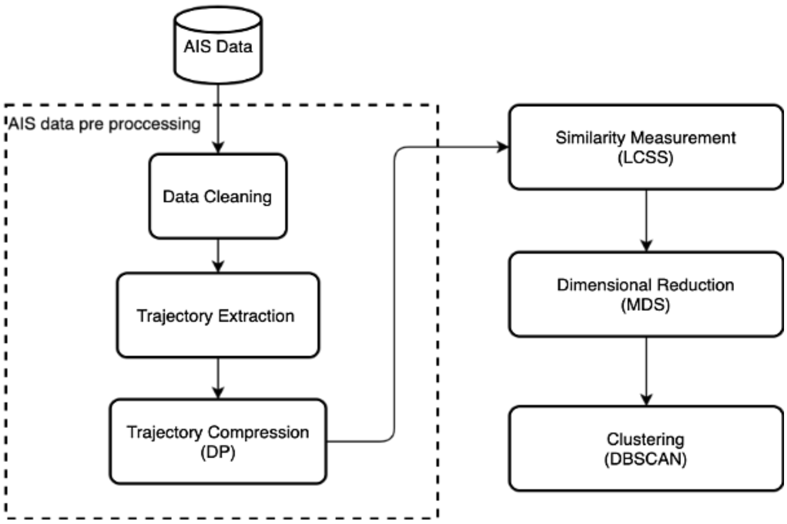


Figure 1. Method overview

2.1. Data preprocessing

Preprocessing is the first step to overcome the problem of AIS data deficiency and make the data ready to be used in trajectory clustering. Base on [13], there are three stages in trajectory clustering preprocessing, there ara cleaning, extraction and compression. At the data cleaning stage, course over ground (COG) and speed over ground (SOG) selections are made. Abnormal data that indicate the vessel is not moving are also eliminated. The SOG selection will also affect the results of the trajectory extraction at a next stage, because it creates a bigger time gap.

Because a vessel can have many trajectories, trajectory extraction cannot be done simply by grouping the vessel's position with MMSI. Therefore, at trajectory extraction phase we trim the trajectory of each MMSI. Referring to research in [18], trajectory trimming is done by measuring the period between trajectories using time threshold. MMSI markers such as 26XXX591 have been replaced by new markers such as 26XXX591-1, 26XXX591-2, and so on, indicating that the data is a single trajectory unit.

The number of points owned by each vessel's trajectory will make the similarity measure process take a long time. To overcome this, trajectory compression can be done by eliminate the redundant coordinate points without losing the trjectorie's original shape. The algorithm used in the trajectory compression process is the DP algorithm. Due its accuracy and speed, the DP algorithm is widely used in compression of trajectories or moving point objects [28].

2.2. Similarity measurement with LCSS

The measurement of the similarity distance between all trajectories is carried out after performing trajectory compression. LCSS is a well-known method for measuring text similarity, while in the context of trajectory similarity measurement, LCSS can solve the noise problem in trajectory [26]. The main idea of LCSS is to calculate Euclidean distances from several points within two trajectoriess in turn. To solve that, LCSS requires threshold parameters ϵ. When measuring the distance of trajectory, A and B. LCSS consider $a_i(a_i \in A)$ and $b_j(b_j \in B)$ is similar if the distance between the trajectories is less than ϵ and LCSS will ignore some points from A and B if the distance of the points exceeds ϵ.

2.3. Dimensional scaling with MDS

After acquiring the distance matrix using LCSS, dimension reduction is carried out to represent 3D data into 2D spatial data. MDS is a dimension reduction approach that preserves an object's core information while converting multidimensional data into a lower-dimensional space. The primary reason for utilizing MDS is to obtain a graphical representation of the data, making it easier to comprehend. There are some other dimensionality reduction techniques such as PCA, factor analysis, and isomaps. However, MDS is the most popular among these techniques due to its simplicity and various application areas [29]. MDS analysis to find spatial maps for objects is based on similarity or difference information between those objects.

2.4. Clustering with DBSCAN

Following the conversion of the distance matrix into spatial data, clustering is conducted with DBSCAN. The measurement of distance with DBSCAN spatial data can be done by calculating the Euclidean distance. Moreover, the DBSCAN algorithm is used to identify clusters and noise with the specified parameters *Eps* and minimum points (*MinPts)*. After completing the clustering process, the cluster labels will be visualized back to each trajectory. DBSCAN is also a density-based clustering algorithm, which scans for a high-density data set to serve as a cluster. DBSCAN does not estimate the density between points for efficiency reasons. Within a radius of the core point, all neighbors are regarded to be part of the same cluster as the core point [30]. The cluster shape generated by DBSCAN is density-dependent, and it is possible to generate arbitrary cluster shapes [31]. A cluster in DBSCAN is defined as the maximum data set connected within that density (density-connected). Membership of each profile is calculated based on the distance formula. Moreover, DBSCAN is considered an unsupervised clustering algorithm because the number of clusters generated is determined by the shape of the data distribution itself, not initialized at the beginning.

3. RESULTS AND DISCUSSION

This study uses datasets from terrestrial AIS receivers at Udayana University. The dataset used has 640,527 rows. Based on MMSI we found 437 vessels. The other attributes of the AIS data used in this study are timestamp, MMSI, latitude, longitude, SOG and COG. The experiment was carried out using M1 Macbook Air. Table 1 is details of the research instrument specifications.

Table 1. Research instrument

Item	Configuration
Number of rows	640,527
Number of vessel (MMSI)	437
AIS dataset	Udayana University terrestrial AIS receiver. Scope from latitude. -8.2 to longitude 116
Dataset stored at	MySQL 8 and .npy file format
Programing language	Python 3.8 with scikit-learn
Hardware spec.	8 Core Apple M1 CPU; 8GB LPDDR4X-4266 MHz SDRAM; 512GB NVMe SSD

3.1. Data preprocessing

There are three steps in the preprocessing stage. Figure 2 visually shows the change of data at each preprocessing step. The first step is the data cleaning, where abnormal data such as empty vessel position attributes, COG values outside 0-360, and out-of-range vessels positions are eliminated. In this study, we aim to identify the trajectory. Therefore, the data with a SOG value below 1.5 will also be eliminated because it shows a vessel is not moving.

The second step is to perform trajectory extraction. After the extraction, it is still necessary to perform data elimination for trajectories that only have a few rows. The data cleaning and trajectory extraction process succeeded in reducing the initial data in Figure 2 (a), which has 640,527 rows and 437 vessels, to Figure 2 (b), which has 127,144 rows and 231 vessels with 405 vessel trajectories.

The last step is to implement the DP algorithm to to compress each trajectory. The epsilon configuration used is 0.001, which is 111m. Figure 2 (c) shows that the number of rows of data can be reduced to 4,225. Visually, the shape of the trajectories maintains the same characteristics as the trajectories before compression. Table 2 shows the breakdown of data changes from the data preprocessing stages.

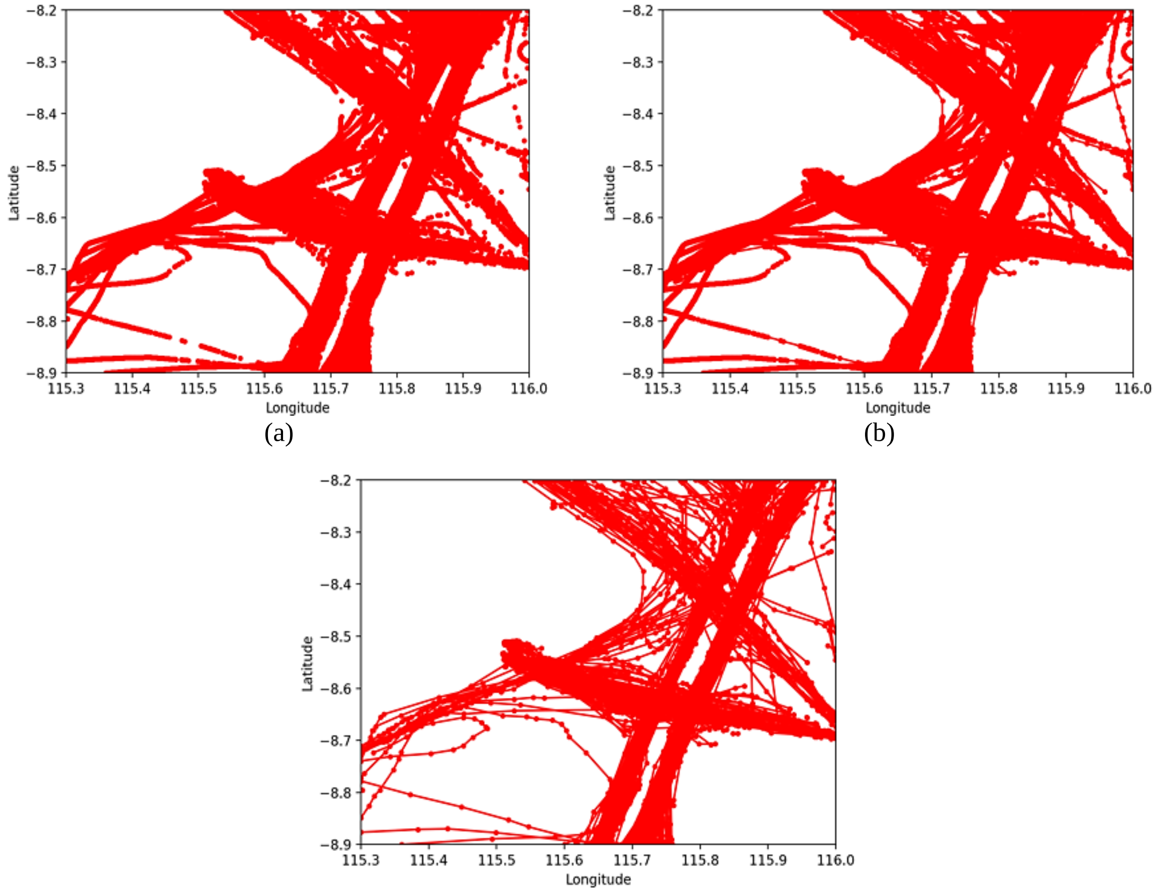


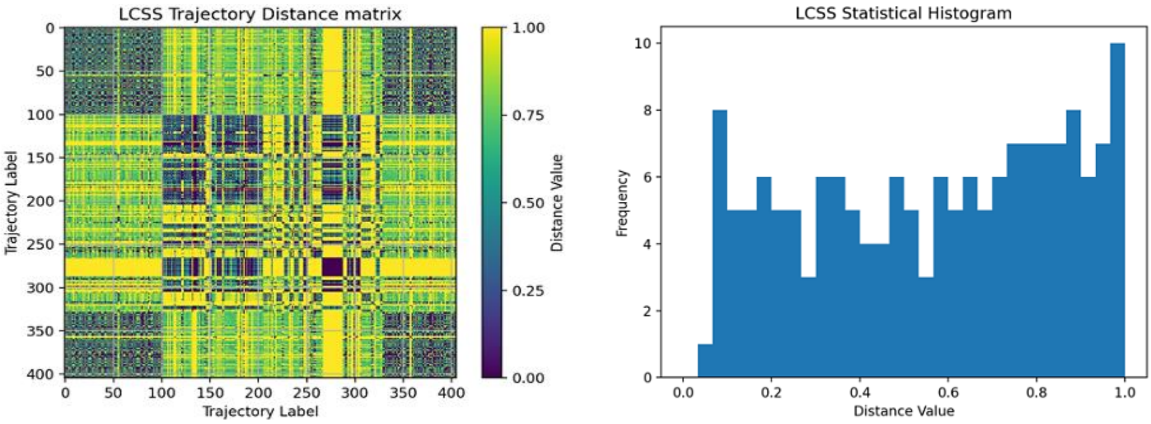
Figure 2. Data preprocessing step from (a) raw AIS data, (b) clean and extracted trajectories and (c) compressed trajectories

Table 2. Data preprocessing result

	RAW	Clean	Compressed
Number of rows	640,527	127,144	4,225
Number of vessel (MMSI)	437	231	231
Number of trajectories	-	405	405

3.2. Similarity measurement with LCSS

The LCSS algorithm is applied to measure the similarity distance between all trajectories in the similarity measurement stage with threshold parameter 0.1. The measured trajectories are trajectories that have been compressed with the DP algorithm. The process to get the distance matrix took 19.414s. Figure 3 (a) is a 2D view of the distance matrix where the x-axis and y-axis are the vessel's trajectory. Figure 3 (a) shows the characteristics of the distance between trajectories. If the distance is close to 0, it is marked with a dark color indicating the similarity of the trajectory characteristics. On the other hand, if the value is greater than 0, it is marked with a light color to show differences between trajectories. In Figure 3 (b), the x-axis is the distance value, and the y-axis is the frequency of the number of passes. Figure 3 (b) also shows the number of similarities between the trajectories for each existing distance value.



(a) (b)

Figure 3. Similarity measurement between all trajectories (a) 2D image from distance matrix and (b) statistical histogram of all distance

3.3. Dimensional reduction with MDS

In the dimensional reduction process, the MDS algorithm converts the distance matrix from 3D data into 2D spatial data. The 2D distance matrix in Figure 3 (a) is 3D data where the x-axis and y-axis are trajectories labels. The z-axis is the value of the distance between trajectories. Figure 4 is the result of the MDS, which represents the data into 2D spatial data.

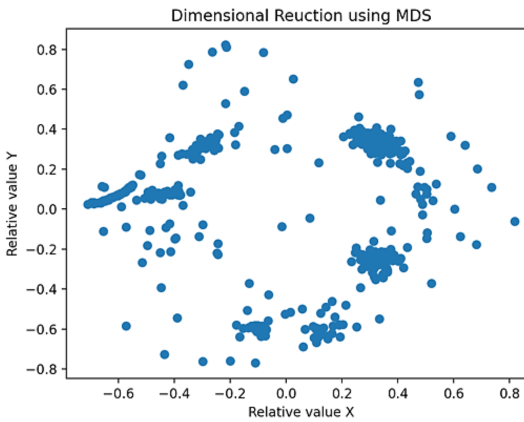
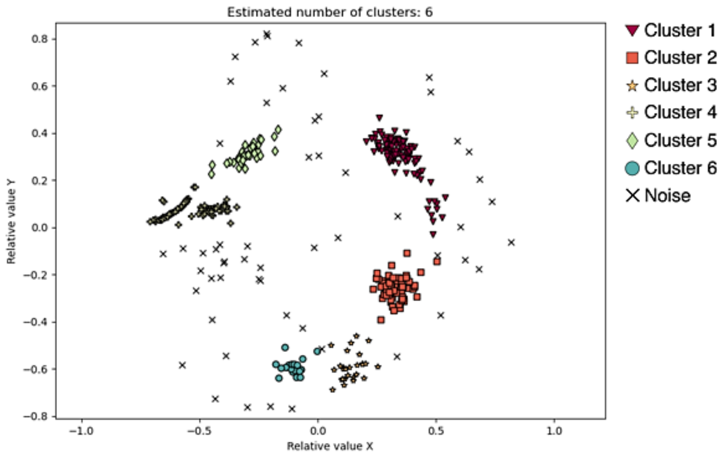


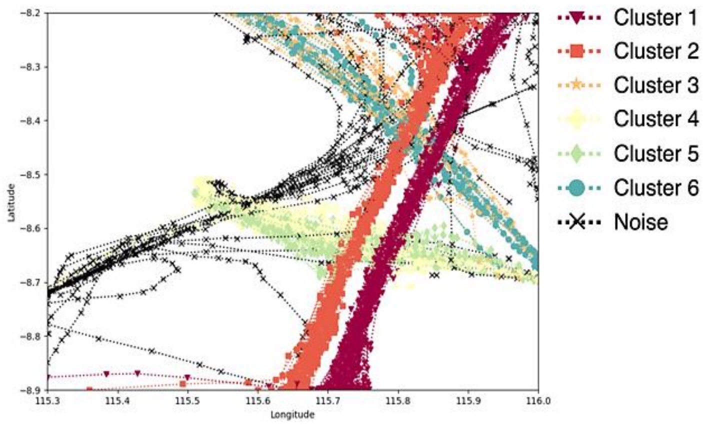
Figure 4. Dimensional reduction MDS spatial representation

3.4. Clustering with DBSCAN

The clustering stage uses the DBSCAN algorithm. The configuration used is *Eps*=0.088 and *MinPts*=9. The data exploited in the clustering process is the spatial data from the MDS in Figure 4, while the obtained result from clustering is shown in Figure 5 (a). The clustering results are then mapped to each trajectory, as shown in Figure 5 (b). Every color shows the trajectories cluster, except the colored black representing noise, where the black trajectories do not fit into any cluster.



(a)



(b)

Figure 5. Clustering result of (a) MDS representation and (b) trajectories labeled by cluster

3.5. Visualization of clustering result

The clustering results using the proposed clustering framework achieved an SC score of 0.533. Thus, the number of successfully generated clusters are 6, while 57 trajectories were found to be noise. The number of trajectories in each cluster can be seen in Table 3.

Each vessels trajectories clusters can be seen in figure 6. The clusters in Figures 6 (a) and 6 (b) show the trajectories that pass through the traffic separation scheme (TSS) in the Lombok Strait. Figure 6 (a) shows vessel traffic moving from south to north, and Figure 6 (b) shows the opposite direction. Figure 6 (c) is the vessel's trajectory from western Indonesia to Lombok. Figures 6 (d) and 6 (e) show the crossing routes that pass through the TSS on Lombok Strait. Figure 6 (d) illustrates the trajectory of vessels going from Lombok to Karangasem Bali, and Figure 6 (e) is for the opposite direction. Figure 6 (f) is the vessel's trajectory from Lombok to western Indonesia. Those figures indicate that the proposed LCSS clustering framework has succeeded in distinguishing trajectories that have different directions even though they have a similar trajectory shape visually.

Table 3. Trajectories in cluster result

No.	Cluster	Number of trajectories
a	1st Cluster	100
b	2nd Cluster	77
c	3rd Cluster	27
d	4th Cluster	88
e	5th Cluster	35
f	6th Cluster	21
g	noise	57

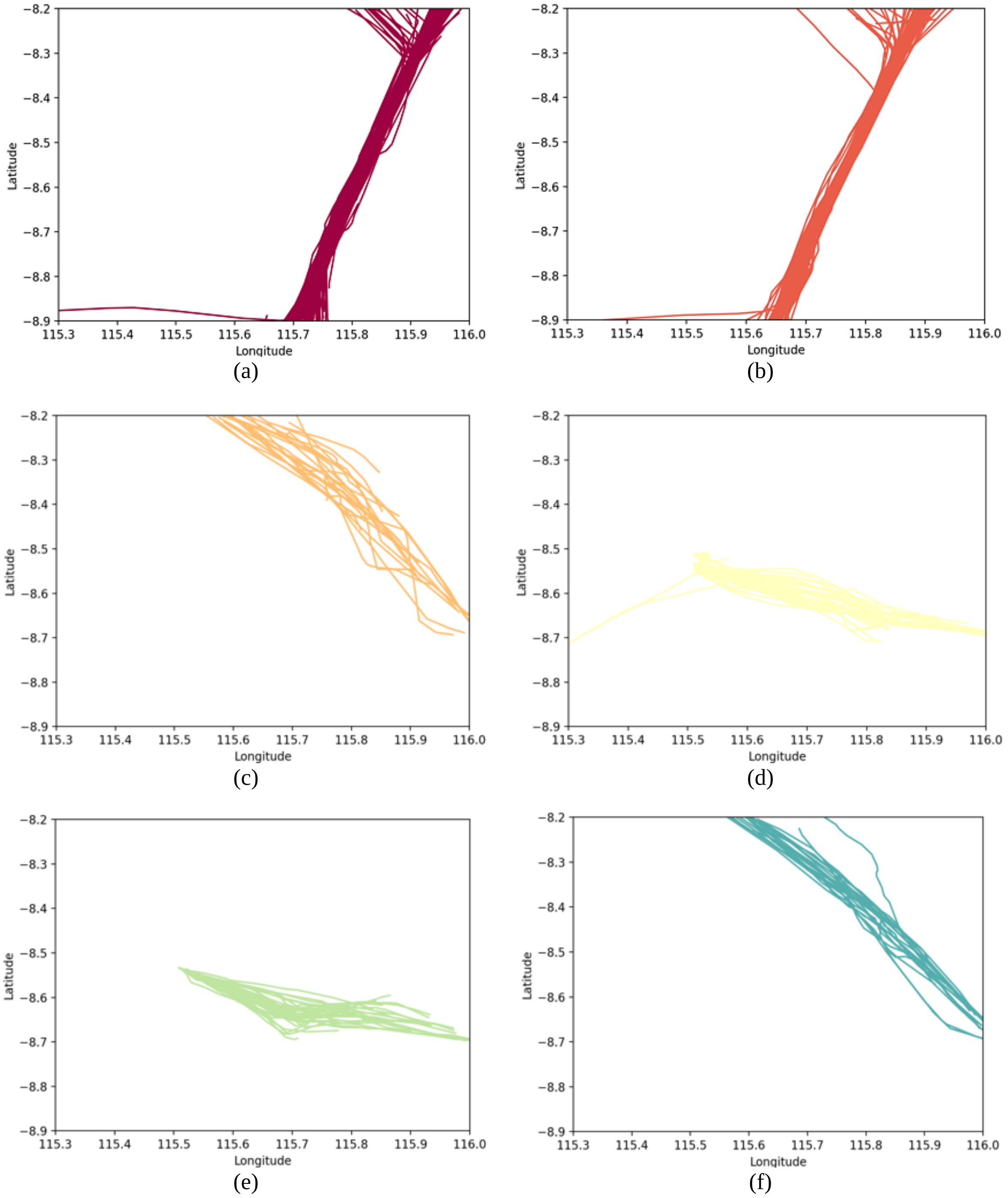


Figure 6. Vessel trajectory cluster of (a) Lombok Strait TSS south to north and (b) is the opposite; (c) Western Indonesia to Lombok; (d) Lombok to Karangasem and (e) is the opposite; and (f) Lombok to Western Indonesia

Figure 7 shows the visual trajectories that are included in the noise cluster. Trajectories that are included in the noise cluster are shorter in length in comparison to trajectories of vessels from southern Bali to northern Bali or vice versa. Those trajectories might be categorized as noise because the amount of data is very small.

3.6. Comparison with different algorithms

Here we provide the comparison of three similarity measurement algorithms, namely DTW, LCSS, and HD. Each algorithm uses the same compressed trajectory data. Three clustering algorithms were also compared, namely DBSCAN K-means and K-medoids. Table 4 shows the comparative description of each method. As shown in Table 4, the clustering process using the HD algorithm is the fastest, with only 18,465s. However, the HD algorithm cannot distinguish trajectories in the opposite direction because it only measures the trajectory distance based on the shape of the trajectory. DTW takes the longest time with a total clustering time of 34,886s. DTW is very affected by abnormal AIS trajectory data, so it cannot provide optimal distance between trajectories. The results of clustering with DTW get the lowest SC score, which is 0.135. The similarity measurement algorithm that can distinguish the direction of the trajectory with a high SC score is LCSS with a total clustering time of 23,468s. The comparison of clustering algorithms is carried out using the results of the most optimal similarity matrix in the previous comparison, namely LCSS. The parameters used in each algorithm are the parameters with the highest SC score. The K-means and K-medoids algorithms cannot identify the noise. Both algorithms get a high SC score while recognizing 4 clusters. LCSS+DBSCAN is the only one that can recognize noise, getting 6 clusters with an SC score of 0.533. This comparison shows that the framework with the proposed combination of algorithms can solve the problem of similarity measurement to noisy AIS data. The proposed framework can also distinguish the direction of the trajectory with a relatively fast total clustering time.

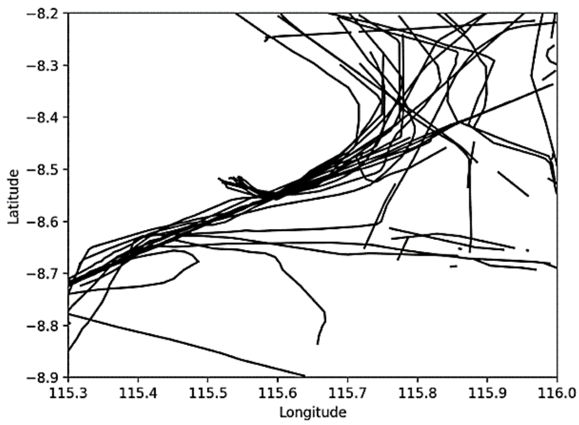


Figure 7. Trajectory noise

Table 4. Comparison results of different algorithms

No.	Proposed method LCSS+DBSCAN	DTW+DBSCAN [9]	HD+DBSCAN [16]	LCSS+K-means	LCSS+K-medoids
Opposite course	Yes	Yes	No	Yes	Yes
Detect noise	Yes	Yes	Yes	No	No
SC score	0.533	0.135	0.498	0.638	0.635
N cluster	6	6	6	4	4
Total time	23,468s	34,886s	18,465s	23,474s	23,472s

4. CONCLUSION

Trajectory clustering based on AIS data requires well-structured preprocessing steps due to the existence of some abnormal data. Moreover, the trajectory cannot be directly clustered with the clustering algorithm alone. Therefore, we propose a framework that combines several algorithms that can process AIS data from scratch to generate clusters. The main contribution of the proposed framework is a well-structured combination of algorithms in preprocessing, similarity measurement, and clustering to construct good quality clusters while minimizing total processing time. Our experiment shows that similarity measurement is the process that takes the longest time, and the chosen trajectory compression with DP significantly accelerates the process. We also observed that the LCSS algorithm is the optimal algorithm in similarity measurement of vessel trajectories based on AIS data. Furthermore, we found the right combination of MDS and DBSCAN for density-based clustering. The comparison in similarity measurement with DTW and HD, and the comparison of clustering with K-means and K-medoids show the performance advantage of the framework with the proposed combination of algorithms. Moreover, the proposed framework can distinguish trajectories in different directions, identify the noise, and produce clusters of good quality with relatively fast total processing time. However, the proposed framework still requires parameter determination for the DP, LCSS, and DBSCAN algorithms. Therefore, our future work will focus on investigating a parameter-free trajectory clustering framework for AIS data.

ACKNOWLEDGEMENTS

This research was supported and funded by the Ministry of Education, Culture, Research and Technology in the Republic of Indonesia with a contract number: B/136-18/UN14.4.A/PT.01.05/2021.

REFERENCES

[1] L. Li, W. Lu, J. Niu, J. Liu, and D. Liu, "AIS data-based decision model for navigation risk in sea areas," *Journal of Navigation*, vol. 71, no. 3, pp. 664–678, May 2018, doi: 10.1017/S0373463317000807.

[2] Z. Yan *et al.*, "Exploring AIS data for intelligent maritime routes extraction," *Applied Ocean Research*, vol. 101, pp. 1–10, Aug. 2020, doi: 10.1016/j.apor.2020.102271.

[3] X. Wu, A. Rahman, and V. A. Zaloom, "Study of travel behavior of vessels in narrow waterways using AIS data – A case study in Sabine-Neches Waterways," *Ocean Engineering*, vol. 147, pp. 399–413, Jan. 2018, doi: 10.1016/j.oceaneng.2017.10.049.

[4] K. T. Seong and G. H. Kim, "Implementation of voyage data recording device using a digital forensics-based hash algorithm," *International Journal of Electrical and Computer Engineering*, vol. 9, no. 6, pp. 5412–5419, Dec. 2019, doi: 10.11591/ijece.v9i6.pp5412-5419.

[5] S. Arasteh *et al.*, "Fishing vessels activity detection from longitudinal AIS data," in *Proceedings of the 28th International Conference on Advances in Geographic Information Systems*, Nov. 2020, pp. 347–356. doi: 10.1145/3397536.3422267.

[6] M. Fournier, R. C. Hilliard, S. Rezaee, and R. Pelot, "Past, present, and future of the satellite-based automatic identification system: Areas of applications (2004–2016)," *WMU Journal of Maritime Affairs*, vol. 17, no. 3, pp. 311–345, Sep. 2018, doi: 10.1007/s13437-018-0151-6.

[7] M. Svanberg, V. Santén, A. Hörteborn, H. Holm, and C. Finnsgård, "AIS in maritime research," *Marine Policy*, vol. 106, pp. 1–10, Aug. 2019, doi: 10.1016/j.marpol.2019.103520.

[8] E. Tu, G. Zhang, L. Rachmawati, E. Rajabally, and G.-B. Huang, "Exploiting AIS data for intelligent maritime navigation: A comprehensive survey from data to methodology," *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, no. 5, pp. 1559–1582, May 2018, doi: 10.1109/TITS.2017.2724551.

[9] H. Li, J. Liu, R. Liu, N. Xiong, K. Wu, and T. Kim, "A dimensionality reduction-based multi-step clustering method for robust vessel trajectory analysis," *Sensors*, vol. 17, no. 8, p. 1792, Aug. 2017, doi: 10.3390/s17081792.

[10] H. Li, J. Liu, K. Wu, Z. Yang, R. W. Liu, and N. Xiong, "Spatio-temporal vessel trajectory clustering based on data mapping and density," *IEEE Access*, vol. 6, pp. 58939–58954, 2018, doi: 10.1109/ACCESS.2018.2866364.

[11] P. Sheng and J. Yin, "Extracting shipping route patterns by trajectory clustering model based on automatic identification system data," *Sustainability*, vol. 10, no. 7, pp. 1–13, Jul. 2018, doi: 10.3390/su10072327.

[12] P. Last, C. Bahlke, M. Hering-Bertram, and L. Linsen, "Comprehensive analysis of automatic identification system (AIS) data in regard to vessel movement prediction," *Journal of Navigation*, vol. 67, no. 5, pp. 791–809, Sep. 2014, doi: 10.1017/S0373463314000253.

[13] I. P. N. Hartawan, I. M. O. Widyantara, A. A. I. N. E. Karyawati, N. I. Er, K. B. Artana, and N. P. Sastra, "AIS data pre-processing for trajectory clustering data preparation," in *2021 IEEE International Conference on Aerospace Electronics and Remote Sensing Technology (ICARES)*, Nov. 2021, pp. 1–5. doi: 10.1109/ICARES53960.2021.9665187.

[14] E. Tu, G. Zhang, S. Mao, L. Rachmawati, and G.-B. Huang, "Modeling historical AIS data for vessel path prediction: A comprehensive treatment," pp. 1–11, Jan. 2020, [Online]. Available: http://arxiv.org/abs/2001.01592

[15] Y. Zhang and G. Shi, "Trajectory similarity measure design for ship trajectory clustering," in *2021 IEEE 6th International Conference on Big Data Analytics (ICBDA)*, Mar. 2021, pp. 181–187. doi: 10.1109/ICBDA51983.2021.9403137.

[16] W. Yitao, Y. Lei, and S. Xin, "Route mining from satellite-AIS data using density-based clustering algorithm," *Journal of Physics: Conference Series*, vol. 1616, no. 1, pp. 1–8, Aug. 2020, doi: 10.1088/1742-6596/1616/1/012017.

[17] R. Zhen, Y. Jin, Q. Hu, Z. Shao, and N. Nikitakos, "Maritime anomaly detection within coastal waters based on vessel trajectory clustering and naïve bayes classifier," *Journal of Navigation*, vol. 70, no. 3, pp. 648–670, May 2017, doi: 10.1017/S0373463316000850.

[18] L. Wang, P. Chen, L. Chen, and J. Mou, "Ship ais trajectory clustering: An hdbscan-based approach," *Journal of Marine Science and Engineering*, vol. 9, no. 6, pp. 1–20, May 2021, doi: 10.3390/jmse9060566.

[19] L. Zhao and G. Shi, "A novel similarity measure for clustering vessel trajectories based on dynamic time warping," *Journal of Navigation*, vol. 72, no. 2, pp. 290–306, Mar. 2019, doi: 10.1017/S0373463318000723.

[20] G. Yuan, P. Sun, J. Zhao, D. Li, and C. Wang, "A review of moving object trajectory clustering algorithms," *Artificial Intelligence Review*, vol. 47, no. 1, pp. 123–144, Jan. 2017, doi: 10.1007/s10462-016-9477-7.

[21] Z. Chen, J. Guo, and Q. Liu, "DBSCAN algorithm clustering for massive AIS data based on the hadoop platform," in *2017 International Conference on Industrial Informatics - Computing Technology, Intelligent Technology, Industrial Information Integration (ICIICII)*, Dec. 2017, pp. 25–28. doi: 10.1109/ICIICII.2017.72.

[22] M. Z. Hossain, M. J. Islam, M. W. R. Miah, J. H. Rony, and M. Begum, "Develop a dynamic DBSCAN algorithm for solving initial parameter selection problem of the DBSCAN algorithm," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 23, no. 3, pp. 1602–1610, Sep. 2021, doi: 10.11591/ijeecs.v23.i3.pp1602-1610.

[23] Y. Chen, S. Tang, N. Bouguila, C. Wang, J. Du, and H. Li, "A fast clustering algorithm based on pruning unnecessary distance computations in DBSCAN for high-dimensional data," *Pattern Recognition*, vol. 83, pp. 375–387, Nov. 2018, doi: 10.1016/j.patcog.2018.05.030.

[24] A. Starczewski and A. Cader, "Determining the eps parameter of the DBSCAN algorithm," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2019, pp. 420–430. doi: 10.1007/978-3-030-20915-5_38.

[25] U. N. Wisesty and T. R. Mengko, "Comparison of dimensionality reduction and clustering methods for SARS-CoV-2 genome," *Bulletin of Electrical Engineering and Informatics*, vol. 10, no. 4, pp. 2170–2180, Aug. 2021, doi: 10.11591/eei.v10i4.2803.

[26] G.-Y. Zhang and J. Zhang, "Trajectory clustering based on trajectory structure and longest common subsequence," in *International Conference on Computer, Electronic Information and Communications (CEIC 2018)*, Aug. 2018, pp. 61–65. doi: 10.12783/dtcse/ceic2018/24525.

[27] A. Octavian, T. Trismadi, and P. Lestari, "The importance of establishing particularly sensitive sea areas in Lombok Strait: Maritime security perspective," *IOP Conference Series: Earth and Environmental Science*, vol. 557, no. 1, pp. 1–12, Aug. 2020, doi: 10.1088/1755-1315/557/1/012013.

[28] L. Zhao and G. Shi, "A method for simplifying ship trajectory based on improved Douglas–Peucker algorithm," *Ocean Engineering*, vol. 166, pp. 37–46, Oct. 2018, doi: 10.1016/j.oceaneng.2018.08.005.

[29] N. Saeed, H. Nam, M. I. U. Haq, and D. B. M. Saqib, "A survey on multidimensional scaling," *ACM Computing Surveys*, vol. 51, no. 3, pp. 1–25, May 208AD, doi: 10.1145/3178155.

[30] E. Schubert, J. Sander, M. Ester, H. P. Kriegel, and X. Xu, "DBSCAN revisited, revisited: Why and how you should (still) use DBSCAN," *ACM Transactions on Database Systems*, vol. 42, no. 3, pp. 1–21, Aug. 2017, doi: 10.1145/3068335.

[31] Y. H. R. *et al.*, "Massively scalable density based clustering (DBSCAN) on the HPCC systems big data platform," *IAES International Journal of Artificial Intelligence (IJ-AI)*, vol. 10, no. 1, pp. 207–214, Mar. 2021, doi: 10.11591/ijai.v10.i1.pp207-214.

BIOGRAPHIES OF AUTHORS

I Made Oka Widyantara received the B.Eng. in electrical engineering from Sepuluh Nopember Institute of Technology (ITS) Surabaya, Indonesia, in 1997. He completed M. Eng degree in the Telecommunication Information System from Bandung Institute of Technology (ITB), Bandung, Indonesia, 2001 and Dr. degree in the multimedia telecommunication from Sepuluh Nopember Institute of Technology (ITS) Surabaya, Indonesia, 2013, He is currently a lecturer in the Department of Electrical Engineering with Udayana University and head of Information Technology Department, Engineering Faculty, Udayana University, Indonesia. His research interests include intelligent signal processing, and image/video compression. He can be contacted at email: oka.widyantara@unud.ac.id.

I Putu Noven Hartawan received the B. Eng degree in Informatics Engineering from STMIK STIKOM Indonesia in 2018. Currently he is pursuing his M. Eng degree at the Postgraduate Program in Electrical Engineering, Udayana University. His interest is research on data mining and software architecture and optimization. He can be contacted at email: novenhartawan@gmail.com.

Anak Agung Istri Ngurah Eka Karyawati received the M. Eng degree in Information Science and Systems Engineering from Ritsumeikan University, Japan, and Dr. degree in Doctoral Program of Computer Sciences, Gadjah Mada University, Indonesia. She is currently a researcher and a lecturer in Informatics Department, Udayana University, Bali, Indonesia. Her research interests include text mining, computational language, natural language processing, information retrieval, and knowledge representation. She can be contacted at email: eka.karyawati@unud.ac.id.

Ngurah Indra Er is a lecturer in the Department of Electrical Engineering of the Udayana University (UNUD), Bali, Indonesia, since 2002. He obtained his PhD from IMT Atlantique, Rennes, France in 2021. He completed his M.Sc. in communications engineering from University of Birmingham, U.K., in 1999 after receiving his B.Eng. in electrical engineering from Sepuluh Nopember Institute of Technology (ITS) Surabaya, Indonesia, a year earlier. His current research interests include Internet of Things (IoT), Internet of Vehicles (IoV), vehicular networks, opportunistic networks, and smart city data collection. He can be contacted at email: indra@unud.ac.id.

Ketut Buda Artana received the M. Sc degree in Marine Engineering from the University of Newcastle Upon Tyne–UK in 1997 and Dr degree in the same field from Kobe University of Mercantile Marine, Japan in 2003, He is currently a professor in Maritime Reliability and Safety, Department of Marine Engineering, Faculty of Marine Technology ITS Surabaya. With colleagues, he has been developing and commercializing AISITS, a real time early warning system for protecting ships, subsea pipeline, subsea cable, and offshore platform. His current position is the head of safety laboratory and senior researcher at the Center of Excellence of Maritime Safety and Marine Installation (PUI-KEKAL). His research interests include maritime risk management, marine safety assessment, reliability engineering, LNG technology. He can be contacted at email: ketutbuda@its.ac.id.