Dock Clustering Management System based on Modified K-Mean Algorithm in Smart Port Services

Ari Wijayanti, Okkie Puspitorini, Nur Adi Siswandari, Haniah Mahmudah and Revfath Risqon Syafaat
Department of Electrical Engineering, Politeknik Elektronika Negeri Surabaya, Indonesia

Keywords: Smart Port, Lora, Modified K-Means Algorithm, Ships.

Abstract: In queueing management system, waiting time is a parameter that indicates times duration when the ship enters the port until it rests on the dock to stevedoring process. The problem in waiting time is taking a long time to place the ships in the suitable dock for unloaded activity. The condition caused by the ship's specifications does not match with the specifications of the dock so the ship has to wait for a long. To overcome this problem, this paper proposed the Modified K-Means Algorithm is used to clustering the ship into the suitable dock. The clustering of the ships is based on Length of All (LOA) or length of the ship, gross tonnage of ships, and commodity type from ship's cargo. Labeling data of dataset according to the specification of Port then they are trained using K – Cross-Validation to determine centroid. The ship's specifications are provided by LoRa (Long Range) device after it is detected first when the ships enter the port area. The LoRa calculates the distance of each ship using the Euclidean Distance Formula and grouping the ships based on minimum distance. The experiments have been conducted with 70 ships, which are clustered by 10 ships on each dock. The result is the accuracy of the Modified K-Means algorithm in clustering ships were reached 91.4%. This percentage indicates that the ships were successfully clustering at the correct dock according to the specifications of the ship and the dock

1 INTRODUCTION

As a nation with thousands of islands, ports are become the most important gate to enter Indonesia territorial. The main role of ports is used as an export and import distribution in the trade sector. The ports in Indonesia have the weakness lies in the quality of infrastructure and port management such as productivity in loading and unloading, severe congestion conditions, and maintaining old customs documents (Suprata, 2020). The cost of distribution services by ships will increase as long as the duration time to park ships over the sea while waiting for the suitable dock is ready to be entered (Nguyen, 2019). This condition causes the ship queueing to get longer in the parking area so that the waiting time increase than before at the port (Unnati, 2017) (*S.P.Singh*,2013). Such conditions lead to the emergence of innovation to create a system to solve this problem. This system serves to issue a queue for ships to be anchored so that it is expected to be able to unravel the queues of ships going to the port. Modified K-Means is one of the simplest and most common clustering methods that can be used in ship queue management. It can group large amounts of data with relatively fast and efficient computation time. However, The results of the cluster formed by the K-means clustering method are very dependent on the value of the initial point cluster initiation. Therefore, the Modified K- Means algorithm is proposed to counter the problems. This algorithm is following predetermined parameters as a previous study in *research* (Emre,2011) (Oyelade*,2010)*.

By implementation of Modified K- Means algorithm in queueing management system, the number of waiting times due to misplaced ships will be overcome and more coordinated to optimize port performance. In previous studies conducted, the vessel detection processwas carried out with an ultrasonic sensor, where this sensor has a limited range to detect vessels (Swapna Ch, 2017) (A Kamalov, 2019) The weakness is the process of detecting objects tends longer and there is no port

Wijayanti, A., Puspitorini, O., Siswandari, N., Mahmudah, H. and Syafaat, R.
Dock Clustering Management System based on Modified K-Mean Algorithm in Smart Port Services.
DOI: 10.5220/0010951200003260
In *Proceedings of the 4th International Conference on Applied Science and Technology on Engineering Science (iCAST-ES 2021)*, pages 692-697
ISBN: 978-989-758-615-6; ISSN: 2975-8246

clustering process first, cause system performance to be less than optimal in determining the port for ships (Unnati, 2017) (*S.P.Singh*,2013). In this research, LoRa (Long Range) sensors are used to detect the arrival of ships into the port at the first with their specification, assumed by a normal condition in the sea environment. At the port gate, the system clustering them and classify the suitable dock by implementing the modified K- Means Algorithm. It affects the efficiency of placement time in the fastest matching process than before so that the determination of the destinationof the ship's dock will be faster and reduce the waiting time list. All of the data in this research are based on Tanjung Perak port container terminal.

This paper is organized as follows: section 1 introduction, section 2 is the theory, Section 3 is the system design and section 4 is the conclusion.

2 MODIFIED K-MEANS ALGORITHM

Modified K - Means Algorithm is one of the most popular algorithms used for clustering data because it can be easily implemented and is the most efficient in terms of execution time. With this algorithm, data of a similar type is tried to be groupedfrom large data sets carried out by repeated calculations. As a result, the computational complexity of this algorithm is very high*(Oyelade,2010).* Several studies have been conducted to minimize this, K-Means algorithm is a sensitive algorithm because the process must determine the correct centroid value as a reference cluster. The error that often occurs in this algorithm is when determining the first centroid, if the first centroid is not appropriate then the cluster results obtained are also not optimal or even an empty cluster. To calculate the distance between data and centroids, the K- Means algorithm uses the Euclidean distance formula (Sharfuddin,2015)

K-Means Algorithm is used to clustering data by calculating the closest distance between the data centersdetermined by each data, the following step K - Means algorithm works in clustering data (Vaishali,2011):

1. Determine the number of K clusters.
2. Determine the centroid randomly
3. Calculate the distance of each data using the Euclidean distance formula
4. Clustering each data by clustering according to the minimum result of the Euclidean distance from eachcentroid
5. Compute new centroid from new clusters
6. Iterate until a convergence condition occurs

Convergence condition is a condition where the members of each cluster do not change after an iteration. Thefollowing Euclidean distance formula is used to calculate the distance of each data by centroid as in equation(1) (M Emre Celebi,2011)

$$\text{Euclidean Distance} = \sqrt{\sum^{n} (Pk - Qk)^2} \quad (1)$$

Where, n = Total Data,

Pk = Value of Centroid

Qk= Value of Each Data

2.1 K – Cross-Validation

K-Fold Cross Validation is primarily used in applied machine learning to estimate the skill of a machine learning model on unseen data. That is, to use a limited sample to estimate how the model is expected to perform in generalwhen used to make predictions on data not used during the training of the model. An example of applying k-fold cross-validation is shown in

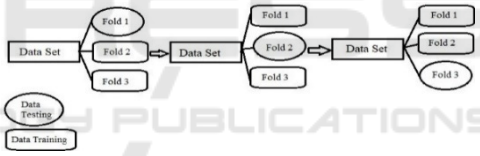

Figure 1: Example of cross k-fold validation.

In thisresearch, the K-Means Algorithm is used for the clustering process in the dataset. Some modifications are made to the K-Means Algorithm, in addition to determining the randomcentroid value in the original K-Means Algorithm, which can causesuboptimal clustering results, so that the centroid generation section also requires modification, namely thelabeling of the dataset. In this research, the dataset comes from Pelindo based on the specifications of the port and ships. First, it has trained the dataset using K-Cross-Validation Method tofind out the accuracy of the centroid value which is generated later after conducting the data training process. Later the centroid generation process is done with the python program.

The following step Modified K - Means algorithm worksin clustering data.

1. Input Dataset from Pelindo
2. Labeling Data of Dataset according to Specification of Port

3. Training Data Using K – Cross-Validation
4. Determine Centroid
5. Calculate the distance of each data using EuclideanDistance Formula
6. Grouping based on the minimum distance

3 SYSTEM DESIGN

The overall design of the system starts when the ship entering the port, the system will determine the ship's dock tocarry out the loading and unloading process. For the clustering process, the system will use Modified K-MeansAlgorithm with parameters that have been obtained from thedatabase system. The parameters used by the system are LOA and type of commodity. Clustering willprevent the ship from parking its ships that are not following the specifications of the ship and dock. After that, the scout ship will pick up the ship to carry out the loading andunloading process at the specified dock. For the entire scheme and flowchart system shown in Figure 1 and Figure 2.

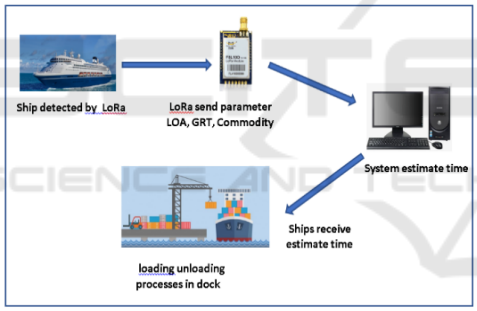

Figure 2: System Scheme.

From the flowchart in Figure 3, The first thing the systemdoes is collecting data collection, where this data collection is a reference for the algorithm that the system is used. The ship will enter the port's gate after it receives the dataset from LoRa sensors. The port will be clusterized using the k-means algorithm with the ship length parameters and the types of commodities that have been adjusted to the previous data. At the end of the process is the ship gets information about a suitable dock for unloading the cargo.

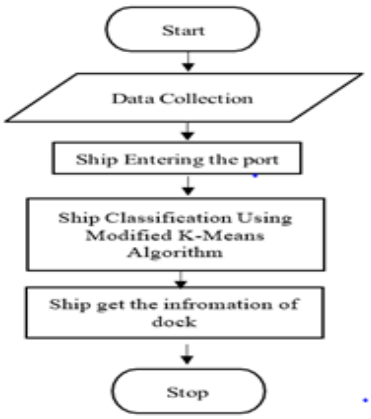

Figure 3: Flowchart System.

3.1 Data Collection

The data that is used as a reference for the Modified K-Means Algorithm in clustering ships is as follows:

- Specification of Ship
- Specification of Dock
- Dock Operational Time
- Dock Facilities
- Dock Service
- Ship loading and Unloading Mechanism

These parameters are needed by the system as a reference in clustering. It is needed by the system as a reference in clustering. Dock facilities and services are important data to match the utility of the ship.

Specification of the ship is primary data to seek a correct dock for parking time as long as the unloading cargos process. During the process, the management pays attention to dock operating time to inform the gate. This part is an important step of the ship loading and unloading mechanism.

Some specifications of the ship that K-Means Algorithm needed to classify ship according to the dock as follows: Length of All (LOA) of Ship means the size of the ship and the commodity types of ship's cargo. Based on the data of the Tanjung Perak port management, the dock types are divided into some utility such as for passenger, dry bulk, liquid bulk, and general cargo named. The Specification of dock services Port as shown in Table 1.

Table 1: Specification of dock services at Tanjung Perak port.

Dock	Commodity Services
Jamrud Utara	Passenger
	Dry Bulk
	General Cargo
Jamrud Barat	General Cargo
	Dry Bulk
Jamrud Selatan	General Cargo
Kalimas	General Cargo
Mirah	Liquid Bulk
	General Cargo
	Container
Berlian Timur	Container
Nilam Timur	Liquid Bulk
	Dry Bulk
	General Cargo
	Container

Table 2: Dock Dimension in Tanjung Perak Port.

Dock	Length (m)	Width (m)
Jamrud Utara	1200	15
Jamrud Barat	217	15
Jamrud Selatan	800	15
Kalimas	2270	15
Mirah	640	15
Berlian Timur	780	15
Nilam Timur	920	15

Figure 4 shows the flowchart of the Modified K-means algorithm for clustering ships. The clustering process of ships needs the dock dimension. Each ship will be matched first in the dimension of the dock before continuing to commodity types of cargos. Loading and unloading cargos mechanism are done inside the suitable dock according to a specified time. Table 2 shows dock dimension in length and width.

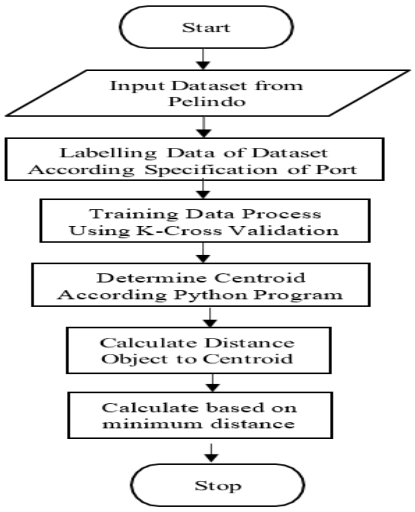

Figure 4: Flowchart Modified K-Means Algorithm to Classify Ship.

3.2 Training Data Process for Ship Clustering

In the K-Means algorithm, there is a step that determinesthe centroid value, for this reason, the data training stage is performed first to see the accuracy of the centroid value generated later. The training data is carried out using the k- cross-validation method to determine the accuracy of the centroid value that will be generated later.

After training as many as 300 data with the K-Means Algorithm to classify port according to ship specifications and use 21 tests with a comparison of the number of a differentnumber of k-folds. In figure 5 shows the process of taking data sets of 300 data from each dock with their commodity has been obtained from Pelindo III Tanjung Perak port container terminal Surabaya.

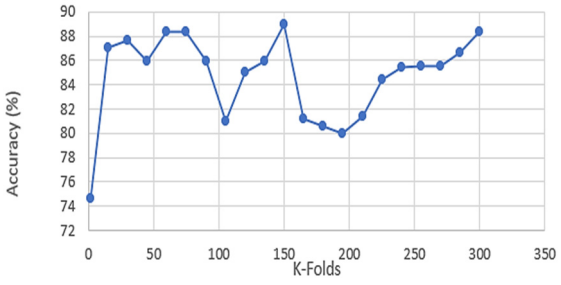

Figure 5: Result of the training data process.

In the process of training the K-Means algorithm in clustering port in Table 3, after testing 21 times where each test changes the value of

k_folds where this value is very influential in seeing the accuracy of the training process andit is found that the average accuracy of 84.658 percent with adata set of 300 data set. Later in generating centroid values that take from the Port data and which will be made the partisan k-means algorithm in determining ship clustering will get a k fold accuracy average of around 84.658%. The percentage means the accuracy is still high and good. According to this process, K-Means Algorithm can be implemented in this system especially in the ship clustering process.

Table 3: Result of training data for ship clustering.

Dataset	K_fold	Accuracy (%)
300	2	74.667
300	15	87
300	30	87.667
300	45	85.926
300	60	88.333
300	75	88.333
300	90	85.926
300	105	80.952
300	120	85
300	135	85.926
300	150	89
300	165	81.212
300	180	80.556
300	195	80
300	210	81.429
300	225	84.444
300	240	85.417
300	255	85.490
300	270	85.556
300	285	86.667
300	300	88.333
Average		84.658

3.3 Centroid Generation Process

For the generation of centroid values for each dock and commodity using the python program with the dataset from Pelindo, the results are as shown in Table 4. This centroid value will be the reference for the k means algorithm in clustering ships according to the specification of ships and the specification of each dock. The graph for the value of centroid for each dock and commodity is shown in figure 6.

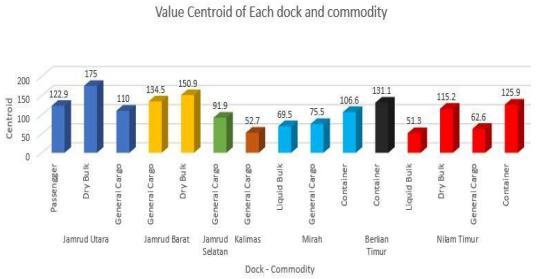

Figure 6: Value of Centroid for Each Dock – Commodity.

Table 4: The results of the centroid generation process.

Doc	Commodity Services	Centroid
Jamrud Utara	Passenger	122.9
	Dry Bulk	175
	General Cargo	110
Jamrud Barat	General Cargo	134.5
	Dry Bulk	150.9
Jamrud Selatan	General Cargo	91.9
Kalimas	General Cargo	52.7
Mirah	Liquid Bulk	69.5
	General Cargo	75.5
	Container	106.6
Berlian Timur	Container	131.1
Nilam Timur	Liquid Bulk	51.3
	Dry Bulk	115.2
	General Cargo	62.6
	Container	125.9

4 RESULT

After conducting experiments with 70 ships, which 10 ships every dock at Jamrud Utara, Jamrud Selatan, Jamrud Barat, Kalimas, Mirah, Berlian Timur and Nilam timur.

Table 5: Result for ships clustering for each dock.

Dock	ships	Success	Fail	Accuracy (%)
Jamrud utara	10	10	0	100
Jamrud barat	10	9	1	90
Jamrud Selatan	10	9	1	90
Kalimas	10	10	0	100
Mirah	10	9	1	90
Berlian Timur	10	9	1	90
Nilam Timur	10	8	2	80
Average				91.4

This success statement indicates that the ship was successfully clusterized at the correct dock according

to the specifications of the ship and also the dock, while the failed statement indicates that the ship was not successfully classified correctly.

Based on Table 5, it can be seen that in the clustering process of 70 ships with 10 ships per dock get different results, at the Jamurd Utara and in Kalimas all 10 ships have been correctly classified so that they get an accuracy of 100%, then at the Jamrud Barat, Jamrud Selatan, Mirah, and Berlian Timur every 9 ships were successfully classified and only one ship failed to be classified correctly to get an accuracy of 90%, then at the Nilam Timur, 8 ships are classified correctly so that only get an accuracy value of 80%, from all dock the average gets an accuracy of 91.4% Graph results of Ships Clustering shown as figure 7

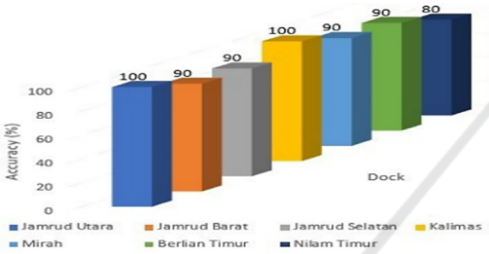

Figure 7: Result accuracy of Ships Clustering.

In the clustering process using modified K-Means algorithm, there are some errors in clustering ships, this is because the centroid value generated between the same commodity with different dock has almost the same value or there is no significant difference in value, this causes errors in clustering ships, especially dock with the specification that has the same type of commodity.

5 CONCLUSIONS

This paper, concludes that the Modified K-Means Algorithm clusterized the ships' accuracy to 91.4% to overcome placement errors that exceed the value of residence time by using LOA parameters and commodity types from the ship specifications. By clustering the ship by the specified dock, a high waiting time value caused by incorrect placement of the ship can be reduced appropriately, so that it can optimize the performance of the port. The future work it is desirable to have higher accuracy by applied and combine with other algorithms.

ACKNOWLEDGEMENTS

Thanks are due to Maritim teams Revfath Risqon Syafaat, Fahmi Nurdin Handy Novian, Dimas Khrisna Ramadhani. and PELINDO III Surabaya for your cooperations

REFERENCES

Unnati R. Raval and Chaita Jani (2016). Implementing & Improvisation of K – Means Clustering Algorithm. *International Journal Computer Science and Mobile Computing,* vol. 5, issue. 5, pg. 191-203

Dr. S. P. Singh and Ms. Asmita Yadav (2013). Study of K – Means and Enhanced K – Means Clustering, *International Journal of Advanced Research in Computer Science.* vol 4, No 10

M Emre Celebi(2011). Improving the performance of K-Means for Color Quantization. *Image and Vision Computing.* Vol 29(4).260- 271.

Oyelade, O.J, and Oladipupo O.O and Obagbuwa I.C (2010). K – Means clustering algorithm for students academic performance. *International Journal of Computer Science and Information Security.* vol 7, No 1.

Sharfuddin Mahmood (2015). A Proposed Modification of K-Means Algorithm. *International Journal of Modern Education and Computer Science,* Vol 6, pp 37-42

Vaishali R.Patel (2011). Modified K-Means Clustering Algorithm. *Computational Intelligence and information technology (CIIT).* pp 307-312.

Sk Ahammad Fahad," A modified K-Means Algorithm for Big Data Clustering (2016). *International Journal of Computer Science Engineering and Technology,* Vol 6, Issue 4,129-132

Swapna Ch, Mukesh KumarNaswal Kishor Y and Niraj (2017). "tracking and emergency detection of inland vessel using GPS-GSM System, *international Conference on Recent Trends in Electronics, Information and communication technology (RTEICT)*

Ahmad Kamalov and suhyun Park (2019). An IoT based Ship Berthing Method using a Set of Ultrasonic Sensors. *MDPI jurnal sensors 2019.* 19,5181;doi: 10.3390/s19235181

Minh-Duc Nguyen, Sung-june Kim, 2019, An Estimation of the Average Waiting Cost of Vessels Calling Container Terminals in Northern Vietnam, *Journal of the Korean Society of Marine Environment & Safety*, Vol. 25, No. 1, pp. 027-033, February 28, 2019, ISSN 1229-3431(Print) / ISSN 2287-3341(Online)

F. Suprata (2020), analysing the cause of idle time in loading and unloading operation at indonesian international port container terminal: port of tanjung priok case study. *Iop Conference Series materials science and engineering 847 012090*, doi: 10.1088/ 1757-899x/847/1/012090