OPEN ACCESS
J. OF INISTA, VOL. 3, NO. 2, PP.043-052, MEI 2021

ISSN : 2622-8106 (ONLINE)
DOI: 10.20895/INISTA.V3I2

Journal of Informatics, Information System, Software Engineering and Applications

The Concept of Big Data Analysis for Maritime Information on Indonesian Waters using K-Means Algorithm

Bita Parga Zen #1

 # Faculty of Informatics, Telkom Institute of Technology Purwokerto
Jl. DI Panjaitan No.128, Purwokerto, Banyumas Regency, Central Java 53147

[1] bita@ittelkom-pwt.ac.id

accepted on 07-07-2021

Abstract

***Abstract*—** Indonesia as an archipelagic country has a strategic geographical location, which is located between two continents and two oceans so that it has many advantages, especially in the maritime sector. Indonesia has a goal of becoming a World Maritime Axis that is responsible for ensuring the security and safety of services based on UNCLO S 1982. One way to achieve this goal is to process Big Data to produce useful maritime information . Data can be obtained from government agencies and international organizations and processed according to big data analytics so that it can be visualized information that can be used by various related parties. The collected data can be clustered using the K-Means algorithm with the aim of dividing the data into several groups. This is especially useful for supporting Indonesia based on UNCLOS 1982 by providing information and ensuring the safety of Indonesian waters and being able to contribute to the Indonesian economy.

Keywords: *maritime, big data, analysis*

I. INTRODUCTION

The Unitary State of the Republic of Indonesia is an archipelagic country consisting of 1/3 land area and 2/3 sea area with a total ocean area of 3,544,743.9 km^2. Indonesia consists of 17,504 islands which lies between latitudes 11°S and 6°N, and longitudes 95°E and 141°E and have a coastline of 54,716 km . Indonesia's territorial waters are directly adjacent to 10 countries, namely Malaysia, Singapore, Thailand, the Philippines, Vietnam, India, Papua New Guinea, Australia, Palau and Timor Leste. Indonesia is a country that is in a tropical region whose climate conditions are influenced by three types of climate, namely the monsoon climate, marine climate and tropical climate. The monsoon climate is influenced by the monsoons which change every 6 months. The tropical climate occurs because Indonesia is located in a tropical area, while the marine climate is due to the fact that Indonesia has a very large sea area and causes evaporation to cause rain.

Indonesia is located between two continents and two oceans and has 4 choke points in the world, namely the Malacca Strait, Sunda Strait, Makassar Strait and Lombok Strait and has 3 ALKI routes. ALKI is Indonesian Archipelago Sea Lanes which have been established through Government Regulation No. 37 of 2002 as a form of Indonesia's responsibility as an archipelagic country and has an obligation to provide routes for international shipping. The ALKI route provides the right of archipelagic channel passage and defines which routes are allowed by ships and aircraft.

Fig 1. Indonesian Archipelago Sea Lanes [1]

Based on Government Regulation no. 37 of 2002, Indonesia established three main ALKIs and several of its branches, namely ALK I, which connected shipping from the South China Sea to the Indian Ocean or vice versa across the Natuna Sea, Karimata Strait, Java Sea and Sunda Strait. ALK II is a voyage from the Sulawesi Sea to the Indian Ocean across the Makassar Strait, Flores Sea and Lombok Strait. Meanwhile ALK III is a voyage from the Pacific Ocean to the Indian Ocean or vice versa, crossing the Maluku Sea, Seram Sea, Banda Sea, Ombai Strait and Savu Sea and is divided into four parts.

The concept of an archipelagic state began to be accepted based on the 1982 United Nations Convention on The Law of the Sea (UNCLOS), article 46 which reads:

“Archipelagic State” means a State constituted wholly by one or more archipelagoes and may include other islands; "archipelago" means a group of islands, including parts of islands, interconnecting waters and other natural features which are so closely interrelated that such islands, waters and other natural features form an intrinsic geographical, economic and political entity, or which historically have been regarded as such.

UNCLOS is the result of a UN conference on the law of the sea which lasted from 1973 to 1982. UNCLOS 1982 is the legal basis for regulating the sea and its use, including the rights and obligations concerning straits and waters used for international waters. These obligations include:

1. The coastal State shall make a clear publication of the hazards associated with navigation. This is also as regulated in Article 24 paragraph (2) UNCLOS which reads as follows:
   The coastal State shall give appropriate publicity to any danger to navigation, of which it has knowledge, within its territorial sea.
2. The coastal State is obliged to provide basic necessities related to shipping, such as lighthouses and rescue facilities for ships requiring assistance;
3. The coastal State may not obstruct foreign ships with rights of peaceful passage [2].
4. The coastal state may not collect fees from foreign ships passing through its territorial sea

Articles 53 and 54 of UNCLOS 1982 state that foreign ships must respect and follow the statutory regulations of an island nation. In addition, UNCLOS 1982 states that archipelagic countries have sovereignty in inland waters, archipelagic waters and territorial seas. On the other hand, based on this statement, Indonesia which combines maritime zones together with the term Indonesian Waters has the consequence that an archipelagic country like Indonesia must accommodate international interests, especially in shipping and flights through archipelagic waters and territorial seas. Geofrey stated that maritime is always related to utilizing the four functions of the sea, namely marine resources, transportation media, physical environment and sovereign territory. This can be achieved by fulfilling three main elements, namely maritime environmental awareness, effective policies and integrated governance [2].

In November 2014 President Jokowi conveyed the concept of the marine sector, namely the World Maritime Axis (PMD) at the East Asia Summit forum in Naypyidaw Myanmar. The World Maritime Axis has the goal of making Indonesia a large, strong and prosperous maritime country through re-manifesting Indonesia's identity as a maritime country, empowering maritime potential for the realization of equal distribution of the Indonesian economy and safeguarding maritime interests and security. In developing the marine sector to realize a World Maritime Axis, there are five pillars that are emphasized, namely rebuilding maritime culture, managing the maritime economy while preserving maritime resources, prioritizing maritime infrastructure development, maritime diplomacy and strengthening Indonesia's maritime defense force. Some things that can be done are maritime development from various aspects such as infrastructure, politics, law, socio-culture, economy and security. In addition, Indonesia needs to uphold the sovereignty of the Indonesian sea area, strengthen and develop maritime networks, revitalize the maritime economy, rehabilitate environmental damage and increase the quality and quantity of human resources in the maritime sector [1].

Based on UNCLOS 1982, Indonesia is obliged to guarantee the security and safety of shipping in Indonesian waters and for the realization of Indonesia as a World Maritime Axis, Indonesia needs a system to manage various maritime information for all stakeholders and law enforcement officers at sea.

II. LITERATURE REVIEW

Big data is defined as a collection of data with a very large and complex amount which can no longer be handled by conventional computer technology [3]. Big data has three characteristics, namely scale (volume), distribution (velocity) and diversity (variety).

Fig 2. Big Data [4]

Big data volume is a form of data whose volume is very large so that it is not sufficient for only one machine, therefore special tools are needed to store and analyze the data. Velocity or data rate refers to how fast the data is generated. Data generated by certain sources can arrive at very high speeds, for example social media data or sensor data. Speed is another important characteristic of big data and the main reason for its exponential growth. High data rates result in a very large volume of data collected in a short span of time. Big data variation refers to the shape of the data. Big data comes in various forms such as structured, unstructured, or semi-structured, including text data, images, audio, video, and sensor data. Big data systems need to be flexible enough to handle a wide variety of data. Veracity refers to how accurate the data is. To extract values from data, the data needs to be sorted to remove noise so that maximum results can be obtained. Therefore, data cleaning is important so that erroneous data can be filtered out. Value or data value refers to the use of data for the purpose to be achieved. The ultimate goal of a big data analytics system is to extract value from data. Data values are also related to the authenticity or accuracy of the data [5].

To perform big data analysis, big data works according to the flow, namely data collection, data preparation, analysis types, analysis modes and visualization. Big data analysis is widely developed and widely used in various fields such as:

1. Commercial and business enterprises: big data can be used to increase efficiency and guide the decision-making process
2. In nature, big data can be used to view and monitor natural processes or changes using satellite imagery data, weather, radar and terrestrial sensing monitoring devices.
3. Governments and the public sector are using massive cloud storage to reduce operating costs, upgrade times and downtime. The data is processed and optimized for data retrieval in times of emergency.
4. Trading, business and economic systems use big data analytics to increase operating margins and take advantage of new opportunities to increase revenue.
5. Social media uses analytical big data to process information on social networks to determine the impact of information, enable the formation of information flows and help understand collective patterns of behavior [6].

Big data is widely used and provides significant benefits in areas such as finance, media, telecommunications and healthcare. However, its application in the maritime sector has been slow. According to the Ericsson report, the maritime industry lags behind other transportation industries in terms of use of information and communication technology. Currently, only a handful of marine users take advantage of big data. There are several benefits that users can get through the use of big data from various sources such as ports and ship movements. The system can analyze these data points to provide shipping information as well as identify efficiencies such as a faster route or preferred port [7].

In August 2015, the Maritime and Port Authority of Singapore (MPA) signed a Memorandum of Understanding (MOU) with IBM to conclude a two-year agreement on big data. The agreement involves developing a platform using IBM's Traffc Prediction tool to estimate ship arrival times and estimate potential traffic jams using fusion analytics. The platform also relies on data mining and anomaly detection using the IBM Incident Detection Module and IBM System G. According to Goh Kwong Heng, CIO MPA, port authorities are planning to invest in big data to improve port operations and activities.

MPA aims to use a data analytics platform to complement its port management system in detecting anomalies and supporting operations and planning processes. The authority also plans to invest in technology such as drones and mobile applications to manage marine accidents and improve the efficiency of its port workers. The port uses a platform to improve productivity and marine safety at a major trans-shipping hub. For example, by gathering information about ship positions and weather data, platforms help avoid accidents by inferring the path the ship is most likely to take in a given situation. It also helps prevent illegal bunkering by detecting unusual movement patterns [8].

Indonesia's vision to become a World Maritime Axis can be implemented by utilizing big data to become information and knowledge that can be used by various stakeholders in obtaining information about the condition of Indonesian waters. Data can be obtained from various related stakeholders, including government agencies, international organizations and private services.

A. Big Data Flow

Big data analysis consists of five stages to process data into useful information that can be used by users.

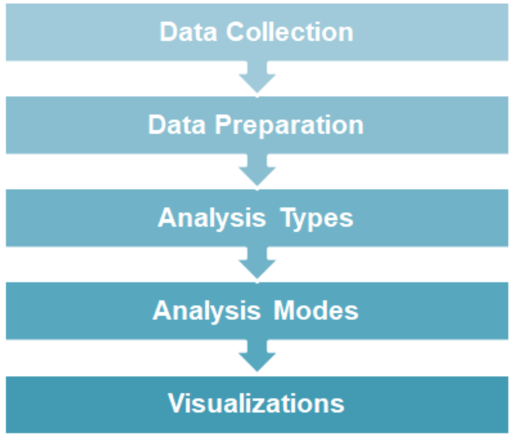

Fig. 3. Big data flow [4]

Data collection is the first step for analytical applications. Before data can be analyzed, the data must be collected and entered into a database. The choice of tools and applications for data collection depends on the data source and the type of data being absorbed. Indonesian maritime data can be obtained from government agencies, other state agencies, international organizations and private services. Government agencies are the Indonesian Navy, Bakamla (Indonesian Maritime Security Agency), the Air and Water Police Corps of the Republic of Indonesia National Police Security Maintenance Agency and the Meteorology, Climatology, and Geophysics Agency (BMKG). International organizations, namely the International Maritime Organization (IMO), International Maritime Bureau (IMB) and ReCAAP which manage and are responsible for marine activities, safety and security of shipping activities. The data obtained is in the form of ship platform data, weather data and shipping traffic condition.

Data can be in the form of text, images, audio or video. Ship platform data is in the form of ship types such as ferries, container ships, tankers, cargo, barges etc. Weather data in the form of rainfall, wind direction, wind speed, temperature etc. Meanwhile, data on shipping traffic conditions are in the form of sea traffic density, ship arrivals and departures, etc.

The second step is data preparation, which is sorting data to avoid data inaccuracies. Data that has not been filtered can have various problems such as corrupted data, missing data, duplication of data, inconsistent abbreviations, typos, incorrect spelling, and incorrect formatting. The data preparation step involves various processes such as data cleaning, data disputing, de-duplication, normalization, sampling and filtering. Data cleaning can detect and resolve problems like corrupted data, lost data, poorly formatted data etc. Data may be concerned with converting data from one raw format to another. For example, when collecting records as raw text files from different sources, it is possible to find inconsistencies in the field separators used in different files. Some files might use commas as field delimiter, others might use tabs as field delimiter. Data wrangling resolves this inconsistency by parsing raw data from different sources and converting it to one consistent format. Normalization is required when data from different sources use different units or scales or have different abbreviations for the same thing. For example, water traffic data reported by some stations might use distances on the kilometer scale while data from other sources uses the mile scale. Filtering and sampling is useful when it comes to processing data that meets certain rules and sorting out data with incorrect values.

The third stage is the type of analysis that will be used in processing data. Analysis can use algorithms according to goals and needs. Various types of analysis that can be used are basic statistics, classification, regression, clustering, recommendation, time series analysis, dimensionality reduction, text analysis, graph analysis, and pattern mining. Figure 4 describes the types of algorithms that can be used according to the chosen analysis method [9].

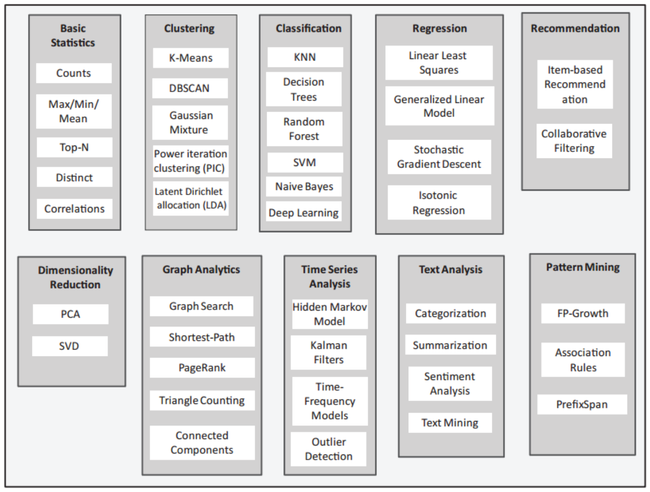

Fig. 4. Type Analysis [4]

This study uses a clustering analysis type with the K-Means algorithm to classify data according to their similarities. Clustering analysis is a type of unsupervised classification, namely the classification of data based on their similarity by identifying similar classes of objects to find the distribution patterns between attributes. Clustering analysis can be used to distinguish each group or class of objects according to the characteristics of each object [10].

III. RESEARCH METHOD

K-means algorithm is an algorithm that can input the number of k clusters and databases containing n object data, and produce a minimum cluster "k". The K-means algorithm accepts a number of "k" inputs; then dividing the data object "n" into "k" clusters, so that the cluster obtained is: the similarity of objects in the same cluster is higher; and objects in different clusters are similar to lower. The cluster similarity is calculated using the "central object" obtained by the mean of each central object cluster [11]. The processing flow is as follows:

1. Randomly select object "k" from n data objects as the center of initial grouping;
2. Calculating the distance between each object and the central object in accordance with the mean (center object) of each object cluster using the euclidean distance.
3. Recalculate the mean (central object) of each new cluster;
4. The center object is calculated repeatedly until the centroid value does not change anymore. [11]

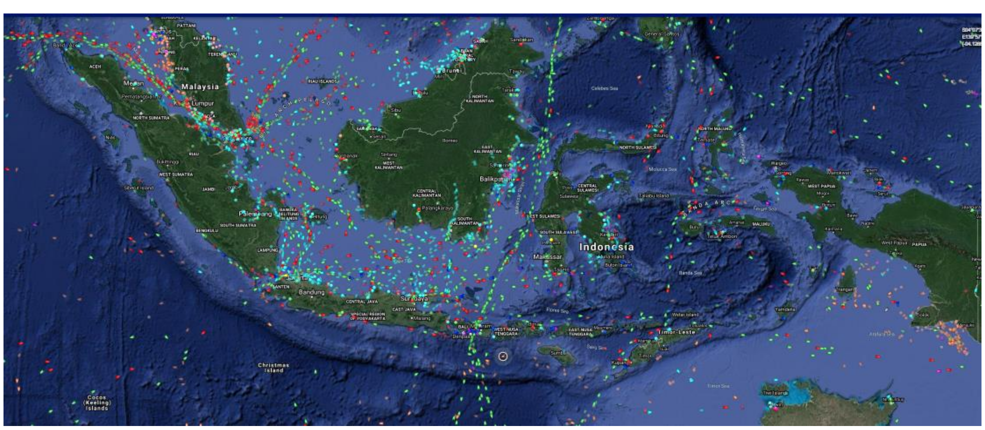

Fig. 5. Indonesian waters traffic [12]

Based on Figure 5, it can be seen that Indonesian marine traffic data used is the coordinate point of the ship's location to determine the point of density of Indonesian marine traffic. The coordinate points are divided into several clusters and calculated using the K-Means algorithm to find the densest point in Indonesian waters.

The final stage is visualization in maritime information in the form of reporting or presentation of maritime information after the data is processed using the chosen analysis method. Visualization is divided into static, dynamic or interactive. Static visualization is used for analysis results which are stored in the serving database and only want to display the results. Dynamic visualization is used for results that are updated regularly. While interactive visualization is used for interaction with users. Synchronization needs to be done so that each stakeholder can access processed maritime information.

The data used can be obtained from AIS as shown in table 1. AIS data provides several data such as MMSI, time and date, latitude, longitude, speed over ground (SOG), course over ground (COG), heading (direction of the ship), vessel name, IMO, Call Sign (ship identity for radio communication/ship call sign) and vessel type.

TABLE 1
AIS DATA, 2020 [13]

MMSI	Base DateTime	LAT	LON	SOG	COG	Heading	Vessel Name	IMO	CallSign	Vessel Type
367149340	2020-01-01T00:00:00	2.996.476	-9.002.724	1.3	10.0	16.0	SYDNEE TAYLOR		WDD4807	31
367687520	2020-01-01T00:00:00	3.020.558	-9.103.578	10.7	124.9	130.0	CHIPPEWA		WDI3361	31
367368170	2020-01-01T00:00:03	4.753.785	-12.232.833	0.6	46.5	141.0	SONJA H		WDE5536	31
367007980	2020-01-01T00:00:05	3.795.154	-12.132.682	0.0	-49.6	511.0	ANGIE M BRUSCO	IMO 5111359	WDC3446	31
367538940	2020-01-01T00:00:05	3.000.258	-9.322.608	3.1	168.0	511.0	RITA ANN		WDG4670	31
367054790	2020-01-01T00:00:07	2.976.044	-9.510.476	0.1	-180.8	511.0	COLT		WDC6330	31
12345678	2020-01-01T00:00:06	3.005.201	-9.054.060	0.0	162.2	511.0	INGRAM TEST UNIT		WWW0000	32
367567020	2020-01-01T00:00:08	3.414.870	-11.920.176	0.0	0.0	511.0	LULAPIN	IMO 8997869	WDG7412	31
367389480	2020-01-01T00:00:11	3.069.094	-8.145.967	0.0	-157.8	511.0	MAVERICK		WDE7139	31
477628100	2020-01-01T00:00:01	3.680.096	-7.522.302	12.8	-128.7	283.0	NINGBO SEAL	IMO 9579066	VRJD3	70
367320010	2020-01-01T00:00:01	4.375.206	-12.423.319	0.9	31.8	511.0	GALWAY BAY	IMO 0000000	WB06953	30
367574180	2020-01-01T00:00:00	2.581.778	-8.012.354	0.0	180.0	511.0	SEAS THE MOMENT		WDG8117	37
367588940	2020-01-01T00:00:00	2.966.370	-9.102.016	0.0	-186.7	511.0	MISS ELIZABETH	IMO 7906851	WDG9553	31
367091120	2020-01-01T00:00:04	3.989.587	-7.519.673	0.0	140.6	511.0	ROBERT E MCALLISTER	IMO 9068574	WDC8669	31
367419340	2020-01-01T00:00:07	3.804.081	-12.213.640	0.0	28.2	248.0	DELTA CATHRYN	IMO 9562219	WDE9621	31
366811570	2020-01-01T00:00:06	3.065.723	-8.804.494	0.0	49.8	352.0	NATALIE H	IMO 7200245	WDA4637	31
367186840	2020-01-01T00:00:10	2.976.239	-9.170.316	5.3	112.7	511.0	KENNETH S SETTOON		WDD7467	31
538002845	2020-01-01T00:00:10	2.608.420	-7.948.273	13.2	189.0	187.0	YASA GOLDEN DARDANEL	IMO 9339985	V7ME9	80
367142710	2020-01-01T00:00:09	3.704.221	-8.918.230	0.0	-153.5	511.0	LEO G LUTGRING		WDD4290	31
367376370	2020-01-01T00:00:00	4.149.156	-8.171.738	0.0	0.0	511.0	WYOMING		WC7559	31
367505650	2020-01-	3.885.223	-9.011.288	0.0	-	310.0	BERNARD G		WDF9486	31

	01T00:00:01				132.6					
367578710	2020-01-01T00:00:02	4.225.984	-7.091.863	9.4	-78.2	351.0	FLYING CLOUD		WDG8551	60
338145626	2020-01-01T00:00:00	4.729.650	-12.242.348	0.0	-49.6	511.0	KNOT RETIRED			37
366999423	2020-01-01T00:00:02	4.051.055	-8.011.168	0.0	-119.1	124.0	EVANICK		AEPP	90
367528690	2020-01-01T00:00:02	5.390.575	-16.651.147	0.1	-175.8	15.0	ALASKAN LADY	IMO 7742358	WDG3692	30
367003870	2020-01-01T00:00:05	3.042.711	-9.119.550	0.0	-176.2	511.0	RYAN	IMO 8976592	WDF2431	31
366772780	2020-01-01T00:00:07	4.762.083	-12.251.473	0.0	-166.5	353.0	WSF WALLA WALLA	IMO 7233151	WYX2158	60
367581680	2020-01-01T00:00:00	3.074.046	-8.804.103	6.5	10.7	10.0	THE DELTA		WDG8848	31
369970628	2020-01-01T00:00:00	4.039.245	-9.137.593	0.0	-161.9	511.0	MOLINE		AEMN	
477214900	2020-01-01T00:00:06	4.813.155	-12.343.617	0.1	157.5	68.0	PTI HERCULES	IMO 9358307	VRBY7	80
367597230	2020-01-01T00:00:07	4.074.065	-7.401.242	0.1	-49.6	269.0	FDNY 343		WDH2404	90
366961530	2020-01-01T00:00:07	3.835.473	-9.035.883	8.9	199.2	210.0	JOE B WYATT		WDB8613	31
367647210	2020-01-01T00:00:07	3.042.585	-9.119.550	0.0	4.5	9.0	ROBERT J SHEA		WDH7341	31
367400420	2020-01-01T00:00:08	4.117.470	-7.318.495	0.0	-146.7	158.0	PT BARNUM	IMO 8654285	WCZ3919	60
367665340	2020-01-01T00:00:13	3.605.423	-8.967.829	0.0	-197.4	511.0	ANGELINE		WDH9170	31
367741530	2020-01-01T00:00:00	2.721.461	-8.251.281	0.0	71.9	511.0	RIVERDANCE		WDI8727	37
366919230	2020-01-01T00:00:06	3.010.341	-9.096.017	0.0	-119.4	263.0	LARRY TILLEY		WY2160	31
367787640	2020-01-01T00:00:14	2.991.861	-9.014.131	0.1	-194.9	511.0	SPIRIT		WDJ5436	31
367001830	2020-01-01T00:00:14	4.620.157	-12.338.729	0.0	0.0	238.0	TERRI L BRUSCO	IMO 7618155	WDC3034	31
367406570	2020-01-01T00:00:22	4.875.540	-12.250.435	0.9	7.1	511.0	YANKEE BOY		WDE8524	30
367540040	2020-01-01T00:00:23	4.763.022	-12.233.213	0.0	45.6	39.0	EXCELLENCE		WDG4781	60
367177660	2020-01-01T00:00:28	4.560.675	-12.267.812	3.1	93.4	511.0	TRIUMPH IV		WBE7851	60
367597330	2020-01-01T00:00:37	4.838.043	-12.251.305	0.0	-49.6	268.0	ADAGIO		WDH2414	37
373746000	2020-01-01T00:00:39	2.987.873	-8.994.732	0.0	132.5	121.0	NEW COMMANDER	IMO 9610652	3EWU5	70
538004188	2020-01-01T00:00:40	3.775.965	-12.234.710	0.0	43.0	153.0	NAVE POLARIS	IMO 9457749	V7VQ6	80
367678850	2020-01-01T00:00:40	3.390.843	-11.845.012	0.1	-196.0	69.0	LELA FRANCO	IMO 9747821	WDI2504	31
367473990	2020-01-01T00:00:01	5.681.015	-13.296.332	0.0	-105.7	223.0	THERESA MARIE	IMO 7613648	WDF6412	30
367641850	2020-01-01T00:00:00	3.424.217	-11.926.448	0.0	0.0	264.0	DON C		WDH6811	90
338997000	2020-01-01T00:00:04	3.683.865	-7.628.181	0.0	0.0	22.0	MCFARLAND	IMO 7739856	AEGB	90
338790000	2020-01-01T00:00:01	4.760.055	-12.233.746	0.0	-135.4	89.0	SALLY FOX	IMO 9776729	WDH7915	60

IV. RESULT AND DISCUSSION

By using coordinate points data which obtained from satellites, the data can be processed to mapping the marine traffic by determining several center points and calculating the closest value. The process is repeated several times until the centroid value does not change. After finding the final value, it found that the ship traffic grouping that useful as shipping information which is very much needed in the maritime sector, especially Indonesia, which is the world's maritime axis.

The last stage is the visualization of maritime information in the form of reporting or presenting maritime information after the data is processed using the selected analytical method. Visualization is divided into static, dynamic or interactive. Static visualization is used for analysis results that are stored in the presentation database and only want to display the results.

The data processed through machine learning, and visualized the position of ships through WebGIS which ships pass through the territory of Indonesia. Then it can be classified which areas are the most densely populated by ships on the islands in Indonesia.

V. CONCLUSION

The use of Big Data in Indonesia's maritime sector is very useful for Indonesia's maritime security, especially in supporting Indonesia's vision as a World Maritime Axis. With the existence of a maritime information system in Indonesian waters, each relevant stakeholder can access information to find out the state of Indonesian waters in real-time. This is especially useful in supporting Indonesia based on UNCLOS 1982 by providing information and ensuring the safety of Indonesian waters and contributing to the Indonesian economy. Big data in the maritime sector which processed by K-means algorithm can be useful as maritime information for marine users.

ACKNOWLEDGMENT

The authors would like to thank to Telkom Institute of Technology Purwokerto for supporting this research to Ideabox competition and Faculty of Informatics, Telkom Institute of Technology Purwokerto for supporting this research.

REFERENCES

[1] Yani, Yanyan & Montratama, Ian. (2018). "Indonesia Sebagai Poros Maritim Dunia: Suatu Tinjauan Geopolitik". Jurnal Pertahanan & Bela Negara. 5. 10.33172/jpbh.v5i2.356.

[2] Afriansyah, Arie. 2015. "Kewenangan Negara Pantai Dalam Mengelola Wilayah Laut". Jurnal Hukum dan Pembangunan Tahun ke-45 Oktober-Desember 2015.

[3] Hurwitz, J., Alan, N., Halper, F., dan Kaufman, M., 2013. Big Data For Dummies. [e-book] John Wiley & Sons, Inc.

[4] Bahga, Arshdeep and Madisetti, Vijay. 2019. "Big Data Science & Analytics: A Hands-On Approach".

[5] Cholissodin, I., Riyandani, E. 2016. "Analisis Big Data" Fakultas Ilmu Komputer, Universitas Brawijaya, Malang.

[6] Karthik Kambatla, Giorgos Kollias, Vipin Kumar, Ananth Grama, Trends in big data analytics, Journal of Parallel and Distributed Computing, Volume 74, Issue 7, 2014, Pages 2561-2573, ISSN 0743-7315, https://doi.org/10.1016/j.jpdc.2014.01.003.

[7] Datamation. 2019. "Big Data 2019: Mining Data for Revenue". QuinStreet, Inc.

[8] Use of Big Data in The Maritime Industry. 2018. Trelleborg Marine Systems and supported, Port Technology.

[9] Toba, Hapnes. (2015). Big Data: Menuju Evolusi Era Informasi Selanjutnya.

[10] Salahudin, Salahudin. (2019). Mengenal Big Data (BD) dan Big Data Analysis (BDA).

[11] G. Wei, Z. Luo and X. Yu, "Clustering Research on Ship Fault Phenomena Based on K-means Algorithm," 2019 Chinese Control And Decision Conference (CCDC), Nanchang, China, 2019, pp. 4412-4415, doi: 10.1109/CCDC.2019.8832666.

[12] https://www.marinetraffic.com/en/ais/home/centerx:117.8/centery:-4.5/zoom:6

[13] https://coast.noaa.gov/htdata/CMSP/AISDataHandler/2020/index.html