Application of Data Mining Technique using K-Medoids in the case of Export of Crude Petroleum Materials to the Destination Country

Fathur Rahman¹, Ihda Innar Ridho¹, M. Muflih¹, Sefto Pratama¹, Mokhamad Ramdhani Raharjo¹, Agus Perdana Windarto²*

¹Universitas Islam Kalimantan Muhammad Arsyad Al Banjari Banjarmasin, Indonesia
²STIKOM Tunas Bangsa, Pematangsiantar, Indonesia

Email: *agus.perdana@amiktunasbangsa.ac.id

Abstract. The purpose of this research is to analyze and implement the data mining technique in the export of crude petroleum materials to the destination country. This is because Indonesia is a member of OPEC (Organization of the Petroleum Exporting Countries) which is one of the largest petroleum exporters in the world. It aims to obtain profits in the form of foreign exchange income obtained by the State. The data source was obtained from the Central Statistics Agency (BPS) with the website https://www.bps.go.id for data for 2017-2018. The calculation process The technique used is clustering with the K-Medoids algorithm. The calculation process is carried out using the help of Rapid Miner tools. The clusters used in this study are 2 namely: high cluster (C1) for export of crude oil materials and low cluster (C2) for crude oil materials. The results of the study stated that the high cluster (C1) consisted of 3 countries (Japan, Thailand and the United States) and the low cluster (C2) consisted of 6 countries (South Korea, Taiwan, China, Singapore, Malaysia and Singapore). It is hoped that the research results will be input and information for the government to rearrange policies in order to increase competitiveness, ensure business certainty and the sustainability of domestic industrial raw materials.

1. Introduction

The State of Indonesia is a member of OPEC (Organization of the Petroleum Exporting Countries) which is one of the largest petroleum exporters in the world. Indonesia is one of the exporting countries to developed and developing countries. The purpose of the exporter is to be able to make a profit. With this export, the government also receives income in the form of foreign exchange. The more exports, the greater the foreign exchange earned by the state. Goods exported by Indonesia consist of petroleum and natural gas (oil and gas) as well as non-oil / gas [1]. Petroleum is one of the commodities that are exported and imported in Indonesia because petroleum is one of the main energy that is widely used in almost every country. As a result of economic growth that continues to increase in other countries resulting in demand for oil continues to increase and make exporting activities run. Exports and imports play an important role in supporting a country's economy. In terms of oil demand, developed countries tend to run slowly even though the economy is going on. The main thing that causes the increase in demand, especially oil, is the increase in the population itself. Therefore it is necessary to review policies to increase competitiveness, ensure business certainty and the sustainability of domestic industrial raw materials specifically for the export of crude petroleum.
materials to destination countries. The policy can be in the form of a mapping of the destination country to export petroleum materials.

There are many computer science techniques that can solve these problems. The branches of computer science are Artificial Intelligence such as data mining [1], [2], Decision Support Systems (SPK) [3]–[6], expert systems [7], Artificial Neural Networks (ANN) [8]–[10], fuzzy logic [11] and others. In this research, data mining becomes a solution because data mining has an approach called clustering. Clustering The main concept emphasized is the search for cluster centers, where cluster centers are determined based on the minimum distance of each data in the cluster center. The data source uses data obtained from the Central Statistics Agency (https://www.bps.go.id). In this case the research focuses on the export of crude oil based on the destination country where the method used is K-Medoids. The purpose of the research is to apply K-Medoids in the export of crude petroleum materials to the destination country so that the results of the study can be input for the government of the country that has the highest priority in the export of crude oil based on the cluster that has been carried out.

2. Methodology

2.1 Data Mining

Data Mining is the process of discovering useful new correlations, patterns and trends by adding a large number of data repositories, using pattern recognition technologies such as static and mathematical techniques [2], [12].

2.2 Clustering Technique

Clustering is a process of grouping records, observations, which have the same object. Cluster technique consists of two methods: hierarchical clustering and non-hierarchical clustering [13], [14].

2.3 K-Medoids Method

This algorithm includes the Partitional Clustering algorithm and is one of the well-known methods. K-Medoids method is better than K-Means method in handling noise and outliers, because this method is not too influenced by outliers or other extreme values [13].

2.4 Research Method

In the application of data mining for the export of crude oil to the destination country, relevant data is needed about it. The data used in this study are the data of Crude Petroleum Exports to Destination Countries from 2017-2018 consisting of 9 countries namely Japan, South Korea, Taiwan, China, Thailand, Singapore, Malaysia, Australia and the United States taken from Central Statistics Agency with official website https://www.bps.go.id. then the data will be processed using the k-medoids method by taking the values of each attribute in the data to classify the crude oil export data.

3. Results and Discussion

In this clustering, the data that has been collected will be calculated in advance based on the data of crude oil export materials in the nine destination countries shown in the following table.

| Table 1. Preliminary data on exports of Crude Petroleum Materials |
|---|---|---|
| Country of destination | 2017 | 2018 |
| Japan | 1582,2 | 1315,4 |
| South Korea | 998,4 | 1001,7 |
| Taiwanese | 1082,0 | 908,0 |
| China | 1468,0 | 442,6 |
| Thailand | 2339,7 | 1819,9 |
| Singapore | 1536,0 | 959,8 |
| Malaysia | 1553,8 | 1148,2 |
| Australia | 1339,4 | 1099,4 |
| United States of America | 1 632,5 | 1 443,4 |
2.5 Centroid Data

In the application of the K-Medoids algorithm, the midpoint value can be determined randomly from the data obtained provided that the cluster created is 2, namely a high export cluster (C1) and a low export cluster (C2). Here are the initial centroids used:

| Table 2. Centroid Data |
|------------------------|
|                      |
| C1 (High Export Cluster) | C2 (Low Export Cluster) |
| United States of America | 1632,5                | 1443,3                |
| South Korea            | 998,4                 | 1001,7                |

Here are the results of the first iteration calculation as shown in the following table:

| Table 3. First Iteration Data |
|-------------------------------|
| Country of destination | 2017 | 2018 | C1  | C2 | Euclidian Distance |
| Japan                     | 1582,2| 1315,4| 16434,3| 8991,49| 16434,3 |
| South Korea               | 998,4| 1001,7| 195733 | 0     | 0     |
| Taiwanese                 | 1082,0| 908,0| 287203,7| 8863,29| 8863,29 |
| China                     | 1468,0| 442,6| 1001765 | 313062,41| 313062,41 |
| Thailand                  | 2339,7| 1819,9| 142459,5| 670792,54| 142459,45 |
| Singapore                 | 1536,0| 959,8| 233965,5| 2293,21| 2293,21 |
| Malaysia                  | 1553,8| 1148,2| 87221,74| 22017,65| 22017,65 |
| Australia                 | 1339,4| 1099,4| 118629,1| 9886,29| 9886,29 |
| United States of America  | 1632,5| 1443,4| 0     | 195732,99| 0     |

Based on table 3 that the cluster process can be done by centroid data in table 2. The cluster process is done by taking the shortest distance from each data that has been processed. From the initial data obtained the grouping in the first iteration as shown in the following table.

| Table 4. Results of the First Iteration Grouping |
|-----------------------------------------------|
| Cluster 1 | Cluster 2 |
| Japan     | South Korea |
| Thailand  | Taiwanese   |
| United States of America | China |
|            | Singapore   |
|            | Malaysia    |
|            | Australia   |

The K-Medoids process stops if the deviation > 0, but if the deviation < 0 then exchange the object with the data cluster to form a new set of objects as the medoid. The process of calculating the deviation by adding up the total new distance value - the total old distance that contains the distance of each object in each cluster with the new medoid. With the number of deviations shown in the following table.

| Table 5. Total First Iteration Distance |
|----------------------------------------|
| C1        | C2 |
| 16434,3   | 0  |
| 142459,5  | 8863,29 |
| 0         | 313062,4 |
| 2293,21   | 22017,65 |
| 9886,29   |      |
| 158893,8  | 356122,9 |
| 515016,6  |      |
Following are the results of data processing in the second iteration as shown in the following table:

| Country of destination | 2017   | 2018   | C1     | C2     | Euclidian Distance |
|------------------------|--------|--------|--------|--------|--------------------|
| Japan                  | 1582.2 | 1315.4 | 255277.8 | 761894.04 | 255277.75         |
| South Korea            | 998.4  | 1001.7 | 670792.5 | 313062.41 | 313062.41         |
| Taiwanese              | 1082.0 | 908.0  | 832819.3 | 216983.16 | 216983.16         |
| China                  | 1468.0 | 442.6  | 1897827  | 0       | 0                 |
| Thailand               | 2339.7 | 1819.9 | 1897826.9 | 0       | 0                 |
| Singapore              | 1536.0 | 959.8  | 740575.7 | 267563.84 | 267563.84         |
| Malaysia               | 1553.8 | 1148.2 | 451966.8 | 497957.16 | 451966.79         |
| Australia              | 1339.4 | 1099.4 | 520120.6 | 431514.84 | 431514.84         |
| United States of America | 1632.5 | 1443.4 | 142459.5 | 1001765.14 | 142459.45        |

The same thing is done as table 4, grouping based on Euclidian Distance as shown in the following table:

| Table 7. Total Second Iteration Distance |
|-----------------------------------------|
| C1                | C2                |
|-------------------|-------------------|
| 255277.8          | 313062.4          |
| 0                 | 216983.2          |
| 142459.5          | 0                 |
| 267563.8          | 451966.8          |
| 431514.8          |                   |
| 397737.2          | 1681091           |
| 2078828.24        |                   |

From the sum of the new distance data - the old distance produces a deviation > 0, then the grouping process stops at the second iteration with the result of the intersection of 1563811.64.

| Table 8. Standard Deviation Results |
|-------------------------------------|
| Standard Deviation = New Distance - Old Distance |
| New Distance | Old Distance | Result           |
| 2078828.24   | 515016.6     | 1563811.64       |

2.6 Rapid Miner Implementation
Following is the data processing using k-medoids on Rapid Miner:

![Figure 1. Input Data](image-url)
By using k-medoids modeling like Figure 1, by initializing the number of clusters as much as 2 pieces, the results obtained with the cluster formed are 2, according to the definition of k value with the number of cluster_0 = 3 items and cluster_1 = 6 items.

The results of the analysis are then tested again using the existing Rapid Miner. By entering the analysis data into Rapid Miner. In Text View it will display the cluster model. Cluster models obtained from the results of testing of data using k-medoids clustering.

Based on Figure 4, the results obtained from manual calculations and implementation with Rapid Miner are the same. The results of the cluster model consist of cluster 0 (high exports) namely 3 items (Japan, Thailand, United States), cluster 1 (low exports) which are 6 items (South Korea, Taiwan, China, Singapore, Malaysia, Australia). From a total of 9 items.

4. Conclusion
Based on the results of the study concluded the following conclusions:
1. The grouping of Crude Petroleum Material Export data has been successfully carried out to the Destination Country using the K-medoids algorithm with 2 clusters, namely for C1 (high exports) totaling 3 data: Japan, Thailand, United States and C2 (low exports) totaling 6 data namely: South Korea, Taiwan, China, Singapore, Malaysia, Australia.
2. With the existence of Rapid Miner software in this study, the accuracy of the data will be good enough for problems that occur related to the export of crude petroleum materials used as samples.

References

[1] W. Katrina, H. J. Damanik, F. Parhusip, D. Hartama, A. P. Windarto, and A. Wanto, “C.45 Classification Rules Model for Determining Students Level of Understanding of the Subject,” J. Phys. Conf. Ser., vol. 1255, no. 012005, pp. 1–7, 2019, doi: 10.1088/1742-6596/1255/1/012005.

[2] A. P. Windarto et al., “Analysis of the K-Means Algorithm on Clean Water Customers Based on the Province,” J. Phys. Conf. Ser., vol. 1255, no. 1, 2019, doi: 10.1088/1742-6596/1255/1/012001.

[3] S. R. Ningsih, R. Wulansari, D. Hartama, A. P. Windarto, and A. Wanto, “Analysis of PROMETHEE II Method on Selection of Lecturer Community Service Grant Proposals,” J. Phys. Conf. Ser., vol. 1255, no. 1, pp. 1–7, 2019, doi: 10.1088/1742-6596/1255/1/012004.

[4] T. Imandasari, M. G. Sadewo, A. P. Windarto, A. Wanto, H. O. Lingga Wijaya, and R. Kurniawan, “Analysis of the Selection Factor of Online Transportation in the VIKOR Method in Pematangsiantar City,” J. Phys. Conf. Ser., vol. 1255, no. 012008, pp. 1–7, 2019, doi: 10.1088/1742-6596/1255/1/012008.

[5] K. Fatmawati et al., “Analysis of Promeethee II Method in the Selection of the Best Formula for Infants under Three Years,” J. Phys. Conf. Ser., vol. 1255, no. 1, 2019, doi: 10.1088/1742-6596/1255/1/012009.

[6] D. R. Sari, N. Rofiqo, D. Hartama, A. P. Windarto, and A. Wanto, “Analysis of the Factors Causing Lazy Students to Study Using the ELECTRE II Algorithm,” J. Phys. Conf. Ser., vol. 1255, no. 1, 2019, doi: 10.1088/1742-6596/1255/1/012007.

[7] Y. Malhotra, “Expert systems for knowledge management: crossing the chasm between information processing and sense making,” Expert Syst. Appl., vol. 20, no. 1, pp. 7–16, 2001, doi: 10.1016/S0957-4174(00)00045-2.

[8] Sumijan, A. P. Windarto, A. Muhammad, and Budiharjo, “Implementation of Neural Networks in Predicting the Understanding Level of Students Subject,” Int. J. Softw. Eng. Its Appl., vol. 10, no. 10, pp. 189–204, 2016.

[9] Budiharjo, T. Soemartono, A. P. Windarto, and T. Herawan, “Predicting tuition fee payment problem using backpropagation neural network model,” Int. J. Adv. Sci. Technol., vol. 120, pp. 85–96, 2018, doi: 10.14257/ijast.2018.120.07.

[10] Budiharjo, T. Soemartono, A. P. Windarto, and T. Herawan, “Predicting School Participation in Indonesia using Back-Propagation Algorithm Model,” Int. J. Control Autom., vol. 11, no. 11, pp. 57–68, 2018.

[11] A. Mulyanto and Abdul Haris, “Penerapan Metode Fuzzy Tsukamoto Untuk Menentukan Jumlah Jam Overtime Pada Produksi Barang di PT Asahi Best Base Indonesia (ABBI) Bekasi,” J. Inform. SIMANTIK Vol.1, vol. 1, no. 1, pp. 1–11, 2016.

[12] D. Hartama, A. Perdana Windarto, and A. Wanto, “The Application of Data Mining in Determining Patterns of Interest of High School Graduates,” J. Phys. Conf. Ser., vol. 1339, no. 1, 2019, doi: 10.1088/1742-6596/1339/1/012042.

[13] B. Supriyadi, A. P. Windarto, T. Soemartono, and Mungad, “Classification of natural disaster prone areas in Indonesia using K-means,” Int. J. Grid Distrib. Comput., vol. 11, no. 8, pp. 87–98, 2018.

[14] A. P. Windarto, “Implementation of Data Mining on Rice Imports by Major Country of Origin Using Algorithm Using K-Means Clustering Method,” Int. J. Artif. Intell. Res., vol. 1, no. 2, pp. 26–33, 2017.