K-Means Algorithm for Clustering Third-party Funds of Conventional Banking

S Ningsih 1 and D Syahputra 1

1 Sekolah Tinggi Ilmu Ekonomi Muhammadiyah Asahan, Kisaran, Indonesia

*Email: supiahningsihSTIEMA@gmail.com

Abstract. Banking business activities in Indonesia are very important for Indonesia's economic growth, one of which is by growing the investment sector through obtaining funds from the general public via third party funds. In this study, the third party funds discussed were third party funds from conventional banking. The purpose of this study is to group the third party funds of Conventional Banking to obtain information on the development of Indonesia's economic growth, especially in conventional banking using the K-means. Research data in the form of conventional banking third party funds were obtained from the Financial Services Authority. The data are divided into 2 clusters: C1 (high) and C2 (low). Results show that 5 years have a low cluster (C1) and 6 years have a high cluster (C2). The findings of the analysis come in the form of information on the grouping of national banking third party funds to provide input to the Government of Indonesia, the Financial Services Authority, and Bank Indonesia in making policies on national banking.

1. Introduction
Banking is everything that concerns a bank, including institutions, business activities, and methods and processes for carrying out its business activities. The essential heart of the economy in Indonesia is banking. Stable can see an economy in a country from banking development in that country [1]. Banking in Indonesia applies a dual banking system, namely conventional banking that uses the interest system and sharia banking that applies the sharia system. Banking business activities in Indonesia are vital for Indonesia's economic growth, one of which is by growing the investment sector through raising funds collected from the general public by third-party funds. In this study, the third-party funds discussed were conventional banking assets from third parties. This type of profit from banking is obtained from traditional bank types that apply the interest system.

This research's relevant goal is to provide input to the Government of Indonesia, Financial Services Authority (FAS), and Bank Indonesia in making policies on conventional banking nationally in the form of data grouping on third party funds. So that with this input, it can further improve banking business activities nationally, which have an impact on the growth of the Indonesia's economy and the community's well-being—obtaining research data in the form of conventional banking third-party funds from the Financial Services Authority through the website https://www.ojk.go.id/. The grouping method used the K-Means Clustering because this method is still effectively used today [2], although there is already an X-Means method which is a development of this method [3]. Apart from clustering, data mining is also often used for data classification problems [4]–[8], which is almost similar to clustering.
Many previous studies related to this research, which discussed grouping using the K-Means algorithm, including research on the grouping of rice plants in Indonesia based on 34 provinces. This study's findings were presented in the form of data on rice plants divided into three clusters, namely the high cluster consisting of 3 provinces, the standard cluster consisting of 23 provinces, and the low cluster consisting of 8 provinces [9]. Subsequent research was carried out to group disaster-prone areas based on provinces in Indonesia. This study's results are data grouping of disaster-prone areas divided into 3 clusters, namely the high cluster consisting of 4 provinces, the standard cluster consisting of 14 provinces, and the low cluster consisting of 16 provinces [10]. Next, the research was conducted to group population density, human development index, open unemployment rate, and school enrollment rates in Indonesia by province. These research results are cluster 1 consisting of 12 provinces, cluster 2 consisting of 6 provinces and cluster 3 consisting of 1 province, cluster 4 consisting of 6 provinces, and cluster 5 consisting of 9 provinces [11]. These related studies are the background of the research to classify this third-party fund for conventional Banking. The results of this research are in the form of information on the grouping of national banking third party funds, which are expected to provide input to the Government of Indonesia, FAS, and Bank Indonesia in making policies on national banking so that this input can further improve national banking business activities which have an impact on the growth of the Indonesian economy and improving people's welfare.

2. Methodology

2.1. Research Methods and Data
The Financial Services Authority of the Republic of Indonesia provided research data in conventional banking third-party funds [12]. K-Means Clustering approach is used in this analysis. K-Means is a clustering algorithm that belongs to the Unsupervised learning category and is used to separate data into multiple classes using a partition scheme. This algorithm accepts data without any class labels as input [13].

2.2. Flowchart
The research flowchart using the K-Means algorithm can be seen in Figure 1.

![Figure 1. Research Flowchart](image-url)
Steps to the cluster using the k-means algorithm are as follows [14].

a. Determine how many clusters (k) there are in the data set.
b. Determine the importance of the center (Centroid).
c. Calculate the closest distance to Centroid for each record.
d. Objects are grouped according to their distance from the nearest Centroid.
e. Iterate from step a to step b until Centroid is maximal.

3. Results and Discussion

3.1. Research data

| Year | 1st Quarter | 2nd Quarter | 3rd Quarter | 4th Quarter |
|------|-------------|-------------|-------------|-------------|
| 2009 | 1,807,947   | 1,846,522   | 1,881,152   | 1,998,594   |
| 2010 | 2,000,719   | 2,115,333   | 2,164,363   | 2,360,279   |
| 2011 | 2,374,001   | 2,461,487   | 2,569,811   | 2,811,086   |
| 2012 | 2,390,771   | 2,478,560   | 2,587,298   | 2,829,782   |
| 2013 | 3,288,625   | 3,420,207   | 3,574,083   | 3,714,488   |
| 2014 | 3,669,863   | 3,886,623   | 4,050,408   | 4,173,170   |
| 2015 | 4,259,117   | 4,381,299   | 4,528,161   | 4,480,322   |
| 2016 | 4,538,309   | 4,644,909   | 4,677,335   | 4,912,483   |
| 2017 | 4,993,877   | 5,123,840   | 5,224,488   | 5,374,070   |
| 2018 | 5,379,459   | 5,485,117   | 5,571,237   | 5,722,404   |
| 2019 | 5,672,886   | 5,799,494   | 5,891,918   | 5,941,722   |

Source: Financial Services Authority, Republic of Indonesia [12]

Using the information in Table 1 and Figure 2, it can be seen that in the last 11 years, conventional banking third-party funds in Indonesia have continued to increase every year. The results of the author's observations for 40 quarters from the 1st quarter of 2019 to the 4th quarter of 2019 can be seen that the lowest third party funds from conventional banking were 1,807,947 billion rupiah in the 1st quarter of 2009 and the highest third party funds from conventional banking were 5,941,722 billion rupiahs in the fourth quarter of 2019.
3.2. Centroid Data
In using K-means, the midpoint or centroid data is obtained. The search process for the midpoint value is carried out by taking the largest (maximum) value for the high-level cluster (C1) and the smallest (minimum) value for the low-level cluster (C2).

| Data Cluster   | 1\textsuperscript{st} Quarter | 2\textsuperscript{nd} Quarter | 3\textsuperscript{rd} Quarter | 4\textsuperscript{th} Quarter |
|----------------|------------------------------|------------------------------|-------------------------------|------------------------------|
| (C1) High Cluster | 5.379.459                    | 5.485.117                    | 5.571.237                    | 5.722.404                    |
| (C2) Low Cluster    | 1.807.947                    | 1.846.522                    | 1.881.152                    | 1.998.594                    |

3.3. Clustering Data
Based on the data in Table 1, data will be grouped into Third Party Funds for Conventional Banking with Rapid Miner. See Figure 3 for the grouping process using the K-Means using Rapid Miner.

Figure 3 depicts the process of linking read Excel with K-Means, as well as the output that will be used with a K = 2 value. The object model contains all of the requisite parameters. In this case, the researcher used a sample of data from Conventional Banking Third Party Funds from 2009 to 2019 based on a quarter (Based on Table 1) with 2 attributes, namely: 1\textsuperscript{st} Quarter, 2\textsuperscript{nd} Quarter, 3\textsuperscript{rd} Quarter and 4\textsuperscript{th} Quarter. The final result of data grouping is seen in Figure 4.

Figure 4 shows that there are two clusters, beginning with cluster 0 (low cluster) and ending with cluster 1 (high cluster). Clusters 0 and 1 each have five and six items, respectively. As shown in Figure 5, we get a graph plot view from Rapid Miner checking.
While the details of the grouping of Conventional Banking Third Party Funds is seen in Figure 6.

![Figure 6. Details of cluster Conventional Banking Third party funds](image)

Based on figure 6, it can be explained that Conventional Banking Third Party Funds are included in cluster 0 (low cluster), namely 2009 to 2013. Meanwhile, Conventional Banking Third Party Funds that are included in cluster 1 (high cluster) are 2014 to 2019.

4. Conclusion

Data mining can be done with data mining. Based on 4 assessment attributes (1st Quarter, 2nd Quarter, 3rd Quarter, and 4th Quarter), the cluster uses 2 clusters, namely: (C1) high cluster and (C2) low cluster. From the clustering results, it can be seen that conventional banking third party funds continue to increase. Information on the grouping of national banking third party funds is expected to provide input to the Government of Indonesia, the Financial Services Authority and Bank Indonesia in making policies on national banking so that it has an impact on the growth of the Indonesian economy and improves the welfare of the community.

References

[1] G. Gopinath, J. C. Stein, S. Avdjiev, L. Gambacorta, and S.-K. Pradhan, “Nber Working Paper Series Banking, Trade, and the Making of a Dominant Currency,” NBER Working Paper Series, vol. 24485, 2018.

[2] J. Hutagalung, N. L. W. S. R. Ginantra, G. W. Bhawika, W. G. S. Parwita, A. Wanto, and P. D. Panjaitan, “COVID-19 Cases and Deaths in Southeast Asia Clustering using K-Means Algorithm,” Journal of Physics: Conference Series, vol. 1783, no. 1, p. 012027, 2021.

[3] N. Arminarahmah, A. D. GS, G. W. Bhawika, M. P. Dewi, and A. Wanto, “Mapping the Spread of Covid-19 in Asia Using Data Mining X-Means Algorithms,” IOP Conf. Series: Materials Science and Engineering, vol. 1071, no. 012018, pp. 1–7, 2021.

[4] I. S. Damanik, A. P. Windarto, A. Wanto, Poningsih, S. R. Andani, and W. Saputra, “Decision Tree Optimization in C4.5 Algorithm Using Genetic Algorithm,” Journal of Physics: Conference Series, vol. 1255, no. 1, pp. 1–6, Aug. 2019.

[5] 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,” Journal of Physics: Conference Series, vol. 1255, no. 1, pp. 1–7, 2019.

[6] H. Siahaan, H. Mawengkang, S. Efendi, A. Wanto, and A. Perdana Windarto, “Application of Classification Method C4.5 on Selection of Exemplary Teachers,” in Journal of Physics: Conference Series, 2019, vol. 1235, no. 1.
[7] I. Parlina et al., “Naive Bayes Algorithm Analysis to Determine the Percentage Level of visitors the Most Dominant Zoo Visit by Age Category,” Journal of Physics: Conference Series, vol. 1255, no. 1, pp. 1–5, 2019.

[8] D. Hartama, A. Perdana Windarto, and A. Wanto, “The Application of Data Mining in Determining Patterns of Interest of High School Graduates,” Journal of Physics: Conference Series, vol. 1339, no. 1, pp. 1–6, 2019.

[9] Sudirman, A. P. Windarto, and A. Wanto, “Data mining tools | rapidminer: K-means method on clustering of rice crops by province as efforts to stabilize food crops in Indonesia,” IOP Conference Series: Materials Science and Engineering, vol. 420, no. 1, pp. 1–8, 2018.

[10] B. Supriyadi, A. P. Windarto, T. Soemartono, and Mungad, “Classification of Natural Disaster Prone Areas in Indonesia using K-Means,” International Journal of Grid and Distributed Computing, vol. 11, no. 8, pp. 87–98, 2018.

[11] A. S. Ahmar, D. Napitupulu, R. Rahim, R. Hidayat, Y. Sonatha, and M. Azmi, “Using K-Means Clustering to Cluster Provinces in Indonesia,” Journal of Physics: Conference Series, vol. 1028, no. 1, pp. 1–6, 2018.

[12] O. J. Keuangan, “Indonesia Banking Statistics,” Financial Services Authority, Republic of Indonesia, vol. 18, no. 1, pp. 1–183, 2019.

[13] A. Wanto et al., Data Mining : Algoritma dan Implementasi. Yayasan Kita Menulis, 2020.

[14] J. Han and M. Kamber, Data Mining : Concepts and Techniques Second Edition. San Francisco: Elsevier, 2006.