

SALES TRANSACTION DATA EVALUATION INFORMATION SYSTEM USING CLUSTERING MODEL

Eko Purwanto^{1*}, Farahwahida binti Mohd², Zalizah Awong Long³, Singgih Purnomo⁴
Duta Bangsa University Indonesia¹, University Kuala Lumpur Malaysia², University Kuala Lumpur
Mayasia³, University Duta Bangsa Malaysia
*Correspondence Email : eko_purwanto@udb.ac.id

ABSTRACT

Sales evaluation exercises at instruments or trading enterprises with many outcomes and components are critical to determining the company's business development. Enterprise managers must know and evaluate sales data through reports on all branches. A trading business that has a department store, so it is essential to assess the sales trade. This sales evaluation knowledge system is necessary to promote store leadership in the deals evaluation strategy, which aims to aid the administration's decision-making process. The dealing data is evaluated using a clustering model to find a group of products selling well, especially in demand and not selling well. The algorithm used in clustering is by using K-Means. The sample data used is sales transaction data of 122 sales transaction data. Based on the results of the Davies Bouldin, it was found that the number of clusters with the best performance was 4 clusters with a Davies Blouldin value of -0.371. This analysis results in the sales data evaluation information system using the clustering model as an instrument for store managers to evaluate deals data at all department stores. The result of this method can be caused by the supervision as data in the decision-making method.

KEYWORDS

clustering, evaluation, K-Means, model, sales, transaction



This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International

INTRODUCTION

Sales are buying and trading interests or benefits supported by two or more parties utilizing a good price tool. Most lower- and upper-class people have often done sales in Indonesia and completed buying and selling transactions.

Sales require monitoring or monitoring in order to show the success achieved from the goals of the organization or agency. The leadership usually carries out monitoring to ensure

the sales results have reached the targets that the performance performed by the employees must achieve. Through monitoring outcomes, the manager can evaluate the sales in each branch.

Regarding sales data on an enterprise engaged in the trade sector with many products and having branches scattered in various regions, it is complex and requires special handling. The company carries out the sales evaluation method to choose the level of enterprise growth of the business.

The clustering model was built using the K-Means method and the Davies-Bouldin Index to evaluate cluster performance (Sari & Purwadinata, 2019).

Customer grouping using the K-Means algorithm is based on transactions, which aims to determine the buying behavior made by customers (Mulyawan et al., 2019).

The K-means clustering algorithm is utilized in the knowledge classification by testing certain cross-validations and dividing the dataset into a few percent for training and the rest for testing, a technique by arranging the group files created by the group model as the conditional variable and the remaining variables as the independent variable (Mohammed & Kassie, 2018).

Classify high profit, high standard, low risk, high focus, and low maintenance customers for supermarkets using hierarchical clustering. Grouping customers is useful when combining business strategies and making important decisions (Rokaha et al., 2019).

Clustering techniques are used to realize customer classification by building commercial databases, getting customer grouping vectors with frequent sets of items, and calculating customer inequality matrices (Li & Chen, 2018).

The k-means segmentation algorithm and C4.5 classification are used for segmenting potential customers based on the Recency Frequency Monetary (RFM) model to increase the percentage of accuracy in customer loyalty classification research, which has an accuracy of 79.33% and Area Under Curve (AUC) 0.831 (Moedjiono et al., 2017).

The customer segmentation model using K-Means grouping can produce a pattern of grouping customer income against their expenses. A cluster is a group of customers with similar characteristics (Alamsyah & Nurriz, 2017).

Customer classification founded on outliers data of customer assets is created as a customer classification model founded on outliers data study of customer assets. Model-based on variables in 4 dimensions, including frequency of transactions, type of product or service traded, number of trades, and age of clients (Zuo & Guo, 2019).

The k-means algorithm is utilized to create classifiers and a two-dimensional grouping segmentation method from the present and potential values presented. Two indicators, comparative size and degree of change in cluster roughness result in differences in segmentation in different periods combined with changing cluster membership of individual customers (Hu & Zhao, 2015).

The K-means clustering algorithm is utilized in grouping deals using e-commerce. The standards used in the analysis are divided into 3 clusters: seller, reseller, and dropship (Aria, 2020).

RESEARCH METHOD

Several steps in the sales evaluation study utilize this clustering model: data cleaning, data integration, data transformation, data mining, and pattern evaluation. Data cleaning is done manually, and the sample data obtained in hard copy is then entered into Microsoft Excel format (.xls) to be imported into the application. The sample data will be selected according to the needs in this study. Combining data in this study was also accomplished

manually. The sample data obtained are combined in one file in Microsoft Excel format (.xls), which will later be used for the data import.

Data transformation has been carried out at this stage, namely at the data cleaning and integration stage. In this study, a process in which clustering is sought, i.e., the calculation method based on K-Means, will create output in the form of groups of goods based on their selling rate. The clustering results will be obtained from the data mining process at this stage.

RESULT AND DISCUSSION

Sample data in the calculation: The author uses only some transaction data as a sample of the design of this system; the sample data is taken from a case study of 122 transaction data.

Table 1. Sample Sales Transaction Data

No.	Product name	Price	Total
1	Bergo Dian Sifon Tumpuk Blik	53000	2
2	Hijab Pashmina Instan Zazkia	64000	5
3	Hijab Instan Pashmina Orchard	45000	6
etc
122	Kerudung Instan Nurjanah	35000	1

The evaluation strategy is performed by segmenting or grouping the Sales of Goods data accessed from the database using a K-Means algorithm clustering process. The author uses the Rapid Miner software to test by utilizing the sales data acquired.

To determine the best number of clusters, the authors look for the performance of the number of clusters by determining the Bouldin Davies; here are the results of the Bouldin Davies from several numbers of clusters with experiments using Rapid Miner Number of clusters 2 with bouldin davies: -0.455, Number of clusters 3 with bouldin davies: -0.478, Number of clusters 4 with bouldin davies: -0.371 and Number of clusters 5 with bouldin davies: -0.447. Based on the results of the Davies Bouldin, it was found that the number of clusters with the best performance was 4 clusters with a Davies Blouldin value of -0.371.

Table 2. Centroid Table

Attribute	Cluster 0	Cluster 1	Cluster 2	Cluster 3
Harga	134615.385	62642.857	87478.261	40227.273
Jumlah	2.462	4.500	3.174	3.886

The cluster or segmentation process with the number of clusters 4 and using the K-Means algorithm. The software used is Rapid Miner, showing the following results: Cluster 0: 13 data, Cluster 1: 42 data, Cluster 2: 23 data, and Cluster 3: 44 data.

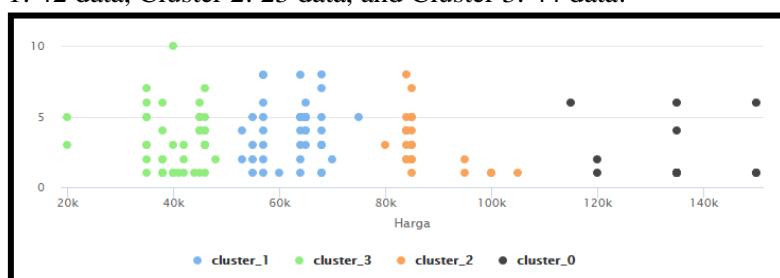


Figure 1. Cluster Model

The sales data evaluation process is displayed as sales data graphs. Sales evaluation as a reference for leaders in the decision-making process. The results of the sales evaluation are shown in Figure 2 and Figure 3.

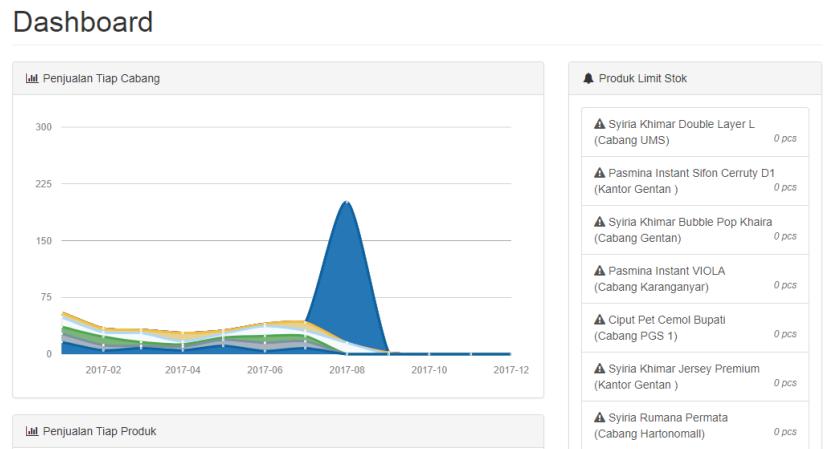


Figure 2. Sales Evaluation Graph by branch

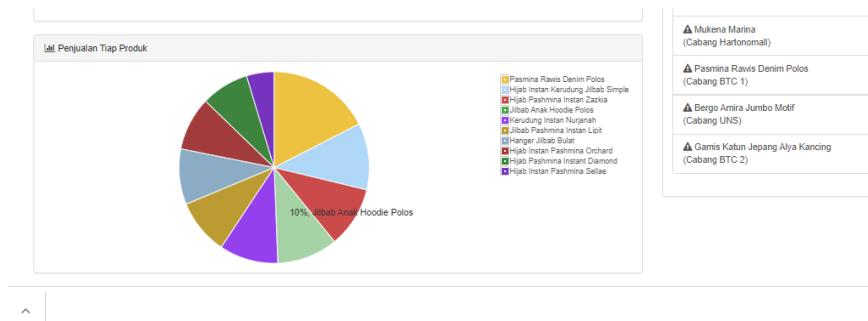


Figure 3. Sales Evaluation Graph by Product

This sales evaluation knowledge method is needed to enable the store leadership in the sales evaluation strategy, which strives to aid the supervision's decision-making process. The evaluation of sales data is carried out utilizing a clustering model to find a group of products that are selling well, are quite in demand, and need to sell better. The algorithm used in clustering is by using K-Means.

CONCLUSION

This sales evaluation knowledge method is needed to enable the store leadership in the sales evaluation strategy, which aspires to aid the supervision's decision-making process. The evaluation of sales data is carried out using a clustering model to find a group of products that are selling well, are quite in demand, and are not selling well. The algorithm used in clustering is by using K-Means. The sample data used is sales transaction data of 122 sales transaction data. Based on the results of the Davies Bouldin, it was found that the number of clusters with the best performance was 4 clusters with a Davies Blouldin value of -0.371. This research results in the sales data evaluation information system utilizing the clustering model as an instrument for shop leaders to evaluate sales data at all branch stores.

The result of this method can be made by the leadership as data in the decision-making process.

REFERENCES

Alamsyah, A., & Nurriz, B. (2017). Monte Carlo simulation and clustering for customer segmentation in business organization. *Proceeding - 2017 3rd International Conference on Science and Technology-Computer, ICST 2017*, 104–109. <https://doi.org/10.1109/ICSTC.2017.8011861>

Aria, R. R. (2020). K-Means to Determine the e-commerce Sales Model in Indonesia. *International Journal of Information System & Technology Akreditasi*, 3(36), 166–172.

Hu, X., & Zhao, S. (2015). A dynamic segmentation method of power customer based on rough clustering. *Chinese Control Conference, CCC, 2015-September*, 8773–8778. <https://doi.org/10.1109/ChiCC.2015.7261026>

Li, B., & Chen, Y. C. (2018). The Application of Clustering Mining Technology in E-Commerce Website. *Proceedings - International Symposium on Parallel Architectures, Algorithms and Programming, PAAP, 2018-December*, 104–107. <https://doi.org/10.1109/PAAP.2018.00025>

Moedjiono, S., Isak, Y. R., & Kusdaryono, A. (2017). Customer loyalty prediction in multimedia Service Provider Company with K-Means segmentation and C4.5 algorithm. *2016 International Conference on Informatics and Computing, ICIC 2016, Icic*, 210–215. <https://doi.org/10.1109/IAC.2016.7905717>

Mohammed, M., & Kassie, A. (2018). Transaction. *2018 2nd International Conference on Trends in Electronics and Informatics (ICOEI), Icoei*, 115–120.

Mulyawan, B., Viny Christanti, M., & Wenas, R. (2019). Recommendation Product Based on Customer Categorization with K-Means Clustering Method. *IOP Conference Series: Materials Science and Engineering*, 508(1). <https://doi.org/10.1088/1757-899X/508/1/012123>

Rokaha, B., Ghale, D. P., & Gautam, B. P. (2019). Enhancement of supermarket business and market plan by using hierarchical clustering and association mining technique. *Proceedings - 2018 International Conference on Networking and Network Applications, NaNA 2018*, 1, 384–389. <https://doi.org/10.1109/NANA.2018.8648716>

Sari, P. K., & Purwadinata, A. (2019). Analysis Characteristics of Car Sales In E-Commerce Data Using Clustering Model. *Journal of Data Science and Its Applications*, 2(1), 68–77. <https://doi.org/10.21108/jdsa.2019.2.19>

Zuo, L., & Guo, J. (2019). Customer Classification of Discrete Data Concerning Customer Assets Based on Data Mining. *Proceedings - 2019 International Conference on Intelligent Transportation, Big Data and Smart City, ICITBS 2019*, 352–355. <https://doi.org/10.1109/ICITBS.2019.00093>