



Segmentation of Credit Card Customers Based on Their Credit Card Usage Behavior using The K-Means Algorithm

Ichwanul Muslim Karo Karo^{*}, Andreyas Yasmanto², Robi Setiawan²

¹Department of Informatics, Universitas Telkom, Indonesia

²Department Life Sciences, Universitas Surya, Indonesia

*Correspondence: E-mail: ichwanulkarokaro@telkomuniversity.ac.id

ABSTRACT

The intensity of credit card customers in making transactions has increased in the last 10 years in Indonesia. This is both a challenge and an opportunity for the Bank. Customer segmentation information is beneficial to reduce bad debts or increasing customer credit card limit capacity. This study aims to segment credit card customers based on their usage behavior with a clustering approach using the K-means algorithm. While the process of evaluating segmentation results using the silhouette index. Based on the experimental results, six groups are the best number of clusters. The six groups are shopping hobbies, payment process at maturity, payment by installments, withdrawing cash, buying expensive goods, and types that rarely use credit cards.

© 2021 Universitas Pendidikan Indonesia

ARTICLE INFO

Article History:

Submitted/Received 13 Oct 2021

First Revised 30 Oct 2021

Accepted 17 Nov 2021

First Available Online 27 Nov 2021

Publication Date 01 Dec 2021

Keyword:

K-Means,

Silhouette index.

1. INTRODUCTION

The era of digital financial development is rapidly linear to the lifestyle of some people in accessing a relatively expensive product. Ease of payment is offered in the form of credit. So that usually, the traders provide choices in payment by installments each period. One payment method in installments is using a credit card (Kumar and Karlina, 2020). The use of credit cards in Indonesia has increased in the last 10 years and it was found that many banks began to be cautious in offering credit card limits due to the inability of prospective debtors to pay debts (Johan and Dewi, 2021).

Of the many factors that determine debtors to use credit cards, some factors that stand out include subjective norms, perceived behavioral control, and perceived benefits (Chien and Devaney, 2001). There are also impulse buying factors that influence credit card users (Cuandra and Kelvin, 2021). For students who use credit cards, there is a relationship between parental influence, financial knowledge, and the student's attitude toward using the credit card (Kashif et al., 2018). There is also a relationship that the number of cards owned negatively related to age and positively related to income level (Jung and Kang, 2021). On the other hand, it was found that the behavior of increasing compulsive shopping was comparable to people with high prestige and those with a low understanding of finance (Khandelwal et al., 2021; Palan et al., 2011).

Several previous studies underlie this research, among others:

1. The research uses the Random Forest algorithm to predict customer satisfaction with credit card services (Yaseen et al., 2020).
2. The research aims to identify qualified credit card customers and develop product quality by creating features that suit their needs to increase customer satisfaction. This study used the k-means and CandRT algorithms (Hassani and Taati, 2020).
3. The research aims to segment customers based on credit card usage behavior in Africa using k-means clustering (Umuhoza et al., 2020).
4. Research aimed at assisting bank management in assessing credit card clients using bidirectional LSTM neural networks by modeling and predicting consumer behavior concerning two aspects: the probability of single and consecutive missed payments for credit card customers (Ala'raj et al., 2021).

2. METHODS

The research has several stages, including data preparation, modeling and evaluation, cluster interpretation, and visualization of the cluster. This dataset is processed using google collab.

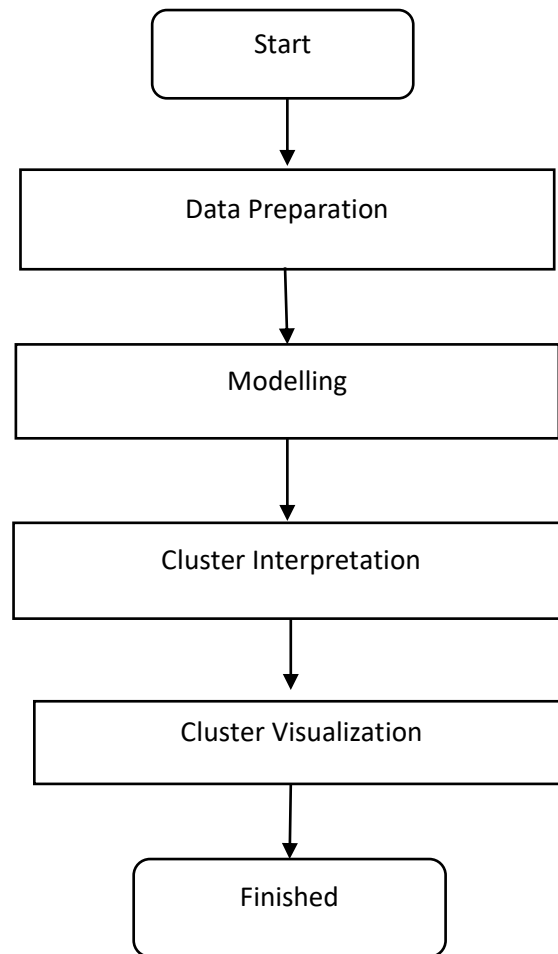


Figure 1. Research Framework.

2.1. Data

The data used in this study was created in 2018 and is available on the site (see <https://www.kaggle.com/arjunbhasin2013/ccdata>). This dataset comprises 8.950 rows (credit card owners) and 18 behavioral variables.

Table 1. Behavioral variables.

No.	Variable Name	Description
1	<i>Cust_id</i>	Identity of the credit card holder
2	<i>Balance</i>	The amount of balance left in their account to make a purchase.
3	<i>Balance Frequency</i>	The intensity with which the balance is updated is scored between 0 and 1 (1 = updated frequently, 0 = not updated frequently).
4	<i>Purchases</i>	The number of purchases made from the account.
5	<i>One Off Purchase</i>	The maximum purchase amount made in a single transaction.

6	<i>Installments Purchases</i>	The number of purchases made on credit.
7	<i>Cash Advance</i>	The user provides the initial cash.
8	<i>Purchases Frequency</i>	The intensity of purchases made, score between 0 and 1 (1 = frequently purchased, 0 = not frequently purchased).
9	<i>One-Off Purchases Frequency</i>	Purchase intensity occurs in one transaction (1 = frequently purchased, 0 = not frequently purchased).
10	<i>Purchases Installments Frequency</i>	The intensity of credit purchases made (1 = often done, 0 = not often done).
11	<i>Cash Advance Frequency</i>	Initial cash intensity paid.
12	<i>Cash Advance Trx</i>	Number of transactions made with cash in advance
13	<i>Purchases Trx</i>	The number of purchase transactions made.
14	<i>Credit Limit</i>	Credit card limit for users.
15	<i>Payments</i>	The amount of payment made by the user.
16	<i>Minimum Payments</i>	The minimum amount of payment made by the user.
17	<i>Prc Full Payment</i>	The percentage of full payment paid by the user.
18	<i>Tenure</i>	Credit card service period for users.

2.2. Data Preparation Phase

The data preparation stage is essential in segmentation (Ziafat and Shakeri, 2016). One of the goals is to standardize the dataset. In this research, there are several data preparation processes carried out.

Missing Value: one of the problems faced in the dataset is the missing data component or missing value. This dataset has a lot of missing data, so special handling is needed for this problem. The most common method to replace missing values is to replace the value with the average value of the attribute (Yadav and Roychoudhury, 2018).

There are many outlier values in the dataset. If the outlier is removed, it will cause the loss of so many data records that the model performance will not be maximized (Ala'raj et al., 2021). So, to overcome this, a range is created to handle extreme values.

2.3. Segmentation Phase

The credit card customer segmentation process will use a Clustering approach. There are many types of clustering algorithms. However, Partitioning Clustering is the best clustering algorithm to determine k clusters (Yadav and Roychoudhury, 2018; Naeem and Wumaier, 2018). One of the well-known partitioning clustering algorithms is the K-means algorithm.

The K-means algorithm is a non-hierarchical cluster analysis method that partitions objects into one or more groups based on the similarity of their characteristics (Muthahharah and Juhari, 2021) so that objects that have closer characteristics are grouped in the same cluster,

while objects that have different characteristics are grouped into other clusters (Khormarudin, 2016). In other words, the K-Means algorithm aims to minimize variation between data in a cluster and maximize variation with data in different clusters (Sulistiyawati and Supriyanto, 2021). The learning algorithm groups the data into k clusters without knowing the target class. This learning is included in unsupervised learning (Urva, 2016). Determining k is assisted using the elbow method (Bholowalia and Kumar, 2018). The following process is to evaluate the segmentation results. The evaluation process uses the Silhouette Index (Si). The stages follow (Gustriansyah et al., 2019).

Table 2. K-Means algorithm.

K-Means Algorithm	
Input	
Data	
Process	
1.	Determine k (number of clusters)
2.	Determine the starting point centroid as many as k
3.	Calculate the distance of each data to the centroid
4.	Put each data into the cluster with the minimum distance
5.	Update the centroid of each cluster
6.	Repeat Steps 3-5 until there is no change in the centroid
Output	
k and each cluster member	

3. RESULT AND DISCUSSION

This section presents and discusses the results of credit card customer segmentation, including determining the number of clusters, evaluating the number of clusters, and interpreting the best cluster.

3.1. Cluster determination

After the data preparation stage, we tested the best number of clusters for the K-means algorithm and obtained the results of the elbow method test with k = 6. This information is the initial reference in the clustering process. However, other k numbers will also be tested,

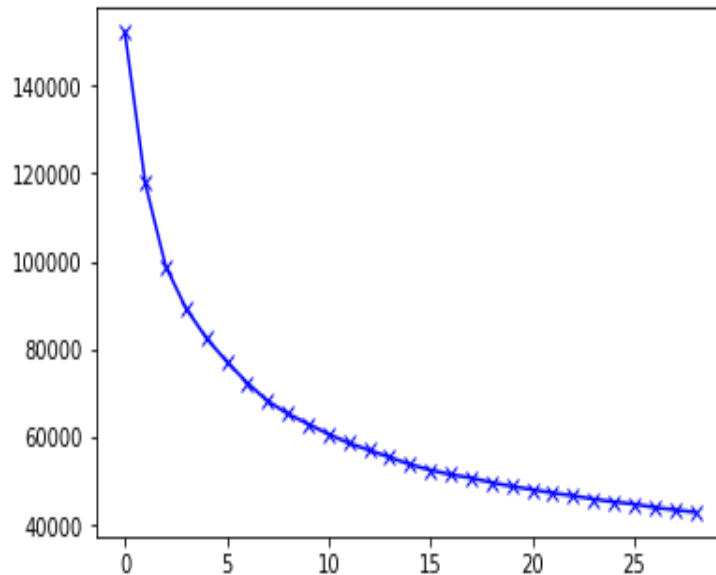


Figure 2. Elbow method results.

3.2. Evaluation of the number of clusters

This section will evaluate each number of clusters tested. The best cluster criteria are determined and interpreted based on the S_i value (**Table 2**). The number of k tested includes $k = 5, 6, 7, 8, 9, 10$.

Table 3. Interpretation of S_i value.

S_i	Description
$S_i > 0,71$	<i>The clusters obtained are powerful</i>
$0,51 < S_i < 0,71$	Data can be clustered
$0,25 < S_i < 0,5$	<i>The clusters obtained are weak</i>
$S_i < 0,25$	No clusters found

Based on the experimental results, the number of clusters with the highest S_i value is $k = 6$. In other words, the best number of clusters obtained is $k = 6$. This result also confirms and strengthens the interpretation results of the elbow method, which provides the same k recommendation. Based on the information in Table 2, the result of $k = 6$ has not produced strong clusters, the data can be clustered, but the cluster members are not coherent. However, if we continue to look for another number of clusters, then the quality of the resulting cluster is not necessarily as good as $k = 6$. The reason is that the $k > 6$ obtained S_i pattern tends to shrink.

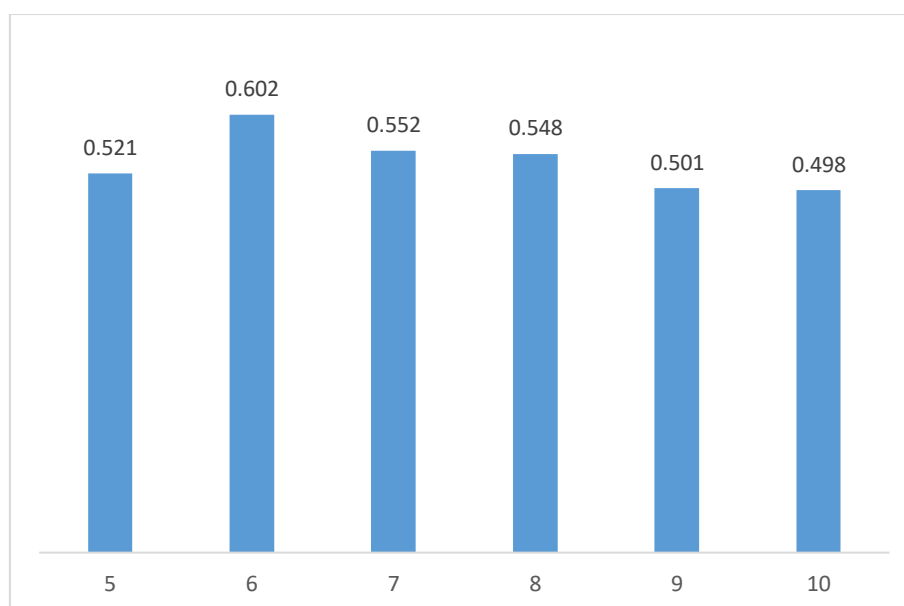


Figure 3. Elbow method results.

3.2. Best cluster interpretation

We use the Principal Component Analysis (PCA) method to visualize the cluster results to transform the data into 2 dimensions (Abdulhafedh, 2021). The visualization can be seen in **Figure 4**.

- Red clusters are credit card customers with behaviors that make all types of purchases. In other words, this type of customer likes to shop.
- Dark blue clusters are credit card customers with due payment behavior. Every credit card usage is paid according to the due date.
- Green cluster is a customer with behavior where every purchase is made in installments.
- Yellow clusters are customers who use credit cards to withdraw cash. In other words, credit card debt.
- Orange cluster is a customer who buys expensive and luxurious goods.
- Purple cluster is a customer with a behavior that does not spend much money. In other words, this type of customer rarely uses credit cards.

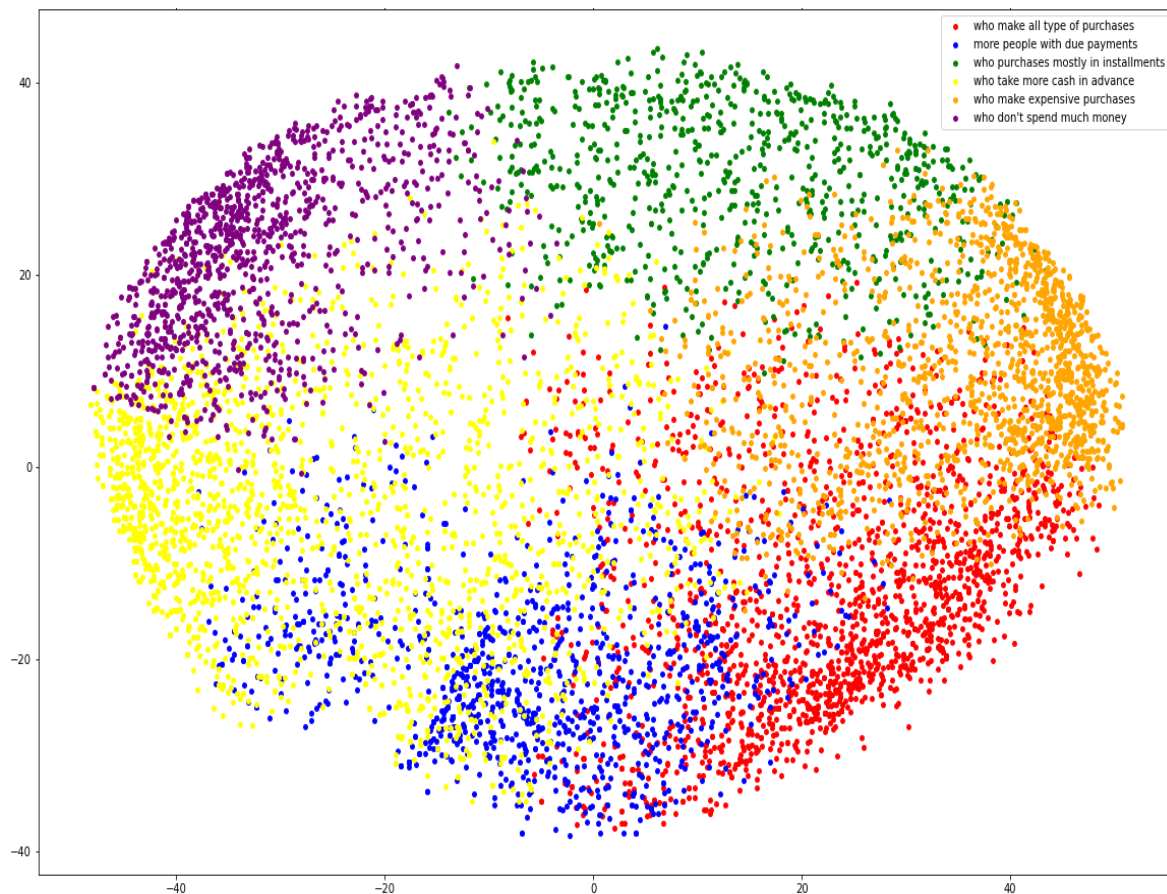


Figure 4. Visualization of customer segmentation.

4. CONCLUSION

Segmentation of credit card customers can be done using a clustering approach with the K-means algorithm. In this research, segmenting customers based on their credit card usage behavior is carried out with the K-means algorithm and the PCA method in the visualization process. The best number of clusters obtained is 6, with the interpretation of quality that is not bad. The six customer groups obtained are shopping hobbies, payment processes at maturity, payment by installments, withdrawing cash, buying expensive goods, and types that rarely use credit cards. In the future, it is possible to improve algorithms and handle many variables to obtain good cluster results.

5. AUTHOR NOTE

The authors declare that there is no conflict of interest regarding the publication of this article. The authors affirm that the paper is free from plagiarism.

6. REFERENCES

- Abdulhafedh, A. (2021). Incorporating K-means, hierarchical clustering and PCA in customer segmentation. *Journal of City and Development*, 3(1), 12-30.
- Ala'raj, M., Abbod, M. F., and Majdalawieh, M. (2021). Modelling customers credit card behaviour using bidirectional LSTM neural networks. *Journal of Big Data*, 8(1), 1-27.

- Bholowalia, P., and Kumar, A. (2014). EBK-means: A clustering technique based on elbow method and k-means in WSN. *International Journal of Computer Applications*, 105(9) 17-24.
- Chien, Y. W., and Devaney, S. A. (2001). The effects of credit attitude and socioeconomic factors on credit card and installment debt. *Journal of Consumer Affairs*, 35(1), 162-179.
- Cuandra, F., and Kelvin, K. (2021). Analysis of influence of materialism on impulsive buying and compulsive buying with credit card use as mediation variable. *Jurnal Manajemen*, 13(1), 7-16.
- Gustriansyah, R., Suhandi, N., and Antony, F. (2020). Clustering optimization in RFM analysis based on K-means. *Indonesian Journal of Electrical Engineering and Computer Science*, 18(1), 470-477.
- Hassani, L., and Taati, E. (2020). Studying product quality by exploring credit card customers behaviour via data mining techniques. *International Journal for Quality Research*, 14(1), 163-164.
- Johan, S., and Dewi, C. E. (2021). Credit limit of unsecured consumer lending: evidence from micro data. *Economics and Finance in Indonesia*, 67(1), 1-10.
- Jung, K., and Kang, M. Y. (2021). Understanding credit card usage behavior of elderly Korean consumers for sustainable growth: Implications for Korean credit card companies. *Sustainability*, 13(7), 3817-3823.
- Kashif, M., Zarkada, A., and Ramayah, T. (2018). The impact of attitude, subjective norms, and perceived behavioural control on managers' intentions to behave ethically. *Total Quality Management and Business Excellence*, 29(5-6), 481-501.
- Khandelwal, R., Kolte, A., Veer, N., and Sharma, P. (2022). Compulsive buying behaviour of credit card users and affecting factors such as financial knowledge, prestige and retention time: a cross-sectional research. *Vision*, 26(2), 172-180.
- Khormarudin, A. N. (2016). Teknik Data Mining: Algoritma K-Means Clustering. *J. Ilmu Komput*, 1-12.
- Kumar, S., and Karlina, L. (2020). Intention to use credit card among college students in Greater Jakarta. *JAAF (Journal of Applied Accounting and Finance)*, 4(1), 49-59.
- Muthahharah, I., and Juhari, A. (2021). A Cluster Analysis with Complete Linkage and Ward's Method for Health Service Data in Makassar City. *Jurnal Varian*, 4(2), 109-116.
- Naeem, S., and Wumaier, A. (2018). Study and implementing K-mean clustering algorithm on English text and techniques to find the optimal value of K. *International Journal of Computer Applications*, 182(31), 7-14.
- Palan, K. M., Morrow, P. C., Trapp, A., and Blackburn, V. (2011). Compulsive buying behavior in college students: the mediating role of credit card misuse. *Journal of Marketing Theory and Practice*, 19(1), 81-96.

- Sulistiyawati, A., and Supriyanto, E. (2021). Implementasi Algoritma K-means Clustering dalam Penentuan Siswa Kelas Unggulan. *Jurnal Tekno Kompak*, 15(2), 25-36.
- Urva, G. (2016). K-Means Clustering to design Radial Basis Function Neural Network (RBFNN) Classifiers. *JURNAL UNITEK*, 9(2), 16-24.
- Umuhoza, E., Ntirushwamaboko, D., Awuah, J., and Birir, B. (2020). Using unsupervised machine learning techniques for behavioral-based credit card users segmentation in Africa. *SAIEE Africa Research Journal*, 111(3), 95-101.
- Yadav, M. L., and Roychoudhury, B. (2018). Handling missing values: A study of popular imputation packages in R. *Knowledge-Based Systems*, 160, 104-118.
- Yaseen, M. K., Raheem, M., and Sivakumar, V. (2020). Credit card business in Malaysia: A data analytics approach. *International Journal of Advanced Computer Science and Applications*, 11(12) 1-10.
- Ziafat, H., and Shakeri, M. (2014). Using data mining techniques in customer segmentation. *Journal of Engineering Research and Applications*, 4(9), 70-79.