Journal of Artificial Intelligence and Engineering Applications

Website: https://ioinformatic.org/

15th February 2025. Vol. 4. No. 2; e-ISSN: 2808-4519

K-Means Clustering Method to Make Credit Payment Groupinhg Efficient

Siti Nur Illah[1*], Nana Suarna[2], Irfan Ali[3], Dodi Solihudin[4]

[1,2,3]STMIK IKMI CIREBON
Jl. Perjuangan No. 10B Majasem, Kesambi, Cirebon, Indonesia
sitinuriiillah@gmail.com[1], ieulala75829@gmail.com[2]*

Abstract

Credit payment management is one of the main challenges in the financial sector, especially in grouping customers based on risk and payment patterns. This study aims to evaluate the K-Means Clustering method in improving the efficiency of credit payment data clustering. The dataset used includes information on payment history, loan amount, tenor, and credit status from financial institutions. The research approach involves data processing stages, application of the K-Means algorithm, and evaluation of results using the Davies-Bouldin Index and Silhouette Score metrics. The results of the analysis show that the K-Means method is effective in identifying customer payment patterns and dividing them into three main clusters: high, medium, and low risk. In addition, this study found that determining the optimal number of clusters using the Elbow Method can improve the accuracy of the clustering results. The resulting model makes a significant contribution to credit risk management, helping financial institutions make strategic decisions related to credit policies and risk mitigation. This study offers practical implications, including increased operational efficiency and predictive ability against potential bad debts. Further studies are recommended to integrate this method with other algorithms to improve the performance of large-scale data analysis.

Keywords: *K-Means Clustering, Credit Grouping, Data Mining, Credit Risk*

1. Introduction

Data clustering systems have become an important part of data analysis, especially in the financial sector. One of the main challenges in managing credit payment data is how to efficiently group customers based on their payment patterns. The K-Means Clustering method is one of the widely used algorithms because of its ability to divide data into several groups quickly and effectively. In the context of credit payment grouping, the application of this method can help financial institutions identify customers with high, medium, or low risk more accurately. This study aims to evaluate the effectiveness of the K-Means Clustering method in improving the efficiency of the credit payment grouping model, so that it can be used as a basis for better decision making for financial institutions.

Grouping credit payments is one of the main challenges faced by financial institutions in managing customer data. In practice, traditional clustering often takes a long time and the results are less accurate because they are unable to handle large amounts of data with complex characteristics. This makes it difficult for financial institutions to identify customer payment patterns correctly, especially in distinguishing customers with high, medium and low risk. The inability to carry out effective grouping can result in errors in decision making, such as providing loans to customers who have a high potential for default or inefficient allocation of resources.

In addition, credit payment data often has unstructured patterns, making it difficult to analyze using conventional methods. As a result, financial institutions may miss opportunities to improve operational efficiency and optimize risk management strategies. The K-Means Clustering method offers a potential solution with its ability to group data automatically and efficiently. However, the application of this method still faces challenges in determining optimal parameters, such as the appropriate number of clusters and handling outliers. Therefore, it is important to explore the application of the K-Means Clustering method to overcome these problems and increase the efficiency and accuracy of the credit payment grouping model. This research aims to answer this challenge by evaluating the effectiveness of K-Means Clustering in the context of credit payment data management.

The K-Means Clustering method has been widely used in previous research for various data analysis purposes, including in the fields of finance and risk management. In research by Gunawan et al. (2019), K-Means is applied to group customers based on their credit payment patterns. This research shows that the K-Means method is effective in distinguishing customers with high, medium and low risk, although adjustments are needed in determining the optimal number of clusters. This study provides the basis that K-Means can be implemented in financial scenarios to assist risk analysis.

2. Research Methods

The research entitled "K-Means Clustering Method to Efficiently Group Credit Payments (Case Study: Credit Payments)" uses a quantitative approach with a focus on the application of the K-Means clustering algorithm. This approach aims to evaluate credit payment patterns in order to group customers based on their payment characteristics. The quantitative approach is used to analyze numerical data obtained from the credit payment system. This data includes information such as loan amount, credit tenor, payment time, and repayment status (on time or late). The data is processed using the K-Means clustering algorithm to produce more efficient grouping.

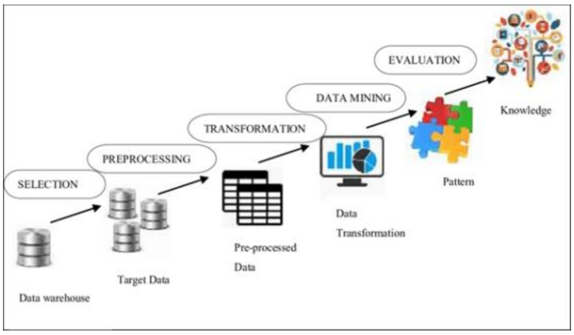

Fig. 1: Knowledge Discovery in Databases (KDD)

2.1. Knowledge Discovery in Databases (KDD)

Knowledge Discovery in Databases (KDD) is a process of discovering useful patterns, information or knowledge from large data sets. The main goal of KDD is to transform raw data into information that can be understood and used for better decision-making. This process involves several stages, from data selection and cleaning, data transformation, data mining, to evaluation and interpretation of results.

1. Selection; At this stage, relevant data is selected from available data sources. The selected data is related to credit payment information, such as loan amount, tenor, repayment history, and credit status.
2. Preprocessing; This stage involves data cleaning, such as handling missing values, detecting data duplication, and removing anomalies. The goal is to ensure that the data used is of good quality.
3. Data transformation; The cleaned data is then transformed into a format suitable for analysis. For example, normalizing numeric values or grouping certain variables to improve analysis accuracy.
4. Data mining; The main stage of the KDD process, namely the application of the K-Means clustering algorithm to group customers based on credit payment patterns. This process produces clusters that represent certain characteristics of customers.
5. Evaluation; The clustering results from the data mining stage are evaluated to ensure their accuracy and relevance to the research objectives. Evaluation can use metrics such as silhouette score or purity.
   These stages are designed to ensure that the data analysis process is carried out systematically, producing meaningful information for decision making.

2.2. K-MEANS

An algorithm that functions to group data based on the similarity of its characteristics. This method was chosen because of its efficient nature in handling large datasets and its ability to provide optimal grouping results. K-Means Clustering works by dividing data into a number of clusters (groups) based on certain centroid values. In the context of this study, this method is used to group customers based on their credit payment patterns so that it can facilitate efficiency analysis and grouping based on risk levels.

3. Results and Discussion

3.1. Results

3.1.1 Data Selection

In the process of data grouping using the K-Means Clustering method, the selection stage is an important step to ensure that the dataset used is relevant and in accordance with the research objectives. This stage aims to filter data based on certain criteria so that only data that is significant to the grouping process is retained. The dataset used in this study is presented in table 1.1 below.

Tabel 1: Dataset

ROW NO.	ID	KECAMATAN	KELURAHAN	TENOR	LASTPAYMENT
1	1	BEBER	CIKANCAS	9	Rp 1.509.000,00
2	2	JAMBLANG	BOJONG WETAN	9	Rp 1.447.000,00
3	3	HARJAMUKTI	ARGASUNYA	12	Rp 1.079.000,00
4	4	PEKALIPAN	PULASAREN	12	Rp 668.950,00
..	..	……….	……….	……….	……….
1034	1034	LEMAHWUNGKUK	PEGAMBIRAN	18	Rp 642.000,00
1035	1035	KEDAWUNG	KEDUNGJAYA	18	Rp 692.000,00
1036	1036	GUNUNG JATI	MAYUNG	24	Rp 430.000,00
1037	1037	WERU	WERU KIDUL	36	Rp 1.192.000,00

1038	1038	KESAMBI	DRAJAT	36	Rp 624.000,00

3.1.2. Prepocessing

In this study, the preprocessing stage is a crucial step in ensuring data quality before clustering using the K-Means Clustering method. One approach used is to utilize the Filter Examples operator in RapidMiner, which aims to filter data according to certain criteria, such as removing empty or irrelevant entries. In addition, the Select Attributes operator is used to select attributes that are significant and relevant to the analysis, so that the processed data becomes more focused and representative.

A. Remove Missing Value

In the preprocessing stage in this study, the Remove Missing Value step is carried out to ensure the quality and integrity of the data that will be used in the clustering process. This step uses the Filter Examples operator in RapidMiner, which is designed to filter data by removing entries that have empty or incomplete values

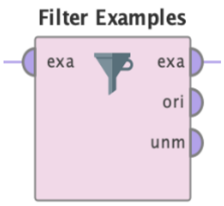

Fig. 2: Filter example

shows the use of Operator Filter Examples in RapidMiner in the data preprocessing process. This operator functions to filter data based on certain criteria, such as eliminating entries that have empty or irrelevant values. This step aims to ensure that the data used in the analysis is cleaner and more consistent, thus supporting a more accurate and efficient clustering process with the K-Means Clustering method.

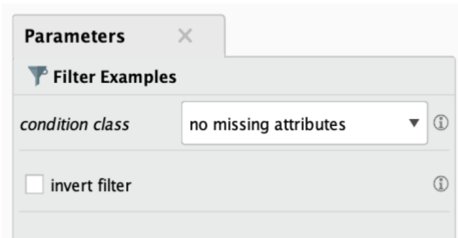

Fig. 3: Filter Parameter Configuration Examples

Shows the parameter configuration in the Filter Operator Examples, where the condition class is set to the no missing attributes option. This setting ensures that only data that does not have empty attributes will be retained in the dataset. With this configuration, the preprocessing process can eliminate incomplete data entries, thereby improving the quality and consistency of the data used for further analysis in the K-Means Clustering method. The results of the dataset filtering process can be seen in the figure below.

Row No.	KECAMATAN	KELURAHAN	TENOR	LASTPAYM...
1	BEBER	CIKANCAS	9	1509000
2	JAMBLANG	BOJONG WE...	9	1447000
3	HARJAMUKTI	ARGASUNYA	12	1079000
4	PEKALIPAN	PULASAREN	12	668950
5	DUKUPUNT...	MANDALA	12	605000
6	KESAMBI	KESAMBI	12	701000
7	KEJAKSAN	SUKAPURA	12	757000
8	LEMAHWUN...	KESEPUHAN	12	662000
9	KEJAKSAN	KEBONBARU	12	430000
10	SUMBER	SUMBER	12	668000
11	HARJAMUKTI	LARANGAN	12	776000

Fig. 4: Results of the Dataset Filtering Process.

Shows the results of data that has gone through the filtering process using the Filter Examples Operator, so that all data entries that have blank values (missing values) have been removed. The data displayed in this image is a clean dataset and is ready to be used for further analysis. This step ensures that the data to be processed is of high quality and can support the grouping process with the K-Means Clustering method optimally.

B. Generate ID

After completing the previous stage, the next step is to use the Generate ID operator in RapidMiner. This stage aims to create a unique attribute that functions as an identifier for each data entry in the dataset. This operator is very useful for ensuring that each row of data has a unique identity that can facilitate further data processing and maintain consistency in the analysis carried out.

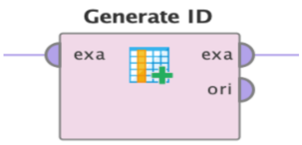

Fig. 5: Generate ID Operator

Figure 5 shows the use of the Generate ID Operator in RapidMiner. This operator is used to create unique attributes that serve as identifiers for each data entry in the dataset. The results of this process can be seen in the image below.

Row No.	id	KECAMATAN	KELURAHAN	TENOR	LASTPA...
95	95	27	61	16	5478000
177	177	5	38	18	3540600
832	832	11	265	54	3375000
941	941	22	42	12	2821000
740	740	18	269	76	2816000
736	736	25	250	75	2674000
642	642	15	152	67	2672000
543	543	23	119	56	2651000
449	449	15	31	25	2640000
425	425	7	7	20	2476000
1020	1020	14	87	18	2436000
945	945	29	199	13	2418000
893	893	30	232	12	2390000
100	100	1	63	16	2322000
327	327	5	24	13	2322000

ExampleSet (1,038 examples,1 special attribute,4 regular attributes)

Fig.6: Dataset With Unique Identifier.

C. Select Attributes

After performing the Filter Examples stage to remove data entries that have blank values, the next step is to use the Select Attributes Operator. At this stage, the operator is used to select the most relevant attributes and have a significant influence on the analysis being performed.

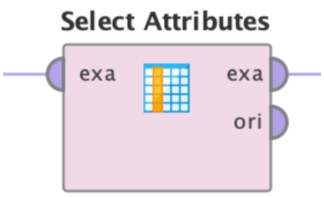

Fig. 7: Results of Dataset Filtering Process.

Figure 7 shows the use of the Select Attributes Operator in RapidMiner. This operator functions to select relevant and significant attributes in the dataset after the filtering process using Filter Examples. By using this operator, only attributes that have an important contribution to the analysis will be retained, while irrelevant attributes can be removed. This process aims to simplify the dataset, improve data processing efficiency, and ensure that only attributes that support the purpose of grouping credit payments are retained in the analysis.

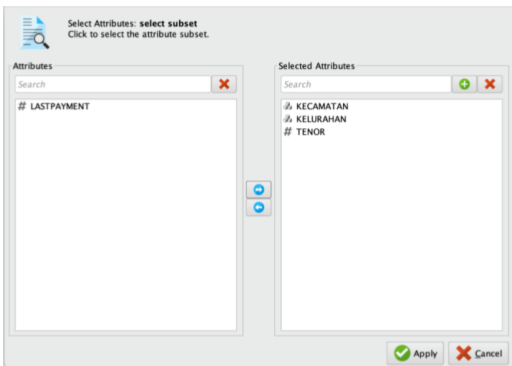

Fig. 8: Selection of Subset Attributes.

Figure 8 shows the configuration of the Select Attributes Operator in RapidMiner, which is used to select relevant attributes in the dataset. In this view, the Sub-district, Village, and Tenor attributes have been selected to be included in the dataset to be used in further analysis. Meanwhile, the LastPayment attribute is not included in the selected attributes. This process allows the selection of a subset of relevant attributes for clustering purposes using the K-Means Clustering method.

Row No.	KECAMATAN	KELURAHAN	TENOR
1	BEBER	CIKANCAS	9
2	JAMBLANG	BOJONG WE...	9
3	HARJAMUKTI	ARGASUNYA	12
4	PEKALIPAN	PULASAREN	12
5	DUKUPUNT...	MANDALA	12
6	KESAMBI	KESAMBI	12
7	KEJAKSAN	SUKAPURA	12
8	LEMAHWUN...	KESEPUHAN	12
9	KEJAKSAN	KEBONBARU	12
10	SUMBER	SUMBER	12
11	HARJAMUKTI	LARANGAN	12

Fig. 9: Results of Select Attributes.

Figure 9 shows the results of the Select Attributes process, where three relevant attributes have been selected for inclusion in the dataset. The selected attributes include District, Village, and Tenor. In addition, this figure also shows that the selection of these attributes does not change the number of rows in the dataset, meaning that the remaining data remains intact at the original number of rows even though only relevant attributes are retained for further analysis. This process aims to simplify the data and focus on attributes that have a significant contribution to the research objectives.

3.1.3 Transformation

After the preprocessing stage is complete, the next step is to convert non-numeric data into numeric so that it can be further processed in clustering algorithms such as K-Means. This process is carried out using the Nominal to Numerical Operator in RapidMiner. This operator functions to convert nominal attributes (such as categories or labels) into numeric attributes, which allows the data to be used in mathematical and statistical algorithms that require input in numeric format.

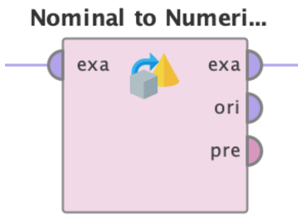

Fig. 10: Nominal To Numerical Operator.

Figure 10 shows the display of Nominal to Numerical Operator in RapidMiner. This operator is used to convert nominal or categorical attributes into numeric format so that they can be further processed in algorithms that require numeric input, such as K-Means Clustering.

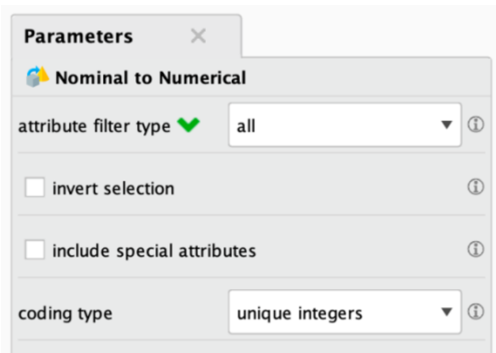

Fig. 11: Parameter Operator Nominal to Numerical.

Figure 11 shows the configuration of the Nominal to Numerical Operator parameter in RapidMiner. In this figure, the Attribute Filter Type is set to "All," which means all nominal attributes in the dataset will be processed to be converted into numeric format. In addition, the Coding Type is set to "Unique Integers," which means each nominal category will be given a unique number as its numeric representation.

With this setting, each category value will be converted into an integer that can be used in the analysis process that requires numeric data, such as the K-Means Clustering algorithm. The transformation results can be seen in the figure below.

Row No.	KECAMATAN	KELURAHAN	TENOR
1	0	0	9
2	1	1	9
3	2	2	12
4	3	3	12
5	4	4	12
6	5	5	12
7	6	6	12
8	7	7	12
9	6	8	12
10	8	9	12
11	2	10	12

Fig. 12: Numerical Transformation Results.

3.1.4 Data Mining

In the Data Mining stage, the K-Means algorithm will be applied to group the data that has been processed in the preprocessing stage. This algorithm is used to divide the dataset into several clusters based on the similarity of patterns in the data, with the aim of finding hidden structures or patterns that can help in further analysis.

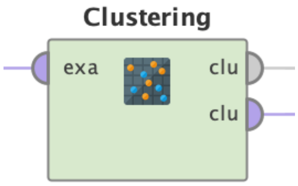

Fig. 13: Operator K-Means.

Figure 13 shows the Clustering Operator (K-Means) in RapidMiner, where the K value is set to 7. In this setting, the data will be grouped into 7 different clusters based on the similarity of patterns in the data. In addition, the performance of each cluster will be measured using the Cluster Distance Performance Operator to evaluate how well the clustering results are performed by the K-Means algorithm, by measuring the distance between data in each cluster. The figure below shows the Cluster Distance Performance operator.

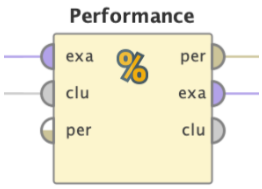

Fig. 14: Cluster Distance Performance Operator.

Figure 14 shows the Cluster Distance Performance Operator in RapidMiner. This operator is used to evaluate the performance of clusters generated by the K-Means algorithm by measuring the distance between data points in each cluster. By using this operator, it can be analyzed to what extent the data in one cluster is distributed homogeneously, as well as how well the separation between the existing clusters is. This evaluation is important to ensure the quality of the clustering performed and can help in selecting the optimal number of clusters. Furthermore, the Subprocess operator will also be used to organize a more complex workflow by compiling several processes in one more structured module.

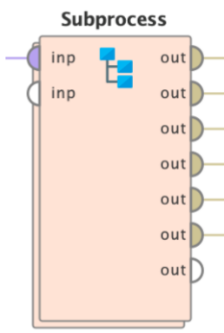

Fig. 15: Subprocess Operator.

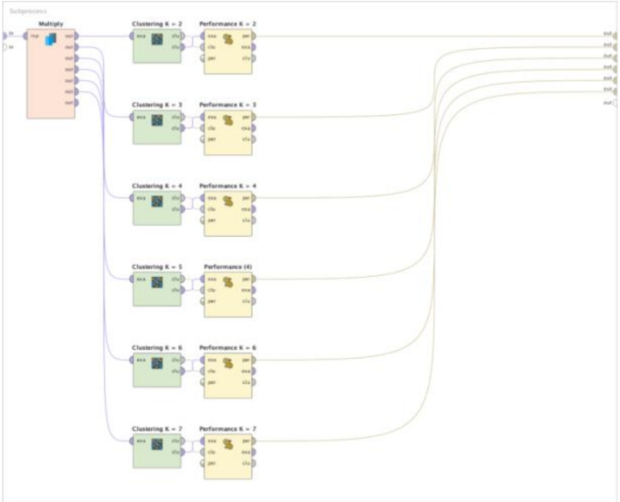

Fig. 16: Application of K-Means in Subprocess.

Figure 16 shows the data mining process carried out to determine the most ideal K value in the K-Means Clustering method in this study. The process involves testing various K values, ranging from K=2 to K=7, with the aim of evaluating clustering performance based on the dataset used. Each test involves dividing the data into groups based on a certain K value, followed by analyzing the performance of the clustering results using relevant evaluation metrics. The results of each test are then compared to identify the K value that provides the most optimal clustering results. This process is expected to produce an ideal K value, thus supporting the research objective in improving the efficiency of credit payment clustering.

3.1.5 Evaluation & Interpretation

At this stage, an analysis is carried out to evaluate the quality of the clustering results in order to determine the most optimal number of clusters (K). This evaluation uses the Davies-Bouldin Index (DBI) metric, which functions as an indicator to measure how well the resulting data is grouped. The DBI values from each test with various K values are presented in Table 4.3, which provides an overview of the effectiveness of each cluster configuration in the clustering process.

Table 2: DBI Value of Each Cluster.

K	Davies Bouldin Index
2	0.199
3	0.253
4	0.279
5	0.279
6	0.281
7	0.286

Table 2 above shows the Davies-Bouldin Index (DBI) values for various numbers of clusters (K) resulting from the clustering process using the K-Means method. The DBI value is used as an indicator of cluster quality, where a lower value indicates a more optimal clustering. Based on the table, the lowest DBI value of 0.199 was obtained at K = 2, which indicates that two clusters provide the best clustering results compared to other numbers of clusters. These results are the basis for consideration in determining the most appropriate number of clusters for the dataset used in this study.

Cluster Model

```
Cluster 0: 326 items
Cluster 1: 712 items
Total number of items: 1038
```

Fig. 17: Cluster Model.

Figure 17 shows the results of data clustering using the K-Means method with a predetermined number of clusters (K), namely K = 2. Based on the cluster model, the data is divided into two groups, Cluster 0 containing 326 items and Cluster 1 containing 712 items, with a total of 1,038 items. These results provide an overview of the data distribution in each cluster, which can be used for further analysis in this study.

PerformanceVector

```
PerformanceVector:
Avg. within centroid distance: 751.805
Avg. within centroid distance_cluster_0: 1050.117
Avg. within centroid distance_cluster_1: 615.218
Davies Bouldin: 0.199
```

Fig. 18: Cluster Vector Performance.

Figure 18 shows the results of clustering performance evaluation using the K-Means method with the number of clusters (K=2). This evaluation involves metrics such as average within centroid distance and Davies-Bouldin Index (DBI) values. The average value of the distance between data points to their centroids was recorded at 751,805, with details of the average distance for Cluster 0 of 1050,117 and Cluster 1 of 615,218. The Davies-Bouldin Index value obtained was 0.199, which indicates that clustering with K=2 produces optimal cluster quality. This information indicates the efficiency of data clustering in the study.

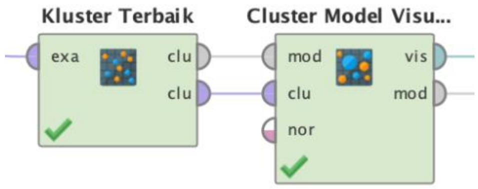

Fig. 19: Best Clusters & Cluster Model Visualizer Operator.

Figure 19 shows the Clustering operator with the best k of 2 and the Cluster Model Visualizer operator used to display the visualization of the best known clusters.

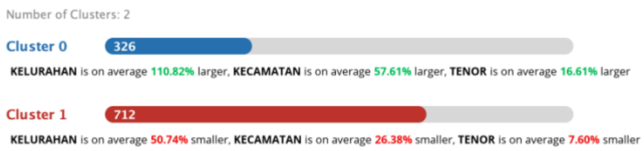

Fig.20: Visualization of Cluster 0 and 1.

Figure 20 illustrates the results of grouping credit payment data using the K-Means method which produces two clusters. Cluster 0 consists of 326 data with the characteristics of an average value of the Village variable which is 110.82% larger, District 57.61% larger, and Tenor 16.61% larger than the overall average. In contrast, Cluster 1 consists of 712 data with an average value of the Village variable 50.74% smaller, District 26.38% smaller, and Tenor 7.60% smaller than the overall average. Based on these results, it can be concluded that Cluster 0 reflects a group with a higher credit payment area and longer tenor, while Cluster 1 shows an area with a lower credit payment rate.

Cluster	KECAMATAN	KELURAHAN	TENOR
Cluster 0	23.439	200.966	43.307
Cluster 1	10.948	46.958	36.184

Fig. 21: Centroid Table.

Figure 21 shows the centroid table of the results of clustering using the K-Means method on credit payment data. This table displays the average value or center position (centroid) for each variable in each cluster. In each Cluster there are 3 variables, the centroid value of Cluster 0, the District variable is 23.439, the Village variable is 200.966, and the Tenor variable is 43.307. Meanwhile, in Cluster 1, the centroid value of the District variable is 10.948, the Village variable is 46.958, and the Tenor variable is 36.184.

This centroid represents the main characteristics of each cluster, where Cluster 0 shows a higher average value for all variables compared to Cluster 1. This difference reflects a clear pattern in the data, which can be used to understand the distribution of credit payments by region (Sub-district and Village) and payment tenor.

Fig. 22: Scatter Plot Visualization.

Figure 22 shows the visualization of the results of grouping credit payment data using the K-Means method in two clusters. This graph uses a scatter diagram type to show the distribution of data based on the Kelurahan variable on the X axis and the Tenor variable on the Y axis. The data is grouped into two clusters, namely Cluster 0 (represented by green dots) and Cluster 1 (represented by blue dots). From this graph, it can be seen that Cluster 0 tends to have a higher Kelurahan value than Cluster 1. In addition, there is variation in the Tenor value, where Cluster 0 mostly has a higher Tenor than Cluster 1. This graph shows a clear distribution pattern, which strengthens the

differences in characteristics between the two clusters. These results can be used to further understand the differences in credit payment segments based on location (Kelurahan) and payment period (Tenor).
Based on the results of the evaluation and interpretation of the data, the K-Means method with the number of clusters (K = 2) proved to produce the most optimal grouping in this study. This is supported by the lowest Davies-Bouldin Index (DBI) value of 0.199, which indicates that the quality of clustering at K = 2 is better compared to the number of other clusters. In addition, centroid analysis and scatter plot visualization reveal clear differences in characteristics between Cluster 0 and Cluster 1, both in terms of the value of the regional variable (Sub-district and District) and the tenor of credit payments. Cluster 0 reflects a group with a higher credit payment area and a longer payment tenor, while Cluster 1 reflects an area with a lower credit payment rate and a shorter tenor. These differences in characteristics provide an in-depth picture of the data distribution pattern, which can be the basis for developing a more efficient and targeted credit payment strategy.
With these results, the K-Means method is not only able to group data effectively, but also provides relevant insights to support decision making based on cluster analysis. This evaluation stage confirms the success of the application of the clustering method in research to improve the efficiency of credit payment grouping.

3.2. Discussion

The results of this study indicate that the learning strategies used significantly increase student engagement and understanding of the learning material. This is in line with the findings of (Ahsina et al., 2022).
which highlights that a collaboration-based approach can increase student learning motivation. In addition, this study supports the view (Fahlevi et al., 2023), that the integration of technology in learning creates a more interesting and interactive learning experience, especially in today's digital era.

These results are also consistent with the study (Holwati et al., 2023), which emphasizes the importance of innovative approaches to improve learning effectiveness.
However, this study differs from the findings (Limia Budiarti & Cendana, 2022), which focus more on the influence of the learning environment on learning outcomes, while this study emphasizes teaching methods.
This study strengthens the view (Saputra & Yusuf, 2024), that the use of learning media that is relevant to student needs contributes to better results. However, this study emphasizes the project-based learning approach more than their focus on digital media.
In addition, (Putra & Hartomo, 2021), noted that individual factors such as students' learning styles also influence the effectiveness of learning, while this study examines the collective effects of teaching methods on the class as a whole.
(Astuti et al., 2022), also showed that adapting the curriculum according to students' needs can improve learning outcomes, which is relevant to this finding. However, this study offers a new contribution by highlighting the direct influence of certain teaching strategies on the mastery of certain skills, in contrast to their general approach.
(Dwididanti & Anggoro, 2022), emphasizes the importance of continuous evaluation to improve the quality of learning, which is relevant as part of the methodology of this study.
(Fitriani et al., 2023), emphasizes that learning success is often influenced by the active role of teachers in managing the class, which is also reflected in this study. However, this study provides additional contributions by presenting empirical data on the influence of teaching approaches on learning outcomes directly.
Meanwhile, research (Ondra Eka Putra & Randy Permana, 2024), emphasizes the importance of personalizing learning, which, although not the main focus of this research, is relevant as a context for understanding

4. Conclusion

This study aims to apply the K-Means Clustering method in grouping credit payment data based on certain characteristics, such as payment patterns, late history, and credit amount. In addition, this study also aims to analyze the efficiency of the K-Means Clustering method in improving the process of grouping credit payment data compared to conventional methods. Based on the results obtained, it can be concluded that this method provides optimal and efficient results in grouping credit payment data, with the Davies-Bouldin Index value indicating very good cluster quality. The following are the conclusions of the results of this study.:
1. This study successfully applied the K-Means Clustering method to group credit payment data based on certain characteristics, such as payment patterns, late history, and credit amount. The clustering results showed two optimal clusters (K = 2) with the lowest Davies-Bouldin Index (DBI) value of 0.199, which reflects very good clustering quality. 2. The K-Means Clustering method is proven to be more efficient than conventional methods in the process of grouping credit payment data. This can be seen from the results of the analysis which reveal clear differences in characteristics between clusters, providing deeper and more relevant insights to support strategic decision making.

References

[1] N. Ahsina, F. Fatimah, and F. Rachmawati, "Analisis Segmentasi Pelanggan Bank Berdasarkan Pengambilan Kredit Dengan Menggunakan Metode K-Means Clustering," *J. Ilm. Teknol. Infomasi Terap.*, vol. 8, no. 3, 2022, doi: 10.33197/jitter.vol8.iss3.2022.883.

[2] M. R. Fahlevi, D. Ridha, D. Putri, and E. Syahrin, "Analisis Pengelompokan Data Pelelangan Barang Dengan Metode K-Means Clustering," *Jurasik (Jurnal Ris. Sist. Inf. dan Tek. Inform.*, vol. 8, no. 1, pp. 53–61, 2023.

[3] Holwati, E. Widodo, and W. Hadikristanto, "Pengelompokan Untuk Penjualan Obat Dengan Menggunakan Algoritma K-Means," *Bull. Inf. Technol.*, vol. 4, no. 3, pp. 408–413, 2023, doi: 10.47065/bit.v4i3.848.

[4] R. Limia Budiarti and G. Cendana, "Klasifikasi Data Nasabah Kredit Pinjaman Menggunakan Data Mining Dengan Metode K-Means Pada Mega Central Finance," *J. Akad.*, vol. 14, no. 2, pp. 88–94, 2022, doi: 10.53564/akademika.v14i2.866.

[5] A. Saputra and R. Yusuf, "Perbandingan Algoritma DBSCAN dan K-MEANS dalam Segmentasi Pelanggan Pengguna Transportasi Publik Transjakarta Menggunakan Metode RFM," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 4, no. 4, pp. 1346–1361, 2024, doi: 10.57152/malcom.v4i4.1516.

[6] A. C. Putra and K. D. Hartomo, "Optimalisasi Penyaluran Bantuan Pemerintah Untuk UMKM Menggunakan Metode Fuzzy C-Means," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 1, no. 10, pp. 474–482, 2021.

[7] N. Astuti, J. N. Utamajaya, and A. Pratama, “Penerapan Data Mining Pada Penjualan Produk Digital Konter Leppangeng Cell Menggunakan Metode K-Means Clustering,” *JURIKOM (Jurnal Ris. Komputer)*, vol. 9, no. 3, p. 754, 2022, doi: 10.30865/jurikom.v9i3.4351.

[8] S. Dwididanti and D. A. Anggoro, “Analisis Perbandingan Algoritma Bisecting K-Means dan Fuzzy C-Means pada Data Pengguna Kartu Kredit,” *Emit. J. Tek. Elektro*, vol. 22, no. 2, pp. 110–117, 2022, doi: 10.23917/emitor.v22i2.15677.

[9] M. N. R. Fitriani, B. Priyatna, B. Huda, A. L. Hananto, and T. Tukino, “Implementasi Metode K-Means Untuk Memprediksi Status Kredit Macet,” *J. Sist. Komput. dan Inform.*, vol. 4, no. 3, p. 554, 2023, doi: 10.30865/json.v4i3.5953.

[10] Ondra Eka Putra and Randy Permana, “Hybrid Data Mining For Member Determination And Financing Prediction In Syariah Financing Saving And Loan Cooperatives,” *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 8, no. 2, pp. 309–320, 2024, doi: 10.29207/resti.v8i2.5683.

[11] B. Jimbaran, “OPTIMIZATION OF K-MEANS CLUSTERING USING PARTICLE SWARM OPTIMIZATION ALGORITHM FOR GROUPING TRAVELER REVIEWS DATA ON TRIPADVISOR SITES a I Made Satria Bimantara, b I Made Widiartha,” vol. 12, no. 1, pp. 1–10, 2023.