Multitek Indonesia: Jurnal Ilmiah
Volume: 20 No. 1 July 2026, Hal. 116 - 127
ISSN: 1907-6223 (print)
ISSN: 2579-3497 (Online)
https://journal.umpo.ac.id/index.php/multitek
https://doi.org/10.24269/mtkind.v20i1.13856

Application Of The C4.5 Method Based On K-Means Segmentation For Loan Risk Classification At Savings And Loan Cooperative

Jihan Martha[1], **Danar Putra Pamungkas**[1], **Made Ayu Dusea Widyadara**[1]
[1] Universitas Nusantara PGRI Kediri
Jl. KH Achmad Dahlan No. 76, Mojoroto, Kec. Mojoroto, Kota Kediri, Jawa Timur 64112
Corresponding Author's Email : jihannmarthaa@gmail.com

Received : May 11, 2026 **Revised : June 7, 2026** **Accepted : June 19, 2026**

Abstract

The Sri Hadi Dharma Savings and Loan Cooperative faces challenges in conducting loan risk analysis, which is still done manually as a result, the decisions made tend to be subjective and inconsistent. This situation has the potential to lead to errors in risk assessment, which could increase the likelihood of non-performing loans. Additionally, the available payment history data is not labeled with risk categories and therefore cannot be directly used in the classification process. This study aims to build a more objective loan risk prediction model by integrating the K-Means and C4.5 algorithms. The data used in this study consists of 7,155 loan payment history records from the 2020-2025 period. The research stages included data preprocessing through cleaning, reduction, creation of a maximum delinquency feature, feature selection, and Min-Max normalization, resulting in 471 unique data points. Next, K-Means was used to form risk groups based on financial characteristics, the clustering results were labeled as low, medium, and high risk categories, then the labeled data were classified using C4.5 with an 80:20 split of training and test data to generate a prediction model. Model evaluation was performed using a confusion matrix with accuracy, precision, recall, and F1-score parameters. The results show that the resulting model is capable of providing predictions with an accuracy rate of 95.79% and generates decision rules that are easy to understand and interpret. Furthermore, the integration of these two methods transforms payment history data into valuable information for identifying members' risk patterns. This model is effective in supporting more objective, transparent, and measurable decision-making in loan risk management at savings and loan cooperatives.

Keyword: *c4.5, k-means, cooperative, risk prediction, loan risk*

INTRODUCTION

Savings and loan cooperatives are one type of non-bank financial institution that plays a vital role in supporting community economic growth by providing financing services to their members [1]. The existence of savings and loan cooperatives contributes to improving members' welfare through flexible and easily accessible savings and loan services [2]. However, in practice, savings and loan cooperatives also face loan risk, namely the possibility of members' delays or failures in meeting their loan repayment obligations. High loan risk can disrupt the savings and loan cooperative's liquidity and operational stability [3].

Loan risk analysis at savings and loan cooperatives is often still conducted manually, based on the experience of staff or management. This approach tends to be subjective, inconsistent, and time-

consuming [4]. The use of data mining techniques can serve as a solution to facilitate an objective analysis process based on loan payment history data. Data mining enables the extraction of patterns from data to support segmentation and prediction [5].

One widely used method for data segmentation is the K-Means clustering algorithm. Previous studies have demonstrated the reliability of this method, such as in the clustering of cooperative members' loans, which was able to identify specific patterns to facilitate the identification of risk categories [6]. Furthermore, the effectiveness of K-Means has also been tested in identifying potential customers at small and medium-sized financial institutions, as well as in classifying credit card user characteristics based on their transaction behavior [7], [8]. K-Means was chosen because it is able to discover clustering patterns in numerical data based on the similarity of characteristics without requiring initial labels [9].

In addition to clustering, classification algorithms are also widely used to predict loan eligibility or risk one such algorithm is C4.5. The main advantage of C4.5 lies in its ability to generate accurate decision trees that remain easy to interpret as a basis for loan approval decisions [10]. Previous research in the cooperative sector has shown that this algorithm performs well in processing both numerical and categorical attributes, making it highly relevant for objectively determining loan eligibility [11], [12], [13].

Based on previous research, most studies have used clustering methods for segmentation or classification methods for prediction separately. There has been little research integrating clustering results as automatic labels to build predictive classification models, particularly for cooperative members' loan payment history data [14]. Therefore, this study proposes the integration of the K-Means and C4.5 algorithms, where K-Means is used to perform automatic loan risk labeling, and the resulting clusters are then used as target classes for the classification process using C4.5.

The objective of this study is to develop a model for predicting the loan risk of cooperative members using the C4.5 algorithm based on the results of K-Means clustering of loan payment history data for the 2020-2025 period, and to evaluate the model's performance using a confusion matrix, accuracy, precision, recall, and F1-score. The findings of this study are expected to assist cooperatives in conducting loan risk analysis more quickly, accurately, and consistently.

RESEARCH METHOD

This study uses a data mining approach to develop a model for predicting loan risk among members of the Sri Hadi Dharma Savings and Loan Cooperative. The method employed is a combination of the K-Means Clustering and C4.5 algorithms.

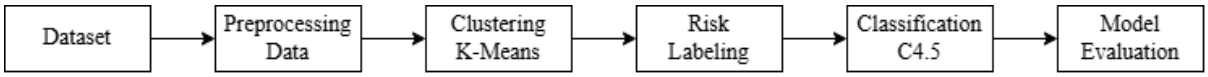

Figure 1. Research Phases

2.1. Dataset

The data used in this study consists of secondary data in the form of members' loan payment histories, obtained directly from the cooperative's management in the form of digital documents, totaling 7,155 records. This data reflects members' loan payment history for the 2020-2025 period. The attributes used in this study include Member ID, Member Name, Disbursement Date, Employment

Status, Loan Amount, Outstanding Loan Balance, Installment Amount, Remaining Tenor, Interest, and Arrears.

2.2 Preprocessing Data

The preprocessing stage is performed to prepare the data before the clustering process. The steps involved include:

1. Data Cleaning
   Removing missing data and outliers.
2. Calculating Maximum Arrears
   Determining the highest arrears value for each member ID.
3. Data Reduction
   Selecting one data point per Member ID with the shortest Remaining Tenor.
4. Feature Selection
   Selecting attributes relevant to loan risk analysis.
5. Data Transformation (Normalization)
   Standardizing the data scale using Min–Max Normalization.

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}} \qquad (1)$$

This stage aims to improve data quality and maintain scale consistency across attributes [15].

2.3 Clustering Using the K-Means Algorithm

At this stage, a clustering process is performed using the K-Means algorithm to group member data into several risk levels based on loan repayment patterns. The main steps in the K-Means algorithm include:

1. Determine the number of clusters (k),
2. Determine the initial centroids,
3. Calculate the distance of the data points from the centroids using Euclidean distance,

$$d(x, y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2} \qquad (2)$$

4. Assign the data points to the nearest cluster,
5. Update the centroids until convergence.

K-Means was applied because it can cluster data based on the similarity of numerical characteristics without requiring initial labels [6], [16]. Cluster stability was then tested using the Adjusted Rand Index (ARI) through several random state experiments to ensure that the clustering results remained consistent. Additionally, Principal Component Analysis (PCA) was used to reduce the data dimensions so that the cluster distribution could be visualized more clearly [17].

2.4 Risk Labeling

The clustering results are then used as the basis for determining the risk labels of member loans. Each cluster formed is interpreted as a risk category as follows:

1. Cluster 0 = Low Risk
2. Cluster 1 = Medium Risk
3. Cluster 2 = High Risk

This labeling aims to convert the clustering results into target variables that can be used in the classification process

3 Classification Using the C4.5 Algorithm

The C4.5 algorithm is used to build a classification model in the form of a decision tree based on labeled data [12]. The resulting model is used to predict the risk category for new member data. The classification process is carried out through the following steps:

1. Determination of Variables
   The data is divided into X (attributes) and Y (risk labels).
2. Data Split (Train-Test Split)
   The dataset is divided into training and test data (80:20) to maintain a balance between model training and testing.
3. Selection of the Best Attributes
   This is performed using the following calculations:
   1. Entropy
      Measures the level of data uncertainty [18].
      $$Entropy(S) = -\sum_{i=1}^{n} - p_i \, log_2 \, p_i \qquad (3)$$
   2. Information Gain
      Measures the influence of an attribute in distinguishing data [18].
      $$Gain(S, A) = Entropy(S) - \sum_{i=1}^{n} \frac{|S_i|}{S} Entropy(Si) \qquad (4)$$
   3. Split Information
      Measures the proportion of data split by an attribute [18].
      $$SplitInfo(S, A) = -\sum_{i=1}^{n} \frac{|S_i|}{|S|} log_2 \left(\frac{|S_i|}{|S|}\right) \qquad (5)$$
   4. Gain Ratio
      Determines the best attribute in a more balanced manner [18].
      $$GainRatio(S, A) = \frac{Gain(S,A)}{SplitInfo(S,A)} \qquad (6)$$
4. Decision Tree Construction
   The best attribute is selected as a node, and the process is repeated until a classification model is formed [18].

4 Model Evaluation

Model evaluation is performed using a confusion matrix with the following metrics:

1. Accuracy
   Used to measure the model's accuracy [19].
   $$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \qquad (7)$$
2. Precision
   Used to measure the accuracy of positive predictions [19].
   $$Precision = \frac{TP}{TP+FP} \qquad (8)$$
3. Recall
   Used to measure the model's ability to detect a specific class [19].
   $$Recall = \frac{TP}{TP+FN} \qquad (9)$$
4. F1-score
   Used to measure the balance between precision and recall [19].
   $$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \qquad (10)$$

Description:

1. TP (True Positive) is the number of positive data points that were correctly predicted.
2. TN (True Negative) is the number of negative data points that were correctly predicted.
3. FP (False Positive) is the number of negative data points that were predicted as positive.
4. FN (False Negative) is the number of positive data points that were predicted as negative.

This evaluation aims to measure the model's performance in classifying members' loan risk levels [20].

RESULTS AND DISCUSSION

This section presents the results of the data mining modeling phase, which utilized a combination of the K-Means Clustering and C4.5 algorithms for classifying member loan risk, along with an analysis of model performance and an interpretation of the resulting decision trees.

3.1. Results of Data Preprocessing

The initial stage of the study involved preparing the raw data so that it was ready to be processed by the K-Means and C4.5 algorithms. A comparison of the data before and after this process is presented in the following table:

Table 1. Data Before Preprocessing

ID	Member Name	Employment Status	Loan Amount	Outstanding Balance	Installment Amount	Interest	Arrears	Remaining Tenor
10	Soepandi/Hentri	Retirees	5400000	1800000	450000	108000	252000	4
26	Pitran Triyono	General	5040000	840000	420000	50400	67500	2
46	Soni	General	9000000	750000	750000	45000	270000	1
46	Soni	General	5040000	2100000	420000	126000	0	5
46	Soni	General	5040000	5040000	420000	302400	0	12

Table 2. Data After Preprocessing

ID	Member Name	Employment Status	Loan Amount	Outstanding Balance	Installment Amount	Remaining Tenor	Maximum Arrears
10	Soepandi/Hentri	Retirees	0.198	0.0842	0.2492	0.1765	0.0571
26	Pitran Triyono	General	0.1834	0.038	0.2308	0.0588	0.0153
46	Soni	General	0.3447	0.0338	0.4338	0	0.0612
59	Juari	General	0.1418	0.0139	0.1785	0	0.0336
60	Didik Subandi	Retirees	0.1002	0.0338	0.1262	0	0.1574

Table 1. Data Before Preprocessing and Table 2. Data After Preprocessing show the changes in data structure and values following a series of sequential processes. The process began with data cleaning to remove null values and outliers from the initial 7,155 data rows. Next, the maximum arrears were calculated for each member ID, where the system identified the highest arrears value from the individual's entire loan history to capture the worst-case risk profile as the basis for classification. Then, data reduction was performed by filtering the dataset down to 471 unique records, where each member ID was represented by only one record based on the shortest remaining tenor to avoid redundancy that could introduce bias into the model. Next, the application of Min-Max Normalization to the selected features Loan Amount, Outstanding Loan Balance, Installment Amount, Remaining Tenor, and Maximum Arrears successfully standardized the value ranges to a 0-1 scale without losing the essence of the original information

3.2. Clustering Results Using K-Means

This stage aims to objectively group cooperative members based on the similarity of their loan characteristics. The number of clusters was set to K=3 to form three main groups, so that the loan risk level of each member can be classified more clearly and systematically.

3.2.1. Distribution of Members by Cluster

Table 3. Distribution of Members by Cluster

Cluster	Number of Members
Cluster 0	323
Cluster 1	95
Cluster 2	53

Table 3. Distribution of Members by Cluster presents the breakdown of 471 unique data points into three distinct groups. The data show that the majority of members are concentrated in Cluster 0, with 323 individuals, while the remainder are divided into two smaller groups: Cluster 1, with 95 individuals, and Cluster 2, with 53 individuals, both of which exhibit more specific characteristics

3.2.2. Data Distribution by Cluster

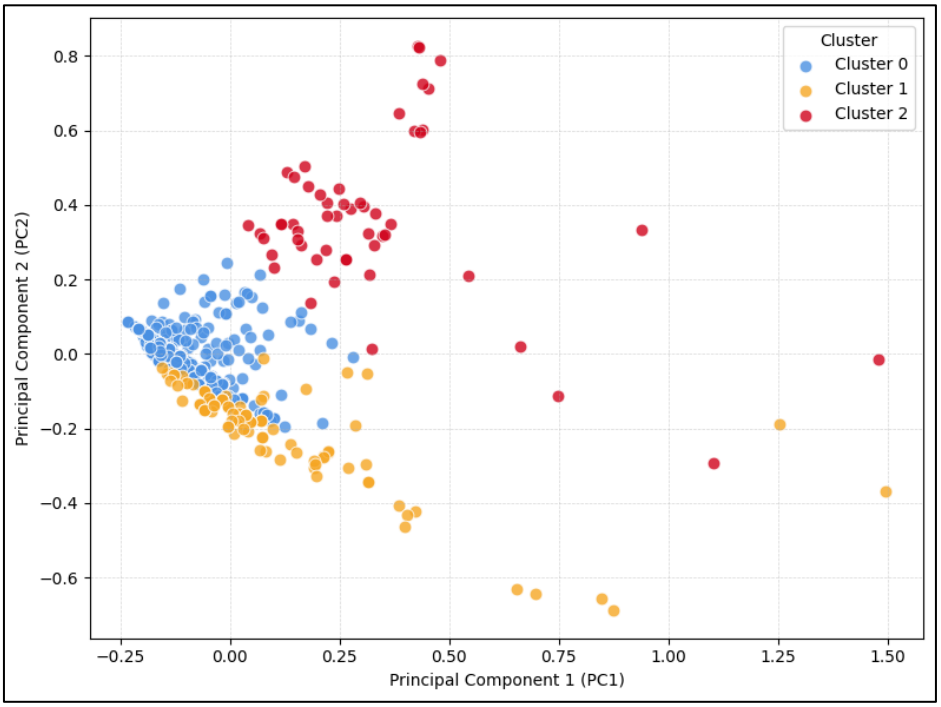

Figure 2. Data Distribution by Cluster

Figure 2. Data Distribution by Cluster shows a visualization of the data distribution resulting from clustering, mapped using Principal Component Analysis (PCA). This graph illustrates a fairly distinct separation between groups, where each point represents the unique profile of a single cooperative member. PC1 is used to represent the largest variation in the data, while PC2 is used to capture the next variation not covered by PC1, allowing the separation patterns between clusters to be visualized more clearly.

3.2.3. Maximum Feature Values for Each Cluster

Table 4. Maximum Feature Values in Each Cluster

Cluster	Loan Amount	Outstanding Balance	Installment Amount	Remaining Tenor	Maximum Arrears
0	9000000	1500000	750000	6	1275000
1	25080000	20900000	1670000	18	3660700
2	20040000	11690000	1670000	10	4414800

Table 4. Maximum Feature Values in Each Cluster shows the highest value of each attribute that forms the basis for pattern formation in each group. The following are the technical characteristics of each cluster:

1. Cluster 0: Represents the group with the most stable financial profile, characterized by the lowest Loan Amount (9.000.000) and Outstanding Loan Balance (1.500.000). Members of this group are in the final phase of repayment (maximum Remaining Tenor of 6), so the low Principal Installment Amount (750.000) is directly proportional to the minimal Maximum Arrears (1.275.000).
2. Cluster 1: This is the group with the highest loan exposure (Loan Amount 25.080.000 and Outstanding Loan Balance 20.900.000). The provision of a long Remaining Tenor of up to 18

months indicates a proportional time buffer for the large loan amount. Although bearing the maximum principal installment Amount (1.670.000), the accumulation of the maximum arrears (3.660.700) is considered a logical consequence of the still very high Outstanding Loan Balance.

3. Cluster 2: Shows the most contrasting conditions because it records the highest Maximum Arrears (4.414.800) among all groups. This occurs because although the Loan Amount (20.040.000) and Outstanding Loan Balance (11.690.000) are lower than in Cluster 1, members are still burdened with the maximum Principal Installment Amount (1.670.000) and a tighter repayment period (maximum Remaining Tenor of 10), thereby triggering more severe payment difficulties.

This mapping of the clustering results provides an empirical basis for the C4.5 algorithm to automatically extract decision rules during the classification process in subsequent research phases.

3.2.4. Cluster Stability Test

Table 5. Results of the Stability Test Using the Adjusted Rand Index (ARI)

Random State Comparison	Adjusted Rand Index (ARI)
Random State 1 vs 10	1.0
Random State 10 vs 20	0.99
Random State 20 vs 30	1.0
Random State 30 vs 42	0.92

Table 5. Results of the Stability Test Using the Adjusted Rand Index (ARI) presents the results of the clustering consistency evaluation through testing of five different random initial states. The results show very high and consistent ARI values approaching 1.0, indicating that the formation of member risk clusters does not change significantly even when the initial centroid positions are randomized. This strong stability ensures that the risk labels generated by K-Means are permanent and valid as a learning basis for the C4.5 algorithm during the classification stage.

3.3. Risk Labeling

This stage serves to transform the results of unsupervised clustering into labeled data to meet classification requirements. Risk labels for each cluster are determined by evaluating the influence of previously identified features.

3.3.1. Feature Correlation with Clusters

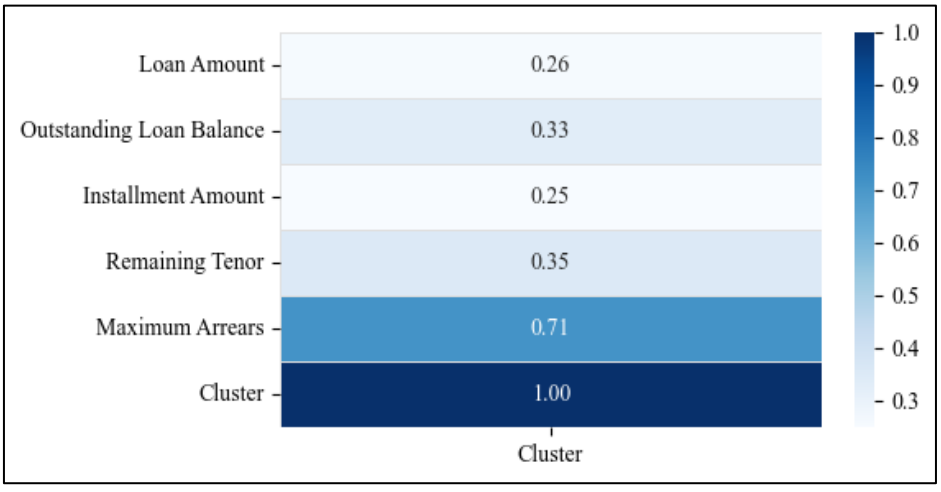

Figure 3. Feature Correlation with Clusters

Figure 3. Feature Correlation with Clusters shows the degree of influence each feature has on the formation of member groups. Based on this visualization, the Maximum Arrears feature has the highest correlation value of 0.71, confirming its dominance as the primary indicator in distinguishing clusters. Other supporting features such as Remaining Tenor (0.35) and Outstanding Loan Balance (0.33) also

contribute significantly to clarifying cluster boundaries, while the Loan Amount (0.26) and Principal Installment Amount (0.25) features complement the data profile with more moderate correlation weights.

3.3.2. Loan Risk Labeling

Table 6. Loan Risk Labeling

Cluster	Risk Label
Cluster 0	Low
Cluster 1	Medium
Cluster 2	High

Table 6. Loan Risk Labeling presents the results of cluster mapping into target classification classes. The determination of these labels aligns with the concept of loan risk management, where the interrelationship of all features particularly Maximum Arrears serves as a key parameter for measuring a member's potential for default. The integration of this labeling effectively provides ground truth for the C4.5 algorithm to extract accurate decision rules in subsequent research stages.

3.4. Classification Results Using C4.5

This stage constitutes the core of loan risk classification modeling, utilizing an 80:20 data split (training and test sets). The applied C4.5 algorithm parameters include the Gain Ratio criterion, a maximum depth of 5 levels, pruning with a confidence of 0.1, and prepruning (minimum gain of 0.01 and minimum leaf size of 2).

3.4.1. Gain Ratio Value

Table 7. Feature Gain Ratio

Features	Gain Ratio
Maximum Arrears	0.3022
Loan Amount	0.2483
Outstanding Balance	0.2359
Remaining Tenor	0.2136

Table 7. Feature Gain Ratio shows that Maximum Arrears was selected as the root node with the highest gain ratio of 0.3022, indicating that this variable has the most significant influence on distinguishing risk levels. Significant contributions are also provided by Loan Amount (0.2483) and Outstanding Loan Balance (0.2359) as strong supporting variables, followed by Remaining Tenor with a value of 0.2136.

3.4.2. C4.5 Decision Tree

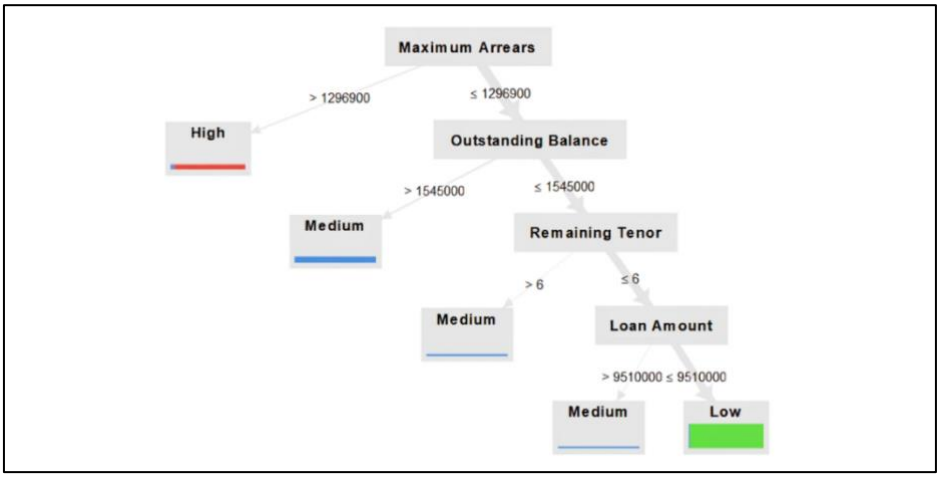

Figure 4. Decision Tree

Figure 4. Decision Tree generates a set of logical rules that form the knowledge base for classifying members' loan risks as follows:

1. Low Risk Rule: A member is classified as Low Risk only if they meet the comprehensive criteria of Maximum Arrears ≤ IDR 1.296.900, Outstanding Loan Balance ≤ IDR 1.545.000, Remaining Tenor ≤ 6 months, and Loan Amount ≤ IDR 9.510.000. This category has the largest data support with 258 members.
2. Medium Risk Rule, obtained through three different branching conditions, namely:
    1. Condition 1: If Maximum Arrears ≤ IDR 1.296.900 and Outstanding Loan Balance > IDR 1.545.000 (56 members).
    2. Condition 2: If Maximum Arrears ≤ IDR 1.296.900, Outstanding Loan Balance ≤ IDR 1.545.000, and Remaining Tenor > 6 months (8 members).
    3. Condition 3: If Maximum Arrears ≤ IDR 1.296.900, Outstanding Loan Balance ≤ IDR 1.545.000, Remaining Tenor ≤ 6 months, but Loan Amount > IDR 9.510.000 (6 members).
3. High Risk Rule: If the Maximum Arrears > IDR 1.296.900, the member is classified as High Risk. Based on the data, there are 42 members in this category, indicating that this Maximum Arrears is a critical indicator of potential loan problems.

This modeling provides the foundation for the C4.5 algorithm to generate decision rules whose performance will be tested during the model evaluation phase.

3.5. Model Evaluation

The evaluation phase is conducted to assess the extent to which the C4.5 model can accurately predict loan risk on previously unseen data (the test set).

3.5.1. Confusion Matrix

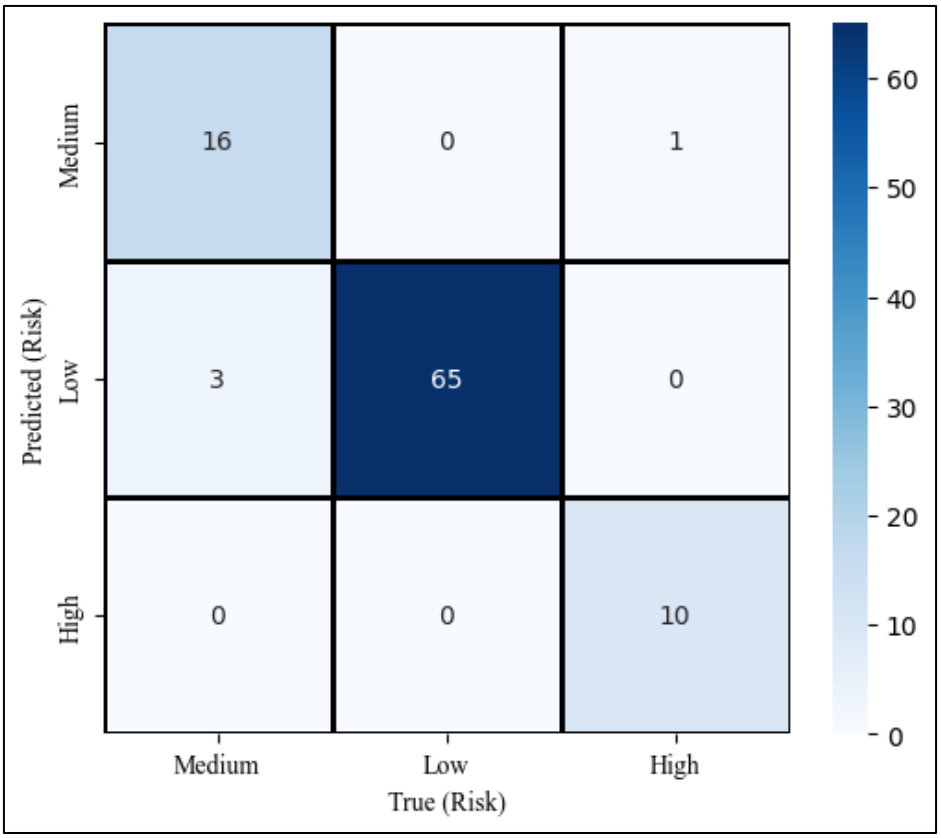

Figure 5. Confusion Matrix

Figure 5. Confusion Matrix shows the distribution of classification results for the test data used in this study. In the low-risk class, the model successfully predicted all 65 data points correctly without any classification errors. For the medium risk class, out of a total of 19 data points, the model correctly predicted 16, although 3 were classified as low risk. Meanwhile, in the high-risk class, the model correctly predicted 10 out of 11 data points, with the remaining 1 data point classified as medium risk.

3.5.2. Classification Report

Table 8. Classification Report

Risk Category	Precision	Recall	F1-Score
Low	0.96	1.00	0.98
Medium	0.94	0.84	0.89
High	1.00	0.91	0.95
Accuracy			**0.96**

Table 8. Classification Report shows that the model achieved an accuracy of 0.96, equivalent to 95.79%, indicating strong classification performance in identifying risk levels based on members' loan repayment patterns. This performance is influenced by the attributes Maximum Arrears and Outstanding Loan Balance which play a significant role in distinguishing between risk classes. The low-risk class has more easily recognizable loan repayment patterns because its payment characteristics are relatively consistent. Meanwhile, the medium-risk class has the lowest recall value because it lies in a transitional state between low and high risk. Consequently, some members exhibit similar loan repayment patterns, causing difficulties in the classification process. In the high-risk class, the model still demonstrates good classification ability because payment characteristics with higher Maximum Arrears values tend to be more easily distinguished from those of other classes. These results indicate that the model can help the credit union identify members based on their loan repayment risk levels more objectively.

The results of this study were compared with several previous studies that used clustering and classification methods in the analysis of cooperative loan risk. Previous research using K-Means was employed to group customers based on loan characteristics and delinquency into risk clusters without a classification stage [6]. Meanwhile, research using C4.5 was used to build a decision tree model for creditworthiness prediction, yielding an accuracy of 93% and an AUC of 0.898 [11]. Additionally, for the SVM, ANN, and Naïve Bayes methods, loan risk analysis was conducted to predict the likelihood of installment payment issues, where SVM achieved 92.0% accuracy, ANN 91.2%, and Naïve Bayes 81.2% through cross-validation, confusion matrix, and ROC curve evaluations [21]. Compared to that study, this research achieved an accuracy of 0.96 (95.79%) through the combination of K-Means and C4.5, indicating that initial clustering helps establish a clearer data structure before classification, allowing C4.5 to extract decision rules in a more structured manner from overlapping financial data and support complex loan risk analysis.

CONCLUSION

This study successfully achieved its objective of developing a model to predict member loan risk at Sri Hadi Dharma Savings and Loan Cooperative through the integration of the K-Means and C4.5 algorithms, which are capable of producing risk analyses that are more objective, structured, and consistent compared to previous manual approaches. This approach makes a tangible contribution to the application of data mining techniques to support data-driven decision-making and generates rules that are easily interpretable in the context of member loan management. The resulting model can also be used to predict risk for new members' data more consistently. Thus, the developed model can serve as a basis for more accurate and measurable considerations in reducing loan risks within the cooperative

environment. However, limitations still exist regarding the scope of data used in this study. Therefore, future research is recommended to include additional attribute variations and to develop the model's implementation into a digital-based system to enable more optimal utilization in the cooperative's operational activities.

REFERENCES

[1] L. Hasan and R. Delzy Perkasa, "PERAN KOPERASI SIMPAN PINJAM DALAM MEMBERDAYAKAN EKONOMI MASYARAKAT," *Jurnal Genta Mulia*, vol. 14, no. 1, Jan. 2023, doi: 10.61290/GM.V14I1.687.

[2] D. P. Ompusunggu, D. R. I. Sutrisno, and A. Hukom, "KONSISTENSI DAN EFEKTIVITAS PERAN LEMBAGA KEUANGAN NON BANK (KOPERASI SIMPAN PINJAM) SEBAGAI PENGGERAK PEREKONOMIAN INDONESIA," *Jurnal Cahaya Mandalika ISSN 2721-4796 (online)*, vol. 4, no. 1, pp. 689–696, Feb. 2023, doi: 10.36312/JCM.V4I1.1449.

[3] W. Wahyudin and R. Purnamasari, "Analisis Risiko Kredit Bermasalah terhadap Return On Equity (ROE)," *Coopetition : Jurnal Ilmiah Manajemen*, vol. 15, no. 2, pp. 305–316, Jun. 2024, doi: 10.32670/COOPETITION.V15I2.4392.

[4] Roviani, D. Supriadi, and I. Dzulfiqar Iskandar, "PREDICTION OF COOPERATIVE LOAN FEASIBILITY USING THE K-NEAREST NEIGHBOR ALGORITHM," *Jurnal Pilar Nusa Mandiri*, vol. 17, no. 1, pp. 39–46, Mar. 2021, doi: 10.33480/PILAR.V17I1.2183.

[5] A. Rifqi and R. T. Aldisa, "Penerapan Data Mining dalam Implementasi Algoritma K-Means Clustering untuk Pelanggan Potensial pada Koperasi Simpan Pinjam," *Building of Informatics, Technology and Science (BITS)*, vol. 5, no. 2, pp. 486−497-486−497, Sep. 2023, doi: 10.47065/BITS.V5I2.4278.

[6] Y. O. Tamba and N. Ekawati, "KLASTERISASI PINJAMAN NASABAH KOPERASI MENGGUNAKAN ALGORITMA K-MEANS," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 8, no. 6, pp. 11107–11114, Nov. 2024, doi: 10.36040/JATI.V8I6.11195.

[7] M. Alfarisi and M. S. Ramadhan, "IMPLEMENTASI K-MEANS CLUSTERING UNTUK MENGIDENTIFIKASI CALON NASABAH POTENSIAL PADA LEMBAGA KEUANGAN KECIL DAN MENENGAH," *JOURNAL OF SCIENCE AND SOCIAL RESEARCH*, vol. 8, no. 3, pp. 3807–3813, Aug. 2025, doi: 10.54314/JSSR.V8I3.4094.

[8] F. D. S. Alhamdani, A. A. Dianti, and Y. Azhar, "Segmentasi Pelanggan Berdasarkan Perilaku Penggunaan Kartu Kredit Menggunakan Metode K-Means Clustering," *JISKA (Jurnal Informatika Sunan Kalijaga)*, vol. 6, no. 2, pp. 70–77, May 2021, doi: 10.14421/jiska.2021.6.2.70-77.

[9] A. Fikri, B. F. Hutabarat, and U. Khaira, "Komparasi K-Means Clustering Dan Complete Linkage Dalam Pengelompokan Penyaluran Pinjaman Financial Technology," *Jurnal Ilmiah Media Sisfo*, vol. 17, no. 2, pp. 228–239, Oct. 2023, doi: 10.33998/MEDIASISFO.2023.17.2.1373.

[10] J. S. Parapat and A. S. RMS, "Data Mining Klasifikasi Data Nasabah Kredit KSU Taman Mandiri Menggunakan Algoritma C4.5," *Jurnal Ilmiah Teknik Elektro Komputer dan Informatika*, vol. 4, no. 2, pp. 144–154, Jan. 2018, doi: 10.26555/JITEKI.V4I2.11416.

[11] D. Puspita, S. Aminah, and A. Arif, "Prediction System for Credit Eligibility Using C4.5 Algorithm," *JOURNAL OF INFORMATICS AND TELECOMMUNICATION ENGINEERING*, vol. 6, no. 1, pp. 148–156, Jul. 2022, doi: 10.31289/JITE.V6I1.7311.

[12] Y. I. Kurniawan, A. Fatikasari, M. L. Hidayat, and M. Waluyo, "PREDICTION FOR COOPERATIVE CREDIT ELIGIBILITY USING DATA MINING CLASSIFICATION WITH C4.5 ALGORITHM," *Jurnal Teknik Informatika (Jutif)*, vol. 2, no. 2, pp. 67–74, Mar. 2021, doi: 10.20884/1.JUTIF.2021.2.2.49.

[13] Sunarti, "Klasifikasi Penentuan Kelayakan Pemberian Pinjaman pada Koperasi Karyawan Menggunakan Algoritma C4.5," *JOINS (Journal of Information System)*, vol. 6, no. 1, pp. 1–8, May 2021, doi: 10.33633/JOINS.V6I1.4330.

[14] A. Ramadhanu, S. Defit, and S. W. Wahhab Kareem, "Hybrid Data Mining with the Combination of K-Means Algorithm and C4.5 to Predict Student Achievement," *International Journal Of Artificial Intelegence Research*, vol. 5, 2021, doi: 10.29099/ijair.v6i1.225.

[15] H. Manurung, "PENGELOMPOKAN UMKM KOTA BINJAI MENGGUNAKAN METODE CLUSTERING K-MEANS UNTUK MENGIDENTIFIKASI POLA PERKEMBANGAN BISNIS," *Jurnal Informatika Kaputama (JIK)*, vol. 8, no. 2, pp. 93–99, Jul. 2024, doi: 10.59697/JIK.V8I2.844.

[16] S. Ocktavia and W. T. Atmojo, "ANALISIS SEGMENTASI PELANGGAN PEMBIAYAAN BERDASARKAN DEMOGRAFI UNTUK MEMPREDIKSI TINGKAT KREDIT MENGGUNAKAN

ALGORITMA K-MEANS," *Jurnal Informatika Teknologi dan Sains (Jinteks)*, vol. 7, no. 1, pp. 394–402, Mar. 2025, doi: 10.51401/JINTEKS.V7I1.5582.

[17] M. Rifqi Mubarok, A. Tri, J. Harjanto, and R. Renaldy, "PENINGKATAN PERFORMA DBSCAN DENGAN REDUKSI DIMENSI PRINCIPAL COMPONENT ANALYSIS (PCA) DALAM KLASTERISASI TINGKAT KEMISKINAN DI INDONESIA," *Jurnal Informatika Teknologi dan Sains (Jinteks)*, vol. 7, no. 3, pp. 1176–1184, Aug. 2025, doi: 10.51401/JINTEKS.V7I3.6129.

[18] S. Amri, M. Alharis, E. Winarno, and A. N. Putri, "Average Gain Ratio and Correlation-Based Feature Selection for Imprecise Classification in Algorithm C4.5," *Engineering, Technology & Applied Science Research*, vol. 15, no. 4, pp. 25275–25279, Aug. 2025, doi: 10.48084/ETASR.10165.

[19] R. Rajunaidi, H. Yuliansyah, S. Sunardi, and M. Murinto, "PREDICTING LOAN ELIGIBILITY WITH SUPPORT VECTOR MACHINE: A MACHINE LEARNING APPROACH," *JURTEKSI (jurnal Teknologi dan Sistem Informasi)*, vol. 11, no. 3, pp. 501–508, Jun. 2025, doi: 10.33330/JURTEKSI.V11I3.3876.

[20] A. Amrin, A. D. Kuswanto, and A. Fauzi, "Optimasi Kinerja Decision Tree C4.5 dengan Metode Backward Elimination pada Sistem Penilaian Kredit," *IMTechno: Journal of Industrial Management and Technology*, vol. 7, no. 1, pp. 1–5, Dec. 2026, doi: 10.31294/IMTECHNO.V7I1.11087.

[21] B. Pernama, H. Dwi Purnomo, and K. Satya Wacana, "Analisis Risiko Pinjaman dengan Metode Support Vector Machine, Artificial Neural Network dan Naïve Bayes," *Jurnal JTIK (Jurnal Teknologi Informasi dan Komunikasi)*, vol. 7, no. 1, pp. 92–99, Jan. 2023, doi: 10.35870/JTIK.V7I1.693.