International Journal of Informatics and Computation (IJICOM)
Vol. 8, No. 1, 2026
ISSN: 2685-8711, E-ISSN: 2714-5263 DOI: 10.35842/ijicom

Intelligent Segmentation of Cooperative Members Using a Hybrid Clustering Approach

I Wayan Sudiantara[1], Ni Wayan Sumartini Saraswati[2], I Putu Agus Eka Darma Udayana[3]

Abstract

Member management in savings and loan cooperatives faces challenges due to heterogeneous member characteristics and increasing credit risk. Member segmentation is essential for data-driven decision-making; however, the mixed nature of cooperative data, which consist of both numerical and categorical attributes, limits the effectiveness of single clustering methods. This study proposes a hybrid clustering framework that sequentially integrates K-Means and K-Modes to generate more comprehensive and interpretable member segments. K-Means is first applied to identify patterns in numerical attributes, whereas the resulting cluster labels are incorporated into K-Modes to enhance the clustering of categorical attributes. The optimal number of clusters was determined using the Elbow Method and Silhouette Analysis. A case study was conducted using 3,216 member records from a savings and loan cooperative, containing demographic, financial, and transactional characteristics. The experimental results indicate that K-Means produced three optimal clusters with a silhouette score of 0.237, while the hybrid framework generated four final member segments with clearer operational interpretations: loyal high-value members, low-risk growth members, seasonal agribusiness members, and economically vulnerable members. The findings demonstrate that the proposed hybrid clustering approach provides more comprehensive and actionable segmentation than single-method clustering, thereby supporting cooperative credit risk management, service development, and data-driven decision-making.

Keywords:
Clustering, K-Means, K-Modes, Machine Learning, Savings and Loan

1. Introduction

Savings and loan cooperatives play an essential role in expanding financial inclusion by providing savings, credit, and other financial services to communities that often have limited access to formal banking institutions. As cooperative membership continues to grow, managing members effectively has become increasingly challenging because individuals exhibit diverse financial behaviors, including differences in savings capacity, borrowing patterns, repayment performance, credit risk, and service utilization. These heterogeneous characteristics make it difficult for cooperatives to apply uniform management strategies without compromising service quality or increasing operational risk. Consequently, member segmentation has emerged as an important analytical approach for identifying groups of members with similar characteristics, enabling cooperatives to formulate more targeted lending policies, personalized services, and effective credit risk management strategies [1].

Despite its importance, member segmentation in savings and loan cooperatives remains a challenging task because cooperative databases typically contain heterogeneous data consisting of both numerical and categorical attributes. Numerical

Corresponding Author: Ni Wayan Sumartini Saraswati (sumartini.saraswati@gmail.com)
1 I Wayan Sudiantara, Institut Bisnis dan Teknologi Indonesia, bimbim.sudiantara@student.instiki.ac.id
2 Ni Wayan Sumartini Saraswati, Institut Bisnis dan Teknologi Indonesia, sumartini.saraswati@gmail.com
3 I Putu Agus Eka Darma Udayana, Institut Bisnis dan Teknologi Indonesia, agus.ekadarma@gmail.com

variables such as savings balance, loan amount, and repayment history coexist with categorical attributes including occupation, membership status, education level, and loan category. This mixed-data structure complicates the application of conventional clustering techniques, which are generally designed for a single data type. Algorithms such as K-Means efficiently partition numerical data but cannot directly process categorical variables, whereas K-Modes effectively clusters categorical data but cannot adequately represent continuous financial attributes. As a result, relying on a single clustering algorithm often produces incomplete member representations and limits the practical value of the resulting segmentation for cooperative management.

The need for effective member segmentation has become more urgent as savings and loan cooperatives face increasing competition from commercial banks and digital financial service providers. Sustainable cooperative performance depends not only on attracting new members but also on maintaining healthy loan portfolios and minimizing credit risk through informed decision-making [2]. Traditional member management approaches that rely on manual assessment or simple demographic grouping are increasingly insufficient for capturing complex behavioral patterns embedded within large transaction datasets. Consequently, data-driven analytical methods and artificial intelligence techniques have gained significant attention as tools for supporting objective decision-making in customer profiling, credit evaluation, and service optimization [3]. Accurate member segmentation enables cooperatives to identify high-value members, recognize potential default risks at an early stage, and develop differentiated financial products that better satisfy the needs of distinct member groups.

To address these challenges, this study proposes a hybrid clustering framework that integrates K-Means and K-Modes sequentially for segmenting savings and loan cooperative members based on mixed-type data. In the first stage, K-Means partitions members according to numerical financial attributes because of its computational efficiency and ability to minimize intra-cluster variance while handling large datasets effectively [4]–[6]. The clustering results obtained from K-Means are subsequently incorporated as an additional feature in the K-Modes algorithm, which performs clustering using categorical member attributes through frequency-based dissimilarity measures and mode-based cluster representations [7], [8]. Furthermore, the optimal number of clusters is determined using the Elbow Method and Silhouette Analysis to achieve a balance between cluster compactness and inter-cluster separation, thereby producing stable and meaningful segmentation results [9]. This sequential integration enables the proposed framework to simultaneously capture financial behavior patterns and socioeconomic characteristics that cannot be represented adequately by a single clustering technique.

The primary contribution of this study lies in the development of a sequential hybrid clustering framework that combines K-Means and K-Modes to segment savings and loan cooperative members using mixed numerical and categorical data within a unified analytical process. Unlike many previous clustering studies that primarily emphasize algorithmic performance or statistical evaluation, this research focuses on producing interpretable member segments that directly support operational decision-making in cooperative management [10]. The resulting clusters are evaluated using cluster quality metrics and subsequently interpreted as practical member categories that can guide lending strategies, credit risk mitigation, product development, and customer relationship management. This practical orientation enhances the usefulness of clustering outcomes for cooperative managers and other non-technical stakeholders.

In addition to its immediate operational benefits, the proposed segmentation framework provides a strong foundation for future artificial intelligence applications in cooperative financial services. The rapid adoption of Generative Artificial Intelligence (GenAI) and intelligent decision-support systems has increased the need for structured, reliable, and interpretable customer representations before advanced recommendation models can be deployed effectively. Since cooperative member databases consist of heterogeneous mixed-type attributes, meaningful segmentation remains an essential preprocessing stage for higher-level AI applications. By integrating K-Means and K-Modes, the proposed approach generates representative member profiles that can support future developments in personalized financial recommendations, intelligent credit assessment, predictive analytics, and AI-driven cooperative management systems.

2. Related Works

Clustering has been widely adopted as an unsupervised learning technique for segmentation and decision support across various application domains. Among existing clustering algorithms, K-Means remains one of the most frequently utilized methods because of its computational efficiency, scalability, and capability to discover hidden structures within numerical datasets. Previous studies have integrated K-Means with supervised learning to improve cluster quality by correcting misclassified samples [11]. Other researchers have employed K-Means to segment e-commerce customers according to transaction behavior for profitability analysis [12], classify provinces based on COVID-19 transmission risk to support public policy formulation [13], and compare its performance with Agglomerative Clustering for financial literacy segmentation among micro, small, and medium enterprises (MSMEs) [14]. Collectively, these studies demonstrate that K-Means effectively identifies homogeneous groups from numerical attributes and provides valuable support for data-driven decision-making.

Although K-Means performs well on continuous numerical variables, its application is limited when datasets contain categorical attributes. To address this limitation, K-Modes was introduced as an extension specifically designed for categorical data through frequency-based dissimilarity measures and mode-based cluster centroids. Previous studies have successfully applied K-Modes to associative classification problems [15], customer segmentation enhanced with boosting techniques [16], urban travel preference analysis [17], and mortality pattern identification based on demographic characteristics [18]. These findings indicate that K-Modes provides more meaningful cluster representations for categorical variables than conventional distance-based clustering algorithms, making it suitable for applications involving qualitative data.

Customer segmentation has also received considerable attention within the financial sector, where clustering techniques are used to improve customer relationship management and credit risk assessment. Several studies have integrated K-Means with the Recency-Frequency-Monetary (RFM) model to classify customer value [19], employed hierarchical clustering to identify potential bank failures [20], and proposed hybrid clustering methods combining K-Means with Density-Based Spatial Clustering of Applications with Noise (KM-DBSCAN) to improve segmentation accuracy and stability [21]. Furthermore, clustering has been utilized to support the development of green financial products by identifying customer groups with distinct financial behaviors [22]. These studies confirm that clustering plays an important role in supporting strategic decisions within financial institutions. However, most existing research has focused primarily on commercial banking environments rather than cooperative-based financial organizations.

Despite the extensive use of clustering methods, several limitations remain. Most previous studies have relied on a single clustering algorithm that processes either numerical or categorical attributes independently, resulting in incomplete representations of mixed-type datasets. Consequently, the generated clusters may fail to capture the complex relationships between financial behavior and demographic characteristics that coexist in cooperative member data. In addition, existing studies generally emphasize algorithmic performance while providing limited interpretation of clustering results for practical cooperative management. Research specifically addressing member segmentation in savings and loan cooperatives also remains relatively scarce, despite the unique operational characteristics and socioeconomic objectives of these institutions. These limitations motivate the development of a hybrid clustering framework that sequentially integrates K-Means and K-Modes to simultaneously exploit numerical and categorical information, producing more representative and operationally meaningful member segments for data-driven cooperative management.

3. Proposed Method

3.1 Research Framework

This study applied a hybrid clustering framework to segment savings and loan cooperative members based on mixed data consisting of numerical and categorical attributes of the members. The research process consisted of several stages, including data collection, data cleaning, preprocessing, cluster evaluation using the Elbow Method and Silhouette Analysis, hybrid clustering implementation, visualization, and cluster insight analysis. The overall research framework is shown in Fig. 1.

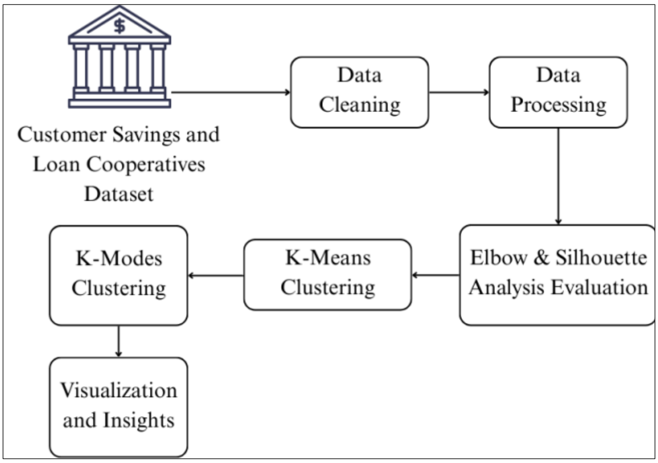


Fig. 1. Research flow

3.2 Hybrid Clustering Method

This This study proposes a sequential hybrid clustering framework that integrates K-Means and K-Modes to segment savings and loan cooperative members using mixed numerical and categorical data. K-Means is first applied to identify financial behavior patterns from numerical attributes, while the resulting cluster labels are incorporated into K-Modes to simultaneously analyze categorical characteristics. This integration overcomes the limitations of using either algorithm independently and produces more representative and interpretable member segments. Consequently, the proposed approach supports data-driven decision-making by enabling more effective credit risk management, personalized services, and member development strategies in savings and loan cooperatives [23], [24]. This algorithm works by randomly determining the number of cluster centers (centroids) and then grouping the data to the nearest centroid based on the Euclidean distance described in Equation (1). The Euclidean distance between a data point and a cluster centroid is calculated as follows:

$$d(x_i\mu_j) = \sqrt{\sum_{q=1}^{p}(x_{iq} - \mu_{jq})^2} \qquad (1)$$

where x_i denotes the i-th member record, μ_j represents the centroid of cluster, x_{iq} is the value of the q-th, numerical attribute of observation i, μ_{jq} is the centroid value of the q-th attribute, and p is the total number of numerical attributes. Smaller distance values indicate a higher similarity between a member and cluster centroid. During the clustering process, each member is assigned to the nearest centroid based on the distance measure..

This process is repeated until the centroid position does not change significantly or until it reaches the maximum number of iterations. The advantages of K-Means clustering are its efficiency and ability to handle large datasets; however, this algorithm is sensitive to outliers and data scale. The resulting K-Means clusters were subsequently used as additional categorical features in the K-Modes clustering stage. K-Modes is a development of K-Means that is specifically designed to handle categorical data. The stages of K-Modes are similar to those of K-Modes. This method uses the mode (the highest value) as the cluster center, where the measure of closeness between the data is calculated based on the number of mismatched categories (mismatch dissimilarity) [27,28]. The dissimilarity between two categorical observations is measured using the simple matching dissimilarity function (2) and (3):

$$d(X,Y) = \sum_{i=1}^{m} \delta(x_i, y_i) \tag{2}$$

$$\delta(x_i, y_i) = \begin{cases} 0, & x_i = y_i \\ 1, & x_i \neq y_i \end{cases} \tag{3}$$

where m enotes the number of categorical attributes. A value of 0 indicates that the two categorical values are identical, whereas a value of 1 indicates a mismatch. The total dissimilarity increases as the number of mismatched categories increases.

Unlike K-Means, which updates cluster centers using arithmetic means, K-Modes updates cluster centers using the mode of each categorical attribute. This mechanism makes K-Modes suitable for representing the dominant categorical characteristics within each cooperative member segment.

3.3 Cluster Evaluation

The optimal number of clusters was determined using two approaches: the Elbow method and Silhouette analysis. Both are used to identify the most optimal clusters [27]. The Elbow method calculates the Within-Cluster Sum of Squares (WCSS) for various values of k (the number of clusters) and then plots the results graphically. The point at which the decrease in WCSS begins to slow significantly is considered the optimal point (i.e., the elbow point). This indicates that increasing the number of clusters after this point does not significantly increase the cluster compactness. The WCSS is mathematically formulated as Equation (4):

$$WCSS = \sum_{j=1}^{K} \sum_{x_i \in C_j} \left\| x_i - \mu_j \right\|^2 \tag{4}$$

Where K is the number of clusters, C_j is the set of observations belonging to cluster j, and is the centroid of cluster j. WCSS measures the total variation in observations within each cluster. Lower WCSS values indicate more compact clusters, meaning that observations assigned to the same cluster are closer to their centroid. In this study, WCSS was used as the basis for the Elbow Method to determine the optimal number of clusters.

Silhouette analysis measures how well a data point fits within its cluster compared to the other clusters. The Silhouette value ranges from -1 to 1, where values closer to 1 indicate that the data are in the correct cluster. This analysis provides additional insight into determining the optimal K in terms of the separation between the clusters. The Silhouette value for each data point i is calculated using (5), where a(i) is the average distance of data i to all other data in the same cluster (intra-cluster distance) and b(i) is the average of the lowest distance between data i and all data in the nearest other cluster (nearest-cluster distance).

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \quad (5)$$

To illustrate the results of this silhouette analysis, a silhouette plot visualization was used for each data point in each cluster. This plot provides an overview of how well each data point is clustered and whether there is a potential overlap between clusters. To implement the Elbow method and Silhouette analysis, various numbers of clusters were used, from K=2 to K=7.

4. Experimental Setup

4.1 Data Collection

The data in this study were collected from members of the Guna Prima Dana Savings and Loans Cooperative in Ungasan, Bali, from 2002 to April 2025. The dataset comprised 3,216 rows with 11 features, including categorical and numeric features. The features used capture the demographic, financial, and transactional characteristics of cooperative members. Therefore, the selection of these features was based on their relevance in representing key aspects of grouping cooperative members, namely, repayment capacity, borrowing behavior, and risk levels. A sample dataset is shown in Fig. 2.

Customer ID	Date of Birth	Monthly Income	Loan Amoount	Interest Rate	Business Type	Occupation	Collateral Type	Loan Status	Membership Start Date	Loan Frequency
17-00074	10/4/1971	4,000,000	40,000,000	10	Restaurant and Food Services	Others	Deposit Account	Current	14/02/2013	1
17-00075	17/08/1989	6,666,667	117,600,000	10	Accommodation Services	Self-employed	Vehicle	Current	25/05/2013	1
17-00078	14/04/1980	14,250,000	1,400,000,000	12	Retail Trade (Food, Beverage, Tobacco Stores)	Others	Deposit Account, Land Certificate, Member Account	Current	30/12/2013	1
17-00079	15/04/1972	5,000,000	100,000,000	10	Personal Services	Self-employed	Deposit Account	Current	26/06/2015	1
17-00082	26/06/1980	6,166,667	50,000,000	11	Primary & Early Childhood Education	Teacher/Lecturer	Vehicle, Member Account, Deposit Account	Current	9/11/2015	1

Fig. 2. Sampel dataset

4.2 Data Cleaning

Before clustering, the dataset underwent a data cleaning process to improve clustering accuracy and reduce bias caused by missing values, duplicate records, and anomalous inputs. Missing values were handled through deletion or rule-based imputation depending on their proportion, while duplicate records were removed to preserve unique observations. Numerical anomalies, such as unrealistic monthly income values, were corrected using median imputation, whereas missing categorical attributes, including loan status, occupation, and collateral type, were assigned the category "Other." Additional validation ensured logical consistency by enforcing a minimum membership age of 17 years and verifying that membership dates were not earlier than the cooperative's establishment date. These preprocessing steps improved data quality and enhanced the reliability of the subsequent clustering analysis [18][19][30].

4.3 Data Preprocessing

The preprocessing stage standardized the dataset through date formatting, numeric type conversion, and consistent categorical labeling to ensure data uniformity. Feature engineering was then performed by deriving the Age and Membership Duration features from temporal attributes, while redundant features, including Customer Code, Date of Birth, and Membership Start Date, were removed. To improve clustering reliability, numerical outliers were treated using Winsorization at the 1st and 99th percentiles, and skewed numerical features were transformed using the log1p function to produce a more stable distribution. These preprocessing steps enhanced data quality and improved the robustness of the subsequent clustering analysis. [21,22,12], [33], [34].

For the application of the K-Means and K-Modes algorithms, feature selection was performed based on the appropriate data type. This separation ensured that each algorithm performed optimally according to the characteristics of the data. The K-Modes feature is supplemented with the clustering results from K-Means. The features are listed in Table 1.

Table 1. Feature engineering

Methods	Features
K-Means	Monthly Income
	Loan Amount
	Interest Rate
	Loan Frequency
	Age
	Membership Duration
K-Modes	Business Type
	Occupation
	Loan Status
	Collateral Type
	K-Means Cluster

The next stage was data scaling for each of the numeric features. Scaling aims to transform the data into the same scale. This was done to avoid the dominance of certain features with large value ranges [35]. In this study, the StandardScaler was implemented for scaling.

5. Result and Analysis

5.1 K-Means

The results of the elbow method are shown in Fig. 3. A substantial decrease in inertia was observed when the number of clusters increased from k = 2 to k = 3. Beyond this point, the reduction in inertia became more gradual, forming a clear elbow at k = 3. This pattern indicates diminishing marginal improvement with additional clusters, whereas the computational complexity continues to increase. Therefore, k = 3 was selected as the optimal number of clusters. This configuration provides a better balance between intra-cluster homogeneity and inter-cluster separation than other values of k.

In addition to using the elbow, the results of the silhouette analysis were used to strengthen the results of the elbow analysis, as shown in the silhouette score and silhouette plot values. The silhouette scores are presented in Table 2, and a silhouette plot is shown in Fig. 4. Based on the computational results, the highest silhouette score was obtained for the k=3 configuration, with a score of 0.2371. This quantitatively confirmed that dividing the data into three groups provided the most optimal separation structure compared to other configurations.

Visual analysis using a silhouette plot provides further insights into cluster quality. At k=3, most samples in each cluster (0, 1, and 2) had positive silhouette coefficients and exceeded the mean (dotted red line). Although there are a few negative values in Cluster 2, indicating the presence of some outliers or data on the border between clusters, the visual width of each silhouette "blade" indicates a fairly proportional distribution of members. Conversely, at higher k values (such as k=4 or k=7), several clusters were very thin and had scores below the mean, indicating over-clustering. Therefore, based on the elbow and silhouette analyses, k=3 was used as the optimal cluster to enter the clustering model.

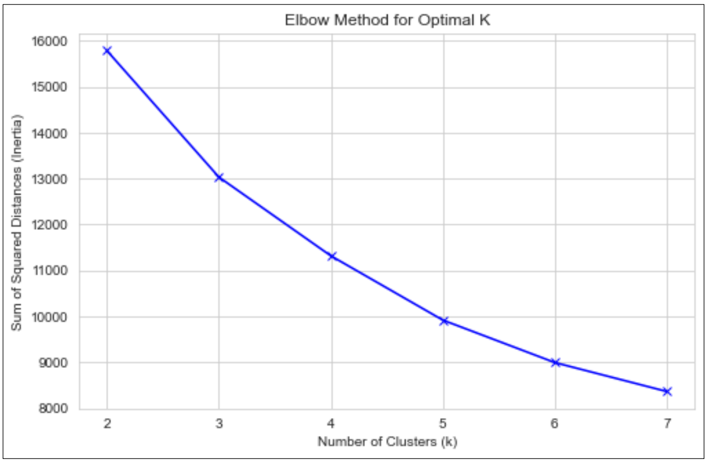


Fig. 3. Elbow method

Table 2. Silhouette score results

Cluster	Silhouette Score
2	0.23069298066402205
3	0.23708960547086105
4	0.20411916530083024
5	0.21863228056776546
6	0.23214163477339733
7	0.2167459943223294

The model results were visualized before being analyzed to determine the characteristics of the cooperative members, as shown in Fig. 5, as well as the cluster characteristics in Table 3.

Table 3. Cluster profile

Cluster	Segment Profile	Key Characteristics (Feature-Based)
0	Loyal and Potential Customers	Large loan limits, longer membership duration, moderate interest rates, and moderate risk levels.
1	New and Young Customers	Relatively younger members with shorter tenure, low risk, and low interest rates.
2	Medium-Risk and Profitable Customers	Smaller loan limits but higher interest rates, associated with higher risk and potentially greater profitability.

Cluster 0 was identified as the Loyal and Potential Customer group. This group has the most established profile of the three clusters, characterized by a significantly higher median loan amount than that of the other two groups. In terms of loyalty, this cluster is dominated by members with long-standing membership (loyalty), indicating historically cultivated transactional trust. Despite their substantial loan exposure, this group is charged moderate interest rates with controlled risk levels. This profile reflects established corporate or individual members who are key contributors to the stability of the cooperative's credit portfolio, where large loan amounts are balanced by a solid track record of loyalty.

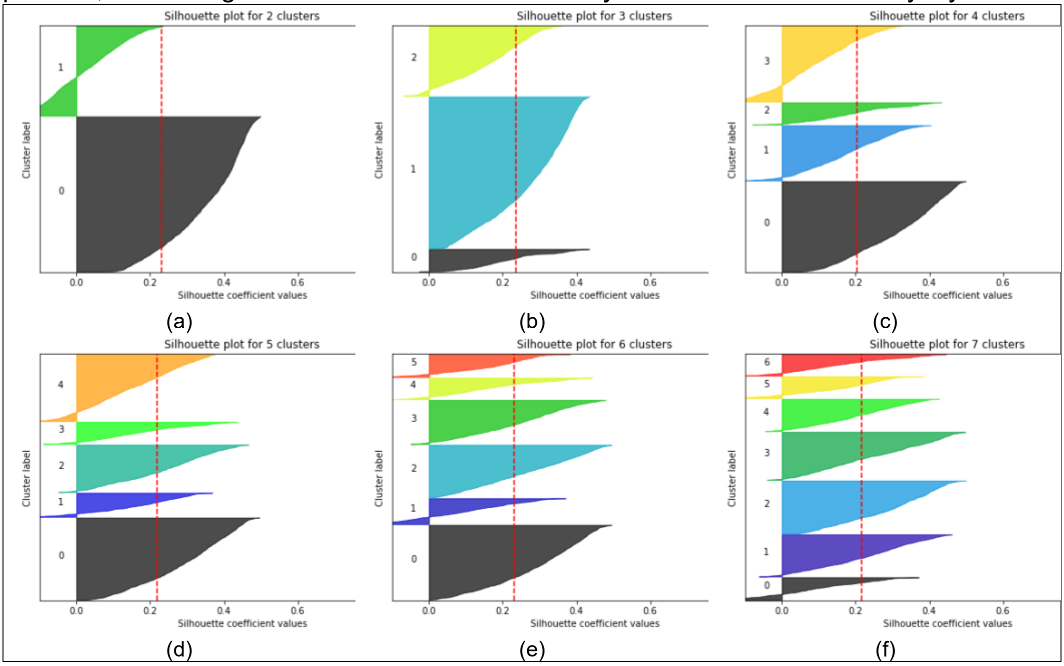
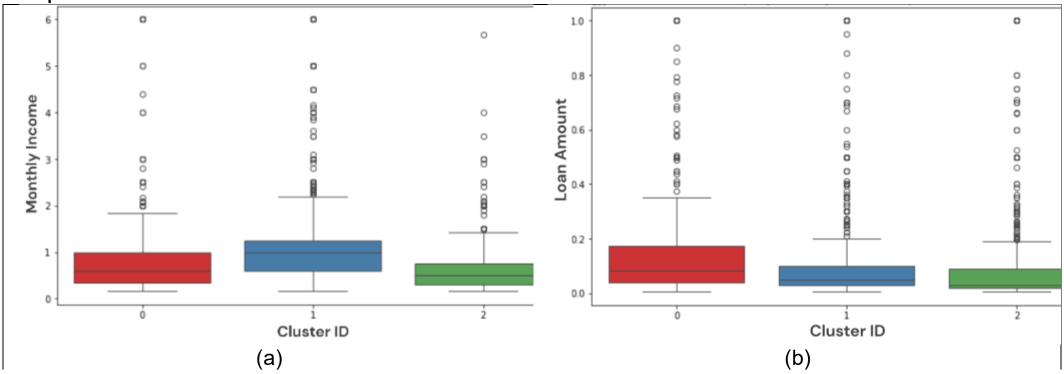


Fig. 4. Silhouette plots for different numbers of clusters (k = 2–7). The plots are arranged sequentially from left to right and top to bottom: (a) k = 2, (b) k = 3, (c) k = 4, (d) k = 5, (e) k = 6, and (f) k = 7.

Cluster 1 is categorized as the New and Young Customer group. Visually, the boxplot shows that members of this cluster are the youngest and have had a very short membership period. Their transactional characteristics tend to be conservative, as evidenced by small loan ceilings and low-interest rates. This indicates that this group is in the early stages of the credit cycle, where institutions implement low-risk policies for members with limited credit histories.

Cluster 2 is defined as a Medium-Risk and Profitable Customers group. Interestingly, although this group has a smaller volume (loan ceiling) than Cluster 0, it contributes significantly to profitability per unit, owing to the highest interest rates. In terms of risk, the boxplot shows that this cluster has higher volatility and risk scores than the other two clusters. This group reflects members willing to take on higher interest burdens to gain access to funding, but the institution compensates by providing a limited ceiling to mitigate the potential for default.



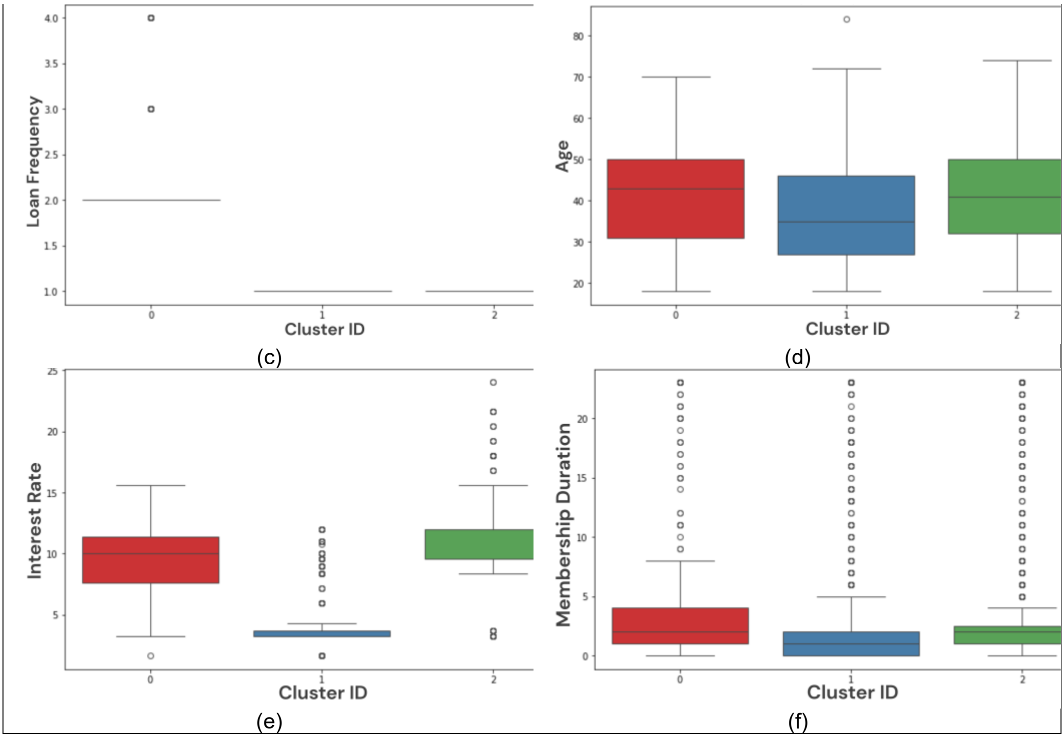


Fig. 5. K-Means clustering visualization based on different variables: (a) monthly income, (b) loan amount, (c) loan frequency, (d) age, (e) interest rate, and (f) membership duration.

5.2 K-Modes

Based on the K-Means results, the three clusters were implemented as additional features of the K-Modes model. Fig. 6 shows the results of the elbow method, along with the silhouette scores listed in Table 4. The elbow analysis is based on changes in the cost or dissimilarity values generated by the K-Modes model. The elbow results indicate a significant decrease in the cost values when the number of clusters increases to k = 4. After this value, the rate of cost reduction tends to plateau, indicating that increasing the number of clusters above k=4 no longer significantly improves the cluster structure but only increases the model complexity. This pattern indicates the formation of an elbow point at k=4, representing a balance between the clustering quality and model efficiency.

Furthermore, the silhouette score was used to strengthen the elbow analysis (Table 4). The silhouette score calculation results show that the silhouette score increases significantly up to k=4 and remains at a relatively high level for larger k configurations. Although the highest silhouette value was obtained at k=6, this increase was accompanied by the formation of an increasingly fragmented cluster structure, potentially reducing the interpretability of the results in the context of operational decision-making in savings and loan cooperatives. Considering the elbow results that showed an elbow point at k=4 and the need for stable and easily interpreted segmentation, this study determined k=4 as the optimal number of clusters at the K-Modes stage.

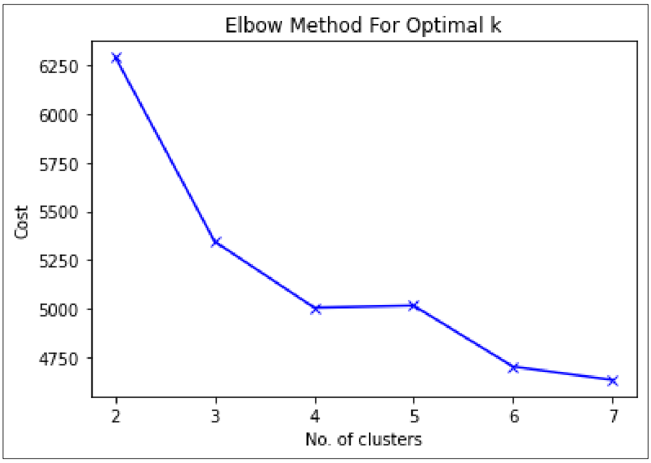


Fig 6. Elbow method

Table 4. Silhouette score results

Cluster	Silhouette Score
2	0.1378171156755118
3	0.3689690182030479
4	0.42532523289999197
5	0.4155245678083523
6	0.5289874218363304
7	0.45166531244313984

The results of the member grouping at the K-Modes stage are then visualized using a treemap for each categorical feature, including collateral type, K-Means cluster, business type, loan category, and occupation, as shown in Fig. 7–11. Treemap visualization was used to provide an overview of the proportion of category dominance in each cluster, so that differences in characteristics between clusters could be observed intuitively and comparatively.

Cluster 0 represents high-value members with established businesses, primarily in the hotel, restaurant, and personal service sectors. Most members belong to the medium-risk, profitable, and loyal customer groups identified by K-Means and commonly use vehicle collateral, indicating stronger financial stability and long-term business potential. Cluster 1 consists mainly of new, low-risk members engaged in personal service and accommodation businesses. Most members provide business certificates (SKU) as collateral, reflecting limited asset ownership and early-stage business development. This segment is well suited for targeted mentoring and capacity-building programs.

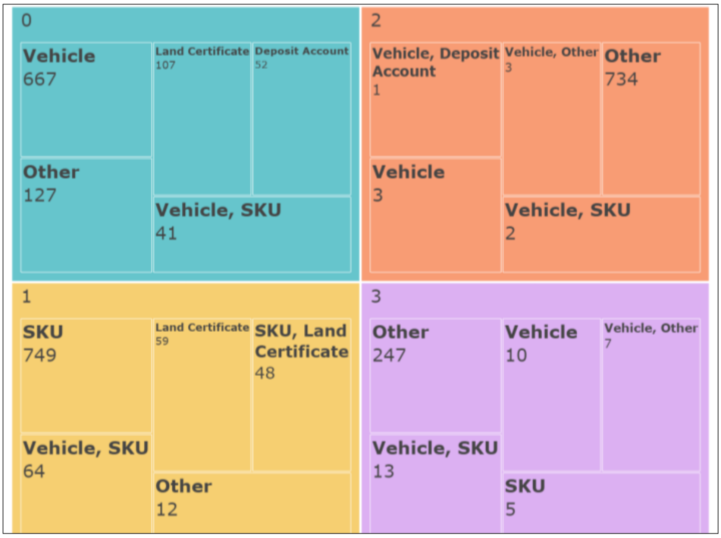


Fig. 7. Collateral type

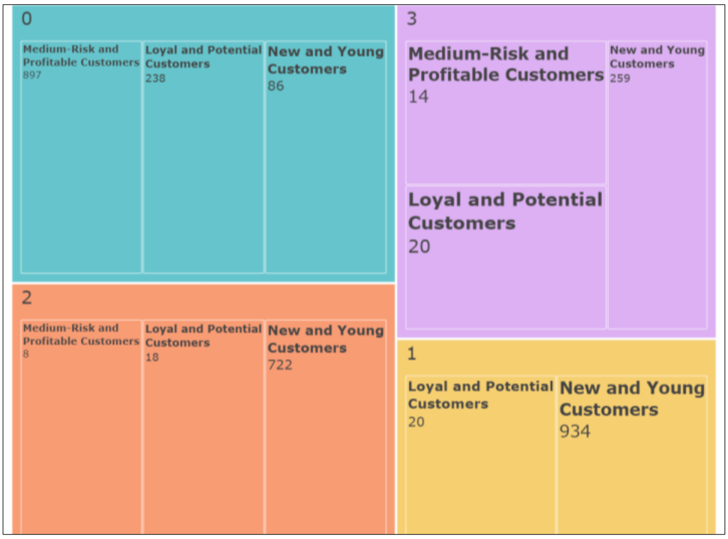


Fig. 8. K-Means Cluster

Cluster 2 exhibited distinct characteristics compared to the previous cluster, with businesses dominated by breeding, beef cattle farming, and pig farming. Members of this cluster are predominantly farmers, with collateral types falling into other categories, such as Integration with the K-Means analysis results, indicating that this cluster is also dominated by new and young customers with low-risk profiles. These characteristics reflect the primary sector-based member segment, which has a seasonal business model and is heavily influenced by external factors. This situation requires a more contextual approach to financing and risk management that reflects the characteristics of agribusinesses.

Fig. 9. Business type

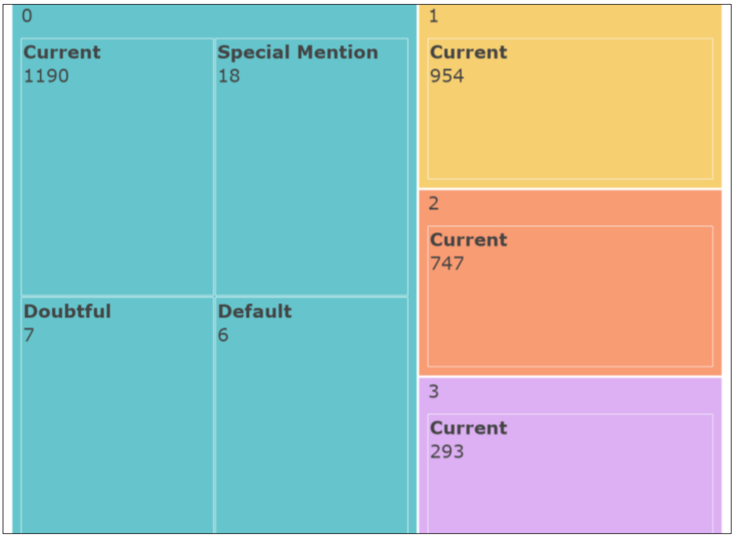


Fig. 10. Loan status

Cluster 3 is dominated by members of retail and pig farming businesses. In terms of occupation, this cluster is predominantly comprised of students, indicating the involvement of young people in small-scale economic activities in the region. The K-Means cluster mapping results indicate that this group also falls within the new and young customer segment with low-risk levels. The collateral types used mostly fall into other categories, indicating limited formal assets for these borrowers. This profile reflects a member segment with limited economic capacity but potential for growth, necessitating a careful, development-based financing approach.

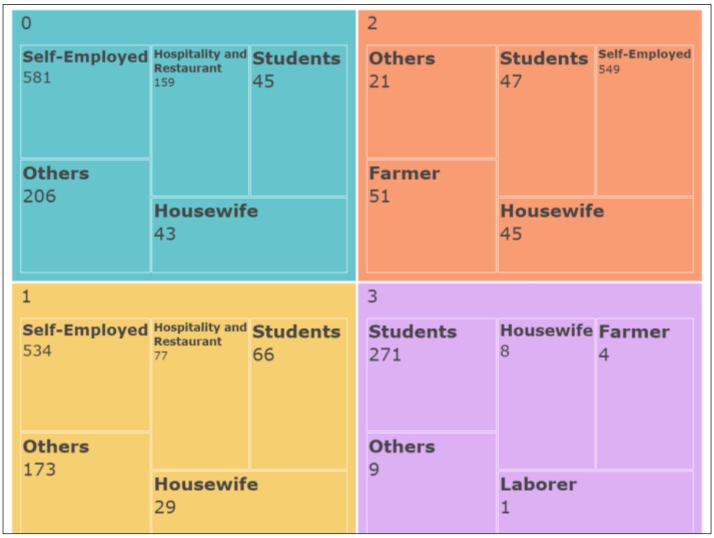


Fig. 11. Occupation

The proposed hybrid clustering framework provides important practical implications for savings and loan cooperatives by enabling data-driven member segmentation for differentiated financing and service strategies. The four identified clusters represent distinct levels of financial maturity, business characteristics, and credit risk, allowing cooperatives to allocate resources more effectively. Established and loyal members can be offered larger financing limits and retention programs, while newer members benefit from gradual financing, financial literacy initiatives, and business development support. Similarly, members engaged in agribusiness require flexible financing schemes aligned with seasonal business cycles, whereas higher-risk members should receive conservative credit exposure combined with regular monitoring. These findings demonstrate that member segmentation can improve credit risk management while supporting sustainable cooperative growth.

From a methodological perspective, the sequential integration of K-Means and K-Modes provides a more comprehensive representation of cooperative members than single-method clustering. K-Means effectively captures variations in numerical financial behavior, whereas K-Modes incorporates categorical socioeconomic characteristics to enrich cluster interpretation. The effectiveness of the proposed framework is demonstrated by the improvement in the Silhouette Score from 0.237 during the K-Means stage to 0.425 after the hybrid clustering process, indicating better cluster cohesion and separation. Unlike previous studies that analyzed numerical and categorical attributes independently, the proposed framework directly links financial behavior with demographic and socioeconomic profiles, producing clusters that are both statistically meaningful and operationally interpretable for cooperative decision-making.

Despite these promising results, several limitations should be acknowledged. The moderate Silhouette Score obtained during the K-Means stage suggests that the numerical data contain overlapping patterns that cannot be completely separated. In addition, the study was conducted using data from a single savings and loan cooperative, which may limit the generalizability of the findings to other institutions with different member characteristics. Future research should validate the proposed framework using larger multi-cooperative datasets and compare its performance with alternative mixed-data clustering algorithms, such as K-Prototypes, DBSCAN, and Gaussian Mixture Models, to further improve segmentation quality and practical applicability.

6. Conclusion

This study proposed a sequential hybrid clustering framework that integrates the K-Means and K-Modes algorithms to segment members of a savings and loan cooperative using mixed numerical and categorical data. The experimental results identified three optimal clusters in the K-Means stage, supported by an Elbow analysis and the highest Silhouette Score of 0.237, while the subsequent K-Modes stage produced four final member segments with an improved Silhouette Score of 0.425, indicating enhanced cluster cohesion and interpretability. The resulting segments comprise loyal high-value members, low-risk growth members, seasonal agribusiness members, and economically vulnerable members, each characterized by distinct financial behaviors, business activities, and socioeconomic profiles.

The proposed hybrid framework provides both methodological and practical contributions to cooperative member management. By combining K-Means and K-Modes sequentially, the framework captures latent financial behavior through numerical attributes while enriching cluster interpretation using categorical socioeconomic characteristics. The resulting segmentation enables cooperatives to implement differentiated financing strategies, including retention programs for loyal members, developmental financing for new members, flexible credit schemes for agribusiness sectors, and conservative lending policies for economically vulnerable members. Consequently, the proposed approach supports more effective credit risk management, targeted resource allocation, and data-driven service planning than conventional clustering methods that analyze numerical or categorical attributes independently.

Although the proposed framework demonstrated promising performance, several limitations remain. The moderate clustering quality obtained during the K-Means stage indicates that numerical member data exhibit partially overlapping patterns, while the evaluation was conducted using data from a single savings and loan cooperative, limiting the generalizability of the findings. Future studies should validate the framework using larger multi-cooperative datasets and investigate alternative mixed-data clustering algorithms, such as K-Prototypes, DBSCAN, Gaussian Mixture Models, and ensemble clustering methods. Furthermore, integrating longitudinal member data, explainable artificial intelligence, and predictive credit-scoring models could transform cooperative member segmentation from a descriptive analytical tool into a proactive decision-support system for sustainable cooperative management.

Acknowledgment

The authors would like to thank the Guna Prima Dana Savings and Loans Cooperative for providing the dataset and supporting this research.

References

[1] Y. Suh, "Discovering customer segments through interaction behaviors for home appliance business," *Journal of Big Data*, vol. 12, no. 57, 2025, doi: 10.1186/s40537-025-01111-y.

[2] V. Untari, E. Waty, A. M. Hidayat, and B. R. Kartawinata, "Strategies to Improve Cooperative Efficiency and Performance Through the Integration of Financial Inclusion , Risk Ratios , and Non-Performing Loans," *Enrichment : Journal of Multidisciplinary Research and Development*, vol. 3, no. 5, pp. 1196–1211, 2025, doi: https://doi.org/10.55324/enrichment.v3i5.415.

[3] M. Lai and T. Hariguna, "Analyzing Company Hiring Patterns Using K-Means Clustering and Association Rule Mining : A Data-Driven Approach to Understanding Recruitment Trends in the Digital Economy," *Journal of Digital Society*, vol. 1, no. 1, pp. 20–43, 2025, doi: https://doi.org/10.63913/jds.v1i1.24.

[4] S. M. Miraftabzadeh, "K-Means and Alternative Clustering Methods in Modern Power Systems," *IEEE Access*, vol. 11, no. September, pp. 119596–119633, 2023, doi:

10.1109/ACCESS.2023.3327640.
[5] H. Liu, J. Chen, J. Dy, and Y. Fu, "Transforming Complex Problems into K-means Solutions," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 7, pp. 1–50, 2023, doi: 10.1109/TPAMI.2023.3237667.Transforming.
[6] A. M. Ikotun and A. E. Ezugwu, "Boosting k-means clustering with symbiotic organisms search for automatic clustering problems," *PLOS ONE*, vol. 17, no. 8, 2022, doi: 10.1371/journal.pone.0272861.
[7] M. Cendana and R. Kuo, "Categorical Data Clustering : A Bibliometric Analysis and Taxonomy," *Machine Learning and Knowledge Extraction*, vol. 6, pp. 1009–1054, 2024, doi: https://doi.org/10.3390/make6020047.
[8] M. Yusran, S. Nuradilla, M. R. Putri, and A. Fitrianto, "Household Clustering in West Java Based on Stunting Risk Factors Using K-Modes and K-Prototypes Algorithms," *Journal of Applied Informatics and Computing (JAIC)*, vol. 9, no. 6, pp. 2986–2999, 2025.
[9] A. P. Dayati, "Matrix Factorization Using LightFM for a Music Recommendation System Based on Emotional and Listening Behavior Awareness," vol. 7, no. 2, 2025, doi: 10.35842/ijicom.
[10] G. S. de Oliveira, F. A. Silva, and R. V. Ferreira, "Explainable Clustering : A solution to interpret and describe," *Journal of Information and Data Management*, vol. 16, no. 1, pp. 170–180, 2025, doi: 10.5753/jidm.2025.4663.
[11] I. Borlea, R. Precup, and A. Borlea, "ScienceDirect Improvement of K-means Cluster Quality by Post Processing Improvement of K-means Cluster Quality by Post Processing Resulted Clusters Resulted Clusters," *Procedia Computer Science*, vol. 199, pp. 63–70, 2022, doi: 10.1016/j.procs.2022.01.009.
[12] K. Tabianan, S. Velu, and V. Ravi, "K-Means Clustering Approach for Intelligent Customer Segmentation Using Customer Purchase Behavior Data," *Sustainability*, vol. 14, no. 12, p. 7243, 2022, doi: https://doi.org/10.3390/su14127243.
[13] D. Abdullah, A. S. Ahmar, R. Rusli, and R. Hidayat, "The application of K-means clustering for province clustering in Indonesia of the risk of the COVID-19 pandemic based on COVID-19 data," *Quality & Quantity*, vol. 56, pp. 1283–1291, 2022, doi: https://doi.org/10.1007/s11135-021-01176-w.
[14] A. Murdiono, H. Pratikto, P. Handayati, A. Winarno, and F. Zen, "Analysis of MSME Financial Literacy Segmentation Using Clustering Methods," *JOIV: International Journal on Informatics Visualization*, vol. 10, no. March, pp. 900–908, 2026, doi: http://dx.doi.org/10.62527/joiv.10.2.4743.
[15] J. Mattiev, M. Davityan, and B. Kavsek, "ACMKC : A Compact Associative Classification Model Using K-Modes Clustering with Rule Representations by Coverage," *Mathematics*, vol. 18, no. 10, p. 3978, 2023, doi: https://doi.org/10.3390/math11183978.
[16] S. R. Niloy, T. M. Hasan, and S. Apu, "Customer Segmentation and Classification using K-Modes Clustering with Ensemble Learning," in *Intelligent Systems and Data Science*, 2024, pp. 3–18. doi: https://doi.org/10.1007/978-981-97-9613-7_1.
[17] M. Babapour and M. V. Corazza, "Urban Commuting Preferences in Italy : Employees ' Perceptions of Public Transport and Willingness to Adopt Active Transport Based on K-Modes Cluster Analysis," *Sustainability*, vol. 17, no. 11, p. 5149, 2025, doi: https://doi.org/10.3390/su17115149.
[18] A. Jimenez-Torres and H. N. Roa, "Unveiling Mortality Patterns in Ecuador: A K-Modes Clustering Approach to Categorical Data Analysis," in *ICT for Intelligent Systems*, 2024, pp. 117–128. doi: https://doi.org/10.1007/978-981-97-5799-2_11.
[19] M. Hosseini, N. Abdolvand, and S. R. Harandi, "Two-dimensional analysis of customer behavior in traditional and electronic banking," *Digital Business*, vol. 2, no. 2, p. 100030, 2022, doi: https://doi.org/10.1016/j.digbus.2022.100030.
[20] N. Owoo and J. Odei-mensah, "Hierarchical clustering-based early warning model for predicting bank failures: Insights from Ghana's financial sector reforms (2017–2019)," *Research in International Business and Finance*, vol. 77, p. 102944, 2025, doi: 10.1016/j.ribaf.2025.102944.
[21] X. Yan, Y. Li, F. Nie, and R. Li, "Bank Customer Segmentation and Marketing Strategies Based on Improved DBSCAN Algorithm," *Applied Sciences*, vol. 15, no. 6. p. 3138, 2025. doi: 10.3390/app15063138.
[22] S. Gáspár, L. Pataki, Á. Barta, G.˝o Thalmeiner, and Z. Zéman, "Consumer Segmentation of Green Financial Products Based on Sociodemographic Characteristics," *Journal of Risk and*

Financial Management, vol. 16, no. 2, p. 98, 2023, doi: https://doi.org/10.3390/jrfm16020098.

[23] T. Hardiani and E. P. Silmina, "Comparative Analysis of K-Means and K-Medoids Algorithms in New Student Admission," vol. 6, no. 2, 2024, doi: 10.35842/ijicom.

[24] M. Raeisi and A. B. Sesay, "A Distance Metric for Uneven Clusters of Unsupervised K-Means Clustering Algorithm," *IEEE Access*, vol. 10, no. July, pp. 86286–86297, 2022, doi: 10.1109/ACCESS.2022.3198992.

[25] Y.-T. Cui and G.-Z. Zhang, "Optimizing community green roof spaces in high-density cities: a K-modes clustering algorithm based analysis of resident preferences and spatial configuration," *Journal of Asian Architecture and Building Engineering*, vol. 24, no. 4, pp. 3052–3069, Jul. 2025, doi: 10.1080/13467581.2024.2366814.

[26] T. Faouzi, L. Firinguetti-Limone, J. M. Avilez-Bozo, and R. Carvajal-Schiaffino, "The α-Groups under Condorcet Clustering," *Mathematics*, vol. 10, no. 5. p. 718, 2022. doi: 10.3390/math10050718.

[27] P. W. Suraya, "Comparing Netflix Content-Based Approaches on Movies and TV Shows," vol. 7, no. 2, 2025, doi: 10.35842/ijicom.

[28] A. Apicella, F. Isgrò, and R. Prevete, "Don't push the button! Exploring data leakage risks inmachine learning and transfer learning," *Artificial Intelligence Review*, vol. 58, p. 339, 2025, doi: https://doi.org/10.1007/s10462-025-11326-3.

[29] M. Alabadla *et al.*, "Systematic Review of Using Machine Learning in Imputing Missing Values," *IEEE Access*, vol. 10, pp. 44483–44502, 2022, doi: 10.1109/ACCESS.2022.3160841.

[30] M. G. Hanna *et al.*, "Ethical and Bias Considerations in Artificial Intelligence/Machine Learning," *Modern Pathology*, vol. 38, no. 3, p. 100686, 2025, doi: https://doi.org/10.1016/j.modpat.2024.100686.

[31] O. E. Karpov *et al.*, "Extreme value theory inspires explainable machine learning approach for seizure detection," *Scientific Reports*, vol. 12, no. 1, p. 11474, 2022, doi: 10.1038/s41598-022-15675-9.

[32] E. Herman, K.-E. Zsido, and V. Fenyves, "Cluster Analysis with K-Mean versus K-Medoid in Financial Performance Evaluation," *Applied Sciences*, vol. 12, no. 16. p. 7985, 2022. doi: 10.3390/app12167985.

[33] C. S. K. Dash, A. K. Behera, S. Dehuri, and A. Ghosh, "An outliers detection and elimination framework in classification task of data mining," *Decision Analytics Journal*, vol. 6, p. 100164, 2023, doi: https://doi.org/10.1016/j.dajour.2023.100164.

[34] A. Q. Md, S. Kulkarni, C. J. Joshua, T. Vaichole, S. Mohan, and C. Iwendi, "Enhanced Preprocessing Approach Using Ensemble Machine Learning Algorithms for Detecting Liver Disease," *Biomedicines*, vol. 11, no. 2. p. 581, 2023. doi: 10.3390/biomedicines11020581.

[35] A. Performance, E. Alshdaifat, D. Alshdaifat, A. Alsarhan, F. Hussein, and S. Moh, "The Effect of Preprocessing Techniques , Applied to Numeric," *Data*, vol. 6, no. 11, 2021.

[36] R. Muñoz-Cancino, C. Bravo, S. A. Ríos, and M. Graña, "On the dynamics of credit history and social interaction features, and their impact on creditworthiness assessment performance," *Expert Systems with Applications*, vol. 218, p. 119599, 2023, doi: https://doi.org/10.1016/j.eswa.2023.119599.

[37] C. Kallandranis, D. Anastasiou, and K. Drakos, "Credit rationing prevalence for Eurozone firms," *Journal of Business Research*, vol. 158, p. 113640, 2023, doi: https://doi.org/10.1016/j.jbusres.2022.113640.

[38] I. W. Sudiantara, N. W. S. Saraswati, and I. P. A. E. D. Udayana, "Uncovering Customer Segments in Savings and Loan Cooperatives Using Clustering Techniques," in *2025 IEEE International Symposium on Consumer Technology (ISCT)*, 2025. doi: 10.1109/ISCT66099.2025.11297252.