

APPLICATION OF CLUSTERING POTENTIAL CREDIT RECIPIENTS USING THE K-MODE METHOD

Yosep Agus Pranoto ¹, Kartiko Ardi Widodo ²

^{1,2} Informatics Engineering, Faculty of Industrial Technology, National Institute of Technology Malang
yaa@lecturer.itn.ac.id

ABSTRACT

This application for clustering potential credit recipients was created as a solution to the problems that arise in the process of applying for a loan in a cooperative. The difficulty arises in determining which applicants are entitled to a loan because of the multiple criteria that are taken into account and the large number of applicants. As a result, the decision-making process takes a long time and is prone to errors. In addition, the decisions made are often subjective. In the development of information technology, a solution can be made by creating a computer-based application using the K-mode method for decision making. This application can be used by related parties to make decisions on the selection of credit acceptance. Based on the tests conducted, the accuracy rate of the application is 95%.

Keyword : cooperative, credit, clustering, K-Mode

1. INTRODUCTION

Cooperatives are business entities consisting of people or legal entities that base their activities on family principles as well as the economic movement of the people. This business entity is considered important in increasing the economic growth of the Indonesian people. Based on this, a cooperative information system can help to facilitate access and be able to support cooperatives as a strong and independent business entity.

Regarding the loan process, many of the loan applicants often complain about the long and complicated loan application process. At present, the maximum time limit from the registration stage to the credit realization stage, set by the management, is a maximum of 7 (seven) working days. From the bank's side, with the increasing rate of delinquent loans, credit initiators must be more careful and selective in providing credit to their members. On the other hand, the cooperative is required to achieve the credit realization target every month. In addition, there is a factor of increasing competition among other competing cooperatives. As a result, the initiators sometimes decide to grant credit without paying attention to aspects of credit risk that should be applied consistently. Thus, a computerized clustering system is needed that can quickly provide information related to the applicant's criteria.

The system to be created is a grouping system for potential loan recipients using the K-mode method. The K-mode method is a method developed from the existing method, namely the K-mean method. The advantage of the K-mode method is that it is able to categorize data and produce more stable clusters and shorter computation time compared to the K-mean method. The existence of an application that uses the K-mode method for grouping potential loan recipients is expected to make it easier for cooperatives to select potential

loan recipients in order to reduce the risk that can harm the cooperative.

2. LITERATURE REVIEW

2.1. K-Mode Method

The flowchart of the K-Mode method used in this application is shown in Figure 1:

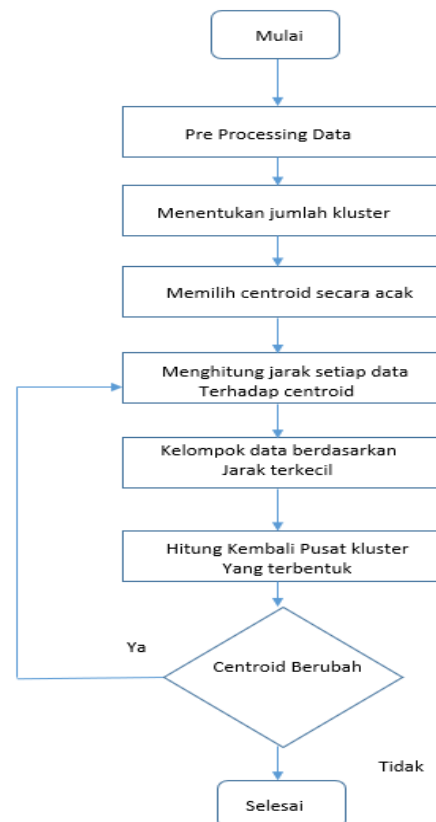


Figure 1. K-mode Flowchart

Start - preprocessing data - determines cluster amount - randomly choose a centroid - calculates the distance of each data to the centroid - data groups based on the smallest distance - recalculate the formed cluster centers - centroid changes - end

The K-mode algorithm is a method developed from the previously existing method, namely the K-mean method. The advantage of this K-mode method is that it is able to categorize data and produce more stable clusters and shorter computation time compared to the K-mean method [1], [2]. The modifications developed in the K-mode method are:

1. Uses simple dissimilarity matching of categorical data features.
2. Replaces the mean with the mode or most frequent value.
3. Uses a frequency-based method to find the mode of a data set.

The algorithm of the K-mode method used to determine clustering is:

1. Select k data for centroid initialization, one for each cluster.
2. Calculate the distance between each object and the cluster mode, assign the object to the cluster whose center has the closest distance to the object, repeat this step until all objects are assigned to the group.

$$D(x, y) = \sum_{j=1}^r \varepsilon(X_j, Y_j) \dots \dots \dots (1)$$

Description:

$D(x, y)$ = x to y data distance

X_j = j^{th} feature value of X

Y_j = j^{th} feature value of Y

R is the number of features and here is the matching value as in the following equation:

$$\varepsilon(X_j, Y_j) = \begin{cases} 0, & X_j = Y_j \\ 1, & X_j \neq Y_j \end{cases} \dots \dots \dots (2)$$

3. Update the mode (as the centroid) of each cluster with the value of the category that appears most often in each cluster.
4. Repeat steps 2 and 3 to fulfil the condition that (a) the data in the cluster does not move, or (b) the position of the centroid does not change..

2.2. Clustering

Clustering is one of the tools in data mining that aims to group objects into clusters. A cluster is a group or set of data objects that are similar to each other in the same cluster and dissimilar to objects in different clusters [3].

2.3. Cooperative

Article 1 Paragraph (1) of Law Number 25 of 1992 concerning cooperatives states that a cooperative is a business entity consisting of individuals or cooperative legal entities by basing its activities on cooperative principles as well as a people's economic movement based on family principles[4].

2.4. Credit

According to Irmayanto (2004: 74), Article 1 paragraph II of Law No. 10/1998 on Amendments to Law No. 7/1992 on Banking states that credit is the provision of money or bills that can be equated with it, based on an agreement or loan and borrowing agreement between the bank and another party, which requires the debtor to repay his debt after a period of time with interest [5].

2.5. Fiduciary Guarantee

A credit agreement that is followed by a fiduciary guarantee agreement in practice must contain details of fiduciary guarantees. The regulation is contained in Act No. 42 of 1999 on Fiduciary Guarantees. The definition of an agreement is as a conformity of will or agreement to an agreement of actions contained therein [6].

2.6. Credit Period

According to Cashmere (2016: 74) the credit period is Every credit given must have a certain period of time, this period includes the agreed credit taking period. It is almost certain that there is no credit that does not have a period of time [7].

3. RESEARCH METHOD

The system flowchart for determining credit recipients is shown in Figure 2:

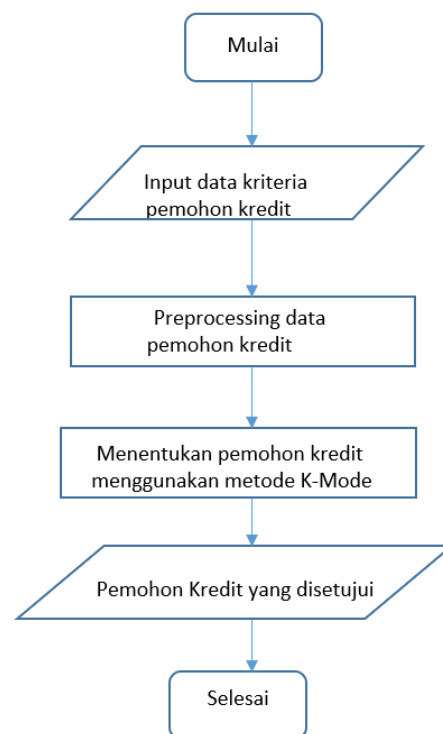


Figure 2. System Flowchart

Start - input credit applicant criteria data – credit applicant preprocessing data – determining credit applicants using K-mode - approved loan applicants

The flow of the credit applicant selection system begins with the input of criteria data from the

applicant, and continues with the preprocessing of the input data. In this process, the values of the criteria data are converted into a category to be processed using the K-mode method. After the conversion process is completed, the process of determining the number of clusters and their centroids is performed, and the minimum distance of each criterion data to the centroid center is calculated. The last step of the system flow is output in the form of decisions about approved loan applicants.

The evaluation criteria used to determine the selection of loan applicants are income, loan amount, credit period, and collateral. While the criteria used in the research object are detailed as in Table 1, Table 2, Table 3 and Table 4 below.

Table 1. Income Criteria

Criterion	Category	Amount
Income	Low	< 2,000,000
	Middle	2,000,000 – 8,000,000
	High	> 8,000,000

Table 2. Credit Amount Criteria

Criterion	Category	Amount
Credit Amount	Small	< 5,000,000
	Middle	5,000,000 – 20,000,000
	Big	> 20,000,000

Table 3. Credit Period

Criterion	Category	Amount
Tenor	Short	< 10 times
	Middle	10 – 20 times
	Long	> 20 times

Table 4. Collateral

Criterion	Category	Amount
Income	Small	< 15,000,000
	Middle	15,000,000 – 50,000,000
	Big	> 8,000,000

As application test data, sample data of 20 people are used, with the results of 11 accepted applicants and 9 rejected applicants taken from the research object as shown in Table 5.

Table 5. Credit Applicant Preprocessing Result Criteria Data

No	Applicant	Income	Credit Amount	Credit Period	Collateral	Result
1	Arman	Low	Small	Short	Big	rejected
2	Dina	Middle	Big	Long	Middle	accepted
3	Rina	Low	Middle	Short	Small	rejected
4	Sintia	Middle	Small	Short	Middle	rejected
5	Vira	Low	Big	Long	Middle	accepted
6	Muklis	High	Middle	Long	Small	accepted
7	Mahdi	High	Kecil	Short	Middle	accepted
8	Roni	Middle	Big	Long	Big	accepted
9	Sukidi	Low	Middle	Short	Middle	rejected
10	Paidi	Middle	Middle	Middle	Small	rejected
11	Misjan	High	Big	Long	Big	accepted
12	Munali	Low	Middle	Middle	Big	rejected
13	Budiyanto	High	Middle	Long	Big	accepted
14	Saidi	High	Big	Middle	Small	accepted
15	Sucipto	Low	Small	Middle	Middle	rejected
16	Buang	Low	Middle	Short	Middle	rejected
17	Muchtar	Middle	Big	Long	Middle	accepted
18	Jumadi	High	Middle	Long	Small	accepted
19	Firman	Middle	Big	Long	Big	accepted
20	Doni	Middle	Small	Short	Small	ditolak

4. RESULT AND DISCUSSION

The application starts by displaying the login page as a user restriction that can manage application data. Users are required to enter their user name and password correctly to enter the main page of the application. On the main page, there is a criteria setting menu, a criteria data input menu, and a K-mode calculation menu. The login form is shown in Figure 3.

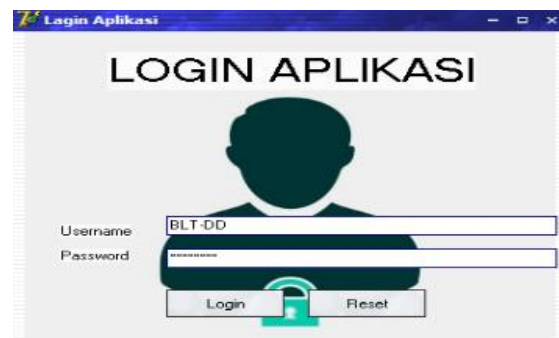


Figure 3. Login Form

The main page of the application is shown in Figure 4.



Figure 4. The main page

The Criteria Data Entry menu displays a form for entering the data of the loan applicant who will later be selected using the K-mode method, as shown in Figure 5.

Figure 5. Data Entry Form

The K-mode calculation menu displays the form for selecting loan applicants using the K-mode method. There are three stages in the K-mode calculation form, namely

- Preprocessing, which is the conversion of the credit applicant's input data into the category data required in the K-mode method process.
- Determining the centroid of the credit application cluster is accepted and the credit application is rejected.
- Perform calculations for credit applicant selection using the K-mode method.
- The K-mode calculation form is shown in Figure 6.

Figure 6. K-mode Calculation Form

To determine the accuracy of the application in calculating the selection of credit applicants using the K-mode method, the application is tested using the sample data contained in Table 5. As for the initial centroid in each cluster is shown in Table 6.

Table 6. Initial Cluster Centroid

Centroid	Income	Credit amount	Credit period	Collateral
Accepted	Big	Small	Long	Big
Rejected	Small	Big	Short	Small

The initial centroid setting in the application is shown in Figure 7.

Figure 7. Initial Centroid Setting in the Application

While the results of the selection of credit applicants using the K-Mode method in the application are shown in Figure 8.

INOMOR KTP	NAMA	KLUSTER
3507062711660002	Aman	Penggajian Kredit ditolak
3507062711660002	Dina	Penggajian Kredit diterima
3507062711660002	Rina	Penggajian Kredit ditolak
3507062711660002	Siria	Penggajian Kredit ditolak
3507062711660002	Via	Penggajian Kredit diterima
3507062711660002	Muklis	Penggajian Kredit diterima
3507062711660002	Mahdi	Penggajian Kredit ditolak
3507062711660002	Rori	Penggajian Kredit diterima
3507061310650002	Sukidi	Penggajian Kredit ditolak
3507062804620002	Padi	Penggajian Kredit ditolak
3507061104680002	Misam	Penggajian Kredit diterima
3507061807610002	Munali	Penggajian Kredit ditolak
3507062309540002	Budyanto	Penggajian Kredit diterima
3507062030690002	Said	Penggajian Kredit diterima
3507062910610002	Sucipto	Penggajian Kredit ditolak
3507062027700002	Buang	Penggajian Kredit ditolak
3507062908660002	Muchtar	Penggajian Kredit diterima
3507062711660002	Jumadi	Penggajian Kredit diterima

Figure 8. Selection Result Form

The Accepted Submission cluster experienced a centroid change twice, while the Rejected Submission cluster experienced a centroid change three times, so the K-mode method was iterated four times. The changes in the centroid of the accepted submission cluster in the application are shown in Figure 9.

Iterasi	Penghasilan	Jumlah Pinjaman	Tenor	Jaminan
1	Besar	Kecil	Lama	Besar
2	Besar	Besar	Lama	Besar
3	Besar	Besar	Lama	Besar

Figure 9. Cluster centroid change Submission accepted

While the change in the centroid of the rejected submission cluster in the application is shown in Figure 10.

Centroid Pengajuan Kredit Ditolak				
Iterasi	Penghasilan	Jumlah Pinjaman	Tenor	Jaminan
1	Kecil	Besar	Pendek	Kecil
2	Kecil	Sedang	Pendek	Kecil
3	Kecil	Sedang	Pendek	Sedang
4	Kecil	Sedang	Pendek	Sedang

Figure 10. Change in cluster centroid Rejected submission

Related to the analysis of K-mode calculations in the application is described as follows:

The first step is to determine the initial centroid of the following Accepted Credit cluster.

Initial Cluster of Accepted Loan Applications

Income = High

Credit Amount = Small

Credit Period = Long

Collateral = Big

Then determine the initial centroid of the rejected application cluster:

Initial Cluster of Rejected Loan Applications

Income = Low

Credit Amount = Large

Credit Period = Short

Collateral = Small

Next, the distance between the 20 data and the centroid of each cluster will be calculated using the K-Mode formula.

Iteration number = 1

distance to cluster 1 = 2 distance to cluster 2 = 2

Entry to Cluster = Credit Application accepted

distance to cluster 1 = 3 distance to cluster 2 = 3

Entry to Cluster = Credit Application accepted

distance to cluster 1 = 4 distance to cluster 2 = 1

Entry to Cluster = Credit Application rejected

distance to cluster 1 = 3 distance to cluster 2 = 3

Entry to Cluster = Credit Application accepted

distance to cluster 1 = 3 distance to cluster 2 = 2

Entry to Cluster = Credit Application rejected

distance to cluster 1 = 2 distance to cluster 2 = 3

Entry to Cluster = Credit Application accepted

distance to cluster 1 = 2 distance to cluster 2 = 3

Entry to Cluster = Credit Application accepted

distance to cluster 1 = 2 distance to cluster 2 = 3

Entry to Cluster = Credit Application accepted

distance to cluster 1 = 4 distance to cluster 2 = 2

Entry to Cluster = Credit Application rejected

distance to cluster 1 = 4 distance to cluster 2 = 3

Entry to Cluster = Credit Application rejected

distance to cluster 1 = 1 distance to cluster 2 = 3

Entry to Cluster = Credit Application accepted

distance to cluster 1 = 3 distance to cluster 2 = 3

Entry to Cluster = Credit Application accepted

distance to cluster 1 = 1 distance to cluster 2 = 4

Entry to Cluster = Credit Application accepted

distance to cluster 1 = 3 distance to cluster 2 = 2

Entry to Cluster = Credit Application rejected

distance to cluster 1 = 3 distance to cluster 2 = 3

Entry to Cluster = Credit Application accepted

distance to cluster 1 = 4 distance to cluster 2 = 2

Entry to Cluster = Credit Application rejected

distance to cluster 1 = 3 distance to cluster 2 = 3

Entry to Cluster = Credit Application accepted

distance to cluster 1 = 2 distance to cluster 2 = 3

Entry to Cluster = Credit Application accepted

distance to cluster 1 = 2 distance to cluster 2 = 3

Entry to Cluster = Credit Application accepted

distance to cluster 1 = 3 distance to cluster 2 = 2

Entry to Cluster = Credit Application rejected

The selection results using the initial centroid (first iteration) are shown in Table 7.

Table 7. Credit Applicant Selection Results using initial centroid (1st iteration)

No	Name	Result
1	Arman	Credit application accepted
2	Dina	Credit application accepted
3	Rina	Credit application rejected
4	Sintia	Credit application accepted
5	Vira	Credit application rejected
6	Muklis	Credit application accepted
7	Mahdi	Credit application accepted
8	Roni	Credit application accepted
9	Sukidi	Credit application rejected
10	Paidi	Credit application rejected
11	Misjan	Credit application accepted
12	Munali	Credit application accepted
13	Budiyanto	Credit application accepted
14	Saidi	Credit application rejected
15	Sucipto	Credit application accepted
16	Buang	Credit application rejected
17	Muchtar	Credit application accepted
18	Jumadi	Credit application accepted
19	Firman	Credit application accepted
20	Doni	Credit application rejected

The results of the calculation in the 1st iteration obtained new values for the centroid in the Accepted Credit cluster and the Rejected Credit cluster, as follows:

Update Cluster of Accepted Credit Applications

Income = High

Credit Amount = Big

Credit Period = Long

Collateral = Large

Update Cluster of Rejected Credit Applications

Income = small

Credit Amount = middle

Credit Period = short

Collateral = small

Iteration will stop if the centroid of each cluster has not changed from the previous iteration. In this study, the process will stop at the 4th iteration with the following analysis:

Cluster Update of Accepted Credit Applications (3rd Iteration)

Income = High

Credit Amount = Big

Credit Period = Long

Collateral = Large

Cluster Update of Rejected Credit Applications (3rd Iteration)

Income = Low

Credit Amount = Medium

Credit Period = Short

Collateral = Medium

Cluster Update of Accepted Credit Applications (4th Iteration)

Income = High

Credit Amount = Big

Credit Period = Long

Collateral = Large

Cluster Update of Rejected Credit Applications (4th Iteration)

Income = Low

Credit Amount = Medium

Credit Period = Short

Collateral = Medium

The data shows that the centroid of each cluster from the third iteration and the fourth iteration has not changed so that in the fourth iteration the process will stop with the credit applicant selection results as shown in Table 8.

Tabel 8. Credit Applicant Selection Results using the 4th Iteration

No	Name	Result
1	Arman	Credit application rejected
2	Dina	Credit application accepted
3	Rina	Credit application rejected
4	Sintia	Credit application rejected
5	Vira	Credit application accepted
6	Muklis	Credit application accepted
7	Mahdi	Credit application rejected
8	Roni	Credit application accepted
9	Sukidi	Credit application rejected
10	Paidi	Credit application rejected
11	Misjan	Credit application accepted
12	Munali	Credit application rejected
13	Budiyanto	Credit application accepted
14	Saidi	Credit application accepted
15	Sucipto	Credit application rejected
16	Buang	Credit application rejected
17	Muchtar	Credit application accepted
18	Jumadi	Credit application accepted
19	Firman	Credit application accepted
20	Doni	Credit application rejected

According to the results of the application tests in Table 8, out of 20 applicants for a loan, 10 applicants were accepted and 10 applicants were rejected. The comparison of the results between the data from the research object and the data from the application test results is shown in Table 9.

Table 9. Comparison of Credit Applicant Selection Results between Real Data and Application Results

No	Name	Real Data	Application Result
1	Arman	rejected	rejected
2	Dina	accepted	accepted
3	Rina	rejected	rejected
4	Sintia	rejected	rejected
5	Vira	accepted	accepted
6	Muklis	accepted	accepted
7	Mahdi	accepted	rejected
8	Roni	accepted	accepted
9	Sukidi	rejected	rejected
10	Paidi	rejected	rejected
11	Misjan	accepted	accepted
12	Munali	rejected	rejected
13	Budiyanto	accepted	accepted
14	Saidi	accepted	accepted
15	Sucipto	rejected	rejected
16	Buang	rejected	rejected
17	Muchtar	accepted	accepted
18	Jumadi	accepted	accepted
19	Firman	accepted	accepted
20	Doni	rejected	rejected

Based on Table 9, there is 1 difference between real data and application results so that the application accuracy level can be calculated according to equation 3.

$$Accuracy = \frac{(20-1)}{20} \times 100\% = 95\%$$

(3)

5. CONCLUSION AND SUGGESTION

After testing the application with a total of 20 data taken from the research object, by setting the initial centroid for the accepted application cluster, namely Income = High, Loan Amount = Small, Credit Period = Long, Collateral = Large, and the initial centroid for the rejected application cluster, namely Income = Small, Loan Amount = Large, Credit Period = Short, Collateral = Small, the selection results were obtained as many as 10 accepted loan applicants and 10 rejected loan applicants with an application accuracy rate of 95%.

For further development, the application can be made based on the website or Android by implementing the KMean clustering method for the selection process.

REFERENCES

- [1] T. Yulianita, D. Istiawan, and A. Statistika Muhammadiyah Semarang, "Implementasi Algoritma K-modes untuk Penentuan Prioritas Rehabilitasi Daerah Aliran Sungai Berdasarkan Parameter Lahan Kritis," *URECOL*, vol. 6, pp. 429–440, 2017.
- [2] N. Sharma and N. Gaud, "K-modes Clustering Algorithm for Categorical Data," 2015.
- [3] N. W. Rohmawati, S. Defiyanti, M. Jajuli, and T. H. Informatika Fakultas Ilmu Komputer Universitas Singaperbangsa Karawang Jl Ronggowaluyo Telukjambe Timur Karawang, "IMPLEMENTASI ALGORITMA K-MEANS DALAM PENGKLASTERAN MAHASISWA PELAMAR BEASISWA," Bandung, Apr. 2015.
- [4] I. G. P. Fajar *et al.*, "Pengembangan Aplikasi Internet Banking Koperasi Simpan Pinjam Berbasis Web," Badung, Nov. 2019.
- [5] S. Zebua, "PENGAMBILAN KEPUTUSAN TERHADAP PEMBERIAN KREDIT YANG DIDASARKAN ATAS INFORMASI AKUNTANSI DAN BUKAN AKUNTANSI (Studi Kasus : PT. BRI Cabang Gunung Sitoli)," *Warta Dharmawangsa*, vol. 15, no. 3, 2021, doi: 10.46576/wdw.v15i3.1356.
- [6] Anak Agung Gede Rizky Pramana, Desak Gde Dwi Arini, and Ni Luh Made Mahendrawati, "PELAKSANAAN PERJANJIAN KREDIT DENGAN JAMINAN FIDUSIA PT.BPR BANK DAERAH GIANJAR KABUPATEN GIANJAR," *Jurnal Preferensi Hukum*, vol. 3, no. 3, 2022, doi: 10.55637/jph.3.3.5594.517-523.
- [7] S. Masyita, "ANALISIS KEPERCAYAAN DAN JANGKA WAKTU KREDIT TERHADAP PENYALURAN KREDIT PADA BANK SULSELBAR MAKASSAR," *PAY Jurnal Keuangan dan Perbankan*, vol. 2, no. 2, pp. 75–82, Dec. 2020, doi: 10.46918/pay.v2i2.754