



Classification of prospective borrowing customers to reduce the risk of bad deposits in sharia cooperatives using the FK-NNC method

Annisa Reza Dhiya'atulhaq^{a,1}, Oliver S. Simanjuntak^{a,2,*}, Heriyanto^{a,3}

^a Department of Informatics, Universitas Pembangunan Nasional Veteran Yogyakarta, 55281, Indonesia

¹ annisarezahaque@gmail.com; ² oliversimanjuntak@yahoo.com*; ³ heriyanto@upnyk.ac.id

* Corresponding Author

ARTICLE INFO

Article history

Received: November 1, 2021

Revised: November 7, 2021

Accepted: November 9, 2021

Keywords

Classification

Sharia Cooperative

Fuzzy K-Nearest Neighbor in every class

ABSTRACT



Objective: Assisting cooperatives in determining the classification of prospective financing members to reduce non-performing deposits in sharia cooperatives

Design/method/approach: The Fuzzy K-Nearest Neighbor in Every Class method is used to classify prospective financing members. System development using the waterfall method.

Results: Based on the implementation and the results of tests carried out using the confusion matrix, the results show that using the Fuzzy K-Nearest Neighbor in Every Class method can classify prospective financing members with an average accuracy rate of 80% with a value of $k=1$ to $k=10$. Stable accuracy results of 80%. It shows that adding k theory to the Fuzzy K-Nearest Neighbor in Every Class method can improve the theory of assigning k values to the previous method, namely K-Nearest Neighbor and Fuzzy K-Nearest Neighbor.

Authenticity/state of the art: Based on previous research carried out, the research themes and characteristics are relatively the same, but in the research conducted, there are differences in terms of the methods used, case study data, preprocessing data, and research outputs. Previous research with the same object, namely the classification of cooperative customers by applying the K-Nearest Neighbor method by determining two classes of classification results, namely traffic jams and smooth, while this study will apply the development of the K-Nearest Neighbor method using the Fuzzy K-Nearest Neighbor in Every method. The class with the output specifies three outcomes: crash, sometimes crash, and smooth. This study uses the data preprocessing stage with fuzzification and data transformation techniques using the min-max normalization method. In contrast, the previous research used the z score normalization method.

This is an open-access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



1. Introduction

Sharia Financing Savings and Loans Cooperative (KSPPS) is a cooperative whose business activities are engaged in savings, loans, and financing following sharia principles (Number 16 /Per/M.KUKM/IX/2015). The advantage of KSPPS is that cooperatives can carry out the distribution of funds and social services simultaneously. However, the problem that often arises in financing funds is frequent bad deposits/installments. One of the contributing factors is the lack of analysis and assessment/customers. Credit analysis errors can lead to credit risks, such as loss of customers,

Classification is the process of finding a model or function that is used to predict an unknown class. Several previous studies related to the classification used the k-nearest neighbor method for classification assessing prospective debtors with five predictor attributes and producing two-class outputs, namely feasible and not feasible with 55 datasets with an average level of accuracy of 81.82 with a k value of 3.5, 7, and 9[1]. Another study using the k-nearest neighbor algorithm method classifies customers' creditworthiness with nine attributes and a value of $k = 1$ by testing ten-fold split validation with accurate results with a confusion matrix of 77.78%[3]. The drawback in using the k-nearest neighbor method is the problematic selection of k values[4]. Other shortcomings have been found in research[5] using the KNN method, which treats all K neighbors in the same way without considering the difference in distance between the test data and neighbors. It is because KNN gets classes from the majority of classes without looking at the proximity of the data. Based on the existing problems, the method uses the potential of fuzzy set theory to handle data uncertainty. According to research [5], using additional fuzzy set theory in KNN can improve overall performance

Fuzzy K-Nearest Neighbor in Every Class is the development of the K-Nearest Neighbor and Fuzzy K-Nearest Neighbor methods. The concept difference lies in the number of k closest neighbors for each class from a test data, not just k closest neighbors such as K-NN/FK-NN. Based on research that has been done in comparing the K-Nearest Neighbor (KNN), Fuzzy K-Nearest Neighbor (FK-NN), Fuzzy K-Nearest Neighbor in Every Class (FK-NNC) method, the results show that using the Fuzzy K-Nearest Neighbor in Every Class method (FK-NNC) provides accurate results that are proven to be higher than the two comparison methods, namely 82% - 97%[4]. Research using the Fuzzy K-Nearest Neighbor in Every Class (FKNNC) method has been carried out by classifying the nutritional status of toddlers with three attributes and two output classes with an accuracy level of 85.19 with a value of $k=8$ [6].

2. Method

The research method used is quantitative. The stages carried out in this research are data collection, data labeling, data preprocessing, classification and testing. The stages in the research can be seen in Fig. 1.

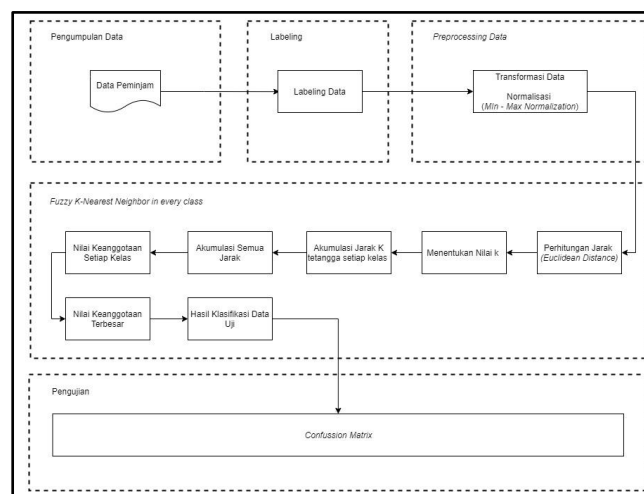


Fig. 1. Research Methodology

2.1. Data Collection

Collecting data in the study used primary data. The primary data used is data taken directly at the research location, namely KSPPS (Sharia Financing Savings and Loans Cooperative) Zatabbaru Sejahtera Mandiri. The data used to classify prospective customers is from member data and borrower data by taking 150 datasets. The data will be divided into 120 training data and 30 test data with a ratio of 80:20 with the same data distribution.

2.2. Data Labeling

Data labeling is giving a class label to the data. The data will be categorized into three classes, namely fluent, substandard, and jammed. Giving class labels based on borrower installment data based on SOPs from the Zatabbaru Sejahtera Mandiri Cooperative can be seen in Table 1.

Table 1. Output Class Label

No	Class Label	arrears
1	Fluent	<1 month
2	Not that smooth	>3-6 months
3	Congested	>6 months

2.3. Data Preprocessing

Preprocessing data is a series of data preparation processes before the data is entered into the classification model. The stage used is to perform data transformation. Data transformation is changing data into a suitable form. Data transformation has several techniques, one of which is to perform data transformation, namely data normalization with the min-max normalization method.

a. Data Representation

Before entering the classification stage, it is necessary to prepare the data first. Ten attributes that will be used in the research are taken. The initial stage to carry out the calculation process is to convert data on the attribute values of the representations that have been determined can be seen in Table 2.

Table 2. Attributes

No	Data Type	Attribute	Attribute Value	Representation
1	interval	Age	Young	0.0 – 1.0
			Old	
			Middle-aged	
2	Nominal	Status	Marry	1
			Single	2
			Death divorce	3
3	Nominal	Membership	Tall	1
			Intermediate	2
			Low	3
3	Nominal	Work	entrepreneur	1
			Private sector employee	2
			Farmer	3
			Trader	4
			Day Worker	5
			civil servant	6
			IRT	7
			Does not work	8

NoData	Type	Attribute	Attribute Value	Representation
4	Nominal	Couple's Job	entrepreneur	1
			Private sector employee	2
			Farmer	3
			Trader	4
			Day Worker	5
			civil servant	6
			IRT	7
			Does not work	8
5	interval	Income	Low	0.0 – 1.0
			Currently	
			Tall	
6	interval	Spouse Income	Low	0.0 – 1.0
			Currently	
			Tall	
7	interval	The number of dependents	A little	0.0 – 1.0
			Currently	
			Lots	
8	Nominal	Homeownership	One's own	1
			contract	2
			ride	3
9	interval	The value of collateral	Small	0.0 – 1.0
			Currently	
			Lots	
10	Label	Financing Status	Fluent	1
			Not that smooth	2
			Congested	3

The first stage of preprocessing is to prepare data by converting data. Of the ten attribute data to be processed, five attributes are converted into fuzzy sets, namely age, income, spouse's income, number of dependents, and guarantee values converted into linguistic variables with fuzzy sets and fuzzy membership degrees obtained.

Fuzzy expressed in the degree of membership used as a determinant of the existence of elements in a set [13]. In contrast to the firm set, which has a value of 0 (no) or 1 (yes), fuzzy logic has a membership value with a range of 0 (zero) to 1 (one). The fuzzy set is expressed by a membership function in the universe U which groups based on linguistic variables. The membership function of the fuzzy set has a degree of membership in a set. The value of the member function is within the range of real numbers in the interval [0,1] [14] expressed in equation (2.1)

$$A: x \rightarrow [0,1] \tag{1}$$

b. Normalization (Min-Max Normalization)

To prevent data values that have different ranges because some data have gone is through the fuzzification process, a data normalization process is carried out where all data will be scaled with a range of values between 0.0 to 1.0. data normalization using the min-max normalization method. In calculating the Euclidean distance, attributes with large values can be more influential when compared to attributes with small values. It is necessary to normalize the data by performing a transformation process on the data[7]. Normalization is a process in data transformation where a numeric attribute is scaled in the range of 0.0 to 1.0 [8]. If there are mixed attributes in the form of numeric and categorical data in the training data, then it is better to use the min-max normalization method [9]. To perform the min-max normalization calculation using the following Equation 2.

$$x_{ij*} = \frac{x_{ij} - x_{min}}{x_{min} - x_{max}} \tag{2}$$

Information:

x_{ij} : attribute value

x_{min} : attribute minimum value

X_{max} : attribute max value

2.4. Fuzzy K-Nearest Neighbor in Every Class

After going through the data preparation process in data preprocessing, enter the classification process stage. The classification method used is the Fuzzy K-Nearest Neighbor in every class algorithm. Fuzzy K-Nearest Neighbor in Every Class (FK-NNC) is a development of the K-Nearest Neighbor (K-NN) and Fuzzy K-Nearest Neighbor (FKNN) methods. This method slightly modifies the Fuzzy K-Nearest Neighbor (FK-NN) method by giving some k closest neighbors to each class label of a test data (each class label has the same number of neighbors as many as the k). This method works because each class label has k closest neighbors, where the closest neighbor of a class label is the training data owned by the class label, which is closest to the test/prediction data. Not the training data that is closest to the test data as in the KNN standard method. The class label in this method is very influential, which aims to reduce the unbalanced class[10]

The use of the Fuzzy K-Nearest Neighbor in Every Class method is based on research that compares the K-Nearest Neighbor (KNN) method, Fuzzy K-Nearest Neighbor (FKNN), Fuzzy K-Nearest Neighbor in Every Class (FKNNC) method. Fuzzy K-Nearest Neighbor in Every Class (FKNNC) gives an accurate result higher than the two comparison methods, 82% - 97%. Research that is expected to have a better level of accuracy in the end[4]. One of the problems K-Nearest Neighbor (KNN) & Fuzzy K-Nearest Neighbor (FKNN) faces is the problematic selection of k, the majority voting method from k neighbors for a considerable k value result significant data distortion. If k is too small, it can result in significant data distortion. Causing the algorithm to be too sensitive to noise [4] The researchers used the Fuzzy K-Nearest Neighbor in Every Class method to correct the weaknesses in the K-NN and FK-NNC methods to train data containing noise and improve the performance of the two previous methods.

The results obtained in implementing the FK-NNC algorithm will get the results of the classification of prospective borrowing customers. The stages in the Fuzzy K-Nearest Neighbor in every class algorithm will be described in the flowchart, which can be seen in Fig. 2.



Fig. 2. Flowchart Method

a. Calculating Distance (Euclidean Distance)

After getting the normalization results, the next step is to find the value of d (distance between training data and test data) for each data using the Euclidean method, which is used to calculate the similarity of a data to find the distance calculation between data [15] using (3) [11].

$$d(x_i, x_j) = \left(\sum_{l=1}^N |x_{il} - x_{jl}|^p \right)^{\frac{1}{p}} \quad (3)$$

Information

d(x_i,x_j): distance between test data and training data

i: 1,2, ... n, number of test data

j: 1,2,3 ..n, the number of training data

N: number of parameters

P: determination of the distance used (p = 2 Euclidean distance)

b. Determining the value of k

After getting the results of the distance between the test data and the training data for each data, determine the value of k (e.g., k=2) to take the closest data for each class. If it is determined that k = 2, then the data taken is kx C (class) which means it will get the two closest data for each class with a total of six data to be calculated

c. Calculating S . Value

Then calculate the value of S as the accumulated distance of k neighbors of each class using (4).

$$S_{ij} = \sum_{r=1}^K d(x_i, x_j)^{-2} \quad (4)$$

Information :

S_{ij}: total distance of test data (x) to all k neighbors from each class (number of training data)

d: total distance of test data (x) to k neighbors in the jth class (done as much as C class)

d. Calculating D Value

After getting the value of S as the accumulated distance of k nearest neighbors of each class, then calculate the value of D as the accumulation of all the distances of CxK neighbors using Equation 5

$$D_i = \sum_{j=1}^C (S_{ij}) \quad (5)$$

Information :

D: the result of accumulated test data distance (x) to each class

C: number of classes

m: weight rank, m > 1

e. Calculating U Nilai Value

After getting the value of D, then calculate the value of data membership in each class with (6).

$$u_{ij} = \frac{S_{ij}}{D_i} \quad (6)$$

Information :

u: membership value

S_{ij}: number of test data distances (x) to all k neighbors of each jth class (sum of training data)

D_i: the result of accumulated test data distance (x) to each class

f. Determining the Largest U Value

The last step is to determine the final result of the test data, select the membership value with the most significant value. The most considerable value will be the result of the classification class using Equation 7

$$y' = \arg \max_{j=1} (u_{ij}) \quad (7)$$

2.5. Testing

Confusion matrix is a method for measuring the classification model. Confusion matrix will give an assessment performance based on true or false objects in the classification [12]. There are four terms in confusion matrix which represents the results of the classification, namely true positive, true negative, false positive and false negative. Table confusion matrix can be seen in Table 3.

Table 3. Confusion Matrix

Predicted Values	Actual Values	
	Positive	negative
Positive	<i>True Positive (TP)</i>	<i>False Positive (FP)</i>
negative	<i>False Negative (FN)</i>	<i>True Negative (TN)</i>

Based on the table can be used to calculate the level of accuracy of the model that has been made. The calculation to get the accuracy value is seen in Equation 7

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN} \times 100\% \quad (7)$$

3. Results and Discussion

In this study, research related to the classification of prospective financing members has been carried out using the Fuzzy K-Nearest Neighbor in every class method. The data used in this study were taken directly at KSPPS Zatabbaru Sejahtera Mandiri Klaten with 150 datasets of members who have applied for financing with paid status. The dataset is divided into 120 training data and 30 test data which is carried out with a value of k at 1 to 10. The process begins with processing the data and preparing the data by taking ten attributes and one output label, then preprocessing the data with transformation using the min max normalization technique, after the data the process is prepared, namely by classifying prospective members in applying for financing using the Fuzzy K-Nearest Neighbor in every class method.

Based on the test results on the classification using the Fuzzy K-Nearest Neighbor in every class method, it shows accuracy with a stable accuracy result of 80% seen from the overall accuracy of the value of k=1 to k=10 the average accuracy shows 80% results.

The results obtained indicate that the use of the Fuzzy K-Nearest Neighbor in every class method can be used to classify prospective members in financing applications with fairly good results. From the results of the research that has been done, it shows that the selection of the value of k in the Fuzzy K-Nearest Neighbor in every class method is easier than the two previous methods, namely K-Nearest Neighbor & Fuzzy K-Nearest Neighbor because the selection of k in this study proved to have stable final accuracy results.

3.1. Interface Results

The results of the design produce an implementation in the form of an interface in the form of a system display when used, which is like Fig. 3.

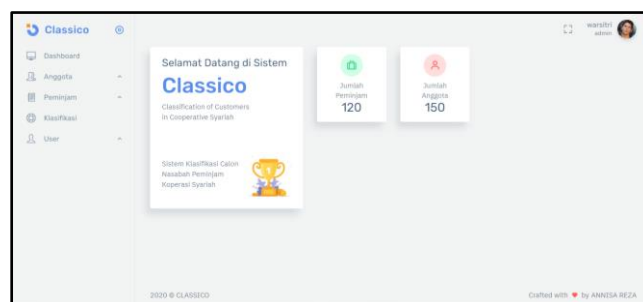


Fig. 3. Classification System Page

3.2. Test

Tests using a confusion matrix will measure the level of classification accuracy using the Fuzzy K-Nearest Neighbor in every class method with a value of $k=1$ to $k=10$. The following is the result of calculating accuracy using a confusion matrix with $k=1$ to $k=10$ in Table 4.

Table 4. Confusion Matrix

Parameter	Fuzzy K-Nearest Neighbor in Every Class									
	K=1	K=2	K=3	K=4	K=5	K=6	K=7	K=8	K=9	K=10
Accuracy	80%	80%	80%	80%	80%	80%	80%	80%	80%	80%

After getting the accuracy results using the confusion matrix with a value of $k=1$ to $k=10$, it was found that the level of accuracy was stable, namely 80%.

4. Conclusion

Based on the research that has been done, it can be concluded that this classification system can predict prospective financing members and the benefits obtained for sharia cooperatives in assessing financing members before providing loans. The Fuzzy K-Nearest Neighbor in every class classification method can perform classification with three class outputs, namely smooth, substandard, and jammed with a stable accuracy of 80%. And the average result up to $k = 10$, which is 80%, shows that the selection of the k value in the Fuzzy K-Nearest Neighbor in every class method is easier because the use of the k value is relatively stable. The comparison of the dataset used is 80:20 where the training data is 120 and the test data is 30.

Suggestions that can be used as further research developments are training data and testing data to be added to determine the effect on classification. Further research is expected to add a weighting method to the Fuzzy K-Nearest Neighbor in every Class so that the classification used will look at the weighting of each variable and will produce a classification that is in accordance with the weighting of the variables. Research on the classification of prospective borrowers focuses on the Fuzzy K-Nearest Neighbor algorithm in every class, so it needs to be compared with other classification algorithms to get better results.

References

- [1] Menarianti, I. (2015). Klasifikasi Data Mining Dalam Menentukan Pemberian Kredit Bagi Nasabah Koperasi. 1(1).
- [2] I. W. Supriana, M. A. Raharja, and P. W. Gunawan, "Sistem Informasi Prediksi Penilaian Kredit Perbankan Menggunakan Algoritma K-Nearest Neighbor Classification," JST (Jurnal Sains dan Teknol., vol. 8, no. 1, p. 44, 2019, doi: 10.23887/jst-undiksha.v8i1.16470.
- [3] H. Annur and M. E. Lasulika, "Klasifikasi Nasabah Kredit Koperasi Menggunakan Algoritma K-Nearest Neighbor," J. Inform. Upgris, vol. 5, no. 2, 2019, [Online]. Available: <http://journal.upgris.ac.id/index.php/JIU/article/view/4343/0>.
- [4] E. Prasetyo et al., "Fuzzy K-Nearest Neighbor in Every Class Untuk Klasifikasi Data," no. Santika, pp. 57–60, 2012.
- [5] D. Li, J. S. Deogun, and K. Wang, "Gene Function Classification Using Fuzzy K-Nearest Neighbor Approach," pp. 644–644, 2008, doi: 10.1109/grc.2007.99.
- [6] R. U. Izzah, S. Wahyuningsih, and F. D. T. Amijaya, "Classification of nutritional status of toddlers using fuzzy k-nearest neighbor in every class (FK-NNC)," J. Phys. Conf. Ser., vol. 1277, no. 1, 2019, doi: 10.1088/1742-6596/1277/1/012050.

-
- [7] E. Budianita and W. Prijodiprodjo, "Penerapan Learning Vector Quantization (LVQ) untuk Klasifikasi Status Gizi Anak," *IJCCS (Indonesian J. Comput. Cybern. Syst.*, vol. 7, no. 2, p. 155, 2013, doi: 10.22146/ijccs.3354.
- [8] H. Junaedi, H. Budianto, I. Maryati, and Y. Melani, "Data Transformation pada Data Mining," *Pros. Konf. Nas. Inov. dalam Desain dan Teknol.*, vol. 7, pp. 93–99, 2011.
- [9] D. T. Larose and C. D. Larose, *Discovering Knowledge in Data: An Introduction to Data Mining: Second Edition*, vol. 9780470908. 2014.
- [10] Santoso, B. Aziz, A.I.S., Zohrahayaty. (2020). *Machine Learning & Reasoning Fuzzy Logic Algoritma, Manual & Rapid Miner*. Yogyakarta : Deepublisher.
- [11] Anton, Howard & Chris Rorres. (2005). *Aljabar Linier Elementer edisi 8*. Jakarta: Erlangga.
- [12] Gorunescu, Florin. 2011. *Data Mining: Concepts, Models, and Techniques*. Verlag Berlin Heidelberg : Springer
- [13] Kusumadewi dan Purnomo. 2010, *Aplikasi Logika Fuzzy untuk Sistem Pendukung Keputusan*, Ed. 2, Graha Ilmu, Yogyakarta.
- [14] Saelan, A. (2009). *Logika Fuzzy. Struktur Diskrit*, 1(13508029), 1–5.
- [15] H. Anton, *Elementary Linear Algebra*, 7th ed. New Jersey: Wiley, 1993.