ISSN: 3021-7636
International Journal of Science Technology and Health

Research Article
Vol. 4 (2), 2026, page 88-99
DOI: https://doi.org/10.69930/ijsth.v4i2.77

Classification of Credit Customers of Mutiara Sharia Savings, Loans and Financing Coorporative Using the KNN Method

Muammar Reza Pahlawan [1,*], Nur Islamuddin [2], Shabarul Mukjizat [3]

[1] Departement of Technology Information, Faculty of Sains and Technology , Institute Sains Dan Teknologi A'syiyah, Indonesia
[2] Departement of Management Faculty of Management, STIE Enam-enam, Indonesia
[3] Departement of Politeknik Teknokrat Internasional, Faculty of Marketing management, Yayasan Pendidikan Muslim Teknokrat, Indonesia

* Correspondence: tintareza@gmail.com

Received : 29 June 2026	Accepted: 28 July 2026	Published: 31 July 2026

Abstract: This study applies the K-Nearest Neighbor (KNN) algorithm to classify the credit status of savings and loan cooperative members into two categories, namely current credit and non-performing credit, based on the variables provided. The data source for this study is data collected by the researcher directly from the primary source, namely the Muharjam Kolaka savings and loan cooperative. Several data collection techniques include: observation methods, interviews, and literature research. The classification process begins with a data preprocessing stage including handling missing data (missing values), data normalization and transforming categorical data into numerical. Overall, the results of the study indicate that the KNN algorithm has good potential as a decision-making tool in the creditworthiness analysis process in cooperatives. This study divided the training data and testing data by 20% or 0.2 where the total number of data used was 614, the testing data became 123, the training data 491 after the 2% division. Next, determine the K value, the K value used was K = 11. In the performance testing phase of the model developed, this study used a confusion matrix, with an F1-score of 90%, a recall of 98%, and a precision of 83%. The accuracy was 0.83, or rounded up to 0.84, or 84%. Both computerized and manual calculations yielded similar results according to the accuracy formula in the confusion matrix.

Keywords: Classification; KNN; Confusion Metrics; Bad Credit; Accuracy

1. Introduction

In this digital age, discussions about the latest technologies such as machine learning, artificial intelligence, and data mining will never cease. Machine Learning is presently most popular and complex computer vision approach that became most prominent in the research and industry. It is presently most favorite innovative area in health care, finance, security, data management, trend analysis and prediction.(Pandey et al., 2008) Classification, for example, is often used for predictions, both in nature and human behavior. For example, human predictions such as predicting whether someone has a rare disease, or predicting

fraudulent behavior in banks, or predictions regarding human habits will then be classified based on their class. In this study, the classification method used is the KNN algorithm, which is quite good at performing a classification. Non-parametric kNN can offer greater robustness to label noise, as it treats all neighbors equally without amplifying the influence of potentially mislabeled instances. In contrast, weighted kNN gives higher importance to closer neighbors, which can be misleading when nearby samples are corrupted (Pfeifer & Kreuzthaler, 2026). Furthermore, KNN can also be used to make predictions regarding health or disease, as previously investigated by (Hatem, 2022) & (Uddin et al., 2022). Classification is a data mining task of predicting the value of a categorical variable (target or class) by building a model based on one or more numerical and/or categorical variables (predictors or attributes). (Gourav Rahangdale et al., 2016). That study notes that various artificial intelligence and machine learning technologies have been employed—particularly over the last two decades—to advance the field of medicine in general, in terms of both diagnosis and treatment. This study was conducted with the aim of predicting bad debt for prospective new cooperative members and testing the performance of the algorithm used to obtain an accuracy value. Thus, this research will help a savings and loan cooperative to address the problem of bad debt. This research will also help cooperative managers in making real decisions to predict whether someone is initially monitored for bad debt or not. This study utilizes Python, a programming language specialized for data and a popular tool for machine learning.(Pedregosa FABIANPEDREGOSA et al., 2011)

2. Methods

The method used in this study is a currently popular machine learning classification method. The algorithm uses K-Nearest Neighbor, which calculates the closest distance between data sets. KNN is a top-ten classification algorithm for data mining.. (Zhang & Li, 2023). the model-free k nearest neighbors (kNNs) method does not have training stage and conducts classification tasks by first calculating the distance between the test sample and all training samples to obtain its nearest neighbors and then conducting kNN classification1(which assigns the test samples with labels by the majority rule on the labels of selected nearest neighbors). Because of its simple implementation and significant classification performance, kNN method is a very popular method in data mining and statistics and thus was voted as one of top ten data mining algorithms. (Zhang et al., 2018)

In pattern recognition, the KNN algorithm is an example-based learning method used to classify objects based on the closest examples in their feature space. Objects are classified based on the majority of their neighbors, where objects will be assigned to the most common class among their K-nearest neighbors. (Boateng et al., 2020), (Pfeifer & Kreuzthaler, 2026)Although KNN has been applied to many classification problems, its use for extracurricular placement in Indonesia is still limited and often from previous studies focused on products, films, courses or academic programs. (Anwar et al., 2026)

The KNN algorithm is an instance-based non-parametric learning method that classifies new samples based on their similarity to existing training instances. (AlOmair et al., 2026).

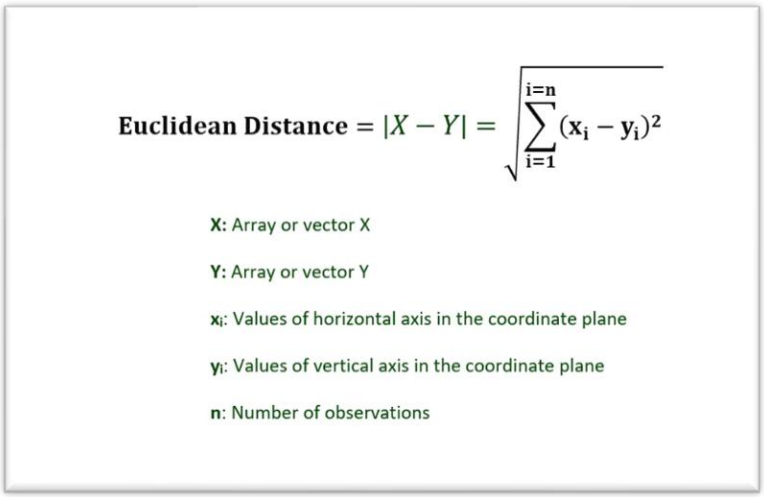


Figure 1. Euclidean Distance

Euclidean distance, or Euclidean distance, is a metric used to measure the straight-line distance between two points in Euclidean space. Euclidean space is a multidimensional space that has the same geometric properties as the two-dimensional or three-dimensional space we are familiar with in everyday life. In Euclidean space, the distance between two points is calculated using the Pythagorean theorem. Mathematically, in nearest neighbor learning, the discrete object classification function is f: Rn->V where V is a finite set {v1,v2,…vs}, namely the set of different categories. The choice of the value of k nearest neighbors is based on the number and degree of distribution in each type of sample, and different values of K can be chosen. (Wang, 2019).. KNN algorithms are ranked among the simplest intelligent algorithms that do not require any learning phase. It is based on calculating the distances between the sampling points to the nearest neighbors of the set of assigned points [19]. The decision is based on the majority vote of the KNN. Many types of distances can be used to decide the nearest neighbors. (Kherif et al., 2021).

All data from this study is included in the appendix, the following is the information obtained from the data:

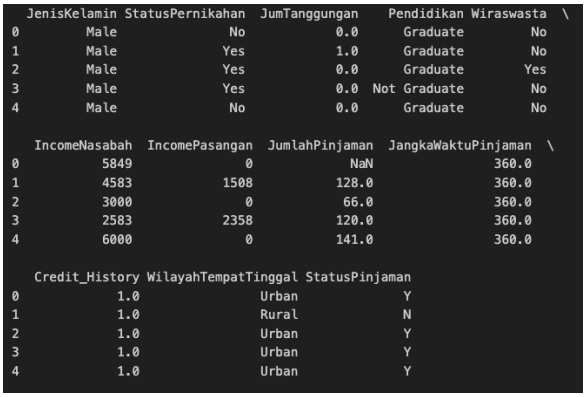
a. ID_Anggota
b. JenisKelamin
c. StatusPernikahan
d. JumTanggungan
e. Pendidikan
f. Wiraswasta
g. IncomeNasabah
h. IncomePasangan
i. JumlahPinjaman
j. JangkaWaktuPinjaman
k. CreditHistory
l. Wilayah
m.StatusPinjaman

The following is a partial overview of the dataset from a study conducted on the feasibility of providing credit to cooperative members. The overview can be seen in the image below:

	ID_Anggota	JenisKelamin	StatusPernikahan	JumTanggungan	Pendidikan	Wiraswasta	IncomeNasabah	IncomePasangan
0	LP001002	Male	No	0.0	Graduate	No	5849	0
1	LP001003	Male	Yes	1.0	Graduate	No	4583	1508
2	LP001005	Male	Yes	0.0	Graduate	Yes	3000	0
3	LP001006	Male	Yes	0.0	Not Graduate	No	2583	2358
4	LP001008	Male	No	0.0	Graduate	No	6000	0

Figure 2. Dataset

In this study, several data distributions or data modifications were performed to facilitate the classification process. This included removing data deemed to have no direct relevance to the desired results , such as Member ID. The following illustrates the changes made to the dataset:



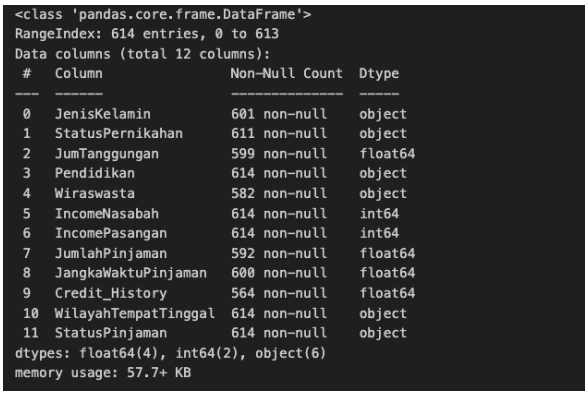
	JenisKelamin	StatusPernikahan	JumTanggungan	Pendidikan	Wiraswasta	\
0	Male	No	0.0	Graduate	No	
1	Male	Yes	1.0	Graduate	No	
2	Male	Yes	0.0	Graduate	Yes	
3	Male	Yes	0.0	Not Graduate	No	
4	Male	No	0.0	Graduate	No	

	IncomeNasabah	IncomePasangan	JumlahPinjaman	JangkaWaktuPinjaman	\
0	5849	0	NaN	360.0	
1	4583	1508	128.0	360.0	
2	3000	0	66.0	360.0	
3	2583	2358	120.0	360.0	
4	6000	0	141.0	360.0	

	Credit_History	WilayahTempatTinggal	StatusPinjaman
0	1.0	Urban	Y
1	1.0	Rural	N
2	1.0	Urban	Y
3	1.0	Urban	Y
4	1.0	Urban	Y

Figure 3. Dataset Missing Value

After removing data deemed unnecessary, this study also checked the dataset to determine if there were any missing values in each column or attribute used. The following are the results of the missing value analysis:



```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 614 entries, 0 to 613
Data columns (total 12 columns):
 #   Column                Non-Null Count  Dtype
---  ------                --------------  -----
 0   JenisKelamin          601 non-null    object
 1   StatusPernikahan      611 non-null    object
 2   JumTanggungan         599 non-null    float64
 3   Pendidikan            614 non-null    object
 4   Wiraswasta            582 non-null    object
 5   IncomeNasabah         614 non-null    int64
 6   IncomePasangan        614 non-null    int64
 7   JumlahPinjaman        592 non-null    float64
 8   JangkaWaktuPinjaman   600 non-null    float64
 9   Credit_History        564 non-null    float64
 10  WilayahTempatTinggal  614 non-null    object
 11  StatusPinjaman        614 non-null    object
dtypes: float64(4), int64(2), object(6)
memory usage: 57.7+ KB
```

Figure 4. Removing Data Deemed

From Figure 4, we can conclude that there are many missing values in the dataset used. For example, we can see that there are three missing data points for marital status. Therefore, one alternative is for the researcher to fill in the missing values. For more details on the missing values in each attribute or column, see the following figure:

```
JenisKelamin            13
StatusPernikahan         3
JumTanggungan           15
Pendidikan               0
Wiraswasta              32
IncomeNasabah            0
IncomePasangan           0
JumlahPinjaman          22
JangkaWaktuPinjaman     14
Credit_History          50
WilayahTempatTinggal     0
StatusPinjaman           0
dtype: int64
```

Figure 5. Missing Value

The analysis process of this research also ensures that the data is truly ready for modeling. However, the next step that can be taken after addressing missing values is to convert categorical data to numeric data. Machine learning models are inherently incapable of processing categorical data, so converting categorical data to numeric data is necessary. Many machine learning models, such as linear regression and support vector machines, are completely incapable of processing categorical data. One technique to change this is One Hot Encoding, also known as dummy variables. One Hot Encoding transforms categorical data by creating a new column for each category.

The collected data, not all information will be included in the experimental phase. Some information has been removed beforehand, leaving only the information that is essential and needed for the classification method to be implemented.

The data that will be used in this trial phase has information consisting of gender, marital status, number of dependents, education, self-employed, income, spouse's income, credit history, and class label which is the loan status, label 0 is said to be not eligible for loan status, and 1 for eligible.

The class label or loan status represents the loan's eligibility, with each number representing a specific status. An explanation of the loan status or class label can be found in the following table:

Table 1. Code class label.

Kode	Label Class
0	Not Fesiable (N)
1	Worthy (Y)

To facilitate the experiment, all categorical data was converted to numeric data. In this case, all genders were converted to numbers. The rules are as follows:

Table 2. Code gender.

Kode	Label Gender
0	Male
1	Female

For education data, which is also text, it is converted into numbers so that it can be used in the classification model calculations. The rules for changing the numbers for education can be seen in Table 3.

Table 3. Code Education.

Kode	Label Education
0	Graduate
1	No Graduate

For self-employed data, the data format is also changed from text to numbers. The following are the provisions for changing the data:

Table 4. Code Self employed.

Kode	Label Self Employeed
0	No
1	Yes

For marital status data, which is text, it has been modified in the preprocessing stage because the classification model will be more usable for processing the data later. The following are the provisions for data modification.

Table 5. Code Marital Status.

Kode	Label Marital Status
0	No
1	Yes

At this stage, the researchers will test the performance of the classification model used. They have developed a simple program using Python. The following is a snippet of the Python program, which imports the required libraries.

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt

from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from sklearn.neighbors import KNeighborsClassifier
from sklearn.metrics import confusion_matrix, classification_report
from sklearn.metrics import f1_score
from sklearn.metrics import recall_score
from sklearn.metrics import accuracy_score
from sklearn.preprocessing import LabelEncoder
```

Figure 6. code of Library Import

Next is a snippet of the Python program used to input the dataset used in this study. Here is the snippet of the Python program:

```
dataset= pd.read_csv('dataset-pinjaman.csv')
print(len(dataset))
dataset.head()
```

Figure 7. code of insert dataset

The following program excerpt shows the process of deleting several columns deemed less important in the classification process in the case the researcher conducted. Data deletion naturally has various reasons. In this study, for example, deleting unimportant attributes and removing several features with too many empty or null data points.

```
dataset = dataset.drop('ID_Nasabah', axis=1)
dataset = dataset.drop('JumlahPinjaman', axis=1)
dataset = dataset.drop('JangkaWaktuPinjaman', axis=1)
dataset = dataset.drop('WilayahTempatTinggal', axis=1)
```

Figure 8. code of column delete

Missing value is a crucial problem, which could compromise the quality ofdata, so missing value estimation is a significant preprocessing step for further experiments. (Yang et al., 2020) The following Python program fragment fills null data with the average or mean value of the target column. The explanation for this is that some values were missing, so the researcher filled them with the average value of the most frequently filled attribute or feature.

```
rata_hsitory = dataset['Credit_History'].mean()
dataset['Credit_History'] = dataset['Credit_History'].fillna(rata_hsitory)
dataset['Credit_History'].isna().sum()
```

Figure 9. code of missing value

Researchers need to change the form of data from text to numbers, so the following simple program snippet represents this, where several columns or attributes are data in text form, and then the following program lines make changes to the data in these columns.

```
dataset['JenisKelamin'] = encoder.fit_transform(dataset['JenisKelamin'])
dataset['StatusPernikahan'] = encoder.fit_transform(dataset['StatusPernikahan'])
dataset['Pendidikan'] = encoder.fit_transform(dataset['StatusPernikahan'])
dataset['Wiraswasta'] = encoder.fit_transform(dataset['Wiraswasta'])
dataset['StatusPinjaman'] = encoder.fit_transform(dataset['StatusPinjaman'])
```

Figure 10. code of encoder

The following is a Python code snippet that splits the dataset to divide the data into testing and training data. In this study, the researchers split the data by 20%. The following is a line of Python code. It's important to note that in machine learning, training data is the data that will be trained, while testing data is the data that will be tested during the model testing process.

3. Results and Discussion

The research conducted revealed that the performance test results for the classification model used by the researchers showed a fairly good score. The accuracy score for this dataset was 84%. The application developed and generated from this research also successfully produced predictions using the Euclidean distance calculation, which is often used in predicting events using the KNN algorithm.

This study cannot be considered a large data set, as it only used hundreds of data sets with 10 attributes. The recall, precision, and f1-score results were 0.45 for class (0), 0.85 for class (1), 0.88 for class (0), and 0.83 for class (1), respectively.

3.1. Testing

In the next testing stage, the researchers conducted a test using a previously determined model, the K-nearest neighbor classification model, commonly known as the K-NN Classifier. In the test, the researchers set the K value to 11, or K=11. The following are the results of the classification performance test using confusion matrices.

3.1.1 Research Stages

The model testing stage, as outlined in the flow above, explains the initial process involved in testing the model chosen by the researcher. The explanation is as follows:

1. Reading the dataset is a process of entering the dataset so that the data can be checked, then carrying out the next process

2. Preprocessing is a step that is always performed in data processing; another term is preparation. In this stage, researchers clean the data, such as removing noise, removing unimportant attributes, and performing an encoder to change the scale of the data.
3. The next step is to divide the data between training and testing data for use in training and testing. In this case, the researchers divided the data by 20%, or 0.2%.
4. The next process is to determine the value of K, the algorithm used in this study requires the value of K as the nearest neighbor value, the value of K in this study is K = 11, this value is obtained from the KNN process carried out previously in the program code line.
5. The next step is to test the model's performance. This is a crucial step because researchers want to see the results of the model's performance. Classification is a task in machine learning. This task is called classification because the computer makes a guess about which category/class the input data will fall into. In this research case, it is between two classes: appropriate or inappropriate. In testing the model's performance, researchers used confusion metrics, which will produce values for accuracy, f1-score, recall, and precision.
6. The final stage is to obtain the accuracy results from the model, there is no benchmark for the best value in testing the model, but if the accuracy is more than 70, then it can be said that the accuracy value is quite good.

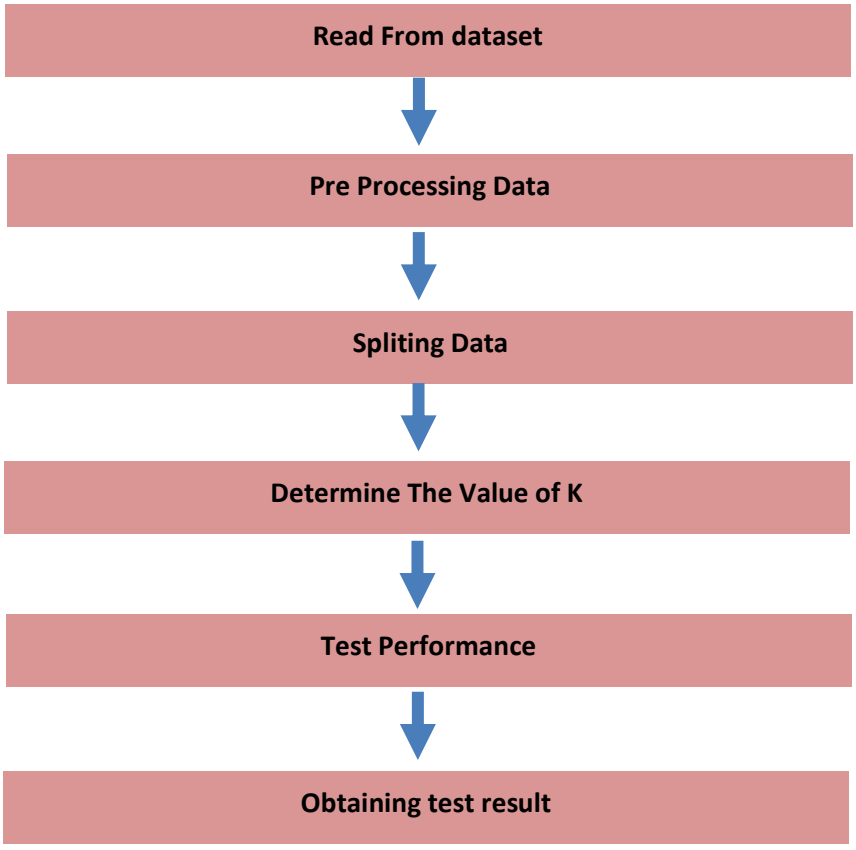


Figure 11. Research Stage

3.1.2. Spiliting Dataset

Before conducting performance testing, researchers divided the data, the distribution of which can be seen in the following table:

Table 6. Distribution Data Training dan Testing.

Data	Count
Training	491
Testing	123
Total	614 Data

In table 6, it can be concluded that from the total of 614 data, the data has been split between training data and testing data by 20%, resulting in a test value of 123 from a total of 614 data, while the remainder is training data of 491 data.

3.1.3. Testing Performance

From the results of this division, a model performance test was carried out with accuracy, f1_score, recall and precision values as shown in the following image:

	precision	recall	f1-score	support
0	0.88	0.45	0.60	33
1	0.83	0.98	0.90	90
accuracy			0.84	123
macro avg	0.86	0.72	0.75	123
weighted avg	0.84	0.84	0.82	123

Figure 12. Result Accuracy

The test results above yielded an accuracy of 84%, which is considered quite good. Furthermore, the researchers also obtained predictions using a confusion matrix table, a popular tool for model testing. The results are as follows:

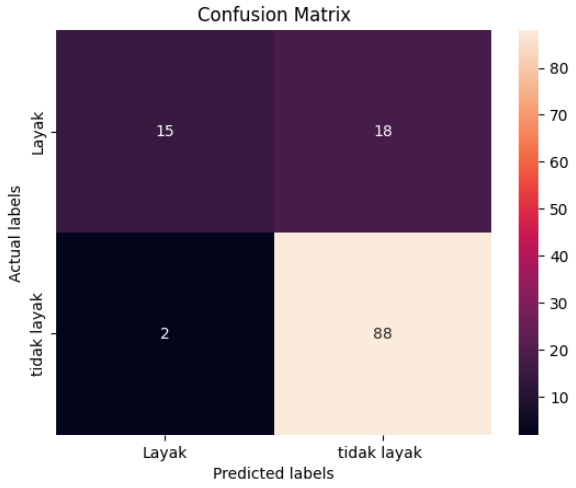


Figure 13. Confusion Metrics

In Figure 3. Assessment metrics clarify the performance of a model. On the off chance that we have a recorded dataset of client agitates, in the wake of preparing the model we need to ascertain the exactness utilizing the test set. (Hemachandran et al., 2021). Shows the visualization or plot of the model test results, if the calculation is carried out according to the formula regarding confusion metrics using TP, TN, FP, FN, then the following matrix results can be interpreted that the number 15 is actually feasible and predicted to be feasible, 88 is actually not feasible and predicted to be not feasible, 18 is actually feasible, but predicted to be not feasible, 2 is actually not feasible, but predicted to be feasible

The following is an overview of the confusion metrics table:

Table 7. Tabel Confusion Metrics

	Worthy	Not Feasible
Worthy	True Positif	False Positif
Not Feasible	False Negatif	True Negatif

	Worthy	Not Feasible
Worthy	TP	FP
Not Feasible	FN	TN

Conclusions

The research conducted revealed that the performance test results for the classification model used by the researchers showed a fairly good score. The accuracy score for this dataset was 84%. The application developed from this research also successfully produced predictions using the Euclidean distance calculation, which is often used in predicting events using the KNN algorithm.

The conclusion of this research is that it has a significant impact on managing data related to cooperative membership. This is achieved through converting categorical data to numerical data and preprocessing. The accuracy results for this study can be considered quite good using the KNN algorithm. However, further developments, such as testing with other algorithm models, are possible. In this study, it cannot be said that the amount of data is quite large because it only uses hundreds of data with a total of 10 attributes. The results of recall, precision, and f1-score in this study each show figures of 0.45 for class (0), 0.85 for class (1), 0.88 for class (0), 0.83 for class (1).

Acknowledgments

The author would like to express his deepest gratitude to all parties who have assisted in the completion of this research.

References

AlOmair, O., M. Zakaria, O., Alabdullatif, M., & Khurshid, Z. (2026). Optimized KNN with domain-informed features and LIME explainability for improved breast cancer classification. *BMC Medical Informatics and Decision Making*, *26*(1). https://doi.org/10.1186/s12911-026-03427-y

Anwar, A. N., Dani, D., & Alim, C. (2026). Analysis of the K-Nearest Neighbor (K-NN) Algorithm Method for Student Extracurricular Recommendations. *Bit-Tech*, *9*(1), 1567–1577. https://doi.org/10.32877/bt.v9i1.4146

Boateng, E. Y., Otoo, J., & Abaye, D. A. (2020). Basic Tenets of Classification Algorithms K-Nearest-Neighbor, Support Vector Machine, Random Forest and Neural Network: A

Review. *Journal of Data Analysis and Information Processing*, *08*(04), 341–357. https://doi.org/10.4236/jdaip.2020.84020

Gourav Rahangdale, Dr. Mahesh Motwani, & Mr. Manish Ahirwar. (2016). Application of k-NN and Naïve Bayes Algorithm in Banking and Insurance Domain. *International Journal of Computer Science Issues*, *13*(5), 69–75. https://doi.org/10.20943/01201605.6975

Hatem, M. Q. (2022). Skin lesion classification system using a K-nearest neighbor algorithm. *Visual Computing for Industry, Biomedicine, and Art*, *5*(1). https://doi.org/10.1186/s42492-022-00103-6

Hemachandran, K., George, P. M., Rodriguez, R. V., Kulkarni, R. M., & Roy, S. (2021). Performance analysis of k-nearest neighbor classification algorithms for bank loan sectors. *Advances in Parallel Computing*, *38*, 9–13. https://doi.org/10.3233/APC210004

Kherif, O., Benmahamed, Y., Teguar, M., Boubakeur, A., & Ghoneim, S. S. M. (2021). Accuracy Improvement of Power Transformer Faults Diagnostic Using KNN Classifier With Decision Tree Principle. *IEEE Access*, *9*, 81693–81701. https://doi.org/10.1109/ACCESS.2021.3086135

Pandey, D., Niwaria, K., & Chourasia, B. (2008). Machine Learning Algorithms: A Review. *International Research Journal of Engineering and Technology*. www.irjet.net

Pedregosa FABIANPEDREGOSA, F., Michel, V., Grisel OLIVIERGRISEL, O., Blondel, M., Prettenhofer, P., Weiss, R., Vanderplas, J., Cournapeau, D., Pedregosa, F., Varoquaux, G., Gramfort, A., Thirion, B., Grisel, O., Dubourg, V., Passos, A., Brucher, M., Perrot andÉdouardand, M., Duchesnay, andÉdouard, & Duchesnay EDOUARDDUCHESNAY, Fré. (2011). Scikit-learn: Machine Learning in Python Gaël Varoquaux Bertrand Thirion Vincent Dubourg Alexandre Passos PEDREGOSA, VAROQUAUX, GRAMFORT ET AL. Matthieu Perrot. In *Journal of Machine Learning Research* (Vol. 12). http://scikit-learn.sourceforge.net.

Pfeifer, B., & Kreuzthaler, M. (2026). Calibrated kNN classification via second-layer neighborhood analysis. *Advances in Data Analysis and Classification*, *20*(1), 145–165. https://doi.org/10.1007/s11634-025-00654-5

Uddin, S., Haque, I., Lu, H., Moni, M. A., & Gide, E. (2022). Comparative performance analysis of K-nearest neighbour (KNN) algorithm and its different variants for disease prediction. *Scientific Reports*, *12*(1). https://doi.org/10.1038/s41598-022-10358-x

Wang, L. (2019). Research and Implementation of Machine Learning Classifier Based on KNN. *IOP Conference Series: Materials Science and Engineering*, *677*(5). https://doi.org/10.1088/1757-899X/677/5/052038

Yang, F., Du, J., Lang, J., Lu, W., Liu, L., Jin, C., & Kang, Q. (2020). Missing Value Estimation Methods Research for Arrhythmia Classification Using the Modified Kernel Difference-Weighted KNN Algorithms. *BioMed Research International*, *2020*. https://doi.org/10.1155/2020/7141725

Zhang, S., & Li, J. (2023). KNN Classification With One-Step Computation. *IEEE Transactions on Knowledge and Data Engineering*, *35*(3), 2711–2723. https://doi.org/10.1109/TKDE.2021.3119140

Zhang, S., Li, X., Zong, M., Zhu, X., & Wang, R. (2018). Efficient kNN classification with different numbers of nearest neighbors. *IEEE Transactions on Neural Networks and Learning Systems*, *29*(5), 1774–1785. https://doi.org/10.1109/TNNLS.2017.2673241