Research Article

Design of K-Nearest Neighbor Algorithm For Classification of Credit Loan Eligibility At Senarak Dana Purwakarta Cooperative

Imay Kurniawan[1*], Purwadi Budi Santoso[2]

[1] Department of Informatic Engineering, Sekolah Tinggi Teknologi Wastukancana Purwakarta, Indonesia
[2] Department of Informatic Engineering, Sekolah Tinggi Teknologi Mandala Bandung, Indonesia

* Corresponding Author: imaykurniawan@wastukancana.ac.id

Abstract: Semarak Dana Cooperative is a savings and loan cooperative located in Purwakarta that has experience lending money to its members as many as 162 money lending transactions. However, there are 26 instances of bad debts. To avoid bad debts, the cooperative needs to classify loans to its members. Classification is done by using the K-Nearest Neighbor (KNN) method based on the attributes of employment, income, age, credit amount, term, and collateral value. Data taken from as many as 162 members are sorted into 2 parts, namely 149 transactions used as training data and 13 transactions used as testing data. In addition, the data is also sorted into two classes, namely 136 current classes and 26 bad classes. The KNN process consists of four stages. First, determine the parameter K nearest neighbor distance. The second stage is to calculate the distance between testing data and training data using Euclidean distance. The third stage sorts the distance data that has been calculated using selection sort in order from the smallest to the largest value of K. The fourth stage calculates the largest number of classes for the largest number of classes set as the classification result class. Implementation using Borland Delphi and Mysql database. The research method was used by applying the Waterfall method. The Waterfall method used is composed of analysis, design, coding, and testing. System design using Unified Modeling Language (UML) by describing use case diagram, activity diagram, and class diagram. Based on the confusion matrix of the KNN classification process, the percentage of accuracy is 77%, precision is 88%, and recall is 78%. These results can be said that the results obtained are quite good, which exceeds 70%.

Keywords: Classification; KNN; training data; testing data; euclidean distance; selection sorting; Borland Delphi; MySQL, Waterfall, UML, confusion matrix

Citation: I.Kurniawan, P. B.Santoso, "Design of K-Nearest Neighbor Algorithm For Classification of Credit Loan Eligibility At Senarak Dana Purwakarta Cooperative", Iota, 2024, ISSN 2774-4353, Vol.04,02.https://doi.org/10.31763/iota.v4i2.742

Academic Editor : Adi, P.D.P

Received : March, 30 2024

Accepted : April, 21 2024

Published : May, 22 2024

1. Introduction

Cooperatives are one of the places to provide loans or credit to its members. Many members apply for loans or credit with different economic conditions which can cause various kinds of problems in the cooperative. One of them is the occurrence of bad credit. One of the services of the Semarak Dana Purwakarta cooperative is to provide money loans on credit to its members, based on previous experience from 162 money lending transactions there were 26 members with bad credit. To minimize the potential for bad debts, an analysis is needed to classify previous data based on the benchmarks of the attributes of employment, income, age, credit amount, loan repayment period, and the value of member collateral. So that in the future it does not cause unwanted credit risks. This study uses the K-Nearest Neighbor (KNN) classification algorithm [1,2,3,4,5] to classify the feasibility of borrowing money on credit from members of the Semarak Dana Purwakarta cooperative.

2. Theory

Design is a series of procedures to translate the results of analysis and a system into a programming language to describe in detail how the components of the system are to describe in detail how the system components are implemented. implemented. The definition of development or building a system is activities to create new systems or replace or improve existing systems either in whole or in part. system that already exists either in whole or in part [1]. Moreover, Classification in data mining is a data learning method to predict the value of a group of attributes. The classification algorithm will produce a set of rules called rules which will be used as indicators to be able to predict the class of data to be predicted [2]. Moreover, The K-Nearest Neighbor (K-NN) algorithm [6,7,8,9,10,11,12] is a method for classifying objects based on learning data that is closest to the object. The K-NN algorithm uses a supervised algorithm [13,14,15,16,17]. The algorithm stages consist of four stages:

1. Determine the parameter K = number of nearest neighbors.
2. Calculate the distance between the data to be evaluated and all training data.
3. Sort the distances formed (ascending order) and determine the closest distance to the Kth order.
4. Find the largest number of classes from the nearest neighbors, and set that class as the evaluated data class. as the evaluated data class.

The classification process using the KNN method can be seen in the following example some data is coming from a questionnaire survey about the classification of tissue paper quality whether good or bad [18,19,20,21,22,23], with training objects using two attributes of durability and strength. Namely, acid resistance and strength using k = 4. Table 1 can be seen as training data.

Table 1. Training Data

X1= durability	X2= strength	Classification
8	4	Good
4	5	Bad
4	6	Bad
7	7	Good
5	6	Bad
6	5	Good

Test Data: The tissue paper will be produced again with attributes X1 = 7 and X2 = 4 without having to spend money to conduct a survey, then it can be classified as good or bad tissue paper.

Step 1: Determine the value of k, where k = 4 is selected

Step 2: Calculate the distance between the data to be evaluated and all training data. The results can be seen in Table 2.

Table 2. Calculating Distance

X1=resilience	X2=strength	Distance
8	4	$(8-7)^2 + (4-4)^2=1$
4	5	$(4-7)^2 + (5-4)^2=10$
4	6	$(4-7)^2 + (6-4)^2=13$
7	7	$(7-7)^2 + (7-4)^2=9$
5	6	$(5-7)^2 + (6-4)^2=8$
6	5	$(6-7)^2 + (5-4)^2=2$

Moreover, Step 3: Sort the smallest data distance to the largest and pair each class, the results can be seen in Table 3.

Table 3. Order of Smallest to Largest Distance

X1= Durability	X2= Power	Distance	Classification
8	4	1	Good
4	5	2	Good
4	6	8	Bad
7	7	9	Good
5	6	10	Bad
6	5	13	Bad

Step 4: Find the largest number of classes from the nearest neighbors, and assign that class as the evaluated data class. Since k=4, then the order 1 to 4 is chosen where the number of good =3 and the number of bad = 1, then it is decided that the result of classification result is good. Credit is the delivery of goods, services, or money from one party (creditor/lender) based on trust to another party (customer or debtor/borrower) with a promise to pay from the credit recipient to the credit provider on the date agreed by both parties [4].

Unified Modeling Language (UML) [24,25] is a standard language for designing software in the form of modeling ranging from enterprise information systems to distributed web-based applications and even to real-time embedded systems [5]. Modeling or designing systems with UML can connect directly to various programming languages, such as Java, C++, or Visual Basic, or even to tables in a relational database or persistent storage of an object-oriented database [5]. In other words, UML models can be applied to many different types of software applications. The term "use case" refers to the complete sequence of events in the system as understood from the user's perspective [6]. In other words, a use case is a set of sequences of events where each sequence represents the interaction of objects outside the system (its actors) with the system itself (and its key abstractions) [5].

Activity diagrams are graphically used to describe a series of Use Case activity flows. Activity Diagrams can be used to model the actions that will be performed when an operation is executed and model the results of these actions [10]. The class diagram is a set of similar objects. An object has an instantaneous state (state) and behavior (behavior). The state of an object is the condition of the object expressed in attributes/properties. While the behavior of an object defines how an object acts and reacts [10].

With Borland Delphi we can develop GUI-shaped applications, in general, Delphi is more widely used for the development of database-based desktop and enterprise applications, but as a general-purpose development tool, it is also capable and used in various types of software development projects. The consideration of using Delphi is because this research requires fast execution to secure text file data where the size of text file data varies from small to large, this can be done by Delphi one of the advantages of Delphi compiler optimization is very fast [8]. MySQL is the most widely used relational database system in the Open Source sector with advantages: fast, stable, easy to learn, compatible with popular OSs, and supported by a variety of programming languages [9]. A confusion matrix is a data set that only has two classes, one class as positive and the other as negative. The confusion matrix contains information comparing the classification result label with the actual label [11]. Seen in Table 4 about the confusion matrix.

Table 4. Confusion Matrix

Class	Positively Classified	Negatively Classified
Positive	TP (True Positive)	TN (True Negative)
Negative	FP (False Positive)	FN (False Negative)

Furthermore, Selection Sort is also an important part of this research. Selection Sort is one method to sort data either the smallest data to the largest or vice versa from the largest data to the smallest. The data sorting of the selection sort method is better than the bubble sort method [14]. The way this method works is based on finding the element with the smallest value, and then exchanging it with the Ith element. Briefly, this method can be explained as follows. In the first step, we look for the smallest data from the first data to the last data. Then, we swap that data with the first data. Thus, the first data now has the smallest value compared to the other data. In the second step, we look for the smallest data from the second to the last data. The smallest data we get we swap with the second data. And so on until all the data is sorted. An example of the Sorting process using the Selection Sort method can be seen in Table 5.

Table 5. Selection Sort Process Example

Iteration to	A [1]	A [2]	A [3]	A [4]	A [5]	A [6]
Initial	22	10	15	3	2	8
I=1,Lok=5	2	10	15	3	22	8
I=2,Lok=4	2	3	15	10	22	8
I=3,Lok=6	2	3	8	10	22	15
I=4,Lok=4	2	3	8	10	22	15
I=5,Lok=6	2	3	8	10	15	22

3. Method

This research method uses the Waterfall model until the fourth stage, 1) Analysis, 2) Design, 3) Coding, and 4) Testing. 3.1 System Analysis 3.1.1 Flowmap of the Current System The following is the flowmap of the current system at Semarak Dana Purwakarta Cooperative: 1. Narrative of the Current System a. Members fill out the loan data form to be submitted b. The leader analyzes the loan from the results of the loan application and loan data. c. If it is not complete, return to the member, if it is complete, continue to register for a credit loan. d. Registration files are archived. 2. Flowmap of Current System The Flowmap of Current System can be seen in Figure 1.

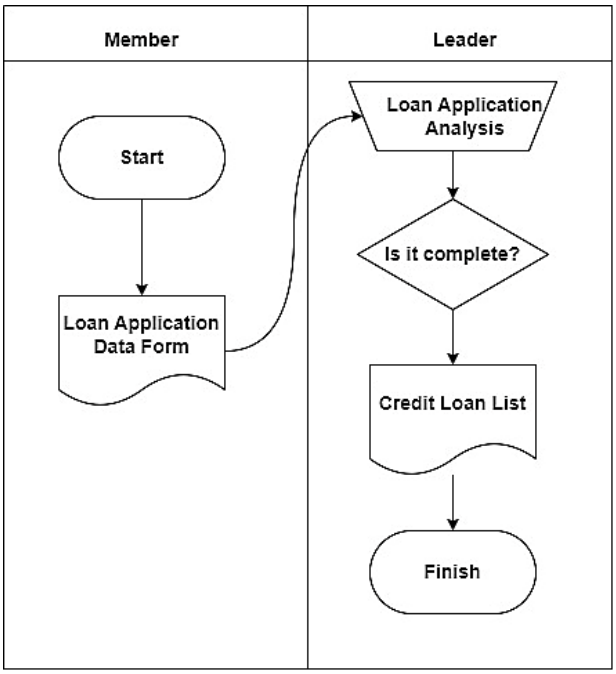


Figure 1. Current System Flowmap

Furthermore, a Flowmap of the Proposed System The following is a flowmap of the system that is being proposed at Semarak Dana Purwakarta Cooperative: 1. Narrative of the proposed system (a). Members fill out the loan data form to be submitted. (b). Administration inputs training data and saves training data. (c). Training data results. (d). Administration inputs testing data and saves testing data (e). The results of the testing data, then the classification process. (f). The results of the classification process, then the testing process. The results of the classification process, then the testing process. and (g). Proposed System Flowmap The Proposed System Flowmap can be seen in Figure 2.

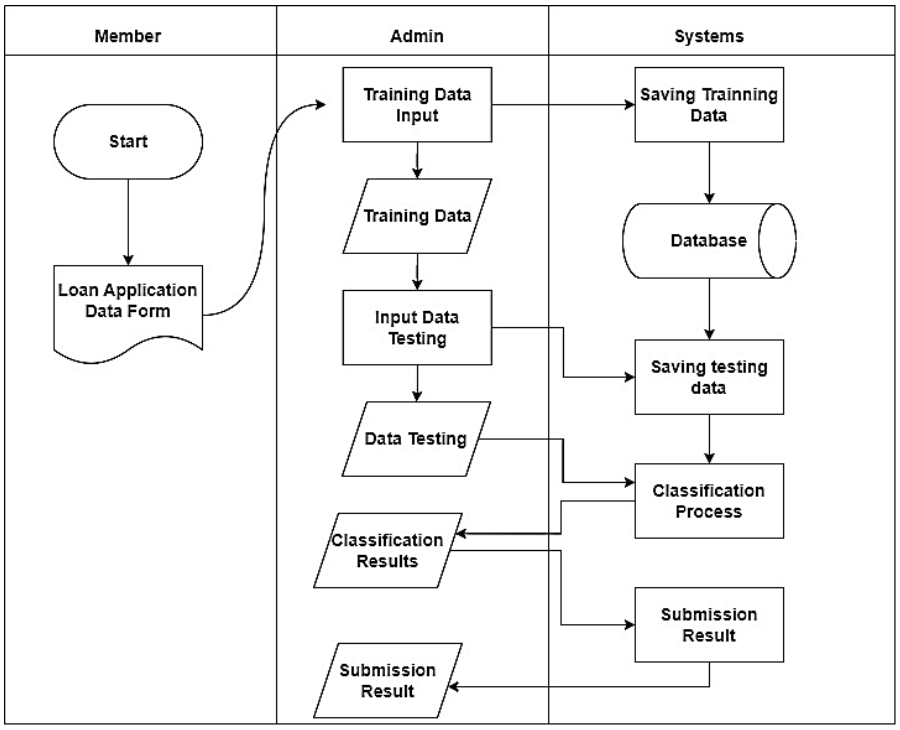


Figure 2. Flowmap of Proposed System

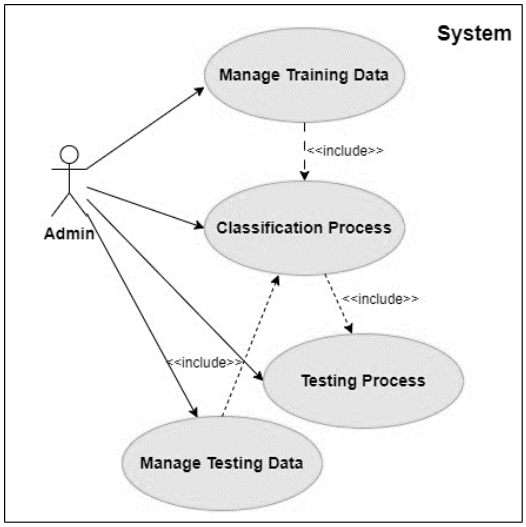
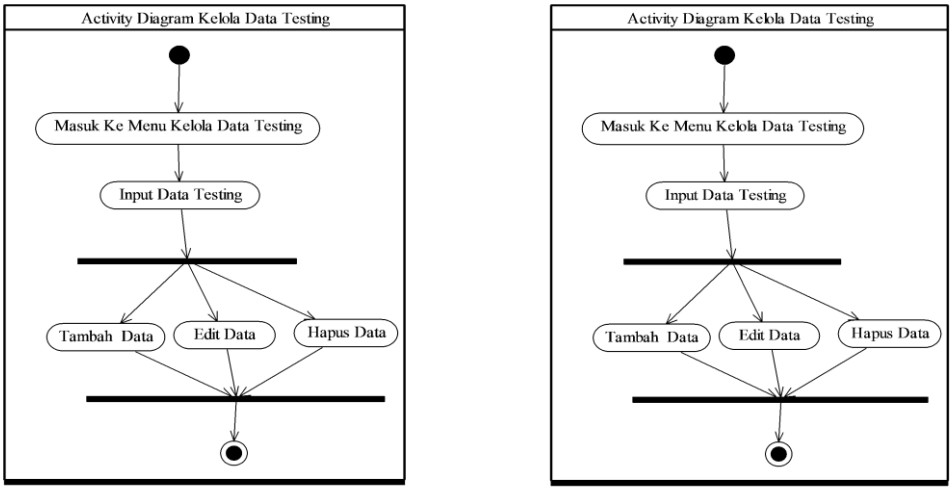
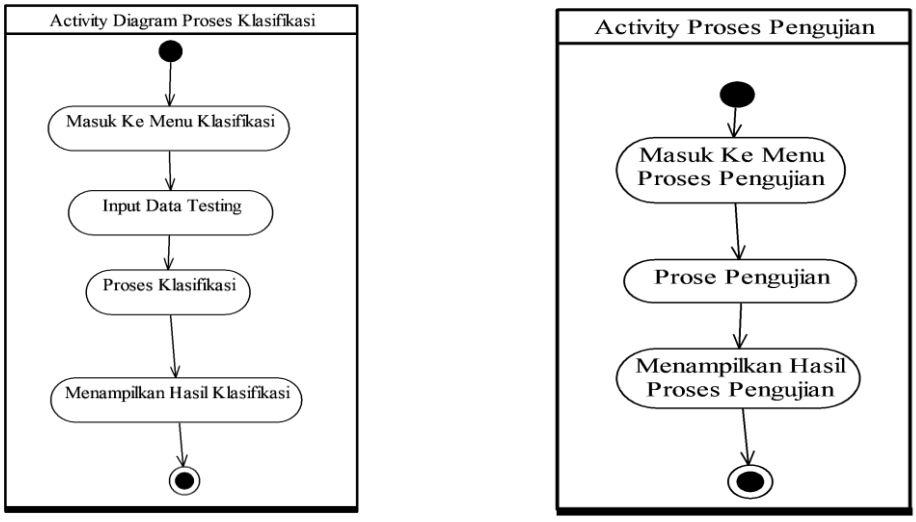


Figure 3. Use Case Diagram



(a) (b)

Figure 4. (a) Activity Diagram Manage Training Data, (b) Activity Diagram Manage Testing Data



(a) (b)

Figure 5. (a) Activity Diagram of Classification Process, (b) Activity Diagram of Testing Process

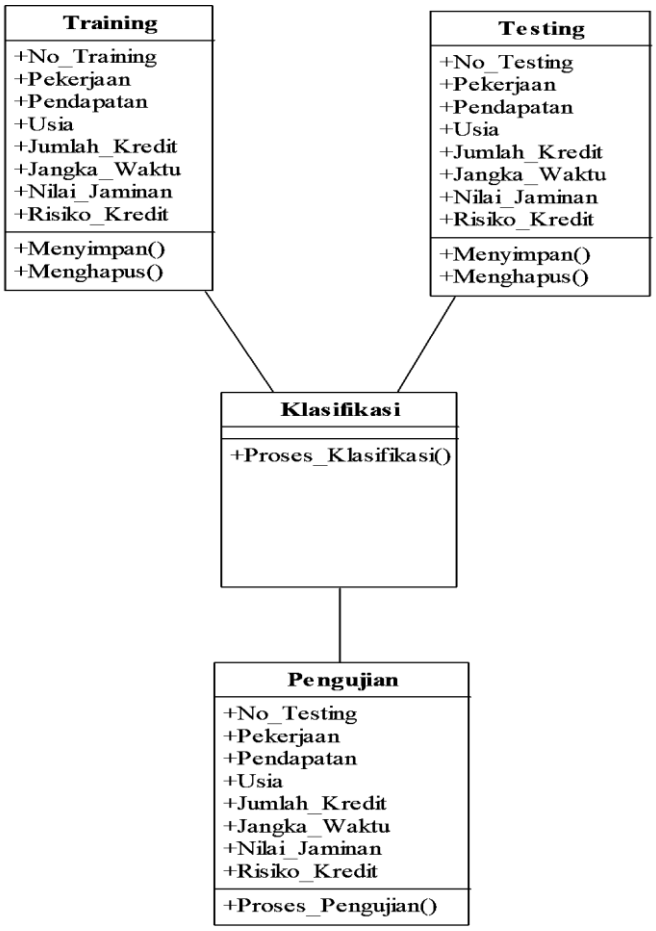


Figure 6. Class Diagram

4. Result and Discussion

In this Result and Discussion session, we will comprehensively discuss how the Graphical User Interface is built. The GUI can be seen from various menus. The menu form includes training data, testing data, classification, and testing can be seen in Figure 7. Moreover, The training data form is used to manage training data which includes adding training data, editing training data, and deleting training data can be seen in Figure 8. The testing data form is used to manage testing data which includes adding testing data, editing testing data, and deleting testing data can be seen in Figure 9.

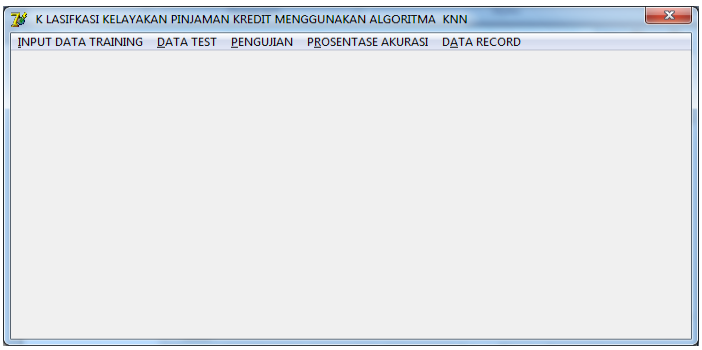


Figure 7. Form Menu

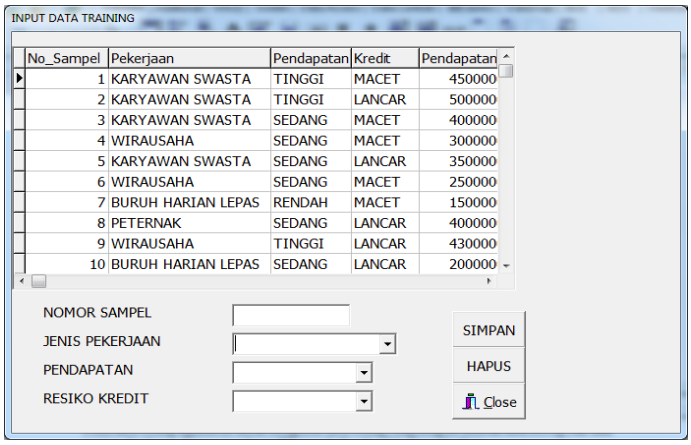


Figure 8. Form Data Training

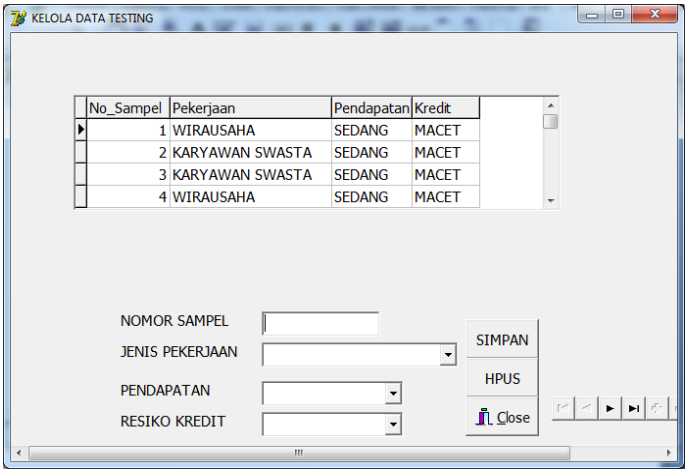


Figure 9. Form Data Testing

The classification form is used to classify credit risk. The classification results can be used to determine whether potential credit recipient customers will have the potential for current credit or bad credit can be seen in Figure 10.

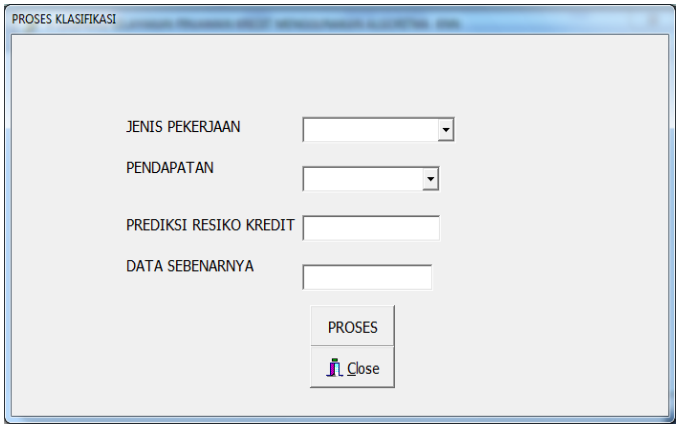


Figure 10. Classification Process Form

Testing Form The test form is used to compare the classification results of the KNN algorithm with the original data, the process can be seen in Figure 11. Moreover, at this stage, the author conducts testing to analyze whether this system still has errors that must be corrected again, this test focuses on the functional specifications of the system, testing can be seen in Tables 6, 7, 8, 9, and 10.

Table 6. Menu Form Testing

No.	Test scenario	Expected results	Conclusion
1	Click on the training data menu	The system displays the training data form	valid
2	Click on the testing data menu	The system will display the testing data form	valid
3	Click on the classification menu	The system displays the classification data form	valid
4	Click the test menu	The system displays the test data form	valid

Table 7. Training Data Form Testing

No.	Test scenario	Expected results	Conclusion
1	Leave the training data input blank or partially inputted and click the save button.	The system will display a message that the training data input has not been inputted all	valid
2	input all training data then click the save button	Training data is stored in the database, then the results are displayed on the training data form.	valid
3	Input training number data and click the delete button	Data is deleted from the database, then the results are displayed in the training data form.	valid

Table 8. Testing Data Testing Form

No.	Test scenario	Expected results	Conclusion
1	Leave the testing data input blank or partially inputted and click the save button.	The system will display a message that the testing data input has not been inputted.	valid
2	input all testing data then click the save button	The testing data is stored in the database, then the results are displayed on the testing data form.	valid
3	Input testing number data then click the delete button	Data is deleted from the database, then the results are displayed on the testing data form.	valid

Table 9. Classification Form Testing

No.	Test scenario	Expected results	Conclusion
1	Leave the testing data input blank or partially inputted and click the process button.	The system will display a message that the testing data input has not been inputted.	valid
2	Input all testing data then click the process	The system will display the results of the KNN algorithm classification stage process. Then the results are stored in the test table	valid

Table 10. Test Form Testing

No.	Test scenario	Expected results	Conclusion
1	Click raffle progress	The system will display the percentage of classification accuracy of the KNN algorithm.	valid

Furthermore, a Measuring the Performance of the KNN Classification Algorithm with a Confusion Matrix. In research on the classification of credit risk of members of the Semarak Dana Purwakarta cooperative using training data of 149 while testing data of 13. Some training data can be seen in Table 11.

Table 11. Data Training

No.	Jobs	Revenue	Age	Total Credit	Period	Value Guarantee	Risk Credit
1.	Private Employee	2 million -5 million IDR	21-35	1 million -3 million IDR	12 months	3 million -6 million IDR	bad debts
2.	Private Employee	2 million -5 million IDR	21-35	3 million -6 million IDR	12 months	3 million -6 million IDR	Current Credit
3.	Private Employee	2 million -5 million IDR	36-50	3 million -6 million IDR	12 months	3 million -6 million IDR	bad debts
4.	Entrepreneurship	2 million -5 million IDR	51-65	< 1 million IDR	12 months	< 1 million IDR	bad debts
5.	Private Employee	2 million -5 million IDR	21-35	1 million - 3 million IDR	12 months	< 1 million IDR	Current Credit
6.	Entrepreneurship	< 2 million IDR	21-35	1 million - 3 million IDR	12 months	< 1 million IDR	bad debts
7.	Labor	< 2 million IDR	36-50	1 million - 3 million IDR	12 months	< 1 million IDR	Current Credit
8.	Breeders	2 million -5 million IDR	36-50	1 million - 3 million IDR	12 months	1 million - 3 million IDR	Current Credit
9.	Entrepreneurship	< 2 million IDR	36-50	3 million -6 million IDR	12 months	3 million -6 million IDR	Current Credit
|	|	|	|	|	|	|	|
|	|	|	|	|	|	|	|
146	ARMED FORCES	2 million -5 million IDR	36-50	3 million -6 million	12 months	3 million -6 million IDR	Current Credit
147	Retired	2 million -5 million IDR	51-65	3 million -6 million	12 months	3 million -6 million IDR	Current Credit
148	Private Employee	2 million -5 million IDR	21-35	3 million -6 million	12 months	3 million -6 million IDR	Current Credit
149	Private Employee	2 million -5 million IDR	21-35	3 million -6 million	12 months	1 million - 3 million IDR	Current Credit

Table 12. Data Testing

No.	Jobs	Revenue	Age	Total Credit	Period	Value Guarantee	Risk Credit
1.	Entrepreneurship	2 million -5 million IDR	21-35	1 million -3 million IDR	3 months	< 1 million IDR	bad debts
2.	Police	2 million -5 million IDR	36-50	3 million -6 million IDR	12 months	3 million -6 million IDR	Current Credit
3.	Breeders	2 million -5 million IDR	36-50	1 million -3 million IDR	3 months	< 1 million IDR	bad debts
4.	Private Employee	>5 million IDR	21-35	1 million -3 million IDR	12 months	3 million -6 million IDR	Current Credit
5.	Entrepreneurship	2 million -5 million IDR	21-35	1 million -3 million IDR	3 months	1 million - 3 million IDR	Current Credit
6.	Private Employee	>5 million IDR	21-35	1 million -3 million IDR	12 months	3 million - 6 million IDR	Current Credit
7.	Entrepreneurship	2 million -5 million IDR	21-35	1 million -3 million IDR	3 months	3 million - 6 million IDR	bad debts
8.	Private Employee	2 million -5 million IDR	21-35	< 1 million IDR	3 months	< 1 million IDR	bad debts
9.	Farmers	2 million -5 million IDR	36-50	3 million -6 million IDR	12 months	3 million -6 million IDR	Current Credit
10.	Police	2 million -5 million IDR	21-35	3 million -6 million IDR	12 months	1 million - 3 million IDR	Current Credit
11.	Labor	2 million -5 million IDR	21-35	< 1 million IDR	3 months	< 1 million IDR	Current Credit
12.	Labor	< 2 million IDR	21-35	1 million -3 million IDR	3 months	< 1 million IDR	Current Credit
13.	Retired	2 million -5 million IDR	66-75	3 million -6 million IDR	12 months	3 million - 6 million IDR	Current Credit

Table 13. Data Training

No.	Jobs	Revenue	Age	Total Credit	Period	Value Guarantee	Risk Credit
1.	1	2	1	2	2	3	bad debts
2.	1	2	1	3	2	3	Current Credit
3.	1	2	2	3	2	3	bad debts
4.	2	2	3	1	1	1	bad debts
5.	1	2	1	2	1	1	Current Credit
6.	2	1	1	2	1	1	bad debts
7.	3	1	2	2	1	1	Current Credit
8.	4	2	2	2	1	2	Current Credit
9.	2	1	2	3	2	2	Current Credit
|	|	|	|	|	|	|	|
|	|	|	|	|	|	|	|
146	9	2	2	3	2	3	Current Credit
147	7	2	3	3	2	3	Current Credit
148	1	2	1	3	2	3	Current Credit
149	1	2	1	3	2	2	Current Credit

Table 14. Data Testing

No.	Jobs	Revenue	Age	Total Credit	Period	Value Guarantee	Risk Credit
1.	2	1	2	1	1	2	bad debts
2.	8	2	2	3	2	3	Current Credit
3.	4	2	2	2	1	1	bad debts
4.	1	3	1	2	2	3	Current Credit
5.	2	2	1	2	1	2	Current Credit
6.	1	3	1	3	2	3	Current Credit
7.	2	2	1	2	1	3	bad debts

No.	Jobs	Revenue	Age	Total Credit	Period	Value Guarantee	Risk Credit
8.	1	2	1	1	1	1	bad debts
9.	5	2	2	3	2	3	Current Credit
10.	7	2	4	3	2	2	Current Credit
11.	3	2	1	1	1	1	Current Credit
12.	3	1	1	2	1	1	Current Credit
13.	7	2	4	3	2	3	Current Credit

Based on the training data of 149 in Table 14, the testing data in Table 15 can be classified as to whether they have the potential for current or bad credit using the KNN classification method.

Table 15. Test data table number 1

No.	Jobs	Revenue	Age	Total Credit	Period	Value Guarantee
1.	2	1	2	1	1	2

The solution steps are as follows: Step 1: select k=7 Step 2: Calculate the distance between the data to be evaluated and all training data. The results can be seen in Table 16.

Table 16. Distance Calculation Result

No.	Distance
1.	$\sqrt{(1-2)^2+(2-1)^2+(1-2)^2+(2-1)^2+(2-1)^2+(3-2)^2} = 2{,}45$
2.	$\sqrt{(1-2)^2+(2-1)^2+(1-2)^2+(3-1)^2+(2-1)^2+(3-2)^2} = 3$
3.	$\sqrt{(1-2)^2+(2-1)^2+(2-2)^2+(3-1)^2+(2-1)^2+(3-2)^2} = 2{,}83$
4.	$\sqrt{(2-2)^2+(2-1)^2+(3-2)^2+(1-1)^2+(1-1)^2+(1-2)^2} = 1{,}73$
5.	$\sqrt{(1-2)^2+(2-1)^2+(1-2)^2+(2-1)^2+(1-1)^2+(1-2)^2} = 1$
6.	$\sqrt{(2-2)^2+(1-1)^2+(1-2)^2+(2-1)^2+(1-1)^2+(1-2)^2} = 1$
7.	$\sqrt{(3-2)^2+(1-1)^2+(2-2)^2+(2-1)^2+(1-1)^2+(1-2)^2} = 1{,}73$
8.	$\sqrt{(4-2)^2+(2-1)^2+(2-2)^2+(2-1)^2+(1-1)^2+(2-2)^2} = 2$
9.	$\sqrt{(2-2)^2+(1-1)^2+(2-2)^2+(3-1)^2+(2-1)^2+(2-2)^2} = 2{,}24$
|	|
|	|
146	$\sqrt{(9-2)^2+(2-1)^2+(2-2)^2+(3-1)^2+(2-1)^2+(3-2)^2} = 7{,}48$
147	$\sqrt{(7-2)^2+(2-1)^2+(3-2)^2+(3-1)^2+(2-1)^2+(3-2)^2} = 5{,}92$
148	$\sqrt{(1-2)^2+(2-1)^2+(1-2)^2+(3-1)^2+(2-1)^2+(3-2)^2} = 2{,}65$
149	$\sqrt{(1-2)^2+(2-1)^2+(1-2)^2+(3-1)^2+(2-1)^2+(2-2)^2} = 2$

Due to space limitations, the results of the calculation of the distance between the 1st testing data and the training data number 10 to 145 are not displayed. By using Borland Delphi, the results of the calculation of the distance between the 1st testing data and the training data number 1 to 149 can be obtained. Steps 3 and 4: sort the data from the smallest to the largest value using the selection sort method then take the 7 smallest values, and then pair each class the results can be seen in Table 17.

Table 17. Distance relationship between Testing data and Training Data

No. Data Training	Distance Between Testing Data and Training Data	Risk Credit
15	1,41	Current Credit
121	1,41	bad debts
143	1,41	Current Credit
4	1,73	bad debts
7	1,73	bad debts
42	1,73	Current Credit
43	1,73	bad debts

Step 5: Find the largest number of classes. credit risk there are two classes, namely current and bad. The number of bad classes (4) is more than the number of current classes. (3). So it can be decided that the classification result is bad. In the same way for the 2nd to 13th testing data, the results can be obtained. Overall can be seen in Table 18.

Table 18. Classification Results of Testing Data

Data Testing	Prediction	Actual
1	bad debts	bad debts
2	Current Credit	Current Credit
3	Current Credit	bad debts
4	Current Credit	Current Credit
5	bad debts	Current Credit
6	Current Credit	Current Credit
7	bad debts	bad debts
8	bad debts	bad debts
9	Current Credit	Current Credit
10	Current Credit	Current Credit
11	Current Credit	Current Credit
12	bad debts	Current Credit
13	Current Credit	Current Credit

From Table 18 the results obtained: TP = 7; TN = 3 FP = 1; FN = 2, For the Confusion Matrix Table can be seen in Table 19.

Table 19. Confusion Matrix

Class	Positively Classified	Negatively Classified
Positive	7	2
Negative	1	3

Accuracy = (TP+TN)/(TP+TN+FP+FN) x 100%= (7+3)/(7+3+1+2) x100% = 77% Precision = TP/(TP+FP) x 100%= 7/(7+1) x100% = 88% Recall = TP/(TP+FN) x 100%= 7/(7+2) x100% = 78%

5. Conclusion

This research has used a variety of methods that can be used to get the most accurate results. The Waterfall method used in this research can show the work steps in this research until the testing stage and it is hoped that the system built can provide accurate results from the predictions that are being built. This waterfall method consists of analysis, design, coding, and testing. System design using Unified Modelling Language (UML) by describing use case diagrams, activity diagrams, and class diagrams, with this UML being able to provide accurate results from the system being built. Then the KNN Algorithm can provide accurate results as well. The KNN algorithm can be used to conduct creditworthiness, and the results of confusion matrix testing obtained an accuracy percentage of 69%, a precision percentage of 87.5%, and a recall percentage of 70%.

Acknowledgments: Thanks to colleagues at the Department of Informatic Engineering, Wastukancana College of Technology Purwakarta, Indonesia, and colleagues at the Department of Informatic Engineering, Mandala College of Technology Bandung, Indonesia, this research is the result of collaboration from fellow lecturers, hopefully, this research can be a good reference for the same discipline.

Author contributions: All authors are responsible for building Conceptualization, Methodology, analysis, investigation, data curation, writing—original draft preparation, writing—review and editing, visualization, supervision of project administration, funding acquisition, and have read and agreed to the published version of the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Y. Manzali, K. A. Barry and M. E. Far, "An improved KNN algorithm based on ensemble methods and correlation," 2023 7th IEEE Congress on Information Science and Technology (CiSt), Agadir - Essaouira, Morocco, 2023, pp. 64-70, doi: 10.1109/CiSt56084.2023.10409922. [Crossref]
2. H.Wang, P.Xu, J.Zhao, 2022, "Improved KNN algorithms of spherical regions based on clustering and region division", Alexandria Engineering Journal, Volume 61, Issue 5, May 2022, Pages 3571-3585, DOI: 10.1016/j.aej.2021.09.004 [Crossref]
3. A. M, A. K, A. K, A. M, and C. K. R, "Accident Prediction Using KNN Algorithm," 2022 Fourth International Conference on Emerging Research in Electronics, Computer Science and Technology (ICERECT), Mandya, India, 2022, pp. 1-5, doi: 10.1109/ICERECT56837.2022.10059746 [Crossref]
4. F.J.Gomez-Gil, et. al, 2024, "Vibration-based monitoring of agro-industrial machinery using a k-Nearest Neighbors (kNN) classifier with a Harmony Search (HS) frequency selector algorithm", Computers and Electronics in Agriculture, Volume 217, February 2024, 108556, doi: 10.1016/j.compag.2023.108556 [Crossref]

5. Q. Gao, X. Ding, C. Li, and Z. T. Qi, "A Gait Recognition of EMG Signal Method Based on Bat Algorithm Optimized KNN," 2023 7th Asian Conference on Artificial Intelligence Technology (ACAIT), Jiaxing, China, 2023, pp. 16-20, doi: 10.1109/ACAIT60137.2023.10528639. [Crossref]
6. H. Xing and R. Zhang, "Gait recognition for exoskeleton robots based on improved KNN-DAGSVM fusion algorithm," 2022 37th Youth Academic Annual Conference of Chinese Association of Automation (YAC), Beijing, China, 2022, pp. 364-369, doi: 10.1109/YAC57282.2022.10023588. [Crossref]
7. Y. Wang, C. Tan, C. Qi, and H. Xiong, "Intelligent Analysis Framework of Power Marketing Big Data based on Multi-Dimensional KNN Algorithm," 2023 3rd International Conference on Smart Data Intelligence (ICSMDI), Trichy, India, 2023, pp. 435-438, doi: 10.1109/ICSMDI57622.2023.00082. [Crossref]
8. Y. Zhao and J. Zhang, "Research on KNN algorithm based on K-means clustering and collaborative filtering hybrid algorithm in AI teaching," 2023 8th International Conference on Information Systems Engineering (ICISE), Dalian, China, 2023, pp. 453-456, doi: 10.1109/ICISE60366.2023.00102. [Crossref]
9. J. Sun, "Research on Evaluating Teaching Quality of Accounting Courses Based on Improved K-nearest Neighbors Algorithm(KNN)," 2023 IEEE 3rd International Conference on Electronic Communications, Internet of Things and Big Data (ICEIB), Taichung, Taiwan, 2023, pp. 371-374, doi: 10.1109/ICEIB57887.2023.10170702. [Crossref]
10. S. G. A, R. Prabha, M. Razmah, T. Veeramakali, S. S and Y. R, "Machine Learning Heart Disease Prediction Using KNN and RTC Algorithm," 2022 International Conference on Power, Energy, Control and Transmission Systems (ICPECTS), Chennai, India, 2022, pp. 1-5, doi: 10.1109/ICPECTS56089.2022.10047501. [Crossref]
11. L. Yao and T. Cao, "Research on Talent Classification Based on Improved KNN Algorithm," 2022 International Symposium on Control Engineering and Robotics (ISCER), Changsha, China, 2022, pp. 252-256, doi: 10.1109/ISCER55570.2022.00050. [Crossref]
12. S. Garg, D. Dixit and S. S. Yadav, "Disease Detection in Plants Using KNN Algorithm," 2022 4th International Conference on Advances in Computing, Communication Control and Networking (ICAC3N), Greater Noida, India, 2022, pp. 938-943, doi: 10.1109/ICAC3N56670.2022.10074491. [Crossref]
13. I. Saini and V. Maheshkar, "Effects of Distance Metrics on Color Recognition using Color Histogram and KNN Algorithm," 2022 IEEE Delhi Section Conference (DELCON), New Delhi, India, 2022, pp. 1-5, doi: 10.1109/DELCON54057.2022.9753105. [Crossref]
14. Q. Zhang, X. Bo, J. Yang and C. Cui, "Indoor Positioning for the Elderly Based on Fuzzy C-Means Clustering and KNN Algorithm," 2024 5th International Seminar on Artificial Intelligence, Networking and Information Technology (AINIT), Nanjing, China, 2024, pp. 236-240, doi: 10.1109/AINIT61980.2024.10581418. [Crossref]
15. H. Zhai, "Improving KNN Algorithm Efficiency Based on PCA and KD-tree," 2022 International Conference on Machine Learning and Knowledge Engineering (MLKE), Guilin, China, 2022, pp. 83-87, doi: 10.1109/MLKE55170.2022.00021. [Crossref]
16. A. Kanan and A. Taha, "Cloud-Based Reconfigurable Hardware Accelerator for the KNN Classification Algorithm," 2022 14th International Conference on Computational Intelligence and Communication Networks (CICN), Al-Khobar, Saudi Arabia, 2022, pp. 308-312, doi: 10.1109/CICN56167.2022.10008343. [Crossref]
17. G. Deivasikamani, A. C, A. T, and R. C. Manoj, "Covid Cough Classification using KNN Classification Algorithm," 2022 International Conference on Applied Artificial Intelligence and Computing (ICAAIC), Salem, India, 2022, pp. 232-237, doi: 10.1109/ICAAIC53929.2022.9793198. [Crossref]
18. B.Lou, et.al, 2022, "Quantitative Timing Strategy Model Based on Improved KNN Algorithm", Procedia Computer Science, Volume 202, 2022, Pages 61-66, DOI: 10.1016/j.procs.2022.04.009 [Crossref]
19. N.B.Oğur, et.al, 2023, "Detection of depression and anxiety in the perinatal period using Marine Predators Algorithm and kNN", Computers in Biology and Medicine, Volume 161, July 2023, 107003, DOI: 10.1016/j.compbiomed.2023.107003, [Crossref]

20. Jianxin Deng, et.al, 2022, "Process parameters design of squeeze casting through an improved KNN algorithm and existing data", Journal of Manufacturing Processes, Volume 84, December 2022, Pages 1320-1330, doi:10.1016/j.jmapro.2022.10.074 [Crossref]
21. C.Cha.Zhu, et.al, 2022, "Multiplicative Fault Detection and Isolation in Dynamic Systems Using Data-Driven K-Gap Metric based kNN Algorithm", IFAC-PapersOnLine, Volume 55, Issue 6, 2022, Pages 169-174, doi:10.1016/j.ifacol.2022.07.124 [Crossref]
22. G.K.Walia and J.S.Sidhu, 2022, "Estimation of optimized window size for hybridized kNN-random forest algorithm based image demosaicing", Materials Today: Proceedings, Volume 66, Part 8, 2022, Pages 3578-3581, doi: 10.1016/j.matpr.2022.07.017 [Crossref]
23. L.Liu, et.al, 2022, "Recognition of gas-liquid flow regimes in a helically coiled tube using wire-mesh sensor and KNN algorithm", International Journal of Multiphase Flow, Volume 154, September 2022, 104144, doi:10.1016/j.ijmultiphaseflow.2022.104144 [Crossref]
24. Ghazi Alkhatib, 2024, "Structured Methodologies vs UML Artifacts Revisited: A perspective from a developing country", Procedia Computer Science, Volume 239, 2024, Pages 2090-2097, doi: 10.1016/j.procs.2024.06.396 [Crossref]
25. Z. Babaalla, et.al, 2024, "Extraction of UML class diagrams using deep learning: Comparative study and critical analysis", Procedia Computer Science, Volume 236, 2024, Pages 452-459, doi: 10.1016/j.procs.2024.05.053 [Crossref]