PREDICTION OF COOPERATIVE LOAN FEASIBILITY USING THE K-NEAREST NEIGHBOR ALGORITHM

Roviani[1*]; Deddy Supriadi[2]; Iqbal Dzulfiqar Iskandar[3*)]

[1]Program Studi Sistem Informasi
[1] Universitas Bina Sarana Informatika, Indonesia
www.bsi.ac.id
oviriviani@gmail.com[1]

[2,3]Program Studi Sistem Informasi
[2,3,] Fakultas Teknik dan Informatika, Universitas Bina Sarana Informatika Kampus Tasikmalaya, Indonesia
www.bsi.ac.id
deddy.dys@bsi.ac.id[2], iqbal.iql@bsi.ac.id [3*)]

(*) Corresponding Author

Abstract—Approval of credit lending to cooperative members without proper feasibility analysis can cause credit problems, cooperatives such as late payment of installments, and an increase in bad credit which can threaten the survival of the cooperative as a provider of lending services. As a solution to minimize the creditworthiness assessment errors for loan funds, research is carried out to analyze the feasibility of loan funds from the data of cooperative members using the data mining method approach and the algorithm used using the K-Nearest Neighbor. The purpose of this research is to predict the feasibility of granting credit with the right decision and to find out the level of evaluation, accuracy, and validation of the effectiveness of the k-NN algorithm on processing creditworthiness application data classifications. After the prediction research was carried out, the data on the eligibility of credit lending applications were conducted at the Bakti Berkah Sukaraja Savings and Loan Cooperative, The data obtained from the accuracy value of the k-nearest neighbor algorithm before being validated has an accuracy of 87.78% with AUC 0.95, after validation with split validation the accuracy decreased slightly by 2% to be 85.71%, while the AUC value in the ROC Curve was 0.836%. Even though there was a decline, it can still be categorized as a good classification. The impact of this research is that besides the accuracy of the k-NN algorithm being validated, the Bakti Berkah Sukaraja Savings and Loan cooperative can predict the feasibility of applying for credit funds, as an effort to reduce the threat of bad credit risk.

Keywords: Accuracy, Creditworthiness, Data mining classification, K-Nearest Neighbor, ROC Curve Split Validation

Abstrak— *Persetujuan Peminjaman kredit pada anggota koperasi tanpa analisa kelayakan yang tepat dapat menimbulkan permasalahan kredit, koperasi seperti terlambat membayar angsuran, dan peningkatan kridit macet yang dapat mengancam keberlangsungan hidup dari koperasi sebagai penyedia layanan jasa peminjaman dana. Sebagai solusi untuk meminimalisir kesalahan assesment kelayakan kredit pinjaman dana, dilakukan riset analisis kelayakan pinjaman dana dari data anggota koperasi dengan pendekatan metode data mining dan algoritma yang digunakan menggunakan K-Nearest Neighbour. Tujuan dari riset ini adalah untuk memprediksi kelayakan pemberian kredit dengan keputusan yang tepat, serta mengetahui tingkat evaluasi, akurasi dan validasi efektifitas algoritma k-NN terhadap pemerosesan klasifikasi data pengajuan kelayakan kredit. Setelah dilakukan riset prediksi data pengajuan kelayakan peminjaman kredit di Koperasi Simpan Pinjam Bakti Berkah Sukaraja, didapatkan data nilai akurasi algoritma k-nearest neigbour sebelum divalidasi memiliki akurasi 87.78% dengan AUC 0.95, setelah dilakukan validasi dengan split validasi akurasi sedikit menurun 2% menjadi sebesar 85,71%, sedangkan Nilai AUC dalam ROC Curve adalah sebesar 0,836%. Meskipun terjadi penurunan namun tetap dapat dikategorikan sebagai good classification. Dampak dari riset ini selain nilai akurasi dari algoritma k-NN sudah tervalidasi, koperasi Simpan Pinjam Bakti Berkah Sukaraja dapat melakukan prediksi kelayakan pengajuan kredit dana, sebagai upaya untuk mengurangi ancaman resiko kredit macet.*

Kata Kunci*: Accuracy, Creditworthiness, Data mining classification, K-Nearest Neighbor, ROC Curve Split Validation*

INTRODUCTION

Cooperatives have a role as business entities that provide services to save or provide loans to each individual from membership, cooperatives that are legal entities based their activities on the principles of cooperatives as well as people's economic movements based on the principle of kinship as referred to in the cooperative laws and regulations. (Simanjuntak, Daulat Freddy, Keri Boru Hotang, 2021).

One of them is a savings and loan cooperative. Savings and loan cooperatives carry out their business activities, only savings, and loan business. The existence of a savings and loan cooperative is one way to get loan funds in developing a business, fulfillment of daily needs so that it has a positive impact on the community, can improve the standard of life, and a good economic cycle. But with the provision of loans to members without proper analysis, can cause cooperative credit problems such as being late in paying installments (Pratama & Purwanto, 2018), bad credit that can threaten the survival of the cooperative (Bahar, 2020) based on credit loan data in 2018, from 70 credit borrowers 64% had bad credit, and in 2019, 76% of 91 credit borrowers had bad credit. (Gani & Fandorann, 2020). Based on the bad credit loan data, the high percentage value of bad credit can be categorized as unhealthy credit lending (Zahra, 2020) because it exceeds the standard safe credit limit of Bank Indonesia, which is 5% (Gubernur Bank Indonesia, 2013). To avoid the risk of the threat of bad credit, it is better if there is a need to analyze member data to determine the eligibility of credit extension to prospective borrower customers by predicting and classifying whether or not members are eligible for funding credit. From this problem. Then research was carried out to predict and classify the creditworthiness of cooperative members with a data mining approach(Bramer, 2020) with the Nearest Neighbor algorithm (I D Iskandar, N Ch Basjaruddin, D Supriadi, Ratningsih, D S Purnia, 2020).

Research related to the creditworthiness analysis of loan funds, classification algorithms, K-Nearest Neighbor has been carried out by several previous researchers, namely research by building a selection application for determining prospective customers for goods selling companies on credit by applying the k-NN algorithm into the program to support retrieval decision (Edy Nasri, 2020), The results of this research conclude that the company can more easily determine prospective debtors who will submit orders for goods by credit. Besides, there is research with an expert system for the selection of prospective motorbike credit debtors with the C.45 classification algorithm (Hartini & Kurahman, 2020) The results of the research conclude that the C.45 algorithm can classify the eligibility of prospective motorcycle credit debtors as a reference to avoid bad credit. Another related research is a comparison of the two k-NN classification algorithms with C5.0 on the prediction of bad credit debtor prospective data in cooperatives (Permana et al., 2020) The results of this research concluded that the accuracy obtained by the C5.0 algorithm was 86.67%, and the KNN algorithm obtained an accuracy of 83.33%. Research classification of bad credit card predictions with the C.4.5 algorithm(Mardhiyah et al., 2020) concludes that the process of predicting the feasibility of applying for credit card data can be done with an accuracy value of 70.93%. Comparative research of classification algorithms C.45 and k-NN for the determination of customer credit in the fuel oil industry (Nuryaman, 2018) The research results conclude the classification results of the C4.5 algorithm with an accuracy of 84% and the AUC graph value of 0.686 and the k-NN algorithm with an accuracy value of 84.86% and an AUC value of 0.692. And research with the title of credit card application classification with k-NN (Yogiek Indra Kurniawan, 2020) concludes the result of a precision value of 92%, a recall value of 83%, and an accuracy value of 93%.

The research that has been done previously regarding the prediction of creditworthiness applications has shortcomings, which only focuses on the results of predictions and evaluations. So that there is no re-validation process on the results of processing algorithms against the data(Nurhasan et al., 2018), It is assumed that the accuracy of the predicted value for the classification results of the k-NN algorithm is not yet valid, due to this shortcoming. Then research was carried out related to the prediction of the feasibility of providing cooperative credit using the K-nearest neighbor algorithm with different case studies and data sets. As well as the results of the k-NN algorithm classification process will be evaluated and validated to ensure that the accuracy value is truly valid

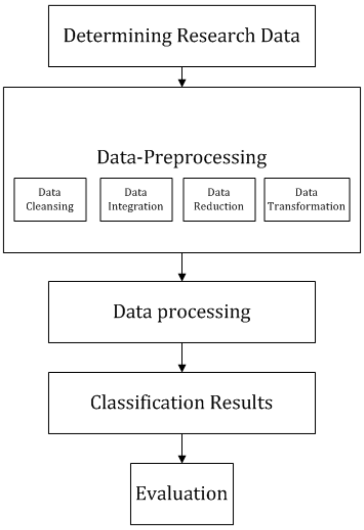
K-Nearest Neighbor classification algorithm (Kumar et al., 2018), Naive Bayes Classifier,(Khajenezhad et al., 2021), and Algoritma C.45 (Ariadi, 2020) included in the classification algorithm that is applied to predict the risk of applying for creditworthiness so that the research is not too broad this research will be focused on the Nearest Neighbor algorithm. (Bazan et al., 2020) The advantages of the K-Nearest Neighbor algorithm are that training is very fast, simple, and easy to learn, resistant to training data that has noise, and is effective if the training data is large.

While the drawbacks of this algorithm are computationally complex with refraction K values, limited memory, and gullibility with irrelevant attributes. While the advantages of naïve Bayes are simple algorithms, faster calculation, and high accuracy. However, the naïve Bayes algorithm also has a weakness where a probability cannot measure how accurate a prediction is. (Pérez-Martín et al., 2018).

Based on these problems. So research was carried out with the title "Data Mining Classification to Determine the Feasibility of Providing Cooperative Credit Using the K-Nearest Neighbor Algorithm. This research scope aims to determine the results of the data mining classification evaluation on the creditworthiness data of cooperative members using the K-nearest Neighbors algorithm, namely the value of accuracy, convention matrix, and ROC Curve, To produce more accurate results, split validation will also be applied (Pham et al., 2020). While the benefits of seeing how to predict and classify the feasibility of giving credit to cooperatives with a data mining approach with Yahoo Nearest Neighbor members, to minimize the risk of being late in paying installments, bad credit that can threaten the survival of the cooperative.

MATERIALS AND METHODS

The method proposed is as follows: data mining classification to determine the feasibility of providing cooperative credit using the k-nearest neighbor algorithm.



Source:(Roviani et al., 2021)

Figure1. Data mining classification framework to determine the creditworthiness of cooperative credit using the proposed k-nearest neighbor algorithm.

Figure 1 is a proposed framework for classifying data mining to determine the feasibility of providing cooperative credit with the K-Nearest Neighbor algorithm. To better understand the stages. Then explain the details of the stages carried out, namely:

First, determine the research data. From the large number of loan submission data, 200 loan submission data are obtained, while the research data are taken is 221 data for a credit application for the cooperative bakti berkah sukaraja from 2016-2017. Second, Data Preprocessing. At this stage, the authors take steps so that the data obtained can become quality data and can be a good input for mining tools. The data preprocessing stages, among others: Data Cleansing At this stage, it is done by filling in the missing values and resolving the inconsistencies. Data Integration,(Miharja et al., 2020) At this stage the authors combine some data from various databases into a new database. Data Reduction At this stage the writer eliminates unnecessary data, Data Transformation. At this stage, the writer converts the data into the appropriate format. Data Processing Using Rapid Miner. At this stage, the data will be classified using the k-NN algorithm with the Rapid Miner mining tools (Deshpande, 2015).

After the classification is carried out, an evaluation will be carried out to see the level of effectiveness of data processing by the k-NN algorithm in the form of accuracy in the confusion matrix, and the predetermined ROC Curve or AUC value. The sample credit application data samples that have been obtained are presented in tables 1, 2, 3, and 4.

Table 1. Example of data set for filing a cooperative loan Bakti Berkah Sukaraja before preprocessing

No. Member	Income 1	Income 2	Spending 1	Spending 2	Income 3
SK00510	750.000	1.500.000	500.000	908.000	842.000
SR01265	-	2.600.000		1.500.000	1.100.000
LG00170	-	4.500.000	-	1.750.000	2.750.000
SR00289	3.000.000	1.040.000	1.500.000	900.000	1.640.000
MK01350	2.500.000	3.750.000	3.000.000	1.300.000	1.950.000
LG00645	1.500.000	1.800.000	1.800.000	-	1.500.000
JG00278	7.800.000	-	4.200.000	900.000	2.700.000
LG00019	2.500.000	3.000.000	1.500.000	1.500.000	2.500.000
JG00174	3.900.000	1.500.000	2.600.000	900.000	1.900.000

Source:(Roviani et al., 2021)

Table 1 is an example of a dataset set of credit applications for cooperative Bakti Berkah Sukaraja before preprocessing consisting of a member number, income 1 is the main source of income for the husband's debtor, attribute income 2 is the source of income from the wife. The attribute of expenditure 1 and expense 2 is the amount of financial expenditure in one month, while income 3 is the amount of net income of the debtor per month.

Table 2. Example of data set for submitting a loan for a debtor's credit fund and guarantee for the cooperative bakti berkah sukaraja before preprocessing

No	No member	Gender	Profession	date	Submission	Necessity	Insurance
1	SK00510	L	Karyawan Swasta	06/02/16	10.000.000	Modal Warung	SPPT
2	SR01265	L	MRT	19/12/16	4.000.000	Pembelian Tanah	Tidak Ada
3	LG00170	P	MRT	06/02/17	35.000.000	Pembelian Rumah	SPPT
4	SR00289	P	MRT	30/01/17	10.000.000	Modal Menjahit	SPPT
5	MK01350	P	MRT	09/09/16	10.000.000	Modal Warung	SPPT
6	LG00645	P	MRT	10/01/17	5.000.000	Modal Bengkel	SPPT
7	JG00278	P	MRT	07/02/17	10.000.000	Warungan Agen LPG	SPPT
8	LG00019	P	MRT	06/02/17	10.000.000	Membeli Kendaraan	SPPT
9	JG00174	P	MRT	09/02/17	8.000.000	Modal Warung	SPPT
10	TR00352	P	MRT	12/01/17	4.000.000	Bordir	SPPT

Source:(Roviani et al., 2021)

Table 2 is a dataset of credit fund submissions and guarantees from debtors before preprocessing is carried out.

Table 3. Data set of eligibility for a credit application for customers of the cooperative bakti berkah sukaraja before preprocessing

No	Loan	Recommend	Date	Disbursement	classification
SK00510	4.000.000	7.000.000	24/02/17	7.000.000	Less
SR01265	-	4.000.000	24/02/17	4.000.000	worth it
LG00170	5.000.000	25.000.000	23/02/17	25.000.000	Less
SR00289	3.000.000	10.000.000	23/02/17	10.000.000	worth it
MK01350	-	10.000.000	23/02/17	10.000.000	worth it
LG00645	3.000.000	4.000.000	22/02/17	4.000.000	Less
JG00278	5.000.000	10.000.000	22/02/17	10.000.000	worth it
LG00019	-	10.000.000	21/02/17	10.000.000	worth it
JG00174	5.000.000	8.000.000	21/02/17	8.000.000	worth it
TR00352	3.000.000	4.000.000	21/02/17	4.000.000	worth it

Source:(Roviani et al., 2021)

Table 3 is an example of a dataset of debtor credit application approval before preprocessing and before being classified by the k-NN algorithm. The data sets that have been preprocessed are presented in Table 4

Table 4. Example of eligibility data set for a credit application for customers of the cooperative Bakti Berkah Sukaraja after preprocessing

No	Profession	Necessity	Insurance	Total Income	Total Expenditures	Loan	Number of Submissions	Time period	classification
1	4	2	2	2.250.000	1.408.000	4.000.000	10.000.000	24	Less
2	2	1	1	2.600.000	1.500.000	0	4.000.000	12	worth it
3	2	1	2	4.500.000	1.750.000	5.000.000	35.000.000	24	Less
4	2	2	2	4.040.000	2.400.000	3.000.000	10.000.000	24	worth it
5	2	2	2	6.250.000	4.300.000	0	10.000.000	24	worth it
6	2	2	2	3.300.000	1.800.000	3.000.000	5.000.000	12	Less
7	2	2	2	7.800.000	5.100.000	5.000.000	10.000.000	18	worth it
8	2	1	2	5.500.000	3.000.000	0	10.000.000	24	worth it
9	2	2	2	5.400.000	3.500.000	500.000	8000.000	12	worth it
10	2	2	2	3.000.000	1.500.000	3.000.000	4.000.000	12	worth it
11	3	1	4	14.500.000	9.200.000	6000000	16.000000	18	?

Source:(Roviani et al., 2021)
Link: https://github.com/iqbaldzi13/Data-Set

Table 4 is an example of a preprocessing credit application dataset, the attributes in table 4 are a combination of table1, table2, and table3 to be predicted using the k-NN algorithm. in column number 11 there is a credit application that is not known to be classified as feasible or not feasible. After preprocessing the data sets were classified using the k-NN algorithm. The K-Nearest Neighbor (K-NN) algorithm is one of the methods used for classification analysis, but in the last few decades, the KNN method has also been used for prediction. (Fatah & Subekti, 2018). The method is to find the closest distance between the data to be evaluated and the closest neighbors in the data. Far or near distance between neighbors is usually calculated based on Euclidean Distance. Where Euclidean distance is the calculation of the distance from 2 points in euclidean space. (Bode, 2017) Euclidean is related to the Pythagorean theory and in this case euclidean can also be applied to more attributes by the formula being

$$D(x+y) = \sqrt{\sum_{i=1}^{n} (x_i + y_i)^2} \quad \text{......................(1)}$$

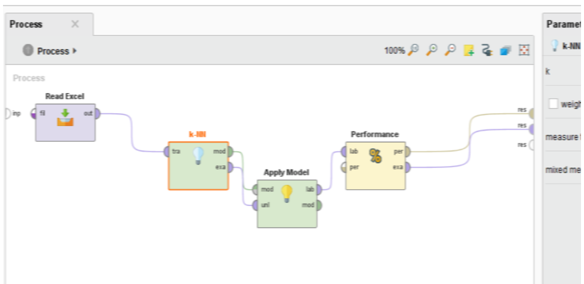
D is the distance between attributes, x is the new attribute data, y is the old attribute data. (Wang, 2019). The stages for calculating the K-NN algorithm are as follows:
Input: training data set D = {(X_i, Y_i), 1≤i≤N}, where X_i is the conditional attribute of the ith sample,
Y_i is the category, new sample X, distance function d.
Output: Category Y of X, For i=1 to N do, Calculate the distance d (X_i, X) between X and Xi; End for Sort the distance and get d (X, X_{i1}) ≤d(X, X_{i2}) ≤... ≤d(X, X_{iN}); Select the first K samples: S= {(X_{i1}, Y_{i1})...(X_{iK}, Y_{iK})}; Count the number of occurrences of each category in S and determine the category Y of X.

RESULT AND DISCUSSION

In table 4 column number 11 there is data for new credit applications, the classification is not yet known whether it is included in the feasible or reduced category. To predict whether it is feasible or reduced it can be processed using the k-NN algorithm by calculating the euclidean distance for data 1, namely:

$$D_1 = \sqrt{(1+2)^2 + (3+4)^2 + (1+2)^2 + (4+2)^2 + (14500000+2250000)^2 + (9200000+1408000)^2 + (6000000+4000000)^2 + (16000000+10000000)^2 + (18+24)^2}$$
D_1 = 15835964,26

Do it up to D10 or data to 10. To speed up the computation process, the next data calculation process will be processed using the Rapid Miner tools with the k-NN algorithm. (Deshpande, 2015) From the data that has been preprocessed before. The types of attributes of each attribute are presented in Table 4.



Source:(Roviani et al., 2021)
Figure 2. The data processing uses the k-NN algorithm without validation

Figure 2. Describing the arrangement of operators used in Rapidminer including reading Excel whose function is to import data and read data that has been stored in an excel file in .xls or .xlsx format, k-NN this operator is one of the calcifications and algorithmic data mining models were chosen in the .xls or .xlsx format. this research. The k-NN operator has the K value of 3, this is because from K-1 to K-10 on K-3 the greatest results are obtained, Apply Model is an operator that functions to apply the model to the dataset, Performance is an operator that functions to evaluate research where the results are in the form of Accuracy, ROC Curve and AUC values. The classification results will be presented in table 5, while the accuracy measurement is presented in table 6.

Table 5. Results of the k-NN classification algorithm

No	Classification	Calculation results
3	less	22748681,28
2	decent	19516659,55
10	decent	18561788,71
6	less	17612495,56
1	less	15835964,26
4	decent	14165154,43
8	decent	13836184,45
9	decent	13427583,55
5	decent	12809078,81
7	decent	9934787,366

Source:(Roviani et al., 2021)

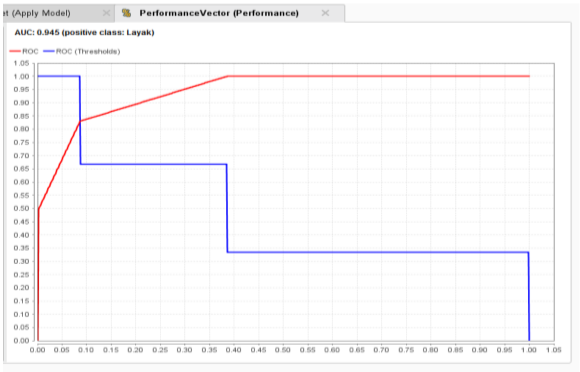
Table 5. is the Results of the k-NN classification algorithm. The classification process will be automatically sorted from the highest to the lowest values can be seen in table 5. Collect classification categories based on the K value, because the K value is 3, then the classification is based on the 3 highest values from table 5. Then the classification of the top 3 values is Less, Eligible, and feasible, by using the KNN see the majority classification based on the K value, namely the majority classification of the 3 highest scores. The majority classification of the top scores is Eligible. So for new data, namely the 11th data can be classified as Feasible meaning that the application is eligible for disbursement according to customer submissions.

Table 6. Results of Confusion Matrix processing without validation

Algorithm	Accuracy	AUC
k-NN	87.78%	0.95

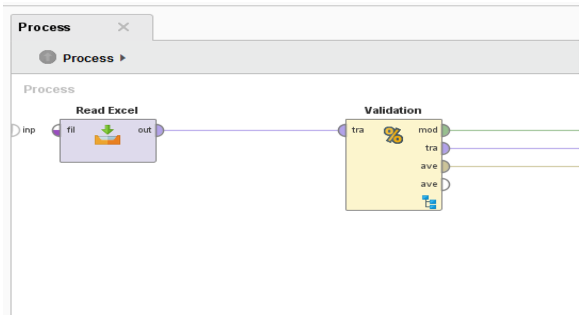
Source:(Roviani et al., 2021)

Table 6 is the Confusion matrix that is useful for evaluating the performance of the classification model carried out by the k-NN algorithm based on the predictive accuracy of a model. Accuracy is expressed as a percentage (%). In this study, the accuracy value was 87.78%.



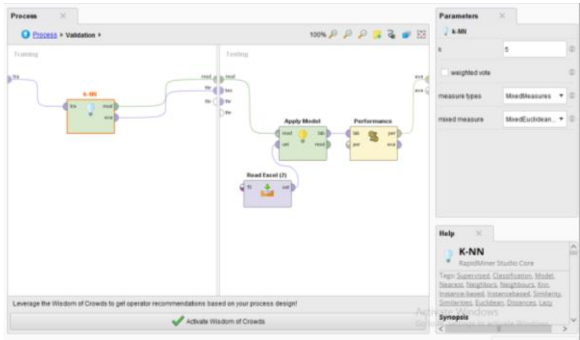
Source:(Roviani et al., 2021)
Figure3. ROC Curve Tanpa Validasi

Figure 3 is the ROC, The accuracy of processing the algorithm on the data is depicted in Figure 3 in the ROC curve (ROC Curve). The value of AUC (Area Under Curve) or the area under the ROC curve (ROC Curve), the larger the area, the better the test results. The result here is 0.945, which means it is an excellent classification. Although the research results are quite high, in the process, it has not been validated, in this study, the validation used is split validation. Split Validation is a validation technique by dividing the data into testing data and validating the model used is valid or not so that the results of the research carried out are more accurate. The split validation operator is presented in Figure 4 and Figure 5.



Source: (Roviani et al., 2021)
Figure 4.Data processing with split validation 1

Figure 4 is a validation operator compiled in rapidminer, this validation process is an attempt to complement the shortcomings of previous research, which only focuses on classification results and evaluation accuracy values. The validation used for this research is Split validation, a multi-tiered split operator validation that has two sub-processes, consisting of a sub-process of testing training data and testing data. The training process is used to study the model from the results of the classification algorithm, then it is applied to testing the data model. when the testing process occurs which is called algorithm validation of data processing. The algorithm used in this research is the k-NN algorithm.



Source:(Roviani et al., 2021)
Figure 5. Data processing multilevel operator that has two sub-processes, consisting of a sub-process of testing training data and testing data.

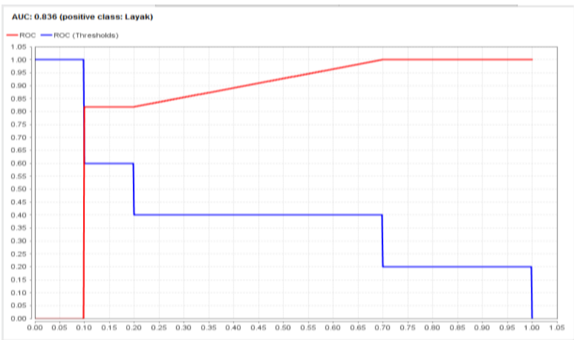
Figure 5 is the k-NN algorithm validation process, the data process is divided into 2, namely training data and testing data, 200 training data for model building, and 21 data testing data used for model testing. In this process, the researcher uses operators such as Read Excel whose function is to import data and read data that has been stored in an excel file in .xls or .xlsx format. The data in the first read excel is 90% of training data. Split Validation is a simple validation operator for evaluating a model where the data has been randomly divided into training data and testing data. Where the comparison ratio used in this study is 90% training data and 10% testing data, this k-NN operator is one of the calcifications and algorithmic data mining models were chosen in this study. In the k-NN operator, the K value is 5, this is because from K-1 to K-10 on K-5 the greatest results are obtained, Read Excel (2) whose function is to import data and read data that has been saved in excel file format .xls or .xlsx. The data in the second read excel is 10% data testing Apply Model is an operator that functions to apply the model to the dataset, Performance is an operator that functions to evaluate research where the results are in the form of Accuracy, ROC Curve and AUC values. The validation results will be presented in table 7.

Table 7. Confusion Matrix accuracy with splits Validation

Algorithm	Accuracy	AUC
k-NN	85.71%	0.836

Source:(Roviani et al., 2021)

Table 7 is the result of the processing of confusion Matrix validation results from the validation of the k-NN algorithm, the accuracy value of the algorithm processing on creditworthiness data in the cooperative of Bakti Blessing Sukaraja after validation is 85.71% and the AUC value is 0.836.



Source:(Roviani et al., 2021)

Gambar6. ROC Curve Dengan Split Validation

Figure 6. Is a ROC Curve with Split Validation used to measure the accuracy of the applied test, depicted in the ROC curve (ROC Curve). The value of AUC (Area Under Curve) or the area under the ROC curve (ROC Curve), the larger the area, the better the test results. The result here is 0.836 which means it is a good classification(Pahlevi, 2020). After comparing the results, the correct results are using split validation where the accuracy value is 85.71% with a K-5 value, and the AUC value is 0.836 which is a good classification. Even though the value is smaller, the results are more accurate because the model used has been validated first.

CONCLUSION

Research on the prediction of creditworthiness to cooperatives using the K-Nearest Neighbor algorithm can be carried out, the results of the research can be concluded that the accuracy value of the k-nearest neighbor algorithm before validation has an accuracy of 87.78% with AUC 0.95, after validation with split validation is carried out on the creditworthiness study of cooperative members. The accuracy decreased slightly by 2% to 85.71%, while the AUC value in the K-nearest neighbor algorithm ROC Curve with split validation in the creditworthiness study of cooperative members was 0.836%. Even though there was a decline, it can still be categorized as a good classification. So that the results of the data mining classification to determine the feasibility of giving cooperative credit using the k-nearest neighbor's algorithm have a high accuracy value, which can be applied to predict the feasibility of applying for credit to customers in cooperatives.

REFERENCES

Ariadi, F. (2020). Analisa Perbandingan Algoritma DT C . 45 dan Naïve Bayes Dalam Prediksi Penerimaan Kredit Motor. *Jurnal Riset Inovasi Bidang Informatika Dan Pendidikan Informatika (KERNEL)*, *1*(1), 1–7.

Bahar, H. (2020). Strategi Penyelesaian Kredit Macet Dan Dampak Terhadap Kinerja Keuangan Pada PT Bank Sulselbar Cabang Barru. *DECISION Jurnal Ekonomi Dan Bisnis*, *1*(2), 178–186.

Bazan, J., Bazan-Socha, S., Ochab, M., Buregwa-Czuma, S., Nowakowski, T., & Woźniak, M. (2020). Effective construction of classifiers with the k-NN method supported by a concept ontology. *Knowledge and Information Systems*, *62*(4), 1497–1510. https://doi.org/10.1007/s10115-019-01391-w

Bode, A. (2017). K-Nearest Neighbor Dengan Feature Selection Menggunakan Backward Elimination Untuk Prediksi Harga Komoditi Kopi Arabika. *ILKOM Jurnal Ilmiah*, *9*(2), 188–195. http://jurnal.fikom.umi.ac.id/index.php/ILKOM/article/view/139

Bramer, M. (2020). Principles of data mining fourth edition. In *Drug Safety* (Vol. 30, Issue 7). https://doi.org/10.1007/978-1-4471-7493-6

Deshpande, V. K. (2015). *Predictive Analytics and Data Mining*. Elsevier Inc.

Edy Nasri, A. S. A. (2020). PENJUALAN BARANG SECARA KREDIT DENGAN ALGORITMA K-NEAREST NEIGHBOR. *Jurnal Sains & Teknologi*, *4*(1), 1–11.

Fatah, H., & Subekti, A. (2018). Prediksi Harga Cryptocurrency Dengan Metode K-Nearest Neighbours. *Jurnal PILAR Nusa Mandiri*, *14*(2), 137–144. https://doi.org/10.33480/pilar.v14i2.30

Gani, A., & Fandorann, U. (2020). Analisis Tingkat Kredit Macet Bumdes Tunas Harapan Desa Simpang Campang Kecamatan Kisam Ilir Kabupaten Oku Selatan. *JETAP Jurnal Akuntansi Dan Bisnis*, *53*(9), 1689–1699. http://journal.unbara.ac.id/index.php/etap/article/view/550

Gubernur Bank Indonesia. (2013). *Peraturan Bank Indonesia Nomor 15/2/Pbi/2013 Tentang Penetapan Status Dan Tindak Lanjut Pengawasan Bank Umum Konvensional* (pp. 1–41). BANK INDONESIA.

Hartini, S., & Kurahman, T. (2020). Sistem Pakar Pemilihan Calon Debitur Kredit Motor Dengan Algoritma C4 . 5 Pada PT . Federal International Finance. *Information System For*

Educators And Professionals, *5*(1), 51–60. http://journal.unbara.ac.id/index.php/etap/article/view/550

I D Iskandar, N Ch Basjaruddin, D Supriadi, Ratningsih, D S Purnia, and T. W. (2020). Popular Content Prediction Based on Web Visitor Data With Data Mining Approach Popular Content Prediction Based on Web Visitor Data With Data Mining Approach. *Journal of Physics: Conference Series*, 1–7. https://doi.org/10.1088/1742-6596/1641/1/012105

Khajenezhad, A., Bashiri, M. A., & Beigy, H. (2021). A distributed density estimation algorithm and its application to naive Bayes classification. *Applied Soft Computing*, *98*, 106837. https://doi.org/10.1016/j.asoc.2020.106837

Kumar, A., Bhatnagar, R., & Srivastava, S. (2018). Analysis of Credit Risk Prediction Using ARSkNN. *Advances in Intelligent Systems and Computing*, *723*(March 2016), 644–652. https://doi.org/10.1007/978-3-319-74690-6_63

Mardhiyah, P. A., Siregar, R. R. A., & Palupiningsih, P. (2020). Klasifikasi Untuk Memprediksi Pembayaran Kartu Kredit Macet Menggunakan Algoritma C4.5. *Jurnal Teknologia*, *3*(1), 91–101. https://aperti.e-journal.id/teknologia/article/view/66

Miharja, J., Putra, J. L., & Hadianto, N. (2020). Comparison Of Machine Learning Classification Algorithm On Hotel Review Sentiment Analysis (Case Study : Luminor Hotel Pecenongan). *Jurnal Pilar Nusa Mandiri*, *16*(1), 59–64. https://doi.org/10.33480/pilar.v16i1.1131

Nurhasan, F., Hikmah, N., & Utami, D. Y. (2018). Perbandingan Algoritma C4.5, KNN, Dan Naive Bayes Untuk Penentuan Model Klasifikasi Penanggung Jawab BSI Entrepreneur Center. *Jurnal Pilar Nusa Mandiri*, *14*(2), 169–174.

Nuryaman, Y. (2018). Komparasi Klasifikasi Penentuan Customer Kredit Menggunakan Algoritma C4.5 dan KNN Pada PT. Citra Semesta Energy. *Jurnal Pilar Nusa Mandiri*, *14*(2), 233–238.

Pahlevi, O. (2020). Data Mining Model for Designing Diagnostic Applications Inflammatory Liver Disease. *Jurnal Dan Penelitian Teknik Informatika*, *5*(1), 51–57. https://doi.org/https://doi.org/10.33395/sinkron.v5i1.10589 e-ISSN

Pérez-Martín, A., Pérez-Torregrosa, A., & Vaca, M. (2018). Big Data techniques to measure credit banking risk in home equity loans. *Journal of Business Research*, *89*(February), 448–454. https://doi.org/10.1016/j.jbusres.2018.02.008

Permana, T., Siregar, A. M., Masruriyah, A. F. N., & Juwita, A. R. (2020). Perbandingan Hasil Prediksi Kredit Macet Pada Koperasi. *Conference on Innovation and Application of Science and Technology*, *3*(1), 737–746. http://publishing-widyagama.ac.id/ejournal-v2/index.php/ciastech/article/view/1970

Pham, B. T., Jaafari, A., Avand, M., Al-Ansari, N., Du, T. D., Hai Yen, H. P., Phong, T. Van, Nguyen, D. H., Van Le, H., Mafi-Gholami, D., Prakash, I., Thuy, H. T., & Tuyen, T. T. (2020). Performance evaluation of machine learning methods for forest fire modeling and prediction. *Symmetry*, *12*(6), 1–21. https://doi.org/10.3390/SYM12061022

Pratama, A. A. S., & Purwanto, I. W. N. (2018). Upaya Restrukturisasi Kredit Bermasalah di PT. Bank Pembangunan Daerah Cabang Gianyar. *Kertha Semaya: Journal Ilmu Hukum*, *6*(4), 1–15.

Roviani, R., Supriadi, D., & Iskandar, I. D. (2021). *Laporan Akhir Penelitian Mandiri: Prediction Of Feasibility Of Cooperative Loans Using K-Nearest Neighbor Algorithm*.

Simanjuntak, Daulat Freddy, Keri Boru Hotang, A. R. (2021). Pelatihan penyusunan laporan keuangan koperasi. *Jurnal Kewirausahaan, Akuntansi, Dan Manajemen TRI BISNIS*, *3*(1), 66–75.

Wang, L. (2019). Research and Implementation of Machine Learning Classifier Based on Research and Implementation of Machine Learning Classifier Based on KNN. *IOP Conference Series: Materials Science and Engineering PAPER*, 1–5. https://doi.org/10.1088/1757-899X/677/5/052038

Yogiek Indra Kurniawan, T. I. B. (2020). Klasifikasi Penentuan Pengajuan Kartu Kredit Menggunakan. *Jurnal Ilmiah Matrik Universitas Bina Darma*, *22*(1), 73–82. https://doi.org/https://doi.org/10.33557/jurnalmatrik.v22i1.843

Zahra, N. L. (2020). Penilaian Tingkat Kesehatan Koperasi Pada Koperasi Simpan Pinjam Mitra Sukses Lestari Malang. *Jurnal Ilmiah Mahasiswa Fakultas Ekonomi Dan Bisnis*, *9*(2), 1–20.