Journal of Artificial Intelligence and Engineering Applications

Website: https://ioinformatic.org/

15th October 2023. Vol. 3. No. 1; e-ISSN : 2808-4519

Implementation of Mechine Learning Eligibility for Customer Credit Payments at Bank BTN Using the K – Nearst Neighbor Algorithm

Ema Sari Suwandi[1*], Relita Buaton[2], Rusmin Saragih[3]

[1,2,3]STMIK Kaputama

emasoewandi@gmail.com [1], bbcbuaton@gmail.com [2], evitha12014@gmail.com [3]*

Abstract

Credit is the provision of money or bills that can be equated with that, based on a loan agreement or agreement between a bank and another party that requires the borrower to pay off the debt after a certain period of time with interest (Government of Indonesia, 1998). In its initial development, credit had a function in stimulating mutual assistance aimed at meeting needs, both in the field of business and meeting daily needs. In developing applications, it is necessary to predict applications at Bank BTN Medan accurately, accurate prediction results are very important in showing the right rating and decision-making in selecting customers. When customers experience arrears, the system used by Bank BTN Medan is still manual and has not applied predication in credit arrears to customers of Bank BTN Medan. Tests carried out in this test use a credit customer dataset from Bank BTN Medan. This study predicts the eligibility of customer credit payments at Bank BTN with the K – Nearst neighbor algorithm. The prediction of the level of smoothness of credit payments is made using K-Nearest Neighbor in order to be able to predict the smoothness of future credit payments.

Keywords: *Mechine Learning, Credit, K – Nrearst Neighbor*

1. Introduction

Credit is the provision of money or bills that can be equated with that, based on a loan agreement or agreement between a bank and another party that requires the borrower to pay off the debt after a certain period of time with interest (Government of Indonesia, 1998).Credit is expected to have a positive socio-economic impact on debtors, creditors or the community. Provision of credit to customers by the bank can experience various problems or risks. The problem or risk that can be experienced by banks in extending credit is the attitude of customers who do not make installment payments on time or delay payments for several months of installment payments which causes bad credit Accurate prediction results are very important in showing ratings and making the right decisions in customer selection. When customers experience arrears, the system used by Bank BTN Medan is still manual and has not applied prediction in credit arrears to Bank BTN Medan customers. Tests carried out in this test use a credit customer dataset from Bank BTN Medan. This study predicts the eligibility of customer credit payments at Bank BTN with the K – Nearst neighbor algorithm.

The prediction of the level of smoothness of credit payments is made using K-Nearest Neighbor in order to be able to predict the smoothness of future credit payments.The K-Nearest Neighbor method has the ability to make a diagnosis by paying attention to the initial data in previous cases and comparing it with new data, this process aims to produce a better and more accurate diagnosis. The advantage of the K-NN method is that it has the power to process training data which has a lot of noise and is very effective if the training data used has a large amount. The more training data, the better the prediction results [1].

2. Research Methods

2.1 Prior Research

Previous research is an attempt by researchers to seek comparisons and to find new inspiration for future research.The first research was conducted by. The results of this study are to optimize the Naive Bayes algorithm with Forward Selection feature selection to be able to increase the results of the accuracy or success rate obtained from predicting credit payments. The data to be used in this study comes from XY Bank in Gorontalo. The data obtained relates to all aspects of credit customers including personal information from customers. The experimental design in this study uses a credit customer dataset. While the analysis used is the Naive Bayes Algorithm Model with Forward Selection feature selection. The prediction of the level of smoothness of credit payments using the Forward Selection-based Naive Bayes

algorithm is able to predict the smoothness of future credit payments. This is evidenced by the acquisition of accuracy values based on Forward Selection Naive Bayes capable of achieving an accuracy value of 71.97% [2].

2.2 Definition of Banks

The origin of the word bank is from the Italian word banca which means a place to change money. In general, the notion of a bank is a financial intermediary institution that is generally established with the authority to receive money deposits, lend money, and issue promissory notes or what is known as a banknote. Bank is a financial institution whose main activity is collecting funds from the public and channeling these funds back to the community and providing other services. Meanwhile, according to Article 1 of Law Number 7 of 1992 concerning banking as amended by Law Number 10 of 1998 quoted by Kristianto, 2011 is a business entity that collects funds from the public in the form of savings and distributes them to the public in the form of credit and other forms in order to improve the standard of living of many people. Commercial Banks are banks that carry out business activities conventionally and or based on Sharia Principles which in their activities provide services in payment traffic. Rural Banks are banks that carry out business activities conventionally or based on Sharia Principles which in their activities do not provide services in payment traffic [3].

2.3 Definition of Customer

According to RI Law no. 10 of 1998 concerning banking contains the types and understanding of customers, in article 1 number 17 it is stated that "the understanding of customers is those who use bank services." According to the banking dictionary, "a customer is a person or entity that has a savings account or loan at a bank." From the above understanding it can be concluded that customers are: People who are used to dealing with or being bank customers (in financial terms) People who are insurance dependents, Comparison Every customer certainly really expects satisfaction from any bank he makes transactions, customer satisfaction is one of the factors it is important for companies to pay attention, because customers/consumers are the main target of a company/banking business entity. Banks must be able to make their customers feel safe and confident to save their funds in the bank. In order for people to want to save their money in the bank, the banking sector provides incentives in the form of remuneration that will be given to the depositor. The remuneration can be in the form of interest, profit sharing, gifts, services or other remuneration. The higher the remuneration provided, the more people will be interested in saving their money. Therefore, the banking sector must provide stimulation and trust so that people are interested in investing their funds [4].

2.4 Definition of Credit

Credit comes from the word credere which means trust. Every economic actor who is given a credit facility is a person trusted by the creditor. This condition is after going through an assessment process on several aspects such as will, motivation, and ability. This understanding needs to be a concern because the trust given by creditors to debtors is an achievement in itself. UU no. 10 of 1998 concerning banking states that "credit is the provision of money or its equivalent based on a loan agreement or agreement between the bank and another party which requires the debtor to pay off the debt according to the time period and the provision of interest". stated that in the provision of credit there are several things, namely the party providing the loan is called the creditor, and the party receiving the loan is called the debtor, the provision of funds, the credit agreement, the credit deadline, the required interest rate, and the risk for the creditor as a result of a number of receipts funds in the future when faced with uncertainty. This principle is a way of conducting an in-depth study of the condition of the prospective debtor which includes an analysis of the character, capacity, capital, collateral, and economic conditions [5].
In daily activities, the word credit is known, starting from credit for glassware sold by mortgage lenders from home to houses or credit in the form of money provided by bonded craftsmen. On a broader scale, it is also known as credit given by leasing and banking companies. Then also already known every credit transactions occur are always related to installments or installments with along with the time period and the amount of installments to be paid. Takers Credit also understands that credit installments already contain it loan principal and interest to be paid. terms used credit takers are called debtors and parties lenders (banks) we call the creditor or in another sense the debtor is recipient of funds while the creditor is the provider of funds [6].

2.5 Algorithm K – Nearst Neighbor

The K-Nearest Neighbor method have the ability to make a diagnosis in this way pay attention to the initial data in the case before and compare with new data, this process aims to yields a better diagnosis as well as accurate. The advantages of the K-NN methodthat is to have inner strength process training data that has lots of noise as well as very effective if the training data used has a large number. More and more training data then the predicted results will be to be better.To use this method need to determine the number of values of K nearest neighbors in doingnew data classification. Furthermore need to follow gradually in application of the K-NN algorithm, as Determining K parameters. Calculate the distance between the data to be tested with all training data. Sort these distances from smallest to the biggest. Determine the distance closest to the value of K. Find and conclude the dominant result closeness to K.
Modification algorithms was raised to correct weaknesses from pre-existing data mining. The weakness of K-Nearest Neighbor lies in the calculation process is carried out almost every time data at the classification stage. To solve this problem then an approach is proposed by combining Naïve Bayes and K-Nearest algorithms Neighbor to overcome the weaknesses of the K Nearest Neighbor method. This algorithm is expected to work improve time effectiveness, accuracy accuracy and long processing time [7].

3. Results And Discussion

3.1 Research methods

In conducting research on this thesis, the authors followed the stages of the methodology in this study, namely as follows:

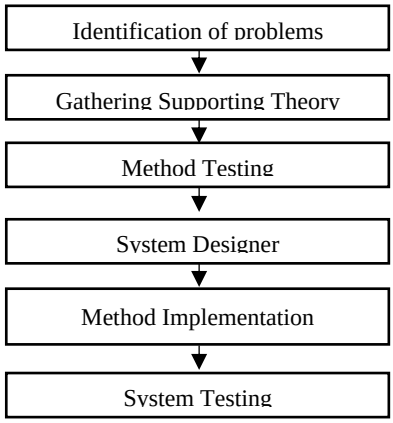

Figure 1: Research Methodology

To clarify the structure of the research methodology above, the author makes the following statement:

1. Problem Identification
   This stage is the initial stage used to identify problems with the aim of observing and looking for problems that are being faced with the object of research, namely seeking predictions on the feasibility of customer credit payments.
2. Gathering Supporting Theory
   Collection of theories related to the subject matter such as theories about predicting credit payments, the methods used and the design applications of the required systems. In this stage, theories are collected from several sources such as journals, articles and other references at Bank BTN Medan.
3. Testing Methods
   At this stage the researcher will test the methods used in the prediction process on the feasibility of customer credit payments with existing guidelines on prediction theories from books and journals related to the subject matter.
4. System designer
   At this stage a system design is carried out for the problem being studied, which can be in the form of a stage for designing the workflow of the system and also designing the design of the interface of the system to be created.
5. Method Implementation
   Implement the methods that have been previously tested with the system design that has been made and do the coding according to the programming language used to make the system.
6. System Testing
   In the final stage, a series of tests are carried out on the system that has been made, tests are carried out in order to find errors in the system and make the necessary improvements.

3.2 Research Supporting Data

The steps for implementing the research method to solve problems with the calculation of the K - Nearest Neighbor algorithm are carried out below:

Step 1 :
Determine Euclidean's Distance. By calculating the distance between the new data and all the training data. With the calculation data in the table below:

Table 1: Supporting Data

NO	NAMA	CBAL	PREBAL	TERM	PD DAYS	INCOME	EXPENDITURE	REPORT RESULTS
1	CJG	89.220.333	91.000.000	240	19	5500000	4000000	Not smooth
2	FS	136.173.281	138.500.000	204	19	4500000	3000000	Not smooth
3	DST	101.436.950	102.500.000	180	19	6700000	2800000	Not smooth
4	SH	137.634.456	138.975.000	240	19	4300000	3200000	Not smooth
5	AL	106.472.268	108.995.000	240	19	5000000	3000000	Not smooth
6	RRW	143.203.192	144.500.000	240	0	7000000	3000000	Fluent
7	FH	143.197.167	144.500.000	180	0	5700000	4000000	Fluent
8	APH	143.195.827	150.500.000	240	0	6000000	3500000	Fluent
9	TM	143.193.614	144.500.000	240	0	6500000	4000000	Fluent
10	SA	143.193.614	144.500.000	180	0	4200000	2500000	Fluent
11	SR	138500000	138500000	204	10	4000000	2000000	?

We are given new data that we will classify, namely CBAL = 138500000, PREBAL = 138500000 , TERM = 204 and PDDAYS = 10, INCOME = 4000000 AND EXPENDITURES = 2000000 So what classification is this new data Current or Not Current.

$\sqrt{(89.220.333 - 138500000)^2} + (91.000.000 - 138500000)^2 + (240 - 204)^2 + (19 - 10)^2 + (5500000 - 4000000)^2 + (4000000 - 2000000)^2 = 68.490.770,03$

$\sqrt{(136.173.281 - 138500000)^2} + (138.500.000 - 138500000)^2 + (204 - 204)^2 + (19 - 10)^2 + (4500000 - 4000000)^2 + (3000000 - 2000000)^2 = 4291109.57$

$\sqrt{(101.436.959 - 138500000)^2} + (102.500.000 - 138500000)^2 + (180 - 204)^2 + (19 - 10)^2 + (6700000 - 4000000)^2 + (2800000 - 2000000)^2 = 51745528.07$

$\sqrt{(137.634.456 - 138500000)^2} + (138.975.000 - 138.500.000)^2 + (240 - 204)^2 + (19 - 10)^2 + (4300000 - 4000000)^2 + (3200000 - 2000000)^2 = 1582653.28$

$\sqrt{(106.472.268 - 138500000)^2} + (108.995.000 - 138.500.000)^2 + (240 - 204)^2 + (19 - 10)^2 + (5000000 - 4000000)^2 + (3000000 - 2000000)^2 = 43569721.62$

$\sqrt{(143.203.192 - 138500000)^2} + (144.500.000 - 138.500.000)^2 + (240 - 204)^2 + (0 - 10)^2 + (7000000 - 4000000)^2 + (3000000 - 2000000)^2 = 8253485.02$

$\sqrt{(143.197.167 - 138500000)^2} + (144.500.000 - 138.500.000)^2 + (180 - 204)^2 + (0 - 10)^2 + (5700000 - 4000000)^2 + (4000000 - 2000000)^2 = 8059365.85$

$\sqrt{(143.195.827 - 138500000)^2} + (150.500.000 - 138.500.000)^2 + (240 - 204)^2 + (0 - 10)^2 + (6000000 - 4000000)^2 + (35000000 - 2000000)^2 = 13380986.18$

$\sqrt{(143.193.614 - 138500000)^2} + (144.500.000 - 138.500.000)^2 + (240 - 204)^2 + (0 - 10)^2 + (6500000 - 4000000)^2 + (4000000 - 2000000)^2 = 8263172.05$

$\sqrt{(143.193.614 - 138500000)^2} + (144.900.000 - 138.500.000)^2 + (180 - 204)^2 + (0 - 10)^2 + (4200000 - 4000000)^2 + (25000000 - 2000000)^2 = 7636754.05$

Note: D is read as distance CBAL, TERM, PDDAYS and PREBAL, INCOME AND EXPENDITURE. We must determine the number of neighbors that we will count (k).

1 Independent Variables, namely variables whose values are not influenced by other variables. In the example data above, the independent variables include CBAL, PREBAL, TERMS and PDDAYS, INCOME, EXPENDITURES.
2 Dependent Variables, namely variables whose values are influenced by other variables. In the example data above, which includes the dependent variable is the Report Results.

Table 2: Results of Euclidean's Distance

EUCLIDEANS DISTANCE		
1	68.490.770,03	Not smooth
2	4.291.109,57	Not smooth
3	51.745.528,07	Not smooth
4	1.582.653,28	Not smooth
5	43.569.721,62	Not smooth
6	8.253.485,02	Fluent
7	8.059.365,85	Fluent
8	13.380.986,18	Fluent
9	8.263.172,05	Fluent
10	7.636.754,05	Fluent

Step 2:
After calculating the distance of each neighbor to the customer, then sort the results by distance, starting from the smallest to the largest in the table below to determine the value of K = 3:

Table 3: Calculating the Distance of KNN Sequential Values

NO	HASIL PREDIKSI	KNN RESULTS		RESPORT RESULTS
1	1.582.653,28	1(K<3)	Yes	Not smooth
2	4.291.109,57	2(K<3)	Yes	Not smooth
3	7.636.754,05	3(K=3)	Yes	Fluent
4	8.059.365,85	4(K>3)	No	Fluent

5	8.253.485,02	5(K>3)	No	Fluent
6	8.263.172,05	6(K>3)	No	Fluent
7	13.380.986,18	7(K>3)	No	Fluent
8	43.569.721,62	8(K>3)	No	Not smooth
9	51.745.528,07	9(K>3)	No	Not smooth
10	68.490.770,03	10(K>3)	No	Not smooth

Step 3:
Take the value (K = 3) of the closest neighbor, then we see each of the neighbors. The meaning of the nearest neighbor value is the distance of the K value that has been taken as the closest distance sample which has been adjusted in the calculation above, so the result is the closest K value in the table below:

Table 4: KNN Results

PREDICTION RESULTS	KNN RESULTS	Smallest Distance	What does KNN mean?
1.582.653,28	1(K<3)	Ya	Tidak Lancar
4.291.109,57	2(K<3)	Ya	Tidak Lancar
7.636.754,05	3(K=3)	Ya	Lancar

After the final calculations have been carried out in the KNN process, the results of applying the above calculations have produced the desired results in the search process for the value of K. Then there are 3 closest data, namely (2), (4) and (10). So it is obtained that Non-Current = 2 and Current = 1. By voting, it is obtained that the customer with number 11 is included in the Non-Current Classification Category. So if the results of the above calculations are carried out, then that the customer will make arrears in the eligibility of credit payments at Bank BTN Medan.

4. Conclusions

The appropriate variables for implementing machine learning are the eligibility of customer credit payments with the criteria used are CBAL (Credit Principal Value), PREBAL (Credit Ceiling), TERM (Monthly Credit Period), PDDAYS (Total Days of Credit Arrear), Income and Expenses. Testing the K – Nearst Neighbor Algorithm for the Implementation of Mechine Learning Appropriateness of Customer Credit Payments at Bank BTN utilizes the results of the K value that has been determined with the prediction results of Current and Non-Current. Results of trials on systems that have been built with the Matlab programming language. For the future, it is hoped that the variables used can be added and adjusted to the needs of the system goals to be better in the needs of Bank BTN Medan. In research with the same topic, it is necessary to try other research methods in supporting Mechine Learning Feasibility of Customer Credit Payments at Bank BTN with Linear Regression Algorithms, Neve Bayessin, SVM and other algorithms. The system in which the application process is carried out can use online and mobile programming languages. The design of the Mechine Learning System for Credit Payment Feasibility using the KNN algorithm can also be developed further into other programming such as online and mobile programming languages.

References

[1] D. M. K. Neighbor, S. Informasi, S. Royal, S. Komputer, and S. Royal, "1 , 2* , 3 1," vol. VII, no. 2, 2021.
[2] M. Hasan, "MENGGUNAKAN ALGORITMA NAIVE BAYES BERBASIS," vol. 9, pp. 317–324, 2017.
[3] R. Irviani and A. Hamid, "Learning Media Application of Knowledge Sharing Room to Improve Student and Community Literacy," no. 06, pp. 1–17, 2022.
[4] P. Kepada, N. Pada, and P. T. Bank, "PERANAN CUSTOMER SERVICE DALAM MENINGKATKAN UTAMA PADANG Yolanda Darma Fernandes , Doni Marlius," pp. 1–12.
[5] M. Bank, Y. Sehat, P. Pt, L. Amelia, and D. Marlius, "CABANG UTAMA PADANG," no. 2007, pp. 1–11.
[6] M. Bank, Y. Sehat, P. Pt, L. Amelia, and D. Marlius, "CABANG UTAMA PADANG," no. 2007, pp. 1–11.
[7] F. Gultom and T. Simanjuntak, "Prediksi Tingkat Kelancaran Pembayaran Kredit Bank Dengan Menggunakan Algoritma Naïve Bayes Dan K-Nearest Neighbor," METHOMIKA J. Manaj. Inform. dan Komputerisasi Akunt., vol. 4, no. 2, pp. 98–102, 2021, doi: 10.46880/jmika.vol4no2.pp98-102.