Jurnal Mantik, 5 (3) (2021) 1791-1798

Published by:Institute of Computer Science (IOCS)

Jurnal Mantik

Journal homepage: www.iocscience.org/ejournal/index.php/mantik/index

Prediction Model of Eligibility of Lending in Credit Banks Using The C4.5 Algorithm and Naive Bayes Method

Indra Maulana[1], Moh. Subchan[2]

[1]Budi Luhur University.
[2]Muhammadiyah University of Banten, Indonesia

E-mail: indramaulanaxe@gmail.com[1], moh.subhan@gmail.com[2]

ARTICLEINFO

Article history:

Received: September 11, 2021
Revised: October 12, 2021
Accepted: November 02, 2021

Keywords:

C4.5 Algorithm,
Naïve Bayes,
Analysis,
Credit.

ABSTRACT

People's credit banks are financial institutions that collect funds from savings and channel them back in the form of credit. One form of credit owned by Rural Banks is installment credit which is intended for customers who want to increase business capital or other needs. To determine quickly and reduce the risk of non-performing loans in lending, To prevent bad loans, accurate forecasting is needed, one of which uses technology in the field of data mining. Naive Bayes predicts future probabilities based on previous experience by studying the correlation of hypotheses which are the class labels that are the target of mapping in the classification and evidence which is the features that are input in the classification model. Data processing based on data mining is expected to be used as a tool in predicting creditworthiness which estimates whether or not an applicant or customer is eligible for credit.

Copyright © 2021 Jurnal Mantik.
All rights reserved.

1. Introduction

People's credit banks are financial institutions that collect funds from savings and channel them back in the form of credit. Under the Banking Law, credit is the provision of money or claims based on an agreement or loan agreement between the bank and the customer, which requires the borrower to repay the debt after a certain period of time. One form of credit owned by Rural Banks is installment credit which is intended for customers who want to increase business capital or other needs. the process of assessing the feasibility of granting credit is a problem for rural credit banks.

For the credit approval process, rural banks must conduct a detailed analysis so that it can be determined whether the granting of credit is feasible or not. Some of the obstacles in the credit approval process at Rural Banks are the inaccurate results of the decisions that result in an increase in non-performing loans and the lack of speed in the results of credit analysis carried out.

This is influenced by the human error factor because there is no standard procedure for analyzing credit. While the data to be analyzed amounted to tens or even hundreds per day, resulting in a very large analysis error rate and took a long time. Based on previous research "Determination of Credit Eligibility with the Naïve Bayes Classifier Algorithm" conducted by Nia Nuraeni, "C4.5 Algorithm for Customer Classification in Predicting Bad Loans" research was carried out by Aldi Zein Pratama and "Implementation of the Naive Bayes Algorithm in Determining Credit Provision" research was conducted by Muhammad Husni Rifqo, Ardi Wijaya. To predict credit worthiness, a data mining model is needed.

2. Method

The research method is a scientific method used to find problems with a specific purpose. The method that will be used is by observing and interviewing the required data, determining attributes and determining the results of data analysis, model analysis and designing and making credit feasibility models. This research on the creditworthiness model uses the C4.5 Algorithm and Naïve Bayes method.

Accredited "Rank 4"(Sinta 4), DIKTI, No. 36/E/KPT/2019, December 13th 2019.

Jurnal Mantik is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0).

2.1 Sample Selection Method

The type of data obtained is primary data directly from the object of research. This data is obtained directly from PT. Rural Bank Kredit Mandiri Indonesia Tangerang Branch in the form of customer data and conducted closed interviews with credit analysts as the author's analysis data in this study, then the data were used as supporting supports and references in applying the creditworthiness model.

2.2 Method of Collecting Data

The data collection method that will be used is by observation and interviewing the required data such as determining attributes related to the creditworthiness model. Determining the type and source of data to obtain truly accurate data is very important. The source of data in this study is credit data taken from PT. Rural Bank Kredit Mandiri Indonesia Tangerang Branch in previous years as a reference to find certain patterns that can be used as determining attributes. The data that can be used in this research are customer data and credit data.

Table1.
Attribute List before Selection

No	Nama Atribut	Nilai Atribut	Keterangan
1	Jns_Kel	W P	Gender
2	Education	SD SMP SLTA SLTP D3 S1 S2 S3	Customer education
3	Status_Married	Married Not married yet	Marital status
4	Number Of Dependents	01-Oct	
5	Home_status	One's own Rent/Contract Other	Status of residence
6	Job_debtor	Employee Entrepreneur Civil servant	Customer Job
7	Position	OWNER EMPLOYEE DRIVER TEACHER OPERATOR SECURITY STAFF	Position in work
8	Type_business_debtor	SERVICE TRADER MANUFACTURE RENT MEDICAL DEVICES BUMN SECURITY SERVICES TRANSPORTATION SERVICES SELLING CLOTHES COOPERATIVE FOOD EDUCATION COUNTRY CHICKEN CUTTING	Type of customer's business

Accredited "Rank 4"(Sinta 4), DIKTI, No. 36/E/KPT/2019, December 13th 2019.

Jurnal Mantik is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0).

No	Nama Atribut	Nilai Atribut	Keterangan
		TRADING ACCOUNTING ADMINISTRATION EMPLOYEE MECHANIC/MAINTENANCE OPERATOR SPV STAFF ENTREPRENEUR	
9	Time_work_debtor	1-30 tahun	Customer working period
10	Purpose_Credit	CONSUMPTION CREDIT WORKING CAPITAL STARTUP CAPITAL RENOVATION	Loan purpose
11	Jns_Facilities	Consumer Credit Working capital Investment	Type of loan facility
12	Ceiling_credit	1.0000.000 - 1.000.000.000	Loan amount
13	Time Period	0-60 bulan	Time period
14	Interest Rate	0-2.00	Interest rate
15	Distance_crowd	1-20 KM	The distance of the guarantee location from the crowd
16	Access_Road	ASPHALT LAND	Access road guarantee location
17	Density_level	CONGESTED LESS SOLID	Population density level at the guaranteed location
18	Wide_road	1 CAR >= 2 CARS ALLEY	Road area at the guarantee location
19	Pert_environment	Fast Stable	Environmental growth at the guarantee site
20	Disaster-prone	NO PROBLEM VULNERABLE	Prone to disaster or not at the guarantee location
21	Flood_area	Free Seldom	Flood area whether or not the location of the guarantee

2.3 Instrumentation

Each research requires several supporting components to launch the research, in this study the supporting components used are Weka software and Rapid Miner software.

Weka software is a desktop-based application which in research is used as a tool to process input data to determine attributes and classifications.

2.4 Analysis Techniques

Descriptive analysis technique is carried out to analyze the data to be carried out on the results of data collection with literature studies, and observations to obtain specifications for the system requirements to be developed. This method will be used to process the existing predictive data in order to produce accurate predictions.

From the data that can be divided into two sets, among others, as training data and test data. The learning outcomes of each method with training data will be compared with the test results using the k-fold cross validation test method to obtain statistical values in the form of accuracy, precision, recall, ROC Curve. The statistical results will be compared to get the best method to be applied in system design.

2.5 Prediction Process Design on Prototype

The processes that will be designed in the system prototype include:

Accredited "Rank 4"(Sinta 4), DIKTI, No. 36/E/KPT/2019, December 13th 2019.

Jurnal Mantik is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0).

a. Excel data import process

The user creates or prepares the required data in the form of .csv and or .xls data formats. then the process of importing data to the system can be run.

b. Preprocessing Process

After the import process is complete, the next stage is the missing value process, which is the process of checking the entire data by the prototype to determine the feasibility of processing the prediction process such as checking criteria and others.

c. Prediction Process

After the data is clean, predictions will be made on the test data using the method with the best accuracy value that has gone through the stages in the analysis process in the research. The prediction results are in the form of predictions of creditworthiness which are stated in the notation of Good and Troubled.

2.6 Research Steps

The steps of this research can be seen in the image below:

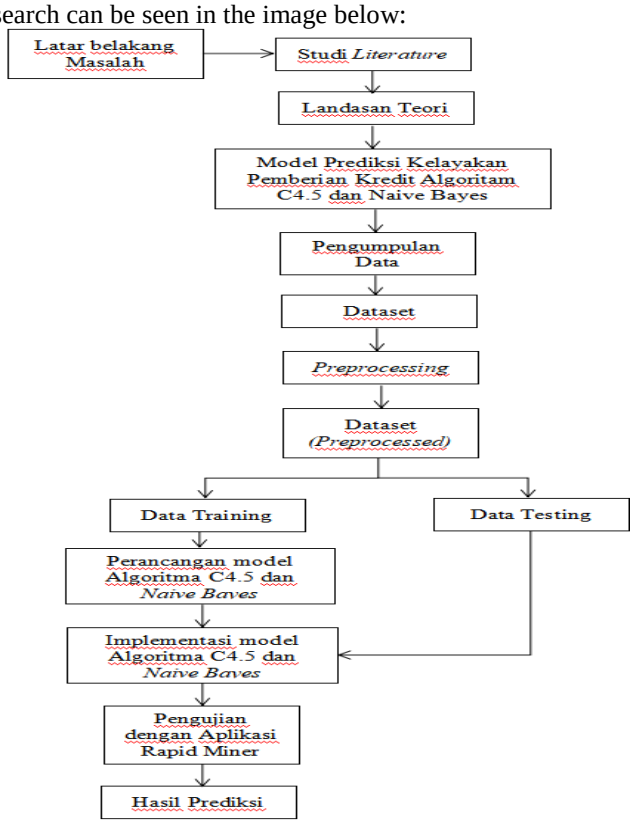

Fig 1. Research Method

3. Result and Analysis

There are management difficulties in checking the feasibility of providing credit to customers at PT. Indonesian Mandiri Credit Rural Bank. So in this study, data mining will be applied using the C4.5 algorithm and the Naïve Bayes algorithm using Rapid Miner tools for creditworthiness. The problem that often arises is the influence of human error factors in credit approvals because there is no standard procedure for creditworthiness approvals carried out by related users. Therefore we need a way to reduce the risk early. To classify customers who deserve credit approval, this study uses the C4.5 algorithm and the Naive Bayes algorithm. This algorithm was chosen because it can process data and produce a good classification. Where is the customer profile data as a parametric for measurement, in order to assist companies in making credit decisions. By having data, it is very important for companies to know potential customers, with information on potential customers, companies can make the right decision to give credit approval.

Accredited "Rank 4"(Sinta 4), DIKTI, No. 36/E/KPT/2019, December 13th 2019.

Jurnal Mantik is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0).

3.1 Data Understanding

In this study, the data for the selection of creditworthiness prediction data were obtained from PT. Rural Bank Mandiri Credit Indonesia which consists of 22 attributes, of which of the 22 attributes will be used 11 attributes, of which 10 are predictor attributes and 1 is the result attribute. In this study, the class label is the credit status attribute.

3.2 Data Preprocessing

The data obtained for this study were 941 customer records in 2017, 2018, 2019 at PT. Indonesian Mandiri Credit Rural Bank. To get quality data, data cleaning, data integration, data transformation and data reduction techniques are used at the preprocessing stage. Data cleaning is used to complete or eliminate incomplete data (missing values), eliminate noisy data, and correct inconsistent data. Data integration is used to integrate data with supporting data. For data transformation, it is the process of converting or merging data into a suitable format for processing in data mining. Data Reduction is used to eliminate incomplete data, eliminate noisy data and correct inconsistent data. Often the data that will be used in the data mining process has a format that cannot be used immediately, therefore the format needs to be changed.

3.3 Data Cleaning

Data Cleaning is a process for cleaning from dirty data. The meaning of dirty data is data that has no value, is inconsistent, irrelevant or there are writing errors when inputting. Data cleaning will also affect the ability of data mining techniques, because the data handled will be reduced in number and complexity. The data used in this study is customer data for 2017, 2018, 2019 totaling 941. In this data there are missing values in the time_term attribute as much as 62 data, interest rate as much as 57 data, road area as much as 38, distance_crowd as much as 10 data, tenure_debtor as much as 21 data, Furthermore, the data that contains missing values are removed.

3.4 Data Integration

Data Integration in this study is to connect customer data with credit data.

3.5 Data Reduction

Data Reduction in this study is to remove unused data related to research needs into a simplified table form as a requirement for application to the C4.5 and Naïve Bayes methods. The data reduction process is carried out by removing as many as 322 noisy data such as 22 sample data shown in the table below.

Table 2 .
Data Noisy

jangka_waktu	suku_bunga	luas_jalan	jarak_keramaian	daerah_banjir
6	0.50	None	\N	None
6	0.50	None	\N	None
24	1.70	None	\N	None
24	1.70	None	\N	None
24	1.70	None	\N	None
12	1.70	None	\N	None
12	1.70	None	\N	None
24	1.80	None	\N	None
24	1.80	None	\N	None
36	1.60	None	\N	None
24	1.70	None	\N	None
24	1.70	None	\N	None
12	1.70	None	\N	None
48	1.70	None	0	None
3	2.00	None	\N	None
3	2.00	None	\N	None
3	2.00	None	\N	None
3	2.00	None	\N	None
24	1.70	None	\N	None
24	1.70	None	\N	None
36	1.60	None	0	None
36	1.60	None	3 KILOMETER ARAH TIMUR	None

Accredited "Rank 4"(Sinta 4), DIKTI, No. 36/E/KPT/2019, December 13th 2019.

Jurnal Mantik is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0).

3.6 Training Data and Testing Data

After preprocessing, the data obtained are 619 records that are ready to be used for training data and testing data. Training data is data that is ready to be mined that has passed data preprocessing. While testing data is data used to test predictive rules obtained from training data. In this study, test data was carried out for the distribution of training data and testing data with several comparisons, namely 50:50, 60:40, 70:30, 80:20 and 90:10. There are 472 in the Good category and 147 in the Troubled category.

Table 3
Comparison of training data and testing data

Percentage comparison of training data and testing data	Acuracy	
	Naive Bayes	C4.5
50:50	76.05%	67.31%
60:40	76.61%	71.77%
70:30	76.34%	69.35%
80:20	76.42%	72.36%
90:10	75.81%	66.13%

3.7 Manual Testing of C4.5 Algorithm Using Ms. Excel

The initial step of the C4.5 algorithm is to find the value of entropy, entropy is used to determine how informative an input attribute is to produce an attribute. First, determine the total entropy value using training data, the number of data is 371 consisting of 283 good credit quality and 88 non-performing credit quality. The number of cases for each attribute and the subset of attributes that will be used to calculate the total entropy and entropy of each attribute can be seen in the table below.

3.8 Software Prototype Design

From the evaluation and validation results above that the C4.5 and Nave Bayes methods have a good level of accuracy and performance so that the rules generated by the C4.5 and Nave Bayes methods can be used as rules for making prototype designs that can make it easier to predict the feasibility of providing credit facility.

3.9 Testing Environment

The test environment provides an overview of the hardware and software specifications used by users in the system testing process, both validation and quality testing. These specifications are obtained in the observation process based on system aspects. The following is a brief summary of the specifications of the hardware and software used by users for the testing process,

a. Hardware instrument in the form of an HP Probook 4330 laptop with the following specifications:

1) Processor: Intel® CPU @ 2.40GHz 2.50 GHz
2) Hard disk: 240 GB
3) RAM: 8.00GB
4) OS : Windows 10

b. Instrument software (software), namely:

1) PHP Programming language
2) MySQL : Database
3) CodeIgniter : Web Framework
4) Weka: Tools that help in testing
5) Rapid Miner: Tools that help in testing
6) Microsoft Excel: Tools that help in calculations

3.10 Summary of Discussion

The dataset used has 619 records which are divided into 60% training data and 40% testing data using attributes resulting from the selection of the Weka software features to get 10 attributes with 1 predictor attribute. The value of Accuracy Model C4.5 is 71.77% and Naïve Bayes is 76.61%. So that the dataset model created is classified into a good classification (good classification).

3.11 Research Implications

The implications of this research revolve around the system aspect and further research aspects. The system aspect is related to operational technical, hardware and software design required. While the advanced research aspect is related to further research that is needed to improve the quality of previous research.

3.12 System Aspect

To apply the results of this study, a good support system is needed, so that interested parties can use the results of this study to predict the feasibility of providing credit facilities. Therefore, adequate facilities and

Accredited "Rank 4"(Sinta 4), DIKTI, No. 36/E/KPT/2019, December 13th 2019.

Jurnal Mantik is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0).

infrastructure are needed consisting of hardware, software (operating systems and applications created at the deployment stage), and other infrastructure in order to provide the best results, hardware and software specifications that can be used in this study have minimum specifications. System Testing.

3.13 Further Research Aspect

There are limitations in this study, so it is hoped that in future studies of the same type the following can be considered:

a. Comparing more methods in data analysis and problem solving, in order to obtain a system that is more effective and efficient in processing or presenting information.
b. Management of research time in order to maximize it, given the short time available.
c. This research can be developed in other classification algorithms, such as the Neural Network algorithm, K-Means or SVM (Support Vector Machine) by adding a feature selection step in order to get optimal results.

4. Conclusion

This study uses 619 customer data at PT. Indonesian Mandiri Credit Rural Bank. For customer data used is customer data in 2017, 2018 and 2019 which is taken from the database of PT. Rural Bank Kredit Mandiri Indonesia Tangerang Branch In the data there are 472 customer data with good credit quality and 147 bad credit quality. The modeling of data mining predictions of creditworthiness has been made a prototype using the PHP programming language with the CodeIgniter framework using the MySql database and calculation tools C4.5 and Naïve Bayes methods. In this study, it was concluded that the 60:40 comparison data used to predict creditworthiness could be processed using the C4.5 and Naïve Bayes methods, the accuracy of C4.5 reached 71.77% and the results of the Nave Bayes accuracy of 76.61% from the results of the study it can be concluded that : 1. Appropriate use when applying the C4.5 and Naïve Bayes algorithms can improve better accuracy in predicting creditworthiness. 2. Information that will be taken by credit analysts or management in deciding the eligibility of customers to obtain credit facilities.

5. References

[1] Ritzkal, R., & Setiadi, D. (2021). Data Storage System Arrival and Departure Airnav Halim Perdana Kusuma Airport. Jurnal Mantik, 5(2), 555-562.
[2] Fadillaha, M. N., & Subchan, M. (2021). Dampak Covid-19 Terhadap Perilaku Konsumen Dalam Penggunaan Marketplace Di Indonesia. Jurnal mitra manajemen, 12(1), 123-130.
[3] Syaputra, A., & Setiadi, D. (2020). Sistem Pakar Diagnosa Kerusakan Sepeda Motor Yamaha Matic Menggunakan Metode Forward Chaining. Jusikom: Jurnal Sistem Komputer Musirawas, 5(2), 126-135.
[4] N. P. Astuti et al., "Vehicle Security System using Short Message Service (SMS) as a Danger Warning in Motorcycle Vehicles," Journal Robot and Control, vol. 1, no. 6, pp. 224–228, 2020, doi: 10.18196/jrc.1642.
[5] Ritzkal,,Syaiful,S., "The application of academic information system measurement software with iso standardization," Proceedings of the International Conference on Industrial Engineering and Operations Management, 2020, (August).
[6] Sibagariang, R., & Riandari, F. (2019). Decision Support System for Determining the Best Wood For the Production Cabinet Using Bayes Method: Decision Support System for Determining the Best Wood For the Production Cabinet Using Bayes Method. Jurnal Mantik, 3(3), 99-103.
[7] Yulias, N., & Widianto, S. R. (2021). Prediction of drinking water facility conditions using the Naive Bayes Algorithm. Jurnal Mantik, 4(4), 2599-2603.
[8] Algoritma, C, Eka Praja, Wiyata Mandala, and Dewi Eka Putri. 2018. "Prediksi Jumlah Pemberian Kredit Kepada Nasabah Di Bank Perkreditan Rakyat Dengan Algoritma C 4.5." 5(1): 70–80.
[9] Gunawan, Rudi. 2019. "Implementasi Data Mining Menggunakan Regresi Linier Berganda Dalam Memprediksi Jumlah Nasabah Kredit Macet Pada BPR Tanjung Morawa." 18(1): 87–91.
[10] Jiang, Yi, Yan Chen, Zhimin Zeng, and Xiangjian He. 2008. "A Bank Customer Credit Evaluation Based on the Decision Tree and the Simulated Annealing Algorithm." Proceedings - 2008 IEEE 8th International Conference on Computer and Information Technology, CIT 2008: 203–6.
[11] L.T, Priyanka, and Neethu Baby. 2013. "Classification Approach Based Customer Prediction Analysis for Loan Preferences of Customers." International Journal of Computer Applications 67(8): 27–31.
[12] Marcos, Hendra et al. 2014. "Implementasi Data Mining Untuk Klasifikasi Nasabah Kredit Bank " X " Menggunakan Classification Rule." : 1–7.

Jurnal Mantik is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0).

[13] Menarianti, Ika. 2015. "Klasifikasi Data Mining Dalam Menentukan Pemberian Kredit Bagi Nasabah Koperasi." Jurnal Ilmiah Teknosains 1(1): 1–10. http://e-jurnal.upgrismg.ac.id/index.php/JITEK/article/view/836.
[14] Murdianingsih, Yuli. 2015. "Klasifikasi Nasabah Baik Dan Bermasalah Menggunakan Metode Naive Bayes." Jurnal Informasi 2015(November): 349–56.
[15] Nuraeni, Nia. 2017. "Penentuan Kelayakan Kredit Dengan Algoritma Naïve Bayes Classifier : Studi Kasus Bank Mayapada Mitra Usaha Cabang PGC." Jurnal Teknik Komputer AMIK BSI (JTK) 3(1): 9–15.
[16] Pratama, Aldi Zein, Laela Kurniawati, Simson Larbona, and Tuti Haryanti. 2019. "Algoritma C4 . 5 Untuk Klasifikasi Nasabah Dalam Memprediksi Kredit Macet." Information System for Educators and Professionals 3(2): 121–30.
[17] Rani, Larissa Navia. 2015. "Klasifikasi Nasabah Menggunakan Algoritma." Jurnal KomTekInfo Fakultas Ilmu Komputer 2(2): 33–38.
[18] Rifqo, Muhammad Husni, and Ardi Wijaya. 2017. "Implementasi Algoritma Naive Bayes Dalam Penentuan Pemberian Kredit." Pseudocode 4(2): 120–28.
[19] Sintia, Syifa, Al Khautsar, Diah Puspitasari, and Prima Mustika. 2018. "Algoritma Naïve Bayes Untuk Memprediksi Kredit Macet Pada Koperasi Simpan Pinjam." 4(2).
[20] ojk.go.id, (2017). Ikhtisar Perbankan. [online] Available at: https://www.ojk.go.id/id/kanal/perbankan/ikhtisar-perbankan/Pages/Lembaga-Perbankan.aspx [Accessed 23 Juni. 2020].

Accredited "Rank 4"(Sinta 4), DIKTI, No. 36/E/KPT/2019, December 13th 2019.

Jurnal Mantik is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0).