March-April 2020
ISSN: 0193-4120 Page No. 849 - 856

Analysis of Potential Customer Predictions at Regional Banks to Reduce Credit Risk

Garaika[1], Muhamad Muslihudin[2]
[1] Department of Management, STIE Trisna Negara, South Sumatera, Indonesia
[2] Department of Information System, STMIK Pringsewu, Lampung, Indonesia
[1]Street MP, Bangsa Raja 27, Belitang, East OganKomeringUlu, South Sumatera, Indonesia
[2]Street WismaRini No. 09 Pringsewu, Lampung, Indonesia

Article Info
Volume 83
Page Number: 849 - 856
Publication Issue:
March - April 2020

Article History
Article Received*: 24 July 2019*
Revised*: 12 September 2019*
Accepted*: 15 February 2020*
Publication*: 12 March 2020*

Abstract:
The development of regional banking provides an opportunity for the public to save and borrow to improve the standard of living and business capital. Analysis of potential customer predictions is carried out to reduce credit risk. The risk of crediting is a credit jam or the bank no longer receives interest and installments regularly. This research was conducted to reduce the risk of credit defaults in Lampung regional banks. This research uses the C45 algorithm as a method that helps determine alternative customers who can receive credit from the Lampung area. The C45 algorithm method is used as a decision tree classification model to help determine alternative customers who can receive credit from the Lampung district. To form a decision tree, attributes needed are used as a reference for classification, in this study six attributes are used, namely Address, Age, gender, marital status, employment, income and electricity bills.

Keywords*: Predictions, customers, Lampung regional banks, credit risk.*

INTRODUCTION

Regional banking is regulated in Act No. 13 of 1962[1] Regional Development Bank is a legal entity based on this Law and its position as a legal entity is obtained by the enactment of its establishment regulation. 2 of 1999 dated 31 March 1999 concerning the Changing of Form of Legal Entity from Regional Company (PD) Lampung Regional Development Bank to Limited Liability Company (PT) Lampung Regional Development Bank (Source: https: //banklampung.co.id/).

The development of regional banking provides opportunities for the public to save and loans to improve living standards and business capital. [2] In addition to savings and loans the bank also provides activities that enable the public to invest, distribute or service and all banking activities that use money as a medium. All of these banking activities will certainly affect the community's economy. The bank itself is a public fund collector which is also aimed at the community so that the services offered by this bank are closely related to community economic activities. In the banking sector, there are also creditors, which are institutions or banks that act as lenders. To reduce the risk of bad funds in creditors, it is necessary to have instruments that are used to predict potential customers and the priority to be given loans by banks.

Based on research [3] the C45 algorithm method is used to determine prospective debtors at a BRI bank that is corrupt, the application of the algorithm method is to make a decision system that can provide an alternative decision for decision-makers in determining the level of risk of giving credit to customers. Research [4] with several classification algorithms that are tried on the training data, the classification model C4.5 algorithm has the highest accuracy value. After being implemented to the test data, the customer decides to reject or the credit is received. Application of C4.5 Algorithm has also been implemented in Bank Customers in

March-April 2020
ISSN: 0193-4120 Page No. 849 - 856

Tasikmalaya with the results [5] classification of customer creditworthiness with a total of 8 attributes resulting in an accuracy of 87.36% is a good level of accuracy compared to the results of the study of customer creditworthiness classification using 4 attributes resulting in 79.50% accuracy. Research [6] with the decision tree method using the C4.5 algorithm in prospective bank customers is expected to process information extraction faster and optimal with greater data capacity so that errors caused in decision making are minimized. Research [7] applying the C45 algorithm with attribute selection so that it can reduce the dimensions of the data, as well as identifying features in the data set by the C45 algorithm method, so this study has a good accuracy of 83.67%.

From some of the research conducted above, the C4.5 algorithm can be a prediction for the identification of banking customers with good accuracy. In this research, the decision tree decision method using the C4.5 algorithm will be used to analyze potential customers at regional banks to reduce customer credit risk

Literature review

A. Data Mining

Turban (2005) Data mining is a term used to reduce the knowledge gathering in a database. Data mining is a process that uses statistical techniques, mathematics, artificial intelligence, and machine learning to interpret and idealize from various large databases.[8].

Larose, (2005) According to Gartner Group data mining is a process of finding meaningful relationships, patterns, and tendencies by examining large amounts of data stored in storage by using pattern recognition techniques such as statistical and mathematical techniques.[9]

Pramudiono, (2006) Data mining is a series of processes to explore the added value of a data set in the form of knowledge that has not been known manually.

Data mining is driven by several factors, among others (Lorose, 2005):

1. Rapid growth in data sets.
2. Data storage in a data warehouse, so that all companies have access to a reliable database.
3. An increase in data access through web navigation and the internet.
4. Business competition pressure to increase market control in economic globalization.
5. Development of software technology for data mining (technology availability).
6. The great development in computing capability and capacity building for storage media.

The statement confirms that data mining automation does not replace human intervention. Humans must participate actively in every phase of the data mining process. The greatness of the capabilities of the data mining algorithm contained in the analysis software that is currently available allows the use of fatal errors. The user might apply an inappropriate analysis of the data set using a different approach. Therefore, an understanding of the statistics and structure of the mathematical model is needed that underlies the work of the software (Larose, 2006).[10]

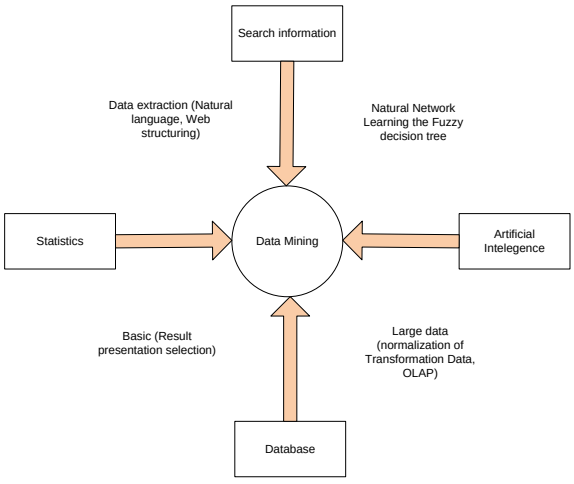

Figure1. The Closest Field of Science

The Relationship with Data Mining is Data mining is not a completely new field. One of the difficulties in defining data mining is the fact that data mining inherits many aspects and techniques from established fields of science in advance. Figure 2.1 shows that data mining has long roots in fields such as artificial intelligence, machine learning,

statistics, databases, and also information retrieval (Pramudiono, 2006). [11]

B. C 4.5 Algorithm

Algorithm C 4.5 is a method for making decision trees based on the training data provided. C 4.5 algorithm is the development of ID3. Some developments carried out in C 4.5 are, among others, able to overcome the missing value, can overcome the continuing data, and pruning. The decision tree is a very powerful and well-known classification and prediction method. The decision tree method converts very large facts into decision trees that represent rules. Rules can be easily understood with natural language. And they can also be expressed in the form of database languages such as Structured Query Language to look for records in certain categories. Decision trees are also useful for exploring data, finding hidden relationships between several potential input variables and a target variable. Because the decision tree combines data exploration and modeling, the decision tree is very good as a first step in the modeling process even when used as the final model of several other techniques.

A decision tree is a structure that can be used to divide large data sets into smaller sets of records by applying a series of decision rules. With each series of divisions, members of the resulting set become similar to each other (Berry danLinoff, 2004). [12]

Research Methodology

C. Algorithm C4.5 method

There are several stages in making a decision tree in C4.5 algorithm they are:

1. Prepare training data, can be taken from historical data that has happened before and has been grouped in certain classes.
2. Determine the root of a tree by calculating the highest gain value of each attribute or based on the lowest entropy index value.

Previously calculated the value of the entropy index first, with the formula:

$$Entropy\,(i) = -\sum_{j=1}^{m} f(i,j).\log 2\, f[(i,j)] \quad (1)$$

Information:
i = case set
m = number of partitions i
f(i,j) = proportion of j to i

Calculate the gain value using the formula:

$$Entropy\ split = -\sum_{i=1}^{p} \frac{n1}{n} .\ IE\,(i) \quad (2)$$

Information:
p = number of attribute partitions
ni = ni proportion to i
n = number of cases in n

Previously calculated the value of the entropy index first, with the formula:

$$Entropy\,(i) = -\sum_{j=1}^{m} f(i,j).\log 2\, f[(i,j)] \quad (3)$$

Information:
i = case set
m = number of partitions i
f(i,j) = proportion of j to i

Calculate the gain value using the formula:

$$Entropy\,(i) = -\sum_{j=1}^{m} f(i,j).\log 2\, f[(i,j)] \quad (4)$$

Information:
p = number of attribute partitions
ni = ni proportion to i
n = number of cases in n

Repeat step 2 until all records are partitioned. The decision tree partition process will stop when:

a. All tuples in the records in the node have the same class.

b. There are no attributes in the partitioned record anymore.

c. There are no records in the empty branch

D. Classification Process

Decision Tree is a classification method that uses a tree structure representation where each internal node is an attribute, each branch is an attribute value,

March-April 2020
ISSN: 0193-4120 Page No. 849 - 856

and each leaf node or terminal node is a class label, and the node the top one is the root node (Han, *et al*, 2006: 291). Decision trees have several ways of determining the size of data in the form of trees, one of which is the C4.5 algorithm. C4.5 algorithm uses gain ratio as a determinant of the root, internal, and leaf.

The calculation of the information needed to classify on tuples D is stated as follows (Han, *et al*, 2006: 297): $(D)=-\Sigma pi$ log2$(pi)mi=1$

Where m is the number of types of value categories in the attribute C, $pi=|Ci,||D|$ is the probability of tuples D that have class Ci.

Suppose there is an attribute A that has v different values $\{a1,2,\ldots av\}$. The A attribute can be used to divide D into v partitions $\{D1,2,\ldots,Dv\}$, where Dj contains tuples D which have a value aj of A. The equation for finding the entropy value from the subset A is as follows (Han, *et al*, 2006: 298): $InfoA(D)=E(A)=\Sigma|Dj||D|vj=1\times Info\ (Dj)$ where $E(A)$ is the entropy of the subset A, v is the number of types of value categories in the subset A, $|Dj||D|$ are the weight of the subset j and the number of samples in the subset that has a value of aj of A, divided by the number of tuples of D.

According to Han, *et al* (2012: 298), the information gain value of the attribute A in the subset D can be calculated by the following equation: $(A)=info(D)-E(A)$

The split information value is used in the search for the gain ratio value to overcome the bias against attributes that have many unique values. The split information and gain ratio equation is stated as follows:

$plitI(D)=-\Sigma|Dj||D|\times log2|Dj||D|vj=1\quad GainRatio\ (A)=Gain(A)SplitInfo\ (A)$

If the attribute has the largest gain ratio value, then the attribute is selected as a split attribute in the decision tree construction (Han, *et al*, 2006: 301).[13][14]

Discussion

Research data

This research data use credit customer data. After conducting a series of stages of data mining, this study uses 22 datasets with 8 attributes consisting of 7 predictor attributes and 1 destination attribute. Meaning predictors used in credit data are Age, Gender, Marital Status, Employment, Income, and Electricity Bill. While the last attribute is as an output in the form of a decision used to distinguish predictions, namely Potential or Potential.

Table1. Customer Credit Data Attribute

Address	Age	Gender	Marital Status	Job	Income	Electricity bills	Classification
Fajaragung	Your	Male	Married	Traders	Great	Are	Potentially
Bumiayu	Middle-aged	Male	Married	Civil servants	Great	Are	Potentially
East Pringsewu	Adult	Male	Unmarried	Private officers	Great	Are	No potential
South Pringsewu	Middle-aged	Male	Married	Private officers	Great	Are	No potential
Pajaresuk	Adult	Male	Unmarried	Private officers	Great	Are	Potentially
East Pringsewu	Adult	Male	Married	Private officers	Great	Are	No potential
Bumiarum	Adult	female	Married	Private	Great	Are	No potential

				officers			
Rejosari	Middle-aged	Male	Married	Civil servants	Great	Are	No potential
Sidoharjo	Adult	Male	Unmarried	Private officers	Great	Are	Potentially
Margakaya	Adult	female	Married	Traders	Great	Are	No potential
Pajaresuk	Middle-aged	Male	Married	Civil servants	Great	Are	Potentially
West Pringsewu	Middle-aged	female	Married	Traders	Great	Are	Potentially
North Pringsewu	Middle-aged	female	Married	Labor	Great	Are	No potential
Podomoro	Adult	Male	Unmarried	Private officers	Great	Are	Potentially
Margakaya	Adult	Male	Unmarried	Traders	Great	Are	Potentially
Podosari	Adult	Male	Unmarried	Civil servants	Great	Are	Potentially
North Pringsewu	Adult	female	Unmarried	Civil servants	Great	Are	Potentially
West Pringsewu	Middle-aged	female	Married	Labor	Great	Are	No potential
East Pringsewu	Middle-aged	Male	Married	Private officers	Great	Are	No potential
Podosari	Your	Male	Married	Labor	Great	Are	No potential
Pajaresuk	Adult	Male	Married	Private officers	Great	Are	No potential
Podomoro	Adult	Male	Married	Private officers	Great	Small	Potentially

B. Sample and Attribute

To form a prediction decision tree for potential customers at regional banks to reduce credit risk, attributes are needed as a reference for classification. The following attributes are used as a reference for classification.

Table1. Attribute

Attribute
Address
Age
Gender
Marital status
Profession
Income
Electricity bills
Classification

After the decision tree is formed, the decision tree is formed after the node has been calculated to determine the entropy value of each attribute and the gain value. This can be seen in the node table below:

Table 2. Table Gain Calculation for Node 1

	Amount	Potentially	No	Entophy	Gain

				potential		
Amount		**22**	**11**	**11**	**1**	
Address						0.782054152
	Fajaragung	1	1	0	0	
	Bumiayu	1	1	0	0	
	East Pringsewu	3	0	3	0	
	Pajaresuk	3	2	1	0.517195221	
	Bumiarum	1	0	1	0	
	Rejosari	1	0	1	0	
	Sidoharjo	1	1	0	0	
	Margakaya	2	1	1	0.405402874	
	West Pringsewu	2	1	1	0.405402874	
	North Pringsewu	2	1	1	0.405402874	
	Podomoro	2	2	0	0	
	Podosari	2	1	1	0.405402874	
	South Pringsewu	1	1	0	0	
Age						0.09225252
	Your	2	1	1	0.405402874	
	Middle-aged	8	3	5	0.877769317	
	Adult	12	7	5	1.011457022	
Gender						0.020242911
	Male	16	9	7	1.053186201	
	female	6	3	3	0.783946123	
Marital Status						0.089086741
	Married	15	5	10	1.002843312	
	Unmarried	7	6	1	0.713920288	
Job						0.308548434
	Traders	4	3	1	0.594674499	
	Civil servants	5	4	1	0.649870822	
	Private officers	10	4	6	0.958388236	
	Labor	3	0	3	0	
Income						0
	Great	22	11	11	1	
Electricity bills						0.029182355
	Are	21	10	11	1.017047056	
	Small	1	1	0	0	

March-April 2020
ISSN: 0193-4120 Page No. 849 - 856

After calculating the node 1 and knowing the value of each attribute and the highest gain, then the 1.1 is calculated as a branch node. After calculating node 1.1 and knowing the entropy value of each attribute and the highest gain, then the node 1.1.2 is calculated as a branch node. After the node calculation is complete, the decision tree is formed based on the results of the node calculation. The following decision tree is formed from the results of the node calculation above.

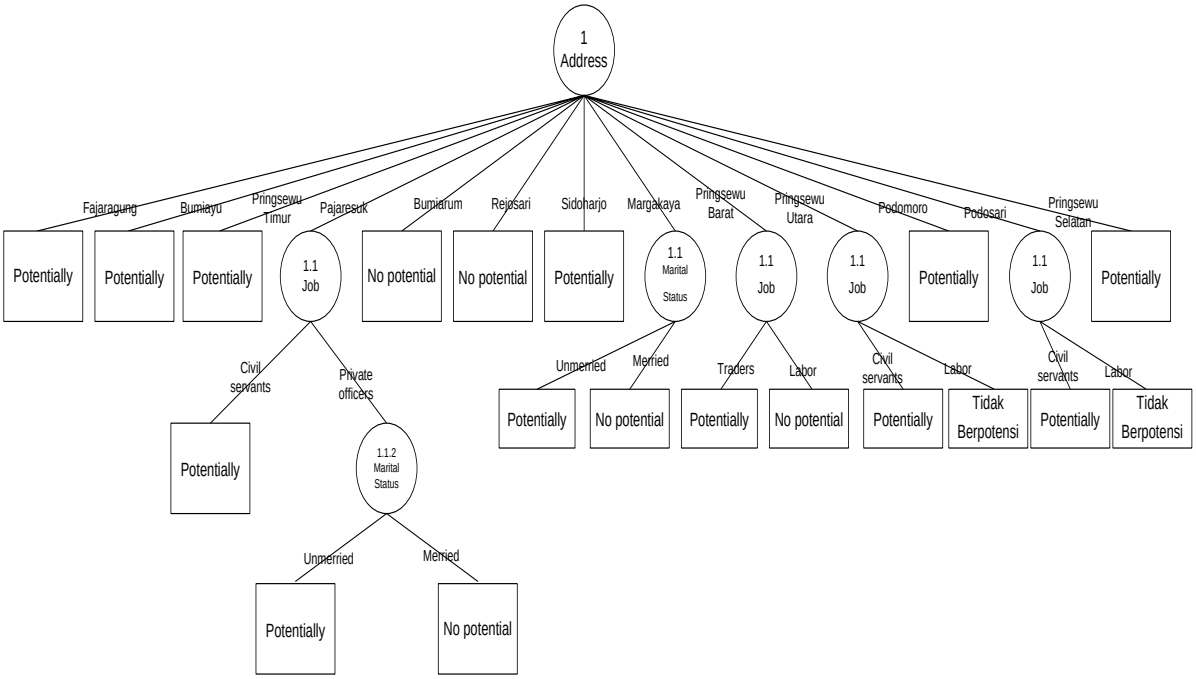

Figure1. Decision tree

Where the rule produced in Figure 1 is:

1. IF Fajaresuk addresses civil servant jobs, THEN Potential
2. IF Fajaresuk addresses the employment of Private Employees Married marital status, THEN No Potential
3. IF Fajaresuk addresses the work of Private Employee marital status, Not Married, THEN Potential
4. IF Margakaya address marital status Not Married, THEN Potential
5. IF address Margakaya marital status Married, THEN No Potential
6. IF addresses West Pringsewu the Trader's work, THEN Potential
7. IF the address of West Pringsewu Labor work, THEN Has No Potential
8. IF North Pringsewu addresses civil servant jobs, THEN Potential
9. IF North Pringsewu addresses Labor work, THEN Has No Potential
10. IF Podosari addresses civil servant jobs, THEN Potential
11. IF Podosari addresses Labor work, THEN Has No Potential

Conclusion

From the decision tree and rule formed where the potential and non-potential status are seen from the attributes in the initial data, namely the Lampung Regional Bank customer data and the results are the Address attribute that becomes the root, the work and the Marriage Status become a branch or node. Classification analysis on Bank Customer data has been successfully carried out with the stages of data preparation, data collection, preprocessing consisting of data cleaning, data integration, data selection, and data transformation, the classification process using the decision tree C4.5 method. The processed

March-April 2020
ISSN: 0193-4120 Page No. 849 - 856

decision tree method can properly identify potential feasibility.

ACKNOWLEDGMENT

Thank you to the STMIK Pringsewu and STIE Trisna Negara for providing funding for this research collaboration conducted in 2019.

REFERENCES

1. P. R. Indonesia, Undang-UndangRepublik Indonesia Nomor 13 Tahun 1962 TentangKetentuan-KetentuanPokok Bank Pembangunan Daerah. 1962, hal. 1–11.
2. B. Lampung, "MeningkatkanLayananBerbasisTeknologi improving," Lampung, 2018.
3. T. A. Muhammad HusniRifqo, "ImplementasiAlgoritma C4.5 UntukMenentukanCalonDebiturDenganMenguku r Tingkat RisikoKreditPada Bank BriCabangCurup," J. Pseudocode, vol. 3, no. 2, hal. 83–90, 2016.
4. H. Marcos dan I. Hidayah, "Implementasi Data Mining UntukKlasifikasiNasabahKredit Bank " X " Menggunakan Classification Rule," Semin. Nas. Teknol. Inf. danMultimed. 2014, hal. 1–7, 2014.
5. C. R. H. Susanto, Neneng Sri Uryani, "Penerapanalgoritma c4.5 untukmenganalisiskelayakanpemberiankreditnasa bah," STMIK TasikMalaya, vol. 4, no. 1, hal. 1–7, 2013.
6. L. N. Rani, "KlasifikasiNasabahMenggunakanAlgoritma C4.5 SebagaiDasarPemberianKredit," J. KomTekInfoFak. IlmuKomput., vol. 2, no. 2, hal. 33–38, 2015.
7. N. N. NandangIriadi, "KajianPenerapanMetodeKlasifikasi Data Mining Algoritma C4.5 UntukPrediksiKelayakanKreditPada Bank Mayapada Jakarta," J. Tek. Komput. AMIK BSI, vol. 2, no. 1, hal. 132–137, 2016.
8. E. Turban, J. E. Aronson, dan T.-P. Liang, "Decision Support Systems and Intelligent Systems," Decis. Support Syst. Intell. Syst., vol. 7, hal. 867, 2007.
9. Y. Zhang, S. Fong, J. Fiaidhi, dan S. Mohammed, "Real-time clinical decision support system with data stream mining.," J. Biomed. Biotechnol.,vol. 2012, hal. 580186, 2012.
10. E. Turban, R. Sharda, dan D. Delen, Decision Support and Business Intelligence Systems. Chapter 6 Artificial Neural Networks for Data Mining, vol. 8th. 2007.
11. M. Muslihudin, "AnalisisPrediksiMahasiswaTidakTepatWaktuM enyelesaikanStudiDenganMenggunakanMetodeA lgoritma C 4.5 (StudiKasus : STMIK Pringsewu)," Tesis IBI Darmajaya, hal. 5–29, 2015.
12. Y. Mardi, "KlasifikasiMenggunakanAlgoritma C4.5," J. Edik Inform., vol. 2, no. 2, hal. 213–219, 2017.
13. A. Srivastava, E.-H. S. H. E.-H. S. Han, V. Singh, dan V. Kumar, "Parallel formulations of decision-tree classification algorithms," Proceedings. 1998 Int. Conf. Parallel Process. (Cat. No.98EX205), vol. 24, hal. 1–24, 1998.
14. J. Han, M. Kamber, dan J. Pei, Data Mining: Concepts and Techniques Third Edition. 2012.