DATA MINING ALGORITHM C4.5 CLASSIFICATION DETERMINATION CREDIT ELIGIBILITY FOR JAYA BERSAMA COOPERATIVES (KORJABE)

Christnatalis [1], Roni Rayandi Saragih[1*], Bobby Christianto Tambunan[1]
[1]Teknik Informatika, Universitas Prima Indonesia
email: [*]ronirayandisaragih@gmail.com

Abstract: This study uses the C4.5 classification algorithm to determine creditworthness, clasification aims to divide the assigned object intoin a number of categories called classes. In this study, the authorusing data mining and C4.5 algorithm as the selection method. The criteria used are loan installments, prospective customer income, termloan time, status of prospective customers. This study resulted in a classification modeldecision tree using the C4.5 algorithm is included in the Excellent category Classification with an accuracy value of 98.33% and a classification error of 1.67%,so that this study uses 70% training data and 30% test data. From resultthe calculation obtained shows that the C4.5 algorithm can be usedto determine the feasibility of granting credit to Koperasi Jaya customers Together (KORJABE).

Keywords: Analysis, Credit Eligibility, C4 Algorithm, Data Mining, Method

Abstrak: Penelitian ini menggunakan metode Algoritma C4.5 klasifikasi untuk menentukan kelayakan kredit, klasifikasi bertujuan untuk membagi objek yang ditetapkan ke dalam satu nomor kategori yang disebut kelas. Dalam penelitian ini, penulis menggunankan data mining dan algoritma C4.5 sebagai metode pemilihannya. Kriteria yang digunakan yaitu , angsuran pinjaman,penghasilan calon nasabah,jangka waktu pinjaman ,status calon nasabah. Penelitian ini menghasillkan model klasifikasi pohon keputusan menggunakan algoritma C4.5 termasuk dalam kategori Excellent Classification dengan nilai akurasi sebesar 98,33% dan klasifikasi eror 1,67%, sehingga penelitian ini kan menggunakan data latih 70% dan data uji 30%. Dari hasil perhitungan yang diperoleh menunjukan bahwa algoritma C4.5 dapat digunakan untuk menen tukan kelayakan pemberian kredit kepada nasabah Koperasi Jaya Bersama (KORJABE).

Kata kunci: Algoritma C4.5, Analisis, Data Mining, Kelayakan Kredit, Metode

INTRODUCTION

Cooperatives have an important role for the Indonesian economy and the development of cooperatives in Indonesia is currently quite fast[1]. In order to develop its activities in the form of loans to cooperative members, it follows the precautionary principle applied to the banking world by conducting credit analysis. to potential borrowers. At the time of providing loans, the cooperative conducts an analysis of the data of cooperative members who have the opportunity to be given loans, and also in providing loans to cooperative members there are also various problems, namely the inability of cooperative members to pay off a number of loans that have been given by the cooperative [2]. In decisions made by cooperatives, cooperatives do not only lend but must reduce loan risk in determining the amount of loans made by prospective customers. In granting credit, it is mandatory to go through several terms to analyze its feasibility. An in-depth analysis of both the data held by the prospective borrower and the authenticity of the supporting data for the feasibility of receiving a loan. The in-depth evaluation of credit awards was known to use the 5Cs, namely character (personality), capacity (ability), capital (capital), condition of economy (economic conditions), collateral (collateral). In addition to the 5C principle in selecting all incoming credits, there are several considerations that become in passing the application, namely completeness of documents, credit risk based on large aspects of income and work history, having a smooth credit history. This is done so that there is an adjustment in the ability to make credit payments, so that things don't happen that make the credit bad [3]. The in-depth evaluation of credit awards was known to

use the 5Cs, namely character (personality), capacity (ability), capital (capital), condition of economy (economic conditions), collateral (collateral). In addition to the 5C principle in selecting all incoming credits, there are several considerations that become in passing the application, namely completeness of documents, credit risk based on large aspects of income and work history, having a smooth credit history. This is done so that there is an adjustment in the ability to make credit payments, so that things don't happen that make the credit bad. The in-depth evaluation of credit awards was known to use the 5Cs, namely character (personality), capacity (ability), capital (capital), condition of economy (economic conditions), collateral (collateral). In addition to the 5C principle in selecting all incoming credits, there are several considerations that become in passing the application, namely completeness of documents, credit risk based on large aspects of income and work history, having a smooth credit history. This is done so that there is an adjustment in the ability to make credit payments, so that things don't happen that make the credit bad. In addition to the 5C principle in selecting all incoming credits, there are several considerations that become in passing the application, namely completeness of documents, credit risk based on large aspects of income and work history, having a smooth credit history [5][6]. This is done so that there is an adjustment in the ability to make credit payments, so that things don't happen that make the credit bad. In addition to the 5C principle in selecting all incoming credits, there are several considerations that become in passing the application, namely completeness of documents, credit risk based on large aspects of income and work history, having a smooth credit history. This is done so

that there is an adjustment in the ability to make credit payments, so that things don't happen that make the credit bad.

The method used in this research is C4.5. classification completion procedure determines creditworthiness, classification aims to divide objects assigned to only one category number which is called Tree class. With customer classification, if there is a problem with the same case, the cooperative only needs to look at the rules that have been formed from the resulting decision tree[7][8]. With a decision tree decision tree using the C4.5 algorithm requires a faster and optimal time. The scope extraction process uses more data capacity, because errors caused in decision making are minimizedThese results were obtained using K-NN based on Forward Selection. The use of the Forward Selection feature in data processing will affect the accuracy of the output obtained. While using the C4.5 classifier algorithm in determining creditworthiness, he obtained an accuracy value using training data of 270 customers of 88.52%. The use of the Forward Selection feature in data processing will affect the accuracy of the output obtained. While using the C4.5 classifier algorithm in determining creditworthiness, he obtained an accuracy value using training data of 270 customers of 88.52%.

The use of the Forward Selection feature in data processing will affect the accuracy of the output obtained. While using the C4.5 classifier algorithm in determining creditworthiness, he obtained an accuracy value using training data of 270 customers of 88.52%.To solve the above problems, the C4.5 algorithm data mining technique is used for decision making in choosing the loan amount. The application of the algorithm is a case-solving solution that is often used in solv-

ing problems in classification techniques that have characteristics, namely the process of determining the entropy value and gain value as a function of the possibility of each criterion as a reference decision that continues to use the decision tree-based output process of the C4 algorithm. 5This is already quite high, therefore it can be concluded that this algorithm is suitable to be implemented in this research regarding the Classification of Determination of Credit Eligibility for the Jaya Bersama Cooperative (KORJABE).

There are previous studies, including C4.5 algorithm to process large amounts of data to find patterns of sales of goods. Based on the decision tree formed, predictions will be made on the data collection of new sales of goods. Thus, Dutalia Supermarket can determine the minimum stock of goods [10]. The results of graduation predictions from this research application can help the study program section find out students' graduation status. This can be a recommendation for taking courses for students for the next semester, such as thesis and internships. With this, students can graduate at least on time [11]. The purpose of this study is to classify dropped out and active students using the C4.5 algorithm as a reference in making policies and actions to reduce the number of dropped out students. The classification results from the C4.5 algorithm were evaluated and validated with a confusion matrix and Receiver Operating Characteristic (ROC) curve to determine the level of accuracy of the C4.5 algorithm in classifying potential dropout students. The results of the evaluation and validation obtained an accuracy rate of 97.75%. The rules obtained from the classification with the C4.5 Algorithm if applied to the new data obtained validation results with an

accuracy rate of 90.0% [12].

METHOD

Types Of Research

This study uses an associative quantitative approach with causality type. The quantitative approach is in the form of an associative causality type, which is a type of research that explains the influence of independent variables on the dependent variable. So, the research conducted is quantitative research, because it can be calculated mathematically, and the purpose of this classification research is to find functions and models that can distinguish or describes a concept or data class to estimate the unknown class of an object.

Materials

The data used in this study is data on credit customers with the attributes of id_customer, name, age, status, installment, time period, and description. The data taken from 2013 to 2020 is 1000 customer data.

Research Framework

The following is an explanation of the stages that will be carried out in the research:

a. Dataset
   This stage is the process of searching for customer data datasets.
b. Data Cleaning.
   At this stage, deletion of data with more than one primary key and deletion of data that is null or missing value, duplicate data.
c. Data Normalization
   It is the stage of forming data relations into normal form to eliminate some of the data that is

still ambiguous. This normal form is a rule that must be met in database relations. This normalization aims to normalize the data that was originally obtained is still messy so that a good and correct data is formed and produces a table with good quality.

d. Data Transformation
   This stage needs to be done because there are several data mining methods such asassociation and clustering methods that require a special data format before being applied. In data transformation, the process of selecting data that is only needed by the data mining used is carried out.This phase determines the quality results of data mining. The purpose of data transformation is also to remove noise or data that contains errors.

e. C4.5 Algorithm
   After processing the data, it is continued with the stages of the C4.5 Algorithm, namely the description tree or decision tree is a classification method that uses a tree structure representation where each internal node is an attribute, and each leaf node is an attribute. nodes) where the terminal node is the class label and the top node is the root node. The decision tree has several ways of determining the size of the data in the form of a tree, one of which is the C4.5 algorithm. The C4.5 algorithm uses the gain ratio as a determinant of roots, interiors and leaves.

f. Evaluation
   The evacuation stage is called the classification stage because at this

stage testing for accuracy will be determined. The testing stage is to see the accuracy results in the classification process of the C 4.5 classifier algorithm.

1. Data Analysis
   The data used in this study are customer credit data from the Jaya Bersama Cooperative (KORJABE), with the following explanation:

a Customer credit data with attributes id_customer, name, age, status, income, installments, period, and payment information given to the administration and then given to the head of the cooperative so that it is in the form of a credit data report customers, current payments and non-performing payments.

b The customer provides credit documents to the administration section, and the administration section provides it to the head or chairman of the cooperative. The head of the cooperative provides reports on credit data for current payment customers and credit data on customers who make non-performing payments to the administration, and the administration provides credit to the cooperative's employees. The criteria for credit delays in cooperatives are special supervision, substandard and non-performing.

RESULT AND DISCUSSION

In this study, primary data is used with data on credit customers with attributes such as nik_customer, age, customer status, income, installments, time period, and information. From the data collected as many as 1000 customer data records that will be processed and analyzed using the C4.5 Algorithm. The software used for this research is Microsoft Excel 2007 in preprocessing the data (preprocessing) and Rapid Miner Studio Version 9.6 in processing and analyzing the C4.5.

In Table 1. The dataset collected by the Jaya Bersama cooperative to determine creditworthiness will be selected for the variables before being processed and analyzed using the C.45 algorithm. Prior to classifying credit customers, the data preprocessing stage is carried out including data cleaning, data integration, data selection, and data transformation. Data cleaning is carried out on data that does not have complete attributes. The next step is to normalize the data by determining the variables that determine eligibility. The variables used are salary ratio, payment period, customer status, age and customer payment information. After the normalization phase, the next phase is the data transformation phase, in this phase the date attribute is removed and the customer name attribute is not used in the analysis. The data transformation can be seen in Table 2.

Table 1. Customer Data Table

Date	Customer Name	Install-ment	Income	Time period	Status	Age	Information
13/10/2013	Novi Adistara	575000	5000000	12	Married	28	Fluent
24/07/2013	Novi Ansari	575000	5000000	12	Married	46	Fluent
07/08/2013	Meiri Aprilina	345000	3000000	6	Married	24	Fluent
15/05/2013	Khairul Fahmi	345000	3000000	12	Married	28	Fluent
21/12/2013	Agus Susilo	2000000	6000000	12	Not married	30	Fluent
03/07/2013	Syah Devi An-driani	2000000	3000000	18	Married	46	Congested
12/02/2013	Nirwana	575000	1500000	12	Not married	24	Congested
08/06/2013	Helmiyanti Fer-nanda SH	345000	1500000	12	Not married	30	Congested
16/06/2013	Nurjannah Nasu-tion	345000	1500000	12	Not married	22	Congested
15/06/2013	Burhan	850000	6000000	12	Married	24	Fluent
09/05/2013	Rouli Ida M Hu-tagaol	515000	3500000	12	Not married	32	Fluent
18/02/2013	Tetti Anna Gultom	345000	3500000	12	Married	24	Fluent
23/06/2013	Iglima	575000	1500000	6	Not married	24	Congested
28/09/2013	Antoni	575000	1500000	12	Not married	26	Congested
---	---	---	---	---	---	---	---
24/12/2013	Subono	1000000	6000000	12	Married	23	Fluent

Table 2. Result of Transformation of Customer Data

Ratio	Time period	Status	Information
25	12	Marry	Fluent
25	12	Marry	Fluent
15	6	Marry	Fluent
12	12	Marry	Fluent
66	12	Not married yet	Fluent
66	18	Marry	Congested
15	12	Not married yet	Congested
14	12	Not married yet	Congested
14	12	Not married yet	Congested
11	12	Marry	Fluent
11	12	Not married yet	Fluent
10	12	Marry	Fluent
10	6	Not married yet	Congested
10	12	Not married yet	Congested
12	---	---	---
66	12	Marry	Fluent

In table 2 The transformed dataset will be processed by applying the C4.5 Algortima using Rapid Miner Studio software. The display of the C4.5 algorithm implementation circuit using the Spilt Validation Operator can be seen in Figure 1.
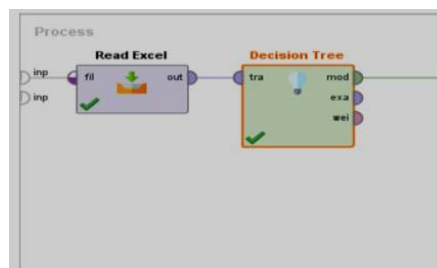

Figure 1. Testing the C4.5 Algorithm with Rapid Miner

1. 0.90 – 1.00 = Excellent Classification
2. 0.80 – 0.90 = Good Classification
3. 0.70 – 0.80 = Fair Classification
4. 0.60 – 0.70 = Poor Classification
5. 0.50 – 0.60 = Failure

Based on the training data and test data, the best C4.5 classification modeling lies in the Linear data sampling model, namely the highest Accuracy value and the lowest Classification Error with an accuracy value of 98.33% and a classification error of 1.67% included in the Excellent Classification category, so that this study uses 70% training data and 30% test data so that it produces knowledge of determining creditworthiness to prospective customers in the form of a decision tree (decision tree).

Based on the decision tree in Figure 3 it can be seen that the time period becomes the main node (root) to determine eligibility in determining credit. With an accuracy rate of 98.33% and a classification error of 1.67%. After the decision tree is formed, the next process is to change the shape of the decision tree into a rule form. This study produces several rules that influence the determination of creditworthiness which is a factor in the
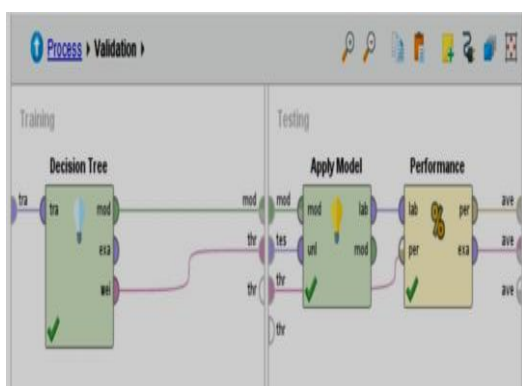

Figure 2. Testing the C4.5 Algorithm with Rapid Miner

In testing the C4.5 algorithm is included in supervised so it requires training data

Table 3. Training Data 70% and Test Data 30%

Data Sampling	Test	
	Accuracy	Classification Errors.
Linear	98.33%	1.67%
Shuffle	95.33%	4.67%
Stratified	95.67%	4.33%
Automatic	98.33%	1.67%

Performance Assessment of Split Validation Operator Accuracy can be classified into 5 categories:

smooth or non-performing of customer payments.

1. and if the ratio is less than 45% to 36% then the payment is stuck.
2. If the salary ratio is more than 45% to 54%, tenure is less than 15 months, married status, then the payment is smooth and if the salary ratio is less than 45% to 36% then the payment is stuck.
3. If the salary ratio is more than 51% to 54%, the term is 9 to 15 months, then the payment is stuck.

4. If the salary ratio is less than 51% to 36%, period of 9 to 15 months, married status, then the payment is stuck.
5. If the salary ratio is more than 45% to 51%, the period of 9 to 15 months, the status is not married, then the payment is smooth.
6. If the salary ratio is less than 45% to 36%, the period of 9 to 15 months, the status is not married, then the payment is stuck.
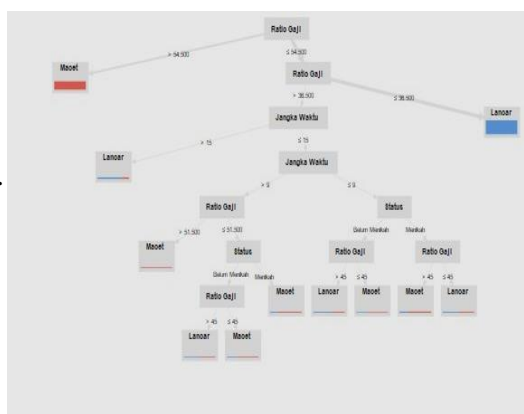

Figure 3. Korjabe Credit Eligibility Decision Tree

Based on the training data and test data, the best C4.5 classification modeling lies in the Linear data sampling model, namely the highest Accuracy value and the lowest Classification with an accuracy value of 98.33% and a classification error of 1.67%. The results of testing the C4.5 algorithm using Rapid Miner tools in determining the accuracy value and error classification value as shown in Figure 3.

		Errors.
Linear	98.33%	1.67%
Shuffle	95.33%	4.67%
Stratified	95.67%	4.33%
Automatic	98.33%	1.67%

In this study, testing was also carried out using the Naive Bayes method with the aim of comparing the value of accuracy and classification of errors. The results of testing the Naive Bayes algorithm after testing using the Split Validation Operator with 70% training data and 30% test data resulted in an accuracy value of 95.33% and an error classification value of 1.67%. And information can be taken from the data tested using Rapid Miner tools that the C4.5 algorithm has a higher accuracy rate than the Naive Baiyes algorithm.

CONCLUSION

Based on the last decision tree, a rule with the following rules is obtained If the salary ratio is less than 45% to 36%, and the period is 9 to 15 months, the status is not married, then the pay-

Table 4. Training Data 70% and Test Data 30%

Data Sampling	Test	
	Accuracy	Classification

ment is declared non-performing. Meanwhile, if the salary ratio is more than 45% to 51%, the period is 9 to 15 months, the status is not married, then the payment is smooth.This study resulted in a decision tree classification model using the C4.5 algorithm which is included in the category Excellent Classification with an accuracy value of 98.33% and 1.67% error classification, so this study uses 70% training data and 30% training data. Data *Mining* with Decision Tree can make a decision by considering each customer acceptance based on customer payments are smooth or bad based on the C4.5 algorithm method

BIBLIOGRAPHY

[1] Kholifah, A. N., & Insani, N. (2016). Classification Analysis on Credit Cooperative X Customers Using Decision Tree C4. 5 Dan Naive Bayes. Journal of Mathematics and Science Education, 5(6), 1–8.

[2] Kurniawan, Y. I., & Angguntina, F. (2018). Application to predict the eligibility of prospective credit members at KSPPS BMT Arta Mandiri Wonogiri using the K nearest neighbor algorithm. JISka, 3(2), 84–94.

[3] Rustiyan & Mustakim, (2018). Application of Fuzzy C Means Algorithm for Problem Analysis of Mandatory Savings for Cooperative Members. Journal of Information Technology and Computer Science, 5(2), 171. https://doi.org/10.25126/jtiik.201 852605

[4] Tree, D., & Credit, K. (2016). EMPLOYEE COOPERATIVE CREDIT FEASIBILITY CLASSIFICATION MODEL WITH Putri's DECISION TREE ALGORITHM. 2010, 263–268.

[5] With, P., Method, C., Naïve, D. A. N., Case, S., Bmt, K., Sidogiri, U. G. T., & Cikarang, C. (2020). 1) , 2). 10, 75–86.

[6] Harlina, S. (2018). Data Mining on Credit Feasibility Determination Using K-Nn Algorithm Based on Forward Selection Data Mining on Credit Feasibility Determination Using K-Nn Algorithm Based on Forward Selection. CCIT Journal, 11(2), 236–244. https://doi.org/10.33050/ccit.v11i 2.591.

[7] Algorithms, D. A. N., & Based, C. (2020). The publication of the analysis of the lending system for customers at the Kroya branch of Artha True Cooperative using data mining and the web-based C4.5 algorithm.

[8] Khasanah, S. N. (2017). Application of the C4.5 Algorithm for Credit Eligibility Determination. Nusa Mandiri Techno Journal, 14(1), 9–14.

[9] Sembiring, M. Ardiansyah. "Penerapan metode decission tree algoritma c45 untuk memprediksi hasil belajar mahasiswa berdasarkan riwayat akademik." *JURTEKSI ROYAL Vol 3 No 1* 3 (2016).

[10] Lalo, Aprilianus Kristianus, Patrisius Batarius, and Yovinia Carmeneja Hoar Siki. "Implementasi Algoritma C4. 5 Untuk Klasifikasi Penjualan Barang di Swalayan Dutalia." *Jurnal*

Teknik Informatika UNIKA Santo Thomas 6.1 (2021): 1-12.

[11]   Kamagi, David Hartanto, and Seng Hansun. "Implementasi Data Mining dengan Algoritma C4. 5 untuk Memprediksi Tingkat Kelulusan Mahasiswa." *Ultimatics: Jurnal Teknik Informatika* 6.1 (2014): 15-20.

[12]   Andriani, Anik. "Penerapan Algoritma C4. 5 Pada Program Klasifikasi Mahasiswa Dropout." *Seminar Nasional Matematika*. 2012.

[13]   Swastina, Liliana. "Penerapan Algoritma C4. 5 Untuk Penentuan Jurusan Mahasiswa." (2013).