Prediction of Credit Card Using the Naïve Bayes Method and C4.5 Algorithm

Ketjie*, Viny Christanti Mawardi¹, Novario Jaya Perdana¹
Informatics Engineering Department, Faculty of Information Technology, Universitas Tarumanagara Indonesia

*ketjie.535150013@stu.unitar.ac.id

Abstract: Prediction credit card submission is a system that is able to provide an assessment of alternatives in order to help credit card applicants in making decisions. Many methods can be used in building a prediction system for credit card submissions. This research will compare two methods in predicting credit card submission, namely the naïve bayes method and c4.5 algorithm, case study of Credit Card Submission Prediction, the results of the research are, knowing the level of accuracy of the two methods. Criteria that form the basis of decision making include age, sex, recent education, marital status, number of dependents, type of company, monthly income, and salary slip. The final result found that the naïve bayes method and the c4.5 algorithm are relatively the same.

1. Introduction
Credit Card is an APMK (Card-Based Payment Instrument) that can be used to make payments for obligations arising from an economic activity, including spending transactions and or to make cash withdrawals. In the development and need of credit cards, certainly as a credit card service applicant requires a system for predicting receipt of employees whether credit card submissions can be accepted. One such solution is to create a Credit Card Submission Prediction application.

There are several methods for predicting credit card. One of them is the Naïve Bayes method and the C4.5 algorithm. Naïve Bayes method and C4.5 algorithm are part of the data mining classification method. Naïve Bayes is a classification using probabilities and statistics to be able to predict future opportunities based on previous experience so that it is known as Bayes theorem. The theorem is combined with "Naïve" which assumes the conditions between attributes are independent of each other[1].

Besides Naïve Bayes, another method used is the C4.5 algorithm. C4.5 Algorithm is one of the algorithms used for classification or grouping. The basis of the C4.5 algorithm is the formation of a decision tree (Decision Tree)[2]. The branch in the decision tree is a classification question and the leaf of the branch is the class in the branch. Expected output in making Predictions for Credit Card Submissions Using the Naïve Bayes Method and C4.5 Algorithm is to provide information on the status of credit card applicants.

2. Methodology
The method used to deal with existing problems aims to show how a data mining model can be used to find out information received or not credit card applicants. There are two methods used in predicting the submission of this credit card, namely the naïve bayes method and the c4.5 algorithm, the two methods will be compared to the results of the accuracy, recall and precision of the formula to calculate the probability and decision tree:

2.1. Naïve Bayes Method
Naïve Bayes is a classification method rooted in the Bayes theorem. The classification method using probability and statistical methods proposed by the British scientist Thomas Bayes, which predicts future opportunities based on experience in the past so that it is known as the Bayes Theorem. The main characteristic of Dr. Naïve Bayes is a very strong assumption (naïve) of independence from each condition / event. The steps of the Naïve Bayes algorithm are as follows[3]:

Ketjie, Viny Christanti Mawardi, Novario Jaya Perdana
Informatics Engineering Department, Faculty of Information Technology, Universitas Tarumanagara Indonesia

1st TICATE 2020
a. Calculate the conditional probability of each attribute by the formula:

\[ P (A \mid B) = \frac{P(A \cap B)}{P(B)} \]  

(1)

Where A is a data sub-attribute of each attribute, B is a data hypothesis A class, P (A \mid B) is a probability of hypothesis A based on condition B, P (B) is the probability of a class B hypothesis, and P (A \cap B) is a data slices from A and B.

b. If given k attributes that are independent of each other, the probability value can be given as follows:

\[ P (X_1, ..., X_k \mid C) = P (X_1 \mid C) * ... * P (X_k \mid C) \]  

(2)

Where P(X1, ..., Xk \mid C) is a probability of attribute X in class C, P (X1 \mid C) is a probability of the 1st attribute of X in class C, and P (Xk \mid C) is the probability of the k-th attribute of X in class C.

c. Normalize the likelihood to find the highest similarity with the formula:

\[ \sum_{i=1}^{n} P(X_1) = \frac{P(X_1)}{P(x_1 \cdot X_1 + ... + P(X_n))} \]  

(3)

Where n is a amount of data in class

2.2. C4.5 Algorithm

C4.5 algorithm is an algorithm of the decision tree method. Decision tree is a classification and prediction method that uses a tree structure representation that contains alternatives to solve a problem step in determining a decision tree with the C4.5 algorithm[3]:

a. Entropy formula:

\[ \text{Entropy} (S) = \sum_{i=1}^{n} p_i * \log_2 p_i \]  

(4)

Where S is a sub-attribute set, n is a number of case partitions in the class, and pi is a proportion from Si to S.

b. Gain formula:

\[ \text{Gain} (S, A) = \text{Entropy}(S) - \sum_{i=1}^{n} \left( \frac{|S_i|}{|S|} \right) * \text{Entropy}(S_i) \]  

(5)

Where S is a case set, A is a attribute, n is a number of partition attribute A (example = n for path attribute 2), | Si | is a number of cases in the registration class, and | S | is a number of cases in class no.

3. Results

This application was made to recommend credit card applicants and compare the naïve bayes method and C4.5 algorithm, this application can be used by the general public as a reference consideration to find out whether or not to be able to get credit card services. The name of this application is "Predicting Credit Card Submissions Using the Naïve Bayes Method and C4.5 Algorithm" using the PHP and MySQL programming languages. The purpose of this application is to compare the naïve bayes method.
and the C4.5 algorithm. This program consists of 5 main modules, namely the main page of credit card predictions, Prediction Form for a Credit Card, Prediction Results, Main Analyst Login Page, and Modul Applicant.

In figure 1, the main page of the credit card prediction, the applicant who will make a credit card prediction before will be given an explanation of what the credit card is and what data is needed before proceeding to the stage of filling out the form to predict the credit card.

In figure 2, we can see the interface of the credit card prediction form. In this module, there are 14 input criteria that must be entered by credit card applicants, then there are eight criteria used to calculate the Naive Bayes method and C4.5 algorithm, namely age, gender, last education, status, number of
dependents, company, income monthly, salary slip. After filling in the data in Figure 2, the credit card prediction results will appear or not, eight criteria can be seen in table 1.

| No. | Name                  | Information                                      |
|-----|-----------------------|--------------------------------------------------|
| 1   | Age                   | Teenager = 17-19                                 |
|     |                       | Adult = 20 - 50                                  |
|     |                       | Elderly = 51<                                    |
| 2   | Gender                | Male                                             |
|     |                       | Women                                            |
| 3   | Last Education        | Diploma University/S1                           |
|     |                       | Master/S2 etc                                    |
| 4   | Marital Status        | Single                                           |
|     |                       | Married                                          |
|     |                       | Widow or Widower                                 |
| 5   | The Number of Dependents | Dependents T0 if 0                             |
|     |                       | T1 if 1                                          |
|     |                       | T2 if 2                                          |
|     |                       | T3< if 3<                                        |
| 6   | Type of Company       | PNS                                             |
|     |                       | BUMN                                            |
|     |                       | Swasta                                          |
| 7   | Monthly Income        | Low if < Rp. 2.000.000                           |
|     |                       | Intermediate if Rp. 2.000.000 – Rp. 3.500.000   |
|     |                       | High if > Rp. 3.500.000                          |
| 8   | Slip Salary           | Yes                                              |
|     |                       | No                                               |

Table 1 Criteria

Figure 3 shows the prediction results of the Naive Bayes method or C45 algorithm in the program, if the credit card applicant is accepted, then the credit card applicant can submit to the bank that serves the credit card.
In Figure 4 can be seen the total training data and testing data that has been input into the program, the total training data received and not, and the total testing data received and not. This page can only be accessed by analysts and assessors, the main function of this page is to view the inputted training, testing and applicant data.

In Figure 5 can be seen the total training data and testing data that has been input into the program, the total training data received and not, and the total testing data received and not. This page can only be accessed by analysts and assessors, the main function of this page is to view the inputted training, testing and applicant data.
In figure 5 can be seen data of credit card applicants who have received or not, this page can only be accessed by users who have access rights such as analysts and assessors, this page serves to edit, delete and add credit card applicant data, on this page can be seen the results Naive Bayes method prediction and C4.5 algorithm. The output of the application "Predicting Credit Card Submissions" can produce viable recommendations and not for credit card applicants. the first submission was tested by 10 users based on 20 training data and subsequent submissions were tested by 50 users.

4. Conclusion
1. Based on the analysis in this study, when using 20 training data and 10 testing data obtained classification accuracy for 10 testing data is 100%
2. In the results of accuracy, recall and precision, it can be concluded that the naïve bayes method and the c4.5 algorithm in the Credit Card Submission Prediction application both have precise results.
3. The results of the Naive Bayes method and the C45 algorithm have the same results due to data limitations, because not all prospective applicants want to share their personal data.
4. Based on the analysis in the credit card prediction application, it can be concluded that this application can provide credit card applicant recommendations.

5. Suggestion
Research carried out certainly can not be separated from a shortage. Therefore, there are some suggestions that the author gives as a reference in future research regarding the classification of data mining using the naïve bayes method and the c45 algorithm:
1. For data sets used in research, it should be obtained directly in a bank
2. Data needs to be added in order to produce a better level of accuracy

6. Reference
[1] George Dimitoglou, James A. Adams, and Carol M. Jim, "Comparison of the c4.5 and a Niave Bayes Classifier for the Prediction of Lung Cancer Survivability", Journal of Computing, Vol. IV, Number 1, (August, 2012), p. 1
[2] Inggrid Kurniawan, data mining, https://sis.binus.ac.id/2019/04/04/data-mining-2/., 14 April 2020
[3] M.Karim and R. Rahman, "Decision Tree and Naïve Bayes Algorithm for Classification and Generation of Actionable Knowledge for Direct Marketing", Journal of Software Engineering and Applications, Vol. VI, Number 4, (2013), p. 1
[4] Muchamad Piko Henry Widiarto, Taking Timely Graduation Patterns at STMIK AMIKOM Yogyakarta Students Using Data Mining Algorithm C4.5, http://repository.amikom.ac.id/files/Publikasi_04.22.0400.pdf, 01 March 2020.