

PAPER • OPEN ACCESS

## The Development of Bank Application for Debtors Selection by Using Naïve Bayes Classifier Technique

To cite this article: S L B Ginting *et al* 2018 *IOP Conf. Ser.: Mater. Sci. Eng.* **407** 012177

View the [article online](#) for updates and enhancements.

You may also like

- [Novel three-dimensional interphase characterisation of polymer nanocomposites using nanoscaled topography](#)  
Mohanad Mousa and Yu Dong
- [The development of bank applications for debtors' selection by using Naive Bayes classifier technique](#)  
S L B Ginting, J Adler, Y R Ginting et al.
- [Nitrogen-Doped Bagasse-Derived Carbon/Low Pt Composite as Counter Electrodes for High Efficiency Dye-Sensitized Solar Cell](#)  
Changli Xiang, Tao Lv, Chika A. Okonkwo et al.



**ECS**  
The  
Electrochemical  
Society  
Advancing solid state &  
electrochemical science & technology

**DISCOVER**  
how sustainability  
intersects with  
electrochemistry & solid  
state science research

# The Development of Bank Application for Debtors Selection by Using Naïve Bayes Classifier Technique

S L B Ginting<sup>1\*</sup>, J Adler<sup>1</sup>, Y R Ginting<sup>1,2</sup> and A H Kurniadi<sup>1</sup>

<sup>1</sup>Computer System Study Program, Universitas Komputer Indonesia, Jl. Dipati Ukur 112-116, Bandung 40132, Indonesia

<sup>2</sup>Department of Mechanical Engineering, Universitas Riau, Kampus Bina Widya KM 12,5, Simpang Baru, Pekanbaru 28293, Indonesia

\*selvia.lorena@email.unikom.ac.id

**Abstract.** The purpose of this study is to create an application that functions automatically with high accuracy when analyzing bank customer data. This needed due to non-performing loans occurring frequently caused by the inaccuracy of credit analysts in the assessment of creditworthiness. This can be seen in the incident occurred in a public bank located in Bandung. This bank does not have the database that serves to accommodate data history and the method used in assessing creditworthiness is merely based on the simple statistical analysis. This leads to reduced accuracy and speed in the decision-making process. This research applies Naïve Bayes Classifier (NBC) method, a Data Mining technique. This helps credit analysts to select customers who are truly eligible to be given credit so that non-performing loan can be avoided. NBC calculates the probability of one class from each group of attributes and determines which class is most optimal. The accuracy of the NBC sampling test from 500 data is 95% compared to the decision made by a credit analyst. It can be concluded that this application is very helpful for credit analysts in recommending customers who are eligible for a loan to the bank's decision maker.

## 1. Introduction

In the banking world, one of the services offered to customers is loaning or credit. Credit will be given to qualified customers in accordance with the terms set by the bank. However, with the large number of customers applying for credits, errors performed by credit analysts occur frequently in selecting customers who are qualified to be given a loan [1,2]. This results in the individual as well as corporate customers' inability to repay credit in a timely manner, which is also known as a non-performing loan. A non-performing loan is a condition wherein a customer is unable to pay the minimum payment that is overdue for more than three months [3]. Due to the presence of this condition, there is a need of a system that is able to make selections rapidly and make predictions accurately by utilizing existing information as a benchmark in the provision of credit to customers who are truly qualified [4,5].

In this research, an application will be developed by applying technology in the field of data mining that is able to function in finding data patterns that have not been known of previously. Data mining is an activity that involves the collection and use of historical data in order to find the regularity of patterns or relationships in large datasets [6,7]. The identification of data pattern is approached by conditional probability. *Naïve Bayes Classifier* (NBC) is one of the classification techniques in data mining that is able to predict the future opportunities based on the experience and then find a function that connects



past data pattern with the desired output [7]. This method will form a classification model after a number of instructional characteristics are given to determine the class that is suitable for an analyzed sample. By using NBC method in this application, it is expected that it is able to classify the data of customers who apply for credit in a large scale and work as a tool for credit analysts in determining whether or not a customer is qualified to receive a loan.

The NBC technique has been applied to other studies. Patil and Sherekar have compared the evaluation results of Naïve Bayes classifier with J48 in the context of bank defaulters [8]. Naïve Bayes classifier has shown good results; however, J48 was found to be more cost-efficient. Ginting and Trinanda found Naïve Bayes was more accurate in book searching for library applications [9,10]. The list of books can be displayed not only based on title, category or author but also based on the descriptions of the books. In his paper, Bustami applied the Naïve Bayes algorithm to classify insurance customer data based on payments history [11]. He used the result to determine future eligible customers.

The purpose of this study is to develop an application by applying the Naïve Bayes Classifier, a well-known data mining technique, to examine the feasibility of an individual or corporate debtor. With the help of this application, banks are expected to be able to decide loan certainty more quickly and accurately. This application can also suppress the operational expenses with the ability to determine the area of credit that becomes the focus of the bank in the future. There have only been few studies commenced that were aimed at bank data analysis, particularly in Indonesia. Thus, the expectation from this study is that the result can be a reference for other studies of the same field.

## 2. Methods

### 2.1. Classification

Data classification is a process that finds the same properties on a set of objects within a database and classifies them into different classes according to the established classification model. The purpose of classification is to find a model of a training set that differentiates attributes into appropriate categories or classes, the model is then used to classify the attributes of which the class are previously unknown [6,7].

### 2.2. Naïve Bayes Technique

Naïve Bayes Classification is a classification using probability and statistics method that predicts future opportunities based on the experience in the past known as Bayes theorem. The theorem is combined with Naïve in which the attribute condition is assumed independent [6,7,11].

### 2.3. Naïve Bayes Classifier Technique

NBC is a classification technique rooted in the Bayes theorem. The main characteristic of the Naïve Bayes Classifier is a very strong assumption (naive) to the independence of each condition or event. Before explaining the Naïve Bayes Classifier, Bayes theorem will be discussed first as, it is the base of the method. On the Bayes theorem, if there are two separate events (eg, A and B), the Bayes theorem is formulated as follows: [11,12,13].

$$P(A|B) = \frac{P(A)}{P(B)} P(A|B) \quad (1)$$

Here the high independence of assumption is used (naive), in which each guideline (F1, F2 ... ..Fn) is independent (independent) of each other. With this assumption, the application of a similarity is valid as follows:

$$P(F_i - F_j) = \frac{P(F_i \cap F_j)}{P(F_j)} = \frac{P(F_i)P(F_j)}{P(F_j)} = P(F_i)$$

For  $i \neq j$ , so that

$$PP(F_i|C, F_j) = P(F_i|C) \quad (2)$$

From the equation above it can be concluded that the naive independence assumption makes opportunity requirements to become possible. Furthermore, the description  $(F_1, F_2, \dots, F_n | C)$  can be simplified as follows:

$$P(F_1, \dots, F_n | C) = P(F_1 | C)P(F_2 | C) \dots P(F_n | C)$$

$$P(F_1, \dots, F_n | C) = \prod_{i=1}^n P(F_i | C) \quad (3)$$

With the above equation, Bayes theorem can be written as follows:

$$P(F_1, \dots, F_n | C) = \frac{1}{P(F_1, F_2, \dots, F_n)} P(C) \prod_{i=1}^n P(F_i | C)$$

$$P(C | F_1, \dots, F_n) = \frac{P(C)}{Z} \prod_{i=1}^n P(F_i | C) \quad (4)$$

The above equation is a model of Naïve Bayes theorem which will then be used in the classification process. The Z presents the constant evidence for all classes on a single sample.

#### 2.4. Data Source

The data used is the historical data of a microcredit bank from 2010-2015 that taken from a database of a public bank in Bandung. The data is kept at the bank and can be obtained when customers apply for credit. This historical data is divided into two: data training and data testing.

### 3. Results and Discussion

#### 3.1. Implementation Calculation Using NBC

The above attributes will determine whether a customer will be given a credit (Eligible/Accepted) or not given the credit (Ineligible/Denied). But before deciding the two classes/categories, two stages are conducted namely the training process and the classification process. In the training phase, the search process of conditional probability value and the search process of prior probability value will be carried out. After the prediction model is built at the training process, the unknown data classification process is then performed.

NBC is applied to several processes that have been described above. The output generated by the system is in the form of a conditional probability value, the prior probability value and the classification accuracy level. While the second is data predicted result namely the credit risk classes (Eligible/Accepted and Ineligible/Denied). The credit risk class used is the largest i.e. Eligible/Accepted obtained if  $P(\text{Eligible} | X) > P(\text{Ineligible} | X)$  where X is a known attribute. Instead Ineligible/Denied is obtained if  $P(\text{Eligible} | X) < P(\text{Ineligible} | X)$ .

When provided a new input, the bank customer data classification can be determined by the following steps:

- a. Counting the number of classes/labels.

$P(Y = \text{Eligible}) = 12/20$ , the number of data “Eligible” on data training is divided by the total number of data

$P(Y = \text{Ineligible}) = 8/20$ , the number of data “Ineligible” on data training is divided by the total number of data

- b. Calculating the number of the same case with the same class.

$P(\text{Age} = 41-55 | Y = \text{Eligible}) = 2/12$

$P(\text{Age} = 41-55 | Y = \text{Ineligible}) = 1/8$

$P(\text{Marital Status} = \text{Married} | Y = \text{Eligible}) = 8/12$

$P(\text{Marital Status} = \text{Married} \mid Y = \text{Ineligible}) = 5/8$   
 $P(\text{The Number of Dependants} = 1\text{-}2 \text{ person} \mid Y = \text{Eligible}) = 3/12$   
 $P(\text{The Number of Dependants} = 1\text{-}2 \text{ person} \mid Y = \text{Ineligible}) = 4/8$   
 $P(\text{Residence} = \text{Other (Renting)} \mid Y = \text{Eligible}) = 1/12$   
 $P(\text{Residence} = \text{Other (Renting)} \mid Y = \text{Ineligible}) = 3/8$   
 $P(\text{The Length of Residence} \Rightarrow 2\text{-}5 \text{ Year} \mid Y = \text{Eligible}) = 2/12$   
 $P(\text{The Length of Residence} \Rightarrow 2\text{-}5 \text{ Year} \mid Y = \text{Ineligible}) = 3/8$   
 $P(\text{The Length of Business} \Rightarrow 5\text{-}10 \text{ Years} \mid Y = \text{Eligible}) = 7/12$   
 $P(\text{The Length of Business} \Rightarrow 5\text{-}10 \text{ years} \mid Y = \text{Ineligible}) = 3/8$   
 $P(\text{The Kind of Business} = \text{Industrial Sector} \mid Y = \text{Eligible}) = 4/12$   
 $P(\text{The Kind of Business} = \text{industrial sector} \mid Y = \text{Ineligible}) = 3/8$   
 $P(\text{The Relationship with the Bank} \Rightarrow 3 \text{ Years} \mid Y = \text{Eligible}) = 5/12$   
 $P(\text{The Relationship with the Bank} \Rightarrow 3 \text{ Years} \mid Y = \text{Ineligible}) = 1/8$   
 $P(\text{The Credit History} \Rightarrow 2 \text{ years} \mid Y = \text{Eligible}) = 5/12$   
 $P(\text{The Credit History} \Rightarrow 2 \text{ years} \mid Y = \text{Ineligible}) = 1/8$   
 $P(\text{Mortgage} = \text{There is Mortgage, CEV} > 100\% \mid Y = \text{Eligible}) = 12/12$   
 $P(\text{Mortgage} = \text{There is Mortgage, CEV} > 100\% \mid Y = \text{Ineligible}) = 4/8$

c. Multiply all variable results Eligible and Ineligible.

$P(\text{Age}/Y) * P(\text{Status Marital}/Y) * P(\text{The Number of Dependants}/Y) * P(\text{Residence}/Y) * P(\text{The Length of Residence}/Y) * P(\text{The Length of Residence}/Y) * P(\text{The Kind of Business}/Y) * P(\text{The Relationship with the Bank}/Y) * P(\text{The Credit History}/Y) * P(\text{Mortgage}/Y)$   
 $= 2/12 * 8/12 * 3/12 * 1/12 * 2/12 * 7/12 * 4/12 * 5/12 * 5/12 * 12/12$   
 $= 0.166 * 0.666 * 0.25 * 0.083 * 0.166 * 0.583 * 0.333 * 0.416 * 0.416 * 1$   
 $= 0.0000128$   
 $P(\text{Age}/N) * P(\text{Status Marital}/N) * P(\text{The Number of Dependants}/N) * P(\text{Residence}/N) * P(\text{The Length of Residence}/N) * P(\text{The Length of Business}/N) * P(\text{The Kind of Business}/N) * P(\text{The Relationship with the Bank}/N) * P(\text{The Credit History}/N) * P(\text{Mortgage}/N)$   
 $= 1/8 * 5/8 * 4/8 * 3/8 * 3/8 * 3/8 * 3/8 * 1/8 * 1/8 * 4/8$   
 $= 0.125 * 0.625 * 0.5 * 0.375 * 0.375 * 0.375 * 0.375 * 0.125 * 0.125 * 0.5$   
 $= 0.0000061$

d. Compare the results of Eligible and Ineligible classes.

From the results above, it appears that the highest probability value is in the  $(P \mid Y)$  class so that it can be concluded that the status of the prospective customer is classified as Eligible.

### 3.2. Test Result

The following test results for credit approval to use applications that have been built (See Figure 1).



Hub. Dengan Bank	Riwayat Kredit	Agunan	Klasifikasi
Tahun	1 - 2 Tahun	Ada Agunan, CEV >= 100...	Tidak Layak
Tahun	1 - 2 Tahun	Ada Agunan, CEV >= 100...	Layak
Tahun	> 2 Tahun	Ada Agunan, CEV >= 100...	Layak
Tahun	> 2 Tahun	Ada Agunan, CEV >= 100...	Layak
Tahun	1 - 2 Tahun	Ada Agunan, CEV >= 100...	Tidak Layak
Tahun	Belum Pernah	Ada Agunan, CEV >= 100...	Layak
Tahun	Belum Pernah	Ada Agunan, CEV >= 100...	Layak
Tahun	Belum Pernah	Ada Agunan, CEV >= 100...	Layak
Tahun	Belum Pernah	Ada Agunan, CEV >= 100...	Layak
Tahun	Belum Pernah	Ada Agunan, CEV >= 100...	Layak
Tahun	> 2 Tahun	Ada Agunan, CEV >= 100...	Layak
Tahun	1 - 2 Tahun	Ada Agunan, CEV >= 100...	Layak
Tahun	Kurang Baik	Ada Agunan, CEV >= 100...	Tidak Layak
Tahun	Kurang Baik	Ada Agunan, CEV >= 100...	Tidak Layak
Tahun	1 - 2 Tahun	Ada Agunan, CEV >= 100...	Layak
Tahun	> 2 Tahun	Ada Agunan, CEV >= 100...	Layak
Tahun	> 2 Tahun	Ada Agunan, CEV >= 100...	Layak
Tahun	> 2 Tahun	Ada Agunan, CEV >= 100...	Layak
Tahun	> 2 Tahun	Ada Agunan, CEV >= 100...	Layak
Tahun	> 2 Tahun	Ada Agunan, CEV < 100%	Layak
Tahun	> 2 Tahun	Ada Agunan, CEV >= 100...	Layak

(a)

Jat. Tinggal	Lama Menetap	Lama Usaha	Jenis Usaha	Hub. Dengan B.	Riwayat Kredit	Agunan	Klasifikasi
Lain (Men...	> 2 - 5 Tahun	> 1 - 5 Tahun	Sektor Jasa	1 - 3 Tahun	1 - 2 Tahun	Ada Agunan, C...	Tidak Layak...
endi...	> 2 - 5 Tahun	> 5 - 10 Tahun	Sektor Industri	1 - 3 Tahun	1 - 2 Tahun	Ada Agunan, C...	Layak...
endi...	> 5 - 8 Tahun	> 5 - 10 Tahun	Sektor Perdag...	1 - 3 Tahun	> 2 Tahun	Ada Agunan, C...	Layak...
endi...	> 5 - 8 Tahun	> 5 - 10 Tahun	Sektor Jasa	1 - 3 Tahun	> 2 Tahun	Ada Agunan, C...	Layak...
di	> 2 - 5 Tahun	> 5 - 10 Tahun	Sektor Pertanian	1 - 3 Tahun	1 - 2 Tahun	Ada Agunan, C...	Tidak Layak...
di	> 2 - 5 Tahun	> 10 Tahun	Sektor Perdag...	> 3 Tahun	Belum Pernah	Ada Agunan, C...	Layak...
di	> 5 - 8 Tahun	> 10 Tahun	Sektor Perdag...	1 - 3 Tahun	Belum Pernah	Ada Agunan, C...	Layak...
di	> 5 - 8 Tahun	> 10 Tahun	Sektor Industri	1 - 3 Tahun	Belum Pernah	Ada Agunan, C...	Layak...
di	> 5 - 8 Tahun	> 10 Tahun	Sektor Industri	1 - 3 Tahun	Belum Pernah	Ada Agunan, C...	Tidak Layak...
Lain (Men...	> 8 Tahun	> 10 Tahun	Sektor Perdag...	> 3 Tahun	> 2 Tahun	Ada Agunan, C...	Layak...
di	> 5 - 8 Tahun	> 5 - 10 Tahun	Sektor Industri	1 - 3 Tahun	1 - 2 Tahun	Ada Agunan, C...	Layak...
endi...	> 2 - 5 Tahun	> 2 - 5 Tahun	Sektor Perdag...	1 - 3 Tahun	Kurang Baik	Ada Agunan, C...	Tidak Layak...

(b)

Hub. Dengan Bank	Riwayat Kredit	Agunan	Klasifikasi
Tahun	1 - 2 Tahun	Ada Agunan, CEV >= 100...	Tidak Layak
Tahun	1 - 2 Tahun	Ada Agunan, CEV >= 100...	Layak
Tahun	> 2 Tahun	Ada Agunan, CEV >= 100...	Layak
Tahun	> 2 Tahun	Ada Agunan, CEV >= 100...	Layak
Tahun	1 - 2 Tahun	Ada Agunan, CEV >= 100...	Tidak Layak
Tahun	Belum Pernah	Ada Agunan, CEV >= 100...	Layak
Tahun	Belum Pernah	Ada Agunan, CEV >= 100...	Layak
Tahun	Belum Pernah	Ada Agunan, CEV >= 100...	Layak
Tahun	Belum Pernah	Ada Agunan, CEV >= 100...	Layak
Tahun	Belum Pernah	Ada Agunan, CEV >= 100...	Layak
Tahun	> 2 Tahun	Ada Agunan, CEV >= 100...	Layak
Tahun	1 - 2 Tahun	Ada Agunan, CEV >= 100...	Layak
Tahun	Kurang Baik	Ada Agunan, CEV >= 100...	Tidak Layak
Tahun	Kurang Baik	Ada Agunan, CEV >= 100...	Tidak Layak
Tahun	1 - 2 Tahun	Ada Agunan, CEV >= 100...	Layak
Tahun	> 2 Tahun	Ada Agunan, CEV >= 100...	Layak
Tahun	> 2 Tahun	Ada Agunan, CEV >= 100...	Layak
Tahun	> 2 Tahun	Ada Agunan, CEV >= 100...	Layak
Tahun	> 2 Tahun	Ada Agunan, CEV < 100%	Layak
Tahun	> 2 Tahun	Ada Agunan, CEV >= 100...	Layak

(c)

Jat. Tinggal	Lama Menetap	Lama Usaha	Jenis Usaha	Hub. Dengan B.	Riwayat Kredit	Agunan	Klasifikasi
Lain (Men...	> 2 - 5 Tahun	> 1 - 5 Tahun	Sektor Jasa	1 - 3 Tahun	1 - 2 Tahun	Ada Agunan, C...	Tidak Layak...
endi...	> 2 - 5 Tahun	> 5 - 10 Tahun	Sektor Industri	1 - 3 Tahun	1 - 2 Tahun	Ada Agunan, C...	Layak...
endi...	> 5 - 8 Tahun	> 5 - 10 Tahun	Sektor Perdag...	1 - 3 Tahun	> 2 Tahun	Ada Agunan, C...	Layak...
endi...	> 5 - 8 Tahun	> 5 - 10 Tahun	Sektor Jasa	1 - 3 Tahun	> 2 Tahun	Ada Agunan, C...	Layak...
di	> 2 - 5 Tahun	> 5 - 10 Tahun	Sektor Pertanian	1 - 3 Tahun	1 - 2 Tahun	Ada Agunan, C...	Tidak Layak...
di	> 2 - 5 Tahun	> 10 Tahun	Sektor Perdag...	> 3 Tahun	Belum Pernah	Ada Agunan, C...	Layak...
di	> 5 - 8 Tahun	> 10 Tahun	Sektor Perdag...	1 - 3 Tahun	Belum Pernah	Ada Agunan, C...	Layak...
di	> 5 - 8 Tahun	> 10 Tahun	Sektor Industri	1 - 3 Tahun	Belum Pernah	Ada Agunan, C...	Layak...
di	> 5 - 8 Tahun	> 10 Tahun	Sektor Industri	1 - 3 Tahun	Belum Pernah	Ada Agunan, C...	Tidak Layak...
Lain (Men...	> 8 Tahun	> 10 Tahun	Sektor Perdag...	> 3 Tahun	> 2 Tahun	Ada Agunan, C...	Layak...
di	> 5 - 8 Tahun	> 5 - 10 Tahun	Sektor Industri	1 - 3 Tahun	1 - 2 Tahun	Ada Agunan, C...	Layak...
endi...	> 2 - 5 Tahun	> 2 - 5 Tahun	Sektor Perdag...	1 - 3 Tahun	Kurang Baik	Ada Agunan, C...	Tidak Layak...

(d)

**Figure 1.** (a) The Result of Data Testing 50 Data (b) The Classification Result on Historical Data of Bank Credit Customers (50 Data) (c) The Result of Data Testing 500 Data (d) The Classification Result on Historical Data of Bank Credit Customers (500 Data).

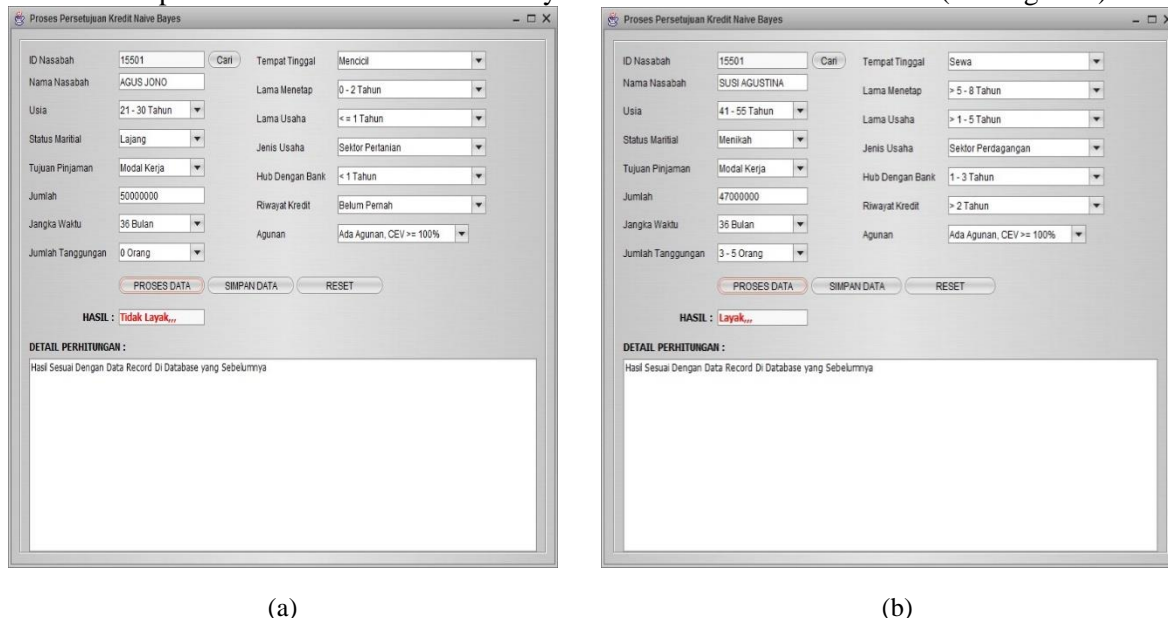
From figure 1, it can be observed that there is a compatibility of the predicted outcome that has been processed using data mining application, figure 1(a) with the factual data (historical data), figure 1(b). On the data testing, there are 38 out of 50 compatible data and 12 data are incompatible with factual data. The percentage of success is:

$$\text{success rate} = \frac{38}{50} \times 100\% = 76\%$$

Based on figure 1(c) and 1(d), there are 475 out of 500 compatible data and 25 data are incompatible. The percentage of success is:

$$\text{success rate} = \frac{475}{500} \times 100\% = 95\%$$

It can be seen that this application is well performing in analyzing the feasibility of customers data who are eligible for loans, however, this application will work optimally when the data provided is relatively larger and more variable. This application can assist credit analyst in recommending customers who deserve personal loan which is decided by the decision maker in that bank (See Figure 2).



**Figure 2.** (a) The Testing Process of 1 Data with “Ineligible” (*Tidak Layak*) Result (b) The Testing Process of 1 Data with “Eligible” (*Layak*) Result.

In figure 2(a), the result obtained is Ineligible (*Tidak Layak*) because the multiplication result of the Eligible class is smaller than multiplication result of the Ineligible class. In figure 2(b), the result obtained is Eligible (*Layak*) because the multiplication result of the Eligible class is bigger than that of Ineligible class. The details of the calculations are described in the previous section.

#### 4. Conclusions

After doing the analysis, design, and testing, it can be concluded as follows:

- Classification technique with NBC can be used to predict credit risk classes quickly.
- If the data training to Nil, Naïve Bayes NBC cannot classify the record, so that the data predicted will be Nil or Ineligible.
- From the results of the testing system, the success rate of data testing as many as 500 data is 95%. Therefore, the applications using an NBC is considered helpful in the process of deciding giving a loan to new customers.
- This study, besides the application, can build the classification of the bank customer's data (who is eligible or ineligible), in the other hand by observing the result of eligibility classification, the bank can also determine which area will be the focus in the next marketing target, so it can minimize the operational cost.

#### References

- [1] Widiyanti N M D Suwendra I W 2014 Pengaruh Penilaian Kredit Terhadap Keputusan Pemberian Kredit pada BPR (The Effects of the Credit Assesment of the Providing Credit on BPR) *E-Journal of Bisma University* **2**(1), pp.40 – 60.
- [2] Kasmir 2006 *Manajemen Perbankan (Th. Management of Banking)* (Jakarta: Raga Grafindo Persada).
- [3] Hasibuan M S P 2009 *Dasar-Dasar Perbankan (The Basic of Banking)* (Jakarta: PT. Bumi

- Aksara).
- [4] Bhambri V 2011 Application of Data Mining in Banking Sector *International Journal of Computer Science and Technology* **2(2)**, pp.199-202.
  - [5] Moin K I and Ahmed D Q B 201. Use of Data Mining in Banking *International Journal of Engineering Research and Applications* **2(2)**,pp.738-742.
  - [6] Han J and Kamber M 2001 *Data Mining: Concepts and Techniques* (USA: Morgan Kaufmann Publishers).
  - [7] Tan P N, Steinbach M and Kamar V 2006 *Introduction to Data Mining* (Boston: Pearson Addison Wesley).
  - [8] Ting S L, Ip W H and Tsang A H 2011 Is Naive Bayes a good classifier for document classification *International Journal of Software Engineering and Its Applications* **5(3)**, pp.37-46.
  - [9] Patil T R and Sherekar SS 2013 Performance analysis of Naive Bayes and J48 Classification Algorithm for Data Classification *International Journal of Computer Science and Applications* **6(2)**, pp.256-261.
  - [10] Ginting S L B and Trinanda R P 2013 Teknik Data Mining Menggunakan Metode Bayes Classifier Untuk Optimalisasi Pencarian Pada Aplikasi Perpustakaan (Data Mining Technique Using Bayes Classifier Method for Search Optimization On Library Application) *Journal of Technology and Information* **1(6)**,pp.37-50.
  - [11] Bustami B 2013 Penerapan Algoritma Naïve Bayes Untuk Mengklasifikasi Data Nasabah Asuransi (Application of Naïve Bayes Algorithm to Classify Insurance Customers Data) *TECHSI-Jurnal Teknik Informatika* **5(2)**, pp. 201 – 220.
  - [12] Zhang H. 2004 The optimality of naive Bayes *AA Journal* **1(2)** pp.3 – 20.
  - [13] Murphy K P 2006 Naive Bayes Classifiers *University of British Columbia* **18**, pp.60 – 70.