RESEARCH ARTICLE   OPEN ACCESS

Decision Support System as A Method Of Determining Feasibility Of Debitural Candidates For Motorcycle Credit Using Naive Bayes Classifier Methods (Study Case: FIF Group Cabang Ciledug)

Noviana Pratiwi[1], Muhammad Haikal Al Wahaby[2], Dr. Krisna Adiyarta M., S.Kom, M.Sc[3]

[1, 2, 3](Department Of Computer Science, Budi Luhur University, Indonesia)

--------------------------------------************************--------------------------------

Abstract:

PT. Federal International Finance (FIF Group) is a group of companies engaged in consumer financing. FIF Group serves financing primarily for retail financing and electronic products. Considering that there are a large number of participants in applying for credit every day, as well as various types of debtor candidates, it requires a system that can handle the issue of evaluating eligibility in applying for credit. One of the factors that causes long-term decision-making activities is in the presentation of data obtained from credit administration activities presented in qualitative form. In this study contains a classification system of risk classes in prospective candidates in submitting motorbike credit using the Naïve Bayes Classifier method. The method is a method of approach to probability in producing a classification for determining a possible grade in an object. From the results of the research that has been done, it is hoped that it can help the company work on the part of credit analysts in selecting prospective borrowers.

*Keywords — **Credit, Risk Class, Classification, Naïve Baiyes Classifier, Decision Support System.***

--------------------------------------************************--------------------------------

I.   INTRODUCTION

At present information technology is growing so rapidly. With this in mind, almost all fields need an ease to handle the information they have had. One of these fields is the field of crediting. Credit terminology comes from the Latin word "Credere" which means trust. On that basis, based on giving credit to a person or business entity is based on a trust. The definition of credit according to Suyatno (2003) quoted from Mardison (2012) is: "Loans or loans that are provided are the provision of money or bills that can be likened to it based on the agreement of the borrowing party between the bank and another party in the event that the borrower is obliged repay the debt after a certain period of time with the amount of interest that has been

determined ". In this study, the data used is motorbike credit data at the Ciledug branch of the International Federal Finance Company (FIF). The data is data that contains debtor data along with the criteria used in determining the feasibility of prospective debtors in the company. The data consists of 4 attributes, namely: A1: Age, A2: Marital Status, A3: Income, A4: Consumptive Debt. A5: length of service, A6: sex, A7: marital status and, A8: other debtor status (guarantor), A9: long stay, A10: property, A11: installment plan, A12: place of residence, A13: work , A14: number of dependents, and A15: telephone. The use of the 15 attributes that do not necessarily produce a classification of credit risk classes with high accuracy when compared to using only some attributes. Therefore, we will look for a combination model of attributes that results in the classification of

risk classes with the highest accuracy. The search process is carried out by experimenting as many combinations of these 15 attributes. Evaluation results from the trial for each combination that has the highest accuracy value will be used as a model for the classification of motor credit risk classes.

II. RELATED WORKS

A.A.1 Credit

Credit comes from the Latin word "credo" which means "I believe", which is a combination of sanskrit "cred" which means "trust" and Latin "do" which means "I place". Getting credit means gaining a trust. On the basis of that trust to someone who needs it, money, goods or services are provided on condition that they repay or give reimbursement within a period of time that has been made in an agreement.

A.A.2 Classification Concept

Classification is a job that assesses data objects to include them in a particular class of available classes. In the classification there are two main jobs done, namely the construction of a model as a prototype to be stored as memory and the use of a model to perform an introduction or classification or prediction on another data object so that it can be known in which class the data object is stored in a model.

A.A.3 The Naïve bayes Classifier Method

The Naïve Bayes Classifier (NBC) is a classification method rooted in a theory often called the Bayes theorem. The main characteristic of the Naïve Bayes Classifier Method is a very strong (naive) assumption that it will be independent of each variable. In other words, the Naïve Bayes Classifier Method assumes that the existence of an

attribute or variable has nothing to do with the existence of an attribute or other variable. The Naïve Bayes Classifier algorithm consists of two stages. That is, in the first stage is the training phase of the set of sample documents or training data and the second stage is the process of classifying documents that have not yet known categories, namely classes. This algorithm utilizes probability theory proposed by a British scientist named Thomas Bayes, which predicts a probability in the future based on previous experience. Because the assumption of these attributes is not conditionally independent, then:

$$Vmap = \frac{argmax}{V_j \in V} \; P(Vj) \prod P(w_k \mid Vj)$$

After obtaining calculations for each category, the selected category is the category that has the largest Vmap value. The P value (Vj) will be determined during the training test, the value of which is based on the equation below:

$$P(V_j) = \frac{|docs_j|}{|contoh|}$$

For P value (W1 | V1) determined by the

$$P(w_k \mid V_j) = \frac{|n_k + 1|}{n + |kosakata|}$$

equation:

A.A.4 Distance

Distance is used to determine the similarity degree or disimilarity degree of two feature vectors. The level of similarity is in the form of a score and based on the score two feature vectors will be said to be similar or not. The distance coefficient that can be used to measure the similarity level of two feature vectors are as follows:

- *Euclidean distance*

Euclidean distance is most commonly used to calculate the similarity of 2 (two) vectors. Euclidean distance calculates the root of the square of the difference 2 (two) vectors (root of square differences between 2 vectors).

$$\sqrt{\sum_{k=1}^{n}(x_{jk} - x_{jl})^2}$$

The smaller the value (k, l) the more similar the two matching vectors are. Conversely, the greater the value (k, l), the more different the two vectors are compared.

- Mean *Euclidean distance*

Mean Euclidean Distance is equal to Euclidean distance, but the final distance on a scale is smaller because the mean difference is used

$$\sqrt{\frac{\sum |x_{jk} - x_{jl}|^2}{n}}$$

B. Literature Review

The literature review used was obtained from several sources used by the author as the basis of this study, including research conducted by Fitri Handayani, Feddy Setio Pribadi (2015) entitled "Implementation of the Naïve Bayes Classifier Algorithm in Automatic Text Classification Community Complaints and Reporting through Call Center 110 services "The results of research conducted using the Niave Bayes Classifier method resulted in a high average accuracy, namely 90% recall and 92% f-measure.

Research conducted by Claudia Clarentina Ciptohartono (2014) with the title "Algoritma Klasifikasi Metode Naïve Bayes Classifier untuk menilai kelayakan kredit".This study produces evidence that the Naïve Bayes algorithm can be applied in providing credit worthiness to BCA Finance Jakarta. In this credit feasibility study, the initial data by pre-processing has an accuracy of 85.57%, but if the initial data is processed and by pre-processing it can produce an accuracy of 92.53%.

Research conducted by Hera Wasiati and Dwi Wijayanti (2014) with the title "Sistem Pendukung Keputusan Dalam Penentuan Kelayakan Calon Tenaga Kerja Indonesia Menggunakan Metode Naïve Bayes Classifier". The research was carried out with 542 data with 362 details included as training data and 180 as test data, resulting in an accuracy of 73.89% and an error of 26.11% so that the exact amount of data was 133 and incorrect was 47.

The research conducted by Yuda Septian Nugroho (2014) entitled "Data Mining Menggunakan Algoritma Naïve Bayes Classifier Untuk Klasifikasi Kelulusan Mahasiswa Universitas Dian Nuswantoro". The results of this study in the form of accuracy of 82.08%. data that lacks complexity causes the model to predict quite accurately.

The research conducted by Arief Junanto (2013) with the title "Algoritma Naïve Bayes Classifier untuk menacari perkiraan waktu studi mahasiswa" This study aims to test with random training data and testing data which results in prediction errors in finding an estimated study time of 20% up to 50% of the number of data records and consistency of training and testing data used can cause high and low levels of error / accuracy. The results of the prediction of the accuracy of students in 2008 using the Naïve Bayes algorithm are predicted to be 254 students "on time" and predicted 4 students "not on time".

Based on a summary review of studies on previous research states that the application of the Naïve Bayes Classifier Method is very

helpful in data processing to find the value needed. It can be shown in terms of time, the accuracy produced using the Naïve Bayes Classifier method is very good and can be said to be in accordance with the expectations of the user, therefore the authors conduct research using the Naïve Bayes Classifier method to make a decision.

C. Conceptual Framework

This study consists of references such as literature studies and observation as a method of data collection - data needed. The problem in this study is that the higher the level of consumer demand for credit, the higher the risk of the company in dealing with consumers who are unable to pay motorcycle loan installments. The purpose of this study is to assist companies in making a decision in choosing the feasibility of prospective borrowers who can receive motorcycle credit.

D. Hypothesis

The hypothesis of this study is that income is estimated to have a significant effect on the application of motorbike debtors, it is suspected that the age category influences the submission of motorbike credit, and it is suspected that the company has difficulty in determining the submission criteria for prospective motor loans. To prove the hypothesis, the writer conducts a trial with the UAT (User Acceptance Test) method as an application test distributed to application correspondents.

III. RESEARCH METHODOLOGY AND DESIGN

The approach in this study is to experiment by testing methods through a prototype system whose effectiveness will be tested using the UAT (User Acceptance Test) and the Naïve Bayes Classifier Method.

A. Method Of Collecting Data

In this study data collection was carried out, from the results of observations made at PT. The Federal International Finance (FIF Group) Ciledug branch obtained the data needed to make a Decision Support System for Debtor Candidates for submission of motor loans. The data obtained is in the form of data of prospective borrowers in the submission of motorbike credit in 2017. From the observation, a system will be made that can categorize data that is appropriate to its class, then a website basis is made so that relevant stakeholders can be accessed directly.

B. Research Object

The intended object of the study is the prospective debtor who will be classified into Eligible or Unworthy credit risk classes, by carrying out the process of data classification at risk of Decent and Unworthy credit by recognizing previously existing data patterns.

C. Research Materials

In this study, the data used is motorbike credit data at the company PT Federal International Finance (FIF) in the Ciledug branch. The data contains prospective debtor data along with the criteria used in determining the debtor's eligibility. The data consists of 4 attributes, namely A1 : Umur, A2 : Status Perkawinan, A3 : Penghasilan, A4 : Hutang Konsumtif.

- a. Umur
  1. Muda (<35 Tahun)
  2. Paruh Baya (35 – 50 Tahun)
  3. Tua (>50 Tahun)

- b. Status Perkawinan
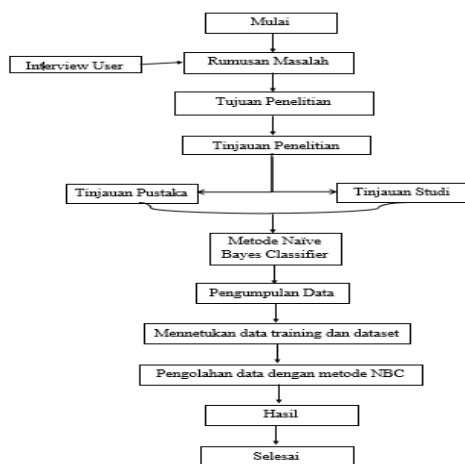 1. Lajang
  2. Menikah

- c. Penghasilan
  1. Rendah (<3 Juta)
  2. Sedang (3 – 10 Juta)

3. Tinggi (>10 Juta)

d. Hutang Konsumtif
   It is qualitative data that shows other debt / other loans.

D. Research Steps

The series of steps that are carried out in a planned and systematic manner to get problem solving is starting from identifying the problem. This stage is the stage to look for problems faced in the topic to be taken.

In an effort to achieve the research objectives that have been established, a research methodology is prepared, while the steps that are prepared are as follows:

1. Explanation of each research step
2. Start the research by looking for problems obtained from interviewing or interviewing users.
3. Make a problem formulation from the results of the interview
4. Determine research objectives
5. Conduct a research review, a research review was obtained from a literature review and study review.
6. Then choose a suitable research method, the authors chose the Naïve Bayes Classifie research method,
7. After that the writer collects data, the data used as research is the company's data in

2017, where the data taken is data from prospective debtor companies in 2017.
8. Choosing training data and testing data
9. After that proceed to processing data, where the first is to do the algorithm processing manually using Microsoft Excel software and then processing the algorithm automatically with the help of WEKA application software.
10. The results of data processing that have been done will make a decision.
11. Finally, do an analysis of the final results obtained from two different calculations, namely manually and automatically, so that a conclusion and appropriate recommendations can be drawn.

IV. RESULTS AND DISCUSSION

In this chapter, the dataset is tested to determine the level of accuracy (truth) of classifying data using the Naïve Bayes Classifier method. Tests are carried out to determine the effect of using parameters.

The first stage of calculating the feasibility search with the Naïve Bayes Classifier method is to look for the probability of each class. For the feasibility of prospective borrowers in a motorbike credit, two classes will be determined, namely class "layak" and "Tidak Layak".

The method of calculation is to find the amount of data with the Feasibility, Decent and Inadequate Risk class of the total Training data or Dataset, then divide it by the total data. The results of these calculations can be seen in the following table:Tabel kelayakan.

TABLE I
TABEL KELAYAKAN

Feasibility Risk Class			
Layak		Tidak Layak	
Layak	10/19	Tidak Layak	9/19

Of the two classes, there are a total of 19 records, of which there are 10 data in the class

category with the risk of "Layak" feasibility, and from 19 records there are 9 data with the risk category of "Tidak Layak" eligibility.

How to find the probability of an attribute is to compare the attributes of data testing with attributes from training data. What is the number of attributes with the risk of "Layak" eligibility in the training data, then divide by the probability of "Layak" class. Likewise, looking for probabilities for classes not.

1. Attributes Umur

Table Umur

Himpunan	Probabilitas Kategori Umur			
	Layak	Tidak Layak	Layak	Tidak Layak
Muda (<35 Tahun)	2	3	2/5	3/5
Paruh Baya (35-50 Tahun)	5	2	5/7	2/7
Tua (>50 Tahun)	3	4	3/7	4/7

Of the 2 classes and with these 3 attributes, there are a total of 19 records, of which there are 2 data with "Layak" class categories and 3 data with "Tidak Layak" categories in Young Characters (<35 Years), from the 7 data there are 5 data with "Layak" class category and 2 data with "Tidak Layak" category on Middle-aged Character (35-50 Years), and from those 7 records there are 3 data with "Layak" and 4 data class categories with category "No Eligible "for Old Character (> 50 Years)

2. Attributes Status

Table Status

Himpunan	Probabilitas Kategori Status			
	Layak	Tidak Layak	Layak	Tidak Layak
Lajang	3	6	3/9	6/9
Menikah	7	3	7/10	3/10

Of the 2 classes and with these two attributes, there are a total of 19 records, of which there are 3 data with "Layak" class categories and there are 6 data with "Tidak Layak" class categories on Single Characters, and from 10 data there are 7 data with "Layak" class category and there are 3 data with "Tidak Layak" class category on Married Character.

3. Attributes Penghasilan

Himpunan	Probability of Income Category			
	Layak	Tidak Layak	Layak	Tidak Layak
Rendah (<3 juta)	3	4	3/7	4/7
Sedang (3-10 juta)	4	3	4/7	3/7
Tinggi (>10 juta)	3	2	3/5	2/5

Table IV.5 Table Atribut Penghasilan

Himpunan	Probability of Income Category			
	Layak	Tidak Layak	Layak	Tidak Layak
Rendah (<3 juta)	3	4	3/7	4/7
Sedang (3-10 juta)	4	3	4/7	3/7
Tinggi (>10 juta)	3	2	3/5	2/5

Of the 2 classes and with these 3 attributes there are a total of 19 records, of which there are 3 data with "Layak" and 4 data class categories with "Tidak Layak" category in Low Character (<3 million), of 7 data there are 4 data with "Eligible" class category and 3 data with "Tidak Layak" class category on Medium Character (3-10 million), and from 5 data there are 3 data with "Layak" class category and 2 data in "Tidak Layak" category "In High Characters (> 10 million.

4. Attributes Hutang Konsumtif

Table IV.6 Table Attributes Hutang Konsumtif

Of the 2 classes and with 3 attributes, there are a total of 19 records, of which there are 5 data with "Layak" class categories and there are 8 data with "Tidak Layak" category in the attributes Yes, of the 6 data there are 5 data with "Layak" class category and 1 data with "Tidak Layak" category.

Manual Calculation
Data Testing : X = (Umur="Paruh Baya (30-50)", Status="Menikah", Penghasilan= "Sedang(3-10 Juta)", Hutang Konsumtif="Ya")

$P(Ci)$
$P(Layak) = 10/19 = 0.52631579$
$P(Tidak\ Layak) = 9/19 = 0.47368421$

$P(X|Ci)$
$P(Umur="Paruh\ Baya\ (30-50\ Tahun)"\ |\ Layak) = 0.71428571$
$P(Umur="Paruh\ Baya\ (30-50\ Tahun)"\ |Tidak\ Layak) = 0.28571429$
$P(Status="Menikah"\ |\ Layak) = 0.7$
$P(Status="Menikah"\ |\ Tidak\ Layak) = 0.3$
$P(Penghasilan="Sedang\ (3-10\ Juta)"\ |\ Layak) = 0.57142857$
$P(Penghasilan="Sedang\ (3-10\ Juta)"\ |Tidak\ Layak) = 0.42857143$
$P(Hutang\ konsumtif="Ya"\ |\ Layak) = 0.38461538$
$P(Hutang\ Konsumtif="Ya"\ |Tidak\ Layak) = 0.61538462$

$P(X|Ci)$
$P(X|Layak) = 0.71428571* 0.7* 0.57142857* 0.38461538= 0.10989011$
$P(X|Tidak\ Layak) = 0.47368421* 0.3* 0.42857143* 0.61538462$
$= 0.02260597$

$P(X|Ci)*P(Ci)$
$P(X|Layak)*P(Layak) = 0.52631579* 0.10989011 = 0.0578369$

Himpunan	Probabilitas Kategori Penghasilan			
	Layak	Tidak Layak	Layak	Tidak Layak
Ya	5	8	5/13	8/13
Tidak	5	1	5/6	1/6

$P(X|\ Tidak\ Layak)*P(Tidak\ Layak)= 0.47368421* 0.02260597$
$= 0.01070809$

With manual calculation, obtained on the

Known (Data Tes)				Searching
Umur	Status	Penghasilan	Hutang Konsumtif	Resiko Kelayakan
Paruh Baya (30-50 Tahun)	Menikah	Sedang (3-10 Juta)	Ya	??

Feasibility Risk attribute with Eligible category has a value of 0.0578369 and with the Feasibility Risk attribute with the Unworthy category has a value of 0.01070809.

The result, people with age: middle age (35-50 years), status: married, income: moderate (3-10 million), consumptive debt: yes, most likely the feasibility risk: worth the biggest value = 0.0578369

V. CONCLUSION

Based on the results of research conducted on the determination of the feasibility of applying for motorbike credit with the Naïve Bayes method at PT Federal International Finance (FIF Group) Ciledug Branch, then the following conclusions can be drawn:

1. The Naïve Bayes method used provides a fast selection process with an easy-to-understand algorithm

2. By using the decision support system application in determining the feasibility of submitting this motorbike credit, the company can easily and quickly make decisions for prospective debtors in applying for motorbike credit.

3. Decision Support System in determining the feasibility risk of applying for a loan with the Naïve Bayes method can be used to assist the company in determining the prospective feasibility of the borrower in submitting motorbike loans to PT Federal International Finance (FIF Group) Ciledug Branch.

REFERENCES

1. Undang – Undang Republik Indonesia Nomor 10 Tahun 1998 tentang Perbankan.
2. Abdullah, Sigit 2013. Penerapan Algoritma Decision Tree C4.5 Untuk Diagnosa Penyakit Stroke Dengan Klasifikasi Data Mining Pada Rumah Sakit Santa Maria Pemalang. Semarang.
3. Bahasa, T.P.K.P.P. dan P., 1989. Kamus Besar Bahasa Indonesia, Jakarta : Balai Pustaka.
4. Desrizal, 2012. Panduan Lengkap PHP Ajax jQuery.Yogyakarta:codingwear.
5. Eko Putri, Riyan dkk 2012. Perbandingan Metode Klasifikasi Naïve Bayes Dan K-Nearest Neighbor Pada Analisis Data Status Kerja Di Kabupaten Demak Tahun 2012. Universitas Diponegoro
6. F. Gorunescu 2011. Data Mining Concept Model Technique. Craiova, Romania: Springer.
7. Fitri Nuraeni, Rahadi Deli Saputra, and Neneng Sri Uryani, 2013.Algoritma C4.5 Untuk Klasifikasi Pola Pembayaran Kredit Motor Pada Perusahaan Pembiayaan (Leasing). Seminar Nasional Informatika.
8. Guritno, S., 2011. Theory and Application of IT Research : metodologi penelitian teknologi informasi Ed.I. Yogyakarta : ANDI.
9. Haris Saputro. 2012. Modul Pembelajaran Prakter Basis Data (MySQL)
10. Hiemstra, D. 2001. Using language models for information retrieval. Disertasi, Ph.D, Dissertation. Center for Telematics and Information Technology. The Netherlands.
11. Ian H. Witten, Eibe Frank, Mark A. Hall 2011 Data Mining: Practical Machine Learning Tools and Techniques (Third Edition), Morgan Kaufmann.
12. Jiawei Han, Micheline Kamber 2006. Data Mining: consept and techniques. San Francisco: Elsevier Inc.
13. M. J. Islam, Q. M. J. Wu, M. Ahmadi and M. A. Sid-Ahmed (2007). Investigating the Performance of Naive- Bayes Classifiers and K- Nearest Neighbor Classifiers. International Conference on Convergence Information Technoloy,
14. M. K. Mukhlis. 2011. "Diagnosa Kemungkinan Pasien Terkena Stroke Dengan Menggunakan Metode Naïve Bayes Dan Metode Jaringan Syaraf Tiruan Berbasis

Web". Surabaya: Institut Negeri Sepuluh Nopember

15. Nugroho, 2013. Kupas Tuntas Adobe Dreamweaver CSS Dengan Pemrograman PHP & MySQL. Yogyakarta, Andi

16. Porter, Michael E, 1993. Competitive Strategy, Techniques for Analysing Industries and Competitord, The Free Press New York.

17. R.H. Sianipar, 2015. HTML5 & CSS3. Penerbit Informatika

18. Suprapto 2008. Bahasa Pemrograman Untuk SMK Kejuruan. Direktorat Jenderal Pendidikan Dasar dan Menengah Departemen Pendidikan Nasional

19. Suyanto, Asep Herman. 2007. Step by Step: Web Design Theory and Practices. Yogyakarta: Andi Ofset. 2012. Definisi dan Pengertian Web Menurut Para Ahli (online). Sumber http://www.sambureki.com/definisi/definisi-dan-pengertian-web-menururt-para-ahli.html.

20. Turban, E. dkk, 2005. Decicion Support Systems and Intelligent Systems.: Andi Offset.

21. Universitas Gunadarma, 2006. Pertemuan ke-6 Java Script.Pemrograman web