International Journal of Science and Research (IJSR)
ISSN: 2319-7064
ResearchGate Impact Factor (2018): 0.28 | SJIF (2018): 7.426

Predicting Customer Loyalty by Using Naïve Bayes

Teguh Iman Hermanto[1], Muhamad Agus Sunandar[2]

[1, 2]STT Wastukancana, Informatics Engineering, Jl. Cikopak No. 53, Purwakarta, Jawa Barat, Indonesia

Abstract: ***The effort to increase customer loyalty is by predicting whether the customer is loyal or not. To carry out company development, a customer development strategy is needed, while to carry out customer development strategies can be done using a data mining approach including classification of customers with the method of naive bayes. This study aims to predict consumer loyalty to telecommunication service products. the method used is CRIPS-DM through the stages of business understanding, understanding data, data preparation, modeling, evaluation and development. The algorithm used for prediction is the Naive Bayes algorithm. The results obtained from this study are customer classification models that can help in customer loyalty.***

Keywords: Naïve Bayes, CRISP-DM, Customer Loyalty, Prediction

1. Introduction

The development of information technology has a positive impact on human life in various fields. Information technology can be used to help solve problems or optimize the activities of individuals or organizations. One area affected is the telecommunication sector. Information technology can be used to determine churn and loyalty customers by paying attention to good customer principles.
Related to many in the communication service industry, business people need to consider strategies to achieve desired profits and win the competition. So that these goals can be obtained from a company must produce an item or service that is needed and desired by consumers. Churn predictions can be used to identify churners earlier before they move, and can help the department to maintain them, so that potential corporate losses can be avoided churn predictions.

By using data mining techniques, existing data can be processed into useful information. The method used in this data mining technique is classification. By using the Naive Bayes algorithm stored data can be managed and produce more useful information. In measuring or classifying customer loyalty to Indosat Ooredoo customers, data mining techniques are used with the Naive Bayes algorithm.

2. Literature Review

a) Data Mining

Data mining is the process of discovering meaningful new correlations, patterns and trens by sifting through large amounts of data stored in repositories,using pattern recognition technologies as well as statistical and mathematical technique.[1]

b) Classification

Classification is a job of assessing data objects to enter them into certain classes. In classification, there are two main jobs carried out, namely the construction and the use of the model. The construction of the model will produce a prototype which then used for carry out recognition / classification / prediction on another data so that it is known in which class the data object is stored in the model.

c) Naïve Bayes Algorithm

Bayesian classification is the classification of statistics that can be used to predict the probability of membership of a class. Bayesian classification is based on the Bayes theorem which has a classification capability similar to the decision tree and neural network. Bayesian Classification has proven to have high accuracy and speed when applied to databases with large data.

The Bayes theorem has the following general form:

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)}$$

X = data with unknown classes
H = the data hypothesis X is a specific class
P(H|X) = probability of hypothesis H based on condition X (*posteriori probability*)
P(H) = probability of hypothesis H (*prior probabilty*)
P(X|H) = probability X is based on conditions in hypothesis
P(X) = probability X

3. Research Method

The method that used in this study is Cross-Industry Standard Process for Data Mining (CRISP-DM). The CRISP-DM was developed in 1996 by analysts representing Daimler Chrysler, SPSS, and NCR. The iterative nature of CRISP-DM is simbolized by the outer circle in Figure 1 bellow :

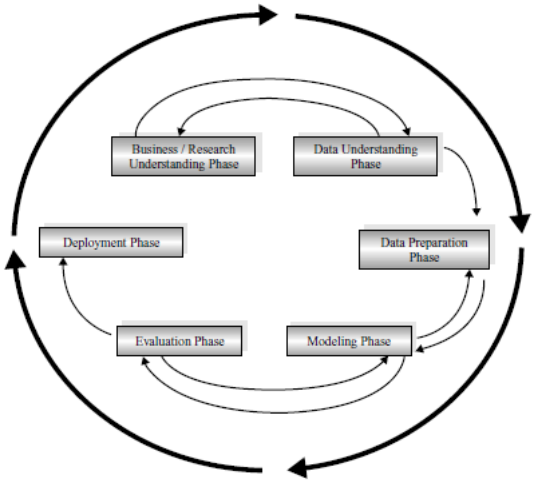

Figure 1: CRISP-DM Phase

The following is an outline of each phase in the CRISP-DM:
1) *Business understanding phase:* The first phase in the CRISP–DM standard process may also be termed the research understanding phase.

a) Enunciate the project objectives and requirements clearly in terms of the business or research unit as a whole.
b) Translate these goals and restrictions into the formulation of a data mining problem definition.
c) Prepare a preliminary strategy for achieving these objectives.

2) *Data understanding phase*
a) Collect the data.
b) Use exploratory data analysis to familiarize yourself with the data and discover initial insights.
c) Evaluate the quality of the data.
d) If desired, select interesting subsets that may contain actionable patterns.

3) *Data preparation phase*
a) Prepare from the initial raw data the final data set that is to be used for all subsequent phases. This phase is very labor intensive.
b) Select the cases and variables you want to analyze and that are appropriate for your analysis.
c) Perform transformations on certain variables, if needed.
d) Clean the raw data so that it is ready for the modeling tools.

4) *Modeling phase*
a) Select and apply appropriate modeling techniques.
b) Calibrate model settings to optimize results.
c) Remember that often, several different techniques may be used for the same data mining problem.
d) If necessary, loop back to the data preparation phase to bring the form of the data into line with the specific requirements of a particular data mining technique.

5) *Evaluation phase*
a) Evaluate the one or more models delivered in the modeling phase for quality and effectiveness before deploying them for use in the field.
b) Determine whether the model in fact achieves the objectives set for it in the first phase.
c) Establish whether some important facet of the business or research problem has not been accounted for sufficiently.
d) Come to a decision regarding use of the data mining results.

6) *Deployment phase*
a) Make use of the models created: Model creation does not signify the completion of a project.
b) Example of a simple deployment: Generate a report.
c) Example of a more complex deployment: Implement a parallel data mining process in another department.
d) *For businesses, the customer often carries out the deployment based on your model.*

4. Research Result

In this reasearch the CRISP-DM phase was carried out only until the evaluation phase. The following are the results of each phase, the results of data mining modeling and their evaluations.

a) Business Understanding
In developing a business and offsetting an increase in consumers, PT. Indosat, Tbk requires a smooth cash flow turnover. Therefore, analysis needs to predict the quality of customer loyalty. Prospective customers who are predicted to have smooth transaction quality will be prioritized to get loyal categories. Based on this, the purpose of data mining modeling is to classify customer data whose results can be used to predict the quality of customer loyalty. In addition, the results of data mining modeling can also be used to find out what factors influence the quality of prospective customers.

Table 1: Attributes of Customer Data

Attribute	Explanation
MSISDN	MSISDN
Gender	Gender
Data Package/Day	Data Package/Day
SMS/Day	SMS/Day
Call/Day	Call/Day
Cost/Day	Cost/Day
Churn/Loyal	Churn/Loyal

1) Data Preparation
Data that has been collected needs to be prepared for the data mining process. Following are the steps that need to be done to prepare data so that it is ready to be processed for data mining modeling:

a) Data Selection
At the stage, data selection attributes will be selected. The selected attributes will be used in data mining process. The attributes to be used are MSISDN, Gender, Data Package/Day, SMS/Day, Call/Day, Cost/Day, Churn/Loyal.

b) Data Cleaning
At this stage, data which have same value in that attribute will be cleaned or deleted. There are 27 data records that have the same value so that data will be deleted. The amount of data used for data mining modeling is 495 records.

c) Data Transformation
At this stage attributes that have a numeric type will be converted into category types to make data mining process easier. The attribute will be grouped into some catagories and attributes were converted are age and loan. The results of grouping attributes can be seen in Table III and Table IV.

Table 2: Grouping of Internet Data Usage

Internet Data	Catagory
100-1000	Low
1100-1500	Medium
1550-2500	high

Table 3: Grouping of SMS Data

SMS	Catagory
5-10	Low
11-20	Medium
21-30	High

Table 4: Grouping of Call Data

Call	Catagory
5-10	Low
11-20	Medium
21-30	High

Table 5: Grouping of Pulse Data

Pulse	Catagory
1000-2000	Low
2100-3000	Medium
3100-5000	High

2) Modeling

Data mining modeling uses Naïve Bayes algorithm. The software used for modeling implementation is Rapidminer. Implementasion of data mining using Rapidminer can be seen on Figure 2. The result of data mining modeling can be seen on Figure 3 and the description of Churn prediction can be seen on Figure 4 and loyalty Prediction can be seen on

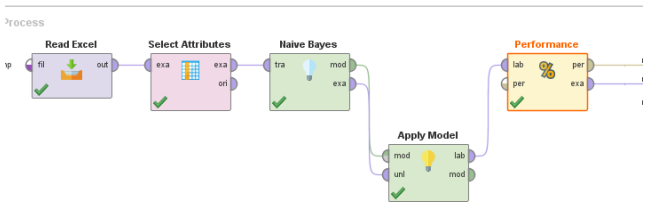

Figure 2: Implementation of Naïve Bayes Algorithm

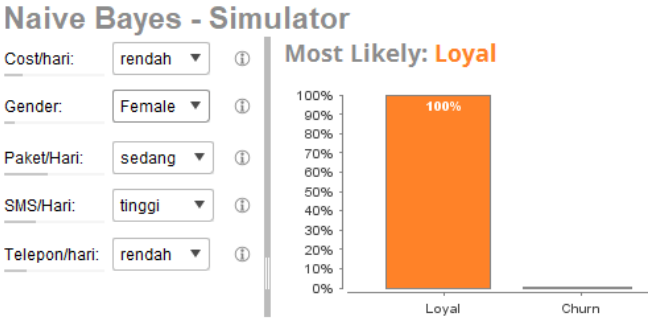

Figure 3: Naïve Bayes as Result of Prediction

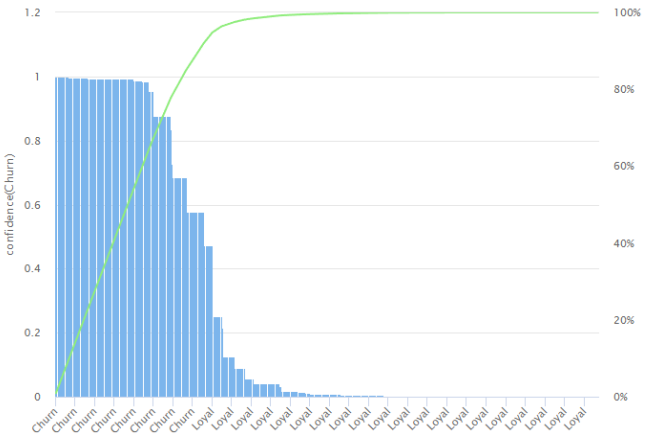

Figure 4: Description of Churn Prediction

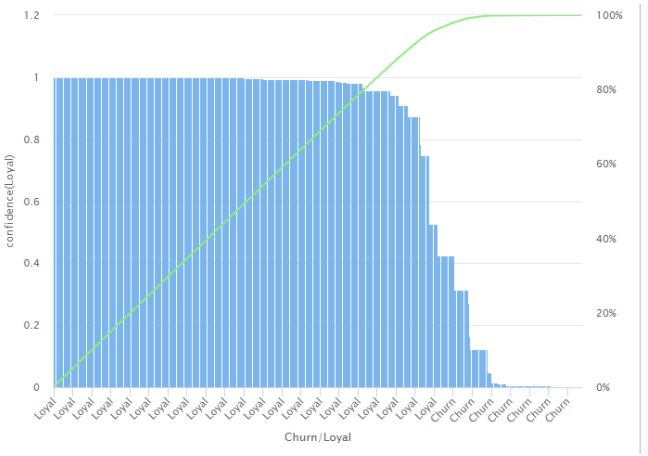

Figure 5: Description of Loyalty Prediction

5. Conclusion

Data mining modeling forms a classification that produces predictions of loyal and non-loyal consumers. It can be concluded that the Naive Bayes algorithm can be used to classify customer quality to predict loyal and non-loyal customers.

References

[1] D. T. Larose, *Discovering Knowledge in Data : An Introduction to Data Mining*. Hokoben, New Jersey: A John Wiley & Sons, Inc., 2005.
[2] E. Prasetyo, *Data Mining : Konsep dan Aplikasi dan Aplikasi Menggunakan Matlab*, 1st ed. Yogyakarta: Andi, 2012.
[3] Kusrini and E. T. Luthfi, *Algoritma Data Mining*, 1st ed. Yogyakarta: Andi, 2009.
[4] M. Yusa, E. Utami, and E. T. Luthfi, "Evaluasi Performa Algoritma Klasifikasi Decision Tree ID3, C4.5, dan CART pada Dataset Readmisi Pasien Diabetes," *InfoSys*, vol. Vol.4, no. No.1, pp. 23–34, 2016.
[5] W. Julianto, R. Yunitarini, and M. K. Sophan, "Algoritma C4.5 Untuk Penilaian Kinerja Karyawan," *Scan*, vol. Vo. IX, no. No. 2, pp. 33–39, 2014.
[6] P. Pitria, "Analisis Sentimen Pengguna Twitter Pada Akun Resmi Samsung Indonesia Dengan Menggunakan Naive Bayes," Universitas Komputer Indonesia, 2014.
[7] H. Leidiyana, "Penerapan Algoritma K-Nearest Neighbor untuk Penentuan Resiko Kredit Kepemilikan Kendaraan Bermotor," *J. Penelit. Ilmu Komputer, Syst. Embed. Log.*, vol. 1, no. 1, pp. 65–76, 2013.
[8] I. H. Witten and E. Frank, *Data Mining: Practical Machine Learning Tools and Techniques*, 2nd ed. San Francisco: Elsevier Inc, 2005.
[9] F. Gorunescu, *Data Mining : Concept, Model, and Techniques*. Berlin: Springer, 2011.
[10] I. Menarianti, "Klasifikasi Data Mining dalam Menentukan Pemberian Kredit Bagi Nasabah Koperasi," *J. Ilm. Teknosains*, vol. 1, no. 1, pp. 1–10, 2015.
[11] Susanti, "Klasifikasi Kredit Menggunakan Metode Decision Tree Pada Nasabah PD BPR BKK Gabus," 2010.

Author Profile

Teguh Iman Hermanto received Bachelor of Information System (M.Kom), Information System Program, Dian Nuswantoro University graduated in 2013. In 2015 he obtained his Master of Information System (M.Kom) degree from Diponegoro University Information Systems Study Program Semarang. currently active as a lecturer at STT Wastukancana Purwakarta.

Muhamad Agus Sunandar, received Bachelor of Engineering (S.T), Informatics Engineering Program, STT Wastukancana graduated in 2012. In 2016 he obtained his Master of Computer (M.Kom) degree from STMIK LIKMI Information Systems Study Program Bandung. currently active as a lecturer at STT Wastukancana Purwakarta.