e-ISSN: 3026-5657
Faculty of Islamic Economics and Business
UIN Prof. K.H. Saifuddin Zuhri Purwokerto
Banyumas, Jawa Tengah - INDONESIA

# Classification Interest Customer XL Internet Card Using Naive Bayes Algorithm at PT XML Tronik

**Dede Yusuf*[1], Zulfikar Yusya Mubarak[2], Nuni Wulansari[3], Annisa Rahayu P.[4], Fajar Nur Wibowo[5], Opi Irwansah[6], Asharryadi Neogroho[7], Atika Mahmuda[8], Nurul Alifah[9]**
[1,2,3,4,5,6,7,8]Universitas Al-Irsyad Cilacap, Indonesia
[9]STMIK Komputama Majenang Cilacap, Indonesia
Author's Correspondence: yusuf.dede17@gmail.com*[1]

**Abstract**
*In the world of trade, competition within a company has become commonplace and makes business people always think about breakthroughs that can ensure the continuity of their business . Basically companies have valuable assets but are rarely used as a reference, these assets are in the form of databases with large numbers. PT.XML Tronik is a company that has been in the technology sector for a long time, one of which is selling GSM Card Providers in Indonesia. The classification method used is the Naïve Bayes Algorithm, where Naïve Bayes is used to classify statistical and probability methods . Method This research uses several stages starting from data collection , analysis sales , data processing using naïve Bayes method and proceed with measurements level accuracy use WEKA application . From the results of data mining using the Naïve Bayes Algorithm, this researcher found patterns of interest in buying XL Card customers at PT.XML Tronik . Based on the results of the analysis carried out, using 20% of the test data with negative and positive labels. So that it can produce a sign ending with an accuracy level of 84.6667% with a precision of 97% and a recall of 83%.*

***Keywords:*** *Islamic Philanthropy, Socio-Economic, Productivity*

## A. Introduction

In the world of trade, competition within a company has become commonplace and makes business people always think about breakthroughs that can ensure the continuity of their business (Budi Santosa, 2007). Big data in a company can obtain new knowledge by utilizing technology that is adapted to current needs, this technology can assist in setting strategies for running its business. It is important to be able to predict consumer interest from within the selection interest goods or services, making it easier to make decisions as well as to determine effective and efficient strategies for consumers.

Basically companies have valuable assets but are rarely used as a reference, these assets are in the form of databases with large numbers. PT.XML Tronik is a company that

has been in the technology sector for a long time, one of which is selling GSM Card Providers in Indonesia. PT.XML Tronik must think of a strategy that can produce superior products that can help the marketing division at that time make decisions in determining marketing strategies. Of the many needs listed above, basically there are also many ways that can be implemented. One of them is by utilizing the database owned by the company in recent years using data mining methods.

Many people know data mining as the KDD (Knowledge Discovery in Database) Model, where data mining is a process of searching for meaningful patterns by looking at large data. Where the large data is stored in a storage and management using statistical techniques in large patterns, with relatively large amounts of data and various data sets.

In data mining, there is one mining method, namely the classification method, where this method focuses on finding a model or function that will explain and differentiate data classes. Where the goal is we can estimate the class of a data object something with an unknown label. The classification stage forms a model that is capable of differentiating data into different classes based on certain rules or functions. The model itself can be an "if-then" rule is a decision tree.(Atmanegara et al., 2022)

The classification method used is the Naïve Bayes Algorithm, where Naïve Bayes is used to classify statistical methods and probabilities that have been published by English scientists, namely Thomas Bayes, this theory is used to predict future opportunities as seen from previous experience, the longer the data used then the results will be more accurate. Until now the theory is known as Bayes' theorem, this theory is combined with the conditional approximation of free attribute reciprocity .

Data mining is part of the knowledge discovery process from data (KDD), where the KDD process stages (Budiman & Ramadina, n.d.)are described as follows:

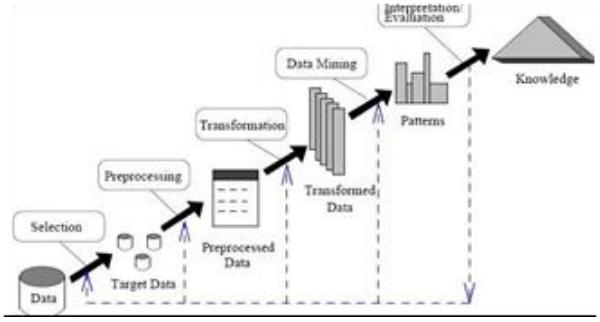



Figure 1 . Data Mining as a KDD Process

Here are some stages of data mining in the KDD process:

1. Data Selection

The first stage is the selection of data Finished on some of the data that has been collected. The database that will be used by the researcher is XL Card sales data as an object of research . From process This results is databases sales of select XL cards in the last 3 years.

2. T transformation

Transformation or normal We call it change , which is one of the processes in changing the initial data type that is adapted to the data requested by data mining.

3. Data mining

The process of data mining is the process of looking for patterns in data selected use method that exist in data mining . This process is carried out using algorithm data mining It is method Algorithm Naive Bayes (Budiman & Ramadina, n.d.).

4. Interpretation /evaluation

Data mining results need to be presented as such easily understood by others. At this stage of the process translate pattern information Which own have earned into the form Which Again easy understood by end users (Bay et al., n.d.) .

Bayes is one technique for predicting based on proballistic/probability-based implementation of simple Bayes theorem assumptions of strong independence. In other words, Naïve Bayes is an independent feature model (Eko Prasetyo, 2012). Prasetyo

explained that there is a connection between Naïve Bayes and classification , the correlation of the hypothesis and the classification evidence , while the hypothesis in Bayes ' theorem is that class labels are the target of mapping in the classification model , evidence is a feature that makes input into the classification model . If X is an input vector containing _ features and Y is a class label , Naïve Bayes is written as P(X|Y). The notation is how to label the probability class Y obtained after Feature X is observed. This notation is called the final probability (posterior probability) for Y, while P(Y) is called the initial probability (prior probability) of Y.(Eko Prasetyo, 2012)

Weka is application Which characteristic open source. Weka is an application that can be a solution for data analysis mining, text mining and predictive analysis. Weka uses several techniques predictive and descriptive in the give view to end user so it can create decision Which best .

## B. Research Method

In general, the stages in solving problems in this study were carried out systematically according to the description in Figure 5. As can be seen, the stages of Data Selection up to Interpretation are KDD processes. Gradually the final data model will be obtained at the expected association pattern.

a. Data collection

To support making this pattern using a learning methodology that is "Classification". Interest in Buying XL Card Customers Using the Naive Bayes Algorithm is the data collected at this stage in the form of XL internet card sales transaction data as a sample. And the data collection will be used as material for analysis. The performance of the data playback algorithm is the Naive Bayes algorithm, the system used for testing data on the main data tool, namely Weka (Cholil et al., 2019)

b. Sales Analysis

In the determine fan card Internet new Which will published is to collect internet card data previously marketed . To solve the problem of determining the interest of the card new will removed from existing problems Researchers need an analytical tool to use to see public interest in buying cards . So from that, the classification method expects to be able to find techniques to differentiate plans with the aim to be able to predict the class of an object Where the label is unknown . By Because That, algorithm naive bayes can estimate opportunity in the period front based on experience in Century before , more so will tested with use Weka. So you can find out consumer interest in internet cards What which will be launched on the market . The initial stage of the Naïve Bayes calculation process , Bayes is working on retrieving training data from sales card XL . As for variable determinant Which used in the classify data internet card that is :

1) Code card

   The card code is a variable type of internet card collected in the internet card code .
2) Quota

   Quota is variable Which collected in the two category It is quota above 10GB and quota under 10GB
3) Period Active

   Active Time is variable Once period use card XL grouped into 3 categories It is 2 week, 2 month , and 7 months
4) Price

   Variable price type price card XL Which grouped in the three category those are high , low , and medium .The following are stages Study with its flowchart format

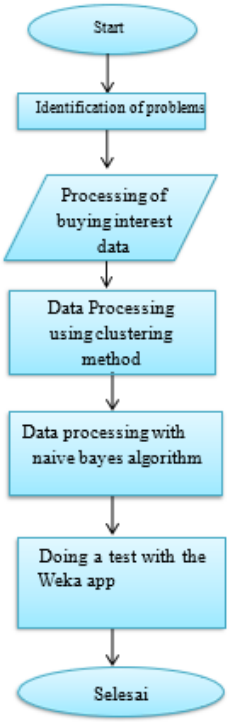



Figure 2. Research Flowchart

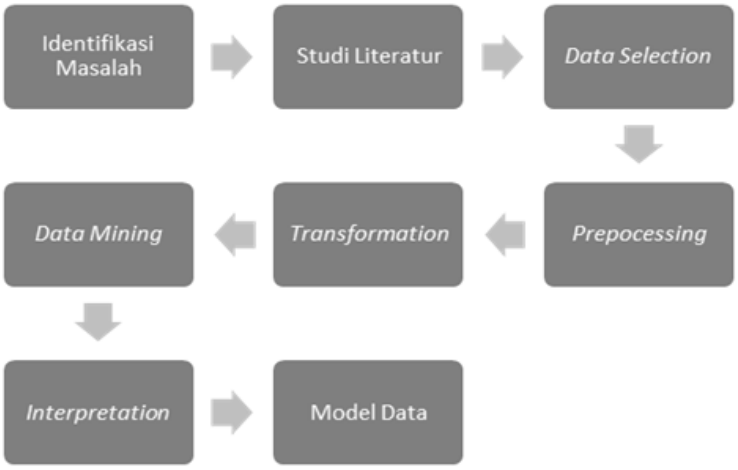



Figure 3 . Research Framework

## C. Results and Discussion

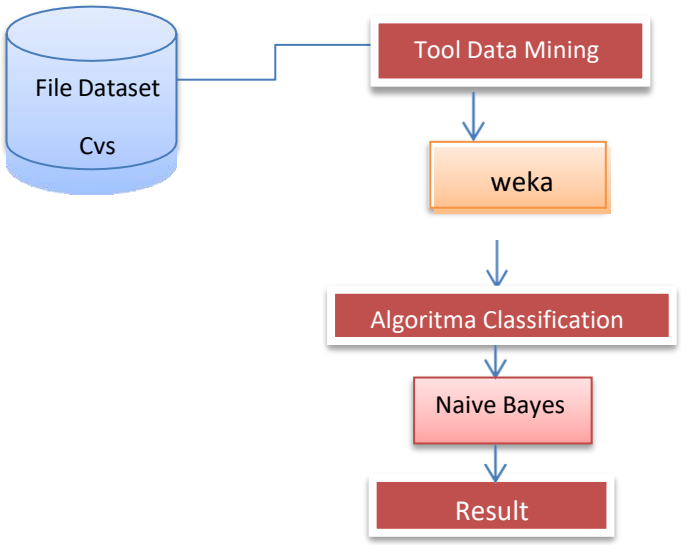
### a. Research Procedures

The research procedure uses the Weka application to determine data sets, and to select data sets to use the Naïve Bayes algorithm in the data mining process (Vembandasamyp et al., 2015). Then record the accuracy of the Naïve Bayes algorithm in the form of tables and graphs

The following is a collection of XL internet card sales transaction data as a sample

| NO | CARD CODE | QUOTA | ACTIVE PERIOD | PRICE | CLASSIFICATION |
|---|---|---|---|---|---|
| 1 | AAA01 | UNDER 10 | 1 MONTH | LOW | INTEREST |
| 2 | AAB02 | ABOVE 10 | 1 WEEK | AT THE MOMENT | NO |
| 3 | AAC03 | UNDER 10 | 1 MONTH | TALL | NO |
| 4 | AAD04 | ABOVE 10 | 1 WEEK | AT THE MOMENT | INTEREST |
| 5 | AAE05 | UNDER 10 | 1 MONTH | TALL | INTEREST |
| 6 | AAF06 | UNDER 10 | 1 MONTH | LOW | INTEREST |
| 7 | AAG07 | UNDER 10 | 1 MONTH | HEIGHT | INTEREST |
| 8 | AAH08 | UNDER 10 | 1 MONTH | LOW | INTEREST |
| 9 | AAI09 | UNDER 10 | 1 MONTH | CREATURE | INTEREST |
| 10 | AAJ10 | UNDER 10 | 1 MONTH | LOW | INTEREST |
| 11 | AAK11 | UNDER 10 | 1 MONTH | CREATURE | INTEREST |
| 12 | AAL12 | UNDER 10 | 1 MONTH | HEIGHT | INTEREST |
| 13 | AAM13 | UNDER 10 | 1 MONTH | CREATURE | INTEREST |
| 14 | AAN14 | UNDER 10 | 1 MONTH | HEIGHT | INTEREST |
| 15 | AAO15 | UNDER 10 | 1 MONTH | LOW | INTEREST |
| 16 | AAP16 | UNDER 10 | 1 MONTH | HEIGHT | INTEREST |
| 17 | AAQ17 | UNDER 10 | 1 MONTH | LOW | INTEREST |
| 18 | AAR18 | UNDER 10 | 1 MONTH | AT THE MOMENT | INTEREST |
| 19 | AAS19 | UNDER 10 | 1 MONTH | LOW | INTEREST |

Table 1 Data Transaction Sale Card XL




Picture 4. Scheme Study

**b. Stages Processing data in the application Weka**

1. This is results from appearance classification results

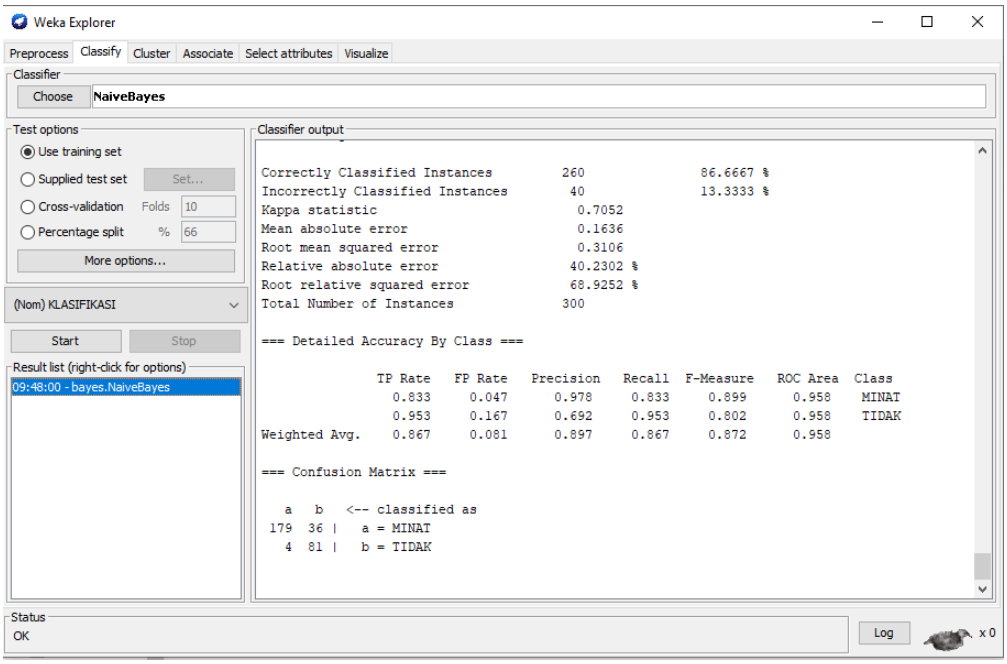



Figure 6. Classification Results

From the above picture it can be seen that the naive Bayes has an accuracy rate of 86.6667% with an error rate of 13.3333%

2. This is results from appearance show vector on part text see

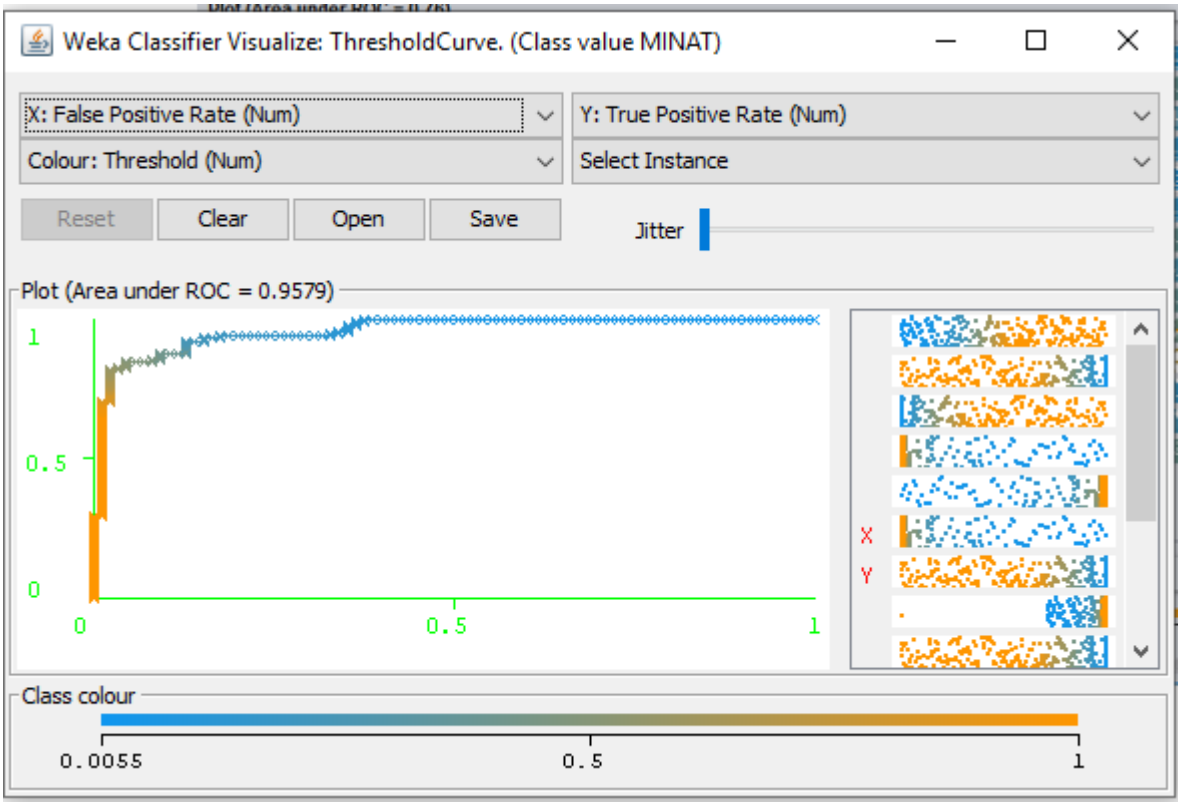



Figure 7. Classification Curve Results

Shows that *the Area Under the Curve* of 0.9579, this means that the level of classification diagnosis _ is at the level of *Classification Good* / Very Good .

## D. lts and Discussion

From the results of data mining using the Naïve Bayes Algorithm, this researcher found patterns of interest in buying XL Card customers at PT.XML Tronik . Based on the results of the analysis carried out, using 20% of the test data with negative and positive labels. So that it can produce a sign ending with an accuracy level of 84.6667% with a precision of 97% and a recall of 83%.

a. Based on the confusion matrix above the errors/errors come from 36 data with an Interest class that are incorrectly interpreted as a non-interest class as well as there

are 4 data with a disinterested class that are incorrectly classified into an interest class

b. Of the 300 existing data , it can be seen that there are 215 data that are interested and 85 data that are not interested.

**References**

Atmanegara, A., Ciptaningsi, R., Rasywir, E., Meisak, D., & Computer, FI (2022). *Journal of Informatics and Computer Engineering (JAKAKOM) Application of Data Mining Algorithm Naive Bayes Classifier To Find Out Customer Purchase Interest For Telkomsel Cards (Rick Cell Simpang Candra)* (Vol. 1, Issue 2).

Bay, SD, Kibler, D., Pazzani, MJ, & Smyth, P. (nd). *The UCI KDD Archive of Large Data Sets for Data Mining Research and Experimentation** . http://www.google.com

Budi Santosa. (2007). *Data Mining : Data Utilization Techniques for Business Purposes* (1st ed.). Scienc e House.

Budiman, I., & Ramadina, R. (nd). Application of Classification Data Mining Functions for Timely Prediction of Student Study Period in Higher Education Academic Information Systems. *IJCCS , x, No. x* (1), 1–5.

Cholil, W., Dalimunthi, AR, & Atika, L. (2019). Data Mining Models in Identifying Patterns of Growth Rates Between Economic Sectors in the Provinces of South Sumatra and Bangka Belitung. *Teknika , 8* (2), 103–109. https://doi.org/10.34148/teknika.v8i2.181

Eko Prasetyo. (2012). *Data Mining: Concepts and Applications Using Matlab* . CV Andi Offset.

Vembandasamyp, K., Sasipriyap, RR, & Deepap, E. (2015). Heart Diseases Detection Using Naive Bayes Algorithm. In *IJISET-International Journal of Innovative Science, Engineering & Technology* (Vol. 2, Issue 9). www.ijiset.com