

✓

# Use of Naïve Bayes classifier algorithm to detect customers' interests in buying internet token

*by Dedy Prehanto*

---

**Submission date:** 05-Oct-2021 08:04PM (UTC+0700)

**Submission ID:** 1665892477

**File name:** Prehanto\_2019\_J.\_Phys.\_Conf.\_Ser.\_1402\_066069.pdf (806.51K)

**Word count:** 2868

**Character count:** 14646

PAPER • **7** OPEN ACCESS

## Use of Naïve Bayes classifier algorithm to detect customers' interests in buying internet token

**6** To cite this article: D R Prehanto *et al* 2019 *J. Phys.: Conf. Ser.* **1402** 066069

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

A promotional banner for the 240th ECS Meeting. The banner features the ECS logo on the left, followed by the text '240th ECS Meeting' in blue, 'Digital Meeting, Oct 10-14, 2021' in black, 'We are going fully digital!' in green, 'Attendees register for free!' in black, and 'REGISTER NOW' in red. On the right side of the banner is a photograph of a diverse group of people in a meeting setting, with a white arrow pointing from the text towards them.

**ECS** **240th ECS Meeting**  
Digital Meeting, Oct 10-14, 2021  
**We are going fully digital!**  
Attendees register for free!  
**REGISTER NOW**

## 7 Use of Naïve Bayes classifier algorithm to detect customers' interests in buying internet token

D R Prehanto<sup>1,\*</sup>, A D Indriyanti<sup>1</sup>, K D Nuryana<sup>1</sup>, S Soeryanto<sup>1</sup> and A S Mubarak<sup>2</sup>

<sup>1</sup> Faculty of Engineering, Universitas Negeri Surabaya, Surabaya, Indonesia

<sup>2</sup> Faculty of Engineering, Universitas Hasyim Asy'ari Jombang, Jombang, Indonesia

\*dedyrahman@unesa.ac.id

**Abstract.** Determining possible technology used in calculating customer buying interest is important for the company as it relates to marketing strategies and policies. The Naïve Bayes Classifier is a possible method to measure customer buying interest. This study aims to analyze the accuracy of this method in calculating customer buying interest of internet token. Data were obtained from past sales and then analyzed using the Naïve Bayes Classifier to predict future opportunities by defining the class of attributes. The classifications used were operator, internet quota, token active period, and product price. This study found that the Naïve Bayes Classifier was provably accurate in calculating customer buying interest as well as comparing the prediction with actual results aside the data training. Results showed that this method successfully classifies 10 products of 858 transactions as data training and 10 products of 115 transactions as data test. Here indicates that all data test and data training analysis showed eight of the tenth meaning that eight of ten predictions were correct or it had 80% accuracy. This study is expected becoming a reference in analyzing data sales and predicting future sales conditions, so it will help to determine appropriate strategy of product development.

### 1. Introduction

Competition in the business world encourages businesses to continue to make breakthrough innovations for the sake of their business continuity. Nowadays cellular store business companies have offered various types of internet packages with various types and prices, but without knowing the amount of consumer interest in the package. This certainly will make consumers turn to companies similar to packages that according to the most profitable consumers [1]. By knowing the interests of the community indirectly will help increase sales turnover or can provide stock according to demand. In addition to knowing the interests of the community, classification can also be made on the sale of goods. By using this classification is able to provide information on the number of purchases of the community.

Classification methods are intended to find models or functions that explain or distinguish concepts or data classes to be able to estimate class and objects whose labels are unknown [2]. In achieving these objectives, the classification process form model that is able to distinguish data into different class classes based on certain rules or functions. Bayes naive classifier algorithm is one of the algorithms found in classification techniques [3]. A study identified empirically relevant features to classify customer satisfaction with the aim of improving hotel services and can help managers understand the satisfaction of guests [4]. The application of the Naive Bayes method is also done help consumers to choose the right vehicle and also to predict the future purchase of the car [5]. The application of the Naive Bayes method to the IndiHome customer classification PT. TELKOM to carry out a marketing



1  
Content from this work may be used under the terms of the Creative Commons Attribution 3.0 licence. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI

Published under licence by IOP Publishing Ltd



cross selling strategy, namely by offering new products [6]. The application of classification with the Bayes classifier naive algorithm in this study is expected to be able to predict or detect potential customer buying interest early on based on past transactions. Thus companies can better understand which sales packages their customers will be interested in.

In general, the purpose of this study is to present an overview and prediction of consumer buying interest in the Internet package pulse. The details of the purpose of this study are, among other things, to be able to classify internet pulses with naïve Bayes classification algorithms, and to determine the level of customer buying interest in customer buying interest in internet package products. So that in this study, the formulation of the problem can be taken, among others, to detect with the Naive Bayes Classifier algorithm in the internet package sales business and classification related to the detection of customer buying interest in internet package sales.

## 2. Method

In the prediction program for customer buying interest in new internet packages, the approach in this study is a quantitative approach, because this research is presented with numbers. According to Saleh which suggests that quantitative research is a lot of research demanded to uncover the numbers, starting from data collection, data processing, interpretation of the data and the appearance of the results [7].

### 2.1. Research design

The author makes a system design that includes making Flowcharts system. At this stage the model is used in the new data, namely internet package sales data from January 2016 to February 2016. Diagram of the research conducted at figure 1.

2.1.1. Flowchart system. This flowchart describes the process carried out by the user in running the system. The flowchart starts when the user enters a username and password which will then be checked by the system, whether the username entered is appropriate or not. If it doesn't match, the system will display a warning to enter the correct username and password. If it is appropriate, then the user will be redirected to the main page, namely alternative data. The user will then enter collection data, preprocessing data, distribution data, here is a decision support system flowchart can be seen below.

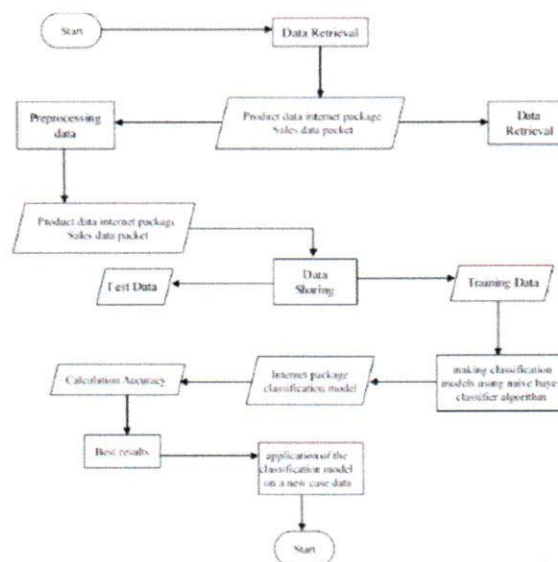


Figure 1. Flowchart system.

## 2.2. Population and sample

The population and sample used in this study is to use data from use data on internet package products sold along with internet package sales data that occurred in May 2015 to February 2016.

## 2.3. Data collection techniques

The techniques used in data collection by the author are interviews, observation, and literature studies. Interviews were conducted with seller and manager.

## 2.4. Data analysis techniques

The author chooses the Simple Additive Weighting method in this study, where this method is centered on weighted summation which will be ranked according to the highest number [2]. This ranking

**2.4.1. Data collection.** Systematic procedure used to collect data is to collect product data along with sales transaction data.

**2.4.2. Data pre-processing.** In the pre-processing phase of the data, data selection is carried out [8]. The selection of data in the selection process consists of 4 attributes, including data on internet package operators, data on the main internet package quota, the active period of packages and the selling price of internet packages. If there is numerical data such as the quota data, price and active period of the internet package, what should be done is to find the mean and standard deviation of each parameter which is numerical data. The equation used to calculate the average value of the calculation (mean) can be seen as follows.

$$\mu = \frac{\sum_{i=1}^n x_i}{n} \text{ or } \mu = \frac{x_1 + x_2 + x_3 + \dots + x_n}{n}$$

Information :

$\mu$  = average count (mean)

$x_i$  = sample value to -i

$n$  = number of sampel

**2.4.3. Data distribution.** At this stage, the data is divided into 2, namely training data (training data) and test data. Training data and test data use data on internet package products sold along with internet package sales data that occurred in May 2015 to February 2016. In the study, the method used in the process of separating training data and test data is divided into 75 to 35 which means 75% of the first data is set as training data (training data) while 35% of the data is used as test data.

**2.4.4. Make a classification model.** In making a classification model, it is necessary to determine attributes. In this study the attributes have been determined which include: internet package operator or provider, internet package quota, internet package active period and internet package selling price. After the attribute is successfully set, the next step is to determine the classification of these products, so that the classification can be done.

**2.4.5. Application using new case data.** At this stage the model is used in the new data, namely internet package sales data from January 2016 to February 2016.

## 3. Results and discussion

The results obtained after conducting this research will be elaborated from each program display form along with the functions of the form [7].



### 5.1. Application of naïve bayes classifier

Naïve Bayes is a classification with probability and statistical methods presented by British scientist Thomas Bayes. that is, predicting future opportunities based on previous experience so that it is known as the Bayes theorem [9]. The greater the amount of training data, the more accurate the results will be. If there is numerical data such as the quota data price and active period of the internet package, all data is continuous, then what must be done is to find the mean and standard deviation of each parameter which is numerical data.

3.1.1. *Training data.* On the application of naïve Bayes classifier, the first thing to do is to determine the training data and test data. The first step is to classify products that have been marketed so that they get a class of "interest" or "no", by determining the average value of the overall sales in the sample data. This study uses a total of 912 training data transactions, from that number the average transaction can be taken by means of total transactions divided by the number of products, which will produce the following values:

$$\frac{\sum x}{n} = \frac{912}{10} = 91.2$$

So that each product has a total transaction of more than or equal to 91.2 can be given a grade "demand" as well as on the contrary be labeled "not interested".

### 3.1.2. Criteria and probability

#### 1) Probability of criteria for internet package operators

Operators in this research are variables that explain the parties acting as internet service providers or commonly called "operators".

**Table 1.** Operators.

Operator	Label Interest	Probability		
		No	No	No
Indosat	3	1	0.75	0.17
Telkomsel	1	2	0.25	0.33
XI	0	3	0.00	0.5
Total	4	6	0.4	0.6

#### 2) Probability of the main internet package quota criteria

In this study quota variables are variables grouped into two categories, namely "high" and "low". "High" quota includes all packages that have a quota of more than 2.5 GB, while quotas with a value of 2.5 GB or quota lower than 2.5 GB are included in the "low" quota category.

#### 3) The probability of the criteria for the active period of the internet package

It is a variable for the duration of the internet package life that can be categorized into 2 categories, namely "low" which is an active period with a duration of less than or equal to 30 days, and a category "high" for an active period of longer than 30 days.

#### 4) Probability of the internet package selling criteria

This variable is a variable in the size of the selling price of the internet package which is categorized into 2 categories, namely high and low. To determine a category from each price, the first to do is calculate the average of the price:

$$\frac{\sum x}{n} = \frac{454000}{10} = 45400$$

The "high" category is the price that is more than the average price of all products, while the average value of the product price is Rp. 45,400. The "low" category is the price that is less than or equal to the average price of all products, while the average value of the product price is Rp. 45,400.

3.1.3. *Testing.* This research I started by entering new data, here the author exemplifies by entering new internet package product data with the following conditions

**Table 2.** Example of entering data.

Sample data	Name	Main quota	Active period	Operator	Price
	Indosat Data 2gb	2 Gb	30 Hari	Indosat	Rp.36..500

From the above test data that has been entered, the second stage is the classification of data. This classification is in accordance with the provisions of the package product data with the provisions described in the previous discussion

**Table 3.** The classification of data.

Sample data	Name	Main quota	Active period	Operator	Price
	Indosat Data 2gb	Low	Low	Indosat	Low

**Start Classification Calculation**

1. Calculates  $P(X|C_i)$  for each Class (label)
2.  $\rho(\text{operator name}|\text{interest}) = p(\text{Indosat}|\text{interest}) = \frac{3}{4} = 0.75$
3.  $\rho(\text{operator name}|\text{not}) = p(\text{Indosat}|\text{not}) = \frac{1}{4} = 0.25$
4.  $\rho(\text{active period}|\text{interest}) = p(\text{low}|\text{interest}) = \frac{4}{9} = 0.4444444444444444$
5.  $\rho(\text{active period}|\text{not}) = p(\text{low}|\text{not}) = \frac{5}{9} = 0.5555555555555556$
6.  $\rho(\text{main quota}|\text{interest}) = p(\text{low}|\text{interest}) = \frac{3}{6} = 0.5$
7.  $\rho(\text{main quota}|\text{not}) = p(\text{low}|\text{not}) = \frac{3}{6} = 0.5$
8.  $\rho(\text{price}|\text{interest}) = p(\text{low}|\text{interest}) = \frac{2}{3} = 0.6666666666666666$
9.  $\rho(\text{price}|\text{not}) = p(\text{low}|\text{not}) = \frac{1}{3} = 0.3333333333333333$

3.2. *Calculation of interest / no class*

$$\rho(\chi|\text{interest}) = 0.75 \times 0.4444444444444444 \times 0.5 \times 0.6666666666666666 = 0.1111111111111111$$

$$\rho(\chi|\text{not}) = 0.25 \times 0.5555555555555556 \times 0.5 \times 0.3333333333333333 = 0.023148148148148147$$

3.3. *Calculation of interest / no class with class label accumulation*

$$\rho(X|C_i) \cdot \rho(C_i) = \rho(X|\text{interest}) \cdot \rho(\text{interest})$$

$$= 0.1111111111111111 \times \frac{4}{10}$$

$$= 0.1111111111111111 \times 0.4$$

$$= 0.044444444444444446$$

$$\rho(X|C_i) \cdot \rho(C_i) = \rho(X|\text{not}) \cdot \rho(\text{not})$$

$$= 0.023148148148148147 \times \frac{6}{10}$$

$$= 0.009259259259259259$$

3.4. *Comparison of each class*

Based on the final calculation by multiplying the opportunity value of the adopted case, we see that the value  $\rho(X|\text{interest}) \cdot \rho(\text{interest})$  is higher than  $\rho(X|\text{not}) \cdot \rho(\text{not})$  that is equal to 0.04444444446 with



0.009259259259 so that it can be concluded that the Internet Card is included in the "Attracted" classification.

#### 4. Conclusion

Based on internet package product sales data used as training data, the Naive Bayes method successfully classifies 10 product data with 858 transaction data as training data and 10 product data with 115 transaction data as test data. From the analysis of all test data and training data, the truth ratio is 8/10, which is 8 out of 10 products with predicted true values. So the naïve Bayes classification method predicts large customer buying interest in internet package products with an 80% accuracy percentage. Users can classify internet package products based on previous sales, by classifying based on each class of all existing attributes. The more sales data that has occurred before, the classification is also more accurate.

#### References

- [1] Permadi G S, Adi K and Gemowo R 2018 Application Mail Tracking Using RSA Algorithm As Security Data and HOT-Fit a Model for Evaluation System *E3S Web of Conferences* **31** 11007
- [2] Saleh A 2015 Implementasi Metode Klasifikasi Naïve Bayes Dalam Memprediksi Besarnya Penggunaan Listrik Rumah Tangga
- [3] Wu J, Pan S, Zhu X, Cai Z, Zhang P and Zhang C 2015 Self-adaptive attribute weighting for Naive Bayes classification. *Expert Systems with Applications ELSEVIER*
- [4] Manuel J, Sánchez-Franco, Antonio N G and Francisco J R C 2018 A naïve Bayes strategy for classifying customer satisfaction: A study based on online reviews of hospitality services *Journal of Business Research*
- [5] Harahap F, Harahap A Y H N, Ekadiansyah, Sari R N, Adawiyah R and Harahap C B 2018 Implementation of Naïve Bayes Classification Method for Predicting Purchase *The 6th International Conference on Cyber and IT Service Management (CITSM 2018)*
- [6] Sfenrianto, Pumamasari I and Bahaweres 2016 Naïve Bayes Classifier Algorithm and Particle Swarm Optimization for Classification of Cross Selling (Case Study: PT TELKOM Jakarta) *ELSEVIER*
- [7] Arikunto S 2002 *Prosedur Penelitian Suatu Pendekatan Praktik* (Jakarta: Rineka Cipta)
- [8] Han J, Pei J and Kamber M 2006 *Data mining: Concepts and techniques* (Elsevier)
- [9] Kusri and Emha T L 2009 *Algoritma data mining* (Yogyakarta: Andi Offset)



# Use of Naïve Bayes classifier algorithm to detect customers' interests in buying internet token

## ORIGINALITY REPORT

19%

SIMILARITY INDEX

16%

INTERNET SOURCES

17%

PUBLICATIONS

0%

STUDENT PAPERS

## PRIMARY SOURCES

1	<a href="https://repository.poliupg.ac.id">repository.poliupg.ac.id</a> Internet Source	6%
2	<a href="https://ejournal.unhasy.ac.id">ejournal.unhasy.ac.id</a> Internet Source	2%
3	"Proceedings of the Second International Conference on the Future of ASEAN (ICoFA) 2017 – Volume 2", Springer Science and Business Media LLC, 2018 Publication	2%
4	A. D. Indriyanti, D. R. Prehanto, T. Z. Vitadiar. "K-means method for clustering learning classes", Indonesian Journal of Electrical Engineering and Computer Science, 2021 Publication	1%
5	D Syafira, S Suwilo, P Sihombing. "Analysis of Attribute Reduction Effectiveness on The Naive Bayes Classifier Method", Journal of Physics: Conference Series, 2020 Publication	1%

[repository.unair.ac.id](https://repository.unair.ac.id)

6	Internet Source	1 %
7	<a href="https://iopscience.iop.org">iopscience.iop.org</a> Internet Source	1 %
8	Adi Wijaya, Surya Darma, Dio Caisar Darma. "Spatial Interaction Between Regions: Study of the East Kalimantan Province, Indonesia", International Journal of Sustainable Development and Planning, 2020 Publication	1 %
9	Pramana Yoga Saputra, Yoppy Yunhasnawa, Windy Fatmila, Faisal Rahutomo et al. "The Development of Classification System of Student Final Assignment Using Naive Bayes Classifier Case Study: State Community Academy of Bojonegoro", International Journal of Engineering & Technology, 2018 Publication	1 %
10	Saruni Dwiasnati, Yudo Devianto. "Utilization of Prediction Data for Prospective Decision Customers Insurance Using the Classification Method of C.45 and Naive Bayes Algorithms", Journal of Physics: Conference Series, 2019 Publication	<1 %
11	<a href="https://speakerdeck.com">speakerdeck.com</a> Internet Source	<1 %



- 12 David Ahmad Effendy, Kusrini Kusrini, Sudarmawan Sudarmawan. "Classification of intrusion detection system (IDS) based on computer network", 2017 2nd International conferences on Information Technology, Information Systems and Electrical Engineering (ICITISEE), 2017  
Publication <1 %
- 
- 13 [www.researchgate.net](http://www.researchgate.net)  
Internet Source <1 %
- 
- 14 [e-jurnal.pnl.ac.id](http://e-jurnal.pnl.ac.id)  
Internet Source <1 %
- 
- 15 Muhammad Ibnu Sa'ad, Doni Bryan, Kusrini, Supriatin. "Decision Support System for Covid19 Affected Family Cash Aid Recipients Using the Naïve Bayes Algorithm and the Weight Product Method", 2020 3rd International Conference on Information and Communications Technology (ICOIACT), 2020  
Publication <1 %
- 
- 16 [jurnalstmiksubang.ac.id](http://jurnalstmiksubang.ac.id)  
Internet Source <1 %
- 

Exclude quotes  On

Exclude matches  Off

Exclude bibliography  On