



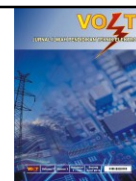
P-ISSN: 2528-5688  
E-ISSN: 2528-5696

**VOLT**

**Jurnal Ilmiah Pendidikan Teknik Elektro**

Journal homepage: [jurnal.untirta.ac.id/index.php/VOLT](http://jurnal.untirta.ac.id/index.php/VOLT)

**Vol. 2, No. 2, October 2017, 125-132**



## **IMPLEMENTATION OF NAIVE BAYES CLASSIFIER ALGORITHM TO EVALUATION IN UTILIZING ONLINE HOTEL TAX REPORTING APPLICATION**

**R Dimas Adityo<sup>1✉</sup>, Herti Miawarni<sup>2</sup>**

<sup>1</sup>Department of Informatics Engineering, Faculty of Engineering  
Universitas Bhayangkara, Surabaya 60231, Indonesia

✉Corresponding author e-mail: [dimas@ubhara.ac.id](mailto:dimas@ubhara.ac.id)

<sup>2</sup>Department of Electrical Engineering, Faculty of Engineering  
Universitas Bhayangkara, Surabaya 60231, Indonesia

*Received: 03 September 2017. Received in revised form: 22 October 2017. Accepted: 23 October 2017*

### **Abstract**

The current implementation of tax reporting regional Pasuruan hotels have used online (Web-based), with the aim of reporting systems can run effectively and efficiency in receiving the financial statements especially from taxpayer property. Pasuruan as one small town quite rapidly in East Java, have implemented role models online tax filing system starting in 2015, with the amount of 6 hotels, there are several classes of hotels ranging from budget class up to class three stars. After the application of the system running for 18 months (2015-2016), from existing data, conducted research on the analysis of the level of compliance of taxpayers reporting incomes in hotel. On the research was designed and built a system to evaluate the level of compliance with the performance from the taxpayer (WP) in the 2nd year (2016) and are classified in categories (1) the taxpayer (WP) very obedient (ST), (2) the taxpayer (WP) is quite obedient (CT), (3) Taxpayers (WP) less obedient (KT). Input data will be processed using the technique of data mining algorithms Naive Bayes Classifier (NBC) to form the table of probability as a basis for the process of classification levels of taxpayer compliance. Based on the results of the measurement, the test results show with accuracy of 50% i.e. 3 taxpayers is the very obedient (ST) to pay taxes. Then from the classification, the study could be made of recommendation solutions to guide the Taxpayer in reporting revenues well and true.

**Keywords:** Classification, Naïve Bayes, Online System

### **INTRODUCTION**

Pasuruan is a town in the province of East Java, Indonesia. The town is located 60 km southeast of Surabaya, capital of East Java

province and 355 km northwest of Denpasar, Bali. The entire region is bordered by the town of Pasuruan Regency Pasuruan. Pasuruan is located on the main line linking the northern coast of the island of Java with Bali island which makes it a city with great economic prospects in the eastern Indonesia region.



Figure 1. Location Map of Pasuruan, East Java

Along with the development of technology and the progress of human civilization that was already growing rapidly in all fields of life. Science and technology nowadays become an integral part of the life of modern society. By leveraging the technology based on the internet (online), financial reporting systems relating to the daily income of a taxpayer (WP) for hotels, the current can be monitored in realtime (Figure 2). In contrast to previous reporting system by the method of self assessment, i.e. a fully reporting system is entrusted by the taxpayers to fill his daily income honestly manually (Adityo & Krisdiyono, 2015).

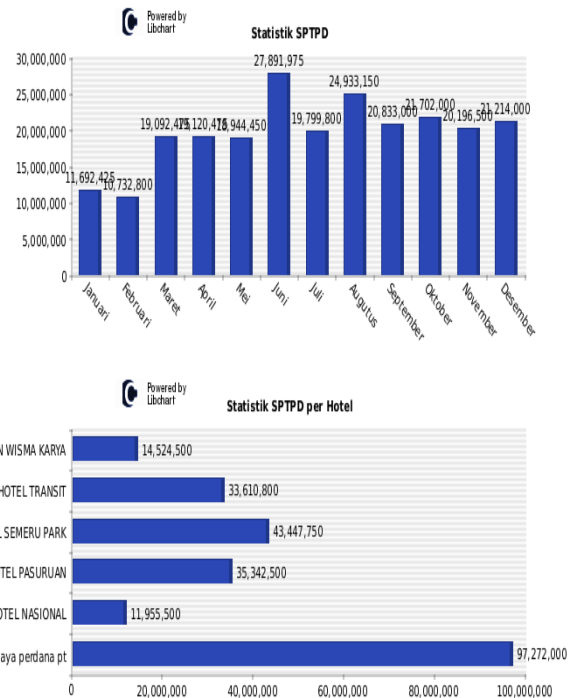


Figure 2. Graphic statistics of Income Acquisition Hotel Year 2016

Naive Bayes Classification method is one of the popular Classification Algorithm, Naive Bayes competitive Performance in the process of classification despite using the assumption of keidependenan attributes (there is no relationship between attributes). Assumption keidependenan this attribute on the data actually rarely happens, but although the assumption keidependenan atribut masses but Naive Bayes classification performance breached quite high, this is proven on a wide range of empirical research.

Classification is the employment rate data object to include it in a particular class of a number of classes that are available. In the classification there are two main work undertaken, namely (1) the construction of the model as a prototype to be stored as memory and (2) the use of the model to perform recognition/classification/prediction on an

object so that other data unknown in the classroom where the data objects in the model that's been kept (Prasetyo, 2014). Classification is a method of data mining can be used to search a set of process model (function) that can describe and differentiate the classes data or concept, whose goal so that the model can be used to predict the class object label is not known or can predict the trend of data that appear in the future. A method of classification also aims to do the mapping data into classes that already defined beforehand based on the value of the attribute data (Prasetyo, 2014). The classification process as shown in Figure 3.

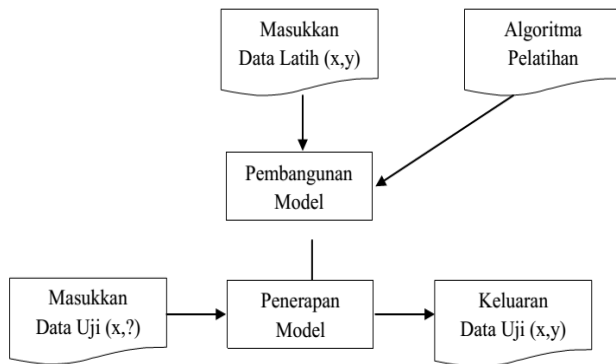


Figure 3. Classification Process

Bayes method is a statistical approach to do an induction inference on the question of classification. This method uses the conditional probability as the basic. In the science of statistics, conditional probability expressed as Figure 4.

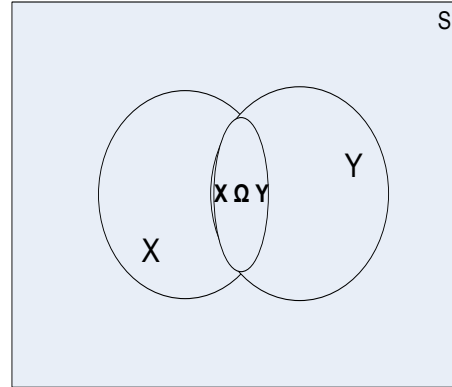


Figure 4. Bayesian Method

The probability of X in Y is the probability interseksi X and Y from the probability of Y, or with other languages  $P(X | Y)$  is the percentage of the number X in y. in addition to the above data such as Naive Bayes method can also handle numerical data in the form. To handle numerical data using Naive Bayes method assuming a normal distribution. As formulated into the following formula :

$$\mu = \frac{1}{n} \sum_{i=1}^n x_i$$

$$\sigma = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \mu)^2}$$

$$f(w) = \frac{1}{\sqrt{2\pi} \sigma} e^{-\frac{(w-\mu)^2}{2\sigma^2}} \tag{1}$$

Calculation of Naïve Bayes on research

For example want to know if an object entered in the requirements for the payment of taxes feasible or not with Naive Bayes algorithm. To set a tax payment will be selected as a value for Classified categories of taxes, has

compiled 10 months. There are four attributes that are used as shown in table 1.

Table 1. Attribute Assignment Tax Payers

Bulan ke-	(C1)	C2	C3	C4
1	100	2	No	Yes
2	200	1	No	Yes
3	500	3	No	Yes
4	600	20	No	No
5	550	8	No	No
6	250	25	Yes	No
7	75	15	Yes	No
8	80	10	No	Yes
9	700	18	Yes	No
10	180	8	Yes	Yes

Description:

- C1 : the value of the tax payments/month
- C2 : the difference in the value of the tax that is deposited with a targeted
- C3 : the daily transactions exceeded
- C4 : the decision to choose the set of such categories of reasonable tax payment or not

The mean and standard deviation for tax payment attributes/month (C1)

$$\mu_{ya} = \frac{100 + 200 + 500 + 80 + 180}{5} = 212$$

$$\mu_{tidak} = \frac{600 + 550 + 250 + 75 + 700}{5} = 435$$

$$\sigma^2_{ya} = \frac{(100 - 212)^2 + (200 - 212)^2 + (500 - 212)^2 + (80 - 212)^2 + (180 - 212)^2}{5 - 1} = 28520,015$$

$$\sigma^2_{ya} = \sqrt{28520,015} = 168,8787$$

$$\sigma^2_{tidak} = \frac{(600 - 435)^2 + (550 - 435)^2 + (250 - 435)^2 + (75 - 435)^2 + (700 - 435)^2}{5 - 1}$$

$$\sigma^2_{tidak} = \sqrt{68624,98} = 261,9637 = 68624,98$$

Mean and variants to attribute the difference deposited with a target of tax (C2)

To calculate the mean value and variants, described into the formula below:

$$\mu_{ya} = \frac{2 + 1 + 3 + 10 + 8}{5} = 4,8$$

$$\mu_{tidak} = \frac{20 + 8 + 25 + 15 + 18}{5} = 17,2$$

$$\sigma^2_{ya} = \frac{(2 - 4,8)^2 + (1 - 4,8)^2 + (3 - 4,8)^2 + (10 - 4,8)^2 + (8 - 4,8)^2}{5 - 1} = 15,699821$$

$$\sigma^2_{ya} = \sqrt{15,699821} = 3,9623$$

$$\sigma^2_{tidak} = \frac{(20 - 17,2)^2 + (8 - 17,2)^2 + (25 - 17,2)^2 + (15 - 17,2)^2 + (18 - 17,2)^2}{5 - 1} = 39,700081$$

$$\sigma^2_{tidak} = \sqrt{39,700081} = 6,3008$$

As for the probability that the difference in tax attributes the increase to the target reasonable category and attribute/not, shown in Table 2 and Table 3.

Table 2. The probability of occurrence of each attribute value exists whether or not the transaction exceeded (C3).

The transaction exceeded	The number of selected events		Probability	
	Yes	No	Yes	Np
Exist	1	3	1/5	3/5
No	4	2	4/5	2/5
Total	5	5	1	1

Table 3. The Probability Of Occurrence Of Each Attribute Value Is Selected For The Fair/No (C4)

Selected for the Reasonable Assignment /not	The number of selected events		Probability	
	Yes	No	Yes	No
Total	5	5	1/2	1/2

**METHOD**

In research, process or data processing stages from beginning to end is described into the following research methods:

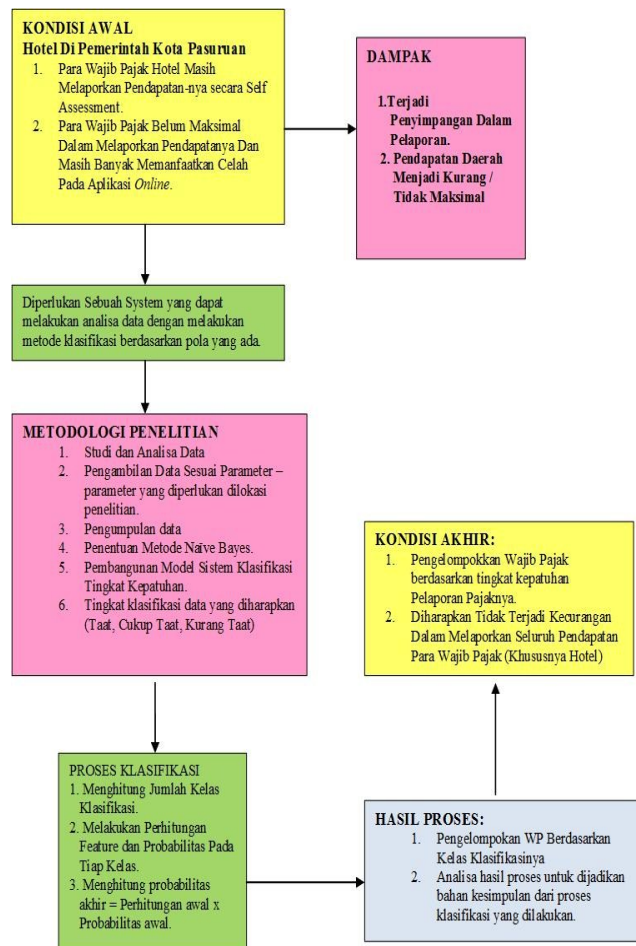


Figure 5. Research Method

Based on Figure 5. This research starts from the initial conditions, namely: (1) in this case the tax payers still reported revenues in the self assessment. Namely by reporting revenues independently sign there is scrutiny from outsiders on the basis of trust, (2) The taxpayer is not a maximum in reporting his income by making use of a loophole in the online tax application. Based on initial conditions is this research needs to be done in determining the level of compliance by taxpayers. Do they belong in the category of 'obey ', 'enough obedient' and 'less obedient ' in reporting his earnings. Therefore required a software application in analyzing the compliance level of taxpayer revenues reported.

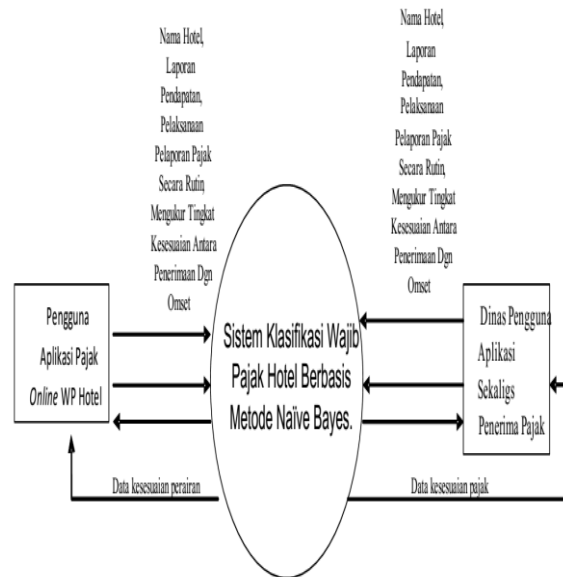


Figure 6. Context Diagram

Figure 6 is a context diagram, the diagram is a diagram that describes the context of the system as a whole. Context diagram consists of two elements, namely hotel taxpayers and office of regional income Pasuruan.

**RESULT AND DISCUSSION**

Pengujian sistem merupakan pengujian dalam memasukkan data ke dalam form – form yang disediakan. Pada tahap ini pengujian dilakukan dengan menggunakan 108 ( 18 bln x 6 Hotel) data training pada sistem. Berdasarkan hasil pengujian dari 108 data training diperoleh hasil bahwa terdapat 97 data yang sesuai dengan kelas sebenarnya. System testing is testing in entering data into the form provided. At this stage of testing is done using 108 (18 months x 6 hotels) data training on the system. Based on the test results from 108 data training obtained the results that there were 97 data corresponding to the actual class.

Table 4. System Testing Result

Data Perolehan Setoran Pajak Selama 18 Bulan Milik 6 WP					
NO	ID WP	Bulan	Pajak Yang Disetorkan (dalam juta)	Target Perolehan (dalam Juta)	Jmlh Laporan Kosong
1	WP 1	Jan	56	99	4
2		Feb	45	90	8
3		Mar	68	95	5
4		Apr	50	94	7
5		Mei	61	96	3
6		Jun	53	97	2
7		Jul	65	92	7
8		Agus	50	93	0
9		Sept	43	95	1
10		Okt	47	91	8
11		Nop	64	93	6
...	...	...	...	...	...
16		Apr	55	94	7
17		Mei	43	92	4
18		Jun	65	96	5
19	WP 2	Jan	56	97	1
20		Feb	45	98	7
...	...	...	...	...	...
25		Jul	55	93	5
26		Agus	43	95	2
...	...	...	...	...	...
...	...	...	...	...	...
108	WP 6	Jun	45	98	7

From data found will be classified by the method of naïve bayes into several classes, where this class is determined based on the value of the specified probability, as for in-class

division of WP: very obedient, observant enough, and less obedient

The accuracy of the results obtained from calculations using the naïve bayes classification metod is:

Table 5. Taxpayer Classification

No	Classifica tion Status	Taxes are Paid in	Earnings Target	Empty Report
1	Very Obedient	68	95	5
2	Quite Obedient	65	96	5
3	Less Obedient	47	91	8

Table 5 is the result of a calculation of data classification of taxpayers using Naïve Bayes Classification. Based on the above data, the following graph of the results of the classification with the Naïve Bayes algorithm. The data shows that the WP 3 is included in the category are very Obedient (ST), 2 in WP Quite Obedient (CT) categories and 1 WP in Less Obedient (KT) categories.

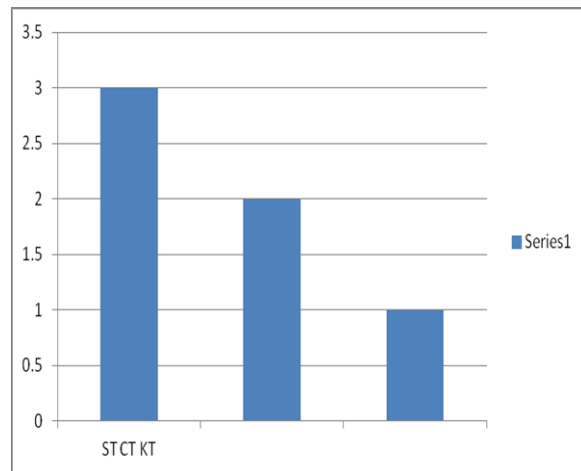


Figure 7. Diagram Of The Classification Results

From the results of the classification shown in Figure 7 can be known that 3 WP is included in the Group of 'very obedient (ST), 2 WP included in the groups 'quite obedient' (CT), and 1 WP included in the group of 'less obedient' (KT).

### CONCLUSION

Based on the results of research and discussion has been done, it can be concluded that the classification of the Naïve Bayes algorithm can be used as one of the methods for classification of the level of observance of taxpayers, support the development of Pasuruan City. Based on the results of the measurement, the test results show with accuracy of 50% i.e. 3 of taxpayers is the very obedient to pay taxes.

As a suggestion, you should research the top system can be implemented in other areas that have potential for tourism, so as to assist regional development and maximize the potential that exists. In addition to the naïve bayes method please use the provided data for comparison with other classification methods.

### ACKNOWLEDGEMENT

Thanks to Universitas Bhayangkara Surabaya and the Directorate of Research and Community Services (DRPM) Kemristekdikti as funders of research with the research contract number 010/SP2H/K2/KM/2017.

### REFERENCES

Adityo, R. D., & Krisdiyono, F. (2015). Sistem Informasi Pelaporan Pajak Hotel Secara Online Pada Dinas Pendapatan Kota

Pasuruan (Jawa Timur). *Prosiding Seminar Nasional Teknologi dan Informatika (SNATIF)*. Kudus: Universitas Muria Kudus.

Bustami, B. (2013). Penerapan Algoritma Naïve Bayes Untuk Mengklasifikasi Data Nasabah Asuransi. *TECHSI-Jurnal Teknik Informatika*, 3(2).

Fitriani, I. R. (2014). Peningkatan Metode Naive Bayes Classification untuk Penentuan Tingkat Keganasan Kanker Payudara Menggunakan Particle Swarm Optimization. (Graduate Thesis). Retrieved from <http://mahasiswa.dinus.ac.id/docs/skripsi/jurnal/13546.pdf>

Han, J., Pei, J., & Kamber, M. (2011). *Data mining: concepts and techniques*. Elsevier.

Jatmika, W. (2009). Deteksi Kanker Payudara Menggunakan Ekstraksi Fitur Statistical Pada Citra Mammogram Berbasis Jaringan Syaraf Tiruan. (Graduate Thesis). Universitas Muria Kudus, Kudus.

Prasetyo, E. (2014). *Data Mining Mengolah Data Menjadi Informasi Menggunakan Matlab*. Yogyakarta: Andi Offset.

Ridwan, M., Suyono, H., & Sarosa, M. (2013). Penerapan Data Mining Untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier. *Jurnal EECCIS*, 7(1), 59-64.

Santosa, B. (2007). *Data Mining Teknik Pe-Manfaatan Data untuk Keperluan Bisnis*. Yogyakarta: Graha Ilmu.

Yuda, N. S. (2014). Data Mining Menggunakan Algoritma Naïve Bayes Untuk Klasifikasi Kelulusan Mahasiswa Universitas Dian Nuswantoro (Studi Kasus: Fakultas Ilmu Komputer Angkatan 2009). *Skripsi, Fakultas Ilmu Komputer, Universitas Dian Nuswantoro, Semarang*.

