IAES International Journal of Artificial Intelligence (IJ-AI)
Vol. 10, No. 2, June 2021, pp. 414~420
ISSN: 2252-8938, DOI: 10.11591/ijai.v10.i2.pp414-420

Measure the effectiveness of information systems with the naïve bayes classifier method

Agung Triayudi[1], Sumiati[2], Saleh Dwiyatno[3], Dentik Karyaningsih[4], Susilawati[5]
[1]Department of Informatic and Communication Technology, Universitas Nasional, Indonesia
[2,4]Department of Informatic, Universitas Serang Raya, Indonesia
[3]Department of Computer Systems Engineering, Universitas Serang Raya, Indonesia
[5]Department of Information System, Universitas Mathla'ul Anwar, Indonesia

Article Info

Article history:

Received Feb 17, 2020
Revised Jan 29, 2021
Accepted Mar 10, 2021

Keywords:

Classification
Data training
Industrial revolution 4.0
Information systems
Naive bayes classifier

ABSTRACT

Technological advances at this time are developing very fast, information systems became the frontline in technological advancements, the need for information systems to support jobs is increasingly high. However, its implementation for users does not have a significant impact, so that it needs to be reviewed and re-evaluated in the use of the information system built. The naive bayes classifier method can provide "effective" and "ineffective" conclusions and is used as material for evaluation and improvement. The purpose of this study is to contribute to measuring the effectiveness of the information system, to solve problems with the naïve bayes classifier method approach which has advantages in the process of classifying data and predicting data. From the test results three times, training has been conducted using 100 data, accuracy value of 84.82% and error 15.18%.

Corresponding Author:

Agung Triayudi
Department of Informatic and Communication Technology
Universitas Nasional
Jl. Sawo Manila, RT.14/RW.3, Ps. Minggu, Kec. Ps. Minggu, Kota Jakarta Selatan, Jakarta, Indonesia
Email: agungtriayudi@civitas.unas.ac.id

1. INTRODUCTION

The rapid development of information technology affects the progress of a company in optimizing performance by creating various information systems [1], [2]. Information system aims to improve service quality to the efficiency of work processes starting from presenting information to support operational functions to management functions [3]. Changes in the process of completing work from a manual system to a modern system where all activities carried out based on electronics known as e-office certainly have many disadvantages in its application. Therefore, an evaluation is needed to measure the effectiveness of information systems in a company. It can be measured using a decision support system [4], [5].

Decision support systems are defined as a system to support the management levels of an organization to make decisions in situations of semi-structured problems [6], [7]. This system is not intended to automate decision making but provides interactive tools that support decision making processes [8], [9]. Naïve bayes is an appropriate method used to measure the effectiveness of information systems, and this method is popularly used in classification techniques [10]. In the research, naive bayes develops very similarly to the multinomial naïve bayes tree method, where one of the data mining techniques for the classification of raw data from customers [11], [12]. The results showed that the multinomial naïve bayes tree (MNBTree) approach had an accuracy rate of 16.26% with 145 features, while the accuracy rate with all 1665 feature datasets was 73.15% [13], [14].

In a previous study, produce a decision support system application with an accuracy level of 81.18% [15]. Therefore, this method was chosen because it was considered simple but produced accurate results [16]. In other studies discuss the sentiment analysis that has been raised in recent scientific studies. The method used is lexicon based and hybrid which is a combination of lexicon based techniques and machine learning. In this study, using the hadoop software framework using the naive bayes and naive bayes complement algorithms methods. Experiments conducted with 8 million reviews were classified as positive, negative and neutral and used various training dataset sizes and around. The performance of the algorithm is measured according to the criteria of precision, accuracy, size-F and calls [17], [18].

In other research on the effectiveness of information systems, the calculation model and evaluation of system effectiveness are based on the analytic hierarchy process (AHP), where the system architecture can be evaluated to support the characteristics of the dataset. [19], [20]. The purpose of this study is to develop a decision support system to analyze the effectiveness of the implementation of information systems using a web-based naïve bayes classifier that can help companies to evaluate and determine the appropriate follow-up. This study uses 8 assessment attributes which consist of ease of use of the system, benefits of the system, speed of completion of work, improvement of work effectiveness, the information generated, speed of information presentation, improvement of service quality, the suitability of the system with needs, and 1 attribute conclusion namely conclusions. From these attributes predictions will be generated in the form of conclusions about the effectiveness of information systems based on previous assessment data.

2. RESEARCH METHOD

The stages of research that will be carried out in making a decision support system to assess information systems using the naïve bayes classifier method are shown in Figure 1.

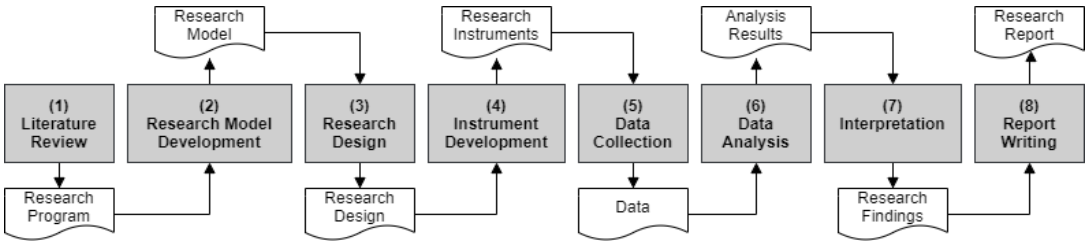

Figure 1. Phase of research

2.1. System requirements analysis

1. Data

   The data needed is questionnaire data about the information system that has been filled by the user.
2. Hardware
3. Software

   The software used for the design of the system include:

a. XAMPP version 3.2.2
b. Notepad++

2.2. Data collection

Data collected in the form of primary data that is a data-assessment questionnaire obtained directly from the data source. Besides, used some literature that was secondary data.

2.2.1. Primary data

Primary data is taken in the form of 100 questionnaire data consisting of 8 assessment criteria written as K1 to K8. Each criterion has 4 members, namely value. Each value has a level from strongly agree (4) to strongly disagree (1). The results of the assessment produce conclusions consisting of effective and Ineffective. Effective results are obtained based on the average rating of 3.00-4.00. Effective results are based on an average rating of <3.00. Effective results are based on an average rating of <3.00. 3 samples of 100 primary data taken are shown in Table 1.

Table 1. Sample of primary data

Name	K1	K2	K3	K4	K5	K6	K7	K8	Conclusion
Sangapan	3	3	3	3	3	3	4	4	Effective
Ratna Kusuma	3	4	3	3	4	4	4	3	Effective
Name	K1	K2	K3	K4	K5	K6	K7	K8	Conclusion
Suri Mulyani	2	3	3	2	3	3	3	3	Ineffective

2.2.2. Secondary data

Literature studies are carried out by studying problems related to the object under study, and getting appropriate references in applying a method, for example by studying several books, journals, and articles related to the problem under study.

2.3. Naïve bayes classifier

Naïve bayes is a classification method based on teorama bayes [21]. The naïve bayes algorithm, which assumes that each variable is independent, independent from each other, has no relationship or correlation that can affect the results [22]. Predictions based on the bayes theorem general formula in (1) [23]:

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)} \quad (1)$$

Once the data is collected, carried out the data processing through data cleansing, data integration, data selection, data transformation [24], and the establishment of the dataset that will be used as training data and testing data [25], [26].

2.4. Analysis of results and conclusions

The final stage is to analyze the accuracy of the results in assessing the effectiveness of the information system, hereinafter referred to as the conclusion.

3. RESULTS AND ANALYSIS

3.1. Data preparation

At this stage, 100 questionnaires were used as training data. Known class of "Effective" as much as 95 data and class "Not Effective" by 5 data. The data consists of eight attributes and one attribute ranking member and the probabilities are shown in Table 2.

Table 2. Table frequencies *(continue)*

No.	Attribute	Member	Frequency
1.	Ease of use of the system	Very easy to use	60
		Easy to use	39
		Difficult to use	1
		Very difficult to use	0
2.	Benefits of the system	Strongly agree	72
		Agree	28
		Disagree	0
		Strongly disagree	0
3.	Speed of completion of work	Very fast	42
		Fast	55
		Not fast	3
		Very fast	0
4.	Increasing the effectiveness of the work	Very true	46
		Right	51
		Not true	3
		Very true	0
5.	Information generated	Strongly agree	57
		Agree	40
		Disagree	2
		Strongly disagree	0
6.	Speed of presenting information	Strongly agree	41
		Agree	57
		Disagree	2
		Strongly disagree	0

Table 2. Table frequencies

No.	Attribute	Member	Frequency
7.	Improved quality of service	Strongly agree	62
		Agree	38
		Disagree	0
		Strongly disagree	0
8.	The suitability of the system to the needs of	Very appropriate	42
		Corresponding	56
		It is not in accordance with	2
		Very appropriate	0
9.	Conclusion	Effective	95
		Ineffective	5

3.2. System testing

System testing is done in two ways, namely by looking at the accuracy of the system and comparing the calculations produced by the system with manual calculations.

1. Test the accuracy of the system

System accuracy testing is done to find out how the naïve bayes classifier method works to provide accuracy and error values [17]. The test was conducted three times on 100 training data with different testing data.

The first test with 100 training data and 25 testing data.
accuracy value = (21/25) *100% = 84%
Error value = (4/25) *100% = 16%
The second test with 100 training data and 40 testing data.
accuracy value = (34/40) *100% = 85%
Error value = (6/40) *100% = 15%
The third test with 100 training data and 55 testing data.
accuracy value = (47/55) *100% = 85.45%
Error value = (8/55) *100% = 14.55%

2. Manual calculations

The second test is done by comparing the results of the system calculation with the results of manual calculations shown in Figure 2.

No	Nama	K 1	K 2	K 3	K 4	K 5	K 6	K 7	K 8
1	Admin	4	4	3	3	3	3	3	4
2	Dinda Audilla	2	2	3	2	1	2	3	2
3	Dita Pramesti	4	3	4	3	4	3	4	4
4	Elis Muhlishoh	4	4	3	3	3	3	4	3

Figure 2. Examples of data test

The stages of testing the manual calculation sample data in Table 3, which are obtained from the primary data sample, are carried out by:

a. Defining variables

Table 3. Definition of variables

X	Value 1	Value 2	Value 3	Value 4
K1	4	2	4	4
K2	4	2	3	4
K3	3	3	4	3
K4	3	2	3	3
K5	3	1	4	3
K6	3	2	3	3
K7	3	3	4	4
K8	4	2	4	3

b. Defining the prior probability

Prior probability that consists of classes effective (H1) and ineffective (H0) based on Table 2 are defined as (2) and (3):

$$P\ (H1)\ /\ P\ (X) = \ 95/100 = \ 0.95 \quad (2)$$

$$P\ (H0)/\ P\ (X) = \ 5/100 = \ 0.05 \quad (3)$$

c. The posterior probability calculation

The criteria for calculating the probability (X) per class (H) in Table 4.

Table 4. Posterior probability

		Value 1		Value 2		Value 3		Value 4	
		H_1	H_0	H_1	H_0	H_1	H_0	H_1	H_0
POSTERIOR PROBABILITIES	K1	0.6315	0	0.0000	0.2	0.6316	0	0.6316	0
	K2	0.7578	0	0.0000	0.2	0.3684	0.8	0.6316	0
	K3	0.5473	0.6	0.3684	0.8	0.6316	0	0.3684	0.8
	K4	0.5157	0.4	0.0000	0.2	0.3684	0.8	0.3684	0.8
	K5	0.3789	0.8	0.0000	0	0.6316	0	0.3684	0.8
	K6	0.5684	0.6	0.0000	0.2	0.3684	0.8	0.3684	0.8
	K7	0.3473	1	0.3684	0.8	0.6316	0	0.6316	0
	K8	0.4421	0	0.0000	0.2	0.6316	0	0.3684	0.8

d. Final probability calculation

The final calculated probability by multiplying the previous probability by the posterior probability of each class shown in Table 5.

Table 5. Final probability

	Value 1	Value 2	Value 3	Value 4
$P (X \mid H_1)$	0.004466368	0	0.00502541	0.001710036
$P (H_1) / P (X)$	0.95	0.95	0.95	0.95
$P (H_1 \mid X)$	0.00424305	0	0.00477414	0.001624534
$P (X \mid H_0)$	0	0	0	0
$P (H_0) / P (X)$	0.05	0.05	0.05	0.05
$P (H_0 \mid X)$	0	0	0	0

From the calculations in Table 6, the P-value (H1 | X) will be compared with P (H0 | X). If the value of P (H1 | X) > P (H0 | X), it can be concluded that the data entered is classified into the Effective class. However, if the value of P (H1 | X) < P (H0 | X), it can be concluded that the data entered is classified into Ineffective classes. The results of comparison of the four assessment samples that have been included as shown in Table 6. This result will be compared with the system prediction results as shown in Table 7. The system prediction results in Table 6 show that the conclusions from the prediction results are the same as manual calculations.

Table 6. Comparison of the calculation results

	Comparison	Conclusion
Value 1	$P (H_1 \mid X) > P (H_0 \mid X)$	EFFECTIVE
Value 2	$P (H_1 \mid X) < P (H_0 \mid X)$	INEFFECTIVE
Value 3	$P (H_1 \mid X) > P (H_0 \mid X)$	EFFECTIVE
Value 4	$P (H_1 \mid X) > P (H_0 \mid X)$	EFFECTIVE

Table 7. System calculation results

No	Name	K1	K2	K3	K4	K5	K6	K7	K8	Conclusion
1	Budi	4	4	3	3	3	3	3	4	EFFECTIVE
2	Andi	2	2	3	2	1	2	3	2	INEFFECTIVE
3	Rita	4	3	4	3	4	3	4	4	EFFECTIVE
4	Dinda	4	3	3	3	3	3	3	3	EFFECTIVE

4. CONCLUSION

The results of the system accuracy test conducted three times using 100 training data resulted in an accuracy value of 84% for 25 test data, 85% for 40 test data, and a maximum value of 85.45% for 55 test data. With a relatively stable value of accuracy and error values and the results of system calculations with relatively similar manual calculations, the decision support system is right to be applied to the naïve bayes classifier method approach to assess the effectiveness of information systems with an average accuracy value of 84.82%.

REFERENCES

[1] Triayudi, Agung, Iskandar Fitri, "Comparison of parameter-free agglomerative hierarchical clustering methods," *ICIC Express Letters*, vol. 12, no. 10, pp. 973-980, 2018, doi: 10.24507/icicel.12.10.973.

[2] Triayudi A, Fitri I. "ALG Clustering To Analyze The Behavioral Patterns Of Online Learning Student," *Journal of Theoretical & Applied Information Technology*. Aug 31, vol. 96, no. 16, 2018.

[3] Triayudi, Agung, Iskandar Fitri, "A new agglomerative hierarchical clustering to model student activity in online learning," *TELKOMNIKA (Telecommunication, Computing, Electronics and Control)*, vol. 17, no. 3, pp. 1226-1235, 2019, doi: 10.12928/telkomnika.v17i3.9425.

[4] Putra, Syopiansyah Jaya, Abd Ahlan, and Mira Kartiwi, "A Coherent Framework for Understanding the Success of an Information System Project," *TELKOMNIKA (Telecommunication, Computing, Electronics and Control)*, vol. 14, no. 1, 2016, doi: 10.12928/telkomnika.v14i1.2711.

[5] A'ang Subiyakto, Abdul Rahman, Mira Kartiwi, and Husni Teja Sukmana, "Influences of the input factors towards the success of an information system project," *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, vol. 13, no. 2, pp. 686-693, 2015, doi: 10.12928/telkomnika.v13i2.1323.

[6] Wibowo, Adhi, and Marti Widya Sari, "Measuring Enterprise Resource Planning (ERP) Systems Effectiveness in Indonesia," *TELKOMNIKA (Telecommunication, Computing, Electronics and Control)*, vol. 16, no. 1, pp. 343-351, 2018, doi: 10.12928/telkomnika.v16i1.5895.

[7] Xie, Xiaoliang, Zhaowen Li, Pengfei Zhang, and Gangqiang Zhang, "Information structures and uncertainty measures in an incomplete probabilistic set-valued information system," *IEEE Access* 7, pp. 27501-27514, 2019, doi: 10.1109/ACCESS.2019.2897752.

[8] Komalasari, Nia, Dina Fitria Murad, Dine Agustine, Muhamad Irsan, Johan Budiman, and Erick Fernando, "Effect of Education, Performance, Position and Information Technology Competency of Information Systems to Performance of Information System," In *2018 International Seminar on Research of Information Technology and Intelligent Systems (ISRITI)*, pp. 221-226. IEEE, 2018, doi:10.1109/ISRITI.2018.8864437.

[9] S. A. B. M. Yusof and M. H. Abdulraheem, "Real Factors which Impact on Decision Making in the E-government," *2015 6th International Conference on Intelligent Systems, Modelling and Simulation*, Kuala Lumpur, pp.252-255, 2015, doi: 10.1109/ISMS.2015.52.

[10] Tripathi, Akarshita, Saumya Yadav, and Rajiv Rajan, "Naive Bayes Classification Model for the Student Performance Prediction," In *2019 2nd International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICICT)*, vol. 1, pp. 1548-1553. IEEE, 2019, doi:10.1109/ICICICT46008.2019.8993237

[11] K P Tripathi, "Decision Support System is a Tool for Making Better Decisions in the Organization," *Indian Journal of Computer Science and Engineering (IJCSE)*, vol. 2, no. 1, pp. 112-117.

[12] Ada, Şükrü & Ghaffarzadeh, Mohsen, "Decision Making based on Management Information System and Decision Support System," *European Researcher*, vol. 93, no. 4, pp. 260-269, March 2015, doi: 10.13187/er.2015.93.260.

[13] Susanti, Aisah Rini, Taufik Djatna, and Wisnu Ananta Kusuma, "Twitter's Sentiment Analysis on Gsm Services using Multinomial Naïve Bayes," *TELKOMNIKA (Telecommunication, Computing, Electronics and Control)*, vol. 15, no. 3, 2017, doi: 10.12928/telkomnika.v15i3.4284.

[14] Hidayatullah, Ahmad Fathan, Chanifah Indah Ratnasari, and Satrio Wisnugroho. "Analysis of Stemming Influence on Indonesian Tweet Classification." *TELKOMNIKA (Telecommunication, Computing, Electronics and Control)*, vol. 14, no. 2, p. 665, 2016, doi: 10.12928/telkomnika.v14i2.3113

[15] S. Sankaranarayanan, T. Pramananda Perumal, "Analysis of Naïve Bayes Classification for Diabetes Mellitus," *International Journal of Computer Sciences and Engineering*, vol. 6, no. 12, pp. 520-524, December 2018.

[16] Kesumawati, Ayundyah; Waikabu, Din, "Implementation Naïve Bayes Algorithm for Student Classification Based on Graduation Status," *International Journal of Applied Business and Information Systems*, vol. 1, no. 2, pp. 6-12, September 2017.

[17] Razaque, F., Soomro, N., Shaikh, S.A., Soomro, S., Samo, J.A., Kumar, N., Dharejo, H., "Using Naïve Bayes algorithm to students' bachelor academic performances analysis," *4th IEEE International Conference of Applied Science and Technology, ICETAS*, 2017. DOI: 10.1109/ICETAS.2017.8277884.

[18] Seref, Berna, and Erkan Bostanci, "Sentiment Analysis using Naive Bayes and Complement Naive Bayes Classifier Algorithms on Hadoop Framework," In *2018 2nd International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT)*, pp. 1-7. IEEE, 2018, doi:10.1109/ISMSIT.2018.8567243.

[19] Jadon, Priyanshu, Deepshikha Bhatia, and Durgesh Kumar Mishra. "A BigData approach for sentiment analysis of twitter data using Naive Bayes and SVM Algorithm." In *2019 Sixteenth International Conference on Wireless and Optical Communication Networks (WOCN)*, pp. 1-6. IEEE, 2019. DOI:10.1109/WOCN45266.2019.8995109.

[20] Xia, Xiaokai, Kai Qu, Jing Shi, Zhiqiang Fan, and Luo Xu, "The construction of effectiveness evaluation model based on system architecture," In *2017 IEEE International Systems Engineering Symposium (ISSE)*, pp. 1-4. IEEE, 2017, doi: 10.1109/SysEng.2017.8088319.

[21] Kapur, Bhrigu; Ahluwalia, Nakin; R, Sathyaraj, "Comparative Study on Marks Prediction using Data Mining and Classification Algorithms," *International Journal of Advanced Research in Computer Science*, vol. 8, no. 3, pp. 171-182, March-April 2017.

[22] Peling, Ida Bagus Adisimakrisna; Arnaman, I Nyoman; Arthawan, I Putu Arich; Janardana, IGN, "Implementation of Data Mining to Predict Period of Students Study Using Naive Bayes Algorithm," *International Journal of Engineering and Emerging Technology*, vol. 2, no. 1, pp. 53-57, January-June 2017.

[23] Harahap, Fitriana, Ahir Yugo Nugroho Harahap, Evri Ekadiansyah, Rita Novita Sari, Robiatul Adawiyah, and Charles Bronson Harahap, "Implementation of Naïve Bayes Classification Method for Predicting Purchase," In *2018 6th International Conference on Cyber and IT Service Management (CITSM)*, pp. 1-5. IEEE, 2018. doi: 10.1109/CITSM.2018.8674324.

[24] Ibrahim, Mohd Naim Mohd, and Mohd Zaliman Mohd Yusoff, "The impact of different training data set on the accuracy of sentiment classification of Naïve Bayes technique," In *2017 IEEE Conference on Open Systems (ICOS)*, pp. 17-20. IEEE, 2017, doi:10.1109/ICOS.2017.8280267.

[25] Chen, Jiangning, Zhibo Dai, Juntao Duan, Heinrich Matzinger, and Ionel Popescu, "Naive Bayes with Correlation Factor for Text Classification Problem," *arXiv preprint arXiv:1905.06115*, 2019.

[26] Rosalina, and Vidila, "Mobile Customer Relationship Management (m-CRM) Application Development in MSMEs Indonesia," In *Journal of Physics: Conference Series*, vol. 1179, no. 1, p. 012090. IOP Publishing, 2019. doi: 10.1088/1742-6596/1179/1/012090.