Analysis of the effect of the lecturer satisfaction with the Naive Bayes Data Mining technique on institutional performance

Siti Aisyah¹*, Preddy Marpaung², Wiwin Aprinai³, Komda Saharja⁴, I Made Yuda Suryawan⁴, Bektı Tufıq Ari Nugroho⁵,⁶, Amin Nurbaedi⁷, Hasrul Azwar Hasibuan⁸, Bernadetha Nadeak⁹, Ahmad Tohir¹⁰

¹Politeknik Negeri Media Kreatif, Indonesia  
²STMIK Pelita Nusantara, Indonesia 
³Departemen Pendidikan Matematika, Universitas Al-Muslim Bireun, Indonesia 
⁴Teknologi Indonesia University, Denpasar, Bali, Indonesia 
⁵Institut Agama Islam Negeri Salatiga, Indonesia  
⁶Universitas Islam Negeri Walisongo Semarang, Indonesia  
⁷PAI, UNU Surakarta, Indonesia 
⁸Universitas Pembangunan Panca Budi, Medan, Indonesia  
⁹Universitas Kristen Indonesia, Jakarta, Indonesia 
¹⁰STKIP Al Islam Tunas Bangsa, Bandar Lampung, Indonesia

Corresponding Email: dedek.sitiaisyah@gmail.com

Abstract. The study aimed to analyze the effect on institutional performance of lecturer satisfaction with data extraction techniques. The solution is the technique of Naive Bayes, where data is obtained through interviews and questionnaires conducted in one of the private institutions in the north-sumatra of Medan. The evaluation criteria are readiness, compassion, reliability and accountability. The tests indicate that the level of accuracy is 85.48% with 81.08% precision, and 93.75% recall value. The Naive Bayes method can also be recommended to predict the degree of satisfaction of the lecturer with institutional performance based on the results of tests using fast miner software.

Keywords: Datamining, Naive Bayes, lecturer satisfaction, institutional performance, prediction, north sumatra, Indonesia.

1. Introduction

One important factor in the comparison of service performance received and anticipated by customers is customer satisfaction [1]. Satisfaction is a person's attitude that can adapt well to work and social relations with workers and others. The degree of satisfaction that is perceived to be a requirement for the fulfillment of desires and expectations, is also derived from perceived satisfaction. This means that the satisfactory achievements of each available service's academic community can be an interlinked evaluation to find out whether the services provided are good or bad [2]. The observations of the satisfaction of lecturers with the institutional performance focus on assessing the quality of services offered to lecturers who conduct or conduct higher education tri-darma. On this basis, predicting is one way to determine the percentage of satisfaction of the lecturer with the institution's performance.

Intelligence artificial is a science which imitates human intelligence and is deployed in computers [3]–[7]. Including Decision support systems [8]–[10], Data mining [11]–[13] and the fuzzy logic [14],
[15], Expert Systems [16] and Artificial Neural Networks [17]–[19], Artificial intelligence has different types of algorithms that can be used in everyday life. The informatics technique that can be classified is data analysis. The Naïve Bayes algorithm is the data mining method used in this study with the classification technique. In previous research the use of the Naïve Bayes predicting algorithm was widely used. One research [20] on the Naïve Bayes method in the prediction of film ratings. The results show that the low-grade distribution model is 0.707 while the high-grade value is 0.293, with an accuracy of 55.80% by 32.41% and 46.7%. The Naïve Bayes algorithm statistical calculations can calculate the likelihood that the old case will be identical with the new. Naïve Bayes needed training data for decisions to reduce computer complexity to simple multiplications at the beginning of the calculation. On the basis of these problems, we hope to measure the progress of the institution to provide services for lecturers through the Naïve Bayes algorithm on satisfaction of lecturers with institutional performance.

2. Methodology

Data obtained by conducting observations in the form of interviews and giving questionnaires from interviews to lecturers at one of the private institutions in the city of Medan, North Sumatra. At this stage, as shown in Figure 1, there are several steps.

![Naïve Bayes Algorithm Flowchart](image-url)

**Figure 1. Naïve Bayes Algorithm Flowchart**

Based on the flow chart in Figure 1, the following can be described:

a) Collecting and entering questionnaire data from the lecturer concerned at a private institution in the city of Medan, North Sumatra.

b) Testing of pre-existing training data.

c) Assess the performance of the algorithm. In the testing process, the performance of the algorithm is tested using a data test set where the data set with the training data set consists of different data.

d) In addition, the implementation of previous data using the naive Bayes algorithm in order to produce accurate accuracy, precision and recall.
e) The system then issues the output in the form of the percentage of satisfaction of the lecturer with the performance of the institution.

3. Results and Discussion

In this case, problem-solving is carried out by conducting an assessment of the satisfaction of lecturers on institutional performance by conducting interviews and distributing questionnaires in the form of several questions to lecturers on institutional performance. The solution provided uses the Naïve Bayes method for data mining. The criteria used for the assessment are Readiness, Empathy, Reliability and Responsibility. The results of the analysis were calculated using software and Microsoft Excel.

| No | Readiness | Empathy | Reliability | Responsibility |
|----|------------|---------|-------------|----------------|
| 1  | 4          | 5       | 7           | 5              |
| 2  | 6          | 6       | 4           | 3              |
| 3  | 6          | 5       | 7           | 5              |
| 4  | 7          | 6       | 4           | 6              |
| 5  | 5          | 7       | 5           | 6              |
| 6  | 6          | 5       | 5           | 6              |
| 7  | 7          | 6       | 4           | 7              |
| 8  | 7          | 4       | 5           | 6              |
| 9  | 5          | 6       | 7           | 4              |
| 10 | 4          | 5       | 6           | 5              |
| 11 | 5          | 7       | 6           | 7              |
| 12 | 6          | 5       | 5           | 7              |
| 13 | 6          | 5       | 7           | 5              |
| 14 | 7          | 6       | 5           | 6              |
| 15 | 6          | 4       | 6           | 5              |
| 16 | 5          | 6       | 5           | 7              |
| 17 | 4          | 7       | 7           | 6              |
| 18 | 5          | 6       | 6           | 5              |
| 19 | 3          | 5       | 5           | 7              |
| 20 | 5          | 7       | 7           | 6              |
| 21 | 6          | 5       | 5           | 7              |
| 22 | 7          | 7       | 6           | 6              |
| 23 | 4          | 5       | 7           | 6              |
| 24 | 5          | 6       | 7           | 5              |
| 25 | 5          | 4       | 5           | 6              |
| 26 | 6          | 6       | 4           | 7              |
| 27 | 7          | 7       | 4           | 6              |
| 28 | 7          | 6       | 5           | 5              |
| 29 | 7          | 7       | 6           | 4              |
| 30 | 6          | 6       | 7           | 4              |

Table 1. Average value of each attribute

The data collected is questionnaire data which is distributed to. The sample data per parameter is grouped into two value intervals as in the following table.

| No | Value Interval | Group |
|----|----------------|-------|
| 1  | 3 – 5          | Low   |
| 2  | 6 -7           | High  |

Table 2. Questionnaire Value Group Per Parameter

Classify the results into two value intervals as in the following table:
Table 3. Classification of Questionnaire Value Results

| No | Value Interval | Classification |
|----|----------------|----------------|
| 1  | 3 – 5          | Not satisfied  |
| 2  | 6 -7           | Satisfied      |

Initial data were obtained using Microsoft Excel to produce the data sets in the following table:

Table 4. Training data

| No | Readiness | Empathy | Reliability | Responsibility | Result         |
|----|-----------|---------|-------------|----------------|----------------|
| 1  | High      | High    | Low         | Low            | Not satisfied  |
| 2  | High      | Low     | High        | Low            | Not satisfied  |
| 3  | High      | High    | High        | High           | Satisfied      |
| 4  | High      | High    | High        | Low            | Satisfied      |
| 5  | High      | High    | Low         | High           | Satisfied      |
| 6  | High      | High    | Low         | High           | Not satisfied  |
| 7  | High      | High    | Low         | Low            | Not satisfied  |
| 8  | High      | High    | High        | High           | Satisfied      |
| 9  | High      | High    | High        | High           | Satisfied      |
| 10 | Low       | Low     | High        | Low            | Not satisfied  |
| 11 | High      | High    | High        | High           | Satisfied      |
| 12 | High      | High    | Low         | High           | Satisfied      |
| 13 | High      | Low     | High        | Low            | Satisfied      |
| 14 | High      | Low     | Low         | High           | Satisfied      |
| 15 | Low       | Low     | High        | Low            | Not satisfied  |
| 16 | Low       | High    | High        | High           | Satisfied      |
| 17 | Low       | High    | Low         | High           | Satisfied      |
| 18 | Low       | High    | Low         | High           | Not satisfied  |
| 19 | Low       | High    | Low         | Low            | Satisfied      |
| 20 | Low       | Low     | High        | Low            | Not satisfied  |
| 21 | High      | High    | High        | Low            | Satisfied      |
| 22 | High      | High    | High        | Low            | Satisfied      |
| 23 | High      | Low     | High        | Low            | Not satisfied  |
| 24 | High      | Low     | High        | Low            | Satisfied      |
| 25 | Low       | High    | Low         | Low            | Not satisfied  |
| 26 | High      | Low     | High        | Low            | Satisfied      |
| 27 | High      | High    | High        | High           | Satisfied      |
| 28 | High      | Low     | Low         | High           | Not satisfied  |
| 29 | Low       | High    | High        | High           | Satisfied      |
| 30 | High      | Low     | Low         | High           | Not satisfied  |

a) Finding the Mean Value
Satisfied Class = 24
Class not satisfied = 6

\[ \mu(\text{Satisfied}(A1)) = \frac{(5+6+6+6+6+4+5+6+5+6+6+5+6+6+5+6+6+5+7+5+6+7+6)}{24} \]
\[ = \frac{136}{24} \]
\[ = 5.6666 \]

\[ \mu(\text{Satisfied}(A2)) = \frac{(4+6+6+6+6+7+6+6+6+5+6+6+5+6+6+5+6+6+5+7+6+6+6+5+5+5)}{24} \]
\[ = \frac{139}{24} \]
\[ = 5.7916 \]

\[ \mu(\text{Satisfied}(A3)) = \frac{(7+6+7+6+6+7+5+6+7+7+6+6+5+6+7+6+7+6+7+6+6+7+7+7)}{24} \]
\[ = \frac{153}{24} \]
\[ = 6.3750 \]

\[ \mu(\text{Satisfied}(A4)) = \frac{(5+7+7+6+7+6+7+5+6+6+7+6+7+6+7+6+4+7+4+7+4+7+6+6+7)}{24} \]
\[ = \frac{149}{24} \]
\[ = 6.2083 \]
\[ \mu(\text{Not satisfied})(A1) = \frac{(5+4+5+6+5+6)}{6} = \frac{31}{6} = 5.1666 \]
\[ \mu(\text{Not satisfied})(A2) = \frac{(6+6+5+5+7+5)}{6} = \frac{34}{6} = 5.6666 \]
\[ \mu(\text{Not satisfied})(A3) = \frac{(4+7+6+5+5+5)}{6} = \frac{33}{6} = 5.500 \]
\[ \mu(\text{Not satisfied})(A4) = \frac{(5+5+6+5+5+6)}{6} = \frac{32}{6} = 5.3333 \]

b) Standard Deviation

\[ \sigma(\text{Satisfied})(A1) = \sqrt{ \frac{(5-5.6666)^2 + (4-5.6666)^2 + (5-5.6666)^2 + (6-5.6666)^2 + (5-5.6666)^2 + (6-5.6666)^2 + (6-5.6666)^2 + (4-5.6666)^2 + (5-5.6666)^2 + (6-5.6666)^2}{24-1} } \]
\[ = \sqrt{0.4926} = 0.2426 \]

\[ \sigma(\text{Not satisfied})(A2) = \sqrt{ \frac{(5-5.1666)^2 + (4-5.1666)^2 + (5-5.1666)^2 + (6-5.1666)^2 + (5-5.1666)^2 + (6-5.1666)^2}{6-1} } \]
\[ = \sqrt{0.5666} = 0.32103556 \]

| Table 5. Capacity Probability Value |
|------------------------------------|
| **Value** | Satisfied | Not satisfied | Satisfied | Not satisfied | Satisfied | Not satisfied | Satisfied | Not satisfied |
| μMean     | 5.6666    | 5.1666        | 5.7916    | 5.6666        | 6.375     | 5.5          | 6.2083    | 5.3333        |
| σDeviasi^2| 0.4926    | 0.5666        | 0.4033    | 0.6666        | 0.4184    | 1.1          | 0.8676    | 0.2664        |
| σDeviasi  | 0.2426    | 0.32103556    | 0.1626    | 0.4443        | 0.175     | 1.21         | 0.7527    | 0.0709        |

The next step is to calculate the probability for the Testing data feature.

| Table 6. Data Testing |
|-----------------------|
| **No** | Readiness | Empathy | Reliability | Responsibility | Result |
| 1      | High      | High    | Low         | Low            | ?      |
| 2      | High      | Low     | High        | Low            | ?      |
| 3      | High      | High    | High        | ?              | ?      |
| 4      | High      | High    | High        | ?              | ?      |
| 5      | High      | High    | Low         | ?              | ?      |
| 6      | High      | Low     | High        | ?              | ?      |
| 7      | High      | Low     | High        | ?              | ?      |
| 8      | High      | High    | High        | ?              | ?      |
| 9      | High      | High    | High        | ?              | ?      |
| 10     | Low       | High    | High        | Low            | ?      |
| 11     | High      | High    | High        | ?              | ?      |
| 12     | High      | Low     | High        | ?              | ?      |
| 13     | High      | Low     | High        | ?              | ?      |
| 14     | High      | Low     | High        | ?              | ?      |
In addition, RapidMiner is used to process the data and to test it with SplitValidation found in RapidMiner. The results of the data set tests are as follows:

![Figure 2. Split Validation](image1)

![Figure 3. Process Validation, Naïve Bayes Method](image2)

After going through the above process, the result of the performance vector is 83.33% Accuracy.

![Figure 4. Performance Vector](image3)

And the Accuracy Performance values are as follows:

![Figure 5. Value of Accuracy Performance](image4)

Information:

a) The number of predictions Eligible and in fact True Eligible is 9 records
b) The number of Predictions of Unfeasible and in fact correct is 3 records.
c) The number of predictions Eligible and in fact true of Unfeasible is 2 records.
d) The number of predictions of Eligible and in fact true of Unreasonable is 16 records
The Accuracy Performance Value shows in the picture that the prediction of the eligible value is 9, with an accuracy rate of 81.82.0%; the incorrect forecast is 10 and its accuracy is 84.25%. The accuracy achieved is therefore 83.33%.

4. Conclusion

According to the results of tests from the data set provided with readiness, empathy, reliability and responsibility attributes, which are based on the Naive Bayes method, an accuracy level of 85.48% with an precision value of 81.80% and a recall value of 93.75% is achieved. Based on the test finding, a lecturer satisfaction with institutional performance can be predicted using the Naïve Bayes Method. Since the accuracy and the recall value of this research are high, it is possible to combine or compare this research with other classification algorithms.

References

[1] H. Hartono, “Mutu Layanan Dari Bagian Layanan Akademik Terhadap Kepuasan Dosen Di Universitas Xyz,” J. Ris. Manaj. dan Bisnis Fak. Ekon. UNIAT, vol. 2, no. 1, pp. 13–24, 2017.
[2] B. Rismayadi and M. Maemunah, “Pengaruh Motivasi Kerja, Kepemimpinan dan Budaya Organisasi Terhadap Kepuasan Kerja Karyawan serta Dampaknya pada Kinerja Perusahaan (Studi kasus pada PT. Concord Indonesia),” J. Manaj. Bisnis Kreat., vol. 2, no. 1, pp. 124–135, 2016.
[3] P. P. P. A. N. W. F. I. R. H. Zer, Masitha, A. P. Windarto, and A. Wanto, “Analysis of the ELECTRE Method on the Selection of Student Creativity Program Proposals,” J. Phys. Conf. Ser., vol. 1255, no. 1, 2019.
[4] Budiharjo, T. Soemartono, A. P. Windarto, and T. Herawan, “Predicting School Participation in Indonesia using Back-Propagation Algorithm Model,” Int. J. Control Autom., vol. 11, no. 11, pp. 57–68, 2018.
[5] Budiharjo, T. Soemartono, A. P. Windarto, and T. Herawan, “Predicting tuition fee payment problem using backpropagation neural network model,” Int. J. Adv. Sci. Technol., vol. 120, pp. 85–96, 2018.
[6] B. Supriyadi, A. P. Windarto, T. Soemartono, and Mungad, “Classification of natural disaster prone areas in Indonesia using K-means,” Int. J. Grid Distrib. Comput., vol. 11, no. 8, pp. 87–98, 2018.
[7] E. Susanto, Y. Novitasari, W. Rahman, and A. P. O. Amane, “Designing Software to Introduce the Musical Instruments,” in Journal of Physics: Conference Series, 2019, vol. 1364, no. 1.
[8] B. Sinamora, A. Amril, I. W. Utami, K. Shanty, and M. Parwati, “Decision support system using PROMETHEE Algorithm Decision support system using PROMETHEE Algorithm,” IOP Conf. Ser. Mater. Sci. Eng. Pap., 2021.
[9] A. Z. Adnan, M. H. Holili, and B. Indrawan, “Analysis of Decision Support System with Analytical Hierarchy Process Method,” J. Phys. Conf. Ser. Pap., 2021.
[10] M. Darmowiyono, W. Yuliyanto, and K. I. Purnomo, “Application of the Simple Multi Attribute Rating Technique (SMART) Method in the selection of thrush medicine products based on consumers Application of the Simple Multi Attribute Rating Technique (SMART) Method in the selection of thrush medicine prod,” J. Phys. Conf. Ser. Pap., 2021.
[11] H. Jatnika, M. Huda, R. R. Amelia, and M. A. Manuhutu, “Analysis of Data Mining in the Group of Water Pollution Areas using the K-Means Method in Indonesia Analysis of Data Mining in the Group of Water Pollution Areas using the K-Means Method in Indonesia,” J. Phys. Conf. Ser. Pap., 2021.
[12] A. Abbas, P. Prayitno, F. Butarbutar, and N. Nurkim, “Implementation of clustering unsupervised learning using K-Means mapping techniques Implementation of clustering unsupervised learning using K- Means mapping techniques,” IOP Conf. Ser. Mater. Sci. Eng. Pap., 2021.
[13] N. Azis, Y. Ali, R. Subekti, P. Junianto, and L. Diner, “Mapping study using the unsupervised learning clustering approach,” IOP Conf. Ser. Mater. Sci. Eng. Pap., 2021.
[14] A. Mukerji, C. Chatterjee, and N. S. Raghuvanshi, “Flood Forecasting Using ANN, Neuro-Fuzzy, and Neuro-GA Models,” *J. Hydrol. Eng.*, pp. 647–652, 2009.

[15] W. Jin, F. Gong, X. Zeng, and R. Fu, “Classification of clouds in satellite imagery using adaptive fuzzy sparse representation,” *Sensors (Switzerland)*, vol. 16, no. 12, 2016.

[16] R. Rahmadewi *et al.*, “Metode Segmentasi Canny pada Citra Rontgen untuk Klasifikasi Penyakit Paru,” *J. Nas. Tek. Elektro*, vol. 1, no. 2, pp. 140–145, 2013.

[17] B. Febriadi, Z. Zamzami, Y. Yunefri, and A. Wanto, “Bipolar function in backpropagation algorithm in predicting Indonesia’s coal exports by major destination countries,” *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 420, no. 1, p. 012087, 2018.

[18] Sumijan, A. P. Windarto, A. Muhammad, and Budiharjo, “Implementation of Neural Networks in Predicting the Understanding Level of Students Subject,” *Int. J. Softw. Eng. Its Appl.*, vol. 10, no. 10, pp. 189–204, 2016.

[19] A. Wanto *et al.*, “Model of Artificial Neural Networks in Predictions of Corn Productivity in an Effort to Overcome Imports in Indonesia,” *J. Phys. Conf. Ser.*, vol. 1339, no. 1, 2019.

[20] R. Wijayatun and Y. Sulistyo, “Prediksi Rating Film Menggunakan Metode Naive Bayes,” *J. Tek. Elektro*, vol. 8, no. 2, pp. 60–63, 2016.