JURNAL EMACS
e-ISSN: 2686-2573
[Engineering, MAthematics and Computer Science] Vol.8 No.3 September 2026: 293-298 DOI: 10.21512/emacsjournal.v8i3.14305

University Students' Feedback Analysis using Naïve Bayes Classification

Yasi Dani
Computer Science Department, School of Computer Science,
Bina Nusantara University
Jakarta, Indonesia 11480
yasi.dani@binus.ac.id

Correspondence: yasi.dani@binus.ac.id

Abstract – *In recent years, student remarks have emerged as an essential source of information for comparing teaching pleasant, getting to know effectiveness, and educational services in higher training institutions. Since maximum remarks records are unstructured and hard to analyze manually because the system is time-ingesting and vulnerable to subjectivity. This study applies the Naïve Bayes category algorithm to automatically analyze and classify college students' remarks into three sentiment classes: Bad, Neutral, and Good. The dataset consisted of 185 student remarks information collected from diverse academic aspects, such as teaching, path content, examinations, laboratory paintings, library centers, and extracurricular activities. The records were processed through cleansing, preprocessing, labeling, encoding, and splitting levels earlier than version training and evaluation. The version overall performance was measured using accuracy, precision, recall, and F1-score metrics. Experimental results confirmed that the version executed an ordinary accuracy of 0.61, with the exceptional overall performance obtained inside the precise magnificence, reaching a precision of 0.67, recall of 0.82, and F1-score of 0.74. The findings display that Naïve Bayes can help automated sentiment evaluation and assist universities in records-driven academic evaluation and choice-making procedures. The proposed technique presents a sensible and green solution for supporting academic institutions display student perceptions, discover provider weaknesses, and constantly improve the pleasantness of academic control strategically.*

Keywords: *artificial intelligence; feedback analysis; education; Naïve Bayes Classifier*

I. INTRODUCTION

In recent years, researchers have been studying about what relevant information can be obtained and analysed from online user feedback in various fields. In software engineering, researchers investigated feedback from app reviews to get some new requirements about the features (Dąbrowski et al., 2020) or to find bugs in the app (Obaidi et al., 2022). To study cultural differences in app store, (Guzman et al., 2018) analysed feedback given by different cultural background users. In business sector, (Wang et al., 2024) studied the sentiment analysis and measured customer's satisfaction for some products by analysing customer's online reviews. In educational field, (Sharma et al., 2018) studied how to classify students' feedback automatically to get significant information for university improvement.

The quality of a university is not only measured by academic achievements but also by the experiences and perceptions of students regarding various aspects of the campus environment. Students who regularly provide reviews or feedback become one of the important indicators in assessing the quality of educational services, covering academic aspects such as teaching and course materials, as well as non-academic aspects such as facilities, administrative services, and the campus environment in general. A large number of reviews collected periodically become a very valuable source of data for continuous evaluation.

However, there are some challenges in analysing user's feedback since there is a large amount of unstructured feedback data. Manual analysis process of thousands of student comments can be very time-consuming and prone to subjectivity. One potential solution to face this challenge is to implement analysis approach based on artificial intelligence (AI). One of the methods that can be used is automatic classification of student reviews into categories such as positive (good), negative (bad), or neutral. This method provides the

Copyright © 2026

university with a fast and clear overview of student sentiments. Kastrati et al. (Kastrati et al., 2021) investigated student feedback sentiment analysis through natural language processing (NLP) and deep learning methods. Alternative techniques, such as the Naïve Bayes classifier, have been utilized to examine user reviews, encompassing Facebook comments on agri-tech startups (Kewsuwun & Kajornkasirat, 2022), and also student feedback from before and after the COVID-19 pandemic (Mujahid et al., 2021).

Naïve Bayes Classification is a preferred machine learning algorithm for sentiment analysis due to its simplicity and remarkable effectiveness in managing extensive amounts of text data. A variety of research has shown the effectiveness of this algorithm in automatically classifying student feedback (Hitesh et al., 2019), (Kumar et al., 2020), (Dani & Ginting, 2024). It functions on a probabilistic framework, making use of Bayes' Theorem, and can produce pretty specific results even when the enter information includes numerous and unstructured language.

This have a look at ambitions to use the Naïve Bayes algorithm for categorizing student feedback in step with sentiment. The statistics will include surveys or open evaluations collected from college students, that allows you to go through a couple of phases, including records accumulating, preprocessing, category, analysis, version development, and accuracy evaluation. The primary aim of this research is to provide the college strategic insights that may assist in shaping rules, elevating coaching high-quality, and basic improving academic offerings.

This paper is organized as follows: Section 2 examines the Naïve Bayes category set of rules. Section 3 describes the studies strategies used. Section 4 offers the findings and tests, while Section 5 concludes with the very last thoughts.

II. METHODS

2.1 Naïve Bayes Classification

Naïve Bayes category is an essential probabilistic technique in supervised machine gaining knowledge of, mainly employed to address category problems (Hand & Yu, 2001). This technique is based totally on Bayes' theorem, an precept from probability idea that lets in us to revise the probability of an speculation as new proof emerges. The term "naïve" emphasizes the algorithm's notion that the capabilities (or attributes) are unbiased of each different given the conditions. despite the fact that this assumption is not always legitimate in sensible conditions, the algorithm normally demonstrates abruptly right performance (Rennie et al., 2003), (Zhang, 2004).

Mathematically, Naïve Bayes uses a probabilistic approach to become aware of the magnificence of a facts point primarily based on its attributes (Rish, 2001), (Kyburg & Pearl, 1991). primarily based on Bayes' theorem, the posterior probability of sophistication C given feature vector X is described by the following formulation:

$$P(C|X) = \frac{P(X|C).P(C)}{\mathrm{P(X)}} \quad (1)$$

Here, represents the likelihood of observing records X given class C, $P(C)$ represents the prior possibility of class C, and $P(X)$ represents the marginal possibility of the characteristic vector X.

The fundamental assumption is that the capabilities are impartial of every different when the elegance label is understood. This substantially streamlines the likelihood component:

$$P(X|C) = \prod_{i=1}^{n} P(x_i|C), \quad (2)$$

where denotes the function of the input vector X. This assumption drastically reduces the computational complexity and allows the model to learn efficaciously even on high dimensional statistics.

Inside the schooling level, the set of rules determines the earlier chance for each class and the conditional chance for every feature relative to the class. Then the model predicts a new instance by calculating the posterior chance of every class and deciding on the best chance class as the expected label.

Although it is predicated on an robust independence assumption, Naïve Bayes has tested tremendous efficacy in actual-global conditions, in particular in excessive-dimensional areas which include herbal language processing. inside the case of textual content type which include analysing feedback from students, the functions inside the dataset usually include words or tokens. The model calculates the opportunity of an textual content belonging to a specific category (e.g., advantageous, terrible, or neutral), then units it to the class with the very best opportunity. there are numerous versions of the algorithm, which include Multinomial Naïve Bayes, Bernoulli Naïve Bayes, and Gaussian Naïve Bayes, which every is designed for unique styles of data distributions. Additionally, this approach is quite easy to apply and comprehend.

2.2 Performance Evaluation

The predictive model is evaluated at several aspects, namely: precision, recall, F1-Score, and accuracy (Dani & Ginting, 2023). These metrics provide a comprehensive understand-ing of the Naïve Bayes Classifier's performance across different sentiment clas-ses (good, neutral, and bad).

i. Precision tells us how many of the predicted positive instances were actual-ly true. The formulation for precision is defined by

$$Precision = \frac{T_P}{T_P + F_P}. \quad (3)$$

ii. Recall describes how many the actual positive instances were truly predict-ed by the model, which is expressed by

$$Recall = \frac{T_P}{T_P + F_N}. \quad (4)$$

iii. F1 score is defined as a weighted average between precision and recall, which is calculated by

$$F1\ score = 2 . \left(\frac{Precision\ .\ Recall}{Precision + Recall}\right). \quad (5)$$

iv. Accuracy measures the overall correctness of the predictive model. It is the ratio of true positive and true negative to the total number of feedback en-tries. The accuracy can be obtained from formula

$$Accuracy \quad = \frac{T_P + T_N}{T_P + F_P + T_N + F_N}, \quad (6)$$

where (true positive) is the number of instances that are predicted as positive and are actually positive, (true negative) is the number of instances that are predicted as negative and are actually negative, (false positive) is the number of instances that are predicted as positive and are actually negative, and (false negative) is the number of instances that are predicted as negative and are actually positive.

2.2 Research Methodology

The research process consist of several steps as presented in Fig. 1. The study begins by analysing the problem and investigating the technique to solve it, including defining the requirement.

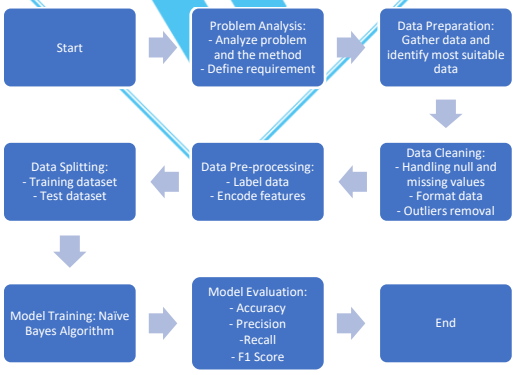


Figure 1. Flowchart of Research Methodology

The next step is data preparation where the dataset is selected after identifying the most suitable dataset. Dataset was taken from the site https://www.kaggle.com/datasets/brarajit18/student-feedback-dataset/code, which consists of 185 student feedback data related to certain aspects within the educational institution. The features data includes the following:

- Teaching (A): Opinions on teaching methods.
- Course Content (B): Assessment of course materials.
- Examination (C): Response regarding exam evaluation.
- Lab work (D): Feedback on laboratory activities.
- Library facilities (E): The quality of library facilities.
- Extracurricular (F): Assessment of extracurricular activities

Table 1 presents the sample data of student feedback in teaching aspect.

Table 1. Sample Data

No	Class Category	Student Feedback in Teaching (A)
1	Neutral	teachers are punctual but they should also give us some practical knowledge other than theortical
2	Good	Excellent lectures are delivered by teachers, and all teachers are very punctual.
3	Bad	lecture delivery is not so good
4	Good	Lectures help in clear our concept
5	Good	lecturers are punctual and can interact them easily.
6	Good	good and punctual
7	Bad	Some of the teachers are unexperienced. Also, they are unfair and baised. Infact many of the teachers are having few or no experience about the subject
8	Neutral	punctuality and lecture is good here but interaction isn't good.
...	…	…
185	Bad	staff is showing partiality between north indians and south indians

The raw data often contains noise, missing values, inconsistent formats, or outliers that can negatively affect model performance. In the next step, the dataset is cleaned by formatting the data consistently and removing missing values and outliers. Data pre-processing is conducted by assigning class labels for supervised learning and encoding categorical variables into numerical form. Each parameter has a value that reflects sentiment: Good (2), Neutral (1), or Bad (0). The ground truth label for each feature in dataset is shown in Table 2.

Table 2. Ground Truth Label for Each Feature in Dataset

	A	B	C	D	E	F
Bad (0)	13	31	25	37	34	12
Neutral (1)	35	26	30	16	24	19
Good (2)	137	128	130	132	127	154

After data pre-processing, the data is separated into two primary sets: a train-ing dataset (70%) and a test dataset (30%). The Naïve Bayes model is developed using the training set, and the test set is kept for assessing the model's effective-ness. Once the model is trained, it is evaluated using the test dataset. Several performance metrics are used to measure the effectiveness of the classification, such as confusion matrix, accuracy, precision, recall, and F1 score.

III. RESULTS AND DISCUSSION

The programming language chosen in this study is Python because it has powerful libraries for data processing and statistical analysis. The data is divided into two parts: 70% of the data, which amounts to 129 records, was used for training the model; the remaining 30% of the data, which is 56 records, was set aside as the test data.

Table 3. Confusion Matrix

		Predicted Values		
		Bad	Neutral	Good
Actual Values	Bad	2	0	9
	Neutral	0	1	6
	Good	7	0	31

After splitting the data and training the model with Naïve Bayes algorithm, we calculate the accuracy of the model using a confusion matrix which is presented in Table 3. The rows of the matrix show actual value of the class, while the col-umn shows the predicted value of the class. From the table, we can see that the model's performance is poor in recognizing the 'Bad' class, since there are only two instances that are correctly predicted as 'Bad', while most of the 'Bad' in-stances are incorrectly predicted as 'Good' (9 instances). Meanwhile, the model also has difficulty in predicting the 'Neutral' class well because only one in-stance is correctly predicted as 'Neutral', while the other seven instances are incorrectly predicted as 'Good'. Finally, the 'Good' class has the best perfor-mance among the other classes, there are 31 instances that are correctly predict-ed, though there are still minor errors where seven instance is incorrectly pre-dicted as 'Bad'.

Table 4. Performance Evaluation

	Accuracy	**Precision**	**Recall**	**F1-score**
Bad	0.61	0.22	0.18	0.20
Neutral	0.61	1.00	0.14	0.25
Good	0.61	0.67	0.82	0.74

Table 4 displays the outcomes for precision, recall, F1-score, and the accuracy of the forecasting model. The visualization of precision, recall, and F1-score for each category can be seen in Fig. 2. The model's overall accuracy is 0.61, indicat-ing that its performance is quite satisfactory.

The explanation of precision, recall, and F1-score outcomes for each class is outlined as follows.

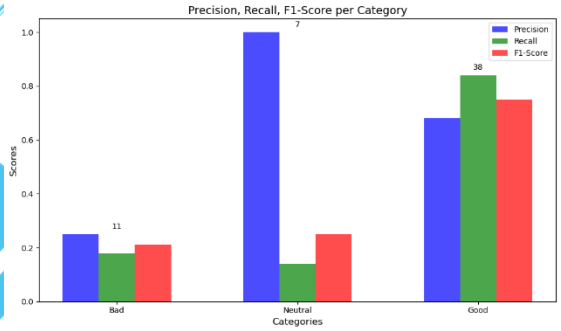


Figure 2. Visualization of Model's Precision, Recall, and F1-Score Comparison

3.1 Bad Class

The precision in Bad class is 0.22 which means that when the model predicts an instance as Bad only 22% is true. This low precision indicates that many in-stances predicted as Bad are actually not in Bad class. The recall is 0.18 which means that only 18% of the actual Bad instances are truly predicted by the mod-el. The F1 score of 0.20 is poor because the precision and recall are low as well.

3.2 Neutral Class

The precision in Neutral class is 1 which means that when the model predicts an instance as Neutral it is 100% true. This high precision indicates that many instances predicted as Neutral are actually in Neutral class. The recall is 0.14 which means that only 14% of the actual Neutral instances are truly predicted by the model. The F1 score of 0.25 is poor because the recall is very low.

3.3 Good Class

The precision in Good class is 0.67 which means that when the model predicts an instance as Good it is 67% true. This relatively good value of precision indi-cates that many instances predicted as Good are actually in Good class. The re-call is 0.82 which means that the model is very good at catching Good instance. The F1 score of 0.74 is strong reflecting a good balance between the precision and recall.

IV. CONCLUSION

This have a look at efficiently implemented the Naïve Bayes type algorithm to investigate and classify college students' comments into fantastic (properly), neutral, and negative (horrific) sentiment classes. The proposed model executed a average accuracy of zero.61, with the first-class overall performance acquired in the properly class, attaining a precision of zero.67, don't forget of zero.82, and F1-rating of zero.seventy four. These consequences imply that the model is capable of identifying fantastic student comments efficaciously and can assist computerized sentiment type for academic assessment purposes.

The model nevertheless confirmed barriers in recognizing the horrific and neutral lessons, where the precision and don't forget values remained distinctly low. This problem is specifically because of the imbalanced distribution of education data, in which the best category dominates the dataset. As a end result, the classifier tends to bias predictions toward the majority class. Consequently, enhancing data balance and increasing the quantity of negative and neutral comments data are necessary to decorate type overall performance.

From a realistic attitude, this studies demonstrates that device learning-based totally sentiment evaluation can assist universities in tracking students' perceptions more successfully and objectively. The device can help academic institutions perceive teaching first-rate problems, compare academic offerings, and assist data-driven selection-making tactics for continuous academic development. destiny studies may additionally consciousness on applying data balancing techniques, the usage of larger datasets, and comparing Naïve Bayes with more advanced device learning or deep learning techniques to achieve higher type overall performance.

AUTHOR'S CONTRIBUTION

Yasi Dani and Maria A. Ginting contributed to the research equally in different areas. Yasi Dani contributed into rewriting half of the introduction, preparing the dataset, providing codes, and providing the repository storage. Maria A. Ginting contributed to the research into rewriting the other half of the introduction, creating workflow charts, and general grammar fixes.

AVAILABILITY DATA AND MATERIALS

The data underlying this study are openly available at *https://www.kaggle.com/datasets/brarajit18/student-feedback-dataset/code*.

REFERENCES

Dąbrowski, J., Letier, E., Perini, A., & Susi, A. (2020). Mining User Opinions to Support Requirement Engineering: An Empirical Study. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 12127 LNCS. https://doi.org/10.1007/978-3-030-49435-3_25

Dani, Y., & Ginting, M. A. (2024). Comparison of Iris dataset classification with Gaussian naïve Bayes and decision tree algorithms. International Journal of Electrical and Computer Engineering, 14(2), 1959-1968. http://doi.org/10.11591/ijece.v14i2.pp1959-1968

Dani, Y., & Ginting, M. A. (2023). Classification of predicting customer ad clicks using logistic regression and k-nearest neighbors. JOIV: International Journal on Informatics Visualization, 7(1), 98-104. https://dx.doi.org/10.30630/joiv.7.1.1017

Guzman, E., Oliveira, L., Steiner, Y., Wagner, L. C., & Glinz, M. (2018). User feedback in the app store: A cross-cultural study. Proceedings - International Conference on Software Engineering. https://doi.org/10.1145/3183428.3183436

Hand, D. J., & Yu, K. (2001). Idiot's Bayes - Not so stupid after all?. International Statistical Review, 69(3). https://doi.org/10.1111/j.1751-5823.2001.tb00465.x

Hitesh, M. S. R., Vaibhav, V., Kalki, Y. J. A., Kamtam, S. H., & Kumari, S. (2019). Real-time sentiment analysis of 2019 election tweets using word2vec and random forest model. 2019 2nd International Conference on Intelligent Communication and Computational Techniques, ICCT 2019. https://doi.org/10.1109/ICCT46177.2019.8969049

Kastrati, Z., Dalipi, F., Imran, A. S., Nuci, K. P., & Wani, M. A. (2021). Sentiment analysis of students' feedback with nlp and deep learning: A systematic mapping study. In Applied

Sciences (Switzerland) (Vol. 11, Issue 9). https://doi.org/10.3390/app11093986

Kewsuwun, N., & Kajornkasirat, S. (2022). A sentiment analysis model of agritech startup on Facebook comments using naive Bayes classifier. International Journal of Electrical and Computer Engineering, 12(3). https://doi.org/10.11591/ijece.v12i3.pp2829-2838

Kumar, R., Sarddar, D., Sarkar, I., Bose, R., & Roy, S. (2020). A Literature Survey On Sentiment Analysis Techniques Involving Social Media And Online Platforms. International Journal of Scientific & Technology Research, 9(05).

Kyburg, H. E., & Pearl, J. (1991). Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. The Journal of Philosophy, 88(8). https://doi.org/10.2307/2026705

Mujahid, M., Lee, E., Rustam, F., Washington, P. B., Ullah, S., Reshi, A. A., & Ashraf, I. (2021). Sentiment analysis and topic modeling on tweets about online education during covid-19. Applied Sciences (Switzerland), 11(18). https://doi.org/10.3390/app11188438

Obaidi, M., Nagel, L., Specht, A., & Klünder, J. (2022). Sentiment analysis tools in software engineering: A systematic mapping study. Information and Software Technology, 151. https://doi.org/10.1016/j.infsof.2022.107018

Rennie, J. D. M., Shih, L., Teevan, J., & Karger, D. (2003). Tackling the Poor Assumptions of Naive Bayes Text Classifiers. Proceedings, Twentieth International Conference on Machine Learning, 2.

Rish, I. (2001). An empirical study of the naive Bayes classifier. IJCAI 2001 Workshop on Empirical Methods in Artificial Intelligence, 22230. https://doi.org/10.1039/b104835j

Sharma, S., Bharti, S. K., & Goel, R. K. (2018). A Frame Study on Sentiment Analysis of Hindi Language Using Machine Learning. International Journal of Trend in Scientific Research and Development (Vol. 2, Issue 4). https://doi.org/10.31142/ijtsrd14397

Wang, Z., Hu, S. jie, & Liu, W. dong. (2024). Product feature sentiment analysis based on GRU-CAP considering Chinese sarcasm recognition. Expert Systems with Applications, 241. https://doi.org/10.1016/j.eswa.2023.122512

Zhang, H. (2004). The optimality of Naive Bayes. Proceedings of the Seventeenth International Florida Artificial Intelligence Research Society Conference, FLAIRS 2004, 2.

In Press