

Sentiment Analysis of Student Feedback for Learning Improvement using Naive Bayes and LSTM (Long Short-Term Memory) Methods

Abdur Rahman, Khairul, Muhammad Iqbal, Andysah Putera Utama Siahaan

Abstrak

This study aims to analyze the sentiment of STMIK Kaputama students' comments using Lexicon-Based Labeling, Multinomial Naive Bayes, and Long Short-Term Memory (LSTM) approaches. The data consisted of student comments related to services, facilities, and learning processes within the campus environment. The research stages included automatic sentiment labeling using an Indonesian sentiment lexicon, text preprocessing (case folding, cleansing, tokenization, padding, and truncating), splitting the data into training and testing sets, training the Naive Bayes model using TF-IDF feature representation, and training the LSTM model using tokenized sequence data. The results show that both models were able to perform sentiment classification effectively on the available testing data. Although the obtained evaluation scores were very high, the results were still influenced by the limited number of data samples and the imbalance of class distribution. Conceptually, LSTM has an advantage in understanding context and word order, while Naive Bayes is simpler and more efficient for small-sized datasets. This study demonstrates that the combination of lexicon-based labeling and sentiment classification methods can be used as an initial approach to support the evaluation of service quality and student satisfaction at STMIK Kaputama.

Keywords: Sentiment analysis, Lexicon-Based Labeling, Naive Bayes, LSTM, TF-IDF, student comments.

Abdur Rahman¹

¹Dept. of Information Technology, Pembangunan Panca Budi University, Medan, Indonesia
e-mail: abdurrahman25071987@gmail.com¹

Khairul², Muhammad Iqbal³, Andysah Putera Utama Siahaan⁴

Email: khairul@dosen.pancabudi.ac.id, muhammadiqbal@dosen.pancabudi.ac.id, andiesiahaan@gmail.com

3rd International Conference Epicentrum of Economic Global Framework (ICEEGLOF)

Theme: Cross-Sector and Cross-Disciplinary Strategies for Achieving the SDGS

<https://proceeding.pancabudi.ac.id/index.php/ICEEGLOF>

Introduction

Learning evaluation is a crucial component in higher education to ensure the quality of teaching. Collecting student feedback through questionnaires typically produces information-rich free-text data. However, manually processing large volumes of data is time-consuming, prone to subjective bias, and inefficient. To address these challenges, applying Natural Language Processing (NLP) through sentiment analysis offers an effective automation solution.

In the field of NLP, Multinomial Naive Bayes is known as a fast and efficient baseline model, despite ignoring word order. On the other hand, Deep Learning-based Long Short-Term Memory (LSTM) is capable of understanding sequential context and long-term dependencies in sentences. While both methods have proven reliable, empirical comparisons of their performance on student feedback data at STMIK Kaputama are still limited.

This study aims to implement and compare the performance of the Naive Bayes and LSTM algorithms based on accuracy, precision, recall, and F1-score metrics. The analysis results are categorized into positive, neutral, and negative sentiments. This research contributes not only to determining the superiority of the model's accuracy but also to providing concrete recommendations for institutions to improve the quality of learning.

Literature Review

Evaluation of the quality of learning and academic services in higher education is an important instrument for maintaining and improving the quality of teaching [1]. One of the main sources of information is the collection of student feedback collected periodically every semester [1]. However, manual analysis of large volumes of text-based review data is relatively time-consuming and prone to subjective bias [1]. To overcome these obstacles, Natural Language Processing (NLP) and text mining techniques are utilized to perform sentiment analysis automatically [1].

In the sentiment analysis pipeline, text preprocessing is a crucial initial stage for cleaning noise in raw text data and improving classification quality [1, 4]. This stage generally includes case folding, cleansing, tokenization, stopword removal, stemming, and feature representation in numerical form [1, 4]. To address the limitations of manual labeling on raw data, a lexicon-based labeling approach is often used as an initial step to automatically convert unstructured opinion text into numerically and categorically labeled data [1].

Several previous studies have applied machine learning and deep learning classification algorithms to process text data and feedback in educational institutions and the public domain. Herdiansyah et al. showed that the Multinomial Naive Bayes algorithm has the advantage of

fast and efficient computation as a baseline model in text opinion classification [4]. Sasmita et al. implemented the Naive Bayes method to develop a sentiment analysis system in evaluating lecturer performance at Ganesha University of Education [4]. In line with that, Irawan et al. applied the Naive Bayes algorithm combined with Term Frequency–Inverse Document Frequency (TF-IDF) feature extraction to analyze student sentiment towards academic and administrative services [4]. Meanwhile, Hidayat et al. tested Naive Bayes to classify student satisfaction levels towards Academic Information System services [4].

Although Naive Bayes has proven to be reliable and efficient for small to medium-sized datasets, this probabilistic algorithm assumes that each word is independent and ignores word order and sentence context [1, 4]. On the other hand, sequence-based deep learning architectures such as Long Short-Term Memory (LSTM) are specifically designed to handle sequential text data [1, 4]. Yosia Wibowo et al. proved that LSTM is able to handle long-term information dependencies and capture sentence context more fully in review text [4]. Nurzaenab et al. also evaluated the performance of LSTM and found its superiority over conventional RNN in understanding contextual relationships between words [4].

Specifically in the realm of higher education service classification and feedback, Saputri et al. compared the performance of Naive Bayes and LSTM to analyze online service

sentiment at the Faculty of Engineering and Vocational Studies, Undiksha [4]. Their study results showed that LSTM was able to provide better performance on complex text due to its ability to understand sentence context, while Naive Bayes remained an efficient and lightweight choice for smaller datasets [4].

Although both methods have been tested in various text classification domains, comparative research directly evaluating the effectiveness of Naive Bayes and LSTM with student feedback data locus in the Academic Information System at STMIK Kaputama is still very limited [4]. This study aims to fill this research gap by comparing the classification performance of the two algorithms based on accuracy, precision, recall, and F1-score metrics, in order to provide data-based recommendations for campus management in improving the quality of the learning process [1, 4].

Research Methodology

This research uses an experimental quantitative approach. The research phase begins with collecting student comment data, followed by automatic labeling using Lexicon-Based Labeling. Text preprocessing includes case folding, cleansing, tokenization, and sequence formation through padding and truncating. The processed data is then divided into training and test data (train-test split). The Naive Bayes model is trained using the TF-IDF word

Abdur Rahman, Khairul,
Muhammad Iqbal, Andysah Putera Utama Siahaan

weight representation, while the LSTM model is trained with a sequential architecture structure. The performance evaluation of both models is measured based on Accuracy, Precision, Recall, and F1-Score.

Results

1. Data Splitting

The dataset used in this study consists of 10 student feedback data sets. The dataset was divided into 80% training data and 20% testing data. Details of the data distribution are shown in Table 1.

Table 1. Dataset Distribution Based on Sentiment

Sentiment	Training Data (80%)	Test Data (20%)	Total
Positive	3	2	5
Neutral	4	0	4
Negative	1	0	1
Total	8	2	10

2. Model Evaluation Results

a. Naïve Bayes

The Naïve Bayes model was trained using 8 training datasets and tested on 2 test datasets. Based on the classification results on the test datasets, the Naïve Bayes model successfully classified all test datasets correctly. The Naïve Bayes model evaluation results are presented in Table 2.

Table 2. Performance Evaluation of Naïve Bayes Model

Evaluation Metrics	Mark
Accuracy	100.00%
Precision	1.00
Recall	1.00
F1-Score	1.00

b. Long Short-Term Memory (LSTM)

The LSTM model was trained over 20 epochs. At the start of training (epoch 1), the training accuracy was 0.50 and steadily increased, reaching 1.00 in the 20th epoch. Meanwhile, the validation accuracy stabilized at 0.83 starting in the 10th epoch.

After training, the LSTM model was tested using test data. The results of the LSTM model performance evaluation can be seen in Table 3.

Table 3. Performance Evaluation of LSTM Model

Metrik Evaluasi	Mark
Accuracy	100.00%
Precision	1.00
Recall	1.00
F1-Score	1.00

Discussion

The results of the comparison of evaluation metrics between the Naïve Bayes and LSTM methods are summarized in Table 4.

Table 4. Performance Evaluation of LSTM Model

Metode	Accuracy	Precision	Recall	F1-Score
Naïve Bayes	100.00%	1.00	1.00	1.00
LSTM	100.00%	1.00	1.00	1.00

Based on Table 4, both methods produced perfect performance values (100% accuracy and an F1-score of 1.00). These homogeneous evaluation results are due to the limited number of test data samples used in the testing (only 2 test data sets).

Although empirically on this limited sample, both algorithms demonstrate similar performance, theoretically and conceptually, the LSTM method is superior to Naïve Bayes for text sentiment analysis. LSTM is able to maintain long-term dependencies and take into account the contextual sequence of words in a sentence, while Naïve Bayes assumes each word is independent of each other (bag-of-words assumption). For further research, it is recommended to increase the volume of test data so that the model performance evaluation can objectively reflect more complex text variations.

Conclusion

A student feedback sentiment analysis approach using a combination of Lexicon-Based Labeling with Naive Bayes and LSTM algorithms was successfully implemented to automatically classify comments. Although both models achieved up to 100% evaluation performance on a limited testing dataset, LSTM conceptually outperformed long and complex text contexts, while Naive Bayes offered a simpler and more efficient process for small to medium-sized data. These results provide an effective starting point for universities to quickly and automatically evaluate the quality of learning, facilities, and services.

References

- [1] Adrian, C., & Choiriah, W. (2025). Analisis Sentimen Pada Kuesioner Evaluasi Dosen Oleh Mahasiswa (EDOM) Fakultas Ilmu Komputer Universitas Lancang Kuning Dengan Metode Naïve Bayes Classifier. 4(1), 173–180. Analisis Sentimen Komentar Universitas di Indonesia Menggunakan Metode Naive Bayes dan SVM_2026.pdf. (n.d.).
- [2] Hamdana, E. N., & Alfahmi, M. B. I. (2021). Pengembangan Sistem Analisis Sentimen Berbasis Java Pada Data Twitter Terhadap Omnibus Law Menggunakan Algoritma Naïve Bayes dan K-Nearest Neighbor (K-NN). Jurnal Informatika Polinema, 7(2), 79–84. <https://doi.org/10.33795/jip.v7i2.688>
- [3] Hartantom, B., Yunita, H. D., Faurian, F., Dirayati, F., Winarko, T., & Marlina, I. (2025). Evaluasi Keterlibatan Mahasiswa Dalam Lingkungan Pembelajaran Daring Menggunakan Natural Language Processing (NLP) dan Analisis Sentimen. Jurnal Algoritma, 22(1), 60–69. <https://doi.org/10.33364/algoritma/v.22-1.2154>.
- [4] Herdiansyah, M. R., Yuliana, A., Informatika, T., Bayes, N., & Sentimen, K. (2024). Naïve Bayes Berdasarkan Komentar Pada Youtube. 8(6), 12454–12459.
- [5] Hidayat, T., Cahyana, R., & Julianto, I. T. (2024). Analisis Sentimen Layanan Sistem Informasi Akademik Mahasiswa Menggunakan Algoritma Naive Bayes. Jurnal Algoritma, 21(1), 119–130. <https://doi.org/10.33364/algoritma/v.21-1.1514>.
- [6] Irawan, D., Budianto, H., & Puteria Wati, D. (2025). Analisis Sentimen Mahasiswa Program Studi Sistem Informasi UNIKU terhadap Pelayanan Akademik dan Administrasi Menggunakan Metode Naive Bayes Classifier. In Buffer Informatika (Vol.

- 12, Number 1).
- <https://journal.fkom.uniku.ac.id/buffer/article/view/573%0Ahttps://doi.org/10.25134/buffer.v12i1.573>.
- [7] Lestari, V. B., & Hutagalung, C. A. (2025). Evaluation of TF-IDF Extraction Techniques in Sentiment Analysis of Indonesian-Language Marketplaces Using SVM, Logistic Regression, and Naive Bayes. In *Journal of Computer Science and Applications* (Vol. 8, pp. 36–44). <https://doi.org/10.21009/j->.
- [8] Nurzaenab, N., Maslihatin, T., Halid, A., Muniar, A. Y., Sabir, F. M., Parawewe, A. M. A. K., Awaliah, N., & Halfiani, H. (2025). Evaluasi Performa Metode Long Short Term Memory (Lstm) Dan Recurrent Neural Network (Rnn) Pada Analisis Sentimen Komentar Pengguna Aplikasi Kitalulus. *Jtriste*, 12(1), 86–92. <https://doi.org/10.55645/jtriste.v12i1.614>.
- [9] Pattiiha, F. S., & Hendry, H. (2022). Perbandingan Metode K-NN, Naïve Bayes, Decision Tree untuk Analisis Sentimen Tweet Twitter Terkait Opini Terhadap PT PAL Indonesia. *JURIKOM (Jurnal Riset Komputer)*, 9(2), 506. <https://doi.org/10.30865/jurikom.v9i2.4016>
- [10] Ramadila, H., Syafitri, E. D., & Zamsuri, A. (2025). Sentiment Detection of Shopee E-Commerce Application Reviews Using Natural Language Processing and Support Vector Machine. *JAISEN: Journal of Advanced Information Systems and Engineering*, 1, 28–38.
- [11] Santoso, H., Armansyah, A., & Desliani, D. (2022). Analisis Sentimen Mahasiswa Terkait Pembelajaran Tatap Muka Menggunakan Metode Naive Bayes Classifier. In *Techno.Com* (Vol. 21, Number 3, pp. 644–654). <https://doi.org/10.33633/tc.v21i3.6262>.
- [12] Saputri, N. K. T. A., Gunadi, I. G. A., & Sunarya, I. M. G. (2024). Analisis Sentimen Pelayanan Daring di Fakultas Teknik dan Kejuruan Universitas Pendidikan Ganesha Menggunakan Algoritma Naïve Bayes dan LSTM. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 4(3), 1120–1129. <https://doi.org/10.57152/malcom.v4i3.1336>.
- [13] Sasmita, A., Pradnyana, G. A., & Divayana, D. G. H. (2022). Pengembangan Sistem Analisis Sentimen Untuk Evaluasi Kinerja Dosen Universitas Pendidikan Ganesha Dengan Metode Naïve Bayes. *JST (Jurnal Sains Dan Teknologi)*, 11(2), 451–462. <https://doi.org/10.23887/jstundiksha.v11i2.44384>.
- [14] Sembiring, A. R., & Dewa, C. K. (2025). Sentiment Analysis On Indonesian Tweets about the 2024 Election. In *Sinkron* (Vol. 9, Number 1, pp. 413–422). <https://doi.org/10.33395/sinkron.v9i1.14481>.
- [15] Simanullang, A. G., Nazara, Y. A., Berutu, S. K. M., Damanik, M. A., & Pardosi, A. P. (2026). Modeling Sentiment Patterns in Indonesian Gojek Reviews with TF-IDF and Support Vector Machine. *JADEN: Journal of Algorithmic Digital Engineering and Networks*, 1(2), 49–63. <https://doi.org/10.65853/jaden.v1i2.121>.
- [16] Tamrin, F., Sudin, S., & Mandar, G. (2023). Analisis Sentimen Menggunakan Speech Recognition Dengan Algoritma Naive Bayes Classifier. *Jurnal Sains Komputer Dan Teknologi Informasi*, 6(1), 60–64. <https://doi.org/10.33084/jsakti.v6i1.5416>.
- [17] Uyun, Q., & Qoiriah, A. (2024). Analisis Sentimen Opini Publik Terhadap Program Merdeka Belajar Kampus Merdeka Dengan Algoritma Naive Bayes - Support Vector Machine (NBSVM). *Journal of Informatics and Computer Science (JINACS)*, 6(02), 589–605. <https://doi.org/10.26740/jinacs.v6n02.p589-605>.

- [18] Wijaya, H., & Hayati, N. (2025). Natural Language Processing (Nlp) Untuk AnalisisSentimen Ulasan Seblak Bandung Pedas Kudus. *Jurnal of Business and Audit Information System*, 8(1), 13–22. <http://journal.ubm.ac.id/index.php/jbase>.
- [19] Yanti, I., & Utami, E. (2025). Analisis Sentimen Pemindahan Ibu Kota Indonesia Menggunakan Word Embedding Word2Vec dan Metode Long Short-Term Memory. *Jurnal Teknik Informatika (Jutif)*, 6(1), 149–158.
- [20] Yosia Wibowo, L., Annisa, N., Ananda Khairunnisa, P., Handrianus Pranatawijaya, V., & Priskila, R. (2024). Implementasi Long Short-Term Memory Dalam Analisis Sentimen Pengguna Aplikasi Twitter Yang Mengandung Ujaran Kebencian. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 8(3), 3170–3174. <https://doi.org/10.36040/jati.v8i3.9654>.
- [21] Yunia Oktari, Rini Novianti, T. A. (2025). J URISMA : Jurnal Riset Bisnis dan Manajemen. *JURISMA: Jurnal Riset Bisnis Dan Manajemen*, XX, 209–227.

