

# Sentiment Analysis of Student Comment on the College Performance Evaluation Questionnaire Using Naïve Bayes and IndoBERT

Wiga Maulana Baihaqi<sup>1\*</sup>, Arif Munandar<sup>2</sup>

<sup>1,2</sup>Universitas Amikom Purwokerto, Central Java, Indonesia

\*corr\_author: arifmunandarr404@gmail.com

**Abstract** – The development of the Internet has played a significant role in various aspects of life and has generated vast amounts of data, including student comments about universities. The challenge in analyzing comment data is the large number of students providing feedback, which makes manual analysis impractical. The purpose of this study is to analyze the performance evaluation of universities by students in terms of positive and negative sentiments, with the aim of assessing the level of student satisfaction with all elements and areas of university operations. This research utilized the Naïve Bayes algorithm and the IndoBERT model to build a classification model based on questionnaire data, starting from the data collection process, data preprocessing, feature extraction, modeling, and evaluation. The results of the IndoBERT model demonstrated the best performance, with an accuracy of 85%. The IndoBERT model effectively recognizes sentiments in text, distinguishing between positive and negative comments regarding university performance.

**Keywords:** Classification, sentiment analysis, college evaluation, IndoBERT, Naïve Bayes

## I. INTRODUCTION

The rapid development of electronic information has permeated various aspects of life including education. In Indonesia, education covers several levels, from basic education to higher education. This has resulted in a considerable amount of data in the field of education, such as student comments on universities [1]. Higher education institutions adhere to Tri Dharma, which must be upheld by all members, including students and lecturers. Students, as a crucial component at the university level, are young adults who have forward-looking perspectives and the freedom to express their opinions and suggestions. Their input is constructive and contributes to the development of both universities and lecturers [2]. Student input regarding the evaluation of lecturers and university performance was collected once

per semester using questionnaires consisting of scales and text messages. Evaluated aspects related to university performance may include facilities, academic services, financial services, student affairs, and collaboration. On the other hand, lecturer evaluations may include teaching methods and materials. The text message section was analyzed to determine the information obtained from numerous students who provided comments and feedback on their satisfaction with the university's performance. The challenge in analyzing comments and feedback data is the large number of students who provide comments, which makes manual analysis impractical. In addition, some feedback can be ambiguous. To date, the data have been examined without understanding the extent of positive and negative responses.

Sentiment analysis is the process of analyzing, identifying, and classifying an individual's opinion that reflects their attitude towards a specific topic or product into positive, negative, or neutral categories. It examines a set of subjective opinions from multiple sources [3] [4]. Sentiment analysis is a field of work in Natural Language Processing (NLP), which is a subdomain of Artificial Intelligence that focuses on processing textual data to examine meaning within the text [2].

Several studies have focused on sentiment analysis. In the study titled [1], a sentiment analysis system was developed for short comments evaluating lecturers by students of the Information Study Program at Pamulang University. The research aimed to create an automated system capable of identifying and classifying student emotions through comments. The K-means clustering algorithm was used to categorize sentiment analysis results into positive and negative sentiments.

Research [5] focused on creating a sentiment analysis application for brands based on a website platform. The goal was to assist users in determining and comparing smartphone brands based on the opinions of other users. In another study [2], sentiment analysis was conducted

on faculty teaching evaluations using the Long Short-Term Memory (LSTM) algorithm. The accuracy achieved using the LSTM method was 91.08%. Furthermore, a study used Deep Learning IndoBERT to analyze consumer perceptions of Gojek, aiming to assess the quality of Gojek's services based on consumer reviews from Twitter. Their research findings indicated a sentiment analysis accuracy of 96% using IndoBERT [6].

Based on the problem description and prior studies, it is necessary to conduct similar research using alternative methods and data. In this study, we developed a system for analyzing student feedback on the performance of higher education institutions using the Naïve Bayes algorithm and the IndoBERT model. The main distinction between this study and previous research is the utilization of TensorFlow and PyTorch software libraries for the IndoBERT model, with a dataset comprised of student comments. Moreover, during the labeling process, researchers employ their own method, which involves automatic labeling using a machine learning algorithm trained on data from a different context, specifically, product reviews. The objective of this research is to categorize student comments and criticisms in the evaluation of higher education institutions' performance into positive and negative classifications with the aim of assessing the level of student satisfaction with performance in all aspects of the institution.

## II. METHOD

There are several steps in the system design process, and the flow of the prediction process is shown in Fig. 1.

### A. Data Collecting

In this study, two types of data were used: primary and secondary. The following is an explanation of the data-collection process.

1) *Primary Data*: The primary data consisted of questionnaires from all students at Amikom Purwokerto University. The questionnaires administered to the students were divided into two types across nine categories. The first type involved a rating scale ranging from 1 to 5, while the second type allowed students to provide comments and criticisms based on the evaluated categories. The purpose of these questionnaires was to evaluate the performance of all the departments at Amikom Purwokerto University. The students completed the questionnaires every semester. The questionnaire data used in this study were collected after the Odd Semester of the 2022/2023 academic year. The data used for analysis and model creation consisted of a free-text section containing comments and criticisms from the students of Amikom Purwokerto University. The questionnaires are divided into nine categories: Learning - Material, Learning - Methods, Learning - Technology, Learning - Evaluation, Learning - Others, Facilities, Student Affairs, Finance, and Collaboration. The total combined primary data from all categories amounted to 3179 rows.

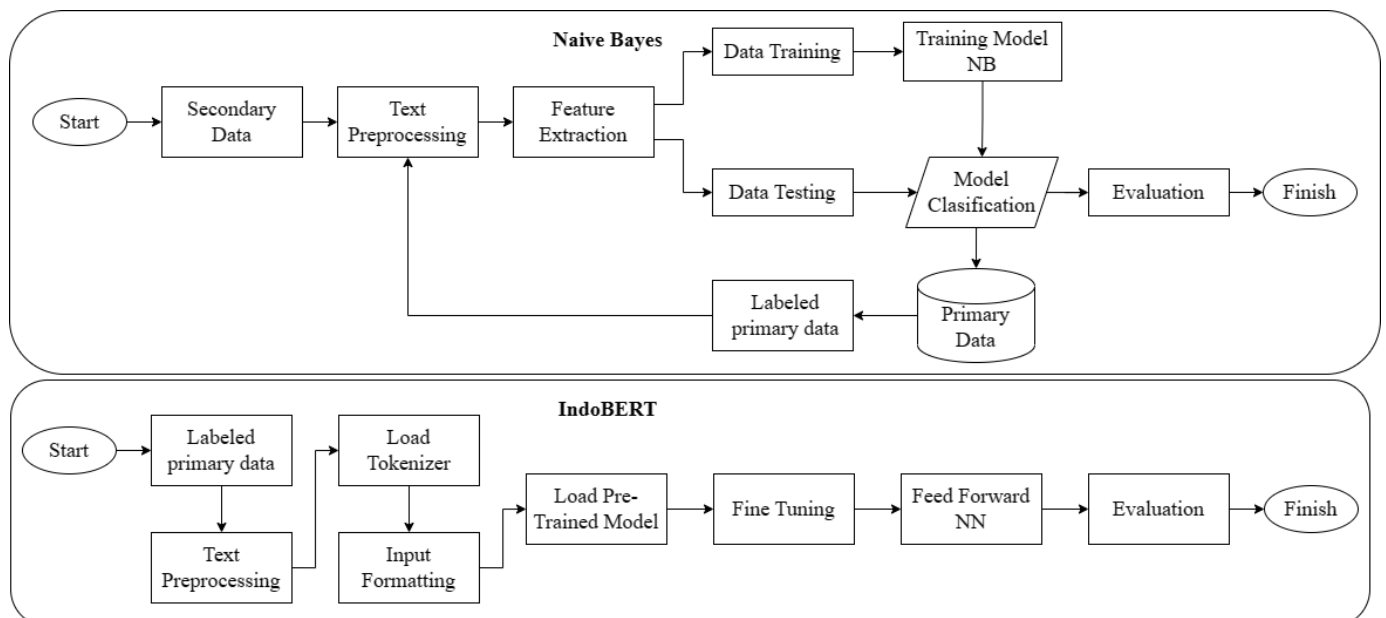


Fig. 1 Research flow

2) *Secondary Data*: The secondary data used in this study were sourced from ([https://github.com/notfound313/sentiment-analysis/blob/main/HP\\_K.csv](https://github.com/notfound313/sentiment-analysis/blob/main/HP_K.csv)). This publicly available dataset comprises 510 rows of electronic product reviews from an e-Commerce platform. The data are utilized to develop a classification model that aims to automate the labeling process for questionnaire data, thereby eliminating the need for manual labeling.

### B. Preprocessing

During the data pre-processing stage, the goal was to clean and structure the data to facilitate the analysis process. The text preprocessing stage is executed automatically using the NLTK (Natural Language Toolkit) library, which is user-friendly and offers access to over 50 corpora and lexical resources [7]. This stage consists of the following phases.

1) *Case Folding*: Converts text to a standard format, allowing only lowercase letters 'a' to 'z' [1], [8], [9].

2) *Tokenizing*: Separates text into words, symbols, characters, or punctuation while removing unnecessary characters [9] – [11].

3) *Removing Stopwords*: Eliminates commonly ignored words, such as adverbs and conjunctions like 'in', 'to', 'from', and other conjunctions [9], [12], [13].

4) *Normalization*: Corrects abbreviations or slang words into actual words, reducing ambiguity [9] [13].

5) *Stemming*: Removes affixes, prefixes, greetings, suffixes, or combinations from the text, retaining only the root word. This phase utilized the Sastrawi library [4], [3], [9], [13].

6) *Labeling*: Assigns labels to the text data, which will be used to train the Naïve Bayes and IndoBERT models. The labeling process employs a pretrained model with secondary data, enabling automatic labeling without manual intervention.

### C. Feature Extraction

The feature extraction process involves converting words into vectors, integers, or float representations [14]. This process utilizes Term Frequency-Inverse Document Frequency (TF-IDF), a technique for weighting individual words within a document [15]. The TF-IDF value is calculated by multiplying the Term Frequency (TF) with the Inverse Document Frequency (IDF), a method that assigns weights to each word in a document [16]. A word is considered more important and given a higher contribution value if it appears frequently within

a single document. Conversely, a word's contribution value is reduced if it is commonly found in multiple documents [13].

Term Frequency expresses the value of a term that frequently appears in a document. The greater the number of occurrences of a term in a document, the higher its weight [13]. The TF formula can be seen in (1).

$$TF = \frac{\text{Word frequency in document}}{\text{Total words in document}} \quad (1)$$

Term frequency can be distributed in a collection of documents using the Inverse Document Frequency (IDF) method. When a term appears more frequently across multiple documents, the IDF value decreases [13]. The IDF formula can be expressed as follows (2).

$$IDF = \text{Log} \frac{\text{Total number of documents}}{\text{Number of documents containing the word}} \quad (2)$$

The feature extraction process is specifically utilized in the Naïve Bayes algorithm, which occurs after data pre-processing. During feature extraction for the Naïve Bayes model, an additional step is executed, which involves feature selection using the SelectKBest library and chi2. This process aims to determine the features with the highest values for training the model, thereby enhancing the evaluation results. Conversely, the IndoBERT model involves only data preprocessing followed by the training phase.

### D. Building Algorithm Model

The modeling process was performed to build a classification model using the Naïve Bayes Classifier algorithm for the review dataset and the IndoBERT model for the questionnaire dataset. Here is a brief description of both algorithms.

1) *Naïve Bayes*: The Naïve Bayes algorithm is a classification method that uses simple probabilities based on Bayes' theorem with a strong assumption of independence among features [3]. This method is well-suited for large datasets, offering fast performance in information classification and achieving high accuracy [4]. The foundation of Naive Bayes in programming is based on the Bayes formula, as shown in (3).

$$P(A|B) = \frac{P(B|A)P(A)}{P_B} \quad (3)$$

Description:

$P(A|B)$  = the likelihood of the occurrence of event A given the occurrence of event B.

$P(B|A)$  = the likelihood of the occurrence of event B given the occurrence of event A.

$P(A)$  = the probability of occurrence of event A irrespective of any other events.

$P(B)$  = the probability of occurrence of event B irrespective of any other events.

2) *IndoBERT*: IndoBERT is a pretrained model that has been trained on four billion Indonesian words and texts obtained from various platforms such as Wikipedia, online posts, video subtitles, and a collection of parallel data called the Indo-4B dataset [17] [18]. BERT (Bidirectional Encoder Representations Transformers) is a multilayer model structurally based on transformer architecture [19]. The transformer mechanism analyzes text by examining the contextual relationships between words through a self-attention mechanism, allowing inputs to interact with each other (self) and determining which should receive greater focus (attention). The sequence representation of words in a sentence is computed by combining different words in the same order using an encoder and a decoder [20].

#### E. Evaluation

The evaluation process is performed to determine the performance of the generated model. The evaluation method that will be used involves the use of a confusion matrix to indicate the classification of the number of correctly classified test data and the number of incorrectly classified test data [4]. The evaluation process includes accuracy, precision, recall and F1 score. Table I shows an example of the Confusion Matrix.

Accuracy refers to the extent to which the predicted value matches the actual value [10]. The accuracy formula can be seen in (4).

$$accuracy = \frac{TP+TN}{TP+FP+FN+TN} \quad (4)$$

Precision is the degree of accuracy of predictions by calculating positive predictions from all prediction data [10]. The precision formula can be seen in (5).

$$precision = \frac{TP}{TP+FP} \quad (5)$$

The recall is the value of the success rate in identifying recognized classes [10]. The recall formula can be seen in (6).

$$recall = \frac{TP}{TP+FN} \quad (6)$$

TABLE I  
CONFUSION MATRIX

Actual Class	Predicted Class	
	Positive	Negative
	Positive	TP
Negative	FP	TN

F1 Score or F-Measure is a combination of precision and recall values or values that represent the overall performance of the system [10]. The F1 Score formula can be seen in (7).

$$f1\ score = 2 \frac{recall \times precision}{recall+precision} \quad (7)$$

### III. RESULT AND DISCUSSION

During the preprocessing stage, the data is cleaned by eliminating blank and duplicate entries, removing punctuation marks, and performing normalization to produce well-organized, analysis-ready data. Weightings are then applied to the data set to train the Naive Bayes algorithm. A testing process is then performed to identify the best performing model. The best performing model is then implemented to categorize positive and negative comments within the LP3M questionnaire data.

#### A. Datasets

The research datasets obtained from LP3M consist of questionnaires assessing the performance of all work areas within Amikom Purwokerto University over a one-semester period. These datasets are characterized by their Indonesian language and unstructured nature. The dataset is divided into nine evaluation categories and initially contains 3,179 rows, which is reduced to 2,671 rows after cleaning. For testing the Naive Bayes algorithm, the dataset is divided into two parts: training data and test data, with an 80:20 ratio. In contrast, for testing the IndoBERT model, the dataset must be divided into three parts: training data, testing data, and validation data. The initial partitioning follows an 80:20 ratio, with 20% of the testing data further partitioned in a 50:50 ratio, resulting in a data set ratio of 80:10:10 for the IndoBERT model. The performance of each data split is evaluated and an example of the data used is shown in Table II.

#### B. Result of Naïve Bayes and IndoBERT Testing

After completing the preprocessing stage, which included cleaning the data by removing punctuation, converting sentences to lower case, removing conjunctions, symbols, numbers, and correcting abbreviations, the next step was to examine the results of testing the Naive Bayes algorithm and the IndoBERT model. The data used to train the model were labeled

with two sentiment classes for each comment, where code 0 represents a negative sentiment and code 1 represents a positive sentiment, as shown in Table II. In the training stage, the data is divided into two parts with a ratio of 80:20 for the Naïve Bayes model, while the IndoBERT model divides the data into three parts with a ratio of 80:10:10. After partitioning the data according to the specified ratios, the data were trained with the respective models, specifically the Naïve Bayes and IndoBERT models. Subsequently, evaluations such as accuracy, recall, precision, and F1 score were performed to evaluate the performance of the models. Fig. 2 shows the accuracy of the IndoBERT model training process using Tensorflow over a total of 5 epochs.

The training progress of the IndoBERT model using TensorFlow is illustrated in Fig. 2. From this graph, it is evident that the model attains a validation accuracy above 80% in the initial epoch and peaks in the 4th epoch with a validation accuracy of 83%. Training the IndoBERT model with TensorFlow on the questionnaire dataset takes roughly 4 minutes in total as in Fig. 3.

TABLE II  
DATASET SAMPLE

Teks	Label
alhamdulillah materi pembelajaran baik	1
asisten dosen tidak membantu mahasiswa	0
ac laboratorium kurang dingin	0
dimohon pembelajaran tepat waktu	0
Internetnya lambat bandwidth sedikit lebih ditingkatkan lagi	0

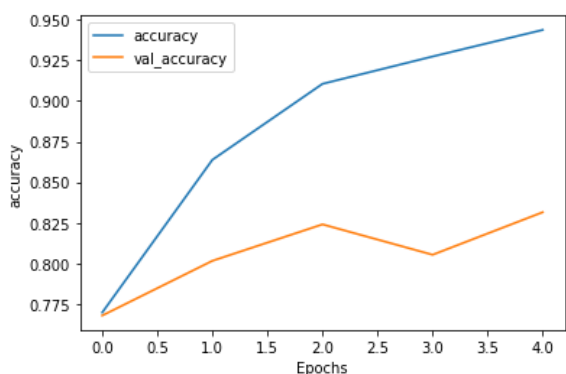


Fig. 2 IndoBERT model training accuracy plot

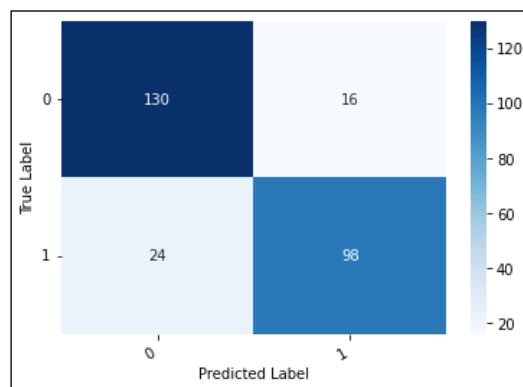


Fig. 3 IndoBERT confusion matrix visualization

In Fig. 3, the IndoBERT model utilizing TensorFlow demonstrates its capabilities by accurately classifying 130 instances as negative labels and 98 instances as positive labels. However, the model incorrectly predicted 16 instances as positive labels when they were negative, and 24 instances as negative labels when they were positive.

The accuracy development graph for training the IndoBERT model with PyTorch is shown in Fig. 4. In the first epoch, the model achieves a validation accuracy of over 69% and drops to its lowest point in the third epoch with a validation accuracy of 67%. On average, training the IndoBERT model with PyTorch on the questionnaire dataset takes approximately 30 minutes per epoch. The IndoBERT model using PyTorch presented in Fig. 5.

In Fig. 5, the IndoBERT model using PyTorch successfully classified the negative label correctly in 102 instances, but made incorrect predictions of the positive label in 44 instances. The positive label was correctly predicted 97 times, but was incorrectly classified as negative 24 times.

In Fig. 6, the Naive Bayes model demonstrates its ability to correctly classify 209 instances as negative,

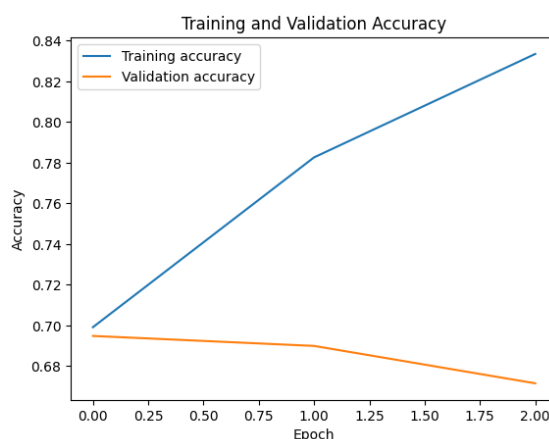
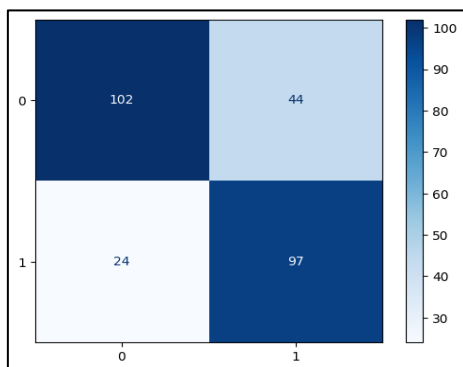


Fig. 4 IndoBERT Training Accuracy Plot

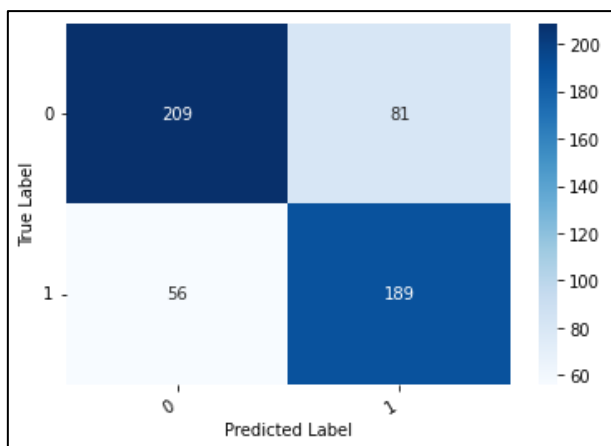


**Fig. 5 IndoBERT Model Confusion Matrix Visualization**

while incorrectly predicting 81 instances as positive. In addition, the model correctly predicts 189 positive instances, but incorrectly identifies 56 instances as negative. Table III shows the confusion matrix calculations for Fig. 3, 5, and 6, which are used to evaluate the performance of the algorithmic model.

The test results for the Naïve Bayes algorithm and the IndoBERT model are shown in Table III. The IndoBERT model using TensorFlow shows superior performance, achieving an accuracy of 85%, a recall of 80%, a precision of 86%, and an F1 score of 83%. In comparison, the IndoBERT model using PyTorch achieves an accuracy of 75%, a recall of 80%, a precision of 69%, and an F1 score of 74%. Conversely, the Naïve Bayes algorithm produces an accuracy of 74%, a recall of 77%, a precision of 70%, and an F1 score of 73%.

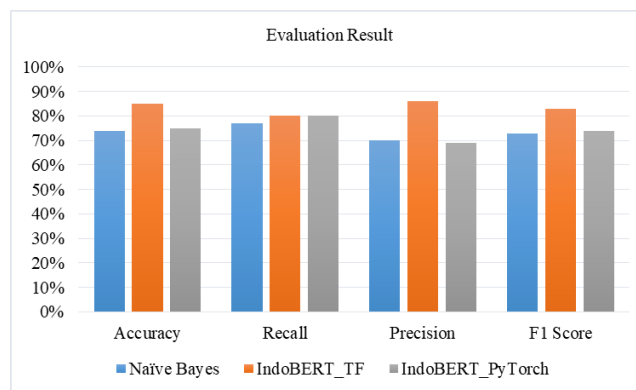
The visualization of the performance results from the testing process of the Naïve Bayes and IndoBERT algorithms can be seen in Fig. 7.



**Fig. 6 Visualization of Confusion Matrix Naïve Bayes**

**TABLE III  
EVALUATION RESULTS**

Algorithm Models	Accuracy	Recall	Precision	F1 Score
Naïve Bayes	74%	77%	70%	73%
IndoBERT (TensorFlow)	85%	80%	86%	83%
IndoBERT (PyTorch)	75%	80%	69%	74%



**Fig. 7 Visualization of Testing Result**

From Fig. 4, it can be observed that the performance of the IndoBERT model using Tensorflow appears to be superior compared to the IndoBERT model using PyTorch and the Naïve Bayes algorithm, as measured by accuracy, recall, precision and F1 Score. Hence, it can be concluded that the IndoBERT model using Tensorflow consistently achieved the highest scores across all tests and showed greater accuracy in classifying the available data, outperforming the IndoBERT model using PyTorch and the Naïve Bayes algorithm.

This study holds significant implications for academic institutions, students, and educators. It provides universities with a comprehensive understanding of student satisfaction in relation to the performance of various components and aspects within higher education institutions. This data can serve as a basis for growth and enhancement in areas that students identify as needing improvement. Furthermore, the study offers a more transparent perspective on how students perceive the value of services and instruction received, allowing universities to better comprehend student needs and elevate their educational experiences.

This study provides an opportunity for students to express their opinions, evaluations, and emotions regarding the performance of higher education institutions. By actively participating in enhancing educational quality, students can improve the learning environment in colleges. Furthermore, institutions can offer more satisfying educational experiences by

assisting students to better comprehend their needs and expectations.

This study offers valuable information to lecturers regarding students' evaluations of their performance. By understanding positive and negative feedback from students, lecturers can assess and modify their teaching strategies to better address their students' needs. The study also encourages self-reflection and continuous improvement in an effort to improve teaching standards. Lecturers can utilize these data as constructive feedback to identify their strengths and weaknesses in facilitating student learning.

In summary, this study holds considerable implications for universities, students, and lecturers. It aims to improve educational standards in higher education institutions, offer valuable feedback to students, and enhance the quality of the instruction provided by lecturers. By utilizing sentiment analysis and classification modeling, this study contributes to increasing student satisfaction and overall performance in higher-education settings.

#### IV. CONCLUSION

Based on the results obtained, it was found that the IndoBERT model using TensorFlow outperformed the IndoBERT model using PyTorch and the Naïve Bayes algorithm in sentiment analysis of comments regarding the performance evaluation of various departments at Amikom Purwokerto University. The test results showed that the IndoBERT model using TensorFlow achieved the best performance, with an accuracy of 85%, recall of 80%, precision of 86%, and F1 score of 83%. The IndoBERT model using PyTorch achieved an accuracy of 75%, recall of 80%, precision of 69%, and F1 score of 74%. Meanwhile, the Naïve Bayes algorithm achieved an accuracy of 74%, recall of 77%, precision of 70%, and F1 score of 73%. These results indicate that the IndoBERT model performs better in the sentiment analysis of comments. This research has demonstrated the effectiveness of using the IndoBERT model for the accurate classification and identification of positive or negative sentiments in comments related to the performance evaluation of departments at Amikom Purwokerto University. In future research, the labeling process could be automated using a dataset with the same contextual discussion. For experimentation, various machine learning algorithms such as SVM, Logistic Regression, and MLP (Multi-Layer Perceptron), as well as deep learning algorithms such as BLSTM, R-CNN, and others, can be compared. In addition, sentiment analysis can be performed using emotion classification, sarcasm detection, and aspect-based sentiment analysis.

#### ACKNOWLEDGEMENT

The researcher expresses sincere gratitude to Amikom Purwokerto University for the support and opportunities provided, both material and non-material, throughout the research. The researcher also extends appreciation to the supervising lecturer who has consistently offered support, advice, and valuable input, enabling the completion of this research within the designated timeframe.

#### REFERENCES

- [1] A. Zein, S. Farizy, and E. Suharyanto, "Sentimen Analisis Pada Komentar Pendek Evaluasi Dosen Oleh Mahasiswa (Edom) Program Studi Sistem Informasi Universitas Pamulang," *J. Ilmu Komput.*, vol. 5, no. 01, pp. 17–23, 2022, [Online]. Available: <https://jurnal.pranataindonesia.ac.id/index.php/jik/article/view/113%0Ahttps://jurnal.pranataindonesia.ac.id/index.php/jik/article/download/113/66>
- [2] M. A. Amrustian, W. Widayat, and A. M. Wirawan, "Analisis Sentimen Evaluasi Terhadap Pengajaran Dosen di Perguruan Tinggi Menggunakan Metode LSTM," *J. Media Inform. Budidarma*, vol. 6, no. 1, pp. 535–541, 2022, doi: 10.30865/mib.v6i1.3527.
- [3] W. Parasati, F. A. Bachtiar, and N. Y. Setiawan, "Analisis Sentimen Berbasis Aspek pada Ulasan Pelanggan Restoran Bakso President Malang dengan Metode Naïve Bayes Classifier," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 4, no. 4, pp. 1090–1099, 2020.
- [4] D. Normawati and S. A. Prayogi, "Implementasi Naïve Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter," *J. Sains Komput. Inform.*, vol. 5, no. 2, pp. 697–711, 2021.
- [5] T. Chandra Harita, R. Kridalukmana, and D. Eridani, "Pengembangan Aplikasi Analisis Sentimen Terhadap Brand Berbasis Web Menggunakan Kerangka Kerja Flask Web-Based Sentiment Analysis Application Development Using Flask Framework," *J. Tek. Komput.*, vol. 1, no. 2, pp. 36–40, 2022, doi: 10.14710/jtk.v1i2.36307.
- [6] N. Mahfudiyah and A. Alamsyah, "Analisis Persepsi Konsumen Terhadap Kualitas Layanan Gojek Menggunakan Sentiment Analysis Dan Topic Modeling Berdasarkan Deep Learning IndoBERT," *e-Proceeding Manag.*, vol. 9, no. 4, pp. 1812–1817, 2022.
- [7] Z. Munawar, Iswanto, D. Widhiantoro, and N. I. Putri, "Analisis Sentimen Covid-19 Pada Media Sosial Dengan Model Neural Machine Translation Analysis of Covid-19 Sentiment on Social Media with Neural Machine Translation Model," *Temat. J. Teknol. Inf. Komun.*, vol. 5, no. 1, pp. 15–20, 2022.
- [8] R. Naquitasia, D. H. Fudholi, and L. Iswari, "ANALISIS SENTIMEN BERBASIS ASPEK PADA WISATA HALAL DENGAN METODE DEEP LEARNING," *J.*

- TEKNOINFO, vol. 16, no. 2, pp. 156–164, 2022.
- [9] S. M. Chamzah, M. Lestandy, N. Kasan, and A. Nugraha, “Penerapan Synthetic Minority Oversampling Technique ( SMOTE ) untuk Imbalance Class pada Data Text Menggunakan KNN,” *Syntax J. Inform.*, vol. 11, no. 02, pp. 56–67, 2022.
- [10] M. Syarifuddin, “Analysis of Public Opinion Sentiment Regarding Covid-19 on Twitter Using the Naïve Bayes and Knn Method,” *Inti Nusa Mandiri*, vol. 15, no. 1, pp. 23–28, 2020.
- [11] I. T. Julianto, “Analisis Sentimen Terhadap Sistem Informasi Akademik Mahasiswa Institut Teknologi Garut,” *J. Algoritma*, vol. 19, no. 1, pp. 458–465, 2022.
- [12] A. P. Giovani, A. Ardiansyah, T. Haryanti, L. Kurniawati, and W. Gata, “Analisis Sentimen Aplikasi Ruang Guru Di Twitter Menggunakan Algoritma Klasifikasi,” *J. Teknoinfo*, vol. 14, no. 2, p. 115, 2020, doi: 10.33365/jti.v14i2.679.
- [13] C. H. Yutika, A. Adiwijaya, and S. Al Faraby, “Analisis Sentimen Berbasis Aspek pada Review Female Daily Menggunakan TF-IDF dan Naïve Bayes,” *J. Media Inform. Budidarma*, vol. 5, no. 2, p. 422, 2021, doi: 10.30865/mib.v5i2.2845.
- [14] M. Ma’ruf, A. P. Kuncoro, P. Subarkah, and F. Nida, “Sentiment analysis of customer satisfaction levels on smartphone products using Ensemble Learning,” *Ilk. J. Ilm.*, vol. 14, no. 3, pp. 339–347, 2022, doi: <http://dx.doi.org/10.33096/ilkom.v14i3.1377.339-347>.
- [15] D. F. Zhafira, B. Rahayudi, and Indriati, “ANALISIS SENTIMEN KEBIJAKAN KAMPUS MERDEKA MENGGUNAKAN NAIVE BAYES DAN PEMBOBOTAN TF-IDF SENTIMENT ANALYSIS OF KAMPUS MERDEKA POLICY USING NAIVE BAYES AND TF-IDF TERM WEIGHTING BASED ON YOUTUBE,” *J. Sist. Informasi, Teknol. Informasi, dan Edukasi Sist. Inf.*, vol. 2, no. 1, pp. 55–63, 2021.
- [16] L. R. Dharmawan, I. Arwani, and D. E. Ratnawati, “Analisis Sentimen pada Sosial Media Twitter Terhadap Layanan Sistem Informasi Akademik Mahasiswa Universitas Brawijaya dengan Metode K- Nearest Neighbor,” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 4, no. 3, pp. 959–965, 2020.
- [17] B. Wilie *et al.*, “IndoNLU: Benchmark and Resources for Evaluating Indonesian Natural Language Understanding,” *CoRR*, vol. abs/2009.0, 2020, [Online]. Available: <https://arxiv.org/abs/2009.05387>
- [18] S. M. Isa, G. Nico, and M. Permana, “Indobert for Indonesian Fake News Detection,” *ICIC Express Lett.*, vol. 16, no. 3, pp. 289–297, 2022, doi: 10.24507/icicel.16.03.289.
- [19] H. Jayadianti, W. Kaswidjanti, A. T. Utomo, S. Saifullah, F. A. Dwiyanto, and R. Drezewski, “Sentiment analysis of Indonesian reviews using fine-tuning IndoBERT and R-CNN,” *Ilk. J. Ilm.*, vol. 14, no. 3, pp. 348–354, 2022, [Online]. Available: <http://dx.doi.org/10.33096/ilkom.v14i3.1505.348-354>
- [20] T. I. Z. M. Putra, Suprpto, and A. F. Bukhori, “Model Klasifikasi Berbasis Multiclass Classification dengan Kombinasi Indobert Embedding dan Long Short-Term Memory untuk Tweet Berbahasa Indonesia (Classification Model Based on Multiclass Classification with a Combination of Indobert Embedding and Long,” *J. Ilmu Siber dan Teknol. Digit.*, vol. 1, no. 1, pp. 1–28, 2022, [Online]. Available: <https://doi.org/10.35912/jisted.v1i1.1509>