Department of Digital Business
Journal of Artificial Intelligence and Digital Business (RIGGS)
Homepage: https://journal.ilmudata.co.id/index.php/RIGGS
Vol. 4 No. 4 (2026) pp: 8030-8044
P-ISSN: 2963-9298, e-ISSN: 2963-914X

Sentiment Analysis on Student Feedback at Universitas Ary Ginanjar Using IndoBERT

Abdul Barir Hakim
[1]Information Systems Study Program, Universitas Ary Ginanjar
[1]abdulbarir.h@uag.ac.id

Abstract

Digital transformation in higher education requires continuous, data-driven evaluation of student satisfaction to ensure service quality and institutional improvement. Student feedback, commonly expressed in free-text form, represents a rich source of information that reflects students' perceptions of academic services. However, the unstructured nature of textual feedback poses challenges for large-scale analysis. This study aims to analyze student feedback sentiment at Ary Ginanjar University using a Natural Language Processing (NLP) approach based on the IndoBERT model. The research methodology follows the CRISP-DM framework, encompassing business understanding, data understanding, data preparation, modeling, evaluation, and deployment. During the data preparation stage, a weak labeling strategy is employed to automatically assign initial sentiment labels using a pre-trained IndoBERT model, followed by manual correction to improve label quality. This human-in-the-loop approach enables the construction of a reliable labeled dataset while reducing annotation effort. The dataset is classified into three sentiment categories: positive, neutral, and negative. Text preprocessing includes normalization, punctuation removal, and tokenization using the IndoBERT tokenizer. Subsequently, IndoBERT is fine-tuned for sentiment classification, and model performance is evaluated using accuracy, precision, recall, and F1-score metrics. To gain deeper insights into dissatisfaction factors, topic modeling using Latent Dirichlet Allocation (LDA) is applied to feedback classified as negative, resulting in five dominant topics. These topics reveal key issues related to teaching structure, learning effectiveness, workload, and classroom management. The findings provide actionable strategic recommendations to support data-driven decision-making and continuous improvement at Ary Ginanjar University.

***Keywords**: Sentiment Analysis, IndoBERT, Natural Language Processing, Topic Modelling, Student Feedback*

1. Introduction

In the current digital era, the quality of higher education services has become a major concern in ensuring student satisfaction, as students represent one of the primary stakeholders in the academic environment. Students evaluate not only the academic quality but also various aspects of campus services, including administrative processes, facilities, information technology support, and staff performance. Therefore, it is essential for higher education institutions to understand students' perceptions and satisfaction levels regarding the services they receive [2].

Systematic processing of student feedback plays a crucial role in supporting continuous improvement within universities. Feedback provided by students reflects their real perceptions of campus services, ranging from academic activities and administrative procedures to supporting facilities. When such feedback is analyzed comprehensively, universities are able to identify students' needs, detect service weaknesses, and formulate data-driven strategies for improvement [3]. Moreover, the use of technologies such as sentiment analysis enhances the efficiency of capturing opinion patterns from large volumes of student comments [4].

The primary benefit of processing student feedback lies in establishing a more objective service evaluation system and improving communication between institutions and students [5]. This effort also contributes to institutional accreditation, as it demonstrates a clear and measurable mechanism for continuous improvement. Conversely, neglecting student feedback may result in persistent dissatisfaction, reduced student loyalty, and a potential decline in the institution's public reputation [3]. Therefore, universities that actively manage and respond to student feedback are better positioned to remain adaptive, responsive, and competitive amid the increasing complexity of higher education.

To enable the automated management and analysis of student feedback, appropriate analytical methods grounded in Natural Language Processing (NLP) are required. NLP is a subfield of artificial intelligence that enables computers to process, interpret, and generate human language in a meaningful way. It has been widely applied in various text-based tasks, including information extraction, text classification, and sentiment analysis. Prior research emphasizes that NLP techniques play a critical role in bridging the gap between human language and machine understanding by facilitating semantic feature extraction and accurate linguistic interpretation [7]

Sentiment analysis is an NLP-based technique designed to identify and categorize opinions or emotional expressions embedded within textual data. This approach has been extensively adopted across various domains, including business, government, and education. Previous studies highlight that sentiment analysis has become a crucial analytical tool for understanding user opinions, particularly in large-scale platforms that generate substantial volumes of textual feedback [8].

As a core technique within Natural Language Processing, sentiment analysis provides an effective approach for uncovering students' opinions and perceptions embedded in textual data. NLP-based methods, particularly IndoBERT, have demonstrated strong capability in accurately classifying sentiment in the Indonesian language [1]. This approach has become increasingly relevant as the majority of student feedback is now collected through online platforms, such as Google Forms, resulting in large volumes of unstructured text data that require automated analysis.

Furthermore, sentiment analysis plays a crucial role in enhancing data-driven decision making within educational institutions. The insights generated from sentiment analysis enable universities to identify service areas perceived as unsatisfactory by students, thereby allowing the formulation of targeted and evidence-based improvement strategies [3]. The adoption of such analytical approaches supports the principle of continuous improvement, which is fundamental to the governance and quality assurance processes of modern higher education institutions.

To systematically interpret sentiment analysis results, particularly in order to gain deeper insights into the factors underlying negative student sentiment, further analysis of negatively classified feedback is required. One widely adopted approach for this purpose is Latent Dirichlet Allocation (LDA), an unsupervised topic modeling technique designed to uncover latent thematic structures within large collections of textual documents. LDA represents each document as a probabilistic distribution over a set of topics, while each topic is characterized as a probabilistic distribution over words. In contemporary text analytics research, LDA remains a prominent exploratory tool for identifying dominant themes in unstructured textual data. Applying LDA to specific data subsets, such as documents labeled with negative sentiment, enables researchers to more effectively identify the underlying sources of dissatisfaction. Recent studies indicate that LDA continues to be relevant and effective, particularly when combined with transformer-based sentiment classification models [6].

As an institution committed to continuous improvement, Universitas Ary Ginanjar requires a data-driven evaluation system to effectively assess and enhance the quality of its services. This study aims to analyze student sentiment toward campus services based on feedback collected through Google Forms, classifying opinions into positive, neutral, and negative categories. Following the identification of negative sentiment, topic modeling is conducted using the Latent Dirichlet Allocation (LDA) method as an exploratory analytical approach to uncover dominant themes that may contribute to student dissatisfaction. The insights derived from this analysis serve as the basis for formulating strategic recommendations that address the identified issues, thereby responding to student feedback reflected in negative sentiment and supporting institutional service improvement.

This study is guided by several research questions aimed at understanding students' perceptions and identifying opportunities for service improvement within Universitas Ary Ginanjar.
The questions are formulated as follows:

1. How do students at Universitas Ary Ginanjar perceive the quality of campus services based on the collected student feedback data?
2. How does sentiment classification using NLP-based sentiment analysis methods categorize student feedback into positive, neutral, and negat aive sentiments?
3. What dominant themes or recurring issues emerge from student feedbacks that were classified as negative?

DOI: https://doi.org/10.31004/riggs.v4i4.4698

4. What strategic recommendations can be proposed to the university based on the sentiment analysis results to enhance the quality of campus services?

This research was conducted in response to the growing need for higher education institutions to systematically analyze large-scale student feedback in order to support data-driven service improvement. Recent studies have demonstrated the effectiveness of sentiment analysis based on Natural Language Processing (NLP), particularly transformer-based models such as IndoBERT, in accurately classifying sentiment in Indonesian-language texts. Wilie et al. introduced IndoNLU as a benchmark for Indonesian NLP and demonstrated that fine-tuned IndoBERT outperforms Multilingual BERT across multiple tasks, including sentiment classification, highlighting the importance of language-specific pre-training [10]. Afifuddin et al. further showed IndoBERT's robustness in handling informal and slang-rich Indonesian Twitter data, achieving an accuracy of 89.7%, which confirms its effectiveness in real-world, noisy text environments [11].

Comparative research by Pratama and Sari revealed that BERT-based models outperform LSTM architectures, with BERT achieving a higher F1-score (90%) in Indonesian sentiment classification, emphasizing the superiority of transformer-based models in capturing contextual semantics [12]. To address data scarcity, Sari et al. applied data augmentation techniques alongside IndoBERT, resulting in a 4–5% accuracy improvement, underscoring the role of data enrichment in model optimization [13]. More recently, Oktaviani et al. demonstrated that fine-tuned IndoBERT is highly effective for academic feedback analysis, achieving up to 92% accuracy, thereby validating its applicability in higher education service evaluation contexts [14].

Furthermore, topic modeling techniques, including Latent Dirichlet Allocation (LDA), have been widely adopted to uncover latent themes in textual data. However, most existing studies treat sentiment classification and topic modeling as separate analytical processes, or apply topic modeling to the entire dataset without focusing on specific sentiment categories. This creates a research gap in understanding the underlying factors that specifically drive negative student sentiment in academic service contexts. To address this gap, the purpose of this study is to integrate IndoBERT-based sentiment analysis with LDA-based topic modeling applied exclusively to negative feedback, enabling a more focused and interpretable exploration of student dissatisfaction. The novelty of this research lies in its sentiment-driven topic modeling approach, which combines state-of-the-art transformer-based classification with targeted exploratory topic analysis to generate actionable strategic recommendations for institutional service improvement.

2. Research Methods

The research methodology refers to the procedures, tools, and research design employed in conducting the study. This methodology is intended to provide a clear, structured, and systematic guideline to ensure that the research process is carried out in an organized and reproducible manner. The stages of this study follow the Cross-Industry Standard Process for Data Mining (CRISP-DM), which is a widely adopted framework for data-driven projects. The overall research workflow based on the CRISP-DM phases is illustrated in Figure 1

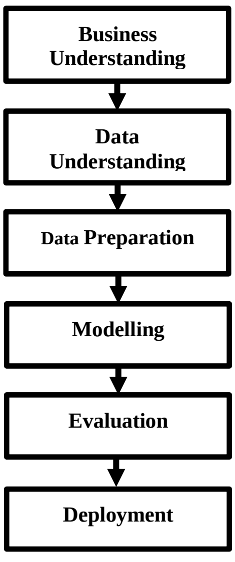

Figure 1 Research methodology workflow following the CRISP-DM phases

DOI: https://doi.org/10.31004/riggs.v4i4.4698

Based on Figure 1**,** this study follows the CRISP-DM (Cross-Industry Standard Process for Data Mining) framework. The research stages are described as follows:

2.1. Business Understanding

The *Business Understanding* phase aims to clearly define the research problem and objectives. In this study, the primary objective is to analyze student feedback sentiment at Universitas Ary Ginanjar using the IndoBERT model. This analysis seeks to capture students' perceptions, levels of satisfaction, complaints, and expectations regarding campus services, as reflected in their written feedback.

2.2. Data Understanding

The Data Understanding phase involves data collection and initial exploration. The dataset used in this study consists of student feedback text data collected via Google Forms during the period from the Odd Semester of the 2021/2022 academic year to the Even Semester Google Forms of the 2023/2024 academic year. This phase also includes preliminary data exploration to identify the volume of feedback collected in each semester and to gain an overall understanding of the data characteristics, ensuring its suitability for subsequent analysis stages.

2.3. Data Preparation

The Data Preparation phase focuses on transforming the collected data into a structured and suitable format for the modeling stage. This phase involves several preprocessing steps, including data cleaning, normalization, labeling, tokenization, and data splitting, as described below:

1. Cleaning is performed to remove irrelevant elements from the text data. This includes eliminating special characters, numbers, punctuation marks, and Indonesian stopwords, as well as removing empty or missing records to ensure data quality and consistency.
2. Normalization aims to standardize the text data. All characters are converted to lowercase to reduce vocabulary variation. In addition, lecturer names are standardized. For cases where a single column contains multiple lecturer names (multi-value attributes), normalization is applied by separating these names into individual rows, ensuring consistent representation across the dataset.
3. Labeling is conducted using a weak labeling approach, where automatic sentiment labeling is initially performed using the IndoBERT model, followed by manual correction to refine the labels. This human-in-the-loop strategy has been shown to be effective in producing high-quality labeled datasets with reduced manual effort and is widely adopted in modern NLP research [9].
   It is important to note that labeling is not applied to the entire dataset. The labeling process is conducted only on data from the Odd and Even Semesters of the 2021/2022 academic year, which are used as the supervised learning dataset for model training.
4. Tokenization is performed using the IndoBERT tokenizer (*indobenchmark/indobert-base-p1*) to generate tokens and token sequences that are compatible with the IndoBERT model architecture.
5. Data Splitting is used to divide the dataset into a training set (80%) and a testing set (20%). Splitting the data into training and testing subsets is a critical step in machine learning to evaluate the model's ability to generalize to unseen data. The 80:20 ratio is widely adopted and recommended in text classification and natural language processing studies, as it provides a balanced trade-off between sufficient training data and reliable performance evaluation. Prior studies have demonstrated that this ratio yields stable performance estimates, particularly for small to medium-sized datasets, while minimizing the risks of underfitting and unreliable evaluation [15][16][17][18].

2.4. Modeling

The modeling phase focuses on building a sentiment classification model using the training dataset. The following steps are conducted:

a. Loading the Pre-trained IndoBERT Model
A pre-trained IndoBERT model is loaded from the HuggingFace Model Hub as the base language model.

b. Adding a Classification Layer
A linear classification layer is added on top of the IndoBERT encoder to enable sentiment classification into predefined classes.

c. Fine-tuning IndoBERT
The IndoBERT model is fine-tuned using the labeled student feedback dataset. This process allows the model to adapt its contextual representations to the specific domain of academic feedback.

d. Hyperparameter Tuning
Several hyperparameters, including learning rate, batch size, and number of training epochs, are tuned to obtain optimal model performance.

2.5. Evaluation

The evaluation phase aims to assess the performance of the trained model using the testing dataset. Model evaluation is conducted by analyzing the confusion matrix and calculating standard classification metrics, namely accuracy, precision, recall, and F1-score.

a. Precision measures the correctness of the model's predictions by indicating the proportion of instances predicted as a given class that truly belong to that class.
b. Recall reflects the model's ability to identify all relevant instances of a particular class, indicating the proportion of actual instances that are correctly detected.
c. F1-score represents the harmonic mean of precision and recall, providing a balanced evaluation of model performance, particularly in cases of class imbalance.
d. Accuracy measures the overall proportion of correctly classified instances across all sentiment classes, including negative, neutral, and positive sentiments.

2.6. Deployment

After the evaluation stage confirms that the model achieves satisfactory performance, the trained model is deployed to perform sentiment analysis on the entire student feedback dataset that has been collected. The predicted sentiment results are then further analyzed and visualized to support interpretation and decision-making. Based on the insights obtained from the sentiment analysis, strategic recommendations are formulated to support service quality improvement at Universitas Ary Ginanjar (UAG). The deployment phase is carried out through the following steps:

a. Applying the Trained Model
The fine-tuned IndoBERT model developed in the modeling stage is applied to the complete student feedback dataset that has passed the data preparation process. As a result, each feedback entry is assigned a predicted sentiment label, namely positive, neutral, or negative.

b. Sentiment Visualization
The sentiment analysis results are visualized using pie charts to provide an overview of sentiment distribution. Additionally, a simple interactive dashboard is developed using Microsoft Excel, allowing sentiment visualizations to be filtered by semester, lecturer, and/or course.

c. Topic Modeling on Negative Sentiment
Subsequently, topic modeling using the Latent Dirichlet Allocation (LDA) method is conducted exclusively on feedback classified as negative. This step aims to gain deeper insights into the underlying factors contributing to negative sentiment expressed by students.

d. Formulation of Strategic Recommendations
Based on the results of sentiment classification and topic modeling, relevant strategic recommendations are formulated and proposed as actionable inputs for UAG to address identified issues and improve the overall quality of academic and administrative services.

DOI: https://doi.org/10.31004/riggs.v4i4.4698

3. Results and Discussions

The results and discussion of this study are presented following the CRISP-DM framework to ensure a structured and systematic analysis.

3.1 Business Understanding

The Business Understanding phase has been elaborated in detail in Chapter 1. In summary, this research aims to perform sentiment analysis on student feedback provided by Universitas Ary Ginanjar. The objective is to identify and understand students' levels of satisfaction, complaints, and expectations toward campus services based on textual feedback data. Furthermore, this stage seeks to generate actionable insights that can be used as strategic recommendations, particularly in addressing issues reflected in negative student feedback. These insights are expected to support data-driven decision-making and continuous service improvement within the institution.

3.2 Data Understanding

The data used for sentiment analysis in this study consist of textual student feedback collected through Google Forms during the period from the Odd Semester of the 2021/2022 academic year to the Even Semester of the 2023/2024 academic year. The feedback represents students' opinions and experiences regarding various aspects of campus services.

Based on the collected dataset, a total of 1,629 records were obtained. An initial exploration of the data revealed that no feedback data were available for the Even Semester of the 2022/2023 academic year, resulting in a gap in the temporal distribution of the dataset. The detailed number of records for each semester is presented in Table 1.

Table 1 Collected Student Feedbacks

Semester	**Jumlah *Record***
Ganjil 2021/2022	643
Genap2021/2022	176
Ganjil 2022/2023	160
Ganjil 2023/2024	413
Genap 2023/2024	237
Total	1629

The collected data are structured in the form of several columns, namely **Semester**, **Timestamp**, **Lecturer Name**, **Course**, and **Feedback (Kesan)**. Each column represents specific information related to the student feedback records. A detailed description of each column is provided in **Table 2**.

Table 2 Attribute Description of the collected Student Feedback

Column	**Descriprion**
Semester	Indicates the academic semester in which the student feedback was collected.
Timestamp	Records the exact date and time when the student submitted the feedback.
Lecturer's Name	Contains the name of the lecturer associated with the course. Academic titles and degrees are excluded to ensure name standardization.
Course	Includes the course code and course title attended by the student.
Feedback (Kesan)	Contains students' impressions and textual feedback regarding the course and learning experience. Sentiment analysis is performed based on this attribute.

4.3 Data Preparation

This stage focuses on preparing the collected data for the modeling phase. Several preprocessing steps were applied to ensure data quality and suitability for sentiment analysis. These steps include data cleaning, normalization, labeling, tokenization, and data splitting. The data preparation process is crucial to reduce noise, improve model performance, and ensure consistent representation of texstual feedback.

DOI: https://doi.org/10.31004/riggs.v4i4.4698

4.3.1 Cleaning

The cleaning process aims to remove irrelevant and noisy elements from the textual data. Specifically, this step involves the removal of special characters**,** numerical values, punctuation marks, and Indonesian stopwords. In addition, records containing empty or missing feedback text were excluded from the dataset.

This process ensures that only meaningful textual information is retained for further analysis, thereby improving the effectiveness of the subsequent tokenization and modeling stages. A summary of the cleaning results is presented in Table 3.

Table 3 Cleaning Results

Semester	Number of Records	Cleaning
Ganjil 2021/2022	643	572
Genap2021/2022	176	164
Ganjil 2022/2023	160	153
Ganjil 2023/2024	413	367
Genap 2023/2024	237	207
Total	**1629**	**1463**

4.3.2 Normalization

The normalization stage was conducted to ensure consistency and uniformity in the textual and categorical data. All text was converted to lowercase to avoid inconsistencies caused by letter casing. In addition, lecturer names were standardized by removing academic titles and unifying naming conventions.

A specific normalization step was applied to handle multi-value lecturer entries, where a single record contained more than one lecturer name. In such cases, the record was normalized by splitting the lecturer names into separate rows, with each row representing a single lecturer associated with the same feedback. This approach ensures accurate representation and prevents ambiguity in lecturer-related analysis.

After completing the normalization process, the total number of records increased to 1,507 records. A summary of the normalization results is presented in Table 4.

Table 4 Normalization Results

Semester	Number of Records	Cleaning	Normalization
Ganjil 2021/2022	643	572	571
Genap2021/2022	176	164	209
Ganjil 2022/2023	160	153	153
Ganjil 2023/2024	413	367	367
Genap 2023/2024	237	207	207
Total	**1629**	**1463**	**1507**

.3.3 Labeling

The labeling process was conducted using a weak labeling approach, where initial sentiment labels were automatically generated using a pre-trained IndoBERT model. This automatic labeling was subsequently refined through manual correction, in which the automatically assigned labels were reviewed and adjusted by the researcher to improve label accuracy.

Not all records were subjected to the labeling process. Labeling was applied only to feedback data from the odd and even semesters of the 2021/2022 academic year, resulting in a total of 780 labeled records. The weak labeling stage initially produced 583 positive, 123 neutral, and 74 negative sentiment labels.

Following this stage, manual correction was performed on 159 records identified as ambiguous or potentially misclassified. After manual refinement, the final labeled dataset consisted of 527 positive, 116 neutral, and 137 negative records. A summary of the labeling results before and after manual correction is presented in Table 5.

Table 5 Labeling Results

Process	Number of Records	Label Positive	Label Neutral	Label Negative
Weak Labeling	780	583	123	74
Manual Correction	780	527	116	137

DOI: https://doi.org/10.31004/riggs.v4i4.4698

4.3.4 Tokenization

The tokenization process was performed using the IndoBERT tokenizer (indobenchmark/indobert-base-p1). This step converts each student feedback text into a sequence of tokens that can be processed by the IndoBERT model. Tokenization is a crucial stage, as the subsequent modeling process relies entirely on these tokenized representations. Each feedback record is transformed into a structured token sequence that captures the lexical and contextual information required for effective sentiment classification.

4.3.5 Data Splitting

The splitting stage aims to divide the labeled dataset into training and testing subsets to support model development and evaluation. An 80:20 ratio was applied, where 80% of the data was allocated to the training set and 20% to the testing set. The training set was used during the modeling phase to fine-tune the IndoBERT model, while the testing set was reserved exclusively for performance evaluation, including the construction of the confusion matrix and the calculation of evaluation metrics.

From the total of 780 labeled records, 624 records were assigned to the training set, and 156 records were allocated to the testing set. A summary of the data distribution is presented in Table 6.

Table 6 Data Splitting Results

Dataset	Percentage	Number of Records
Training Set	80%	624
Testing Set	20%	156
Total		**780**

4.4 Modeling

The modeling phase was conducted by fine-tuning a pre-trained IndoBERT model using the labeled training dataset. The model architecture consists of the IndoBERT base model with an additional linear classification layer on top to perform multi-class sentiment classification.

The training process was implemented using the PyTorch framework with support from the HuggingFace Transformers library. The following training parameters were applied:

- Learning rate: 2e-5
- Batch size: 8
- Number of epochs: 4
- Optimizer: AdamW
- Loss function: Cross-Entropy Loss

These hyperparameters were selected based on common best practices for fine-tuning transformer-based models, aiming to balance training stability, convergence speed, and generalization performance. The model was trained exclusively on the training set, which consists of manually validated sentiment labels.

4.5 Evaluation

Model evaluation was performed using the testing dataset, which was not involved in the training process. A confusion matrix was constructed to analyze the classification results across the three sentiment classes: positive, neutral, and negative. The confusion matrix provides a detailed overview of correct and incorrect predictions for each class and is presented in Figure 2.

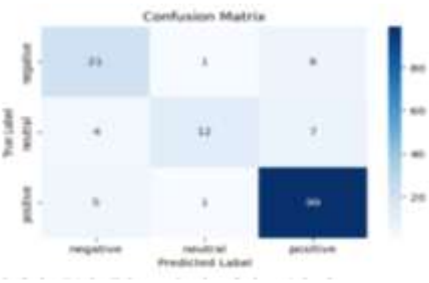

Figure 2 Confusion Matrix of Testing Set Results

DOI: https://doi.org/10.31004/riggs.v4i4.4698

To quantitatively assess the model's performance, several standard evaluation metrics were calculated, including Accuracy, Precision, Recall, and F1-Score. These metrics collectively offer a comprehensive evaluation of the model's effectiveness, particularly in handling class imbalance and capturing sentiment distinctions within student feedback data. The results of these metric calculations are presented in **Figure 3**, offering a comprehensive overview of the model's effectiveness in classifying student feedback sentiments.

	precision	recall	f1-score	support
negative	0.70	0.75	0.72	28
neutral	0.86	0.52	0.65	23
positive	0.88	0.94	0.91	105
accuracy			0.85	156
macro avg	0.81	0.74	0.76	156
weighted avg	0.85	0.85	0.84	156

Figure 3 Metrics Calculation Results

4.5.1 Recall

The recall value for the negative class is 75%, indicating that 75% of all instances that truly express negative sentiment were correctly identified by the model, while the remaining 25% were misclassified into other classes, primarily the neutral class. This result suggests that the model demonstrates a reasonably strong capability in capturing student complaints or critical feedback. However, the misclassification of 25% of negative instances remains a concern, particularly when the primary objective of the system is early detection of dissatisfaction or service-related issues, as a portion of critical feedback may remain undetected.

For the neutral class, the recall score is 52%, meaning that only slightly more than half of the neutral feedback instances were correctly classified, while the remaining 48% were incorrectly predicted as either positive or negative. This indicates that the model struggles to distinguish neutral sentiment, especially in feedback that is ambiguous, descriptive, or lacks explicit emotional expressions. Low recall for the neutral class is a common phenomenon in sentiment analysis, as neutral sentiment often overlaps semantically with both positive and negative expressions.

In contrast, the positive class achieved a recall score of 92%, indicating that the vast majority of truly positive feedback instances were successfully identified by the model. This demonstrates the model's strong effectiveness in recognizing expressions of satisfaction or appreciation from students. The risk of missing positive opinions is therefore minimal, which can be attributed to the dominant proportion of positive samples in the dataset and the more explicit linguistic patterns typically used to convey positive sentiment.

4.5.2 Precision

The precision score for the negative class is 70%, indicating that 70% of the instances predicted by the model as negative truly belong to the negative sentiment category, while the remaining 30% represent misclassifications, primarily originating from the neutral or positive classes. This result suggests that the model still exhibits a relatively high rate of error when identifying negative sentiment, which may lead to false alarms, where feedback that is not actually negative is incorrectly classified as negative. Such behavior is understandable given the relatively small proportion of negative samples in the dataset and the inherently ambiguous nature of complaint-related language.

For the neutral class, the precision value reaches 86%, meaning that 86% of the instances predicted as neutral are indeed neutral. This indicates that the model is fairly reliable when assigning the neutral label, and misclassifications into the neutral class occur relatively infrequently. However, despite the high precision, the recall for the neutral class remains low, suggesting that many truly neutral instances are not successfully identified as neutral by the model.

The positive class achieves a precision score of 88%, indicating that 88% of the feedback predicted as positive genuinely reflects positive sentiment. This demonstrates that the model is highly reliable in identifying positive sentiment, with a relatively low rate of false positive predictions. The strong performance for this class is largely influenced by the dominance of positive samples in the dataset, allowing the model to more effectively learn and generalize the linguistic patterns associated with expressions of satisfaction.

4.5.3 F1-Score

The F1-score for the negative class is 72%, indicating a reasonably good balance between precision and recall in detecting negative sentiment. This result suggests that the model is relatively consistent in identifying negative feedback, both in terms of prediction accuracy and coverage of complaint-related instances. However, there remains room for improvement, as misclassifications and missed negative cases still occur. This limitation is particularly important because negative sentiment often represents service-related issues or student dissatisfaction that require institutional attention.

For the neutral class, the F1-score reaches 65%, reflecting limited and less stable performance. This relatively low value is primarily influenced by the low recall for neutral sentiment, despite its relatively high precision. The model frequently struggles to distinguish neutral feedback from positive or negative sentiment, especially when the language used is ambiguous or informational in nature. This finding indicates the need for further strategies, such as increasing the amount of neutral data, refining label definitions, or reconsidering the role of the neutral class in the classification scheme.

In contrast, the positive class achieves a high F1-score of 91%, demonstrating that the model performs very well and consistently in classifying positive sentiment. Both precision and recall values are high, resulting in minimal classification errors. This indicates that the model has effectively learned the linguistic patterns associated with positive expressions, leading to accurate and stable predictions.

4.5.4 Accuracy

The overall accuracy of the model is 85%, meaning that 85% of the test data were correctly classified, while the remaining 15% were misclassified.

Based on the evaluation metrics discussed above, it can be concluded that the IndoBERT-based sentiment classification model demonstrates strong overall performance, as reflected by its relatively high accuracy. The model excels in identifying positive sentiment, but still exhibits limitations in recognizing neutral and negative sentiment. These limitations are largely influenced by class imbalance and the ambiguous nature of sentiment expressions in textual feedback. Nevertheless, the model is considered sufficiently reliable for general sentiment analysis tasks and can be further improved through additional data collection, class balancing techniques, or more advanced analytical approaches focused on negative sentiment.

4.6 Deployment

Although the evaluated model did not achieve exceptionally high performance, it can be categorized as sufficiently reliable for conducting sentiment analysis. Therefore, the trained model was subsequently deployed to analyze the entire student feedback dataset that had been collected. The predicted sentiment results were further analyzed and visualized to facilitate interpretation. Based on these outcomes, strategic recommendations were formulated to support Universitas Ary Ginanjar (UAG) in improving the quality of its academic services in accordance with the identified sentiment patterns.

The results obtained at each step of this deployment phase are described as follows.

4.6.1 Application of the Model to the Entire Dataset

The sentiment classification model developed during the modeling phase was applied to the complete dataset after it had undergone the data preparation process. As a result, sentiment labels—positive, neutral, or negative—were assigned to all student feedback entries. In this stage, sentiment prediction was performed on a total of 1,507 records.
Table 7 presents the overall sentiment analysis results obtained from the full dataset.

Table 7 Sentiment Analysis Results

Sentiment	**Number of Records**
negative	280
neutral	211
positive	1016
Total	**1507**

4.6.2 Visualization and Dashboard Development

After the trained model was applied to the entire dataset and sentiment labels were predicted for all records, a sentiment analysis results table was generated. The distribution of sentiment categories was then visualized using pie charts to provide an intuitive overview of the sentiment proportions. An example of the sentiment visualization is presented in Figure 4.

DOI: https://doi.org/10.31004/riggs.v4i4.4698

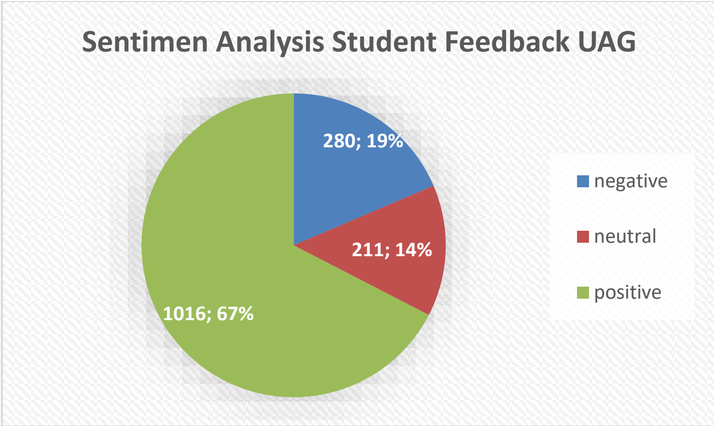

Figure 4 Sentiment Visualization

To further support exploratory analysis, a simple dashboard was developed using spreadsheet-based tools, enabling filtering and comparison of sentiment distributions based on academic semester, lecturer, and course. This interactive visualization approach facilitates stakeholders in identifying sentiment trends and patterns across different academic contexts, thereby supporting data-driven decision-making. An example of the developed dashboard interface is presented in Figure 5, illustrating the interactive visualization of sentiment analysis results.

Figure 5 Screencapture of the simple dashboard

4.6.3 Topic Modeling Using LDA for Negative Sentiment

The next step involved applying topic modeling using the Latent Dirichlet Allocation (LDA) method, specifically on feedback classified as negative, in order to gain deeper insights into the underlying factors contributing to student dissatisfaction.

LDA-based topic modeling was conducted on 280 records of negative sentiment. The optimal number of topics was determined using the coherence score as an evaluation criterion. The best-performing model achieved a coherence value of 0.39, indicating that the extracted topics exhibit a reasonably consistent semantic structure. The resulting LDA model generated five topics, each represented by the top eight most significant words. The extracted topics are illustrated in Figure 6.

Figure 6 Topic Modeling Results with LDA

DOI: https://doi.org/10.31004/riggs.v4i4.4698

4.6.3 Topic Modeling Using LDA on Negative Sentiment

The next stage of the deployment phase involved conducting topic modeling using Latent Dirichlet Allocation (LDA) specifically on feedback classified as negative sentiment. This step was performed to gain deeper insights into the underlying factors that contribute to student dissatisfaction.

LDA topic modeling was applied to 280 records of negative sentiment feedback. The optimal number of topics was determined based on the coherence score, which is commonly used to evaluate the semantic consistency of generated topics. The best-performing model produced a coherence score of 0.39, indicating that the extracted topics were reasonably coherent and interpretable. The final model generated five topics, each represented by eight dominant keywords, as illustrated in Figure 4.5.

4.6.3.1 Topic 0: Course Scheduling and Material Organization

Dominant keywords such as schedule, material, examples, and future indicate student concerns related to:

- inconsistent or frequently changing class schedules,
- learning materials that are perceived as poorly structured,
- insufficient examples provided during lectures, and
- expectations for improvement in future course delivery.

Interpretation:
Students express the need for better organization of learning materials and more stable, well-managed class schedules.

4.6.3.2 Topic 1: Difficulty Understanding Course Content

Keywords including material, understand, direct, and class reflect issues such as:

- difficulties in comprehending course materials,
- a lack of interactive learning processes,
- the need for direct explanations or demonstrations, and
- classes that do not adequately support student understanding.

Interpretation:
Students struggle to understand course content due to teaching methods that are perceived as insufficiently effective or engaging.

4.6.3.3 Topic 2: Workload and Lecturer Responsiveness

The appearance of words such as lecturer, assignments, class, and offline suggests concerns regarding:

- an excessive number of assignments,
- expectations related to lecturer performance,
- requests for more offline (face-to-face) classes, and
- issues in lecturer–student communication.

Interpretation:
Students perceive the assignment workload as overly demanding and expect improved responsiveness and communication from lecturers.

4.6.3.4 Topic 3: Alignment Between Course Content and Learning Objectives

Keywords such as course, subject, suggestions, students, and materials indicate:

- misalignment between course materials and course objectives,
- ineffective use of class time, and
- a need for better-structured lecture sessions.

DOI: https://doi.org/10.31004/riggs.v4i4.4698

Interpretation:
There are issues related to the alignment of instructional materials with course objectives and the overall structure of lecture sessions.

4.6.3.5 Topic 4: Lack of Practical Learning Components

Dominant terms including material, practice, learning, and understanding highlight:

- difficulties in understanding theoretical concepts,
- learning processes that are overly theoretical,
- a strong demand for more practical activities, and
- unclear explanations during lectures.

Interpretation:
Students express a strong preference for more practice-oriented learning, as theoretical explanations alone are perceived as insufficient.

4.6.3.6 Summary of Topic Analysis

Based on the five extracted topics, the primary sources of student dissatisfaction can be summarized as follows:

1. Quality of material delivery
2. Difficulty in understanding learning content
3. Excessive assignment workload
4. Lecturer performance and communication
5. Lack of practical and interactive learning activities

Overall, the findings indicate that students expect learning experiences that are:

- more structured,
- more interactive,
- more practice-oriented,
- easier to understand, and
- supported by better scheduling and classroom management.

4.6.4 Formulation of Strategic Recommendations

Based on the sentiment analysis and topic modeling results, several strategic recommendations are proposed to address the identified issues and improve service quality at Universitas Ary Ginanjar:

1. Develop standardized and well-structured learning modules complemented by practical laboratory or case-based materials.
2. Provide training programs for lecturers focused on interactive and student-centered teaching methods.
3. Strengthen practice-oriented learning by incorporating real-world cases and applied exercises.
4. Improve lecture scheduling and classroom management to ensure consistency and effectiveness.
5. Offer additional remedial or support classes for students who experience difficulties in understanding course materials.

These recommendations are expected to support continuous improvement efforts and enhance the overall learning experience for students.

DOI: https://doi.org/10.31004/riggs.v4i4.4698

4. Conclusions and Recommendation

This study investigated student perceptions of campus services at Universitas Ary Ginanjar (UAG) by analyzing textual student feedback using Natural Language Processing (NLP) techniques. Based on the results, several conclusions can be drawn in accordance with the research questions.First (RQ1), the analysis reveals that students' perc eptions of campus services are heterogeneous, reflecting varying levels of satisfaction and dissatisfaction across academic and non-academic aspects. Student feedback indicates that perceptions are shaped not only by instructional quality but also by supporting factors such as class management, clarity of material delivery, lecturer–student interaction, scheduling, and the relevance of course content to practical applications. This finding suggests that student perceptions are formed holistically through their overall academic experience. Second (RQ2), sentiment classification using an IndoBERT-based NLP model demonstrates strong performance in categorizing student feedback into positive, neutral, and negative sentiments. The model achieved a high overall accuracy and performed particularly well in identifying positive and negative sentiments, indicating that explicit opinions expressed by students can be effectively detected. However, the neutral sentiment class exhibited relatively higher misclassification rates, reflecting linguistic ambiguity and overlap between neutral and other sentiment expressions. The sentiment distribution shows that 67% (1,016 records) of the feedback is positive, 19% (280 records) is negative, and 14% (211 records) is neutral, indicating that the majority of students are generally satisfied with campus services. Third (RQ3), the application of Latent Dirichlet Allocation (LDA) topic modeling to negative sentiment feedback successfully uncovered key themes underlying student dissatisfaction. The dominant topics are related to unstructured course delivery, difficulties in understanding learning materials, limited classroom interaction, excessive assignment workload, insufficient alignment between theory and practice, and issues in scheduling and class management. These findings indicate that negative sentiments are primarily associated with pedagogical and managerial aspects of the learning process. Finally (RQ4), based on the identified negative sentiment themes, several strategic recommendations can be formulated for institutional improvement. The findings emphasize the need for systematic enhancement of teaching quality through better instructional planning, more interactive and practice-oriented learning approaches, and improved lecturer pedagogical competencies. Additionally, improvements in scheduling and class management are essential to foster a more conducive learning environment. Overall, the study demonstrates that NLP-based sentiment analysis using modern language models such as IndoBERT provides an effective and scalable approach for data-driven evaluation of academic services in higher education institutions.

Reference

[1] F. Koto, F. P. Rahmaningtyas, and R. Mahendra, "IndoBERT: A pretrained language model for Indonesian," *arXiv preprint arXiv:2009.05387*, 2020.

[2] R. D. Septiani and A. N. Hidayanto, "Analisis Sentimen Mahasiswa Terhadap Sistem Informasi Akademik Menggunakan Naïve Bayes," *Jurnal Sistem Informasi*, vol. 16, no. 1, pp. 1–8, 2020.

[3] D. A. Putri and Z. Fatimah, "Sentiment Analysis of Student Feedback to Improve Academic Services," *Journal of Information Systems Engineering and Business Intelligence*, vol. 7, no. 1, pp. 32–40, 2021, doi: 10.20473/jisebi.v7i1.23037.

[4] R. M. Sari and P. Lestari, "Implementasi Analisis Sentimen dalam Evaluasi Layanan Akademik Mahasiswa," *Jurnal Teknologi Informasi dan Komunikasi Pendidikan*, vol. 11, no. 2, pp. 102–112, 2023.

[5] Y. Utami and R. A. Nugroho, "Penggunaan Text Mining untuk Menilai Kepuasan Mahasiswa terhadap Fasilitas Kampus," *Jurnal Sistem Informasi dan Komputerisasi Akuntansi*, vol. 8, no. 2, pp. 87–95, 2019.

[6] R. Albalawi, T. H. Yeap, and M. Benyoucef, "Using topic modeling methods for short-text data analysis," *Information Processing & Management*, vol. 60, no. 1, 103167, 2023.

[7] Y. Guo, X. Zhang, and C. Lu, "Natural Language Processing in Artificial Intelligence," *ACM Computing Surveys*, vol. 53, no. 5, pp. 1–35, 2020.

[8] W. Medhat, A. Hassan, and H. Korashy, "Sentiment analysis algorithms and applications: A survey," *Applied Computing and Informatics*, vol. 17, no. 1, pp. 1–19, 2021.

[9] A. Ratner et al., "Snorkel: Rapid training data creation with weak supervision," *VLDB Journal*, vol. 29, no. 2, pp. 1–23, 2020.

[10] B. Wilie et al., "IndoNLU: Benchmark and resources for evaluating Indonesian natural language understanding," in *Proc. AACL-IJCNLP*, 2020, pp. 843–857, doi: 10.18653/v1/2020.aacl-main.85.

DOI: https://doi.org/10.31004/riggs.v4i4.4698

[11] M. Afifuddin, D. P. Sari, and A. Nugroho, "Sentiment analysis on Indonesian Twitter data using IndoBERT," *International Journal of Advanced Computer Science and Applications*, vol. 13, no. 5, pp. 225–232, 2022.

[12] R. A. Pratama and P. Sari, "Comparative study of BERT and LSTM for Indonesian sentiment analysis," *Procedia Computer Science*, vol. 179, pp. 451–458, 2021.

[13] F. M. Sari, S. Dewi, and E. Prasetyo, "Improving sentiment classification of Indonesian reviews using IndoBERT and data augmentation," *Indonesian Journal of Artificial Intelligence and Data Science*, vol. 2, no. 2, pp. 89–97, 2023.

[14] F. Oktaviani, R. Ramadhani, and A. Putri, "Fine-tuning IndoBERT for academic feedback sentiment analysis," *Journal of Educational Data Mining*, vol. 16, no. 1, pp. 55–68, 2024.

[15] Berrar, D. (2023). Performance measures and model evaluation in machine learning. *Pattern Recognition Letters, 165*, 77–85. https://doi.org/10.1016/j.patrec.2022.12.002

[16] Géron, A. (2022). *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow* (3rd ed.). O'Reilly Media.

[17] Raschka, S., Liu, Y., & Mirjalili, V. (2022). *Machine Learning with PyTorch and Scikit-Learn*. Packt Publishing.

[18] Sarker, I. H. (2021). Machine learning: Algorithms, real-world applications and research directions. *SN Computer Science, 2*(3), 160. https://doi.org/10.1007/s42979-021-00592-x

DOI: https://doi.org/10.31004/riggs.v4i4.4698