Indonesian Journal of Data and Science

Volume 6 Issue 3
ISSN 2715-9936
https://doi.org/10.56705/ijodas.v6i3.377

Research Article

Sentiment Analysis of Student Comments on Facilities and Infrastructure at Instiki Using Retrieval Augmented Generation

Ni Putu Juliana Dewi [1,*], I Kadek Dwi Gandika Supartha [2], I Putu Yoga Indrawan [3], Ketut Jaya Atmaja [4]

[1] Institut Bisnis dan Teknologi Indonesia, Bali 80225, Indonesia, dewijuliana750@gmail.com
[2] Institut Bisnis dan Teknologi Indonesia, Bali 80225, Indonesia, gandika.supartha@instiki.ac.id
[3] Institut Bisnis dan Teknologi Indonesia, Bali 80225, Indonesia, yoga.indrawan@instiki.ac.id
[4] Institut Bisnis dan Teknologi Indonesia, Bali 80225, Indonesia, ketutjayaatmaja@instiki.ac.id

Correspondence should be addressed to Ni Putu Juliana Dewi; dewijuliana750@gmail.com

Received 18 October 2025; Accepted 28 December 2025; Published 31 December 2025

Abstract:
This research was conducted to analyze the sentiment of student comments on infrastructure facilities at the Indonesian Institute of Business and Technology (INSTIKI) to overcome the problem of comment analysis that was previously done manually. The data used is in the form of student comments in 2024. The method used in this study is Retrieval Augmented Generation (RAG) with data labeling using Lexicon-Based. The test was carried out on three Large Language Models (LLMs), namely indobenchmark/indobert-base-p1, TinyLlama/TinyLlama-1.1B-Chat-v1.0, and w11wo/indonesian-roberta-base-sentiment-classifier. The test results showed that the indobenchmark/indobert-base-p1 model produced the highest accuracy of 80% in both test sessions compared to other models. The TinyLlama/TinyLlama-1.1B-Chat-v1.0 model produced 60% accuracy in session 1 and 65% in session 2, while the w11wo/indonesian-roberta-base-sentiment-classifier model produced 60% accuracy in both test sessions. The difference in the performance of these three LLMs shows that the model's understanding of Indonesian can affect the results of sentiment predictions.

Keywords: Sentiment Analysis; Facilities and Infrastructure; Retrieval Augmented Generation (RAG); Lexicon-Based; Student Comments.

1. Introduction

The availability of campus facilities plays a role in increasing student satisfaction, which in turn can positively impact learning motivation. In the context of higher education, the quality of campus facilities, along with the quality of teaching, are important factors influencing students' motivation to learn [1]. Campus facilities include various physical facilities and infrastructure provided by the university to support both academic and non academic student activities. These include adequate lecture halls, laboratories to support practical activities, a library with a comprehensive collection, sports facilities, and other supporting facilities such as meeting rooms and recreation areas for students to interact and socialize [2].

The Indonesian Institute of Business and Technology (INSTIKI), established on April 18, 2008 under the Wahana Widya Wisesa Denpasar Foundation, is a university in Bali that focuses on information technology. The development of campus facilities is essential to support the learning process and enhance student comfort and satisfaction. As the main users, students are directly affected by the condition of campus facilities and infrastructure. Currently, INSTIKI provides various facilities to support academic activities, including laboratories, study rooms, libraries, Wi-Fi, sports facilities, parking areas, canteens, and other supporting amenities.

Students play an important role in providing input in the form of criticism and suggestions to support improvements to educational facilities. To provide criticism and suggestions from students regarding campus facilities and infrastructure, INSTIKI provides a campus facility in the form of a questionnaire that will be at the end of each semester, students can provide comments regarding the condition of available campus facilities. Based on the results of interviews conducted with Ir. Anak Agung Gede Bagus Ariana, S.T., M.T., as Head of the Quality Assurance Department of the Faculty of Informatics Technology, information was obtained that in analyzing questionnaire comments still uses a manual method, where student comments in Excel are read one by one to be grouped based on the category of existing facilities and infrastructure. This method has a drawback because it requires a long time to analyze thousands of comments related to available facilities and infrastructure. This problem can be overcome by applying data mining technology, in this case, sentiment analysis. Through student comments, it is necessary to conduct sentiment analysis to determine student assessments of facilities and infrastructure on the INSTIKI campus and so that in the future it can help the institution in making improvements to existing facilities and infrastructure.

Sentiment analysis, also known as opinion mining or emotional intelligence, is a method for gathering and extracting information from unstructured text data. The goal of sentiment analysis is to identify and determine the tendency of opinions contained in user-generated text, whether they are positive, neutral, or negative [3], [4], [5], [6]. Based on comments from students regarding the facilities and infrastructure of the INSTIKI, this sentiment analysis is used to classify existing opinions. To find out the sentiments contained in a text, labeling with Lexicon-Based is used.

The lexicon-based approach is a method in sentiment analysis that uses a collection of sentiment words that have been assigned specific weights. This lexicon contains emotionally charged words and negation words, where each word represents a sentiment polarity, namely positive or negative. Once the lexicon is determined, words in the analyzed text are matched with entries in the sentiment lexicon. Next, the weights of the identified sentiment words are summed to obtain a sentiment score, which is used as the basis for classifying the text's sentiment into positive, negative, or neutral [7], [8].

Retrieval Augmented Generation (RAG) is an approach that aims to improve a models ability to generate text for high knowledge tasks by leveraging relevant information obtained from external data sources. By incorporating this information into the generation process, RAG can produce more precise and contextual output [9]. Therefore, by using the Retrieval Augmented Generation (RAG) method, the system will be able to identify the sentiments contained in a comment.

Based on the description above, sentiment analysis is necessary to determine whether student sentiment tends to be positive, negative, or neutral regarding INSTIKI infrastructure. Therefore, the researcher is interested in conducting a sentiment analysis of INSTIKI student comments to gain a better understanding of student opinions regarding INSTIKI infrastructure. This information can then be used as input for improving or developing infrastructure.

.

2. Method

Retrieval Augmented Generation (RAG) is the method used in this study to conduct sentiment analysis. The research data in the form of comments is then processed through the preprocessing stage. A Lexicon-Based approach is then used to label comment data based on sentiment and group it into negative, positive, and neutral categories. Once labeled sentiment, the comment data is stored as a Knowledge Base (KB) and analyzed using RAG. During the retrieval phase, the system retrieves relevant comments from KB and presents them as context for the Large Language Model (LLM). Based on the given context, the LLM is then used to generate sentiment from the input comments. There are three Large Language Models (LLMs) used, namely, indobenchmark/indobert-base-p1, TinyLlama/TinyLlama-1.1B-Chat-v1.0, and w11wo/indonesian-roberta-base-sentiment-classifier.

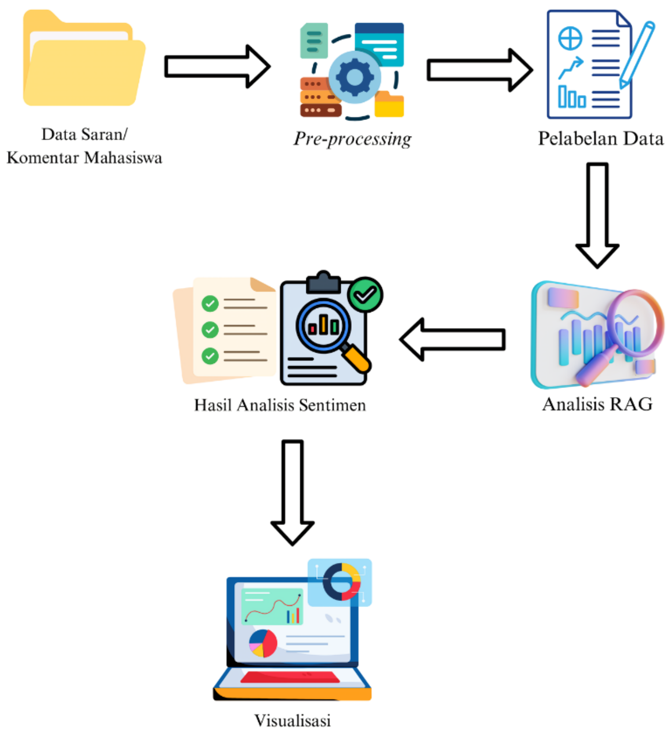

Figure 1. System Overview

Figure 1 is an overview of the system for sentiment analysis, it can be seen that comment data will be input into Google Colab and will be preprocessing to clean the data. After preprocessing, the data will be labeled using Lexicon-Based and analyzed using the Retrieval Augmented Generation (RAG) method to find out the sentiment labels contained in the text. After obtaining the results of the sentiment analysis, data visualization will be carried out so that the information contained in the document can be read and easy to understand.

Student Suggestions/Comments Data

Data collection was conducted through interviews with Mr. Agung Ariana. The resulting dataset consisted of 2,394 student comments.

Preprocessing Data

Data preprocessing is one of the major steps in the data mining process, it goes through various steps such as data cleaning, data integration, data selection, and data information [10]. This step plays a crucial role in the sentiment analysis process, as it standardizes the text data and eliminates irrelevant or noisy elements [11]. The preprocessing steps carried out in this study are as follows:

Case Folding

Case folding is a process that transforms all characters in a word into a consistent letter case, most commonly lowercase [12]. This step is the process of converting all capital letters in a document to lowercase to handle all terms consistently in text analysis, as text varies in its use of letters. This is done to improve the accuracy of data processing and analysis which will also make it easier to find and organize related data [13].

Cleaning

Cleaning is a process that involves identifying, correcting, or eliminating errors, inconsistencies, and missing values in a dataset [14]. In this process, The elements removed can include HTML tags, URLs, numbers, special characters, emojis, or excess whitespace. For example, data scraped from websites or social media often contains links, hashtags, or metadata that can interfere with the analysis process. By removing these elements, the text becomes cleaner, easier to process, and focuses more on meaningful content. Furthermore, text cleaning also helps reduce the amount of data to be processed, thereby improving the efficiency of machine learning models [7]. It's as simple as removing components that don't have any significance in a document.

Normalization

Normalization is a crucial step in Natural Language Processing (NLP), transforming raw text data into a same form that is easily processed by computer systems. Natural language text is generally unstructured and contains various variations, such as the use of informal language and differences in word forms. Therefore, before carrying out computational processes such as text analysis, machine translation, or information retrieval, the text must first be normalized. Through this process, words with the same meaning or root can be treated consistently [7].

Tokenizing

Tokenization is the goal of this process is to divide sentences into smaller pieces called "tokens", whether that is words or phrases [15]. This step is crucial because NLP models are unable to process entire texts as single units. By breaking text into manageable components, tokenization allows computational systems to analyze and process language data more effectively [7].

Filtering

Filtering or Stopwords are function words that generally carry little semantic meaning and do not express sentiment. However, they frequently appear in texts and should be removed to reduce dimensionality, lower computational costs, and improve system performance [16]. Removing stopwords helps minimize noise in the data and allows the analysis to focus on more meaningful content. In tasks such as information retrieval or sentiment analysis, stopwords can mask the true intent or emotional tone of a sentence. Their removal enables algorithms to better emphasize words that directly convey meaning [7]. Examples of stopword removal are the words "and", "that", "in", and others.

Stemming

Stemming is is the process of changing affixed words into root words [17]. The goal is to ensure that multiple forms of words with similar meanings are treated as one word when analyzed. Stemming involves removing affixes from words using specific rules, sometimes resulting in words that are not standard words [7].

Data Labeling

Lexicon-based approaches to sentiment analysis focus on building a sentiment lexicon, a collection of words containing sentiments and scores for each sentiment. The lexicon is constructed by selecting emotionally charged words and negation words. Each word in the lexicon is assigned a weight representing the strength and polarity of its sentiment, whether positive or negative. Next, words in the analyzed text are matched with words in the sentiment lexicon. The weights of the found sentiment words are then calculated and summed to obtain a sentiment score, which is used as the basis for determining the sentiment category of the text, whether positive, negative, or neutral [7], [8]. **Table 1** is some examples of comment data results that have gone through the Lexicon-Based labeling stage.

Table 1. Data Labeling Results

Comments	Label
mohon fasilitas parkir luas mahasiswa susah cari parkir hujan parkir banjir licin bahaya mahasiswa kali mahasiswa jatuh sana ban motor gelincir	negative
kipas lounge baik panas banget terimakasih	positive
luas area parkir tingkat baru tinggi bahaya gempa	neutral
kamar mandi kadang air mati closetnya kamar mandi bagus cuma baak tolong perhati toilet gedung terima kasih	negative

Comments	Label
ruang kelas mohon tingkat duduk meja kuat topang wifi mohon tingkat kualitas kencang	positive

RAG Analysis

Data that has been labeled using Lexicon-Based will produce a knowledge base (KB) that is used in the RAG process. Retrieval Augmented Generation (RAG) is a hybrid approach that combines information retrieval techniques with generative language models to enhance the accuracy and relevance of generated responses. This method improves the capability of Large Language Models (LLMs) by retrieving pertinent information from external sources, allowing the models to produce more factual outputs. By integrating external knowledge memory or Knowledge Based (KB), RAG helps overcome the inherent limitations of standalone generative models. The retrieved contextual information is incorporated into the models context window, enabling the generation of responses that are more accurate, up-to-date, and contextually grounded [18], [19].

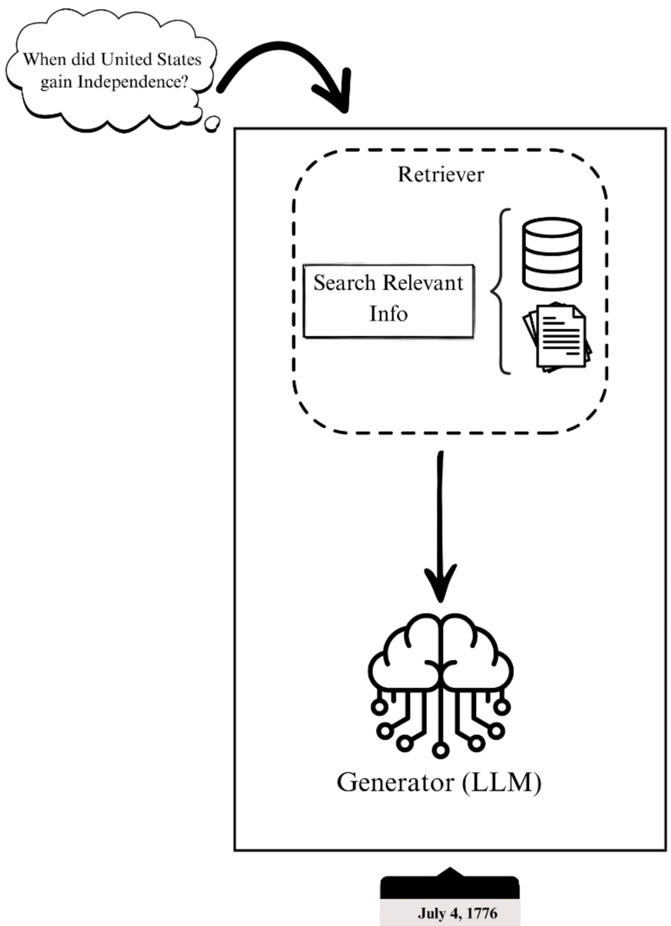

Figure 2. RAG Architecture

Figure 2 illustrates how the Retrieval Augmented Generation (RAG) system works in answering questions. Questions entered by the user are processed by the retriever component, which searches for relevant information from external data sources that serve as a knowledge base. The search results are then fed into the generator component, a natural language based generative AI model. This model combines the retriever's results with the generative AI model's ability to generate answers. The generative model generates an accurate answer, “July 4, 1776” based on the information gathered during the retriever process.

Visualization

In order for the data from the analysis to be easily understood, this study conducts data visualization, where visualization means expressing ideas or feelings using the form of images, writings (words and numbers), maps, graphs, and so on. Data visualization is a method for quickly understanding and interpreting data, as the human brain processes information presented visually more effectively than text. The primary goal of data visualization is to convey information clearly and efficiently, allowing users to easily identify patterns, trends, and insights contained within the data [20].

3. Result and Discussion

Data Preprocessing Results

Table 2. Data Preprocessing Results

Comments Before Preprocessing	Comments After Preprocessing
Area parkir belakang sepertinya kurang aman terutama pada saat musim hujan datang, yang dimana area turunan lumayan licin dan masih banyak area genangan air dan lumpur membuat sepatu mahasiswa agak kotor ketika menginjak lantai, sekian saja kritik saya Terima Kasih 🥰🥰🥰	*area, parkir, aman, musim, hujan, mana, area, turun, lumayan, licin, area, genang, air, lumpur, sepatu, mahasiswa, kotor, injak, lantai, sekian, kritik, terima, kasih*
Semoga fasilitas di instiki seperti internet bisa di gunakan secara lancar tanpa ada gangguan apapun dan selalu memenuhi kebutuhan belajar bagi mahasiswa	*moga, fasilitas, instiki, internet, lancar, ganggu, apa, penuh, butuh, ajar, mahasiswa*
Mohon parkir diperluas karena sangat sulit untuk mencari parkir dan juga jalan masuk belakang mohon dipasang paving agar lebih nyaman saat lewat terimakasih	*mohon, parkir, luas, sulit, cari, parkir, jalan, masuk, mohon, pasang, paving, nyaman, terimakasih*
TOLONG UNTUK FASILITAS KELAS SEPERTI AC DIPERBAIKI, PADA KELAS R324 AC MATI TOTAL KELAS PENGAP MENGGANGGU AKTIVITAS BELAJAR, MEMBUAT TIDAK NYAMAN DAN PUSING	*tolong, fasilitas, kelas, ac, baik, kelas, ac, mati, total, kelas, pengap, ganggu, aktivitas, ajar, nyaman, pusing*
Salam hormat untuk kampus tercinta INSTIKI, parkiran sangat sempit dan tidak nyaman tolong diperbesar dan diperbaiki parkirannya	*salam, hormat, kampus, cinta, instiki, parkir, sempit, nyaman, tolong, besar, baik, parkir*

Table 2 shows the results of preprocessing comment data before further analysis. Preprocessing steps such as case folding, cleaning, normalization, tokenizing, filtering, and stemming transform comments that originally consisted of capital letters, non-standard terms, and irrelevant sections into cleaner, more structured text. The purpose of this preprocessing is to facilitate data analysis during the sentiment labeling process and the subsequent application of methods.

Labeling Lexicon-Based

Table 3. Labeling Lexicon-Based

Lexicon Label Summary	
Label	**Amount of Data**
positive	1,129
negative	683
neutral	582

A summary of sentiment labeling results using a Lexicon-Based approach to comment data is shown in **Table 3**. Based on the labeling results, there were 1,129 comments with positive sentiments, 683 with negative sentiments, and 582 with neutral sentiments. These results illustrate the distribution of comment sentiment after the labeling process prior to further analysis. The results of this labeling will produce a Knowledge Base (KB) that functions as a source of knowledge that will later be used as a reference or reference by Retrieval Augmented Generation ((RAG) to analyze the sentiment contained in a comment.

Retrieval Augmented Generation (RAG)

In this study, test comments are considered new data or data that has not been seen during the test. The RAG system then uses the available knowledge base to perform the data retrieval process before generating sentiment labels. With this method, the test results show how well the RAG model can perform sentiment analysis on new comments without being affected by the previously identified data. The data analyzed was carried out in 2 sessions, namely 5 data in session 1 and 20 data in session 2.

Table 4. indobenchmark/indobert-base-p1 Model

KB sentiment	Original Text from Knowledge Base	RAG results
negatif	ada sedikit saran untuk parkir ke atas ya itu di upgrade lah pak tidak...	NEGATIF
negatif	fasilitas ya lengkap kadang *wifi*nya susah di akses...	
negatif	mungkin area parkir ya bisa di atur lebih luas apalgi kalau hujan jad...	
negatif	area parkir licin ketika hujan kadang parkiran susah dicari akses bela...	
negatif	parkirannya mungkin lebih luas agar tidak sulit untuk parkir dan untuk...	

Table 4 shows the results of the application of the Retrieval Augmented Generation (RAG) method using model 1, namely indobenchmark/indobert-base-p1. This model generates a classification of sentiment into negative categories using comments that have been preprocessed and labeled based on Lexicon-Based as context.

Table 5. TinyLlama/TinyLlama-1.1B-Chat-v1.0 Model

KB sentiment	Original Text from Knowledge Base	RAG results
negatif	*areal parkir yang kurang luas saat hujan sedikit licin...*	NEGATIF
negatif	*parkiran kurang memadai karena sekarang musim hujan tanjakan...*	
positif	*tempat parkir diperluas lagi jalan di belakang diperbaiki aga...*	
negatif	*parkirannya mungkin lebih luas agar tidak sulit untuk parkir...*	
negatif	*jalan menuju parkiran diperbaiki dan diperluas lagi karena...*	

Table 5 shows the results of the application of the RAG method using model 2 TinyLlama/TinyLlama-1.1B-Chat-v1.0. The system retrieves some relevant comments from the Knowledge Base (KB) and uses them as knowledge context so that the model can identify sentiment in the comments. Sentiment on the input comments is classified as negative by the TinyLlama/TinyLlama-1.1B-Chat-v1.0 model.

Table 6. w11wo/indonesian-roberta-base-sentiment-classifier Model

KB sentiment	Original Text from Knowledge Base	RAG results
negatif	*areal parkir yang kurang luas saat hujan hujan sedikit licin dan becek...*	NEGATIF
negatif	*parkirannya mungkin lebih luas agar tidak sulit untuk parkir dan untuk...*	
negatif	*saat hujan sangat sulit untuk parkir di lantai atas karena kondisi tan...*	
negatif	*parkiran kurang memadai karena sekarang musim hujan tanjakannya licin...*	
negatif	*untuk tempat parkir bagian depan sudah sangat baik tetapi untuk parkir...*	

Table 6 shows the results of the application of the RAG method using model 3, namely w11wo/indonesian-roberta-base-sentiment-classifier. The system finds 5 negatively labeled data contexts that are relevant to the input data, which will be used as a reference to support the sentiment classification model. The model classifies comments into negative categories.

Visualization

Confusion Matrix

The Confusion Matrix is used to display a visualization of the performance of the RAG method with several LLMs used in this study. Through this visualization, the prediction patterns generated by each LLMs model can be easily seen. The vertical axis (row) is the actual label of the comment data, while the horizontal axis (column) is the prediction label generated by the model. The following are the results of the evaluation of the 3 models used.

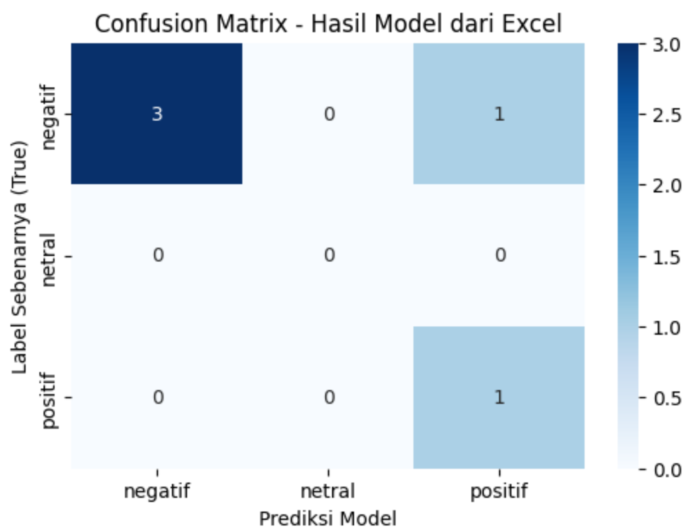

Figure 3. indobenchmark/indobert-base-p1 Model

Figure 3 shows the results of the evaluation visualized using the confusion matrix for the indobenchmark/indobert-base-p1 model resulting in an accuracy of 80%, precision of 90%, recall of 80%, and an f1-score of 81.9%. In the results of the tests carried out on this model, there were 3 data that had a negative label and the model correctly predicted as negative, 1 data that had a positive label and the model correctly predicted as positive, and 1 data that had a positive label, but the model predicted as negative.

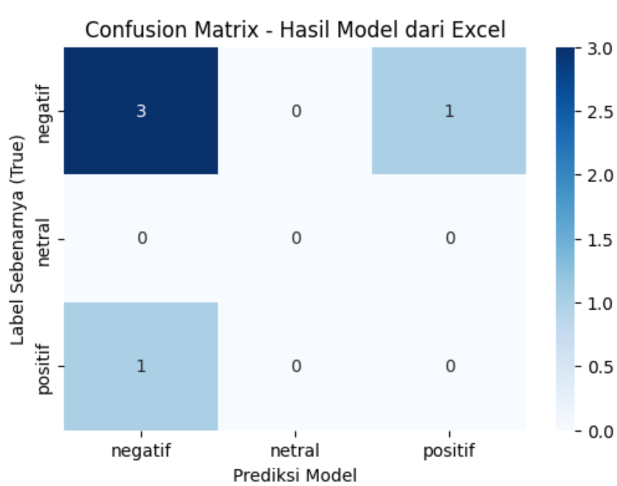

Figure 4. TinyLlama/TinyLlama-1.1B-Chat-v1.0 Model

The results of the evaluation visualized with the confusion matrix for the TinyLlama/TinyLlama-1.1B-Chat-v1.0 model can be seen in **Figure 4**. The results of the evaluation resulted in an accuracy, precision, recall, and f1-score of 60%. All evaluation metrics produce the same value due to the lack of test data, and then have the same amount of FN (False Negative) and FP (False Positive). Where FN is the data that is actually positive, but the model predicts as negative and FP is the data that is actually negative, but the model predicts as positive. For example, out of 5 data tested, FN and FP have the same amount of data, namely 1 data, then the results of the matrix evaluation, namely accuracy, precision, recall, and f1-score will have the same value.

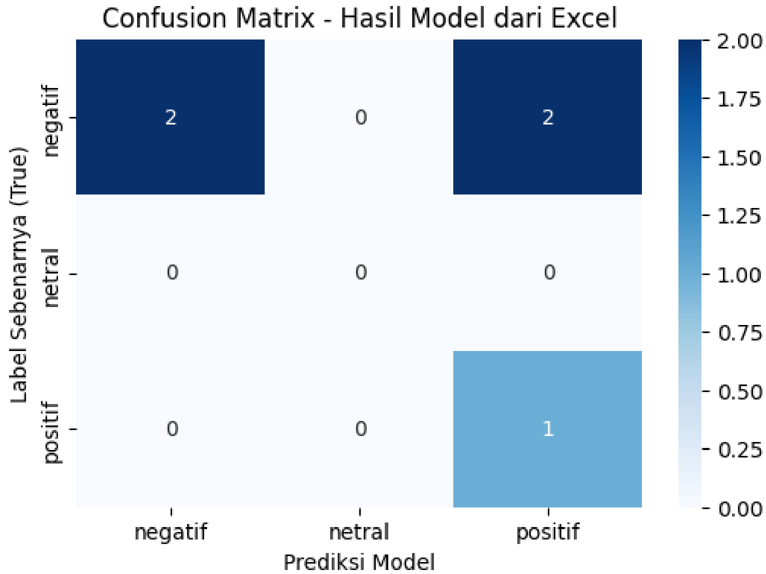

Figure 5. w11wo/indonesian-roberta-base-sentiment-classifier Model

Figure 5 shows the results of the evaluation using the w11wo/ indonesian-roberta-base-sentiment-classifier model visualized with a confusion matrix. This model yields a 60% accuracy value indicating that overall, the accuracy level for the classification is at an intermediate level, an accuracy of 86.7% indicating that the model tends to be accurate in predicting sentiment, a 60% recall indicating that the model has not identified a good amount of sentiment data, and a 63.3% f1-score.

Table 7. Comparison of Evaluation Results of 3 Large Language Models (LLMs)

Confusion Matrix Testing	Session 1			Session 2		
	Model 1	**Model 2**	**Model 3**	**Model 1**	**Model 2**	**Model 3**
Accuracy	80%	60%	60%	80%	65%	60%
Precision	90%	60%	86.7%	100%	100%	100%
Recall	80%	60%	60%	80%	65%	60%
F1-Score	81.9%	60%	63.3%	88.9%	78.8%	75%

Table 7 shows the results of the comparison of Accuracy, Precision, Recall, and F1-Score using the RAG method with 3 different LLMs, namely the indobenchmark/indobert-base-p1 model (Model 1), TinyLlama/TinyLlama-1.1B-Chat-v1.0 (Model 2), and w11wo/ indonesian-roberta-base-sentiment-classifier (Model 3). In the test session 1, the highest accuracy was possessed by the indobenchmark/indobert-base-p1 model which was 80%. In the test session 2, the highest accuracy was also possessed by the indobenchmark/indobert-base-p1 model, which was 80%. It can be concluded that in the tests carried out in session 1 with a total of 5 comment data and session 2 with a total of 20 comment data, the indobenchmark/indobert-base-p1 model was superior to other models.

WordCloud

WordCloud is used to display the words that appear most often in a dataset. WordCloud makes it easy to understand the data in terms of trends and patterns in comments. The data to be displayed is divided into three categories, namely Positive, Negative, and Neutral based on sentiment distribution so that WordCloud data can provide appropriate information.

Figure 6. WordCloud of Positive Sentiments

Figure 6 shows that words with larger appearances than others include "*bagus*" (good), "*saran*" (suggestion) , "*fasilitas*" (facilities), "*mahasiswa*" (students) and others. This indicates that these words frequently appear in comments, potentially indicating a positive assessment of the facilities available on campus. Furthermore, words with less prominent appearances indicate a lower frequency of appearance in comments.

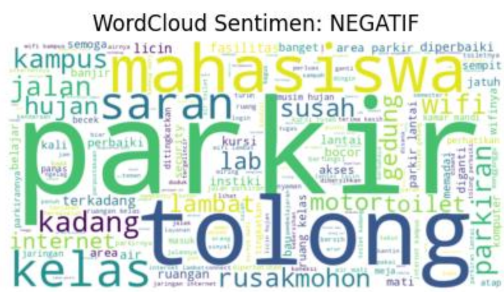

Figure 7. WordCloud of Negative Sentiments

Figure 7 clearly shows that the most frequently occurring words are prominent, such as "*tolong*" (please), "*kadang*" (sometimes), "*parkir*" (parking), "*susah*" (difficult), and so on. This indicates that these words appear most frequently in negative comments. The word "*parkir*" (parking) has the largest size, indicating a parking problem. Meanwhile, words that are less prominent indicate that they don't appear frequently in comments.

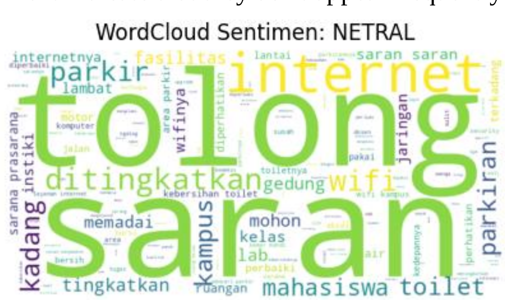

Figure 8. WordCloud of Neutral Sentiments

Figure 8 shows that the most prominent words are "*tolong*" (please) and "*saran*" (suggestion) indicating that these words have a high frequency of occurrence because they frequently appear in comments labeled as neutral sentiment. Meanwhile, the other words, which are less prominent and have a smaller size, indicate that they have a low frequency because they rarely appear in comments labeled as neutral sentiment.

Bar Chart

Bar charts show the percentage of data obtained for each sentiment. Bar charts are needed to show the number of positive, negative, and neutral sentiments that can be categorized using several LLM models. The following are bar charts for the three LLM models used in this study.

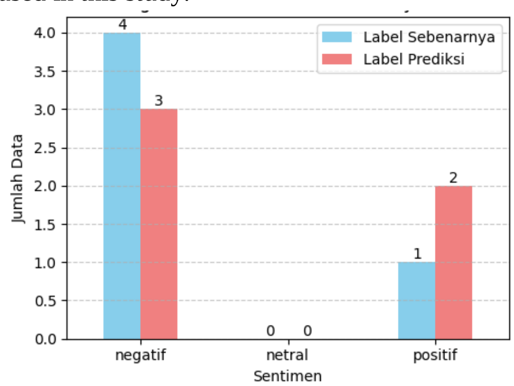

Figure 9. Bar Chart of indobenchmark/indobert-base-p1 model

Figure 9 shows the results of a comparison of sentiment data based on the predicted labels generated by the LLM model and the actual labels. Four data points had negative sentiment, but only three were predicted as negative by the model. One data point had positive sentiment, but two were predicted as positive by the model.

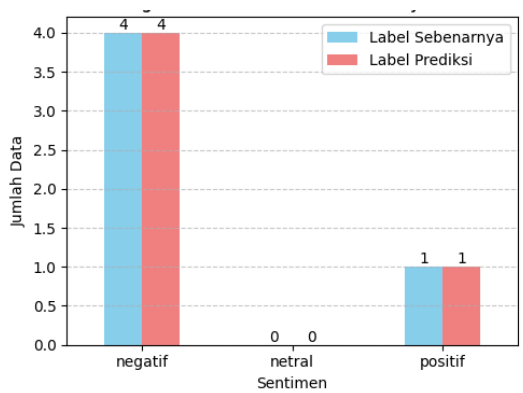

Figure 10. Bar Chart of TinyLlama/TinyLlama-1.1B-Chat-v1.0 model

Figure 10 shows the results of a comparison of sentiment data based on the predicted labels generated by the LLM model and the actual labels. Four data points have negative sentiment, and four data points were predicted as negative by the model. One data point has positive sentiment, and one data point was predicted as positive by the model.

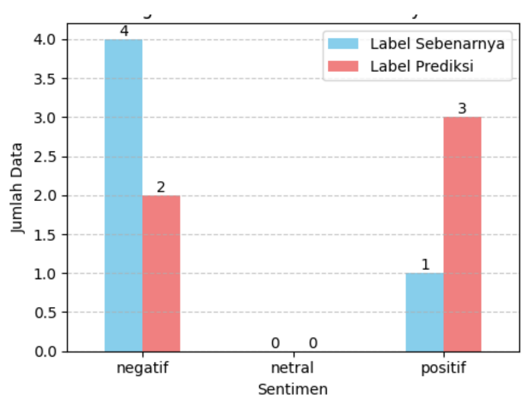

Figure 11. Bar Chart model w11wo/indonesian-roberta-base-sentiment-classifier

Figure 11 shows the results of a comparison of sentiment data based on the predicted labels generated by the LLM model and the actual labels. Four data points had negative sentiment, but only two were predicted as negative by the model. One data point had positive sentiment, but three were predicted as positive by the model.

4. Conclusion

Based on the results of the research that has been conducted, it can be concluded that the Retrieval Augmented Generation (RAG) method can be applied in the sentiment analysis process. The application of this method is carried out through data pre-processing and labeling stages using the Lexicon-Based method, so that the system is able to automatically classify student comments into negative, positive, and neutral sentiment categories. Furthermore, the performance of the RAG method shows different results for each Large Language Models (LLMs) used. The test results show that the indobenchmark/indobert-base-p1 model produces an accuracy of 80% in both test sessions, while the TinyLlama/TinyLlama-1.1B-Chat-v1.0 model produces an accuracy of 60% in the first session and 65% in the second session, and the w11wo/ indonesian-roberta-base-sentiment-classifier model produces an accuracy of 60% in both test sessions. These differences in performance indicate that the model's ability to understand Indonesian, especially mixed text and informal language, can affect the results of sentiment prediction.

Based on the evaluation results, the indobenchmark/indobert-base-p1 model was proven to have the highest evaluation score compared to other models. Further research is recommended to automatically process and analyze comments after the questionnaire is completed by integrating the Retrieval Augmented Generation (RAG) method into a real-time questionnaire system. This development will enable institutions to monitor and evaluate facility and infrastructure issues more quickly and sustainably.

References:

[1] S. Yahya, M. Aditya Domili, D. Sartika Tahir, M. Wahid, S. Basri, and B. Mandiri, "The Effect of Teaching Quality and Campus Facilities on Student Learning Motivation," *Eastasouth J. Learn. Educ.*, vol. 1, no. 02, pp. 36–43, Jul. 2023, doi: 10.58812/ESLE.V1I02.100.

[2] S. Nayma *et al.*, "The Effect of Campus Facilities Quality and Technology Support on Student Satisfaction," *J. Akuntansi, Manajemen, dan Perenc. Kebijak.*, vol. 2, no. 3, pp. 1–11, Jan. 2025, doi: 10.47134/JAMPK.V2I3.586.

[3] N. Farawahida Abdullah and N. Haizum Abd Rahman, "Consumer Acceptance and Perceptions of Electric Vehicles in Malaysia Using Sentiment Analysis," *J. Qual. Meas. Anal. JQMA*, vol. 21, no. 2, pp. 41–51, 2025, doi: 10.17576/jqma.2102.2025.04.

[4] D. A. Kristiyanti and S. Hardani, "Sentiment Analysis of Public Acceptance of Covid-19 Vaccines Types in Indonesia using Naïve Bayes, Support Vector Machine, and Long Short-Term Memory (LSTM)," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 7, no. 3, pp. 722–732, Jun. 2023, doi: 10.29207/RESTI.V7I3.4737.

[5] A. Erkan and T. Gungor, "Analysis of Deep Learning Model Combinations and Tokenization Approaches in Sentiment Classification," *IEEE Access*, vol. 11, pp. 134951–134968, 2023, doi: 10.1109/ACCESS.2023.3337354.

[6] M. Rodríguez-Ibánez, A. Casánez-Ventura, F. Castejón-Mateos, and P. M. Cuenca-Jiménez, "A review on sentiment analysis from social media platforms," *Expert Syst. Appl.*, vol. 223, p. 119862, Aug. 2023, doi: 10.1016/J.ESWA.2023.119862.

[7] G. Patil and V. Chaplot, *Natural Language Processing in Real-World Systems*. SK Research Group of Companies, 2025.

[8] A. Neustein, P. N. . Mahalle, P. Joshi, and G. R. Shinde, *AI, IoT, big data and cloud computing for Industry 4.0*. Springer, 2023.

[9] G. Xiong, Q. Jin, Z. Lu, and A. Zhang, “Benchmarking Retrieval-Augmented Generation for Medicine,” *Proc. Annu. Meet. Assoc. Comput. Linguist.*, pp. 6233–6251, 2024, doi: 10.18653/V1/2024.FINDINGS-ACL.372.

[10] V. Singh, V. K. Asari, S. Kumar, and R. B. Patel, *Computational Methods and Data Engineering: Proceedings of ICMDE 2020, Volume 1*. Springer Nature, 2020.

[11] K. L. Tan, C. P. Lee, K. M. Lim, K. L. Tan, C. P. Lee, and K. M. Lim, “A Survey of Sentiment Analysis: Approaches, Datasets, and Future Research,” *Appl. Sci. 2023, Vol. 13*, vol. 13, no. 7, Apr. 2023, doi: 10.3390/APP13074550.

[12] J. Zizka, F. Darena, and A. Svoboda, *Text mining with machine learning : principles and techniques*. CRC Press, 2020.

[13] M. Isnan, G. N. Elwirehardja, and B. Pardamean, “Sentiment Analysis for TikTok Review Using VADER Sentiment and SVM Model,” *Procedia Comput. Sci.*, vol. 227, pp. 168–175, Jan. 2023, doi: 10.1016/J.PROCS.2023.10.514.

[14] K. Mahasivabhattu, *Ultimate Machine Learning With ml.net : Build, Optimize, and Deploy Powerful Machine Learning Models for Data-Driven Insights with ML.NET, A*. Orange Education Pvt. Ltd, 2024.

[15] M. Yousef and A. ALali, “Analysis and Evaluation of Two Feature Selection Algorithms in Improving the Performance of the Sentiment Analysis Model of Arabic Tweets,” *Int. J. Adv. Comput. Sci. Appl.*, vol. 13, no. 6, pp. 705–711, Jun. 2022, doi: 10.14569/IJACSA.2022.0130683.

[16] H. T. Duong and T. A. Nguyen-Thi, “A review: preprocessing techniques and data augmentation for sentiment analysis,” *Comput. Soc. Networks 2020 81*, vol. 8, no. 1, pp. 1-, Jan. 2021, doi: 10.1186/S40649-020-00080-X.

[17] N. Umar, M. A. Nur, T. Informatika, and H. Makassar, “Application of Naïve Bayes Algorithm Variations On Indonesian General Analysis Dataset for Sentiment Analysis,” *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 6, no. 4, pp. 585–590, Aug. 2022, doi: 10.29207/RESTI.V6I4.4179.

[18] P. J. K. Singh, *MASTERING RETRIEVAL-AUGMENTED GENERATION*. BPB PUBLICATIONS, 2025.

[19] D.-S. Huang, H. Chen, B. Li, and Q. Zhang, *Advanced Intelligent Computing Technology and Applications 21st International Conference, ICIC 2025, Ningbo, China, July 26-29, 2025, Proceedings, Part X*. Springer, 2025.

[20] Food and Agriculture Organization off the United Nations, *Farm Data Management, Sharing and Services for Agriculture Development*. FAO, 2021.