JURNAL LOCUS: Penelitian & Pengabdian
Volume 4 No. 9 September 2025
E-ISSN 2829-7334| P-ISSN 2829-5439
Hompage: https://locus.rivierapublishing.id/index.php/jl

Sentiment Analysis of Reviews on the ChatGPT Application using the Long Short-Term Memory Method

Dias Aziz Pramudita, Helmi Imaduddin, Synta Nur Afiana Azizah, Isnawati Muslihah
[1,2,3] Universitas Muhammadiyah Surakarta, Indonesia
[4]Institut Seni Indonesia, Indonesia
Email: dap207@ums.ac.id, hi776@ums.ac.id, a710230007@student.ums.ac.id, isnawati.muslihah@isi-ska.ac.id

Abstract:

The integration of Information and Communication Technology (ICT) in the era of Education 5.0 has significantly transformed human interaction and learning processes. A notable advancement in this context is the application of Artificial Intelligence (AI) in education, which offers substantial benefits but also presents challenges such as data privacy and security concerns. Chatbots based on the Generative Pretrained Transformer (GPT) model, such as ChatGPT, exemplify this AI application. This study aims to analyze user sentiment toward the ChatGPT application on the Google Play Store using the Long Short-Term Memory (LSTM) method. The research employed a quantitative approach with data preprocessing steps including data cleaning, tokenization, and normalization. A total of 193,749 reviews, with an average rating of 4.9, were considered, from which 8,927 comments were selected for analysis. The LSTM model achieved an accuracy of 76.8%, precision of 71%, recall of 70%, and an F1-score of 76%. The findings indicate that the ChatGPT application is generally well-received by users, suggesting that it can be effectively integrated into the educational curriculum. Proper implementation can enhance the learning process while maintaining the essential role of teachers, ensuring that AI serves as a supportive educational tool rather than a replacement.

Keywords— sentiment analysis, google play reviews, chatgpt, long short-term memory

Corresponding: Dias Aziz Pramudita
E-mail: dap207@ums.ac.id

INTRODUCTION

Information and Communication Technology (ICT) is considered a primary tool supporting the learning process. It has transformed the landscape of education with the emergence of Education 5.0, where ICT significantly influences human interaction and work methods (Dewanto, 2023; Minaswati, 2023; Misnawati Misnawati, 2023; Rochim, 2024; Sandy et al., 2023; Suparno, 2019). Artificial Intelligence (AI) is the simulation of human intelligence processes by machines, especially computer systems. These processes include learning (the acquisition of information and rules for using the information), reasoning (using rules to reach approximate or definite conclusions) and self-correction (Akbar, 2023; Karyadi, 2023; Merentek et al., 2023; Pratama et al., 2022; Susanto, 2023). The application of AI in education has provided many significant benefits in the learning process (Manongga et al., 2022). However, the existence of AI also raises a number of challenges, such as data privacy issues, gaps in technology access, and major risks related to data security. Therefore, it is important to consider ethical aspects and develop appropriate guidelines for the use of AI in educational contexts.

One of the outcomes of advancements in AI technology is the development of chatbots using the GPT (Generative Pretrained Transformer) model. This language model is trained on large datasets to generate text that closely resembles human language (Fahriza & Riza, 2023). The use of chatbots in education must be carefully managed and utilized wisely to mitigate potential negative impacts. Therefore, sentiment analysis is necessary to understand public responses and acceptance of chatbots in the ChatGPT application (Panjaitan & Manurung, 2022).

The use of sentiment analysis leveraging Natural Language Processing (NLP) is a process for analyzing the messages, opinions, views, or emotions contained in text or documents (Nurrohmat & SN, 2019). NLP encompasses a variety of language expressions, requiring identification and representation to understand the phrases conveyed in the text (Panikar et al., 2022). This research aims to analyze the sentiment of reviews on the ChatGPT application available on Google Play by developing a system using the Long Short-Term Memory (LSTM) method. This approach allows us to represent public acceptance of artificial intelligence, specifically ChatGPT, through positive and negative sentiments. The findings can subsequently be used to enhance the educational sector, particularly in the learning process.

The rapid development of artificial intelligence in today's technological era has changed many aspects, bringing new challenges and opportunities. The presence of artificial intelligence has brought significant efficiency improvements to various sectors, including education. Artificial intelligence like ChatGPT has great potential to facilitate the learning process more easily and effectively in the field of Education. However, the development of chatbots using models such as GPT (Generative Pre-trained Transformer) also carries the potential for negative impacts that need to be considered. The use of chatbots in education must be carefully regulated and managed wisely to reduce the potential for adverse impacts.

Therefore, the rapid growth in the field of artificial intelligence must be approached wisely, considering social aspects, ethics and long-term implications. Sentiment analysis aids in understanding public responses and acceptance of chatbots within the ChatGPT application. This understanding allows stakeholders in educational institutions to plan the integration of this technology into the curriculum, enhancing the quality of learning without diminishing the essential role of teachers in the educational process.

Previous studies have explored the application of AI in education, particularly regarding the use of chatbots. Wang, Zhang, and Liu (2023) highlighted how chatbots can enhance student engagement and learning effectiveness through adaptive AI-driven interactions; however, their study was limited to measuring academic performance without considering broader user sentiment. Similarly, Li and Chen (2022) focused on sentiment analysis of AI-based learning platforms to evaluate user experience, but their study was restricted to school-internal web-based platforms, providing limited insight into public perceptions on commercially available applications.

The aim of this research is to identify positive and negative sentiments of ChatGPT users as indicators of AI acceptance in education, and to provide recommendations for educational institutions on ethical and effective integration of AI technologies. Consequently, this study contributes not only to understanding public perceptions of AI, but also to the development of policies and strategies for AI implementation grounded in societal sentiment analysis.

RESEARCH METHOD

The research employed a quantitative approach with data preprocessing steps including data cleaning, tokenization, and normalization. The system generally consists of six main parts, namely data collection, preprocessing, sentence conversion, vector conversion, sentiment classification, and system accuracy testing (Panjaitan & Manurung, 2022). The initial stage in system architecture is to retrieve data through the data scraping process. Review data was collected from the ChatGPT app available on Google Play, which was then grouped into two types of sentiment: positive and negative. The next step is data pre-processing to ensure the review data is clean so that the classification and sentiment determination process is more accurate, which includes several steps such as Casefolding, data filtering, tokenization, and slangword conversion. After the preprocessing stage, the next step is to convert the sentences into a format suitable for system input. This process involves converting pre-processed data into a numeric representation. Next, the words that have been converted will be changed into vectors, with the vector values taken from the training results using the Indonesian Wikipedia corpus and the word2vec technique. The next stage is the process of classifying novel reviews using the Long Short-Term Memory (LSTM) model, used to identify review sentiments on new data or testing data. The final stage is to test the accuracy of the system for novel data reviews, which involves calculating accuracy, precision, recall, and f-measure values. All the processes that occur in the system can be seen with the explanation in Figure 1.

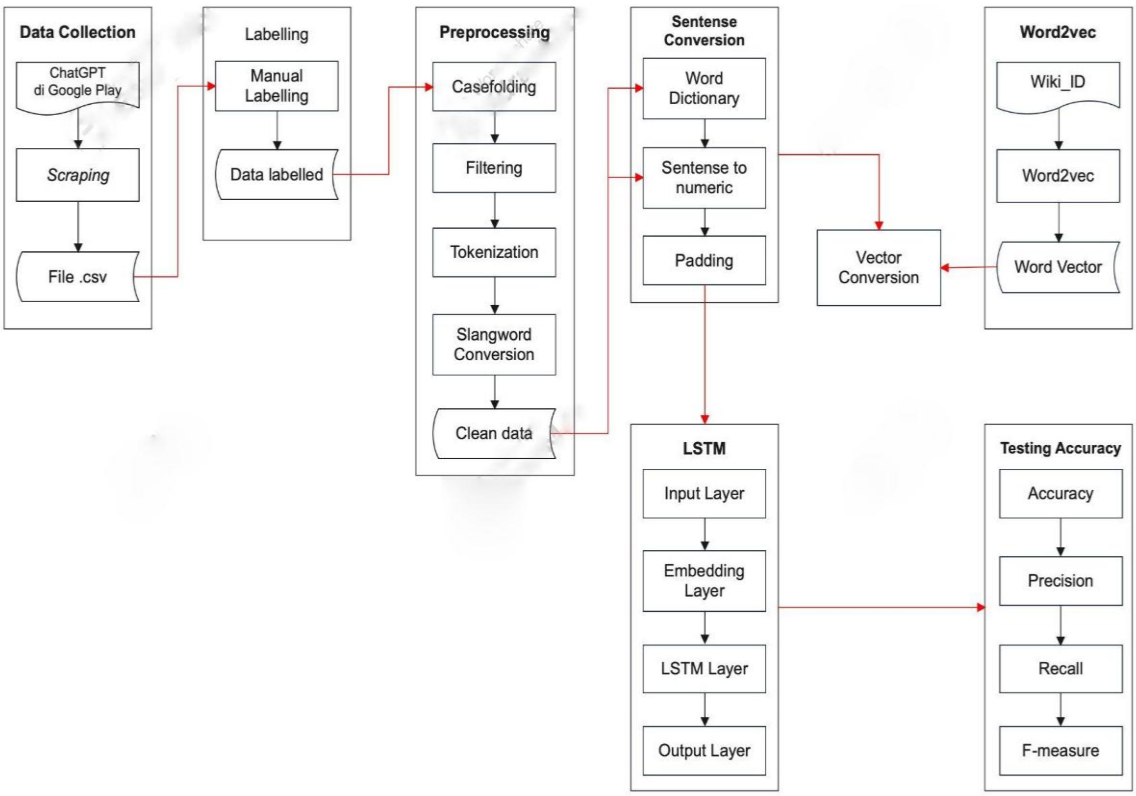

Figure 1. System Architecture

The data collected is in the form of reviews related to the ChatGPT application on Google Play with the site address: https://play.google.com/store/apps/details?id=com.openai.chatgpt

Data collection in this study was carried out by scraping the ChatGPT application available on Google Play. The scraping process is carried out using a library called google_play_scraper, with the help of the Python programming language on Google Colab, then using a transformer to carry out sentiment analysis on the reviews (Maulida et al., 2024). Furthermore, the process of saving scraping data into CSV format will automatically be saved in Google Drive. The steps in the ChatGPT application review scraping process are shown in Figure 2..

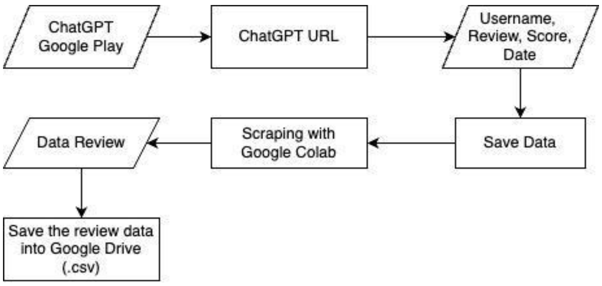

Figure 2. Scraping Process

Based on the results of data collection, there were 8,927 reviews collected on the ChatGPT application. This data is then manually labeled and divided into two sentiment categories: positive and negative. However, in this study, the number of positive reviews outnumbered negative reviews.

In order to address this class imbalance problem, undersampling and oversampling strategies have been employed. Under-sampling reduces the number of positive reviews to balance the data distribution, while over-sampling increases the number of negative reviews (Almuhaya et al., 2024). In this context, the under-sampling method was chosen to ensure each sentiment category had 1000 reviews, creating a balanced data distribution.

RESULTH AND DISCUSSION

This section discusses the results of sentiment classification testing of the model that has been built. Sentiment classification testing is carried out by measuring the accuracy, precision, recall, and f-measure values from the calculation of the Long Short-term Memory and Naïve Bayes methods.

The total review data used is 2000 data with 1000 data each for positive and negative sentiment. The training data used is 80% of the total data, which will be processed with k-fold. While 20% of the total data is used as testing data.

Classification testing is done by measuring the accuracy, precision, recall, and f-measure values obtained by comparing each manually labeled review with the results of the Long Short-term Memory method calculation carried out by the system. The number of reviews that match the results of the Long Short-term Memory method calculation by the system with manual labeling will affect the accuracy, precision, recall and f-measure values obtained. The greater the number of matching reviews, the higher the accuracy, precision, recall and f-measure values obtained. There are two Long Short-term Memory method architectures tested, namely LSTM 1 layer and LSTM 2 layers which will be compared with the naïve bayes method.

LSTM Classification testing with 1 layer

The results of the classification test of the ULASAN ChatGPT sentiment classification calculation using the Long Short-term Memory 1 layer method produced the best overall accuracy with an accuracy of 74.56% with the best testing parameters, namely the CBOW architecture, the number of neurons 150, the number of epochs 150, the L2 regularization value 0.2 and the sigmoid activation function.

LSTM Classification testing with 2 layers

The overall system test results of sentiment classification calculations using the 2-layers LSTM method show that the accuracy value is better than 1-layer LSTM. The overall accuracy value obtained is 76.80%. This means that increasing the number of layers in LSTM can increase the accuracy of sentiment classification.

The best overall accuracy results with an accuracy of 76.80% with the best testing parameters, namely the CBOW architecture, the number of neurons 100 per layer, the number of epochs 100, the L2 regularization value 0.2 and the sigmoid activation function.

Naive Bayes Classification Testing

Testing with the Naïve Bayes method uses the same data as the data tested in the Long Short-Term Memory method, namely 1600 training data (80% of data) and 400 test data (20% of data). The data has also gone through the same preprocessing process while the feature extraction used is TF (Term-Frequency). The training data will be processed using k-fold which is divided into 5 parts. The best Naive Bayes method test results with an accuracy of 70.50%.

Comparison of Accuracy Results

Comparison of the accuracy results of sentiment classification calculations using the LSTM 1 layer, LSTM 2 layers, and Naïve Bayes methods is shown in Table 3. Sentiment classification calculations using the Long Short-Term Memory method have better accuracy values than the Naïve Bayes method.

Table 1. Accuracy results

Method	Accuracy (%)	Precision (%)	Recall (%)	F-Measure (%)
LSTM 1 Layer	74.56	75	75	74
LSTM 2 Layer	76.80	77	76	76
Naïve Bayes	70.50	71	70	71

CONCLUSION

The research concluded that sentiment analysis of ChatGPT reviews can effectively be performed using the Long Short-Term Memory (LSTM) method, which outperforms the Naïve Bayes method in accuracy (76.80% vs. 70.50%), precision (77% vs. 71%), recall (76% vs. 70%), and F-measure (76% vs. 71%). Additionally, a 2-layer LSTM model yields better performance than a 1-layer LSTM, indicating that increasing the number of layers enhances accuracy, with the 2-layer model achieving 76.80% accuracy compared to 74.56% for the 1-layer. Future research could explore optimizing LSTM architectures further or applying other deep learning models to improve sentiment classification accuracy and adapt to evolving user feedback on AI applications.

REFERENCES

Akbar, J. S. (2023). Penerapan kecerdasan buatan (AI) dalam pembelajaran kimia. *Jurnal Pendidikan Sains, 7*(3).

Almuhaya, B., Saha, B., Kaur, M., Bazel, M. A., & Mohammed, R. (2024). Comparative analysis of machine learning algorithms for Arabic sentiment analysis on imbalanced social media data. In *2024 ASU International Conference in Emerging Technologies for Sustainability and Intelligent Systems (ICETSIS)* (pp. 1362–1367). https://doi.org/10.1109/ICETSIS61505.2024.10459389

Dewanto, A. C. (2023). Risiko dan mitigasi penggunaan kecerdasan buatan dalam bidang pendidikan. *Prosiding Konferensi Ilmiah Pendidikan, 4*(2018).

Fahriza, M. N., & Riza, N. (2023). Analisis sentimen pada ulasan aplikasi ChatGPT menggunakan metode klasifikasi K-Nearest Neighbor (KNN) systematic literature review. *Jurnal Mahasiswa Teknik Informatika, 7*(2).

Karyadi, B. (2023). Pemanfaatan kecerdasan buatan dalam mendukung pembelajaran mandiri. *Educate: Jurnal Teknologi Pendidikan, 8*(2).

Manongga, D., Rahardja, U., Sembiring, I., Lutfiani, N., & Yadila, A. B. (2022). Dampak kecerdasan buatan bagi pendidikan. *ADI Bisnis Digital Interdisiplin (ABDI Jurnal), 3*(2), 110–124.

Maulida, N., Suarna, N., & Prihartono, W. (2024). Analisis ulasan sentimen aplikasi mobile JKN dengan algoritma support vector machine berbasis particle swarm optimization.

Merentek, T. C., Usoh, E. J., & Lengkong, J. S. J. (2023). Implementasi kecerdasan buatan ChatGPT dalam pembelajaran. *Jurnal Pendidikan Tambusai, 7*(3).

Minaswati, M. (2023). ChatGPT: Keuntungan, risiko, dan penggunaan bijak dalam era kecerdasan buatan. *Prosiding Seminar Nasional Pendidikan, Bahasa, Sastra, Seni, dan Budaya, 42*(4).

Misnawati, M. (2023). ChatGPT: Keuntungan, risiko, dan penggunaan bijak dalam era kecerdasan buatan. *Prosiding Seminar Nasional Pendidikan, Bahasa, Sastra, Seni, dan Budaya, 2*(1). https://doi.org/10.55606/mateandrau.v2i1.221

Nurrohmat, M. A., & SN, A. (2019). Sentiment analysis of novel review using long short-term memory method. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems), 13*(3), 209. https://doi.org/10.22146/ijccs.41236

Panikar, R., Bhavsar, R., & Pawar, B. V. (2022). Sentiment analysis: A cognitive perspective. In *2022 8th International Conference on Advanced Computing and Communication Systems (ICACCS)* (pp. 1258–1262). https://doi.org/10.1109/ICACCS54159.2022.9785027

Panjaitan, N., & Manurung, N. (2022). Sentiment analysis of Google Classroom application using support vector machine. In *2022 6th International Conference on Electrical, Telecommunication and Computer Engineering (ELTICOM)* (pp. 117–120). https://doi.org/10.1109/ELTICOM57747.2022.10038262

Pratama, R. A., Supani, A., & Firdaus, A. (2022). Pemanfaatan media pembelajaran 3 dimensi untuk materi kecerdasan buatan dalam mata kuliah kecerdasan buatan. *Jurnal Laporan Akhir Teknik Komputer, 2*(1).

Rochim, A. A. (2024). Kecerdasan buatan: Resiko, tantangan dan penggunaan bijak pada dunia pendidikan. *Antroposen: Journal of Social Studies and Humaniora, 3*(1). https://doi.org/10.33830/antroposen.v3i1.6780

Sandy, F., Adi Palangi, W., Liling, D., Putra Pratama, M., Studi, P., Pendidikan, T., Keguruan, F., & Pendidikan, I. (2023). Impelentasi penggunaan kecerdasan buatan dalam pendidikan tinggi. *Seminar Nasional Teknologi Pendidikan UKI Toraja.*

Suparno, P. (2019). Menyikapi penggunaan artificial intelligence (AI, kecerdasan buatan) dalam pendidikan fisika. *Seminar Pendidikan Nasional.*

Susanto, E. (2023). Analisis implementasi kecerdasan buatan dalam pembelajaran. *Sindoro Cendekia Pendidikan, 1*(8).