International Journal of Computing and Digital Systems
ISSN (2210-142X)
Int. J. Com. Dig. Sys. #, No.# (Mon-20..)

# Leveraging ALBERT for Sentiment Classification of Long-Form ChatGPT Reviews on Twitter

**Wanda Safira[1], Benedictus Prabaswara[1], Andrea Stevens Karnyoto[2] and Bens Pardamean[1,2]**

*[1] Computer Science Department, BINUS Graduate Program - Master of Computer Science, Bina Nusantara University, Jakarta, Indonesia, 11480*
*[2] Bioinformatics and Data Science Research Center, Bina Nusantara University Jakarta, Indonesia, 11480*

*E-mail address: wanda.safira@binus.ac.id, benedictus.prabaswara@binus.ac.id, andrea.stevens@binus.edu, bpardamean@binus.edu*

*Received ## Mon. 20##, Revised ## Mon. 20##, Accepted ## Mon. 20##, Published ## Mon. 20##*

**Abstract:** Sentiment analysis of content created by users on social media sites reveals important information on public attitudes toward upcoming technologies. Researchers have challenges understanding these impressions, ranging from cursory evaluations to in-depth analyses. Focusing on detailed, long-form reviews exacerbates the difficulty in achieving accurate sentiment analysis. This research addresses the challenge of accurately analyzing sentiments in lengthy and unstructured social media texts, specifically focusing on ChatGPT reviews on Twitter. The study introduces advanced natural language processing (NLP) methodologies, including Fine-Tuning, Easy Data Augmentation (EDA), and Back Translation, to enhance the accuracy of sentiment analysis in lengthy and unstructured social media texts. The primary objective is to evaluate the effectiveness of the ALBERT transformer-based language model, in sentiment classification. Results demonstrate that ALBERT, when augmented with EDA and Back Translation, achieves significant performance improvements, with 81% and 80.1% accuracy, respectively. This research contributes to sentiment analysis by showcasing the efficacy of the ALBERT model, especially when combined with data augmentation techniques like EDA and Back Translation. The findings highlight the model's capability to accurately gauge public sentiments towards ChatGPT in the complex landscape of lengthy and nuanced social media content. This advancement has implications for understanding public attitudes towards emerging technologies, with potential applications in various domains.

***Keywords: Sentiment Analysis, AlBERT, Natural Language Processing, ChatGPT, Long-Form Review***

## 1. Introduction

Sentiment analysis of social media posts is crucial for understanding public perceptions and opinions on emerging technologies like ChatGPT [1], [2]. The tremendous amount of content users create on these platforms provides enormous helpful information for assessing public opinion. However, lengthy and unstructured text from platforms like Twitter poses challenges for accurate classification [3]. Texts containing long sentences make it difficult to analyze sentiment, leading to less accurate results. This research addresses sentiment analysis for texts with long sentences, utilizing advanced natural language processing (NLP) methodologies for improved accuracy.

Social media platforms are invaluable digital spaces connecting users and facilitating online social interactions [4], [5]. By providing a platform for users to communicate with each other, these digital spaces become vibrant hubs where individuals can not only stay current on news and information about today's world but also actively engage in 24-hour digital social exchanges [6], [7]. Through these dynamic networking platforms, people can virtually participate in a wide range of social activities and access up-to-date content from anywhere at any time, fostering a rich and interconnected online community [8], [9].

ChatGPT, short for Chat Generative Pre-trained Transformer, is a prominent model in natural language processing (NLP) developed by OpenAI. It is built on the Transformer architecture and has been trained on various internet materials to produce human-like responses to specified cues [10]. ChatGPT has gained attention for its capacity to comprehend and reply to natural language requests. It is an effective tool for a variety of NLP tasks, including question-answering, summarization, sentiment

*E-mail:author's email*

analysis, and more [11]. The model's versatility and potential applications have sparked significant interest and research in AI and NLP [12].

Twitter opinions can be categorized into positive, neutral, or negative sentiments through Sentiment Analysis. Sentiment Analysis is a Natural Language Processing (NLP) subfield that automatically classifies subjective text opinions. It detects if an opinion or viewpoint carries an underlying negative, positive, or neutral tone [13]. This is accomplished by applying NLP techniques to process the linguistic context of the text and determine its emotional inclination or polarity [14]. For example, when a user tweets a ChatGPT review, Sentiment Analysis can automatically evaluate whether the opinion is praiseful or critical based on the text composition. The system then outputs the results of its sentiment categorization as either a positive, negative, or neutral label on the opinion. This allows large volumes of tweets to be efficiently sorted and tagged by underlying sentiment [15].

The rise of transformer-based language models has marked a significant advancement in sophistication and accuracy [16]. Among these models, ALBERT: A Lite BERT for Self-Supervised Learning of Language stands out as a prominent contender, celebrated for its efficient parameter reduction techniques and robust performance across various NLP tasks [17]. Nevertheless, to understand ALBERT's effectiveness and competitive edge comprehensively, it is imperative to compare it with other transformer models such as BERT, XLNet, DistilBERT, and ELECTRA. ALBERT can outperform these models based on its architecture and reduce running time; therefore, the name Lite version of BERT. While ALBERT excels in optimizing efficiency without compromising performance, each model brings distinct strengths and innovations to the table [17]. BERT, renowned for its groundbreaking masked language modeling approach, has established a standard for contextual language comprehension [18]. XLNet, with its bidirectional context understanding, has garnered recognition for capturing intricate contextual dependencies [19]. DistilBERT offers a lightweight alternative with its compact architecture [20]. Meanwhile, ELECTRA introduces a fresh perspective on adversarial training through its discriminator model [21].

In this study, we contribute to evaluating the capability of the ALBERT model in conducting sentiment analysis effectively, especially when faced with datasets containing numerous words in each data instance. We utilize ALBERT, a leading deep learning model, based on numerous comparative studies highlighting its superior performance [22]. Additionally, we contribute to investigating various methodologies to improve the accuracy of the ALBERT model on datasets characterized by many lexical items per data entry. The methodologies we will explore include Fine-Tuning and Data Augmentation techniques such as Easy Data Augmentation and Back Translation [23]. Here are our main contributions can be outlined as follows:

- Utilization of ALBERT, a prominent deep learning model, based on comparative studies showcasing its superior performance.
- Investigation of methodologies to enhance ALBERT's accuracy on datasets with many lexical items per data entry.
- Exploration of Fine-Tuning and Data Augmentation techniques, including Easy Data Augmentation and Back Translation.
- Evaluation of the ALBERT model with performance evaluation metrics, confusion matrix, and results from test data analysis.

This paper is organized as follows. Section 2 provides an extensive review of relevant literature, highlighting the challenges posed by the nature of social media posts and introducing transformer-based language models such as ALBERT. In Section 3, we present the methodology, including the flowchart of the research process, dataset details, and pre-processing steps, delve into the detailed architecture of ALBERT, fine-tuning strategies, and the implementation of Easy Data Augmentation and Back Translation techniques. Performance evaluation metrics and results are discussed in Section 4, followed by the presentation of confusion matrices and sentiment prediction results in Section 4. Finally, Section 5 concludes the paper by summarizing key findings, contributions, and implications for sentiment analysis, particularly in emerging technologies like ChatGPT.

## 2. Literature Review

Sentiment analysis on social media has become crucial for evaluating public sentiment, especially regarding emerging technologies like ChatGPT [24]. However, unstructured and lengthy posts on platforms like Twitter pose accuracy challenges [25]. Lengthy texts with extended phrases make sentiment categorization more manageable, leading to decreased performance. Sentiment analysis, a vital technique in natural language processing (NLP), assesses subjective opinions as positive, neutral, or negative based on linguistic context [26], [27]. Transformer models such as ALBERT, BERT, DistilBERT, XLNet, and ELECTRA have propelled improvements in sentiment analysis through contextual language modeling and efficiency [28]. ALBERT, known for its efficient parameter reduction techniques, significantly advances sentiment analysis [17].

The paper "ALBERT: A Lite BERT for Self-Supervised Learning of Language Representations" introduces a novel approach to address the memory limitations and training speed of BERT models by significantly reducing the number of parameters while maintaining or even improving performance on various natural language processing tasks [17]. ALBERT achieves state-of-the-art results on benchmarks like GLUE, RACE, and SQuAD by focusing on self-supervised learning and emphasizing intersentence coherence. The innovative techniques proposed in ALBERT enhance training efficiency and demonstrate the potential for creating more powerful language representations through advanced self-supervised learning methods [17]. The research contributes to the ongoing exploration of optimizing large-scale language models for practical applications in NLP tasks.

Alamoudi and Alghamdi present a novel approach to deep learning and word embeddings are used to classify sentiment and analyze aspects of Yelp reviews [29]. The paper addresses a research gap by evaluating the efficacy of machine learning, deep learning, and transfer learning models for sentiment classification tasks on Yelp datasets, which have yet to be substantially researched [29]. Additionally, the paper introduces a new unsupervised technique for aspect extraction based on semantic similarity and pre-trained language models, offering a more straightforward and cost-effective alternative to traditional supervised learning methods. Based on the research, ALBERT outperforms other tested models while maintaining a higher accuracy than its base BERT model. The research contributes to the field by enhancing the understanding of sentiment analysis in online reviews and providing practical implications for businesses and decision-makers seeking to extract valuable insights from customer feedback on platforms like Yelp.

Fine-tuning and data augmentation techniques are also evaluated to increase the precision of transformer-based language models such as ALBERT, BERT, DistilBERT, XLNet, and ELECTRA [28], [30].

Fine-tuning is a strategic approach in deep learning where the weights of a neural network are initialized with pre-trained models, and the model is subsequently refined or "fine-tuned" on a specific task or dataset of interest [31]. This method capitalizes on the knowledge gained by the pre-trained model on a large and diverse dataset, enabling the network to learn general features and representations [31]. By starting with these pre-existing weights, the model benefits from the wealth of information encoded in the initial parameters [32].

Data augmentation is a powerful data-level technique often used to improve the effectiveness of machine learning models. It has found use in the realm of natural language processing (NLP), extending its benefits to text augmentation [33]. In the context of NLP tasks, data augmentation involves the creation of diverse variations of existing textual data, thereby expanding the training dataset and enhancing the model's ability to generalize to different linguistic patterns [23]. This approach proves especially valuable when working with limited labeled data, as it mitigates the risk of overfitting and helps build more robust and adaptable NLP models [34].

## 3. METHODS

An overview of how the research will be conducted in a flowchart. Figure 1 shows the steps for this research involving collecting the dataset, preprocessing the data that has been collected, initializing and fine-tuning the model, following the data is augmented, and then evaluating the model by analyzing the results.

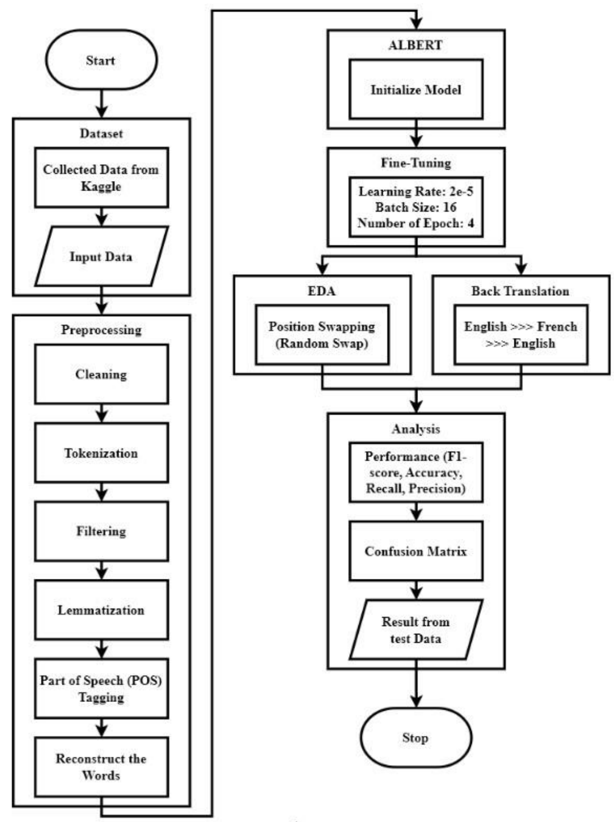


Figure 1. Flowchart Sentiment Analysis

### A. Dataset

The dataset used in this study originates from Kaggle.com. This dataset is a collection of data from the Twitter platform, containing opinions submitted by individuals who have used ChatGPT. Each dataset entity has been assigned a classification label, namely Positive, Neutral, and Negative.

Prior research conducted by Kaggle user Sujal Neupane involved testing the DeBERTa model, resulting in an evaluation accuracy of 28.7%, a performance level considered suboptimal for a transformer model.

Based on Table I, the long-format dataset consists of several data entities labeled as 10539 Positive, 10539 Neutral, and 10539 Negative. Further analysis of this dataset is expected to provide in-depth insights into opinions and sentiments surrounding the use of ChatGPT.

TABLE I. DATA DISTRIBUTION

| Polarity | Number of Data |
|---|---|
| Positive | 10539 |
| Neutral | 10539 |
| Negative | 10539 |

The data analysis divides the dataset into three categories: long, medium, and short, using quartiles. The specifications are as follows: the long category has a word range from 25 to 64 words, the medium category has a word range from 15 to 24 words, and the short category has a word range under 15 words. In this context, only the dataset with the long category is selected because it is considered the most relevant to achieving the objectives of this study.

TABLE II. DATA SPLIT

| Segment | Number of Data |
|---|---|
| Training | 18492 |
| Validation | 6563 |
| Test | 6562 |

As shown in Table II, the dataset in this study is split into three segments: Training Data, Validation Data, and Testing data. The Training Data contains 60% of the original dataset, while the remaining data for Validation and Testing data are split in half, so both Validation and Testing data are 20%.

### B. Pre-Processing

Preprocessing is necessary in sentiment analysis to prepare and transform raw text data before feeding it into machine learning models [35]. The effective preprocessing analysis makes patterns in the text more apparent for the model to learn from and improves accuracy on unseen real-world data [36], [37]. The primary procedures of the pre-processing stage are outlined as follows:

- Cleaning: Takes in a text and runs several regex substitutions to remove Twitter handles, hashtags, retweet marks, links, and non-word characters.
- Tokenization: Breaks text into individual tokens (words) using nltk word tokenizer.
- Filtering: Filters out stop words and keeps only significant tokens.
- Lemmatization: Lemmatizes tokens to their root form to standardize terms and Converts words like "writing" to "write" or “Studying” to “Study.”
- Part Of Speech (POS) Tagging: Tags each token with its part-of-speech like noun, verb, or adjective, and keeps only significant POS tags for sentiment analysis.
- Reconstruct the words: Joins the preprocessed list of tokens into a cleaned sentence string and prepares final output for feeding into sentiment analysis models.

Following the execution of various pre-processing steps detailed earlier.

TABLE III. CLEAN DATA COMPARISON

| Uncleaned Data | Cleaned Data |
|---|---|
| #ChatGPT will replace @Google as search engine as it's only limited to crawling and not understanding exact search. Not atleast as effectively as this model. Internet is about to be so much cognitively advance soon @sama | replace search engine exact search at least effectively internet much cognitively advance soon |
| keep seeing tiktoks about how chatgpt is gonna ruin all education and so im putting in some of my old homework problems to see how it does and so far its doing stuff but just completely WRONG ðŸ˜ like thermo stuff is gonna be done by hand for a few more years at least | keep ruin education old homework problem far stuff completely wrong thermo stuff hand year least |
| It's amazing how people on Twitter are excited about #ChatGPT. If you think media manipulation is bad just wait for the AI manipulation IF it becomes mainstream. Don't know why @elonmusk is happy about it... After all that Twitter Files show... | twitter medium manipulation bad wait ai manipulation mainstream know happy twitter file show |

Table III shows that the preprocessing task shows excellent results from the uncleaned data. The preprocessing task makes the uncleaned data more assertive and more understandable for the model to learn.

### C. ALBERT

The ALBERT framework is a comprehensive development of the BERT (Bidirectional Encoder Representations from Transformers) architecture aimed at increasing efficiency and scalability in language representation [12]. A key innovation lies in introducing parameter sharing across layers, reducing redundancy, and rendering the model more parameter-efficient. This is

International Journal of Computing and Digital Systems
ISSN (2210-142X)
Int. J. Com. Dig. Sys. #, No.# (Mon-20..)

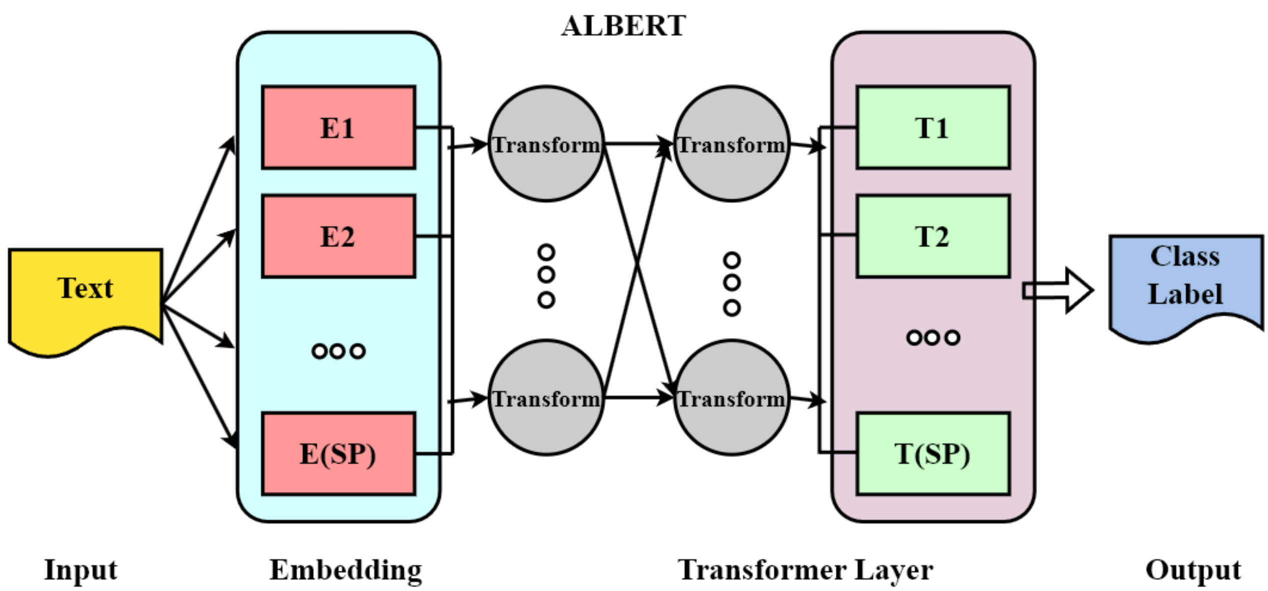


Figure 2. ALBERT Architecture

achieved by sharing weights and biases within the feedforward neural network (FFN) across layers [12]. The embedding layer is factorized into token embeddings and token-type embeddings, reducing the overall number of parameters.

Cross-layer parameter sharing is also introduced, further optimizing parameter utilization. The training objective involves Sentence Order Prediction (SOP), wherein the model predicts the correct order of sentences in a document, facilitating improved contextual understanding.

Figure 2 shows the overall architecture of ALBERT. ALBERT uses multiple transformer layers interspersed with a more comprehensive particular transformer layer to learn contextual relationships between words in the input text. The output embeddings are fed to a task-specific classifier. The goal is to achieve BERT-level performance with significantly fewer parameters.

$$h_i = LN(h_{i-1} + FNN(LN(h_{i-1} \cdot W_i + b_i))) \quad (1)$$

Equation (1) is the equation that represents the critical parameter-sharing idea in ALBERT where:

1. $h_i$ is the output layer of $i$
2. $LN$ is layer normalization
3. $FNN$ is a feedforward neural network
4. $W_i$ and $b_i$ are the shared weights and biases across layers

The parameter sharing is achieved through $W_i$ and $b_i$ which are the same for all layers.

The key idea behind ALBERT is parameter sharing across layers. In BERT, each layer has its own set of parameters, leading to many parameters in the model. ALBERT minimizes the number of parameters by sharing them between layers. This enables for a more parameter-efficient model while maintaining performance.

*D. Fine-Tuning*

Fine-tuning is a transfer method of learning that involves retraining a previously learned neural network on a fresh dataset or task [31]. The model in this research was tuned using the Adam optimizer, and the fine-tuned parameters are:

**Learning Rate: 2e-5**
**Batch Size: 16**
**Number of Epoch: 4**

We utilize fine-tuning to adapt a pre-trained ALBERT language model to our specific text classification task. ALBERT is a transformer-based model trained on a large corpus of unlabelled text data using a novel permutation language modelling objective [28].

*E-mail:author's email*

**International Journal of Computing and Digital Systems**
ISSN (2210-142X)
Int. J. Com. Dig. Sys. #, No.# (Mon-20..)

TABLE IV. COMPARISON OF MODEL PERFORMANCE METRICS

| Model | Accuracy | F1-Score | Precision | Recall |
|---|---|---|---|---|
| LSTM + Random Forest | 55% | 53% | 55% | 54% |
| Naive Bayes | 72% | 65% | 71% | 66% |
| Electra + EDA | 78.9% | 79.2% | 79.8% | 78.9% |
| XLNet + EDA | 78.7% | 78.4% | 78.7% | 78.4% |
| BERT + EDA | 79.3% | 79.5% | 80.1% | 79.3% |
| DistilBERT + EDA | 77.1% | 77.5% | 79.3% | 77.1% |
| ALBERT + EDA (Proposed Method) | **81%** | **81.2%** | **81.5%** | **81%** |
| ALBERT + Back Translation (Proposed Method) | 80.1% | 80.8% | 81.4% | 80.1% |

### *E. Easy Data Augmentation (EDA)*

EDA, which stands for Easy Data Augmentation, is a simple yet effective approach for improving the robustness and preventing overfitting text classification models. It involves applying four straightforward data transformation operations to the training data:

- Synonym Substitution: Instead of randomly selecting non-halt terms and replacing them with arbitrary synonyms, randomly choose n non-stop words in each sentence and replace them with random synonyms.
- Random Insertion: Select an arbitrary synonym for an arbitrary non-stop word in the sentence. Insert that synonym at an arbitrary placement within the sentence. Redo n times.
- Position Swapping: Randomly select two words from the sentence and switch their locations. Repeat n times.
- Word Dropping: Randomly eliminate each word in the sentence with probability p.

In this research, we will use Position Swapping (Random Swap) as it results in better performance and accuracy than other methods. Using Random Swap from Easy Data Augmentation on the training dataset, the original training length is 18,492 and the augmented length is 36,984.

### *F. Back Translation*

Back translation augmentation is a powerful technique employed in natural language processing, text data is translated into another language and then translated back into the original language [38]. After using Back Translation, the augmented length of the data becomes 20583, initially 18492. This means the Back Translation added 2092 new text data that has been translated to French and back to English.

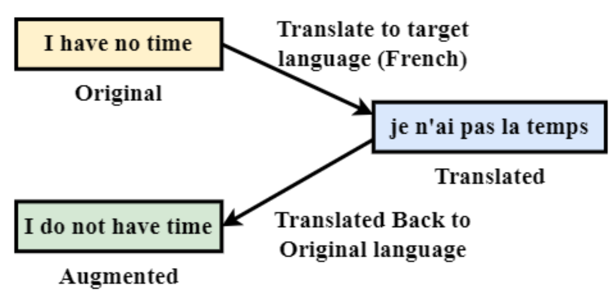


Figure 3. Back Translation Process

Figure 3 shows how Back Translation proceeds by translating the original data to the target language and translating it back to the original language, producing different sentences with the same meaning.

## 4. RESULT AND ANALYSIS

### *A. Performance Evaluation*

The model performance analysis for sentiment analysis of long-form texts reveals notable trends and insights. Among the models evaluated, ALBERT demonstrates strong performance, particularly when augmented with Easy Data Augmentation (EDA) or Back Translation techniques, achieving 81% and 80.1% accuracy, respectively. Table IV compares model performance metrics for sentiment analysis of long-form texts.

*E-mail:author's email*

International Journal of Computing and Digital Systems
ISSN (2210-142X)
Int. J. Com. Dig. Sys. #, No.# (Mon-20..)

The performance of Naive Bayes and LSTM + Random Forest from previous research demonstrates noteworthy characteristics. Naive Bayes, characterized by its simplicity and efficiency, achieved an accuracy of 72%, indicating a reasonable level of predictive capability. However, its F1-Score, Precision, and Recall metrics are comparatively lower at 65%, 71%, and 66% respectively. Notably, BERT and ELECTRA also exhibit competitive performance, with accuracies ranging from 79.3% to 79.5%. Furthermore, integrating data augmentation techniques consistently enhances model performance across various metrics, including F1-Score, Precision, and Recall.

However, models like XLNet and DistilBERT show slightly lower accuracy and F1-Score, with 78.7% and 77.1%, respectively, indicating the potential for improvement. Fine-tuning strategies or alternative data augmentation methods may enhance the performance of these models. These findings underscore the practical utility of transformer-based models in sentiment analysis of long-form text.

### B. *Training Accuracies*

Training accuracies measure how well a sentiment analysis model fits the data it was trained on. While training accuracy is essential for tracking model progress throughout development, validation accuracy should drive model selection to avoid overfitting. Figures 4 and 5 show the training accuracy from the ALBERT model using EDA and Back Translation.

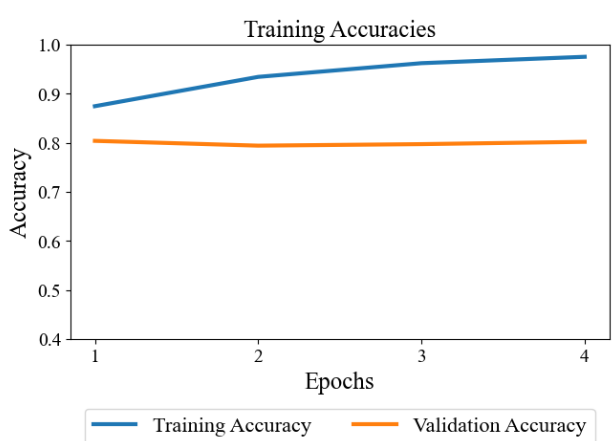


Figure 4. The Training Accuracy of ALBERT with EDA.

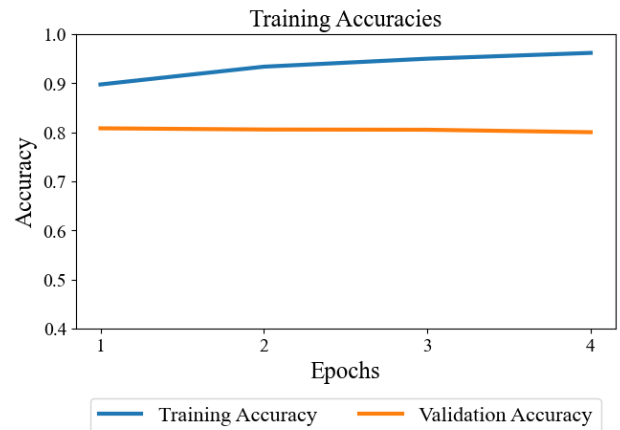


Figure 5. The Training Accuracy of ALBERT with Back Translation.

### C. *Confusion Matrix*

Based on the accuracy results of the model, Figure 6 and Figure 7 show the confusion matrix of the ALBERT model using EDA and Back Translation. Mapping words into negative, neutral, and positive, the confusion matrix displays a list of words that the model accurately and inaccurately classifies into each sentiment category.

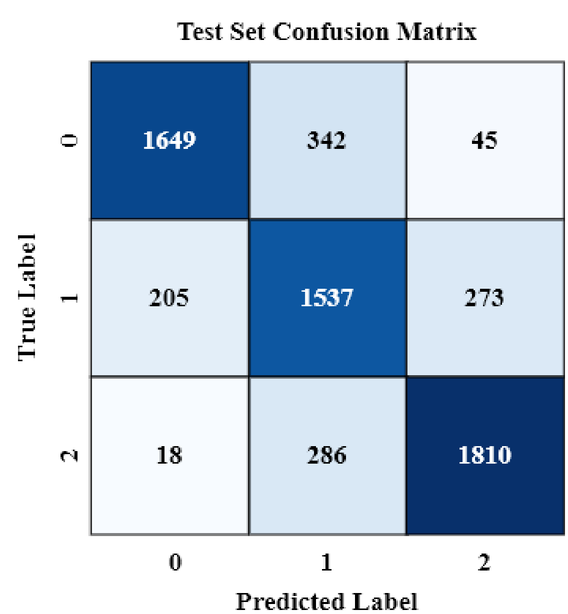


Figure 6. The Confusion Matrix of ALBERT with EDA.

*E-mail:author's email*

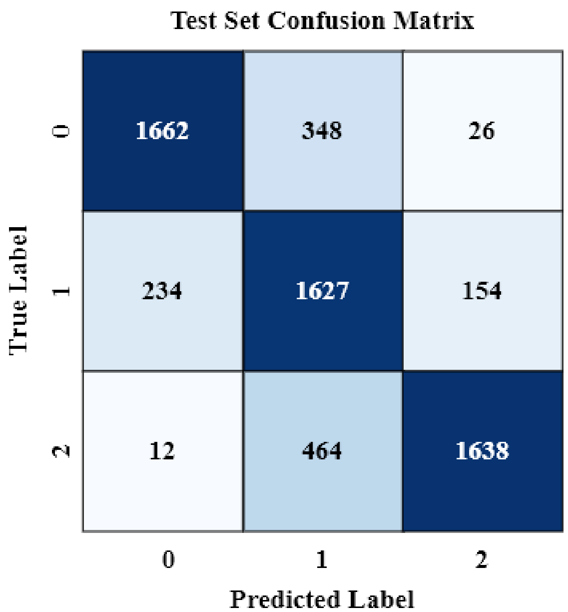


Figure 7. The Confusion Matrix of ALBERT with Back Translation.

### *D. Result from Test Data*

After conducting the analysis, Table V and Table VI display the sentiment prediction results on the ALBERT with EDA and Back Translation test data. Testing with this new data is performed to validate further how accurately the model predicts sentiments in data it has not seen before.

TABLE V. SENTIMENT PREDICTION RESULT ON ALBERT WITH EDA TEST DATA

| **Polarity** | **Text** |
|---|---|
| Positive | I have been fascinated my #OpenAi release of the conversational GPT-3 model, #ChatGPT. This incredible language model, finely tuned for conversations, creates human-like responses and is capable of remembering the conversation and building upon the stored knowledge. |
| Neutral | My daughter just asked me how to convert .tpl file to .abr file. Not knowing what they were, I was able to give her step-in-step instructions under 10 seconds. Between me and #ChatGPT, I am better, with ... https://t.co/daprlCr1XI |
| Negative | OpenAi GPT3 and beyond, this is the single most threat to computer programming jobs. \n\nTrust me, itâ€™s a long road ahead for devs. \n\nThe ChatGPT should be a source of worry for Alphabet. |

TABLE VI. SENTIMENT PREDICTION RESULT ON ALBERT WITH BACK TRANSLATION TEST DATA

| **Polarity** | **Text** |
|---|---|
| Positive | ChatGPT is neat, but using open source code in closed source coding is not; know your attribution:\n"Output generated by code generation features of our Services, including OpenAI Codex, may be subject to third party licenses, including, without limitation, open source licenses." |
| Neutral | This article was written 90% by @OpenAI's ChatGPT. I was working on a SwiftUI app today and prompted it to help me change the background color. Initially the answer was incorrect, but after suggesting that it use a ZStack, the answer was ðŸ’¯.\n\nðŸ¤– ðŸ“šâœ• ï¸• \n\nhttps://t.co/JNFXtLiOGu |
| Negative | ChatGPT is really bad at math. It can't even pick the larger of two numbers.\n\nBut it's not just for humanitiesâ€”it's actually really good at writing code.\n\nSo, running with @sjwhitmore's latest thread on building your own ChatGPT-like tool, I made ChatGPT's math-nerd alter ego. |

## 5. CONCLUSION

This research focuses on using the ALBERT model to analyze sentiments in long Twitter reviews about ChatGPT. It highlights the need for advanced natural language processing (NLP) methods to improve sentiment analysis accuracy due to the complexity of social media content. Results show that ALBERT, combined with data augmentation techniques, achieves high accuracy rates of 81% and 80.1%. This study demonstrates the effectiveness of ALBERT in handling lengthy social media texts and emphasizes the importance of data augmentation for better model performance. It suggests that transformer-based models like ALBERT are valuable for understanding public opinions on emerging technologies like ChatGPT.

## ACKNOWLEDGMENT

The authors would like to express their gratitude towards the ChatGPT Sentiment Analysis dataset collectors and Sujal Neupane's contributions, reviewers, and peers for insightful feedback and research participants who unintentionally contributed through Kaggle comments to support the execution of this research.

## REFERENCES

[1] Y. Su and Z. J. Kabala, “Public Perception of ChatGPT and Transfer Learning for Tweets Sentiment Analysis Using Wolfram Mathematica,” *Data*, vol. 8, no. 12, Art. no. 12, Dec. 2023, doi: 10.3390/data8120180.

[2] A. A. Hidayat and B. Pardamean, "Count time series modelling of Twitter data and topic modelling: A case of Indonesia flood events," *Procedia Comput. Sci.*, vol. 227, pp. 805–812, Aug. 2023, doi: 10.1016/j.procs.2023.10.586.

[3] F. Kateb and J. Kalita, "Classifying Short Text in Social Media: Twitter as Case Study," *Int. J. Comput. Appl.*, vol. 111, pp. 1–12, Feb. 2015, doi: 10.5120/19563-1321.

[4] J. Davis, "Social Media," 2016. doi: 10.1002/9781118541555.wbiepc004.

[5] K. Purwandari, J. W. C. Sigalingging, T. W. Cenggoro, and B. Pardamean, "Multi-class Weather Forecasting from Twitter Using Machine Learning Aprroaches," *Procedia Comput. Sci.*, vol. 179, pp. 47–54, Jan. 2021, doi: 10.1016/j.procs.2020.12.006.

[6] K. Purwandari, R. B. Perdana, J. W. C. Sigalingging, R. Rahutomo, and B. Pardamean, "Automatic Smart Crawling on Twitter for Weather Information in Indonesia," *Procedia Comput. Sci.*, vol. 227, pp. 795–804, Aug. 2023, doi: 10.1016/j.procs.2023.10.585.

[7] I. Nurlaila, R. Rahutomo, K. Purwandari, and B. Pardamean, "Provoking Tweets by Indonesia Media Twitter in the Initial Month of Coronavirus Disease Hit," in *2020 International Conference on Information Management and Technology (ICIMTech)*, Bandung, Indonesia: IEEE, Aug. 2020, pp. 409–414. doi: 10.1109/ICIMTech50083.2020.9211179.

[8] Y. K. Dwivedi *et al.*, "Metaverse beyond the hype: Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy," *Int. J. Inf. Manag.*, vol. 66, p. 102542, Oct. 2022, doi: 10.1016/j.ijinfomgt.2022.102542.

[9] K. Purwandari, A. S. Perbangsa, J. W. C. Sigalingging, A. A. Krisna, S. Anggrayani, and B. Pardamean, "Database Management System Design for Automatic Weather Information with Twitter Data Collection," in *2021 International Conference on Information Management and Technology (ICIMTech)*, Aug. 2021, pp. 326–330. doi: 10.1109/ICIMTech53080.2021.9535009.

[10] P. P. Ray, "ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope," *Internet Things Cyber-Phys. Syst.*, vol. 3, pp. 121–154, Jan. 2023, doi: 10.1016/j.iotcps.2023.04.003.

[11] B. Lund and T. Wang, "Chatting about ChatGPT: How may AI and GPT impact academia and libraries?," *Libr. Hi Tech News*, vol. 40, Feb. 2023, doi: 10.1108/LHTN-01-2023-0009.

[12] D. Cortiz, *Exploring Transformers models for Emotion Recognition: a comparision of BERT, DistilBERT, RoBERTa, XLNET and ELECTRA*. 2022, p. 234. doi: 10.1145/3562007.3562051.

[13] S. Chaurasia and S. Sherekar, "Sentiment Analysis of Twitter Data by Natural Language Processing and Machine Learning," 2023, pp. 59–70. doi: 10.1007/978-981-99-2768-5_6.

[14] K. Svensson and J. Hagelbäck, "Sentiment Analysis With Convolutional Neural Networks".

[15] S. Pipin, F. Sinaga, S. Winardi, and M. Hakim, "Sentiment Analysis Classification of ChatGPT on Twitter Big Data in Indonesia Using Fast R-CNN," *J. MEDIA Inform. BUDIDARMA*, vol. 4, pp. 2137–2148, Oct. 2023, doi: 10.30865/mib.v7i4.6816.

[16] B. G. Bokolo and Q. Liu, "Deep Learning-Based Depression Detection from Social Media: Comparative Evaluation of ML and Transformer Techniques," *Electronics*, vol. 12, no. 21, p. 4396, Oct. 2023, doi: 10.3390/electronics12214396.

[17] Z. Lan, M. Chen, S. Goodman, K. Gimpel, P. Sharma, and R. Soricut, "ALBERT: A Lite BERT for Self-supervised Learning of Language Representations." arXiv, Feb. 08, 2020. doi: 10.48550/arXiv.1909.11942.

[18] S. R *et al.*, "Analyzing Sentiments Regarding ChatGPT Using Novel BERT: A Machine Learning Approach," *Information*, vol. 14, no. 9, Art. no. 9, Sep. 2023, doi: 10.3390/info14090474.

[19] T. Han, Z. Zhang, M. Ren, C. Dong, X. Jiang, and Q. Zhuang, "Text Emotion Recognition Based on XLNet-BiGRU-Att," *Electronics*, vol. 12, no. 12, Art. no. 12, Jan. 2023, doi: 10.3390/electronics12122704.

[20] V. Sanh, L. Debut, J. Chaumond, and T. Wolf, "DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter." arXiv, Feb. 29, 2020. doi: 10.48550/arXiv.1910.01108.

[21] K. Clark, M.-T. Luong, and Q. V. Le, "ELECTRA: PRE-TRAINING TEXT ENCODERS AS DISCRIMINATORS RATHER THAN GENERATORS," 2020.

[22] S. F. N. Azizah and W. Widiarto, "Performance Analysis of Transformer Based Models (BERT, ALBERT and RoBERTa) in Fake News Detection".

[23] C. Shorten, T. M. Khoshgoftaar, and B. Furht, "Text Data Augmentation for Deep Learning," *J. Big Data*, vol. 8, no. 1, p. 101, Jul. 2021, doi: 10.1186/s40537-021-00492-0.

[24] K. R. Manda, "Sentiment Analysis of Twitter Data Using Machine Learning and Deep Learning Methods".

[25] 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, Aug. 2023, doi: 10.1016/j.procs.2023.10.514.

[26] K. Purwandari, R. Rahutomo, J. W. C. Sigalingging, M. Ajy Kusuma, A. Prasetyo, and B. Pardamean, "Twitter-Based Text Classification Using SVM for Weather Information System," in *2021 International Conference on Information Management and Technology (ICIMTech)*, Aug. 2021, pp. 27–32. doi: 10.1109/ICIMTech53080.2021.9534945.

[27] K. Purwandari, T. W. Cenggoro, J. W. C. Sigalingging, and B. Pardamean, "Twitter-based classification for integrated source data of weather observations," *IAES Int. J. Artif. Intell. IJ-AI*, vol. 12, no. 1, Art. no. 1, Mar. 2023, doi: 10.11591/ijai.v12.i1.pp271-283.

[28] S. Casola, I. Lauriola, and A. Lavelli, "Pre-trained transformers: an empirical comparison," *Mach. Learn. Appl.*, vol. 9, p. 100334, Sep. 2022, doi: 10.1016/j.mlwa.2022.100334.

[29] E. Alamoudi and N. Alghamdi, "Sentiment classification and aspect-based sentiment analysis on yelp reviews using deep learning and word embeddings," *J. Decis. Syst.*, vol. 30, pp. 1–23, Jan. 2021, doi: 10.1080/12460125.2020.1864106.

[30] E. Zhang, A. Cheok, Z. Pan, J. Cai, and Y. Yan, "From Turing to Transformers: A Comprehensive Review and Tutorial on the Evolution and Applications of Generative Transformer Models," *Sci*, vol. 5, p. 46, Dec. 2023, doi: 10.3390/sci5040046.

[31] P. D. Alfano, V. P. Pastore, L. Rosasco, and F. Odone, "Top-tuning: A study on transfer learning for an efficient alternative to fine tuning for image classification with fast kernel methods," *Image Vis. Comput.*, vol. 142, p. 104894, Feb. 2024, doi: 10.1016/j.imavis.2023.104894.

[32] M. Umer *et al.*, "ETCNN: Extra Tree and Convolutional Neural Network-based Ensemble Model for COVID-19 Tweets Sentiment Classification," *Pattern Recognit. Lett.*, vol. 164, pp. 224–231, Dec. 2022, doi: 10.1016/j.patrec.2022.11.012.

[33] B. Li, Y. Hou, and W. Che, "Data augmentation approaches in natural language processing: A survey," *AI Open*, vol. 3, pp. 71–90, Jan. 2022, doi: 10.1016/j.aiopen.2022.03.001.

[34] O. Litake, N. Yagnik, and S. Labhsetwar, "IndiText Boost: Text Augmentation for Low Resource India Languages." arXiv, Jan. 23, 2024. Accessed: Feb. 08, 2024. [Online]. Available: http://arxiv.org/abs/2401.13085

[35] D. Sudigyo, A. A. Hidayat, R. Nirwantono, R. Rahutomo, J. P. Trinugroho, and B. Pardamean, "Literature study of stunting supplementation in Indonesian utilizing text mining approach," *Procedia Comput. Sci.*, vol. 216, pp. 722–729, Jan. 2023, doi: 10.1016/j.procs.2022.12.189.

[36] T. N. Ph.D and A. Amalanathan, "DATA PREPROCESSING IN SENTIMENT ANALYSIS USING TWITTER DATA," vol. 3, pp. 89–92, Jul. 2019.

[37] R. Rahutomo, F. Lubis, H. H. Muljo, and B. Pardamean, "Preprocessing Methods and Tools in Modelling Japanese for Text Classification," in *2019 International Conference on Information Management and Technology (ICIMTech)*, Aug. 2019, pp. 472–476. doi: 10.1109/ICIMTech.2019.8843796.

[38] M. Fadaee and C. Monz, "Back-Translation Sampling by Targeting Difficult Words in Neural Machine Translation," in *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, E. Riloff, D. Chiang, J. Hockenmaier, and J. Tsujii, Eds., Brussels, Belgium: Association for Computational Linguistics, Oct. 2018, pp. 436–446. doi: 10.18653/v1/D18-1040.

**Wanda Safira**, A first-year student pursuing a Master's in Computer Science at Bina Nusantara University, is incredibly passionate about advancing natural language processing. She aims to focus her studies on developing innovative machine learning and deep learning models for sentiment analysis and text classification.

**Benedictus Prabaswara**, A first-year student pursuing a Master's degree in Computer Science at Bina Nusantara University, is incredibly passionate about advancing natural language processing. He aims to focus his studies on developing innovative machine learning and deep learning models for sentiment analysis and text classification.

**Andrea Stevens Karnyoto** earned his Bachelor's degree from STMIK Dipanegara, a Master's degree from Hasanuddin University, and a Ph.D. from Harbin Institute of Technology. His Ph.D. studies focused on natural language processing, including Large Language Modeling (LLM), the latest transformers architecture, topic-based methods, and graph neural networks. Aside from working in Artificial Intelligence (AI), he has a web and desktop programming background. During his career in computer science, he created many information systems for various organizations. In addition, he has received two grants for research and two for community service from the Ministry of Education, Culture, Research, and Technology.

**Bens Pardamean** has over thirty years of global experience in information technology, bioinformatics, and education, including a strong background in database systems, computer networks, and quantitative research. His professional experience includes being a practitioner, researcher, consultant, entrepreneur, and lecturer. His current research interests are developing and analyzing genetic data in cancer and genome-wide association studies (GWAS) for agriculture genetic research. After successfully leading the Bioinformatics Research Interest Group, he currently holds a dual appointment as the Director of the Bioinformatics & Data Science Research Center (BDSRC) and as a Professor of Computer Science at the University of Bina Nusantara (BINUS) in Jakarta, Indonesia. He earned a doctoral degree in informative research from the University of Southern California (USC), a master's degree in computer education, and a bachelor's degree in computer science from California State University, Los Angeles.