Public sentiment analysis on facebook towards information security using the albert model with data augmentation

Fadlillah Mukti Ayudewi*, Arridho Ramadhan Firdaus

Information Technology Study Program, Faculty of Science and Technology, Universitas ‘Aisyiyah Yogyakarta, Indonesia
Email: fadlillahmuktiayudewi@unisayogya.ac.id

Abstract

Information security has become a strategic issue in the digital era, particularly with the increasing risks of data breaches and privacy threats on social media platforms. This study aims to analyze public sentiment on Facebook comments related to information security by leveraging ALBERT (A Lite BERT) combined with Back Translation data augmentation to overcome class imbalance. The dataset consists of 1,000 comments categorized into negative, neutral, and positive sentiments, processed through text pre-processing, augmentation, and classification stages. The experimental setup employed a stratified data split of 80% training, 10% validation, and 10% testing to ensure balanced class distribution. Evaluation using accuracy, precision, recall, and F1-score demonstrates that the proposed model achieves 81% accuracy with a macro F1-score of 0.80, where the neutral class shows the most stable performance. The findings indicate that ALBERT with data augmentation effectively improves sentiment classification performance. This research contributes to the development of adaptive sentiment analysis systems to support policymakers in formulating strategies related to digital information security.

Keywords: Sentiment Analysis, ALBERT, Data Augmentation, Back Translation, Information Security, NLP

1. Introduction

Information security is one of the strategic issues that has gained increasing attention in the digital era. The rapid growth of internet and social media usage has escalated the risks of data breaches, misuse of personal information, and threats to digital privacy (Aldean et al., 2021). This situation has positioned information security as a widely discussed topic in the public sphere, whether in the form of criticism, complaints, or support for certain policies.

Facebook, as the social media platform with the largest number of users in Indonesia (We Are Social, & Kepios (2023), has become a primary space for public discussions on issues related to information security. The comments expressed on this platform reflect a wide range of sentiments, from support to opposition toward policies related to digital privacy. However, there is still no structured sentiment map available to capture and analyze public responses to these issues. This highlights the need for an analytical system capable of processing large volumes of comments quickly, accurately, and efficiently.

From an Islamic perspective, the importance of verifying information is also emphasized in the Qur’an, as stated in Surah Al-Hujurat (49:6) Badan Litbang dan Diklat Kementerian Agama RI. (2019) : “O you who have believed, if there comes to you a disobedient one with information, investigate, lest you harm a people out of ignorance and become, over what you have done, regretful.” This verse underscores the obligation to verify the truth of information before disseminating it, which aligns with the fundamental principles of information security in the digital era.

Furthermore, the Qur’an also emphasizes the importance of protecting privacy in Surah An-Nur (24:27) Badan Litbang dan Diklat Kementerian Agama RI. (2019), which states that one should not enter another person’s house without prior permission. This value can be drawn as an analogy in the digital context, where respect for data privacy is an essential aspect of information security.

Sentiment analysis based on Natural Language Processing (NLP) is one of the widely used approaches to automatically understand public perceptions (Putranta et al., 2023). Transformer-based models such as BERT and its variants have proven effective in capturing deep contextual meaning in language (Syahriani et al., 2010). One of its developments is ALBERT (A Lite BERT), which provides a lighter architecture while maintaining high accuracy (Lan et at 2020), (Kharisudin & Putra, 2022). Nevertheless, the limited availability of data and the imbalance of sentiment class distribution remain significant challenges in applying NLP models to real-world datasets. To address these issues, data augmentation techniques such as synonym replacement and back-translation can be employed to enrich training data variations and improve model performance (Julianto & Wibowo, 2024) (Witanti & Dian, 2022).

Based on this background, this study focuses on developing a public sentiment analysis system related to information security issues on Facebook by utilizing the ALBERT model and data augmentation approaches. The main objectives of this research are (1) to build an NLP-based sentiment analysis model to classify public comments into positive, neutral, and negative categories, (2) to address data limitations through the application of text augmentation techniques, and (3) to generate a structured mapping of public opinion in the form of informative visualizations. This study is expected to provide a meaningful contribution to the development of adaptive and applicable methods for monitoring public opinion while supporting policymakers in formulating strategies for information security in the digital domain.

2. Metode

This study used a dataset containing 1000 comments from the Facebook platform to test the performance of the ALBERT model in sentiment classification. This dataset consists of user comments which are then categorized into three classes: negative, neutral, and positive. The flow of the research stages is shown in Figure 1.

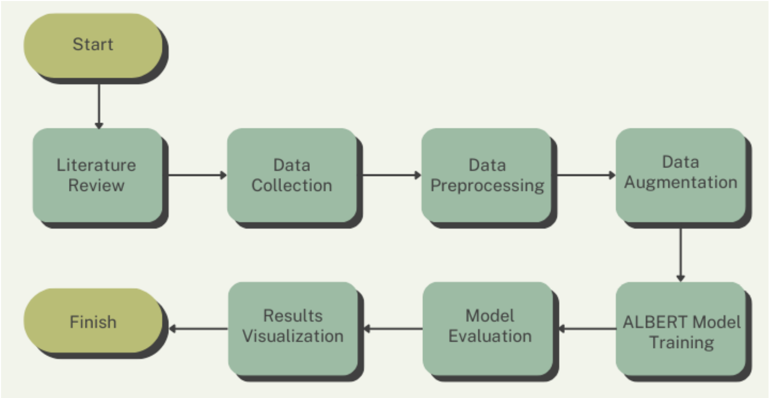


Figure 1. Research flow of sentiment analysis using ALBERT with data augmentation.

Figure 1 illustrates the research flow, which begins with dataset collection, followed by preprocessing for text handling, and then the augmentation stage aimed at enriching data variation. Next, feature extraction is carried out using the ALBERT model Lan et al (2020), followed by the sentiment classification stage. The classification results are then evaluated using metrics to assess the model's performance (Maulidiana et al., 2024).

3. Literature Review

Research on sentiment analysis based on social media has been widely conducted, particularly in the context of understanding public opinion on social, political, and digital security issues. Aldean et al. (2021), examined public sentiment regarding the issue of COVID-19 vaccination on Twitter using the Random Forest Classifier method and found that the majority of public opinion tended to be influenced by negative sentiments related to vaccine effectiveness. This indicates that social media has become a primary space for the formation of public opinion concerning strategic issues.

From the methodological perspective, Natural Language Processing (NLP) serves as the main foundation in sentiment analysis research. Kharisudin et al. (2022), compared Support Vector Machine (SVM), Naïve Bayes, and Logistic Regression for sentiment analysis of Tokopedia application users, with SVM achieving the best performance. Another study by Julianto et al. (2024) analyzed public opinion related to the free internet program using Naïve Bayes, while Witanti & Dian (2022) demonstrated the effectiveness of SVM in classifying sentiments toward COVID-19 vaccination.

Aldean et al., (2021), conducted a comparative analysis between SVM and Naïve Bayes on the case of public opinion regarding the Supreme Court's ruling, with results showing that SVM had higher accuracy. Similarly, Widyadhana et al. (2023) also applied SVM to measure public sentiment toward public services at Polres Ponorogo, and the findings indicated that the model was capable of capturing opinion tendencies effectively.

In terms of methodology, Liu (2012), Medhat et al. (2014) emphasized the importance of both machine learning-based and lexicon-based approaches in sentiment analysis. Syahriani et al. (2010) highlighted the application of Naïve Bayes in the Indonesian language context, specifically on Facebook comments regarding presidential candidates. Hadikristanto et al. (2023) further expanded research into the issue of data breaches, employing Naïve Bayes and SVM to identify public opinion on Twitter.

However, one of the main challenges in sentiment analysis is data limitations and class imbalance. Several studies have highlighted the need for data augmentation techniques to improve model performance, although their application remains limited to specific topics such as healthcare (Witanti & Dian, 2022). On the other hand, Lan et al. (2020) developed ALBERT, a transformer-based model that is lighter than BERT but still maintains high performance, making it suitable for applications with limited computational resources.

Beyond technical aspects, Islamic values also provide a normative foundation in information management. The Qur'an emphasizes the importance of verifying news in QS. Al-Hujurat (49:6) (Badan Litbang dan Diklat Kementerian Agama RI. (2019), to prevent harm caused by invalid information. The principle of safeguarding privacy is also emphasized in QS. An-Nur (24:27) (Badan Litbang dan Diklat Kementerian Agama RI. (2019), which is relevant to efforts in protecting personal data confidentiality in the digital era.

Based on previous studies, it can be concluded that: (1) sentiment analysis is an important method for understanding public opinion on strategic issues, (2) machine learning algorithms such as SVM and Naïve Bayes have proven effective in sentiment classification, (3) transformer-based models such as ALBERT offer potential to enhance analysis performance with computational efficiency, and (4) Islamic values provide an ethical basis for ensuring the validity and privacy of information. These points serve as the foundation for this research in developing a sentiment analysis system on public opinion regarding information security in social media.

3.1. Dataset Collection

The dataset was collected by extracting comments from Facebook on specific topics relevant to the research. The dataset was divided into three subsets: training data (80%), validation data (10%), and testing data (10%). Facebook was chosen as the data source due to its high level of user interaction and the diversity of opinions that reflect public sentiment (Alfatah, 2024).

3.2. Pre-Processing

The pre-processing stage was carried out to prepare the data for analysis. The steps included:

a. Case Folding : converting all characters to lowercase.
b. Remove Punctuation: removing irrelevant punctuation marks.
c. Tokenization : splitting text into tokens/words.
d. Stopword Removal : eliminating common words that carry little meaningful information.
e. Label Encoding : assigning labels to the data according to sentiment categories (positive, neutral, negative).

The results of pre-processing served as the primary input before proceeding to the augmentation stage.

3.3. Teknik Augmentasi

To address class imbalance in the dataset, this study applied the Back Translation data augmentation technique, which involves translating text into another language and then back into Indonesian to generate new sentence variations [10]. This technique is expected to provide more diverse data, thereby improving model performance in sentiment classification.

3.4. Experiment

The experiment was conducted by applying the augmentation technique using Back Translation, which has been shown to be effective in enhancing textual data diversity (Setiono & Sari, 2025) (Natasya, 2023). Augmentation process was integrated with the ALBERT model as the primary method for sentiment classification.

The data was split using a stratified scheme: 80% training data, 10% validation data, and 10% testing data, ensuring that class distribution (positive, neutral, negative) remained balanced across subsets. This stratified scheme aligns with common practices in machine learning to maintain the quality of model evaluation and to avoid bias caused by uneven class distribution (Setiono & Sari, 2025).

3.5. Feature Extraction and Model

This study employed ALBERT (A Lite BERT for Self-Supervised Learning of Language Representations) as the primary model for feature extraction and classification (Lan et al., 2020). ALBERT was chosen because it has fewer parameters than BERT while still being capable of producing high-performance language representations (Lan et al., 2020) (Setiono & Sari, 2025).

The process begins with pre-processed text input, followed by tokenization using the ALBERT tokenizer with the addition of special tokens [CLS] and [SEP]. The resulting embeddings (token, position, and segment embeddings) are processed through the ALBERT encoder with a self-attention mechanism. The representation of the [CLS] token is then used for sentiment classification, passed into a fully connected layer, and followed by the softmax activation function to generate class probabilities (positive, neutral, negative).

3.6. Evaluation

Model evaluation was carried out using several common metrics in NLP research (Dalianis, 2018) namely:

a. Accuracy : measures the percentage of correct predictions.
b. Precision : measures the correctness of positive predictions.
c. Recall : measures the model’s ability to detect positive classes.
d. F1-Score : mean of precision and recall.

These metrics were computed using a confusion matrix, which provides a detailed view of the model’s performance for each class (positive, neutral, negative) (Sun & Sun, 2017).

4. Results and Discussion

4.1. Dataset Distribution and Augmentation Process

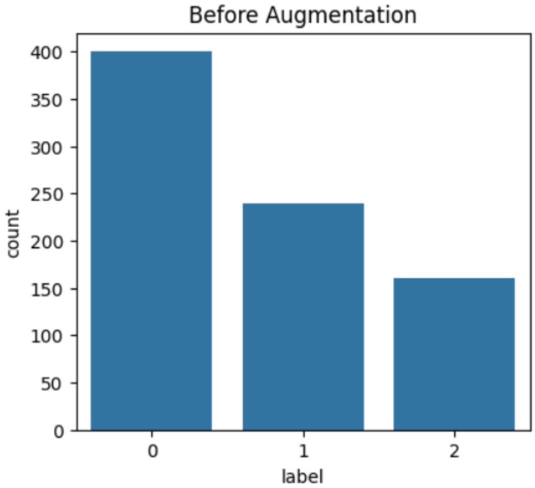


Figure 2. Dataset before Augmentation

Figure 2 illustrates the dataset distribution before augmentation. Label 0 represents the negative sentiment class, label 1 represents the neutral class, and label 2 represents the positive class. It is evident that the initial dataset was imbalanced, with the negative class (0) dominating at around 400 comments, followed by the neutral class (1) with approximately 240 comments, and the positive class (2) with around 170 comments.

This imbalance may reduce model performance, particularly in recognizing classes with relatively fewer data (positive and neutral). Therefore, this study applied the Back Translation technique as a data augmentation method. By translating text into another language (e.g., English) and then back into Indonesian, new variations of user comments were generated without altering their original meaning. The augmentation process aimed to:

a. Balance the number of data samples across classes, resulting in a more proportional distribution.
b. Increase the diversity of data fed into the ALBERT model training process.
c. Reduce the risk of model bias toward the majority class.

After augmentation, the dataset distribution was expected to become nearly balanced across the three classes, thereby enabling a more optimal training process. Visualization of the augmentation results is shown in Figure 3.

4.2. Augmentation Results

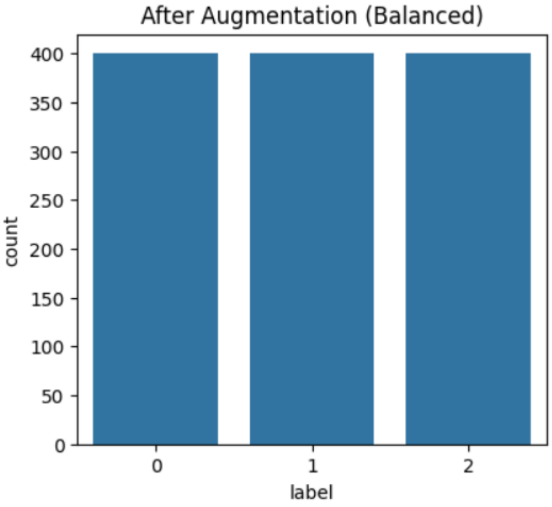


Figure 3. Augmentation Results

Figure 3 shows the dataset distribution after applying the Back Translation augmentation method. It can be observed that the number of samples in each class became balanced, with approximately 400 comments in each sentiment class: negative (0), neutral (1), and positive (2).

The comparison between Figure 2 (before augmentation) and Figure 3 (after augmentation) demonstrates that the augmentation technique successfully addressed the data imbalance. This is crucial since models tend to become biased toward the majority class when the dataset is imbalanced.

With balanced data, the ALBERT model had an equal opportunity to learn representations from each sentiment class. This process was expected to improve accuracy, precision, recall, and F1-score, particularly for the minority classes (positive and neutral) that previously had fewer samples.

4.3. Model Training Results

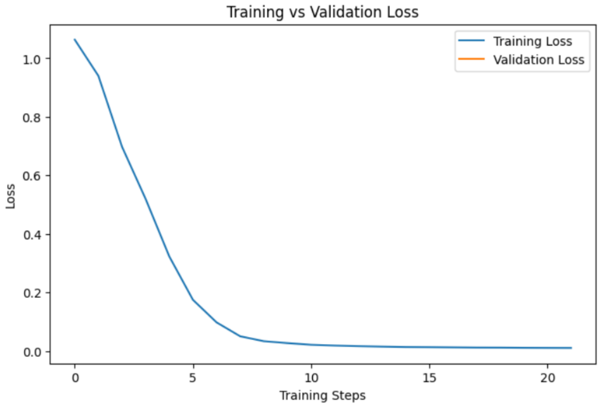


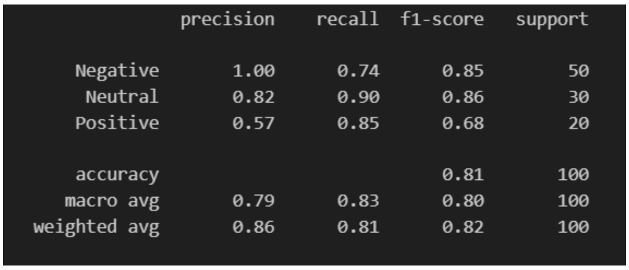
Figure 4. Training and Validation Loss Graph

The training of the ALBERT model was conducted using the augmented dataset. Figure 4 presents a comparison between training loss and validation loss at each training step. It can be seen that the training loss decreased consistently from the beginning and approached near zero after several training iterations, indicating that the model effectively learned patterns from the data.

Additionally, the validation loss remained low and stable, suggesting that the model did not suffer from significant overfitting. In other words, the model's performance on the validation data was aligned with its performance on the training data, demonstrating good generalizability to unseen test data.

These results indicate that the use of the Back Translation augmentation technique successfully enriched data diversity, thereby enabling a more optimal training process for the ALBERT model.

4.4. Model Evaluation



```
              precision    recall  f1-score   support

    Negative       1.00      0.74      0.85        50
     Neutral       0.82      0.90      0.86        30
    Positive       0.57      0.85      0.68        20

    accuracy                           0.81       100
   macro avg       0.79      0.83      0.80       100
weighted avg       0.86      0.81      0.82       100
```

Figure 5. Classification Report

The evaluation was conducted using an external dataset to measure the model's generalization ability. The classification report and confusion matrix are presented in Figure 5. The model achieved an accuracy of 81%, with a macro average precision of 0.79, recall of 0.83, and F1-score of 0.80. The weighted average scores were also high, with an F1-score of 0.82, indicating consistent performance despite differences in class distribution. Class results are as follows:

a. Negative Class obtained perfect precision (1.00) but relatively lower recall (0.74), meaning the model was highly precise in predicting negative sentiment but still misclassified some negative data as positive.
b. Neutral Class demonstrated the most balanced performance, with precision of 0.82, recall of 0.90, and F1-score of 0.86. This shows that the model was highly effective in recognizing neutral text.
c. Positive Class had lower precision (0.57), although recall was relatively high (0.85). This indicates that misclassification still occurred when the model attempted to distinguish positive text from other classes.

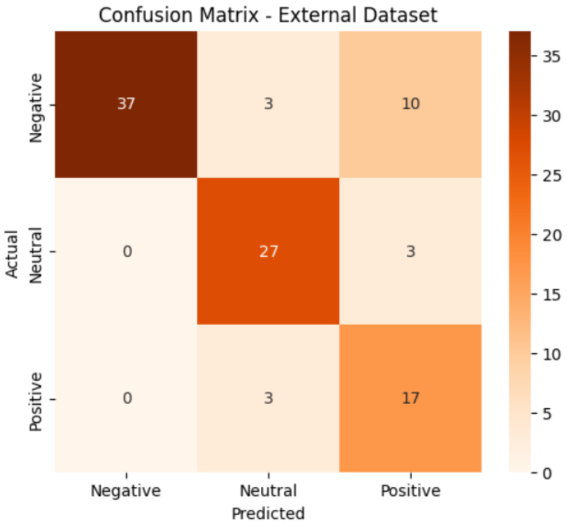


Figure 6. Confusion Matrix

The confusion matrix shows that most misclassifications occurred between the Negative and Positive classes, whereas the Neutral class was more stable and consistently predicted.

Overall, the evaluation results indicate that the ALBERT model with data augmentation achieved good and balanced performance in sentiment analysis, though improvements are still needed particularly for the Positive class.

5. Conclusion and Suggestions

This study demonstrates that the combination of Back Translation data augmentation and the ALBERT model is effective for sentiment analysis. The evaluation results achieved an accuracy of 81% with a macro F1-score of 0.80, where the Neutral class was more stable compared to the Positive class. For future development, it is recommended to explore other augmentation methods, employ different transformer-based models, and expand the dataset to further optimize model performance and enhance its applicability in real-world systems.

References

Aldean, P., et al. (2021). Analisis sentimen masyarakat terhadap vaksinasi Covid-19 di Twitter menggunakan metode random forest classifier (studi kasus: vaksin Sinovac). Dalam *Prosiding Seminar Nasional Teknologi dan Rekayasa*. Surabaya, Indonesia.

We Are Social, & Kepios. (2023). *Digital 2023: Indonesia*. We Are Social.

Badan Litbang dan Diklat Kementerian Agama RI. (2019). *Al-Qur'an dan terjemahannya*. Kementerian Agama RI.

Putranta, Y. S. V., Rachmad, B., & Prasetyo, W. (2023). Analisis sentimen masyarakat terhadap kebijakan penghapusan subsidi BBM pada media sosial Twitter menggunakan algoritma naïve Bayes classifier dengan ekstraksi fitur n-gram TF-IDF. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer (J-PTIIK)*.

Syahriani, I., Yusuf, A. A., & Yulianto, T. S. (2010). Sentiment analysis of Facebook comments on Indonesian presidential candidates using naïve Bayes with feature selection. Dalam *Journal of Physics: Conference Series*.

Lan, Z., et al. (2020). ALBERT: A lite BERT for self-supervised learning of language representations. Dalam *Proceedings of the International Conference on Learning Representations (ICLR)*. Addis Ababa, Ethiopia.

Kharisudin, I., & Putra, M. I. (2022). Analisis sentimen pengguna aplikasi marketplace Tokopedia pada situs Google Play menggunakan metode support vector machine (SVM), naïve Bayes, dan logistic regression. Dalam *PRISMA (Prosiding Ilmiah Mahasiswa) Universitas Negeri Semarang*.

Julianto, M. R., & Wibowo, Y. A. T. (2024). Analisis sentimen respon publik terhadap program internet gratis di platform X melalui pendekatan algoritma naïve Bayes. *Jurnal Indonesia: Manajemen Informatika dan Komunikasi (JIMIK), 5*(3).

Witanti, A., & Dian, H. (2022). Analisis sentimen masyarakat terhadap vaksinasi COVID-19 pada media sosial Twitter menggunakan algoritma support vector machine (SVM). *Jurnal Sistem Informasi dan Informatika (Simika), 5*(1).

Maulidiana, D. R., Hidayat, M. F., & Ramadhan, D. G. P. D. (2024). Comparative analysis of SVM and NB algorithms in evaluating public sentiment on Supreme Court rulings. *Jurnal Sisfokom, 13*(2).

Widyadhana, F. K., Sari, N. Y., & Rachmad, B. (2023). Sentiment analysis pada opini masyarakat terhadap pelayanan publik Polres Ponorogo menggunakan metode support vector machine. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer (J-PTIIIK), 7*(7).

Liu, B. (2012). *Sentiment analysis and opinion mining*. Morgan & Claypool Publishers.

Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. *Ain Shams Engineering Journal, 5*(4), 1093–1113. https://doi.org/10.1016/j.asej.2014.04.011

Hadikristanto, W., & Tri, A. (2023). Implementasi algoritma naïve Bayes dan support vector machine tentang pembobolan dan kebocoran data di Twitter. *BIT Jurnal: Jurnal Teknologi Informasi, 5*(2).