Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 4, October 2024
DOI : https://doi.org/10.33395/sinkron.v8i4.14040
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Sentiment Analysis on KPU Performance Post-2024 Election via YouTube Comments Using BERT

Nafiatun Sholihah[1]**, Ferian Fauzi Abdulloh**[2]*****, Majid Rahardi**[3]
[1,2,3]Universitas AMIKOM Yogyakarta, Indonesia
[1]nafiatun@amikom.ac.id, [2]ferian@amikom.ac.id, [3]majid@amikom.ac.id

Submitted : Aug 18, 2024 | **Accepted** : Aug 29, 2024 | **Published** : Oct 02, 2024

Abstract: This research aims to analyze public sentiment regarding the performance of the General Election Commission after the 2024 presidential election using the BERT (Bidirectional Encoder Representations from Transformers) model. Given the General Election Commission's crucial role in maintaining election integrity and the importance of transparency in Indonesian democracy, understanding public opinion through sentiment analysis is essential. Data was collected from YouTube comments, a platform increasingly popular for public expression. The analysis process began with data preprocessing, including case folding, text cleaning, tokenization, and stop word removal. The BERT model was then applied to classify the sentiment of the comments, with the model's performance evaluated using 10-fold cross-validation. The evaluation results showed that the first fold (k=1) achieved the best performance with an accuracy of 96%, precision of 96%, recall of 96%, and an F1-score of 96%, indicating the model's effectiveness in accurately classifying sentiment. In contrast, the ninth fold (k=9) exhibited the lowest accuracy at 86% with other metrics also lower, suggesting performance instability potentially caused by data variability. Accuracy and loss graphs confirmed that the first fold experienced consistent accuracy improvements and significant loss reduction, while the ninth fold showed performance fluctuations. This study provides valuable insights into public sentiment regarding the General Election Commission performance, with BERT demonstrating significant potential for sentiment analysis on social media platforms like YouTube.

Keywords: BERT, ML, NLP, Optimization, Sentiment Analysis

INTRODUCTION

The General Elections Commission (KPU) of Indonesia holds significant responsibility in ensuring the integrity and transparency of elections, which are fundamental pillars of democracy in Indonesia (KPU, 2013) .Evaluating the performance of the KPU after the 2024 presidential election is a key focus, given its crucial role in determining the future direction of democracy in Indonesia. Public trust and credibility towards the KPU are essential, learning from past issues of corruption, collusion, and nepotism during the New Order era. Transparency and accountability are crucial. In the digital age, public opinions about government performance often emerge on platforms like YouTube (Rahman Isnain et al., 2021; Thomas et al., 2021). Therefore, sentiment analysis of YouTube comments regarding the KPU's performance post-2024 presidential election is important to understand public perception.

Public satisfaction or dissatisfaction with the KPU’s performance post-2024 presidential election can be seen as a reaction to the gap between expectations and reality. Opinions expressed on platforms like YouTube contain emotional viewpoints that can be analyzed to uncover sentiments such as happiness, sadness, or anger. This sentiment analysis method allows for a deeper understanding of the public's emotional responses, though there are challenges in standardizing emotions. This analysis is valuable for election organizers to comprehend public feedback. With the massive volume of public opinions on digital media, challenges arise in understanding each opinion and obtaining accurate insights (Halim et al., 2022). This research will utilize the BERT algorithm to classify sentiments in YouTube comments related to the 2024 presidential election and KPU performance. Data will undergo preprocessing including case folding, text cleaning, tokenizing, and removal of stop words before analysis. Model performance will be measured using 10-fold cross-validation and evaluated with a confusion matrix.

*Ferian Fauzi Abdulloh

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 4, October 2024
DOI : https://doi.org/10.33395/sinkron.v8i4.14040
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Previous research on KPU performance analysis includes two studies. The first, using the Naïve Bayes method, involved a dataset of 2,475 entries, consisting of 321 positive, 409 neutral, and 1,745 negative entries. The confusion matrix testing using Python programming showed a negative accuracy rate of 74.5%. This indicates that the public's trust in the KPU's performance leading up to the 2024 election could be considered low (Fuad Amirullah et al., 2023). The second study used Naïve Bayes with Particle Swarm Optimization (PSO) and processed a dataset of 1,000 entries. Evaluation using the confusion matrix revealed an accuracy of 73.67% for Naïve Bayes without PSO and 78.33% for Naïve Bayes with PSO. This shows that the accuracy of the Naïve Bayes algorithm with PSO is superior compared to the one without PSO (Putra et al., 2023).

The two previous studies on KPU performance were based on comments from social media X, but it cannot be denied that YouTube is now the most popular video-sharing platform. YouTube is one of many multimedia-sharing social networks. According to a report from We Are Social, as of October 2023, the number of active YouTube users globally has reached 2.49 billion. In Indonesia, there are 139 million YouTube users, making it the fourth largest country in terms of YouTube users worldwide. YouTube, owned by Google, ranks second among the most popular social media platforms globally, after Facebook (Mutia Annur, 2023). Opinions expressed by YouTube viewers can be valuable assets for the growth of companies or YouTubers. The method used to analyze these comments is known as sentiment analysis or opinion mining. Although not new, understanding this field remains crucial given the value of viewer.

Sentiment analysis is a part of Natural Language Processing (NLP) aimed at identifying or analyzing opinions or sentiments (Alifia Putri & Al Faraby, 2020). This is achieved using computational and statistical techniques to recognize, extract, and process information from text related to sentiment. Thus, sentiment analysis can help understand people's perceptions and responses to a topic, product, or service, and assist in decision-making in various fields, such as business and politics. However, drawing hidden conclusions from viewer comments is challenging, especially with conventional methods. Traditional approaches to sentiment analysis often face limitations in understanding language and context. For example, negative reviews may arise from misunderstandings, while praise could contain irony. In such cases, the BERT (Bidirectional Encoder Representations from Transformers) approach has introduced new innovations. BERT is a Google-developed language model that examines words and understands the relationships between words and sentence context (Alaparthi & Mishra, 2020). Pre-trained BERT models that undergo fine-tuning for sentiment analysis have achieved accurate results. Evaluations have shown that the SmallBERT model, fine-tuned with a dataset of 515,000 hotel reviews over 5 epochs, achieved an accuracy of 91.40%, precision of 90.51%, recall of 90.51%, and an F1 score of 90.51% when evaluated with manually labeled datasets. Visualization of sentiment comparison shows a dominance of positive sentiment, performed using Tableau (Vidya Chandradev et al., 2023).

Bidirectional Encoder Representations from Transformers (BERT) is an advanced algorithm in natural language processing (NLP) that enhances computer understanding of human language context and nuances (Pardamean & Pardede, 2021). BERT is effective in sentiment analysis for interpreting emotions in text, such as positive, negative, or neutral sentiments. Its advantage lies in its ability to understand the context of words within sentences, providing more accurate sentiment analysis compared to traditional algorithms. BERT is useful for handling large volumes of social media data, facilitating efficient understanding of public opinions. However, implementing BERT requires careful parameter selection, thorough evaluation, and testing to ensure optimal performance and data suitability. (Aljabar et al., 2024) highlights BERT's uniqueness in understanding context from both directions, producing rich meaning vector representations. Using the Hugging Face library and BERT model, 5,000 lines of comments were analyzed. Results showed 72% positive sentiment, 15% neutral, and 13% negative (Nayla et al., 2023). The findings of the research reveal that the BERT algorithm outperforms Random Forest and Naïve Bayes, achieving an accuracy of 0.92 and an average macro F1-score of 0.92 (Husin, 2023). This research focuses on the analysis and comparison of two classification methods: Support Vector Machine (SVM) optimized with Particle Swarm Optimization (PSO) (Sholihah et al., 2023). This research will analyze sentiments regarding the KPU's performance post-2024 presidential election using BERT. Indonesian comments will be processed through case folding, text cleaning, tokenizing, and removal of stop words. The preprocessing output will be classified using BERT, and model performance will be evaluated using 10-fold cross-validation and confusion matrix.

With the increasing popularity of YouTube, sentiment analysis of comments on this platform has become increasingly relevant. According to the We Are Social report, YouTube has a significant number of active users, including in Indonesia, with 139 million active users. This video platform ranks second among the most popular social media platforms globally, after Facebook. Opinions expressed by YouTube viewers can be valuable assets for company or YouTuber growth. Sentiment analysis using BERT allows for a more profound and accurate understanding of public opinions. BERT, with its ability to comprehend word context within sentences, can provide more accurate sentiment analysis compared to conventional methods. The implementation of BERT in this research is expected to offer valuable insights into public sentiment regarding the KPU's performance post-2024 presidential election.

*Ferian Fauzi Abdulloh

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 4, October 2024
DOI : https://doi.org/10.33395/sinkron.v8i4.14040
e-ISSN : 2541-2019
p-ISSN : 2541-044X

METHOD

In this study, the research object will be comments on YouTube videos with the keyword "KPU" in the form of Indonesian text, words, or sentences taken from YouTube social media. The text will then undergo sentiment analysis to identify and classify comments or sentiments (Positive, Negative, or Neutral) regarding the KPU's performance. Figure 1 below will illustrate the stages of data testing conducted in the experimental design.

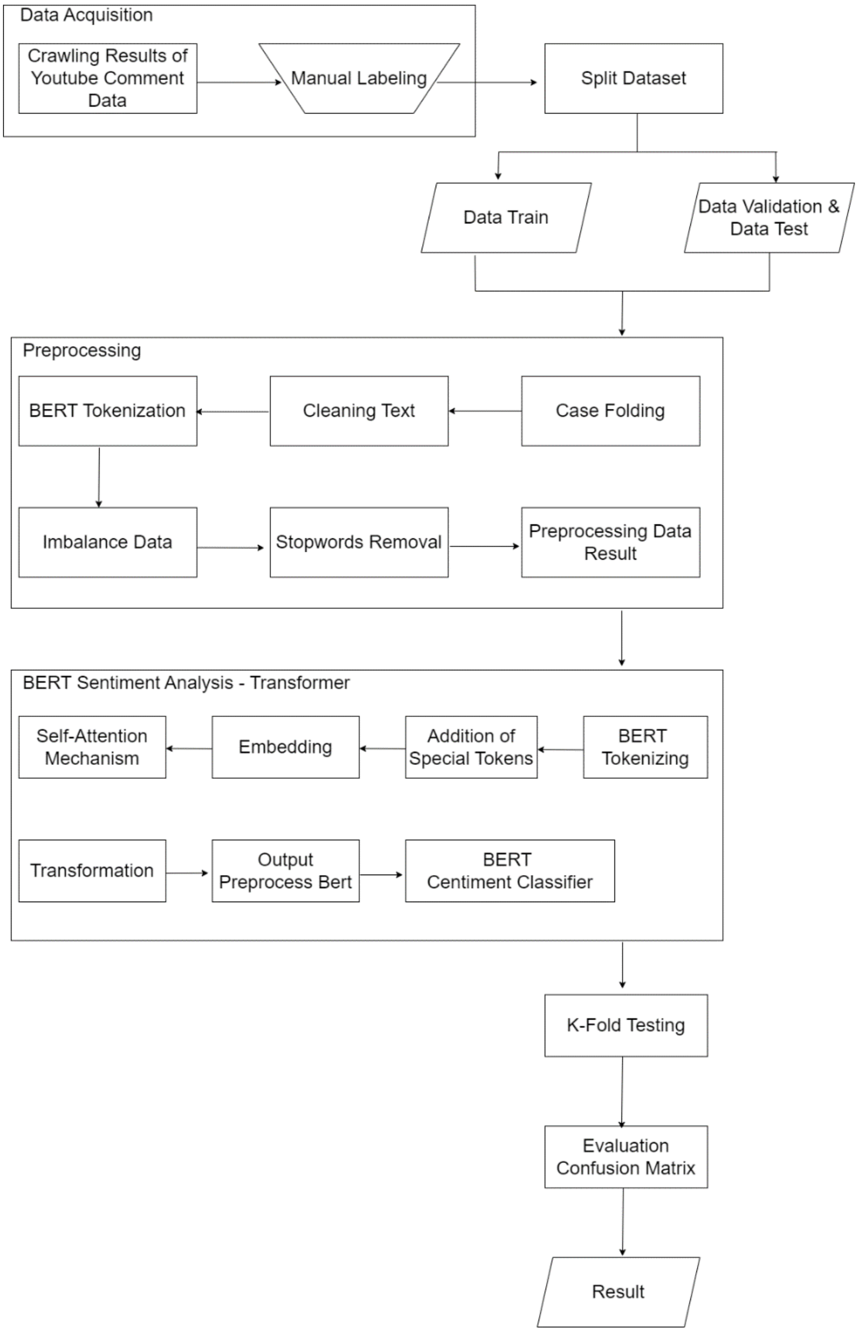

Fig 1. experimental design

Data Acquisition

The method in this study consists of data crawling and manual labeling.

Crawling

The data collected from the YouTube platform used the Python programming language with libraries such as google-api-python-client. The data retrieval process was conducted using crawling techniques, which involve extracting semi-structured comments from YouTube by utilizing the API provided by YouTube. Approximately

*Ferian Fauzi Abdulloh

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 4, October 2024
DOI : https://doi.org/10.33395/sinkron.v8i4.14040
e-ISSN : 2541-2019
p-ISSN : 2541-044X

1,000 comments were collected, including attributes such as the date and time the comment was posted, the user who posted it, and the content of the comment itself. Data collection was carried out over one month, in April 2024, focusing on comments containing the keywords “KPU” and “Pilpres 2024” from videos related to the performance of the General Elections Commission (KPU) after the 2024 presidential election.

Manual Labeling

Next, the labeling (classifying) stage of the data is conducted manually to identify sentiments (Positive, Negative, or Neutral). During this manual labeling stage, at least 1,200 data points will be used.

Preprocessing

Data preprocessing aims to prepare clean data for research by performing steps such as ignoring irrelevant attributes, testing, and modifying data to facilitate understanding, ensuring data consistency, handling missing values, and reducing redundancy in the data (Kusrini, 2009). One of the stages in data preprocessing is text processing, which aims to convert unstructured text documents into structured data ready for use. The steps in text processing include:
1). Case Folding
This step is necessary to convert uppercase letters in the text to lowercase to ensure uniformity.
2). Cleaning Text
The cleaning process involves removing words that are considered irrelevant to reduce noise in the classification process.
3). Tokenizing
Tokenizing is the process of separating text data into individual tokens or units. Generally, a set of characters in a text will be split into words during tokenization.
4). Imbalance Data
In this process, the dependent variable will be checked to see the balance between positive, neutral, and negative data. If data imbalance is found, the SMOTE technique will be applied. The SMOTE technique is employed to address data imbalance by utilizing oversampling and evaluating its impact on decision boundaries within the given context (Aprilliandhika & Fauzi Abdulloh, 2024).
5). Stopword Removal
Stopwords are words that are considered to have no significant meaning, such as "kalau", "di", "dan", "dengan", "oleh", "sebagainya". Therefore, stopword removal is the process of eliminating stopwords that do not add value to the data being used.

BERT – Transformer

After preprocessing, the transformer process in the context of the BERT (Bidirectional Encoder Representations from Transformers) model proceeds as follows:
1). BERT Tokenizing: This stage involves converting words in the text into tokens that can be processed by the model. BERT uses a special tokenizer that breaks text into sub-word tokens, allowing the model to handle unknown or new words more effectively.
2). Adding Special Tokens: BERT adds special tokens at the beginning and end of each sequence. For example, each sequence starts with a special [CLS] token used for classification tasks and ends with a [SEP] token to separate sentences or segments.
3). Embedding: Each token is then converted into an embedding vector. This process involves three types of embeddings: token embeddings, which convert tokens into vectors; segment embeddings, which distinguish between different sets in tasks involving two sets (such as questions and answers); and position embeddings, which provide information about the token's position in the sequence. The combination of these three embeddings is used as the input to the model.
4). Self-Attention Mechanism: In this stage, the self-attention mechanism allows tokens to weigh and combine information from other tokens in the same sequence to obtain richer context. This enables the model to better understand the relationships and context between words.
5). Transformation: Through a series of transformer layers, each consisting of self-attention mechanisms and feed-forward neural networks, the model processes the token embeddings, gradually updating and refining them to obtain a deeper and more contextual representation of each token.
6). Output: The final representation of tokens generated by the transformer can be used for various NLP tasks, such as sentiment classification, text comprehension, and more. For classification tasks, the representation of the [CLS] token is often used as an aggregate representation of the entire sequence to make predictions.

*Ferian Fauzi Abdulloh

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 4, October 2024
DOI : https://doi.org/10.33395/sinkron.v8i4.14040
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Overall, the transformer process in BERT allows the model to gain a deep understanding of the text by considering the bidirectional context of each token in the text, making it highly effective for complex natural language processing tasks.

Classification

Using a pre-trained BERT model or fine-tuning the BERT model on the collected and preprocessed dataset. This process will adapt the BERT model for the context of sentiment analysis on YouTube comments related to the performance of the KPU.

Testing

K-fold cross-validation is a validation technique used to evaluate the effectiveness of a model in making predictions by minimizing the bias that may arise from the selection of non-representative training samples. This technique involves randomly dividing the dataset into k parts, or "folds." During the validation process, the model is run k times, where in each iteration, one part of the data is chosen as the test set and the remaining parts as the training set. This way, each part of the data has the opportunity to be used as both a test set and a training set. This allows the model to be tested and trained using all combinations of data subsets, resulting in more robust and accurate outcomes.

Evaluation

The final step, the effectiveness of the classification results will be assessed using accuracy scores obtained from evaluation using the Confusion Matrix. This evaluation process utilizes the Confusion Matrix to assess performance, comparing the predicted classes with the actual classes. Using the structure of the Confusion Matrix, the accuracy of the predictions is calculated, providing an overview of how accurately the model classifies the data. This matrix enables the calculation of important evaluation metrics such as accuracy, precision, and recall, using formulas (1), (2), (3), and (4).

$$Accuracy = \frac{\mathrm{TP+TN}}{\mathrm{TP+TN+FP+FN}} \quad (1)$$

$$Precision = \frac{\mathrm{TP}}{\mathrm{FP+TP}} \quad (2)$$

$$Recall = \frac{\mathrm{TP}}{\mathrm{FN+TP}} \quad (3)$$

$$F1 - Score = 2 * \frac{\mathrm{Precision+recall}}{\mathrm{Precision+recall}} \quad (4)$$

RESULT

Crawling Data

In this research, YouTube comments data were collected using the Google Client API. The process began with API configuration, which involved creating and setting up credentials in the Google Cloud Console to obtain the necessary API key. After collecting the YouTube comments data using the Google Client API, which resulted in 1200 comments, the next step was to convert the data into CSV format for easier further analysis. CSV (Comma-Separated Values) format is a popular choice due to its simplicity and ease of integration with various data analysis tools. In the CSV file, the comments data is organized in the following column: 'text_comment'. The collected YouTube comments data were converted into CSV format using Python libraries such as pandas. The data was structured as a dictionary, which was then converted into a DataFrame before being saved as a CSV file.

Manual Labeling

In the manual labeling phase, each collected comment will be evaluated manually to determine its sentiment. Evaluators will read each comment and classify it as 'Positive', 'Neutral', or 'Negative'. The results of this evaluation will be recorded in a new column in the CSV file named 'Sentiment'. This 'Sentiment' column will contain the sentiment labels for each comment, ensuring that each data entry has an accurate sentiment assessment that can be used as training data for sentiment analysis models such as BERT. The results of the manual labeling for the dataset are shown in Table 1 below.

*Ferian Fauzi Abdulloh

e-ISSN : 2541-2019
p-ISSN : 2541-044X

Tabel 1. Sample Dataset on Manual Labeling

Text	Sentiment
Kebanyakan alasan, ketua KPU brengsek tak bisa kelola proses pemilu secara professional dan independen. Audit forensik SIREKAP harus dilakukan untuk menjelaskan kebodohan komisioner KPU. Usut terus kebohongan KPU tentang penyimpanan data di server milik alibaba di singapure.	0
Muak penuh kebohongan jahat kalian bohong semoga anda di laknat 7 turunan	0
Kalau memang ada kecurangan sebaiknya pemilu sekarang ini diulang saja agar bisa mencapai hasil perhitungan sesuai dengan kenyataan yang ada dilapangan dan perlu penjagaan super ketat agar tidak.ada.lagi kecurangan kecurangan yang dilakukakan oleh calon tertentu melalui orang2.suruanya kalau masih ada kecurangan muda2han Allah memusnakan para pelaku kecutangan tersebut	1
Semangat... Indonesia negara yg hebat	2
Terima kasih KPU dan semua pihak pemilu 2024 sukses	2
Untuk ketua kpu dan jajarannya, ingat kematian itu tidak bisa ditolak, bisa datang kapan saja. Dan ingat pula apa yang telah anda lakukan terhadap rakyat, yang akan di kenang selamanya, dan pasti nya berimbas pada keturunan anda.	1

Preprocessing

Preprocessing is a crucial step in building artificial intelligence models. In this research, several processes were performed, including removing null values, applying case folding, removing punctuation and emojis, eliminating duplicate data, separating letters and numbers in a word, removing leading dashes or spaces, addressing class imbalance through oversampling techniques, and performing tokenization.

This phase is part of preprocessing known as data cleaning, where data with null values are identified and removed. Removing null values means eliminating rows that are empty or do not have any value. The results before and after removing null values are shown in Table 2.

Table 2. Result before and after data cleaning

	Text	sentiment
Before	1233	1231
After	1231	1231

Table 2 presents data before and after data cleaning. In the 'text' column, there were null values, while the 'sentiment' column did not contain any null values. After checking, the next step was to perform data cleaning to remove rows containing null values, resulting in the removal of 2 rows with null values.

Following the data cleaning process, the next step is case folding. The purpose of case folding is to standardize all letters to lowercase, eliminating the differences between uppercase and lowercase letters that could affect text analysis. By applying case folding, words like "KPU", "kpu", and "Kpu" will all be converted to "kpu". The results of case folding are shown in Table 3.

*Ferian Fauzi Abdulloh

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 4, October 2024 e-ISSN : 2541-2019
DOI : https://doi.org/10.33395/sinkron.v8i4.14040 p-ISSN : 2541-044X

Table 3. Result Case Folding

Before	Kinerja KPU dalam Pilpres 2024 sangat baik
After	kinerja kpu dalam pilpres 2024 sangat baik

Table 3 shows the process of converting all characters or letters to lowercase, including both previously uppercase and already lowercase letters.

The next step is handling data imbalance, which is present in the 'sentiment' column. To address this issue, SMOTE (Synthetic Minority Over-sampling Technique) is used to apply an oversampling method to manage the data imbalance.

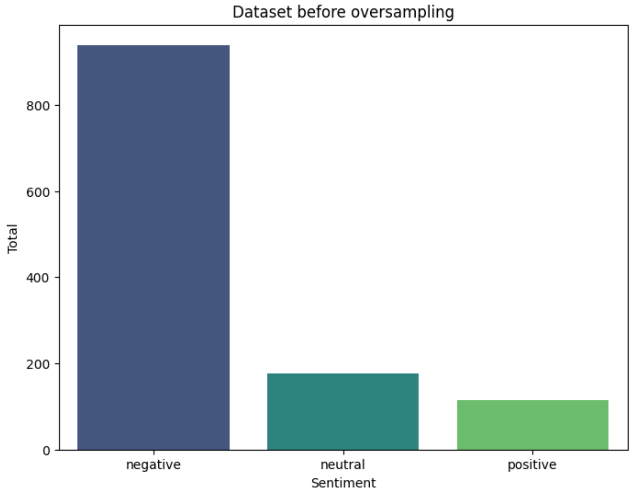

Fig 2. Before SMOTE

Fig 2 shows the imbalance between the negative, neutral, and positive data, where negative data comprises 900 entries, neutral data has fewer than 200 entries, and positive data has 100 entries. To address this issue, the SMOTE method was used, resulting in a more balanced data distribution as shown in Fig 3. Before applying SMOTE, the neutral and positive data had fewer than 200 entries each, but after applying SMOTE, the data became more balanced. With the SMOTE method, synthetic data is generated for the minority classes by averaging the nearest neighbors. This helps increase the number of samples in the minority classes, making their distribution closer to that of the majority class and creating a more balanced dataset. It is important to note that SMOTE implementation was carried out without diminishing the quality of the original data, ensuring that the dataset still reflects the original variation and complexity.

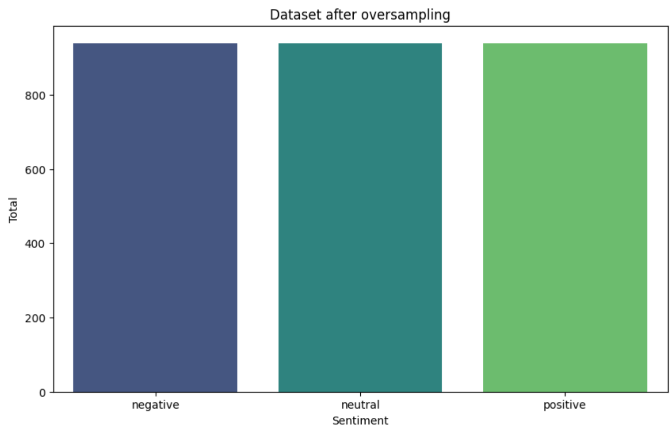

Fig 3. Result after SMOTE

*Ferian Fauzi Abdulloh

SinkrOn
JURNAL & PENELITIAN TEKNIK INFORMATIKA
Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 4, October 2024
DOI : https://doi.org/10.33395/sinkron.v8i4.14040
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Overall, the data preprocessing stage is a fundamental step in this research to ensure the accuracy, reliability, and consistency of the dataset before applying it to the BERT model. By conducting thorough preprocessing, the resulting dataset is better prepared and suitable for training the BERT model, leading to a more accurate and reliable model.

Evaluation

In this research, testing was performed using k-fold cross-validation with values of k ranging from 1 to 10. Each fold was used to evaluate the performance of the BERT model in classifying YouTube comment sentiments. The testing was conducted with a maximum of 3 epochs. Each iteration of k-fold cross-validation involved training and testing the model, with the data split into 80% for training, 10% for validation, and 10% for testing. The results of the k-fold cross-validation are presented in Table 3.

K	Accuracy	Precision	Recall	F1-score
1	96%	96%	96%	96%
2	92%	91%	92%	91%
3	94%	94%	94%	94%
4	92%	92%	91%	91%
5	90%	89%	89%	89%
6	94%	94%	94%	93%
7	93%	93%	93%	93%
8	94%	94%	94%	94%
9	86%	87%	86%	86%
10	89%	91%	89%	89%

Table 3. Kfold Result

The testing results, the first fold (k=1) achieved the highest accuracy of 96%, followed by the third fold (k=3), the sixth fold (k=6), and the eighth fold (k=8), each with an accuracy of 94%. These values indicate that in the first fold, the BERT model performed exceptionally well in classifying YouTube comment sentiments, achieving the highest accuracy among all tested folds. The optimal performance in the first fold suggests that the BERT model learned very effectively from the training data and was able to generalize well to the test data. The high accuracy, along with other metrics such as precision, recall, and F1-score, all showing high values (all 96%), indicates that the model is not only accurate in its predictions but also consistent in identifying sentiments precisely. This can be observed in the differences in the confusion matrix between k-fold 1 and k-fold 9, as shown in Fig 4.

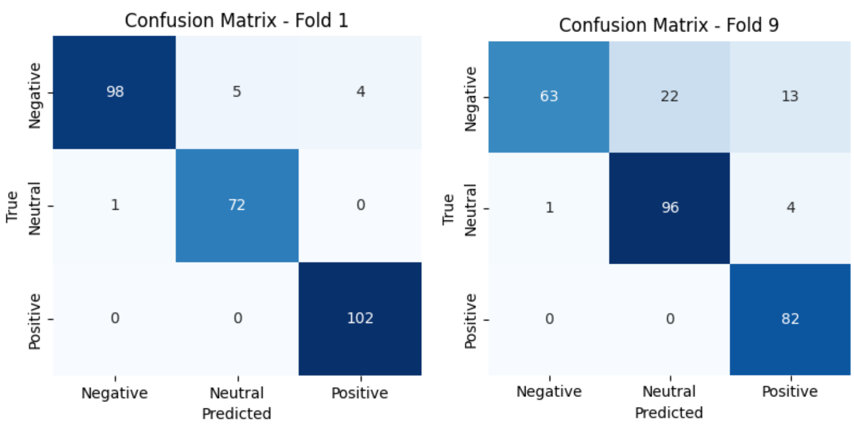

Fig 4. Confussion marix kfold 1 and kfold 9

Contrast, the ninth fold (k=9) exhibited relatively lower results with an accuracy of 86%, and other metrics were also lower compared to the other folds. This indicates instability in the model's performance, which could be due to variations in the data or differences in the distribution within the data subsets used in that fold. The visualizations of accuracy and F1-score for each k-fold are shown in Fig 5.

*Ferian Fauzi Abdulloh

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 4, October 2024
DOI : https://doi.org/10.33395/sinkron.v8i4.14040
e-ISSN : 2541-2019
p-ISSN : 2541-044X

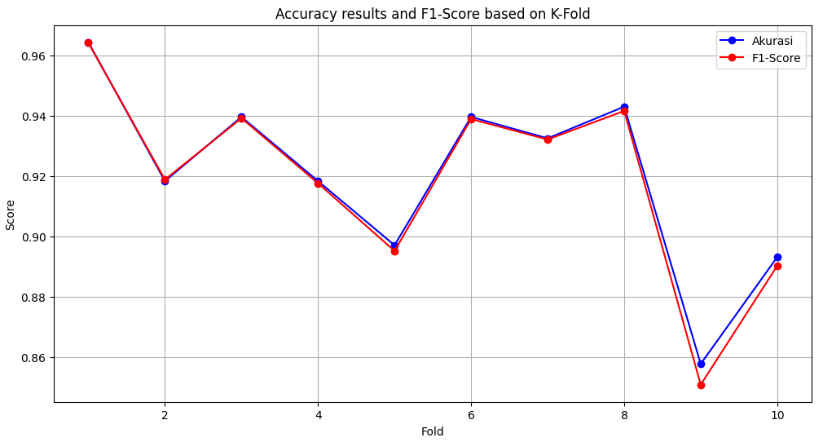

Fig 5. Visual Accuracy and F1 – Score From kfold 1 – 10

In Figure 5, the first fold (k=1) demonstrates the best performance, achieving an accuracy of 96%, with precision, recall, and F1-score all at 96%. This result indicates that the BERT model can classify YouTube comment sentiments with high accuracy in this fold. The second fold (k=2) has an accuracy of 92%, with precision, recall, and F1-score at 91%, 92%, and 91%, respectively. Although slightly lower than the first fold, these metrics still reflect strong performance. The third fold (k=3) reaches an accuracy of 94%, with precision, recall, and F1-score all at 94%, indicating performance almost comparable to the first fold. The fourth fold (k=4) shows an accuracy of 92%, with precision at 92%, recall at 91%, and F1-score at 91%, indicating a slight decrease in performance compared to the first fold. The fifth fold (k=5) records an accuracy of 90%, with precision, recall, and F1-score at 89%, showing a more noticeable decline in model performance, as evidenced by increased training and validation loss. The sixth fold (k=6) achieves an accuracy of 94%, with precision, recall, and F1-score at 94% and 93%, respectively, showing good performance stability with consistent loss reduction. The seventh fold (k=7) shows an accuracy of 93%, with precision, recall, and F1-score all at 93%, indicating a slight decrease from the previous folds. The eighth fold (k=8) reaches an accuracy of 94%, with precision, recall, and F1-score all at 94%, demonstrating good performance stability, similar to the third and sixth folds. The ninth fold (k=9) has the lowest accuracy at 86%, with precision at 87%, recall at 86%, and F1-score at 86%, reflecting significant classification errors and performance fluctuations during the epochs. The tenth fold (k=10) achieves an accuracy of 89%, with precision at 91%, recall at 89%, and F1-score at 89%, showing better results than the ninth fold but not as high as the first fold. It indicates better performance stability compared to the ninth fold but remains less optimal than the first fold.

DISCUSSIONS

The results of this study indicate that the BERT model is able to consistently classify public sentiment related to KPU performance with very good performance, especially in the first fold, with an accuracy, precision, recall, and F1-score of 96%. This finding is consistent with previous research by (Aljabar et al., 2024) which found that the BERT-based approach is very effective in sentiment analysis, especially for Indonesian language data. In this study, BERT successfully captured important patterns in YouTube comments, which supported KPU's performance, as also discussed by (Muhammad Arief Rahman et al., 2019) which underlined the importance of managing public sentiment on social media platforms.

However, it should be noted that in the ninth fold, the model's performance decreased, with an accuracy of 86%, precision of 87%, recall of 86%, and F1-score of 86%. This decrease may be due to the imbalance of data distribution, which is a common challenge in sentiment analysis (Pradany & Fatichah, 2016), which found that uneven data distribution often causes a decrease in model performance in classification tasks. This factor is reinforced who stated that variations in data characteristics can cause difficulties for the model in maintaining prediction consistency.

Furthermore, the limitations of this study include the limited number of epochs (only three epochs), which may not be enough to achieve optimal convergence in several folds, (Halim et al., 2022). They showed that increasing the number of epochs can often improve model performance in classification tasks. Nevertheless, the results of the first to eighth folds support the argument that the BERT model is consistently able to capture the deep meaning of the text, stating that transformer-based models such as BERT are effectively able to process natural language data. In this study, the attention mechanism used by BERT. Allows the model to understand a broader context, which contributes to excellent results in most folds.

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 4, October 2024
DOI : https://doi.org/10.33395/sinkron.v8i4.14040

e-ISSN : 2541-2019
p-ISSN : 2541-044X

CONCLUSION

This research evaluates the effectiveness of the BERT model in analyzing sentiments related to the performance of the General Election Commission based on YouTube comments. Through testing using k-fold cross-validation with values of k ranging from 1 to 10, we gained in-depth insights into the performance of the BERT model in sentiment classification. The results show that the first fold achieved the highest accuracy of 96%, indicating that the BERT model is highly effective in classifying sentiments with a high degree of accuracy for this data. However, despite the excellent results in the first fold, there was variation in performance across other folds, with the lowest accuracy recorded in the ninth fold (86%). This variation suggests that while BERT is robust in sentiment classification, challenges in performance consistency remain. The performance drop in some folds may be attributed to uneven data distribution or differences in data characteristics across folds. This research provides significant insights into the strengths of the BERT model in sentiment analysis, particularly in the context of KPU performance. The results highlight opportunities for further improvements, such as enhancing preprocessing techniques or adjusting the model to increase performance consistency. Understanding the strengths and limitations of the BERT model provides a foundation for advancing its application in sentiment analysis.

ACKNOWLEDGMENT

We extend our deepest gratitude and appreciation to DRTPM/LLDIKTI5 and LPPM-Universitas Amikom Yogyakarta for their support and funding of this research, under the contract numbers with LPPM: 029/KONTRAK-LPPM/AMIKOM/VI/2024 dated June 19, 2024; LLDIKTI: 0609.20/LL5-INT/AL.04/2024 dated June 14, 2024; and DRTPM with LLDIKTI: 107/E5/PG.02.00.PL/2024 dated July 11, 2024. We also wish to acknowledge the contributions of all team members and other parties involved in this research..

REFERENCES

Alaparthi, S., & Mishra, M. (2020). *Bidirectional Encoder Representations from Transformers (BERT): A sentiment analysis odyssey*.

Alifia Putri, C., & Al Faraby, S. (2020). *Analisis Sentimen Review Film Berbahasa Inggris Dengan Pendekatan Bidirectional Encoder Representations from Transformers*. *6*(2), 181–193.

Aljabar, A., Karomah, B. M., Kunci, K., & Bert, : (2024). Mengungkap Opini Publik: Pendekatan BERT-based-caused untuk Analisis Sentimen pada Komentar Film. In *Journal of System and Computer Engineering (JSCE) ISSN* (Vol. 5, Issue 1).

Aprilliandhika, W., & Fauzi Abdulloh, F. (2024). *Comparison of K-Nearest Neighbor and Support Vector Machine Algorithm Optimization With Grid Search Cv on Stroke Prediction Komparasi Optimasi Algoritma K-Nearest Neighbor Dan Support Vector Machine Dengan Grid Search Cv Pada Prediksi Stroke*. *5*(4), 991–1000.

Fuad Amirullah, Syariful Alam, & M.Imam Sulistyo S. (2023). Analisis Sentimen Terhadap Kinerja KPU Menjelang Pemilu 2024 Berdasarkan Opini Twitter Menggunakan Naïve Bayes. *STORAGE: Jurnal Ilmiah Teknik Dan Ilmu Komputer*, *2*(3), 69–76. https://doi.org/10.55123/storage.v2i3.2293

Halim, F., Liliana, & Gunadi, K. (2022). Ringkasan Ekstraktif Otomatis pada Berita Berbahasa Indonesia Menggunakan Metode BERT. In *Jurnal Infra* (Vol. 10, Issue 1).

Husin, N. (2023). Komparasi Algoritma Random Forest, Naïve Bayes, dan Bert Untuk Multi-Class Classification Pada Artikel Cable News Network (CNN). In *Jurnal Esensi Infokom : Jurnal Esensi Sistem Informasi dan Sistem Komputer* (Vol. 7, Issue 1). https://doi.org/10.55886/infokom.v7i1.608

KPU. (2013). *Peraturan Komisi Pemilihan Umum Nomor 20 Tahun 2018*. https://peraturan.bpk.go.id/Details/173318/peraturan-kpu-no-23-tahun-2018

Kusrini, L. T. E. ,. (2009). *Algoritma Data Mining. Buku Algoritma Data Mining*.

Muhammad Arief Rahman, Herman Budianto, & Esther Irawati Setiawan. (2019). Aspect Based Sentimen Analysis Opini Publik Pada Instagram dengan Convolutional Neural Network. *Journal of Intelligent System and Computation*, *1*(2), 50–57. https://doi.org/10.52985/insyst.v1i2.83

Mutia Annur, C. (2023). *Indonesia Peringkat Keempat Pengguna YouTube Terbanyak Dunia*. Databoks. https://databoks.katadata.co.id/-/statistik/08ebe16c8ac6904/indonesia-peringkat-keempat-pengguna-youtube-terbanyak-dunia

Nayla, A., Setianingsih, C., & Dirgantoro, B. (2023). Deteksi Hate Speech Pada Twitter. In *eProceeding of Engineering* (Vol. 10, Issue 1).

Pardamean, A., & Pardede, H. F. (2021). Tuned bidirectional encoder representations from transformers for fake news detection. *Indonesian Journal of Electrical Engineering and Computer Science*, *22*(3), 1667–1671. https://doi.org/10.11591/ijeecs.v22.i3.pp1667-1671

Pradany, L. N., & Fatichah, C. (2016). Analisa Sentimen Kebijakan Pemerintah Pada Konten Twitter Berbahasa Indonesia Menggunakan Svm Dan K-Medoid Clustering. *Scan : Jurnal Teknologi Informasi Dan Komunikasi*, *11*(1). https://doi.org/10.33005/scan.v11i1.635

*Ferian Fauzi Abdulloh

Putra, T. D., Utami, E., & Kurniawan, M. P. (2023). Analisis Sentimen Pemilu 2024 dengan Naive Bayes Berbasis Particle Swarm Optimization (PSO). In *Explore* (Vol. 13, Issue 1). https://doi.org/10.35200/ex.v11i2.13

Rahman Isnain, A., Indra Sakti, A., Alita, D., & Satya Marga, N. (2021). SENTIMEN ANALISIS PUBLIK TERHADAP KEBIJAKAN LOCKDOWN PEMERINTAH JAKARTA MENGGUNAKAN ALGORITMA SVM. *JDMSI*, *2*(1), 31–37.

Sholihah, N., Fauzi Abdulloh, F., & Rahardi, M. (2023). Optimasi Analisis Sentimen terhadap Kinerja Direktorat Jenderal Pajak Indonesia Melalui Teknik Oversampling dan Seleksi Fitur Particle Swarm Optimization. In *Smart Comp: Jurnalnya Orang Pintar Komputer* (Vol. 12, Issue 4). https://doi.org/10.30591/smartcomp.v12i4.5814

Thomas, S., Informasi, T., & Shanti Bhuana, I. (2021). Studi Analisis Metode Analisis Sentimen pada YouTube. *JIFOTECH (JOURNAL OF INFORMATION TECHNOLOGY)*, *1*(1).

Vidya Chandradev, I Made Agus Dwi Suarjaya, & I Putu Agung Bayupati. (2023). Analisis Sentimen Review Hotel Menggunakan Metode Deep Learning BERT. In *Jurnal Buana Informatika* (Vol. 14, Issue 02). https://doi.org/10.24002/jbi.v14i02.7244