Journal of Business, Social and Technology
Volume 7, Issue 3, 1063-1073
e_ISSN: 2807-6362
https://bustechno.polteksci.ac.id/index.php/jbt/index
DOI: doi.org/10.59261/jbt.v7i3.673

Public Opinion Sentiment Analysis of the #BoikotTrans7 Hashtag Case on YouTube Using the Naive Bayes Classifier Method

Mateus Jemboin
Institut Bisnis dan Teknologi Indonesia, Indonesia

***Desak Made Dwi Utami Putra**
Institut Bisnis dan Teknologi Indonesia, Indonesia

I Gede Made Yudi Antara
Institut Bisnis dan Teknologi Indonesia, Indonesia

I Gusti Agung Indrawan
Institut Bisnis dan Teknologi Indonesia, Indonesia

I Gede Iwan Sudipa
Institut Bisnis dan Teknologi Indonesia, Indonesia

***Corresponding author:**
Desak Made Dwi Utami Putra, Institut Bisnis dan Teknologi Indonesia, Indonesia.
✉desak.utami@stiki-indonesia.ac.id

Article Info:
Article history:
Received: May 23, 2026
Revised: June 15, 2026
Accepted: June 17, 2026

Keywords:
Sentiment Analysis, #BoikotTrans7, Naive Bayes, YouTube, SMOTE.

Abstract

Background: The emergence of the hashtag #BoikotTrans7 on social media was triggered by public reaction to the broadcast of the "Xpose Uncensored" program on Trans7, which was perceived as offensive to the dignity of Islamic boarding schools (*Pesantren*) and religious figures. This issue quickly developed into widespread public discourse on digital platforms, particularly in YouTube comments.
Objective: This study aims to analyze public sentiment on YouTube to identify patterns of opinion and public response regarding the #BoikotTrans7 controversy.
Methods: Data were collected using the YouTube Data API v3 through a data crawling process from October 15 to December 4, 2025, resulting in 10,490 comments. Sentiment analysis was conducted using the Naïve Bayes classifier with TF-IDF (Term Frequency–Inverse Document Frequency) feature extraction. To improve model performance, class imbalance was addressed using the Synthetic Minority Over-sampling Technique (SMOTE), and hyperparameter optimization was performed using GridSearchCV.
Results: The Naïve Bayes model achieved an accuracy of 78.46% in classifying sentiments into positive, negative, and neutral categories. The findings indicate that positive sentiment dominated YouTube comments, largely originating from users supporting Trans7 rather than boycott supporters. This suggests a discrepancy between viral hashtag narratives and actual public opinion on YouTube. Word cloud visualization highlights dominant keywords such as "*Pesantren*," "*Kyai*," "*Santri*," and "*maaf*," indicating that religious and cultural elements strongly shape the discourse surrounding the controversy.
Conclusion: The study provides insights for broadcasting evaluation and contributes to the development of sentiment analysis and text mining methodologies in social media research.

To cite this article: Jemboin, M., Putra, D. M. D. U., Antara, I. G. M. Y., Indrawan, I. G. A., & Sudipa, I. G. I. (2026). Public opinion sentiment analysis of the #BoikotTrans7 hashtag case on YouTube using the Naive Bayes classifier method. *Journal of Business, Social and Technology, 7*(3), 1063–1073. https://doi.org/10.59261/jbt.v7i3.673

INTRODUCTION

Media constitutes a fundamental channel for public information dissemination, encompassing both electronic and print forms (Facchinetti, 2021; Garnham, 2020). Electronic media delivers content through audio-visual means, while print media relies on written text and

static images (Aminu & Aliyu, 2024). Television, as a prominent form of electronic media, shapes public discourse through diverse programming, including news, talk shows, quiz programs, and films, which are produced and broadcast continuously (Putra et al., 2021). Trans7 is one of Indonesia's private television stations that used to be known as TV7.

This program began on March 22, 2000, and its existence was announced in the Supplement to State Gazette Number 8687 of 2001 which was issued on December 28, 2001 as PT Duta Visual Nusantara Tivi Tujuh. On August 4, 2006, the Kompas Gramedia group established a strategic partnership with the group (now known as CT Corp) and on December 15, 2006, TV7. One of the interesting shows from the Trans7 Tv station is "Xpose Uncensored", which is an infotainment program. This event is part of "Celebrities" or "Celebrities in the News".

With the infotainment format, "Xpose Uncensored" has a more relaxed delivery style. This can be seen from the way this show reports news about the world of celebrities. However, "Xpose Uncensored" not only addresses the lives of artists, like most other infotainment programs, but also addresses emerging national issues. Initially, the name of the program was "Celeb Expose", later changed to "Celebrity Expose". Finally, the current name is "Xpose Uncensored". One of the events that is currently the public discussion is the #BoikotTrans7 case. This event began with a broadcast aired by the national television channel Trans 7 on October 13, 2025, namely Xpose Uncensored. This reaction mainly emerged from the students and the *Pesantren* community, who felt disturbed by one of the episodes of the Xpose show aired by Trans7.

This episode features a title that is considered challenging: "The students just drink milk and have to squat, is life in the cottage really like this?". The video quickly spread on various platforms, such as TikTok, Instagram, X, and YouTube, and immediately received strong reactions from the public. Many people think that the content degrades the dignity of Kiai, students, and Islamic boarding schools in general. Although the program may have been intended as a form of social criticism, the narrative presented was considered unobjective. The show is seen as directing a negative view of the lives of students without providing an opportunity for the *Pesantren* to provide explanations, which can risk causing misperceptions among the public. Although many parties demanded and requested the revocation of Trans 7's broadcasting license, there were also some who supported it because they considered that the show represented their concerns about the dark side of life at the *Pesantren*. This shows that this case raises pro and con opinions between those who feel aggrieved and those who thank Trans 7 for the truth that has been revealed.

The author gathers public opinions through comments related to the issue of hashtags #BoikotTrans7 on one of the platforms that has a lot of impressions on the topic, namely YouTube. YouTube is a site that allows people to openly save, watch, and share videos. YouTube is the best place to share videos from all over the world, ranging from short videos, tutorials, vlogs, short films, movie trailers, music, education, animation, entertainment, news, TV, and various other interesting information. The increasing number of smartphone and internet users has also caused videos on YouTube to be more diverse (Zhou et al., 2016).

In the current digital era, public opinion expressed through social media has become a critical determinant for companies, media institutions, and policymakers in formulating responsive strategies (Grossman, 2022; Mohammadi & Jafari, 2025). Broadcasting companies must be sensitive to public sentiment, given that negative perceptions can directly affect their reputation, viewership ratings, and regulatory standing (Jonkman et al., 2020). The #BoikotTrans7 phenomenon demonstrates how a single broadcast event can rapidly escalate into a large-scale public controversy, making it imperative for media institutions to monitor and understand the distribution of public sentiment in real time. The ability to objectively analyze such digital discourse through computational methods provides an evidence-based foundation for corporate decision-making and content policy reform (Labafi et al., 2022).

Sentiment analysis is a branch of Natural Language Processing (NLP) that computationally identifies and categorizes opinions expressed in text, determining whether the expressed attitude toward a topic, product, or issue is positive, negative, or neutral (Liu, 2012). The process typically involves several sequential stages: (1) data collection from a target platform, (2) text preprocessing to clean and normalize raw data, (3) feature extraction to convert text into

numerical representations suitable for machine learning, (4) model training and classification, and (5) evaluation of model performance using metrics such as accuracy, precision, recall, and F1-score. In the context of social media analysis, sentiment analysis enables researchers to quantify large volumes of unstructured user-generated content and extract actionable insights from public discourse.

METHOD

Data Collection (Crawling Data)

Primary data in the form of comments was collected from the YouTube platform using the YouTube Data API v3 with the Python programming language through Google Colaboratory. Google Colaboratory, or Colab for short, is a research and development platform provided by Google for free. Colab provides a cloud-based working environment that allows users to write and execute Python code within a Jupyter Notebook environment without the need for additional configuration or installation (Hartati et al., 2024). Data collection was carried out from four YouTube videos related to #BoikotTrans7 case with the keywords: "Trans7 boycott case", "Trans7 boycott case", "Pros and cons of Trans7 boycott case", and #BoikotTrans7 hashtags. The data was collected in the period from October 15 to December 4, 2025, resulting in a total of 10,490 comments stored in the format of the CSV file to be processed. The following is a table of video sources that are the data collection of comments regarding the Trans 7 boycott case.

Table 1. Comment Source

No	Channel	Video Title
1.	CNN	"Trans7 admits negligence in its broadcast about Islamic boarding schools"
2.	Viva.co.id	TREMBLE! Trans7 President Director apologises in front of the House of Representatives and KPI
3.	Profit Story	*Pesantren* VS TRANS7 controversy - deliberately to insult *Kyai*?
4.	Guru Gembul	PROS AND CONS OF TRANS 7 BOYCOTT

Preprocessing data

According to Prasetija (2022), preprocessing is the first step in the classification of texts aimed at preparing written data before further processing. This activity includes various data modifications to improve the quality and uniformity of the text, such as removing special symbols, normalizing writing, and eliminating unrelated words. By carrying out preprocessing, the text is transformed into a neater and more structured format, so that the resulting information becomes more efficient and ready to be used in the next stage of analysis and modeling (Aufar et al., 2023). The preprocessing stage is carried out through the following six steps:

1. Cleaning: Removing irrelevant characters such as punctuation, emoticons, URL links, and custom symbols.
2. Case Folding: Converting the entire text to lowercase for uniformity.
3. Word Normalization: Changing non-standard words, abbreviations, and slang into standard words according to KBBI.
4. Tokenizing: Breaking sentences into tokens in the form of individual words.
5. Stopword Removal: Removing meaningless common words such as "and", "or", "which", and the like.
6. Stemming: Reducing all word forms to their root form using the Sastrawi library, a Python library designed for Indonesian language stemming.

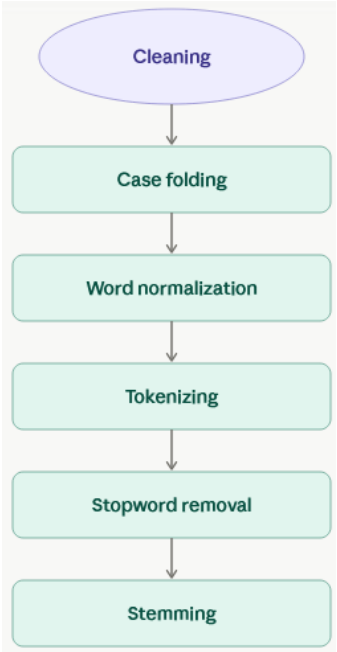

Figure 1. Data Processing Flow

Data Labeling

Data labeling was carried out automatically using the InSet Lexicon, an Indonesian sentiment dictionary containing a list of words and their polarity values. The labeling rules were defined as follows: if the polarity score was greater than 0, comments were labeled as Positive; if equal to 0, they were labeled as Neutral; and if less than 0, they were labeled as Negative. The auto-labeling results were then validated by linguists to ensure the accuracy and consistency of the labels, particularly for comments containing ambiguous meanings or context-specific expressions.

More specifically, the labeling criteria applied in this study were defined as follows. A comment was classified as Positive when its cumulative polarity score according to the InSet Lexicon was greater than zero, indicating that positive sentiment words outweighed negative ones in the comment (e.g., comments expressing support for Trans7, praise for the *Pesantren* community, or approval of the station's apology). A comment was classified as Negative when the polarity score was less than zero, meaning negative sentiment words were dominant (e.g., comments condemning Trans7, demanding license revocation, or expressing disapproval of the broadcast). A comment was classified as Neutral when the polarity score equaled zero, reflecting either a balanced mixture of positive and negative words or the absence of sentiment-bearing terms (e.g., factual statements, informational queries, or comments containing only neutral particles).

Model Implementation

The dataset was divided by an 80:20 ratio, namely 80% (7,894 data) for training and 20% (1,974 data) for testing using the stratified split method of the Python scikit-learn library, with random_state=42 to maintain reproducibility.

Model Evaluation

The evaluation of model performance was carried out using a confusion matrix measuring 3×3 with the following metrics: (1) Accuracy, measuring the percentage of correct predictions overall; (2) Precision, measuring the accuracy of positive predictions; (3) Recall, measuring the model's ability to detect actual classes; and (4) F1-Score, the harmonic average of precision and recall.

RESULTS AND DISCUSSION

Results

Application of Analytics

Sentiment analysis is the process of analyzing text to determine the sentiments expressed, such as positive, negative, or neutral. Sentiment analysis can be used for a variety of purposes, such as understanding public opinion, measuring customer satisfaction, and detecting fraud (Maulana et al., 2024). All the applications of analysis in the study are to find out the public opinion regarding the case of the boycott hashtag7 by using a machine learning method, namely the naïve bayes classifier which aims to divide the community or public opinion into three sentiments, namely positive, negative, and neutral. The results of the classification will be used to see public opinion in general regarding the trans 7 boycott case.

This research was conducted using the Python programming language and was run through the Google Colab Platform, and using CSV-formatted data files. The raw data collected from the YouTube platform comprised a total of 10,490 comments retrieved from 4 video sources. Prior to analysis, the raw dataset underwent a comprehensive text preprocessing pipeline to remove noise including punctuation, emoticons, URLs, and irrelevant symbols, resulting in a clean dataset of 10,490 records ready for classification. In this study, an 80:20 split ratio, with 80% of the data (7,894 records) used for training and 20% (1,974 records) for testing. Next, the data preprocessing process is carried out to clean the text of irrelevant elements before it is included in the classification process, the preprocessing stages of which will be described in the following description.

Data was collected through a crawling process from the Youtube Platform using the Youtube API tool. Data was collected using the keywords "Trans7 boycott case", "Trans7 boycott case", "Pros and cons of Trans 7 boycott case", during the period 13 October - 04 December 2025. The crawled data is stored in the form of a CSV file and input into the Google Colab workflow for initial processing. Furthermore, the data is processed using the Python programming language by going through text preprocessing stages such as cleansing, case folding, tokenizing, filtering, and stemming. After the text cleanup process, data labeling is carried out using a lexicon-based library approach to group sentiment. The output of this system resulted in 3 classifications being positive as many as 6698 (63.85%), negative as many as 2664 (25.39%), and neutral as many as 1128 (10.75%).

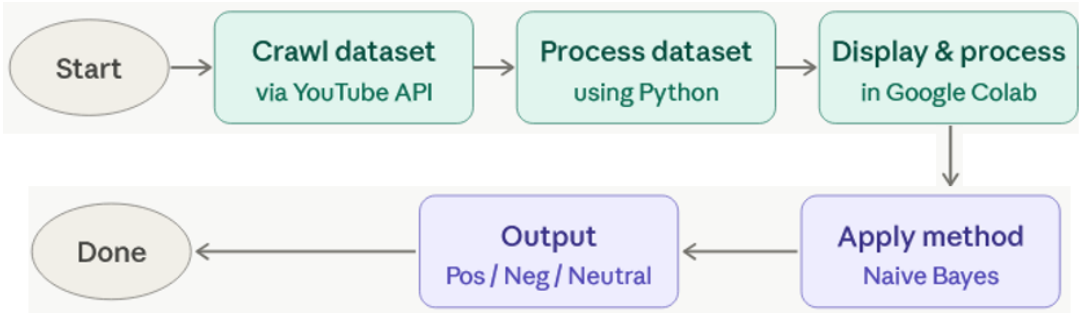

Figure 2. Data Management Process

The entire data analysis flow in Figure 1 began with the project preparation stage and the setup of the work environment using Google Colaboratory (Colab), a platform that supports data processing based on the Python programming language. The raw data, in the form of a CSV dataset that had been successfully collected from the YouTube platform, then underwent a series of preprocessing stages, which included cleaning, case folding, tokenizing, filtering, and stemming. After the data was declared clean, it proceeded to the analysis stage using the Naïve Bayes method, where the model was developed based on data that had gone through the cleaning process to classify each comment. In the final stage, a classification output was produced that grouped public opinion into three sentiment categories, namely positive (support), negative (opposition/criticism), and neutral (information that does not contain strong sentiment). The

entire process produced outputs that were used for reporting, data visualization, and the preparation of policy recommendations in this study.

Training Model Naïve Bayes

At this stage, a model training process was carried out to classify sentiment using the Naive Bayes algorithm, this algorithm was chosen because it is efficient and suitable for text analysis, especially in handling text classification with a feature in the form of word occurrence frequency (TF-IDF).

The selection of the Naive Bayes algorithm as the sole classification method in this study was based on its demonstrated suitability for large-scale text classification tasks. Naive Bayes is particularly effective when working with high-dimensional feature spaces, as is typically the case with TF-IDF-weighted textual data, and has been shown to achieve competitive performance in Indonesian-language sentiment analysis tasks (Hartati et al., 2024; Kevin et al., 2024). Furthermore, given the relatively large dataset (10,490 comments), computational efficiency was an important consideration. While studies comparing multiple methods such as Support Vector Machine (SVM) and Naive Bayes often demonstrate performance trade-offs Maulana (2024), the primary focus of this study is on the optimization pipeline involving SMOTE and GridSearchCV rather than algorithmic comparison, which represents a direction for future research.

Naive bayes classify is a classification method based on Bayes' theorem assuming that the features do not affect each other. This method is often utilized in text grouping due to its ease, speed, and adequate level of accuracy for large amounts of data (Bisono and Zulherry, 2025). To address the potential problem of class imbalance in sentiment data, the SMOTE (Synthetic Minority Over-sampling Technique) technique was implemented as part of the training pipeline.

The results of the study showed that the SMOTE (Synthetic Minority Oversam-pling Technique) technique was able to make the model more stable and have better performance for measurements in terms of AUC, balance, and MCC (Erlin et al., 2022). This process aims to improve the representation of minority classes and help better learning models from existing data. The evaluation stage of the Naive Bayes model was carried out using test data to measure the classification performance of each sentiment category, which is summarized in the Classification Report. After the parameter tuning process with GridSearchCV and the implementation of SMOTE, the model achieved an accuracy of 78.46% on the test data. The average values for all metrics (Macro Average) are Precision 0.76, Recall 0.70, and F1-score 0.72. Meanwhile, the Weighted Average shows a Precision of 0.88, a Recall of 0.78, and an F1-score of 0.78, indicating the model's fairly good performance in predicting sentiment. Here is a picture of the results of the accuracy of the naïve Bayes model using Smote.

```
Akurasi Model Naive Bayes (setelah tuning dengan SMOTE): 0.7846

Classification Report (setelah tuning dengan SMOTE):
              precision    recall  f1-score   support

          -1       0.62      0.75      0.68       533
           0       0.70      0.46      0.56       225
           1       0.88      0.85      0.87      1340

    accuracy                           0.78      2098
   macro avg       0.73      0.69      0.70      2098
weighted avg       0.79      0.78      0.78      2098
```

Figure 3. Naïve Bayes Accuracy Results with Smote

Confusion Matrix Visualization Results

The confusion matrix is a table that shows the amount of test data that is correctly classified and the amount of test data that is incorrect (Prasetija et al., 2022). The visualization below presents the confusion matrix obtained from the prediction results of the Naive Bayes

model. This model has gone through a hyperparameter tuning process using GridSearchCV and handling imbalance classes with the SMOTE technique, as well as using an 80:20 data division between the training data and the test data. Below is a screenshot of the script along with the results of the confusion matrix visualization which is the output of the model testing process with 80:20 data division. The Confusion Matrix is a table with 4 (four) different combinations of predicted values and actual values.

In performance measurement using the Confusion Matrix, there are 4 (four) terms as a representation of the results of the classification process. The four terms are True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN). Where the True Negative (TN) value is the amount of negative data that is correctly detected, while False Positive (FP) is negative data but is detected as positive data. Meanwhile, True Positive (TP) is positive data that is detected to be true. False Negative (FN) is the opposite of True Positive, so the data is positive, but it is detected as negative data (Endah & Encis, 2021). The following is the result of the 80:20 confusion matrix.

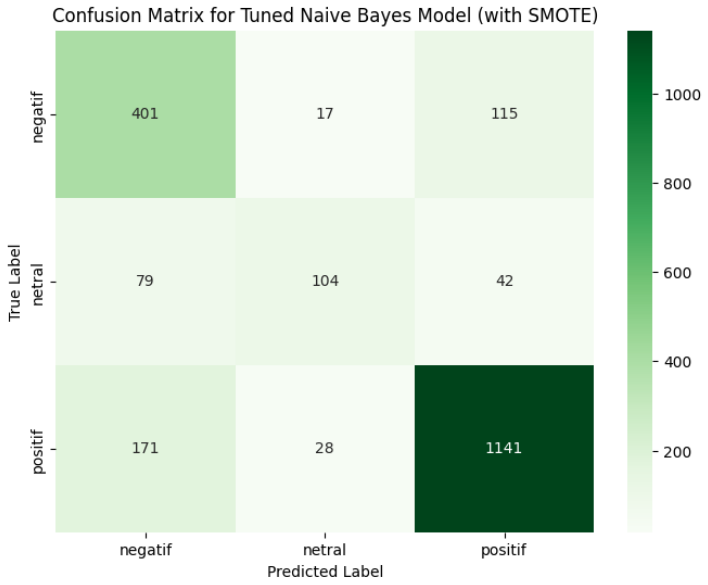

Figure 4. Visualization of the Confusion Matrix

The Confusion matrix image above shows the performance of the Naive Bayes model in classifying sentiment data into three categories: positive, negative, and neutral. The image is divided into three rows and three columns: 1) The line represents the actual (actual) label. 2) The columns represent the predicted results of the model.

Each cell in the table contains the amount of data that falls into that category. The higher the number on the row = column (diagonal part), the better the model's accuracy to that category.

Discussion

The dominance of positive sentiment (63.85%) in the dataset carries a significant interpretive implication that directly contradicts the expected narrative of the #BoikotTrans7 movement. Rather than reflecting anti-Trans7 sentiment, the majority of YouTube comments express support for the television station or seek to contextualize the controversy. This outcome may be explained by several factors: (1) YouTube's algorithmic tendency to surface content from established media channels that may attract audiences already sympathetic to Trans7's perspective; (2) the presence of counter-narrative voices from community members who felt the boycott movement was an overreaction; and (3) the nature of the selected video sources (CNN, Viva.co.id, Profit Story, Guru Gembul), which may have attracted audiences with diverse or moderating viewpoints. This finding underscores the risk of interpreting social media trends at face value, as hashtag virality does not necessarily correspond to the dominant direction of public

sentiment on a given platform. Compared to studies on similar boycott phenomena on Twitter/X and Instagram, which tend to show stronger negative sentiment alignment with boycott movements Maulana (2024), YouTube's comment dynamics appear more nuanced, reflecting platform-specific audience behaviors.

Visualizing Labeling Using InSet Lexicon

Lexicon-based is a part of unsupervised machine learning (Artana et al., 2023). Lexicon based is a sentiment analysis method that works by matching words in text with a sentiment lexicon, which is a list of words that have been labeled/scored positively, negatively, or neutrally before. The following is a visualization of the distribution of data labeling using the lexicon-based method with a total of 10,490 data: 1) Positive dominated with 6,698 data (63.85%) 2) Negative as many as 2,664 data (25.39%). 3) Neutral as many as 1,128 data (10.75%)

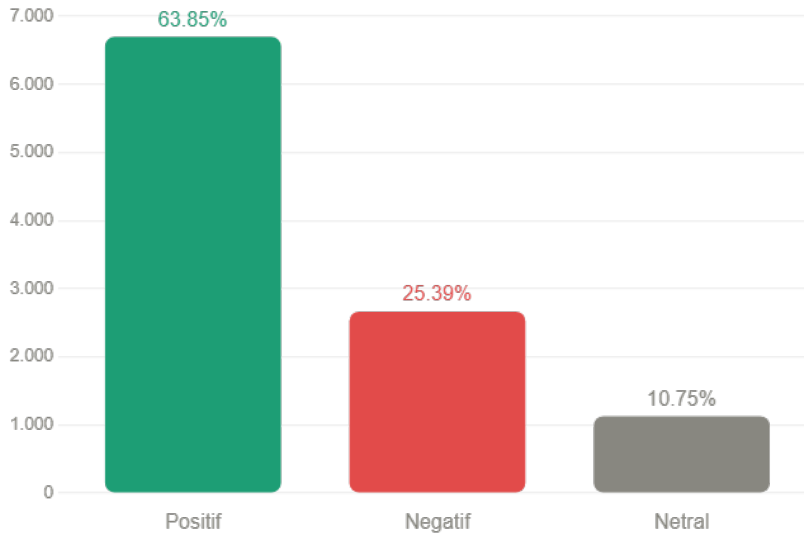

Figure 5. Data labeling visualization

Wordcloud

Usage Word Cloud In sentiment analysis, it has proven to be an effective method to uncover public perception. Word Cloud is a form of representation Visual which shows the frequency of occurrence of words in a text, thus helping researchers in quickly identifying main themes and topics from large text data. In the context of sentiment analysis, Word Cloud play a role in understanding the distribution and level of occurrence of words that have positive or negative emotional connotations in public conversation (Agusia et al., 2024).

The visualization of Word Cloud in this study is an effective tool for recognizing key keywords in YouTube comments, based on sentiment categories (positive, negative, and neutral). Word cloud is a form of visualization that displays words from the provided text. In this representation, the font size is given proportional to the frequency of occurrence of the word, where the words that appear more frequently will have a larger font size. These visualizations have a usefulness for developers to gain insight into the positive and negative aspects of the text (Kevin et al., 2024).

By using specially filtered words (including informal terms and common particles in Indonesian), the goal of visualizing the word cloud is to eliminate words that are less meaningful. This allows Word Cloud to better highlight terms that shape and reflect the topic and analyze those sentiments. Through Word Cloud, it is possible to visually understand the characteristic patterns of each sentiment; for example, words that are large in Word Cloud 'positive' indicate positive expressions that appear frequently, as well as for 'negative' and 'neutral' sentiments, which provide a clear picture of the characteristics of the text that distinguish each type of sentiment. The following are the results of the visualization of word cloud in this study:

The Word Cloud visualizations in this study reveal distinct thematic patterns across the

three sentiment categories. In the positive Word Cloud, dominant terms such as "*Pesantren*", "*Kyai*", "*Santri*", and "*maaf*" (apology) indicate that supportive comments frequently center on religious institutions and forgiveness narratives, suggesting that many positive commenters were either defending Islamic boarding school communities or accepting Trans7's public apology. The prominence of the word "*maaf*" is particularly noteworthy, as it signals that a substantial portion of the public acknowledged the broadcaster's remorse and chose to de-escalate the controversy. In the negative Word Cloud, terms associated with protest, institutional criticism, and calls for accountability dominate, reflecting the concerns of the *Pesantren* community and its supporters who felt their dignity had been compromised. In the neutral Word Cloud, informational keywords dominate, suggesting that a segment of commenters engaged with the topic in a factual or observational manner without expressing strong emotional polarity. Overall, the Word Cloud analysis confirms that the #BoikotTrans7 controversy was fundamentally rooted in religious and cultural identity, with public discourse revolving around the perceived treatment of Islamic educational institutions by mainstream media.

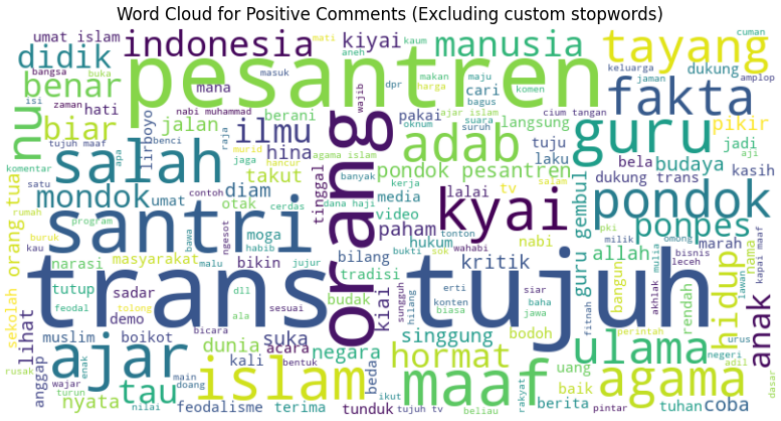

Figure 6. Wordcloud Positive

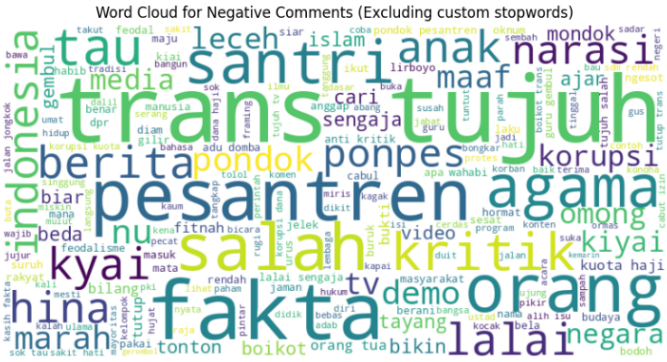

Figure 7. Wordcloud Negative

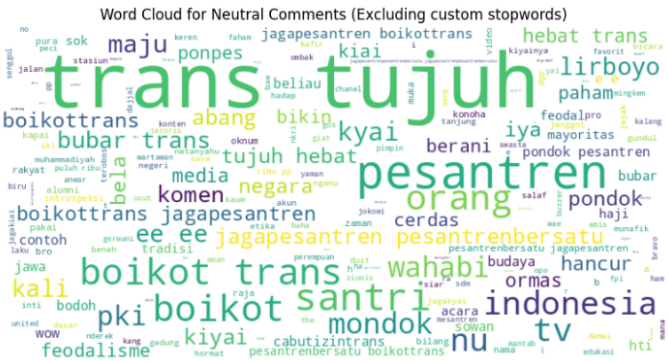

Figure 8. Wordcloud Neutral

CONCLUSION

This study aims to analyze sentiment towards the phenomenon of #BoikotTrans7 hashtags on the YouTube platform, which was triggered by the controversy of the "Xpose Uncensored" program. Through data collection using YouTube Data API v3, 10,490 comments were obtained which then went through a series of strict text preprocessing stages, including cleansing, case folding, normalization, tokenizing, Stopword removal, and stemming to ensure optimal data quality. The implementation of the Synthetic Minority Over-sampling Technique (SMOTE) proved to be a crucial component in overcoming the problem of class imbalance in the dataset, while the use of GridSearchCV for hyperparameter tuning succeeded in improving the performance of the model to reach an accuracy level of 78.46% with a precision value of 0.88%.

The findings of the study show that the positive sentiment that dominates the dataset comes from community groups that support Trans7, not the boycott movement itself. This indicates that although the #BoikotTrans7 hashtag was widely discussed, most of the public comments on YouTube are in favor of the television station. The findings through Word Cloud visualization confirm that public discourse is also colored by keywords such as "*Pesantren*", "*Kyai*", "*Santri*", and "facts", which reflect the sensitivity of religious and cultural issues in this controversy. Overall, this research not only provides an academic contribution to the application of optimized text mining methods, but also provides practical insight for the broadcasting industry that the public response on social media is not always in line with the narrative of the growing boycott movement. Therefore, data-based sentiment analysis is an important instrument in understanding public opinion objectively in the digital era.

ACKNOWLEDGEMENT

The author would like to express his deepest gratitude to the Informatics Engineering Study Program, Indonesian Institute of Business and Technology for the support of academic facilities provided during the research process. The author also expressed his sincere gratitude to Mrs. Desak Made Dwi Utami Putra as Supervisor I who has spent her time, energy, and thoughts in providing very meaningful direction, guidance, and input so that this research can be completed properly. Hopefully all forms of support and guidance that have been given will be given in a proper reply.

AUTHOR CONTRIBUTION STATEMENT

All authors contributed collaboratively to this research. Mateus Jemboin was responsible for data collection, preprocessing, model implementation, and drafting the original manuscript. Desak Made Dwi Utami Putra contributed to research conceptualization, methodology development, supervision, and critical manuscript revision. I Gede Made Yudi Antara assisted in data validation, analysis interpretation, and result evaluation. I Gusti Agung Indrawan contributed to literature review support and methodological refinement, while I Gede Iwan Sudipa participated in overall supervision and final manuscript review. All authors have read, revised, and approved the final version of the manuscript and agree to be accountable for all aspects of the work.

REFERENCES

Agusia, P., Manurung, M. U. A., Calista, V., & Mawardi, V. C. (2024). Pemanfaatan Word Cloud Pada Analisis Sentimen Dalam Menggali Persepsi Publik. *Seminar Nasional CORISINDO*, 25–30.

Aminu, S., & Aliyu, A. M. (2024). Impact of Printed and Audio-Visual Media on Students' Academic Performance in Secondary Schools in Katsina State. *UMYU Journal of Educational Research*, *12*(2), 73–82.

Artana, I. K., Pradnyana, A., & Darmawiguna, I. G. (2023). Analisis Sentimen Twitter untuk Menilai Kesiapan Pembelajaran Tatap Muka Terbatas dengan Inset Lexicon dan Levenshtein Distance. *Jurnal Pendidikan Teknologi Dan Kejuruan*, *20*(2), 200–209.

Aufar, A. F., Mochamad Alfan Rosid, Eviyanti, A., & Astutik, I. R. I. (2023). Optimizing Text Preprocessing for Accurate Sentiment Analysis on E-Wallet Reviews. *JICTE (Journal of*

Information and Computer Technology Education), *7*(2), 42–50. https://doi.org/10.21070/jicte.v7i2.1650

Bisono, A. T., & Zulherry, A. (2025). Analisis Sentimen Game Genshin Impact untuk Mengetahui Reaksi dan Harapan Pemain Menggunakan Metode Naïve Bayes. *Sudo Jurnal Teknik Informatika*, *4*(2), 183–193. https://doi.org/10.56211/sudo.v4i2.1131

Endah Fauziningrum, M. P., & M. Pd Encis Indah Suryaningsih, S. T. (2021). Evaluasi Dan Prediksi Penguasaan Bahasa Inggris Maritim Menggunakan Metode Decision Tree Dan Confusion Matrix (Studi Kasus Di Universitas Maritim Amni). *Angewandte Chemie International Edition, 6(11), 951–952.*, 5–24.

Erlin, E., Desnelita, Y., Nasution, N., Suryati, L., & Zoromi, F. (2022). Dampak SMOTE terhadap Kinerja Random Forest Classifier berdasarkan Data Tidak seimbang. *MATRIK : Jurnal Manajemen, Teknik Informatika Dan Rekayasa Komputer*, *21*(3), 677–690. https://doi.org/10.30812/matrik.v21i3.1726

Facchinetti, R. (2021). News discourse and the dissemination of knowledge and perspective: From print and monomodal to digital and multisemiotic. *Journal of Pragmatics*, *175*, 195–206.

Garnham, N. (2020). The media and the public sphere. In *The information society reader* (pp. 357–365). Routledge.

Grossman, E. (2022). Media and policy making in the digital age. *Annual Review of Political Science*, *25*, 443–461.

Hartati, T., Sohadi, R. T., Tohidi, E., & Wahyudin, E. (2024). Jurnal Informatika dan Rekayasa Perangkat Lunak Penerapan Algoritma Naive Bayes pada Analisis Sentimen Ulasan Aplikasi Whoosh-Kereta Cepat Di Google Play Store. *Jurnal Informatika Dan Rekayasa Perangkat Lunak*, *6*(1), 244–249.

Jonkman, J. G. F., Boukes, M., Vliegenthart, R., & Verhoeven, P. (2020). Buffering negative news: Individual-level effects of company visibility, tone, and pre-existing attitudes on corporate reputation. *Mass Communication and Society*, *23*(2), 272–296. https://doi.org/10.1080/15205436.2019.1694155

Kevin, K., Enjeli, M., & Wijaya, A. (2024). Analisis Sentimen Pengunaaan Aplikasi Kinemaster Menggunakan Metode Naive Bayes. *Jurnal Ilmiah Computer Science*, *2*(2), 89–98. https://doi.org/10.58602/jics.v2i2.24

Labafi, S., Ebrahimzadeh, S., Kavousi, M. M., Abdolhossein Maregani, H., & Sepasgozar, S. (2022). Using an Evidence-Based Approach for Policy-Making Based on Big Data Analysis and Applying Detection Techniques on Twitter. *Big Data and Cognitive Computing*, *6*(4), 160.

Liu, B. (2012). *Sentiment analysis and opinion mining*. https://doi.org/10.2200/S00416ED1V01Y201204HLT016.

Maulana, B. A., Fahmi, M. J., Imran, A. M., & Hidayati, N. (2024). Analisis Sentimen Terhadap Aplikasi Pluang Menggunakan Algoritma Naive Bayes dan Support Vector Machine (SVM). *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, *4*(2), 375–384. https://doi.org/10.57152/malcom.v4i2.1206

Mohammadi, R., & Jafari, J. S. (2025). The Role of Social Media in Shaping Regulatory Policies for Digital Businesses: Business and Policy Perspectives. *Future of Work and Digital Management Journal*, *3*(1), 1–10.

Prasetija, Z. R. N. S., Romadhony, A., & Setiawan, E. B. (2022). Analisis Pengaruh Normalisasi Teks pada Klasifikasi Sentimen Ulasan Produk Kecantikan. *E-Proceeding of Engineering*, *9*(3), 1769–1775.

Putra, D., Ilhaq, M., Desain, P., Visual, K., Jl, P. P., No, B. R., Jl, P., Royong, G., & Palembang, U. (2021). *Pemahaman dasar film dokumenter televisi*. *6*(2), 86–91.

Zhou, R., Khemmarat, S., Gao, L., Wan, J., & Zhang, J. (2016). How YouTube videos are discovered and its impact on video views. *Multimedia Tools and Applications*, *75*(10), 6035–6058.