Utilizing Sentiment Analysis for Reflect and Improve Education in Indonesia

Henderi[1,*], Asro[2], Agus Sulaiman[3], Tri Basuki Kurniawan[4], Deshinta Arrova Dewi[5], Mashal AlQudah[6]

[1,2,3]*University of Raharja, Indonesia*

[4] *Faculty of Science and Technology, Universitas Bina Darma, Indonesia*

[5] *Faculty of Data Science and Information Technology, INTI International University, Malaysia*

[6]*Institute of Public Administration, Digital Transformation & Information Programs Department, Riyadh, Saudi Arabia*

(Received: August 13, 2024; Revised: October 6, 2024; Accepted: November 20, 2024; Available online: December 28, 2024)

Abstract

This study explores the potential of sentiment analysis in providing valuable insights into education in Indonesia based on comments from the YouTube platform. Utilizing the Naive Bayes Classifier method, this research analyzed 13,386 processed comments out of 17,920 original comments. The results show that 53.8% of comments were negative, while 28.5% were positive, and 17.7% were neutral, reflecting diverse perspectives on existing educational issues. The Accuracy of this model reached up to 72.51% with testing on various sample sizes (10%-30%), indicating the model's effectiveness in identifying sentiments. Although the model tends to classify comments as unfavorable, this opens opportunities for introspection and improvement within the educational system. Further analysis with a word cloud revealed dominant keywords, indicating areas that require more attention in public discussions about education. By leveraging this sentiment analysis, the study offers practical and valuable guidance for policymakers to reflect on and enhance educational strategies and policies in Indonesia. This research measures public reactions and aims to foster more constructive and inclusive discussions about the sustainable development of education in Indonesia.

Keywords: Component, Sentiment Analysis for Education, Naive Bayes Classifier, Youtube Comments, Education Quality

1. Introduction

Over the past decade, social media has become an essential public platform where individuals from diverse backgrounds share their views and feelings on various social topics, including education. As a fundamental element of national progress, education often becomes a focal point in public discussions in Indonesia. Additionally, there is a noticeable trend toward integrating phonics learning with the development of reading comprehension skills, which has been shown to positively influence educational success in later stages. As a key pillar of nation-building, education remains a critical topic in Indonesia's public discourse [1]. In Indonesia, internet usage has reached significant levels, with over 63 million people actively participating online [2]. Sentiment analysis, a rapidly growing branch of data science, offers tools to interpret large-scale data from social media to uncover public perceptions and reactions. In this context, as a popular social media platform, YouTube often records and presents various viewpoints on Indonesia's education system through comments on uploaded videos.

The approach using the Naive Bayes Classifier in sentiment analysis has proven effective in examining and interpreting extensive text data, providing a deeper understanding of public sentiment. Previous studies have shown that sentiment analysis can provide insights for policy-making and strategic adjustments in the public and private sectors. In the public sector, strategic planning is highly beneficial as a starting point for nonprofit or public organizations based on common sense, which sometimes dominates every decision-making process, such as education [3]. In this study, we adopt this method to analyze YouTube comments, which serve as a digital representation of public opinion, to reflect on and

***Corresponding author: Henderi (henderi@raharja.info)**
DOI: https://doi.org/10.47738/jads.v6i1.527

improve education policies in Indonesia. The results of this research note the dominance of negative comments towards education in Indonesia, indicating a broad scope for improvement in the education system. The other study [4] also shows that negative sentiment in online comments often reflects an urgent need for change and adaptation to issues considered necessary by the public. Through in-depth analysis, this study aims to explore specific areas requiring attention while proposing actionable solutions to address these issues. Furthermore, this research seeks to uncover positive and neutral sentiments as sources of ideas and initiatives that can be implemented to enrich educational practices in Indonesia.

Thus, through this analytical approach, this study provides valuable contributions to the discussion and development of education policies that are more responsive to the needs and expectations of the Indonesian people. Sentiment analysis not only measures public reactions but also paves the way for a more productive and inclusive dialogue about the future of education in Indonesia.

2. Related Works

Sentiment analysis has become an essential tool across various domains, including politics, education, public services, and traffic management, providing valuable insights into public opinion and aiding in improving strategies and policies. According to multiple studies, this analytical method can significantly contribute to reflecting and enhancing the educational landscape in Indonesia. This approach enables students to regulate their cognitive abilities or non-cognitive skills and identify their weaknesses [5], allowing for enhancements in future activities. One approach to helping students stay in their special education is a stronger foundation of content knowledge, academic skills, and non-cognitive skills [6].

Sentiment analysis is vital in understanding opinions in education and A.I. ethics. Research has emphasized that improving the quality of educators is essential for better educational outcomes [7]. It underscores the need for continuous professional development, aligning with sentiment analysis to identify areas needing improvement in Indonesian education. In the broader context, sentiment analysis also gauges public opinion on A.I. ethics. The study employs Naive Bayes Classifier and TF-IDF to analyze 1,138 social media posts, categorizing them into positive, neutral, and negative sentiments. The model achieves 71% accuracy with average Precision, Recall, and F1-Score values of 66%, 71%, and 64%, respectively, highlighting the method's effectiveness in capturing public sentiment on

A.I. ethics [7].

They further highlighted the utility of sentiment analysis in education by applying the Naive Bayes Classifier to predict student graduation outcomes, achieving a high accuracy rate of 97.63% [8]. According to their study, predictive modeling in education can serve as a valuable tool for assessing and enhancing various aspects of the educational system in Indonesia. The other report has conducted a comparative study of Naive Bayes, Support Vector Machine (SVM), and Logistic Regression models in sentiment analysis related to Indonesian immigration [9]. According to their research, Logistic Regression outperformed Naive Bayes and SVM, with an accuracy of 77%. While focused on immigration, their comparative insights into the effectiveness of different models are valuable for the current study, as they highlight the importance of selecting the most appropriate model for sentiment analysis in educational contexts.

The other researchers [10] have developed a hybrid machine-learning model to analyze sentiment and assess satisfaction with Turkish universities using Twitter data. Their approach, which combines conventional machine learning, deep learning, and BERT-based transformers, effectively handles the linguistic complexities of Turkish text. The model, BERT-BiLSTM-CNN, demonstrated superior Accuracy in sentiment analysis, making it a valuable tool for evaluating public satisfaction with educational initiatives.

The insights drawn from these studies are directly relevant to the current research on sentiment analysis to reflect and improve education in Indonesia. The application of Naive Bayes, SVM, and Logistic Regression across various contexts—ranging from A.I. ethics to political sentiment and educational outcomes—demonstrates the versatility and effectiveness of these models in extracting actionable insights from large datasets. By leveraging the methodologies and findings from these studies, the current research can identify critical areas within the Indonesian education system that require improvement. According to a study comparing COVID-19 patient recovery rates in Indonesia using Naive

ISSN 2723-6471

Bayes and PSO-enhanced Naive Bayes algorithms, it has reported an accuracy of 94.07% for Naive Bayes and 95.56% for PSO-based Naive Bayes [11].

The research demonstrated the utility of Naive Bayes in predicting student outcomes, highlighting its potential for analyzing public sentiment toward educational policies [8]. The other researcher has conducted a comprehensive systematic literature review on sentiment analysis, exploring various methods, applications, and the challenges faced in the field [12]. It addresses the increasing importance of sentiment analysis as a tool for automatically identifying opinions and emotions from vast online comments. The study highlights the significant role of artificial intelligence technologies in enhancing sentiment analysis and discusses the challenges and limitations researchers encounter. Additionally, it offers valuable insights and guidance for scholars and practitioners in selecting appropriate methodologies and best practices for conducting sentiment analysis.

The other report [13] has provided a comprehensive survey on integrating sentiment analysis with graph neural networks (G.N.N.s) for stock prediction. Their review highlights the growing interest in using G.N.N.s to enhance stock prediction accuracy by incorporating sentiment analysis from various sources like news, social media, and financial reports. The study emphasizes the potential benefits of this integration, discusses challenges in data collection and preprocessing, and outlines the limitations of current approaches, offering valuable insights for future research in this interdisciplinary field.

A comprehensive review of the application of sentiment analysis techniques to assess public sentiment towards sustainability initiatives was conducted [14]. Their research utilized a variety of machine learning and deep learning models to analyze tweets related to environmental sustainability, comparing pre-trained models like V.A.D.E.R. and

B.E.R.T. with traditional methods such as Logistic Regression and SVM. The study highlights the variability in model performance, emphasizing the need for carefully selected tools tailored to the specific context of sustainability. This work underscores the potential of sentiment analysis to inform policies and actions to foster environmental responsibility.

Investigating the relationship between sentiment analysis and token returns during Initial Coin Offerings (I.C.O.s) that involve 391 tokens, it was found that traditional I.C.O. ratings often fail to predict actual returns. That sentiment analysis of whitepapers offers limited insights [15]. However, introducing a new I.C.O. index and analyzing sentiment from tweets significantly improved the understanding of factors affecting six-month token returns. The research highlights the potential of machine learning models as a more transparent and effective alternative to conventional I.C.O. ratings, offering new perspectives on capital raising in the blockchain industry.

The research developed a framework combining time-series analysis with text mining to analyze direct-message conversations on social media platforms. This innovative approach allows investigators to efficiently identify significant changes in sentiment within digital communication trails, such as SMS and WhatsApp, thereby reducing the time and resources needed for in-depth analysis in digital forensics. The framework enhances the ability to pinpoint critical moments in conversations, making it a valuable tool in the field [16]. The research about a comprehensive review of the advancements and challenges in NLP-based sentiment analysis was conducted by Jim et al. [17]. The study explores various application domains, preprocessing techniques, datasets, and evaluation metrics, offering insights into the use of machine learning, deep learning, and pre-trained models for sentiment analysis. It also examines recent experimental results, highlights limitations, and discusses the challenges faced in the field. The review concludes by proposing future research directions to address these challenges, providing a thorough understanding of sentiment analysis and its evolving landscape.

The quality of Bengali YouTube content on floating agriculture and viewers' interactions was analyzed by Chakma et al. [18]. The study assessed 57 videos, finding that only 26.32% were high quality. Despite the limited number of high-quality videos, viewer engagement was generally positive, with many views, likes, and comments. The research highlights the current state of YouTube content on floating agriculture, providing insights for content creators to improve the quality of educational videos in this domain. Else research [19] examines the rise of anti-Roma racism on YouTube during the 2016 U.K. Referendum, focusing on how this racism, particularly "entitlement racism," is expressed in viewer comments. The study highlights the normalization of such rhetoric and calls for improved content moderation to address this growing issue.

Meanwhile, the study is about improving sentiment analysis in the Bengali language by using transformer-based models [20]. In the research, the transformer-ensemble model achieves 95.97% accuracy, outperforming recent approaches, and addresses the challenges of limited resources in Bengali text processing. Moreover, similar research examined the role of sentiment analysis in education, highlighting its use in improving pedagogy through artificial intelligence techniques [21]. The review addresses various sentiment analysis levels, annotation methods, and challenges, offering insights into future research in educational sentiment analysis.

3. Research Method

The methodology used to survey recent work on sentiment analysis and opinion mining for public security is described [22]. This study's methodology ensures that the research is conceptual and goal-oriented [23]. It involves a structured sequence of eight steps, ranging from data acquisition to applying the Naive Bayes method for analysis. This study employs sentiment analysis to explore public opinions on Indonesian education by examining comments on the YouTube platform.

3.1. Research Stages

3.1.1. Problems Formation and Data Collection

This stage is the first and crucial step in the research. Conducting preliminary observations and a comprehensive literature review identifies problems related to education in Indonesia. This literature review analyzes journals, articles, and other relevant sources to identify existing research gaps and formulate appropriate research questions. This step is essential for determining the research focus and providing a clear direction for the subsequent analysis process. After that, the data collection was conducted. In this stage, the data used in this research was collected from comments posted on YouTube videos related to education in Indonesia. Utilizing the YouTube Data API, the data collection process was automated and efficient, resulting in 17,920 comments. These comments represent public opinion and serve as this study's primary data source for sentiment analysis.

3.1.2. Preprocessing and Data Labeling

Data preprocessing is a crucial phase in this study as it ensures that the data used for subsequent analysis is clean and ready for examination [24]. This stage involves several essential procedures. Initially, raw data, which often contains unwanted elements such as URLs, HTML codes, and non-alphanumeric characters, is thoroughly cleaned to remove disruptive components like hashtags, usernames, and duplicate data. After cleaning, all text is converted to lowercase through a process known as case folding to ensure consistency and prevent unnecessary word duplication. Next, the data undergoes tokenization, breaking sentences into individual words while eliminating irrelevant symbols and numbers. According to the Indonesian dictionary, this is followed by word normalization, where non-standard words or abbreviations are adjusted to their standard forms. Subsequently, stop word removal is executed, eliminating words that do not add significant meaning, such as conjunctions and ordinary words that do not substantially affect the sentence context. The stemming process is also implemented to reduce words with affixes to their root forms, using a dictionary tailored to the dataset findings and Indonesian language standards. After completing these steps, the resulting dataset contains 13,386 clean comments ready for further analysis. This thorough preprocessing is vital for ensuring the data quality used in modeling and lays a solid foundation for conducting sentiment analysis. In data labeling, the cleaned comments are classified into three sentiment categories: positive, negative, and neutral. Trained human annotators perform this classification to ensure that each comment is consistently and accurately assessed according to sentiment. Accurate labeling is critical to training an effective predictive model.

3.1.3. Modeling with Naïve Bayes and Validation

After labeling, the Naive Bayes model is used for sentiment classification. This model was chosen for its proven ability to handle large volumes of text data and classify quickly and efficiently. Based on the training from the labeled dataset, the model is expected to predict the sentiment category of new comments. After the model is trained, validation is performed to assess its effectiveness using evaluation metrics such as Accuracy, Precision, and Recall. This evaluation is usually performed on a subset of data not used during the training phase, such as 10%, 20%, and 30%, to test the model's performance under varied conditions and ensure that the model is reliable and not overfitting.

3.1.4. Analysis, Interpretation, and Data Visualization

The results of the model classification are then analyzed to gain a deeper understanding of public sentiment towards educational issues in Indonesia. This analysis includes evaluating the distribution of sentiments and identifying themes or topics that frequently appear in the comments. Finally, the analysis results are presented visually through word clouds and pie charts, which not only facilitate the understanding of trends and patterns in the data but also present the research findings in an attractive and accessible way for policymakers, other researchers, and the public.

3.2. Naïve Bayes Algorithms

Naive Bayes is a Bayesian graphical model that assigns nodes to each feature or characteristic. It is termed "naive" because it disregards the prior distribution of the parameters and assumes the independence of all features or rows [25]. This approach is widely recognized for its simplicity and effectiveness in text classification tasks. The algorithm calculates the probability for each class based on the given features and selects the class with the highest probability as the predicted output. The basic formula for Naive Bayes is generally described as Eq.1.

$$P(C|x) = PcxP(c)P(x) \tag{1}$$

In Equation (1), Px| is the posterior probability of class C given the predictor x; Px|C is the likelihood, which is the probability of predictor x given class C; P.C. is the prior probability of class C, and Px is the prior probability of predictor x.

3.3. Implementation of Naïve Bayes

In the Python implementation, the Scikit-learn library provides the MultinomialNB module, which is suitable for text classification with word frequency features. Naive Bayes has been proven to have high Accuracy and speed when used in large databases [11], [26]. The implementation process includes text vectorization, model training, and sentiment prediction. Text vectorization converts text into numerical format using TfidfVectorizer, which calculates word frequencies and assigns weights to words based on how informative they are across documents. Model training is the Naive Bayes model trained using label data, and sentiment prediction is the trained model used to predict sentiment on test data.

3.4. Model Evaluation

After the implementation process, the model is evaluated using a confusion matrix and performance metrics such as Accuracy, Precision, Recall, and F1-Score. Confusion Matrix is a table that shows the frequency of predicted and actual class labels. It helps identify the types of errors made by the model. Accuracy is the ratio of correct predictions to the total number of predictions. It shows the closeness of the classification results to the actual value [12]. Accuracy is obtained by comparing the correctly classified data with the overall data. It also obtained Precision, Recall, and F1-Ccore [13], [27]. Precision is the ratio of accurate optimistic predictions to the total predictions that are declared positive, Recall is the ratio of accurate optimistic predictions to the total actual positive cases in the data, and F1-Score is the harmonic mean of Precision and Recall, providing a balance between the two metrics. The formulas for the metrics of Accuracy are in (2), Precision is in (3), Recall is in (4), and F1 Score in (5).

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+TN+FN} \tag{2}$$

$$\text{Precision} = \frac{TP}{TP+FP} \tag{3}$$

$$\text{Recall} = \frac{TP}{TP+FN} \tag{4}$$

$$\text{F1 Score} = 2x\left(\frac{\text{Precision x Recall}}{\text{Precison+Recall}}\right) \tag{5}$$

ISSN 2723-6471

In Equations (2), (3), (4), and (5), True Positive (T.P.) are cases where positive data is correctly classified as positive by the model, True Negative (T.N.) in cases where harmful data is correctly classified as unfavorable by the model, False Positive (T.P.) cases where harmful data is incorrectly classified as positive by the model, and False Negative (F.N.) in cases where positive data is incorrectly classified as unfavorable by the model [14]. In the context of sentiment analysis, there are three categories (positive, negative, and neutral). We can add TNu (True Neutral) in cases where the model correctly classifies neutral data as neutral [26].

3.5. Analysis

The analysis results are presented visually to facilitate understanding and data interpretation. They are word clouds, pie charts, bar charts, and square visualization of confusion matrices. The Word Clouds display frequently occurring words in comments, making it easier to identify significant issues in discussions about education; the Pie Chart presents the proportion of sentiments (positive, negative, and neutral) in the dataset, providing a quick overview of the general sentiment distribution; Bar Chart is used to show the number of comments per month from the crawling results, depicting trends in discussions over time and responses to specific events. Square Visualization of Confusion Matrix visualizes the model's classification performance, displaying true positives, true negatives, false positives, and false negatives, making it easier to evaluate the model's Accuracy [28], [29].

4. Results and Discussion

Data Collection This study used the Naive Bayes Classifier method to analyze the sentiment of 13,386 processed comments from 17,920 original comments received via the YouTube platform. The results explore public perceptions of education in Indonesia, emphasizing the impact of sentiment on educational discussions and policies. Figure 1 below is a visual representation of the data collected from January to July before processing.

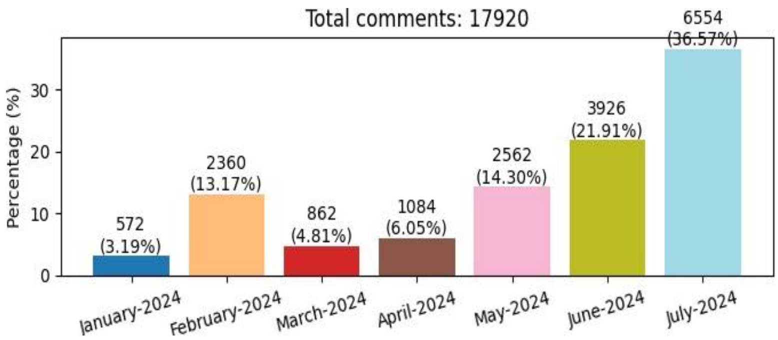

Figure 1. Crawling results Sentiment distribution

The analysis of the sentiment distribution from the comments shows that the majority of comments have a negative sentiment (53.8%), followed by positive comments (28.5%) and neutral comments (17.7%). It indicates significant dissatisfaction or areas of concern among the public regarding the current education system. Figure 2 show pie charts depict the proportions of positive, negative, and neutral comments based on the sentiment analysis. These charts help to understand the general public's views on education issues in Indonesia.

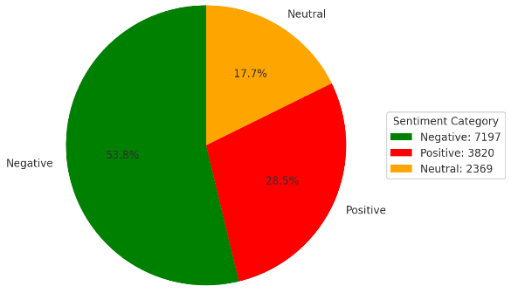

Figure 2. Sentiment distribution

 ISSN 2723-6471

Confusion matrix analysis and performance metrics to evaluate the model used, we use the confusion matrix obtained for each model and its R.O.C. curve [15]. Figure 3 below are the Naive Bayes results for a test size of 10% with an accuracy of 72.51%.

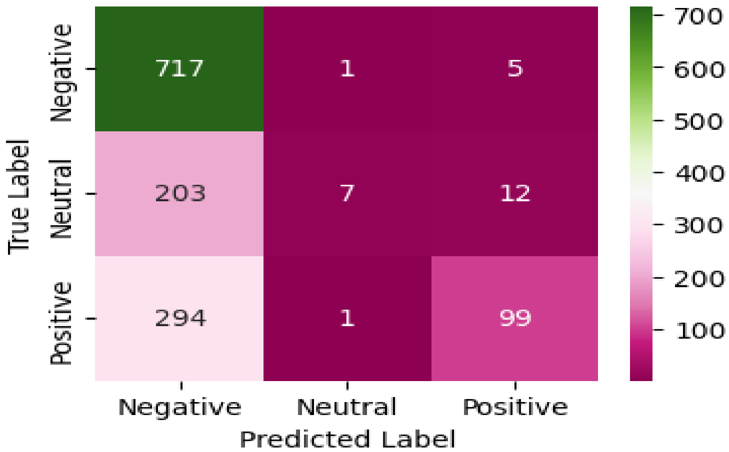

Figure 3. Confusion matrix

The classification results of calculations for the 10% data test used in (2), (3), (4), and (5) were conducted. The Accuracy is 72.51%, Precision is 0.853, Recall is 0.251, and $F1$ Score is 0.388. That means the Naive Bayes model tested achieved an accuracy of 72.51%. It indicates that most of the predictions made are accurate. This model has a high precision rate of 85.30%, which suggests that most predictions are optimistic. However, the model has a weakness in Recall, which is only 25.10%. It indicates that the model needed to have identified many positive cases that were present. Also, the $F1$ Score is recorded at a low 38.8%, indicating that the model still needs improvement to balance Precision and the ability to detect all true positives. The Naive Bayes results for a test size of 20% with an accuracy of 72.40% are shown in figure 4 and figure 5.

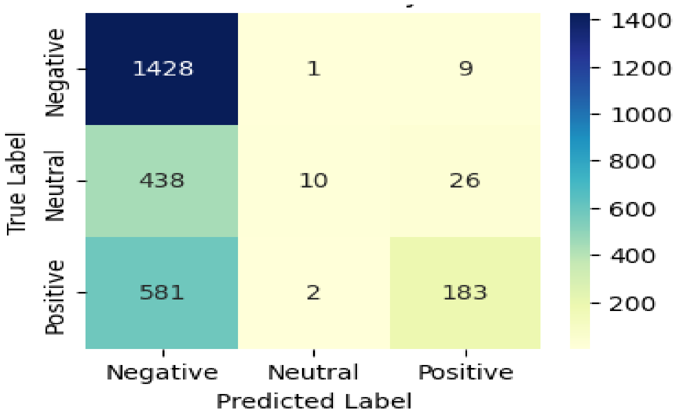

Figure 4. Confusion matrix

Classification Report

	precision	recall	f1-score	support
Negative	0.58	0.99	0.74	1438
Neutral	0.77	0.02	0.04	474
Positive	0.84	0.24	0.37	766
accuracy			0.61	2678
macro avg	0.73	0.42	0.38	2678
weighted avg	0.69	0.61	0.51	2678

Figure 5. Classification report

Also, the classification results of calculations of the 20% data test use (2), (3), (4), and (5) were conducted. The Accuracy is 72.40%, Precision is 0.839, Recall is 0.239, and F1 Score is 0.372. That means the Naive Bayes model tested with a 20% test sample size achieved an accuracy of 72.40%, demonstrating a relatively high level of Accuracy in its predictions. The model's precision rate is 83.9%, indicating that most optimistic forecasts are correct.

However, the model has limitations in Recall, which is only 23.9%, indicating that the model needed to identify the number of positive cases. The result of the F1 Score obtained is 37.2%. This implies that improvements are still required to effectively balance precision and recall detecting positive cases. Figure 6 is a detailed explanation of the results obtained from the Naive Bayes model, which was tested with a 30% test size and an accuracy of 72.25%, and the classification report in figure 7.

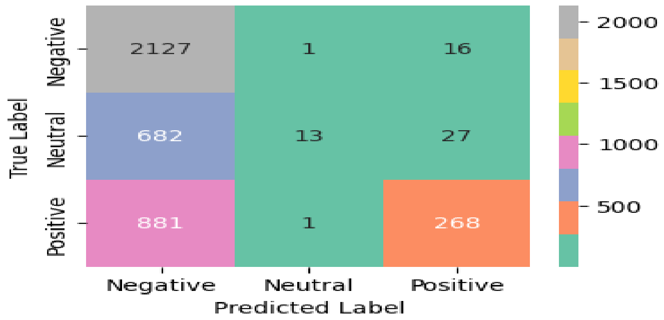

Figure 6. Confusion matrix

Classification Report

	precision	recall	f1-score	support
Negative	0.58	0.99	0.73	2144
Neutral	0.87	0.02	0.04	722
Positive	0.86	0.23	0.37	1150
accuracy			0.60	4016
macro avg	0.77	0.41	0.38	4016
weighted avg	0.71	0.60	0.50	4016

Figure 7. Classification report

The classification results of calculations of the 30% data test use (2), (3), (4), and (5). The Accuracy is 72.25%, Precision is 0.862, Recall is 0.233, and F1 Score is 0.367. That implies the Naive Bayes model with a 30% sample size results in an accuracy of 72.25%, showing a relatively high prediction accuracy. The model's Precision reached 86.2%, indicating high effectiveness in identifying optimistic predictions with few errors. However, the model has a weakness that needs improvement, only 23.3%, indicating a lack of effectiveness in identifying all existing positive cases. This results in a low F1 score of 36.7%, indicating the need for improvement in balancing Precision and positive case detection. The confusion matrix analysis for each test data subset size shows that the model can identify negative comments but struggles to accurately classify positive and neutral comments. This may reflect a bias in the training data or a tendency in how positive and neutral comments are articulated [16]. Table 1 compares accuracy values from the performance of each algorithm.

Table 1. Comparison Results of the Test

Comparison Results				
Data Test %	**Accuracy**	**Precision**	**Recall**	**F1-score**
10	72.551101	0.853448	0.251269	0.388235
20	72.398392	0.839450	0.238903	0.371951
30	72.247225	0.861736	0.223043	0.366872

Word Cloud and word frequency visualization data visualization through word clouds and frequency graphs provide deep insights into the dominant themes and topics in discussions about education in Indonesia. The detailed explanation of these two types of visualizations is Positive Word Cloud, Which displays words like "guru" (teacher), "didik" (educate), and "sekolah" (school), showing the positive aspects often associated with education in Indonesia. The word "guru" appears 1,580 times, indicating respect or satisfaction with the role of teachers in education as seen in figure 8.

Figure 8. Word clouds for positive

Then, the Negative Word Cloud (see figure 9) Highlights words like "enggak" (no), "tidak" (not), and "masalah" (issue), with "enggak" appearing 3,184 times, indicating the prevalence of negative sentiment toward specific issues in education. Neutral Word Cloud: Contains words like "anak" (child), "bangsa" (nation), and "tri", which tend to be descriptive or general, reflecting neutral discussions about education.

Also, the Neutral Word Cloud contains words like "anak" (child), "bangsa" (nation), and "tri," which tend to be descriptive or general and reflect neutral discussions about education, as shown in figure 10. The following will show the frequency of the Positive Frequency Graph: The word "guru" dominates with a frequency of 1,580, followed by "ajar" (teach) and "sekolah" (school). It indicates a focus on teaching and the school environment as positive aspects. Neutral Frequency Graph: The word "anak" appears most frequently, followed by "bangsa" and "tri," indicating topics often discussed in a more objective or neutral context. The negative tops words also showed in figure 11.

Figure 9. Word clouds for negative

Figure 10. Word clouds for neutral

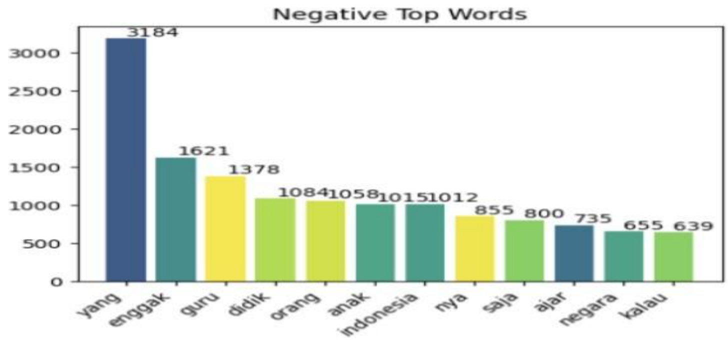

Figure 11. Frequency graphs for negative

Neutral Frequency Graph: The word "anak" appears most frequently, followed by "bangsa" and "tri" indicating topics often discussed in a more objective or neutral context.

In this research, the discussion sentiment analysis of YouTube comments reveals that most public responses to education in Indonesia are negative, with a striking tendency toward words like "enggak" and "masalah" frequently appearing. That indicates significant issues within the education system, mainly related to the quality of teaching and infrastructure. Although the Naive Bayes Classifier model achieved a stable accuracy above 72%, the finding of low Recall on positive comments indicates that the model is less effective in capturing the more subtle positive nuances, possibly due to the lack of frequency or diversity in positive expressions in the data. These findings provide important insights for policymakers to identify and address areas that require improvement in education.

Studying public sentiment allows policymakers to gain a deeper understanding of the aspects considered harmful by the public and to strengthen the elements that receive positive responses. That encourages the revision and development of strategies and policies that more effectively target the real issues perceived by the Neutral Frequency Graph: The word "anak" appears most frequently, followed by "bangsa" and "tri" indicating topics often discussed in a more objective or neutral context.

5. Conclusion

This study investigated public sentiment toward education in Indonesia through sentiment analysis of YouTube comments, using the Naive Bayes Classifier as the primary analytical tool. A total of 17,920 comments were collected, of which 13,386 were successfully processed after thorough preprocessing stages. The sentiment distribution revealed that the majority of comments, 53.8%, were negative, reflecting dissatisfaction with various aspects of the education system, including teaching quality, infrastructure, and policy implementation. Meanwhile, 28.5% of comments were positive, highlighting elements of education that are appreciated, such as teacher roles and school environments, while the remaining 17.7% of comments were neutral, indicating objective or descriptive discussions. The Naive Bayes Classifier achieved a commendable accuracy of 72.51%, demonstrating its effectiveness in sentiment identification. Precision scores of over 85% for some test cases indicate that the model is particularly adept at avoiding false positives, making it reliable for optimistic sentiment detection. However, the model's recall, which measures its ability to identify all relevant sentiments, remains low at 0.251. This indicates that the model struggles to capture the full range of positive and negative nuances present in the comments. The F1 score of 0.388 further highlights the need for improvement in balancing precision and recall for comprehensive sentiment detection.

Analysis of word clouds and frequency graphs provided deeper insights into public perceptions. Positive word clouds highlighted terms like "guru" (teacher) and "didik" (educate), suggesting respect and appreciation for educators. Conversely, negative word clouds prominently featured words such as "enggak" (no) and "masalah" (issue), indicating dissatisfaction and the prevalence of challenges within the education system. Neutral discussions frequently revolved around general topics such as "anak" (child) and "bangsa" (nation), which reflect broader societal concerns. The findings from this research are significant for policymakers. The high proportion of negative sentiments underscores the urgent need to address systemic issues, including resource allocation, policy inefficiencies, and infrastructure deficits. Positive sentiments point to opportunities to strengthen aspects that are working well, such as improving teacher support and training. Furthermore, the neutral sentiment reflects a baseline of public discussion, providing additional context for targeted policy interventions. A hybrid machine learning model tailored to Indonesia's educational context could be developed to help policymakers gain a deeper understanding of public sentiment and address key issues in education policy. The hybrid model to analyze sentiment can be implemented using Scrumban [27], [28], or a hybrid Lean and Agile methods [29].

6. Declarations

6.1. Author Contributions

Conceptualization: H., A., and A.S.; Methodology: T.B.K.; Software: H.; Validation: A., A.S., and D.A.D.; Formal Analysis: H., T.B.K., and D.A.D.; Investigation: A.S., and M.A.; Resources: T.B.K.; Data Curation: D.A.D.; Writing Original Draft Preparation: H., A., M.A., and T.B.K.; Writing Review and Editing: A.S. and D.A.D.; Visualization: H.; All authors have read and agreed to the published version of the manuscript.

6.2. Data Availability Statement

The data presented in this study are available on request from the corresponding author.

6.3. Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

6.4. Institutional Review Board Statement

Not applicable.

6.5. Informed Consent Statement

Not applicable.

6.6. Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] G. F. Medina-Hinostroza, A. L. Nina-Medina, P. Nina-Medina, and S. Tasayco-Barrios, "Learning to Read and Write at The Initial Level," *International Journal of Religion*, vol. 5, no. 9, pp. 434–441, May 2024, doi: 10.61707/ag8n6v26.

[2] N. Azizah, "Public Sentiment Analysis on the 2024 Presidential Election Using Naive Bayes Classifier (NBC) and Support Vector Machine (SVM) on Social Media Data," in *Prosiding Seminar Internasional Batch 1 Amal Insani Foundation*, vol. 2024, no. 1, pp. 1–8, 2024.

[3] B. M. A. S. A. Bangkara, P. J. Pattiasina, E. Fatmawati, A. Heryani, and A. Damayanto, "Relevance of education policy and implementation in Indonesia," *Linguistics and Culture Review*, vol. 6, no. 1, pp. 216–232, Feb. 2022, doi: 10.21744/lingcure.v6ns5.2156.

[4] K. Kanclerz, P. Milkowski, and J. Kocon, "Cross-lingual deep neural transfer learning in sentiment analysis," *in Procedia Computer Science*, vol. 2020, no. 1, pp. 128–137, 2020, doi: 10.1016/j.procs.2020.08.014.

[5] H. Sapulete, F. Sopacua, and V. Sopacua, "The Analysis of Students' Metacognitive Skills in Physics through Problem-Solving Strategies in Physics Education Students," *Jurnal Penelitian Pendidikan IPA*, vol. 10, no. 7, pp. 4453–4460, Jul. 2024, doi: 10.29303/jppipa.v10i7.7102.

[6] R. Prabaswara, J. Lemantara, and J. Jusak, "Classification of Secondary School Destination for Inclusive Students using Decision Tree Algorithm," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 7, no. 5, pp. 1–6, Aug. 2023, doi: 10.29207/resti.v7i5.5081.

[7] P. Y. Saputra, A. D. W. Sumari, Y. W. Syaifudin, and V. S. M. Navalino, "Understanding People Opinion on Artificial Intelligence Ethics through Machine Learning-based Sentiment Analysis," *International Journal of Future Trends in Engineering*, vol. 2023, no. 1, pp. 1–10, 2023.

[8] A. Meiriza, E. Lestari, P. Putra, A. Monaputri, and D. A. Lestari, "Advances in Intelligent Systems Research," *International Journal of Advances in Research*, vol. 2020, no. 1, pp. 1–5, 2020.

[9] P. Assiroj, A. Kurnia, and S. Alam, "The performance of Naïve Bayes, support vector machine, and logistic regression on Indonesia immigration sentiment analysis," *Bulletin of Electrical Engineering and Informatics*, vol. 12, no. 6, pp. 3843–3852, Dec. 2023, doi: 10.11591/eei.v12i6.5688.

[10] A. B. Alawi and F. Bozkurt, "A hybrid machine learning model for sentiment analysis and satisfaction assessment with Turkish universities using Twitter data," *Decision Analytics Journal*, vol. 11, no. 1, pp. 1–10, Jun. 2024, doi: 10.1016/j.dajour.2024.100473.

[11] A. F. Watratan, E. Utami, and A. D. Hartanto, "Comparison of Naive Bayes and PSO-Based Naive Bayes Algorithms for Prediction of Covid-19 Patient Recovery Data in Indonesia," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 7, no. 4, pp. 809–816, Aug. 2023, doi: 10.29207/resti.v7i4.4893.

[12] Y. Mao, Q. Liu, and Y. Zhang, "Sentiment analysis methods, applications, and challenges: A systematic literature review," *King Saud bin Abdulaziz University Journal*, vol. 2024, no. 4, pp. 1–10, Apr. 2024, doi: 10.1016/j.jksuci.2024.102048.

[13] N. Das, B. Sadhukhan, R. Chatterjee, and S. Chakrabarti, "Integrating sentiment analysis with graph neural networks for enhanced stock prediction: A comprehensive survey," *Decision Analytics Journal*, vol. 2024, no. 3, pp. 1–12, Mar. 2024, doi: 10.1016/j.dajour.2024.100417.

[14] T. Anderson, S. Sarkar, and R. Kelley, "Analyzing public sentiment on sustainability: A comprehensive review and application of sentiment analysis techniques," *Natural Language Processing Journal*, vol. 8, no. 9, p. 100097, Sep. 2024, doi: 10.1016/j.nlp.2024.100097.

[15] P. Rasivisuth, M. Fiaschetti, and F. Medda, "An investigation of sentiment analysis of information disclosure during Initial Coin Offering (ICO) on the token return," *International Review of Financial Analysis*, vol. 95, no. 10, pp. 1–10, Oct. 2024, doi: 10.1016/j.irfa.2024.103437.

[16] M. Harris, J. Jacobson, and A. Provetti, "Sentiment and time-series analysis of direct-message conversations," *Forensic Science International: Digital Investigation*, vol. 49, no. 6, pp. 1–8, Jun. 2024, doi: 10.1016/j.fsidi.2024.301753.

[17] J. R. Jim et al., "Recent advancements and challenges of NLP-based sentiment analysis: A state-of-the-art review," *Natural Language Processing Journal*, vol. 6, no. 3, p. 100059, Mar. 2024, doi: 10.1016/j.nlp.2024.100059.

[18] K. Chakma, U. B. Ruba, and S. Das Riya, "YouTube as an information source of floating agriculture: Analysis of Bengali language contents quality and viewers' interaction," *Heliyon*, vol. 8, no. 9, p. e10719, Sep. 2022, doi: 10.1016/j.heliyon.2022.e10719.

[19] P. Breazu, "Entitlement Racism on YouTube: White injury—the license to humiliate Roma migrants in the U.K.," *Discourse, Context and Media*, vol. 55, no. 10, p. 100718, Oct. 2023, doi: 10.1016/j.dcm.2023.100718.

[20] Md. N. Hoque et al., "Exploring transformer models in the sentiment analysis task for the under-resource Bengali language," *Natural Language Processing Journal*, vol. 2024, no. 7, p. 100091, Jul. 2024, doi: 10.1016/j.nlp.2024.100091.

[21] T. Shaik, X. Tao, C. Dann, H. Xie, Y. Li, and L. Galligan, "Sentiment analysis and opinion mining on educational data: A survey," *Natural Language Processing Journal*, vol. 2, no. 3, p. 100003, Mar. 2023, doi: 10.1016/j.nlp.2022.100003.

[22] M. S. Md Suhaimin, M. H. Ahmad Hijazi, E. G. Moung, P. N. E. Nohuddin, S. Chua, and F. Coenen, "Social media sentiment analysis and opinion mining in public security: Taxonomy, trend analysis, issues and future directions," *King Saud bin Abdulaziz University Journal*, vol. 2023, no. 10, pp. 1–15, Oct. 2023, doi: 10.1016/j.jksuci.2023.101776.

[23] M. K. Anam, T. A. Fitri, A. Agustin, L. Lusiana, M. B. Firdaus, and A. T. Nurhuda, "Sentiment Analysis for Online Learning using The Lexicon-Based Method and The Support Vector Machine Algorithm," *ILKOM Jurnal Ilmiah*, vol. 15, no. 2, pp. 290–302, Aug. 2023, doi: 10.33096/ilkom.v15i2.1590.290-302.

[24] M. F. Abdillah and K. Kusnawi, "Comparative Analysis of Long Short-Term Memory Architecture for Text Classification," *ILKOM Jurnal Ilmiah*, vol. 15, no. 3, pp. 455–464, Dec. 2023, doi: 10.33096/ilkom.v15i3.1906.455-464.

[25] A. Daza, N. D. González Rueda, M. S. Aguilar Sánchez, W. F. Robles Espíritu, and M. E. Chauca Quiñones, "Sentiment Analysis on E-Commerce Product Reviews Using Machine Learning and Deep Learning Algorithms: A Bibliometric Analysis and Systematic Literature Review, Challenges and Future Works," *Journal of Intelligent Information Management*, vol. 2024, no. 11, pp. 1–12, Nov. 2024, doi: 10.1016/j.jjimei.2024.100267.

[26] S. R. Velu, V. Ravi, and K. Tabianan, "Multi-lexicon classification and valence-based sentiment analysis as features for deep neural stock price prediction," *Sci.*, vol. 5, no. 1, p. 8, 2023.

[27] M. Alqudah and R. Razali, "An empirical study of Scrumban formation based on the selection of scrum and Kanban practices," *Int. J. Adv. Sci. Eng. Inf. Technol.*, vol. 8, no. 6, pp. 2315–2322, 2018.

[28] M. K. Alqudah, R. Razali, and M. K. Alqudah, "Agile methods selection model: a grounded theory study," *Int. J. Adv. Comput. Sci. Appl.*, vol. 10, no. 7, pp. 1-12, 2019.

[29] M. K. Alqudah, R. Razali, M. K. Alqudah, M. N. Al Dalaien, H. M. Alabool, and H. A. Alkhazaleh, "A grounded theory of selecting lean and agile practices for software development," *J. Softw.: Evol. Process*, vol. 36, no. 4, pp. 1-13, 2024.