Indonesian Journal of Data and Science

Volume 7 Issue 1
ISSN 2715-9936
https://doi.org/10.56705/ijodas.v7i1.382

Research Article

Comparison of Naïve Bayes and SVM in Sentiment Analysis of ChatGPT for Learning on X and YouTube

Ni Putu Eka Swari [1]; Ni Wayan Jeri Kusuma Dewi [2]; Ni Ketut Utami Nilawati [3]; Aniek Suryanti Kusuma [4]; Ni Luh Wiwik Sri Rahayu Ginantra [5]

[1] Institut Bisnis Dan Teknologi Indonesia, Denpasar, Indonesia, ekasuari02@gmail.com
[2] Institut Bisnis Dan Teknologi Indonesia, Denpasar, Indonesia, wayan.kusumadewi@instiki.ac.id
[3] Institut Bisnis Dan Teknologi Indonesia, Denpasar, Indonesia, utami.nilawati@instiki.ac
[4] Institut Bisnis Dan Teknologi Indonesia, Denpasar, Indonesia, anieksuryanti@instiki.ac.id
[5] Institut Bisnis Dan Teknologi Indonesia, Denpasar, Indonesia, wiwik@instiki.ac.id

Correspondence should be addressed to Ni Putu Eka Swari; ekasuari02@gmail.com

Received 28 November 2025; Accepted 18 Jan 2026; Published 30 March 2026

© Authors 2026. CC BY-NC 4.0 (non-commercial use with attribution, indicate changes).
License: https://creativecommons.org/licenses/by-nc/4.0/ — Published by Indonesian Journal of Data and Science.

Abstract:
The rapid development of artificial intelligence technology has encouraged users to actively express opinions on social media platforms such as X and YouTube, including discussions on the use of ChatGPT as a learning support tool. This study aims to analyze public sentiment toward the use of ChatGPT in learning contexts by comparing the performance of the Naïve Bayes and Support Vector Machine (SVM) classification methods. A total of 5,500 comments from platform X and 5,543 comments from YouTube were collected through a crawling process using relevant keywords during the period from January 2023 to December 2025. The data were preprocessed and labeled into three sentiment classes (positive, negative, and neutral) using a lexicon-based approach with the INSET Lexicon. Feature extraction was conducted using the Term Frequency–Inverse Document Frequency (TF-IDF) method, and the dataset was divided into training and testing sets with an 80:20 ratio. Model performance was evaluated using accuracy, precision, recall, and F1-score. The results show that the SVM classifier consistently outperformed the Naïve Bayes method on both platforms. On platform X, SVM achieved an accuracy of 76.67%, while Naïve Bayes obtained 74.60%. On YouTube, SVM achieved an accuracy of 73.10%, significantly higher than Naïve Bayes at 62.04%. These findings indicate that SVM is more effective for sentiment analysis of social media data related to the use of ChatGPT in learning environments.

Keywords: Sentiment Analysis, Chat GPT, Naïve Bayes, Support Vector Machine (SVM), Social Media.

1. Introduction

The rapid development of artificial intelligence (AI) has led to its widespread adoption across various domains, including education [1]. In recent years, AI-based tools have increasingly been used to support learning activities by assisting students in understanding course materials, completing assignments, and improving learning efficiency [2]. One prominent example is ChatGPT, an advanced language model developed by OpenAI that is capable of generating human-like responses and providing explanations across diverse subject areas [3]. Previous studies have reported that AI-powered educational tools can offer personalized learning support, although concerns related to ethics, academic integrity, and data privacy remain significant challenges [4].

The growing use of ChatGPT in educational contexts has generated diverse public responses. On the one hand, ChatGPT is perceived as a helpful tool that enables quick access to information and enhances students' understanding of learning materials [5]. On the other hand, concerns have been raised regarding students' overreliance on AI, which may reduce independent learning, critical thinking, and creativity [6]. These contrasting perceptions are frequently expressed on social media platforms, where users openly share their opinions and experiences related to the use of AI

tools in education [7]. As a result, social media data provide a valuable source for examining public sentiment toward ChatGPT as a learning support tool.

Sentiment analysis has been widely applied to analyze public opinions expressed in textual data on social media [8]. Machine learning methods such as Naïve Bayes and Support Vector Machine (SVM) are among the most commonly used classifiers for sentiment analysis tasks due to their effectiveness and computational efficiency [9]. However, most previous studies focus on single-platform datasets or general sentiment topics, while cross-platform sentiment analysis related to the use of ChatGPT in educational contexts remains limited, particularly for datasets derived from different social media platforms with distinct characteristics [10].

Based on this research gap, this study aims to compare the performance of Naïve Bayes and Support Vector Machine (SVM) in sentiment analysis of public opinions regarding the use of ChatGPT for learning on two social media platforms, namely X and YouTube. Sentiment labeling is performed using the Indonesian INSET Lexicon, and model performance is evaluated using accuracy, precision, recall, and F1-score metrics. Accordingly, the research question addressed in this study is: Which classification method, Naïve Bayes or SVM, demonstrates better performance in analyzing sentiment toward ChatGPT usage for learning across different social media platforms? This study contributes by providing empirical evidence on cross-platform sentiment characteristics and comparative classifier performance in the context of AI-supported learning, thereby supporting the selection of appropriate machine learning methods for educational data analysis.

2. Method

This study employs a comparative quantitative research design using text mining and machine learning approaches to analyze user sentiment toward the use of ChatGPT as a learning support tool. The data consist of user comments collected from social media platforms X and YouTube through a crawling process using keywords related to ChatGPT and learning during the period from January 2023 to December 2025. A total of 5,500 comments were collected from platform X and 5,543 comments from YouTube. The overall research workflow is illustrated in the flowchart presented in **Figure 1**.

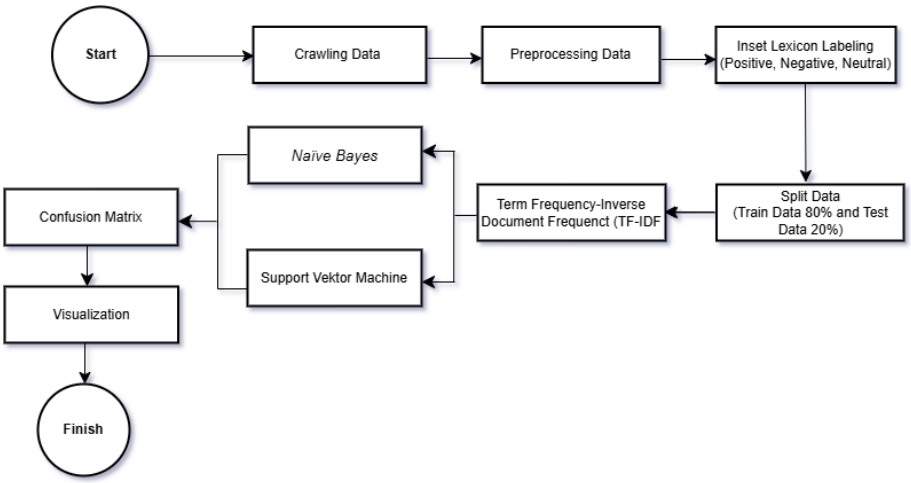

Figure 1. Research Flow

The collected data were subjected to a preprocessing stage to remove irrelevant elements, followed by sentiment labeling using the Indonesian INSET Lexicon and feature weighting using the Term Frequency–Inverse Document Frequency (TF-IDF) method. The labeled data were classified into three sentiment classes: positive, negative, and neutral. The dataset was then divided into training and testing sets using an 80:20 ratio prior to the classification process. Sentiment classification was performed using the Naïve Bayes and Support Vector Machine (SVM) methods. The performance of both methods was evaluated using a confusion matrix, from which accuracy, precision, recall,

and F1-score were calculated. The results of both classifiers were subsequently compared to determine the best-performing method for sentiment analysis across different social media platforms.

Data Selection

The data used in this study consist of user comments related to the use of ChatGPT as a learning support tool, collected from social media platforms X and YouTube. Data collection was conducted through a crawling process using keywords related to ChatGPT and learning, such as "ChatGPT for learning," "ChatGPT education," "ChatGPT homework," and their Indonesian equivalents. These keywords were applied uniformly across both platforms to ensure consistency in data retrieval. The crawling process aimed to capture public discussions and opinions regarding the use of ChatGPT in educational contexts. Data were collected during the period from January 2023 to December 2025.

Table 1. Data Volume After Crawling and Cleaning Stages

Category	Platform X	YouTube
Crawling Result	5,500	5,543
After Duplicate Remove	3,270	5,394
After Filtering Relevance	2,163	2,301

Based on the crawling process, a total of 5,500 comments were obtained from platform X and 5,543 comments from YouTube. To ensure data quality and reproducibility, a structured and sequential data-cleaning procedure was applied. The first step involved duplicate removal to eliminate repeated comments originating from reposts, retweets, or identical content. The second step consisted of relevance filtering, in which comments unrelated to the use of ChatGPT as a learning support tool were excluded. This filtering process removed spam content, promotional messages, out-of-context discussions, and comments without explicit opinions related to learning or educational use of ChatGPT. The relevance assessment was conducted based on keyword presence and contextual relevance to learning activities.

After the data cleaning process, the final dataset consisted of 2,163 comments from platform X and 2,301 comments from YouTube. This gradual filtering process reduced noise in the dataset and ensured that the data used in this study accurately represent user sentiment toward ChatGPT usage in learning environments, thereby supporting the reliability and validity of subsequent sentiment analysis and classification results. A comparison of the data volume before and after each data-cleaning stage is presented in **Table 1**.

Preprocessing

Preprocessing is an important step in this study that aims to prepare textual data before further analysis is conducted. At this stage, raw and unstructured comment data are processed to become more organized and consistent. The preprocessing process is carried out using Google Colaboratory by utilizing Python libraries that support text processing [11]. All preprocessing steps are applied comprehensively to the research dataset to ensure that the data are ready for the feature weighting and sentiment classification stages.

Comment data preprocessing is conducted prior to the classification process to prepare the data for sentiment analysis. Through this stage, raw and diverse textual data are simplified to improve computational efficiency and enable the classification model to perform more effectively. In addition, preprocessing aims to remove irrelevant words or elements, standardize word forms and letter cases, and reduce the number of terms that do not contribute to sentiment determination. In this study, the preprocessing stages include emoji conversion to transform emojis into textual representations, case folding to convert all characters to lowercase, cleansing to remove unnecessary characters, normalization to standardize word forms, tokenization to split text into word units, stopword removal to eliminate common words with minimal sentiment relevance, and stemming to reduce words to their root forms [12], [13]. All preprocessing steps are applied consistently to the research dataset to ensure that the data are ready for subsequent analysis processes.

InSet Lexicon Labeling

In this study, sentiment labeling was conducted using a lexicon-based approach with the Indonesian Sentiment Lexicon (INSET), which is specifically designed to identify sentiment polarity in Indonesian-language text. The INSET lexicon consists of sentiment-bearing words annotated with polarity scores, enabling automated sentiment scoring without requiring manually labeled training data. During the labeling process, each comment was tokenized and matched against the INSET dictionary. The polarity scores of all matched words were summed to obtain an overall sentiment score for each comment. Based on this score, comments were classified into three sentiment categories: positive (score > 0), negative (score < 0), and neutral (score = 0). This rule-based thresholding mechanism is widely used in Indonesian sentiment analysis studies because it is transparent, reproducible, and suitable for large-scale social media datasets where manual annotation is costly and time-consuming [14], [15].

Despite its practicality, lexicon-based sentiment labeling has inherent limitations, particularly in handling contextual nuances such as negation, sarcasm, and implicit sentiment. Several recent studies emphasize that lexicon-based methods may introduce noise when applied to informal social media text, but they remain effective when combined with thorough preprocessing and appropriate feature extraction. Prior research on Indonesian sentiment analysis demonstrates that the INSET lexicon provides reliable baseline sentiment labels and is frequently used to support machine learning-based classification tasks. In this study, neutral sentiment labels were retained as a separate class to preserve the natural distribution of user opinions regarding ChatGPT usage in learning contexts, thereby supporting robust multiclass sentiment classification and fair model evaluation in subsequent stages [16].

TF-IDF

Term Frequency–Inverse Document Frequency (TF-IDF) is a feature-weighting method commonly used in text mining and sentiment analysis to represent textual data in numerical form. TF-IDF measures the importance of a word in a document by considering how frequently the word appears in that document and how rare it is across the entire document collection. This approach helps reduce the influence of commonly occurring but less informative words, while emphasizing terms that carry stronger semantic or sentiment information. Previous studies in Indonesian-language sentiment analysis have shown that TF-IDF is effective for extracting discriminative features from social media text and improving classification performance when used with traditional machine learning classifiers [17], [18].

The TF-IDF weight for a term wi in a document d is calculated as follows:

$$\boldsymbol{TF} * \boldsymbol{IDF} = \boldsymbol{TF}(\boldsymbol{wi}, \boldsymbol{d}) * \boldsymbol{IDF}(\boldsymbol{wi}) \quad (1)$$

$$\boldsymbol{IDF}(\boldsymbol{wi}) = \frac{\boldsymbol{N}}{\boldsymbol{DF}(\boldsymbol{wi})} \quad (2)$$

In this study, TF-IDF was applied after preprocessing and sentiment labeling to transform user comments into numerical feature vectors for classification. This representation enables classifiers such as Naïve Bayes and Support Vector Machine to focus on sentiment-relevant terms while minimizing noise from non-informative words. Prior research confirms that TF-IDF remains a reliable and efficient feature extraction technique for Indonesian social media sentiment analysis tasks [19], [20].

Naïve Bayes

Naïve Bayes is a probabilistic classification algorithm based on Bayes' theorem that estimates the probability of a document belonging to a particular class. Despite assuming conditional independence among features, Naïve Bayes remains widely used in text classification due to its simplicity, computational efficiency, and strong performance on high-dimensional textual data [21], [22].

In this study, the Multinomial Naïve Bayes (MNB) variant was employed, as it is particularly suitable for text data represented using term frequency or TF-IDF features. MNB models word occurrence distributions within documents and has been shown to be effective for sentiment classification of user-generated content [23]. The posterior probability of a class Ci given an input document X is calculated as follows:

$$P(Ci|X) = \frac{P(X|Ci).P(Ci)}{P(X)} \quad (3)$$

The neutral sentiment class was treated as an independent category and included in both the training and evaluation stages to support multi-class sentiment classification. This approach prevents bias toward polarized sentiments and allows the classifier to learn patterns associated with comments that do not express clear positive or negative opinions. During evaluation, the neutral class was retained in the confusion matrix and performance metrics to ensure a fair and comprehensive assessment of classification performance across all sentiment categories.

Support Vector Machine

Support Vector Machine (SVM) is a supervised learning algorithm widely used for text classification tasks due to its ability to handle high-dimensional feature spaces, such as those generated by TF-IDF representations. SVM operates by constructing an optimal hyperplane that maximizes the margin between classes in the feature space, allowing the model to achieve good generalization performance on unseen data [24].

In sentiment analysis, SVM is frequently adopted because of its effectiveness in handling sparse and high-dimensional textual data, as well as its competitive performance compared to other traditional machine learning classifiers in text mining studies [25]. Its robustness makes SVM particularly suitable for classifying user-generated content such as social media comments and online reviews.

In this study, a linear kernel SVM was employed to reduce model complexity while maintaining stable classification performance, considering that the input features were derived from TF-IDF weighting. Key parameters, including the kernel type and penalty parameter (C), were explicitly defined to improve method reproducibility. [24] Furthermore, SVM was implemented in a three-class sentiment classification setting (positive, negative, and neutral), where the neutral class was treated as an independent category during both training and evaluation to ensure a fair multi-class assessment [26].

Confusion Matrix

Confusion Matrix is an evaluation tool used to measure the performance of classification models by comparing predicted class labels with actual class labels. It provides a detailed breakdown of correct and incorrect predictions for each class and is widely applied in text classification and sentiment analysis studies. Unlike simple accuracy metrics, the confusion matrix enables a more comprehensive assessment of model behavior, especially in multi-class classification scenarios where class imbalance may occur [27].

In multi-class sentiment classification, such as positive, negative, and neutral categories, the confusion matrix is structured so that rows represent the actual sentiment classes and columns represent the predicted sentiment classes generated by the model. Correct predictions appear along the diagonal of the matrix and are referred to as True Positives (TP) for each respective class. Misclassified instances appear in the off-diagonal cells and are treated as False Negatives (FN) for the corresponding actual class. This formulation allows clear evaluation of how well the model distinguishes between sentiment categories and serves as the basis for calculating accuracy, precision, recall, and F1-score in a multi-class setting [28].

Table 2. Confusion Matrix for Three-Class Sentiment Classification

Actual \ Predicted	Positive	Negative	Neutral
Positive	TP	FN	FN
Negative	FN	TP	FN
Neutral	FN	FN	TP

Based on **Table 2**, the confusion matrix was applied to evaluate sentiment classification performance across three sentiment classes. The neutral sentiment class was treated as an independent category and included consistently during both training and evaluation stages. This approach ensures fair multi-class evaluation and prevents bias toward dominant sentiment classes, which is particularly important in sentiment analysis of user-generated content where neutral opinions naturally occur [29].

3. Result and Discussion

Based on the total polarity score obtained from the INSET lexicon, each comment was classified into one of three sentiment categories: positive (score > 0), negative (score < 0), or neutral (score = 0). This threshold-based lexicon labeling approach enables an objective grouping of user opinions by aggregating sentiment-bearing words within each comment. Such a method has been widely applied in Indonesian sentiment analysis studies, as it provides interpretable sentiment categories that closely align with human judgment and public opinion patterns [17], [18]. By applying this labeling scheme, the overall sentiment tendencies of users toward ChatGPT as a learning support tool can be systematically examined across different social media platforms.

Table 3 presents the sentiment distribution of user comments on Platform X and YouTube after the INSET lexicon labeling process.

Table 3. Sentiment Distribution After INSET Lexicon Labeling

Sentiment Category	Platform X	YouTube
Positive	557	908
Negative	1,463	1,173
Neutral	143	220
Total	2,163	2,301

Based on **Table 3**, negative sentiment is the most dominant category on both platforms. On Platform X, negative sentiment accounts for 1,463 comments, indicating a strong tendency for users to express critical opinions regarding the use of ChatGPT as a learning support tool. Positive sentiment appears in 557 comments, while neutral sentiment represents the smallest proportion with 143 comments, suggesting that most users express explicit opinions rather than neutral statements.

In contrast, the sentiment distribution on YouTube appears more balanced. Although negative sentiment remains the largest category with 1,173 comments, the difference compared to positive sentiment (908 comments) is relatively small. This pattern suggests that YouTube users express more diverse perspectives, including both critical and supportive views toward the use of ChatGPT in learning contexts. Neutral sentiment on YouTube accounts for 220 comments, indicating a slightly higher presence of neutral expressions compared to Platform X.

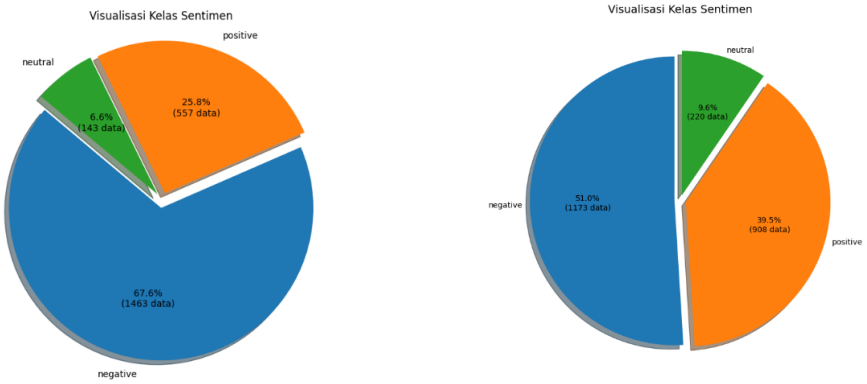

Figure 2. INSET Labeling Result

Figure 2 visualizes the sentiment distribution of user comments on Platform X and YouTube after INSET lexicon labeling. The figure highlights a clear contrast in user expression patterns across platforms, where Platform X is characterized by a strong dominance of negative sentiment, while YouTube exhibits a more balanced distribution between negative and positive sentiments. The presence of neutral sentiment is relatively limited on both platforms, indicating that most users express explicit opinions. Overall, the visualization reinforces the differences in how users articulate their views toward ChatGPT as a learning support tool across social media platforms.

Confusion Matrix

The confusion matrix was employed to evaluate the performance of the Naïve Bayes and Support Vector Machine (SVM) models in classifying sentiment into three categories: negative, neutral, and positive. This evaluation compares the actual sentiment labels with the predicted labels generated by each model, thereby providing insight into classification accuracy as well as misclassification patterns for each sentiment class. The confusion matrix is particularly useful in multi-class sentiment classification, as it allows performance analysis at the class level, including classes that are inherently difficult to distinguish, such as neutral sentiment.

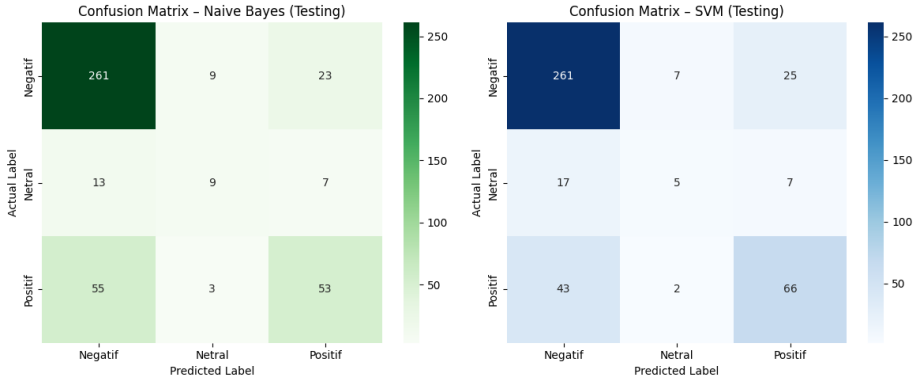

Figure 3. Confusion Matrix Results for Naive Bayes and SVM Platform X

Figure 3 illustrates the confusion matrix results for sentiment classification on the Platform X dataset. The Naïve Bayes model correctly classified 261 negative sentiment instances, but several misclassifications were observed, with 23 negative instances predicted as positive and 9 as neutral. In the positive class, 53 instances were correctly classified, while 55 were misclassified as negative, indicating overlap between negative and positive expressions in user comments. The SVM model demonstrated stronger performance, correctly classifying 261 negative instances and showing improvement in the positive class with 66 correctly classified instances. However, both models struggled to accurately classify neutral sentiment, as reflected by the low number of correct predictions for this class, suggesting that neutral expressions on Platform X are less distinctive and more difficult to separate from polarized sentiments. This pattern indicates that SVM achieves better class-level discrimination, particularly in terms of precision and recall for polarized sentiments, although neutral sentiment remains challenging. Overall, the results indicate that SVM outperforms Naïve Bayes on the Platform X dataset, particularly in distinguishing positive sentiment.

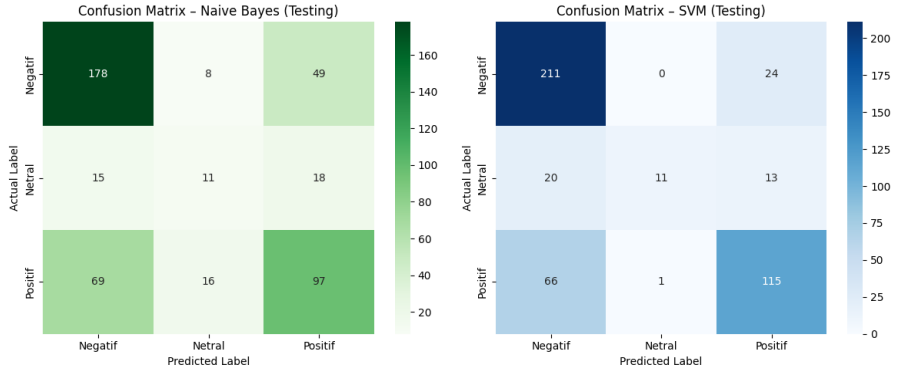

Figure 4. Confusion Matrix Results for Naive Bayes and SVM Platform YouTube

Figure 4 shows the confusion matrix results for sentiment classification on the YouTube dataset. The Naïve Bayes model correctly classified 178 negative sentiment instances and 97 positive instances, but misclassifications were observed between negative and positive sentiments, indicating overlap in user expressions. The SVM model achieved better performance, with 211 correctly classified negative instances and 135 correctly classified positive instances, demonstrating its stronger ability to separate dominant sentiment classes on YouTube. However, similar to the results on Platform X, both models showed limited performance in identifying neutral sentiment, as reflected by the relatively low number of correct neutral predictions. Overall, these results confirm that SVM consistently outperforms Naïve Bayes on the YouTube dataset, while neutral sentiment remains the most challenging class to classify accurately.

Model Evaluation

Based on the evaluation results on Platform X, the Support Vector Machine (SVM) method demonstrates superior performance compared to Naive Bayes across all evaluation metrics. In terms of accuracy, SVM achieves 76.67%, while Naive Bayes attains 74.60%, indicating that SVM is more consistent overall in producing correct sentiment predictions. Regarding precision, SVM records a value of 74.68%, which is higher than Naive Bayes at 72.92%, suggesting that SVM provides more accurate sentiment classification and is better at reducing misclassification errors. The visualization of sentiment prediction results is presented in **Table 4** and **Figure 5**.

Table 4. Naive Bayes and SVM Performance Platform X

Metric	Naïve Bayes (%)	SVM (%)
Accuracy	74,60	76,67
Precision	72,92	74,68
Recall	74,60	76,67
F1-Score	73,21	75,28

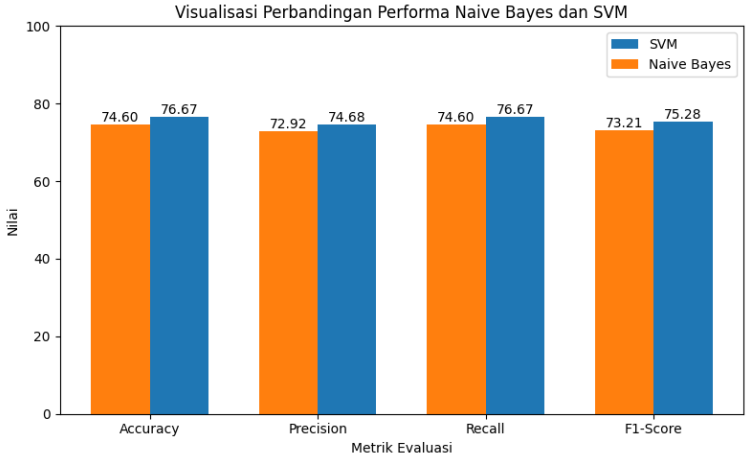

Figure 5. Visualization of Naive Bayes and SVM Performance Comparison Platform X

For the recall metric, SVM again outperforms Naïve Bayes with a value of 76.67%, whereas Naïve Bayes achieves 74.60%, indicating that SVM is more capable of identifying sentiment instances that belong to their correct classes, including those with more complex and overlapping textual characteristics. In terms of F1-score, SVM achieves 75.28%, which is higher than Naïve Bayes at 73.21%, reflecting a better balance between precision and recall. This result suggests that SVM provides more stable and reliable classification performance when handling high-dimensional text data and varying sentiment distributions on Platform X.

Overall, these findings indicate that SVM is more effective and reliable for sentiment classification on Platform X data, while Naïve Bayes can still serve as a comparative baseline despite its relatively lower performance, particularly in handling complex text patterns and data imbalance. The superior performance of SVM in this study can be attributed to its ability to handle high-dimensional TF-IDF features without relying on the word independence assumption used by Naïve Bayes. Moreover, the diverse and unstructured nature of text on Platform X allows SVM to separate sentiment classes more effectively, leading to more stable classification results.

Based on the evaluation results on the YouTube platform, the Support Vector Machine (SVM) method demonstrates superior performance compared to Naïve Bayes across all reported evaluation metrics. In terms of accuracy, SVM achieves 73.10%, while Naïve Bayes attains 62.04%, indicating that SVM is more consistent in correctly classifying the sentiment of YouTube comments. Regarding precision, SVM records a value of 74.83%, which is substantially higher than Naïve Bayes at 60.98%, suggesting that SVM is more effective in reducing misclassification errors between sentiment classes. The visualization of sentiment prediction results is presented in **Table 5** and **Figure 6**.

Table 5. Naive Bayes and SVM Performance Platform YouTube

Metric	Naïve Bayes (%)	SVM (%)
Accuracy	62,04	73,10
Precision	60,98	74,83
Recall	62,04	73,10
F1-Score	61,31	71,37

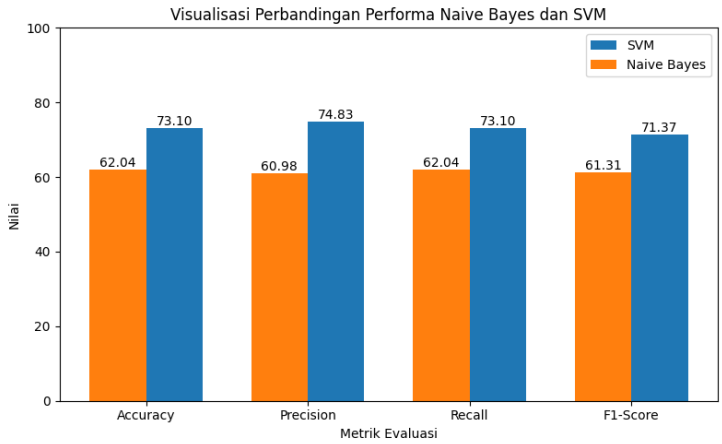

Figure 6. Visualization of Naive Bayes and SVM Performance Comparison Platform YouTube

From the precision perspective, SVM recorded a value of 74.83%, which is considerably higher than that of Naïve Bayes at 60.98%. This indicates that SVM is more accurate in assigning the correct sentiment labels to YouTube comments, thereby reducing misclassification errors between sentiment classes that often overlap in user-generated text. Regarding recall, SVM again outperformed Naïve Bayes with a value of 73.10%, while Naïve Bayes achieved only 62.04%. This result suggests that SVM is more effective in identifying sentiment instances that truly belong to each class, even when dealing with diverse expressions and varying comment lengths.

For the F1-score metric, SVM obtained a value of 71.37%, which is higher than Naïve Bayes at 61.31%, indicating that SVM achieves a better balance between precision and recall, an important aspect in multi-class sentiment classification where class imbalance and semantic ambiguity are common. A higher F1-score suggests that SVM performs more consistently across sentiment categories rather than focusing on dominant classes only. Overall, these results demonstrate that SVM provides more stable and reliable performance for sentiment classification on YouTube data compared to Naïve Bayes, particularly in handling linguistic variability, informal expressions, and complex textual patterns that frequently appear in longer user-generated comments.

Visualization of Sentiment Word Cloud

Figure 7 presents the Word Cloud visualization of positive, negative, and neutral sentiment classes on Platform X, highlighting distinctive word usage patterns across sentiment categories. In the positive sentiment class, words such as "*bantu*", "*bagus*", "*pakai*", "chatgpt", and "*banget*" appear prominently, indicating users' positive experiences, perceived usefulness, and satisfaction with ChatGPT as a learning support tool. In contrast, the negative

sentiment class is dominated by words such as "*enggak*", "*tidak*", "*salah*", and "*ribet*", which reflect user complaints, perceived difficulties, or mismatches between expectations and actual usage experiences. Meanwhile, the neutral sentiment class contains frequently occurring words such as "*jadi*", "*aku*", "*pakai*", "*tanya*", and "*beli*", which carry minimal emotional polarity and mainly represent information sharing, questioning, or general discussion. Overall, the Word Cloud visualization provides qualitative insight into discussion tendencies within each sentiment class on Platform X and serves as a complementary interpretation of the sentiment classification results rather than a definitive indicator of overall sentiment polarity.

Figure 7. WordCloud Visualization Sentiment Classes Platform X

Figure 8 presents the Word Cloud visualizations for positive, negative, and neutral sentiment classes based on user comments from the YouTube platform. The visualization illustrates the most frequently occurring words in each sentiment category, where word size reflects frequency of occurrence. In the positive sentiment class, words such as "chatgpt", "bantu", and "ajar" appear dominantly, indicating generally positive perceptions of ChatGPT as a learning support tool among YouTube users. In contrast, the negative sentiment class is characterized by frequent occurrences of words such as "*enggak*" and "salah", reflecting user criticism, misunderstandings, or dissatisfaction related to ChatGPT usage. Meanwhile, the neutral sentiment class is dominated by commonly used words without strong emotional connotations, suggesting activities such as asking questions, sharing information, or participating in general discussions. This visualization supports the sentiment distribution results on YouTube and strengthens the qualitative interpretation of the classification outcomes.

Figure 8. WordCloud Visualization Sentiment Classes Platform YouTube

The Word Cloud analysis indicates that discussions related to the use of ChatGPT as a learning medium involve diverse user perspectives across sentiment classes on both platforms. Positive sentiment is associated with perceived usefulness and learning support, while negative sentiment highlights confusion, errors, or inconvenience arising from unmet expectations or improper usage. Neutral sentiment primarily reflects informational exchanges and contextual discussions without strong emotional orientation. These findings suggest that user sentiment toward ChatGPT is strongly influenced by usage context and learning needs. Therefore, the Word Cloud visualization should be interpreted as a complementary analytical tool that illustrates dominant discussion patterns rather than as an absolute measure of sentiment polarity.

Discussion

The results of this study indicate that the Support Vector Machine (SVM) method consistently outperforms Naïve Bayes in classifying sentiment related to the use of ChatGPT for learning on Platform X and YouTube. SVM achieves higher accuracy, precision, recall, and F1-score on both platforms, with a more pronounced advantage on YouTube, where the data are more diverse and imbalanced. These performance differences align with findings from recent sentiment analysis research showing that SVM often provides more stable classification results than Naïve Bayes in social media text contexts, largely due to the effectiveness of SVM in handling high-dimensional TF-IDF feature

spaces and flexibility with complex textual patterns [30], [31] . Studies comparing SVM and Naïve Bayes across various domains, including marketplace reviews and public opinion data, also find that SVM tends to outperform Naïve Bayes in terms of accuracy, precision, and F1-score due to its ability to maximize margin separation between classes [32].

Moreover, comparative research has reported that while Naïve Bayes is computationally efficient and suitable as a baseline classifier, its assumption of feature independence can limit its performance in contexts where feature interactions are complex - a limitation less pronounced in SVM models [33]. Although Naïve Bayes can yield satisfactory results in specific applications, multiple recent studies suggest that SVM generally provides stronger performance for multi-class sentiment tasks involving unstructured user comments from social media platforms, partly because SVM's decision boundaries better handle overlapping class distributions.

Word Cloud visualizations in this study further support these findings by illustrating distinct language patterns across sentiment classes. Dominant words in the positive sentiment class emphasize perceived usefulness and support for learning, while negative sentiment words reflect confusion or dissatisfaction. Neutral sentiment primarily contains informational or general discussion terms. These qualitative insights suggest that sentiment polarity is influenced more by the context of ChatGPT usage in learning rather than by the technology itself, a pattern similarly noted in other sentiment analysis research that combines quantitative classification metrics with lexical visualization [34].

Overall, the study's findings confirm that SVM provides more stable and reliable sentiment classification compared to Naïve Bayes for comments related to ChatGPT usage on social media, especially in high-dimensional and imbalanced datasets. Future research could explore additional sentiment categories, larger multilingual corpora, or advanced models such as transformer-based architectures (e.g., BERT) to improve classification accuracy further and practical relevance for AI-based learning development. Additionally, future studies should consider enhancing evaluation methodologies by incorporating cross-validation and reporting performance metrics at the class level to address challenges in neutral sentiment classification.

4. Conclusion

This study examined public sentiment toward the use of ChatGPT as a learning support tool on social media platforms X and YouTube by comparing the performance of the Naïve Bayes and Support Vector Machine (SVM) classifiers using accuracy, precision, recall, and F1-score as evaluation metrics. The results consistently demonstrate that SVM outperforms Naïve Bayes across both platforms, with stronger performance observed on the YouTube dataset, which is characterized by greater linguistic diversity and class imbalance. These findings indicate that SVM is more robust in handling high-dimensional text features and complex sentiment distributions under the experimental settings applied in this study. The Word Cloud analysis further supports the quantitative results by revealing distinct language patterns across positive, negative, and neutral sentiment classes, suggesting that user sentiment is shaped primarily by the context and manner in which ChatGPT is used in learning activities rather than by the technology itself. Overall, the findings confirm that SVM provides a more stable and reliable approach for sentiment classification of social media data related to educational technology, while Naïve Bayes remains suitable as a comparative baseline method within the defined experimental scope.

References:

[1] M. F. Saleh and R. Imanda, "Public Sentiment Analysis of the Free Meal Program: A Comparison of Naive Bayes and Support Vector Machine Methods on the Twitter (X) Social Media Platform," *J. Appl. Informatics Comput.*, vol. 9, no. 1, pp. 131–139, 2025, doi: 10.30871/jaic.v9i1.8895.

[2] S. Ngarifatul Khofiyah and P. Subarkah, "Comparison of Naive Bayes and SVM in Public Opinion Sentiment Analysis on Platform X," *J. Teknol. Inf. Univ. Lambung Mangkurat*, pp. 125–138, 2025, doi: 10.20527/jtiulm.v10i2.478.

[3] M. Alawida, S. Mejri, A. Mehmood, B. Chikhaoui, and O. Isaac Abiodun, "A Comprehensive Study of ChatGPT: Advancements, Limitations, and Ethical Considerations in Natural Language Processing and Cybersecurity," *Inf.*, vol. 14, no. 8, 2023, doi: 10.3390/info14080462.

[4] R. F. P. Pratama and W. Maharani, "Comparative Analysis of Naive Bayes and SVM for Improved Emotion Classification on Social Media," *Edumatic J. Pendidik. Inform.*, vol. 9, no. 1, pp. 11–20, 2025, doi: 10.29408/edumatic.v9i1.29087.

[5] A. Göçen, M. M. Ibrahim, and A. U. I. Khan, "Public attitudes toward higher education using sentiment analysis and topic modeling," *Discov. Artif. Intell.*, vol. 4, no. 1, 2024, doi: 10.1007/s44163-024-00195-4.

[6] Y. Mao, Q. Liu, and Y. Zhang, "Sentiment analysis methods, applications, and challenges: A systematic literature review," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 36, no. 4, p. 102048, 2024, doi: 10.1016/j.jksuci.2024.102048.

[7] S. Yang, Y. Dong, and Z. G. Yu, "ChatGPT in Education: Ethical Considerations and Sentiment Analysis," *Int. J. Inf. Commun. Technol. Educ.*, vol. 20, no. 1, pp. 1–19, 2024, doi: 10.4018/IJICTE.346826.

[8] N. F. B. Casillano, "Education in the ChatGPT Era: A Sentiment Analysis of Public Discourse on the Role of Language Models in Education," *J. Eval. Educ.*, vol. 5, no. 4, pp. 144–154, 2024, doi: 10.37251/jee.v5i4.1151.

[9] L. Lu *et al.*, "Healthcare professionals and the public sentiment analysis of ChatGPT in clinical practice," *Sci. Rep.*, vol. 15, no. 1, pp. 1–11, 2025, doi: 10.1038/s41598-024-84512-y.

[10] D. Smith-Mutegi, Y. Mamo, J. Kim, H. Crompton, and M. McConnell, "Perceptions of STEM education and artificial intelligence: a Twitter (X) sentiment analysis," *Int. J. STEM Educ.*, vol. 12, no. 1, 2025, doi: 10.1186/s40594-025-00527-5.

[11] M. A. Palomino and F. Aider, "Evaluating the Effectiveness of Text Pre-Processing in Sentiment Analysis," *Appl. Sci.*, vol. 12, no. 17, 2022, doi: 10.3390/app12178765.

[12] M. Siino, I. Tinnirello, and M. La Cascia, "Is text preprocessing still worth the time? A comparative survey on the influence of popular preprocessing methods on Transformers and traditional classifiers," *Inf. Syst.*, vol. 121, no. March 2023, p. 102342, 2024, doi: 10.1016/j.is.2023.102342.

[13] A. F. Aufar, Mochamad Alfan Rosid, A. Eviyanti, and I. R. I. Astutik, "Optimizing Text Preprocessing for Accurate Sentiment Analysis on E-Wallet Reviews," *JICTE (Journal Inf. Comput. Technol. Educ.*, vol. 7, no. 2, pp. 42–50, 2023, doi: 10.21070/jicte.v7i2.1650.

[14] X. Ding, B. Liu, and P. S. Yu, "A holistic lexicon-based approach to opinion mining," *WSDM'08 - Proc. 2008 Int. Conf. Web Search Data Min.*, no. February 2008, pp. 231–239, 2008, doi: 10.1145/1341531.1341561.

[15] A. Nurkasanah and M. Hayaty, "Feature Extraction using Lexicon on the Emotion Recognition Dataset of Indonesian Text," *Ultim. J. Tek. Inform.*, vol. 14, no. 1, pp. 20–27, 2022, doi: 10.31937/ti.v14i1.2540.

[16] Y. Fauziah, B. Yuwono, and A. S. Aribowo, "Lexicon Based Sentiment Analysis in Indonesia Languages : A Systematic Literature Review," *RSF Conf. Ser. Eng. Technol.*, vol. 1, no. 1, pp. 363–367, 2021, doi: 10.31098/cset.v1i1.397.

[17] G. Popoola, K. K. Abdullah, G. S. Fuhnwi, and J. Agbaje, "Sentiment Analysis of Financial News Data using TF-IDF and Machine Learning Algorithms," *2024 IEEE 3rd Int. Conf. AI Cybersecurity, ICAIC 2024*, 2024, doi: 10.1109/ICAIC60265.2024.10433843.

[18] H. Barus, I. N. Fajri, and Y. Pristyanto, "Sentiment Classification Analysis of Tokopedia Reviews Using TF-IDF, SMOTE, and Traditional Machine Learning Models," *J. Appl. Informatics Comput.*, vol. 9, no. 5, pp. 2552–2561, 2025, doi: 10.30871/jaic.v9i5.10524.

[19] Andi Riswawan, "Implementing TF-IDF and Logistic Regression for Sentiment Analysis of YouTube Comments on the iPhone 16," *J. Teknol. Dan Open Source*, vol. 7, no. 2, pp. 277–284, 2024, doi: 10.36378/jtos.v7i2.4753.

[20] F. Fitriana and H. Setiawan, "Performance Analysis of SVM In Emotion Classification: A Comparative Study Of TF-IDF and Countvectorizer," *J. Embed. Syst. Secur. Intell. Syst.*, vol. 6, no. 2, pp. 133–145, 2025, doi: 10.59562/jessi.v6i2.8396.

[21] H. Guo, J. Sun, J. Luo, Y. Peng, and C. Ye, "Thickness-Related Fault Diagnosis of Steel Strip Based on W-KPLS Method Considering Mechanism Weight Optimization," *Appl. Sci.*, vol. 12, no. 9, 2022, doi: 10.3390/app12094491.

[22] D. Shalikha and A. Alamsyah, "Improved Accuracy of Naïve Bayes Algorithm and Support Vector Machine Using Particle Swarm Optimization for Menstrual Cup Sentiment Analysis on Twitter," *J. Adv. Inf. Syst. Technol.*, vol. 4, no. 2, pp. 139–148, 2023, doi: 10.15294/jaist.v4i2.59561.

[23] C. Dewi, R. C. Chen, H. J. Christanto, and F. Cauteruccio, "Multinomial Naïve Bayes Classifier for Sentiment Analysis of Internet Movie Database," *Vietnam J. Comput. Sci.*, vol. 10, no. 4, pp. 485–498, 2023, doi: 10.1142/S2196888823500100.

[24] Heri Suroyo and E. J. Pratama, "Comparison of Text Representation Methods for Sentiment Analysis Using Support Vector Machine," *J. Adv. Inf. Ind. Technol.*, vol. 7, no. 1, pp. 21–30, 2025, doi: 10.52435/jaiit.v7i1.610.

[25] T. Ahmed Khan, R. Sadiq, Z. Shahid, M. M. Alam, and M. Mohd Su'ud, "Sentiment Analysis using Support Vector Machine and Random Forest," *J. Informatics Web Eng.*, vol. 3, no. 1, pp. 67–75, 2024, doi: 10.33093/jiwe.2024.3.1.5.

[26] G. Mutanov, V. Karyukin, and Z. Mamykova, "Multi-class sentiment analysis of social media data with machine learning algorithms," *Comput. Mater. Contin.*, vol. 69, no. 1, pp. 913–930, 2021, doi: 10.32604/cmc.2021.017827.

[27] J. Opitz, "A Closer Look at Classification Evaluation Metrics and a Critical Reflection of Common Evaluation Practice," *Trans. Assoc. Comput. Linguist.*, vol. 12, no. 2018, pp. 820–836, 2024, doi: 10.1162/tacl_a_00675.

[28] X. Chen, D. Esserman, and F. Li, "Competing risks regression for clustered survival data via the marginal additive subdistribution hazards model," *Stat. Neerl.*, vol. 78, no. 2, pp. 281–301, 2024, doi: 10.1111/stan.12317.

[29] E. Kahya Özyirmidokuz, B. Molu Elmas, and E. A. Stoica, "AI-Based Sentiment Analysis of E-Commerce Customer Feedback: A Bilingual Parallel Study on the Fast Food Industry in Turkish and English," *J. Theor. Appl. Electron. Commer. Res.*, vol. 20, no. 4, p. 294, 2025, doi: 10.3390/jtaer20040294.

[30] M. M. Danyal, S. S. Khan, M. Khan, M. B. Ghaffar, B. Khan, and M. Arshad, "Sentiment Analysis Based on Performance of Linear Support Vector Machine and Multinomial Naïve Bayes Using Movie Reviews with Baseline Techniques," *J. Big Data*, vol. 5, pp. 1–18, 2023, doi: 10.32604/jbd.2023.041319.

[31] H. Setyawan, L. M. Azizah, and A. Y. Pradani, "Sentiment Analysis of Public Responses on Indonesia Government Using Naïve Bayes and Support Vector Machine," *Emerg. Inf. Sci. Technol.*, vol. 4, no. 1, pp. 1–7, 2023, doi: 10.18196/eist.v4i1.18681.

[32] N. Z. B. Jannah and K. Kusnawi, "Comparison of Naïve Bayes and SVM in Sentiment Analysis of Product Reviews on Marketplaces," *Sinkron*, vol. 8, no. 2, pp. 727–733, 2024, doi: 10.33395/sinkron.v8i2.13559.

[33] M. Hamad and V. Prevelakis, "SAVTA: A hybrid vehicular threat model: Overview and case study," *Inf.*, vol. 11, no. 5, pp. 1–22, 2020, doi: 10.3390/INFO11050273.

[34] J. Nabila, Ida Ayu Devian Branitasandhini Putra, and Heri Wijayanto, "Sentiment Analysis on Starbucks Reviews: Implementation of K-Nearest Neighbors and Support Vector Machine," *Infact Int. J. Comput.*, vol. 9, no. 02, pp. 79–86, 2025, doi: 10.61179/infact.v9i02.761.