Jurnal Ilmu Siber dan Teknologi Digital (JISTED)
ISSN: 2986-7312, Vol 5, No 1, 2026, 35-47 https://doi.org/10.61401/jisted.v5.n1.p35-47.2026

Comparison of Naïve Bayes and SVM for Geopolitical News Sentiment on IHSG Volatility

M. Chesta Adabi Alfarido[1*], Isnandar Agus[2]
Institut Informatika dan Bisnis Darmajaya, Bandar Lampung, Indonesia[1,2]
chestasmkn1geta@gmail.com[1*], isnandaragus@mail.darmajaya.ac.id[2]

Article History
Received on 04 June 2026
1st Revision on 13 June 2026
2nd Revision on 24 June 2026
3rd Revision on 29 July 2026
4th Revision on 18 August 2026
Accepted on 28 August 2026

Abstract

Purpose: This study aims to analyze public sentiment toward geopolitical news related to the conflict between Iran and the United States and examine its relationship with the volatility of the *Indeks Harga Saham Gabungan* (IHSG).

Research Methodology: This study used comment data from YouTube, collected through purposive sampling. The data were processed through preprocessing stages, including case folding, text cleaning, tokenization, normalization, stop-word removal, and stemming. Feature extraction was performed using the TF-IDF method. Classification was conducted using the Naïve Bayes and Support Vector Machine (SVM) algorithms implemented in Python. The dataset consisted of 998 entries, divided into 798 training and 200 testing data.

Results: The results indicate that SVM achieved the best performance, with an accuracy of 82.5%, which is slightly higher than that of Naïve Bayes at 81%. In addition, SVM obtained a higher macro-average F1-score (0.72) than Naïve Bayes (0.61), demonstrating better performance in classifying all classes, particularly minority classes.

Conclusions: SVM was more effective in this study, as it provided a more balanced classification performance, although it showed a higher indication of overfitting compared to Naïve Bayes.

Limitations: This study is limited by a relatively small dataset, imbalanced class distribution, and the use of a single data source (YouTube), which may not fully represent the broader public perception.

Contributions: This study identifies effective text-based sentiment classification methods in geopolitical contexts and provides insights into the relationship between public sentiment and IHSG volatility.

Keywords: *Classification, Geopolitics, Naïve Bayes, Sentiment Analysis, SVM*

How to Cite: Alfarido, M. C. A., & Agus, I. (2026). Comparison of Naïve Bayes and SVM for Geopolitical News Sentiment on IHSG Volatility. *Jurnal Ilmu Siber dan Teknologi Digital*, 5(1), 35-47.

1. Introduction

The development of information technology has accelerated the rapid and widespread dissemination of news, including information related to global geopolitical issues. One of the most prominent issues that continues to attract global attention is the conflict between Iran and the United States, which has persisted over a long period and undergone various phases of escalation (Noviyanto et al., 2025).

During the study period, tensions escalated again, marked by military confrontations and disruptions in strategic routes such as the Strait of Hormuz, which plays a crucial role in global energy distribution. This situation has led to increased economic uncertainty and heightened risk perception in global financial markets, ultimately influencing capital market movements across countries (Emmanuel, 2026).

The *Indeks Harga Saham Gabungan* (IHSG) serves as the primary indicator reflecting the performance of the capital market in Indonesia (Hafifa et al., 2025). Its movement is closely monitored by investors, as it provides a basis for investment decisions such as buying, holding, or selling stocks. The IHSG is inherently volatile and influenced by various internal and external factors. Stock market volatility represents the magnitude of fluctuations in asset returns over time and is widely used as an indicator of financial risk and uncertainty in the market (Tong et al., 2023). Internal factors include macroeconomic conditions such as inflation, interest rates, and exchange rates, while external factors encompass global stock index movements, commodity prices, and international political and security conditions (Wijaya & Agustin, 2015). Therefore, changes in global sentiment particularly those driven by geopolitical events can significantly affect IHSG volatility (Han et al., 2025).

In this context, public sentiment toward geopolitical news plays a crucial role in shaping investor perceptions and behavior. Sentiments whether positive, negative, or neutral can influence market expectations and contribute to stock market dynamics (Zhao et al., 2025). Sentiment analysis, as a subfield of natural language processing, enables the systematic extraction and classification of opinions from textual data (Liu, 2012). In this study, sentiment data are obtained from YouTube comments related to geopolitical news, which represent real-time public reactions to global events.

However, studies that explicitly examine the relationship between geopolitical news sentiment and stock market volatility particularly the IHSG remain limited. Existing research on sentiment analysis predominantly focuses on domestic issues or consumer opinions, providing limited insight into how global geopolitical sentiment influences the Indonesian capital market. Therefore, this study aims to analyze sentiment in geopolitical news related to the conflict between Iran and the United States using machine learning approaches and examine its relationship with IHSG volatility. Specifically, this study applies and compares classification models, namely Support Vector Machine (SVM) and Naïve Bayes, to evaluate their effectiveness in sentiment classification. This study is expected to provide a more comprehensive understanding of how global sentiment influences financial market dynamics in Indonesia.

2. Literature Review and Hypotheses Development

2.1 Natural Processing Languange

Natural Language Processing (NLP) is a field of artificial intelligence that enables computers to understand and process human language (Qin et al., 2026). Several NLP stages used in this study include:

1. Preprocessing: This process involves cleaning the data. It includes actions such as converting text to lowercase, filtering, removing stopwords, and stemming.
2. The significance level of words in a document compared to a collection of documents (corpus) is measured using a term weighting method known as term frequency–inverse document frequency (TF-IDF).

2.2 Data Mining

Data mining is a technique used to extract correlations, trends, and anomalies from large databases to predict specific outcomes. Knowledge Discovery in Databases (KDD) is another term for data mining

and is grounded in the disciplines of statistics, artificial intelligence, and machine learning (Zhang, 2021).

2.3 Text Processing

The preprocessing stage was conducted to clean and prepare text data for subsequent analysis (Allahyari et al., 2017). The process consisted of case folding, text cleaning, tokenization, word normalization, stopword removal, and stemming. Case folding converted all text to lowercase to standardize word representation and minimize feature duplication caused by capitalization differences (Teller, 2000). For example, "Saham Naik" was converted to "saham naik." Text cleaning removed irrelevant elements, including URLs, non-alphabetic characters, and excessive spaces, to reduce noise in the dataset (Vijayarani et al., 2015). For instance, "ihsg naik!!! https://link.com" was transformed into "ihsg naik."

Tokenization divided the cleaned text into individual word units, or tokens, to facilitate word-level analysis (Webster & Kit, 1992). For example, "saham naik hari ini" was segmented into ["saham", "naik", "hari", "ini"]. Word normalization subsequently converted informal words and abbreviations into their standard forms using a manually constructed slang dictionary. This step reduced lexical variation and ensured more consistent feature representation (Roy et al., 2021). Examples included converting "gak" to "tidak" and "yg" to "yang."

Stopword removal eliminated common words considered non-informative for sentiment analysis using the default Sastrawi stopword list supplemented with manually selected terms (Sonkusale et al., 2023). For example, "saham yang bagus" was reduced to "saham bagus." Finally, stemming was performed using Sastrawi to convert words into their root forms, consolidate morphological variants, and reduce feature dimensionality (Tala, 2003). For instance, "kenaikan" was converted to "naik," while "berjualan" was converted to "jual."

2.4 Labeling

The labeling process in this study was conducted manually (manual annotation). This method is widely used in sentiment analysis because it can produce more accurate and valid labels compared to automated methods. Research by van Atteveldt, van der Velden, and Boukes (2021) shows that the best performance in sentiment analysis is still achieved through human annotation.

2.5 Feature Extraction

After the preprocessing stage, the text data were transformed into a numerical representation using the term frequency-inverse document frequency (TF-IDF) method, which measures the importance of a word based on its frequency and distribution across documents (Manning et al., 2008).

Feature extraction was performed by considering both single words and word combinations while limiting the number of features to maintain a balance between information richness and model complexity. The feature representation is built on the training data and then applied to the test data to prevent data leakage, resulting in a numerical matrix that is used as an input for the model (Lahkar & Singh, 2022).

The TF (term frequency) Formula 1 is as follows:

$$tf_{t,d} = thenumberofoccurrencesoftermtindocumentd \quad (1)$$

Description:

- t = term/word
- d = dokument

- *tft,d*= number of occurrences of term *t* in document *d*

The inverse document frequency (IDF) Formula 2 is:

$$idf_t = log\left(\frac{N}{df_t}\right) \quad (2)$$

Description:

- *idft* = inverse document frequency weight
- *N*= total number of documents
- *dft*= number of documents containing term *t*

2.6 Naïve Bayes Classifier

Naïve Bayes is a probabilistic classification method based on Bayes' theorem, using a generative approach that models the posterior probability of a class based on observed features. This model Formula 3 assumes that each feature is conditionally independent given the class, thereby simplifying computations without significantly reducing performance (Webb, 2010).

$$P\left(XC\right) = \prod_{i=1}^{n} P(x_i|C) \quad (3)$$

Description:

- *P(C|X)* = probability of class *C* given data *X* (posterior)
- *(PX|C)* = probability of data *X* given class *C*
- *P(C)* = prior probability of class *C*
- *P(X)* = total probability of the data

2.7 Support Vector Machine Classifier

Support Vector Machine (SVM) is a supervised learning algorithm that works by finding an optimal separating hyperplane to divide data into two or more classes. The goal of SVM is to maximize the margin or the distance between the separating hyperplane and the nearest data points from each class. SVM performs well on large datasets, such as news text, and is often considered more accurate than simple probabilistic-based models Formula 4 (Fung & Mangasarian, 2002).

$$w.x + b = 0 \quad (4)$$

Description:

- *w* = weight vector that determines the orientation of the line/hyperplane
- *x* = input feature (data to be classified)
- *b* = bias (the shift of the line from the origin)

2.8 Model Evaluation Techniques

Model evaluation is an important step in assessing the performance of a machine learning model. One commonly used evaluation tool is the confusion matrix, which displays the performance of a

classification model by comparing the model's predictions with the actual labels. The confusion matrix consists of four main elements: True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN). Various evaluation metrics can be calculated based on this confusion matrix, such as accuracy, precision, recall, and F1-score (Ahludzikri et al., 2025).

2.9 Related Research

Fachreza et al. (2023) focus on the sentiment classification of Twitter users regarding Indonesia capital relocation policy using the Gaussian Naïve Bayes method. The dataset consists of 500 manually labeled tweets categorized into supportive and non-supportive sentiments. The research applies comprehensive text preprocessing and TF-IDF weighting before classification. The results indicate that the model achieves optimal performance at 80% feature usage, with 82.0% accuracy, 76.9% precision, and 100% recall.

Anam et al. (2023) analyzed public sentiment on online learning using a combination of lexicon-based methods and Support Vector Machine (SVM). A total of 1,225 Indonesian tweets were processed using preprocessing techniques, TF-IDF weighting, and SMOTE to address data imbalance. The findings show that integrating lexicon-based labeling (InSet Lexicon) improves model performance, increasing accuracy from 90% to 95%.

Millennianita et al. (2024) compared the Naïve Bayes Classifier (NBC) and Support Vector Machine (SVM) for sentiment classification on Twitter related to the Kim Garam bullying case. Using 1,150 tweets and applying preprocessing, TF-IDF, and TextBlob-based labeling, the results demonstrated that SVM outperformed NBC, achieving the highest accuracy of 97.8% compared to 93.9% under an 80:20 data split.

Saddam et al. (2023) examined public sentiment on flood disaster management in Jakarta using a support vector machine (SVM). The dataset consisted of 28,778 tweets collected via SNScrape and processed through extensive preprocessing and TF-IDF feature extraction. Using 10-fold cross-validation, the model achieved strong performance with 88.6% accuracy, 88.6% precision, and 89.4% recall.

Darmawan (2024) conducted sentiment analysis on the stock acquisition of Tokopedia by TikTok using the Naïve Bayes method based on YouTube comments. This study employed preprocessing stages, TF-IDF weighting, and evaluation using k-fold cross-validation and confusion matrices. The results showed that the best accuracy reached 80.51% when negation words were retained. Additionally, the study found that differences in negation usage did not significantly affect the classification results, and that the ratio of training and testing data influenced the model performance.

Leandro and Fianty (2025) conducted sentiment analysis on TikTok application reviews obtained from the Google Play Store. The data were collected using web scraping techniques and processed through preprocessing stages, including data cleaning, case folding, tokenization, stopword removal, and stemming. Feature extraction was performed using the TF-IDF method, and the dataset was divided into training and testing data with an 80:20 ratio. The study applied Naïve Bayes and Support Vector Machine (SVM) algorithms for classification. The results showed that SVM achieved better performance with an accuracy of 88.76%, compared to Naïve Bayes, which achieved an accuracy of approximately 84%.

Sela and Sonita (2025) also highlighted the importance of preprocessing and feature extraction in improving model performance. The results indicated that text preprocessing steps were effective in reducing noise and enhancing data consistency, while TF-IDF contributed to better feature weighting. Additionally, the evaluation using metrics such as accuracy, precision, recall, F1-score, and AUC demonstrated that SVM provided more reliable classification results compared to Naïve Bayes, particularly in distinguishing between sentiment classes.

2.10 Hypothesis Development

Based on the theoretical framework and previous studies, the following hypotheses are proposed.

H_1: Geopolitical news sentiment derived from YouTube comments significantly affects IHSG volatility during the study period

H_2: SVM outperforms Naïve Bayes in classifying geopolitical-related sentiments in YouTube comments

3. Research Methodology

3.1 Research Flow Diagram

The research flow diagram illustrates the sequence of stages carried out in this study, starting from data collection to model evaluation. Each stage was systematically designed to ensure the accuracy and effectiveness of the sentiment analysis process.

The overall research stages are shown in Figure 1.

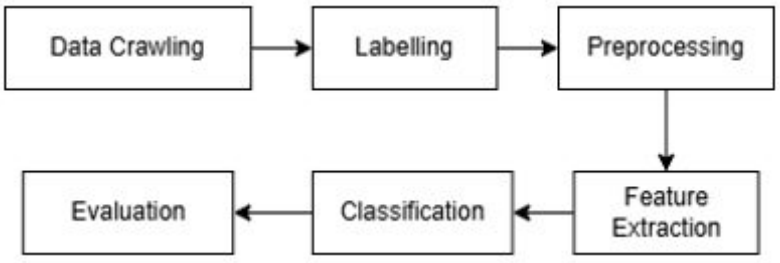


Figure 1. Research Stages

This study utilizes YouTube comments as the primary data source. The data were collected from February to April 2026, following the escalation of geopolitical tensions marked by the disruption of the Strait of Hormuz and the subsequent decline in the IHSG. These events are treated as real-world occurrences during the study period, and the collected data capture public responses reflected in online discussions. This study adopts an observational approach, analyzing sentiment from real-world data collected during the study period without incorporating predictive or simulation-based modeling.

3.1.1 Data Crawling

Data were collected from YouTube comments on videos related to geopolitical news and financial market discussions published by KompasTV and CNN Indonesia. All comments within the defined study period (February–April 2026) were included for analysis. No keyword-based filtering or time-of-day constraints were applied, as the dataset was constructed based on relevant video selection during the observed geopolitical event period. Duplicate and irrelevant comments were removed during the preprocessing to ensure data quality. The stage is the initial phase in which the dataset used for subsequent processes is collected. The collected data include video titles, channel names, and user comments obtained from YouTube, specifically from the KompasTV and CNN Indonesia channels, which serve as the main source for analyzing public sentiment toward market conditions during the conflict period. The total raw data obtained in this study amounted to 1,037 entries.

Next, a duplicate removal process was performed based on the comment content to avoid data redundancy. After this process, the dataset was reduced to 1,020 entries, which were used for the next stage of analysis. The results of the data crawling process are presented to illustrate the structure and characteristics of the collected datasets. The data consist of video titles, channel names, and user comments obtained from the KompasTV and CNN Indonesia YouTube channels, which serve as the primary source for sentiment analysis. A sample of the dataset after preprocessing and duplicate removal is presented in Figure 2.

	A	B	C
1	Judul Video	Komentar	Label
2	CNN Indonesia - BREAKING	HAPUS SEMUA DAFTAR SAHAM SAMPAH!	Negatif
3	CNN Indonesia - BREAKING	PERKETAT ATURAN IPO!	Positif
4	CNN Indonesia - BREAKING	JANGAN JADIKAN SAHAM CUMA ALAT BUAT NGERAMPOK DUIT RAKYAT!	Negatif
5	CNN Indonesia - BREAKING	GMN NEGARA MAJU KALO INVESTOR DIKADALIN	Negatif
6	CNN Indonesia - BREAKING	IHSG perusahaannnya gak ada yg jelas.... Fundamentalnya gak ada cuman dipompom sama band	Netral
7	CNN Indonesia - BREAKING	Hahahaha...negara sarang penyamun, di smua lini semua tukang tipu tukang sulap	Negatif
8	CNN Indonesia - BREAKING	URUS PINJOL JUGA NOH.!!!!	Netral
9	CNN Indonesia - BREAKING	JANGAN RUSAK MENTAL WARGA INDONESIA.!!!"	Negatif
10	CNN Indonesia - BREAKING	Putus2 min reupload lagi yg clear napa	Netral
11	CNN Indonesia - BREAKING	Pejabat OJK kapan?😂	Netral
12	CNN Indonesia - BREAKING	welcome resesi .. pada cabut sebelum lebih parah lagi ..	Negatif
13	CNN Indonesia - BREAKING	OJK tidak merasa perlu bertanggung jawab? Dasar ...	Netral
14	CNN Indonesia - BREAKING	Jelek banget sih vidio Lu putus²	Negatif
15	CNN Indonesia - BREAKING	Tugas mereka dan bandar ngarungin ritel Indonesia hahaha,mau mencoba ngarungin asing wkwk	Netral
16	CNN Indonesia - BREAKING	Wajah mereka perlihatkan seperti orang tidak bersalah atas anjlok ihsg, orang busuk jadi penguru	Negatif
17	CNN Indonesia - BREAKING	Hrs di periksa kpk seluluruh direksi.di duga ada manipulasi harga saham" konglo secara tidak waja	Negatif
18	CNN Indonesia - BREAKING	Budek...emang	Netral
19	CNN Indonesia - BREAKING	Negara jawa semuanya ga ada yg beres 😂	Netral
20	CNN Indonesia - BREAKING	Tentunya investor harus tenang... dan rasional..	Netral

Figure 2. Data Crawling Results

Furthermore, the distribution of the collected data shows that 559 entries were obtained from the CNN Indonesia channel, whereas the remaining 461 entries were collected from the KompasTV channel. This distribution reflects the contribution of each source to providing data for sentiment analysis.

3.1.2 Labeling

The labeling process in Table 1 was conducted manually by human annotators based on the sentiment interpretation of each comment. The collected data were then categorized into three sentiment classes: positive, negative, and neutral.

Table 1. Labeling Examples

Comment	Label
HAPUS SEMUA DAFTAR SAHAM SAMPAH!	Negative
Buat jangka panjang oke banget sih,,, cuman nyangkut modal nya 😂	Positive
ada apa nih	Neutral

3.1.3 Preprocessing

At this stage, the data undergo several processes, namely case folding, text cleaning, tokenization, normalization, stopword removal, and stemming.

Table 2. Preprocessing Results

Comment	Filtering Result
HAPUS SEMUA DAFTAR SAHAM SAMPAH!	hapus semua daftar saham sampah
OJK tidak merasa perlu bertanggung jawab? Dasar ...	ojk tidak rasa perlu tanggung jawab dasar
Negara jawa semuanya ga ada yg beres 😂	negara jawa semua tidak beres

Based on Table 2, comment data that initially contained capital letters, punctuation, emoticons, and informal words were transformed into a simpler and more consistent form. This process includes case folding, text cleaning, word normalization, stop word removal, and stemming, resulting in cleaner data that focus on the main information. Tokenization is performed by breaking the text into word units (tokens). For example, the text "hapus semua daftar saham sampah" is converted into "hapus," "semua," "daftar," "saham," and "sampah." The tokenized output is subsequently used as input for feature extraction using TF-IDF representation

4. Results and Discussions

This section describes the implementation of the Naïve Bayes and SVM algorithms on the dataset. The process begins with labeling the dataset, which consists of 1,020 entries. After going through the preprocessing stage, some data became empty due to the text cleaning process and were therefore removed. As a result, the number of data used in the modeling stage was reduced to 998 entries. The dataset was then divided into 798 training data and 200 testing data. The dataset was divided using an approximate 80:20 ratio to ensure sufficient data for training while maintaining a representative portion for testing. This split is commonly used to balance model learning and evaluation reliability.

Before presenting the classification results, model training was conducted using both algorithms with the same dataset and feature extraction method to ensure a fair comparison between Naïve Bayes and SVM. The results of the Naïve Bayes classification are presented to evaluate the performance of the model on the dataset. This evaluation included accuracy, precision, recall, and F1-score, which provided a comprehensive overview of the model's effectiveness in classifying sentiment. A detailed classification report is presented in Figure 3.

```
=== NAIVE BAYES ===
Train Accuracy: 0.8796992481203008
Test Accuracy : 0.81
              precision    recall  f1-score   support

     Negatif       0.67      0.24      0.35        25
      Netral       0.81      0.98      0.89       146
     Positif       0.93      0.45      0.60        29

    accuracy                           0.81       200
   macro avg       0.80      0.56      0.61       200
weighted avg       0.81      0.81      0.78       200
```

Figure 3. Naïve Bayes Classification Report

The classification results using the Support Vector Machine (SVM) model are presented to demonstrate its performance in analyzing sentiment within the dataset. The evaluation highlights the model's ability to distinguish between sentiment classes using key metrics such as precision, recall, and F1-score, along with overall accuracy. These results reflect how well the SVM model generalizes the data and handles the classification tasks. The complete performance report is shown in Figure 4.

```
=== SVM ===
Train Accuracy: 0.9523809523809523
Test Accuracy : 0.825
              precision    recall  f1-score   support

     Negatif       0.50      0.52      0.51        25
      Netral       0.88      0.90      0.89       146
     Positif       0.83      0.69      0.75        29

    accuracy                           0.82       200
   macro avg       0.74      0.70      0.72       200
weighted avg       0.83      0.82      0.82       200
```

Figure 4. SVM Classification Report

Table 3. Accuracy Comparison

Model	Train Accuracy (%)	Test Accuracy (%)	Gap (%)
Naïve Bayes	87.9	81.0	6.9

Model	Train Accuracy (%)	Test Accuracy (%)	Gap (%)
SVM	95.2	82.5	12.7

Based on Table 3, Naïve Bayes shows a relatively small gap between training and testing accuracy (6.9 %), indicating that the model has good generalization ability and minimal overfitting. In contrast, SVM has a larger gap (12.7%), suggesting an indication of overfitting, although it achieves higher testing accuracy. Therefore, Naïve Bayes is more stable, whereas SVM performs better in terms of accuracy. This behavior indicates that although SVM is capable of capturing more complex patterns in the training data, it is also more prone to overfitting compared to Naïve Bayes. Conversely, Naïve Bayes demonstrates more consistent performance between training and testing, suggesting better generalization on unseen data.

Table 4. Comparison of Precision, Recall, dan F1-Score

Model	Kelas	Precision	Recall	F1-Score
Naïve Bayes	Negatif	0.67	0.24	0.35
Naïve Bayes	Netral	0.81	0.98	0.89
Naïve Bayes	Positif	0.93	0.45	0.60
SVM	Negatif	0.50	0.52	0.51
SVM	Netral	0.88	0.90	0.89
SVM	Positif	0.83	0.69	0.75

Based on Table 4, Naïve Bayes demonstrates a very strong performance in the neutral class, achieving a recall of 0.98, but shows a significant decline in the negative and positive classes. This indicates that the model tends to be biased toward the majority class. The observed performance differences are likely influenced by the imbalanced distribution of the dataset, where the neutral class dominates. As a result, Naïve Bayes tends to favor the majority class, whereas SVM shows better capability in handling minority classes.

In contrast, the SVM model exhibits a more balanced performance across all classes. In the negative class, SVM achieves a recall of 0.52, which is higher than that of Naïve Bayes at 0.24. A similar pattern is observed in the positive class, where SVM attains a recall of 0.69, compared with 0.45 for Naïve Bayes. This improvement indicates that SVM is more sensitive to recognizing patterns in minority classes, resulting in a more balanced classification performance. Therefore, it can be concluded that SVM is more effective in handling minority classes, whereas Naïve Bayes tends to perform better on the dominant class.

Table 5. Macro Average Comparison

Model	Precision	Recall	F1-Score
Naïve Bayes	0.80	0.56	0.61
SVM	0.74	0.70	0.72

As shown in Table 5, SVM has a higher macro-average F1-score than Naïve Bayes, with 0.72 compared to 0.61. This indicates that SVM demonstrates more consistent performance across all classes without being overly influenced by the dominance of a particular class. In contrast, Naïve Bayes tends to produce uneven performance across different classes.

Overall, the findings of this study highlight important insights regarding both the performance of classification models and the relationship between geopolitical sentiment and market volatility. Based on the comparative evaluation, the SVM model demonstrates superior performance compared to Naïve Bayes in classifying geopolitical-related sentiment from YouTube comments. This is reflected in its higher test accuracy and macro-average F1-score, as well as its more balanced performance

across all sentiment classes, particularly in handling minority classes. Although Naïve Bayes shows more stable generalization with a smaller gap between training and testing accuracy, SVM provides more reliable classification outcomes in the context of imbalanced data. Therefore, Hypothesis 2 (H_2) is supported.

In contrast, the results indicate that the sentiment derived from YouTube comments does not significantly affect IHSG volatility during the observed period. Although this study specifically focuses on comments discussing the decline of the IHSG due to geopolitical conflicts, the sentiment analysis results are not sufficient to empirically validate this relationship. This suggests that public sentiment expressed in YouTube comments may not directly translate into measurable market reactions or that its influence is relatively weak compared to other dominant macroeconomic and financial factors. Therefore, Hypothesis 1 (H_1) is not supported.

Several factors may explain this outcome. First, the data source consists of user-generated content on YouTube, which is often informal, subjective, and not necessarily representative of broader market sentiment. Such data may contain noise, sarcasm, or irrelevant opinions that reduce the accuracy of sentiment representation. Second, the labeling process was conducted manually, introducing a degree of subjectivity that may affect the consistency and reliability of the classification results. Third, market volatility, such as the IHSG, is influenced by complex and multifactorial dynamics, including global economic conditions, policy changes, and institutional investor behavior, which are not fully captured by social media sentiment alone. These findings indicate that sentiment derived from YouTube comments has limited capability in explaining stock market volatility, particularly in the context of IHSG movements influenced by geopolitical issues.

5. Conclusions

5.1 Conclusion

Based on the results of the study, it can be concluded that the objective of comparing the performance of the Naïve Bayes and Support Vector Machine (SVM) algorithms in the sentiment analysis of geopolitical-related YouTube comments has been successfully achieved. The experimental results show that SVM provides a better overall classification performance compared to Naïve Bayes, particularly in handling minority classes and producing more balanced evaluation metrics. Therefore, Hypothesis 2 (H_2) is supported.

In contrast, the findings indicate that sentiment derived from YouTube comments does not have a significant effect on IHSG volatility during the study period. This suggests that the sentiment expressed in user-generated content may not directly reflect or influence stock market movements. Therefore, Hypothesis 1 (H_1) is not supported. Overall, while SVM is more effective for sentiment classification tasks, the use of YouTube-based sentiment alone is insufficient to explain fluctuations in IHSG volatility.

5.2 Research Limitations

This study had several limitations that may have affected the results. First, the dataset used was relatively limited, consisting of 998 data points, which may have influenced the generalization capability of the model. Second, the imbalanced data distribution, with a dominance of the neutral class, causes the model to be biased and less optimal in recognizing minority classes. Third, the labeling process was conducted manually, which may have introduced subjectivity and affected the consistency of the classification results. In addition, the data source was limited to comments from YouTube, which may not fully represent public perception across various platforms and may contain noise, informal language, and irrelevant information.

5.3 Suggestions and Directions for Future Research

Based on these limitations, future research should use larger and more diverse datasets by incorporating multiple data sources, such as Twitter, online news, and discussion forums, to improve

data representativeness. Furthermore, techniques for handling imbalanced data, such as resampling or class-weighting methods, can be applied to enhance the performance of minority classes. Future studies may also explore more advanced classification methods, such as ensemble learning or deep learning, and perform model parameter optimization to improve accuracy. Additionally, integrating sentiment analysis with other economic variables can provide a more comprehensive understanding of the relationship between public sentiment and stock market volatility.

Acknowledgement
The author would like to express sincere gratitude to the supervisor for the guidance, support, and valuable feedback throughout the research. Appreciation is also extended to all parties who contributed to this study.

Author Contributions
MCAA contributed to all aspects of the study, including conceptualization, data collection, preprocessing, modeling, analysis, and manuscript writing. IA contributed to the supervision, guidance, and review of the manuscript.

References
Ahludzikri, F., Herwanto, R., Rz, A. A., Agus, I., & Irianto, S. Y. (2025). Web-based research article classification using the random forest algorithm. *Jurnal Ilmu Siber dan Teknologi Digital*, *4*(1), 15-31. https://doi.org/10.35912/jisted.v4i1.5547

Allahyari, M., Pouriyeh, S., Assefi, M., Safaei, S., Trippe, E. D., Gutierrez, J. B., & Kochut, K. (2017). A brief survey of text mining: Classification, clustering and extraction techniques. *arXiv preprint*. https://doi.org/10.48550/arXiv.1707.02919

Anam, M. K., Fitri, T. A., Agustin, A., Lusiana, L., Firdaus, M. B., & Nurhuda, A. T. (2023). Sentiment analysis for online learning using the lexicon-based method and the support vector machine algorithm. *ILKOM Jurnal Ilmiah*, *15*(2). https://doi.org/10.33096/ilkom.v15i2.1590.290-302

Darmawan, N. I. (2024). Analisis sentimen terhadap akuisisi saham tokopedia oleh tiktok menggunakan naïve bayes berdasarkan komentar youtube jurnal pengembangan teknologi informasi dan ilmu komputer. Retrieved from https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/13896

Emmanuel, P. (2026). The impact of the us–Israel–Iran war on global corporate profitability: Evidence from energy, transportation, and supply chain disruptions in 2026. https://doi.org/10.13140/RG.2.2.35575.59043

Fachreza, M. R. D., Suhartono, & Yaqin, M. A. (2023). Klasifikasi sentimen masyarakat terhadap proses pemindahan Ibu Kota Negara (IKN) Indonesia pada media sosial twitter menggunakan metode naïve bayes. *JISKA (Jurnal Informatika Sunan Kalijaga)*, *8*(3), 243-251. https://doi.org/10.14421/jiska.2023.8.3.243-251

Fung, G., & Mangasarian, O. L. (2002). Incremental support vector machine classification. Paper presented at Proceedings of the 2002 SIAM International Conference on Data Mining (SDM).

Hafifa, N., Maisaroh, S., Umaira, P. N., & Pangestoeti, W. (2025). Dampak penurunan IHSG terhadap stabilitas ekonomi publik indonesia: Analisis penyebab, implikasi, dan strategi kebijakan. *Socius: Jurnal Penelitian Ilmu-Ilmu Sosial*, *2*(12). https://doi.org/10.5281/zenodo.15703818

Han, L., Ning, D., & Xin, S. (2025). Geopolitical risk and economic growth: Evidence from ASEAN countries. *Finance Research Letters*, *88*, 109102. https://doi.org/10.1016/j.frl.2025.109102

Lahkar, B., & Singh, J. (2022). Twitter text sentiment analysis: A comparative study on unigram and bigram feature extractions international research journal of engineering and technology. Retrieved from https://www.irjet.net/archives/V9/i8/IRJET-V9I870.pdf

Leandro, J. O., & Fianty, M. I. (2025). Evaluation of sentiment analysis methods for social media applications: A comparison of support vector machines and naïve bayes. *JOIV: International Journal on Informatics Visualization*, *9*(2), 796-807. https://doi.org/10.62527/joiv.9.2.2905

Liu, B. (2012). Sentiment Analysis and Opinion Mining. *Morgan & Claypool*. https://doi.org/10.1007/978-3-031-02145-9

Manning, C. D., Raghavan, P., & Schütze, H. (2008). *Introduction to information retrieval*.

Millennianita, F., Athiyah, U., & Muhammad, A. W. (2024). Comparison of naive bayes classifier and support vector machine methods for sentiment classification of responses to bullying cases on twitter. *Journal of Mechatronics and Artificial Intelligence*, *1*(1), 11-26. https://doi.org/10.17509/jmai.v1i1.69959

Noviyanto, H., Burhanuddin, A., & Abdullah, N. (2025). Security and sovereignty in the strait of hormuz: The Iran-USA Conflict. *Journal of Peace, Security and Democracy*, *1*(1), 87-105. https://doi.org/10.63280/jpsd.v1i1.42245

Qin, L., Chen, Q., Feng, X., Wu, Y., Zhang, Y., Li, Y., Li, M., Che, W., & Yu, P. S. (2026). Large Language Models Meet NLP: A Survey. *Frontiers of Computer Science*, *20*(11), 2011361. https://doi.org/10.1007/s11704-025-50472-3

Roy, A., Ghosh, S., Ghosh, K., & Ghosh, S. (2021). An unsupervised normalization algorithm for noisy text: A case study for information retrieval and stance detection. *Journal of Data and Information Quality (JDIQ)*, *13*(3), 1-25. https://doi.org/10.1145/3418036

Saddam, M. A., Dewantara, E. K., & Solichin, A. (2023). Sentiment analysis of flood disaster management in jakarta on twitter using support vector machines. *Sinkron :Jurnal dan Penelitian Teknik Informatika*, *7*(1), 470-479. https://doi.org/10.33395/sinkron.v8i1.12063

Sela, T., & Sonita, A. (2025). Comparison of naive bayes and support vector machine algorithms in sentiment analysis. *Jutisi : Jurnal Ilmiah Teknik Informatika dan Sistem Informasi*, *14*(2), 973. https://doi.org/10.35889/jutisi.v14i2.2727

Sonkusale, L., Chaturvedi, K. K., Sharma, A., Lal, S. B., Farooqi, M. S., Lama, A., Mishra, D. C., Joshi, P., & Kumar, M. (2023). Evaluating text preprocessing methods for discovering quality topics to improve the information retrieval mechanism acta scientific computer sciences volume. Retrieved from https://www.researchgate.net/publication/389503309

Tala, F. (2003). A study of stemming effects on information retrieval in Bahasa Indonesia. Retrieved from https://www.researchgate.net/publication/2919045

Teller, V. (2000). Speech and language processing: An introduction to natural language processing. *Computational Linguistics, and Speech Recognition*. https://doi.org/10.1162/089120100750105975

Tong, C., Huang, Z., Wang, T., & Zhang, C. (2023). The effects of economic uncertainty on financial volatility: A comprehensive investigation. *Journal of Empirical Finance, 73*, *73*, 369-389. https://doi.org/10.1016/j.jempfin.2023.08.004

van Atteveldt, W., van der Velden, M. A. C. G., & Boukes, M. (2021). The validity of sentiment analysis: Comparing manual annotation, crowd-coding, dictionary approaches, and machine learning algorithms. *Communication Methods and Measures*, *15*(2), 121-140. https://doi.org/10.1080/19312458.2020.1869198

Vijayarani, S., Ilamathi, M. J., & Nithya, M. (2015). Preprocessing techniques for text mining-an overview. *International Journal of Computer Science & Communication Networks*, *5*(1), 7-16.

Webb, G. I. (2010). Naïve Bayes. In C. Sammut & G. I. Webb (Eds.), *In C. Sammut & G. I. Webb (Eds.), Encyclopedia of Machine Learning (pp. 713–714). Springer US* (pp. 713-714).

Webster, J. J., & Kit, C. (1992). Tokenization as the initial phase in NLP. Paper presented at COLING 1992 volume 4: The 14th international conference on computational linguistics.

Wijaya, T. S. J., & Agustin, S. (2015). Faktor-faktor yang mempengaruhi nilai ihsg yang terdaftar di bursa efek Indonesia. *Jurnal Ilmu dan Riset Manajemen (JIRM)*.

Zhang, Y. (2021). Sales forecasting of promotion activities based on the cross-industry standard process for data mining of e-commerce promotional information and support vector regression. *Journal of Computers*, *32*(1), 212-225. https://doi.org/10.3966/199115992021023201018

Zhao, C., Kang, L., Xi, X., Du, S., & Li, J. (2025). Investor sentiment and stock market volatility: Exploring the relationship using sentiment analysis of stock bar comments. *Finance Research Open*, *1*(3), 100016. https://doi.org/10.1016/j.finr.2025.100016