Jurnal Teknik Informatika Vol. 18 No. 2, October 2025 (257-269)
ISSN: p-ISSN 1979-9160 (Print)| e-ISSN 2549-7901 (Online)
DOI: https://doi.org/10.15408/jti.v18i2.46447

JURNAL TEKNIK INFORMATIKA

Homepage : http://journal.uinjkt.ac.id/index.php/ti

Public Sentiment and GoTo Stock Price Movement in Indonesia: A Null-Relationship Study Using Naïve Bayes and Non-Parametric Measures

Dita Pramesti[1*], Hanif Fakhrurroja[2], Rahma Karina M.[3]

[1,2,3]Information System Study Program, School of Industrial Engineering, Telkom University
[1,2,3]Main Campus (Bandung Campus), Jl. Telekomunikasi No.1, Bandung 40257, West Java, Indonesia

Article:
Accepted: June 30, 2025
Revised: April 03, 2025
Issued: October 30, 2025

***Correspondence Address:**
ditapramesti@telkomuniversity.ac.id

ABSTRACT

The expiration of the lock-up period for PT GoTo Gojek Tokopedia Tbk's shares led to a sharp stock price decline and public discourse on Twitter. This study aims to examine the statistical relationship between public sentiment and GoTo's stock price movement in Indonesia. Tweets were classified into positive or negative sentiment using the Naïve Bayes classifier, selected for its computational efficiency on large-scale textual data. The model achieved 70% accuracy, with a precision of 82% and F1-score of 75%. The sentiment polarity was then compared with stock trends across 39 distinct trading periods using four non-parametric statistical tests: Chi-Square ($p = 0.6398$), Cramer's V (0.014), Goodman-Kruskal's Lambda (0.053), and Mann-Whitney U test ($p = 0.8994$). None of these tests showed a statistically significant association between sentiment polarity and stock price movement. These findings highlight that while public sentiment may reflect short-term public interest, it does not reliably capture the market's behavioral dynamics—especially in cases of investor decisions driven by broader macroeconomic or institutional factors. Sentiment data, therefore, should be considered as a complementary, rather than primary indicators in stock price analysis.

Keywords : *naïve bayes; public sentiment; stock price movement; non-parametric statistics; GoTo; twitter.*

Jurnal Teknik Informatika Vol. 18 No. 2, October 2025 (257-269)
ISSN: p-ISSN 1979-9160 (Print)| e-ISSN 2549-7901 (Online)
DOI: https://doi.org/10.15408/jti.v18i2.46447

1. INTRODUCTION

In the current digital era, investment decisions are no longer solely based on financial reports or technical analysis. The dynamics of public opinion widely disseminated through social media also play a significant role in shaping market perceptions [1]. Twitter, widely used in Indonesia, serves as a platform for investors and the public to express views on corporate performance [2], [3]. This dynamic was evident in the case of GoTo, a tech company formed by the merger of Gojek and Tokopedia on May 17, 2021, creating an integrated digital ecosystem. GoTo conducted its IPO in April 2022. Following the end of the lock-up period in late November, the stock experienced a sharp decline, prompting strong public reactions on Twitter [4], [5], [6].

Twitter offers a rich source of emotional and perceptual data for tracking public responses to stock-related events through sentiment analysis and Natural Language Processing (NLP) [7], [8], [9]. Evolving sentiment analysis methods have shown that tweet volume, user influence, and the integration of Twitter data with machine learning models—such as Support Vector Machine (SVM) and Logistic Regression—can enhance stock market prediction performance [10], [11], [12]. More recent approaches integrate aspect-based sentiment analysis (ABSA) with statistical tools—such as Pearson correlation, Granger causality, and uncertainty coefficient—to derive interpretable sentiment-price relationships [13]. Some models even achieve strong predictive power without explicit sentiment labels, as demonstrated in domain-specific forums [14].

Advanced techniques like neural networks, XGBoost, LSTM, and Spearman correlation have shown improved performance in modeling the nonlinear and time-sensitive relationship between sentiment and stock prices [15], [16], [17]. According to behavioral finance, investor sentiment plays a central role in driving short-term price volatility. Psychological biases such as herding behavior, overreaction, and narrative framing often cause markets to move independently of financial fundamentals [18].

Prior studies indicate that public sentiment from social media can influence stock movements, especially when integrated with machine learning models. However, most focus on forecasting without statistically evaluating the relationship between sentiment and market trends. Few have applied non-parametric methods or examined sentiment across distinct price periods.

Indonesia's capital market, largely driven by retail investors active on platforms like Twitter, exhibits a unique behavioral dynamic—where sentiment tends to be more reactive and less tied to fundamentals than in developed markets.

To address this gap, this study seeks to answer the following question: Does public sentiment on Twitter significantly relate to GoTo's short-term stock price movements in Indonesia? Unlike predictive approaches, this study adopts a correlational and distributional perspective using four non-parametric statistical tests—Chi-Square, Cramer's V, Goodman-Kruskal Lambda, and Mann-Whitney U—on sentiment represented both categorically (positive/negative) and numerically (Net Sentiment Score). The findings provide early empirical evidence of social media sentiment as a contextual, non-traditional indicator for investment decisions in emerging markets.

Based on the formulated research question and the need to statistically evaluate the relationship between sentiment and market movement, this study hypothesizes that there is no significant relationship between public sentiment and the direction of GoTo's stock price trend. This null hypothesis reflects the assumption that sentiment, as expressed on Twitter, does not meaningfully align with stock price movements. Alternatively, if a statistically significant relationship is observed, the hypothesis would be rejected in favor of the proposition that public sentiment does indeed relate to stock trend direction.

2. METHODS

This study adopts the Knowledge Discovery in Databases (KDD) methodology as the primary framework to systematically extract insights from large-scale data. KDD offers a structured process for transforming raw data into interpretable patterns, making it suitable for data mining tasks. As shown in Figure 1, the workflow involves collecting and analyzing sentiment and stock data to evaluate the

Jurnal Teknik Informatika Vol. 18 No. 2, October 2025 (257-269)
ISSN: p-ISSN 1979-9160 (Print)| e-ISSN 2549-7901 (Online)
DOI: https://doi.org/10.15408/jti.v18i2.46447

relationship between public opinion and GoTo's stock movements.

Data were sourced from two platforms: Twitter and Yahoo Finance. From Twitter, relevant tweet texts and metadata (e.g., timestamp, engagement) were collected for sentiment analysis. Concurrently, stock data from Yahoo Finance were used to determine the direction of price trends, categorized as either upward or downward.

Sentiment classification was performed using the Naive Bayes algorithm, producing two types of outputs: (1) categorical labels (positive/negative), and (2) a numerical Net Sentiment Score (NSS), ranging from -1 to 1, reflecting the overall sentiment strength.

To investigate the relationship between sentiment and stock price trends, two non-parametric statistical approaches are utilized. The Chi-Square test, Cramer's V, and Goodman-Kruskal Lambda are applied to evaluate the association between sentiment categories (positive/negative) and stock trend direction (upward/downward). In parallel, the Mann-Whitney U test is employed to determine whether a significant difference exists in the distribution of Net Sentiment Score (NSS) between upward and downward trend groups. Overall, this workflow demonstrates that the study not only utilizes text classification techniques but also integrates them with statistical evaluation methods to investigate the potential role of social media sentiment as a non-traditional indicator in investment decision-making.

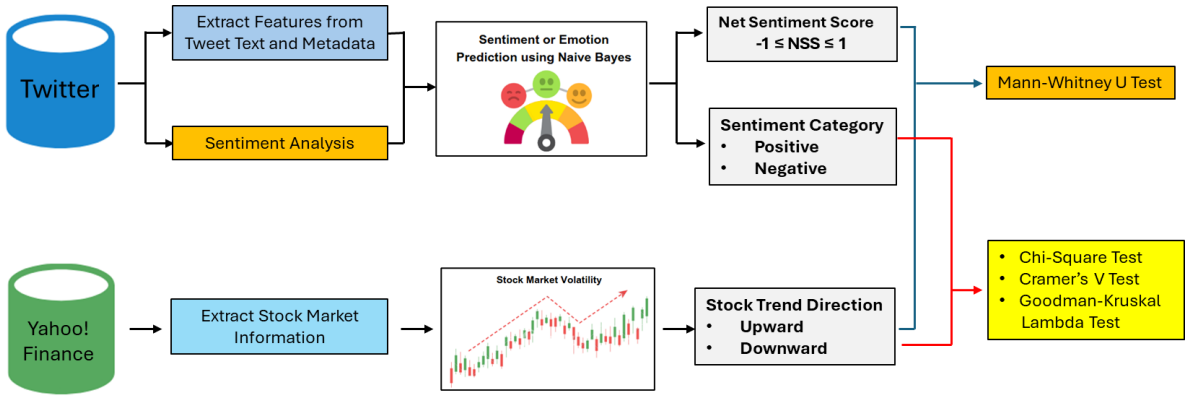


Figure 1. *Problem-solving workflow to analyze the relationship between twitter sentiment and stock trend direction (adapted and modified from* [10]*)*

2.1. Data Selection

This study utilizes two datasets: comment data, hereinafter referred to as tweets, and stock price data of the GoTo company. Tweets were collected using the Twitter API through keyword-based crawling, employing terms such as "Goto", "Gojek", "Tokopedia", "Gojek Tokopedia goto", "*Saham* Goto", "*Saham* Gojek", and "*Saham* Tokopedia". The dataset spans from December 2022 to March 2023.

Stock data for GoTo (GOTO.JK) were sourced from Yahoo Finance for the same period, including variables such as date, open, high, low, close, adjusted close, and volume. The highest price recorded was IDR 141 (Dec 1, 2022), and the lowest was IDR 83 (Dec 26, 2022), with a sharp decline in December followed by notable gains in January–February. A sample is presented in Table 1.

Table 1. *Sample of GoTo stock price*

Date	Open	High	Low	Close	Adj Close	Volume
1/2/2023	91	94	91	93	93	897258100
1/3/2023	93	97	93	95	95	2761729900
1/4/2023	95	97	93	96	96	1816437300

2.2. Data Preprocessing

Before proceeding to the preprocessing stage, the tweet data were labeled into three categories: positive, negative, and neutral. After labeling, tweets with neutral sentiment were removed from the dataset, leaving only those labeled as positive and negative. Following the removal of neutral labels, a total of 26,840 tweets remained, consisting of 17,221 positive tweets and 9,619 negative tweets.

After the labeling process, preprocessing began with a data cleansing step, which involved removing URLs, usernames,

punctuation marks (such as .,/!:;’”|/><{}[]), hashtags, digits (0–9), icons, whitespace, and images. These entities were deemed unnecessary for the modeling stage. The next step was case folding, which converted uppercase letters in tweets to lowercase.

Preprocessing began with tokenization, followed by normalization using a custom slang dictionary. Stopwords and colloquial terms were removed, then stemming reduced words to their root forms. Emojis and special characters were filtered using regex and emoji libraries. However, the preprocessing pipeline did not address deeper semantic challenges such as sarcasm, irony, or negation detection. The informal and context-rich nature of social media language presents challenges that may affect the accuracy of sentiment classification. These limitations are acknowledged as potential sources of noise in this study.

2.3. Data Transformation

At this stage, Term Frequency-Inverse Document Frequency (TF-IDF) weighting was applied to assess the importance of words in the tweet dataset. This technique evaluates a word’s relevance within a document relative to its frequency across the corpus [19]. Term Frequency (TF), shown in Equation 1, represents the ratio of a word’s occurrence to the total word count in a tweet.

$$\mathrm{TF(i,j)} = \frac{\mathrm{n_{i,j}}}{\sum_{\mathrm{k}} \mathrm{n(k,j)}} \quad (1)$$

As shown in Equation 2, IDF is determined by the ratio between the total number of documents (D) and the document frequency $\mathrm{df_{(d,t)}}$, which represents how many documents include the term *t*.

$$\mathrm{IDF} = \ln\left(\frac{1+\mathrm{D}}{1+\mathrm{df_{(d,t)}}}\right) \quad (2)$$

The TF-IDF score is obtained as the product of TF and IDF. Once the weighting is complete, the tweet data is segmented into training and testing subsets for model development.

2.4. Data Mining

To address class imbalance prior to modeling, the SMOTE-Tomek method was applied. This hybrid technique combines SMOTE (Synthetic Minority Over-sampling Technique) and Tomek links to balance the dataset [20]. SMOTE generates synthetic samples by interpolating between a minority instance and its k-nearest neighbors [21], increasing minority class representation. Tomek links then identify borderline instances from different classes; if certain conditions are met, overlapping majority class samples are removed, refining the class boundary.

Modeling was performed using the Naïve Bayes algorithm, a probabilistic classifier based on Bayes’ theorem that assumes feature independence [22]. The classification process follows Equation 3, where $P(h|X)$ denotes the posterior probability of class h given fetures X, and $P(x_i|h)$ reflects the likelihood of feature x_i in class h**.** The total probability product is shown in Equation 4.

$$P(h|X) = P(h) \cdot \prod_{i=1}^{n} P(x_i|h) \quad (3)$$

$$\prod_{i=1}^{n} P(x_i|h) = P(x_1|h) \cdot P(x_2|h) \cdot \cdots \cdot (x_n|h) \quad (4)$$

While more complex classifiers such as SVM or Random Forest are available, this study deliberately selected Naïve Bayes due to its effectiveness in short-text classification and its suitability for large-scale social media data. Moreover, the focus of this research lies in evaluating the relationship between classified sentiment and stock price trends, rather than optimizing classification performance. Thus, Naïve Bayes serves as a reliable and interpretable baseline model for the analytical goals of this study.

2.5. Evaluation

The model was evaluated using a confusion matrix consisting of True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN). These values were used to calculate standard performance metrics, including accuracy (overall correctness), precision (positive prediction quality), recall (sensitivity), and F1-score (harmonic mean of precision and recall).

Accuracy indicates the proportion of correctly classified instances among all data records, reflecting the model’s overall classification performance. Precision refers to the fraction of predicted positive instances that are actually positive. Recall, on the other hand, measures the proportion of actual positive instances that the model correctly identifies, representing the true positive rate. The F1-score

Jurnal Teknik Informatika Vol. 18 No. 2, October 2025 (257-269)
ISSN: p-ISSN 1979-9160 (Print)| e-ISSN 2549-7901 (Online)
DOI: https://doi.org/10.15408/jti.v18i2.46447

is the harmonic mean of precision and recall, balancing the trade-off between the two metrics. While precision captures the correctness of positive predictions, recall quantifies the model's ability to detect true positives [23].

2.6. Non-Parametric Analysis

Non-parametric tests were selected due to the categorical nature of key variables (sentiment polarity and stock trend direction) and the non-normal distribution of Net Sentiment Scores (NSS). Parametric tests such as t-test or Pearson correlation are not suitable under these conditions, hence non-parametric alternatives were adopted for their robustness in such scenarios.

The Chi-Square test evaluates whether there is a statistically significant association between sentiment polarity (positive/negative) and stock trend direction (up/down), based on the difference between observed and expected frequencies (Equation 5).

$$\chi^2 = \sum_{i=1}^{r}\sum_{j=1}^{k}\frac{(O_{ij}-E_{ij})^2}{E_{ij}} \quad (5)$$

The values r and k represent the number of rows and columns in the contingency table, respectively. Meanwhile O_{ij} and E_{ij} denote the observed and expected frequencies in cell-(i,j). Once the value of χ^2 is obtained, it is compared to the critical value of the Chi-Square distribution at a given degree of freedom and significance level (in this study, α=0.05). If the p-value is below the significance level of 0.05, the association is considered significant [25], [26].

To complement the Chi-Square test, Cramer's V (Equation 6) is used to measure the strength of association [26]. Its value ranges from 0 (no association) to 1 (perfect association), with conventional thresholds categorizing the strength as weak (≥0.1), moderate (≥0.3), or strong (≥0.5) [27].

$$V = \sqrt{\frac{\chi^2}{n\cdot min(k-1,r-1)}} \quad (6)$$

The value of χ^2 is derived from the Chi-Squared test, n represents the grand total of observations, while k and r denote the number of columns and rows, respectively.

In this study, Lambda is used to assess how well public sentiment categories (positive/negative) predict stock price direction (up/down). The test measures the reduction in classification error when the independent variable is known [28], as shown in Equation 7.

$$\lambda(Y|X) = \frac{E_1-E_2}{E_1} \quad (7)$$

E_1 refers to classification errors made without sentiment information, while E_2 represents errors made when sentiment is known. To compare sentiment intensity, the Mann-Whitney U test (Equation 8) is used to assess differences in Net Sentiment Scores (NSS) between uptrend and downtrend periods. This non-parametric test is suitable for comparing two independent groups when normality cannot be assumed [29].

$$U = n_1n_2 + \frac{n_1(n_1+1)}{2} - R_1 \quad (8)$$

The values n_1 and n_2 represent the number of samples in each group, while R_1 denotes the sum of ranks in the first group. The U value is then converted into a p-value, and if the p-value<0.05, it indicates a significant difference between the two sentiment categories.

3. RESULTS AND DISCUSSION

This section presents the research findings, starting from sentiment classification modeling using the Naïve Bayes algorithm to the analysis of the relationship between public sentiment and stock price movements using non-parametric measures.

3.1. Modeling Result

After TF-IDF transformation, the data were split into 70% training and 30% testing. The training set was imbalanced (12,092 positive vs. 6,759 negative tweets), posing a risk of model bias. To mitigate this, the SMOTE-Tomek method was applied to balance the classes and enhance separability by oversampling minority instances and removing overlapping samples.

Post-resampling, each class contained 11,932 tweets, forming a balanced training set of 23,864 samples. The remaining 8,052 tweets constituted the test set used for evaluating the Naïve Bayes classifier.

Jurnal Teknik Informatika Vol. 18 No. 2, October 2025 (257-269)
ISSN: p-ISSN 1979-9160 (Print)| e-ISSN 2549-7901 (Online)
DOI: https://doi.org/10.15408/jti.v18i2.46447

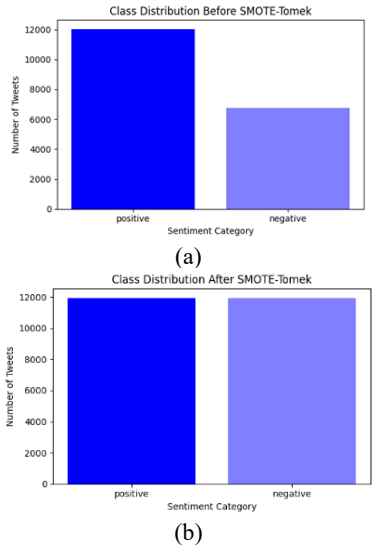


Figure 2. *Comparison of class distribution: (a) before and (b) after applying SMOTE-Tomek*

Figure 2 presents the class distribution before and after applying SMOTE-Tomek resampling. Initially, the dataset exhibited a strong imbalance, which was corrected through hybrid resampling to ensure equal representation of both sentiment classes.

3.2. Performance Evaluation

Model performance was evaluated using the classification results of the Naïve Bayes algorithm. A confusion matrix (Figure 3) was generated from the test data to compute accuracy, precision, recall, and F1-score.

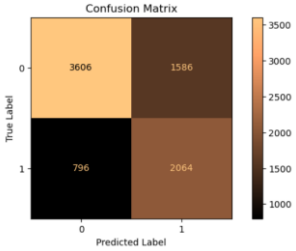


Figure 3. *Heatmap confusion matrix*

Based on the confusion matrix, the model correctly classified 3,606 tweets as positive (TP) and 2,064 as negative (TN). It misclassified 1,586 positive tweets as negative (FN) and 796 negative tweets as positive (FP). Detailed classification metrics derived from these values are presented in Table 2.

Table 2. *Model performance*

Metric	Score
Accuracy	0.70
Precision	0.82
Recall	0.69
F1-Score	0.75

The classification results from the testing data yielded an accuracy of 70%, a precision of 82%, a recall of 69%, and an F1-score of 75%. These results were obtained from the performance evaluation on the testing dataset. Based on the calculations performed, it can be concluded that the accuracy score of 0.70 (or 70%) reflects the level of correctness and reliability of the prediction results in this study. The precision score of 0.82 (or 82%) indicates the degree of consistency between the actual data and the data classified by the system. The recall score of 0.69 (or 69%) shows that the system has a relatively high success rate in identifying relevant instances. Lastly, the F1-score of 0.75 (or 75%) demonstrates that both the precision and recall values obtained are reasonably good. Furthermore, the 70% accuracy achieved by the Naïve Bayes classifier is considered acceptable for baseline sentiment classification tasks, particularly in informal and noisy text environments such as social media. Previous studies in Indonesian-language sentiment analysis using similar models report accuracy levels typically ranging between 65% and 75%, depending on data quality and preprocessing techniques [22].

3.3. Relationship Analysis Between Public Sentiment and Stock Movement

The classification of comment data using the Naïve Bayes Classifier algorithm produced prediction results based on the testing dataset. These results reflect the sentiment classification of the tested comment data, in which each entry was predicted as either positive or negative. Table 3 presents a structured summary of aggregated sentiment and stock trend data, organized across clearly defined “Stock Movement Periods”. Each row represents a contiguous interval, ranging from a single trading day to multiple consecutive days, during which the stock exhibited a consistent directional trend—either an increase (uptrend) or a decrease (downtrend). This event-based aggregation approach was chosen to ensure that public sentiment is evaluated within the actual market context rather than fixed time slices, allowing more interpretable insights.

Jurnal Teknik Informatika Vol. 18 No. 2, October 2025 (257-269)
ISSN: p-ISSN 1979-9160 (Print)| e-ISSN 2549-7901 (Online)
DOI: https://doi.org/10.15408/jti.v18i2.46447

Table 3. *Period-Based aggregation of stock movement and sentiment polarity*

No.	Stock Movement Period	Stock Trend	Positive Tweets	Negative Tweets	NSS	Sentiment Classification
1	01/12/2022 – 12/12/2022	Down	458	493	-0.0368	Negative
2	13/12/2022	Up	49	44	0.0538	Positive
3	14/12/2022	Down	21	19	0.0500	Positive
4	15/12/2022	Up	13	31	-0.4091	Negative
…	…	…	…	…	…	…
38	13/03/2023	Up	124	168	-0.1507	Negative
39	14/03/2023 – 20/03/2023	Down	24	35	-0.1864	Negative

Over the entire observed timeframe, from December 1, 2022, to March 20, 2023, a total of 39 such periods were identified. Each reflects a distinct phase of uninterrupted price movement, serving as the foundation for the sentiment aggregation process. The columns labeled "Positive Tweets" and "Negative Tweets" report the total number of sentiment-classified tweets posted during each period. These figures represent the aggregared expression of public opinion collected via social media. Sentiment classification obtained as the predicted output from a Naïve Bayes classifier trained on labeled data. The model was applied to testing data to generate sentiment labels based on tweet content.

From these aggregated values, the net sentiment score (NSS) was calculated using Equation 9 [30]. The NSS ranges from -1 to +1, where positive values indicate the dominance of positive sentiment, and negative values reflect stronger negative sentiment.

$$NSS = \frac{N_{positive} - N_{negative}}{N_{positive} + N_{negative}} \quad (9)$$

Here, $N_{positive}$ and $N_{negative}$ denote the number of tweets classified as expressing or negative sentiment, respectively. Based on this score, each period was assigned a "Sentiment Classification": positive, if NSS>0; negative, if NSS<0; and neutral, if NSS=0, but no neutral case appeared in this dataset excerpt.

To enable a clearer analysis of sentiment trends relative to stock movement, sentiment was categorically labeled as Positive or Negative based on the Net Sentiment Score (NSS), while stock direction was classified as Uptrend or Downtrend. A contingency table, as illustrated in Table 4, was constructed to summarize the joint frequency distribution, showing the number of periods each sentiment–trend combination occurred.

Table 4. *Contingency table*

		Sentiment Category		Total
		Positive	Negative	
Stock Drection	Up	14	5	19
	Down	16	4	20
Total		30	9	39

As shown in the table, there were 19 periods classified as "uptrend" and 20 as "downtrend", with the sentiment categories distributed accordingly. This contingency table serves as the basis for the statistical tests that follow.

A Chi-squared test of independence was then performed to assess whether there is a significant association between sentiment polarity and stock trend. Based on the contingency table, the Chi-Square test was performed using Equation (9). The test produced a Chi-Square statistic of 0.2190 and a p-value of 0.6398. Since the p-value exceeds the significance level of 0.05, it indicates that there is no statistically significant association between sentiment polarity and stock trend direction.

To measure the strength of any existing relationship, Cramer's V was calculated using Equation 6 and resulted in a value of 0.014, suggesting an extremely weak association between sentiment and movement.

To further evaluate whether sentiment can be used to predict stock movement, the Goodman-Kruskal Lambda coefficient was calculated. In this analysis, the dependent variable is stock trend, and the independent variable is sentiment category. The lambda was calculated in Equation 10.

$$\lambda(\text{stock trend}|\text{sentiment category}) = \frac{19-(14+4)}{19} = 0.053 \quad (10)$$

The lambda value obtained was 0.053, which means that knowing the sentiment category reduces the error in predicting stock movement by only 5.3% compared to simply guessing based on the most frequent outcome.

Jurnal Teknik Informatika Vol. 18 No. 2, October 2025 (257-269)
ISSN: p-ISSN 1979-9160 (Print)| e-ISSN 2549-7901 (Online)
DOI: https://doi.org/10.15408/jti.v18i2.46447

Because the reduction in error is minimal (just 5.3%), this indicates that sentiment does not provide meaningful predictive power in determining whether the stock will move up or down.

As a comparison, a second lambda test was also performed to evaluate the reverse direction: predicting sentiment category based on stock movement. In this case, λ(sentiment category|stock trend)=0.000 indicating that knowing the direction of stock movement does not provide any advantage in predicting whether public sentiment was positive or negative. This asymmetry illustrates that while sentiment provides a very limited signal for predicting stock movement, the reverse is not true: stock direction does not help explain public sentiment in this dataset.

To further explore this relationship, an analysis was conducted to determine whether the Net Sentiment Score (NSS)—a continuous variable derived from the proportion of positive and negative comments—differs significantly between uptrend and downtrend periods. This score provides a more nuanced view of public sentiment intensity over time.

Given the nature of the variables, the Mann-Whitney U test was employed as a non-parametric method suitable for comparing NSS distributions across two independent groups without assuming normality. The test produced a U-statistic of 185.0 with a p-value of 0.8994, indicating no statistically significant difference between sentiment intensity during rising and falling stock periods. Since the p-value exceeds the conventional $\alpha = 0.05$ threshold, we fail to reject the null hypothesis that both distributions originate from the same population.

A boxplot in Figure 4 was used to visualize the distribution of NSS across the two categories. While the median NSS for uptrend periods appears marginally higher, both distributions show substantial overlap in their interquartile ranges, and each contains negative outliers. These results further support the conclusion that sentiment intensity, as measured by NSS, does not vary systematically with the direction of stock price movements.

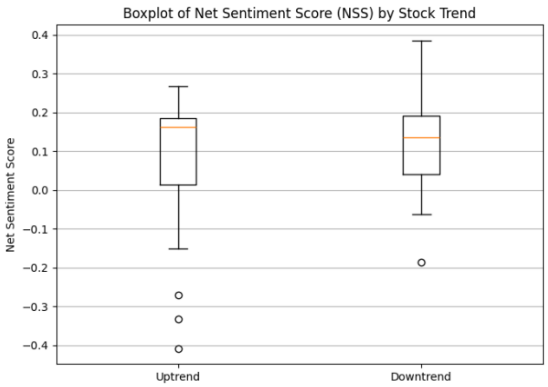
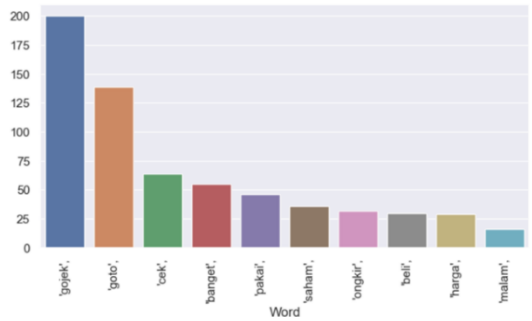


Figure 4. *Boxplot of Net Sentiment Score (NSS) by stock trend*

This visual evidence aligns with the statistical findings: despite minor differences in central tendency, there is no meaningful separation in NSS distributions between stock movement directions. In summary, the Mann-Whitney U test, supported by visual inspection, confirms that net sentiment score does not significantly vary according to stock trend direction. This finding aligns with the results of the previous categorical analysis, suggesting that sentiment, whether treated as a classification or continous measure, does not exhibit a strong relationship with market movement within the observed timeframe.

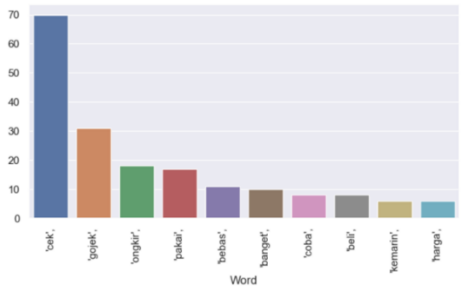
3.4. Visualization

The visualization aims to identify frequently occurring words in the comment data, which facilitates the interpretation and understanding of the information contained within the comments. As a sample, four periods were selected based on Table 3, consisting of two periods during an upward stock trend and two periods during a downward stock trend.



(a)

Jurnal Teknik Informatika Vol. 18 No. 2, October 2025 (257-269)
ISSN: p-ISSN 1979-9160 (Print)| e-ISSN 2549-7901 (Online)
DOI: https://doi.org/10.15408/jti.v18i2.46447

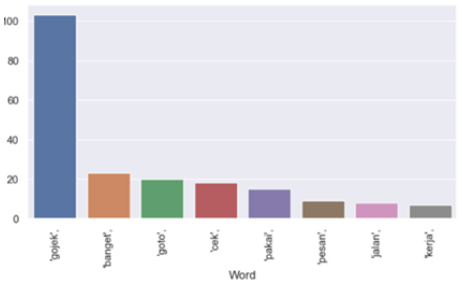


(b)

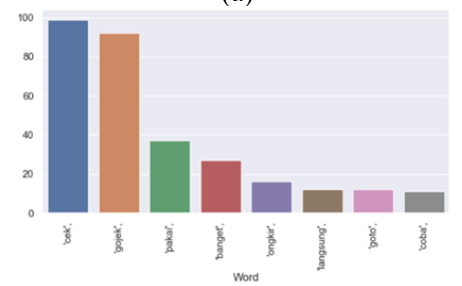
Figure 5. *Word frequency during: (a) downward stock trend with negative sentiment category, and (b) downward stock trend with positive sentiment category*

Figure 5(a) shows the word frequency distribution during Period 1 (01/12/2022–12/12/2022), when the stock trend was declining and sentiment was predominantly negative. Dominant terms included "gojek," "goto," "*cek* (check)," "*saham* (stock)," and "*harga* (price)," with "*saham* (stock)" co-occurring with "GoTo" in 4.4% of the 951 tweets—indicating that investment-related topics were part of the public discourse.

In contrast, Figure 5(b) displays word frequencies from Period 9 (30/12/2022–01/01/2023), which also featured a downward trend but positive sentiment. Although commonly used words remained similar, the co-occurrence of "*saham* (stock)" and "GoTo" dropped to 1.3% of 262 tweets, suggesting that positively toned discussions were less focused on financial performance.



(a)



(b)

Figure 6. *Word frequency during: (a) upward stock trend with positive sentiment category, and (b) upward stock trend with negative sentiment category*

Figure 6(a) illustrates word frequencies during Period 34 (03/03/2023–08/03/2023), a rising stock trend accompanied by predominantly positive sentiment. Frequent terms such as "gojek," "*cek* (check)," "*coba* (try)," and "goto" reflect engagement with GoTo's services, but notably, there was no co-occurrence of "*saham* (stock)" with "GoTo," suggesting minimal focus on investment topics despite the favorable trend.

In contrast, Figure 6(b) depicts Period 38 (21/03/2023–28/03/2023), which also saw a rising trend but with dominant negative sentiment. Commonly used words—like "*banget* (very)," "*pesan* (order)," "*kerja* (work)," and "*pakai* (use)"—indicated service-related complaints or user dissatisfaction. The co-occurrence of "*saham* (stock)" and "GoTo" appeared in only 1.5% of 292 tweets, reinforcing the view that public sentiment, even when negative, may not center on financial aspects during upward stock movements.

Based on the above descriptions and visualizations, it can be interpreted that public perceptions (whether positive or negative) of a company's condition do not always reflect actual market conditions. These results also reinforce the notion that, even during periods of rising stock prices, investment-related topics are not the primary focus of public opinion—even when sentiment is negative. The word distribution visualizations also support the results of the Chi-Squared Test and the Goodman-Kruskal Lambda, as the emerging sentiments were not consistently aligned with the upward or downward trends of stock prices.

3.5. Discussions

This study examined whether public sentiment expressed on social media is associated with the direction of GoTo's stock price movement. Sentiment was analyzed both as a categorical variable (positive/negative) and a numerical score (Net Sentiment Score or NSS), while stock movement was classified into uptrend and downtrend periods.

The statistical results indicated no significant relationship between sentiment and stock trends, regardless of how sentiment was represented. Accordingly, the null hypothesis—which states that Twitter sentiment has no meaningful relationship with stock price direction—was not rejected. The association

was weak, and sentiment showed minimal predictive power. Similarly, sentiment intensity measured by NSS did not differ significantly between uptrend and downtrend periods, as supported by both statistical tests and boxplot visualization.

Several factors may explain these findings. First, the effect of sentiment may be delayed or indirect, and thus not captured in aggregated trend-based periods. Second, stock prices are likely influenced by multiple factors—such as macroeconomic indicators, investor behavior, official news, and government policies—that go beyond public sentiment. While this study did not directly integrate macroeconomic or fundamental data, public news sources during the observation period included coverage of rising interest rates, tech-sector corrections, and regulatory uncertainties in Indonesia—all of which may have influenced GoTo's stock performance independently of social media sentiment. Incorporating such contextual data is suggested for future research to more comprehensively isolate sentiment effects.

Third, the Naïve Bayes classifier used in this study may not adequately capture complex expressions like sarcasm or irony, which are common in social media. Moreover, aggregating sentiment into multi-day trend periods may obscure short-term fluctuations more sensitive to market sentiment shifts.

Additionally, the sentiment classifier achieved an accuracy of 70%, which, while acceptable for basic sentiment detection, may introduce a degree of noise into the aggregated sentiment scores. A higher accuracy could potentially improve the signal quality of sentiment data and sharpen the observed relationship with stock price movement. Therefore, future studies could benefit from conducting sensitivity analysis to assess whether improved classification performance would yield different statistical outcomes.

Theoretically, these findings suggest that although sentiment analysis has become a popular approach in financial market studies, its results are not always consistent across different contexts. The use of sentiment metrics as a standalone indicator appears to be insufficient for predicting short-term stock price movements. For practitioners, particularly investors or market analysts, this finding serves as a reminder that relying solely on public sentiment data, without incorporating technical and fundamental indicators, may lead to less accurate decision-making. In other words, sentiment should be used as a complementary tool, rather than a substitute for other forms of market information.

3.6. Limitations and Future Work

The finding that public sentiment does not have a significant relationship with stock price movement contradicts several previous studies. For instance, study [10] reported that tweet volume influenced prediction accuracy, while [13] demonstrated interpretable links between sentiment aspects and stock prices using ABSA with statistical techniques like Pearson correlation and Granger causality. These results suggest that fine-grained sentiment modeling may offer greater predictive value than our broader approach.

Several factors may explain this discrepancy. The data scope, limited to one stock and a relatively short observation period, may restrict the generalizability of the findings. In addition, the classification approach used remains relatively basic and does not yet capture more complex emotional or contextual nuances. This study also recognizes inherent limitations associated with using Twitter as a data source. Although Twitter is widely used in Indonesia, its user base may not accurately reflect the broader population of retail investors. The presence of automated accounts (bots), echo chambers, and opinion leaders can skew sentiment distributions, potentially amplifying certain narratives disproportionately. Moreover, without demographic validation, it remains unclear how representative Twitter sentiment truly is of actual investor behavior in emerging markets such as Indonesia.

Future research should consider expanding the dataset across multiple stocks and longer observation windows, incorporating technical indicators (e.g., RSI) into multivariable models, and applying more advanced classification models such as ensemble methods or transformer-based architectures. Additionally, integrating real-time data and market context may help uncover lag effects or indirect influences of sentiment on price dynamics.

CONCLUSION

This study examined whether public sentiment on Twitter could predict the direction of GoTo's stock price movement. A Naïve Bayes classifier achieved 70% accuracy in sentiment classification. However, statistical tests—including Chi-Square (p = 0.6398), Cramer's V (0.014), Goodman-Kruskal Lambda (0.053), and Mann-Whitney U (p = 0.8994)—found no significant relationship between sentiment and stock trends. Based on these findings, the null hypothesis, which posits no meaningful association between Twitter sentiment and stock price movement, was not rejected.

These results suggest that sentiment analysis alone may be insufficient for short-term stock price forecasting, particularly for early-phase IPO stocks such as GoTo. While public opinion can be a valuable information source, its predictive value appears limited in isolation. Practitioners are advised to complement sentiment signals with other indicators, such as trading volume, price volatility, or macroeconomic news, for more reliable market insight.

REFERENCES

[1] L. Kuerzinger and P. Stangor, "The Relevance and Influence of Social Media Posts on Investment Decisions of Young and Social Media-Savvy Individuals — An Experimental Approach Based on Tweets," *Journal of Behavioral and Experimental Finance*, vol. 44, 2024, doi: https://doi.org/10.1016/j.jbef.2024.101005.

[2] Z. Zhao and X. Li, "Social Media and Family Investment Behavior," *Finance Research Letters*, vol. 61, 2024, doi: https://doi.org/10.1016/j.frl.2023.104945.

[3] S. K. Khatik, R. Joshi, and V. K. Adwani, "Inferring the Role of Social Media on Gen Z's Investments Decisions," *Journal of Content Community and Communication*, vol. 14, no. 8, pp. 309–317, Dec. 2021, doi: 10.31620/JCCC.12.21/24.

[4] E. Djatmiko and J.-M. Lagoda, "How Can Retail and Institutional Investors Impact GOTO's Price and Valuation," *International Journal of Advanced Research in Economics and Finance*, vol. 4, no. 3, pp. 165–179, Oct. 2022, doi: 10.55057/ijaref.2022.4.3.15.

[5] D. Alanudin and M. M. Robbani, "Harmonizing Social Impact and Corporate Success: The Nexus of GOTO's Contribution to Society and Profitability," *Jurnal Indonesia Sosial Sains (JISS)*, vol. 5, no. 08, pp. 2131–2138, Sep. 2024, doi: 10.59141/jiss.v5i08.1222.

[6] L. Talans and A. M. A. F. Minardi, "Behavior of Stock Prices Due to The Lock-Up Period Expiration in IPOs and Follow-ons," *Revista Contabilidade & Finanças*, vol. 32, no. 86, pp. 331–344, Aug. 2021, doi: 10.1590/1808-057x202112150.

[7] A. Yadav, "Sentiment Analysis Using Twitter Data," *International Journal for Research in Applied Science and Engineering Technology (IJRASET)*, vol. 11, no. 5, pp. 5833–5837, May 2023, doi: 10.22214/ijraset.2023.52899.

[8] T. Sharma and P. Kaushik, "Leveraging Sentiment Analysis for Twitter Data to Uncover User Opinions and Emotions," *International Journal on Recent and Innovation Trends in Computing and Communication*, vol. 11, no. 8s, pp. 162–169, Aug. 2023, doi: 10.17762/ijritcc.v11i8s.7186.

[9] J. Gulati, D. Sethi, S. Kumar, and F. Choudhary, "An Analysis and Research on Sentiments of Twitter Data," in *2021 3rd International Conference on Advances in Computing, Communication Control and Networking (ICAC3N)*, Greater Noida, India: IEEE, Dec. 2021, pp. 32–37. doi: 10.1109/ICAC3N53548.2021.9725702.

[10] C. Saravanos and A. Kanavos, "Forecasting Stock Market Volatility using Social Media Sentiment Analysis," *Neural Comput & Applic*, Dec. 2024, doi: 10.1007/s00521-024-10807-w.

[11] P. Koukaras, C. Nousi, and C. Tjortjis, "Stock Market Prediction Using Microblogging Sentiment Analysis and Machine Learning," *Telecom*, vol. 3, no. 2, pp. 358–378, May 2022, doi: 10.3390/telecom3020019.

[12] R. Gupta and M. Chen, "Sentiment Analysis for Stock Price Prediction," in *2020 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR)*, Shenzhen, Guangdong, China: IEEE, Aug. 2020, pp. 213–218. doi: 10.1109/MIPR49039.2020.00051.

[13] K. Ong, W. van der Heever, R. Satapathy, E. Cambria, and G. Mengaldo, "FinXABSA: Explainable Finance through Aspect-Based Sentiment Analysis," in *2023 IEEE International Conference on Data Mining Workshops (ICDMW)*, Dec. 2023, pp. 773–782. doi: 10.1109/ICDMW60847.2023.00105.

[14] A. Derakhshan and H. Beigy, "Sentiment Analysis on Stock Social Media for Stock Price Movement Prediction," *Engineering Applications of Artificial Intelligence*, vol. 85, pp. 569–578, Oct. 2019, doi: 10.1016/j.engappai.2019.07.002.

[15] S. V. Kolasani and R. Assaf, "Predicting Stock Movement Using Sentiment Analysis of Twitter Feed with Neural Networks," *Journal of Data Analysis and Information Processing*, vol. 08, no. 04, pp. 309–319, 2020, doi: 10.4236/jdaip.2020.84018.

[16] Padmanayana, Varsha, and Bhavya K, "Stock Market Prediction Using Twitter Sentiment Analysis," *International Journal of Scientific Research in Science and Technology*, pp. 265–270, Jul. 2021, doi: 10.32628/CSEIT217475.

[17] D. Valle-Cruz, V. Fernandez-Cortez, A. López-Chau, and R. Sandoval-Almazán, "Does Twitter Affect Stock Market Decisions? Financial Sentiment Analysis During Pandemics: A Comparative Study of the H1N1 and the COVID-19 Periods," *Cogn Comput*, vol. 14, no. 1, pp. 372–387, Jan. 2022, doi: 10.1007/s12559-021-09819-8.

[18] Z. Song, "Behavioral Biases in The Cryptocurrency Market: A Study on The Impact of Investor Sentiment on Price Anomalies," *Journal of Applied Economics and Policy Studies*, vol. 18, no. 2, pp. 35–39, Apr. 2025, doi: 10.54254/2977-5701/2025.21938.

[19] S. Jain, S. K. Jain, and S. Vasal, "An Effective TF-IDF Model to Improve the Text Classification Performance," in *2024 IEEE 13th International Conference on Communication Systems and Network Technologies (CSNT)*, Jabalpur, India: IEEE, Apr. 2024, pp. 1–4. doi: 10.1109/CSNT60213.2024.10545818.

[20] W. Chandra, B. Suprihatin, and Y. Resti, "Median-KNN Regressor-SMOTE-Tomek Links for Handling Missing and Imbalanced Data in Air Quality Prediction," *Symmetry*, vol. 15, no. 4, p. 887, Apr. 2023, doi: 10.3390/sym15040887.

[21] Z. A. Faridzi, D. Pramesti, and R. Y. Fa'rifah, "A Comparison of Oversampling and Undersampling Methods in Sentiment Analysis Regarding Indonesia Fuel Price Increase Using Support Vector Machine," in *2023 International Conference on Advancement in Data Science, E-learning and Information System (ICADEIS)*, Bali, Indonesia: IEEE, Aug. 2023, pp. 1–6. doi: 10.1109/ICADEIS58666.2023.10270851.

[22] H. Fakhrurroja, T. Maeza Chiqamara, F. Hamami, and D. Pramesti, "Sentiment Analysis of Local Water Company Customer Using Naive Bayes Algorithm," in *2024 IEEE International Conference on Industry 4.0, Artificial Intelligence, and Communications Technology (IAICT)*, BALI, Indonesia: IEEE, Jul. 2024, pp. 168–173. doi: 10.1109/IAICT62357.2024.10617472.

[23] J. Xu, Y. Zhang, and D. Miao, "Three-way Confusion Matrix for Classification: A Measure Driven View," *Information Sciences*, vol. 507, pp. 772–794, Jan. 2020, doi: 10.1016/j.ins.2019.06.064.

[24] A. F. Siegel, "Chi-Squared Analysis," in *Practical Business Statistics*, Elsevier, 2012, pp. 507–522. doi: 10.1016/B978-0-12-385208-3.00017-1.

[25] A. Agresti, C. Franklin, and B. Klinkenberg, *Statistics: The Art of Science of Learning from Data*, 4. ed. Boston: - Pearson, 2017.

[26] M. Mendikowski, B. Schindler, T. Schmid, R. Möller, and M. Hartwig, "Improved Techniques for Training Tabular GANs Using Cramer's V

Statistics," *Proceedings of the Canadian Conference on Artificial Intelligence*, Jun. 2023, doi: 10.21428/594757db.4c0ffb71.

[27] A. Agresti, *Introduction to Categorical Data Analysis*. John Willey & Sons, Inc., 2007.

[28] T. O. Kvålseth, "Measuring Association Between Nominal Categorical Variables: An Alternative to the Goodman–Kruskal Lambda," *Journal of Applied Statistics*, vol. 45, no. 6, pp. 1118–1132, Apr. 2018, doi: 10.1080/02664763.2017.1346066.

[29] T. W. MacFarland and J. M. Yates, "Mann–Whitney U Test," in *Introduction to Nonparametric Statistics for the Biological Sciences Using R*, Cham: Springer International Publishing, 2016, pp. 103–132. doi: 10.1007/978-3-319-30634-6_4.

[30] N. A. Vidya, M. I. Fanany, and I. Budi, "Twitter Sentiment to Analyze Net Brand Reputation of Mobile Phone Providers," *Procedia Computer Science*, vol. 72, pp. 519–526, 2015, doi: 10.1016/j.procs.2015.12.159.