

Sentiment Analysis for Starbucks Stock Price Prediction Using VADER and ARIMAX

Widyanti Rahayu^{1*}, Gabriel Olivia Yvonne Manurung², Eti Dwi Wiraningsih², Vera Maya Santi¹

¹Study program of Statistics, Universitas Negeri Jakarta, Indonesia

²Study Program of Mathematics, Universitas Negeri Jakarta, Indonesia

wrahayu@unj.ac.id

ABSTRACT

Article History:

Received : 06-04-2026

Revised : 17-06-2026

Accepted : 20-06-2026

Online : 01-10-2026

Keywords:

Sentiment Analysis;

VADER;

ARIMAX;

Stock Price Forecasting.



The escalation of the Israel–Palestine conflict in October 2023 generated extensive public reactions worldwide, particularly on social media platforms. On October 9, 2023, a statement supporting Palestine posted by the Starbucks Workers United union on X (Twitter) prompted legal action from Starbucks. Although the dispute was related to trademark and intellectual property issues, public interpretation associated the lawsuit with opposition to pro-Palestinian support, leading to boycott campaigns in several countries. This situation coincided with a notable decline in Starbucks' stock performance, including an eleven-day consecutive decrease in December 2023. Given the growing influence of online discussions on financial markets, social media sentiment may provide valuable information for stock price forecasting. Using a quantitative time-series forecasting approach, this study aims to analyze public sentiment toward Starbucks associated with the boycott movement, develop an ARIMAX model incorporating sentiment as an exogenous variable for stock price forecasting, and evaluate the predictive performance of the proposed model. A total of 41,211 English-language posts containing the keyword “boycott Starbucks” were collected from X between October 2023 and December 2024, while stock market data were obtained from Yahoo Finance through the *yfinance* library. Sentiment classification was conducted using the VADER approach, followed by ARIMAX modeling and forecast evaluation using Mean Absolute Percentage Error (MAPE) and Mean Squared Error (MSE). The results indicate that negative sentiment dominated the discussions (54%), followed by positive (43.3%) and neutral sentiment (2.7%), with the highest activity occurring in December 2023. Based on the minimum AIC and BIC values, ARIMAX(3,0,3) was selected as the best model and satisfied all diagnostic assumptions. Forecasting for the period January 2–16, 2025 produced a MAPE of 1.97% and an MSE of 3.58, indicating high forecasting accuracy. The findings highlight the potential of incorporating social media sentiment as an external information source in ARIMAX-based stock price forecasting, particularly during periods of heightened public attention triggered by major socio-political events.



Crossref

<https://doi.org/10.31764/jtam.v10i4.39002>



This is an open access article under the **CC-BY-SA** license

A. INTRODUCTION

Stock investment represents a popular form of investment for obtaining financial gains, with opportunities for capital gain and dividend income. Investors require well-developed strategies to minimize risk, particularly through predicting stock price movements, which is crucial for determining optimal timing for buying or selling shares (Sharma et al., 2021). Equity price movements are shaped by multiple determinants, notably a firm's financial condition and the collective sentiment emerging from news coverage, social media discourse, and other

information channels (Khan et al., 2020). In the digital era, social media has become a primary platform for real-time opinion expression, significantly influencing public perception and stock market performance (Ivanenko, 2023; Mawarnie et al., 2024).

The boycott movement represents one phenomenon affecting public sentiment, reflecting negative sentiment that can decrease stock prices through reduced consumer purchasing power and disrupted financial performance (Albrecht et al., 2013). The Palestine-Israel conflict escalation on October 7, 2023, triggered international boycotts against companies perceived as supporting Israel, with Starbucks Corporation becoming a major target after suing Starbucks Workers United over a "Solidarity with Palestine!" post. This sparked a viral boycott movement on TikTok with the hashtag #boycottstarbucks, reflecting the role of social media in amplifying political consumerism and translating it into measurable sales and stock market impacts (Fekrazad et al., 2022). The resulting stock price decline was recorded as the longest consecutive drop since the company's IPO, consistent with evidence that boycott events produce significant negative abnormal returns on targeted firms (Ranco et al., 2015). Starbucks' stock price decreased from \$107.21 in November 2023 to \$93.89 in May 2024, with first quarter of 2024 revenue declining (Investing.com, 2025), demonstrating that public sentiment significantly influences market performance.

Although numerous studies have investigated the role of social media sentiment in stock market forecasting, several limitations remain. First, most previous studies focused on general market sentiment, corporate earnings announcements, or routine financial news, while limited attention has been given to sentiment generated by large-scale global boycott movements that emerge from geopolitical events (Kumar & Prabhu, 2018). Such events tend to create prolonged public reactions, collective consumer behavior, and reputational risks that differ from conventional market information. Second, existing studies generally evaluate sentiment effects on stock prices without specifically examining how boycott-related sentiment evolves over time and interacts with stock market dynamics during periods of sustained public pressure (Borg & Boldt, 2020; Hutto & Gilbert, 2014). Third, while ARIMAX and sentiment-based forecasting models have been widely applied in financial prediction, empirical evidence integrating social media sentiment, stock opening prices, and trading volume within a unified forecasting framework for the Starbucks boycott case remains scarce (Ibrahim & Alturfi, 2022; Wang & Vo, 2025; Pokhrel & Adhikari, 2023; Wang & Vo, 2025; Sharma et al. 2021). Consequently, further investigation is required to understand whether sentiment associated with a global boycott campaign can provide meaningful information for stock price forecasting in a high-profile multinational corporation.

This study aims to analyze public sentiment toward Starbucks on the social media platform X during the boycott movement related to the Israel–Palestine conflict, to develop a stock price forecasting model using the ARIMAX approach by incorporating sentiment, opening price, and trading volume as exogenous variables, and to evaluate the predictive performance of the proposed model. In addition, this study seeks to provide empirical evidence regarding the potential contribution of social media sentiment as an external information source in forecasting stock price movements during periods of heightened public attention and reputational pressure.

B. METHODS

1. Data Sources and Research Variable

This study employed a quantitative research approach with a time-series forecasting design. The research aimed to investigate the relationship between public sentiment on social media and Starbucks stock price movements during the boycott period, as well as to develop a forecasting model using the Autoregressive Integrated Moving Average with Exogenous Variables (ARIMAX) method. The analysis integrated sentiment data derived from Twitter (X) with historical stock market data to evaluate the potential contribution of sentiment indicators to stock price prediction.

This study draws upon two main datasets spanning the period from October 1, 2023, to December 31, 2024. The first dataset comprises sentiment data collected from Twitter (X) through web scraping using the keyword “boycott starbucks,” yielding 41,211 English-language tweets. For optimal VADER performance, tweets were preprocessed by limiting each sentence to no more than 50 words. The second dataset contains historical stock prices of Starbucks (Close, Open, High, Low, Volume), automatically retrieved via the Python *yfinance* library, totaling 317 observations. The datasets were subsequently merged by matching observation dates, yielding 267 aligned records spanning October 2, 2023, to December 30, 2024, to analyze the linkage between public sentiment and stock price movements. Furthermore, stock price data covering 10 active trading days (January 2–16, 2025) were collected for out-of-sample forecast evaluation, with weekends and public holidays excluded from the analysis.

Exogenous variables included the VADER compound score, Open price, and trading Volume. Sentiment measured public attitudes toward the Starbucks boycott, Open reflected initial market value, and Volume captured trading intensity. These variables were chosen for their relevance in explaining price movements and sentiment-driven market dynamics.

2. Research Procedure

The research methodology followed a systematic approach consisting of the following steps:

- a. Tweets containing the keyword “boycott starbucks” were scraped from Twitter, while historical stock prices of Starbucks were retrieved using the Python *yfinance* library.
- b. Tweet data were preprocessed through cleaning (removal of URLs, mentions, and special characters), tokenization, stopword elimination, and lemmatization to obtain base word forms.
- c. VADER was used to classify sentiments and calculate compound scores ranging from -1 (most negative) to +1 (most positive), computed as (Hutto & Gilbert, 2014):

$$\text{Compound Score} = \frac{x}{\sqrt{x^2 + \alpha}} \quad (1)$$

where x is the aggregate valence score and α is a normalization parameter ($\alpha = 15$). To assess the validity of the sentiment classification results, a random subset of tweets was manually annotated and compared with the sentiment labels generated by VADER. The manual labeling process categorized tweets into positive, negative, and neutral classes

based on their contextual meaning. Classification performance was evaluated using accuracy, precision, recall, and F1-score metrics.

- d. Sentiment analysis results were merged with stock price data based on date.
- e. The Variance Inflation Factor (VIF) was computed to evaluate the presence of multicollinearity among the explanatory variables (Holzhauer et al., 2016):

$$VIF = \frac{1}{1 - R_i^2} = \frac{1}{\text{Tolerance}} \quad (2)$$

where R_i^2 is the coefficient of determination obtained from the auxiliary regression. The auxiliary regression is performed by regressing one independent variable against all other independent variables. $VIF > 10$ indicates problematic multicollinearity.

- f. Data were divided into 80% training set (214 observations, October 2, 2023 - October 10, 2024) and 20% testing set (53 observations, October 11, 2024 - December 30, 2024).
- g. ARIMAX model identification was applied by first forming a regression model using the best subset regression method.
- h. Variance stationarity was assessed using the Box-Cox test, and a Box-Cox transformation was applied when necessary (Wei, 2006):

$$y(\lambda) = \begin{cases} \frac{(y^\lambda - 1)}{\lambda} & \text{if } \lambda \neq 0 \\ \ln(y) & \text{if } \lambda = 0 \end{cases} \quad (3)$$

where y is original data and λ is transformation parameter. The ADF test was used for mean stationarity (J. Wang et al., 2021):

$$\Delta y_t = \alpha + \beta t + \gamma y_{t-1} + \sum_{i=1}^p \delta_i \Delta y_{t-i} + \epsilon_t \quad (4)$$

where Δy_t is first difference of y_t , α is constant term, β is coefficient of time trend, t is time trend, γ is coefficient of lagged level, p is lag order, δ_i is coefficients of lagged differences, and ϵ_t is error term. If non-stationarity was detected, transformation or differencing was performed.

- i. Regression diagnostics were conducted to verify that model assumptions were met.
- j. ACF and PACF plots of regression residuals were examined to determine initial parameters (p, q) for the ARIMAX model.
- k. The ARIMAX(p, d, q) model was estimated (Tiao & Box, 1981):

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \epsilon_t - \theta_1 \epsilon_{t-1} - \theta_2 \epsilon_{t-2} - \dots - \theta_q \epsilon_{t-q} + \sum_{i=1}^b \beta_i X_{i,t} \quad (5)$$

where y_t is stock closing price at time t , p is order of non-seasonal AR terms, q is order of non-seasonal MA terms, b is number of exogenous variables incorporated in the model, $X_{i,t}$ is exogenous variables at time t , β_i is coefficients of the covariates, and ϵ_t white noise process.

- l. Model selection was based on the minimum AIC and BIC values to obtain the most optimal model (Akaike, 1998; V. M. Santi et al., 2019, 2021, 2022, 2025, 2026; Schwarz, 1978)

$$AIC = -2 \ln(L) + 2k \text{ and } BIC = -2 \ln(L) + k \ln(n) \quad (6)$$

where L denotes the maximized likelihood function of the model, k represents the total number of estimated parameters, and n corresponds to the sample size.

- m. Diagnostic procedures were carried out to verify the adequacy and reliability of the fitted model. The Jarque-Bera test assessed normality (Thadewald & Büning, 2007):

$$JB = \frac{n}{2} \left(S_w^2 + \frac{(K_r - 3)^2}{4} \right) \quad (7)$$

where JB denotes the Jarque-Bera test statistic, S is the sample skewness, K is the sample kurtosis. The Ljung-Box test verified white noise residuals (Cryer & Chan, 2008; V. Santi et al., 2023):

$$Q = n(n+2) \sum_{k=1}^K \frac{\hat{\rho}_k^2}{n-k} \quad (8)$$

where Q is ljung-box test statistic, n is number of residuals, $\hat{\rho}_k$ is ACF of the residuals at lag k , k is maximum lag. The Breusch-Pagan test was used to detect heteroscedasticity (Breusch & Pagan, 1979):

$$\chi^2 = n \times R^2 \quad (9)$$

where χ^2 is LM test statistic, n is the sample size, R^2 is coefficient of determination obtained from the auxiliary regression.

- n. The ARIMAX model was evaluated by comparing its forecasts against actual closing prices.
- o. Forecasting was conducted using the selected ARIMAX model for 10 active trading days (January 2-16, 2025), excluding weekends and holidays.
- p. MAPE (Laurente Blanco & Machaca Hanco, 2020) and MSE (Chicco et al., 2021) evaluated prediction accuracy:

$$MAPE = \frac{100\%}{n} \sum_{t=1}^n \left| \frac{y_t - \hat{y}_t}{y_t} \right| \text{ and } MSE = \frac{1}{n} \sum_{t=1}^n (y_t - \hat{y}_t)^2 \quad (10)$$

where y_t is actual value at time t , $\hat{y}_t$ is predicted value at time t , n is number of observations in the testing set.

C. RESULT AND DISCUSSION

1. Sentiment Distribution

After preprocessing and labeling all collected tweets with sentiment categories, the analysis showed as shown in Table 1.

Table 1. Sentiment Distribution of Tweets Related with the Keyword 'Boycott Starbucks'

Sentiment	Percentage	Count
Negative	54%	22,257
Positive	43.3%	17,841
Neutral	2.7%	1,113

This distribution clearly indicates that public perception toward Starbucks during the boycott period was predominantly critical and negative. To further understand how sentiment evolved over time, monthly sentiment distributions were examined as shown in Figure 1.

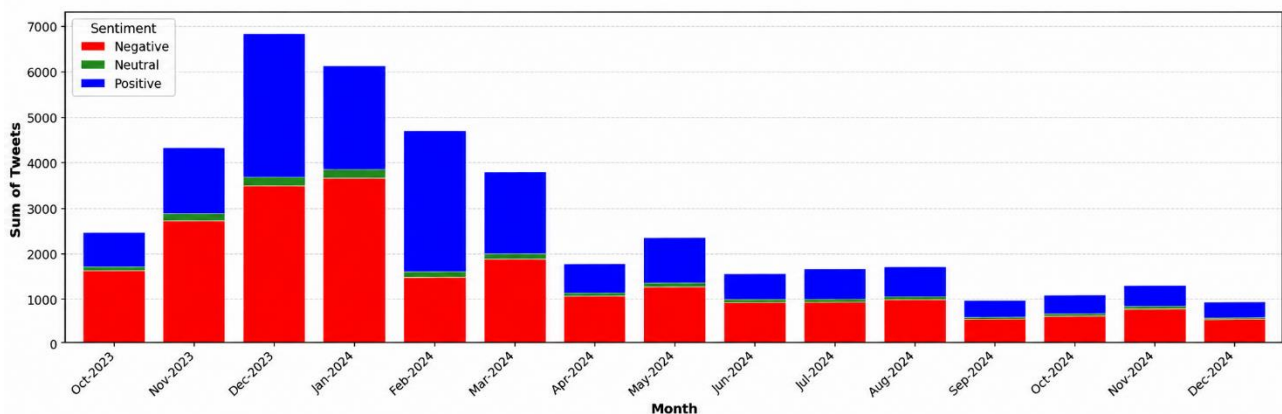


Figure 1. Classification of Tweets with the Keyword 'boycott starbucks' by Month

Monthly sentiment trends show sustained negative sentiment, with peak activity in December 2023, shortly after major boycott-related events. Sentiment fluctuations aligned with several notable declines in Starbucks stock price, consistent with evidence that sustained negative social media sentiment leads to lower stock returns on targeted firms (Fekrazad et al., 2022). To complement the sentiment distribution analysis, a word frequency examination was conducted to identify the most word appearing in the discussion as shown as Figure 2 excluding the words 'boycott' and 'starbucks' as these are the search keywords, as well as noise such as K-Pop entities, other products, and common words.

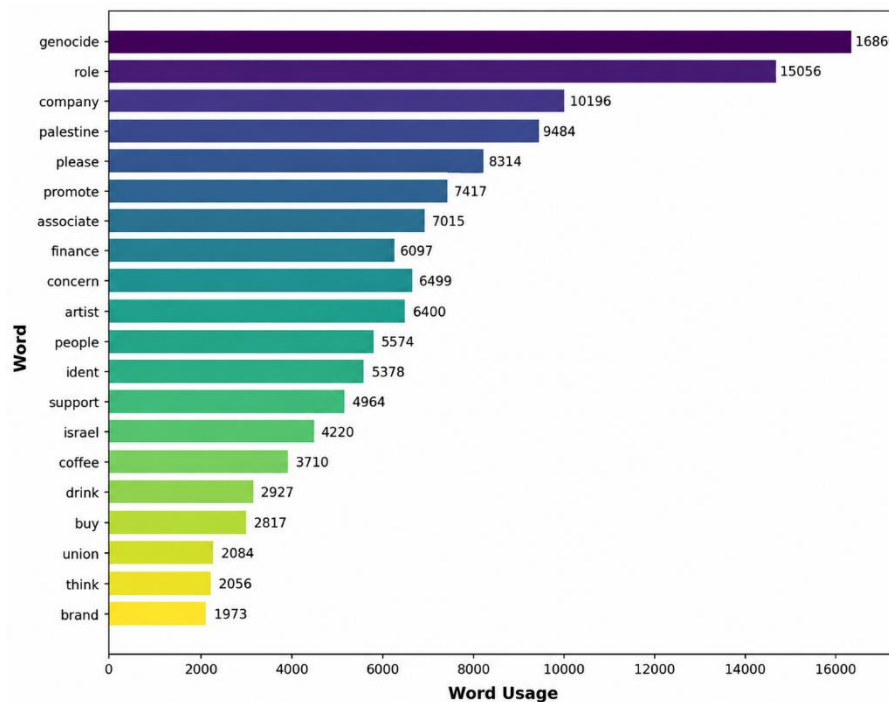


Figure 2. Top 20 Most Frequently Occurring Words

The analysis revealed dominant words such as “genocide,” “company,” and “Palestine,” highlighting the strong political framing embedded in boycott-related conversations. To further explore the linguistic characteristics within each sentiment category, separate word clouds were generated for negative and positive tweets as shown as Figure 3 excluding the words 'boycott', 'starbucks', noise such as K-Pop entities, other products, and common words.

The visual analysis indicates that tweets labeled as positive frequently convey pro-boycott narratives, highlighting an inherent limitation of VADER, which assigns polarity through lexicon-based rules rather than contextual semantic interpretation. (Borg & Boldt, 2020) A random sample of 500 tweets was manually labeled by two independent annotators. The VADER classifier achieved an accuracy of 86.6%, precision of 83.7%, recall of 86.9%, and F1-score of 87.3%.



Figure 3. Word Cloud (a) Negative Sentiment, (b) Positive Sentiment

2. Multicollinearity Testing and Data Preparation

Prior to estimating the ARIMAX model, a multicollinearity assessment was performed to verify that the exogenous variables (Open, Volume, and Sentiment) were not highly intercorrelated. The correlation matrix in Figure 4 indicates weak pairwise associations: -0.09 (Open–Sentiment), -0.18 (Open–Volume), and -0.06 (Sentiment–Volume), all well below the 0.8 threshold typically signaling serious multicollinearity. VIF values were also under 10 for each predictor, confirming the absence of substantial multicollinearity.

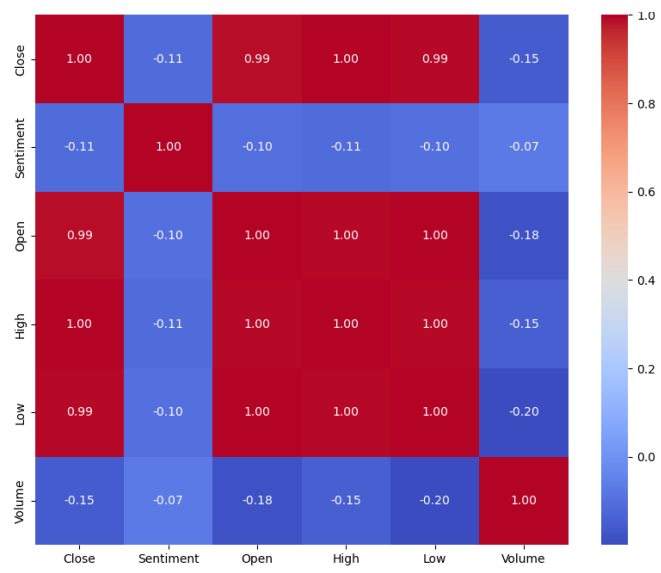
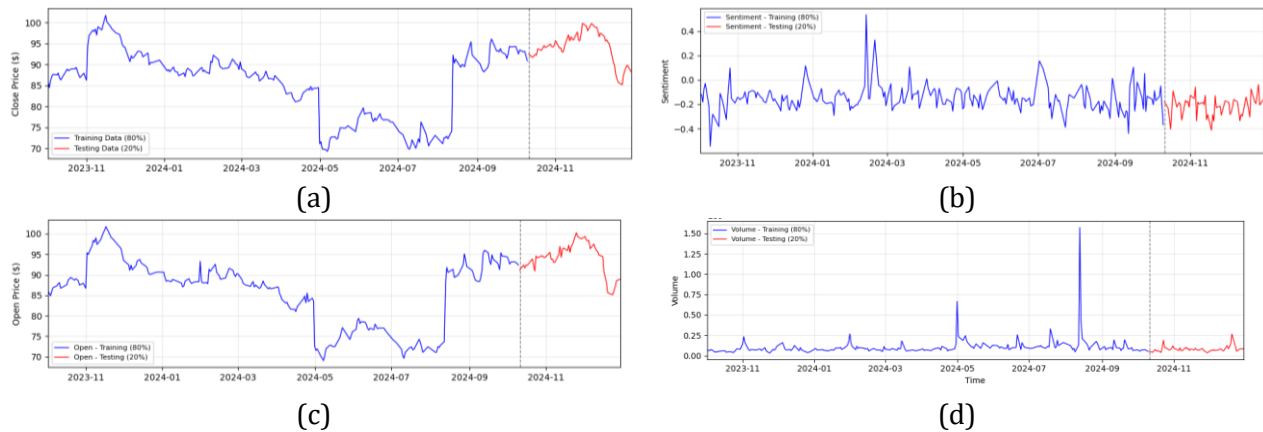


Figure 4. Correlation Matrix Between Variables

Open, Volume, and Sentiment were included as exogenous variables for their theoretical and empirical relevance: Open reflects daily price direction, Volume indicates trading intensity, and Sentiment captures social media influence. Including Open, strongly correlated with Close, prevented underfitting while maintaining model parsimony.

3. Data Split

The dataset (267 observations) was divided into 80% training (214; October 2, 2023–October 10, 2024) and 20% testing (53; October 11–December 30, 2024). This split balances model estimation and out-of-sample evaluation. The ARIMAX model was estimated using the training dataset, whereas the testing dataset was employed to evaluate its forecasting performance (see Figure 5).

**Figure 5.** Plot of Data Training and Data Testing

(a) Plot of Close Price, (b) Plot of Sentiment, (c) Plot of Open Price, (d) Plot of Volume

4. ARIMAX Model Identification from Regression Residuals

Before ARIMAX modeling, a regression using training data was built with Close as the dependent variable and Open, Volume, and Sentiment as predictors via best subset regression.

Table 2. Best Subset Regression Model

Model	Predictor
1	X_1
2	X_1, X_2
3	X_1, X_2, X_3

Table 3. Best Subset Regression Summary

Model	R-Square	Adj. R-Square	Pred R-Square	C(p)	AIC	SBIC
1	0.9801	0.9800	0.9797	12.7222	666.3153	58.8499
2	0.9811	0.9810	0.9781	3.3625	657.1123	49.8811
3	0.9813	0.9810	0.9778	4.0000	657.7283	50.5742

Best Subset Regression Summary (continued)

Model	SBC	MSEP	FPE	HSP	APC
1	676.4132	276.7389	1.3053	0.0061	0.0203
2	670.5762	263.8747	1.2503	0.0059	0.0194
3	674.5581	263.4281	1.2539	0.0059	0.0195

Model 3, including Open, Volume, and Sentiment, was identified as the preferred specification due to its superior R^2 , Adjusted R^2 , and lowest MSEP, despite Model 2 having a slightly lower AIC. The resulting regression equation is:

$$\text{Close} = 0.7514 - 0.9863X_1 + (2.081 \times 10^{-8})X_2 - 0.8109X_3$$

Model 3, which incorporated sentiment scores alongside Open price and Volume, produced the lowest AIC and BIC values among all candidate models. Although this study does not compare ARIMAX with a pure ARIMA model, the progressive improvement observed across the three model specifications provides empirical support for the relevance of sentiment variables in forecasting Starbucks stock prices during the boycott period.

5. Regression Model Stationarity Testing

Based on the Box-Cox plot, the λ value is 1.21 with a 95% confidence interval, indicating the model is stationary in variance as shown in Figure 6.

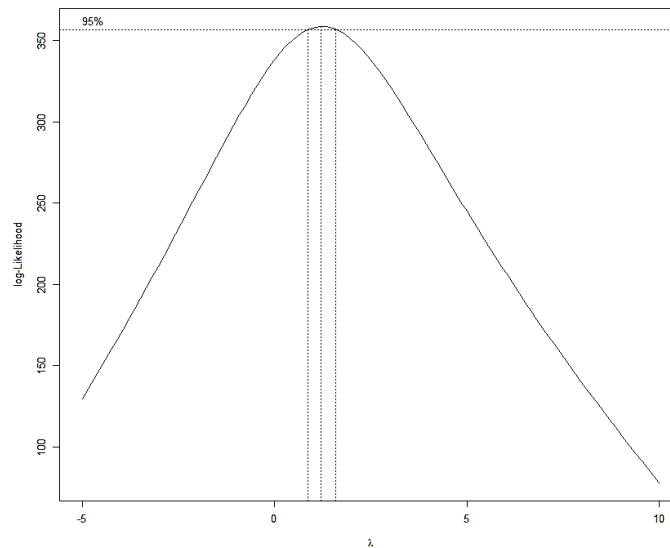


Figure 6. Box-Cox Plot of Regression Residuals Model

The ADF test ($p = 0.01 < 0.05$) confirms mean stationarity without differencing, while ACF and PACF plots show cut-offs at lags 3 and 3–4, respectively (Figure 7).

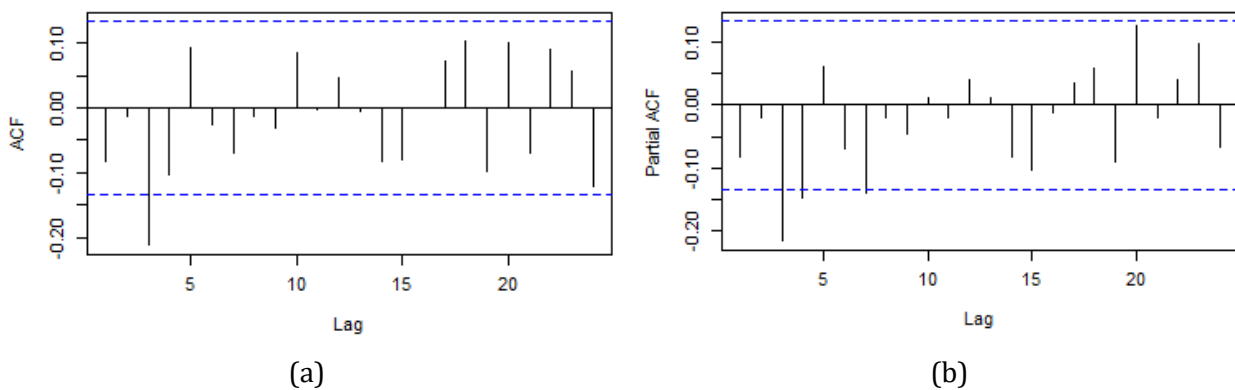


Figure 7. (a) ACF Plot of Residuals, (b) PACF Plot of Residuals

Figure 8 diagnostics show residuals are approximately normal, homoscedastic, and free of significant autocorrelation, as supported by Q–Q, residuals-versus-fitted, ACF, and PACF plots.

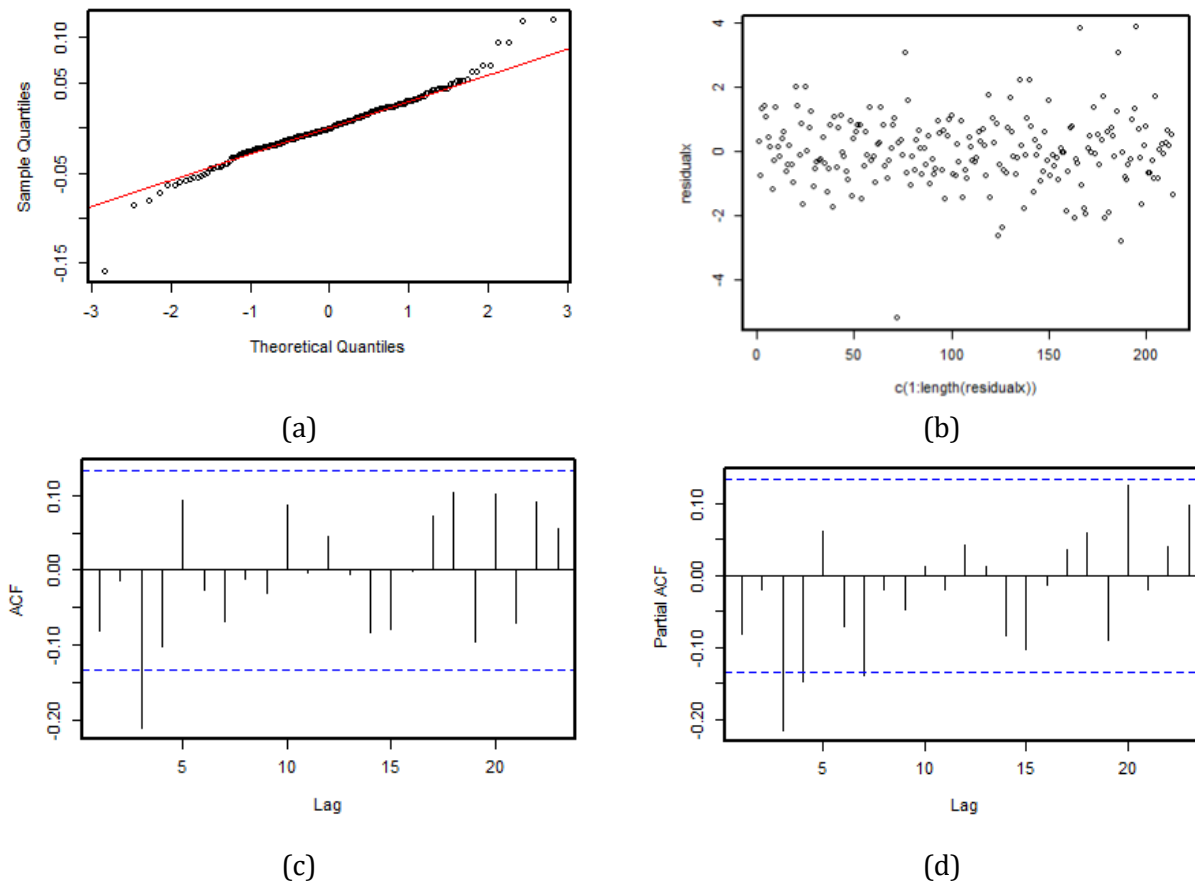


Figure 8. Residual Diagnostic Plots of the Regression Model: (a) Normal Q-Q Plot, (b) Residual vs Index Plot, (c) ACF Plot of Residuals, (d) PACF Plot of Residuals

6. ARIMAX Model Selection and Estimation

Analysis of regression residuals ACF and PACF plots lead to seven candidate ARIMAX models as shown in Table 4.

Table 4. Estimation and Significance Testing of Parameters

Model	Parameter	Parameter Value	p-value	Conclusion	AIC	BIC
ARIMAX (2,0,2)	ϕ_1	-0.12464	0.4985	No	645.4673	675.7611
	ϕ_2	0.69183	6.556×10^{-6}	Significant		
	θ_1	-0.12272	0.5134	No		
	θ_2	-0.87728	2.950×10^{-6}	Significant		
ARIMAX (2,0,3)	ϕ_1	0.36602	0.0547098	No	645.0058	678.6655
	ϕ_2	-0.13917	0.5665165	No		
	θ_1	-0.52443	0.0038780	Significant		
	θ_2	0.16918	0.4959231	No		
	θ_3	-0.34333	3.034×10^{-4}	Significant		
ARIMAX (2,0,4)	ϕ_1	-0.26250	0.51647	No	645.3013	682.3270
	ϕ_2	-0.057124	0.85107	No		
	θ_1	0.11305	0.77732	No		
	θ_2	-0.029970	0.92960	No		
	θ_3	-0.32366	1.222×10^{-4}	Significant		
	θ_4	-0.25136	0.04463	Significant		
	ϕ_1	0.52052	0.064584	No	644.3339	677.9937

Model	Parameter	Parameter Value	p-value	Conclusion	AIC	BIC
ARIMAX (3,0,2)	ϕ_2	0.26646	0.231408	No		
	ϕ_3	-0.13097	0.058456	No		
	θ_1	-0.73137	0.008756	Significant		
	θ_2	-0.26863	0.335116	No		
ARIMAX (3,0,3)	ϕ_1	0.51956	2.351×10^{-9}	Significant	637.0925	674.1182
	ϕ_2	-0.62330	1.835×10^{-11}	Significant		
	ϕ_3	0.54832	1.269×10^{-8}	Significant		
	θ_1	-0.71498	2.2×10^{-16}	Significant		
	θ_2	0.63206	2.2×10^{-16}	Significant		
ARIMAX (4,0,2)	ϕ_1	-0.50350	0.044563	Significant	650.4747	687.5004
	ϕ_2	-0.34649	0.208369	No		
	ϕ_3	-0.27501	0.005616	Significant		
	ϕ_4	-0.23222	0.003591	Significant		
	θ_1	0.38896	0.134445	No		
	θ_2	-0.27875	0.335695	No		
ARIMAX (4,0,3)	ϕ_1	0.51694	1.233×10^{-8}	Significant	638.9796	679.3713
	ϕ_2	-0.60626	1.025×10^{-7}	Significant		
	ϕ_3	0.52591	4.363×10^{-5}	Significant		
	ϕ_4	-0.026443	0.738	No		
	θ_1	-0.72304	2.2×10^{-16}	Significant		
	θ_2	0.63045	2.365×10^{-15}	Significant		
	θ_3	-0.90741	2.2×10^{-16}	Significant		

Parameter estimation using Maximum Likelihood method identified ARIMAX(3,0,3) as the best model with AIC = 637.0925 and BIC = 674.1182, the lowest among all candidates. All parameters were significant ($p - values < 0.05$). Therefore, the model equation is:

$$Y_t = 0.51956Y_{t-1} - 0.62330Y_{t-2} + 0.54832Y_{t-3} + 0.99154X_{1t} + (2.0406 \times 10^{-8})X_{2t} - 2.0850X_{3t} + \varepsilon_t - 0.71498\varepsilon_{t-1} + 0.6320\varepsilon_{t-2} - 0.91708\varepsilon_{t-3}$$

7. Diagnostic Checking

Three diagnostic tests were performed:

- Normality Test: The Jarque–Bera test produced a p-value < 0.05 , suggesting a formal departure from normality. Nevertheless, the Q–Q plot indicates that 61.2% of residuals (131/214) align with the 45° reference line, while 38.8% (83/214) show moderate deviations in the tails. This pattern suggests approximate normality adequate for practical forecasting purposes (see Figure 9).

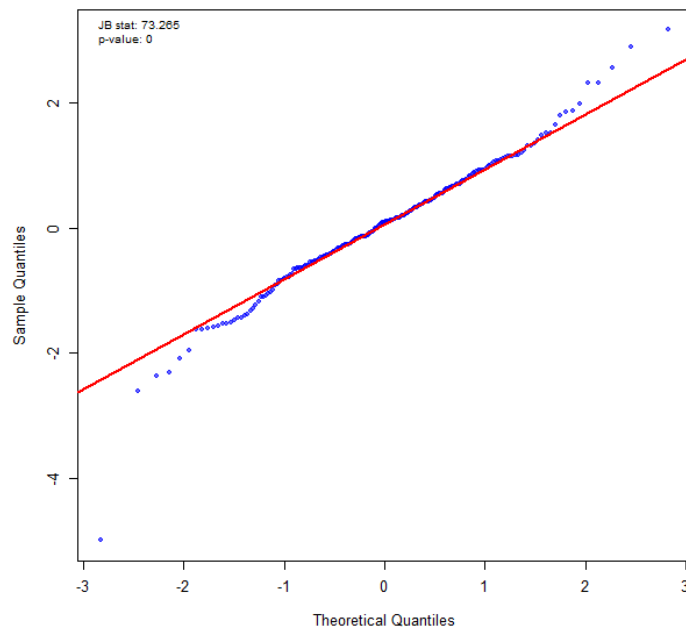


Figure 9. Histogram Plot of Normality Test Results Model ARIMAX(3,0,3)

- b. Independence Test: The Ljung–Box statistics at lags 10, 15, and 20 yielded p-values of 0.9444, 0.9843, and 0.9589, respectively, all above 0.05, indicating no residual autocorrelation. Consistently, the ACF and PACF diagnostics indicate that all lag coefficients lie within the respective confidence intervals (Figure 10).

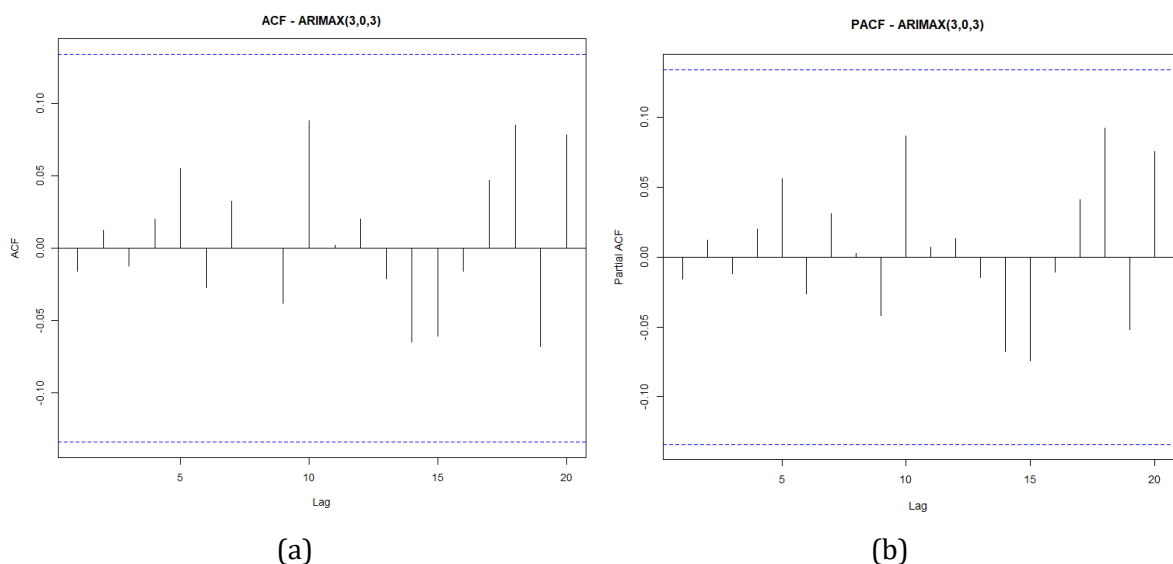


Figure 10. ACF and PACF Plots of the Model ARIMAX(3,0,3) Residuals
(a) ACF Plot, (b) PACF Plot

- c. Heteroscedasticity Test: The ARCH test at lags 10, 15, and 20 produced p-values of 0.992, 0.9894, and 0.9989, respectively, confirming homoscedastic residual variance. As illustrated in Figure 11, residuals oscillate randomly around zero, while the squared residuals display stable dispersion, providing visual support for variance constancy. The Portmanteau Q and Ljung–Box diagnostics further show residual behavior consistent

with white noise, aside from minor extreme fluctuations. Collectively, these results indicate stable variance and support the model's reliability for forecasting.

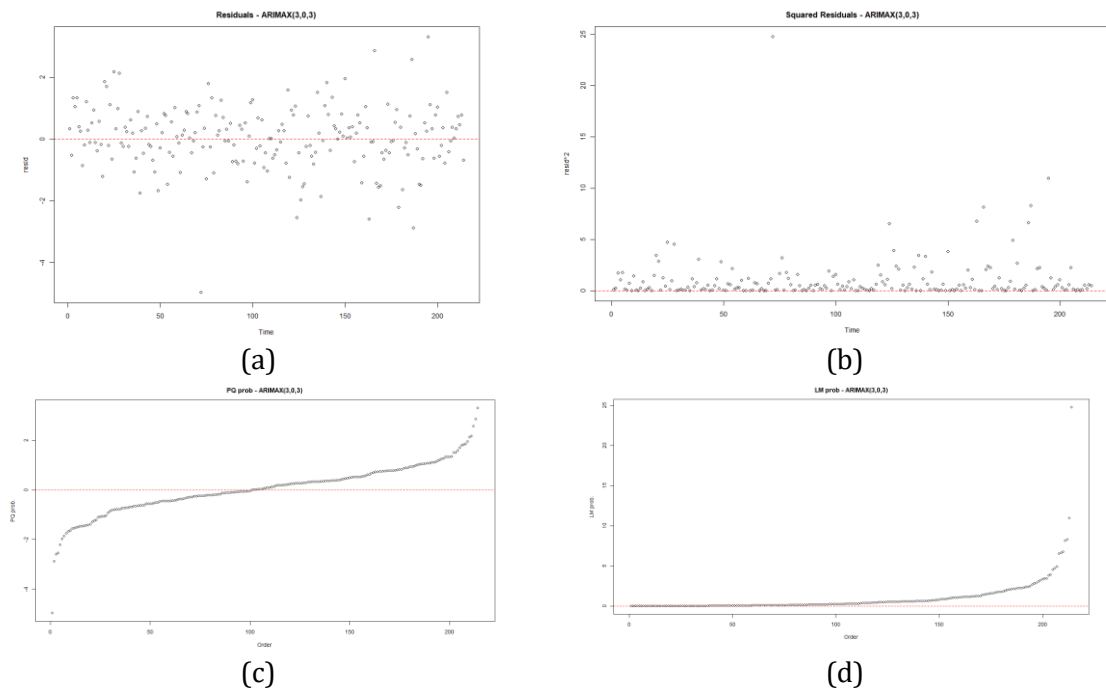


Figure 11. Heteroskedasticity Test Results of Model ARIMAX(3,0,3) (a) Residual Plot, (b) Squared Residuals, (c) Portmanteau Test Probabilities, (d) Lagrange Multiplier Test Probabilities

These results confirm the ARIMAX(3,0,3) model adequately meets statistical assumptions.

8. ARIMAX Model Evaluation on Testing Data

ARIMAX(3,0,3) model was used on 53 days of testing data from October 1 to December 30, 2024. Figure 12 shows that forecast values generally follow actual data patterns, with some variations at specific observation points.



Figure 12. Plot of Starbucks Close Price Using ARIMAX(3,0,3) on Data Testing

Out-of-sample evaluation produced $\text{MAPE} = 1.10\%$ and $\text{MSE} = 1.896$. Based on the criterion proposed by Laurente Blanco and Machaca Hancoco (2020), a MAPE below 10% reflects excellent predictive performance. Accordingly, the $\text{ARIMAX}(3,0,3)$ specification exhibits very high accuracy in forecasting Starbucks stock prices on the testing set. The low MSE further indicates minimal squared prediction errors, implying close alignment between estimated and observed values.

9. Forecasting Results and Model Evaluation

Following testing evaluation, the $\text{ARIMAX}(3,0,3)$ model was applied for forecasting Starbucks stock close prices for 10 days ahead with only weekday trading (excluding weekends and holidays) for January 2-16, 2025. The findings are summarized in Table 5 and in Figure 13.

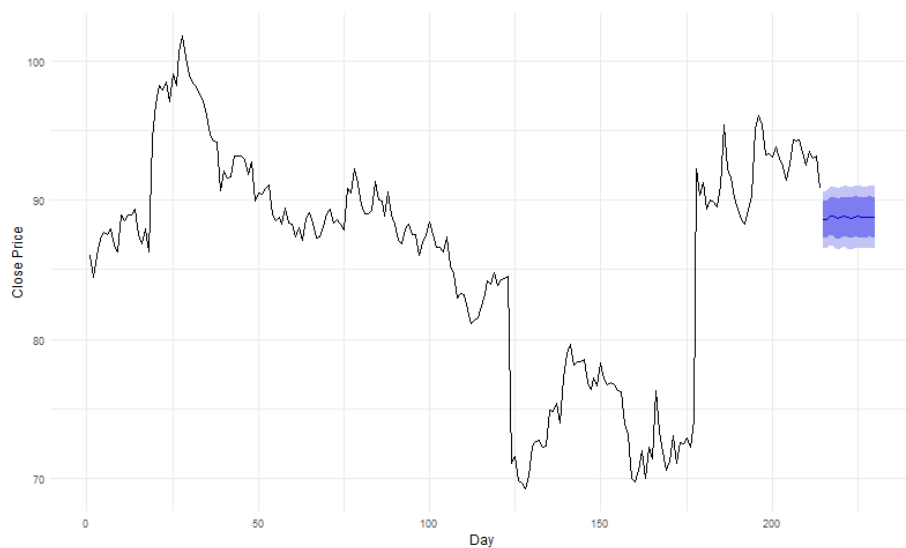


Figure 13. Forecast Plot of Starbucks Close Price Using $\text{ARIMAX}(3,0,3)$

The forecast values, their corresponding actual observations, errors, and squared errors per day are summarized in Table 5.

Table 5. Data Testing vs Actual Close using $\text{ARIMAX}(3,0,3)$

Date	Actual Close	Forecast	Error	Squared Error
1/2/2025	89.786	88.622	1.163	1.354
1/3/2025	90.185	88.869	1.316	1.732
1/6/2025	90.721	88.732	1.988	3.954
1/7/2025	90.438	88.852	1.586	2.516
1/8/2025	90.205	88.774	1.430	2.045
1/10/2025	89.864	88.777	1.086	1.179
1/13/2025	91.091	88.736	2.355	5.547
1/14/2025	90.662	88.790	1.872	3.505
1/15/2025	90.672	88.794	1.878	3.527
1/16/2025	91.987	88.752	3.235	10.466

Forecasts exhibited stable dynamics that closely tracked the observed closing price trend, although prediction errors increased as the forecast horizon extended, reflecting limitations of the $\text{ARIMAX}(3,0,3)$ specification for longer-term projections. The evaluation yielded $\text{MAPE} =$

1.97% and MSE = 3.58, both well below the 10% benchmark, confirming strong predictive performance when incorporating Open, Volume, and Sentiment as exogenous inputs. The relatively low MSE indicates small squared deviations between predicted and actual values, supporting the model's accuracy. Overall, the model provides reliable short-term forecasts with acceptable precision, despite minor deviations over extended horizons.

D. CONCLUSION AND SUGGESTIONS

The analysis of 41,211 tweets containing the phrase “boycott starbucks” between October 2023 and December 2024 showed that negative sentiment was dominant (54%), followed by positive (43.3%) and neutral sentiment (2.7%). Discussion intensity reached its highest level in December 2023, reflecting heightened public attention during the boycott campaign. Frequently occurring keywords, including *genocide*, *company*, and *palestine*, suggest that Starbucks was strongly associated with the broader Israel–Palestine discourse.

Among the candidate models, ARIMAX(3,0,3) incorporating Open price, trading Volume, and sentiment score demonstrated the best performance, achieving the lowest AIC (637.0925) and BIC (674.1182) values while satisfying all diagnostic assumptions. Forecasts for the period January 2–16, 2025 produced a MAPE of 1.97% and an MSE of 3.58, indicating reliable predictive performance for Starbucks stock prices during the study period.

Several limitations should be acknowledged. Sentiment data were collected solely from X (Twitter), VADER may not fully capture contextual nuances such as sarcasm, and the analysis focused on a single boycott-related event. Future studies should integrate data from multiple social media platforms, extend the observation period, and explore advanced forecasting and sentiment-analysis techniques, including LSTM, GRU, BERT, RoBERTa, and multimodal approaches, to improve predictive capability and sentiment representation.

ACKNOWLEDGEMENT

The authors sincerely acknowledge the Faculty of Mathematics and Natural Sciences, Universitas Negeri Jakarta, for funding this research through the competitive research grant program.

REFERENCES

- Akaike, H. (1998). Information Theory and an Extension of the Maximum Likelihood Principle. In K. and K. G. Parzen Emanuel and Tanabe (Ed.), *Selected Papers of Hirotugu Akaike* (pp. 199–213). Springer New York. https://doi.org/10.1007/978-1-4612-1694-0_15
- Albrecht, C.-M., Campbell, C., Heinrich, D., & Lammel, M. (2013). Exploring why consumers engage in boycotts: Toward a unified model. *Journal of Public Affairs*, 13(2), 180–189. <https://doi.org/10.1002/pa.1473>
- Borg, A., & Boldt, M. (2020). Using VADER sentiment and SVM for predicting customer response sentiment. *Expert Systems with Applications*, 162, 113746. <https://doi.org/10.1016/j.eswa.2020.113746>
- Breusch, T. S., & Pagan, A. R. (1979). A Simple Test for Heteroscedasticity and Random Coefficient Variation. *Econometrica*, 47(5), 1287–1294. <https://doi.org/10.2307/1911963>
- Chicco, D., Warrens, M., & Jurman, G. (2021). The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *PeerJ Computer Science*, 7, e623. <https://doi.org/10.7717/peerj-cs.623>
- Cryer, J. D., & Chan, K. S. (2008). *Time Series Analysis: With Applications in R* (2nd ed.). Springer.

- Fekrazad, A., Harun, S. M., & Sardar, N. (2022). Social media sentiment and the stock market. *Journal of Economics and Finance*, 46(2), 397–419. <https://doi.org/10.1007/s12197-022-09575-x>
- Holzhauser, H. M., Lu, X., McLeod, R., & Wang, J. (2016). RiskTRACK: The Five-Factor Model for Measuring Risk Tolerance. *The Journal of Risk Finance*, 17(4), 428–445. <https://doi.org/10.1108/JRF-04-2016-0054>
- Hutto, C., & Gilbert, E. (2014). VADER: A parsimonious rule-based model for sentiment analysis of social media text. *Proceedings of the International AAAI Conference on Web and Social Media*, 8(1), 216–225. <https://doi.org/10.1609/icwsm.v8i1.14550>
- Ibrahim, M., & Alturfi, A. (2022). Impact of exogenous and endogenous variables in time series analysis, exchange rate in turkey as case study. *International Journal of Statistics and Applied Mathematics*, 7(5), 10–14. <https://doi.org/10.22271/maths.2022.v7.i5a.874>
- Investing.com. (2025). Starbucks Corporation Income Statement. <https://id.investing.com/equities/starbucks-corp-income-statement>
- Ivanenko, V. (2023). Assessing Social Media's Impact on Stock Market Predictions: Financial Sentiment Embedding Approaches. *The 13th International Conference on Dependable Systems, Services and Technologies (DESSERT)*, 1–6. <https://doi.org/10.1109/DESSERT61349.2023.10416463>
- Khan, W., Malik, U., Ghazanfar, M. A., Azam, M. A., Alyoubi, K. H., & Alfakeeh, A. S. (2020). Predicting stock market trends using machine learning algorithms via public sentiment and political situation analysis. *Soft Computing*, 24(15), 11019–11043. <https://doi.org/10.1007/s00500-019-04347-y>
- Laurente Blanco, L. F., & Machaca Hanco, R. W. (2020). Modeling and forecasting international tourism demand in Puno - Peru. *Revista Brasileira de Pesquisa Em Turismo*, 14, 34–55. <https://doi.org/10.7784/rbtur.v14i1.1606>
- Mawarnie, H. P., Widodo, & Nugraheni, M. (2024). Sentiment Analysis Twitter (X) Users Towards the Boycott of Pro-Israel Brand Using Support Vector Machine Method with SMOTE Approach. *2024 IEEE International Conference on Control & Automation, Electronics, Robotics, Internet of Things, and Artificial Intelligence (CERIA)*, 1–6. <https://doi.org/10.1109/CERIA64726.2024.10914878>
- Pavan Kumar M R, & Prabhu, J. (2018). Role of sentiment classification in sentiment analysis: a survey. *Annals of Library and Information Studies (ALIS)*, 65(3), 196–209. <https://doi.org/10.56042/alis.v65i3.21598>
- Pokhrel, A., & Adhikari, R. (2023). Leveraging Exogenous Insights: A Comparative Forecast of Paddy Production in Nepal Using ARIMA and ARIMAX Models. *Economic Review of Nepal*, 6(1), 52–69. <https://doi.org/10.3126/ern.v6i1.67970>
- Ranco, G., Aleksovski, D., Caldarelli, G., Grčar, M., & Mozetič, I. (2015). The effects of twitter sentiment on stock price returns. *PLoS ONE*, 10(9). <https://doi.org/10.1371/journal.pone.0138441>
- Santi, V. M., Baihaqi, A., & Siregar, D. (2025). Modeling Poverty In West Java Province Using Negative Binomial Regression With Penalized Smoothly Clipped Absolute Deviation. *Barekeng*, 19(4), 2557–2570. <https://doi.org/10.30598/barekengvol19no4pp2557-2570>
- Santi, V. M., Notodiputro, K. A., Indahwati, & Sartono, B. (2022). Restricted Maximum Likelihood Estimation For Multivariate Linear Mixed Model In Analyzing Pisa Data For Indonesian Students. *Barekeng*, 16(2), 607–614. <https://doi.org/10.30598/barekengvol16iss2pp607-614>
- Santi, V. M., Notodiputro, K. A., Indahwati, Sartono, B., Susetyo, B., & Kurnia, A. (2026). Two-Level Random Effects in Multivariate Linear Mixed Model for Longitudinal Data. *Engineering Letters*, 34(1), 368–379. https://openurl.ebsco.com/EPDB%3Agcd%3A16%3A619312/detailv2?sid=ebsco%3Aplink%3Ascholar&id=ebsco%3Agcd%3A190639054&crl=f&link_origin=www.google.com
- Santi, V. M., Notodiputro, K. A., & Sartono, B. (2019). Variable selection methods applied to the mathematics scores of Indonesian students based on convex penalized likelihood. *Journal of Physics: Conference Series*, 1402(7). <https://doi.org/10.1088/1742-6596/1402/7/077096>
- Santi, V. M., Notodiputro, K. A., Sartono, B., & Rahyu, W. (2021). Generalized Linear Mixed Models by penalized Lasso in modelling the scores of Indonesian students. *Journal of Physics: Conference Series*, 1869(1). <https://doi.org/10.1088/1742-6596/1869/1/012140>

- Santi, V., Wahyu, R., & Hadi, I. (2023). Forecasting the Value of Indonesia's Oil and Gas Imports Using Seasonal Autoregressive Integrated Moving Average Model. *BAREKENG: Jurnal Ilmu Matematika Dan Terapan*, 17, 2047–2058. <https://doi.org/10.30598/barekengvol17iss4pp2047-2058>
- Schwarz, G. (1978). Estimating the Dimension of a Model. *The Annals of Statistics*, 6(2), 461–464. <https://doi.org/10.1214/aos/1176344136>
- Sharma, A., Tiwari, P., Gupta, A., & Garg, P. (2021). Use of LSTM and ARIMAX algorithms to analyze impact of sentiment analysis in stock market prediction. *Intelligent Data Communication Technologies and Internet of Things: Proceedings of ICICI 2020*, 57, 377–394. https://doi.org/10.1007/978-981-15-9509-7_32
- Thadewald, T., & Büning, H. (2007). Jarque–Bera Test and its Competitors for Testing Normality – A Power Comparison. *Journal of Applied Statistics*, 34(1), 87–105. <https://doi.org/10.1080/02664760600994539>
- Tiao, G. C., & Box, G. E. P. (1981). Modeling Multiple Time Series with Applications. *Journal of the American Statistical Association*, 76(376), 802–816. <https://doi.org/10.2307/2287575>
- Wang, J., Ji, T., & Li, M. (2021). A Combined Short-Term Forecast Model of Wind Power Based on Empirical Mode Decomposition and Augmented Dickey-Fuller Test. *Journal of Physics: Conference Series*, 2022(1), 12017. <https://doi.org/10.1088/1742-6596/2022/1/012017>
- Wang, P.-C., & Vo, T. T. H. (2025). Stock price prediction based on dual important indicators using ARIMAX: A case study in Vietnam. *Journal of Intelligent Systems*, 34(1), 20240101. <https://doi.org/10.1515/jisys-2024-0101>
- Wei, W. (2006). Time Series Analysis: Univariate and Multivariate Methods. In *TIME SERIES ANALYSIS* (2nd ed.).