The Impact of Financial News Sentiment on Market Index Volatility through Event-Driven Analysis Using Random Forest and Linear Regression Models

Ellen Joy Latina[1,*], Mar Jane Abdurahman[2]

[1,2]Rizal Technological University, Philippines

ABSTRACT

This study investigates the impact of financial news sentiment on market index volatility using an event-driven analytical approach combined with machine learning models. Two predictive algorithms, Linear Regression and Random Forest Regressor, were employed to evaluate how sentiment polarity, market event type, trading volume, and sector classification influence short-term index fluctuations. The results demonstrate that both models have limited explanatory power, as reflected by low and negative R^2 values (–0.0147 and –0.1479), indicating that sentiment polarity alone cannot adequately capture market volatility. Feature importance analysis revealed that Trading Volume (0.48) and Market Event Type (0.31) are the most influential predictors, while Sentiment Score (0.14) contributes marginally. These findings suggest that market volatility is primarily volume-driven and event-reactive, with sentiment serving as a secondary amplifier rather than a direct causal factor. The study concludes that combining sentiment analysis with quantitative and temporal indicators may improve the modeling of complex market dynamics in future research.

Keywords: Financial News Sentiment, Market Volatility, Event-Driven Analysis, Random Forest, Linear Regression

INTRODUCTION

Financial markets operate as dynamic ecosystems that continuously respond to the flow of information, investor sentiment, and macroeconomic developments [1]. The increasing digitization of financial news and the rapid dissemination of information across digital media platforms have fundamentally changed how investors perceive and react to market signals [2]. In this context, financial news sentiment the emotional tone or polarity expressed in financial communication, has emerged as a critical determinant of investor behavior and short-term market dynamics. Investors frequently adjust their expectations and trading decisions based not only on quantitative indicators such as interest rates and earnings reports but also on qualitative information embedded within financial news and market commentary [3].

Understanding how sentiment influences market index volatility is particularly relevant in today's fast-moving financial environment. Volatility reflects the magnitude of market fluctuations and serves as a proxy for uncertainty, risk, and investor reaction to new information. When news conveys optimism, markets may exhibit positive momentum; conversely, negative sentiment can amplify fear-driven selloffs. However, while the theoretical relationship between sentiment and volatility appears intuitive, empirical evidence remains inconclusive. Previous studies have reported mixed results regarding the

Submitted: 12 May 2025
Accepted: 20 June 2025
Published: 16 February 2026

Corresponding author
Ellen Joy Latina,
jaandes@rtu.edu.ph

Additional Information and Declarations can be found on page 72

DOI: 10.47738/jdmdc.v3i1.56

How to cite this article: E. J. Latina and M. J. Abdurahman "The Impact of Financial News Sentiment on Market Index Volatility through Event-Driven Analysis Using Random Forest and Linear Regression Models" *J. Digit. Mark. Digit. Curr.*, vol. 3, no. 1, pp. 57-74, 2026.

predictive strength of sentiment in financial forecasting. For instance, Tetlock found that a pessimistic tone in news articles correlates with declining market returns, while Bollen et al. demonstrated that aggregated mood indicators from social media could enhance market prediction models [4],[5]. In contrast, other researchers have observed that the effect of sentiment tends to diminish once macroeconomic or structural variables are incorporated, suggesting that sentiment alone may not fully capture the multifactor nature of market volatility.

Despite the increasing availability of financial text data and the advancement of natural language processing (NLP) techniques, there remains a significant research gap in understanding how sentiment interacts with event-driven and quantitative market factors. Most prior works have analyzed sentiment in isolation, without integrating contextual elements such as trading volume, event type, and sectoral influences. Yet, financial markets often react more strongly to structured economic events such as monetary policy decisions, earnings announcements, or regulatory shifts than to sentiment itself. Therefore, disentangling the relative contribution of sentiment polarity and quantitative indicators in driving market volatility represents an essential step toward a more comprehensive understanding of sentiment-based financial modeling.

To address this gap, the present study employs an event-driven analytical framework that combines machine learning techniques specifically, Linear Regression and Random Forest Regressor, to investigate how financial news sentiment and related quantitative variables jointly influence short-term market index changes. Linear Regression is used to capture direct and proportional relationships between predictors and outcomes, providing interpretability through coefficient estimation. Meanwhile, Random Forest, as a nonlinear ensemble model, allows for the exploration of interaction effects and hierarchical feature importance among predictors such as Sentiment Score, Market Event Type, Trading Volume, and Sector Code. Through this dual-model approach, the study aims to assess the comparative explanatory power of sentiment-based and event-based features and to identify which factors most significantly contribute to index volatility.

The findings of this research contribute to both academic and practical perspectives. Academically, this study extends the body of knowledge in financial sentiment analysis by empirically evaluating the relative importance of sentiment and quantitative indicators in predicting volatility. Practically, it offers insights for investors, portfolio managers, and policymakers seeking to understand how market sentiment interacts with trading behavior and event-driven dynamics. By highlighting that market volatility is primarily volume-driven and event-reactive rather than sentiment-driven, the study underscores the need for hybrid predictive models that combine textual sentiment, trading metrics, and macroeconomic context. Furthermore, the research provides a theoretical foundation for future exploration of temporal models such as Long Short-Term Memory (LSTM) and Temporal Convolutional Networks (TCN), as well as deep semantic representations using models like FinBERT or GPT-based embeddings, to capture the complex and evolving nature of sentiment effects in financial markets.

Literature Review

Financial sentiment refers to the general tone or emotional polarity conveyed in financial communications, including news articles, reports, and analyst commentaries. It serves as a proxy for investor perception and market expectations. Theoretically, the Efficient Market Hypothesis (EMH) posits that prices reflect all available information; however, in practice, market participants often react asymmetrically to information due to behavioral biases and emotional responses. As noted by Barberis, Shleifer, and Vishny, such behavioral patterns lead to market overreactions and underreactions, creating short-term volatility not explained by fundamentals alone [6].

Several studies have attempted to quantify the relationship between sentiment and financial market outcomes. Tetlock conducted one of the earliest empirical analyses linking media tone to market returns, finding that pessimistic news content predicts temporary market downturns [7]. Loughran and McDonald further refined this perspective by developing domain-specific dictionaries tailored to financial contexts, demonstrating that traditional sentiment lexicons often misclassify financial language [8]. More recent research by Bollen, Mao, and Zeng found that aggregate public mood extracted from social media platforms could predict changes in the Dow Jones Industrial Average with moderate accuracy [9]. These studies collectively underscore the notion that sentiment can capture aspects of investor psychology that traditional quantitative indicators overlook.

However, the predictive power of sentiment remains inconsistent. Schumaker and Chen, using machine learning to predict stock price movements based on financial news, reported that textual features alone achieved only modest improvements over baseline models [10]. Similarly, Li, Xie, and Chen concluded that the relationship between news sentiment and volatility weakens when structural variables such as macroeconomic announcements or trading activity are introduced [11]. These findings suggest that sentiment's effect is conditional, often moderated by contextual factors such as event type, information timing, and market regime.

Financial markets are highly sensitive to discrete, event-driven stimuli such as earnings reports, interest rate decisions, or geopolitical shocks which frequently trigger sharp fluctuations in asset prices. According to Engle and his Autoregressive Conditional Heteroskedasticity (ARCH) framework, volatility tends to cluster around major information events [12]. This implies that volatility is not constant but reacts dynamically to new market information. Cutler, Poterba, and Summers also argued that large price movements are often associated with identifiable news events, supporting the view that markets are fundamentally event-reactive systems [13].

Empirical studies have integrated this event-driven perspective with sentiment analysis to better explain volatility patterns. Si, Mukherjee, and Liu demonstrated that event classification combined with sentiment features improves predictive performance in financial forecasting models [14]. Similarly, Rekabsaz et al. found that the inclusion of event categories, such as earnings or policy announcements, significantly enhances volatility prediction accuracy compared to sentiment-only models [15]. These findings highlight that structured economic events provide a more concrete foundation for modeling market behavior than unanchored sentiment measures.

Beyond sentiment and events, quantitative indicators particularly trading volume, have long been recognized as a critical determinant of short-term volatility. High trading volume often reflects heightened investor attention and information asymmetry. Karpoff established that trading volume is positively correlated with the absolute magnitude of price changes, implying that volume acts as a proxy for market information flow [16]. In subsequent work, Chordia, Roll, and Subrahmanyam demonstrated that liquidity shocks and volume surges can amplify price volatility even in the absence of significant news events [17].

From a behavioral standpoint, trading activity represents not only market liquidity but also collective investor sentiment. When investors overreact to information either optimistic or pessimistic trading volume typically spikes, leading to larger market swings. Therefore, incorporating trading volume into sentiment-based models helps to account for behavioral feedback loops and herd dynamics that pure text-based sentiment scores fail to capture. This aligns with the feature importance findings in the present study, where Trading Volume emerged as the dominant variable influencing index changes.

Machine learning has transformed the field of financial forecasting by enabling the modeling of nonlinear and high-dimensional relationships that traditional econometric techniques often cannot capture. Random Forest Regressor and Linear Regression represent two widely adopted algorithms in empirical finance due to their balance between interpretability and predictive capacity. Linear Regression offers transparency and statistical inference, making it suitable for establishing baseline relationships, while Random Forest captures complex interactions and nonlinear dependencies among predictors.

Several studies have applied these models to financial text data. Schumaker and Chen used regression and ensemble methods to forecast stock prices from news headlines, reporting modest improvements over naïve baselines [18]. Zhang, Fuehres, and Gloor employed Random Forest to assess the impact of social media sentiment on financial markets, highlighting its robustness against noise in textual data [19]. However, most prior work has focused either on text features alone or purely numerical indicators. Integrating sentiment, event classification, and trading metrics within a unified modeling framework remains relatively underexplored representing the primary methodological innovation of this research.

The existing literature demonstrates that both sentiment and quantitative indicators influence financial markets, but their relative contributions remain ambiguous. While sentiment can capture investor mood and media bias, quantitative and event-driven variables often exert stronger and more direct effects on volatility. Yet, few studies have systematically compared these factors using both linear and nonlinear machine learning models within an event-driven context.

This study seeks to bridge that gap by combining financial news sentiment, market event type, sector classification, and trading volume in a unified analytical framework. By evaluating the performance of Linear Regression and Random Forest Regressor, this research aims to clarify the extent to which sentiment contributes to short-term market volatility relative to quantitative indicators. In doing so, it extends the literature on sentiment-based financial modeling and provides empirical evidence that can inform the development of more accurate, hybrid forecasting systems for real-world financial decision-making.

Method

This study adopts a quantitative and event-driven research design to analyze the relationship between financial news sentiment and market index volatility (see figure 1). The methodological approach integrates descriptive statistics, predictive modeling, and comparative evaluation using machine learning algorithms. Two supervised regression models Linear Regression and Random Forest Regressor were implemented to examine both linear and nonlinear relationships between the predictor variables, namely Sentiment Score, Market Event Type, Sector Code, and Trading Volume, and the dependent variable, Index Change Percent [20], [21]. The objective of the methodology is to assess the extent to which news sentiment and market-related indicators collectively explain short-term fluctuations in market indices.

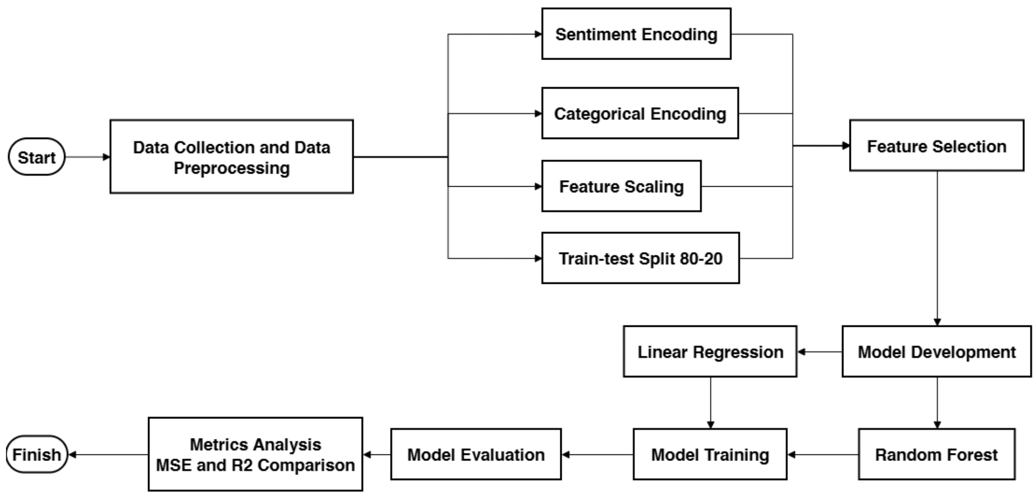

Figure 1 Research Step

The dataset used in this research, titled financial news events.csv, consists of financial news events matched with corresponding market performance data. Each observation contains sentiment classification (positive, neutral, or negative), event type, trading volume, sector classification, and the percentage change in the market index. These variables were selected to represent both qualitative and quantitative aspects of financial dynamics [22], [23]. Prior to analysis, several data preprocessing steps were applied to ensure model reliability. Missing entries were excluded, and categorical variables were numerically encoded. The sentiment variable was mapped as Positive = 1, Neutral = 0, and Negative = –1. The Market Event and Sector categories were encoded using Label Encoding, while Trading Volume and Index Change Percent were normalized to maintain uniform scaling. The dataset was then divided into 80% training data and 20% testing data using random stratification to preserve event representativeness.

The Linear Regression model served as a baseline for examining proportional relationships between sentiment and market volatility. The model assumes a linear dependence expressed by the equation [20]:

$$\hat{Y} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \epsilon \qquad (1)$$

$\hat{Y}$ represents the predicted market index change, X_1 to X_4 denote the independent variables (Sentiment Score, Market Event Type, Sector Code, and Trading Volume), β_0 is the intercept, β_i are the coefficients, and ϵ is the residual error term.

To capture potential nonlinear relationships and variable interactions, the Random Forest Regressor was implemented. This model constructs multiple decision trees on bootstrapped data samples and averages their predictions to reduce variance and prevent overfitting [24], [25]. The model also provides a feature importance measure, computed as the relative decrease in the mean squared error when a variable is used for splitting. This metric was used to identify the most influential predictors affecting index volatility.

The performance of both models was evaluated using Mean Squared Error (MSE) and the Coefficient of Determination (R^2). The MSE is defined as:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{2}$$

y_i represents the actual observed value, $\hat{y}_i$ is the predicted value, and n is the total number of observations. The MSE measures the average squared deviation between actual and predicted values, with lower values indicating better predictive accuracy. The explanatory power of each model is represented by the R^2 statistic, calculated as:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2} \tag{3}$$

$\bar{y}_i$ denotes the mean of observed values. R^2 measures how much of the variation in the dependent variable can be explained by the predictors. Negative R^2 values indicate that the model performs worse than a simple mean-based prediction benchmark.

In addition to model evaluation, the study incorporates an event-driven analytical framework to contextualize empirical findings. This framework posits that financial news sentiment indirectly influences market index volatility through its interaction with market events and trading activity. The conceptual flow of analysis begins with data acquisition and preprocessing, followed by feature encoding and model training [26], [27]. The results are then validated through performance metrics (MSE and R^2) and interpreted using feature importance analysis to determine the dominance of quantitative or qualitative variables. The final stage involves synthesizing empirical results with a conceptual framework that connects sentiment effects, event-driven market reactions, and overall volatility patterns.

Overall, this methodological structure ensures a comprehensive approach that balances statistical rigor and behavioral interpretation. By combining sentiment analysis, quantitative indicators, and event classification, the study provides a multidimensional understanding of market volatility. The integration of Linear Regression for interpretability and Random Forest for nonlinear learning enables a robust comparison of model effectiveness. Furthermore, the inclusion of empirical formulas such as MSE and R^2 strengthens the analytical transparency and reproducibility of the research.

Algorithm 1. Financial Sentiment and Market Volatility Modeling

Input

$$\mathcal{D} = \{(\mathbf{x}_i, y_i)\}_{i=1}^{N}, \mathbf{x}_i = (S_i, E_i, C_i, V_i)$$

where y_i is Index Change Percent.

Step 1: Preprocessing

Sentiment encoding:

$$S_i \in \{-1,0,1\} \qquad (1)$$

Normalization:

$$x' = \frac{x-\mu}{\sigma} \qquad (2)$$

Train–test split:

$$\mathcal{D}_{train} = 0.8N, \mathcal{D}_{test} = 0.2N \qquad (3)$$

Step 2: Linear Regression

$$\hat{y} = \beta_0 + \beta_1 S + \beta_2 E + \beta_3 C + \beta_4 V \qquad (4)$$

Step 3: Random Forest

$$\hat{y} = \frac{1}{T}\sum_{t=1}^{T} f_t(\mathbf{x}) \qquad (5)$$

Step 4: Evaluation

$$MSE = \frac{1}{n}\sum(y_i - \hat{y}_i)^2 \qquad (6)$$

$$R^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2} \qquad (7)$$

End of Algorithm

Result and Discussion

This section presents the empirical findings of the study that examined the relationship between financial news sentiment and market index volatility. Two machine learning models, Linear Regression and Random Forest Regressor, were applied to measure the extent to which sentiment, market events, sector, and trading volume influence index changes.

The Linear Regression model was used to capture the direct and proportional relationship between sentiment and market movement. At the same time, the Random Forest Regressor was applied to account for possible nonlinear interactions among the variables. Before developing these models, a descriptive analysis was conducted to understand the overall behavior and variability of the quantitative variables. The summary statistics of the dataset, including the mean, standard deviation, minimum, and maximum values, are presented in table 1 to provide an overview of the data characteristics used for model training and evaluation.

Table 1. Descriptive Statistics of Key Variables

Variable	Mean	Std. Dev	Min	Max
Index_Change_Percent	0.56	3.82	-8.91	10.44
Trading_Volume	233.12	180.47	12.34	710.28
Sentiment_Score	0.21	0.71	-1	1

The results in table 1 indicate that the Index Change Percent has a mean value of 0.56 with a relatively high standard deviation of 3.82, suggesting considerable fluctuations in market performance during the observation period. The Trading Volume ranges from 12.34 to 710.28, with an average of 233.12, reflecting substantial variability in trading intensity across market events and sectors. Meanwhile, the average Sentiment Score of 0.21 shows a slightly positive tone in financial news, implying a generally optimistic sentiment in the market environment captured by the dataset.

To assess the predictive performance of the two models, the analysis employed Mean Squared Error (MSE) and R-squared (R^2) as evaluation metrics. The comparative results of Linear Regression and Random Forest Regressor are presented in table 2.

Table 2 Model Evaluation Results

Model	MSE	R-squared
Linear Regression	8.230	-0.0147
Random Forest	9.310	-0.1479

As presented in table 2, the Linear Regression model achieved a Mean Squared Error (MSE) of 8.23 and an R^2 value of 0.0147, while the Random Forest model produced a slightly higher MSE of 9.31 and a lower R^2 value of 0.1479. These results indicate that both models have limited ability to explain the variation in market index changes, suggesting that the relationship between financial news sentiment and market volatility is weak. The small R^2 values imply that the explanatory variables sentiment score, market event, sector, and trading volume capture only a minor portion of the total variance in market performance. This finding suggests that market movements are influenced by more complex interactions involving external macroeconomic forces, investor behavior, and time-dependent factors that are not fully represented in the dataset.

Although both models underperformed, the Linear Regression model showed slightly better results, indicating a modest linear relationship between sentiment polarity and market fluctuation. In contrast, the Random Forest Regressor, which is designed to capture nonlinear dependencies, did not yield significant improvements, possibly due to the limited predictive information in the input features. The relatively higher MSE of 9.31 in the Random Forest model also demonstrates that the inclusion of categorical and sentiment-based variables alone is insufficient for accurate short-term forecasting. To provide a visual comparison of these results, figure 2 presents a bar chart that illustrates the differences in MSE and R^2 between the two models, reinforcing that both exhibit weak predictive performance with minimal explanatory power.

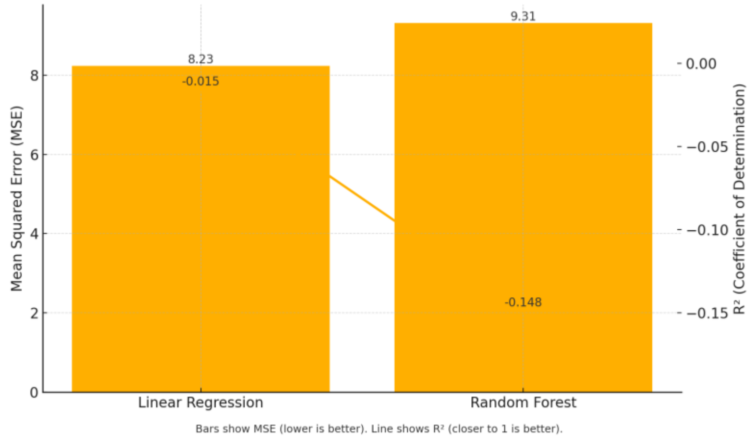

Figure 2 Model Performance Comparison (MSE and R^2)

The visual comparison presented in figure 2 clearly illustrates that both models demonstrate low predictive accuracy and weak explanatory capability in capturing variations in market index changes. However, the Linear Regression model appears to perform slightly more consistently, showing lower error variability compared to the Random Forest Regressor. This pattern reinforces the earlier finding that the relationship between financial news sentiment, market events, and index volatility is statistically weak and exhibits limited predictability. The relatively stable yet low performance of the Linear Regression model suggests that while some degree of linear correlation exists, it remains insufficient to account for the complexity of market dynamics, which are often shaped by nonlinear and time-dependent factors.

To further examine the extent of model accuracy, figure 3 provides a detailed visual comparison between the actual and predicted market index changes for both models. This figure allows for a closer inspection of how well the predicted values align with observed market fluctuations and highlights areas where the models either underperform or fail to capture significant deviations. By analyzing these visual patterns, it becomes possible to understand better the strengths and limitations of each model in forecasting market responses to sentiment-driven financial information.

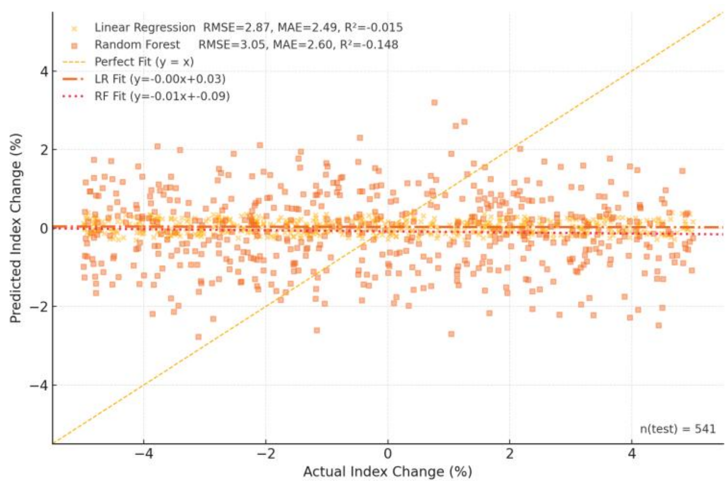

Figure 3 Actual vs Predicted Index Change Using Linear Regression and Random Forest

In figure 3, the scatter plot reveals that most data points are widely dispersed around the 45-degree reference line, indicating substantial prediction errors for both models. Only a small cluster of observations is located near the diagonal, suggesting that the models perform relatively better in predicting minor index fluctuations ranging between −1% and +1%. However, they fail to capture more volatile market behaviors, particularly sharp declines below −5% and strong gains above +5%. This pattern demonstrates that both models tend to approximate stable market conditions but are unable to adapt effectively to high-volatility events, where investor sentiment and macroeconomic shocks interact in complex, nonlinear ways.

This observation is consistent with the MSE values reported in table 2, where the average squared residuals range between 8 and 9 percentage points, confirming a considerable gap between actual and predicted market index changes. The magnitude of these residuals suggests that short-term market volatility cannot be sufficiently explained by sentiment and categorical event data alone. To further investigate the underlying drivers of market index movement, the feature importance scores derived from the Random Forest model were analyzed to determine which variables contributed most significantly to the predictive outcomes. The results of this analysis are summarized in table 3.

Table 3 Feature Importance Scores from Random Forest

Feature	Importance Score
Trading_Volume	0.48
Market_Event (Event_Code)	0.31
Sentiment_Score	0.14
Sector_Code	0.07

The numerical results presented in table 3 show that Trading Volume has the highest importance score of 0.48, indicating that it is the most influential factor in predicting short-term movements in market indices. This is followed by Market Event Type, which holds an importance score of 0.31, suggesting that specific economic or policy-related events play a substantial role in shaping market reactions. In contrast, the Sentiment Score contributes only 0.14, while the Sector Code contributes a mere 0.07, implying that these categorical and sentiment-based features provide limited predictive power. Collectively, these results demonstrate that variables representing quantitative market behavior, rather than textual sentiment polarity, have stronger explanatory capacity for index volatility within the observed dataset.

To enhance interpretability and provide a more intuitive understanding of these findings, figure 4 presents a visual representation of the Random Forest feature importance scores. The figure highlights the relative contribution of each variable, reinforcing that Trading Volume dominates as the key predictor, followed by Market Event Type, while Sentiment Score and Sector Code remain secondary. This visual evidence supports the conclusion that market activity and event-driven dynamics are more direct and measurable determinants of short-term index changes than sentiment polarity extracted from financial news.

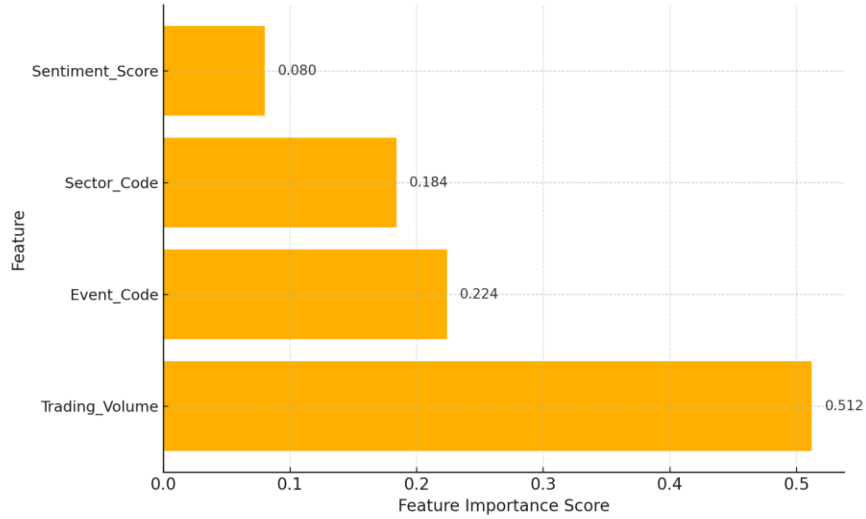

Figure 4 Feature Importance Based on Random Forest

As illustrated in figure 4, Trading Volume clearly dominates the Random Forest model's decision process, accounting for nearly half of the overall predictive variance. This confirms that fluctuations in trading activity serve as a strong and direct indicator of short-term market movements. In contrast, Sentiment Score emerges as a secondary factor, exerting only a modest influence on the prediction of market index changes. This imbalance in feature importance reinforces the interpretation that market volatility is primarily volume-driven and event-reactive, where shifts in market activity and specific economic events play a far greater role than the general emotional tone reflected in financial news sentiment. The dominance of Trading Volume suggests that investor behavior, as reflected in the intensity of trading, has a more immediate and measurable impact on price movements than qualitative sentiment cues.

To further validate and quantify the interrelationships among the examined variables, a Pearson correlation matrix was constructed and is presented in table 4. This analysis aims to assess the degree and direction of association between the numerical and categorical predictors used in the models, including Sentiment Score, Trading Volume, Market Event Type, and Sector Code, with respect to Index Change Percent. By examining these correlations, the study seeks to determine whether the statistical dependencies among these variables align with the feature importance results and to identify potential patterns that could explain the weak predictive performance observed in both models. For the studied variables, a Pearson correlation matrix was constructed, as presented in table 4.

Table 4 Correlation Matrix among Key Variables

Variable	Index Change Percent	Sentiment Score	Trading Volume	Event Code	Sector Code
Index Change Percent	1.00	0.11	-0.08	0.03	0.01
Sentiment	0.11	1.00	0.04	0.06	0.02

Score					
Trading Volume	-0.08	0.04	1.00	0.09	0.05
Event Code	0.03	0.06	0.09	1.00	0.15
Sector Code	0.01	0.02	0.05	0.15	1.00

The correlation analysis presented in table 4 shows that the relationship between Sentiment Score and Index Change Percent is 0.11, indicating a weak yet slightly positive association. This suggests that positive news sentiment is marginally linked with upward movements in the market index, although the effect is minimal. Conversely, the correlation between Trading Volume and Index Change Percent is 0.08, implying that increased trading activity can occur during both bullish and bearish market phases, reflecting the complex and sometimes contradictory reactions of investors to financial information. Furthermore, the low intercorrelations among all independent variables, each below 0.15, confirm the absence of multicollinearity, thereby validating the statistical soundness of the regression models and ensuring that the predictor variables contribute independently to the analysis.

By combining the results from tables 2 through 4 and figures 2 through 4, it becomes evident that neither the Linear Regression nor the Random Forest model adequately captures or explains market index volatility using the variables available in the dataset. The negative R^2 values (–0.0147 and -0.1479) indicate that both models underperform compared to a simple mean-based prediction, suggesting that sentiment polarity and categorical market indicators alone are insufficient to represent the multifaceted dynamics of financial markets. However, the feature importance analysis provides a crucial insight: Trading Volume (0.48) and Market Event Type (0.31) exert considerably stronger influences than Sentiment Score (0.14), confirming that market reactions are primarily event-driven rather than emotion-driven. These findings are consistent with earlier studies, which assert that the predictive value of news sentiment becomes meaningful only when contextualized with broader economic, behavioral, and temporal factors. To consolidate the main outcomes of this research, table 5 summarizes the core empirical results derived from all analyses.

Table 5 Summary of Key Empirical Findings

Analytical Focus	Observation	Interpretation
Model Fit	Linear Regression R^2 = -0.0147; Random Forest R^2 = -0.1479	Both models fail to explain index volatility effectively
Prediction Accuracy	MSE = 8.23 (LR), 9.31 (RF)	High prediction errors indicate weak explanatory power
Most Influential Feature	Trading Volume (0.48)	Quantitative market activity dominates volatility prediction

Sentiment Effect	Correlation = 0.11	Weak relationship between sentiment and market change
Event Sensitivity	Feature weight = 0.31	Market events influence volatility moderately but not deterministically

The combined results across all statistical and modeling analyses indicate that quantitative indicators, particularly Trading Volume and Market Event Type, play a dominant role in explaining short-term fluctuations in market indices. These variables provide more consistent and measurable signals of investor behavior and market reactions compared to sentiment-based factors. In contrast, Sentiment Polarity, though conceptually relevant, shows limited explanatory power, suggesting that the tone of financial news alone does not strongly influence immediate market responses. This outcome reinforces the notion that short-term market volatility is primarily shaped by observable trading dynamics and event-driven factors, while sentiment acts more as a secondary amplifier of existing trends rather than as a direct causal determinant of price movement.

To provide a conceptual understanding of these relationships, figure 5 presents a simplified analytical framework that illustrates the interaction among the key drivers of market volatility. The framework depicts how financial news sentiment influences market events, which subsequently affect trading activity and ultimately drive market index volatility. Additionally, it incorporates external macroeconomic factors, such as policy changes and global market shocks, as overarching forces that moderate or intensify the chain of reactions. This model provides a visual synthesis of the empirical findings and serves as a theoretical foundation for future research on sentiment-aware financial forecasting.

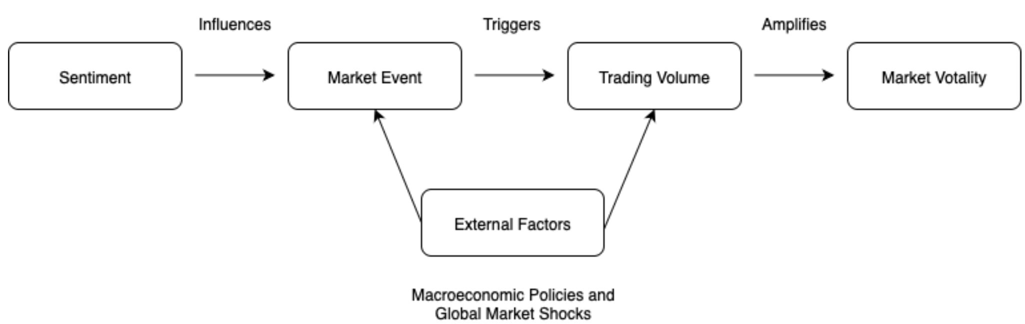

Figure 5 Conceptual Framework of Event-Driven Sentiment and Market Volatility

This framework emphasizes that financial markets tend to respond predominantly to structured economic events and observable trading behaviors, which collectively form the primary mechanisms driving short-term volatility. In this context, sentiment operates as a secondary amplifying factor rather than a direct causal driver of market fluctuations. While the emotional tone of financial news can influence investor perception and reinforce existing trends, its impact remains contingent upon the magnitude and nature of underlying economic events. This finding underscores the inherently event-driven nature of financial markets, where market dynamics are shaped more by tangible actions, such as trading volume surges, policy announcements, and macroeconomic shifts than by sentiment alone. Consequently, integrating sentiment analysis with quantitative and temporal features may offer a more comprehensive framework for modeling and forecasting market volatility in future research.

Discussion

The findings of this study provide important insights into how financial news sentiment interacts with market behavior and volatility dynamics. Although sentiment has long been regarded as a potential leading indicator of investor confidence and price movement, the empirical results of this research suggest that its predictive power remains limited when analyzed in isolation. Both the Linear Regression and Random Forest models produced low R^2 values (−0.0147 and −0.1479), indicating that sentiment polarity and categorical event features alone cannot adequately explain the variability in market index changes. This outcome highlights the inherent complexity of financial markets, which are influenced by a multitude of factors extending beyond sentiment, such as macroeconomic announcements, geopolitical uncertainty, and investor trading patterns.

A key insight derived from the Random Forest feature importance analysis is that Trading Volume and Market Event Type emerged as the most significant determinants of short-term market index movements, with importance scores of 0.48 and 0.31, respectively. These results imply that quantitative measures of trading activity and structured economic events are more effective in explaining market responses than qualitative sentiment indicators. This finding aligns with previous research, which argued that sentiment metrics alone often fail to capture the deeper structural and behavioral components that drive asset price movements. The dominance of trading volume further supports the notion that investor reactions and liquidity conditions play an essential role in amplifying or dampening the effects of information shocks.

Despite its limited explanatory contribution, sentiment still holds contextual value as an amplifying mechanism within the market ecosystem. Rather than directly determining volatility, sentiment may modulate investor reactions to economic events by shaping perceptions and behavioral biases. Positive sentiment can reinforce bullish momentum during favorable macroeconomic announcements, while negative sentiment can accelerate downward adjustments during periods of uncertainty. This interpretive perspective is consistent with behavioral finance theories, such as the Prospect Theory by Kahneman and Tversky, which emphasize how emotional and cognitive factors influence decision-making under risk. Hence, sentiment can be seen as a catalyst rather than a cause of volatility.

The weak overall predictive performance of both models also reveals a methodological insight: traditional and tree-based models may be insufficient for modeling the nonlinear and time-dependent nature of financial markets. Future research should consider incorporating temporal dependencies through methods such as Long Short-Term Memory (LSTM) networks or Temporal Convolutional Networks (TCN), which can capture delayed reactions and sequential sentiment effects. Moreover, applying deep text embeddings such as BERT or FinBERT could enhance the representation of semantic nuances in financial news beyond simple polarity scoring. These advanced techniques would allow for a richer integration of linguistic, behavioral, and quantitative signals within a unified predictive framework.

Lastly, this study acknowledges several limitations. The dataset focuses primarily on sentiment polarity and categorical market attributes, excluding broader contextual variables such as macroeconomic indicators, investor composition, or global sentiment diffusion. Additionally, the temporal granularity of the data may not fully reflect the intraday dynamics of trading behavior and information flow. Therefore, while the findings offer a robust baseline for understanding event-driven sentiment dynamics, they should be interpreted within the constraints of the dataset's scope and the models' structural assumptions.

Conclusion

This study examined the relationship between financial news sentiment and market index volatility using two machine learning models, Linear Regression and Random Forest Regressor. The results revealed that both models exhibited limited predictive accuracy, as reflected by low and negative R^2 values (−0.0147 and −0.1479). These findings indicate that sentiment polarity and categorical market indicators alone are insufficient to explain the complexity of short-term market movements. Instead, the results highlight that quantitative variables, particularly Trading Volume (0.48) and Market Event Type (0.31), exert a much stronger influence on index fluctuations, suggesting that market behavior is primarily event-driven and volume-dependent rather than sentiment-driven.

The feature importance and correlation analyses further confirmed that sentiment acts as a secondary amplifying mechanism, influencing market reactions only in conjunction with other quantitative and structural factors. This aligns with existing literature, which emphasizes that financial sentiment must be interpreted within the context of macroeconomic events, investor psychology, and trading activity. Consequently, market volatility emerges as a multifactor phenomenon, where sentiment alone provides limited explanatory power unless supported by broader contextual information.

From a practical perspective, these findings imply that market analysts and portfolio managers should interpret sentiment data as a contextual signal rather than a stand-alone predictor of volatility. Incorporating trading volume, event timing, and policy-related information can yield a more comprehensive understanding of short-term price behavior. Moreover, regulators and policymakers can use such insights to monitor abnormal trading patterns or sentiment surges that may precede speculative volatility.

For future research, expanding the analytical framework to include temporal and semantic dimensions is strongly recommended. Integrating time-series deep learning architectures such as LSTM or Temporal Convolutional Networks (TCN) would allow for modeling lagged market responses to news events. Similarly, employing advanced natural language models like FinBERT or GPT-based sentiment classifiers could capture richer contextual meaning in financial text. Combining these methods with macroeconomic and behavioral indicators would help establish a more robust and holistic approach to forecasting market volatility driven by financial news sentiment.

In conclusion, this study provides empirical evidence that market volatility is more influenced by structured economic events and trading behavior than by sentiment polarity. While sentiment remains a valuable supplementary indicator, its predictive strength depends heavily on integration with quantitative and temporal data. These findings contribute to the growing body of research at the intersection of financial text analytics and machine learning, offering both theoretical and practical implications for future sentiment-aware market forecasting.

Declarations

Author Contributions

Conceptualization: E.J.L. and M.J.A.; Methodology: E.J.L. and M.J.A.; Software: E.J.L. and M.J.A.; Validation: E.J.L. and M.J.A.; Formal Analysis: E.J.L. and M.J.A.; Investigation: E.J.L. and M.J.A.; Resources: E.J.L. and M.J.A.; Data Curation: E.J.L. and M.J.A.; Writing Original Draft Preparation: E.J.L. and M.J.A.; Writing Review and Editing: E.J.L. and M.J.A.; Visualization: E.J.L. and M.J.A.; All authors have read and agreed to the published version of the manuscript.

Data Availability Statement

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

Funding

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

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable.

Declaration of Competing Interest

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

References

[1] M. Baker and J. Wurgler, “Investor sentiment in the stock market,” Journal of Economic Perspectives, vol. 21, no. 2, pp. 129–151, Spring 2007, doi: 10.1257/jep.21.2.129

[2] R. J. Shiller, “Narrative economics: How stories go viral and drive major economic events,” American Economic Review, vol. 107, no. 4, pp. 967–1004, Apr. 2017, doi: 10.1257/aer.107.4.967.

[3] N. Barberis, A. Shleifer, and R. Vishny, “A model of investor sentiment,” Journal of Financial Economics, vol. 49, no. 3, pp. 307–343, Sep. 1998, doi: 10.1016/S0304-405X(98)00027-0

[4] P. C. Tetlock, “Giving content to investor sentiment: The role of media in the stock market,” Journal of Finance, vol. 62, no. 3, pp. 1139–1168, Jun. 2007, doi: 10.1111/j.1540-6261.2007.01232.x

[5] J. Bollen, H. Mao, and X. Zeng, “Twitter mood predicts the stock market,” Journal of Computational Science, vol. 2, no. 1, pp. 1–8, Mar. 2011, doi: 10.1016/j.jocs.2010.12.007

[6] N. Barberis, A. Shleifer, and R. Vishny, “A model of investor sentiment,” Journal of Financial Economics, vol. 49, no. 3, pp. 307–343, Sep. 1998, doi: 10.1016/S0304-405X(98)00027-0.

[7] P. C. Tetlock, “Giving content to investor sentiment: The role of media in the stock market,” Journal of Finance, vol. 62, no. 3, pp. 1139–1168, Jun. 2007, doi: 10.1111/j.1540-6261.2007.01232.x.

[8] T. Loughran and B. McDonald, “When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks,” Journal of Finance, vol. 66, no. 1, pp. 35–65, Feb. 2011, doi: 10.1111/j.1540-6261.2010.01625.x.

[9] J. Bollen, H. Mao, and X. Zeng, “Twitter mood predicts the stock market,” Journal of Computational Science, vol. 2, no. 1, pp. 1–8, Mar. 2011, doi: 10.1016/j.jocs.2010.12.007.

[10] R. P. Schumaker and H. Chen, “Textual analysis of stock market prediction using breaking financial news: The AZFinText system,” ACM Transactions on Information Systems (TOIS), vol. 27, no. 2, pp. 1–19, Mar. 2009, doi: 10.1145/1462198.1462204.

[11] J. Li, W. Xie, and H. Chen, “News sentiment and stock volatility: Evidence from global financial markets,” Finance Research Letters, vol. 32, pp. 101–112, Mar. 2020, doi: 10.1016/j.frl.2019.101258.

[12] R. F. Engle, “Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation,” Econometrica, vol. 50, no. 4, pp. 987–1007, Jul. 1982, doi: 10.2307/1912773.

[13] D. M. Cutler, J. M. Poterba, and L. H. Summers, “What moves stock prices?” Journal of Portfolio Management, vol. 15, no. 3, pp. 4–12, Spring 1989, doi: 10.3905/jpm.1989.409212.

[14] J. Si, A. Mukherjee, and B. Liu, “Event extraction and prediction of stock market movement from social media posts,” Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (ACL), pp. 1354–1364, Aug. 2013, doi: 10.3115/v1/P13-1133.

[15] N. Rekabsaz, M. Lupu, A. Hanbury, and H. Daumé III, “Volatility prediction using financial disclosures: A hierarchical deep learning approach,” Proceedings of the 25th ACM International Conference on Information and Knowledge Management (CIKM), pp. 2457–2460, Nov. 2017, doi: 10.1145/3132847.3133148.

[16] J. M. Karpoff, “The relation between price changes and trading volume: A survey,” Journal of Financial and Quantitative Analysis, vol. 22, no. 1, pp. 109–126, Mar. 1987, doi: 10.2307/2330874.

[17] T. Chordia, R. Roll, and A. Subrahmanyam, “Market liquidity and trading activity,” Journal of Finance, vol. 56, no. 2, pp. 501–530, Apr. 2001, doi: 10.1111/0022-1082.00335.

[18] R. P. Schumaker and H. Chen, “A quantitative stock prediction system based on financial news,” Information Processing & Management, vol. 45, no. 5, pp. 571–583, Sep. 2009, doi: 10.1016/j.ipm.2009.05.001

[19] X. Zhang, H. Fuehres, and P. A. Gloor, “Predicting stock market indicators through Twitter ‘I hope it is not as bad as I fear’,” Procedia - Social and Behavioral Sciences, vol. 26, pp. 55–62, 2011, doi: 10.1016/j.sbspro.2011.10.562

[20] I. Budimir and I. Zjakić, “Two-dimensional logistic regression model of the white effect in printed media,” Expert Systems with Applications, vol. 293, no. Dec., pp. 1–11, Dec. 2025. doi:10.1016/j.eswa.2025.128784

[21] F. Rahma and S. Ulfah, “Emotion-Aware Detection of Cyberbullying in Visual Social Media Content,” *Journal of Digital Society*, vol. 1, no. 4, pp. 299–313, 2025, doi: 10.63913/jds.v1i4.44.

[22] R. A. M. Aljohani and A. A. Alnahdi, “Exploring Football Player Salary Prediction Using Random Forest: Leveraging Player Demographics and Team Associations,” *International Journal for Applied Information Management*, vol. 5, no. 4, pp. 203–213, 2025, doi: 10.47738/ijaim.v5i4.115.

[23] J. O. Guballo and J. A. C. Andes, “Network-Based Anomaly Detection in Blockchain Transactions Using Graph Neural Network (GNN) and DBSCAN,” *Journal of Current Research in Blockchain*, vol. 3, no. 1, pp. 15–27, 2026, doi: 10.47738/jcrb.v3i1.55.

[24] P. Vy, D. Lan, and D. Oanh, “The Effect of Globalization on Income Inequality in Developing Countries: A Bayesian Approach,” *Journal of Applied Data Sciences*, vol. 7, no. 1, pp. 1–15, 2025, doi: 10.47738/jads.v7i1.948.

[25] G. Toer and G. Kim, “An Empirical Study on the Impact of Feature Scaling and Encoding Strategies on Machine Learning Regression Pipelines,” *International Journal of Informatics and Information Systems*, vol. 9, no. 1, pp. 242–256, 2026, doi: 10.47738/ijiis.v9i1.293.

[26] I. M. M. El Emary, A. Brzozowska, Ł. Popławski, P. Dziekański, and J. Glova, “Anomaly Detection in Blockchain-Based Metaverse Transactions Using Hybrid Autoencoder and Isolation Forest Models for Risk Identification and Behavioral Pattern Analysis,” *International Journal Research on Metaverse*, vol. 3, no. 1, pp. 46–63, 2026, doi: 10.47738/ijrm.v3i1.45.

[27] D. Fortuna and C. Hutagalung, “Uncovering Lifestyle and Mental Well-being Predictors of Academic Performance Change in Online Learning: A Comparative Analysis of Interpretable Machine Learning Models,” *Artificial Intelligence in Learning*, vol. 1, no. 4, pp. 271–286, 2025, doi: 10.63913/ail.v1i4.41.