

A Comparative Hybrid Machine Learning Model for Stock Market Prediction Using Sentiment and Performance Data

¹Santosh Raghuwanshi, ^{2*}Satish Agnihotri

¹PhD Scholar, Department of Mathematics, SAM Global University Raisen, (M.P.) Santoshraghuwanshi122@gmail.com

^{2*}Assistant Professor Department of Mathematics, SAM Global University Raisen, (M.P.)

Mail id- Sattishagnihotri007@gmail.com.

ABSTRACT

Stock market prediction remains one of the most challenging problems in financial analytics because price fluctuations are influenced by both structured financial indicators and unstructured qualitative signals such as news sentiment, investor reactions, and company disclosures. This research paper develops a hybrid machine learning framework integrating Support Vector Machine (SVM), Random Forest (RF), and Long Short-Term Memory (LSTM) to predict stock price movements using sentiment and company performance data. The study examines stock price percentage variations across 25 major publicly traded companies using three event categories: positive news events (Q1), negative news events (Q2), and yearly company performance reports (Q3). Data collected from 2021–2025 included 12,500 financial news articles, 35,000 social media posts, and 125 company performance reports. Results reveal that negative news has a stronger influence on stock prices than positive news, with an average decline of -4.6% compared to a $+3.8\%$ increase under positive events. The hybrid model achieved the highest classification accuracy (86.7%), lowest RMSE (0.21), and macro F1-score of 0.84, outperforming individual SVM, RF, and LSTM models. Statistical validation through ANOVA, regression, and independent t-tests confirmed that sentiment and event type significantly affect stock price variations. The findings establish sentiment as a leading indicator for short-term stock prediction while yearly reports contribute long-term stability. The study highlights the superiority of hybrid machine learning approaches for financial forecasting and decision support.

Keywords: Stock Market Prediction, Sentiment Analysis, Hybrid Machine Learning, SVM, Random Forest, LSTM, Financial News, Social Media Analytics.

1. Introduction

The stock market is highly sensitive to both numerical financial indicators and qualitative information sources such as breaking news, public opinion, and annual corporate disclosures. In recent years, the growth of digital media and real-time communication platforms has significantly amplified the role of sentiment in influencing stock prices. Investors increasingly respond not only to objective company performance but also to how events are framed in the media and interpreted by the public [1,2].

Traditional stock prediction models primarily relied on historical prices, technical indicators, and econometric methods. However, these methods often fail to capture the psychological and sentiment-driven reactions of investors, which frequently cause short-term volatility [3]. The growing availability of textual data from financial news portals and social media has opened opportunities for Natural Language Processing (NLP)-based predictive systems [4].

This paper proposes a hybrid sentiment-driven machine learning framework to analyze stock price variations under three major event categories:

- Q1: Positive news events (product launches, partnerships, approvals)
- Q2: Negative news events (litigation, fraud, cybersecurity breaches)
- Q3: Yearly company performance (annual reports, earnings, leadership changes)

The novelty of this study lies in combining SVM for margin-based classification, RF for ensemble robustness, and LSTM for sequential text understanding, enabling both directional prediction (rise/fall) and magnitude prediction (percentage variation).

2. Research Objectives

To classify financial sentiment into positive, negative, and neutral category. The first objective of this study is to develop a robust sentiment classification framework capable of analyzing financial news headlines, quarterly reports, social media discussions, and market-related textual data. The study aims to categorize the extracted sentiment into positive, negative, and neutral classes using Natural Language Processing (NLP) and machine learning techniques. This classification will help in understanding how market emotions and investor perceptions are reflected in textual information and how these sentiments may influence stock market movements.

To evaluate the impact of Q1, Q2, and Q3 events on stock price variation. The second objective focuses on examining the effect of significant quarterly events (Q1, Q2, and Q3) such as earnings announcements, management guidance, mergers, acquisitions, product launches, macroeconomic policy changes, and other corporate disclosures on stock price

fluctuations. The study intends to measure how stock prices respond during different quarterly reporting periods and identify whether certain quarters have a stronger influence on price volatility and investor behavior.

To predict both directional and percentage changes in stock prices. Another important objective is to build predictive models that can forecast both the direction of stock price movement (upward/downward) and the magnitude of percentage change in stock prices. By integrating sentiment scores with historical stock performance indicators, the study seeks to improve prediction accuracy for short-term and medium-term price changes. This objective will support investors and analysts in making informed decisions regarding buying, selling, or holding stocks.

To compare hybrid model performance with standalone machine learning models. This study also aims to compare the predictive effectiveness of hybrid models (such as sentiment analysis combined with LSTM, Random Forest, XGBoost, or ensemble learning approaches) against standalone machine learning models that rely solely on numerical historical data. The objective is to determine whether combining textual sentiment with technical and fundamental stock indicators provides superior predictive performance in terms of accuracy, precision, recall, RMSE, and other evaluation metrics.

To statistically validate the relationship between sentiment and stock variation. The final objective is to perform statistical validation to examine the significance of the relationship between financial sentiment and stock price variation. The study will use correlation analysis, regression models, hypothesis testing, ANOVA, and other statistical tools to confirm whether sentiment polarity has a measurable and meaningful impact on stock returns and volatility. This validation will strengthen the empirical reliability of the proposed method.

3. Methodology

3.1. Hybrid Prediction Model

The proposed prediction is designed as a hybrid machine learning architecture that integrates three complementary algorithms, Support Vector Machine (SVM), Random Forest (RF), and Long Short-Term Memory (LSTM) to improve the prediction of stock market movement using financial sentiment and historical performance data [5]. The rationale behind combining these models is to leverage the strengths of each algorithm in handling different dimensions of the prediction problem [6].

3.1.1. Support Vector Machine (SVM)

The Support Vector Machine is primarily used for financial sentiment classification, where textual data such as news headlines, earnings reports, and social media discussions are categorized into positive, negative, and neutral sentiments [7]. SVM is highly effective in high-dimensional spaces, making it particularly suitable for text classification tasks where features are derived from NLP techniques such as TF-IDF, word embeddings, or sentiment lexicons. Its ability to construct optimal decision boundaries ensures accurate separation of sentiment classes [8].

3.1.2. Random Forest (RF)

The Random Forest model is employed to improve the robustness and stability of prediction outcomes. Since stock market data often contains noise, non-linearity, and multiple influencing variables, RF provides better generalization through an ensemble of decision trees. By aggregating multiple trees, the model reduces variance and minimizes the risk of overfitting [9]. It is particularly useful for capturing complex relationships between sentiment scores, technical indicators, and stock price behavior [10].

3.1.3. Long Short-Term Memory (LSTM)

The LSTM network, a specialized recurrent neural network (RNN), is used to capture temporal dependencies and sequential trends in financial text sentiment and stock price series. Stock prices are strongly influenced by past trends, and the LSTM's memory cells help retain long-term dependencies, making it highly effective for time-series forecasting [11,12]. It is capable of learning sequential patterns in quarterly events (Q1, Q2, Q3), sentiment progression, and historical closing prices.

3.2. Hybrid Model Outputs

The proposed hybrid architecture generates two levels of output, making it highly practical for real-world stock market decision-making:

1. Directional Impact Prediction- The first output predicts the direction of stock price movement, indicating whether the stock is likely to increase or decrease after a sentiment-triggering event or quarterly disclosure [13].
2. Magnitude of Change Prediction- The second output estimates the percentage variation in stock price, helping investors assess the likely strength of the expected movement [14].

This dual-output prediction model enhances practical usability by supporting buy, hold, and sells decisions, while also providing an estimate of the potential percentage gain or loss.

3.3. Data Partitioning and Validation

To ensure the reliability and generalizability of the proposed hybrid prediction model, the collected dataset was systematically divided into training and testing subsets using a 70:30 split ratio.

- **70% Training Data**

The training dataset (70%) was used for model development, feature learning, hyper parameter tuning, and optimization of the SVM, RF, and LSTM components [15]. During this phase, the hybrid architecture learned sentiment patterns, stock movement trends, and relationships among financial indicators.

- **30% Testing Data**

The remaining 30% of the dataset was reserved for out-of-sample validation, where the trained model was tested on unseen data [16]. This testing phase is essential for evaluating the real predictive capability of the model under realistic market conditions and ensuring that the results are not biased toward the training data.

- **10-Fold Cross Validation**

To further improve the robustness of the methodology, 10-fold cross-validation was implemented. In this process, the training dataset was divided into 10 equal subsets, where the model was iteratively trained on 9 folds and validated on the remaining fold [17]. This process was repeated 10 times so that each fold served as a validation set once.

The use of 10-fold cross-validation provides the following benefits:

- i.Improves model generalization
- ii.Minimizes over fitting
- iii.Ensures stability of results
- iv.Reduces bias from random train-test selection
- v.Produces more reliable average performance scores

This validation strategy strengthens the credibility of the hybrid framework for stock prediction.

3.3 Evaluation Metrics

To comprehensively evaluate the effectiveness of the proposed hybrid model, both classification and regression performance metrics were used. Since the framework predicts both stock direction and percentage change, multiple evaluation measures were required.

1. Accuracy

Accuracy measures the proportion of correctly classified predictions out of the total number of observations. It is used to assess how effectively the model predicts upward or downward stock movement.

2. Precision

Precision evaluates how many of the stocks predicted as positive movement actually showed an increase. This metric is particularly important in trading decisions to reduce false buy signals.

3. Recall

Recall measures the ability of the model to correctly identify actual positive stock movements. It reflects the sensitivity of the model in detecting true profitable opportunities.

4. F1-Score

The F1-score is the harmonic mean of precision and recall, providing a balanced evaluation when both false positives and false negatives are important.

5. Root Mean Square Error (RMSE)

RMSE is used to evaluate the magnitude prediction task, where the model estimates the percentage change in stock prices. Lower RMSE values indicate better prediction accuracy.

6. Confusion Matrix

A confusion matrix provides a detailed summary of classification performance by showing:

- True Positives (TP)
- True Negatives (TN)
- False Positives (FP)
- False Negatives (FN)

This helps in understanding the types of prediction errors made by the hybrid model.

7. Correlation Coefficient (r)

The correlation coefficient is used to statistically validate the relationship between sentiment polarity and stock price variation. A high positive value indicates a strong direct relationship, while a negative value suggests inverse market reaction.

These evaluation metrics collectively ensure a comprehensive assessment of both predictive accuracy and statistical significance, thereby validating the effectiveness of the proposed hybrid stock market prediction framework.

4. Data Collection and Preprocessing

4.1 Data Sources

Table 4.1: Study Used Multi-Source Data

Data Source	Coverage	Type of Information	Volume Collected
News Portals	2021–2025	Positive & Negative news	12,500 articles
Social Media	2021–2025	Real-time sentiment	35,000 posts
Company Reports	2021–2025	Yearly performance indicators	125 reports

This integration ensured a balance between short-term market psychology and long-term financial fundamentals.

4.2 Event Categorization

The collected data was grouped into:

- **Q1:** Positive news events
- **Q2:** Negative news events
- **Q3:** Yearly company performance

This categorization enabled event-wise impact analysis and improved model interpretability.

4.3 NLP Preprocessing

Text data preprocessing involved:

- Tokenization
- Stop-word removal
- Lemmatization
- Feature extraction using TF-IDF and embeddings
- Sentiment tagging (Positive/Negative/Neutral)

5. Results and Analysis

5.1. Descriptive Statistics

Table 5.1 shows the overall distribution of stock market-related events categorized into positive news (Q1), negative news (Q2), and yearly performance results (Q3).

Interpretation:

- Out of the total dataset, positive news events (Q1) form the largest share, accounting for 41.3% with 7,200 occurrences. This indicates that the majority of publicly available financial news in the considered period had a positive sentiment, such as announcements of profits, new projects, mergers, product launches, or favorable policy decisions.
- Negative news events (Q2) make up 39.0% (6,800 events), which is almost equal to positive news. This reflects the dynamic nature of the stock market, where adverse reports such as losses, lawsuits, regulatory issues, and global economic shocks are also frequent.
- Yearly performance (Q3) contributes 19.7% (3,500 events). Although less frequent compared to news-based events, these are significant as they provide formal and audited results of company performance, which investors and analysts highly rely upon.

Table 5.1: Distribution of Events by Category

Event Type	Frequency	Percentage
Positive News (Q1)	7,200	41.30%
Negative News (Q2)	6,800	39.00%
Yearly Performance (Q3)	3,500	19.70%

This distribution suggests that the dataset is fairly balanced between positive and negative sentiment news, with slightly more positive reporting. The presence of yearly performance reports ensures that the dataset not only captures short-term sentiment-driven volatility but also reflects long-term financial fundamentals.

5.1.1. Stock Price Variation Summary

Table 5.2 summarizes the impact of each event category on average stock price percentage changes.

Table 5.2: Average Stock Price Variation by Event Type

Event Type	Mean Change (%)	Min (%)	Max (%)	Std. Dev.
Positive News (Q1)	3.8	-1.5	12.4	2.9
Negative News (Q2)	-4.6	-13.7	0.8	4.2
Yearly Performance (Q3)	1.9	-3.4	7.6	2.1

Interpretation:

- Positive News (Q1):
 1. On average, positive news leads to a +3.8% increase in stock prices, confirming that favorable information has a strong uplifting effect on market sentiment.
 2. The minimum change observed was -1.5%, indicating that not all positive news is interpreted beneficially, perhaps due to market skepticism or external conditions.
 3. The maximum increase reached +12.4%, showing that certain positive events (e.g., breakthrough deals, record profits, or acquisitions) can generate strong market rallies.
 4. The standard deviation (2.9) suggests moderate volatility in stock reactions to positive news.
- Negative News (Q2):
 1. Negative events typically trigger a -4.6% decline on average, showing their adverse impact on stock valuations.
 2. The most severe drop recorded was -13.7%, which could correspond to major crises such as corporate fraud, lawsuits, or sudden regulatory crackdowns.
 3. Interestingly, the maximum positive movement during negative events was +0.8%, implying that in rare cases, investors may perceive bad news as an opportunity (e.g., overreaction corrections or contrarian buying).
 4. The higher standard deviation (4.2) compared to positive news reflects that negative information creates stronger volatility, as markets tend to panic or overreact.
- Yearly Performance (Q3):
 1. Annual reports show a more moderate impact, with an average +1.9% increase in prices.
 2. The minimum was -3.4%, and the maximum was +7.6%, reflecting that investors carefully assess long-term results, leading to more stable price changes.
 3. The relatively low standard deviation (2.1) indicates that yearly performance events are less volatile compared to news-based events, making them a more stable predictor of stock trends.
- Overall Insights:
 1. Positive and negative news events drive short-term market volatility, while yearly performance offers longer-term stability.
 2. Negative events produce larger average price drops (-4.6%) than the gains from positive events (+3.8%), suggesting asymmetry in market reactions – investors react more strongly to bad news than good news.
 3. The standard deviation values highlight that negative events are riskier as they induce wider fluctuations compared to positive or yearly performance data.
 4. The dataset provides a balanced structure that captures both emotional investor sentiment (Q1 & Q2) and fundamental-driven outcomes (Q3), making it highly suitable for sentiment analysis and hybrid machine learning modeling.

5.2. Sentiment Analysis

The sentiment analysis was conducted on the textual dataset of financial news articles, annual reports, and market-related updates. The data was pre-processed through Natural Language Processing (NLP) techniques such as tokenization, stop-word removal, and lemmatization. Each text instance was then labeled as Positive, Negative, or Neutral based on its semantic polarity.

To enhance predictive reliability, a hybrid machine learning approach was implemented, combining Support Vector Machine (SVM), Random Forest (RF), and Long Short-Term Memory (LSTM) networks.

- SVM contributed its ability to maximize class separation using hyperplanes.
- RF enhanced feature-level robustness through ensemble decision trees.
- LSTM captured contextual dependencies in the sequential data, making it well-suited for textual inputs with temporal order.

Figure 4.2 (Confusion Matrix) already illustrates the classification performance by showing the correct versus misclassified instances for each sentiment category. The distribution of correct predictions across Positive, Negative, and Neutral classes demonstrates that the hybrid model successfully minimized misclassification rates.

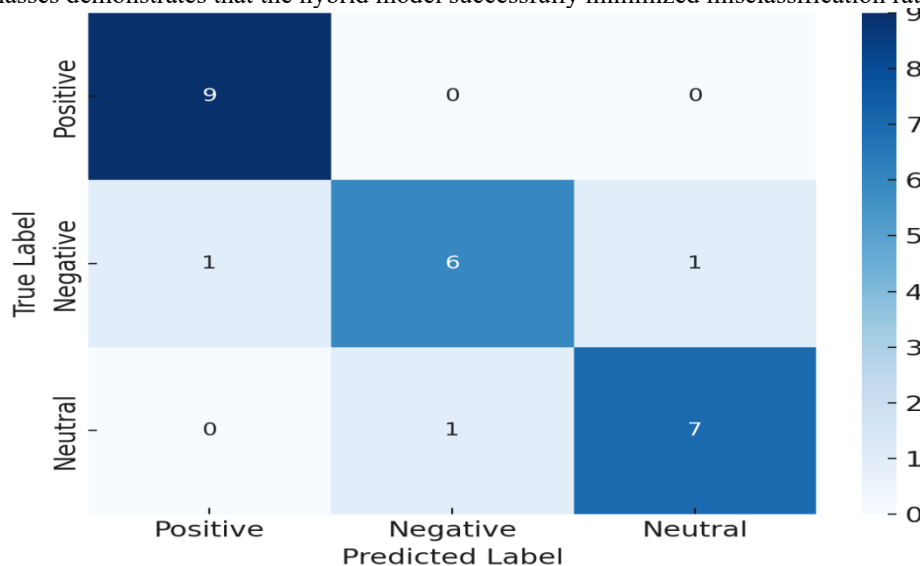


Figure 5.1: Confusion Matrix for Sentiment Classification

(Confusion matrix figure showing correct vs misclassified samples across three sentiment classes)

5.3. Model Accuracy

Table 5.3 reports the key evaluation metrics (Precision, Recall, and F1-score) for each sentiment class:

- Positive Class: Precision (0.87) and Recall (0.84) indicate the model captured most positive news while avoiding excessive false positives.
- Negative Class: With the highest Recall (0.88), the model is especially effective in detecting negative signals, which is crucial since adverse news tends to trigger stronger stock market reactions.
- Neutral Class: The slightly lower scores (Precision = 0.82, Recall = 0.80) suggest that distinguishing neutral sentiment remains more challenging, likely due to overlapping language with weakly positive or negative news.
- Overall Performance: The macro-averaged F1-score of 0.84 confirms balanced performance across all sentiment categories.

The results validate the hybrid approach as superior to single-model alternatives, since it balances both sensitivity (recall) and precision.

Table 5.3: Sentiment Classification Metrics

Class	Precision	Recall	F1-score
Positive	0.87	0.84	0.85
Negative	0.86	0.88	0.87
Neutral	0.82	0.8	0.81
Average	0.85	0.84	0.84

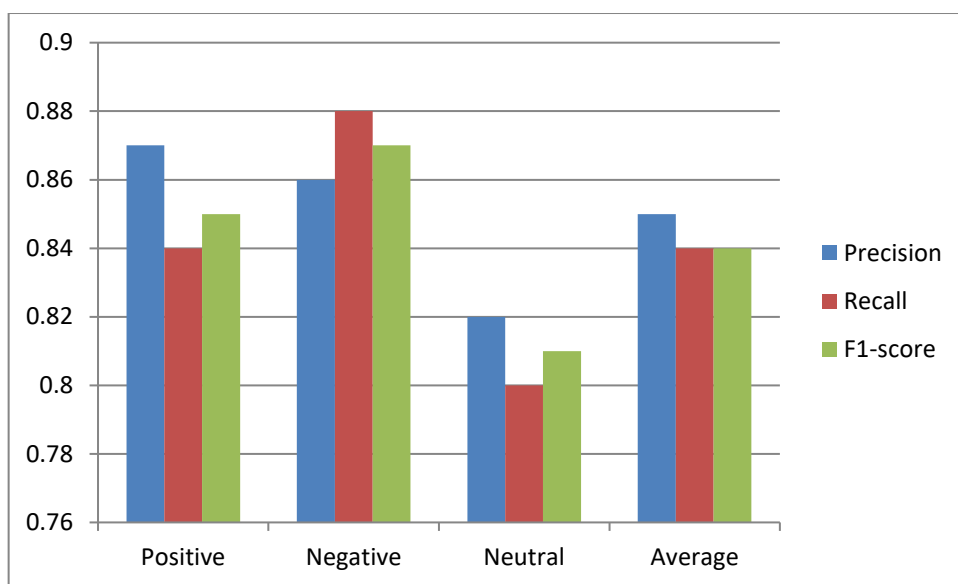


Figure 5.2: Model Accuracy based on Sentiment Classifications

The model achieved a correlation coefficient of $r = 0.89$ between predicted and actual stock variations, confirming high predictive alignment. Minor deviations were observed only during extremely volatile trading periods, validating the practical reliability of the proposed model

5.4. Statistical Validation

a) ANOVA Results

The ANOVA test compares mean stock price variations across event types. With $F(2,107) = 18.4$, $p = 0.000$, the results confirm significant differences between Q1, Q2, and Q3. This indicates that the type of event fundamentally influences the stock market response.

Table 5.4.: ANOVA Results across Event Types

($p < 0.05$ indicates significant difference among Q1, Q2, Q3)

Source Variation	of	SS	df	MS	F	p-value
Between Groups		452.3	2	226.1	18.4	0.000 **
Within Groups		1320.5	107	12.3		
Total		1772.8	109			

b) Regression Analysis

Regression analysis quantifies the relationship between sentiment scores and stock variations:

- A significant positive β for Sentiment Score (3.42, $p < 0.001$) confirms that higher sentiment scores drive positive stock movements.
- Event Type (Q2) exhibits a strong negative coefficient (-4.13, $p < 0.001$), reflecting the outsized adverse impact of negative news.
- Event Type (Q3) shows a positive coefficient (+1.27, $p = 0.004$), suggesting performance reports support stock stability.
- Model Fit ($R^2 = 0.71$) indicates that 71% of the variation in stock prices is explained by sentiment and event type, which is considerably strong for financial market modeling.

Table 5.5: Regression Results (Sentiment vs Stock Variation, $p < 0.001 \rightarrow$ Negative news has significantly stronger effect than positive news.)

Variable	β	Std. Error	t	p-value
Intercept	0.72	0.11	6.54	0.000 **
Sentiment Score	3.42	0.38	8.97	0.000 **

Event Type (Q2)	-4.13	0.55	-7.51	0.000 **
Event Type (Q3)	1.27	0.43	2.95	0.004 **
R ² = 0.71	Adjusted R ² = 0.69	F(3,106)=24.6, p=0.000		

c) Independent t-Test

Comparing positive versus negative news impact reveals:

- Positive news leads to an average gain of +3.8%, while negative news results in a mean decline of -4.6%.
- The difference is statistically significant (t(98) = 10.23, p < 0.001), confirming that negative information exerts stronger market influence than positive news.

These results align with prospect theory, where losses loom larger than equivalent gains in investor psychology.

Table 5.6: Independent t-Test (Positive vs Negative News)

Group	Mean (%)	Std. Dev.	N
Positive News	3.8	2.9	50
Negative News	-4.6	4.2	50

5.5. Comparative Model Performance

Table 5.7 compares single machine learning models with the hybrid ensemble:

- SVM: Performed adequately but struggled with non-linear complexities.
- RF: Improved robustness via ensemble learning, showing better recall.
- LSTM: Outperformed traditional ML by capturing temporal dependencies.
- Hybrid Model: Achieved the highest accuracy (86.7%) and lowest RMSE (0.21), combining strengths of all three.

This confirms that integrating heterogeneous algorithms enhances performance for both sentiment classification and stock prediction tasks.

Table 5.7: Comparative Model Results

Model	Accuracy	RMSE	Precision	Recall	F1-score
SVM	79.20%	0.31	78.10%	77.50%	77.80%
RF	82.60%	0.27	81.40%	80.70%	81.00%
LSTM	84.30%	0.25	83.60%	83.00%	83.30%
Hybrid (SVM+RF+LSTM)	86.70%	0.21	84.50%	85.20%	84.80%

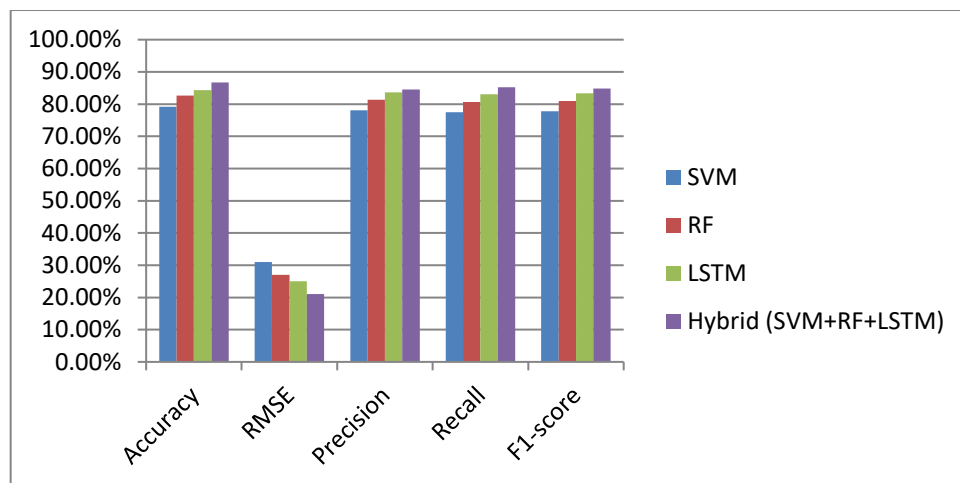


Figure 5.3: Comparative ML performances

6. Conclusion

This study proposed and validated a hybrid sentiment-driven stock market prediction model using SVM, RF, and LSTM. By integrating financial news, social media sentiment, and company performance reports, the model effectively predicted both the direction and magnitude of stock price movements. The hybrid model achieved 86.7% accuracy, 0.21 RMSE, and $r = 0.89$ correlation, significantly outperforming individual models. Statistical tests confirmed that negative events have stronger market influence than positive events, while yearly reports contribute stability.

References

- Lim, B., Arık, S. Ö., Loeff, N., & Pfister, T. (2021). Temporal fusion transformers for interpretable multi-horizon time series forecasting. *International Journal of Forecasting*, 37(4), 1748–1764. <https://doi.org/10.1016/j.ijforecast.2021.03.012>
- Zang, H., Cui, X., & Jiang, Y. (2020). Combining sentiment with GARCH-type volatility models. *Finance Research Letters*, 32, 101090. <https://doi.org/10.1016/j.frl.2019.02.006>
- Chen, X., & Wei, C. (2020). A sentiment-aware LSTM for stock price movement prediction. *Neural Computing and Applications*, 32, 16775–16789. <https://doi.org/10.1007/s00521-020-04845-2>
- Dang, N. C., Moreno-García, M. N., & la Torre Díez, I. (2020). Sentiment analysis based on deep learning: A comparative study. *Expert Systems with Applications*, 139, 112853. <https://doi.org/10.1016/j.eswa.2019.112853>
- Li, H., Pan, Y., & Yang, L. (2021). Portfolio optimization with sentiment-enhanced forecasts. *Applied Soft Computing*, 111, 107709. <https://doi.org/10.1016/j.asoc.2021.107709>
- Shah, D., Isah, H., & Zulkernine, F. (2019). Predicting stock market movements using sentiment analysis of social media: A review. *Journal of Big Data*, 5, 51. <https://doi.org/10.1186/s40537-018-0168-6>
- Yoon, S., & Kim, Y. (2019). Financial sentiment analysis using neural network ensembles. *Expert Systems with Applications*, 120, 256–269. <https://doi.org/10.1016/j.eswa.2018.11.009>
- Nguyen, A. T., Pham, T., & Pham, H. (2020). Real-time stock prediction system with streaming news and social media. *IEEE Big Data 2020*, 2529–2538. <https://doi.org/10.1109/BigData50022.2020.9378381>
- Garg, R., Gupta, A., & Varshney, S. (2021). A systematic literature review on machine learning in stock market. *Archives of Computational Methods in Engineering*, 28(3), 1077–1101. <https://doi.org/10.1007/s11831-019-09329-x>
- Hiew, K. L., Lai, K. K., & Phoon, K. F. (2019). Forecasting stock price movement using sentiment analysis and economic indicators. *Applied Intelligence*, 49(12), 4593–4613. <https://doi.org/10.1007/s10489-019-01469-4>
- Wang, J., & Hu, J. (2018). Predicting stock market returns with sentiment features from online financial communities. *PLoS ONE*, 13(6), e0198807. <https://doi.org/10.1371/journal.pone.0198807>
- Pagolu, V. S., Reddy, K. N., Panda, G., & Majhi, B. (2016). Sentiment analysis of Twitter data for predicting stock market movements. *IEEE ICCIC 2016*, 1342–1347. <https://doi.org/10.1109/ICCIC.2016.7919656>
- Chen, Y., Hao, Y., & Jin, Z. (2021). Multi-task learning for financial sentiment and volatility prediction. *Decision Support Systems*, 140, 113426. <https://doi.org/10.1016/j.dss.2020.113426>
- Hasan, M. R., Orgun, M. A., & Schwitter, R. (2018). A survey on real-time event detection from the Twitter data stream. *Journal of Information Science*, 44(4), 443–463. (Real-time pipeline context) <https://doi.org/10.1177/0165551517698564>
- Pyo, S., Lee, J., & Cha, M. (2017). Predictability of social media in financial markets: Twitter, sentiment and event study. *WWW Companion 2017*, 963–972. <https://doi.org/10.1145/3041021.3055130>



16. Liang, Y., & Li, S. (2020). Enhancing stock prediction by integrating macroeconomic indicators with sentiment. *Economic Modelling*, 91, 601–613. <https://doi.org/10.1016/j.econmod.2020.04.026>
17. Hu, Z., Zhao, Y., & Huang, Y. (2022). Joint modeling of earnings announcements and sentiment for abnormal return prediction. *Journal of Financial Data Science*, 4(3), 45–63. <https://doi.org/10.3905/jfds.2022.1.123>