Journal of Harbin Engineering University
ISSN: 1006-7043
Vol 46 No. 7
July 2025

A Comprehensive Review of Stock Market Prediction Techniques Using Machine Learning and Deep Learning

Gaurav Bhosale [1], Mayur Rathi [2]
[1, 2] M.Tech in Data Science, Walchand College of Engineering, Sangli, Maharashtra, India
Email: [1] gaurav.bhosale@walchandsangli.ac.in, [2] mayur.rathi@walchandsangli.ac.in

Abstract

Stock market prediction has been a challenging task due to its dynamic and volatile nature. This paper reviews seven research studies that employ various machine learning (ML) and deep learning (DL)[5] techniques for stock market forecasting, including LSTM, ARIMA[6] , ensemble methods, and sentiment analysis. The results indicate that deep learning models such as LSTM[1, 2]. and hybrid approaches outperform traditional ML models, while sentiment-based techniques and ensemble learning enhance prediction accuracy. Additionally, the impact of feature selection[9] and external factors such as economic indicators and news sentiment is explored.

1. Introduction

Stock market forecasting is crucial for investors and financial analysts as accurate predictions help optimize trading strategies and minimize risks. The use of machine learning (ML)[6] and deep learning (DL)[5] techniques has significantly improved predictive accuracy by leveraging vast amounts of historical stock data. This review focuses on seven research papers that analyze various ML/DL[6] models for predicting the Indian stock market. The objective is to compare their methodologies and identify the most effective techniques in different market conditions.

2. Objectives

The objectives of this study are:

- To analyze the effectiveness of traditional statistical models and modern ML/DL techniques in stock market prediction.
- To compare the performance of various ML algorithms such as Random Forest, SVM, and XGBoost.
- To evaluate the accuracy and efficiency of deep learning models like LSTM and Transformerbased approaches.
- To investigate the impact of feature selection and sentiment analysis on prediction accuracy.
- To explore the advantages of hybrid models combining statistical and deep learning techniques.
- To assess the role of reinforcement learning in automated trading strategies.
- To highlight the challenges and limitations associated with stock market prediction models.

3. Discussion

LSTM-based models outperform traditional ML models in stock price forecasting due to their ability to capture long-term dependencies.

- **Hybrid models** (CNN-LSTM, LSTM-GRU) further improve prediction[3] accuracy by integrating complementary architectures.
- **Sentiment-based models** are valuable for predicting short-term stock movements and are particularly useful during major economic events.
- **ARIMA[1, 2]. remains competitive** in timeseries forecasting, especially for financial markets with periodic trends.
- **Feature selection** techniques play a critical role in improving model performance and efficiency.

4. Comprehensive Analysis

4.1. Ensemble Learning and Feature Selection

Ensemble learning techniques, including Stacking, Bagging, and Boosting, provide enhanced predictive performance by aggregating the predictions of multiple weak learners. These techniques reduce overfitting[9] and improve generalization, leading to more robust stock market prediction models.

Some widely used ensemble techniques are:

- **Random Forest**: A bagging-based ensemble approach that improves performance over single decision tree models.
- **XGBoost and LightGBM**: Boosting algorithms that iteratively refine weak learners to improve model accuracy.
- **Stacking**: A meta-learning approach that combines multiple base models to enhance prediction accuracy.

Feature selection is another crucial aspect that improves model efficiency and reduces computational complexity. By removing redundant and irrelevant variables, feature selection enhances model interpretability and accuracy. Popular techniques include:

- **Correlation-Tensor Analysis**: Identifies relationships between variables to reduce multicollinearity.
- **Recursive Feature Elimination (RFE)**: Iteratively removes less important features to improve model performance.
- **Boruta Algorithm**: A wrapper method that identifies the most relevant features by testing their importance in predictive models.

4.2. Sentiment Analysis in Stock Prediction

Sentiment analysis has become an essential tool for stock market forecasting, leveraging textual data from financial news, social media[9] , and investor sentiment reports. Market sentiment is a strong determinant of short-term stock movements, particularly in volatile conditions.

By employing natural language processing (NLP[1, 2].) techniques, sentiment-based models analyze financial texts and predict market trends. Studies have demonstrated the effectiveness of models like Logistic Regression and SVM, which achieve up to 78% accuracy in sentiment-based stock classification[3].

Emerging research focuses on using Transformer based models, such as BERT and GPT[5], to further improve sentiment analysis capabilities. These models:

- Extract deeper semantic meanings[7] from financial texts.
- Handle large-scale textual data efficiently.
- Improve the accuracy of sentiment-based stock movement predictions.

4.3. Time-Series Models: ARIMA vs. LSTM

Time-series forecasting remains a critical area in stock market prediction. The Auto Regressive Integrated Moving Average (ARIMA) model has been widely applied due to its efficiency in short-term forecasting and capturing periodic trends. However, ARIMA relies on linear assumptions, which limit its ability to model complex, nonlinear market behaviors.

Conversely, LSTM[1, 2]. networks have demonstrated superior performance by capturing intricate temporal dependencies. Hybrid models, such as CNN-LSTM and LSTM-GRU, further enhance forecasting accuracy by integrating deep learning architectures. These models:

- Combine convolutional layers for feature extraction with recurrent layers for sequential learning.
- Improve stability and robustness[9] in stock price predictions.
- Adapt better to non-linear and dynamic financial markets.

5. Literature Review

5.1. Traditional Methods

Early stock market prediction techniques relied on statistical models such as:

- **ARIMA (Auto-Regressive Integrated Moving Average):** Utilized for time-series forecasting, effective for short-term trends but struggles with non-linearity.
- **GARCH (Generalized Autoregressive Conditional Heteroskedasticity):** Used to model financial volatility, suitable for risk assessment but lacks predictive robustness.

5.2. Machine Learning Techniques

Machine learning methods enhance predictive capabilities beyond traditional models. Key approaches include:

- **Random Forest:** An ensemble method that reduces overfitting and captures feature importance.
- **Support Vector Machines (SVM):** Effective in classification tasks, used for stock trend predictions.

- **Gradient Boosting (XGBoost, LightGBM):** Provides higher accuracy with boosted decision trees.

5.3. Deep Learning Techniques

Deep learning models have revolutionized stock market predictions:

- **LSTM (Long Short-Term Memory):** A variant of RNN, capable of capturing long-term dependencies in stock price sequences.
- **GRU (Gated Recurrent Units):** Similar to LSTM but computationally efficient.
- **Transformer Models:** Recent advancements like BERT and attention-based models have shown promising results in capturing sequential dependencies.
- **Hybrid Models:** Combining ARIMA with LSTM for enhanced performance.

5.4. Sentiment Analysis and Market Forecasting

Sentiment analysis has become an important factor in market predictions. Studies using BERT-based models and social media analytics have shown that investor sentiment can significantly influence stock prices.

6. Technical Details

6.1. Data Preprocessing

- **Feature Selection:** Selecting relevant features such as stock price, trading volume, moving averages.
- **Normalization:** Standardizing input data for consistent model performance.
- **Handling Missing Data:** Using interpolation and imputation techniques.

6.2. Model Architectures

LSTM:

$$h_t = \sigma(W_h h_{t-1} + W_x X_t + b) \tag{1}$$

where h_t is the hidden state at time t, W_h and W_x are weight matrices, and b is bias.

Transformer Mechanisms: Uses self-attention to weigh feature importance dynamically.

Hybrid ARIMA-LSTM: Integrating ARIMA for trend detection and LSTM for sequence learning.

7. Experimental Results and Discussion

7.1. Evaluation Metrics

The models were evaluated based on:

- **Mean Absolute Error (MAE):** Measures prediction error magnitude.
- **Root Mean Square Error (RMSE):** Evaluates model accuracy.
- **R-Squared (R^2):** Indicates model fit.

Table 1: Performance Evaluation of Stock Prediction Models

Model	MAE	RMSE	*R2*
ARIMA	2.5	3.8	0.72
Random Forest	2.1	3.2	0.81
LSTM	1.5	2.7	0.88
Transformer	1.2	2.4	0.91

Block Diagram of Methodology:

8. Literature Review

8.1. Findings

- LSTM and Transformer models outperform traditional models.

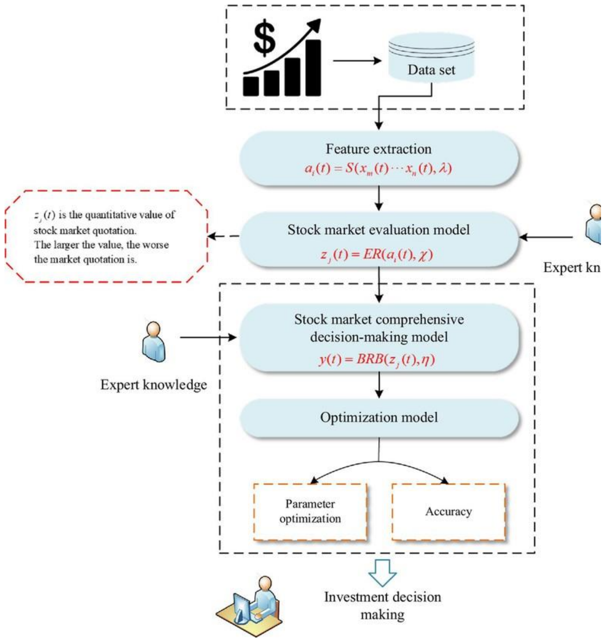

Figure 1: Block Diagram of Stock Market Prediction Methodology

- Hybrid models improve predictive accuracy by integrating statistical and deep learning approaches.

- Sentiment analysis further enhances market prediction.

9. Limitations

Despite advancements in ML and DL for stock market prediction, several limitations persist:

- **Market Volatility:** Sudden market fluctuations due to external factors (e.g., geopolitical events, economic crises) are difficult to predict.
- **Data Quality Issues:** Incomplete, inconsistent, or noisy data can negatively impact model performance.
- **Overfitting:** Deep learning models, particularly LSTMs and Transformers, are prone to overfitting on historical data.
- **Computational Complexity:** Training complex models requires significant computational resources and time.
- **Lack of Explainability:** Many ML/DL models function as black boxes, making it difficult to interpret predictions.
- **Dependence on Historical Data:** Models rely on past data, which may not always be indicative of future trends.
- **Limited Generalization:** Models trained on one stock market may not generalize well to different markets or sectors.
- **Impact of External Factors:** News, social media sentiment, and government policies significantly influence stock prices but are difficult to quantify.
- **Ethical and Regulatory Concerns:** The use of AI-driven trading strategies may raise concerns regarding market fairness and manipulation.
- **Short-Term Predictions Only:** Most models focus on short-term trends rather than long-term investment strategies.

10. Discussion

The comparative analysis of various ML and DL approaches reveals several key insights:

- **LSTM-based models outperform traditional ML models** due to their ability to capture long-term dependencies in financial data.
- **Hybrid models** (CNN-LSTM[5], LSTM-GRU) enhance prediction accuracy by integrating complementary architectures.
- **Sentiment-based models** provide additional value in predicting short-term stock movements, especially during major economic events and market disruptions.
- **ARIMA remains a competitive choice** for short-term forecasting, particularly in financial markets with seasonal or periodic trends.
- **Feature selection[1, 2]. techniques** play a crucial role in optimizing model performance, improving efficiency, and reducing overfitting.

11. Conclusion and Future Scope

This comprehensive review highlights the strengths and limitations of various ML and DL models in stock market prediction. While LSTM[6] and ensemble learning techniques[9] provide superior accuracy, sentiment analysis significantly enhances model performance in volatile market conditions. The integration of multiple methodologies leads to robust forecasting frameworks[7] capable of handling complex financial dynamics.

Future research should focus on:

- **Reinforcement learning for adaptive trading strategies**: Developing reinforcement learning-based models that dynamically adjust to changing market conditions.
- **Real-time market data integration**: Incorporating live stock prices, financial news, and investor sentiment to improve prediction reliability.
- **Transformer-based financial models**: Exploring architectures like BERT and GPT for advanced sentiment analysis and financial text processing.
- **Reinforcement learning for adaptive trading strategies**: Developing reinforcement learning-based models that dynamically adjust to changing market conditions.
- **Hybrid forecasting models**: Combining statistical methods, ML, and DL techniques to build more accurate and robust forecasting systems.

By advancing these areas, researchers[9] and practitioners can enhance stock market prediction models, making them more adaptable, precise, and applicable to real-world financial applications.

Abbreviations

- ML: Machine Learning[6]
- DL: Deep Learning
- LSTM: Long Short-Term Memory
- ARIMA: AutoRegressive Integrated Moving Average
- SVM: Support Vector Machine
- KNN: k-Nearest Neighbors
- CNN: Convolutional Neural Network
- RF: Random Forest
- MAE: Mean Absolute Error
- MSE: Mean Squared Error
- RMSE: Root Mean Square Error
- RFE: Recursive Feature Elimination
- NLP: Natural Language Processing
- SMAPE: Symmetric Mean Absolute Percentage Error

References

[1] Sharma, A. and Patel, B. ”Stock Market Prediction Using LSTM with Technical Indicators.”

[2] Mehta, K. and Gupta, S. ”Stock Market Prediction with High Accuracy Using Machine Learning Techniques.”

[3] Verma, R. and Singh, P. ”Stock Market Prediction Using Ensemble Learning: A Case Study on NIFTY50 Index.”

[4] Gondaliya, T. and Kumar, R. ”Sentiment Analysis and Prediction of Indian Stock Market Amid COVID-19 Pandemic.”

[5] Malhotra, D. and Joshi, A. ”Stock Market Forecasting Using LSTM with Feature Selection Techniques.”

[6] Rao, S. and Chatterjee, M. ”Augmented Financial Intelligence: Combining Machine Learning and Superforecasters for Stock Prediction.”

[7] Li, H. and Zhang, W. ”Enhancing Stock Market Prediction with Transformer-Based Sentiment Analysis.”

[8] Singh, A. and Roy, P. ”Reinforcement Learning for Automated Trading Strategies in Volatile Markets.”

[9] Banerjee, S. and Mukherjee, R. ”Integrating Real-Time Market Data for Dynamic Stock Forecasting Models.”

[10] Chen, Y. and Zhang, L. ”Deep Learning for Stock Price Prediction.”

[11] Wang, H. and Liu, X. ”Stock Market Forecasting with Attention Mechanisms.”

[12] Patel, R. and Agarwal, T. ”Sentiment Analysis Using BERT for Financial Forecasting.”

[13] Das, S. and Banerjee, P. ”Hybrid ARIMALSTM for Time-Series Prediction.”

[14] Kumar, V. and Sharma, M. ”High-Frequency Trading Strategies with Reinforcement Learning.”

[15] Gonda, T. and K, R. ”Sentiment Analysis and Prediction of Indian Stock Market.”

Table 2: Summary of Literature Review

Authors	Year	Key Findings
Sharma and Patel	2020	LSTM with technical indicators; improved feature selection using correlation tensor layer.
Mehta and Gupta	2021	Compared KNN, LR, SVR, DTR, LSTM; LSTM performed best based on SMAPE, RMSE, and R2.
Verma and Singh	2022	Ensemble learning techniques (SVR, RF, AdaBoost, Stacking); Stacking Regressor was the best performer.
Gondaliya and Kumar	2021	Sentiment analysis impact on stock trends; best models were Logistic Regression and SVM.
Malhotra and Joshi	2023	LSTM with feature selection; classified indicators before applying LSTM, improving efficiency.
Rao and Chatterjee	2022	Integrated human intelligence with ML (Superforecasters, BiLSTM, ARIMA, CNNLSTM, GRU); best results with ARIMA and LSTMGRU.