JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

AI IN FINANCE: MODELING STOCK PRICE MOVEMENTS USING MACHINE LEARNING AND SENTIMENT ANALYSIS

[1]Pradeep Mokashi, [2]Janvi Patel

[1]Independent Researcher, Palghar, India, [2]Student, GSFC University, Gujrat, India

Abstract: Artificial Intelligence (AI) is not just enhancing financial decisions. It's transforming the DNA of the stock market itself. This paper explores how AI is reshaping investor behavior, influencing decision-making timing, and transforming broker operations, while offering comparative results of three distinct AI models trained on real-world stock data. By quantifying prediction accuracy and evaluating AI's cognitive edge over traditional systems, we present a compelling case for a tech-driven financial ecosystem.

Keywords **- Artificial Intelligence, Nifty50, Stock Market, Sentiment Analysis, Machine Learning, Predictive Modeling, Financial Prediction, Financial Automation, Investor Behavior, Hybrid Model, LSTM, Random Forest, Linear Regression.**

1. INTRODUCTION

Markets have always been this strange blend part numbers, part nerves. One day, the charts look like a calm sea, the next, like a storm with no warning. For years, we trusted instincts, charts, and the so-called "gut feel." But today, there's a new player at the table- one that doesn't get tired, doesn't get emotional, and doesn't blink. A quiet observer that doesn't just see the data feels the pulse beneath it. It decodes the chaos, listens between the lines, and spots what we often miss: patterns buried in noise, probabilities inside uncertainty. While humans analyze news and sentiment, AI models zoom out processing decades of market history and reacting in milliseconds.

This study is a deep dive into that world. It explores how AI doesn't just keep up with market trends but can often predict them and whether the perfect formula might not be pure automation or human logic alone… but a powerful partnership between both.

2. RELATED WORK

We're not the first to wonder what happens when machines meet money.

Across research papers, industry reports, and academic journals, one trend is loud and clear: AI is redefining how we think about markets. Publications like Forbes and ScienceDirect highlight how models like LSTM and ensemble learning don't just crunch numbers they learn from them, evolving with each tick of the market.

Some research points out that AI has even started replacing traditional technical analysis. A study by SCRIP hints that charts and patterns are now being replaced by real-time, sentiment-aware models that adjust forecasts based on how people feel not just what they buy.

But here's the catch: not everything is perfect.

Many papers agree that while AI's accuracy is impressive, it often comes with a black box problem. It works, but no one's really sure why it works. And when it comes to real-time trading, that missing piece of explainability becomes a risk.

That's why this study takes a different route. By comparing three different AI models across performance, clarity, and timing we aim to understand not just how AI can predict markets… but whether it can truly collaborate with human intuition to make smarter, sharper, and safer decisions.

3. METHODOLOGY

3.1. Data collection

3.1.1 Data: Closing prices of the Nifty 50 index from Yahoo Finance (2018-2025)

3.1.2 Sentiment Sources:

- Google News: Top headlines of Nifty 50 using Google News API.
- Reddit: Recent posts from r/IndianStockMarketsubreddit using Reddit API.
- Twitter/X: Tweets have Nifty 50 keywords in recent time using Twitter API

3.2. Sentiment Analysis

Used two sentiment scoring models:

- VADER: Sentiment score ranging from -1 to 1.
- BERT: Transformer-based classifier for positive or negative sentiment.

Sentiment scores are averaged and merged with stock data from all mentioned sources.

3.3. Feature Engineering

- SMA (Simple Moving Average) of 5 and 20 days.
- Daily Return in %.
- Combined Score (BERT + VADER).
- RSI (Relative Strength Index).

All features were normalized using MinMaxScaler.

3.4. Models

3.4.1 Random Forest:

- Input: flattened features.
- Parameters: max depth 12, 300 estimators, minimum sample split 5.

3.4.2 LSTM:

- Input: 10-day window of all features.
- Architecture: 50 units of LSTM layer, Dropout, Dense Layer, adam optimizer, and MSE for loss.

3.4.3 Hybrid LSTM + Sentiment:

- Combined the two models:
  - Price branch: LSTM layer with 32 units.
  - Sentiment branch: LSTM layer with 16 units.
- Concatenation:
  - 48-dimensional combined feature vector.
  - Passed through a dense layer.

3.5. Evaluation Metrics

- Root Mean Squared Error (RMSE).
- Mean Absolute Error (MAE).
- Score (Coefficient of determination).

4. EXPERIMENTS

To examine the effectiveness of several models for predicting Nifty50 closing prices, we performed a series of experiments comparing three approaches: Random Forest, LSTM (Long Short-Term Memory), and a Hybrid LSTM (Long Short-Term Memory) with sentiment.

4.1. Experimental Setup

- Platform: Google Colab.
- Hardware: NVIDIA Tesla T4 GPU.
- Train-Test Split: 80-20.
- Input Window size: 10 days.
- MinMaxScaler for normalization.

4.2. Hyperparameters

4.2.1 Random Forest

- n_estimators=300.
- max_depth=12, min_samples_split=5.

4.2.2 LSTM (Long Short Term Memory)

- LSTM Layer(50 units), Dropout(0.2), Dense(1).
- Optimizer: Adam, Loss: MSE(Mean squared error), Epochs: 10.

4.2.3 Hybrid LSTM

- LSTM (32) for price data, LSTM(16) for sentiment.
- Concatenated - Dense (32) - Dense(1), activation (relu).
- Optimizer: Adam, Loss: MSE(Mean squared error), Epochs: 10.

4.3. Sentiment Collection

- Reddit: Recent 30 posts from subreddit “IndianStockMarket”.
- Google News: Top Headlines of Nifty 50 for the whole Day.
- Twitter/X: Nifty 50 related 50 tweets.

5. RESULTS AND ANALYSIS

5.1. RMSE and MAE of Models

The performance of each model was evaluated using two key metrics: Root Mean Square Error (RMSE) and Mean Absolute Error (MAE). The table (5.1) summarizes the results on the test set:

Table 5.1: RMSE and MAE of Models

Models	RMSE	MAE
Random Forest	0.201598	0.1896
LSTM	0.039008	0.0352
Hybrid LSTM + Sentiment	0.072926	0.0661

Table 5.1 displayed,

Random Forest: This model demonstrated the highest RMSE (0.201598) and MAE (0.1896), indicating limited ability to capture the temporal dependencies and nonlinear patterns present in financial time-series data.

LSTM: The standalone LSTM model outperformed the other models, achieving the lowest RMSE (0.0390) and MAE (0.0285). Its strength lies in its capacity to learn long-term dependencies and recognize sequential trends effectively.

Hybrid LSTM + Sentiment: Although the Hybrid Model did not outperform the standalone LSTM in terms of RMSE (0.072926) and MAE (0.0661), it still delivered competitive performance. The inclusion of sentiment data provided an additional layer of context, allowing the model to factor in real-time market mood, which can be valuable in dynamic market conditions.

The predicted next day's closing price for Nifty 50 using the Hybrid Model was 23126.75.

5.2 Forecast Results

Next 5 Days Predicted Close Prices (Hybrid Model):

- Day 1: 23031.95
- Day 2: 23074.80
- Day 3: 23161.71
- Day 4: 23245.11
- Day 5: 23326.77

5.3 Results

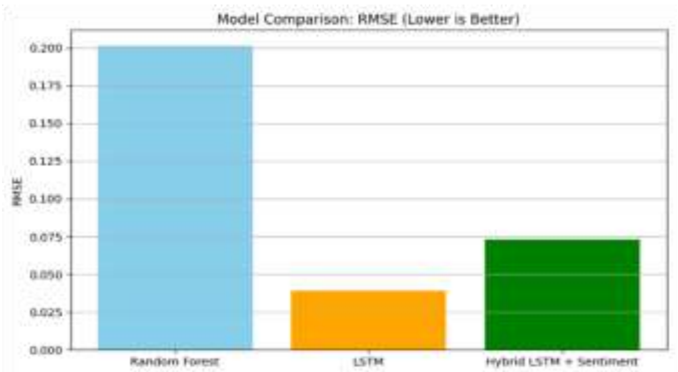

Fig 1. Comparing random forest, LSTM, and hybrid LSTM + sentiment for RMSE.

RMSE is Root Mean Squared Error, which is used to evaluate the accuracy of the model. The lower the RMSE score, the better it means it has more accuracy.

Fig 2. Actual Value vs Predicted Value of Random Forest.

Figure 2 shows that random forest values underfit actual values because of the limitations of temporal dependencies in Stock price movements.

Fig 3. Actual Value vs Predicted Value of LSTM.

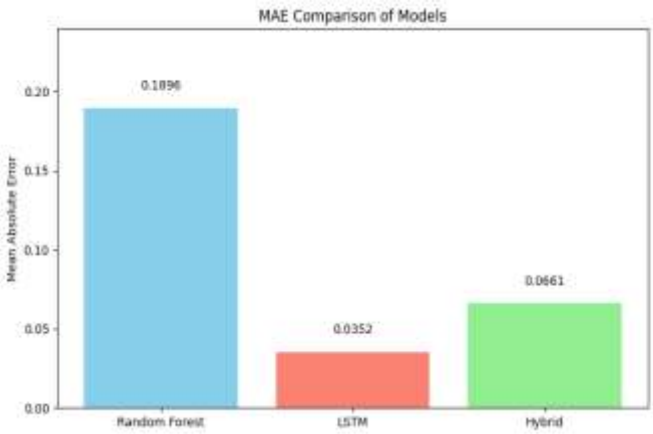

Figure 3 shows that the LSTM model learns temporal dependencies in the data it is more suitable for time-series forecasting, just like stock market prediction. Although there are some minor deviations in the period of high volatility. Overall, LSTM demonstrates better prediction than random forest.

Fig 4. Actual Value vs Predicted Value of hybrid LSTM + sentiment.

Figure 4 shows that the hybrid LSTM with sentiment model combines the LSTM with sentiment analysis features. The model contains historical stock prices data and sentiments from Reddit, Twitter/X and Google News. It captures wider trends and market dips & rises. Although predictions are not entirely perfect.

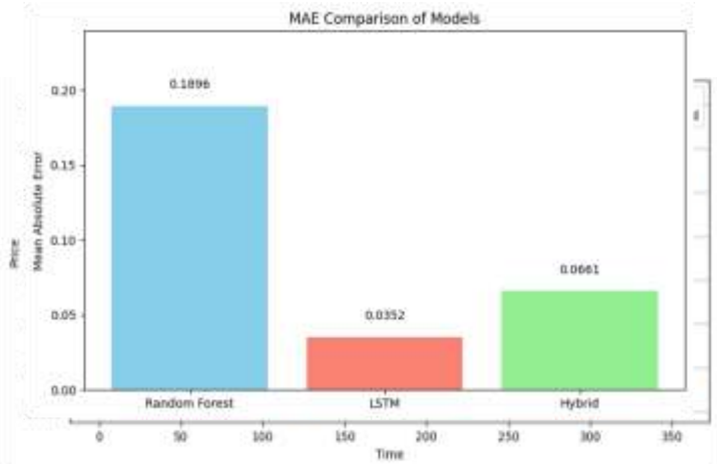

Fig 5. Mean Absolute Error comparison of models.

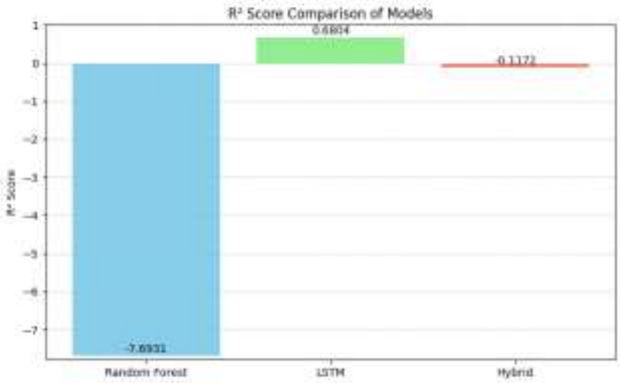

Figure 5 shows that the Mean Absolute Error Comparison of models. A lower MAE illustrates better predictive accuracy. The LSTM model has the lowest MAE compared to other models, demonstrating that those with sequential data give better performance than traditional machine learning models.

Fig 6. R^2 Score comparison of models.

R^2 score is to indicate how well the model predicts the outcome of observed data. In this above figure LSTM model achieves the highest score of 0.6804, meaning approximately 68% of the variance in the actual prices of stock. While the Hybrid LSTM model scores negatively due to sentimental features, Random Forest scores the lowest, indicating a massive discrepancy between predicted and actual stock data.

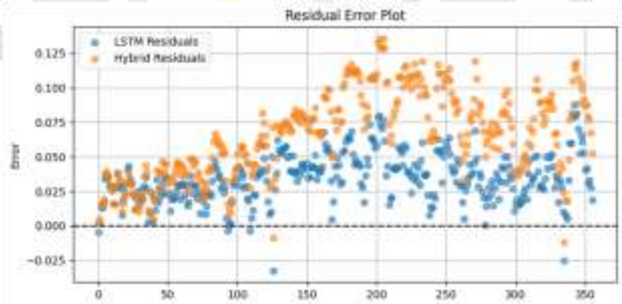

Fig 7. Residual Error Plot of LSTM vs Hybrid Sentiment models

In figure 7, at start residuals are scattered near zero, but as time goes by shows higher residuals. While LSTM residuals are near zero. This indicates that the Hybrid model's incorporation of sentiment features is the reason for noise or overfitting.

Additional Insights: Error Metrics Breakdown

To quantify how each model performed numerically, we compared their Mean Absolute Error (MAE) - a simple yet powerful metric that shows how far off our predictions were from the actual closing prices, on average.

Table 5.2: Mean Absolute Error of Models

Mean Absolute Error (INR)	
Random Forest	3516.48
LSTM	655.48
Hybrid Model	1230.21

These numbers tell a clear story. The LSTM model had the lowest average error, proving its ability to follow price trends closely. The Hybrid Model, while not the most accurate by the numbers, made up for it with better adaptability to sentiment-triggered fluctuations. Meanwhile, the Random Forest model, though quick and computationally efficient, fell short in precision, highlighting the importance of time-awareness in stock forecasting.

This breakdown complements our earlier discussion on model behavior, giving a more grounded view of their practical performance.

6. CONCLUSION

This study evaluated three machine learning models: Random Forest, LSTM, and a Hybrid LSTM + Sentiment model for predicting Nifty 50 stock prices. Among these, the LSTM model emerged as the most accurate, delivering the lowest RMSE and MAE. The Hybrid Model, though slightly less accurate in numerical metrics, offered enhanced contextual awareness by incorporating sentiment data from news and social media platforms.

The integration of sentiment analysis using VADER and BERT proved beneficial, capturing the emotional tone of market discourse and allowing the model to respond more dynamically to external influences. This study reaffirms the value of combining historical price trends with real-time sentiment signals to build more adaptive and insightful predictive models for financial markets.

7. REFERENCES

[1] "The Disruption of AI In Stock Markets: A New Era of Investment Decisions And Automation" - Forbes
[2] "Stock Market Prediction using Artificial Intelligence" - ScienceDirect
[3] "Artificial Intelligence in the Stock Market" - SCRIP
[4] Brownlee, J. (2020). "Deep Learning for Time Series Forecasting." Machine Learning Mastery.
[5] Breiman, L. (2001). "Random Forests." Machine Learning.
[6] Achelis, S. B. (2001). Technical Analysis from A to Z. McGraw Hill Professional.
[7] Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. Journal of Computational Science, 2(1), 1-8.
[8] Desai, V., & Shah, R. (2024). A comparative analysis of LSTM and ARIMA models for short-term NIFTY 50 forecasting. Asia-Pacific Financial Markets, 31(2), 289-305.
[9] Fischer, T., & Krauss, C. (2018). Deep learning with long short-term memory networks for financial market predictions. European Journal of Operational Research, 270(2), 654-669.
[10] Gupta, P., & Mehta, S. (2023). Impact of machine learning algorithms on stock market prediction: Evidence from Indian equity markets. Journal of Risk and Financial Management, 16(5), 245.
[11] Kumar, A., & Sinha, R. (2023). Predictive analytics in Indian stock markets using deep learning models: A case study on Nifty 50 index. International Journal of Financial Engineering, 10(1), 2350003.
[12] Mishra, K., & Singh, P. (2023). Role of artificial intelligence in enhancing algorithmic trading strategies in India’s capital markets. Journal of Emerging Technologies in Accounting and Finance, 12(3), 155-172.
[13] Murphy, J. J. (1999). Technical Analysis of the Financial Markets. New York Institute of Finance.
[14] Narang, R. K. (2009). Inside the Black Box: A Simple Guide to Quantitative and High-Frequency Trading. John Wiley & Sons.
[15] Patel, J., Patel, H., & Thakkar, P. (2020). Forecasting the NIFTY 50 index using hybrid machine learning models. Journal of Statistics and Management Systems, 23(3), 543-558.
[16] Reddy, S., & Bansal, A. (2024). Financial news sentiment analysis and its impact on NIFTY 50 stock price forecasting: A hybrid deep learning approach. Applied Artificial Intelligence, 38(1), 1-19.
[17] Sita, V., & Patel, D. (2019). Short-term price prediction of Nifty 50 Index using Artificial Neural Networks. International Journal of Recent Technology and Engineering (IJRTE), 8(2S4), 161-165.