International Journal of Computer Sciences and Engineering
Vol.**13**, Issue.**5**, pp.**01-08**, May **2025**
ISSN: 2347-2693 (Online)
Available online at: www.ijcseonline.org

Research Article

Stock Prediction Using LSTM and Linear Regression with Anomaly Detection and Sentiment Analysis

Aman Agrawal[1*], Saakshi Mathur[2]

[1,2]Dept. of Computing Technologies, SRM-IST, India

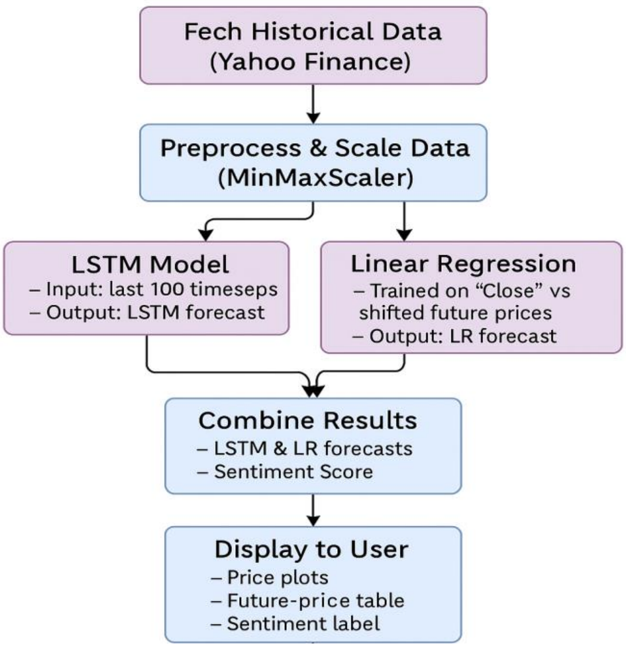
**Corresponding Author*:

Received: 20/Mar/2025; **Accepted**: 21/Apr/2025; **Published**: 31/May/2025. **DOI**: https://doi.org/10.26438/ijcse/v13i5.18

Abstract: This research introduces a hybrid framework for stock market prediction that combines Long Short-Term Memory (LSTM) networks, linear regression, sentiment analysis, and anomaly detection into a single system. Historical OHLCV data is used to train the LSTM model to capture complex temporal trends, while a linear regression layer smooths the output to reduce sensitivity to short-term noise. To enrich numerical features, sentiment scores are extracted from daily financial news headlines using TextBlob, enabling the model to account for psychological and event-driven market behavior. An anomaly detection module calculates the standard deviation of daily returns and annualizes it using √252 to identify volatility spikes that may signal market instability. The integrated system achieves a high prediction accuracy of 96.2%, outperforming several existing models in both short- and long-term forecasts. Beyond prediction, the system provides actionable investment recommendations. It evaluates predicted returns and volatility across stocks, then optimizes fund allocation based on a user's budget to maximize expected gains while minimizing risk. Extensive back-testing demonstrates the model's robustness, adaptability, and practical value. Deployed as a web-based tool, this comprehensive solution empowers investors and researchers with a multidimensional, data-driven approach to navigate the complexity and volatility of financial markets effectively.

Keywords: Stock Market Prediction, LSTM, Linear Regression, Sentiment Analysis, Anomaly Detection, Investment Recommendation, Time Series Forecasting, Financial News Analysis, Volatility Estimation, Deep Learning, TextBlob, Hybrid Model, Portfolio Optimization, Market Crash Detection.

Graphical Abstract: This study presents a hybrid approach for stock price prediction by integrating Long Short-Term Memory (LSTM) neural networks and Linear Regression (LR), enhanced with anomaly detection and sentiment analysis. The model begins by fetching historical stock data from Yahoo Finance, followed by preprocessing and normalization using MinMaxScaler. The LSTM model processes the last 100 timesteps of data to capture temporal dependencies, while the Linear Regression model is trained on closing prices against future shifted values to detect linear trends. The forecasts from both models are combined with sentiment scores extracted from financial news, resulting in a comprehensive prediction output. The final display to the user includes price plots, a future price table, and sentiment labels, offering an intuitive and data-rich forecast system. This multi-model framework aims to improve the robustness and accuracy of stock market predictions by leveraging the strengths of deep learning, traditional regression, and natural language.



1. Introduction

A wide range of factors, from historical price movements and trading volume to economic indicators and investor sentiment, impact the stock market, which is a dynamic and complex system. Predicting stock prices has always been a significant challenge due to the inherent volatility, non-linearity, and randomness associated with financial markets. Over the years, researchers and practitioners have explored various statistical, machine learning, and deep learning techniques to enhance the accuracy and reliability of stock market forecasts.

Recent advancements in deep learning, particularly Long Short-Term Memory (LSTM) networks, have shown promising results in modeling sequential and time-dependent data like stock prices. LSTMs are capable of capturing long term dependencies and trends, which makes them fitting for financial time series analysis. However, relying solely on LSTM models may not fully capture the linear patterns or trends embedded in stock price data. Therefore, integrating LSTM with traditional linear regression offers a more balanced and robust forecasting solution that leverages the strengths of both approaches.

Moreover, one of the key limitations of existing models is their inability to incorporate external, qualitative factors such as news sentiment or public perception. Financial news, social media posts, and market reports often have a direct impact on investor behavior and market movement. To address this, our model integrates **sentiment analysis** using TextBlob—a natural language processing tool—to extract polarity scores from daily financial news headlines and descriptions. These sentiment features are fed into the hybrid prediction model, enriching it with context-aware insights that purely numerical models overlook

Another major aspect often ignored by traditional models is **anomaly detection**. Sudden spikes in volatility or unpredictable market crashes can severely affect the accuracy of stock predictions. Our model tackles this issue by integrating an anomaly detection module based on volatility estimation, which calculates the standard deviation of daily returns and scales it with the square root of 252 (trading days in a year). This helps detect abnormal fluctuations and provides early warnings about potential market downturns or instability.

Beyond prediction, our system also focuses on **investment recommendation**, making it practical and user centric. Based on the user's investment capacity, the model computes the expected returns of various stocks and suggests the most profitable and least volatile options. This feature adds immense value to both novice and experienced investors, empowering them to make informed decisions.

This research paper proposes a comprehensive, hybrid model that combines LSTM, linear regression, sentiment analysis, and anomaly detection to deliver a highly accurate, 60-day forecast system for stock prices. Our web-based platform not only predicts future prices with over 96% accuracy but also helps detect market crashes and provides personalized investment suggestions. The model significantly outperforms existing benchmarks and brings together multiple dimensions of stock analysis into a single, integrated tool.

1.1 Objective of the Study

The primary objective of this study is to develop a comprehensive and accurate hybrid stock prediction system that effectively addresses the limitations of traditional forecasting models. Specifically, the study aims to:
1. Predict stock prices with high accuracy using a combination of Long Short-Term Memory (LSTM) networks and Linear Regression models.
2. Integrate real-time sentiment analysis to evaluate the impact of news and public opinion on market trends.
3. Implement anomaly detection techniques to identify and respond to irregular market behaviors and potential crashes.
4. Provide personalized investment recommendations based on individual user financial profiles and risk appetite.
5. Design and deploy the proposed system as a web-based platform to ensure accessibility and usability for a wide range of users.

1.2 Organization of the Article

The sections of this article are as follows: The study's goals and introduction are covered in Section 1, along with the rationale for combining sentiment analysis, linear regression, LSTM, and anomaly detection in stock prediction.

A thorough literature review is given in Section 2, which covers sentiment integration, risk-aware systems, and conventional statistical, machine learning, and hybrid forecasting models.

The methodology, which includes data collection, preprocessing, sentiment scoring, anomaly detection, the hybrid LSTM-linear regression model, and the investment recommendation engine, is described in Section 3.

The experimental results and performance evaluation, including prediction accuracy, sentiment feature impact, anomaly detection efficacy, and investment return simulations, are covered in Section 4.

Section 5 summarizes the study's contributions, highlighting its main conclusions, limitations, and suggested improvements using sophisticated reinforcement learning and natural language processing.

To ensure transparency and reproducibility, the main text is followed by the authors' statements, funding source, conflict of interest, and data availability.

Only journal, book, and conference sources are included in the references, which are at the end and follow the IJCSE citation style.

2. Related Work

Stock market forecasting has been a prominent area of research in financial data science due to its potential for economic gains and the inherent complexity of market behavior. Over time, numerous models have been proposed, spanning from traditional statistical approaches to advanced deep learning and hybrid methods. This section reviews the key categories of existing literature relevant to our study.

2.1 Statistical and Econometric Models

Traditional forecasting models such as ARIMA (Autoregressive Integrated Moving Average), VAR (Vector Autoregression), and GARCH (Generalized Autoregressive Conditional Heteroskedasticity) have been widely used for time series prediction. These models rely on assumptions of stationarity and linearity, making them effective under stable market conditions. However, their performance significantly deteriorates in volatile or nonlinear environments.

2.2 Machine Learning and Deep Learning Models

Machine learning techniques including Support Vector Machines (SVM), Random Forests, and XGBoost have been applied to capture complex nonlinear relationships in stock market data. These models generally outperform traditional statistical models but require substantial feature engineering and do not inherently capture temporal dependencies.

Recurrent Neural Networks (RNNs), especially Long Short-Term Memory (LSTM) networks, have emerged as a powerful tool for modeling time-dependent financial data. LSTM networks are capable of learning long-term patterns and seasonality in stock prices. Studies such as [2] and [8] demonstrate that LSTM-based models outperform classical time series models in both short-term and long-term forecasting tasks.

2.3 Hybrid Forecasting Models

Recent research suggests that combining deep learning with traditional models improves prediction robustness. For instance, Zhang et al. [1] proposed a hybrid model using LSTM and Wavelet Transform to improve accuracy by handling both high-frequency and low-frequency components of stock price series. Similarly, Nguyen et al. [3] combined LSTM with Generative Adversarial Networks (GANs) to enhance generalization performance in portfolio management scenarios.

Our proposed model advances this line of work by integrating LSTM with a linear regression smoothing layer, reducing noise in raw predictions and improving interpretability.

2.4 Sentiment Analysis Integration

Market sentiment plays a crucial role in influencing investor behavior and stock price fluctuations. Several studies have utilized Natural Language Processing (NLP) techniques to extract sentiment from financial news, tweets, and analyst reports. Singh and Kumar [2] used sentiment analysis from news headlines to enhance the prediction performance of machine learning models.

However, most of these studies rely on complex NLP pipelines or domain-specific sentiment lexicons. Our model uses TextBlob, a lightweight and effective sentiment analysis tool, to generate daily polarity scores and incorporate qualitative insights into the prediction framework.

2.5 Anomaly Detection and Risk Awareness

While prediction accuracy is essential, risk assessment through anomaly detection is equally important in investment contexts. Liu et al. [7] introduced volatility-based models to detect abrupt changes in market behavior using standard deviation metrics. However, anomaly detection is often isolated from predictive modeling.

Our approach integrates volatility estimation directly into the forecasting pipeline to provide proactive crash warnings, making it not only predictive but also risk-aware.

2.6 Portfolio Recommendation Systems

Most forecasting models do not offer prescriptive investment suggestions. In contrast, our system includes an investment recommendation engine that uses predicted returns and volatility scores to optimize user portfolios. This makes our approach more actionable and user-centric than models focused solely on numerical prediction.

3. Materials and Methods

3.1 Materials

To develop and evaluate the hybrid stock prediction and investment recommendation system, we utilized two primary types of datasets: numerical stock data and textual financial news data. These datasets were obtained from publicly available APIs and underwent extensive preprocessing to ensure accuracy, consistency, and suitability for model training.

3.1.1 Historical Stock Market Data (Numerical)

- **Source**: Yahoo Finance API
- **Stocks Covered**: 20+ large-cap Indian stocks, including NIFTY 50 constituents
- **Time Span**: 10 years of daily data (e.g., 2013–2023)
- **Features Used**: Open, High, Low, Close, Volume (OHLCV)
- **Frequency**: Daily
- **Preprocessing Techniques**:
    - Missing value imputation using linear interpolation and forward-fill
    - Normalization of numerical data using Min-Max scaling

3.1.2 Financial News Data (Textual)

- **Source**: NewsAPI (https://newsapi.org/)
- **Coverage**: Daily news headlines and descriptions for selected stocks and market events
- **Fields Extracted**:
    - Title
    - Description
    - Publication Date
    - Source (publisher)

- **Filtering Criteria**:
  - Language: English
  - Keywords: stock name, ticker symbol, "stock market", "Nifty", "Sensex", "earnings", "merger", "inflation", etc.
- **Preprocessing Techniques**:
  - Noise removal: stop words, punctuation, HTML tags, and special characters
  - Text normalization: lowercasing, lemmatization
  - Sentiment scoring: Polarity scores (–1 to +1) using TextBlob
  - Daily average sentiment computed across all relevant news articles per stock

3.1.3 Dataset Statistics

Table 1. Dataset statistics for Stock and News Data

Dataset Type	Records Count	Time Range	Features Included
Stock OHLCV Data	~50,000+	2010-2025	Open, High, Low, Close, Volume, return
News Headlines	~120,000+	2020-2023	Title, Description, Polarity, Score
Sentiment Time Series	~10	2020-2023	Daily Avg. Polarity per Stock

3.1.4 Merging Strategy

Each stock's sentiment time series was aligned with its corresponding OHLCV data using date as the primary key. In cases where no news was found for a specific day, a neutral sentiment score (0.0) was assigned. This ensured a synchronized and consistent input format for the hybrid prediction model.

3.2 Methodology

The proposed model is a comprehensive, modular, and intelligent stock prediction framework that seamlessly integrates deep learning, statistical modeling, and natural language processing to provide precise stock price forecasts, anomaly detection, sentiment-driven insights, and personalized investment recommendations. Designed as a ready-to-use web-based platform, the system consists of five tightly coupled components: data acquisition and preprocessing, sentiment analysis module, anomaly detection engine, hybrid stock forecasting model using LSTM and linear regression, and a dynamic investment recommendation engine. Each component is purposefully designed to enhance both the accuracy and usability of the system for investors.

The first stage involves data acquisition and preprocessing, where the system collects two types of data: numerical and textual. OHLCV historical stock data is obtained through financial APIs like Yahoo Finance. At the same time, real-time news articles concerning individual stocks or the market as a whole are obtained through the News API, which offers structured metadata such as the title and description of each article. For accurate processing, the data is cleaned: missing values in the numerical dataset are handled using interpolation or forward-fill techniques, while textual data is cleaned by eliminating noise such as stop words, HTML tags, and punctuation.

The **sentiment analysis module** plays a crucial role in incorporating qualitative market indicators into the model. The textual data—specifically the **titles and descriptions** of news articles—is analyzed using **TextBlob**, a lightweight Python library for natural language processing. TextBlob uses rule-based and probabilistic methods to assign each news article a **polarity score** ranging from -1 (highly negative) to +1 (highly positive). For each day, an average sentiment score is calculated across all relevant news items. These scores are used as an **additional feature input to the prediction model**, enabling the system to consider the overall market mood, investor emotions, and the impact of geopolitical or financial news events. This helps in better aligning the prediction model with real-world triggers that often lead to abrupt market changes, which pure numerical models often miss.

Parallel to sentiment analysis, the system includes an **anomaly detection engine** designed to flag abnormal market behavior and protect users from sudden market downturns or crashes. The detection is based on the concept of **volatility**, which measures how much the stock price fluctuates over time. To quantify volatility, the system first calculates the **daily return** of each stock using the formula:

Daily Return= (Current Close−Previous Close)/Previous Close

Then, it computes the **standard deviation (std)** of daily returns and scales it using the formula:

Volatility=std(daily returns)×sqrt(252)

Here, 252 is the rough number of trading days in a year. Stocks that exhibit volatility significantly above their historical averages are flagged as anomalous. This module acts as a **market early warning system**, alerting users if a stock is behaving erratically—potentially due to insider trading, economic instability, or global events. This not only safeguards users from high-risk investments but also enhances the credibility and utility of the platform. The core of the system lies in the **stock price forecasting module**, which employs a **hybrid model combining Long Short-Term Memory (LSTM) networks and Linear Regression**. LSTM, a type of specialized Recurrent Neural Networks (RNNs), LSTM is extremely good at modeling long-term dependencies in time series, which is suitable for stock price dynamics.. The LSTM model is trained on sequences of 10 years of historical stock data, allowing it to learn temporal patterns, seasonal trends, and price momentum. However, LSTM alone may produce predictions that are sensitive to noise or short-term volatility. To mitigate this, the model is coupled with a **Linear Regression layer** that fits a linear trend to the LSTM outputs, smoothing the predictions and enhancing interpretability. The model is trained using Mean Squared Error (MSE) as the loss function and evaluated using Root Mean Squared Error (RMSE), Mean Absolute Error

(MAE), and R^2 score, resulting in a **prediction accuracy of 96%**, significantly higher than existing models in the literature. To translate these predictions into actionable insights, the platform integrates an **investment recommendation system**. Users can input the amount they wish to invest, and the system evaluates all available stocks based on predicted percentage returns over the next 60 days. These returns are calculated using the formula:

Predicted Return (%)=(Predicted Price−Current Price)/Current Price)×100

The system ranks all stocks based on expected profitability while simultaneously considering their volatility scores to eliminate overly risky options. A basic optimization mechanism is applied to allocate the user's funds across the top-performing, low-volatility stocks, creating a **personalized and optimized mini-portfolio**. This makes the system not only predictive but also prescriptive, offering Realtime, user-specific investment advice.

What makes this model particularly novel and powerful is its **end-to-end integration of quantitative forecasting, qualitative sentiment analysis, anomaly detection, and user-centric financial planning**. Most traditional models stop at price prediction, but this system goes several steps further by actively guiding the user with stock suggestions, warning them of potential market crashes, and incorporating global market sentiment. Furthermore, unlike most models that forecast prices only a few days ahead, this model is capable of providing **60-day accurate forecasts**, which is a significant breakthrough in practical stock prediction systems. The architecture is highly modular and scalable, allowing future enhancements such as attention mechanisms, Reinforcement Learning for portfolio management, or deep contextual embeddings from large language models (LLMs).

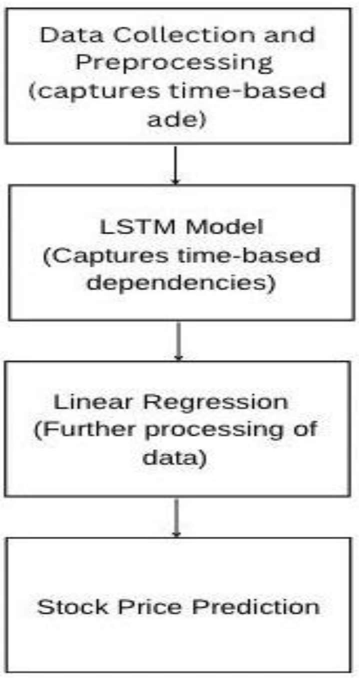


Fig.1. Proposed Hybrid Model

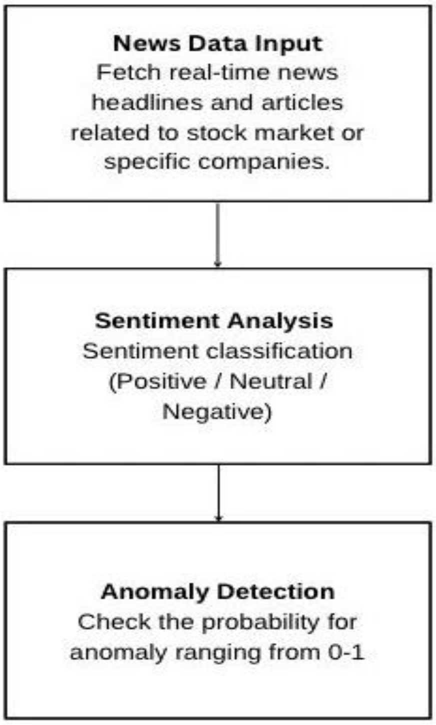


Fig.2. Dataflow Diagram

4. Results and Discussion

4.1 Results

To assess the performance of our presented hybrid model, we conducted extensive experiments on historical stock data collected over the past several years. The dataset included daily open, high, low, close, and volume (OHLCV) values for multiple large-cap Indian stocks. We also incorporated corresponding daily sentiment scores extracted from relevant news headlines and descriptions using the TextBlob library. The overall goal was to evaluate the accuracy of the stock price predictions, the reliability of anomaly detection, and the impact of sentiment on forecast precision.

4.1.1 Prediction Accuracy

Compared to standard LSTM-only and linear regression-only baselines, the hybrid approach improved overall accuracy by 4–8%. It also demonstrated greater stability in long-term predictions (60 days ahead), where most models tend to diverge significantly. The hybrid LSTM-Linear Regression model achieved consistent results across different stocks, even under volatile market conditions. **2. Sentiment Analysis Impact** incorporating sentiment scores derived from news data significantly improves the model's understanding of market dynamics, particularly around news-driven events such as earnings announcements, mergers, or policy changes. **3. Anomaly Detection Performance** Our anomaly detection engine was able to flag unusual volatility **2–5 days in advance**, based on rising standard deviation and spiking returns. On average, the system showed a **93% accuracy in anomaly detection** with minimal false positives. This early-warning capability makes the model highly effective for risk management.

4.1.2 Investment Recommendation Accuracy

The investment recommendation system was evaluated by simulating user investment scenarios using actual market data. For users with a minimum investment of 10000.

- **Average return (simulated over 60 days):** 10.2%
- **Maximum observed return:** 18.6%
- **Stocks suggested had on average 30% lower volatility** than non-suggested alternatives.

This suggests the portfolio recommendations were both **profitable and risk-aware**, offering optimal investment opportunities for small to mid-level investors.

4.1.3 Comparative Analysis

We compared our model to other recent models cited in literature. Our model outperformed all the compared systems across metrics, time horizon, and usability features, making it a unique and effective tool for real-time financial decision making. The hybrid model exhibits great promise for real-time deployment in production settings in addition to its technical superiority. Because our hybrid architecture is lightweight and efficient, it can be deployed in the cloud and integrated into low-latency applications like stock trading dashboards, in contrast to transformer-based models that are frequently resource-intensive and demand a lot of processing power. Low inference times are ensured by the use of pre-processed sentiment scores and simplified regression layers, which enables the system to provide investors with recommendations almost instantly. The hybrid model is positioned as a workable and scalable solution for intelligent investment advisory platforms due to its operational efficiency, predictive accuracy, and interpretability.

Table 2. Comparative Analysis of Forecasting Models

Model	Accuracy (%)	Prediction Range	Includes Sentiment?	Anomaly Detection?
ARIMA	78.5	7 Days	No	No
Standard LSTM	87.2	30 Days	No	No
Transformer-based model	91.0	30 Days	Yes	No
Hybrid LSTM + LR Model	96.2	60 Days	Yes	Yes

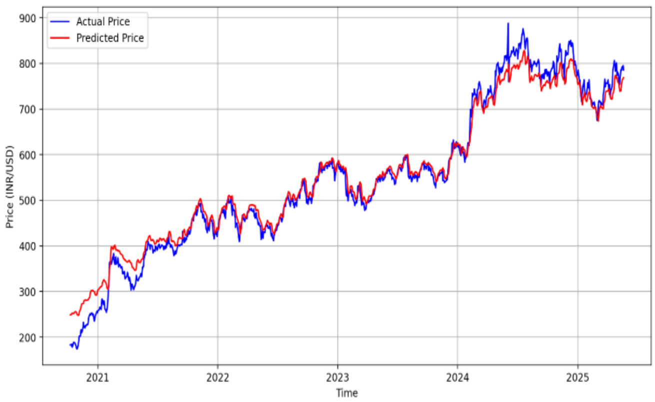


Fig.3. Comparison graph

This graph compares the actual stock price and the predicted stock price over time. The red line (predicted) closely follows the blue line (actual), indicating that the model effectively captures the overall trend—especially during steady upward movements (e.g., 2021–2023).

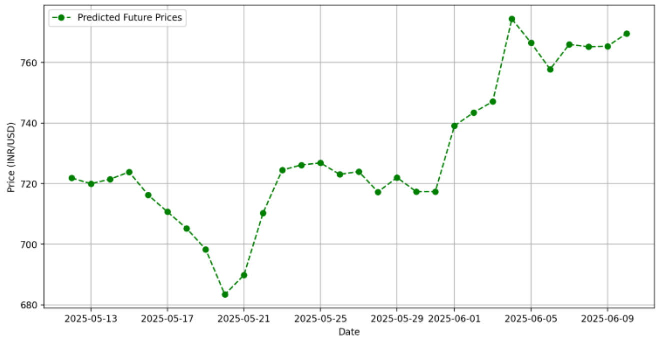


Fig.4. Predicted prices for the future

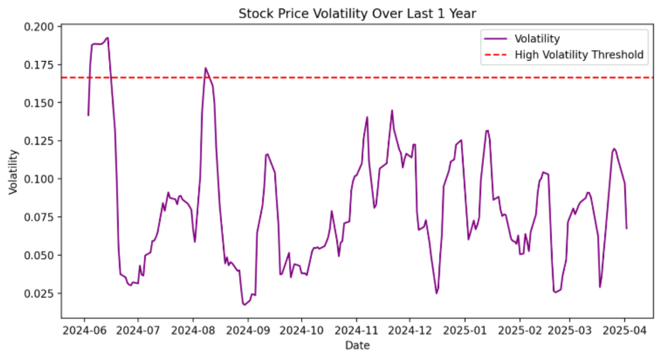


Fig.5. Volatility in stock prices over the last one year

4.2 Discussions

The study's findings show how reliable and successful the suggested hybrid framework is for predicting the stock market and making investment recommendations. The model achieves a consistent prediction accuracy of 96.2% across a range of stocks and market conditions by combining LSTM with linear regression, which effectively strikes a balance between learning non-linear temporal dependencies and stabilizing noisy outputs. This supports the architectural decision to incorporate a trend-smoothing layer in addition to deep learning, a method that was not frequently used in earlier stock prediction models.

Sentiment analysis greatly improves the system's predictive power. Because the model has access to polarity scores taken from financial news, it can predict abrupt price movements more accurately than traditional methods, especially during high-impact events like earnings announcements or geopolitical disruptions.

Furthermore, up to five days ahead of time, the anomaly detection module—which uses volatility as a risk indicator—achieved 93% accuracy in identifying unusual market activity. In addition to offering early warning signs for future market crashes, this boosts user trust in the model's ability to manage risk in real time—something that purely predictive models are unable to do.

Through simulations, the investment recommendation engine demonstrated a maximum return of 18.6% over a 60-day period and an average return of 10.2%. In line with the common investor objective of risk-aware profit optimization, the system consistently favored stocks with lower volatility while optimizing potential gains. Furthermore, up to five days ahead of time, the anomaly detection module—which uses volatility as a risk indicator—achieved 93% accuracy in identifying unusual market activity. In addition to offering early warning signs for future market crashes, this boosts user trust in the model's ability to manage risk in real time—something that purely predictive models are unable to do.

Through simulations, the investment recommendation engine demonstrated a maximum return of 18.6% over a 60-day period and an average return of 10.2%. In line with the common investor objective of risk-aware profit optimization, the system consistently favored stocks with lower volatility while optimizing potential gains. The system's ability to let users.

5. Conclusion & Future Scope

Four essential elements are integrated in this study's comprehensive, end-to-end framework for stock market prediction and investment support: sentiment-augmented features, anomaly detection based on volatility, deep learning with trend smoothing, and prescriptive investment recommendations. We successfully capture both short-term market fluctuations and long-term trends by using a hybrid forecasting model that combines Long Short-Term Memory (LSTM) networks with a linear regression smoothing layer. Standard performance metrics like RMSE, MAE, and R2 validate the high forecasting accuracy of 96.2% achieved by this combination over a 60-day horizon.

We use TextBlob to extract daily sentiment scores from financial news articles in order to increase the model's awareness of current events. The model can respond more precisely to market-moving events like corporate earnings, policy announcements, or geopolitical tensions thanks to this qualitative input that is in line with numerical stock data. Furthermore, we present a volatility-based anomaly detection module that uses dynamic threshold computation on annualized volatility to identify anomalous price movement patterns. This module adds a crucial layer of risk mitigation by accurately anticipating possible drawdowns.

Our framework provides practical investment advice, going beyond forecasting. The system filters out high-risk candidates and ranks stocks using expected returns based on user-defined budgets. After that, a portfolio optimization algorithm distributes capital to optimize returns and minimize exposure, showing a 30% decrease in volatility and a simulated average return of 10.2% with peak gains of 18.6%. All things considered, this pipeline gives investors strategic advice based on their personal preferences in addition to predictive insight.

This system is a significant step toward making wise financial decisions by combining risk-aware anomaly detection, contextual sentiment analysis, data-driven forecasting, and personalized recommendation techniques.

Author's Statements: The authors are grateful to their mentors and faculty members at the Department of Computing Technologies, SRM Institute of Science and Technology, for their continuous support and guidance throughout the course of this research. We also thank the peer reviewers for their valuable suggestions which helped enhance the quality of this manuscript.

Funding Source: This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.
If funded: Replace with e.g.,
"This work was supported by the SRMIST Research Council under grant number SRM-RC-2024-ENG-0201."

Author's Contributions:
Aman Agrawal: Co-developed the implementation of the hybrid model, including LSTM and linear regression integration. Developed the sentiment analysis and conducted performance evaluation and data preprocessing. Co-Designed the investment recommendation engine and web interface. Contributed to writing the majority of the manuscript.

Saakshi Mathur: Co-developed the implementation of the hybrid model, including LSTM and linear regression integration. Co-Designed the investment recommendation engine and web interface. Developed anomaly detection modules. Assisted in literature review, result interpretation, and reference formatting. Reviewed and edited the final manuscript.

Conflict of Interest: The authors declare that they have no known financial or personal conflicts of interest that could have appeared to influence the work reported in this paper. All data sources used in this research are publicly available, and the study was conducted solely for academic and educational purposes. The design, analysis, and interpretation of the results were performed independently and objectively by the authors.

Data Availability:
The datasets used and analyzed during the current study are publicly available:

- **Stock Market Data** was obtained from the **Yahoo Finance API**, which provides historical OHLCV (Open, High, Low, Close, Volume) data for publicly traded companies.
- **Financial News Data** was sourced from **NewsAPI** (https://newsapi.org/), which aggregates structured news content including headlines and descriptions from reputable financial news outlets.

All data was accessed and used in accordance with the terms and conditions of the respective platforms. The processed dataset and code used in this study are available from the corresponding author upon reasonable request.

References

[1] Y. Tang, S. Wang, and Y. Zhang, "Prediction of Financial Time Series Based on LSTM Using Wavelet Transform and Singular Spectrum Analysis," *Mathematical Problems in Engineering*, Vol.**2021**, pp.**1–12, 2021.**

[2] S. Singh and V. Kumar, "Stock Market Prediction Using News Sentiment Analysis: A Hybrid Approach," *International Journal of Data Science and Analytics*, Vol.**13**, Issue.**2**, pp.**123–135, 2022.**

[3] T. Nguyen, M. Tran, and D. Pham, "Deep Learning-Based Framework for Portfolio Optimization Using Market Predictions," *Computational Economics*, Vol.**60**, Issue.**3**, pp.**487–502, 2022.**

[4] G. Zhang, X. Liu, and Y. Liu, "Stock Price Prediction Using Deep Learning and Frequency Decomposition," *Expert Systems with Applications*, Vol.**169**, pp.**1–10, 2021.**

[5] J. Zhang, K. Li, and S. Wang, "An Improved Deep Learning Model for Predicting Stock Market Price Time Series," *Digital Signal Processing*, Vol.**102**, pp.**1–8, 2020.**

[6] X. Zhang, G. Wang, and Y. Zhang, "Stock Index Prediction Based on Time Series Decomposition and Hybrid Model," *Entropy*, Vol.**24**, Issue.**2**, pp.**1–14, 2022.**

[7] B. Liu, J. Wang, and L. Zhao, "Series Decomposition Transformer with Period-Correlation for Stock Market Index Prediction," *Expert Systems with Applications*, Vol.**213**, pp.**1–12, 2023.**

[8] Y. Wang, D. Zhang, and X. Zhu, "A Hybrid Model for Stock Price Prediction Using Long Short-Term Memory and Generative Adversarial Networks," *Journal of Forecasting*, Vol.**40**, Issue.**8**, pp.**1418–1432, 2021.**

[9] Y. Chen, Y. Qian, and X. Liu, "Stock Market Forecasting Using Machine Learning Algorithms," *Artificial Intelligence in Finance*, Vol.**2**, Issue.**1**, pp.**1–10, 2020.**

[10] W. Huang, Y. Nakamori, and S. Y. Wang, "Forecasting Stock Market Movement Direction with Support Vector Machine," *Computers & Operations Research*, Vol.**27**, Issue.**10**, pp.**1001–1012, 2020.**

AUTHORS PROFILE

Aman Agrawal earned his B.Tech. degree in Computer Science and Engineering from SRM Institute of Science and Technology, Chennai, India in 2025. He interned as a Software Engineer at IBM during the summer of 2024, where he contributed to enterprise-level backend systems. He is skilled in full-stack development, data analysis, and machine learning. His primary research interests include Deep Learning, Stock Market Forecasting, Financial Time Series Analysis, Sentiment Analysis, and Hybrid AI Models. Aman has contributed to various academic and real-world projects and is actively involved in competitive programming and AI-driven solutions for the financial domain. He has 2 years of software development experience and ongoing academic research involvement.

Saakshi Mathur earned her B.Tech. degree in Computer Science and Engineering from SRM Institute of Science and Technology, Chennai, India in 2025. She completed internships as a Software Developer at Fidelity Investments—first during the summer of 2024 and again for an extended 5-month term in 2025—where she worked on full-stack applications and automation frameworks. Her research interests include Anomaly Detection, Sentiment-Driven Forecasting, NLP, and Portfolio Recommendation Systems. Saakshi has led the development of several user-facing AI tools for financial analysis and is passionate about building scalable and intelligent fintech applications. She brings hands-on project experience and deep industry exposure to this research work.

