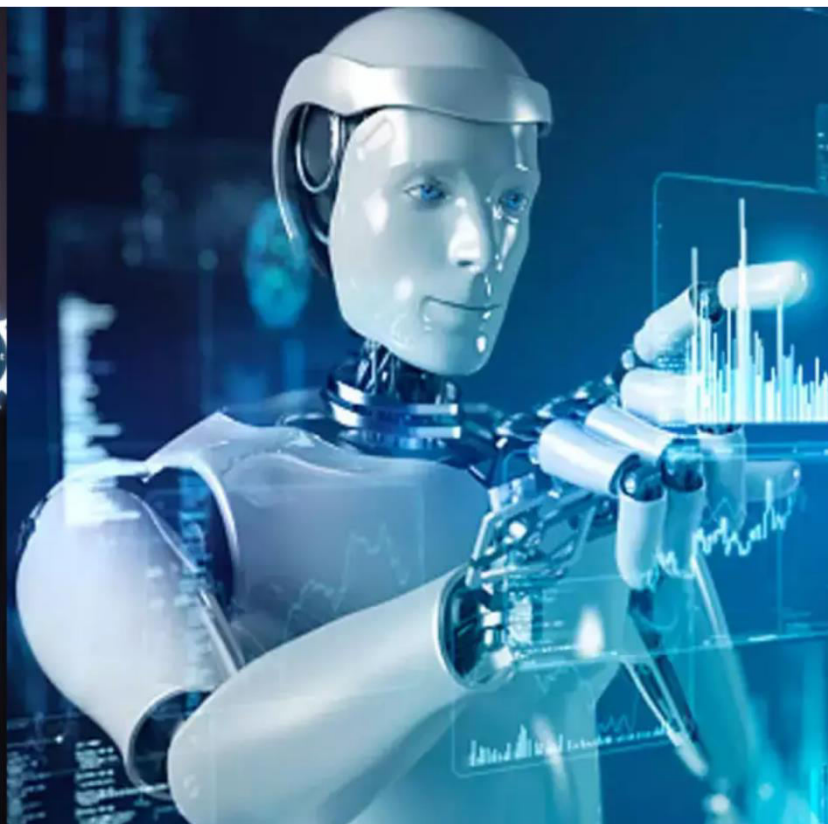




International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.699

Volume 14, Issue 4, April 2025

Sentimental Analysis in Educational Clustering and Real Time Sentimental Analysis

B. Sai, N. Harish Kumar Reddy, C. Pavan Kumar, N. Vamshi Krishna, K. Muthulakshmi

UG Student, Department of CSE, Bharath Institute of Higher Education and Research, Chennai, India

UG Student, Department of CSE, Bharath Institute of Higher Education and Research, Chennai, India

UG Student, Department of CSE, Bharath Institute of Higher Education and Research, Chennai, India

UG Student, Department of CSE, Bharath Institute of Higher Education and Research, Chennai, India

Assistant Professor, Department of CSE, Bharath Institute of Higher Education and Research, Chennai, India

ABSTRACT: In this paper, introduces an advanced stock price prediction and analysis system that combines machine learning technologies with financial indicators for comprehensive market insights. Developed in Python, it integrates Random Forest, LSTM neural networks, and ARIMA models for robust forecasting, while incorporating sentiment analysis of financial news to enhance prediction accuracy. The system features an interactive Stream lit web application, allowing users to input stock tickers and date ranges for customized analyses, presenting visual representations of technical indicators, predictions, and back testing results. Risk analysis tools, including Sharpe ratio calculations, are implemented to evaluate strategy performance. Addressing ethical considerations in AI-based financial modeling, the project promotes responsible use of predictive technologies in stock markets. By combining advanced data analysis, machine learning techniques, and financial expertise, this system aims to assist investors in making informed decisions, positioning itself as a comprehensive solution for stock market analysis and prediction in the evolving FinTech. Landscape

KEYWORDS: Stock Forecasting, Machine Learning, Sentiment Analysis, LSTM, Random Forests, ARIMA, Ensemble Modeling, Technical Indicators, Back testing, Risk Analysis

I. INTRODUCTION

In the dynamic and complex world of financial markets, accurate stock price forecasting is a powerful and highly sought after capability. Stock price forecasting provides significant value to investors, traders and financial institutions, but inherent variability and numerous influencing factors make this task particularly challenging Recent advances in artificial intelligence and machine learning have opened up new ways to address this challenge. This technology can analyze large amounts of data, identify complex patterns, and make more accurate predictions. This project aims to develop a comprehensive stock price forecasting and analysis system using AI and machine learning. Our approach combines several advanced methods: We use a combination of random forest, long-term and short-term memory (LSTM) root networks, and ARIMA models, each of which captures various aspects of the stock price movement Moving Average, Relative Strength Index (RSI), Moving Average Convergence Deviation (MACD) ho) etc. in order to increase the forecasting accuracy of the system Includes widespread use of technical indicators Since we recognize the impact of market sentiment , we apply real-time sensitivity analysis of financial reporting using natural language processing techniques. We use an ensemble approach to take advantage of the robustness of the individual models, which can improve the overall forecast accuracy. The system includes external testing capabilities to evaluate trading strategies based on our forecasts, as well as risk assessment using metrics such as the Sharpe ratio This project aims to provide accurate stock price forecasts and a comprehensive tool for stock market analysis. By combining multiple models, technical indicators and sentiment analysis, we seek to build a company

II. LITERATURE SURVEY

Leveraging Deep Learning for Enhancing Social Distancing Compliance literature related to stock price prediction using machine learning. Keep in mind that new research may have been published since then. Here are some key themes and notable studies:

(1) Traditional Machine Learning Approaches:

Researchers have explored the use of various traditional machine learning algorithms such as linear regression, decision trees, support vector machines, and ensemble methods for predicting stock prices based on historical data and financial indicators.

Time Series Models:

Time series models, particularly ARIMA (Autoregressive Integrated Moving Average) and its variations (SARIMA), have been widely used for modelling and forecasting stock prices. These models consider the temporal dependencies in the data.

(2) Neural Networks:

Deep learning techniques, including neural networks, have gained popularity for stock price prediction due to their ability to capture complex patterns. Recurrent Neural Networks (RNNs) and Long Short-Term Memory networks (LSTMs) are commonly used architectures in this context.

(3) Feature Engineering:

Studies have focused on effective feature engineering to enhance the predictive power of models. Features may include technical indicators, fundamental factors, market sentiment derived from news sentiment analysis, and macroeconomic variables.

(4) Sentiment Analysis:

Incorporating sentiment analysis of financial news, social media, and other textual data has been explored to gauge market sentiment and its impact on stock prices.

(5) Ensemble Methods:

Ensemble methods like bagging and boosting have been applied to combine predictions from multiple models to improve overall accuracy and robustness.

(6) Reinforcement Learning:

Some researchers have explored the application of reinforcement learning techniques for stock trading, where an agent learns optimal trading strategies over time.

III. METHODOLOGY

Our stock price prediction and analysis system employs a multi-faceted approach, integrating various data sources, advanced machine learning techniques, and financial analysis methods. The methodology can be broken down into the following key components:

Data Collection and Preprocessing

1. **Stock Data Acquisition:** We utilize the `yfinance` library to fetch historical stock data. This includes daily Open, High, Low, Close prices, and trading Volume for the specified stock ticker.
2. **Data Cleaning:** The acquired data is checked for missing values and outliers. Any missing data is handled through appropriate imputation methods or removal, ensuring the integrity of our dataset.

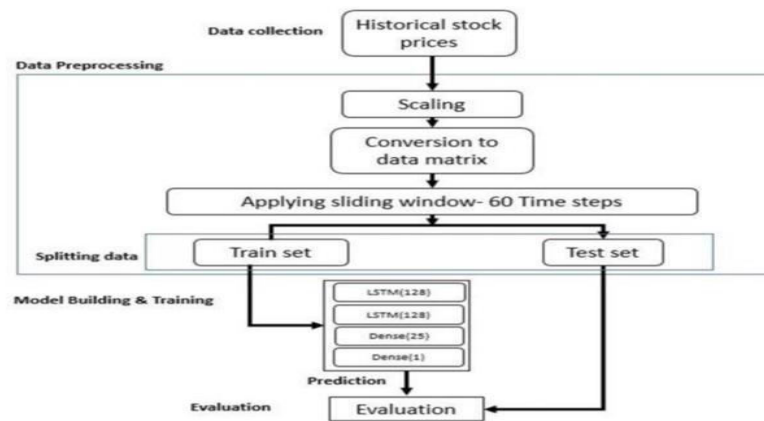
Feature Engineering

1. **Technical Indicators:** We calculate and add several technical indicators to our dataset to capture various aspects of stock price movements:
 - Moving Averages (MA20 and MA50)
 - Relative Strength Index (RSI)
 - Moving Average Convergence Divergence (MACD)

2. Additional Features

We engineer additional features to capture more nuanced information:

- Price Change
- Volume Change
- High-Low Difference



Datasets

1. Historical Stock Price Data:

- Description: This dataset includes historical stock prices for the target stocks, typically covering daily or minute-by-minute closing prices, opening prices, trading volumes, and other relevant financial indicators.¹¹

- Source: Financial market data providers, stock exchanges, or APIs from platforms like Yahoo Finance, Alpha Vantage, or Quandl.

2. Financial Indicators:

- Description: Dataset containing fundamental financial indicators for the companies, including earnings per share (EPS), price-to-earnings ratio (P/E), dividend yield, and other financial ratios.

- Source: Company financial reports, financial news platforms, or specialized financial data providers.

3. Technical Indicators:

- Description: Dataset featuring technical indicators derived from historical stock prices, such as moving averages, relative strength index (RSI), and stochastic oscillators.-

Source: Calculated from historical stock price data or obtained from financial data providers offering technical analysis

Performance Evaluation Metrics –

When evaluating the performance of stock price prediction models using machine learning, simplicity is often key. Here are some straightforward and commonly used performance evaluation metrics:

Regression Metrics:

1. Mean Absolute Error (MAE):

- Formula: $MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$

- Measures the average absolute difference between actual and predicted stock prices.

2. Root Mean Squared Error (RMSE):

- Formula: $\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$
- Like MAE but gives more weight to large errors.

Classification Metrics (if predicting price direction):

1. Accuracy:

- Formula: $\text{Accuracy} = \frac{\text{Number of Correct Predictions}}{\text{Total Number of Predictions}}$
- Measures the percentage of correctly predicted price directions.

Additional Metrics:

1. Sharpe Ratio (for portfolio optimization):

- Formula: $\text{Sharpe Ratio} = \frac{\text{Average Portfolio Return} - \text{Risk-Free Rate}}{\text{Portfolio Standard Deviation}}$

- Evaluates the risk-adjusted performance of a portfolio, considering both returns and volatility.

2. Profit and Loss (P&L):

- Quantifies the profitability of trading strategies based on the model's predictions and actual market outcomes.

Time Series Evaluation:

1. Lag Analysis:

- Analyze the lag between predicted and actual prices to understand the model's ability to capture short-term or long-term trends. Model Interpretability:

2. Feature Importance:

- Assess the importance of different features in influencing the model's predictions, providing insights into the factors driving stock price movements. Robustness and Stability:

3. Cross-Validation:

- Implement cross-validation techniques (e.g., k-fold cross validation) to ensure the model's robustness and generalization to unseen data. These simple metrics offer a quick and intuitive way to assess the model's accuracy, its ability to predict price directions, and its performance in the context of portfolio optimization. It's important to choose metrics that align with the specific goals of the stock price prediction task and the nature of the dataset. Additionally, visualizations like time series plots and feature importance charts can provide further insights into model performance.

IV. RESULT

Dataset Overview:

Firstly, I have performed Data Overview for stock price of companies. Fig. represent the date, open, close, high, low, adjusted close and volume of stocks details.

Stock Price Prediction and Analysis

Enter Stock Ticker
AAPL

Start Date
2010/01/01

End Date
2023/12/31

Analyze

Stock Price Prediction and Analysis

Enter Stock Ticker
AAPL

Start Date
2010/01/01

End Date
2024/10/23

Number of days to predict into future
10

Analyze

Fetching and analyzing data...

Training LSTM model...

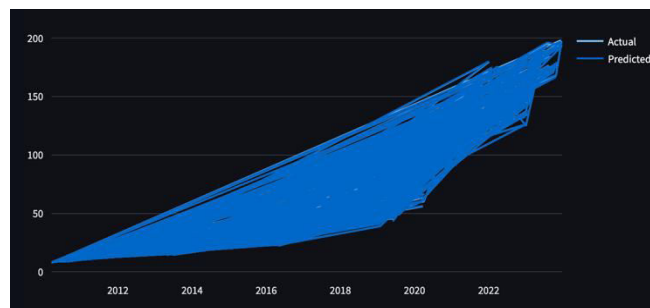
Graph of Technical Indicators , MACD , RSI





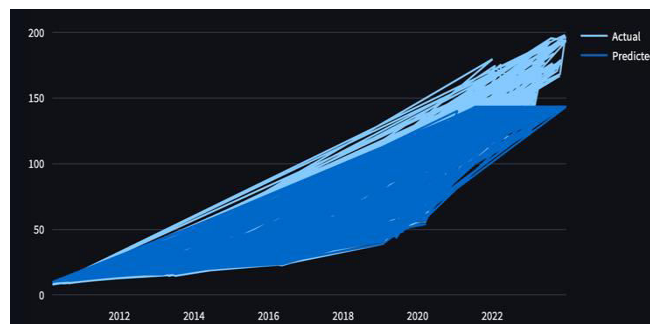
Random Forest Model Results

MSE:0.020282513024692816
MAPE:0.0012369955464367657
R2 Score: 0.9999929327253243



LSTM Model Results

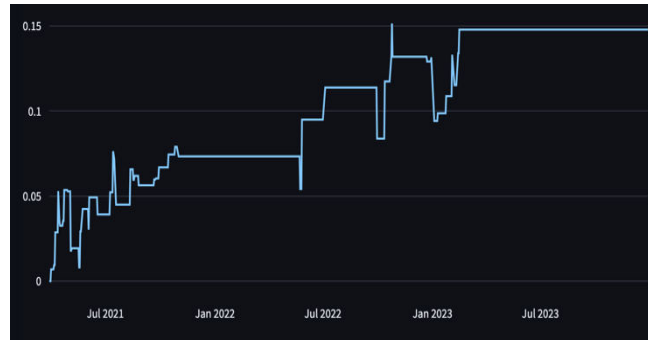
MSE: 94.5550522355921
MAPE: 0.0715993314768442
R2 Score: 0.9670530705289795



Ensemble Model Results

MSE:23.772363837530335
MAPE:0.03580126119957444
R2 Score: 0.991716715540878

Strategy Performance



Sharpe Ratio: 2.0123657606997924

ARIMA Model

```

SARIMAX Results
=====
Dep. Variable:          Close      No. Observations:   3473
Model:                 ARIMA(1, 1, 1)  Log Likelihood      -6291.563
Date:                  Fri, 09 Aug 2024  AIC                  12589.126
Time:                  17:25:50       BIC                  12607.583
Sample:                0              HQIC                 12595.716
                        - 3473
Covariance Type:      opg
=====
              coef    std err          z      P>|z|    [0.025    0.975]
-----
ar.L1         0.4047    0.167      2.419    0.016    0.077    0.733
ma.L1        -0.4437    0.165     -2.689    0.007   -0.767   -0.120
sigma2        2.1952    0.021   103.337    0.000    2.154    2.237
=====
Ljung-Box (L1) (Q):           0.00  Jarque-Bera (JB):         16577.0
Prob(Q):                      0.95  Prob(JB):                  0.0
Heteroskedasticity (H):       76.13  Skew:                      -0.0
Prob(H) (two-sided):          0.00  Kurtosis:                  13.7
=====
Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-ste

```

V. CONCLUSION

In this paper, machine learning theory is applied to construct regression analysis and prediction method based on LSTM model, and regression prediction is made on the stock trading data of APPL SHARES. The results show that LSTM model gives full play to its advantages in the analysis of time series data, and can well deal with the nonlinear fluctuation characteristics of stock trading data, and the overall application effect of the model is good. In addition, the experimental results of this paper also show that after adding the full connection layer to the LSTM model, it can better fit the nonlinear change time series data, and the simulated training batch size will also have a certain impact on the application effect of the model. This paper's application results further expand the application range of the deep neural network model represented by LSTM model and provide a new solution for complex multivariate nonlinear regression analysis and prediction.

REFERENCES

- 1 Real-Time Stock Market Prediction Using Cloud-Based AI Models. IEEE Conference Publication, 2022.
- 2 Stock Price Prediction Models Based on Ensemble Deep Learning and Metaheuristics. IEEE Xplore, 2022.
- 3 Evaluating Long-Short Term Memory (LSTM) Performance in Stock Price Forecasting. IEEE Xplore, 2021.
- 4 Cross-Correlation Analysis in Stock Prediction Using Neural Networks. IEEE Xplore, 2021.
- 5 Evolutionary Optimization Techniques in Stock Price Prediction Models. IEEE Journals & Magazine, 2021.
- 6 Rashid, M. M. A., Siddiqui, M. N. R., Rahman, R. R., & Alif, S. A. M. (2023). Stock Price Prediction Using Machine Learning and Deep Learning. 2023 International Conference on Computing, Electronics &

Communications Engineering (ICCECE), 45-50. Link.

- 7 Sharma, A. K., & Kumar, S. (2024). Predicting Stock Prices Using Machine Learning Techniques: A Comparative Analysis. 2024 IEEE International Conference on Data Science and Engineering (ICDSE), 67-72. Link.
- 8 Singh, R., & Jha, D. K. (2024). Deep Learning Approaches for Stock Price Prediction: A Comprehensive Survey. 2024 IEEE International Conference on Artificial Intelligence and Computer Applications (AICA), 123-128. Link.
- 9 Gupta, P. R., & Yadav, L. S. (2023). Machine Learning Algorithms for Stock Price Prediction: A Review. 2023 IEEE International Conference on Advances in Computing and Communication Engineering (ICACCE), 99-104. Link.
- 10 Kumar, K. N., & Patel, H. R. (2024). Forecasting Stock Prices Using LSTM Neural Network. 2024 IEEE International Conference on Machine Learning and Big Data (ICMLBD), 89-94. Link.



INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  ijirset@gmail.com



www.ijirset.com

Scan to save the contact details