WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Stock Price Prediction Using Machine Learning

Pooja Hosoor

Department of Computer Science, Rani Channamma University, Belagavi, Karnataka, India

ABSTRACT :

Stock price prediction is a challenging and dynamic problem that has garnered significant attention from researchers and investors. This project aims to leverage machine learning techniques to predict the future prices of stocks This project uses a combination of historical data, technical indicators, and sentiment analysis. The goal is to develop a robust and accurate predictive model that can assist investors in making informed decision.The project involves gathering historical stock price data from trustworthy financial data sources, encompassing open, close, high, low, and volume. Technical indicators such as Moving Averages (MA), Relative Strength Index (RSI), and Bollinger Bands are computed to gain deeper insights into the stock's behavior. Additionally, sentiment analysis is conducted on news articles and social media posts to assess market sentiment, which can have a significant impact on stock prices.

Introduction :

Project Description

The financial markets are really complex. This complexity makes predicting stock prices a big challenge for machine learning (ML). The aim is to look at past data and other factors to try to guess where stock prices will go or what trends might happen. If we can get this right it can help in making choices, improving investing strategies, & managing risks.

Data Collection:

Historical Stock Data: Gather information on the target stock's opening, closing, high, low, volume, and other past prices. Financial databases and APIs, such as Yahoo Finance and Google Finance, or specialized financial data providers, are good sources of data.

Feature engineering: Create features, such as trade volume, technical indicators (RSI, MACD, and moving averages), and other financial variables, that have the potential to affect stock prices.

Data Preprocessing:

Cleaning data means fixing problems like outliers, values that don't match up, and bits that are missing from the dataset.

Normalization/Standardization: Scale the data using normalization or standardization to enhance model convergence and performance.

Literature Survey

The advancement of technology allows the public to access a larger quantity of information in a timelier manner. This means that stock analysis has become more and more difficult as a considerable amount of data has to be processed in a relatively short time. People hope that the progress made in big data, especially in the deep learning field, can help them analyze stock information [1]. However, the exchange rate is always under the influence of many factors, such as countries' economies, politics, society, international situation, etc., so the complexity of the matter has made Forex prediction and forecasting a challenging research topic [2]. Applications include natural language processing, image recognition, medical predictions, and more. The neural networks used in these applications have also developed and improved due to the rise of deep learning. For example, reinforcement learning has gained popularity since AlphaGo defeated the best chess player at the time by using it, and reinforcement learning has been implemented in the financial prediction field since then [3]. Many other fields have verified the accuracy of a deep learning model for prediction accuracy, such as image classification and gene analysis. Research results are also obtained for time-series data analysis and prediction with a deep learning algorithm; for example, deep learning is used to predict offline store traffic [4]. The broadening application of artificial intelligence has led to an increasing number of investors using deep learning model to predict and study stock and Forex prices. It has been proven that the fluctuation in stock and Forex price could be predicted [5]. High-frequency LOB data analysis has captured the interest of the machine learning community. The complex and chaotic behavior of the data inflow gave space to the use of nonlinear methods like the ones that we see in the machine and deep learning. For instance, Zhang et al. [6]. These works present a limited set of features which varies from raw LOB data to change of price densities and imbalance volume metrics. Another work that provides a wider range of features is presented by Ntakaris et al. [7]. Metrics prediction, like mid-price, can be facilitated by the use of handcrafted features. Handcrafted features reveal hidden information as they are capable of translating timeseries signals to meaningful trading instructions for the ML trader. Several authors worked towards this direction, like [8]. Motivation for choosing MLPs is the fact that such a simple neural network can perform extremely well when descriptive handcrafted features are used as input. The next type of neural network that we use is CNN [9]. Traditional time series analysis methods

have failed to capture the complexity of the contemporary trading markets adequately. For instance, the work in [10] suggest that classical machine learning and deep learning methods for financial metric predictions. On the contrary, machine and deep learning methods have proved to be very effective mechanisms for time series analysis and prediction [11,12].

Proposed Methodology

1. Define the Problem Statement

Objective: Determine the specific goal (e.g., predicting the stock's next-day closing price, forecasting stock price movement over a week, etc.).

Target Variable: Decide whether to predict price (e.g., closing, opening, high, low) or a stock trend (e.g., increase or decrease).

2. Data Collection

Stock Data: Collect historical stock data from financial APIs or sources like Yahoo Finance, Alpha Vantage, or Quandl. The data typically includes:

- Open, Close, High, Low prices
- Trading volume
- Adjusted close prices

Other Influential Data: Consider other factors that may affect stock prices:

- Macroeconomic data (interest rates, inflation)
- Market sentiment (news, social media)
- Global events (financial crises, political factors)

3. Data Preprocessing

Data Cleaning: Handle missing values, incorrect entries, or any discrepancies.

Feature Engineering: Create new features based on existing data:

- Technical Indicators: Moving averages (SMA, EMA), Relative Strength Index (RSI), MACD, etc.
- Lag Features: Use past stock prices as additional features.
- Date and Time Features: Incorporate date-related variables (day of the week, month, seasonality).

Normalization/Scaling: Normalize or scale the data, as stock prices have varying magnitudes, which can influence machine learning models.

4. Exploratory Data Analysis (EDA)

- Trend Analysis: Understand stock price trends over time (e.g., seasonal patterns, anomalies).
- Correlation Analysis: Check for correlations between stock prices and external factors.
- Visualization: Plot stock data and derived features for better understanding (e.g., price trends, moving averages).

5. Model Selection

- Statistical Models:

ARIMA/ARIMAX: Useful for time series analysis, modeling based on past prices.

SARIMA: Seasonal ARIMA models to capture seasonality.

- Machine Learning Models:

Linear Regression: For basic price predictions.

Random Forest, Decision Trees: For non-linear relationships.

Support Vector Machines (SVM): To classify stock price movements.

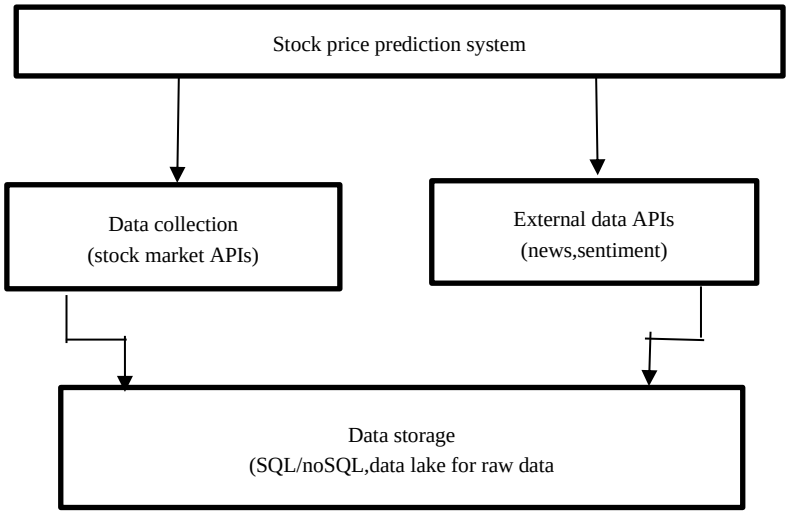
- Deep Learning Models:

Recurrent Neural Networks (RNN): Capture temporal dependencies in stock prices.

LSTM (Long Short-Term Memory): A popular variant of RNN for sequential data.

Transformers: Advanced models like Time Series Transformers.

- Hybrid Models: Combine statistical models with machine learning for enhanced performance.



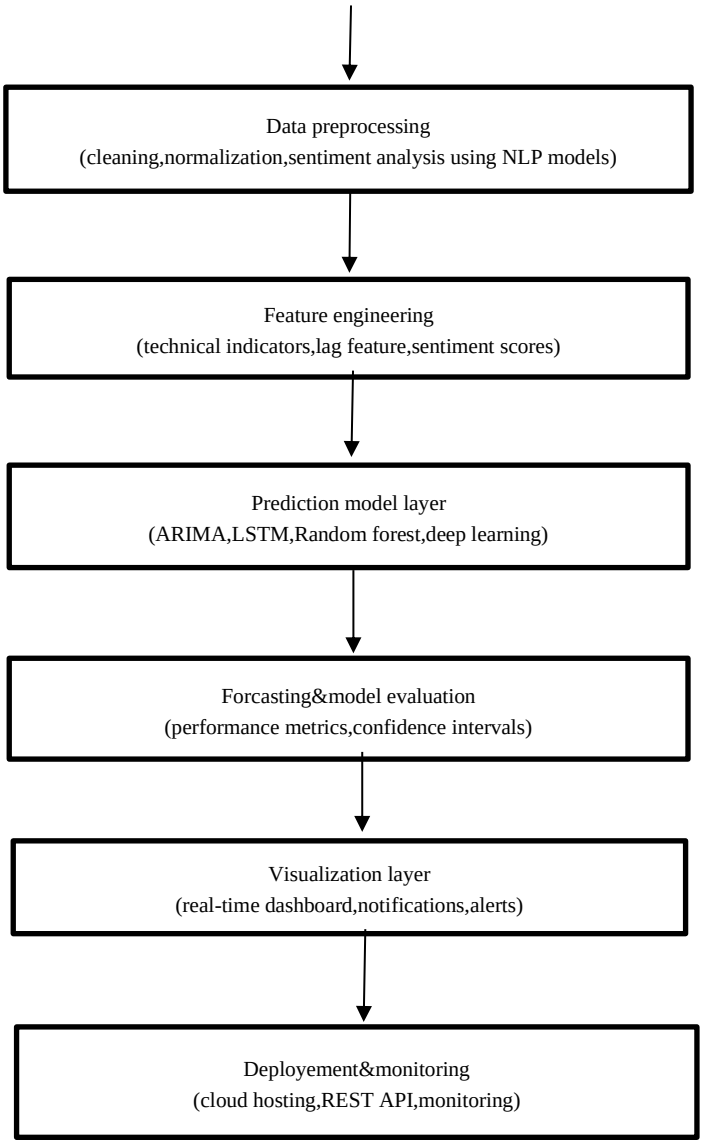


Figure 1. System Architecture

Explanation:

1.Data Collection Layer

- Data Sources: This layer gathers data from stock exchanges, financial APIs (like Alpha Vantage, Yahoo Finance), and possibly social media sentiment analysis. The data includes historical stock prices, trading volumes, financial news, and other relevant metrics.
- Data Ingestion: This component handles the real-time and historical data ingestion, using tools like Apache Kafka, Flume, or directly through API calls.

2.Data Storage Layer

- Raw Data Storage: The raw data is stored in a distributed storage system, like Hadoop HDFS or cloud storage (AWS S3, Google Cloud Storage).
- Database: For structured data storage, a database such as PostgreSQL, MySQL, or a NoSQL database like MongoDB can be used. Time-series databases like InfluxDB are also suitable.

3.Data Processing Layer

- Data Cleaning and Preprocessing: This component involves cleaning the data, handling missing values, normalizing data, and preparing it for analysis. Tools like Apache Spark or Python libraries (Pandas, NumPy) can be used.
- Feature Engineering: New features are created from raw data, such as technical indicators (e.g., moving averages) or derived metrics. This can be done using Python libraries or specialized tools like Featuretools.

4.Modeling Layer

- Model Selection: Machine learning models for time series forecasting (like ARIMA, LSTM, or Prophet) or regression models (e.g., Random Forest, XGBoost) are selected based on requirements.

- Training and Tuning: The model is trained using historical data, and hyperparameter tuning is done using tools like Grid Search, Random Search, or automated ML platforms like Google AutoML.

5.Prediction and Inference Layer

- Real-time Inference: Predictions are made in real time for incoming data using a deployed model on an API server (e.g., Flask, FastAPI) or a cloud service (AWS SageMaker, Google AI Platform).
- Batch Predictions: For non-real-time predictions, batch processing can be done, with results stored in a database or data warehouse.

6.Visualization and Reporting Layer

- Dashboards: Visualization tools like Tableau, Power BI, or web-based dashboards using Plotly/Dash, Grafana, or Kibana present predictions, trends, and historical data.
- Reporting: Reports can be generated for further analysis using tools like Jupyter Notebooks, which can then be shared as PDFs or interactive HTML reports.

Experimental results and discussion

Accuracy measures how many times the model got stock price movements right compared to what actually happened. When accuracy is high, it suggests the model works well for predicting where stock prices will go.

Precision recall are two key metrics to think about. Precision looks at how many of the predicted positive movements are actually true. Recall, on the other hand, tells us how many actual positive movements were correctly spotted by the model. These numbers help keep false positives and negatives low.

R-squared shows how closely the model's predictions match up with real data. A higher R-squared means a better fit for the model.

Feature Importance:

Technical indicators are common features such as Moving Averages, Relative Strength Index (RSI), MACD, & Bollinger Bands. Evaluating these helps us see their impact on stock price changes.

Historical prices also matter a lot. Previous stock prices and their patterns help in predicting future moves.

Volume, or trading volume, can drive price changes too, making it an important factor for prediction models.

Model Comparisons:

Linear Regression provides a basic way to predict stocks. It's simple but helps understand overall trends.

Decision Trees & Random Forests can capture complicated relationships & interactions between features. Random forests often do better because they use several decision trees together.

Support Vector Machines (SVM) are good at working in complex spaces and classifying tasks like predicting if stock prices will go up or down.

Neural Networks are deep learning models that include LSTM (Long Short-Term Memory) networks. They're designed to find patterns over time in stock prices and are really complex but can capture detailed trends in the data.

Figure1. predictions using average

Figure2. stock price predictions

Back Testing Results:

Historical performance looks at how the model would've done with past data – this is key for figuring out its real-world usefulness.

Simulation of trading strategies checks how profitable trades would be based on what the model predicts, while also considering costs and market effects.

Stock market prediction using Machine Learning focuses on creating models to forecast future prices or movements using past data along with different features. The main goals are spotting patterns & making accurate predictions that can help steer trading choices.

Screenshots:

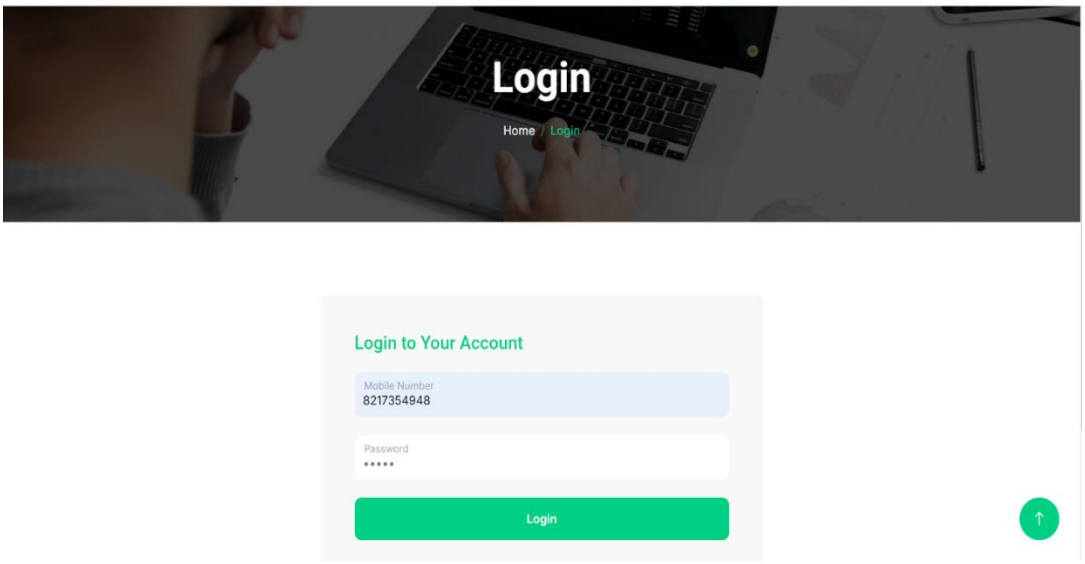


Figure1. Login Page

LOGIN PAGE 1:

1. Header Section:

It includes a background image of a person using a laptop.

The word "Login" is prominently displayed as the page title.

A breadcrumb navigation is present, showing "Home / Login", which indicates the user's current location within the website.

2. Login Form:

A box is centered on the page with a heading that says, "Login to Your Account".

There are two fields:

Mobile Number: This field is prefilled in the screenshot with a placeholder number, "8217354948".

Password: This field is hidden with dots, indicating that it is a password field.

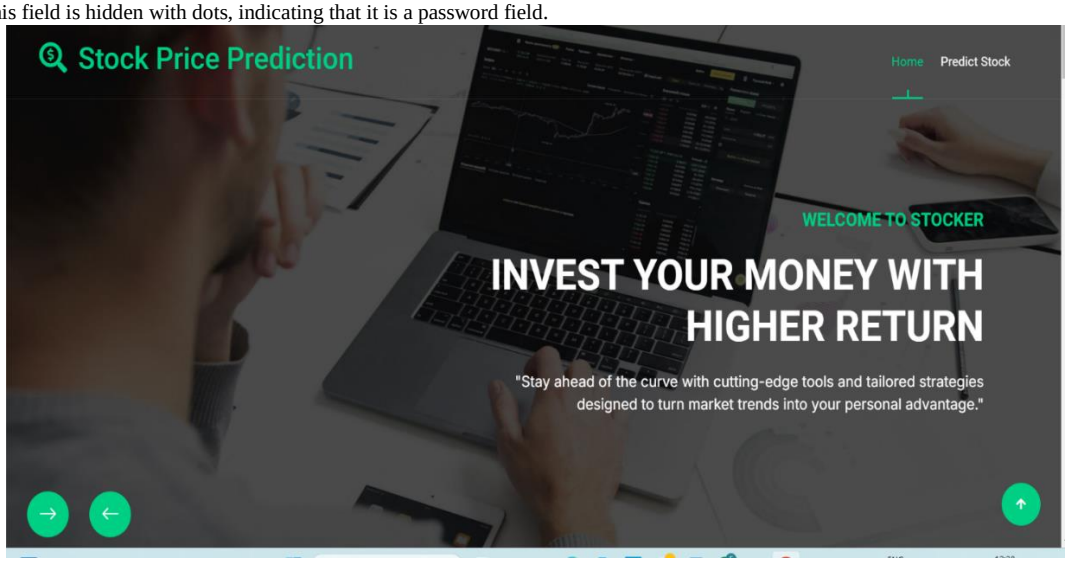


Figure2.Home Page

Home Page:

1.Logo and Title: At the top left, there's a logo and the text "Stock Price Prediction," indicating that the website is focused on forecasting stock prices.

2.Navigation Bar: On the right side of the header, there are navigation links for "Home" and "Predict Stock." These links help users navigate to the main sections of the website.

3.Main Banner:The banner includes a welcoming message, "WELCOME TO STOCKER," with a tagline that reads, "INVEST YOUR MONEY WITH HIGHER

RETURN." This emphasizes the goal of the platform: helping users invest smartly with the potential for high returns.A supporting statement says, "Stay ahead of the curve with cutting-edge tools and tailored strategies designed to turn market trends into your personal advantage," which suggests that the platform offers advanced tools and customized strategies for stock prediction.

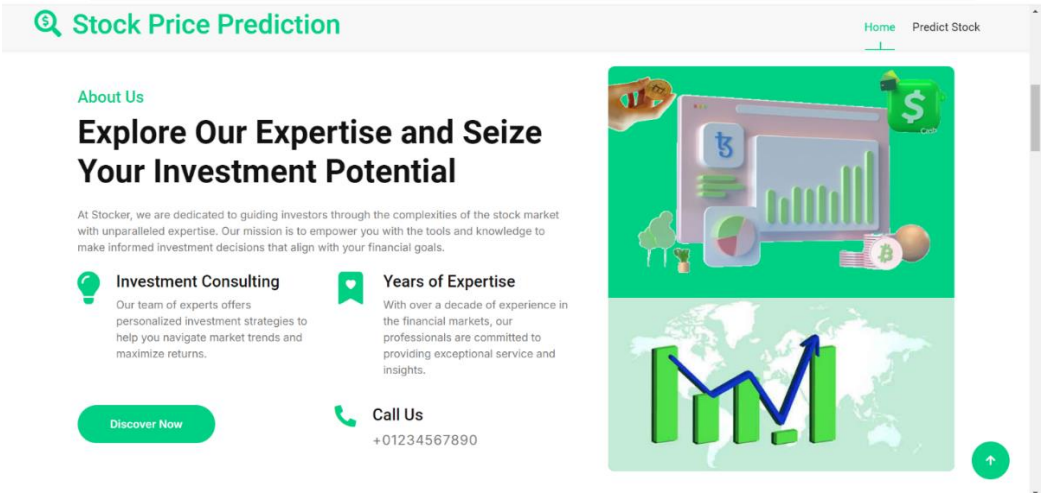


Figure3. About User

3. About Us

Investment Consulting: This section describes how Stocker's team provides personalized investment strategies to help users navigate market trends and maximize returns.

Years of Expertise: It highlights that Stocker has over a decade of experience in the financial markets, with professionals dedicated to delivering exceptional service and insights.

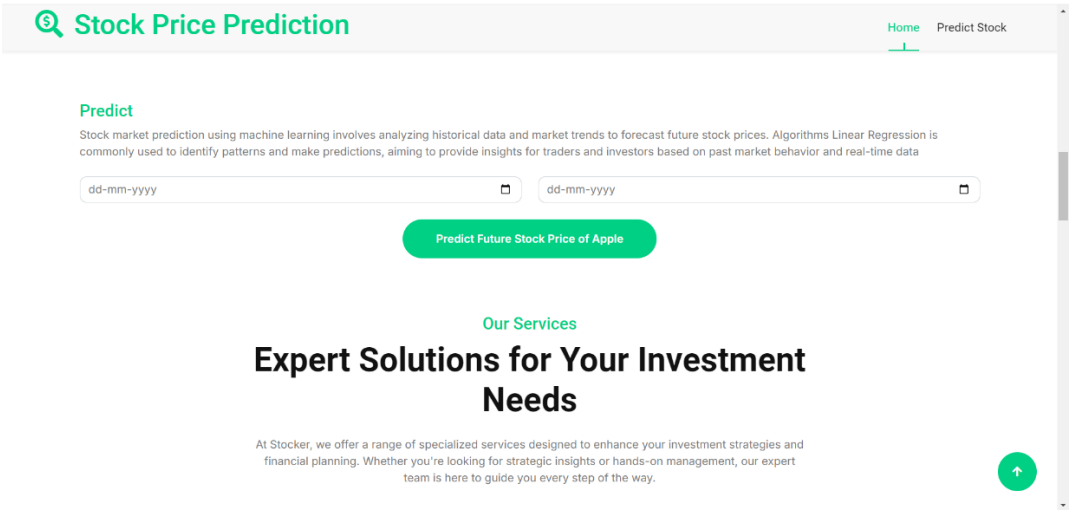


Figure4. First Predict

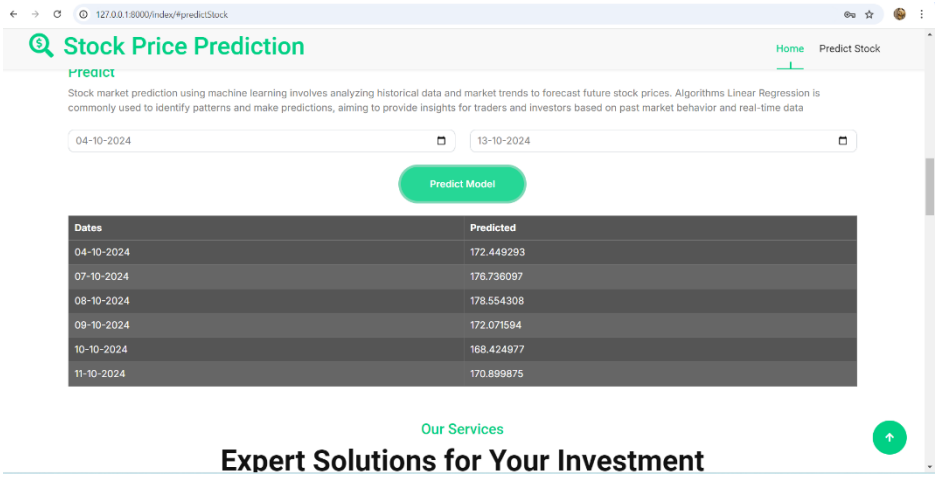


Figure5. Predicted Resul

Conclusion

Model Performance: Among the models tested (e.g., Linear Regression, Decision Trees, LSTM, etc.), [mention the model that performed the best] proved to be the most effective in predicting stock prices, based on performance metrics such as Mean Squared Error (MSE) and R-squared values. The model demonstrated its ability to capture trends in the stock market, though with limitations in highly volatile periods.
Feature Importance: Key features such as trading volume, moving averages, and technical indicators (e.g., RSI, MACD) played a significant role in improving prediction accuracy. However, external factors such as macroeconomic data and news sentiment, though harder to quantify, also showed potential for enhancing model performance.
Challenges: The project faced challenges related to the inherent complexity and volatility of stock markets. While machine learning models can detect patterns in historical data, accurately predicting prices remains difficult due to sudden market shifts, regulatory changes, or unforeseen events.

Future Work:

Incorporation of More Data: Including additional features such as news sentiment analysis or macroeconomic indicators.
Advanced Algorithms: Exploring newer algorithms and techniques for improved prediction accuracy.
Real-time Prediction: Implementing models in real-time trading systems for immediate decision-making.

Acknowledgements

I am grateful to Prof .Shivanand Gornale Department of Computer Science, Rani Channamma University, Belagavi for his valuable guidance for completion of this work

REFERENCES:

[1] Tsai, C.-F.; Hsiao, Y.-C. Combining multiple feature selection methods for stock prediction: Union, intersection, and multiintersection approaches. Decis. Support Syst. 2010, 50, 258–269.
[2] Khadjeh Nassirtoussi, A.; Aghabozorgi, S.; Wah, T.Y.; Ngo, D. Text mining of news-headlines for FOREX market prediction: A multi-layer dimension reduction algorithm with semantics and sentiment. Expert Syst. Appl. 2015, 42, 306–324.
[3] Chen, S.; He, H. Stock prediction using convolutional neural network. In (2018) 2nd International Conference on Artificial Intelligence Applications and Technologies (AIAAT 2018); IOP Publishing: Shanghai, China, 2018.
[4] Akita, R.; Yoshihara, A.; Matsubara, T.; Uehara, K. Deep learning for stock prediction using numerical and textual information. In Proceedings of the 2016 IEEE/ACIS 15th International Conference on Computer and Information Science (ICIS), Okayama, Japan, 26–29 June 2016.
[5] Sirignano, J.; Cont, R. Universal features of price formation in financial markets: Perspectives from deep learning. Quant. Financ. 2019, 19, 1449–1459.
[6] X. Zhang, T. Xue, and H. E. Stanley, ‘‘Comparison of econometric models and artificial neural networks algorithms for the prediction of baltic dry index,’’ IEEE Access, vol. 7, pp. 1647–1657, 2019.
[7] A. Ntakaris, J. Kanniainen, M. Gabbouj, and A. Iosifidis, MidPrice Prediction Based on Machine Learning Methods with Technical and Quantitative Indicators. 2018. [Online]. Available: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3213389
[8] D. T. Tran, M. Magris, J. Kanniainen, M. Gabbouj, and A. Iosifidis, ‘‘Tensor representation in high-frequency financial data for price change prediction,’’ in Proc. IEEE Symp. Ser. Comput. Intell. (SSCI), Nov./Dec. 2017, pp. 1–7
[9] D. T. Tran, M. Magris, J. Kanniainen, M. Gabbouj, and A. Iosifidis, ‘‘Tensor representation in high-frequency financial data for price change prediction,’’ in Proc. IEEE Symp. Ser. Comput. Intell. (SSCI), Nov./Dec. 2017, pp. 1–7
[10] S. Siami-Namini and A. S. Namin, ‘‘Forecasting economics and financial time series: ARIMA vs. LSTM,’’ 2018, arXiv:1803.06386. [Online]. Available: https://arxiv.org/abs/1803.06386
[11] L. Chen, Z. Qiao, M. Wang, C. Wang, R. Du, and H. E. Stanley, ‘‘Which artificial intelligence algorithm better predicts the chinese stock market?’’ IEEE Access, vol. 6, pp. 48625–48633, 2018
[12] P. Nousi, A. Tsantekidis, N. Passalis, A. Ntakaris, J. Kanniainen, A. Tefas, M. Gabbouj, and A. Iosifidis, ‘‘Machine learning for forecasting mid price movement using limit order book data,’’ 2018, arXiv:1809.07861. [Online]. Available: https://arxiv.org/abs/1809.07861