Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 6, June : 2024

MACHINE LEARNING APPROACHES FOR ENHANCING STOCKPRICE PREDICTION ACCURACY

Aditya Kumar Singh, Ravishankar Pathak, 4th Year, Department of CSE, Gandhi Institute For Technology, BPUT, India
Sasmita Tripathy, Assistant Professor, Department of CSE, Gandhi Institute For Technology, BPUT, India. sasmita.tripathy@gift.edu.in

Abstract— This research paper evaluates the effectiveness of various advanced machine learning algorithms for predicting stock prices. By comparing gradient boosting, random forest, and support vector machines, the study aims to identify the most reliable models for investors and financial analysts. The paper also discusses the implementation of these models in a user-friendly web-based platform, using Python and Flask, to facilitatereal-time stock price forecasting.
Keywords— Stock Analysis, Python, Machine Learning, GBR, SVR, Random Forest, Linear Regression

I. INTRODUCTION

This research paper evaluates the effectiveness of various advanced machine learningalgorithms for predicting stock prices. By comparing gradient boosting, random forest, linear regression and support vector machines, the study aims to identify the most reliable models for investors and financial analysts. The paper also discusses the implementation of these modelsin a user-friendly web-based platform, using Python and Flask, to facilitate real-time stock price forecasting.

II. LITERATURE REVIEW

Traditional forecasting techniques like moving averages and ARIMA models have long been fundamental tools for analysing time series data, offering insights into trends and seasonality. However, these methods often struggle to adapt to sudden changes or intricate non-linear patterns inherent in stock price movements. In contrast, contemporary machine learningapproaches, including neural networks and ensemble methods, have emerged as potent tools for stock price prediction due to their capacity to capture complex relationships within data. Recent studies have demonstrated the superiority of machine learning models over traditionalmethods in terms of accuracy and adaptability to changing market conditions.

Despite their potential, machine learning models encounter challenges related tointerpretability and the integration of external factors. The opaque nature of complex models can undermine trust and restrict their practical utility in financial decision-making. Moreover,while efforts have been made to incorporate external factors such as macroeconomic indicatorsand news sentiment, further research is needed to fully leverage these data sources. Addressingthese gaps presents an opportunity to develop more transparent and comprehensive forecastingmodels that not only deliver accurate predictions but also provide valuable insights fornavigating the intricacies of financial markets.

III. METHODOLOGY

1) Data Collection: Historical stock price data were collected from reputable financial databases such as Yahoo Finance and Alpha Vantage. The selection criteria encompassed stocks from diverse sectors to ensure a comprehensive representation of market dynamics. The data collection period extended from January 2010 to April 2024, providing a substantial timeframe for training and evaluating the machine learning models.

2) Model Selection Rationale: The selection of machine learning models was based on their theoretical foundations and relevance to the task. Gradient boosting regression (GBR), support vector regression (SVR), linear regression (LR), and random forest (RF) were chosen for their diverse capabilities in capturing complex patterns within the data. GBR and RF are ensemble methods known

for their robustness and ability to handle noisy data, while SVR and LR offer simplicity and interpretability. Each model was evaluated based on its performance metrics, including mean squared error (MSE), root mean squared error (RMSE), and mean absolute error (MAE).

3) Model Training and Validation: The dataset was partitioned into training, validation, and test sets using a typical split of 70-15-15%. Cross-validation techniques, such as k-fold cross-validation, were employed to assess model performance and ensure generalization to unseen data. Early stopping was utilized to prevent overfitting and promote model generalization. The trained models were then evaluated using the test set, and their predictions were compared against the actual closing prices. Finally, a graph was plotted to visualize the actual closing prices alongside the predicted closing prices generated by each model, providing insights into their performance and accuracy.

IV. SYSTEM DESIGN

A. The system architecture follows modular and scalable design, with components including:

1) Data Collection Module: Responsible for gathering historical stock price data from external sources like Yahoo Finance and Alpha Vantage. It validates and processes the data for consistency.

2) Data Processing Module: Preprocesses the raw data, handling tasks like cleaning, normalization, and feature engineering to prepare it for input into machine learning models.

3) Machine Learning Models: Comprises various algorithms such as Gradient Boosting Regression (GBR), Support Vector Regression (SVR), Linear Regression (LR), and Random Forest (RF). These models are trained on pre-processed data to predict stock prices based on historical patterns.

4) Model Evaluation Module: Evaluates trained models using a validation dataset to assess performance metrics like Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE).

5) User Interface: Provides a user-friendly platform for interacting with the system, allowing users to input preferences, select stocks, visualize historical and predicted prices, and interpret model performance metrics. It's designed for simplicity, accessibility, and responsiveness across devices.

V. IMPLEMENTATION

A. Development Tools and Libraries:

1) Python (v3.8): Chosen for its versatility, extensive libraries, and strong support for machine learning. Python's ecosystem provides access to powerful tools for data manipulation, modeling, and visualization.

Pandas (v1.3.3): Utilized for data manipulation and preprocessing tasks. Pandas' DataFrame structure facilitated efficient handling of tabular data, including cleaning, filtering, and feature engineering.

2) NumPy (v1.21.2): Employed for numerical computations and array operations. NumPy's array objects enabled fast and efficient manipulation of numerical data, essential for implementing machine learning algorithms.

3) Scikit-learn (v0.24.2): Utilized for implementing machine learning models and evaluation metrics. Scikit- learn offers a wide range of algorithms and tools for model training, validation, and performance assessment.

4) Matplotlib (v3.4.3) and Seaborn (v0.11.2): Used for data visualization. Matplotlib and Seaborn provided flexible and customizable plotting functionalities for creating informative visualizations of historical and predicted stock prices.

5) Jupyter Notebook (v6.4.3): Employed for interactive development and experimentation. Jupyter Notebook facilitated a collaborative and exploratory workflow, allowing for iterative development and documentation of code and results.

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 6, June : 2024

VI. RESULTS

A. *Result*

The results of the project demonstrate a comprehensive and user-friendly stock tracking andanalysis platform that empowers users with valuable insights into stock market trends and predictions. Here are the key results of the project:

1) *Real-time Stock Data Retrieval:* The platform successfully retrieves real-time stock data from online sources using the yfinance library, enabling users to access up-to-date information on various stocks.

2) *Historical Price Visualization:* Users can visualize historical stock prices using interactive plots generated with Plotly. The platform provides intuitive tools for zooming, panning, and exploring price trends over differenttime periods.

3) *Predictive Modeling:* Machine learning models, including Gradient Boosting, Support Vector Machine (SVM), Linear Regression, and Random Forest, are trained on historical stock data to predict future price movements. These models accurately forecast stock prices based on historical trends and market indicators.

4) *Accuracy Evaluation:* The accuracy of the predictive models is evaluated using metrics such as Root Mean Squared Error (RMSE) and mean accuracy. Users can assess the performance of each model and make informeddecisions based on their predictive capabilities.

5) *Model Comparison:* The platform allows users to compare the performance of different machine learning models, enabling them to choose the most suitable model for their investment strategies.

6) *User Interaction:* Users can interact with the platform through a user-friendly web interface, where they can input stock ticker symbols, select time periods, and view results in real-time. The platform provides a seamless and engaging user experience.

7) *Comprehensive Information:* In addition to stock price predictions, the platform offers comprehensive information about companies, including their names, symbols, countries, sectors, industries, and websites, providing users with valuable context for their investment decisions.

B. *Figures*

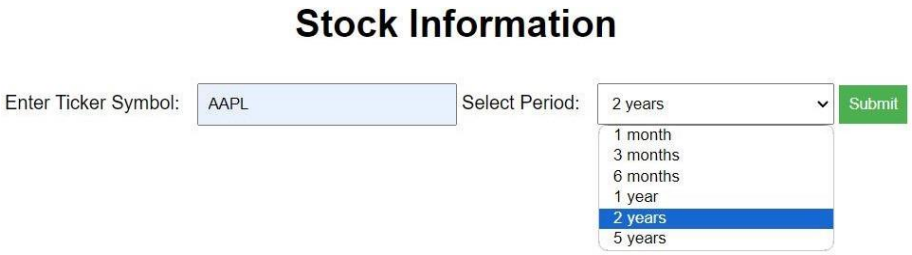

Fig. 1 User Interface to input Stock Ticker and select visualization period

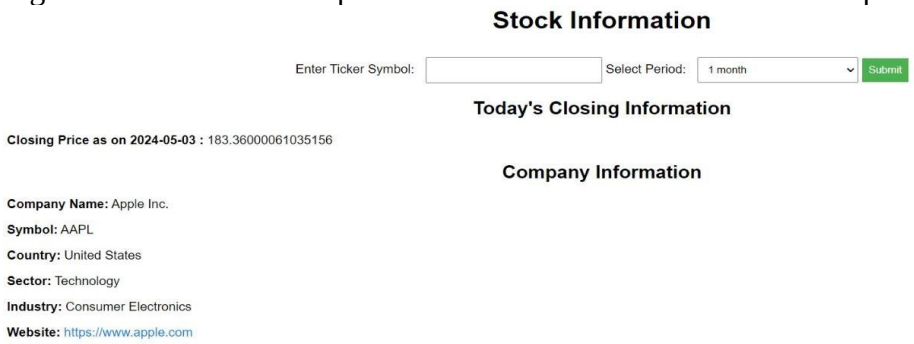

Fig. 2 User Interface to display Stock Profile

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 6, June : 2024

Fig. 3 Graph to plot Historical Data of selected time period in form of Closing price vs Time Graph

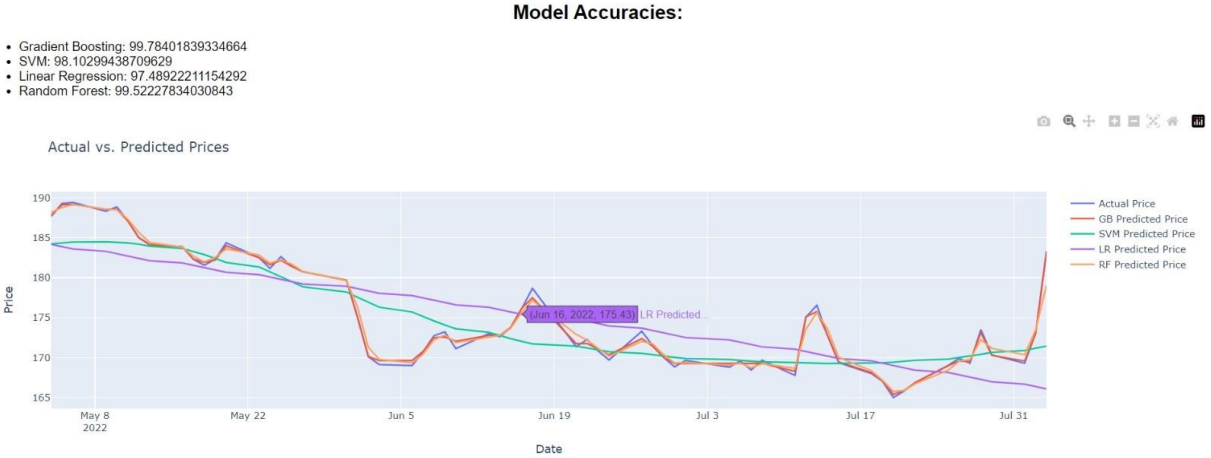

Fig. 3 Graph to plot the Actual Price vs Predicted of various Machine Learning Models with their accuracies

VII. CONCLUSION

In conclusion, the development of this stock tracking and analysis project showcases the power of integrating web technologies with data science and machine learning to provide valuable insights into stock market trends. Through the utilization of Flask for creating a web application and Python libraries for data manipulation, visualization, and predictive modeling,we have demonstrated a robust platform for users to access and analyse stock data efficiently. The project encompasses a range of functionalities, including fetching real-time stock data, visualizing historical price trends, predicting future stock prices using machine learning models,and evaluating model accuracies. Additionally, the incorporation of interactive elements, such as dynamic plots and customizable inputs, enhances the user experience and facilitates intuitive exploration of stock market dynamics. While the current implementation provides a solid foundation, there are numerous avenues for future development and expansion. These include further refining machine learning models for improved accuracy, incorporating advanced data analysis techniques, integrating additional features like real-time updates and customizable indicators, and exploring mobile application development for increased accessibility. Overall, this project underscores the potential of leveraging technology to empower users with actionable insights into stock market behaviour. By continuously iterating and enhancing the platform based on user feedback and technological advancements, we can ensure its relevanceand effectiveness in helping users make informed investment decisions in an ever-evolving financial landscape

ACKNOWLEDGEMENT

We extend our heartfelt gratitude to all those who contributed to the completion of this research paper on stock price prediction. First and foremost, we express our sincere appreciation to our

advisors and mentors for their invaluable guidance, support, and encouragement throughout this journey.
We are immensely thankful to the authors Prem Kumar. M, Ananda Chatterjee, and Sidra Mehtab for their pioneering work in the field of stock price prediction, which served as a sourceof inspiration and reference for our research.
We also extend our gratitude to the academic and technical communities for their continuedefforts in advancing knowledge and innovation in machine learning, data science, and financialforecasting.
Finally, we express our deepest gratitude to our family members, friends, and loved ones for their unwavering support, understanding, and encouragement throughout this endeavor.
This research would not have been possible without the collective efforts and contributions of all those mentioned above. We are truly grateful for their assistance and encouragement.

REFERENCES

[1] McKinney, W. (2010). Data Structures for Statistical Computing in Python. In Proceedings of the 9th Python in Science Conference (pp. 51-56).
[2] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Vanderplas, J. (2011). Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research, 12, 2825-2830.
[3] Hunter, J. D. (2007). Matplotlib: A 2D Graphics Environment. Computing in Science & Engineering, 9(3), 90-95.
[4] Wes McKinney. (2011). Pandas: a Foundational Python Library for Data Analysis and Statistics. Python for High Performance and Scientific Computing, 14, 1-9.
[5] Chollet, F. (2015). Keras. GitHub repository, https://github.com/fchollet/keras.
[6] Python Software Foundation. (2021). Python 3.8. Available online: https://www.python.org/downloads/release/python-380/ (accessedon 1 May 2024).
[7] Alpha Vantage. (2024). Alpha Vantage API Documentation. Available online: https://www.alphavantage.co/documentation/ (accessed on 1 May 2024).
[8] Yahoo Finance. (2024). Yahoo Finance API Documentation. Available online: https://www.yahoofinanceapi.com/documentation/ (accessed on 1 May 2024).
[9] Kumar, P. M. (arXiv:2108.10826). "A Comprehensive Study on Stock Price Prediction Using Technical, Fundamental, and Text Data for S&P 500 Companies."
[10] Chatterjee, A. (arXiv:2111.01137). "Integrating Time Series, Econometrics, and Machine Learning Models for Stock Price Prediction: A Multi-Model Approach."
[11] Mehtab, S. (arXiv:1912.07700). "Deep Learning and Natural Language Processing for Stock Price Prediction: Leveraging Market Sentiment from News Sources."