Stock Market Prediction Using Machine Learning

[1]Aditiya Munjal, [2]Ganesh Shinde, [3]Fazal Mehmood, [4]Dr .Vaishali Anaspure, [5]Atharva Chavan

Savitribai Phule Pune University, G H Raisoni College of Engineering and Management Pune, India

[1]aadi.munjal07@gmail.com, [2]ganeshshinde7936@gmail.com, [3]azizifazal883@gmail.com, [4]vaishali.anaspure@raisoni.net, [5]chavanatharava126@gmail.com

Abstract - **The stock market is a curve that is erratic. The stock market is a difficult and unstable place to make predictions. Predicting the stability of future market stocks is the primary goal of the topic's persuasion. Numerous scholars have conducted investigations into the trajectory of future market development. Since data in stocks fluctuates, data is a crucial source of efficiency. Influence the prediction's accuracy on the same possibilities. Machine learning has entered the scene for the deployment and prediction of training sets and data models in the recent trend of stock market prediction technologies. Machine learning makes use of various prediction models and algorithms to anticipate and automate necessary tasks. The study focuses on using LSTM and regression to make predictions.**

Keywords - Stock Market, Prediction, Machine Learning.

I. INTRODUCTION

The stock market has captured the interest of investors all around the world. It is a complex and dynamic financial structure. It acts as a marketplace where people and organizations can buy stock in firms, symbolizing ownership and possibly earning substantial returns. Still, it is extremely difficult to forecast the stock market's erratic swings.

A complex undertaking with the goal of predicting future stock prices, stock market prediction is a popular choice among investors. In order to identify trends and make wise investment decisions, this method entails evaluating a variety of variables, such as market sentiment, corporate performance, and economic data. Although there is no denying the appeal of making significant profits by making precise stock market predictions, these undertakings are fundamentally uncertain due to the market's intrinsic complexity. Accurate forecasting is difficult due to the multitude of internal and external factors influencing stock values. The

Goal of stock market prediction propels research and innovation despite its inherent difficulties. In an attempt to decipher the mysteries surrounding market movements, researchers and analysts utilize a wide range of methodologies, from sophisticated machine learning algorithms to conventional fundamental analysis. Even with continuous improvements, making consistently accurate stock market forecasts is still a difficult task. An enormous challenge is presented by the market's dynamic character and the abundance of influencing elements. Despite this, there is still a lot of research being done on stock market prediction because it has the potential to yield large profits and because people will always be curious to learn more about the financial world.

II. RELATED WORK

One notable contribution of their review is the identification of emerging trends and promising avenues for future research in stock market prediction. By synthesizing findings from a range of studies, Gandhmal and Kumar offer a comprehensive perspective on the strengths and limitations of different techniques, aiding researchers and practitioners in making informed choices when selecting predictive models for financial markets [1] The systematic analysis of stock market prediction techniques, providing insights into computational approaches for financial forecasting. Their work emphasizes the importance of methodical evaluation and identifies emerging trends.[2] The knowledge base for cognition-driven sentiment analysis. The work introduces a common and common sense knowledge base, potentially enriching the understanding of market sentiment in predictive models.[3] Evaluated multiple classifiers for stock price direction prediction in their work published in Expert Systems with Applications . The study provides insights into the performance of different classifiers, offering valuable considerations for selecting appropriate models in stock market prediction[4]. Presented a deep learning-based

 DOI : 10.35291/2454-9150.2023.0239

approach for fresh produce market price prediction . Their work showcases the applicability of deep learning in predicting market prices for agricultural products, suggesting potential insights for similar applications in financial markets [5]. "Deep Learning for Financial Prediction" , presenting insights on Circuits and Systems in Digital Enterprise Technology. Their work contributes to the understanding of deep learning applications specifically in the financial domain [6]. A "Convolutional LSTM Network" for precipitation nowcasting, introducing machine learning approaches in temporal prediction [7].

III. SYSTEM ARCHITECTURE

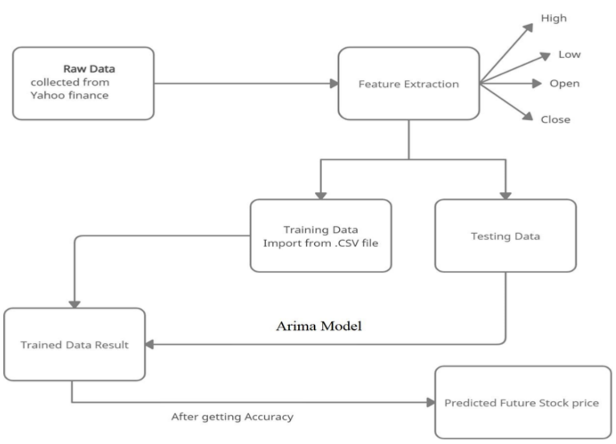

Fig. 3.1- Module 1: User Authentication and Management

Module 1: User Authentication and Management

Submodule1.1: User Sign Up Page using HTML,CSS & Javascript

1 .Create a User Sign Up Page:

- Develop a web page using HTML,CSS and Javascript for user registration.
- Include input fields for first name,last name, email, password, and confirm password.
- Implement form validation to ensure proper input.

2. Handle User Registration:

- On form submission, validate the input data and perform necessary checks (e.g., unique email).
- If the input data is valid, send a request to the backend server to create a new user account.

3 . Display Success/Failure Messages:

- Show a success message upon successful user registration.
- Display appropriate error messages if any validation or server-side errors occur.

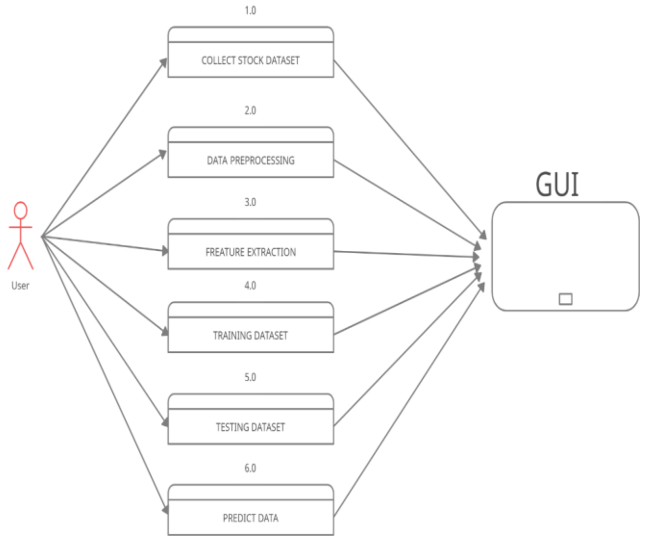

Fig. 3.2- Module 2: Backend Connectivity with Django

Module 2 : Backend Connectivity with Django

Submodule 2.1 : Django

1. Setup Django with Redis:

- Install Redis on your server or local machine.
- Create a new Django project or use an existing one

2. Connect to Redis from Python::

- Using Python connect to Redis using the redis-py library. Install it using pip.

Submodule 2.2: Dashboard For Stock Prediction:

- Design an interactive dashboard for users to make stock predictions.
- Display real-time stock market data and relevant statistics.
- Implement user-friendly data visualization tools (e.g., charts, graphs).
- Integrate a machine learning model for stock price predictions
- Allow users to input stock symbols or select from a list.

Submodule 2.3 Machine Learning Model Integration:

- Train a machine learning model (e.g., LSTM, Random Forest) using historical stock data.
- Integrate the trained model into the dashboard for making predictions.
- Provide users with the ability to customize prediction parameters (e.g., time horizon)

 DOI : 10.35291/2454-9150.2023.0239

ISSN : 2454-9150 Vol-09, Issue-09, Dec 2023

Submodule 2.4: Prediction Results:

- Display predicted stock prices and trends in a visually appealing manner.
- Include confidence scores or prediction intervals to indicate model uncertainty
- .Enable users to compare predictions with actual stock prices.

Submodule 2.5: Future Enhancements:

- Discuss potential future enhancements, such as mobile app development, additional prediction features, or integration with financial news sources.

Module 3:Machine Learning Models

1) ARIMA(Auto Regressive Integrated Moving Average)

ARIMA Model Overview:

- ARIMA is a widely used time series forecasting model that combines AR (AutoRegressive), I (Integrated), and MA (Moving Average) components.Model Parameters (p, d, q):
- ARIMA is defined by three parameters: p, d, and q.
    - p: Lag order for autoregressive terms.
    - d: Degree of differencing to achieve stationarity.
    - q: Lag order for moving average terms.

Stationarity:

- ARIMA requires data to be stationary, achieved through differencing.

Model Fitting and Validation:

- Fit the model using historical data.

Forecasting:

- Use the trained ARIMA model to predict future values.

Limitations:

- ARIMA may struggle with complex data patterns and non-linearity.

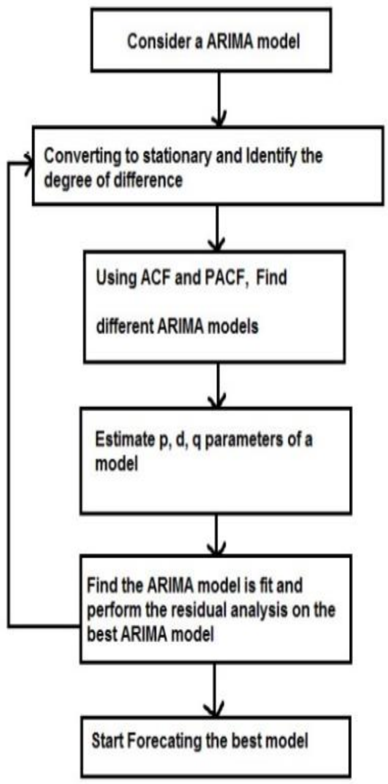

Fig. 3.3- Module 3: Machine Learning Model

IV. IMPLEMENTATION

These are the Machine Learning Algorithms implemented during the building of the project.

4.1. LINEAR REGRESSION

Linear regression is a fundamental machine learning algorithm employed to examine the connection between two variables. It can be categorized into two types: positive and negative. Positive linear regression manifests when both variables simultaneously increase or decrease. Conversely, negative linear regression occurs when one variable increases while the other decreases. Single linear regression (SLR) represents the most basic form of linear regression. Multiple linear regression (MLR) serves as an extension of SLR, capable of handling more than one independent variable. MLR additionally takes into account residual errors, which represent the discrepancies between the actual values of the dependent variable and the predicted values.

4.2. POLYNOMIAL REGRESSION

Non-linear regression in the form of polynomial regression overcomes the drawbacks of linear models by allowing for complicated curves and a range of estimations that are not possible with linear regression. To generalize unknown dependencies within restrictions, it uses nth polynomial values, which improves system integrity and reliability while reducing mistakes. Compared to linear regression, it is more effective and efficient since it takes into account unique

 DOI : 10.35291/2454-9150.2023.0239

values and is widely relevant in scenarios like tissue growth ratio determination and epidemics. In constraint analysis, this method improves accuracy and user understanding by creating prediction models based on strong dependencies while maintaining data integrity during processing.

4.3. ARIMA MODEL

Based on time series data, the ARIMA model—a well-known instrument in economic forecasting—is used to anticipate stock prices. Preprocessing, training the model, and data collecting are the steps in its execution. Establishing the regression and moving average orders inside the model is aided by figuring out the differencing, auto-correlation, and partial auto-correlation orders. The three primary parts of ARIMA are Moving Average (MA) based on historical prediction mistakes, Integrated (I) which guarantees stationarity by differencing, and Auto-Regression (AR) which uses lag values as predictors. Together, these components make up the ARIMA (p, d, q) model, in which 'p' stands for AR order, 'd' for differencing, and 'q' for MA order. In the end, ARIMA converts time series data into a stationary format so that it may be effectively analyzed and predicted.

4.4. NESPY

NSEpy is a library for extracting historical and real-time data from the National Stock Exchange's website. The API for this library is designed to be as basic as possible. Python and the Scipy stack are excellent tools for data analysis, and NSEpy's major goal is to provide analysis-ready data-series for use with the Scipy stack. NSEpy is compatible with the Technical Analysis library like MACD, RSI.

4.5.CELERY

Task queues are a way to distribute work across multiple threads or machines. A task is a unit of work that goes into a task queue. Task queues are regularly monitored by dedicated worker processes for new work to accomplish. Celery communicates with clients and workers via messages, with a broker acting as a middleman. To begin a job, the client adds a message to the queue, which is then delivered to a worker by the broker. Multiple workers and brokers can be used in a Celery system, allowing for high availability and horizontal scaling.

4.6.RADIS SERVER

Redis is a powerful key-value store that is free source and ideal for developing high- performance, scalable web applications.

Redis is distinguished by three primary characteristics.

1.Redis stores its database entirely in memory, with the disc serving only as a backup.

2.When compared to other key-value data stores, Redis has a large number of data types.

3.Redis is capable of replicating data to an unlimited number of slaves.

4.7.EDA ALGORITHM

The four main parts of the EDA algorithm are replacement, sampling, learning, and selection. Through the use of AMP optimization domain methodologies, these components can be improved. Cyber-EDA is presented in this study, which also summarizes new features for each EDA component.

4.8. THRESHOLDING WITH SHORT TERM MEMORY

This approach addresses biases from restricted sampling in Thresholding with Short-Term Memory by tracking chosen solution frequencies in the current population. Adaptive thresholds are used in this approach to balance the frequencies of the solutions, avoiding oversampling and the dominance of the most common solution. Overcounts are dispersed among different frequencies. In order to guarantee proportionate evolution representation, the adaptive threshold modifies in response to finished function evaluations. The present threshold is set by increasing the previous value in proportion to the evolution that has been completed, where 'n' scales function evaluations to the threshold value. This successfully preserves variety and reduces bias resulting from small sample and population sizes.

4.9. COMPARATIVE STUDY OF IMPLEMENTED MODULES

Sr. no	Implemented Algorithms and Model	Description	Accuracy Percentage (%)
1.	Linear Regression	Performed analysis on data elements to define a single entity, i.e., result value of stock predicted.	79 % to 81 %
2.	Arima Model	Variable difference ordering and algorithm implemented to increase the accuracy rates	91 % to 93 %
3.	EDA Algorithm	Diversification is generated and thresh holding of short-term memory is done	89% to 92 %

V. RESULT & DISCUSSION

Module 1: User Authentication and Management

1)Login Page:

In this figure login page are there. This page is used to login existing user, if user is not available in database it won't login. If user is valid then it allowed user to login to dashboard page. There is "Register Here" button which redirect user to registration form where user can register themselves

 DOI : 10.35291/2454-9150.2023.0239

International Journal for Research in Engineering Application & Management (IJREAM)
ISSN : 2454-9150 Vol-09, Issue-09, Dec 2023

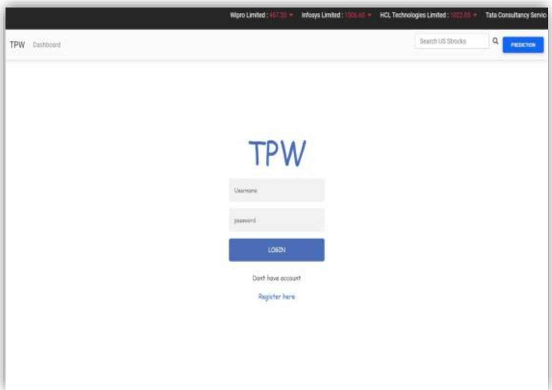

Fig 5.1 : Login Page

2)Registration Page:

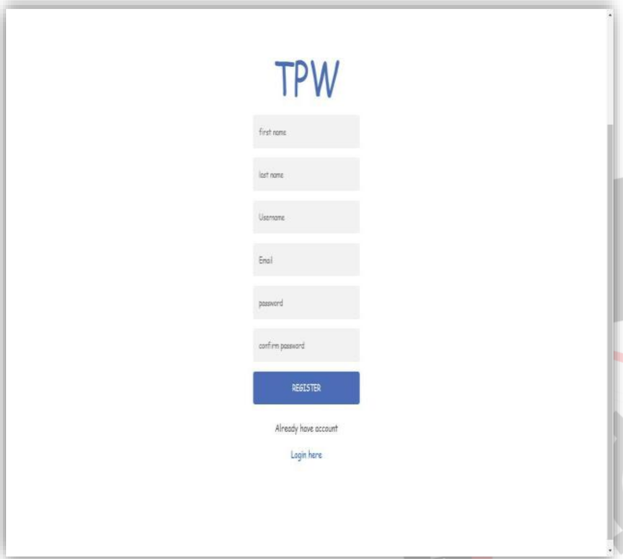

Fig 5.2 : Registration Page

This is Registration Page, where user need to choose valid Username, valid Email and valid Password. After entering valid information user can register in to database. After successfully register, user will be redirected to login page, where user need to login using the credential.

Module 2: Backend Connectivity with Django

1)In above figure there are list of various stock where user can see details about specific stock.

It also shows price change, previous close, Day high, Day Low, Market Price.

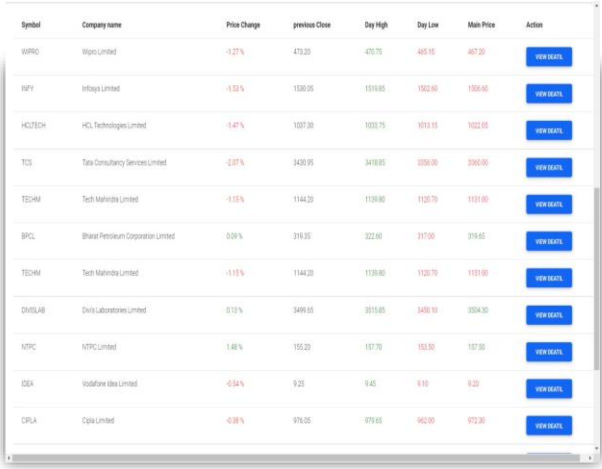

Fig 5.3 : List of Stock

2) In above figure we are displaying the chart showing fluctuations in costs of IBM shares, there are other options shown (1m, 3m, 6m, YTD, 1y, All) these options display the stocks of different timeline with candlestick and volume on that specific day. On right side user can choose specific date using "From – To" option, using month and date.

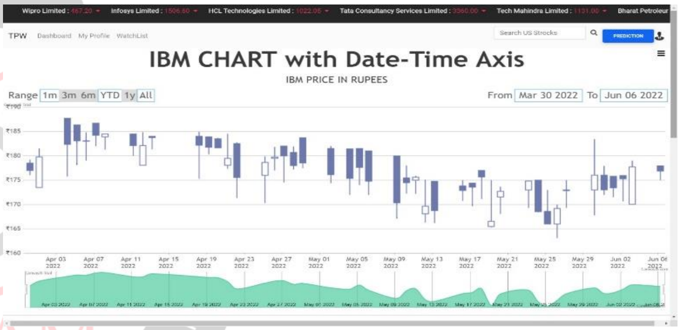

Fig 5.4 : IBM stock displaying chart

Module 3: Machine Learning Models

In above figure the actual prediction is showing. Different stock indicates in different color line. The line shows that where stock goes in upcoming year, and according to future prediction it shows whether user need to buy it sell the stock.

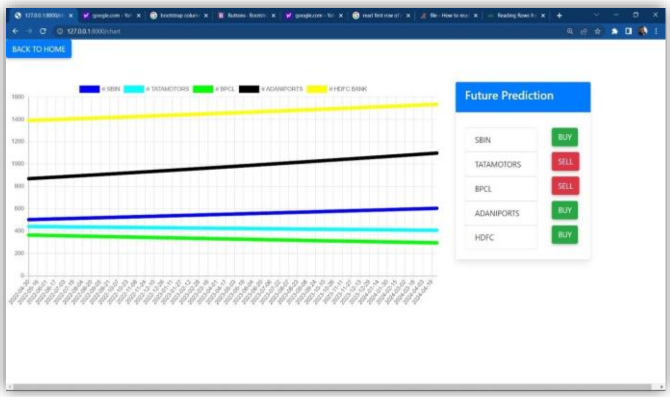

Fig 5.5 : Prediction

 DOI : 10.35291/2454-9150.2023.0239

International Journal for Research in Engineering Application & Management (IJREAM)
ISSN : 2454-9150 Vol-09, Issue-09, Dec 2023

VI. CONCLUSION & FUTURE SCOPE

In summary, stock is an unpredictable system that tracks chain segments and also exhibits surprising interdependence. It is described as a curve that is constantly shifting, causing prices to go from low to high and vice versa.

The accuracy is compromised when one reliance is left unchecked because of the increased integration of the same with other dependents. Since the real prediction cannot be made for any fiscal day and is always changing and flipping the tables, accuracy is not the word used to describe overstock. It is more practical and flexible in nature when there are more component assets and dependencies, which makes prediction even more difficult.

The hit, profit, or gain rate is computed for the approximate value after taking it into account.

The project incorporates a number of sophisticated machine learning algorithms that are integrated and put into practice. The output produced by these algorithms is displayed to the user as a graph, which helps them understand the scenario and decide whether to invest and profit from it.

The suggested program processes the dataset's or the.csv file's raw collection of data. After data has been cleaned and purified, it is processed further to provide useful results. The output is shown as a graph on the screen following the computation of the mean.

VII. REFERENCES

[1] D. P. Gandhmal and K. Kumar, "Systematic analysis and review of stock market prediction techniqes," Computer Science Review, vol. 34, 2019, doi: 10.1016/j.cosrev.2019.08.001.

[2] Patel, J., Shah, S., Thakkar, P., & Kotecha, K. (2015). Predicting stock and stock price index ovemet using Trend Deterministic Data Preparation and machine learning techniques. Expert Systems with Applications, 42(1), 259-268.

[3] [3]EsCambria, D. Olsher, and D. Rajagopal, ³Senticnet 3: a common and common-sense knowledge base for cognition-driven sentiment analysis,´ AAAI, pp. 1515±1521, 2014.

[4] M. Ballings, D. V. D. Poel, N. Hespeels, and R. Gryp, "Evaluating multiple classifiers for stock price direction prediction," Expert Systems with Applications, vol. 42, no. 20, pp. 7046–7056 2015.

[5] L. Nassar, I. E. Okwuchi, M. Saad, F. Karray and K. Ponnambalam, "Deep Learning Based Approach for Fresh Produce Market Price Prediction," 2020 International Joint Conference on Neural Networks (IJCNN), Glasgow, United Kingdom, 2020, pp. 1-7, doi: 10.1109/IJCNN48605.2020.9207537.

[6] V. Unadkat, P. Sayani, P. Kanani and P. Doshi, "Deep Learning for Financial Prediction," 2018 International Conference on Circuits and Systems in Digital Enterprise Technology (ICCSDET), Kottayam, India, 2018, pp. 1-6, doi:10.1109/ICCSDET.2018.8821178.

[7] Convolutional LSTM, LSTM, Fully Connected doc- X. Shi, Z. Chen, H. Wang, D. Yeung, W. Wong, and W. Woo, "Convolutional LSTM network: A machine learning approach for precipitation nowcasting," CoRR, vol. abs/1506.04214,

 DOI : 10.35291/2454-9150.2023.0239