A Novel Approach Stock Price Predication Using Machine Learning

[1]Dr. Ronak Panchal, [2]Miss. Khusbu Champaneria, [3]Dr. Ankit Patel
[1] Adjunct Professor of 1st Author,[2]Student Of VBTMCA of 2nd Author,[3]Professor of 3rd Author
[1]Department of Information and Communication Technology of 1st Author, Surat, India
[2]Department of Computer Science, Vidyabharti Trust College of MCA of 2nd Author, Surat, India
[3]Department of Computer Science, Vidyabharti Trust College of MCA of 3rd Author, Surat, India

Abstract:

In this research work, we have provided various machine learning algorithms and using them we have predicted stock market predication and recommendation in given yahoo finance(yfinance) dataset. Apart from this, we have done Exploratory Data Analysis, Statistical Information, stock price data visualization, predication and recommendation display on UI using Python Streamlit framework in which we have done various charts of stock Indicators on relations between various fields. The technical indicator of provide valuable insights into stock price volatility, trend direction, and potential reversals in financial markets. We have used machine learning algorithms include linear regression, logistic regression, decision trees, random forests, support vector machines, k-means for Stock Price Prediction and Recommendation like stock share is buyable, sellable or hold.

Keywords: Machine Learning Algorithms, Technical indicator, Stock Price Prediction, Exploratory Data Analysis, Streamlit.

1.Introduction to Stock Market and Data Science:

In recent years, the intersection of data science and algorithms has revolutionized various fields, ranging from healthcare to finance, from marketing to entertainment. This introduction sets the stage for understanding the fundamental concepts and significance of data science and algorithms in today's technological landscape as a stock market. The stock market is known for its ever-changing, unpredictable, and complex nature. Forecasting stock prices presents a formidable challenge due to a multitude of factors and analyzing trends from past years can provide valuable insights for anticipating market movements, thereby assisting investors in making informed decisions and maximizing their investment returns. This is an interesting topic for research and we hope to achieve significant growth.

Overview of Machine Learning Algorithms

Machine learning is a branch of artificial intelligence (AI) focused on developing algorithms that enable computers to learn and improve from experience without being explicitly programmed. It involves the creation of models that can analyze data, identify patterns, and make decisions or predictions. Machine learning algorithms are trained using large datasets, allowing them to recognize complex relationships and make accurate predictions or classifications. Key types of machine learning include supervised learning, where models learn from labeled data; unsupervised learning, where models identify patterns in unlabeled data; and reinforcement learning, where models learn by interacting with an environment and receiving feedback. Machine learning has applications across various fields, including healthcare, finance, marketing, and robotics, and continues to drive innovation in AI technology.

Supervised Learning	Unsupervised Learning
Supervised Learning can be used for 2 different types of problems i.e. regression and classification	It can be used for two different types of problems like clustering and association
Here, input and output data is provided.	Here, only input data is provided.
Here, the data is labelled.	Here, the data is not labelled.
Here, results are more accurate.	Here, the results are less accurate.
Here, feedback is also accepted.	Here, no feedback is taken.

Table 1:Difference between supervised and unsupervised learning

Supervised Learning:

Supervised learning is a machine learning paradigm where algorithms learn from labeled data. Labeled data consists of input-output pairs, where the algorithm learns to map input data to corresponding output labels. It involves a training phase where the algorithm learns from examples, and a prediction phase where it makes predictions on new, unseen data. It can be categorized into two main types: Classification and Regression. Compared to unsupervised learning, supervised learning is less complex and tends to yield more accurate results. In Classification, the goal is to categorize input data into distinct classes or groups, while Regression aims to predict continuous values based on input features And other Regression algorithms are utilized in scenarios such as predicting prices or weather conditions, where continuous predictions are required.

Linear Regression:

Linear Regression is a fundamental statistical method used for modeling the relationship between a dependent variable and one or more independent variables. It assumes a linear relationship between the input variables and the output, making it suitable for predicting continuous values. There are primarily three types of linear regression:

1. **Simple Linear Regression:** Simple Linear Regression involves only two variables: one independent variable (predictor) and one dependent variable (response). It assumes a linear relationship between the predictor and the response, represented by a straight line.
2. **Multiple Linear Regression:** Multiple Linear Regression extends the concept of simple linear regression to more than one independent variable. It assumes a linear relationship between the dependent variable and multiple independent variables.
3. **Polynomial Regression:** Polynomial Regression is a form of linear regression where the relationship between the independent variable and the dependent variable is modeled as an nth-degree polynomial.

Logistic Regression:

Logistic Regression is a widely-used machine learning algorithm for predicting categorical outcomes based on independent variables. It's specifically designed for binary classification tasks, where the outcome is either 0 or 1, yes or no, true or false, etc. The model produces predictions on a sigmoid curve, which smoothly transitions between the two possible outcomes. Under the umbrella of Classification, Logistic Regression encompasses three main categories: Binomial, Multinomial, and Ordinal, each suited for different types of categorical data.

Random Forest:

Random Forest is a versatile and powerful machine learning algorithm used for both classification and regression tasks. It operates by constructing a multitude of decision trees during training and outputs the mode (for classification) or mean prediction (for regression) of the individual trees. Random Forest mitigates overfitting by aggregating predictions from multiple trees, resulting in robust and accurate models. It's particularly effective for handling high-dimensional datasets with complex relationships between features and outcomes.

II. RELATED WORKS

Predicting stock prices presents a significant challenge due to their unpredictable and fluctuating nature, often influenced by various factors and time variations. In recent times, researchers have turned to machine learning techniques to aid in making trading decisions within the stock market. In this overview, we'll briefly examine some of the research endeavours in this domain.

III. WORKFLOW:

This project deal with gather historical data on stock prices from YFinance, along with relevant factors that may influence price movements such as market indices and company-specific metrics. Conduct EDA to gain insights into the relationships between different variables and their impact on stock prices. This may involve visualizations such as line chart, scatter plots, and correlation matrices. Create new features or transform existing ones to improve the predictive power of the model. This could include technical indicators, sentiment analysis of news articles, or lagged variables. Choose appropriate machine learning or statistical models for stock price prediction. Commonly used models include linear regression, support vector machines (SVM), decision trees and random forest. Model is split the data into training and testing sets, and train the chosen model on the training data. Evaluate the performance of the trained model using appropriate metrics such as mean squared error (MSE), mean absolute error (MAE),r2_score or accuracy. This helps assess how well the model generalizes to unseen data. Fine-tune hyperparameters of the model or experiment with different algorithms to improve performance further. Once the model is trained and evaluated satisfactorily, use it to make predictions on new or unseen data. Deploy the trained model into a production environment where it can be used to generate real-time predictions or insights. Continuously monitor the model's performance and retrain it periodically with updated data to ensure its accuracy and relevance over time and recommendation of the data buyable, sellable or hold the stock in the market.

2. Literature review:

According to Kabita Sahoo and Abhaya Kumar Samal have described the different data exploration data analysis and visualize data using some charts. [2]
As Stated by Tahir Sher1, Abdul Rehman [3] research work the data accessed from the Yahoo Finance API belongs to two stock exchanges. The data set is divided into three phases: the training, validation, and testing phases.
In Based on Tran Phuoc[1,2], Pham Thi Kim Anh[1,2], Chien V. Nguyen have aimed of this study are to predict the stock price trend in the stock market in an emerging economy. The corresponding technical analysis indicators for each stock code include: simple moving average (SMA), convergence divergence moving average (MACD), and relative strength index (RSI); and the secondary data from VN-Index and VN-30 stocks [4].

In consonance with SUBHA SINGH, SREEDEVI GUTTA have been learned the trends with huge amount of data provided and train themselves to predict the movement of the stock price [5].

As reported by Amit Gupta, T.J. Nagalakshm have predicted the end result of an incident supported the link between variables obtained from the data-set. statistical regression is one kind regression employed in Machine Learning. [6]

According to Poornima S P, Priyanka C N are going to apply KNN method and linear regression for predicting the stocks. The stock holders can invest confidently based on the results obtained from the model. [7]

As Mehar Vijha, Deeksha Chandola have been utilized for predicting the next day closing price for five companies belonging to different sectors of operation. [8]

As said by Paresh Shrikhande[1], Raghu Ramani[2] have scaling processed split to avoid the information leakage that could have leaked testing period data into training period data. [9]

Hence , Aarti Puthran, Ashutosh Shukla have using past prices of stock and news headlines together to predict the stock prices. This paper will give recommendations on stock buy /sell on the basis of current news [10]

3.Implementation diagram and analysis

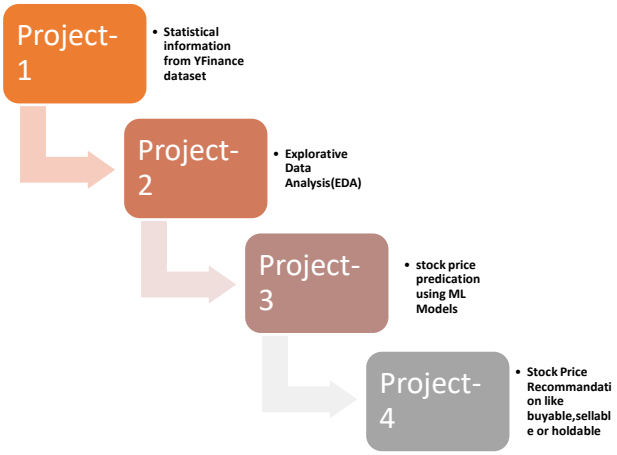

Figure 1: Phases of Stock Price Project

Technical Analysis:

Another prevalent approach utilized in stock market modeling and prediction is technical analysis. This methodology relies heavily on historical market data, particularly price and volume metrics. By scrutinizing these quantitative aspects, technical analysts aim to discern patterns and trends that may offer insights into future market movements. [4] These are mathematical calculations based on historical price and volume data. Indicators include moving averages, relative strength index (RSI), stochastic oscillators, MACD (Moving Average Convergence Divergence), and Bollinger Bands. These indicators help identify trends, momentum, overbought or oversold conditions, and potential trend reversals.

Input Data:

Real-time data for stocks is obtained from Yahoo Finance and serves as input for the model. Using the seaborn Python library, the fluctuations in these stocks' data over the past year are visualized. The visualization encompasses the data from exactly one year ago to the present date.

Analysis Of Dataset:

MACD (Moving Average Convergence Divergence):

MACD is a momentum indicator showing the relation between two moving averages of a security's price. It includes the MACD line, signal line, and histogram. The MACD line is the difference between short-term and long-term EMAs. Traders utilize MACD for trend, momentum, and buy/sell signals.

EMA (Exponential Moving Average Indicator):

EMA calculates the Exponential Moving Average (EMA) of a given time series, which is more responsive to recent price changes compared to Simple Moving Average (SMA). It applies a smoothing factor to the previous EMA value and the current price, giving more weight to recent prices. Traders use EMA to identify trends, support and resistance levels, and generate buy or sell signals, particularly in fast-moving markets.

SMA (Simple Moving Average Indicator):

SMA calculates the Simple Moving Average (SMA) of a time series by taking the arithmetic mean of a set number of prices over a defined period. It smooths out price data over time, assigning equal importance to each data point. SMA is frequently employed to spot trends, support/resistance levels, and to trigger buy/sell signals. Yet, it can lag in volatile markets.

Bollinger bands:

A Bollinger band is an indicator that provides a range within which the price of an asset typically trades. The width of the band increases and decreases to reflect recent volatility. The closer the bands are to each other – or the 'narrower' they are – the lower the perceived volatility of the financial instrument. The wider the bands, the higher the perceived volatility.

Closing Price:

We get the data directly from Yahoo Finance's server, and it covers various stocks for a year. We use Python's matplotlib and pyplot library to analyze this data. Specifically, we look at the stock prices of companies like Tesla, PepsiCo, Cisco, and Marriott International. We visualize how much stock these companies are trading each day, which is their sales volume.

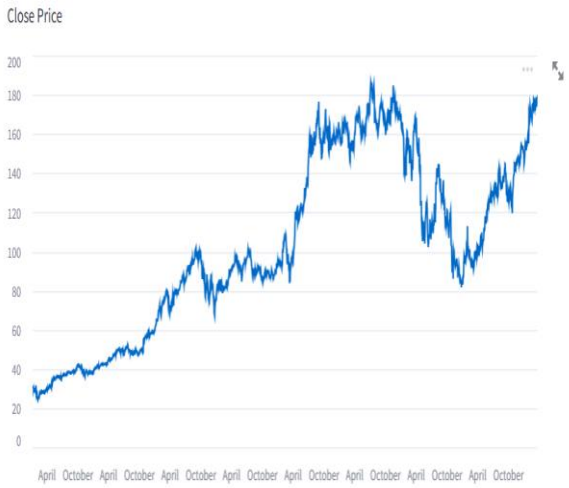

Figure 2: Closing Price

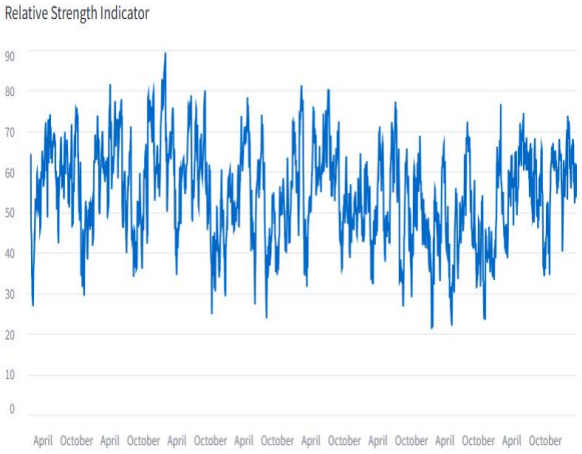

Figure 3: Relative Strength Indicator

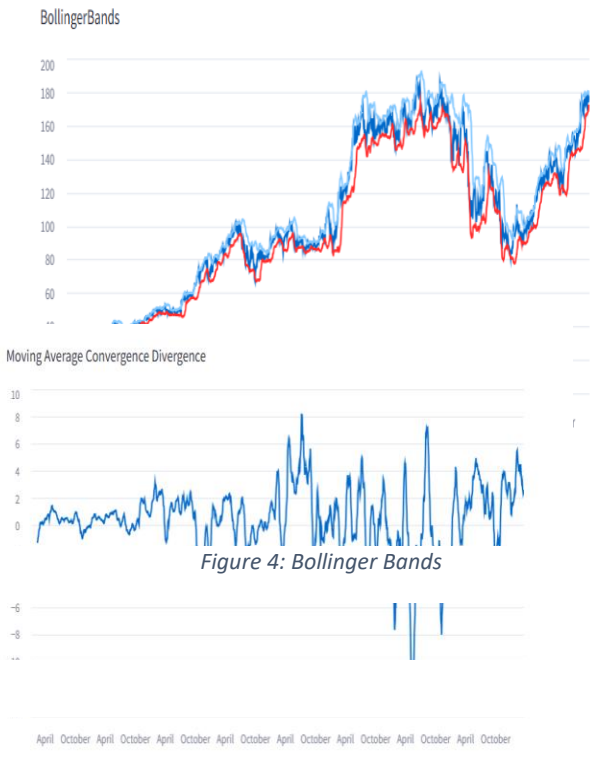

Figure 4: Bollinger Bands

Figure 5: Moving Average Convergence Divergence

Figure 6: Simple Moving Average

Figure 7: Expoenetial Moving Average

Model Building:

- After completing data analysis, we proceeded with model development.
- Utilizing the SKlearn library in Python, we imported a Linear Regression model and applied it to our dataset using a pipeline.
- Subsequently, we partitioned the data into training and testing sets, fitting the pipeline to the training data first.
- Following this, we made predictions on the dependent variable (Y) and evaluated the model's performance on the testing data.
- Assessing the accuracy using the R-squared score revealed a remarkably high accuracy rate of 98%.

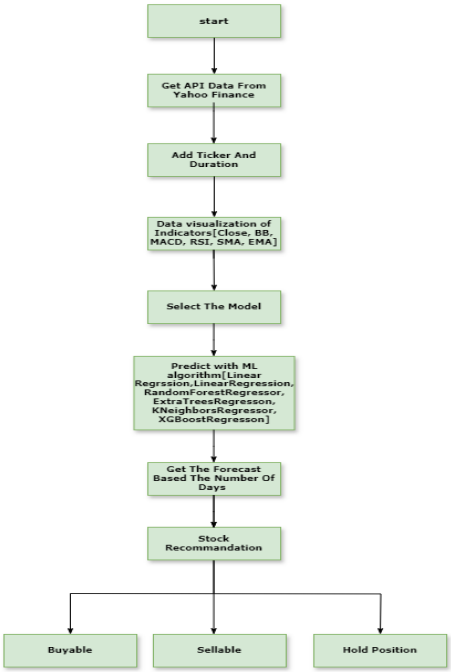

Figure 5: Model Building

4. Future use and conclusion:

Future Use:

Stock Price Prediction using machine learning algorithm helps you discover the future value of company stock and other financial assets traded on an exchange. The entire idea of predicting stock prices is to gain significant profits. Predicting how the stock market will perform is a hard task to do. There are other factors involved in the prediction, such as physical and psychological factors, rational and irrational behaviour, and so on. All these factors combine to make share prices dynamic and volatile. This makes it very difficult to predict stock prices with high accuracy.

Conclusion:

This project leveraged data analysis and machine learning algorithms to devise a scalable model for price prediction in India. Crafting an efficient machine learning algorithm necessitated rigorous training, testing, and evaluation of multiple algorithms. Each experiment was conducted within the streamlit environment due to its efficient training capabilities. we aim to gather additional data and explore training with different algorithms. Moreover, there's potential to transform our project into a feature-rich mobile application, catering to a wider user base. This evolution would enable real-time processing capabilities, enhancing the utility and accessibility of the program.

5. Reference:

Supervised vs Unsupervised Learning - Javatpoint

6. Acknowledge:

I am thankful to Dr. Ronak Panchal and Dr. Ankit Patel sir for guidance. GitHub link is here:

7. Bibliography:

Bibliography

[1] javatpoint, “difference-between-supervised-and-unsupervised-learning,” javatpoint, [Online]. Available: https://www.javatpoint.com/difference-between-supervised-and-unsupervised-learning. [Accessed 10 03 2024].

[2] A. K. S. Kabita Sahoo, “Exploratory Data Analysis using Python,” *Journal of Innovative Technology and Exploring Engineering (IJITEE),* pp. 1-10, 2019.

[3] T. Sher1, “Exploiting Data Science for Measuring the Performance of Technology Stocks,” *Department of Creative Technologies, Air University,* vol. 76, pp. 1-17, 2023.

[4] 2. P. T. K. A. C. V. N. Tran Phuoc1, “Applying machine learning algorithms to predict the stock price trend in the stock market-The case of Vietnam,” *Humanities & Social Science Communication,* pp. 1-18, 2024.

[5] S. G. SHUBHA SINGH, “Stock Prediction Using Machine Learning,” *Department of Computer Science and Information System California State University San Marcos* , vol. 9, p. 7, 2021.

[6] T. Amit Gupta, “Stock Price Prediction using Linear Regression in Machine Learning,” *International Journal of Innovative Technology and Exploring Engineering (IJITEE),* vol. 8, no. 12, p. 4, 2019.

[7] P. C. N. Poornima S P, “Stock Price Prediction using KNN and Linear Regression,” *International Journal of Engineering and Advanced Technology (IJEAT),* vol. 8, p. 4, 2019.

[8] ,. D. C. Mehar Vijha, “Stock Closing Price Prediction using Machine Learning Techniques,” *International Conference on Computational Intelligence and Data Science (ICCIDS 2019),* p. 8, 2020.

[9] R. R. Paresh Shrikhande, “Stock Market Analysis and Prediction,” *International Journal for Research in Applied Science & Engineering Technology (IJRASET)* , vol. 10, p. 12, 2022.

[10] A. S. Aarti Puthran, “Stock Market Prediction and Recommendation System,” *International Journal Of Creative Research Thoughts,* vol. 10, p. 6, 2022.

[11] S. G. SHUBHA SINGH, “Stock Prediction Using Machine Learning,” *WSEAS TRANSACTIONS on COMPUTER RESEARCH ,* vol. 9, p. 7, 2021.

8. AUTHORS PROFILE:

Dr. Ronak Panchal, Prof. Ronak Panchal, was born in 1986, received the PhD degree in computer science from Sabarmati University, Ahmedabad. He is an Adjunct Professor at Vidyabharti Trust College of MCA, Bardoli, Surat. He is working as senior data scientist in IT MNC company, Pune. His main research interests lie in Data science, Semantic Web and Ontology.

Miss. Khusbu Champaneria, was born in 1999, She is studying from Vidyabharti Trust College Of MCA, Bardoli, Surat. Also, she is working as trainee as Data Scientist in Banana SoftInfoTech, Bardoli, Surat.