A
Mini Project Report on

Stock Prediction and Analysis Using Machine Learning

Submitted in partial fulfillment of the requirements
for the degree of
BACHELOR OF ENGINEERING
IN
Computer Science & Engineering
Artificial Intelligence & Machine Learning

by

Soham Shelke - 23106044

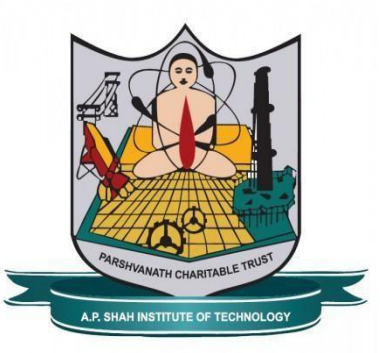
Atharva Tawde - 23106122

Atharva Samant-23106057

Raunak Panigrahi - 23106092

Under the guidance of

Prof. Shraddha Dalvi



Department of Computer Science & Engineering
(Artificial Intelligence & Machine Learning)
A. P. Shah Institute of Technology
G. B. Road, Kasarvadavali, Thane (W)-400615
University Of Mumbai
2024-2025

Parshvanath Charitable Trust's
A. P. SHAH INSTITUTE OF TECHNOLOGY
(Approved by AICTE New Delhi & Govt. of Maharashtra, Affiliated to University of Mumbai)
(Religious Jain Minority)

CERTIFICATE

This is to certify that the project entitled "**Stock Prediction and Analysis Using Machine Learning"** is a bonafide work of Soham Shelke (23106044), Atharva Tawde (23106122), Atharva Samant (23106057), Raunak Panigrahi (23106092) submitted to the University of Mumbai in partial fulfillment of the requirement for the award of **Bachelor of Engineering** in **Computer Science & Engineering (Artificial Intelligence & Machine Learning).**

Prof. Shraddha Dalvi	Dr. Jaya Gupta
Mini Project Guide	Head of Department

Parshvanath Charitable Trust's
A. P. SHAH INSTITUTE OF TECHNOLOGY
(Approved by AICTE New Delhi & Govt. of Maharashtra, Affiliated to University of Mumbai)
(Religious Jain Minority)

Project Report Approval

This Mini project report entitled "**Stock Prediction and Analysis Using Machine Learning"** by **Soham Shelke, Atharva Tawde, Atharva Samant and Raunak Panigrahi** is approved for the degree of ***Bachelor of Engineering*** in ***Computer Science &Engineering***, (AI & ML) ***2024-25***.

External Examiner: _____

Internal Examiner: _____

Place: APSIT, Thane

Date:

Declaration

We declare that this written submission represents my ideas in my own words and where others' ideas or words have been included, I have adequately cited and referenced the original sources. I also declare that I have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in my submission. I understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission hasnot been taken when needed.

Soham Shelke	Atharva Tawde	Atharva Samant	Raunak Panigrahi
(23106044)	(23106122)	(23106057)	(23106092)

ABSTRACT

This project explores the application of machine learning techniques to predict stock prices, an area of significant interest in the financial sector. The primary objective is to develop a model that accurately forecasts the closing price of a selected stock, enabling investors to make informed decisions. We employed various machine learning algorithms, including Linear Regression and Decision Tree, to analyze historical stock data. The dataset was sourced from Kaggle and included key features such as opening price, highest and lowest prices, volume, and closing price.

Model Evaluation
The models were evaluated using metrics such as Mean Squared Error (MSE), Mean Absolute Error (MAE), and R-squared (R^2). The Linear Regression model demonstrated a high R^2 value of 0.9972, indicating excellent predictive performance. In contrast, the tree-based models showed moderate R^2 values, suggesting that while the decision tree captured some relationships, it had room for improvement in generalization.

Key Insights and Future Work
Our findings highlight the importance of selecting the appropriate model for stock price prediction and the potential of linear models in capturing market trends. Future work will focus on optimizing tree-based models and exploring additional features to enhance predictive accuracy.

Index

CHAPTER 1

INTRODUCTION

1. INTRODUCTION

Stock market prediction is a critical aspect of financial analysis that involves forecasting the future value of a company's stock or other financial instruments traded on an exchange. Accurate predictions enable investors to make better decisions, maximizing returns while minimizing risks. However, the stock market is influenced by numerous factors, including economic indicators, market sentiment, and global events, making it highly volatile and challenging to predict.

In recent years, machine learning has emerged as a powerful tool for analyzing and predicting stock prices. By leveraging historical data and identifying patterns, machine learning models can provide insights that traditional statistical methods might miss. These models can learn complex relationships between various features, such as opening price, closing price, volume, and other market indicators.

This project aims to develop a robust machine learning model to predict stock prices with high accuracy. We have explored different algorithms, including Linear Regression, and DecisionTree, to understand their effectiveness in predicting stock prices. The dataset used for this study was sourced from Kaggle, containing historical stock data, which was preprocessed and analyzed to extract meaningful features.

Why Use Machine Learning for Stock Market Prediction?

Traditional methods of stock market analysis, such as technical and fundamental analysis, rely heavily on historical data and human expertise. While these methods are still valuable, they often struggle to account for the complex and non-linear relationships between various market factors. Machine learning, on the other hand, excels at identifying patterns and correlations in large datasets, making it a powerful tool for predicting stock prices.

Machine learning algorithms can process vast amounts of data, including historical stock prices, trading volumes, financial reports, and even news articles and social media sentiment. By analyzing this data, ML models can uncover hidden patterns and make predictions about future price movements.

Objectives:

- To explore and compare different machine learning algorithms for stock price prediction.
- To develop a predictive model that can forecast stock prices accurately.
- To analyze the performance of the models using evaluation metrics such as MSE, MAE, and R^2.

Motivation: The motivation behind this project stems from the increasing reliance on data-driven decisions in the financial industry. Accurate stock price predictions can lead to significant financial gains and improve investment strategies. By harnessing the power of machine learning, this project seeks to contribute to the development of more reliable and effective prediction models.

CHAPTER 2

LITERATURE SURVEY

2. LITERATURE SURVEY

2.1 History

1. **Early Approaches**:
    - Fundamental Analysis: Investors used financial metrics to assess company health.
    - Technical Analysis: Traders analysed historical price data and patterns to forecast future movements.

2. **Statistical Methods**:
    - Regression analysis emerged as a tool for modelling stock price relationships with economic indicators and past prices.

3. **Linear Regression Development**:
    - Linear Regression became a key method for predicting stock prices, with numerous studies validating its effectiveness despite limitations.

4. **Advancements in Computation:**
    - Increased computational power led to more complex models, incorporating multiple regression and eventually machine learning techniques.

5. **Data Availability:**
    - The rise of online trading and enhanced data access allowed for broader application of statistical methods in stock analysis.

6. **Recent Trends:**
    - Today, stock price prediction includes advanced algorithms and deep learning, while Linear Regression remains a foundational approach.

2.2 Review

1.Stock Price Prediction using Machine Learning
Publisher: IEEE
Authors: B N Varaprasad; Ch. Kundan Kanth; G. Jeevan; Y. Kalyan Chakravarti
Published in & by: 2022 International Conference on Electronics and Renewable Systems(ICEARS).

In the current era stock price prediction plays a key role for prediction of future data with respect to training the past data by using machine learning or deep learning technologies. Building a model and then passing the past data as input that is as training data to the model based on the results acquired need to consider an algorithm which gives better accuracy and response time and segmentation. In this paper for estimating the stock values we are considering LSTM and Regression models of Machine Learning. Factors considered are opening values of stock; closing values of stock, lower and higher values of stock and volume.

2.Stock Market Prediction using Supervised Machine Learning Techniques
Publisher: IEEE
Authors: Zaharaddeen Karami Lawal; Hayati Yassin; Rufai Yusuf Zakari
Published in & by: 2020 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE)

Stock price prediction is one of the most extensively studied and challenging glitches, which is acting so many academicians and industries experts from many fields comprising of economics, and business, arithmetic, and computational science. Predicting the stock market is not a simple task, mainly as a magnitude of the close to random-walk behavior of a stock time series. Millions of people across the globe are investing in stock market daily. A good stock price prediction model will help investors, management and decision makers in making correct and effective decisions. In this paper, we review studies on supervised machine learning models in stock market predictions. The study discussed how supervised machine learning techniques are applied to improve accuracy of stock market predictions. Support Vector Machine (SVM) was found to be the most frequently used technique for stock price prediction due to its good performance and accuracy. Other techniques like Artificial Neural Network (ANN), K-Nearest Neighbor (KNN), Naïve Bayes, Random Forest, Linear Regression and Support Vector Regression (SVR) also showed a promising prediction result.

3.Stock Market Prediction using Machine Learning Techniques
Publisher: IEEE
Authors: Aditi Gupta; Akansha; Khushboo Joshi; Madhu Patel;
Published in & by: 2023 International Conference on Power, Instrumentation, Control and Computing (PICC)

One of the most researched and difficult issues that affects so many academics and industry specialists from various departments of economics, business, mathematics, and computing science is stock price prediction. It is challenging to make accurate predictions about the stock market, primarily due to the stock time series near resemblance to random walks. This paper will examine various machine learning and artificial intelligence (AI) approaches to stock price prediction. This study gives a comprehensive assessment of 22 research publications that recommend various approaches, such as computation techniques, machine learning algorithms, performance metrics, and top journals. Research questions are used to guide the selection of the studies. As a result, these chosen research are assisting in the discovery of ML methods and their dataset for stock market prediction. Due to its excellent performance and accuracy, LSTM (Long Short Term Memory) was discovered to be the technique utilized the most commonly for stock price prediction. Numerous other methods, including CNN (Convolutional Neural Networks), RNN (Recurrent Neural Networks), SVM (Support Vector Machines), RF (Random Forests), and SVR (Support Vector Regressions), also produced encouraging prediction outcomes.

4.Stock market prediction
Publisher: IEEE
Author: Radu Iacomin
Published in & by: 2015 19th International Conference on System Theory, Control and Computing (ICSTCC)

In a financially volatile market, as the stock market, it is important to have a very precise prediction of a future trend. Because of the financial crisis and scoring profits, it is mandatory to have a secure prediction of the values of the stocks. Predicting a non-linear signal requires advanced algorithms of machine learning. The literature contains studies with different machine learning algorithms such as ANN (artificial neural networks) with different feature selection. The results of this study will show that the algorithm of classification SVM (Support Vector Machines) with the help of feature selection PCA (Principal component analysis) will have the success of making a profit.

CHAPTER 3

PROBLEM STATEMENT

3.PROBLEM STATEMENT

The primary challenge in financial markets is predicting stock prices with high accuracy due to their inherent volatility and the influence of numerous external factors. Traditional methods often struggle to capture complex patterns and interactions within the data, leading to suboptimal forecasting performance.

Problem:

To address these challenges, this project aims to develop and evaluate machine learning models for predicting stock prices based on historical data. Specifically, the problem is to create a predictive model that can accurately forecast future stock prices using various features, such as historical prices, trading volume, and other market indicators.

CHAPTER 4

EXPERIMENTAL SETUP

4.EXPERIMENTAL SETUP

4.1 Hardware Setup

- **Minimum Laptop Specifications for Stock Price Prediction:**

1. Processor: Intel Core i5 (8th gen) / AMD Ryzen 5
2. RAM: 8 GB (minimum), 16 GB recommended
3. Storage: 256 GB SSD
4. OS: Windows 10/11 or Ubuntu Linux
5. Battery: 6-8 hours
6. Screen Size: 13-15 inches
7. Connectivity: USB 3.0/Type-C, Wi-Fi 5

4.2 Software Setup

The experimental setup outlines the processes and tools used to develop, train, and evaluate the machine learning models for stock price prediction. It includes details on Programming language used, IDE, libraries, and frameworks.

1.Programming Language:

- Python

2.Integrated Development Environment (IDE):

- Jupyter Notebook

3.Libraries:

- yfinance: To fetch historical stock data.
- pandas: For data manipulation and analysis.
- matplotlib: For data visualization.
- scikit-learn: For applying Linear Regression and Decision Trees.
- numpy: For numerical operations.

4.**Frameworks:**

- Flask: For deploying the machine learning model as a web application.

CHAPTER 5

PROPOSED SYSTEM & IMPLEMENTATION

5.PROPOSED SYSTEM AND IMPLEMENTATION

5.1 Block diagram of proposed system

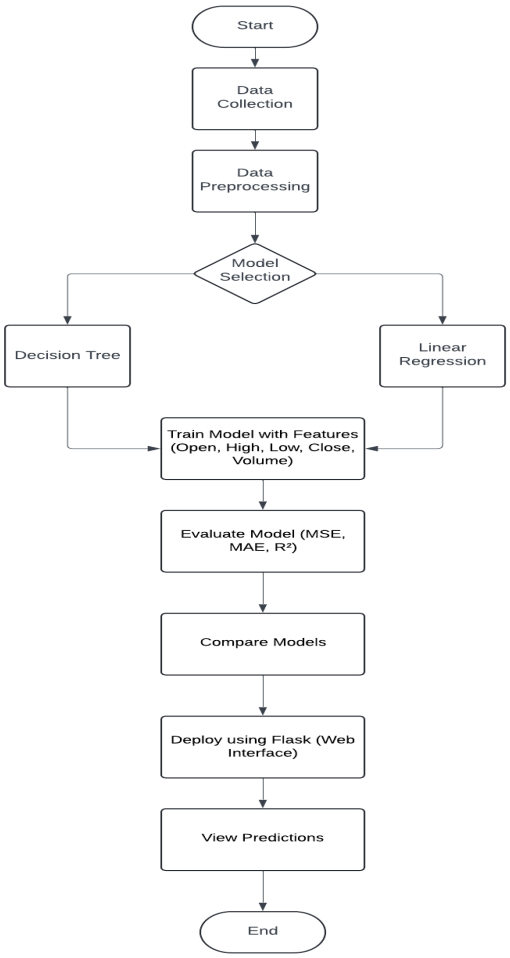


Fig-1 : Block Diagram Of Proposed System

5.2 Description of block diagram

1.Data Collection

- **Source:** Kaggle dataset containing historical stock prices and trading volumes.

2.Data Preprocessing:

- **Cleaning:** Handling missing values and outliers.
- **Feature Selection:** Identifying relevant features that influence stock prices.

3.Model Selection:

- **Linear Regression:** Linear regression is a supervised learning algorithm used to model the relationship between a dependent variable (target) and one or more independent variables (predictors). The goal is to find a linear equation that best describes the data.

  In its multiple form (Multiple Linear Regression), it assumes a linear relationship between the dependent variable Y (Closing Price) and multiple independent variables X(Open, High, Low, Close) The relationship is represented as: Y = m1x1 + m2x2..+ b.

- **Decision Tree:** A Decision Tree is a supervised learning algorithm used for regression and classification tasks. In regression, it models the relationship between the dependent variable Y(Closing Price) and multiple independent variables X (Open, High, Low, Volume) by splitting the data into branches based on feature values. The model predicts by dividing the dataset into smaller subsets based on the feature that minimizes prediction error. The final prediction is the average of target values (e.g., Closing Price) at the leaf nodes. The relationship is represented as a series of decision rules, making it suitable for capturing non-linear relationships.

4.Model Training and Evaluation:

- **Training:** Building and training the models using the processed data.
- **Evaluation:** Using metrics such as Mean Squared Error (MSE), Mean Absolute Error (MAE), and R-squared (R^2) to assess model performance.

5. Deploy Using Flask:

In this project, Flask is used as the web framework to deploy the trained machine learning model and provide a simple user interface to interact with the model. Flask serves as the bridge between the machine learning model and the users, allowing the stock price prediction model to be accessed through a web application.

Key Challenges in Stock Market Prediction

- **Data Quality and Availability**: The accuracy of ML models heavily depends on the quality and quantity of data available. Incomplete or noisy data can lead to poor predictions.

- **Market Volatility**: Stock markets are influenced by unpredictable events such as political changes, natural disasters, and economic crises, making accurate predictions challenging.

- **Overfitting**: ML models may become too complex and start fitting noise in the data rather than actual patterns, leading to poor generalization to new data.

5.3 Implementation :

```
from sklearn.linear_model import LinearRegression
from sklearn.model_selection import train_test_split
import numpy as np
```

```
x = data[['ema_50','rsi_7','rsi_14','cci_7','cci_14','sma_50','sma_100','ema_100','macd','bollinger','TrueRange','atr_7','atr_14']]
y = np.array(data['close']).reshape(-1,1)
```

```
x_train, x_test, y_train, y_test = train_test_split(x,y,test_size=0.2)
```

```
model = LinearRegression()
```

```
model.fit(x_train,y_train)
```

```
LinearRegression
LinearRegression()
```

```
y_pred = model.predict(x_test)
```

```
import matplotlib.pyplot as plt
```

```
actual_data = np.array(y_test).reshape(-1,1)
predicted_data = np.array(y_pred).reshape(-1,1)
```

```
plt.figure(figsize=(10, 6))
plt.plot(actual_data, label='Actual Values', color='b')
plt.plot(predicted_data, label='Predicted Values', color='r', linestyle='--')
plt.xlabel('Sample Index')
plt.ylabel('Value')
plt.title('Actual vs Predicted Values')
plt.legend()
plt.show()
```

Fig-2 : Model Training

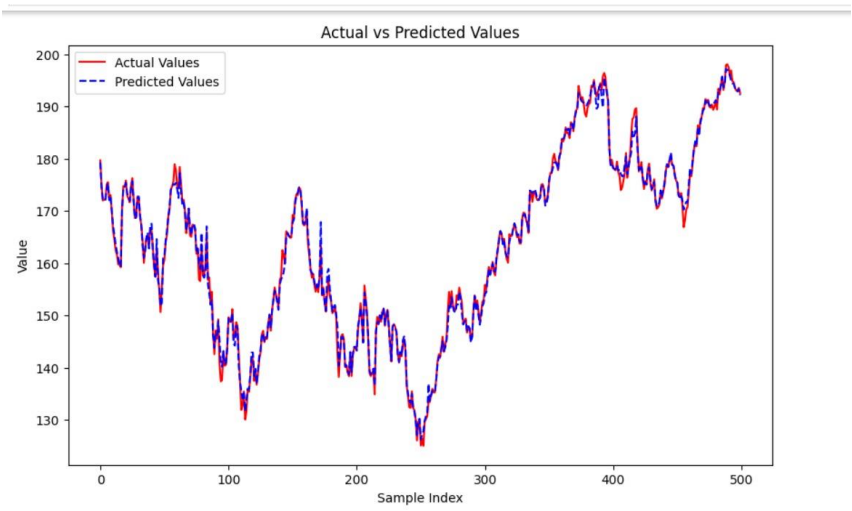


Fig-3: Visualization

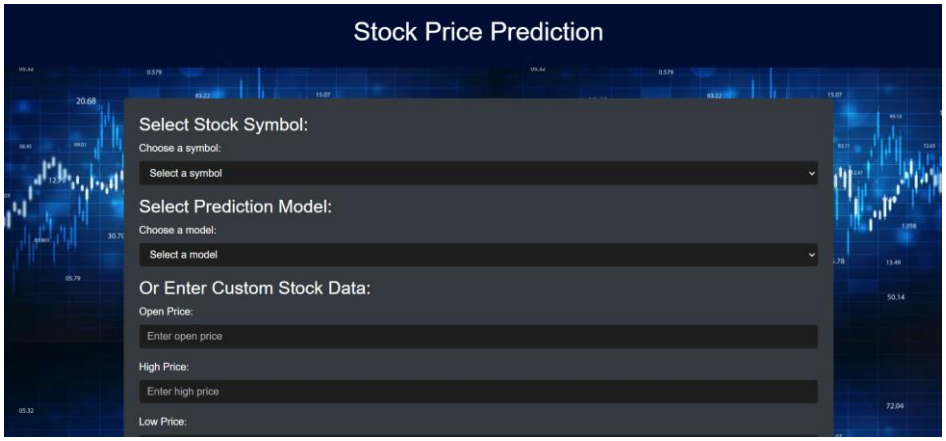


Fig-4: Interface

Fig-5: Actual vs Predicted Graph

5.4 Advantages:

Stock market prediction and analysis using machine learning (ML) offer several advantages, making it a powerful tool for investors, traders, and financial institutions. Here are some key benefits:

1.Improved Accuracy

Pattern Recognition: ML models can analyze vast amounts of historical data to identify patterns and trends that may not be apparent to human analysts. This leads to more accurate predictions of stock prices and market movements.

Complex Data Handling: ML can process and analyze various types of data simultaneously, such as historical prices, trading volumes, news sentiment, and economic indicators, leading to a more comprehensive and accurate prediction model.

2.Speed and Efficiency

Real-Time Analysis: ML algorithms can process and analyze data in real-time, enabling traders and investors to make quick decisions based on current market conditions.

Automation: Machine learning models can automate the analysis and prediction process, reducing the time and effort required for manual analysis and allowing traders to focus on strategy execution.

3.Handling Large Datasets

Big Data Processing: ML models are capable of processing and analyzing massive datasets that include historical prices, financial statements, social media sentiment, and other relevant factors. This ability to handle big data is crucial for making informed predictions in the complex environment of the stock market.

Incorporation of Alternative Data: Machine learning models can incorporate alternative data sources, such as social media, news, and even satellite imagery, providing a more comprehensive view of market sentiment and potential price movements.

CHAPTER 6

CONCLUSION

6.CONCLUSION

Stock market prediction using machine learning is a major advancement in financial technology, offering enhanced tools to analyse data and predict market trends. Machine learning models can uncover patterns, process real-time data, and adapt to changing conditions, giving investors a competitive edge.

While traditional methods are still useful, machine learning improves price forecasting, risk management, and opportunity identification. It also reduces costs by automating analysis. However, challenges like data quality, overfitting, and unpredictable events can affect accuracy.

In conclusion, though not perfect, machine learning significantly enhances market analysis and trading strategies, with growing importance in finance.

7.REFERENCES

1. B. N. Varaprasad, C. Kundan Kanth, G. Jeevan and Y. K. Chakravarti, "Stock Price Prediction using Machine Learning", 2022 International Conference on Electronics and Renewable Systems(ICEARS).

2. Z. K. Lawal, H. Yassin and R. Y. Zakari, "Stock Market Prediction using Supervised Machine Learning Techniques: An Overview," 2020 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE)

3. A. Gupta, Akansha, K. Joshi, M. Patel and V. Pratap, "Stock Market Prediction using Machine Learning Techniques: A Systematic Review," 2023 International Conference on Power, Instrumentation, Control and Computing (PICC)

4. E. Matthes, “Python Crash Course”. San Francisco, California: No Starch Press, 2019.

5. A. Géron, “Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow”. 2nd ed., California: O'Reilly Media, 2019.

6. S. Raschka and V. Mirjalili, “Python Machine Learning”. 3rd ed., Birmingham, UK: Packt Publishing, 2019.