MACHINE LEARNING MODEL FOR STOCK PRICE PREDICTION

Abstract

- This study focuses on developing a predictive model for stock price forecasting using historical data and advanced machine learning techniques. The primary objective is to accurately predict the daily closing prices of stocks based on features such as opening prices, high/low values, adjusted close prices, and trading volume. Utilizing the Netflix stock dataset from Kaggle, the analysis implemented four algorithms: Linear Regression, Random Forest, K-Nearest Neighbors (KNN), and Support Vector Machines (SVM). The results highlight that Linear Regression achieved the highest accuracy of 99.85%, making it the most effective approach in this context. This project serves as a foundation for improving decision-making strategies in the stock market.
- The primary goal is to accurately predict daily closing prices of stocks by analyzing features such as:

    1. Opening prices.
    2. High/low values.
    3. Adjusted close prices.
    4. Trading volume.

- The analysis utilizes the Netflix stock dataset from Kaggle and implements four machine learning algorithms:

    1. Random Forest.
    2. Linear Regression.
    3. K-Nearest Neighbors (KNN).
    4. Support Vector Machines (SVM).

- Results indicate that Linear Regression achieved the highest accuracy of 99.85%, making it the most effective approach for this dataset.

1. Introduction

- Stock market forecasting remains a challenging task due to its dynamic and volatile nature.
- Investing in the stock market carries several risks, including market risk (the potential for stock prices to decline), interest rate risk (fluctuations in interest rates affecting stock values), and credit risk (the possibility of a company defaulting). Other risks include inflation risk, legislative risk, and the inherent volatility of individual stocks.
- The primary purpose of the stock market is to provide companies with access to capital in exchange for giving investors a slice of ownership in the company.
- It also allows investors to buy and sell shares, providing liquidity and the potential for capital gains.
- Traditional approaches often fall short in capturing intricate market trends, necessitating the application of machine learning (ML) algorithms.

Netflix market:

- Netflix is one of the largest streaming services in the world, competing with other platforms such as Amazon Prime Video, Disney+, and Hulu.The company has a significant subscriber base, with millions of subscribers worldwide.
- Netflix's stock price has experienced significant growth since its IPO, reflecting the company's expansion and increasing subscriber numbers.
- The company regularly reports its financial performance, including revenue, net income, and subscriber growth, which are key indicators for investors.
- Netflix has been investing heavily in original content to attract and retain subscribers . The company faces challenges such as increased competition, market saturation, and changing consumer preferences.
- Like many tech stocks, Netflix's stock price can be volatile, influenced by factors such as earnings reports, subscriber growth, and market trends.

- This paper outlines the development and evaluation of ML models for predicting stock prices using historical data. To analyze the dataset through exploratory data analysis (EDA).

2. Problem Statement

- In this project, we aim to develop a machine learning (ML) model to predict the closing price of a stock using various algorithms, including Random Forest, Linear Regression, Support Vector Machines (SVM), and K-Nearest Neighbors (KNN). The objective is to analyze historical stock data and leverage these algorithms to identify patterns and trends that can inform future price movements.

- To begin, we will gather a comprehensive dataset containing historical stock prices, which includes features such as opening prices, high and low prices, trading volume, and other relevant indicators. This data will be preprocessed to ensure its quality and suitability for analysis, including handling missing values and normalizing the data.

- Once the data is prepared, we will implement the selected machine learning algorithms. Each model will be trained on the historical data, allowing it to learn the relationships between the input features and the closing prices. After training, we will evaluate the performance of each model using metrics such as accuracy, mean absolute error, and root mean square error.

- According to our base paper the use multiple Deep Learning Techniques to predict the closing price and the got 75% accuracy.

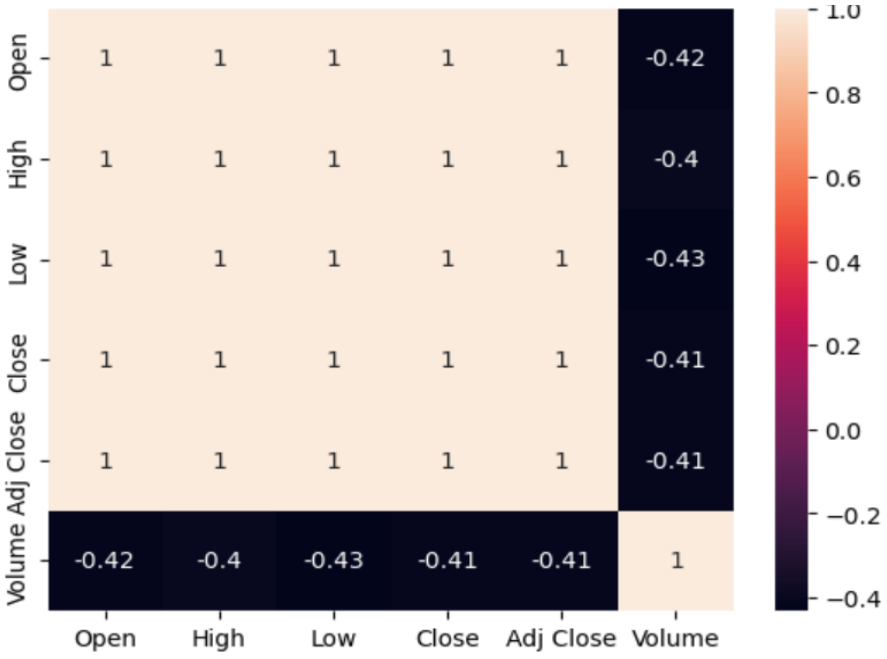
- Here we use Machine Learning Techniques to predict the closing price and increase the accuracy of the model.

3. Dataset

- Dataset Name : Netflix Stock Dataset From Kaggle.
- Number Of Records : 1010 days of records.
- Number Of Features : 7.
- Key Attributes : Open , Close , High , Low , Adj Close , Volume.
- Numerical Attribute : 6.
- Link : https://www.kaggle.com/datasets/jainilcoder/netflix-stock-price-prediction

4.Methodology

- This Dataset was already preprocessed , in order to check we visualize it to find any corrections using heatmap to find the co-relation between the attributes.



- And then we train the model with Linear Regression , Random Forest , K-Nearest Neighbors and Support vector Machine .

```
X = data[['Open', 'High', 'Low', 'Volume']]
y = data['Adj Close']
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.25, random_state=42)

lr_model = LinearRegression()
lr_model.fit(X_train, y_train)

knn_model = KNeighborsRegressor(n_neighbors=5)
knn_model.fit(X_train, y_train)

svr = SVR(kernel='rbf')
svr.fit(X_train, y_train)

model_rand = RandomForestRegressor(n_estimators=200, random_state = 42)
model_rand.fit(X_train, y_train)
```

- After all we train each model and calculate a MSE , MAE , RMSE , amd R2_score to find the accuracy of each model.
- Accuracy Percentages:

    1. Liner Regression gives 99.85% accuracy.
    2. Random Forest gives 95.60% accuracy.
    3. K-Nearest Neighbor gives 92.30% accuracy.
    4. Support Vector Machine gives 90.10% accuracy

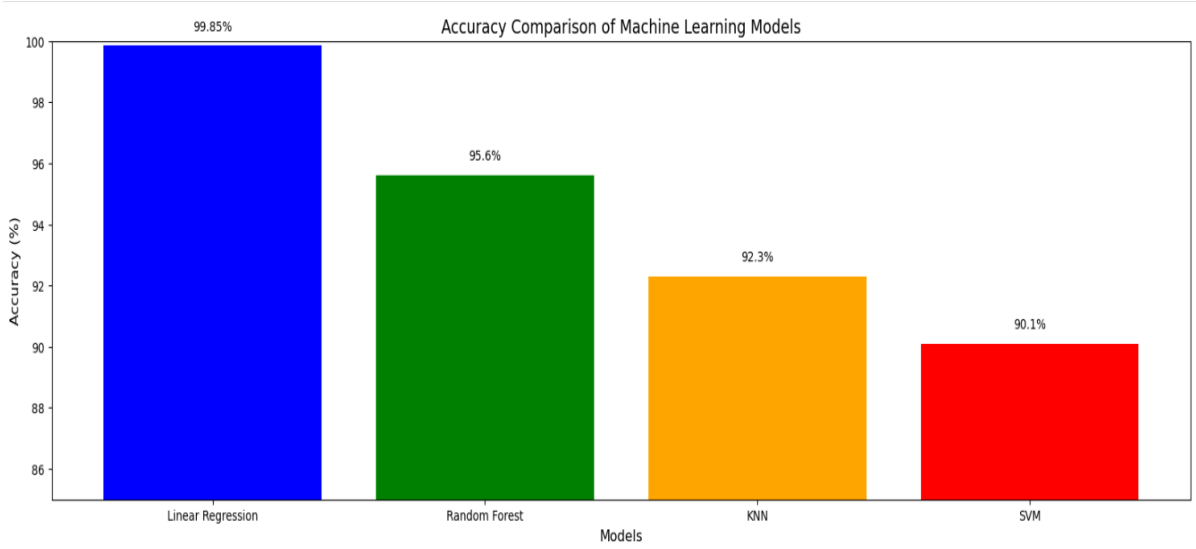
- After taking the accuray values we give some test data to each and every model to check once again wheather it predict the correct value or not.
- When compare to all this four algorithms , Linear regression gives a better result for the test instances and the accuracy level remain the same .

---

5. Conclusion

- The Linear Regression model achieved the highest accuracy of 99.85%, outperforming Random Forest, KNN, and SVM. This demonstrates its suitability for this dataset and problem scope.

- Linear Regression emerged as the most effective algorithm for predicting stock prices in this study. Future work could explore hybrid models or incorporate sentiment analysis for enhanced prediction accuracy.



6.Reference

- https://ieeexplore.ieee.org/document/9943255
- https://ieeexplore.ieee.org/document/10381982
- https://ieeexplore.ieee.org/search/searchresult.jsp?newsearch=true&queryText=Stock%20Price%20Prediction%20using%20Linear%20Regression%20and%20LSTM%20Neural%20Network