A

Project Report on

Stock Market Prediction

Submitted in partial fulfillment of completion of the course

Advanced Diploma in IT, Networking and Cloud Computing

Submitted by:

Anju Luthra

Vidushi Garg

Year 2022-2024

In

NSTI Noida (W)

D-1, Block D, Sector 1, Noida, Uttar Pradesh 201301

Abstract

This project delves into the dynamic realm of stock market analysis, utilizing the powerful yfinance library alongside essential data science tools such as numpy, pandas, and matplotlib. The integration of machine learning, specifically employing the scikit-learn library for linear regression, enhances our ability to make informed predictions. Evaluation metrics, including mean squared error, mean absolute error, and R-squared, contribute to a comprehensive assessment of model performance. Moreover, the project explores the application of confusion_matrix for assessing classification accuracy. Through this synthesis of financial data and advanced analytical techniques, the project aims to empower users with a robust toolkit for insightful stock market analysis.

Acknowledgement

We would like to express our sincere gratitude to several individuals and IBM organization for supporting us throughout our diploma study. First, we wish to express our sincere gratitude to our Edunet mentor, Miss Mala Mishra and Miss Ankita Shukla for their enthusiasm, patience, insightful comments, helpful information, practical advice and unceasing ideas that have helped us tremendously at all times in our study and writing of this project report.

We also wish to express our sincere thanks to our course co-ordinator, Mr. D.A.Guruvulu ,for his valuable suggestions and guidance.

We extend our sincere thanks to our respected Head of the division Smt. Shashi Mathur [JD NSTI(W) NOIDA], for allowing us to use the facilities available.

In addition, we are deeply indebted to the Ministry of Skill Development & Entrepreneurship and IBM for granting us the diploma course. Their technical and financial support has enabled us to complete our diploma course studies successfully

ADVANCE DIPLOMA IN IT NETWORKING & CLOUD COMPUTING

The Advanced Diploma in IT Networking and Cloud Computing program offered by NSTI (W) Noida in collaboration with Edunet Foundation is a comprehensive course designed to equip students with advanced skills in information technology and cloud computing. This program covers a wide range of topics, including Computer Networking, Database Management, Virtualization, Cloud Technologies, and Cybersecurity. Students will gain hands-on experience through practical labs, workshops, and real-world projects, enabling them to excel in the rapidly evolving IT industry. Upon completion of the program, Graduates will have a strong foundation in both IT Fundamentals and Cloud Computing, making them highly sought-after professionals in the field.

Project Requirements

Project Name	**Stock Market Analysis**
Languages Used	**Python**
Editor	**Jupyter Notebook, Google Colab**
Web Browser	**Google Chrome, Microsoft Edge**

Team Composition and Workload Division

Vidushi	Data Analysis, Synopsis
Anju Luthra	Data Analysis, Synopsis

Tables of Content

1. Introduction to Problem

In the dynamic landscape of financial markets, making informed investment decisions is critical for maximizing returns and mitigating risks. The sheer volume of available data, coupled with the complexity of market dynamics, necessitates robust analytical tools. This project addresses the challenge of comprehensively analyzing stock market data, emphasizing data visualization and predictive modeling.

The problem at hand involves leveraging historical stock data to gain insights into market trends and patterns. This analysis aims to go beyond traditional methods by incorporating data visualization techniques such as candlestick charts, moving averages, and volume analysis. Furthermore, the project seeks to predict future stock prices using both elementary models, like simple moving averages and linear regression, and more advanced time series forecasting techniques with the Prophet library.

By delving into these aspects, the project aims to equip investors and analysts with a practical and data-driven approach to navigate the complexities of financial markets. The focus is on providing a toolkit that empowers users to visualize historical performance, identify potential investment opportunities, and make informed decisions in the ever-evolving landscape of stock trading.

2. Requirements

3.1 Technology Stack

Python: High-level programming language used for server-side scripting.

Google Colab: short for Colaboratory, is a free, cloud-based platform provided by Google that allows users to create, share, and run Jupyter notebooks. It offers pre-installed libraries for data science and machine learning, provides free access to GPUs and TPUs, and integrates seamlessly with Google Drive for collaborative and easily shareable coding projects

3.2 Hardware

Laptop/ Computer

3.3 Software

Operating System (OS)

Version Control System

Text Editors and Integrated Development Environments (IDEs)

3. Overview

The data analysis project aims to investigate and derive meaningful insights from a specific dataset. It involves collecting, cleaning, and processing raw data to uncover patterns, trends, and correlations. Using statistical methods and visualization tools, the project seeks to provide a comprehensive understanding of the data, enabling informed decision-making. The analysis may involve exploring relationships between variables, identifying outliers, and creating predictive models. Throughout the project, a systematic approach is followed, including hypothesis testing and validation of results. The ultimate goal is to offer actionable recommendations or conclusions based on the data findings. The project typically employs programming languages such as Python or R, along with tools like Jupyter Notebooks, to facilitate a transparent and reproducible analytical workflow. Overall, the data analysis project serves to extract valuable insights, enhance understanding, and support evidence-based decision-making in a given domain.

4. Project Module

1. Import the required libraries.
2. Load/ Read the Dataset
3. Prepare EDA
4. Do Visualizations
5. Do the Prediction
6. Evaluation of Model

6 Sample Screenshots

```
#importing necessary libraries
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
%matplotlib inline
import seaborn as sns
import yfinance as yf
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error , mean_absolute_error ,r2_score
```

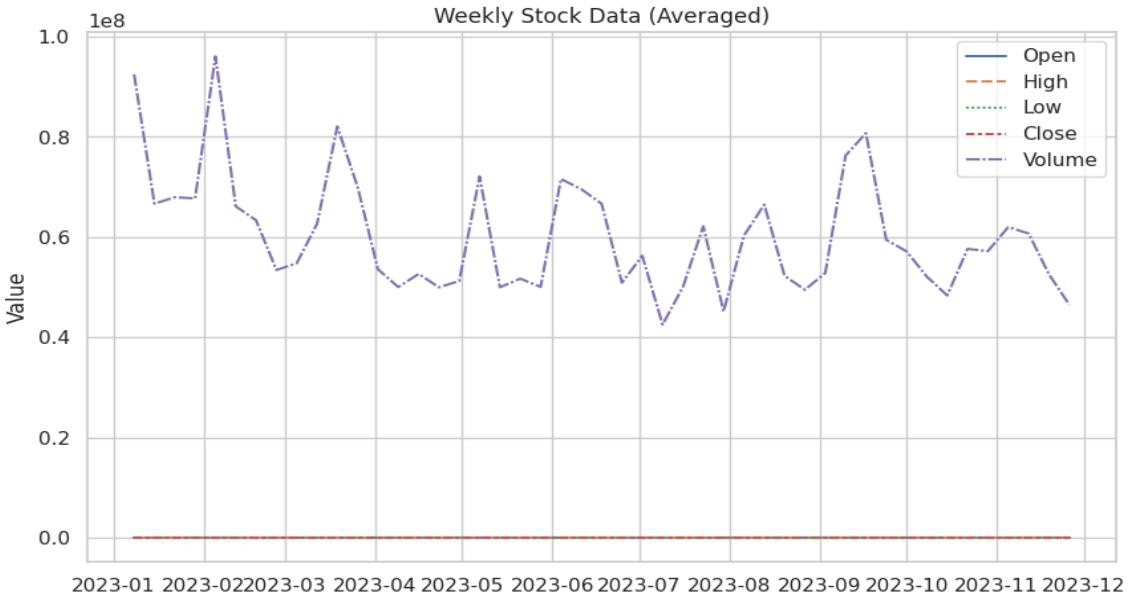
```
[ ] #importing dataset
    b = yf.Ticker(input('Enter ticker name : '))
    a = b.history(period="max")
```

```
Enter ticker name : AAPL
```

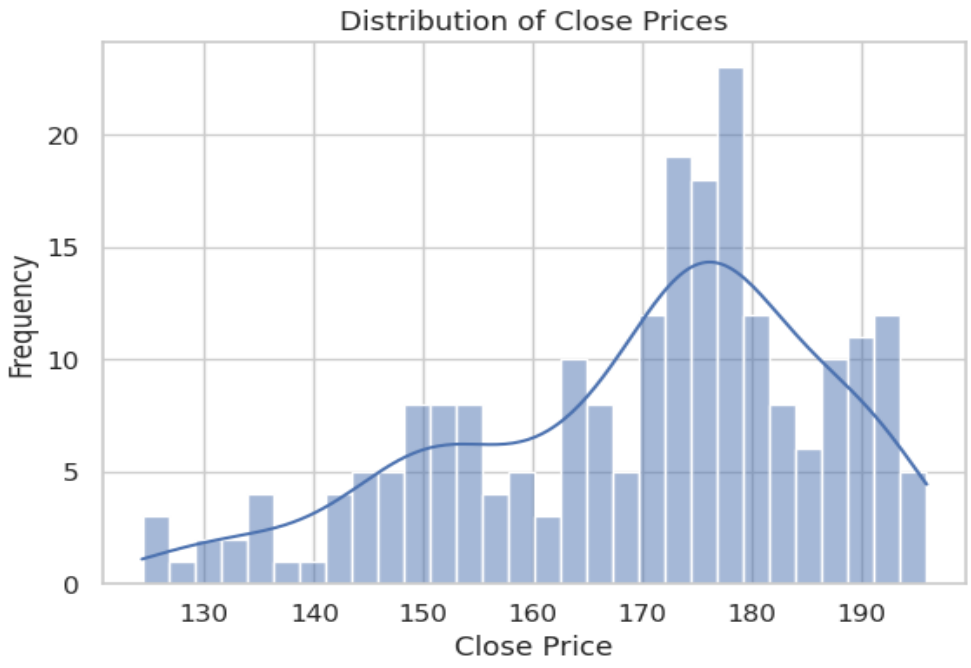
```
start_date = input("Enter start date  ")

data = a[(a.index >= start_date)]

# Print the filtered data
print(f"Data for the ticker number you gave" , b )
print(" ")
print(data)
```

```
Enter start date  2023-01-01
Data for the ticker number you gave yfinance.Ticker object <AAPL>

                                 Open        High         Low       Close  \
Date
2023-01-03 00:00:00-05:00  129.555841  130.172390  123.479803  124.374802
2023-01-04 00:00:00-05:00  126.184691  127.944857  124.384755  125.657639
2023-01-05 00:00:00-05:00  126.423361  127.059803  124.066539  124.325089
2023-01-06 00:00:00-05:00  125.309594  129.565795  124.195816  128.899521
2023-01-09 00:00:00-05:00  129.744788  132.668449  129.168010  129.426559
...                               ...         ...         ...         ...
2023-11-14 00:00:00-05:00  187.699997  188.110001  186.300003  187.440002
2023-11-15 00:00:00-05:00  187.850006  189.500000  187.779999  188.009995
2023-11-16 00:00:00-05:00  189.570007  190.960007  188.649994  189.710007
2023-11-17 00:00:00-05:00  190.250000  190.380005  188.570007  189.690002
2023-11-20 00:00:00-05:00  189.889999  191.910004  189.880005  191.449997

                              Volume  Dividends  Stock Splits
Date
2023-01-03 00:00:00-05:00  112117500        0.0           0.0
2023-01-04 00:00:00-05:00   89113600        0.0           0.0
2023-01-05 00:00:00-05:00   80962700        0.0           0.0
2023-01-06 00:00:00-05:00   87754700        0.0           0.0
2023-01-09 00:00:00-05:00   70790800        0.0           0.0
...                              ...        ...           ...
2023-11-14 00:00:00-05:00   60108400        0.0           0.0
2023-11-15 00:00:00-05:00   53790500        0.0           0.0
2023-11-16 00:00:00-05:00   54412900        0.0           0.0
2023-11-17 00:00:00-05:00   50922700        0.0           0.0
2023-11-20 00:00:00-05:00   46505100        0.0           0.0

[223 rows x 7 columns]
```

```
sns.set(style="whitegrid")
plt.figure(figsize=(10, 6))
sns.lineplot(data=weekly_data)
plt.xlabel('Week')
plt.ylabel('Value')
plt.title('Weekly Stock Data (Averaged)')
plt.show()
```




```
sns.histplot(df['Close'], bins=30, kde=True)
plt.title('Distribution of Close Prices')
plt.xlabel('Close Price')
plt.ylabel('Frequency')
plt.show()
```



7 Source Code

Predicting whether to buy or sell the choosen stock using Linear Regression Model

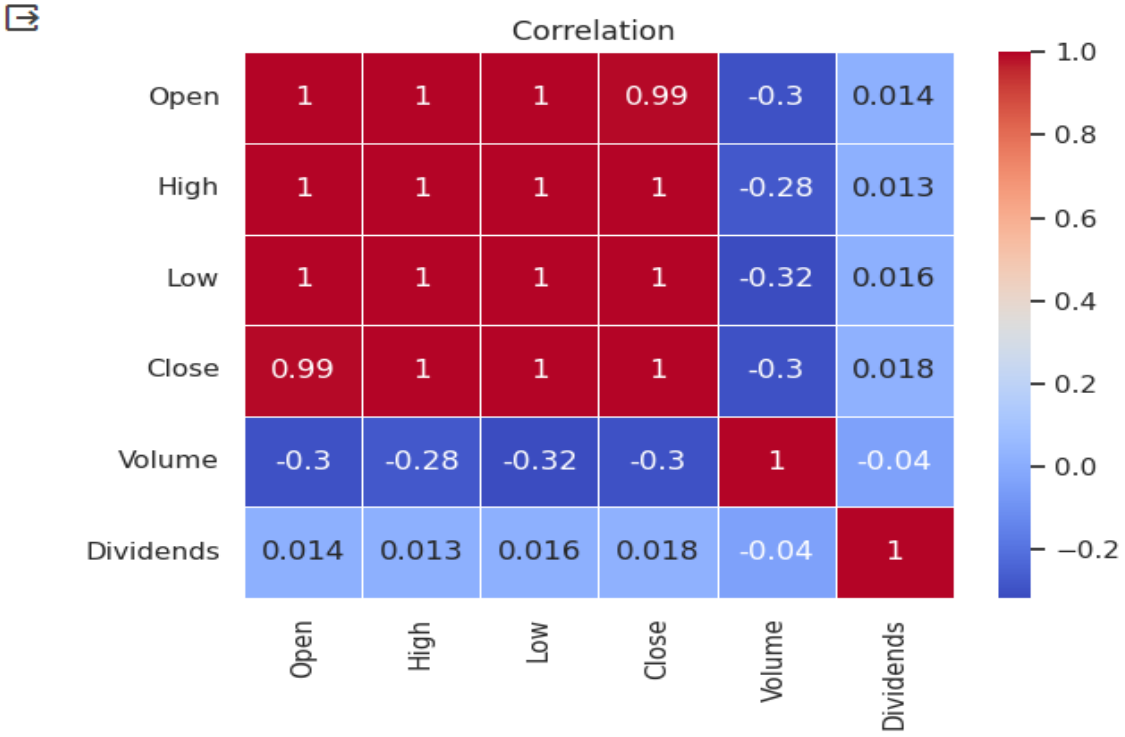
```
[31] df['Next_day_Open'] = df['Open'].shift(-1)
     # Drop the last row as it will have NaN for 'Next_day_Open'
     df = df.dropna()
```

```
X = df[['Open']]  # Features
y = df['Next_day_Open']  # Target
# Split the data into training and test sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
```

```
[33] model = LinearRegression()
     model.fit(X_train, y_train)
```

```
▾ LinearRegression
LinearRegression()
```

```
correlation = df.corr()
sns.heatmap(correlation, annot=True, cmap='coolwarm', linewidths=.5)
plt.title('Correlation ')
plt.show()
```



8 Future Scope

The future scope of stock market analysis includes increased use of machine learning and AI, advancements in quantitative finance, integration of blockchain technology, reliance on alternative data sources, automation with robo-advisors, emphasis on cybersecurity, and exploration of decentralized finance (DeFi). Additionally, there will likely be a continued focus on global economic factors, ESG criteria, and educational platforms for democratizing access to stock market insights.

9 Conclusion

In conclusion, the evolving landscape of stock market analysis presents exciting opportunities for leveraging advanced technologies and data-driven strategies. The integration of machine learning, blockchain, and big data analytics is poised to enhance predictive modeling and decision-making. The increasing role of alternative data sources and automation tools reflects a shift towards more comprehensive and efficient market insights. As the financial industry embraces decentralized finance and emphasizes factors like ESG criteria, the future holds a dynamic and interconnected ecosystem. However, it's crucial to adapt to regulatory changes, prioritize cybersecurity, and promote financial education to ensure a well-informed and resilient approach to stock market analysis in the years to come.

10 References

https://finance.yahoo.com/

THANK YOU