Stock Price Prediction and Investment Recommendations through Machine Learning Analysis

Author:

Hasibul Islam

Dept. of Computer Science & Engineering

Jahangirnagar University

Dhaka, Bangladesh

devhasibulislam@gmail.com

Sadia Khanum

Dept. of Computer Science & Engineering

City University

Dhaka, Bangladesh

Co-Author:

Wary Hossain Rabby

&

MD Rafiqul Islam

Dept. of Computer Science & Engineering

Jahangirnagar University

Dhaka, Bangladesh

A.S.S.M.Q-E-Elahy

Department of Applied Statistics and Data Science

Jahangirnagar University,

Dhaka, Bangladesh

Gias Uddin

Department of Computer Science & Engineering

Uttara University,

Dhaka, Bangladesh

Md. Emran Sikder

Department of Computer Science & Engineering

Daffodil International University,

Dhaka, Bangladesh

ACKNOWLEDGEMENT

ABSTRACT

We're researching how our thesis can help guess stock prices and suggest smart investment moves. We start by checking if the current stock prices are right, looking at both the percentage and money differences. We also predict the prices tomorrow, showing the real-time and guessed numbers and explaining how much they differ. After that, we give practical advice in three categories: Sell, Hold, and Buy, so people can make smart choices. We also look at what happens if the stock prices are guessed wrong and how it affects people's investment portfolios.

Chapter 1

Introduction

1.1 Introduction

In our thesis, firstly we check if current stock prices are right, looking at the percentage and money differences. Then, we try to guess what the prices will be tomorrow, comparing real-time and guessed numbers and explaining the differences. Our suggestions are simple: Sell, Hold, or Buy, to help investors make smart choices. We also look at what happens if our guesses are wrong and how it affects people's investments.

In this chapter, the motivation behind the research is introduced. After that, we will present the objectives of our thesis. This thesis includes the significance of the problem and problem statement in detail. Then we will present the contributions and significance of the statements. The chapter ends with a short description of the organization of the thesis.

1.2 Motivation

We need to know how our money in stocks can go up or down, both in percentage and actual value. Trying to guess what the stock price will be tomorrow includes looking at trends and how the market is doing. When deciding whether to sell, keep, or buy stocks, it depends on things like how the market is changing and what you want to achieve with your investments. Also, if we make a mistake in predicting, it can lead to losing money. So, it's smart to think carefully, make informed choices, and maybe get advice from experts in handling the ups and downs of the stock market.

1.3 Objectives

- Assess the accuracy of current stock prices.

- Determine correctness in percentage and actual value.

- Predict next-day stock prices in real-time.

- Display and compare predicted and actual prices in percentage and amount.

- Provide clear recommendations for buying shares: Sell, Hold, or Buy.

- Examine the impact of incorrect share prices on the overall investment portfolio.

1.4 Contribution to Knowledge and Statement of Significance

This thesis significantly advances financial knowledge by:

- Enhances understanding of stock price accuracy through machine learning analysis

- Provides valuable insights into next-day price predictions with real-time comparisons.

- Contributes practical recommendations for buying shares at critical stages.

● Highlights the significant impact of incorrect share prices on investment portfolios.

1.5 Thesis Organization

We have divided our thesis into five chapters. In Chapter 1, Introduction and some related works are reviewed. Chapter 2 gives a Literature review that indicates the related work to our thesis. Chapter 3, explains our research methodology and the mechanism of how we work. Chapter 4, presents an analysis of the results of our thesis applied to our dataset and discussion. Chapter 5, represents a Conclusion of our thesis.

Chapter 2

Literature Review

2.1 Introduction

In this part, we will explain similar works, an overview of the research, and some of the research's obstacles. We will cover other study papers and their work's methodology and correctness. We give a summary of stock price analysis around the world. We will go into how we improve the present price accuracy, next-day price prediction and buying recommendations, and impact on the portfolio.

2.2 Related Work

We mentioned some papers related to our work. A Davis, C. K. [1] Exploring the intersection of machine learning, quantitative portfolio choice, and mispricing in financial markets. This abstract highlights the potential of advanced algorithms to identify mispriced assets and their impact on optimizing portfolio selection strategies, offering valuable insights for investors and researchers alike. Gu, A., Viens, F.G. and Yi, B. [2] This topic explores ideal risk-sharing and investment approaches for insurers facing mispricing and uncertainty in models, enhancing financial stability and maximizing returns. Tu, J. and Zhou, G. [3] This research explores the integration of economic objectives into Bayesian priors, addressing parameter uncertainty in portfolio choice, and offering valuable insights for decision-making in financial contexts. Ang, A., Papanikolaou, D. and Westerfield, M.M. [4] The thesis explores how people make investment choices, considering illiquid assets that are harder to sell. It shows that uncertainty about the duration of illiquidity increases risk aversion, leading to reduced allocation in both liquid and illiquid assets. Investors are willing to sacrifice 2% of their wealth to hedge against rare illiquidity crises. Ben-David, I., Drake, M.S. and Roulstone, D.T. [5] This study examines how companies make acquisition decisions based on investor perceptions of over or undervaluation (measured using short interest). Overvalued firms are 54% more likely to acquire other companies using their stock, while undervalued ones perform better in cash acquisitions. Misvaluation influences merger strategies and outcomes. Cvitanić, J., Lazrak, A., Martellini, L., and Zapatero, F. [6] This study explores how to make the best investment choices when we don't know all the information, and shows that learning about expected returns can significantly improve investment decisions. It finds that following analysts' advice is not very helpful in making profitable investments. Liu, J. and Longstaff, F.A. [7] How a cautious investor should invest their money when there's a chance to make easy profits through arbitrage (buying and selling assets to take advantage of price differences). Doukas, J.A., Kim, C.F. and Pantzalis, C. [8] Stocks with higher risk often have bigger price differences due to the challenges arbitrageurs face in trading them. Sørensen, C. [9] Smart investing using stocks and bonds, suggesting zero-coupon bonds for protection. Utilizes meanvariance approach for optimal wealth growth. Stulz, R.M. [10] How people from different countries decide where to invest and how it affects investment returns. It reviews theories, empirical tests, and their significance in international finance. Davis, C.K. [11] Why mispricing can occur in the stock market and how it affects investments and prices.
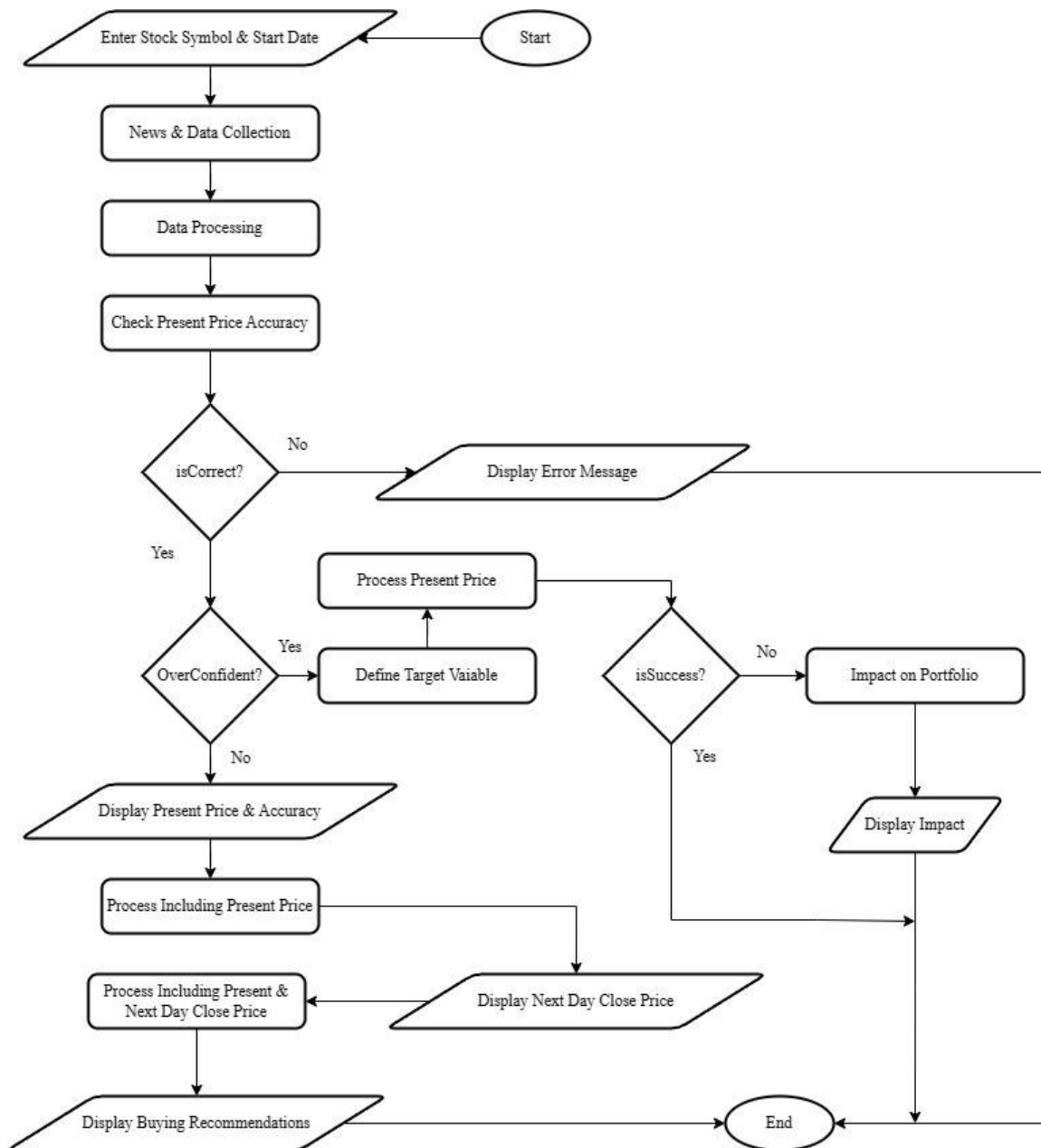
Chapter 3

Research Methodology

3.1 Introduction

In this part, we use Decision Tree Regression, Decision Tree Classification, Gradient Boosting, ARIMA, Random Forest Regression and to analyze the data set. We have also visualized the data with various attribute features.

3.2 Working Procedure

We convert our working procedure into the following:

3.2.1 Data Pre-processing

The data processing represents evaluating the accuracy of present stock prices by determining correctness in percentage and amount. It also encompasses predicting next-day prices in real-time, comparing actual and predicted values. Additionally, it includes processing data to provide practical recommendations for buying shares at different stages and assessing the impact of incorrect share prices on investment portfolios.

3.2.2 Model Development and Validation

● Decision Tree Regression

  o Used for Predict Current Price Accuracy of a share where we got 98.46% accuracy and 1.54% loss by 70% data for training, 15% data for testing & 15% data for validation.

● Random Forest Regression

  o Used for Predict Current Price Accuracy of a share where we got 99.01% accuracy and 0.99% loss by 80% data for training, 10% data for testing & 10% data for validation.

● AutoRegressive Integrated Moving Average (ARIMA)

  o Used for Predict Next Day Close Price of a share where we got 98.19% accuracy and 1.54% loss by 70% data for training, 15% data for testing & 15% data for validation.

● Gradient Boosting

  o Used for Predict Stock Buying Recommendation of a share where we got recommendation like Buy/Sell/Hold by 70% data for training, 15% data for testing & 15% data for validation.

● Decision Tree Classification

Fig: 3.2.1 Methodology Flowchart

  o Used for Predict Portfolio Impact from a share where we got 99.17% accuracy and 0.83% loss by 80% data for training, 10% data for testing & 10% data for validation.

3.3 Comparison Table

We carefully studied nine thesis papers and found that ours is better because it gives more efficient results. This means our way of doing things and what we discovered are really important. It makes our thesis stand out and adds something valuable to what others have already done.

Authors	Works	Methods	Datasets	Accuracy
Cvitanić, J., Lazrak, A., Martellini, L. and Zapatero, F.	Dynamic stock buying recommendation	SVMs, ANNs & KNN	Data.gov & Interactive Brokers	48.0%

(2006)				
Gu, A., Viens, F.G. and Yi, B. (2017)	Reinsurance and Investment strategies	Neural Networks & SVM	Tingo & Quandl	87.012%
Davis, C. K. (2022)	Portfolio Choice & Mispricing	Linear Regression	AlphaQuerry	98.054%
Liu, J. and Longstaff, F.A. (2004)	Losing money on arbitrage	LSTM, GARCH & Genetic Algorithm	Interactive Brokers & Data.gov	40.0%
Doukas, J.A., Kim, C.F. and Pantzalis, C. (2010)	Arbitrage risk and stock mispricing	LSTM & SVM	Interactive Brokers	62.082%
Davis, C.K. (2022)	Misprediction on next close price	HMM & PCA	Alpha Vantage & Yahoo Finance	98.50%
Cvitanić, J., Lazrak, A., Martellini, L. and Zapatero, F. (2006)	Dynamic stock buying recommendation	SVMs, ANNs & KNN	Data.gov & Interactive Brokers	48.0%
Ang, A., Papanikolaou, D. and Westerfield, M.M. (2014)	Illiquid assets pose	GBM	Tingo & Data.gov	72.004%
Ben-David, I., Drake, M.S. and Roulstone, D.T. (2015)	Mispricing in short interest rates	RNN, LSTM & SVM	Interactive Brokers	82.082%
Our Study	Stock price forecast with recommendations & impact	DTR, RFR, ARIMA, GB & DTC	Alpha Vantage, IEX Cloud & Yahoo & Google	99.025%

			Finance	

Table: Comparison Table

Chapter 4

Result Analysis and Discussion

4.1 Introduction

First, we will discuss how and where we have collected data. After that, we explain the dataset that we used and explain how we set up the environment for implementing the proposed system. Lastly, we explain the result analysis and discussion then the accuracy of the present price and next-day price prediction and Buying recommendations and impact on the portfolio.

4.2 Data Collection

We have used an existing dataset that has been collected from Real-time data. The name of the dataset is Yahoo Finance (API Token), Alpha Ventage (Secret Key) & IEX Cloud (Public Key & Private Key).

4.3 Data Preprocessing

● Handle missing values.

● Calculate additional indicators for OHLCV.

● Calculate relevant financial matrices:

  o Moving averages

  o Relative strength index (RSI)

  o Volatility

● Outliers' normalization

● Data splitting (80% training, 10% testing & 10% validation)

4.4 Dataset Information

In our thesis, these data include:

● Stock Prices:

  o Daily or periodic records of the stock's historical prices.

● Financial Metrics:

  o Information like earnings, revenue, and other financial indicators.

● Technical Indicators:

  o Calculated values based on stock price patterns, helping analyze trends.

11

● Economic Factors:

  o Data on broader economic conditions influencing the stock market.

● Market Sentiment:

  o Analysis of news and social media sentiment impacting market behavior.

● Trading Activity:

  o Volume data represents the number of shares traded.

● Time Series Information:

  o Sequential data reflects how stock prices change over time.

● Binary Labels:

  o Indicators showing if stock prices went up, down, or remained unchanged.

● Company Events:

  o Information on company-specific occurrences affecting stock values.

● External Influences:

  o Factors like global events or geopolitical changes affecting financial markets.

4.5 Experimental Setup

To evaluate the performance and effectiveness of our thesis, we applied several algorithms and models. The thesis was carried out on a computer with Windows 11 and we used Google collab and python programming language.

The experimental setup for our thesis involves several key components:

● Gather historical stock price data, including relevant financial indicators and market data.

● Clean and preprocess the data, handling missing values, normalizing numerical features, and encoding categorical variables.

● Identify key features influencing stock prices through analysis and domain knowledge.

● Divide the dataset into training and testing sets to train the model on one subset and evaluate its performance on another.

● Choose machine learning models suitable for stock price prediction, such as Random Forest, ARIMA, or Decision Trees.

● Fine-tune model parameters for optimal performance using techniques like grid search or random search.

● Train the selected models on the training dataset, allowing them to learn patterns and relationships.

● Assess the models' performance on the testing dataset using appropriate metrics like Mean Absolute Error (MAE) or classification accuracy.

● Develop investment strategies based on model predictions, considering risk tolerance and portfolio optimization.

● Simulate the impact of recommended trades on a portfolio, analyzing returns, and risks.

● Validate the models by applying them to historical data to see how well they would have performed in the past. This part represents how

● Implement risk management strategies to mitigate potential financial losses associated with model predictions.

4.6 Result Analysis

The total experiment analysis has been carried out. Here we have shown the validation result that we achieve when our system predicts present price accuracy and next-day price prediction, and displays buying recommendations and portfolio impact. We have used an existing dataset that has been collected from Real-time data. The name of the dataset is Yahoo Finance (API Token), Alpha Vantage (Secret Key), and IEX Cloud (Public Key & Private Key). We have used some algorithms like ARIMA, Random Forest Regression, Decision Tree Regression, and Decision Tree classification.

4.6.1 For Predicting Present Price Accuracy
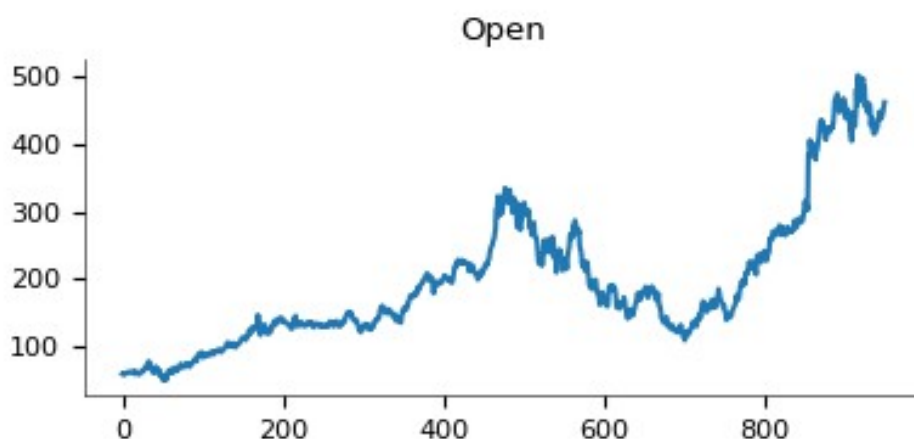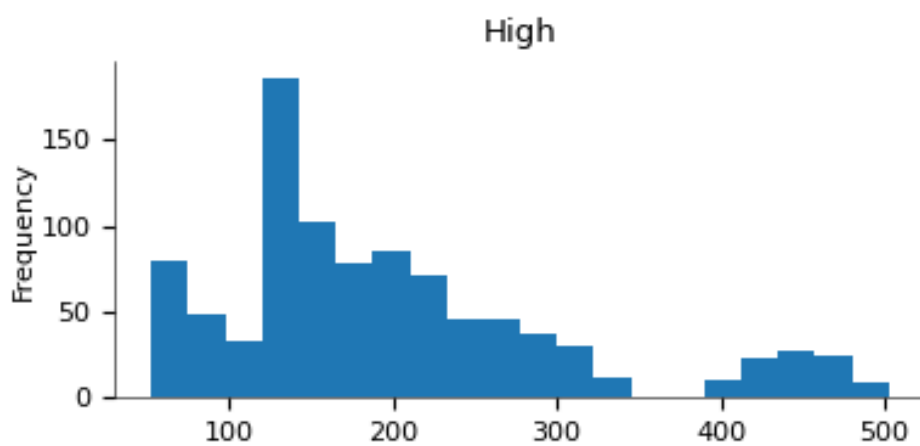

Fig: 4.6.1.1 Open Price

13
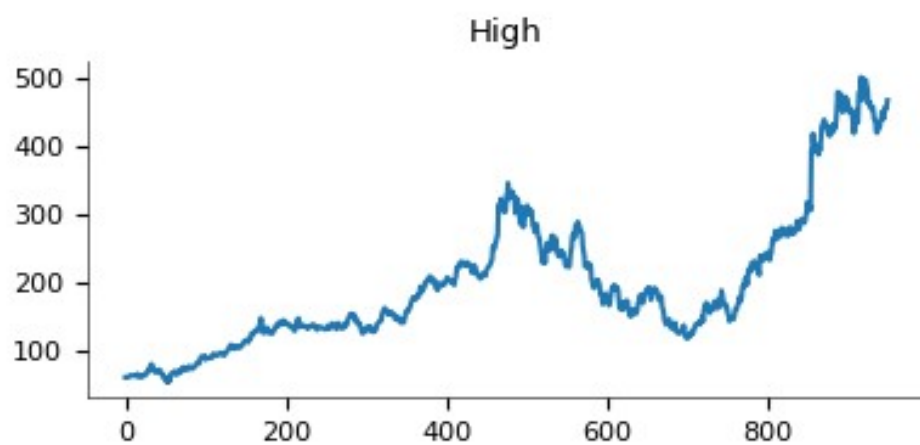
Fig: 4.6.1.1 Highest Price
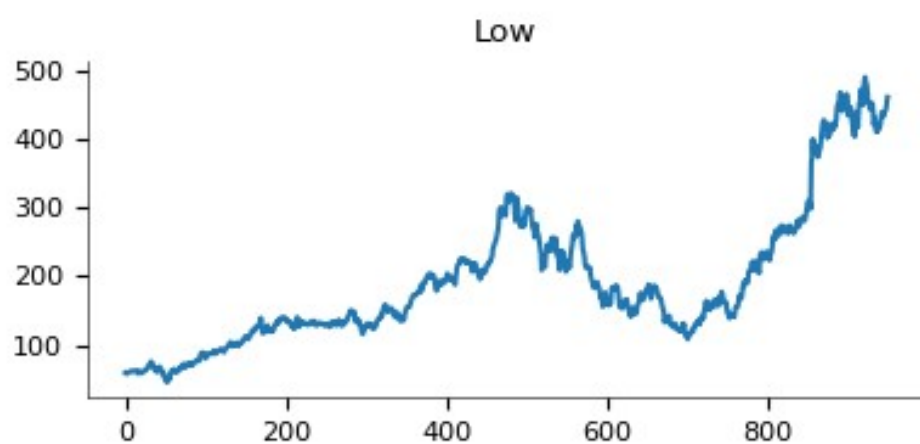
Fig: 4.6.1.2 Lowest Price
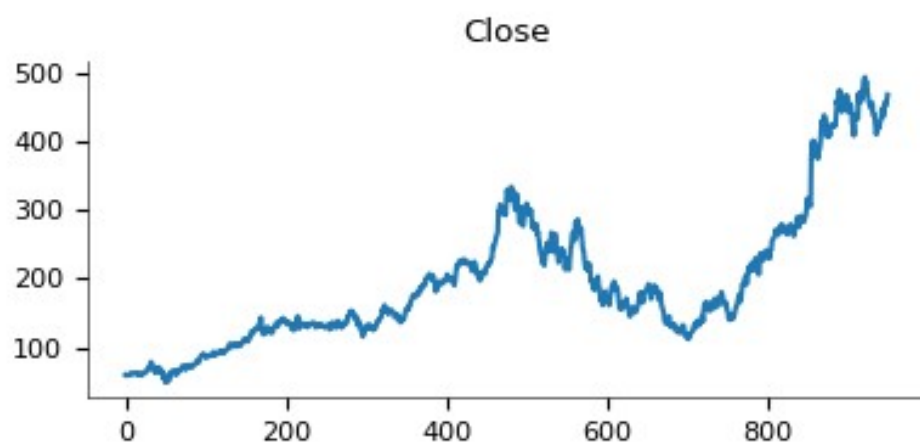
Fig: 4.6.1.3 Close Price
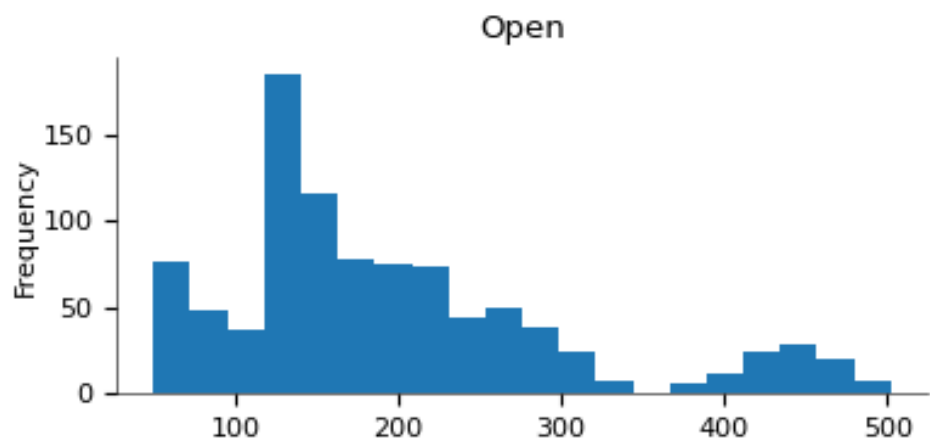
Open

Fig: 4.6.1.4 Open Price
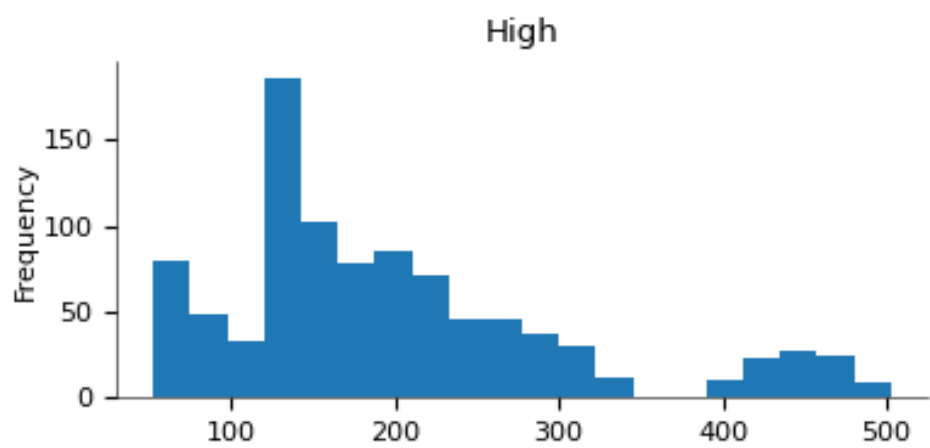
High

Fig: 4.6.1.5 Highest Price

Low
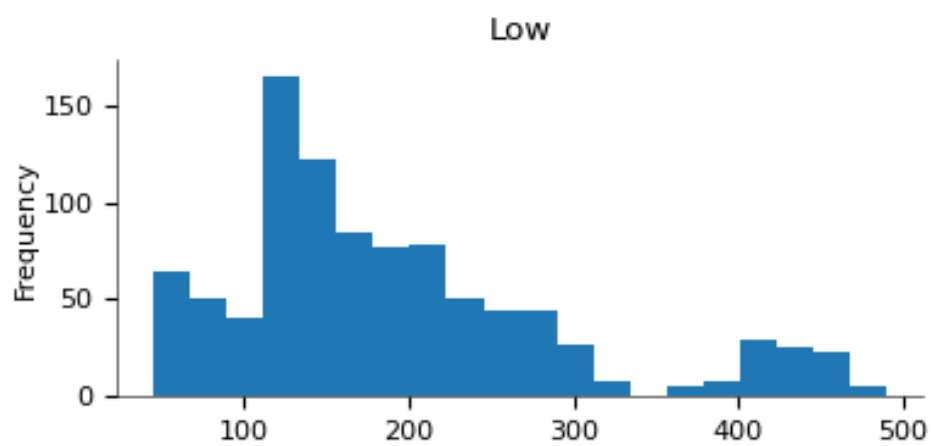
Fig: 4.6.1.6 Lowest Price
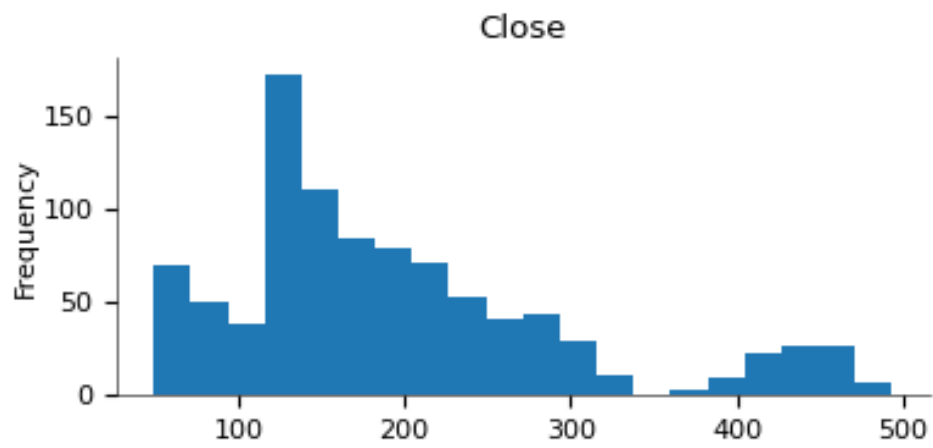
Fig: 4.6.1.7 Close Price
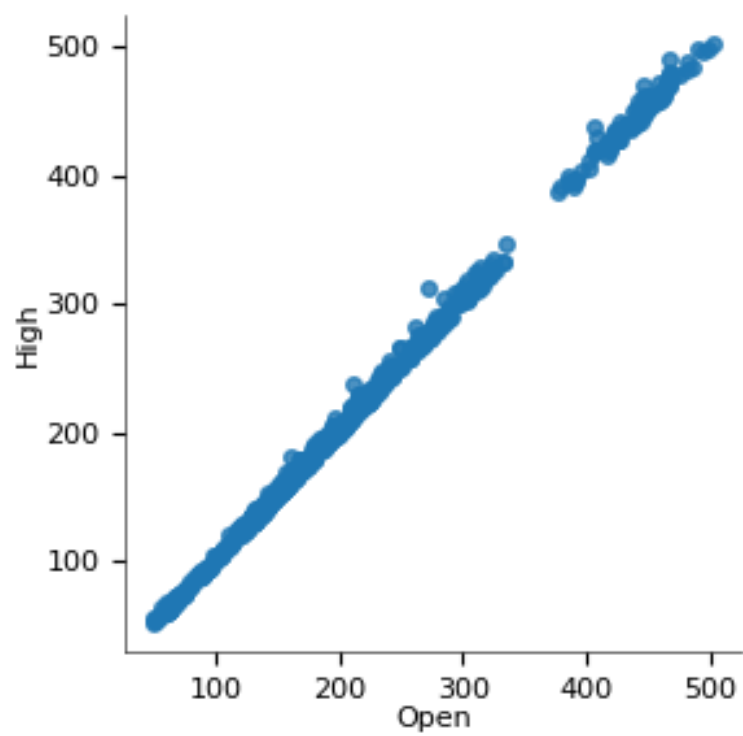
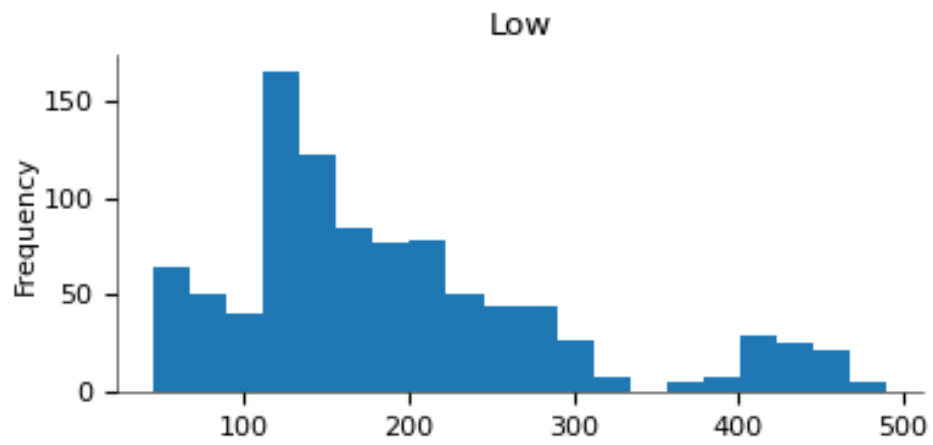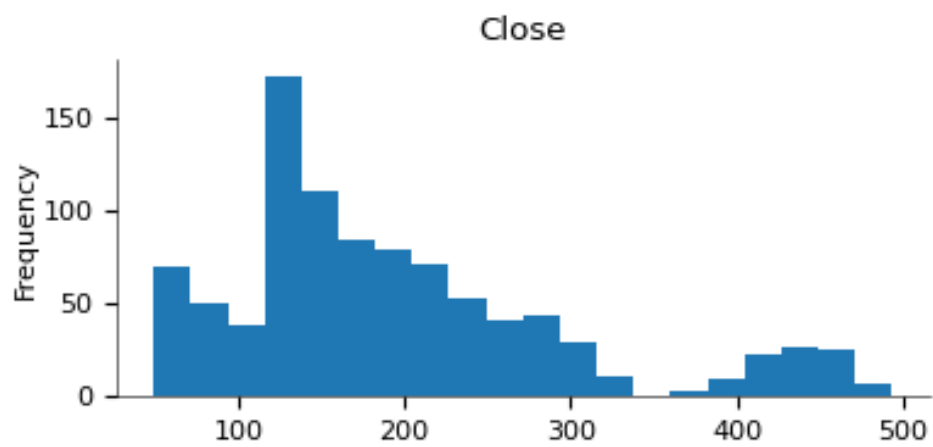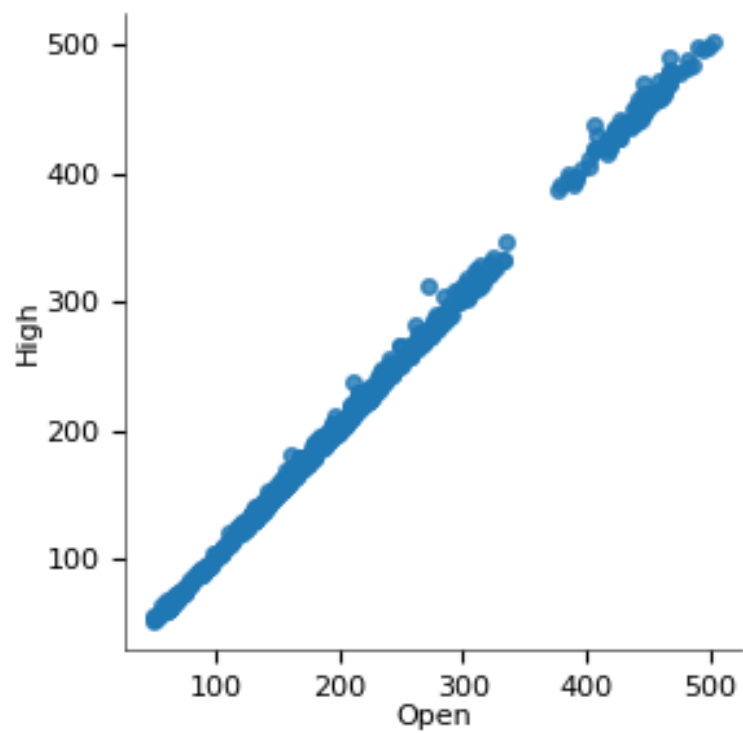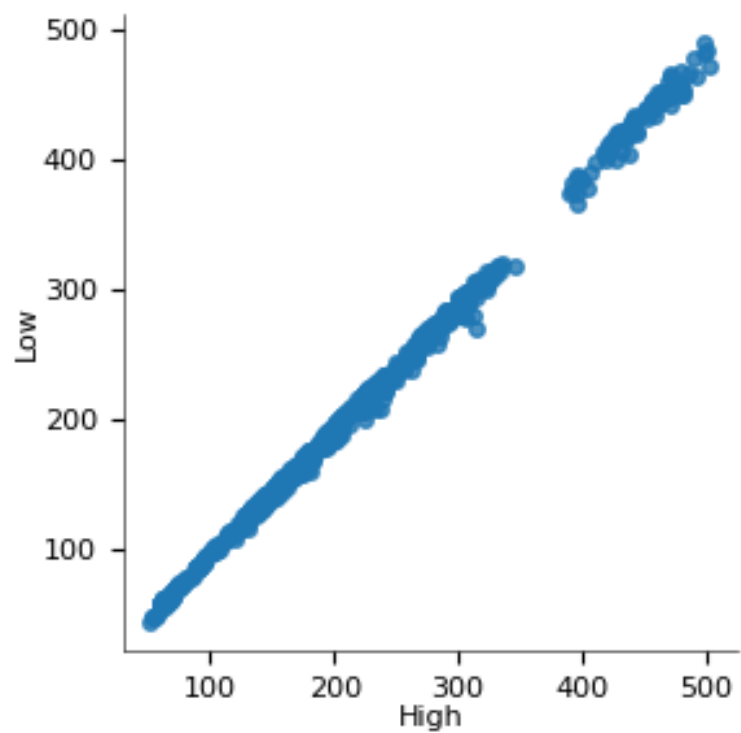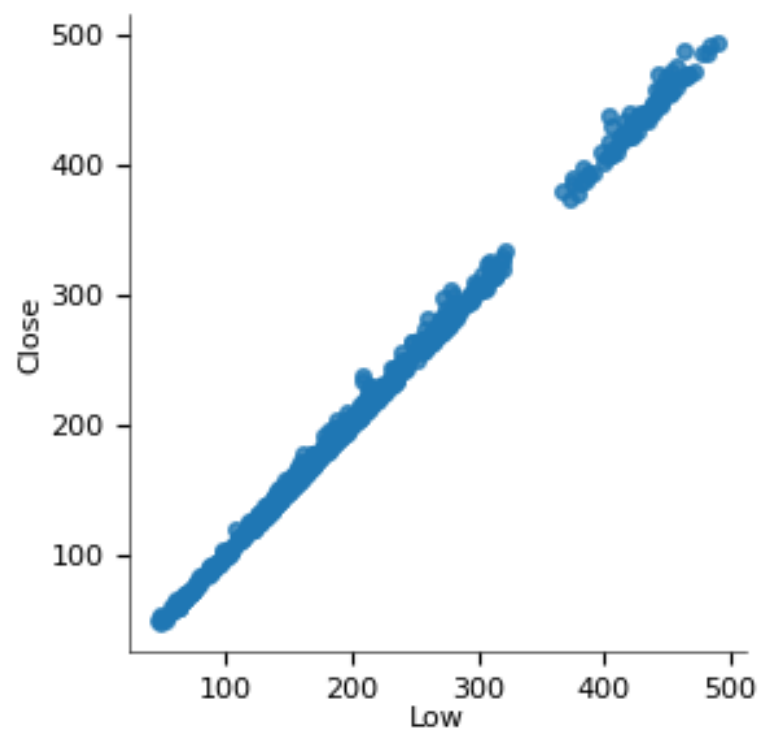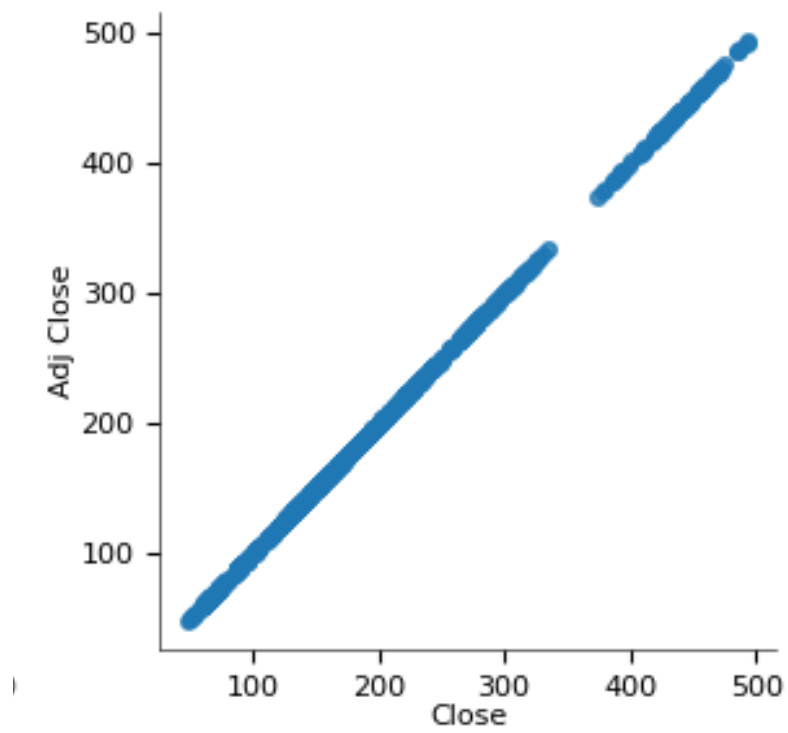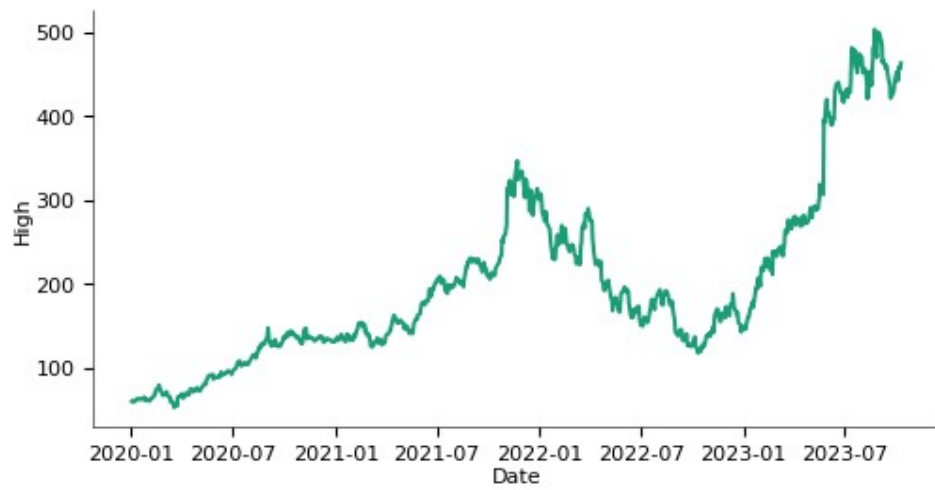

Fig: 4.6.1.8 Open Price
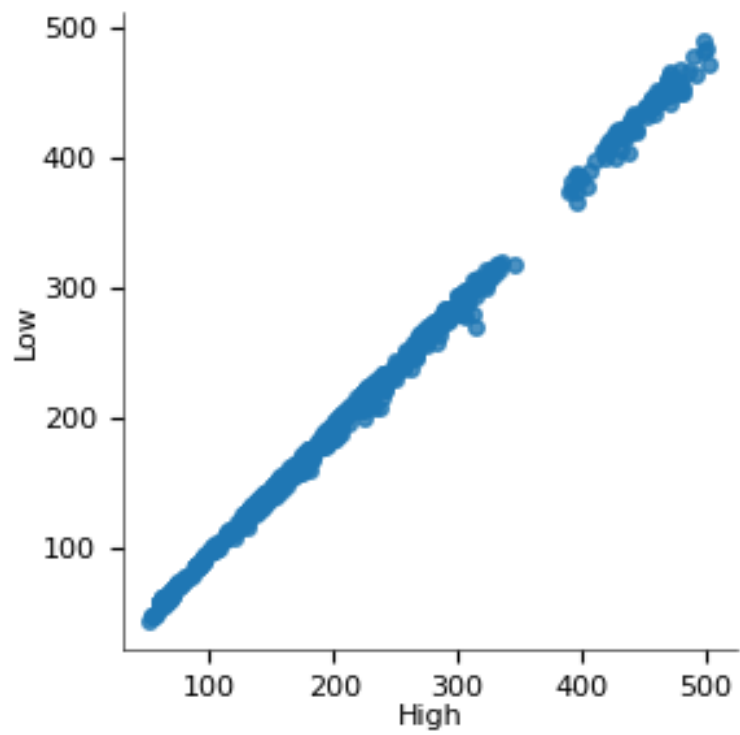
Fig: 4.6.1.9 Highest Price
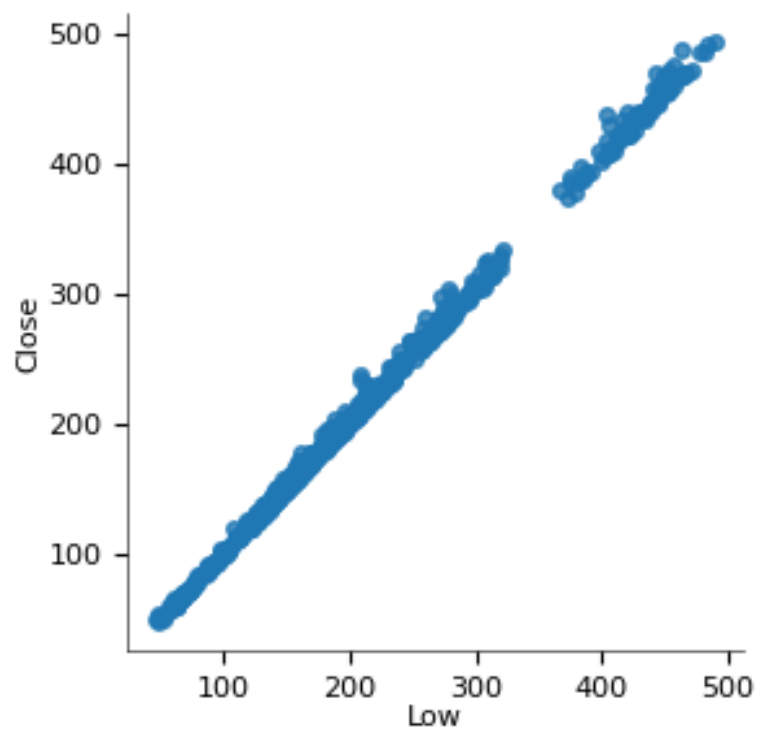
Fig: 4.6.1.10 Lowest Price
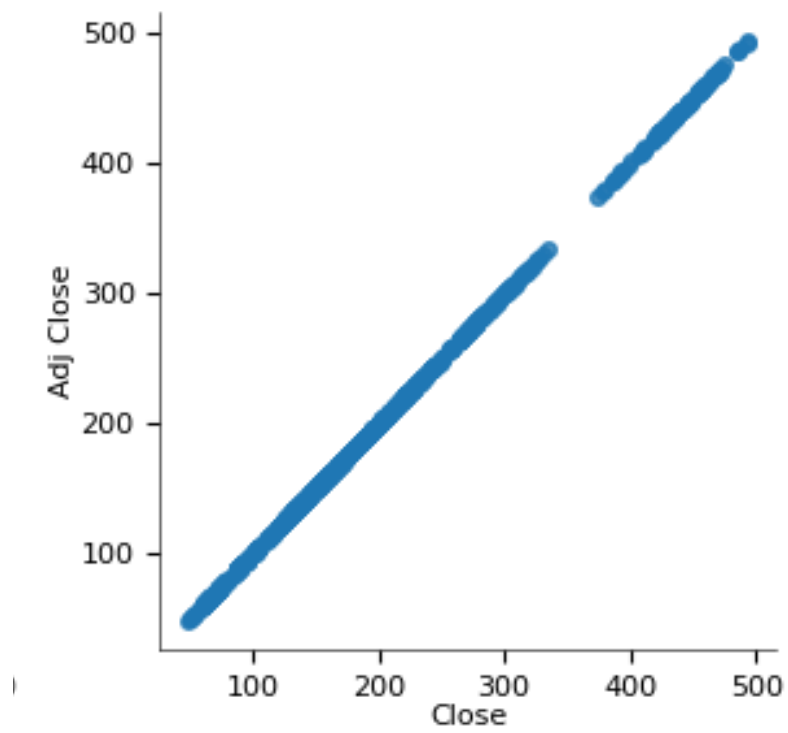
Fig: 4.6.1.11 Close Price
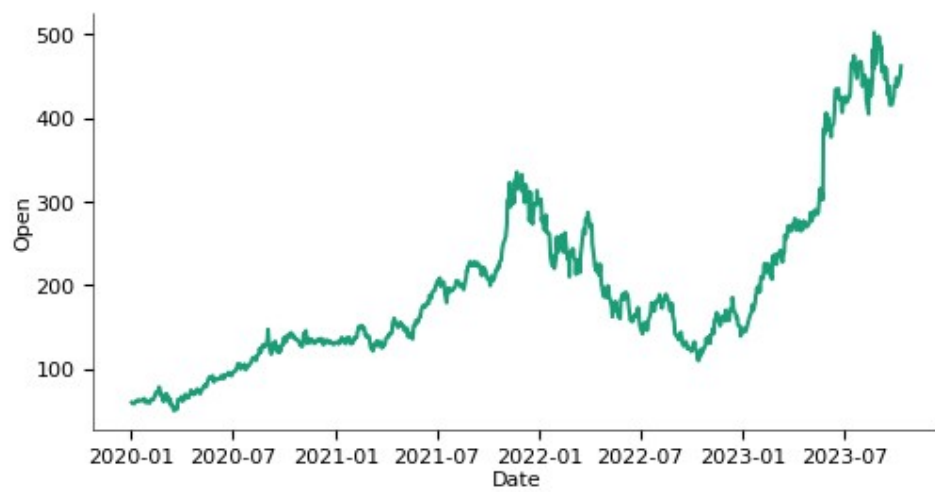
Fig: 4.6.1.12 Open Price
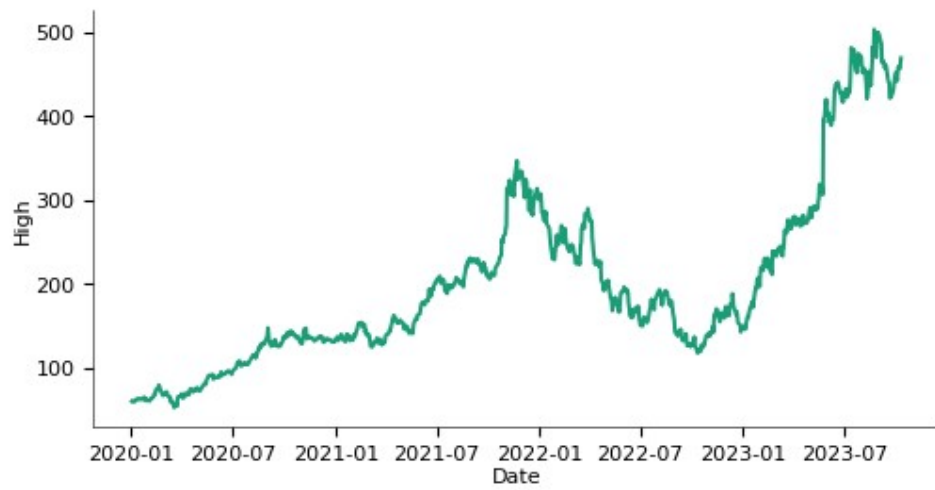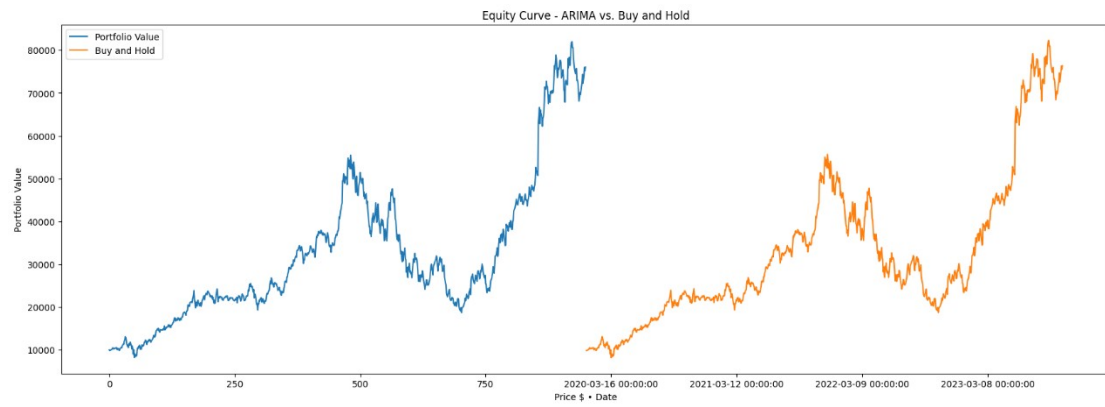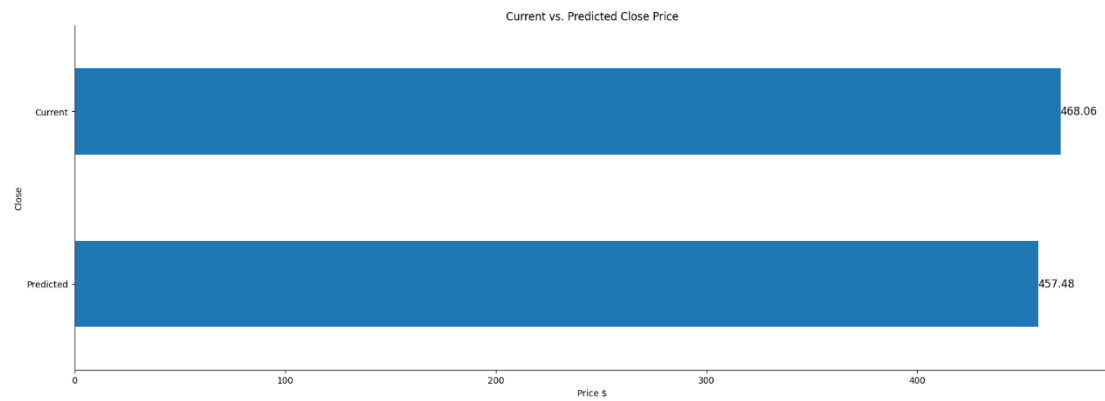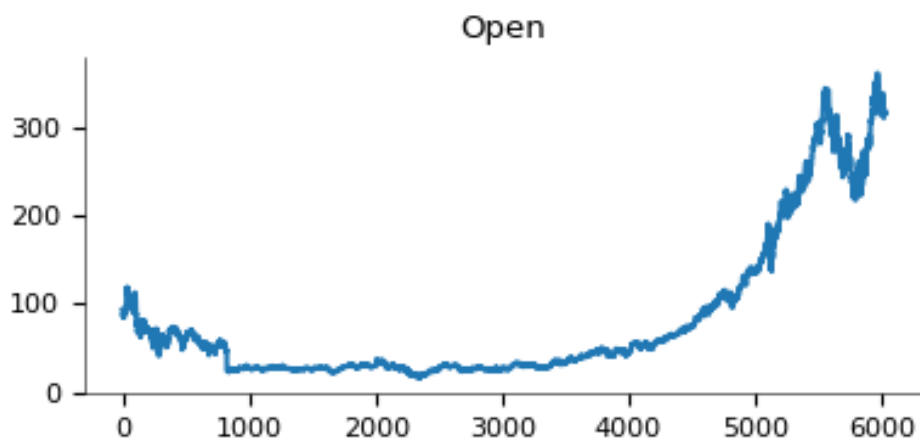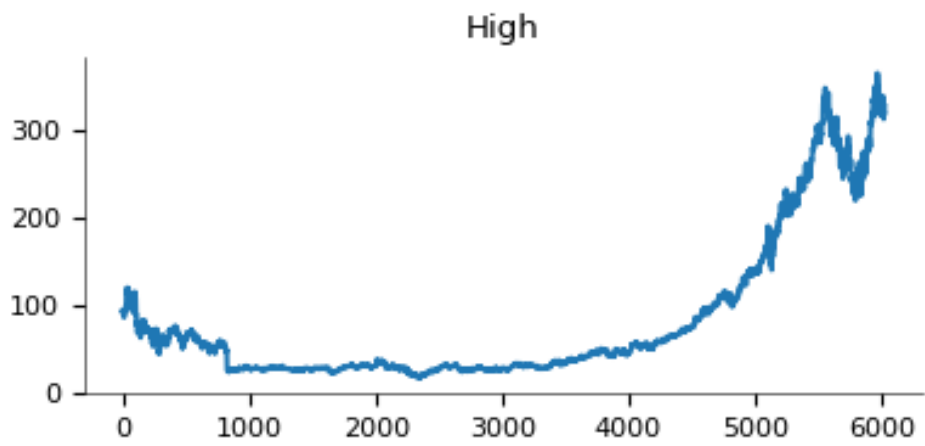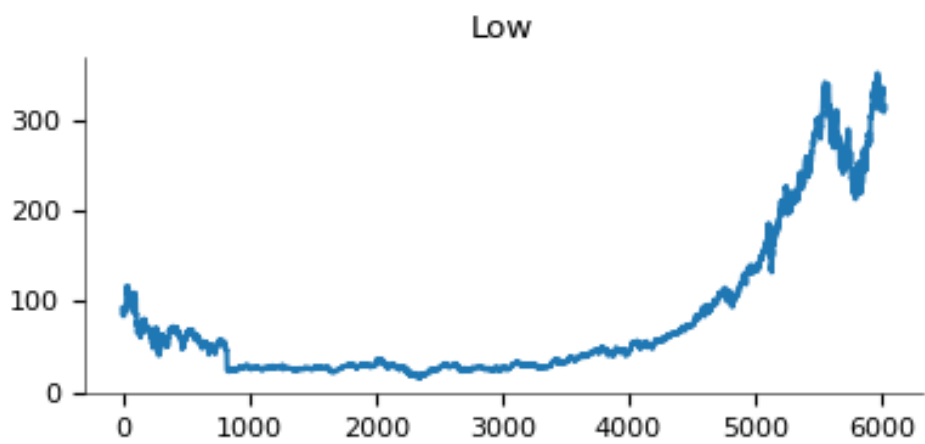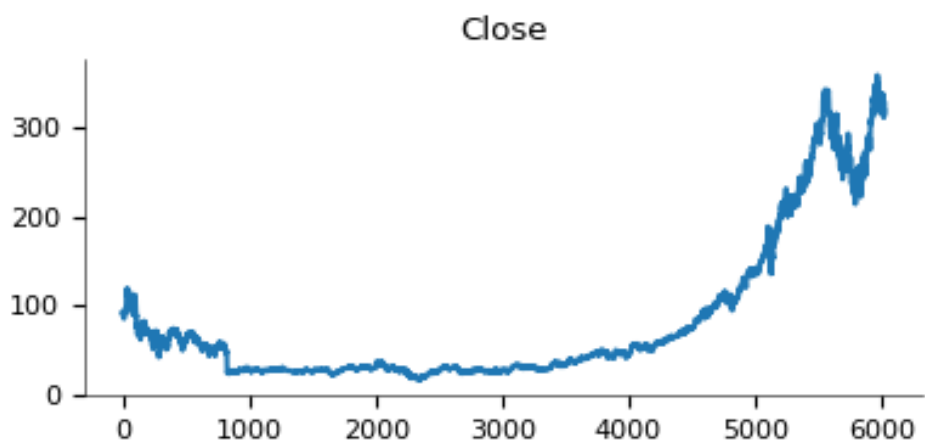
Fig: 4.6.1.13 Highest Price
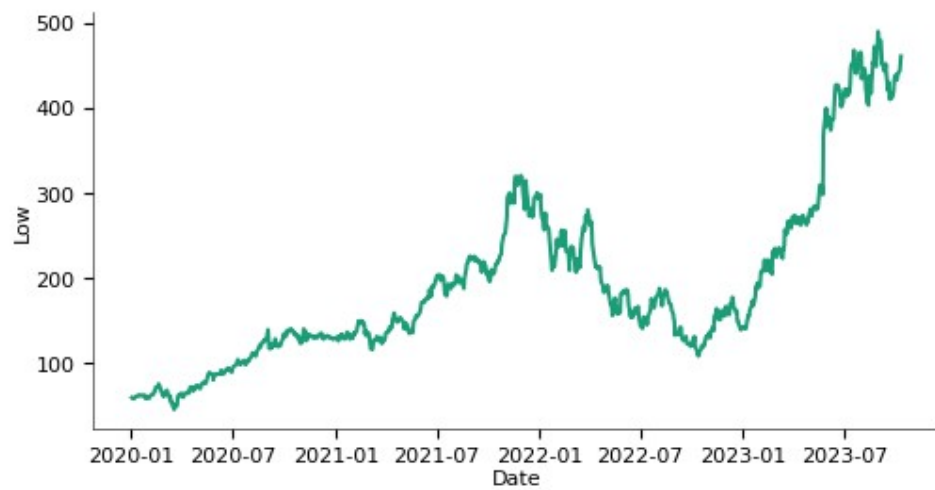
Fig: 4.6.1.14 Lowest Price

Fig: 4.6.1.15 Close Price
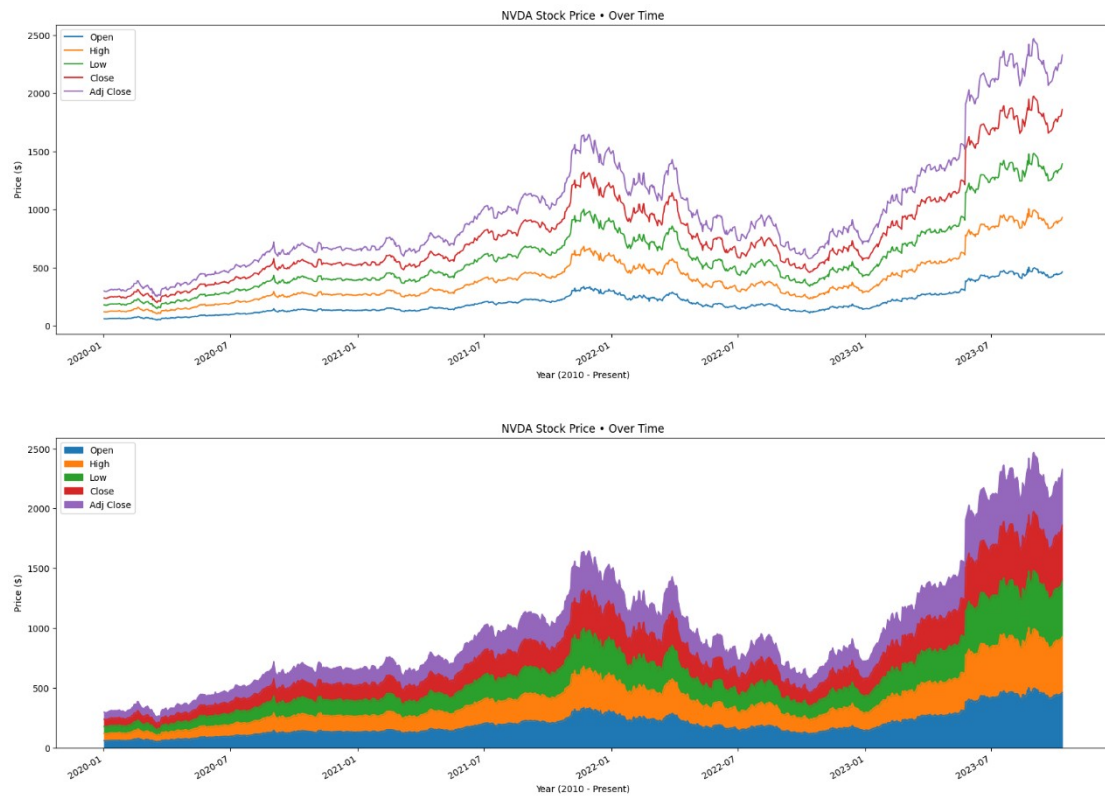
Fig: 4.6.1.16 Fluctuate Over Time Period
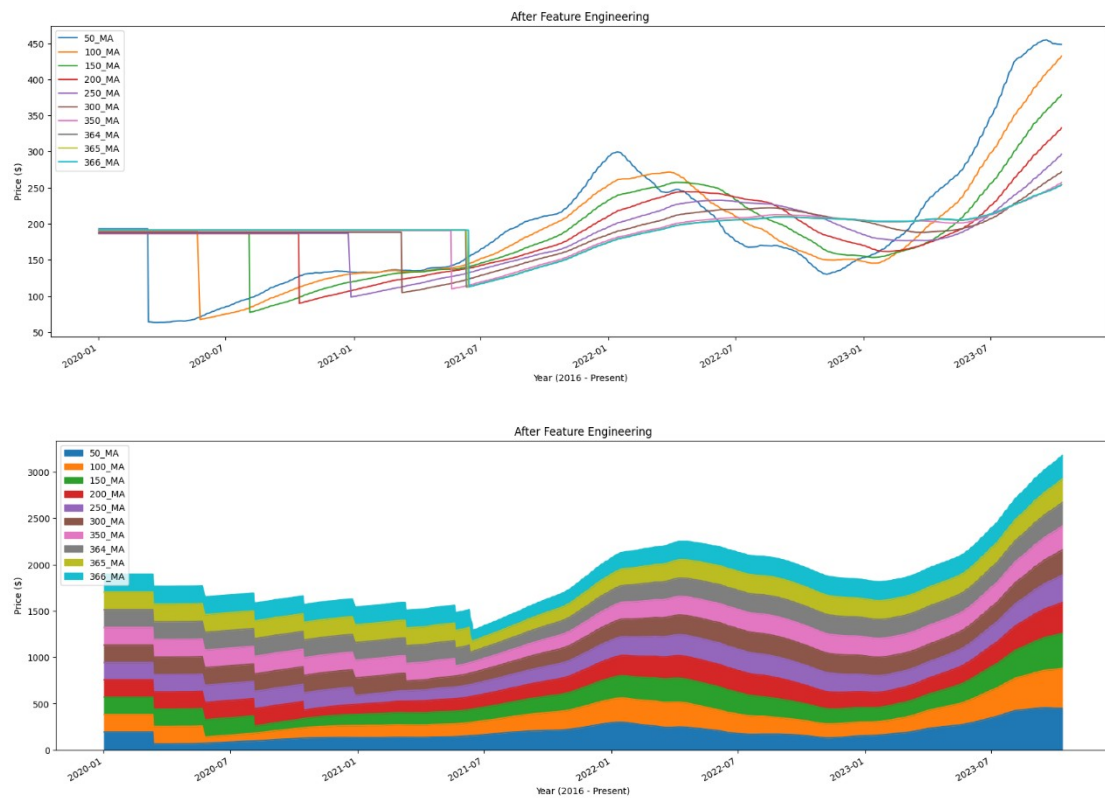
Fig: 4.6.1.17 Fluctuate Over Feature Engineering

Candlestick Chart for NVDA

Fig: 4.6.1.18 Prices in Candlestick Pattern

NVDA's Present Price Accuracy displaying on reducing:

		Finance Name	Finance Price $	Predicted Price $	Reducing Accuracy %
0		Yahoo Finance	468.06	467.792	0.06
1		CNBC Finance	469.55	467.792	0.38
2		Google Finance	468.06	467.792	0.06
3		Bloomberg Finance	468.06	467.792	0.06
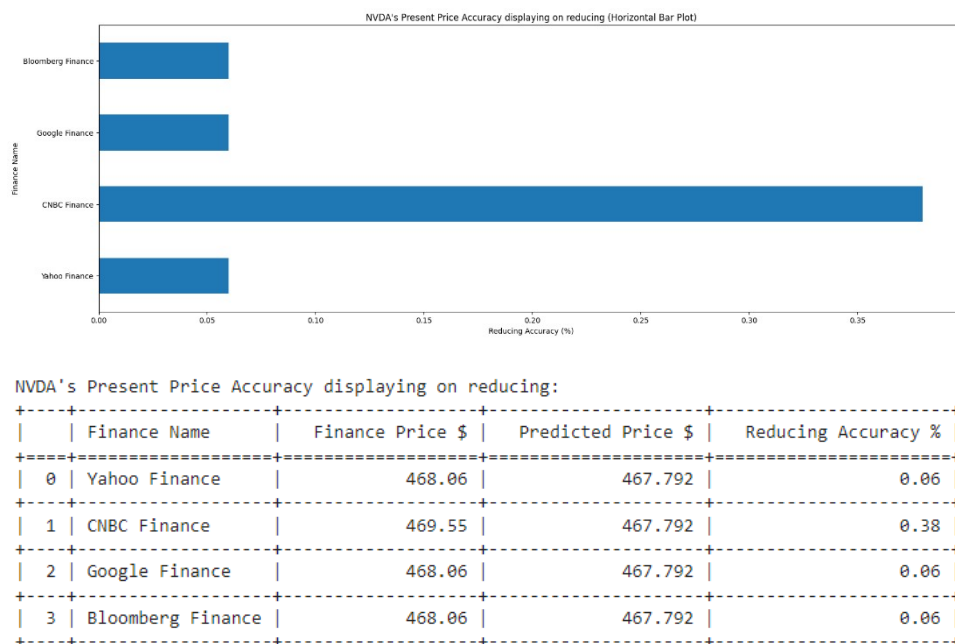
Fig: 4.6.1.19 Outcomes

21

4.6.2 For Predicting Next Day Close Price
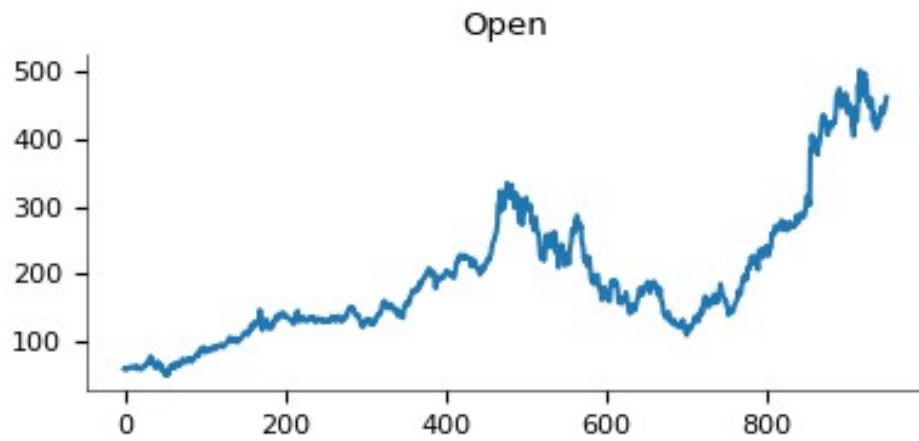
Fig: 4.6.2.1 Open Price
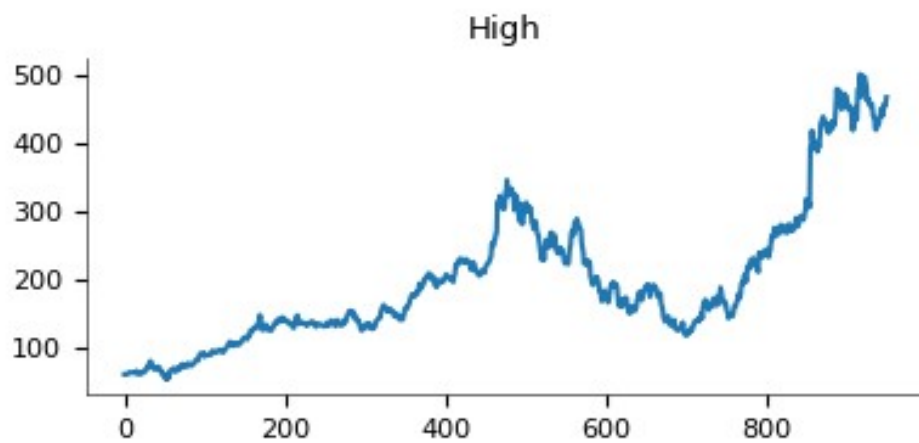
Fig: 4.6.2.2 Highest Price
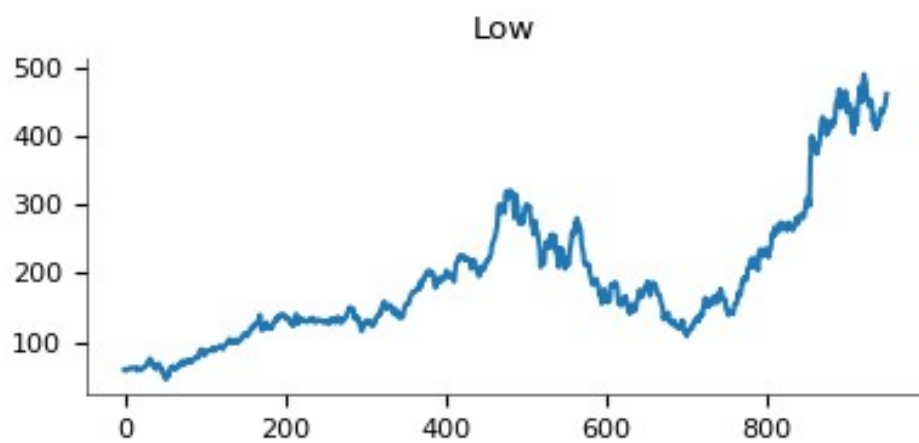
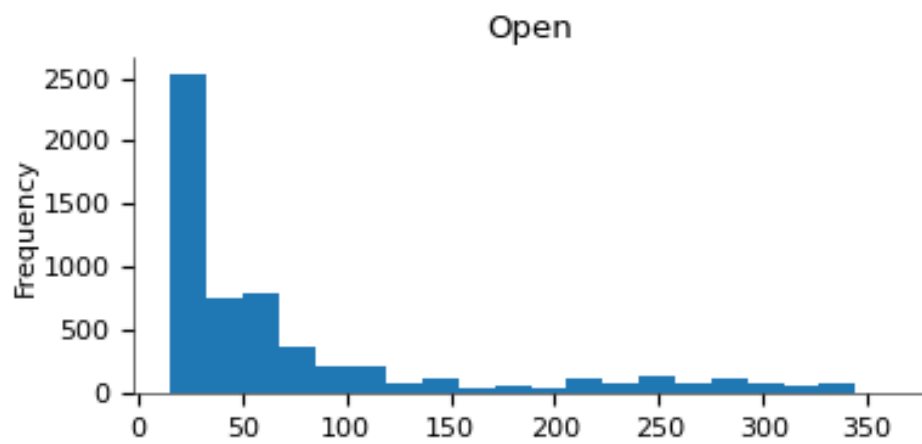Fig: 4.6.2.3 Lowest Price
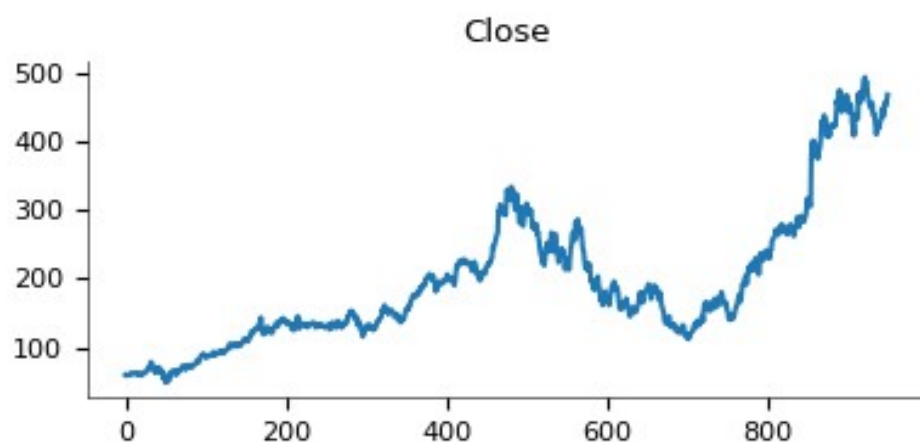
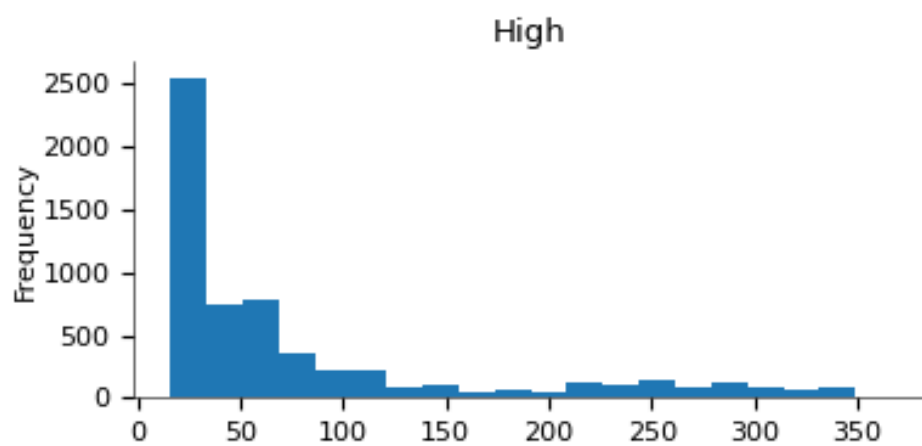Fig: 4.6.2.4 Close Price
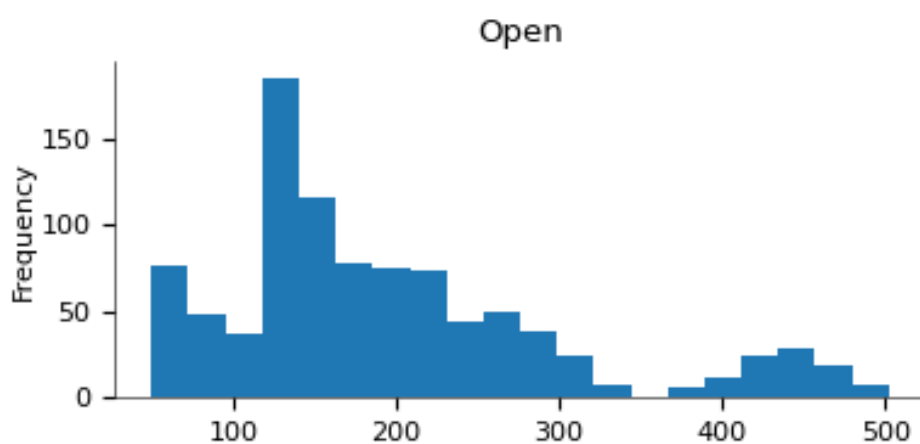
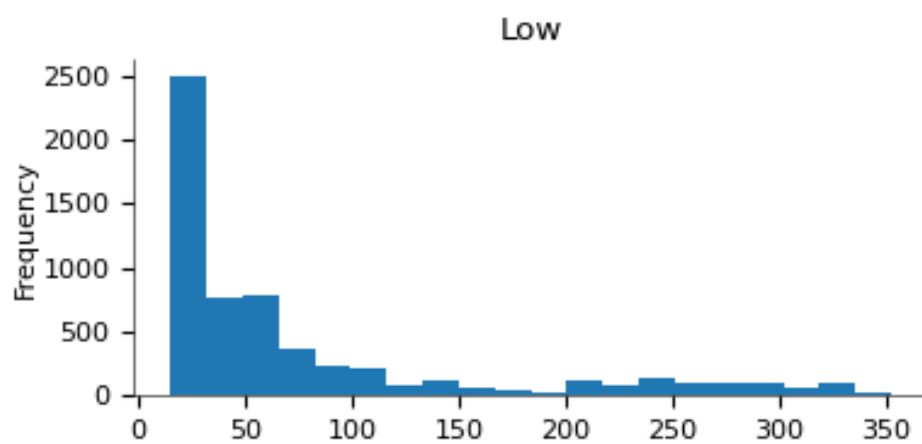Fig: 4.6.2.5 Open Price

Fig: 4.6.2.6 Highest Price
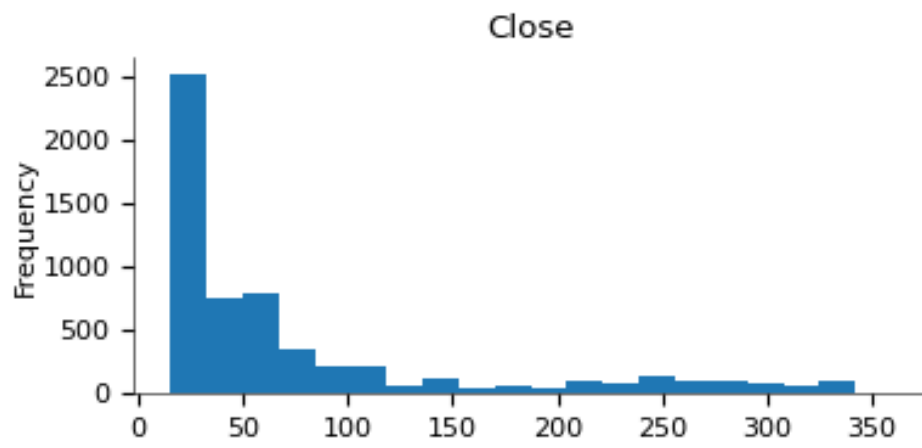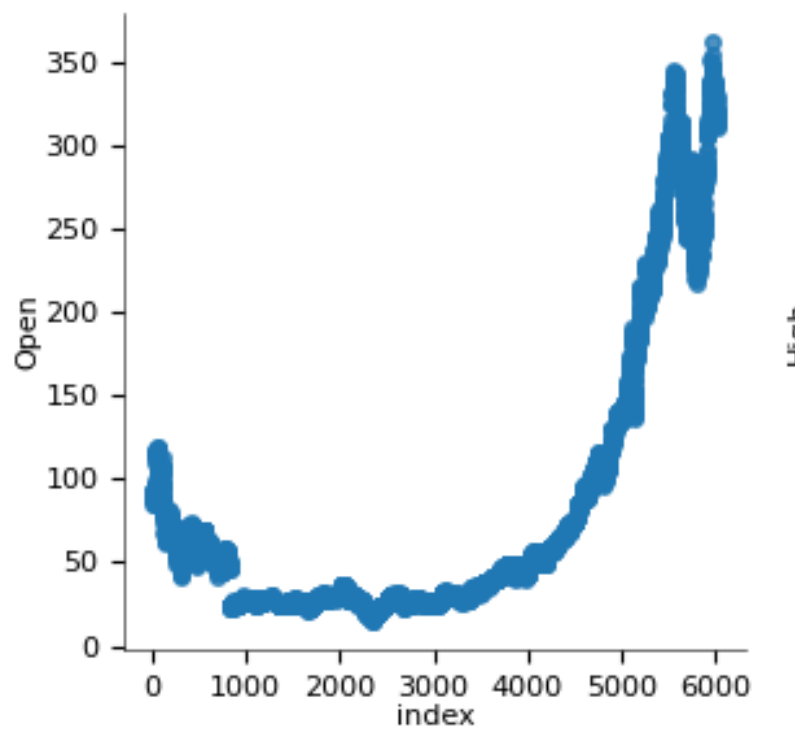
Fig: 4.6.2.7 Lowest Price

Fig: 4.6.2.8 Close Price

Fig: 4.6.2.9 Open Price

Fig: 4.6.2.10 Highest Price

Fig: 4.6.2.11 Lowest Price

Fig: 4.6.2.12 Close Price

Fig: 4.6.2.13 Open Price

Fig: 4.6.2.14 High Price
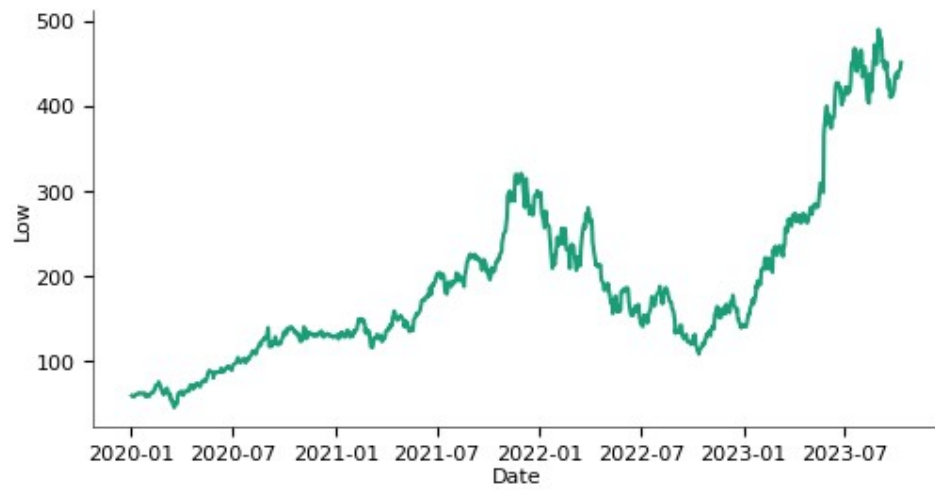
Fig: 4.6.2.15 Lowest Price
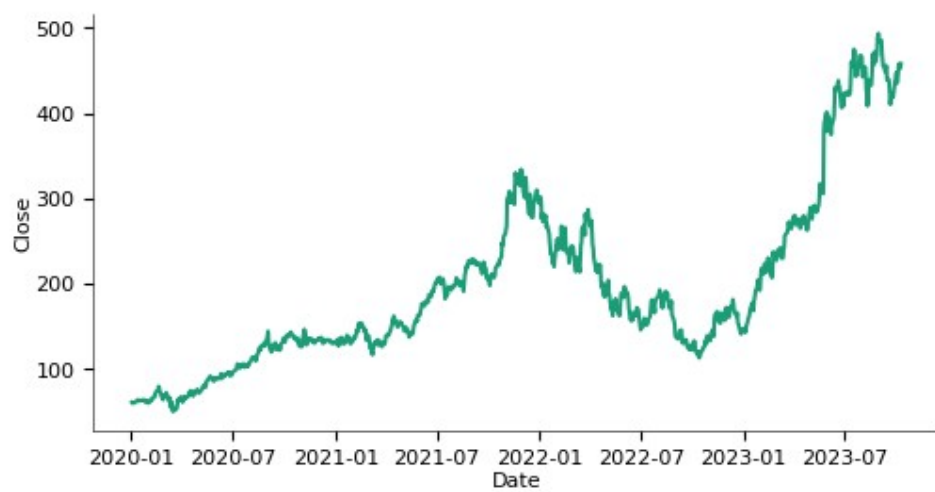
Fig: 4.6.2.16 Close Price

Fig: 4.6.2.17 Equity Curve


```
+----+-----------+-----------+--------------+
|    | Close     |  Price $  | Accuracy %   |
+====+===========+===========+==============+
|  0 | Current   |   468.06  | N/A          |
+----+-----------+-----------+--------------+
|  1 | Predicted |   457.48  | -2.26        |
+----+-----------+-----------+--------------+
```

Fig: 4.6.2.18 Outcomes

4.6.3 For Predicting Stock Buying Recommendation

Fig: 4.6.3.1 Open Price

28

Fig: 4.6.3.2 Highest Price

Fig: 4.6.3.3 Lowest Price

Fig: 4.6.3.4 Close Price

Fig: 4.6.3.5 Open Price

Fig: 4.6.3.6 High

Fig: 4.6.3.7 Lowest Price

Fig: 4.6.3.8 Close Price

Fig: 4.6.3.9 Open Price
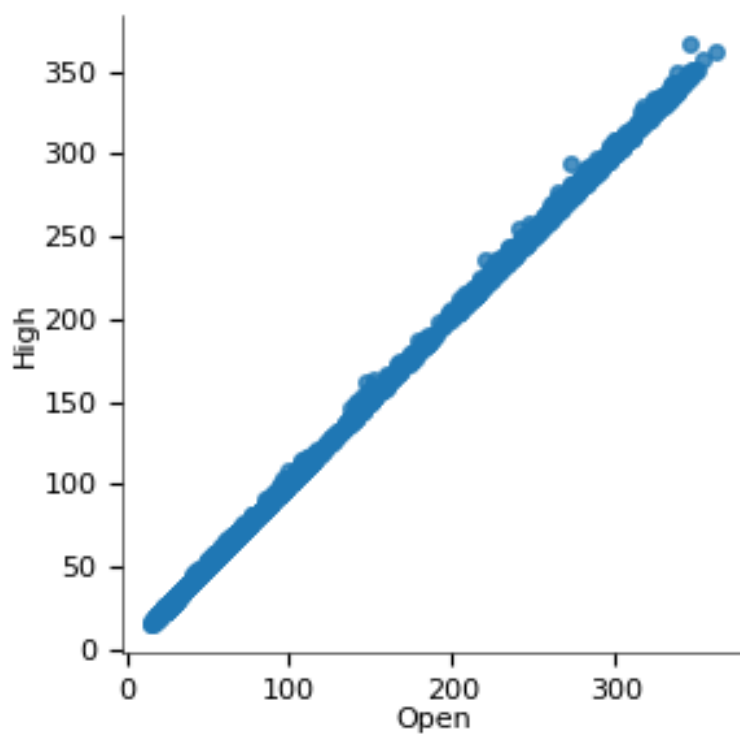
Fig: 4.6.3.10 Highest Price
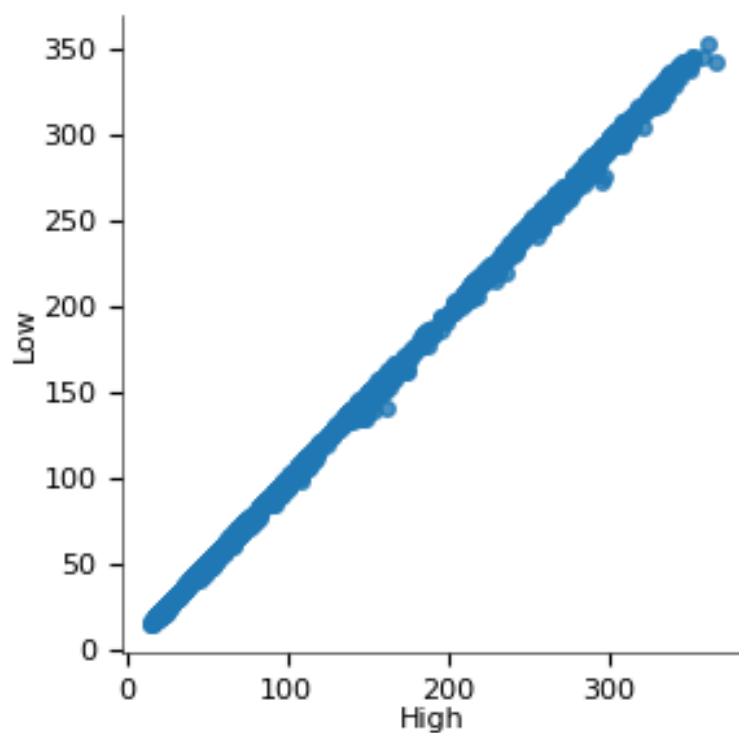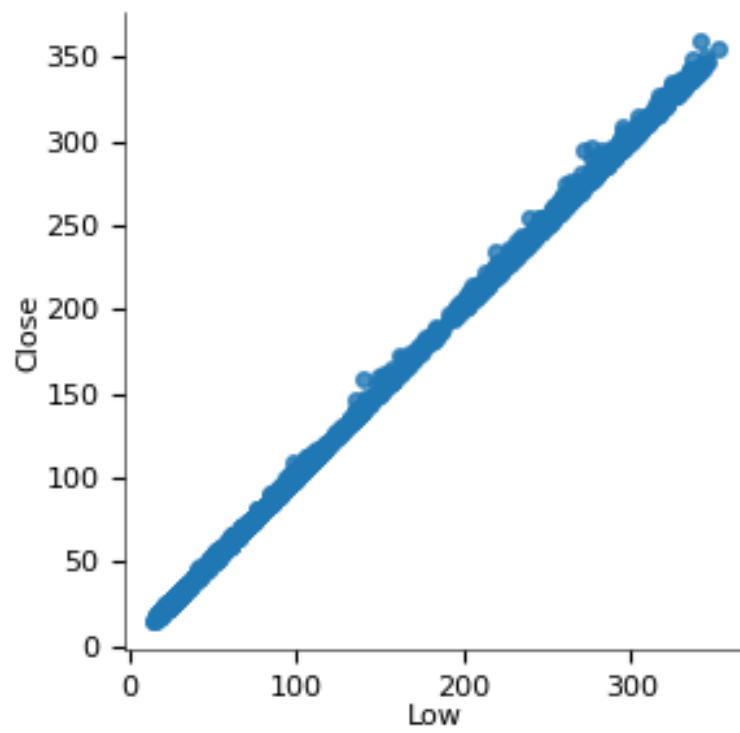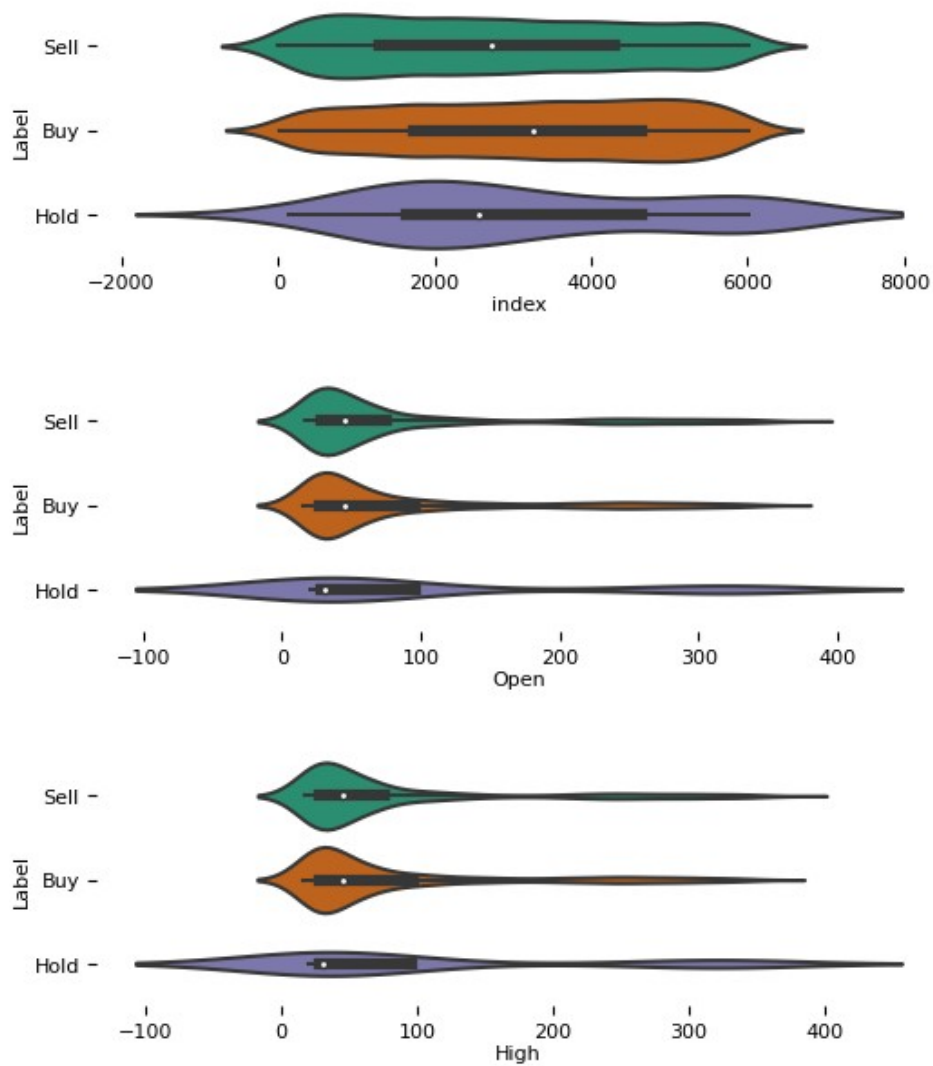

Fig: 4.6.3.11 Lowest Price
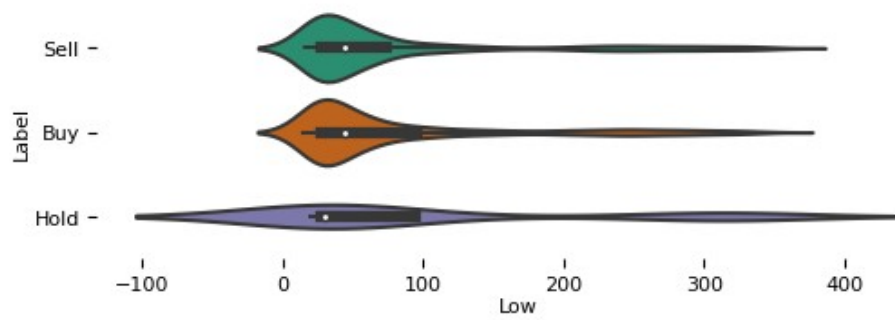
Fig: 4.6.3.12 Close Price
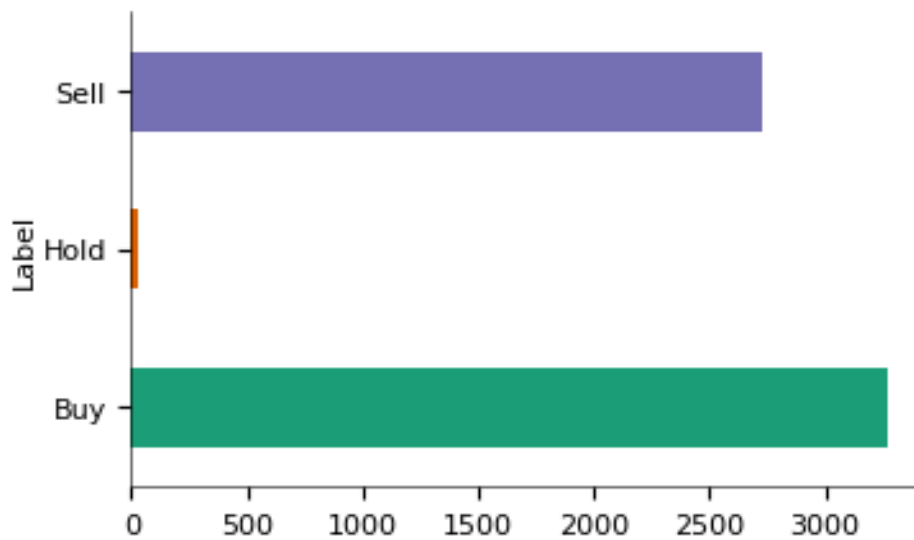
Fig: 4.6.3.13 Faceted Distributions

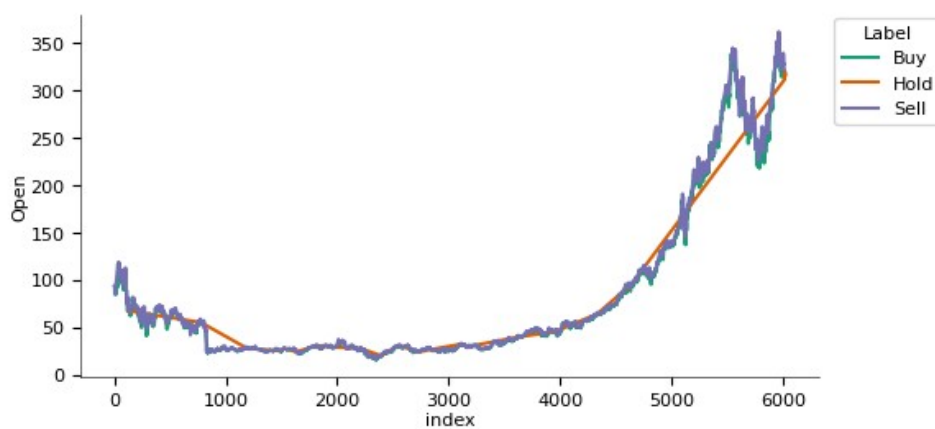Fig: 4.6.3.14 Categorical Distributions

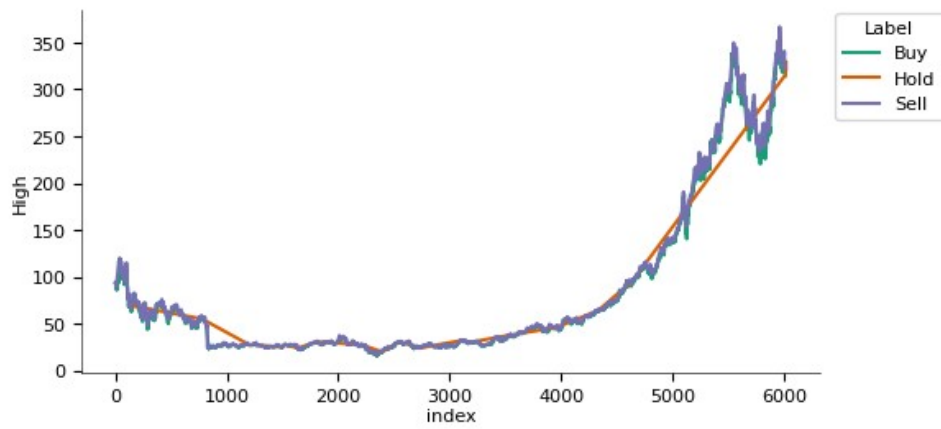Fig: 4.6.3.15 Open Time Series

34

Fig: 4.6.3.16 High Time Series
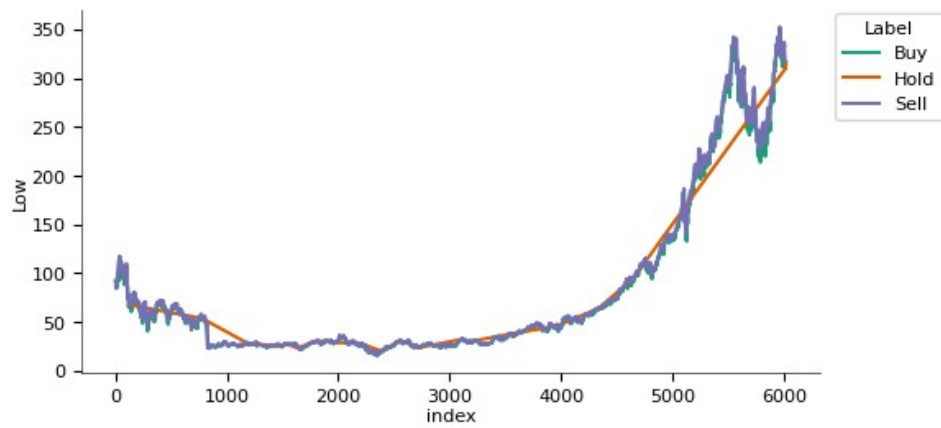
Fig: 4.6.3.17 Low Time Series
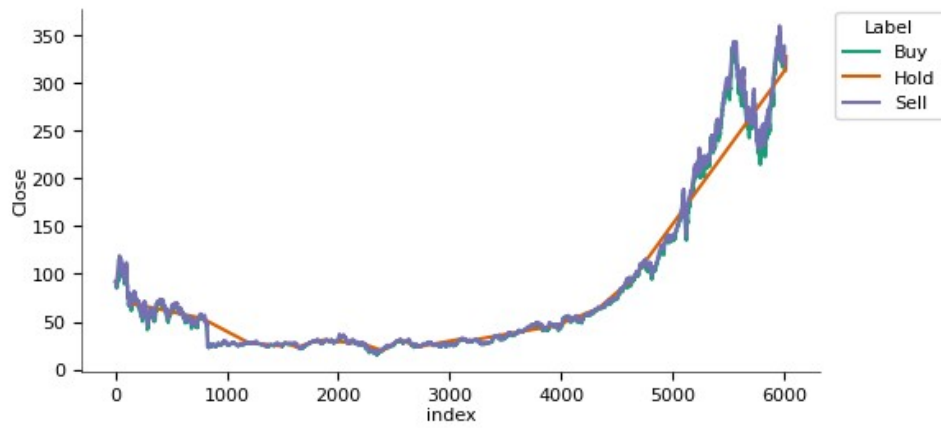
Fig: 4.6.3.18 Close Time Series
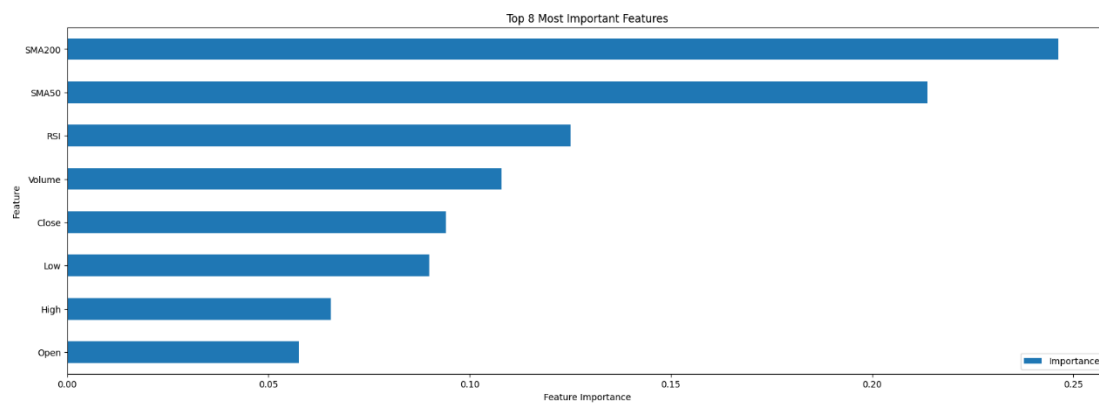
Fig: 4.6.3.19 Features Analysis

Fig: 4.6.3.19 Outcomes

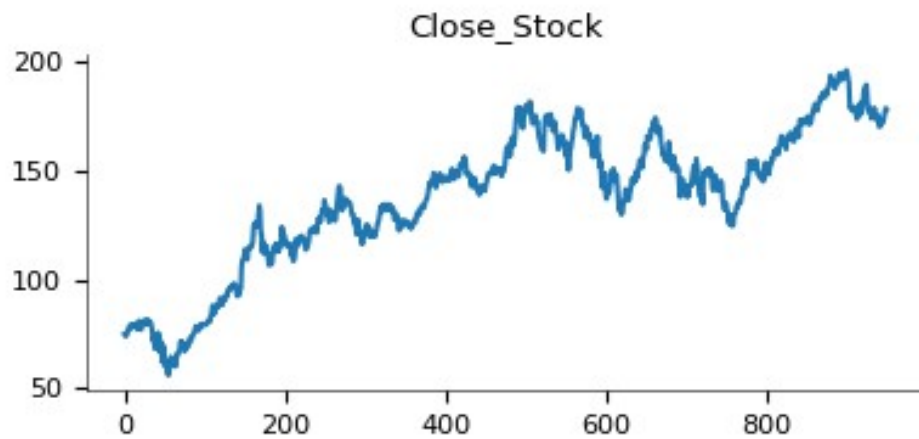**4.6.4 For Predicting Investors' Portfolio Impact**

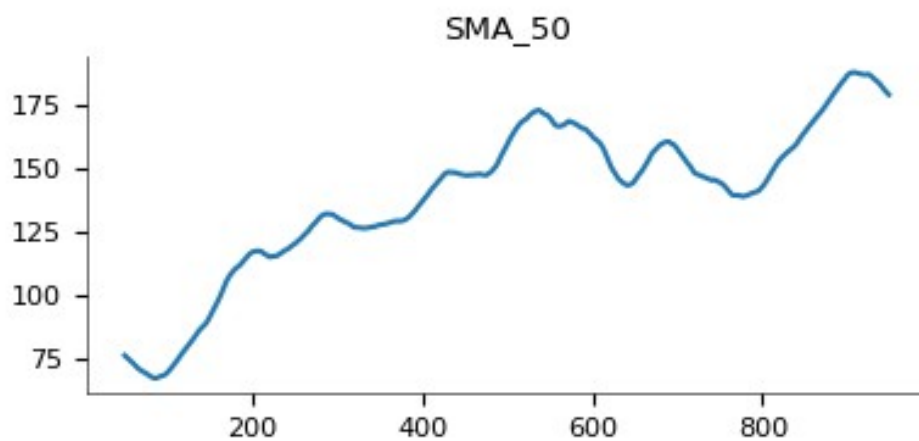Fig: 4.6.4.1 Closing Price

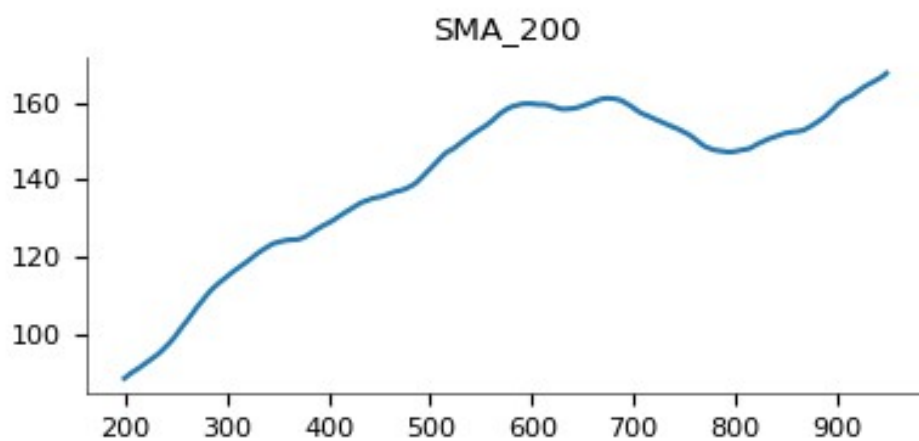Fig: 4.6.4.2 Simple Moving Average for 50 Days
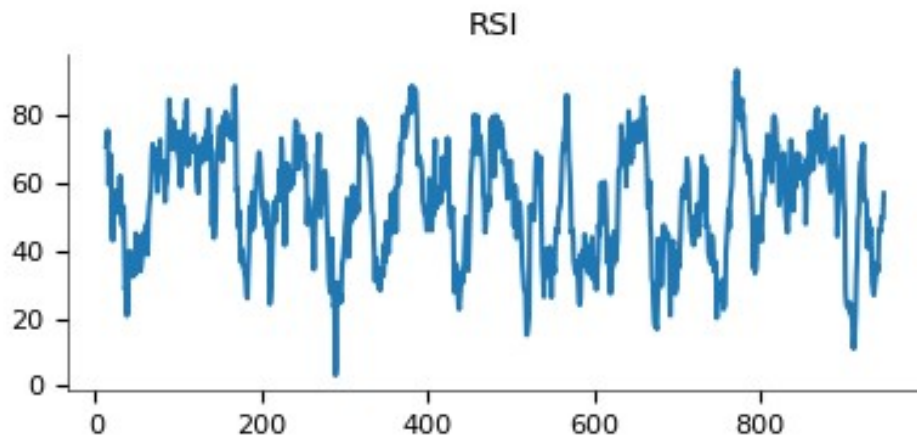

Fig: 4.6.4.3 Simple Moving Average for 200 Days
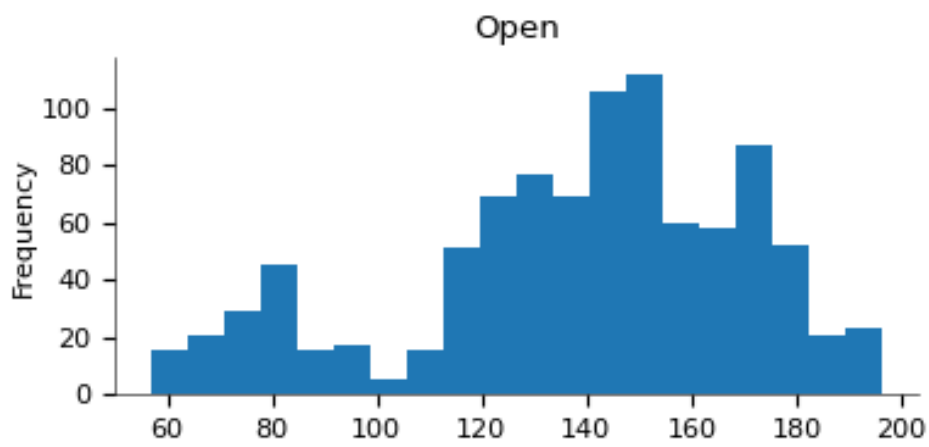
Fig: 4.6.4.4 Relatively Strength Index
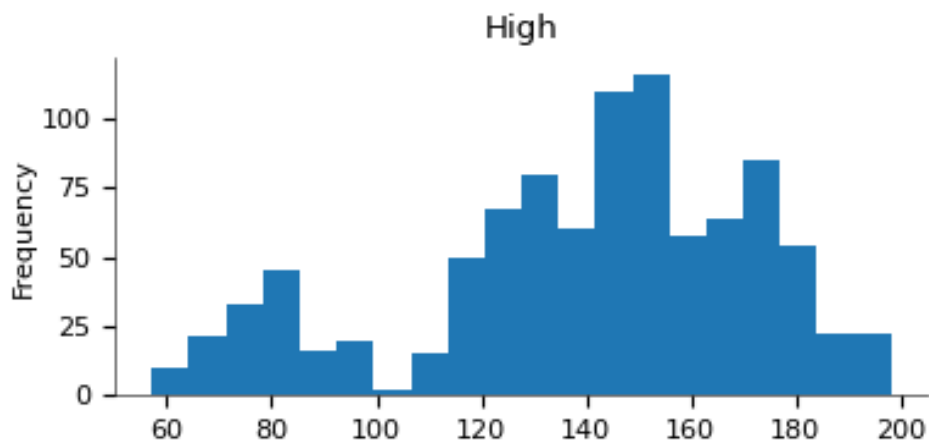

Fig: 4.6.4.5 Open Price

Fig: 4.6.4.6 Highest Price
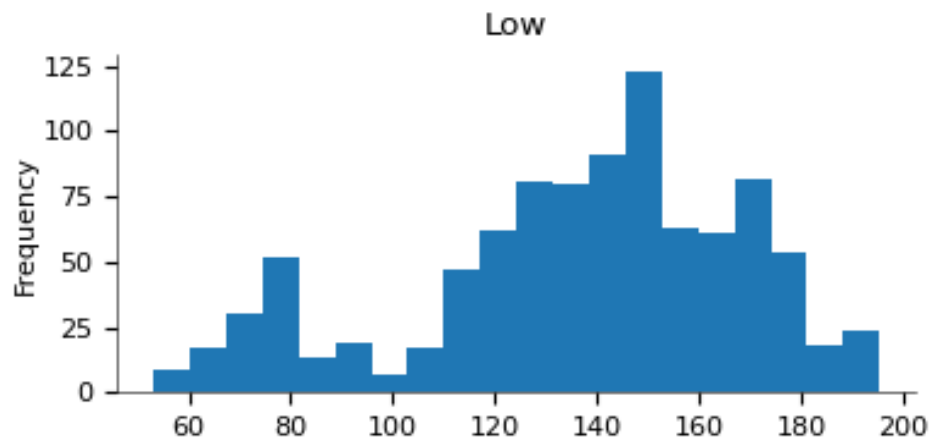
Fig: 4.6.4.7 Lowest Price
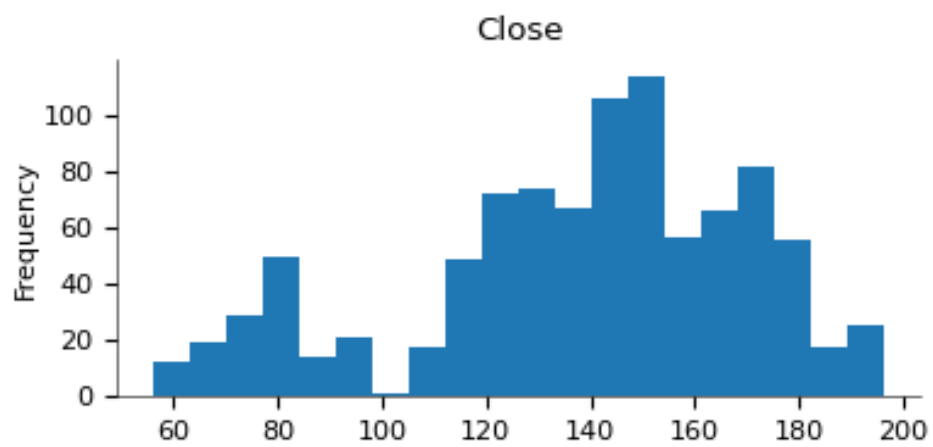
Fig: 4.6.4.8 Close Price
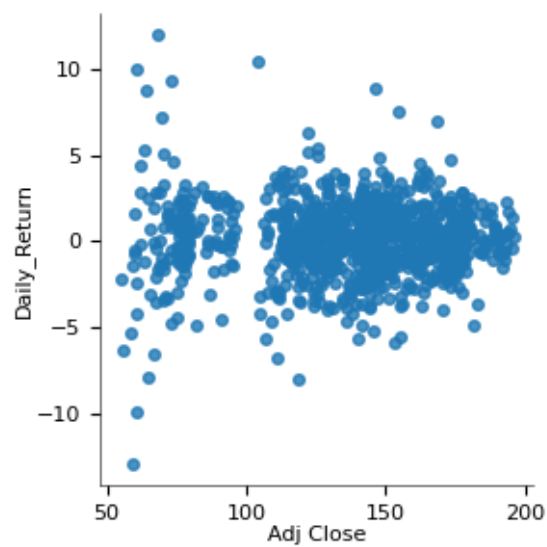
Fig: 4.6.4.9 Daily Return

Fig: 4.6.4.10 Adj Close
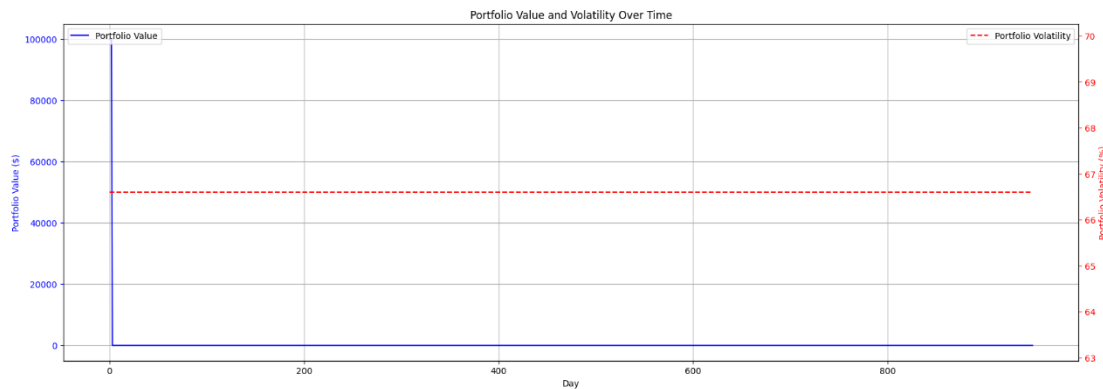
Fig: 4.6.4.11 Portfolio Value & Volatility Over Time

```
Initial Investment: $100000.00
Portfolio Volatility (Risk): 66.60%
```

Fig: 4.6.4.12 Outcomes

4.7 Discussion

We carefully check if current stock prices are right, evaluating accuracy in both percentage and actual amounts. Moving to Next-Day Prediction Insights, we compare real-time and predicted prices, focusing on differences in percentage and amount. In Strategic Buying Recommendations, we unveil the model's advice at key stages: Sell, Hold, and Buy. The final part, Portfolio Impact Examination, delves into the fallout of inaccurate share price predictions on the investment portfolio.

Chapter 5

Conclusion

5.1 Conclusion

We looked at predicting stock prices and giving investment advice using machine learning, finding some important information. Checking how accurate the current prices are showed that the model can tell if they're right, measuring correctness in percentages and real amounts. Predicting prices for the next day demonstrated that the model is good at comparing what it thinks will happen to what really happens, showing the differences in percentages and amounts. The advice on when to buy or sell, called Strategic Buying Recommendations, gave practical suggestions at important times: Sell, Hold, and Buy. However, we also researched how getting the share prices wrong could affect an investment portfolio, emphasizing the need to improve the model for better accuracy and to reduce risks. Looking ahead, making our machine learning system better is crucial for navigating the complexities of the stock market.

Reference

[1] Davis, C. K. (2022). Machine learning, quantitative portfolio choice, and mispricing (Doctoral dissertation, The University of Chicago).

[2] Gu, A., Viens, F.G. and Yi, B., 2017. Optimal reinsurance and investment strategies for insurers with mispricing and model ambiguity. Insurance: Mathematics and Economics, 72, pp.235-249.

[3] Tu, J. and Zhou, G., 2010. Incorporating economic objectives into Bayesian priors: Portfolio choice under parameter uncertainty. Journal of Financial and Quantitative Analysis, 45 [4] pp.959-986.

[4] Ang, A., Papanikolaou, D. and Westerfield, M.M., 2014. Portfolio choice with illiquid assets. Management Science, 60(11), pp.2737-2761.

[5] Ben-David, I., Drake, M.S. and Roulstone, D.T., 2015. Acquirer valuation and acquisition decisions: Identifying mispricing using short interest. Journal of Financial and Quantitative Analysis, 50(1-2), pp.1-32.

[6] Cvitanić, J., Lazrak, A., Martellini, L. and Zapatero, F., 2006. Dynamic portfolio choice with parameter uncertainty and the economic value of analysts' recommendations. The Review of Financial Studies, 19(4), pp.1113-1156.

 [7] Liu, J. and Longstaff, F.A., 2004. Losing money on arbitrage: Optimal dynamic portfolio choice in markets with arbitrage opportunities. Review of Financial Studies, pp.611-641.

[8] Doukas, J.A., Kim, C.F. and Pantzalis, C., 2010. Arbitrage risk and stock mispricing. Journal of Financial and Quantitative Analysis, 45(4), pp.907-934.

[9] Sørensen, C., 1999. Dynamic asset allocation and fixed income management. Journal of Financial and Quantitative Analysis, 34(4), pp.513-531.

[10] Stulz, R.M., 1995. International portfolio choice and asset pricing: An integrative survey. Handbooks in operations research and management science, 9, pp.201-223.

[11] Davis, C.K., 2022. Machine learning, quantitative portfolio choice, and mispricing (Doctoral dissertation, The University of Chicago.