

Project Report On

“Stock Price Decision Forecasting”

Submitted by
Rohan Upadhyay
SJSU ID: 016672682

Under The Guidance Of
Professor Mahima Agumbe Suresh
Department of Computer Engineering
May 2023

Acknowledgement

I do acknowledge guidance and support of professor Mahima Agumbe Suresh, department of Computer Engineering, San Jose State University. I am thankful for the professor's assistance in the project and grateful for providing me proper direction to work on the project. I am also thankful for my classmates and teammates for their continuous support and encouragement.

Abstract

The historical data of the Tokyo Stock Market is used for the training and prediction of the price of stock data. The data is of high frequency trading of stocks. Different methods are used for prediction of decisions based on forecasting price. Now these models can be trusted to use for making the strategies of stock exchange in real time.

Contents

No.	Topics	Page
1	Introduction	2
2	Methodology	2
3	Project Plan	4
4	Future Work	4
5	Conclusion	4
6	References	5

Introduction

The historical dataset is used for the price forecasting the decision of buy or sell for the stocks. The data related to securities is found, then outlier data for the closing price is removed. This way primary data analysis is done first. Data exploration is performed and obtained knowledge is used for the process of future predictions. Finally, various ARIMA time series models are used and compared to obtain better prediction results.

Methodology

Firstly, data analysis is done on the historical data of the Tokyo Stock Market[2] using pandas, numpy, matplotlib and more libraries. The data is cleaned and meanwhile the data exploration is done for deciding the methods to use for decision forecasting. The figure 1 shows the analysis on the 'Close' price at the end of each day. From the graph and figure 2 represents the Close price target returns over day. Then targets are distributed and analyzed for future use. From looking at the data statistics, histogram, skew and kurtosis, the returns are fairly normally distributed. Thus it motivates for trying to fit linear regression as a simple baseline for the further implementations. Thus simple linear regression is performed along with which training and test dataset is decided. As target data is stationary we use ARIMA time series model for the prediction. At first the AR model is used and it is cutting off as seen in autocorrelation data present in figure 3. From the observations it is decided to use the ARIMA(2,0,0) model along with other ARIMA models and compare the accuracies.

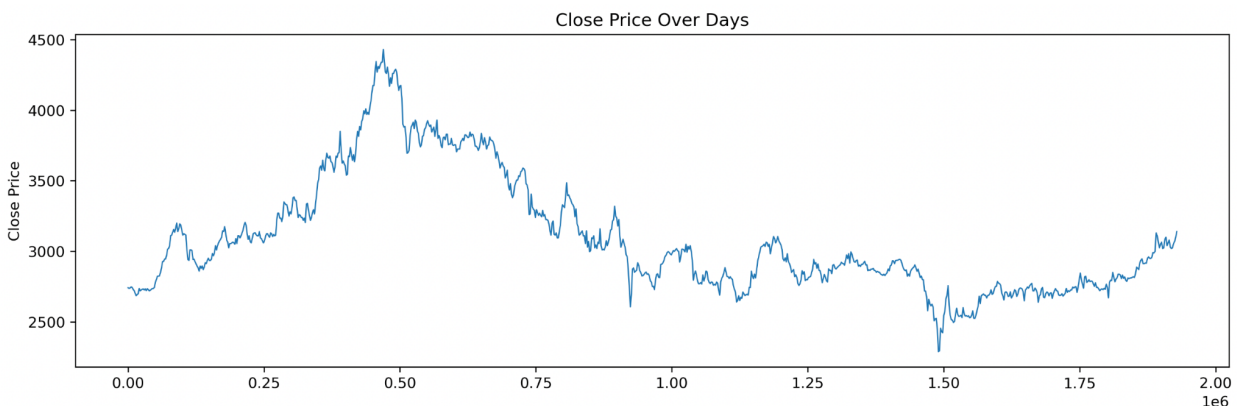


Figure 1.: Closing price over days

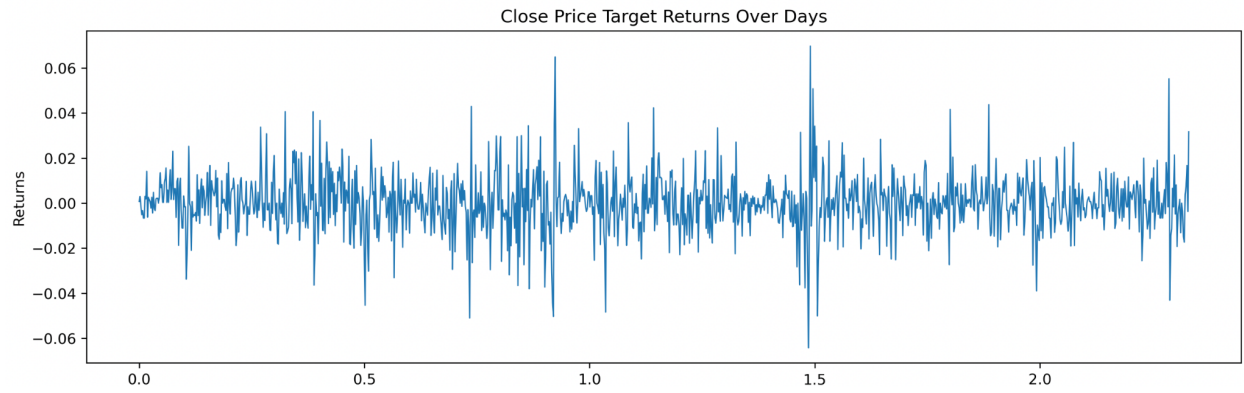


Figure 2.: Close Price Target Returns

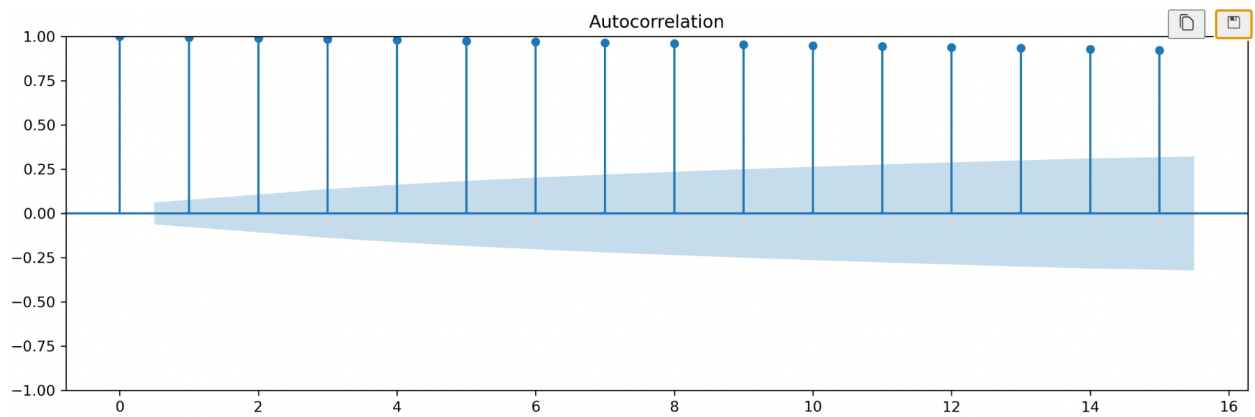


Figure 3: Autocorrelation

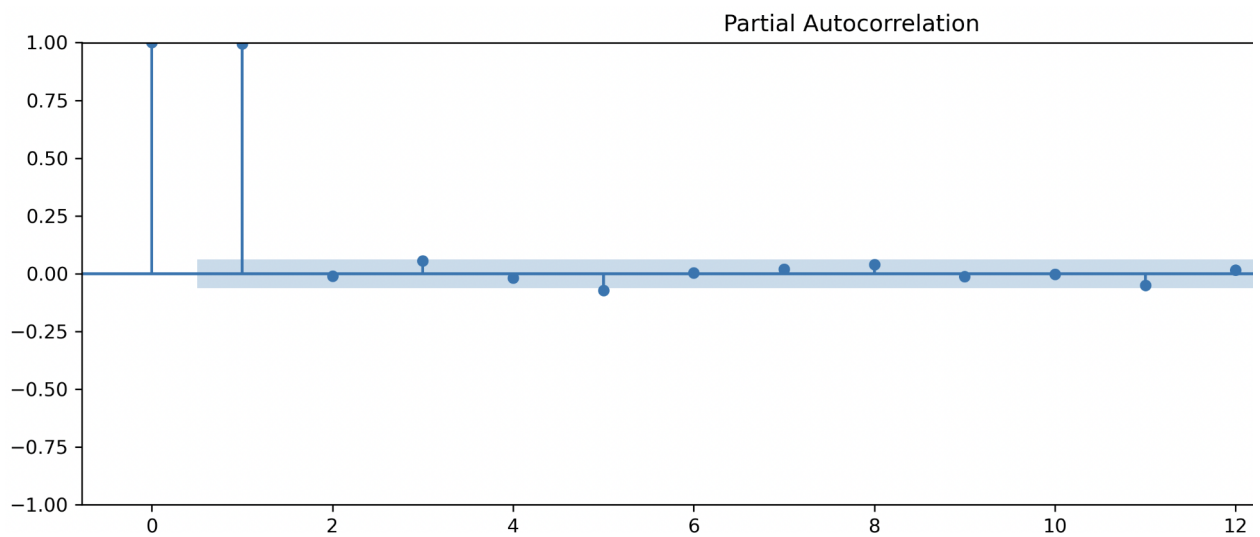


Figure 4: Partial Autocorrelation

Project Plan

The plan is to clean the data set and do data exploration first. After completion of data exploration we do move forward to define the problem and the proposed method of solving it. The aim is to develop an accurate model for the prediction of the stock market data based on JPX Tokyo Stock Market Data[2]. The task is to predict the Buy/Sell decision so it seems to be a classification problem. But it is more important to find out the exact moment at which the accuracy of this prediction is high. The final goal is to make profits using this trained model in real time. So it becomes a regression problem. Now the Linear Regression is performed on the labeled dataset. The ARIMA models are used for prediction purposes. Now this models are compared to find out the most accurate one.

Future Work

The future plan is to increase the accuracy of model aiming better prediction. Using the current model we may obtain the current performance data and define the work goals for further improvements. Now, this way we may also develop our electronic trading strategies. Finding accurate outcomes we may also apply this model to perform in the real market. Improvement in such a field is highly valuable and may lead doors towards the advancement of prediction strategies. Meanwhile, other data models are also to be explored to improve the accuracy.

Conclusion

The ARIMA models are used for the better prediction of the points when the trading call will be more successful. The provided model is trained to do predictions for buy and sell of data on a predicted point of time for securing the profits. The trained model uses regression analysis for forecasting these accurate predictions time points. The intensity of price target return signals are used to define these points. All in all, ARIMA time series model, ARIMA(2, 0, 0) seems to be the best model for 2 predictions with higher train correlation and comparable test correlation.

References

- [1] [Online]. Available: <https://otexts.com/fpp2/stationarity.html>. [Accessed: 15-May-2023].
- [2] “JPX Tokyo Stock Exchange Prediction”, *Kaggle*. [Online]. Available: <https://www.kaggle.com/competitions/jpx-tokyo-stock-exchange-prediction/discussion/317038>. [Accessed: 15-May-2023].
- [3] “Introduction to ARIMA models”, [Online]. Available: <https://people.duke.edu/~rnau/411arim.htm>. [Accessed: 15-May-2023].
- [4] “GitHub - 0ROHAN0/CMPE255_HW: This repository is about the work of Data Mining .”, *0ROHAN0/CMPE255_HW: This repository is about the w*. [Online]. Available: https://github.com/0ROHAN0/CMPE255_HW. [Accessed: 15-May-2023].
- [5] “Learn Pandas Tutorials”, [Online]. Available: <https://www.kaggle.com/learn/pandas>. [Accessed: 15-May-2023].