Stock Price Prediction Project Report

Zhiheng Chang, Janelle Zeng

❖ OVERVIEW

There is an unstoppable growing interest in predicting markets for years among economists, policymakers, academics and market makers, and investors. The prediction of a stock market may work as a recommendation/warning program for investors or shareholders. The most important factor of stock price prediction is the forecasting accuracy. No doubt that the stock price prediction will essentially helps on investors' selection on earning the maximum profits with a high predictive accuracy.

❖ PROBLEM

The Classic advise: "buy low, sell high" does not provide enough context to help investors to make proper investment decisions. Although advanced mathematics and statistics does help some time, computer sciences is more helpful by implementing different algorithms because computers can make trading decisions way faster than humans could make. The intention of this project is telling the investor when is the right time to buy/sell to maximize the profits.

❖ DATA SELECTION

For this project, The stock price data is obtained from Yahoo Finance. The data starts from december 10th of year 2012 to december 10th of the current year (2019). The data set (csv file) will be uploaded from the local machine. The csv contains seven columns: Date, Open, High, Low, Close, Adjust Close, and Volume. The price of the first trade for any listed stock is the Open price. The last price at which a stock trades during a regular trading session is the Close price. The highest/lowest closing price of a stock over the past 52 weeks is the High/Low price. Adjusted closing price amends a stock's closing price to accurately reflect that stock's value after accounting for any corporate actions. In this project, the adjusted closing price will be mainly focused on. The program will analyse historical adjusted closing price and predict future adjusted closing price so the 'Adj Close' column will be extracted from the original data set and set index as its corresponding date.

❖ APPROACH

For this project, two machine learning algorithm will be imported from sklearn and used to predict future stock price: Support Vector Machine and Linear regression. Both are commonly used and popular in stock price predicting. SVM is a supervised machine learning algorithm which can be used for both classification or regression challenges. Linear regression is also a supervised learning which models a target prediction value based on independent variables. A new column called 'Prediction' is created which holds the target or dependent variable through duplicate the 'Adj Close' column but shifted up by n days (n is the number of days user want to predict into the future and it is used throughout the program). The independent variable x is the 'Adj Close' column and 'Prediction' column is the dependent variable y. Then split the data into 80% training and 20% testing, adn create both SVM and LR model fit with trained data. SVM gets a confidence score of 0.975 and LR gets a confidence score of 0.971 which both models are really close to the best score of 1.0. Lastly, the predict() method will be used to predict the next n target values based on the 'Adj Close' data.

❖ SUMMARY

The theoretical prediction, however, can not be absolutely, 100% accurate even with a best algorithm because in real world circumstances, there are many more factors to consider rather than just the Adjusted closing price such as event shocks like 911.

❖ REFERENCES

1. https://finance.yahoo.com/quote/FB/history?period1=1355115600&period2=1575954000&interval=1d&filter=history&frequency=1d
2. https://medium.com/@randerson112358/predict-stock-prices-using-python-machine-learning-53aa024da20a
3. https://medium.com/machine-learning-101/chapter-2-svm-support-vector-machine-theory-f0812effc72#:~:targetText=A%20Support%20Vector%20Machine%20(SVM,hyperplane%20which%20categorizes%20new%20examples.
4. https://byteacademy.co/blog/time-series-python