FINAL REPORT

Do-Young Chung

03/12/2021

Statement of the problem

The stock market is where investors connect to buy and sell investments to make a large amount of money. Nowadays, more people are turning to stock market investment for profits and even entertainment. But the problem is that most likely average day trader loses money. Furthermore, as the cryptocurrency market rises and enjoys unprecedented success, now investment is very common in our life and the intrinsic values of mathematics and computing to investment are emerging. Within this situation, I used ARIMA (autoregressive integrated moving average) model to predict future points in the price of time series data.

Purpose

The purpose of the program is to explore a dataset of large company stocks, and analyze them in order to boost up the probability of winning profits by predicting prices using the ARIMA module. The program analyzes stock market data to predict the following price through five stages: Data Set Selection, Data Set Cleaning, Data Set Visualization, and Data Set Prediction. Ultimately, compare the accuracy of predicted data by using different size of data and different types of data.

Dataset

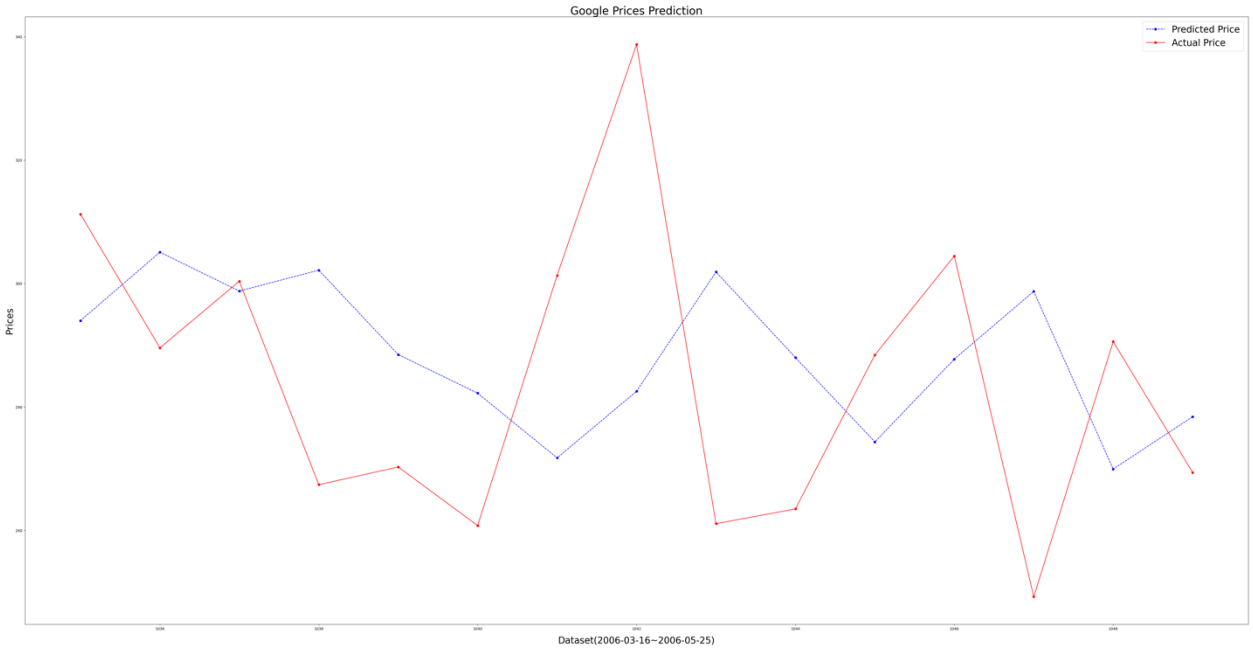
- Google (2006-01-03 to 2017-12-29)
- Amazon (2006-01-03 to 2017-12-29)

Technology/Package

- Python (Google Collab)
- NumPy and Pandas Data frame
- Data Set Visualization Tools (matplotlib, lag_plot, bootstrap_plot)
- Data Set Prediction Tools (ARIMA, mean_squared_error)

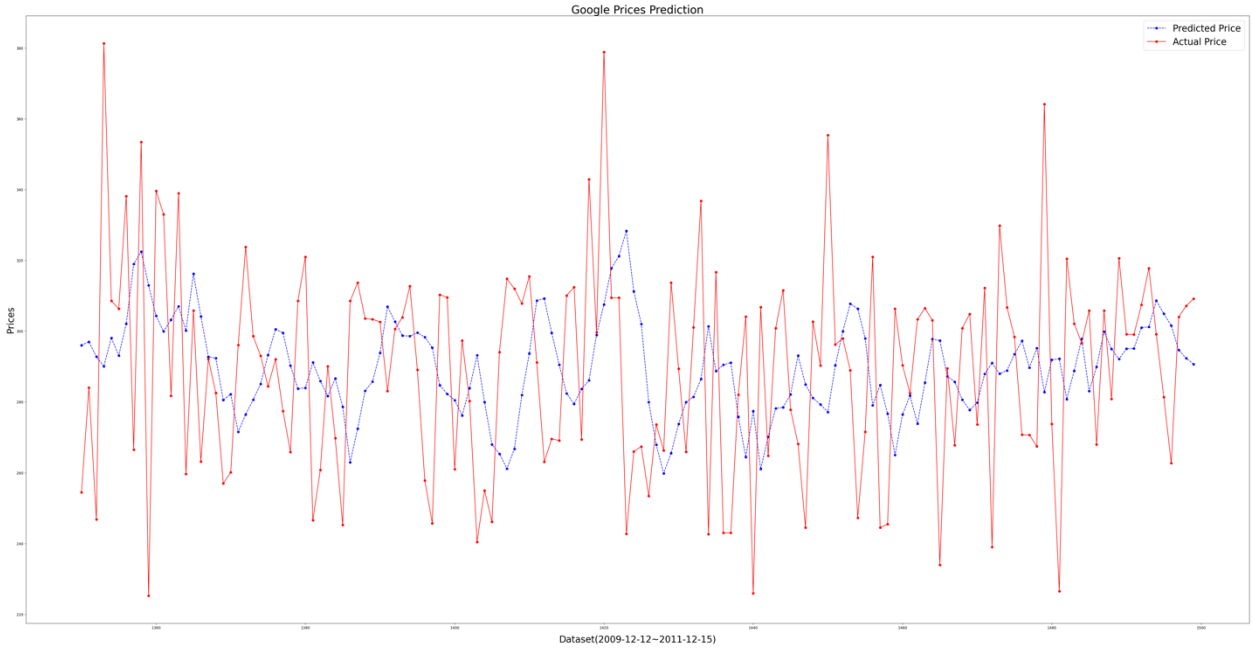
Result

50 Size Google Stock Data (2006-03-16 to 2006-05-25)



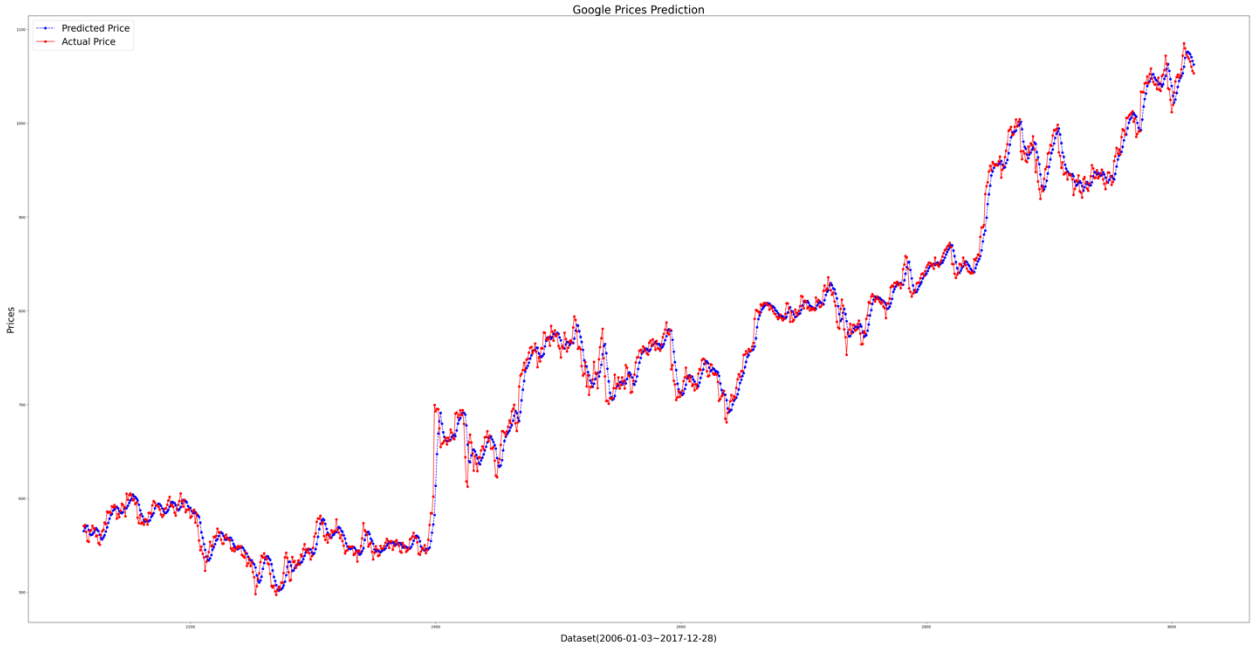
The accuracy of the prediction data is 54.743104437698776%

500 Size Google Stock Data (2009-12-12 to 2011-12-15)



The accuracy of the prediction data is 75.29703101252647%

3018 Size Google Stock Data (2006-01-03 to 2017-12-29)

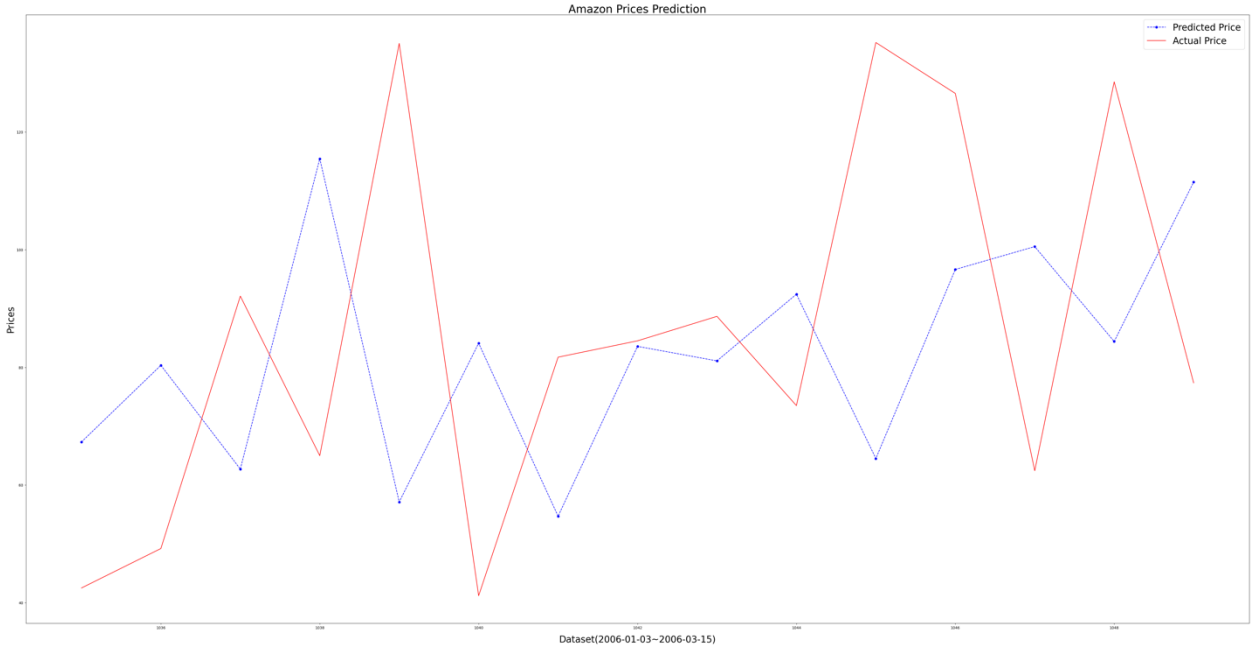


The accuracy of the prediction data is 98.38578746878025%

Google Size Accuracy Table

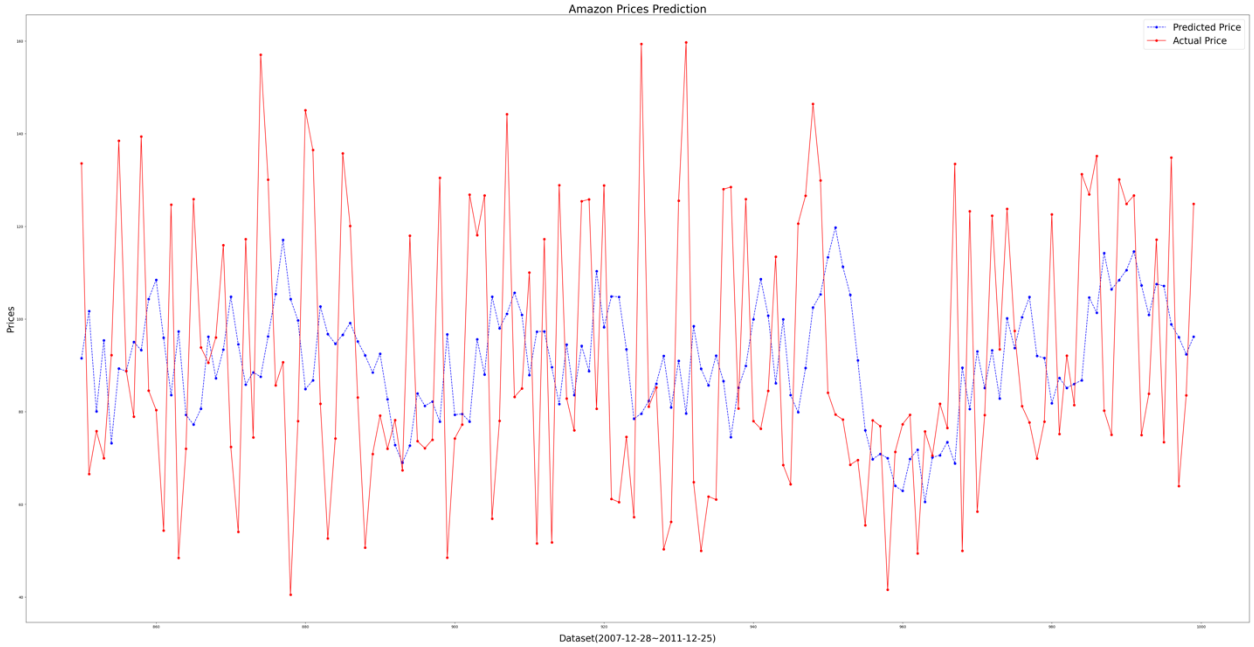
Size	50	500	3018
Accuracy (%)	63.2%	70.5%	98.4

50 Size Amazon Stock Data (2006-01-03 to 2006-03-15)



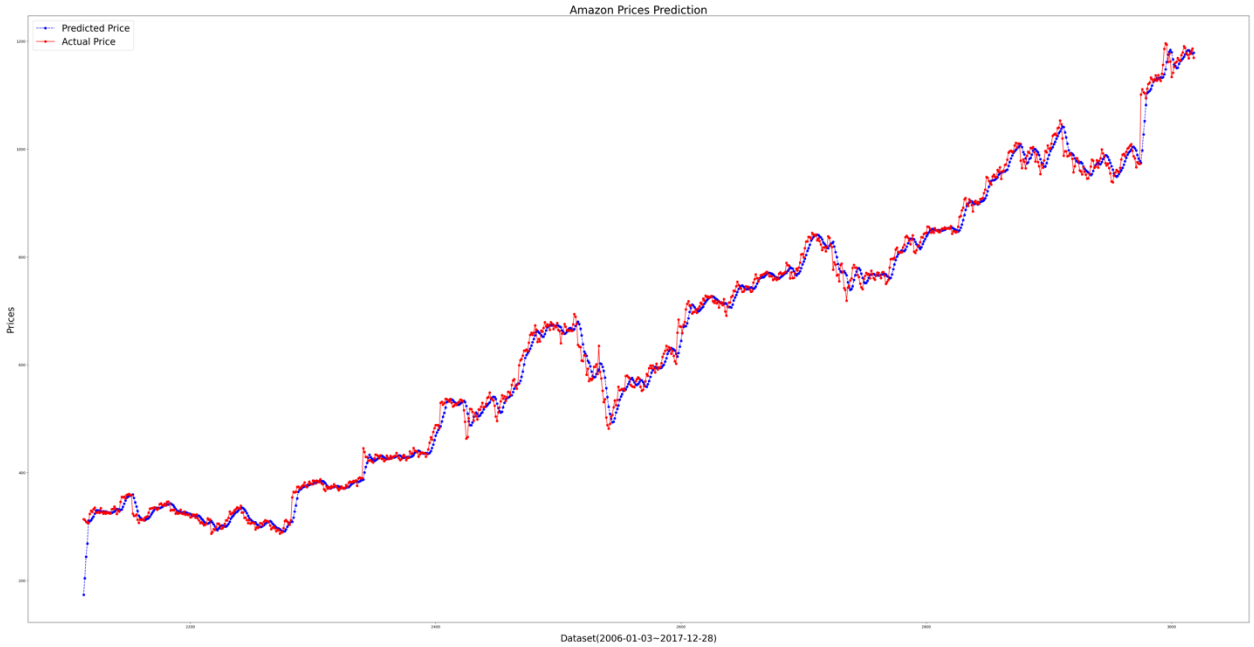
The accuracy of the prediction data is 54.86585128609253%

500 Size Amazon Stock Data (2007-12-28 to 2009-12-21)



The accuracy of the prediction data is 67.9201803636624%

3018 Size Amazon Stock Data (2006-01-03 to 2017-12-29)



The accuracy of the prediction data is 98.00405754201196%

Amazon Size Accuracy Table

Size	50	500	3018
Accuracy (%)	54.9%	67.9%	98.0%

Compare

Google Size Accuracy Table

Size	50	500	3018
Google Accuracy (%)	63.2%	70.5%	98.4
Amazon Accuracy (%)	54.9%	67.9%	98.0%
Difference (absolute %)	8.3%	2.6%	0.4%

As the size of data that is put to train increases, the accuracy of the algorithm also increases.