Springboard Capstone Project Two
Stock Price Prediction using Time Series Analysis

Author: Anshul Dikshit

Date: November 21, 2018

Outline

- Introduction & Problem Statement
- Dataset
- Analysis & Findings
- Machine Learning through different Time-Series Analysis algorithms.
- Conclusions

Introduction & PROBLEM STATEMENT

Making Money in stock market is every investor's dream. There are a lot of high valued companies listed on S&P and NASDAQ which can yield high profits if their stocks are traded wisely.

THE PROBLEM

Stock market prediction is a very complex process as it depends on various factors including market sentiments, IPOs, social media announcements and others.

| 1 day | 5 days | 1 month | 6 months | YTD | **1 year** | 5 years | Max |

2,000

1,500

1,000

Mar 2018   Jul 2018   Nov 2018

Data Set

The data has been taken from macrotrends.com which provides a starting point dataset consisting **Amazon** records starting from May 1997. I've tried to go through the process of understanding the individual variables in the data by presenting beautiful, clear, and interactive data visualizations along with some approaches to their interpretation.

Test dataset:

| Variable Name	Variable description
date	Date on which trade was done
open	Opening value of the stock for that day
high	High value of the stock during that day
low	Low value of the stock during that day
close	Closing value of the stock
volume	Total number of stocks traded

Train dataset:

| Variable Name	Variable description
date	Date on which trade was done
open	Opening value of the stock for that day
high	High value of the stock during that day
low	Low value of the stock during that day
close	Closing value of the stock
volume	Total number of stocks traded

Prediction using Moving Average
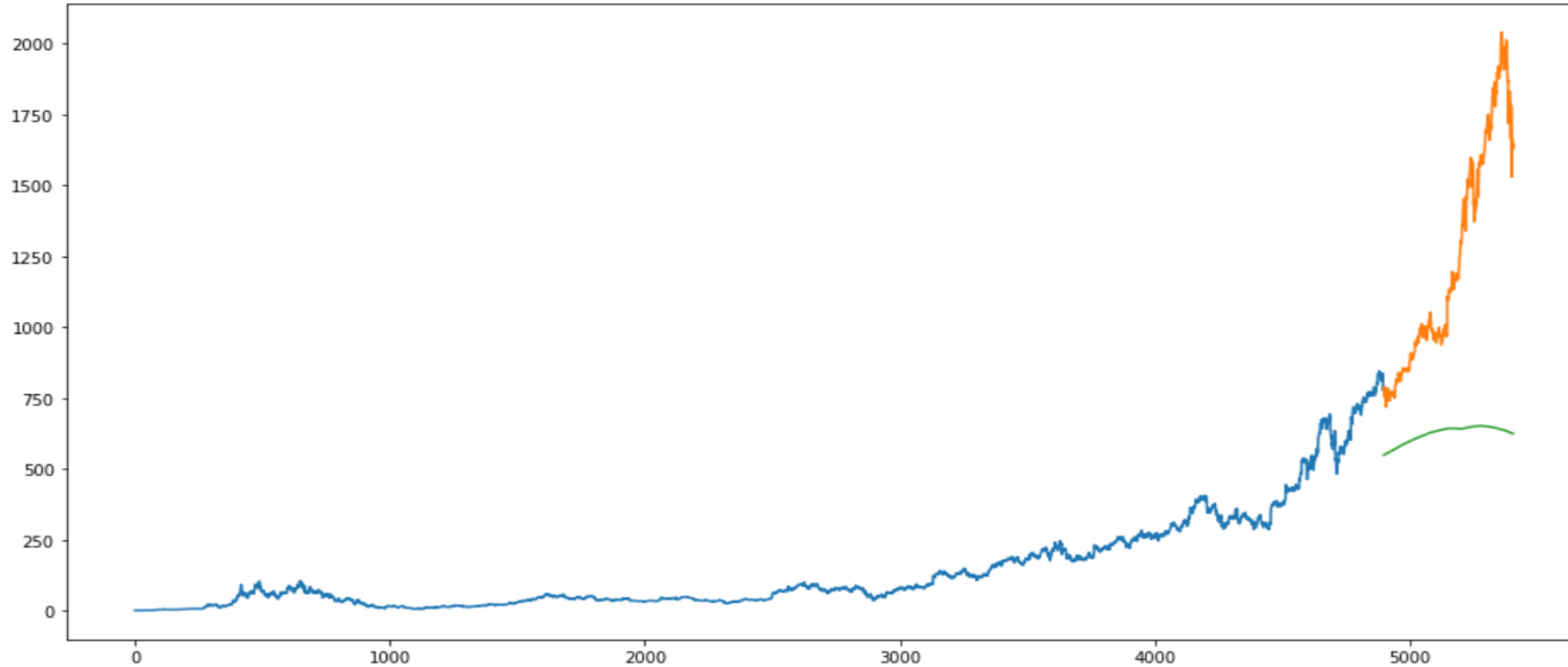
The RMSE value is close to 717 and the results are not very promising (as you can gather from the plot). The predicted values are of the same range as the observed values in the train set (there is an increasing trend initially and then a slow decrease).

Prediction using Linear Regression
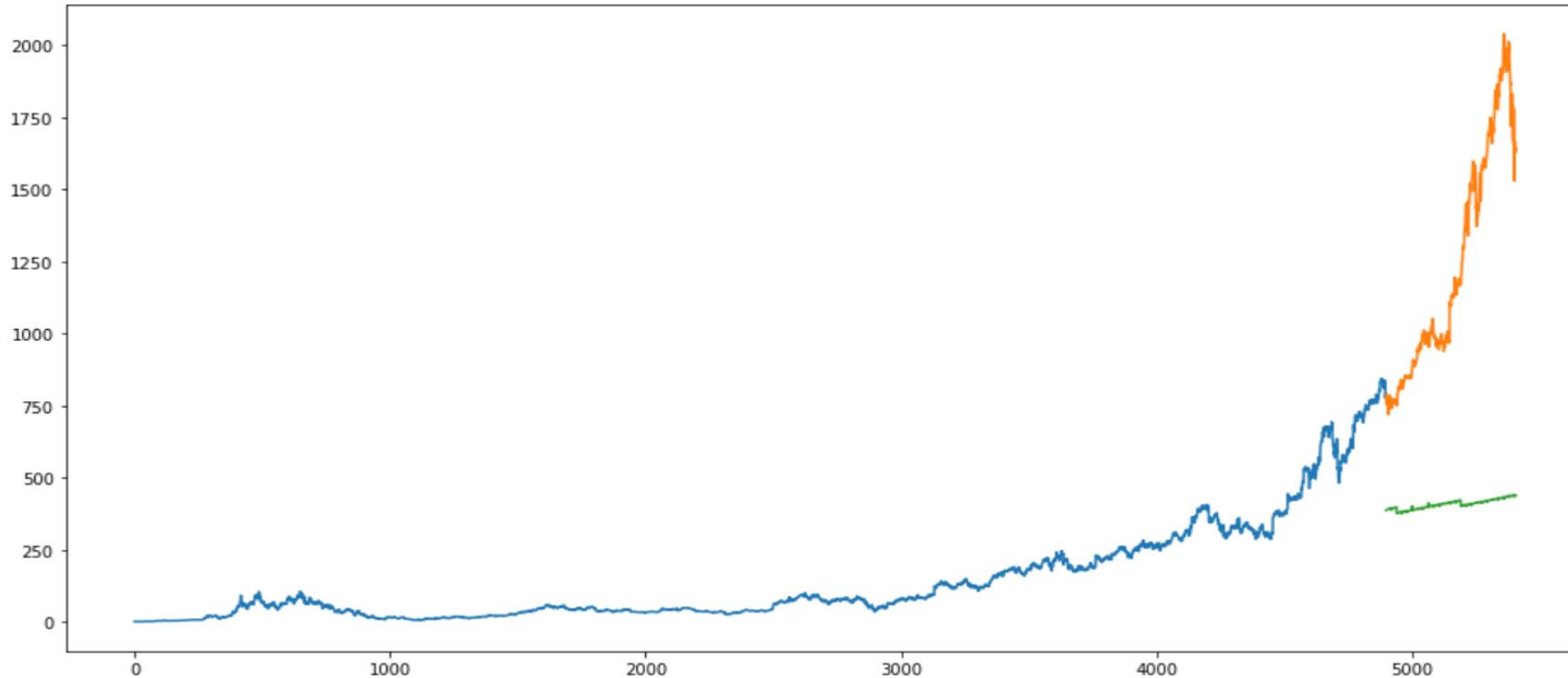
The RMSE value is higher than the Moving Average, which clearly shows that linear regression has performed poorly. Linear regression is a simple technique and quite easy to interpret, but there are a few obvious disadvantages. One problem in using regression algorithms is that the model overfits to the date and month column. Instead of taking into account the previous values from the point of prediction, the model will consider the value from the same date a month ago, or the same date/month a year ago.

Prediction using **k-Nearest Neighbours**
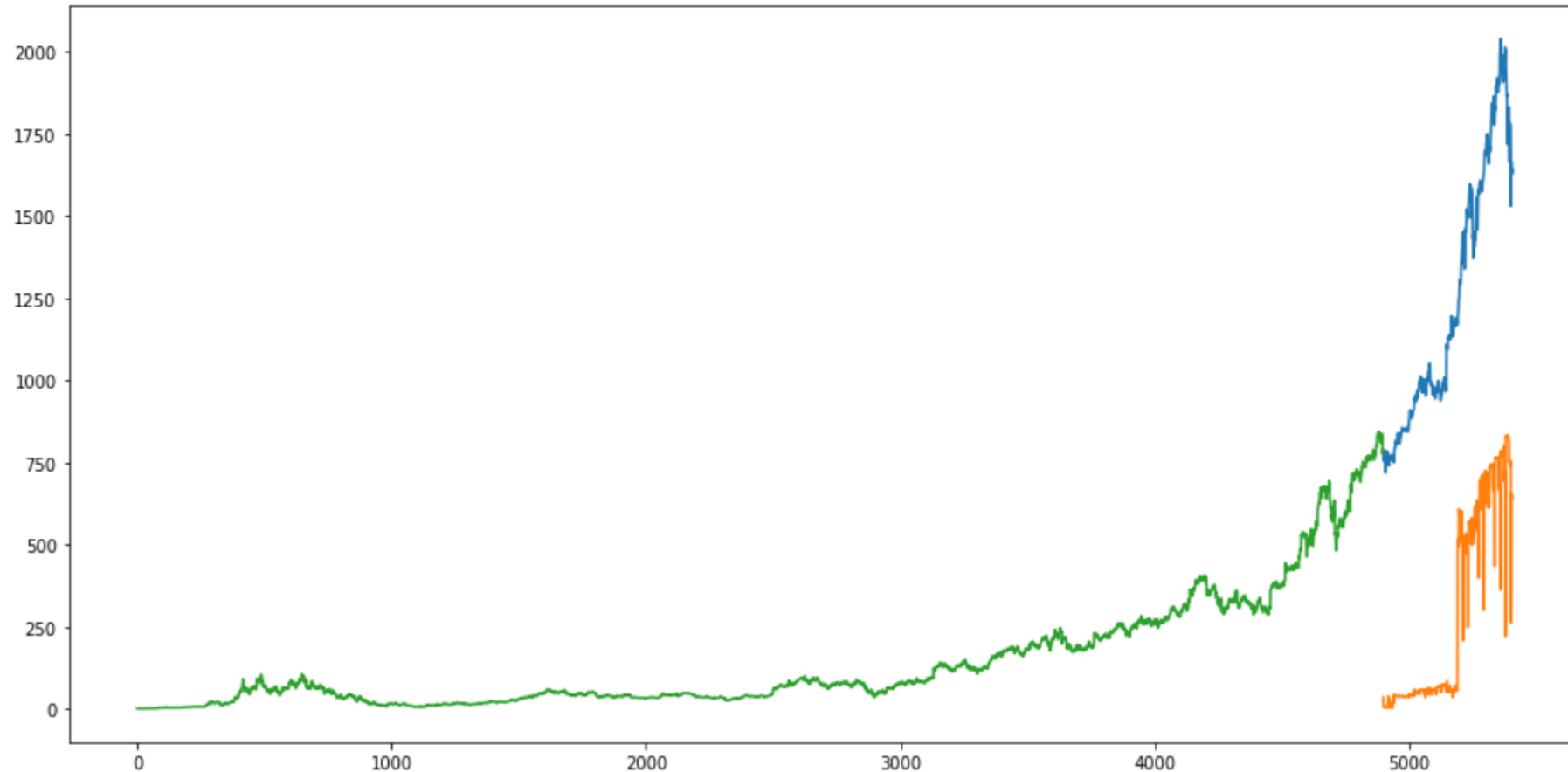
Once again a huge RMSE of 957.12, means that k-Nearest Neighbours also did not perform well. The RMSE value is almost similar to the linear regression model, however we see that its output pattern is not very different from the actual ones. But its only the pattern, not the actual numbers. We will now move on to the next method for forecasting and see how it works.

Prediction using **Long Short Term Memory (LSTM)**
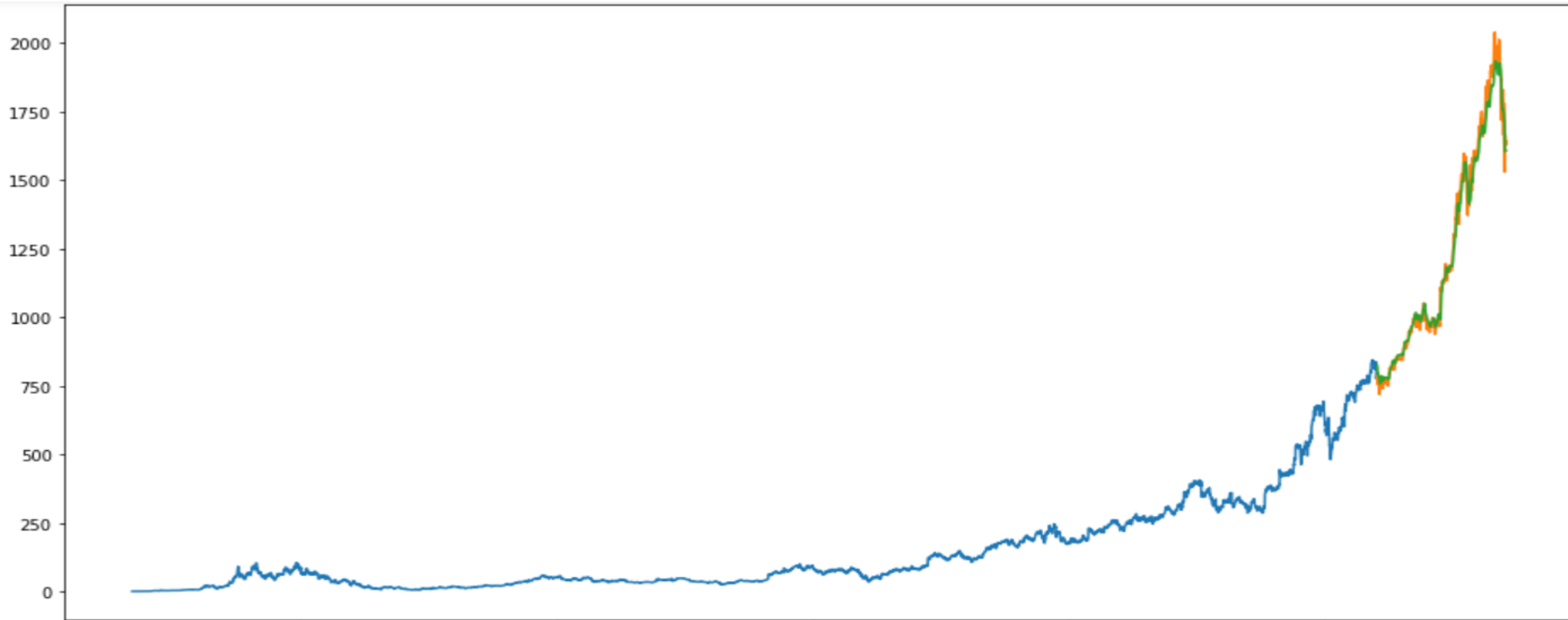
RMSE of 40.25!!!! That's a great reduction in RMSE. LSTM has easily outshone any algorithm we saw so far. The LSTM model can be tuned for various parameters such as changing the number of LSTM layers. But are the predictions from LSTM enough to identify whether the stock price will increase or decrease? Certainly not!

As I mentioned at the start of the presentation, stock price is affected by the news about the company and other factors like demonetization or merger/demerger of the companies. There are certain intangible factors as well which can often be impossible to predict beforehand.

As an end note, I will try to find a way to use these tangible / intangible factors as a further enhancement on my project.