Subject: Business Analytics

Project Paper Title: Stock Price Prediction

Author: Arusyak Hakobyan

Abstract

This paper discusses time series forecasting, specifically stock price prediction by using methods such as moving average, ARIMA, and LSTM. It explains what is time series, then what is financial analysis and why it matters to be able to predict the stock price in the next time period. We first give an introduction to all methods, in brief, to prepare the ground for the topics discussed in this paper. Then we consider their applications, provide in-depth explanations, and compare the results.

Keywords: financial analysis, stock, price, prediction, time series, deep learning, artificial intelligence, ARIMA, LSTM

Introduction

Time series prediction is an important area of machine learning. When tracking business metrics, monitoring industrial processes we are facing with time series data. These data can be described as a sequence of observations stored in time order. So time-series analysis is a method to analyze time-series data to get meaningful characteristics of that data and generate other useful insights which can be applied in a business situation(Lyla, 2019). Time series forecasting is a technique used across many fields of study not only in business. It is greatly used in geology, economics and many more. The method basically predicts future events by analyzing the trends of the past, having an assumption that future trends will be similar to historical ones. The various applications of time-series forecasting include weather forecasting, shape detection, earthquake prediction(What is time series forecasting? - Definition from WhatIs.com).

In this paper, we are going to concentrate on financial analysis. It is the process of assessing a company's performance and making recommendations about how it is going to perform in the future(Financial Analysis - Overview, Guide, Types of Financial Analysis). To keep it short, it uses historical data to generate projections about the future. Financial analyses provide insights into the organization's current and future state. Thus it is a critical aspect of the company's success as it highlights both the weaknesses and strengths which directly affect competitiveness(What is Financial Analysis?)

The objective of stock price prediction is to find the future value of company stock. Predicting successfully the future price can result in a significant profit. The EMH(efficient-market hypothesis) states that the stock price is unpredictable as it fully reflects all available information. Generally, there are three methods for predicting the stock price. First

is fundamental analysis, which is only concerned with the company's fundamentals. The second one is technical analysis, which is the "opposite" of the first one, meaning it is not concerned with the company that underlies the stock. It predicts the future price of a stock based solely on the past trends of the data. And the last one is machine learning, which is using ANNs(common forms: RNNS, TDNN,) or GAs(Stock market prediction, 2019).

Related work

I have taken some courses from datacamp which gave clear understanding of how to work with financial data. The courses I completed are:

https://www.datacamp.com/courses/intro-to-python-for-finance

https://www.datacamp.com/courses/intro-to-financial-concepts-using-python

https://www.datacamp.com/courses/importing-managing-financial-data-in-python

I read many articles in Medium which gave me ideas such as

1. https://medium.com/ml-everything/predicting-the-stock-market-p-hacking-and-why-you-should-be-bullish-90fddc583838
    - trying simple method before getting into neural networks.
2. https://blog.goodaudience.com/introduction-to-1d-convolutional-neural-networks-in-keras-for-time-sequences-3a7ff801a2cf
    - further development of this paper could be CNNs.
3. https://www.datacamp.com/community/tutorials/lstm-python-stock-market
    - LSTM is a good choice for financial data analysis

And one of the articles served me as a guide for implementing:

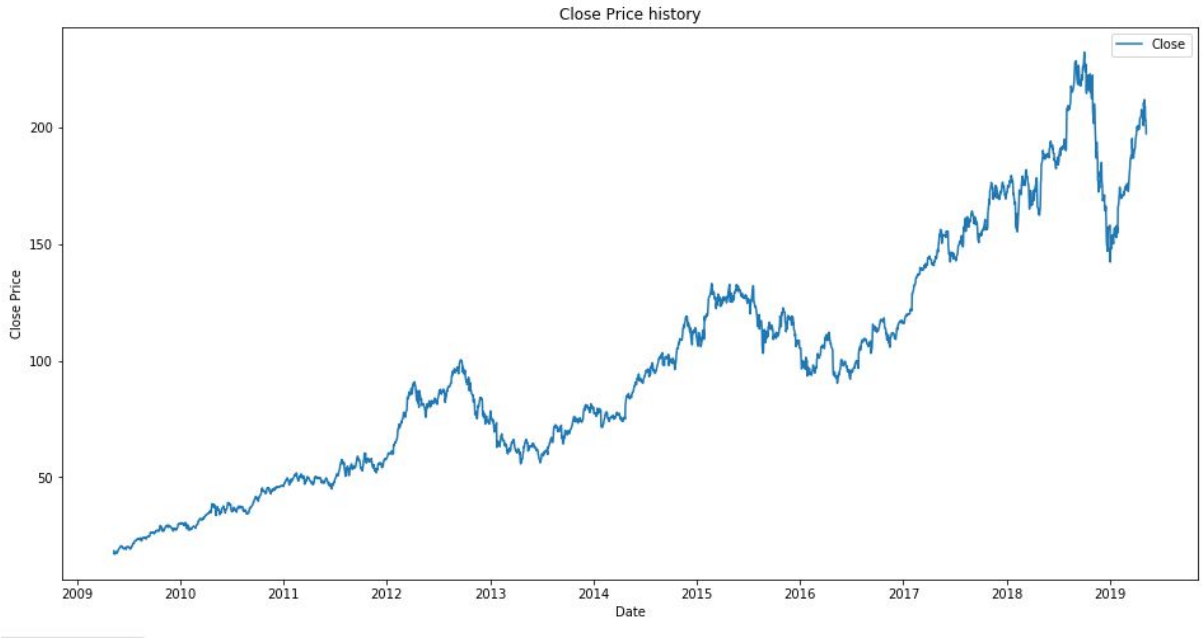
https://www.analyticsvidhya.com/blog/2018/10/predicting-stock-price-machine-learningnd-deep-learning-techniques-python/

Dataset and features

The dataset we are going to work with is the historical data of Apple Inc,, which is an is an American multinational technology company founded back in 1976. We have taken daily data for 10 years starting from 2009/11/05 to 2019/11/05. Overall containing 2518 observations and 7 features. The features Open and Close represent the starting and closing price at which the stock is traded on a particular date. High, Low and Adj Close stand for the maximum, minimum, and the last price of the share for the specified date, respectively.

Date	Date	Open	High	Low	Close	Adj Close	Volume
2009-05-11	2009-05-11	18.195715	18.708570	18.160000	18.510000	12.295175	101164700
2009-05-12	2009-05-12	18.508572	18.530001	17.607143	17.774286	11.806481	152370400
2009-05-13	2009-05-13	17.601429	17.717142	17.054285	17.070000	11.338662	148992900
2009-05-14	2009-05-14	17.111429	17.647142	17.100000	17.564285	11.666988	111956600
2009-05-15	2009-05-15	17.474285	17.802856	17.372858	17.488571	11.616698	91891800

Date	Date	Open	High	Low	Close	Adj Close	Volume
2019-05-06	2019-05-06	204.289993	208.839996	203.500000	208.479996	207.680222	32443100
2019-05-07	2019-05-07	205.880005	207.419998	200.830002	202.860001	202.081787	38763700
2019-05-08	2019-05-08	201.899994	205.339996	201.750000	202.899994	202.121628	26339500
2019-05-09	2019-05-09	200.399994	201.679993	196.660004	200.720001	199.949997	34908600
2019-05-10	2019-05-10	197.419998	198.850006	192.770004	197.179993	197.179993	41183400



Methodological approach

We are going to use a couple of approaches for predicting stock price. We will start from a simple model called moving averages. Then we will make it a little bit complex by implementing autoregressive integrated moving average(ARIMA) and lastly, we will try using artificial recurrent neural network architecture called long short-term memory(LSTM).

Experiments and results

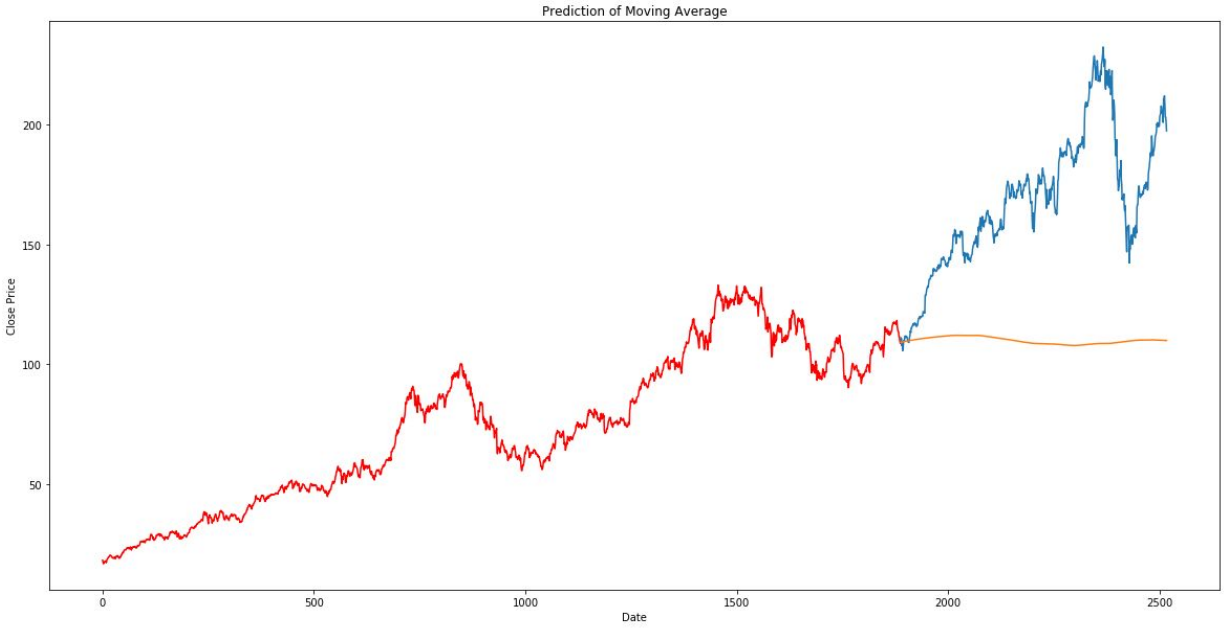
- ***A simple method using averages***

One of the most common things we use for measuring something is the average. For example, teachers usually calculate the average marks of each student to get their overall performance score, or we are calculating the average temperature of the past week to compare with today's temperature. So for simplicity let's first start from the simple method.

Our averaging algorithm is constructed so that the predicted closing price for each day is the average of previously observed values. We are not going to use a simple average, instead, we

are using the moving average technique that is our predicted value will be considered as an input for the next iteration.

As we can see from the graph our prediction is far from reality, however, it succeeds in predicting the increasing trend. The training data range is tiny compared to that of testing. And as our predictions lay in training data range, our moving average algorithm has not very promising results. The RMSE confirmes the failure of our model, by being more than 60



- **ARIMA**

  One of the most popular statistical methods for time series forecasting is ARIMA. It takes into account the past values to predict future values as all of the moving average did, however, there are some important parameters to be specified in ARIMA:

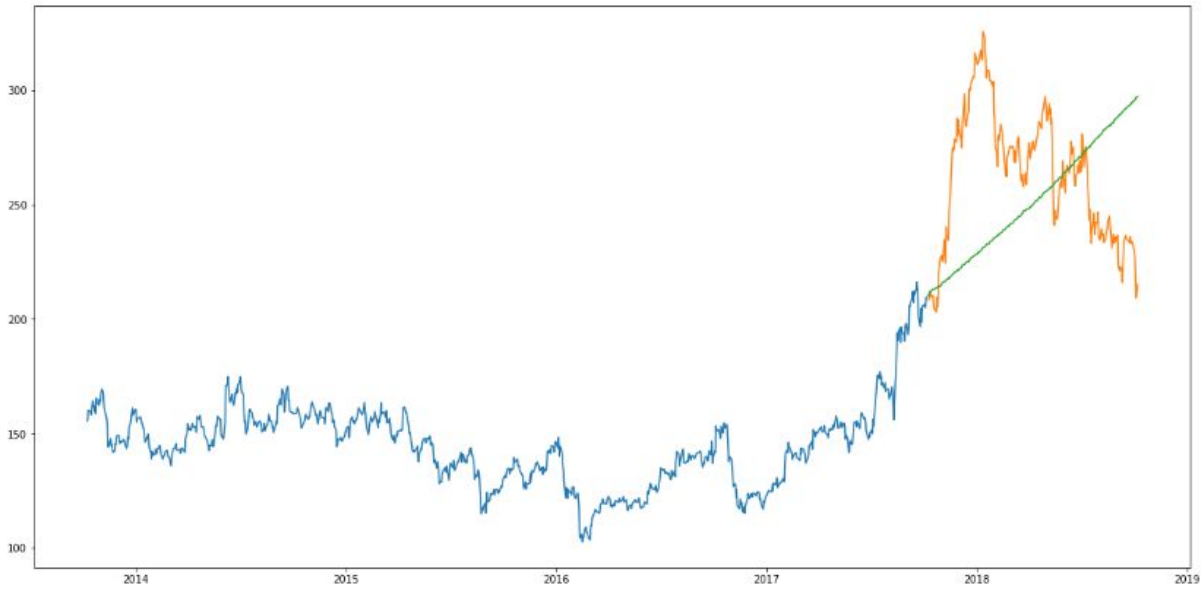
  ➔ p - number of time lags of the autoregressive model

  ➔ q - order of the moving-average model

  ➔ d - differencing order

As tuning of the parameter consumes lots of time, we will use auto ARIMA which automatically selects the best combination of parameters that provide the least possible error(Autoregressive integrated moving average, 2019)(Fuqua School of Business).

As we can see from the graph below, ARIMA has outperformed the moving average algorithm described above, as here the range of prediction and real values are the same. However, here also the prediction only succeeded in capturing the increasing trend and nothing more. The RMSE is close to 50, twice better than in our first approach.

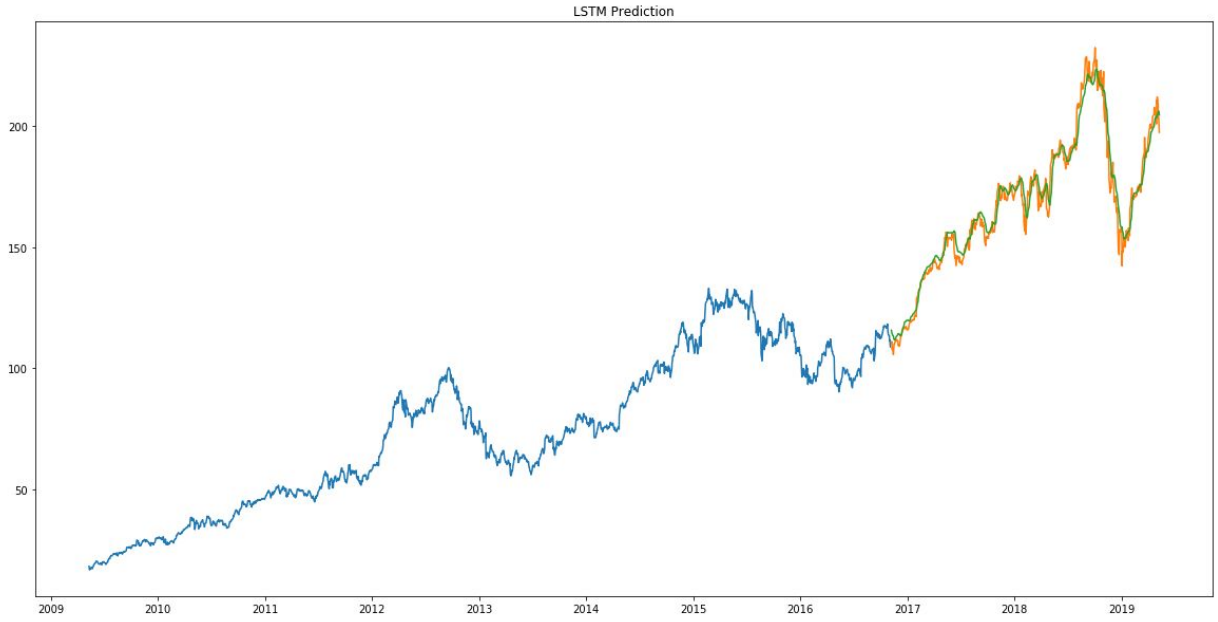


- **LSTM**

LSTM is a special kind of RNN which is capable of learning long-term dependencies. Practically the model's default behavior is remembering information for long periods(What is LSTM?, 2018). For understanding what is LSTM, let us first get the concept of recurrent neural networks(RNNs). All RNNs have the form of a chain of repeating modules of the neural

network. In standard RNNs, this repeating module will have a very simple structure, such as a single tanh layer. Yet, LSTMs has a different structure of the repeating module, instead of having a single neural network layer, there are four interacting ones(Understanding LSTM Networks).

The plot below shows that the predictions and actual results are in the same range and have the same shape. Even though the prediction is not a perfect one, the model has reached the state of understanding the movement of the future price fluctuation. The is a work for improving the prediction of the fluctuation rate. But overall the model has RMSE of around 5.



Review and future plans

- **Review/*Comparison***

  While trying three different methods, we managed to improve the performance in each approach. In case of moving average algorithm, we had RMSE of 60 and then changed that into ARIMA we got a almost twice better result with RMSE of 40, and our last step

of introducing neural networks, LSTM gave us 8 times better than ARIMA with RMSE of around 5.

- ***Future plans***

For later analysis, it would be great to see how other deep learning methods such as convolutional neural networks perform in financial data prediction. Moreover, not only one-dimensional CNNs, but also multi-dimensional ones should be taken into account.

References

Autoregressive integrated moving average. (2019, April 17). Retrieved from https://en.wikipedia.org/wiki/Autoregressive_integrated_moving_average

Financial Analysis - Overview, Guide, Types of Financial Analysis. (n.d.). Retrieved from https://corporatefinanceinstitute.com/resources/knowledge/finance/types-of-financial-analysis/

Fuqua School of Business. (n.d.). Retrieved from https://people.duke.edu/~rnau/411arim.htm

Lyla, Y (2019, January 02). A Quick Start of Time Series Forecasting with a Practical Example using FB Prophet. Retrieved from https://towardsdatascience.com/a-quick-start-of-time-series-forecasting-with-a-practical-example-using-fb-prophet-31c4447a2274

Stock market prediction. (2019, May 10). Retrieved from https://en.wikipedia.org/wiki/Stock_market_prediction

Understanding LSTM Networks. (n.d.). Retrieved from https://colah.github.io/posts/2015-08-Understanding-LSTMs/

What is Financial Analysis? (n.d.). Retrieved from https://www.financialplannerworld.com/what-is-financial-analysis/

What is LSTM? (2018, April 19). Retrieved from https://hub.packtpub.com/what-is-lstm/

What is time series forecasting? - Definition from WhatIs.com. (n.d.). Retrieved from https://whatis.techtarget.com/definition/time-series-forecasting