INFO 7390 - Advances in Data Sciences and Architecture

Stock Price Prediction using MLP and ARIMA

Ashish Bais (001221790)

ABSTRACT

Stock Price Prediction has always been a hot topic of discussion and research for investment firm, brokerage firms and other financial institutions due to highly volatile nature of stock prices. Many institutions and researchers have tried to develop better predictive model over the years. With more and more people interested in stock and securities investments, the need of stock price prediction system is more than ever before.

The purpose of this project is to develop a stock price prediction model to accurately predict the stock prices based on the historical stock market data. The dataset used consist of the largest first 500 stocks by market capitalization termed as Standard and Poor's 500 (S&P500) listed at New York Stock Exchange (NYSE) with total market capitalization of $23.7 trillions as of April 2018.

For better predicting the stock prices of S&P500 stocks, Multilayer Perceptron (MLP) and Autoregressive Integrated Moving Average (ARIMA) to calculate the accuracy of predicting the stock prices as a function of time have been applied and tested in this paper. Published stock data for S&P500 stocks was used for training and developing stock price prediction model. The results of developed prediction model show a high accuracy for prediction of future stock prices based on the historical S&P500 stock data.

Introduction

Stock Price Prediction has been a dream of all investors since its inception of Stock Markets. Today, stocks are widely considered as a valuable investment alternative by both financial experts as well as common people. However, efficiently investing in stock is a complex process and needs the understanding and critical analysis of multiple factors. It is a time-consuming process and involves the complexities often acts as deterrents in the stock investment world. Thus, forecasting the stock price is regarded as one of the most difficult task to accomplish in financial and investment industry due to complex nature of stock markets.

Generally, the prediction model/system helps users with identification of well informed choices from large pool of options by reducing the complexity of decision making process for them. It provides vital information to concerned user by using set of data analysis algorithms and techniques. Stock Price Prediction system works on the same principal.

Stock Price Prediction has been a central issue and goal of many research organizations, investment firms, and financial institutions from many years. Various formal and informal techniques have been developed and utilized to predict stock prices and market trends. Techniques ranging from hand written plots to modern Long short-term memory (LSTM) have been tried to resolve this problem. Gaussian kernelled SVM provided the 64% accuracy in predicting stock prices. The accuracy of 50% is achievable using neural network solutions. Use of this advanced approach have been helpful in few ways however still lacks the accuracy to predict the stock prices and has not been entirely successful.

In this paper, the prediction model was aimed at enhancing the profitability opportunities for users by predicting the future stock prices based on historical S&P500 stock market data. Multilayer Perceptron (MLP) and

Autoregressive Integrated Moving Average (ARIMA) were applied to train and test the prediction model to forecast the stock prices in terms of accuracy in predicting future stock prices with respect to time.

Multilayer Perceptron (MLP)

MLP is a feed forward artificial neural network which maps set of input data onto a set of appropriate output data. It consists of multiple layers (input layer, no or more hidden layers and output layer) with fully connected to next layer with neurons, variables, placeholders, and non-linear activation functions. Each layer passing its output to next layer as input. MLP uses supervised learning algorithm called as backpropagation.

A basic architecture of feed forward MLP is described below.

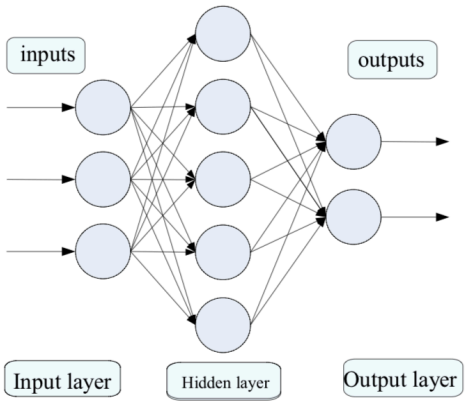

Figure 1: Architecture of a feed forward MLP

For stock price prediction model,
- Input is historical stock market data
- Each circle represents neuron which processes input stock data
- Output of each neuron from input model sent to each neuron in next hidden layer
- Three hidden layers in project model
- Output is the future prices of stocks

Following steps were performed for developing stock price prediction model using MLP:
- Importing S&P500 data
- Setting variable
- Assigning placeholder
- Initializing weight and bias
- Assigning input, hidden (3) and output layers
- Calculating and minimizing error
- Selecting training and testing data
- Calculating error and accuracy

Autoregressive Integrated Moving Average (ARIMA)

Autoregressive Integrated Moving Average (ARIMA) models are statistical analysis models which uses time series to predict future trends. It most widely used as a robust and efficient model for short term predictions in financial time series forecasting. This model is already being extensively used in the field of finance and economics and regarded as most prominent model in financial forecasting. ARIMA model was used in predicting the future stock prices based on the historical stock market data for S&P500.

ARIMA model, also referred as Box-Jenkin methodology was introduced by Box and Jenkins in 1970. It involves the set of activities for identifying, estimating and diagnosing the model with time series data for estimating the short-term forecasts. Using ARIMA model, the future price of stocks can be represented as a linear combination of past values and past errors as

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \ldots + \phi_p Y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \ldots - \theta_q \varepsilon_{t-q}$$

where
Yt is the actual stock value;
Et is error at time t;
Oi and Oj are coefficients
p and q are autoregressive and moving average integers respectively.

Prediction model identification, parameters and variables estimation, and testing were the steps involved in building ARIMA model for stock price prediction.

ARIMA was used to identify best performing stocks for a particular period.

Following steps were performed for developing prediction model using ARIMA:

- Selecting stocks from dataset
- Setting ARIMA model
- Allocating train and test data
- Training model
- Predicting forecast stock prices
- Calculating mean square error between predicted and test stock prices
- Calculating the accuracy for predicting stock prices as a function of time
- Identifying best performing stock based on prices as a function of time

DATASET

For this project, we have utilized published Standard and Poor's 500 (S&P500) stock data. This dataset represents the stock prices of largest first 500 stocks listed at New York Stock Exchange (NYSE) based on their market capitalization. These stocks together had market capitalization of $23.7 trillions (As of April 2018).

Dataset	No of stocks	Size (Rows, Columns)
S&P 500	500	41266, 501

Dataset was already cleaned and prepared with no missing values. Dataset was split into training and test data. 80% of total dataset was used as training data and remaining 20% of dataset was used for testing.
The dataset was available and used in csv format.

RESULTS

Results using MLP Model:

The result of applying the MLP on S&P500 dataset are given below:

The final error in predicting the future stock prices was 0.951563%.

The accuracy for predicting the future stock prices was 99.0484373644%.

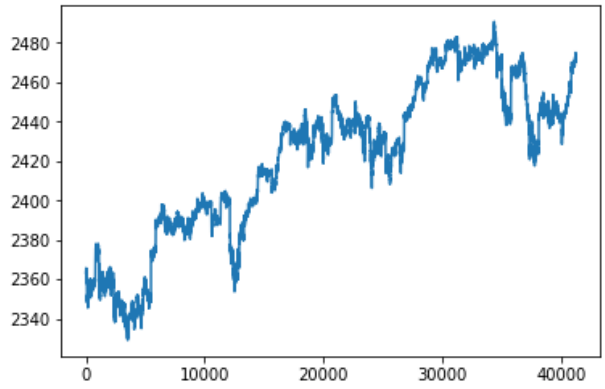

Figure 2: Plot of stock price for all line items of stock 1

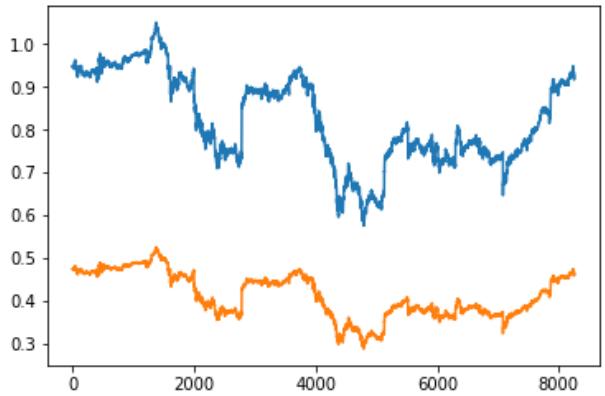

Figure 3: Plot of test price and predicted price of stock 1

Please note that difference in plot of accuracy is due to scaling. Though the difference is negligible, scale makes it look bigger.

Results of ARIMA Model:

With ARIMA model, calculated the error, accuracy and mean forecast value for below selected first 5 stocks.

Result of Stock 1:

```
Mean Square error for stock 1 is: 9.91366371956
Accuracy percentage for stock 1 is: 99.6561157857
Mean forecast value of stock 1 is: 2400.93986332
```

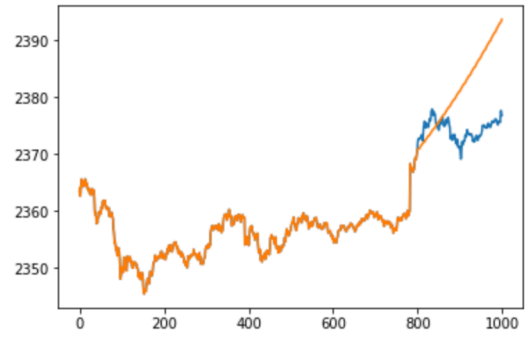

Figure 4: Plot of test price and predicted price of stock 1

Function of Accuracy with respect to time for stock 1:
2.922e-07 x^3 - 9.338e-05 x^2 + 0.004247 x + 99.89

Result of Stock 2:

```
Mean Square error for stock 2 is: 0.576715963766
Accuracy percentage for stock 2 is: 98.8095277832
Mean forecast value of stock 2 is: 41.1355424515
```

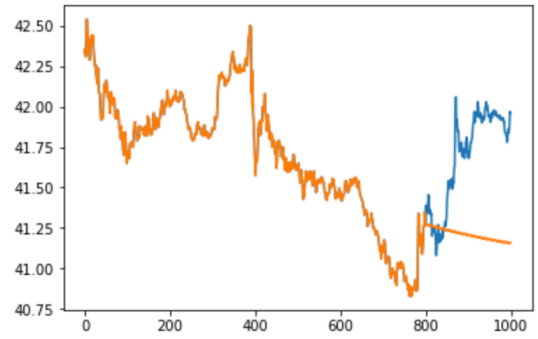

Figure 5: Plot of test price and predicted price of stock 2

Function of Accuracy with respect to time for stock 2:
8.611e-07 x^3 - 0.0001911 x^2 - 0.003318 x + 99.96

Result of Stock 3:

```
Mean Square error for stock 3 is: 0.297550819552
Accuracy percentage for stock 3 is: 99.8188711269
Mean forecast value of stock 3 is: 146.121254015
```

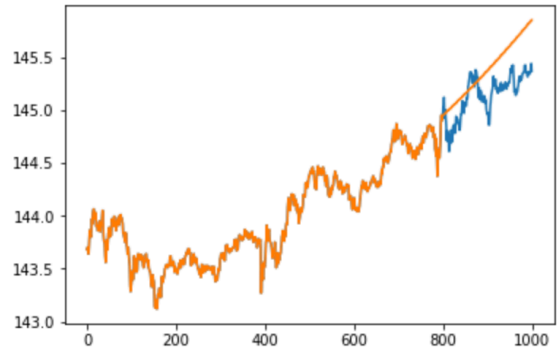

Figure 6: Plot of test price and predicted price of stock 3

Function of Accuracy with respect to time, for this stock is:
9.369e-08 x^3 - 3.664e-05 x^2 + 0.00279 x + 99.84

Result of Stock 4:

```
Mean Square error for stock 4 is: 0.23897150894
Accuracy percentage for stock 4 is: 99.8400083245
Mean forecast value of stock 4 is: 131.723542088
```

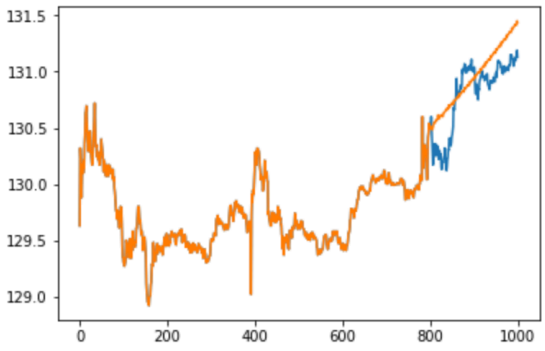

Figure 7: Plot of test price and predicted price of stock 4

Function of Accuracy with respect to time, for this stock is:
-1.099e-07 x^3 + 1.867e-05 x^2 + 0.0002829 x + 99.78

Result of Stock 5:

```
Mean Square error for stock 5 is: 0.816637474904
Accuracy percentage for stock 5 is: 99.0874934817
Mean forecast value of stock 5 is: 82.7256086548
```

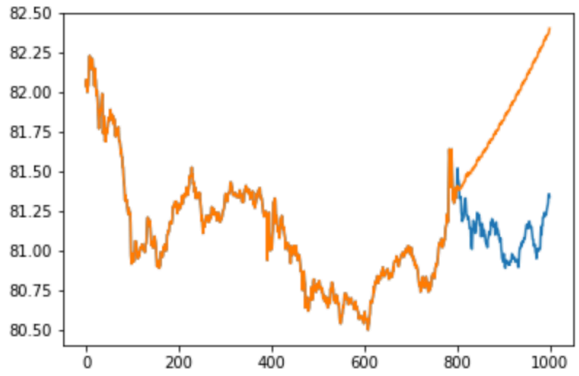

Figure 8: Plot of test price and predicted price of stock 5

Function of Accuracy with respect to time, for this stock is:
$2.035e\text{-}07\ x^3 - 3.328e\text{-}05\ x^2 - 0.007708\ x + 99.89$

Using ARIMA model, identified the best performing stock based on mean forecast value of stock. For five stocks analyzed, the result of best performing stock was stock no 1.

CONCLUSION

The stock price prediction models using MLP and ARIMA are presented in this paper. The accuracy of predicting the future stock prices with respect to time function was calculated using both models independently. From above result, concluded that the accuracy for predicting the future stock prices using MLP was 99.05% and the average accuracy of predicting future stock prices using ARIMA was 99.44%. Thus, ARIMA model was more accurate than MLP model in predicting future stock prices.

The mean error in predicting the future stock prices using MLP was 0.95% and the mean error in predicting the future stock price using ARIMA was 1.97%. Thus, the mean error in predicting the future stock prices is more in ARIMA model than MLP model.

Further, noted that time required for predicting future stock prices using MLP was around 12-15 seconds. However, the time required for predicting future stock prices using ARIMA was around 10 mins. Hence, the MLP model is much faster in predicting future stock prices than ARIMA model. For MLP model, the curve of accuracy of predicting the future stock prices is plateau in less time than ARIMA model.

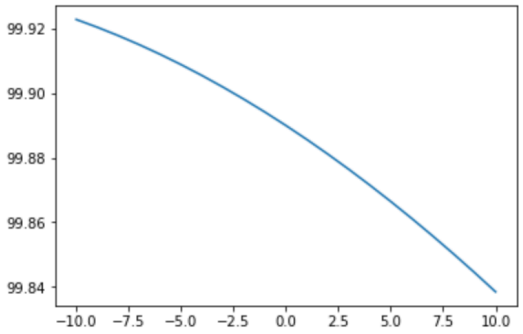

Figure 9: Plot of accuracy with respect to time for MLP

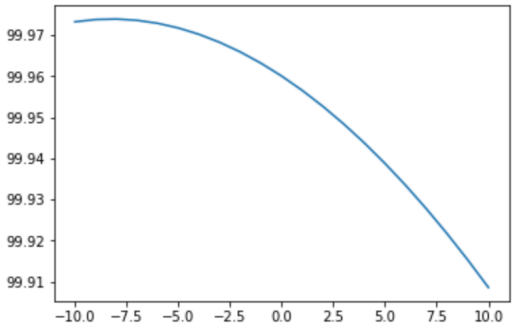

Figure 10: Plot of accuracy with respect to time for ARIMA

Figure (9) and (10) confirms that the accuracy of predicting the future stock prices was dropped quickly in MLP model than in ARIMA model.

Thus, if accuracy of predicting the future stock prices is more important, ARIMA model is a better choice.

REFERENCES

[1] Volodymyr Turchenko, Patrizia Beraldi, Francesco De Simone, Lucio Grandinetti "Short-term stock price prediction using MLP" September 2011.

[2] A. Victor Devadoss, T. Anthony Alphonnse Ligori "Forecasting of Stock Prices Using Multilayer Perceptron" December 2013.

[3] Mahdi Pakdaman Naeini, Hamidreza Taremian, Homa Baradaran Hashemi "Stock Market Value Prediction Using Neural Networks", 2010

[4] Ayoele A. Adebiyi, Aderemi O. Adewumi, Charles K. Ayo "Stock Price Prediction Using ARIMA Model" 2014

[5] Lin H. "Stock Market Prediction Investigation" 2013

[6] Arik S., Eryilmaz B. 2013. Stock Prediction and Portfolio Optimization

[7] https://en.wikipedia.org/wiki/Multilayer_perceptron

[8] Jason Brownlee "How to create ARIMA model for time series forecasting in python" January 2017