National University of Computer and Emerging Sciences

AI Term Project

CS-461

Stock Market Prediction

Group Members:

L-174467 Tehseen Ullah Kahloon
L-174378 Ahmed Mamoon Tahir
L-174250 Hammad Shahid

Course Instructor:

Dr. Kashif Zafar

Abstract:

Predicting stock exchange values is a topic that has garnered a lot of interest from various researchers. In this regard, the role of machine learning and general artificial intelligence has been coherently studied for the purpose of making accurate predictions. Concepts like the Support Vector Machine are quite popular when it comes to this domain. In the following project, we analyze the effectivity of prediction models, taught as per the course syllabus, for the stock exchange. The methods we focused on include the Multi-layer Perceptron, Linear Regression and the Recurrent Neural Network (specifically LSTM). We also decide to explore other algorithms outside those taught in the coursework, and therefore, incorporated Random Forest into our analysis. This paper will highlight our findings and provide a coherent interpretation.

Introduction:

Indeed, providing accurate predictions on the behavior of stocks is a challenging task. A lot of factors play a role and these pertain to both the seemingly random behavior and the physical aspects. These factors amalgamate whereby propagating volatility in values and countering accuracy.

- **Stock Market and Stock Exchange?**

  The stock market is simply a place where actions can be performed on shares of companies. It is a market, like any other, where the involved companies have opened their shares to the general public; hence, enabling them to buy and sell these shares. It follows a proper manner of conduct in the form of exchanges and OTC methods (Over the Counter). These exchanges operate on the national and international level with various trading emplacements for proper transactions.

  Stock exchange is classified as being a part of the Stock market. It does not exist on its own, rather it’s a part of the larger system. Trading in the stock market pertains to subjecting shares to various transactions through multiple stock exchanges in the market.

- **Predictions on the Stock Market?**

  When we say ‘predicting the stock market’, we are essentially talking about the value/equity of a share (translated as a stock). It is an important part of how businesses throughout the world operate. Forecasting prevalently decides the standing of the company in relation to its competitors.

Problem:

In this part of the paper, we will strictly define our problem statement and the domain under which we will be working:

- **Fundamental Analysis:**
  Fundamental Analysis involves analyzing the company's future profitability on the basis of its current business environment and financial performance.
- **Technical Analysis:**
  Technical Analysis, on the other hand, includes identifying trends in the stock exchange marketplace.

As you might have guessed, our focus will be on the technical analysis part. We'll be using a dataset from various sources. The profit or loss calculation is usually determined by the closing price of a stock for the day; hence we will consider the closing price as the target variable.

Stock Market Analysis:

Analysis is performed vigorously, in different forms; so as to provide accurate and reliable predictions.

Using Artificial intelligence and Machine Learning:

A number of artificial intelligence and machine learning techniques have been used over the past decade to predict the stock market. Genetic Algorithm, Neural Networks, Linear Regression, Long-Short Term Memory, Moving Average, K-nearest Neighbors, and many more models. And there is various dataset available online on Kaggle, Quandl, etc. Now, we need to select models and get dataset for our models.

Corpus:

The corpus contains data on the TATA Beverages stock. It was taken from Kaggle and all the methods have been tested on the same dataset.

	Date	Open	High	Low	Last	Close	Total Trade Quantity	Turnover (Lacs)
0	2018-09-28	234.05	235.95	230.20	233.50	233.75	3069914	7162.35
1	2018-09-27	234.55	236.80	231.10	233.80	233.25	5082859	11859.95
2	2018-09-26	240.00	240.00	232.50	235.00	234.25	2240909	5248.60
3	2018-09-25	233.30	236.75	232.00	236.25	236.10	2349368	5503.90
4	2018-09-24	233.55	239.20	230.75	234.00	233.30	3423509	7999.55

Models and Comparison:

For comparison purposes, root mean square error for each method has been taken into account. This value will provide a clear insight into the performance and effectivity of each process, on a relative scale.

Linear Regression:

Linear Regression is a popular machine learning technique which aims to provide us with a relationship between an independent variable and a dependent variable(s). Simple Linear Regression has been applied on the given problem. Basically liner regression works by finding the best fit line between the two variables. As the name suggests, the relationship being derived is to be linear, in nature. Hence, the relating equation is generated by first deducing the slope, and then the intercept.

The root mean squared value of this implementation is as follows:

```
print('Root Mean Squared Error is: ',np.sqrt(np.mean(np.power((np.array(testY)-np.array(predictions)),2))))
Root Mean Squared Error is:  34.42194737691777
```

MLP:

Moving on, we decided the analyze the impact of a multi-layer perceptron on the corpus. A multilayer perceptron is classed as a feed forward neural network. It consists of various neurons in the overall structure; divided into the input layer, the output layer and the hidden layer which performs the most important part of the implementation. The neurons are connected to each other and data is propagated as per the activation function. This data propagation is highly dependent of the weights of the architecture, therefore, there is a need to optimize them. In our implementation, the optimization algorithm being used was back propagation. This algorithm works by gradient descent and starts at the output layer, moving backwards.

The root mean squared value of this implementation is as follows:

```
# Find Root Mean Squared Error

np.sqrt(np.mean(predictions-testY)**2)

8.495644771475897
```

LSTM:

Long Short Term Memory is a variant of a Recurrent Neural Network (RNN) but proves to be better in the sense that it accounts for the disappearing gradient values. Essentially, it makes use of memory (for information) in its process. Like the basic RNN, LSTM works ideally in processing and predicting data that is sequential.

In our implementation, the following root mean squared value was generated (for a very small number of epoch):

```
# Find Root Mean Squared Error

np.sqrt(np.mean(p-testY)**2)
```

```
6.5503176171305135
```

Results and Conclusion:

Method	RMSE
Linear Regression	34.4
Multilayer Perceptron	8.50
LSTM	6.55

From the summary table above, we can see that LSTM clearly outperforms the rest of the algorithms when it comes to stock prediction for the targeted closing values of stock. Perhaps its dominance lies in its ability to fine tune parameters the given task. The simpler multilayer perceptron also gave somewhat promising results. However, linear regression performed extremely poorly. This clearly indicates that Linear Regression is not suited for the task of Stock Prediction.

Additional Work:

In our efforts to further our knowledge regarding this interesting domain, we came across several literatures which supported a method called **Random Forest** as being the best for stock prediction. This method works by generating decision trees and then combines them together to give the prediction. It falls under the category of supervised learning algorithms and indeed, produced the most ideal and promising predictions when compared to algorithms falling in the coursework. The root mean squared error for this algorithm is as follows:

```
# Find Root Mean Squared Error

np.sqrt(np.mean(predictions-testY)**2)
```

```
0.40544348894349397
```

References:

- https://www.analyticsvidhya.com/blog/2018/10/predicting-stock-price-machine-learningnd- deep-learning-techniques-python/
- https://github.com/mwitiderrick/stockprice/blob/master/NSE-TATAGLOBAL.csv
- https://ieeexplore.ieee.org/abstract/document/8212715
- https://www.researchgate.net/profile/Kumkum_Garg2/publication/238747905_A_Hybrid_Machin e_Learning_System_for_Stock_Market_Forecasting/links/00b7d53b4cfc215d81000000/A-Hybrid- Machine-Learning-System-for-Stock-Market-Forecasting.pdf
- https://scholar.google.com.pk/scholar?q=stock+market+analysis+using+machine+learning&hl=en &as_sdt=0&as_vis=1&oi=scholart
- https://www.datacamp.com/community/tutorials/lstm-python-stock-market
- https://www.investopedia.com/terms/s/stockmarket.asp
- https://capital.com/stock-market-prediction-definition
- https://www.investopedia.com/terms/s/stockmarket.asp
- https://www.ijert.org/study-of-machine-learning-algorithms-for-stock-market-prediction