Stock Price Prediction Using Data Analytics

Mrs. Teena Varma [#1], Myron D'souza [#2], Hasan Husssain[#3], Ricky Mathews [#4]

[1]Teena Varma: Professor, Dept. of Computer Engineering, Mumbai University
[2]Myron D'souza: Student, Dept. of Computer Engineering, Mumbai University
[3]Hasan Husssain: Student, Dept. of Computer Engineering, Mumbai University
[4]Elvis Braganza: Student, Dept. of Computer Engineering, Mumbai University
[5]Ricky Mathews: Student, Dept. of Computer Engineering, Mumbai University

---***---

Abstract: *The objective of this project is to show a way how to successfully predict the stocks for the next years to come. This system is going to work on a google dataset this system and algorithm is a RNN(Recurrent Neural networks) as there is a need of such algorithms that can predict a step ahead thus this model used in this project does and similarly this model can be used to predict other many dataset to show us the future outcomes not only for stocks but raw data come in greater uses. Here we are working on a google dataset. Predict its future stock*

Keywords: prediction, algorithm, recurrent, neural, networks, stork market,

Introduction

Stock market prediction is the act of trying to predict for system the future value of the company stock or other financial instrument traded on an exchange. when successful prediction of company then stock's future price could give significant profit to a company or share market investors. And also tells us which companies may end up not with a good stock. Which also helps share market investors.

There is a lot of complications and financial indicators and also the fluctuation of the stock market, is very useful to which makes for the investors to understand, analyze to make a good profit . But, with help of technology opportunity to maximize the profit from stock market investment has increased as the stork market works in such a way it is its backbone. Experts with help of technology can make better decisions over stock market prediction.

A stock of a company, is all the shares into which ownership of the corporation is divided. Shares are collectively known as stock. A single share of the stock represents fractional ownership of the corporation to the total number of shares. It entitles one to a part of that company's earning and asset. Which also determine whether the company can make profit down the road

So, RNNS or Recurrent Neural Networks have proven to be one of the most powerful models for naturally processing sequential data and LSTM or long short-term memory is one of the most successful one RNN architecture. The LSTM introduces a different type memory cells a unit of computation for replaces the traditional artificial neurons in the systems hidden layers of the interconnected network. With these memory cells networks working together hand in hand are able to effectively associate certain memories and input remote in time. Therefore, they suit to adopt such a structure of the data dynamically over time with high prediction capacity.

1. The project or model needs to be made without any money spent at the table. 2. The model used to predict a certain prediction while only using very little memory. 3. The model used for train also should run with minimum RAM and processing power. 4. The dataset for the model should be selected keeping the above following parameters considered. 5. The model needs to be trained and tested in the as the following and the output is a prediction the result can vary in real world situation

1. Literature survey

As per the said to load an actually dataset of google and would also minimum memory consumption and be the most appropriate for our model. For the purpose we analyzed various time intervals of the datasets like at what the stock closed and what it opened at. The highs and lows during the day and the at most the volume on that date this dataset should have every day from a year to year mostly more than 3

1.1 DATA PRE-PROCESSING

This deep learning system is based like these on various libraries like numpy, pandas, skitlearn, matplotlib, datetime and also keras The data to be trained is first read its few time and time and then are displayed. Data preprocessing small step is carried out as each stepare very important , whether the data has any zero or null values or not. Checking for different datatypes of different columns and converting the object data type into float data type.as this algorithm are easy to work with other data types we consider a Seven day rolling mean concept which means in the simple example that for every single stock prediction we look seven days back collect for all the transactions that stay in this range so as to get the average of our column. This concept cannot result in any output for first six ambiguity of language is what makes fields. And creating first data frame to which it is of the training set.

Data preprocessing involves data minimization where we reduce a part of the not so importance for the numerical data. Then data transform and normalization is done.

Then data is cleaned and filling the null values in the mean then data set is divided into training and testing sets. Min max scalar is used here to transform features by scaling each of them to also a set range and here our feature will be range 0 to 1.

A data structure is created with 60 time stamps(60 days) and one output. Data of first 60 days is taken and then made prediction for the 61st day and so on, (2-61, result on 62nd). And reshape the data. Now, we chose such features that need to be put into neural networks which are high, low, date, close, volume. Importing keras library and packages is the first step building RNN.

Training a neural network from here data is going to fed to the neural network and trained for prediction. Here we assignment of random biases and weights to our model is done. This is our LSTM model which is composed of sequential input layer followed by three LSTM layers and a dense layer with enough activation and then finally a denser output layer with also linear activation function. Dropout is regularization technique for reducing the over-fitting in neural networks, it drops out a single unit in neural networks. Then we have our three layers and one output layer which 1 unit. Dropouts also decrease some error percentage in prediction.

After Compiling RNN with help of optimizer call Adaptive Movement Estimator. The Adam optimizer certainly combines with the help of two optimizers ADAgrad and RMSprop. Now, we can final fit RNN to such a training set. Epoch equals hundred and for the batch size is 32, epoch is frame of time, in machine learning it indicates number of passes will through the entire training data set .each Batch size is then number of training examples for utilized in a single iteration.

The output generated by RNN is compared with all the target value, the error or the difference between the target and the obtained value which is then minimized by using back propagation. algorithm which adjusts the weight and the biases of the network.

Now predictions are made and visualize the results, getting the real stock price of year to be predicted(2012). With help of the function 'iloc' rows and columns are selected in order of number in order that they appear in a single data frame. Finally, to get a predicted stock price, merging of training set and test set is should done so as to ensure both the sets values are present Rewriting data and X test list is given to predicted stock price. Taking inverse transform to get data out from matrix form. Then with help of we can matplotlib, we get visualized form of real stock and predicted stock, both rising up gradually or not.

1.2 IMPLEMENTATION OF PROPOSED METHODOLOGY

Before data was used for various ways pre-processing of data is necessary while handling trading or of any dataset they needed to be processed through and through. The various types options present to process this type of condition like:

1. Ignore the days where there is no trading and choose the days with trading.

2. Assign 0 to the days along with zero trading.

3. Build a model that can minimize the value for no trading.

A model should not be any determine the value for the days with no trading because that may change the values of calculation a significantly to the final error in the networks. Initial experiments were conducted, using techniques 1 and 2 above, and it was found that, technique 1 showed in lower error values. As a result, technique 1, from the above list, has been employed.

Libraries used are

Numpy-Numpy it is library for the Python programming language, adding support for large, multi-dimensional arrays and matrices and along with a large collection of high-level mathematical and functions to operate with these arrys.

Matplotlib- Matplotlib is a plotting library for the Python programming language which makes your dataset visualize and its numerical mathematics extension also NumPy.

Pandas- In computer programming, pandas is a software library written for the Python programming language for data preprocessing and analysis.

Scikit-learn- is a free software machine learning library for the Python programming language. It features various classification, regression and clustering algorithms including support vector machines.

RNN- A recurrent neural network (RNN) is a class of artificial neural networks where connections between nodes form a directed graph along a temporal sequence.like the dendroid

Keras- Keras is an open-source neural-network library written in Python. which are easy to use

2. FIGURES AND TABLES

2.1 Now that we have one generalized Trend, we create another function based on the loss window for seasonal extraction called STL (Seasonal, Trend and Loss) as shown in fig 5. The overview of both functions is shown in Fig 4
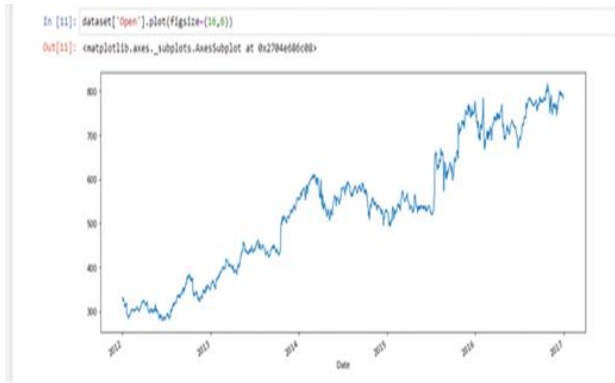
Fig 1: Summary of polynomial trend

Fig 2: Polynomial Trend

Fig 5: Time series decomposition of data
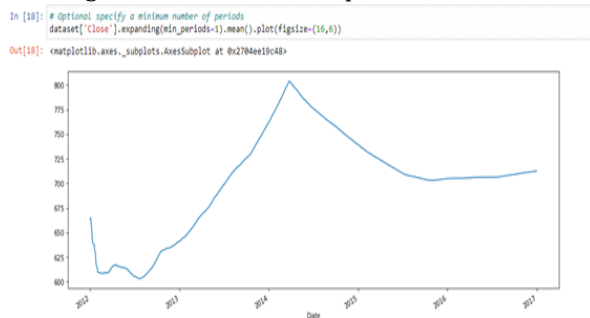
Fig 6: Polynomial and Time series trend with Actual data

3. Result

For accuracy we took data of 2012 and performed prediction for the rest of the time for which we had data. After that we compared the result of each method with the actual value of the stock and result was obtained. We predicted the open price of the stock and best method to forecast the opening price of stock is feed forward neural network. We also observed that for different stock different methods can provide better result and this is true for different types of prices (Open, low, high and low) too

- **Conclusion**

The popularity of stock market trading is growing rapidly which is encouraging researchers to find out new method for prediction which is not only helpful to researchers but also helps invertors or any person dealing with the stock market. In order to help predict the stock a forecasting model with a good accuracy is required. And in this project we have used one of the most precise predicting technology using RNN and LSTM units which help investors or analyst or any person interested in investing the stock market by providing them good knowledge of future situation of the stock market.

FUTURE SCOPE

The future scope of this algorithm is less as much advance algorithm are used in this place but the principal is the same as not only this process takes time and also processing power also so to tackle this the data is already processed and stored in servers

And when access is needed devices fetch the data as there is no time to do this process again and again.
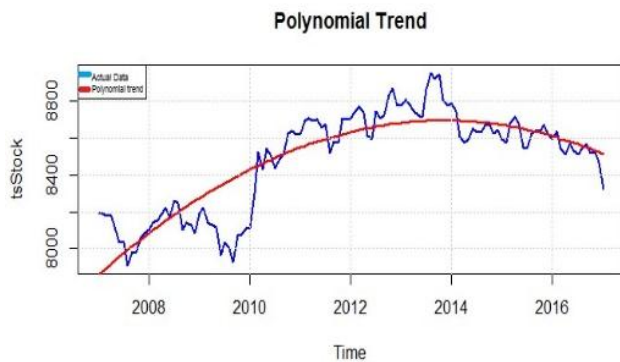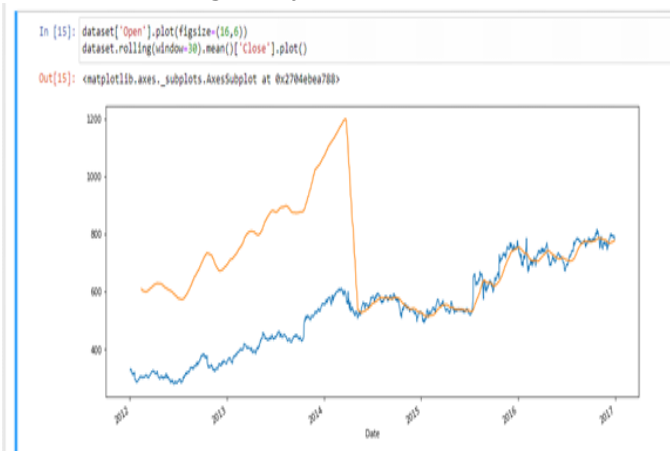
REFERENCES

[1] Definition of "buy low, sell high". Available from: http://www.thefreedictionary.com. [cited 27th july 2017]

[2] Keras. Available: https://keras.io. [cited 27th july 2017]

[3] reLU.Available from: https://github.com/Kulbear/deep-learning-nanofoundation/wiki/ReLU-and-Softmax-Activation-Functions. [cited 27th july 2017]

[4] Adam optimizer. Available from: https://arxiv.org/abs/1412.6980. [cited 27th july 2017]

[5] Yahoo Finance. Available from: http://www.finance.yahoo.com. [cited 27th july 2017] Google Finance. Available from:https://www.google.com/finance. [cited 27th july 2017]

www.geeksforgeeks.com