STOCK PRICE PREDICTION

MSDA3050-01-S21 APPLIED MACHINE LEARNING

SHREYA BANERJEE

PROF. BARNY SANCHEZ

APRIL 25, 2021

ABSTRACT

In stock markets, the aim is to be able to predict future values for a company's financial stocks. In recent times for stock market prediction, technologies such as machine learning algorithms are used which works and relies deeply on past values of stocks market for predictions by performing training and testing on previous stock data. This paper anchors on Long short Term Memory (LSTM) which is an advanced version of recurrent neural network (RNN) in which previous information of the previous state is considered.

PROBLEM STATEMENT

Stockbrokers and investors have been facing extreme volatility for stock investments due to the current situations of the pandemic and observing market performance. Machine Learning models provide a pathway for curious investors like me to study and predict the performances for the invested stocks to avoid loss brackets. In the past, due to unseen circumstances my shares have faced a sudden dip in value per share. I see a clear benefit of stock price prediction as it helps to make profits and understand the market trends better. My machine learning model will be executed to predict the values of the stocks for the next 30 days from the collected stock price values of stipulate time period. The volatility of stocks highly depends on buy orders and the relationship of the sell orders. If number of buy orders increase than sell orders, the price of the stocks increase and vice versa. Alongside these factors physical and behavioral patterns make a major difference making it more complex for the prediction of stock prices for each stock.

Stock market analysis is widely categorized into two:

- Fundamental Analysis- In this analysis, company's profit analysis is calculated through external factors such as the current financial performance by analyzing the quarterly earnings and financial statements/balance sheets.
- Technical Analysis- For technical analysis, prices of past and present for demand and supply patterns are observed. Charts and other statistical analysis sum up for analyzing stocks overtime and to understand its predictions.

SOLUTION

For predicting stock prices, considering the factors affecting the stability of the buy and sell values make it very difficult. But machine learning algorithms are a good approach towards understanding the factorial behavior and predicting values for the intended stocks and shares.

For this project, the data source for the prediction of stock prices is going to be taken from public listed source. Stacked Long Short-Term Memory (Stacked LSTM) is the model which will be used for the prediction analysis. Stock price prediction is used for determining the stock performance for progressing years for the investors to be able to reap the most benefits from the profits earned through the correct investing cycle. Through the process of web scrapping, the data will be transformed into a suitable data frame which will be used for predicting the value for the share.

DATASET CREATION

To harvest my data for "Apple.Inc" stocks , I am using two approaches to explore multiple ways

of collecting and harvesting data. The first approach for harvesting data is using Bloomberg. After logging in Bloomberg, filters were applied before the extraction of raw stock price data was performed. For example, selection of the data range, daily periodic interval, and important parameters such as Open and close price, high, low and volume. The entire filtered data is then processed to remove any unwanted column values and then exported into a csv file. For my second approach, for acquiring stock data is using API key from 'Tiingo' which is a free platform that can be used to collect data using unique API key with a limit of 50 requests per day. Tiingo provides historical end of day prices on equities and exchange-traded fund data it also has the feature of extraction for real time data. The dataset will be created using an imported package 'pandas_datareader' which is a data reader provided through pandas library which will give time series dataframe object with informational columns of the stocks. 'df.get_data_tiingo ('AAPL', api_key=key)' is used to acquire the data which is then converted to csv file using 'df.to_csv' function. To observe and understand the data first I will check the size of the data, column details and plot the data which will be used for the analysis. As LSTM is sensitive to the scale of data, MinMax scaler from sklearn preprocessing library with a feature range between 0 and 1 will be used. Splitting dataset into 70% of the data will be used for training and the other 30% will be used for testing purpose.

MODEL EVALUATION

In the beginning, the model is fitted on a training dataset, which is an example to fit the parameters of the model. Practically the training dataset often consists of an input vector pairs for the values. The training dataset is running with the current model and the target value is compared to result obtained for the input vector for training set. The first step in making good

predictions is training a model. In my model, the data is split into x_train and y_train then further converted to numpy arrays and reshaped it. Before creating the stacked LSTM, the input is reshaped from [samples,timestamps] to [samples, timestamp, features] which is required. Finally, the test dataset is used to provide an unbiased evaluation of a final model fit on the training dataset. In my model, the dataset used is x_test and y_test which is further converted.

VISUALIZATION

The training of the model will be done by the train set, the validation set will tune the model hyperparameters(X_train,Y_train) and the performance of the model for the final prediction values. Before the root mean squared error between actual and predicted values on the validation set will be calculated, transformation to the original form of the split matrix will be done. RMSE performance metrics are calculated:

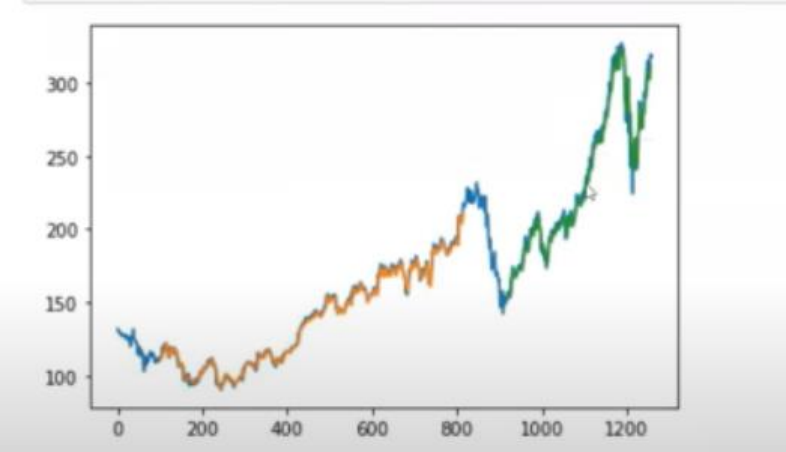
From sklearn.metrics import mean_squared_error

Math.sqrt(mean_squared_error(ytrain,train_predict)

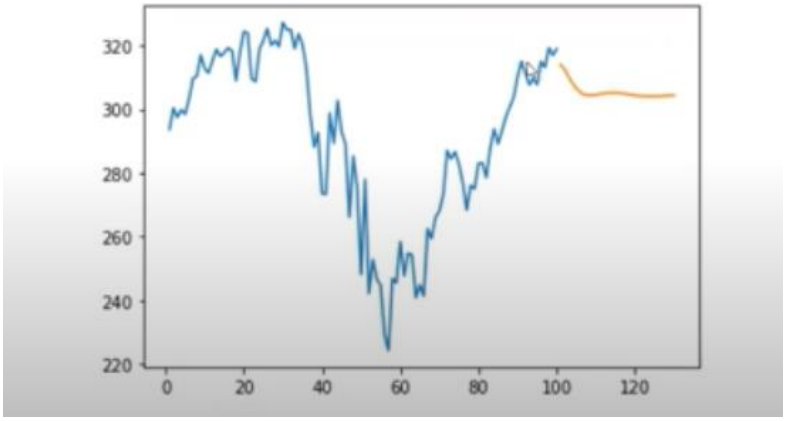
Math.sqrt(mean_squared_error(y_test,test_predict)

The first graph in fig(1) is produced to show the training, validation, and testing for the LSTM prediction

The second graph in fig(2) shows the outcome for the predicted values for the next 30 days



Fig(1)



Fig(2)

PROJECT CHALLENGES

LSTM has the capability to deal with past sequential data problems making it an effective model for predicting forward stock values properly. It is very important that the price of previous day stock values is available.

REFERENCES

Hiransha M.; Gopalakrishnan, E.A. Menon, V.K.;Soman, K.P.(2018). NSE Stock Market Prediction Using Deep-Learning Models. Procedia Comp.Sci 2018, 132, 1351-1453

Sidra Mehtab, Jaydip Sen (2020). Stock Price Prediction Using CNN and LSTM- Based Deep Learning Models, Machine Learning, 239-247, 2014.

T. Gao, Y. Chai and Y. Liu, "Applying long short term memory neural networks for predicting stock closing price," (2017) 8th IEEE International Conference on Software Engineering and Service Science (ICSESS), Beijing, 2017, 575-578,2017.

Mostafa, M.: Forecasting Stock Exchange Movements Using Neural Networks(2010): Empirical Evidence from Kuwait. Expert Systems with Applications, 6302 – 6309, 2010

Sen, J.: Stock Price Prediction Using Machine Learning and Deep Learning Frameworks. (2018) In: Proceedings of the 6th International Conference on Business Analytics and Intelligence, 4933-1932, 2018.