Syeda Zainab Akhtar

(MLND Capstone Project)

May ,2020

Domain Background-

As a student of finance, I've always been intrigued with the world's financial markets. I also wanted to follow up on my MLND project on this topic. I've heard from various stock predictors on the stock exchange that ML forecasts anything for a single company or supply of the day. I have a personal stake in microprocessor corp. which is a microprocessor business and also a GPU supplier to my network. Moreover, owing to the competitive stock sector, the usage of ML is more required to prevent the tragedy of the 2008 market crash.

Problem statement

The strategy deals with the prediction of a specified stock on NASDAQ where then the progress will be identified and subsequent decisions about the near future of the company will be made. This will offer the investor an idea whether to spend or not.

This project is designed to use neural networks to ensure a continuous monitoring of stock ups and downs. The performance is calculated by a long study of stock market.

Even for modern libraries, such as the tensor flow developed by Google, these problems can be easily solved while running on large data sets. I will like to model financial economies specifically, because as a business trader you can equally decide when to purchase stocks and when to sell them in order to earn a profit. That is where the simulation of time series works.

I can use the time-series paradigm known as the Strong Short-Term Memory. LSTM models are strong, particularly for long-term memory retention. To predict the price of my favorite stock.

Data sets and inputs-

I have used data of set NASDAQ from

Kaggle(https://www.kaggle.com/gunhee/amdgoogle/version/1)to this project.The project is able to predict everyday final price of stocks after rigorous training and finally testing.

Data pre- processing----

1.Data discretization
2.Data integration

The data sets used completely continuous because numeric variables that have an infinite number of values between any two values.Also a continuous variable can be numeric or date/time. For example, the length of a part or the date and time a volume or stock point is received.

There are 2336 values provided by kaggle in AMD data set from may 2009 to september
2018. The tables include volumes of stocks each day,opening and closing points per share.

Solution statement-

The main approach will be to use LSTMs because LSTMs are very successful in forecasting sequence problems because they are capable of processing past information. It is important in our situation, because the previous price of the stock is necessary for the estimation of the future demand.
The major output will be in Jupyter notebook (.ipynb format)and will be supported using various python modules.

Libraries used for LSTM from keras

```
yrecurrent_activation='hard_sigmoid', use_bias=True,
kernel_initializer='glorot_uniform', recurrent_initializer='orthogonal',
bias_initializer='zeros', unit_forget_bias=True, kernel_regularizer=None,
recurrent_regularizer=None, bias_regularizer=None,
activity_regularizer=None, kernel_constraint=None,
recurrent_constraint=None, bias_constraint=None, dropout=0.0,
recurrent_dropout=0.0, implementation=1, return_sequences=False,
return_state=False, go_backwards=False, stateful=False, unroll=False))
```

Libraries like pandas ,matplotlib, numpy also used.
The overall estimate is for the study of the current and expected asset values.

Personal results I'd love to get the forecast sharing point of the day later utilizing the impact of the training data sets. It's also going to be good to see the market growth of the sector.

Benchmark models-

The model will mainly be evaluated using continuous data in such a way that it is checked for final estimation using linear regression. Enhancement and performance will be analyzed using recurrent neural network models. The LSTM has back-spreading and resource control features that enable it handle large data sets. The neural net definition is the biggest feature for me where the norm should have been re-evaluated again to make final prediction.

Metrics used-

For analyzing model various metrics used are-

1. RMSE(root mean squared evaluation)-It is very good model for large continuous data .Unlike the absolute mean error it severely punishes large Error.

It is:

$$((Y_{pred}-Y_{obs})^2/n))^{1/2}=\text{Error}$$

Y_{pred}-linear regression output

Y_{obs}-Neural network

output n - total input

d/d(E)=gradient descent

Project Design-

1. Data sets-

a.Data set of AMD from kaggle.

b .use pandas to import data set. eg-

d=pd.read_csv('Filename)

c . Analyzing data sets from top and tail.

d . split dat in 3:7of testing: train

2. Develop primary and secondary models-

LSTM(primary)and linear regression secondary model.Curves will be plotted to increase the visualization using graphs.

3. Libraries used -Keras,numpy,pandas ,matplotlib,scikit learn etc.

References-

1. https://www.kaggle.com/gunhee/amdgoogle/version/1
2. https://ieeexplore.ieee.org/abstract/document/23959/metrics#metrics
3. https://www.kaggle.com/johanvandenheuvel/lstm-model-of-stockdata/comments
4. https://www.quantcon.com/

5. https://www.kdnuggets.com/2018/11/keras-long-shor t-term-memory-lstm-model-predict-stock-prices.html

6. https://www.datacamp.com/community/tutorials/lstm-python-stock-market