

# **Analytics GC (2022)**

## **Submitted by**

Samarth Pusalkar - 19D180029

Rachakonda Sravan Kumar- 180040076

Alakh Agrawal - 200040018



**Indian Institute of Technology, Bombay**

*12th February 2022*

# Research

Our approach to the problem statement was to first understand the data and look at the commonalities between Open, Close, High and Low.

After normalizing the data using these four features we then went on to try out multiple models. We tried out Neural Network Sequential Modelling and LSTM.

**Sequential models** are the machine learning models that input or output sequences of data. Sequential data can include text streams, audio clips, video clips, time-series data and etc.

It suggests a systematic, sequential approach to the problem.

**Long short-term memory(LSTM)** is an artificial recurrent neural network architecture used in the field of deep learning. Unlike standard feedforward neural networks, LSTM has feedback connections. It can process not only single data points but also entire sequences of data.

Long Short-Term Memory (LSTM) networks are a type of recurrent neural network capable of learning order dependence in sequence prediction problems.

We later tried **Logistic Regression** from sklearn and found high accuracy. Logistic regression is a process of modeling the probability of a discrete outcome given an input variable. Logistic regression is a useful analysis method for classification problems, where you are trying to determine if a new sample fits best into a category.

Scikit-learn has a highly optimized version of logistic regression implementation, which supports multiclass classification task.

So for normalizing the data we took the open, high, low and close and created four values from it which were-

```
x1=(close-open)/(high-low)*10  
x2=close/open*10  
x3=high/close*10  
x4=open/low*10
```

Then we created a simple sequential model and fitted the data after deciding which model fits best.

The model is supposed to get filtered input but for the unlabeled parts given we made prediction for all of them as we didn't know how to give y predictions only for specific input and not confuse you with which 'y' is where.

That's it!

# Conclusion

After looking at various models and playing around with them we found the Deep Learning based Sequential Neural Network model accuracy to be approximately 53.

Looking at the input features and output class, we concluded that  $y=1$  refers to the scenario when the price of the stock increases the next day whereas,  $y=0$  means the price will drop or remain same.