

# **Data Science**

## **(Project Report)**

TY Computer Engineering 2021-22 Div.:

1 Batch: T5

### **Group Members:**

1. Isha Rathi (111903037)
2. Mihir Malani (111903046)
3. Nikita Tipule(111903051)

### **Proposed Topic: Stock Market Prediction**

#### **Introduction**

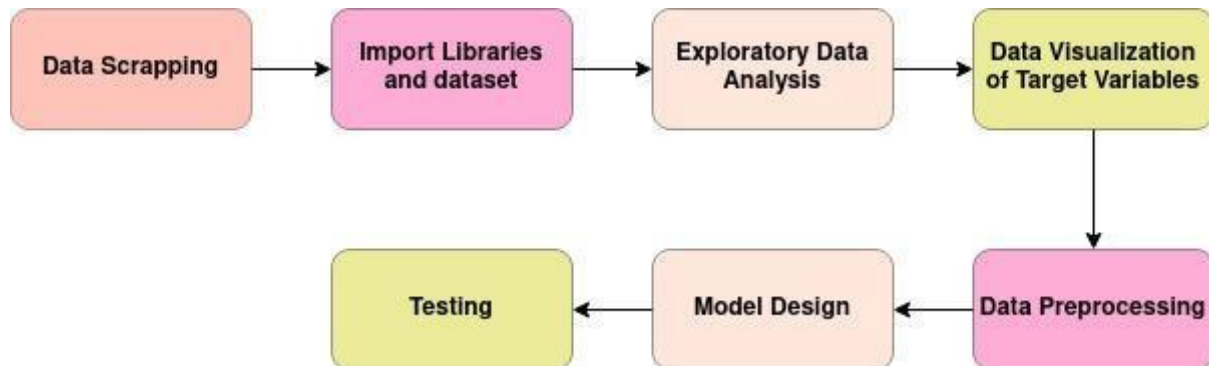
The stock market broadly refers to the collection of exchanges and other venues where the buying, selling, and issuance of shares of publicly held companies take place.

In simple words, a share indicates a unit of ownership of the particular company. If you are a shareholder of a company, it implies that you as an investor, hold a percentage of ownership of the issuing company. As a shareholder you stand to benefit in the event of the company's profits, and also bear the disadvantages of the company's losses.

**The successful prediction of a stock's future price could yield significant profit.**

## **Project Pipeline/implementation**

The various steps involved in this project are -



### **Data Scrapping:**

- Data scraping, also known as web scraping, is the process of importing information from a website into a spreadsheet or local file saved on your computer.
- There are various techniques to scrap data from web. In our project, we had scrapped data using beautifulsoup from Wikipedia website.

### **Import Necessary Dependencies Dataset:**

- Various Machine Learning Libraries like Quandl, Pickle, pandas, Keras, seaborn , Requests and sklearn should be imported there to carry out the implementation process.
- We can import the Scrapped Dataset by reading and loading it from the CSV file.

### **Exploratory Data Analysis:**

- Exploratory Data Analysis involves exporting the five top records of data & columns/features in data.
- Also, it involves finding the length & shape of the dataset, gaining data information like Datatypes, checking for null, unique and number of target values.

## **Data Visualization:**

- Data visualization is the graphical representation of information and data. It is a particularly efficient way of communicating when the data is numerous as of our dataset.
- For our project we used the following visualization techniques to get information about dataset
  - Pair Plot
  - Scatter Plot

## **Data Preprocessing:**

- Data cleaning can be applied to filling in missing values, remove noise, resolving inconsistencies, identifying and removing outliers in the data.
- Data preprocessing is a predominant step in machine learning to yield highly accurate and insightful results. Greater the quality of data, greater is the reliance on the produced results. Incomplete, noisy, and inconsistent data are the properties of large real-world datasets. Data preprocessing helps in increasing the quality of data by filling in missing incomplete data, smoothing noise and resolving inconsistencies.
- We have perform pre-processing steps on dataset like checking for null values, dropping the columns that were not required for the prediction from csv file and standardizing the values to appropriate scales.

## **Model Building:**

In the problem statement we have used LSTM ( Long Short-Term Memory ) Network model which is special kind of RNN ( Recurrent Neural Network) model. It is capable of learning order dependance in sequence prediction problems.

## **Conclusion**

The actual and Predicted values were close enough thus signifying the accuracy of the system.

## **References**

- [https://en.wikipedia.org/wiki/Stock\\_market\\_prediction](https://en.wikipedia.org/wiki/Stock_market_prediction)
- [https://en.wikipedia.org/wiki/NIFTY\\_50](https://en.wikipedia.org/wiki/NIFTY_50)