**Springboard Data Science Career Track 2020 Milestone Report 1**

# **Stocks Sentiment Analysis Using AI**

**Mukkul Kurwa**

**December 2020**

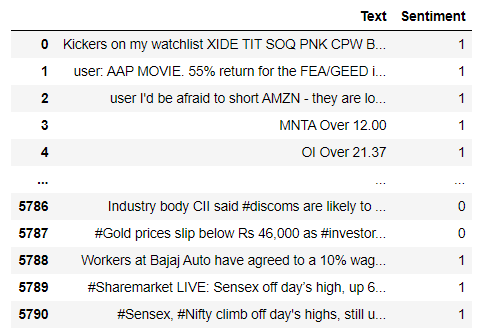
**Project Overview:** We live in a world where we are constantly bombarded with social media feeds, tweets, and news articles. This huge data could be leveraged to predict people sentiment towards a particular company or stock. Natural language processing (NLP) works by converting words (text) into numbers. These numbers are then used to train an Al/ML model to make predictions. Al/ML-based sentiment analysis models, can be used to understand the sentiment from public tweets, which could be used as a factor while making a buy/sell decision of securities.



**Problem Statement:** Al/ML-based sentiment analysis models, can be used to understand the sentiment from public tweets, which could be used as a factor while making a buy/sell decision of securities.

**Metrics:** Once the model is trained, I need to test its performance on the testing data-set. The model has never seen this information before; as a result, the testing data-set allows me to determine whether or not the model will be able to generalize to information that wasn't used during its training phase. I have used some of the metrics provided by Scikit-learn for this purpose such as classification reports and accuracy score.

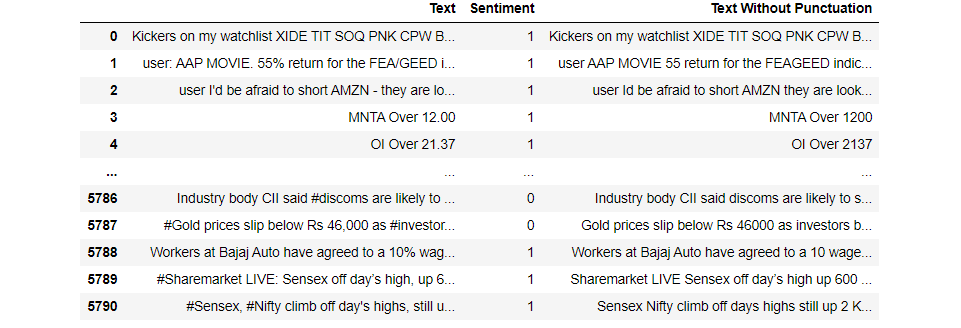
# Data Exploration: Our data-set has 1 feature known as Text upon which we will be training our model which in turn predict the sentiment.



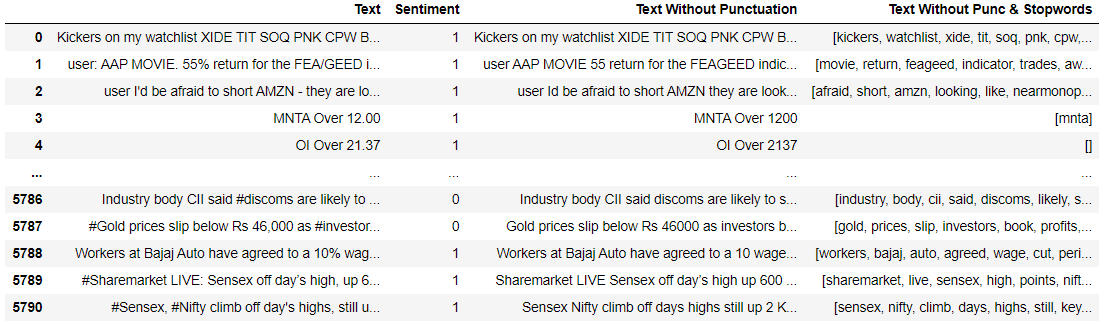
**Data Cleaning:**

1. **What kind of cleaning steps did you perform?**

* Removed punctuation from the text.

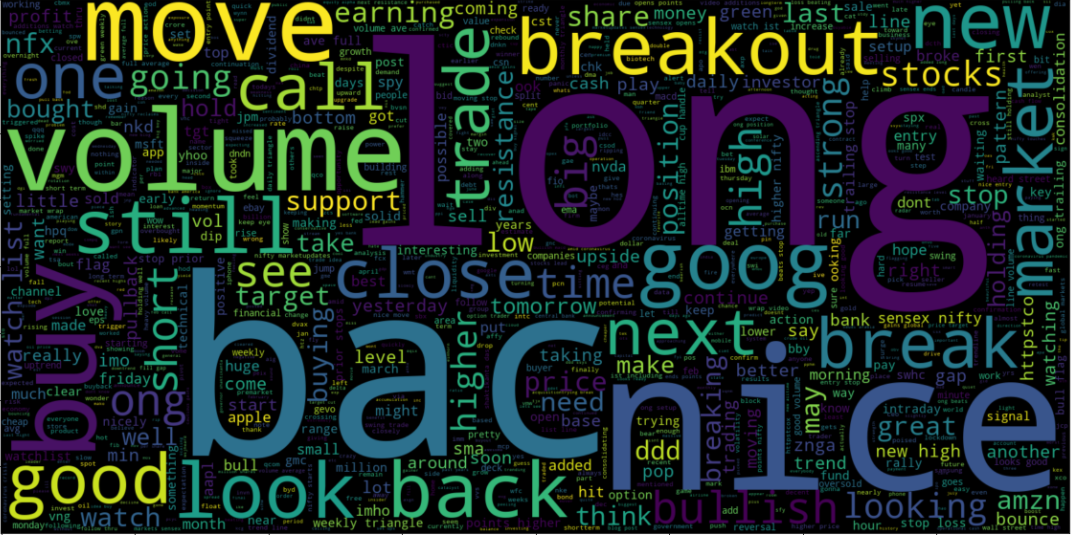


* Removed the stop-words.

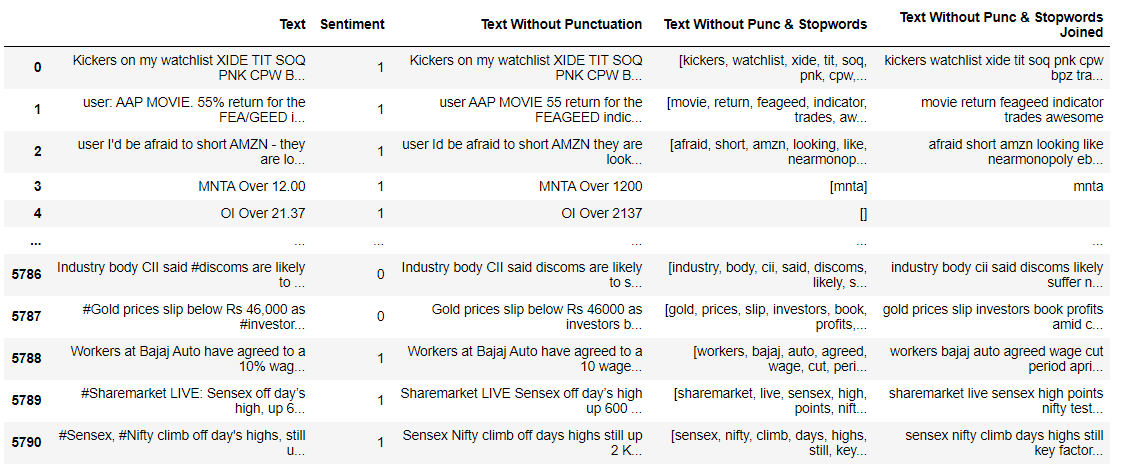


**Data Visualization:**

* Plotted the Word-cloud as follows:



* The cleaned data after visualization:



* The distribution for the number of words in the text:

