**DATA 557 A Project: Milestone 2 Proposal**

**Project name:** Indian stock markets data: Analysis of indicators and market sentiments

**Students Involved:**

1. Sagnik Ghosal
2. Arjun Sharma
3. Raman S V
4. Adithyaa V

**Project concept and goals:**

Stock markets are a rich source of properly constructed data. We have access to at least 10+ years of daily data for various stocks and indices. The data here includes details around stock prices such as opening prices, closing prices and daily highs and lows. Further, we also have access to volumes of stocks traded on a daily basis. For decades now, scientists and laymen alike have been using statistical methods to analyze and even predict stock market performance. While the adage of “past performance is not indicative of future results” is rather famous, we want to try and analyze if there are indeed any particular set of methods that can predict stock market performance at a high level of confidence. For instance, if we identified a method based on movements of RSI (relative strength indicator) that accurately predicted a buy or sell position in a stock and if the method here returns overall net positive trade value if bought and sold based on it, we would consider this a successful method.

In terms of goals, we want to be to able to identify at least 2 methods that have had excellent success on positive trades over the past 10 years of data. Further, we would also want to be able to look at major movements for big companies or indices and identify the sentiments surrounding these movements from both the analysts and retailers point of view. For instance, if the Tesla stock fell 20% in a day, we want to analyze what the mood of the market was around this stock both a day before the crash and a day after. We want to illustrate how big stock brokerages give out calls on a stock and what their reactions are when said calls go awry.

**Datasets:**

We have identified the below dataset from the Indian stock exchange – National Stock Exchange. They provide us with a daily dump of the index’s constituents price movements. The data is available from the site – <https://www1.nseindia.com/products/content/equities/equities/archieve_eq.htm>.

We will attempt to automate the pull of this data all the way till a point in 2000s. Most of the files have the necessary columns for analysis – stock name, high price, low price, open price, closing price and previous closing price. Using this, we will be able to create a repository of around 15 years of data for all the major stocks/companies that we will shortlist.

We would not need a lot of initial clean up as long as we identify the dates the columns change over, if at all there was a change in columns.

Further, we have only 3 datatypes per file for the relevant fields – Text (stock name), decimal (stock prices and volumes), date (timestamp for these values).

In terms of sentiment analysis, we will be looking at twitter feeds for the stocks/indices and try to gauge interest by measuring tweet activity.

We are trying to locate a website/source that would give us access to the historical stock recommendations by major brokerages. If we do not get this data, we would not continue with the sentiment analysis and would instead focus on identifying newer indicators and their correlation with stock prices at a weekly, monthly and yearly aggregations of stock price.