The answer to “What is the likely outcome of a given scenario?” depends on an analysis of multiple training scenarios. It involves the selection of appropriate predictors and creating a model to provide an estimate of the probable outcome. With well-defined and structured variables, this task is fairly existent in practice. But what if the scenarios we have are not in any structure but rather in the form of text data? The first step here – interpret the text! With the industry dealing with voluminous amounts of text data in the form of reports, news, forms, views, it is essential to have a methodology to interpret the data as much as possible avoiding any loss of information. At the same time it is important that it is done by means of an automated engine so that the process is intuitive, fast, consistent, and is a scientific and systematic way rather than human manual judgement. This ensures a higher quality of interpretation at a lower cost with reduced time. This brings the need for text and sentiment analytics.

### METHODOLOGY

- **Aggregation of a words score:** The score for each word is aggregated across all text data to come up with one from which the keywords are identified.

- **Breaking text into words:** The text data is cleaned for punctuations, prepositions, and stopwords. Then stemmed to get only the root of the words.

- **Known final outcomes:** Known final outcomes are monitored and their stocks very quickly. But to invest more, they take time to analyse and hence the movements are not immediately reflected.

- **Financial markets are primarily driven by investor sentiments and how the investors perceive a news item to be positive or to be a negative one.**

- **A database of historical news items is created by scraping web pages.** At the same a database of half hourly stock price data is created. Each news item is mapped to the next 10 half hourly returns that would have been witnessed after the announcement of the news.

- **A score is assigned to each item by the formula:**
  
  \[ \text{Maximum Return} \times \text{Maximum Positive streak} + \text{Minimum Return} \times \text{Maximum Negative streak} \]

- **A score is assigned to each item by the formula:**
  
  \[ \text{Maximum Return} \times \text{Maximum Positive streak} + \text{Minimum Return} \times \text{Maximum Negative streak} \]

- **A score is assigned to each item by the formula:**
  
  \[ \text{Maximum Return} \times \text{Maximum Positive streak} + \text{Minimum Return} \times \text{Maximum Negative streak} \]

- **A score is assigned to each item by the formula:**
  
  \[ \text{Maximum Return} \times \text{Maximum Positive streak} + \text{Minimum Return} \times \text{Maximum Negative streak} \]

- **A score is assigned to each item by the formula:**
  
  \[ \text{Maximum Return} \times \text{Maximum Positive streak} + \text{Minimum Return} \times \text{Maximum Negative streak} \]

### OBSERVATIONS

- **Negative news is quicker to show its effects on the stock prices than a positive news** – a possible reason for this is that investors get worried about possible losses and sell off their stocks very quickly. But to invest more, they take time to analyse and hence the movements are not immediately reflected.

- **In case of document inference, modelers never accept a premise. They just “do not have sufficient evidence to reject it”**. As a result, they avoid the presence of positive words. Hence just like stock price movement, the accurate detection rate of negative words is higher than that of positive words.