**Dow Jones Index Prediction**

**Context:**

In this use case, you are tasked to predict if the Dow Jones Industrial Average index will close higher or lower based on news.

**Data:**

The dataset provided is comprised of the following:

1. **Training\_Combined\_News\_DJIA.csv**: This is the training data set and it contains a table that combines both stock and news data and have the following columns:

* Date : The date where the news have been taken
* Label: “1” if the DJIA closed higher or same. “0” if the DJIA closed lower. This the variable to predict
* Top1 to Top25: Top 25 historical news headlines news from Reddit WorlNews based on reddit users’ votes for each date.

The training data is from 2008 to 2014

1. **Test\_Combined\_News\_DJIA.csv**: This is your test data set and it contains the same columns as the training file.

The test data is from 2015 to 2016

**Guidance:**

* What type of information is relevant that might affect the stock market?

### Can you summarize through a visualization the type of information that is relevant to the stock market (e.g. key words with high impact, key words with low impact)?

* What features can you create from the news to build a predictive model?
* Can you think about creating more independent variables (feature engineering)?
* Is this a classification or regression task? What models/algorithms will be suitable for such Task?
* How can you use embedding for this use case? Can you implement your idea?

### Scoring:

### By 2:30pm (on the day of the Hackathon), a new file “Test\_Combined\_News\_DJIA \_with\_score” will be released and it will have the stock movement for the testing period.

### You will need to use this to calculate the Accuracy and provide it to the judges.