Kaggle

My kaggle account name is bhargavvader (https://www.kaggle.com/bhargavvader).
The challenge we will be looking at is the challenge titled: “Two Sigma: Using News to Predict Stock Movements”. The sponsor of the challenge is Two Sigma, a New York city based hedge fund. A hedge-fund primarily focusing on technological methods for its trading strategies, their website describes them as using cutting edge machine learning and artificial intelligence methods.

The project involves making predictions of stock market returns based on newspaper data. The idea is to use indicators in the news to make more accurate predictions on what stock market returns would be. The market data is provided by Intrino and the news data is provided by Thomson Reuters. The submissions will be evaluated based on how well the market return value is predicted; this is done by first asking the user to calculate a confidence value between -1 and 1 based on how well the stock is predicted to perform, after multiplying it with a ‘*assetCode’* value given to us. The final submission score is the mean of this value divided by the standard deviation of the predicted values. Submissions are required to have the following variables: time, assetCode, confidenceValue.

The honour code mentions that no outside data is to be used to create the model, and that one cannot abuse the competition infrastructure to gain an unfair advantage.

The competition is a two stage competition, where during the first stage participants will build models and the leaderboard only reflects scores over a historical period. After this stage finished, the code is frozen and the evaluations occurs based on future data. The start date is 9/25/2018, and there is a team merger deadline on 1/2/2019. After 1/8/2019 the code will be frozen and the final results are announced on 7/15/2019.

It seems quite clear how this competition will be very beneficial for Two Sigma: they are a hedge fund company actively using a wide variety of data sources to make investments which in turn bring in their profits. Text analysis is a field of data science which is difficult to make very accurate predictions with given the rather subjective nature of the data, but by making this an open problem it is possible to receive a wide variety of submissions using this textual data in a new way potentially leading upto very interesting results, which will in turn boost the profits and performance of the company.