

Second Capstone Proposal

Problem Statement: The problem trying to be solved in this project is to predict the stock market using news. News can influence people to make a decision to sell or buy a stock. Some news are more important than others, for example, earnings reports or bankruptcy filings. Such news can not only influence someone to buy or sell a stock but it can also skew the price of a stock indefinitely! The goal behind this project is to find a correlation between the news, sentiment, and the price a stock trades at.

Dataset: There are two datasets used in this project. The first has information on the company name, the open price, the closed price, the volume traded and various other past price information. The second dataset has information on news' headlines, news' sentiment, words counts novelty counts and various other counts information. This data, coming from kaggle, is very clean and most likely won't need any wrangling of data. There will still be a check for nulls or empty rows of data as well as a distinct count analysis over the entire dataset.

<https://www.kaggle.com/c/two-sigma-financial-news/data>