

Project Fletcher Proposal

Topic: Clustering Companies Based On Their Media

Alexy Abramson

I) Abstract

The subject that I will address with my project is to analyze the relationship between the news stories written about companies that sell goods to the public and the change in revenue for these companies (which is a proxy for consumer engagement with a company). The project will seek to analyze how companies can be categorized based on the coverage that they receive in the media, the change in their revenues, company-specific factors such as industry, and economic factors.

II) Methodology

The text data obtained from the project will be analyzed using Natural Language Processing, and in particular, a sentiment analysis will be conducted on the news stories written about the particular organizations. The results from the NLP data processing will then be analyzed using unsupervised machine learning techniques to flesh out any trends that may be present in the data, and subsequent conclusions will be made based off of the results of that data analysis.

III) Data

The project will use three types of data in order to conduct the analysis for the project. The first type of data is news stories about the studied companies which will come from news feed APIs. The second type of data is the data on revenue of the studied companies, which will come from governmental filings about the companies revenues filed with the SEC, including information about their revenues, other financial factors, and information about the industry that the company was in. The third type of data will be general economic data, such as price changes in the economy and in general, to isolate the impact of those changes on a company's revenues from the aforementioned factors.