

Project Summary

I) Design

The project sought to take publicly-traded companies on the Standard and Poors 500 Index, the media coverage that those companies received in a fiscal quarter, and the change in consumer engagement during that quarter, and then group the companies into different categories based on those headlines and change in sales.

In order to analyze the press coverage, the top headlines from the fiscal quarter were analyzed as this is where the majority of the information that consumers would see about a company would be. In order to measure the projected impact of a headline, each headline was analyzed using sentiment analysis.

In order to measure consumer engagement with a company, the change in sales in that fiscal quarter as compared to the previous fiscal quarter was used since this is the most direct way that consumers engage with a company.

The relevant data was extracted from two main sources, one for the headlines and another for the sales data. The news headlines were compiled from Google News, and the top headlines were selected since those were the headlines that a consumer would be most like to see, if they even saw all of those headlines. The sales data was compiled from the quarterly or annual reports that companies are required to submit to the Security Exchange Commission (SEC) every three months or every year respectively. The time period used for the analysis was the second quarter of 2017, that is between April and June of 2017.

II) Tools

A) Python - Programming Language

- 1) Pandas (Data Analysis)
- 2) Numpy (Feature Transformation)
- 3) NLTK (Natural Language Processing)
- 4) Scikit-learn (Modeling)
 - a) Count Vectorizer
 - b) Standard Scaler
 - c) Principal Component Analysis
 - d) K-Means Clustering
- 5) Textblob
- 6) Matplotlib (Graphic Design)

III) Data

FEATURE	CONTEXT	SOURCE
Company Headlines	The headlines from news articles written about each company.	Myriad news sources via Google News
Sales Data Q1 2017	The amount of sales made in the first quarter of 2017.	SEC
Sales Data Q2 2017	The amount of sales made in the second quarter of 2017.	SEC

IV) Results

The results of the analysis showed that the companies, when analyzed based on sales data sorted themselves into six main groups. One was a group that had a neutral headline sentiment and little change in sales. A second group had a positive headline sentiment and a slight decrease in sales. A third group had a negative sentiment and a decrease in sales. A fourth group had negative headline sentiment and an increase in sales. A fifth group had a positive headline sentiment and an increase in sales. And a sixth and final group, which consisted of outliers, had positive headline sentiment but a short decline in sales.

For the model used, it was determined that the most important factors that were used to group the companies into these groups were the overall sentiment of headlines about a company and also the frequency of headlines that had a positive sentiment.

V) Next Steps

There are several steps that will be taken in the next phase of the analysis. The first will be to incorporate companies from additional countries to see if there is a global trend or if different countries have different relationships between headline sentiment and change in sales. The second will be to add additional domestic companies to see if there is a difference between larger and smaller companies in the United States. Furthermore, with additional companies, it will also allow for a greater degree of topic modeling which can add another set of features. In addition, more fiscal quarters will be analyzed to account for both different economic conditions and different seasons to see if those factors have any impact.