

The background of the image is a dark blue, high-tech digital interface. It features a prominent candlestick chart with orange and white bars, overlaid with a white line graph that trends upwards. A person's hand, wearing a white shirt cuff and a dark suit sleeve, is visible on the right side, pointing towards the chart. The overall aesthetic is professional and data-driven.

Improving Market Sector Sentiment Analysis through Subreddit Classification

By Jesse Villines

The background of the slide is a complex, blue-toned collage of financial data. It features multiple overlapping line charts with various colored lines (white, yellow, blue) representing different market trends. There are also bar charts and numerous numerical data points scattered across the image, some in white and some in yellow. The overall aesthetic is that of a high-tech financial dashboard or a data visualization from a trading platform. The text is overlaid on a semi-transparent white rectangular area in the center.

Problem Statement:

**Can sentiment analysis for market information
be optimized by observing key features in
news?**

r/StockMarket - Reddit's front page of the stock market, financial news.

About Community

Stock market news, Trading, investing, long term, short term traders, daytrading, technical analysis, fundamental analysis and more. We cover it all at r/stockmarket.

618k

Members

1.3k

Online



Created Jul 9, 2008

r/Stocks - Investing and trading for all

About Community

Almost any post related to stocks is welcome on /r/stocks. Don't hesitate to tell us about a ticker we should know about, but read the sidebar rules before you post. Check out our wiki and Discord!

1.0m

Investors/Traders

8.5k

online, analyzing figures



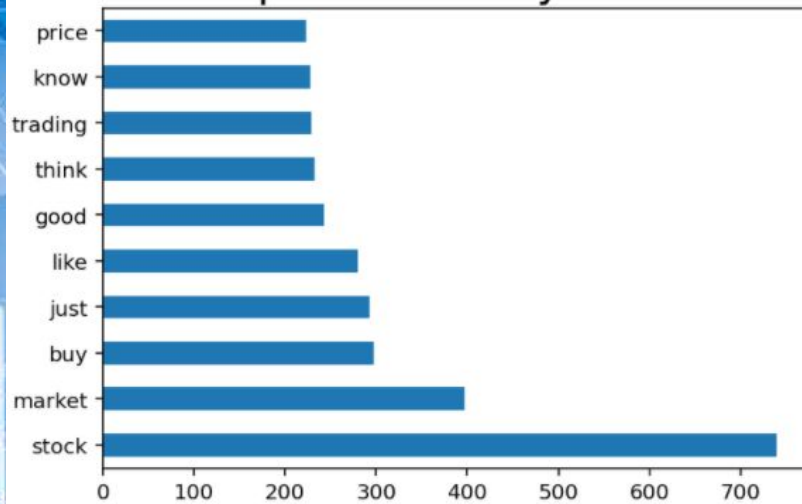
Created Jun 27, 2008

The Dataset:

20,036 posts, evenly distributed between r/Stocks and r/StockMarket.

Making a Baseline
Accuracy of 50%

Top 10 Words by Count



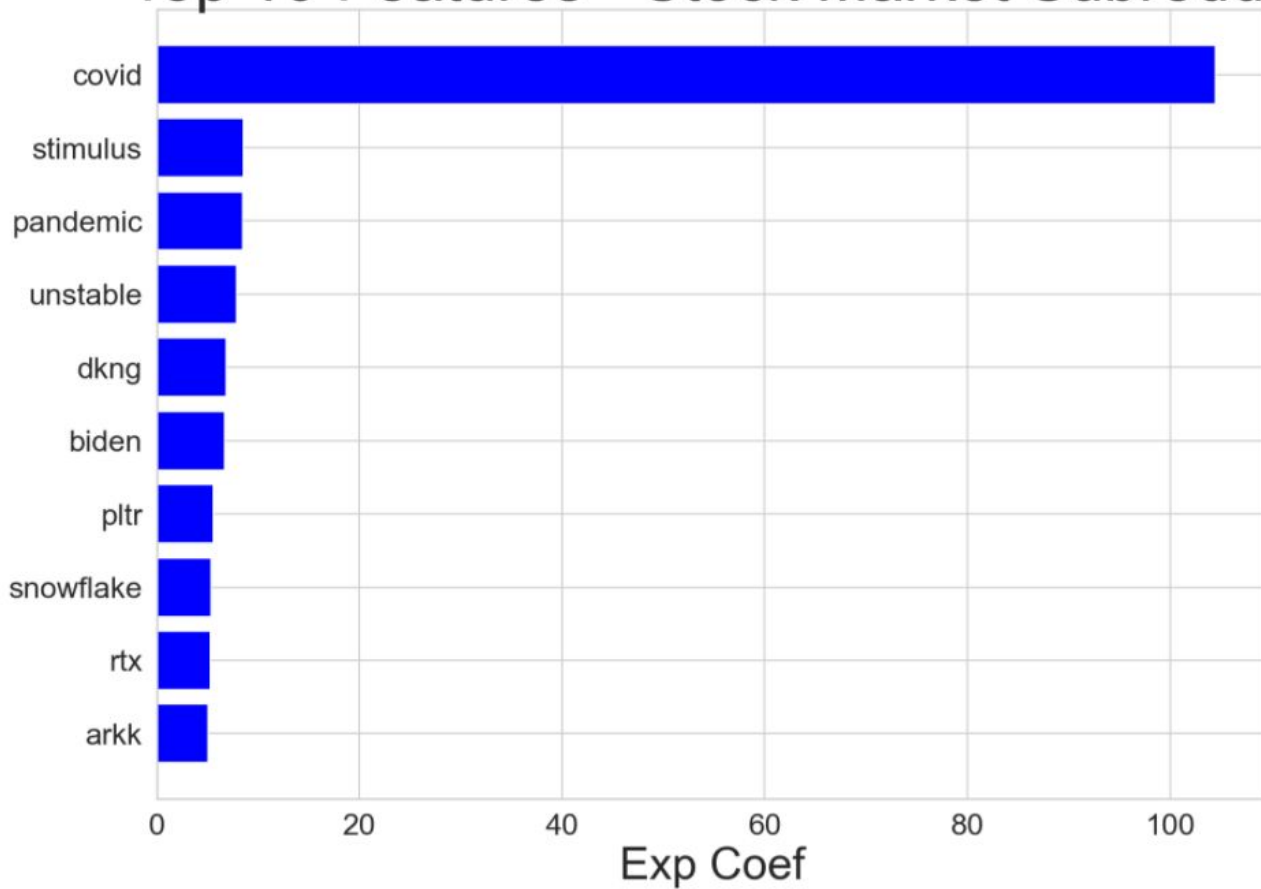
Model Features:

- Title + Post Body
- Sentiment Score
 - A score from -1 to 1 based on how negative or positive the text was.

Model Evaluation

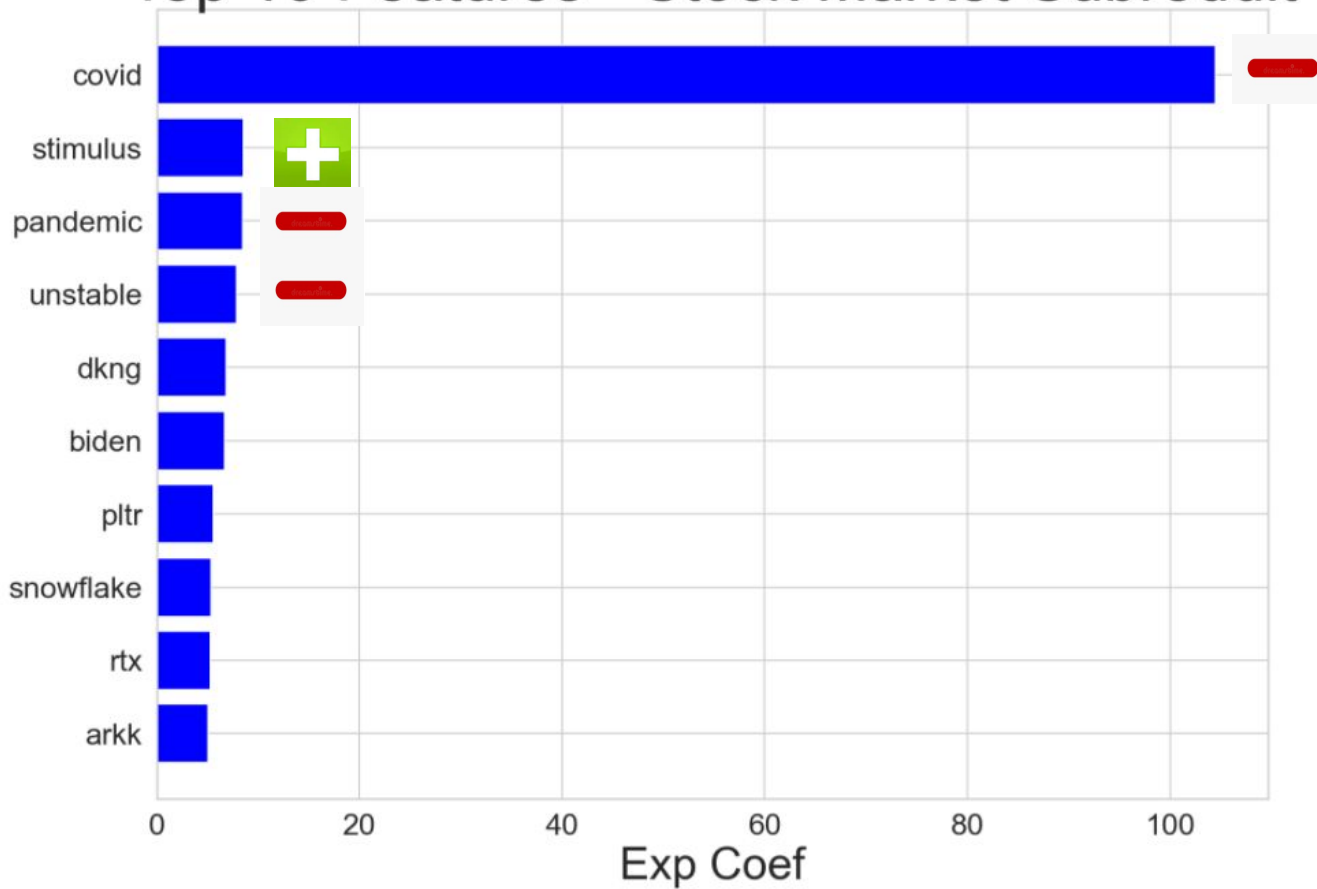
Model	Accuracy Train	Accuracy Test	Precision	Recall	F1-Score	Specificity
Logistic Regression	0.91	0.70	0.72	0.65	0.69	0.74
Random Forest	0.95	0.70	0.72	0.66	0.69	0.75
Ada Boost	0.72	0.70	0.76	0.59	0.66	0.82
SVC	0.85	0.71	0.76	0.61	0.67	0.80

Top 10 Features - Stock Market Subreddit



**Top
feature:
covid**

Top 10 Features - Stock Market Subreddit



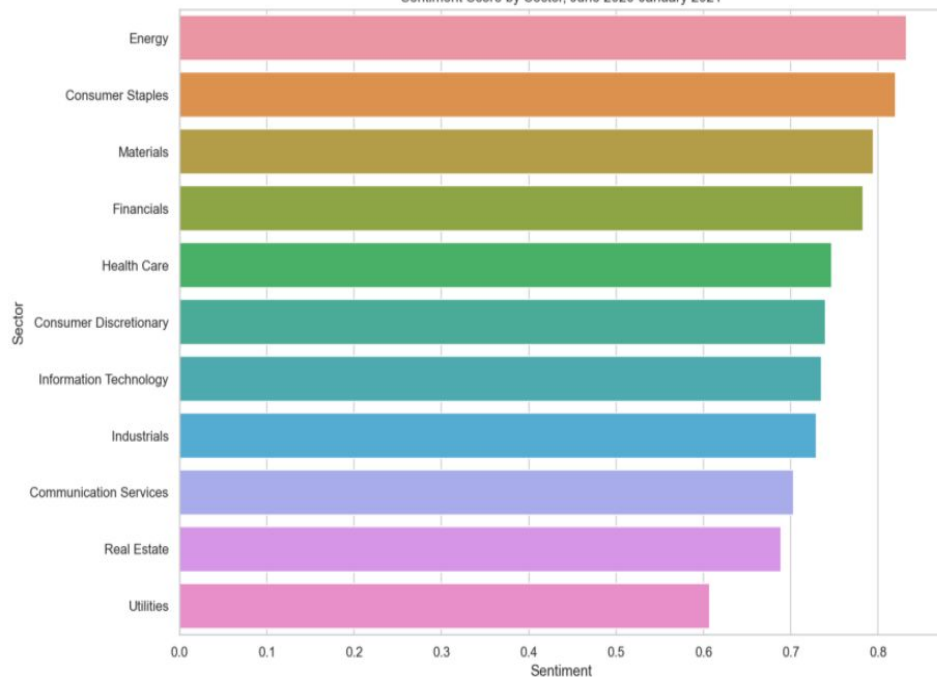
**Top
feature:
covid**

Market Sentiment by Sector

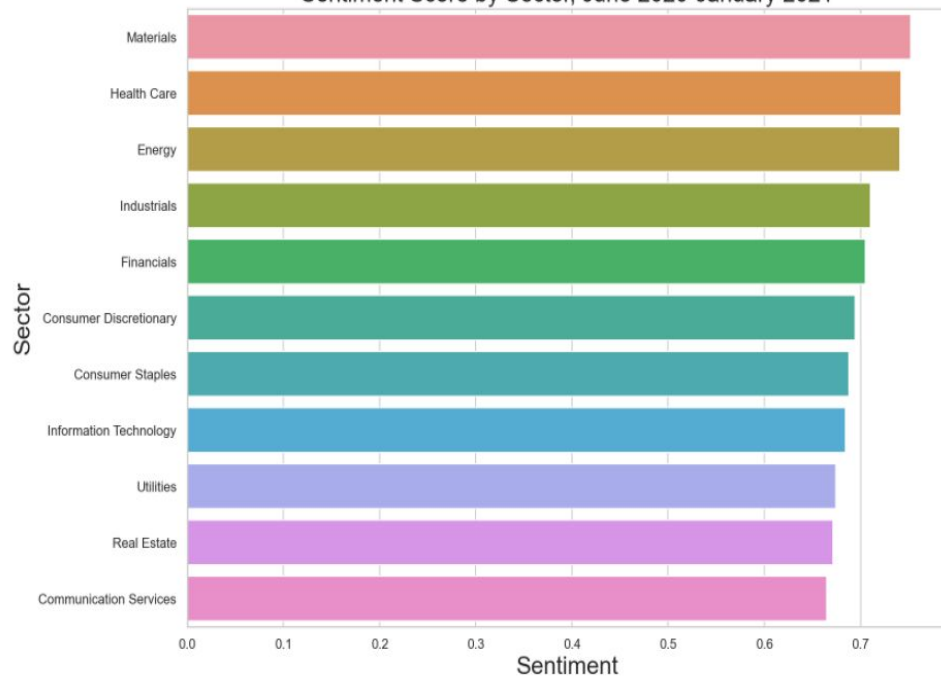
Before

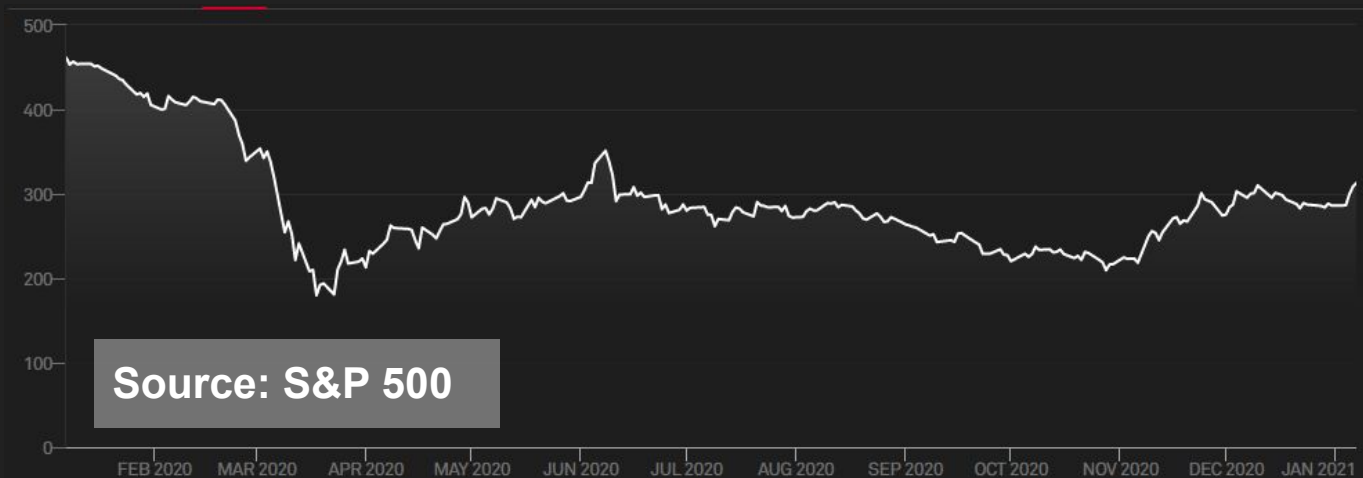
After

Sentiment Score by Sector, June 2020-January 2021



Sentiment Score by Sector, June 2020-January 2021

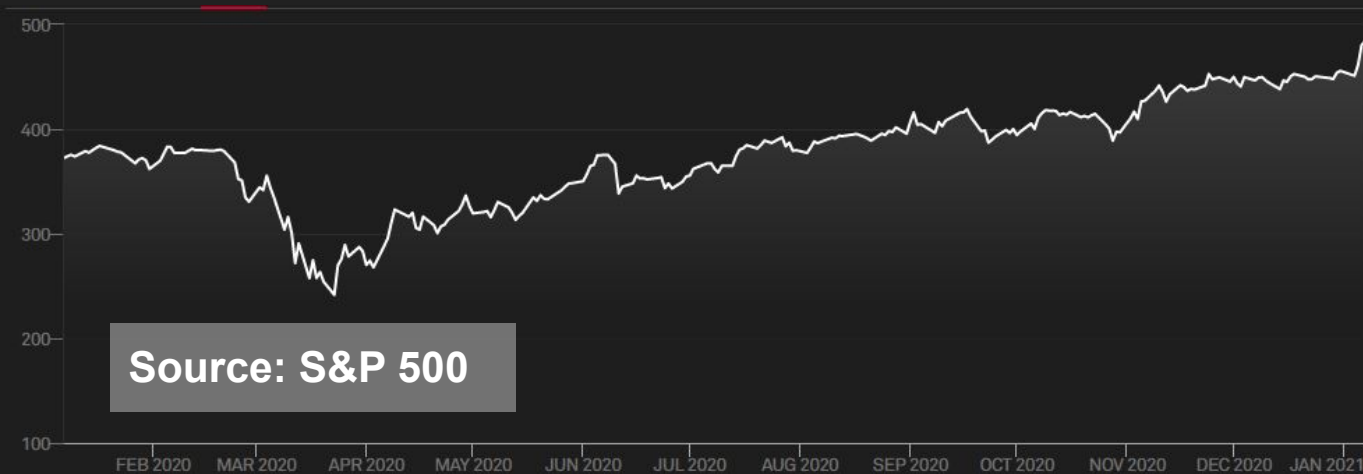




Energy

-32.15%

1 YR RETURNS



Materials

29.98%

1 YR RETURNS

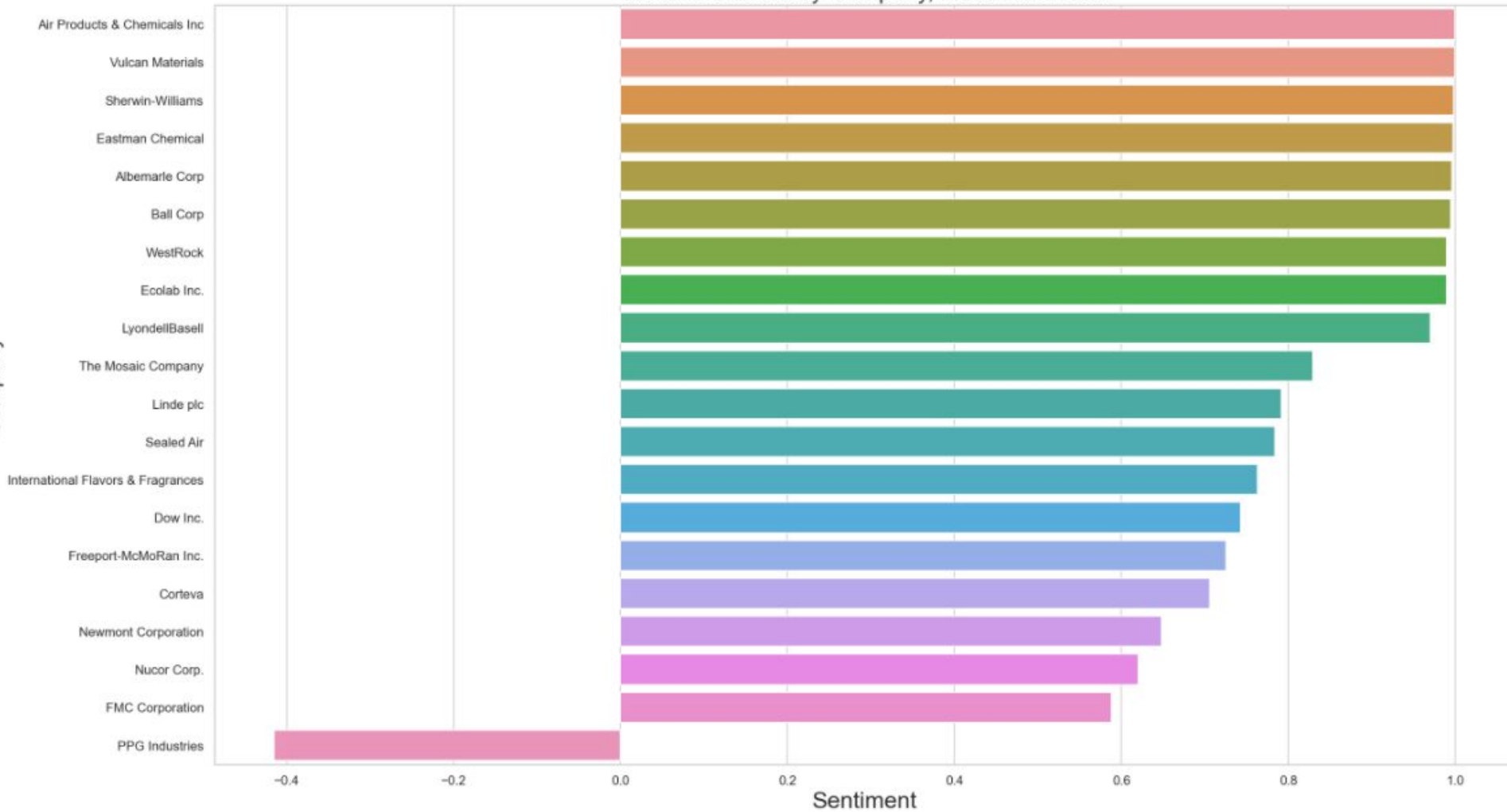
Recommendations and Next Steps:

- Compare normal news to market news to pick more market specific words to assign a sentiment.
- More sources official (Bloomberg, WSJ, etc.) and unofficial (twitter, facebook, blogs) to capture better overall market sentiment.
- Utilize market cap weighted groupings
- Classify into leaders and laggards. Which reports are predictive and which are descriptive.

Additional Slides

Sentiment Score by Company, Materials Sector

Company



Sentiment Score by Company, Energy Sector

