Marc Etter, Yichi Yang, David Chang, Olivia Wang
Adam Meyers
Natural Language Processing
8 May 2023

Using Economic News Headlines SA to Predict the Economic Situation

Abstract:

Sentiment analysis for news articles is a growing field that has the potential to provide valuable insights into public opinion and market trends. In this paper, we propose a novel approach to sentiment analysis that draws on resources from various news websites to generate a comprehensive and diverse dataset. Our method uses natural language processing techniques to extract sentiment from news articles, and then compares these results against economic data such as GDP, S&P 500, and unemployment rate.

The goal of this research is to explore the relationship between news sentiment and economic indicators, and to identify any correlations or patterns that may exist. By analyzing sentiment across multiple news sources and comparing it to economic data, we can gain a more nuanced understanding of how public sentiment affects economic trends.

The crux of this paper is an experiment that filters a general news dataset by relevance to the economy and performs sentiment analysis on the filtered article headlines in order to predict economic indicators over monthly time intervals. Based on this data, we find that sentiment in economic headlines over the time period 2008-2019 is correlated with the change in economic indicators (GDP, Unemployment, S&P 500 index) over the same time-period. Moreover, because our derived correlation outperforms the baseline general news sentiment in regard to its correlation with economic data, we believe that this correlation is substantive.

To evaluate our approach, we conducted a series of experiments using data from multiple news sources and economic indicators. Our results demonstrate that sentiment analysis can provide valuable insights into market trends and that there is a strong correlation between news sentiment and economic indicators. Our findings suggest that incorporating sentiment analysis into economic forecasting models could lead to more accurate and reliable predictions.

Overall, our research highlights the importance of considering public sentiment when analyzing economic trends and demonstrates the potential of sentiment analysis for improving economic forecasting and decision-making.

Problem Statement and Introduction:

A Google Trends query on “media bias” search interest in the United States shows an increasing trend since 2010, with a peak around the 2020 US Presidential election. In light of the increasing political and social importance of the topic of media bias in the United States, it is prudent to devise a means to objectively measure bias. To do so, one must measure the distance between the sentiments expressed in a text and the ground truth, a task that is complicated by the subjective nature of many issues in the media cycle. However, this subjectivity problem does not apply to economic conditions in the United States, whose ground truth may be ascertained through data published by the Bureau of Labor Statistics. As a result, we decide to evaluate the bias by finding the correlation between the subjective news from the media and the objective economic indicator.

Due to the availability of factual data on the economy, we choose to investigate the extent of news outlets’ bias on reporting on economic conditions. In order to gain the comprehensive economic situation, we choose the most typical economic indicators:

unemployment rate, GDP, and S&P 500. We use language models to detect the sentiment (positive, negative) and analyze the relationship between sentiment and the economic data .

Related Work:

The project is largely motivated by a previous paper, "Longitudinal analysis of sentiment and emotion in news media headlines using automated labeling with Transformer language models" by David Rozado, Ruth Hughes, and Jamin Halberstadt. According to its abstract, this paper explored the "sentiment and emotion in 23 million headlines from 47 news media outlets" (Rozado, Hughes, Halberstadt 1), to seek the pattern governing readers' "[decision] whether to engage more in-depth with an article's content." The said project utilizes a Transformer language model, SiEBERT, a derivative of the larger RoBERTa-large Transformer architecture, to provide multi-label classification of headlines.

While Rozado, Hughes, and Halberstadt's conclusion focused on the "increase of sentiment negativity in headlines across news media outlets," it motivated the conception of this project to be redone through an economic lens. Rozado, Hughes, and Halberstadt found a pattern of downward trending sentiment across headlines of popular news media outlets. Through this lens, this project instead aims to discover potential trends with economic data potentially being explained by changes in sentiment of news headlines.

Dataset and reference to Longitudinal Analysis Paper:

(1) Data Collection

For headline data, we acquired 23+ million URLs of the article's headlines from the Longitudinal Analysis Paper by Rozado, Hughes, and Halberstadt as the base data. It contains the data ranging from 2000 to 2019 and including the URLs, years of news, the first and last words of the headline, and the sources(47 news media outlets). The category of news is not restricted to the economy. Firstly, we get the 200,000 samples from the base data as the control group. The 200,000 data mainly concentrates on the news after 2007. Reuters, the Guardian, and Dailymail are the top three news media with higher proportions. Next, we created the Economic Relevance Filter to gain the 400,000 economic-related data. After that, we sampled 400,000 data to have about 47,000 data used for sentiment analysis. Above 99% data is after 2007. Under this dataset, Reuter, Fox, and WallStreetJournal become the top three news media. Due to the lack of headline and specific date, we scrape necessary data through the urls for the following processing.

For annotation data, we manually annotate the economic related data as our training data and testing data to find the accuracy of the Economic Relevance Filter. The human sentiment annotations for a small subset of headlines used as ground truth to evaluate the Sentiment Analysis System.

For economic indicators, we collect the yearly, monthly data of unemployment rate and S&P 500, and yearly data of GDP ranging from 2000 to 2019. The data is mainly from the World Bank and Stooq. We plot the data for finding the relationship and divergence.

(2) Data processing

This project focuses on the headline aspect of news articles. Our intuition led us to believe that as the writer, the headline would represent the first impression of their article. Thus, it was important to communicate the sentiment in this title. What's more, in order to make comparisons, we also need the date data to assure the time span of the data. As a result, the following describes the steps used to extract the headline text and date from a large quantity of news article URLs.

For the headline, first, web scraping was used. Using the Scrapy python library, a web crawler was implemented to parse through the URL set. Some websites had anti-crawling functionalities built into their websites. For example, requests to the Washington Post timed out after many consecutive tries. However, for sites like CNN and Fox Business, web scraping was ideal as the headline could be separated from the rest of the article. Second, we addressed the issue with sites not returning headlines to the web scraper. We noticed that the majority of the articles had their headlines in the URL itself. For example, a Wall Street Journal article about the Canadian dollar has the URL below.

`https://www.wsj.com/articles/canadian-dollar-drifts-lower-amid-lull-in-domestic-data-1376337850?tesla=y`.

Since WSJ has a pay wall that prohibits non-authenticated users from viewing the article, our web scraping tool did not parse the headline. The alternative we chose to implement for URLs like these was a regular expression to extract the presumed article name from the URL. In this example, it would return “canadian dollar drifts lower amid lull in domestic data”.

For the date, because the Scrapy python library does not support extracting the date, we simply use the BeautifulSoup python library to extract the text content for that website and perform the search of date and time. Similarly, we use the regular expression for solving the non-returned data through the URL itself. For example, Newsmax has the URL https://www.newsmax.com/finance/streettalk/ibm-revenue-earnings-profit/2018/01/18/id/838044/. The regular expression helps us extract the date, which is “2018/01/18” in this case.

Lastly, we make a combination and transfer the data type to the unified format of using a regular expression on the url and scraping from the url as our final dataset.

(3) Potential Data Bias

In order to obtain our economic news dataset, we subset the Rozardo, Hughes, and Halberstadt (RHH) Dataset by running it through the economic-relevance filter described in the Approach Section. We found that our filter was disproportionately likely to identify economic news from three sources (Reuters, Fox Business, and Wall Street Journal), and observations after 2007. The economic filter may have been reproducing unforeseen biases in our annotations.

Left, the temporal distribution of observations in the RHH Dataset (Figure 1, Figure 2); right, the share of observations by news source in the Economic News Dataset (Figure 3, Figure 4).

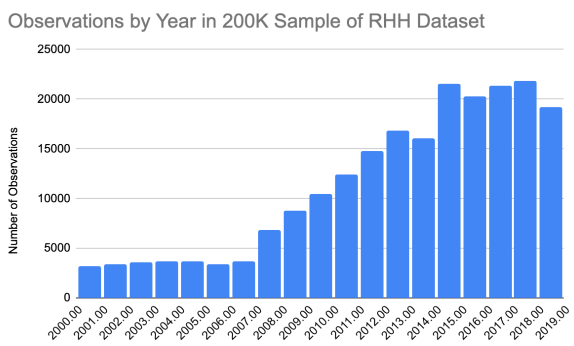


Figure 1

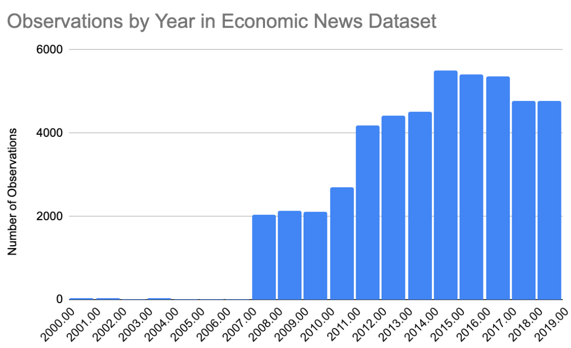


Figure 2

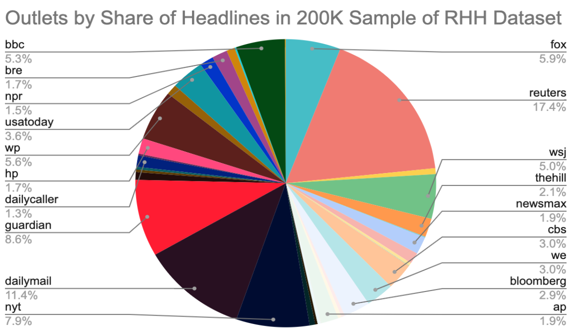


Figure 3

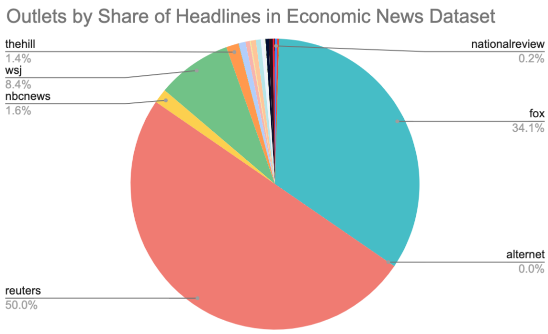


Figure 4

Thus, to reintroduce a diversity of sources into our Economic News Dataset, we first generated a large sample of more than 7 million headlines for RHH Dataset in order to maximize the number of observations from low-observation sources. Filtering this output produced 400,000 economic related headlines, disproportionately composed of Fox Business and Reuters observations. Setting a quota of 1,000 observations per month per source, the 400,000 large output was further sampled until the final Economic News Dataset of roughly 47,000 observations was produced. The resulting 47,000-observation dataset was both less time-consuming to work with and restored some diversity in news sources to the Economic News Dataset.

Annotations, Economic Relevance Filter:

To determine the trend in the sentiment of economic news from a general news dataset of 23 million headlines, we first needed to devise a means to filter out non-economic news. To this end, we trained a machine learning system using scikit-learn to predict which headlines were economically related using a manually annotated training dataset of size n=2141 (0.01% of the dataset). For each headline, we assigned a binary score: ‘1’ if the headline is economic news, and ‘0,’ if not. We determined that a headline was economic news if it satisfied any of the following criteria: the headline contains macroeconomic news, business news, or news concerning the stock market.

To determine the accuracy of our annotations, three group members annotated an annotation dataset of n=582 samples, which was created by sampling every fourth headline from the training dataset. The three annotators agreed 87.5 percent of the time. The Kappa was 0.833, indicating good agreement.

The performance metrics of the model trained on 2141 manually annotated headlines with a 90%-10% train-test split are shown below:
(‘1’ denotes that a headline is economic news; ‘0’ denotes not economic news)
The below results is the output of the system after training on a representative sample of 0.01% of the dataset

```
Accuracy: 0.9196261682242991
Precision: 0.9018324390272794
Recall: 0.9196261682242991
F1 score: 0.8940562756611742
              precision    recall  f1-score   support

           0       0.92      0.99      0.96       489
           1       0.67      0.13      0.22        46

    accuracy                           0.92       535
   macro avg       0.80      0.56      0.59       535
weighted avg       0.90      0.92      0.89       535
```

It is noteworthy that the recall of ‘1,’ (the proportion of economic-related headlines that the model identified) is low, at 0.13. However, low recall does not pose an issue, given the large size of the dataset (23 million headlines). Secondly, the model shows a high precision of 0.92 at the more important task of determining which headlines are not economically related.

Despite the precision of 0.67 on ‘0’ (economic news), the annotating standard for economic news was sufficiently conservative that a larger number of headlines not annotated as economic news could also be considered economic news. For this reason, a reasonable observer can see that the overwhelming majority of headlines returned by the system are economically related. Shown below is a random sample from 50,000 economically-related headlines.

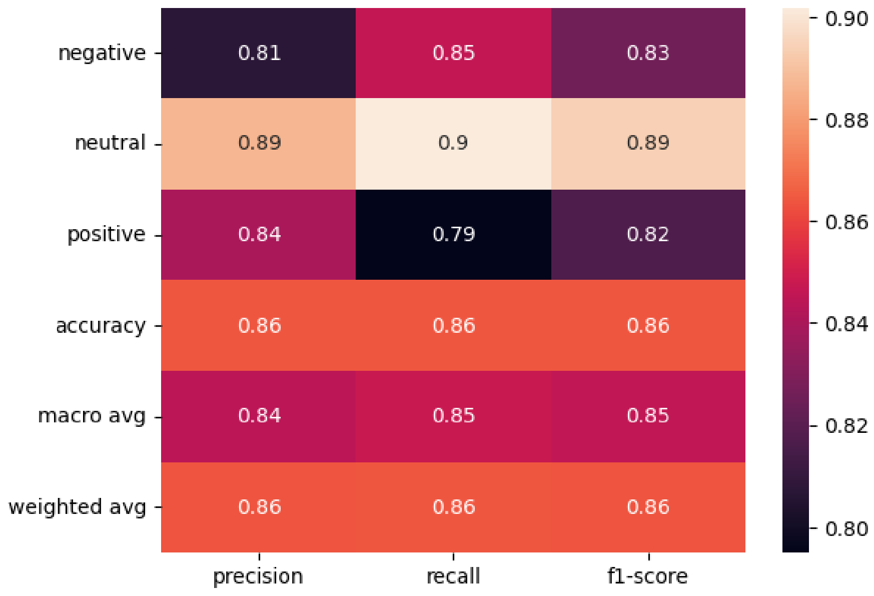
"CORRECTED-US STOCKS-Wall St tumbles, led by chipmakers; S&P lowest since May"
Regal Entertainment Q2 Estimates Cut At B. Riley Due To Weak Domestic Box Office
"Analyst View: China's Xi promises to open economy, lower tariffs"
Consolidate $30K Credit Card Debt: Good Idea?
Dividend ETF Beats S&P 500 For A Decade
U.S. Stocks Step Higher; Bond Markets See Lower Rates Longer
"BRIEF-Apple Is Said To Plan Giant High-End iPhone, Lower-Priced Model - Bloomberg"
"UPDATE 1-Canadian Oil Sands profit rises, cuts distribution"
EURO GOVT-Bunds hold near highs on Greek debt swap nerves
"ECB buying supports Italian, Spanish bonds"
Sony Books Highest Q1 Profit Since 2006 on Strong Sensor Sales
Australia shares set for a strong start on U.S. growth hopes
Gold Prices Steady as Dollar Advances
"NCR says 3Q earnings, revenue below estimates; lowers 2014 revenue outlook; shares tumble"
Consumer watchdog tightens mortgage rules on banks
5 Key Takeaways From Danaher Corporation's Earnings Report
Rising Debt Equals Higher Taxes Now and in the Future

Sentiment Analysis Component

The sentiment analysis model, being a big part of the study, needed to accept an input term in the form of a news headline and be able to output the relevant amounts of positive, neutral, and negative sentiment in the input. It was decided early on that a multi-label classification model would be better suited for such a task as opposed to a multi-class model as the latter would only generate the discrete result (e.g. is it positive or negative).

After settling on the purpose of the system, the project decided to utilize a fine-tuned Bidirectional Encoder Representation from Transformers (BERT) model developed by Google researchers in 2018. BERT is based on the transformer architecture and is designed to pre-train deep bidirectional representations of language that can be fine-tuned for a wide range of natural language processing tasks. A masked language modeling objective is used during training which enables BERT to predict words based on the surrounding context. With access to contextual patterns from phrases before as well as after its locality, it could more accurately predict the meaning of the masked words. The base model was trained from the BooksCorpus, a 800 million word corpus, as well as Wikipedia, another 2500 million words.

The model used to fine tune the BERT model was provided by a dataset on Kaggle, “Sentiment Analysis for Financial News (Ankur Sinha).” First, a multi-label classification model was initialized from the HuggingFace Transformers library where the BERT-base-cased model was selected. To evaluate the performance of the model, the `train_test_split()` function was used from the sklearn library. For the training set, 25% of the model was committed to evaluation while the remaining 75% was used to train the classification model. The results of the evaluation are below.



The model has 86% accuracy which was adequate at judging whether a headline had positive, neutral or negative sentiment. Since this project has such a high amount of data points, the model mapping the actual trend of the economic data was expected to be higher.

Results & Analysis:

(1)

Our sentiment system produced scores for negative, neutral, and positive sentiment corresponding to the amount of each sentiment detected. To classify each headline with a discrete sentiment, we chose the greatest of these three scores and then encoded the resulting classification by assigning -1 to negative sentiment, 0 to neutral sentiment, and 1 to positive sentiment.

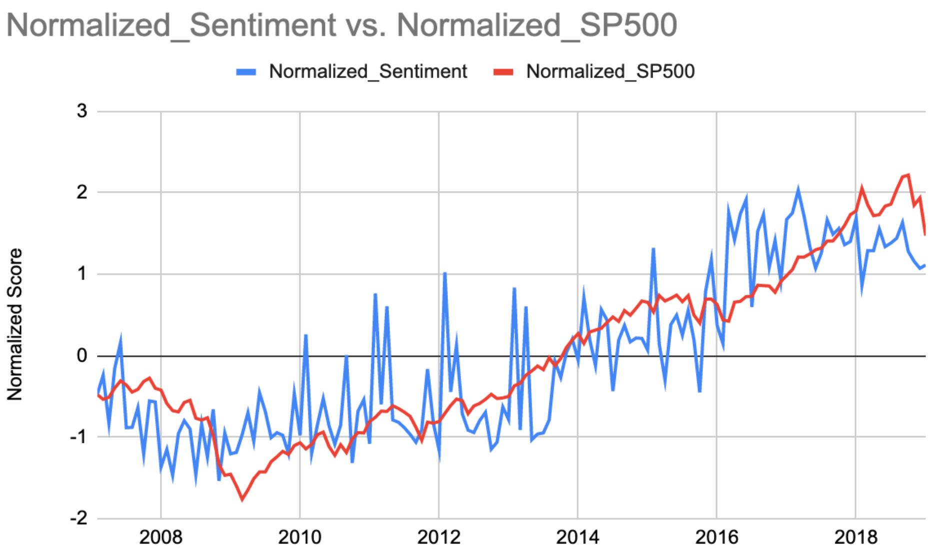
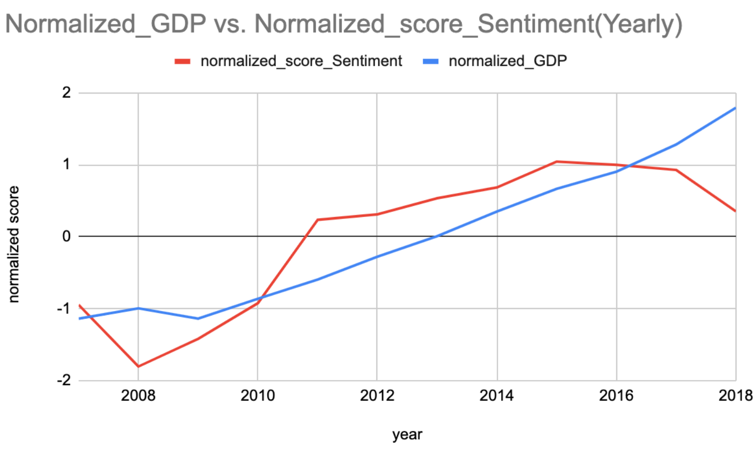


Figure 5.1: The R-squared of Normalized Sentiment Score against Normalized_SP500 is 0.695.

By plotting the average of these sentiment scores for all observations over monthly and yearly intervals, we were able to track trends in economic news sentiment. We find a correlation between economic news sentiment and three economic indicators: GDP, unemployment rate, and the S&P 500 stock index. Figure 5.1 shows the monthly average of Economic News Sentiment, plotted against the S&P 500 index.



The Economic News dataset consists of 47,986 observations, some of which have incomplete data. The data plotted over yearly intervals includes all observations in the dataset for which the headline was available (n=47,192). Data plotted over monthly intervals shows the sentiment trend for observations for which we were able to obtain the headline in order to perform sentiment analysis AND month of publication, either by web scraping or by parsing the url (n=40,395).

Figure 5.2:

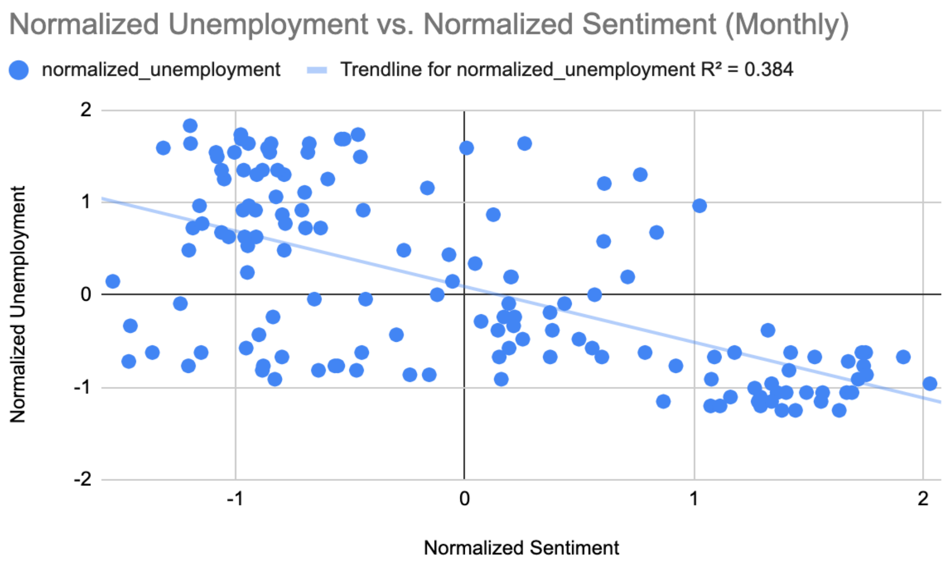


Figure 6: scatterplot showing the correlation between economic news sentiment and unemployment rate (monthly intervals)

(2) Comparison of Sentiment Trend with Baseline:

Below, we plot our economic news sentiment against (Normalized_score_Sentiment (Yearly) the economic indicators on the left (Figure 7, 9, 11). On the right, we plot general news sentiment against economic indicators (Figure 8, 10, 12). This comparison is intended as a baseline with which to compare our economic news sentiment trend. For each chart, the economic sentiment score (left) is a superior predictor of economic indicators than the general news baseline (right).

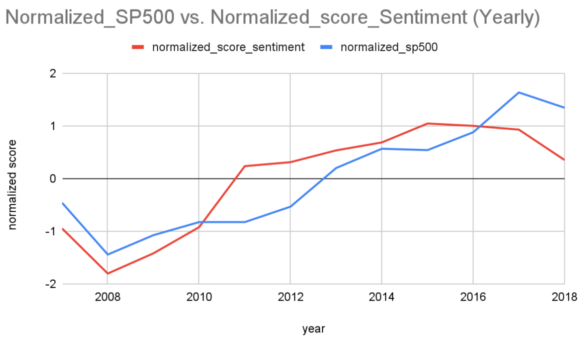


Figure 7

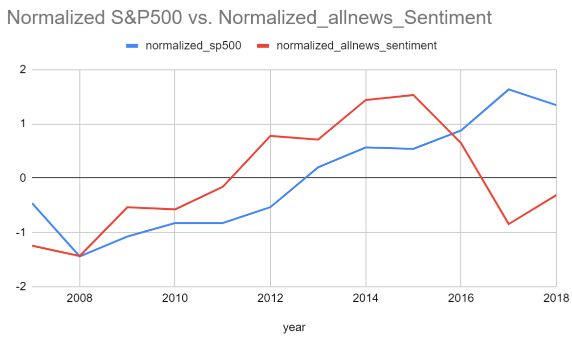


Figure 8

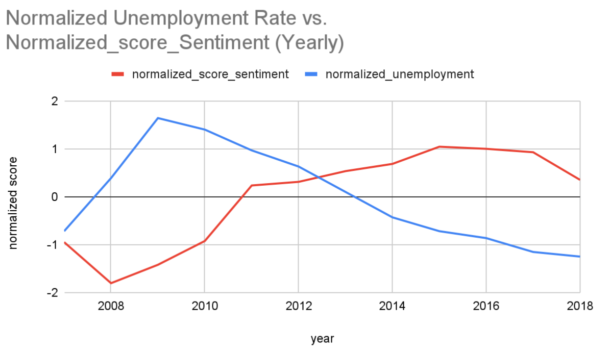


Figure 9

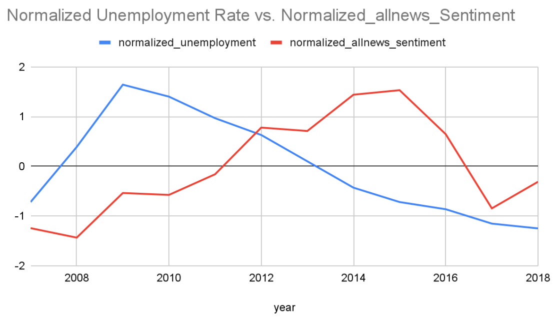


Figure 10

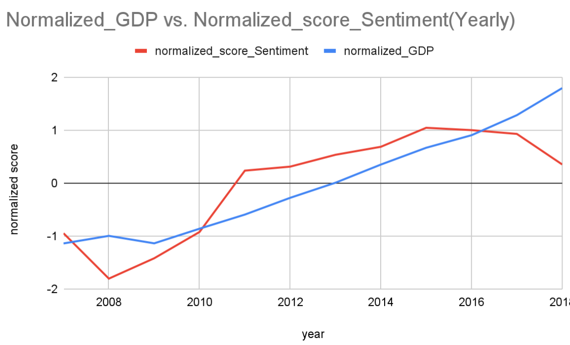


Figure 11

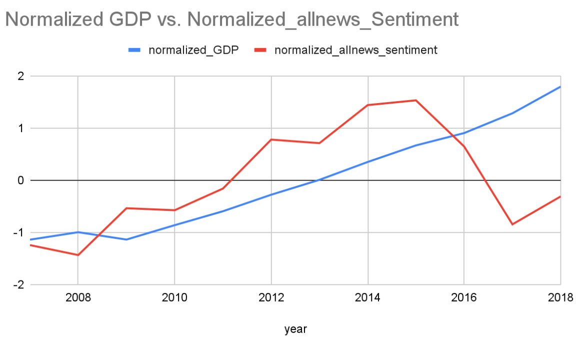


Figure 12

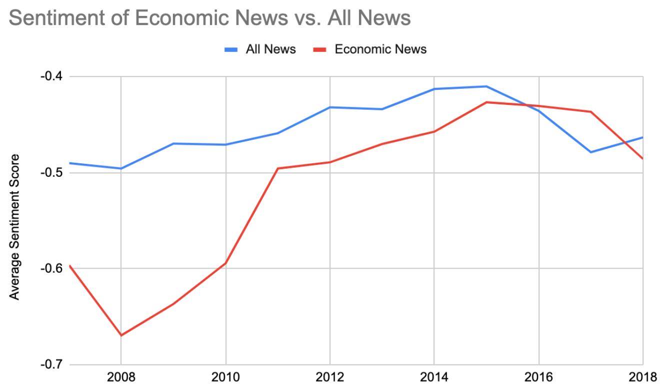
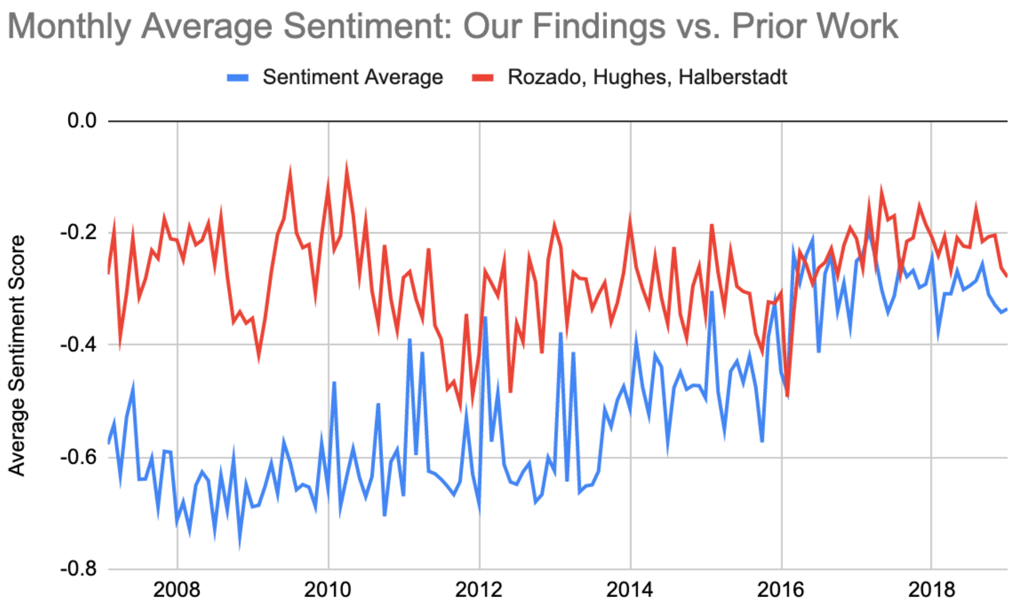


Figure 13

Figure 13 shows the yearly average sentiment for economic news versus all news in the RHH Dataset. We find that economic news sentiment was far lower than general news at the beginning of the Great Recession, but converged with general news sentiment by the end of the 2010s.

(3) Comparison with Prior Work

We compared our sentiment scores for the headlines in the Economic News Dataset with the sentiment found from Rozardo, Hughes, and Halberstadt (RHH) for the same observations. In this analysis, we reconciled the disparate classification of sentiment in our dataset and in the RHH Dataset. Whereas our sentiment score discriminates between negative (-1), neutral (0), and positive (1) sentiment for a given headline, the RHH Dataset assigns six emotions (disgust, sadness, anger, neutrality, joy, and surprise). To compare our scores with RHH, we mapped their emotions onto negative, neutral, and positive as follows (disgust: -1, anger: -1, sadness: -1, neutrality: 0, surprise: 0, joy: 1).



We found that whereas our scores showed an increasing trend in the timeframe 2007-2019, the corresponding RHH sentiment showed no trend. This is likely because Rozardo, Hughes, and Halberstadt sought to classify general news headlines by emotion, whereas our model classifies between positive economic news and negative economic news.

The more relevant nature of our binary classification to the task of measuring positive and negative economic conditions may cause the increased correlation between our Economic News Sentiment Score and the economic indicators relative to general news sentiment.

Future Work

As stated in the problem statement, we wish to find the correlation between economic data and the sentiment results of news articles. However, there are two main difficulties we encountered: first, our Economic News Dataset does not capture full diversity in news sources in the RHH Dataset. In addition, we were only able to scrape the month of publication for 84% of the Economic News Dataset, limiting our analysis over monthly time intervals. For the remainder of the dataset, we were only able to obtain the year of publication. Second, anti web scraping techniques were used by many sites, including Wall Street Journal, ABC News, The Hill and Washington Examiner. Due to access difficulties, we did not obtain as much information as we originally planned. Second, the dataset did not include enough news articles and URLs after 2018. As the figures shown, the sentiment scores range from 2007 to 2018. If data after 2018 could be obtained, we could analyze the news trend before and during COVID-19 pandemic, which would provide more insights on how global epidemics could affect the sentiment and economy in the US.

The difficulty caused by anti-crawler protection could be resolved by using proxies, which helps avoid detection or blocking. In the future, we want to continue the project by keeping adding data and obtaining the information from the latest news. In addition, using a personal laptop for such a huge dataset takes longer than we expected. Using computers from the lab would be helpful for running scraper and sentiment analysis more efficiently.

An indicator for the US economy can be developed using the data collected. In the future, we wish to train a prediction model that computes GDP, unemployment rate, S&P 500 and other economic indicators based on the news sentiment results in the past few years. Another task that could be done in the future is increasing the manually annotated headlines. Due to the time restriction, we were able to manually annotate 2500 headlines. However, by increasing the ground truth data, the model accuracy could be improved.

References

Design and evaluation of SentiEcon: A fine-grained ... - researchgate. Accessed May 9, 2023. https://www.researchgate.net/publication/341411688_Design_and_Evaluation_of_SentiEcon_a_fine-grained_EconomicFinancial_Sentiment_Lexicon_from_a_Corpus_of_Business_News.

Anonymized. “Data for Manuscript: ‘Longitudinal Analysis of Sentiment and Emotion in News Media Headlines Using Automated Labelling with Transformer Language Models.’” Zenodo, July 29, 2021. https://zenodo.org/record/5144113#.ZFnQw-zMI6A.

Combining social cognitive theories with linguistic features for multi ... Accessed May 9, 2023. https://blender.cs.illinois.edu/paper/sentiment_2012.pdf.

Rozado, David, Ruth Hughes, and Jamin Halberstadt. “Longitudinal Analysis of Sentiment and Emotion in News Media Headlines Using Automated Labelling with Transformer Language Models.” PLOS ONE. Accessed May 9, 2023. https://journals.plos.org/plosone/article?id=10.1371%2Fjournal.pone.0276367.

Hu, Jiahui, Patrick Paroubek, and Dirk Schumacher. "Annotation model and corpus for opinionated economy and finance narrative detection." In Proceedings of the 3rd Financial Narrative Processing Workshop, pp. 61-66. 2021.