Predicting the Federal Funds Rate: A Machine Learning Model Using Natural Language Processing on Federal Reserve Documents, Macroeconomic Indicators, and Company Reporting Sentiment Data

Henry Irwin, Jackson Battey, Jon Jafarnia
Data Mining & Knowledge Discovery
Fall 2023

Abstract

What is predictive of the future Fed Funds Rate? Using Federal Reserve member speeches, FOMC minutes, and FOMC statements, we extract lexical and semantic features through natural language processing. We pull macroeconomic indicators from Yahoo Finance, FRED, and Quandl. Finally, we scrape and analyze sentiment for 10-K and 10-Q filings for the 30 companies in the Dow Jones Industrial Average. We gather this data from 2000 to 2023. We then analyze our four features—lexical features, semantic features, macroeconomic indicators, and microeconomic sentiment—evaluating their predictiveness of the Fed Funds Rate 60 days in the future. Specifically, we analyze the predictiveness of features pertaining to two questions: will the Fed Funds Rate increase, decrease, or remain the same, and to what degree will the Fed Funds Rate change? To evaluate these questions, we use a logistic regression classification model and a linear regression model, respectively. We find evidence for Fed member speeches, FOMC minutes, and FOMC statements predicting both classification and degree. We find evidence for macroeconomic indicators predicting classification. Finally, we find evidence for microeconomic indicators predicting both classification and degree.

Keywords: Federal Funds Rate, Macroeconomics, Microeconomics, Natural Language Processing, Company Reporting Sentiment Data, Loughran-McDonald Dictionary, TF-IDF Vectorization, Google News Word2Vec Model

I. Introduction

The Federal Funds Rate (Fed Funds Rate) is the target interest rate set by the Federal Open Market Committee (FOMC): the policymaking body of the Federal Reserve (Fed). Specifically, it is the rate at which depository institutions—commercial banks and credit unions—borrow and lend their excess reserve balances to each other overnight. The word “target” is important here as the Fed Funds Rate is simply a target rate for which banks should lend money to each other, the Fed cannot enforce adherence to the suggested rate.

Congress states the goal of the Fed as: “maximum employment, stable prices, and moderate long-term interest rates.” This has become commonly known as the dual mandate, as the forces of inflation and unemployment generally work in opposite directions. An increase in spending associated with a booming economy, creates jobs but can increase inflation. Similarly, a decrease in spending, associated with a weak or recessionary economy, decreases inflation but destroys jobs. In order to maintain sustained growth and a healthy economy, the Fed targets an unemployment rate of ~4.4% and an inflation rate of ~2%.

The Fed manipulates the Fed Funds Rate to accomplish its goals. The Fed Funds Rate essentially serves as the risk-free lending rate, commonly known as the price of liquidity or cash. If the Fed Funds Rate is low, liquidity is cheap, meaning people can easily borrow and spend money. If the Fed Funds Rate is high, liquidity is expensive, meaning people cannot easily borrow and will not spend money. Thus, the Fed Funds Rate is manipulated to manipulate the amount of spending in the economy to accomplish the Fed’s long-term inflation and unemployment goals.

Accurately forecasting the Fed Funds Rate provides a significant advantage when investing in any asset, whether it be stocks or bonds. It also provides a good understanding of what to expect from the general economy in the near future. In this paper, we seek to create a model that forecasts the Fed Funds Rate 60 days in the future, and understand what the Federal Reserve considers when making decisions regarding the Fed Funds Rate. Specifically, we analyze our ability to predict the Fed Funds Rate using each of our four features—lexical features from our Federal Reserve corpus, semantic features from our Federal Reserve corpus, Macroeconomic Data, and Microeconomic Sentiment Data (from 10-Ks and 10-Qs).

II. Literature Review

Loughran and McDonald (2011)

Timothy Loughran and Bill McDonald's paper, “When is a Liability not a Liability? Textual Analysis, Dictionaries, and 10‑Ks,” is a pivotal work in sentiment analysis of a company's 10-K—an annual filing of a company’s financial performance. The authors introduce the Loughran-McDonald dictionary, a groundbreaking financial dictionary tailored for sentiment analysis within financial texts. Specifically, Loughran and McDonald’s paper demonstrates how word categorization into positive and negative sentiment groups creates an understanding of the tone and sentiment conveyed in financial documents. Our study makes use of this dictionary to quantify the tone and sentiment of FOMC minutes as well as 10-K documents from the 30 companies in the Dow Jones.

Jones (1972)

Spärck Jones' seminal work introduces the measure of inverse document frequency (IDF), demonstrating its utility in differentiating words and other terms (such as n-grams) based on their specificity within a collection of documents. Specifically, it reveals the significance of IDF in its accounting for the discriminative power of terms in a corpus. The paper formed the basis for the now commonly used TF-IDF vectorization of texts and documents. This method represents terms in vector space, through assigning weights to words in a document based on their frequency in the document and rarity across a collection of documents. The vectorization allows for feature extraction methods such as cosine similarity on textual data. Our model and research make use of TF-IDF vectorization and cosine similarity as two of our main methods of feature extraction, and are integral to our analysis of FedSpeak.

III. Data

Federal Reserve Speeches, FOMC Minutes, and FOMC Statements

The Federal Reserve keeps a historical record of its members' speeches on economics and economic policy, as well as the minutes and press statements from each FOMC meeting. We scrape the Federal Reserve's website obtaining all available speeches, FOMC minutes, and FOMC statements since 2000. These texts provide a corpus with which we can analyze FedSpeak and its predictive value on the Federal Funds Rate.

a. Lexical Features

Specifically, we use TF-IDF vectorization to represent these documents in vector space. TF-IDF vectorization uses the importance of a word in a document relative to its corpus, assigning higher scores to terms that are frequent in a document when they are rare across its corpus. These lexical features are integral to the use of mathematical modeling on text.

b. Semantic Features

Furthermore, we wanted to make sure to include semantic features—the meaning and sentiment of the text—in our model. We extracted semantic embeddings from each Fed document using the Google News Word2Vec model. Word2Vec models are trained by obscuring a word and using the context in which the word appears to make a prediction about what the word is. This results in a computer learning the context or meaning of a word—a word's semantics. The Google News Word2Vec model was trained on a large corpus of Google News articles and, as a result, has a large number

Macroeconomic Indicators

Using Python APIs for Quandl, Yahoo Finance, and Fred, we scrape macroeconomic data on the Effective Federal Funds Rate, 2-year Treasury Yield, Real Gross Domestic Product, Core Consumer Price Index, Price Consumption Expenditures Index, Unemployment Rate, Savings Rate, Retailers Sales, Manufacturing Purchasing Managers' Index, University of Michigan Consumer Sentiment, S&P 500 Index, Volatility Index, and Liquidity (defined as the volume traded of the S&P 500). We impute the missing data using the most recent value. We chose the most recent value because missing data is often due to a lack of reporting–some measures are only published monthly or quarterly.

Furthermore, we wanted the model to be able to learn the Fed's mindset of relativity. Therefore, we calculated the 1-year percent change of each macroeconomic indicator, such that relativity is included for each measure.

Microeconomic Sentiment

The Securities and Exchange Commission keeps a historical record of all public company filings since their initial public offering. This database provides access to all annual (10-K) and quarterly (10-Q) filings of a company's financial performance, given its CIK number (Central Index Key given by the SEC). Our research takes the CIK of the most recent 30 companies in the Dow Jones and scrapes all 10-Ks and 10-Qs, providing a corpus from which we can extract features to use in our model.

To extract sentiment from these filings, we count the number of positive and negative words in each document according to the Loughran-McDonald dictionary. We then create a sentiment score using the following formula:

$$Sentiment\ Score\ =\ \frac{(Positive\ Word\ Count - Negative\ Word\ Count)}{(Positive\ Word\ Count + Negative\ Word\ Count)}$$

IV. Methods

Using our four features—TFIDF lexical features from Federal Reserve text, Google News Word2Vec semantic embeddings from Federal Reserve text, microeconomic sentiment, and macroeconomic indicators—we ran regression and classification models, comparing them to a baseline model to assess their effectiveness.

Baseline Model

Our baseline for regression was a simple autoregressive model in which our only feature was the current Fed Funds Rate. Our baseline for classification was a dummy model that predicted the most frequent Fed Funds Rate movement: increase, decrease, or remain the same. All four of our features were able to outperform the baseline.

Cross Validation & Evaluation

Next, we tested all possible combinations of features using 10-fold cross-validation on the dataset. We did this for both regression and classification. Cross-validation is incredibly important to ensure our model is generalizable to unseen data and verifies its performance isn't due to a certain split of the data.

For regression, our main metric was mean squared error as we wanted to penalize large deviations from the actual rate. We also included mean absolute error to understand how each model performed when all deviations from the actual rate are treated equally.

For classification, our main metric was weighted F1 score. We decided to use weighted F1 as it is a combination of precision and recall. Its consideration of false positives and false negatives ensures a more comprehensive evaluation, especially in scenarios where class distributions are unequal. We also made sure to include an accuracy measure as a baseline.

Classification and Regression Models

For our regression model, we predict the change in the Federal Reserve Effective Funds Rate from the current rate to the rate 60 days in advance, using a linear regression model.

For our classification model, we predict if the rate would increase, decrease, or remain the same using a logistic regression model.

V. Results

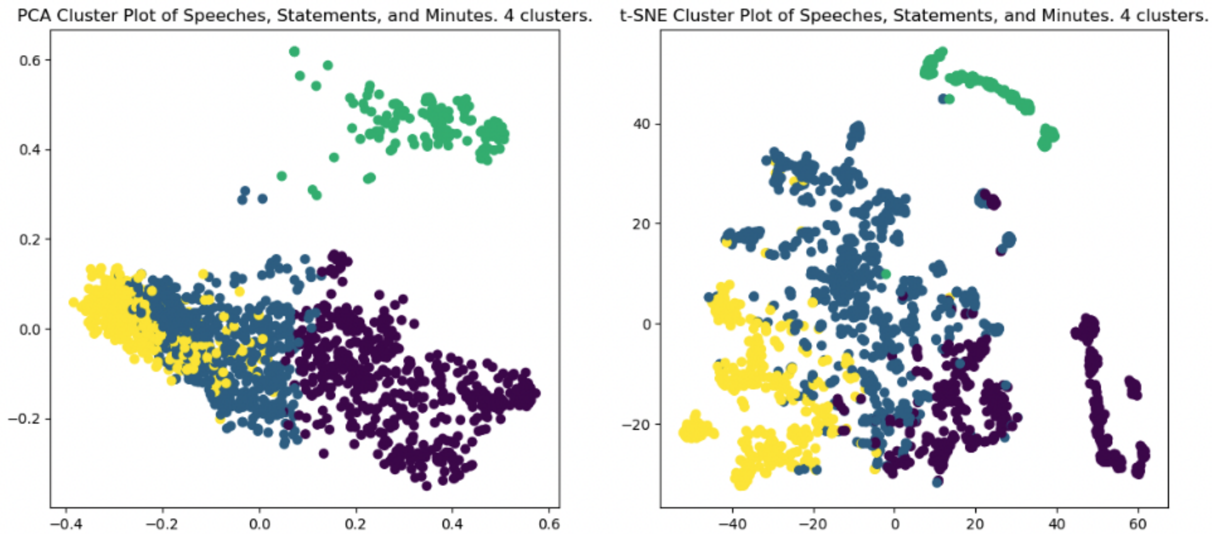
Federal Reserve Speeches, FOMC Statement, and FOMC Minutes Analysis

To analyze and understand our text features, we employ clustering. After initial TF-IDF vectorization, we apply unsupervised k-means clustering to Minutes, Statements, and Speeches. We also cluster individually on minutes and statements, as well as speeches. We visualize our results using PCA and T-SNE plots. PCA plots reduce the dimensionality of our data while preserving variance, and T-SNE plots explore nonlinear relationships between data points.

An analysis of these clusters reveals that FOMC Statements and Minutes are much more linear as they tend to strictly deal with policy regarding the Fed Funds Rate, while the Federal Reserve member speeches are very non-linear as they tend to cover a variety of subject matters at any given time.

a. Minutes, Statements, Speeches

Figure 1: Minutes, Statements, and Speeches Clusters



Generally, in Figure 1, we see clear clusters, meaning our lexical features are likely pretty solid. However, there is some blurriness between clusters, meaning multiple clusters share characteristics. Furthermore, the T-SNE plot exhibits less blurriness, indicating the presence of non-linear relationships within the data. To investigate where this nonlinearity is coming from, we break our corpus into two smaller corpora: minutes and statements vs. speeches.

b. Minutes and Statements

Figure 2: Minutes and Statements Clusters

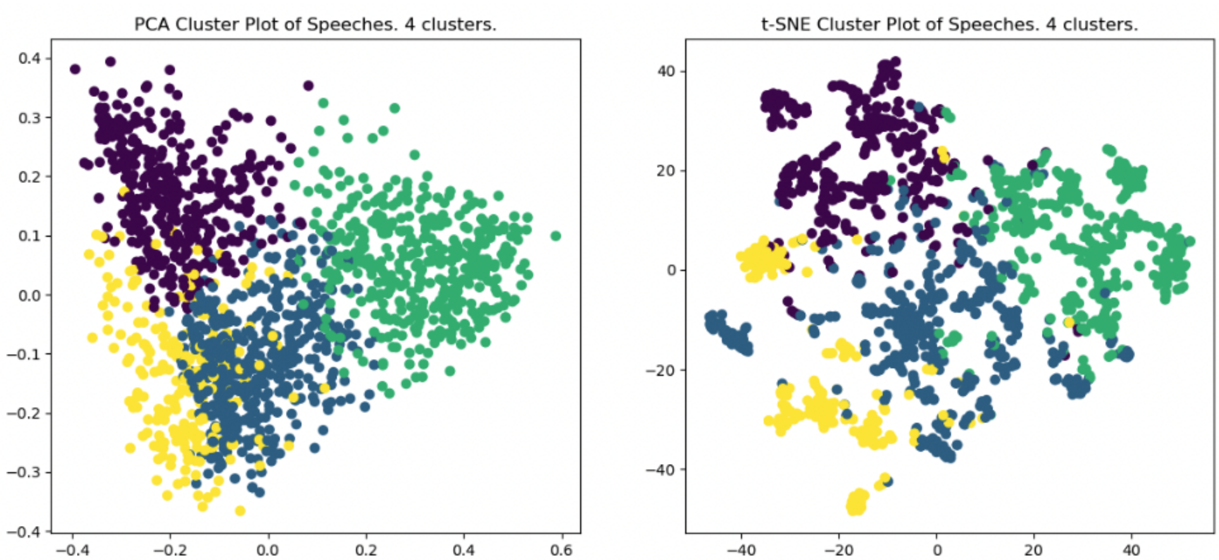
In Figure 2, the clusters using minutes and statements appear to be much more linear. It is important to note that the sum of the variance between clusters plateaus around three clusters instead of four clusters as seen in Figure 1. Interestingly, in our classification models, we use three categories: increase, decrease, and remain the same. While we cannot guarantee that the three clusters correlate to the classifications we used in our models, there is meaningful evidence for this correlation. Specifically, looking at the top 10 words in each cluster, we see evidence for our three categories.

c. Cluster 1 (Decrease): basis, Alan, foreseeable, today, sustainable, Roger, discount, believes, growth, jr
d. Cluster 2 (Increase): structure, payment, conditions, securities, tools, reports, supervision, bank, data, financial
e. Cluster 3 (Remain the Same): continued, board, remained, growth, meeting, prices, period, quarter, members, participants

Words such as “sustainable” and “growth” correlate to speeches where the Fed wants to foster sustainable growth through a decrease in the Fed Funds Rate. Words such as “supervision,” “financial,” “data,” and “conditions” correlate to speeches in which the Fed wants to supervise economic conditions and financial data due to concern over the economy, likely correlating to increases in the Fed Funds Rate. Finally, words such as “continued,” “remained,” and “quarter” likely correlate to periods in which the Fed’s policy will remain the same as they are waiting for more data to be released in the next quarter.

f. *Speeches*

Figure 3: Speeches Clusters



In Figure 3, the clusters using speeches appear to be very blurry and nonlinear. This makes a lot of sense, as these speeches tend to be on a wide variety of topics with little to no guidelines. An analysis of the top 10 words in each cluster provides evidence for clustering on speech topic.

- Cluster 1: firms, credit, institutions, liquidity, management, Basel, banking, capital, banks, risk
- Cluster 2: states, monetary, text, percent, capital, payments, policy, education, growth, productivity
- Cluster 3: growth, prices, rates, labor, percent, FOMC, rate, monetary, policy, inflation
- Cluster 4: housing, small, loans, lending, CRA, communities, credit, mortgage, banks, community

Cluster 1 appears to be speeches regarding risk management. Cluster 2 appears to be speeches regarding sustainable growth. Cluster 3 appears to be speeches regarding monetary policy. Cluster 4 appears to be speeches regarding the housing and mortgage market.

Macroeconomic Indicator Analysis

Figure 1 and Figure 2 show correlation matrices of the macroeconomic indicators. Using these figures, we can make inferences as to what features will be predictive of the Fed Funds Rate. We notice that gross domestic product (Real GDP), inflation (Core CPI & PCE), unemployment rate, savings rate, consumer sentiment, and liquidity are all highly correlated with the Fed Funds Rate. Interestingly, the 1-year change in Manufacturing PMI is also highly correlated with the Fed Funds Rate. These correlations are supported by the underlying economics behind the Federal Reserve's decisions to change the Fed Funds Rate, and we expect these features to be important in crafting a predictive model.

Figure 4: Macroeconomic Indicators Correlation Matrix

	Eff_Fed_Funds	Real_GDP	Core_CPI	PCE	Unemployment	Savings_Rate	Retail_Sales	Manufacturing_PMI	Consumer_Sent	Liquidity	Volatility	SP_500
Eff_Fed_Funds	1.00	-0.43	0.38	-0.45	-0.53	-0.45	-0.01	-0.21	0.41	-0.48	-0.10	-0.22
Real_GDP	-0.43	1.00	0.16	0.99	-0.22	0.49	-0.18	0.42	-0.24	0.55	-0.18	0.89
Core_CPI	0.38	0.16	1.00	0.18	-0.64	-0.12	-0.35	-0.21	-0.07	-0.14	0.10	0.39
PCE	-0.45	0.99	0.18	1.00	-0.18	0.52	-0.17	0.41	-0.30	0.57	-0.14	0.91
Unemployment	-0.53	-0.22	-0.64	-0.18	1.00	0.33	0.19	-0.11	-0.46	0.33	0.29	-0.37
Savings_Rate	-0.45	0.49	-0.12	0.52	0.33	1.00	0.10	0.13	-0.15	0.41	0.18	0.51
Retail_Sales	-0.01	-0.18	-0.35	-0.17	0.19	0.10	1.00	0.03	0.21	-0.11	-0.18	-0.19
Manufacturing_PMI	-0.21	0.42	-0.21	0.41	-0.11	0.13	0.03	1.00	0.08	-0.01	-0.50	0.43
Consumer_Sent	0.41	-0.24	-0.07	-0.30	-0.46	-0.15	0.21	0.08	1.00	-0.55	-0.42	-0.16
Liquidity	-0.48	0.55	-0.14	0.57	0.33	0.41	-0.11	-0.01	-0.55	1.00	0.35	0.35
Volatility	-0.10	-0.18	0.10	-0.14	0.29	0.18	-0.18	-0.50	-0.42	0.35	1.00	-0.14
SP_500	-0.22	0.89	0.39	0.91	-0.37	0.51	-0.19	0.43	-0.16	0.35	-0.14	1.00

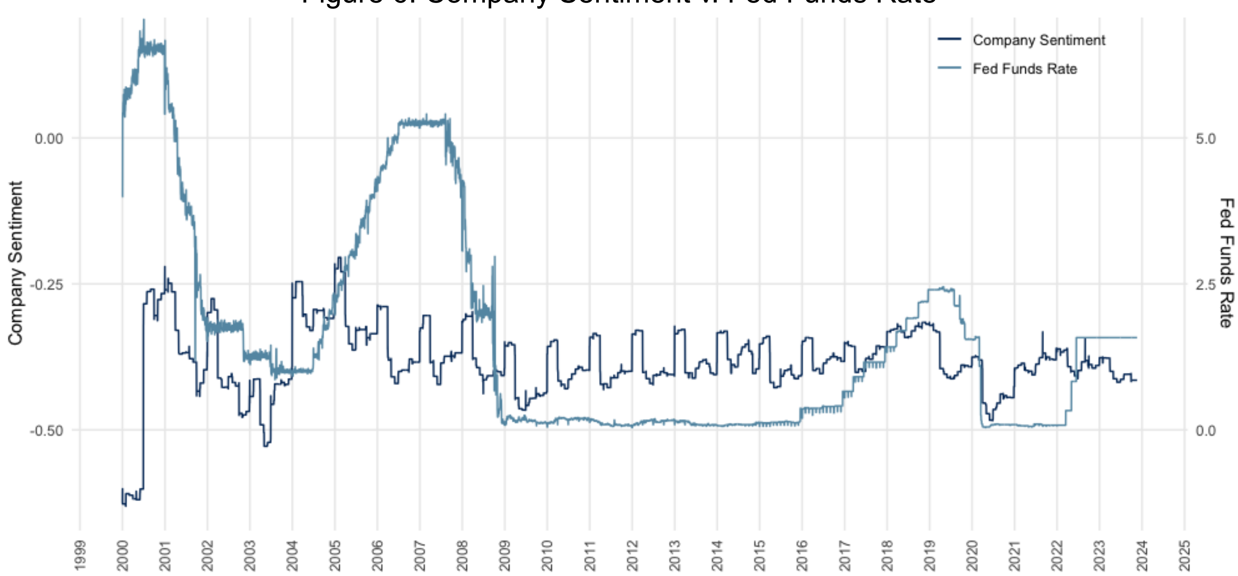
Figure 5: Macroeconomic Indicators (%Δ from 1 Yr Ago) Correlation Matrix

	Eff_Fed_Funds	Real_GDP_%Δ	Core_CPI_%Δ	PCE_%Δ	Unemployment_%Δ	Savings_Rate_%Δ	Retail_Sales_%Δ	Manufacturing_PMI_%Δ	Consumer_Sent_%Δ	Liquidity_%Δ	Volatility_%Δ	SP_500_%Δ
Eff_Fed_Funds	1.00	0.19	0.06	0.26	-0.11	-0.14	-0.05	-0.28	-0.14	0.26	0.03	-0.13
Real_GDP_%Δ	0.19	1.00	0.16	0.87	-0.80	-0.68	-0.13	0.23	0.28	-0.23	-0.41	0.50
Core_CPI_%Δ	0.06	0.16	1.00	0.22	-0.28	-0.18	-0.06	-0.41	-0.25	0.02	0.08	-0.07
PCE_%Δ	0.26	0.87	0.22	1.00	-0.68	-0.62	-0.08	0.19	0.12	-0.14	-0.37	0.44
Unemployment_%Δ	-0.11	-0.80	-0.28	-0.68	1.00	0.80	0.17	-0.01	-0.29	0.26	0.39	-0.31
Savings_Rate_%Δ	-0.14	-0.68	-0.18	-0.62	0.80	1.00	0.13	-0.11	-0.23	0.24	0.44	-0.27
Retail_Sales_%Δ	-0.05	-0.13	-0.06	-0.08	0.17	0.13	1.00	0.03	0.04	0.00	-0.04	0.09
Manufacturing_PMI_%Δ	-0.28	0.23	-0.41	0.19	-0.01	-0.11	0.03	1.00	0.40	-0.28	-0.35	0.51
Consumer_Sent_%Δ	-0.14	0.28	-0.25	0.12	-0.29	-0.23	0.04	0.40	1.00	-0.35	-0.51	0.46
Liquidity_%Δ	0.26	-0.23	0.02	-0.14	0.26	0.24	0.00	-0.28	-0.35	1.00	0.51	-0.37
Volatility_%Δ	0.03	-0.41	0.08	-0.37	0.39	0.44	-0.04	-0.35	-0.51	0.51	1.00	-0.51
SP_500_%Δ	-0.13	0.50	-0.07	0.44	-0.31	-0.27	0.09	0.51	0.46	-0.37	-0.51	1.00

Microeconomic Sentiment Analysis

Figure 3 shows a plot of the average sentiment score across the 30 most recent companies in the Dow Jones Industrial Average plotted against the effective Fed Funds Rate since 2000. What becomes apparent when looking at the chart is the importance of this feature to our model. Company sentiment is clearly predictive of the effective Fed Funds Rate—when sentiment increases or decreases, the effective Fed Funds Rate follows suit shortly thereafter. Interestingly enough, the Federal Reserve has been criticized recently for not raising rates fast enough in response to the COVID recovery, leading to rampant inflation, and this can be seen in the figure, where there is an increased lag between an increase in company sentiment and the increase in the effective Fed Funds Rate around 2020 to 2022.

Figure 6: Company Sentiment v. Fed Funds Rate



Model Results

Figure 7: Classification Results

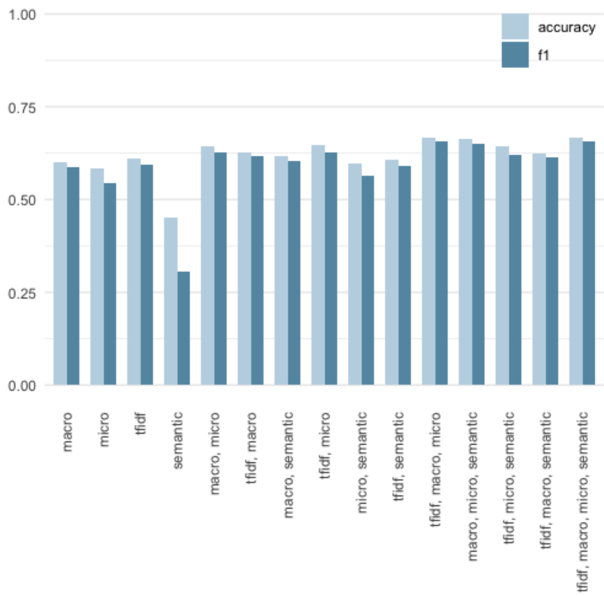
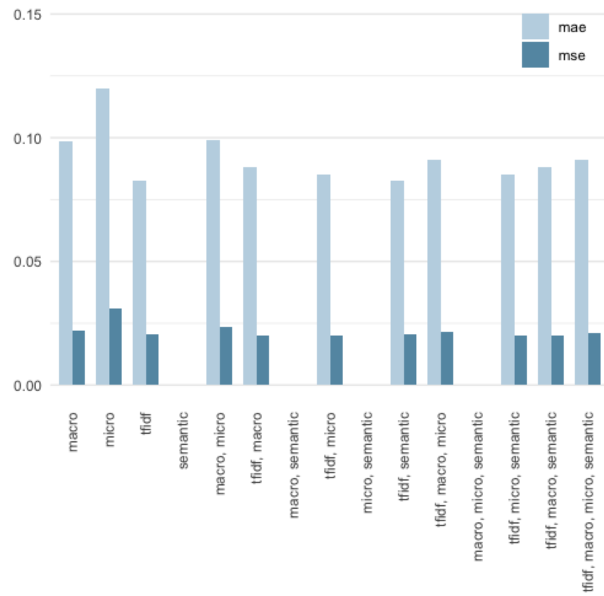


Figure 8: Regression Results



Worse models are excluded in figure 8 to allow for a reasonable scale of the y-axis

Table 1: Model Results

Features				Regression Metrics		Classification Metrics	
tfidf	**macro**	**micro**	**semantic**	**mse**	**mae**	**f1**	**accuracy**
0	0	0	**1**	0.2823	0.3560	0.3071	0.4513
0	0	**1**	0	0.0312	0.1197	0.5435	0.5835
0	0	**1**	**1**	0.5674	0.5127	0.5651	0.5985
0	**1**	0	0	0.0218	0.0984	0.5881	0.6009
0	**1**	0	**1**	0.4057	0.4340	0.6050	0.6159
0	**1**	**1**	0	0.0234	0.0992	0.6282	0.6433
0	**1**	**1**	**1**	7.2206	1.6564	0.6507	0.6632
1	0	0	0	0.0204	0.0826	0.5930	0.6111
1	0	0	**1**	0.0204	0.0826	0.5908	0.6086
1	0	**1**	0	0.0199	0.0851	0.6258	0.6485
1	0	**1**	**1**	0.0199	0.0850	0.6202	0.6435
1	**1**	0	0	0.0202	0.0880	0.6160	0.6259
1	**1**	0	**1**	0.0201	0.0880	0.6129	0.6234
1	**1**	**1**	0	0.0213	0.0911	0.6560	0.6682
1	**1**	**1**	**1**	0.0213	0.0909	0.6560	0.6682

Upon combining our features, we found that TF-IDF vectorization of text from the Federal Reserve, microeconomic sentiment, and Word2Vec semantic features created the best linear regression model. Specifically, this model had a mean squared error of 0.0199 and a mean absolute error of 0.0850.

We found that TF-IDF vectorization of text from the Federal Reserve, macroeconomic indicators, microeconomic sentiment, and Word2Vec semantic features creates the best logistic regression classification model. Specifically, this model has an F1 score of 0.650 and an accuracy of 0.6682.

VI. Discussion

Our in-depth analysis of our features reveals that macroeconomic indicators are important for the question: will the Fed Funds Rate increase, decrease, or remain the same? This follows directly from the purpose of the Fed: the Fed is targeting an unemployment rate and an inflation rate, and thus is tracking macroeconomic data and adjusting the Fed Funds Rate to hit those targets.

Microeconomic sentiment appears to be predictive of classification and degree. Interestingly microeconomic sentiment is more important than macroeconomic indicators for the question: by what degree will the Fed Funds Rate change? This is an incredibly interesting finding. Essentially, this points to the Fed focusing less on the macroeconomic measures and more on how the macroeconomics has affected businesses when deciding the degree of the Fed's possible policy changes.

Finally, we find that Fed speeches, FOMC statements, and FOMC minutes are predictive in both questions of classification and magnitude of change. This confirms the sentiment that the Fed has begun to use its interactions with the public as a way to manage expectations of the market. The Fed manages public expectations such that changes to rate policy do not catch the public off guard and cause economic turmoil, such as what occurred in the Taper Tantrum of 2013.

Future studies should analyze if audio and facial expressions of Federal Reserve speeches are predictive of the Fed Funds Rate. Furthermore, we were limited by the amount of data we could gather on FOMC speeches, so future research should look to expand the corpus of text analyzed.

VII. Citations

Academic Papers

Loughran, Tim, and Bill McDonald. "When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks." The Journal of Finance, vol. 66, no. 1, 2011, pp. 35-65.

Sparck Jones, Karen. "A Statistical Interpretation of Term Specificity and Its Application in Retrieval." Journal of Documentation, vol. 28, no. 1, 1972, pp. 11-21.

Articles

Medium. "How to Download and Scrape 10-K Filings from SEC Edgar." https://medium.com/@jan_5421/how-to-download-and-scrape-10-k-filings-from-sec-edgar-b0d245fc8d48

Towards Data Science. "Fedspeak: How to Build a NLP Pipeline to Predict Central Bank Policy Changes." https://towardsdatascience.com/fedspeak-how-to-build-a-nlp-pipeline-to-predict-central-bank-policy-changes-a2f157ca0434

Data Sources

Board of Governors of the Federal Reserve System. www.federalreserve.gov
FRED API. Federal Reserve Bank of St. Louis. https://fred.stlouisfed.org/docs/api/fred/
Quandl API. Quandl. https://data.nasdaq.com/tools/api
Yahoo Finance API. Yahoo. https://developer.yahoo.com/api/

Websites

ChatGPT. OpenAI. https://chat.openai.com/
Stack Overflow. Stack Exchange Inc. https://stackoverflow.com/