

# Predicting FOMC Actions using ML and NLP

Jeremy Lao - jjl359  
NYU Computer Science

John Reynolds - jr4716  
NYU Computer Science

April 20, 2019

## Abstract

*The Federal Open Market Committee (FOMC) meets throughout the year to set the target Federal Funds rate. The Federal Funds rate is one of the primary monetary policy tools, and it impacts the cost of borrowing money globally. This paper explores ML and NLP methods to predict the potential outcome of an upcoming FOMC using past FOMC statements and meeting minutes to train ML and NLP models.*

## 1 Introduction

In this paper we methods in Natural Language Processing and Machine Learning to predict Federal Open Market Committee (FOMC) rate actions (hold or change) using text from FOMC meeting minutes, Board of Governors speeches, and FOMC post-meeting statements. Since this is work in sentiment analysis we created a simulator to gain insight into the sensitivity of detecting sentiment by randomly generating words with from a mix of distributions containing hawkish or dovish(positive) words, negative words, and commonly used words. That simulator showed accuracy could hit levels of 100 percent given enough training data and high percentage of sentiment (i.e., positive) words to overall text.

### 1.1 What is the FOMC

The FOMC sets the Federal Funds target range (target rate prior to 2009). The permanent members are the Board of Governors of the Federal Reserve System, President of the Federal Reserve Bank of New York, and the rest of the seats are rotated through the Presidents of the other Federal Reserve Banks. There are twelve Reserve Banks and the Board of Governors oversees the activity, operations, and policies of the Federal Reserve System.

### 1.2 Why Does the Federal Funds Target Range Matter?

The Federal Funds Target Range is set by the FOMC. The Federal Funds Target Range is the range where the Federal Funds Effective rate (calculated as the volume weighted median of eligible transactions reported on FR 2420) is expected to fix on a daily basis. The Federal Funds rate is the primary monetary policy instrument of the Federal Reserve System, and it influences the level of interest rates domestically and globally. For example, if the Federal Funds effective rate were 5 percent, then the interest rate on a 30 year mortgage would be 5 percent or greater.

## 2 Methodology

### 2.1 Data

We primarily utilized three sources for our data, they are outlined below.

The time frame of our study covered 2000 to 2019. The primary source of our textual data were FOMC statements, FOMC meeting minutes, and Board of Governors speeches from [www.federalreserve.gov](http://www.federalreserve.gov). For FOMC statements between 2000 and 2008 we utilized Stanford’s FOMC corpus. The Stanford FOMC corpus only has text prior to 2008 and it does not have Board of Governor speeches.

The data for the Federal Funds target rate and target range (post 2009) is from FRED St. Louis. FRED offers a wealth of economic data and information to promote economic education and enhance economic research. FRED is updated regularly and allows 24/7 access to regional and national financial and economic data.

## 2.2 Scraping and Text Pre-Processing

We collected data from the Federal Reserve website and pre-processed the data for the document-term matrix.

We utilized Python packages *beautifulsoup4*, *re*, and *urllib* to scrape contextual data from the Federal Reserve website. Our scraping algorithm used regular expressions to handle and remove extraneous *html* and *javascript* text collected by the page scraper. We had to handle non UTF-8 characters upfront and remove them at this stage altogether. We collected nearly one-thousand documents from the Federal Reserve website.

After scraping the data, we pre-processed the text by removing all punctuation, ensuring proper spacing between words, setting all words to lower case, and making all numbers *d* to reduce the dimensionality document-term matrix. We also used Regex to detect direct references to the Federal Funds target rate/range and transformed those references (a mixture of numbers, punctuation, and the word ‘percent’) to *ddpercentrate*. Since we were modeling the sentiment of FOMC rate action, this was one of our strategies to directly capture the target rate in the document-term matrix.

This formed the bulk of our text pre-processing and it allowed us to essentially utilize the stored text in the document-term matrix for NLP and ML models.

## References

- [1] Hansen,Stephen, McMachon, Michael *Transparency and Deliberation with the FOMC:a Computational Linquistics Approach*, CEPR Discussion Paper, (2014)