Bloomberg INDG Datathon Challenge

Identifying Shifting Policy Priorities in the Federal Register.

Introduction[#]

The [Federal Register] is "the daily journal of the United States government." The register contains official announcements from public agencies including proposed and final rules. As such, the Federal Register is an important data point in reflecting the regulatory priorities of federal agencies.

Description[#]

The dataset contains rules and proposed rules published by federal agencies responsible for regulating financial markets in the United States from January 1, 2001 through July of 2021. The agencies included are:

- Commodity Futures Trading Commission (CFTC)
- Office of the Comptroller of the Currency (OCC)
- Consumer Financial Protection Bureau (CFPB)
- Federal Deposit Insurance Corporation (FDIC)
- The Federal Reserve Board (FRB)
- Securities and Exchange Commission (SEC)

The dataset contains roughly 3600 documents and additional metadata elements can be accessed via the [Federal Register API].

Access the dataset and other resources [here]. Within the folder, there is starter code inside a Jupyter notebook that should aid you in trying to parse the files inside the docs folder.

Judging[#]

As regulatory priorities of federal agencies evolve over time, new regulatory issues and topics will emerge. A useful output of this challenge would include a method, model, or tool to help identify these emerging issues as they occur and classify them in future Federal Reserve publications.

- What were the common regulatory priorities of these agencies from 2001 through 2006?
- What new topics and issues emerged from 2007 through 2012? (And what topics went awaay?)
- How significantly do topics shift between administrations? (Bush-Obama in 2008-09, Obama-Trump in 2016-17)
- How many times, on average, would a topic need to be discussed before it is flagged as an emerging topic?

Suggestions[#](#)

1. Take panels of data, a few years at a time, and analyze them against the next panel of data.
2. Unsupervised clustering would likely be useful here to identify topic clusters.
3. A tool to assist in manual review of identified clusters would also be helpful. (Perhaps a way of tracking what clusters were helpful and which clusters were not.)
4. Events for testing methods and approaches include the housing crisis of 2008 (and prior/subsequet rulemaking leading up to it) and the still ongoing COVID-19 crisis. A good unsupervised approach would identify both events as new clusters.

Prizes[#](#)

1st Place: Nintendo Switch

2nd Place: Eddie Bauer Softshell Jacket with the Bloobmerg INDG logo