

# 230ZB Notes

September 28, 2024

So overall idea I was thinking was to have GPT forecast which GICS {sector|industry group|industry} will have the highest returns over the next day (not tied to the idea in case anyone else has something they'd rather do). Currently, the model is running at the industry group level (which has more categories than the sector level but far less than the industry level). I was thinking maybe more granularity may be better, since it's such a short term strategy.

All three files should run top to bottom right now. File 1 `1_get_data.ipynb` basically just deals with getting the raw data. It does the following:

1. Get S&P constituent data (GICS membership, ticker, returns, date, etc.) and is has updated holdings as the constituents come in and out of the S&P.
2. Gets all the FOMC statements, scrapes them, and removes noise.

Then the next file, `2_construct_fine_tune_dataset.ipynb` constructs the datasets needed to finetune the model. If the FOMC statements and their date is the data, the correct label given our objective is the GICS {sector|industry group|industry} that has the highest and lowest returns over the next day. This file gets those correct labels, structures the data in the expected format for the fine tuning process, and then ensures the data is in the proper format.

The last file I pushed is the `3_gics_strategy.ipynb`. This one basically fine tunes the model, runs the model, and constructs the portfolio. I'd say this is kind of just a rough draft version that still needs some work. Also, the instructions for this is quite vague.