

1) Which tasks have been completed?

I have successfully completed code to download 10-K filings for a given list of companies from the Energy sector, and used code to automatically extract the section of risk factors from those filings. Occasionally for some companies, the code may miss the risk section and extract the previous text, so I am still working on improving the accuracy of the information retrieval task.

2) Which tasks are pending?

I still need to analyze the text extracted, and find a way to visualize the results in a meaningful way. After watching the week 12 lectures, I think netPLSA may be useful in my case for identifying systematic risks related to the industry. (Thanks to the reviewers for the great questions, they are very inspiring. To answer some questions from previous feedback, I am not only trying to identify risks related to single companies, but rather, how they collectively reflect the risk of certain industries. I also know that risk factors will likely be negative in terms of sentiments, but I think it is interesting to see the degree to which different companies vary in the negativity of the sentiments, or how that sentiment changes over time, if we compare multiple years of 10-K filings). I still have to do some research regarding which packages to use for analyzing that.

Aside from that, some other things include documentation and presentation of results. I will plan such that I have enough time to work on those.

3) Are you facing any challenges?

I am a little bit worried about my time. I suppose the second part won't take tremendously long as I have the entire thanksgiving to work on that... If it unfortunately does, I will reduce the scope of my project to focus on the algorithm and be simple when it comes to visualizing the results.