

CS410 Progress Report

Team Starks

Saad Rasheed - srashee2 -

Javier Huamani - huamani2

Sai Allu - allu2

Which tasks have been completed?

Our team has currently taken the The New York Times Annotated Corpus dataset and extracted the data given to us into more useful excel and CSV files. Our team has taken these csv files and have used different libraries (gensim, pandas, sklearn, and nltk) to create a topic model.

Which tasks are pending?

To successfully recreate the paper our team still needs to find the stock time series data and run that against our code, while refining the code to find the optimal execution. Lastly, we have to implement measures of quality such as a Granger test.

3) Are you facing any challenges?

The stock time series data seems to be not referenced in the paper. We also have an issue with deterministic / non-deterministic results when running our code.