

CS 410 Final Project Proposal: Sentiment Analysis on Financial Statements

Team Member:

Hui Huang – huih3@illinois.edu

Motivation:

A 10-K is a comprehensive report filed annually by a publicly-traded company about its financial performance required by SEC. Information documented in the 10-K includes the history, organizational structure, financial statements, earnings per share, and any other relevant data. As the 10-K document contain information which are not captured by the financials, such as the expectations about the future and risk measures. It's interesting to see if the text information extracted helps in investment decisions for the investors.

Plan:

Data: Get 10K documents from SEC website. (4h-6h as SEC has limits on the number of calls that can be made to the website)

Methods: The bag-of-words model/RNN model to model the sentiments for the companies. (15h Need to get familiar with PyTorch Framework for RNN)

Performance Evaluation: Back test on historical stock returns using the sentiments extracted. (3-6h Need to build a stock return factor model)

Expected outcome: Sentiments extracted is an alpha factor for the stock returns ☺

Programming language: Python