



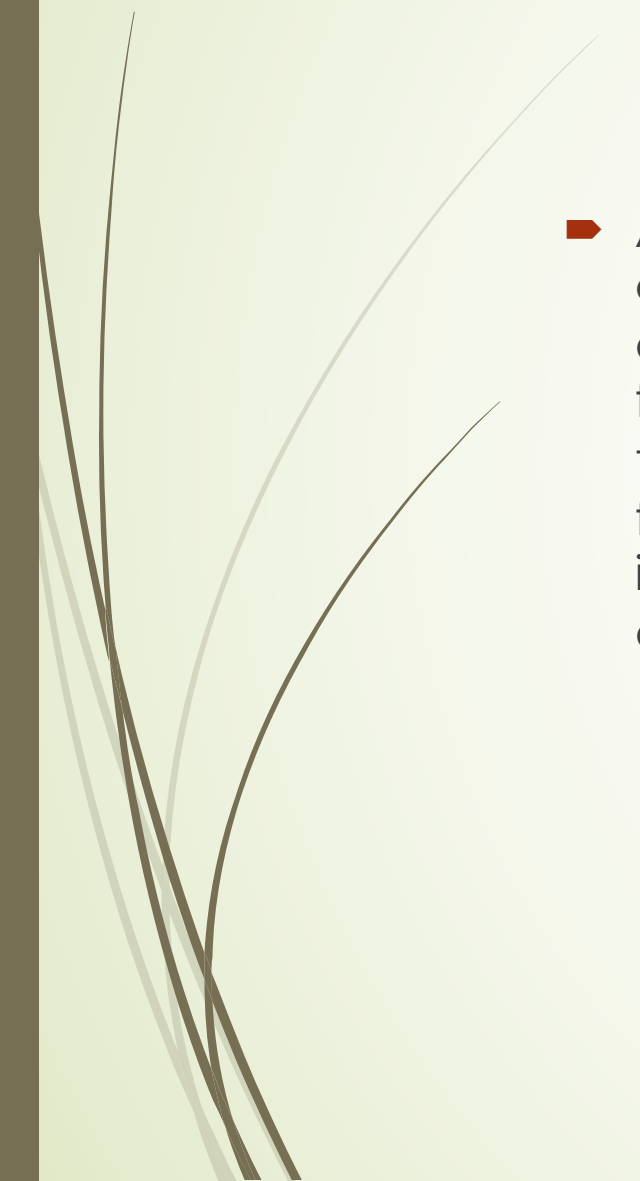
# Sentiment Analysis on Financial Statements (Free Topics Track

CS 410: Text Information Systems Final Course Project

Team member: Hui Huang – [huih3@illinois.edu](mailto: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.
- 



# Set up



- It's better to run the jupyter notebook in Google Colab so that you can take advantage of GPU.
- Unzip the finbertProsus.zip and put it in the working directory
- sentiment\_scores.csv is the saved prediction using finBERT
- All files in tean project.zip are 10Ks downloaded from EDGAR database and should be put in the working directory as well.
- loughran\_mcdonald\_master\_dic\_2016.csv is the Loughran McDonald Sentiment Word Lists and should be put in the working directory
- adj\_close.csv is adjusted closing prices for the tickers from yahoo finance and should be put in the working directory as well

# Data – 10-K filings from EDGAR database

ABT\_2021-02-19\_10K.txt — Edited

ITEM 1A. RISK FACTORS

In addition to the other information in this report, the following risk factors should be considered before deciding to invest in any of Abbott's securities. Additional risks and uncertainties not presently known to Abbott, or risks Abbott currently considers immaterial, could also affect Abbott's actual results. Abbott's business, financial condition, results of operations, or prospects could be materially adversely affected by any of these risks.

Business and Operational Risks

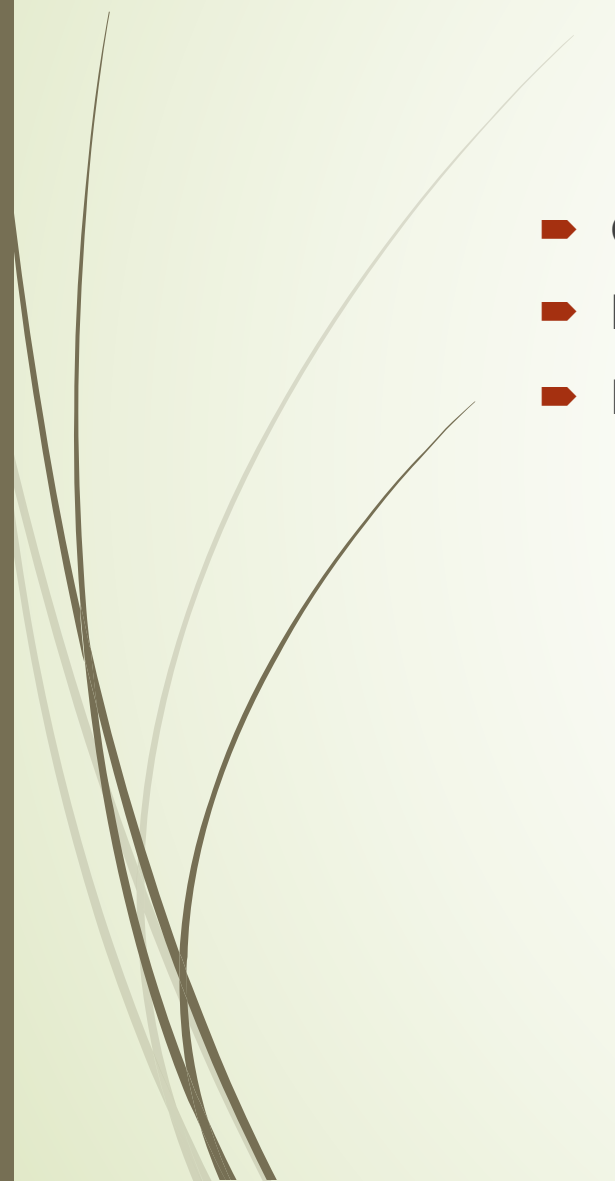
Abbott may acquire other businesses, license rights to technologies or products, form alliances, or dispose of or spin-off businesses, which could cause it to incur significant expenses and could negatively affect profitability.

Abbott may pursue acquisitions, licensing arrangements, and strategic alliances, or dispose of or spin-off some of its businesses, as part of its business strategy. Abbott may not complete these transactions in a timely manner, on a cost-effective basis, or at all, and may not realize the expected benefits. If Abbott is successful in making an acquisition, the products and technologies that are acquired may not be successful or may require significantly greater resources and investments than originally anticipated. Abbott may not be able to integrate acquisitions successfully into its existing business or transition disposed businesses efficiently, and could incur or assume significant debt and unknown or contingent liabilities. Abbott could also experience negative effects on its reported results of operations from acquisition or disposition-related charges, amortization of expenses related to intangibles and charges for impairment of long-term assets. These effects could cause a deterioration of Abbott's credit rating, result in increased borrowing costs and interest expense, and decrease liquidity.

Abbott depends on sophisticated information technology systems and maintains protected personal data, and a cyber attack or other breach affecting these information technology systems or protected data could have a material adverse effect on Abbott's results of operations.



# Preprocessing the Data

- Clean Up
  - Lemmatize
  - Remove Stopwords
- 



# Analysis on 10ks

- **Loughran McDonald Sentiment Word Lists + Bag of Words**
  - **FinBERT**
- 



# Evaluations



- Factor Returns
- Basis Points Per Day per Quantile
- Sharpe Ratio of the Alphas





# Self-Evaluation

- Manged to deliver what's planned.
- The results show that the sentiment factors extracted using both Bag of Words and finBERT don't deliver great sharpe ratio, it can be because the 10Ks are filed once a year and the market has already priced the information disclosed in the 10Ks.
- However, surprisingly, the sentiment factors extracted using Bag of Words are more consistent than those using finBERT. finBERT is a more advanced model based on transformers. It's supposed to be able to capture context better but didn't give a better result in this case.