Saucy Stocks
Kelsey Hawkins, Thomas Kudey, Michael Costanzo
Nov 14th, 2023

Datasets Used & Target Variables

1. Kelsey - created my own datasets, pulling from multiple sources such as Yahoo Finance and github repositories containing historical data pulled from YFinance in 2016-2017. Also grabbed 3 companies conference transcript (2017) from github repo. Self-downloaded shareholder reports from the last 4 quarters for top 100 companies & one year of historical stock data per company. There are no target variables as the models will be unsupervised.
   a. https://github.com/kkhawk20/EarningsCall_Dataset
   b. https://github.com/kkhawk20/MAEC-A-Multimodal-Aligned-Earnings-Conference-Call-Dataset-for-Financial-Risk-Prediction
   c. https://drive.google.com/drive/folders/1mHmm85q6tzB-V0usK4tMwQmzMnPX1p6P?usp=sharing
   d. https://github.com/kkhawk20/MGSC310-Final-Project
2. Thomas - The following github repo is comprised of two scripts and two generated csv's. Each script behaves in a similar process. First is web scraping tickers, and then hitting Yahoo Finance's API . Next it stores the data as a dataframe. Lastly it generates a csv with the dataframe. The first csv generated has the closing stock market price for the top 25 S&P500 companies each day for the last 5 years. The second csv has the date of the quarterly reports for the same companies in the same time interval. Target variable is the stock price.
   a. https://github.com/tkcpsc/Quarterly_Market_Analytics
3. Michael -The first dataset tracks every NBA team's record and win percentage by year since the founding of the League. The second is a record of fan attendance, split by home and away games. If the team is winning games and fans are interested in going to games, does that have any effect on the stock prices of the sponsors?
   a. https://data.world/gmoney/nba-team-records-by-year
   b. https://data.world/gmoney/nba-team-annual-attendance
   c. https://help.yahoo.com/kb/SLN2311.html

Business Question/s & Methods

(Michael) Does sponsoring a stadium correlate with performance of the team / stadium its attributed to? Model: Taking previous sponsored data and team records, fan base size, city, etc and running a regression on stock data to see if there is a relationship between these.

I found datasets relating to NBA team performance and annual fan performance. I plan to cross reference this with datasets on the stocks of companies that sponsor the teams directly or the stadium that they play in. I wonder If you can predict a sports teams performance, can you also predict their sponsors stock price? The question a business might ask is "should we sponsor a team/stadium?" The NBA isn't the only sport that people care about but It is the one that I know the best and I think there is enough data there for a starting point.

(Thomas)How does the timing of release for a quarterly stakeholder report, and its contents, affect a company's stock price before and after? Model: Use stock prices, stakeholder reports, the terminology reports (maybe use embeddings?), and mentioning of different subjects or profitability?

I created my own data set by hitting Yahoo Finance's api and generating time series data for the goal of predicting the effect of the the quarterly report on the target stock price. Im going to use a neural network for this application. A Neural Net will be able to detect non linear changes and patterns due to it being a time series set. Neural networks detects patterns better than sorts of regressions and trees.

(Kelsey) Does the content of stakeholder reports, specifically mentioning being affected by environmental issues, affect stock prices? Model: Use stock prices, stakeholder reports, the terminology of those reports, the location of these companies and their headquarters.

Using a **Language Model (BERT)**, I will train the model on the text data gathered from the companies quarterly press reports for 2023, or as currently released as of November 20th, 2023. For the stock data, I will build a Random forest, XGBoost, LSTM(Long Short-Term Memory) network, and a basic neural network on the stock data. My goal is to create these models and compare the years 2023 and 2016 to see how the outcomes differ and the economy changed through the years from the LM model to these models.