Predicting Market Values of Privately Held Companies

- For this project data would be collected from:

    A. **SEC Filings using their free EDGAR (Electronic Data Gathering, Analysis, and Retrieval system) database.** This should be more than enough data, as filings go back to Q3 1994.

- My model would be a **web scraper** and would use a **RandomForestClassifier.**

- Stretch goal would be to **predict market values in $** instead of just classifying companies based on whether or not they are likely to be large-, mid-, or small-cap.

- My observations will include **filings of all publicly traded companies** over a select amount of time (15 years maybe), and the target will be whether the company will fall into a category (**large-, mid-, or small-cap**) .

Additional Notes

- Understanding how to use the EDGAR database could be challenging - some learning will be involved
- Tons of information to collect - not sure if this is feasible

Tracking COVID-19/News/Political Disinformation through Social Media

- For this project data would be collected from:

  A. **Twitter using their free API** (or if I am going the covid route, there is already a covid-19 stream developed by twitter for developers - the problem with this is it is giving upwards of ten million tweets per day)

  B. **Politifact** allows access to a database of fake and real news stories.

- My model would be a **daily scraper** and would use a **PassiveAggressiveClassifier** as this allows us to take in a lot of data and discards them right away after use, without storing them in the memory.

- Stretch goal would be to include a **real time email/txt update** when a fake news story/ tweet comes out.

- My daily observations will include **each separate tweet/article,** and the target will be binary (**whether or not the story is true**).

Additional Notes

- If going the covid route, there is already a covid-19 stream developed by twitter for developers - the problem with this is it is giving upwards of ten million tweets per day.
- Working with a PassiveAggressiveClassifier is new - not sure where to start
- Also would not know how to send out updates through python for my stretch goal

Predicting Stock Prices

- For this project data would be collected from:
  A. **Yahoo Finance**
  B. There are many other APIs that could be used, however the majority of these are paid services.
- My model would be a **daily scraper** and would use a **Regression model** to predict the stock price.
- Stretch goal would be to include a **daily email/txt update** with stock predictions.
- My daily observations will include **daily/historical stock prices,** and the target will be a stocks **predicted value**.

Additional Notes

- Major problem is that stock prices are usually not predictable
- Also would not know how to send out updates through Python for my stretch goal

THANK YOU