Project Plan

Approach

Our team has been using an Agile methodology. Our project and sprint management is being done on Jira. https://thebuffsofwallstreet.atlassian.net/jira/software/projects/CAP/boards/1

GANTT Chart:

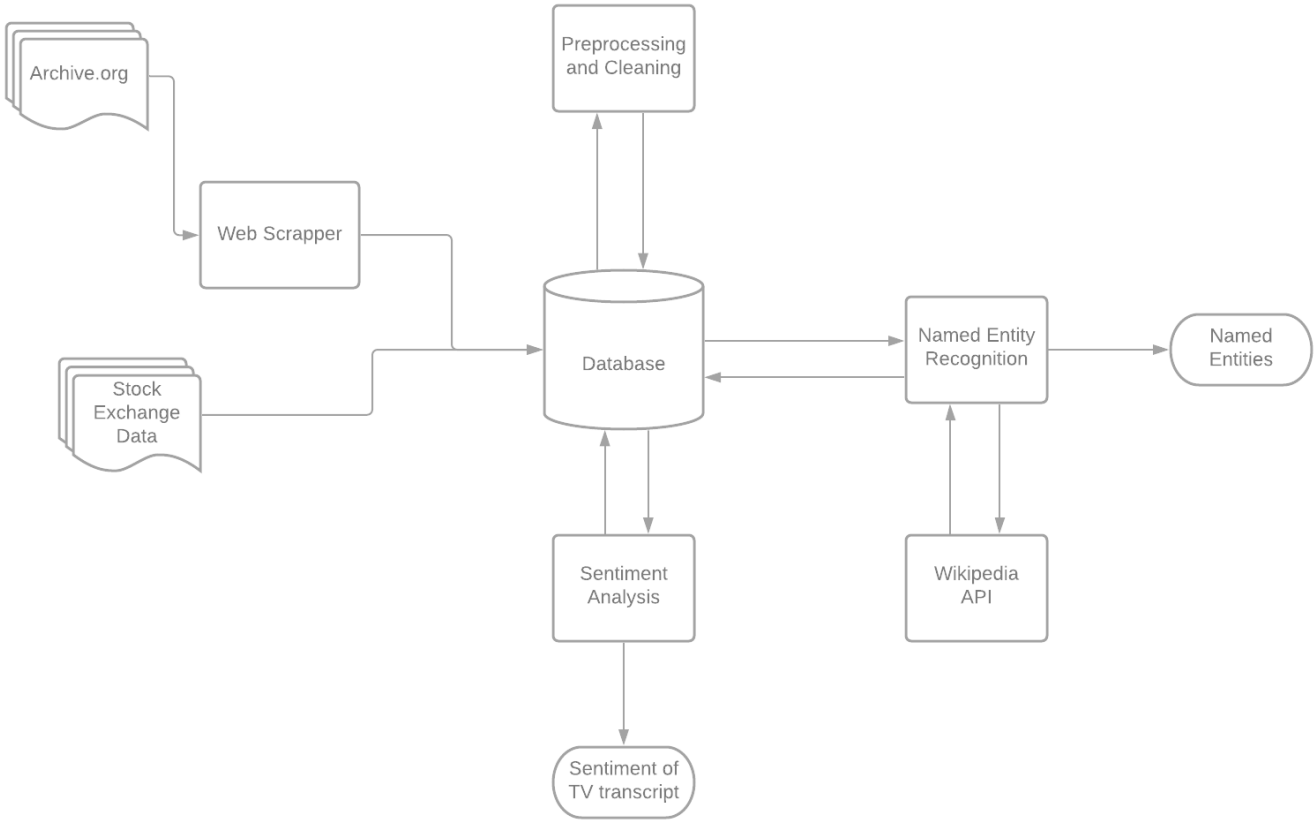
https://thebuffsofwallstreet.atlassian.net/jira/software/projects/CAP/boards/1/roadmap

WBS:

Features

Stock Price Sentiment Model

- Create data system to efficiently perform functions on the Summit high performance compute cluster, taking advantage of many CPUs and GPUs.
- Create design documents to formalize our data architecture and support future maintainers.
- Scrape tv transcript data from the web and store in a database.
- Clean and process data.
- Named Entity Recognizer. Identify snippets relevant to a given company.
- Create a sentiment classifier. Consider other dimensions like ‘authority’ or ‘subjectivity’.
- Measure intensive vs extensive properties (sentiment vs frequency). Model should be robust to changes in show schedules that might affect aggregate sentiment.
- Use methods for insights in stock trading through automated process.
- Mature database and data processing pipeline to maintain up-to-date metrics.
- Get database and code running on CU compute cluster.



Moody Manual Optical Character Recognition

- Apply YOLO computer vision model to parse words from images of Moody's manual pages.

Twitter Analytics

- Obtain twitter data form api.
- Match tweets to news segments in tv. Use similarity measures or sentence encodings.
- Investigate Twitter data for correlation between Trump's Twitter and Fox News.

Campaign contribution analytics

- Obtain campaign contribution data.
- Analyze data.
- Investigate campaign contributions to determine a company's political leanings.
- Perform community detection algorithms (like big CLAM) to identify types of contributors.

Prioritization of Features

1. Stock Price Sentiment Model
2. Twitter Analytics
3. Moody Manual Optical Character Recognition
4. Campaign contribution analytics