hotOrNot: Tracking company attractiveness and profitability from news streams

Euan O'Neill - 2312420o

December 18, 2020

1 Status report

1.1 Proposal

1.1.1 Motivation

Thousands of articles and news stories are published every day on many different companies putting them in both a positive and negative light. A system that could garner the current sentiment and attractiveness of a company through analysing these many articles and new stories could help a stockbroker or financial trader identify what companies are a worthwhile investment and which ones are not

1.1.2 Aims

The aim of this project is to develop a system which will collect news articles and stories from various sources, extract what companies are being discussed and the sentiment towards them. The sentiment will be used to measure the attractiveness of a company and predict its future attractiveness and stock price. Additionally this system will have a web app which allows it user to view collected information of different companies. The effectiveness of the project will be evaluated by comparing the predicted attractiveness and price to the actual attractiveness and price.

1.2 Progress

- News crawler setup and deployed on university open shift that collects articles from various news sources
- Basic entity identification working on both articles and sentences
- Sentence extraction performed from collected articles
- Sentiment analysis on individual sentences
- Sentences turned into data points and used for predicting stock prices
- Backend api build along with time series and postgres databases to store information
- Django front end build with homepage and company page
- All components integrated and working together

1.3 Problems and risks

1.3.1 Problems

- Many different components to system meant alot of time was committed to researching and learning new plugins / libraries.
- The initial build of the backend components could cause too much strain on the database therefore they had to be rebuilt using a more sophisticated architecture.

1.3.2 Risks

- Many different components to project **Mitigation:** Continue to separate concerns by having different components act as micro services to main system.
- Predictions are not particularly accurate. **Mitigation:** Continue improvements of overall system and measure accuracy improvements to created metrics.
- Time issues in how long it takes for the system to process articles and sentences.

1.4 Plan

1.4.1 Semester 2

- Week 1: Improve Search System
  Better search system for finding related companies.
- Week 2-5: Improve Prediction and evaluate system
  Using prediction metrics incrementally improve overall system in order to improve accuracy of stock predictions **Deliverable:** Measure of accuracy improvements from different improvements to the system
- Week 6-7: Stretch Goals
  Try to achieve any additional stretch goals while still improving prediction along with ironing out any bugs with the system
- week 8 Final Implementation:
  Iron out any remaining issues with the project. **Deliverable:** Polished working version of the project ready
- week 9-10: Focus on dissertation
  Deliverable: First whole draft of dissertation submitted to supervisor