Practicum Proposal:

Harmon Hawkins (434-409-4498, [harmon.hawkins@gmail.com](mailto:harmon.hawkins@gmail.com))

Title:

Sentiment Analysis of Airline News Posts, and How That Relates to Stock Price and General Airline Feeling

Description:

While working as a developer at American Airlines I have become familiar with different airlines and their reputations in the media. During a meeting my team reflected on how the technology and processes that we were implementing compared to other airlines. While there was no real way of knowing we do know that certain airlines have a reputation with the media for being tech leaders and ahead of the curve compared to other airlines. Does this type of reputation affect the general perception of different airlines? In articles that specifically highlight an airline’s technology is there differing sentiment in the article or in the comments about certain airlines? Depending on time this could be expanded to looking at stock price of different airlines after an article is published and seeing how sentiment from the article or the response influences stock price, if any.

Data:

Much of the data that I plan to use will be scraped from the web. If available and time permitting, data and articles going back to the start of 2019 will be collected. This can be adjusted as needed. I would like to limit the number of sites that I scrape to around 3 to 4. I would primarily like to look at 2-3 popular airline blogs and one major news source, but this could change depending on how hard it is to scrape each site. I would like to get at least 10-15 articles and their comments from each source, and ideally more than that. I would also primarily like to focus on articles that revolve around technology or new processes at the airline.

Analysis:

The primary analysis of this project after the web scraping will be using text mining to do sentiment analysis on the titles of the articles, the actual article, and the comment sections of that article if available. The sentiment analysis aims to look at how different airlines are written about in different new sources and what the general sentiment is about the article and the airline is around the time of publication. Once this data has been collected, it could be combined with stock prices and financial numbers to see if there is a financial impact to the articles.

Anticipated Difficulties:

I think some of the web scraping could be difficult and take some time to develop especially when scraping different websites. To attempt to combat this I wanted to build some extra time into my proposed timeline to allow for any additional time that may be needed. I could also cut down on the number of different websites data is collected from or the number of airlines that are being looking at.

Timeline:

Week:

1. Proposal
2. Web Scraping
   1. Pick sites that will be scraped and airlines to look at.
   2. Start writing code to scrape different websites.
3. Web Scraping
   1. Finish/Adjust any code as needed
   2. Finish collecting data
4. Data Cleaning
   1. If needed continue collection of data
   2. Clean data, break down article body or comments as necessary
5. Analysis
   1. Start analyzing data and exploring any interesting features
6. Analysis/Visualization
   1. Finish analyzing data
   2. Start creating visualizations/dashboards that highlight interesting findings
7. Visualization
   1. Finish any lose ends and wrap up project
   2. If there is extra time, do any desired additional analysis
8. Presentation
   1. Present findings