William Hiatt
Deven Biehler
Gabriel Sams
Sprint 3 Report
Sprint Dates: February 20th - March 24th

Sprint 2 Retrospection: The majority of the tasks during sprint 2 involved exploring the tools and code necessary to build a working model prototype. Sprint 1 was mostly planning and researching while sprint 2 involved using all that research to start creating a model. The team completed all administration work such as the project report solution approach, sprint 2 report, fix/update project report, and the project presentation video. In addition the team completed tasks such as research, explore, and implement SpaCy, get a data set, fix/troubleshoot web scraping tool, explore and implement Textblob, explore Vader, and explore/implement Gensim.

What went right: The sprint planning was very good. We had plenty of time to complete the tasks and also had a little flexibility to add new tasks that came up. Communication between the team throughout the sprint was seamless. The team met with the sponsor almost every week of the spring to go over the project and get valuable insight from the sponsor. The sprint had very few hiccups and any issues that were presented the team was able to quickly fix.

Lessons Learned/Improvement: There were a couple improvements that we implemented and learned from the last sprint. First the scale of the project. The team was able to get a better idea of the scale of the project compared to the last sprint. Once the team had a good idea of the scale we were able to break the project up to start implementing some of the tools we had been researching. Previously it was hard to figure out where to start because we didn't have a very solid grasp on the scale of the project. Another improvement compared to the last spring report was reaching out to the sponsor for feedback on the administrative work. This allowed the team to make sure that they were articulating the goal of the project accurately. It was also good as it allowed the team and sponsor to check that everyone has the same end goal for the project. The team ran into several bugs and issues in the code when exploring each of our respective tools/functions that we had to deal with during Sprint 2. These were solved by troubleshooting and investing time in Sprint 3 towards making the tools work well together.

Sprint-3 Planning: Sprint 3 was focused on preparing a final alpha prototype model to build upon in the 2nd semester of this project, as well as to present to the sponsor at our final presentation meeting. We made this the primary goal of this sprint because we needed a presentable model by the end of the semester, and since this is the last sprint of the semester we needed a final point to stop at that would be easy to reenter in the next semester. The major tasks involved in developing the prototype model was to implement the exploration performed in Sprint 2 on sentiment analysis, web scraping and article pre-processing, and topic modeling of articles. Once they were implemented separately they were combined into a single prototype model that performs all functions on the single articles fed in from our input list. We also decided to revise our project report for the end of the semester as well as completing any necessary

documentation for the project before the semester ends to have a solid understanding of our previous work next semester.

Our specific task planning for Sprint 3 were the following tasks: fix web scraping bug, fine tune web scraping, add subjectivity score for every article, fine tune sentiment scale, filter out social media, edit Project Report, implement SpaCyTextblob into the working alpha prototype, research viable tools and libraries for web scraping and sentiment analysis, write final project report and prepare sponsor presentation, run Topic Modeling on Scraped Data, and research ways of extracting important topics from data.

Sprint Task Assignments:

William: Fix webscraping bug, fine tune webscrapping, add subjectivity score, fine tune sentiment scale, filter out social media, edit/write final project report, prepare sponsor presentation.

Gabriel: Edit Project Report, Implement SpaCyTextblob into the working alpha prototype, Research viable tools and libraries for web scraping and sentiment analysis, Write final project report and prepare sponsor presentation.

Deven: Run Topic Modeling on Scraped Data, Write Final Project Report and Prepare Sponsor Presentation, and Research ways of extracting important topics from data

Sprint Task Details
William Hiatt Sprint Task Details
Task: Fix webscraping bug
Implementation: Previously our model was only scraping the first URL within our attached .CSV file. After looking into it the issue was resolved and now the model will scrape every URL that is within the .CSV file.

Task: Fine tune webscrapping
Implementation: The model was having issues collecting a lot of garbage data (such as advertisements and "related stories" titles. After some research and experimenting we switched from BeautifulSoup to Newspaper3k. Now webscrapped data is strictly from the body of the article and excludes all the garbage that was scrapped previously.

Task: Add subjectivity score
Implementation: This was an easy implementation as Textblob already has this feature in addition to it's sentiment analysis. Implementing it was as simple as calling the method when the sentiment method was called and outputting it to the table.

Task: Fine tune sentiment scale

Implementation: Previously the articles were ranked with a positive or negative sentiment. The scale was adjusted to include additional parameters such as neutral, neutral negative, neutral positive, positive, and negative. In addition these scores are based off the average of all the scores instead of it being based off of 0.

Task: Filter out social media
Implementation: The dataset contains some links to social media sites such as Twitter, Facebook, and YouTube. These have been filtered out so they receive a "Social Media" label instead of receiving a sentiment score.

Task: Write/edit final project report
Implementation: Wrote the staging and deployment plan section as well as part of the alpha prototype section. Also helped clean up any errors in the report and expanded on why we used some of the tools we chose.

Task: Prepare sponsor presentation
Implementation: Worked witht he team to create our sponsor presentation. This was used to demo the model to the sponsor.

Gabriel Sams Sprint Task Details
Task: Edit Project Report
Implementation: According to the client's recommendations, I revised the Project Report from Introduction through Solution Approach. Also included more citations and references to previous/related works.

Task: Implement SpaCyTextBlob as Sentiment Analysis tool in prototype model
Implementation: I took the exploration from Sprint 2 and applied it to the working prototype model to gather per-article sentiment analysis, per-named entity sentiment analysis, and the functions to use for per-topic sentiment analysis. Per-topic sentiment analysis is incomplete, but only needs to be applied to the working topic analysis code to function. Estimated completion is by the end of semester or very start of the 2nd semester.

Task: Research alternative tools/functions for web scraping and sentiment analysis.
Implementation: I performed research and reported to my team on new viable tools that would make the web scraping of articles easier. I also researched the SpaCyTextblob library to find the best alternatives of the default sentiment analyzer that was being used.

Task: Write Final Project Report and Prepare Sponsor Presentation
Implementation: I wrote the test plan section and the future work section of the project report as well as contributed to the alpha prototype demonstration to finalize the project report for this semester. I also helped the team write the script and prepare slides for our final sponsor presentation.

Deven Biehler Sprint Task Details

Task: Run Topic Modeling on Scraped Data
Implementation: For every input article, Using an LDA model from Gensim topic modeling package, extracted each topic and produced a number of words in order from greatest to least relevant in a topic. Using the coherence score the LDA, fine tuned the number of topics generated out of a given article.

Task: Write Final Project Report and Prepare Sponsor Presentation
Implementation: I wrote the Functional Requirements, Non Functional Requirements, Literature and Existing solutions slides, and the Gensim Demo section. I also assisted with the alpha prototype description and demonstration in the final project report for the semester.

Task: Research ways of extracting important topics from data
Implementation: I explored different ways to try to get a coherent topic from the data. This involved extracting nouns, removing different types of stop words, and tuning the dataset as best as I could. Through my exploration, I used PyLDAvis LDA visualizer to visualize the coherence of the topics.