Project Status Report 5

Team 3

NewsVerse

Software Engineering
Fall 24

CS 673 A1

Damanjit Singh
Abigail Gualda
Stephen Yang
Dewei Wang
Yuhang Wang

Important Decisions and Team Weekly SCRUM Meeting Notes Captured here:

Important Team Documents:

- Slack Link: https://cs673a1.slack.com/archives/C07L83B3249
- Google Folder: 📁 CS673 Team 3: NewsVerse WebApp
- Jira Link:
  https://bu-team3-swe673.atlassian.net/jira/software/projects/SCRUM/boards/1/backlog
- Figma Link:
  https://www.figma.com/design/EStFv0CsKg7HSsz05Vvkbn/News-Verse?node-id=2-4705&node-type=frame&t=R5pJdfTDQRUUayIW-0
- Weekly Team Meeting:
  - Wednesday 6:00 - 7:45pm
  - Zoom Link:
    https://bostonu.zoom.us/j/92986961141?pwd=2PsIEJaSs4AAVc7mfAXL3VkEbRNj15.1&from=addon

SCRUM Mtg Notes:

Attendance: Stephen, Abby, Singh (sick) , Dewei , and Yuhang
Date: Nov 6, 2024
Meeting: Weekly Work Review and Turn in Assignments Mtg
Agenda:
- Talk about bugs that have risen since we last met
  - Challenge 1: There are too many resources being pulled from out data – such as there is one title that has over 100 articles attached to it
    - Solution: we are only going to display the news source and its calculated biases
  - Challenge 2: The category page needs to be reviewed because of the lack of keywords
    - Dewei made a version with keywords and a keyword search and he would like for the rest of the team to review it
    - Solution: Data modification that will allow us to have a way of separating the top news source keywords in comparison to the top news source keywords
  - Challenge 3: By now our JSON data contains too much information that we do NOT need – all we need is the base source and title of the article – everything else is irrelevant to what we need
- End Last Sprint
- Begin New Sprint

○ Assign tasks to teammates

Discussions:
- Review the new and improved landing page
  ○ Have reviewed it and would like to do another version to finalize it in terms of font, color, and size.
- Start to prepare the unit testing *****
  ○ This will be covered when we meet in person

Individual Contributions:
- Abigail Gualda
  This week, I contributed to the project organization by managing task assignments on the Jira board, distributing responsibilities among team members, and outlining our final two to three sprints. I also completed the collection and organization of news source logos to facilitate their display on the frontend interface of the website. Additionally, I developed a detailed explanation for users that clarifies the bias metric presented alongside each article, aiming to enhance transparency and user understanding. Furthermore, I finalized the logo design for NewsVerse. My upcoming task involves designing logos for each news category, which will be instrumental in finalizing the landing page's design.

- Damanjit Singh
  This week, I focused on enhancing our NewsVerse platform by testing content retrieval and summarization for link previews, implementing hyperlinks on news source logos to provide direct access to original articles, and developing filter functionality with keyword recognition for the search bar to improve search precision and user experience.

- Stephen Yang
  This week, I finished the final steps of AI semantics development. This functionality is integrated into the back-end pipelines. Once the news is stored in the tables, it will automatically trigger ChatGPT API to execute AI semantics, which returns the groups of news, the topic and summary of the group, and the media bias scoring of each news. To ensure the interface would display the exact content from the back end and be free from bugs, I coordinated with Dewei regarding the data schema. Apart from that, I go through the news sources stored

in the database to prepare a list of news names for Abigail to collect news logos.

- Dewei Wang
  This week, I continued to focus on front-end web design. I changed some of the display effects on the web page and added some dynamic effects to the logo. I updated several minor grammar errors and tried to connect the website's front-end and database. I made some adjustments to the data format to make the web page better read and display content. I also added some keyword searches. The remaining work includes fine-tuning of some details and data preprocessing. Overall, the front-end has been basically completed.

- Yuhang Wang
  This week, I focused on creating our own LLM to summarize news content for our website. I decided to choose an already pretrained text model and fine-tuned it to better suit our news summarization task. Since we want our website to deliver unbiased summaries, we need the model to remain as objective as possible in its text summaries. I found a public dataset, processed the data, and fine-tuned the model accordingly. Overall, the main structure of the model is complete. The remaining work is to integrate it with our configuration files and compare its summarization quality against ChatGPT. In the future, we might also quantize the model to ensure faster and more efficient inference for summarizing the content we aim to deliver.

GitHub Repo Link:
- https://github.com/stephenyang0215/NewsVerse