Bot || !

Team Members: Patrick Canny, Lane Gramling, Liam Ormiston, Damian Vu, Taylor Walenczyk

Project Synopsis (1-25 words)

"Bot || !" leverages existing and user-generated data to classify pathogenic social media accounts. It's applications include identification and removal of spam accounts that influence public opinion.

Project Description (150-250 words)

Our team decided to take on this project because the problems it wrestles with are imminent and far-reaching. The widespread use of pathogenic social media accounts has a significant impact on the news people see every day; furthermore, these accounts can drastically impact public opinion on important global events. Our project seeks to push back against this trend by providing a tool for identifying and related pathogenic accounts.

Additionally, the complexity of this project affords us the opportunity to develop our skills across the tech stack. This project involves building out a slick user-experience, robust data storage systems, and a novel analytics engine. Accordingly, we will integrate several codebases, pushing us to improve our DevOps, project management, and software engineering skills. At the end of the day, we expect to have a fully functioning application that allows users to vote on whether or not they believe an account is a spam account and sends that data to a classification tool that can judge whether or not an account is pathogenic with a reasonable level of accuracy. If we implement the core functionality highlighted above ahead of schedule, we will add more features.

Project Milestones

Our team has identified several key milestones in the creation of our product. The first will be creating a high-level project timeline that we can use to plan our work around. We will have this high-level workflow decided before the start of November. We will also aim to have UML documents, including some preliminary wireframes, created for all project components by the start of November as well. By the end of the first semester, we hope to have a very simple dataflow set up for our project. This would include a minimally viable mobile/web application with user registration, as well as an established database of account data.

When we move into the implementation phase, the milestones will be broken down on a monthly basis. In February, the application should be able to register and store users in a database, along with any necessary information about those users. We will also have a functioning build system set up, in order to ensure that our product is tested and functional with each change. By the end of March, our application frontend will be more refined and fully featured. Our Internal API will be ready to transfer data across each piece of the application to the backend toolkit, which will process that data and continually classify accounts. Finally, by the end of April, we will have deployed an application that is robust and able to effectively classify newly created accounts as pathogenic or not.

Project Budget

Item	Price	Quantity	Timeline	Description	Purpose
Dell Optiplex 790 Refurbished high-performance MiniTower	$237.89	5	Needed near the beginning of development to aggregate data	Important specs: 8GB RAM, 2.220GB storage, 3.4Ghz Intel process Full description	Hardware for a small HDFS cluster for data storage and processing
Total	**$1189.45**				

Work Plan

Liam - Application developer, front-end engineering

Damian - Deployment/build engineer, data visualization, internal API developer

Lane - Backend engineering, internal API developer, data engineering

Pat - Data engineering, data modeling, application development

Taylor - Backend engineering, data modeling, Hadoop developer

References

"Early Identification of Pathogenic Social Media Accounts" (Alvari, Shaabani, Shakarian)