Enhancing Social Media Governance with Policing Bots - Milestone 6 Evaluation

—

Authors:

- Cody Manning - cmanning2020@my.fit.edu
- Gabriel Silva - jsilva2021@my.fit.edu
- Liam Dumbell - ldumbell2021@my.fit.edu
- Nickolas Falco - nfalco2020@my.fit.edu

Client:

- Dr. Khaled Slhoub - kslhoub@fit.edu

Progress of current milestone (Task Matrix):

Task	Cody	Gabriel	Liam	Falco	To Do
Finalize the detection algorithms	15%	10%	50%	25%	N/A
Create the decide module	10%	40%	30%	20%	N/A
Finalize the maliciousness algorithms	5%	40%	20%	5%	N/A
Test the framework as a whole	40%	40%	10%	10%	N/A
Create developer / User Manual	70%	10%	20%	0%	N/A
Final Demo	10%	50%	15%	25%	N/A

Discussion for each accomplished task for the milestone:

- Task 1: The algorithms used to detect bots have been altered to increase the accuracy, which is a result of dialing back the weight of the first algorithm used,

and increasing the weight of the second algorithm used. We were able to make the results of testing come back with less false NEGATIVES this way. Interestingly, this did not change the rate of false positives. We are pretty happy with the detection, but if we had more time we would have continued adding additional methods for it.

- Task 2: The decide module was the easiest thing we added, it was the final ‘D’ of the “three D’s” we decided on at the beginning of the project. It requires a human decision though, as Reddit does not have an automated report system (that we have the ability to access). The module essentially reads the input from the first two sections. If a user is deemed to be a bot, and THEN is deemed to be malicious, we give the user a link to Reddit’s reporting page, with a note to review the information manually.
- Task 3: For the maliciousness, we toned down the numbers for when the maliciousness is flagged. For example, the amount of profanity used needed to throw a warning for maliciousness. We feel like this is something that would need to be tweaked on a subreddit-by-subreddit basis. There are other tweaks we would like to make, like searching for repetition on messages across multiple subreddits.
- Task 4: This was the big focus for the final milestone. We have been working on making sure that the data we collect and analyze is accurate. The methodology we used to test for false positives (human beings that we incorrectly deem to be bots) and false negatives (bots that we incorrectly deem to be humans) was testing our tool against a list of known humans and a list of known bots. We took a percentage of the incorrect amount of data and have been using it to tweak our algorithms. The percentage of false positives is about 15% and the percentage of false negatives is about 20%. We have had a list of known bots for a while, but we added more to it for this task. We also compiled a list of known humans. This was a much more complicated task than the known bots. The methodology for that was scraping the moderator list for as many of the biggest subreddits as possible. On Reddit, a moderator that is a bot is almost always stated as such in the name (SpamBOT, for example). These two lists were the backbone of our testing.
- Task 5: A user/developer manual was created for the project. We focus on the developer side mostly, as this framework is mostly focused on people who will be using it from this standpoint. This is the primary reason why there isn’t a GUI, as the standard Reddit user really shouldn’t be using this tool unless they are curious about the validity of another user they are talking to. This tool would mostly be used by a moderation team on Reddit, possibly as an attachment for their specific subreddit.

- Task 6: We have created a demo video that goes over all of the aspects of the project, showcasing the multitude of modules and features the project is capable of.

Discussion of contribution for the current milestone:

- Cody Manning: Cody was focused on testing primarily. He compiled the lists of known bots and known humans to test the accuracy of the program. Running the program against the user lists found a small handful of bugs that he fixed. Cody also worked on efficiency and improving the in-place algorithms. Cody spent an extensive amount of time creating the user/developer manual. Cody also wrote this document.
- Gabriel Silva: Gabriel was responsible for fixing up our maliciousness algorithms. A bunch of bugs came up in testing that Gabriel was in charge of fixing. He also helped a lot in the testing process and was the person primarily responsible for creating the final demo. He also brainstormed the application of the deciding module.
- Liam Dumbell: Liam was responsible for working on improving the detection algorithms based on the result of testing accuracy. Liam also helped finalize the database systems we used, to properly condense our data into a form that was easily readable and useful for the sake of the project. Liam also helped with the creation of the user/developer manual and the Final demo video.
- Nickolas Falco: Falco's primary focus was code readability/refactoring. He spent a lot of time cleaning up the GitHub repository and working on general professionalism for the future development of the project. Falco had a hand in the creation of the decide module, the increase in accuracy of the detection module, and the final demo.

Lessons Learned:

This project has been quite an ordeal, and we aren't truly finished working on it, but for the sake of completeness, these are a few things that the team has learned in the creation of our framework:

- **Have a clear plan, and stick to it**: Planning is like 70% of software development. Make sure that everyone on the team is on the same page, and that if a new feature/module is desired everyone should understand and agree on the changes. Doing this extensive planning may seem like an arduous task, but it ensures that everyone is on the same page and work doesn't get delayed. We had a misstep in this section, as originally the framework was supposed to be run on Twitter. We didn't plan for the heavy restrictions the Twitter API had though,

and we were forced to shift gears to Reddit to complete the project. If we had researched the API's available, we may have been able to avoid this delay.

- **Start early**: Software development is a lengthy process, involving a ton of different ideas and moving pieces working in tandem. Make sure you start early, not just to catch bugs but to also plan.
- **Understand API Limits**: Reddit, like many other APIs, has rate limits to prevent abuse. It's crucial to understand these limits and design your application to handle them gracefully. We ran into quite a few issues with daily throttling, which at times would restrict our ability to make substantial changes to the framework. This kind of thing is also important for efficiency's sake. There is only so much you can improve when you have to rely on an API to receive data. We spent a bunch of time trying to improve runtime on a system that was not limited by anything we had control over. Also important is proper research on what you can and can't access with an API. There were several pieces of information that we would substantially benefit from having, but were unavailable to us with the information we had from the API.
- **Test early**: This is something that we were lacking. Make sure you have a system in place to ensure the accuracy of your results and data earlier rather than later. This didn't end up hurting us too badly, as we got lucky and had good results even after testing late; but this is something that could have been a complete disaster if our accuracy was bad the entire time we were doing the project.

Dates of meeting with Client:

- 4/12/2024

Dates of meeting with Faculty Advisor:

- 4/12/2024

Evaluation by Faculty Advisor:

- Faculty Advisor: detach and return this page to Dr. Chan (HC 214) or email the scores to pkc@cs.fit.edu
- Score (0-10) for each member: circle a score (or circle two adjacent scores for .25 or write down a real number between 0 and 10)

Cody Manning	0	1	2	3	4	5	5.5	6	6.5	7	7.5	8	8.5	9	9.5	10
Gabriel Silva	0	1	2	3	4	5	5.5	6	6.5	7	7.5	8	8.5	9	9.5	10
Liam Dumbell	0	1	2	3	4	5	5.5	6	6.5	7	7.5	8	8.5	9	9.5	10
Nickolas Falco	0	1	2	3	4	5	5.5	6	6.5	7	7.5	8	8.5	9	9.5	10

Faculty Advisor Signature _____ Date: _____