**Adrian Butiu**

**Andrew Dallarte**

**Ricardo Gayagoy**

**Tell something about your Scraping experiences.**

In our experience, we have experimented with the use of both API and Selenium Web Scraping. Our first attempts to our api went smoothly, but for Selenium, we have to figure out first what webpage security we have to deal with, as we are aware of anti-scraping methods of well-known sites. Thankfully, there is minimal protection towards extracting texts of subreddits, which expanded our time to scrape as early as possible.

**What are your pain points and how did you resolve it?**

During the development of our scraping tool, we first inspected Reddit’s HTML structure and how it handles our target data. What we saw was a shadow DOM that may hinder us from creating our script, as we need to know how to bypass and retrieve the post content we need. Thankfully, we have found another div class that’s above the shadow DOM and contains the post content of the Reddit post. We now utilize Selenium for our scraping tool as it is what we are most familiar with, and let it scroll on several Reddit posts in the gaming subreddits to get the score (upvotes/downvotes) and the post content.

**What techniques have you used?**

We just used normal selenium, which scrapes the content we want by finding the parent div that contains all the elements we need. We also used a model called RoBERTa that gives us a dependent variable or the sentiments; however, we are still unsure how to use the variable as we lack experience and knowledge to do so for now. We are planning to learn more through consultation.

**What are the lessons learned?**

Based on what we did, we learned how to scrape using Selenium even if the elements are hidden by the shadow DOM. We also learned how to utilize the graphics card as we used Cuba for the reddit api. We also learned how to gain sentiments using Roberta

**If you changed your topic, what is the revised approved topic? Tell something about the revised topic?**

The revised version of our topic shifted from sentiment analysis to classification analysis of different machine learning models. Instead of focusing only on identifying sentiments, the new topic explores how various models perform when applied to the same dataset. It allows us to compare their predictions, measure their accuracy, and analyze the differences in the results they produce. This change provides us with deeper insights into model behavior, helping us understand which model is most suitable for a specific task.