

Week 4 Meeting Notes

1. Call to Order

A meeting was held at 33 Charles St East on 30 September 2022.

2. Attendees

1. Regan
2. Riley
3. Jared

3. Open Issues

1. What are our strengths, and experience with programming?
 - a. Regan: Data science, machine learning. Did a small front end web development project 2 years ago and will not be confident to develop an intricate webpage for this project. Has interned in a hedge funds where he analyzed data and made prediction models using python.
 - b. Riley: Similar to Regan, except that he has more experience with data science. Adept in cleaning data, visualizing data and dash. Similarly, has made prediction models for a hedge fund using python.
 - c. Jared: Utilized AI algorithms and cloud heavily during his previous internship. Has used Docker in the past for projects.
2. What are our academic commitments and goals?
 - a. Regan is a varsity squash player who trains 4 times a week and is keen to learn more about project management in large software companies.
 - b. Riley has joined several clubs and aims to meet new friends there. Also, he wishes to become a better software developer and be more well-versed with how projects are developed in larger software companies.
 - c. Jared wishes to explore Canada more during his time as an exchange student here at UofT and he wishes to become more familiar with the best practices of project management to develop software.
3. What is one problem which is interesting or meaningful to us?

A common hobby which all 3 of us enjoy is watching movies and videography. Riley mentioned that it was also very interesting that while many movies (such as the ones created by Marvel Studios) make tons of money, there are also many out there which are not profitable. We deliberated on a couple of reasons:

- a. Was the marketing on social media bad?
- b. Was the choice of cast/directors be the reason?
- c. Could the genre of the movie affect the gross revenue?
- d. Was the budget of the movie too low/high for the movie?

Since there could be many reasons why a movie is profitable, we wanted to narrow down to a few selected features which are most important. This is so that in the future, any new and upcoming directors who intend to create their own films or movies can perhaps make use of these selected features so as to increase the probabilities that their product will be profitable!

- 4. Is there a dataset which we will be able to download for free online?

We require a large dataset which has features of many movies over a long course of history such that this can be large enough to form our test and train sets. We later found a dataset on Kaggle, IMDB 5000 dataset which has a wide range of features and data of movies all the way back to 1920s. This will mean that it will be suitable for us to analyze and create a prediction model for the project.

- 5. What are the possible tech stacks and tool chains that are available which will achieve our desired outcomes?

Our tech stack will require us to do some form of data cleaning, analysis and visualization. This leaves us with the option of Python, Scala and R. We came to a conclusion that since our 3 of us have had the most experience with writing code in Python, and Scala while being on the most popular functional languages, actually runs on JVM which makes it ideal for working with high volume datasets, none of us had any experience with writing Scala. Furthermore, Python has some really good libraries such as plotly, seaborn and dash which are great for data visualization.

Our choice of container, we have many options, from Docker, Buildan and Podman. Just by analyzing these 3 options, we concluded that Docker will be the best for our project.

- a. Buildan and Podman only runs on Linux and there is no convenient wrapper for Windows and MacOS which Docker has.
- b. There is no Docker Compose replacement for Podman.

- c. Docker has been used by Jared in the past. This will mean that he will be able to guide both Regan and Riley on its use. More time can then be spent on developing the project instead of learning a new software.

6. Preliminary project development plan and milestones.

We discussed and broke down the milestones and development plan of our project. In Weeks 3-4, we have to streamline the goal of our project, find a reliable dataset which fulfills the requirements of this goal, complete initial data cleaning and visualization to understand the dataset we are working with. Lastly, we will have to push a toy (preliminary) application that is functional and fulfills the requirements stated by Mike. Containerize it and ensure it works on all devices.

Weeks 5-7, we have to dive deeper into creating a prediction model for the problem at hand, experimenting with different machine learning algorithms and optimizing parameters to ensure that we have found the best possible set of algorithms and parameters to achieve our desired outcomes. We need to visualize the data and results to ensure that the outcome is meaningful for users. Lastly, we need to improve the user interface such that anyone can easily input their parameters and obtain meaningful results.

Weeks 8-9, we have to undergo testing, sorting of bugs and triage any bugs. We have to look out for any patterns if they exist and ensure that the whole project runs according to our plan without any unexpected outcomes.

7. Discuss the responsibilities and immediate actions to be completed by each member to achieve the next milestone.

Our next milestone will be to implement the software. After discussion, the following will be the responsibilities and immediate actions by each team member.

- a. Riley: Clean dataset, and start an exploratory data analysis for this dataset such that Regan will be able to capitalize on this before implementing his Machine Learning algorithms.
- b. Regan: Document and ensure meeting minutes and assignment submission items are all in order while waiting for Riley to finish data cleaning and exploratory data analysis. Will implement machine learning algorithms to solve the given problem.

- c. Jared: Containerize everything and ensure that the deployment of the software is smooth and bugless.