

ALY 6040

Module 6: Individual Reflection

Professor Reilly

Group 7: Sean McLean, Catherine Smereena Dommaty, Susheel Reddy Nelvoy

The project between the three of us had some issues in the beginning when there were some miscommunications about figuring out what kind of dataset we should use for the project. I am naturally more comfortable with being more of a collaborator than a leader, so it was a bit of a challenge for me to take the initiative most of the time. That aspect of the project was frustrating, especially when the lack of communication was there, but leadership is a skill that is valuable and so I worked on improving my leadership qualities for this project. Overall, it was a learning experience in that regard so something to take with me for future coursework and in my future roles in the workforce.

For the project I chose the Netflix dataset to analyze, Catherine did the exploratory data analysis of the dataset, and for the data mining types we chose mine which was association mining and Catherine's which was clustering. We thought they were the two strongest types for the data we were working with. I wrote the first draft and final draft of the report and included Catherine's EDA (Exploratory Data Analysis) and clustering analysis to the final product. The presentation slides were supposed to be done by Susheel but he had issues completing them, so I stepped in and created them in the end. He did the classification data mining techniques, but we decided not to use them in the project because we thought the other two methods were stronger in their findings. We communicated afterwards on what slides to do and I thought they both did an impressive job with the presentation.

From the association mining techniques, I employed methods such as frequent itemset analysis, high-confidence rules, subset rule removal, and visualizations. Some of my key results were finding subgroups of rules that had high confidence values and lower frequencies that I think added some interesting insights into the dataset. I used most of my results in the report and the parts I did not were minor aspects that were not pertinent information.