



**Individual Reflection**

Trieu Vo

College of Professional Studies, Northeastern University

ALY 6040: Data Mining

Prof. Joseph M. Reilly

**My role**

My name is Trieu Vo, a member of group 4 (Spotify Track Popularity Analysis and Prediction). This is an extremely interesting project with numerous techniques that can be applied and learned. Because I have a bachelor's degree in computer science, my strengths are in programming, data analysis, and model building. In this project, I and two other members, Ashu and Abid, rotate through different roles, each contributing more or less to different aspects of the project. We discuss how to select the best dataset for the project, as well as the requirements of each assignment and the professor's requirements. We had to schedule the time and place to work on the project because we all had other classes at the same time. Every week, we plan to meet at Ashu's house on Sunday to work on the project together. My primary responsibility as a programmer is to translate the project's ideas into lines of code and models. And, from lines of code, we create visualization charts to help the story reach the audience. I arrange and divide each part of the project's code for each team member so that they have the best connection and are completed by the most skilled person. At the same time, I check other team members' code and add data pipeline ideas to make the process go more smoothly. I'm also a big fan of using Kaggle in teams so that everyone can work on and edit projects online while staying in sync. I am also in charge of creating two models, XGBoost and Random Forest, as well as fine-tuning hyperparameters to improve model accuracy. At the end of the project, I and two other team members collaborate to write the report, create the Powerpoint presentation, and use storytelling techniques to deliver the best performance to the audience.

**Team roles**

Ashu is an expert at analyzing business problems and providing strategic direction. He researched Spotify, discovered the issues it was experiencing, and devised solutions and questions to address those issues. He has worked in India for many years and has field experience in a variety of projects. His assistance was invaluable in allowing our team to concentrate on the most pressing issues. At the same time, he assisted in arranging a location for us to meet weekly and discuss new project ideas. His visualization abilities are also exceptional, and he contributed several charts to our project. He is also proficient in report writing and worked with us to complete the final report. Ashu is also a friendly person who makes everyone around him feel good.

Abid is the group's leader as well as a very funny and sociable individual. He exudes confidence and has a lot for me to learn. He was the one who shaped the project's direction and came up with ideas to make our stories more accessible to the audience in this project. He is also an expert in data visualization and provides fascinating insights into the data. He is also very sociable and funny, which makes us feel more at ease while working hard. His modeling skills are also quite good, and he contributed to the team's Logistic Regression model. He is also conscientious and hardworking. During difficult and stressful times, Abid frequently cracks jokes to lighten the mood, allowing us to relax and focus on solving each project difficulty. I like how confident and open he is. He offers many interesting ideas for making presentations simpler, easier to understand, and to the point when creating a Powerpoint presentation. Overall, he is an upbeat, cheerful, and enthusiastic individual.

**My Analysis & Techniques**

In this project, I completed several analyses, such as the data description at the start of the project, to help the team understand each feature of the data more thoroughly. I also use Pandas, Matplotlib, and Seaborn to generate and analyze each data's histogram and boxplot to assist the team in understanding outliers and noise in the data. I use Linear Correlation to determine which properties to include in the model. I propose creating a data pipeline for each data type to help transform features faster and save time. I built and tested two models, XGBoost and Random Forest, and fine-tuned the hyperparameters to get the best results from the model. I also calculated feature importance to determine which characteristics are useful for predicting song popularity. Finally, I contribute to the creation of reports and presentations in order to tell the best story possible to the audience.