Stephanie Benavidez

DSC680-T301 (2241-1)

Applied Data Science

**Week 1: Milestone 1 – Project Proposal and Data Selection**

For this first project of this semester, I thought it would be neat to do a classification model or models on music genres. According to an article by Ashley King, “the top streaming genres of 2023 so far are as follows: R&B/Hip-Hop, Rock, Pop, Latin, Country, Dance/Electronic, World Music, Christian/Gospel, Classical, and Jazz” (King, 2023). Growing up, my family listened to about half of these genres, especially on long road trips. My family even would turn it into a game to see if we could guess what the title of the song was, and who sang it to kill time for a long drive. Doing these activities as a child has allowed me to enjoy all genres of music, though there may be some genres that I still do not like.

With this project, I will be using audio files of the different genres, and trying to run different models to see if it can classify which genre the audio file is. According to (Music Classification: Beyond Supervised Learning, Towards Realworld Applications, 2023) , music classification “is a music information retrieval (MIR) task whose objective is the computational understanding of semantics.” Though there are four different music classification tasks, this project will just be focused on genre classification. Some of the research questions that I would like to try to explore are as follows:

- Will the model or models be able to understand the quality of the audio to choose the correct genre?

- Are there enough datasets that have great audio files to predict the genre?
- Can I get at least an 80% accuracy level within the classification models used?

For the datasets, currently, I have not found the actual datasets that I will be using for this project. Though I am sure I will be using Kaggle, or other data science websites to find the proper datasets. With any data set, there will be manipulation and cleaning to remove unnecessary data. Each data set should have at least ten genres of music with the audio files to use in the model or models. From research there are existing GTZAN datasets that have the majority of the information that will be used within this project, however, I just need to look through these data sets and decide which ones to use.

The analysis methods I have thought about using for this project are multiclass SVM, K-Means Clustering, K-Nearest neighbors, or a convolutional neural network. Though I may not be able to do all of these analysis methods in this project, my goal is to do at least two or three of them. As with any project, after you have the data sets, and try to analyze the data, the methods tend to change. Hopefully, out of these four, the results are good, and nothing will be necessary to change.

The ethical considerations to consider are the data sets audio files not having excellent quality, therefore, not allowing the project to reach the potential that is needed. There may also not be enough resources available to have the proper data sets to have a successful project. Lastly, time is shorted to four weeks, and may not produce the outcomes as desired.

For the challenges or issues that may be faced within this project, the main concern is the data. Data always plays a huge role in the outcome of the project. This

main concern is also why I have not fully chosen which data sets to use. Secondly, make sure your coding to clean and prepare the data, as well as each model is correct. Being new to some of these models, the coding can be a little tricky, not allowing the desired results to be formulated. Lastly, a lack of resources can be difficult in any project, which also makes it difficult to get the desired results.

In conclusion, I hope to learn more about classification models and which ones work best to help fulfill the desired outcomes with this project. Though throughout these semesters in our data science program, we have gone through them, the more practice we have the better it will be for future data science jobs. Hopefully, I will be able to produce the results and get good accuracy for each genre.

## References


King, A. (2023, July 22). *Top Streaming Genres of 2023 So Far & Consumption Habits.* Retrieved from https://www.digitalmusicnews.com: https://www.digitalmusicnews.com/2023/07/22/top-streaming-genres-of-2023-so-far-and-consumption-habits/

*Music Classification: Beyond Supervised Learning, Towards Realworld Applications*. (2023). Retrieved from https://music-classification.github.io: https://music-classification.github.io/tutorial/part1_intro/what-is-music-classification.html