Rachael Enders

Faith Mathew

Cameron Schneider

Shannon Wasson

CIS 635

May 31, 2024

Project Progress Report

- **Progress so far**
  - *Completed Tasks: Detail what portions of the project you have successfully completed up to this point.*
    - So far, we have started working on the data preprocessing portion of the project. Specifically, the data cleaning portion has been completed. This portion involved removing the track ID section, since it was unnecessary labeling, removing all missing data points, and checking for duplicates. After doing the data cleaning, since we want to understand more the data distribution of the different audio features before applying clustering algorithms, we came up with three histograms using matplotlib library, which gave us foundation for more complex analyses.
  - *Data: Provide links to the datasets you're utilizing.*
    - https://www.kaggle.com/datasets/maharshipandya/-spotify-tracks-dataset
- **Challenges**
  - *Encountered Difficulties: Discuss any significant challenges you have faced during the project. This could include technical difficulties, conceptual misunderstandings, or data issues.*
    - One challenge was trying to get the dataset to load into Google Colab. It would not load since it was only available in a CSV file and there was not a URL file available.
  - *Describe how you addressed these challenges. If some challenges remain unresolved, outline your proposed approach to tackle them moving forward.*
    - The only way to get the dataset uploaded was by manually uploading the file every time the program was run.
- **Collaboration (if applicable)**
  - *How frequently does your group meet?*
    - The group meets asynchronously via Discord, text, and blackboard.
  - *Does everyone feel like each group member is contributing meaningfully to the project?*
    - Everyone seems to be contributing equally and working together to answer any questions anyone may have. So far our group has worked efficiently

and well together. Being a group asynchronously, everyone has adapted and shown up and contributed.

- **Next Steps**
  - *Remaining Tasks: Outline what still needs to be accomplished on your project.*

Implementing the clustering algorithms for grouping Spotify tracks based on various features.

  - *What is your plan to complete the remaining tasks?*

Selecting the appropriate clustering algorithms, implementing them and evaluating the clustering results, and then integrating the clustering into the recommendation system.

  - *Do you foresee any potential challenges?*

When determining the right number of clusters.