

Project Proposal

The question I want to answer in my project is how accurately I can classify songs into their respective genres using audio features. I will use SQLite to create a database to manage my songs and hopefully use a Discord bot for user interaction. Specifically, I want to explore how effective audio features like tempo, energy, and lyrics for genre classification and how these results can be stored, queried, and shared interactively.

To do this, I will first gather a dataset of songs and their genres using Kaggle or Spotify API. These features will include key attributes such as tempo, energy, and danceability. I will then use Genius to collect the lyrics for each song to add that as a feature. Everything will then be processed and stored in an SQLite database for efficient management.

Next, I will build a machine learning model to classify songs into genres. I plan to use models like Random Forest or CNNs, depending on the performance during the testing phase. Once trained, the model will predict genres based on the extracted audio features, and the results will also be stored in the SQLite database. I plan to achieve this through numerous functions that so hopefully the only thing that needs to be given to the bot is the song and artist.

The final step involves integrating a Discord bot. Users will be able to interact with the bot by uploading songs for genre prediction or querying the database for existing results. For example, users can upload a song file through a Discord command, and the bot will extract features, classify the song, and return the predicted genre. The bot will also allow users to search for songs by title or artist to view classification results from the database.

I anticipate that using SQLite and Discord will add layers of interactivity and accessibility to the project. The SQLite database will act as a repository for storing song metadata, features, and predictions, making the data reusable for future enhancements. The Discord bot will allow real-time interaction.

For visualizing the results, I plan to create interactive graphs showing the distribution of genres, the performance of the classification model, and feature importance. These visualizations will be created using libraries like Plotly or Matplotlib.

I think the largest challenge will come creating the Discord bot and when tying all the pieces together. However, by breaking the project into pieces and making sure to create each part with the next/last in mind should make sure everything goes smoothly.