Project Proposal

**October 23rd 2024**

# Problem Statement

Spotify recently launched a new AI-supported playlist called "daylist," which detects the day and time when you open it. Then search your listening history to generate a playlist based on what you typically enjoy at that moment. It's an amazing feature, and we've been enjoying it a lot. Additionally, the description section of the playlist includes a short explanation listing the genres it contains. However, this information is often inaccurate.

# Proposed Solution

We propose developing a machine learning model that can accurately classify the genres of songs in the "daylist" playlist. This model would ensure that the listed genres in the playlist's description match the actual content, providing a more accurate and personalized listening experience.

# Dataset

We plan to use the “Prediction of Music Genre” dataset from Kaggle, as it provides a rich set of audio features that Spotify provides. This metadata contains important information such as key, danceability, energy and popularity. These can enable us to train models effectively for music genre classification. This dataset’s diversity, and size make it suitable for developing predictive machine-learning algorithms.

# Model

We intend to use the Random Forest Model. The data size is 7.54 MB, which is considered small to medium. This fits as Random Forests do not require large amounts of data to achieve high accuracy compared to other options such as Neural Networks. Additionally, compared to Individual Decision Trees, Random Forest Models enhance the accuracy of genre predictions since they are less sensitive to anomalies or outliers in the data, and reduce overfitting. Underrepresented genres can also be handled through techniques such as Class Weighting, which will ensure that minority genres are accurately classified.

# Evaluation

We will be using F1-Score and Confusion Matrix as the evaluation metrics for our music genre classification. An imbalance in the dataset can impact model performance; the F1-Score effectively balances precision and recall, making the evaluation crucial for underrepresented genres. Moreover, the Confusion Matrix will help visualize results and identify genres with frequent misclassifications, ensuring a comprehensive evaluation.