

# CSE/ECE 343 : Machine Learning Project Proposal

## Title : Automated Music Playlist Generator

Arka Sarkar  
arka18222@iiitd.ac.in

Pankil Kalra  
pankil18061@iiitd.ac.in

Daksh Thapar  
daksh18137@iiitd.ac.in

### 1 Motivation

With the growing popularity of music streaming services like **Spotify**, **Apple Music** and **Wynk**, the number of songs have skyrocketed globally. Thus creating personalized playlists for users has become tedious and challenging as it involves individual listening to various songs and categorizing them based on their audio features. The objective is to sort songs with similar musical characteristics into playlists automatically. Modern machine learning techniques and visualization tools should help us find accurate models that categorize millions of songs into user playlists based on song choices.

### 2 Related Work

Playlist generation is a broad problem which has been approached in many different ways:

1. Automated Playlist generation from Personal Music Libraries [1] by Diana Lin and Sampath Jayarathna provides a method to use K-means clustering, Affinity Propagation and DBSCAN on various audio features (such as "danceability", "energy", "acousticness") to generate an automated playlist.
2. A Comparison of Playlist Generation Strategies for Music Recommendation and a New Baseline Scheme [2] by Geoffray Bonnin and Dietmar Jannach uses multiple methods such as KNN, Popularity-based, Same artists - greatest hit, Collocated artists - greatest hit, Content-based approaches to generate playlists.
3. Music Playlist Generation based on Community Detection and Personalized PageRank [3] by Bangzheng He, Yandi Li and Bobby Nguy uses song similarity graph approaches and apply community detection and Personalized PageRank to generate playlists.

### 3 Timeline

A Tentative 12 week timeline :

**Week 1-2** : Data Collection (including Scraping).

**Week 3** : Pre-processing and Data Visualization.

**Week 4** : Feature Extraction.

**Week 5** : Feature Analysis, Selection, Correlation, HeatMaps.

**Week 6** : Logistic Regression, Support Vector Machines.

**Week 7** : Decision Trees, Random Forest.

**Week 8** : K- Nearest Neighbours, K- Shortest Path.

**Week 9** : Analysis and performance of models.

**Week 10** : Hyperparameter Tuning, Check for model Overfitting and Underfitting.

**Week 11** : Report Writing.

**Week 12** : Buffer.

### 4 Individual Tasks

Tasks	Team Member/s
Data Collection	Daksh and Pankil
Pre-processing and Data Visualization	Arka and Pankil
Feature Extraction	Arka and Daksh
Analysis of Features (Selection , correlation, heatmaps. etc....)	Arka, Daksh and Pankil
Logistic Regression and Support Vector Machines	Daksh
Decision Trees, Random Forest	Arka
K-Nearest Neighbours, K - Shortest Path	Pankil
Hyperparameter Tuning , Check for Overfitting and underfitting of models, and selection of best model	Arka, Daksh and Pankil
Report Writing	Arka, Daksh and Pankil

### 5 Final Outcome

The objective is to generate automated song playlists for users using Label Classification and Graph based techniques. We will identify specific audio features that have highest correlation to our training dataset, use Topic modelling techniques like LDA[4] for song lyrics and compare various Machine Learning models using our feautres based on metrics like Accuracy, Precision, F1 score and NMI for our classification problem. Graph-based approaches will be used to generate the playlist, measure the similarity between our predicted song and the actual song labels and will provide us meaningful metrics to visualize playlists.

The project will help us identify important audio features and song lyric collections which simulate playlist generation process for users. This can find key applications among music enthusiasts and in mobile applications to make song collections for not only individuals, but for age groups and people with similar interests for song genres and artists.

### References

[1] Automated Playlist generation from Personal Music Libraries by Diana Lin and Sampath Jayarathna.

[2]A Comparison of Playlist Generation Strategiesfor Music Recommendation and a New Baseline Scheme by Geoffray Bonnin and Dietmar Jannach.

[3] Music Playlist Generation based on CommunityDetection and Personalized PageRank by Bangzheng He, Yandi Li and Bobby Nguy.

[4]Topic Modeling with LSA, PLSA, LDA lda2Vec by Joyce Xu.