

# Clustering Music by Genres Using Supervised and Unsupervised Algorithms

## Machine Learning Project Report

Mrittika Das | A25CE09011

Anjali Sanjay Kashide | 25AI06001

Anwasha Chakraborty | 25CL06002 | Group Leader

Sayali Abhay Khamitkar | 22CS01052



Indian Institute of Technology Bhubaneswar

# 1. Data Collection and Feature Extraction

Our group collaboratively built a balanced music dataset covering four genres: **pop**, **rock**, **classical**, and **jazz**. Each member collected **40 audio samples** from their assigned genre. All clips were standardized to **30 seconds** to ensure uniformity.

We jointly completed the entire preprocessing pipeline. This included downloading audio files, converting them into MP3/WAV formats, resampling to a fixed rate, converting to mono, and trimming/padding to achieve consistent duration.

Feature extraction was performed using **Mel-spectrogram-based features (20 Mel bands)** generated with *librosa*. These features capture frequency-domain characteristics of the audio. Each genre was stored as a separate CSV file, and later all four files were merged into a **single unified dataset** for analysis.

## 2. Music Genre Classification Using CART

After preprocessing, all audio files were converted to consistent WAV format and their Mel-spectrogram features were normalized. Using the combined feature dataset, we trained a **CART (Decision Tree)** classifier to categorize songs into the four genres. An **80/20 train-test split** was used for evaluation.

**Model Accuracy: 0.75**

### Classification Summary

- **Pop** and **Jazz** showed the best recall (1.00 and 0.88), meaning most samples were correctly identified.
- **Classical** (0.62 recall) and **Rock** (0.50 recall) were more often misclassified, though both had high precision (1.00).
- Overall performance shows reasonable separability of genres using Mel-band features.

### Most Important Features

The classifier found features **f1**, **f7**, **f20**, **f11**, and **f2** to be the strongest contributors. These Mel-band components likely capture dominant frequency patterns unique to each genre.

### Result Interpretation

The decision tree indicates that a few Mel-band features hold significant discriminative power. Pop and jazz tracks show clearer Mel-spectrogram patterns, while classical and rock have overlapping frequency ranges, causing confusion. The overall 75% accuracy shows good but not perfect separability.

### 3. K-Means Clustering

We conducted clustering using **K-Means**, both **without** and **with PCA**, to evaluate how well the genres naturally cluster in feature space.

#### i) K-Means Without PCA

Using the original 20 Mel features:

- Silhouette Score: **0.49**
- Rand Index: **0.47**
- ARI: **0.11**
- NMI: **0.31**
- Purity: **0.43**

##### Interpretation:

The clusters were compact but poorly aligned with true genres. Most samples collapsed into a single cluster (mostly classical), showing that **mean Mel-band features alone are insufficient for clean unsupervised separation**.

#### ii) K-Means With PCA (Top 10 Components)

- Silhouette Score: **0.18**
- Rand Index: **0.75**
- ARI: **0.36**
- NMI: **0.45**
- Purity: **0.66**

##### Interpretation:

PCA improved alignment with true genres, giving clearer clusters. Although geometric separation remained weak, the purity and RI values show that PCA retains useful genre structure. Still, unsupervised methods struggle due to overlap in frequency content across genres.

### 4. Music Genre Classification Using SVM

We also trained a **Support Vector Machine (SVM)** classifier using the same dataset. After cleaning labels, imputing missing values, and applying **StandardScaler**, we trained an RBF-kernel SVM.

#### SVM Results

- **Overall Accuracy: 96.87%**
- **Classical & Jazz:** perfect precision, recall, F1 = **1.00**
- **Pop:** precision = **1.00**, recall = **0.83**
- **Rock:** recall = **1.00**, precision = **0.88**

## Result Interpretation

The SVM substantially outperformed CART and K-Means. Its ability to learn non-linear boundaries allowed it to capture complex differences in Mel-spectrogram patterns. This demonstrates that **supervised learning is far more effective than unsupervised clustering** for this task, given labeled data.

## 5. Contributions :

Mrittika Das (A25CE09011)- Data Collection for Jazz genre samples, followed by feature extraction+ ppt+ report (section 1)

Anjali Sanjay Kashide (25AI06001)- Data Collection for Pop genre samples, followed by feature extraction+ model training and testing using CART+ report (section 2)

Anwesha Chakraborty (25CL06002)- Data Collection for Rock genre samples, followed by feature extraction+ clustering using PCA+ K-Means clustering + report (section 3)

Sayali Abhay Khamitkar (22CS01052)- Data Collection for Classical genre samples, followed by feature extraction+ model training and testing using SVM + report (section 4)