Report project Python for AI:Classification for Music Genres

1-Problem description

**Dataset**:

music\_genre.csv(<https://www.kaggle.com/datasets/vicsuperman/prediction-of-music-genre>)

-As I understand the dataset is acquired from Spotify music app

- 50005 datapoints(5 of which is empty rows)

-Number of features are 18

Numerical features include: instance\_id, popularity, acousticness, danceability, duration\_ms, energy, instrumentalness, liveness, loudness, speechiness, tempo, valence

Categorical features include: artist\_name, track\_name, key, mode, obtained\_date, music\_genre

-X∈R50005x17

**Labels**:

-Target Variable: ‘music\_genre’

-Classes (Genres): 10 (Electronic, Anime, Jazz, Alternative, Country, Rap, Blues, Rock, Classical, Hip-Hop)

-Label Distribution: Each genre has 5000 instances.

- y∈{Electronic, Anime, Jazz, Alternative, Country, Rap, Blues, Rock, Classical, Hip-Hop}

Histogram:

All genres are evenly represented with 5000 instances each.

**Goal:**

-To predict the music genre of a track using various classification algorithms.

-Aim to achieve a performance of at least 60% accuracy.(as there are 10 classes to determine and the nature of the dataset is taken into consideration goal was set to a reasonable ratio)

2- Data preprocessing & Feature engineering

**Feature engineering steps:**

**Data cleaning steps:**

**Pre-processing steps:**

**Data splits, tried different ways:**

3- Methods

4-Results

5-Conclusion