Lab 1: Music Genre Classification

1 Files included

- o GenreAll.arff – arff dataset file with global song Essentia descriptors
- o GenreAll.csv – csv dataset file with global song Essentia descriptors
- o features_30_sec.csv – csv file with global song descriptors
- o features_3_sec.csv - csv file with 3sec-window song descriptors

2 Background

Automatic genre classification using machine learning techniques is pivotal in managing and enriching large-scale text and music collections. In textual data, assigning genre labels—such as promotional or legal—enhances information retrieval by enabling users to filter documents based on their specific needs, thereby improving search relevance and efficiency. This automated labeling also aids in the curation of web corpora, ensuring a diverse and representative dataset for training language models, which is essential for advancing natural language processing applications. In the realm of music, genre classification facilitates the organization of extensive audio libraries, supporting personalized recommendation systems and efficient content-based music retrieval. By analyzing audio features, machine learning models can accurately categorize tracks into genres like classical, jazz, or hip-hop, enhancing user experience in music streaming services. Overall, the integration of automatic genre classification streamlines the management of vast information repositories, making them more accessible and user-centric.

3 Tasks

Implement the following tasks using a Google Colab notebook.

Task 1. Using the file with song Essentia descriptors (GenreAll file), your aim is to train a model with the highest accuracy possible. Explore different machine learning algorithms (e.g, Logistic regression, k-NN, decision trees, ANN, and ensemble methods). Check for overfitting and, if necessary, perform feature selection and tune the algorithm's parameters. Evaluate the model using 10-fold cross-validation.

Task 2. Same as Task 1, but this time using the features provided in the CSV files *features_30_sec.csv* and *features_3_sec.csv*.

Submitting your answer

The Lab can be solved individually or in teams of two people (1 submission per team). Submission is through the Aula Global. Submissions should contain a PDF file with your name(s), the link to your Google Colab notebook (make sure you give me access to it) and a summary and discussion of the results. Provide all the results for every algorithm and parameters/features choices you made. The Lab submission deadline will be discussed in class.