Classify Music Genre !!!

Abstract

The goal of this project is to classify the data acquired from [MachineHack](https://machinehack.com/) using different techniques and apply the knowledge received from SADAIA Bootcamp in one finalized project.

Design

The project idea was taken from [MachineHack](https://machinehack.com/) , the aim is to classify the data into 11 different classes (11 different Music genre) using two types of machine learning algorithms

One VS One multi-class with Logistic regression and **CatBoostClassifier** .

Data

The dataset contains 17996 records with 17 features . two of are related to artist and track names .

The rest are related to song properties such as Loudness , energy , length of the song .

Some of the features include null values and this is handled by using two different approaches .

Algorithm

Models

Logistic regression with OneVSOne multi-class

CatBoost

Model Evaluation and Selection:

The entire training dataset of was split into 70/30 train vs. holdout

In this project , different scores used to show the performance of the selected models .

Scores

For Logistic Regression :

The highest accuracy for this model was 0.50

For CatBoost :

The highest accuracy for this model was 0.70 .

Tools

- Pandas for data manipulation
- Scikit-learn for modeling , scaling the data and filling null values
- texthero for NLP visualization
- Matplotlib and Seaborn for plotting
- catboost for modeling

Communication

Initially I used only logistic regression and ended up with 0.50 accuracy then I looked for the winners of this competition and select one of the top ranking and apply its solution and use it as comparison through the notebook.