Music Lyrics Classifier

Introduction

This project involves the creation of a music lyrics classifier using Machine Learning models. The aim is to categorize lyrics into different music genres based on their lyrical content. Such a classifier could have a variety of applications, such as helping to understand the distinguishing features of different genres, or assisting music platforms in automated categorization of music content.

Dataset

The dataset for this project has been sourced from Kaggle, and can be found at the following link: https://www.kaggle.com/code/levkharlashkin/music-classification/input. The dataset consists of songs spanning various genres, with each song's lyrics stored in textual form.

Implementation

The classifier is implemented in the Jupyter notebook named Classifier.ipynb. The notebook includes data preprocessing, exploratory data analysis, and model training and evaluation stages. We have employed Naive Bayes, Decision Tree, and Random Forest classifiers to predict the music genre based on the lyrics.

To run the classifier, you need to open the Classifier.ipynb notebook in Jupyter, and run the cells in sequence from the top.

Prerequisites

Ensure that you have the following installed:

- Python 3
- Jupyter
- Libraries including pandas, numpy, scikit-learn, and nltk

Usage

1. Download the repository and navigate to the folder via the command line.
2. Launch Jupyter by typing jupyter notebook in your command line and pressing enter.
3. Open the Classifier.ipynb notebook.
4. Ensure the dataset is available in the same directory as the notebook, or modify the data loading cell to match your dataset location.
5. Run the cells in the notebook from top to bottom.

Please note that some cells might take a while to execute, especially the model training cells, due to the volume of data and the complexity of the models.

Results

The results of the classifier can be found in the final sections of the Jupyter notebook. We have evaluated the model performance using various metrics including accuracy score, confusion matrix, and classification report. For a detailed analysis of the results,

please refer to the notebook.