Indian Institute Of Technology Jodhpur

IML-CSL2010

Music genre classification

Revalla Manjunadh **Angothu Jeevan kumar**

Pamarthi Rohith **Gaddam Sai Charan Reddy**

(B23MT1033,B23ME1042,B23ES1007,B23BB1019)

Abstract

The digital music landscape is vast, presenting listeners with an overwhelming array of choices. This project focuses on developing a robust music genre classification system to facilitate music exploration. Our classification system utilizes various machine learning algorithms, including Logistic Regression, SVM, MLP, decision tree and KNN.

We rigorously evaluate these algorithms using real-world datasets, demonstrating their effectiveness in accurately categorizing songs into distinct genres. Through comprehensive assessments and user studies, we highlight the strengths of the classification system, providing a structured approach for users to navigate and discover different music genres. This system aims to enhance user engagement and simplify the exploration of the diverse digital music landscape.

Contents

*1 introduction

In the era of digital music streaming, listeners are faced with an overwhelming variety of song choices. This abundance of options, while a testament to the diversity of musical content available, presents a challenge: how can listeners efficiently discover new songs that align with their current interests? This project introduces a music genre classification system to address this challenge, providing a structured approach to navigating and exploring different musical genres.

> 1.1 problem statement

The digital music landscape offers a vast array of songs and artists. However, users often struggle to efficiently discover new music that
resonates with their tastes, primarily due to the challenge of navigating this extensive library. This project addresses this issue through a **Music Genre Classification System**. Recognizing the significance of music genres in facilitating song discovery, we have developed a classification system that categorizes songs into distinct genres. By leveraging machine learning algorithms such as Logistic Regression, SVM,MLP,KNN,GNB and Decision trees this system provides a structured approach for exploring music by genre.

Our project aims to:

- Facilitate genre-based music exploration through a robust music genre classification system, helping users discover songs within their preferred genres.

- Enhance user engagement and satisfaction by providing structured genre classifications that improve the overall music streaming experience.

By implementing and evaluating this classification system, we aim to transform how listeners interact with and discover music on streaming platforms, catering to both casual listeners and genre enthusiasts.

>1.2 report structure

The structure of this report has been carefully crafted to offer a thorough overview of our project's development, methodology, and findings. It is organized as follows:

- Section 2: Approaches Tried - This section explores the various machine learning algorithms we examined. We discuss the rationale behind each choice, the methodologies used, and the challenges encountered, along with a detailed comparison of their strengths and weaknesses.

• Section 3: Experiments and Results- Here, we describe the experiments carried out to assess the performance of the algorithms. We outline the dataset, experimental settings, and evaluation metrics, followed by a detailed analysis of the results.

• Section 4: Summary - This section concludes the report, summarizing key findings and contributions. We reflect on the system's performance and its potential industry impact.

• Appendix A: Contribution of Each Member - In this appendix, we outline the contributions of each team member, including their roles, completed tasks, and unique insights.

*2. Approaches tried

In our pursuit of designing an effective classification system, we explored various machine learning algorithms. Each algorithm was fine-tuned using grid search to identify the optimal hyperparameters for maximizing accuracy. Below, we provide brief notes on the different approaches we tested

> 2.1 classifier

2.1.1 Logistic Regression

Logistic Regression is a widely used statistical model for classification tasks that predicts the probability of a categorical outcome. By modeling the relationship between dependent and independent variables through the logistic (sigmoid) function, it outputs probabilities that are then used to classify observations. Unlike linear regression, logistic regression maps outputs between 0 and 1, making it effective for binary and multiclass classification. It is valued for its simplicity, efficiency, and interpretability, especially when there is a linear relationship between the predictors and the log-odds of the response variable. Regularization techniques like L1 and L2 help prevent overfitting and manage multicollinearity, making the model more robust. The code snippet provided evaluates the model using 5-fold cross-validation, which offers a comprehensive measure of its performance and helps assess how well it generalizes to unseen data by reporting cross-validation scores and training accuracy.

2.1.2 Support Vector Machine (SVM)

Support Vector Machine (SVM) is a powerful and versatile supervised learning algorithm used for both classification and regression tasks. It works by finding the optimal hyperplane that best separates data points of different classes in the feature space, maximizing the margin between the closest points (support vectors) of each class. This maximization ensures high generalization capability, making SVM particularly effective for complex datasets where class boundaries are not linear. By utilizing kernel functions, such as linear, polynomial, and radial basis function (RBF), SVM can map input features into higher-dimensional spaces, enabling it to solve non-linearly separable problems. The code snippet provided uses cross_val_score to assess the performance of an SVC model with 5-fold cross-validation, ensuring that the model's stability and predictive performance are tested across multiple subsets of the training data.

2.1.3 Decision Tree

A Decision Tree is a popular and interpretable supervised learning algorithm used for both classification and regression tasks. It functions by recursively splitting the dataset into subsets based on the most significant features, creating a tree-like structure where each node represents a decision rule, and each leaf node represents an outcome or class. This step-by-step partitioning allows the model to capture complex relationships in the data, making it flexible and easy to understand. However, without proper tuning, decision trees can overfit, leading to poor generalization on unseen data. To mitigate this, techniques like pruning or setting constraints such as maximum depth can be applied. The code provided uses cross_val_score with 5-fold cross-validation to evaluate the DecisionTreeClassifier, helping assess its performance consistency across different training subsets and ensuring a robust understanding of the model’s predictive capabilities.

2.1.4 MLP Classifier

A Multi-Layer Perceptron (MLP) is a type of artificial neural network commonly used for supervised learning tasks, including classification and regression. MLPs consist of an input layer, one or more hidden layers, and an output layer, with each layer containing interconnected nodes (neurons) that use nonlinear activation functions to model complex relationships within the data. The model in the provided code is configured with two hidden layers, each comprising 54 neurons, and is trained with a maximum of 1000 iterations to ensure convergence. MLPs leverage backpropagation and optimization algorithms, such as stochastic gradient descent, to iteratively update weights and minimize the loss function. This allows them to capture intricate patterns that simpler models might miss. The cross_val_score function is applied with 5-fold cross-validation to evaluate the performance of the MLPClassifier, ensuring the model's effectiveness and generalization are assessed across multiple subsets of the training data.

2.1.4 KNN classifier

K-Nearest Neighbors (KNN) is a non-parametric, instance-based learning algorithm employed for classification tasks, which categorizes data points based on the classes of their nearest neighbors in the feature space using distance metrics, typically Euclidean distance. In this project, KNN was utilized to classify music genres by analyzing extracted audio features from the dataset. The model's performance was evaluated through 5-fold cross-validation, which involves partitioning the dataset into five subsets to ensure that each portion serves as a validation set once while the others are used for training. The highest cross-validation score achieved reflects

KNN's effectiveness in accurately predicting music genres, demonstrating its ability to generalize well to unseen data. Its simplicity and interpretability make KNN a valuable tool for classification tasks, especially in scenarios where data patterns are inherently local and clustered.

2.1.5 Gaussian naïve bayes classifier

Gaussian Naive Bayes (GNB) is a probabilistic classifier that applies Bayes' theorem with the assumption that the features are normally distributed. This algorithm is particularly effective for classification tasks, especially when the dataset has a clear distribution of classes.

In this project, GNB was employed to classify music genres based on audio features extracted from the dataset. The model was evaluated using cross-validation, specifically a 5-fold approach, which involves dividing the dataset into five subsets. Each subset serves as a validation set while the remaining subsets are used for training, ensuring that the model is tested against various portions of the data to assess its performance.

The maximum cross-validation score obtained during this evaluation process serves as an indicator of the model's effectiveness in generalizing to unseen data. This score provides insight into how well the GNB model can accurately predict music genres based on the learned features, demonstrating its potential as a reliable classification tool in music genre analysis.

*3 Experiments and Results

In this section, we detail the experiments conducted to evaluate the performance of the various machine learning algorithms for our classification task. We also present the results obtained from these experiments.

>3.1 Dataset

The GTZAN dataset, collected in 2000-2001 from various sources, contains 1,000 audio files. It consists of 10 genres, each containing 100 audio files. Each audio file is 30 seconds long, providing sufficient data for feature extraction. The files are in .wav format, and the total size of the dataset is approximately 1.2GB. The GTZAN dataset is commonly used in music genre classification research and has been shown to yield strong results.

>3.2 Experimental Setting

We split the dataset into training and testing sets for each algorithm using an 80:20 ratio. The training set was used to train the model, while the testing set was used to evaluate its performance. To fine tune the hyperparameters for each algorithm, we employed grid search with cross-validation.

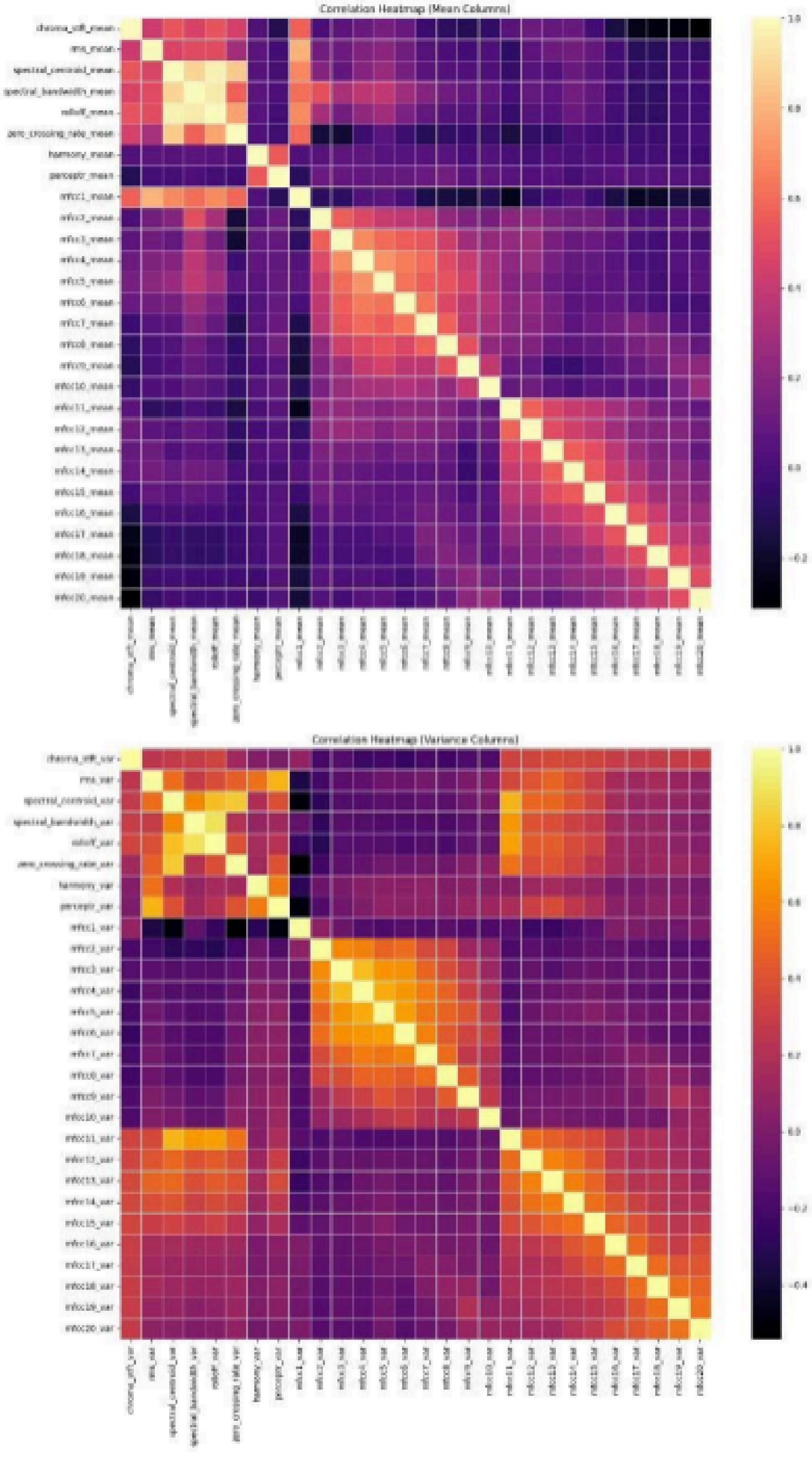
>3.3 Comparison of Results

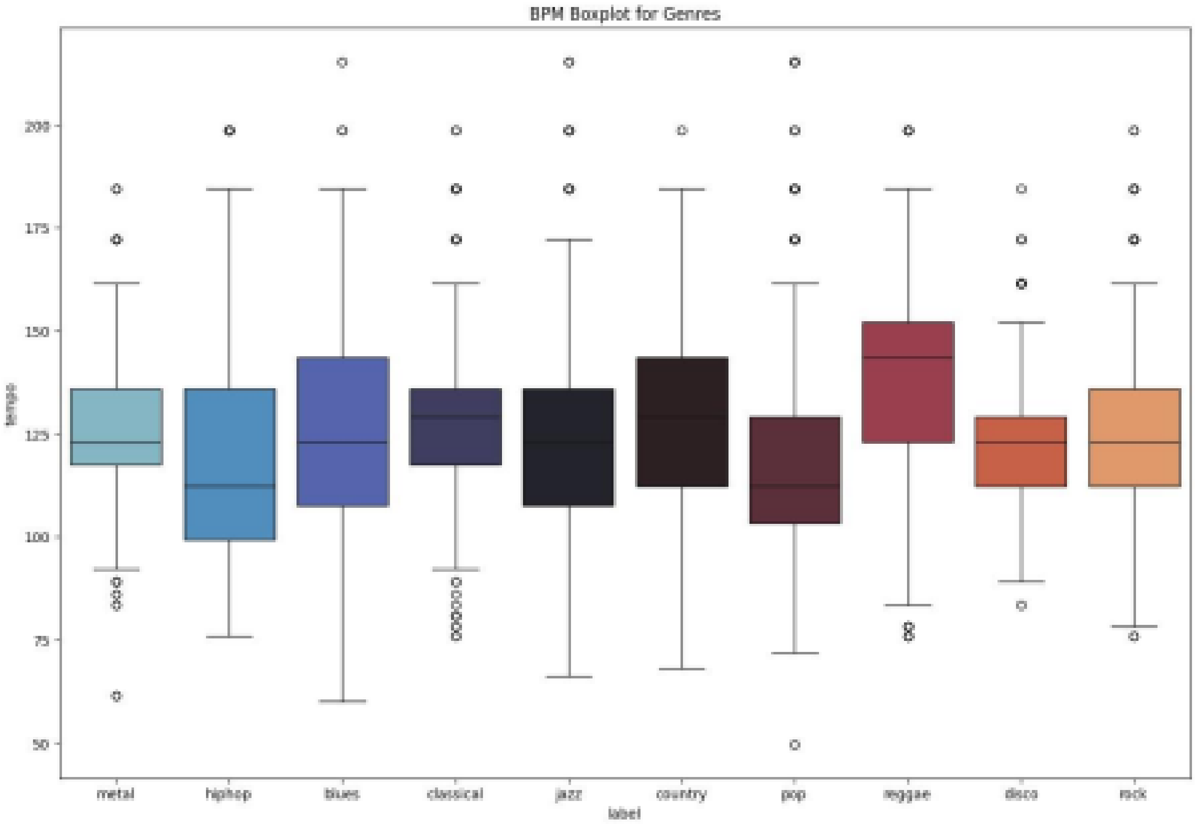
The cross-validation scores for various classifiers are as follows :

- Logistic Regression CV: 0.4809017
- MLP classifier: 0.21164684
- Support Vector Machine CV: 0.2926066
- KNN Classifier CV: 0.281779
- Gaussian Naive Bayes CV:0.4304511
- Decision Tree CV:0.6489096

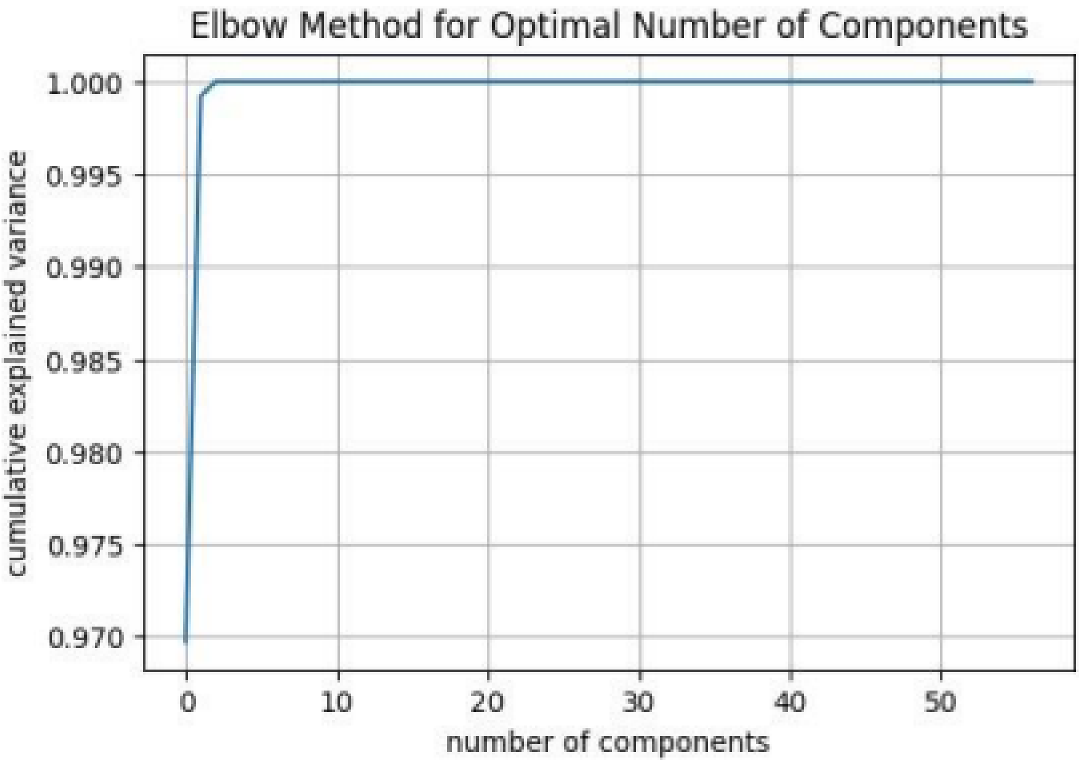
Plots for classifier:

*Figure 2: covariance of the features

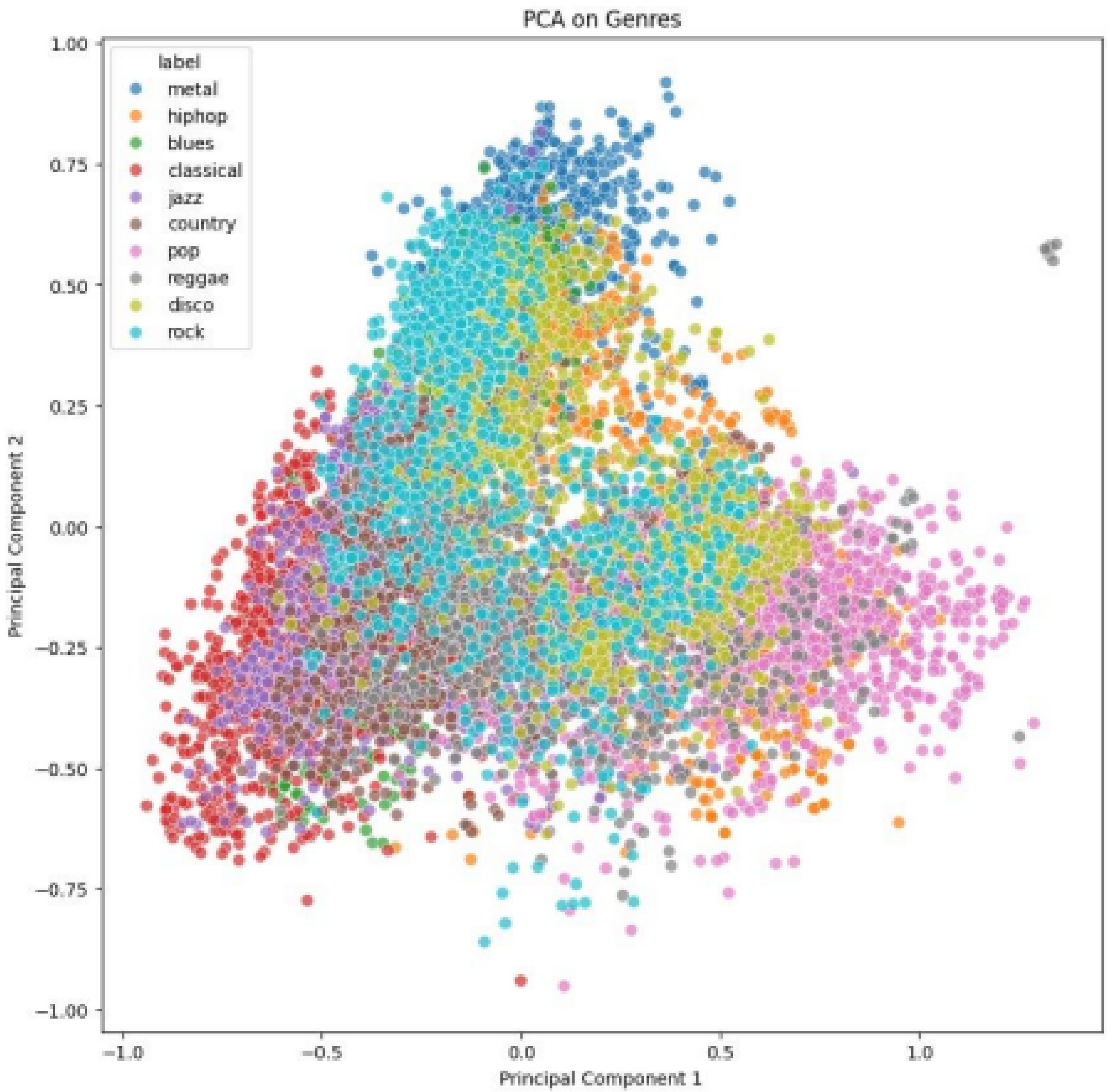




(a)boxplots of data



(b)Number of principal components for PCA and LDA

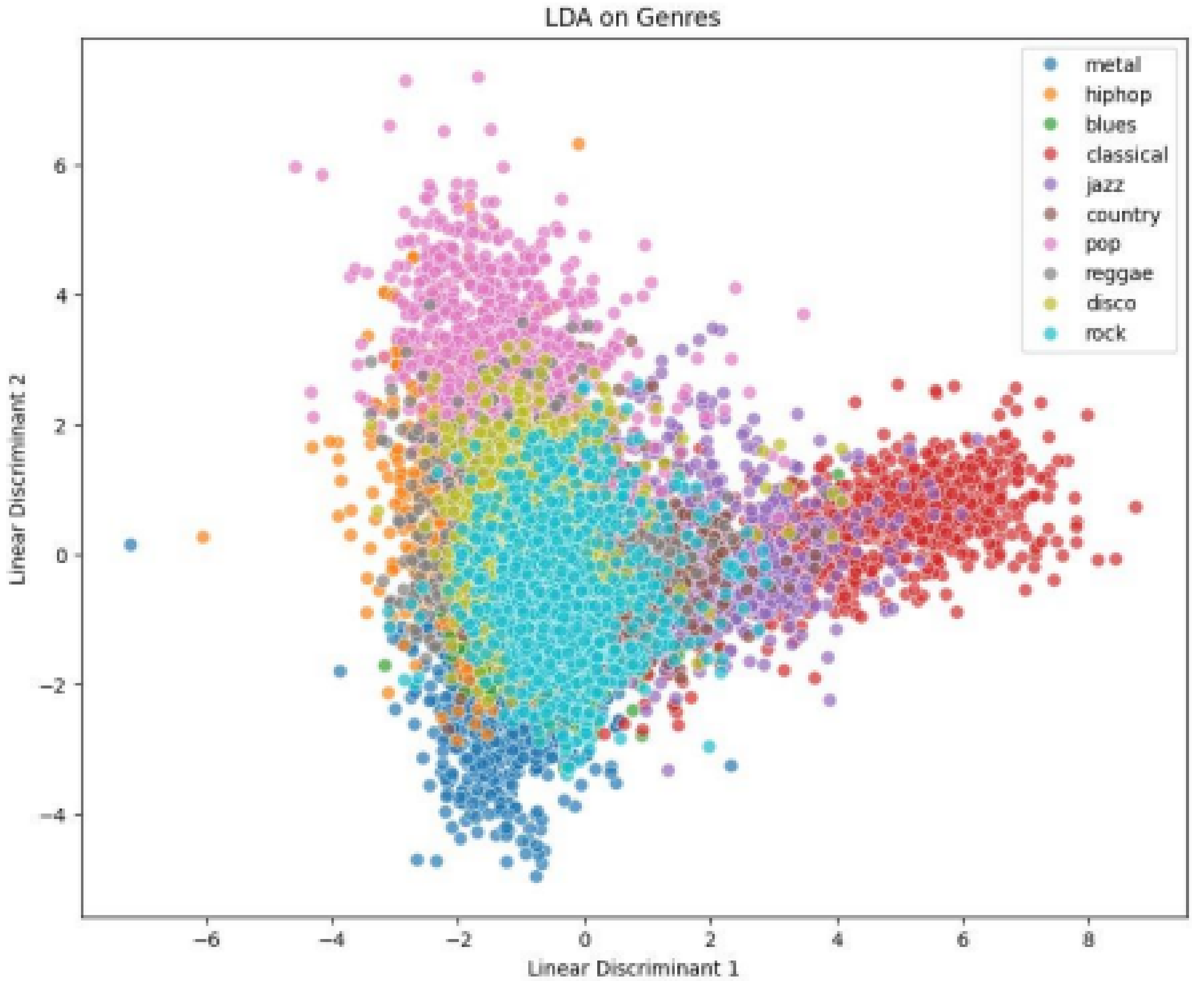


(c) PCA on genres

Principal Component Analysis (PCA) is a powerful dimensionality reduction technique used to transform high-dimensional data into a lower-dimensional space while retaining as much variance as possible. In the provided code, PCA is employed to reduce the feature set from the original dataset

��

X into two principal components, which are constructed to maximize variance and minimize information loss. Before applying PCA, the data is first normalized using Min-Max scaling to ensure that all features contribute equally to the analysis, as PCA is sensitive to the scale of the data. The transformed features, labeled as "Principal Component 1" and "Principal Component 2," are then combined with the original class labels to facilitate visualization. A scatter plot is created to illustrate the distribution of the genres in the two dimensional PCA space, where data points are color-coded according to their respective labels. This graphical representation aids in identifying patterns, clusters, and separations between different genres, providing valuable insights into the underlying structure of the dataset. The scatter plot is saved as an image file for further analysis and reporting.



(d) LDA on genres

Linear Discriminant Analysis (LDA) is a statistical method used for dimensionality reduction and classification, particularly when the classes are well-separated. In the provided code, LDA is applied to transform the original feature space X into a lower-dimensional space with two linear discriminants, LD1 and LD2, which maximize the separation between different classes represented by the labels y. The transformation captures the directions in which the classes differ the most, making it easier to visualize and analyze the data. The resulting components are organized into a new DataFrame, which includes the original labels for reference. Additionally, a scatter plot is generated to visually represent the LDA results, with points colored according to their respective classes, allowing for an intuitive understanding of the distribution and separation of genres in the reduced feature space. The plot is saved as an image file for further analysis and reporting.

4. summary

The Music Classification System project aimed to improve the user experience on music streaming platforms by providing an accurate system for classifying music genres. Using machine learning techniques such as collaborative and content-based filtering, the system classifies music tracks based on audio features and user behavior. The project explored several machine learning algorithms, including Logistic Regression, SVM, Random Forest, and Light GBM. Evaluations conducted on the GTZAN and Spotify Top Tracks datasets showed that Decision trees classifier delivered the best performance. A hybrid approach incorporating similarity metrics, clustering, PCA was also implemented to enhance the system's classification accuracy. The project’s results demonstrate its potential to improve music categorization and user engagement. Future work could focus on refining the system and scaling it to handle a broader range of music and diverse genres.