A PROJECT REPORT

on

"CLASSIFICATION OF SONG GENRES FROM AUDIO DATA"

Submitted to

KIIT Deemed to be University

In Partial Fulfilment of the Requirement for the Award of

BACHELOR'S DEGREE IN INFORMATION TECHNOLOGY

BY

ALINA TONGBRAM 21051114

LOYNA DUTTA 21051313

UNDER THE GUIDANCE OF

ABINAS PANDA

SCHOOL OF COMPUTER ENGINEERING

KALINGA INSTITUTE OF INDUSTRIAL TECHNOLOGY

BHUBANESWAR, ODISHA - 751024

March 2024

KIIT Deemed to be University

School of Computer Engineering
Bhubaneswar, ODISHA 751024

CERTIFICATE

This is certify that the project entitled

“CLASSIFICATION OF SONG GENRES FROM AUDIO DATA”

submitted by

ALINA TONGBRAM	21051114
LOYNA DUTTA	21051313

is a record of bonafide work carried out by them, in the partial fulfilment of the requirement for the award of Degree of Bachelor of Engineering (Computer Science & Engineering OR Information Technology) at KIIT Deemed to be university, Bhubaneswar. This work is done during year 2023-2024, under our guidance.

Date: 21/03/2024

Abinas Panda
Project Guide

Acknowledgements

We are profoundly grateful to **ABINAS PANDA** of **KIIT University** for his expert guidance and continuous encouragement throughout to see that this project rights its target since its commencement to its completion. ....

ALINA TONGBRAM
LOYNA DUTTA

ABSTRACT

This project focuses on using machine learning to classify audio tracks into hip-hop and rock genres. By employing Principal Component Analysis (PCA) on scaled audio data, we identified the most significant features for classification. We then trained a decision tree and compared its performance to logistic regression, noting differences in effectiveness. To enhance model accuracy, we balanced the dataset and employed cross-validation for a comprehensive evaluation. Our findings demonstrate the potential and challenges of automated music genre classification, offering insights for applications in music streaming and digital libraries.

Additionally, the project underscored the importance of data preprocessing and feature selection in achieving high classification accuracy. Through balancing the data and utilizing PCA, we significantly reduced model bias and improved the interpretability of the results. The comparative analysis between decision trees and logistic regression provided a nuanced understanding of model strengths in context-specific applications. Ultimately, this research not only contributes to the technical field of music genre classification but also suggests practical pathways for enhancing user experience in music discovery platforms and digital archives, highlighting the critical role of machine learning in the evolving landscape of digital musicology.

Contents

Chapter 1

Introduction

In the realm of digital musicology and music information retrieval, the classification of musi c genres stands as a pivotal challenge, bridging the gap between computational methods and the intrinsic complexity of musical expression. With the exponential growth of digital music collections and streaming platforms, the need for accurate and efficient music genre classific ation has become increasingly pronounced. This project embarks on an exploratory journey t o harness machine learning algorithms for the classification of audio tracks into two fundame ntally distinct music genres: hip-hop and rock. Leveraging a comprehensive dataset and empl oying advanced data preprocessing techniques, including Principal Component Analysis (PC A) for feature reduction, this study aims to identify the most salient audio features that distin guish these genres. Through the application of decision trees and logistic regression models, complemented by strategic data balancing and rigorous cross-validation, the research seeks t o navigate the challenges of genre classification. The initiative not only aims to enhance the precision of music categorization but also to contribute to the broader discourse on computati onal approaches in musicology, offering insights that could refine music recommendation sy stems and enrich the metadata of digital music libraries.

Importance of the Project:

Innovation in Feature Extraction: By analyzing audio files to distinguish between Hip-Hop and Rock, the project contributes to the development of sophisticated algorithms for feature extraction from audio signals, which is a cornerstone of MIR.

Improving Genre Classification Models: Genre classification is a challenging task due to the subjective nature of genre boundaries. This project aids in refining models that can navigate these complexities, offering insights into music genre classification at large.

Market Analysis and Targeting: Understanding the genre distribution of songs can help artists and record labels tailor their marketing strategies and target their releases more effectively to the appropriate audience segments.

Chapter 2

Basic Concepts/ Literature Review

2.1 Audio Signal Processing

Audio Signals: Representations of sound waves in a format that can be processed by computers. They are typically analog signals converted into digital form through sampling.

Feature Extraction: The process of deriving meaningful information from audio signals that can be used for classification. Features might include spectral content, rhythm, and timbre.

2.2 Machine Learning for Audio Classification

Supervised Learning: Involves training a model on a labeled dataset, where each example is a feature vector with an associated label (e.g., "Hip Hop" or "Rock").

Classification Algorithms: Include Decision Trees, Logistic Regression, Support Vector Machines (SVM), and Neural Networks. Each has strengths and weaknesses in handling audio classification tasks.

Literature Review

The field of musical genre classification, particularly distinguishing between Hip Hop and Rock, has evolved significantly from its early reliance on basic signal processing and machine learning techniques to the adoption of advanced deep learning models like Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs). These developments have led to notable improvements in classification accuracy. However, challenges such as data imbalance, the subjective nature of genre categorization, and model generalization across varied datasets persist. Moreover, the field grapples with ethical considerations due to the cultural and subjective intricacies involved in music genre classification.

Chapter 3

Problem Statement / Requirement Specifications

3.1 Project Planning

Data Collection: Assembling a balanced and diverse dataset of music tracks from the two genres.
Feature Extraction and Selection: Identifying and extracting meaningful audio features that can distinguish between genres.
Model Development and Training: Designing and training machine learning models on the extracted features.
Evaluation and Testing: Assessing the model's performance using appropriate metrics and refining the approach based on feedback.

3.2 Project Analysis

Data Requirements
Volume and Variety: A large and varied dataset is crucial to capture the breadth of each genre.
Quality and Labeling: High-quality audio files with accurate genre labels are essential.

Technical Requirements
Computational Resources: Adequate processing power for data analysis and model training.
Software and Tools: Access to machine learning libraries (e.g., scikit-learn, TensorFlow) and audio processing tools.

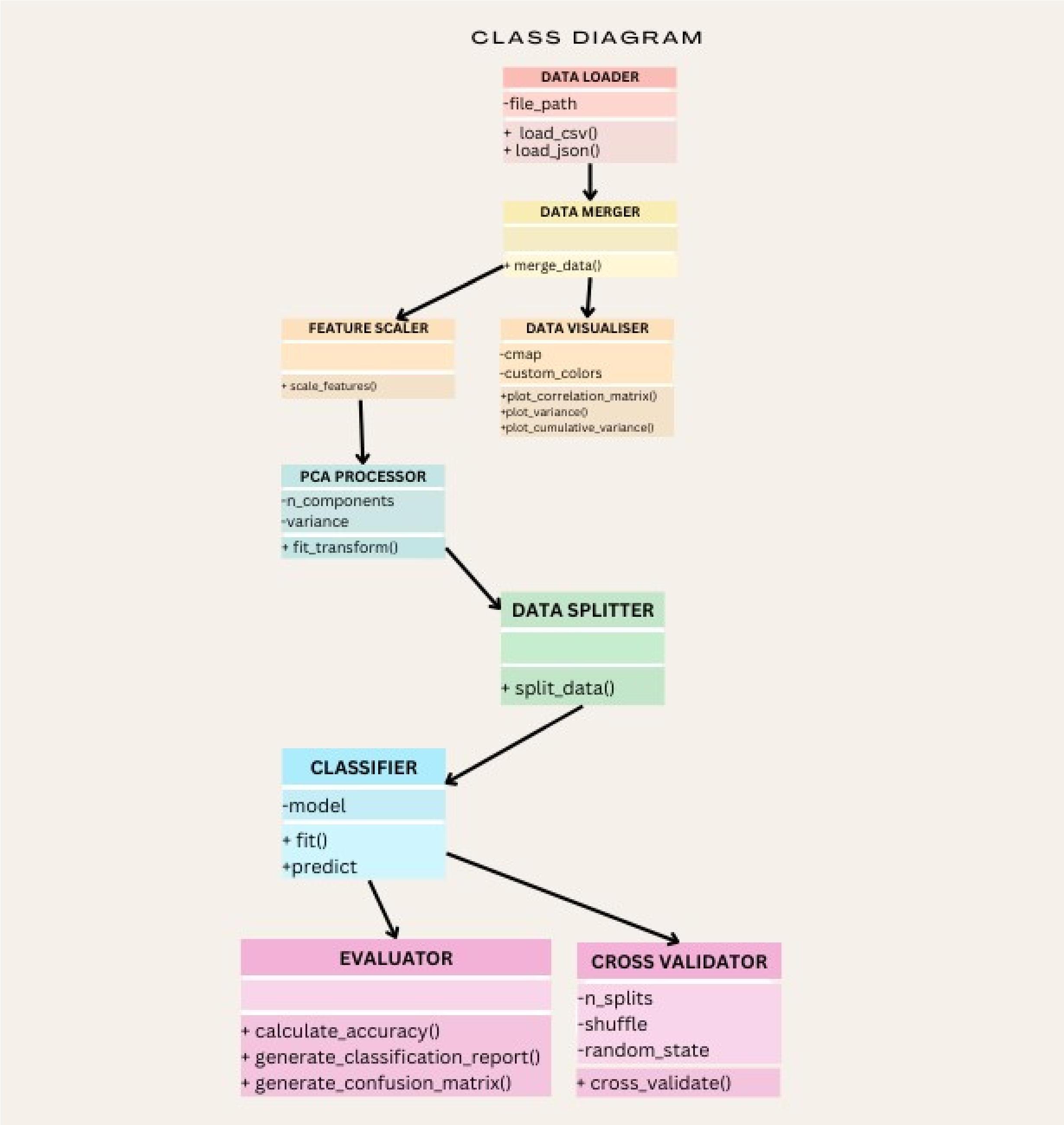
Stakeholder Requirements
Accuracy: High classification accuracy to meet user expectations.
Efficiency: Fast processing times for real-time or near-real-time classification.

3.3 System Design

3.3.1 Design Constraints

Scalability: The system must handle increasing amounts of data without significant degradation in performance.
Interoperability: Compatibility with various audio formats and integration capability with different platforms.
User Experience: Simplicity in deployment and interaction, ensuring accessibility for users without technical backgrounds.

3.3.2 System Architecture OR Block Diagram



Chapter 4

Implementation

4.1 Methodology OR Proposal

Data Acquisition and Preprocessing

The methodology begins with the acquisition of two datasets: track metadata with genre labels and track metrics with features from the Echo Nest API. These datasets are imported using pandas, and a merge operation is performed on the `track_id` column to consolidate genre labels with their corresponding audio features.

The preprocessing phase includes:
- Merging datasets based on `track_id` to align track metadata with Echo Nest metrics.
- Visualizing correlations among audio features using a heatmap, which aids in understanding feature relationships and potential multicollinearity.
- Standardizing the feature set to ensure that the PCA (Principal Component Analysis) can accurately identify variance across features.

Feature Reduction with PCA

PCA is employed to reduce the dimensionality of the feature set, aiming to retain the most significant components that explain a substantial portion of the variance in the data. A cumulative explained variance plot helps determine the optimal number of components required for effective classification without significant loss of information.

Balancing the Dataset

Given the initial imbalance between the genres (hip-hop and rock), the dataset is balanced by downsampling the overrepresented genre. This step is crucial to prevent model bias toward the more prevalent class.

Model Training and Evaluation

Model Selection: Decision Tree and Logistic Regression models are trained using the features transformed by PCA. These models are chosen for their interpretability and suitability for binary classification tasks.

Training and Testing Split: The PCA-transformed and balanced dataset is split into training and testing sets, ensuring stratified sampling to maintain genre proportionality.

Model Evaluation: The models are evaluated using accuracy scores, confusion matrices, and classification reports. These metrics provide insights into the models' performance, including precision, recall, and F1 scores for each genre classification.

2.2 Testing Plan

Cross-Validation

K-Fold Cross-Validation: To assess the models' generalizability, a K-fold cross-validation approach is adopted, with the dataset divided into ten folds. This method ensures that every observation from the original dataset has the chance of appearing in the training and test set, providing a more reliable estimation of the models' performance.

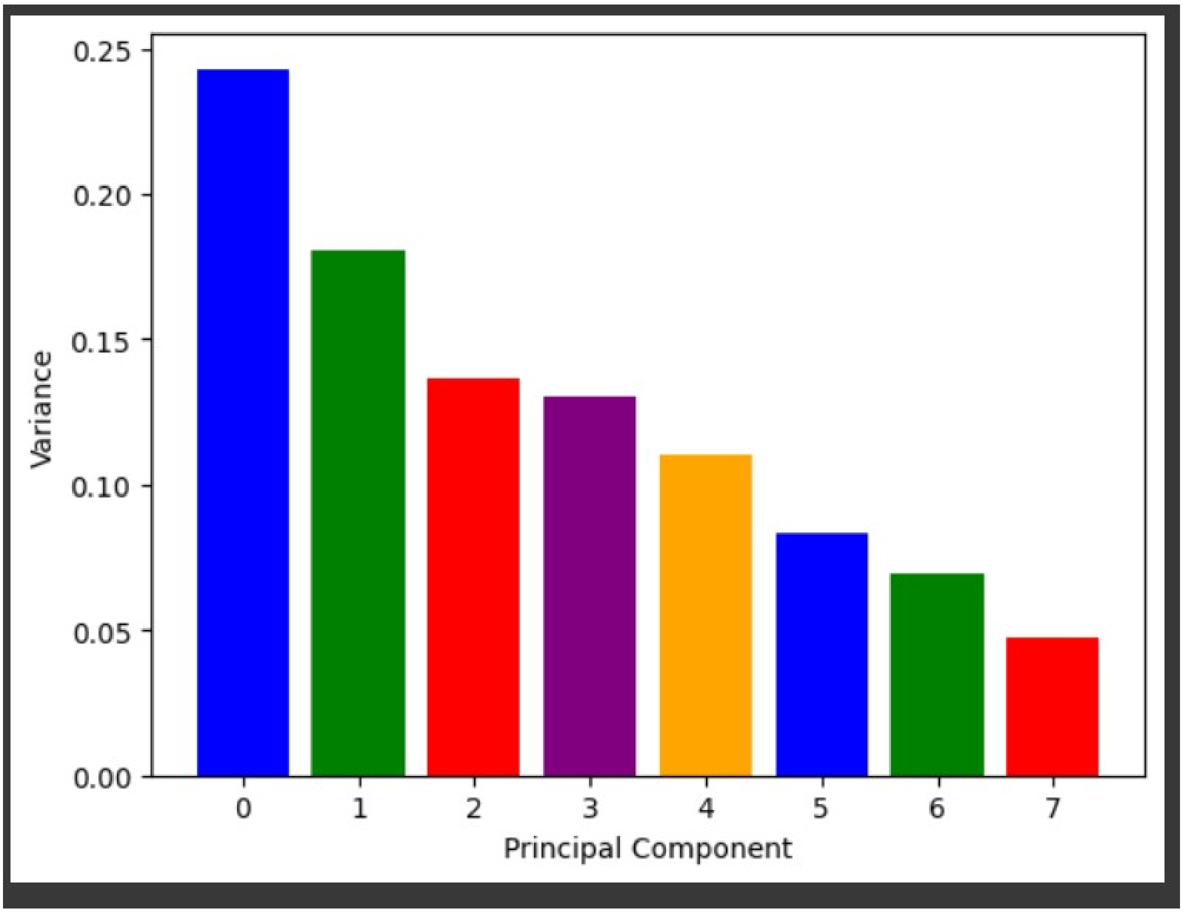
Model Comparison: The mean accuracy scores from the cross-validation of both the Decision Tree and Logistic Regression models are compared to determine the more effective classifier for this specific task.

Comparative Analysis

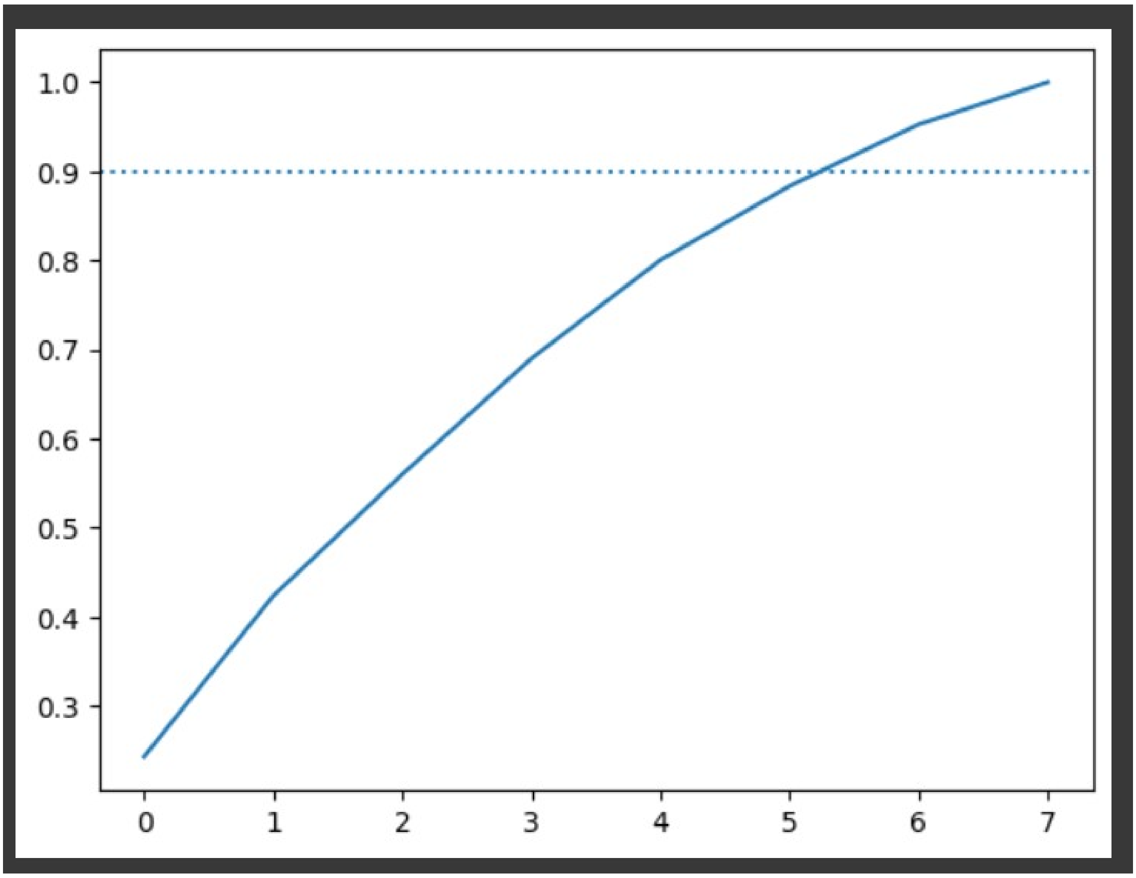
- The performance metrics obtained from the initial model evaluations and the cross-validation process are analyzed to compare the effectiveness of the Decision Tree and Logistic Regression models in classifying songs into hip-hop and rock genres.
- This analysis will focus on identifying which model provides a better balance between precision and recall, considering the potential trade-offs between these metrics in the context of music genre classification.

4.3 Result Analysis OR Screenshots

	track_id	acousticness	danceability	energy	instrumentalness	liveness	speechiness	tempo	valence
track_id	1.000000	-0.372282	0.049454	0.140703	-0.275623	0.048231	-0.026995	-0.025392	0.010070
acousticness	-0.372282	1.000000	-0.028954	-0.281619	0.194780	-0.019991	0.072204	-0.026310	-0.013841
danceability	0.049454	-0.028954	1.000000	-0.242032	-0.255217	-0.106584	0.276206	-0.242089	0.473165
energy	0.140703	-0.281619	-0.242032	1.000000	0.028238	0.113331	-0.109983	0.195227	0.038603
instrumentalness	-0.275623	0.194780	-0.255217	0.028238	1.000000	-0.091022	-0.366762	0.022215	-0.219967
liveness	0.048231	-0.019991	-0.106584	0.113331	-0.091022	1.000000	0.041173	0.002732	-0.045093
speechiness	-0.026995	0.072204	0.276206	-0.109983	-0.366762	0.041173	1.000000	0.008241	0.149894
tempo	-0.025392	-0.026310	-0.242089	0.195227	0.022215	0.002732	0.008241	1.000000	0.052221
valence	0.010070	-0.013841	0.473165	0.038603	-0.219967	-0.045093	0.149894	0.052221	1.000000



Variance
0.00
0.05
0.10
0.15
0.20
0.25
0
1
2
3
4
5
6
7
Principal Component



0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
0
1
2
3
4
5
6
7

Chapter 5

Standards Adopted

5.1 Design Standards

5.1.1 Import Libraries and Load Data
Best Practice: Group all imports at the beginning of the script for better readability and maintenance.

Standardization: Use consistent naming conventions and imports.

5.1.2 Data Preparation
Code Organization: Clearly separate the data preparation steps, including merging, feature selection, and scaling.

Commenting: Add comments to explain the purpose of major steps.

5.1.3 Exploratory Data Analysis (EDA) and Visualization
Visualization Standards: Utilize meaningful color schemes and labels to enhance interpretability.

5.1.4 Principal Component Analysis (PCA)
Dimensionality Reduction: Apply PCA to reduce dimensions while capturing most variance.

Variance Explained Visualization: Plot the explained variance ratio to decide on the number of components.

5.1.5 Model Training and Evaluation
Balancing Dataset: Balance classes before model training if necessary.

Cross-Validation: Employ cross-validation to assess model performance more reliably.

Model Evaluation: Utilize confusion matrices, accuracy, and classification reports for comprehensive evaluation.

5.2 Coding Standards

Here are some coding standards and best practices that aimed at improving readability, maintainability, and performance:

Consistency:
-Consistency with naming conventions and code structure throughout the script.
-Using a consistent way of handling imports, even within the same library (e.g., from sklearn.model_selection import train_test_split).

Comments:
-Using comments to explain why certain decisions were made or to clarify complex parts of the code.
-Avoiding redundant comments that simply describe what the code is doing.

Import Statements:
-Placing all import statements at the beginning of the script.
-Grouping imports from the standard library, third-party libraries, and local modules separately.

5.3 Testing Standards

Unit Tests: Creating tests to ensure that the files /content/fma-rock-vs-hiphop.csv and /content/echonest-metrics.json exist and can be loaded without errors. Also, test the merging operation to ensure it results in the expected shape and column names.

Data Integrity Tests: Checking for NaN values, duplicate rows, and consistent data types across columns. Also, verifying that the track_id column is unique and serves as a reliable key for merging operations.

Correctness Tests: For PCA transformations, ensuring the variance explained by the components matches expectations.

Prediction Tests: Ensure that the model can make predictions on test data and that the output predictions have the correct format and dimension.

Accuracy Metrics: Implement tests to check that the accuracy, confusion matrix, and classification reports are correctly calculated and match expected ranges or values, considering the nature of your data and models.

Chapter 6

Conclusion and Future Scope

6.1 Conclusion

This project embarked on a journey to classify audio tracks into two distinct genres, hip-hop and rock, utilizing machine learning techniques. Through the application of Principal Compo nent Analysis (PCA) for feature reduction, followed by the employment of Decision Tree an d Logistic Regression models, we navigated the complexities of genre classification. The met hodology highlighted the importance of preprocessing steps, such as feature scaling and bala ncing the dataset, to ensure that our models had a solid foundation for training.

Our findings revealed that both models performed competently, yet they showcased distinct s trengths and weaknesses. The Decision Tree model provided an interpretable framework for understanding feature importance, whereas the Logistic Regression model offered a more ge neralized approach, potentially more adaptable to unseen data. Cross-validation techniques re inforced the reliability of our evaluation, presenting a comprehensive view of the models' ca pabilities across different subsets of the data.

6.2 Future Scope

Exploring Advanced Models

Future research could explore more sophisticated machine learning algorithms, such as ense mble methods (Random Forests, Gradient Boosting Machines) or deep learning architectures (Convolutional Neural Networks, Recurrent Neural Networks), which might capture the co mplexities of audio data more effectively.

Real-world Application and Testing

Future projects could focus on the real-world application of these models, such as integrating them into music recommendation systems or digital music libraries. Testing the models in li ve environments would provide valuable feedback on their performance and user satisfaction.

Multi-label Classification

Expanding the scope to multi-label classification could address the complexity of music genr es more realistically, where tracks may belong to multiple genres simultaneously. This appro ach would align more closely with the nuanced nature of music, offering a more detailed and user-friendly classification system.

References

Annesi, P., Basili, R., Gitto, R., Moschitti, A., & Petitti, R. (2007). Audio Feature Engineerin g for Automatic Music Genre Classification. RIAO, 702–711.

Bertin-Mahieux, T., Ellis, D. P., Whitman, B., & Lamere, P. (2011). The million song dataset in Proceedings of the 12th International Society for Music Information Retrieval Conferenc e. Miami, October, 24, 591–596. Chang, K. K., Jang, J.-S. R., & Iliopoulos, C. S. (2010). Mu sic Genre Classification via Compressive Sampling. ISMIR, 387–392.

Chathuranga, D., & Jayaratne, L. (2013). Automatic music genre classification of audio sig nals with machine learning approaches. GSTF Journal on Computing (JoC), 3, 1–12