MAR ATHANASIUS COLLEGE OF ENGINEERING, KOTHAMANGALAM

Initial Project Report

MUSIC GENRE IDENTIFICATION USING MACHINE LEARNING

Done by

AKHIL ROCK BABU

Reg No: MAC23MCA-2010

Under the guidance of

Prof. Manu John

ABSTRACT

The project aims to create a music genre classification system using machine learning algorithms, specifically k-Nearest Neighbors (KNN), Support Vector Machine (SVM), and Naive Bayes (NB), to accurately predict the genre of unseen audio files using the GTZAN dataset and Python Librosa package.

The music industry leverages genre classification for enhanced user experience, personalized content, and targeted audience targeting in recommendation engines. This system is crucial in organizing digital music libraries, enhancing user experience, and enhancing the effectiveness of music producers and marketers.

Machine learning algorithms have been found to be effective in classifying music genres. Studies have shown that SVM with an RBF kernel is accurate by 87.5%, while KNN outperforms other models with an accuracy of 92.69%, surpassing CNN's 72.40%. KNN's high accuracy and efficiency make it a leading algorithm for this task.

The GTZAN dataset will be used to evaluate three machine learning algorithms for identifying music genres: KNN, SVM, and NB. The KNN algorithm will be evaluated for accuracy, computational efficiency, and robustness. SVM will be assessed for high-dimensional data handling, while Naive Bayes will be considered for its rapid and probabilistic approach. The goal is to find the most accurate algorithm.

The dataset consists of 10,000 records of pre-extracted features from 1,000 audio tracks, each 30 seconds long, trimmed into 3-second segments. This results in a detailed and comprehensive dataset that captures the essence of each music genre. The dataset is balanced, with tracks categorized into 10 genres: blues, classical, country, disco, hip-hop, jazz, metal, pop, reggae, and rock. Features such as Mel Frequency Cepstral Coefficients (MFCCs), chroma features, zero-crossing rate, and tempo will be extracted using the Librosa package. These features will be normalized and used to train and evaluate the machine learning models, ensuring a comprehensive and reliable classification system.

References:

1. Ndou, N., Ajoodha, R., & Jadhav, A. (2021). Music Genre Classification: A review of Deep-Learning and Traditional Machine-Learning Approaches. 2021 IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS). https://doi.org/10.1109/iemtronics52119.2021.9422487

2. Ghildiyal, A., Singh, K., & Sharma, S. (2020). Music Genre Classification using Machine Learning. 2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA). https://doi.org/10.1109/iceca49313.2020.9297444

3. Setiadi, D. R. I. M., Rahardwika, D. S., Rachmawanto, E. H., Sari, C. A., Irawan, C., Kusumaningrum, D. P., Nuri, N., & Trusthi, S. L. (2020). Comparison of SVM, KNN, and NB Classifier for Genre Music Classification based on Metadata. 2020 International Seminar on Application for Technology of Information and Communication (iSemantic). https://doi.org/10.1109/isemantic50169.2020.9234199

Submitted By:

Akhil Rock Babu   Faculty Guide:

Reg No: M23CA011   Prof. Manu M John

2023 –25 Batch MCA Department, MACE   Associate Professor S3 MCA

INTRODUCTION

In the rapidly evolving landscape of digital music, the efficient categorization of music genres is pivotal for enhancing user engagement and optimizing content delivery. Music genre classification facilitates personalized recommendations, organizes extensive digital music libraries, and supports targeted marketing strategies. As such, developing a robust and accurate music genre classification system is essential for both users and industry professionals.

This project seeks to address this need by employing advanced machine learning techniques to classify music genres. Specifically, it explores the performance of three prominent algorithms: k-Nearest Neighbors (KNN), Support Vector Machine (SVM), and Naive Bayes (NB). The evaluation will be based on the GTZAN dataset, a well-regarded benchmark in music genre classification research, and the Python Librosa package, a powerful tool for audio analysis.

The GTZAN dataset comprises 10,000 pre-extracted feature records from 1,000 audio tracks, each 30 seconds in duration and segmented into 3-second intervals. This dataset encompasses 10 distinct genres: blues, classical, country, disco, hip-hop, jazz, metal, pop, reggae, and rock. Key features of interest include Mel Frequency Cepstral Coefficients (MFCCs), chroma features, zero-crossing rate, and tempo. These features provide a comprehensive representation of the audio data, capturing essential musical characteristics that are critical for genre classification.

Using the Python Librosa package, features from the GTZAN dataset will be meticulously extracted and normalized. This preprocessing step is crucial for ensuring the accuracy and reliability of the subsequent machine learning models. The KNN, SVM, and Naive Bayes algorithms will then be applied to this processed data, each evaluated for its effectiveness in predicting music genres.

Prior studies have demonstrated varying levels of success among these algorithms. SVM with an RBF kernel has achieved an accuracy of approximately 87.5%, showcasing its capability to handle complex data patterns. KNN, however, has consistently delivered superior performance with an accuracy of 92.69%, outperforming other models, including Convolutional Neural Networks (CNNs) which achieved an accuracy of 72.40%. This high accuracy makes KNN a particularly strong candidate for music genre classification. Naive Bayes, with its probabilistic approach, offers rapid classification capabilities, making it an attractive option for certain applications.

In this project, the KNN algorithm will be assessed for its precision, computational efficiency, and scalability. SVM will be evaluated for its performance with high-dimensional feature spaces, and Naive Bayes will be analyzed for its speed and probabilistic classification strengths. By comparing these algorithms, the project aims to identify the most effective method for genre classification, contributing to advancements in automated music analysis and enhancing the overall functionality of music recommendation systems.

The findings of this project will not only provide insights into the relative strengths of these machine learning algorithms but also offer practical recommendations for their application in real-world music classification tasks. Ultimately, the goal is to develop a system that significantly improves the accuracy and efficiency of music genre classification, thereby enhancing user experiences and supporting the broader music industry.

LITERATURE REVIEW

Paper 1: Music Genre Classification: A Review of Deep-Learning and Traditional Machine-Learning Approaches

The paper compares how well deep learning and traditional machine learning can classify different types of music. It uses traditional algorithms like K-Nearest Neighbors (KNN), Support Vector Machine (SVM), and Naive Bayes (NB) as well as deep learning algorithms like Convolutional Neural Networks (CNN). KNN was 92.69% more accurate than SVM and CNN. The review shows that KNN is very accurate but sensitive to noise, SVM is effective but computationally expensive, and CNN requires a lot of data and resources. The authors suggest studying hybrid models in the future. They suggest using standardized data and evaluation methods to make studies more comparable and reliable.

Title of the paper	Ndou, N., Ajoodha, R., & Jadhav, A. (2021). Music Genre Classification: A review of Deep-Learning and Traditional Machine-Learning Approaches. 2021 IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS). https://doi.org/10.1109/iemtronics52119.2021.9422487
Area of work	This review is about music information retrieval, especially genre classification.
Dataset	The review looks at different datasets, such as the GTZAN dataset, which is used to compare music genre classification models. The GTZAN file has 60 columns and 10000 entries. https://www.kaggle.com/sets/andradaolteanu/gtzan-dataset-music-genre-classification
Methodology/Strategy	The paper looks at different ways to classify music. This means talking about how to find features, like MFCCs and chroma features, and comparing how well different classification algorithms work.
Algorithm	Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Convolutional Neural Network (CNN)
Result/Accuracy	**KNN: 92.69%** CNN: 72.40% SVM :80.80%
Advantages	This compares different methods and shows how machine learning can help find and classify features more accurately with extracted features.
Future Proposal	Suggests more research into hybrid models that combine deep learning with traditional methods and explore more diverse datasets for better generalization.

Paper 2: Music Genre Classification using Machine Learning

The paper focuses on using machine learning algorithms for music genre classification. The authors used a dataset of audio tracks and extracted features to train their models. They explored the performance of algorithms such as Support Vector Machine (SVM) with a Radial Basis Function (RBF) kernel, K-Nearest Neighbors (KNN), and Random Forest (RF). The results showed that SVM with an RBF kernel achieved an accuracy of 87.5%. The paper discusses the benefits of using SVM for its high accuracy and ability to handle non-linear data, although it can be computationally expensive. The authors suggest future research should consider the integration of more advanced feature extraction techniques and the use of hybrid models to enhance classification performance.

Title of the paper	Ghildiyal, A., Singh, K., & Sharma, S. (2020). Music Genre Classification using Machine Learning. 2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA). https://doi.org/10.1109/iceca49313.2020.9297444
Area of work	This research falls under the domain of music information retrieval, specifically targeting the classification of music into genres using machine learning algorithms.
Dataset	The study utilizes the GTZAN dataset, which is a standard dataset for music genre classification containing 60 columns and 10000 entries. https://www.kaggle.com/datasets/andradaolteanu/gtzan-dataset-music-genre-classification
Methodology/Strategy	The methodology involves feature extraction from the audio tracks using techniques like MFCCs, chroma features, and others. The features that are extracted are used to train machine learning models to classify music genres.
Algorithm	Support Vector Machine (SVM), K-Nearest Neighbors (KNN) and Random Forest (RF).
Result/Accuracy	**SVM: 87.5%** RF: 69.6% KNN: 66.4%
Advantages	The study compares different machine learning algorithms, showing which ones work best for music genre classification.
Limitations	The paper mentions potential overfitting issues with more complex models and points out the limitations of the GTZAN dataset, which may not be large enough to capture all the nuances of music genres.
Future Proposal	For future work, the authors suggest using deep learning techniques that have been successful in other studies for figuring out music. They also recommend using larger and more diverse datasets to make the classification models more robust and general.

Paper 3: Comparison of SVM, KNN, and NB Classifier for Genre Music Classification based on Metadata

The paper compares the effectiveness of three machine learning algorithms—Support Vector Machine (SVM), K-Nearest Neighbors (KNN), and Naive Bayes (NB)—for music genre classification based on metadata. The authors used a specific dataset containing metadata of various music genres to train and evaluate these models. The results showed that SVM with a Radial Basis Function (RBF) kernel achieved an accuracy of 80%. KNN and NB also performed well, with KNN achieving competitive results but slightly lower accuracy than SVM. The paper highlights the strengths of SVM in handling high-dimensional data and its robustness, though it is computationally intensive. KNN's simplicity and effectiveness are noted, but it is sensitive to noise and outliers. NB is recognized for its efficiency and simplicity but is less accurate in this context. The authors propose further research into optimizing these algorithms and combining them with advanced feature extraction methods to improve classification accuracy.

Title of the paper	Setiadi, D. R. I. M., Rahardwika, D. S., Rachmawanto, E. H., Sari, C. A., Irawan, C., Kusumaningrum, D. P., Nuri, N., & Trusthi, S. L. (2020). Comparison of SVM, KNN, and NB Classifier for Genre Music Classification based on Metadata. 2020 International Seminar on Application for Technology of Information and Communication (iSemantic). https://doi.org/10.1109/isemantic50169.2020.9234199
Area of Work	The paper falls under the domain of music information retrieval, specifically focusing on genre classification based on metadata.
Dataset	The study uses metadata features extracted from Spotify music dataset from www.crowdai.org
Methodology/Strategy	The authors extracted metadata features and applied three classifiers—SVM, KNN, and NB—to classify music genres. They compared the performance of these classifiers to determine which one is most effective for this task.
Algorithm	Support Vector Machine (SVM) K-Nearest Neighbours (KNN) Naive Bayes classifier (NB)
Result/Accuracy	**SVM: 80%** KNN: 75.61% NB: 75.05%
Advantages	The study compares multiple classifiers to find the best one for music genre classification based on metadata.
Limitations	Using only metadata might not capture the full complexity of music genres. This limitation suggests that the classifiers might not work as well as they would with a more complete feature set.
Future Proposal	The authors suggest that future research could use metadata and audio features to improve classification accuracy by using the strengths of both types of data.

LITERATURE SUMMARY

	TITLE	DATASET	ALGORITHM	ACCURACY
PAPER 1	Ndou, N., Ajoodha, R., & Jadhav, A. (2021). Music Genre Classification: A review of Deep-Learning and Traditional Machine-Learning Approaches. 2021 IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS). https://doi.org/10.1109/iemtronics52119.2021.9422487	GTZAN dataset 10,000 Records of 60 features.	K-Nearest Neighbors (KNN) Convolutional Neural Network (CNN) Support Vector Machine (SVM)	**KNN: 92.69%** CNN: 72.40% SVM :80.80%
PAPER 2	Ghildiyal, A., Singh, K., & Sharma, S. (2020). Music Genre Classification using Machine Learning. 2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA). https://doi.org/10.1109/iceca49313.2020.9297444	GTZAN dataset 10,000 Records of 60 features.	Support Vector Machine (SVM) Random Forest (RF) K-Nearest Neighbors (KNN)	**SVM: 87.5%** RF: 69.6% KNN: 66.4%
PAPER 3	Setiadi, D. R. I. M., Rahardwika, D. S, Rachmawanto, E. H., Sari, C. A., Irawan, C., Kusumaningrum, D. P., Nuri, N., & Trusthi, S. L . (2020). Comparison of SVM, KNN, and NB Classifier for Genre Music Classification based on Metadata. 2020 International Seminar on Application for Technology of Information and Communication (iSemantic). https://doi.org/10.1109/isemantic50169.2020.9234199	The study uses metadata features extracted from Spotify music dataset from www.crowdai.org	Support Vector Machine (SVM) K-Nearest Neighbours (KNN) Naive Bayes classifier (NB)	**SVM: 80%** KNN: 75.61% NB: 75.05%

PROPOSED MODEL

Music Genre Classification Using k-Nearest Neighbors (KNN)

Introduction

Music genre classification is an important part of music information retrieval systems. It affects music recommendation engines, playlist automation, and music library organization. k-Nearest Neighbors (KNN) has shown great performance with short audio features.

Objective

To develop and optimize a music genre classification system using the k-Nearest Neighbors (KNN), Support Vector Machine (SVM), and Naive Bayes (NB) algorithms, and compare their accuracy and efficiency to demonstrate the superior performance of the best algorithm among the three.

Background and Motivation:

Recent research highlights the effectiveness of various algorithms in music genre classification, with some achieving higher accuracy than many traditional and deep learning models. For example, in the study by Ndou, Ajoodha, & Jadhav (2021), KNN achieved an impressive accuracy of 92.69%, significantly outperforming Convolutional Neural Networks (CNNs). This study aims to explore and compare the performance of three different algorithms: k-Nearest Neighbors (KNN), Support Vector Machine (SVM), and Naive Bayes (NB), in terms of their accuracy and efficiency in music genre classification.

Why Choose KNN, SVM, and NB:

- **High Accuracy:** KNN, SVM, and NB have all demonstrated high accuracy in various classification tasks, making them strong candidates for music genre classification.

- **Simplicity and Efficiency:** These algorithms are easy to implement and computationally efficient, making them suitable for real-time applications.

- **Effectiveness with Short-Duration Features:** KNN, SVM, and NB perform exceptionally well with short-duration audio features, which are crucial for timely and accurate genre classification.

Methodology:

1. Data Collection:

Use the GTZAN dataset and extract relevant audio features, such as Mel Frequency Cepstral Coefficients (MFCCs), chroma, and spectrograms, using the Python Librosa package, and collect the preprocessed feature data into a CSV file.

2. Model Development:

1. Choosing the Algorithms: Based on literature review, k-Nearest Neighbors (KNN), Support Vector Machine (SVM), and Naive Bayes (NB) algorithms are chosen due to their high accuracy and efficiency.

2. Training the Models: The KNN, SVM, and NB models are trained using the feature matrix and corresponding labels. Cross-validation is used to optimize the parameters for each algorithm.

3. Validation: The performance of each model is validated using 10-fold cross-validation to ensure robustness and avoid overfitting.

4. Hyperparameter Tuning: Each model is fine-tuned by adjusting hyperparameters to achieve the best performance.

## 3. Evaluation:

- Perform 10-fold Cross-Validation: This ensures the robustness and reliability of each model.

- Evaluate the Models: The models are evaluated based on accuracy, precision, recall, F1-score, and computational efficiency. The results are compared to determine the best-performing algorithm.

2. Prediction Process:

1. Input: An unseen audio file is input into the system.

2. Feature Extraction: Features are extracted from the input audio file using the same methods (Librosa) as during training.

3. Normalization: The extracted features are normalized to match the scale of the training data.

4. Model Prediction: The normalized feature vector is passed to the trained models (KNN, SVM, and NB). Each model makes a prediction based on its specific algorithm.

5. Class Label Assignment: The models predict the genre of the input audio file based on their respective algorithms:

   o KNN: Calculates the distance between the input vector and all training samples, identifying the k-nearest neighbors and predicting the genre based on the majority class among them.

   o SVM: Uses the optimized hyperplane to classify the input vector.

   o NB: Uses the probability distributions learned during training to classify the input vector.

PIPELINE DIAGRAM:
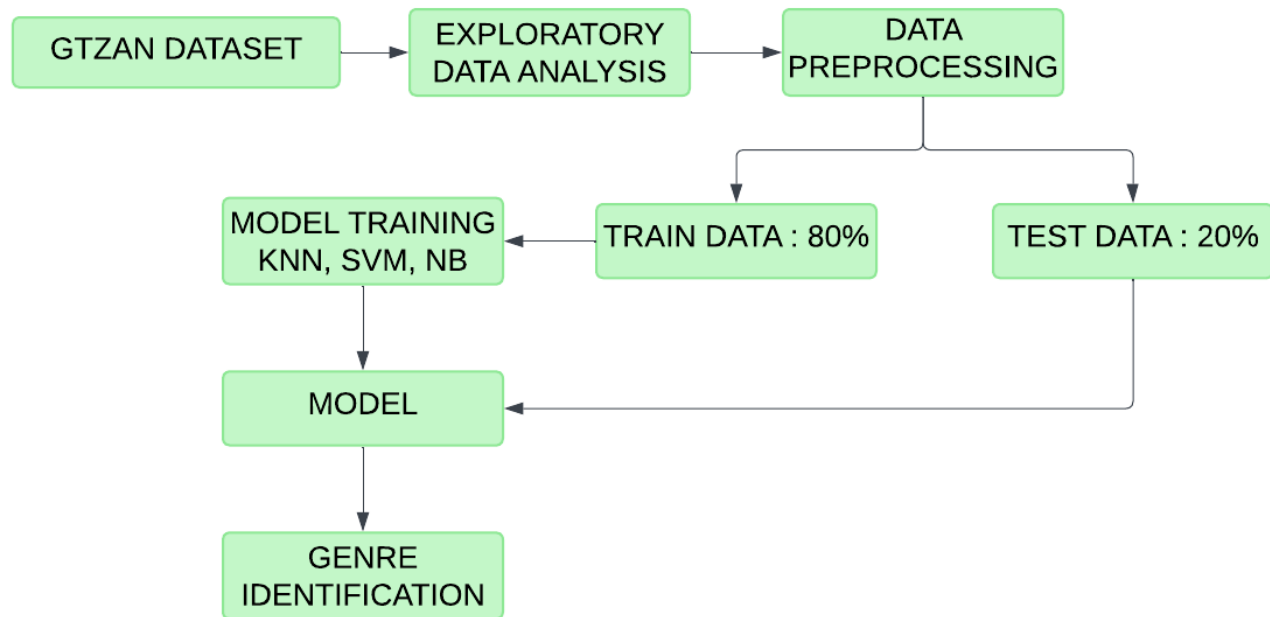
Conclusion:

This project leverages the GTZAN dataset and compares three algorithms—k-Nearest Neighbors (KNN), Support Vector Machine (SVM), and Naive Bayes (NB)—to develop a high-accuracy music genre classification system. By systematically extracting and normalizing relevant audio features, training robust models, and optimizing their parameters, the system is capable of accurately predicting the genre of new, unseen audio files. The combination of high accuracy, computational efficiency, and scalability makes this approach a powerful solution for music genre classification, with potential applications in music recommendation systems, automated playlist generation, and digital music libraries.

DATASET DESCRIPTION

Dataset Overview:

The GTZAN dataset is a widely used benchmark for music genre classification tasks. It contains 1,000 audio tracks each 30 seconds long, divided into 10 genres: blues, classical, country, disco, hiphop, jazz, metal, pop, reggae, and rock. Each genre has 100 tracks, making the dataset balanced and suitable for classification tasks.

Source:

The GTZAN dataset is publicly available on Kaggle and was originally compiled by George Tzanetakis in 2002. It is a go-to dataset for researchers and practitioners working on music genre classification.

Dataset Link: https://www.kaggle.com/datasets/andradaolteanu/gtzan-dataset-music-genre-classification

filename	length	chroma_stft_mean	chroma_stft_var	rms_mean
blues.00000.0.wav	66149	0.3354063630104065	0.091048293	0.13040502369403
blues.00000.1.wav	66149	0.3430653512477875	0.086146526	0.11269924789667
blues.00000.2.wav	66149	0.34681475162506104	0.092242889	0.13200338184833
blues.00000.3.wav	66149	0.3636387884616852	0.086856157	0.13256472349166
blues.00000.4.wav	66149	0.33557942509651184	0.088128544	0.14328880608081
blues.00000.5.wav	66149	0.3766697347164154	0.089702107	0.13261780142784

Features:

For music genre classification, various audio features can be extracted from the raw audio files. Using the Python Librosa package, we can extract the following key features:

- Chroma Feature (chroma_stft): 12 coefficients representing the energy distribution across the 12 different pitch classes.

- Root Mean Square Value (rms): Represents the power of the audio signal.

- Spectral Centroid (spectral_centroid): Indicates where the center of mass of the spectrum is located.

- Spectral Bandwidth (spectral_bandwidth): Measures the width of the band of frequencies.

- Spectral Contrast (spectral_contrast): The difference in amplitude between peaks and valleys in the sound spectrum.

- Spectral Rolloff (spectral_rolloff): The frequency below which a specified percentage of the total spectral energy lies.

- Zero Crossing Rate (zero_crossing_rate): The rate at which the signal changes sign.

- Harmony and Perceived Pitch (harmony, perceptr): Represent harmony and pitch features.

- Tempo (tempo): The estimated tempo of the music.

-MFCCs (mfcc1-mfcc20): 20 coefficients representing the Mel Frequency Cepstral Coefficients.

Class Labels:

The dataset is labeled with 10 distinct music genres:

1. Blues

2. Classical

3. Country

4. Disco

5. Hiphop

6. Jazz

7. Metal

8. Pop

9. Reggae

10. Rock