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

TOPIC: MUSIC GENRE IDENTIFICATION USING MACHINE LEARNING

Music Genre Identification is an interesting and challenging task in the field of machine learning and audio signal processing. This project aims to group music tracks into different types using Machine Learning Algorithms. With the amount of digital music available, Automated Genre Identification becomes more important for various applications, such as music recommendation systems, music libraries, and streaming services.

In this project, we use machine learning algorithms to classify music Genres based on the features extracted from audio files. The GTZAN dataset is used for training and evaluating. You can find it on Kaggle. Our system uses a raw audio file. Audio features are taken from the audio signals and are important because they show what the music is like and help classify it into different types of music.

We use three different machine learning algorithms to classify the Genres: K-Nearest Neighbors (KNN), Support Vector Machine (SVM), and Naive Bayes classifier (NB). Each model is trained on the features in the GTZAN dataset. The performance of these models is then compared to see which algorithm gives the best Identification. Once trained, the models can classify new music tracks into their respective Genres based on the extracted audio features. The system predicts the type of music for the song you put in. This project is important because it could help the music industry. A good Genre Identification can make music streaming services better by giving users personalized recommendations and organizing music libraries more efficiently. It can also help with musicological research by making it easier to group large collections of music.

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

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
Reg No: M23CA011
2023 –25 Batch MCA Department, MACE

Faculty Guide:
Prof. Manu M John
Associate Professor S3 MCA

LITERATURE REVIEW

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

This paper explains how to classify music genres using both deep learning and traditional machine learning methods. The study looks at the strengths and weaknesses of each approach and talks about how accurate they are and how useful they are.

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

This paper uses machine learning techniques to classify music genres. The study looks at algorithms and how well they can classify music into different genres.

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: 76.4%** 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

This paper compares the performance of three different classifiers for music genre classification based on metadata. The study wants to find out which program is the best at identifying different types of music using information about their history.

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: 76.4%** 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.

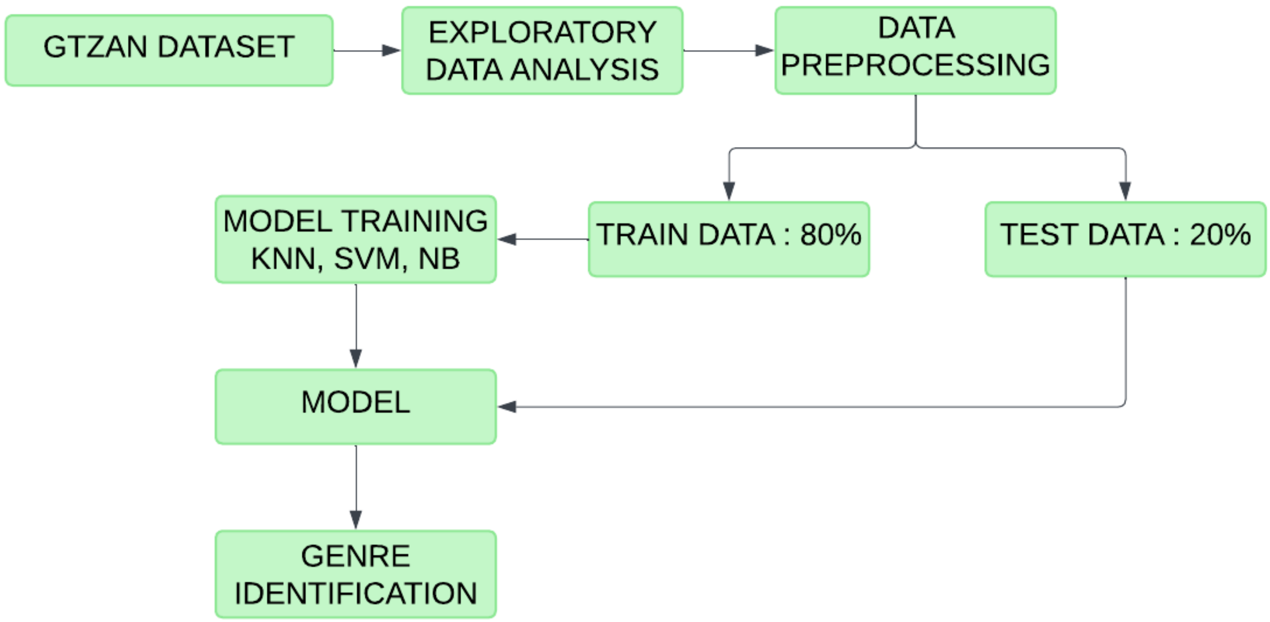
3. Evaluation:

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