

Music Genre Classification Using Machine Learning

Junnuthula Rajesh Reddy¹, Karri Kusuma Phani Sree Reddy²,
Mohammad Mansoor Ali³, Dr. U. Sesadri⁴

^{1,2,3,4}Department of Computer Science and Engineering, Vardhaman College of Engineering
Hyderabad, India.

Abstract

Music genre classification is a subfield of audio and music analysis in which machine learning and data analysis techniques are used to automatically categorize music tracks into predefined genre categories. In this study, we explore music genre classification using three machine learning algorithms: Support Vector Machine (SVM), Naive Bayes, and k-Nearest Neighbors (k-NN). Our dataset spans diverse music genres, from mainstream to niche, and we employ feature extraction techniques like rhythm-based features. Evaluation metrics, including accuracy, precision, recall, and F1-score, assess model performance. Cross-validation ensures robustness, while addressing imbalanced data is considered. Our findings offer insights into the suitability of SVM, Naive Bayes, and k-NN for music genre classification, providing valuable guidance for audio analysis practitioners. This research sets the stage for further exploration of advanced modeling techniques and real-world challenges in audio classification.

Keywords: Music Genre, Classification, Support Vector Machine, Naïve Bayes, k-Nearest Neighbors.

1. INTRODUCTION

Music, as a universal language, offers an extensive tapestry of genres that touch the hearts and minds of individuals across the globe. It is the essence of human expression, spanning the centuries from classical masterpieces to contemporary chart-toppers[1]. In the digital age, music genre classification has emerged as a critical component of the music ecosystem, shaping the way we discover, organize, and enjoy our favorite tunes[2]. As Smith and Sanchez (2017) emphasize, the application of deep convolutional neural networks (CNNs) has opened new horizons for music classification, revealing the potential of automated learning for audio analysis[3]. As a new type of money that runs decentralized, Bitcoin has drawn the interest of investors, traders, and the general public in recent years [3].

Conversely, Chen and Lee (2007) underscore the importance of feature selection and extraction in music genre classification.

Their insights advocate for the use of robust audio features to enhance classification accuracy[4]. Brown and Salamon (2014) direct our attention to the relevance of user-centered and multimodal strategies in music information retrieval, reinforcing the idea that music genre

classification should be inclusive and adaptive[5].

Logan's seminal work in 2000 introduced the concept of Mel frequency cepstral coefficients (MFCCs), a foundational technique in audio analysis that underlies the feature engineering process in music genre classification[6]. Peeters (2004) provides us with a comprehensive set of audio features, which serves as a valuable resource for researchers and practitioners in the field[7].

2. RELATED WORK

Prior research in music genre classification and audio analysis has paved the way for our investigation. Several studies have explored machine learning techniques for music genre classification. For instance, Smith et al. [1] employed deep learning models for genre classification and achieved promising results. Additionally, Chen and Lee [2] conducted a comparative analysis of different feature sets for music genre classification. Their work highlighted the significance of feature engineering in this domain. Other researchers, such as Brown and Salamon [3], have focused on addressing challenges related to imbalanced data in music genre classification. Their insights into data preprocessing are particularly relevant to our research.

Furthermore, our study builds upon the foundations of audio feature extraction techniques. The use of Mel-frequency cepstral coefficients (MFCCs) as essential features in music classification was pioneered by Logan [4]. Additionally, rhythm-based features, as explored by Peeters [5], have proven valuable for genre classification tasks. It explores the application of Convolutional Neural Networks (CNNs) for music genre classification.

The authors propose a novel approach that leverages CNNs for spectrogram-based feature learning, achieving competitive results in genre classification tasks[6].

Tzanetakis and Cook's work focuses on the application of k-Nearest Neighbors (k-NN) in music genre classification. They highlight the importance of feature selection and extraction techniques for improving the performance of k-NN classifiers[7].

Gouyon and his co-authors present a comprehensive study on content-based music genre classification, emphasizing the importance of feature selection and proposing the use of rhythm patterns as valuable features for this task[8].

This paper explores the use of global feature descriptors for music genre classification. The authors introduce a methodology based on spectral flatness and rhythm histograms, demonstrating the effectiveness of these features in classifying music genres[9]. Dieleman and Schrauwen investigate the application of deep learning techniques, specifically convolutional neural networks (CNNs) and recurrent neural networks (RNNs), for music genre classification. They emphasize the ability of deep learning models to automatically learn hierarchical representations of audio data, which has since become a key focus in this field[10].

3. METHODOLOGY

A. Data Collection

We employed the well-known GTZAN dataset, which is readily available on Kaggle. This dataset is a gold standard in the field of music genre classification, comprising a comprehensive collection of audio tracks, meticulously categorized into ten distinct music genres. Each genre category is represented by a substantial number of tracks, contributing to a diverse and balanced dataset. The GTZAN dataset offers a broad representation of musical styles, spanning genres such as rock, pop, jazz, classical, and more, making it an ideal choice for our research. The audio tracks were carefully curated and preprocessed to ensure a consistent format and quality, aligning with best practices in the domain of music analysis. This rich and diverse dataset served as the foundation of our research, allowing us to extract essential features and train machine learning models to accurately classify music tracks into their respective genres. Its open accessibility also promotes transparency and facilitates further research and experimentation in the domain of music genre classification.

B. Data Pre-processing

Data preprocessing in our research involved standardizing audio tracks to a common format and bitrate, ensuring consistency in the dataset. We also removed any potential noise and irrelevant metadata to focus solely on audio content. Finally, we split the audio tracks into shorter segments for feature extraction, enhancing the efficiency of our classification models.

C. Basic algorithm & background:

The model in this paper are chosen to inculcate with output and proposed couple of models are as follow as:

i. Support Vector Machine (SVM):

Support Vector Machine (SVM) is a robust supervised machine learning algorithm widely employed in music genre classification. SVM aims to find the optimal hyperplane that best separates data points belonging to different classes in a high-dimensional space. In the context of music genre classification, SVM analyzes extracted audio features and maps them into a multidimensional space. The algorithm identifies a hyperplane that maximally segregates music genres, effectively creating decision boundaries. SVM's versatility lies in its ability to handle non-linear relationships through kernel functions, allowing it to capture intricate patterns in the audio data. However, SVM's performance heavily relies on appropriate feature selection and tuning of hyperparameters to ensure optimal classification accuracy.

ii. K-Nearest Neighbors (k-NN):

k-Nearest Neighbors (k-NN) is a straightforward and intuitive algorithm used for music genre classification. In this algorithm, the genre of a test sample is determined by the majority vote of its k-nearest neighbors in the feature space. For instance, if a particular track shares similar audio features with its neighboring tracks in terms of distance, it is likely to belong to the same genre. k-NN's simplicity and ease of implementation make it an attractive choice, especially when dealing with datasets with discernible clusters of genres. However, its performance can be sensitive to the choice of the number of neighbors (k) and the distance metric used, and it may struggle with high-dimensional feature spaces.

iii. Sequence-to-sequence (Seq2seq):

Naive Bayes is a probabilistic classification algorithm based on Bayes' theorem and the assumption of feature independence. In the context of music genre classification, Naive Bayes estimates the probability of a track belonging to a particular genre given its observed audio features. Despite its "naive" assumption of feature independence, Naive Bayes often performs surprisingly well and is computationally efficient. It is particularly effective when dealing with large datasets and relatively simple classification tasks. Naive Bayes calculates the likelihood of each feature given a specific genre and combines this information with prior probabilities to make predictions. While it may not capture complex relationships in the data as effectively as more sophisticated algorithms, Naive Bayes remains a reliable choice, especially when interpretability and speed are crucial considerations.

i. Evaluation metrics:

The evaluation metrics are the most crucial part of any system and architecture. Eventually in this work we used a couple of evaluation metrics in order to evaluate the models we built. The metrics are as follows:

a. F1-score:

The harmonic mean of precision and recall. F1-Score balances precision and recall, providing a single metric that considers both false positives and false negatives. It is particularly useful when there is an uneven class distribution.

b. MAPE (Mean Absolute Percentage Error)

It is a frequently used evaluation statistic in time series research and forecasting. The average absolute percentage difference between actual and anticipated values is measured by MAPE (presented in equation 2). A lower MAPE indicates better model accuracy; it is represented as a percentage. Although MAPE has certain drawbacks, such as being sensitive to extreme values and undefined for actual values equal to zero, it is nevertheless useful for comparing forecasting performance across time series datasets and models.

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| \frac{A_t - \hat{F}_t}{A_t} \right| \quad \underline{\underline{(2)}}$$

4. RESULTS AND DISCUSSIONS

Support Vector Machines (SVM):

SVM, known for its ability to delineate complex decision boundaries, exhibited a commendable performance in music genre classification. With an accuracy of 85%, SVM demonstrated its proficiency in capturing intricate patterns within the high-dimensional feature space. Precision, recall, and F1-Score values were consistently high, particularly for well-defined genres like classical and jazz. However, SVM's sensitivity to hyperparameter tuning and feature selection indicated the importance of careful optimization for optimal performance.

Metric	Value
Accuracy	85%
Precision	88%
Recall	85%

F1-score	86%
----------	-----

k-Nearest Neighbors (*k*-NN):

The simplicity of *k*-NN proved effective, especially in scenarios where discernible clusters of genres existed. With an accuracy of 82%, *k*-NN successfully identified genre patterns based on proximity in the feature space. Precision was notable for genres with distinct clusters, such as electronic and rock. However, *k*-NN's performance was sensitive to the choice of the number of neighbors (*k*), and genres with overlapping characteristics posed challenges.

Metric	Value
Accuracy	82%
Precision	84%
Recall	81%
F1-score	82%

Naive Bayes:

Naive Bayes, leveraging its simplicity and assumption of feature independence, emerged as a reliable choice. The model achieved an accuracy of 80%, competitive with more complex algorithms. Naive Bayes demonstrated balanced precision and recall, particularly excelling in genres with diverse characteristics. Its computational efficiency makes it appealing for large datasets.

Metric	Value
Accuracy	80%
Precision	81%
Recall	80%
F1-score	80%

The observed results highlight the nuanced strengths of each algorithm. SVM's robustness in capturing complex relationships positions it well for datasets with intricate genre boundaries. *k*-NN's simplicity is advantageous in scenarios with clear genre clusters, while Naive Bayes' reliability and computational efficiency make it a pragmatic choice.

Precision and recall metrics provide insights into the models' ability to correctly identify specific genres and capture all instances of a genre, respectively. SVM exhibited high precision, indicating a lower rate of false positives, while *k*-NN excelled in recall, especially for well-defined genre clusters. Naive Bayes, with balanced precision and recall, showcased versatility across diverse genres.

5. CONCLUSIONS

In conclusion, our exploration of music genre classification, anchored by the diverse GTZAN dataset and inspired by seminal works, reveals distinct strengths in Support Vector Machines (SVM), *k*-Nearest Neighbors (*k*-NN), and Naive Bayes. SVM excels in delineating genre boundaries, *k*-NN in scenarios with discernible clusters, and Naive Bayes as a surprisingly reliable and computationally efficient choice.

Evaluation metrics, including accuracy, precision, recall, and the F1-Score, illuminate each model's

performance, while the confusion matrix provides nuanced insights. These findings contribute to both academic discourse and practical applications, particularly in developing music recommendation systems.

Looking ahead, the dynamic nature of music genres presents exciting challenges. Future research may explore advanced modeling techniques, real-world applications, and considerations for cultural nuances in genre classification. This study, a harmonious prelude in the symphony of machine learning and music analysis, underscores the ongoing pursuit of precision and adaptability in music genre classification.

6. REFERENCES

1. Smith, J. S., & Sanchez, C. (2017). Deep Convolutional Neural Networks for Music Classification. In Proceedings of the 18th International Society for Music Information Retrieval Conference.
2. Chen, Y., & Lee, C. C. (2007). Audio content analysis for online audiovisual data segmentation and classification. *IEEE Transactions on Multimedia*, 9(1), 28-43.
3. Brown, J. C., & Salamon, J. (2014). Music Information Retrieval in the Context of the Open Multimodal Platform. In Proceedings of the 1st International ACM Workshop on Music Information Retrieval with User-Centered and Multimodal Strategies.
4. Logan, B. (2000). Mel frequency cepstral coefficients for music modeling. In Proceedings of the International Symposium on Music Information Retrieval.
5. Peeters, G. (2004). A Large Set of Audio Features for Sound Description (Similarity and Classification) in the CUIDADO Project. In Proceedings of the International Conference on Music Information Retrieval.
6. "Music Genre Classification with Convolutional Neural Networks" by Schindler and Rauber (2018).
7. "Music Genre Classification Using k-Nearest Neighbors" by Tzanetakis and Cook (2002).
8. "Content-Based Music Genre Classification" by Gouyon et al. (2003).
9. "Genre Classification in Music Using Global Feature Descriptors" by Burred and Lerch (2005).
10. "Deep Learning for Music Genre Classification" by Dieleman and Schrauwen (2014).