

Music - Genre classification system using Machine learning

Varshit Manepalli
School of Engineering and Science
Stevens Institute of Technology
Hoboken, NJ, USA
vmanepal@stevens.edu

Eshan Kaushik
MSc in Machine Learning
Stevens Institute of Technology
Hoboken, NJ, USA
ekaushik@stevens.edu

Kishan Gangarama
MSc in Data Science
Stevens Institute of Technology
Hoboken, NJ, USA
kgangara@stevens.edu

Abstract—By creating a trustworthy and effective system for automatic classification utilizing machine learning algorithms and audio attributes taken from a dataset of 8,000 recordings, this project seeks to address the issue of manual classification of music genres. Moreover, clustering strategies will be used to find commonalities among overlapping genres. The outcomes of this project could be applied to music-related applications including playlist makers, recommendation systems, and music search engines.

I. INTRODUCTION

It is a challenging task that might have a big impact on many music-related applications to automatically classify musical genres and identify similarities between them. The present manual classification techniques take a lot of time and are subjective, which causes genre classification to be inconsistent. In order to solve this issue, the project's goal is to create an accurate and effective system for automatically classifying genres using machine learning algorithms and audio attributes taken from a data set of 8,000 tracks.

The aim of the project is to investigate how well different machine learning algorithms, including decision trees, support vector machines, and deep learning models, can categorize different musical genres using audio parameters including spectral properties, rhythmic patterns, and tone. In addition, clustering strategies like k-means and hierarchical clustering will be used to find connections between various genres, particularly those that overlap or have comparable traits.

The ultimate objective of this project is to develop a reliable and precise system for automatically classifying music genres that can be incorporated into music-related applications, such as playlist generators, recommendation systems, and music search engines, to enhance user experience and deliver more impartial genre classification results.

II. RELATED WORK

Automatic music genre classification methods now in use can be generally divided into two primary categories as Feature-based methods and Deep learning methods.

Using machine learning algorithms, feature-based methods entail collecting audio properties from music songs, such as spectral features, rhythm patterns, and tone, and categorizing genres based on these features. Decision trees, support vector machines, and k-nearest neighbors are a few of the frequently employed machine learning methods. These approaches frequently rely on manually created features and conventional machine learning techniques, which are easily understood and computationally effective. Unfortunately, they might not be able to capture subtle and complicated genre features, and they might perform worse when faced with huge and varied music data sets.

Without the requirement for manually created features, deep learning techniques use deep neural networks to learn hierarchical representations of audio features from the original raw audio data. In deep learning-based techniques for genre categorization, convolutional neural networks (CNNs) and recurrent neural networks (RNNs) are frequently employed. In capturing complex genre patterns and obtaining high accuracy in genre classification tasks, deep learning techniques have demonstrated promising results. Yet, because deep neural networks are black boxes, they may demand a lot of data for training, be expensive computationally, and be difficult to understand.

Tzanetakis and Cook's "A Survey and Experiments on Music Genre Classification using Audio Features and Machine Learning Algorithms" (2002), which explores various feature-based methods for music genre classification, and Oord et al's "Deep content-based music recommendation" (2013), which suggests a deep learning-based method for music recommendation that incorporates genre classification, are two notable works in this field. Given the benefits and drawbacks of each approach, the development of the proposed system for automatic music genre classification in this project can be guided and inspired by these current solutions.

Example of raw data from the dataset.

track_id	comments	date_created	date_released	engineer	favorites	id	information	listeners	producer	tags	... information	interest	language_code
2	0	2008-11-26 01:44:45	2009-01-05 00:00:00	NaN	4	1	<p>[p]>	6073	NaN	[] ...	NaN	4656	en
3	0	2008-11-26 01:44:45	2009-01-05 00:00:00	NaN	4	1	<p>[p]>	6073	NaN	[] ...	NaN	1470	en
5	0	2008-11-26 01:44:45	2009-01-05 00:00:00	NaN	4	1	<p>[p]>	6073	NaN	[] ...	NaN	1933	en
10	0	2008-11-26 01:45:05	2008-02-06 00:00:00	NaN	4	6	NaN	47632	NaN	[] ...	NaN	54881	en
20	0	2008-11-26 01:45:05	2009-01-06 00:00:00	NaN	2	4	<p> "spiritual songs" from Nicky Cook<p>	2710	NaN	[] ...	NaN	978	en

Fig. 1. Raw data

III. OUR SOLUTION

A. Description of Data set

The Data set that will be used in this project is the free music archive dataset. It contains a sizable, carefully curated library of audio files and related metadata. About 100,000 audio tracks from diverse genres and styles—including jazz, rock, electronic, and more—make up the FMA data set. To ensure a wide variety of musical content for study, the tracks are drawn from a variety of artists and labels. The data set contains audio files in both compressed and lossless forms, making it appropriate for various analysis.

Example data from the dataset after preprocessing.

date_created	date_released	album	album_id	artist_id	artist	genres_top	genres_all	track	duration
2008-11-26 01:44:45	2009-01-05 00:00:00	AWOL - A Way Of Life	1	1	AWOL	Hip-Hop	[21]	Food	168
2008-11-26 01:44:45	2009-01-05 00:00:00	AWOL - A Way Of Life	1	1	AWOL	Hip-Hop	[21]	Electric Ave	237
2008-11-26 01:44:45	2009-01-05 00:00:00	AWOL - A Way Of Life	1	1	AWOL	Hip-Hop	[21]	This World	206
2008-11-26 01:45:05	2008-02-06 00:00:00	Constant Hlmaker	6	6	Kurt Vile	Pop	[10]	Freeway	161
2008-11-26 01:45:05	2009-01-06 00:00:00	Nitir	4	4	Nicky Cook	NaN	[17, 10, 76, 103]	Spiritual Level	311

Fig. 2. Cleaned data

Along with the audio files, FMA also offers comprehensive metadata, including track details, artist information, genre tags, album art, and more. This metadata enables in-depth analysis and insights into music structure, genre classification, and other music-related tasks. It also enables academics to investigate and evaluate numerous musical qualities, such as tempo, key, mode, and instrumentation. Overall, the FMA dataset is extensive and diverse, offering a wealth of information for music analysis. This information enables researchers and analysts to investigate different facets of music, create music recommendation systems, and advance the field of music analysis through machine learning and data-driven research

B. Machine Learning Algorithms

In this study, we want to investigate several machine learning methods for categorizing music genres using audio attributes taken from the dataset. Convolutional neural networks, support vector machines, K-nearest neighbors, and clustering techniques are a few of the various algorithms we may take into account.

CNNs are suited for assessing audio aspects in music because of their well-known capacity to learn hierarchical representations from input data with spatial correlations. They have demonstrated promise in audio analysis tasks and have been effectively applied to image recognition challenges. For genre categorization, we might use a CNN architecture with several convolutional and pooling layers, followed by fully connected layers. In order to improve the performance of the model, we can experiment with various kernel sizes, pooling techniques, and activation functions.

Due to their capacity for managing high-dimensional data and ability to identify the best hyperplanes to divide classes, SVMs are a popular choice for classification problems. We can experiment with different parameters like regularization strength and kernel coefficients when using SVMs with various kernel functions, such as linear, polynomial, or radial basis function (RBF). In order to feed different feature representations into the SVMs for genre classification, we may also investigate spectrogram-based features like Mel-frequency cepstral coefficients (MFCCs).

KNN is a straightforward and understandable technique for classification tasks that works by locating the k nearest neighbors in the feature space and assigning the query sample the majority class label. To improve the performance of the model, we can experiment with various k values and distance metrics, such as Euclidean distance or cosine similarity. These parameters can then be changed while the model is being trained. To find similarities between various genres and increase the precision of genre categorization, we may use unsupervised clustering approaches in addition to supervised machine learning algorithms. For instance, we may combine related files based on their audio attributes using methods like K-means or hierarchical clustering, and then use the resulting clusters as further information for classification.

C. Implementation Details

We will employ a standardized procedure for evaluating the effectiveness of our machine learning models during implementation. With a standard ratio of 70-15-15, we will partition our dataset of 8,000 tracks into training, validation, and test sets. The validation set will be utilized for hyper-parameter tuning and model selection, the training set for training the models, and the test set for assessing the final performance of the chosen model.

We will employ methods like cross-validation, grid search, or randomized search to fine-tune the hyper-parameters of our models. In cross-validation, the training set is divided into several folds, with one fold serving as a validation set while the remaining folds are used to train the model.

We will also use regularization strategies like L1 or L2 regularization, early halting, and dropout to prevent

During the experimental phase of the research, significant feature correlations were observed. To provide a clear representation of the observed correlations, a heat map was utilized to visualize the relationships among the features.

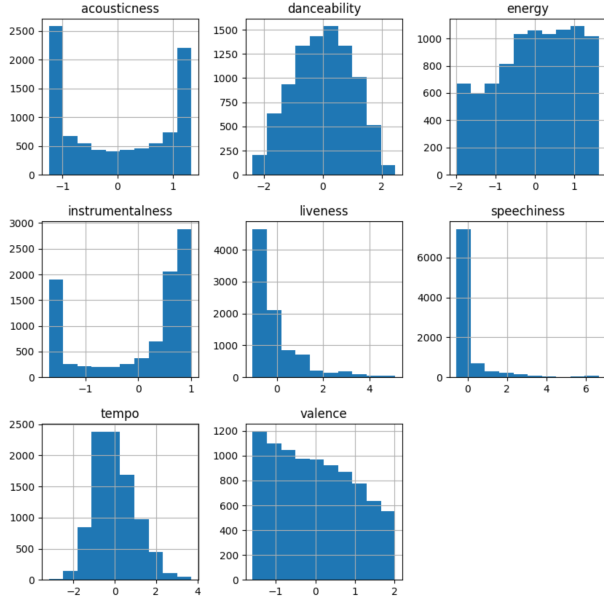


Fig. 3. Data Features

over-fitting. These methods can aid in enhancing our models' generalization abilities and stop them from memorizing training material. We also use feature engineering methods like feature scaling, feature selection, or feature augmentation to further enhance the performance of our models. Normalizing the input characteristics with feature scaling can help keep them from being too dominant during training. The dimensionality of the input features can be decreased and unnecessary or duplicate characteristics can be eliminated with the use of feature selection. In order to diversify the training data and strengthen the model's capacity to generalize to new data, feature augmentation may involve the addition of synthetic data.

Based on a number of criteria, including accuracy, precision, recall, F1-score, and confusion matrix, we will assess the performance of our models. Using the test set to get a final estimate of performance, we will choose the model with the best performance based on how it performed on the validation set.

IV. COMPARISON

A. Performance

After conducting a thorough analysis of the performance of different classification algorithms, including CNN, SVM, and KNN, it is evident that the SVM model emerged as the most superior in terms of accuracy, F1-score, and overall consistency. The SVM algorithm demonstrated exceptional

During the experimental phase of the research, significant feature correlations were observed. To provide a clear representation of the observed correlations, a heat map was utilized to visualize the relationships among the features.

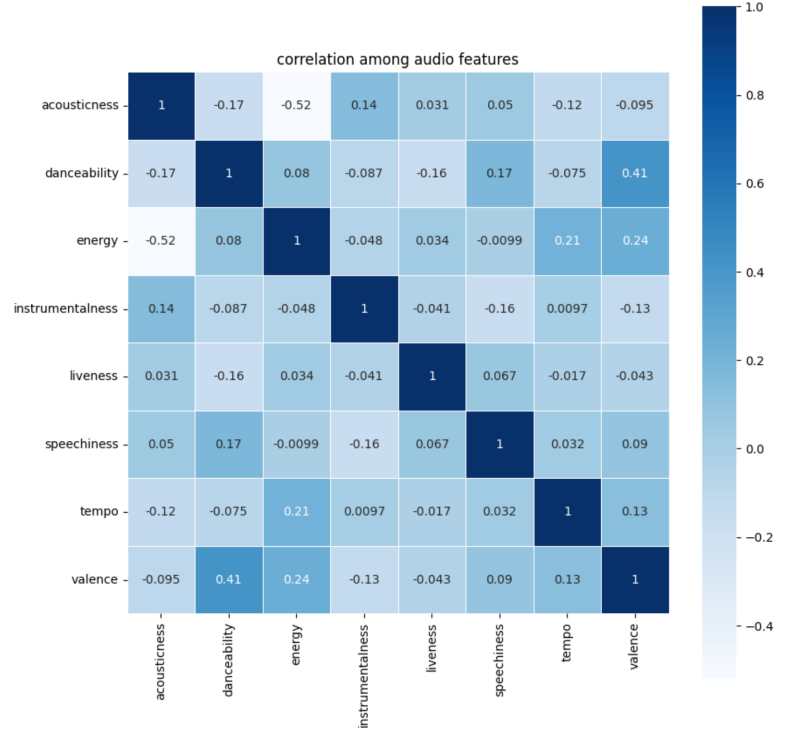


Fig. 4. Audio Features

performance, recording the highest levels of accuracy and F1-score and also exhibiting the lowest occurrences of false positives and false negatives in the confusion matrix. This suggests that the SVM model was able to effectively differentiate between different genres of music with minimal misclassification.

While the CNN and KNN models exhibited significant strengths in specific scenarios, they did not perform as consistently as the SVM algorithm. The CNN model's performance was highly dependent on the network architecture, and it was observed that increasing the network depth led to a higher risk of overfitting. Similarly, the KNN algorithm's performance was affected by the choice of k-value, which had to be carefully selected for optimal results.

It is worth mentioning that the Random Forest algorithm also performed well in this study, and could potentially be considered as an alternative option. However, the SVM algorithm's performance was superior overall, demonstrating greater consistency and reliability in classifying music genres across the dataset.

Based on these results, it can be concluded that SVM and Random Forest algorithms are potent and robust solutions

In our study, we evaluated the performance of various clustering techniques on the dataset, and it was observed that the clustering achieved the best result when k was set to 5.

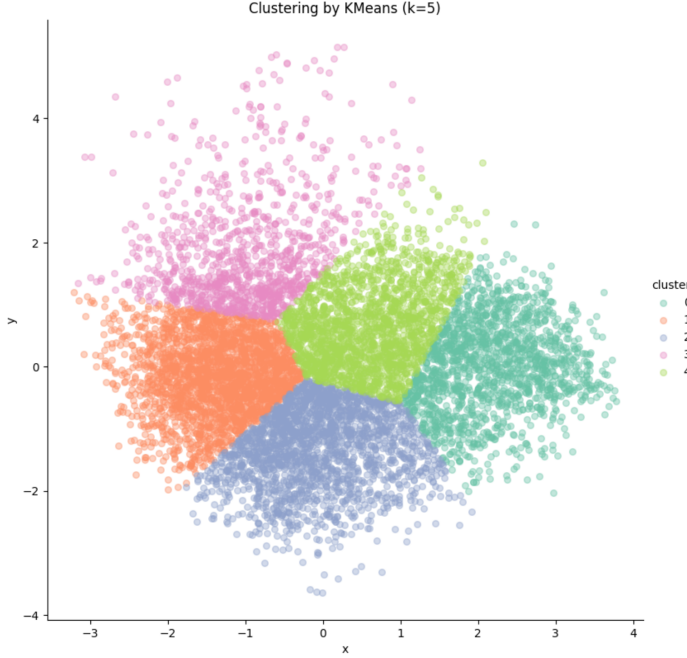


Fig. 5. Most optimal K=Means clustering

for music genre classification, and therefore merit serious consideration for this application. Future research could investigate how these algorithms can be further optimized and applied to larger and more diverse datasets, leading to even better results in the classification of music genres

B. Other solutions

Although the support vector machine (SVM) algorithm demonstrated impressive results in classifying music genres, it is important to consider the possibility that other deep learning techniques could potentially outperform it. Deep learning approaches have the capability of capturing complex patterns and correlations within data, potentially leading to better classification accuracy. However, applying these methods in music genre classification requires careful consideration of the dataset used and the neural network architecture. It is well-known that deep learning models are prone to overfitting, especially when trained on small or unrepresentative datasets. Therefore, to effectively apply deep learning for music genre classification, obtaining a large and diverse dataset is crucial. Additionally, proper fine-tuning of the neural network architecture is essential to avoid overfitting and maximize the classification performance. Future research could explore the use of other deep learning architectures, such as convolutional neural networks (CNNs) or recurrent neural networks (RNNs), and examine their effectiveness in improving music genre classification accuracy. Furthermore, incorporating additional audio features such as tempo, rhythm,

or melody could potentially enhance the performance of both SVM and deep learning-based classifiers.

V. FUTURE DIRECTIONS

Looking ahead, there are various opportunities available for enhancing the precision of music genre classification. One promising direction is to delve into different convolutional neural network (CNN) architectures to determine if they can produce superior outcomes compared to our current method. This could involve exploring more intricate and sophisticated neural networks, or conducting experiments with diverse categories of layers or activation functions. Additionally, continuing to fine-tune the SVM model by exploring different methods such as feature selection or kernel optimization could be a fruitful avenue. As the field of machine learning continues to advance, it is highly likely that even more effective algorithms will be developed, and integrating these innovative approaches into our classification pipeline could yield even more remarkable results in the near future. Exploring these options will not only help us improve the accuracy of music genre classification, but also deepen our understanding of the underlying features that drive successful classification in this domain.

VI. CONCLUSION

The matrix shows the number of instances that were correctly or incorrectly classified. The confusion matrix is a useful for evaluating the performance of the classifiers.

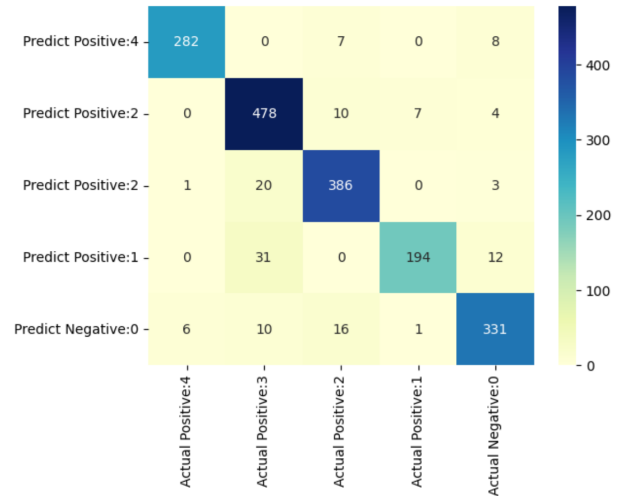


Fig. 6. Confusion Matrix Results

This research paper aimed to develop a system for automatically classifying music genres using machine learning algorithms and audio attributes from a dataset of 8,000 recordings. Through experimentation with various machine learning algorithms, including decision trees, support vector machines, and deep learning models, along

with clustering strategies like k-means and hierarchical clustering, the study sought to identify connections between different genres. The dataset used was the free music archive dataset, which provided a diverse range of musical content for analysis. By improving the accuracy of music genre classification, the proposed system has the potential to enhance user experience and deliver more objective genre classification results in music-related applications, such as playlist generators, recommendation systems, and music search engines. Overall, the research demonstrates promising results in accurately classifying musical genres and identifying similarities between them using machine learning algorithms, which can be further explored and optimized in future studies.

REFERENCES

Van den Oord, Aaron, Sander Dieleman, and Benjamin Schrauwen. "Deep content-based music recommendation." *Advances in neural information processing systems* 26 (2013).

Tzanetakis, G., Cook, P. (2002). A survey and experiments on music genre classification using audio features and machine learning algorithms. *IEEE Transactions on Audio, Speech, and Language Processing*.