

EXECUTIVE SUMMARY

Music Genre Classification Model: Empowering Data-Driven Music Curation

The music industry relies on personalized experiences to retain audiences in a competitive streaming landscape. Accurate genre classification is essential for improving track recommendations, curating playlists, and enhancing listener satisfaction, benefiting streaming platforms, record labels, and artists. This project developed a scalable model to classify music tracks into predefined genres using detailed audio features such as Mel-frequency cepstral coefficients (MFCCs), spectral centroid, and rhythmic patterns. By evaluating machine learning algorithms, including K-Nearest Neighbors (KNN), Random Forest, and Convolutional Neural Networks (CNN), the project identified the most effective approach for genre prediction. The model delivers high accuracy and reliability, offering significant value for content curation, personalized recommendations, and market insights.

Data Overview

The dataset employed in this project is the GTZAN Genre Collection, a widely recognized resource in music genre classification research. It comprises 1,000 audio tracks, each with a duration of 30 seconds, systematically categorized into 10 distinct genres. Each genre contains 100 tracks, ensuring balanced representation across categories.

Genres Included: Blues, Classical, Country, Disco, Hip-Hop, Jazz, Metal, Pop, Reggae, and Rock.

Dataset Composition

The dataset includes:

- **Raw Audio Files:** 1,000 audio tracks in WAV format, each 30 seconds long, sampled at 22,050 Hz in mono.
- **Extracted Features Dataset:** Pre-computed numerical features for each track, encapsulating essential audio characteristics such as tempo, timbre, and harmonic content.

Data Preprocessing

Preprocessing involved handling missing values, normalizing continuous variables, and extracting features to capture each track's unique characteristics. Key features were engineered from the raw audio files to enhance classification accuracy, focusing on their relevance to genre differentiation.

Feature Engineering

Librosa was used to engineer the features for this model. The library extracts audio features by applying signal processing techniques to convert raw sound data into numerical representations. The key features were selected to capture the timbral, rhythmic, and harmonic properties of the audio signals.

Feature	Description
Chroma_STFT	Energy distribution across 12 pitch classes, capturing harmonic structure
RMS (Root Mean Square Energy)	Reflects audio signal intensity, useful for detecting changes in dynamics
Spectral Centroid	Represents the "center of mass" of the spectrum, indicating brightness
Spectral Bandwidth	Width of the spectrum band, distinguishing fuller sounds from narrower ones
Zero Crossing Rate	Rate of sign changes in the signal, distinguishing smooth from noisy textures
Tempo	Speed of the track, crucial for identifying rhythmic patterns
MFCCs (Mel-Frequency Cepstral Coefficients)	Captures timbral characteristics through spectral shape analysis

For each extracted feature, both the **mean** and **variance** were calculated to capture complementary aspects of the feature's behavior. A detailed breakdown of the constructed dependent variables can be found in the [Appendix](#) section.

Methodology

To classify music tracks into genres, three machine learning models were developed and evaluated: **K-Nearest Neighbors (KNN)**, **Random Forest**, and **Convolutional Neural Networks (CNN)**. Each model was assessed using accuracy, precision, recall, and F1-score to ensure a robust evaluation of their performance.

Model	Accuracy	Precision	Recall	F1-Score
KNN	91.99 %	0.92	0.92	0.92
Random Forest	88.29 %	0.88	0.88	0.88
CNN	56.11 %	0.54	0.56	0.54

Model Evaluation

For a music genre classification model, achieving fewer wrong predictions (minimizing both false positives and false negatives) is critical to maintaining user satisfaction. Misclassifications can lead to inaccurate recommendations, reducing user engagement and trust. In this context:

- **Precision:** Ensures recommended tracks align with the expected genre.
- **Recall:** Prevents relevant tracks from being overlooked during classification.
- **F1-Score:** Balances precision and recall, making it the most relevant metric for ensuring overall reliability in genre classification systems.

Model Selection

The K-Nearest Neighbors (KNN) model was selected as the final model due to its superior performance across all metrics, achieving the highest accuracy (91.99%) and F1-score (0.92). Its straightforward implementation and high predictive capability make it well-suited for this use case, particularly given the size and structure of the dataset.

Optimized Model Equation

The KNN model uses the Euclidean distance metric to classify tracks by identifying the majority genre among the closest k neighbors in the feature space. The optimal value of k was determined to be 1 through hyperparameter tuning using grid search, ensuring the model's efficiency and accuracy.

Key Findings

Top Features Identified by KNN

Analysis revealed that certain audio features play a pivotal role in genre classification.

1. **Percussive Mean:** Highlights rhythmic elements, essential for genres like hip-hop.
2. **MFCC4 Mean & MFCC3 Mean:** Encodes tonal and textural variations critical for differentiating genres like classical or jazz.
3. **Harmony Variance & Percussive Variance:** Reflect the balance and dynamics within tracks, providing insights into the complexity of certain genres.

These findings emphasize the importance of rhythmic and timbral properties in distinguishing genres and demonstrate the ability of the model to capture subtle audio differences.

Complex Genre Relationships

Some genres with overlapping characteristics, such as jazz and blues or reggae and rock, required more nuanced modeling techniques. This highlights the potential for exploring hierarchical or multi-label classification to further enhance the model's performance.

Data Quality Impact

The model's strong performance was largely dependent on the quality of the dataset. The use of well-labeled, balanced data combined with robust preprocessing and feature selection processes ensured consistent and accurate classification results. These steps are critical for achieving high reliability in real-world applications.

Learning Curve

The learning curve for the optimized KNN model demonstrated consistent improvement in cross-validation accuracy as the size of the training data increased. This confirms that the model is not overfitting and highlights its scalability. The analysis suggests that further performance gains are possible with larger datasets, reinforcing the model's adaptability for broader industry applications.

Business Implications and Scope

The genre prediction model offers transformative opportunities for the music industry, enhancing efficiency, personalizing user experiences, and driving strategic decisions. Its scalability and accuracy make it valuable across streaming platforms, record labels, media synchronization, and analytics. Key applications include:

1. **Personalized Music Recommendations:** Streaming platforms can refine recommendation systems by analyzing listening patterns and introducing diverse genres, increasing user engagement and retention.
2. **Targeted Genre Identification:** Record labels and artists can align new releases with suitable audiences, optimizing marketing strategies and ensuring content resonates with listeners.
3. **Automated Playlist Curation:** The model enables dynamic, genre-based playlist creation, saving time while offering more personalized and granular listening options.
4. **Efficient Licensing and Synchronization:** Media supervisors can quickly locate genre-specific tracks, streamlining music selection for films, advertisements, and other projects.
5. **Tracking Genre Trends:** Platforms and labels can monitor evolving genre preferences, investing strategically in rising trends and tailoring campaigns for maximum audience impact.
6. **Audience Demographic Matching:** By combining genre predictions with demographic data, marketers can design highly targeted campaigns to reach specific listener groups effectively.

Conclusion

This genre prediction model empowers the music industry to automate workflows, improve user experiences, and adapt to market trends. By integrating this tool, businesses can enhance operational efficiency, deliver personalized content, and make informed, data-driven decisions in an evolving and competitive landscape.

APPENDIX

Feature	Description
label	Categorical variable indicating the genre of the track
mfcc1_mean to mfcc20_mean	Mean values of the 20 Mel-Frequency Cepstral Coefficients (MFCCs), capturing timbral textures
mfcc1_var to mfcc20_var	Variance values of the 20 MFCCs, highlighting dynamic tonal variations
chroma_stft_mean	Mean of energy distribution across 12 pitch classes, representing harmonic structure
chroma_stft_var	Variance of energy distribution across pitch classes, capturing fluctuations in harmony
rms_mean	Mean Root Mean Square (RMS) energy, indicating average signal intensity
rms_var	Variance of RMS energy, reflecting dynamic changes in intensity
spectral_centroid_mean	Mean of the spectral centroid, representing the average brightness of the audio
spectral_centroid_var	Variance of the spectral centroid, highlighting shifts in brightness over time
spectral_bandwidth_mean	Mean of spectral bandwidth, describing the average width of the audio spectrum
spectral_bandwidth_var	Variance of spectral bandwidth, capturing tonal richness changes
rolloff_mean	Mean of spectral rolloff, representing the frequency where most energy is concentrated
rolloff_var	Variance of spectral rolloff, indicating energy distribution changes
zero_crossing_rate_mean	Mean zero crossing rate, showing the average rate of signal sign changes (smooth vs noisy).
zero_crossing_rate_var	Variance of zero crossing rate, identifying changes in noisiness across the signal

harmony_mean	Mean of harmonic components in the signal, reflecting tonal balance
harmony_var	Variance of harmonic components, capturing variations in tonal structure
perceptr_mean	Mean of percussive components, emphasizing rhythmic elements in the track
perceptr_var	Variance of percussive components, capturing fluctuations in rhythm
tempo	The tempo of the track, measured in beats per minute (BPM)