Machine Learning Classification Techniques Applied in modern-day Music Recommendation Systems

Benjamin Hsin *

Shanghai American School (Puxi Campus), Shanghai, China

* Corresponding Author Email: Benjamin01px2024@saschina.org

Abstract. Music streaming platforms have increased in demand over recent years.Due to the covid-19 pandemic and lockdowns, there has been a substantial increasein users in music streaming platforms, where the music streaming market grew by $7.47 billion. Hence, the importance of recommendation systems in these platforms. The goal of this research paper is to address and provide insights into the behind the scenes to recommendation systems. To have the most accurate results, this paper will be based on pre-existing GTZAN audio samples as datasets for different classifiers and compare the results of accuracy and consistency of the model's outcome for each genre. Confusion matricesare the finest way to determine results from each classifier. To identify limitations, advantages, and whether or not each classifier provides the most ideal outcomes that satisfy user demands. This process will include processes such as the sample input, feature extraction, and finally, put through classifiers.

Keywords: K-nearest-neighbors, Convolutional Neural Network, support vector machine, Decision Trees, Feature Extraction, Music Classification.

1. Introduction

In today's society, personalized recommendation systems are becoming more demanding for the benefit of companies and users' experience. Because of this matter, I have taken the opportunity to explore the in-depth machine learning processes that allow music streaming platforms to implement the personalized recommendation of music to users [1]. Music recommendation systems are a powerful tool/feature to music streaming platforms such as Spotify, which is why since 2019, they have improved their systems which gave them leverage over other platforms. This research mentions different classifiers used in machine learning algorithms that sort music pieces into genres, having the end goal of allowing a better-personalized user experience. This includes features such as: being able to sort music into popular music pieces currently, musical genres, and make changes according to the user's recent music choices. The current industry model for music classification is called the multi-layer neural network classifier, which is used to separate classical and pop music. The audio features extracted are the average amplitude of Fourier Transform coefficients within different sub-bands. This report will first introduce ways to overcome the main struggles or areas of improvement in classification, then conduct a comparison between different models and identify why one outweighs the other in accuracy, and finally present the analysis and the suitable model that will be most beneficial in future developments.

2. Related Work

Cory McKay's [2] experiment uses k-Nearest-Neighbors, which has resulted in an overall success rate of 84.8% achieved for parent genres and 57.8% for subgenres across five training runs. Feature extraction played a significant role in this experiment which helped extract features from audio samples and provides extracted features to the classifier which then categorizes each piece of audio samples to match the labeled datasets (genres). Each music has its unique features which are why this process is necessary to help differentiate each sample from another.

On the contrary, there are issues regarding issues faced in music classification systems. The biggest problemis consistency, it is difficult for both humans and computers to classify music pieces under each genre consistently. According to Cory McKay [2], claiming that genre classification has the potential to contribute to the theoretical understanding of how each piece of music is classified under each genre.

Additionally, A research team from Cornell University [3] researchedhow the "Dislike" button will have the potential of improving Spotify's song recommendations. On average, users are 20 percent more likely to enjoy a song when 400,000 likes and disliked songs are trained in an algorithm, compared to Spotify's current algorithm which is only trained upon the number of likes. Sasha Stoikov, a senior research associate at Cornell Financial Engineering Manhattan, concluded that although the like button allows users to find songs they enjoy, it also has a greater chance of recommending songs they will not enjoy. Sasha co-authored with Hongyi Wen, a doctoral student at Cornell Tech, on a new system called "Piki", this system selects music from a database with 5 million songs, since this is a newly proposed system, they motivate new users to rate 25 songs and receive $1 in return. Because of this incentive, they were able to keep truthful ratings throughout the whole process.The problem with Spotify's algorithm impacts not only listeners but also new artists because their songs don't have enough likes to be recommended to a wide audience [3]. For Stoikov, future research and applications of Piki should be incorporated into other streaming platforms such as Netflix [3].

3. Methodology

3.1. Data

For this research, the GTZAN dataset will be used, providing one-hundred thirty seconds of audio files as sample data for each genre type. 3 main steps will be processed: audio file data processing, feature extraction, and classification. The dataset will be in the format of audio samples from 10 different genres into 10 genres (Blues, Classical, Country, Disco, HipHop, Jazz, Metal, Pop, Reggae, and Rock). The GTZAN genre collection dataset has been chosen specifically for this research because it is commonly used as a benchmark dataset in the field of machine learning, in this case, genre classification processes.

3.2. Methods

3.2.1 KNN

An algorithm that makes observations for similar data features for each class (genre) and classifies that observation according to the matching class [4]. K-nearest-neighbors (KNN) is a technique under supervised learning, it can calculate the distance between two data points which is also called the k number of labeled data points. KNN is commonly used for these tasks such as music genre classification, where the model is fitted with the training set and later evaluated with the test set, which then stores its performance score. This means that throughout all the sample audio samples, this process will be repeated multiple times to classify each data sample to its belonging genres.

3.2.2 CNN

CNN has its superiority over other neural networks in the sense of performance with image, speech, or audio signal inputs [5]. Three main layers: 1. Convolutional Layer 2. Pooling Layer 3. Fully connected (FC) Layer. According to results from [6], a 92.93 percent accuracy responded for CNN models, onthe contrary, Ndiatenda suggests KNN models provided a 92.69 percent accuracy. Despite both CNN and KNN systems returning with similar accuracies, it is also true that CNN models are exceptionally better at extracting features from audio samples, therefore, allowing CNN systems to stand out from KNN systems, especially for future applications.

3.2.3 Feature Extraction

Feature extraction performs frame processing on the original audio signal, then performs calculations based on the mathematical statistical significance of the features of each file, and finally uses the calculated results as training data for the classifier in form of vectors [7]. According to current methods, analysis methods such as the time-domain method are used to analyze and count the waveform state of musical signals from the time dimension, in turn converting frequency-domain analysis methods convert these signals to frequency domain through Fourier transform, collecting features that will be used for training data suitable for modeling [7]. Feature selection removes unnecessary features after features have already been extracted in earlier steps. This is done to either improve accuracy or boost the performance of high-dimensional datasets. To put it simply, feature extraction transforms arbitrary data into numerical features whereas feature selection is applied in addition to these numerical features [8].

Moreover, Feature Extraction is essential, especially for implementing genre classification techniques in recommendation systems. For instance, feature extraction increases the accuracy of models by extracting features from input data, making it easier for the classifier to categorize each audio sample.

4. Results

For the process of classification to operate efficiently and accurately, we first need to understand the important role of feature extraction and feature selection, in addition, understand when to use which. Feature extraction performs frame processing on the original audio signal, then performs calculations based on the mathematical statistical significance of the features of each file, and finally uses the calculated results as training data for the classifier in form of vectors [7]. According to current methods, analysis methods such as the time-domain method are used to analyze and count the waveform state of musical signals from the time dimension, in turn converting frequency-domain analysis methods convert these signals to frequency domain through Fourier transform, collecting features that will be used for training data suitable for modeling [7].

Feature selection removes unnecessary features after features have already been extracted in earlier steps. This is done to either improve accuracy or boost the performance of high-dimensional datasets. To put it simply, feature extraction transforms arbitrary data into numerical features whereas feature selection is applied in addition to these numerical features [8].

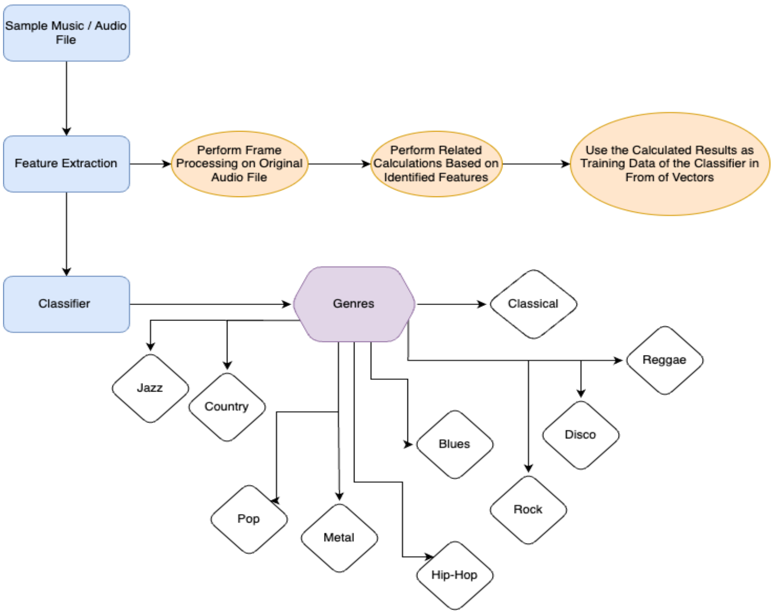

Figure 1. Model process

Figure 1 shows the model process. The general model of how each sample music will be processed and classified under certain genres. The blue boxes represent the main factors supplied for this model to work efficiently. The orange boxes present insight into how audio samples will go through the feature extraction process.

There will be three main parts for music classification works in machine learning:

(1) Audio samples will be presented and imported for raw data.

(2) Audio samples will go through feature extraction, where the main purpose of this process is to improve the accuracy of the results which is essential for a recommendation system where we want to provide precise results.

(3) Lastly, classifiers such as CNN or KNN will be used to categorize audio samples into individual genres. This process is very reliant on feature extraction.

Loss Value: This is the value representing the prediction error within each model. To do this, we use the BCE (Binary Crossentropy), this technique is specifically used for classification tasks, where the output value should be passed through a sigmoid activation where a range of output is 0-1 [9]. Reconstruction loss is the difference between the output image and the real image. Whereas KL loss is the measurement of the difference between two probability distributions.

Accuracy: This is to show how accurate the model is at classifying each sample piece into its significant genre/class. This model provided 96% accuracy.

Confusion matrices are a summary of the number of correct and incorrect predictions made by a classifier. This is exceptionally useful in helping a business decide which system to use for its streaming platform.

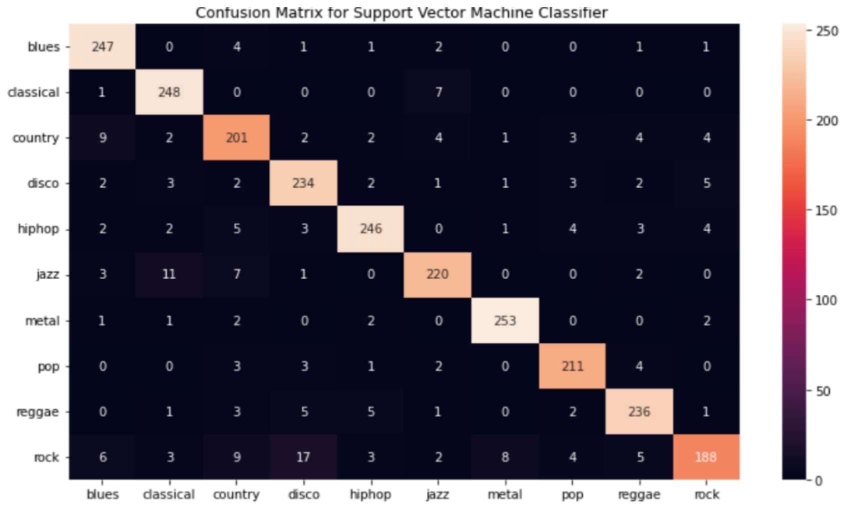

Figure 2. Confusion Matrix for Support Vector Machine (SVM) [10]

Figure 2 shows the confusion matrix for SVM from Kaggle, which is compared with our results. On the y and x-axis are each genre tested for this classifier. According to the results shown for the classifier SVM, the overall accuracy of this model is 89%.

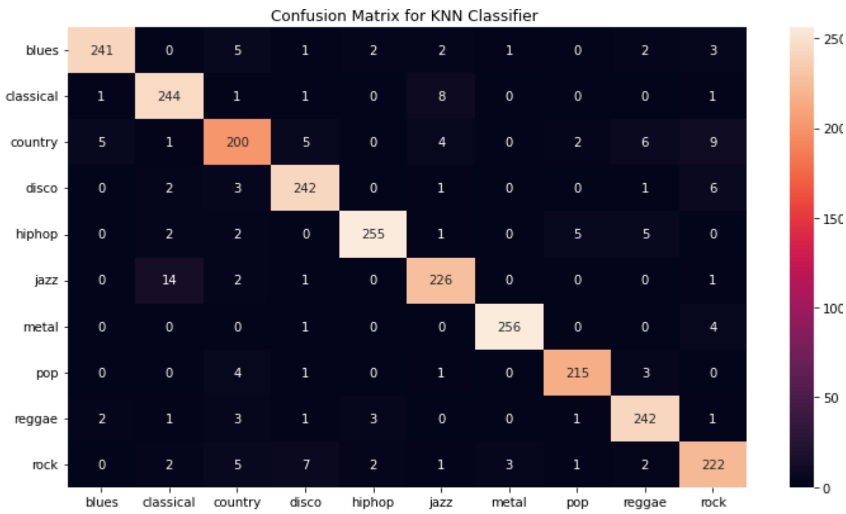

Figure 3. Confusion Matrix for KNN Classifier

Figure 3 shows the confusion matrix for KNN model. The same y and x-axis data are presented. According to this confusion matrix, the result of the KNN classifier is 92% accurate.

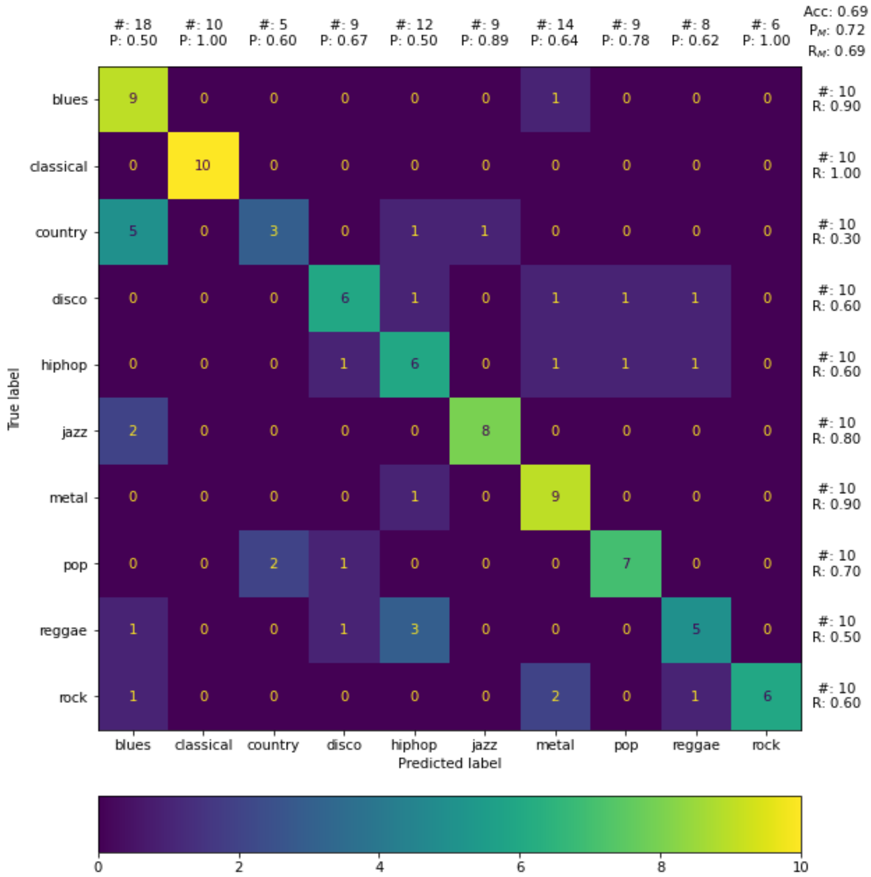

Figure 4. Confusion Matrix for CNN

Figure 4 shows the confusion matrix for CNN Classifier. According to this confusion matrix, the final accuracy for this classifier is 62%.

5. Discussion

Although the most effective and accurate model implementation for the usual recommendation systems that we use now can safely conclude that KNN satisfies current demands. As early as 2013, Hernani Costa and Luis Macedo at the University of Coimbra, Portugal, have reasoned their view on how having an implication emotion-based recommender system can help overcome the problem of information overload. First of all, how can an emotion-based recommendation system be beneficial for music streaming platforms? According to an article by Shahram Heshmat [11], an associate professor at the University of Illinois at Springfield, discussing the psychology behind the effects of music on emotions. He claims, "Listening to music is an easy way to alter mood or relieve stress.” A handful of research suggests the benefits of listening to music and how it causes the release of neurotransmitters associated with pleasure such as dopamine. An emotion-based recommendation system grants users a more personal recommended playlist based on their emotions. Instead of using a KNN classifier for these systems, an approach of using CNN is present. In January of 2021, the popular music streaming platform Spotify has granted a patent that permits them the use of speech recognition to determine a user’s emotional state, gender, age, and accent; with these factors in play, Spotify will now be able to create more specific and personalized recommendations for the user [12]. As this system develops over time, it could achieve more accurate results and have the possibility to be applied to more streaming services and to identify specific music pieces just by playing the song for the service to recognize that song for the user to listen to.

6. Conclusion

Having more personalized music playlists created for the user allows music streaming businesses to gain customers' attention and high customer satisfaction. Without recommendations, users will find it harder to discover their favorite tracks on the platform, therefore, losing interest. This research aims to identify and analyze classification models and potential flaws with the current model and propose a probable way to resolve the issue. Disregarding the generic use of classification, these techniques could also be used to provide studies on the perception and properties of music. I have achieved my goal of this project by first explaining how a music genre classification system works and how it is implemented into streaming platforms, the importance of personalized recommendation systems for the benefit of user's experience, and finally explained why CNN would be a much ideal technique for the future of recommendation systems. To further develop my research, as my knowledge on this topic extends, emotion-based recommendation systems will be the next step in my research to find the next steps for more personalized music recommendation systems that will greatly benefit users in the future.

References

[1] Muvi.com. 2022. Music Recommendation System – Revolutionizing the Music Streaming Industry. https://www.muvi.com/blogs/music-recommendation-system.html.

[2] McKay, C. (2004). Issues in Automatic Musical Genre Classification. In Proceedings fo the McGill Graduate Students Society Symposium.

[3] Adam Conner-Simons. 2021. The 'Dislike' button would improve Spotify's recommendations. https://news.cornell.edu/stories/2021/09/dislike-button-would-improve-spotifys-recommendations.

[4] Towardsdatascience.com. 2022. A Simple Introduction to K-Nearest Neighbors Algorithm. https://towardsdatascience.com/a-simple-introduction-to-k-nearest-neighbors-algorithm-b3519ed98e.

[5] IBM. 2021.Convolutional Neural Networks. https://www.ibm.com/cloud/learn/convolutional-neural-networks.

[6] Arsh Chowdhry. 2021. Music Genre Classification Using CNN – Chowdhry Arsh. https://www.clairvoyant.ai/blog/music-genre-classification-using-cnn#:%7E:text=The%20accuracy%20we%20achieved%20for,helps%20in%20the%20classifying%20process.

[7] Zhang, J. (2021). Music feature extraction and classification algorithm based on deep learning. Scientific Programming, 2021.

[8] Y. Chen, J. Pei and D. Li, "DETPro: A High-Efficiency and Low-Latency System Against DDoS Attacks in SDN Based on Decision Tree," ICC 2019 - 2019 IEEE International Conference on Communications (ICC), 2019, pp. 1-6, doi: 10.1109/ICC.2019.8761580.

[9] Yashi Shukla. 2022. Genre Classification using CNN. https://medium.com/@yashi4001/genre-classification-using-cnn-dcbc109b6d1d#:%7E:text=The%20dataset%20consists%20of%2010,important%20feature%20of%20audio%20files

[10] Kaggle. 2022. Music Classification, Recommendation, and Generation. https://www.kaggle.com/code/lkergalipatak/music-classification-recommendation-generation. Accessed 31 July 2022.

[11] Costa, H., Macedo, L. (2013). Emotion-Based Recommender System for Overcoming the Problem of Information Overload. In: , et al. Highlights on Practical Applications of Agents and Multi-Agent Systems. PAAMS 2013. Communications in Computer and Information Science, vol 365. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-38061-7_18.

[12] Savage, M. 2021. Spotify wants to suggest songs based on your emotions. https://www.bbc.com/news/entertainment-arts-55839655.