**MUSIC GENRE CLASSIFICATION**

A Project-II Report

Submitted in partial fulfillment of requirement of the

Degree of

**BACHELOR OF TECHNOLOGY in COMPUTER SCIENCE & ENGINEERING**

BY

**Aditya Manthankar**

**EN18CS301012**

Under the Guidance of

**Mr. Vivek Gurjar (AI Team Lead, Linkites Infotech Pvt Ltd)**

**Mr. Ashish Kumawat (Professor Medicap University)**



**Department of Computer Science & Engineering**

**Faculty of Engineering**

**MEDI-CAPS UNIVERSITY, INDORE- 453331**

**MAY 2022**

**MUSIC GENRE CLASSIFICATION**

A Project-II Report

Submitted in partial fulfillment of requirement of the

Degree of

**BACHELOR OF TECHNOLOGY in COMPUTER SCIENCE & ENGINEERING**

BY

**Aditya Manthankar**

**EN18CS301012**

Under the Guidance of

**Mr. Vivek Gurjar (AI Team Lead, Linkites Infotech Pvt Ltd)**

**Mr. Ashish Kumawat (Professor Medicap University)**



**Department of Computer Science & Engineering**

**Faculty of Engineering**

**MEDI-CAPS UNIVERSITY, INDORE- 453331**

**MAY 2022**

**Report Approval**

The project work **Music Genre Classification** is hereby approved as a creditable study of an engineering/computer application subject carried out and presented in a manner satisfactory to warrant its acceptance as prerequisite for the Degree for which it has been submitted.

It is to be understood that by this approval the undersigned do not endorse or approved any statement made, opinion expressed, or conclusion drawn there in; but approve the “Project Report” only for the purpose for which it has been submitted.

Internal Examiner

Name: Mr. Ashish Kumawat

Designation: Assistant Professor

Affiliation

External Examiner

Name:

Designation

Affiliation

**ii**

**Declaration**

I/We hereby declare that the project entitled **Music Genre Classification** submittedin partial fulfillment for the award of the degree of Bachelor of Technology in ‘<Name of Department>’ completed under the supervision of **Mr. Ashish Kumawat, Assistant Professor, Department of Computer Science and Engineering,** Faculty of Engineering, Medi-Caps University Indore is an authentic work.

Further, I/we declare that the content of this Project work, in full or in parts, have neither been taken from any other source nor have been submitted to any other Institute or University for the award of any degree or diploma.

**Aditya Manthankar**

**Name of the student**

**Signature with date**

**iii**

**Certificate**

I/We, **Ashish Kumawat** certify that the project entitled **Music Genre Classification** submittedin partial fulfillment for the award of the degree of Bachelor of Technology by **Aditya Manthankar** istherecordcarried out by him/them under my/our guidance and that the work has not formed the basis of award of any other degree elsewhere.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Ashish Kumawat

Department of Computer Science and Engineering

Medi-Caps University, Indore

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

<Name of External Guide (If any)>

<Name of the Department>

Name of the Organization

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Dr. Pramod S. Nair

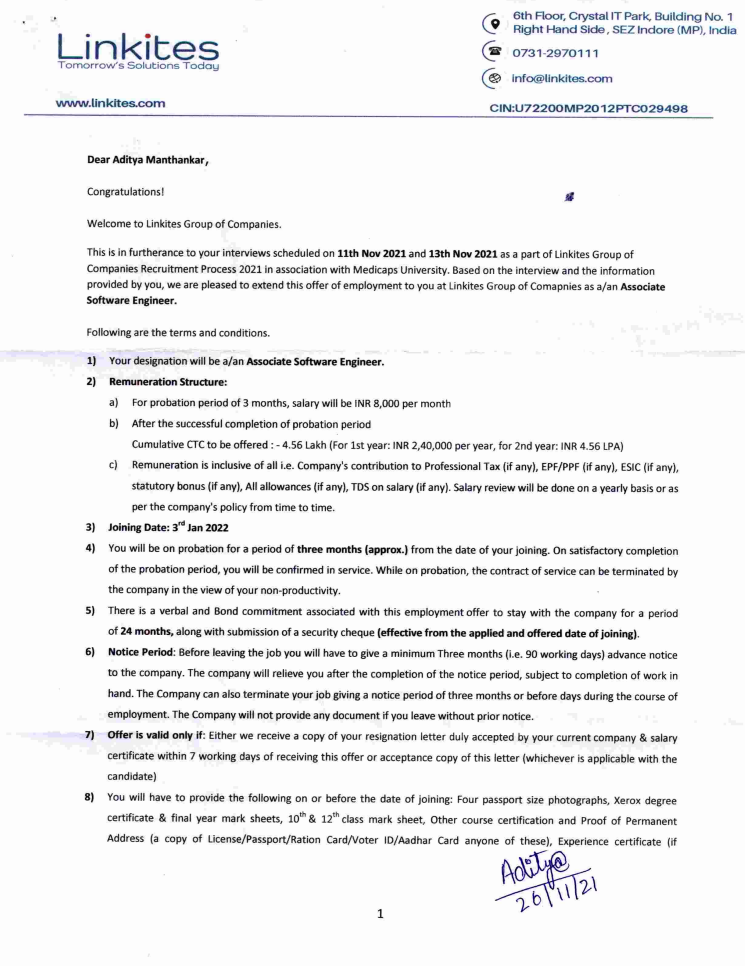
Head of the Department

Computer Science & Engineering

Medi-Caps University, Indore

**iv**

**Offer Letter of the Project work-II/Internship**

****

**v**

****

**vi**

**Completion Certificate/ Letter**

****

**vii**

**Acknowledgements**

I would like to express my deepest gratitude to Honorable Chancellor, **Shri R C Mittal,** who has provided me with every facility to successfully carry out this project, and my profound indebtedness to **Prof. (Dr.) Dilip K. Patnaik,** Vice Chancellor, Medi-Caps University, whose unfailing support and enthusiasm has always boosted up my morale. I also thank **Prof. (Dr.) D K Panda,** Pro Vice Chancellor, **Dr. Suresh Jain,** DeanFaculty of Engineering, Medi-Caps University, for giving me a chance to work on this project. I would also like to thank my Head of the Department **Dr. Pramod S. Nair** for his continuous encouragement for betterment of the project.

I express my heartfelt gratitude to my **External Guide, Mr. Vivek Gurjar**, AI Team Lead, Linkites Infotech Pvt. Ltd as well as to my Internal Guide, Mr**. Ashish Kumawat,** Professor, Department of Computer Science and Engineering, Medi Caps University, without whose continuous help and support, this project would ever have reached to the completion.

I would also like to thank to my team at Linkites Mr. Prabhat Jogdande, Mr. Jay Agrawal, Ms. Priyal Mandloi who extended their kind support and help towards the completion of this project.

It is their help and support, due to which we became able to complete the design and technical report. Without their support this report would not have been possible.

**Aditya Manthankar**

B.Tech. IV Year

Department of Computer Science & Engineering

Faculty of Engineering

Medi-Caps University, Indore

**viii**

**Abstract**

Music plays a very important role in people’s lives. Music brings like-minded people together and is the glue that holds communities together. Communities can be recognized by the type of songs that they compose, or even listen to. Genre classification is an important task with many real-world applications. As the quantity of music being released on a daily basis continues to sky-rocket, especially on internet platforms such as Soundcloud and Spotify – a recent study suggests that tens of thousands of songs were released every month on Spotify – the need for accurate meta-data required for database management and search/storage purposes climbs in proportion. Being able to instantly classify songs in any given playlist or library by genre is an important functionality for any music streaming/purchasing service, and the capacity for statistical analysis that correct and complete labeling of music and audio provides is essentially limitless. We implemented two classification algorithms admitting two different types of input. We experimented k-nearest neighbors and a basic feed-forward network.

For the input to our algorithms, we experimented with both raw amplitude data as well as transformed Mel-spectrograms of that raw amplitude data. We then output a predicted genre out of 10 common music genres.

**ix**

**Table of Contents**

|  |  |  |
| --- | --- | --- |
|  |  | **Page No.** |
|  | Report Approval | ii |
|  | Declaration | iii |
|  | Certificate | iv |
|  | Offer Letter of the Project work-II/Internship | v |
|  | Completion letter/certificate | vi |
|  | Acknowledgement | vii |
|  | Abstract | viii |
|  | Table of Contents | ix |
|  | List of figures | x |
|  | List of tables | xi |
|  | Abbreviations | xii |
|  | Notations & Symbols | xiii |
| Chapter 1 | **Introduction** |  |
|  | 1.1 Introduction | 1 |
|  | 1.2 Literature Review |  |
|  | 1.3 Objectives |  |
|  | 1.4 Significance |  |
|  | 1.5 Research Design |  |
|  | 1.6 Source of Data |  |
|  | 1.7 Chapter Scheme |  |
| Chapter 2 | **Report on Present Investigation** |  |
|  | 2.1 Experimental Set-up |  |
|  | 2.2 Procedures Adopted |  |
| Chapter 3 | **(Main Chapter)** |  |
|  | 3.1 Implementation |  |
| Chapter 4 | **Results and Discussions** |  |
| Chapter 5 | **Summary and Conclusions** |  |
| Chapter 6 | **Future scope** |  |
|  | Appendix |  |
|  | Bibliography |  |

**x**

**List of Figures**

|  |  |  |
| --- | --- | --- |
| **Figure Number** | **Title Page** | **Page Number** |
| 1.1 | Machine Learning Flow |  |
| 1.2 | Supervised Learning |  |
| 1.2.1 | Deep Learning Process |  |
| 2.1.1 | Spectrograms |  |
| 2.1.2 | Spectral Roll-Off |  |
| 2.1.3 | Chroma Feature |  |
| 2.1.4 | Zero Crossing Rate |  |
| 4 | Test Accuracy |  |

**x**

**Abbreviations**

The following abbreviations are used:

|  |  |
| --- | --- |
| **Abbreviations** | **Meaning** |
| CNN | Convolution Neural Network |
| MFCC | Mel-frequency cepstral coefficients |
|  |  |
|  |  |

**xi**

**Chapter-1**

1. **Introduction**

Music is characterized by giving them categorical labels called genres. These genres are created by humans. A music genre is segregated by the characteristics which are commonly shared by its members. Typically, these characteristics are related to the rhythmic structure, instrumentation, and the harmonic content of the music. To categorize music files into their respective genres, it is a very challenging task in the area of Music Information Retrieval (MIR), a field concerned with browsing, organizing and searching large music collections. Classification of genre can be very valuable to explain some interesting problems such as creating song references, tracking down related songs, discovering societies that will like that specific song, sometimes it can also be used for survey purposes. Automatic musical genre classification can assist humans or even replace them in this process and would be of a very valuable addition to music information retrieval systems. In addition to this, automatic classification of music into genres can provide a framework for development and evaluation of features for any type of content-based analysis of musical signals.

The concept of automatic music genre classification has become very popular in recent years as a result of the rapid growth of the digital entertainment industry. Dividing music into genres is arbitrary, but there are perceptual criteria that are related to instrumentation, structure of the rhythm and texture of the music that can play a role in characterizing particular genre. Until now genre classification for digitally available music has been performed manually. Thus, techniques for automatic genre classification would be a valuable addition to the development of audio information retrieval systems for music.

**1.1 Introduction to Machine Learning:**

The term machine learning was first introduced by **Arthur Samuel** in **1959**. We can define it in a summarized way as:

“Machine learning enables a machine to automatically learn from data, improve performance from experiences, and predict things without being explicitly programmed.”

## How does Machine Learning work?

## A Machine Learning system **learns from historical data, builds the prediction models, and whenever it receives new data, predicts the output for it**. The accuracy of predicted output depends upon the amount of data, as the huge amount of data helps to build a better model which predicts the output more accurately.

## 

Figure 1.1 Machine Learning Flow

# **Types of Machine Learning**

The major recognized categories of machine learning are: **supervised learning, unsupervised learning, reinforcement learning and reinforcement learning.**

**1. Supervised learning:** The dataset being used has been pre-labelled and classified by users to allow the algorithm to see how accurate its performance is.

**2. Unsupervised learning:** The raw dataset being used is unlabeled and an algorithm identifies patterns and relationships within the data without help from users.

**3. Semi-supervised learning:** The dataset contains structured and unstructured data, which guide the algorithm on its way to making independent conclusions. The combination of the two data types in one training dataset allows machine learning algorithms to learn to label unlabeled data.

**4. Reinforcement learning:** The dataset uses a “rewards/punishments” system, offering feedback to the algorithm to learn from its own experiences by trial and error.

**1.1.3 Supervised Learning**

Supervised learning is a type of machine learning method in which we provide sample labelled data to the machine learning system in order to train it, and on that basis, it predicts the output. The system creates a model using labelled data to understand the datasets and learn about each data, once the training and processing are done then we test the model by providing sample data to check whether it is predicting the exact output or not. Some very practical applications of supervised learning algorithms in real life, include: Face Detection, Signature recognition, Spam detection and Weather forecasting.

The working of Supervised learning can be easily understood by the below example and diagram:

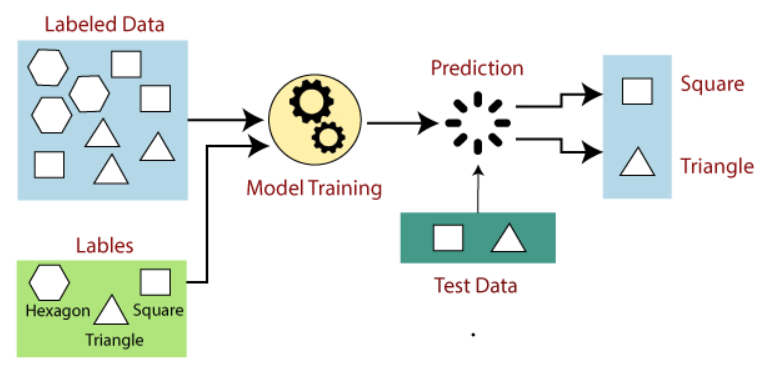
****

Figure 1.2 Supervised Learning

Supervised learning can be further divided into two types of problems:

1. **Regression:** Regression algorithms are used if there is a relationship between the input variable and the output variable. It is used for the prediction of continuous variables, such as Weather forecasting, Market Trends, etc.

Below are some popular Regression algorithms which come under supervised learning:

* Linear Regression
* Regression Trees
* Non-Linear Regression
* Bayesian Linear Regression
* Polynomial Regression

1. **Classification:** Classification algorithms are used when the output variable is categorical, which means there are two classes such as Yes-No, Male-Female, True-false, etc.

Here are a few popular classification algorithms:

* Linear Classifiers
* Support Vector Machines
* Decision Trees
* K-Nearest Neighbor
* Random Forest

**1.2 Literature Review**

**1.2.1 Introduction to Deep Learning:**

Deep learning can be considered as a subset of [machine learning](https://www.simplilearn.com/tutorials/machine-learning-tutorial/what-is-machine-learning). It is a field that is based on learning and improving on its own by examining computer algorithms. While machine learning uses simpler concepts, deep learning works with artificial neural networks, which are designed to imitate how humans think and learn.

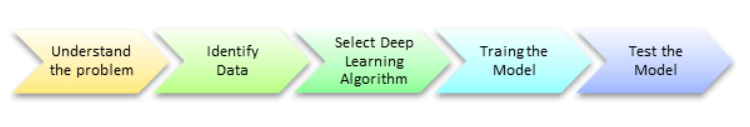


Figure 1.2.1 Deep learning Process

## 1.2.2 What Is a Neural Network?

A neural network is a series of algorithms that endeavors to recognize underlying relationships in a set of data through a process that mimics the way the human brain operates.

## 1.2.3 Types of Deep Learning Networks

## Most common types of neural networks or deep learning networks are:

### **1. Feed Forward Neural Network:** A feed-forward neural network is none other than an [Artificial Neural Network](https://www.javatpoint.com/keras-artificial-neural-networks), which ensures that the nodes do not form a cycle. In this kind of neural network, all the perceptrons are organized within layers, such that the input layer takes the input, and the output layer generates the output.

### **2. Recurrent Neural Network:** [Recurrent neural networks](https://www.javatpoint.com/keras-recurrent-neural-networks) are yet another variation of feed-forward networks. Here each of the neurons present in the hidden layers receives an input with a specific delay in time. The Recurrent neural network mainly accesses the preceding info of existing iterations.

### **3. Convolutional Neural Network:** [Convolutional Neural Networks](https://www.javatpoint.com/keras-convolutional-neural-network) are a special kind of neural network mainly used for image classification, clustering of images and object recognition. DNNs enable unsupervised construction of hierarchical image representations. To achieve the best accuracy, deep convolutional neural networks are preferred more than any other neural network.

**1.3 Objectives**

The idea behind this project is to see how to handle sound files in python, compute sound and audio features from them, run Machine Learning Algorithms on them, and see the results.

In a more systematic way, the main aim is to create a machine learning model, which classifies music samples into different genres. It aims to predict the genre using an audio signal as its input.

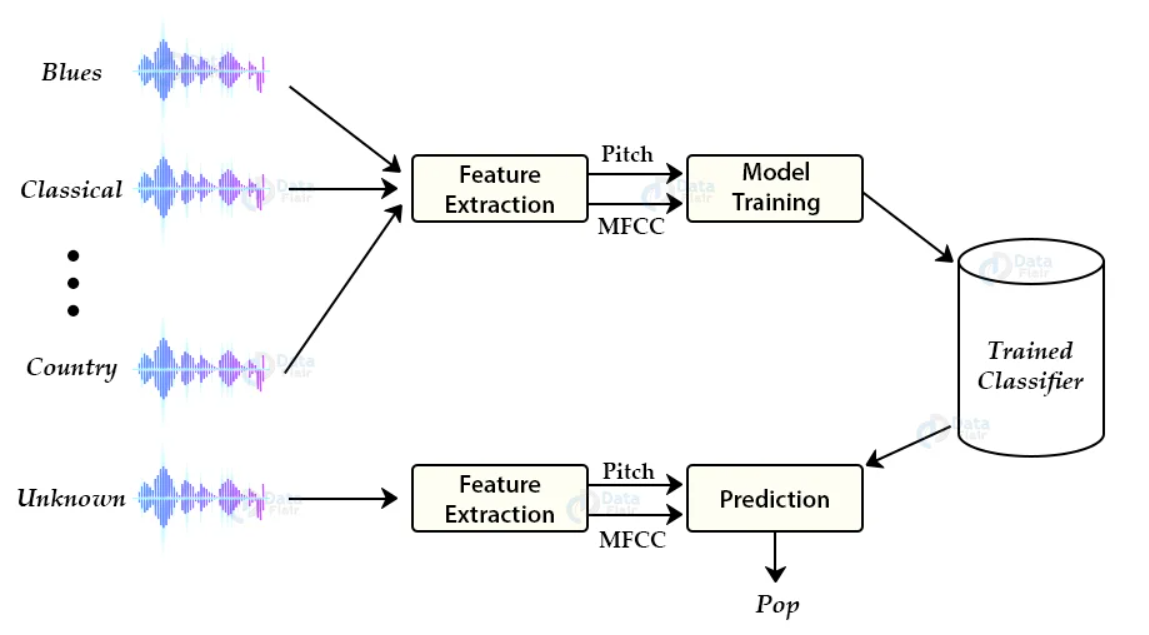
The objective of automating the music classification is to make the selection of songs quick and less cumbersome. If one has to manually classify the songs or music, one has to listen to a whole lot of songs and then select the genre. This is not only time-consuming but also difficult. Automating music classification can help to find valuable data such as trends, popular genres, and artists easily. Determining music genres is the very first step towards this direction.

**1.4 Significance**

Classification of genre can be very valuable to explain some interesting problems such as creating song references, tracking down related songs, discovering societies that will like that specific song, sometimes it can also be used for survey purposes.

Music plays a very important role in people’s lives. Music brings like-minded people together and is the glue that holds communities together. Communities can be recognized by the type of songs that they compose, or even listen to. Different communities and groups listen to different kinds of music. One main feature that separates one kind of music from another is the genre of the music.

**1.5 Research Diagram**

****

**1.6 Source of Data**

The dataset used in this project is GTZAN genre collection dataset was collected in 2000-2001. It consists of 1000 audio files each having 30 seconds duration. There are 10 classes (10 music genres) each containing 100 audio tracks. Each track is in .wav format. It contains audio files of the following 10 genres:

* Blues
* Classical
* Country
* Disco
* Hip-hop
* Jazz
* Metal
* Pop
* Reggae
* Rock

The dataset folder contains following contents:

* genres original - A collection of 10 genres with 100 audio files each, all having a length of 30 seconds (the famous GTZAN dataset, the MNIST of sounds)
* images original - A visual representation for each audio file. One way to classify data is through neural networks. Because NNs (like CNN, what we will be using today) usually take in some sort of image representation, the audio files were converted to Mel Spectrograms to make this possible.
* 2 CSV files - Containing features of the audio files. One file has for each song (30 seconds long) a mean and variance computed over multiple features that can be extracted from an audio file. The other file has the same structure, but the songs were split before into 3 seconds audio files (this way increasing 10 times the amount of data we fuel into our classification models).

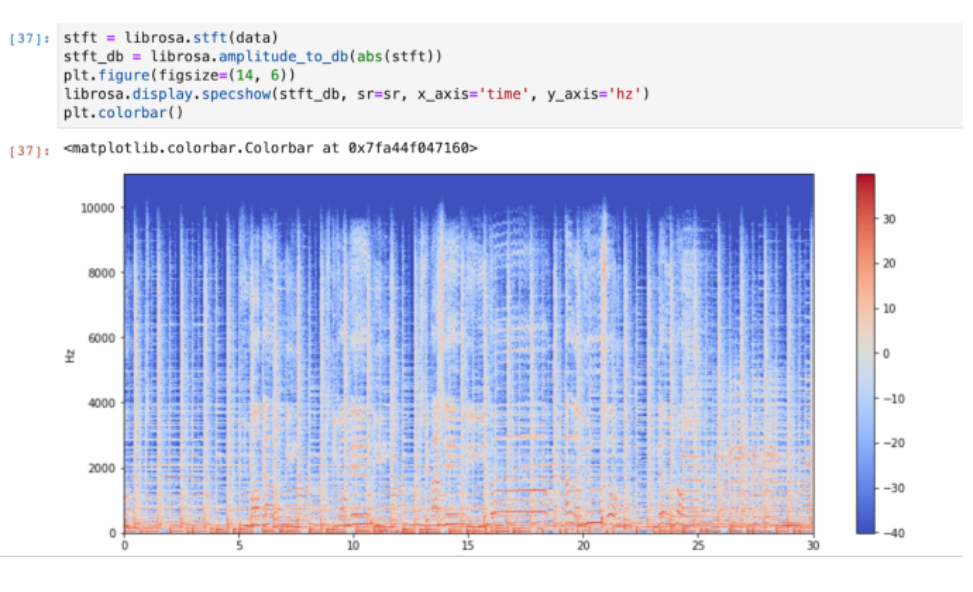
**Chapter 2**

**2.1 Experimental Set-up**

### **2.1.1 Spectrograms**

A spectrogram is a visual way of representing the signal loudness of a signal over time at various frequencies present in a particular waveform. Not only can one see whether there is more or less energy at, for example, 2 Hz vs 10 Hz, but one can also see how energy levels vary over time.

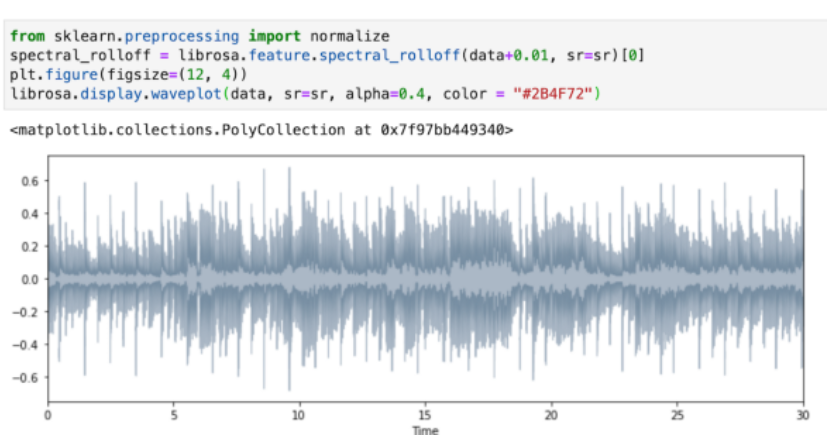
Spectrograms are sometimes called sonographs, voiceprints, or voicegrams. When the data is represented in a 3D plot, they may be called waterfalls. In 2-dimensional arrays, the first axis is frequency while the second axis is time.

****

### **Figure 2.1.1 Spectrograms**

### **2.1.2 Spectral Roll-Off**

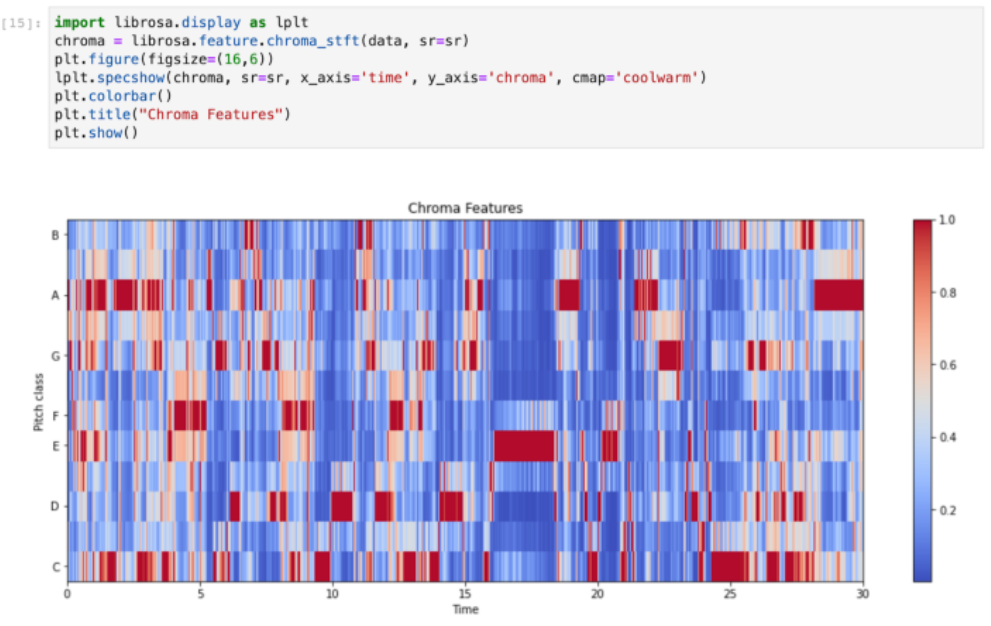
Spectral Roll off is the frequency below which a specified percentage of the total spectral energy, (e.g. 85%) lies.



### Figure 2.1.2 Spectral Roll-Off

### **2.1.3 Chroma Feature**

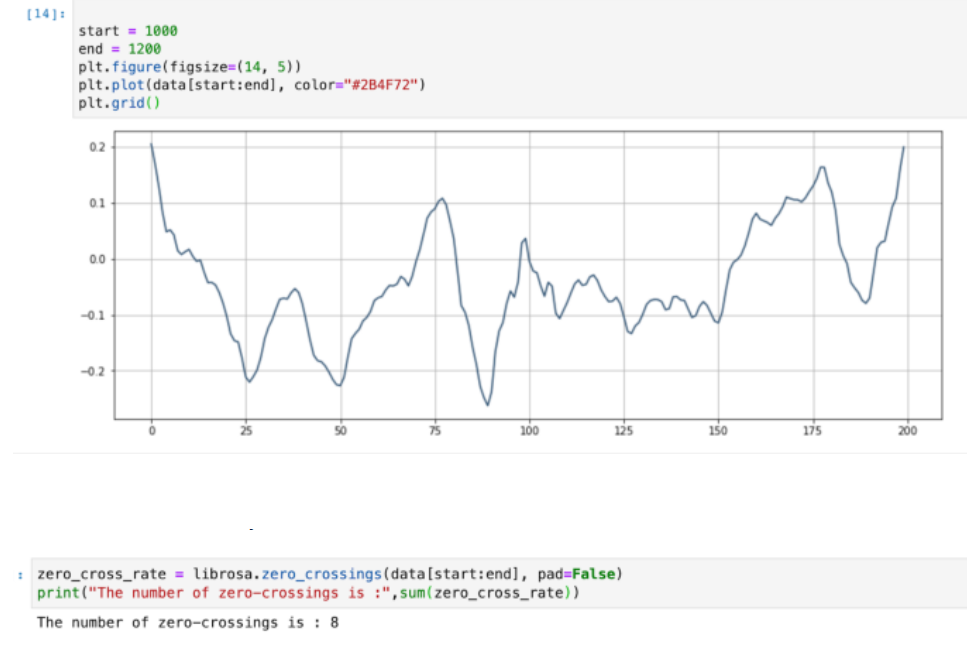
It is a powerful tool for analyzing music features whose pitches can be meaningfully categorized and whose tuning approximates to the equal-tempered scale. One main property of chroma features is that they capture harmonic and melodic characteristics of music while being robust to changes in timbre and instrumentation



### Figure 2.1.3 Chroma Feature

### **2.1.4 Zero Crossing Rate**

[Zero crossing](https://en.wikipedia.org/wiki/Zero-crossing_rate) is said to occur if successive samples have different algebraic signs. The rate at which zero-crossings occur is a simple measure of the frequency content of a signal. Zero-crossing rate is a measure of the number of times in a given time interval/frame that the amplitude of the speech signals passes through a value of zero.



### Figure 2.1.4 Zero Crossing Rate

**2.2 Procedure Adopted**

A CNN is a feed-forward network, that is, input examples are fed to the network and transformed into an output; with supervised learning, the output would be a label, a name applied to the input. That is, they map raw data to categories, recognizing patterns that may signal. A feedforward network is trained on labeled data until it minimizes the error it makes when guessing their categories. With the trained set of parameters (or weights, collectively known as a model), the network sallies forth to categorize data it has never seen. A trained feedforward network can be exposed to any random collection of audio files. That is, a feedforward network has no notion of order in time, and the only input it considers is the current example it has been exposed to. Feedforward networks are amnesiacs regarding their recent past; they remember nostalgically only the formative moments of training.

* + 1. **Operations Of CNN**

Each block in a CNN consists of the following operations:

• Convolution: This step involves a matrix filter (say 3x3 size) that is moved over the input data which is of dimension (say 4x4). The filter is first placed on the raw audio data matrix and then we compute an element-wise multiplication between the filter and the overlapping portion of the data, followed by a summation to give a feature value.

• Pooling: This method is used to reduce the dimension of the feature map obtained from the convolution step. By max pooling with say 2x2 window size, we only retain the element with the maximum value among the 4 elements of the feature map that are covered in this window. We move this window across the feature map with a predefined stride.

• Non-linear Activation: The convolution operation is linear and in order to make the neural network more powerful, we need to introduce some non-linearity. For this purpose, we can apply an activation function such as Rectifier Linear Unit (ReLU) on each element of the feature map.

The model consists of 3 convolutional blocks (conv base), followed by a flatten layer which converts a 2D matrix to a 1D array, which is then followed by a fully connected layer, which outputs the probability that a given image belongs to each of the possible classes.

The final layer of the neural network outputs the class probabilities (using the softmax activation function) for each of the eight possible class labels

**Chapter 3**

**3.1 Implementation**

**3.1.1 Music Genre Classification**

As mentioned, the data set which we are going to use here was downloaded has about 100 songs under each of the 10 labels or genres. Each of the songs are 30 seconds long. I split every audio file into 10 segments with each segment being three seconds long.

Hence, the number of songs under each label is now 1000, which is a decent number to train the model to achieve good accuracy.

Now that we have our data ready, we need to extract the features which will be suitable to feed into our network. The feature extraction will be done by using MFCCs. Librosa is used to extract the features from each of the audio segments. We create a dictionary with the label or category of the genre as the key and all the extracted features from all the 1,000 segments as an array of features under that label.

Once we do this in a loop for all 10 categories, we dump the dictionary into a JSON file. This JSON file thus becomes our dataset on which the model will be trained.

Moving into the coding for dataset preprocessing, we first define the number of segments and sample rate of each segment. The sample rate is required to know the playback speed of the song. Here, we keep it constant for every segment:

We then create a loop in which we open up every song file from every genre folder and split it into 10 segments. We then extract the MFCC features for each segment and append it to the dictionary under the genre name (which is also the folder name).

Before we build the model, we have to load it into our program and split it into training and testing. This is done by opening the JSON file which we created in the last section and converting it into NumPy arrays for easy computation.

After loading the data, we prepare the data and split it into train and test sets as mentioned earlier. This is done by using the following sklearn’strain\_test\_split function

Next, the CNN network is created using TensorFlow.

 The size of the input layer depends on the size of the MFCC coefficient which we are passing as an argument. You can experiment with more hidden layers and test the accuracy.

Once all the methods and functions have been defined, it’s time to call them and train our classification model.

First, we pass the test date percentage and validation data percentage. Validation data is some part of the training data, with which the model is not trained and is used to validate the model. The validation set tells us whether the data is performing well or not after the training is done. Next, we the compile CNN network. Compiling is used to add the optimizer (which defines the learning rate) and the loss calculating function.

After compiling, model.fit() is used to train the model on our data. The training time depends upon your system hardware.

We don’t want to keep training our model in order to test it. So, after training, we save the model so we can use this saved file to predict on our new data. At the end of the training, you can see the accuracy achieved.

We load the saved model and define the different classes or genres. The model will predict a number from 0 to 9, and each number will represent a genre as defined during training.

The model predicts a genre for each and every segment of the input song. The most predicted genres combine all the predictions of all the sliced segments of a particular input song and gives the final prediction.

**Chapter -4**

**Result and Discussion**

The best performance in terms of accuracy is observed for the CNN model that uses only the spectrogram as an input to predict the music genre with a test accuracy of 78.94. Also, there were some false predictions by the model when provided with completely new audio files The reason behind this low-test accuracy rate could be the limited dataset of 1000 audio tracks. Increase in the dataset and better hardware system hardware might improve the accuracy of these models

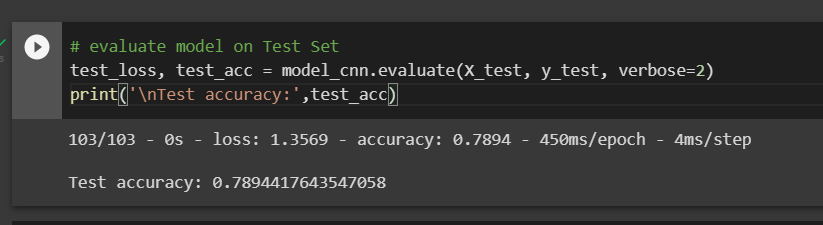
****

Figure 4 Test Accuracy

**Chapter -5**

**Summary and Conclusion**

This research work provides the details of an application which performs Music Genre Classification using Machine Learning techniques. The application uses a Convolutional Neural Network model to perform the classification. A Mel Spectrum of each track from the GTZAN dataset is obtained. This is done by using the librosa package of python. A piece of code is implemented which performs classification of huge database of songs into their respective genres.

For the GTZAN dataset, the model we used achieved a training accuracy of about 98% and validation accuracy of 79. Python was the language used to develop the model. A number of packages such as Keras, Numpy, pandas were used to build the model. Experiment is done on the google Colab platform. TensorFlow package is used for deep-learning.

From results described above it can be concluded that a coarse classification of music into four genres can be done easily by machines. And the confidence with which it is done is also significant. The work described in this report can be extended to include further genres like Techno etc. Another direction would be to classify music by artist/composer. This we believe will be more difficult than classifying into different genres; an analysis of the kind described in the report may not be enough to differentiate between different artists/composers.

**Chapter-6**

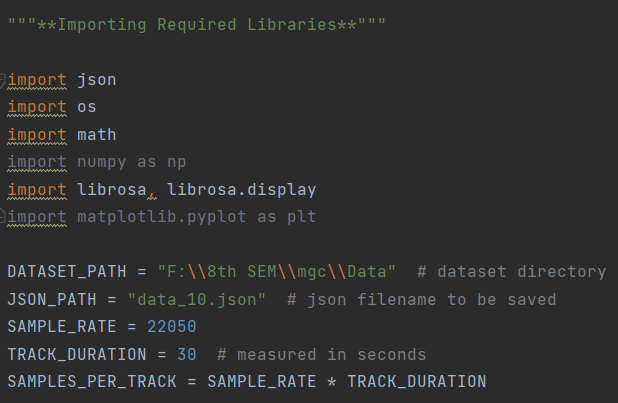
**Future Scope**

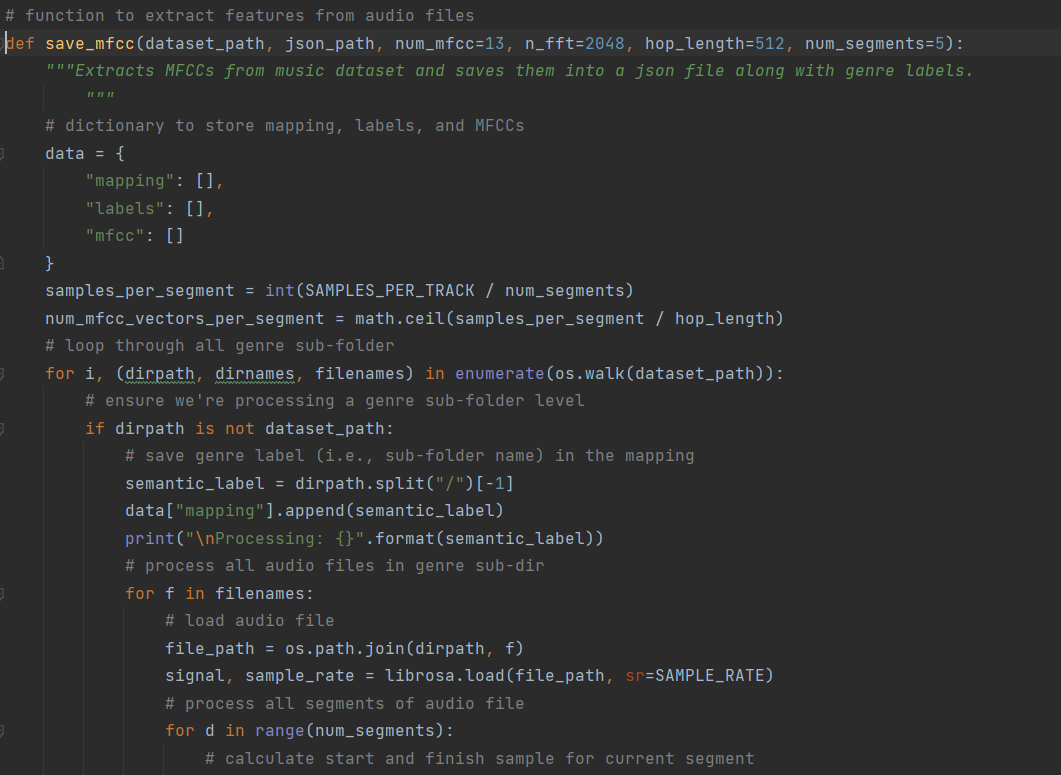
In the future, we hope to experiment with other types of deep learning methods, given they performed the best. Given that this is time series data, some sort of RNN model may work well (GRU, LSTM, for example). We are also curious about generative aspects of this project, including some sort of genre conversion (in the same vein as generative adversarial networks which repaint photos in the style of Van Gogh, but for specifically for music). Additionally, we suspect that we may have opportunities for transfer learning, for example in classifying music by artist or by decade.

The extension of this work would be to consider bigger data sets. Also, with time the style represented by each genre will continue to change. So, the objective for the future will be to stay updated with the change in styles of genres and extending our software to work on these updated styles. This work can also be extended to work as a music recommendation system depending on the mood of the person.

**Appendix**

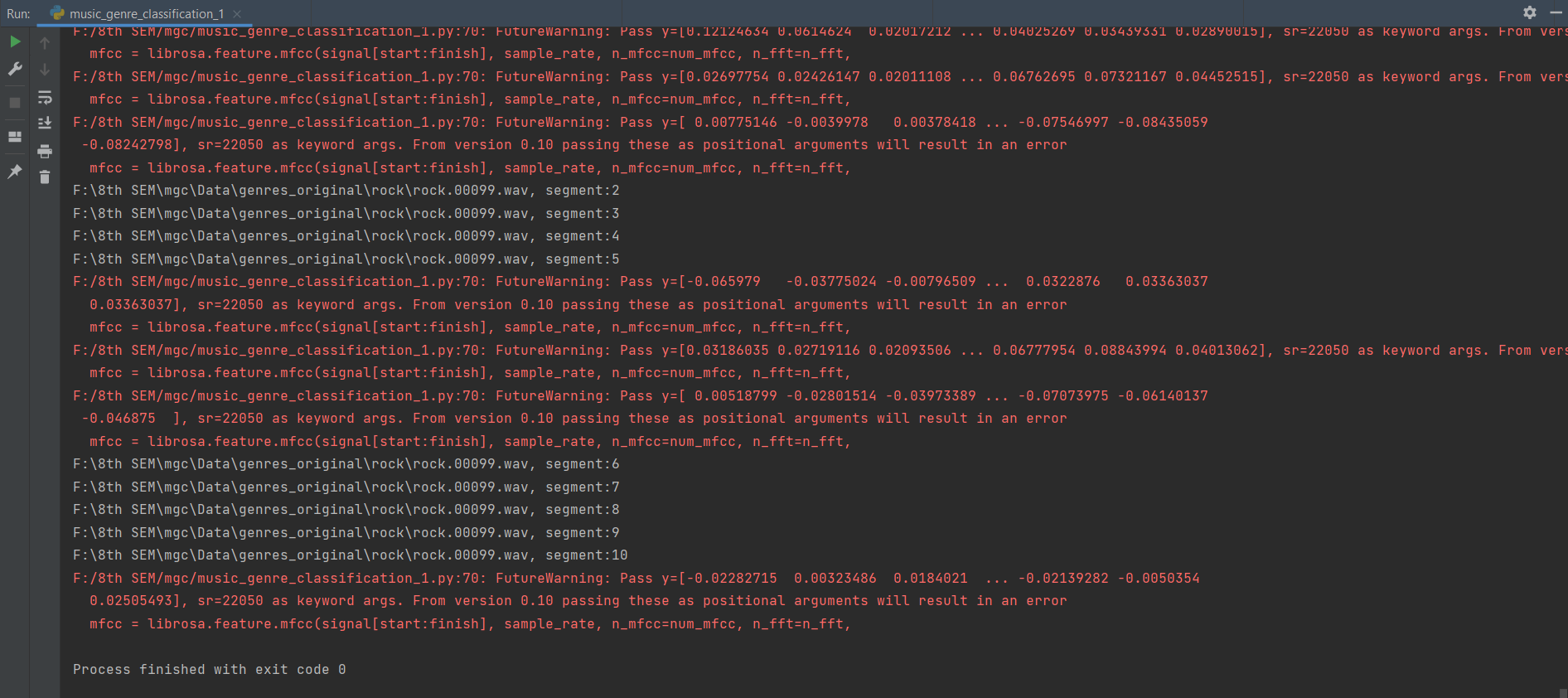
**Feature Extraction:**

****

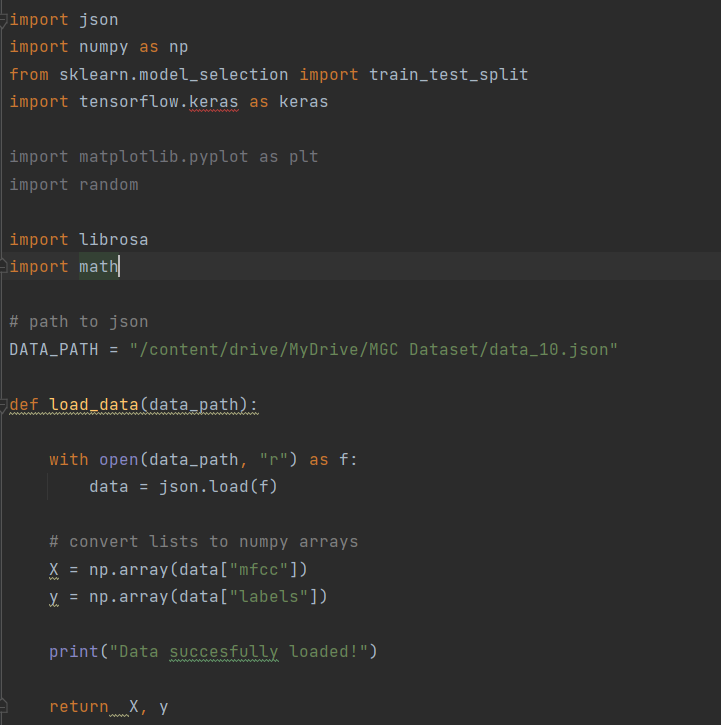
****

****

**Output:**

****

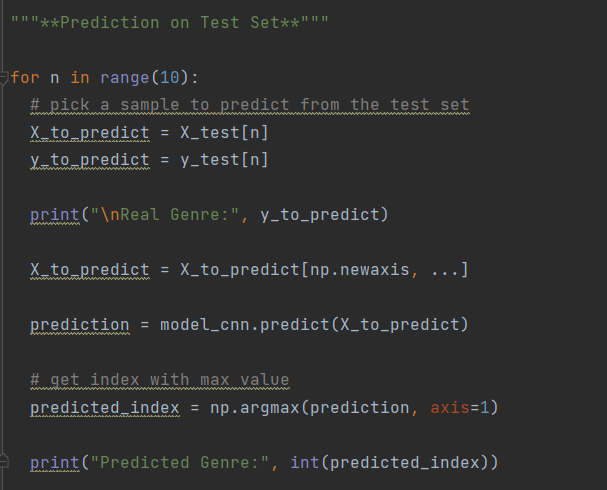
**Model Building and Training:**

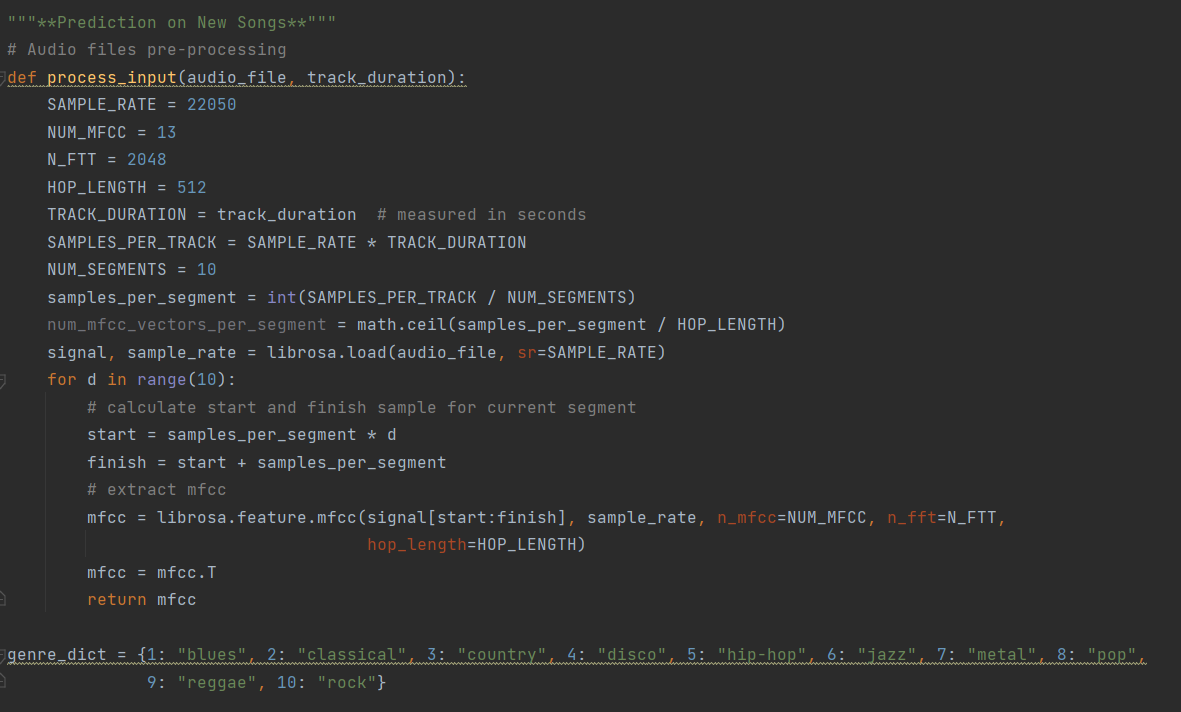
****

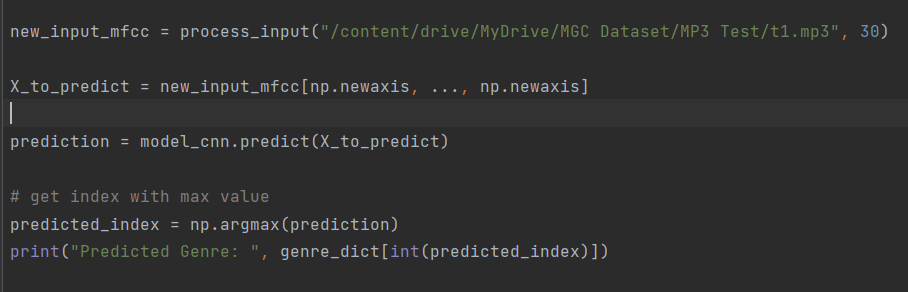
****

****

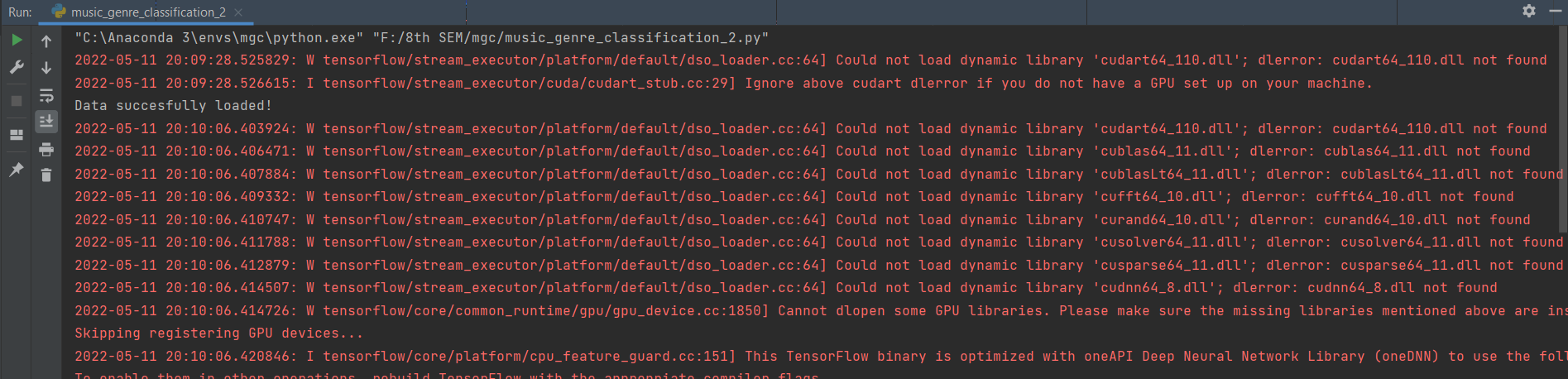
****

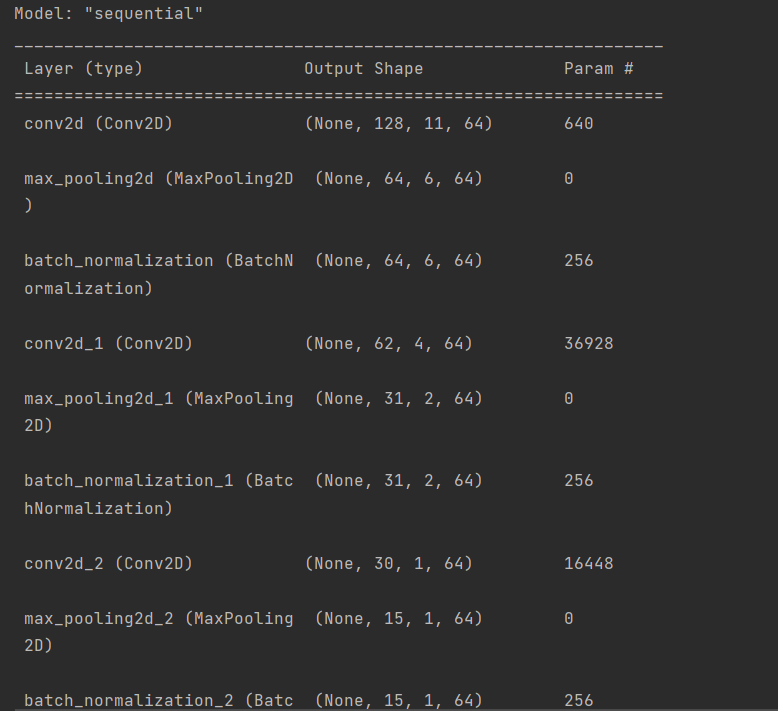
****

****

****

**Output:**





**Bibliography**

[1] Snigdha Chillara, Kavitha A S, Shwetha A Neginhal, Shreya Haldia, Vidyullatha K, Music “Genre Classification using Machine Learning Algorithms: A comparison”

[2] R. Thiruvengatanadhan, “Music Classification using MFCC and SVM”

[3] Rajeeva Shreedhara Bhat, Rohit B. R, Mamatha K , “Music Genre Classification”

[4] Derek A. Huang, Arianna A. Serafini, Eli J, “Music Genre Classification”