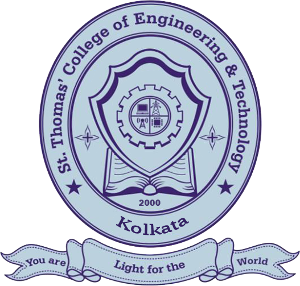
**St. Thomas’ College of Engineering and Technology**



**Music Genre Classification**

**Prepared By**

**Akash Bose (Univ. Roll: 12200217061, Reg. No.: 171220110067)**

**and**

**Anirban Pal (Univ. Roll: 12200217055, Reg. No.: 171220110073)**

**Under the guidance of**

**Dr. Arijit Ghosal­**

**(Assoc. Prof, IT Dept)**

**Dr. Ranjit Ghosal­**

**(Asst. Prof, IT Dept)**

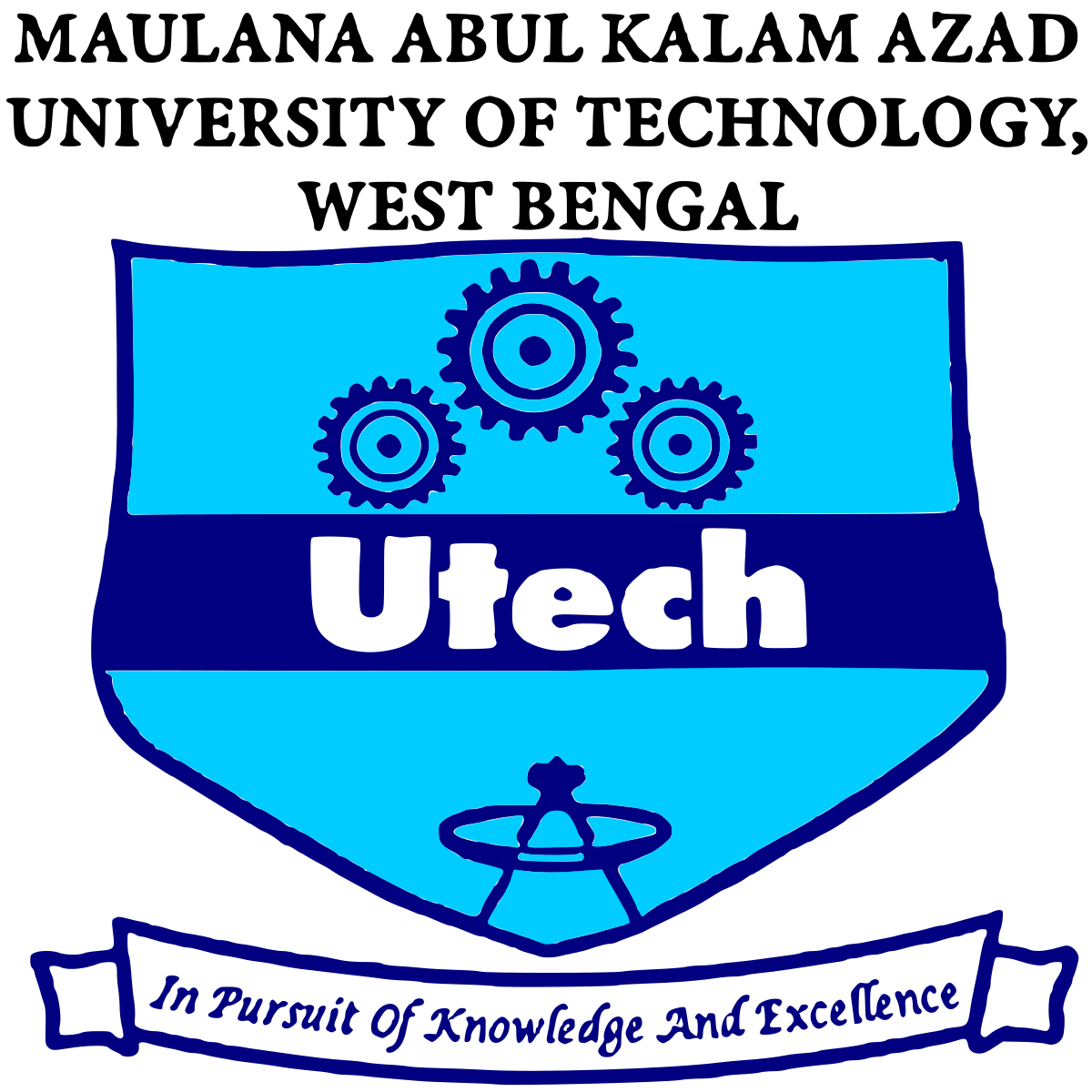
**Project Report**

**Submitted in partial fulfillment of the requirement for the degree of B.Tech in Information Technology**

**Department of Information Technology**

**Affiliated to**

**Maulana Abul Kalam Azad University of Technology, West Bengal**



**June, 2021**

**This is to certify that the work in preparing the project entitled Music Genre Classification has been carried out by Akash Bose and Anirban Pal under my guidance during the session 2020-2021 and accepted in partial fulfillment of the requirement for the degree of Bachelor of Technology in Information Technology.**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Dr. Arindam Chakravorty**

**Head, Department of Information Technology**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Dr. Arijit Ghosal Dr. Ranjit Ghoshal**

**Department of Information Technology Department of Information Technology**

**Acknowledgement**

I am over helmed in all humbleness and gratefulness to acknowledge my depth to everyone who have helped me to put these ideas, well above the level of simplicity and into something concrete. I would like to express my special thanks of gratitude to Dr. Arijit Ghosal for his immense help in guiding us throughout. I would also like to take the opportunity to thank Dr. Ranjit Ghoshal for mentoring and guiding us. I also would like to thank my parents and friends who have motivated me and inspired me during these difficult times. The project would not have been achievable if not for the aforementioned assist.

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

**Akash Bose Anirban Pal**

**Vision and Mission of the College**

**Vision:**

To evolve itself into an industry oriented research based recognized hub of creative solutions in various fields of engineering by establishing progressive teaching-learning process with an ultimate objective of meeting technological challenges faced by the nation and the society

**Mission:**

* To create opportunities for students and faculty members in acquiring professional knowledge and developing social attitudes with ethical and moral values
* To enhance the quality of engineering education through accessible comprehensive, industry and research oriented teaching-learning process
* To satisfy the ever-changing needs of the nation for evolution and absorption of sustainable and environment friendly technologies

**Vision and Mission of the Department**

**Vision:**

To promote the advancement of learning in Information Technology through research-oriented dissemination of knowledge which will lead to innovative applications of information in industry and society.

**Mission:**

* To incubate students, grow into industry ready professionals, proficient research scholars and enterprising entrepreneurs.
* To create a learner-centric environment that motivates the students in adopting emerging technologies of the rapidly changing information society.
* To promote social, environmental and technological responsiveness among the members of the faculty and student.

**PEO:**

PEO1: Exhibit the skills and knowledge required to design, develop and implement IT solutions for real life problems.

PEO2: Excel in professional career, higher education and research.

PEO3: Demonstrate professionalism, entrepreneurship, ethical behavior, communication skills and collaborative team work to adapt the emerging trends by engaging in lifelong learning.

**PO: Project Mapping with Program Outcomes**

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **PO1** | **PO2** | **PO3** | **PO4** | **PO5** | **PO6** | **PO7** | **PO8** | **PO9** | **PO10** | **PO11** | **PO12** |
| 3 | 3 | 2 | 3 | 3 | 3 | - | 3 | 2 | 3 | 3 | 3 |

**Correlation levels 1, 2 or 3 as defined below:**

**1: Slight (Low) 2: Moderate (Medium) 3: Substantial (High)**

**Justification:**

1. Engineering knowledge: Application of the knowledge of mathematics, science, engineering fundamentals, and an engineering specialization to the solution of complex engineering problems is required in this project, hence, it satisfies PO1.
2. Problem analysis: Identifying, formulating, reviewing research literature, and analyzing complex engineering problems reaching substantiated conclusions using first principles of mathematics, natural sciences, and engineering sciences is highly required in this project, hence, it satisfies PO2.
3. Design/development of solutions: Designing solutions for complex engineering problems and designing system components or processes that meet the specified needs with appropriate consideration for the public health and safety, and the cultural, societal, and environmental considerations is required to be done in this project, thus, PO3 stands applicable.
4. Conduct investigations of complex problems: Using research-based knowledge and research methods including design of experiments, analysis and interpretation of data, and synthesis of the information to provide valid conclusions is required to a great extent. PO4 stands applicable.
5. Modern tool usage: Appropriate modern tools and methods are used. Hence PO5 stands applicable.
6. The engineer and society: Application of reasoning informed by the contextual knowledge to assess cultural issues and the consequent responsibilities relevant to the professional engineering practice is recognized in this project, hence, PO6 stands applicable.
7. Environment and sustainability: This project’s impact is more cultural and technology based, hence, PO7 is not applicable.
8. Ethics: Applying ethical principles and commitment to professional ethics and responsibilities and norms of the engineering practice is required. PO8 is applicable.
9. Individual and team work: Functioning effectively as an individual, and as a member or leader in diverse teams, and in multidisciplinary settings: this is applicable here.
10. Communication: Communicating effectively on complex engineering activities with the engineering community and with society at large, such as, being able to comprehend and write effective reports and design documentation, making effective presentations, and giving and receiving clear instructions: PO10 is substantially applicable in this project.
11. Project management and finance: Demonstrating knowledge and understanding of the engineering and management principles and applying these to one’s own work, as a member and leader in a team, to manage projects and in multidisciplinary environments: this is highly applicable in this project, hence PO11 has high correlation level (3)
12. Life-long learning: Recognizing the need for, and having the preparation and ability to engage in independent and life-long learning in the broadest context of technological change: this is relevant with respect to this project, hence, PO12 has high correlation level of 3.

**PSO: Project Mapping with Program Specific outcomes**

|  |  |  |
| --- | --- | --- |
| **PSO1** | **PSO2** | **PSO3** |
| 3 | - | 3 |

**Correlation levels 1, 2 or 3 are as defined below:**

**1: Slight (Low) 2: Moderate (Medium) 3: Substantial (High)**

**Justification:**

1. This project requires me to apply programming knowledge to build an efficient and effective solution of the problem with an error free, well documented and reusable code, user friendly interface and well-organized database. Hence, the project substantially satisfies PSO1.
2. Creation of multimedia enabled web solutions using information in different forms for business, education and the society at large isn’t applicable in this project, hence, PSO2 stands irrelevant/ not applicable.
3. Understanding and analysing a big complex problem and decomposing it into relatively smaller and independent modules algorithmically is done in this project, hence, PSO3 is satisfied substantially.

**Index Page**

**Contents Page No.**

1. **Introduction . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 1**
   1. **Problem Statement**
   2. **Problem Definition**
   3. **Objective**
   4. **Literature Survey/ Background Study**
   5. **Brief Discussion on Problem**
2. **Concepts and Problem Analysis . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 1**
3. **Data Preparation . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 2**
   1. **Dataset**
   2. **Feature Extraction**
   3. **Feature Selection**
   4. **Data Processing**
4. **Classification Models . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 6**
   1. **K-Nearest Neighbours**
   2. **Logistic Regression**
   3. **Support Vector Machines**
   4. **Decision Trees**
   5. **Random Forest**
   6. **Artificial Neural Network**
   7. **Convolutional Neural Network**
5. **Testing . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 13**
6. **Deployment . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 15**
7. **Conclusion . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 17**
8. **Future Planning . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 17**

**Annexure: Gantt chart for project planning**

**Reference**

**1. Introduction**

**1.1 Problem Statement:** Music Genre Classification using Machine Learning.

**1.2 Problem Definition:** The different genres of music are identified based on their characteristic features with the help of different classification techniques.

**1.3 Objective:** The aim of this project is to build a model which correctly recognizes the genre of the music based on their characteristic audio features using Machine Learning. This project extends to building an interface which predicts the genre of a given music file using that model.

**1.4 Literature Survey:**

1. Miguel Flores Ruiz de Eguino, “Deep Music Genre”, Stanford University, 2016.
2. D. A. Huang, A. A. Serafini, E. J. Pugh, “Music Genre Classification”, Stanford University, 2019.
3. R. S. Bhat, Rohit B. R., Mamatha K. R., “Music Genre Classification”, BMS College of Engineering, 2020.
4. M. Haggblade, Y. Hong, K. Kao, “Music Genre Classification”, Stanford University, 2011.
5. S. Zhang, H. Gu, R. Li, “Music Genre Classification: Near-Realtime vs Sequential Approach”, 2019.

**1.5 Brief Discussion on Problem:** With the advent of digital devices and media streaming platforms, music has become ever more popular. People have access to millions of songs by numerous artists spanning over different genres. The average user has evolved his taste for a particular genre rather than a taste for a single band or an artist. This taste for the likeable genres has heightened the use of music playlists based on genres. Companies nowadays use music classification, either to be able to place recommendations to their customers such as Spotify and SoundCloud or simply as a product like Shazam. Determining music genres is the first step in that direction. Machine Learning techniques have proved to be quite successful in extracting trends and patterns from the large pool of audio data.

**2. Concepts and Problem Analysis**

This project is an application of Machine Learning. Machine Learning is the use of algorithms so as to compel a system to learn and adapt on its own from previous experiences and draw inferences from patterns in data. A general machine learning block diagram is illustrated in Fig 1. A typical machine learning problem starts with data acquisition followed by pre-processing and preparation of the data suitably. The prepared data is used for extracting features and the dataset is split into training and testing sets. The machine learning model is trained using this training set and the trained model is tested using the test set. This is followed by model evaluation which involves calculating the accuracy and precision of the trained model. This evaluation can be further used to fine tune the model until a desired accuracy is achieved that can be deployed

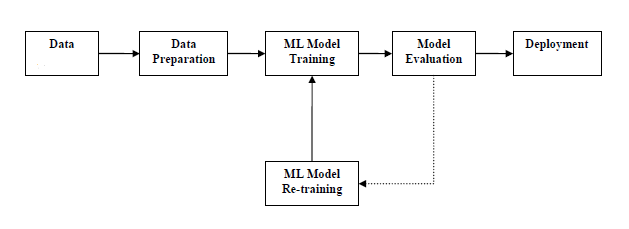


Fig. 1 A typical machine learning block diagram

**3. Data Preparation:**

**3.1 Dataset:**

The dataset used in this project is the GTZAN Genre Collection Dataset. It has been designed and written by George Tzanetaki. The dataset has been taken from the popular software framework MARSYAS[1]. Marsyas (Music Analysis, Retrieval and Synthesis for Audio Signals) is an open source software framework for audio processing with specific emphasis on Music Information Retrieval applications.

Dataset consists of 1000 audio tracks each 30 seconds long. It contains 10 genres (Blues, Classical, Country, Disco, Hip-Hop, Jazz, Metal, Pop, Reggae and Rock), each represented by 100 tracks. The tracks are all 22050 Hz Mono 16-bit audio files in .wav format.

**3.2 Feature Extraction:**

Sound is represented in the form of an **audio** signal having parameters such as frequency, bandwidth, decibel etc. A typical audio signal can be expressed as a function of Amplitude and Time. The sound excerpts are digital audio files in .wav format.).A typical audio processing process involves the extraction of acoustics features relevant to the task at hand, followed by decision-making schemes that involve detection, classification, and knowledge fusion. Thankfully we have some useful python libraries which make this task easier.

In this project Librosa [2] has been used to read the digital audio .wav files. Librosa is a python package for music and audio analysis. It provides the building blocks necessary to create music information retrieval system. The package has the necessary functions to extract the features that are used for the genre classification

**3.2 Feature Selection:**

Each and every audio signal consists of a load of features that can be extracted. However the characteristics that are relevant to the problem have been only considered for the extraction. This removes overfitting of the data, improves accuracy and also reduces training time. The major features used are described below.

**3.2.1 Zero Crossing Rate (ZCR):** ZCR is the rate of sign-changes along a signal, i.e., the rate at which the signal changes from positive to negative, or vice-versa. The ZCR values are in the range of 0-1. The values are related to frequencies. High ZCR value corresponds to a high frequency signal portion and vice-versa. ZCR is a good feature to discriminate among audio-silence/noise, speech/noise. It usually has higher values for metal and rock genres. So, ZCR has been considered in this project.

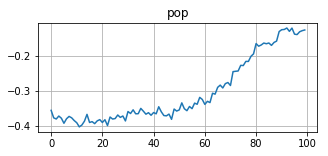
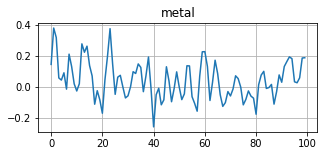
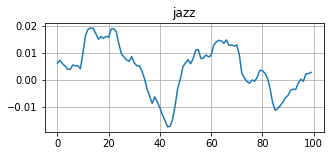
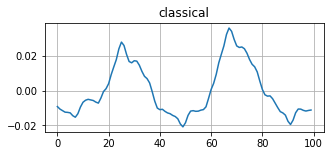
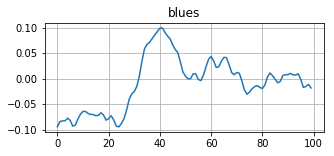


Fig 2 Variation of ZCR among different genres

**3.2.2 Spectral Centroid:** It indicates where the “centre of mass” for a sound is located and is calculated as the weighted mean of the frequencies present in the sound. If two songs, one from a blues genre and the other belonging to metal are compared ,the blues genre song which is the same throughout its length, will have the spectral centroid lie somewhere near the middle of its spectrum while that for a metal song would be towards either of its end.

**3.2.3 Spectral Roll-Off:** It is a measure of the shape of the signal. It represents the frequency below which a specified percentage of the total spectral energy, e.g. 85%, lies. This can help distinguish between a rock or metal, where the specified percentage of energy will correspond to a higher frequency to that of blues or classical where the spectral energy is concentrated at lower frequencies.

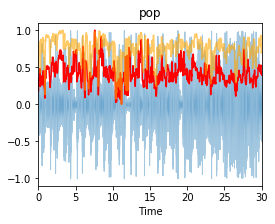
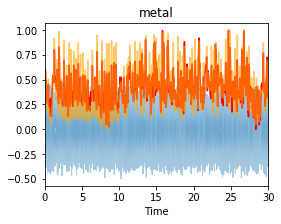
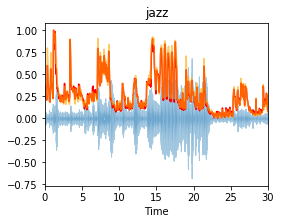
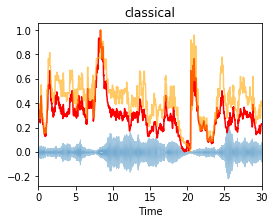
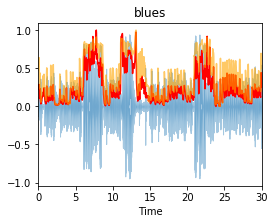


Fig 3 Variation of Spectral Rolloff (Orange) and Spectral Centroid (Red)

**3.2.4 Mel Frequency Cepstral Coefficients:** The Mel frequency cepstral coefficients (MFCCs) of a signal are a small set of features (usually about 10–20) which concisely describe the overall shape of a spectral envelope. It models the characteristics of the human voice. The MFCCs encode the timbral properties of the music signal by encoding the rough shape of the log-power spectrum on the Mel frequency scale.

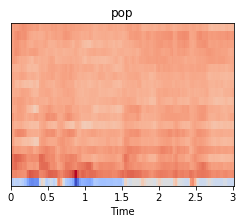
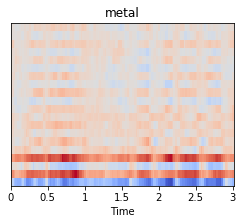
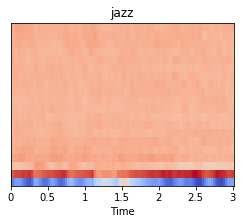
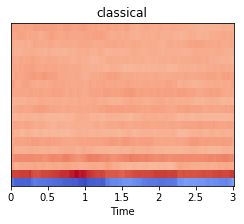
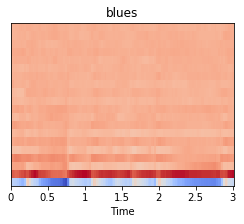


Fig 4 Variation of MFCC among different genres

**3.2.5 Chroma Features:** Chroma features are an interesting and powerful representation for music audio in which the entire spectrum is projected onto 12 bins representing the 12 distinct semitones (or chroma) of the musical octave.  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.

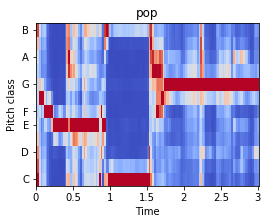
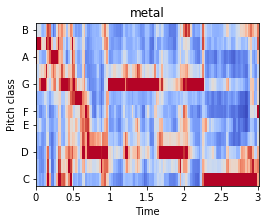
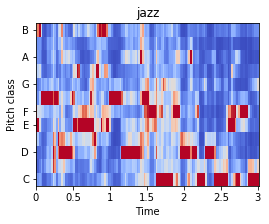
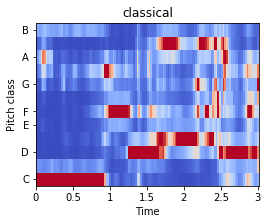
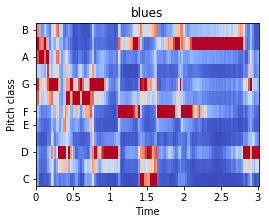


Fig 5 Chroma features of different genres

**3.2.6 Mel-Spectrogram:** The Mel-frequency cepstrum (MFC) is a representation of the short-term power spectrum of a sound, based on a linear cosine transform of a log power spectrum on a nonlinear Mel scale of frequency. Mel-Spectrogram is a spectrogram with the Mel scale expressed in dB as its y-axis. The Mel-spectrograms have been used for the CNN classification model.

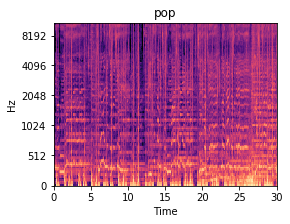
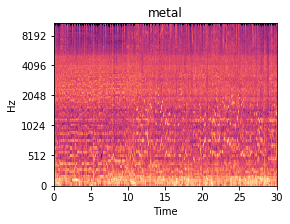
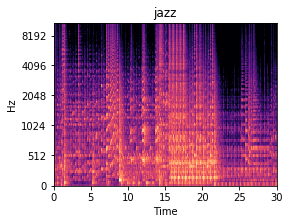
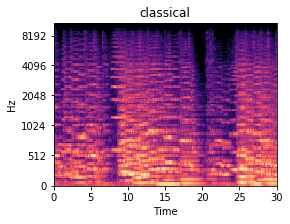
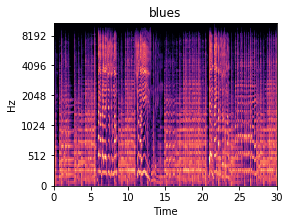


Fig 6 Mel-spectrograms of different genres

**3.2.6 Other Features:** Some other features like Spectral Bandwidth, Harmony, Percussive and Tempo were also considered for the classification purposes.

**3.4 Data Processing:** The raw data extracted was required to be processed before it could be used for the training purposes. This was done in order to correct for distortions and errors like radiometric errors in the data. A lot of the features were scaled to make the data on the same scale. Methods like the Standard scaler and min-max scaler from the Scikit-learn library were employed for feature scaling. Also while preparing the dataset for testing with external audios data processing was done for converting the audios to 16bit mono 22050 Hz .wav files. This was done using Audacity.

**4. Classification Models:**

In machine learning, classification refers to a predictive modeling problem where a class label is predicted for a given example of input data. A model will use the training dataset and will calculate how to best map examples of input data to specific class labels. As such the training dataset must be sufficiently representative of the problem and have many examples of each class label. The models used for the classification are described below.

**4.1 K-nearest Neighbours:** The k-nearest neighbours (KNN) algorithm is a simple, easy-to-implement supervised machine learning algorithm that can be used to solve classification problems. In k-nearest neighbours algorithm an object is classified by a vote of its neighbours, with the object being assigned to the class which is most common among its k nearest neighbours. If k = 1, then the object’s class is that of the single nearest neighbour.

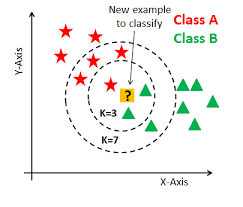


Fig. 7: Diagram of knn classification

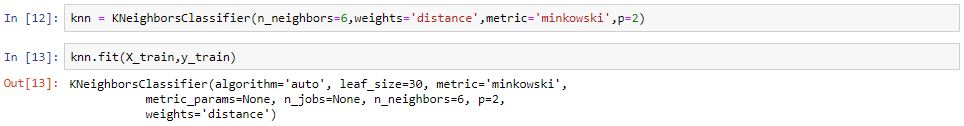


Fig. 8: Knn model using scikit-learn

**4.2 Logistic Regression:** Logistic regression is named for the function used at the core of the method, the logistic function.In order to map predicted values to probabilities, we use the sigmoid function. The function maps any real value into another value between 0 and 1. In machine learning, we use sigmoid to map predictions to probabilities. The sigmoid curve can be represented with the help of following graph. We can see the values of y-axis lie between 0 and 1 and crosses the axis at 0.5.

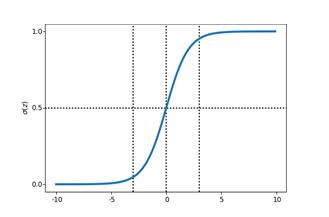


Fig. 9: Diagram of sigmoid function f(x) = 1/(1 + e-x)

We then calculate the likelihood of parameters. Likelihood is the probability of data, given a model and specific parameter values i.e. it measures the support provided by the data for each possible value of the regression coefficients. Finally we calculate the cost function and apply gradient descent algorithm to update the regression coefficients to achieve minimum cost value.

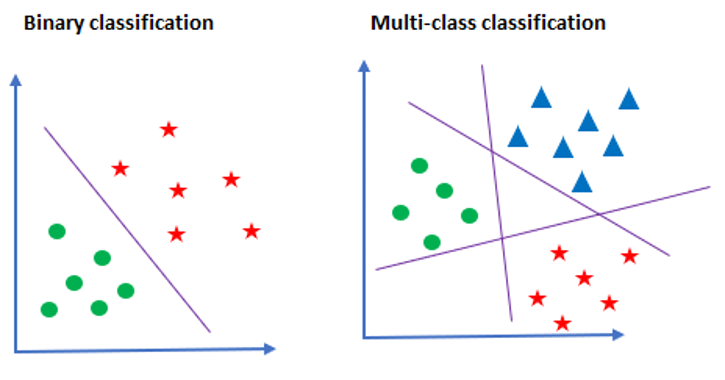


Fig 10: Diagram representing OneVOne and OneVRest Classification

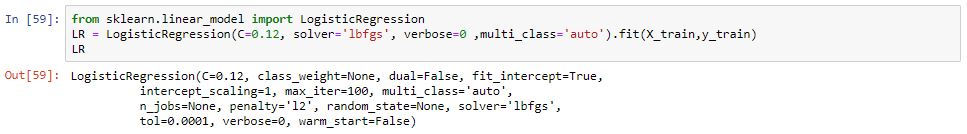


Fig 11: LR model using scikit-learn

**4.3 Support Vector Machine:** Support Vector Machine is one of the most preferred classification algorithms used for classification. SVM finds a hyper-plane that creates a boundary between two different classes of data. In two dimensional space, this hyper-plane is just a line. In SVM, each data is plotted in an n-dimensional space, where n is the number of features. Next we calculate the optimal hyper-plane to separate the data. SVM is usually used for binary classification but for multiclass classification we can create a SVM classifier for each class and the classifier with the highest score is chosen as the output of the SVM.[3]

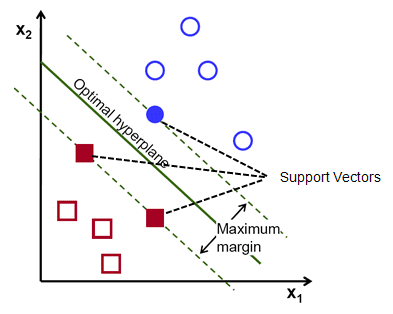


Fig 12: Diagram representing svm classification

There are some important parameters in support vector classifier. The kernel is selected based on the type of data used and the type of transformation. The gamma value determines how far the influence of a single training example reaches during transformation. The ‘c’ parameter determines the extent of regularization applied on the data.

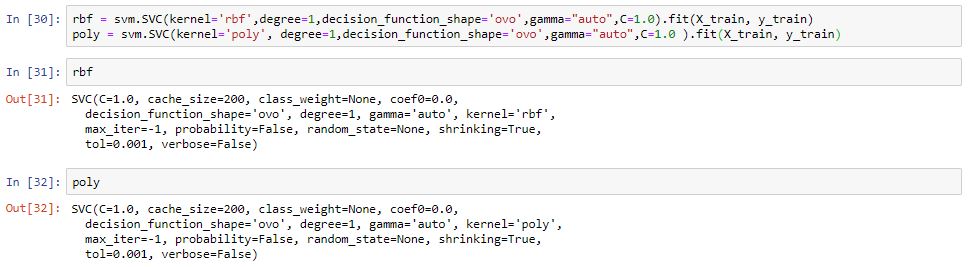


Fig 13: SVM model using scikit-learn

**4.4 Decision Trees :** Decision trees are a supervised method used for classification of data. It creates a model that can predict a target value by learning from a set of decision rules inferred from the data itself. The greater the depth of the tree the greater is its complexity. Decision trees are easier to interpret and understand. Decision tree has a flowchart like tree structure where each internal node represents a test decision, each branch represents an outcome of a decision, and each leaf node holds a class label.

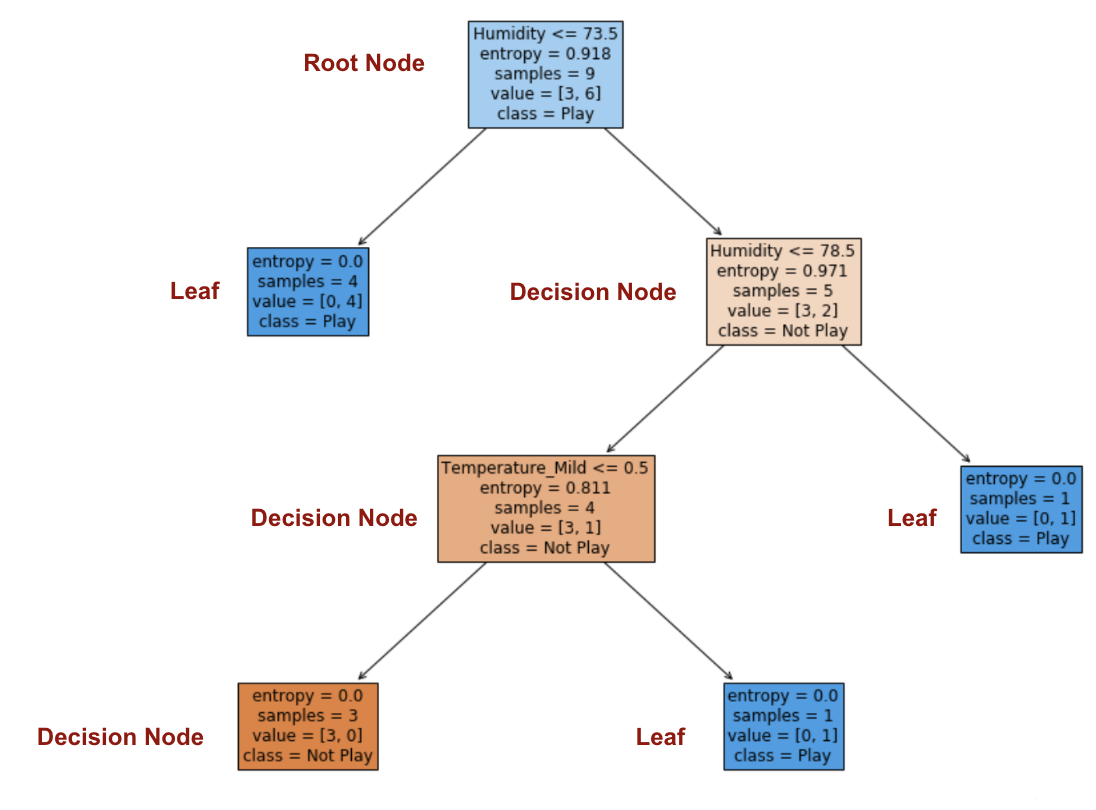


Fig 14: Diagram demonstrating how decision tree works

There are several methods use to evaluate decision tree splitting. One of them is Information Gain which is equal to 1 – Entropy where Entropy determines the purity of the node. Entropy = -∑i = 1 to n (pilog2pi). Another is Gini Impurity which is equal to 1 – Gini. Gini determines the probability of correctly labeling a randomly chosen element. Gini = ∑i = 1 to n (pi)2.

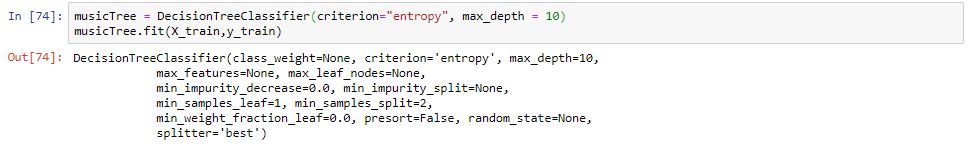


Fig 15: DT model using scikit-learn

**4.5 Random Forest:** The main idea behind random forest is that it is a collection of large number of decision trees that operate together to predict a class. Each individual decision tree in the random forest predicts a class and the class with the most number of votes becomes the predicted output of the random forest model. In our random forest model using scikit-learn the total number of trees in the forest i.e. estimators = 1000, the splitting evaluation method i.e. criterion is chosen as ‘gini’ and max\_depth is set to 10.

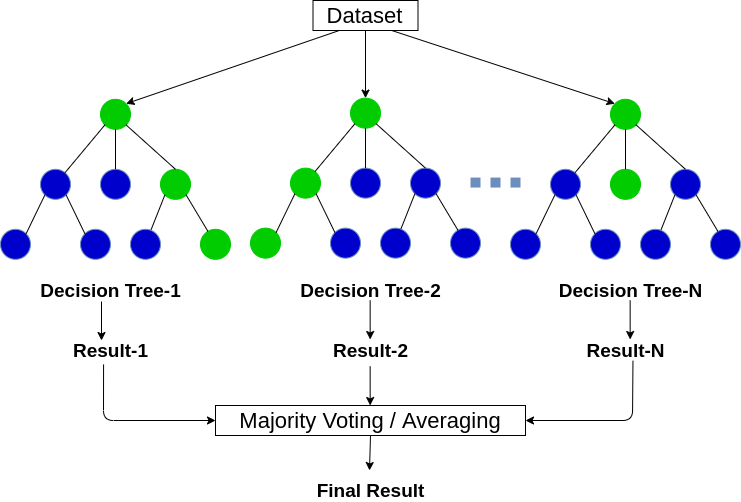


Fig 16: Diagram demonstrating working principle of random forest

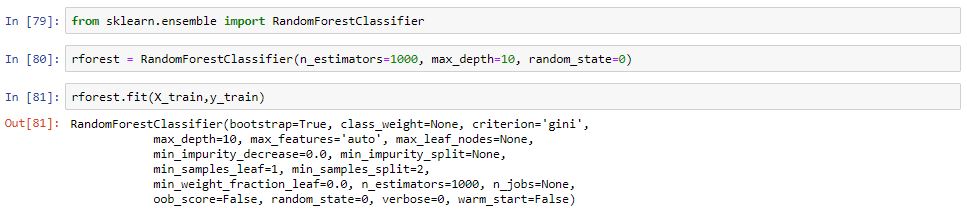


Fig 17: Random Forest model using scikit-learn

**4.6 Artificial Neural Network:** An artificial neural network is like a human brain which consists of several nodes or neurons interconnected like a web. An artificial neural network consists of an input layer, an output layer and one or more hidden layers. The input layer receives several input data of different shapes and forms, the hidden layers extract various information from the input data and the model learns about the information presented and finally the output layer predicts the output. During supervised phase the Ann model computes the error of the predicted values from the actual values. Then it uses backpropagation to update the weights and biases of the connections between the processing units unit the error value is minimum. Each neuron or processing unit takes an input value, then processes the data using weights and biases and finally passes it through an activation function and converts the input into a more useful output.[4]

The following image shows how a neuron or processing unit works.

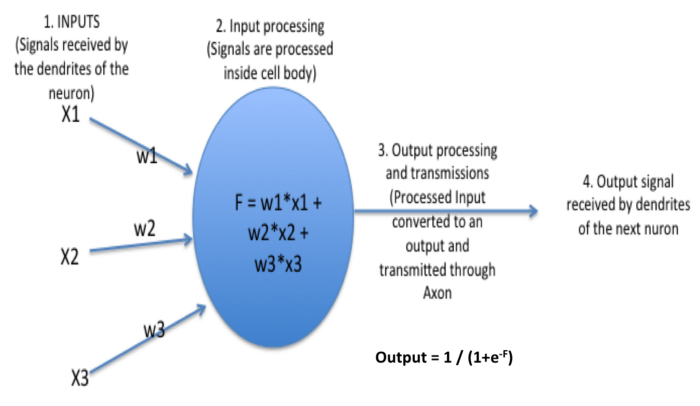


Fig 18: Diagram of a single processing unit

The following diagram shows how an artificial neural network looks like.



Fig 19: Diagram of an artificial neural network

The following diagram represents our artificial neural network model created using Keras.

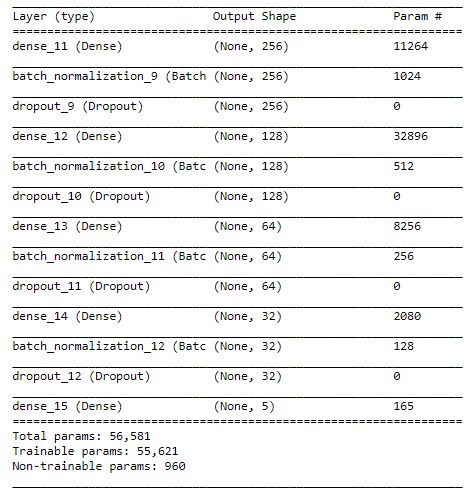


Fig 20: Keras regression model.

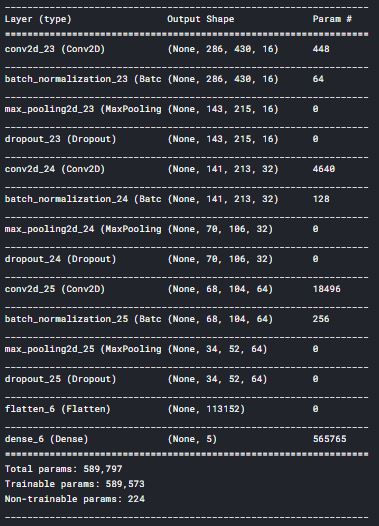
**4.7 Convolutional Neural Network:** Convolutional neural network is a deep learning algorithm that takes in images(in this case 286x430 sized image of a mel-spectrogram of an audio signal) as its input, extracts various features and trains the model updating its weights and biases at each layer until the error at the output is minimized. The following diagram represents a general onvolutional neural network.



Fig 21: Diagram of a onvolutional neural network

Convolution is a process where several filters extract high-level features such as edges from the input images. Max Pooling returns the maximum value from a portion of the convolved image to reduce the spatial size. Batch normalization normalizes the output of previous layer so that mean is 0 and variance is 1 and then rescales it to stabilize the data. Fully connected layer flattens the image into a single vector and finally we apply softmax function to predict the output. Following image shows our keras cnn model.

Fig 22: Keras cnn model



**5. Testing:**

GTZAN dataset has 100 audio files of each genre. We have also segmented the 30sec audio files into 10sec and 3sec audio files so that for each genre the dataset size increases to 300 audio files and 1000 audio files respectively for each segment. We have calculated the training and validation accuracy of each model for five genres i.e. blues, classical, jazz, metal, and pop. We then selected the best model based on validation accuracy and then tested our model with unknown data i.e. audio files external to GTZAN dataset. The following table shows a comparative study of accuracy achieved by different classifiers.

|  |  |  |
| --- | --- | --- |
| Models | Segment | Validation Accuracy |
|  | 30sec | 90% |
| KNN | 10sec | 97% |
|  | 3sec | 96% |
|  | 30sec | 89% |
| LR | 10sec | 92% |
|  | 3sec | 93% |
|  | 30sec | 89% |
| SVM( kernel = rbf ) | 10sec | 94% |
|  | 3sec | 95% |
|  | 30sec | 89% |
| SVM ( kernel = poly ) | 10sec | 92% |
|  | 3sec | 92% |
|  | 30sec | 78% |
| DT | 10sec | 84% |
|  | 3sec | 88% |
|  | 30sec | 87% |
| RF | 10sec | 92% |
|  | 3sec | 94% |
|  | 30sec | 88% |
| Keras Regression | 10sec | 95% |
|  | 3sec | 96% |
| Keras CNN | 30sec | 81% |

Table 1: Comparative analysis of accuracy achieved by different classifiers.

From the above table it is clear that k-nearest neighbours classifier and our keras regression model gave the best results. The following images represent the confusion matrix created using the actual and the predicted music genre classes for both the knn and keras model.

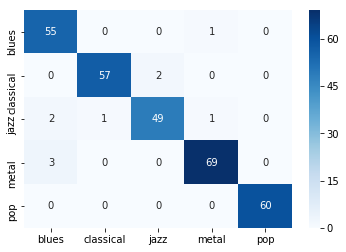
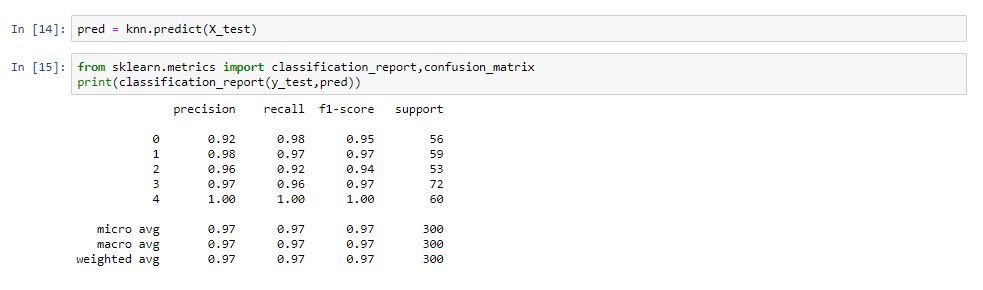


Fig 23: KNN accuracy (segment=10sec) Fig 24: KNN confusion matrix (segment=10sec)

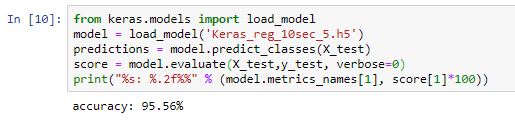


Fig 25: Keras regression model accuracy (segment=10sec)

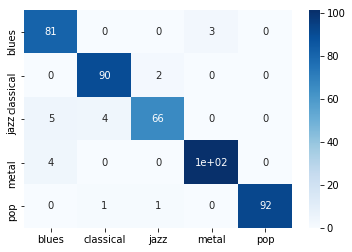


Fig 26: Confusion Matrix for Keras model output (segment=10sec)

We created a custom dataset consisting of 29 songs that are unknown to the dataset. We converted the songs to 22050Hz 16 bit mono .wav files before testing using Audacity software. We then created a dataset by extracting various features using Librosa. Finally we used keras regression model to test and see how good it is working for data unknown to the dataset.

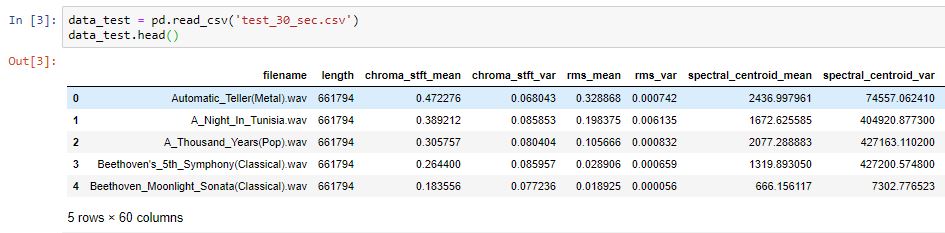


Fig 27: Custom testing dataset

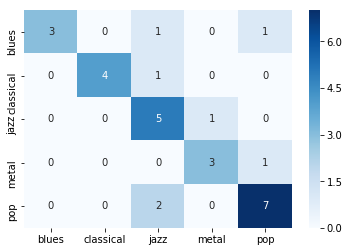


Fig 28: Confusion matrix demonstrating keras regression model performance for above dataset

**6. Deployment:**

We created a web application using HTML, CSS, Flask, Jinja2, Python, Scikit-learn libraries and Keras. The web application is deployed on Heroku a cloud application platform. The web application accepts an audio file as an input then shows the predicted genre class using previously trained models. This is a link to our web application: <https://music-genre-detection.herokuapp.com> . The following images are snapshots of our web application.

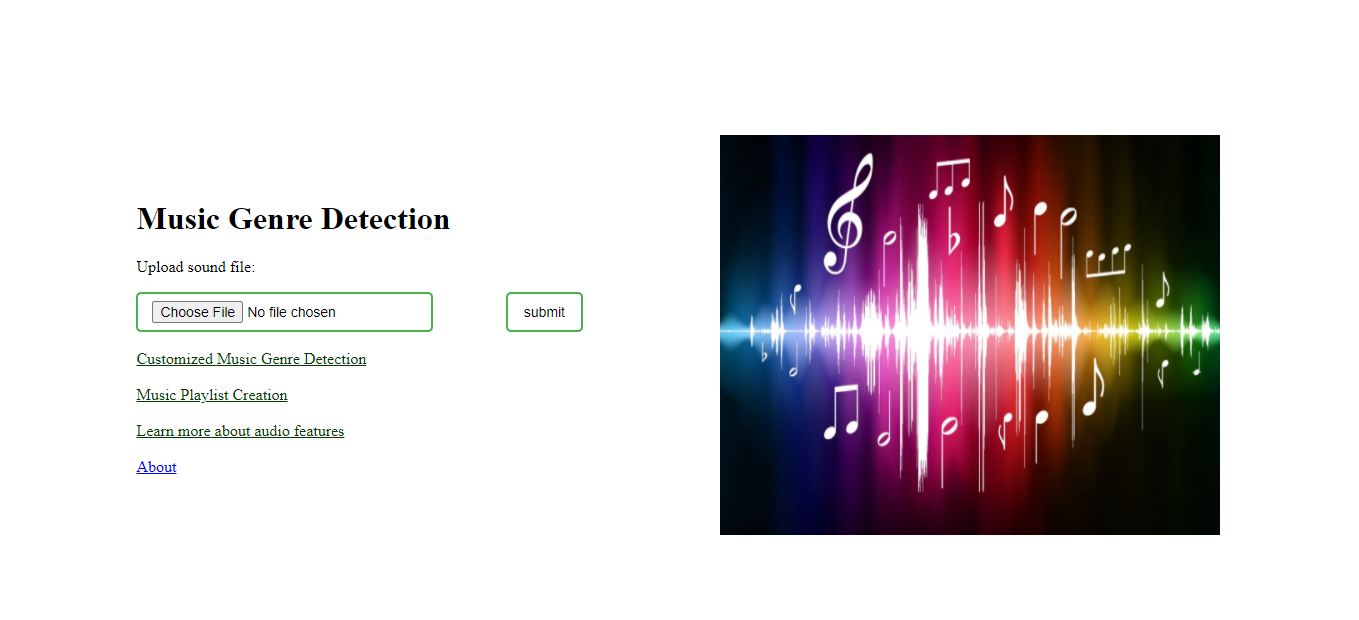


Fig 29: Home page of our web application.

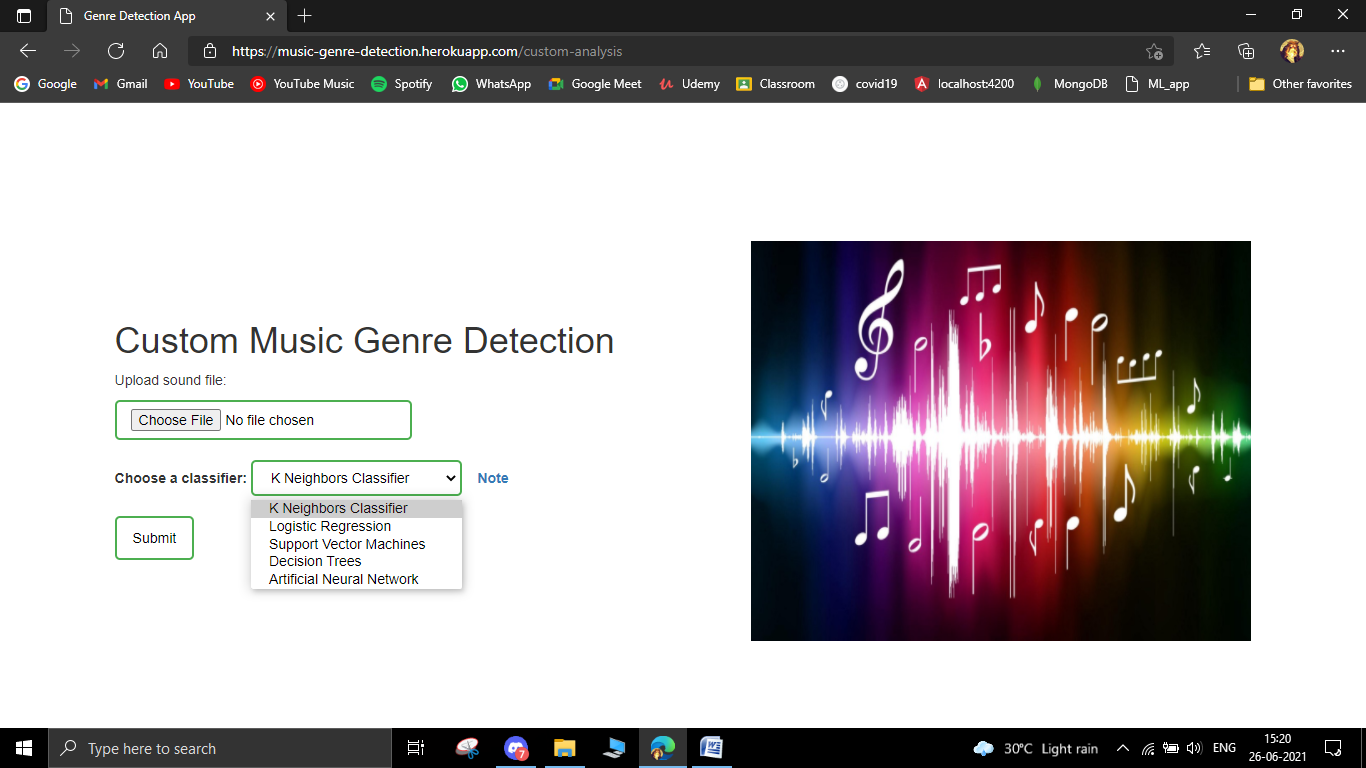


Fig 30: User can choose his/her own classifier.

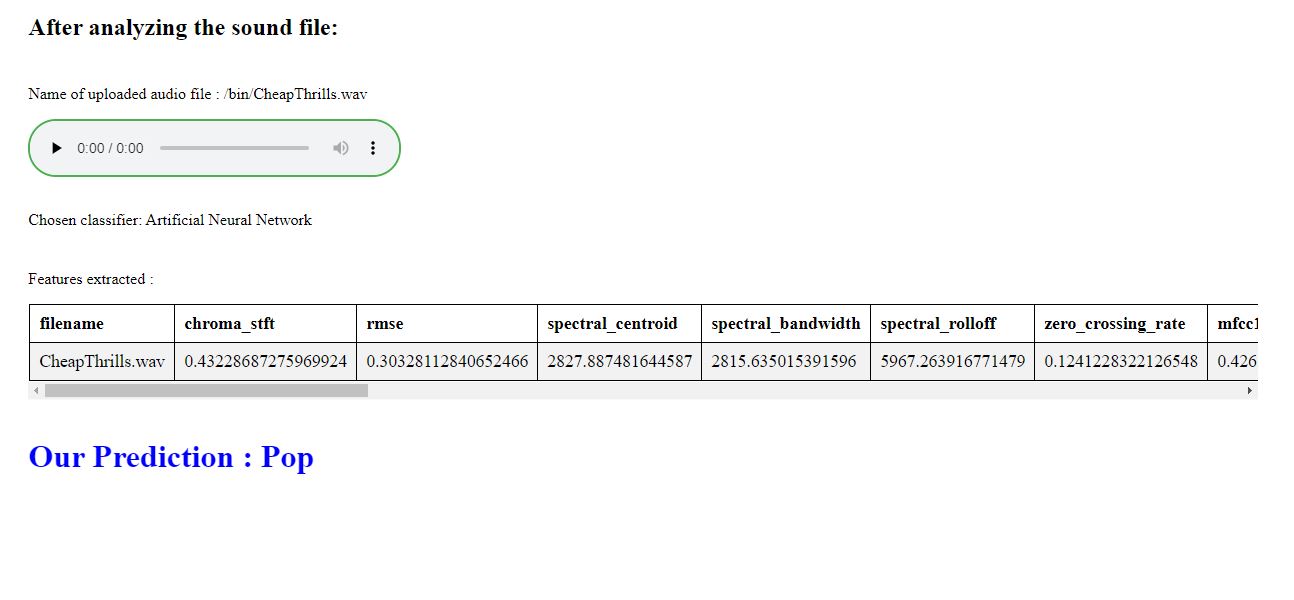


Fig 31: Webpage displaying genre prediction.

**7. Conclusion**

The project is an application of machine learning which a trending technology is. This project aiming at categorizing the music based on genres finds various applications. We worked with only five genres and used classification techniques like KNN, Logistic Regression, Support Vector Machine, Decision Tree, Random Forest, Keras Regression Model and Convolutional neural network. We are satisfied with the accuracy achieved by different classification models.

**8. Future Planning:**

The classification has been done with five genres. So we need to work on building a model that can give a good accuracy while working with all ten genres. Also we need to see how we can use LSTM to improve our accuracy. We can work on a model in future that works as an ensemble consisting of multiple classifiers and the most voted prediction will be shown as the output. We also need to improve the web application in terms of response time i.e. it will take less time to process the uploaded audio file to predict the output. We can also include another feature in the web application that can create a playlist based on genre.

**Gantt Chart for Project Planning:**

|  |  |
| --- | --- |
| **REQUIREMENT ANALYSIS,**  **FEASIBILITY STUDY** | |
|  |  | **LITERATURE SURVEY** |
|  |  |  | **DATA ACQUISITION & FEATURE EXTRACTION** | |
|  |  |  |  |  | **CODING** |
|  |  |  |  |  |  | **TESTING & ANALYSIS** |
|  |  |  |  |  |  |  | **DEPLOYMENT**  **&**  **MAINTAINENCE** |
| **SEPTEMBER**  **2020** | **OCTOBER 2020** | **NOVEMBER**  **2020** | **DECEMBER 2020** | **JANUARY**  **2021** | **FEBRUARY**  **2021** | **MARCH-APRIL**  **2021** | **MAY-JUNE**  **2021** |

Fig 32: Gnatt Chart

**References**

[1] “ Musical genre classification of audio signals ” by G. Tzanetakis and P. Cook in IEEE Transactions on Audio and Speech Processing 2002.

[2]Documentation of Librosa, a python package for audio handling. Retrieved from https://librosa.org/doc/latest/index.html

[3] “Understanding Support Vector Machine algorithm from examples” by Sunil Ray. Retrieved from https://www.analyticsvidhya.com/blog/2017/09/understaing-support-vector-machine-example-code/

[4] “Introduction to Neural Networks, Advantages and Applications” by Jahnavi Mahanta. Retrieved from https://towardsdatascience.com/introduction-to-neural-networks-advantages-and-applications-96851bd1a207