Sumukh Y S, 2021, 9:4
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

Music Emotion Detection Using Machine Learning

Sumukh Y S, Sai Gagan N V, Srujan M S, Subhanshu Singh, Asst. Prof. Mrs. Shubha T V
Department of Computer Science and Engineering,
SJB Institute of Technology, Bengaluru, India
yssumukh@gmail.com, gagannvs@gmail.com, sudhanshusingh00712345@gmail.com, srujanms43@gmail.com, Shubhatv@sjbit.edu.in

Abstract- Music has a large influence on the listeners' emotion. Classification of music based on emotion can play an important role in various aspects. People dealing with anxiety can get some help by listening to calm and relaxing music. One's whole day can be made by listening to happy music. There are many such advantages of classifying music based on emotions. For this to happen, we need to recognize the emotion in music. In this paper, we go through a probabilistic classification approach to recognize the emotion in music using Gaussian Naïve Bayes' classifier.

Keywords- Machine Learning, Music Emotion, Sentiment Analysis.

I. INTRODUCTION

Machine Learning is one of the fields that rapidly developing. There are various areas where machine learning can be used. One of those areas is Sentiment analysis. We can analyze sentiments/ emotions from various data like image, text, audio, etc.

Machine Learning can be used on audio files for various applications. AI Audio is one of the fields within Artificial Intelligence that have seen significant growth in recent days.

In this paper we are focusing on sentiment analysis of music. Music can also be considered as a way to convey or express emotions. Music can also influence one's emotion. If a person listens to happy music early in the morning, then his/her whole day goes well as they will carry the happy mood throughout the day. If someone is having a rough day, they can listen to calm music to feel better. One can up his/her speed of working by listening to energetic music. Music can be used for various purposes from simple relaxation to psychological treatments.

We have come up with a project that involves music, emotion and machine learning. If we need to use music based on the emotion they convey, we need to detect the emotion in the music has. As there are millions of songs available, it is practically impossible for humans to listen and analyze the emotion. Also, each human has different perception. So, we have built a machine learning model that can recognize the emotion that music conveys.

II. DESIGN AND ARCHITECTURE

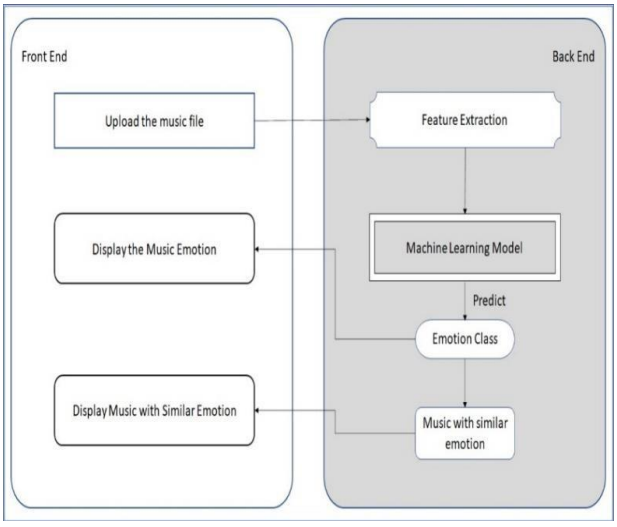

Fig 1. System Architecture.

1. Upload the Music File:
Initially, in the front end, we upload the song. We can specify the type as audio in the input tag in HTML code in order to avoid the other types of files. There is a folder named 'Upload' where the uploaded songs get stored in the server side. The folder's content can be cleared after every use.

International Journal of Science, Engineering and Technology

An Open Access Journal

2. Feature Extraction:
We extract the features of the audio file uploaded by the user after the song is uploaded. The features that are used to build the dataset are the features that we extract from the user uploaded music file.

The features include mfccs, chroma energy normalizations, etc. which are described in the below sections.

3. Machine Learning Model:
We have built a machine learning model using the dataset we've built as described in below sections. We have used Gaussian Naïve Bayes Classifier to predict the Emotion Class based on the Thayer's Emotion Space. Display the music emotion: The emotion class that is predicted by the ML model is then displayed on screen along with the name of the song.

4. Display Music with Similar Emotion:
We have a selected set of music for each emotion class stored in the server side. Based on the emotion predicted by the ML model, we display the songs with similar emotion which can be played from the front end.

III. METHODOLOGY

We created a part of the dataset using the data of the PMEmo dataset 2019 version. The PMEmo dataset is a standard dataset for music classification available in the internet. The PMEmo dataset comes with the continuous Arousal and Valence values for more than 700 songs. We calculated the mean of Arousal and Valence values for each entry in the dataset. It comes along with the music files which are used to build the dataset.

We extracted the required features which are mentioned below, from the music files of PMEmo dataset using a python library called librosa. We then used the Thayer's Emotion Space representation to cluster these music files into 4 emotion classes namely, Happy and Excited, Sad and depressed, calm and Relaxed, Angry and Anxious.

We plotted a graph where each song is represented as a dot with mean- valence as x-coordinates and mean-arousal as y-coordinates. We clustered the songs by considering the quadrants they lie on. This way we built the dataset and we added around 500 music files' features with known emotions to enhance the accuracy.

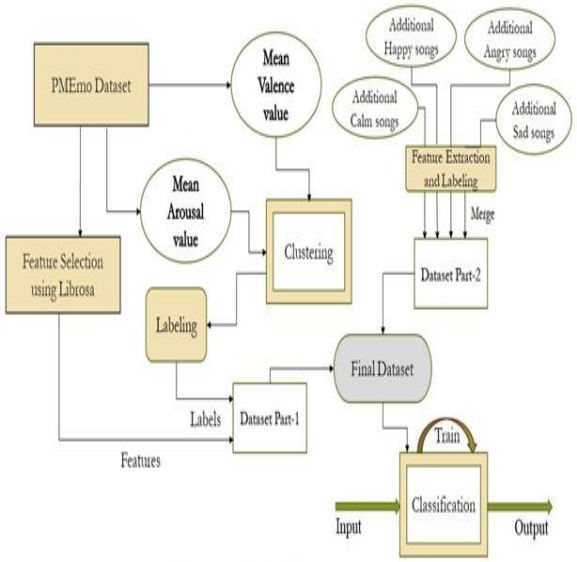

Fig 2. Methodology.

1. Separation of Harmonic and Percussive Signals:

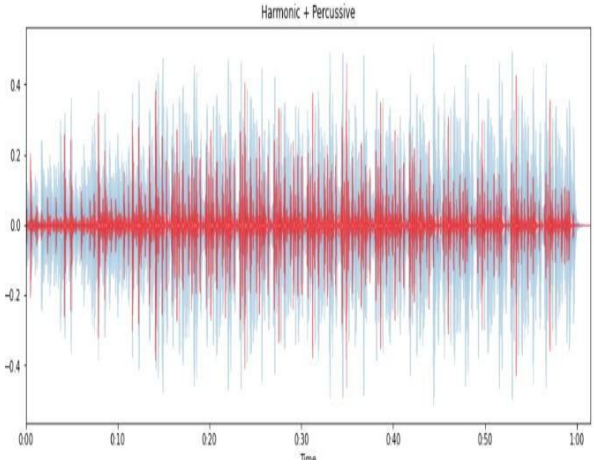

Fig 3. Harmonic and Percussive Signals.

Songs are a combination of Harmonic and Percussive audio signals. Harmonic signals have horizontal structure. Percussive signals have vertical structure. We need only Harmonic signals for extracting the features.

2. Music Information Retrieval (Music features used):
Rhythmic features provide the main beat and its strength of a music track. Several beat-tracking algorithms have been proposed to estimate the main beat and the corresponding strength. Pitch features, mainly derived from the pitch histogram, describe the melody of the music.

International Journal of Science, Engineering and Technology

An Open Access Journal

Timbral features are generally characterized by the properties related to instrumentations or sound sources such as music, speech, or environment signals.

The features used to represent the timbral texture of a music track include zero crossing, spectral centroid, spectral flux, spectral roll off, Mel-frequency cepstral coefficients (MFCC), Daubechies wavelet coefficients histograms (DWCH), and octave-based spectral contrast (OSC), etc. Visualization of these features for sample music is shown as follow.

3. Beat (for Tempo):

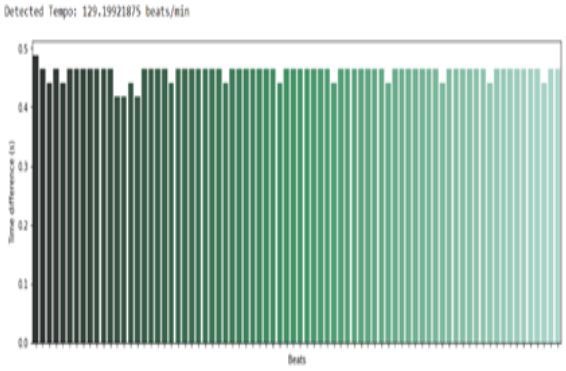

Fig 4. Beat Extraction.

We extract beats for the estimation of tempo in beats per minute and beats per second, and an array of frame numbers corresponding to detected beat events. Frames here correspond to short windows of the signal (y), each separated by hop length = 512 samples.

Since v0.3, librosa uses centered frames; so that the kth frame is centered on sample k. we will use the average tempo of the song as a feature.

4. Chroma Energy Normalized (CENS):

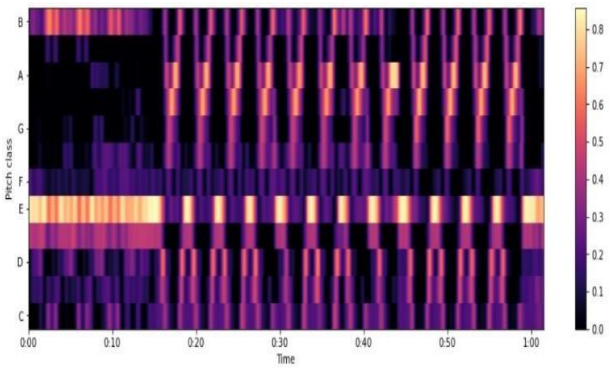

Fig 5. Chroma Energy Normalized.

Calculate the Chroma Energy Normalized (CENS) for the audio file. A 12- element representation of the spectral energy where the bins represent the 12 equal- tempered pitch classes of western-type music (semitone spacing).

5. MFCCs:

Mel-frequency cepstral coefficients (MFCCs) are coefficients that collectively make up an MFC. They are derived from a type of cepstral representation of the audio clip (a nonlinear "spectrum-of-a-spectrum").

The difference between the cepstrum and the Mel-frequency cepstrum is that in the MFC, the frequency bands are equally spaced on the Mel scale, which approximates the human auditory system's response more closely than the linearly spaced frequency bands used in the normal cepstrum. This frequency warping can allow for better representation of sound, for example, in audio compression.

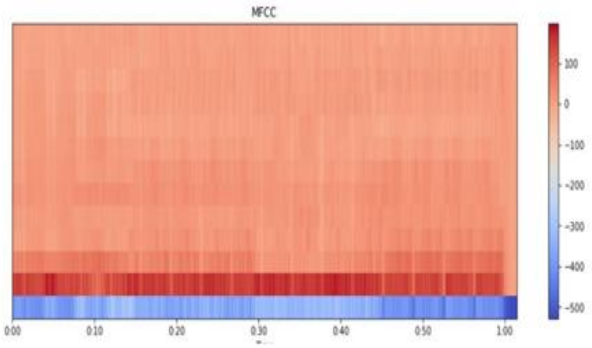

Fig 6. MFCCs.

6. Spectral Centroid:

The spectral centroid is a measure used in digital signal processing to characterize a spectrum. It indicates where the "center of mass" of the spectrum is. Perceptually, it has a robust connection with the impression of "brightness" of a sound.

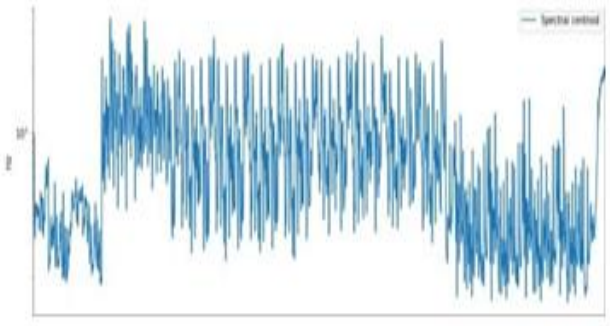

Fig 7. Spectral Centroid.

International Journal of Science, Engineering and Technology

An Open Access Journal

7. Spectral Contrast:

OSC was developed to represent the spectral characteristics of a music piece. It considers the spectral peak and valley in each sub-band separately.

In general, spectral peaks correspond to harmonic components and spectral valleys correspond to non-harmonic components or noise in a music piece. Therefore, the difference between spectral peaks and spectral valleys will reflect the spectral contrast distribution.

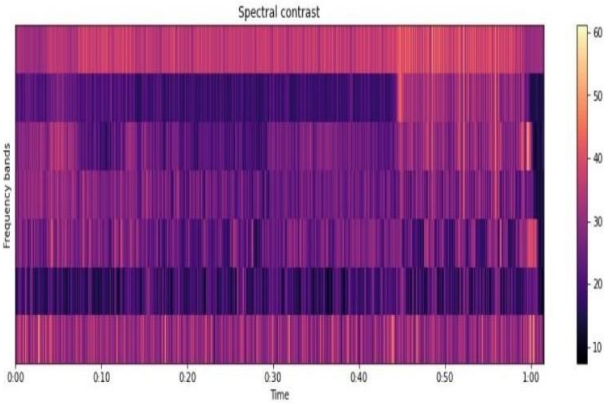

Fig 8. Spectral Contrast.

8. Spectral Roll off:

Spectral roll off point is defined as the Nth percentile of the power spectral distribution, where N is usually 85% or 95%.

The roll off point is the frequency below which the N% of the magnitude distribution is concentrated. This measure is useful in distinguishing voiced speech from unvoiced: unvoiced speech has a high proportion of energy contained in the high-frequency range of the spectrum, where most of the energy for voiced speech and music is contained in lower bands.

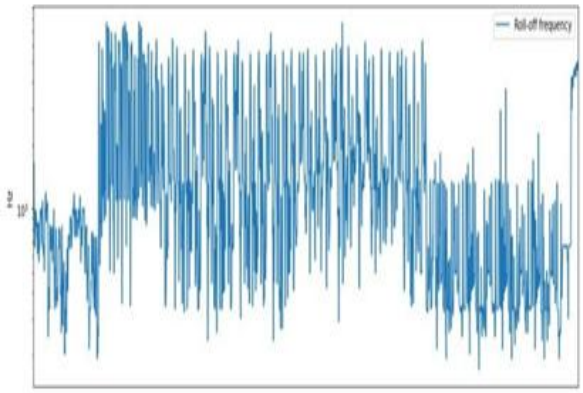

Fig 9. Spectral Roll off.

9. Zero Crossing Rate:

A zero-crossing point is a point in a digital audio file where the sample is at zero amplitude. At any other point, the amplitude of the wave is rising towards its peak or sinking towards its trough.

The zero-crossing rate is the rate of sign-changes along a signal, i.e., the rate at which the signal changes from positive to negative or back. This feature has been used heavily in both speech recognition and music information retrieval, being a key feature to classify percussive sounds.

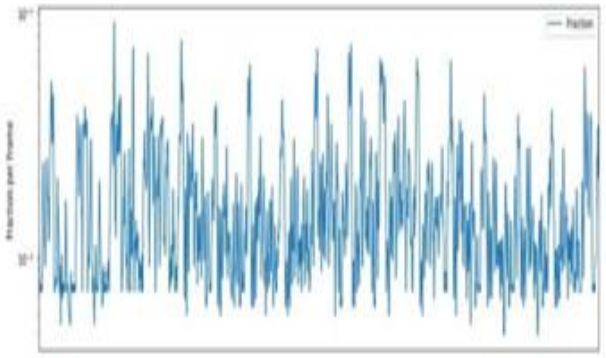

Fig 10. Zero Crossing Rate.

Timbral features are generally characterized by the properties related to instrumentations or sound sources such as music, speech, or environment signals.

The features used to represent the timbral texture of a music track include zero crossing, spectral centroid, spectral flux, spectral roll off, Mel- frequency cepstral coefficients (MFCC), Daubechies wavelet coefficients histograms (DWCH), and octave-based spectral contrast (OSC), etc.

IV. CLASSIFICATION

There are 4 Emotion classes: Happy and Excited, Anxious and Angry, Sad and Depressed, Calm and Relaxed. Each Emotion class is considered as a target in the final dataset. As the music features are continuous in nature, we have used Gaussian Naïve Bayes' Classifier for final prediction of emotion class.

V. CONCLUSION

Recognizing musical emotion remains a challenging problem as there are various human emotions. Different people perceive music in different ways.

International Journal of Science, Engineering and Technology

An Open Access Journal

Emotion perceived by one person can be slightly different from the emotion perceived by another person. We have come up with an idea of classifying the music based on Thayer's emotion model, which consists of 4 emotion classes which cover almost all the emotions.

We have built a machine learning model that can classifies music without using lyrics and our model predicts the emotions with an accuracy of 88.49%. The model can be further enhanced by considering the lyrics of the songs; by doing so, we may have to build an ML model compatible to multiple languages.

REFERENCES

[1] Bischoff, K., Firan, C.S., Paiu, R., Nejdl, W., Laurier, C., Sordo, M.: Music mood and theme classification - a hybrid approach. In: Proc. of the 12th International Society for Music Information Retrieval (ISMIR) Conference, pp. 657–662 (2011).

[2] Kim, J.H., Lee, S., Kim, S.M., Yoo, and W.Y.: Music mood classification model based on Arousal-Valence values. In: Proc. of the 2nd International Conference on Advancements in Computing Technology (ICACT), pp. 292–295 (2011).

[3] Review of data features-based music emotion recognition methods - by Xinyu Yang, Yizhuo Dong, Juan Li, 2017.

[4] The AMG1608 dataset for music emotion recognition - by Yu-An Chen, Yi-Hsuan Yang, Ju-Chiang Wang, 2015.

[5] An Analysis of Low-Arousal Piano Music Ratings to Uncover What Makes Calm and Sad Music So Difficult to Distinguish in Music Emotion Recognition - by Yu Hong, Chuck-Jee Chau, and Andrew horner, 2017.