Report

Genre predictor of Songs

1. Introduction

We are in an age in which we have millions of option for music and limited time to decide whether we like or dislike a song, it would be sad if we miss out on some excellent songs, because we didn't tumble upon them accidently, while creating a collection which we will be using for years to come.

Genre recommendation, can be used in music recommendation, automated radio, something new to explore and many more options.

The purpose of this project is to find relation between songs and its genre, and create a model which can predict whether a song belongs to Folk, Electronic, Pop or Rock genre.

2. Related Work

T.Feng uses 13 features per segment, to classify the music in 4 genres using GTZAN dataset. T. Feng generates more dataset out of initial dataset of GTZAN. And applying Boltzman machine predicts the music genres.

Miguel Flores, in his paper on Deep Music Genre uses GTZAN dataset to predict the genre. The paper mainly focuses on result generated using convolution neural network by varying the epoch. The dataset is similar to T.Feng's of 13 features MFCC dataset per segment.

3. Data Processing

Major step in a machine learning application is collection and modeling of dataset.

3.a Data Pre-Processing:

A song in a mp3 format, must first be converted to a machine-readable format i.e. numbers, so a mp3 is converted to a wav format. The current waveform obtained is a stereo format, the data produced in stereo format is combination of different channels, that's data is produced by various sources. The stereo waveform then is converted to a mono form, produced by one source.

3.b Data Cleaning:

A music in human term can be defined as loudness, pitch and timbre. Loudness is intensity of the sound and is measured dB, pitch is rate of vibration produced by the sound and

timbre is used to differentiate between the sound produced by different musical instrument, when the pitch and loudness are the same.

Loudness has the same mapping for computers. Pitch of a sound is converted to 12 chroma features and Timbre is converted to 12 MFCC features. Loudness, Pitch and Timbre are normalized. These features combine to a total of 25 features, but these features are numerical representation of the music at particular time frame or in segment. A music is collection of segments, in this step a collection of 500 such segments stored. Thus, a total of 12500 columns for each music.

3.c Data Storage:

FMA: A Dataset For Music Analysis (Cornell University Library), is collection 30 sec music sound clip. For the purpose of this application, I am utilizing the fma_large dataset, which is collection of 8000 tracks of 30s. This set is randomly divided into 60% training and 40% test data. A multithreaded application converts these sound clips in mp3 format and store a chunk of 1000 songs in each csv file, unbalanced genres mapped to the songs are stored in another file in JSON format.

4. Modeling the Data

Cleaned data is then combined into dataframe, for further processing. Genre detection is an example of multiclass classifier, following models are created for the data obtained:
a) K-nearest neighbors
b) Decision Tree
c) Random Forest

5. Verifying the Results

Here,
Genre_1 = Folk
Genre_2 = Electronic
Genre_3 = Pop
Genre_4 = Rock

	N_{Genre_1}	N_{Genre_2}	N_{Genre_3}	N_{Genre_4}	
N_{Genre_1}	N_{P1}	N_{FP2}	N_{FP3}	N_{FP4}	N_{R1}
N_{Genre_2}	N_{FP1}	N_{P2}	N_{FP3}	N_{FP4}	N_{R2}
N_{Genre_3}	N_{FP1}	N_{FP2}	N_{P3}	N_{FP4}	N_{R3}
N_{Genre_4}	N_{FP1}	N_{FP2}	N_{FP3}	N_{P3}	N_{R4}
	N_{s1}	N_{s2}	N_{s3}	N_{s4}	N_T

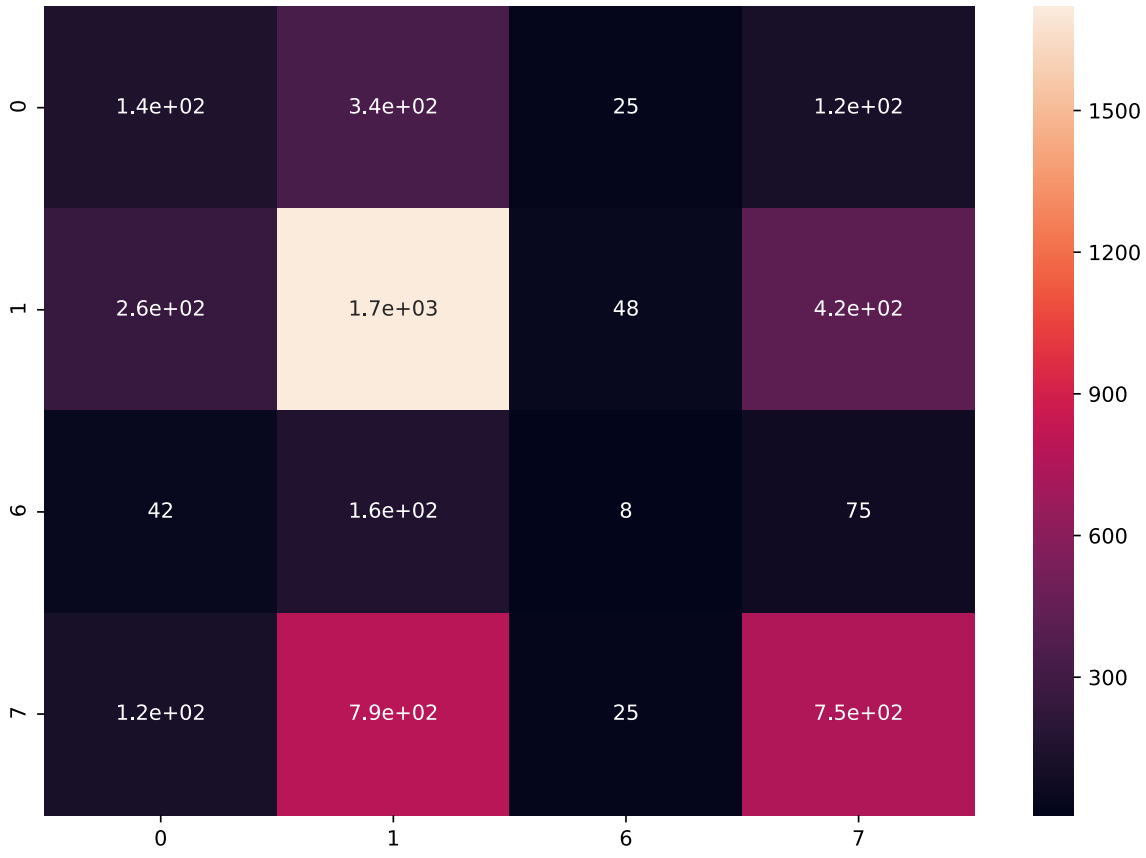
Report

Step 1) Calculate the accuracy of the algorithm:
Accuracy = ($N_{P1} + N_{P2} + N_{P3} + N_{P4}$) / N_T
Step 2) Calculate Precision and Recall, for each Genre
Precision = N_{P1} / (N_{S1})
Recall = N_{P1}/ N_{R1} for Genre 1,
similarly we can compute precision and recall for remaining genres.

K-nearest neighbors:

In K-nearest neighbors, we create models with variety of neighbors to find the model most accurate prediction.

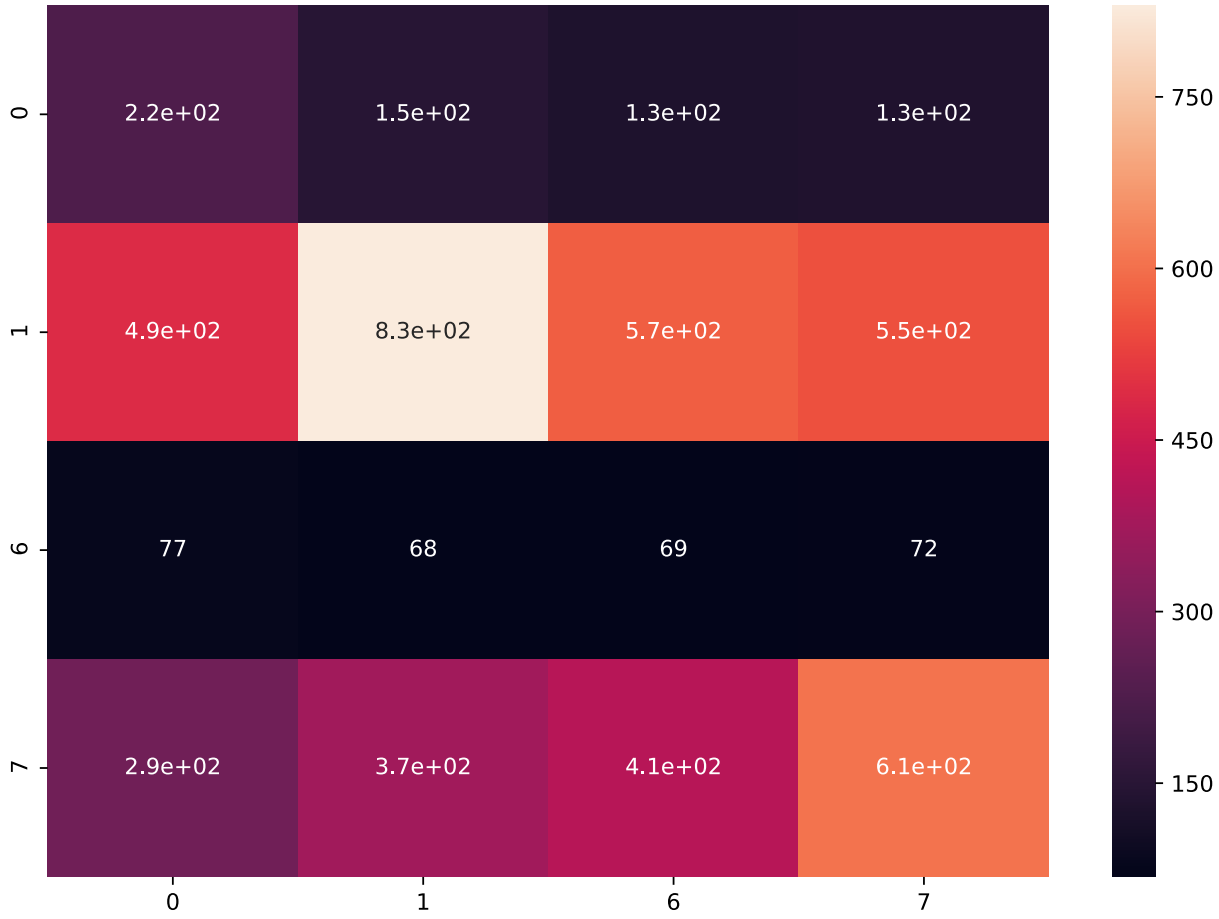
The best accuracy is provided by number of neighbors = 8, with accuracy = 52%



Report

	Folk	Electronic	Pop	Rock
Precision	25%	57%	7%	55%
Recall	22%	70%	3%	45%

Decision Tree:



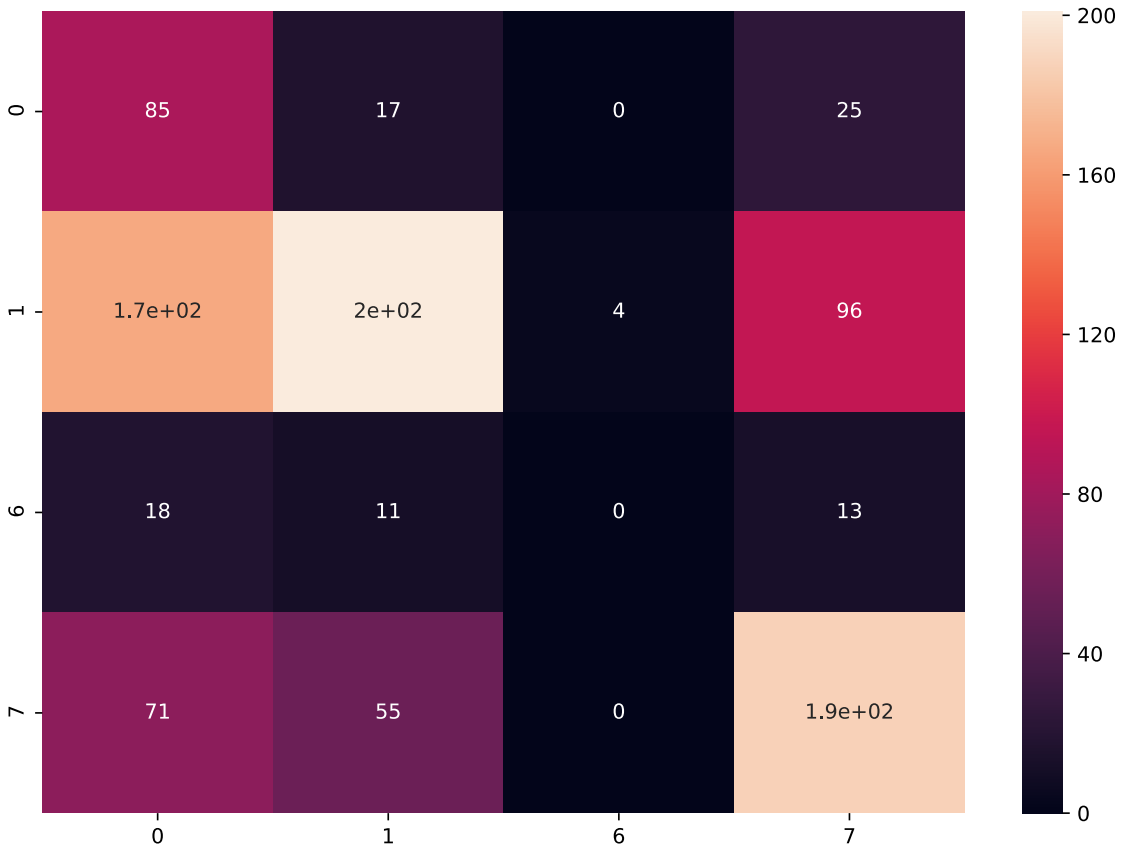
The accuracy generated by decision Tree is 34%
Precision and Recall from above heat map are:

	Folk	Electronic	Pop	Rock
Precision	20%	59%	6%	45%
Recall	35%	34%	5%	36%

Report

Random Forest:

The accuracy generated by Random Forest is: 50%



Precision and Recall from the above heat map is:

	Folk	Electronic	Pop	Rock
Precision	25%	70%	0%	59%
Recall	67%	42%	0%	60%

6. Conclusion:

Random Forest predict the genre most accurately, and with a higher precision and recall than the remaining two algorithms. The value for Random Forest can be further improved by increasing its maximum parameter value. The reason for low accuracy, precision and

recall is amount of data and similarity between the genres.

7. **Future:**

Future of genre prediction is to improve the accuracy even further, by collecting more data. Another curious case is the use case of multi-label and multi-class genre prediction.