**SONG CLASSIFICATION AND RECOMMENDATION**

**Aparna Pavithran Naveenraj Palanisamy**

**May 2016**

**1. Introduction**

**1.1 Background**

Music is an art, which is enjoyed by almost all sorts of people. It has the power to change the mood of us, release stress, cure diseases etc. But different people listen to music in different ways based on their choice of selection. If you have a collection of songs, which is played by a music player in a shuffled order, will you like a rock song followed by a melody? No one will like to listen to music from one end to another extreme type of music.

We are proposing a solution for the above scenario, by recommending a song from the given song lists based on the type or genre of the current playing one. We will be doing unsupervised and supervised learning to associate songs to different clusters by EM algorithm, K means algorithm, Naïve Bayes algorithm and SVM algorithm.

However, recommending next music has to depend on so many things like music genre, language, lyrics, singer, particular band, musical instrument and a lot more. Considering all these factors are out of scope for our project. Hence we considered only genre of the song. However, the project can be extended in lot more ways.

**2. Problem Definition**

Substantial number of songs in multiple languages gets released to online music stores and streaming servers or uploaded to archives around the world each day, and all these songs need to be classified. Users have a preference to listen to a certain genre of music, which is driven by their mood. Predicting users and coming up with a tailored playlists, which suits their mood, is a challenge. Current generation of players does not have the capability to classify and play based on related music patterns.

**2.1 Task Definition**

For this proposed work, we used EM algorithm to classify songs to different clusters based on the frequency of the song.

We will give labeled song for each genre and based on that we will classify cluster to genre. If user is playing one song from particular cluster, then based on its genre, will suggest list of songs, which is related to that. Hence user can pick from the list and play the next. For example: If a user is listening to hip hop song then song with more hip hop probabilities among the lists will be suggested to user to play next.

Music is enjoyed by all kind of people, but the choice varies. On an average, American user listens 4 hours of music per day.  Pandora users listen to more than 60,000 hours of music. Preferred music genre among users varies. Suggesting songs based on users interests will give interest to the users and hence streaming more songs.

**2.2 Data Collection**

Songs are collected from 6-7 genres, around 650 songs, around 100 songs of each genre. Songs are collected from various sources.

**2.3 Feature Set**

Feature set is generated from songs by considering different factors like Zero crossing rate- time based domain feature, Centroid/ mean/ slope of spectrum, short-term energy, short term entropy, spectral entropy, spectral flux, spectral roll off, MFCC etc. These are different types of feature extraction techniques used. In total 72 features were generated and reduced to 44 by removing features, which are not contributing much for the classification.

**2.4 Algorithm Definition**

We have used four different algorithms here. They are 2 supervised algorithms – Naïve Bayes algorithm and Support Vector Machine (SVM) algorithm – and 2 unsupervised algorithm – EM algorithm and K Means algorithm. Our main focus is on EM algorithm. We can get probability of next playable list from either EM algorithm or Naïve Bayes algorithm.

EM algorithm: EM algorithm is commonly used method computational pattern recognition community which enables determine maximum-likelihood in given data set when the data is incomplete or has missing values of the parameters. When data set supplied contains missing values, due to problems with or limitations of the observation process. The likelihood function has to be simplified by making certain assumptions like existence of and values for additional missing (or hidden) parameters. This is specifically applicable when optimization of likelihood function is analytically intractable.

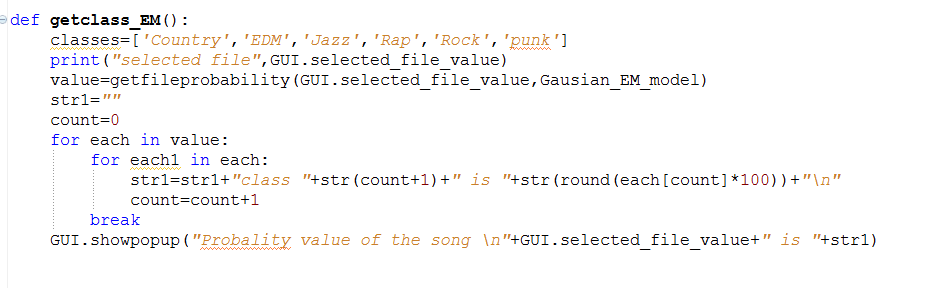
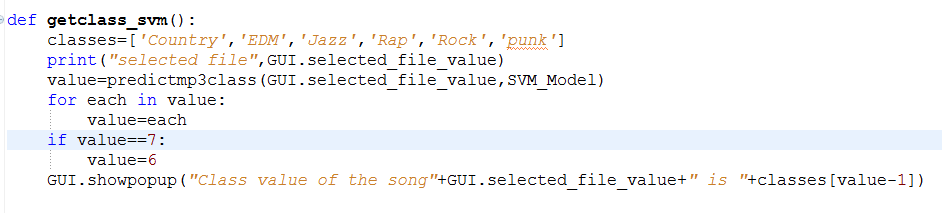


Fig 1. EM algorithm

  
Fig 2. SVM algorithm

**3. Experimental Evaluation**   
  
**3.1 Methodology**

Songs were in .mp3 format, which was converted to .wav format. From the .wav format of song each frame is read then from each frame, frequency is extracted and then from the frequency following features are extracted - Zero crossing rate- time based domain feature, Centroid/ mean/ slope of spectrum, short-term energy, short term entropy, spectral entropy, spectral flux, spectral roll off, MFCC.

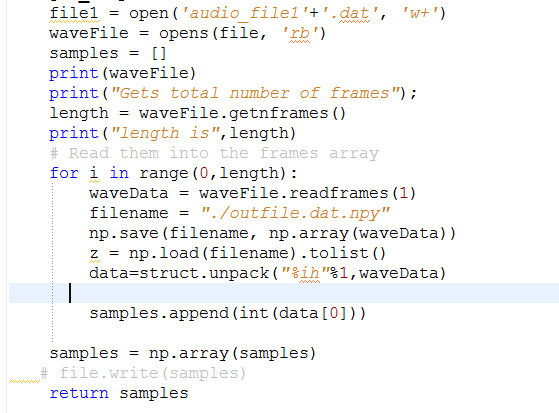


Fig 3. Reading Wav file and each Frame

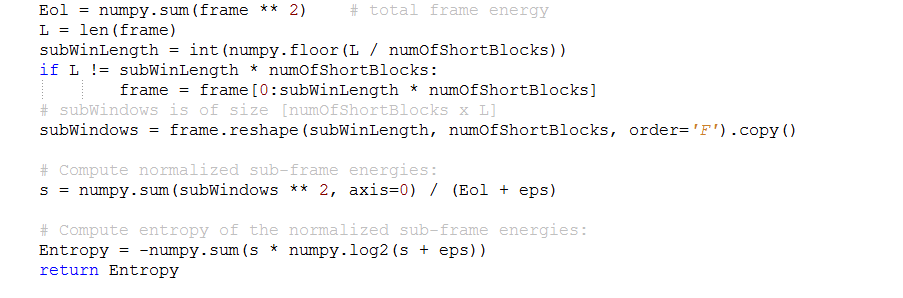


Fig 4. Spectral Entropy Extraction

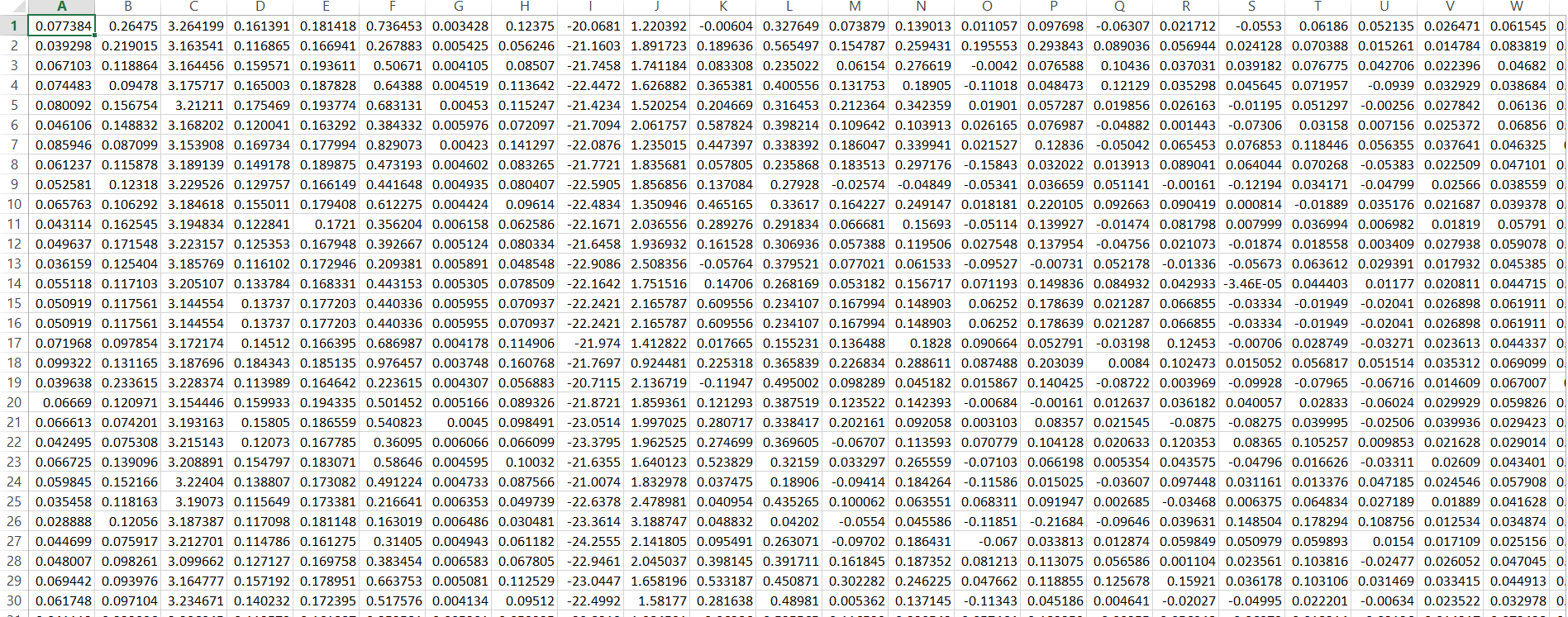


Fig 5. Extracted Features

A data set is generated from this and it is used as the input to all machine learning algorithms to produce desired result. The same dataset is used as train and test set by cross validation. Probability of next song has chosen by doing EM and Naïve Bayes algorithms. Based on the probability next high probable songs are listed out for playing next.

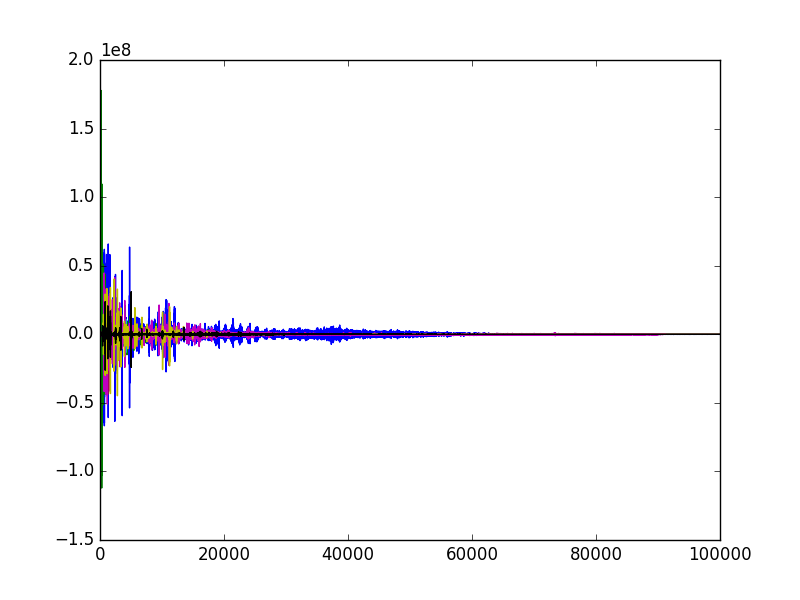
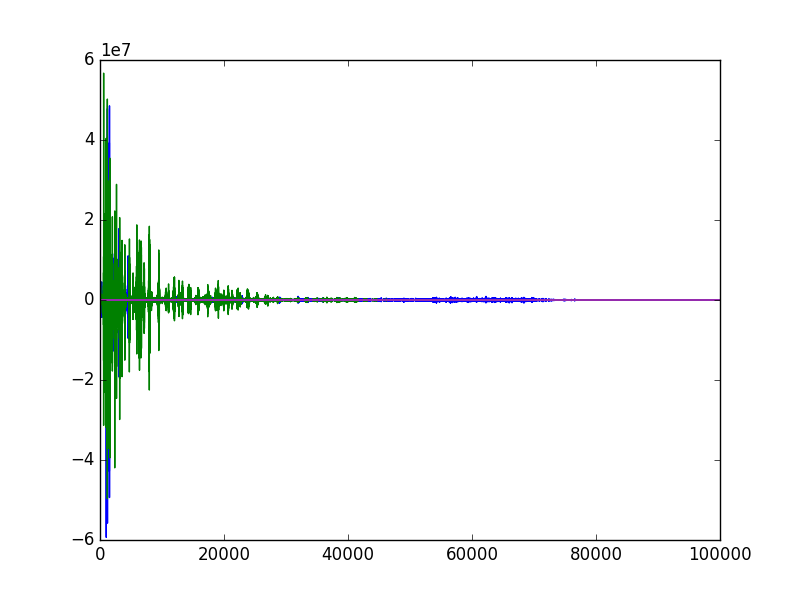
 

Fig 6. Frequency frame- COUNTRY Fig 7. Frequency frame- JAZZ

**3.2 Evaluating method**

We have used cross validation for evaluating our work. Cross-validation, sometimes called rotation estimation is a model validation technique for assessing how the results of a statistical analysis will generalize to an independent data set. It is mainly used in settings where the goal is prediction, and one wants to estimate how accurately a predictive model will perform in practice. The goal of cross validation is to define a dataset to "test" the model in the training phase (i.e., the validation dataset), in order to limit problems like over fitting, give an insight on how the model will generalize to an independent dataset. We have used k-fold cross validation-a type of cross validation method- method here with different k values.

**3.3 Basic Data flow**

**Basic Classification Tasks**

**Music**

**Musical Data Collection**

**Feature Extraction**

**Machine Learning**

**Classifications**

Fig 8. Basic Song Classification Data flow

**3.4 Architecture**

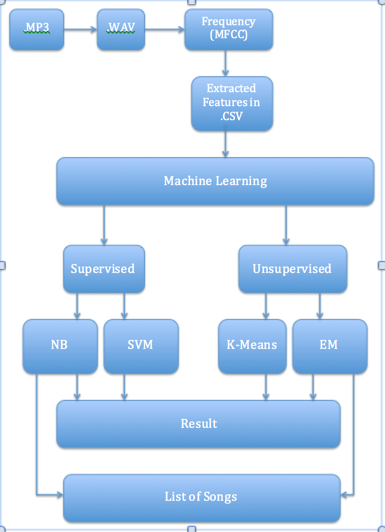


Fig 9. System Architecture

**3.5 Results**

Algorithms used to cluster songs were EM and K means algorithms. Whereas supervised algorithms used were Naïve Bayes and SVM. SVM showed 56 percentages accuracy whereas Naïve Bayes showed 76 percentages accuracy. Accuracy is less because now days it is difficult to classify songs to only one genre as one song might contain more genres. In such cases EM will be the best algorithm than K means as it does soft clustering. And that is exactly what we are looking for. We need more probable songs to play next from the given list, which are more matching with the currently playing one. SVM showed better accuracy of 66 percentages when used different kernels like RBF, sigmoid and polynomial.

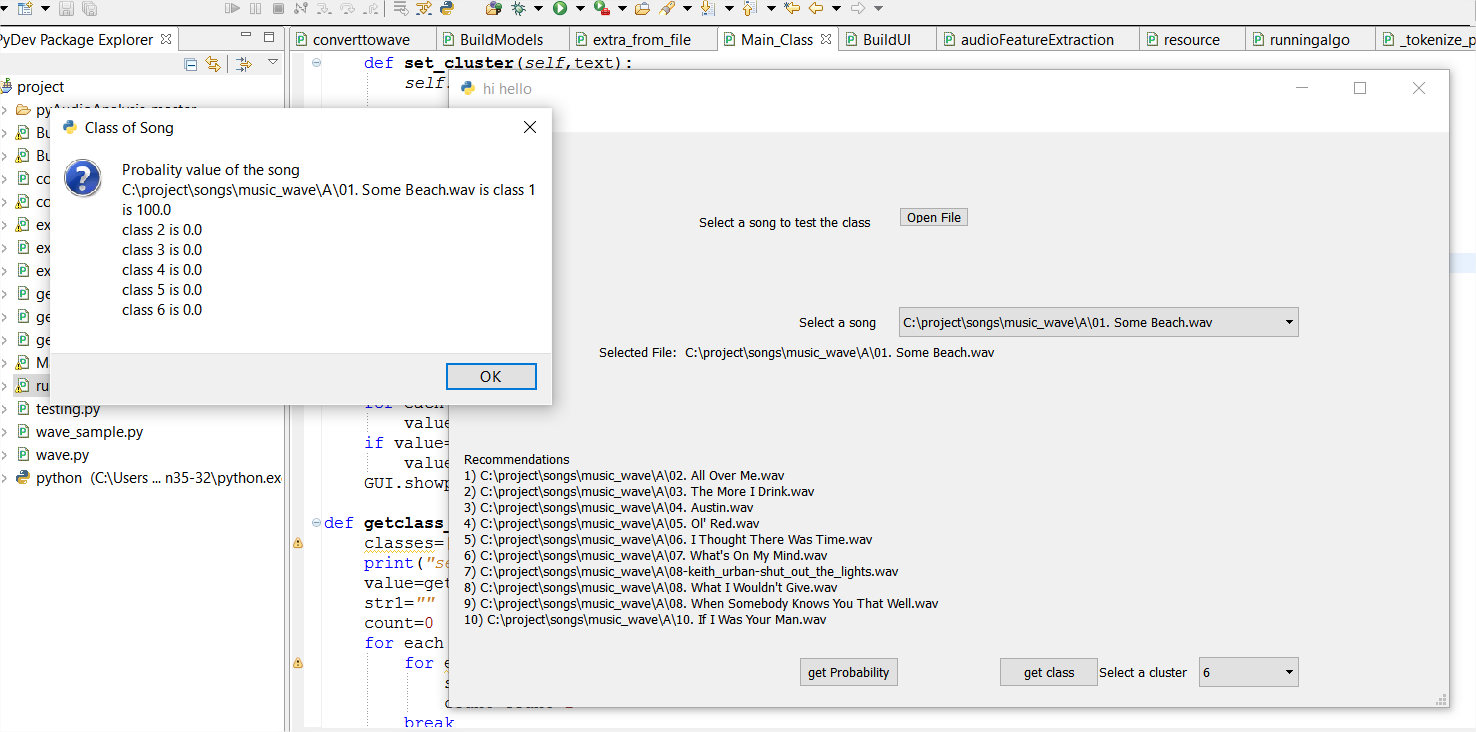


Fig 10. Suggested lists

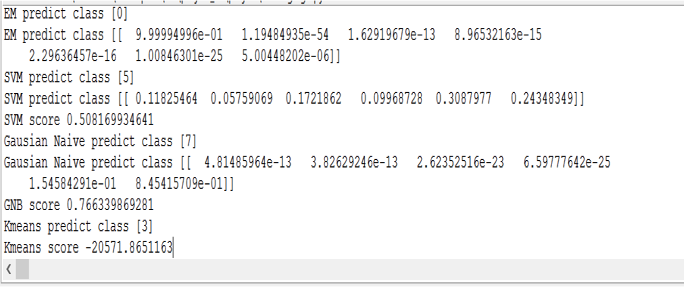


Fig 11. Accuracy

**3.6 Discussion**

Strength is using EM to recommendations and also we have compared the results of supervised learning with SVM and Naïve Bayes. In unsupervised both hard and soft classification is done.

**4. Related Work**

There were a lot of works done in this field for classifying songs with supervised algorithms. But there was nothing much with unsupervised algorithms, especially Expectation Maximization (EM) algorithm. Also we are proposing a new logic of predicting next probable songs which user will like.  
  
**5. Future Work**

In this project we have considered only genre-based classification. Considering more features for classification like language, lyrics, singer, particular band, musical instrument and a lot more can be used to classify to extend this project.   
  
**6. Conclusion**   
In unsupervised algorithms, EM algorithm worked better than K means due to soft clustering. EM was the perfect algorithm for our project because rather than doing hard clustering, EM will even give the probability of being a song in another cluster even when it is clustered into one category. In supervised algorithm, Naïve Bayes algorithm showed better accuracy than SVM algorithm.

**7. References**

1. ‘Clustering Music by Genres Using Supervised and Unsupervised Algorithms’ Kyuwon Kim, Wonjin Yun, and Rick Kim
2. ‘Musical Genre Classification of Audio Signals’,George Tzanetakis and Perry Cook,IEEE Transactions on Speech and Audio Processing, 10(5), July 2002
3. ‘Analysis and Clustering of Musical Compositions using Melody-based Features’ Isaac Caswell Erika Ji December 13, 2013
4. Thierry Bertin-Mahieux, Daniel P.W. Ellis, Brian Whitman, and Paul Lamere (2011). ‘The Million Song Dataset’. In Proceedings of the 12th International Society for Music Information Retrieval Conference (IS-MIR 2011), 2011.
5. [11] S. Li, “Content-based classification and retrieval of audio using the nearest feature line method,” IEEE Trans. Speech Audio Processing, vol. 8, pp. 619–625, Sept. 2000.
6. ‘Smart Music Rating Predictor’ Scott Chao, Huang Song, Liqun Yang.
7. ‘Classifying Parts Of Songs’ Nick Colonnese December 10,2010.
8. ‘Can Song Lyrics Predict Genre?’ Danny Diekroeger Stanford University [danny1@stanford.edu](mailto:danny1@stanford.edu)
9. ‘Song genre classification via training on segment chromas and MFCCs’ Caleb Jordan Stanford University [grnstnrd@stanford.edu](mailto:grnstnrd@stanford.edu), Alex Cope Stanford University [alexcope@stanford.edu](mailto:alexcope@stanford.edu).
10. ‘Predicting Hit Songs with MIDI Musical Features’ Keven (Kedao) Wang Stanford University [kvw@stanford.edu](mailto:kvw@stanford.edu)
11. ‘Algorithmic Clustering of Music Based on String Compression’, Rudi Cilibrasi, Paul Vita and Ronald de Wolf.
12. Thierry Bertin-Mahieux, Daniel P.W. Ellis, Brian Whitman, and Paul Lamere (2011). The Million Song Dataset. In Proceedings of the 12th International Society for Music Information Retrieval Conference (IS-MIR 2011), 2011.