iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 5 **Issue:** VI **Month of publication:** June 2017

DOI:

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)

Automatic mood classification of Indian Popular music

Vikas Verma[1]
[1]Assistant Professor School of Engg. &Technolog Jaipur National University, Jaipur Raj. India

Abstract ***Music has been an inherent part of human life when it comes to recreation; entertainment and much recently, even as a therapeutic medium. The way music is composed, played and listened to has witnessed an enormous transition from the age of magnetic tape recorders to the recent age of digital music players streaming music from the cloud. What has remained intact is the special relation that music shares with human emotions. We most often choose to listen to a song or music which best fits our mood at that instant. In spite of this strong correlation, most of the music software's present today is still devoid of providing the facility of mood-aware play-list generation. This increase the time music listeners take in manually choosing a list of songs suiting a particular mood or occasion, which can be avoided by annotating songs with the relevant emotion category they convey. The problem, however, lies in the overhead of manual annotation of music with its corresponding mood and the challenge is to identify this aspect automatically and intelligently. Our focus is specifically on Indian Popular Hindi songs. We have analyzed various data classification algorithms in order to learn, train and test the model representing the moods of these audio songs and developed an open source framework for the same. We have been successful to achieve a satisfactory precision of 70% to 75% in identifying the mood underlying the Indian popular music by introducing the bagging (ensemble) of random forest approach experimented over a list of 4600 audio clips.***

I. INTRODUCTION

THE well-known German philosopher Friedrich Nietzsche" once quoted a famous line: Without music, life would be a mistake". Music has always been an inherent part of recreation of human life. Music is not just useful for entertainment, but studies have shown that listening the right music does play an important role in healing; rejuvenating and even inspiring human mind in difficult conditions such as is widely studied and demonstrated by the field of Music Therapy [3]. With the rapidly increasing technology and the new advent in latest multimedia gadgets, music has reached almost every individual's personal gadget may it be a laptop, music player or a cell phone. The music which in the olden days was limited to live concerts, performances or radio broadcasts is now available at everyone's finger tips within few clicks. Music has thus become very easily accessible and available. However, the music database is ever increasing and the list will go so long that it would not be wrong to say that we might hear a couple or more completely new and never-heard music pieces every single day. Today, the overall music collection can count to a few millions of records in the whole world and still continue to increase every day. With so much of variety of music easily available, we humans do not always listen to the same type of music all the time. We have our interests, favorite artists, albums, music type. To put simply, we have our personal choices and more importantly, even our choice might differ from time to time. This choice is very much naturally governed by our emotional state at that particular instant. The relation between musical sounds and their innocence on the listeners' emotion has been well studied and is evident from the much celebrated papers such as that of Hevner [1] and Farnsworth [2]. The papers described the conference paper formatting requirements is to use this document as a template and simply type your text into it. Experiments which eventually substantiated a hypothesis that music inherently

II. INTRODUCTION TO MUSIC FEATURE

As it is a well-established fact that music indeed has an emotional quotient attached with it, it is very essential to know what are the intrinsic factors present in music which associate it with a particular mood or emotion. A lot of research has been done and still going on in capturing various features from the audio file based on which we can analyze and classify a list of audio files. Audio features are nothing but mathematical functions calculated over the audio data, in order to describe some unique aspect of that data. In the last decades a huge number of features were developed for the analysis of audio content. Dalibor Mitrovic and team [4] have analyzed various state-of-the-art audio features useful for content- based audio retrieval. Audio features were initially studied and explored for application domains like speech recognition [5]. With upcoming novel application areas, the analysis of music and general purpose environmental sounds gained importance. Different research fields evolved, such as audio segmentation, music

©IJRASET: All Rights are Reserved

information retrieval (MIR), and environmental sound recognition (ESR). Each of these areas developed its specific description techniques (features). Many audio features have been proposed in the literature for music classification. Different taxonomies exist for the categorization of audio features. Weihs et al. [6] have categorized the audio features into four subcategories, namely short-term features, long-term features, semantic features, and compositional features. Scaringella [7] followed a more standard taxonomy by dividing audio features used for genre classification into three groups based on timbre, rhythm, and pitch information, respectively. Each taxonomy attempts to capture audio features from certain perspective. Instead of single-level taxonomy, Zhouyu Fu and team [8] unify the two taxonomies and present a hierarchical taxonomy that characterizes audio features from different perspectives and levels. From the perspective of music understanding, we can divide audio features into two levels, low-level and mid-level features along with top-level labels. Low-level features can be further divided into two classes of timbre and temporal features. Timbre features capture the tonal quality of sound that is related to different instrumentation, whereas temporal features capture the variation and evolution of timbre over time. Low-level features are the basis description of the audio data, for instance, tempo, beats per minute and soon. On the contrary, mid-level features are derived by using these basic features to provide the music related technical understanding such as rhythm, pitch which in turn is perceived by the humans as genre, mood, which form the top-level of the taxonomy.

III.MUSIC MOOD MODEL AND AUDIO FEATURES

Various experts in the fields of psychology, musicology has come with models describing human emotions. One of the most ancient of the experiments done by Hevner [1] helped in categorizing various adjectives into different groups each representing a class of mood. The model was more of a categorical model wherein a list of adjectives representing same emotion was grouped together. Russell [9] later came up with the circumflex model representing human emotions on a circle with each mood category plotted within the circle separated from other categories along the polar co-ordinates. Thayer [10] too came up with a dimensional model plotted along two axes (Stress versus energy) with mood represented by a two-dimensional co-ordinate system and lying on either of the two axes or the four quadrants formed by the two-dimensional plot. JungHyun Kim and team [11] proposed an Arousal-Valence (A-V) based mood classification model for music recommendation system. The collected music mood tags and A-V values from 20 subjects were analyzed and the A-V plane was classified into eight regions depicting mood by using k-means clustering algorithm. Their work shows that some regions on the A-V plane can be identified by representative mood tags like previous mood models, but some mood tags are overlapped in almost all regions. Akase and group [12] discuss an approach for the feature extraction for audio mood classification. In this task the timbral information has been widely used, however many musical moods are characterized not only by timbral information but also by musical scale and temporal features such as rhythm patterns and bass-line patterns. Their paper proposed the extraction of rhythm and bass-line patterns, and these unit pattern analysis are combined with statistical feature extraction for mood classification. In combination with statistical features including MFCCs and musical scale feature, the affectivity of the features was verified experimentally.

McKay and team [13] developed "jAudio" an open-source audio feature extraction framework that includes implementations of 26 core features, including both features proven in MIR research and more experimental perceptually motivated features. jAudio places an even greater emphasis on implementations of meta-features and aggregators that can be used to automatically generate many more features from core features (for instance , standard deviation, derivative, running mean etc.) that can be useful for music analysis. The tool has been quite useful and widely accepted for music analysis research. Dalibor Mitrovic and team's work [14] deals with the statistical data analysis of a broad set of state-of-the-art audio features and low-level MPEG-7 audio descriptors. The investigation comprises of data analysis to reveal redundancies between state-of-the-art audio features and MPEG-7 audio descriptors. The work employs Principal Components Analysis, which reveals low redundancy between most of the MPEG-7 descriptor groups. However, there is high redundancy within some groups of descriptors such as the Basic Spectral group and the Timbral Spectral group. Redundant features capture similar properties of the media objects and should not be used in conjunction. The paper provides a good insight on the choice of audio features for analysis.

IV.MOOD(EMOTION) MODELS

Mood Models are generally studied by two approaches

A. Categorical approach

This introduces distinct classes of moods which form the basis for all other possible emotional variations.

©IJRASET: All Rights are Reserved

Volume 5 Issue VI, June 2017
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

B. Dimensional approach

This classifies emotions along several axes such as valence (pleasure), arousal (activity), potency (dominance) and so on. This is generally the most commonly used approach in music applications Human psychologists have done a great deal of work and proposed a number of models on human emotions. Musicologists have too adopted and extended a few of the inuential models that we will be navigating through.

The six universal emotions defined by Ekman [15] anger, disgust, fear, happiness, sadness, and surprise, are well known in psychology. However, since they were designed for encoding facial expressions, some of them may not be suitable for music (for instance, disgust), and some common music moods are missing (for instance, calm or soothing).

V. AUDIO FEATURES

The key components of a classification system are feature extraction and classifier learning [16]. Feature extraction addresses the problem of how to represent the music pieces to be classified in terms of feature vectors or pair-wise similarities. Many audio features have been proposed in the literature for music classification. Different taxonomies exist for the categorization of audio features. Weihs etal. [6] has categorized the audio features into four subcategories, namely short- term features, long term features, semantic features, and compositional features. Scaringella [7] followed a more standard taxonomy by dividing audio features used for genre classification into three groups based on timbre, rhythm, and pitch information, respectively. Each taxonomy attempts to capture audio features from certain perspective. Zhouyu Fu [8] characterizes the audio features into two levels: low-level and middle-level features as seen in Figure 4.1. Our audio feature selection is inspired by this two-tier taxonomy of audio features.

Low-level features although not closely related to the intrinsic properties of music as perceived by human listeners, form the basic features which can be used to derive the mid-level features providing a closer relationship and include mainly three classes of features, namely rhythm, pitch, and harmony as seen in the Figure1 . In our work we focus only on the low-level audio features which can be further categorized as :-

A. Timbral features

These capture the tonal quality of sound that is related to different instrumentation. Timbre" is the quality of a musical note or sound or tone that distinguishes different types of sound production, such as voices and musical instruments, string instruments, wind instruments and percussion instruments. The physical characteristics of sound that determine the perception of timbre includes spectrum and envelope.

B. Temporal features

These capture the variation and evolution of timbre over time. In this work, more focus is laid on the instantaneous timbre values rather than their temporal variation, although not completely ignored.

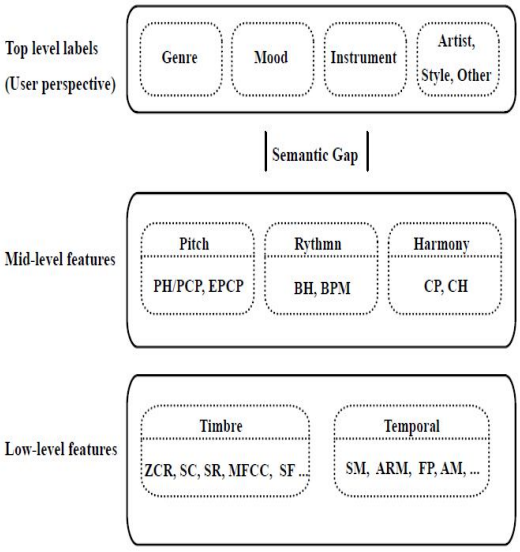


Figure 1: Audio Features Taxonomy

©IJRASET: All Rights are Reserved

Volume 5 Issue VI, June 2017
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

These low-level features are extracted using various signal processing techniques like Fourier transform, Spectral Cepstral analysis, autoregressive modeling and similar computations. We follow the MPEG-7 standardization [28] and make use of the jAudio [13] and Marsyas [17] open source tools for extracting selective timbral spectral and temporal audio features from the music pieces. The features are extracted and consolidated for each music piece in a standard Attribute relation file format (ARFF) [18] so as to make it easy for mining the relations between these features with respect to the corresponding mood of the audio files.

After a careful study and survey of various experts papers and publications, our current consolidated list of selected and extracted features is as mentioned below. The list below names just each form of audio feature, but the feature vector is comprised of its actual value as well as corresponding meta-features such as standard deviation, mean, logarithm wherever required as identified by McKay and team [13].

VI. MINING MOOD FROM AUDIO FEATURES

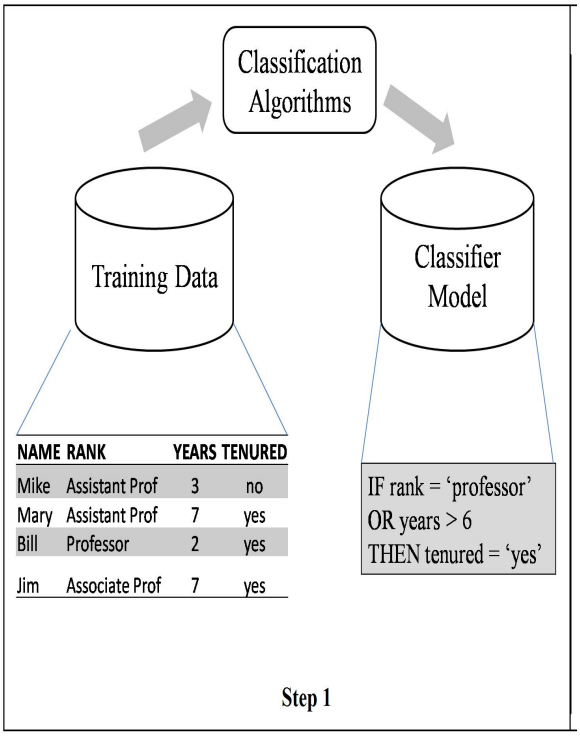
A. Classification

This work involves learning the mood aspect of music by analyzing the various feature vectors extracted from each audio file. The learning done thus can facilitate in identifying which specific category of mood a particular audio file belongs to, provided its fixed set of audio features are available. Of all the functionalities of data mining just described, classification" and cluster analysis" seem to be the best methods of discovering the mood information from the music feature data set. Also, as witnessed in most of the literature survey, classification algorithms have always proved to be quite effective as compared to others in analyzing the mood or genre aspect of music data-sets so far. Our own experimentation too has revealed a quite higher performance of classification algorithms as compared to clustering algorithms. Hence, we opt for the classification techniques of mining this music feature data-set with a supervised learning approach.

The Figure 2 shows a general process of classification. It is a two-step process:-

B. First

This is also called as a \learning step" or training phase" which involves learning of a mapping or a function y = f(X), that can predict the associated class label y of a given tuple X. In this view, we wish to learn a mapping or function that separates the data classes. Typically, this mapping is represented in the form of classification rules, decision trees, or mathematical formulae. This mapping or function is generally termed as a Classification Model". As seen in the step 1 of Figure 2, each row of



©IJRASET: All Rights are Reserved

Volume 5 Issue VI, June 2017
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

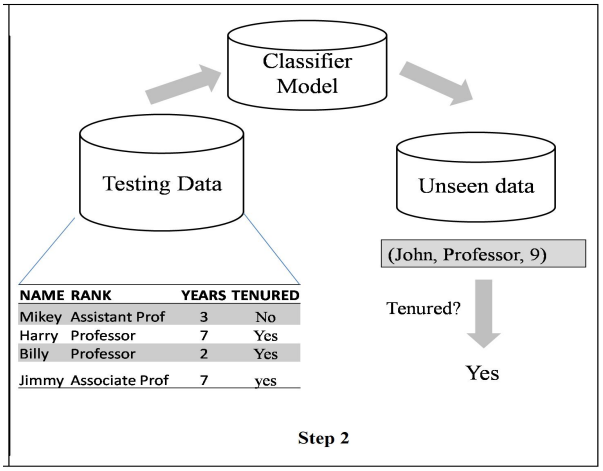


Figure 2: Classification process

the table represents the tuple X. The function f(X) is learnt as a process of training by using classification algorithms, and corresponding rule is stored in the classifier model. This rule helps in predicting if the person represented in the tuple X is tenured (yes) or not (no) depending upon the values of various attributes of the tuple.

C. Second

This model is used for classification. The model is evaluated against the test data-set in order to predict the class label of each data instance as has been learned from the model. The results are compared with actual classes of the test data and accordingly decided whether the model is accurate enough to classify the test data. If the model is acceptable, it can be used further for classifying data with unknown classes. As seen in the step 2 of Figure 2, the classifier model evaluates over an unknown tuple X by applying the function f(X) learnt in order to predict its outcome.

VII. OUR APPROACH: BAGGING OF RANDOM FORESTS

In this work we present an additional hierarchy of ensemble by generating an ensemble of Random Forests using bootstrap aggregation also known as bagging.

The Algorithm for the same can be explained as below.

A. Algorithm: Bagging of Random forest

Data: Training set, N forests = Number of forests

Result: Majority vote of classification

1) Initialization;
2) For i 1 to N forests d
3) Select a new bootstrap sample from training set
4) Generate a Random forest on this bootstrap with Un-pruned random tress as mentioned in Algorithm1
5) Save the Random Forest built
6) Record the majority vote of classification among The trees
7) En
8) Return the majority vote of classification among the Random Forests;

For growing Random Trees, the randomly sampled data attributes are split on the basis of “Gini Index" which has shown better results when working with CART trees. Gini Index basically measures the impurity of the data set D at hand and is given by the formula:-

$$Gini(D) = 1 - \sum_{i=1}^{m} p_i^2$$

©IJRASET: All Rights are Reserved

Volume 5 Issue VI, June 2017
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

Where pi is the probability that a tuple in D belongs to class Ci. The sum is computed over m classes. The Gini index considers a binary split for each attribute. For each split, a weighted sum of the impurity of each resulting partition is calculated. Suppose dataset D is split into partition D1 and D2, on the basis of attribute A1 then the Gini Index of attribute A1 for splitting dataset D is given by :-

$$Gini_{A_1}(D) = \frac{|D_1|}{|D|}Gini(D_1) + \frac{|D_2|}{|D|}Gini(D_2)$$

The Gini Index is computed for all the eligible splitting attributes and the reduction in impurity of Gini index for each attribute is calculated by the formula:-

$$\Delta Gini_{A_1} = Gini(D) - Gini_{A_1}(D)$$

The attribute maximizing the above mentioned reduction in impurity is selected as the splitting attribute. The approach mentioned in Algorithm 2 has not only shown a rise in accuracy of classification of music data-sets as compared to traditional Random forest approach, but also has shown a consistent better performance as compared to other classification techniques

VIII. MOOD IDENTIFICATION SYSTEM

A. Mood Model Selection

A list of 2500 Indian popular songs which are well-known were chosen by surveying the songs most liked by majority of the people. A short experiment similar to Hevner's [1] was conducted with the help of a panel of 5 music listeners wherein each member of the panel was independently supposed to listen to a 30 second clip of each of the 2500 songs and note down the best adjective(s) that they think describes the song emotion aptly. The panel constituted of one Music expert, two avid music listeners and two common music listeners. The adjectives collected were then grouped together depending on the similarity and the music clip under consideration. A total of five groups of moods were categorized each covering a group of adjectives of songs with similar emotional quotient. These five categories of moods form our mood model as shown in the table.

1) *System Overview:* The Mood Identification system is the main engine which would help identify the mood of given music or audio files. This system is designed as an open source software system. The system would generally be a part of the back-end in most of the applications whose result can be used by the application layer on top to utilize the information in the required way. The system has two-fold objectives as mentioned below:-

The system should have a provision of analyzing music files and learn the classifier models associated with them
It should be able to predict the class of mood that a particular audio file or music belongs to.

Table 1: Mood Model: Indian popular Hindi music

Mood Category	Adjectives represented
Happy	cheerful, funny, comic, happy, jovial
Sad	depressed, frustrated, angry, betrayal, withdrawal, serious
Silent	peaceful, calm, silent, nostalgia, slow-paced
Excited	danceable, celebration, fast-track, excited, motivational, inspirational
Romantic	love, romantic, playful

©IJRASET: All Rights are Reserved

Volume 5 Issue VI, June 2017
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

An abstract view of the Mood Identification system as seen from a users' perspective is as shown in the Figure 3. The system accepts music files as input from the user and returns the respective mood associated with each file to the end-user.

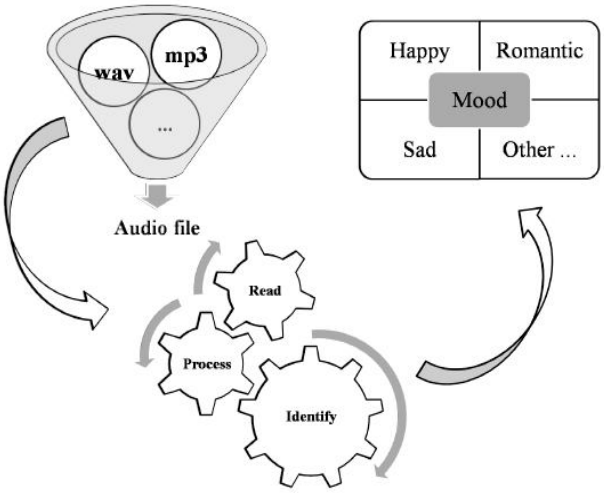


Figure 3: Mood Recognition System

B. Mood Identification System

This is the main processing unit of the whole system and is responsible for mining the mood from the music data-set obtained as input from the audio feature extractor module. It comprises of the actual implementation of algorithms. The module has two important roles to perform as mentioned below:-

1) Mood Learner: In this case, the input received is a training data-set of music features with the Mood attribute manually updated by the domain experts, from the point of view of training. The Mood learner can make use of the existing mining algorithms or newly written algorithms, provided they follow the convention and framework laid down by Weka tool [19]. Thus, this module can serve as the experimenter so that user - analyst or researcher - can utilize it to try various algorithms to mine mood aspect of the underlying music data-set. The classifier model learnt can thus be saved so that it can be utilized for further evaluation purpose. The output of this part of the module generally serves useful to end-users who are analysts or researchers, keen to understand and tune the machine learning aspect of this whole process. Using this model, the classifier model for bagging of Random forest approach was trained and store after careful evaluation and comparisons with other comparable models. Mood learning is generally one-time activity. Once done, the model is saved and can be re-used for evaluations any number of times. However, depending upon the user preference, the learning can be made iterative to improve accuracy with the most updated music data which evolves over time to a great extent. This change, however, might require few codes.

2) Mood Detector: In this case, the music data-set received as input will have some dummy data in the \Mood" attribute as this feature is not known ans is expected to be predicted by this module. The Mood detector then evaluates the data-set under consideration against the mood classifier model that has been saved. The evaluation results in predicting the mood for every 30 second music clip that was fed to the system by the user. In case a whole song was fed instead by the user, the system returns the maximum voted mood from the moods predicted for all of the clips derived from that song. The output of this module is generally used by the end-user application such as a mood-annotator or any Music information retrieval application or even the end-user he/she. Although the module helps in detecting the mood of the music under consideration, the whole and sole control of accepting or rejecting this decision can be always given to the end-user with some minor enhancements to the code.

IX. EXPERIMENTS AND RESULTS

The project involved a lot of rigorous experimentation from data mining point of view. In addition to it, the preparation and pre-processing involved in carrying out the experimentation is also worth mentioning. This section describes the experimental apparatus, flow and results obtained during the experimentation for music mood identification process.

A. Experimental Setup

The apparatus included:

1) A huge diverse personal music collection of Indian popular music in mp3 or

©IJRASET: All Rights are Reserved

wav format.

2) Open-source tools and libraries for audio processing.
3) Open-source data mining framework - Weka [19].
4) Music Mood Identification System.
5) Panel of five people - one Music expert, two avid Music listeners, two common music listeners
6) One workstation for software development, assembly and execution

B. Data Collection

The data collection involved personal music collection of Indian popular Hindi songs. Only those songs which are generally popular and famous among the people were selected and care was taken to ensure there is a good mix of collection of songs spawning across each of the five mood classes. Only songs belonging to MP3or WAV format were shortlisted in alignment with the scope of the project.

C. Training and Testing

The datasets in each stage were subjected to a range of various existing classification algorithms under numerous runs and folds. Those algorithms showing a bias towards only specific class labels or performing very low were discarded thereby subjecting the dataset to a 66%-34% training-testing split learning and evaluation for all the algorithms. Following are top 11 algorithms which have shown top comparable results during this experimentation:-

1) Naïve Bayes
2) Support Vector Machines
3) J48 (ID3 algorithm implementation)
4) Random Tree
5) Random Forest
6) REPTree
7) Simple CART (Classification and Regression Trees)
8) Bagging of Random Trees
9) Bagging of Random Forests
10) Bagging of simple CART
11) Bagging of REPTree

D. Results

The 11 classification algorithms were evaluated with respect to four evaluation measures for each of the datasets generated:-

1) *Receiver Operating Characteristic (ROC):* It shows the trade-off between the true positive rate and the false positive rate. It is a two-dimensional plot with vertical axis representing the true positive rate and horizontal axis representing the false positive rate. A model with perfect accuracy will have an area of “1". The area under the ROC curve is a measure of the accuracy of the model. It ranks the test tuples in decreasing order: the one that is most likely to belong to the positive class appears at the top of the list .The closer to the diagonal line (i.e., the closer the area is to 0.5), the less accurate is the model. Area under ROC was mainly used in signal detection theory and medical domain where in it was said to be the plot of Sensitivity verses 1- Specificity which is a similar plot as defined earlier. For each of the five classes of mood model, area under ROC is calculated and more the value nears “1", more accurate the classification is. Confusion Matrix: The columns of the confusion matrix represent the pre- dictions, and the rows represent the actual class.

Correct predictions always lie on the diagonal of the matrix. Equation 1 shows the general structure of confusion matrix.

$$\begin{bmatrix} TP & FN \\ FP & TN \end{bmatrix} \quad (1)$$

wherein, True Positives (TP) indicate the number of instances of a class that were correctly predicted, True Negatives (TN) indicate the number of instances NOT of a particular class that were correctly predicted NOT to belong to that class. False Positives (FP) indicate the number of instances NOT belonging to a class were incorrectly predicted belonging to that class and False Negatives

©IJRASET: All Rights are Reserved

Volume 5 Issue VI, June 2017
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

(FN) indicate the number of instances that were in- correctly predicted belonging to other class. Though the confusion matrix gives a better outlook on how the classifier performed than accuracy, a more detailed analysis is preferable which are provided by the further metrics. Since, in this case we have five mood classes, the confusion matrix will be a 5 X 5 matrix with each diagonal representing the True positives.

2) Recall: Recall is a metric that gives a percentage of how many of the actual class members the classifier correctly identified. (FN + TP) represent a total of all minority members. Recall is given by equation 2.

$$Recall = \frac{TP}{TP + FN} \quad (2)$$

3) Precision: It gives us the total the percentage of how many of a particular class instances as determined by the model or classifier actually belong to that particular class. (TP + FP) represents the total of positive predictions by the classifier. Precision is given by equation 3.

$$Precision = \frac{TP}{TP + FP} \quad (3)$$

Thus in general it is said that Recall is a Completeness Measure and Precision is a Exactness Measure. The ideal classifier would give value as 1 for both Recall and precision but if the classifier gives higher (closer to one) for one of the above metrics and lower for the other metrics in that case choosing the classifier is difficult task. In such cases some other metrics as discussed further are suggested in the literature.

4) F-Measure: It is a harmonic mean of Precision and Recall. We can say that it is essentially an average between the two percentages. It really simplifies the comparison between the classifiers. It is given by the equation 4.

$$F - Measure = \frac{2}{(\frac{1}{Recall} + \frac{1}{Precision})} \quad (4)$$

Figures 1, 2, 3, 4 depict the performance of the algorithms with reference to the four measures namely, AUROC, Recall, Precision and Fmeasure. From each of the results, it can be seen that Bagging (ensemble) approach of classification tree algorithms like Random Forest, Random Tree and Simple CART showed better results as compared to other algorithms, and Bagging of Random Forest topped among all consistently.

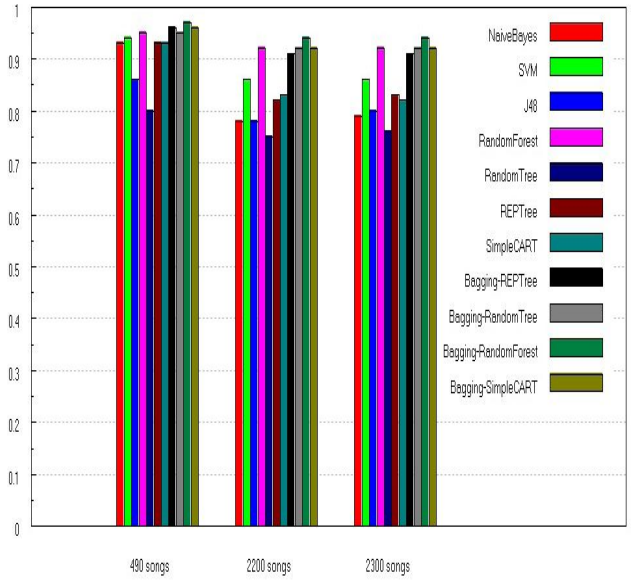


Figure 1: Area under ROC statistics

©IJRASET: All Rights are Reserved

Volume 5 Issue VI, June 2017
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

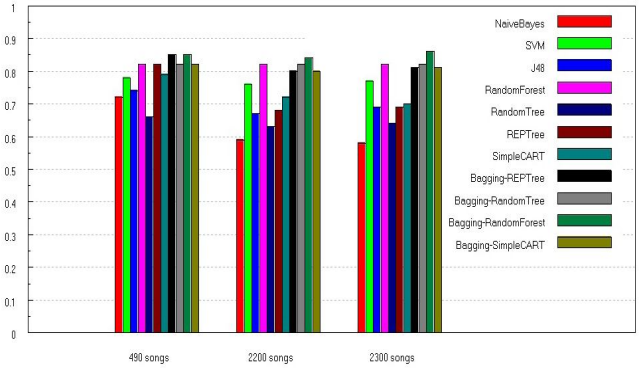


Figure 2: Recall statistics

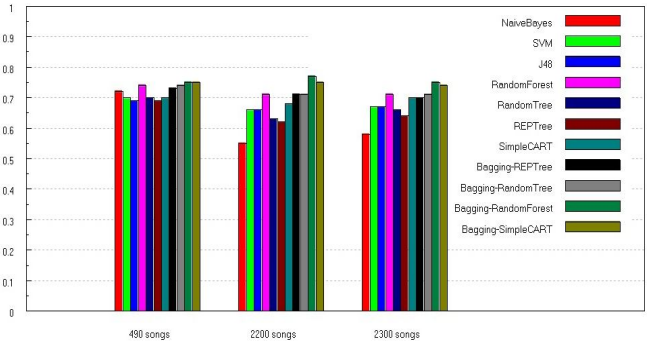


Figure 3: Precision statistics

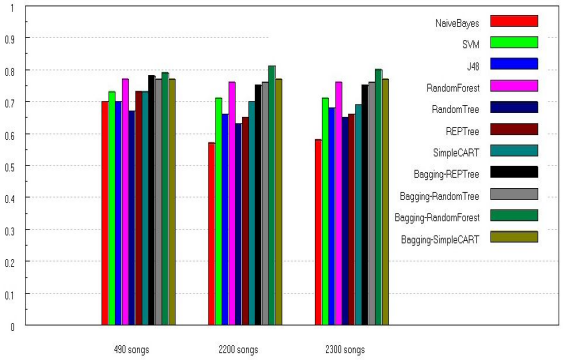


Figure 4: F-measure statistics

Table 2: Experimental Results on Test Dataset of 2938 music clips

TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Mood Class
0.964	0.106	0.751	0.964	0.845	0.991	excited
0.805	0.021	0.914	0.805	0.856	0.978	happy
0.77	0.006	0.971	0.77	0.859	0.967	romantic
0.822	0.019	0.867	0.822	0.844	0.977	sad
0.871	0.038	0.849	0.871	0.86	0.983	silent
0.853	0.042	0.867	0.853	0.853	0.98	← Avg.

©IJRASET: All Rights are Reserved

Volume 5 Issue VI, June 2017
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

Table 3: Experimental Results on Test Dataset of 2938 music clips

a	b	c	d	e	← Classified As
704	16	1	0	9	a = excited
69	**511**	7	15	33	b = happy
94	16	**470**	10	20	c = romantic
34	5	1	**314**	28	d = sad
36	11	5	23	**506**	e = silent

The Table 2 shows the evaluation results obtained for the said metrics after performing a test run a dataset of 2938 music clips belonging to Indian popular Hindi music. The table shows the classification performance for each of the mood category defined in the mood model. The last row represents the metric values with weighted average taken over all the classes. The Table 3 displays the confusion matrix for the evaluation of the Test dataset of 2938 music clips belonging to Indian popular music. As seen from the matrix, the diagonal elements marked bold are the correctly identified data instances and denote the True positives. From the data seen in the matrix, following can be inferred:-

Total number of instances: 2938

Number of correctly classified instances: 2505 (85.26%)

Number of incorrectly classifier instances: 433 (14.74%)

X. CONCLUSION AND FUTURE WORK

The Bagging of Random Forest approach thus performed much better as compared to not just other decision tree based algorithms but other classification algorithms as well. This was a new observation in case on analysis of Indian popular music unlike western music where SVM and neural network algorithms dominated the classifier accuracy. The classification performance achieved seems satisfactory so far thus making it useful for use in real applications. The open source framework developed as a part of the project also serves as a common framework for music data mining analysis in terms of an end-to-end solution. Although the current approach has proved satisfactory results, we consider this as just a first step in exploring Indian popular music and it opens avenues for further research and developments in this are to bring more efficient results.

A. Future Work

The path forward involves a further cycle of experimentations and refinement of the audio features and the mood categories if required, so as to enrich the dataset in addition to increased number and variety of songs to extract further valuable information for mood learning. During this development cycle the mood model also has a chance to likely undergo some changes to suit best the Indian Song scenarios. The current system is capable of recognizing the mood of songs of 30 second duration. Further this can be extended to derive the mood of the entire song by collectively recognizing and weighing the moods recognized for the 30-second trimmed clips of the song. In future, this system can be extended to other genres of Indian songs like Hindustani classical, Carnatic music with changes involving audio features and classification techniques. Customization of this system to non-Indian songs cannot be ruled out as well after a thorough experimentation. Since some of the moods songs like Hindustani classical, Carnatic music with changes involving audio features and classification techniques. Customization of this system to non-Indian songs cannot be ruled out as well after a thorough experimentation. Since some of the moods represented by Indian popular music are very much governed by expressions, which are very well conveyed through lyrics, lyrics analysis in combination with audio features can make the system much stronger with a better accuracy.

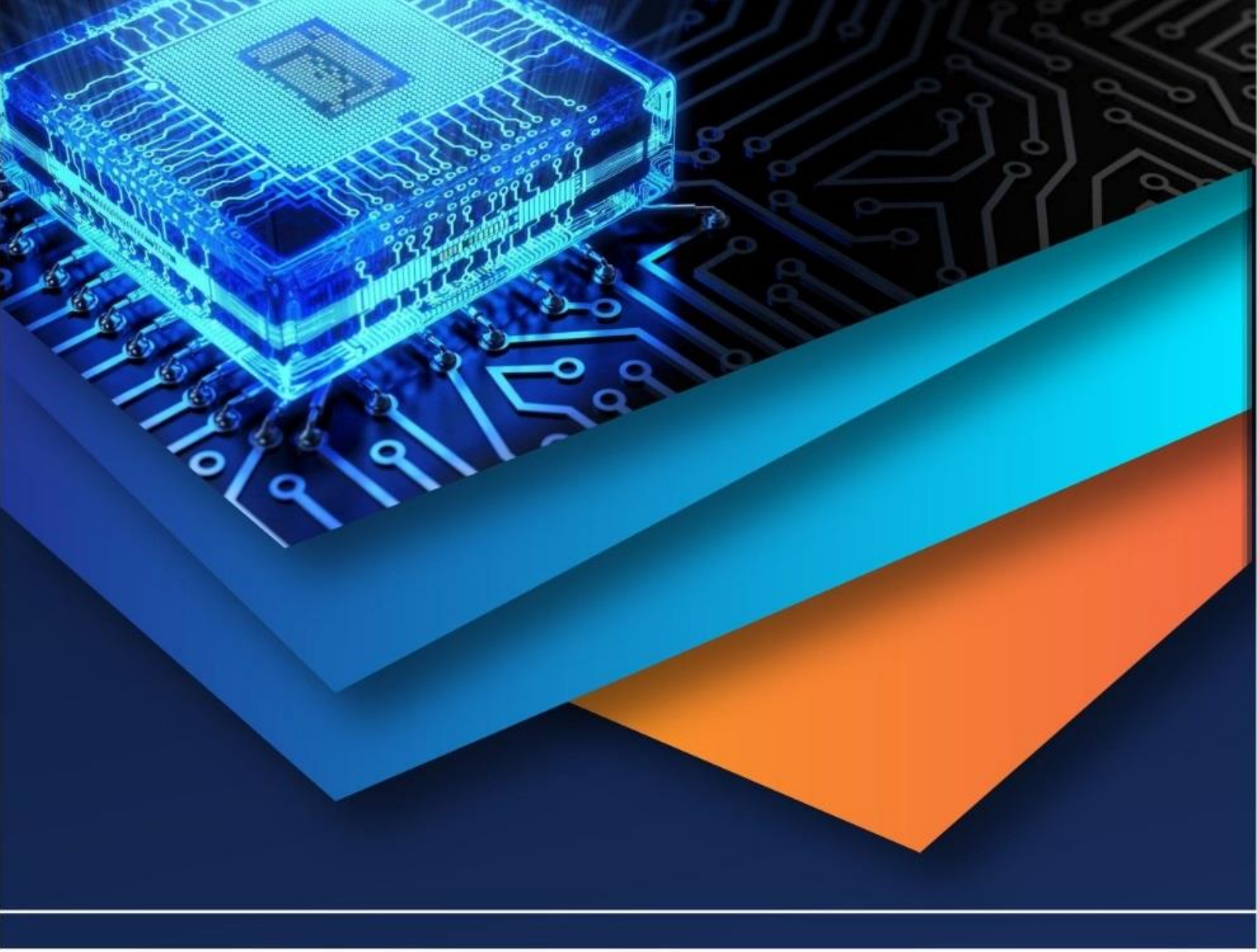
REFERENCES

[1] Kate Hevner, (1936), \Experimental studies of the elements of expression in music, American Journal of Psychology, 48:246268.
[2] Paul R. Farnsworth, (1958), \The social psychology of music, The Dryden Press.
[3] Mirenkov, N., Kanev, K., Takezawa, H., (2008), Quality of Life Supporters Employing Music Therapy", Advanced Information Networking and Applica-Applications - Workshops(AINAW).
[4] Dalibor Mitrovic, Matthias Zeppelzauer, Horst Eidenberger, (2007), \Analysis of the Data Quality of Audio Descriptions of Environmental Sounds", Journal of Digital Information Management, 5(2):48.
[5] Capurso, A., Fisichelli, V. R., Gilman, L., Gutheil, E. A.,Wright, J. T., (1952),Music and Your Emotions", Liveright Publishing Corporation.

©IJRASET: All Rights are Reserved

[6] Weihs, C., Ligges, U., Morchen, F., Mullensiefen, D., (2007), Classification in music research.", Advance Data Analysis Classification, vol. 1, no. 3, pp. 255291.
[7] Scaringella, N., Zoia, G., Mlynek, D., (2006), Automatic genre classification of music content, A survey.", IEEE Signal Processing Magazine, vol. 23, no. 2, pp. 133141.
[8] Zhouyu Fu, Guojun Lu, Kai Ming Ting, Dengsheng Zhang, (2011), A Survey of Audio-Based Music Classification and Annotation", IEEE Transactions on multimedia, Vol. 13, No. 2.
[9] Russell J. A., (1980), A circumflex model of affect", Journal of Personality and Social Psychology, 39: 1161-1178.
[10] Thayer, R. E., (1989), \The Bio-psychology of Mood and Arousal", New York: Oxford University Press.
[11] JungHyun Kim, Seungjae Lee, SungMin Kim, WonYoung Yoo, (2011), \Music Mood Classification Model Based on Arousal-Valence Values", ICACT2011, ISBN 978-89-5519-155-4.
[12] Tsunoo, E., Akase, T., Ono, N., Sagayama S., (2010), \Musical mood classification by rhythm and bass-line unit pattern analysis", Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP).
[13] McEnnis, D., McKay, C., Fujinaga, I., Depalle P., (2005), \jAudio: A feature extraction library", Proceedings of the International Conference on Music Information Retrieval. 6003.
[14] Dalibor Mitrovic, Matthias Zeppelzauer, Horst Eidenberger, (2007), \Analysis of the Data Quality of Audio Descriptions of Environmental Sounds", Journal of Digital Information Management, 5(2):48.
[15] Ekman P., (1982), \Emotion in the Human Face", Cambridge University Press, Second ed.
[16] Duda, R. O., Hart P. E., (2000)\Pattern Classification", New York Press:Wiley.
[17] Marsyas, http://opihi.cs.uvic.ca/marsyas.
[18] Attribute-Relation File Format, http://www.cs.waikato.ac.nz/ml/weka/arff .html
[19] Weka, http://www.cs.waikato.ac.nz/ml/weka/

©IJRASET: All Rights are Reserved



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)