A

Dissertation

on

Mood Identification on the basis of lyrics and audio clips

Submitted in partial fulfillment of the Requirement

For the Award of the Degree of

Master of Technology

in

Computer Science and Engineering

by

Meenu

University Roll No. 2K13/CSE/11

Under the Esteemed Guidance of

Dr. Kapil Sharma

Associate Professor, Software Engineering Department, DTU

2013-2015

COMPUTER ENGINEERING DEPARTMENT

DELHI TECHNOLOGICAL UNIVERSITY

DELHI-110042, 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 suits our mood at that instant. In spite of this strong correlation, most of the music softwares present today are 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. This task is to automatically mark a song using affective labels in an emotion set specified by psychologists. In the last few years, it has attracted more and more attention and wide range of related researches have been carried out. We take the same inspiration forward and contribute by making an effort to build a system for automatic identification of mood underlying the audio songs by mining their spectral, temporal audio features. We will built a hybrid system that will depict the mood of the song on the basis of lyrics as well as music. Our focus is specifically on India 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 music by introducing successful to achieve a satisfactory precision of 70% to 75% in identifying the mood.

ACKNOWLEDGEMENT

I would like to express my deep sense of gratitude to my project supervisor Dr. Kapil Sharma for providing the opportunity of carrying out this project and being the guiding force behind this work. I am deeply indebted to him for the support, advice and encouragement he provided without which the project could not have been a success.

I am grateful to Dr O.P. Verma, HOD, Computer Engineering Department, DTU for his immense support. I would also like to acknowledge Delhi Technology University library and staff for providing the right academic resources and environment for this work to be carried out.

Last but not the least I would like to express my sincere gratitude to my parents and friends for constantly encouraging me during the completion of this work.

Meenu

University Roll No. :2K13/CSE/11

M.Tech (Computer Science and Engineering)

Department of Computer Engineering

Delhi Technological University

Delhi-110042

Computer Engineering Department
Delhi Technological University
Delhi-110042
www.dce.edu

CERTIFICATE

This is to certify that the dissertation titled "**Mood Identification on the Basis of Lyrics and Audio Clips**" is bonafide record of work done by **Meenu**, **Roll No. 2K13/CSE/11** at **Delhi Technological University** for the partial fulfillment of the requirement for the degree of Master of Technology in Computer Science and Engineering. This project is carried out under my supervision and has not been submitted elsewhere, either in part or full, for the award of any other degree or diploma to the best of my knowledge and belief.

(Dr. Kapil Sharma)

Date:_______

Associate Professor and Project Guide

Department of software Engineering

Delhi Technological University

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Correlation between Mood and Music

"Music heals everything" says the philosopher Hemeto Pascoal, implying that music is essential part of our lives. Moreover, music can act as a healing agent if listened in a specific way. In the old days; music was limited to radio broadcasts or live concerts. But as the technology is increasing new multimedia gadgets are coming in the market which is easily accessible to almost every human being, maybe it's a laptop or an ipad or mobile phone. Everyone can have the latest music at his fingertips within few seconds, result of which is a large music database. Music collection is increasing day by day with many music pieces releases every single day. So, there are much variety of music available in one's music device based on artists, albums, music type etc. or we can say we have our personal choices which too may differ time to time. This choice is much more influenced by the emotional behavior of a person at a particular instant of time. Hevner (1) and Farnsworth studied the relation between music and their effect on human beings. The papers store, retrieve our music files based on different tags like artists, group, movie or genre. To select a piece of music according to one's mood at that particular instant we need an extra search attribute which is mood, because all the above stated attributes cannot precisely depict the mood. There are many difficulties depicting the mood of a song. The reasons for this are as follows:

- First, subjectivity of the mood of the music. The emotion of the music can be affected by surrounding environment, mood, cultural background or personality.
- Second, ambiguity in finding the adjectives of mood. For example happy and excited can be referred to same song.
- Third, it is perplexing as to how music stimulates emotions. We could not find out that what inherent property of music initiates a specific emotional response.

1.2 Introduction to Music Features

There are intrinsic feature present in music which helps in determining the mood or emotion attached with that particular song. So, it is very important to know those features in order to catch the mood of song. For capturing these various features, based on researches going on we can categorize and analyze the song. Audio feature can be defined as the statistical identities that are computed over music pieces to gain information about that data. Various audio features have characterized till date. Dalibor Mitrovic and his partners (2) have analyzed various states to art audio features, which are important for content based audio retrieval.

For the application domain such as speech recognition (3) audio feature were primarily studied and discovered. After some time, the analysis of audio files gained pace and importance. Different research fields came into existence like music information and retrieval (MIR), audio segmentation, environmental sound recognition (ESR). Each field has their own type of features. There are various audio features for music classification which are proposed. Audio feature were classified into four classes by Weihs et al. (4) which are compositional features, semantic features and long term features. Zhouyu Fu and Team (5) present a hierarchical taxonomy that symbolizes audio features from dissimilar perspective level. A was followed by Scaringella (6) more specific classification which divides features determining genres classification into the group which give the information about timbre, rhythm and pitch. So, we can split audio features into two basic levels which are low level and mid-level features. Temporal and timbre features are the subclass of the low level features. Tonal value of song that is linked to different instrumentation is caught by timbre feature. Low level audio features are zero crossing rate, magnitude spectrum, spectral roll-off, spectral centroid etc. Temporal features describe the deviation and progress of timbre with time. Low level features are the basic description of audio data, such as, tempo, beats per minute etc. whereas mid-level features uses rhythm, pitch, mood which is important of them all. The low level features make a base for these mid level features.

Experts have studied various features in the past years and had made a great progress in standardized some features. MPEG7 (3) can be example of such features which provide slow level audio features list. It also offers procedures and means to obtain the features. Complex mathematical and signal processing which converts digital audio data into features is used by the audio feature extraction. These features are represented by numbers. Orthogonal features are the demand of the research which provides descriptors with extreme discrepancy for the input data.

1.3 Data Mining and Machine Learning

To find interesting and useful patterns from the large dataset is called data mining. It analyzes the large dataset from database. Data mining is interdisciplinary and a new field which involves various tasks like analyzing data and then to classify it. Classification involves generalization of an existed structure among the data that is already assigned some label. For forecasting the class of a new and unknown data this structure can be used. Clustering is another important feature of data mining which involves discovering groups and structures of data. Our current problem of mood identification is a data mining problem because we have to classify so many music pieces. In machine learning things learn when they change their behavior in a way that makes them perform better in the future. Initially pattern recognition technique was used for mood detection. MIDI files were used by Wang et al. (4) to extract features and used an SVM to classify music into 6 main classes which were robust, lyrical, restless, joyous, gloomy and sober. Li et al (7) divided emotions into thirteen categories but in six classes only. After that they tried MARSYAS collected audio feature and used support vector machine or learning and determining of mood. Liu etc. (8) used mood recognition system, which is hierarchical. He used a Gaussian mixture model to characterize the feature dataset and to classify the audio files, a Bayesian classifier is used.

So we can sum up that may be its genre classification or instrument classification or mood classification, data mining techniques proved much effective in analyzing and categorizing music.

1.4 Motivation

We can search a song by different tags such as artist, name of song, album, band, genre etc. but this type of search is restricted when it comes to search a song according to the mood of a person at that instant. So, a new parameter had to be added to so that the person can choose a song according to his mood. But it is not that simple to implement this idea so we have to interpret the music type with this mood tag. Most of the time this tag is manually edited which is prone to any errors. So we want to build a system that automatically detects the mood for a particular song and for a particular instant.. It will lower the errors and manual work will be avoided. Music can be organized in a better way according to the user's mood, with this way. Many different sources came into the picture after researches, with different solutions to the same problem that is identifying mood from song's text file which provides effective data which is different from that is fetched from music piece. So, we will make a hybrid model that combines both audio features and lyrics to find out the mood of a song effectively. Moreover we are working on Bollywood music which has many flaws. Like may be a song is sad and it would be interpreted as calm. Another problem associated with Bollywood music is ambiguous meaning of the words. But we find it as a challenge, and will try to work with the Bollywood music.

1.5 Thesis objective and scope

Researchers have worked on non-Indian music. As we all know music is subjective to cultural backgrounds so Bollywood music has different aspects to deal with. We want to build a mood identification system for Bollywood music by analyzing various aspects of music pieces, and lyrics. For generation of such system we need to achieve these goals:

- To build a mood identification system for Bollywood music.
- To develop an open-source framework that can help exploring and testing music data with different machine learning and data mining algorithm.

We can achieve the goal with these sub goals:

- Various mood identification associated with Bollywood music
- Identification of music features that helps in depicting the mood
- Developing the means for obtaining the audio features.
- Naming and constructing the data mining technique.
- Design implementing and testing.

The scope of our thesis is only Bollywood music.

1.6 Thesis Outline

Chapter 2 is about description of important papers and literature that are referred and used in our work. Chapter 3 is about the mood identification models. In chapter 4 we have discussed about data mining features associated with mood. Chapter 5 is about proposed work. In chapter 6 contains the results and analysis of mood identification systems. Chapter 7 is applications and future use of the project.

Chapter 2

Literature Survey

Many researchers in the fields of psychology have proposed various models which expresses human emotions. Hevner (1) specifies eight different class of mood, but the adjectives of same type of mood are kept together. Russell (9) plotted the emotions in a circle representing different moods on its circumference which varies in a small amount from each other. After that 'Thayer' (10) came up with a different model in which he considered the two dimension stress and energy as the axes of the two dimension space. Vahida Z Attar influenced by Thayer's model defined the low level audio features, being a super class of timbral and temporal features. They come up with an accuracy of 75% to 80% while F-measure had in range 70% to 75%. They worked with Indian popular music and used RandomTree and simpleCart algorithms. The area under the ROC curve was found to be 0.91 to 0.94. Yi Liu and Yue considered the audio features such as Rhythm, Timbre and Intensity. They used support vector machine as classifier and studied Gaussian Mixture Model. With Relief, SFS, fisher and active, feature section algorithm accuracy was found to be 84%. For Microarray data mining Hanna M.Hussain, Khaled Benkrid, Huseyin Seker and Ahmet T Erdogan designed FPGA. Using k mean algorithm for microarray data speeds up the analysis of datasets. Shang Lei implemented Information gain, a feature selection algorithm. Fewzee, Pouria and Karray Fakri implemented person's correlation coefficient classification algorithm. This algorithm predicts the continuous emotional content. Vallabha Hamiholi has presented a method to depict mood for Bollywood music. It is based on Thayer's model consisting four mood types which are as follows:- Exuberance, anxiety, serene and depression. Audio features which are related to intensity, timbre and rhythm are extracted from the audio files. Success rate in this was achieved in this model was found to be 60%. When the same model was applied to western music accuracy achieved was 40%.

The mood classification model for music recommendation system was proposed by Jung Hyun Kim and Won Young Yoo. They evaluated moods tags and AV values from 20 subjects and categorized music mood on AV plane into 8 regions with the use of k-means clustering algorithm. They presented that some zones can be recognized by representative mood tags but some mood tags are overlain all zones. In Zhouyu Fu, Guojun Lu, Kai Ming Ting (5) and Dengsheng Zhang have proposed the recent development in music classification.

Yongkai Zhao, Deshum Yang, Xiaoou Chen came up with a model that deals with the audio files, lyrics and features extracted from MIDI form. B.Han and group proposed SMERS which uses SVM. Chia-Chu Liu and his partners devised a system for pop music, using hierarchical framework. Pang and Lee (2008) defined sentiment polarity, and political viewpoints. They called the lyrics as bag of words, part of speech tags, position in the documents, higher order n grams. Stylometric analysis is authorship attribution, text genre identification, and classification. Rudman suggests that more than one thousand stylometric features have been defined. Subasic and Huettner made a word lexicon. Pang and Lee (2008) categorized the lexicon building method into two sets of supervised and unsupervised learning. Supervised lexicon can be defined as a group of language model comparison; F-score feature ranking and SVM feature ranking. To start with few seed words and label them, as they are already known. Other words are like synonym or other similar words made the corpus. Alm(2008) included a set covering the aspects like syntactic, rhetoric, lexical, and orthographic. Alm included in the feature set dimensional lexical scores calculated from ANEW (Affective Norms for English Words) word list. This list contains 1034 unique words in with scores in three aspects which are valence, arousal and dominance. However, on the basis of text prediction does suggest possible features for consideration in this research.

Chapter 3
Music Mood Models

Most of the literature dealing with music and psychology tells us that music mood is subjective and the mood of the same music piece can be interpreted differently by different individual. However, it is seen that there are considerable agreements about the moods underlying the music belonging to similar cultural and ethnic context (11). Thus music belonging to similar cultural background has a better chance of consensus among the various persons in interpreting the mood of the song. Our work limits the scope to India popular music which falls under a common platform thus increasing the chances of similar interpretations of the music among the individuals, when it comes to understanding the mood. In order to classify songs according to their mood, it is essential to identify the list of moods which a song can be categorized into. This chapter explores the various mood i.e. happy, sad, excited, calm etc. models that have been proposed and proven constructive in categorizing music as per the emotions.

3.1 Relation between Music and Mood

Many distinct mood categories and taxonomies were proposed. There are many different words that describe the mood and most of them are similar to each other. These words may be adjectives, so there is the need of normalizing the terms used, since there is not standard and uniform mood taxonomy. Categorical and dimensional are two main mood nomenclatures.

The first one which is the categorical approach consists of several different classes of emotional states whereas, dimensional models maps emotional states along several axes.

Categorical models has to be further sub dived. We may call them discrete and continuous. What is the difference between the two? Let us take the case of high arousal and high valance quadrant. Now several moods like happy, excited or pleasure can be mapped into this quadrant. This may lead to ambiguity. In discrete models it is assumed that some finite numbers of moods are possible. In case of continuous model every point I the quadrant is considered as a separate state/

3.2 Mood Models

Mood Models are generally studied by two approaches:-

- Categorical approach: in this approach, it defines different classes of emotions which again form the basis for the various moods.
- Dimensional approach: This classifies mood on the basis of several parameters such as emotions along several such as pleasure, arousal, and potency and so on. This is generally the most commonly used approach in music applications

A number of models have been proposed by psychologists on human emotions. Musicologists have too adopted and extended some models that we will be checking later. The six universal emotions depicted by Ekman (12) are disgust, fear, sadness, happy, fear and anger. Since these were basically intended for facial expressions basically, all of these may not be fit for audio files.

3.2.1 Hevner's Model

Hevner mapped a group of 67 adjectives divided into eight different emotional categories (each one containing from 6 to 11 adjectives), which were mapped into a circular model. The main emotions in each emotional category are dignified, sad, dreamy, serene, graceful, happy, exciting and vigorous (from group 1 to 8, respectively) (Meyers, 2007). Hevner also made pioneering studies that tried to relate emotions and moods to music – six musical features which are rhythm, harmony, tempo, pitch ,mode and melody were studied, with main conclusion that music and emotions were indeed connected and that music clearly carries an emotional meaning (Laar, 2005).

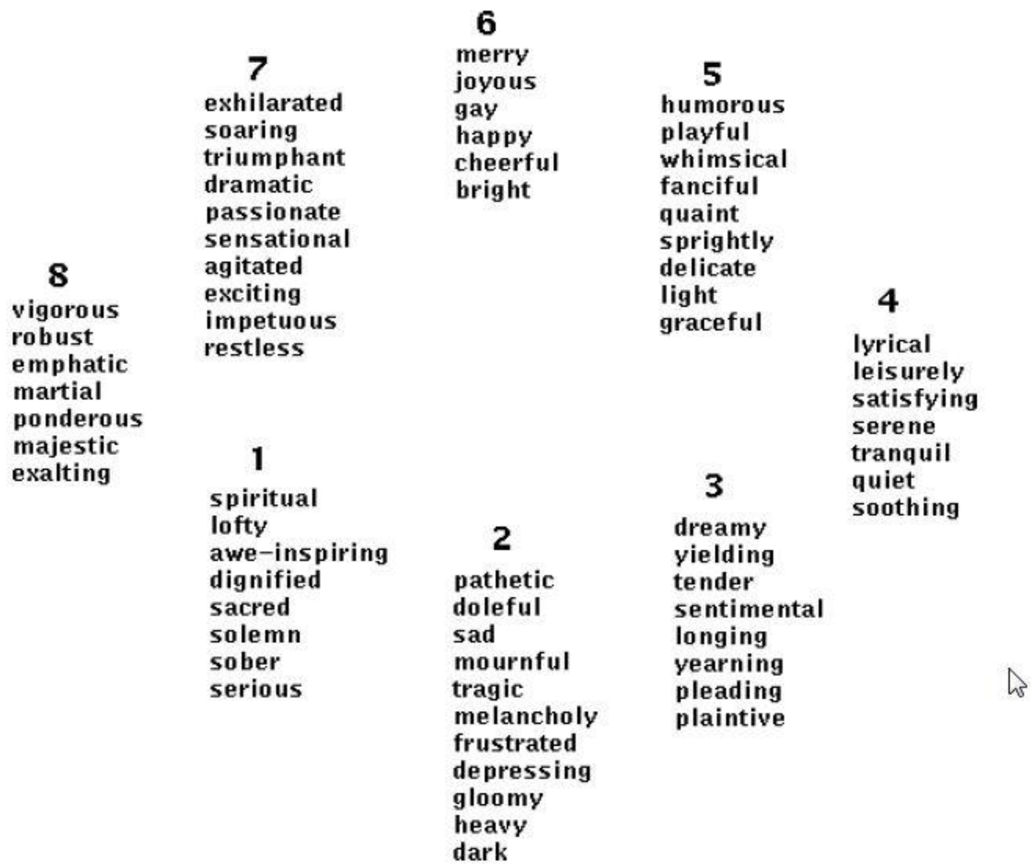

Figure 1 Hevner's Model

3.2.2 Russell's Model

Both Ekmans and Hevners models belong to Categorical Models. On the contrary, James Russell (9)came up with a model of emotions arranging 28 adjectives in a circle on two-dimensional bipolar space (arousal - valence). This model helped in separating and keeping away the opposite emotions. Figure below depicts the Russell’s model which has-been adopted in a extensive musical psychology studies (9) (13) (14).

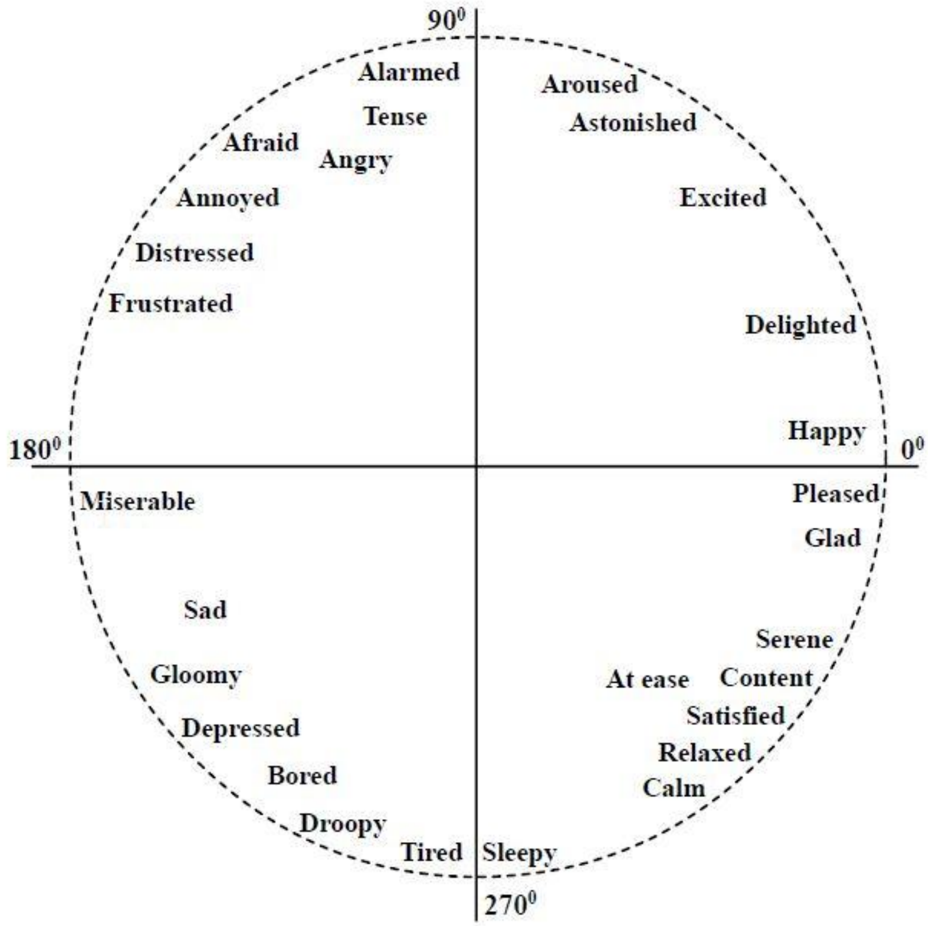

Figure 2 Russell's Model

3.2.3 Thayer's Model

A popular and more recent approach is based on the simple Robert Thayer's Model (15) (8) a dimensional model where a musical piece can be classified in one of four categories mapped into a four-quadrant, bi-dimensional space. This approach has its advantages, since although simple it relates two different dimensions effectively, and so each category could be the result of some combination of the valence and arousal components (16). It describes the mood with two factors on the basis of Stress and energy and then splits mood into four different groups which forms the four quadrants in 2D space. Two-dimensional space: Contentment, Depression, Exuberance and Anxious/Frantic. In this model, Contentment implies to be happy and calm music; Depression implies to be calm and anxious music Exuberance refers toand energetic music; and energetic music is defined by the Anxious. These four clusters have the clear definition of four different moods. This categorization of music emotion in four classes provides a temporary music model and it is broadly adopted in mood recognition studies. Figure 3 depicts the Thayers model for mood.

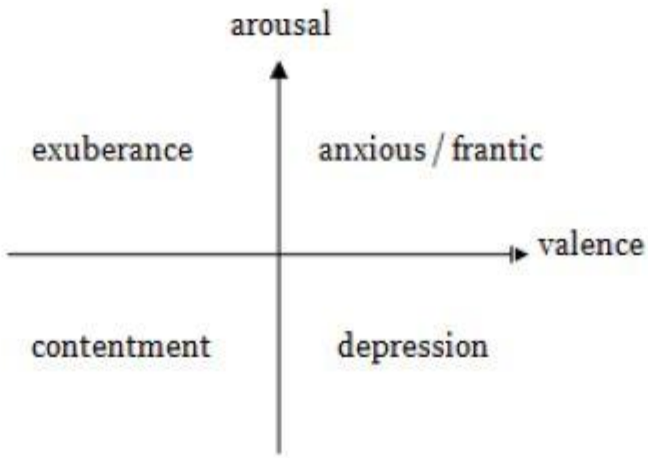

Figure 3 Thayer's Model

3.2.4 Tellegen-Watson-Clark Model

In 1999, the Tellegen-Watson-Clark model was proposed. This is a more complex model, as it can be seen in Figure 4. It contains many more moods and emotions than Thayer's model. The idea behind this dimensional model is to relate positive/negative affective rate as one dimension being pleasantness/unpleasantness and other is engagement/disengagement (which are 45º rotated) as another. This is more understandable looking at Figure below, where the model is represented.

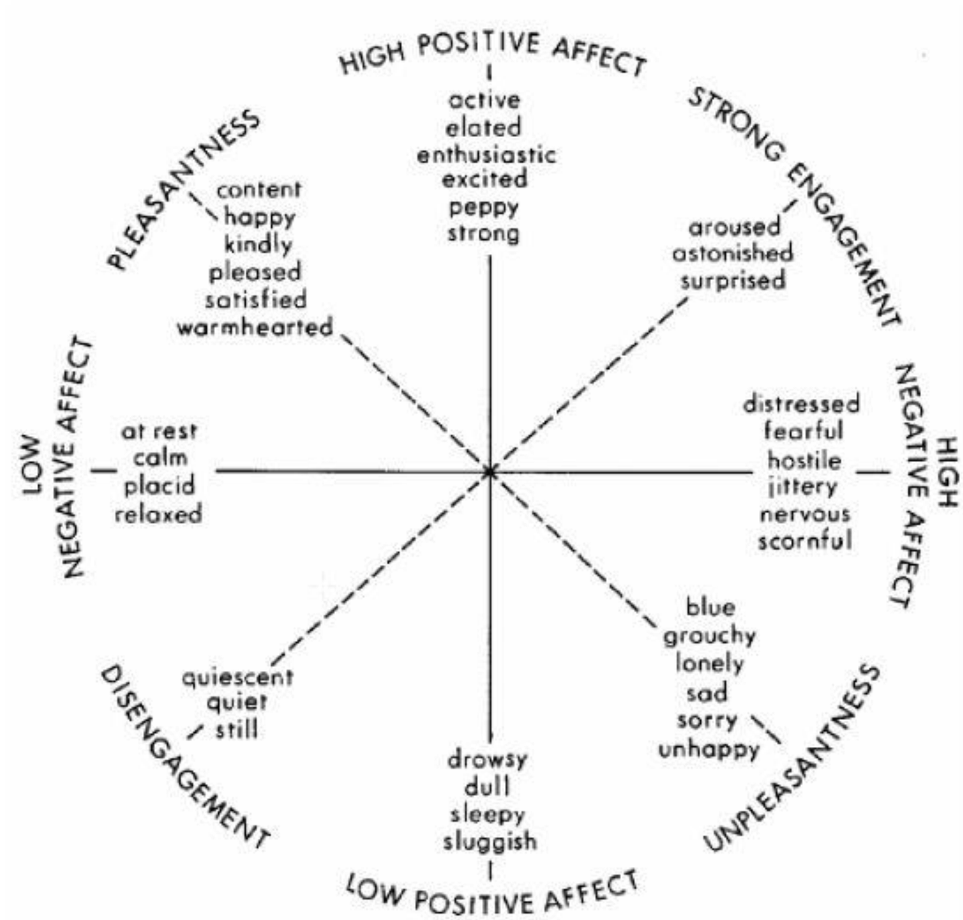

Figure 4 Tellegen-Watson-Clark Model

3.2.5 Indian Classical model: Navras

For consideration of Indian Bollywood Music, we need to know the traditional mood model that is prevalent since ancient times in the Indian Classical Music which forms the base for India music. Navras as it is name in Sanskrit, it means nine sentiments. This model sums up all the major categories of emotions into total nine classes. These nine sentiments are depicted in the Figure below.

Figure 5 Indian Classical model: Navras

Studying the advantages and short-coming of various models learned so far and taking into consideration the Indian popular music scenario, to come up with one of the mood models exactly from the existing ones mentioned in the literature cannot do justice in selecting the mood categories. Hence, we try to put forth a simple mood model covering majority of mood aspects after careful study and experiments as would be witnessed in coming chapters.

Chapter 4

Data Mining and Machine Learning

4.1 Data Mining Overview

In real life we come across vast quantity and type of data. To understand better some examples can be quoted. A scientist dealing with weather has a data for 365 days. For simplicity it can be assumed that data consists of several attributes like daily temperature, humidity, rainfall etc. Now in raw form this data is hardly useful because the probability of event that a useful conclusion will be drawn by manually looking at it. So data is useful only when some patterns can be drawn out from it and use these patterns to gain advantages in future. Data mining deals with the first part i.e. finding patterns from existing data whereas machine earning deals with the second part i.e. first using such information for training and classifying the raw input and draw conclusion on it.

Data mining has several components. First, data should be electronically stored. Second, process of finding these patterns should be automated. Third, pattern found should be meaningful. Now meaningful can be interpreted in several terms. The most common interpretation is that it should be economically useful.

Machine learning is a process by which there is an improvement in the behavior of the machines and this leads to the betterment in performance next time.

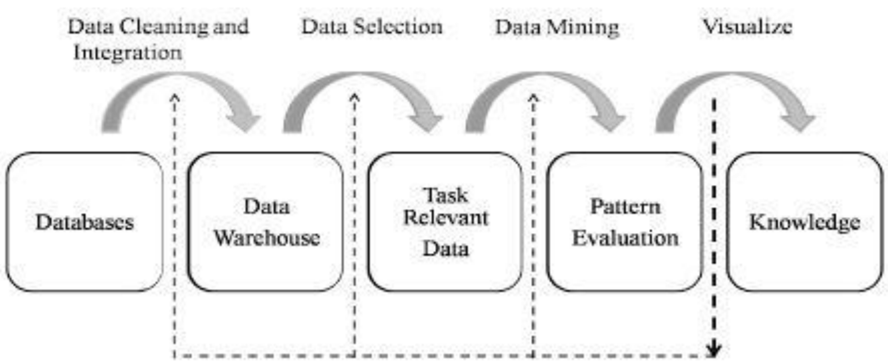

Figure 6 Data Mining

4.2 Output: Knowledge Representation

The output knowledge can be represented in various forms. Some of these are tables, linear models, trees, rules, clusters etc. Each one of these has their own qualities. Trees are based on “divide-and-conquer” method whereas in linear models weights are attached to attributes. So in case of linear models main focus is on finding the weights. In case of cluster the output is in the form of a diagram. For the purpose of this project tree based classification model has been used. This is described in next section.

4.2.1 Classification using a Decision Tree

Decision tree method is one of the several methods that is used in data mining and machine learning. Tree consists of a root node, branches and leaves. Internal nodes represent attributes where as branches represents outcome of a test. Leaf node represents class label. But this depends on the data we are classifying. In some cases the leaf nodes may give probability distribution.

Tree models are of two types: classification trees and regression trees. Target variable in case of classification tree can only a finite set of values whereas in case of regression tree it can take continuous values. In other words output in case of classification tree is a “class” whereas in case of regression tree it can be a real number.

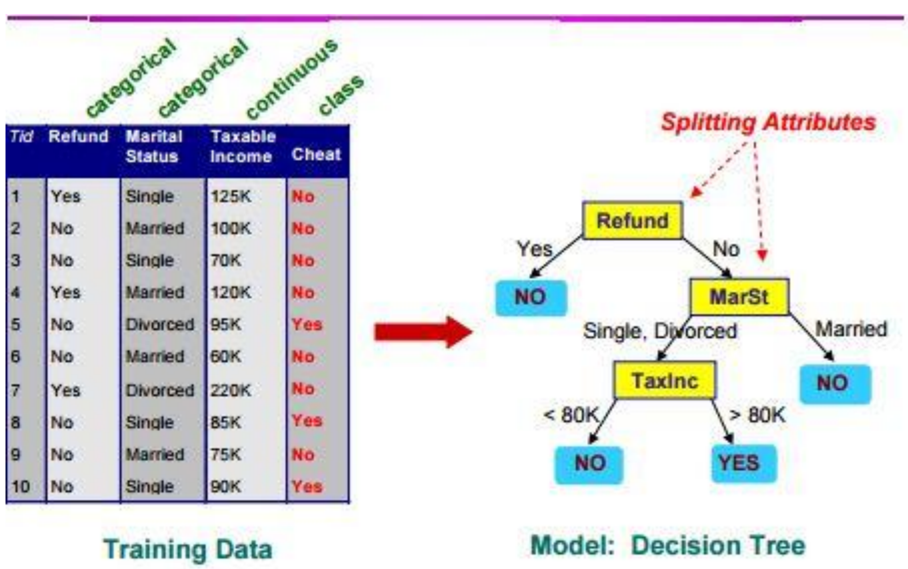

Figure 7 Classification Using Tree

Why Decision trees?

There are several reasons why decision tress should be used for classification. Some of them include less domain knowledge requirement, ability to handle high dimensional data, generally fast learning and classification, output easily understandable by humans and accuracy is generally high.

4.2.2 Classification using C4.5 Algorithm

C4.5 is an algorithm that was developed by the efforts of Ross Quinlan starting from 1970s. It is a tree classifier. Earlieran algorithm known as ID3 was developed by Quinlan. C4.5 is an extended version of this ID3 algorithm. C4.5 develops a decision tree. One a tree is generated it can be used to solve classification problems.

Figure 8 shows the pseudocode for the algorithm. For the purpose of generation of tree C4.5 make use of concept “information entropy”. ID3 also makes use of the same concept.

C4.5 is an improvement over ID3. It can handle can continuous and discrete data. Also, it can handle missing attributes. After creation of the tree, the again goes through the tree and removes some useless branches. These branches are replaced by leaf nodes. This process is known as pruning.

Open source java implementation of C4.5 algorithm is J48. This is included in weka-an open source data mining tool.

Later on C5.0 was created by Quinlan. It has several advantages over C4.5. It is faster than C4.5 and also takes less memory than C4.5. In case of C5.0 decision trees that are produced are smaller than C4.0.

Algorithm 1.1 C4.5(D)

Input: an attribute-valued dataset D

Tree = {}
if D is "pure" OR other stopping criteria met **then**
    terminate
end if
for all attribute $a \in D$ **do**
    Compute information-theoretic criteria if we split on a
end for
a_{best} = Best attribute according to above computed criteria
Tree = Create a decision node that tests a_{best} in the root
D_v = Induced sub-datasets from D based on a_{best}
for all D_v **do**
    $\text{Tree}_v = \text{C4.5}(D_v)$
    Attach Tree_v to the corresponding branch of Tree
end for
return Tree

Figure 8 C4.5 Algorithm

Freely available open source software weka was used for data mining purpose for this project.

Chapter 5

Proposed Hybrid Mood Classification System

5.1 Mood Model Selection

The mood models were studied from the psychological perspective as reflected in the literature review. If music emotions are mapped on proposed dimensional models then these different emotions could be plotted in two dimensional spaces. After that one has to group together the various emotions to get different clusters of moods. Also it should be remembered that large number of different moods may confuse the end user, so a reasonable number of moods should be used to build a model.

Some emotions like anger, surprise, horrific cannot be fully described by music alone, so proposed model should use easily identifiable and easily expressible moods.

Cluster1	Cluster2	Cluster3	Cluster4
Excited	Delighted	Calm	Sad
Astonished	Happy	Relaxed	Gloomy
Aroused	Pleased	Satisfied	Depressed

Figure 9 Modified MIREX Model

In my work, a modified version of MIREX mood cluster model has been used. Cluster containing mood as "anger" was removed because in Indian tradition showing anger is not the main purpose of music and secondly anger cannot be represented by audio and lyrics also and face expression and stress play an important role.

5.2 Mood Classification Model Using Audio Features

5.2.1 Audio Features

Feature extraction and classifier Learning is the key components of a classification system (17). Feature extraction helps in representing the music pieces to be classified in terms of feature vectors or pair-wise similarities. In the literature, many audio features have been proposed for classification of songs. For categorization of audio features various taxonomies are present. There are various audio features for music classification which are proposed. Weihs et al. (4)

have categorized the audio features into subcategories like long term features, semantic features and compositional features. Zhouyu Fu and Team (5)present a hierarchical taxonomy that characterizes audio features from different perspective level. More standard taxonomy which divides audio features used for genre classification into the group based on timbre, rhythm and pitch information respectively is also followed. So, we can divide audio features into two basic levels which are low level and mid-level features. Low level features are of two types timbre and temporal features. Timber features depicts the tonal quality. Tonal quality is related to the instrument tones. Low level audio features are zero crossing rate, magnitude spectrum, spectral roll-off, spectral centroid and many more. Variation and evolution of timbre can be depicted by temporal feature. Low level features are the basic description of audio data, such as, tempo, beats per minute etc. whereas mid-level features uses rhythm, pitch, mood which is important of them all. The low level features make a base for these mid-level features.

Experts have studied various features in the past years and many of the features have been standardized. MPEG7 (3)can be example of such features which provides a list of low level audio features. It also provides the techniques and tools to extract the features. Complex mathematical and signal processing which converts digital audio data into features is used by the audio feature extraction. These features are represented by numbers. Orthogonal features are the demand of the research which provides descriptors with a high variance for the underlying data.

Low-level features although not directly related to the fundamental properties of music as perceived by human listeners form the basic features which can be used to derive the mid-level features. Mid-level features include mainly three classes of features, namely rhythm, intensity, and timbre.

Audio Intensity: Audio signal intensity is very important when it comes to mood detection. Audio files with high intensity relate to *Exuberance* and *Anxious* whereas with low intensity relate to *Serene* and *Depression*. The above observation about signal intensity has also been noted in several works (18) .

Timbre:.Timbre makes a particular musical piece sound different from another, even if they have the same pitch and loudness(19). Timbre plays an important role in human perception of music.

Tones with many higher harmonics are related to *Anxiousness* and *Exuberance*, whereas with few, low harmonics can be associated with *Serenity* and *Depression* (20).

Rhythm: Human mood response is dictated by the tempo and rhythm periodicity present in a musical piece (18). Through rhythm features like Beat Sum, Strongest Beat and Strength of Strongest Beat we also can get some information about the emotion of the song i.e. positive or negative. Usually Fast songs are happier than slow ones. We extracted rhythm features including Beat Sum, Strongest Beat and Strength of Strongest Beat. In several research works it has been established that these features play an important role in determining the mood of the music (21). Regular rhythm with fast tempo may be perceived as expressing exuberance, while irregular rhythm with slow tempo conveys anxiousness.

5.2.2 Feature List

Following is the list of selected features:

- Magnitude spectrum
- Power spectrum
- Spectral centroid
- Spectral rolloff point
- Spectral flux
- Compactness
- Spectral variability
- Root mean square
- Fraction of low energy windows
- Zero crossing
- Beat sum
- Strength of strongest beat
- MFCC
- LPC
- Strongest frequency via zero crossings
- Strongest frequency via spectral centroid

Some of the above features are one dimensional and some are multi-dimensional. So a consolidated feature vector was prepares. Both running mean and standard deviation was calculated for above listed features.

5.2.3 Audio Pre-processing

Each of the song that was to be used for listening test was split and an audio clip of 30 second was extracted from it using freely available software FormatFactory. 30 second has been proved a good duration and in most cases it is easy to correlate a particular song to a particular mood.

For the standardization purpose, each of the 30 second music clips was converted to WAV format. Again freely available software Format Factory was used for this purpose. This was necessary so as to ensure that each file that is processed was treated equally.

5.2.4 Preparation of ground truth

For the purpose of building this proposed system a list of 320 Indian popular Bollywood songs were selected. Songs for this experiment were collected mostly for popular Indian music websites. Personal collection of songs was also used. To prepare the ground truth a small listening experiment was conducted with the help of four people. People in the panel were made to listen to a 30 second audio clip and predict the mood of the same clip. Some of the points and problems faced that are worth mentioning are as follows:

- Some 30 second audio clips contain lyrics. Sometimes it is possible that mood expressed by audio and lyrics are different. For example a music clip having high valance consists of lyrics that belongs to sad class. So, we tried that lyrics of the songs were ignored and only audio was used to predict the mood of the song. This procedure of ignoring lyrics has been followed in some other works.
- The people in the people were already familiar with few songs. In this situation they knew about the lyrics of the song. This must have had an impact on their decision indirectly regarding the mood of the song.
- We considered only 30 second clips. It is seen that in Hindi music mood of the song changes few times during the course of full song. Also few songs are slow and it is not possible to come at a conclusion by listening to a 30 second clip.

- We considered 320 songs out of which nearly 80 songs were considered from each class to avoid class imbalance problem.
- The people in the panel were from same age group. They may have had an impact on the quality of the ground truth.
- If more than 1 people predicted the mood which was different from others prediction, the song was dropped from the list.

5.2.5 Audio Feature Extraction

JAudio was used for the feature extraction process. It is a toolkit developed in JAVA language for music features extraction. Also, this toolkit is available freely for research purpose. The extracted features can be broadly classified in these categories: timbre, intensity, rhythm. The first three sets can express mood information to some degree and are necessary for mood identification. Complete list of the extracted features is given in some other previous sections. JAudio can produce output in arff format that can be later used by Weka.

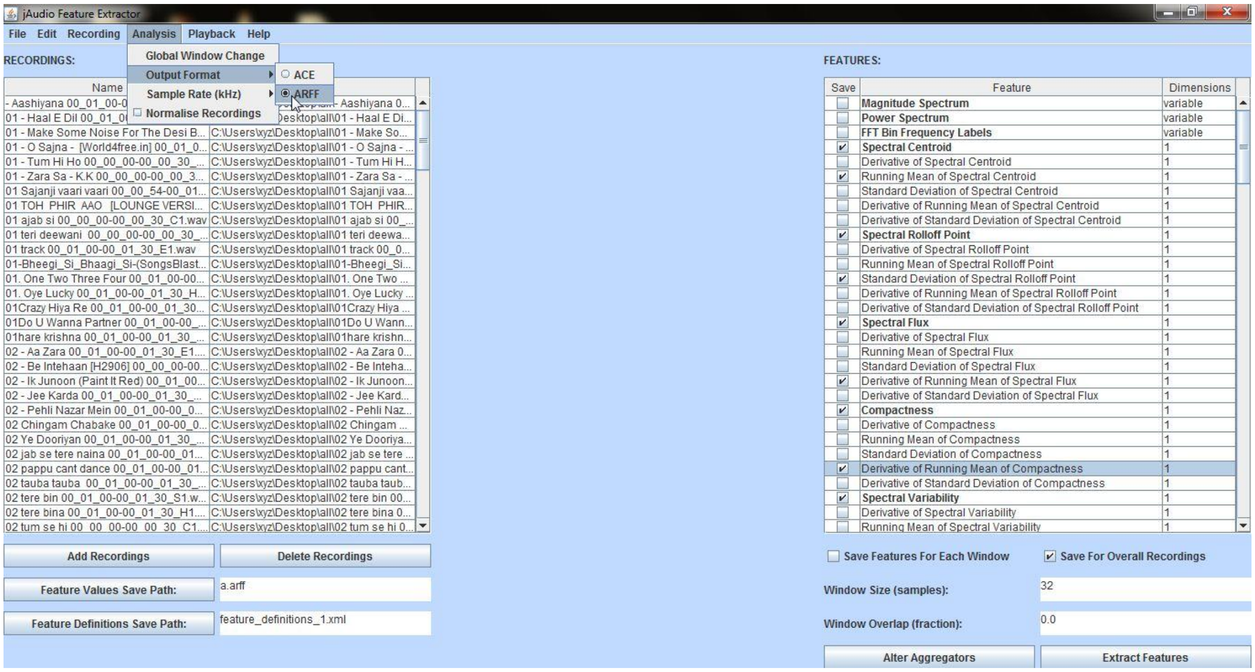

Figure 10 jAudio GUI

5.2.6 Data-set generation

The feature vectors thus extracted from the attributes of the each music clip - which can be called as a data instance. These feature vectors computed in the memory are stored in a file following the standard ARFF file format understood by most of the data mining tools like Weka. An additional feature named mood was manually appended to this file. This mood attribute was decided by a listening test as described in an earlier section.

5.2.7 Feature Selection

Generally it is observed that dataset prepared have many low information features. The primary job of a classifier is to classify the instances of the dataset in various classes as specified by the nominal attribute or class attribute. A low information feature is one which doesn't help in classifying the instances. This is because the value for this attribute is similar for the instances belonging to different classes. If such features are removed data set can be made smaller. It is also easy for the classifier to make decision regarding classification.Information gain helps to determine which attribute in a given dataset of training feature vectors is most useful for discriminating between the classes to be learned (22).

These are several features that are more or less common across all clusters (or "classes) and therefore contribute little information to the classifier for classification process. Low information features can decrease the efficiency of the model if taken collectively.

Eliminating low information features increases the performance of model and brings clarity by removing duplicate noisy data. But using higher information features it is ensured that model takes less time for training and testing and also consumes less memory.

5.2.8 Mood Identification

This mood identification process makes use of the extracted features and consists of two parts i.e."Mood Learner" and "Mood Detector". The data set was randomly partitioned in two parts. One consisting of 80% instances was used for training purpose and other consisting of 20 % was used for testing purpose. Various models were generated using various classification algorithms. Training is one-time activity. Once completed, the model can be saved and can be re-used for evaluations later. The complete result of various models and test dataset is shown in later chapter.

5.3 Mood Classification Model Using Lyrics

Based on the examination of some previous works (23) (24), it was found that lyrics are a weak source of mood information. But lyrics when combined with some other model like Audio based classification model can improve the efficiency. This can be understood but this simple logic.

Two song , one belonging to class “calm” and another one belonging to “sad “ class may have slow pace and audio features based classification model may not accurately classify them. But a distinction can be made between the two by analyzing the lyrics because one song that belongs to sad class has high probability of having words that create a saddening effect like “tutna”, “bewafa” etc. In simpler words lyrics can be used to distinguish between “positive emotional category” and “negative emotional category”.

5.3.1 Collection and Preprocessing

To the best of my knowledge there is no central lyrics database for Bollywood songs**.** So, one has to depend on various websites for the collection of lyrics. Lyrics were extracted from internet by querying the google search engine with keywords like “track name” + “singer name”. In earlier this method has showed high recall but low precision (25). Some major problems were faced during this process. Some of them are mentioned below.

- Standardization of lyrics was a major problem. Different websites have different formats for lyrics. Some website like “geetmanjsha” doesn’t consider the repeating lines. It means irrespective of how many times a line is repeated in a song (in one go) it is written once.
- The above problem creates another problem during the classification. How many times a particular word has appeared in the song is a major factor in determining the mood of the song.
- Different websites have used different spellings for same word. It mean “pyar” and “pyaar” are same. But classifier cannot distinguish between them.
- Another problem was of word fragmentation. It means that the word “abcxyz” was mentioned as two separate words “abc” and “xyz”.
- Many lyrics contained the extra information (nearly useless for my purpose) like “written by” , “address of the webpage” , instrumentation (e.g., “(SOLO PIANO)”), and/or the performing artists.

  Some common patterns that were observed are as follows:

- Any of the words in () or [].
- For more refer to the page.
- “Written by” at the end of the song
- “Fade to end” at end of a lyric which represents the sound effect.

- Lines with “http://” or “www”.
- Annotation of part of song which is followed by a number indicating times how many times the segment has been repeated: e.g., [Chorus (2x)] etc.

Despite above mentioned hurdles during preprocessing stage lyrics were extracted and normalization process was carried out manually i.e. an effort was made to standardize the lyrics but this manual effort has its own limits.

5.3.2 Data Set Generation and Model Building

Before discussing the model building method it is necessary to focus on some inherent shortcoming associated with Hindi lyrics.

First, Hindi is not a homogeneous language like English. When a writer writes a song may word from local language are also included. This creates a problem during classification.

Second, stemming is not efficient in case of Hindi lyrics. For example “raat” and “ratiya” both means the same i.e. night. But classifier cannot identify it. So, stemming algorithms are not very helpful. Considering above disadvantages lyrics based mood classification model was created.

Some basic lyrics features are as follows:

- Content words (Content): all words except function words, without stemming;
- Content words with stemming : stemming is the process of combining words that have the same roots;
- Part-of-speech (POS) tags: such as noun, verb, proper noun, etc.
- Function words (FW): In case of text information retrieval these are also called “stopwords” in text information retrieval.

For each of the feature types, more than one representational model can be compared. Some are: 1) Boolean; 2) term frequency; 3) normalized frequency. In a Boolean representation model, is is checked if a particular word is present in lyrics or not. In case of term frequency and normalized frequency as the name suggests, it is observed how many times a term is present.

Weka provides a function for string to words or tokens conversion. Final model is based on term frequency and normalized frequency model. “N-gram” tokenize algorithm or stemming algorithms were not used due to problems mentioned in previous sections.

5.4 Hybrid Model for Mood Classification

Hybrid models can be used to improve the performance of the classifier. They are generally used to integrate heterogonous data sources. They work best in the situation when one can possibly make up for mistakes of other. For this data should be of sufficiently diverse nature. Several types of hybrid models have been proposed for mood classification. Some examples are as follows:

- Audio + Lyrics which have shown some to high improvement
- Audio + Tags which have shown good improvement
- Audio + Images based hybrid model using album art to derive associations to mood

In my case, Audio and lyrics based hybrid model was created.

5.4.1 Two Hybrid Methods

Two popular hybrid methods have been proposed to combine information retrieved from various sources. The easiest and simplest hybrid model is one in which feature concatenation takes place i.e. more than one feature vectors are combined together to form one feature vector and the classification algorithms run on the combined features set.

The other method is to combine the output of the individual classifier which is based on different sources of information. The combination can be either on the basis of averaging or by multiplying. This is known as “late fusion model”. The above discussed two methods are also known as “serial hybrid model” and “parallel hybrid models” respectively.

In my case, first approach was followed. A combined feature vector was created by combining audio feature vector and lyrics feature vector. Then classification algorithm was used to create a classification model based on this combined feature set. Results of the experiment are summarized in next chapter

Chapter 6

Experiments and Results

A lot of rigorous experimentation was involved in the project if considered from data mining point of view. To carry out experimentation a lot of preparation and preprocessing was required. In this chapter summary of experimental apparatus and result that were obtained are presented. Complete details of the preprocessing are presented somewhere else.

6.1 Summary of Experimental Setup

6.1.1 Mood Classification Using Audio Features

To carry out the experiment following preparations were made:

- Collection of popular Indian music for various websites.
- Conversion to standard format (wav) and clipping to obtain a 30s clips.
- Preparation of ground truth by conducting a listening test.
- Extraction of features using freely available tool Jaudio.
- Preparation of basic classification model using open source data mining tool weka.
- Refinement of classification model using feature selection and information gain algorithms.
- Testing using final refined model.

6.1.2 Mood Classification Using Lyrics

To carry out the experiment following preparations were made:

- Collection of lyrics of the songs from various websites.
- Preprocessing and conversion to standard format (removal of extra alphabetic characters and website references).
- Preprocessing using weka that is required for text classification.
- Preparation of basic classification model.
- Refinement and testing of model.

6.1.3 Mood Classification Using Both Audio and Lyrics (Hybrid Model)

To carry out the experiment following preparations were made:

- Creation of a Hybrid features vector using both audio and Lyrics feature vectors.
- Creation and refinement of classification model
- Testing of classification and result recording.

(The details of each step involved in this process and various shortcomings are mentioned at various places)

6.2 Results

6.2.1 Evaluation Metrics

Different classification models were created using different classification algorithms and were evaluated with respect to following evaluation measures for each datasets generated:

- *Receiver Operating Characteristic (ROC)*: In a two dimensional plot Y axis (vertical axis) represents the true positive rate and X axis (horizontal axis) represents the false positive rate. ROC curve shows the tradeoff between TPR and FPR. If the model has perfect accuracy, it will have an area of "1".The area under the ROC curve is the measure of the accuracy of the classification model. In other words more the area under ROC curve more the accuracy of model. Test tuples are ranked in decreasing order. If the tuple is most likely to classified to positive class, it appears at the top of the list. More the area under curve closer to "0.5" (i.e. the curve is closer to the diagonal) less accurate is the model. For each of the four classes of mood (excited, happy, calm, sad) ROC is calculated. More the value closer to "1" more accurate is the classification model.

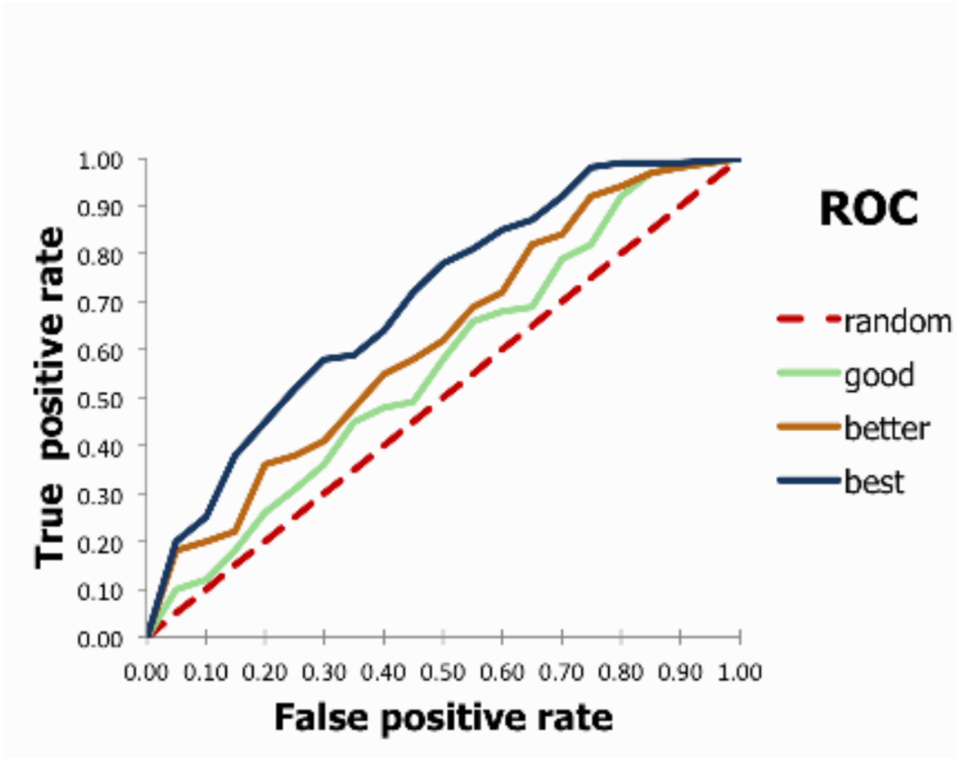

Figure 11 ROC

- *Confusion Matrix:* Confusion matrix is n by n matrix. “n” is the number of classes. General structure of a confusion matrix is shown below by figure 11. Rows are representing actual class where columns are representing predicted class.

		PREDICTED CLASS	
		affinity_card = 1	affinity_card = 0
ACTUAL CLASS	affinity_card = 1	516	25
	affinity_card = 0	10	725

Figure 12 Confusion Matrix for a binary classifier

- True Positives (TP):The number of instances of a class correctly predicted.
- True Negatives (TN): the number of instances NOT of a particular class that were correctly predicted NOT to belong to that class.
- False Positives (FP):The number of instances that didn't belong to a class but incorrectly predicted belonging to that class.
- False Negatives (FN): The number of instances incorrectly predicted belonging to other class.

Since, in my case I have four mood classes; the confusion matrix will be a 4 X 4matrix where diagonal represents the True positives. From the above numbers following can be easily calculated:

- True Positive Rate: $TPR = \frac{positive\ \ correctly\ \ classified}{total\ \ positives} = \frac{TP}{P}$
- False Positive Rate: $FPR = \frac{negatives\ \ incorrectly\ \ classified}{total\ \ negatives} = \frac{FP}{N}$

- *Recall:* It is defined by following equation:

$$Recall = \frac{TP}{TP + FN}$$

It gives a percentage of how many of the actual class members were correctly classified by the classifier.

- *Precision:* It is given by following equation:

$$Precision = \frac{TP}{TP + FP}$$

i.e. proportion of the instances which truly have class "x" divided by total instances classified as class "x".

- *F-Measure:* Harmonic mean of Precision and Recall is known as F-Measure. Comparison between the classifier is simplified to a great extent. It can also be said that it is an average between two percentages. It is given by the following formula:

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

Next section summarizes the result of the experiment for different mood models and different classifiers.

6.2.2 Evaluation Metrics – Graphical Representation

6.2.2.1 True Positive

1.True Positive					
	J48	**Random Forest**	**Naive Bayes**	**Bagging**	**Voting**
Audio	52.5	67.5	61.5	65	66.25
Audio Features	58.75	67.5	66.25	62.5	67.5
Lyrics	62.5	64.5	65.4	65	64.5
Hybrid	74.1	79.1	71.6	77.08	80.08

Table 1 True Positive rate for different classifiers (in %)

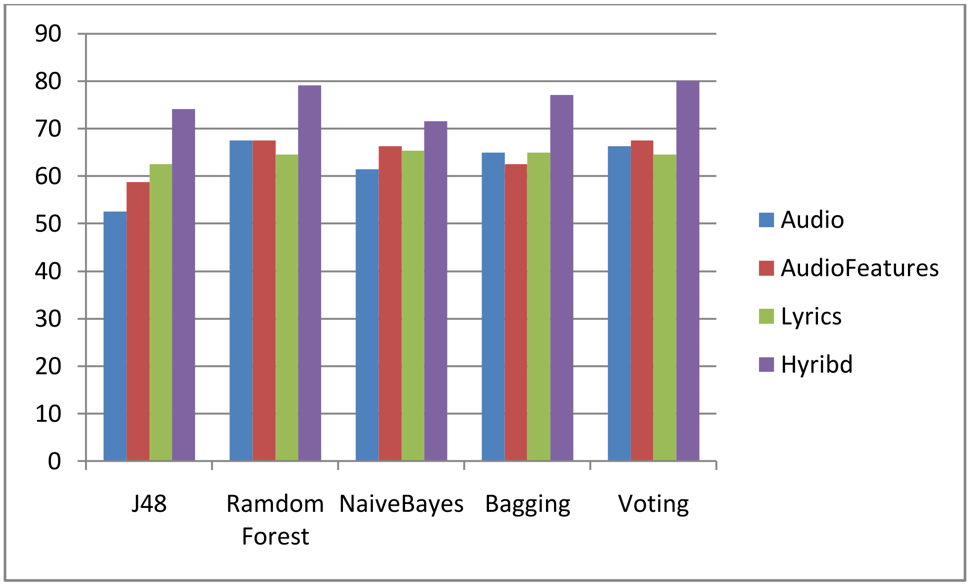

Figure 13 True Positive Rate – Bar Diagram

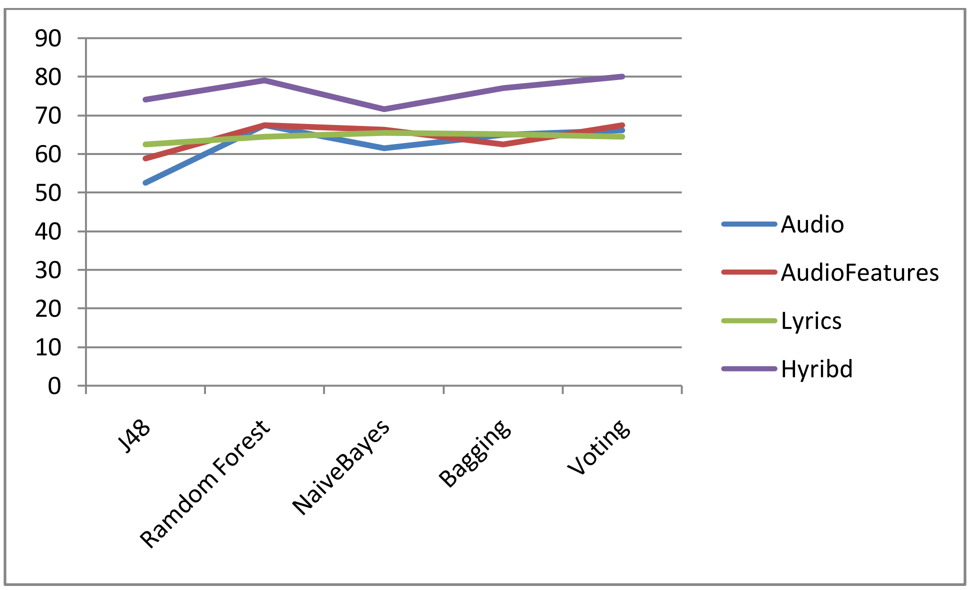

Figure 14 True Positive Rate-Line Diagram

6.2.2.2 Precision (Weighted Average)

2.Precision Weighted Avg.					
	J48	**Random Forest**	**Naïve Bayes**	**Bagging**	**Voting**
Audio	0.538	0.675	0.604	0.639	0.659
Audio Features	0.586	0.677	0.678	0.614	0.683
Lyrics	0.624	0.685	0.651	0.647	0.652
Hybrid	0.743	0.79	0.724	0.77	0.806

Table 2 Precision (Weighted Average)

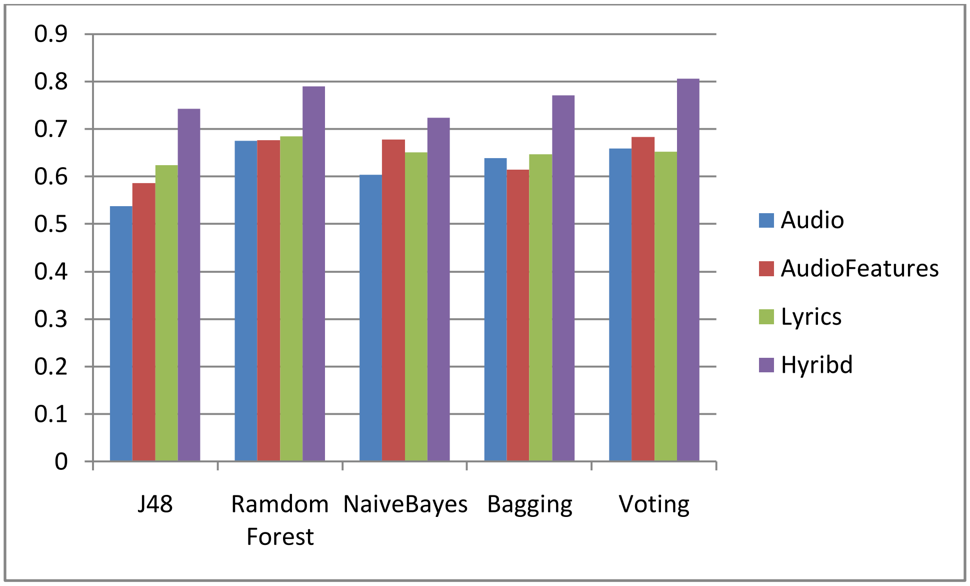

Figure 15 Precision for different classifiers

6.2.2.3 Recall (Weighted Average)

3.Recall Weighted Avg.					
	J48	**Random Forest**	**Naïve Bayes**	**Bagging**	**Voting**
Audio	0.538	0.675	0.613	0.65	0.663
Audio Features	0.588	0.675	0.663	0.625	0.675
Lyrics	0.625	0.646	0.654	0.646	0.65
Hybrid	0.742	0.792	0.717	0.771	0.808

Table 3 Recall for different classifiers (Max value "1")

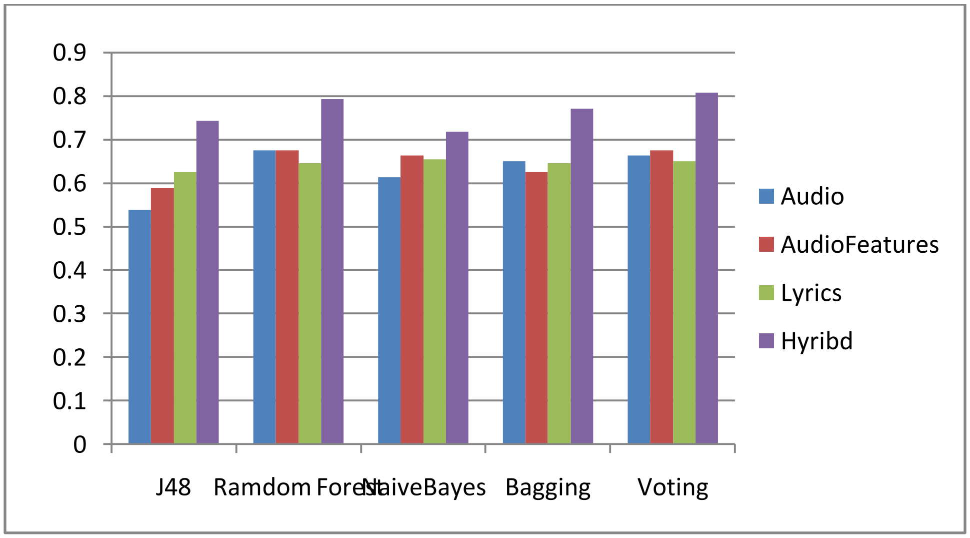

Figure 16 Recall for different classifiers

6.2.2.4 F-Measure (Weighted Average)

4.F-Measure Weighted Avg.					
	J48	**Random Forest**	**Naïve Bayes**	**Bagging**	**Voting**
Audio	0.533	0.655	0.589	0.639	0.649
Audio Features	0.582	0.657	0.664	0.616	0.664
Lyrics	0.622	0.649	0.648	0.643	0.649
Hybrid	0.739	0.79	0.71	0.769	0.805

Table 4 F-Measure for different classifiers (Max value "1")

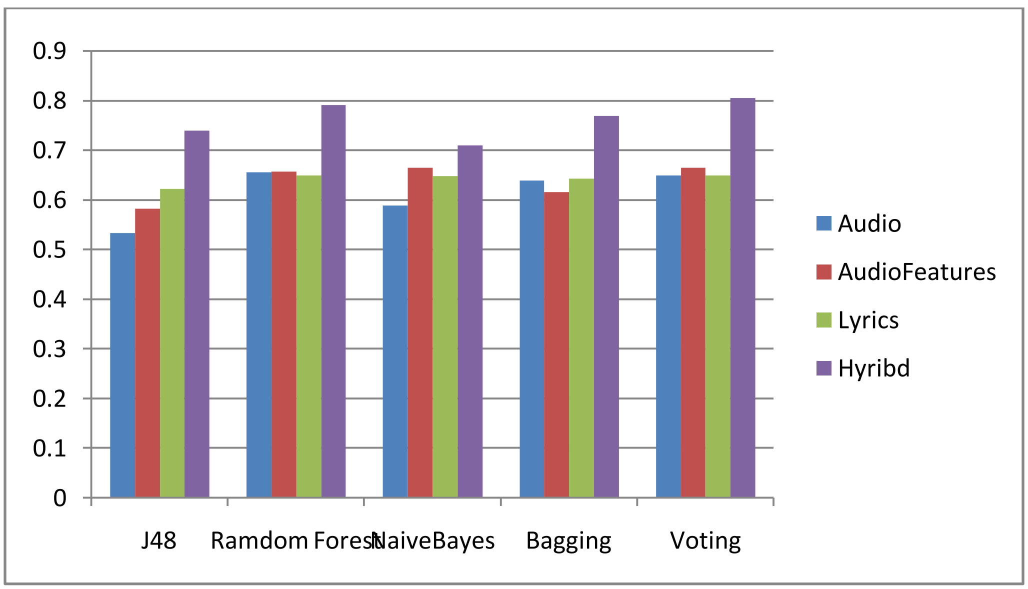

Figure 17 F-Measure for different classifiers

6.2.2.5 ROC (Weighted Average)

5.ROC Weighted Avg.					
	J48	**Random Forest**	**Naïve Bayes**	**Bagging**	**Voting**
Audio	0.708	0.873	0.831	0.835	0.849
Audio Features	0.727	0.896	0.852	0.849	0.884
Lyrics	0.76	0.878	0.812	0.87	0.842
Hybrid	0.834	0.947	0.909	0.926	0.952

Table 5 ROC for different classifiers (Max value "1")

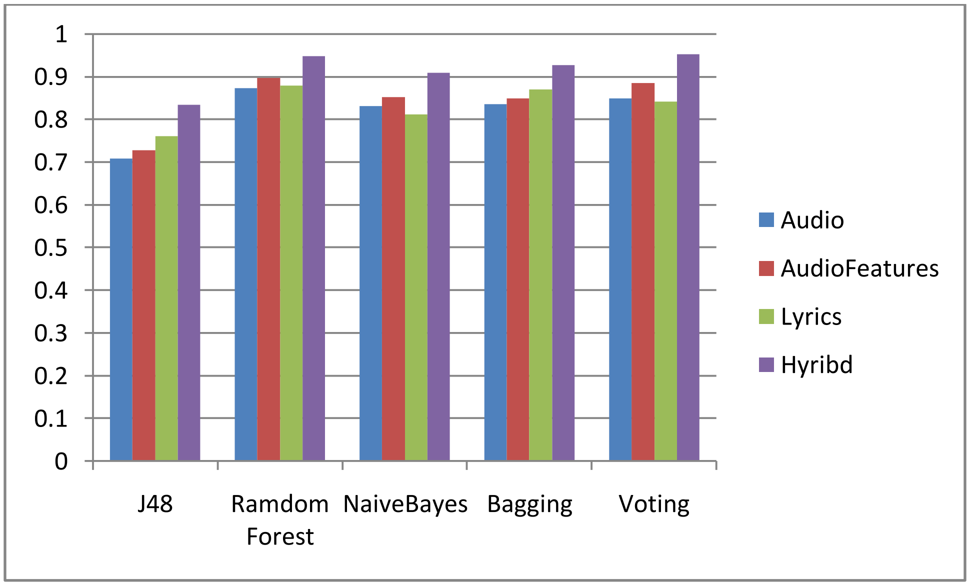

Figure 18 ROC for different classifiers

6.2.2.6 Precision (For Different Classes, Classifier J48)

6.Precision Class Wise J48				
	Audio	**Audio Feature**	**Lyrics**	**Hybrid**
Calm	0.789	0.941	0.611	0.912
Excited	0.5	0.591	0.7	0.755
Happy	0.5	0.421	0.527	0.653
Sad	0.381	0.455	0.684	0.667

Table 6 Precision for Different Classes, Classifier J48 (Max value "1")

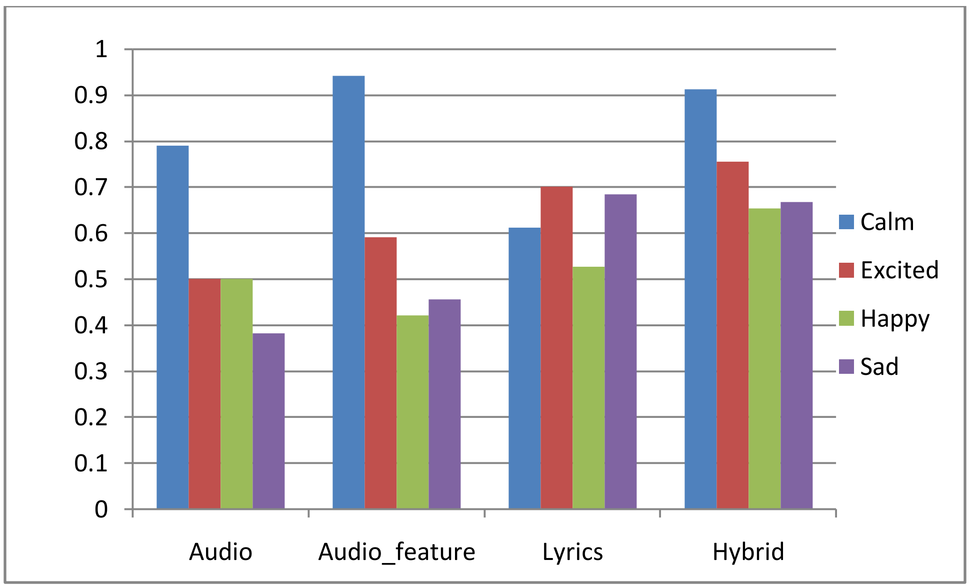

Figure 19 Precision for Different Classes, Classifier J48

6.2.2.7 ROC (For Different Classes, Classifier J48)

7.ROC Class Wise J48				
	Audio	**Audio Feature**	**Lyrics**	**Hybrid**
Calm	0.935	0.924	0.759	0.932
Excited	0.728	0.742	0.785	0.836
Happy	0.617	0.579	0.669	0.754
Sad	0.584	0.716	0.817	0.819

Table 7 ROC for Different Classes, Classifier J48 (Max value “1”)

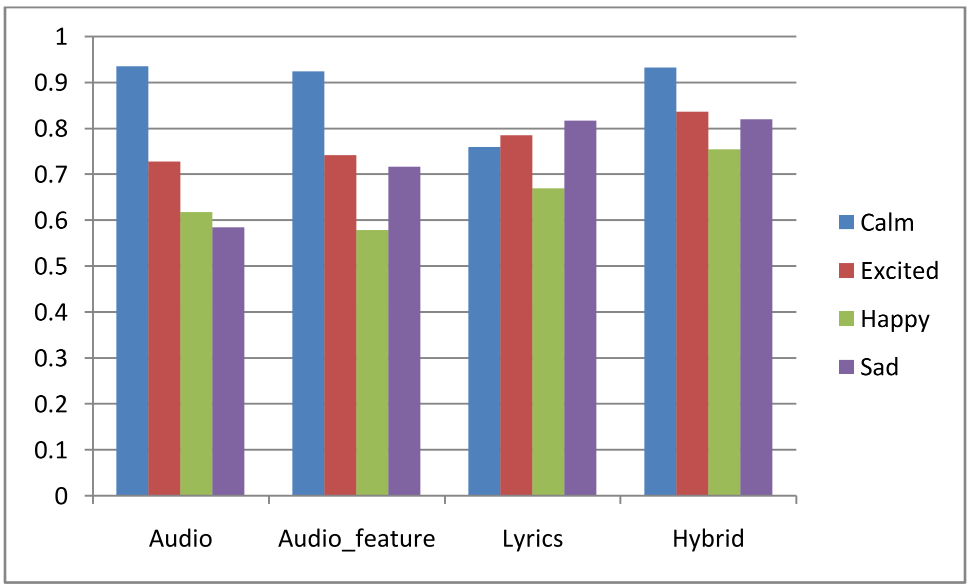

Figure 20 ROC for Different Classes, Classifier J48

6.2.2.8 Recall (For Different Classes, Classifier J48)

8.Recall Class Wise J48				
	Audio	**Audio Feature**	**Lyrics**	**Hybrid**
Calm	0.833	0.889	0.579	0.912
Excited	0.529	0.684	0.614	0.702
Happy	0.36	0.32	0.5	0.582
Sad	0.444	0.556	0.779	0.794

Table 8 Recall for Different Classes, Classifier J48 (Max value "1")

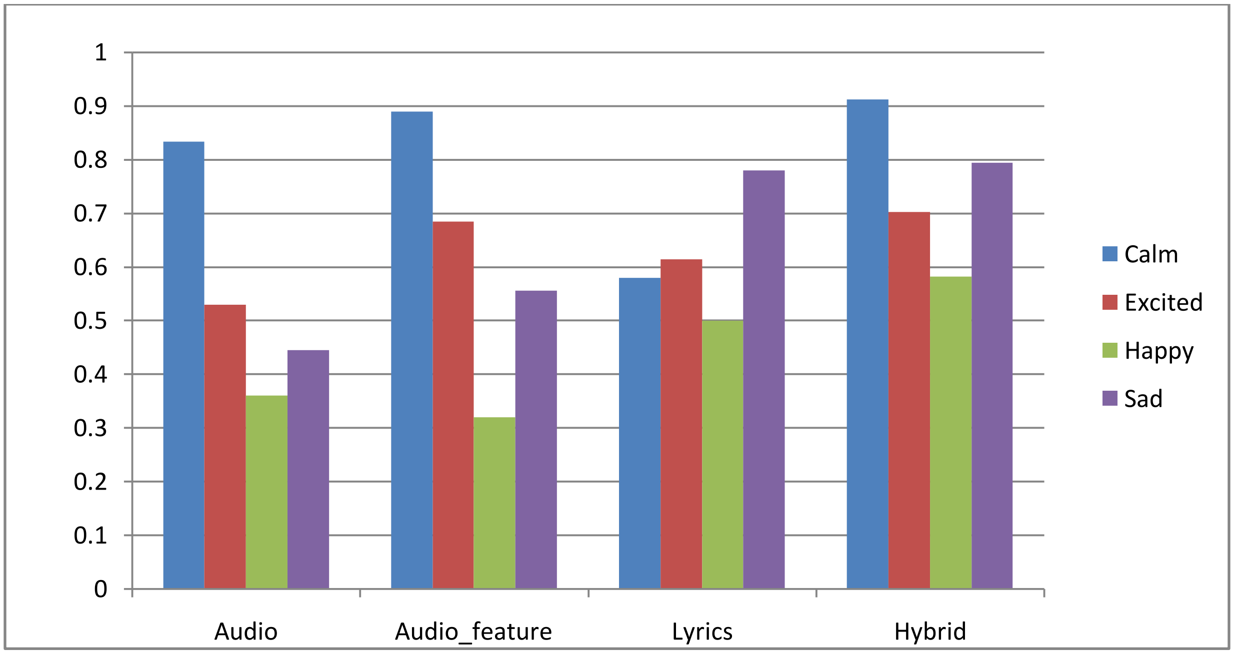

Figure 21 Recall for Different Classes, Classifier J48

6.2.3 Interpretation of Results:

If the data presented in previous section is analyzed carefully some conclusion that can be drawn are as follows:

1. The accuracy in case of hybrid model is significantly higher than other models where only audio features or only lyrics were used for classification purpose (Figure 44).
2. Information gain algorithm results in improved performance of the classifier (Figure 17).
3. As far as class wise results are concerned, accuracy in case of class “calm” is very high (more than 90%).This can be seen in Figure 20.
4. Generally the accuracy of class “sad” is low (nearly 50%).
5. In Hybrid model there is significant improvement in the accuracy of class “sad” (Figure 20).
6. The improved performance in hybrid model is mainly because of improved accuracy of class “sad”.
7. Most of the instances that are incorrectly classified are classified in adjacent class like “excited” being classified as “sad”.

It can be safely concluded that a hybrid model results in better classification of songs based on mood. There is notable improvement in the accuracy of class “sad”.

Chapter 7
Applications and Future Work

7.1 Applications

The quantity of music available online or offline is increasing day by day. So if one looks from the point of view of user choosing and selecting music to listen is a difficult task. Also now a days music is not something that is just listened in free time. It is an integral part of people life. Music now a day is also a gifting item in standalone form and in combination with other things.

For better understanding a situation can be considered. At various channels of FM songs are dedicated. These dedicated songs generally reflect a particular mood or a particular moment when that mood was expressed.

Now taking a leaf out of above example user can select songs or music related to a particular mood and can gift it in the form of "digital greeting card". In a world where quantity of music is increasing at a rapid rate, this mood selection application combined with a search engine can be of great significance.

Also, this system can be advance to include natural sounds and videos. Live speech mood detection can also be one possible of application. Following are some possible applications:

7.1.1 MIR (Music Information retrieval)

MIR is of interdisciplinary nature that deals with retrieval of useful data or features from music. It creates large scale collection of music materials which varies in audio, symbolic and metadata forms, which is secure and accessible. It uses knowledge from areas as signal processing, machine learning, information and music theory. So our system can help in auto categorization of music.

7.1.2 Intelligent Automatic Music Composition

Composing music has been a manual task these days. As technology is expanding, various software and devices are available in the market which helps in composing the songs automatically on basis of different genre. Such composition system needs information and data about the music along with the pattern recognition and signal processing. This system will help in building such composition system.

7.1.3 Music Therapy System

Music therapy is a method of curing people who are mentally disable or ill. In this process of curing experts examines the patient and suggests his/her a list of songs or music piece on daily basis. Songs are selected on the basis of mood parameter for different patients. So, generating such list will require the mood parameter which will be provided by our mood identification system. Lastly, the information provided by this system would pair well with data from sites such as Last.FM, where the listening patterns of a user are recorded and analyzed. These patterns permit ones to know the mood effects and evolution of their music and listening habits. This could likewise be applied to artists and bands.

7.1.4 Suggestions Regarding Music

The system can be used to give suggestion regarding music on websites by using past music download data as an input. This feature has been used successfully on some e-commerce websites and also for news websites. This can be extended to music websites.

7.2 Future Work

We successfully mapped Bollywood Indian songs with their mood using audio features. Then lyrics were considered to improve to efficiency of the classifier and to distinguish between few mood models. In this way a hybrid model of mood classification was created and tested. Results seem satisfactory when various performance matrices are considered. In crude terms the accuracy was more than 70%. There has not been much work on classification of Bollywood music as reflected in the literature survey. We think that there is considerable scope of improvement in several areas which may lead to increased efficiency and credibility of the classification system. Some starting points regarding future work are discussed below and many more can up as our understanding for Bollywood music and mood classification improves.

Some basic points regarding future work are as follows:

- As it is pointed out somewhere else in the report that a modified MIREX model was used for the classification purpose. One class which may be referred as “anger” class was dropped from the model. In our experiment and some other experiments related to Bollywood music mood classification it has been seen that number of songs in this category has been very low. Also during listening test people hardly classified songs under this category even after when they were given an option. Also, in Indian cinema anger is not represented through music. So first and foremost the mood models need to be evaluated in the context of Indian music and a mood model should be proposed that is simple and also suits our requirements. On things should be noted that a model with too many mood categories may look good from researcher point of view but it is certainly not good from user's point of view. Because more choices may lead to confusion. So a balanced model which is neither too specialized nor too generalized should be developed for Indian Bollywood music.

- As it has been pointed out that during listening test due to some constraints, all the participants were from the same age group. A more diverse panel should be chosen for listening test which preferably have members from different age groups. Some specialist can also be added to the group if possible.
- Only a 30s second clip was used to identify the mood of song but as it has been pointed out that mood of some songs change several time during the course of the song. More discussion is required to come at conclusion that what should be done with such songs.
- What can be possible features set for the classifier? To find this out more work is required with respect to features selection algorithms.
- There is always a scope for improvement when lyrics based classification is considered. We have ignored n-gram tokens for classification purpose. In some researches it has been pointed out that n-gram token lead to a more efficient classification system. So there is a scope for using n-gram tokens for classification purposes.
- Stemming was a problem as far as Hindi song lyrics are considered. Different algorithms should be tested to check which one gives the best results.
- As far as hybrid system is considered, we used serial hybrid system. A late fusion model is said to be more efficient. This system can be extended to have late fusion hybrid music.
- Regional songs often find way in the playlist of Indian listeners. Some examples are Punjabi and Bhojpuri songs. So if possible classifier model should be extended to classify this kind of regional music.

Bibliography

1. *Experimental studies of the elements of expression.* **Henver, Kate.** 1936, American Journal of Psychology.

2. *Analysis of the Data Quality of Audio Descriptions of Environmental Sounds.* **Mitrovic, Dalibor, Zeppelzauer, Matthias and Eidenberger, Horst.** 2007, Journal of Digital Information Management.

3. *"Fundamentals ofspeech Recognition".* **Rabiner, L., Juang, B.** 1993.

4. *"classification in music research".* **Wang C., Ligges U., Morchen F., Mullensiefen D.** 2007.

5. *"A Survey of Audio based Music Classification and Annotation"* . **Zhouyu Fu, Guojun Lu, Kai Ming Ting, Dengsheng Zhang.** 2011.

6. *Äutomatic genre classification of music content, A survey".* **Scaringella N., Zoia G., Mlynek D.** 2006, The journal of the Acoustical society of Japan.

7. *"Detecting emotion in music".* **M., Ogihara and Li T.** 2003.

8. **Lu, L, Liu, D and Zhang, H.** Automatic mood detection and tracking of music audio signals. IEEE Transactions on Audio, Speech, and Language Processing. 2006. pp. 5-18.

9. *A circumplex model of affect.* **Russell, J.A.** 1980, Journal of Personality and Social Psychology, pp. 1161-1178.

10. *"The Bio-psychology of Mood and Arousal".* **Thayer, R.E.** 1989.

11. *psychological foundations of musical behaviour.* **Radocy E., boyle J. D.** 1988.

12. **Ekman, P.** *Emotion in the Human Face.* s.l. : Cambridge University Press , 1982.

13. *Continuous response to music using a two dimensional emotion space.* **Schubert, E.** 1996. Proceedings of the 4th International Conference of Music Perception and Cognition.

14. **Tyler, P.** Developing A Two-Dimensional Continuous Response Space for Emotions Perceived in Music. 1996.

15. **Meyers, O.C.** *A mood based music classification and exploration system.* Massachusetts Institute of Technology. Cambridge: MA, USA : s.n., 2003. Master's thesis.

16. **Theyer, R.E.** *The Bio-psychology of Mood and Arousal.* New York : Oxford University Press, 1989.

17. *"pattern Classification".* **Duda, R.O. and Hart P. E.** 2000.

18. *Play it again with feeling: computer feedback in musical communication of emotions.* **Juslin, P.N., et al.** 12(1), 2006, Journal of Experimental Psychology: Applied, pp. 79-95.

19. *"The art of recording, understanding and craftingthe mix".* **Moylan, William.**

20. *Expression, perception, and induction of musical emotions: A review and a questionnaire study of everyday listening.* **Patrik, N. Juslin and Petri, Laukka.** 2004, Journal of New Music Research, pp. 217-238.

21. *“Spectral and Temporal Periodicity Representations of Rhythm for the Automatic Classification of Music Audio Signal".* **Peeters, G.** 2011.

22. *Computational methods of feature selection.* **Liu H., Motoda H.** 2008.

23. *Multimodal Music Mood Classification using Audio and Lyrics.* **Jens Grivolla, Cyril Laurier , Perfecto Herrera.** 2008.

24. *"Lyrics text mining in music mood classification".* **Xiao Hu, J. Stephen Downien and Andreas F. Ehmann.** 2009.

25. *"Mutiple lyrics alignment: automatic retrieval of song lyrics".* **Knees P., Schedl, M. , Widmer, G.** 2005. 6th international Symposium on music information retrieval of song lyrics.

26. *abc.* **zcs.** 2009, cvb.

27. *Towards multi-modal music emotion classification.* **Yang, Y.-H., et al.** 2008. In Proceedings of Pacific Rim Conference on Multimedia (PCM’08).

28. *Music emotion classification:A fuzzy approach.* **Yang, Y., Liu, C. and Chen, H.** Santa Barbara: CA, USA : ACM Press, 2007. In Proceedings of the 14th annual ACM international conference on Multimedia. pp. 81-84.

29. *Disambiguating music emotion using software agents.* **Yang, D. and Lee, W.** 2004. In Proceedings of the 5th International Conference on Music Information Retrieval (ISMIR'04).

30. **Voong, Michael and Beale, Russell.** Music organization using colour synaethesia. *CHI '07 extended abstracts on Human factors in computing factors.* New York:NY, USA : ACM Press, 2007, pp. 1869-1874.

31. *Beyond Timbral Statistics: Improving Music Classification Using Percussive Patterns and Bass Lines.* **Tsunoo, E., et al.** 2011, Audio, Speech, and Language Processing, pp. 1003-1014.

32. *Multi-label classification of music into emotions.* **Trohidis, K., et al.** 2008. In Proceedings of the 9th International Conference on Music Information Retrieval (ISMIR’08).

33. **Stone, P. J.** *General Inquirer: a Computer Approach to Content Analysis.* Cambridge : M.I.T. Press.

34. **Sloboda, J.A. and Juslin, P.N.** *Music and Emotion: Theory and Research.* New York : Oxford University Press, 2001.

35. *Tempo and beat analysis of acoustic musical signals.* **Scheirer, E.** 1998, J.Acoust. Soc. Amer., vol. 103, no. 1, pp. 588601.

36. **Pohle, T., Pampalk, E. and Widmer, G.** Evaluation of frequently used audio features for classification of music into perceptual categories. In Proceedings of the 4th International Workshop on Content-Based Multimedia Indexing. 2005.

37. *jAudio: A feature extraction library.* **McEnnis, D., et al.** 2005. Proceedings of the International Conference on Music Information Retrieval.

38. —. **McAnnis, D., et al.** 2005. Proceedings of the International Conference on Music Information Retrieval.

39. **Mayor, L.B.** *Emotion and Meaning in Music.* Chicago : University of Chicago Press, 1956.

40. *Combination of audio and lyrics features for genre classification in digital audio collections.* **Mayer, R., Neumayer, R. and Rauber, A.** 2008. In Proceedings of the 16th ACM International Conference on Multimedia.

41. *Multimodel music mood classification using audio and lyrics.* **Laurier, C., Grivolla, J. and Herrera, P.** 2008. In Proceedings of the International Conference on Machine Learning and Applications.

42. **JungHyun, Kim, et al.** Music Mood Classiication Model Based on Arousal-Valence Values. 2011.

43. *Perceptual and cognitive applications in music information.* **Huron, D.** 2000, Proc. Int. Symp. Music Information Retrieval (ISMIR).

44. *Lyric-based song emotion detection with affective lexicon and fuzzy clustering method.* **Hu, Y., Chen, X. and Yang, D.** 2009. In Proceedings of the 10th International Conference on Music Information Retrieval (ISMIR'09).

45. *Music and mood: where theory and reality meet.* **Hu, X.** 2010. In Proceedings of iConference 2010.

46. *The 2007 MIREX audio mood classification task: Lessons lerned.* **Hu, X, et al.** Philadelphia, PA : s.n., 2008. Proc. of the Intl. Conf. on Music Information Retrieval.

47. *Language feature mining for music emotion classification via supervised learning from lyrics.* **He, H., et al.** 2008. Proceedings of Advances in the 3rd International Symposium on Computation and Intelligence (ISICA 2008).

48. **Haring, B.** Beyond the Charts: Mp3 nd the Digital Music Revolution. s.l. : JM Northern Media LLC, 2000.

49. **Downie, J.S.** The music information retrieval evaluation exchange (2005-2007): A window into music information retrieval research. s.l. : Acoustical Science and Technology, 2008. Vol. 29, pp. 247-255. 4.

50. **Das, Atin and Das, Pritha.** *Classification of Different Indian Songs Based on Fractal Analysis.* s.l. : Complex Systems Publications, 2005.

51. *"Machine learning approaches for mood classification of songs toward music search engine.* **Dang, Trung-Thanh and kiyoaki-shirai.** 2009. Ïnternational Confrence on Knowledge and System Engineering.

52. *Tutorial on music recommendation.* **Celma, O. and Lamere, P.** Vienna, Austria : s.n., 2007. Eighth International Conference on Music Information Retrieval.

53. *Music and Mood.* **Carolyn, J. Murrock.** 2005, Psychology of Moods, pp. 144-155.

54. **Capurso, A., et al.** *Music and Your Emotions.* s.l. : Liveright Publishing Corpoation, 1952.

55. **Capurso, A., et al.** *Music and Your Emotions.* s.l. : Liveright Publishing Corporation, 1952.

56. **Bradley, M. M. and Lang, P. J.** *Affective Norms for English Words (ANEW): Stimuli, Instruction Manual and Affective Ratings.* University of Florida. 1999. Tecnical Report C-1.

57. *Music mood and theme classification - a hybrid approach.* **Bischoff, K., et al.** 2009. In Proceedings of the 10th International Conference on Music Information Retrieval (ISMIR'09).

58. *How do you feel about "Dancing Queen"? Deriving mood and theme annotations from user tags.* **Bischoff, K., et al.** 2009. In Proceedings of Joint Conference on Digital Libraries (JCDL'09).

59. *Emomusic-Classifying music according to emotion.* **Baum, Doris.** Kosice : s.n., 2006. Proceedings of the 7th Workshop on Data Analysis.

60. *Recognition of emotion in Japanese, Western, and Hindustani music by Japanese listeners.* **Balkwill, L.-L., Thompson, W. F. and Matsunag, R.** 2004, Japanese Psychological Research.

61. **Anderson, Chris.** The Long Tail . 2004.