Music Mood Classification Using Audio Features

S.P.M.Harinda Lakmal

2018

Music Mood Classification Using Audio Features

A dissertation submitted for the Degree of Master of Computer Science

S.P.M.Harinda Lakmal
University of Colombo School of Computing
2018

Declaration

The thesis is my original work and has not been submitted previously for a degree at this or any other university/institute.

To the best of my knowledge it does not contain any material published or written by another person, except as acknowledged in the text.

Student Name : S.P.M.Harinda Lakmal

Registration Number : 2014/MCS/044

Index Number : 14440441

____

Signature: Date: 12.07.2018

This is to certify that this thesis is based on the work of

Mr. /~~Ms~~. S.P.M.Harinda Lakmal

under my supervision. The thesis has been prepared according to the format stipulated and is of acceptable standard.

Certified by :

Supervisor Name : Dr. D.A.S. Athukorala

____

Signature: Date: 12.07.2018

Abstract

Music emotions and moods are subjective and vary between people. Moods are categorized in to different groups where mainly divided into categorical and dimensional models. No widely accepted framework has been emerged due to ambiguity among the different moods. This work involved predicting four basic and major moods: anger, happy, neutral and sad which categorized under categorical model. Preparing ground truth dataset for analysis was part of the work due to unavailability of such dataset which are confirmed by expert judgment. Training dataset consists 800 music samples which includes 200 samples for each mood where testing dataset consists 100 samples by including 25 samples for each mood.

930 audio features are extracted through MIRToolbox using MATLAB and jAudio for feature selection. WEKA has been used as main tool for pre-processing, feature selection, experiment and classification tasks in addition to the Python sklearn. Supervised discretize, unsupervised normalize and unsupervised replace missing value filters have been used to pre-processes the original dataset. Three distinct original, discretized and normalized datasets are used to feature selection with WEKA attribute and wrapper subset evaluators. 59 datasets obtained through feature selection is prepared for classification with eight classification algorithms namely Naïve Bayes, LibSVM, SMO, IBK, AdaBoostM1, Bagging, 48 and Random Forest. WEKA experimenter is used to upload, configure algorithms, perform classification and compare the output results.

According to the analysis, support vector machine algorithm LibSVM exhibited highest prediction accuracy while Bagging exhibits the highest average performance and the minimum standard deviation among the accuracy. LibSVM reported highest accuracy deviation as well as the lowest prediction results. Among three distinct datasets, only discretized datasets provided the better accuracy. During comparison, IBK, AdaBoostM1 and IBK not exhibit the prediction accuracy over 75%. Algorithms and discretized datasets predicated over 75% has been considered to perform optimization with different parameter setting obtain the better output results. Based on the results, SMO results over 77% accuracy under *NormalizedPolyKernel* and random seed seven.

Acknowledgment

I take this opportunity to forward my gratitude to the UCSC management and the senior lecturers to provide me the opportunity to follow the master's degree program in computer science in this reputed institute even my basic degree with management information systems.

Also, I grateful to all our well qualified lecturers who provide the knowledge, experience and guidance towards to achieve this well recognition qualification to our entire life.

I grateful to Dr. Ajantha Athukorala, who is my project supervisor and my machine learning lecturer, the support, courage and guidance provided me to achieve this goal. He is flexible and matured person who supported me to even select this project title and encouraged me in difficult situations.

The courage and support given by my loving wife Dr.Dilini Priyadarshani and three little kids, Sethika Disas (7), Nethika Vihas (4) and Jayeli Saveathma (11/2) during this period.

My parents always arranged comfortable environment to me to proceed with studies by taking my responsibilities towards the family.

My batch mates, especially Jagath Ariyarathne and Sampath Dilhan, the assistance provided me during this period.

Table of Contents

List of Figures

List of Tables

Chapter 1: Introduction

We listen to music when we wake up till go to the sleep and this has become essential part of our lives. For many, music is a life time partner which brings joy to us, motivate us and be with us through our worries and difficult times. No one can find any culture through the past human history which not had music. Music can consider as special feature in every human society which has more beyond the entertainment. In addition to the entertainment, music reaches us on intellectual, social and emotional levels, but many describe this as spiritual [8]. Fundamentally, music is a combination of sound where the use of melodic, harmonic, and rhythmic devices in music can induce a psychological state which is very personal to everyone. "Music is tone organized toward beauty" (William Arms Fisher), "Music is sonorous air." (Ferruccio Busoni - Italian composer) are some definitions given for "What is music?" [8]

1.1 Motivations of the Project

Usually we select music which are best suite with the present mental situation or conditions and depend on the personality of each person. Increase and maintaining human productivity of different tasks in stressful environment is also challenging. Music is considered as energetic mood controller and helps in improving the mood and state of the person which in turn increase the productivity. Continuous music play requires creating and managing personalized song playlist which is a time-consuming task.

Due to rapid growth and use of internet and other information distribution channels we are experiencing the expansion of personal and public music collections. It is expected that this massive growth will never go down and increase year by year which required new tools and techniques with more advanced and flexible than past. This enables individuals more capable of searching or browsing large music sets based on their specific needs.

Generally, music is arranged based on Artist, Album, Band, Genre, Year, Bitrates etc. [5] [6] [7]. When required, each time we need to purposefully select the songs which match our interest and this may become very boring and unpleasant experience. Also, when selecting songs from relatively very large collection this become more tedious and difficult. Based on the above parameters it is difficult to classify songs in to different moods naturally because those are unable to use adequately to extract the music features. Therefore, majority of the music applications are still unable to provide mindful playlist to the end user. It would be very helpful if the music player itself selects a song according to the current mood of the user.

Therefore, new research fields such as Music Information Retrieval (MIR) are gaining more awareness due to problems we are facing when extracting useful music information. Thus main goal of MIR is to find different ways to extract information from data in songs and make searching, classification and indexing of songs much faster and precise. To facilitate that, MIR involves search categories based on music metadata such as title, artist, album and criteria like mood, taste and emotion, that are subjective and may vary between individuals and cultures [5] [6] [7].

Most music mood classification systems are based on lyrics classifications while performance and accuracy are still behind the music mood classification systems expressed by means of musical features. Therefore, in order to improve the overall performance of the music mood detection, integrated multi model system can be used (audio + text/lyrics).

The objective of this project is to investigate how audio files with different formats to be processed and the possible tools and techniques available to extract the audio features effectively to categorize the audio into different moods.

Then use of feature selection to select features that aid in classification by giving better classification results than the other features. Identify the good features for classification will leads to reduce the computational costs and allows maintaining the accuracy level.

Investigate classifier that must be computationally efficient with less complexity for the classification process which is subsequent process to the feature extraction. Classification used to accurately label the audio using the features selected. Then nature of the unknown audio is classified under a known class of audio moods.

Finally, this project effort focusses on two main areas, researching and software engineering which are challenging and motivating me to proceed.

Following benefits will open when ability to search songs based on emotion criteria:

1. **Individuals**: could use some face detection system or technique to capture the current mood or emotional state of the individual and pass it to automatically paly the songs that best suite with the recognized mood.
2. **Film industry**: could use these capabilities to find songs that match the planned scene, such as fear, anger, joy and other kind of emotions [6].
3. **Radio and televisions**: could help when finding the right music when client requests music with different feelings or mood settings [6].

4. **Call centers**: to maintain the satisfaction level when customers are waiting in line. Songs can be dynamically change if there are unexpected queue in call center or in any difficult situation to manage the customers [6].
5. **Researches**: open opportunities to heavily research on finding music based on criteria that are subjective to individuals with more accurately.

1.2 Objective of Project:

The objective of this project is to categorized songs based on subjective criteria such as mood which vary between individuals and investigate how audio files with different formats to be processed and the possible tools and techniques available to extract the audio features effectively to categorize the audio into different moods.

Then use of feature selection to select features that aid in classification by giving better classification results than the other features. Identify the good features for classification will leads to reduce the computational costs and allows to maintain the accuracy level.

Investigate classifier that must be computationally efficient with less complexity for the classification process which is subsequent process to the feature extraction. Classification used to accurately label the audio using the features selected. Then nature of the unknown audio is classified under a known class of audio moods.

1.3 Related Projects (Others who are proposing logically related work or previous projects which were based for your proposal):

1. The Mood Based Music System will be of great advantage to users looking for music based on their mood and emotional behavior. It will help reduce the searching time for music thereby reducing the unnecessary computational time and thereby increasing the overall accuracy and efficiency of the system [19].

2. Study on music mood classification using audio and lyrics information. Analysis is able to classify the lyrics significantly, but mood classification performance is still quite inferior to that of audio-based techniques. Therefore, integrating this in a multimodal system (audio+text) allows an improvement in the overall performance. They demonstrate that lyrics and audio information are complementary and can be combined to improve a classification system [20].

3. Due to enhancement in technology in the recent years lots of music is available as handy media for various devices. Thus there is an urgent need of analyzing music for storage, indexing and retrieval. In this paper we aimed at classifying music on the basis of their moods. We have identified three moods for this purpose: happy, angry and sad. These cognitive styles have few things in common. Identifying and extracting these features is a challenging problem. On the basis of our observations and literature review we have identified the eight features namely energy, entropy, zero-crossing rate, spectral rolloff, spectral centroid, spectral flux, RMS of signal and MFCC. After extracting features, we have used neural network-based training for classification. We have used Matlab 7.0.1 neural net tool for this purpose. We have populated a database of 150 songs consisting of 50 songs of each category. In this population 90 songs are used for training set and 60 for testing. The experimental results demonstrate the effectiveness of our classification system. We have obtained an overall accuracy of nearly 75%. The complete system is developed in MATLAB 7.0.1 including the GUI of the system [21].

1.4 Scope of Project:

- The song is given as input which can be an mp3, wav file format. It is going to be processed by the system and given as an input to the feature extractor.
- Using audio feature extraction tools/techniques the selective audio features are extracted. A feature vector containing all these features is extracted for each of the music file. The feature vectors extracted from the attributes of each music file will be stored in a flat file format which can be understood by the data mining tools.
- Among high-dimensional feature sets, apply feature selection algorithms to identify the most compact but discriminative feature subset which has an impact to the classification accuracy and computational effort.
- Apply classification algorithms and any applicable technique to classify songs into predefined set of classes or labels. Refer Figure I.01for high level architecture of the scope of the project

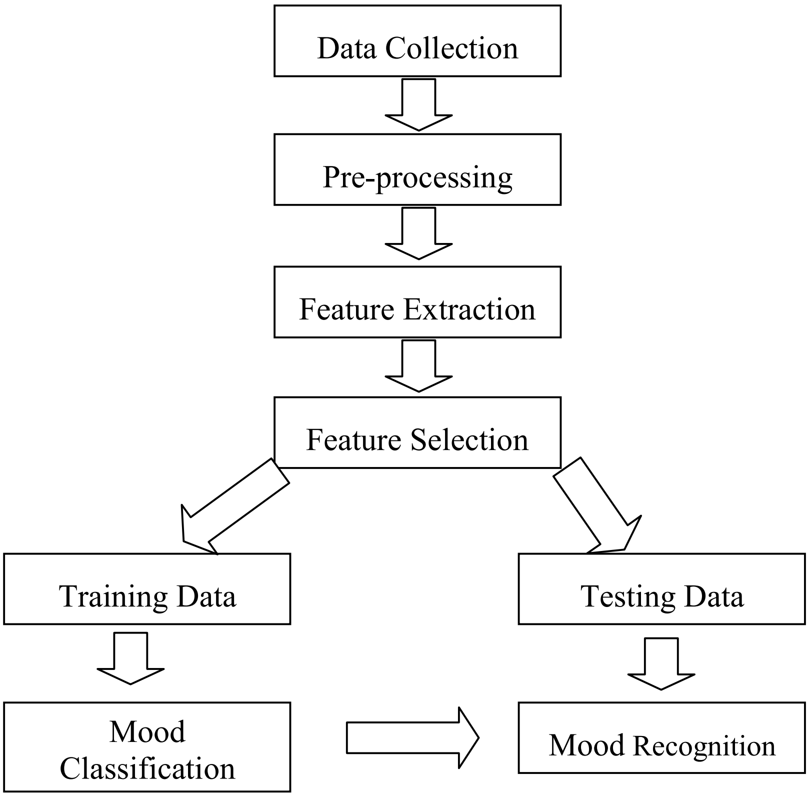


Figure I.01: High level architecture of the scope of the project

Structure of the Dissertation:

a. Title Page
b. Declaration
c. Abstract
d. Acknowledgement
e. Table of Contents
f. List of Figures
g. List of Tables
h. Introduction
i. Literature Review
j. Analysis and Design
k. Evaluation and Result
l. Conclusion
m. References
n. Appendix

Chapter 2: Literature Review

2.1 Definitions of emotion:

An emotion is a feeling comprising physiological and behavioral (and possibly cognitive) reactions to internal and external events [22].

An emotion is a complex psychological event that involves a mixture of reactions: (1) a physiological response (usually arousal), (2) an expressive reaction (distinctive facial expression, body posture, or vocalization), and (3) some kind of subjective experience (internal thoughts and feelings [23].

The process of emotion . . . is initiated when one's attention is captured by some discrepancy or change. When this happens, one's state is different, physiologically and psychologically, from what it was before. This might be called a "state of preparedness" for an emotion . . . The process almost always begins before the name [of the emotion is known] and almost always continues after it [24].

"The concept of emotion . . . refer[s] to (1) emotional syndromes, (2) emotional states, and (3) emotional reactions. An emotional syndrome is what we mean when we speak of anger, grief, fear, love and so on in the abstract. . . . For example, the syndrome of anger both describes and prescribes what a person may (or should) do when angry. An emotional state is a relatively short term, reversible (episodic) disposition to respond in a manner representative of the corresponding emotional syndrome. . . . Finally, and emotional reaction is the actual (and highly variable) set of responses manifested by an individual when in an emotional state: . . . facial expressions, physiological changes, overt behavior and subjective experience [25].

Emotion is a complex set of interactions among subjective and objective factors, mediated by neural/hormonal systems, which can (a) give rise to affective experiences such as feelings of arousal, pleasure/displeasure; (b) generate cognitive processes such as perceptually relevant effects, appraisals, labeling processes; (c) activate widespread physiological adjustments to the arousing conditions; and (d) lead to behavior that is often, but not always, expressive, goal-oriented, and adaptive [26].

2.2 Definitions of Mood

A mood is a relatively long lasting emotional state. Moods differ from simple emotions in that they are less specific, less intense, and less likely to be triggered by a stimulus or event [Thayer, 1989]. [6]

2.3 Mood vs. Emotion

Music can have *(1)Mood,* where the state and/or quality of a feeling associated to the track (e.g. happy, sad, aggressive) and *(2)Theme,* which refers to context or situations that best fit when listening to the track (e.g. party time, Christmas, at the beach) [16]. Emotion is "temporary and evanescent" while mood is "relatively permanent and stable" according to the most influential first work formally analyzing music and mood using psychological methodologies by *Meyer's Emotion and Meaning in Music* [7]. Also, emotion can be *expressed* (feeling that are "intrinsic" to a given track) or *induced* (feelings that the listener associates with a given track) by music [16]. In music psychology, both *emotion* and *mood* has been used as affective effects of music, but emotion tend to be more popular. But researches in MIR and existing music repositories tend to use *mood* over *emotion* as metadata type to store and organized music. (e.g.: AllMusicGuide [1] and AMP[2]) [7]. Mood and emotions are subjective, varying from person to person and culture to culture. [6] [7] [10] [11]. Different persons have different perceptions of the same stimulus and often use some of these different words to describe similar experiences. Unfortunately, there is no standard, widely accepted, mood taxonomy. Therefore, understanding the existent models to represent emotions and moods and identifying better fits for this project can be seen as one important task.

2.4 Music Mood Categories

Several theoretical models have been proposed for identification of emotions and moods that can be grouped in to two major approaches: *categorical models* or *dimensional/parametric models* [6] [7] [16]. Categorical models consist of several discrete mood categories/clusters or states of emotion, such as anger, fear, happiness and joy (Hevner's adjective circle) [6] [7]. Dimensional models use several axes to map emotions to a plan [6] [7]. The most frequent approach is using two axes (e.g. arousal-valence (AV) or energy-stress), with some cases of a third dimension (Thayer's Model of Mood).

2.4.1 Six Basic Emotions – Categorical Model (The theory of American psychologists Paul Ekman and Wallace V. Friesen) [1]

Ekman and Friesen identified six basic emotions based on studying the isolated culture of people from the Fori tribe in Papua New Guinea in 1972. Following six basic emotions were identified: Anger, Disgust, Fear, Happiness, Sadness, Surprise.

Figure L.01: Use of the six basic emotions in practice[1]

[1] https://managementmania.com/en/six-basic-emotions

2.4.2 Plutchik's Wheel of Emotions – Categorical Model

Plutchik's eight primary emotions are Joy, Trust, Fear, Surprise, Sadness, Anticipation, Anger, and Disgust. For example, rage is the stronger form of anger while annoyance is the weaker (Plutchik, 2001a).

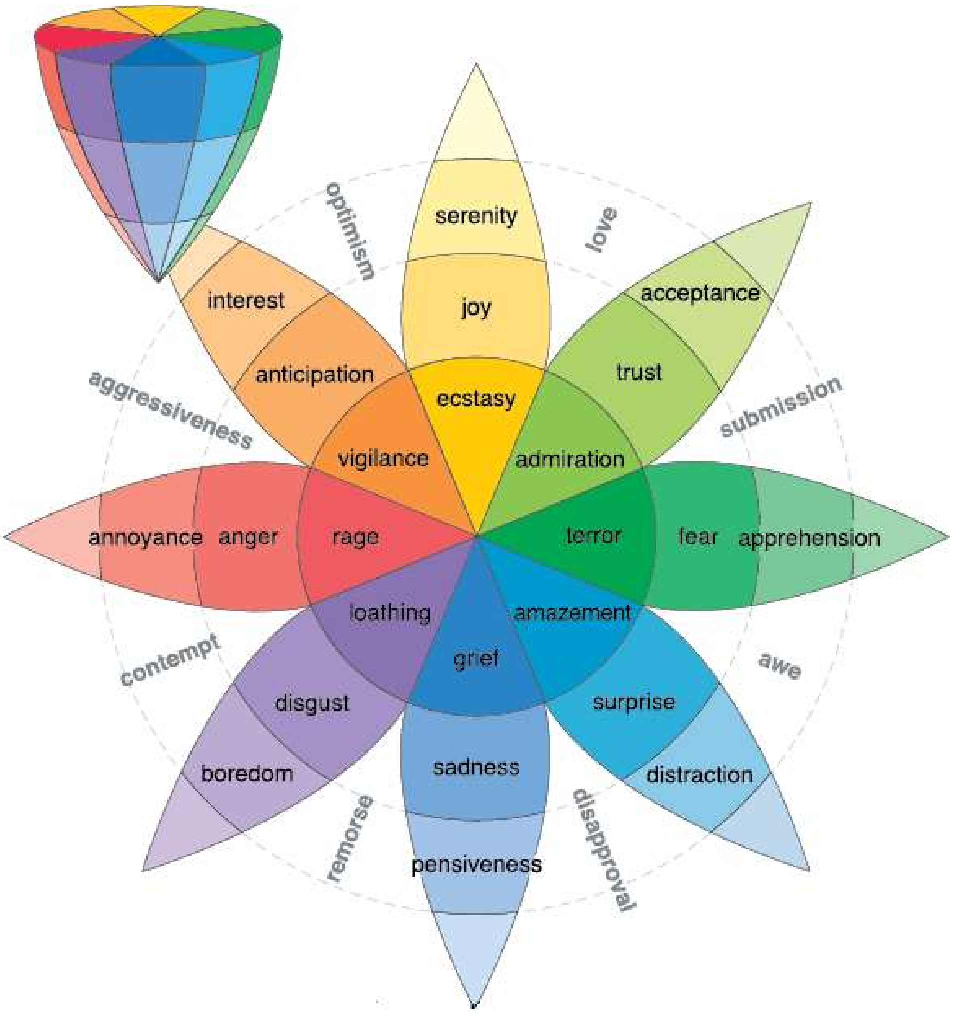


Figure L.02: Plutchik's Wheel of Emotions is analogous to a color wheel[1]

[1] http://www.6seconds.org/2017/04/27/plutchiks-model-of-emotions/

2.4.3 Hevner's Adjective Circle (Categorical Model)

*Kate Hevner*is best known by her research in music psychology, being one of the first to do research on the subject of music mood. She concluded that music and emotions are intimately connected, with music always carrying emotional meaning in it. As a result, she introduced an emotion (adjective) list; known as Hevner's Adjective Circle. Hevner's list is composed by 67 different adjectives such as serene, tranquil, and quiet which are organized in eight different groups/clusters in a circular way (Figure L.03). The adjectives within each cluster are close in meaning and the meanings of adjacent clusters would differ slightly. But the difference between clusters increases step by step until a cluster at the opposite position is reached [6] [7].

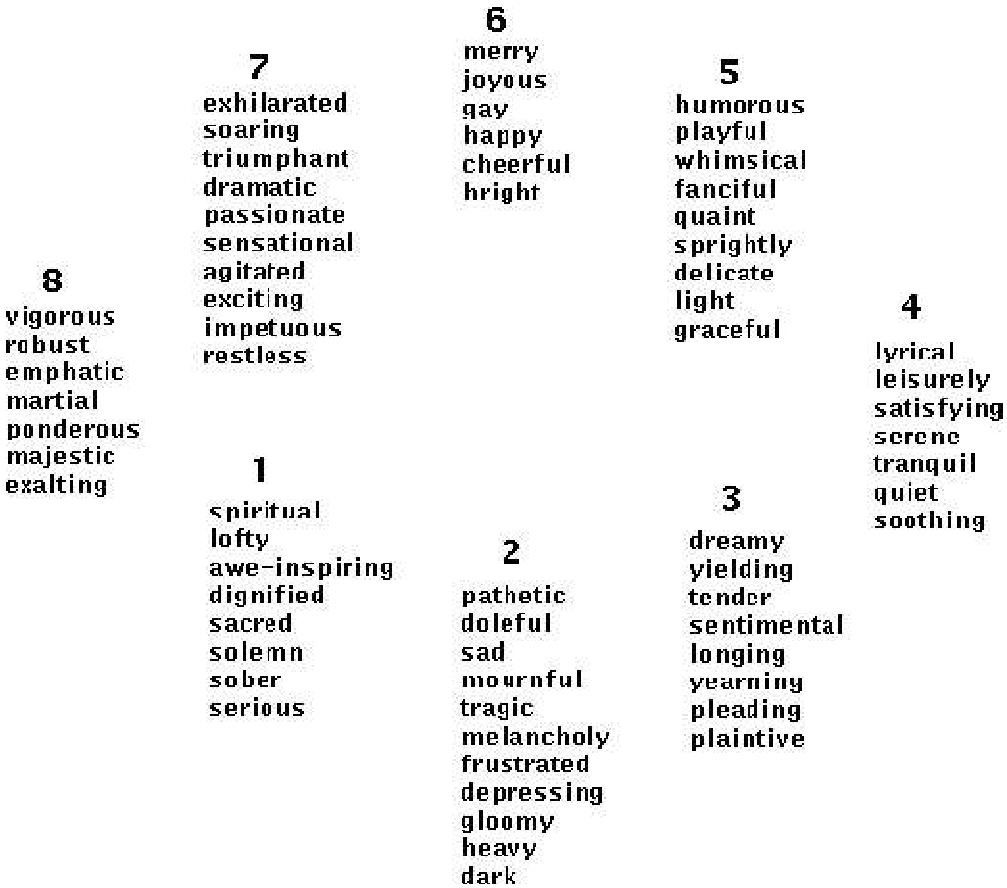


Figure L.03: Hevner's adjective circle

2.4.4 Music Information Research Evaluation eXchange (MIREX)

The IMIRSEL has derived a set of 5 mood clusters from the AMG mood repository (Hu & Downie 2007) and mood clusters effectively reduce the diverse mood space into a tangible set of categories. These 5 clusters are derived by performing clustering on a co-occurrence matrix of mood labels for popular music from the All Music Guide (AMG) [16].

Cluster #	**Moods**
Cluster_1	passionate, rousing, confident, boisterous, rowdy
Cluster_2	rollicking, cheerful, fun, sweet, amiable/good natured
Cluster_3	literate, poignant, wistful, bittersweet, autumnal, brooding
Cluster_4	humorous, silly, campy, quirky, whimsical, witty, wry
Cluster_5	aggressive, fiery, tense/anxious, intense, volatile, visceral

Table 01: MIREX Mood Clusters [16]

2.4.5 Thayer's Model of Moods

This is dimensional model where emotions are positioned in a continuous multidimensional space. The most influential dimensions contain such as Valence (happy-unhappy), Arousal (active-inactive) and Dominance (dominant-submissive). Two-dimensional mood model [Thayer, 1989], offering a simple but effective way to represent mood. Thayer adopted a distinct approach than Hevner's adjectives list, stating that mood depends on two factors: Stress (happiness/anxiety) and Energy (calm/energy) combined in a two-dimensional axis forming four different quadrants [6] [7]: (Figure L.04)

1. *Contentment*: representing calm and happy music
2. *Depression*: referring to calm and anxious music
3. *Exuberance*: referring to happy and energetic music
4. *Anxiety*: representing frantic and energetic music

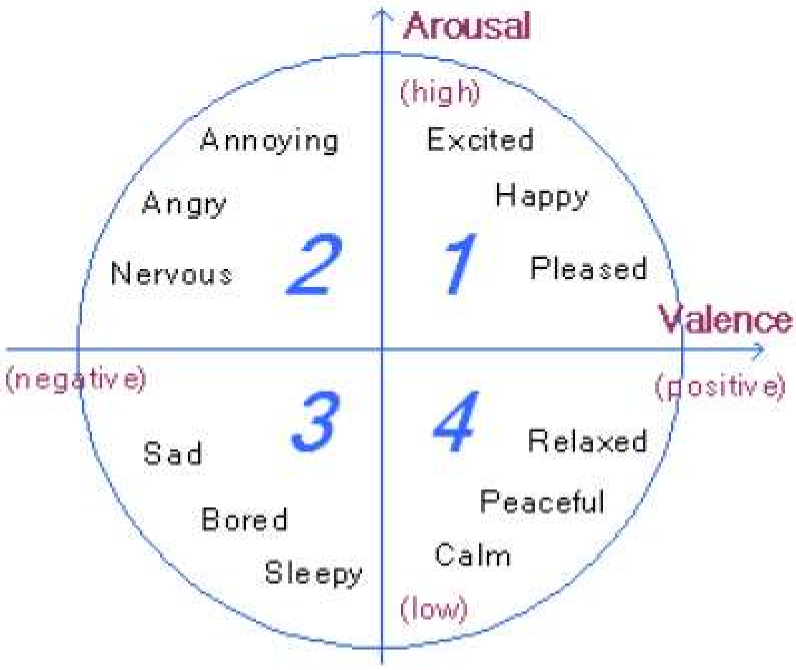


Figure L.04: Thayer’s arousal-valence emotion plane

Weak aspects of this model are its low granularity, not having a high number of well-defined different emotions. On the other hand, its simplicity is an advantage, making it less ambiguous.

2.4.6 Russell’s model of Moods

Russell’s model has combination of valence and arousal dimensions and places 28 emotions denoting adjectives on a circle (Figure L.05) [7]. This is most noted dimensional model [16] where emotions exist on a plane along independent axes. This model put one axis to represent the *Arousal* level, which denotes the intensity/loudness in form of high (active) and low values (inactive) and the other axis to represent *Valence*, which is an appraisal of polarity [16].

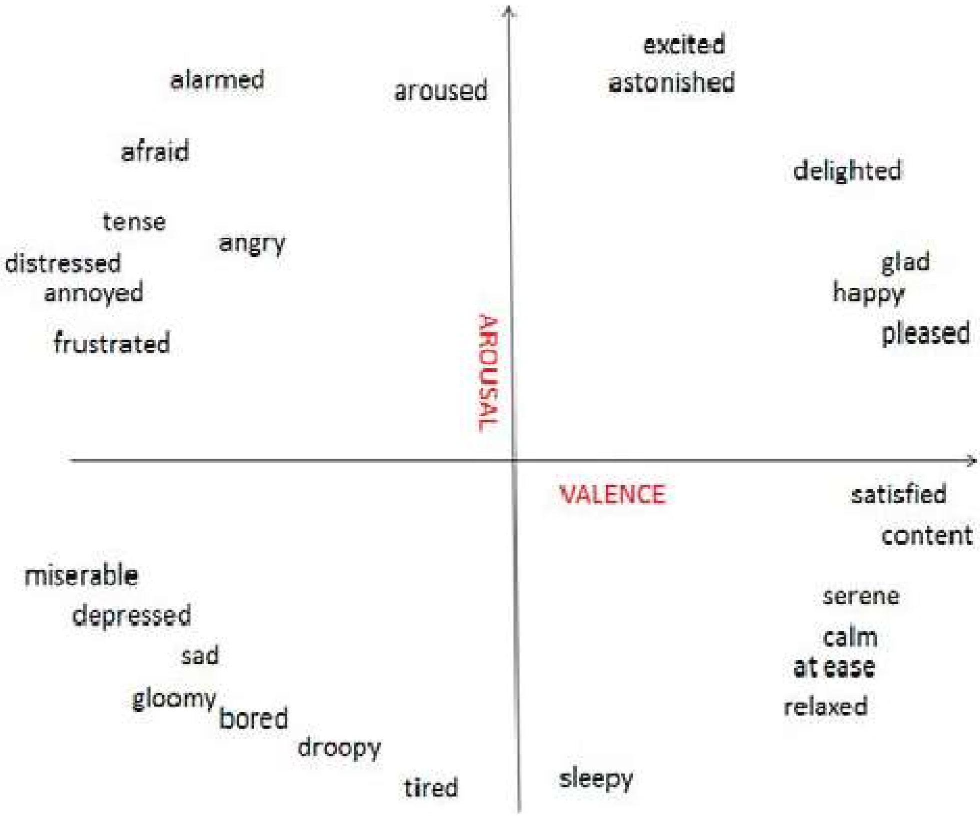


Figure L.05: Russell's model with two dimensions: arousal and valence

2.4.7 The Geneva Emotion Music Scale (GEMS) model – 2008[1]

The GEMS (Figure L.06) is the latest and first model and rating instrument that specifically designed to capture the richness of musically evoked emotions. This model consist 9 dimensions and 45 emotion labels which can be used for music research studies. Also, there is a shorter scale of GEMS with 9 dimensions and 25 music labels.

[1] http://www.zentnerlab.com/content/musically-evoked-emotions

Zentner, M., Grandjean, D., & Scherer, K. (2008). Emotions evoked by the sound of music: Characterization, classification, and measurement, *Emotion*

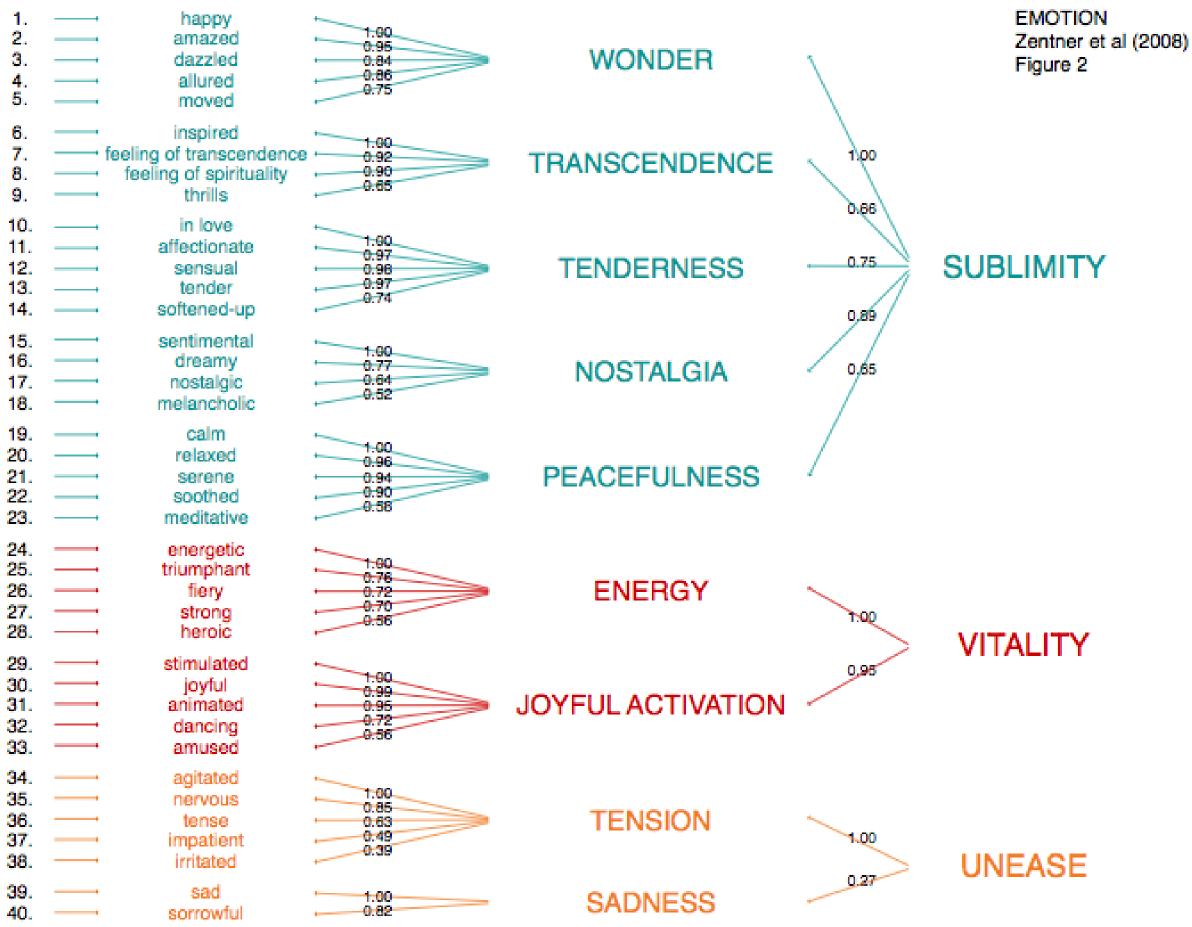


Figure L.06: The Geneva Emotion Music Scale (GEMS) model

However, each cluster has a high number of different emotions features, most of them being of similar or very close meaning [6]. This ambiguity raises several difficulties in discriminating one from another to obtain the “ground truth” [6]. Therefore, this study will consider major moods that can be used to easily identify and discriminate among each other such as sad, happy, angry, Neutral/Clam.

Most Musical Information Retrieval (MIR) systems are especially designed for automatic analysis of the musical databases. Majority of MIR systems index available musical databases based on song title or artist name where improper indexing can result in an incorrect search. More effective systems extract important features from audio and classify the audio to its genre based on those features. Therefore, those MIR systems have low-level or mid-level audio feature extraction capabilities which build applications involving automatic classification such as speech/music discrimination, music genre or mood recognition etc.

2.5 Possible sources for music moods

Two distinct sources *intrinsic* and *extrinsic* can be identified as main sources for music moods [7]. Intrinsic characteristics refer to specific structural or metadata characteristics of music while extrinsic referring to semantic contexts which is outside the music such as some social tagging [7]. Therefore, MIR to be more effective, combination of both music content as well as information shared by people who listen the music is important [7]. To this research work combining moods given by social tagging or other extrinsic sources are not considered.

2.6 Properties of the Representation Levels

There are multiple possible ways of representing music information in technical systems, and with computers. Such representations are chosen according to relevant viewpoints on the music content to match the target system functions. The term "Representation" refers to the way of information is represented internally in the system [2], i.e. essentially data structures.

Music Representation Types

There are two main distinct and complementary ways of representing music content in technical systems has been using for several decades.

1. **Audio signal representations**, resulting from the recording of sound sources or from direct electronic synthesis [2].
2. **Symbolic representations**, representations of discrete musical events such as notes, chords, rhythms, etc. [2]

The symbolic representation is content-aware and describes events in relation to formalized concepts of music (music theory), whereas the signal representation is a blind, content-unaware representation, thus adapted to transmit any, nonmusical kind of sound, and even non-audible signals [2].

Even digitized through sampling, the signal representation appears as a continuous flow of information, both in time and amplitude, whereas the symbolic representation accounts for discrete events, both in time and in possible event states (e.g. pitch scales). Low bandwidth control parameters, such as MIDI continuous controllers, are also part of this category [2].

It should also be noted that despite various existing methods for coding audio signals, be them analog or digital, even in compressed form, they all refer to and enable the reconstruction of the same representation of audio signals as amplitude functions of time [2].

Definition of Bit Size, Sample Rate, Channels, Data and header terms will help to understand the representation of audio files (See Appendix B).

2.7 Audio Features and Feature Extraction

Audio feature can be any lower dimensional representation which can be used to interpret the audio signal. Features can be either low-level (there is no any direct meaning) or high-level (humans refer terms to represent music such as tempo). These features can be use directly to compute the results or compute derived features which have more meaningful. In order to classify the audio signals to pre-define set of classes, these features in the audio signal to be extracted. Feature extraction is process of computing a compact numerical representation that represents the characteristics present on the audio signals. This is necessary because audio signals carry too much redundant and/or irrelevant information and computation can be done on a frame by frame basis or within segments, sounds or tracks. Audio feature extraction is most important thing in music mood recognition and good feature set is a must for classification. Feature extraction requires an in-depth understanding of the signal processing theories.

Audio features are metadata which extracted to do some useful classifications. There are three major categories of metadata [03].

(1) Editorial metadata: music reviews, inputs by music experts.
(2) Acoustic metadata: objective set of musical features obtained through analysis of an audio
(3) Cultural metadata: documents produced by a specific society or environment

Computers are unable to directly deal with continuous-time (CT) waveforms [13]. Then CT waveform needs to be sampled at regular time intervals before it can be stored and processed by a computer [13]. The sampling operation converts the CT signal into a discrete-time (DT) signal or sequence [13]. Following diagram illustrate the sign wave and sampling waveform with 16000 samples per second (Fs = 16000 Hz (16000 samples/sec)). Figure L.07.

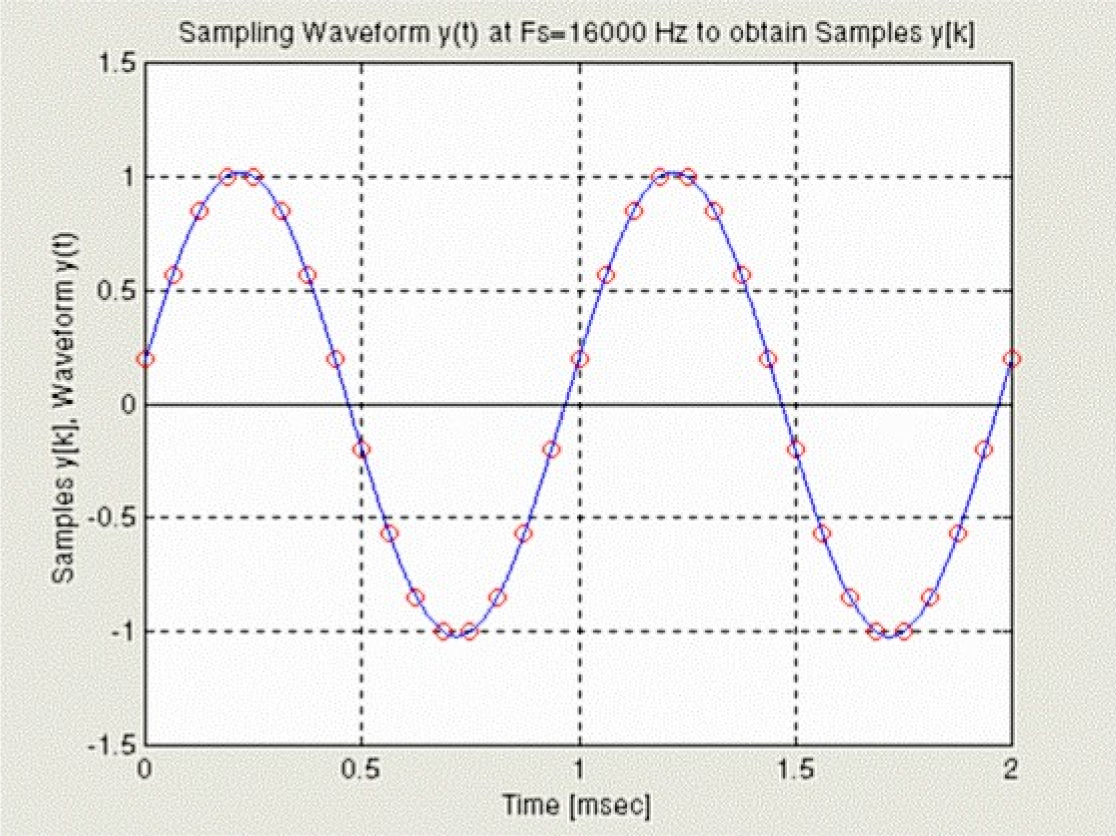


Figure L.07: Sampling of 1000 Hz Sine wave [13]

Usually audio features are extracted by breaking the DT (discrete-time) audio signal into windows or frames which may vary between 10-50 milliseconds length. Based on that following two approaches can be taken to process the signal [12].

1. *Single Vector Mode* [12]
    - Take whole file as texture window (A texture window is a long-term segment in the range of seconds containing several analysis windows).
    - Each file represents only single audio feature vector and subject to have only one classification.
    - Provides better description of the signal because feature vector generated for each window is measured variation in time of each feature.
    - Does not provide real-time classification.
2. *Texture Window Mode* [12]
    - Suppose to defining shorter texture windows
    - Make several class decisions along each file, one for each texture window.
    - At the end of the file the decisions are averaged to obtain a final class decision.
    - Provide real-time classification

Sound possesses three properties - intensity or loudness, pitch, and quality or timbre. These properties are which one sound may be distinguished from another.

Intensity/Loudness

The intensity of a sound depends on the density of the medium through which the sound is transmitted and the amplitude of the sound waves which reach the ear. The intensity of sound decreases as the distance from the source of sound increases [14].

Pitch

Pitch is the property of sound which determines whether the sound is high or low. Pitch is determined by the number of vibrations per second made by the sounding body or the frequency of the waves. Comparatively slow vibrations produce a low sound, while rapidly vibrating substances produce a high-pitched sound [14].

Quality/Timbre

It is possible to detect the difference in the sound produced by two bodies of different composition, but with the same intensity and pitch. The property of sound by which we can distinguish this difference is called quality. Quality depends upon the form of the vibrations or form of the waves [14].

According to [15] there are mainly four ways to distinguish the extractable features of the audio signal.

1. The **steadiness** or **dynamicity** of the feature
   Value represent which is extracted from the signal at a given time or value that represent signal behavior over the time such as mean, standard derivation etc.

2. The **time extent** description provided by the feature
   Extracting some features applies only part of the audio (e.g. attack of the sound) where some other features apply to entire signal (e.g. loudness). Based on this, time extend descriptions can be distinguish as

   Global descriptors: values are extracted or computed for the whole signal. If signal is segmented or divided in to frames, then each segment should be non-overlapping and required localized value of previous segment. (E.g. attack duration of the sound).

Instantaneous descriptors: values are extracted or computed for each time frame which varies signal length between 10-60 milliseconds. (E.g. spectral centroid).

3. The **abstractness** of the feature

   What feature represent. Either feature is abstract or concrete? This represent whether feature is mathematical process with little concrete meaning or does it represent something we can understand from experience?

4. The **extraction process** of the feature

   This can further distinguish into four categories

   a. Features that are directly computed from the audio signal (e.g. zero crossing rate).
   b. Features that are extracted or computed after transform the signal into FFT, etc. e.g. spectral characteristics.
   c. Features that are extracted from a model such as sinusoidal model or source/filter model. E.g. harmonicity/noisiness.
   d. Features that try to mimic the output of the ear system.

Audio signal features can be categorized [15] in to following groups. Figure L.08

- **The Temporal features**

  Temporal features are time domain features which are simple to extract and have easy physical interpretation. These features can be either global or instantaneous descriptors which computed from waveform or signal energy envelope. E.g., the energy of signal, zero crossing rate, maximum amplitude, minimum energy, etc.

- **The Energy features**

  These are features which refer to energy contents of the signal and are instantaneous descriptors. E.g. global energy, harmonic energy, noise energy.

- **The Spectral features**

  The spectral features are frequency based instantaneous descriptors which are obtained by converting the time-based signal into the frequency domain using the Fourier Transform, e.g. spectral centroid, spectral flux, spectral density, spectral roll-off, etc. These features can be used to identify the notes, pitch, rhythm, and melody.

- **Harmonic features**

  Instantaneous features that are calculated from sinusoidal modeling (Sinusoidal modeling is an analysis of digital audio signals to decomposed into sinusoids component. This process involves detecting and extracting sinusoids/sign curves from the original signal). E.g. harmonic/noise ratio, harmonic energy ratio, harmonic deviation.

- **Perceptual features**

  This involve with finding audio features based on the human perception. E.g. loudness, sharpness.
- **Statistical features**

  Features that are statistically calculated from other extracted features such as mean, standard deviation. There is possibility of extract statistical features from almost all features [6].

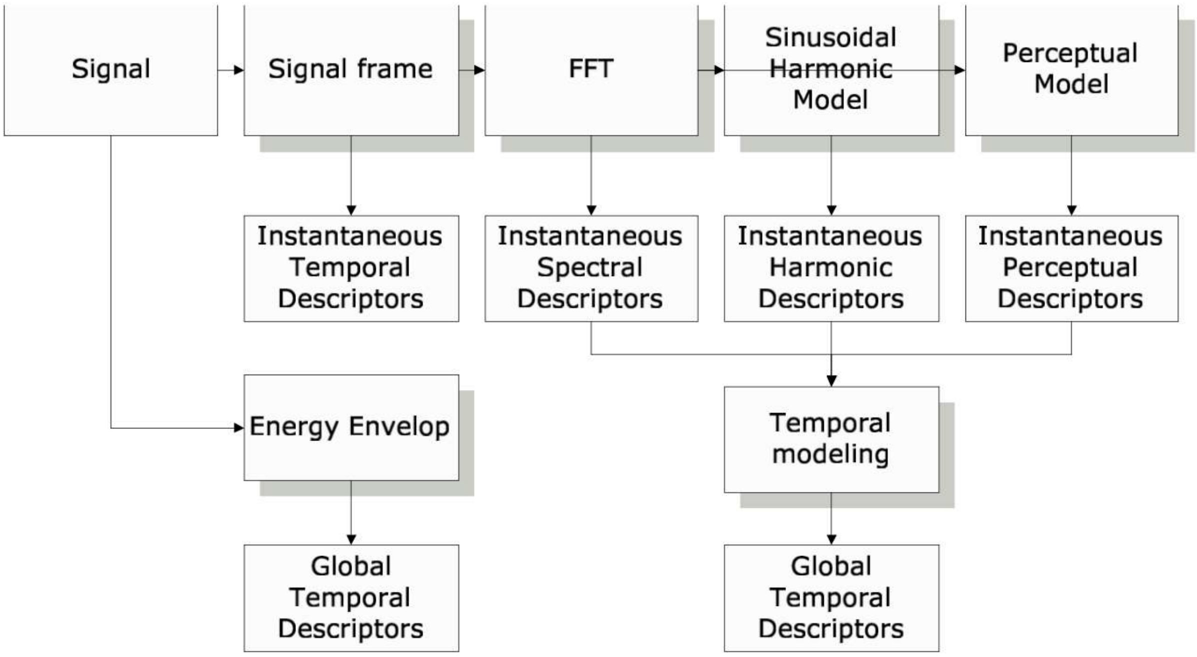


Figure L.08: Detail of temporal, energy, spectral, harmonic and perceptual descriptor's extraction process [15]

Audio features described under above feature categories are used to classify the music into different perceptual categories [03] such as tempo, mood (happy/neutral/sad), emotion (soft/neutral/aggressive), complexity, and vocal content [03].

2.7.1 Audio Features

This section will describe the selected audio features for the analysis, tools and techniques used to extract the features and their basic descriptions. Mainly MIRToolbox 1.7 and jAudio 3.0 of jMIR is used to extract the features. MIRToolbox is Mathlab extensions and required Mathlab to run and extract the MIRToolbox supported features. JAudio has GUI which can be easily extracting the selected features either to XML of ARFF formats.

2.7.1.1 Dynamic Features

RMS (Root-Mean-Square Energy)

The global energy of the signal can be computed simply by taking the root average of the square of the amplitude, also called root-mean-square (RMS). Square each value of the signal's voltage and current, add up the squares and divide by the number of samples to find the average square or mean square. Then take the square root of that. [18]

$$x_{RMS} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} x_i^2} = \sqrt{\frac{x_1^2 + x_2^2 + \cdots}{n}}$$

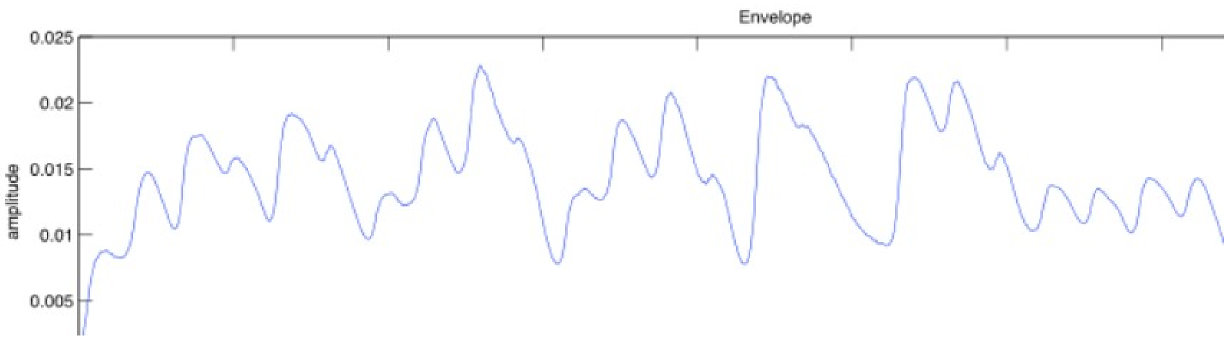


Figure L.09 Amplitude of the signal over the time [18]

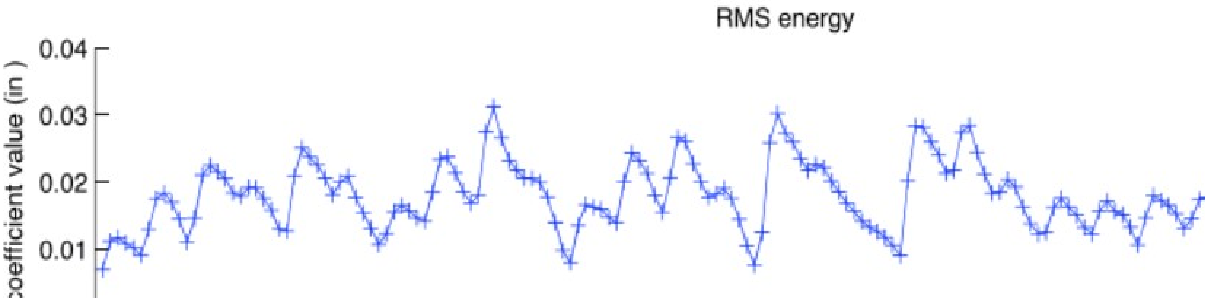


Figure L.10 Energy curve of the signal [18]

Low Energy

The energy curve can be used to get an assessment of the temporal distribution of energy, to see if its remains constant throughout the signal, or if some frames are more contrastive than others. One way to estimate this consists in computing the low energy rate [6].

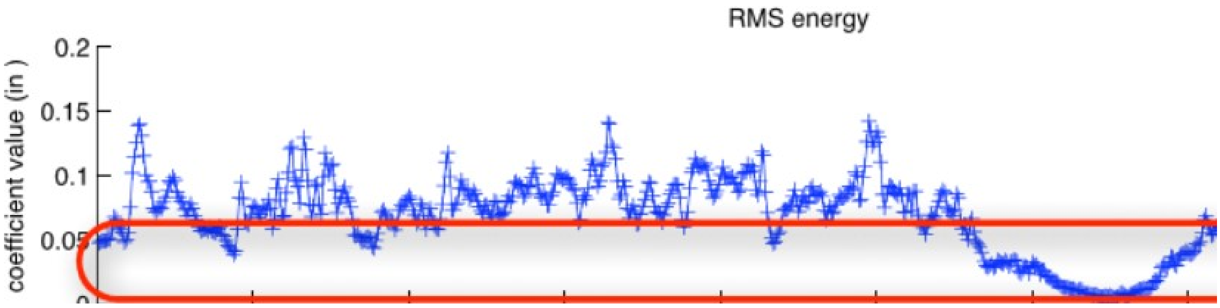


Figure L.11 Low Energy [18]

Frames that have quite constant high energy and fewer with very low energy will lead to have low low-energy rate

2.7.1.2 Rhythmic Features

Tempo

Tempo can be defined as the pace or speed at which a section of music is played. Tempos, or tempi, help the composer to convey a feeling of either intensity or relaxation. Tempo can be considered as the speedometer of the music. Typically, the speed of the music is measured in beats per minute, or BPM. Usually tempo is estimated by detecting periodicities from the event detection curve. [18]

Autocorrelation

Autocorrelation is another way to evaluate periodicities in signals either through audio waveform, a spectrum, or an envelope of the signal [18]. This involved correlation of a signal with a delayed version of itself as a function of delay to find similarity of a signal. The analysis of autocorrelation is a mathematical tool for finding repeating patterns, such as the presence of a periodic signal obscured by noise or identifying the missing fundamental frequency in a signal implied by its harmonic frequencies [Wikipedia]. A perfectly periodic signal will have its highest autocorrelation value when j is the same as its period.

$$R_{xx}(j) = \sum_n x_n \overline{x}_{n-}$$

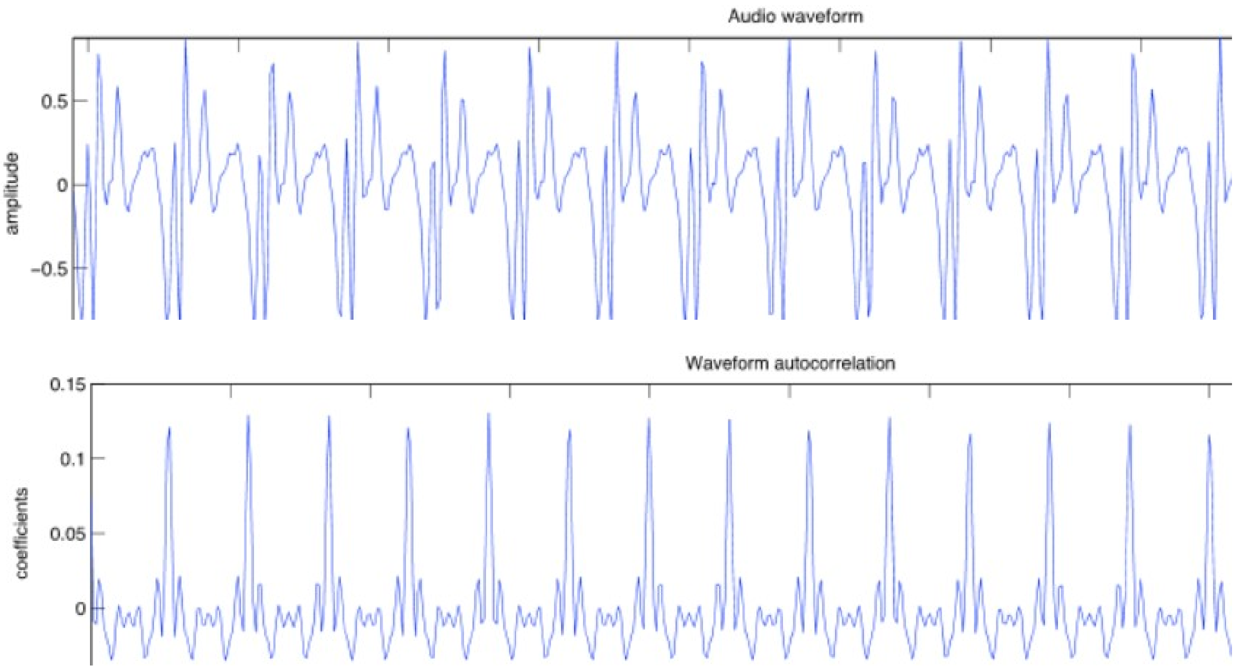


Figure L.12 Autocorrelation of a signal [18]

Onset

Onset is also referred to as temporal location of events which is another way of determining the tempo. This is achieved based on computation of detection curve to show the successive bursts of energy corresponding to the successive pulses. A peak picking is automatically performed on the detection curve.

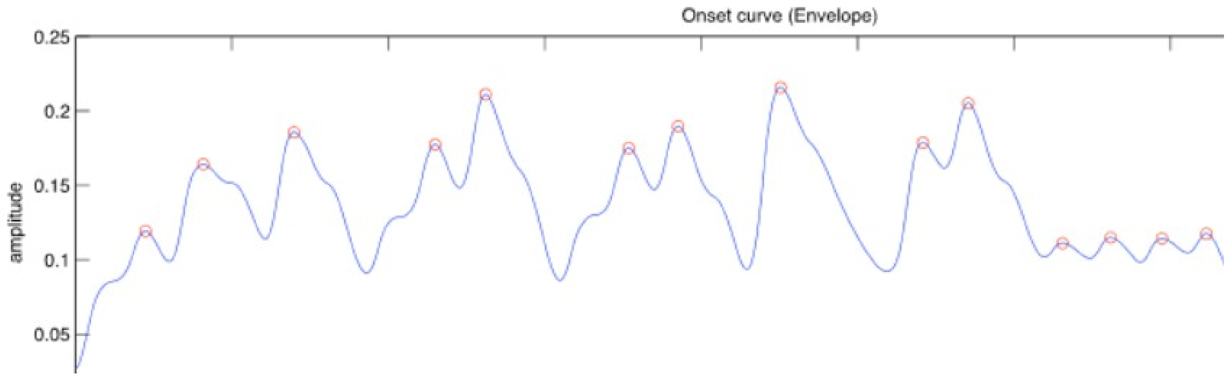


Figure L.13 Onset Curve [18]

2.3.1.3 Timbre Features

Attack Time

Attack time is the estimation of temporal duration for a signal to rise to its peak (e.g. attack time in amplitude).

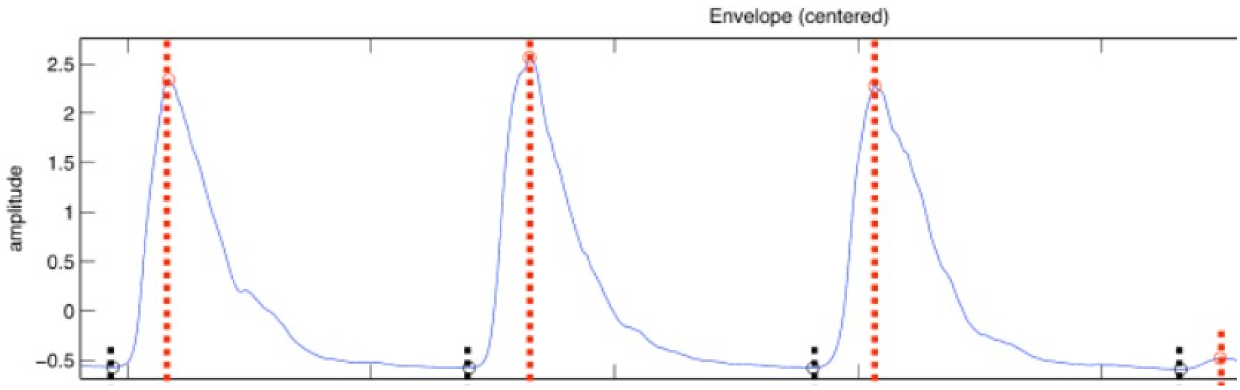


Figure L.14 Attack time in amplitude [18]

Attack Slope

Attack slope is another description of the attack phase. It consists on calculating the average slope of the entire attack phase, from its start to the peak.

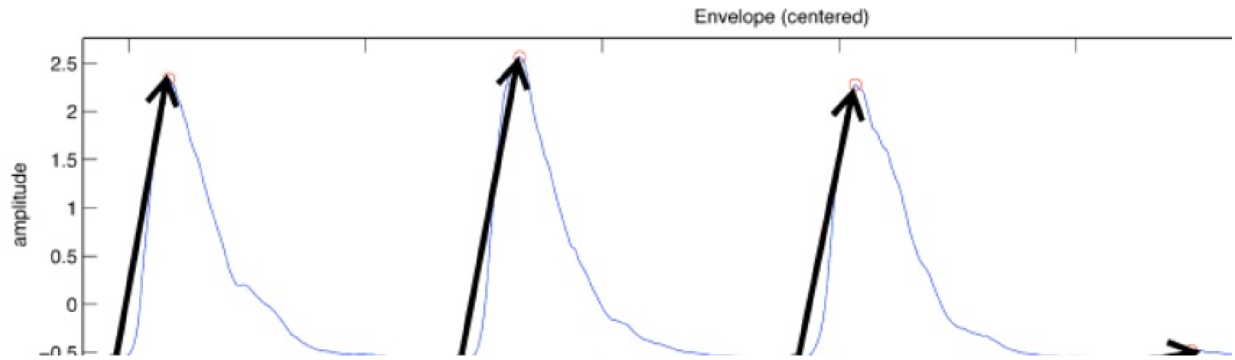


Figure L.15 Attack slope [18]

Spectral Centroid

An important and useful description of the shape of a spectrum can be obtained using its moments. The first moment, called the mean, is the geometric center (centroid) of the distribution and is a measure of central tendency for the random variable [18]. Higher centroid values correspond to "brighter" textures with more high frequencies. It is defined as the center of gravity of the magnitude spectrum of the STFT [6].

$$\mu_1 = \int x f(x) dx$$

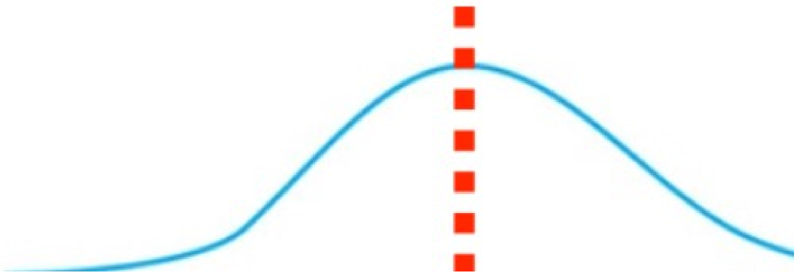


Figure L.16 Spectral Centroid [18]

Spectral Spread

This is second central moment of the spectrum distribution which gives the variance. Being the squared deviation of the random variable from its mean value, the variance is always positive and is a measure of the dispersion or spread of the distribution [18].

$$\sigma^2 = \mu_2 = \int (x - \mu_1)^2 f(x) dx$$

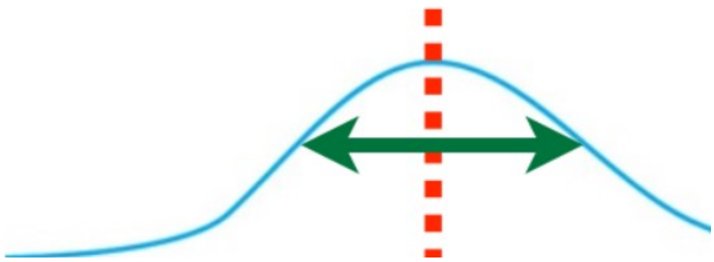


Figure L.17 Spectral Spread of the distribution [18]

Spectral Skewness

Skewness is a measure of the symmetry of the distribution. The skewness can have a positive value when distribution is said to be positively skewed with a few values much larger than the mean and therefore a long tail to the right. A negatively skewed distribution has a longer tail to the left. [18]

$$\gamma^3 = \frac{\int (x - \mu)^3 \cdot p(x)}{\sigma^3}$$

X are observed data (e.g. the frequencies of the spectrum), P(x) are the probabilities to observe X (e.g. the amplitudes of the spectrum).

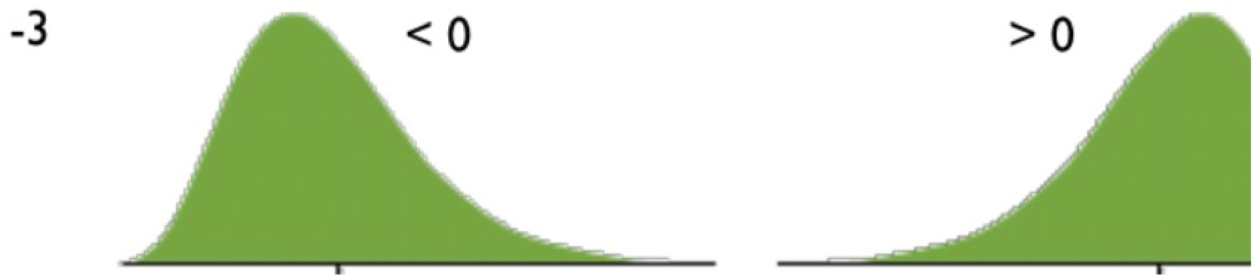


Figure L.18 Spectral Skewness [18]

Spectral Kurtosis (Flatness)

Described as sharpness of the peak of a frequency distribution curve. Kurtosis is a measure of whether the data are heavy-tailed or light-tailed relative to the normal distribution. Then, data sets with high kurtosis tend to have heavy tails, or outliers. Data sets with low kurtosis tend to have light tails, or lack of outliers.

$$\gamma^4 = \frac{\int (x-\mu)^4 \cdot p(x}{\sigma^4}$$

Figure L.19 Spectral Skewness [18]

Spectral Roll off

The frequency below which 90% of the magnitude distribution of the spectrum is concentrated. This is a measure of the amount of the right-skewedness of the power spectrum. The spectral roll off point is the fraction of bins in the power spectrum at which 85% of the power is at lower frequencies. Default is fixed to 85% and some other framework used 95%. [18]

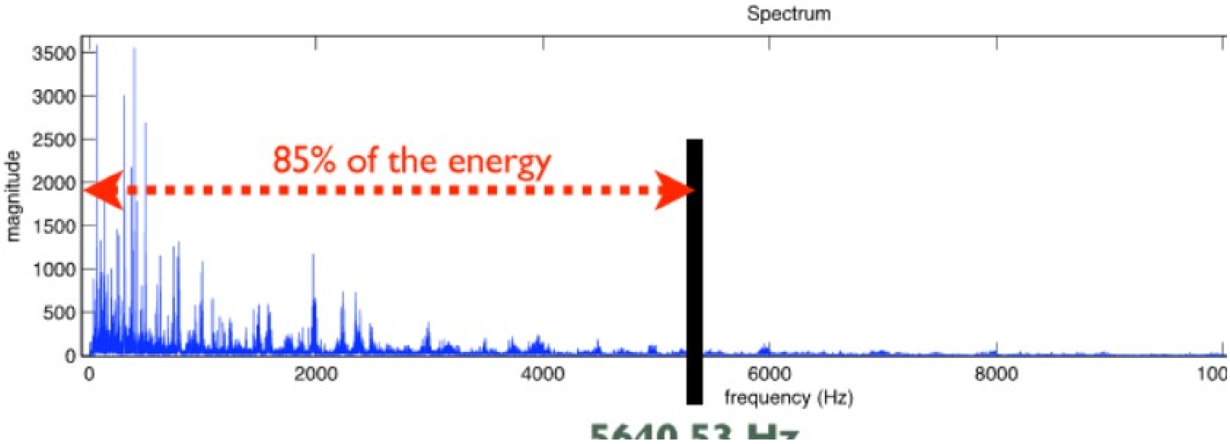


Figure L.20 Spectral Roll Off [18]

Entropy of Spectrum

Compute the relative Shannon (1948) entropy of the input which used in information theory and calculated based on following equation [18],

$$H(X) := -\sum^{n} p(x_i) \log$$

where b is the base of the logarithm. Entropy offers a general description of the input curve p and indicates whether it contains predominant peaks or not [18].

Spectral Flatness (SF)

Indicates whether the distribution is smooth or spiky, and results from the simple ratio between the geometric mean and the arithmetic mean [18]:

$$\frac{\sqrt[N]{\prod_{n=0}^{N-1} x(}}{(\sum_{n=0}^{N-1} x(n)}$$

This is usually measured by decibels and provides a way to quantify how noise-like a sound is, as opposed to being tone-like [Wikipedia]. A high SF indicates that the spectrum has a similar amount of power in all spectral bands and would sound similar to white noise, and the graph of the spectrum would appear relatively flat and smooth. A low spectral flatness indicates that the spectral power is concentrated in a relatively small number of bands – this would typically sound like a mixture of sine waves, and the spectrum would appear "spiky"[6].

Roughness

Roughness is also described as Sensory Dissonance which related to the beating phenomenon that occurs whenever a pair of sinusoids is close in frequency [18] [6]. Estimation of roughness depending on the frequency ratio of each pair of sinusoids represented as follows:

Figure L.21 Sensory Dissonance depending on frequency ratio [18]

Spectral Irregularity

Spectral irregularity or spectral peaks variability describe the degree of variation of the successive peaks of the spectrum [18]. Two distinct calculations are available with MIR Toolbox;

1. Based on (Jensen, 1999), where the irregularity is the sum of the adjoining partials, square of the difference in amplitude between adjoining partials. (Default approach) [18]

$$\left(\sum_{k=1}^{N}(a_k - a_{k+1})^2\right) / \sum$$

2. Based on (Krimphoff et al., 1994), where the irregularity is the sum of the amplitude minus the mean of the preceding, same and next amplitude [18]

$$\sum_{k=2}^{N-1}\left| a_k - \frac{a_{k-1} + a_k +}{3} \right.$$

Brightness

Brightness, also called as High frequency energy consists in fixing a minimum frequency value/ cut-off frequency and measuring the amount of energy above that frequency. The result is expressed as a number between 0 and 1 [18] [6].

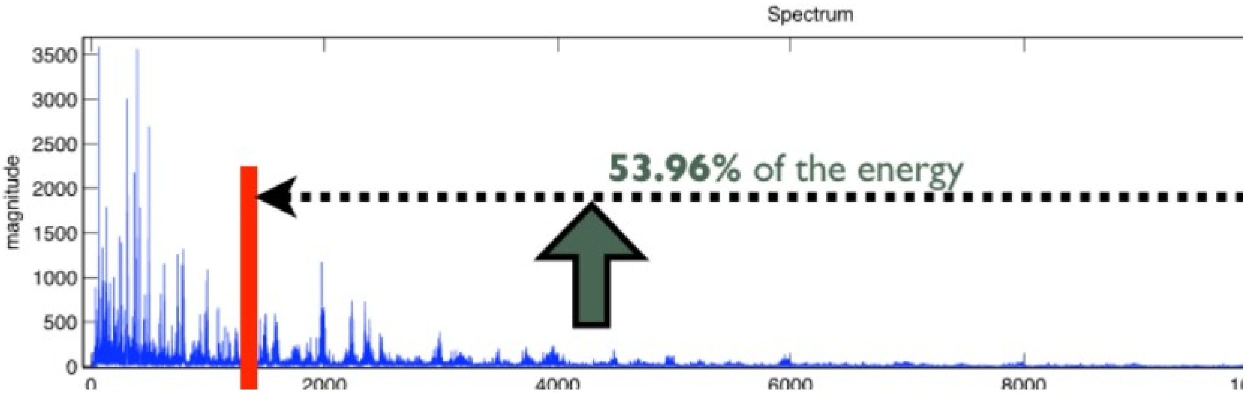


Figure L.22 Brightness/High Frequency Energy [18]

Mel Frequency Cepstral Coefficient (MFCC)

In automatic speech recognition, identifying components of the audio signal that can be used to recognize the verbal content and discarding all the other stuff which carries information like background noise are important. The main point of speech recognition is that the sounds generated by a human are filtered by the shape of the vocal tract including tongue, teeth. This shape determines what sound comes out. Possibility of determining the shape accurately, give an accurate representation. The shape of the vocal tract manifests itself in the envelope of the short time power spectrum, and the job of MFCCs is to accurately represent this envelope. MFCCs offer a description of the spectral shape of the sound [18]. The frequency bands are positioned logarithmically (on the Mel scale), which approximates the human auditory system's response more closely than the linearly-spaced frequency bands [6].

The calculation of this feature works as follows. After taking the log amplitude of the magnitude spectrum, the FFT bins are grouped and smoothed according to the perceptually motivated Mel-frequency scaling. Then, to decorrelate the resulting feature vectors, a Discrete Cosine Transform is performed. Usually, only the first 13 components are returned since most of the signal information tends to be concentrated in a few low-frequency components of the discrete cosine transform (DCT). These 13 coefficients are mostly used for speech representation but states that the first five coefficients are adequate for music representation [6].

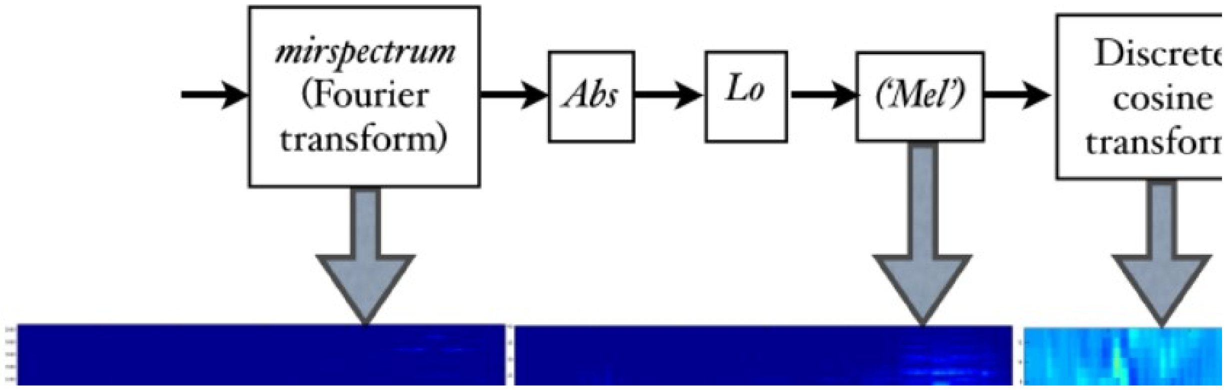


Figure L.23 Brightness/High Frequency Energy[18]

Deltas and Delta-Deltas MFCC

Also known as differential and acceleration coefficients respectively and improve the performance when combine with the MFCC.

Zero-Crossing Rate (ZCR)

The rate of sign-changes of the signal during the duration of a frame or number of times the signal crosses the X-axis [18]. This can be used as a simple indicator of noisiness. As an example, heavy metal music, due to guitar distortion and heavy percussion, will tend to have much higher zero crossing values than classical music [6]

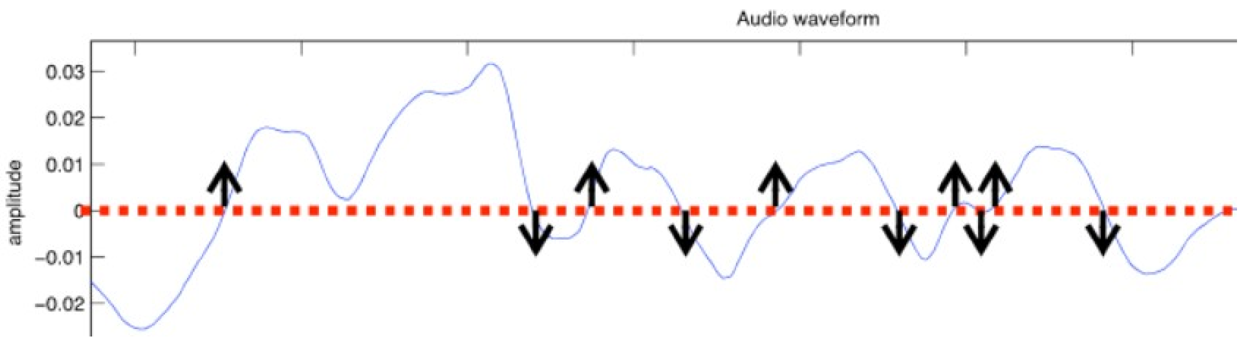


Figure L.24 Zero Crossing Rate - Waveform crossing the X-Axis [18]

Spectral Flux

Spectral flux is a measure of how quickly the power spectrum of a signal is changing, calculated by comparing the power spectrum for one frame against the power spectrum from the previous frame. This is also called as distance between successive frames. The spectral flux is not dependent upon overall power can be used to determine the timbre of an audio signal, or in onset detection among other things.

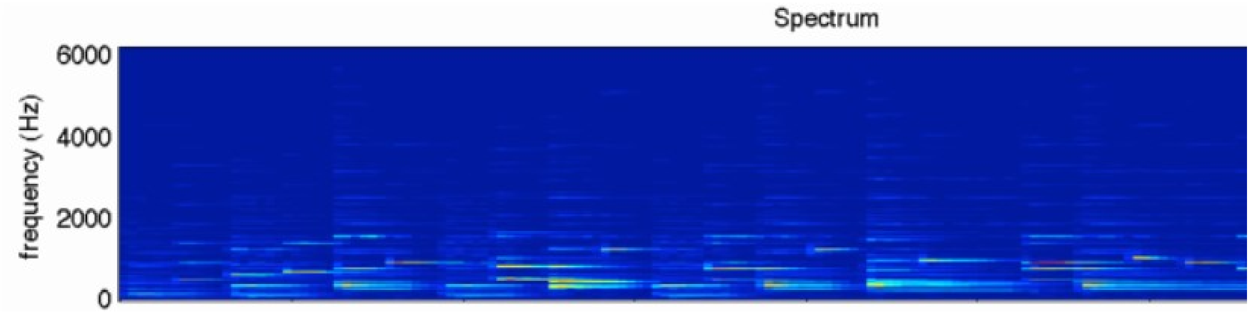


Figure L.25 Spectrum with beginning of frames' temporal locations [18]

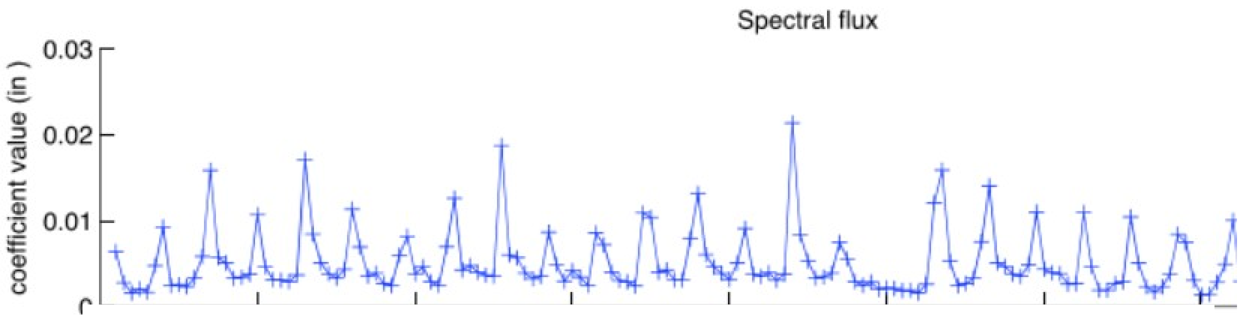


Figure L.26 Spectral flux [18]

The peaks in the curve indicate the temporal position.

Compactness

A measure of the noisiness of a signal. Found by comparing the components of a window's magnitude spectrum with the magnitude spectrum of its neighboring windows.

Spectral Variability

The standard deviation of the magnitude spectrum. This is a measure of the variance of a signal's magnitude spectrum.

Strongest Beat

The strongest beat in a signal, in beats per minute, found by finding the strongest bin in the beat histogram.

Beat Sum

The sum of all entries in the beat histogram. This is a good measure of the importance of regular beats in a signal.

Strength of Strongest Beat

How strong the strongest beat in the beat histogram is compared to other potential beats.

LPC

Linear predictive coding (LPC) is a tool used mostly in audio signal processing and speech processing for representing the spectral envelope of a digital signal of speech in compressed form, using the information of a linear predictive model [Wikipedia]. LPC is a method of speech coding that constructs a digital filter that will model a short segment (called a frame) of a speech waveform using many fewer bits than the explicit waveform itself requires. LPC determines the portions of the speech frame (often about 20 ms long) that are voiced from the voiceless parts. For the voiced portions, it measures the F0 (based on the distance between apparent pulses). It also measures the mean amplitude of the frame

https://www.cs.indiana.edu/~port/teach/641/signal.proc.html

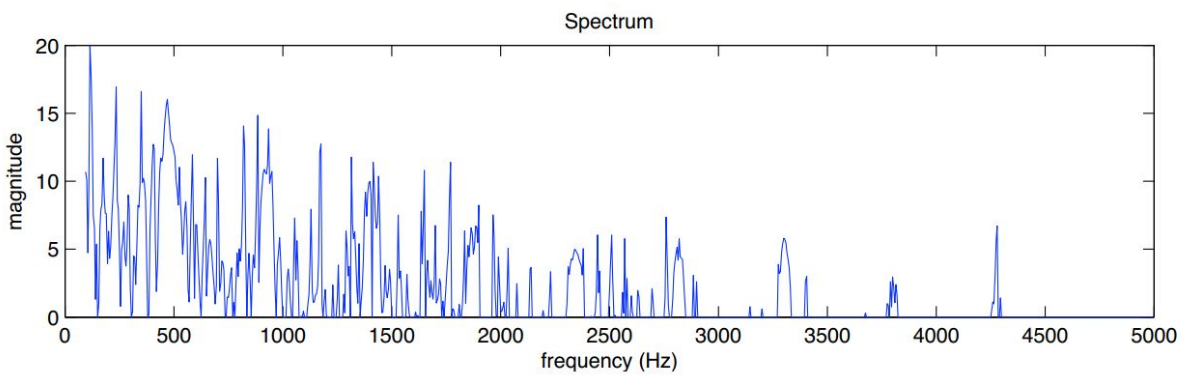
Method of Moments

Statistical Method of Moments of the Magnitude Spectrum.

2.7.1.4 Tonality Features

Chromagram

Chromagram also called as Harmonic Pitch Class Profile which shows the distribution of energy along the pitches or pitch classes. The Chromagram is a redistribution of the spectrum energy along the different pitches (i.e.,“chromas”) [18]



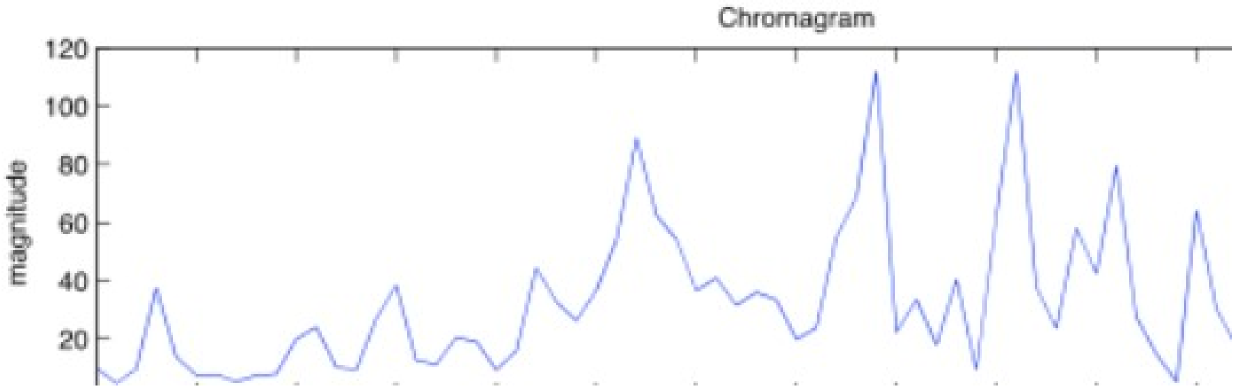


Figure L.27 Chromagram [18]

Key Clarity

Gives a broad estimation of tonal center positions and their respective clarity [18]

Mode

Estimates the modality, the difference between major and minor keys and, returned as a numerical value between -1 and +1: the closer it is to +1, the more major the given excerpt is predicted to be, the closer the value is to -1, the more minor the excerpt might be [18] [6].

Key Strength

Key strength also known as probability of key candidate computes the key strength, a score between -1 and +1 associated with each possible key candidate, by computing correlation coefficients between the Chromagram [18].

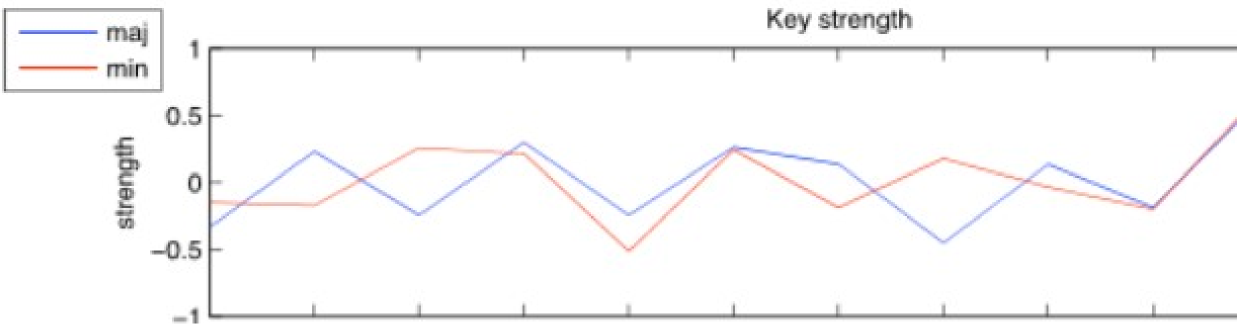


Figure L.28 Correlation coefficients for each different tonality candidate [18]

Harmonic Change Detection Function (HCDF)

The Harmonic Change Detection Function (HCDF) is the flux of the tonal centroid [18]

2.7.2 Feature Extraction

Usually when mining raw music data for analysis, it is considered as too large, noisy and redundant. Therefore, input signal is transformed into small space of variables that simplify the analysis process. This can be achieved through Fourier Transform (FT) and Sort-Time Fourier Transform (STFT) where signal is filtered in to bunch of circular paths and Fourier transform in to short time segments respectively. Every circular path required size, speed and starting angle of the circle which map the amplitude, frequency and phase of the sinusoidal. The combination of all cycles represents the audio signal. STFT is required in order proceed with block-based processing. Most of the time, one feature is not sufficient for analysis. Thus, required combining more features together which built the feature vector or multi-dimensional space. Frequently mid-level and low-level features are extracted for analysis (Figure L.09).

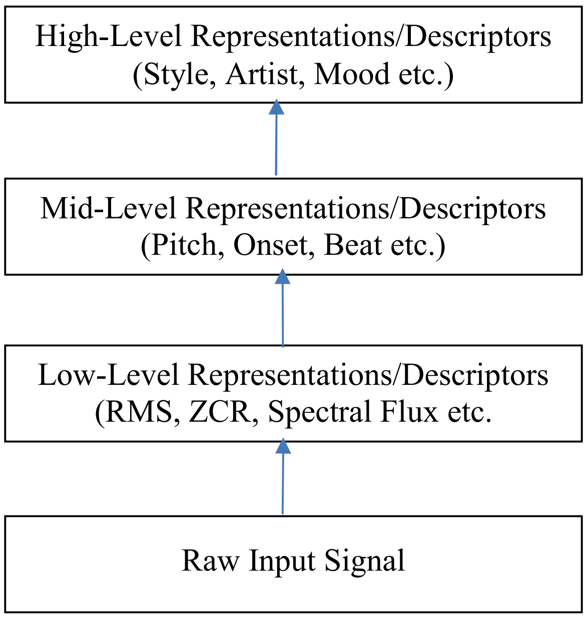


Figure L.29: Music Feature Extraction Stack

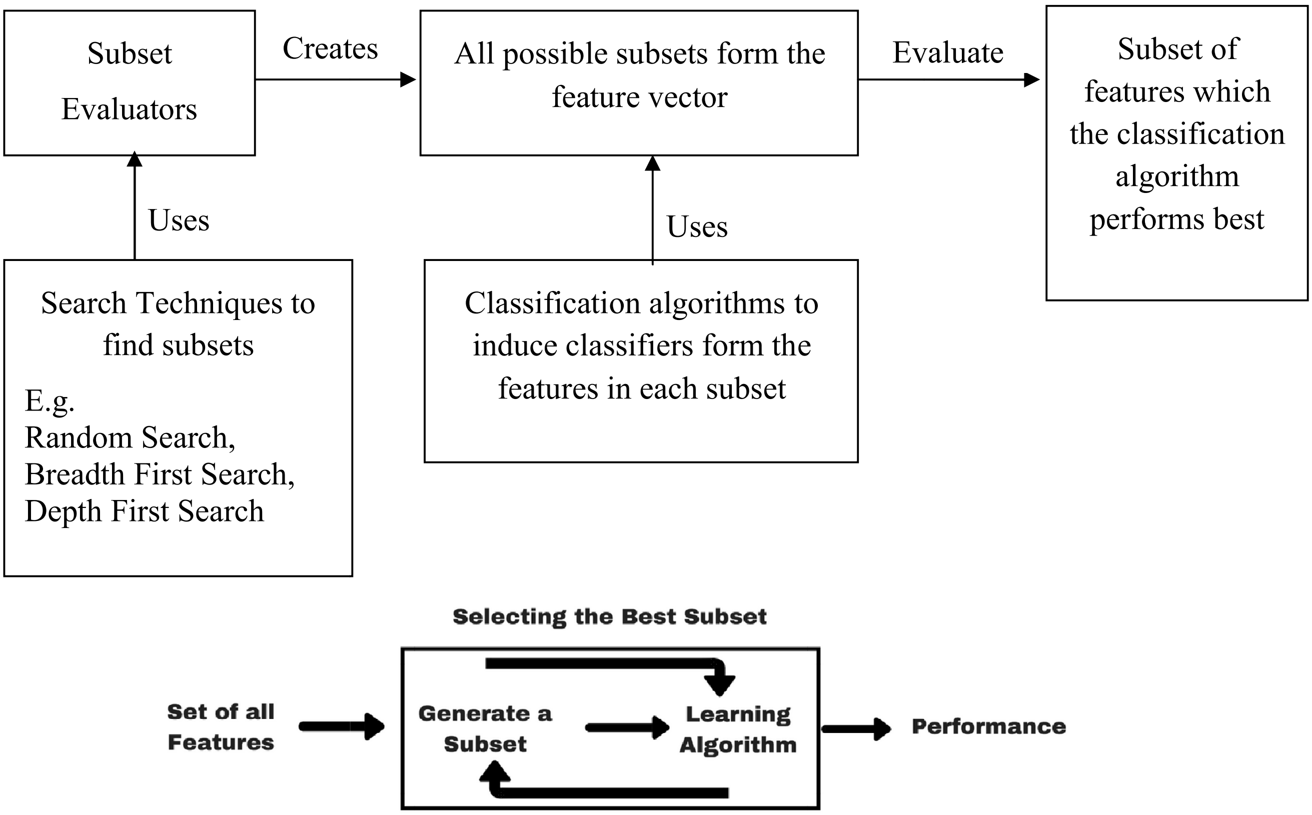
2.8 Feature Selection

Machine learning is all about algorithms applies to solve the specific classification problems. Success of using machine learning is not depend entirely on the selection of best algorithm, but mainly on the feature creation and feature selection. If put garbage as input for machine learning or any other algorithm, garbage results come out. This is the simple rule which machine learning works. When have large number of features, feed algorithm with features that are most important or relevance to undergoing task. This process affects to provide better results, reduce training and evaluation time and reduce the hassle when deal with large set.

Therefore, feature selection provides a compact feature subset which has an interesting trade-off between classification accuracy and computational effort.

There are mainly two feature selection methods, (1) Wrapper methods which are embedded in to the learning or classification process (Figure L.10) (2) Filter methods which used as part of the pre-processing steps (Figure L.11).

2.8.1 Wrapper Method Approach



https://www.analyticsvidhya.com/wp-content/uploads/2016/11/Wrapper_1.png

Figure L.30: Wrapper Method Approach for Feature Selection

In wrapper methods, there are commonly used search methods for feature subset selection

1. **Forward Selection**

   This is an iterative method which starts the model with no features and keeps adding features in each iteration that best fit for the performance of the model. Process terminated when not improves the performance when adding new features. This is related BestFirst search algorithm in WEKA with forward search direction.

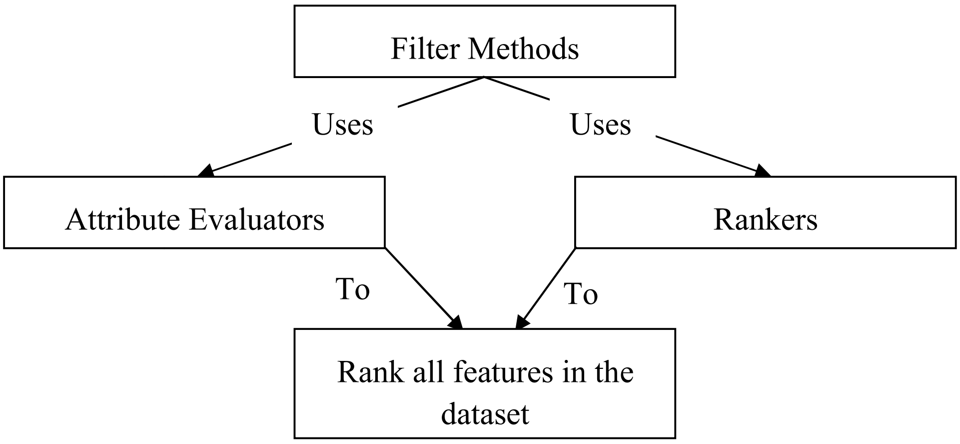
2. **Backward Elimination**

   This is an iterative method which start model with all features and keep removing features that are least significant to improve the overall model performance. Process terminated when not improves the performance when removing new features. This is related BestFirst search algorithm in WEKA with backward search direction.

3. **Recursive Feature Elimination**

   This is a greedy optimization algorithm which creates models and keeps aside the best or the worst performing feature at each iteration. Each iteration create model with left features until all features are evaluated. This is related GreedySetpwise search algorithm in WEKA.

2.8.2 Filter Method Approach



https://www.analyticsvidhya.com/wp-content/uploads/2016/11/Filter_1.png

Figure L.31: Filter Method Approach for Feature Selection

Number of features required to select from the feature vector can be defined and omit one feature at a time that has lower ranks and see the predication accuracy of the classification algorithm.

Selection of attributes is critically important because it is affected to the ability of successfully and meaningfully modeling the problem or not. Including redundant attributes in dataset misleading the modeling algorithms and tend to have predications to be skewed. Also keeping irrelevant attributes leads have an overfitting situation which negatively impact to the modeling power of the algorithm and can caused to predictive accuracy. Therefore, it is important to remove redundant and irrelevant attributes from the dataset before evaluating through algorithms. Then feature selection or attribute selection is processes of selecting best subset of features from the dataset which tend have highest accuracy. *Reducing overfitting* (reduces less opportunity or noise attributes), *improve accuracy* (less misleading attributes tend for more accuracy) and *reducing training time* (less data tend train algorithms faster) and reducing the complexity for easy interpretation are the major benefits of feature selection.

Following tools can be used to feature selection

- WEKA
- using scikit-learn in Python
- Using Caret R package in R

In WEKA, feature selection can be achieved through two steps

1. Attribute Evaluator: each attribute/feature in the dataset is evaluated in the context of the target variable or the class attribute based on the selected evaluator either filter or wrapper. Some attribute evaluation techniques require specific search method to be used. E.g.CorrelationAttributeEval technique used Ranker search method.
2. Search Method: used to navigate different combinations of attributes/subsets in the dataset to arrive on a short list of chosen features.

2.9 Feature Classification

Classification is a process of automatically assigning a category or label from several categories or labels to an individual item based on the item's characteristics or features. For this research work, individual items are music samples with 30 seconds length and characteristics are the features extracted from samples. Categories/labels are classes defined for categorized the samples as anger, happy, neutral and sad. Complexity of the classification depends on the relationship between extracted features and the classes. In any classification task, identify the problems and building the ground truth dataset is the initial and most challenging step. Main problems of this work can be identified as subjective nature of the mood perceptions, time consuming and labor-intensive mood annotation process, different mood models and relatively small datasets. As moods are vague, there is no straightforward or base model for represent. Thus, according to the literature almost all papers used their own or different models, feature sets, extraction tools and methods, feature selection algorithms and classification algorithms which makes difficult for comparison.

There are two major classification methods (1) Supervised and (2) unsupervised where supervised learning has priori knowledge dataset to train the model and get the desired output whereas unsupervised learning does not have any initial dataset and outcome/grouping is based on the clustering data into different classes. In supervised learning model, can predict the target label after sufficient training on the ground truth dataset and model find the way to accurately find the target respective to the inputs. When target is different set of distinct classes or labels, then it is referred as classification problem and when target is continuous values, then it became as regression problem. In unsupervised learning, algorithm is working only on the inputs to find the relationships and build the different clusters for inputs. Then new inputs can be put in to the appropriate cluster. Supervised learning methods considered as fast, accurate and must be able to generalize which is ability to predict correct result with unseen data. Most common problem associated this learning is overfitting where algorithm is working well with training data and poorly with new data. To carry out good classification work there supposed to have three datasets:

1. Training set – used for model training and predict the known target
2. Validation set – used to tune the model and estimate the model prediction error
3. Test set – used only to access the performance of the classifier and should never use during the training and validation.

Neural Networks, decision trees, support vector machines are commonly used supervised learning algorithms while k-means, self-organizing maps are the commonly used unsupervised learning algorithms.

Chapter 3: Analysis and Design

The project "Music Mood Classification Using Audio Features" involve with subject domains such as audio signal processing, feature extraction techniques, feature selection and classification algorithms. There is vast literature to capture to understand the complete project and identify the way forward with best set of tools and techniques. Main areas such as behavior and properties of audio signals, available features of audio signal, tools and techniques available to process and extract audio features are studied deeply and engaged with understanding, comparing and selecting of feature selection and classification techniques and algorithms. This is more challenging because there is no any standard to follow when processing audio signals to annotate music moods.

Initial proposal mentioned that mood detection can be achieved via clustering algorithms. But when proceed through the literature review, identified that mood detection is possible via classification algorithms because testing data is annotating to predefined set of mood labels through supervised learning.

When proceed through the literature following challenges and difficulties are faced.

1. Finding sample data sets: dataset should include at least more than 200 music samples with assigned mood labels to use as training and testing. Most data sets are not released for public.
2. Inability to find standard way to follow: each researcher follows their own ways to analyze the music moods. There is no any standard set for music features which can be considered to extract for mood detection. Most researchers combined different sources for feature extraction such as audio contents and lyrics.

Most universities had music research laboratories dedicated to facilitating music and audio related research projects and share the knowledge and do the experiments. For this project all details and knowledge can only obtain through the research papers and resources available on Internet.

When analyze music modes via computational model, essentially required to access data which contain in the music file as well as their corresponding mood attached with the music. Then classifier model can be used to check whether our model result the corresponded mood of song or not. For the given music file, model should always match to one of predefined label/mood under training data. This project uses single-class labeling to assigns one model label to music piece. But, people might be annotated same music file with different moods based on their mood preference. Then there should be a mechanism to maintain the different moods annotate by music and rate the mood based on the most frequent search. Below illustrates the planed approach for implementation based on the gathered literature so far.

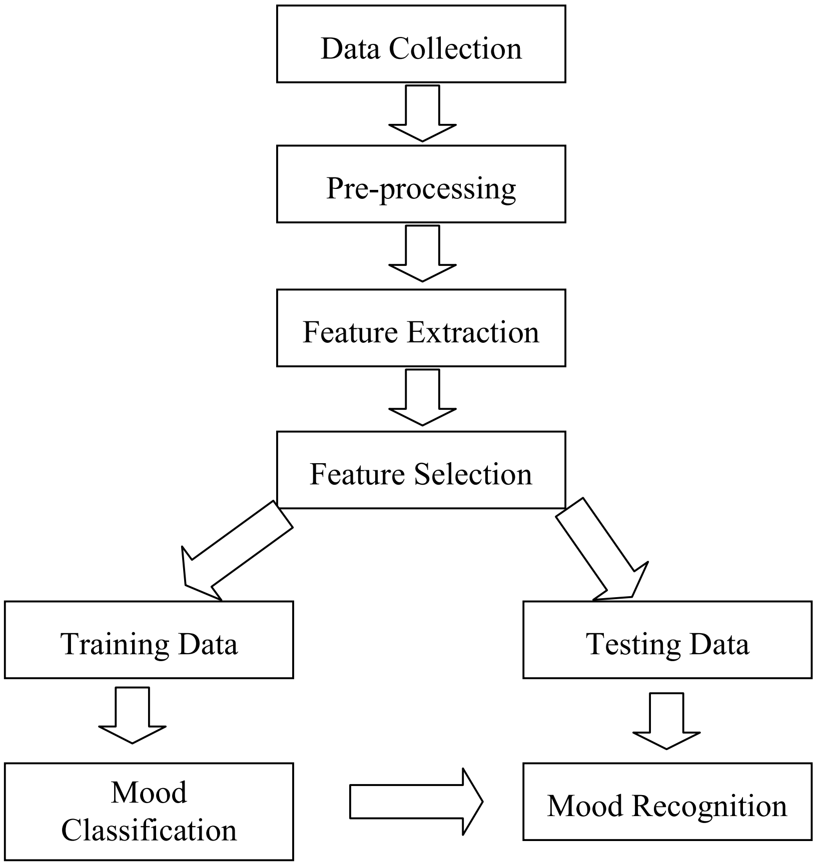


Figure D.01: Planed Approach for Implementation

Align with the literature; music moods can be considered through mainly two types of emotion models, categorical or dimensional. Lack or not availability of standardization among existing models, oversimplification, ambiguity and dissimilarity leads to problem of incompatibility between models and caused selection of suitable model more challenging [17]. This project mainly focuses on the categorical approach and select four basic moods happy, sad, angry and neutral/relaxed as distinct mood labels for classification. The objective of this project is to categorized songs based on music mood which is subjective criteria vary between individuals.

3.1 Data Collection

Creating an emotionally annotated dataset for classification research is not an easy and challenging task. Accuracy and success of the research work is highly depending on the correctly annotated music samples. The main obstacle is the subjective and uncertain nature of the music perception. Making a success and trustworthy dataset required human experts or evaluators to assign each song to one or more music moods. Volume of the sample leads to make this process time consuming and end up with high cost. Also, another major obstacle is that unavailability of the common representation model for the music emotion and large and professional datasets are not available to the public. As a result, majority of researches experiment with small datasets which is less than 1000 songs.

For this research work, manually prepared the sample dataset due to unavailability of annotated and available adequate datasets in public and poor response for the requests made by asking private datasets. All music pieces are downloaded from www.us.audionetwork.com web site as streaming MP3 files. Effort given for download music through www.audiosparx.com has failed because each music has proprietary voice message periodically which may impact to the analysis. *Ant Video Downloader 3.1.24,* which is free and open source Firefox browser add-on is used to download audio files. Each category consists 200 samples for training and 20 samples testing which total up as 800 for training and 80 for testing.

3.2 Per-processing for feature extraction

All music files consists MPEG 1 Layer 3, stereo channel mode with 48,000 sample rates. *Exiftool* which is free and open source command line tool used to extract the metadata for sound files. Then *FFmpeg 3.4,* which is command line free and open source leading multimedia framework is used to extract the first 30 seconds from audio files and convert that first 30 seconds audio to WAVE file format. [1,2]The conversion from MP3 to WAVE is done because MP3 files are compressed and contain 10-50 milliseconds silent gap between beginning and end of the file due to the compression algorithms and that is referred as encoder delay. [1,2]But WAVE files are uncompromised quality files with lossless, uncompressed, broadcast CD quality music which better fits for analysis tasks. [1,2]Also WAVE file of 44,100 kHz 16 bit has full frequency response up to 22 KHz where MP3 cut off around 18 KHz while possibility of having humans to listen up to 22KHz. Therefore, converted files WAVE is considered for feature extraction process.

3.3 Feature Extraction

There is no any proper guideline found during literature to identify the features that are mostly relevant for each music mood. Therefore, all possible mid-level and low-level [17] music features that could be extracted from both jAudio from JMIR and MIRToolbox tools are considered. MIRToolbox has special statistical MATLAB function *mirfeatures,* which computes large set of features and return them on a structure array at once instead of issuing sperate functions declared for each feature. Then following steps carried out to compute and convert features to WEKA supported ARFF format.

1. Used MATHLAB to execute *mirfeatures*function to compute features of 800 music samples with 30 seconds length. Average processing time for music sample vary between 120s – 150s.
2. MIRToolbox*mirexport*MATLAB function to export computed figures from structure array from MATHLAB to tab delimited text file format.
3. Microsoft Excel Spread Sheet application to convert tab delimited output of *mirexport* to CSV (comma separated value) format.
4. WEKA ArffViewer option to convert CSV format to ARFF format.

---

[1] https://www.premiumbeat.com/blog/when-to-use-wav-files-when-to-use-mp3-files-what-is-the-difference-between-the-two-formats/
[2] http://www.audioanimals.co.uk/news/why-wav-is-better-than-mp3

JAudio supported computing and extracting features directly to the ARFF format. Then WEKA Simple CLI command line interface used to merge computed features from both MIRToolbox and jAudio to together. "java weka.core.Instances merge <File1><File2>><Output File> command issued through Simple CLI to combine above two ARFF files with different number of instances and same number of attributes (Name and type should match). As a summary, 930 features (including feature with feature vectors) extracted from both tool where 890 from the MIRToolbox and 40 from the jAudio.

Following table illustrate the summary of extracted features and related tool used to extract the feature (Table 02).

Tool	Feature Category	Main Feature	Sub Feature	MIR Command	Sub Features / Feature Vector	Noof Features
M	Dynamics	RMS		*Mirrms**	Mean,Std,Slope,PeriodFreq,PeriodAmp,PeriodEntropy	06
M	Fluctuation	Peak		*mirfluctuation mirpeak*	PeakPosMean,PeakMagMean	02
M	Fluctuation	Centroid		*mircentroid*	Mean	01
M	Rhythm	Tempo		*mirtempo**	Mean,Std,Slope,PeriodFreq,PeriodAmp,PeriodEntropy	06
M	Rhythm	Envelope Autocorrelation	PeakPos	*mirtempo**	Mean,Std,Slope,PeriodFreq,PeriodAmp,PeriodEntropy	06
			PeakMag	*mirtempo**	Mean,Std,Slope,PeriodFreq,PeriodAmp,PeriodEntropy	06
M	Rhythm	Attack Time		*mirattacktim*	Mean,Std,Slope,PeriodFreq,PeriodAmp,PeriodEntropy	06
M	Rhythm	Onset Curve (Envelope) - Time		*mirevents with envelope curve*	PeakPosMean,PeakMagMean	02
M	Rhythm	Attack Slope		*mirattackslope*	Mean,Std,Slope,PeriodFreq,PeriodAmp,PeriodEntropy	06
M	Rhythm	Onset Curve (Envelope) - Slope		*mirattackslope*	PeakPosMean,PeakMagMean	02
M	Spectral	Spectral Centroid		*Mircentroid**	Mean,Std,Slope,PeriodFreq,PeriodAmp,PeriodEntropy	06
M	Spectral	Brightness		*Mirbrightness**	Mean,Std,Slope,PeriodFreq,PeriodAmp,PeriodEntropy	06
M	Spectral	Spectral		*Mirspread**	Mean,Std,Slope,PeriodFreq,Pe	06

		Spread			riodAmp,PeriodEntropy	
M	Spectral	Spectral Skewness		*Mirskewness**	Mean,Std,Slope,PeriodFreq,PeriodAmp,PeriodEntropy	06
M	Spectral	Spectral Kurtosis		*Mirkurtosis**	Mean,Std,Slope,PeriodFreq,PeriodAmp,PeriodEntropy	06
M	Spectral	Rolloff (Threshold 95)		*mirroloff** *95 % threshold*	Mean,Std,Slope,PeriodFreq,PeriodAmp,PeriodEntropy	06
M	Spectral	Rolloff (Threshold 85)		*mirroloff** *85 % threshold*	Mean,Std,Slope,PeriodFreq,PeriodAmp,PeriodEntropy	06
M	Spectral	Entropy of Spectrum		*Mirentropy**	Mean,Std,Slope,PeriodFreq,PeriodAmp,PeriodEntropy	06
M	Spectral	Spectral Flatness		*mirflatness**	Mean,Std,Slope,PeriodFreq,PeriodAmp,PeriodEntropy	06
M	Spectral	Roughness (1)		*Mirroughness**	Mean,Std,Slope,PeriodFreq,PeriodAmp,PeriodEntropy	06
M	Spectral	Spectrum (Roughness 2)	PeakPos	*Mirroughness** *Peaks used for estimation*	Mean,Std,Slope,PeriodFreq,PeriodAmp,PeriodEntropy	06
			PeakMag		Mean,Std,Slope,PeriodFreq,PeriodAmp,PeriodEntropy	06
M	Spectral	Spectral Irregularity (Irregularity 1)		*Mirregularity**	Mean,Std,Slope,PeriodFreq,PeriodAmp,PeriodEntropy	06
M	Spectral	Spectrum (Irregularity 2)	PeakPos	*Mirregularity** *Peaks used for estimation*	Mean,Std,Slope,PeriodFreq,PeriodAmp,PeriodEntropy	06
			PeakMag		Mean,Std,Slope,PeriodFreq,PeriodAmp,PeriodEntropy	06
M	Spectral	MFCC		*Mirmfcc**	Mean (1-13) Feature Vector	13
M	Spectral	MFCC		*Mirmfcc**	Std (1-13) Feature Vector	13
M	Spectral	MFCC		*Mirmfcc**	Slope (1-13) Feature Vector	13
M	Spectral	MFCC		*Mirmfcc**	PeriodFreq (1-13) Feature Vector	13
M	Spectral	MFCC		*Mirmfcc**	PeriodAmp (1-13) Feature Vector	13
M	Spectral	MFCC		*Mirmfcc**	PeriodEntropy (1-13) Feature Vector	13
M	Spectral	MFCC	Mel-Spectrum		Mean (1-40) Feature Vector	40
M	Spectral	MFCC	Mel-Spectrum		Std (1-40) Feature Vector	40
M	Spectral	MFCC	Mel-Spectrum		Slope (1-40) Feature Vector	40
M	Spectral	MFCC	Mel-Spectrum		PeriodFreq (1-40) Feature Vector	40
M	Spectral	MFCC	Mel-		PeriodAmp (1-40) Feature	40

			Spectrum		Vector	
M	Spectral	MFCC	Mel-Spectrum		PeriodEntropy (1-40) Feature Vector	40
M	Spectral	Delta-MFCC (1)		*delta-MFCC**	Mean (1-13) Feature Vector	13
					Std (1-13) Feature Vector	13
					Slope (1-13) Feature Vector	13
					PeriodFreq (1-13) Feature Vector	13
					PeriodAmp (1-13) Feature Vector	13
					PeriodEntropy (1-13) Feature Vector	13
M	Spectral	Delta-MFCC (2)		*delta-MFCC**	Mean (1-13) Feature Vector	13
					Std (1-13) Feature Vector	13
					Slope (1-13) Feature Vector	13
					PeriodFreq (1-13) Feature Vector	13
					PeriodAmp (1-13) Feature Vector	13
					PeriodEntropy (1-13) Feature Vector	13
M	Spectral	Delta-Delta-MFCC (1)		*delta-delta-MFCCs**	Mean (1-13) Feature Vector	13
					Std (1-13) Feature Vector	13
					Slope (1-13) Feature Vector	13
					PeriodFreq (1-13) Feature Vector	13
					PeriodAmp (1-13) Feature Vector	13
					PeriodEntropy (1-13) Feature Vector	13
M	Spectral	Delta-Delta-MFCC (2)		*delta-delta-MFCCs**	Mean (1-13) Feature Vector	13
					Std (1-13) Feature Vector	13
					Slope (1-13) Feature Vector	13
					PeriodFreq (1-13) Feature Vector	13
					PeriodAmp (1-13) Feature Vector	13
					PeriodEntropy (1-13) Feature Vector	13
M	Timbre	Zero-Crossing Rate (ZCR)		*mirzerocross**	Mean,Std,Slope,PeriodFreq,PeriodAmp,PeriodEntropy	06
M	Timbre	Lowenergy		*mirlowenergy*	Mean	01

M	Timbre	RMS Energy			Mean,Std,Slope,PeriodFreq,PeriodAmp,PeriodEntropy	06
M	Timbre	Spectral Flux		*mirflux**	Mean,Std,Slope,PeriodFreq,PeriodAmp,PeriodEntropy	06
M	Tonal	Peak Chromagram	PeakPos	*mirchromagram*	Mean,Std,Slope,PeriodFreq,PeriodAmp,PeriodEntropy	06
			PeakMag		Mean,Std,Slope,PeriodFreq,PeriodAmp,PeriodEntropy	06
M	Tonal	Centroid of Chromagram		*mirchromagram*	Mean,Std,Slope,PeriodFreq,PeriodAmp,PeriodEntropy	06
M	Tonal	Key Clarity		*Mirkey** *Second output*	Mean,Std,Slope,PeriodFreq,PeriodAmp,PeriodEntropy	06
M	Tonal	Mode		*Mirmode**	Mean,Std,Slope,PeriodFreq,PeriodAmp,PeriodEntropy	06
M	Tonal	Key Strength		*Mirkey**	Mean (1-12) Feature Vector	12
					Std (1-12) Feature Vector	12
					Slope (1-12) Feature Vector	12
					PeriodFreq (1-12) Feature Vector	12
					PeriodAmp (1-12) Feature Vector	12
					PeriodEntropy (1-12) Feature Vector	12
M	Tonal	Harmonic Change Detection Function		*Mirhcdf**	Mean,Std,Slope,PeriodFreq,PeriodAmp,PeriodEntropy	06
J	Spectral	Compactness			Std,Avg	02
J	Spectral	Spectral Variability			Std,Avg	02
J	Spectral	Strongest Beat			Std,Avg	02
J	Spectral	Beat Sum			Std,Avg	02
J	Spectral	Strength of Strongest Beat			Std,Avg	02
J	Spectral	LPC			Std (0-9), Avg (0-9)	20
J	Spectral	Method of Moments			Std (0-4), Avg (0-4)	10
						930

MIRToolBox=M, jAudio=J, *=Frame based

Table 02: Features Extracted from MIR Toolbox and jAudio

3.4 Feature Selection

"Any classification system is only as good as the features that it receives" [17], not all features selected for extraction may not be relevance to classification algorithm to make the prediction. Include all extracted features without selecting the best feature set for classification may lead to take high computation overhead, time complexity etc. Therefore, best decision is to check the relevance of each selected feature to participate for classification algorithm.

When analyze the original dataset all most all values are decimal figures which may not work well with some feature selection and classification algorithms. Also, there is a possibility of improving the performance when rescale the attributes in the dataset. Then, prepared two additional datasets by applying following filters through WEKA:

Discretize Filter: supervised attribute filter which converts numeric attributes in the dataset in to nominal attributes

Normalize Filter: unsupervised attribute filter which normalizes all numeric attributes in the dataset in to integer values in between 0 and 1 as default.

Used WEKA attribute selection option to get the feature subsets under three datasets (original, normalized and discretized) using attribute and wrapper feature selection algorithms.ClassifierAttributeEval(CAE), CorrelationAttributeEval(COR), GainRatioAttributeEval(GRA), InfoGainAttributeEval(ING), OneRAttributeEval(ONR), PrincipalComponents(PCA), ReliefFAttributeEval(RAE), SymmetricalUncertAttributeEval(SYM) attribute filter evaluators with *Ranker* search method and CfsSubsetEval(CSE), ClassifierSubsetEval(CSE), WrapperSubsetEval(WSE) subset evaluators with *GreedyStepwise* and *BestFirst* have been used to identify the feature subsets under each dataset. Following table illustrate the number of features selected from each evaluator under three datasets.

#	Type	Attribute Evaluator	Search Method	Classifier	Threshold	Original Full	Original 10-Fold	Discretize Full	Discretize 10-Fold	Normalized Full	Normalized 10-Fold
1	A	CAE	Ranker	Naive Bayes	0.15	42		76		43	43
2	A	CAE	Ranker	J48	0.15	45		77		48	42
3	A	CAE	Ranker	IBK	0.15	0*	0*	78	74	0**	0**
4	A	CAE	Ranker	LibSVM	0.15	37	35	79	77	44	47
5	A	COR	Ranker	-	0.15	248		235		248	249
6	A	GRA	Ranker	-	0.15	78	87	79	81	78	87
7	A	ING	Ranker	-	0.15	149	146	146	148	149	146
8	A	ONR	Ranker	OneR	0.15	0*	0*	930	930	0*	0*
9	A	PCA	Ranker		0.15	99	N/S	181	N/S	99	N/S
10	A	SYM	Ranker		0.15	14	19	14	14	14	15
11	A	RAE	Ranker	-	0.15	0*	0*	11	11	0*	0*
12	W	CSE	GreedyStepwise		0.15	119		112		119	99***
14	W	CSE	BestFirst (Forward)		0.15	118		111		118	99***
16	W	CSE	BestFirst (Forward)	Naive Bayes	0.15	14		20		32	10***
18	W	CSE	BestFirst (Forward)	J48	0.15	26		36		22	1***
20	W	CSE	GreedyStepwise	Naive Bayes	0.15	12		20		12	
22	W	CSE	GreedyStepwise	J48	0.15	25		34		13	0***
24	W	WSE	BestFirst (Forward)	Naive Bayes	0.15	10		14		18	

26	W	WSE	BestFirst (Forward)	J48	0.15	IM		16		17	
28	W	WSE	GreedyStepwise	Naive Bayes	0.15	10		14		18	
30	W	WSE	GreedyStepwise	J48	0.15	IM		16		14	

**	Average merit is below 0.15
*	No attributes found to work with! Or no features ranked to threshold limit
***	Considered features that are count upwards 5 folds (Over 50%)
IM	Insufficient Memory
A	Attribute
W	Wrapper (Subset)

Table 03: Number of Features Selected with Each Evaluator Under Three Datasets

During the feature selection process observed that evaluators with given classifier took average 10-40 minutes to process while evaluators with given classifier and 10-fold cross validation took average 30-150 minutes to process. Also, wrapper subset evaluators took considerably more time than the attribute evaluators.

3.4 Classification

Initial three datasets (Original, Discretize and Normalized) and all feature subsets extracted through feature selection process are taken to evaluate with eight classification algorithms considered through WEKA. These algorithms represent the major classification algorithms categories defined in WEKA. All algorithms run initially with their default parameters and configurations. Below listed the algorithms considered for classification (Table 04).

	Algorithm Category	**Algorithm**
01	Bayes	**Naïve Bayes.** Naive Bayes classifier using estimator classes. [WEKA]
02	Functions	**Lib SVM.** A wrapper class for the libsvm library. This wrapper supports the classifiers implemented in the libsvm library, including one-class SVMs [WEKA]
03	Functions	**SMO**. Algorithm for training a support vector classifier. This implementation globally replaces all missing values and transforms nominal attributes into binary ones. [WEKA]

04	Lazy	**IBk**. K-nearest neighbours classifier. [WEKA]
05	Meta	**AdaBoostM1**. Boosting a nominal class classifier using the Adaboost M1 method. [WEKA]
06	Meta	**Bagging**. Class for bagging a classifier to reduce variance. [WEKA]
07	Trees	**J48.** Class for generating a pruned or unpruned C4.[WEKA]
08	Trees	**Random Forest**. Class for constructing a forest of random trees.[WEKA]

Table 04: Algorithms Selected for Classification

Naming convention used to identify the datasets are described as follows:

01DA_CLA_RNK_NB_76

A B C D E F G

A	Dataset group number
B	Dataset type **D**=Discretized Attributes; **N**=Normalized Attributes; **O**=Original Attributes
C	Attribute selection algorithm category **A**=Attribute Selection Algorithms **W**=Wrapper/Subset Selection Algorithms
D	Attribute selection algorithm **CLA** = ClassifierAttributeEval **COR** = CorrelationAttributeEval **GNR** = GainRatioAttributeEval **IFG** = InfoGainAttributeEval **ONR** = OneRAttributeEval **RFL** = ReliefFAttributeEval **PCA** = PrincipalComponents **SYM** = SymmetricalUncertAttributeEval **CFS** = CfsSubsetEval **CLA** = ClassifierSubsetEval

	WRP = WrapperSubsetEval
E	Search Algorithm **RNK** = Ranker **GRD** = GreedyStepwise **BFF** = BestFirst (Forward)
F	Classification Algorithm **NB** = Naive Bayes **J48** **IBK** = K-nearest neighbor **SVM** = Support Vector Machine **ONR** = OneR **NOC** = No Classification Used
G	Number of attributes available in the dataset based on the attribute selection

Execution and comparison of classification results performed with following four steps

1. 8 classification algorithms with 59 datasets with missing values and 66% training split.
2. 8 classification algorithms with 59 datasets with missing values and 80% training split.
3. 8 classification algorithms with 59 datasets without missing values and 66% training split.
4. 8 classification algorithms with 59 datasets without missing values and 66% training split

Following Table 05 (Train percentage of 66%) illustrate the predication accuracy of each selected algorithm under 59 datasets with missing values. WEKA *Experimenter* tool used to upload the required datasets, classification algorithms and compare the prediction accuracy.

	Dataset	**bayes. NaiveBayes**	**functions. LibSVM**	**functions. SMO**	**lazy. IBk**	**meta. AdaBoostM1**	**meta. Bagging**	**trees. J48**	**trees. RandomForest**
1	01DA_CLA_RNK_NB_76	63.93	68.09	63.24	61.69	65.07	70	59.67	68.93
2	01NA_CLA_RNK_NB_43	62.13	62.79	68.12	60.26	63.6	68.38	58.35	68.49
3	01OA_CLA_RNK_NB_42	62.1	25.66	68.05	61.18	64.52	68.57	57.94	67.46
4	02DA_CLA_RNK_J48_77	64.01	67.83	64.04	62.61	64.82	69.82	59.6	69.82
5	02NA_CLA_RNK_J48_48	60.81	61.47	68.38	56.03	63.57	67.13	59.08	66.14
6	02OA_CLA_RNK_J48_45	60.96	27.98	68.68	56.69	63.79	67.5	57.1	66.8
7	03DA_CLA_RNK_IBK_78	64.04	68.24	63.24	61.18	65.92	69.63	59.6	69.6
8	04DA_CLA_RNK_SVM_79	64.08	67.98	63.09	61.07	64.96	69.67	59.71	69.49

9	04NA_CLA_RNK_SVM_44	61.73	63.12	69.08	62.32	65.55	68.46	59.01	69.12
10	04OA_CLA_RNK_SVM_37	54.04	44.63	66.95	58.35	61.84	66.58	55.85	66.8
11	05DA_COR_RNK_NOC_235	68.35	73.05	71.1	66.32	67.87	71.4	58.93	71.14
12	05NA_COR_RNK_NOC_248	64.12	65.74	69.96	60.44	64.67	68.57	56.1	69.19
13	05OA_COR_RNK_NOC_248	64.26	24.52	69.96	60.44	65.63	68.71	55.92	69.26
14	06DA_GNR_RNK_NOC_79	64.08	66.51	65.66	62.32	66.1	66.65	60.11	66.88
15	06NA_GNR_RNK_NOC_78	59.96	58.53	63.38	58.79	62.13	65.66	55.33	66.29
16	06OA_GNR_RNK_NOC_78	60	26.91	63.31	58.79	62.21	65.59	55.33	65.26
17	07DA_IFG_RNK_NOC_146	65.04	70.33	66.84	64.49	67.46	69.63	59.3	69.85
18	07NA_IFG_RNK_NOC_149	61.07	61.99	69.37	61.95	66.25	68.42	57.61	68.53
19	07OA_IFG_RNK_NOC_149	61.18	24.52	69.34	61.95	65.26	68.46	58.01	67.83
20	08DA_ONR_RNK_ONR_930	70.77	72.98	76.4	69.6	68.86	71.88	57.76	72.02
21	09DA_PCA_RNK_NOC_181	69.3	74.41	67.87	32.5	65.04	71.43	56.47	63.57
22	09NA_PCA_RNK_NOC_99	56.1	65.96	60.29	28.64	59.45	66.88	51.14	61.43
23	09OA_PCA_RNK_NOC_99	56.14	65.96	60.29	28.64	59.78	67.13	51.14	60.99
24	10DA_SYM_RNK_NOC_14	61.07	61.47	60.7	55.11	56.76	58.6	55.74	57.72
25	10NA_SYM_RNK_NOC_14	56.8	56.76	58.31	53.35	56.91	61.14	53.31	60.11
26	10OA_SYM_RNK_NOC_14	56.91	41.21	58.35	53.35	56.69	61.03	53.27	60.63
27	11DA_RLF_RNK_NOC_11	66.07	65.59	64.15	59.67	59.89	62.98	60.11	62.13
28	12DW_CFS_GRD_NOC_112	76.62	77.24	71.29	70.44	69.38	74.12	59.23	73.46
29	12NW_CFS_GRD_NOC_119	67.72	69.78	73.05	63.42	68.27	71.51	58.27	71.47
30	12OW_CFS_GRD_NOC_119	67.83	24.52	72.98	63.42	67.65	71.58	57.13	71.88
31	13DW_CFS_BFF_NOC_111	76.58	77.21	71.73	70.7	68.93	74.12	59.26	73.9
32	13NW_CFS_BFF_NOC_118	67.61	69.82	72.65	63.31	67.68	71.4	58.49	71.47
33	13OW_CFS_BFF_NOC_118	67.76	24.52	72.57	63.31	67.46	71.95	57.39	71.88
34	14DW_CLA_BFF_NB_20	75.74	71.29	72.13	64.15	63.46	71.58	61.99	70.15
35	14NW_CLA_BFF_NB_32	68.38	60.66	64.23	25.18	61.47	66.87	55.74	66.69
36	14OW_CLA_BFF_NB_14	66.62	39.67	64.82	32.57	60.55	67.35	56.29	66.91
37	15DW_CLA_BFF_J48_36	67.72	68.79	65.22	59.04	64.23	69.63	59.85	69.3
38	15NW_CLA_BFF_J48_22	54.34	51.18	58.24	44.3	57.43	61.32	50.4	60.66
39	15OW_CLA_BFF_J48_26	52.76	29.56	59.89	43.82	56.36	62.76	49.38	61.76
40	16DW_CLA_GRD_NB_20	75.74	71.29	72.13	64.15	63.46	71.58	61.99	70.15
41	16NW_CLA_GRD_NB_12	67.87	61.54	63.49	30.37	59.3	65.92	56.32	65.29
42	16OW_CLA_GRD_NB_12	67.13	40.44	65.63	32.5	60.74	67.13	57.1	66.69
43	17DW_CLA_GRD_J48_34	67.5	68.86	66.95	60.37	63.86	69.67	60	68.6
44	17NW_CLA_GRD_J48_13	54.34	53.75	59.04	46.99	56.03	61.76	51.91	61.32
45	17OW_CLA_GRD_J48_25	53.49	29.67	59.71	45.04	57.13	62.28	50.04	62.13
46	18DW_WRP_BFF_NB_14	73.42	69.3	70.04	61.91	61.91	66.29	61.51	65.11
47	18NW_WRP_BFF_NB_18	70.7	65.92	66.65	54.01	62.76	67.43	54.82	67.06
48	18OW_WRP_BFF_NB_10	67.21	29.63	67.1	52.02	61.84	66.95	58.35	66.8
49	19DW_WRP_BFF_J48_16	67.13	66.58	68.64	55.33	61.36	63.16	65.77	61.76
50	19NW_WRP_BFF_J48_17	63.82	63.97	65.59	36.21	64.08	68.35	62.76	68.9
51	20DW_WRP_GRD_NB_14	73.42	69.3	70.04	61.91	61.91	66.29	61.51	65.11
52	20NW_WRP_GRD_NB_18	70.7	65.92	66.65	54.01	62.76	67.43	54.82	67.06
53	20OW_WRP_GRD_NB_10	67.21	29.63	67.1	52.02	61.84	66.95	58.35	66.8
54	21DW_WRP_GRD_J48_16	67.13	66.58	68.64	55.33	61.36	63.16	65.77	61.76
55	21NW_WRP_GRD_J48_14	63.93	63.57	65.29	33.46	64.63	67.72	62.32	67.94
56	PrincipalComponents_930	56.14	65.96	60.29	28.64	59.78	67.13	51.14	60.99

57	Original_930	63.27	24.52	67.35	56.95	65.07	69.04	54.63	69.6
58	Discretized_930	70.77	72.5	76.36	69.6	68.24	71.69	57.32	71.47
59	Normalized_930	63.42	64.71	67.35	56.95	64.85	69.04	55.11	68.97
	Average	**64.63**	**56.65**	**66.63**	**54.39**	**63.23**	**67.68**	**57.40**	**66.92**
	Max	**76.62**	**77.24**	**76.4**	**70.7**	**69.38**	**74.12**	**65.77**	**73.9**
	Min	**52.76**	**24.52**	**58.24**	**25.18**	**56.03**	**58.6**	**49.38**	**57.72**
	Standard Deviation	**5.95**	**17.13**	**4.49**	**12.19**	**3.50**	**3.37**	**3.57**	**3.81**

Table 05: Predication Results of 8 Classification Algorithms with 59 Datasets with missing values

According to the results, best accuracy 77.24 is reported by LibSVM classification algorithm with discretized dataset of 112 attributes which is extracted by wrapper CfsSubsetEval feature selection algorithm with GreedyStepwise search algorithm. Also, only NaiveBayes, LibSVM and SMO provided the accuracy more than 75% with discretized datasets. Maximum average accuracy of 67.68 and minimum standard deviation 3.37 is given by the Bagging classifier.

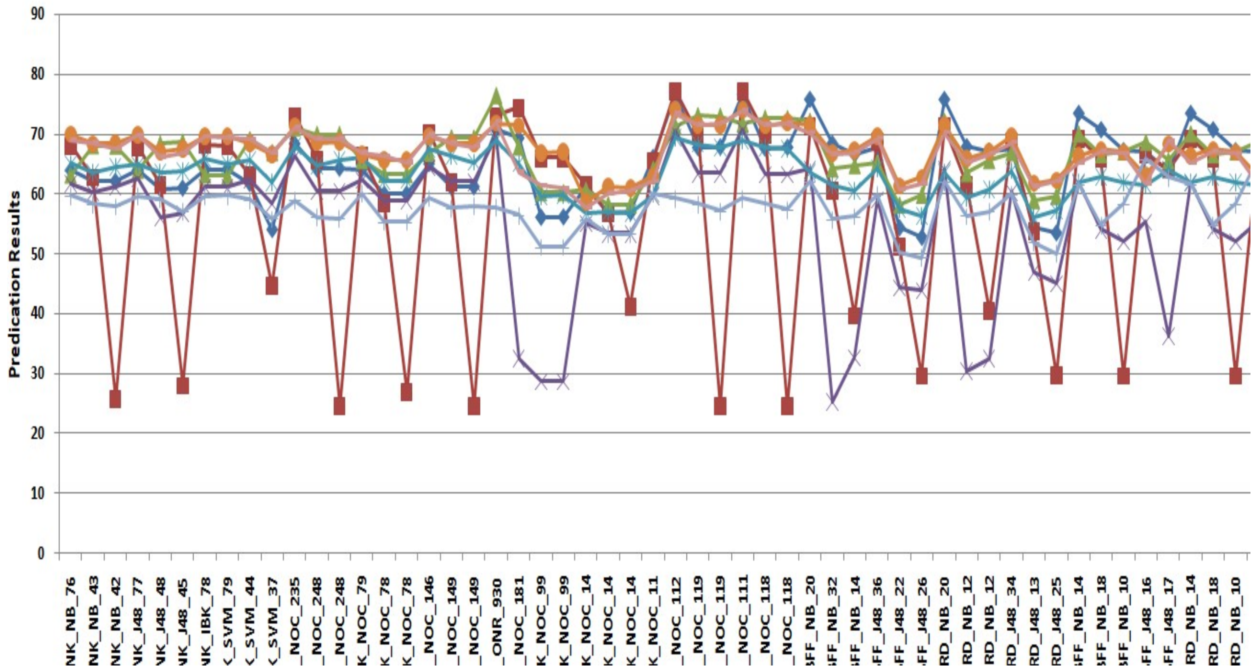


Figure D.02: Graphical view of predication results of 8 classification algorithms with 59 datasets

Following Table 06 (Train percentage of 80%) illustrate the predication accuracy of each selected algorithm under 59 datasets with missing values. WEKA *Experimenter* tool used to upload the required datasets, classification algorithms and compare the prediction accuracy.

	Dataset	**bayes. NaiveBayes**	**functions. LibSVM**	**functions. SMO**	**lazy. IBk**	**meta. AdaBoostM1**	**meta. Bagging**	**trees. J48**	**trees. RandomForest**
1	01NA_CLA_RNK_NB_43	62.88	65	70.63	61.06	63.94	70.44	57.44	69.75
2	01OA_CLA_RNK_NB_42	63.56	25.75	70.06	61.5	65.44	70.19	58.38	69.88
3	02DA_CLA_RNK_J48_77	64.94	69.63	65.31	62.81	65.25	71.25	62.5	70.5
4	02NA_CLA_RNK_J48_48	63.19	63.88	69.25	57.69	65.06	68.5	57.56	67.81
5	02OA_CLA_RNK_J48_45	63.38	28.75	70.25	59.19	63.63	69.06	57.69	68.13
6	03DA_CLA_RNK_IBK_78	65.44	70.06	65.19	61.5	66.5	71.25	62.31	72.06
7	04DA_CLA_RNK_SVM_79	65.13	69.94	64.75	61.38	65.94	71.69	62.38	71.06
8	04NA_CLA_RNK_SVM_44	63.5	64.94	71.5	61.88	66	71.19	60.5	71.06
9	04OA_CLA_RNK_SVM_37	54.69	46	68.88	58.5	62.63	67.69	57.25	67.31
10	05DA_COR_RNK_NOC_235	69.44	74.63	70.31	66.25	69.25	73.06	61.38	73.38
11	05NA_COR_RNK_NOC_248	65.63	68.25	71.25	61.75	67.38	70.19	56.88	70.75
12	05OA_COR_RNK_NOC_248	65.63	24.31	71.25	61.75	66.94	70	57.31	70.56
13	06DA_GNR_RNK_NOC_79	66.13	68.5	64.94	62.5	66.25	68.69	61.44	68.75
14	06NA_GNR_RNK_NOC_78	60.88	61.81	64.69	59.44	63.88	68.81	56	67.94
15	06OA_GNR_RNK_NOC_78	60.88	26.63	64.62	59.44	65.63	68.63	56.56	68.5
16	07DA_IFG_RNK_NOC_146	66.75	71.75	66.88	64.94	67.88	72.06	60.75	71.75
17	07NA_IFG_RNK_NOC_149	62.94	63.94	71.31	62	66.44	70.88	57.19	71.19
18	07OA_IFG_RNK_NOC_149	62.88	24.31	71.31	62	66.81	70.75	56.94	70.25
19	08DA_ONR_RNK_ONR_930	72	74.38	77.31	69.94	67.69	72.69	58.13	72.94
20	09DA_PCA_RNK_NOC_181	69.69	76.06	68.63	33.06	65.63	71.81	58.25	67.38
21	09NA_PCA_RNK_NOC_99	55.5	67.5	62.44	27.44	60.19	67.19	50.88	61.56
22	09OA_PCA_RNK_NOC_99	55.5	67.5	62.31	27.44	58.88	68.31	50.81	62
23	10DA_SYM_RNK_NOC_14	61.56	62.38	61.44	57	58.81	59.81	56.88	59.5
24	10NA_SYM_RNK_NOC_14	56.63	58.25	59.69	52.63	58	62.81	55.25	60.5
25	10OA_SYM_RNK_NOC_14	56.75	42.06	59.56	52.63	57.69	63.56	55.25	62
26	11DA_RLF_RNK_NOC_11	66.5	65.5	64.88	60.38	61.81	63.63	60.25	63.06
27	12DW_CFS_GRD_NOC_112	76.81	77.56	70.81	70.19	70.81	75.56	62.75	75.13
28	12NW_CFS_GRD_NOC_119	67.94	71.44	73.81	63.31	68.94	72.88	60.75	72.63
29	12OW_CFS_GRD_NOC_119	67.81	24.31	73.88	63.31	66.56	73.19	59.56	73.63
30	13DW_CFS_BFF_NOC_111	76.88	77.5	71.25	71.06	70.81	74.69	61.88	75.81
31	13NW_CFS_BFF_NOC_118	67.88	71.19	74.38	63.38	69.63	73.06	60.63	74.31
32	13OW_CFS_BFF_NOC_118	67.75	24.31	74.38	63.38	67.38	72.94	59.25	72.56
33	14DW_CLA_BFF_NB_20	76.63	73.06	73.81	65.63	65.38	73.44	63.38	71.94
34	14NW_CLA_BFF_NB_32	69.75	63.19	65	25.88	60.94	67.25	55.31	66.25
35	14OW_CLA_BFF_NB_14	68.44	39.81	65.44	33.19	62	68.44	58.44	68.63
36	15DW_CLA_BFF_J48_36	68.44	69.88	69.19	60.19	64.81	70.19	60.38	70.63

37	15NW_CLA_BFF_J48_22	54.31	53.63	61.19	44.38	57.25	62.75	51.44	61.88
38	15OW_CLA_BFF_J48_26	52.19	27.94	61.13	42.5	56.19	64.25	50.25	62.63
39	16DW_CLA_GRD_NB_20	76.63	73.06	73.81	65.63	65.38	73.44	63.38	71.94
40	16NW_CLA_GRD_NB_12	70.31	64	65.06	31.69	60.63	66.56	57.19	65.81
41	16OW_CLA_GRD_NB_12	68.5	40.94	66.13	33.63	60.75	68.44	58.75	68.44
42	17DW_CLA_GRD_J48_34	68.19	70.06	69.75	62.38	64.75	69.88	61.31	70.19
43	17NW_CLA_GRD_J48_13	55	53.94	60.19	48.56	58.44	63.25	54.06	62.81
44	17OW_CLA_GRD_J48_25	53.38	27.94	61.69	44.5	57.56	64.19	49.81	62.88
45	18DW_WRP_BFF_NB_14	75.31	70.75	71.06	62.63	62.13	67.88	63.25	66.88
46	18NW_WRP_BFF_NB_18	71.56	67.56	67.25	53.13	63.88	67.94	54.5	67
47	18OW_WRP_BFF_NB_10	68.81	28.19	67.75	52.81	62.13	67.63	60.44	67.69
48	19DW_WRP_BFF_J48_16	68.13	67.56	69.75	56.38	62	63.19	66.81	62.56
49	19NW_WRP_BFF_J48_17	65.75	65.44	65.88	35.31	66.31	69.75	63.56	69.44
50	20DW_WRP_GRD_NB_14	75.31	70.75	71.06	62.63	62.13	67.88	63.25	66.88
51	20NW_WRP_GRD_NB_18	71.56	67.56	67.25	53.13	63.88	67.94	54.5	67
52	20OW_WRP_GRD_NB_10	68.81	28.19	67.75	52.81	62.13	67.63	60.44	67.69
53	21DW_WRP_GRD_J48_16	68.13	67.56	69.75	56.38	62	63.19	66.81	62.56
54	21NW_WRP_GRD_J48_14	65.75	65.25	66.38	33.31	66.31	69.5	64.94	69
55	PrincipalComponents_930	55.5	67.5	62.31	27.44	58.88	68.31	50.81	62
56	Original_930	62.94	24.31	66.69	56.13	66.19	69.81	55.81	70.25
57	Discretized_930	72	73.69	77.25	69.94	70.25	72.13	58.5	73.19
58	Normalized_930	63.06	67.44	66.63	56.13	65.5	70.69	56.75	70.31
	Average	**65.54**	**57.54**	**67.87**	**54.60**	**64.06**	**69.00**	**58.54**	**68.28**
	Max	**76.88**	**77.56**	**77.31**	**71.06**	**70.81**	**75.56**	**66.81**	**75.81**
	Min	**52.19**	**24.31**	**59.56**	**25.88**	**56.19**	**59.81**	**49.81**	**59.5**
	Standard Deviation	**6.30**	**18.00**	**4.39**	**12.48**	**3.65**	**3.43**	**4.02**	**4.07**

Table 06: Predication Results of 8 Classification Algorithms with 59 Datasets without missing values

According to the above results, best accuracy 77.56 is reported by LibSVM classification algorithm with discretized dataset of 112 attributes which is extracted by wrapper CfsSubsetEval feature selection algorithm with Greedy Stepwise search algorithm. Also, only NaiveBayes, LibSVM, SMO, Bagging and Randomforest classifiers provided the accuracy of more than 75% with discretized datasets. Maximum average accuracy of 69.00 and minimum standard deviation 3.43 is given by the Bagging classifier.

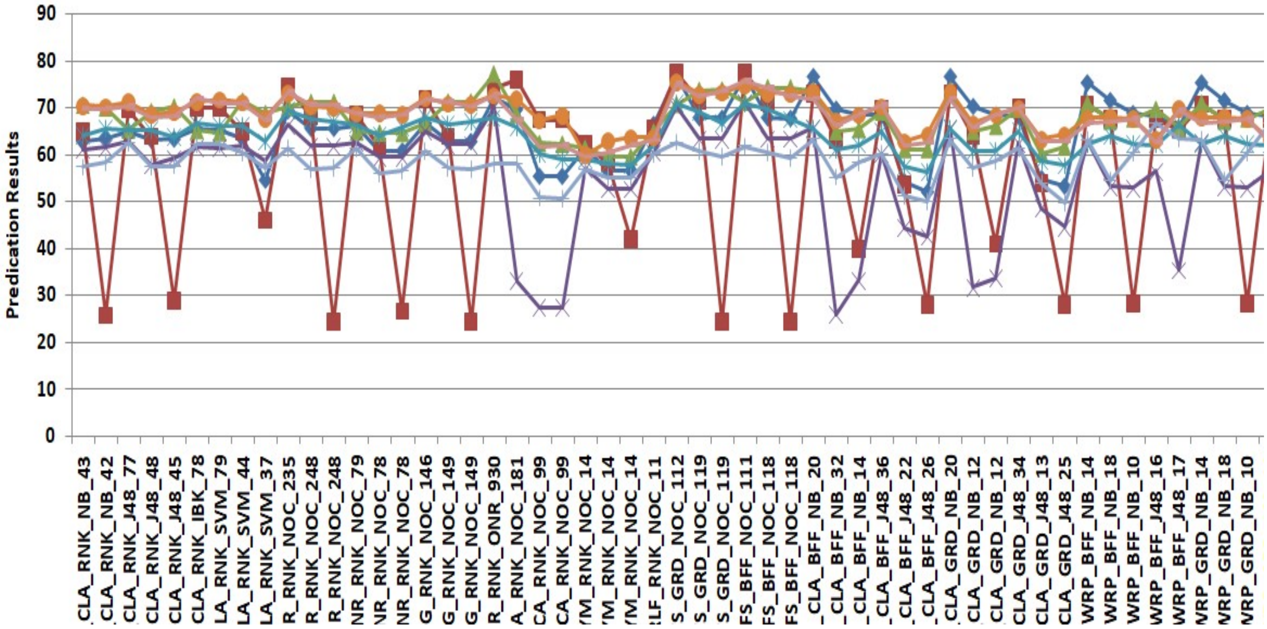


Figure D.03: Graphical view of predication results of 8 classification algorithms with 59 datasets

Following Table-07 illustrate the summary view of classification accuracy obtained over 75% with their relevant dataset name, percentage split used and the prediction accuracy.

Datasets with missing values		NaiveBayes		LibSVM		SMO		IBk		AdaBoostM1		Bagging	
		66%	80%	66%	80%	66%	80%	66%	80%	66%	80%	66%	80%
1	08DA_ONR_RNK_ONR_930					76.4	77.3						
2	09DA_PCA_RNK_NOC_181				76.1								
3	12DW_CFS_GRD_NOC_112	76.62	76.8	77.2	77.6								75.
4	13DW_CFS_BFF_NOC_111	76.58	76.9	77.2	77.5								
5	14DW_CLA_BFF_NB_20	75.74	76.6										
6	16DW_CLA_GRD_NB_20	75.74	76.6										
7	18DW_WRP_BFF_NB_14		75.3										
8	20DW_WRP_GRD_NB_14		75.3										
9	Discretized_930					76.4	77.3						
Datasets without missing values													
1	08DA_ONR_RNK_ONR_930_NM					76.4	77.3						
2	09DA_PCA_RNK_NOC_181_NM				76.1								
3	12DW_CFS_GRD_NOC_112_NM	76.62	77	77.2	77.6								75.
4	13DW_CFS_BFF_NOC_111_NM	76.65	76.9	77.2	77.5								75.
5	14DW_CLA_BFF_NB_20_NM	75.7	76.6										

Table 07: Summary view of classification accuracy obtained over 75%.

According to the above result, observed that only discretized datasets has the predication accuracy over 75% and IBK (K-Nearest Neighbor), J48 and AdaBoostM1 did not predict over 75% under any dataset. (Refer APPENDIX F for predication results of 8 classification algorithms under datasets without missing values).

Below graphs describes the average, maximum, minimum and standard deviation of predicted accuracy of eight algorithms under 66% training splits with and without missing values and 80% training splits with and without missing values for 59 datasets.

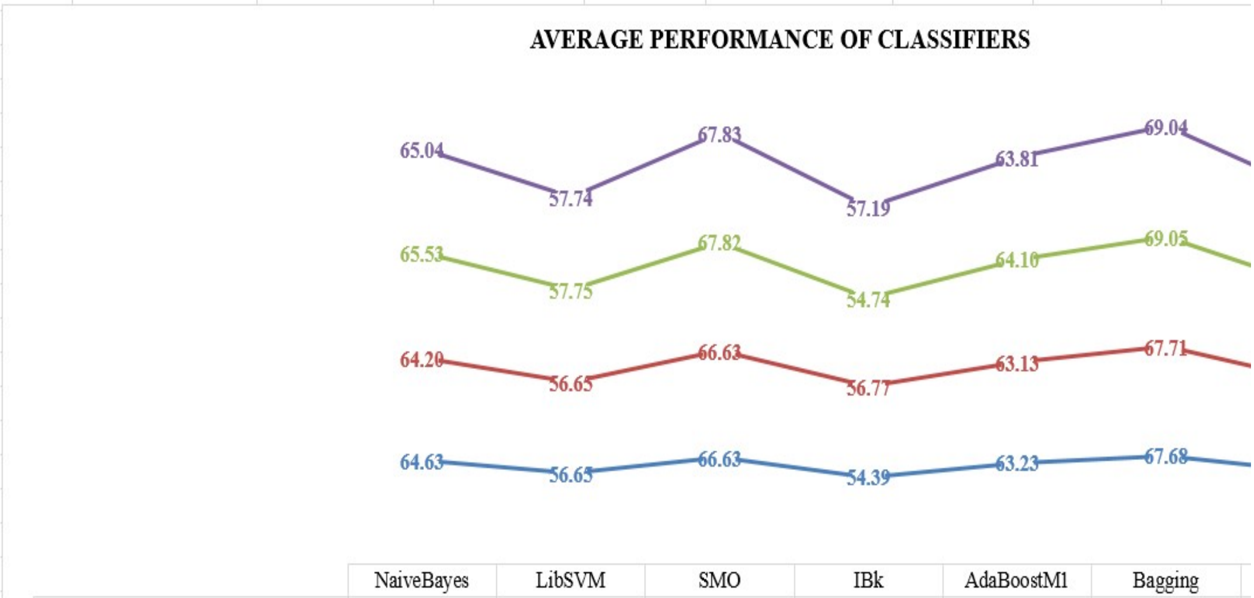


Figure D.04: Average performance of the classifiers

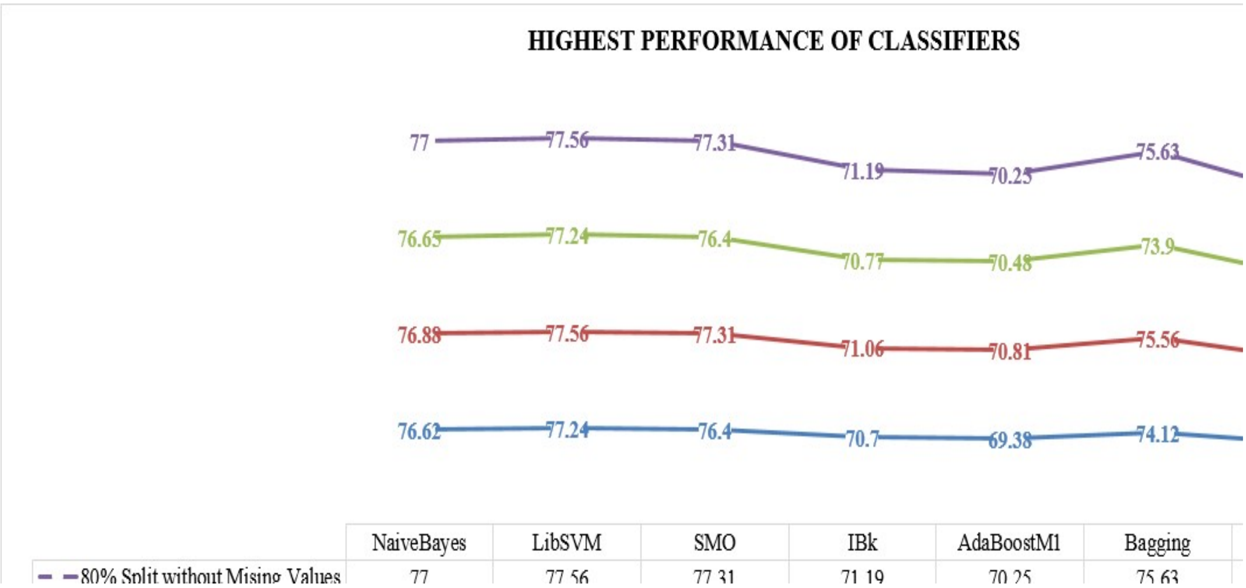


Figure D.05: Highest performance of the classifiers

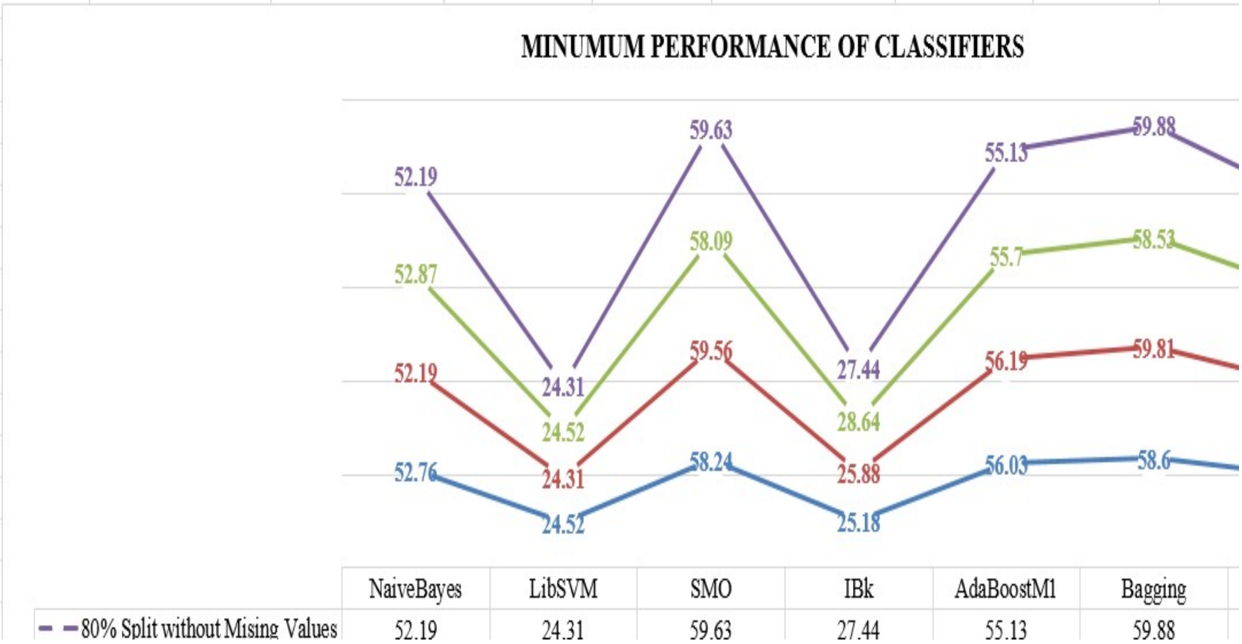


Figure D.06: Minimum performance of the classifiers

Even though LibSVM has recorded highest accuracy (77.56%), it also participates to produce the lowest prediction accuracies (24.31%) among the other algorithms.

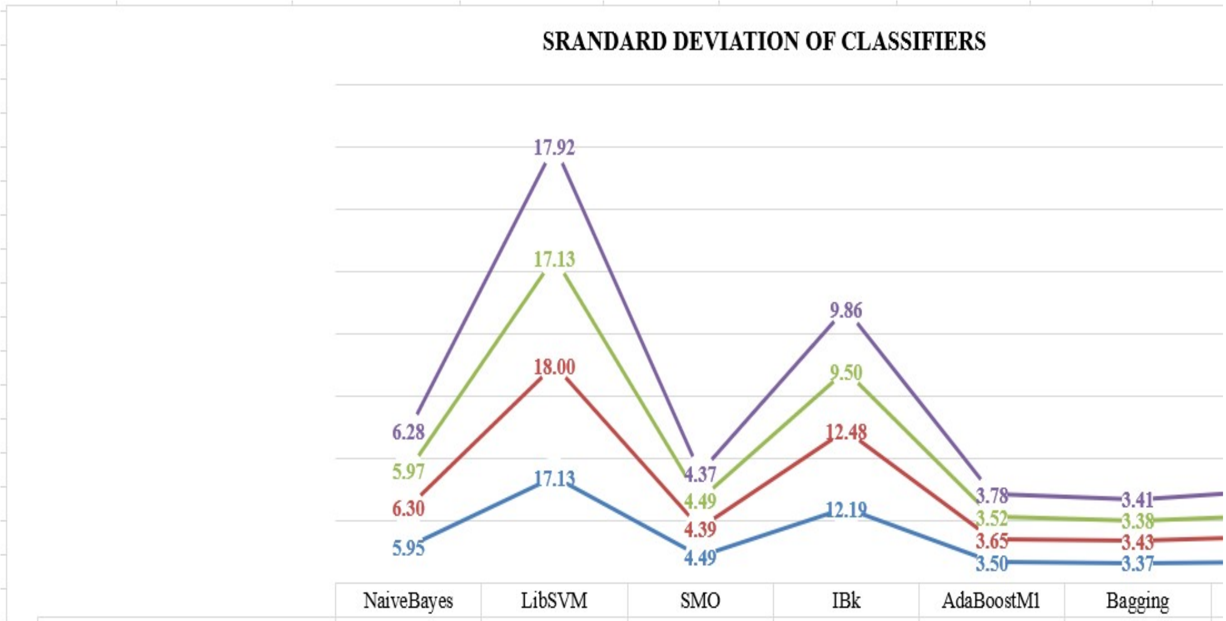


Figure D.07: Standard Deviation of the classifiers

According to the above graphs, LibSVM has highest and lowest predictive accuracy where Bagging has highest average predictive accuracy and the lowest standard deviations of the accuracies.

According to the above analysis, NaiveBayes, LibSVM, SMO, Bagging and RandomForest classification algorithms predicted over 75% accuracy among 18 discretized datasets. Therefore, it is further decided to analyze the 18 datasets with above five algorithms under different parameter settings to check the possibility for improvements.

1. Naïve Bayes

Naïve bayes with 10-fold cross validation performed with its default parameters and changing the *useKernelEstimator* and *useSupervisedDiscretization* parameters. Following table and graph describes the prediction accuracy under different settings. No accuracy reported over 77.56% which is the highest value obtained so far. Highest 3% accuracies are highlighted.

Dataset	NaiveBayes ''	NaiveBayes -K	NaiveBayes -D
08DA_ONR_RNK_ONR_930	⇨ 69.95	⇨ 69.95	⇨ 69.95
08DA_ONR_RNK_ONR_930_NM	⇨ 70.05	⇨ 70.05	⇨ 70.05
09DA_PCA_RNK_NOC_181	⇨ 70.1	⇩ 59.56	⇨ 67.17
09DA_PCA_RNK_NOC_181_NM	⇨ 70.1	⇩ 59.56	⇨ 67.17
12DW_CFS_GRD_NOC_112	⇧ 75.89	⇧ 75.89	⇧ 75.89
12DW_CFS_GRD_NOC_112_NM	⇧ 75.98	⇧ 75.98	⇧ 75.98
13DW_CFS_BFF_NOC_111	⇧ 75.9	⇧ 75.9	⇧ 75.9
13DW_CFS_BFF_NOC_111_NM	⇧ 75.95	⇧ 75.95	⇧ 75.95
14DW_CLA_BFF_NB_20	⇧ 76.77	⇧ 76.77	⇧ 76.77
14DW_CLA_BFF_NB_20_NM	⇧ 76.75	⇧ 76.75	⇧ 76.75
16DW_CLA_GRD_NB_20	⇧ 76.77	⇧ 76.77	⇧ 76.77
16DW_CLA_GRD_NB_20_NM	⇧ 76.75	⇧ 76.75	⇧ 76.75
18DW_WRP_BFF_NB_14	⇧ 73.74	⇧ 73.74	⇧ 73.74
18DW_WRP_BFF_NB_14_NM	⇧ 73.75	⇧ 73.75	⇧ 73.75
20DW_WRP_GRD_NB_14	⇧ 73.74	⇧ 73.74	⇧ 73.74
20DW_WRP_GRD_NB_14_NM	⇧ 73.75	⇧ 73.75	⇧ 73.75
Discretized_930	⇨ 69.95	⇨ 69.95	⇨ 69.95
Discretize_930_NM	⇨ 70.05	⇨ 70.05	⇨ 70.05
Parameter Changes	**NaiveBayes ''**	**NaiveBayes -K**	**NaiveBayes -D**
useKernelEstimator	FALSE	TRUE	FALSE
useSupervisedDiscretization	FALSE	FALSE	TRUE

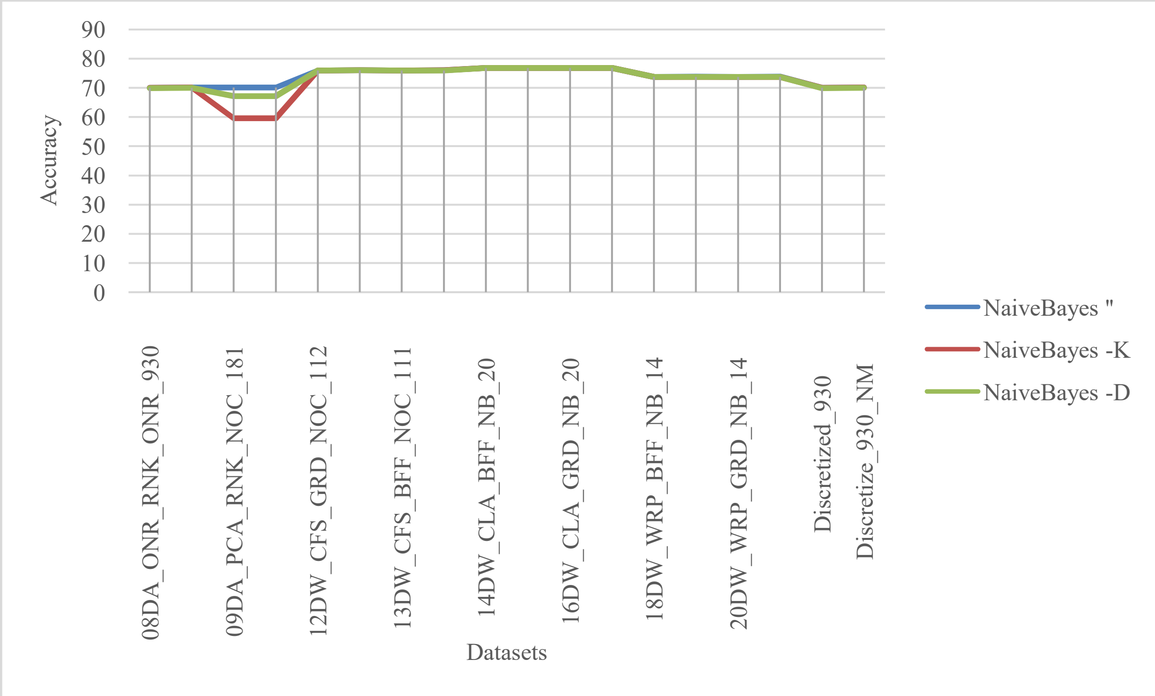


Figure D.08: Naïve bayes performance analysis with different parameter settings

2. LibSVM

LibSVM with 10-fold cross validation performed with its default parameters and changing the *kernel* and *seed* parameters. Following table and graph describes the prediction accuracy under different settings. No accuracy reported over 77.56% which is the highest value obtained so far. Highest 3% accuracies are highlighted.

	Dataset	LibSVM '-S 0 -K 2	LibSVM '-S 0 -K 0	LibSVM '-S 0 -K 1	LibSVM '-S 0 -K 2	LibSVM '-S 0 -K 3
1	08DA_ONR_RNK_ONR_930	72.96	76.48	68.8	72.96	71.34
2	08DA_ONR_RNK_ONR_930_NM	72.96	76.48	68.81	72.96	71.34
3	09DA_PCA_RNK_NOC_181	73.1	76.42	71.1	73.1	75.69
4	09DA_PCA_RNK_NOC_181_NM	73.1	76.42	71.1	73.1	75.69
5	12DW_CFS_GRD_NOC_112	77.01	70	73.86	77.01	76.19
6	12DW_CFS_GRD_NOC_112_NM	77.01	70	73.86	77.01	76.19
7	13DW_CFS_BFF_NOC_111	77.04	70.23	73.89	77.04	76.19
8	13DW_CFS_BFF_NOC_111_NM	77.04	70.23	73.89	77.04	76.17
9	14DW_CLA_BFF_NB_20	71.39	73.69	72.16	71.39	70.43
10	14DW_CLA_BFF_NB_20_NM	71.41	73.7	72.18	71.41	70.46
11	16DW_CLA_GRD_NB_20	71.39	73.69	72.16	71.39	70.43
12	16DW_CLA_GRD_NB_20_NM	71.41	73.7	72.18	71.41	70.46
13	18DW_WRP_BFF_NB_14	68.65	69.53	65.48	68.65	67.91
14	18DW_WRP_BFF_NB_14_NM	68.66	69.54	65.5	68.66	67.93
15	20DW_WRP_GRD_NB_14	68.65	69.53	65.48	68.65	67.91
16	20DW_WRP_GRD_NB_14_NM	68.66	69.54	65.5	68.66	67.93
17	Discretized_930	72.23	76.48	68.7	72.23	70.56
18	Discretize_930_NM	72.21	76.48	68.71	72.21	70.58
	Parameter Changes	**LibSVM '-S 0 -K 2**	**LibSVM '-S 0 -K 0**	**LibSVM '-S 0 -K 1**	**LibSVM '-S 0 -K 2**	**LibSVM '-S 0 -K 3**
	kernal	radial	linear	polynomiyal	radial	sigmoid
	seed	1	7	7	7	7

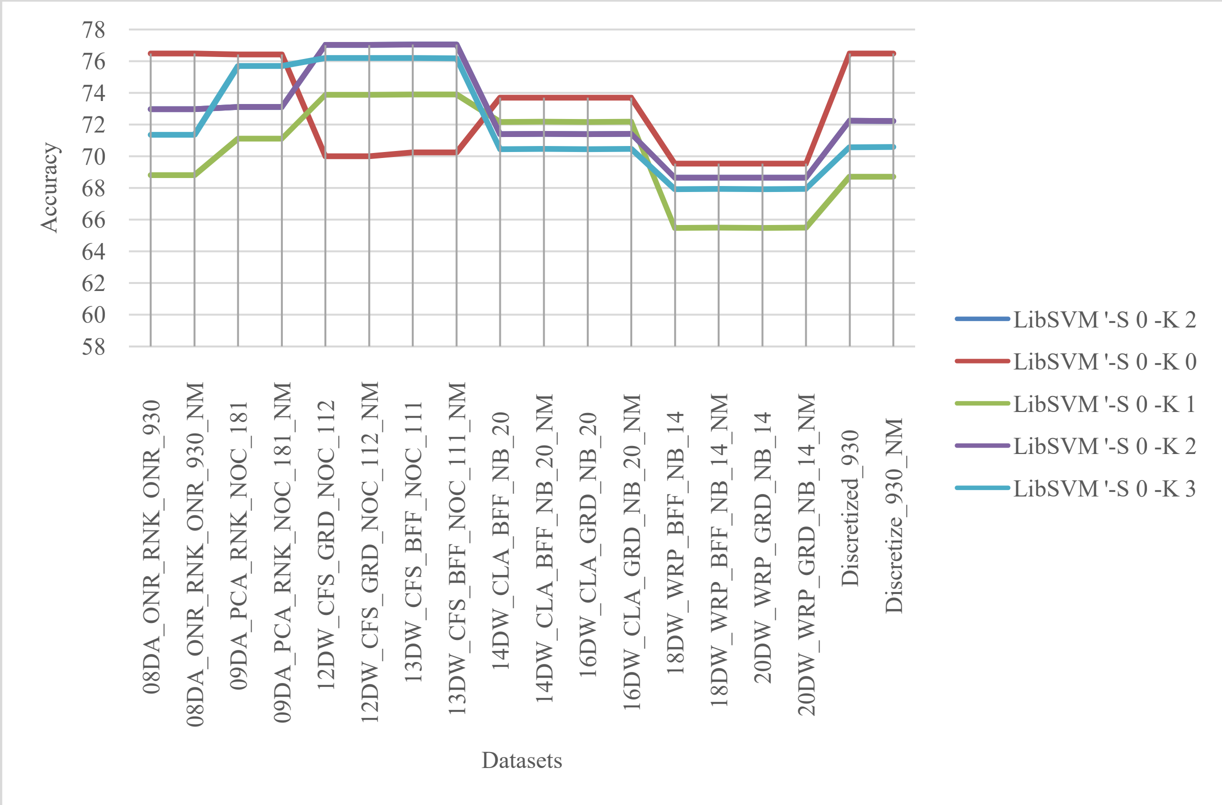


Figure D.09: LibSVM performance analysis with different parameter settings

3. Random Forest

RandomForest with 10-fold cross validation performed with its default parameters and changing the *computeAttributeImportance*and *seed* parameters. Following table and graph describes the prediction accuracy under different settings. No accuracy reported over 77.56% which is the highest value obtained so far. Highest 3% accuracies are highlighted.

	Dataset	RandomForest -S 1	RandomForest -S 7	RandomForest -S 7 attribute-importance
1	08DA_ONR_RNK_ONR_930	70.84	70.76	70.76
2	08DA_ONR_RNK_ONR_930_NM	71.14	71.2	71.2
3	09DA_PCA_RNK_NOC_181	66.54	66.33	66.33
4	09DA_PCA_RNK_NOC_181_NM	66.54	66.33	66.33
5	12DW_CFS_GRD_NOC_112	73.26	73.24	73.24
6	12DW_CFS_GRD_NOC_112_NM	73.23	73.21	73.21
7	13DW_CFS_BFF_NOC_111	73.3	73.09	73.09
8	13DW_CFS_BFF_NOC_111_NM	73.58	72.96	72.96
9	14DW_CLA_BFF_NB_20	70.33	70.16	70.16
10	14DW_CLA_BFF_NB_20_NM	70.45	70.04	70.04
11	16DW_CLA_GRD_NB_20	70.33	70.16	70.16
12	16DW_CLA_GRD_NB_20_NM	70.45	70.04	70.04
13	18DW_WRP_BFF_NB_14	65.6	65.34	65.34
14	18DW_WRP_BFF_NB_14_NM	65.88	65.65	65.65
15	20DW_WRP_GRD_NB_14	65.6	65.34	65.34
16	20DW_WRP_GRD_NB_14_NM	65.88	65.65	65.65
17	Discretized_930	70.84	70.96	70.96
18	Discretize_930_NM	71.24	70.52	70.52
	Parameter Changes	**RandomForest -S 1**	**RandomForest -S 7**	**RandomForest -S 7 attribute-**
	computeAttributeImportance	FALSE	FALSE	TRUE
	seed	1	7	7

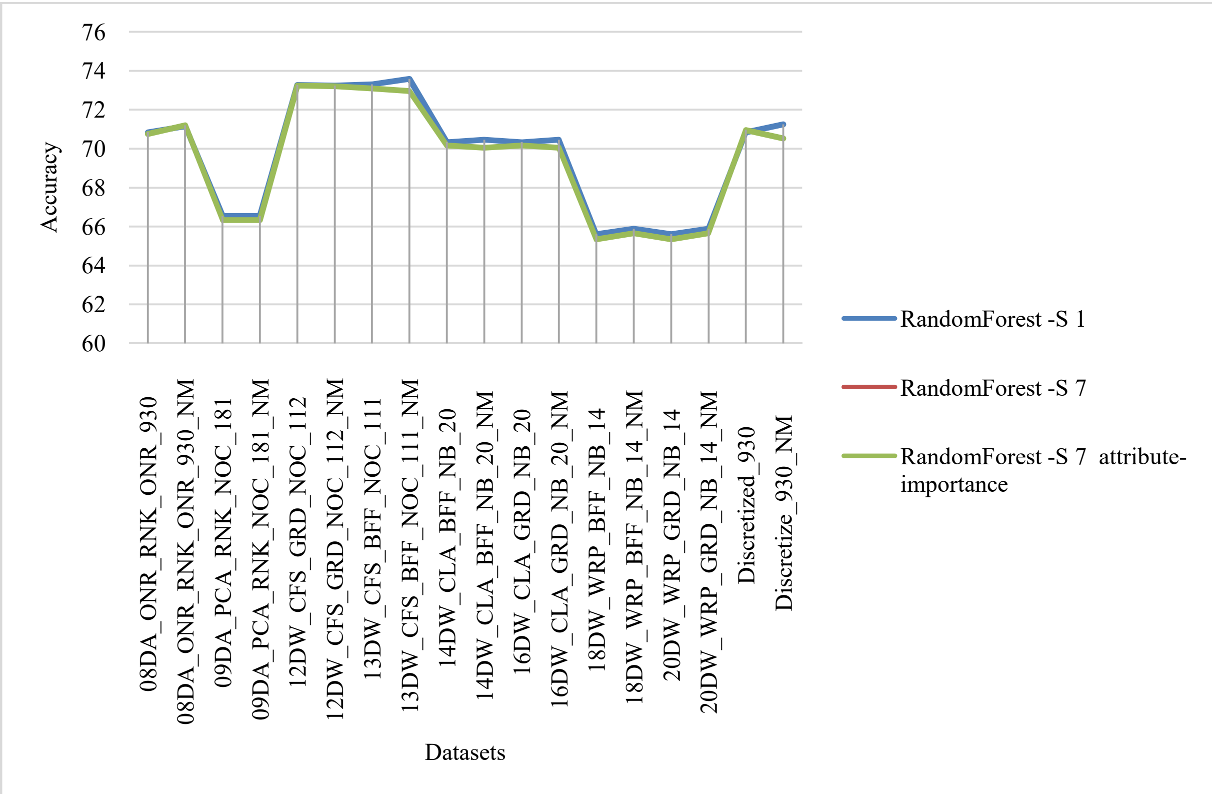


Figure D.10: RandomForest performance analysis with different parameter settings

4. SMO

SMO with 10-fold cross validation performed with its default parameters and changing the *kernel* and *seed* parameters. Following table and graph describes the prediction accuracy under different settings. Accuracy reported over 77.56% which is the highest value obtained so far. Highest 3% accuracies are highlighted.

	Dataset	SMO PolyKernel S1	SMO PolyKernel S7	SMO NormalizedPolyKernel S7	SMO Puk S7	SMO RBFKernel S7
1	08DA_ONR_RNK_ONR_930	**76.64**	**76.65**	**75.54**	**68.51**	**72.81**
2	08DA_ONR_RNK_ONR_930_NM	76.65	76.64	75.54	68.52	72.8
3	09DA_PCA_RNK_NOC_181	69.05	69.06	73	73.61	72.55
4	09DA_PCA_RNK_NOC_181_NM	69.05	69.06	73	73.61	72.55
5	12DW_CFS_GRD_NOC_112	69.99	70	77.86	74.29	77.61
6	12DW_CFS_GRD_NOC_112_NM	69.99	70	77.89	73.97	77.65
7	13DW_CFS_BFF_NOC_111	70.32	70.34	77.97	74.17	77.6
8	13DW_CFS_BFF_NOC_111_NM	70.32	70.32	77.96	74.27	77.64
9	14DW_CLA_BFF_NB_20	73.51	73.47	73.45	72.44	70.29
10	14DW_CLA_BFF_NB_20_NM	73.49	73.49	73.49	72.5	70.3
11	16DW_CLA_GRD_NB_20	73.51	73.47	73.45	72.44	70.29
12	16DW_CLA_GRD_NB_20_NM	73.49	73.49	73.49	72.5	70.3
13	18DW_WRP_BFF_NB_14	69.64	69.63	69.24	67.59	68.06
14	18DW_WRP_BFF_NB_14_NM	69.65	69.63	69.24	67.61	68.07
15	20DW_WRP_GRD_NB_14	69.64	69.63	69.24	67.59	68.06
16	20DW_WRP_GRD_NB_14_NM	69.65	69.63	69.24	67.61	68.07
17	Discretized_930	76.62	76.62	75.54	68.51	72.81
18	Discretize_930_NM	76.65	76.61	75.54	68.52	72.8
		SMO PolyKernel S1	SMO PolyKernel S7	SMO NormalizedPolyKernel	SMO Puk S7	SMO RBFKernel
	kernal	PolyKernel	PolyKernel	NormalizedPolyKernel	Puk	RBFKernel
	seed	1	7	7	7	7

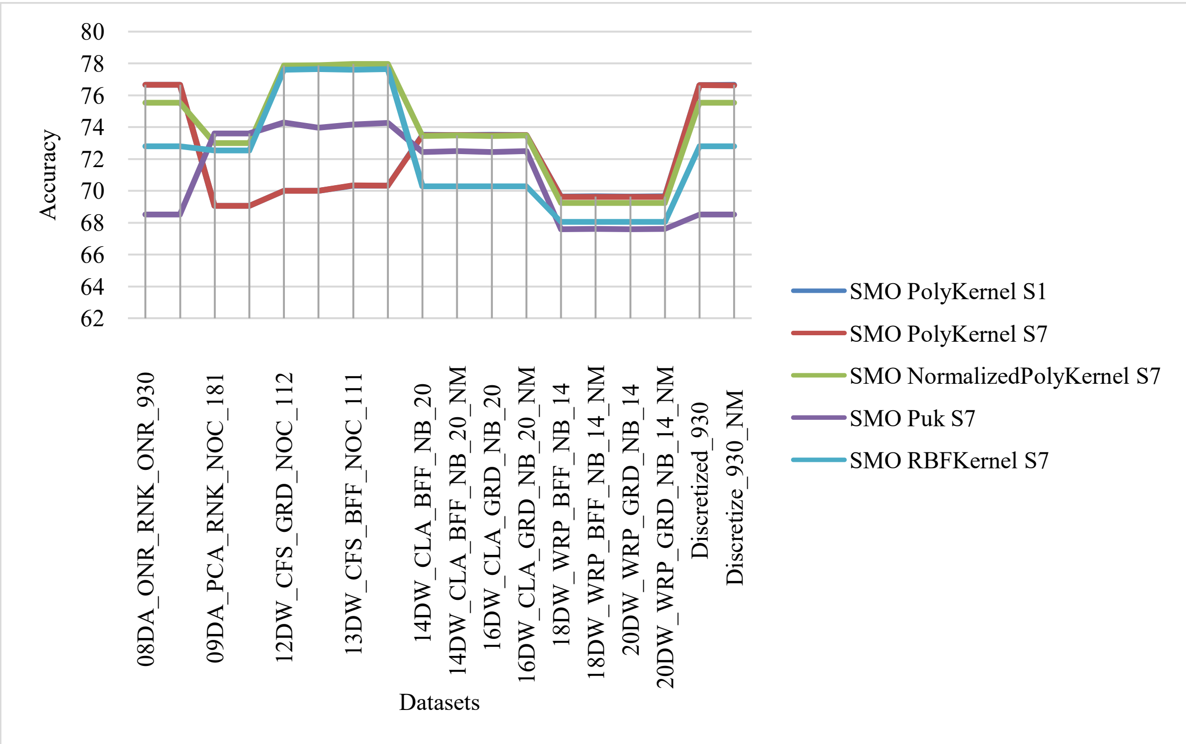


Figure D.11: SMO performance analysis with different parameter settings

5. Bagging

Bagging with 10-fold cross validation performed with its default parameters and changing the *classifier* and *seed* parameters. Following table and graph describes the prediction accuracy under different settings. Accuracy reported over 77.56% which is the highest value obtained so far. Highest 3% accuracies are highlighted.

	Dataset	Bagging S1 REPTree	Bagging S7 REPTree	Bagging S7 DecisionStump	Bagging S7 HoeffdingTree	Bagging S7 J48	Bagging S7 RandomForest
1	08DA_ONR_RNK_ONR_930	66.8	66.85	43.32	70.36	69.34	70.79
2	08DA_ONR_RNK_ONR_930_NM	67.04	67.16	43.46	70.39	69.3	70.77
3	09DA_PCA_RNK_NOC_181	67.41	67.31	48.7	69.5	66.24	72.55
4	09DA_PCA_RNK_NOC_181_NM	67.41	67.31	48.7	69.5	66.24	72.55
5	12DW_CFS_GRD_NOC_112	67.3	67.82	43.32	76.02	70.37	73.47
6	12DW_CFS_GRD_NOC_112_NM	67.24	67.98	43.46	76.13	70.4	73.36
7	13DW_CFS_BFF_NOC_111	67.35	67.74	43.32	75.96	70.44	73.31
8	13DW_CFS_BFF_NOC_111_NM	67.25	67.96	43.46	76.05	70.28	73.4
9	14DW_CLA_BFF_NB_20	66.04	66.26	43.29	76.39	66.56	71.19
10	14DW_CLA_BFF_NB_20_NM	66.14	66.23	43.26	76.4	66.75	71.24
11	16DW_CLA_GRD_NB_20	66.04	66.26	43.29	76.39	66.56	71.19
12	16DW_CLA_GRD_NB_20_NM	66.14	66.23	43.26	76.4	66.75	71.24
13	18DW_WRP_BFF_NB_14	65.44	64.58	43.24	73.55	64.01	66.71
14	18DW_WRP_BFF_NB_14_NM	65.37	64.65	43.21	73.55	64.01	66.74
15	20DW_WRP_GRD_NB_14	65.44	64.58	43.24	73.55	64.01	66.71
16	20DW_WRP_GRD_NB_14_NM	65.37	64.65	43.21	73.55	64.01	66.74
17	Discretized_930	66.61	66.76	43.32	70.36	69.19	70.48
18	Discretize_930_NM	67.1	67.03	43.46	70.39	69.44	70.81
		Bagging S1	Bagging S7	Bagging S7 DecisionStump	Bagging S7 HoeffdingTree	Bagging S7 J48	Bagging S7 RandomForest
	Classifier	REPTree	REPTree	DecisionStump	HoeffdingTree	J48	RandomForest
	Seed	1	2	3	4	5	6

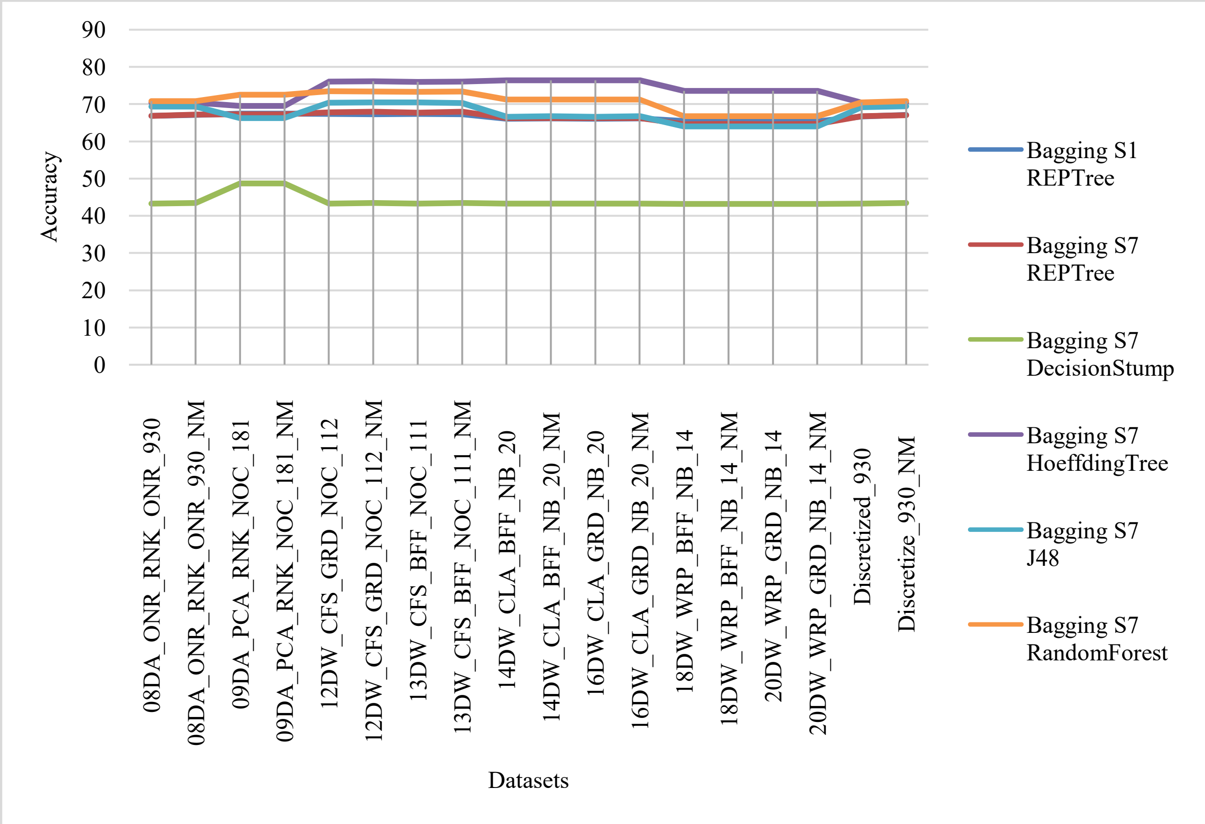


Figure D.12: Bagging performance analysis with different parameter settings

Following graphs illustrate the differences between accuracy predicted among the three dataset categories, 1. Original (Datasets with originally extracted values) and two additional datasets generated based on attribute filtering (2) Normalized and (3) discretized with eight classification algorithms under with and without missing values.

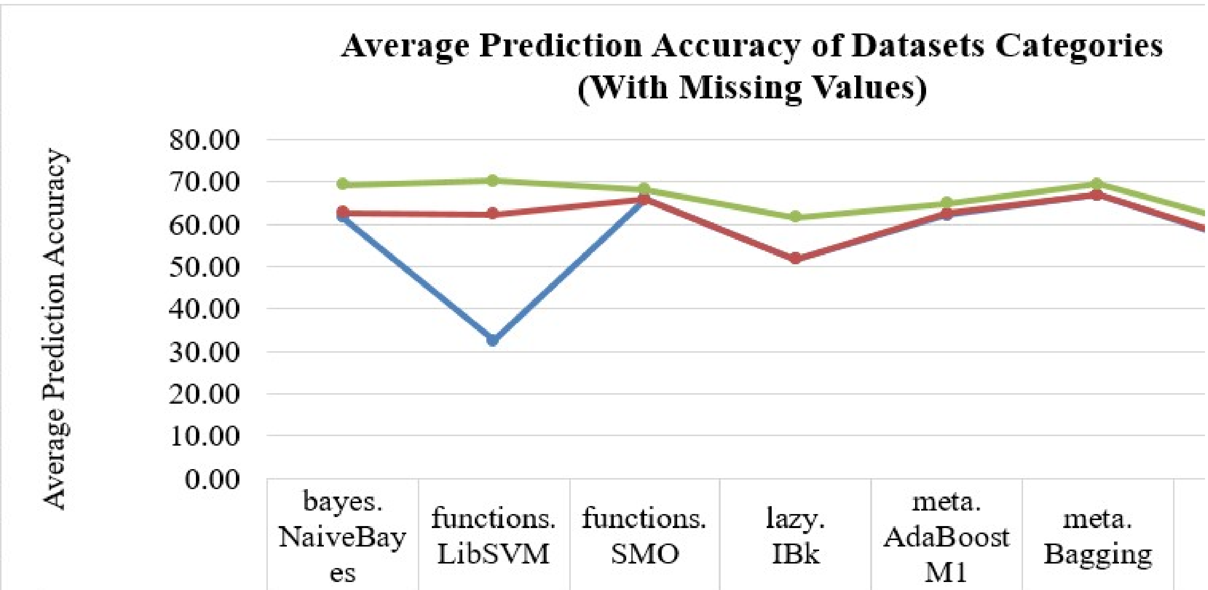


Figure D.13: Average Prediction Accuracy of Datasets Categories (With Missing Values)

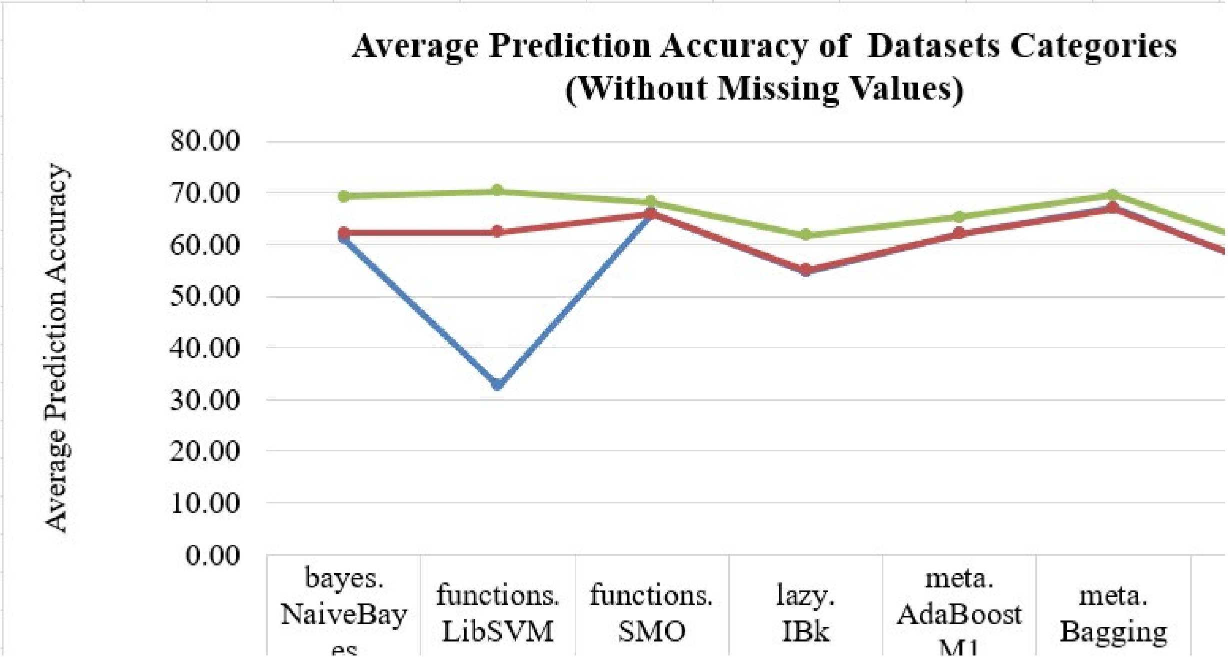


Figure D.14: Average Prediction Accuracy of Datasets Categories (Without Missing Values)

According to the above graphs, datasets with discretized values obtain the highest prediction accuracy than original and normalized.

Based on the findings obtained through WEKA analysis, SVM classification algorithm with discretized datasets provided the highest predication accuracy. Therefore, to get the better accuracy further, classification task is extended to analyze with Python. For Python analysis following are considered.

1. Considered SVM as main classification algorithm based on WEKA findings and Logistic Regression in addition to that.
2. Considered discretized datasets which provided highest accuracy with WEKA findings. Python does not support to analyze attributes with categorical data which discretized datasets has and the missing values. Therefore, for Python analysis, normalized datasets without missing values are considered corresponded to the discretized datasets.
3. 10-fold cross validation is performed with seed 7.

Following graph illustrate the accuracies obtained and the performance differences.

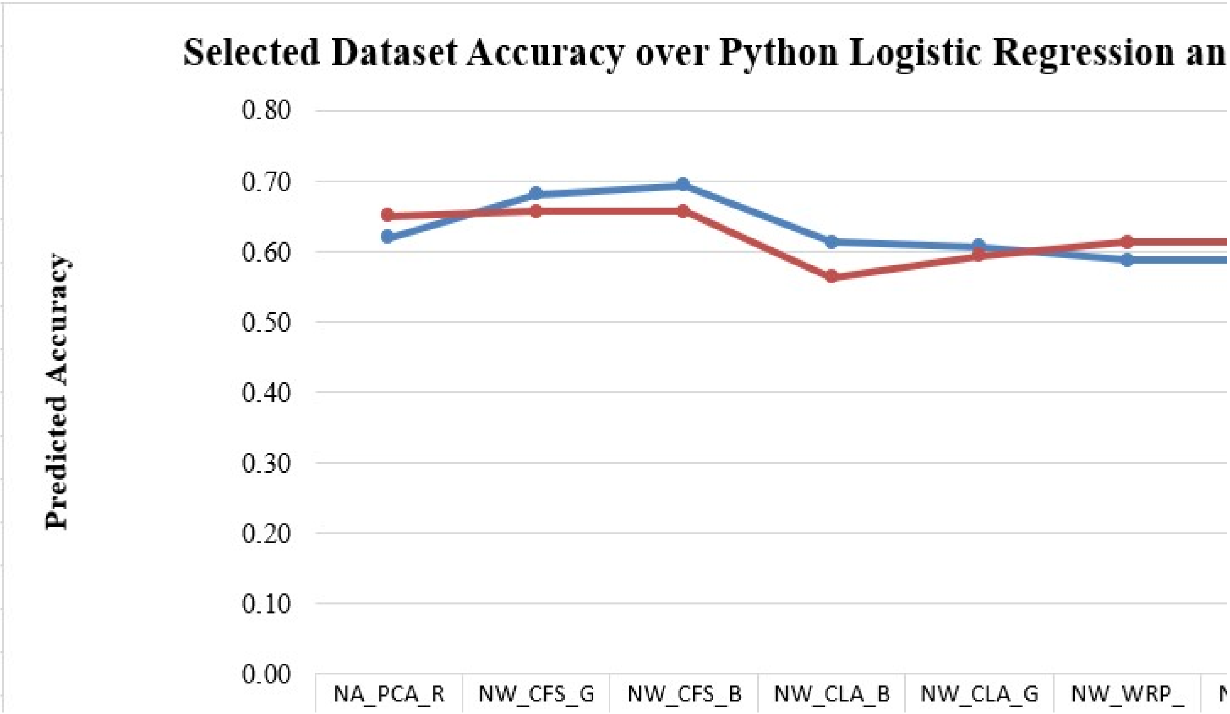


Figure D.15: Selected Dataset Accuracy over Python Logistic Regression and SVM

As a summary 930 audio features are extracted through MIRToolbox using MATLAB and jAudio for feature selection using 800 samples. WEKA has been used as main tool for pre-processing, feature selection, experiment and classification tasks in addition to the Python sklearn. Supervised discretize, unsupervised normalize and unsupervised replace missing value filters have been used to pre-processes the original dataset. Three distinct original, discretized and normalized datasets are used to feature selection with WEKA attribute and wrapper subset evaluators. 59 datasets obtained through feature selection is prepared for classification with eight classification algorithms namely Naïve Bayes, LibSVM, SMO, IBK, AdaBoostM1, Bagging, 48 and Random Forest. WEKA experimenter is used to upload, configure algorithms, perform classification and compare the output results.

According to the analysis, support vector machine algorithm LibSVM exhibited highest prediction accuracy while Bagging exhibits the highest average performance and the minimum standard deviation among the accuracy. LibSVM reported highest accuracy deviation as well as the lowest prediction results. Among three distinct datasets, only disseized datasets provided the better accuracy. During comparison, IBK, AdaBoostM1 and IBK not exhibit the prediction accuracy over 75%. Algorithms and discretized datasets predicated over 75% has been considered to perform optimization with different parameter setting obtain the better output results. Based on the results, SMO resulted over 77% accuracy under *NormalizedPolyKernel* and random seed 7.

Chapter 4: Evaluation and Result

Through the analysis, experiment results show that datasets with discretized values used for classification has obtained the better prediction accuracy than the datasets with originally extracted and normalized values. Among the eight algorithms used, only LivSVM and SMO algorithms which are based on support vector machine provided accuracy over 77% and SMO resulted more accuracy when perform with parameter optimization. Thus, to evaluation and obtain the results using testing datasets SMO algorithm is used.

Testing Sample

To test the selected algorithm, testing sample with 100 instances which included 25 instances of each category is used. New dataset is prepared by following the entire steps which used to obtain the training dataset. Class attribute of the dataset is kept blank to determine the predicting mood class via final model.

Algorithm for final model

SMO with *NormalizedPolyKernel* kernel parameter and *random seed* 7 used to build the final model. To build the final model, filtered classifier in WEKA which support to configure the filtering and classification algorithm has been used. This is because when evaluate the testing dataset with finalized model trained using discretized datasets encountered "Train and Test data are not compatible" issue. The root cause for this issue was when discretized the datasets using supervised discretize filtering algorithm, it changes the original attributes' data type to categorical data type based on the values available in each attribute. These categorical data types are subjective to change over the distribution of the values pertaining to each attribute. Hence both training and testing datasets with same attributes have different categorical data types after discretization pre-processing which validated when WEKA perform the model evaluation with supplied test sets. To avoid that, filtered classifier is used which support discretizing datasets before calling the classification algorithm and not bound the categorical data types for the final model.

Four models build using following datasets to perform the evaluation.

DW_CFS_GRD_NOC_112	Dataset with 112 attributes selected using wrapper classifier subset evaluation feature selection algorithm with greedy search method over discretized dataset which has missing values
DW_CFS_GRD_NOC_112_NM	Dataset with 112 attributes selected using wrapper classifier subset evaluation feature selection algorithm with greedy stepwise search method over discretized dataset without missing values
DW_CFS_BFF_NOC_111	Dataset with 111 attributes selected using wrapper classifier subset evaluation feature selection algorithm with bestfirst search method over discretized dataset with missing values
DW_CFS_BFF_NOC_111_NM	Dataset with 111 attributes selected using wrapper classifier subset evaluation feature selection algorithm with bestfirst search method over discretized dataset without missing values

java weka.filters.unsupervised.attribute.Remove filter used over simple CLI option in WEKA to extract the relevant feature subsets from the original datasets. WEKA Explorer with Classify option used to train and build the model over training datasets and evaluate the testing datasets. When build the final model full training set used to train the model.

Following table illustrate the SMO algorithm predicted accuracy during the training stage over four discretized datasets and the accuracy provided during testing/evaluation stage.

Dateset used to build the model	Training	Testing					
	SMO NormalizedPolyKernel, Seed 7 with 800 sample size	Anger		Happy		Neutral	
		Correctly Predicted Instances	%	Correctly Predicted Instances	%	Correctly Predicted Instances	%
DW_CFS_GRD_NOC_112	77.86	19/25	76	16/25	64	16/25	64
DW_CFS_GRD_NOC_112_NM	[illegible]	19/25	76	16/25	64	16/25	64

Table 08: Accuracy provided during testing/evaluation stage.

According to the above evaluation, all four models built over four datasets have obtained the overall 70% prediction accuracy.

Chapter 5: Conclusion

This research work intended to predict music mood of a music piece which is subjective and ambiguous to person to person. This is still emerging research area due to subjective nature of moods/emotions, ambiguous definitions of mood categories and finding a ground truth data is more challenging, costly and time-consuming task. During this work faced difficulty to find ground truth dataset which is done by previous works that helps to build the strong baseline to measure the accuracy. Then new dataset created with downloading 1000 songs through online store which consists 250 songs in each major category. To understand the features of music, need to understand the basics of digital signal processing which is highly related with mathematics and complex by nature. No research work found by directly promising features that are directly involved with mood prediction and the algorithms. Thus, most researches have been used different features and algorithms based on their preferences and combined lyrics, social tagging and expert judgment to improve the accuracy. This work only involved with features extracted through digital audio for analysis and extracted features as much as possible and selects good feature subsets using feature selection.

WEKA data mining and machine learning tool and the python programming language with sklearn are mainly used to perform the classification tasks and tried cover main algorithm categories defined in WEKA. Data per-processing, feature selection and performance compression among major algorithms with different datasets obtain though feature selection is used to analyze and obtain the better results. According to the analysis, Support Vector Machine algorithms including LibSVM and SMO obtained the better results over pre-processed datasets using supervised discretize filtering algorithms.

For the future works, building ground truth dataset by combining audio features, expert judgment and social annotation tags will be a comprehensive and more experiment tasks. Also incorporate neural networks with deep learning for audio mood classification will be good research work which is not covered through this work.

REFERENCES

[1]. Cyril Laurier,JensGrivolla,PerfectoHerrera,Multimodal Music Mood Classification Using Audio and Lyrics,IEEE,22 December 2008

[2]. Hugues Vinet, The Representation Levels of Music Information, IRCAM, 1, place Igor Stravinsky, F-75004 PARIS – FRANCE,hugues.vinet@ircam.fr,http://www.ircam.fr, 1999

[3]. Owen Craigie Meyers, A Mood-Based Music Classification and Exploration System, Master of Science in Media Arts and Sciences at the MASSACHUSETTS INSTITUTE OF TECHNOLOGY,June 2007

[4]. Susheel Sharma, Rakesh Singh Jadon, Mood Based Music Classification, Department of Computer Applications MITS, Gwalior, M.P. 474005, India

[5]. XIAO HU, IMPROVING MUSIC MOOD CLASSIFICATION USING LYRICS, AUDIO AND SOCIAL TAGS, Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Library and Information Science in the Graduate College of the, University of Illinois at Urbana-Champaign, 2010, Urbana, Illinois

[6]. Renato Eduardo Silva Panda, AUTOMATIC MOOD TRACKING IN AUDIO MUSIC, master's in informatics engineering MSc. Thesis, July 2010

[7]. Xiao Hu, Music and Mood: Where Theory and Reality Meet, Graduate School of Library and Information Science, University of Illinois at Urbana-Champaign,501 E. Daniel St. Champaign, IL, 61820, xiaohu@illinois.edu,2010

[8]. https://www.psychologytoday.com/blog/the-power-music/201608/what-is-music-exactly

[9]. https://www.didjshop.com/BasicMusicalHarmony.html

[10]. http://www.multiweb.cz/twoinches/mp3inside.htm#FrameHeaderA, inside MP3 inside

[11]. http://www.topherlee.com/software/pcm-tut-wavformat.html, Digital Audio - Creating a WAV (RIFF) file

[12]. Hariharan Subramanian, AUDIO SIGNAL CLASSIFICATION,M.Tech. Credit Seminar Report, Electronic Systems Group, EE. Dept, IIT Bombay, Submitted November 2004

[13]. https://ecee.colorado.edu/~mathys/ecen1200/sound/digisound2006_2pp.pdf, Digitized Sound. Telecommunications 1. P. Mathys. Sampling of Waveforms

[14]. http://electriciantraining.tpub.com/14182, WAVE PROPAGATION

[15]. Geoffroy Peeters, A large set of audio features for sound description (similarity and classification) in the CUIDADO project, Ircam, Analysis/Synthesis Team, http://www.ircam.fr/, 2004.

[16]. Sebastian Napiorkowski, Music mood recognition: State of the Art Review, MUS-15, July 10, 2015

[17]. Mark Bukowski, Music Mood Classification: From a Music Piece to a Computed Mood, MUS-15, July 9, 2015

[18]. Olivier Lartillot, MIRtoolbox 1.7, User's Manual, University of Oslo, Norway, Department of Musicology, olartillot@gmail.com, June 3rd, 2017

[19]. Karan Mistry1, Prince Pathak2, Prof. Suvarna Aranjo,Mood based Music Player, International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395 -0056,Volume: 04 Issue: 3 | Mar -2017 www.irjet.net p-ISSN: 2395-0072

[20]. Cyril Laurier, JensGrivolla, Perfecto Herrera, Multimodal Music Mood Classification Using Audio and Lyrics,IEEE,22 December 2008

[21]. Susheel Sharma1, Rakesh Singh Jadon2, Mood Based Music Classification, 1 Department of Computer Applications MITS,Gwalior, M.P. 474005, India, IJISET - International Journal of Innovative Science, Engineering & Technology, Vol. 1 Issue 6, August 2014

[22]. Sternberg, R. In Search of the Human Mind, 2nd Ed.Harcourt, Brace, 1998

[23]. Nairne, J. S. Psychology, The Adaptive Mind. 2nd Ed. Wadsworth, 2000

[24]. Ellsworth PC. William James and emotion, is a century of fame worth a century of misunderstanding? Psychological Review, 1994

[25]. Averill, J. R., Chon, K. K., & Hahn, D. W,Emotions and creativity, East and West. Asian Journal of Social Psychology,2001

[26]. Kleinginna, P. R. Jr., & Kleinginna, A. M, A categorized list of emotion definitions, with suggestions for a consensual definition. Motivation and Emotion,1981

APPENDIX A – MOST COMMONLY USED AUDIO FORMATS AND THEIR STRUCTURES

MP3 Music File Format [10]

Can be expressed as follows where TGAs are optional

[TAG v2] Frame1 Frame2 Frame3... [TAG v1]

- MP3 file is divided into a small block - frames.
- Each frame has constant time length 0.026 sec.
- But size of one frame (in Bytes) varies according to bitrate. E.g. For 128kbps it is (normally) 417 Bytes and for 192kbps 626 Bytes.
- The first 4 Bytes of each frame is frame header and the rest is audio data.
- Frame header consists of information about frame (bitrate, stereo mode...)
- Because of frames are independent items, each of them can have its own characteristics? E.g. in Variable Bitrate files, where each frame can have different bitrate
- Frame header has following structure (each letter is one bit):
  AAAAAAAA AAABBCCDEEEEFFGH IIJJKLMM

AAAAAAAA AAA	Frame synchronizer	All bits are set to 1. It is used for finding the beginning of frame
BB	MPEG version ID	**00** MPEG Version 2.5 **01** Reserved **10** MPEG Version 2 **11** MPEG Version 1 In most MP3 files these values should be 11
CC	Layer	**00** Reserved **01** Layer III **10** Layer II **11** Layer I In most MP3 files these values should be 01 (because MP3 = MPEG 1 Layer 3)
D	CRC Protection	**0** Protected by CRC **1** Not protected
EEEE	Bitrate index	**0000**free 1000 112 **0001** 32 **1001** 128 **0010** 40 **1010** 160 **0011** 48 **1100** 224 **0100** 56 **1101** 256 **0101** 64 **1110** 320 **0110** 80 **1111** bad **0111** 96

		All values are in kbps.
FF	Sampling rate frequency index	**00** 44100 **01** 48000 **10** 32000 **11** Reserved All values are in Hz. In most MP3 files this value should be 00.
G	Padding	**0** Frame is not padded **1** Frame is padded Padding is used to fit the bitrates exactly
H	Private bit	can be freely used for specific needs of an application e.g. it can execute some application specific events
II	Channel	**00** Stereo **01** Joint Stereo (Mostly used in MP3) **10** Dual (Also known as Dual mono; 2 separate channels) **11** Mono (single channel)
JJ	Mode extension (only if Joint Stereo is set)	Intensity Stereo MS Stereo **00** off off **01** on off **10** off on **11** on on Tells which mode for Joint Stereo is used.
K	Copyright	**0** Audio is not copyrighted **1** Audio is copyrighted
L	Original	**0** Copy of original media **1** Original media
MM	Emphasis	**00** None **01** 50/15 **10** Reserved **11** CCIT J.17 Tells if there are emphasized frequencies above cca. 3.2 kHz.

WAV file format [11]

The WAVE file format is a subset of Microsoft's RIFF specification for the storage of multimedia files. A RIFF file starts out with a file header followed by a sequence of data chunks. A WAV (RIFF) file is a multi-format file that contains a header and data.

The header of a WAV (RIFF) file is 44 bytes long and has the following format:

Positions	Sample Value	Description
1 - 4	"RIFF"	Marks the file as a riff file. Characters are each 1-byte long.
5 - 8	File size (integer)	Size of the overall file - 8 bytes, in bytes (32-bit integer).
9 -12	"WAVE"	File Type Header. it always equals "WAVE" for audio
13-16	"fmt "	Format chunk marker. Includes trailing null
17-20	16	Length of format data as listed above
21-22	1	Type of format (1 is PCM) - 2-byte integer
23-24	2	Number of Channels - 2-byte integer
25-28	44100	Sample Rate - 32-byte integer. Common values are 44100 (CD), 48000 (DAT). Sample Rate = Number of Samples per second, or Hertz.
29-32	176400	(Sample Rate * BitsPerSample * Channels) / 8.
33-34	4	(BitsPerSample * Channels) / 8.1 - 8-bit mono2 - 8 bit stereo/16 bit mono4 - 16 bit stereo
35-36	16	Bits per sample
37-40	"data"	"data" chunk header. Marks the beginning of the data section.
41-44	File size (data)	Size of the data section.
Sample values are given above for a 16-bit stereo source.		

APPENDIX B – DEFINITIONS

Bit Size

Bit size determines how much information can be stored in a file. For most of today's purposes, bit size should be 16 bits. 8-bit files are smaller (1/2 the size) but have less resolution [11].

Bit size deals with amplitude. In 8-bit recordings, a total of 256 (0 to 255) amplitude levels are available. In 16 bits, a total of 65,536 (-32768 to 32767) amplitude levels are available. The greater the resolution of the file is, the greater the realistic dynamic range of the file. CD-Audio uses 16-bit samples [11].

Sample Rate

Sample rate is the number of samples per second. CD-Audio has a sample rate of 44,100. This means that 1 second of audio has 44,100 samples. DAT tapes have a sample rate of 48,000. When looking at frequency response, the highest frequency can be 1/2 of the sample rate [11].

Channels

Channels are the number of separate recording elements in the data. For a real quick example, one channel is mono, and two channels are stereo [11].

Data

The data is the individual samples. An individual sample is the bit size times the number of channels. For example, a monaural (single channel), eight-bit recording has an individual sample size of 8 bits. A monaural sixteen-bit recording has an individual sample size of 16 bits. A stereo sixteen-bit recording has an individual sample size of 32 bits. Samples are placed end-to-end to form the data. So, for example, if have four samples (s1, s2, s3, s4) then the data would look like: s1s2s3s4 [11].

Header

The header is the beginning of an audio file. The header is used to provide specifications on the file type, sample rate, sample size and bit size of the file, as well as its overall length [11].

Melody

Melody is the part of a song that would sing along with. When looking at the music, it would be the top line or voice. It is in the foreground [9] [6].

Harmony

Harmony is the other music notes that go along with the melody but aren't in the foreground. Harmony is more than one note that when sounding together creates a chord or something that sounds nice together [9] [6].

Rhythm

Rhythm is the duration that each note is played. It has nothing to do with the pitch (the highness or lowness of a note). It has everything to do with the beat and the combinations of lengths of notes [9] [6].

Sound Waves [14]

Requirement for sound: Three basic elements for transmission and reception of sound must be present before a sound can be produced. They are

(1) The source (or transmitter),

(2) A medium for carrying the sound (air, water, metal, etc.),

(3) The detector (or receiver).

Terms used in wave motion: There are special terms concerning the waves CYCLE, WAVELENGTH, AMPLITUDE, and FREQUENCY. Please refer figure A1 and A2.

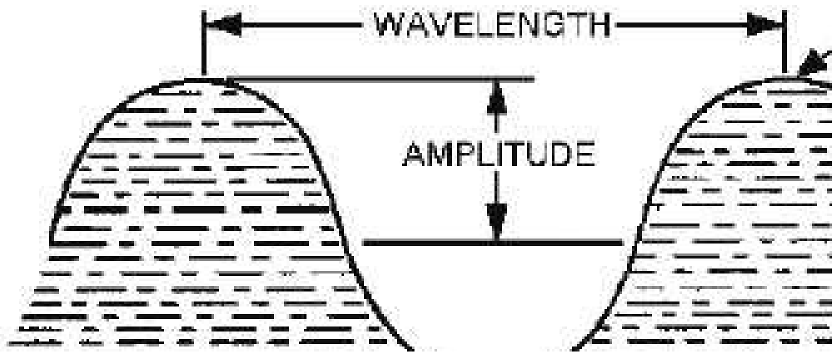


Figure A1: Elements of a wave [14]

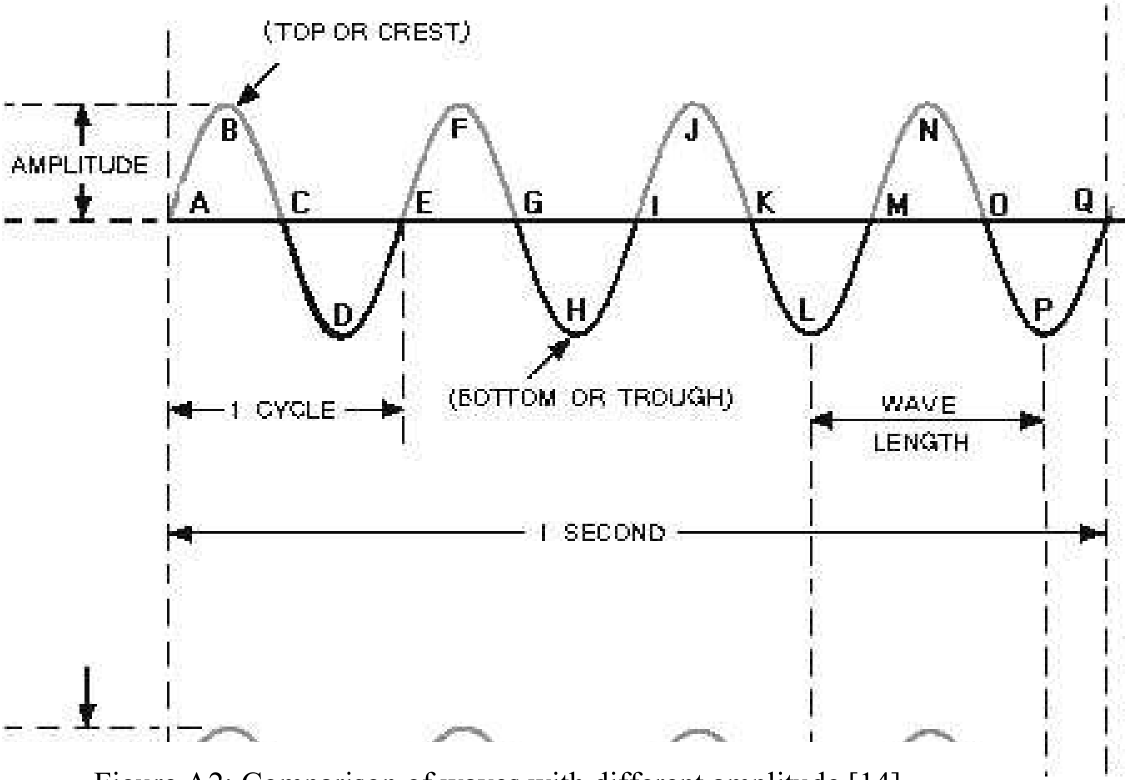


Figure A2: Comparison of waves with different amplitude [14]

Cycle

Points A, B, C, D and E contain one complete cycle of a wave. Each cycle has a maximum value above and a minimum value below the reference line. The portion above the reference line (between points A and C) is called a POSITIVE ALTERNATION and the portion below the reference line (between points C and E) is known as a NEGATIVE ALTERNATION. The combination of one complete positive and one complete negative alternation represents one cycle of the wave. The peak of the positive alternation (maximum value above the line) is referred as the TOP or CREST, and the peak of the negative alternation (maximum value below the line) is called the BOTTOM or TROUGH.

Wavelength

Distance from the leading edge (either TOP or BOTTOM) of one cycle to the corresponding point on the next cycle. Wavelengths vary from an inch at extremely high frequencies to many miles at extremely low frequencies. By default, wavelengths express in meters.

Amplitude

Two waves may have the same wavelength, but with the different size of the crests. Compare wave 1 and wave 2 of figure A2. The height of a wave between crest and the reference line is called the AMPLITUDE of the wave. The amplitude of a wave gives a relative indication of the amount of energy the wave transmits. A continuous series of waves, such as A through Q, having the same amplitude and wavelength, is called a train of waves or WAVE TRAIN.

Frequency

The number of vibrations, or cycles, of a wave train in a unit of time is called the FREQUENCY of the wave train and is measured in HERTZ. E.g. If 5 waves pass a point in one second, the frequency of the wave train is 5 cycles per second.

What is sound?

"Sound is the sensation of hearing", "range of compression-wave frequencies to which the human ear is sensitive". To distinguish the frequencies used in the audible range and the outside of that, following terms are used.

INFRASONIC	:	Sounds below 15 hertz
SONIC	:	Sounds between 15 hertz -10,000 hertz. (Normal hearing range of the humans are 20 hertz -20,000 hertz)
ULTRASONIC	:	Sounds above 20,000 hertz

Music genre

Conventional category that identifies some pieces of **music** as belonging to a shared tradition or set of conventions. It is to be distinguished from **musical** form and **musical** style, although in practice these terms are sometimes used interchangeably. (https://en.wikipedia.org/wiki/Music_genre)

Tempo

Used to describe the timing of music, or the speed at which a piece of music is played. For example, a soothing song would be described as a slow tempo song.
(https://www.vocabulary.com/dictionary/tempo)

APPENDIX C – MISSING ATTRIBUTES OF THE DATASET

Below table illustrate attributes which has missing values with their percentage out of 930 attributes and 800 instances. Used WEKA Simple CLI interface with java weka.core.Instances<filename> option to obtain the statistics.

Attribute ID	Name	Type	Nom	Int	Real	Missing	Unique	Dist
4	dynamics_rms1_RMSenergy_P	Num	0%	11%	88%	4 / 1%	76 / 10%	172
5	dynamics_rms1_RMSenergy_P	Num	0%	0%	100%	3 / 0%	775 / 97%	786
9	fluctuation_centroid1_Spe	Num	0%	0%	2%	784 / 98%	16 / 2%	16
13	rhythm_tempo1_Tempo_Perio	Num	0%	0%	44%	450 / 56%	13 / 2%	61
14	rhythm_tempo1_Tempo_Perio	Num	0%	0%	44%	450 / 56%	346 / 43%	348
19	rhythm_tempo2_Envelopeaut	Num	0%	0%	55%	357 / 45%	15 / 2%	59
20	rhythm_tempo2_Envelopeaut	Num	0%	0%	55%	357 / 45%	435 / 54%	439
25	rhythm_tempo2_Envelopeaut	Num	0%	0%	81%	149 / 19%	13 / 2%	67
26	rhythm_tempo2_Envelopeaut	Num	0%	0%	81%	149 / 19%	637 / 80%	644
31	rhythm_attack_time1_Attac	Num	0%	0%	4%	772 / 97%	20 / 3%	24
32	rhythm_attack_time1_Attac	Num	0%	0%	4%	772 / 97%	26 / 3%	27
39	rhythm_attack_slope1_Atta	Num	0%	0%	4%	771 / 96%	23 / 3%	26
40	rhythm_attack_slope1_Atta	Num	0%	0%	4%	771 / 96%	27 / 3%	28
47	spectral_centroid1_Spectr	Num	0%	28%	70%	15 / 2%	74 / 9%	138
48	spectral_centroid1_Spectr	Num	0%	0%	98%	15 / 2%	769 / 96%	777
53	spectral_brightness1_Brig	Num	0%	20%	79%	2 / 0%	83 / 10%	155
54	spectral_brightness1_Brig	Num	0%	0%	100%	2 / 0%	784 / 98%	791
59	spectral_spread1_Spectral	Num	0%	50%	48%	17 / 2%	27 / 3%	69
60	spectral_spread1_Spectral	Num	0%	0%	98%	17 / 2%	763 / 95%	773
65	spectral_skewness1_Spectr	Num	0%	38%	61%	11 / 1%	47 / 6%	108
66	spectral_skewness1_Spectr	Num	0%	0%	99%	11 / 1%	771 / 96%	780
77	spectral_rolloff951_Rollo	Num	0%	28%	70%	21 / 3%	72 / 9%	132
78	spectral_rolloff951_Rollo	Num	0%	0%	97%	21 / 3%	759 / 95%	769
79	spectral_rolloff951_Rollo	Num	0%	0%	100%	1 / 0%	740 / 93%	769
80	spectral_rolloff851_Rollo	Num	0%	0%	100%	1 / 0%	785 / 98%	792
83	spectral_rolloff851_Rollo	Num	0%	17%	82%	10 / 1%	82 / 10%	158
84	spectral_rolloff851_Rollo	Num	0%	0%	99%	10 / 1%	774 / 97%	782
89	spectral_spectentropy1_En	Num	0%	14%	32%	427 / 53%	39 / 5%	77
90	spectral_spectentropy1_En	Num	0%	0%	47%	427 / 53%	363 / 45%	368
103	spectral_roughness1_Rough	Num	0%	0%	100%	1 / 0%	773 / 97%	786
104	spectral_roughness2_Spect	Num	0%	0%	100%	1 / 0%	785 / 98%	792

113	spectral_roughness2_Spect	Num	0%	13%	87%	1 / 0%	66 / 8%	160
114	spectral_roughness2_Spect	Num	0%	0%	100%	1 / 0%	785 / 98%	792
121	spectral_irregularity1_Sp	Num	0%	0%	100%	1 / 0%	745 / 93%	772
122	spectral_irregularity2_Sp	Num	0%	0%	100%	1 / 0%	785 / 98%	792
131	spectral_irregularity2_Sp	Num	0%	13%	87%	1 / 0%	65 / 8%	161
132	spectral_irregularity2_Sp	Num	0%	0%	100%	1 / 0%	785 / 98%	792
173	spectral_mfcc1_MFCC_Perio	Num	0%	20%	71%	73 / 9%	87 / 11%	154
186	spectral_mfcc1_MFCC_Perio	Num	0%	9%	86%	41 / 5%	676 / 85%	683
332	spectral_mfcc2_Mel-Spectr	Num	0%	16%	84%	1 / 0%	81 / 10%	163
333	spectral_mfcc2_Mel-Spectr	Num	0%	15%	85%	1 / 0%	71 / 9%	154
360	spectral_mfcc2_Mel-Spectr	Num	0%	17%	83%	2 / 0%	88 / 11%	182
361	spectral_mfcc2_Mel-Spectr	Num	0%	18%	82%	1 / 0%	94 / 12%	182
362	spectral_mfcc2_Mel-Spectr	Num	0%	18%	82%	2 / 0%	91 / 11%	182
363	spectral_mfcc2_Mel-Spectr	Num	0%	19%	81%	3 / 0%	85 / 11%	172
364	spectral_mfcc2_Mel-Spectr	Num	0%	18%	81%	3 / 0%	79 / 10%	173
365	spectral_mfcc2_Mel-Spectr	Num	0%	17%	82%	5 / 1%	79 / 10%	169
366	spectral_mfcc2_Mel-Spectr	Num	0%	17%	82%	6 / 1%	87 / 11%	166
367	spectral_mfcc2_Mel-Spectr	Num	0%	17%	82%	7 / 1%	86 / 11%	170
368	spectral_mfcc2_Mel-Spectr	Num	0%	18%	81%	10 / 1%	74 / 9%	159
369	spectral_mfcc2_Mel-Spectr	Num	0%	18%	81%	10 / 1%	86 / 11%	163
370	spectral_mfcc2_Mel-Spectr	Num	0%	17%	81%	14 / 2%	84 / 11%	156
371	spectral_mfcc2_Mel-Spectr	Num	0%	18%	80%	14 / 2%	75 / 9%	148
372	spectral_mfcc2_Mel-Spectr	Num	0%	0%	100%	1 / 0%	785 / 98%	792
373	spectral_mfcc2_Mel-Spectr	Num	0%	0%	100%	1 / 0%	785 / 98%	792
400	spectral_mfcc2_Mel-Spectr	Num	0%	0%	100%	2 / 0%	782 / 98%	790
401	spectral_mfcc2_Mel-Spectr	Num	0%	0%	100%	1 / 0%	785 / 98%	792
402	spectral_mfcc2_Mel-Spectr	Num	0%	0%	100%	2 / 0%	784 / 98%	791
403	spectral_mfcc2_Mel-Spectr	Num	0%	0%	100%	3 / 0%	783 / 98%	790
404	spectral_mfcc2_Mel-Spectr	Num	0%	0%	100%	3 / 0%	781 / 98%	789
405	spectral_mfcc2_Mel-Spectr	Num	0%	0%	99%	5 / 1%	781 / 98%	788
406	spectral_mfcc2_Mel-Spectr	Num	0%	0%	99%	6 / 1%	778 / 97%	786
407	spectral_mfcc2_Mel-Spectr	Num	0%	0%	99%	7 / 1%	781 / 98%	787
408	spectral_mfcc2_Mel-Spectr	Num	0%	0%	99%	10 / 1%	778 / 97%	784
409	spectral_mfcc2_Mel-Spectr	Num	0%	0%	99%	10 / 1%	778 / 97%	784
410	spectral_mfcc2_Mel-Spectr	Num	0%	0%	98%	14 / 2%	772 / 97%	779
411	spectral_mfcc2_Mel-Spectr	Num	0%	0%	98%	14 / 2%	774 / 97%	780
569	spectral_dmfcc2_MFCC_Peri	Num	0%	20%	71%	73 / 9%	87 / 11%	154
582	spectral_dmfcc2_MFCC_Peri	Num	0%	9%	86%	41 / 5%	676 / 85%	683
767	timbre_zerocross1_Zero-cr	Num	0%	13%	87%	2 / 0%	69 / 9%	166
768	timbre_zerocross1_Zero-cr	Num	0%	0%	100%	1 / 0%	783 / 98%	791
774	timbre_lowenergy2_RMSener	Num	0%	11%	88%	4 / 1%	77 / 10%	173
775	timbre_lowenergy2_RMSener	Num	0%	0%	100%	3 / 0%	775 / 97%	786

776	timbre_lowenergy2_RMSener	Num	0%	0%	100%	1 / 0%	764 / 96%	781
777	timbre_spectralflux1_Spec	Num	0%	0%	100%	1 / 0%	785 / 98%	792
782	timbre_spectralflux1_Spec	Num	0%	0%	100%	1 / 0%	755 / 94%	777
783	tonal_chromagram_peak1_Ch	Num	0%	0%	100%	1 / 0%	781 / 98%	790
786	tonal_chromagram_peak1_Ch	Num	0%	1%	40%	470 / 59%	143 / 18%	207
787	tonal_chromagram_peak1_Ch	Num	0%	0%	41%	470 / 59%	322 / 40%	326
788	tonal_chromagram_peak1_Ch	Num	0%	0%	100%	1 / 0%	623 / 78%	707
789	tonal_chromagram_peak1_Ch	Num	0%	100%	0%	1 / 0%	3 / 0%	4
792	tonal_chromagram_peak1_Ch	Num	0%	0%	0%	799 /100%	1 / 0%	1
793	tonal_chromagram_peak1_Ch	Num	0%	0%	0%	799 /100%	1 / 0%	1
794	tonal_chromagram_peak1_Ch	Num	0%	0%	100%	1 / 0%	4 / 1%	8
795	tonal_chromagram_centroid	Num	0%	0%	100%	1 / 0%	785 / 98%	792
798	tonal_chromagram_centroid	Num	0%	0%	10%	720 / 90%	56 / 7%	66
799	tonal_chromagram_centroid	Num	0%	0%	10%	720 / 90%	78 / 10%	79
804	tonal_keyclarity1_Keyclar	Num	0%	4%	81%	126 / 16%	127 / 16%	231
805	tonal_keyclarity1_Keyclar	Num	0%	0%	84%	126 / 16%	660 / 83%	667
840	tonal_mode2_Keystrength_S	Num	0%	0%	100%	1 / 0%	783 / 98%	791
849	tonal_mode2_Keystrength_P	Num	0%	3%	97%	4 / 1%	244 / 31%	424
851	tonal_mode2_Keystrength_P	Num	0%	4%	96%	3 / 0%	270 / 34%	442
852	tonal_mode2_Keystrength_P	Num	0%	3%	96%	3 / 0%	227 / 28%	394
853	tonal_mode2_Keystrength_P	Num	0%	2%	97%	4 / 1%	255 / 32%	424
854	tonal_mode2_Keystrength_P	Num	0%	2%	98%	6 / 1%	256 / 32%	433
855	tonal_mode2_Keystrength_P	Num	0%	2%	98%	4 / 1%	237 / 30%	407
856	tonal_mode2_Keystrength_P	Num	0%	2%	98%	4 / 1%	252 / 32%	428
857	tonal_mode2_Keystrength_P	Num	0%	2%	98%	2 / 0%	224 / 28%	406
858	tonal_mode2_Keystrength_P	Num	0%	2%	97%	5 / 1%	224 / 28%	399
859	tonal_mode2_Keystrength_P	Num	0%	2%	98%	1 / 0%	246 / 31%	417
860	tonal_mode2_Keystrength_P	Num	0%	3%	97%	6 / 1%	245 / 31%	411
861	tonal_mode2_Keystrength_P	Num	0%	0%	100%	4 / 1%	777 / 97%	786
863	tonal_mode2_Keystrength_P	Num	0%	0%	100%	3 / 0%	781 / 98%	789
865	tonal_mode2_Keystrength_P	Num	0%	0%	100%	2 / 0%	782 / 98%	790
866	tonal_mode2_Keystrength_P	Num	0%	0%	99%	5 / 1%	779 / 97%	787
867	tonal_mode2_Keystrength_P	Num	0%	0%	100%	4 / 1%	780 / 98%	788
868	tonal_mode2_Keystrength_P	Num	0%	0%	100%	4 / 1%	776 / 97%	786
869	tonal_mode2_Keystrength_P	Num	0%	0%	100%	1 / 0%	785 / 98%	792
870	tonal_mode2_Keystrength_P	Num	0%	0%	99%	4 / 1%	782 / 98%	789
871	tonal_mode2_Keystrength_P	Num	0%	0%	100%	1 / 0%	785 / 98%	792
872	tonal_mode2_Keystrength_P	Num	0%	0%	99%	5 / 1%	777 / 97%	786
881	tonal_mode2_Keystrength_P	Num	0%	0%	100%	1 / 0%	777 / 97%	788
882	tonal_mode2_Keystrength_P	Num	0%	0%	100%	1 / 0%	769 / 96%	784
883	tonal_mode2_Keystrength_P	Num	0%	0%	100%	1 / 0%	769 / 96%	784
884	tonal_mode2_Keystrength_P	Num	0%	0%	100%	1 / 0%	775 / 97%	787

885	tonal_hcdf1_HarmonicChang	Num	0%	0%	100%	1 / 0%	785 / 98%	792
886	tonal_hcdf1_HarmonicChang	Num	0%	0%	100%	1 / 0%	783 / 98%	791
887	tonal_hcdf1_HarmonicChang	Num	0%	0%	100%	1 / 0%	785 / 98%	792
888	tonal_hcdf1_HarmonicChang	Num	0%	9%	91%	1 / 0%	87 / 11%	202
889	tonal_hcdf1_HarmonicChang	Num	0%	0%	100%	1 / 0%	781 / 98%	790
890	tonal_hcdf1_HarmonicChang	Num	0%	0%	100%	1 / 0%	738 / 92%	768

APPENDIX D – TOOLS

1. **Ant Video Downloader 3.1.24** (https://www.ant.com/)
    - Firefox browser plug-in
    - Used for streaming audio download.
    - Command Line Tool
    - Open Source
    - This tool used to download streaming audio mp3 from us.audionetwork.com to prepare the dataset with four emotion categories.
2. **ExifTool by Phil Harvey 10.68** (https://www.sno.phy.queensu.ca/~phil/exiftool)
    - Used to Reading Meta information of mp3 files
    - Command Line Tool
    - Open Source
3. **FFmpeg 3.4** (https://www.ffmpeg.org)
    - Leading multimedia framework, able to decode, encode, transcode, mux, demux, stream, filter and play pretty much anything that humans and machines have created.
    - Used to extract different audio lengths from mp3 files and convert MP3 to WAV format.
    - Command Line Tool
    - Open Source
4. **jMIR_3_0_developer** (http://jmir.sourceforge.net/jAudio.html)
    - jAudio of jMIR is used to extract selected audio features from sound files.
    - Has command line and well as GUI version for feature extraction. This support feature extraction to either XML or WEKA supported ARFF format.
    - Open Source
5. **MIRtoolbox1.7**

    https://www.jyu.fi/hytk/fi/laitokset/mutku/en/research/materials/mirtoolbox)

    - MIRtoolbox offers an integrated set of functions written in Matlab, dedicated to the extraction from audio files of musical features. Majority of features have been extracted through functions provided by this tool. This tool has separate functions to extract each feature and common function such as *mirfeatures* to extract all related features at once.
    - Integrated with MatLab
    - Open Source

6. **Weka 3.8.2** (https://www.cs.waikato.ac.nz/ml/weka/downloading.html)
   - Weka is data mining software that uses a collection of machine learning algorithms.
   - This tool is used to data pre-processing, feature selection and classification.
   - Open Source
7. **Mathlab R2017b** (https://in.mathworks.com/)
   - Analyzing data, developing algorithms, or creating models
   - Mainly used to run the MIRtoolbox1.7 Matlab tools to extract and analyze the audio features.
   - Commercial software. Used 30 days trial version.
8. **Audacity (https://www.audacityteam.org/)**
   - Audacity® is free, open source, cross-platform audio software for multi-track recording and editing.
   - Provide rich set of features under Recording, Import and Export, Sound Quality, Editing, Accessibility, Audio Effects, Plug-ins and sound analysis

APPENDIX E – FILTER AND WRAPPER ATTRIBUTE SELECTORS USED IN WEKA

Filter Type	Feature Selection Method	Search Algorithm	Description
Attribute	ClassifierAttributeEval	Ranker	Evaluates the worth of an attribute by using a user-specified classifier. [WEKA]
Attribute	CorrelationAttributeEval	Ranker	Used to select most relevant attributes by calculating the correlation between each attribute and the output variable. Then select attributes that have a moderate-to-high positive or negative correlation (close to -1 or 1) and drop attributes with a low correlation (value close to zero). [WEKA]
Attribute	GainRatioAttributeEval	Ranker	Evaluates the worth of an attribute by measuring the gain ratio with respect to the class. [WEKA]
Attribute	InfoGainAttributeEval	Ranker	Calculate information gain/entropy of each attribute for the output variable. Entry values vary from 0 (no information) to 1 (maximum information). Those attributes that

			contribute more information will have a higher information gain value and can be selected, whereas those that do not add much information will have a lower score and can be removed. [WEKA]
Attribute	OneRAttributeEval	Ranker	Evaluates the worth of an attribute by using the OneR classifier. [WEKA]
Attribute	PrincipalComponents	Ranker	Performs a principal components analysis and transformation of the data. It is a way of identifying patterns in data and expressing the data in such a way as to highlight their similarities and differences. once have found these patterns in the data, can compress the data. [WEKA]
Attribute	ReliefFAttributeEval	Ranker	Evaluates the worth of an attribute by repeatedly sampling an instance and considering the value of the given attribute for the nearest instance of the same and different class. [WEKA]
Attribute	SymmetricalUncertAttributeEval	Ranker	Evaluates the worth of an attribute by measuring the

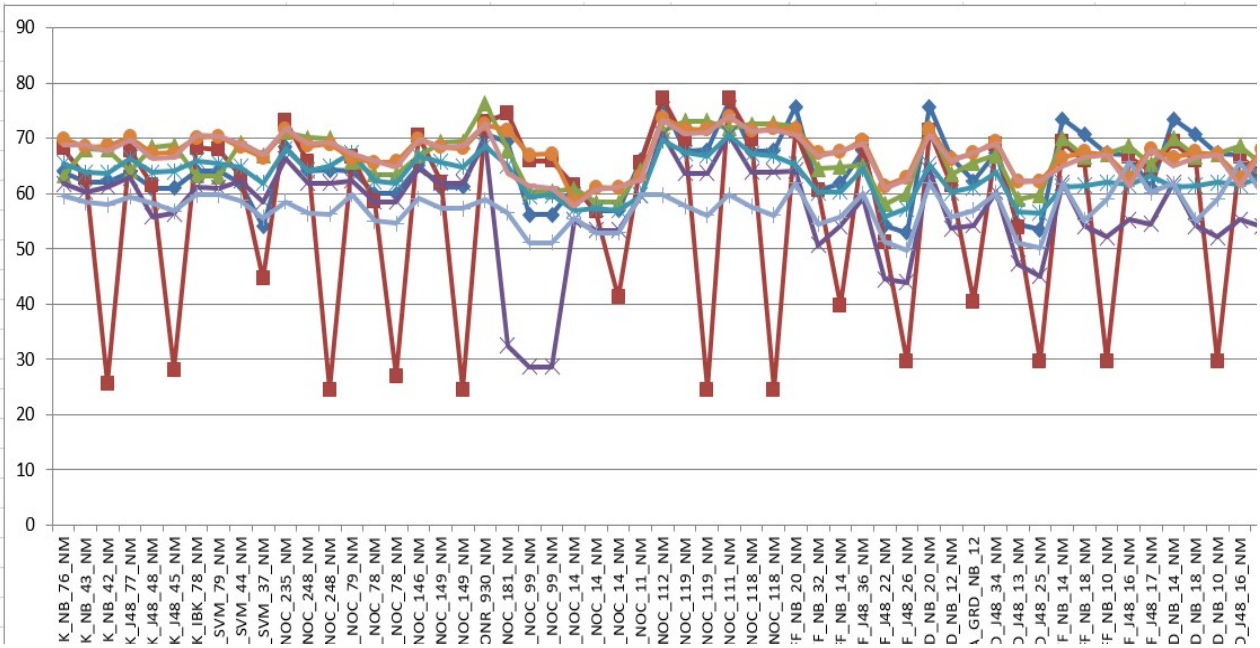
			symmetrical uncertainty with respect to the class. [WEKA]
Subset	CfsSubsetEval	GreedyStepwise BestFirst	Evaluates the worth of a subset of attributes by considering the individual predictive ability of each feature along with the degree of redundancy between them. [WEKA]
Subset	ClassifierSubsetEval	GreedyStepwise BestFirst	Evaluates attribute subsets on training data or a separate hold out testing set. Uses a classifier to estimate the 'merit' of a set of attributes. [WEKA]
Subset	WrapperSubsetEval	GreedyStepwise BestFirst	Evaluates attribute sets by using a learning scheme. [WEKA]

APPENDIX F – PREDICATION RESULTS OF 8 CLASSIFICATION ALGORITHMS UNDER DATASETS WITHOUT MISSING VALUES

Following Table (Train percentage of 66%) illustrate the predication accuracy of each selected algorithm under 59 datasets without missing values. WEKA *Experimenter* tool used to upload the required datasets, classification algorithms and compare the prediction accuracy.

	Dataset	bayes. NaiveBayes	functions. LibSVM	functions. SMO	lazy. IBk	meta. AdaBoostM1	meta. Bagging	trees. J48	trees. RandomForest
1	01DA_CLA_RNK_NB_76_NM	63.86	68.13	63.24	61.73	65.44	69.82	59.49	68.86
2	01NA_CLA_RNK_NB_43_NM	62.13	62.79	68.09	60.26	63.82	68.42	58.35	68.68
3	01OA_CLA_RNK_NB_42_NM	62.1	25.66	67.98	61.18	63.71	68.57	57.98	67.65
4	02DA_CLA_RNK_J48_77_NM	63.93	67.87	64.15	62.9	66.32	70.18	59.45	69.67
5	02NA_CLA_RNK_J48_48_NM	60.88	61.43	68.42	55.85	63.75	67.21	58.31	66.21
6	02OA_CLA_RNK_J48_45_NM	60.99	28.05	68.75	56.32	64.01	67.39	56.84	66.47
7	03DA_CLA_RNK_IBK_78_NM	64.04	68.27	63.27	61.18	65.96	70.07	59.74	70.55
8	04DA_CLA_RNK_SVM_79_NM	64.15	67.98	63.09	60.99	65.48	70.33	59.78	70.33
9	04NA_CLA_RNK_SVM_44_NM	61.73	63.12	69.12	62.32	64.82	68.38	58.75	69.01
10	04OA_CLA_RNK_SVM_37_NM	54.01	44.6	66.95	58.46	61.87	66.43	55.63	66.95
11	05DA_COR_RNK_NOC_235_NM	68.42	73.05	71.1	66.29	68.01	71.54	58.46	71.73
12	05NA_COR_RNK_NOC_248_NM	64.15	65.74	70.04	61.84	64.15	68.57	56.36	69.04
13	05OA_COR_RNK_NOC_248_NM	64.23	24.52	70	61.84	64.89	68.9	56.25	69.19
14	06DA_GNR_RNK_NOC_79_NM	64.08	66.54	65.66	62.32	67.39	66.4	59.82	67.17
15	06NA_GNR_RNK_NOC_78_NM	59.96	58.53	63.42	58.53	62.21	65.59	55	65.96
16	06OA_GNR_RNK_NOC_78_NM	60	26.91	63.42	58.53	61.84	65.7	54.6	64.78
17	07DA_IFG_RNK_NOC_146_NM	65	70.37	66.84	64.56	66.8	69.89	59.08	69.71
18	07NA_IFG_RNK_NOC_149_NM	61.07	62.02	69.26	61.91	65.7	68.38	57.39	68.24
19	07OA_IFG_RNK_NOC_149_NM	61.18	24.52	69.38	61.91	64.74	68.2	57.32	68.57
20	08DA_ONR_RNK_ONR_930_NM	70.88	72.98	76.4	69.38	68.57	72.46	58.82	72.5
21	09DA_PCA_RNK_NOC_181_NM	69.3	74.41	67.87	32.5	65.04	71.43	56.47	63.57
22	09NA_PCA_RNK_NOC_99_NM	56.1	65.96	60.29	28.64	59.45	66.88	51.14	61.43
23	09OA_PCA_RNK_NOC_99_NM	56.14	65.96	60.29	28.64	59.78	67.13	51.14	60.99
24	10DA_SYM_RNK_NOC_14_NM	61.1	61.47	60.85	55.11	56.88	58.53	55.44	57.83
25	10NA_SYM_RNK_NOC_14_NM	56.8	56.76	58.35	53.35	57.32	61.07	52.94	60.74
26	10OA_SYM_RNK_NOC_14_NM	56.91	41.18	58.38	53.35	56.91	60.96	52.9	61.18
27	11DA_RLF_RNK_NOC_11_NM	66.07	65.59	64.12	59.63	59.67	63.09	59.89	62.46
28	12DW_CFS_GRD_NOC_112_NM	76.62	77.24	71.32	70.77	69.63	73.68	59.74	73.27
29	12NW_CFS_GRD_NOC_119_NM	67.68	69.78	73.05	63.71	67.43	71.69	57.68	70.88
30	12OW_CFS_GRD_NOC_119_NM	67.79	24.52	73.09	63.71	66.87	71.51	56.07	71.1
31	13DW_CFS_BFF_NOC_111_NM	76.65	77.17	71.73	70.66	70.48	73.9	59.78	74.01

32	13NW_CFS_BFF_NOC_118_NM	67.57	69.89	72.54	63.86	67.1	71.32	57.65	71.29
33	13OW_CFS_BFF_NOC_118_NM	67.72	24.52	72.61	63.86	66.87	71.8	56.03	71.43
34	14DW_CLA_BFF_NB_20_NM	75.7	71.29	72.1	64.08	65.26	71.62	61.84	70.51
35	14NW_CLA_BFF_NB_32_NM	60.51	60.66	64.41	50.7	60.18	67.21	54.49	66.8
36	14OW_CLA_BFF_NB_14_NM	61.84	39.74	64.85	54.15	60.26	67.54	55.7	67.76
37	15DW_CLA_BFF_J48_36_NM	67.68	68.79	65.18	59.19	64.45	69.6	59.93	69.08
38	15NW_CLA_BFF_J48_22_NM	54.26	51.21	58.09	44.49	55.7	61.29	51.32	60.85
39	15OW_CLA_BFF_J48_26_NM	52.87	29.56	59.93	43.86	57.32	62.83	49.71	62.39
40	16DW_CLA_GRD_NB_20_NM	75.7	71.29	72.1	64.08	65.26	71.62	61.84	70.51
41	16NW_CLA_GRD_NB_12_NM	66.51	61.54	63.49	53.71	60.18	66.1	55.74	65.81
42	16OW_CLA_GRD_NB_12	62.21	40.48	65.55	54.23	61.1	67.28	56.88	67.35
43	17DW_CLA_GRD_J48_34_NM	67.43	68.86	66.95	60.29	63.82	69.45	60.04	69.04
44	17NW_CLA_GRD_J48_13_NM	54.34	53.75	58.97	47.28	56.69	62.13	51.1	62.32
45	17OW_CLA_GRD_J48_25_NM	53.49	29.67	59.74	45.04	56.47	62.21	50.15	62.35
46	18DW_WRP_BFF_NB_14_NM	73.42	69.3	69.93	61.91	61.25	66.43	61.73	65.04
47	18NW_WRP_BFF_NB_18_NM	70.63	65.92	66.69	54.15	61.4	67.54	55.11	66.62
48	18OW_WRP_BFF_NB_10_NM	67.21	29.63	67.1	52.06	61.95	67.02	58.9	66.91
49	19DW_WRP_BFF_J48_16_NM	66.99	66.62	68.64	55.22	61.58	62.9	65.74	61.32
50	19NW_WRP_BFF_J48_17_NM	60.99	64.04	65.29	54.52	63.2	68.01	60.18	67.87
51	20DW_WRP_GRD_NB_14_NM	73.42	69.3	69.93	61.91	61.25	66.43	61.73	65.04
52	20NW_WRP_GRD_NB_18_NM	70.63	65.92	66.69	54.15	61.4	67.54	55.11	66.62
53	20OW_WRP_GRD_NB_10_NM	67.21	29.63	67.1	52.06	61.95	67.02	58.9	66.91
54	21DW_WRP_GRD_J48_16_NM	66.99	66.62	68.64	55.22	61.58	62.9	65.74	61.32
55	21NW_WRP_GRD_J48_14_NM	61.21	63.6	65.11	53.97	63.05	67.9	60	67.65
56	PrincipalComponen_930_NM	56.14	65.96	60.29	28.64	59.78	67.13	51.14	60.99
57	Discretize_930_NM	70.88	72.5	76.36	69.38	67.43	71.84	57.87	72.06
58	Normalize_930_NM	63.2	64.71	67.39	59.45	64.3	69.12	55.66	69.67
59	Original_930_NM	63.24	24.52	67.39	59.45	64.74	68.75	54.82	68.38
	Average	**64.20**	**56.65**	**66.63**	**56.77**	**63.13**	**67.71**	**57.21**	**66.99**
	Max	**76.65**	**77.24**	**76.4**	**70.77**	**70.48**	**73.9**	**65.74**	**74.01**
	Min	**52.87**	**24.52**	**58.09**	**28.64**	**55.7**	**58.53**	**49.71**	**57.83**
	Standard Deviation	**5.97**	**17.13**	**4.49**	**9.50**	**3.52**	**3.38**	**3.49**	**3.76**



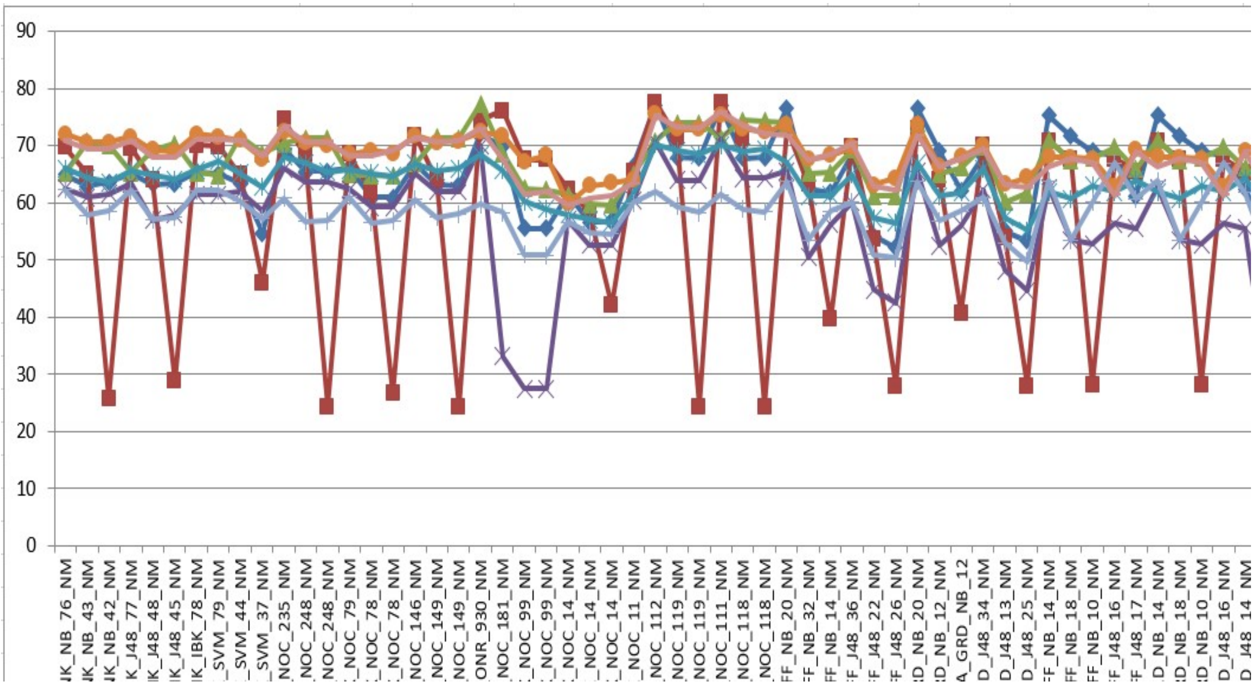
APPENDIX F: Graphical view of predication results of 8 classification algorithms with 59 datasets without missing values (66% split)

According to the results, best accuracy 77.24 is reported by LibSVM classification algorithm with discretized dataset of 112 attributes which is extracted by wrapper CfsSubsetEval feature selection algorithm with GreedyStepwise search algorithm. Also, only NaiveBayes, LibSVM and SMO provided the accuracy more than 75% with discretized datasets. Maximum average accuracy of 67.71 and minimum standard deviation 3.38 is given by the Bagging classifier.

Following Table (Train percentage of 80%) illustrate the predication accuracy of each selected algorithm under 59 datasets without missing values. WEKA *Experimenter* tool used to upload the required datasets, classification algorithms and compare the prediction accuracy.

	Dataset	bayes. NaiveBayes	functions. LibSVM	functions. SMO	lazy. IBk	meta. AdaBoostM1	meta. Bagging	trees. J48	trees. RandomForest
1	01DA_CLA_RNK_NB_76_NM	64.88	69.75	65.19	62.5	65.94	72	62.31	70.88
2	01NA_CLA_RNK_NB_43_NM	62.88	65	70.63	60.94	64.25	70.5	57.75	69.44
3	01OA_CLA_RNK_NB_42_NM	63.56	25.75	70.13	61.5	63.56	70.56	58.56	69.5
4	02DA_CLA_RNK_J48_77_NM	65	69.63	65.31	63.31	65.63	71.5	62.19	70.75
5	02NA_CLA_RNK_J48_48_NM	63.19	63.88	69.25	57	64.81	69.25	57.13	67.88
6	02OA_CLA_RNK_J48_45_NM	63.31	28.81	70.31	57.94	64	68.88	57.44	68.06
7	03DA_CLA_RNK_IBK_78_NM	65.38	70.06	65.25	61.38	65.69	72	62.19	71.06
8	04DA_CLA_RNK_SVM_79_NM	65.06	69.88	64.75	61.38	67.25	71.56	62.13	71.19
9	04NA_CLA_RNK_SVM_44_NM	63.5	64.94	71.5	61.88	65.13	70.94	60.31	70.88
10	04OA_CLA_RNK_SVM_37_NM	54.63	46	68.63	58.56	62.69	67.69	57.19	68.19
11	05DA_COR_RNK_NOC_235_NM	69.38	74.56	70.31	65.94	68.06	72.44	60.81	73.31
12	05NA_COR_RNK_NOC_248_NM	65.56	68.25	71.25	63.75	66.94	70.38	56.75	70.63
13	05OA_COR_RNK_NOC_248_NM	65.56	24.31	71.25	63.75	65.44	70.06	56.81	70.63
14	06DA_GNR_RNK_NOC_79_NM	66.06	68.5	64.94	62.38	65.75	68.56	61.06	68.19
15	06NA_GNR_RNK_NOC_78_NM	60.88	61.81	64.75	59.25	65.19	69.06	56.44	68.12
16	06OA_GNR_RNK_NOC_78_NM	60.88	26.63	64.62	59.25	64.63	68.69	56.88	69.25
17	07DA_IFG_RNK_NOC_146_NM	66.69	71.81	66.88	65.13	66.81	71.81	60.56	71.37
18	07NA_IFG_RNK_NOC_149_NM	63.06	63.94	71.38	62	65.56	70.5	57.38	70.69
19	07OA_IFG_RNK_NOC_149_NM	63	24.31	71.31	62	66	70.75	58.06	70.94
20	08DA_ONR_RNK_ONR_930_NM	72.13	74.38	77.31	69.94	68.44	72.38	59.69	73
21	09DA_PCA_RNK_NOC_181_NM	69.69	76.06	68.63	33.06	65.63	71.81	58.25	67.38
22	09NA_PCA_RNK_NOC_99_NM	55.5	67.5	62.44	27.44	60.19	67.19	50.88	61.56
23	09OA_PCA_RNK_NOC_99_NM	55.5	67.5	62.31	27.44	58.88	68.31	50.81	62
24	10DA_SYM_RNK_NOC_14_NM	61.56	62.38	61.38	57	57.88	59.88	56.56	59.5
25	10NA_SYM_RNK_NOC_14_NM	56.63	58.25	59.75	52.63	57.13	63	54.69	60.75
26	10OA_SYM_RNK_NOC_14_NM	56.69	42	59.63	52.63	56.5	63.44	54.56	61.19
27	11DA_RLF_RNK_NOC_11_NM	66.5	65.5	64.94	60.31	61.5	64	59.94	63.37
28	12DW_CFS_GRD_NOC_112_NM	77	77.56	70.81	70.63	70.25	75.63	61.94	75.31
29	12NW_CFS_GRD_NOC_119_NM	67.88	71.38	73.88	63.88	69.06	72.88	59.31	73.31
30	12OW_CFS_GRD_NOC_119_NM	67.81	24.31	73.94	63.88	68.44	72.88	58.25	73.06
31	13DW_CFS_BFF_NOC_111_NM	76.94	77.5	71.25	71.19	70.06	75.38	61.44	75.88
32	13NW_CFS_BFF_NOC_118_NM	67.81	71.19	74.38	64.31	68.88	73	58.88	73.44
33	13OW_CFS_BFF_NOC_118_NM	67.88	24.31	74.31	64.31	69.38	72.31	58.38	72.13
34	14DW_CLA_BFF_NB_20_NM	76.63	73	73.88	65.56	66.94	73.63	63.69	71.75
35	14NW_CLA_BFF_NB_32_NM	62.44	63.19	65.13	50.44	61.19	67.56	53.5	67.44
36	14OW_CLA_BFF_NB_14_NM	61.81	39.56	65.31	56.25	61.31	68.31	58.56	68.5
37	15DW_CLA_BFF_J48_36_NM	68.5	69.88	69.19	60.06	64.94	69.69	60.06	70.44
38	15NW_CLA_BFF_J48_22_NM	54.19	53.63	61.19	44.75	57.31	63.06	50.88	62.81
39	15OW_CLA_BFF_J48_26_NM	52.19	27.94	61.13	42.56	56.44	64.31	50.31	62.13
40	16DW_CLA_GRD_NB_20_NM	76.63	73	73.88	65.56	66.94	73.63	63.69	71.75
41	16NW_CLA_GRD_NB_12_NM	68.94	64	65.06	52.5	61.13	66.38	56.88	66.19
42	16OW_CLA_GRD_NB_12	62.19	40.69	66.13	56	61.75	68.06	58.56	67.62

43	17DW_CLA_GRD_J48_34_NM	68.19	70.06	69.81	62.25	65.94	70.06	60.94	69.38
44	17NW_CLA_GRD_J48_13_NM	55.06	53.94	60.13	48.19	57.38	63.25	53	63.19
45	17OW_CLA_GRD_J48_25_NM	53.38	27.94	61.69	44.5	55.13	64.38	49.69	62.56
46	18DW_WRP_BFF_NB_14_NM	75.31	70.81	71.06	62.63	61.94	68	63.87	66.31
47	18NW_WRP_BFF_NB_18_NM	71.69	67.56	67.25	53.44	60.69	67.88	53.5	67.63
48	18OW_WRP_BFF_NB_10_NM	68.94	28.19	67.75	52.75	63	67.63	60.06	67
49	19DW_WRP_BFF_J48_16_NM	68.13	67.63	69.75	56.31	61.88	62.88	66.94	61.31
50	19NW_WRP_BFF_J48_17_NM	61.13	65.56	65.88	55.5	63.94	69.19	60.63	68.88
51	20DW_WRP_GRD_NB_14_NM	75.31	70.81	71.06	62.63	61.94	68	63.87	66.31
52	20NW_WRP_GRD_NB_18_NM	71.69	67.56	67.25	53.44	60.69	67.88	53.5	67.63
53	20OW_WRP_GRD_NB_10_NM	68.94	28.19	67.75	52.75	63	67.63	60.06	67
54	21DW_WRP_GRD_J48_16_NM	68.13	67.63	69.75	56.31	61.88	62.88	66.94	61.31
55	21NW_WRP_GRD_J48_14_NM	62.31	65.25	66.13	55.5	64.31	69	61.75	69
56	PrincipalComponen_930_NM	55.5	67.5	62.31	27.44	58.88	68.31	50.81	62
57	Discretize_930_NM	72.13	73.69	77.25	69.94	69.38	72.56	59.38	72.81
58	Normalize_930_NM	62.94	67.44	66.75	59.25	65.06	69.88	58	70.19
59	Original_930_NM	63.13	24.31	66.75	59.25	66.13	70.25	57.62	70.13
	Average	**65.04**	**57.74**	**67.83**	**57.19**	**63.81**	**69.04**	**58.34**	**68.21**
	Max	**77**	**77.56**	**77.31**	**71.19**	**70.25**	**75.63**	**66.94**	**75.88**
	Min	**52.19**	**24.31**	**59.63**	**27.44**	**55.13**	**59.88**	**49.69**	**59.5**
	Standard Deviation	**6.28**	**17.92**	**4.37**	**9.86**	**3.78**	**3.41**	**3.99**	**4.05**



APPENDIX F: Graphical view of predication results of 8 classification algorithms with 59 datasets without missing values (80% split)

According to the above results, best accuracy 77.56 is reported by LibSVM classification algorithm with discretized dataset of 112 attributes which is extracted by wrapper CfsSubsetEval feature selection algorithm with GreedyStepwise search algorithm. Also, only NaiveBayes, LibSVM, SMO, Bagging and Randomforest classifiers provided the accuracy of more than 75% with discretized datasets. Maximum average accuracy of 69.04 and minimum standard deviation 3.41 is given by the Bagging classifier.