Classification of Music Based on Correlation between Mood, Linguistic and Audio Features.

SUBMISSION DATE: 24.05.18

SUBMITTED BY:

Md. Mashrur Bari Sobhan (16373015)

Department of Computer Science and Engineering

Supervisor:

Moinul Zaber, Ph.D

Associate Professor

Department of Computer Science and Engineering

University of Dhaka.

Declaration

I hereby certify that this material, which I now submit for assessment on the programme of Master of Engineering in Computer Science and Engineering is entirely my own work, that I have exercised reasonable care to ensure that the work is original, and does not to the best of my knowledge breach any law of copyright, and has not been taken from the work of others save and to the extent that such work has been cited and acknowledged within the text of my work.

Signature of Supervisor **Signature of Author**

Moinul Zaber, Ph.D
Associate Professor
Department of Computer Science and Engineering
University of Dhaka.

Md. Mashrur Bari Sobhan
(16373015)

ABSTRACT

The emergence of music in recent times has been enviable. Some people consider music to be an integral part of their regular lives, while others sometimes even consider music to be some divine inspiration setting the mood for them for the rest of the day. For such people, a well-trimmed precise playlist of the songs that they would love to listen to, based on genre or mood of the songs, is priceless. Genre of an individual song is very much available, as that information is mostly provided within the song, but getting to judge the mood of the song is much more of a challenge. If it is a challenge itself for one distinct song, then one can easily imagine the hassle that a person faces when selecting a playlist of songs from a huge library of music. This ultimately gives rise to the importance of the classification of music based on the mood of the individual songs.

This project establishes such a method, which ultimately works with a combination of features, such as the linguistic and audio features of a song to classify a song according to the mood the song represents or is appropriate for. These features are then used in conjunction with several metrics to find out their relevance or relationships and measured for validation purposes.

Acknowledgement

Firstly, I would like to thank my supervisor Dr. Moinul Zaber of the Computer Science and Engineering Department at Dhaka University. He has encouraged me to take the necessary steps to make sure this project is a success, steered me in the right direction and has been a constant source of inspiration for me in every step of the way.

I would also like to personally thank my fellow classmates Mr. Nibras Ar Rakib and Mr. S. M. Zamshed Farhan, as without my numerous conversations and sessions of brain storming with them this project would not see the day of light today.

TABLE OF CONTENTS

1. CHAPTER 1 - INTRODUCTION 8
1.1 Background 8
1.2 Motivation 8
1.3 Problem statement 9
1.4 Solution 9
1.5 Methodology 9
1.6 Who are the stakeholders? 10
1.7 Outline of the project 10
2. CHAPTER 2 - LITERATURE REVIEW **11**
2.1 How Music can be linked with mood 11
2.1.1 Affective Norms for English Words (ANEW) Ratings 12
2.1.2 Using Valence and Arousal to define Mood 15
3. CHAPTER 3 – SYSTEM ANALYSIS AND DESIGN **17**
3.1 Requirement Analysis 17
3.1.1 Million Song Dataset 17
3.1.2 MusiXmatch Dataset 19
3.1.3 Deezer API 19
3.2 Design and Development 20
3.2.1 Machine Learning Algorithms 20
3.2.2 Web Application 22
3.2.3 Machine learning using Python 23
4. CHAPTER 4 - IMPLEMENTATION **25**
4.1 Workflow 25
4.2 Step by step implementation and results 25
4.2.1 Technology overview 26
4.2.2 Song Selection and the following processes 27
4.2.3 Final results and analysis 28
4.3 Potential improvements 31
5. CHAPTER 5 - CONCLUSION & FUTURE WORK **33**
REFERENCES **35**

LIST OF FIGURES

Figure 1: Figure showing a brief outline of the report 10

Figure 2: 2-dimensional space representation of emotion 12

Figure 3: 24 emotion terms mapped valence-arousal emotion scale 16

Figure 4: List of 55 provided fields for each track within MSD 18

Figure 5: Support Vector Machine 21

Figure 6: Workflow of the Project 25

Figure 7 & 8: A basic layout of the web application 26

Figure 9 & 10: Homepage (device) Song selection from the web application 27

Figure 11: Mood defined by the selected song, as predicted by the system 28

Fig 12: Comparison between Classification using both Linear Regression and Support Vector Machine 31

LIST OF TABLES

Table 1: the actual mood and the predicted mood defined by each of the songs when Linear Regression algorithm is applied 29

Table 2: the actual mood and the predicted mood defined by each of the songs when Support Vector Machine is applied 30

CHAPTER 1

Introduction

1.1 Background

Today the music industry is a multi-billion dollar industry, where numerous song tracks are continuously being created, updated, uploaded and shared by old and upcoming artists every minute. People, whatever they might be doing during the order of the day, are continuously listening to their favourite tracks to suit their mood. Music has become such an essential part of a human being's life that, today a morning just does not seem to get going without listening to music.

So, thriving on this constant yearning for music, this phenomenon has given rise to many start-ups with the intention of making digital music available on portable devices such as ipods, iphones, mobile devices, etc. Companies such as Deezer, Spotify, SoundCloud, etc. have become quite famous within a very short period of time, while constantly serving the purpose of their listeners by making available a constant stream of the listeners' favourite music. As these platform owners get more and more creative in their approach to their listeners – by offering custom made playlists, user specific subscription packages, to name a few – the listeners are getting more and more hooked onto these platforms as time goes by.

These online music streaming providers are consistently looking for different ways of making their online services more intuitive for the users. Playlists based on artists, genres, locations, etc. are very much available in the hands of the listeners. However, getting what one wants only gives rise to more and more aspects of needs not even known to human being at an initial stage. Such matters give rise to a whole new world of problems that such providers must take care of on a timely basis, to remain ahead of others in an extremely competitive world of music.

1.2 Motivation

Currently in platforms providing music for listeners there are plenty of easily available parameters using which a listener can trim his/her music playlist, e.g. artist names, songs names, genre of music, location of artists, etc. But one very important aspect that is currently missing is listening to music as per one's current mood. As such, today if one searches for

music on youtube or any other platforms, the best one can get is some relative information about a song which might define the mood of the song very slightly. But this ultimately never gives a definitive clue about the mood a song defines and thus gives rise to the necessity to do just that.

1.3 Problem Statement

As mentioned before, music can very well set the tune of the day for a human being. It can inspire a listener, it can revive a person if he/she feels down or may even provide condolence to a grieving person. For example, a listener is currently in need of something to inspire him/her to overcome a challenge on random day. So, in such cases, the very first step towards achieving that for people is turning to music. Now, while randomly searching through an enormous list of songs, it is very difficult for the person to recognise which songs might inspire him/her and then making a playlist out of it in the process. Therefore, as of current scenario, this is a huge problem, as such online music streaming services do not have a method of classifying music according to mood. Therefore, this becomes a tedious job any listener and ultimately results in the loss of interest of the person in question, while the ultimate purpose is not served.

1.4 Solution

This project aims to give a solution to the problem addressed earlier by classifying songs according to mood. The methods applied in achieving such classification therefore be used in different platforms to classify each and every song as per mood, provide mood tags to each of the sound tracks and thus enable a listener to filter from an enormous list of sound tracks to get the songs he/she needs as per his/her mood.

1.5 Methodology

This project takes help of an API, namely Deezer API, to get the information required for a list of songs. Using these information the system built for the purpose of this project learns about these songs. Thus, once a song within the system is selected, the system takes in a few pre-defined characteristics from the selected song and using its learning classifies the song according to mood, i.e. declares the mood defined by the song.

1.6 Who are the stakeholders?

There are millions of listeners signing up or signing in at different digital music providers' platforms every minute from around the world. As mentioned before, while genre based or artist based music tracks can be customised as per listeners' wish, it is currently impossible to customise a playlist based on a listener's preference. Using the methods applied to the system built for this project, any system can learn from a predefined set of music, using a few lyrical characteristics of the music tracks, and then define a particular track's mood efficiently. As this is done, it then gets easier for a listener to filter out the tracks he/she does not need and make a fully customised playlist of his/her preference with ease.

1.7 Outline of project

Figure 1 shows a brief overview of the report structure. To meet the objectives, a literature review was carried out and the background literature is provided in details within Chapter 2. The part of the literature review focuses on a broad discussion on how music can be linked with mood and the parameters required to quantify music according to mood.

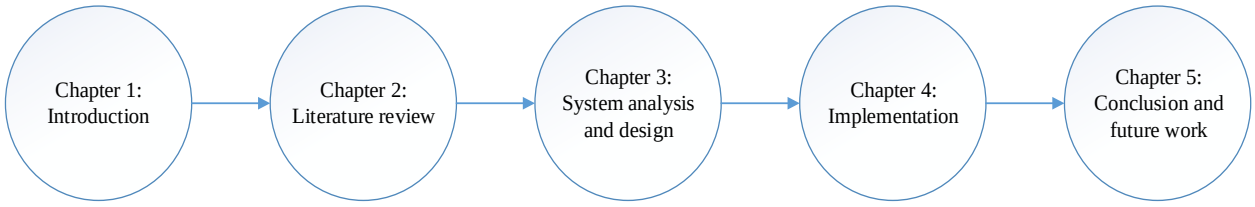

Fig 1: Figure shows a brief outline of the report

Chapter 3 defines the system analysis of the project. The functional and non-functional requirements, including the data model, source of data, classification methods, and system design of the proposed tools are introduced in Chapter 3. The implementation of the web application, user interface evolvement, the results and analysis of the results along with potential improvements is the subject of Chapter 4. Chapter 5 concludes the report by discussing the findings and proposing future research endeavour in the area of classification of music by mood using other characteristics of music.

CHAPTER 2

Literature Review

2.1 How Music can be linked with mood

For ages music has continued to be a vital part of our lives. We have been listening to music since the early ages of the history of human being's existence. The trend has never stopped, rather grew more and more with time, encompassing the whole of the earth gradually with time. Not too many of us remember vinyl or 8tracks. We are now a streaming generation with our iTunes and Spotify apps a click away on our smart devices. The music industry has also changed quite a bit over the last couple of decades, all thanks to new technology. As discussed before, instead of making records available to be bought from music stores, records and albums these days are uploaded and made available at digital music records platforms such as Spotify, Deezer, SoundCloud, iTunes, etc.

Thus, with time the tendency towards listening to more and more music has only increased, so much so that, music has effectively been successful in defining a listener's mood at any particular moment in time. While listening to music may bring greater health benefits, creating it can be an effective therapy, too. Example includes a unique orchestra for people with dementia helping it improve their mood and boost their self-confidence, according to researchers at the Bournemouth University Dementia Institute (BUDI) in Dorset, U.K. The orchestra is one of several BUDI research projects that aims to demonstrate how people with dementia can still learn new skills and have fun. Eight people with dementia and seven caregivers participated in the project, along with students and professional musicians.

A 2013 study in the Journal of Positive Psychology found that people who listened to upbeat music could improve their moods and boost their happiness in just two weeks. In the study, participants were instructed to try to improve their mood, but they only succeeded when they listened to the upbeat music of Copland as opposed to the sadder tunes of Stravinsky [1]. On the other hand, even sad music brings most listeners pleasure and comfort, according to recent research from Durham University in the United Kingdom and the University of Jyväskylä in Finland, published in PLOS ONE. Conversely, the study found that for some people, sad music can cause negative feelings of profound grief. The research involved three

surveys of more than 2,400 people in the United Kingdom and Finland, focusing on the emotions and memorable experiences associated with listening to sad songs [2].

Now that it has been established that there is a tangible connection between music and mood of a listener, there needs to be a way to quantify music as per mood. This ultimately gave rise to the Affective Norms for English Words (ANEW) ratings.

2.1.1 Affective Norms for English Words (ANEW) Ratings

The most common method of quantifying an emotional state is by associating it with a point in a 2-dimensional space with valence (attractiveness/aversion) and arousal (energy) as dimensions, as stated by Russell [4]. An example of such a 2-dimensional space representation is provided below:

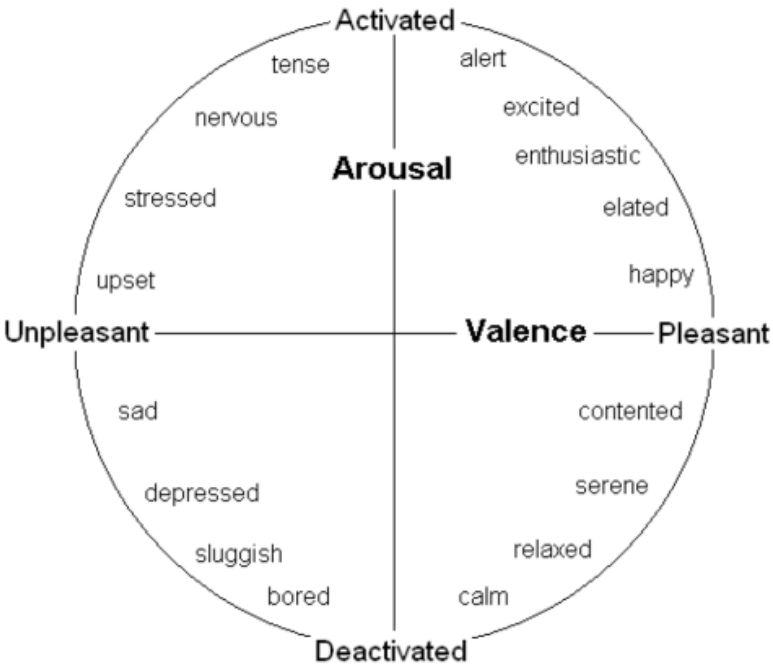

Fig 2: 2-dimensional space representation of emotion

In such a space a high valence value corresponds to positive mood, such as ‘happy’ or ‘elated’, and high arousal values depict energetic state, e.g. ‘excited’. For a particular song its valence and arousal values can be plotted on this 2-dimensional space to make conclusive deductions about the song’s mood representations. The mean valence and arousal values of the words can be found by taking the help of the Affective Norms for English (ANEW) dictionary.

The Affective Norms for English Words (ANEW) was developed to provide a set of normative emotional ratings for a large number of words in the English language [3]. It is a commonly used set of 1,034 words characterised on the affective dimensions of valence, arousal, and dominance. Traditionally, studies of affect have used stimuli characterised along either affective dimensions or discrete emotional categories, but much current research draws on both of these perspectives. The existence of these affective collections should help in comparing results across different investigations of emotion, as well as in allowing replication within and across research labs assessing basic or applied problems in the study of emotion [3]. So, in simple words, human evaluations of the 'happiness' level of individual words can be drawn directly using the Affective Norms for English Words (ANEW) study.

During this study to form ANEW, to assess the three dimensions of pleasure, arousal, and dominance, the Self-Assessment Manikin (SAM), an affective rating system originally devised by Lang (1980) was used [3]. Bradley and Lang (1990) determined that the SAM correlates well with factors of pleasure and arousal obtained using the longer, verbal Semantic Differential Scale (Mehrabian and Russell, 1974. The graphic SAM figures comprise bipolar scales that depict different values along each emotional dimension. For the pleasure dimension, SAM ranges from a smiling, happy figure to a frowning, unhappy figure; to represent the arousal dimension, SAM ranges from an excited, wide-eyed figure to a relaxed, sleepy figure. For the dominance dimension, SAM ranges from a large figure (in control) to a small figure (dominated).

In this version of SAM, the subject can bubble in the location corresponding to any of the 5 figures on each scale, or between any two figures, which results in a 9-point rating scale for each dimension. In addition to the ScanSAM version of SAM, there is a booklet paper-and-pencil version, as well as a dynamic computer version written for IBM-compatible written and distributed by Ed Cook at the University of Alabama-Birmingham (e.g., Cook, Atkinson, & Lang, 1987). The computer SAM scale uses a 20-point scale, rendering more points along each dimension than the ScanSAM or paper-and-pencil versions. Using SAM, subjects have rated the words currently in the ANEW on the dimensions of pleasure, arousal, and dominance. There are several characteristic features of the resulting space.

First, these stimulus materials evoke reactions across the entire range of each dimension: mean pleasure ratings for these words range from very unpleasant to very pleasant, and are distributed fairly evenly across the space. Secondly, a wide range of arousal levels are elicited by these materials. For items rated as neutral in valence (i.e., those occurring at and near the midline of the valence dimension), arousal ratings do not attain the high levels

associated with either pleasant or unpleasant materials. The distribution of the ANEW words in affective space is quite similar to that obtained with IAPS and IADS stimuli, suggesting that this space is representative of the affective distribution of materials from a number of different modalities.

In general, each experiment presented 100-150 different words [3]. Each set of words was prepared on an 8 1/2 x 11 sheet of paper that presented individual words in 4 columns and 14 rows (i.e. 56 words per sheet). The words were sequentially numbered in rows from 1 to n. For each set of words, two forms were prepared which counterbalanced the order in which any specific word was rated, and the items surrounding any specific word. SAM ratings of pleasure, arousal, and dominance were collected in a self-paced procedure, in which the subject silently read each word, and then bubbled in their emotional ratings on the ScanSam sheet.

Subjects. Introductory Psychology class students, balanced for gender, participated as part of a course requirement [3].

Design. Subjects were run in groups ranging in size from 8 to 25, with the male:female ratio no more than 1:2 (or 2:1) for any single group session. The ScanSAM version of the SAM instrument was used. In this format, the (unlabelled) dimensions of pleasure, arousal and dominance are graphically rendered, with each of the 3 dimensions ordinally scaled with 5 figures. The subject can bubble in 1 of 9 circles on the ScanSam sheet, with bubbles between graphic figures indicating a level intermediate between the 2 graphically depicted levels for each dimension [3].

Materials and Equipments. Among the words currently included in the ANEW are approximately 150 words used in Mehrabian & Russell (1974) and 450 words used in Bellezza, Greenwald, & Banaji, (1986) [3].

Procedure. Subjects received the sheets containing the words to-be-rated, and a packet of ScanSam rating sheets. Instructions regarding ScanSam were given, and each subject worked at his or her own pace rating the words in the word booklet [3].

Now, using the values from the ANEW dictionary, calculations can be made using the method as described below [5].

If we consider l_i= (w_1, w_2...w_n) be the i^{th} lyric, consisting of n_i words and that in our complete collection of lyrics of the songs we have a set of {l_1, l_2 ... l_m} lyrics; then the valence, v_i, and arousal, a_i, of the lyric i can calculated as,

$$\boldsymbol{v_i = \frac{1}{ni}\sum_{j=1}^{ni} V(w_n),\ a_i = \frac{1}{ni}\sum_{j=1}^{ni} A(w_n),\ i = 1...m,}$$

Where *'V'* and *'A'* are the mean values of valence and arousal, respectively, for each of the words. Thus, using this method described above, any music track can be quantified using its lyrical characteristics.

2.1.2 Using Valence and Arousal to define Mood

Now the word emotion itself can be understood in a variety of ways, and it is used colloquially for a range of things varying from mild to intense, simple to complex, brief to extended, and private to public. In the science of emotion, or affective science, different words are used to distinguish different, related phenomena [6]:

- *Affect* is a wide umbrella word, often used to encompass the different phenomena concerning *valence*, i.e. positive and negative, internal states.
- *Attitudes* are the most stable beliefs held by an individual about the valence of things.
- *Moods* are passing and more long-term states, often not directly or simply related to a specific cause.
- *Emotions* are the most short-lived reactions and responses to events and situations, reflecting the current goals, attitudes and mood of the individual, and they work to *appraise* the situation. Emotions can further be explained as the conscious *feeling*, the *behaviour* an emotion causes, and its *physiological manifestation*.

The exact set and amount of basic emotions is debated, and no clear solution exists to determine the best set. But in one of the frameworks a method was proposed by J. R. Fontaine et al. [7], with wide set of emotion categories which identifies 24 emotion terms that are commonly found in emotion research and everyday language, which are mapped onto the 2-dimensional emotion space in the Figure below:

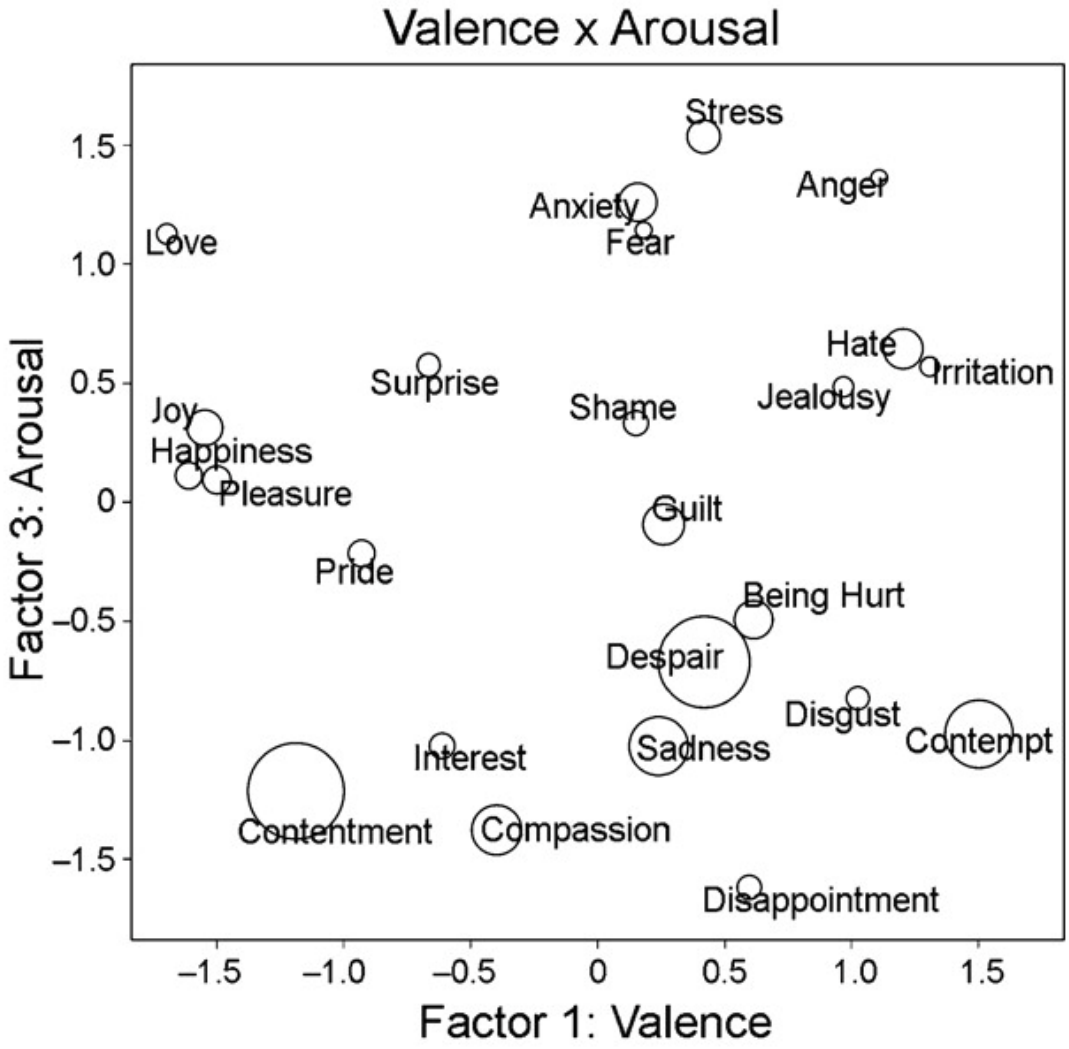

Fig 3: 24 emotion terms mapped onto the 2-dimensional valence-arousal emotion scale, based on answers from subjects speaking three different languages: English, Dutch and French. The size of each circle represents the mean distance between the ratings of different language groups, i.e. a smaller circle means that the ratings were more consistent across languages. Image from J. R. Fontaine et al. [7].

Therefore, using this 2-dimensional valence-arousal emotion scale and the valence-arousal values previously received from the ANEW dictionary for a music track, the music track can be classified according to mood.

CHAPTER 3
System Analysis and Design

3.1 Requirement Analysis

As defined before, using the valence and arousal values for a particular sound track, the track can be classified according to mood. Therefore, to carry out this activity for this project, a dataset of songs is required. There are numerous dataset sources available on the internet, which have been made publicly available for free, which consist of a list of characteristics for numerous songs. But, obviously not all these characteristics are necessary to serve the purpose of this project, while, in most cases the dataset was found to consist of data which of no relevance to the intentions of this project at all. Therefore, after careful considerations, two such sources of dataset for music were found to be of interest.

3.1.1 Million Song Dataset

The Million Song Data (MSD) set is a collection of audio features and metadata for a million popular tracks [8]. It is merely an attempt to help researchers by providing a large-scale dataset. The MSD contains metadata and audio analysis for a million songs that were legally available to The Echo Nest. The songs are representative of recent western commercial music. The main purposes of the dataset are [8]:

- To encourage research on algorithms that scale to commercial sizes,
- To provide a reference dataset for evaluating research,
- As a shortcut alternative to creating a large dataset with APIs (e.g. The Echo Nest's)
- To help new researchers get started in the MIR field

The dataset contains an enormous list of music tracks and each of the music tracks contains characteristics as provided below:

analysis_sample_rate	artist_7digitalid
artist_familiarity	artist_hotttnesss
artist_id	artist_latitude
artist_location	artist_longitude
artist_mbid	artist_mbtags
artist_mbtags_count	artist_name
artist_playmeid	artist_terms
artist_terms_freq	artist_terms_weight
audio_md5	bars_confidence
bars_start	beats_confidence
beats_start	danceability
duration	end_of_fade_in
energy	key
key_confidence	loudness
mode	mode_confidence
num_songs	release
release_7digitalid	sections_confidence
sections_start	segments_confidence
segments_loudness_max	segments_loudness_max_time
segments_loudness_start	segments_pitches
segments_start	segments_timbre
similar_artists	song_hotttnesss
song_id	start_of_fade_out
tatums_confidence	tatums_start
tempo	time_signature
time_signature_confidence	title
track_7digitalid	track_id
year	

Fig 4: List of 55 provided fields for each track within MSD [8]

While this is a great attempt from the publishers at making all these data available for free only for research purposes, this alone does not suffice the needs of this project. As it can be seen from the list above, the MSD does not provide any lyrical features for a particular track. Therefore, to achieve that, this list of tracks is required to be cross matched with another particular dataset namely MusiXmatch.

3.1.2 MusiXmatch Dataset

The MusiXMatch dataset provides lyrics for many Million Song Dataset tracks [9]. The lyrics come in bag-of-words format: each track is described as the word-counts for a dictionary of the top 5,000 words across the set. The dataset comes in two text files, describing training and test sets. The split is done according to the split for tagging, see tagging test artists. There are 210,519 training bag-of-words, 27,143 testing ones. They also provide the full list of words with total counts across all tracks so one can measure the relative importance of the top 5,000.

Now, using this dataset along with the MSD dataset one can make a combined dataset of music tracks containing the required lyrical features. Yet, this does not serve the purpose of this project, since, using the lyrical features, the number of times each word occurs within the lyrics, a time and resource consuming calculation is thus required to be carried out for each tracks to give an approximate value of valence and arousal for each of the tracks. This ultimately becomes a tedious task and gives rise to more and more complexities along the way. Therefore, while looking for an alternative way to find all these features together at one place, a solution namely Deezer API came along.

3.1.3 Deezer API

Deezer is an Internet-based music streaming service. It allows users to listen to music content from record labels including Sony, Universal Music Group, and Warner Music Group on various devices online or offline. Deezer has made a few methods available for developers to make use of their resources to get data about their huge library of songs. One of these resources is the Deezer API.

Deezer API is a simple api that provides a nice set of services to build up web applications allowing the discovery of Deezer's music catalogue [10]. Using the parameters as per requirement, the api provides a response in json format containing information such as, valence, arousal, song name, Million Song Database ID, artist name, etc.

Since, the Deezer API serves the purpose of this project, this is therefore chosen as the data source, since there is enough data present within this system, approximately more than 14000 songs, provides all the information as required for this project.

3.2 Design and Development

Once the data requirement is fulfilled for the project, the next step is to ensure pre-processing of the data. Now since for each of the music tracks the valence and arousal values are available, the dataset is then cross matched with the 24 emotion terms mapped onto the 2-dimensional valence-arousal emotion scale, as introduced in Section 2.1.2. The challenge here was that, for all the songs the valence and arousal values did not fall within the range consisted within the emotions provided. Thus, the music tracks for which these values did not exist were filtered out, and for the remaining tracks, consisting around 350 songs, the mood is loaded onto the dataset.

At this point the target is load up the system built with this data, containing the valence and arousal values of the provided songs and use some machine learning algorithm to help the system learn about these characteristics of the provided songs. Once the system learns about these characteristics from the training data provided, then it will be able to correctly predict the mood defined by a particular song.

3.2.1 Machine Learning Algorithms

For the purpose of this project two separate machine learning models were selected, Linear Regression model and Support Vector Machine.

Linear regression is perhaps one of the most well-known and well understood algorithms in statistics and machine learning. It is a linear model, i.e. a model that assumes a linear relationship between the input variables (x) and the single output variable (y).

More specifically, that y can be calculated from a linear combination of the input variables (x).When there is a single input variable (x), the method is referred to as simple linear regression. When there are multiple input variables, literature from statistics often refers to the method as multiple linear regression.

The representation is a linear equation that combines a specific set of input values (x) the solution to which is the predicted output for that set of input values (y). As such, both the input values (x) and the output value are numeric.

For example, in a simple regression problem (a single x and a single y), the form of the model would be:

$$y = B0 + B1*x$$

In higher dimensions when there are more than one input (x), the line is called a plane or a hyper-plane. The representation therefore is the form of the equation and the specific values used for the coefficients (e.g. B0 and B1 in the above example).

Support Vector Machine (SVM) is a supervised machine learning algorithm which can be used for both classification and regression challenges. However, it is mostly used in classification problems. This algorithm plots each data item as a point in n-dimensional space (where n is number of features you have) with the value of each feature being the value of a particular coordinate. Then, it performs classification by finding the hyper-plane that differentiate the two classes very well as provided below in the figure.

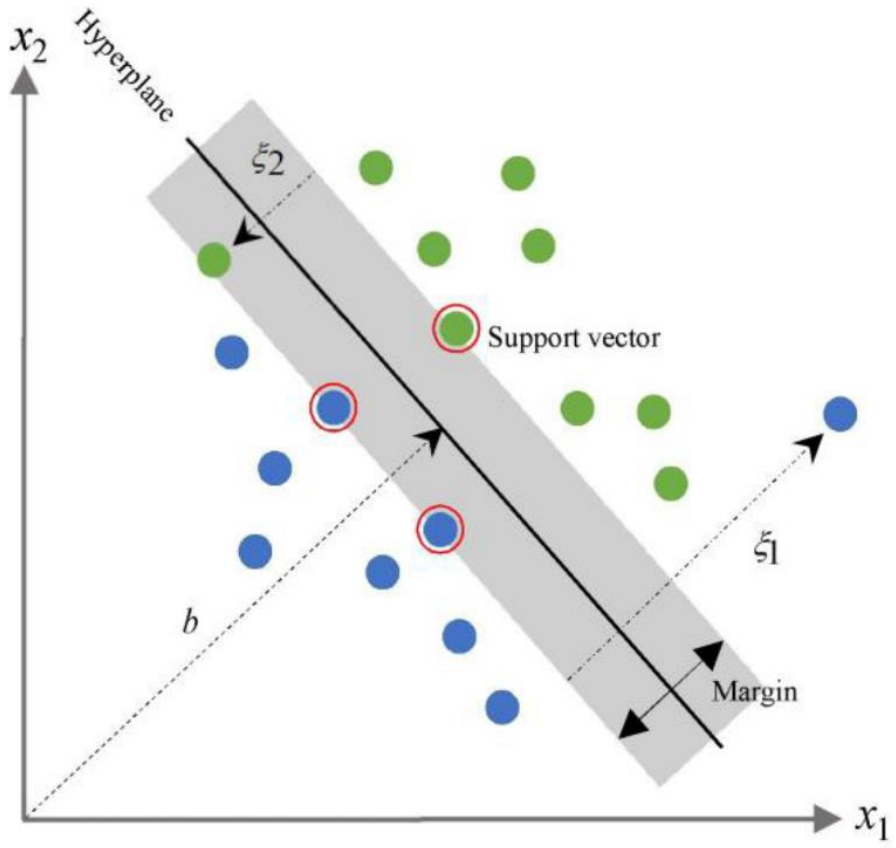

Fig 5: Support Vector Machine

The hyper-plane splits the input variable space. In SVM, a hyper-plane is selected to best separate the points in the input variable space by their class, either class 0 or class 1. In two-dimensions one can visualise this as a line and that all of the input points can be completely separated by this line. For example:

B0 + (B1 * X1) + (B2 * X2) = 0

Where the coefficients (B1 and B2) that determine the slope of the line and the intercept (B0) are found by the learning algorithm, and X1 and X2 are the two input variables. The distance between the line and the closest data points is referred to as the margin. The best or optimal line that can separate the two classes is the line that as the largest margin. This is called the Maximal-Margin hyper-plane. The margin is calculated as the perpendicular distance from the line to only the closest points. Only these points are relevant in defining the line and in the construction of the classifier. These points are called the Support Vectors. They support or define the hyper-plane. The hyper-plane is learned from training data using an optimisation procedure that maximises the margin. The advantages of support vector machines are:

- Effective in high dimensional spaces.
- Still effective in cases where number of dimensions is greater than the number of samples.
- Uses a subset of training points in the decision function (called support vectors), so it is also memory efficient.
- Versatile: different Kernel functions can be specified for the decision function. Common kernels are provided, but it is also possible to specify custom kernels.

The disadvantages of support vector machines include:

- If the number of features is much greater than the number of samples, avoid over-fitting in choosing Kernel functions and regularisation term is crucial.
- SVMs do not directly provide probability estimates, these are calculated using an expensive five-fold cross-validation.

3.2.2 Web Application

Once the dataset is ready and filled up, consisting of the required data only, ready to be trained, the next step is to build an application. For the purpose of this project, this application is a web application. The web application is made up of a simple form designed using Bootstrap and made functional using Javascript. The form is very much an easy-to-use and an intuitive one, consisting of a few details and a dropdown menu containing a list of songs. Once a song at random is chosen and the form is submitted, the system takes in the input consisting of the

selected song and its valence-arousal values. A script is then run which runs the machine learning algorithm on the training dataset, learns from the dataset and tries to predict the selected song's defining mood using the song's valence and arousal values.

3.2.3 Machine learning using Python

The machine learning model implemented for the purpose of this project is built using python. Python is an interpreted high-level programming language for general-purpose programming. Python comes with a huge amount of inbuilt libraries. Many of the libraries are for Artificial Intelligence and Machine Learning. Some of the libraries are Tensorflow (which is high-level neural network library), scikit-learn (for data mining, data analysis and machine learning), pylearn2 (more flexible than scikit-learn), etc.

So, within the system built for this project, once the form in the web-application is submitted, a python script is run. The libraries used for the purpose of this project within python are defined below:

a) **NumPy:** NumPy is the fundamental package for scientific computing with Python. It contains among other things:

- a powerful N-dimensional array object
- sophisticated (broadcasting) functions
- tools for integrating C/C++ and Fortran code
- useful linear algebra, Fourier transform, and random number capabilities

Besides its obvious scientific uses, NumPy can also be used as an efficient multi-dimensional container of generic data. Arbitrary data-types can be defined. This allows NumPy to seamlessly and speedily integrate with a wide variety of databases.

b) **pandas:** *pandas* is an open source, BSD-licensed library providing high-performance, easy-to-use data structures and data analysis tools for the Python programming language.

c) **sys:** system-specific parameters and functions

d) **scikit-learn:** offers simple and efficient tools for data mining, data analysis and machine learning

e) **LabelEncoder:** Encode labels with value between 0 and n_classes-1.

f) **OneHotEncoder:** Encode categorical integer features using a one-hot aka one-of-K scheme.

g) **linear_model:** Ordinary least squares Linear Regression.

h) **train_test_split:** Split arrays or matrices into random train and test subsets

i) **svm:** provides support vector machine classification functionalities in python

After running the script, the system gets trained and then tries to predict the mood the song, selected previously while submitting the form from the web-application, defines.

CHAPTER 4

Implementation

4.1 Workflow

The brief overview of the workflow representing the steps taken during the course of the project is provided below:

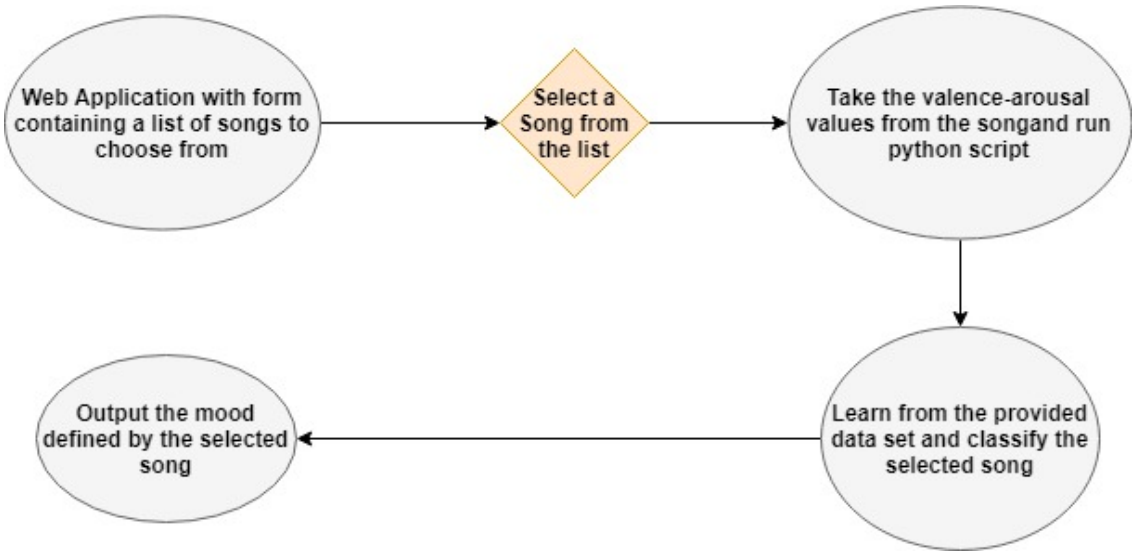

Fig 6: Workflow of the Project

4.2 Step by step implementation and results

As pointed out within the workflow in the previous section, after the pre-processing of the data, containing only the required characteristics (fields) of the music tracks, it is time to move onto the web application.

4.2.1 Technology overview

The Web application has been designed mainly using a custom Bootstrap form, with basic CSS to give it lovely interface, and some Javascript to make it much easy to use. Provided below are two figures showing a basic layout of the web application.

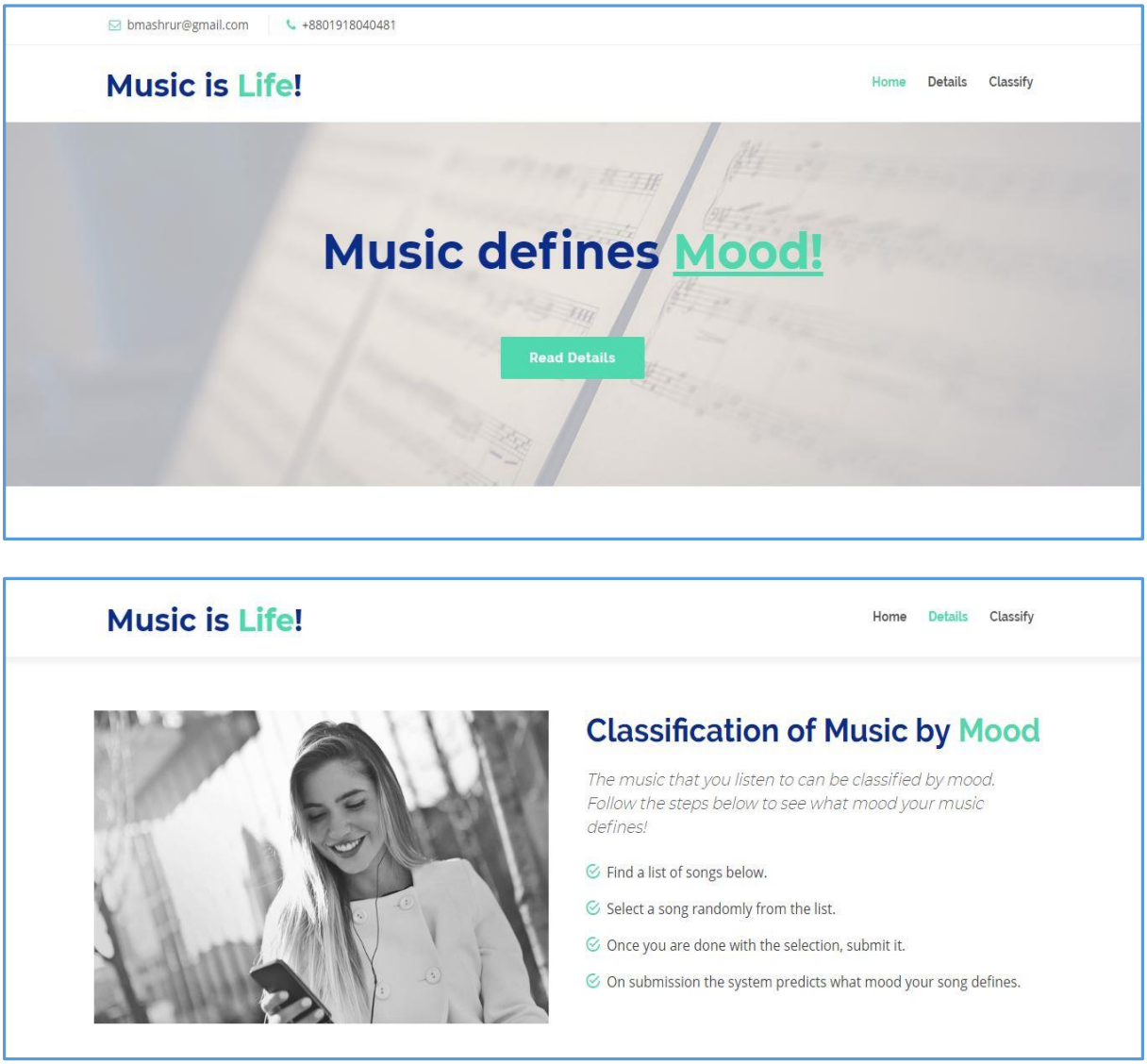

Fig 7 & 8: A basic layout of the web application

In the backend the website mainly uses PHP to help the data flow. The whole web application is a very simple, easy to use, user interactive two page application that basically takes input at one page and shows the output onto the other page.

The machine learning portion is carried out by using Python, with the help of the libraries mentioned in Section 3.2.3. This is where the main work takes place.

4.2.2 Song Selection and the following processes

After browsing through the contents of the web application, once the user reaches almost near the end of the page, the user gets a dropdown menu where he/she can choose a song from the list contained within the dropdown menu. This is shown on the figures below:

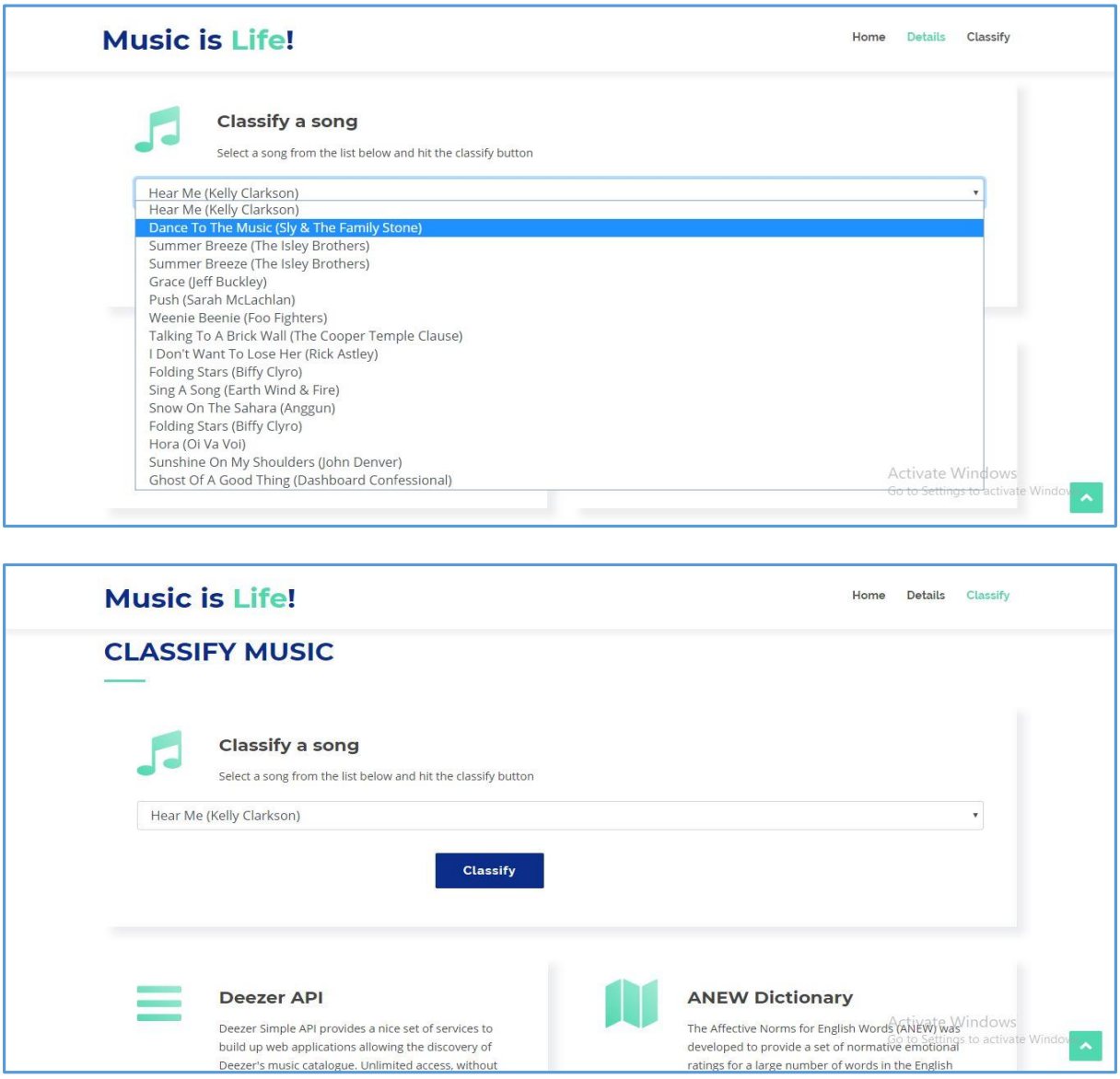

Fig 9 & 10: Song selection from the web application

Once a song has been selected randomly from the dropdown menu, and the form is submitted using the “Classify” button, the system receives the song and the valence-arousal values along associated with the song.

After receiving the valence-arousal for the song, the system runs a python script. Here the system attempts to read through a csv file containing the original dataset containing around 345 songs’ valence-arousal values and the mood defined by those individual songs. Using machine learning algorithms, such as Linear Regression and Support Vector Machine (SVM), the system learns about those songs characteristics. Using this learning the system then runs the valence-arousal values fetched from the selected songs against its learnings and then predicts what mood the selected song defines.

4.2.3 Final results and analysis

Once the system predicts the mood defined by the selected song, it then fetches another web page and prints the predicted value, i.e. mood defined by the selected song, out onto that web page. This layout is shown below:

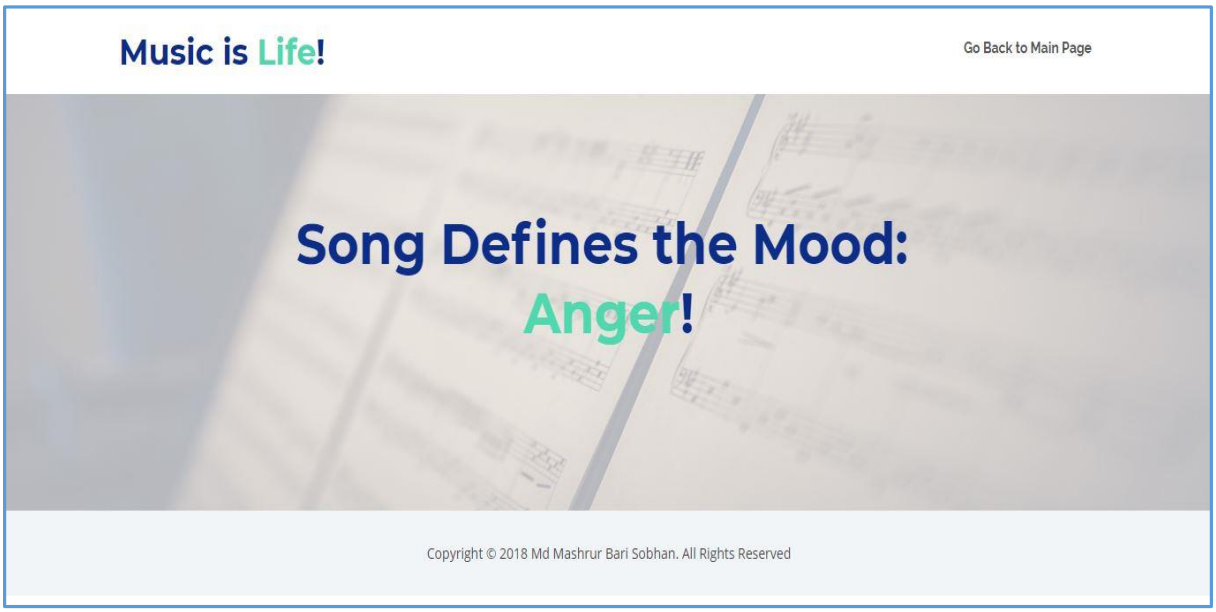

Fig 11: Mood defined by the selected song, as predicted by the system

All looks fantastic at this point, and the overall system behaves as per wish. Yet, there is a catch here. The prediction made by the system differs when different machine learning algorithms are applied.

For the purpose of this project, there were two machine learning algorithms applied onto learning the dataset, namely Linear Regression and Support Vector Machine (SVM).

For analysis purposes, a list is provided below showing the actual mood and the predicted mood defined by each of the songs when Linear Regression algorithm is applied:

song_name	artist	valence	arousal	mood	mood_predicted
Hear Me	Kelly Clarkson	-1.52	0.45	Joy	**Joy**
Dance To The Music	Sly & The Family Stone	1.18	1.39	Anger	Compassion
Summer Breeze	The Isley Brothers	0.89	-0.44	Being Hurt	Happiness
Grace	Jeff Buckley	-0.91	-0.23	Pride	Joy
Push	Sarah McLachlan	1	0.54	Jealousy	Despair
Weenie Beenie	Foo Fighters	0.19	0.35	Shame	Happiness
Talking To A Brick Wall	The Cooper Temple Clause	-1.35	0.14	Pleasure	Joy
I Don't Want To Lose Her	Rick Astley	1.28	1.44	Anger	Being Hurt
Folding Stars	Biffy Clyro	-0.73	0.58	Surprise	Interest
Sing A Song	Earth Wind & Fire	1.23	1.56	Anger	Being Hurt
Snow On The Sahara	Anggun	0.18	0.38	Shame	Happiness
Hora	Oi Va Voi	0.46	-0.6	Despair	Interest
Sunshine On My Shoulders	John Denver	0.24	1.22	Fear	**Fear**
Ghost Of A Good Thing	Dashboard Confessional	-0.46	-0.94	Interest	Joy

Table 1: Table shows the actual mood and the predicted mood defined by each of the songs when Linear Regression algorithm is applied

From the table above it can be seen that, in the case of Linear Regression, the result comes out to be correct on only two occasions. Thus, it can be deduced that, for the case of this dataset, Linear regression performs poorly and therefore other algorithm needs to be tested out to see if better results can be achieved with those

Hence comes the use of the machine learning algorithm, Support Vector Machine (SVM). When the system uses SVM, the outputs of the system are listed out below:

song_name	artist	valence	arousal	mood	mood_predicted
Hear Me	Kelly Clarkson	-1.52	0.45	Joy	Happiness
Dance To The Music	Sly & The Family Stone	1.18	1.39	Anger	**Anger**
Summer Breeze	The Isley Brothers	0.89	-0.44	Being Hurt	**Being Hurt**
Grace	Jeff Buckley	-0.91	-0.23	Pride	**Pride**
Push	Sarah McLachlan	1	0.54	Jealousy	**Jealousy**
Weenie Beenie	Foo Fighters	0.19	0.35	Shame	Fear
Talking To A Brick Wall	The Cooper Temple Clause	-1.35	0.14	Pleasure	Happiness
I Don't Want To Lose Her	Rick Astley	1.28	1.44	Anger	**Anger**
Folding Stars	Biffy Clyro	-0.73	0.58	Surprise	Pride
Sing A Song	Earth Wind & Fire	1.23	1.56	Anger	**Anger**
Snow On The Sahara	Anggun	0.18	0.38	Shame	Fear
Hora	Oi Va Voi	0.46	-0.6	Despair	**Despair**
Sunshine On My Shoulders	John Denver	0.24	1.22	Fear	**Fear**
Ghost Of A Good Thing	Dashboard Confessional	-0.46	-0.94	Interest	**Interest**

Table 2: Table shows the actual mood and the predicted mood defined by each of the songs when Support Vector Machine is applied

As it can be seen from the outputs above, SVM performs admirably better than Linear Regression. Below is a bar chart showing the comparison between the performances for both the algorithms.

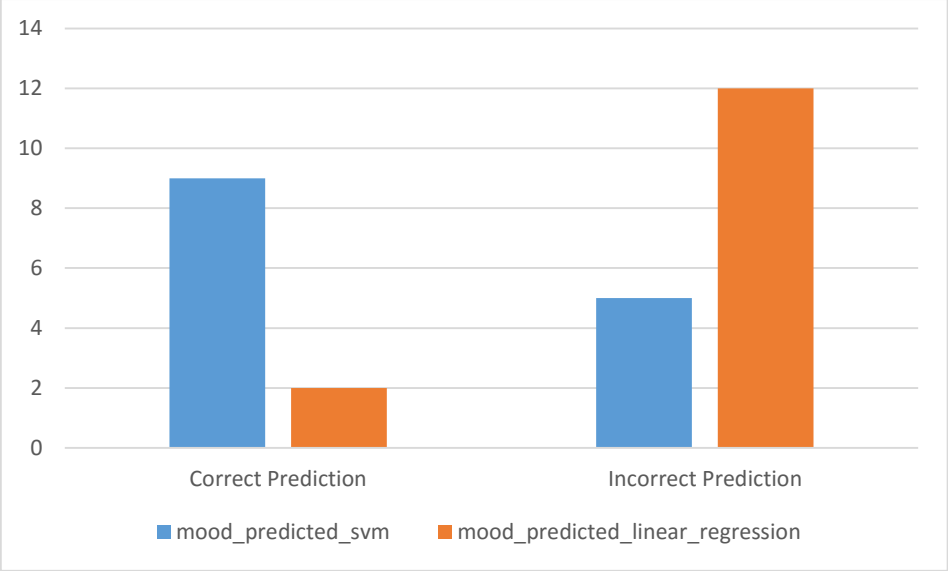

Fig 12: Comparison between Classification using both Linear Regression and Support Vector Machine

Choice of a machine learning algorithm to be used to learn about a particular dataset (training) and then providing predictions (testing) depends on the many aspects, and one of the major aspects is the dataset itself. While comparing these results, a deduction can be made that, the provided dataset is not data won't be linearly separable, which is why Linear Regression might not be performing as well as SVM.

4.3 Potential improvements

There is always a room improvement in method and/or model, and the model proposed in this project is by no means any different. There is a definite potential for further improvements here. Since, the Million Song Dataset, introduced with Section 3.1.1 of this report, contains a lot of characteristics of a song, 55 fields to be exact, and thus contains a few very interesting characteristics of a music track. Some of these fields are loudness, tempo, danceability, etc which can further define the mood proposed by a music track. These fields have not be considered to be important for the purpose of this project since, currently there are no available validating data provided which ultimately gives an indication as to what value of loudness, tempo or danceability defines which mood. Therefore, due to time and resource constraints, these factors have not been taken under consideration. But this project certainly

give rise to the idea and potential of how these factors can even further define mood of a song, and with the availability of further resources can be worked with in the future, and hence improve this project as a whole.

CHAPTER 5
Conclusion & Future Work

As a form of art and cultural activity, music is a major part of the daily life of many people. The demand for music consumption has created a billion-dollar global music industry, which encompasses music production and distribution, major record companies such as – Warner Chappell, Universal, and Sony/ATV being the largest music publishers in the world – live concerts, along with other music-related activities. In the U.S., the music industry was estimated to generate about 17.2 billion U.S. dollars in 2016. Forecasts show a slight growth in the coming years; by 2021, it is expected that the music industry revenue in the U.S. will total over 22.6 billion U.S. dollars. This phenomenon is apparently predicted to rise even further after that. Thus, the importance of music to people's lives is very well established by now.

With the rise of availability of music and the sharing of music digitally, the phenomenon has created a constant yearning within the millions of listeners listening to music every minute around the globe. This has ultimately given rise to the constant need for evolution within the intuitiveness and user interaction in the digital music streaming platforms such as Spotify, SoundCloud, Deezer, etc. Therefore, it does not need be defined any further as to how much of potential an innovation in this sector carries.

This project is just a step towards making such an innovation a reality and realising that potential eventually. Currently, music enthusiasts can customise their music playlists as per genre, artist names, band names, locations, etc. But with time gradually, people have started querying about how they can customise a music playlist according to their moods. This project is a baby step towards solving that problem and thus creating a revolution in the process.

As discussed before, the results achieved through carrying out this project can be further improved with time, by considering plenty of other characteristics of individual music tracks. But the challenge remains in the validation process for both the training and testing data. A crowd sourced response can be an important step towards achieving that, but the process is a very time consuming and a resource consuming one. The potential for such a system is huge, as one day a person might not even have to input his/her mood himself/herself, where the

system might be able to suggest him/her a playlist of songs by detecting or predicting his/her mood over a particular period of time.

REFERENCES

1. Yuna L. Ferguson & Kennon M. Sheldon (2012) Trying to be happier really can work: Two experimental studies, The Journal of Positive Psychology, 8:1, 23-33, DOI: 10.1080/17439760.2012.747000
2. Tuomas Eerola & Henna-Riikka Peltola (2016) Memorable Experiences with Sad Music—Reasons, Reactions and Mechanisms of Three Types of Experiences, PLOS ONE, https://doi.org/10.1371/journal.pone.0157444
3. Bradley, M. M., & Lang, P. J. (1999). "Affective norms for English words (ANEW): Instruction manual and affetctive ratings". Technical Report C-1, TheCenter for Research in Psychology, University of Florida.
4. J.A. Russell. "A circumplex model of affect". *Journal of personality and social psychology, 39(6):1161, 1980.*
5. McVicar, M., Freeman, T., De Bie, T.: "Mining the correlation between lyrical and audio features and the emergence of mood". *In: Proc.* ISMIR. pp. 783–788 (2011)
6. Valtteri Wikström (2015), Emotion Transfer Protocol, Experiments in emotion transmission, http://vatte.github.io/etp/
7. Fontaine, Johnny RJ, Klaus R Scherer, Etienne B Roesch, and Phoebe C Ellsworth. 2007. "The World of Emotions Is Not Two-Dimensional." Psychological Science 18 (12). SAGE Publications: 1050–57.
8. Thierry Bertin-Mahieux, Daniel P.W. Ellis, Brian Whitman, and Paul Lamere. The Million Song Dataset. *In Proceedings*: 12th International Society for Music Information Retrieval Conference (ISMIR 2011), 2011.
9. The MusiXmatch Dataset (Aug-2015), http://labrosa.ee.columbia.edu/millionsong/musixmatch
10. Deezer API, https://developers.deezer.com/api
11. https://github.com/deezer/deezer_mood_detection_dataset