**EMOTION AWARE MUSIC RECOMMENDATION SYSTEM**

*A*

*Project Report*

*Submitted in partial fulfilment of the*

*Requirements for the award of the Degree of*

**BACHELOR OF ENGINEERING**

IN

**COMPUTER SCIENCE & ENGINEERING**

By

*Under the guidance of*

**A close-up of a logo

Description automatically generated with medium confidence**

**Department of Computer Science & Engineering**

**Vasavi College of Engineering (Autonomous)**

**(Affiliated to Osmania University)**

**Hyderabad-500 031**

**Department of Computer Science & Engineering**

**A close-up of a logo

Description automatically generated with medium confidence**

**DECLARATION BY THE CANDIDATE**

**Vasavi College of Engineering (Autonomous)**

**(Affiliated to Osmania University)**

**Hyderabad-500 031**

**Department of Computer Science & Engineering**

**A close-up of a logo

Description automatically generated with medium confidence**

**BONAFIDE CERTIFICATE**

**Acknowledgement**

**Abstract**

Music is an essential part of our regular life. It cheers us up and makes us feel better. Not all forms of music are appropriate for every mood. Furthermore, ever-growing digital music catalogues make it virtually impossible to recollect a specific tune that fits the present emotion. Besides that, due to the enormous number of songs accessible, people are frequently perplexed when selecting a track. This necessitates the development of a context-sensitive music recommendation system.

Therefore, we present a context-aware music recommendation system that assists in identifying the user's current emotion and suggesting music which is relevant to that emotion. We have come up with a comprehensive strategy to improve user preference prediction; our technique integrates context and emotion elements and strives to give users a more convenient, intuitive, and pleasurable listening experience. Finally, we discuss the evaluation and performance metrics and results of our research.

**7. Table of Contents**

**Abstract 5**

**List of Tables 8**

**List of Figures 9**

1. **Introduction 12**
   1. Facial Emotion Identification **13**
   2. Music Recommendation **16**
      1. Collaborative Filtering **16**
      2. Content based Filtering **16**
2. **Overview of proposed system 17**
   1. Emotion Identification **17**
   2. Music Recommendation **18**
3. **System design 19**
   1. Emotion Identification Module **19**
   2. Music Tagging & Recommendation Module **20**
   3. User-Interface Module **21**
4. **References 59**

**List of Tables**

**Table 2.1.1 :** Compare & Contrast of Emotion DetectionModels **17**

**List of Figures**

**Figure 1.1.1 :** Basic model of how Neuronsin ANN work **13**

**Figure 1.1.2 :** Processing of a digital image byNeural Network **14**

**Figure 1.1.3 :** Emotion Identification by aNeural Network **15**

**Figure 1.1.4 :** Neural Network involvingmore hidden layers **15**

**Figure 1.2.1 :** Collaborative Filtering **16**

**Figure 1.2.2 :** Content-based Filtering **16**

**Figure 3.1.1 :** Emotion Identification module **19**

**Figure 3.2.1 :** Music Recommendation module **20**

**Figure 3.3.1 :** User-Interface Module **21**

**1. Introduction**

Music is universal and easily accessible in our everyday lives, thanks to the tremendous growth of digital music technologies. People listen to music on various sources to improve their mood and enliven the environment. However, due to the enormous number of songs accessible, people are frequently perplexed when selecting a track.

People enjoy listening to different genres of music according to their mood. So, Naive recommendations do not work as intended.This necessitates the development of a context-sensitive music recommendation system.Context-aware recommendation systems also need accumulation of additional data, which can be difficult to come by but has shown to be worthwhile in some cases.

Context-aware music recommendation systems build on the principle that various personality traits connect with distinct item attributes (for example, acoustic qualities or musical genres) and that users in different emotional states and moods prefer various sorts of things.

Emotion-aware Recommendation systems have yielded encouraging results, demonstrating that using emotional states and reactions, recommendation algorithms may improve prediction accuracy and refine personalisation.

The emotions of users may be identified in a variety of ways. However, they all fall into one of two types. Face expressions, keystrokes, and mouse-click patterns are all examples of implicit ways of detecting emotion. Explicit methods, on the other hand, take direct input from the user.For the identification of emotion, we choose to utilise an approach based on facial expressions.

In the next sections, several Facial-Emotion Identification techniques, machine-learning models, and recommendation algorithms will be reviewed and contrasted. We further demonstrate that the emotion-identification method used has no effect on the recommendation algorithms.

**1.1. Facial Emotion Identification**

The task of recognising emotion from a user's face data is known as facial emotion identification. A Neural Network is typically used to accomplish the task. A Neural Network is a network of Interconnected perceptrons.

**Working of an Artificial Neural Network:**

A single neuron can either fire or not, so its activity level can be represented as a one or a zero. One neuron receives its information from a bunch of other neurons, but because the strength of these connections varies, each one might be assigned a different weight. Some connections are excitatory, so they have positive weights, while others are inhibitory, so they have negative weights.

The approach to find out whether a certain neuron fires is to multiply the activity of each input neuron by its weight, then add them all together. The neuron fires if their sum is larger than a value called the bias; if it is less than that, the neuron does not fire.

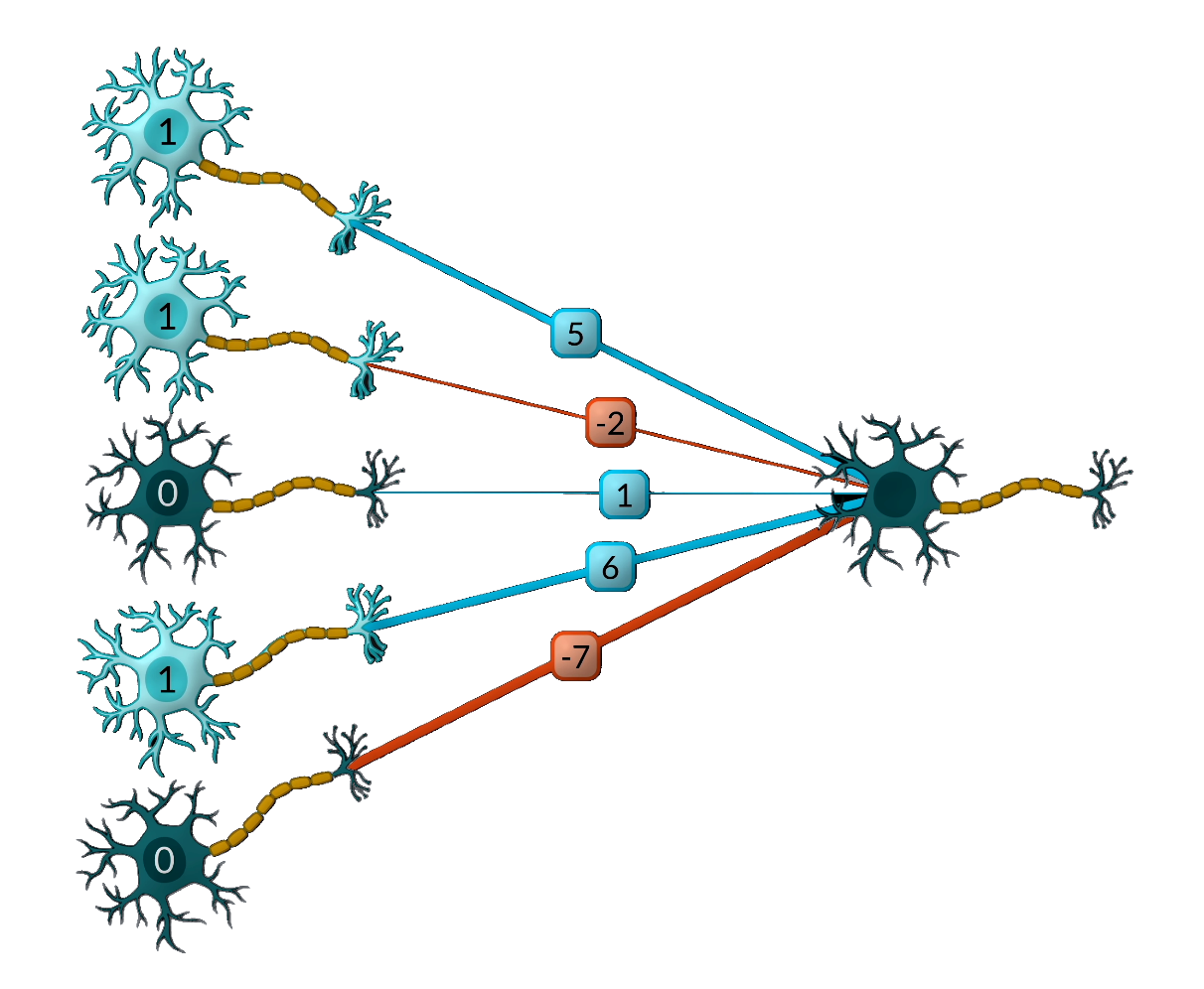


Figure1.1.1: Basic model of how Neurons in Artificial Neural Networks work.

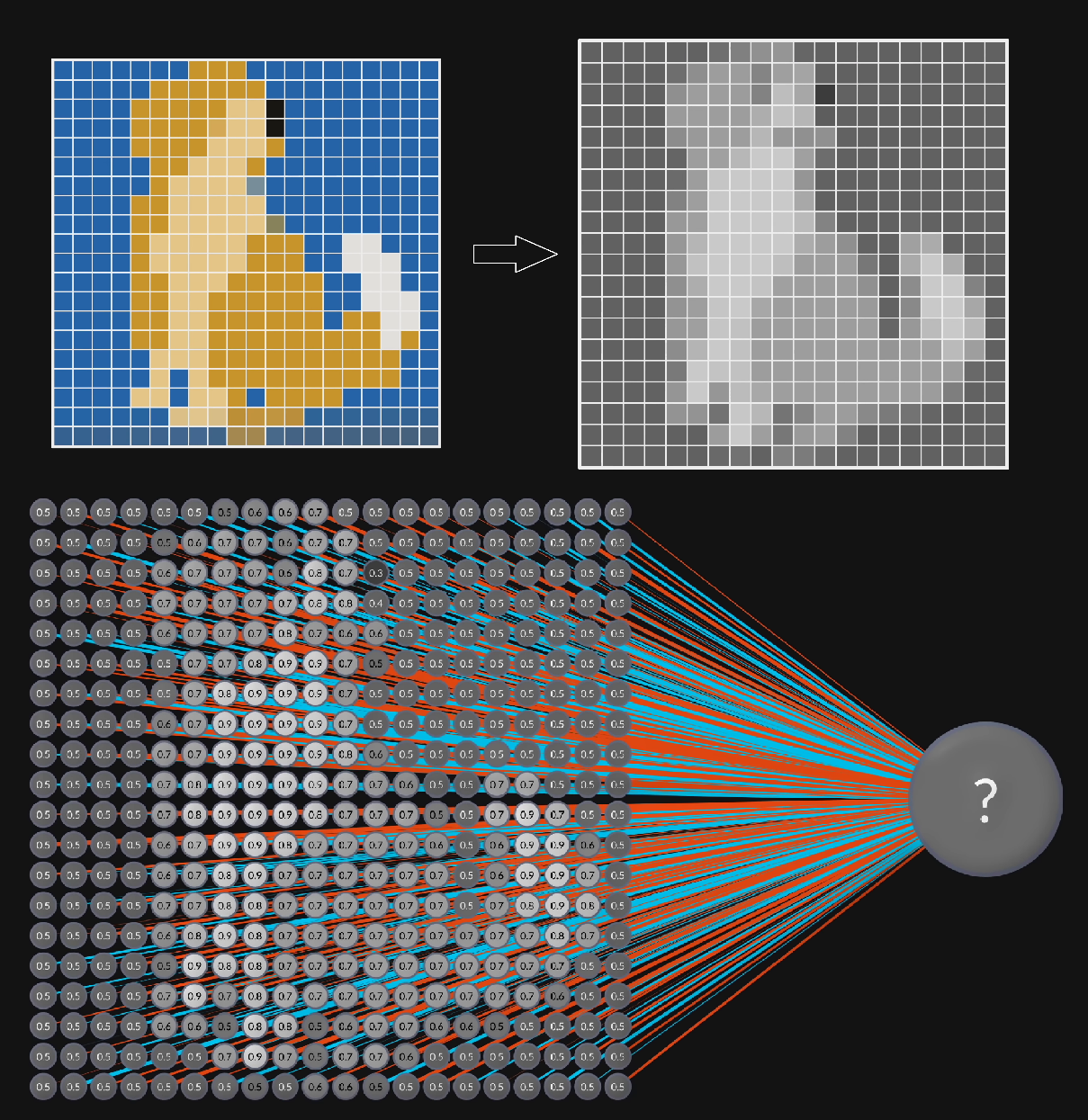


Figure1.1.2: Processing of a digital image by a Neural network.

The Neural Network accepts an M x N matrix as input, which represents a picture with M x N pixels.Each pixel functions like a neuron, with the brightness of the pixel determining its activation.In most cases, the activation is normalised to a number between zero and one. Each of these neurons is linked to the next collection of neurons via their own adjustable weights, which is referred to as a layer.

Depending on the bias and the summation of the product of weights and activation, each of the neurons in the following layer will either fire or not.At the end of the process, the output layer will have a number of neurons, generally only one of which will fire.

The outcome of the Neural Network is represented by the output layer.Depending on what the Neural Network interprets the image as, just one neuron in the output layer will fire.The training algorithm will alter the weights associated with each neuron whenever an output neuron fires when it shouldn't.Similarly, when an output neuron does not fire when it should, the weights are modified.

The weights associated with the neurons in the hidden layer are continuously adjusted by the algorithm until the neural network properly recognises all of the training pictures. The process of adjusting the weights is called Back Propagation.

A picture containing text

Description automatically generated

Figure1.1.3: Emotion Identification by a Neural Network.

For more precise training, the Neural Network becomes denser.Processing a single image takes roughly 700,000,000 mathematical operations because of all the large Matrix Multiplications.This creates a huge need for faster processing equipment, particularly GPUs (graphics processing units) that are designed for rapid, simultaneous computations.

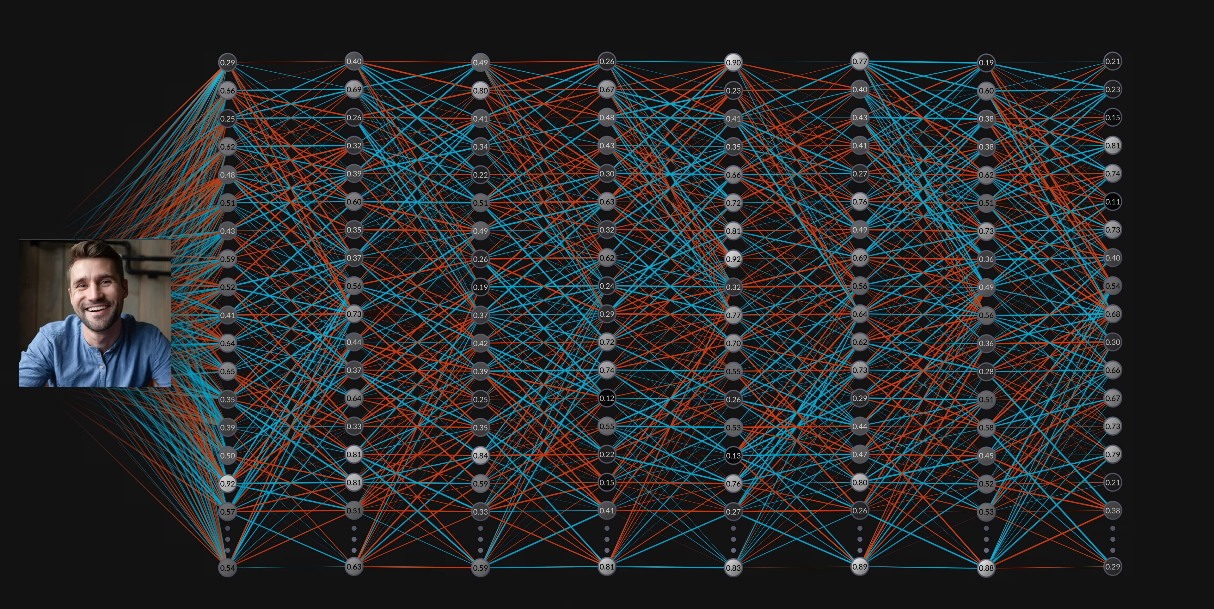


Figure1.1.4: A Neural Network involving more hidden layers.

The contrasts, benefits, and limitations of the top three Neural Networks, **Convolutional Neural Networks**, **Deep Neural Networks**, and **Multi-task Cascaded Convolutional Neural Networks** are the topic of discussion of our project.

**1.2. Music Recommendation**

The ordinary individual makes a range of purposeful media consumption selections on any given day.As we manoeuvre these omnipresent options in our increasingly diversified environment, recommendation system algorithms that have become pervasive in our lives point us in the right direction.

The fundamental goal of recommendation algorithms is to assess user data so that personalised recommendations may be made.

**1.2.1. Collaborative Filtering**

To provide suggestions, collaborative filtering looks for similarities between users and products at the same time. Collaborative filtering is often used with very big data sets. The collaborative filtering strategy has the benefit of not relying on machine comprehensible material, thus it can properly recommend complicated items without requiring a "knowledge" of the item itself.

Recommended to him

Similar Users

She Listened

Figure1.2.1: Collaborative Filtering

**1.2.2. Content-based Filtering**

Another prominent way for constructing recommender systems is content-based filtering.Based on the user's prior activities or explicit input, content-based filtering uses item attributes to recommend additional items that are similar to what they like.

He Listened

to him

Music

Similar

Recommended

Figure1.2.2: Content-based Filtering

**2.Overview of Proposed System**

The Proposed system consists of two phases:

2.1. Emotion Identification

2.2. Music Recommendation

**2.1. Emotion Identification**

Neural Networks are typically used for face classification and emotion identification. After extensive research and experimentation on different neural network models such as the CNN model, MTCNN model and DNN model, the following observations were made:

|  |  |  |  |
| --- | --- | --- | --- |
| **Parameter** | **CNN** | **Deep Face (DNN)** | **FER (MTCNN)** |
| Accuracy | High | Low | High |
| Train Time | Low | High | High |
| Validation Time | Low | High | High |
| Advantages | Speed | Light Weight | Self-alignment of Face |
| Disadvantages | Large Training Data | Low Accuracy | High Train time |
| Common | Unable to detect rare facial expressions like Disgust | | |

Table2.1.1: Compare & Contrast of Different Emotion Detection Models.

The MTCNN model has the highest accuracy compared to other models. Although the train time of MTCNN is high, it can be reduced using GPUs. Hence, we propose the MTCNN model for emotion identification.

In the proposed system, the MTCNN model takes the user’s face as input and outputs the dominant emotion detected.

**2.2. Music Recommendation**

Content-based filtering is used to recommend music. Since there are no publicly available datasets on users’ emotions while listening to songs, collaborative filtering cannot be used. Hence content-based filtering approach is proposed for music recommendation.

The proposed approach involves identifying the emotion of each song from the MuSe dataset. Each track in the dataset has valence, arousal, and dominance values, which represent the overall emotion of the song. K-means clustering is done after determining the initial centroids based on the VAD range of the dataset.

After clustering, we get 7 clusters, each representing one of the following seven emotions: happy, sad, angry, neutral, surprise, disgust, and fear. All the tracks with the same emotion get grouped into the corresponding cluster. Based on this, the songs in the dataset get annotated with their corresponding emotion.

Based on the emotion identified from the previous phase and using the tagged music dataset, top k tracks with the same emotion get recommended to the user. These get displayed as Spotify widgets, which can be played by clicking the widget.

**3.System Design**

The System consists of the following modules

1. Emotion Identification module

2. Music Tagging and Recommendation module

3. User Interface module

**3.1. Emotion Identification Module**

In this phase, we detect and identify the user's emotion. A Multi-task Cascaded Convolutional Network (MTCNN) model, trained on the FER Dataset, is used to identify the emotion. MTCNN has an added advantage of self-alignment of face compared to other models.

The MTCNN model takes the face as input and outputs the dominant emotion identified. The neural network is composed of numerous layers of weighted nodes. Each layer processes the input and feeds it on to the next layer. The node weights are adjusted during propagation. The node with the largest weight in the output layer correlates to the dominant emotion.

The model performs on par with the current state-of-the-art systems and has reported accuracy of 92.19%.

Figure3.1.1: Emotion Identification Module.

Happy

**3.2. Music Tagging and Recommendation Module**

The next phase involves tagging music with its corresponding emotion. The MuSe dataset consists of over 90,000 tracks along with their valence, arousal, and dominance values.Valence represents the pleasantness dimension, Arousal represents the intensity dimension, and dominance represents the control dimension. However, the dataset doesn’t consist of the emotion of the track. Hence each track needs to be tagged with its corresponding emotion.

Tagging involves clustering based on VAD values. VAD values are floating-point coordinates in 3-dimensional space, which represent the overall emotion of the song. VAD values are relative and can vary based on the range of the dataset. Based on the range of the MuSe dataset, VAD values for the seven emotions are determined.

K-Means clustering is used to group songs having same emotion. Initial centroids of the 7 clusters are obtained based on the VAD range. K-Means Clustering is then performed with the initial centroids.7 Clusters are obtained at the end, each identifying an emotion.

Based on the emotion identified from module 1 and using the tagged music dataset, top k tracks with the same emotion get recommended to the user.

VALENCE

AROUSAL

DOMINANCE

Figure3.2.1: Music Recommendation Module.

**3.3. User Interface Module**

Through a web-based interface, the user interface module coordinates all interactions between users and our system. Users can capture their emotions using the camera module. Based on the user's emotion top 15 tracks are displayed as Spotify widgets, which get played on clicking the widget.

Diagram

Description automatically generated

Figure3.3.1 User-Interface Module.

**References**

1. Sunitha M, Adilakshmi T (2018) Music recommendation system withuser-based and item-based collaborative filtering technique. In NetworkingCommunication and Data Knowledge Engineering: Springer.
2. H.-C. Kwon and M. Kim. Lyrics-based emotion classification using feature selection by partial syntactic analysis. 2011 IEEE 23rd International Conference on Tools with Artificial Intelligence (ICTAI 2011), 2011.
3. J. A. Speck, E. M. Schmidt, B. G. Morton, and Y. E. Kim. A comparative study of collaborative vs. traditional musical mood annotation. In A. Klapuri and C. Leider, editors, ISMIR, University of Miami, 2011.
4. K. Bischoff, C. Firan, R. Paiu, W. Nejdl, C. Laurier and M. Sordo. Music Mood and Theme Classification a Hybrid Approach. In Proceedings of the 10th International Society for Music Information Retrieval Conference, ISMIR 2009.
5. Hyung, Z. Park, J.S. Lee, K. Utilizing context-relevant keywords extracted from a large collection of user-generated documents for music discovery. Inf. Process. Manag. 2017.
6. M. V. Zaanen and P. Kanters. Automatic mood classification using tf\*idf based on lyrics. In Proceedings of the 11th International Society for Music Information Retrieval Conference, Utrecht, The Netherlands, August 9-13, 2010.
7. R. Malheiro, R. Panda, P. Gomes, and R. Paiva. Bi-modal music emotion recognition: Novel lyrical features and dataset. In 9th International Workshop on Music and Machine Learning MML2016 in conjunction with the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases ECML/PKDD 2016.
8. Ji-Oh Yoo Han-Saem Park and Sung-Bae Cho. L. Wang. A context-aware MRS using fuzzy Bayesian networks with utility theory. In FSKD 2006.
9. Giz H. He, J. Jin, Y. Xiong, B. Chen, W. Sun, and L. Zhao. Language Feature Mining for Music Emotion Classification via Supervised Learning from Lyrics. Springer Berlin Heidelberg, Berlin, Heidelberg, 2008.
10. Y. Hu, X. Chen, and D. Yang. Lyric-based Song Emotion Detection with Affective Lexicon and Fuzzy Clustering Method. In ISMIR, 2009.
11. C. Laurier, J. Grivolla, and P. Herrera. “Multimodal Music Mood Classification Using Audio and Lyrics. 2008 Seventh International Conference on Machine Learning and Applications, 2008.
12. F. H. Rachman, R. Sarno, and C. Fatichah, “Music Emotion Classification based on LyricsAudio using Corpus based Emotion,” Int. Journal of Electrical and Computer Engineering (IJECE), 2018.
13. R. Malheiro, R. Panda, P. Gomes, and R. Paiva. Music emotion recognition from lyrics: A comparative study. In 6th International Workshop on Machine Learning and Music, 2013.
14. Shlok Gilda, Husain Zafar, Chintan Soni, and KshitijaWaghurdekar. Smart music player integrating facial emotion recognition and music mood recommendation. In 2017 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET). IEEE, 2017.
15. I. Bakker, T. van der Voordt, P. Vink, and J. de Boon. Pleasure, arousal, dominance: Mahrabian and russell revisited. Current Psychology, 2014.
16. Emotion based Music Recommendation System (<https://github.com/Udhay-Brahmi/Emotion-based-music-recommendation-system>).
17. Retinaface (<https://github.com/serengil/retinaface>).
18. FER (<https://github.com/justinshenk/fer>).
19. Realtime Emotion Detection (<https://github.com/neha01/Realtime-Emotion-Detection>).
20. Emotion Detection (<https://github.com/dhruvpandey662/Emotion-detection>).
21. Keras Optimizers (<https://keras.io/api/optimizers/>).
22. Keras Metrics (<https://keras.io/api/metrics/>).
23. Keras Losses (<https://keras.io/api/losses/>).
24. Future Computers will be Radically Different

([https://www.youtube.com/watch?v=GVsUOuSjvcg](https://www.youtube.com/watch?v=GVsUOuSjvcg&ab_channel=Veritasium)).