How to Cite:
Bongirwar, V. K., & Potnurwar, A. (2022). Song recommendation using speech emotion recognition. *International Journal of Health Sciences*, *6*(S1), 10428–10434.
https://doi.org/10.53730/ijhs.v6nS1.7498

Song recommendation using speech emotion recognition

Vrushali K. Bongirwar
Department of Computer Science and Engineering, Shri Ramdeobaba College of Engineering and Management, Nagpur
Correspondence author email: bongirwarvk@rknec.eu

Archana Potnurwar
Priyadarshini College of Engineering, Nagpur
Email: archanapotnurwar@gmail.com

***Abstract*---**Music has become an integral part of our day-to-day life. With the advancement in technology, lots of businesses are nowadays using recommendation systems for their benefit like in e-commerce, selling books, movies, and video recommendations. This helps businesses to get monetary benefits and the users to get better services. One of the best ways to suggest songs would be according to the people's mood. In this project, we aim to interpret the user's mood at a particular time using what they speak and then suggest songs based on the analysis done by the system. The best thing about music is that nothing can be more relaxing than a pleasant melody. Due to all these things, we choose to do this Song Recommendation system according to the mood of the user.

***Keywords*---**Emovibe, Sentiment Analysis, Sequential Model, Tensorflow, Keras, MongoDB

Introduction

Sentiment analysis is the process of using natural language processing, text analysis, and statistics to analyze one's sentiment. In today's dynamic world, entertainment is thekey to survival. Everyone needs an entertainment source to calm out the stress in their life. Sentiment analysis can be used to make our entertainment selection easy. People are expressive, they write what they feel. Whenever we listen to a song, we can feel the music. The reason we feel the music is because we are feeling that same energy and same emotion. Don't we get tired of finding the music we want to hear at moment? Some are songs we don't even know about! Keeping this in mind we are presenting a solution, that helps and suggests songs to the users based on their mood. Emovibe is developed to provide

an interesting source of entertainment. This system is trained to recommend songs based on the mood of the user. The sole purpose of this system is to recommend songs that a user wishes to hear. The system provides many songs to selectfrom.

Related Work

In [1] the author introduces a new approach of feature extraction, using DBNs in DNN to extract emotional characteristics in voice signals automatically, about feature extraction in speech emotion identification challenges. The author in paper [2] studies the creation and validation of a collection of hybrid acoustic features based on prosodic and spectral data for improving speech emotion recognition and the created acoustic qualities have made it simple to distinguish eight different types of human emotions in recent projects thereafter. The study in [3] sought to employ an inception net to solve the problem of emotion recognition also various databases were investigated, and the IEMOCAP database was used as the dataset for my experiment. TensorFlow was used to train my model.

The author in [4] presents the approach of the utilization of the human worker's emotions in a human-robot collaboration context. The effect of the human's trust and comfort levels during collaborative work was mirrored in this utilization. The system uses [5] conventional metrics along with a user evaluation test. The document contains a brief personal story associated with request for the song which is considered as a input to the system assuming the story is closely related to the song's information and background. The system is based on the text processing and analysis extracted from the large corpus congaing information about the story and related songs.

Some of the recommendation systems [6] are designed to take features from the facial expression and then apply the machine learning algorithms for prediction of the mood as the facial expression plays a vital role in the prediction of human emotion.

Proposed Methods

The technologies used for making this project are ReactJSforthefrontend, Flask RESTAPI asthemiddleware, Python in the backend, and MongoDB Atlas as the database. The implementation of the project and the proposed methods are as follows-

Dataset and Data Cleaning

To train our model, we have used a mixed dataset from Kaggle. The dataset consists of 65,000 data rows in the format sentence, emotion. In data cleaning, Used RE (Regular Expression)librarytoremovetwitterhandleslike'@sam'from text and to remove every thing but words containing only small and capital letters. Used Multi Label Binarizer to encode the output labels for our data (One-hot encoding). The label for eachsentencewillbeanarraywith1attheindexoftheemotion it is classified as and the rest of the indexes as 0. Used word Lemmetizer from NLTK to convert words with similar root words. E.g.: Playing, Plays, Played get

converted to Play. Removed all the stop words from our sentences using stop words available in NLTK. For giving the input to our model we decided on a vocab size of 10000 and converted the sentences as input to an array of length10000.

Model Training

The cleaned data was split into training and testing data in the ratio 3:1 (75% training and 25% testing) and was given as input to the sequential model from Keras-Tensorflow. The model has 3 dense layers of neural networks. Layer one has the input shape as vocab size i.e., 10000 and 50 neurons with Relu as the activation function. Layer 2 has 35 neurons and Relu as an activation function. The last layer has 10 neurons and a sigmoid as the activation function. The optimizer used for training the model is Adam and the learning is performed at a learning rate of0.01.Thetrainingandtestingaccuracyas86%and76%.The model gives an array of the length of 10 as output with each element as the probability of the text being of that particular emotion. We take the index of the maximum value from our output and classify the text as emotion at thatindex.

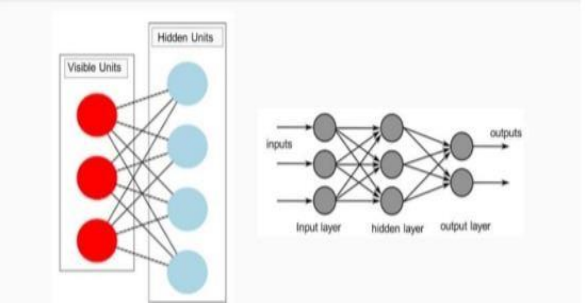

Figure 1: Model Architecture

Input Through Speech or Text

The user enters the input through text or by recording the audio. The audio is then converted to text using voice-to-text conversion in ReactJs. This text is sent to the backend using Flask REST API.

Input to the Model

The trained model is used to predict the emotion based on the text that is received from the front end. This model classifies the text into one of nine emotions (Sadness, Happy, Worry, Relief, Fear, Surprise, Neutral, Angry, Love).

Preparing a Songs Database

Twenty-five songs per emotion with the respective YouTube links were added to the MongoDB collection. The following figure shows the snapshot of the Mongo cluster created –

Figure 2: Snapshot of MongoDB Collection

Rendering Songs According to Classified Emotion

We then find all the songs from our MongoDB Atlas database corresponding to the emotion predicted. We then display three random songs on our front end according to the emotion classified. The following flowchart (Fig. 3) gives an overview of the entire process:

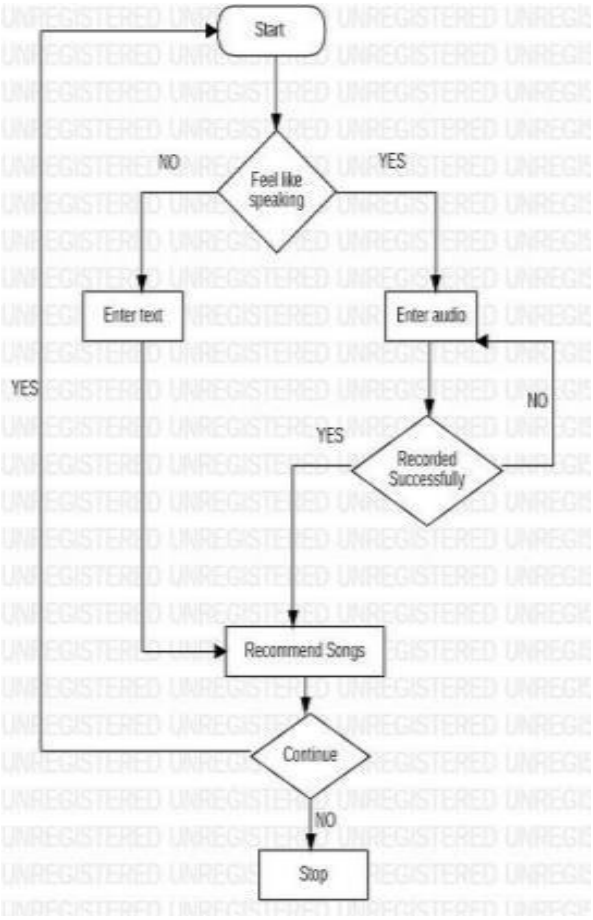

Figure 3: Flowchart Showing Working

Output

Every application has its working homepage which will give the proper User Interface for theshowcase or Display purpose. In the beginning, the user is navigated to the First page i.e., homepage as shown below:

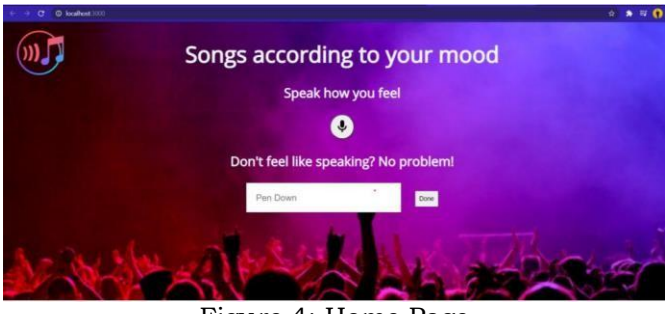

Figure 4: Home Page

As the application will consider the audio / text input and then process that to recognize the emotion of the user. And further according to the recognized emotion the Song of the same category is suggested by the application to the User. So, for audio input the processing is shown in the subsequent figure which displayed the page when the user records their audio through the microphone.

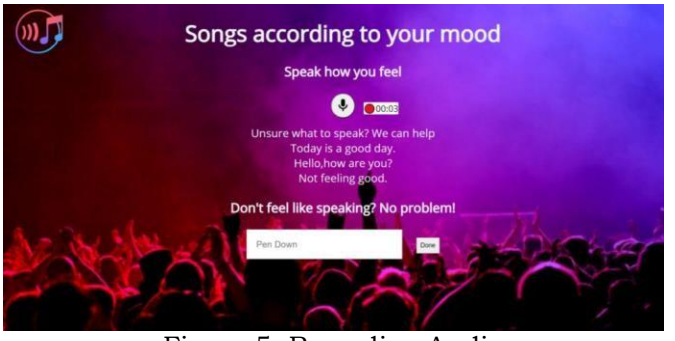

Figure 5: Recording Audio

Similarly, if the user input is considered in the form of textual data, the application is designed to recognize the Emotion from the text and displayed the following page.

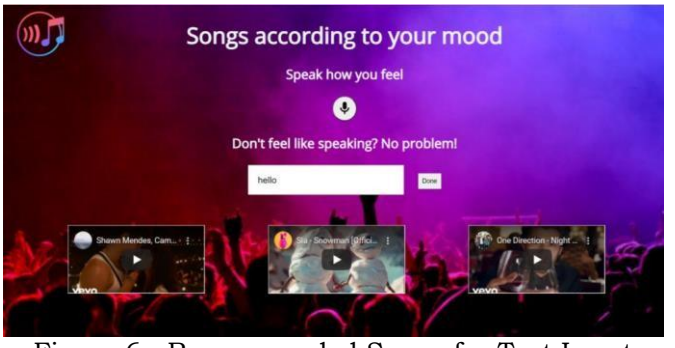

Figure 6: Recommended Songs for Text Input

Conclusions

Recommendation systems is very vital area in today's era. We have proposed the Song recommendation system which uses Emotion recognition as a key parameter for the recommendation of the songs as per the mood of the user. In our application development we have designed the system first. After understanding the Scope of the application, we have developed the data set which contains the datafrom multiple datasets from different sources. We have considered this resultant merged dataset to train our model. After merging datasets, we started working on our model and studied different models in machine learning. After understanding and comparing each model based on the Literature, we have selected the Sequential model for the further implementation.After the training phase, the model is tested and updated the weights for the better accuracy.

The web pages are developed in react. The database is designed in the MangoDB as a next phase and finally the system gets integrated and developed an application for the Song recommendation based on the human emotion at that specific time. To integrate python with MongoDB we used some inbuilt python libraries. To integrate react with python we used the flask integration.This results in the proper recommendations suggested as per the category or class of the user's emotion classified by the model. The accuracy of the application is noted as 83.41%.

References

1. Huang, C., Gong, W., Fu, W. and Feng, D., 2014. Research of speech emotion recognitionbased on deep belief network and SVM. Mathematical Problems in Engineering, 2014.
2. Kerkeni, L., Serrestou, Y., Mbarki, M., Raoof, K. and Mahjoub, M.A., 2018, January. SpeechEmotion Recognition: Methods and Cases Study. In ICAART (2) (pp. 175-182).
3. Roopa, S.N., Prabhakaran, M. and Betty, P., 2018. Speech emotion recognition using deeplearning. Int. J. Recent Technol. Eng.
4. Aouani, H. and Ayed, Y.B., 2020. Speech emotion recognition with deep learning. ProcediaComputer Science, 176, pp.251-260.
5. Hyung, Ziwon& Lee, Kibeom& Lee, Kyogu. (2014). Music recommendation using textanalysis on song requests to radio stations. Expert Systems with Applications: An International Journal. 41. 2608-2618. 10.1016/j.eswa.2013.10.035.
6. ShanthaShalini. Ka ,Jaichandran. Ra ,Leelavathy. S A, Raviraghul. R ,Ranjitha. J. and Saravanakumar. Na, "Facial Emotion Based Music Recommendation System using computer vision and machine learning techiniques" Turkish Journal of Computer and Mathematics Education Vol.12 No.1 (2021), 912-917