Proceedings of the 6th International Conference on Communication and Electronics Systems (ICCES-2021)
IEEE Xplore Part Number: CFP21AWO-ART; ISBN: 978-0-7381-1405-7

Emotion Based Music Playlist Recommendation System using Interactive Chatbot

Amrita Nair, Smriti Pillai, Ganga S Nair, Anjali T
Department of Computer Science and Engineering
Amrita Vishwa Vidyapeetham
Amritapuri, India
theamritanair@gmail.com, smritipillai.smriti@gmail.com, gangasnair2009@gmail.com, anjalit@am.amrita.edu

2021 6th International Conference on Communication and Electronics Systems (ICCES) | 978-1-6654-3587-1/21/$31.00 ©2021 IEEE | DOI: 10.1109/ICCES51350.2021.9489138

***Abstract*—Music is an integral part of our lives. However, since the social media platforms like TikTok and Instagram have a huge influence on the music charts worldwide, users are exposed solely to mainstream music, therefore the recommendations on music streaming platforms are not very personalized. An emotion-based recommendation system permits the users to listen to music based on their emotions. Existing systems use audio signals using the CNN approach [1] and collaborative filtering [2] to recommend songs based on the user's history. The proposed research work develops a personalized system, where the user's current emotion is analyzed with the help of the chatbot. The chatbot identifies the user's sentiment by asking some general questions. Based on the input provided by the user, a score is generated for each response, which adds up to a final score; this score is used to generate the playlist. The proposed recommendation system utilizes the Spotify platform and API for the playlist generation and recommendation.**

***Index Terms*—Recommendation System, Sentiment Analysis, LSTM, Bidirectional LSTM, Chatbot, Convolutional Neural Network**

I. Introduction

Music is a universal language. It has been a crucial part of our lives since the beginning of time. We listen to music when we're having a bad day, we listen to it when we have a great day. Music inspires and enlightens us. From chirping of birds to drum percussion, from harp to riffs of an electric guitar, music has different forms of expression. Music connects people regardless of their religion, caste, and creed. It brings people together and has a huge influence on our lives. Music isn't just an art form or a language, but it also affects the human mind and body. It stimulates our minds. According to studies, music has therapeutic properties, and music therapy programs help with anxiety, dementia, stress management, and self-confidence. [3] [4] According to a paper published in the journal Neuroscience [5], personalized music-based interventions is encouraged for the treatment of brain disorders associated with abnormal mood and emotion-related brain activity. [6] In today's era, especially since the rapid growth of streaming platforms and applications like TikTok, the way users consume music has changed. Music is not judged by its quality, but rather by its popularity. This hinders quality music from underrated artists reaching the people. Therefore, due to the current music trends, it's not always necessary that the user finds music that he relates to. And as research shows that personalized music can have a positive impact on the human mind, it is important to have access to music based on our moods. The use of Human-Computer Interaction (HCL) has been one of the most sought-after ways that help the computer understand humans better. The most popular way of interaction is using chatbots, which are very engaging and user-friendly. Chatbots can be trained in such a way where for every response the user gives, they understand the context and respond accordingly to them. This makes the system efficient and gives the proper understanding of the user. Spotify is one of the world's largest streaming platforms with 345 million active users. [7]It provides a web API with full access to music data. In the API, each song is associated with attributes like Energy (tells us about the energy of the song), Valence (tells us if the music is cheerful and happy), danceability, acoustic-ness, etc. Such attributes help us better understand the overall mood of the song.

II. Related Work

This section compares the existing recommendation systems and the models they have followed to achieve particular targets in their paper. These papers helped us understand how these algorithms can be made more efficient, and thus helped us choose which models can be considered for our system.

In one of the papers, Naveen Kumar KS et al, evaluated the performance of linear and non-linear text representation techniques for sentiment analysis. They classified emotion into two categories, positive and negative using the tone and expression towards a particular topic. Confusion matrix was used to measure the performance of their model. The model is trained using TF-IDF and Glovec. Feature extraction is done by using vectorization because the Twitter Dataset is text data. The TF-IDF is used here because they're considering the features from 10,000 to 40,0000 which is generally performed using Deep Learning algorithms. Glovec is an unsupervised learning algorithm for obtaining vector representations for words. Deep Learning algorithms like LSTM and GRU have been used for training the dataset, and LSTM offers the best accuracy of 75.3%. [8]

In another paper, S. Srinivasan et al, designed two Deep Convolutional Neural Network models. These models were combined with pre-trained models like VGG19, Xception and they were trained on three different datasets. Each convolutional layer was followed by the ReLU activation layer and a

TABLE I
COMPARISON OF EXISTING WORKS

Paper Title	Dataset	Sentiment Analysis	Playlist recommendation	Model Used
Amrita-CEN-SentiDB 1: Improved Twitter Dataset for Sentimental Analysis and Application of Deep learning [8]	Twitter Dataset	Sentiment- Positive and negative using the tone and expression towards a particular topic	No	LSTM
Emotion based Music Recommendation System [9]	Images of facial expressions	Emotions were classified as Happy, Angry, Surprise, and Sad faces	Yes (Without personalized suggestions)	Facial Action Coding System
Emotion Detection in Hinglish Hindi+English Code-Mixed Social Media Text [10]	12,000 Hindi-English mixed texts	Emotions- Happy, Sad, and Anger	No	CNN-BiLSTM
A personalized music recommendation system using convolutional neural networks approach [1]	Million song dataset	No	Yes	CNN
Emotion Based Music Playlist Recommendation System using Interactive Chatbot	Twitter Dataset	Sentiment- Positive, Negative or Neutral	Yes (Including personal favourites)	Bidirectional LSTM

2 layers of max-pooling layer. Next the dropout regularization was used which helped with flattening the output. Finally,a sigmoid activation function was used. In sigmoid activation, the input to the function is transformed into a value between 0.0 and 1.0. Cross models are also employed in their work where the features are extracted from the final hidden layer of the first model and used them on ML classifiers. They also used Random Forest, Linear Support Vector Machine, K-Nearest Neighbor, and AdaBoost as classifiers in their model. They also used a Cost-sensitive learning method to handle data imbalance. [11]

In one of the papers the authors, J. James Anto Arnold et al, suggested a method where the person's face was recorded using the webcam. The recorded video was then divided into frames. The facial expressions obtained from the webcam were then pre-processed converting it into a sequence of Action Units (AUs). Using combinations of the 64 AUs, Facial Action Coding System (FACS) helped describe all the facial expressions. After Feature Extraction, the emotions were classified as Happy, Angry, Surprise, and Sad faces. These emotions were identified and music was suggested accordingly. [9]

In another work, Premjith Ba et al, used a dataset of 12,000 Hindi-English mixed texts and tried to classify them into emotions Happy, Sad, and Anger. Totally five models have been experimented with like CNN, LSTM, CNN-LSTM, BiLSTM and CNN-BiLSTM for classification where CNN-BiLSTM achieved 83.21% classification accuracy. The first experiment was done with 1D-CNN and the study indicated that 1D-CNN gave good results in NLP classification tasks. When CNN alone was used LSTM layers were removed from the model. LSTM and BiLSTM have the ability to memorize patterns with a sequence which can have significant importance while analyzing the text. At last, the CNN was used with LSTM and BiLSTM models because CNN is capable of abstracting features and lessen the complexity of training LSTM or BiLSTM. Results showed that CNN-BiLSTM performed better when compared to the rest of the models. [10]

III. PROPOSED METHODOLOGY

Our approach is depicted in the block diagram in Fig 1. It starts with the chatbot interacting with the user to understand their current mood. In existing systems, they used the Singular Value Decomposition(SVD) based feature for sentiment prediction where they compared classification performed using SVM via linear, polynomial and, RBF kernel, Naive Bayes, Simple logistics, Random forest and chose the best model. [12] In our project, the chatbot asks some general questions based on topics like sports, weather, etc. For every response the user gives, the score associated with the response is taken into account for understanding the overall emotion of the conversation, then the chatbot replies to the user based on the polarity of the sentence. Polarity, a function from the Text Blob, which is a Python library used for text processing and other tasks involved in natural language processing, helps in understanding the sentiment of a statement. The polarity score is a float within the range [-1.0, 1.0]. [13]

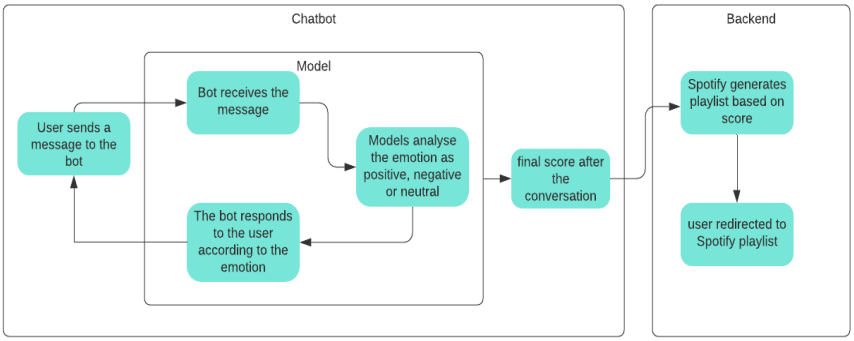

Fig. 1. Block Diagram

For calculating the overall score, the user input is fed into a model which classifies whether the user's response was 'Positive', 'Negative' or 'Neutral'.

A. Dataset Description and Text Preprocessing

For the sentiment training aspect, our models were trained using the Twitter Dataset from Kaggle which is a collection of around 25,000 tweets from Twitter and other websites. The tweets are classified into Positive, Negative, and Neutral. 40% of the data is labeled as Neutral, 31% labeled as Positive and the rest is Negative.

The first step is to preprocess and organize the data, as the dataset contains missing data and has inconsistencies. The best ways to preprocess inconsistent data is by tokenization and removing stopwords. [14] Text preprocessing will make the data more readable and understandable, and thus making it easy and efficient for the model to process. This leads to more accurate outcomes.

Since the dataset consists of tweets by various people on the Internet, they express their reactions better using various special characters, punctuation marks, and emoticons. And since Twitter is used by people globally who speak different languages, the purpose and meaning of these punctuations could differ, thus the emoticons and the punctuations become immaterial and hence are removed. For uniformity, all the tweets are converted into lowercase. Tweets with no content are deleted. As the tweets are classified into three sentiments, they are labeled such that 0 stands for Neutral, 1 stands for Positive and -1 stands for Negative. Further, the library Tweet-Preprocessor is used to clean, parse and tokenize tweets.

The preprocessing of the tweets is followed by the training of the model. After inferring on the existing recommendation systems and their successes, three different models - LSTM, Bidirectional LSTM, and 1-D Convolutional Neural Network are taken into consideration. Each of them is trained on the dataset, to classify the tweets into one of the 3 sentiments. The best performing model is then chosen for the calculation of the score.

B. Long Short Term Memory

Long Short Term Memory Networks otherwise known as LSTM networks are a type of recurrent neural network that handle long-term dependencies. They are extensively used for text, classification, time series prediction, speech recognition, emotion detection, etc. For our project, we're using LSTM as a part of sentiment analysis. Unlike recurrent neural networks, LSTMs can remember information for a longer duration. They do this with the help of gates. These gates help in the easier flow of information, and can also regulate whether to keep a piece of information or not.

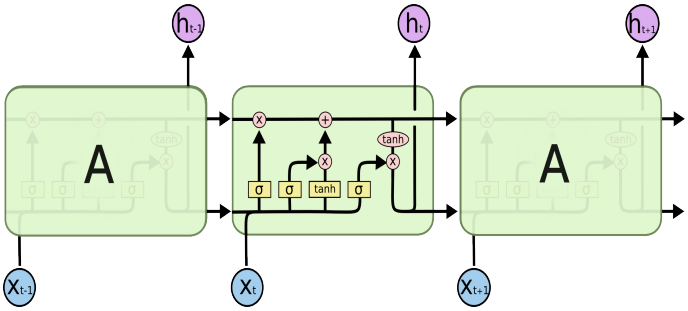

Fig. 2. Source: Understanding LSTM Networks – colah's blog [15]

- Forget Gate: This gate helps in determining whether a piece of information will be useful. If not, it is pulled out from the cell state. This is done by passing past and current information into the sigmoid function. The output from the function helps us establish whether the data has to be forgotten. This optimizes the performance.
- Input Gate: This gate is in charge of inclusion of useful information to the cell states.
- Output Gate: This gate regulates and decides the value for the next state.

The cell states in the network help transfer relevant information. Each gate acts a neural network, and they help in deciding the relevance of the information. For our project, we want to make sure we consider the important bits of information from the conversation the user has with the chat-bot, as it helps us in calculating a better and more accurate score.

Values from the initial and hidden state are passed to the sigmoid activation function, which helps to bring the values into a range from 0 to 1. The output generated from the previous layer is then passed through the tanh activation function, which helps in re-scaling values from -1 to 1. Both these outputs are then multiplied. Based upon the ultimate value, the value that should be taken by the hidden state is decided. This hidden state value is used for predicting the output. Similarly, the new cell state and value of the new hidden state are carried over for the next steps. This helps LSTMs to retain information for a longer period. In our project, we want to make use of every response given by the user for the computation of the score, thus we chose LSTM as one of the models.

C. Bidirectional LSTM

Bi-LSTM Layer considers the sequential internal relationship against each word in the input sentence given which means that the meaning of each word depends on the meaning of its previous words. However, in bi-directional LSTM, the process runs both forward and backward direction, i.e. first the meanings are derived in a left to right manner such that the meaning of the words lean on the meanings of the words preceding it and then in a right to the left manner, such that the meaning of each word depends on the meanings of words after it. This step is said to work like a memory for the system. It is usually thought that Bi-directional LSTMs learn faster than one directional LSTM, even though it mostly depends on the task at hand.

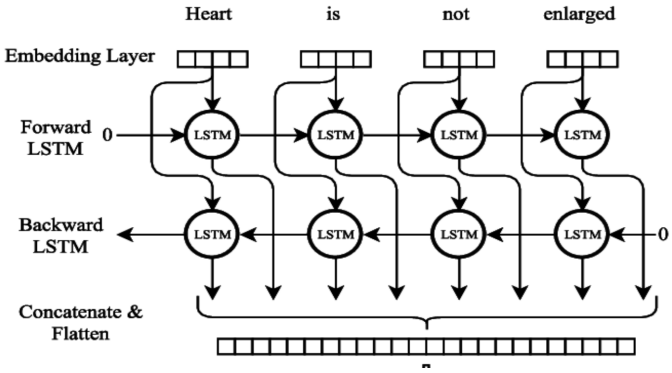

Fig. 3. Source: BiDirectional LSTM Explained [16]

D. Convolutional 1D Neural Network

Convolutional Neural Networks (CNN) are feed-forward networks that were developed for image classification problems, and it uses feature learning where it takes 2-D input representing an image's pixels and color channels. This process can also be applied to 1-D sequences of data inputs. The pre-processing needed in a CNN is much lower as compared to other classification algorithms. It can learn the filters/characteristics with enough training given to them. In a CNN model, each input image is passed through a series of convolution layers with filters, also known as Kernals, pooling layers, fully connected layers (FC), and an activation function such as Softmax function is used to classify an object and bring the values within the range [0, 1]. The most important role of a CNN is to take in image inputs and reduce them in size without losing their important features which help with a good prediction of the model.

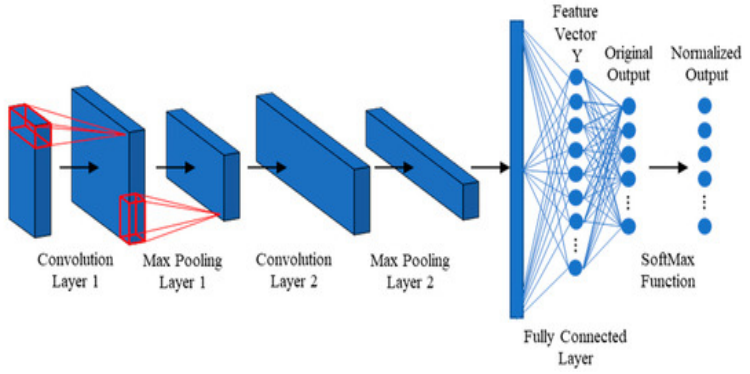

Fig. 4. Source: 1D Convolutional Neural Network Architecture [17]

CNN's are known for feature extraction mostly in images that are in the form of matrices, but they have shown good results when it comes to text classification where the input becomes a 1D array and the network uses 1D convolutional and pooling operations. [18]

After training the models, it was found that the Bidirectional LSTM performed the best among the three with an accuracy of 79.64%. (Fig. 5) Thus, Bidirectional LSTM was used to analyze the sentiment of the user's responses, and emotion from each response was accumulated to compute a final score in the range of 0 to 1, score less than 0.5 being negative/sad emotion and score greater than 0.5 being positive/happy.

After the score is computed, we're redirected to the page where the user is prompted to press the "Create Playlist" button.

For generating playlists, we're using the Spotify Web API. Spotify API provides endpoints using which we can get metadata about the artists, albums, and tracks. Using the API, with selective authorization, we can create playlists and save them in the user's profile. More features are explained in the documentation. [19]

The user is asked to log in with their Spotify account or using an email. If the user is an existing Spotify user, then we recommend songs from both his most listened to artists as well as new artists. This ensures that the artist can discover new artists and music, as well as can enjoy his favorites. After the artists are selected, underrated tracks are chosen along some top tracks and then the program selects tracks that are within a certain mood range.

This is done using attributes like Valence, Energy, Danceability, and Acoustic-ness. In the Spotify API, every track is associated with these attributes. The definition of these attributes according to the official documentation. [19]

- Energy - It is a measure from 0.0 to 1.0 and represents a perceptual measure of intensity and activity.

- Valence - Valence is a measure from 0.0 to 1.0 describing the musical positiveness conveyed by a track. Tracks with high valence sound more positive (e.g. happy, cheerful, euphoric), while tracks with low valence sound more negative (e.g. sad, depressed, angry).
- Danceability - Danceability describes how suitable a track is for dancing based on a combination of musical elements including tempo, rhythm stability, beat strength, and overall regularity. A value of 0.0 is least danceable and 1.0 is the most danceable.
- Acousticness - A confidence measure from 0.0 to 1.0 of whether the track is acoustic. 1.0 represents high confidence the track is acoustic.

Since Energy and Valence help us better understand the sentiment of the track, these were the key attributes that helped in classifying songs. After the tracks are chosen, a playlist is created and the user can view the playlist on Spotify.

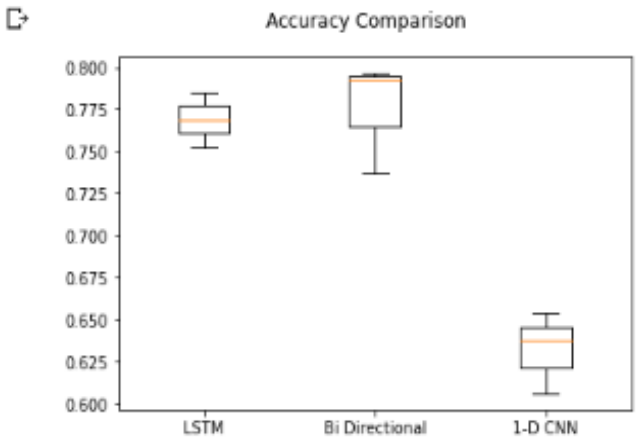

Fig. 5. Accuracy comparison of the three models

IV. Experimental Results

We have used three models for our project, namely, LSTM, Bidirectional LSTM and Convolutional 1D Neural Network. The models were trained on the Twitter dataset. The dataset was preprocessed, and after training the models, we got an accuracy of 77.89% for LSTM model, 79.29% for Bidirectional LSTM model and 63.75% for Convolutional 1-D Neural Network model. From this we chose Bidirectional LSTM as the best model and was further used to analyze the emotion of the user.

After opening the application user can have a conversation with the chatbot and based on the sentiment analysed a score is calculated which is passed to generate a custom playlist.

When a user had a conversation as shown in Fig.6, the emotion was detected as negative i.e., sad and a score of 0.109 was generated as output. From this score, a playlist consisting of sad and slow songs based on the user's favourite artists and language was generated (shown in Fig.6). Here, since the user already had a Spotify account, tracks by his favorite artists as well as new artists were taken in the playlist.

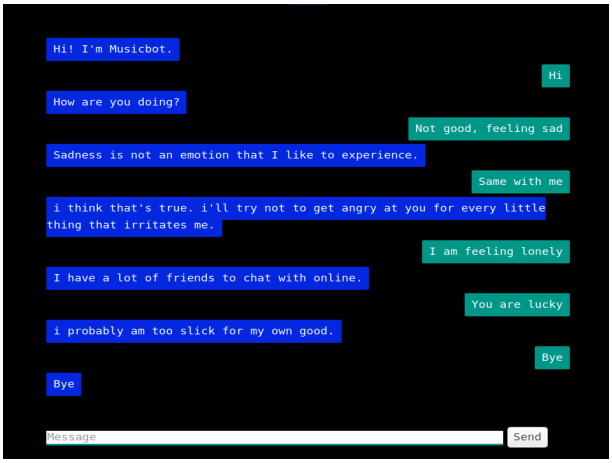

Fig. 6. App UI showing Chat with user

V. Conclusion

In this project, we have designed and developed a music recommendation system based on the person's emotion which is being identified with the help of a chatbot. We used three models, namely, LSTM, Bidirectional LSTM, and 1-D CNN, and trained them on the Twitter dataset, wherein Bidirectional LSTM gave the best results, which was further used for identifying and classifying the user's emotion to a certain category ie. positive, negative and neutral. Currently, the user's response is classified into three categories of emotion - Positive, Negative, and Neutral. As part of our future work, the emotional spectrum can be expanded by introducing more emotions like happy, sad, frustrated, confused, excited and so on which gives a better perception of the user's emotion. As of now, the user is only questioned from certain domains like weather, sports, favorite music, etc. The chatbot can be improved by adding more questions from various other topics. The chatbot gives a vague response to the user, but with proper training, we would be able to develop a chatbot that would give a fitting response to the user's input.

References

[1] S. Chang, A. Abdul, J. Chen and H. Liao, "A personalized music recommendation system using convolutional neural networks approach," 2018 IEEE International Conference on Applied System Invention (ICASI), 2018, pp. 47-49, doi: 10.1109/ICASI.2018.8394293.
[2] E. Shakirova, "Collaborative filtering for music recommender system," 2017 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (EIConRus), 2017, pp. 548-550, doi: 10.1109/EIConRus.2017.7910613.
[3] Research reveals pain and pleasure of sad music, https://www.dur.ac.uk/news/newsitem/?itemno=28329
[4] Music as an aid for postoperative recovery in adults: a systematic review and meta-analysis, https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(15)60169-6/fulltext.
[5] Journal Of Neuroscience, https://www.jneurosci.org/
[6] Interaction between DRD2 variation and sound environment on mood and emotion-related brain activity , https://www.sciencedirect.com/science/article/abs/pii/S0306452216306236
[7] Spotify — Company Info, https://newsroom.spotify.com/company-info/

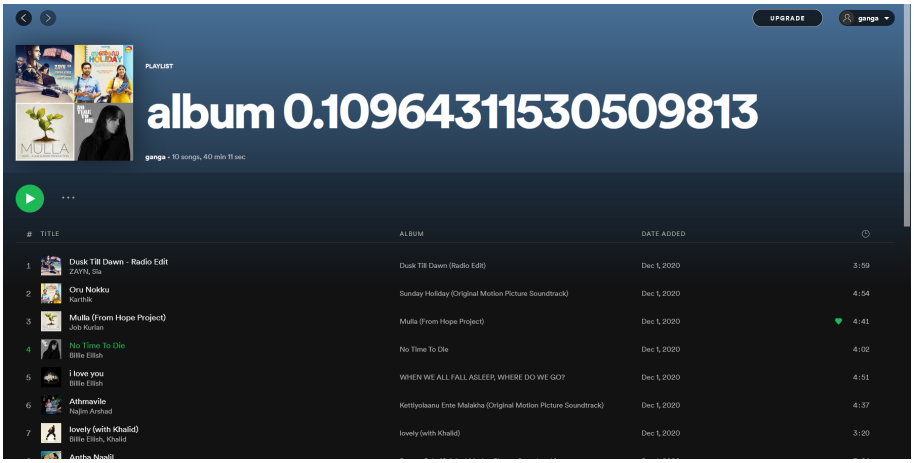

Fig. 7. Playlist for emotion detected as sad

[8] K. S. Naveenkumar, R. Vinayakumar and K. P. Soman, "Amrita-CEN-SentiDB 1: Improved Twitter Dataset for Sentimental Analysis and Application of Deep learning," 2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT), Kanpur, India, 2019, pp. 1-5, doi: 10.1109/ICCCNT45670.2019.8944758.
[9] "Emotion based Music Recommendation System", H.Immanuel James, J.James Anto Arnold, J.Maria Masilla Ruban, M.Tamilarasan,R.Saranya, International Research Journal of Engineering and Technology(IRJET), Volume:06 Issue:03—Mar 2019
[10] T Tulasi Sasidhar, Premjith B, Soman K P, "Emotion Detection in Hinglish(Hindi+English) Code-Mixed Social Media Text",Procedia Computer Science,Volume 171,2020,Pages 1346-1352,ISSN 1877-0509, https://doi.org/10.1016/j.procs.2020.04.144.
[11] S. Srinivasan et al., "Deep Convolutional Neural Network Based Image Spam Classification," 2020 6th Conference on Data Science and Machine Learning Applications (CDMA), Riyadh, Saudi Arabia, 2020, pp. 112-117, doi: 10.1109/CDMA47397.2020.00025.
[12] Thara, S., and S. Sidharth. "Aspect based sentiment classication: Svd features."; In 2017 International Conference on Advances in Computing, Communications and Informatics(ICACCI), pp. 2370-2374. IEEE, 2017.
[13] Tutorial: Quickstart — TextBlob 0.16.0 documentation https://textblob.readthedocs.io/en/dev/quickstart.html#sentiment-analysis
[14] Anjali, T., T. R. Krishnaprasad, and P. Jayakumar.;A Novel Sentiment Classification of Product Reviews using Levenshtein Distance.; In 2020 International Conference on Communication and Signal Processing (ICCSP), pp. 0507-0511. IEEE, 2020.
[15] Understanding LSTM Networks – colah's blog, http://colah.github.io/posts/2015-08-Understanding-LSTMs/
[16] BiDirectional LSTM Structure, https://paperswithcode.com/method/bilstm
[17] 1-D Convolutional Neural Network structure, https://www.mdpi.com/1424-8220/20/4/1059/htm
[18] Kim, Yoon. (2014). Convolutional Neural Networks for Sentence Classification. Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing. 10.3115/v1/D14-1181.
[19] Spotify Documentation, https://developer.spotify.com/documentation/