Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 218 (2023) 1196–1208

Procedia Computer Science

www.elsevier.com/locate/procedia

International Conference on Machine Learning and Data Engineering

# LyEmoBERT: Classification of lyrics' emotion and recommendation using a pre-trained model

Revathy. V. R[1], Anitha S. Pillai[2], Fatemah Daneshfar[3]

[1]*Research Scholar, School of Computing Sciences, Hindustan Institute of Technology and Science, Chennai, India*
[2]*Professor, School of Computing Sciences, Hindustan Institute of Technology and Science, Chennai, India.*
[3]*Assistant Professor, Department of Computer Engineering, University of Kurdistan, Sanandaj, Iran.*

**Abstract**

Music plays a significant role in evoking human emotions. Thanks to the quick proliferation of smartphones and mobile internet, music streaming applications and websites have made the music emotion recognition task even more active and exciting. However, music emotion recognition faces significant challenges too. These include inaccessibility of data, unavailability of large data volume, and the lack of other emotionally relevant features. While emotionally relevant features can be identified by analyzing lyrics and audio signals, the availability of datasets annotated with a lyrical emotion remains a constant challenge. This study uses the Music4All dataset to evaluate the lyrical features relevant for the identification of four important human emotions - happy, angry, relaxed, and sad. This was done with the help of several machine learning algorithms based on a semantic psychometric model. A transfer learning approach was also used to understand the feelings of the lyrics from an in-domain dataset and then predict the emotion of the target dataset. Further, it was observed that the BERT model improves the overall accuracy of the model (92%). A simple lyrics recommender system is also built using the Sentence Transformer model.



*Keywords:* Emotion classification; BERT; music lyrics recommendation; Music Information Retrieval; Emotion detection from lyrics; transformer approach.

## 1. Introduction

Since ancient times, music has always played an important role in our life. It helps evoke human emotions and deeply influences our moods, thoughts, and social interactions. Music adds to our social and cultural elements and influences us in many diverse ways.
In a musical environment, humans express their moods through various different emotions, such as happiness, sadness, tenderness, aggressiveness, etc. Thus, music can be called a medium to communicate emotions. Music affects us in so many different ways that today music emotion recognition has become an active part of the research. Being a subdomain of music information retrieval (MIR), music emotion recognition is extremely important [36].


10.1016/j.procs.2023.01.098

Thanks to the proliferation of smartphones and mobile internet, today music streaming websites and music information retrieval systems are on a quick rise. Several aspects of music emotion recognition have been studied and solved by researchers over the last few years.
Music emotion recognition has many advantages. It is helping researchers find out the prominent feeling or state of mind that music brings to a listener. It also helps in the identification of a listener's taste in music and aids music streaming systems (YouTube) and music recommender systems (Spotify) in automatic playlist generation.

Music mood recognition is a continuous discovery process. In it, the emotions associated with a musical piece are identified through different means. These include audio analysis, lyrical text analysis, and more. Most of the music classification research at present is based on analyzing the audio signals and the features of the music. However, even though music audio and lyrics are closely connected to each other, our brain operates the audio signals, i.e., the harmonic (tunes), and the semantic features of the lyrics separately [2]. Therefore, ideally, the investigation of music emotion should start from the lyrics as they contain prominent features that represent emotionally dependent and relevant information.

The new era of music recommendation systems uses information retrieval methods. They leverage collaborative approaches, including listeners' history, to improve recommendations [25]. However, due to the lack of an unpopular song recommendation, such recommender systems often suffer sparsity problems. Also, they often fail to recommend new songs or songs that lack user interaction. Due to this, less widely known tracks often get missed [2]. To work perfectly, music retrieval systems need a deep understanding of music. It can help find similar songs in a much better way. Here, lyrics and audio, which are a vital part of the song, play an inevitable role. Lyrics and audio of the music evoke emotion and help in understanding the emotional theme of a song. They can help better understand the user's emotional preferences.

This paper investigates the methods that are best suitable for labelling text data as well as the models that ensure efficient emotion classification. Lyrics provide deep information about a song. They help to reveal the meta-data features of the song. For example, the type of the song, its sentiment, and the message that it wants to give. This work aims to automate the same.

For this purpose, a classifier is built to predict the emotion of the song. It can suggest whether a song is happy, sad, relaxed, or angry, only by analysing its lyrics. Such a classifier will find increasing use in the music industry. It can help to automatically classify and form playlists in music recommender systems. This paper proposes an improvised approach to label lyrics data using emotionally relevant features and pre-trained word embeddings. This work also highlights the importance of pre-trained models and how they can be utilized for transfer learning from a corpus dataset to the lyric dataset. The pre-trained model is suitable for predicting songs' emotions, and this prediction can be applied to solve several challenges in Music Information Retrieval (MIR) tasks. The proposed model can even be re-trained to perform multi-class text classification and improve performance. It suggests future motivations to improvise the performance of text-based emotion recognition tasks and is good for exploring and adding more emotions to the text.

This paper is organized in the following way: the related work is discussed in section 2, a background of the research is in section 3, describing the dataset used for the work is in section 4, explaining the proposed model in section 5, and recommendation of the lyrics is explained in section 6, the discussion about the results and comparison is shown in section 7, which is then followed by acknowledgment section and references.

## 2. Related work

Recently, several studies have concentrated on the multi-class classification of text including lyrics [1, 4, 5, 8, 15, 19, 32, 33]. In [3] the semantic analysis of lyrics was proved as an efficient way to recognize emotion. Combining lyric and audio features does not assure better mood prediction in spite of training on lyrics features [11]. The model proposed in [20] uses multi-modal aspects of the music for the classification. Their work has shown that combining audio, MIDI and lyrics information can alleviate the glass ceiling problem. Authors of [18] also concluded that among the multi-modal classification of audio, MIDI, and lyrics of the music emotion, the mode that was the least performed was lyrics because of its structural features. In [18] used audio and lyrics for mood classification. The authors used

state-of-the-art NLP techniques such as TF-IDF to identify lyrics' features. The mood classification based on lyrics features reached low accuracy by applying SVM, Naïve Bayes, and KNN methods. Lyrics combined with audio give better results in their study [18].

A multi-class emotion classification approach of lyrics is discussed in [11] and also a labelled emotion dataset is available as part of this work. They experimented an out-of-the-domain dataset using the pre-trained model and further used it to classify songs based on emotions – joy, sadness, etc. However, the Naïve Bayes approach offered better results as compared to the BERT model [11]. In [20], the authors analyzed ways to categorize the emotions of lyrics which was demonstrated based on Russell's circumplex Model (Figure 1). In [30, 32, 34] authors worked on text-based recommendations. In [32] authors performed a comparative analysis on whether lyrics or audio is important for generating recommendations. In [34], text recommendations are performed based on transfer learning using the BERT model. In [30], BERT is used to generate sequential recommendations. In [32] authors performed lyric-based recommendations based on their similarities.

## 3. Background

### *3.1 Emotional aspect of lyrics for emotion detection*

Natural language processing (NLP) employs many linguistic and computational algorithms to identify various human language interpretations in any form of input including text, audio, or many more. Various notable research contributions are available in support. These include emotion detection, sentiment detection, and polarity detection. Emotion detection and categorization have been studied in both dimensional and discrete spaces. The discrete area of emotions is derived from feelings that are influenced by human beings' biological processes and culture. According to Paul Ekman, happiness, anger, surprise, sadness, fear, and disgust are expressive emotions of a human being [12]. This research work does emotion classification according to the dimensional model of emotions. The dimensional space considers text, audio, and video, and maps their features into the respective emotion apt for the context.

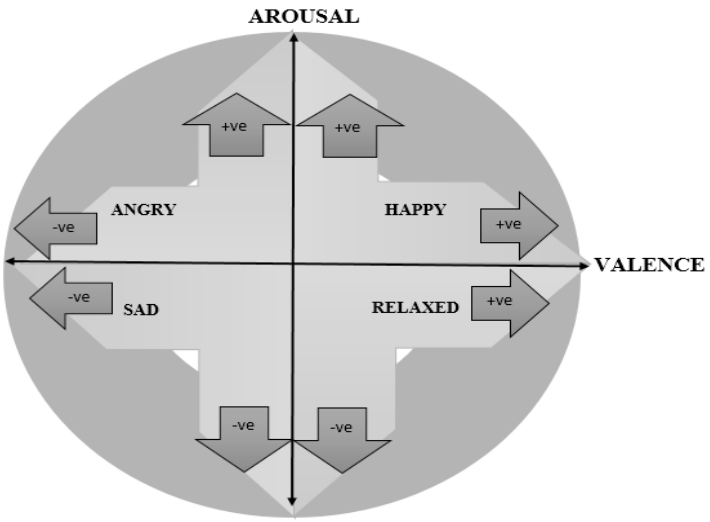


Fig. 1 The dimensional model of James Russell [28] [36]

The literature points to several studies that can be referred in order to recognize the human emotions under multi-modality. Valence and arousal are the important aspects of most dimensional space models. The Russell's Circumplex model , Plutchik's model [22] and the vector mode are some widely used dimensional models [31]. This research work applies the Russell's Circumplex model [28] and considers valence and arousal dimensions for emotion mapping. The emotions from the text (lyrics here) are thus being identified according to Russell's dimensions [36].

## 4. Dataset

The Music4All dataset [29] contains with multi-modal data, apt for several music information retrieval applications such as recommender system, playlist generation, and emotion detection. The dataset contains more than 1 lakh samples of song information including lyrics, genre, language, tags and 30 seconds of audio clips. The lyrics were available in a text file format.

Preprocessing was done as follows. First, 28K lyrics files were randomly selected and converted into spreadsheet format and the I.D. of the file was used to store it in a unique row format. 12 feature columns were selected from lyrical metadata information and were further used for feature selection for emotion detection.

For annotation of emotions, the MER dataset [11] has lyrics mapped into four quadrants happy, sad, angry and relaxed. As, the authors of the MER dataset have used several features of lyrics like stylistic, semantic, and structure of the song for classification of songs, this research considers that the MER dataset is rich in information about lyrical emotions. Thus this research work uses knowledge acquired from the MER dataset to train the Music4All dataset

## 5. The proposed model

### *5.1 The LyEmoBERT architecture*

The architecture of the LyEmoBERT is illustrated in Fig. 3. The initial part of the LyEmoBERT model is applying transfer learning. The MER dataset is used as a source of knowledge and further, the transfer of knowledge is done on the Music4All (destination) dataset [29]. After that, the BERT model is used to train the destination dataset and re-prediction of the lyrical emotions is done. The main objective lies in labelling the Music4All lyrics using emotions and then using the lyrics' label to produce recommendations using the Sentence Transformer model.

### 5.2. *Emotion categorization using Russell's model*

To label the Music4All dataset [29] using the emotion label the initial move is to identify emotionally relevant features. The *valence* feature is the available emotionally relevant feature and is discussed in [20]. The *valence* refers to the positivity of the music.

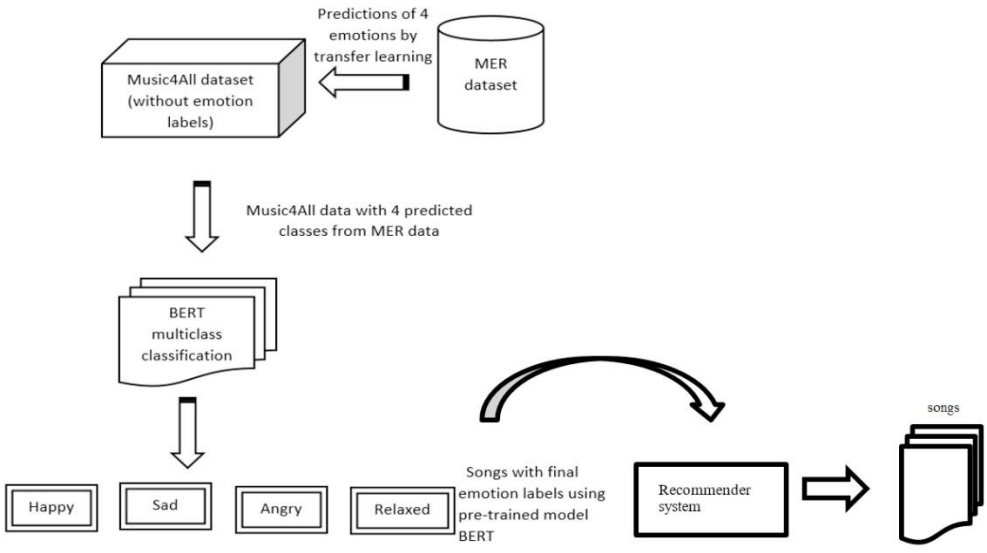


Fig. 2. The proposed LyEmoBERT architecture [36]

multi-class classification using the *valence* and *energy* features in the dataset was conducted in the previous research by the authors of this paper [27]. Here, the music4all dataset was annotated by emotion labels which were calculated by categorizing the valence and energy columns in the x-y coordinate of Russell's Arousal-Valence Plane. The highest accuracy achieved for this approach was 99.7%, but without considering the lyrical features. Owing to this, it was required to do an extension to that approach in [27] to study the use of emotionally relevant lyrical features which is carried out in this paper.

5.2.1 *Transfer learning for predicting the emotions*

The initial step in identifying the emotions of lyrics in this paper is through the learning-based approach. The learning-based approach used here assumes that the source dataset has all the keywords related to the four quadrants of Russell's plane. The in-domain dataset contains 211, 205, 205 and 150 happy, angry, sad and relaxed songs with corresponding emotional keywords (Fig. 1). The CountVectorizer algorithm converts the text keywords into vector space by calculating the frequency of each word in the lyrics [35] [13]. TF-IDF vectorizer was applied to convert the keyword into a feature matrix. For building the classifiers, state-of-the-art algorithms such as MultiNomial Naïve Bayes, Logistic regression, and Linear SVM were used. The following figure shows the results of the Naïve Bayes approach:

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| 0.0 | 0.800000 | 0.837209 | 0.818182 | 43 |
| 1.0 | 0.942857 | 0.825000 | 0.880000 | 40 |
| 2.0 | 0.472973 | 0.897436 | 0.619469 | 39 |
| 3.0 | 1.000000 | 0.030303 | 0.058824 | 33 |
| accuracy | 0.677419 | 0.677419 | 0.677419 | 0 |
| macro avg | 0.803958 | 0.647487 | 0.594119 | 155 |
| weighted avg | 0.797163 | 0.677419 | 0.622466 | 155 |

Fig. 3. Results of Naïve Bayes approach

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| 0.0 | 0.850000 | 0.790698 | 0.819277 | 43 |
| 1.0 | 0.818182 | 0.900000 | 0.857143 | 40 |
| 2.0 | 0.568182 | 0.641026 | 0.602410 | 39 |
| 3.0 | 0.555556 | 0.454545 | 0.500000 | 33 |
| accuracy | 0.709677 | 0.709677 | 0.709677 | 0 |
| macro avg | 0.697980 | 0.696567 | 0.694707 | 155 |
| weighted avg | 0.708192 | 0.709677 | 0.706507 | 155 |

Fig. 4. Results of Linear SVM approach

The results in figure 3 show that classes 0 and class 1 which denote happy and angry had given satisfactory predictions as their precision, recall, and f1-score measures are (80%, 83%, 81%) and (94%, 82%, 88%) respectively. Class 2 which denotes sad emotion had low precision of 47%, a high recall of 90% and an f1 score of 62 %. Class 3 denotes relaxed emotion and had 100% precision but had 30 % recall and 59 % f1-score. The micro accuracy achieved from this classifier is 67%. Next, a linear SVM classifier was built. The SVM classifier gave the results as shown in figure 4. The results in figure 4 show that classes 0 and class 1 which denote happy and angry had better predictions as their precision, recall, and f1-score measures are (85%, 79%, 82%) and (81%, 90%, 86%) respectively. Class 2 which denotes sad emotion had low precision of 56%, a recall of 64% and an f1 score of 60 %. Class 3 denotes relaxed emotion and had 56% precision but had 45% recall and 50 % f1-score. The micro accuracy achieved from this classifier is 71%. The next algorithm used for the classification task is logistic regression. The results of logistic regression are shown in figure 5.

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| 0.0 | 0.756098 | 0.720930 | 0.738095 | 43 |
| 1.0 | 0.777778 | 0.875000 | 0.823529 | 40 |
| 2.0 | 0.605263 | 0.589744 | 0.597403 | 39 |
| 3.0 | 0.483871 | 0.454545 | 0.468750 | 33 |
| accuracy | 0.670968 | 0.670968 | 0.670968 | 0 |
| macro avg | 0.655752 | 0.660055 | 0.656944 | 155 |
| weighted avg | 0.665783 | 0.670968 | 0.667398 | 155 |

Fig. 5. Logistic regression classifier results on MER dataset

The results in figure 5 show that classes 0 and class 1 which denote happy and angry had good predictions as their precision, recall, and f1-score measures are (76%, 72%, 74%) and (78%, 88%, 82%) respectively. Class 2 which denotes sad emotion had low precision of 61%, a recall of 59% and an f1 score of 60 %. Class 3 denotes relaxed emotion and had 48% precision but had 45% recall and 47 % f1-score. The micro accuracy achieved from this classifier is 67%. Among the three classifiers, the linear SVM gave better accuracy, recall, and f1-score for all four classes of emotion. Owing to this, for predicting emotions in the Music4All dataset, linear SVM model results were considered. For predicting the Music4All emotions, NaiveBayes (TF-IDF and Count Vectorizer), Random Forest with Count Vectors, Random Forest with Word-Level TF-ID, Xgboost, Count Vectors and Xgboost, WordLevel TF-IDF algorithms were used and the results displayed in table 1 were obtained.

Table 1: Results of predicting emotion labels in the Music4All dataset

| Prediction Model | Accuracy | Precision | Recall | F1_Score |
|---|---|---|---|---|
| Naïve Bayes, Count Vectors [9] | 0.72 | 0.72 | 0.72 | 0.72 |
| Naïve Bayes, TF-IDF Vectors [9] [7] | 0.67 | 0.67 | 0.67 | 0.67 |
| RF, Count Vectors [16] | 0.70 | 0.70 | 0.70 | 0.70 |
| RF, WordLevel TF-IDF [16] | 0.67 | 0.67 | 0.67 | 0.67 |
| Xgboost, Count Vectors [7] | 0.66 | 0.65 | 0.66 | 0.66 |
| Xgboost, WordLevel TF-IDF [7] | 0.63 | 0.63 | 0.63 | 0.63 |

The Naïve Bayes model's prediction labels were considered emotion annotations for the Music4All dataset. Using the learning-based (transfer learning) approach, the emotion of the Music4All label is predicted. But still, the accuracy needs to be improved. For achieving that, it is required to analyze the distribution of each emotion class when predictions have been completed.

5.2.2 *Multi-class classification*

The multi-class classification method used in this work classifies the lyrics based on 4 different classes: class 0-Happy, class 1 - angry, class 2 - Sad, and class 3 - relaxed. The prediction of lyrics emotion using the in-domain domain dataset and re-training of the dataset using BERT model for emotion prediction are crucial parts of the proposed model, LyEmoBERT. And thus hybrid methodology including BERT embeddings with the integration of transferred knowledge shown in Figure 2 are the main contribution of the proposed model. A prototype of a novel recommendation strategy using the cosine similarity measure and keyword similarity is also proposed in this work. The distribution of classes in the music4All dataset is as follows: 35.7 % for happy songs, 11.4 % for related songs, 16.2 % for angry songs and 36.7 % for sad songs. The labelled dataset is now tested for pre-trained model BERT Large uncased, which gives higher accuracy than the state-of-the-art models. The results with evaluation accuracy Bert-uncased achieved is 91.5% through two-fold approaches transfer learning from MER corpus dataset and pre-training method.

### 5.3 *Tokenization*

Tokenization here means conversion of the lyrics sentences to the BERT tokens which is BERT acceptable format. A sentence “You are young” is converted into tokens: [‘you’, ‘young’] after BERT tokenization, which means each word in the sentence [which are not stop words] is separated into different BERT tokens.

### 5.4 *The LyEmoBERT’s configuration*

The LyEmoBERT's experiment was carried out using BERT-Base, Uncased model with 12-layer [10] in Google Colab. Different batch sizes of 128, 64 and 32 have been experimented with. Learning rates were taken with different values such as 5e-5, 3e-5 and 2e-5. And the training epochs chosen to be 3.0 The overfitting issue is prevented by adding a dropout layer. The average accuracy achieved by the LyEmoBERT model by running the experiment for 10 times using the above-mentioned batch sizes and learning rates has resulted as 92%. 96 % of precision, f1score and recall was achieved. The detailed results are given in Table 3.

Table 2: Multi-class classification results [36].

| Classes | precision | recall | f1-score | support |
|---|---|---|---|---|
| 0 | 0.82 | 0.83 | 0.82 | 1620 |
| 1 | 0.76 | 0.73 | 0.75 | 665 |
| 2 | 0.77 | 0.77 | 0.77 | 1583 |
| 3 | 0.72 | 0.67 | 0.69 | 525 |
| macro avg | 0.76 | 0.76 | 0.76 | 4393 |
| weighted avg | 0.78 | 0.78 | 0.78 | 4393 |

The class happy, class sad and the class angry showed almost the same or little low scores for all evaluation matrices. Under the batch size of 32 and 5e-5 learning rate, the relaxed emotion class achieved 72 per cent of accuracy which is slightly better than the earlier score. For the remaining batch sizes and the other learning rates, the accuracy and precision score was 69%.

The confusion matrix shown in figure 6 displays the analysis between the actual and the predicted values of the four different lyrical emotion classes. The diagonally represented cells can be interpreted as the true positive (T.P.) values. The Happy class showed a T.P value of 1300, the angry class exhibited a T.P. value of 480, the Sad class showed a T.P. value of 1200, and the relaxed class showed a T.P. value of 360. From the further interpretation of the confusion matrix, the True Negative value for the happy class is 2433 and False Positive value is 286 and the False Negative (F.N.) value is 283. The number of instances of happy and sad exists more than the other emotions in the test dataset. This could be a probable reason for the *angry* and *relaxed* emotions achieving comparatively less accuracy.

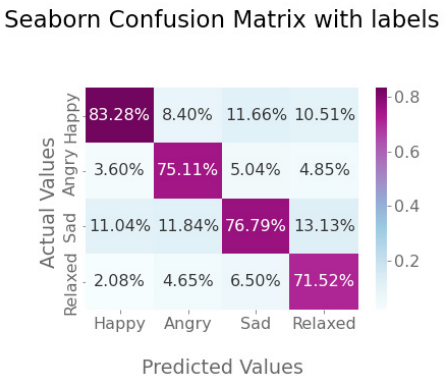


Fig.6. Confusion matrix

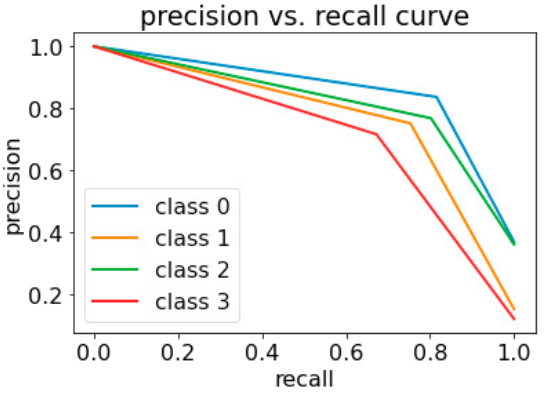


Fig. 7 Precision-Recall curve[

The quality of predictions produced by LyEmoBERT can be deeply visualized with the help of graphical representation. The significance of the prediction (precision) and the relevancy of the outcomes (recall) can be illustrated using the precision-recall curve. This curve depicts the outcomes under variated thresholds.

The plot displayed in figure 7 shows the prediction quality of the LyEmoBERT model. The happy class has occupied the large area and has highest precision and recall scores. The second largest area is occupied by the class sad its precision and recall scores are larger than classes angry and the class relaxed. The third-largest area is occupied by the class angry and the least area is occupied by the class relaxed.
Figure 8 displays the standard ROC (Receiver operating characteristic) plot.

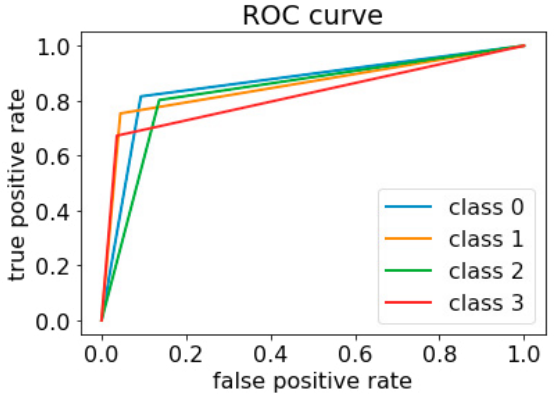


Fig. 8. ROC curve

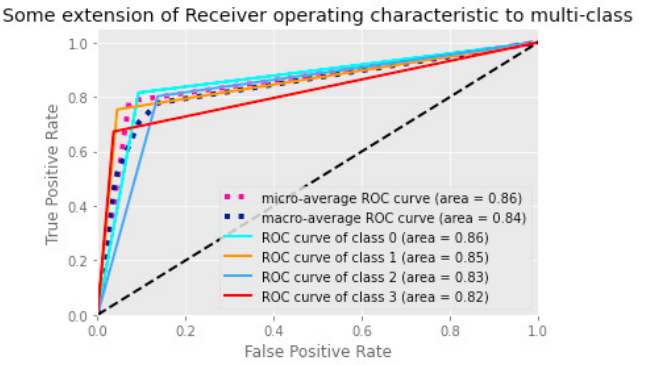


Fig. 9. ROC curve with micro and macro average

The class which exhibit a curve nearing to the top left corner is the one which is the best performer. Similarly, when a line nears to the dotted line in the middle is a worst performer with TPR equivalent to FPR. Interpreting the figure 8 and figure 9, the cyan coloured line indicating the happy class line is the best predictor having an area of 0.86. AUC and ROC plots interpretation can be concluded by the fact that the LyEmoBERT is a promising model
The percentage of predicted class instances for each emotion in descending order is as follows: the happy class, the sad class; the angry class, and the relaxed class with 35.8%, 16%, 36.5%, and 11.6%.

, and

## 6. Lyrics Recommendations

This section discusses the recommendation approach used for generating recommendations based on emotion labels predicted from the Ly-BERT multi-class classification task. The initial task performed for building a recommender system was to identify a transformer model which can find similar music based on emotion through the user's input. The sentence transformer model named *paraphrase-MiniLM-L6-v2* was used for this task. The reason is these models are the best performers in identifying similar keywords and have been applied in paraphrasing and plagiarism detection. This sentence transformer model will convert the keywords in the text into the tensor format to create the embeddings in the lyrics corpus. The query submitted by the user is also converted to a tensor and a query embedding will be created based on the query keywords. The next step is to find the cosine similarity between the query embeddings and the lyrics corpus embeddings. This is performed by the *pytorch_cos_sim* function of PyTorch's util package. Also, to retrieve Top K similar lyrics with the highest similarity score, *torch.topk* function was used. These two processes are shown the figure 10:

```
Query: happy

Top 5 most similar lyrics in corpus:

id
9k3XJ7Srb3wDsUIr
lyrics      dont walk away pretending everythings ok dont care know theres use lies become
truths dont carebert_labels
0
Name: 2604, dtype: object
id
0CCR0R8sEGMDQSsU
lyrics      happy birthday buddha happy hunting happy hiding happy new year hope get
together bert_labels
3
Name: 54, dtype: object
id
D00kRZ59TKyEd5iq
lyrics      save face know youve got one change ways youre young boy one day youll man
oh girl
bert_labels
0
Name: 3475, dtype: object
```

Fig. 10. Query submitted and retrieved results

After that, top recommendations are displayed with the emotion labels predicted early. The following figure displays the top 3 recommendations for lyrics using query word “sad”.

| | lyrics | bert_labels |
|---|---|---|
| 4141 | oh sing sad songs im one tenderly bring soft sympathy ive begun see way clear plain stop fall lay tear pain tear thats oh want says knows moments rare suppose true goes say dont care ah knows maybe sing sake song think decide shes wrong shed like think cruel knows thats lie would tool allowed cry oh sorrow real even though cant change plan could see feel know shed understand oh actually think im blame really believe word mine could relieve pain cant see grieves shes blindly deceived shame ma... | 2 |
| 285 | losing several steps ahead strange anything anyway move strain feels helpless trap build without pulled away say matters quiet face heartbreak shatter okay well train nothing new see saying mean pulled away imagine talks last night never bringing every day want think us dancing something theres dream doesnt sound good pulled away say matters quiet face heartbreak shatter okay well train nothing new see saying mean pulled away pulled away say matters quiet face heartbreak shatter okay well tr... | 2 |
| 1485 | asked content world cherished bring darkened place contemplate perfect future stand utter words tide hate losing sight im sorry seething words say impress ive become anathema soul cant say youre losing always tried keep tied world know leading please tears sympathy cant say losing must though know could take tears sympathy gracefully respectfully facing conflict deep inside confined losing control could change gracefully respectfully ask please dont worry dont turn back dont turn away cant s... | 2 |

Fig. 11. Sample lyrics recommendation

The figure 11 shows the top 3 lyrics based on the sad keyword and cosine similarity between the lyrics. Also, the bert_labels column shows the emotion prediction using multi-class classification. It can be inferred that the label emotion is shown as 2 for all lyrics. As emotion class 2 denotes *sad* and the query input was also given as *sad*, the above figure proves that the multi-class classification task has given satisfactory results as both input query and predicted emotion stands the same for the top 3 lyrics. Now to build user-song recommendations, a deep dense neural network was used. The lyrics with predicted emotions were 4392 in the count. Initially, for building the pivot table for a recommendation, user-ids, emotion labels, title, album and the song count were merged into a single CSV file.

For user-song recommendation emotion labels annotated using valence-energy columns as well as BERT-emotion models were also considered. Using these features, each song was grouped according to the users' listening count and a listening score was obtained for them using the deep dense neural network of 5 layers. The first input layer contained 64 nodes, the second layer (first dropout layer) contains 64 nodes, the next hidden layer contains 32 nodes, the third hidden layer contains 16 layers and the final prediction layer contains a single output node The *relu* activation function was used in all nodes as it is found to be apt for the prediction task of the recommender system. The dataset does not contain user ratings. Owing to this, the recommendations are generated by considering the listening score. The training and validation losses were plotted across each epoch and the mean absolute error was plotted during each epoch shown in figure 13 and figure 14. The training loss was 0.1367, the mean squared error was 0.1367, the validation loss was 0.1350 and the mean squared error for validation and test set were 0.1350 and 0.1381 respectively (figure 12 and 13). Training loss seems to be decreasing after 4 epochs. Here mean squared error shows the predicted listening score and actual listening score. The mean squared error (MAE) is calculated using the following formula:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|p_i - r_i| \quad (1)$$

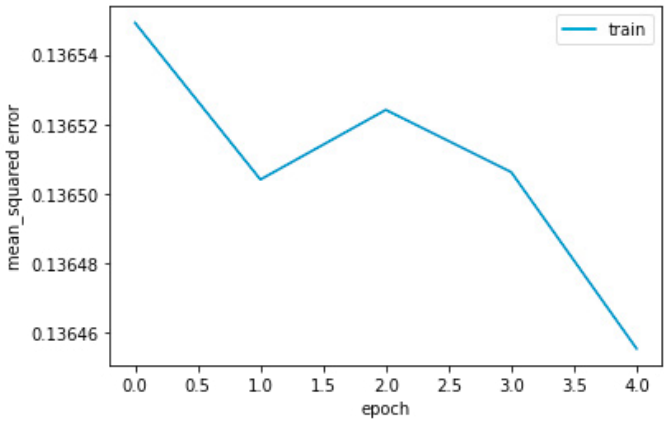


Fig. 12. Training loss

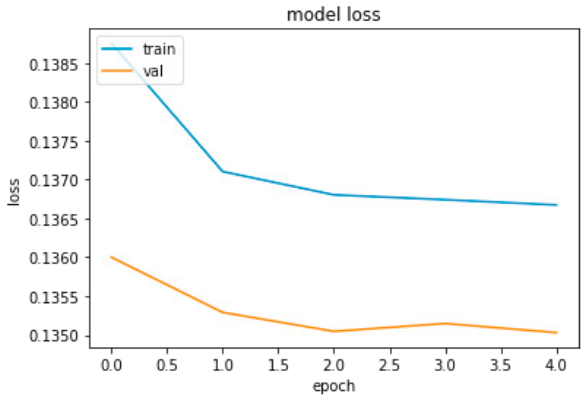


Fig. 13. Training and validation loss

Top K recommendations based on *sad* emotions are shown in the following figure. Discounted Cumulative Gain (DCG) was used as the metric to identify the relevance of ranking the quality of the listening score. After calculating the DCG, Ideal Discounted Cumulative Gain (IDCG) and Normalized Discounted Cumulative Gain (NDCG) were measured and the NDCG graph is plotted as shown below:

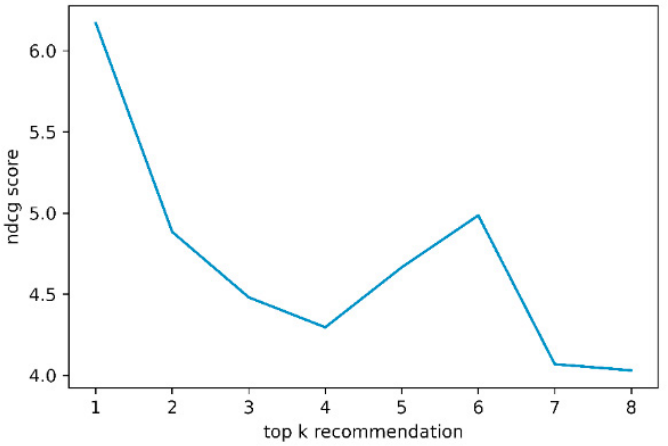


Fig. 14. NCDG curve

The plot in the figure 14 illustrate the NCDG score and a relevant rank of 80%, when k=6 for top k recommendations.

## 7. Discussion

The proposed model LyEmoBERT class has achieved the highest accuracy for the class *happy* and the lowest accuracy was shown by the *relaxed* class. The imbalanced test case samples are found to be the reason behind the low accuracy. Being the objective of annotating the unlabeled dataset and emotion-based classification given preference, the variations in the accuracy scores because of the low training sample are not taken as a major concern. The proposed model is performing better than the work discussed in [29] where the emotion classes were not appropriately identified. This work considers more relevant features from the lyrics and is able to learn unseen lyrics.

The LyEmoBERT's experiments on the validation part showcased overall accuracy and f1 score of 96%, the recall score value as 96%, TN of 558 counts, TP of 1112 counts, and the false negative and false positive counts equivalent to 46 and 41 respectively. These outcomes prove that the LyEmoBERT model has performed better and is shown diagrammatically in figure 11. The test data is showing a different story based on the fact that the happy: sad ratio is equivalent, and the angry and relaxed class sample counts are comparatively very less, as seen in figure 14.

The findings of the performance evaluation can be concluded to the point that imbalanced classes play a vital role in the classification task. But the proposed work showcase a proven improved result when compared to [29]. The final outcome of this work is a subset of the Music4All dataset with emotion label annotation which can be used for emotion recognition of lyrics. A prototype of the emotion-based lyrics recommender system is developed by learning the corpus embeddings. When compared to some recent works such as [24], the proposed model in this work achieved an overall accuracy of 92 % on evaluation. Some other recent works are also compared in table 3 given below:

Table 3: Comparison

| Related works and Proposed model | Results |
|---|---|
| Multi-Task Symbolic Music Neural Network(MT-SMNN) [24] | Accuracy: 67.58, Macro F1: 0.664 |
| Residual-Inception Blocks based MER [17] | Accuracy: 87.64 |
| MEC based on valence and energy features [29] | Valence and energy's MSE score of 0.19 according to the AV plane |
| Emotion-aware Chinese song emotion recommendation[6] | F1-score 88% |
| Songs Recommendation[14] | F1-score 77% |
| Lyrics recommendation based on lyrics similarity [21] | F1 score 74% |
| Proposed Model (LyEmoBERT) | **Accuracy: 92%, F1 Score 96%** |

As shown in the above table, the proposed model has the highest accuracy of 92% among other state-of-the-art approaches through the transfer learning approach.

## 8. Conclusions

In this research work, a multi-class classification of lyrical emotions and recommendations based on those emotions were carried out. The LyEmoBERT model obtained better results when compared to existing work on the Music4All dataset [29]. The significant improvement in terms of accuracy for each emotion class is between 67% and 82%. The integration of an in-domain music corpus for emotion classification based on the same emotion dimension proves to be an efficient transfer learning approach. Considering this work as an initial step, and in the future more techniques to build an efficient emotion classifier have to be planned. The proposed work can contribute well to emotion-based information retrieval such as recommendation systems. More emotions and different types of dimensional planes need to be explored in the future to extend the LyEmoBERT. The main continuation of this work is to enhance the LyEmoBERT results by balancing or augmenting the available data.

### Acknowledgements

We thank the authors Igor Andre Pegoraro Santana, Leonardo Catharin, Juliano Donini, Rafael Biazus Mangolin, Valeria Delisandra Feltrim, Fabio Pinhelli, Yandre Maldonado e Gomes da Costa, and Marcos Aurelio Domingues from the State University of Maringa (UEM), Brazil who permitted to access to the Music4All dataset for this research work.

## References

[1] Agrawal Y, Shanker RG, Alluri V (2021). “Transformer-based approach towards music emotion recognition from lyrics.” *In: European Conference on Information Retrieval* 167-175.

[2] Balakrishnan A, and Dixit K (2014). Deepplaylist: using recurrent neural networks to predict song similarity *Stanford University*: 1-7.

[3] Barthet M, Fazekas G, Sandler M (2012). “Music emotion recognition: From content-to context-based models.” *In: International symposium on computer music modeling and retrieval* 2012: 228-252.

[4] Çano, E.(2018), ”Text-based sentiment analysis and music emotion recognition”. *arXiv preprint arXiv:1810.03031*: 1-131.

[5] Chang WC, Yu HF, Zhong K, Yang Y, Dhillon IS (2020). “Taming pretrained transformers for extreme multi-label text classification”. *In: Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining*, 3163-3171.

[6] Chen X, and Tang TY (2018). “Combining content and sentiment analysis on lyrics for a lightweight emotion-aware Chinese song recommendation system. *In: Proceedings of the 2018 10th International Conference on Machine Learning and Computing 2018*: 85-89.

[7] Chen, T. and Guestrin, C. (2016),” Xgboost: A scalable tree boosting system”. In Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining: 785-794.

[8] Choi K. (2021) “Bimodal Music Subject Classification via Context-Dependent Language Models”. *In Diversity, Divergence, Dialogue: 16th International Conference, iConference 2021* (**12645)**: 68-77.

[9] Cortes, C. and Vapnik, V. (1995). “Support-vector networks. Machine learning”, **20(3)**:273-297.

[10] Devlin, J.; Chang, M.W.; Lee, K.; Toutanova, K (2018). “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding” *2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*: 4171–4186.

[11] Edmonds, D. and Sedoc, J (2021) Multi-Emotion Classification for Song Lyrics”. In Proceedings of the Eleventh Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis: 221-235.

[12] Ekman, Paul (1992). "Facial Expressions of Emotion: New Findings, New Questions". *Psychological Science*. **3 (1)**: 34–38.

[13] Goldberg, Y.; Levy, O (2014). word2vec Explained: Deriving Mikolov et al.'s Negative-Sampling Word-Embedding Method.

[14] Gupta V, Jeevaraj S, Kumar S (2021). “Songs Recommendation using Context-Based Semantic Similarity between Lyrics”. In2021 IEEE India Council International Subsections Conference (INDISCON): 1-6.

[15] Jin, D., Jin, Z., Zhou, J. T., and Szolovits, P (2020). “Is BERT Really Robust? A Strong Baseline for Natural Language Attack on Text Classification and Entailment.” *Proceedings of the AAAI Conference on Artificial Intelligence*, **34(05)**: 8018-8025.

[16] L. Breiman (2001). Random forests. Machine Learning, **45(1)**:5–32.

[17] Liao Jr, Yi (2022). "A Music Playback Algorithm Based on Residual-Inception Blocks for Music Emotion Classification and Physiological Information." *Sensors* **22(3)** : 1-22

[18] Malheiro, R., Panda, R., Gomes, P. J., & Paiva, R. P. (2013). "Music emotion recognition from lyrics: A comparative study". In 6th International Workshop on Music and Machine Learning–MML 2013–in conjunction with the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases

[19] Michaela Vystrčilová and Ladislav Peška. (2020) "Lyrics or Audio for Music Recommendation?" In Proceedings of the 10th International Conference on Web Intelligence, Mining and Semantics (WIMS 2020). Association for Computing Machinery: 190–194.

[20] Panda, R. E. S., Malheiro, R., Rocha, B., Oliveira, A. P., & Paiva, R. P.(2013) "Multi-modal music emotion recognition: A new dataset, methodology and comparative analysis." *In 10th International Symposium on Computer Music Multidisciplinary Research (CMMR 2013)*: 570-582.

[21] Patra BG, Das D, Bandyopadhyay S. (2017) "Retrieving similar lyrics for music recommendation system". *In: Proceedings of the 14th International Conference on Natural Language Processing (ICON-2017)*: 290-297.

[22] Plutchik, R (2001). "The nature of emotions: Human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice." *American scientist* **89(4)**:344-350. 2001

[23] Qaiser, S.; Ali, R (2018). Text Mining: Use of TF-IDF to Examine the Relevance of Words to Documents. International Journal of Computer Applications. **180**

[24] Qiu, Jibao, C. L. Chen, and Tong Zhang (2022). "A Novel Multi-Task Learning Method for Symbolic Music Emotion Recognition." arXiv preprint arXiv:2201.05782, 2022.

[25] Rastogi P, Singh V (2016). "Systematic evaluation of social recommendation systems: Challenges and future." *International Journal of Advanced Computer Science and Applications.*:7(4).

[26] Revathy V.R and Anitha S.P (2021). "Binary Emotion Classification of Music using Deep Neural Networks" Lecture Notes in Networks and Systems. Proceedings of the 13th International Conference on Soft Computing and Pattern Recognition. **417**: 484–492

[27] Revathy V.R., Pillai A.S (2022). "Multi-class classification of song emotions using machine learning: Comparative analysis" ICIETE 2022, IEEE Xplore (In Press).

[28] Russell, J.A. (1980)" A circumplex model of affect*". Journal of personality and social psychology*, **39(6):**1161.

[29] Santana, I.A.P., Pinhelli, F., Donini, J., Catharin, L., Mangolin, R.B., Feltrim, V.D. and Domingues, M.A.(2020) "Music4all: A new music database and its applications." *In 2020 International Conference on Systems, Signals and Image Processing (IWSSIP)* 399-404.

[30] Sun, F., Liu, J., Wu, J., Pei, C., Lin, X., Ou, W. and Jiang, P., (2019) "BERT4Rec: Sequential recommendation with bidirectional encoder representations from transformer." *In Proceedings of the 28th ACM international conference on information and knowledge management*: 1441-1450.

[31] Tweney, Ryan D. (2003) "Wilhelm Wundt in History: The making of a scientific psychology." :318-319.

[32] Vystrčilová, M. and Peška, L. (2020). "Lyrics or Audio for Music Recommendation*". In Proceedings of the 10th International Conference on Web Intelligence, Mining and Semantics*: 190-194.

[33] Z. Gao, A. Feng, X. Song and X. Wu (2019) "Target-Dependent Sentiment Classification With BERT," *in IEEE Access*, **7**: 154290-154299.

[34] Zeng, Z., Xiao, C., Yao, Y., Xie, R., Liu, Z., Lin, F., Lin, L. and Sun, M (2021) "Knowledge Transfer via Pre-training for Recommendation: A Review and Prospect". *Frontiers in Big Data*, **4**: 1-11.

[35] Zhang, Y.; Jin, R.; Zhou, Z.H (2010). "Understanding bag-of-words model: A statistical framework". *International Journal of Machine. Learning and Cybernetics* **1**: 43–52.

[36] Revathy, V R.; Anitha S P.; Fatemah Daneshfar. "LyBERT: Multi-class classification of lyrics using Bidirectional Encoder Representations from Transformers (BERT), 05 April 2022, PREPRINT (Version 1).