

An Automatic Ensemble Multi-Model Chat Context Music Recommender Model

Jaisharma. K

Assistant Professor, Department of CSE
Saveetha School of Engineering, Saveetha Institute of
Medical and Technical Sciences
Chennai, Tamil Nadu, India.
k.jaisharma@gmail.com

S. Thirumal

Assistant Professor, Department of CSE
Vels Institute of Science Technology and Advanced Studies
Chennai, Tamil Nadu, India.
selvarajthirumal@gmail.com

Abstract—Music has the ability to change the human's vibes, emotions and moods. Most people have the habit of listening to music while multitasking and travelling. In this article, focused on developing a model to recommend music automatically based on the text message sent in the chat applications. The propose an Ensemble Multi-Model Chat Context Music Recommender, which is influenced by BiLSTM Encoder to categories the user chat text emotion, user personalized characteristic features using Demographic information and the text emotion, personalized traits and available music playlist using Integrated DeepCNN sub models mapped to give immense user experience. The detailed working and implementation results in 98.55% accuracy with less loss ratio and performed better than other models for WhatsApp-Chat dataset.

Keywords—BiLSTM; Chat; Commercial Context; DeepCNN; Demographic; Enterprises; Environmentally Music; Recommendation; Sound Technologies; Text

I. INTRODUCTION

In today's computer driven environment, the most convenient way for communication are chat applications, people use text and emojis to express their views in the social networking platforms and it becomes inevitable. These social networking platforms constantly generate huge amounts of data which opens up a gate for the researchers to explore more on those data. People also feel these social networks as a vent to express their feelings. so, the analysis of these data helps to identify the people's emotions based on the word's authors used to express. As emotion detection may be used to the behaviors or the mood of the person at a particular time, already in the digital world there are some streaming platforms which use expressions of the people to suggest wish lists [1].

The wish lists can be a watchlist or playlist of songs, in the upper hand the musical industry has evolved much in the size of the economy. It is expected that the compound annual growth would be 14.7 % from 2022 to 2023 which was 29.45 billion in 2021. As the smartphone users are increasing day by day the market size of music will also go up accordingly. It is also observed that people hear music based on their mood and emotions, and it was clearly evident during the pandemic like covid -19 people who were in the quarantine have faced more moods so authors were suggested to listen to music. There are some results that show music has more potential for induction

of future emotional states which is termed as Inductive system [2].

The Inductive system assumes the perceived emotions are always consistent and induced emotions are not consistent. In these systems mainly have two main views one is the system view and the other is user view, the system views basically describe the physiological theories based on user interactions and the user view gives content interpretation. Emotions elicited in users will not always go hand in hand with the perceived emotions [2], [3].

As part of Intelligent music recommendation several studies show that emotions have a vital role in it. It has been found that there are very few attempts to create musical models based on the emotions and their changes. In this paper, exploration is done in the direction to focus on creating an intelligent music recommendation system based on the user's emotions expressed in text over a period of time, in order to recommend such a system machine learning algorithm are used [3].

The propose a model to recommend music based on the text message by the intended users. This model helps to automate the musical automation to choose the music playlist for hustles free user music experience. This model can be used for music listeners, where the music will be played according to the text used to type in the chat application. Also, can be used for commercialization of the music streaming applications for the prime users [4].

The goal of this work is as follows: (i) To categories the user text emotion using BiLSTM Encode as Anxious, Calm, Contentment, Depression, Energetic, Exuberant, Frantic and Happy. (ii) To estimate the use of demographic scores about personality traits, cognitive styles, cognitive abilities and domain experience during the registration process. (iii) To map the user categorized emotion with the music playlist such as Pop, Rock, Dance, Jazz, Hip-Hop, Metal, Blues and Folk using Integrated DeepCNN. This article is organised as follows: section 2 Literature Survey, Section 3 Proposed Work, Section 4 Dataset and Experimental Setup, Section 5 Discussion and Section 6 Conclusion.

II. LITERATURE SURVEY

Roberto et.al proposes a music recommendation system which provides support the interaction process by understanding the emotions induced based on the reinforcement learning (RL) in various scenarios of Markov Decision Process and for decision making authors used Go Explore method but the limitation in this work is collection of the rating for every item as the user may show less interest in giving it and it also requires more and attention and effort for recommendation generation and becomes more challenging in collaborative filtering [5].

Roberto and Alfonso Guarino propose a system for gym environments with a music atmosphere with Machine learning methods and genetic algorithms like Regression methods (including Random Forest) - Non-parametric Kruskal-Wallis H tests show improved Gym Intelligence users' training quality and comfort [6].

Vedhas Pandit et al., presents a shallow convolution Neural Network based emotion recognition which is potentially fast and robust. This work uses the concepts of trained filter weights visualisations and facial action as a unit of activation for demonstrating the generalizability of the approach across different cultures. Also it shows that the proposed approach can be practically applied to pain intensity prediction [7].

Tingting Wanga et al., presented a probabilistic model which retrieves the interests of the user and underlying topics from the chat logs of the health care. To have clear understanding about the participants in the chat groups of online healthcare, the system provides more details of them by exploring the individual interests and underneath information in the discussion topics by using hidden variables separately for each. This model is benchmarked using Author Topic Model (ATM) and latent Dirichlet allocation (LDA) which is also evaluated on a real time online chat group of pregnant women. The results of the model in the real time shows a best result to outperforms various benchmark models and more effective in retrieving the knowledge from social media [8].

Mehmet Bilal et al., presents a new method for recognizing human emotions using EEG signals and transfer learning with deep learning models. The proposed method shows promising results in emotion recognition while listening to music, with the best classification result achieved using the VGG16 model. It is demonstrated that pre-trained deep learning models can be effectively used for the problem of recognizing human emotions, even with limited original training data. Also, it highlights the importance of EEG-based emotion recognition in various fields such as human-computer interaction, health, safety, and education [9].

Vybhav Chaturvedi et al., provided a systematic review of existing research in the field physiological signal based emotion recognition and music mood detection in humans. This work shows a serious investigation of models used for music mood assessing and emotion detection techniques. This work has brought out the more underlying information in the field of detection of emotion of human and music and more about the associated problems [10]. Byeong-jun Han et al., presented a system to recognize the emotions which extracts

the seven distinct features from music using a support vector machine and maps them into eleven types of emotion categories based on Thayer's two dimensional emotion model.

The polar representation of the classifier using a support vector machine has shown a convincing result with accuracy of 94.55% which is higher than other classification algorithms [11].

Mohammad Mehedi Hassan et al., proposes the use of an unsupervised deep belief network (DBN) for feature extraction from physiological signals, which are then combined with statistical features to classify human emotions. The DBN-based model shows clear discriminative features for the "Happy" emotion, achieving 100% classification accuracy. However, it faces difficulties in discriminating neutral emotions from other affective states, resulting in the lowest classification accuracy of 53% for neutral emotions [12].

Xinzhi Wang et al., introduces a methodology that brings out the emotion correlations in the emotion recognition system using the deep learning models. It also contributes in tracking the development in the emotions and understanding the emotion complexities. It also brought out the highlights of emotion analysis of crucial applications and attracted attention of researchers and provided directions to extract emotion correlations [13].

Ankush Chatterjee et al., highlights the relatively new research area of classifying textual dialogues based on emotions and mentions that most work in emotion detection has focused on image-based emotion recognition and this work focuses on understanding emotions in text, specifically in chat conversations, and does not address other forms of communication or contexts [14].

III. PROPOSED WORK

Automation is used to simplify the human needs as it is needed with less manual effort, by keeping in mind this research work is performed for automatic mindset-based songs recommended from the conversation text. Similar to the online shopping products recommendation, in work the model will suggest or recommend the songs based on the conversation happening between the two persons via the mobile chat applications. This application-oriented model will help the users to make the text communication in a smoother way, in a good pleasant inner peace. In Ensemble Multi-Model, three sub-level models are designed namely Linguistic feature extraction using BiLSTM Encoder, demographic user profile extractor and DeepCNN Music Recommender Integration model.

Figure 1, represents the basic block flow diagram of the proposed work Ensemble Multi-Model Chat Context Music Recommender. The chat conversation between the user is analyzed using API to extract the text as shown below.

In the first stage work, used the text conversation from the dataset, the corpus preprocessing cleaning process such as Case Folding, Tokenization, Filtering, Stemming and Normalization were performed before extracting the features. The features and

pattern discovery are made with the support of BiLSTM Encoder, which has the essence of NLP formulations [15].

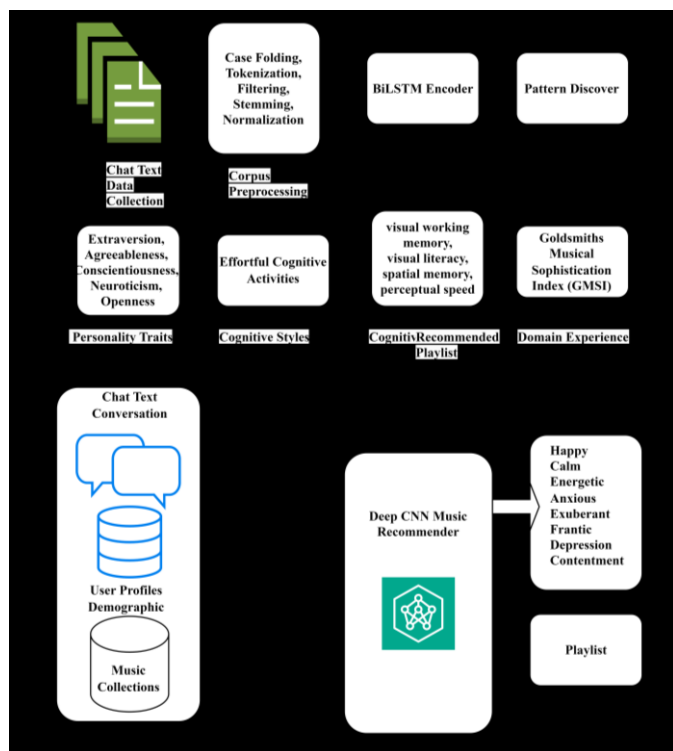


Figure 1. Block diagram of Ensemble Multi-Model Chat Context Music Recommender

In the second stage of the model, which collects the information about the texting users to build the demographic data such as user experience, abilities and personality. The third work is the integration model of stages and to recommend the music playlist according to the text characteristics using DeepCNN model.

A. Algorithm 1: Ensemble Multi-Model Chat Context Music Recommender

Inputs:

Required 1: Chat Text Corpus collection

Required 2: User specific profile information

Required 3: Music playlist collections

Outputs: Recommended Musical playlist

Start

Define variables S, t, R, M

Initialization t=0 time sequence

while Message M belongs to String S **do**

 Train ibatch size, target [redacted] Y

end while

Set hyperparameter tuning

while node_dimen=100, 130, 300, maxi_feat=1000, epoch=27, ibatch_size=32 **do**

 Vector_dimen M[m1, m2, m3,...mn] encode BiLSTM

 Identify similarity pattern P=i=1n z1 X z2

return P

end while

while estimate GMSI(G, i, j)

```

for personality trait and cognitive path
    [redacted]
    path_score=intelligence(path_index)
end for
return MLE path_score
while pattern and path_score do
    for pattern from M do
        [redacted]
    end for
    for path score from R do
        [redacted]
    end for
end while
while [redacted] and [redacted] do
    for t=i+1 iterations do
        Determine fitness for music_cooridates MC=range 0 to 1
        Cross-EntropyLoss
        [redacted]
    end for
end while
return the best music recommended playlist-based chat
conversation
    
```

The step-by-step processes involved in proposed work Ensemble Multi-Model Chat Context Music Recommender provided in Algorithm 1, which encompass the BiLSTM encoder, Deymographic user profile and Integrated DeepCNN model processes.

B. Linguistic feature extraction using BiLSTM Encoder

Bidirectional Long Short-Term Memory (BiLSTM) encoder is the efficient model to train the Natural Language Processing (NLP) using the RNN variant. The LSTM suffers in predicting the information only from the past sequence of words and RNN suffers in vanishing gradient problems, sometimes exploration problems causing. To overcome such issues, BiLSTM encoders were used, in which forward and reverse flow used to predict form past text and as well from the future text typed by the user. Another purpose of using BiLSTM encoder is effectively handling variable input text sequences, without any restrictions. During the preprocessing Case Folding, Tokenization, Filtering, Stemming and Normalization.

The case folding ensures the standard form of text from the given input, for more sensible operation it converts all text in lowercase. So that the converted text can be used for various text manipulation operations such as comparison, sorting, searching and pattern identifications. The unstructured text meanings were extracted by chunking the complete sentence into small meaningful single units using the tokenization libraries. The tokenization can be performed in different levels such as word level, sentence level and character level according to the necessity of the BiLSTM encoding process. The filtering stage supports the part-of-speech (POS) tagging, Named Entity Recognition (NER) during the subsequent processes. Additionally, this stage helps to clean the noisy and

irrelevant data, which also deals with removal of stop words and special characters. In this BiLSTM encoder using the filtering process, labels the input text under two categories such as flagged and not-flagged. The text in corpus in English labelled as not flagged and text available in Hindi labelled as flagged. For more accurate understanding of the text, the root form of the given text was identified by removing the prefix and suffix with the Stemming process. Along with the above-mentioned normalization process, lengthy sentences are padded with uniform string length to perform the task more consistently and effectively.

The BiLSTM Encoder model architecture diagram with the number of nodes used for model building is represented in Figure 2. The initial input layer consists of 100 forward and backward nodes, in this layer performing the preprocessing tasks such as removal of stop words, punctuations and special characters.

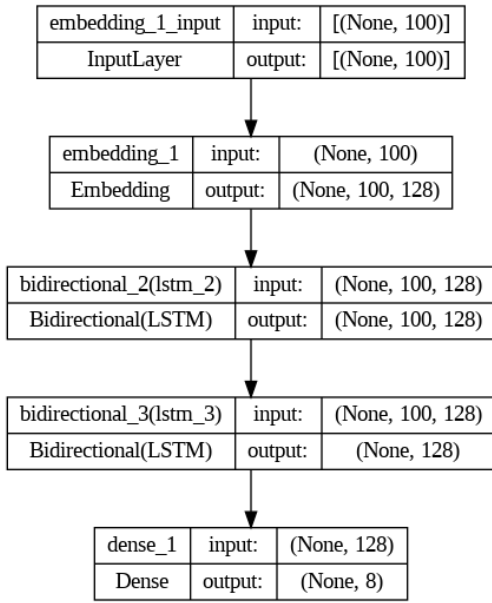


Figure 2. BiLSTM encode model architecture

The second layer consists of a ReLU activation function with 128 nodes for each unidirectional flow for dimensionality reductions. The text features are extracted in the third layer with 128 BiLSTM nodes, which perform padding, dimensionality and normalizations. The next subsequent layer max-pooling, this forms the sentence vectorization based on the relationships identified using POS tagging and NER. The dense layer uses the softmax function to categories the text by concatenating the 8 fully connected nodes. The output of the softmax layer is stored in the vector format, which contains the features and patterns of the given input text message.

$$f_t = \text{sigmoid}(w_f [h_{t-1}, x_t] + b_f) \quad (1)$$

$$i_t = \text{sigmoid}(w_i [h_{t-1}, x_t] + b_i) \quad (2)$$

$$\tilde{C}_t = \tanh(w_c [h_{t-1}, x_t] + b_c) \quad (3)$$

$$C_t = f_t \times C_{t-1} + i_t \times \tilde{C}_t \quad (4)$$

$$O_t = \text{sigmoid}(w_o [h_{t-1}, x_t] + b_o) \quad (5)$$

$$h_t = O_t \times \tanh(C_t) \quad (6)$$

The above Equations 1 to 6 show the mathematical functions used in BiLSTM Encoder to determine the text features and patterns.

C. Demographic User Profile Extractor

Information about the person plays an important role in all aspects in this digital world, it is noted that most of the applications, websites, portals, etc are collecting the consumers data. The collected data can be used for personalizing the needs and expectations of the registered users. In this second model, which incorporates personality traits, cognitive styles, cognitive abilities and domain experience scores to make decisions. The human personality information is collected and marks rated based on their feedback responses. The human personality is qualified with parameters such as extraversion which deals with how a person interacts with talkative and social behavioral activities. The person's individuality, cooperative and kind attitude behavior with another community called Agreeableness. The conscientiousness of the person is measured in terms of the ability to be goal-oriented and methodological-oriented for a problem. The emotional stability score is measured in terms of positivism to tackle problems and the openness score measures the ability of idea and practical exposure to handle problems.

The effortful cognitive activities score the mental stability of the person, this will represent the users problem-solving skills, attention making abilities and learning with decision making abilities. The cognitive ability score relates the thinking abilities of the person in terms of visual manipulation, comprehending the information with visuals, re-collecting environmental information and quick decision-making abilities. The music related information is collected using Goldsmiths Musical Sophistication Index (GMSI), where the person's musical knowledge, expertise behaviors were scaled. The GMSI considers the factors such as active engagement, perceptual abilities, musical training, singing abilities and musical emotions of the registered users. The collected information namely personality traits, cognitive styles, cognitive abilities and domain experiences assessed within the scale from 0-10 in the vector matrix.

D. Integrated DeepCNN Music Recommender model

This Integrated DeepCNN Music Recommender model has three inputs, the first input is taken from BiLSTM Encoder in vector matrix format containing the text features and patterns to the context of the communication. The second input taken from the Demographic User Profile in the vector matrix format containing the registered user preferences collected from the consumers. The third input is the latest music playlist fetched using music applications Web API (e.g Spotify Web API). To create multileveled equations, it may be necessary to treat the equation as a graphic and insert it into the text after y paper is styled.

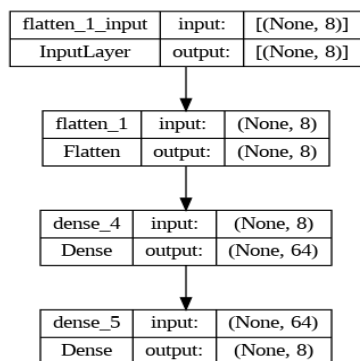


Figure 3. Integrated DeepCNN model architecture

Figure 3 shows the model architecture, the three different inputs were feeder using 8 input nodes, which is integrated in the dense layer and has the signum activation function. Based on the vector matrix input, the signum activation functions the output ranging -1 to 1. From the threshold value the Integrated DeepCNN model categories the user's current mood such as Anxious, Calm, Contentment, Depression, Energetic, Exuberant, Frantic and Happy. According to the mood the music recommender has updated the latest playlist categories such as Pop, Rock, Dance, Jazz, Hip-Hop, Metal, Blues and Folk effectively.

E. Dataset and Experimental Setup

The dataset is downloaded from Kaggle dataset repository, titled as "WhatsApp-Chat" which consist of text messages in English and Hindi. There were 12412 conversation text instances available, along with mobile number of users, text messages exchanged, year, month, day, h and minute attributes of each instance. The Whatsapp message scrapped using API by a private user from his personal account, the corpus is stored in csv format of data from 2018 to 2022. To execute this research, used Google Colab Pro subscription with the specification of 32GB RAM, GPU power of T4, P100 and K80, accessed via Google Chrome browser. The base machine used to access has 8GB RAM, with intel i5 processor, and coded with the latest python programming packages.

IV. RESULT

The proposed model is executed for 30 epochs with 12412 instances of the dataset, 70:20:10 ratio is used for training, testing and validation respectively. The minimum sample size required for justification is 708 using the Gpower sample size calculator, took 12412 samples in all epochs for existing comparison. The measured the model's performance in the aspect of accuracy given in equation (7).

$$accuracy = \frac{TP + TN}{TP + FN + TN + FP} \quad (7)$$

The result of this research experimentation is shown in Table 1, in which proposed model archives an accuracy of 98.58%, whereas the LRA-DNN with 94.77%, DNN with 92.13%, CNN with 90.42% and ANN with 87.69%. The

obtained accuracy values in table 1 are cumulative averages measured for 30 epochs.

TABLE I. COMPARATIVE RESULT OF MUSIC RECOMMENDATION MODELS IN TERMS OF ACCURACY

Model Used for Music Recommendation	Accuracy (%)
Ensemble Multi-Model (Proposed)	98.58
Leaky Relu activated Deep Neural Network [16]	94.77
Deep Neural Network [17]	92.13
Convolutional Neural Network [18]	90.42
Artificial Neural Network [19]	87.69

The graphical representation of the performance is given in Fig. 4, the x-axis all models were taken and for y-axis accuracy scales were used to plot the values of all models.

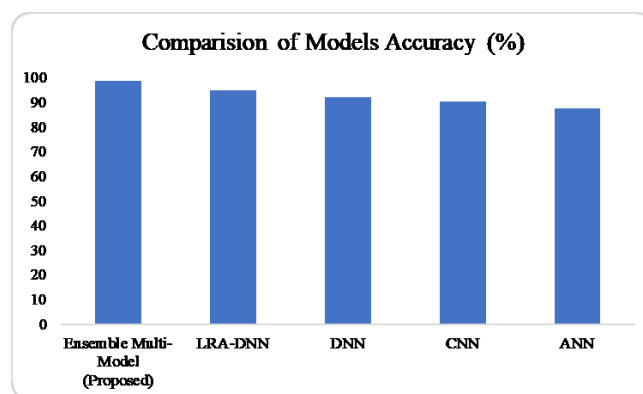


Figure 4. Comparative Bar Graph analysis of proposed model with existing models

V. DISCUSSION

Based on the result accuracy achieved, the proposed model out performs other models referred in this article namely LRA-DNN, DNN, CNN and ANN for music recommendation to the considered dataset. Figure 5 shows the significance of the model loss with respect to the epochs considered, the increase in epoch size gradually decreasing the model error rate. In model, use 30 epochs, initially the error rate starts near to 0.05 and reduces to 0.03 close to the last epoch of the model.



Figure 5. Proposed model validation accuracy with loss for 30 epoch

Similar to that, the Fig. 6 shows the marker line graph of model accuracy, during the first epoch the model accuracy was 97.41%. When the model epoch increases till 30, the accuracy reaches to 98.55% and there is no improvement after the 17th epoch, the model prevails to 98.55% throughout the epochs.

The predicted user preferred music collections statistically analyzed using confusion matrix as shown in Fig. 7. Since model has multiple categories prediction such as anxious, calm, contentment, depression, energetic, exuberant, frantic, and happy of the user, the prediction performance is measured with true positive, true negative, false positive and false negative.

The Fig. 7, shows multiple perspectives from the predicted output, this also helps to finetune the model performance by adjusting the threshold and weightage of the model.

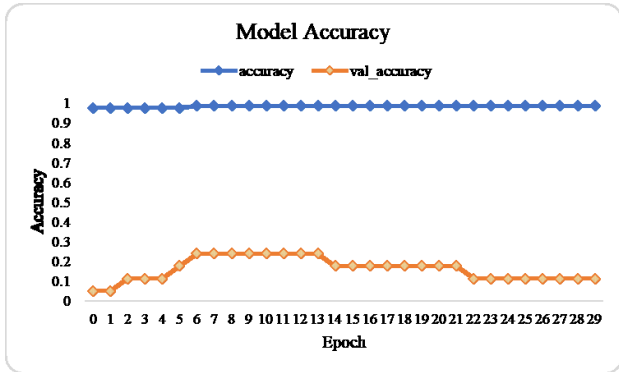


Figure 6. Proposed model performance measured in accuracy for 30 epoch

The important part of work is to recommend the soothing music playlist according to the text context the user typing in the chat application. In model used a playlist like pop, rock, dance, jazz, hip-hop, metals, blues and folk using Web Spotify API interface. The Integrated DeepCNN model correlates the user's mindset based on text and recommends the music according to that automatically.

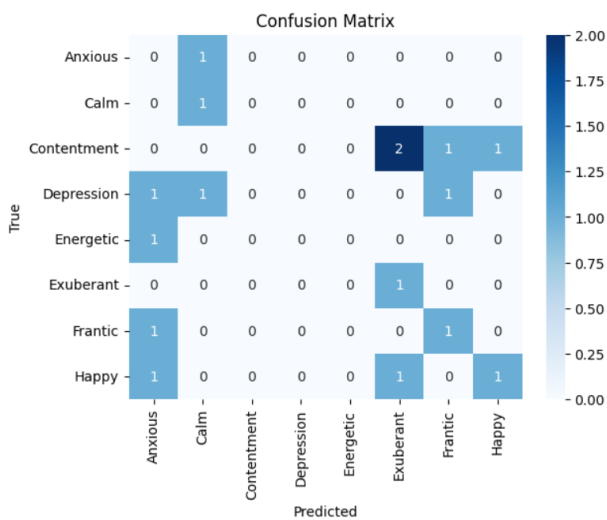


Figure 7. Confusion Matrix for chat context emotion classification

The comparison result is represented in Fig. 8, the left side bar graph is the actual user interested song during the specific

time and the right sidebar shows the recommended playlist category by model. Based on table 1, models prediction level is 98.55% accurate and various figures represented in this article, shows model works efficiently.

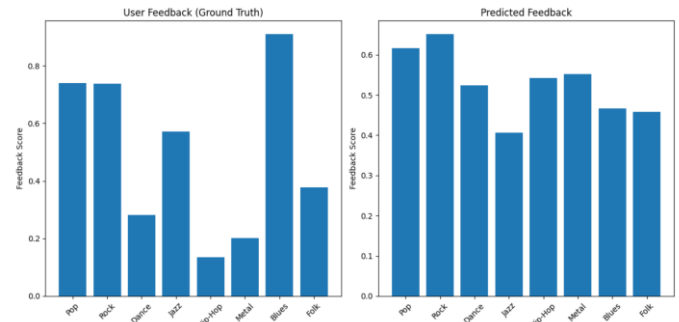


Figure 8. Music recommended bar chart comparison with actual and predicted

The factors that improved the performance of the model are summarized as follows: The usage of BiLSTM Encoder for preprocessing handles the input data in sequential text communication message, which is relatively tagged with time series analysis. The stacked layers and its number nodes, reduces the overhead problems with effective minimal layers. The weightage of each node is adjusted according to the required learning rate from BiLSTM forward and backward flow of data. The demographic user profile contains the measurable abilities and qualities of the user, this makes the precise decision-making factor used in this model. The integrated DeepCNN involves identifying more depth patterns from the complex analysis. As well, the DeepCNN transforms the playlist according to the predicted text classification using the data augmentation process. The model uses self-attention to select the high probability portion of the 8 node inputs, which support the model to explore long-range dependencies.

The author [20], used CNN based music emotion classification, and transformed the musical profile into linear space matrix. The author address and improve the classification accuracy using CNN, which captures the local and global music information. Author Yajian Zhou [21], proposed a Text-CNN model to handle the issues related to short text; authors used a semantic extension framework. The context-level extracts the word level N-Gram and context level form the 1D convolution, also tackles the vanishing and gradient immigration issues. In this article, authors used a self-attention model with CNN to annotate the music profiles and authors addressed labelling tasks of music files [22]. The Neural Network is widely used for many research activities, one such using the social tags to annotate the emotions in the music files, but authors used very limited short segments [23]. The use of BiLSTM and VGG16 for analyzing music genre, authors used disparities in spectrums as the music feature and also worked in single and multiple labelled categories [24]. Based on the above article's references, the proposed work is also considered as authenticated research to compete with the global standards.

Even Though the proposed works automates the music recommendation for the chat users, there are some limitations. Some of them rarely focus on large quality text chat conversations and this work is designed for English language

users. The model's computation has some complexity in parallelization of multiple threads at time. In future, this work can be further expanded to work with native speakers to handle large scale data.

VI. CONCLUSION

Most of the music recommendations available are collaborative-based such as memory and model based, but very research works exist using content-based recommendation. In this article, proposed a content-based model, using BiLSTM encoder, demographic user profile and Integrated DeepCNN. The proposed work Ensemble Multi-Model Chat Context Music Recommender, analysed and categorised the user typed text using BiLSTM Encoder, the user's demographics information was collected before user started to use the application, the music playlist genie updated periodically using web API. The Integrated DeepCNN model consolidates and suggests suitable songs according to the chat user's emotions. The performance of the proposed model is measured in accuracy and achieved 98.55%, compared with other existing models, model performs well to the mentioned dataset. Based on the performance, the suggest that research can be useful for advanced music recommendation using user typed text messages.

REFERENCES

- [1] R. De Prisco, A. Guarino, D. Malandrino, and R. Zaccagnino, "Induced Emotion-Based Music Recommendation through Reinforcement Learning," *NATO Adv. Sci. Inst. Ser. E Appl. Sci.*, vol. 12, no. 21, p. 11209, Nov. 2022.
- [2] Chung, Jaemin, Jiho Lee, and Janghyeok Yoon. "Understanding music streaming services via text mining of online customer reviews." *Electronic Commerce Research and Applications* 53 (2022): 101145.
- [3] Anbalagan, E., Sasikumar, S., Kumar, M.G.V., Paramesh, J. and Sriram, K.P., 2024, July. Advancing Personalized Recommendation Systems with a Groundbreaking Collaborative Filtering Algorithm Driven by Machine Learning. In 2024 Second International Conference on Advances in Information Technology (ICAIT) (Vol. 1, pp. 1-6). IEEE.
- [4] N. Bacanin et al., "Hybridized sine cosine algorithm with convolutional neural networks dropout regularization application," *Sci. Rep.*, vol. 12, no. 1, pp. 1–20, Apr. 2022.
- [5] De Prisco, Roberto, et al. "Induced Emotion-Based Music Recommendation through Reinforcement Learning." *Applied Sciences* 12.21 (2022): 11209.
- [6] De Prisco, Roberto, et al. "Providing music service in ambient intelligence: experiments with gym users." *Expert Systems with Applications* 177 (2021): 114951.
- [7] Sivapriyan, R., Santhakumar, D., Naik, P.K., Banu, E.A., Jaaywel, W.S. and Safi, H., 2024, May. A Composite of Design of Collaborative Filtering Models for Stay Recommendation. In 2024 4th International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE) (pp. 245-247). IEEE.
- [8] Thapaswini, G. and Gunasekaran, M., 2024, October. CaSR-Net: An Ensemble Model for Optimized Crop Recommendation Using Machine Learning. In 2024 IEEE 6th International Conference on Cybernetics, Cognition and Machine Learning Applications (ICCCMLA) (pp. 188-194). IEEE.
- [9] Er, Mehmet Bilal, Harun Çiğ, and İbrahim Berkan Aydılek. "A new approach to recognition of human emotions using brain signals and music stimuli." *Applied Acoustics* 175 (2021): 107840.
- [10] V. Chaturvedi, A. B. Kaur, V. Varshney, A. Garg, G. S. Chhabra, and M. Kumar, "Music mood and human emotion recognition based on physiological signals: a systematic review," *Multimedia Systems*, vol. 28, no. 1, pp. 21–44, Apr. 2021.
- [11] Kathiravan, M., Irumporai, A., Buvaneshvari, R. and Sreesubha, S., 2024, November. A Novel Approach to Enhance Movie Recommendations through Hybrid Collaborative Filtering and Sentiment Analysis. In 2024 5th International Conference on Data Intelligence and Cognitive Informatics (ICDICI) (pp. 1225-1231). IEEE.
- [12] Hassan, Mohammad Mehedi, et al. "Human emotion recognition using deep belief network architecture." *Information Fusion* 51 (2019): 10-18.
- [13] X. Wang, L. Kou, V. Sugumaran, X. Luo, and H. Zhang, "Emotion Correlation Mining Through Deep Learning Models on Natural Language Text." <https://doi.org/10.1109/TCYB.2020.2987064> (accessed Nov. 20, 2023).
- [14] Jaisharma, K., Deepa, N. and Devi, T., 2024, May. Monitoring AI-Based Processing for Predicting Poisonous Gas Emissions in Smart Cities Using Novel Temporal Dynamics Prediction Model. In 2024 Second International Conference on Data Science and Information System (ICDSIS) (pp. 1-6). IEEE.
- [15] Bharath, B., and B. B. Beenarani. "A novel activity recommendation for tourists based on location using recurrent neural network in comparison with convolutional neural network with improved accuracy." In *AIP Conference Proceedings*, vol. 3252, no. 1, p. 020039. AIP Publishing LLC, 2025.
- [16] Wibawa, Putu Widyantara Artanta, and Cokorda Pramatha. "Systematic Literature Review: Machine Learning Methods in Emotion Classification in Textual Data." *Jurnal Sisfokom (Sistem Informasi dan Komputer)* 12.3 (2023): 425-433.
- [17] Singh, Jagendra. "An efficient deep neural network model for music classification." *International Jnal of Web Science* 3.3 (2022): 236-248.
- [18] Jaisharma, K., Deepa, N. and Devi, T., 2024, December. Automated Score Calculation of Similar Words from Language Disabled Children using Hybrid Enhanced Latent Semantic Index. In 2024 International Conference on Communication, Computing, Smart Materials and Devices (ICCCSMD) (pp. 1-6). IEEE.
- [19] D. Han, Y. Kong, J. Han, and G. Wang, "A survey of music emotion recognition," *Frontiers of Computer Science*, vol. 16, no. 6, pp. 1–11, Jan. 2022.
- [20] F. Ke, "Intelligent Classification Model of Music Emotional Environment Using Convolutional Neural Networks," *J. Environ. Public Health*, vol. 2022, Aug. 2022, doi: 10.1155/2022/7221064.
- [21] Senthil, K., G. Teja, and R. Srudhakerthi. "Performance Metrics and Accuracy Assessment of Machine Learning Models for Crop Recommendation." In 2024 International Conference on Electronic Systems and Intelligent Computing (ICESIC), pp. 105-109. IEEE, 2024.
- [22] Z. Miao and C. Cheng, "Construction of multimodal music automatic annotation model based on neural network algorithm," in *Seventh International Conference on Mechatronics and Intelligent Robotics (ICMIR 2023)*, SPIE, Sep. 2023, pp. 482–488.
- [23] N. He, "Neural Networks for Music Emotion Recognition and Social Tags Emotion Representation," 2023. Accessed: Nov. 16, 2023. [Online]. Available: <http://hdl.handle.net/10453/171926>.
- [24] Hongdan, Wang, et al. "An intelligent music genre analysis using feature extraction and classification using deep learning techniques." *Computers and Electrical Engineering* 100 (2022): 107978.