Ensemble Tweets Emotion Detection Model Using Transformer Based Architecture, Support Vector Machine and Long Short-Term Memory

Opeyemi Aderike Abisoye(✉), Abdullahi Bala, Solomon Adelowo Adepoju, Oluwaseun Adeniyi Ojerinde, and John Kolo Alhassan

Federal University of Technology Minna, PMB 65, Minna, Niger State, Nigeria
o.abisoye@futminna.edu.ng

Abstract. In this current age of Fourth Industrial Revolution (4IR), there is an exponential growth in public generated data such as mobile data, business data, social media data, Internet of Things (IoT) data, cyber security data which are in form of image, video and text, due to the incessant usage of social media. This available textual data is frequently adopted and significantly important for extracting information such as user's sentiments, and emotions. Considering the complexity and large amount of textual data, the adoption of various machine learning and deep learning model for the analysis of emotion has not yet attained optimum accuracy. Recently, Bidirectional Encoder Representational from Transformer Based Architecture (BERT) are achieving state of art accuracy. Hence, this study adopts an ensemble-based model using Bidirectional Encoder Representational from Transformer (BERT-Large), Long Short-Term Memory (LSTM) and Support Vector Machine (SVM) for detecting user's emotion. The three trained model are loaded from the local repository and stack together by comparing their predictions and selecting the majority vote approach. This study performs emotional analysis on imbalanced tweets of six (6) different classes, which includes; sadness, anger, love, surprise, fear, and joy. The experiment shows that the voting of BERT prediction and Ensemble model perform better than the other models with to an optimum accuracy of 93%, 93% respectively. BERT-Large performed well as a standalone model and also the ensemble techniques for prediction of multi-social platforms in real time usage.

Keywords: Emotions · BERT-Large · Support Vector Machine · Long short-term memory · Ensemble

1 Introduction

Recent emanation of various social media platform such as WhatsApp, Facebook, Instagram, and twitters has resulted to large amount of publicly available text corpus or dataset [1, 2]. Considering the fact that social media user tends to feel more comfortable expressing their feelings or opinions during chatting session, hence social media data is tag to be more realistic or rich in emotional context [3]. Different opinion of customer feedback on a particular product can facilitate the quick identification of issues

O. Arsenyeva et al. (Eds.): STUE 2023, LNNS 807, pp. 14–25, 2023.
https://doi.org/10.1007/978-3-031-46874-2_2

or improvement that is essential for better customer services. Recently, researcher has focus on various techniques that is capably of analyzing user sentiment and emotion in written form automatically. This approach is known has sentiment analysis or emotion analysis [4, 5].

Furthermore, the analysis of sentiment is described as the process of automating, identifying and deducing user opinion from a written or documented text, this also involve classification of user opinion into either positive, negative of neutral sentiment [5, 6]. While emotion analysis is a type of sentiment analysis but with a wider range of classification class, it involves identifying user emotion been express in text format [7]. Based on research it is reveal that the task of emotion analysis is more challenging than that of sentiment analysis, this is due to greater number of possible emotions classifications [1, 8].

Emotions plays a crucial role in every form of human communication, which often influence an individual perception of an experience, topic or event [9, 10]. User (customer) opinions and feedback on various products, online or offline can be collated, via various mediums such as comments, reviews, message forums, and polls [4]. The gathered data can take the form of either audio, video or text [11, 12]. However, the analysis of human emotion from textual data is a field of study uses Natural Language Processing (NLP), computational linguistic and text analysis to extract subjective information from source materials. The detection of emotion using Natural Language processing is a sub domain of sentiment analysis which primarily focuses on extracting emotional information of the user (writer) from text data. Emotion detection aims to pinpoint specific emotions such as happiness, sadness, love, and fear [13]. Ensemble methods involve the combination of two or more machine learning (ML) or hybridization of ML and deep learning (DL) techniques to achieve better performance and optimal results [14, 15]. These methods take advantage of the strengths of multiple techniques to improve overall performance. The Ensemble methods may include a prediction unit and an optimization unit for more accurate output. They are flexible and they provide higher capabilities in comparison to single methods, and ensemble method becoming increasingly popular due to their potentials [16, 17]. In other words, ensemble base learning involves combining multiple machine learning models which perform intelligent task collectively. Common instance of ensemble models includes

1. Netflix: this system adopts hundreds of machine learning model for predicting possible preferences of movies for user.
2. IBM Watson: this ensemble-based model is the first prediction model to won the quiz game of jeopardy.
3. Google: in recent model google adopt ensemble-based techniques for developing a neural machine translation system [18].

In this digital age exorbitant textual data evolves from emotions and there are computational complexities among them; thus, we need a good machine learning and deep learning model for the analysis of emotion that has not yet attained optimum accuracy.

2 Literature Review

Considering the research work of [9] which focuses on identifying emotions in Roman Urdu text. Despite Roman Urdu is frequently utilize for information exchange on social media platforms, there is limited research on emotion detection in this language. The primary challenge is the lack of benchmark corpora for detecting user emotion from text, which is crucial for various NLP tasks. The analysis of user emotion from a text corpus has various significant benefit such as improving product quality, dialog systems, investment trends, and mental health. The research developed a comprehensive corpus of 18k sentences from various domains and annotated with 6 different categories to focus on the emotional polarity of Roman Urdu sentences. Based on the experimental result its identify that the SVM prediction model had the best F-1 score measure among the applied algorithms.

The study by [1] investigates the effective use of BERT models thus, the bidirectional encoders for the analysis of user sentiment and emotion detection using data from user Tweet. Two different classifiers were adopted for the two tasks, and the performance of the models was evaluated using real-world tweet datasets. Based on the experiments carried out its revealed that the adopted models were able to achieve 0.92, 0.90% accuracy in term of detecting and analyzing user emotion and sentiment.

Paper [13] identifies that the automatic detection of user emotions for their tweet generated data has numerous real-life applications. This research transforms an imbalanced data into a balance data by including a neutral class to a benchmark dataset that contain 4 categorical classes of user emotion thus, joy, sadness, fear, and anger. The study adopts SVM and BERT model for the detection and identification of user emotion using the balanced dataset. A novel ensemble model was proposed by merging the BERT and SVM models. The experiments carried out shows that the proposed model achieved an optimum intelligent of 0.91 in recognizing emotions in tweets.

Sentiment analysis approach is examined by [6] on Yelp restaurant reviews, IMDB movie reviews, and data collected from Twitter. Word-2-Vec, the BERT model, Bag-of-word along with the TF-IDF feature extraction was the word embedding techniques used by the researchers. The sentiment analysis model was developed using techniques this includes; the LSTM, Naïve Bayesian, CNN and SVM model. Model performance was evaluated using metrics such as, F-1 (F), Precision, ROC curve, and Accuracy. In respect to the result, it can be concluded that the word embedding generated by the BERT architecture perform best, and the authors recommend the use of BERT approach to address similar issues.

Paper [8] introduced Bayesian machine learning model for detecting emotions from social media news articles. This model takes into account the hidden connection or relationship that map two sentences together in a document and can detect emotions at both the document and sentence level. The developed model outperforms the existing method on the level of sentence dis-ambiguity and document dis-ambiguity and sentence-level emotion detection, as shown in experiments on public corpora.

Considering the work of [10] introducing a new framework for detecting human emotions in a smart advance city using the application of Internet of a thing to facilitate the smart environment. The developed approach is named VEmoBar, capable of creating some barrier that can virtually senses human emotions through wireless signals and reflections. A problem is defined to determine the optimal placement of VEmoBars to maximize cumulative accuracy. A novel approach and system initialization are proposed to solve this problem, and the results are evaluated through simulations in various scenarios. The article also discusses future issues and challenges for implementing this framework in smart cities.

In this research thesis the author introduces a prediction model that is capable of predicting human emotion from real time image capturing. The model utilizes the deep CNN and it's evaluated against 8 different datasets. However, the researchers [15] thus, minimize the parameters of the network by 909 compared to Vanilla CNN and 509 in respect to current state-of-the-art research. The model achieved 74% accuracy, which is an improvement over previous models with reduced computational complexity. This model can be useful in fields such as elderly care, child therapy, and babysitting where robots need to understand human emotions to provide more customized assistance.

Paper [4] describes the system submitted by the WSSC Team to the EmoEvalEs@IberLEF 2021 emotion detection competition. The proposed model combines transformer embedding with topic information and offense features to classify emotions in social media text. The results showed that the model outperforms the benchmark models that are state of art. A weighted average 0.66% F1 score was achieved with 0.67 accuracy.

These researchers [19] introduces a new recommendation system which incorporates user's emotions with interests to provide personalized product suggestions. Unlike existing approaches, which rely on past feedback, similarity of other users' buying patterns, or a combination of both, the proposed system does not require a large amount of data. Instead, it captures a user's eye-gaze and facial expressions as they browse a website using an inexpensive webcam. Eye-gaze detection is done by extracting the pupil center of both eyes, and calculating a reference point using a joint probability. Facial expression analysis is done by analyzing landmark points on the face to determine the user's emotion. Both methods work in near real-time, allowing the system to provide intelligent recommendations on-the-fly without the need for user feedback or buying patterns.

3 Research Methodology

3.1 Data Collection

This study utilized the available emotion data from a popular data science platform repository (Hugging Face). It's a company that provides an NLP platform for featuring pre-trained models, fine-tuning libraries, and development tools. The platform is well-known for its transformer-based models like BERT and Generative Pretrained Transformer 2 (GPT-2), which have demonstrated excellent performance in multiple NLP tasks. It allows developers to easily integrate NLP into applications and projects through APIs and libraries. Additionally, Hugging Face offers a model hub where users can access pre-trained models, customize them using their own data, and share them with others [20, 21].

Furthermore, the adopted dataset was downloaded from the Hugging Face repository using the *datasets* python API module, and all the training, testing, and validation data samples were downloaded into the Google Colab environment in a CSV format. The downloaded data was archive in the Google Drive for future reference and easy importation into the Google Colab for further usage using the python *panda's* library. It was essential to merge the entire dataset into a single data for proper usage as depicted in Fig. 1. The dataset includes twenty thousand (20,000) emotional tweet data points, with two (2) columns of features which include the tweet text and the corresponding emotion. Figure 1 depicts the sample data point from the downloaded tweet emotions dataset. It's clearly visualized that the *text* column contains various user tweets information while the *label* column contains various type of emotions such has sadness, anger, love, fear and joy. Moreover, the table shows the total number of columns thus, 2 and the total number available row or data entries thus, 20,000 rows.

```
]: data_full
]:
```

	text	label
0	i didnt feel humiliated	sadness
1	i can go from feeling so hopeless to so damned...	sadness
2	im grabbing a minute to post i feel greedy wrong	anger
3	i am ever feeling nostalgic about the fireplac...	love
4	i am feeling grouchy	anger
...	...	...
1995	im having ssa examination tomorrow in the morn...	sadness
1996	i constantly worry about their fight against n...	joy
1997	i feel its important to share this info for th...	joy
1998	i truly feel that if you are passionate enough...	joy
1999	i feel like i just wanna buy any cute make up ...	joy

20000 rows × 2 columns

Fig. 1. The emotion tweets dataset.

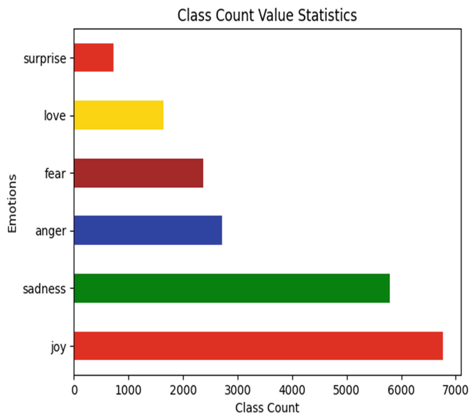

Fig. 2. Tweets emotion classes count.

3.2 Data Exploration

Every data science research requires comprehensive understanding about the dataset for extraction of meaningful features from the dataset. Based on analysis it identifies that the label column contains six (6) different classes, this includes; sadness, anger, love, surprise, fear, and joy. This is visually illustrated using the bar chart in Fig. 2. It identifies that most of the data-sample belong to the *joy* label class with over 6,000 data sample, which is followed by the *sadness* label data-sample with over 5,000 sample, up to *surprise* label sample with minimum number of data sample of 1,000.

In respect to the word cloud in Fig. 3, the frequency of word distribution or frequency in the emotion tweet dataset is explore using the application programming interface for word cloud [22]. Emotional words such as "feel", "feeling", "really", "think", and "make", are the most occurring words in the tweet emotion corpus. The word cloud representation uses text size to indicate how often a word appears in a corpus. Hence, the larger the text size in the figure the more frequently the word appears in the text corpus.

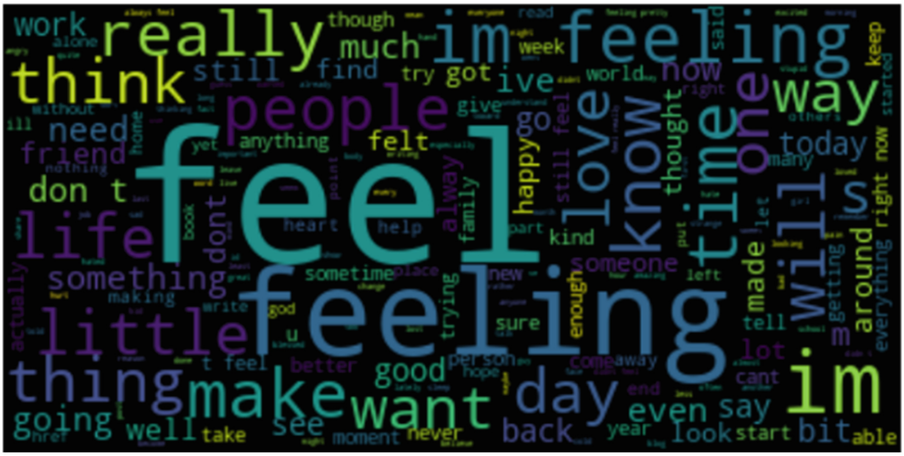

Fig. 3. Word cloud visual word frequency.

3.3 Ensemble BERT-Large, SVM and LSTM

The conceptual diagram of the proposed ensemble model is illustrated in Fig. 4 which combines the Bidirectional Encoder Representational from Transformer (BERT-Large), Support Vector Machine (SVM), and Long Short-Term Memory (LSTM). The illustration of the proposed model is depicted diagrammatically in detail using Fig. 5.

The model uses sub models for predictions and this includes the BERT-Large, SVM, and LSTM model. As depicted in the Fig. 4, the dataset is loaded from the local repository into the Google Colab and training samples are feed into the three models (BERT, LSTM, and SVM) for training. Each machine learning model is trained individually, and the

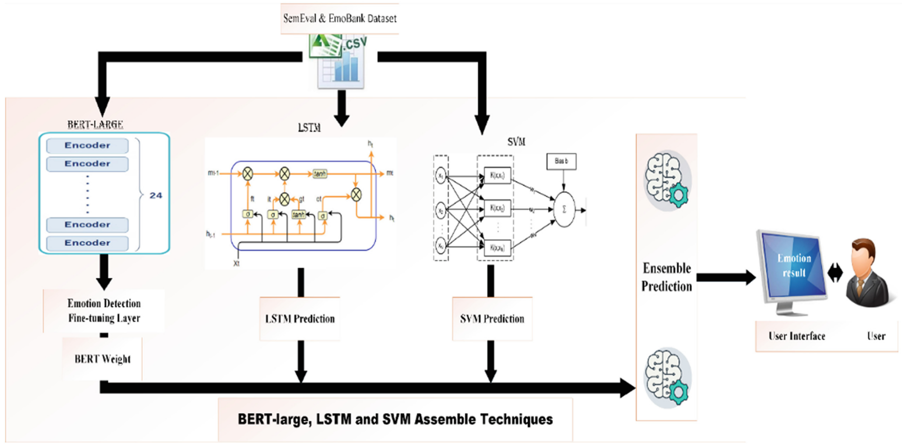

Fig. 4. Conceptual illustration of the proposed ensemble model.

resulting trained models are saved locally on the system directory for application and ensemble usage. The saved models are then loaded, and the prediction outcomes of each model are utilized in an ensemble technique known as voting. The final prediction outcome can be viewed by the user through the user interface.

3.4 Material and Tool

This study adopt various software tools for the development of the proposed model. Then the entire coding is carried out using the Google Collaboratory notebook environment and jupyter notebook, because it provides access to GPU and TPU processing power [23]. Other software tools used Microsoft Visio for drawing the various conceptual diagram. Python is choosing as the programming language of choice due to is richness in data science module.

4 Results and Discussion

The actual implementation and training of the ensemble model was carried out and the various performance evaluation metrics considered includes the Accuracy, precision, recall and F1-score. Finally, the model was evaluated with the existing study for performance comparison. Each model (SVM, LSTM and BERT) is trained separately and later ensemble for emotion prediction.

4.1 Support Vector Machine Model

It's essential to carry out some preprocessing on the textual data before it can be utilized by the SVM machine learning model. Hence, this study adopts count vectorization to transform the textual data into numerical dictionary corpus. Based on parameter tuning (cross fold validation) its identify that the best parameter combination is set to $c = 0.1$, *gamma* = 0.1 and *kernel* = ‘linear’. The preprocessing and training is programmatically achieved using the predefined methods in the sk-learn. The SVM classification report is depicted in Fig. 5.

4.2 Long Short-Term Memory Model

This requires a separate preprocessing technique for proper deep learning utilization of text data. The tensor flow API is use in perform text tokenization, sequencing and padding before they are passed into the deep learning architecture. The LSTM model is trained for 10 epochs and for every iteration the accuracy, loss and validation accuracy were computed. However, accuracy of 90% is achieved after the 10 iterations of training. The classification report is visualized in Fig. 6, putting into considering the various performance metrics.

```
print(report)

              precision    recall  f1-score   support

           0       0.89      0.89      0.89       519
           1       0.81      0.87      0.84       484
           2       0.89      0.93      0.91      1348
           3       0.81      0.70      0.75       328
           4       0.93      0.93      0.93      1169
           5       0.93      0.65      0.77       152

    accuracy                           0.89      4000
   macro avg       0.88      0.83      0.85      4000
weighted avg       0.89      0.89      0.88      4000
```

Fig. 5. SVM classification report

4.3 Bidirectional Encoder Representational from Transformer (BERT) Model

The BERT model weight is download from the hugging face repository and fine turned by adding the emotion classification heads using TensorFlow layers. The architectural structure is shown in the Fig. 4. The model defines two input layers, thus the input ids and the attention mask input with 70 maximum word length. The next layer is the BERT layer follow by a one-dimensional max pooling layer, then we have the dense layer, one dropout layer, another dense layer and finally an output layer. After 10 epochs of training the model was able to achieve an accuracy of 93% as shown in Fig. 7.

Generally, the result gotten from the model (SVM, LSTM, BERT) training are summarize in Table 1.

From the result summary in Fig. 8 along with the baseline literature we can see that both BERT-Large and ensemble techniques perform best in term of prediction accuracy of 93% in Table 1. It evaluates the performance comparison between the proposed method

```
]: from sklearn import metrics
   class_report = metrics.classification_report(test_data.label ,
   confu_matrix = metrics.confusion_matrix(test_data.label, predi
   print(class_report)

                 precision    recall  f1-score   support

              0       0.91      0.95      0.93       831
              1       0.92      0.84      0.88       697
              2       0.96      0.95      0.95      1980
              3       0.84      0.88      0.86       507
              4       0.97      0.96      0.96      1755
              5       0.77      0.85      0.81       230

       accuracy                           0.93      6000
      macro avg       0.89      0.91      0.90      6000
   weighted avg       0.93      0.93      0.93      6000
```

Fig. 6. BERT classification report.

```
              precision    recall  f1-score   support

           0       0.85      0.90      0.88       552
           1       0.81      0.91      0.86       473
           2       0.95      0.91      0.93      1339
           3       0.80      0.86      0.83       336
           4       0.97      0.92      0.94      1152
           5       0.78      0.71      0.74       148

    accuracy                           0.90      4000
   macro avg       0.86      0.87      0.86      4000
 ighted avg       0.91      0.90      0.90      4000
```

Fig. 7. LSTM classification report.

Table 1. SVM, LSTM, BERT-Large ensemble classification report.

S/N	Model	Accuracy
1	SVM	89%
2	LSTM	90%
3	BERT-Large	93%
4	Ensemble model (SVM, LSTM, BERT-L)	93%

and the existing model. Based on the table its identify that the proposed model is more efficient in term of accuracy than the existing model.

This research introduces a multi-social ensemble model based on BERT-Large, SVM and LSTM. In consideration of the experimental result carried out, developed ensemble model was evaluated using standard performance metrics of accuracy and performance efficiency. The SVM model achieved an accuracy of 83%, while LSTM achieved 90% accuracy, BERT-Large achieved 93% accuracy and the utilization of the three model in

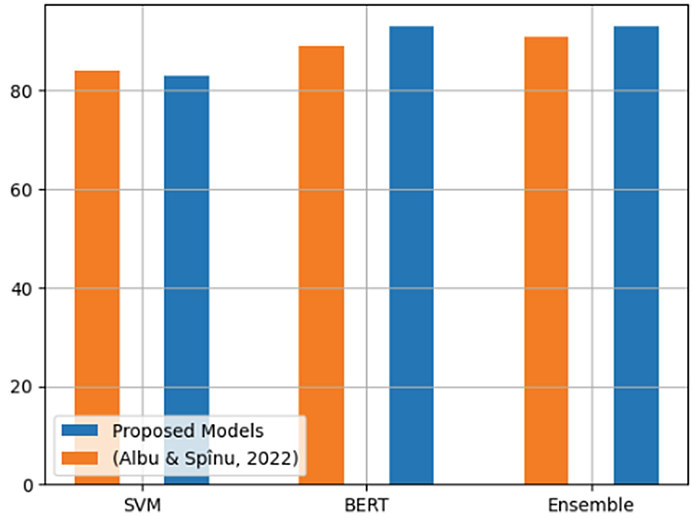

Fig. 8. Performance comparism (existing and proposed model).

an assemble approach also achieve an accuracy of 93%. It identifies that the BERT-Large perform best due to his high richness in contextual information. However, the adoption of ensemble approach in respect to BERT-Large alone does not significantly improve the accuracy. In conclusion the proposed model was able to achieved an optimum accuracy of 93%. This research encourages the use of standalone BERT-Large and Ensemble model to detect tweets.

5 Conclusion

This study has developed a multi-social emotion detection model using BERT and ensemble the BERT-Large, SVM with LSTM model to achieve utmost emotion classification accuracy. The findings from this research study gave us an indication of potency of BERT-Large version of Transformer based Architecture and Ensemble Model in detecting tweets on social media platforms based on emotional analysis. This study has established that the BERT and Ensemble Model are the current state-of-art model.

This study shows that the analysis of user-generated content provides valuable information about the opinion of users for a large variety of topics and products, allowing firms to address typical marketing problems for better customer services. For instance, the evaluation of customer satisfaction or the measurement of the impact of a new marketing campaign on brand perception. Moreover, the analysis of customers' opinions helps business owners to find out possible issues and can possibly suggest new interesting features about a certain product and this can be a driver to attain success.

References

1. Chiorrini A, Diamantini C, Mircoli A, Potena D (2021) Emotion and sentiment analysis of tweets using BERT. CEUR Workshop Proc 2841:17
2. Gnanavel S, Duraimurugan N, Jaeyalakshmi M et al (2021) A live suspicious comments detection using TF-IDF and logistic regression. Ann Rom Soc Cell Biol 25(5):4578–4586
3. Shu K, Mahudeswaran D, Wang S et al (2020) Fakenewsnet: a data repository with news content, social context, and spatiotemporal information for studying fake news on social media. Big Data 8(3):171–188. https://doi.org/10.1089/big.2020.0062
4. Vitiugin F, Barnabo G (2021) Emotion detection for spanish by combining LASER embeddings, topic information, and offense features. CEUR Workshop Proc 2943:8
5. Zhang L, Wang S, Liu B (2018) Deep learning for sentiment analysis: a survey. Wiley Interdisc Rev: Data Min Knowl Discov 8(4):e1253. https://doi.org/10.1002/widm.1253
6. Başarslan MS, Kayaalp F (2021) Sentiment analysis on social media reviews datasets with deep learning approach. Sakarya Univ J Comput Inform Sci 4(1):35–49 (2021). https://doi.org/10.35377/saucis.04.01.833026
7. Singh M, Jakhar AK, Pandey S (2021) Sentiment analysis on the impact of coronavirus in social life using the BERT model. Soc Netw Anal Min 11:33. https://doi.org/10.1007/s13278-021-00737-z
8. Nandwani P, Verma R (2021) A review on sentiment analysis and emotion detection from text. Soc Netw Anal Min 11:81. https://doi.org/10.1007/s13278-021-00776-6
9. Tang D, Zhang Z, He Y et al (2019) Hidden topic-emotion transition model for multi-level social emotion detection. Knowl-Based Syst 164:426–435. https://doi.org/10.1016/j.knosys.2018.11.014
10. Majeed A, Mujtaba H, Beg MO (2020) Emotion detection in roman Urdu text using machine learning. In: Proceedings of the 35th IEEE/ACM international conference on automated software engineering workshops. ACM, pp 125–130. https://doi.org/10.1145/3417113.3423375
11. Kim H, Ben-Othman J, Cho S, Mokdad L (2019) A framework for IoT-enabled virtual emotion detection in advanced smart cities. IEEE Netw 33(5):142–148. https://doi.org/10.1109/MNET.2019.1800275
12. Minaee S, Azimi E, Abdolrashidi A (2019) Deep-sentiment: sentiment analysis using ensemble of CNN and Bi-lstm models. arXiv preprint https://doi.org/10.48550/arXiv.1904.04206
13. Bartneck C, Lyons MJ, Saerbeck M (2017) The relationship between emotion models and artificial intelligence. arXiv preprint https://doi.org/10.48550/arXiv.1706.09554
14. Albu IA, Spînu S (2022) Emotion detection from tweets using a BERT and SVM ensemble model. UPB Sci Bull, Ser C: Electr Eng Comput Sci 84(1):63–74
15. Mosavi A, Salimi M, Faizollahzadeh Ardabili S et al (2019) State of the art of machine learning models in energy systems, a systematic review. Energies 12(7):1301. https://doi.org/10.3390/en12071301
16. Jaiswal S, Nandi GC (2020) Robust real-time emotion detection system using CNN architecture. Neural Comput Appl 32(15):11253–11262. https://doi.org/10.1007/s00521-019-04564-4
17. Ardabili S, Mosavi A, Várkonyi-Kóczy AR (2019) Advances in machine learning modeling reviewing hybrid and ensemble methods. In: Várkonyi-Kóczy A (eds) Engineering for sustainable future. INTER-ACADEMIA 2019. LNNS, vol 101. Springer, Cham, pp 215–227. https://doi.org/10.1007/978-3-030-36841-8_21
18. Salman R, Alzaatreh A, Sulieman H (2022) The stability of different aggregation techniques in ensemble feature selection. J Big Data 9(1):51. https://doi.org/10.1186/s40537-022-00607-1

19. Serban IV, Sankar C, Germain M et al (2018) A deep reinforcement learning chatbot (short version). arXiv preprint https://doi.org/10.48550/arXiv.1801.06700
20. Jaiswal S, Virmani S, Sethi V et al (2019) An intelligent recommendation system using gaze and emotion detection. Multimedia Tools Appl 78:14231–14250. https://doi.org/10.1007/s11042-018-6755-1
21. Yu S, Chen Y, Zaidi H (2020) A financial service chatbot based on deep bidirectional transformers. arXiv preprint https://doi.org/10.48550/arXiv.2003.04987
22. Wolf T, Debut L, Sanh V et al (2020) Transformers: state-of-the-art natural language processing. In: Proceedings of the 2020 conference on empirical methods in natural language processing: system demonstrations. Association for Computational Linguistics, pp 38–45. https://doi.org/10.18653/v1/2020.emnlp-demos.6
23. Derczynski L, Ritter A, Clark S, Bontcheva K (2013) Twitter part-of-speech tagging for all: overcoming sparse and noisy data. In: Proceedings of the 2013 international conference recent advances in natural language processing (RANLP). Association for Computational Linguistics, pp 198–206