International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 05 | May - 2023 SJIF 2023: 8.176 ISSN: 2582-3930

TEXT CLASSIFICATION INTO EMOTIONAL STATES USING BERT TECHNIQUE

[#1]Mr.M.Thirunavukkarasu,
Assistant Professor, Dept of CSE,
SCSVMV(Deemed to be university),
Tamilnadu, India
mthiru@kanchiuniv.ac.in

[#2]P.Sri Vani Praneetha,
Student, Dept of CSE,
SCSVMV(Deemed to be university),
Tamilnadu, India
sushmanaiduthokala@gmail.com

[#3]T.Sushma Naidu,
Student, Dept of CSE,
SCSVMV(Deemed to be university),
Tamilnadu, India
srivp02@gmail.com

ABSTRACT - **Sentiment analysis is now regarded as one of the most crucial fields of study within Natural Language Processing (NLP), as customers, business owners, and other stakeholders can benefit from knowing the implicit or explicit attitudes conveyed in social media contents. The term "aspect-based sentiment analysis" (ABSA) was created to capture aspect attitudes stated towards particular review features because researchers realized that the generic sentiments derived from the textual contents were insufficient.. Put forth a transfer learning-based strategy to address the aforementioned issues with current ABSA techniques. To thoroughly assess the suggested methodologies, we create an experiment that uses data from many fields.**

KEYWORDS: Algorithm, ABSA, BERT, Emotional States, Text, NLP, SMS.

I. INTRODUCTION:

Experts from a variety of disciplines, including natural language processing, computational linguistics, and computational intelligence, have been interested in the classification of opinions, sentiments, and emotional states. Formal and informal writings can both be subjected to computer analysis. Whereas the informal textual material consists of SMS, chat, and social media posts, the formal textual content includes poetry, novels, essays, plays, and official/legal documentation. Text categorization, often known as text classification, is the process of giving free-text documents a predetermined category. It has significant real-world applications and can present conceptual perspectives of document collections. For instance, news articles are frequently categorized by subject areas (topics) or geographic codes; academic papers are frequently categorized by technical areas and subareas; and patient reports in healthcare organizations are frequently indexed from a variety of perspectives using taxonomies of disease categories, types of surgical procedures, insurance reimbursement codes, and so forth. Spam filtering, in which email communications are divided into the two categories of spam and non-spam, is another common use for text categorization.

II. PROBLEM STATEMNT:

The formal text is dense, making it difficult to identify and categorize emotional experiences. For instance, the poetry "Love Philosophy" by Shelley's verse and the sunlight clasps the earth, and the moonbeams kiss the sea: portrays the feeling of love. It takes a lot of time and effort to manually identify the poet's expressed emotional states in poetry. Sentiment analysis is the method of automatically extracting the opinions of writers and characterizing them according to their polarity: positive, negative, or neutral. The aim of emotion analysis, on the other hand, is to identify the emotion that is expressed in the text. Given the wider range of classes and the more nuanced variations between them, this task is typically more challenging than sentiment analysis. Although both lexicon-based and learning-based approaches have been

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 05 | May - 2023 SJIF 2023: 8.176 ISSN: 2582-3930

used to tackle similar challenges in the literature, the latter has performed better in terms of classification. Large deep learning models have been the focus of recent research because of this. Such models need enormous corpora of annotated data, which are typically hard to find and expensive to develop, in order to be accurately trained.

III. LITERATURE SURVEY:

IV. EXISTING METHOD:

With the current method, human beings manually identify the text's emotions by examining its content. This method takes a long time to analyze the content. Disadvantages are Low Accuracy, Time Consuming , High Complexities

V. PROPOSED METHOD:

In our suggested approach, we've developed a BERT-based system that uses deep learning to identify the text's emotional states. The suggested method divides the material into many emotional states, such as neutral, joy, fear, sadness, and fury. The diagram below shows the proposed system's block diagrams. Advantages High Accuracy, Time Saving, Low complexities, High Relatability.

VI. SYSTEM ARCHITECTURE

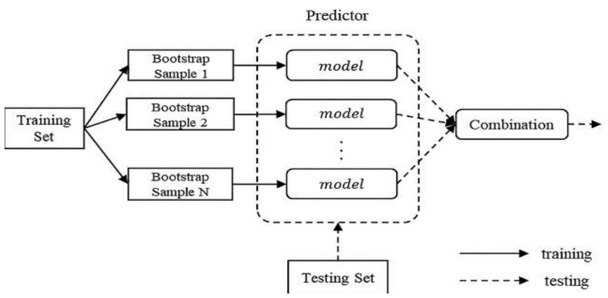


FIG.1.SYSTEM RCHITECTURE

VII. METHODOLOGY:

In the BERT training phase, the model learns to predict whether the second sentence in a pair will come after another in the original document by receiving pairs of sentences as input. The second sentence in 50% of the input pairings during training is the subsequent sentence in the original text, and the second sentence in the remaining 50% is a randomly chosen sentence from the corpus. The underlying presumption is that the second phrase will not be connected to the first.

Before entering the model, the input is processed as follows to aid the model in differentiating between the two sentences during training:

- At the start of the first sentence, a [CLS] token and a [SEP] token are added.
- Each token has been embedded with a sentence that represents either Sentence A or Sentence B. Sentence embeddings and token embeddings both share a concept of two words.
- Each token is given a positional embedding to identify where it belongs in the sequence.

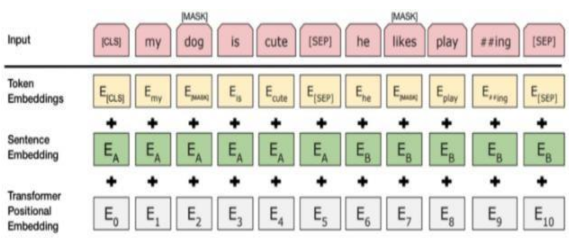


FIG. Next Sentence Prediction (NSP)

The subsequent actions are taken in order to determine whether the second statement is, in fact, related to the first:

- The entire input sequence is processed by the Transformer model.
- The output of the [CLS] token is transformed into a

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 05 | May - 2023 SJIF 2023: 8.176 ISSN: 2582-3930

21-shaped vector using a simple classification layer (learned matrices of weights and biases) .

• Using Softmax to get the Is Next Sequence probability. In order to minimize the combined loss function of the two techniques, Masked LM and Next Sentence Prediction are learned jointly when training the BERT model. It's not too difficult to use BERT for a particular task, BERT merely adds a thin layer to the basic model but can be applied to a wide range of linguistic tasks:

• By adding a classification layer on top of the Transformer output for the [CLS] token, classification tasks like sentiment analysis are carried out similarly to Next Sentence classification.

• In Question Answering tasks (like SQuAD v1.1), the program must indicate the correct response in the text sequence after receiving a question about it. A Q&A model can be trained using BERT by learning two additional vectors that indicate the start and finish of the response.

• In Named Entity Recognition (NER), the software must identify the different sorts of entities in a text sequence

VIII. ALGORITHM

System:

1. **Create Dataset**: The dataset of text is examined which is in a CSV format and divided into training and testing sets.
2. **Pre-processing:** The considered dataset is done word to vector, padding, tokenization, and embedding.
3. **Training**: Train our model with BERT using the pre-processed training.

User:

By providing the text, assess and forecast the feeling.

STEPS FOR EXECUTING THE PROJECTS:

1. Install the necessary applications
2. Creating the unique model.
3. Putting the dataset in use.
4. The dataset is being pre-processed.
5. Using BERT to train the unique model.
6. Conducting text emotion prediction.

IX. RESULTS :

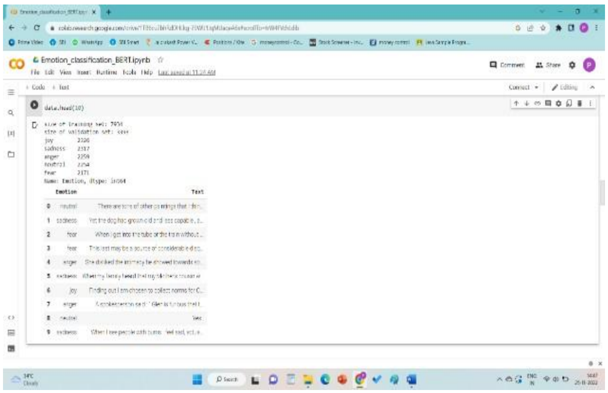
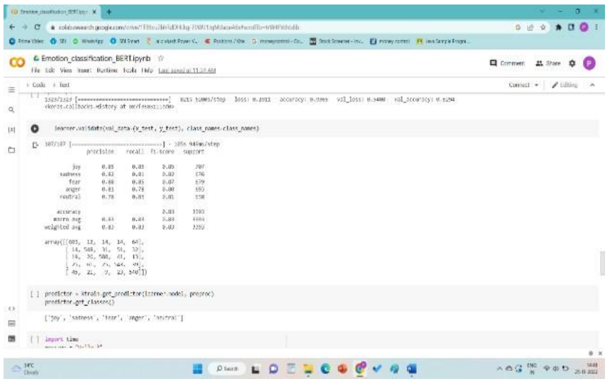


FIG.2.RESULT OF TOP 10 LINES



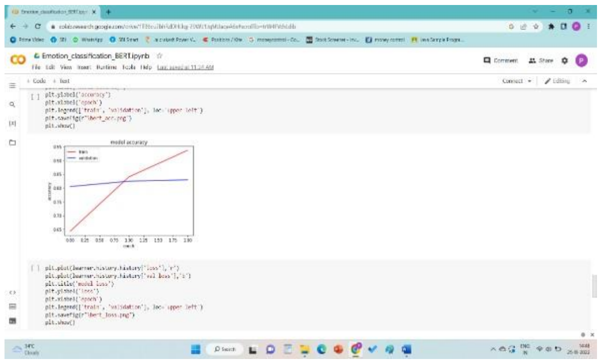


FIG.3.PREDICTION OF EMOTION

ACCURACY TABLE:

Number of characters in each Epoch : 1323

Epoch Number	Accuracy
1	80.49 %
2	82.43%
3	82.94%

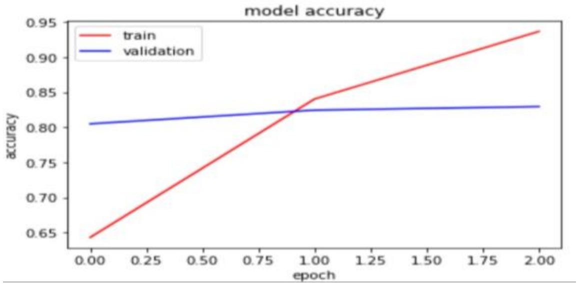


FIG.4.ACCURACY GRAPH

X.CONCLUSION :

In the proposed study, we developed a system that uses deep learning-based BERT to identify different emotional states in text. The suggested method categorizes the text according to various emotional states, including neutral, joy, fear, sadness, and rage. We have successfully developed a system in this application that is used to categorize the emotions in the text that was taken. The BERT algorithm is used to carry out the training process. On the other side, the user can assess and verify their training accuracy outcomes.

REFERENCES

[1] Z. Zhang, Y. Zhang, and T. Zhou, 2018, ‘‘Medical knowledge attention enhanced neural model for named entity recognition in Chinese EMR,’’ in Chinese Computational Linguistics and Natural Language Processing Based on Naturally Annotated Big Data. Cham, Switzerland: Springer, pp. 376–385.

[2] J. Wang, L.-C. Yu, K. R. Lai, and X. Zhang, 2018, ‘‘Dimensional sentiment analysis using a regional CNN-LSTM model,’’ in Proc. 54th Annu. Meeting Assoc. Comput. Linguistics, pp. 225–230.

[3] Q. Zhou and H. Wu, ‘‘NLP at IEST 2019: BiLSTM-attention and LSTM attention via soft voting in emotion classification,’’ in Proc. 9th Workshop Comput. Approaches Subjectivity, Sentiment Social Media Analysis, pp. 189–194.

[4] D. Stojanovski, G. Strezoski, G. Madjarov, I. Dimitrovski, and I. Chorbev, 2021, ‘Deep neural network architecture for sentiment analysis and emotion identification of twitter messages’, Multimedia Tools Appl., vol. 77, no. 24, pp. 32213–32242.

[5] S. Rani and P. Kumar, Apr. 2021, ‘‘Deep learning-based sentiment analysis using convolution neural network,’’ Arabian J. Sci. Eng., vol. 44, no. 4, pp. 3305–3314.