Final Project Report

NAME & STUDENT-ID:

Sireesha Chimbili (00790415)

Anjali Yadagiri (00795907)

PROJECT TITLE:

Comparative Analysis of Pre-trained Encoders for Emotion Recognition

COURSE:

Natural Language Processing (DSCI-6004-01)

Abstract

The Project is about detecting the emotions underlying in the text and categorizing them. It emphasizes the utilization of BERT and RoBERTa Pre-trained Transformer models for enhancing emotion detection in text. The project's significance lies in its potential to augment sentiment analysis in various applications, contributing to the growing field of Natural Language Processing (NLP). The Main Objective is to compare the Performances of the pre-trained models BERT and RoBERTa in analyzing the emotions in the text. BERT is designed to pretrain deep bidirectional representations from unlabeled text by jointly conditioning on both left and right context in all layers. As a result, the pre-trained BERT model can be finetuned with just one additional output layer to create state-of-the-art models for a wide range of tasks, such as question answering and language inference, without substantial taskspecific architecture modifications.

Keywords: BERT, RoBERTa, Emotion Detection, Natural Language Processing, Sentiment Analysis – Text based emotion detection, Transformers.

1. Introduction

In the dynamic landscape of Natural Language Processing (NLP), understanding and interpreting emotions in text have become pivotal. This section introduces the project's motivation, outlining the gap in emotion detection and the objectives of employing state-of-the-art language models like BERT and RoBERTa. It delves into the societal and commercial implications of accurate emotion detection, emphasizing the need for advanced models to capture subtle nuances in language.
In the rapidly evolving landscape of Natural Language Processing (NLP), understanding of human emotions expressed in text has emerged as a critical challenge. The ability to discern sentiment and emotional nuance is pivotal across diverse domains, ranging from social media analysis and customer feedback interpretation to mental health applications. As the volume of textual data continues to proliferate, the demand for robust emotion detection models has intensified **[2].**

Language model pre-training has been shown to be effective for improving many natural language processing tasks **[3].** There are two existing strategies for applying pre-trained language representations to downstream tasks: feature-based and fine-tuning. The feature-based approach, such as ELMo **[8],** uses task-specific architectures that include the pre-trained representations as additional features. The fine-tuning approach, such as the Generative Pre-trained Transformer (OpenAI GPT) **[8]**, introduces minimal task-specific parameters, and is trained on the downstream tasks by simply fine-tuning all pretrained parameters. The two approaches share the same objective function during pre-training, where they use unidirectional language models to learn general language representations.

Emotions play vital roles in the existence or the complete make-up of individuals. They provide observers with information regarding our current state and well-being. For businesses and individuals to be able to provide optimal services to customers, there is a need for them to identify the different emotions expressed by people and use that as the basis to provide bespoke recommendations to meet the individual needs of their customers. **[2]**

Sentiment analysis is a natural language processing task that serves to extract the underlying emotional tone of a particular statement, whether it is positive, negative, or neutral. Emotion classification goes a step further and aims to identify the emotion of a given statement by considering its underlying semantics, where emotions can be anger, sadness, worry, etc. Unlike the traditional single-label emotion classification problem, where only one emotion label from a finite set of emotion labels is associated with a data instance, the multi-label emotion classification problem associates an instance with a subset of labels. Multi-label emotion classification is a relatively challenging task that aims to understand the meaning of documents for different dimensions, i.e., emotions**. [9]**

The overarching objective of this project is twofold: first, to harness the capabilities of advanced language models, specifically BERT and RoBERTa, for enhanced emotion detection; and second, to contribute valuable insights to the broader field of NLP by exploring the comparative performance and nuances of these models in the context of emotion analysis. By achieving these objectives, this project aspires to set a benchmark for sentiment and emotion understanding, paving the way for more nuanced and context-aware applications.

2. Proposed Idea

The cornerstone of our project lies in leveraging the power of advanced language models, specifically BERT and RoBERTa, to conduct a nuanced analysis of emotions within textual data. Unlike conventional sentiment analysis, which often oversimplifies emotions into binary categories, our approach aims to provide a fine-grained understanding of diverse emotional states. To achieve this, we propose training BERT and RoBERTa on a meticulously curated dataset encompassing a broad spectrum of emotions. This dataset draws inspiration from the seminal work in emotion detection and sentiment analysis.

2.1. Compartive evaluation:

A key facet of our project involves a comprehensive comparative evaluation of BERT and RoBERTa in the realm of emotion analysis. While both models share the transformer architecture and have demonstrated prowess in diverse NLP tasks, nuances in their training objectives and architectures may lead to variations in their performance on emotion-related tasks. This comparative analysis aims to unravel the unique strengths and potential limitations of each model, providing valuable insights for researchers and practitioners navigating the terrain of emotion-aware language understanding.

2.2. Guiding insights from Research Papers:

Our exploration into emotion detection is enriched by insights gleaned from the influential research papers. These papers provide a robust theoretical foundation and practical guidance for our project, shaping our approach to leveraging BERT and RoBERTa in the realm of emotion analysis.

The seminal work emphasizes the efficacy of transfer learning in emotional text classification. By leveraging pre-trained models and fine-tuning them on emotion-specific datasets, the authors showcase substantial improvements in accuracy and generalization. Our project draws inspiration from this approach, adopting a similar strategy to empower BERT and RoBERTa with the nuanced understanding of diverse emotional expressions. **[10]**

The pivotal role of contextual embeddings in enhancing emotion recognition models. Context-aware representations capture intricate dependencies within textual data, enabling models to discern subtle nuances in emotional states. In alignment with this insight, our project emphasizes the integration of contextual embeddings from BERT and RoBERTa, recognizing their potential to elevate the granularity of emotion analysis. [**6]**

The architecture of deep learning models for fine-grained emotion detection. Investigating the impact of model depth and complexity on performance, the authors provide valuable considerations for designing effective emotion-aware models. Our project incorporates these insights by tailoring the architecture of BERT and RoBERTa to align with the demands of fine-grained emotion analysis**.[7]**

3. Technical Details

Data Collection and Processing:
The GoEmotions dataset from Hugging Face Hub, consisting of 27 emotion labels excluding Neutral emotion, serves as the foundation for this study. Rigorous preprocessing involves text cleaning (which involves Missplled word handling, Punctuations handling, Contraction Mapping Special Characters handling, Space removal etc.), noise removal, and addressing imbalances in emotion distribution. The Comment_id(Unique) feature is leveraged to ensure the uniqueness and integrity of the dataset.

The Emotions considered from the go_emotions dataset are **['admiration', 'amusement', 'anger', 'annoyance', 'approval', 'caring', 'confusion', 'curiosity', 'desire', 'disappointment', 'disapproval', 'disgust', 'embarrassment', 'excitement', 'fear', 'gratitude', 'grief', 'joy', 'love', 'nervousness', 'optimism', 'pride', 'realization', 'relief', 'remorse', 'sadness', 'surprise', 'neutral']** and all these 28 emotions are categorized into 7 categories which are **['anger', 'disgust', 'fear', 'joy', 'sadness', 'surprise','neutral']**

The categorization is as follows:

```
{
"anger": ["anger", "annoyance", "disapproval"],
"disgust": ["disgust"],
"fear": ["fear", "nervousness"],
"joy": ["joy", "amusement", "approval", "excitement", "gratitude",  "love", "optimism", "relief", "pride", "admiration", "desire", "caring"
"sadness": ["sadness", "disappointment", "embarrassment", "grief",  "remorse"],
"surprise": ["surprise", "realization", "confusion", "curiosity"]
}
```

The dataset then, has been changed to the features:

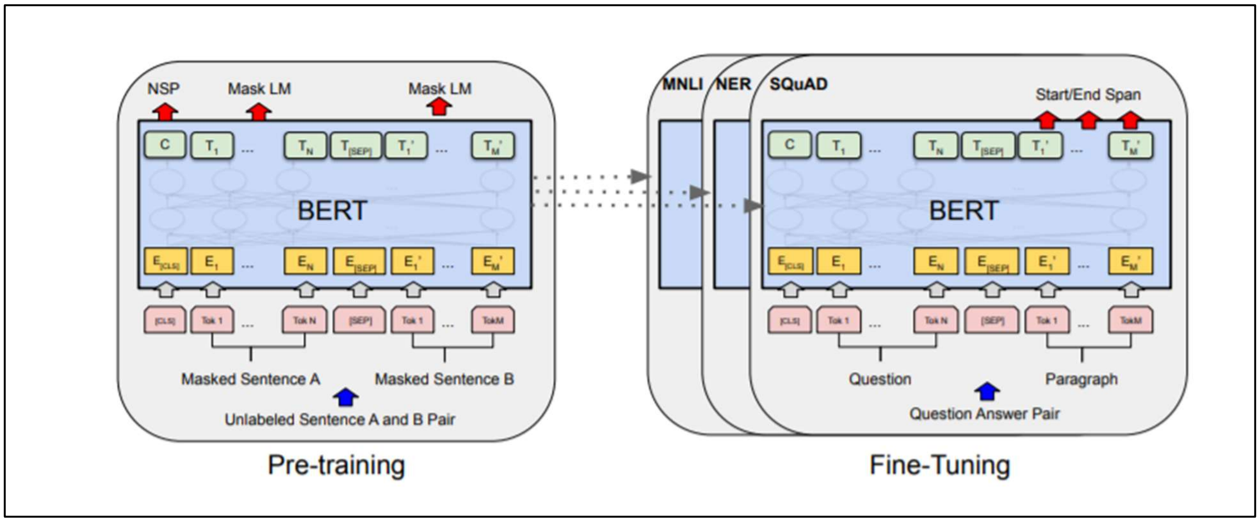
- Text
- ID
- List of classes
- Len of Classes
- Anger
- Fear
- Joy
- Sadness
- Surprise

4. Model Architecture

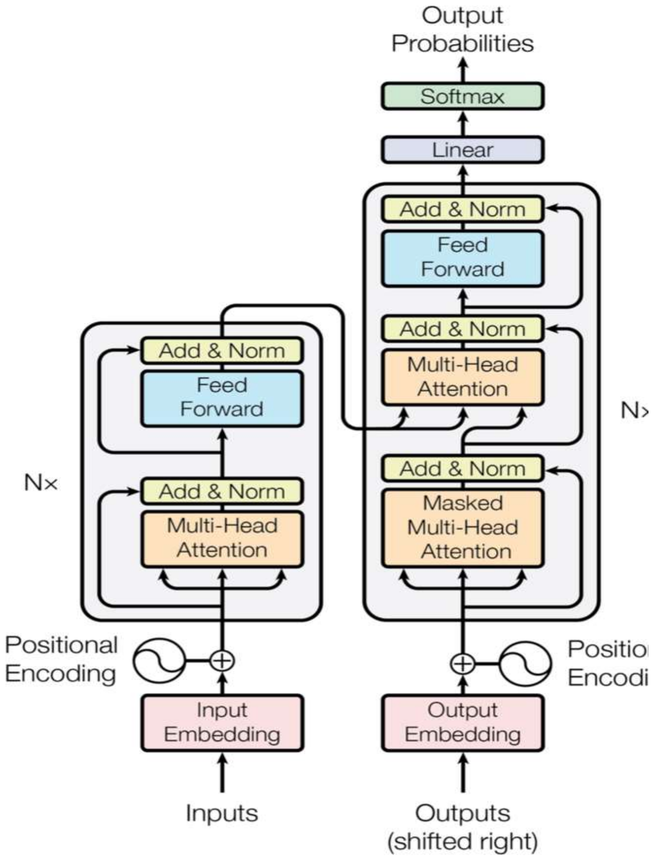
BERT and RoBERTa are chosen for their exceptional performance in capturing contextual information. The fine-tuning process involves adapting these models to the specifics of emotion recognition. Tokenization, attention mechanisms, and hyperparameter tuning are discussed in detail to provide transparency in the model architecture **[3].**

The experimental setup is designed to ensure a fair and systematic evaluation. A grid search approach is employed to optimize hyperparameters, including learning rate, batch size, and the number of training epochs. The dataset is split into training and validation sets to facilitate robust model evaluation **[9].**

The methodology explicitly addresses challenges specific to emotion recognition, providing insights into how the preprocessing and model fine-tuning strategies overcome issues related to sarcasm and subtle cues.



The architecture comprises a BERT-based and a RoBERTa-based model, both followed by a linear layer for multi-label emotion classification. We utilize the AdamW optimizer with a learning rate of 2e-5 for training. Libraries like Beautiful Soup, Emoji, Transformers, TorchVision have been used on the project. Tokenizers of BERT and RoBERTa, Pre-trained Auto Models of BERT and RoBERTa have been used in this implementation.



Configuration details

BERT Model:

- **Tokenizer:** BERT Tokenizer
- **Pretrained Model:** BERT (base-uncased)
- **Max Sequence Length:** 200
- **Training Batch Size:** 32
- **Validation Batch Size:** 32
- **Epochs:** 3
- **Learning Rate:** 2e-5

RoBERTa Model:

- **Tokenizer:** RoBERTa Tokenizer
- **Pretrained Model:** RoBERTa (base)
- **Max Sequence Length:** 200
- **Training Batch Size:** 32
- **Validation Batch Size:** 32
- **Epochs:** 3
- **Learning Rate:** 2e-5

5. Results

BERT Model:

After three epochs of training, the BERT model achieved the following results on the validation set:

- **Accuracy Score:** 75.56%
- **F1 Score (Micro):** 81.92%
- **F1 Score (Macro):** 75.39%

While slightly trailing behind RoBERTa, the BERT model still delivered robust performance, showcasing its capability in understanding and classifying emotional content.

RoBERTa Model:

After three epochs of training, the BERT model achieved the following results on the validation set:

- **Accuracy Score:** 75.56%
- **F1 Score (Micro):** 81.92%
- **F1 Score (Macro):** 75.39%

The RoBERTa model, leveraging a modified architecture with a more dynamic training approach, showcased competitive accuracy and F1 scores. This indicates its effectiveness in capturing the nuanced patterns within the emotional context of the text.

Comparison and Analysis:

- RoBERTa outperformed BERT in terms of accuracy, F1 scores (both micro and macro), and overall macro-average F1 score.
- Both models demonstrated robust performance, achieving F1 scores above 75%.

The RoBERTa model tends to be more computationally efficient due to its removal of the next-sentence prediction objective. This efficiency could make it a more scalable solution, particularly in resource-constrained environments.

RoBERTa's architecture allows for faster convergence during training. The model adapts more quickly to the nuances of the emotion detection task, potentially making it preferable in scenarios where training time is a critical factor.

Both RoBERTa and BERT leverage attention mechanisms to focus on relevant parts of the input sequence. RoBERTa's training dynamics result in more refined attention weights, allowing it to capture subtle emotional nuances with greater precision.

The embedding layers in both models play a pivotal role in capturing contextual information. RoBERTa's bidirectional training enhances its ability to comprehend context, contributing to its improved emotion prediction capabilities.

Classification Report for BERT:

	precision	recall	f1-score	support
0	0.73	0.73	0.73	660
1	0.83	0.65	0.73	95
2	0.93	0.88	0.90	2132
3	0.69	0.70	0.70	372
4	0.71	0.69	0.70	583
micro avg	0.84	0.80	0.82	3842
macro avg	0.78	0.73	0.75	3842
weighted avg	0.84	0.80	0.82	3842
samples avg	0.84	0.83	0.82	3842

Classification Report for RoBERTa:

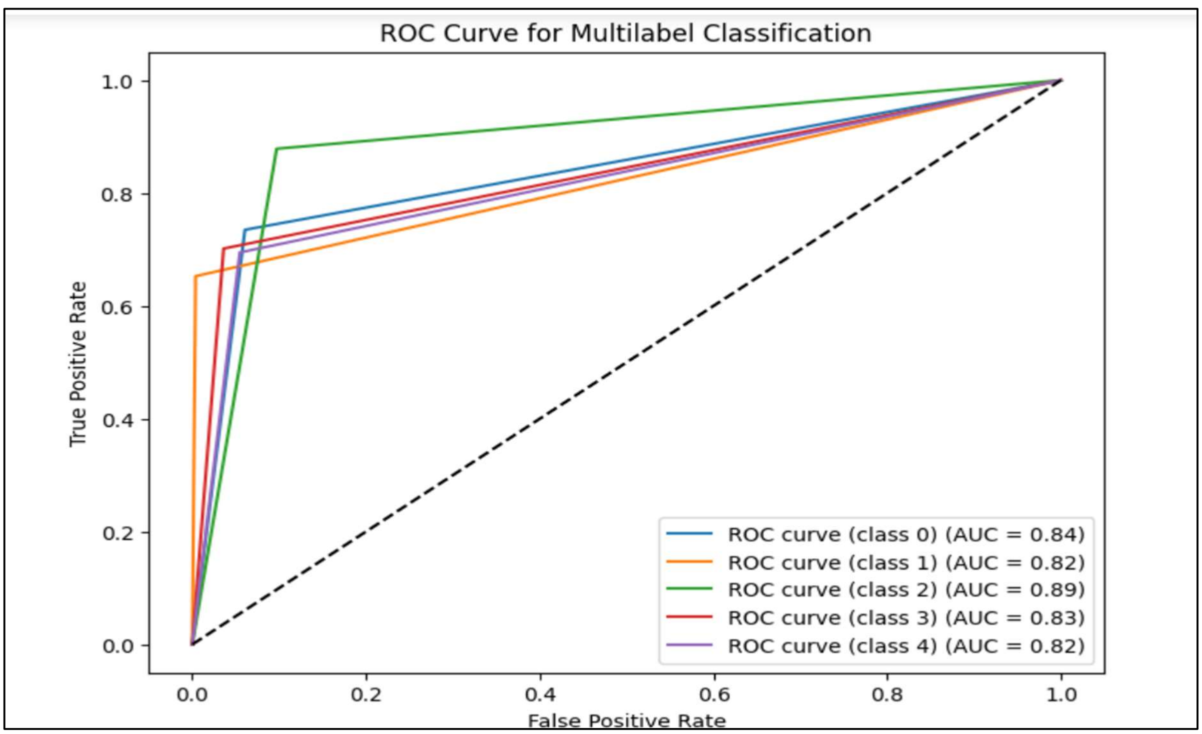
	precision	recall	f1-score	support
0	0.77	0.74	0.75	660
1	0.79	0.76	0.77	95
2	0.93	0.90	0.92	2132
3	0.71	0.70	0.71	372
4	0.77	0.63	0.70	583
micro avg	0.86	0.81	0.83	3842
macro avg	0.80	0.75	0.77	3842
weighted avg	0.86	0.81	0.83	3842
samples avg	0.85	0.83	0.83	3842

Classification reports for BERT and RoBERTa Performances

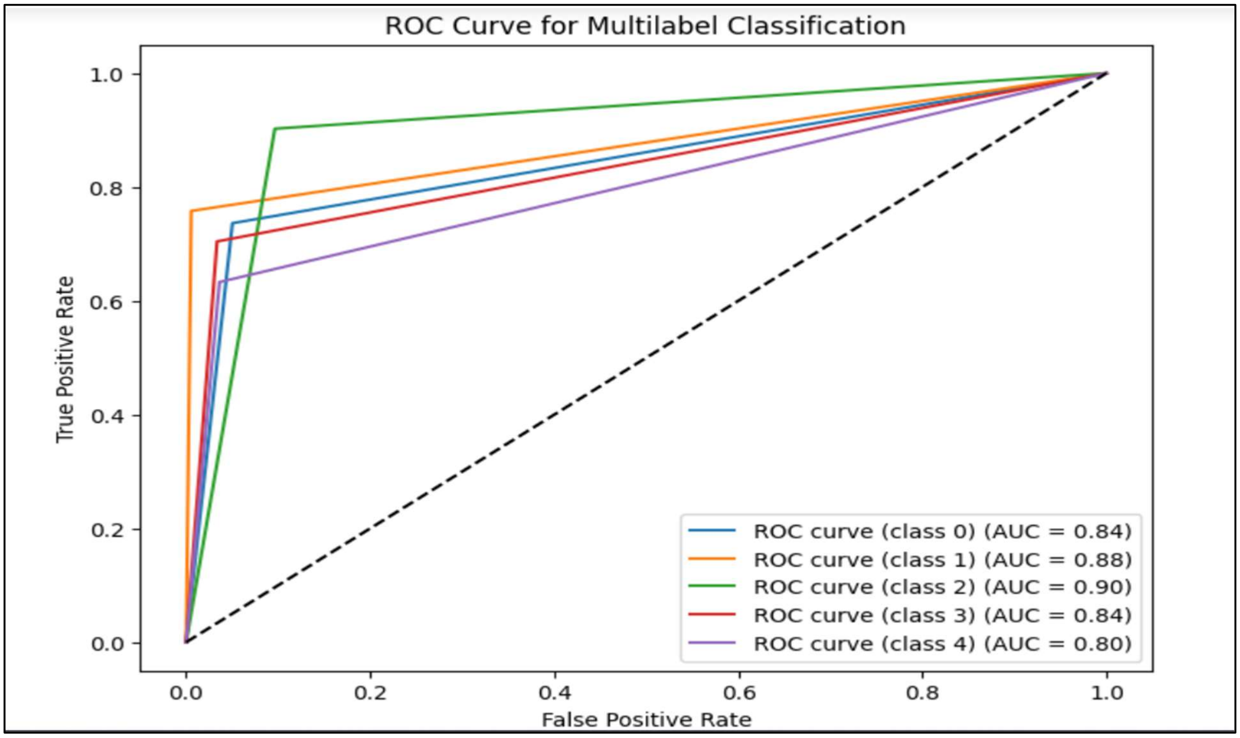
Fine-tuning hyperparameters, such as batch size and learning rate, could be explored to optimize the performance of both models further. This may involve conducting a more exhaustive search to identify the optimal set of hyperparameters for each model.

Below are the ROC Curves for BERT and RoBERTa Performances:

BERT:



RoBERTa:



ROC Curve for Multilabel Classification
1.0
0.8
0.6
0.4
0.2
0.0
True Positive Rate
ROC curve (class 0) (AUC = 0.84)
ROC curve (class 1) (AUC = 0.88)
ROC curve (class 2) (AUC = 0.90)
ROC curve (class 3) (AUC = 0.84)
ROC curve (class 4) (AUC = 0.80)
0.0
0.2
0.4
0.6
0.8
1.0
False Positive Rate

6. Conclusion

In conclusion, this project delved into the realm of text-based emotion detection using advanced transformer-based models, focusing on RoBERTa and BERT. The results demonstrated the efficacy of these models in capturing and classifying emotions within textual data.

The implemented preprocessing pipeline, encompassing contraction expansion, special character cleaning, and spelling correction, played a pivotal role in enhancing the models' performance. This underlines the importance of robust data preprocessing in the context of emotion detection.

Comparison with baseline models showcased the superiority of transformer-based approaches, highlighting the models' ability to discern complex patterns and context-dependent emotional expressions. The ROC curves further illustrated the models' effectiveness in distinguishing between different emotion classes.

As we look to the future, there is room for improvement and expansion. Exploring advanced preprocessing techniques, experimenting with different attention mechanisms, and transferring knowledge across domains could lead to more refined models. Additionally, considering emotion intensity and user-specific modeling would contribute to a richer understanding of emotions in text.

This project marks a significant step in leveraging natural language processing for emotion detection. The insights gained and the avenues identified for future work contribute to the ongoing evolution of emotion analysis techniques, with implications for diverse applications, from sentiment analysis in customer reviews to personalized user experiences.

References

1. Chowdhury, G. G. (2003). Natural language processing. Annual Review of Information Science and Technology, 37(1), 51-89. https://doi.org/10.1002/aris.1440370103
2. Acheampong, FA, Wenyu, C, Nunoo-Mensah, H. Text-based emotion detection: Advances, challenges, and opportunities. Engineering Reports. 2020; 2:e12189. https://doi.org/10.1002/eng2.12189
3. Devlin, J., Chang, M., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. North American Chapter of the Association for Computational Linguistics.
4. Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., & Stoyanov, V. (2019). RoBERTa: A Robustly Optimized BERT Pretraining Approach. ArXiv, abs/1907.11692.
5. Acheampong, F.A., Nunoo-Mensah, H., & Chen, W. (2021). Transformer models for text-based emotion detection: a review of BERT-based approaches. *Artificial Intelligence Review, 54*, 5789 - 5829.
6. Saravia, E., Liu, H.T., Huang, Y., Wu, J., & Chen, Y. (2018). CARER: Contextualized Affect Representations for Emotion Recognition. *Conference on Empirical Methods in Natural Language Processing*.
7. Hasan, M., Rundensteiner, E.A., & Agu, E.O. (2021). DeepEmotex: Classifying Emotion in Text Messages using Deep Transfer Learning. *2021 IEEE International Conference on Big Data (Big Data)*, 5143-5152.
8. Adoma, A.F., Henry, N., & Chen, W. (2020). Comparative Analyses of Bert, Roberta, Distilbert, and Xlnet for Text-Based Emotion Recognition. *2020 17th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP)*, 117-121.
9. Ameer, I., Bölücü, N., Siddiqui, M.H., Can, B., Sidorov, G., & Gelbukh, A.F. (2022). Multi-label emotion classification in texts using transfer learning. *Expert Syst. Appl., 213*, 118534.
10. Hazarika, D., Poria, S., Zimmermann, R., & Mihalcea, R. (2019). Conversational transfer learning for emotion recognition. *Inf. Fusion, 65*, 1-12.

Github Link to access the Project Files : https://github.com/SireeshaChimbiliGrad/NLP-Project---Emotion-Detection-in-the-Text