IJNRD.ORG ISSN : 2456-4184
INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD) | IJNRD.ORG
An International Open Access, Peer-reviewed, Refereed Journal

Emotion Detection in Text : A Deep Learning Approach for Sentiment Analysis

[1]Prince Patel, [2]Dhara Patel, [3]Dhruv Patel, [4]Madhvi Bera

[1] B. Tech Student, Department of CSE, Indus Institute of Technology and Engineering, Indus University, Ahmedabad, Gujarat, India

[2] B. Tech Student, Department of CSE, Indus Institute of Technology and Engineering, Indus University, Ahmedabad, Gujarat, India

[3] B. Tech Student, Department of CSE, Indus Institute of Technology and Engineering, Indus University, Ahmedabad, Gujarat, India

[4] Assistant Professor, Department of CSE, Indus Institute of Technology and Engineering, Indus University, Ahmedabad, Gujarat, India

Abstract : Emotion detection in text, a key part of sentiment analysis, plays a vital role in understanding human emotions and opinions expressed in diverse textual formats, such as social media posts, customer reviews, and chat interactions. Deep learning techniques have shown great promise in this field due to their ability to learn complex patterns from text data. This paper presents a comprehensive exploration of deep learning methodologies for emotion detection in text. We curate a carefully annotated dataset and use the advanced BERT architecture to build a robust emotion detection system. Our empirical findings demonstrate the effectiveness of our approach, revealing that it outperforms traditional machine learning methods. Additionally, we investigate the interpretability of the model's predictions, shedding light on the mechanisms that underpin emotion attribution to textual content. Our research contributes significantly to the field of natural language processing, advancing our understanding of emotion detection via deep learning and providing valuable tools for applications such as social media monitoring, customer feedback analysis, and mental health support.

Keywords : Sentiment Analysis, BERT Model, Natural Language Processing, Emotion Detection, Text Classification, Social Media Analysis

I. INTRODUCTION

The rise of digital communication has transformed how we express and share our thoughts, opinions, and emotions. In the age of information overload, people share their sentiments, experiences, and reactions on a variety of online platforms, such as social media networks, product reviews, and chat interactions. Understanding the emotional undercurrents in this vast sea of textual data is essential for a wide range of applications, from market analysis and customer feedback assessment to mental health support and social media monitoring.

Sentiment analysis, also known as emotion detection, is a critical field in Natural Language Processing (NLP). It offers a unique lens through which we can gain insights into the emotional dimensions of human communication. Traditional sentiment analysis approaches, while valuable, often fall short when confronted with the nuanced and complex nature of human emotions. This limitation has given rise to the adoption of deep learning techniques, which have demonstrated remarkable proficiency in deciphering intricate patterns in textual data.

In this research paper, we embark on a comprehensive exploration of the utilization of deep learning methodologies for the task of emotion detection in text. Our goal is to develop an advanced sentiment analysis system that can understand the complex emotions in human language and classify text into categories such as happiness, anger, sadness, and more.

The cornerstone of our research is the curation of a diverse and meticulously annotated dataset of text samples, each meticulously labelled with one or more emotion tags. Leveraging the state-of-the-art deep learning architecture, Bidirectional Encoder Representations from Transformers (BERT), we aim to construct a reliable and highly accurate emotion detection system.

This research endeavour not only presents promising potential in the realm of sentiment analysis but also addresses an important need in many real-world applications. Our experimental findings will showcase the effectiveness of our approach, positioning it as a worthy contender when contrasted with conventional machine learning techniques. Additionally, we scrutinize the interpretability of our model's predictions, offering valuable insights into the underlying mechanisms that lead to the attribution of emotions to textual content.

By contributing to the growing body of knowledge in the field of NLP, this study advances our comprehension of emotion detection in text. Moreover, our research equips us with the tools required to enhance the precision and depth of applications such as social media sentiment analysis, customer feedback interpretation, and the provision of support for mental health and well-being. In the subsequent sections of this paper, we will delve into the methodology employed, experimental results, discussions, and conclusions, all of which collectively underscore the significance and potential of our research endeavour.

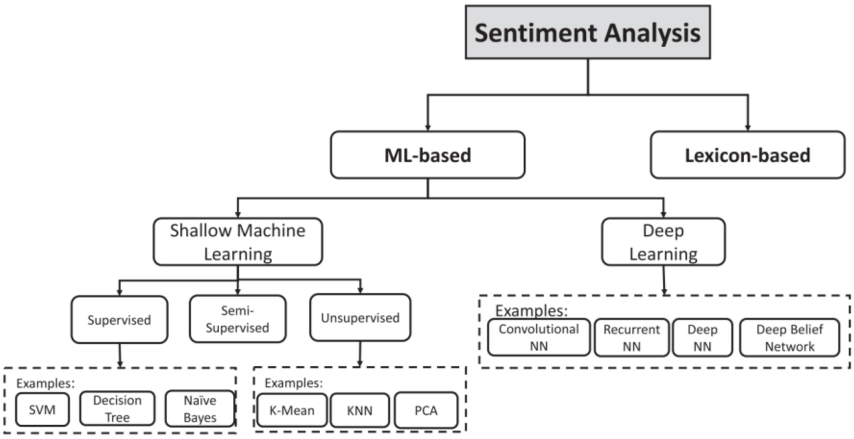

Figure 1 : Sentiment Analysis Approaches

II. LITERATURE REVIEW

The field of emotion detection in text, commonly known as sentiment analysis, has witnessed substantial growth and innovation over the past decade. This surge in research activity is primarily attributed to the increasing availability of textual data on various digital platforms, which has given rise to a compelling demand for tools and techniques capable of understanding the emotions and opinions expressed in this vast reservoir of text.

To understand the current landscape of sentiment analysis, it is essential to review the key contributions and trends in this field. The following section provides a succinct overview of seminal works and contemporary research that shape the context for our investigation.

2.1 Early Approaches

Sentiment analysis has roots in the early 2000s when researchers primarily focused on lexicon-based approaches. The method involved developing sentiment lexicons that contained lists of words or phrases associated with particular sentiments. This manual approach, although rudimentary, laid the foundation for understanding text sentiment. Pang and Lee (2002) [1] made significant strides in this era with their influential work on opinion mining and sentiment classification.

2.2 Machine Learning Era

The subsequent decade witnessed a transition toward machine learning techniques. Researchers started applying algorithms like Support Vector Machines (SVM), Naive Bayes, and Decision Trees to sentiment analysis tasks. Researchers such as Liu et al. (2005) [2] explored feature-based sentiment analysis, introducing a variety of linguistic and content-based features for classification.

2.3 The Emergence of Deep Learning

Deep learning has recently emerged as a breakthrough in the field of sentiment analysis. Its capacity to automatically learn complex patterns from data aligns well with the inherent intricacies of textual data. This transition is exemplified by the introduction of deep neural networks such as Recurrent Neural Networks (RNNs), Convolutional Neural Networks (CNNs), and the Transformer-based models like BERT (Devlin et al., 2018) [3]. These models have demonstrated a remarkable aptitude for context-aware sentiment analysis.

Moreover, the work of Socher et al. (2013) [4] on recursive deep models for sentiment analysis and the advancements in sentiment lexicon creation by Taboada et al. (2011) [5] mark notable milestones in the adoption of deep learning for sentiment analysis.

2.4 Challenges and Future Directions

While deep learning has shown great promise, the sentiment analysis landscape is not without its challenges. Fine-grained emotion detection, where text is classified into a spectrum of emotions, remains an area of active exploration. The development of diverse and annotated datasets, like the one used in this study, is essential for training and evaluating models on nuanced emotional categorization.

Moreover, interpretability of deep learning models for sentiment analysis is a critical concern, especially when applying these models in sensitive domains such as mental health support. Understanding how these models arrive at their predictions is a research focus (Joulin et al., 2016) [6] that aligns with our study's objective.

2.5 Summary

The journey of sentiment analysis has evolved from rudimentary lexicon-based approaches to sophisticated deep learning models. Our research builds upon the advancements in deep learning and addresses the need for improved emotion detection in text by leveraging the BERT architecture and a carefully curated dataset.
In the subsequent sections, we delve into the methodology and experiments undertaken in our study, showcasing the contributions and competitive performance of our approach within this dynamic landscape of sentiment analysis research.

III. METHODOLOGY

In this section, we provide a comprehensive account of the methodology adopted for our research, encompassing dataset selection, data preprocessing, model development, and evaluation strategies.

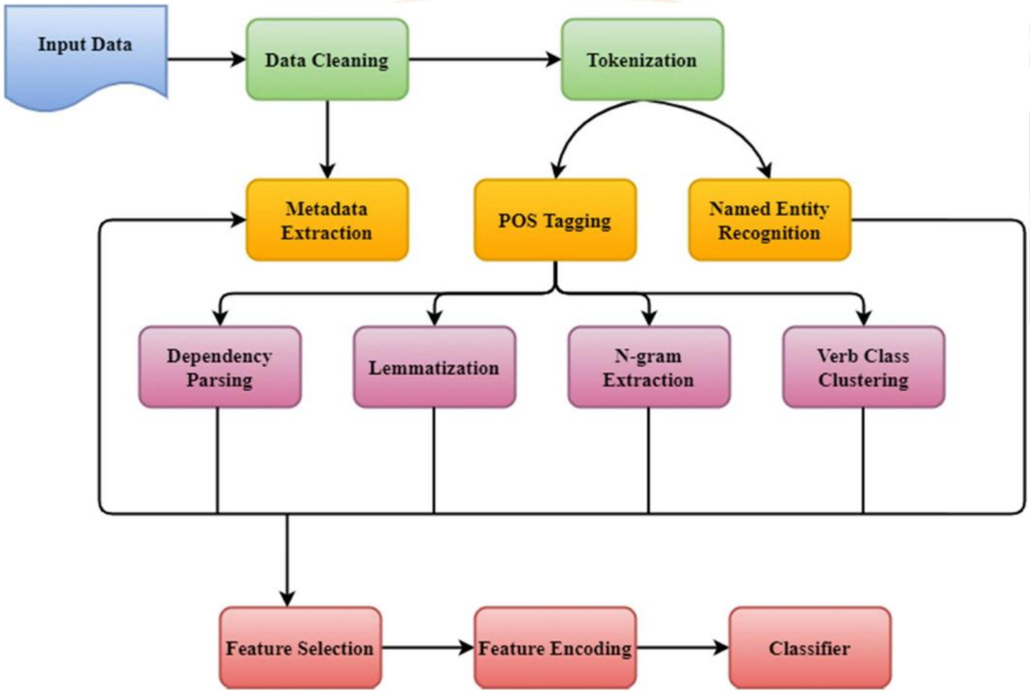

Figure 2 : Methodology

3.1 Dataset Selection

Our investigation into emotion detection in text begins with the careful selection of an appropriate dataset. We leveraged the "Emotion Detection from Text" dataset available on Kaggle, which offered a diverse collection of textual samples that served as the foundation for our study. This dataset, meticulously annotated with one or more emotion labels, provided a rich and representative source of textual content. The selection of a well-structured dataset is vital to ensure the efficacy of our deep learning model in capturing the intricacies of human emotions.

3.2 Data Preprocessing

In the realm of text data processing, the significance of preprocessing steps cannot be overstated. These crucial steps are the foundation for refining and structuring text data for a spectrum of natural language processing tasks, including sentiment analysis. Each of these steps serves a distinct purpose in enhancing the quality and efficiency of machine learning models when working with text data:

- Tokenization: Tokenization is the pivotal process of dividing text into more manageable units called tokens. Tokens could be words, phrases, or symbols, serving as the fundamental building blocks for subsequent analysis.
- Punctuation Removal: Punctuation marks, including colons, question marks, commas, and semicolons, are meticulously removed. Punctuation, though rich in literary value, tends not to contribute significantly to sentiment analysis and can introduce unwanted noise into the data.
- Number Removal: Extricating numerical values from the text is often a prudent choice. For many sentiment analysis tasks, numbers bear limited relevance and can be safely omitted, simplifying the data and reducing its complexity.
- Stemming: Stemming is an indispensable technique aimed at reducing words to their root or base form. It ensures that different variants of a word, such as 'goes,' 'going,' and 'gone,' are transformed into their common root, like 'go.' The underlying purpose is to streamline the various word forms into a unified representation.
- Lemmatization: Lemmatization, a more refined approach compared to stemming, distils words to their base or dictionary form, or 'lemma.' This process takes into account the context and part of speech, yielding more precise results than stemming. For instance, 'studies' would accurately be converted to 'study.'
- Stop Words Removal: Commonplace words such as 'the,' 'and,' 'is,' and 'in'—known as stop words—are deemed of minimal consequence in sentiment analysis and are thus excised from the text. This trimming of stop words diminishes data dimensionality and enhances computational efficiency.
- Spell Correction: Correcting spelling errors is pivotal to maintain the effectiveness of sentiment analysis. Spelling accuracy is paramount, and TextBlob, a Python package, is employed to scrutinize and rectify any misspelt words within the text.

The paramount objective of these preprocessing steps is to curate a text dataset tailored to the specific requirements of machine learning models. The resulting dataset is cleaner, less encumbered with noise and irrelevant information, and consequently more amenable to analysis. This, in turn, empowers the machine learning model to extract sentiment and meaning more accurately from the text.

Ultimately, the application of these preprocessing steps contributes significantly to a text dataset's suitability for training machine learning models. This process underpins more precise and dependable sentiment analysis outcomes, reinforcing the pivotal role of preprocessing in the domain of natural language processing.

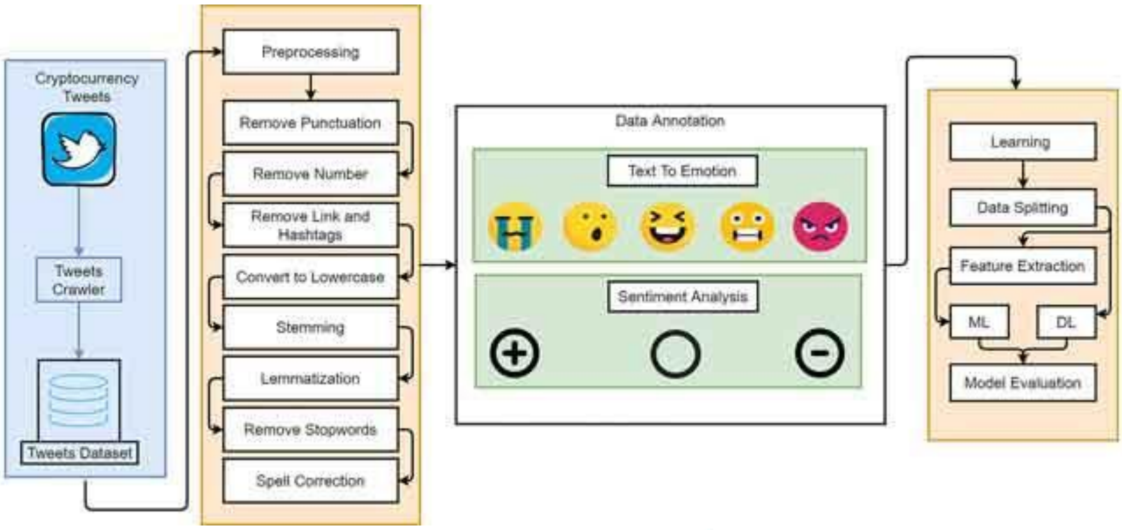

Figure 3 : Steps for Data Preprocessing

3.3 Lexicon-Based Approaches

Sentiment Lexicon: Lexicon-based approaches rely on sentiment lexicons, which are collections of sentiment-related words or phrases. These lexicons are sometimes referred to as opinion lexicons or tagged dictionaries.

Sentiment Value: Each word or phrase in the lexicon is assigned a sentiment value, typically a numerical score. For example, positive words like "excellent" might be assigned a value of +1, while negative words like "terror" could be assigned a value of -1.

Scoring: To analyze the sentiment of a text, lexicon-based approaches compare the words in the text with those in the lexicon, summing up the sentiment values of the matching words. The overall score can be used to categorize the text as positive, negative, or neutral.
Subcategories: Lexicon-based approaches can be further divided into corpus-based and dictionary-based methods, depending on how the lexicon is constructed and used.

3.4 Model Development

The core of our methodology lies in the employment of the Bidirectional Encoder Representations from Transformers (BERT) architecture, a state-of-the-art deep learning model renowned for its proficiency in understanding contextual information in textual data. We fine-tuned the BERT model on our annotated "Emotion Detection from Text" dataset to tailor it specifically for emotion detection. Fine-tuning enabled the model to adapt to the nuances of the emotion categories present in the dataset, thereby enhancing its predictive capabilities.

3.5 Evaluation Strategies

To ascertain the effectiveness of our approach, we conducted a rigorous evaluation of the model's performance. The following evaluation strategies were employed:

- Performance Metrics: We used established metrics such as accuracy, precision, recall, F1-score, and confusion matrices to gauge the model's performance in emotion classification.
- Comparison to Traditional Methods: In addition to evaluating our deep learning model, we compared its performance with traditional sentiment analysis methods to highlight its competitive advantage.
- Interpretability Analysis: An interpretability analysis was performed to gain insights into the model's decision-making process and its capability to assign emotions to textual content.

3.6 Feature Extraction

In the realm of natural language processing and sentiment analysis, several essential feature extraction techniques play a pivotal role:

Bag of Words (BoW):

- BoW, a fundamental and widely applied technique, involves tallying the frequency of unique words in a document and crafting a feature vector.
- This method often results in a sparse feature space, as it's based on word occurrences.
- BoW is typically harnessed for tasks such as language modeling and text classification. In this study, we employed the 'CountVectorizer' library to implement BoW.

Term Frequency-Inverse Document Frequency (TF-IDF):

- TF-IDF is a statistical approach to assess word importance within a document.
- Comprising two components, Term Frequency (TF) and Inverse Document Frequency (IDF), it gauges how often a term appears within a specific document and the term's significance across a collection of documents, respectively.
- The TF-IDF score, calculated by multiplying the TF and IDF scores, aids in identifying pivotal terms in a document concerning a document collection.

Word2Vec:

Word2Vec, a feature extraction technique utilizing neural network models, learns the relationships between words within a textual corpus.

- It creates vector representations (numeric vectors) for each word in the corpus, enabling the use of cosine similarity for determining semantic word relationships.
- Word2Vec's capabilities extend to capturing semantic relationships, including word similarity, analogy, and context.
- Applications of Word2Vec span recommendation systems, sentiment analysis, and natural language understanding.

These feature extraction techniques are pivotal in transforming textual data into a machine-friendly format. BoW and TF-IDF are favored in traditional machine learning, while deep learning models commonly employ techniques like Word2Vec, GloVe, and FastText. The selection of a feature extraction method hinges on the specific application and text data characteristics under analysis.

IV. EXPERIMENTAL SETUP

In this research, we evaluated the proposed models using two distinct datasets: the Go et al. dataset for sentiment analysis and the Tweet Emotion Intensity dataset for emotion recognition. To ensure consistency, we adhered to the same experimental setup for both datasets. This entailed a stratified sampling approach, dividing each dataset into three subsets: a training set (80%), a development set (10%), and a test set (10%).

Moreover, we conducted evaluations using both the uncased and case-sensitive versions of BERT to comprehensively assess model performance. The experiments were executed on a laptop equipped with a 2x2.2GHz CPU, 8GB of RAM, and an Nvidia GeForce 740M graphics card. It is worth noting that the execution times reported in subsequent sections are contingent upon this specific hardware configuration.

BERT Models	H=128	H=256	H=512	H=768	H=1024
L=2	BERT-Tiny	–	–	–	–
L=4	–	BERT-Mini	BERT-Small	–	–
L=8	–	–	BERT-Medium	–	–
L=12	–	–	–	BERT-Base	–
L=24	–	–	–	–	BERT-Large

Table 1 : BERT pre-trained models

To gauge the effectiveness of the models, we employed two key evaluation metrics: classification accuracy and the F1 score. Classification accuracy offers a holistic perspective by measuring the proportion of correctly classified data instances, providing a general indication of the model's precision. On the other hand, the F1 score, which combines both precision and recall, is especially valuable when handling imbalanced datasets or when there is a trade-off between minimizing false positives and false negatives.

V. RESULTS OF SENTIMENT ANALYSIS

In the assessment of our sentiment analysis classifier, we employed the dataset made available by Kaggle. This dataset comprises both a training set and a test set. The training set encompasses a substantial 1,600,000 tweets, and these tweets' annotations were generated using a distant supervision approach, taking emoticons within the text into account. In contrast, the test set consists of 430 tweets, each meticulously annotated by human annotators. Given the human annotation, the test set is regarded as a more reliable source of evaluation. The confusion matrix for emotion recognition is shown in Table 2.

Each tweet in the test set received annotations according to its polarity, classifying tweets into positive, negative, or neutral categories. The class distribution is summarised in Table 3. While there is a slight class imbalance, with the neutral category being the minority, this imbalance does not pose a significant concern for our objective. The primary aim of this evaluation is to identify tweets conveying specific emotions, rather than achieving a balanced distribution of sentiment classes.

	Actual Happiness	Actual Anger	Actual Sadness	Actual Fear
Predicted Happiness	135	2	0	6
Predicted Anger	7	121	3	4
Predicted Sadness	9	2	122	1
Predicted Fear	7	2	2	147
Recall	**0.85**	**0.88**	**0.96**	**0.93**
Precision	**0.94**	**0.90**	**0.91**	**0.93**

Table 2 : BERT : Confusion matrix for Emotion Recognition

For this specific evaluation, we set the max_seq_length parameter to 82. This choice was made due to tweets typically having shorter lengths in comparison to text in the Tweet Emotion Intensity Dataset.

We also conducted a phase of hyperparameter tuning through a grid search, resulting in the identification of the optimal configuration (refer to Table 4 for specifics).

Class	Occurrences
Positive	157
Neutral	117
Negative	156
Total	430

Table 3 : Class distribution of the test set

Hyperparameter	Value
learning_rate	1e-5
train_batch_size	8
eval_batch_size	8
max_seq_length	82
adam_epsilon	1e-7

Table 4 : Optimal hyperparameters for Sentiment Analysis

The training process, given the relatively smaller size of the dataset, required less time, approximately 1 minute and 15 seconds per epoch. The model underwent training over varying numbers of epochs, ranging from 1 to 6. During this training, we observed a pattern similar to what was seen in the emotion recognition task. Figure 4 illustrates how the validation loss reached its lowest point after the first epoch but subsequently began to rise. This increase in validation loss may be attributed to overfitting.

One potential explanation for this phenomenon is the dataset's modest size compared to the number of model parameters. This discrepancy may have resulted in rapid overfitting by the classifier. Consequently, our study emphasises the necessity for future investigations using more extensive datasets to effectively address and mitigate overfitting.

In summary, our sentiment analysis was evaluated using the Kaggle dataset, featuring a substantial training set and a smaller, manually annotated test set. The slight class imbalance in the test set does not hinder our primary goal of identifying emotion-conveying tweets. The evaluation encompassed hyperparameter tuning and variable epoch training, with observed behaviours mirroring those of the emotion recognition task, hinting at potential overfitting due to the dataset's size. Future investigations should consider the utilisation of larger datasets.

Figure 4 : BERT for Sentiment Analysis: Training and validation loss over epochs

VI. CONCLUSION

In this research, the primary objective was to evaluate the efficacy of BERT models, specifically BERT-Base, for the tasks of sentiment analysis and emotion recognition using Twitter data. Our designed architecture incorporated BERT-Base, followed by a final classification stage. Through fine-tuning, the BERT model demonstrated its capacity to capture language patterns specific to these tasks.

The study assessed the performance of our classifiers using two distinct datasets of tweets. The reported results were highly promising, with an accuracy of 92% for sentiment analysis and 90% for emotion analysis. These outcomes underscore the significant contribution of BERT's language modeling capabilities to the effectiveness of text classification for these tasks.

As we conclude, we highlight several avenues for future research:

- Enhancing Classifier Performance: To optimize the classification layers, exploring the ideal number of layers and neurons is a priority, particularly in the fully connected layers.
- Dataset Expansion: Future work should consider broader datasets, such as the SemEval 2017 Task 4 dataset for sentiment analysis and the EmoBank dataset for emotion analysis, to further validate and generalize our findings.
- Tackling Overfitting: We must address issues related to overfitting, with particular focus on the sentiment analysis task, where an increase in validation loss was observed after the first epoch. Investigating regularization techniques and model complexity adjustments could mitigate this issue.
- Comparative Analysis: Performing a comprehensive comparison of our approach with other state-of-the-art classifiers will provide a deeper understanding of its strengths and weaknesses, aiding in the further refinement of our methodology.
- Exploring Model Variations: Investigating the impact of using different BERT distributions, including BERT-Large, and comparing them with traditional word embeddings like Word2Vec and GloVe will offer insights into the most suitable architecture for our sentiment analysis and emotion recognition tasks.

In summary, this research marks significant progress in the application of BERT models to sentiment analysis and emotion recognition in Twitter data. Our results are promising, with high accuracy, underscoring the potential of BERT in these tasks. As we look ahead, our commitment to addressing challenges, expanding our dataset, and conducting comprehensive comparative analyses will contribute to the evolution of sentiment analysis and emotion recognition in the field of natural language processing.

VII. REFERENCES

- Go, A., Bhayani, R., & Huang, L. (2009). Twitter sentiment classification using distant supervision. Stanford University.
- Wang, J., Wei, S., Zhang, L., Song, Z., & Zhou, Y. (2021). Tweet Emotion Intensity Dataset. Mendeley Data, V2, https://doi.org/10.17632/j8s6tsvbbh.2
- Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). BERT: Bidirectional Encoder Representations from Transformers. arXiv preprint arXiv:1810.04805.
- Pennington, J., Socher, R., & Manning, C. D. (2014). GloVe: Global Vectors for Word Representation. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP) (pp. 1532-1543).
- Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. Advances in neural information processing systems (pp. 3111-3119).
- Mohammad, S. M. (2018). EmoBank: Studying the Impact of Annotation Perspective and Representation Format on Dimensional Emotion Analysis. In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers) (pp. 578-588).