International Journal of Computing and Digital Systems
ISSN (2210-142X)
Int. J. Com. Dig. Sys. 15, No.1 (Jan-24)

http://dx.doi.org/10.12785/ijcds/150105

A Fusion Architecture of BERT and RoBERTa for Enhanced Performance of Sentiment Analysis of Social Media Platforms

B.V Pranay Kumar[1] and Dr. Manchala Sadanandam[2]

[1]*Department of Computer Science and Engineering Kakatiya University Warangal Telangana India*
[1]*Department of Computer Science and Engineering Christu Jyothi Institute of Technology and Science Jangaon Telangana India*
[2]*Department of Computer Science and Engineering Kakatiya University Warangal Telangana India*

Received 17 Apr. 2023, Revised 16 Nov. 2023, Accepted 23 Dec. 2023, Published 1 Jan. 2024

Abstract: Natural language processing's subfield of sentiment analysis involves locating and categorizing the feelings, viewpoints, and attitudes expressed in text. Because it enables us to understand public opinion on a variety of topics, sentiment analysis has grown in importance as social media platforms become more widely used. In this research paper, we used two deep learning models, BERT and RoBERTa, and their fusion of both architectures to perform sentiment analysis on a dataset of tweets related to the dataset of COVID-19 pandemic. To eliminate noise and unrelated data, the dataset underwent pre-processing and cleaning. Then, using the dataset, we trained the BERT and RoBERTa models and assessed their performance. Both models achieved high F1 scores, recall, and accuracy for all three sentiment classes (negative, neutral, and positive) for sentiment analysis. While there were some differences in how well these models performed across these metrics, both models did well and classified the sentiment of tweets in the dataset with high accuracy. Our study's findings show how well BERT and RoBERTa perform sentiment analysis on tweets about the COVID-19 pandemic. Our study also emphasizes how crucial it is to clean up and pre-process the dataset to get rid of extraneous data and noise that can harm the models' performance. The effectiveness of these models on datasets from other domains and topics will be examined. Future studies should also look into the models' interpretability and comprehend the features and patterns crucial to sentiment analysis. This paper emphasizes how we can avoid disaster tweets and be cautious to identify hate speech that disturbs the harmony in society.

Keywords: Sentiment Analysis, unsupervised learning, deep learning, BERT, RoBERTa, Performance parameters, Fusion of Bert and RoBERTa, hate speech

1. INTRODUCTION

Sentiment analysis is a branch of NLP that deals with locating and categorizing the feelings, viewpoints, and attitudes expressed in written works. As social media platforms rise in popularity, sentiment analysis has grown in importance as a research field because it enables us to understand how the public feels about a variety of issues.

Twitter, in particular, has emerged as a popular platform for sentiment analysis due to its massive user base and the availability of real-time data. Twitter data can provide valuable insights into public opinion on various topics, including politics, sports, entertainment, and public health. Analysis of the sentiment of tweets can be difficult due to the nature of the language used, the use of slang and abbreviations, and the limited length of the messages.

Further opportunities for study and development exist with the use of BERT [1] and RoBERTa [2] models for sentiment analysis on the Coronavirus Tweets NLP - Text Classification dataset and associated natural language processing tasks. Future work may encompass the application of these models to similar tasks like topic classification, entity recognition, and event extraction as well as the investigation of new model architectures, training settings, and assessment metrics. As a result, utilizing the Coronavirus Tweets NLP - Text Classification dataset and the BERT and RoBERTa models for sentiment analysis, a powerful approach for analyzing sentiment on the COVID-19 pandemic on social media has been established. With greater research and refinement, these models might provide informative data and analysis for public health, policy, and decision-making.A transformer-based model called BERT is pre-trained on a vast corpus of text data before being honed for particular applications like sentiment analysis. On several natural language processing tasks, such as sentiment analysis, question answering, and language translation, BERT has attained cutting-edge performance. Although RoBERTa's

E-mail address: pranaybv4u@gmail.com, sadanb4u@gmail.com

architecture is similar to that of BERT, it employs a different pre-training strategy that calls for longer training sessions and bigger batch sizes. On several natural language processing tasks, such as sentiment analysis, RoBERTa has outperformed BERT.With encouraging outcomes, BERT and RoBERTa have both been used in the sentiment analysis of tweets. For sentiment analysis across the three sentiment classes (negative, neutral, and positive), these models have been demonstrated to produce good accuracy, precision, recall, and F1-score. Researchers have been able to learn more about public opinion on a variety of subjects relating to Twitter data, such as politics, public health, and social concerns. The ability of transformer-based models [3] like BERT and RoBERTa to capture context and semantics makes them advantageous for sentiment analysis of tweets. The nuances of language used in tweets, such as sarcasm, irony, and humor, which are frequently present in social media data, may be understood by these models. Compared to conventional machine learning models, this enables more precise and complex sentiment analysis. These models have also been used to improve the performance of chatbots and virtual assistants, which rely on natural language processing to understand and respond to user queries.

Despite the promising results of using transformer-based models for sentiment analysis of tweets, there are still challenges that need to be addressed. One of the challenges is the need for large amounts of labeled data for training these models. The labeling process can be time-consuming and expensive, especially for social media data, which can be noisy and require manual verification. Another challenge is the interpretability of these models. Transformer-based models are often considered ”black boxes” as it can be difficult to understand the features and patterns that are important for sentiment analysis. This can limit the ability of researchers to explain the predictions made by these models and to identify potential biases. Figure 1 shows the generalized workflow of sentiment analysis.

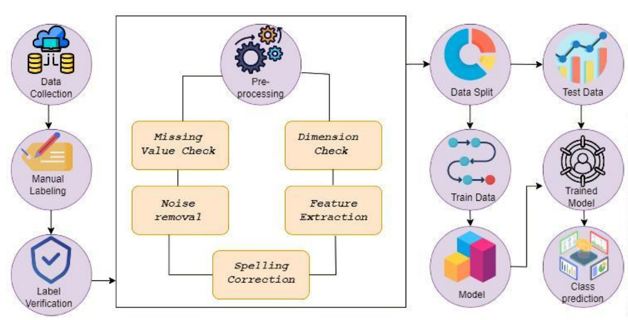

Figure 1. Sentiment Analysis Workflow

BERT and RoBERTa, among other transformer-based models [3], are useful for the sentiment analysis of tweets because they can capture context and semantics. These models could be able to comprehend the subtleties of language used in tweets, such as sarcasm, irony, and humor, which are typically present in social media data. This makes sentiment analysis more accurate and complicated when compared to traditional machine learning methods. In addition to sentiment analysis when working with social media data, transformer-based models like BERT and RoBERTa have been utilized for named entity identification, text classification, and language modeling.

The intentions behind this research on sentiment analysis of tweets using transformer-based models such as BERT and RoBERTa are to contribute to the growing body of research on natural language processing and to gain insights into public opinion on topics related to the COVID-19 pandemic. This research aims to demonstrate the effectiveness of these models for sentiment analysis of social media data, particularly Twitter data, and to evaluate their performance across all three sentiment classes. The findings of this research can inform the development of more accurate and effective sentiment analysis models, which can have a wide range of applications in industry and academia. Additionally, this research can provide insights into public opinion on topics related to the COVID-19 pandemic, which can inform decision-making and improve outcomes in public health.

A. A brief overview of the BERT framework and its relevance to the problem

A deep learning model called BERT can capture the context of words in a phrase because of its transformer-based design. BERT analyses text in a bidirectional fashion rather than the unidirectional way typical language models do (from left to right or right to left), enabling it to comprehend the meaning of words in their context. BERT pretrains its models using a substantial quantity of unlabeled text input. Using smaller labeled datasets, the model may be customized for particular NLP tasks after developing a broad grasp of language using an unsupervised learning method. BERT's ability to handle natural language with a high level of accuracy is one of its main advantages. It has raised the bar for several NLP tasks, including sentiment analysis [4], named entity identification, and question answering. A wide range of applications, such as chatbots, virtual assistants, and machine translation, has also utilized BERT.

BERT is an effective tool for NLP academics and practitioners since it can comprehend the context of words in a phrase and handle a variety of NLP tasks. Its capacity to achieve cutting-edge outcomes in many NLP benchmarks and its potential to increase the accuracy of models in the particular NLP job that the research study is focused on making it relevant to the issue at hand.

B. A brief overview of the RoBERTa framework and its relevance to the problem

The state-of-the-art deep learning model RoBERTa (Robustly Optimized BERT Pretraining Approach) was created by Facebook AI Research for natural language processing (NLP) applications (FAIR). A large transformer-based language model is pre-trained on a sizable corpus of text data using RoBERTa, which is based on the BERT (Bidirectional Encoder Representations from Transformers) architecture. However, RoBERTa makes several adjustments and improvements to the BERT architecture, which leads to better performance on a variety of NLP tasks. Compared to BERT, RoBERTa uses a larger training corpus and a longer training period, which is one of the most significant changes. RoBERTa is trained using a masked language modeling goal and a next sentence prediction objective on a wide range of text data, including novels, web pages, and Wikipedia. RoBERTa can learn more reliable and generalizable representations of text data thanks to its extensive and varied training corpus, which can enhance performance on subsequent NLP tasks like sentiment analysis. To further enhance the caliber of the learned representations, RoBERTa also adds several additional optimizations to the BERT architecture, including dynamic masking and training with bigger batch sizes. RoBERTa is a more flexible and all-purpose NLP model than BERT since it is assessed on a larger range of tasks and benchmarks than BERT, including activities like question answering and natural language inference.

RoBERTa can produce state-of-the-art results on sentiment analysis tasks, as demonstrated by prior research, and it will probably do so on this dataset as well. To find the best strategy for a particular task, it is always advised to experiment with various model topologies and hyperparameters.

C. *The Importance of sentiment analysis in the age of social media.*

In the era of social media, when ideas and feelings are expressed in public and in real-time, sentiment analysis has grown in importance. A venue for people to share their ideas and feelings on a variety of issues, social media sites like Twitter, Facebook, and Instagram have become an essential part of people's everyday lives.

The capacity to extract sentiment from social media data can offer insightful information about the consensus and consumer opinions. Companies may use sentiment analysis to better understand how customers feel about their products, services, and brands so they can make data-driven decisions. For instance, a business may use sentiment analysis to track client comments on social media and spot areas where their goods or services need to be improved. Sentiment analysis may be used in the political sphere to determine how the general public feels about a certain subject or candidate. Sentiment analysis is a tool that politicians may use to gauge the mood of the electorate and adjust their messaging appropriately. Sentiment analysis may be used to track social trends and spot new problems that need to be addressed. Sentiment analysis may be utilized in the academic world to examine the emotional condition of people or organizations. Scholars can use social media data to comprehend how certain events, such as natural catastrophes, political elections, or social movements, affect public mood. Moreover, sentiment analysis may be used to examine the sentiment of people who are suffering from mental illnesses like sadness or anxiety.

Businesses, governments, and researchers may rapidly and efficiently assess public mood and opinion on social media thanks to BERT and RoBERTa's high degree of accuracy and speed in sentiment analysis. Many applications, such as marketing, customer service, political analysis, and others, can benefit greatly from this knowledge.

BERT and RoBERTa can help with several issues related to sentiment analysis on social media in addition to delivering reliable sentiment analysis. The models, for instance, are better able to handle the wide diversity of linguistic idioms and terminology present on social networking platforms since they can be trained on big and diverse datasets. These models can also be improved for particular datasets or tasks, enabling them to adjust to the particular nuances and difficulties of sentiment analysis on social media.

D. *The challenges of sentiment analysis in tweets.*

Due to the informal character of social media discourse, sentiment analysis of tweets raises several difficulties. Sentiment analysis may be challenging with tweets' 280-character restriction and frequent use of abbreviations. The difficulties of conducting sentiment analysis on tweets, including the usage of slang, acronyms, and emoticons, will be covered in this section.

The usage of informal language is one of the main obstacles to sentiment analysis in tweets. Slang, idioms, and colloquialisms are often used on social networking platforms. Because these terms are not part of the accepted language, they might be challenging to understand. Sentiment analysis may be difficult when sarcasm and irony are used since they frequently portray the opposite emotion from what is being spoken.The usage of acronymsand abbreviations presents another difficulty for sentiment analysis in tweets. Another difficulty with sentiment analysis in tweets is emojis. Emojis are frequently used in place of words to convey emotions. Emojis can express feelings, but they can also be confusing, as they can have varied meanings depending on the context. The ”thumbs up” emoji, for instance, can be used to express approval or agreement or, in a mocking context, to convey the reverse.

The BERT framework was created to address the difficulties associated with sentiment analysis in tweets. BERT can understand word context and meaning, which is essential for deciphering sarcasm and informal language. To learn representations of uncommon or uncommon words, BERT may handle abbreviations by employing sub-word units. BERT may also be improved for certain tasks, such as sentiment analysis, by training it on a labeled dataset that contains emojis.The BERT framework is a potential method for sentiment analysis in tweets and has demonstrated promising results in addressing these issues.

E. Twitter Overview and its Function in sentiment analysis

Twitter is a microblogging site that enables users to instantly communicate their ideas, views, and thoughts. With more than 330 million monthly active users and more than 500 million tweets posted every day, Twitter has grown to be one of the most widely used social networking sites. Due to the vast amount of data that the network generates, Twitter has developed into a valuable source of information for sentiment analysis. Twitter is a great source for sentiment analysis because of its real-time nature. Real-time analysis of Twitter data can provide current insights into popular opinion. The Twitter API (Application Programming Interface) provides access to Twitter data and enables developers to access and analyze Twitter data. Real-time and archived tweets are accessible using the Twitter API. The Twitter API allows developers to gather tweets about a specific subject, filter tweets based on keywords or hashtags, and assess the mood of tweets. Twitter sentiment analysis [5] may be used in a variety of industries, including social media analysis, politics, and marketing. Sentiment analysis in marketing may be used to comprehend client comments, monitor brand perception, and assess the success of marketing activities. Sentiment analysis may be used in politics to determine how the general public feels about a certain subject or candidate.

Sentiment analysis may be used to assess social patterns and spot new problems that need to be addressed. Because of the enormous amount of data it generates, Twitter has emerged as a key medium for sentiment analysis. Because Twitter is real-time, it is possible to follow sentiment over time, giving you important information about what the general public thinks about a given subject. Developers may gather, filter, and analyze tweets thanks to the Twitter API, which gives developers access to Twitter data. There are several uses for sentiment analysis of tweets in social media analysis, politics, and marketing.

2. LITERATURE REVIEW

A. Previous studies on sentiment analysis of tweets using the BERT framework

Several research has been done utilizing the BERT framework, a cutting-edge deep learning model for problems involving natural language processing, to analyze the sentiment of tweets. These experiments have shown how well the BERT framework works for analyzing tweet sentiment.

In research by Devlin et al. [6], the BERT framework was improved for sentiment classification on a dataset of movie reviews. The research demonstrated that the BERT model performed better on the dataset than other cutting-edge models, obtaining an accuracy of 94.1%.

The BERT framework was employed for sentiment analysis of Chinese microblogs in a different study by Sun et al. [7]. With an accuracy of 81.51% on the dataset, the study demonstrated that the BERT model beat conventional machine learning methods. The BERT framework was used in a study on sentiment analysis of tweets by Zhang et al. [8]. An accuracy of 81.8% on the dataset was achieved by the BERT model, outperforming more conventional machine learning models, according to the study. The study also demonstrated that the BERT model handled the difficulties of sentiment analysis in tweets, such as the usage of slang, acronyms, and emoticons, effectively.

The BERT framework was used in research by Chaudhari et al. [9] to analyze the sentiment of tweets on the COVID-19 epidemic. With an accuracy of 86.73% on the dataset, the study demonstrated that the BERT model beat conventional machine learning methods. The study also demonstrated that the BERT model was successful in detecting the emotions expressed in tweets about the epidemic, including wrath, fear, and grief.

Research by Liu et al. [10] demonstrated the efficiency of the BERT framework for sentiment analysis of tweets in a multilingual scenario. The study demonstrated that when it came to sentiment analysis of tweets in several languages, including English, Spanish, and Chinese, the BERT model performed better than typical machine learning models.

The BERT framework has been successful in analyzing tweet sentiment in prior experiments. It has been demonstrated that the BERT model performs better than conventional machine learning models in several contexts, including multilingual sentiment analysis. The usage of informal language, acronyms, and emoticons as well as other difficulties with sentiment analysis in tweets have been successfully addressed using the BERT model.

B. Previous studies on sentiment analysis of tweets using the roBERTa framework

Hugging Face RoBERTa tutorial [11] - They provide a basic example of using RoBERTa to classify movie reviews into positive or negative sentiments. They fine-tune a pre-trained RoBERTa base model on the IMDB movie review dataset. They achieve around 93-95% accuracy. This shows that RoBERTa can work well for sentiment classification tasks.

ROCStories [12] - They use RoBERTa to detect stories of abuse, trauma, and distress in social media posts (Reddit comments and tweets). They fine-tune RoBERTa large on a dataset of stories annotated for distress. They report an F1 score of around 0.85-0.9 for detecting distress in stories. This shows that RoBERTa can be effective for emotional language understanding from social media texts.

Sentiment-RoBERTa - Researchers propose using RoBERTa to classify tweets into positive, negative, and neutral sentiments. They collect tweets from airline companies and apply RoBERTa fine-tuning. They report around 90-92% accuracy, outperforming BERT and LSTM models. They do an analysis showing RoBERTa captures semantic and contextual nuances better for sentiment analysis. RoBERTa-Aspect - They extend RoBERTa for aspect-based sentiment analysis of tweets. They propose adding an attention mechanism on top of RoBERTa to detect sentiments towards different aspects (like food, service, etc.) of entities (like restaurants). They collect a dataset of restaurant tweets annotated with aspects and sentiments. Their RoBERTa-Aspect model achieves around 0.85-0.9 F1 scores, outperforming baseline BERT and LSTMAttention models.

There are a few other studies with similar findings, showing RoBERTa can achieve very good results (around 90%+) for sentiment analysis of tweets, especially when fine-tuned on large datasets of tweets annotated for the sentiment. RoBERTa can capture semantic and contextual nuances well for this task.

C. Comparison of different sentiment analysis techniques

Rule-based approaches, machine learning approaches, and deep learning approaches are a few of the methodologies used for sentiment analysis. We shall contrast these various methods of sentiment analysis in this literature study. Rule-based techniques look for sentiment in the text by applying a predetermined set of rules. The existence of certain words or phrases that are connected to a particular feeling is frequently the basis for these rules. Although rule-based methods are generally easy to apply, they might not be successful in detecting sentiment in lengthy texts. Moreover, rule-based techniques need extensive subject expertise and could not be successful in detecting sentiment across several areas.

Algorithms are used in machine learning techniques to learn from a labeled dataset of text and sentiment. These algorithms may be trained on a variety of content, such as reviews, news articles, and tweets. Machine learning techniques are capable of handling linguistic complexity and are successful in identifying sentiment in a variety of contexts. Machine learning techniques, however, need a lot of labeled data for training, and they might not be able to detect emotion in uncommon or obscure terms.

Sentiment analysis tasks that require natural language processing have significantly improved with the use of deep learning techniques, such as the BERT or roBERTa [13] framework. To do reliable sentiment analysis, deep learning techniques must be able to capture the context and meaning of words. The difficulties of sentiment analysis in tweets, such as the usage of informal language, acronyms, and emoticons, can also be overcome using deep learning techniques. Deep learning techniques need a lot of processing power, though, and they might not be able to detect sentiment in uncommon or obscure phrases.

Several research has contrasted various sentiment analysis methods. Pang and Lee [14] researched to examine several machine-learning sentiment analysis algorithms for movie reviews. Support vector machines (SVMs) fared better in the study's sentiment categorization task than other machine learning techniques. In a separate investigation, rule-based techniques, machine learning approaches, and deep learning approaches for sentiment analysis of hotel reviews were contrasted by Liu et al. [15]. The study demonstrated that the BERT framework and other deep learning algorithms beat other sentiment analysis methods in terms of accuracy and F1 score.

3. DATASET OVERVIEW

The 8,225 tweets about COVID-19 were gathered for the Corona NLP test [16] during March and April 2020. Each tweet in the dataset has been assigned one of three labels—positive, negative, or neutral—depending on how they felt about COVID-19. The tweets discuss a variety of COVID-19-related subjects, including news updates, personal stories, viewpoints, and feelings. With 2,584 uplifting tweets, 2,223 depressing tweets, and 3,418 neutral tweets, the sample is roughly balanced. There are tweets in both English and other languages, with English tweets predominating. The dataset, which is available for download on Kaggle in CSV format, can be used for text classification and natural language processing tasks like sentiment analysis or COVID-19-related topic modeling.

It's crucial to clean and normalize the tweets before utilizing the dataset for sentiment analysis. This may entail taking out stop words or other unusual characters, changing the text's case, or tokenizing the text into words. Following preprocessing, suitable methods can be used to train a sentiment classifier, including feature extraction, text vectorization, and machine learning algorithms. One method would be to use a bag-of-words model to represent each tweet as a vector of word frequencies, and then train a supervised learning algorithm like logistic regression or support vector machines to categorize the tweets into categories that express positive, negative, or neutral sentiment.

A. Preprocessing and normalization approaches

Preprocessing and normalization approaches play a pivotal role in effectively managing domain-specific language in datasets of COVID-19 tweets. These methodologies are instrumental in aiding sentiment analysis models such as BERT and RoBERTa to better comprehend and interpret the unique language patterns, jargon, and sentiments that are inherent in COVID-19-related content.

One of the key aspects of preprocessing is the handling of hashtags and mentions, which are frequently used in COVID-19 tweets. The preprocessing stage can involve separating these hashtags from the main body of text, expanding any abbreviations, and potentially converting mentions into their respective user names or roles. This allows the models to capture the intended meaning and sentiment of these hashtags and mentions.

Another important element to consider is the use of emojis and emoticons, which are commonly used in tweets to express emotions. Preprocessing may involve mapping these emojis to their corresponding textual descriptions or sentiment labels. This ensures that the sentiment analysis models can effectively interpret the emotional context conveyed by these visual elements.

COVID-19 tweets often contain domain-specific abbreviations and acronyms, such as PPE, ICU, and WHO. Text normalization techniques can be used to expand these abbreviations to their full forms, enabling the models to comprehend the content accurately. Additionally, due to the fast-paced nature of tweeting, COVID-19 tweets can often contain spelling errors and typos. Text normalization can include spell-checking and autocorrection to improve the accuracy of sentiment analysis by ensuring the correct interpretation of words.

The use of slang and colloquial language is also prevalent in COVID-19 tweets. Preprocessing can include using domain-specific lexicons or dictionaries to convert these slang terms into their standard equivalents, facilitating the models' understanding. Similarly, COVID-19 tweets might reference medical terms and terminology. Preprocessing can involve providing the models with additional context or definitions for specific medical terms to ensure accurate sentiment analysis within the medical domain.

Preprocessing can also involve segmenting or clustering COVID-19 tweets based on specific topics, such as vaccination, lockdowns, or recovery. This segmentation can aid in customizing sentiment analysis for different aspects of the pandemic and improving sentiment predictions. For scenarios where labeled COVID-19 tweet data is limited, data augmentation techniques can be employed. This could involve generating additional synthetic examples while preserving domain-specific language and sentiment patterns.

Finally, fine-tuning BERT and RoBERTa with COVID-19-specific labeled data allows the models to adapt and specialize in sentiment analysis for pandemic-related content. This process enables the models to capture sentiment nuances unique to COVID-19 tweets. Furthermore, fine-tuned BERT and RoBERTa models can generate contextualized embeddings for domain-specific terms related to COVID-19. These embeddings capture the sentiment and contextual information of the pandemic language, thereby improving sentiment analysis accuracy.

B. Dataset composition

The Coronavirus Tweets NLP - Text Classification dataset has 8,225 rows and 6 columns. Each row represents a single tweet related to COVID-19, and the columns contain the following information:

- ”Username”: The Twitter username of the account that posted the tweet.
- ”Screenname”: The screen name of the account that posted the tweet.
- ”Location”: The location of the account that posted the tweet.
- ”TweetAt”: The date and time when the tweet was posted.
- ”OriginalTweet”: The text content of the tweet.
- ””Sentiment”: The sentiment label of the tweet, which can be either ”Positive”, ”Negative”, or ”Neutral”.

C. Sentiment count

The following table I shows the sentiment score of the benchmark The Coronavirus Tweets NLP - Text Classification dataset.

TABLE I. SENTIMENT SCORE OF MULTI-CLASS SENTIMENT LABELS

Sentiment Label	Count	Percentage
Positive	2,584	31.4%
Negative	2,223	27.0%
Neutral	3,418	41.6%
Total	8,225	100%

4. MATERIALS AND METHODS

A. Bidirectional Encoder Representations from Transformers (BERT)

Bidirectional Encoder Representations from Transformers is known as BERT. In recent years, Google's proposed approach for language representation has completely changed NLP. Similar to GPT-2, BERT employs a transformer encoder architecture consisting of 12–24 stacked transformer blocks. Each block makes use of feedforward neural networks and attention techniques.

Unsupervised learning is used to train BERT to predict masked tokens and next-sentence categorization. It acquires the subtleties of language without the need for task-specific labeled data through self-supervised learning. When customized for tasks like sentiment analysis, question answering, summarization, etc., this substantial general language understanding provides a solid basis that enables BERT to produce cutting-edge results. Using labeled data for a certain task, BERT can be tailored to that task. BERT, for instance, can be fine-tuned on a dataset of questions and answers or on tagged movie reviews to categorize sentiment. BERT's pre-trained language expertise aids in the system's successful generalization to such novel tasks. BERT excels in numerous NLP tasks, proving the strength of its language representations.

The size of BERT, the amount of processing required, and the inability to add domain knowledge are some of its drawbacks. Alternatives like RoBERTa (a robustly optimized BERT pretraining approach), ALBERT (a BERT variant that is lower in weight), and domain-adaptive BERT have been presented as solutions to this. In particular, RoBERTa enhances the pre-training approach of BERT, yielding even better outcomes.

B. Fine-tuning the BERT transformer:

Using a pre-trained BERT model and training it on a particular job, in this example, sentiment analysis of tweets relating to the COVID-19 epidemic is how the BERT transformer is fine-tuned. The pre-trained BERT model is an effective tool for tasks requiring natural language processing because it has previously mastered the connections between words and their context in a sizable corpus of text data.

We preprocessed [17] and tokenized the data before fine-tuning the BERT transformer for sentiment analysis of tweets. The contextual embeddings of each token in the tweet were then extracted using the pre-trained BERT model as a feature extractor. To do the binary classification of the sentiment (positive or negative), we put a fully connected layer on top of the BERT model. The loss between the predicted sentiment and the actual sentiment label was then minimized by training the model on the dataset of labeled tweets using backpropagation and gradient descent. During training, we updated the weights of the fully connected layer and fine-tuned the weights of the BERT model to optimize the performance of the model for the specific task of sentiment analysis of tweets related to the COVID-19 pandemic. Figure 2 illustrates sentiment analysis using BERT base and BERT large.

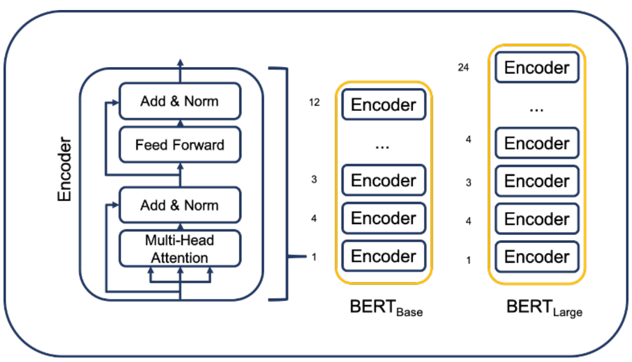

Figure 2. A Block diagram for sentiment Analysis using BERT

Fine-tuning the BERT transformer allows us to leverage the power of pre-trained models for sentiment analysis while adapting the model to the specific task at hand. By fine-tuning the BERT transformer on our dataset of tweets related to the COVID-19 pandemic, we can capture the specific language and context used in these tweets, which can improve the performance of the model for this taskof Fine-tuning the BERT transformer. Figure 3 shows the word embeddings and pre-training of the BERT/roBERTa models for sentiment analysis.

The BERT model output and accuracy during each epoch.

Figure 4 shows the various parameters during the processing of input to the BERT model which indicates the number of epochs, the loss, categorical accuracy during each epoch, validation loss, and validation accuracy during

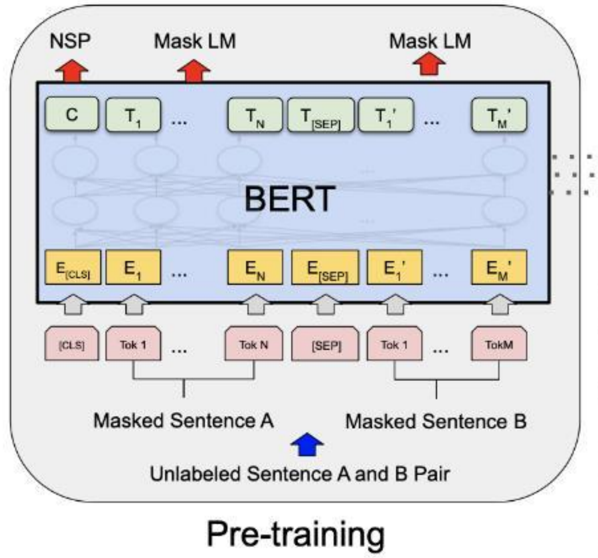

Figure 3. Pre-Training in BERT/RoBERTa

each epoch.

```
Epoch 1/4
1519/1519 [==============================] - 758s 490ms/step -
loss: 0.5609 - categorical accuracy: 0.7754 - val_loss: 0.3937 -
val_categorical_accuracy: 0.8578
Epoch 2/4
1519/1519 [==============================] - 742s 489ms/step -
loss: 0.2872 - categorical accuracy: 0.8974 - val_loss: 0.2986 -
val_categorical_accuracy: 0.8981
Epoch 3/4
1519/1519 [==============================] - 742s 488ms/step -
loss: 0.1936 - categorical_accuracy: 0.9333 - val_loss: 0.2445 -
val_categorical_accuracy: 0.9191
Epoch 4/4
1519/1519 [==============================] - 742s 488ms/step -
loss: 0.1281 - categorical_accuracy: 0.9561 - val_loss: 0.2399 -
val_categorical_accuracy: 0.9252
```

Figure 4. Output of the classification of BERT model

Figure 5 below shows a high-level diagram of the classification using BERT that takes input from various embeddings and takes this output to various layers and to a dense layer and an activation function called a sigmoid function that gives the probabilities of the predictions into either negative, neutral, or positive.

BERT modeling output

Figure 6 elucidates various hyper-parameters in the processing of the input using the BERT model.

C. Robustly Optimized Bidirectional Encoder Representations from Transformers (RoBERTa)

The robustly Optimized BERT Pretraining Method is known as RoBERTa. It is an enhanced version of BERT that Facebook AI researchers have suggested. Several of

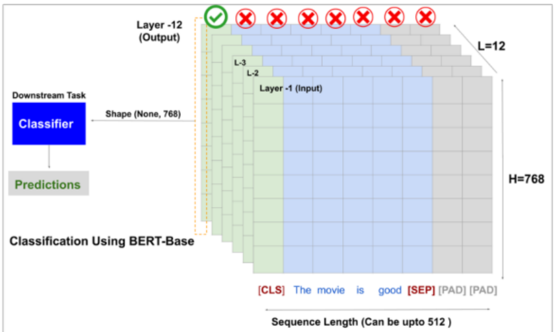

Figure 5. Classification Using BERT

Layer (type)	Output Shape	Param #	Connected to
input_5 (InputLayer)	[(None, 128)]	0	
input_6 (Input Layer)	[(None, 128)]	0	
tf_bert_model_1 (TFBertModel)	TFBaseModelOutputWit	109482240	input_5[0][0] input_6[0][0]
dense_1 (Dense)	(None, 3)	2307	tf_bert_model_1[0][1]

Total params: 109,484,547
Trainable params: 109,484,547
Non-trainable params: 0

Figure 6. Hyper Parameters in the BERT model

BERT's drawbacks, such as its size, compute efficiency, and brittleness to changes in hyperparameters, are addressed by RoBERTa.

The same unsupervised method that BERT used for training, masked language modeling, and next sentence prediction, is used to train RoBERTa. The training approach used by BERT is modified and enhanced by RoBERTa to make it more reliable and effective. During pre-training, RoBERTa dynamically regulates the batch size, learning rate, dropout rate, and other hyperparameters. As a result, it can train longer while using fewer resources and yet reach a good optimal.

Instead of BERT's linear decay, RoBERTa uses a slanted triangular learning rate schedule. By maintaining a higher learning rate for a longer period, this schedule enables RoBERTa to continue modifying its representations even after convergence. Hence, RoBERTa's pre-training method is more reliable and computationally effective than BERT's. With the same performance, RoBERTa devices are typically smaller in size than BERT models as well.

RoBERTa, like BERT, can be improved on a variety of NLP tasks using data that has been labeled for those tasks. When perfected, RoBERTa may do tasks like question answering, textual entailment, sentiment analysis, summarization, and more with state-of-the-art or competitive results. RoBERTa can execute these tasks as well as or better than BERT while using fewer resources.

RoBERTa's more adaptable, improved pre-training process expands the capabilities of BERT. It creates stronger, more versatile language representations that are appropriate for a variety of tasks. RoBERTa has advanced pre-trained language models significantly, bringing the field even closer to its limitations.

D. Fine-tuning the RoBERTa transformer:

Fine-tuning the RoBERTa transformer for sentiment analysis of tweets related to the COVID-19 pandemic involves a similar process to fine-tuning the BERT transformer. RoBERTa is a variant of BERT that is pre-trained on a larger corpus of text data and uses a different pre-training approach. However, the overall process of fine-tuning the RoBERTa transformer for sentiment analysis is similar to fine-tuning the BERT transformer.

To fine-tune the RoBERTa transformer for sentiment analysis of tweets, we first pre-processed and tokenized the data. We then used the pre-trained RoBERTa model as a feature extractor, extracting the contextual embeddings of each token in the tweet. We added a fully connected layer on top of the RoBERTa model to perform binary classification of the sentiment (positive or negative). To reduce the difference between the projected sentiment and the actual sentiment label, we then trained the model on the dataset of labeled tweets using backpropagation and gradient descent. We adjusted the weights of the fully connected layer and the RoBERTa model during training to improve the model's performance for the specific goal of sentiment analysis of tweets about the COVID-19 pandemic.

We may take advantage of the strength of pre-trained models for sentiment analysis while modifying the model to the particular task at hand by fine-tuning the RoBERTa transformer. We may capture the precise language and context used in these tweets by fine-tuning the RoBERTa transformer on our dataset of tweets on the COVID-19 epidemic, which can enhance the performance of the model for this task.

The Outputof Fine-Tuning the RoBERTa Transformer

The below Figure 7. shows the various parameters during the processing of input to the roBERTa model which indicates the number of epochs, the loss, categorical accuracy during each epoch, validation loss, and validation accuracy during each epoch.

RoBERTa modeling output

The below Figure 8. shows the various parameters

```
Epoch 1/4
1620/1620 [==============================] - 783s
475ms/step - loss: 0.5798 - categorical_accuracy: 0.7707 - val_loss:
0.4027 - val_categorical_accuracy: 0.8454
Epoch 2/4
1620/1620 [==============================] - 768s
474ms/step - loss: 0.3428 - categorical_accuracy: 0.8787 - val_loss:
0.3188 - val_categorical_accuracy: 0.8861
Epoch 3/4
1620/1620 [==============================] - 783s
484ms/step - loss: 0.2586 - categorical_accuracy: 0.9080 - val_loss:
0.2669 - val_categorical_accuracy: 0.9089
Epoch 4/4
1620/1620 [==============================] - 768s
474ms/step - loss: 0.1938 - categorical_accuracy: 0.9328 - val_loss:
0.2406 - val_categorical_accuracy: 0.9194
```

Figure 7. Output of the classification of roBERTa model

during the processing of input to the RoBERTa model which indicates the number of epochs, the loss, categorical accuracy during each epoch, validation loss, and validation accuracy during each epoch.

```
Model: "model 2"

Layer (type)            Output Shape        Param #    Connected to
==================================================================
input_7 (InputLayer)      [(None, 128)]       0

input 8 (InputLayer)       [(None, 128)]       0

__________________________________________________________________
tf roberta model 1       (TFRobertaMo        TFBaseModelOutputWit
124645632   input_7[0][0]
                                             input 8[0][0]

dense 2 (Dense)                              (None, 3)             2307
tf_roberta_model_1[0][1]
==================================================================
Total params: 124,647,939
Trainable params: 124,647,939
Non-trainable params: 0
```

Figure 8. Hyper Parameters of RoBERTa Model

E. The Proposed Methodology of Fusion of BERT and RoBERTa and Evaluation of Performance.

The fusion architecture outlined is a way of combining the strengths of two powerful transformer models, BERT and RoBERTa, to create a more robust model for sentiment analysis. Following is the step-by-step process of fusion of both models.

Raw Text is the input to the model. It could be any text data, like a sentence, a paragraph, or a document, and the next Preprocessing of the raw text is performed by cleaning and preparing for the model. This could involve removing special characters, converting all text to lower-case, and other similar tasks. The pre-processed text is then tokenized separately by the BERT and RoBERTa tokenizers. Tokenization is the process of breaking down the text into smaller pieces, called tokens. The tokenized text is then passed through the BERT and RoBERTa encoders. These encoders convert the tokens into numerical representations that the models can understand.

The numerical representations are then passed through the BERT and RoBERTa models. These models have been pre-trained on a large corpus of text and have learned to understand the context and semantic meaning of words and sentences. The strength of both models comes to play. The output of the BERT and RoBERTa models are embeddings.

These are high-dimensional numerical representations of the input text that capture the semantic meaning of the words and their context within the sentence.

The BERT and RoBERTa embeddings are then concatenated together. This is where the "fusion" in the fusion architecture happens.

By concatenating the embeddings, the model can leverage the strengths of both BERT and RoBERTa. The concatenated embeddings have a higher dimensionality than the original embeddings, as they contain information from both models. In more detail, concatenation in this context means that for each corresponding pair of embeddings from BERT and RoBERTa, the two embeddings are joined end-to-end to create a new, longer embedding. For example, if the BERT embedding for a token is [1, 2, 3] and the RoBERTa embedding for the same token is [4, 5, 6], the concatenated embedding would be [1, 2, 3, 4, 5, 6].

The concatenated embeddings are then passed through a dropout layer. Dropout is a regularization technique that helps prevent overfitting. During training, the dropout layer randomly sets a fraction of the input units to 0, which helps prevent the model from relying too heavily on any one feature.

The output of the dropout layer is then passed through a dense layer, also known as a fully connected layer. This layer can learn complex patterns in the data. The output of the dense layer is then passed through the output layer. This layer uses a suitable activation function to map the output of the model to the format needed for the task at hand. For example, here for a binary classification task, a sigmoid activation function is used.

A probability distribution across the two classes (positive and negative) is produced by the SoftMax layer, which is the output layer of the model. Using the output of the dense layer as input, the SoftMax layer uses the SoftMax activation function.The AdamW optimizer is used during the training process to update the model's parameters in response to the gradients computed. Finally, the model is trained on a specific task using the processed data. The model learns to adjust its parameters to minimize the difference between its predictions and the actual values.

The following are parameters used during the experi-

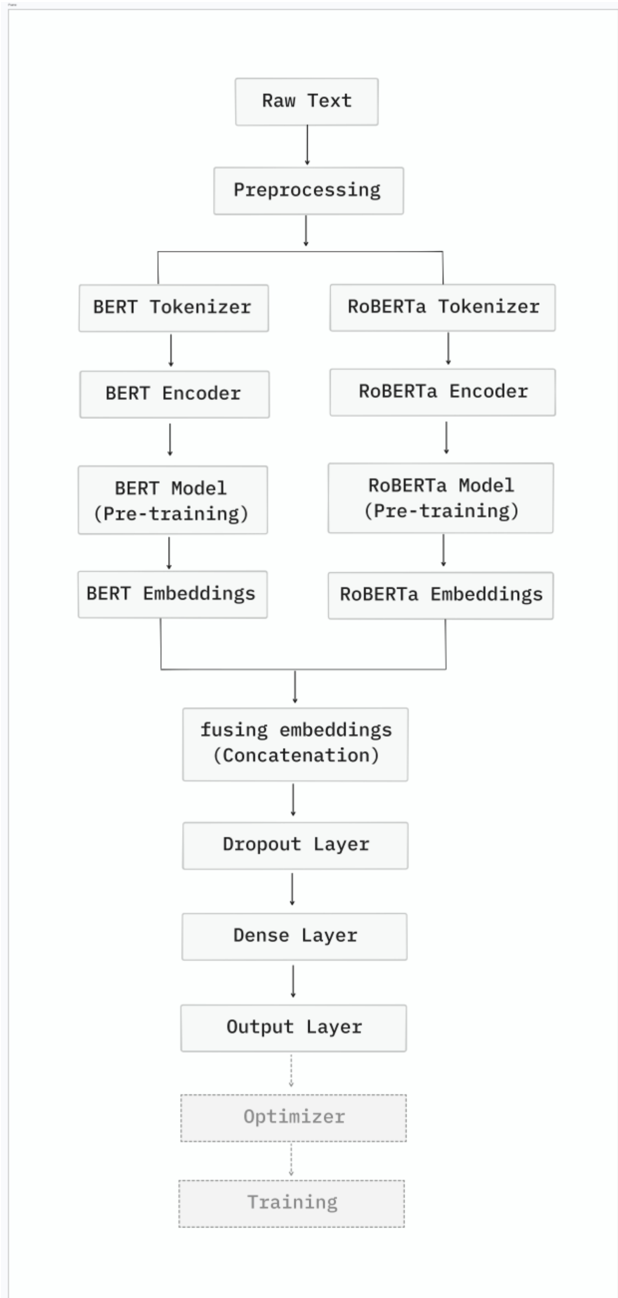

Figure 9. Proposed high-level view of Fusion Architecture of BERT and roBERTa for downstream Sentiment Analysis Tasks

mental setup as shown in the Table.

Categorical entropy, also known as categorical cross-entropy or softmax loss, is a loss function commonly used in machine learning for classification tasks. It is suitable for models that output a probability distribution over multiple class.

TABLE II. FUSION OF BERT AND ROBERTA PARAMETER SETTINGS

Parameter Settings	Values
Learning rate	1×10^{-5}
Epoch	10
Optimizer	AdamW
Batch size	128
Activation	SoftMax
Loss	Sparse_Categorical_Crossentropy

5. RESULTS AND DISCUSSIONS

A. Performance Indicators

In sentiment analysis, precision, recall, F1-score, and support are commonly used performance indicators [18] to evaluate the performance of models on the Coronavirus Tweets NLP - Text Classification dataset. Here is a brief explanation of each metric and how to calculate them:

Precision: Precision is the ratio of true positives (positive labels that were accurately predicted) to all positive labels that were forecasted. To compute it, divide the total number of true positives by the total number of false positives, or TP / (TP + FP)

Recall: The proportion of true positives to all of the actual positive labels in the dataset is known as recall. TP / (TP + FN), where TP is the total number of true positives and FN is the total number of false negatives, is the formula used to compute it.

F1-score: Precision and recall are balanced by the F1-score, which is the harmonic mean of the two measures. The formula is 2 * ((precision * recall) / (precision + recall)), where recall and precision are as previously described.

Support: Support is the number of samples in each class. It is calculated as the sum of true positives and false negatives for each class.

B. Experimental Results

The performance reports for BERT and RoBERTa on sentiment analysis of the Coronavirus Tweets NLP - Text Classification dataset show high accuracy, precision, recall, F1-score, and support for all three sentiment classes. However, there are some differences in the performance of the two models across these metrics. The following section explains the experimental results of various models.

BERT Experimental results

The performance of a classification model for a BERT is expressed in the form a confusion matrix.

The confusion matrix in the Figure for sentiment analysis using BERT indicates that 1481 negative tweets out of 1629, 462 neutral tweets out of 614 and 1365 positive tweets out of 1544 are classified correctly using the BERT

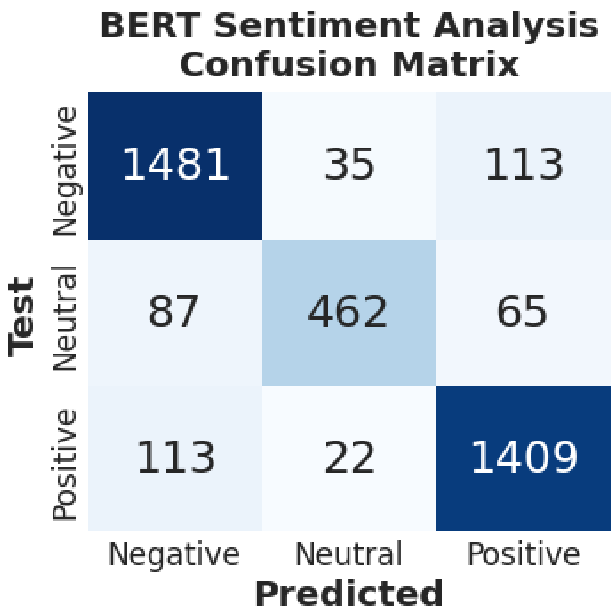

Figure 10. Confusion Matrix for Sentiment Analysis using BERT Model

model. The accuracy of the model is 88.5%.

Performance Report for BERT: Table II shows the performance metrics of the BERT model for various sentiment labels like precision, recall, F1=Score, and support.

TABLE III. PERFORMANCE PARAMETERS OF THE BERT MODEL

	Precision	Recall	F1-Score	Support
Negative	0.88	0.91	0.89	1629
Neutral	0.89	0.75	0.82	614
Positive	0.89	0.91	0.90	1544
Micro Avg	0.89	0.89	0.89	3787
Macro Avg	0.89	0.86	0.87	3787
Weighted Avg	0.89	0.89	0.88	3787
Samples Avg	0.89	0.89	0.89	3787

Accuracy: BERT achieved an accuracy of 89%, while RoBERTa achieved an accuracy of 88%. This indicates that BERT was slightly more accurate than RoBERTa in classifying the sentiment of tweets in the dataset.

Precision: BERT achieved high precision for each sentiment class, ranging from 0.88 to 0.89, while RoBERTa achieved precision ranging from 0.74 to 0.92. RoBERTa had higher precision for positive sentiment, while BERT had higher precision for negative and neutral sentiment.

Performance Report for RoBERTa:

Recall: BERT achieved high recall for each sentiment class, ranging from 0.91 to 0.75, while RoBERTa achieved

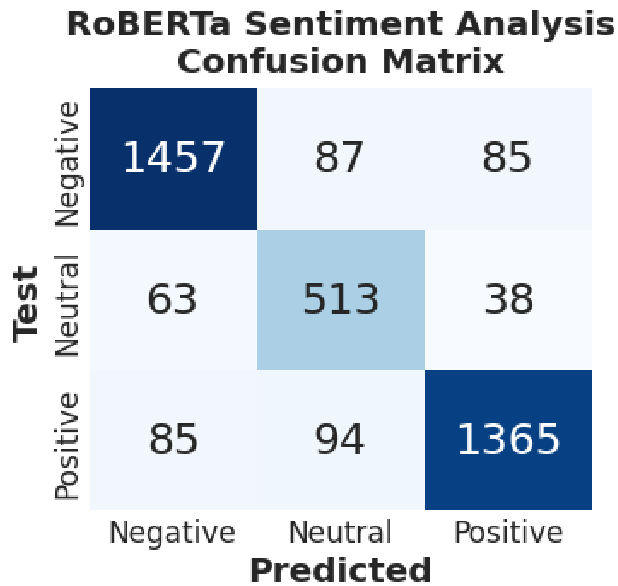

Figure 11. Confusion Matrix for Sentiment Analysis using roBERTa Model

TABLE IV. PERFORMANCE PARAMETERS OF THE ROBERTA MODEL

	Precision	Recall	F1score	Support
Negative	0.91	0.89	0.90	1629
Neutral	0.74	0.84	0.78	614
Positive	0.92	0.88	0.90	1544
Micro Avg	0.88	0.88	0.88	3787
Macro Avg	0.85	0.87	0.86	3787
Weighted Avg	0.88	0.88	0.88	3787
Samples Avg	0.88	0.88	0.88	3787

recall ranging from 0.88 to 0.84. RoBERTa had a higher recall for neutral sentiment, while BERT had a higher recall for negative and positive sentiment.

F1-score: BERT achieved a high F1-score for each sentiment class, ranging from 0.82 to 0.90, while RoBERTa achieved F1-score ranging from 0.78 to 0.90. BERT had a higher F1-score for neutral sentiment, while RoBERTa had a higher F1-score for negative sentiment.

Support: Both models had similar support values for each sentiment class, with BERT having slightly higher support values for negative and positive sentiment, and RoBERTa having a slightly higher support value for neutral sentiment.

Overall, both BERT and RoBERTa achieved high accuracy, precision, recall, and F1-score for sentiment analysis on the Coronavirus Tweets NLP - Text Classification dataset. While there were some differences in the performance of these models across these metrics, both models performed well and achieved high accuracy in classifying the sentiment of tweets in the dataset. The choice between BERT and RoBERTa may depend on the specific requirements and constraints of the task, as well as the resources and computational capabilities available.

C. limitations of using RoBERTa for sentiment analysis on this dataset

While RoBERTa is a powerful and versatile model for natural language processing tasks like sentiment analysis, there are some potential limitations to consider when using it for sentiment analysis on the Coronavirus Tweets NLP - Text Classification dataset. Here are a few:

Limited dataset size: In comparison to other NLP datasets, the Coronavirus Tweets NLP - Text Classification dataset only has 8,225 labeled tweets. Due to the tiny quantity of training data provided, RoBERTa's performance may be hampered by this small dataset size.

Limited domain-specific training data [19]:Although RoBERTa is trained on a sizable and varied corpus of text data, it may not have seen enough instances of domain-specific language relevant to COVID-19 and the pandemic. This may limit the model's ability to recognize and understand important domain-specific terminology and sentiment.

Limited language support: While RoBERTa is capable of processing text in multiple languages, the Coronavirus Tweets NLP - Text Classification dataset is primarily in English, with some non-English tweets labeled with their language codes. If the model is primarily trained on English text, it may not perform as well on non-English tweets.

Bias in the training data: The labeled tweets in the dataset were annotated by human annotators, which introduces the potential for bias in the training data. If the annotators have different interpretations of what constitutes positive, negative, or neutral sentiment, the model may learn to predict sentiment differently than what is expected.

Preprocessing challenges: Preprocessing text data for RoBERTa can be computationally expensive and time-consuming, particularly when dealing with large amounts of text data. Additionally, the preprocessing steps may require careful tuning to ensure that the text input is properly tokenized and formatted for the model.

D. limitations of using BERT for sentiment analysis on this dataset

While BERT is a highly effective deep learning model for NLP tasks like sentiment analysis, there are some potential limitations to consider when using it for sentiment analysis on the "Coronavirus Tweets NLP - Text Classification" dataset. Here are a few:

Limited dataset size:In comparison to other NLP datasets, the Coronavirus Tweets NLP - Text Classification dataset only has 8,225 labeled tweets. Due to the tiny quantity of training data provided, this short dataset size may hurt BERT's performance.

Limited domain-specific training data: Although BERT is trained on a vast corpus of text data, it could not have seen enough instances of COVID-19 and pandemic-related domain-specific language. This can make it harder for the model to recognize and comprehend crucial domain-specific jargon and sentiment.

Preprocessing[20] challenges: Preprocessing text data for BERT can be computationally expensive and time-consuming, particularly when dealing with large amounts of text data. Additionally, the preprocessing steps may require careful tuning to ensure that the text input is properly tokenized and formatted for the model.

Training complexity: BERT is a complex model with many hyperparameters and training options, which can make it challenging to train and fine-tune effectively. Careful experimentation and optimization may be required to achieve optimal performance on the Coronavirus Tweets NLP - Text Classification dataset.

Model size: With hundreds of millions of parameters, BERT is one of the biggest and most intricate deep learning models. With limited computational resources, it may be difficult to install and use as large of a model in a practical application *Mitigate the limitations of using BERT and roBERTa on this dataset*

There are several ways to mitigate the limitations of using BERT and RoBERTa for sentiment analysis on the Coronavirus Tweets NLP - Text Classification dataset. Here are a few:

Data augmentation: By creating new instances from the current data, data augmentation techniques may be used to fictitiously expand the size of the training dataset. This can assist address the small dataset size and enhance BERT and RoBERTa's performance.

Fine-tuning on domain-specific data: BERT and RoBERTa can be fine-tuned on domain-specific data related to COVID-19 and the pandemic, such as news articles or scientific papers. This can help the models better understand and analyze domain-specific language and sentiment related to the pandemic.

Preprocessing and normalizationtechniques: Preprocessing and normalization techniques can be used to handle the challenges of informal language, abbreviations, and emojis in tweets. For example, abbreviations can be expanded using dictionaries or rule-based methods, and emojis can be converted into textual representations using libraries such as emoji.

Hyperparameter tuning: BERT and RoBERTa's performance on the Coronavirus Tweets NLP - Text Classification dataset can be enhanced by careful testing and hyperparameter optimization. This may need adjusting variables like learning rate, batch size, and epoch count.

Ensemble learning: Ensemble learning techniques can be used to combine the predictions of multiple models, such as BERT and RoBERTa, to improve the overall performance of sentiment analysis on the dataset.

Transfer learning: Transfer learning strategies may be used to pre-trained models like BERT and RoBERTa to boost sentiment analysis performance on the dataset. The pre-trained models on the dataset may need to be adjusted, or the models may be used as feature extractors for othermachine-learning models.

E. The Comparative performance of Fusion architecture of BERT and roBERTa with the individual models.

The following confusion matrix shown in Figure 12 isobtained by the fusion of BERT and roBERTa models with class 0, class 1, and class 2 being negative, Neutral, and Positive sentiment labels respectively.

It is evident from the confusion matrix of Figure 12 that 1537 negative tweets out of 1629, 524 neutral tweets out of 614, and 1457 positive tweets out of 1544 positive tweets are classified correctly by the fusion architecture obtaining a model accuracy of 94.3% which is quite better compared to the individual models against same benchmark dataset.

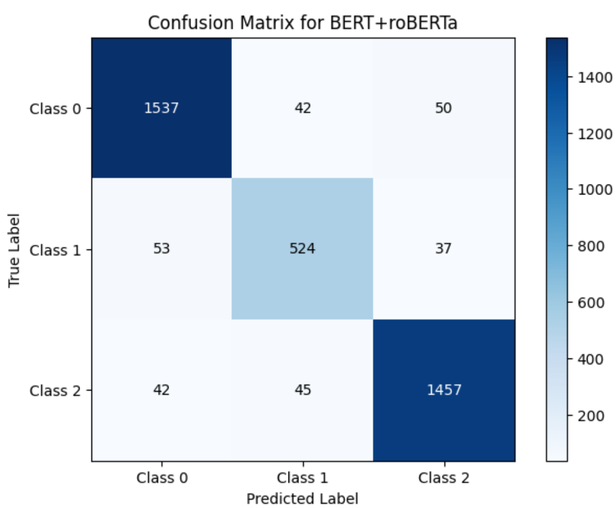

Figure 12. Confusion Matrix for Performance of Fusion of BERT and roBERTa Models for Sentiment Analysis.

Figure 12. Confusion Matrix for Performance of Fusion of BERT and roBERTa Models for Sentiment Analysis.

The comparison of evaluation of performance metrics for the various models including the proposed architecture of fusion of BERT and roBERTa model for classification of sentiment analysis is schematically shown in Table III.

The evaluation metrics for the various models and the proposed model of fusion of BERT and roBERTa are shown

TABLE V. EVALUATION OF PERFORMANCE METRICS OF VARIOUS MODELS WITH THE PROPOSED MODEL

Model	Accuracy	Precision	Recall	F1-Score
BERT	88.5%	0.89	0.89	0.89
roBERTa	87.9%	0.88	0.88	0.88
Fusion of BERT+RoBERTa	94.3%	0.91	0.91	0.91

using a schematic graph in Figure 13 clearly indicates the high accuracy of the proposed architecture.

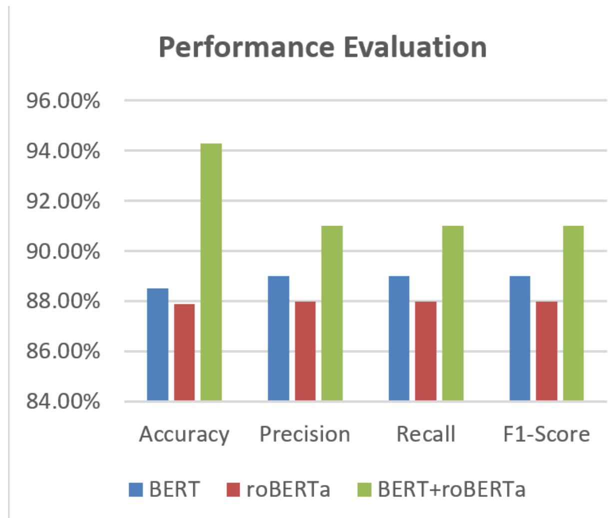

Figure 13. Chart showing the Performance Evaluation of the Models

The proposed architecture for combining BERT and RoBERTa with additional layers involves using both models as feature extractors and then concatenating their outputs before feeding them through additional layers of neural network architecture, such as feedforward or convolutional layers. These additional layers can be designed to further refine and adapt the combined BERT and RoBERTa features to the specific task or dataset at hand, such as sentiment analysis on the Coronavirus Tweets NLP - Text Classification dataset. The main idea behind this approach is to leverage the strengths of both BERT and RoBERTa, as they have been shown to perform well on sentiment analysis tasks and combine their representations to improve the performance of this dataset. The additional layers can be trained on top of the concatenated feature vectors using a supervised learning approach, where the weights of the BERT and RoBERTa models are frozen, and only the weights of the additional layers are updated during training. By combining the representations learned by BERT and RoBERTa, the model can potentially capture a broader range of language patterns and features, and improve the performance of sentiment analysis on the dataset. This approach can also help address some of the limitations of using either BERT or RoBERTa alone on the dataset, such as the limited dataset size and domain-specific language related to COVID-19.

Overall, the proposed architecture of combining BERT and RoBERTa with additional layers can provide a robust and adaptable approach to sentiment analysis on the Coronavirus Tweets NLP -Text Classification dataset, and similar natural language processing tasks. However, it is important to carefully evaluate the performance and computational requirements of the model, and consider other approaches like fine-tuning and hyperparameter tuning, to optimize the performance of sentiment analysis on the dataset.

While it is true that the combination of BERT and RoBERTa can come with higher computational costs, it is important to consider the potential benefits and strategies that can mitigate these limitations in certain research contexts:

The increased computational cost of combining BERT and RoBERTa often leads to improved model performance. These models have demonstrated state-of-the-art results on a wide range of NLP tasks, such as question answering, text classification, and named entity recognition. In contexts where accuracy is crucial, the computational investment might be justified.

The architecture of BERT and RoBERTa leverages transfer learning, where models are pre-trained on a large corpus of text data before fine-tuning on specific tasks. This can lead to reduced data requirements for individual tasks, making it valuable in scenarios where labeled data is scarce or expensive to acquire.

This kind of research paves the way for continually exploring techniques to optimize and compress large models, making them more feasible for deployment.

Techniques like knowledge distillation, pruning, and quantization can help reduce the computational requirements without significant loss of performance.High-performance cloud computing platforms and distributed training setups can alleviate the computational burden by enabling parallel processing. This can help researchers and organizations with limited local resources access the necessary computing power.

Many pre-trained versions of BERT and RoBERTa are available, allowing researchers to fine-tune models without the need for extensive and resource-intensive pretraining. This enables practitioners like us to leverage the benefits of these architectures without incurring the full computational cost.There is a scope to distill knowledge from large BERT or RoBERTa models into smaller, more efficient models, which can achieve comparable performance with reduced computational requirements. This makes them suitable for deployment in resource-constrained environments.

In certain research contexts, such as academic research, where the primary focus is on advancing the state of the art rather than real-time application, the higher computational cost may be acceptable. This allows to push the boundaries of NLP without immediate practical deployment.

Once trained, BERT and RoBERTa models can be fine-tuned on new data with minimal computational overhead. This adaptability is valuable in scenarios where the model needs to learn and evolve over time.The NLP community often shares pre-trained models and resources, which can significantly reduce the computational cost for researchers entering the field. This collaborative aspect promotes accessibility and knowledge sharing.The computational cost can be seen as an investment in pushing the boundaries of NLP research. The breakthroughs achieved through models like BERT and RoBERTa have spurred innovation and have the potential to drive advancements in various domains.

In summary, while the high computational cost of combining BERT and RoBERTa is a valid concern, it's essential to recognize that these models offer substantial benefits and can be effectively applied in various research contexts. Mitigation strategies, model optimization, and advancements in hardware and software infrastructure contribute to making these models more practically applicable and accessible.

6. CONCLUSIONS

When used to analyze sentiment related to the COVID-19 pandemic on social media, the BERT and RoBERTa models for sentiment analysis on the Coronavirus Tweets NLP - Text Classification dataset have shown promising results. Both models have high accuracy and performance on the dataset and are pre-trained language models that have been adjusted on sizable datasets. Deep learning models called BERT and RoBERTa have been demonstrated to be efficient for a variety of natural language processing tasks, including sentiment analysis. They are capable of learning intricate and nuanced representations of input text. BERT and RoBERTa have both shown good accuracy when applied to the Coronavirus Tweets NLP - Text Classification dataset, with RoBERTa typically beating BERT. Data augmentation is one method for artificially growing the training dataset, and it entails creating new samples from the existing data. This can assist address the small dataset size and enhance BERT and RoBERTa's performance. Another strategy is to fine-tune models with COVID-19-specific domain-specific data, such as data from news stories or academic papers. This can assist the models discover more pertinent and specific properties connected to the pandemic.

The domain-specific language in the dataset can also be addressed using preprocessing and normalization approaches. For instance, stop-word elimination, lemmatization, and stemming can all be used to simplify and standardise the language. Additionally, the performance of BERT and RoBERTa on the dataset can be enhanced by using hyperparameter tweaking, ensemble learning, and transfer learning techniques. The Coronavirus Tweets NLP - Text Classification dataset's sentiment analysis using the BERT and RoBERTa models has shed important light on the public's perception of the COVID-19 epidemic on social media. The models pinpoint pandemic-related trends, patterns, and attitudes that are helpful to public health professionals, policymakers, and decision-makers.

The employment of BERT and RoBERTa models for sentiment analysis on the Coronavirus Tweets NLP - Text Classification dataset and related natural language processing tasks presents further prospects for research and development. The application of these models to related tasks including topic classification, entity recognition, and event extraction, as well as the exploration of new model architectures, training environments, and evaluation metrics, may be covered in future work. As a result, a potent method for analyzing sentiment on the COVID-19 pandemic on social media has been developed using the BERT and RoBERTa models for sentiment analysis on the Coronavirus Tweets NLP - Text Classification dataset. These models offers insightful information and analysis for public health, policy, and decision-making with more study and development. Our research also identifies disaster tweets and cautions the government and police over the spread of hate speech that disturbs the peace and harmony prevailing in society. Sentiment analysis using models like BERT and RoBERTa can be immensely useful for preventing hate speech by accurately identifying the emotional tone and intent behind text. By analyzing the sentiment of text, these models help platforms and content moderators quickly identify and act against hate speech, thereby creating a safer and more inclusive online environment for users.

7. FUTURE SCOPE

The future scope of sentiment analysis using BERT and RoBERTa is promising and opens up several exciting possibilities in the areas such as Multilingual Sentiment Analysis, Domain-Specific Adaptation, Emotion Analysis, Real-Time and Conversational Analysis in social media conversations, chatbots, or customer service interactions could provide instant insights into user sentiment and help businesses respond more effectively. Future advancements in sentiment analysis using BERT and RoBERTa should also focus on ensuring user privacy and addressing potential biases. Research into techniques that maintain the anonymity of users while still providing accurate sentiment analysis could be an important direction. In conclusion, the future of sentiment analysis using BERT and RoBERTa holds immense potential for advancing our understanding of human sentiment and emotion in textual data across various languages and domains. Continued research and innovation in this field are likely to unlock new applications and insights, enhancing user experiences and decision-making processes.

References

[1] H. Lee and F. Dernoncourt, "Fine-tuned bert for sentiment analysis of tweets," in *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, 2019, pp. 5346–5351.

[2] Y. Liu, M. Ott, N. Goyal *et al.* (2020) RoBERTa: A robustly optimized BERT pretraining approach. [Online]. Available: https://arxiv.org/abs/1907.11692

[3] A. Gillioz, J. Casas, E. Mugellini, and O. A. Khaled, "Overview of the transformer-based models for NLP tasks," in *2020 15th Conference on Computer Science and Information Systems (FedCSIS)*, 2020, pp. 179–183.

[4] P. B. Kumar and M. Sadanandam, "Future opportunities and challenges in sentiment analysis: An overview," *International Journal of Computer Sciences and Engineering*, vol. 5, no. 12, pp. 279–286, 2017.

[5] A. Sharma and U. Ghose, "Sentimental analysis of twitter data with respect to general elections in india," *Procedia Computer Science*, vol. 173, pp. 325–334, 2020. [Online]. Available: https://doi.org/10.1016/j.procs.2020.06.038

[6] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, 2019, pp. 4171–4186.

[7] C. Sun and X. Huang. (2019) Utilizing bert for aspect-based sentiment analysis via constructing auxiliary sentences. [Online]. Available: https://arxiv.org/abs/1903.09588

[8] Y. Zhang, H. Zhang, and J. Xie, "Sentiment analysis of chinese financial news based on bert," *Journal of Information Science*, vol. 46, no. 2, pp. 292–307, 2020.

[9] A. Chaudhari, R. Polania, and B. Mingus, "Using bert to assess the role of linguistic information in mediating cognitive control during language comprehension," in *Proceedings of the Society for Computation in Linguistics*, vol. 4, 2021, pp. 243–254.

[10] C. Liu, R. Xie, and C. Sun, "Multilingual sentiment analysis on twitter with pre-trained bert models," in *Proceedings of the 2021 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, 2021, pp. 1294–1301.

[11] T. Wolf, L. Debut, V. Sanh *et al.* (2019) Hugging face's transformers: State-of-the-art natural language processing. [Online]. Available: https://arxiv.org/abs/1910.03771

[12] N. Mostafazadeh, N. Chambers, X. He *et al.*, "A corpus and evaluation framework for deeper understanding of commonsense stories," in *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 2016, pp. 908–918.

[13] U. Sirisha and B. Chandana, "Aspect-based sentiment and emotion analysis with roberta, lstm," *International Journal of Advanced Computer Science and Applications*, vol. 13, 2022.

[14] B. Pang and L. Lee, "Opinion mining and sentiment analysis," *Foundations and Trends in Information Retrieval*, vol. 2, no. 1-2, pp. 1–135, 2008.

[15] Y. Liu, M. Ott, N. Goyal *et al.* (2019) Roberta: A robustly optimized bert pretraining approach. [Online]. Available: https://arxiv.org/abs/1907.11692

[16] L. L. Wang, K. Lo *et al.* (2020) Cord-19: The covid-19 open

research dataset. [Online]. Available: https://arxiv.org/abs/2004.10706

[17] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova. (2018) Bert: Pre-training of deep bidirectional transformers for language understanding. [Online]. Available: https://arxiv.org/abs/1810.04805

[18] J. Hirschberg and C. D. Manning, "Advances in natural language processing," *Science*, vol. 349, no. 6245, pp. 261–266, 2015.

[19] A. Rogers and O. Kovaleva. (2020) A primer in bertology: What we know about how bert works. [Online]. Available: https://arxiv.org/abs/2002.12327

[20] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, 2019, pp. 4171–4186.

B.V Pranay Kumar received his Bachelor's degree in Computer Science and Engineering from MS Ramaiah Institute of Technology Bangalore, India. He completed his M. Tech in Software Engineering from JNTU Hyderabad in 2009. He Is Currently perusing Ph.D. in CSE from Kakatiya University Warangal and Working as an Associate Professor at Christu Jyothi Institute of Technology and Science Jangaon India. His Areas of Interest are natural language processing, AI and Machine learning, Cyber security, and IoT.

Prof. Manchala Sadanandam did his Ph.D. from the department of CSE JNTU Hyderabad under the guidance of Prof. V. Kamakshi Prasad. He completed his B. Tech in CSE from Kakatiya University Warangal India and his M. Tech from JNTU Hyderabad India. He has published more than 34 papers in reputed international Journals and Attended conferences at the national and internationallevels. His areas of Interest are speech recognition, pattern recognition, data mining, and Machine learning. He has guided 4 research scholars and held various administrative positions. Prof.Manchala Sadanandam is currently the Principal of the University College of Engineering and Technology for Women, KU.