# Abstract

In order to categorize emotions, this work uses preprocessing methods and a modified RoBERTa model. Text samples with seven different emotions from the goEmotion dataset were utilized for training and testing. Among the preparation operations carried out on the text data were the removal of stop words, punctuation, special characters, emojis, URLs, and numbers. The improved RoBERTa model, which now contains three more fully connected layers, was fine-tuned using optimized hyperparameters. Increased accuracy was seen in the trial outcomes when compared to the baseline methodology. The paper lays the framework for future research in this field by highlighting the relevance of preprocessing and model customization in order to increase the effectiveness of emotion categorization.

# Introduction

Recent years have seen a rise in the importance of sentiment analysis research, which focuses on predicting emotions from textual data (Devlin et al., 2018; Liu et al., 2019). The ability to accurately identify and categorize emotions from text is extremely important in a variety of fields, such as the study of customer reviews, social media monitoring, and mental health evaluation. With the help of a tailored RoBERTa model enhanced with extra classification layers, the research work presented here tries to address this important issue. Our goal is to improve the precision of emotion classification in textual data by utilizing deep learning techniques.

It is vital to understand how to correctly infer emotions from writing. Human communication is fundamentally impacted by emotions, which also have an impact on social relationships, decision-making, and general wellbeing. As a result, the capacity to recognize and comprehend emotions from textual data offers important insights into people's feelings and views. Furthermore, precise emotion recognition has wide-ranging effects on a variety of disciplines, including marketing, psychology, and human-computer interaction.

The limitations in current research studies on emotion prediction are what inspired us to propose this investigation. There is always room for improvement, especially in terms of accuracy, even though prior techniques have produced excellent results (Haddi et al., 2013). Our goal is to outperform the current benchmarks by tweaking the RoBERTa model's hyperparameters and adding further classification layers. We seek to improve the overall performance of emotion prediction tasks by resolving these problems.

To ensure the model receives high-quality input data, our suggested method follows a thorough pretreatment pipeline. Stop words, punctuation, special characters, emojis, URLs, and numerical values are all removed using this process (Haddi et al., 2013). We aim to improve the model's capacity to reliably detect emotional content in textual input by removing noise and unimportant information.

The goEmotion dataset, which offers labelled instances for the six main emotions as well as a class for normal samples, yields a total of seven emotion classes (Go et al., 2009), and was used to assess the efficacy of our method. We got outstanding results, with an accuracy of 67%, through precise hyperparameter optimization and intensive experimentation. This notable improvement over the current state-of-the-art result of 64% shows the effectiveness of our specialized RoBERTa model. These outcomes demonstrate the potential of our method for precisely predicting emotions in textual data while also validating its efficacy.

In conclusion, the goal of our research is to develop speech and sentiment analysis for emotion detection. We have significantly increased accuracy by utilizing a tailored RoBERTa model and adding extra classification layers. By highlighting the potential of deep learning approaches in emotion classification and highlighting the importance of precise emotion prediction for a variety of applications, this work adds to the body of current information. Overall, our findings highlight the need of utilizing cutting-edge approaches to improve emotion analysis and offer insightful information on human mood and communication.

# Related Work

In terms of predicting emotions from textual data, sentiment analysis has undergone significant improvements in recent years. We review the most recent scientific studies that have impacted the subject of emotion prediction in this part.

Due to the large volume of user-generated content and the predominance of emotions in online dialogues, sentiment analysis on social media sites has become a popular research topic. Emotions were used as noisy labels to train sentiment classifiers in a Twitter sentiment classification approach (Go et al., n.d.) developed. We were able to perform comprehensive sentiment analysis on Twitter data using this technique, providing us with insights into the sentiments and state of mind of the general population.

The preprocessing of textual input is a vital component of emotion prediction. To improve the effectiveness of sentiment classification models, (Haddi et al., 2013) examined the impact of text preprocessing in sentiment analysis. They underlined the significance of reducing noise and unimportant data. Their results showed how strategies like stopping words, removing punctuation, and managing unique characters can increase the precision of emotion prediction.

The creation of transformer-based models for natural language processing is one of the most important contributions to the discipline. BERT (Bidirectional Encoder Representations from Transformers), a pre-training method, was introduced by (Devlin et al., 2018), revolutionizing sentiment analysis and emotion prediction challenges. The accuracy and performance of BERT have significantly improved as a result of its capacity to gather contextual information and word dependencies.

(Liu et al., 2019) proposed RoBERTa, a robustly optimized BERT version that significantly improved language representation capabilities, building on the success of BERT. The pre-training approach used by RoBERTa, which included dynamic masking and extensive data augmentation, greatly enhanced the model's capacity to recognize subtle linguistic patterns. The foundation for further studies in emotion prediction has been established by these transformer-based models.

Although deep learning models have dominated the industry, other fields have also seen breakthroughs. (Xu et al., n.d.), for instance, investigated the use of graph convolutional networks (GCNs) for emotion classification. GCNs can better predict emotions by capturing syntactic and semantic information by representing the dependence links between words as a graph structure.

Contextualized word embeddings have moreover been quite important in emotion prediction tests. In order to enhance the performance of emotion categorization, (Torres, 2018) suggested a novel method that blends contextual embeddings from BERT with conventional word embeddings. A better understanding of the emotional content of text is made possible by the integration of contextual and static word representations.

The significance of reducing bias in emotion prediction algorithms has come to light in recent years. An adversarial debiasing methodology was proposed by (Odbal et al., 2022) to reduce bias in sentiment analysis. Their method increases fairness and increases the accuracy of emotion prediction models by explicitly modeling and limiting the impact of sensitive variables, such as gender or ethnicity.

In order to forecast emotions, the utilization of multi-modal data has also become a potential avenue. For the purpose of categorizing emotions, (Delbrouck et al., 2020) suggested a multimodal transformer model that incorporates textual and visual input. Their method captures both linguistic and visual cues by simultaneously modeling text and image elements, leading to increased emotion prediction accuracy.

In conclusion, since 2019, the field of emotion prediction has made considerable strides. Sentiment analysis and emotion classification tasks have been greatly improved by transformer-based models like BERT and RoBERTa. Researchers have also looked into various approaches, such as bias reduction, contextualized word embeddings, graph convolutional networks, and the incorporation of multimodal data. Collectively, these investigations improve the precision, accuracy, and understanding of emotion prediction models.Top of Form

# Approach

## Background

Natural language processing and sentiment analysis both involve the significant task of emotion classification from textual input. In order to enable applications like sentiment analysis, consumer feedback analysis, and social media sentiment assessment, it entails comprehending and classifying the emotional content contained in text. The capacity to correctly categorize emotions in text offers important insights on the feeling, actions, and intentions of people.

The classification of emotions in the past has depended on manually created characteristics and machine learning methods. The subject has been completely transformed by recent developments in deep learning, particularly transformer-based models. By capturing contextual dependencies and semantic representations of text, these models, including BERT (Bidirectional Encoder Representations from Transformers) and its variation RoBERTa, have achieved exceptional success in a variety of natural language processing tasks.

In order to encode the associations between the words in a text sequence, transformers use self-attention mechanisms. As a result, they can provide contextualized word embeddings and capture long-range dependencies. The contextualized embeddings increase the accuracy of emotion categorization by allowing the model to comprehend the intricacies and subtleties of emotional expressions in text.

On emotion-labeled datasets, pre-trained transformer models are typically fine-tuned in the context of emotion classification. The model is fine-tuned by using task-specific data and training it on a particular activity to enable it to acquire emotion-specific properties. This method makes use of the transformer model's pre-trained information and modifies it for the domain of emotion categorization.

Concerns about prejudice and fairness in the labelling of emotions have also drawn more attention recently. prejudices may appear as a result of social prejudices present in the labelled datasets or imbalances in the training data. To improve the fairness and inclusivity of emotion categorization algorithms, efforts have been made to reduce biases and achieve equitable predictions across various demographic groups.

Our strategy was to create a cutting-edge emotion classification model capable of capturing the nuanced emotional expressions in the content while promoting accuracy and precision by expanding on the developments in transformer models and taking bias mitigation strategies into account.

## Proposed Approach

Our goal in this project was to create a useful method for categorizing emotions utilizing the goEmotion data and a specialized RoBERTa model. The steps we took included choosing the dataset, text preprocessing, customizing the model, adjusting the hyperparameters, and evaluating the results.

The goEmotion dataset, which includes text samples labelled with one of seven emotion classifications, was first chosen. We were able to capture the complexity and diversity of emotions in our model training because to the dataset's wide range of emotional expressions.

On the goEmotion dataset, we preprocessed the text to ensure the best model performance. Stop words, punctuation, special characters, symbols, URLs, and digits have to be eliminated. To enable the model to concentrate on the significant information connected with each emotion, it was intended to remove noise and unimportant information from the text.

The RoBERTa model, a transformer-based language model renowned for its cutting-edge performance in natural language processing applications, served as the basis for the core architecture. On top of the underlying model, we built three fully connected layers to specifically adapt the RoBERTa model for emotion categorization. The model was able to learn emotion-specific information and generate precise predictions thanks to the addition of these extra layers, which each had 256, 128 and 7 neurons.

The model's hyperparameters were then adjusted to maximize performance. In order to do this, various hyperparameter settings were tested, including batch size, learning rate, optimizer, and loss function. The learning rate governed how quickly the model altered its parameters based on calculated gradients, while the batch size determined how many text samples were handled in each iteration during training. To effectively direct the training process and promote precise emotion predictions, we also investigated various optimizers, such as Adam or stochastic gradient descent (SGD), and assessed various loss functions, such as categorical cross-entropy.

During the training phase of our specific RoBERTa model, we loaded the preprocessed goEmotion dataset. The chosen loss function was trained into the model by backpropagation parameter modifications. To evaluate the effectiveness of our model, we employed a range of metrics, including accuracy, precision, recall, and F1-score. The F1-score offers a balanced evaluation of total performance, accounting for both recall and precision. Recall evaluated the model's capacity to properly identify every instance of a particular emotion class, while precision defined the model's ability to accurately classify instances within specific emotion classes. Accuracy offered an overall measure of accuracy.

This comprehensive approach was used to produce an emotion categorization model that was trustworthy and precise. Our strategy makes use of the goEmotion dataset, text preprocessing techniques, customization of the RoBERTa model, hyperparameter fine-tuning, and assessment metrics to ensure the model's performance in capturing the intricacies and variety of emotions in textual data.

In conclusion, our method successfully classified emotions by combining the advantages of the goEmotion dataset with the RoBERTa model, which was then fine-tuned using hyperparameters. Giving insightful details about the emotional content of text data was crucial in order to enable applications like sentiment analysis, customer feedback analysis, and social media sentiment monitoring.

# Experiments

## Dataset

A benchmark dataset for emotion categorization is the goEmotion dataset. It includes a sizable selection of text examples and the labels for the accompanying emotions. The dataset includes a wide variety of emotional expressions, allowing for the creation and assessment of reliable emotion categorization models. The text examples in the goEmotion dataset reflect actual emotional expressions in a variety of contexts and subjects. We can train our model to correctly classify emotions in many circumstances because the dataset offers a wide range of emotional expressions.

A total of 58,000 text samples, mostly from the social media site Reddit, have been collected for the goEmotion dataset from several online sources. One of seven emotion labels—joy, sadness, anger, surprise, fear, disgust, or the neutral class—is annotated next to each text sample. This extensive investigation of various emotional states is made possible by the rich assortment of emotion labels. The goEmotion dataset is divided into training, and test sets to ensure accurate evaluation. The emotion categorization model is trained using a training set that contains about 43,000 samples. During the model training phase, the training set, which contains roughly 7,500 samples, is also used to adjust the model's hyperparameters and track its effectiveness. Finally, the model's generalization and performance on untested data are evaluated using the test set, which also includes about 7,500 samples.

The goEmotion dataset contains emotions expressed across a variety of domains and subjects. This variety ensures that the model is exposed to numerous environmental cues and linguistic patterns associated with distinct emotions, making the dataset more realistic of real-world emotional expressions.

## Baseline Methodology

In this study, we compare our proposed method for categorizing emotions to a baseline method that uses a pre-trained RoBERTa model with parameter tweaking. The baseline method also utilizes the goEmotion dataset, providing a useful basis for comparison.

The RoBERTa model, a transformer-based language model well renowned for its outstanding performance in a variety of natural language processing applications, is used as the first step in the baseline technique. The pretrained model can capture detailed contextual representations of language because it was trained on a huge corpus of text data.

Ensemble learning is used to further improve the performance of the baseline technique. To provide a more reliable and accurate outcome, ensemble learning includes training numerous models and integrating their predictions. The baseline methodology uses ensemble learning to gather different viewpoints and enhance the overall functionality of the emotion categorization model.

The foundational methodology also includes parameter adjustment in addition to ensemble learning. The process of parameter tuning entails making the model's hyperparameters as efficient as possible. To ensure the best outcomes, hyperparameters like learning rate, batch size, and optimizer are fine-tuned. The baseline methodology tries to improve the model's ability to capture the subtleties of various emotions by carefully choosing and changing these hyperparameters.

The baseline methodology uses the goEmotion dataset for training and evaluation, much like our method does. A variety of text examples from this dataset are provided, each tagged with one of seven different emotion labels. We can directly compare the effectiveness of our approach to the baseline approach under comparable circumstances by using the same dataset.

The baseline methodology acts as a benchmark for assessing the efficacy of our suggested strategy. We may evaluate the improvement brought about by our modified RoBERTa model, preprocessing methods, and hyperparameter fine-tuning by comparing our results with the baseline methodology.

In conclusion, the baseline methodology makes use of the goEmotion dataset in addition to a pretrained RoBERTa model with ensemble learning and parameter adjustment. This methodology acts as a standard by which we measure the effectiveness of our strategy. We may assess the effectiveness and improvements of our contributions to the field of emotion categorization by comparing the results acquired using our proposed strategy with those obtained using the baseline methodology.

## Evaluation Measures

In this study, we used a number of evaluation measures to evaluate how well our emotion categorization algorithm performed. Accuracy, precision, recall, and F1-score were some of these measurements.

Accuracy evaluates the overall accuracy of the model's predictions by comparing the percentage of correctly classified cases to all instances in the dataset. By assessing the proportion of instances belonging to a certain emotion class that were correctly anticipated, out of all instances belonging to that emotion class, precision emphasizes the significance of precise forecasts. Recall, which is also known as sensitivity, measures the proportion of examples out of all cases that are correctly categorized as falling under a particular emotion class. The F1-score is a balanced metric that combines precision and recall into a single figure to provide an overall assessment of the model's performance.

By considering these evaluation criteria, we were able to gain greater insight into the predictability of our model, the precision of specific emotion recognition, and the trade-off between precision and recall. By comparing our results to the baseline technique, these indicators allowed us to evaluate the effectiveness of our strategy and highlight the advancements in emotion classification.

## Results

During our studies, we assessed the effectiveness of our emotion classification model using a variety of preprocessing methods. Below is a summary of each experiment's findings.:

**Baseline Methodology:** We achieved a 61% accuracy using the baseline methodology, which combined ensemble learning and parameter optimization with a pre-trained RoBERTa model. This baseline method was used as our starting point for comparison.

**Customized Model:** Next, we modified the RoBERTa model for emotion classification by including three brain layers. The tailored model's performance, however, produced a somewhat lower accuracy of 59%. This result demonstrated that, in comparison to the baseline methodology, the model's classification performance was not enhanced by the addition of the neural layers.

**Preprocessing + Baseline:** We combined the baseline methodology with preprocessing methods on the text data to further improve the performance of the model. Stop words, punctuation, special characters, icons, URLs, and numerals were among the preprocessing operations. The improved accuracy of 64% obtained from this combined strategy shows the value of preprocessing in improving the model's classification performance.

**Preprocessing + Customized Model:** Finally, we preprocessed the customized model with the additional neural layers using the same methods. The preprocessing and customized model together produced the highest accuracy, 67.01%. This result showed that, among all the conducted experiments, the joint use of preprocessing approaches and the customized model produced the best outcomes for emotion classification.

**Results Comparison:** We can see that the customized model produced an accuracy of 0.59, which was somewhat less accurate than the baseline methodology's 0.61. This shows that the performance of the model as compared to the first technique was not considerably enhanced by the addition of the three neural layers. The accuracy rose to 0.64 when preprocessing methods were added to the baseline methodology, though. This shows that improving the model's classification performance involved preprocessing techniques such removing stop words, punctuation, and special characters. The customized model and preprocessing procedures worked well together to produce an accuracy of 0.6701 and the best outcomes. This demonstrates how preprocessing and model customization both work to increase the precision of emotion classification.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Model** | **Preprocessing** | **Accuracy** | **Precision** | **Recall** | **F1-score** |
| Baseline | Before | 0.618 | 0.613 | 0.612 | 0.646 |
| Customized | Before | 0.592 | 0.594 | 0.589 | 0.584 |
| Baseline | After | 0.646 | 0.595 | 0.559 | 0.559 |
| Customized | After | 0.673 | 0.674 | 0.678 | 0.675 |

The acquired results can be ascribed to a variety of different causes. The ability of the model to learn and generalize is impacted both by the selection of the dataset and the quality of the dataset itself, including the size of the dataset and the variety of emotions it contains. The addition of three neural layers to the architecture of the model has an effect on the capacity of the model to extract meaningful features for the purpose of emotion classification. The use of preprocessing techniques, such as the removal of noise and irrelevant information, can lead to a more accurate representation of the data. By tweaking the hyperparameters, the model's best performance may be guaranteed. The choice of pertinent assessment criteria, such as accuracy, precision, recall, and F1-score, provides a thorough study of the model's performance. The outcomes are a result of taking into consideration and maximizing the effects of each of these factors individually. We can observe that the modified model yielded an accuracy that was much lower, 0.59, compared to the baseline methodology's 0.61. This shows that the model's performance did not significantly increase after the three neural layers were added compared to the method that was utilized as the baseline. However, once we added preprocessing methods to the baseline methodology, the accuracy increased to 0.64. This suggests that the preprocessing procedures, which included deleting stop words, punctuation, and special characters, played a substantial role in enhancing the model's capability to classify data. When we paired the preprocessing methods with the specialized model, we obtained the best results possible, which resulted in an accuracy of 0.6701. This illustrates that preprocessing and the customization of the model are both useful ways to improve the accuracy of emotion classification.

# Conclusion

To sum up, the main goal of our research was to classify emotions using a customized RoBERTa model and preprocessing methods. We have out comprehensive tests to assess various strategies and their effects on the precision of emotion classification. Our research has shown that combining preprocessing methods with model customization can significantly boost classification performance.

According to our findings, the modified RoBERTa model combined with preprocessing methods, such as deleting stop words, punctuation, and special characters, produced a maximum accuracy of 67.01%. This demonstrates how crucial data pretreatment is to raise the standard and applicability of the input data for tasks involving emotion categorization.

The performance of the customized model was also contrasted in our trials with that of the baseline strategy, which made use of a pre-trained RoBERTa model with ensemble learning and parameter adjustment. The integration of the three completely connected layers was intended to improve the model's capacity to detect and categorize emotions, even if the customized model's accuracy was somewhat lower than the baseline.

There are various potential areas for further development. First, further higher performance might result from further adjusting the model's hyperparameters and architecture. Incorporating more contextual data, like user demographics or temporal aspects, could improve the model's comprehension of emotions in various circumstances. Additionally, experimenting with other pretraining strategies or using more sophisticated language models, like GPT-3, may enhance the precision and generalizability of the emotion categorization model. Lastly, a more thorough investigation of emotion classification across many domains would be made possible by increasing the dataset to include a wider spectrum of emotions and distinct textual sources.

In conclusion, our research advances the field of emotion categorization by shedding light on the effects of model customization and preprocessing methods. The outcomes pave the way for additional developments in emotion analysis and its uses in a variety of fields, including social media monitoring, consumer sentiment analysis, and mental health support systems.

# References

Delbrouck, J.-B., Tits, N., Brousmiche, M., & Dupont, S. (2020). *A Transformer-based joint-encoding for Emotion Recognition and Sentiment Analysis*. 1–7. https://doi.org/10.18653/v1/P17

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *NAACL HLT 2019 - 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference*, *1*, 4171–4186. https://arxiv.org/abs/1810.04805v2

Go, A., Bhayani, R., & Huang, L. (n.d.). *Twitter Sentiment Classification using Distant Supervision*. Retrieved June 14, 2023, from http://tinyurl.com/cvvg9a

Haddi, E., Liu, X., & Shi, Y. (2013). The Role of Text Pre-processing in Sentiment Analysis. *Procedia Computer Science*, *17*, 26–32. https://doi.org/10.1016/J.PROCS.2013.05.005

Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., Stoyanov, V., & Allen, P. G. (2019). *RoBERTa: A Robustly Optimized BERT Pretraining Approach*. https://arxiv.org/abs/1907.11692v1

Odbal, Zhang, G., & Ananiadou, S. (2022). Examining and mitigating gender bias in text emotion detection task. *Neurocomputing*, *493*, 422–434. https://doi.org/10.1016/J.NEUCOM.2022.04.057

Torres, J. (2018). *EmotionX-JTML: Detecting emotions with Attention*. 56–60. http://www.

Xu, P., Liu, Z., Indra Winata, G., Lin, Z., & Fung, P. (n.d.). *EmoGraph: Capturing Emotion Correlations using Graph Networks*.