Project Report

Majed Alrogi

1. Introduction and Motivation

In today's digital age, social media is a vital platform for communication and expression, yet it often presents challenges for individuals with disabilities that impair their understanding of emotional context in written text.

My project is motivated by the need to make social media more accessible and inclusive, particularly for users with conditions such as Autism Spectrum Disorder (ASD), which can hinder the interpretation of social nuances.

By implementing semantic analysis using BERT on a Twitter dataset, I aim to solve this problem. My goal is to help users better grasp the emotional undertones of social media posts, thereby enhancing their online interactions and experiences.

3. Data Analysis and Pre-processing

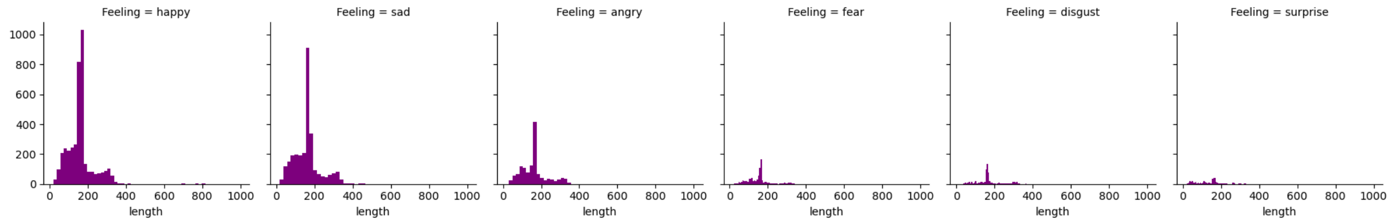
My dataset comprised 9870 unique tweets. The primary goal of pre-processing was to prepare the data in a way that makes it suitable for semantic analysis using BERT.

Before preparing the data, I conducted exploratory data analysis (EDA) to gain insights into my dataset and guide my modeling strategy.

I noticed that the tweet length was roughly about 180 characters for most tweets and no truncation was needed on the dataset. The length was roughly the same across labels.

A notable challenge in my dataset was the imbalance in emotional representations – primarily between Happy and Sad sentiments, compared to Surprise. To address this, I used the idea of class weights to ensure that in training, the weight of each label would be roughly equal. To decide the weights I used sklearn.utils.class_weight.compute_class_weight() to get preferable weights.

Before feeding the data into BERT, the tweets were tokenized, which is a process of converting the continuous text into discrete words or tokens. This step is crucial as BERT operates on token-level input.



BERT:

BERT, which stands for Bidirectional Encoder Representations from Transformers, is a complex model in the field of natural language processing (NLP). Developed by Google, BERT has revolutionized how machines understand human language. Its key mechanism is to process language in a bidirectional manner, meaning it considers the context from both the left and right sides of a target word within a sentence. This is a significant advancement over previous models, which typically processed text in a linear, unidirectional way.

The core architecture of BERT is based on the Transformer, which uses attention mechanisms to understand the relationships between different words in a sentence. Unlike traditional sequential models, such as LSTM (Long Short-Term Memory), BERT can capture more complex contextual relationships in text. This makes it highly effective for tasks like sentiment analysis, question-answering, and, crucially for my project, interpreting emotional nuances in social media posts.

In my project, BERT's capability to discern emotions is leveraged to enhance the accessibility of social media for users with disabilities. By training BERT on a dataset of Twitter posts, the model learns to identify and interpret various emotional tones embedded in short, often informal, messages. This training enables BERT to predict the emotional context of unseen tweets, with the goal of assisting users in understanding the underlying sentiments of social media content.

The reason I picked BERT for my project is its pre-training on a vast amount of data, including all of Wikipedia. This pre-training allows BERT to develop a deep understanding of language structure and common patterns. This allows me to finetune BERT to allow it to use its already learned knowledge and apply it to semantic analysis

There are two versions of BERT, BERT base and BERT large. They differ in the amount of parameters they have. BERT base has 110 million parameters while large has 340 million parameters. I did not require too many parameters as my task was relatively straight forward.

7. Evaluation Metrics

Evaluating the performance of my BERT-based semantic analysis model is critical to ensure its effectiveness in interpreting emotions in tweets. To measure this performance, I employed several key evaluation metrics, each offering a unique perspective on the model's capabilities.

Accuracy: This is the most straightforward metric, representing the proportion of correctly predicted tweets out of all tweets in the test set. Accuracy gives a general idea of the model's overall performance but can be misleading in the context of imbalanced datasets.

Precision: Precision is particularly important in my context as it measures the proportion of correctly predicted positive observations (for each emotion) to the total predicted positive observations. High precision implies a low rate of false positives, which is crucial in avoiding the misinterpretation of a tweet's emotional tone.

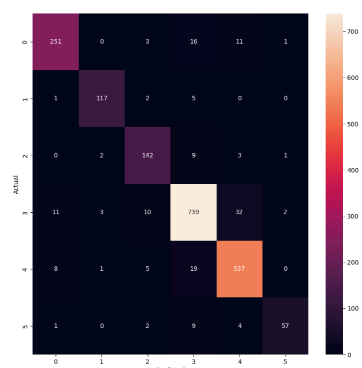
Recall (Sensitivity): Recall assesses the model's ability to correctly identify actual positives from the dataset. For my project, this means the model's ability to correctly detect the presence of a specific emotion in a tweet. A high recall is essential for ensuring that the model does not overlook subtle emotional cues.

F1 Score: The F1 Score is a balanced measure that considers both precision and recall. It is the harmonic mean of precision and recall, providing a single metric that speaks to the balance between the two. Given the potential trade-offs between precision and recall, the F1 Score is a critical metric for evaluating my model.

Confusion Matrix: Beyond these numerical metrics, I also utilized a confusion matrix for each emotion category. This matrix provides a visual and quantitative way of understanding not just the successes of the model, but also its errors, such as confusing sadness with anger or happiness with surprise.

Scores can be found below:

```
Accuracy: 0.9196606786427146
Recall: 0.9196606786427146
Precision: 0.9200252296562192
F1 Score: 0.9194570385971799
```



There are multiple avenues that I could expand upon if I were able to work on it further. One step could be to expand the dataset to get a more comprehensive representation of twitter. This could help BERT generalize better.

Another step that could increase accessibility is incorporating more languages like Arabic. This would make the model accessible to a huge proportion of people.

I would also like to increase the amount of emotions to capture more nuanced emotions like sarcasm and anxiety. This would allow users to understand a wider variety of emotions.

Overall I really enjoyed this project and it helped me learn more about NLP and transformers which was a topic that I wanted to learn more about after the brief introduction in class.