

**Natural Language Processing**

**DSCI 6004**

**Final Project on**

**Emotion Classification in Twitter Messages Using Pre trained Models**

**Done by**

**Krishna sai Bhiravajosyula**[**(kbhir2@unh.newhaven.edu)**](mailto:kbhir2@unh.newhaven.edu)

**Contents**

**Abstract:**……………………………………………………………………………………3

**Introduction:**……………………………………………………………………………….3

**Related Work:**……………………………………………………………………………...4

**Methods:**…………………………………………………………………………………….4

Data cleaning…………………………………………………………………………...5

Exploratory data analysis……………………………………………………………….5

Model sampling and splitting…………………………………………………………...5

Model Architecture……………………………………………………………………..6

Model training and evaluation for Distilbert and gpt-2…………………………………7

**Results:**……………………………………………………………………………………….8

Model Performance………………………………………………………………………8

**Discussion:**…………………………………………………………………………………..13

**Conclusion:**………………………………………………………………………………….13

**References:**………………………………………………………………………………….13

**Abstract:**

This report focuses on emotional classification of twitter messages using natural

language processing techniques to classify emotions of text into six categories such as love,

anger, sadness, joy, fear and Surprise. With the help of pre-trained models from hugging face

libraries like GPT-2, DistilBert, we can effectively develop a classification model to tailor

human emotions for business and social media platforms. This method helps attain to study

human emotions and can help leverage practical real-world applications such as customer

improvement and segmentation and support mental health behavior. We used evaluation

metrics such as Accuracy, F1-score, Precision and Recall evaluating the performance of two

models to compare which model is more opted for classification problems.

**Introduction**

As social media dominates the web because of interactive connection among people

to share their feelings, it has generated lot of text data in the form of posts, captions and

content. This text has intense emotions associated with it from various people. This study

focuses on people emotions into six most common categories to analyze and tailor the

services of twitter platform according to interests of people. The goal of this project was to

develop a classification model for people to answer the emotional state using advanced

transformers architecture models like gpt2 and Distilbert which are trained on large corpus of

data with millions of parameters and fine tuned to specific tasks including classification often

the most crucial task of Natural language Processing.

These models are well suited to find similarity in context and can understand the text

data with great accuracy and efficiency. The model GPT-2 is trained on 1.5 billion

parameters for performing tasks like text generation, machine translation and text

classification. Model like DistilBert is trained on 66 million parameters is flexible and

lightweight easy to integrate with applications specifically fine-tuned for sentiment analysis

and text classification. The text is Pre-Processed, and models are imported from hugging face

converted to Dataset to perform prediction on text data using evaluation metrics along with

hyperparameter values and then are validated on unseen text data to compare their predictions

with actual labels.

**Related Work:**

Emotion classification has always been the most crucial research in extensive fields

like medical and retail. Previously the classification models relied on machine learning

algorithms like support vector machine and naïve bayes, as mentioned in Pang and lee

(2008) research paper. These algorithm approaches would include manual feature

engineering tasks and struggled with characteristic expressive data. Then the introduction of

word embeddings like Word2Vec published in Mikelov et al (2013) paper showed great

improvement in capturing sematic relationships between words and their meanings.

The introduction of transformers models revolutionized text data for every task

including text classification, text generation and translation. BERT, introduced by Devlin et

al in 2019 showed potential strength of bi-directional transformers in capturing complex

meanings of text data. One variant of Bert model which is DistilBert a lightweight version of

Bert has gained significance because of its flexible architecture and efficiency which made it

Prominent for real life applications.

While many studies showed a gap between generative models like GPT2 and bert

models for multi class emotional classification. This project aims to close in the gap between

two models when worked on same data by capturing metrics like accuracy and F1 score to be

most crucial metrics in classification tasks. This project also focuses on various

hyperparameters to optimize and evaluate the model for specific implications.

**Methods**

**Data Cleaning**

The data set used in this project is named emotional classification extracted from

Kaggle website which is a pool of dataset for machine learning and deep learning tasks.

The dataset has three columns named ‘unnamed’ , ‘text’ which has information of the

user’s prompt and ‘label’ ranging from 0 to 5 for all six emotions which are ‘sadness (0)’,

‘joy (1)’, ‘love (2)’, ‘anger (3)’, ‘fear (4)’, ‘surprise (5)’. The first step is to ensure drop

irrelevant column named ‘unnamed’ from my dataset.



**Exploratory data Analysis**

We have plotted a bar chart to compare the labels in our dataset. This analysis shows the

distribution of categories along the dataset to prevent any bias.

A graph of blue bars

Description automatically generated with medium confidence

This bar chart illustrates that labels named sadness and joy are more than the other named

labels. Since we are using twitter social media platform dataset sadness and joy are most common

emotions in the social media posted by the users.

**Model sampling and splitting**

The dataset is sampled to load 30,000 rows and split into 70 percent as training and 30

percent as validation data. This uses sci-kit learn library importing train\_test\_split from

model\_selection.

A screenshot of a computer program

Description automatically generated

**Model Architecture**

**Two architectures are used to compare performance**

1. **DistilBert:** This is a smaller efficient version of BERT and it uses a bi-directional attention

mechanism. It considers both left and right words in text making it highly efficient at

understanding deep contexts of words in a text. The model is imported using transformers

library and loaded with model-name which is name of the model and token which is Distilbert

tokenizer for text tokenization and model which is distilbertforsequenceclassification to

classify the text in our dataset along with training arguments to train the model with .train()

function.

**A computer code with black text

Description automatically generated with medium confidence**

**2.) GPT-2:** This is also transformers-based model developed by open AI for text generation. It

uses its adaptive knowledge of pre trained datasets to classify emotion labelled dataset. Unlike bert

model, this is a large language model trained with billions of parameters. The model classifies text

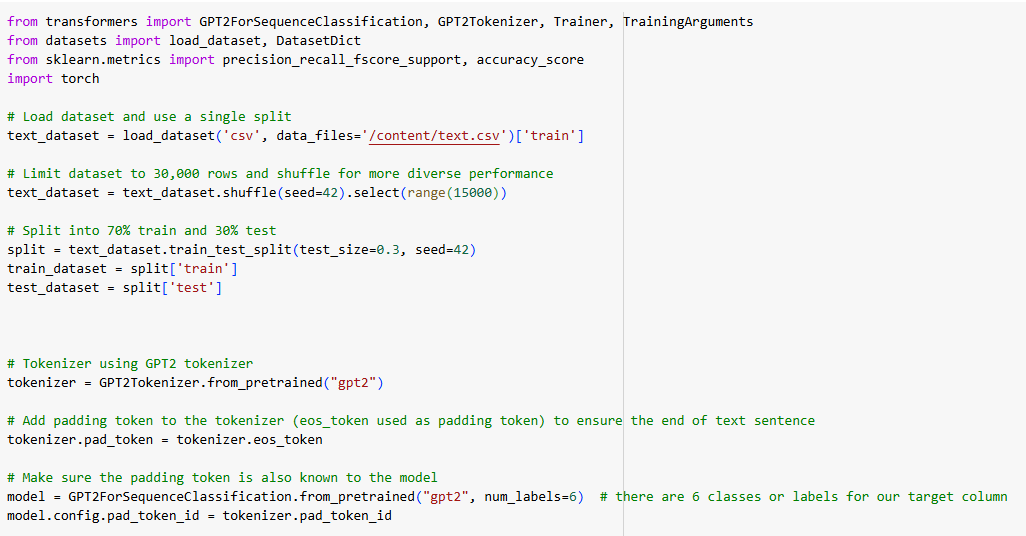
data by framing every word as text-to-text transformation based on user’s input and it generates

output by understanding complex relationships between words and context. The model is imported

using transformers library along with tokenizer and model for tokenizing text and model name for

sequence classification. only difference between both models is we also import pad-token to define

the maximum length of sentence and assign an id to pad-token.



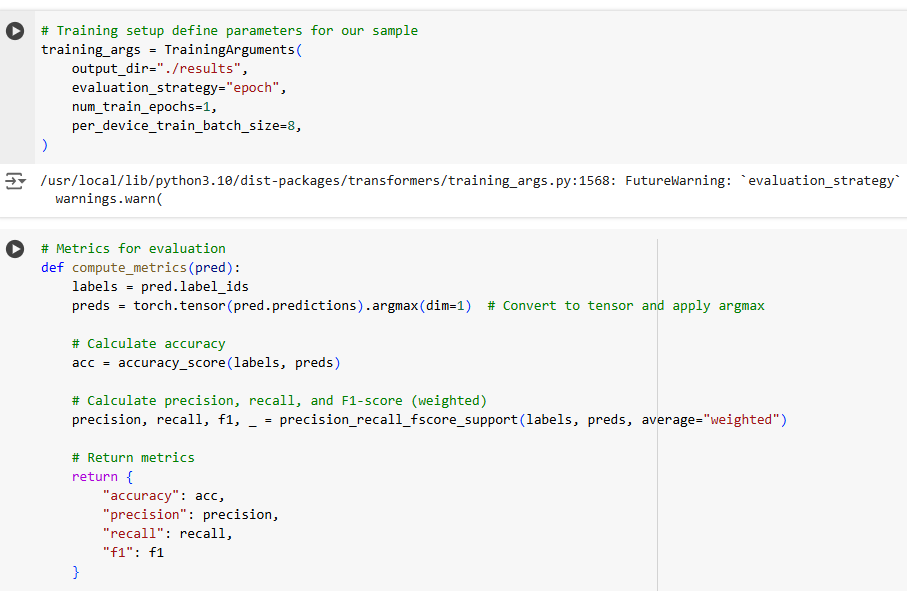
**Model Training and evaluation for both Distilbert and GPT-2**

The model first converts text into tokens and converts tokens into tensors using a function

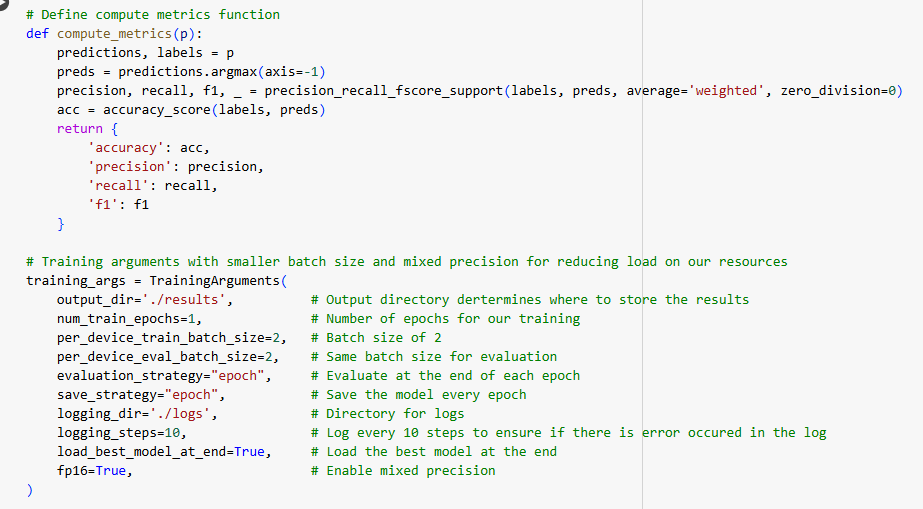
named tokenize function. Then followed we define training arguments which has info on the output to

be stored in result directory, defining batch size and number of epochs to iterate the model.

**Distilbert**



**Gpt-2**



**Results**

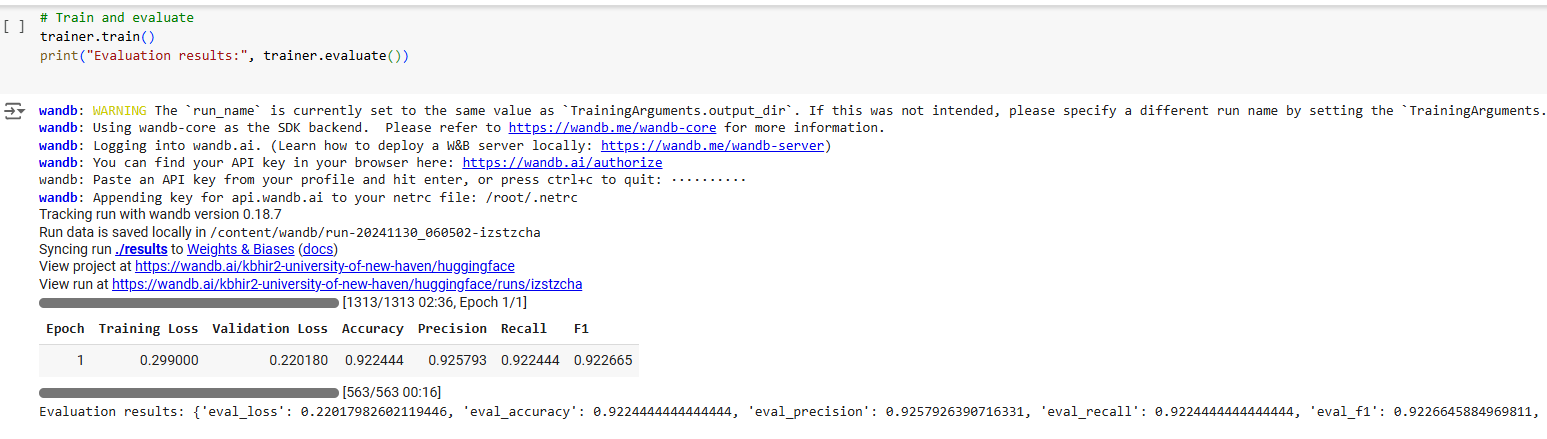
**Model performance**

**Bert Model:** As the model is trained on our dataset it showed a promising result of 92% accuracy

effectively at classifying the dataset into emotions based on their labels. Here are the training results

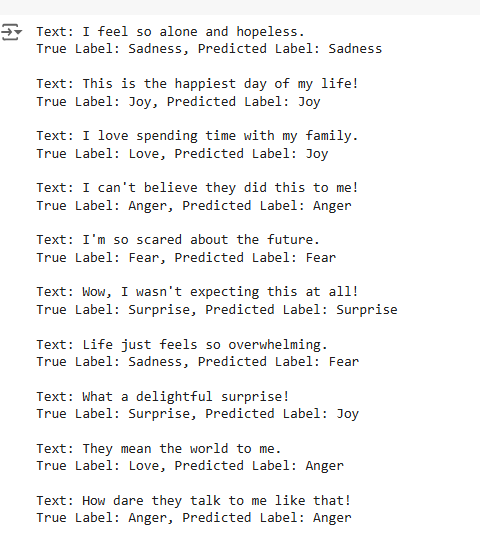
of our model. We took considerations regarding accuracy, f1-score, precision, recall as our metrics

since this is a classification problem.



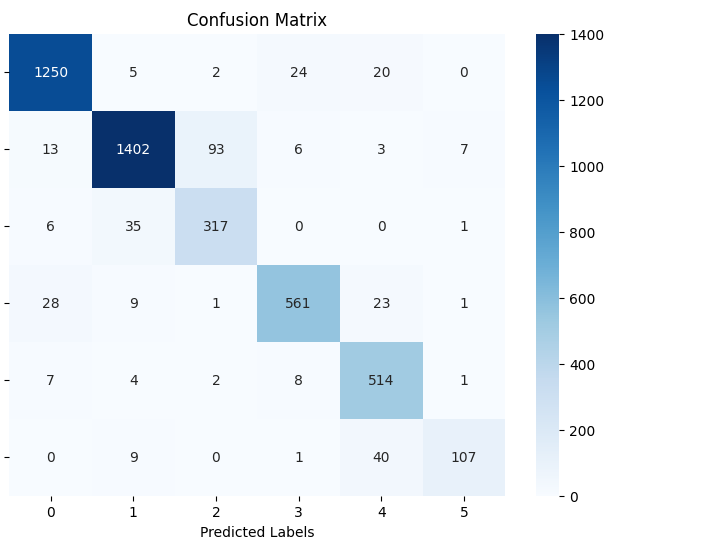
The model is also made to predict unseen data performing well. Here are the results of our

model on unseen data.



**Confusion Matrix for DistilBert:** The confusion matrix is crucial for finetuning as it depicts all

correct and incorrect predictions made by the model.



The visual shows

1. **True Positives (TP):**

  - **Sadness (0):**1250

   - **Joy (1):** 1402

   - **Love (2):**317

   - **Anger (3):**561

   - **Fear (4):** 514

   - **Surprise (5):**107

These are the cases the model correctly predicted the emotion.

**False Positives (FP):**

   - **Sadness (0):** 51

   - **Joy (1):** 122

   - **Love (2):**42

   - **Anger (3):**62

   - **Fear (4):**22

   - **Surprise (5):**50

These are the cases the model incorrectly predicted the emotion.

**False Negatives (FN):**

   - **Sadness (0):**54

   - **Joy (1):** 62

   - **Love (2):** 322

   - **Anger (3):**594

   - **Fear (4):** 552

   - **Surprise (5):**110

   These are the cases the model missed the emotion, predicting wrong label.

**True Negatives (TN):**

   - **Sadness (0):**4839

   - **Joy (1):** 4608

   - **Love (2):** 5513

   - **Anger (3):** 4977

   - **Fear (4):** 5106

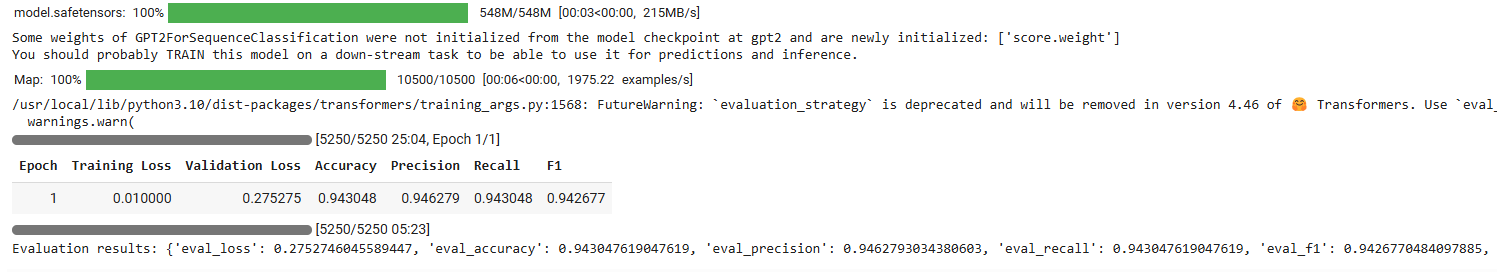
   - **Surprise (5):** 5897

These are the cases the model correctly predicted that the sample did not belong to the given emotion.

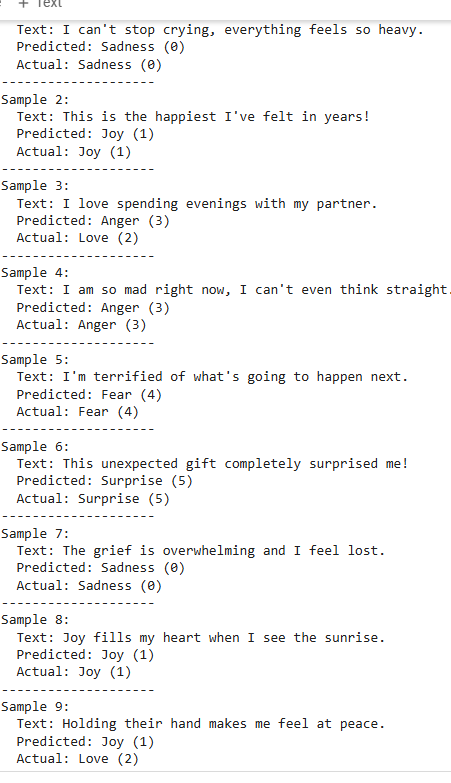
**GPT-2 Model:** The model uses gpt2sequenceclassification token to classify the text based on the

labels. The model performed slightly better accuracy of 94% since it is trained on 10 to 15 times more

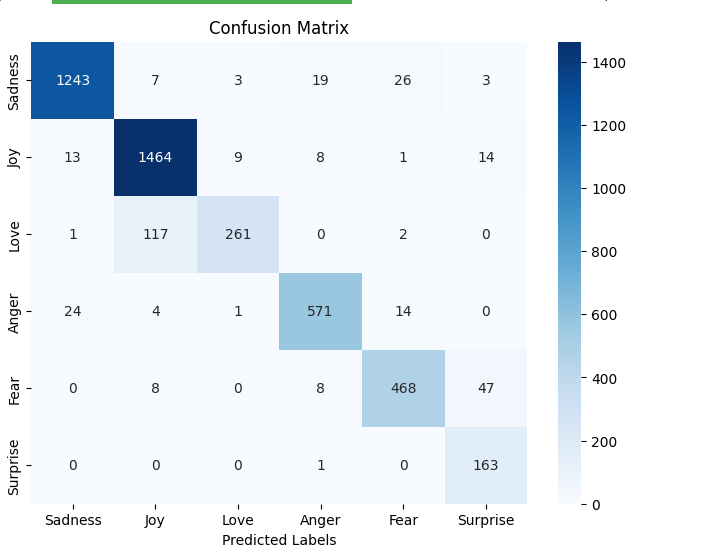
parameters than distilbert model.



Here are the results on unseen data.



**Confusion matrix for GPT-2**



The values for **True Positives (TP)**, **False Positives (FP)**, **False Negatives (FN)**, and **True Negatives (TN)** based on the confusion matrix:

**True Positives (TP):**

**Sadness (0):** 1243

**Joy (1):** 1464

**Love (2):** 261

**Anger (3):** 571

**Fear (4):** 468

**Surprise (5):** 163

These are the cases the model correctly predicted the emotion.

**False Positives (FP):**

**Sadness (0):** 13

**Joy (1):** 9

**Love (2):** 0

**Anger (3):** 14

**Fear (4):** 8

**Surprise (5):** 1

These are the cases the model incorrectly predicted the emotion.

**False Negatives (FN):**

**Sadness (0):** 24

**Joy (1):** 8

**Love (2):** 2

**Anger (3):** 0

**Fear (4):** 8

**Surprise (5):** 0

These are the cases the model missed the emotion, predicting wrong label.

**True Negatives (TN):**

**Sadness (0):** 2877

**Joy (1):** 2676

**Love (2):** 3894

**Anger (3):** 3572

**Fear (4):** 3673

**Surprise (5):** 3993

These are the cases the model correctly predicted that the sample did not belong to the given emotion.

**Discussion**

The performance of two models shows little difference of about 2% in their metrics even on

Complex text because of their size. The distilbert showed promising results as it took less time to

evaluate and flexible in using low resources can be deployed for real life applications.

While GPT-2 took more to load and training process as it has transformed data word by word

to understand the context and categorize the emotions in the text but the development of gpt-2

finetuned model gpt2sequenceclassifier showed that pre-trained knowledge on large corpus of data

can be useful for classification tasks. The incorrect predictions were significantly less than the

Distilbert model as seen in confusion matrix as it designed for multipurpose tasks.

**Conclusion**

This project aims to build an emotion classificational model by utilizing two advanced

transformers architecture models Distilbert, gpt-2. The results of the project showed that gpt-2 is

superior as it can more complex data while distilbert showed its strength in its flexibility and

efficiency as it utilizes less time with compromising accuracy.

The two models effectively showed our purpose to use them in real life applications such as

social media, health monitoring applications and retail applications for personalized interest to users

improving business growth.

**References**

Pang, B., & Lee, L. (2008). opinion mining and sentiment analysis. Foundations and Trends in Information Retrieval.

Mikolov, T., Yih, W. T., & Zweig, G. (2013). Distribution representation of words and phrases and their compositionality. Proceedings of NIPS 2013.

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: pre training of deep bi-directional transformers for language understanding. Proceedings of NAACL 2019.

**GitHub Link**

[**https://github.com/Manaskarthik28/NLP\_final\_project**](https://github.com/Manaskarthik28/NLP_final_project)