Unveiling Mental Health Insights through Social Media Analysis

DATS 6312 – Natural Language Processing
Udbhav Kush

Table of Contents

Introduction

In 2020, the World Health Organization (WHO) cast a stark light on the global prevalence of mental disorders, revealing that nearly 1 billion people, constituting just over one in ten of the world's population, grapple with the complexities of mental health. Closer to home, the United States witnessed a striking statistic in 2021, where more than one in five adults—equivalent to a staggering 57.8 million individuals—were reported to be living with a mental health condition. Amidst these staggering numbers, a profound societal shift has unfolded, as a significant majority now turns to social media as their primary conduit for communication. This transformation positions social media platforms not only as hubs for connection but also as unique arenas for gauging and understanding the subtle nuances of mental health indicators in an increasingly interconnected world. As our digital landscapes evolve, the intersection between mental health and social media becomes an ever more critical frontier for exploration and analysis.

Data

Dataset Used

Upon formulating our problem statement, we embarked on a comprehensive literature review to investigate the various solutions that have been applied to address this challenge. During our exploration, we encountered a research paper available that aligns closely with our specified requirements. Recognizing the compatibility of their data with our research objectives, we opted to leverage the data presented in this research paper as a valuable foundation for our own investigation.

The tweets were classified into 10 different classes based on mental health disorders. The classes were: EATING DISORDER, SCHIZOPHRENIA, OCD, PTSD, ANXIETY, BIPOLAR, AUTISM, DEPRESSION, ADHD, and CONTROL.
The Control Group was the non-diagnosed group of users.

Splits

We did a standard split of training-test and dev. The training set was used for training the model and the test set was used to make decisions while training the model. The dev set was untouched and only evaluated right before the project presentation.

My Contributions

Training Script

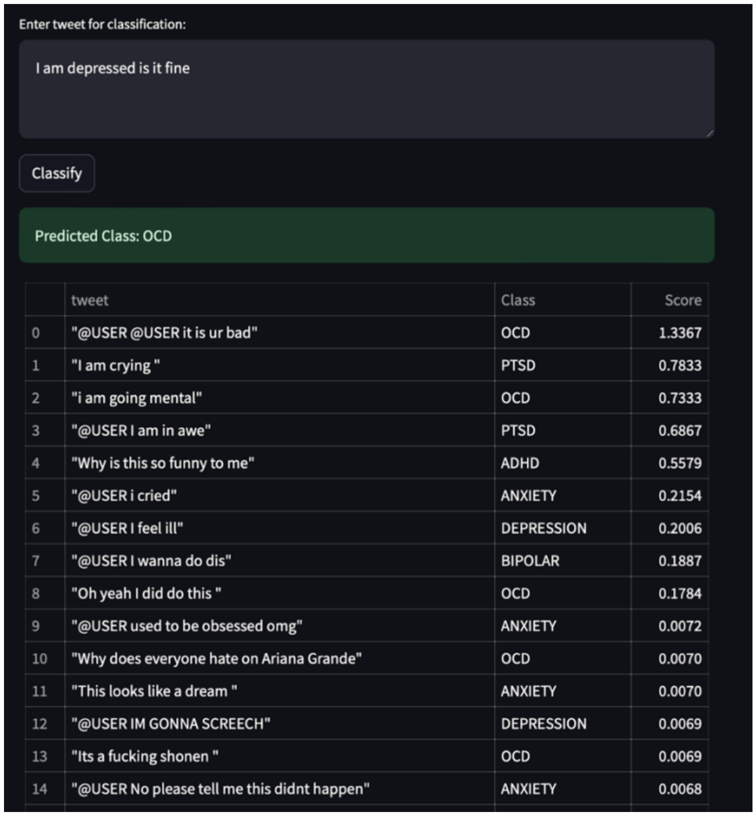
I was responsible for creating the general training script that I adapted from the Deep Learning Course. This was done from cross-compatibility of the model and made debugging easy for all the teammates.

Train DistillBert on the Spanish Dataset

I was responsible for training the DistillBert model on the Spanish tweets dataset. Initially, we tried the "bert-base-multilingual-cased" model to train the whole dataset of English and Spanish together. However, the model did not give good results on the initial few epochs and the training time was very high as we were training the complete dataset.

I used a finetuned distillbert model "dccuchile/distilbert-base-spanish-uncased-finetuned-xnli" to train the Spanish tweet dataset.

App Using Streamlit



Enter tweet for classification:

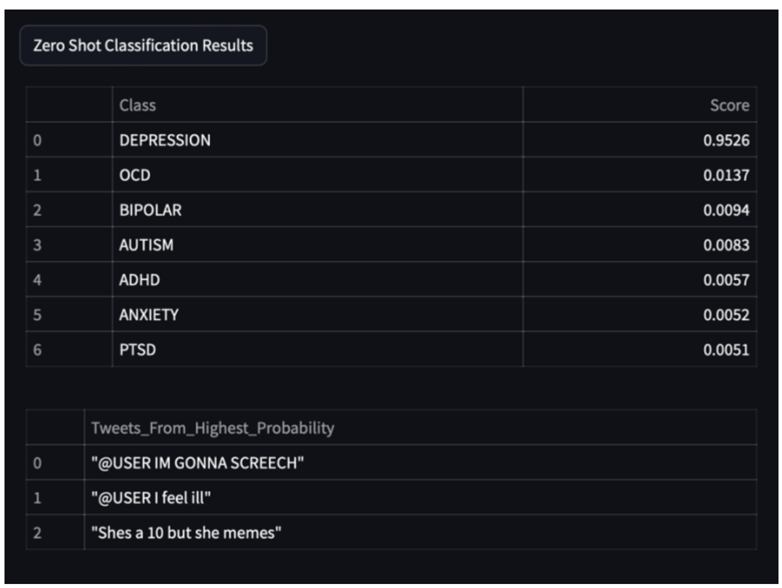
I am depressed is it fine

Classify

Predicted Class: OCD

	tweet	Class	Score
0	"@USER @USER it is ur bad"	OCD	1.3367
1	"I am crying "	PTSD	0.7833
2	"i am going mental"	OCD	0.7333
3	"@USER I am in awe"	PTSD	0.6867
4	"Why is this so funny to me"	ADHD	0.5579
5	"@USER i cried"	ANXIETY	0.2154
6	"@USER I feel ill"	DEPRESSION	0.2006
7	"@USER I wanna do dis"	BIPOLAR	0.1887
8	"Oh yeah I did do this "	OCD	0.1784
9	"@USER used to be obsessed omg"	ANXIETY	0.0072
10	"Why does everyone hate on Ariana Grande"	OCD	0.0070
11	"This looks like a dream "	ANXIETY	0.0070
12	"@USER IM GONNA SCREECH"	DEPRESSION	0.0069
13	"Its a fucking shonen "	OCD	0.0069
14	"@USER No please tell me this didnt happen"	ANXIETY	0.0068

Figure 1. Classification of Mental Disorder



Zero Shot Classification Results

	Class	Score
0	DEPRESSION	0.9526
1	OCD	0.0137
2	BIPOLAR	0.0094
3	AUTISM	0.0083
4	ADHD	0.0057
5	ANXIETY	0.0052
6	PTSD	0.0051

	Tweets_From_Highest_Probability
0	"@USER IM GONNA SCREECH"
1	"@USER I feel ill"
2	"Shes a 10 but she memes"

Figure 2. Zero-Shot Classification Results

I also tried creating a psychotherapist counselor chatbot for further assistance to the user. However, the size of the chatbot was too high and I could not integrate it well with the Streamlit app.

RESULTS

Model	Accuracy	F1Score
dccuchile/distilbert-base-spanish-uncased-finetuned-xnli	0.61	0.59

The model did not perform very well because the Twitter data is not proctored. So, some tweets from all the classes seemed similar.

CONCLUSION

Our project cannot successfully classify English language and Spanish language tweets using the BERT-based classification. Reason is similar context in different classes.

FURTHER IMPROVEMENTS

- The development of a mental health assistance chatbot to provide necessary support for the user proved to be challenging for us. However, we plan to try to incorporate this in the future.

- We can also try implementing the langdetect model to detect the language in which a tweet was posted and classify it accordingly.

- We could try langdetect model to detect the language in which the tweet was posted and then use the classification model accordingly.

REFERENCES

- Dccuchile/distilbert-base-Spanish-uncased-finetuned-xnli at main. dccuchile/distilbert-base-spanish-uncased-finetuned-xnli at main. (n.d.). https://huggingface.co/dccuchile/distilbert-base-spanish-uncased-finetuned-xnli/tree/main

- Dccuchile/distilbert-base-Spanish-uncased-finetuned-xnli at main. dccuchile/distilbert-base-spanish-uncased-finetuned-xnli at main. (n.d.). https://huggingface.co/dccuchile/distilbert-base-spanish-uncased-finetuned-xnli/tree/main

- M. E. Villa-Pérez, L. A. Trejo, M. B. Moin and E. Stroulia, "Extracting Mental Health Indicators from English and Spanish Social Media: A Machine Learning Approach," in IEEE Access, doi: 10.1109/ACCESS.2023.3332289.

Percentage of code copied from the internet: Around 30% of code copied from the internet.