LAMA at SMM4H 2024: Experimenting with Transformer-based and Large Language Models for Classifying Effects of Outdoor Spaces on Social Anxiety in Social Media Data

Falwah AlHamed
Department of Computing,
Imperial College London
London, UK
f.alhamed20@imperial.ac.uk

Julia Ive
Queen Mary
University of London
London, UK
j.ive@qmul.ac.uk

Lucia Specia
Department of Computing,
Imperial College London
London, UK
l.specia@imperial.ac.uk

Abstract

Social Anxiety Disorder (SAD) is a common condition, affecting a significant portion of the population. While research suggests spending time in nature can alleviate anxiety, the specific impact on SAD remains unclear. This study explores the relationship between discussions of outdoor spaces and social anxiety on social media. We leverage transformer-based and large language models (LLMs) to analyze a social media dataset focused on SAD. We developed three methods for the task of predicting the effects of outdoor spaces on SAD in social media. A two-stage pipeline classifier achieved the best performance of our submissions with results exceeding baseline performance.

1 Introduction

Social anxiety disorder (SAD) is a prevalent anxiety disorder that affects up to 12% of the population at some point in their lives (Kessler et al., 2005). Interestingly, social media platforms like Reddit have become a space for people with SAD to connect, share their experiences, and seek advice on managing symptoms. While research suggests that spending time outdoors in natural environments can be beneficial for alleviating anxiety in general (Barton and Pretty, 2010; Berman et al., 2008), little is known about the specific impact of such environments on SAD. This study investigates the effect of outdoor spaces on social anxiety using social media posts as a source of data.

2 Data and Task Description

This work leverages the dataset provided by the organizers of SMM4H (Xu et al., 2024). The data originates from the r/socialanxiety subreddit on Reddit and consists of 3,000 annotated social media posts. The training set comprises 1,800 posts, each labelled with a code based on the user's sentiment towards the mentioned nature-related spaces or activities. A four-class labelling scheme was employed: **positive**: the space/activity benefits the user's well-being; **neutral**: mention of nature without a clear impact; **negative**: the space/activity has a negative impact; and **unrelated**: posts where the nature keyword is metaphorical or unclear in meaning. The task is four-class classification on these labels.

3 Methods

3.1 Models Exploration

We address this multi-class classification task by leveraging several transformer-based and LLMs on the provided training dataset. While the task definition specifies four classes, there is natural division into two broader categories: Related and Unrelated. The Related category encompasses posts where the mentioned outdoor space is relevant to the user's experience, and the effect can be positive, neutral, or negative. Conversely, the Unrelated category consists solely of posts where the outdoor space reference is metaphorical and has no bearing on the user's experience. We performed experiments utilizing five models: BERT, RoBERTa, MentalBERT, GPT3.5, and LlamA. For each model, we conducted three distinct experiments:

Experiment 1: Binary Classification: The model classifies posts into: Related and Unrelated classes.

Experiment 2: Multi-Class Classification within Related class: The model focuses solely on "related" posts and further categorizes them into three classes: positive, neutral, and negative.

Experiment 3: Four-Class Classification: This model classifies posts into all classes (positive, neutral, negative, and unrelated).

The primary objective of this exploration was to identify the optimal model for each experiment, which would then be used in subsequent stages of our submission models design. For transformer-based models, we utilized the Hugging Face transformers library (Wolf et al., 2019) with

Proceedings of The 9th Social Media Mining for Health Applications (#SMM4H) Workshop and Shared Tasks, pages 95–97
August 15, 2024

Experiment	BERT			RoBERTa			Mental BERT			Llama			GPT		
	P	R	F1	P	R	F1	P	R	F1	P	R	F1	P	R	F1
Experiment 1	0.97	0.97	0.97	0.98	0.98	0.98	**0.99**	**0.98**	**0.98**	Random Labels			0.70	0.60	0.58
Experiment 2	0.66	0.67	0.62	**0.71**	**0.66**	**0.68**	0.83	0.54	0.53	N/A					
Experiment 3	0.73	0.77	0.73	**0.82**	**0.71**	**0.73**	0.63	0.65	0.64	N/A					

Table 1: Results of best model exploration experiments (bold fonts signify the best results)

Submission	F1 Score	Precision	Recall	Accuracy
Our RoBERTa 4-classes	0.543	0.607	0.534	**0.651**
Our 2-Stages Pipeline	**0.545**	**0.633**	**0.536**	0.636
Our Multi-task Learning	0.321	0.346	0.326	0.448
All Teams Mean	0.519	0.565	0.538	0.575

Table 2: Results of submitted systems (bold fonts signify the best results)

model cards (`bert-base-uncased`, `roberta-base`, and `mental/mental-bert-base-uncased`. These models received only the post text as input and were trained to predict a pre-defined class label depending on the experiment. LLMs employed a zero-shot learning approach, where prompts incorporating both the post text and relevant keywords were fed to the models. We experimented with various prompts and the best performing prompt is " `Do you think the person is really [keyword] in this paragraph or just a metaphor? return 1 if you think they are doing this otherwise, return 0 Text:[post_text]`.

3.2 Methods Description

Based on experiment results, we selected the top-performing models for each task and employed them to develop three submission methods:

Four-Class Classification: This method utilizes the chosen model from Experiment 3 to classify posts into all four original classes (positive, neutral, negative, and unrelated).

Two-Stage Pipeline: This approach leverages two models in a sequential pipeline. The first stage employs the model selected from Experiment 1 for the binary classification of posts into Related and Unrelated. Subsequently, posts classified as "Related" are passed to the model chosen from Experiment 2, which performs a three-class classification into positive, neutral, and negative.

Multi-Task Learning: This method employs a multi-task learning model trained on two tasks simultaneously. The first task is a binary classification of all (posts) into Related and Unrelated categories. The second task is a multi-class classification focusing solely on posts classified as "Related" in the first task, categorizing them into positive, neutral, and negative.

4 Results

The evaluation metric for this task is the macro-averaged F1-score over all 4 classes. Table 1 summarizes the performance of the models across various experiments, evaluated using precision, recall, and F1-score metrics. MentalBERT emerged as the most effective model for classifying posts into "Related" and "Unrelated" categories. On the other hand, RoBERTa achieved the best performance in both classifying "Related" posts into positive, neutral, and negative classes and in the four-class classification task. LLMs performance in classifying posts as related/unrelated was disappointing, thus it was discarded from the remaining experiments. Based on these results, we selected RoBERTa for the final submission employing the four-class classification approach. Additionally, the pipeline approach for the submission utilized MentalBERT for the initial binary classification of posts into "Related" and "Unrelated" categories and RoBERTa model for the following three-class classification into positive, neutral, and negative. RoBERTa was also employed in the multi-task approach. Results of our submission provided by task organizers is shown in Table 2.

5 Conclusion

This study investigated the connection between social anxiety and outdoor spaces on social media. We employed transformer-based and LLMs classifiers on the provided dataset. Our results exceeded the baseline, and the pipeline approach achieved the highest performance as evaluated by the task organizers.

References

J Barton and J Pretty. 2010. Tenacious nature: A review of the evidence for greater well-being within natural environments. *Journal of Environmental Psychology*, 30(4):380–390.

Marc G Berman, John Jonides, and Stephen Kaplan. 2008. The cognitive benefits of interacting with nature. *Proceedings of the National Academy of Sciences*, 105(48):19127–19132.

Ronald C Kessler, Pamela Berglund, Olga Demler, and Elizabeth E Walters. 2005. Prevalence, severity, and comorbidity of twelve-month dsm-iv disorders in the national comorbidity survey replication. *Archives of General Psychiatry*, 62(6):617–627.

Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Rémi Louf, Morgan Funtowicz, et al. 2019. Huggingface's transformers: State-of-the-art natural language processing. *arXiv preprint arXiv:1910.03771*.

Dongfang Xu, Guillermo Lopez Garcia, Lisa Raithel, Rolland Roller, Philippe Thomas, Eiji Aramaki, Shuntaro Yada, Pierre Zweigenbaum, Sai Tharuni Samineni, Karen O'Connor, Yao Ge, Sudeshna Das, Abeed Sarker, Ari Klein, Lucia Schmidt, Vishakha Sharma, Raul Rodriguez-Esteban, Juan Banda, Ivan Flores Amaro, Davy Weissenbacher, and Graciela Gonzalez-Hernandez. 2024. Overview of the 9th social media mining for health applications (#SMM4H) shared tasks at ACL 2024. In *Proceedings of The 9th Social Media Mining for Health Research and Applications Workshop and Shared Tasks*, Bangkok, Thailand. Association for Computational Linguistics.