

Classifying Social Media Users Before And After Depression Diagnosis Via Their Language Usage: A Dataset And Study

The 2024 Joint International Conference on Computational Linguistics, Language
Resources and Evaluation (LREC-COLING 2024)

Falwah Alhamed, Julia Ive, Lucia Specia

Department of Computing, Imperial College London



Introduction

- According to the NHS, one in four people experiences mental health problems
- Mental illness can significantly impact individuals' quality of life
- Users on social media often share updates about their daily lives, including moods and feelings
- Most studies focus on the classification of users suffering from depression versus healthy users, or on the detection of suicidal thoughts
- In this paper, we aim to understand and model linguistic changes that occur when users transition from a healthy to an unhealthy state.



Contribution

- The first English dataset of textual posts by **the same users** before and after reportedly being diagnosed with depression
- A lexicon for finding posts with symptoms of depression
- Empirical work comparing multiple predictive models (based on SVM, Random Forests, BERT, RoBERTa, GPT-3, GPT-3.5, Bard, and Alpaca) built using our dataset for the task of classifying user posts as before and after depression.



+

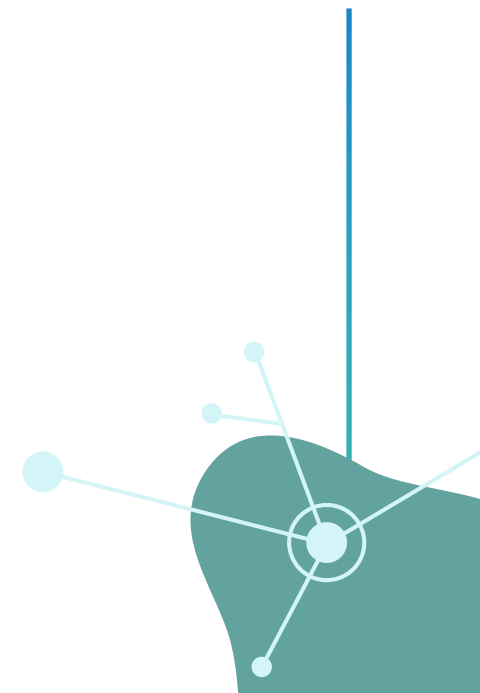
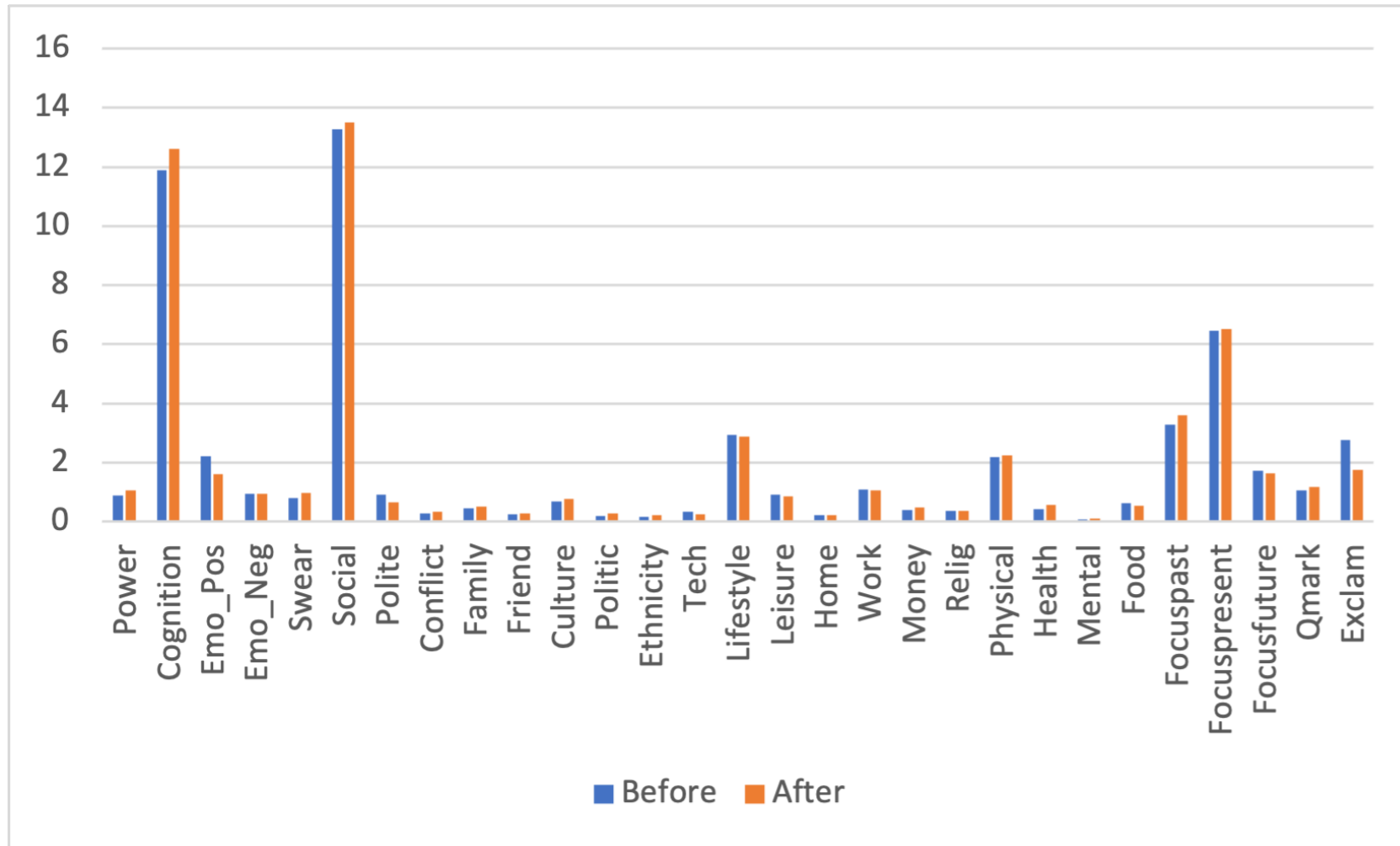
Dataset Collection

○

- First, we collected tweets with the search words ‘I was diagnosed with depression on [specific month and year]’
 - 2034 posts (tweets) were pulled and manually inspected
- Only tweets for users who mentioned a specific month and year of diagnosis were selected
 - 120 users.
- Then for each user, we retrieved posts for three years before the diagnosis date and three years after the diagnosis date
 - 1.9 million posts (tweets) [Final Dataset]

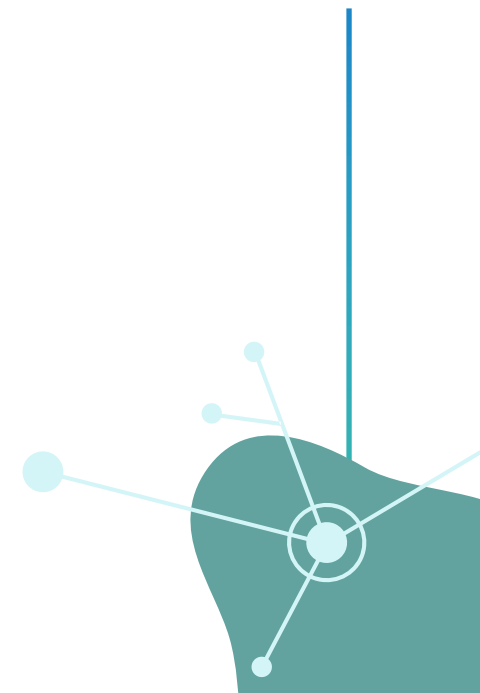


Data Analysis - LIWC



Data Analysis - POS

POS	Before Diagnosis	After Diagnosis
Noun, common (NN)	19.24%	15.8%
Noun, plural (NNS)	2.59%	3.04%
Noun, proper (NNP)	1.32%	1.77%
Verb, base form (VB)	3.14%	3.85%
Verb, present tense (VBZ)	2.11%	2.41%
Verb, past tense (VBD)	1.89%	2.25%
Adverb (RB)	3.91%	4.61%
Determiner (DT)	4.86%	5.83%
Pronoun (PRP)	3.73%	4.49%
Adjective (JJ)	8.24%	8.04%
Preposition (IN)	5.62%	6.90%



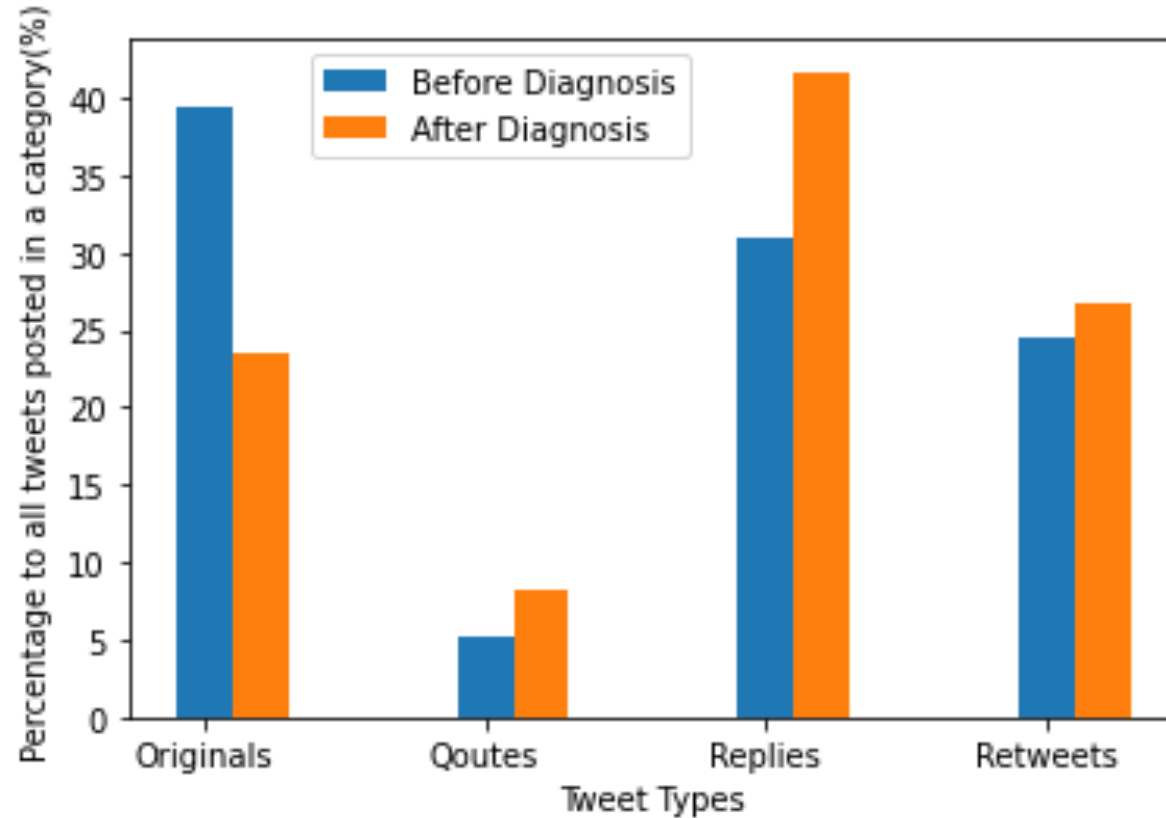
Data Analysis - Posting Frequency

Original: Posts written by the user in his/her timeline.

Quoted: Posts are originally written by someone else and the user quotes this tweet and adds a comment then posts it in his/her timeline.

Replied: Posts are posted by someone and the user replies to this specific tweet.

Retweeted: Posts are originally written by someone else and the user reposts it in his/her timeline.



Lexicon Building and Data Filtering

- We created a lexicon that can be used to screen posts for depression symptoms
- The lexicon contains 598 words related to depression symptoms according to the Center for Epidemiologic Studies Depression (CES- D) validated scale
- We started by building up a lexicon from existing words related to the depression scale questions and symptoms which can be categorized into: poor appetite and eating disturbance, feeling down and depressed, concentration problems, feeling tired or having little energy, sleep disturbance, loss of interest, self-blame and shame, loneliness, and suicidal thoughts. For each category, we created a list of relevant words.

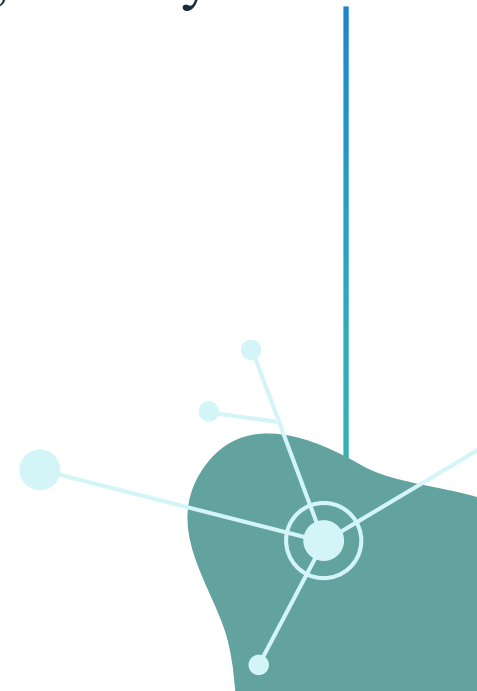


Data Chunks Creation

- Posts are combined into sets of chunks for each user. Each chunk contains posts for a one-week period to reflect the clinically validated scales of depression such as the CES-D.
- The final dataset contains 28k chunks that are balanced between classes, namely “Before” and “After” depression diagnosis.

	Number	Avg. length (in words)
Users	120	-
All Tweets	1,969,645	14
Filtered Tweets	1,213,061	14
Chunks	28,697	433

Dataset statistics (a chunk contains posts for one week).



- +
○

Classification Models

The classification unit is a **chunk** of tweets and the same user will have some chunks labelled as positive (after depression) and negative (before depression)

- **Classical ML Models**

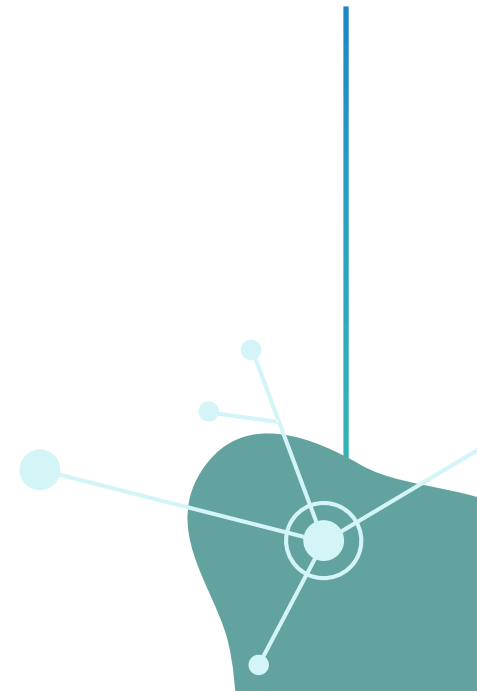
- SVM
- RF

- **Transformer-based Models**

- BERT
- RoBERTa
- MantalBERT

- **Large Language Models (LLMs) [zero-shot]**

- GPT: GPT-3 “text-curie-001” and GPT-3.5 “text-davinci-003”
- Google Bard (Gemini)
- Alpaca



- +
○

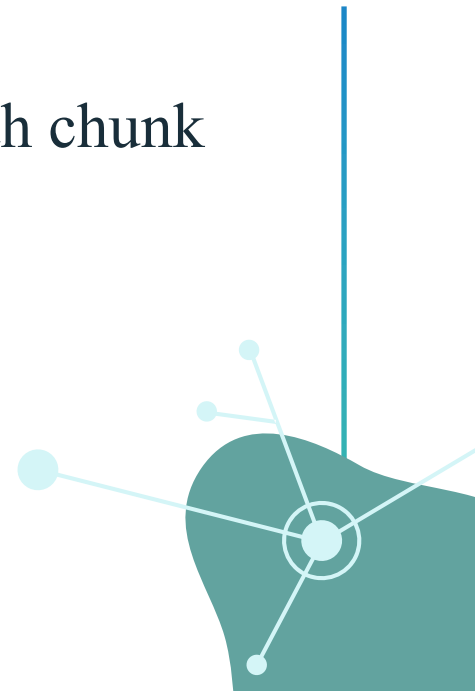
Experimental Settings

Lexicon Filtering

We run two sets of experiments, with and without filtering with lexicon.

Chunk Length

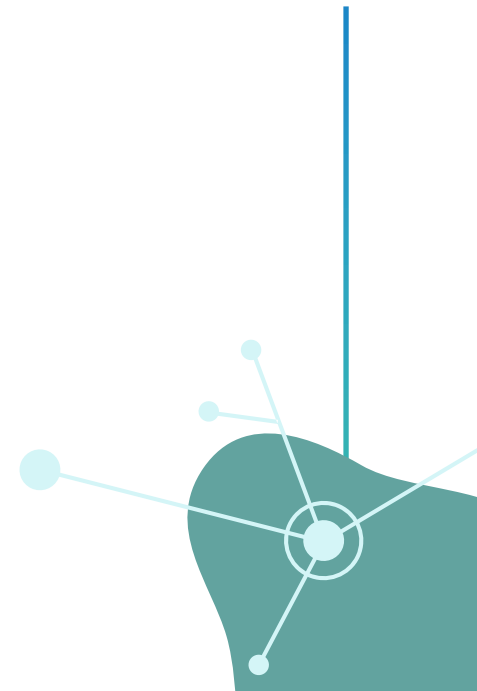
We run two sets of experiments, the first with full length and the other with chunk length trimmed to GPT maximum which is 2049 tokens.



Results - Classical and Transformers-based Models

Model	Accuracy	Precision	Recall	F-1
SVM	0.49	0.49	0.49	0.44
SVM-Filtered	0.60	0.50	0.60	0.48
RF	0.50	0.50	0.50	0.50
RF-Filtered	0.60	0.63	0.60	0.45
BERT	0.90	0.91	0.90	0.90
BERT-Filtered	0.98	0.98	0.98	0.98
RoBERTa	0.97	0.97	0.97	0.97
RoBERTa-Filtered	0.98	0.98	0.98	0.98
MentalBERT	0.96	0.96	0.96	0.96
MentalBERT-Filtered	0.98	0.98	0.98	0.98

Results for classical and transformers-based models on classifying data on full dataset and lexicon-filtered dataset.



Results - LLMs

Prompt	GPT-3 Response
Do you think the person who wrote this text is depressed	I believe this person is depressed because of the use of emojis and the long rant about how they are "no Gangsta" and "a pleasure to meet them." Additionally, the text is very wordy and has a lot of parentheses.
Do you think the person who wrote this text is depressed? answer yes or no	The person who wrote this text may be depressed.
Do you think the person who wrote this text is depressed? return a probability percentage	The author of this text is likely depressed.
To what extent do you think the person who wrote this text is depressed?	The person who wrote this text may be depressed according to some measures because they mention feeling mentally exhausted after a long day and then talk about enjoying a dinner walk with their dogs and watching TV.
To what extent do you think the person who wrote this text is depressed? answer with one word only	The person who wrote this text is likely depressed.
Classify if the person who wrote this text is depressed	The person who wrote this text likely suffers from depression due to the chronic use of "old people" as derogatory terms, the use of the phrase "sometimes you just need a little chocolate in your life" to mask the user's underlying need for alcohol, and the admission that the user has had a long day.
Classify if the person who wrote this text is depressed, reply with one word only	Depressed

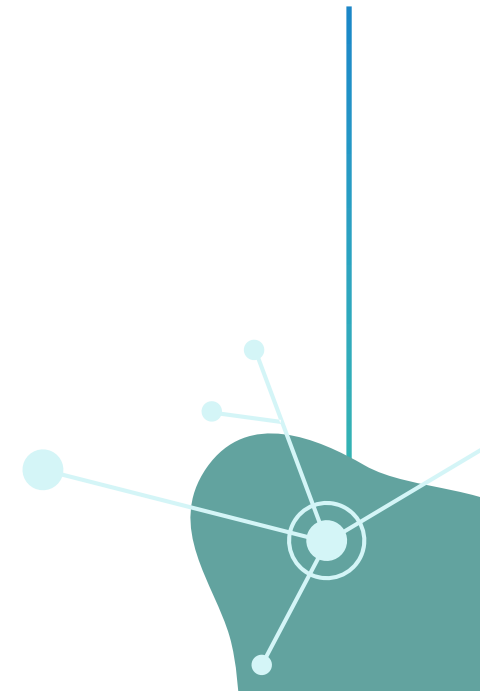
Example of GPT-3 "text-curie-001" responses to our prompts



- +
 - - ## Results – All Models

Model	Accuracy	Precision	Recall	F-1
SVM-Filtered	0.60	0.50	0.60	0.44
RF-Filtered	0.60	0.63	0.60	0.45
BERT-Filtered	0.97	0.97	0.97	0.97
RoBERTa-Filtered	0.95	0.96	0.95	0.95
GPT-Filtered "Text-Curie-001"	0.51	0.51	0.247	0.32
BARD-Filtered	0.46	0.5	0.285	0.36
Alpaca-Filtered	Hallucinating			

Comparison between results for all models with chunk length trimmed to 2049 tokens.



- +
○

Discussion

Lexicon Filtering

Our results in Table 4 show that using our lexicon to filter the dataset improved results according to all metrics, especially in precision and recall.

Chunk Length

Our experiment shows that chunking data rows into one-week chunks that align with the clinical periodicity of depression diagnosis validated questionnaires yields significantly improved results compared to previous studies that applied the same model for post- level classification

LLMs

While LLMs are powerful language models known for their NLP capabilities, LLMs have faced challenges in certain mental health classification tasks where they have not performed as effectively as transformer-based models



Thank you!

Any Questions?

Contact Falwah Alhamed on f.alhamed20@imperial.ac.uk