

Analyzing Language Patterns for Depression Detection on Social Media: Insights from Reddit Data and Machine Learning Techniques

Madan Mohan Tito Ayyalasomayajula¹

¹Affiliation not available

April 18, 2024

Abstract

Language can disclose a lot about a person's emotional condition, social standing, and even personality. In the proposed research, the authors looked at the words that are most helpful in determining if a person is depressed or not. Additionally, the authors tried to include a variety of dataset features to see if they could help us form a more accurate judgement. The authors obtained the data from the social networking site Reddit. It includes erratic messages and comments made by a group of depressed people. Given that it performed the best, the authors used logistic regression using the Term Frequency-Inverse Document Frequency (TFIDF) for the proposed model. One of the factors that contributed to a marginal speed improvement was the average time between two consecutive postings or comments. In F1-score, the proposed model fared better than other models utilising the same dataset. The diagnosis of mental disease is turning into a significant issue as people's understanding of the value of mental health grows. Many psychiatrists found it challenging to make a diagnosis of mental disease in a patient due to the complexity of each mental disorder, making it impossible to start the patient on the right course of treatment before it was too late. However, the significance of integrating into people's daily life has produced a setting in which further knowledge about a patient's mental illness may be acquired.

1. Introduction

Depression is a common mental disorder whose importance is occasionally overlooked. It can cause serious problems, incapacity, psychotic episodes, and, in rare cases, suicide if it is not properly controlled. The World Health Organization (WHO) estimates that depression affects more than 264 million people worldwide. Similar to the majority of mental illnesses, early detection may be very helpful for prevention. As technology develops, people are spending more and more time online, and their behaviour on various websites may reveal a lot about them. Excellent indicators of personality, emotional or social status, as well as mental health, can be found in language. It should not be surprising that people who are depressed use a lot of words and phrases that convey negative emotions, especially negative adjectives and adverbs like "lonely" and "sad". The use of pronouns like "me", "myself", and "I" by depressed people is particularly intriguing because it shows that they are frequently more self-focused [1]. The authors can frequently learn useful information about a person's mental condition from the content of their social media accounts. But the technology that is now used to treat depression only acts in response. Tracking internet users, for instance, can identify some risks, but warnings only sound when the person is ill or when anything immoral or insulting is done. The authors think a better option for early detection is to use Natural Language Processing (NLP) techniques on typical social network posts. This study's main goal is to suggest a method for using NLP to assess whether a person is having depressive thoughts or intentions. By extracting phrases most frequently used by Reddit users suspected of having depressive tendencies as well as other group-specific features, the authors attempted to achieve this. The dataset used in this study was first reported by Losada and Crestani

[2]. The dataset, which is described in more detail in Section 3, consists of numerous Reddit comments and posts written by different individuals. The authors used numerous Machine Learning (ML) models along with several strategies to achieve this, which are more fully described in Section 4.

The authors' four main contributions to this paper are as follows:

1. The authors look into how certain phrases or phrase combinations might be used to recognise depression.
2. As a purely functional feature, the authors look into the effects of comment/post-time frequency.
3. The authors look into the effects of comment and post length as a feature only.
4. The authors look into the effect of user posts' emotions as a feature.

The rest of the paper is structured as follows: The works that are connected to our strategy are detailed in section 2. The empirical investigation, dataset description, and pre-processing are all included in section 3. The problem statement is given in section 4. Models that have been suggested are presented in section 5. The outcomes of the suggested technique are addressed in section 6. Future research is mentioned in section 7, and section 8 of the paper is where we draw some conclusions.

2. Related Works

Studies such as Zaghoulani [3] has looked into a big social media corpus for identifying youth depression. The author suggests developing a sentiment analysis-based linguistically annotated corpus to study teen online behaviour across the MENA region. The authors want to eventually compile a sizable user base with more precise self-reported sadness signals. A sentiment analysis and emotion recognition-based automated psychometric analyzer for healthcare has been looked at by Vij [4]. They asserted and came to the conclusion that the main objective of the proposed work is to develop a self-service medical kiosk or a psychometric analyzer with fast computational linguistics capabilities that can produce a brief, concise summary of the patient's emotional health based on previous records, medications, and treatments. Almouzni et al. [5] looked into finding Arabic Twitter users that were depressed. They claimed to have developed a prediction model based on the identification of depressed individuals using an Arabic sentiment analysis employing supervised learning to assess whether a user's tweet is depressed or not. They found that sad people are more socially isolated. Priya et al. [6] have suggested ML algorithms for predicting stress, depression, and anxiety in contemporary life. They claimed that in this study, ML algorithms were employed to evaluate five different levels of stress, depression, and anxiety. They discovered that random forest has the highest accuracy (91% and 89%). A research has been done by Feuston et al. [7] on how mental illness is expressed on Instagram. They explained how their individual histories, viewpoints, and experiences with mental illness and health had an impact on how they understood the findings. Murnane et al. [8] have suggested designing technology to support long-term mental health management social ecologies. They paid close attention to the patient's perspective as well as the many viewpoints and experiences. A new class of collaborative informatics infrastructures and interfaces aimed at enabling the social ecologies of personal data activities will be built using the widely applicable design principles they gave. Pater et al. [9] have looked at a study of a case involving eating disorder patients who used digital self-harm indicators. Future studies, according to the report's authors, might examine post-intervention data and contrast it with pre-intervention data to assess changes in patients' online identity presentations. Among methods for identifying depression in college students, Xu et al. [10] suggested employing contextually-filtered characteristics and routing behaviour. In this paper, they present a unique association rule mining-based technique for automatically producing contextually filtered features that performs better than existing feature selection techniques for a depression diagnosis. Psycholinguistic patterns in social media texts have been presented by Trifan et al. [11], which aid in our understanding of depression. They are eager to discuss other psycholinguistic components with those who can shed light on them through clinical papers in a subsequent investigation. In a work that Mathur et al. [12] suggested, suicidal intent was estimated using temporal psycholinguistic clues. This study fills a gap by combining qualitative and quantitative approaches to examine the effects of enhancing text-based suicidal ideation identification.

There aren't many statistics and publications about depression despite the fact that it's a serious mental health problem. Problems with NLP are common today. The reason for this is likely a general lack of interest in the subject. Another problem is that because the topic is quite subjective, classifying such specific behavioural patterns may be challenging. The work by Losada and Crestani [2] offers excellent insight into this issue. Their dataset is the first to be used in research on language use and depression. The details of this dataset, which was also used in this investigation, are described in Section 3. An Early Risk Detection Error (ERDE) measure was established by Losada and Crestani [2] as a fresh evaluation statistic for their methodology. This metric is concerned with the speed at which affirmative circumstances can be found and the accuracy of assessments. In a fascinating study, Wang et al. [13] employed sentiment analysis to assess whether or not a user was depressed. It is advised that word and artificial regulations be used to determine each micro-depressive blog's propensity. After that, a framework for detecting depression is created using the suggested approach and ten psychologically confirmed traits of depressed people. Since social networks have a lot of text information, many researchers are seeking to build models based on ever expanding data. In the years to come, using NLP with such a benefit to address the growing depression problem may be adequate to delve further into melancholy and provide doctors with fresh and intriguing information. Another excellent method for assessing the mental health and suicide risk of a community was provided by Benton et al. [14]. It has been demonstrated that gender modelling improves accuracy in tasks involving social media text. For 10 prediction tasks, the authors of these developed neural Multi-Task Learning (MTL) models. The outcomes of their model demonstrated that choosing the MTL thinks that employing the appropriate selection of auxiliary activities for a certain mental state might result in a significantly better model. For situations with the fewest data points, the model dramatically improves. The most significant finding for our purposes was that gender prediction does not adhere to the two aforementioned rules but rather improves performance as a measure of a supplemental task.

3. Empirical Study

The description of the dataset and the pre-processing are given in this section.

3.1 Dataset Description

The Reddit posts and comments in this dataset are from 892 different individuals. The remaining people served as a control group, with 137 people receiving treatment for depression. The Reddit API limit appears to be 1000 posts and 1000 comments per user as shown in Figure 1. Additionally, both posts and comments are chronologically ordered, which is crucial given the goal of early depression identification. One XML file was created for each user and used to build the collection. Only those who have publicly admitted to having a diagnosis of depression are classified as depressives, while users who come to Reddit via sub-Reddits devoted to the topic are not depressed; instead, they are interested in learning more about depression because someone close to them is experiencing it. Each entry is identified by:

- TITLE
- ID NUMBER
- TEXT
- DATE

The 486 train subjects, 83 of whom are positive, and the 406 test subjects, 54 of whom are positive, make up the datasets' train-test split.

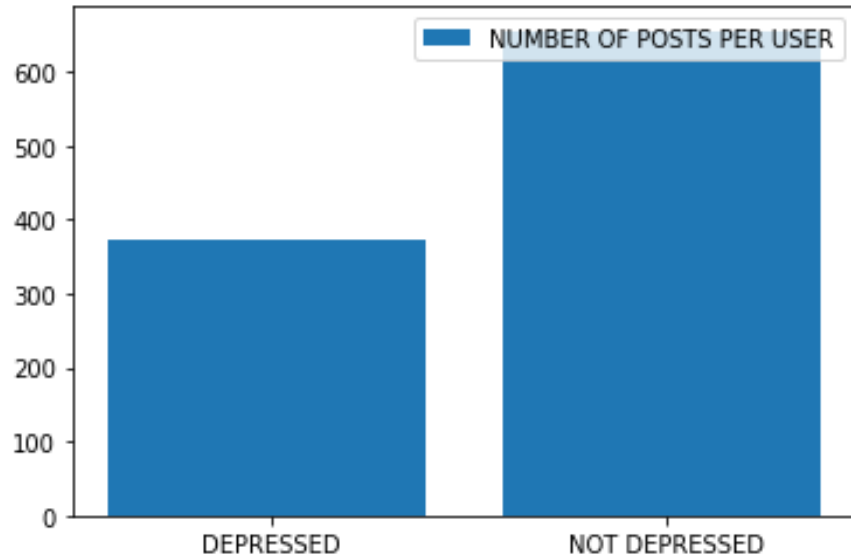


Figure 1: Number of posts made by each user, from those who are and are not depressed

3.2 Dataset Preprocessing

It is essential to clean data before using it to generate model features, especially for Reddit data, which consists of several different sets of data. The authors' initial step was to concatenate the most recent N titles or sentences for each user, where N is the minimum number of posts necessary for the suggested model to work correctly. The time between two consecutive posts or comments made by the same person was another type of data that the authors were able to retrieve. After grouping the data, the authors removed all text-based links, capitalised nothing, and created "bag of words" representations of the sequences. The authors then created a lemma for each term using WordNetLemmatizer.

4. Problem Statement

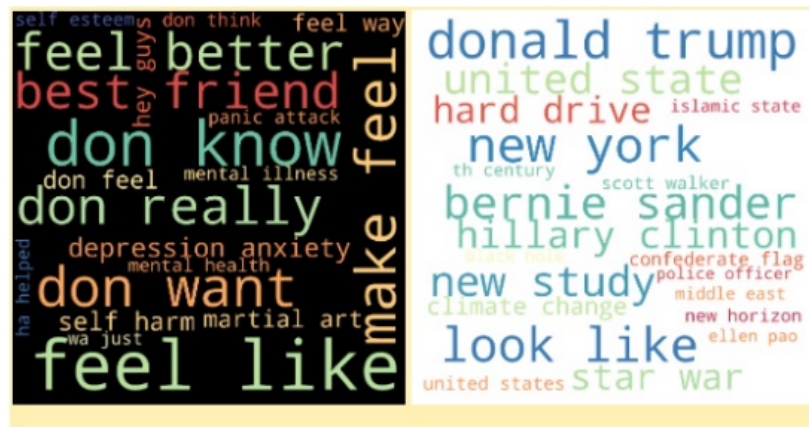
When looking through depression-related literature, the authors came across a few claims that purported to define depressed people and their involvement on social media. The use of absolute terms, negative emotion in tweets, the quantity of postings made by each user, and the intervals between posts were all examined by the authors. The time of posting is another interesting finding about depressive users on social media, however these claims were already investigated in 2019 research by Banovic et al. [15], and they were found to be unimportant, at least for this dataset. Researchers may observe a number of features and their behaviours in the material in this area. The data analysis demonstrated that several hypotheses do in fact point to significant differences between groups that are depressed and those that are not. Absolutist terms (such as absolutely, entirely, completely, etc.) don't differ significantly between groups, but average words per post and the number of comments per user do [16].

5. Models

The authors used numerous models that take into account the factors discussed above. The TF-IDF, which is essentially a statistic that determines how important a specific word is to a document in a collection or corpus, is used by the authors to vectorize words. The fact that some words appear more frequently than others can be explained by the fact that the TF-IDF value increases in proportion to how frequently a word appears in the text and is offset by the number of documents in the corpus that contain the term. For the classification challenge, the authors used a logistic regression classifier, which delivered positive outcomes on a number of text classification tasks. The baseline models and these models were contrasted. The researchers used two different naive baselines:

Random guesser: A fundamental model that makes arbitrary predictions about a user’s level of depression. Each estimate has an equal chance of being chosen.

Stratified random guesser: Another simple model with the minor addition that the estimate is not entirely random. It uses the percentage of depressed people in the train set to determine whether or not the user will be depressed.



6. Results

In terms of F1 scores on the first 10, 100, and 500 postings, this model performs noticeably better than the one described by Losada et al [2]. Additionally, this model outperforms the F1 scores of Banovic et al. [15] for all related data subsets. The average time between postings was the only feature that significantly improved the performance of the basic LR + TF-IDF model, despite the fact that data analysis provided fascinating insights as shown in Table 1. Other additional characteristics, however, only added needless noise that reduced model accuracy on the test set.

Table 1: F1 results from various methods

Model	F1 @ 5	F1 @ 10	F1 @ 20	F1 @ 50	F1 @ 100	F1 @ 200	F1 @ 500
STRATIFIED F1	0.2099	0.2099	0.2099	0.2099	0.2099	0.2099	0.2099
RANDOM F1	0.1921	0.1921	0.1921	0.1921	0.1921	0.1921	0.1921
LR TF-IDF	0.4322	0.5421	0.5901	0.651	0.6981	0.6852	0.6852
LR TF-IDF+SENT	0.2511	0.4911	0.5412	0.5911	0.6555	0.6312	0.6312
LR TF-IDF+POST LEN	0.3511	0.5001	0.5611	0.6311	0.6711	0.6512	0.6512
LR TF-IDF+POST LRN	0.3581	0.5011	0.5711	0.6100	0.6523	0.6411	0.6411
LR TF-IDF AVG DIFF BETWEEN POSTS	0.4312	0.5177	0.5661	0.6411	0.7011	0.6811	0.6811

The suggested model outperforms both of the previous models, therefore it's important to think about what the authors can take away from it and whether they can identify any trends in the diagnosis of depression. The authors can identify the words and phrases that have the most impact on a user's choice of category by using the logistic regression model. To better illustrate what the model found as the words that contribute the most to classification, the authors produced a word cloud using these terms. The larger the term, the bigger the donation. It is clear from this unigram word cloud that the authors would quickly link certain terms to depressed people, therefore it will be interesting to see how the model arranges words for both classes. To see how closely bigram concepts match human intuition, the authors may also reevaluate them. This data supports the idea that people who are depressed focus far more on themselves than on others.

7. Future Works

The size and complexity of the dataset presented the biggest difficulty in creating the model. Although it is an expensive process, expanding the dataset will improve model performance. Since the time of postings was the sole factor taken into account in addition to users' posts, it stands to reason that more information about users would be helpful to academics as well. The work of Benton et al. [14] on gender representation as an additional attribute may be useful. Alternatively, the authors could use it directly to Long Short-Term Memory (LSTM) networks, which are the main component of most cutting-edge models because they can start making use of long sequential input data as well as its going to order by wanting to avoid gradient vanishing and using recurrent connections. The authors could try using word embeddings currently averaging across all sayings as input towards the logistic regression model.

8. Conclusion

Depression is a mental condition that can result in more serious problems or even suicide if it is not properly and swiftly treated. Since a complete history of postings might provide crucial information to assist in better diagnosing patients, this type of research may be advantageous to psychiatrists. The TF-IDF characteristics and other characteristics that the authors believed might aid in a better diagnosis of depression were used to test the logistic regression model's base model. Only the response time between posts improved the fundamental model on particular data subsets among the features the authors used, which also included post sentiment, post count, post length, and average time between posts. In comparison to the model described by Losada et al. [2], the authors' F1 score using that model was significantly higher. As can be seen in section 5, the main goal was to identify words or groups of words that aid in the diagnosis of

depression. Using deep learning models, incorporating further features not included in this paper, like the user's gender, or experimenting with different word embeddings could all improve the work.

References

- [1] W. Bucci and N. Freedman, "The language of depression," *Bull. Menninger Clin.*, vol. 45, no. 4, pp. 334–358, 1981, Accessed: Nov. 18, 2022. [Online]. Available: <https://search.proquest.com/openview/d804439a2c70467603bbdf0c20a3f31a/1?pq-origsite=gscholar&cbl=1818298>
- [2] D. E. Losada and F. Crestani, "A test collection for research on depression and language use," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2016, vol. 9822 LNCS, pp. 28–39. doi: 10.1007/978-3-319-44564-9_3.
- [3] W. Zaghouani, "A Large-Scale Social Media Corpus for the Detection of Youth Depression (Project Note)," in *Procedia Computer Science*, 2018, vol. 142, pp. 347–351. doi: 10.1016/j.procs.2018.10.483.
- [4] A. Vij and J. Pruthi, "An automated Psychometric Analyzer based on Sentiment Analysis and Emotion Recognition for healthcare," in *Procedia Computer Science*, 2018, vol. 132, pp. 1184–1191. doi: 10.1016/j.procs.2018.05.033.
- [5] S. Almouzni, M. Khemakhem, and A. Alageel, "Detecting Arabic Depressed Users from Twitter Data," in *Procedia Computer Science*, 2019, vol. 163, pp. 257–265. doi: 10.1016/j.procs.2019.12.107.
- [6] A. Priya, S. Garg, and N. P. Tigga, "Predicting Anxiety, Depression and Stress in Modern Life using Machine Learning Algorithms," in *Procedia Computer Science*, 2020, vol. 167, pp. 1258–1267. doi: 10.1016/j.procs.2020.03.442.
- [7] J. L. Feuston and A. M. Piper, "Beyond the coded gaze: Analyzing expression of mental health and illness on instagram," *Proc. ACM Human-Computer Interact.*, vol. 2, no. CSCW, Nov. 2018, doi: 10.1145/3274320.
- [8] E. L. Murnane, T. G. Walker, B. Tench, S. Volda, and J. Snyder, "Personal informatics in interpersonal contexts: Towards the design of technology that supports the social ecologies of long-term mental health management," *Proc. ACM Human-Computer Interact.*, vol. 2, no. CSCW, Nov. 2018, doi: 10.1145/3274396.
- [9] J. A. Pater, B. Farrington, A. Brown, L. E. Reining, T. Toscos, and E. D. Mynatt, "Exploring indicators of digital self-harm with eating disorder patients: A case study," *Proceedings of the ACM on Human-Computer Interaction*, vol. 3, no. CSCW. Association for Computing Machinery, Nov. 01, 2019. doi: 10.1145/3359186.
- [10] X. Xu *et al.*, "Leveraging Routine Behavior and Contextually-Filtered Features for Depression Detection among College Students," *Proc. ACM Interactive, Mobile, Wearable Ubiquitous Technol.*, vol. 3, no. 3, pp. 1–33, Sep. 2019, doi: 10.1145/3351274.
- [11] A. Trifan, R. Antunes, S. Matos, and J. L. Oliveira, "Understanding depression from psycholinguistic patterns in social media texts," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2020, vol. 12036 LNCS, pp. 402–409. doi: 10.1007/978-3-030-45442-5_50.
- [12] P. Mathur, R. Sawhney, S. Chopra, M. Leekha, and R. Ratn Shah, "Utilizing temporal psycholinguistic cues for suicidal intent estimation," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2020, vol. 12036 LNCS, pp. 265–271. doi: 10.1007/978-3-030-45442-5_33.
- [13] X. Wang, C. Zhang, Y. Ji, L. Sun, L. Wu, and Z. Bao, "A depression detection model based on sentiment analysis in micro-blog social network," in *Lecture Notes in Computer Science (including subseries Lecture*

Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) , 2013, vol. 7867 LNAI, pp. 201–213. doi: 10.1007/978-3-642-40319-4_18.

[14] A. Benton, M. Mitchell, and D. Hovy, “Multitask learning for mental health conditions with limited social media data,” in *15th Conference of the European Chapter of the Association for Computational Linguistics, EACL 2017 - Proceedings of Conference* , 2017, vol. 1, pp. 152–162. doi: 10.18653/v1/e17-1015.

[15] L. Banovićbanović, V. F. Fatorić, and D. Rakovac, “How soon can we detect depression?” Accessed: Nov. 18, 2022. [Online]. Available: https://www.fer.unizg.hr/_download/repository/TAR-2019-ProjectReports.pdf#page=7

[16] M. Al-Mosaiwi and T. Johnstone, “In an Absolute State: Elevated Use of Absolutist Words Is a Marker Specific to Anxiety, Depression, and Suicidal Ideation,” *Clin. Psychol. Sci.* , vol. 6, no. 4, pp. 529–542, Jul. 2018, doi: 10.1177/2167702617747074.