JOURNAL OF
EMERGING INVESTIGATORS

Article

Depression detection in social media text: leveraging machine learning for effective screening

Jenna Aram Shin[1]*, Andrew Gunhee Kim[1]*, Zuhawn Sung[2]

[1] Asia Pacific International School (APIS), Seoul, South Korea

[2] Fermi National Accelerator Laboratory, Batavia, Illinois, USA

* These authors contributed equally

SUMMARY

Depression is a significant mental health concern that affects millions of people worldwide. Early diagnosis and intervention are crucial for effective treatment. However, identifying depression symptoms can often be challenging. From the current new wave of social media platforms, we hypothesized that examining texts in social media is an enabling pathway to describe a person's mental condition, compared to conventional protocol applied in clinical diagnosis. In this study, we investigated the detection of depression-related patterns on social media platforms. We utilized advanced classification algorithms in order to extract key linguistic features and emotional expressions from social media texts. By applying two supervised classification machine learning algorithms, decision trees and random forests, we were able to distinguish differences between the depression and non-depression conditions. We applied natural language processing, an AI-powered language model which is the basis of ChatGPT. Our study suggests that screening mental disorders is feasible only by looking at texts. Hence, we suppose that our developed method would be applicable to evaluate mental disorders by leveraging a person's activity level. It is also expected that this way would further improve the accuracy of depression diagnosis by considering both physical and psychological symptoms.

INTRODUCTION

Currently, psychologists and mental disorder researchers define 'depression' like depressive disorder as a severe mood-related mental disorder (1, 2). Like other mentor disorders, various genetic factors, such as biological features, brain nerve system and chemistry, hormones, and inherited traits, are involved in determining depression (3,4). The human brain can be easily affected by depression. Depression is associated with changes in the levels of neurotransmitters and hormones like monoamine and dopamine, which can affect the brain's neural activity (5, 6). The study of the Global Burden of Disease from 1990 to 2017 indicates that the cases of depressive disorder rapidly increased from 162 to 241 million at an increasing rate of 49.86% (7). In 2020, a major depressive episode appeared in 21.0 million adults, almost 8.4% of the total adult population in the US, according to the report of the National Survey on Drug Use and Health. Female adults are much more determinantal to depression than the case of males, and those aged 18-49 years are much more outstanding than those aged greater than 50 years (8). Between 2010 and 2018, the economic burden on adults with depressive disorder in the United States experienced a substantial increase, surging from $237 billion to $326 billion, representing a significant growth of approximately 38% (9). This estimate includes direct and indirect overall health-related expenses.

The person who suffers from depression shows several severe symptoms associated with personal feelings, a way of thinking, and daily activities (1, 3). The symptoms can be classified into two different categories: mood-related and physical symptoms. The former includes feelings of sadness, anhedonia (lack of interest or enjoyment), anger, loss of appetite, and erratic sleep habits, and, as a consequence of mood-related symptoms, the latter include constant fatigue, muscle aches, insomnia, headaches, back pain, limb pain, and gastrointestinal problems. In addition, since the decreased function of the labor (or work) can be initiated by depression (3), older individuals suffering from depression experience more pronounced challenges in terms of memory loss and response in comparison to young adults with depression.

The symptoms of depression are often not easily noticeable, even in the case of a psychologist evaluating the patient via a well-developed medical prognosis procedure using The Diagnostic and Statistical Manual of Mental Disorders (DSM-5-TR), a reference of diagnostics criteria for mental illness (10). The primary signs include continuous and chronic experiences of the following suffering such as persistent sad, anxious or empty mood, feelings of hopelessness, irritability, frustration, restlessness, guilt, worthlessness, difficulty concentrating or remembering, thoughts of death or suicide or suicide attempts, and isolating from family and friends (1,2). However, the diagnosis of depression in a clinical way has been disputed and criticized due to over-medication and -pharmaceuticalisation (5). The incongruity between patients' distress expression and their emotional resistance to medicalized discourse also leads to inappropriate diagnoses from medical professionals (11). Hence, an earlier diagnosis of depression is crucial for effective treatment and reduction of social and economic costs.

https://doi.org/10.59720/24-008

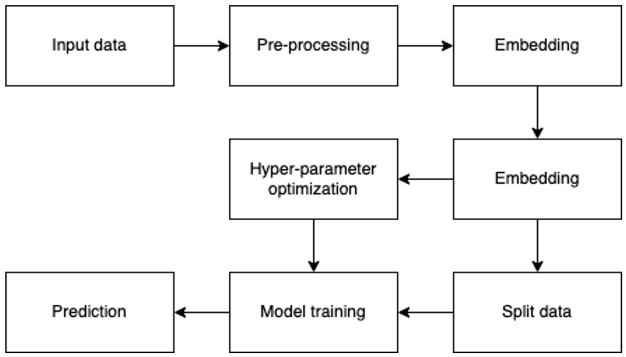

Figure 1: Workflow chart of the process applied in this study. This chart represents the sequence of data analysis based on Python code written in Google CoLab. Details of each process are explained in the main text.

Therefore, it is highly demanding to develop more effective depression screening methods. In the recent social media wave, patients express their anxiety and depression mood via social networking services (SNS) (12). This tendency becomes further remarkable with the usage of advanced smart mobile devices. Using this technical advantage, one can acquire various factors for individual psychological parameters by monitoring individual activities related to mental features. A meta-analysis on the relative association between SNS use and potential depression showed that depression symptoms are related to increasing SNS usage time (12). Trend clustering analysis also suggested a distinct pattern between social media use and depression symptoms (13). Statistics show that 32% of adolescents have positive experiences with depression-related SNS posts (14).

Although these studies do not suggest any clinical meanings, we hypothesized that sophisticated classification of texts in SNS can be used as a tool to distinguish someone who suffers from depression. Machine learning can treat tremendous datasets and draw rational predictions based on the learning of unknown datasets, such as SNS tweets. Hence, we anticipated that if significant amounts of individual texts are assessed regarding their importance on depression evaluation via machine learning, fast and more accurate depression screening would be expected compared to the case by conventional clinical protocol (10). This study focused on effectively recognizing depression in SNS texts by utilizing machine learning algorithms. We classified sentimental text tweets by text labeling and specified high-frequency sentences that represent depression. After vectorization of text embedding with the optimized hyper-parameters, decision tree (DT) and random forest (RF) models were applied to predict accuracy and F1-score (a metric of accuracy). The results showed remarkably high accuracies, suggesting that evaluation only with SNS text tweets can identify individuals who experience depression. The importance rank of the features, which are the individual words extracted from the original dataset, indicates that "depression" is the most crucial word associated with "depressive mental disorder". And

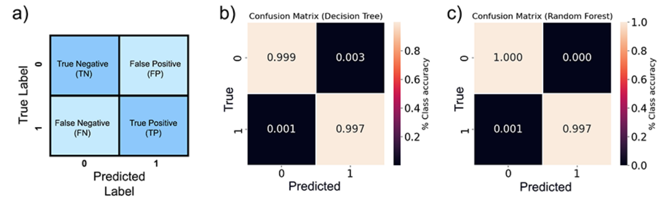

Figure 2: Confusion matrix (a) terms of binary 2 x 2 matrix, (b) Decision Tree, and (c) Random Forest from our ML analysis. From (a), predictive accuracy values are easily noticeable for each ML algorithm.

"anxiety," "emoji for depression," and "anxiety depression" are the following. Our study suggests that evaluating word expressions posted on SNS would be a criterion for leveraging a personal emotional level associated with mental disorders.

RESULTS

In this study, the words on social media, called tweets, were applied for personal sentimental expression. The dataset was downloaded via Kaggle website (15). It consists of a total of 10,282 tweets categorized into three major columns such as index, message, and label. The binary in the label indicates the depressive status of the tweeters who wrote the message. This status is a survey response by the individuals, which does not indicate any actual medical diagnosis. Among 10314 people, 8000 were labeled with 0, indicating they were not depressed, and 2314 were labeled with 1, indicating depression. These are input expressions by the tweeters.

Since the dataset contains labels with two classes, it is naturally a binary classification problem. We applied machine learning algorithms to classify different sentiment expressions from raw tweet texts (**Figure 1**). The workflow chart of the applied process is shown in **Figure 1**. We wrote Python codes on the Google Colab platform, a Google version of the Jupyter notebook. The code is shared through the link (16). Pre-data processing was initially performed to clean out the raw data, and text embedding was carried out for hyperparameter optimization after splitting the data into train and test sets. Two supervised classification machine learning algorithms, decision trees (DT) and random forest (RF) (17), were applied for prediction analysis. The prediction accuracy was evaluated by following the references (18–20) described in the Method.

The confusion matrix is a table that summarizes the performance of a classification algorithm, as the output can be two or more classes. It predicts a categorical label for each input stance. For binary classification, it will be a 2 x 2 matrix, as shown in **Figure 2a**. The confusion matrices obtained from our machine learning (ML) algorithm applications (decision tree and random forest) are shown in **Figure 2b and 2c**. The outcomes of the maximized value of true positive (TP) and true negative (TN) for both DT and RF were close to "1", meaning

Decision Tree		Random Forest	
Accuracy	F1 score	Accuracy	F1 score
0.99915	0.99729	0.99958	0.99865

Table 1: Predictive accuracy acquired from two ML algorithms.

https://doi.org/10.59720/24-008

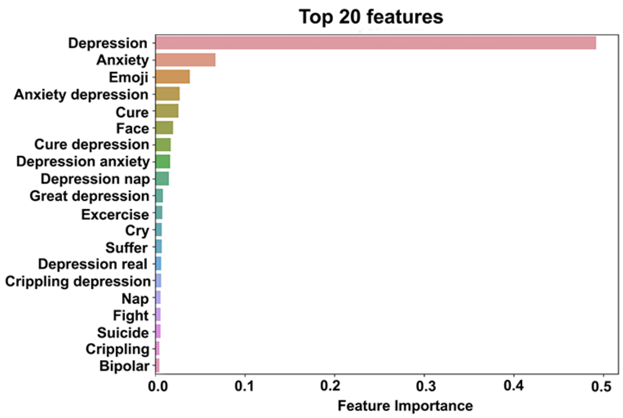

Figure 3: Feature importance of the top 20 words (terms) in Random Forest model. The top 20 words are identified based on two ML algorithms, and their importance for prediction in Random Forest method is listed.

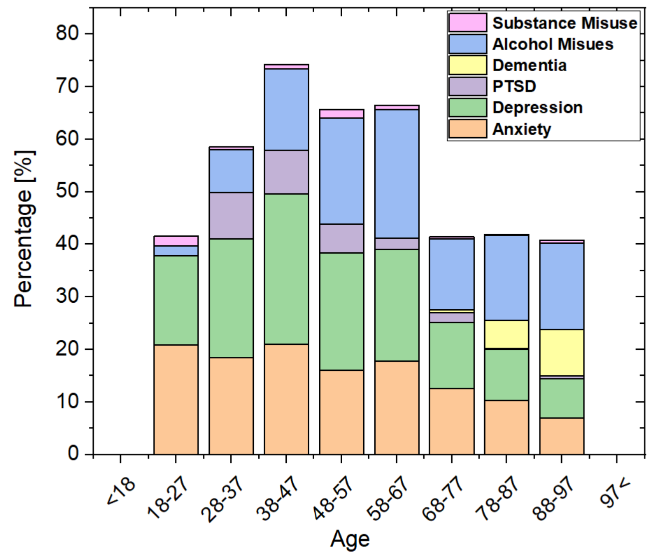

Figure 4: Percentage of each mental disorder split by age group of the registered veterans in UK (total patient 2449, 88% male, 12% female). This plot was reconstructed using data from Reference 15 to emphasize the significance of each mental disorder in stock profile format using Python numpy.

that prediction with those two models is ultimately optimized for a classification problem. Naturally, false classification is close to "0".

The results of our ML analysis performance in **Table 1** describe the accuracy and F1 score for each model: 0.99915, 0.99729 for Decision Tree, and 0.99958, 0.99865 for Random Forest, respectively. Unfortunately, we could not perform a statistical evaluation due to the limited availability of similar data sets. However, there was likely no significant difference between the two models based on either accuracy or F1 score. Those values are high enough for prediction accuracy. It is likely that our prediction applications were highly accurate in identifying people who suffer from depression only with SNS text tweets.

Furthermore, we evaluated the importance of the features that contributed to the accuracy and F1-score of the models. In this case, the features are the extracted words from the original tweet texts after filtering. **Figure 3** shows the top 20 words (features) ordered by their importance based on RF modeling. It is easily seen that "depression" is the most important word for prediction compared to others. Anxiety, emoji, anxiety depression, and cure follow in a sequence, but their contributions are not significant, likely 20% of the word "depression." Namely, our models give a high weight to the term "depression" to identify depression candidate tweets.

DISCUSSION

In this study, we showed that "depression" is the most critical sentimental word to describe the depression status in the SNS tweets. Our study suggests that texting on social media plays a pivotal role in increasing the diagnosis accuracy of mental disorders like depression because a person who suffers from the disorder would naturally write sentimental feelings on the social networking service (SNS) platform. A significantly increased rate of depression is considerably attributed to inappropriate SNS usage, and analogous mental disorders like anxiety, stress, obsessive-compulsive disorder, attention-deficit/hyperactivity disorder, and alcohol overuse and dependency can also reflected in problematic SNS use (21). Indeed, this trend is significant with the usage of mobile devices (22). Hence, if the SNS texts are interpreted in a proper way, an effective diagnosis of depression can be realized, and if combined with the systematic medical process, it may lead to cost reduction compared to the case in the medical approach is only applied. A psychiatrist or behavioral doctor diagnoses a person with depression based on a clinical manual like the DSM-5 (10). However, this clinical approach is time-consuming and requires multiple steps to diagnose accurately, including training, administration (asking many questions), scoring, and consensus coding. Indeed, psychological over-medication and medicine overdose can also be achieved by improper clinical diagnosing from the patient's resistance (11).

By evaluating sentimental text tweets for predictive screening methodology with machine learning algorithms, we were able to validate the prediction effectiveness and efficiency of the depression screening. With the 5-fold cross-validation for hyper-parameter optimization, we estimated prediction accuracy and F1-score (a metric of accuracy in machine learning). We evaluated the extent of the importance of the features, which are the individual words extracted from the original dataset. The result shows remarkably high accuracies, implying that our technical approach enables us to determine if the person suffers from depression solely based on SNS texts. "Depression" is the most crucial word that can determine depressive disorder. In addition, "anxiety", "emoji for depression", and "anxiety depression" were found to be the next important words in sequence. This implies that using these specific words is a strong positive symptomatic indication of depression, from which clinical professionals could make further precision diagnoses. However, the tweets applied in this study are from a publicly available database, so that detailed demographic information is lacking. Thus, it restrained our studies from specifying age or gender-related "sentimental words" that would allow us to investigate detailed depression-representative texts. Risk factors associated with the prevalence and occurrence of depression vary with age

https://doi.org/10.59720/24-008

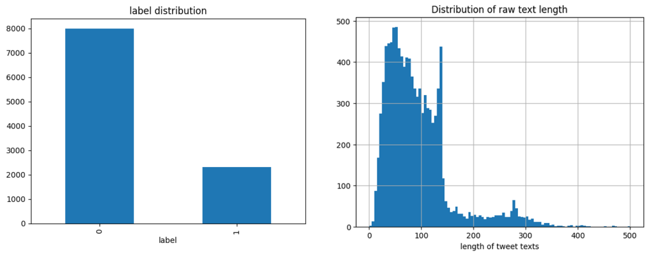

Figure 5: Distributions of a) label and b) lengths of a row text dataset. 0 indicates healthy condition, while 1 does depression one. Majority of tweet lengths are between 0 to 150.

and gender, even with occupations and social groups (23, 24).

Recently emerging Chat GPT is intrinsically organized based on natural language understanding (NLU), which was applied in this study. Text embedding is the process of transforming texts into arrays of numbers, which is a basic process step of NLU. To analyze text in an accurate way, it's advisable to use a tool specifically dedicated to NLU (25). Considering this point, our technical approach would be useful for enhancing the feasibility of ChatGPT application in mental clinic fields.

In fact, our results suggest that anyone can be readily aware of an individual who suffers from depressive disorder if the person writes the words of "depression" or "anxiety" on an SNS. However, our approach provides a precise technical direction on how a psychiatrist or behavioral doctor can effectively and efficiently diagnose mental disorder patients with the current popular culture of sharing personal information on social networking platforms. Social media such as X (Twitter), Facebook, discussion forums, and microblogs are long-standing platforms since they have become popular for expressing and recording individuals' personalities, feelings, moods, thoughts, and behaviors (26). Our study suggests a simple but rational approach that a behavioral health doctor or psychiatrist can use to expedite the diagnosis of patients who suffer from mental disorders.

There is an interesting lesson about how depression impacts a specific group. We reconstructed the percentage profile of mental disorders depending on age group collected in March 2021 by UK patient primary healthcare medical (23). As shown in **Figure 4**, depression prevails in most cases among mental health disorders, including anxiety and depression, post-traumatic stress disorder (PTSD), and alcohol misuse. The rate is even higher than the case of veterans who have combat experiences like PTSD. This is an unambiguous example of how much significant depression is compared to other mental disorders and how much its impact varies with the type of a group. Depression can accelerate the government's financial burden and, as a result, the overall health system of the country can be risky.

As a future project, we proposed to focus on a specific group that can be easily affected by environmental experience. It would be based on strong cluster analysis to classify individual group categories regarding mental disease. Additionally, we propose facilitating smart wearable sensors.. The activity of a person who suffers from depression is significantly lower than the case of a healthy person (27). Currently, 97.63 million global users use smart wearables, and the number of users is forecasted to increase by 740.53 m in 2029 (28). Its medical application is also highly demanding because medical practitioners employ wearables to measure simple activity associated with parameters for medical treatment (29, 30). From these facts, if SNS text analysis is combined with physical activity investigation using smart wearables, we can expect an expedite diagnosis of mental disorder can be achieved.

We confronted several analytical limitations in this study. As aforementioned, we could not interpret the demographic effect on our depression screening due to a lack of information. In addition, since our study is based on tweets with already-expressed depression status by the individual, screening unknown patients was not performed, which could be important in diagnosing random patients. Hence, investigating undetermined SNS tweets with demographic information would be beneficial for further clarifying the advantage of machine learning applications for depression screening.

MATERIALS AND METHODS

Obtaining the social media dataset

Machine learning (ML) is the process of using data and algorithms that gradually enhance accuracy by emulating how humans learn. The algorithm is trained on a dataset of examples and learns to recognize patterns in the data that can be used to make precision predictions and judgments. DT and RF supervised classification ML models were used in this study (31). A decision tree is an algorithm that utilizes the decision-making process by employing a tree-like model of choices and their likely effects. Random Forest is an ensemble model that uses many small trees trained independently and gathers the results from the trees and reports the final prediction as the majority vote of the individual trees. The dataset applied was downloaded via the Kaggle website (15). It consists of a total of 10,282 tweets categorized with three major attributes, such as index, message, and label for the depressive status expressed by the individuals. For this study, we wrote Python code in the platform of Google Colab (16).

Dataset filtering

We first assessed the label distribution of the words and text length in the social media (tweets) dataset, as shown in **Figure 5a and 5b**. The label distribution reveals the balance of labels. Since training machine learning models work best on a balanced dataset, we examined whether our data's label balance was good. In our case, the skewness is not quite bad, so we decided to use them as they are.

Most texts had below 200 characters and some texts had 500 characters. We inspected the longer texts and concluded that they are mostly long URLs or meaningless characters like commas, at sign, backslash, misspelled, etc. Thus, we decided to filter those texts out by applying a cut, the text length should be less than 180. **Figure 6** shows the distribution of a text's word count before and after filtering. The figure sheds light on how many words are contained in the text and the impact of character length filtering. As a result, we used filtered texts for the next step. Text data preprocessing is an important step in preparing text input data to transform into a form that machines can digest.

Figure 7 displays word clouds that represent the term frequency appearing in the texts. Figure 7a highlights the key

https://doi.org/10.59720/24-008

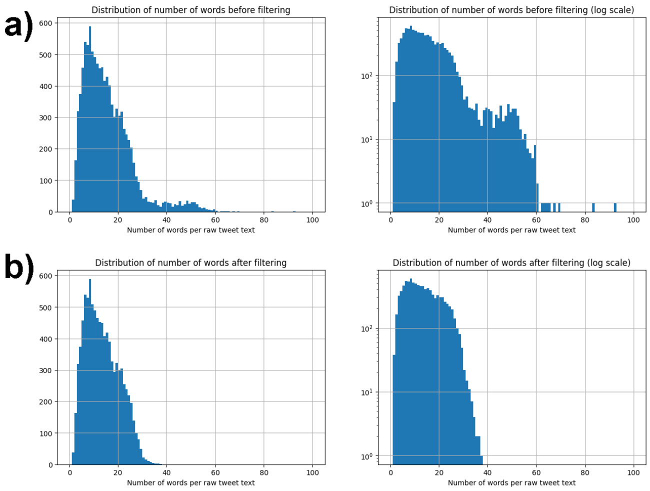

Figure 6: Number of words per raw tweet text in (a) before filtering and (b) after filtering. Both linear (left) and log (right) scales at y coordinate are compared to show tail data.

Figure 7: Word clouds (a) with and (b) without depression. The size of characters represents relative frequency for each conditions.

terms that are frequently linked to topics related to depression. Some words stand out more than others, indicating their importance and relevance in the context of depression. A word cloud without depression is shown in Figure 7b, giving a more positive word.

Data preprocessing and text embedding

Since text data preprocessing is necessitated to make text input suitable for machine learning model analysis, a Python package called spacy (18) was applied. During preprocessing, words are split in a sentence for the next step (word tokenization), punctuation is removed, stop words are structural words such as "a" and "the" are usually removed in the preprocessing, URL or email addresses are removed, and special characters such as emoji or user identifiers are removed.

After preprocessing, each sentence is transformed into a real-valued vector in a process called text embedding (18). We used TfidfVectorizer from sklearn (19) to transform our sentences to vectors. It includes a bag of words vectorizing step and a term frequency - inverse document frequency (TFIDF) weighting step. TFIDF is roughly defined as term frequency multiplied by inverse document frequency. Some frequent words in the documents can carry little meaningful information, so it is insufficient to reflect the meaning of texts. Inverse document frequency is used to reweight the term frequency so that too-frequent words receive small weights while the weights of infrequent words are enhanced (32). We use the following non-default parameters of TfidfVectorizer, which are ngram (the range of n-grams to be extracted) = (1,2), max_df (ignore terms that the document frequency is higher than the given value) = 0.5, and min_df (ignore terms that the document frequency is lower than the given value) = 0.001.

Parameter optimization for machine learning

The choice of the algorithm parameters was based on heuristics. We tried a few sets of parameters and didn't find much difference in the outcomes. So, we chose the ones with reasonable and good rounding numbers. The vectorized inputs, together with labels, are split into train and test samples. We use the sklearn train_test_split function to sample them. The split ratio of train to test sample is 0.75 to 0.25. When both algorithms (Decision Tree and Random Forest) are applied, we optimized hyper-parameters using brute-force grid search method. We could not optimize all hyper-parameters due to lack of time and computing resources. Thus, we started with the parameters that are sensitive to the model performance, which are max_depth: maximum depth of the tree, and min_sample_leaf: minimum number of samples required to split an internal node. For grid search, we use GridSearchCV provided by sklearn which utilizes cross validation techniques. The default cross validation is a 5-fold method. The 5-fold method utilizes all datasets without splitting train and test. For each fold, 80% of the dataset are used for training and 20% are used for testing. Optimization is to find the hyper-parameters to produce the best accuracy score. Once GridSearchCV finds the optimal hyper-parameters, we train the models by setting those hyper-parameters. The performances of training models are measured by accuracy score and f-1 score. Finally, the models are trained based on features in the data. We can see how much each feature contributes to the model.

Validation of machine learning model

To execute the performance evaluation of each model applied, we decided to use a simple metric, "accuracy". Following the literature (33), we defined it as a simple ratio by accuralely classifying the sample from the total number of the applied sample. However, the asymmetric structure is a typical feature of the practical dataset, which means a classification can not be evenly attributed, namely unbalanced. Thus, we also facilitate other metrics to counterbalance the dataset. It is an "F1 score" with validation determined from the prediction and recall (34). This metric is based on harmonical measuring of these two features (recall and precision): recall, called sensitivity, can denote the relevancy by the fraction of the correctly retrieved samples with false negative sample for the true positive accuracy class, but precision can depict the relevancy of the fraction for the retrieved samples only with the positive class, typically called positive predictive value. Using this F1 score, we were able to produce a suitable

$$Accuracy\ =\ Number\ of\ correct\ predictions/Total\ number\ of\ predictions$$

$$Accuracy\ =\ \frac{TP\ +\ TN}{TP\ +\ TN\ +\ FP\ +\ FN}$$

performance with the introduction of various errors. These definitions and relations are described as follows (33)

$$Precision\ =\ \frac{TP}{TP\ +\ FP}$$

$$Recall\ =\ \frac{TP}{TP\ +\ FN}$$

$$F1\ score\ = 2 * \frac{Precision\ *\ Recall}{Precision\ +\ Recall}$$

where TP = True Positives, TN = True Negatives, FP = False Positives, and FN = False Negatives.

Received: December 19, 2023
Accepted: May 2, 2024
Published: March 25, 2025

REFERENCES

1. Hammen, C., and E. Watkins. “Depression.” *Taylor & Francis*, 2018, pp 1-3, http://doi.org/10.4324/9781315542805.
2. “Depression.” for NIH (National Institute of Mental Health), www.nimh.nih.gov/health/topics/depression. Accessed March 2024
3. Gelenberg, A. J. “A review of the current guidelines for depression treatment” *J Clin Psychiatry*. 71(7): e15. Jul. 2010, https://doi.org/10.4088/JCP.9078tx1c.
4. “Anxiety Disorders.” for verywell health, www.verywellhealth.com/depression-and-anxiety-signs-symptoms-and-treatment-5191284, Medically revied by M Machintry, Published 06 Feb. 2024.
5. Dumit, J., “Is It Me or My Brain? Depression and Neuroscientific Facts.” *Journal of Medical Humanities*, 24, pp. 35–47, Jun. 2003, https://doi.org/10.1023/A:1021353631347.
6. Sigh B. S., “The utility of PET imaging in depression”, *Front Psychiatry*. vol. 15, no. 1322118, 22 Apr 2024, https://doi.org/10.3389/fpsyt.2024.1322118.
7. Liu, Q., et al., “Changes in the global burden of depression from 1990 to 2017: Findings from the Global Burden of Disease study.” *Journal of Psychiatric Research*, vol. 126, pp. 134-140, July 2020, https://doi.org/10.1016/j.jpsychires.2019.08.002.
8. “Annual report National Survey on Drug use and Health (NSDUH).” Releases by Substance Abuse and Mental Health Service Administration (SAMHSA: https://www.samhsa.gov/), www.nimh.nih.gov/health/statistics/major-depression, accessed Mar 2024.
9. Greenberg, P. E., et al., “The Economic Burden of Adults with Major Depressive Disorder in the United States (2010 and 2018).” *Pharmacoeconomics*, vol. 39 no. 6, pp. 653-665, Jun 2021, https://doi.org/10.1007/s40273-021-01019-4.
10. “Diagnostic and Statistical Manual of Mental Disorder (DSM-5-TR).” by American Psychiatric Association, May 18, 2013, ISBN 978-0-89042-554-1.
11. Kokanovic, R., et al., “Depression: the ambivalence of diagnosis.” *Sociology of Health and Illness*, vol. 35, issue 3, pp. 377-390, Aug. 2012, https://doi.org/10.1111/j.1467-9566.2012.01486.x.
12. Cunningham, S, “Social Media and Depression Symptoms: a Meta-Analysis.” Res Child Adolesc Psychopathol, vol. 49, pp. 241-253, Jan. 2021, https://doi.org/10.1007/s10802-020-00715-7.
13. Ariel, S., et al., "Social Media Use and Depression and Anxiety Symptom: A Cluster Analysis.” *American Journal of Health Behavior*, vol. 42, no. 2, pp.116-218, Mar 2018, https://doi.org/10.5993/AJHB.42.2.11.
14. “Social Media Statistics, Social Media and Anxiety Among Youth”, ConsumerNotice.org, www.consumernotice.org/personal-injury/social-media-harm/statistics/. Modified 15 Aug 2023.
15. “Sentimental Analysis for Tweets, Kaggle.com.” Database of tweets to perform Sentimental Analysis. www.kaggle.com/datasets/gargmanas/sentimental-analysis-for-tweets.
16. “Depression_data_machine_ learning-Jenna.ipynb.”, Google colab written by Jenna Shin and Andrew Kim, colab.research.google.com/drive/1RJY5GH4HESSfJjSRBCVRyBaw5P0Y507D?usp=sharing.
17. Wu, J. D., et al., “Privately Evaluating Decision Trees and Random Forests.” Cryptology, *Proceedings on Privacy Enhancing Technologies*, vol. 4, pp. 335–355, 2016, https://doi.org/popets-2016-0043.
18. “Language Processing Pipelines.” For spaCy, Industrial-Strength Natural Language Process, in python, https://spacy.io/usage/processing-pipelines#processin, Access Mar 2024.
19. “A Guide on Word Embeddings on NLP.” for www.turing.com/kb/guide-on-word-embeddings-in-nlp, Accessed 15 Jan 2024.
20. “Scikit-learn: Machine Learning in Python.” Pedregosa, et al., JMLR 12, pp. 2825-2830, 2011.
21. Weinstein, AM, “Problematic Social Networking Site use-effects on mental health and the brain” *Front Psychiatry*, vol. 19, no. 13:1106004, Jan 2023, https://doi.org/10.3389/fpsyt.2022.1106004.
22. Dietrich, J. J., et al., “High cellphone use associated with greater risk of depression among young woman aged 15-24 years in Soweto and Durban, South Africa.” *Glob Health Action*, vol. 14, no. 1, 1836792, Aug 2021, https://doi.org/10.1080/16549716.2021.1936792.
23. Finnegan, A., and R. Randles, “Prevalence of common mental health disorders in military veterans: using primary healthcare data.” *BMJ Mil Health*, vol. 22, no. 169, pp. 523-528, Nov 2023, https://doi.org/10.1136/bmjmilitary-2021-002045.
24. Cuijpers, P., et al., “Psychotherapy for Depression Across Different Age Groups: A Systematic Review and Meta-Analysis.” *JAMA Psychiatry*, vol. 77 no. 7, pp. 694-702, Jul 2020, https://doi.org/10.1001/jamapsychiatry.2020.0164.
25. Wolf, T., et al., “Transformers: State-of-Art Natural

https://doi.org/10.59720/24-008

Language Process." *Proceedings of the 2020 EMNLP (Systems Demonstrations)*, pp. 38-45, Oct 2020, https://doi.org/10.18653/v1/2020.emnlp-demos.
26. Kaplan, A. M., and M. Haenlein, "Users of the world, unite! The challenges and opportunities of Social Media." *Business Horizons*, vol. 53 no. 1, pp. 59-68. Jan 2010, https://doi.org/10.1016/j.bushor.2009.09.003.
27. Dinas, P. C., et al., "Effects of exercise and physical activity on depression." *IrJ Med Sci* 180, pp. 319-325, Nov 2020, https://doi.org/10.1007/s11845-010-0633-9.
28. "Number of users of smartwatches worldwide from 2020 to 2029.", Statista, www.statista.com/forecasts/1314339/worldwide-users-of-smartwatches.
29. Beendikt, J. B., et al., "Wearable technology in orthopedic trauma surgery - An AO trauma survey and review of current and future applications." *Injury*, vol. 53, no. 6, Jun 2022. pp. 1961-1965, https://doi.org/10.1016/j.injury.2022.03.026.
30. Lamyae, S., et al. "A systematic review of gamification in e-Health." *J Biomed Inform*, vol. 71, pp. 31–48, Jul 2017, https://doi.org/10.1016/j.jbi.2017.05.011.
31. Jehad, A., et al., "Random Forests and Decision Trees." *IJCSI International Journal of Computer Science*, vol. 9, issue 5, no. 3, Sep 2012, ISSN (Online): 1694-0814.
32. "Term Frequency", Constellate, constellate.org/docs/what-is-tf-idf, Published 23 May 2024.
33. Reda, Y. and Dustin, A, "Probabilistic Extension of Precision, Recall, and F1 score for more thorough evaluation of classification models." *Proceedings of the First workshop on evaluation and comparison of NLP systems*, https://doi.org/10.18653/v1/2020.eval4nlp-1.9.
34. Sasaki, Y., "The truth of the F-measure.", version: 26th, Oct 2007, https://www.cs.odu.edu/~mukka/cs795sum09dm/Lecturenotes/Day3/F-measure-YS-26Oct07.pdf.