An Empirical Analysis of Mental Health Detection using Machine Learning

Navneet Singh[1], Jagrit Aggarwal[1], Satya Ranjan Jena[1], Ishleen Kaur[2]

VIPS-TC School of Engineering and Technology, GGSIPU, Delhi, India[1]
SGTB Khalsa College, University of Delhi, Delhi, India[2]

Corresponding author: Navneet Singh, Email: navneetsingh.871931.aesl@gmail.com

Mental health disorders continue to pose significant challenges to individuals and societies worldwide. It serves itself to be an issue faced by all well-being, hence necessitating efficient and accurate methods for their detection and diagnosis. This study focuses on empirical mental health detection by analyzing the Reddit app's comments. Social media is a multifaceted platform providing updates regarding the news and events, and to mutually connect with our close well-being. Social media offers unfiltered information about the individual's thoughts and emotions, providing a rich source for observing and detecting mental health challenges that people face nowadays. The research focuses on the evaluation of the comments using word embedding techniques like Fast Text, Word2Vec, and Bag of Words, for a better understanding of the textual data. Furthermore, several machine learning models are employed to perform sentiment analysis on the preprocessed data. The models utilized in this study include Random Forest, Multinomial Naïve Bayes, Logistic Regression, and XGBoost classifier, a member of the gradient boosting algorithm family. By leveraging machine learning models, the study provides valuable insights into efficient and accurate diagnostic processes for specific mental health disorders, particularly PTSD and anxiety. The highest accuracy achieved by our analysis is 79.471%.

Keywords: Anxiety, Machine learning, Mental health, Natural language processing, Social media, Post-Traumatic Stress Disorder (PTSD).

2023. In Satyasai Jagannath Nanda & Rajendra Prasad Yadav (eds.), *Data Science and Intelligent Computing Techniques*, 361–371. Computing & Intelligent Systems, SCRS, India. https://doi.org/10.56155/978-81-955020-2-8-32

1 Introduction

In recent years, the significance of mental health has witnessed a substantial surge mainly due to the SARS-CoV-2 pandemic resulting in heightened attention and dedicated research efforts understanding the profound impact of mental health on individuals' families and societies has fostered an enhanced comprehension of patterns and effective strategies to address mental health-related concerns [1]. This study aimed to ascertain the prevalence and state of mental health while exploring emerging trends within this field. However, an exceptional increase in the awareness of issues that are related to mental health is witnessed, along with their effect on the person suffering from it. Early detection and timely treatment are considered crucial in effectively combating these mental health challenges, leading to a transformative paradigm shift.

The domain of mental health has been significantly influenced by notable advancements in science and technology. New approaches such as machine learning, big data analytics, natural language processing, and data analytics are being actively implemented and used for profound insights into an individual's mental health condition. These state-of-the-art tools empower researchers to acquire valuable information, driving significant progress in comprehending and managing mental health disorders. The accessibility and provision of mental health treatment have been completely transformed by the development of digital mental health interventions, such as smartphone applications, online counseling platforms, and virtual support communities [2]. The pandemic's unparalleled difficulties, including protracted times of loneliness, dread, and uncertainty, have significantly negatively influenced mental health [3]. The requirement of treatment for mental health-related issues and detecting them has become more efficient and adaptable in the crisis, which has caused a focus on mental health support networks.

Furthermore, the fast growth of social media platforms in recent years has fundamentally changed how we interact with one another, share information, and communicate [4]. Social media now has billions of users and is vital to our everyday life. Social importance and scholarly interest in the connection between mental health and social media are growing [5]. Social media plays a crucial role in detecting whether an individual is suffering from a mental health condition. It offers chances for interpersonal interaction, self-expression, and access to a variety of knowledge and tools about mental health. With the emergence of online forums and support groups, people dealing with mental health issues now have a place to belong and a forum to share their stories [6]. Looking at it differently, mounting data points to the possibility that social media use leads to be detrimental to mental health. According to studies, using social media in an unhealthy or excessive way might worsen symptoms of sadness, anxiety, loneliness, and low self-esteem [7]. These negative outcomes have been linked to numerous elements, including social comparison, cyberbullying, and ongoing exposure to curated and idealized portrayals of others' lives. First, it can guide initiatives and plans meant to encourage social media users' good mental health. Researchers and mental health providers can create some evidence-based recommendations for appropriate social media use by pinpointing the pathways through which social media affects mental well-being [8].

Nevertheless, there are still considerable gaps and difficulties in the realm of mental health. By assessing the present condition of mental health, finding new trends, and investigating creative methods to improve mental health assistance, this study seeks to add to the corpus of existing information [9]. The remaining paper is structured as follows: Section 2 presents an overview of the related studies conducted in the area. The methodology of the paper is presented in Section 3. The results of the study are presented and discussed in Section 4. Finally, Section 5 concludes the study.

1.2 Related Works

In recent years, anxiety and Post-Traumatic Stress Disorder (PTSD) have become serious issues in India's mental health. These illnesses are more common because of the nation's diversified population, growing urbanization, and sociocultural pressures [10] Post-traumatic stress disorder (PTSD) is caused by trauma, but is caused by many problems, including stress, stress, anxiety, and social anxiety. A recent study conducted in India revealed an alarming incidence of PTSD and anxiety disorders [11]. They believe that most of the Indian population suffers from PTSD and anxiety symptoms. Socioeconomic inequalities, urbanization, exposure to trauma, and cultural variables all increase the risk and progression of these disorders [12].

Anxiety and PTSD have a significant impact on people's daily activities and quality of life. These issues affect many areas of life, including relationships, productivity, and health. Lack of access to mental health services, lack of knowledge about mental illness, and cultural stigma and misconceptions about mental illness prevent people from receiving appropriate support and treatment [13]. In terms of available therapies, psychotherapy, especially cognitive-behavioral therapy (CBT), has shown promise in treating PTSD and anxiety. However, there aren't enough qualified mental health experts available, especially in rural areas [14].

Stakeholders including policymakers, mental health professionals, and the community are working to raise mental health awareness, reduce stigma, and increase access to mental health services as they recognize the urgent need for stress and PTSD in India [15]. Plenty of researchers from around the world have examined the public health mental crisis that occurred during COVID-19 [16]. A study done in China in 2020 found that over half of the participants had a moderate-to-severe psychological impact and that a third of the individuals experienced moderate-to-severe anxiety [17]. Several studies are being conducted to evaluate the pandemic's psychological effects on the Saudi population. For instance, during the Saudi Arabian lockdown at the height of the outbreak, Albagmi and his colleagues evaluated the prevalence of anxiety and associated factors. The poll received a total of 3,017 responses from Saudi Arabia's five major areas.

The findings showed that during the COVID-19 pandemic, 19.6% of the respondents experienced moderate to severe anxiety [18]. This research focused gradually on the Reddit dataset by considering the comments surrounded by mental health obstacles. The steps performed are as follows: the preprocessing of data is in Step1, in Step2, word embedding techniques were taken into consideration followed by TF-IDF vectorizer used in Step3 by each word embedding technique, in Step4 machine learning algorithms were applied (see Fig.1).

2 Methods

This research focused gradually on the Reddit dataset by considering the comments surrounded by mental health obstacles. As social media is a realm where people openly share their thoughts and emotions, social media comments are best chosen for consideration. The steps performed are as follows:

Step1
The preprocessing of data including tokenization, stemming, and lemmatization, was performed., The total data was divided into training and testing datasets.

Step2
Word embedding techniques like Word2Vec, Fast Text, and Bag of Words, were taken into consideration to be applied to the textual data. Three of them were applied to the dataset to examine the best out of them.

Step3
In this, after the Word embedding techniques the TF-IDF (Term Frequency – Inverse Document Frequency) vectorizer is used for attaining high accuracy and for deep understanding of the word

important for consideration. TF-IDF was one of the best vectorization techniques to be applied to the textual dataset.

Step4
After that, machine learning algorithms like Random Forest, Multinomial Naïve Bayes, and Logistic Regression them was applied to the dataset.

Step5
Furthermore, the accuracy of all the models was taken out for comparison of the techniques.

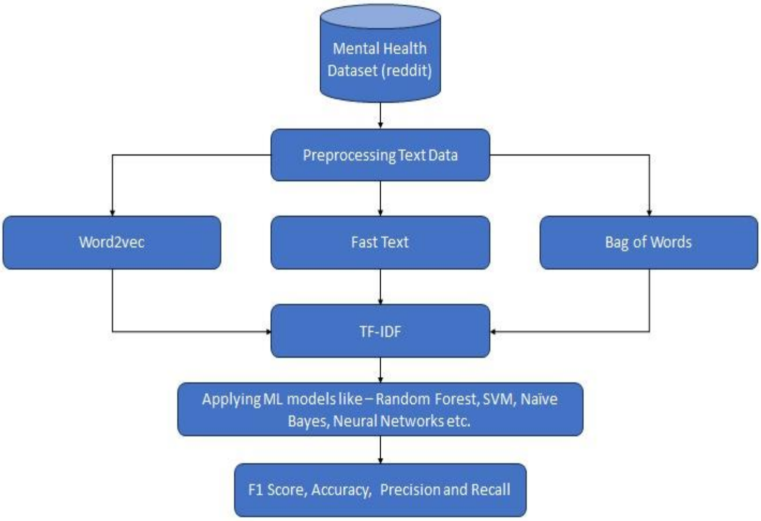
Mental Health Dataset (reddit)
Preprocessing Text Data
Word2vec
Fast Text
Bag of Words
TF-IDF
Applying ML models like – Random Forest, SVM, Naïve Bayes, Neural Networks etc.
F1 Score, Accuracy, Precision and Recall

Fig. 1. The framework of sentimental analysis.

2.1 Data Description

In this section, we focused on providing a detailed description of our dataset for a comprehensive understanding of our research. The dataset played a crucial role in the empirical analysis between different machine learning algorithms and word embedding techniques. The dataset was obtained from Kaggle. Kaggle is an optimal platform for data science and machine learning, and acknowledging the source is important for transparency. The dataset used in the study is 'Mental Health Subreddits' obtained from Kaggle (https://www.kaggle.com/code/lizakonopelko/mental-health-subreddits/notebook). The dataset used in our research was created by Liza Konopelko.

This 'Mental Health Subreddits' dataset provides information on the comments of Reddit apps that they are categorized to be in the class of 'Depression' or 'PTSD'. Social media apps are one of the greatest sources of retaining information which makes this dataset more pertinent to be used for analysis. The dataset consists of a training set and a testing set where the training set includes the data on which machine learning models are to be trained to contain the subset of the whole dataset like subreddit, text, and datasets labeled to depression or PTSD only, while the test set is used for evaluating the machine learning models. The dataset aims to develop models that can predict the optimal word embedding techniques, machine learning models, and vectorizers suitable for performing the sentimental analysis with the highest accuracy.

2.2 Data Preprocessing

The first step in our research involved collecting a large-scale dataset of user comments from diverse online platforms. The dataset encompassed a wide range of topics, ensuring the inclusion of various mental health indicators. To prepare the collected textual data for analysis, a series of preprocessing steps were employed. These steps included tokenization, which breaks down the text into individual words or tokens, and removal of stop words to eliminate commonly occurring, non-informative words. Additionally, stemming or lemmatization techniques were applied to reduce words to their base form, standardizing the text and aiding in feature extraction.

Feature extraction plays a vital role in transforming raw text into meaningful representations for predictive modeling. In our research, we employed various NLP techniques for feature extraction. Such techniques were the Bag-of-Words (BoW) model, the Fast text embedding model, and the Word2vec embedding model, which represents each comment as a vector of word frequencies or presence indicators. Term Frequency-Inverse Document Frequency (TF-IDF) was used to weigh the importance of each word in the corpus. In our study, the methodology and classification techniques can vary depending on the specific goals and requirements of the task Here are several approaches and categorization strategies that are frequently utilized in PTSD and anxiety predictive analysis.

2.3 Classification techniques

In our study, the methodology and classification techniques can vary depending on the specific goals and requirements of the task. Here are several approaches and categorization strategies that are frequently utilized in PTSD and anxiety predictive analysis.

Supervised Learning: In this method, a machine learning model is trained using labeled data in which each sample of the text is assigned a sentiment label. The model can categorize new, unlabelled data after learning patterns and characteristics from the classified data. Multinomial Naive Bayes, Logistic Regression, and Random Forest are common supervised learning methods used in sentiment analysis.

Multinomial Naive Bayes

The simplest classifier, Nave Bayes, performs quick classification on the label data. For instance, the Nave Bayes method is frequently used in phishing. On emails we received, the phishing filter classifier assigns a label of "phishing" or "Not Phishing." Nave Bayes Classifier builds machine learning architectures using pieces that are put together according to similarity and adhere to Bayes' law [19].

The Nave Bayes classification is represented mathematically in the following equations.

P(k|a) is equal to p(k/a) p(k)p(a). (1)

P(k|A) equals p(a1/k) p(a2/k). × · · · .. × p(an/k). (2)

Where P(k|a) in Eq. (1) means posterior probability, P(k) means class prior probability, P(a|k) means likelihood and P(a) indicates predictor.

Logistic regression

Logistic regression is a popular machine-learning algorithm used for binary classification tasks. It is a supervised learning algorithm that predicts the probability of an input from a given class. The main idea of logistic regression is to use a logistic function to model the relationship between input properties and binary output. The logistic function also called the sigmoid function, assigns values between 0 and 1 to all real numbers. This output can be interpreted as the probability that the input belongs to the positive class.

Random Forest

Random Forest is a popular machine-learning algorithm that is used for both classification and regression tasks. It is an ensemble method that combines multiple decision trees to make predictions. Each tree is built independently, and during the training process, it considers a random subset of features for each split. To make predictions using a Random Forest, each tree in the ensemble independently predicts the output, and the final prediction is determined by majority voting (for classification) or averaging (for regression) over all the individual tree predictions. They can handle large datasets with high-dimensional feature spaces, and they are resistant to overfitting.

XG Boost

Extreme Gradient Boosting, also known as XGBoost, is a potent machine-learning method renowned for its outstanding performance and scalability. It is a member of the family of algorithms known as gradient boosting, which consecutively trains a group of weak prediction models to produce a strong prediction model.

Gradient-based optimization

By repeatedly including weak learners in the ensemble, XGBoost improves the objective function. Each new learner's direction is determined using gradient information. Concerning the predictions generated by the current ensemble, XGBoost calculates the gradients of the loss function in each iteration. Then, to reduce gradients and enhance overall prediction, a new learner is fitted

3 Results

3.1 Evaluation Metrics

Precision:

It shows the effectiveness of the model in determining PTSD and anxiety support. Precision in this study refers to the degree to which the model can classify letter samples that represent characters, events, or emotions associated with PTSD and anxiety.

Precision is the percentage of true positive predictions compared to all positive predictions of the model.

$$\text{Precision} = \text{True Positives} / (\text{True Positives} + \text{False Positives})$$

Recall:

Consciousness is important for removing the most important examples of PTSD and stress from the literature. Retrieval assesses the model's ability to find and summarize the literature describing PTSD and anxiety-related behaviors in this study, ensuring the quality of attention to detail. High recall can lead to a greater analysis of thought patterns, suggesting a lower incidence of missing grades associated with PTSD and anxiety.

Return measures the ratio of the true probability of all good events in the data.

Recall = True Positives / (True Positives + False Negatives)

F1 Score:

The F1 score is a composite and regression analysis that provides a fair measure of the theoretical performance of the analysis algorithm. It takes into account the fact (correctness) of the model detecting relevant events and returns (ability) to capture a significant portion of current important events in documents. In this study, an emotional assessment model with a high F1 score will balance accuracy and recall and ensure accuracy and precision in the identification and classification of PTSD and personality pattern-related stress. It is especially important when the data is unequal, for example, one class has more instances than the other.

F1 Score = 2 * (Precision * Recall) / (Precision + Recall)

3.2 Results and Analysis

This section shows the results of our study, focusing on the key findings and outcomes derived from the analysis of the dataset. Table 1 shows the accuracy of the machine learning models with the respective word embedding techniques.

Table 1. Accuracies of all the machine learning models.

ML Models	Fast text + tf-idf	Word2vec	Bagofword + tf-idf
Random forest	70.512	55.089	69.112
Multinomial Naïve Bayes	73.201	70.437	60.861
xgboost	64.421	50.898	57.892
Logistic regression	79.471	53.293	68.501

As shown in Table 1, all three-word embedding techniques like BagofWords (BoW), Fast text, and Word2vec were tested along with the Machine learning models like Random Forest, Logistic Regression, Multinomial Naive Bayes, and XGBoost. The Fast text focuses on n-gram characters to perform better word embedding on textual data. Fast text easily associates the new words with the words based on the training dataset [20]. Hence, it had given the highest accuracy. BagofWords focuses on the frequency of individual words present in the text, the frequency of each word is independent of other words and lacks the focus on order or structure of the textual data hence, it lacks in understanding the meaning of the sentence. Word2vec creates a vector for each word and works based on vectors. The Tf-Idf vectorizer is widely used for weighing the schema of words based on their importance to determine the sentiment. A graphical representation of the results is also shown in Figure 2.

Logistic regression is the algorithm commonly used for classification and has been successfully used in a variety of applications [21-22]. It has also been regarded as one of the best algorithms for sentiment analysis to determine the sentiment from the textual data [23]. Multinomial Naive Bayes is a popular algorithm used for text classification tasks, particularly when dealing with discrete features like word counts or frequencies. Random Forest is a machine-learning model which consists of multiple decision trees. XGBoost is a gradient booster algorithm used for machine learning tasks.

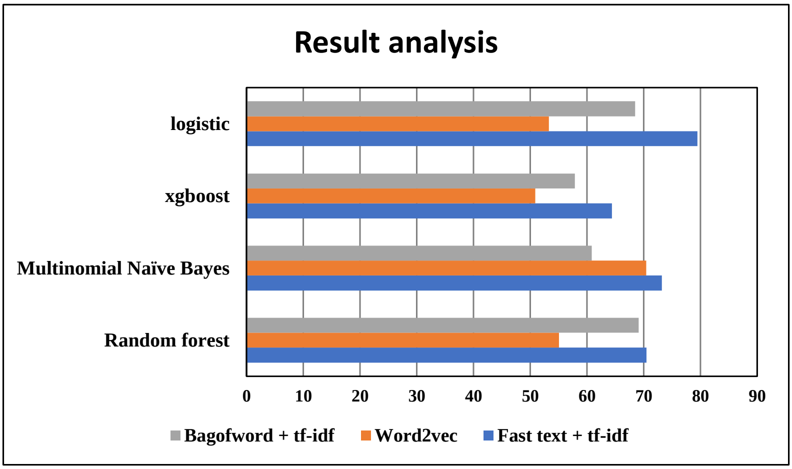

Fig. 2. Comparative result analysis

Fig.2 shows the accuracy of all the word embedding techniques applied with the machine learning models taken into consideration in this study. To get the best path for performing the empirical analysis on the textual dataset. The highest accuracy achieved by performing this analysis is 79.471%, and the lowest is 53.239%.

4 Discussion

Our study delves into the empirical analysis of mental health detection through machine learning algorithms applied to Reddit app comments. We aim to shed light on efficient diagnostic processes for mental health disorders, primarily focusing on Post-Traumatic Stress Disorder (PTSD) and anxiety.

The contemporary landscape has witnessed a remarkable surge in the importance of mental health, catalyzed by events such as the SARS-CoV-2 pandemic [1]. This heightened awareness has underscored the need for innovative strategies to comprehend and address mental health issues. In this context, our research contributes substantially by offering insights into efficient diagnostic processes. We strategically leveraged social media as the foundation of our analysis, acknowledging its dual role in mental health. It serves both as a source of mental health support and, conversely, as a potential exacerbator of mental health challenges [2].

Our study's cornerstone lies in the utilization of word embedding techniques to enhance the representation of textual data. We assessed FastText, Word2Vec, and Bag of Words, each offering a unique approach. Among these, FastText emerged as the most effective technique in our study, as it adeptly handles n-gram characters and associations with new words [20]. The effectiveness of our approach becomes more apparent when we explore the machine learning models applied to sentiment analysis. These models included Random Forest, Multinomial Naïve Bayes, Logistic Regression, and XGBoost. The importance of the right combination of techniques for effective sentiment analysis becomes evident. For instance, Logistic Regression, when coupled with the FastText word embedding method, exhibited remarkable accuracy (79.471%) [23].

Our findings align with previous research, affirming the multifaceted relationship between social media and mental health. Social media indeed plays a dual role, offering a platform for support and information sharing, while also potentially contributing to mental health challenges [4,5]. The influence of socioeconomic factors, urbanization, and trauma exposure on mental health, particularly in diverse populations like India, is well-documented [10,11]. Despite the notable advancements, significant gaps and challenges persist in the realm of mental health. Our study contributes significantly by offering data-driven insights into current mental health trends and patterns [9]. These insights have the potential to guide evidence-based recommendations for healthy social media use, thereby mitigating the negative effects associated with excessive or unhealthy online behavior [8].

5 Conclusion

In conclusion, the research conducted for analyzing the sentiment analysis using the dataset taken from Reddit, which consists of social media comments sheds light on the sufferings of people on online platforms. By using the word embedding techniques followed by the TF-IDF vectorizer and machine learning algorithms, the study aims to create an accurate and efficient model. The findings of this research have paved the way for a more precise approach to performing sentiment analysis.

Among the different paths explored, the combination of FastText, TF-IDF, and Logistic Regression emerged as the most suitable for sentiment analysis, yielding an accuracy of 79.471%. This outcome signifies the effectiveness of this model in accurately identifying and understanding the sentiments expressed in textual data.

Overall, the research on sentiment analysis has demonstrated the efficiency of machine learning models when applied to textual datasets, facilitating a better understanding of the sentiments expressed by individuals facing mental health issues. The insights gained from this study can contribute to the development of tools and techniques for analyzing sentiment in online platforms, ultimately aiding in the support and well-being of individuals experiencing mental health challenges.

References

[1] Reiss F, Meyrose AK, Otto C, Lampert T, Klasen F, Ravens-Sieberer U. Socioeconomic status, stressful life situations and mental health problems in children and adolescents: Results of the German BELLA cohort-study. PLoS One. 2019 Mar 13;14(3):e0213700. doi: 10.1371/journal.pone.0213700. PMID: 30865713; PMCID: PMC6415852.

[2] Fairburn CG, Patel V. The impact of digital technology on psychological treatments and their dissemination. Behav Res Ther. 2017 Jan;88:19-25. doi: 10.1016/j.brat.2016.08.012. PMID: 28110672; PMCID: PMC5214969Author, F., Author, S., Author, T.: Book title. 2nd ed. Publisher, Location (1999).

[3] Javed B, Sarwer A, Soto EB, Mashwani ZU. The coronavirus (COVID-19) pandemic's impact on mental health. Int J Health Plann Manage. 2020 Sep;35(5):993-996. doi: 10.1002/hpm.3008. Epub 2020 Jun 22. PMID: 32567725; PMCID: PMC7361582.

[4] Primack, B. A., Shensa, A., Sidani, J. E., Whaite, E. O., yi Lin, L., Rosen, D., et al. (2017). Social media use and perceived social isolation among young adults in the US. Am. J. Prev. Med. 53, 1–8. doi 10.1016/j.amepre.2017.01.010.

[5] Pantic I. Online social networking and mental health. Cyberpsychol Behav Soc Netw. 2014 Oct;17(10):652-7. doi: 10.1089/cyber.2014.0070. Epub 2014 Sep 5. PMID: 25192305; PMCID: PMC4183915.

[6] Popat A, Tarrant C. Exploring adolescents' perspectives on social media and mental health and well-being – A qualitative literature review. Clinical Child Psychology and Psychiatry. 2023;28(1):323-337. doi:10.1177/13591045221092884.

[7] Emily B. O'Day, Richard G. Heimberg, Social media use, social anxiety, and loneliness: A systematic review, Computers in Human Behavior Reports, Volume3, 2021, 100070, ISSN2451-9588,
https://doi.org/10.1016/j.chbr.2021.100070.(https://www.sciencedirect.com/science/article/pii/S245195882100018X).

[8] Vaingankar JA, van Dam RM, Samari E, Chang S, Seow E, Chua YC, Luo N, Verma S, Subramaniam M. Social Media-Driven Routes to Positive Mental Health Among Youth: Qualitative Enquiry and Concept Mapping Study. JMIR Pediatr Parent. 2022 Mar 4;5(1):e32758. doi: 10.2196/32758. PMID: 35254285; PMCID: PMC8933808.

[9] Wainberg ML, Scorza P, Shultz JM, Helpman L, Mootz JJ, Johnson KA, Neria Y, Bradford JE, Oquendo MA, Arbuckle MR. Challenges and Opportunities in Global Mental Health: A Research-to-Practice Perspective. Curr Psychiatry Rep. 2017 May;19(5):28. doi: 10.1007/s11920-017-0780-z. PMID: 28425023; PMCID: PMC5553319.

[10] Trivedi JK, Sareen H, Dhyani M. Rapid urbanization - Its impact on mental health: A South Asian perspective. Indian J Psychiatry. 2008 Jul;50(3):161-5. doi: 10.4103/0019-5545.43623. PMID: 19742238; PMCID: PMC2738359.

[11] Gilmoor AR, Adithy A, Regeer B. The Cross-Cultural Validity of Post-Traumatic Stress Disorder and Post-Traumatic Stress Symptoms in the Indian Context: A Systematic Search and Review. Front Psychiatry. 2019 Jul 4;10:439. doi: 10.3389/fpsyt.2019.00439. PMID: 31333512; PMCID: PMC6620607.

[12] Sareen J. Posttraumatic stress disorder in adults: impact, comorbidity, risk factors, and treatment. Can J Psychiatry. 2014 Sep;59(9):460-7. doi: 10.1177/070674371405900902. PMID: 25565692; PMCID: PMC4168808.

[13] Corrigan PW, Watson AC. Understanding the impact of stigma on people with mental illness. World Psychiatry. 2002 Feb;1(1):16-20. PMID: 16946807; PMCID: PMC1489832.

[14] Kar N. Cognitive behavioral therapy for the treatment of post-traumatic stress disorder: a review. Neuropsychiatr Dis Treat. 2011;7:167-81. doi: 10.2147/NDT.S10389. Epub 2011 Apr 4. PMID: 21552319; PMCID: PMC3083990.

[15] Wainberg ML, Scorza P, Shultz JM, Helpman L, Mootz JJ, Johnson KA, Neria Y, Bradford JE, Oquendo MA, Arbuckle MR. Challenges and Opportunities in Global Mental Health: A Research-to-Practice Perspective. Curr Psychiatry Rep. 2017 May;19(5):28. doi: 10.1007/s11920-017-0780-z. PMID: 28425023; PMCID: PMC5553319.

[16] Plenty of researchers from around the world have examined the public health mental crisis that occurred during COVID-19.

[17] Que J, Shi L, Deng J, Liu J, Zhang L, Wu S, Gong Y, Huang W, Yuan K, Yan W, Sun Y, Ran M, Bao Y, Lu L. Psychological impact of the COVID-19 pandemic on healthcare workers: a cross-sectional study in China. Gen Psychiatr. 2020 Jun 14;33(3):e100259. doi: 10.1136/gpsych-2020-100259. PMID: 32596640; PMCID: PMC7299004.

[18] Albagmi FM, AlNujaidi HY, Al Shawan DS. Anxiety Levels Amid the COVID-19 Lockdown in Saudi Arabia. Int J Gen Med. 2021 May 31;14:2161-2170. doi: 10.2147/IJGM.S312465. Erratum in: Int J Gen Med. 2022 Apr 19;15:4091-4092. PMID: 34103971; PMCID: PMC8180301.

[19] Awasthi, A., Goel, N. Phishing website prediction using base and ensemble classifier techniques with cross-validation. *Cybersecurity* **5**, 22 (2022). https://doi.org/10.1186/s42400-022-00126-9.

[20] Asudani, D.S., Nagwani, N.K. & Singh, P. Impact of word embedding models on text analytics in a deep learning environment: a review. *Artif Intell Rev* (2023). https://doi.org/10.1007/s10462-023-10419-1

[21] Kaur I, Kapoor N.: Token based approach for cross project prediction of fault prone modules. In 2016 International Conference on Computational Techniques in Information and Communication Technologies (ICCTICT), pp. 215-221. IEEE, New Delhi (2016).

[22] Doja MN, Kaur I, Ahmad T. Age-specific survival in prostate cancer using machine learning. Data Technologies and Applications 54(2), 215-34 (2020).

[23] Mohammad Aman Ullah, Syeda Maliha Marium, Shamim Ara Begum, Nibadita Saha Dipa, An algorithm and method for sentiment analysis using the text and emoticon, ICT Express, Volume 6, Issue 4, 2020, Pages 357-360, ISSN 2405-9595, https://doi.org/10.1016/j.icte.2020.07.003.