



# A Survey on Psychological disorders and threats detection using the Twitter dataset

**Name of 1<sup>st</sup> Author:** Dyanaraj N Thorat, **Designation of 1<sup>st</sup> Author:** Student

**Name of 2<sup>nd</sup> Author:** Shubham C Jadhav, **Designation of 2<sup>nd</sup> Author:** Student

**Name of 3<sup>rd</sup> Author:** Pranav P Joshi, **Designation of 3<sup>rd</sup> Author:** Student

**Name of College/Institute:** Sinhgad Academy of Engineering, Pune

**City:** Pune, **Country:** India

**Name of Guide:** Prof. Suvarna Bahir

**Abstract:** This One in four people within the world is going to be suffering from mental or neurological disorders in their lives. According to a study 450 million people currently suffer from mental or neurological conditions, placing mental disorders among the leading causes of ill-health and disability worldwide. (WHO) [1]. The number of individuals suffering from the mental disease is on the rise and with it the burden on health and social care use, also because of the loss of both productivity and quality-adjusted life-years. Social media platforms became a source of 'in the moment' daily exchange, with topics including well-being and psychological state. An increasing number of social networking sites makes people more engaged in their virtual life quite ever and at an equivalent time, the quantity of knowledge people put online is gigantic and also heaven for researchers conducting their research. People tend to place their thoughts online to share with the entire world which also includes suicidal thoughts. Teens are engaged considerably during a virtual world of social media like Twitter which we are going to be covered in this survey. There are various techniques, studies conducted on this topic but we are focusing on the following four topics Mental health characterization, Sentiment analysis, Cyberbullying, suicide detection.

**IndexTerms - mental disorders, Suicide, Cyberbullying, sentiment computing, Text Mining, Machine Learning, Twitter.**

## I. INTRODUCTION

Mental and substance use disorders are the leading explanation for years lived with disability worldwide. Social media is changing the way people self-identify as suffering from a disorder and how they communicate with others.[4] There are nearly 2.8 billion active social media users globally, of which 2.5 billion can access their accounts on mobile devices. Suicide is the second leading explanation of death globally among people 15 to 29 years of age. Almost 800,000 people die due to suicide annually. A prior suicide attempt is that the most vital risk factor for suicide within the general population.[2] studying popular social media platforms such as Twitter holds the key to understanding what concerns patients, rather than clinicians, most, say researchers. Twitter is a growing social media site associated with cyberbullying and suicidal activities. Also studying Sentiment analysis (SA) is a strategy used for self-detection of viewpoints combined in written text. SA has been mainly used to classify text to determine the writer's stand/alignment towards a certain topic. When posting a status on the internet, people often use dialects of the language. So, in this survey, we are going to cover different papers mainly on four topics: Mental health, sentiment analysis, depression detection, cyberbullying. Twitter currently has 353 million monthly active users [3]. This is why we are focusing on research papers that are using Twitter as their dataset.

### Paper 1:

Sentiment analysis (SA), is a strategy used for the self-detection of viewpoints combine in written text. SA has been mainly used to classify text to determine the writer's stand/alignment towards a certain topic. People often use language dialects over the internet to post a status. Understanding the dialect in which a text snippet is written is crucial for insight into what has been written. This work focuses on comparing different techniques on a million records of tweets to classify which contains positive sentiments and which contains negative ones.

### Paper 2:

The purpose of the approach is to decompose each post in terms and compare them automatically to predefined suspicious terms by similarity distance calculation. The proposed approach is based on the calculation of a distance to distinguish suspicious posts for detecting suspicious profiles within social media. It is intended to identify more significant suspicious profiles and also to improve the system to improve security.

Paper 3:

Mental and substance use disorders are the leading cause of years lived with disability worldwide. Social media is changing the way people self-identify as suffering from a disorder and how they communicate with others. Studying popular social media platforms such as Twitter holds the key to understanding what concerns patients, rather than clinicians, most, say researchers. The researchers propose an approach to automatically identify posts related to mental health on social media.

Paper 4:

There are nearly 2.8 billion active social media users globally, of which 2.5 billion can access their accounts on mobile devices. Suicide is the second leading cause of death globally among people 15 to 29 years of age. Close to 800,000 people die due to suicide every year. A prior suicide attempt is the most important risk factor for suicide in the general population. It is a misconception that suicide and depression affect mostly the poor.

About half of the world's population is now living on the internet. Twitter is a growing social media site associated with cyberbullying and suicide.

**II. LITERATURE SURVEY**

According to the author **George Gkotsis**, social media platforms have become a source of 'in the moment' daily exchange, with topics including well-being and mental health. In their study, they analyzed posts from the social media platform Reddit and developed classifiers to recognize and classify posts related to mental illness according to 11 disorder themes. Using a neural network and deep learning approach, they could automatically recognize mental illness-related posts in our balanced dataset with an accuracy of 91.08% and select the correct theme with a weighted average accuracy of 71.37%.

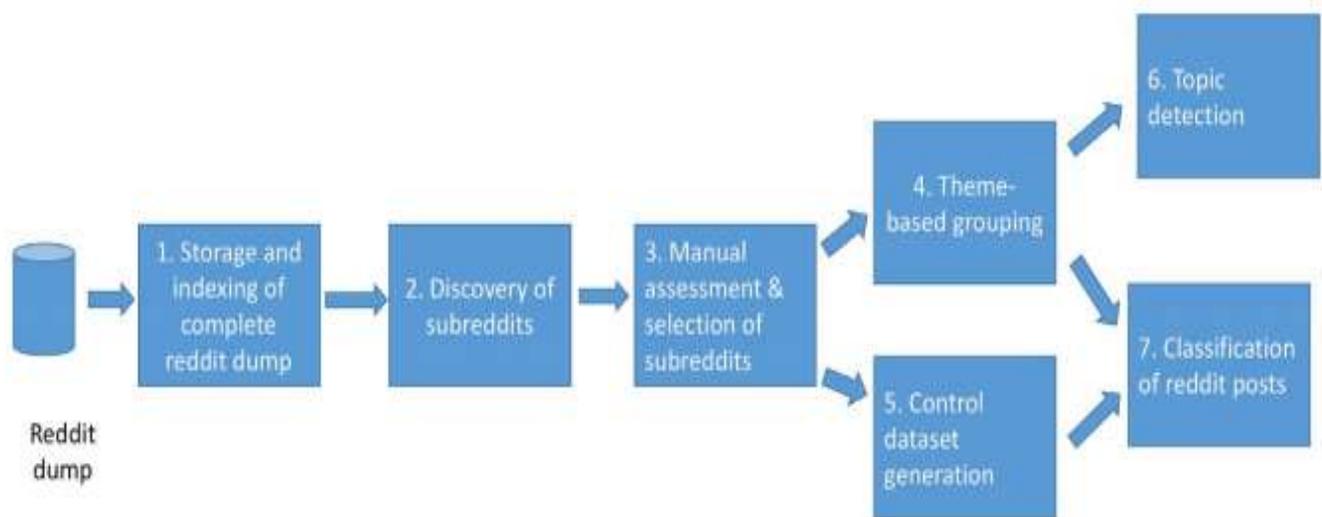


fig.: 1

There are some Limitations also they found that some of the themes are highly interrelated and not always distinguishable as separate and exclusive classes. Another limitation of their approach is the number of mental health themes that are used in the multiclass classification task. [4]

According to the author Firoj Fattulal Shahare, Nowadays social data is growing at a rapid pace, it plays a significant role in every aspect of life. The current web world range of users uses social media and social networks to browse and read news-connected information. They tend to propose a replacement methodology to try and do the drought sentiment analysis for news data. Their proposed method uses a Bayes and Levenshtein algorithm to determine the emotion into different social media news data categories.

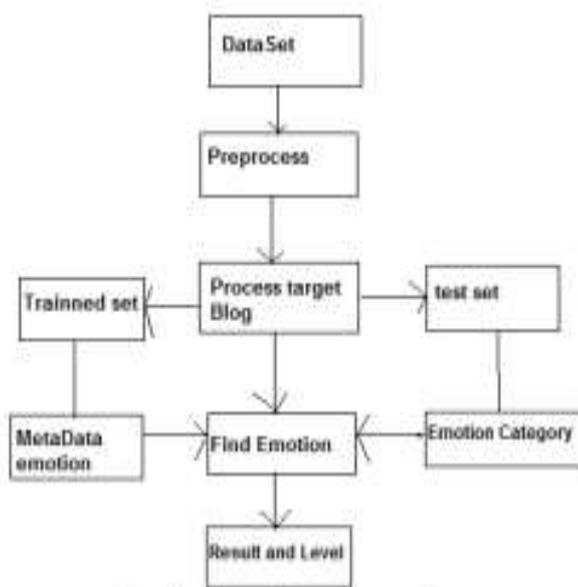


fig.: 2

With the help of their algorithm, they can easily distinguish an emotional text from news events and data. Emotion is not easy to find from large data because users use several texts to define an emotion but it's not easy to define in a category, but their approach uses a specific way to find this text from a larger amount of data.[5]

According to the author Ghelmar Astoveza, their study aims to classify suicidal tweets using Artificial Neural Network (ANN), and the data collected were annotated as "non-risky" or "risky," and was classified correctly by a trained classifier. It was found in this study using unigrams and bigrams to describe a complete process of classifying tweets. The data gathered was then put in an excel file for data labeling to increase the accuracy of the model. Their study found that 91% of the non-risky tweets were classified correctly and only 9% misclassifying 9% of those labeled as 'risky' It was also found that 65% of risky tweets were labeled correctly and 35% were misclassified as 'unsafe' and 65% were labeled as 'unacceptable'. [2]

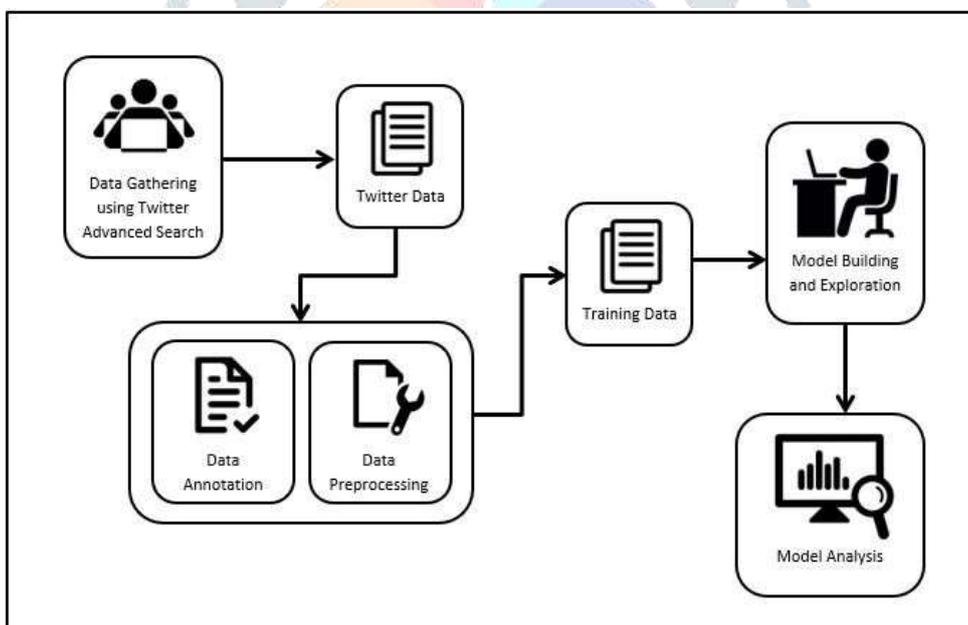


fig.: 3

According to the author Salim Alami, exponential development in online social media allows users around the globe the possibility to share and communicate information and ideas freely. So to develop user profiles of cybercriminals, text mining techniques are an effective way to detect and predict criminal activities in microblog posts. This emerging media has become a dominant communication tool and it has been used as a communication channel in several events, especially "The Arab Spring" and the Boston Attack. The evaluation of their proposed approach is done within real posts.

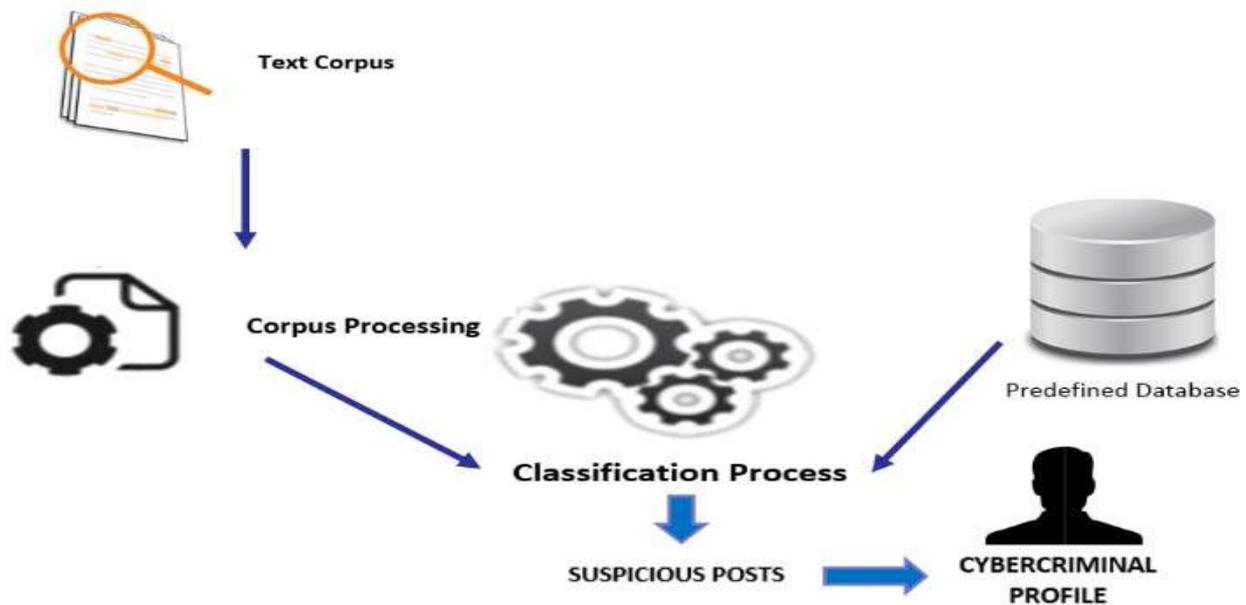


fig.: 4

They Conclude that the proposed approach is based on a calculation of a similarity distance to detect and predict criminal activities. The purpose of this approach is to decompose each post in terms and compare them automatically to a predefined suspicious terms database. For future work, they are planning to improve the system in terms of execution time, developing new scoring methods, and using other knowledge resources to improve precision rates. According to the author, their proposed solution presents an idea of a global research project including an automatic system for detecting suspicious profiles in the social media profiles, through which we can uncover suspicious behavior.[6]

### III. DISCUSSION

Although the proposed methods can be a proof-of-concept for similar implementations and have shown promising results, we believe that the results' accuracy could be even greater. In evaluation using real-world data, when the proposed method shows 91% accuracy, it means that based on the matching process, on average, in 91% of the cases the model has achieved 100% relevant results, which are the users who have self-report statements. This is in a situation where we did not limit the user domain to a small community (as was done in many studies on Reddit and Facebook). In addition, there were no other user's health records (e.g., personality traits) available for validation except user statements.

Also, we randomly sampled a total of 200 profiles (100 from each class) and asked three experts<sup>15</sup> to determine which profile could be considered as a reference to depression symptoms following the DSM-5 criteria. Each profile was assigned with one of the following labels: 1) no depression reference is expressed, 2) one or more depression references are expressed, and 3) unable to make a judgment. Then we chose the most agreed one. Accordingly, we found that in many retrieved cases the system retrieved profiles that might have been prone to depressive patterns of thinking but had not (yet) reported the self-statements of depression. This indicated that if we were less strict in the evaluation phase, the overall accuracy would be significantly improved.

On the other hand, the results of the analysis confirm that the automatic data collection can be a practical, cost-effective, and less time-consuming approach than the traditional ways. This method can also be used for similar issues (such as suicide prevention frameworks or investigation of other mental disorders) with a slight modification. We also point out that in some cases the combination of features will improve the evaluation results, but there has to be a trade-off between increasing metrics and the features involved. For example, after examining the cross-combinations, in some cases, combining three features leads to about a 1% increase in accuracy, which is not an optimal choice due to the feature dimensions, and using lightweight approaches (like what was described in the previous section) is a more appropriate choice.

### IV. CONCLUSION & FUTURE SCOPE

The study of mental health and its disorder and its impact on society is important. From the survey, we see that day by day active users on social media are increasing, and people are expressing their emotions on social media so it is necessary to study the data and design a system to detect their mental health condition in early stage and take actions accordingly. Which results in saving lots of lives. By combining Mental health characterization, Sentiment analysis, Cyberbullying, suicide detection we can create a single system that will work for all mentioned conditions and can improve the time complexity.

This study aims to detect depressed users on Twitter using self-report diagnosis. We first tried to provide a road map for mental disorder prediction via social data mining and then present a new framework by automatically collecting the data from the diagnosed and control user's profile. We extracted the desired features from the dataset and applied the preprocessing steps. Then we examined the distribution of data in defined categories to identify the differences between two groups and replicate previous findings. We tried to characterize the connection between depression and language use through lexicon analysis and NLP techniques, and we have also introduced a set of new features that have not been further explored in previous research, including bio-text and features resulting from the analysis of user profile picture and banner image. We reduced the feature size to identify the effective ones by correlation analysis. Afterward, we applied nine classification models on the features and compared the results using three

evaluation metrics: accuracy, F1-score, and area under the ROC curve. Thus, our findings suggest a relationship between depression symptoms and almost all the studied features, but tweets word bigrams and bio-character four-grams were identified as two important ones that in cross-validation using Catboost/GB model and MLP achieved the accuracy of 91% and 83%, respectively. Also, the F1-score of these two features was 0.89 and 0.82, which outperform the reported results with LR and SVM (common classifiers in the literature). Also in the case of dimensionality reduction techniques, in an alternative approach, we used SVD for feature selection, which led to the smaller set of features, and by comparing the implementation results, we showed that the correlation-based method leads to better outcomes. Furthermore, due to the importance of the interpretation of features in this study, the correlation-based method was a more proper choice.

We believe that the proposed mechanism can be implemented for other mental disorders with a slight change in the initial filtering phase. This framework can also be used as a lightweight method in the form of clinical decision support systems to facilitate diagnosis decisions or as suicide and self-harm prevention tools in social network platforms. We mentioned in the discussion section that if we were less strict in the evaluation, the results could be improved. In addition, although we tried to use all the important features in the user profile, there is other information that we believe can increase the accuracy of the diagnosis. This information includes hashtags and context information (e.g., tweets time, which showed a meaningful difference in a study on Sina Weibo [63]). For this reason, we also stored this information in the generated dataset to be used for future studies.

Finally, it is good to point out that behavioral/derived data are not limited to reported statements and profile information but includes all the activities and actions performed by the user. Such as the content that the user normally follows the games that he plays on the social network, or even the time he spends on the platform. In the presence of clinical information and patient approval, some of this information can be explicitly recorded and collected by the API or the crawler, and others implicitly by developing the application on the platform or installing add-ons in the user's browser. The development of such tools and the study of the collected data can create a new generation of analyses and open new gates to social network analysis in the field of diagnosing the user's mental state. Accordingly, the use of clinical information, semantic similarity techniques, transformer-based machine learning models, and more image analysis approaches are among the issues that will be addressed in our future work.

## V. REFERENCES

- i. <https://www.who.int/news/item/28-09-2001-the-world-health-report-2001-mental-disorders-affect-one-in-four-people#:~:text=One%20in%20four%20people%20in,ill%2Dhealth%20and%20disability%20worldwide>.
- ii. G. Astoveza, R. J. P. Obias, R. J. L. Palcon, R. L. Rodriguez, B. S. Fabito and M. V. Octaviano, "Suicidal Behavior Detection on Twitter Using Neural Network," TENCON 2018 - 2018 IEEE Region 10 Conference, 2018, pp. 0657-0662, doi: 10.1109/TENCON.2018.8650162.
- iii. <https://www.statista.com/statistics/282087/number-of-monthly-active-twitter-users/>
- iv. Gkotsis, G., Oellrich, A., Velupillai, S. et al. Characterisation of mental health conditions in social media using Informed Deep Learning. Sci Rep 7, 45141 (2017).
- v. F. F. Shahare, "Sentiment analysis for the news data based on the social media," 2017 International Conference on Intelligent Computing and Control Systems (ICICCS), 2017, pp. 1365-1370, doi: 10.1109/ICCONS.2017.8250692.
- vi. S. Alami and O. Elbeqqali, "Cybercrime profiling: Text mining techniques to detect and predict criminal activities in microblog posts," 2015 10th International Conference on Intelligent Systems: Theories and Applications (SITA), 2015, pp. 1-5, doi: 10.1109/SITA.2015.7358435.