

# Leveraging Stacking Model to Identify Depression

## A Hybrid Approach

Sonal Singh

Data Scientist

Fractal Analytics Bengaluru, India

**Abstract**—Depression has become one of the most common mental illnesses, and the widespread use of social media provides new ideas for detecting various mental illnesses. At present, the existing research mostly uses a single detection method. This paper proposes to use machine learning technology to detect users of depressive patients based on user-shared content and posting behaviors on social media. This paper proposes feature selection and ensemble strategy using stacking model for detecting depressed users. This method is a data-driven, predictive approach for early detection of depression.

**Keywords**—*Social Media Analytics, Depression Detection, Machine Learning (ML), Support Vector Machine (SVM), Naive Bayes, Decision Tree, Feature Selection*

### I. INTRODUCTION

Depression, or depressive disorder, is a common disease. According to the World Health Organization (WHO), the number of people with depression was estimated at more than 300 million affected worldwide. Depression may severely impact Ill-being and functioning at work, school, and family, and can even lead to self-harm. While we feel moody, sad or low from time to time, few people encounter these emotions seriously, for drawn out stretches of time (weeks, months or even years) and in some cases, with no apparent reason. Depression is something other than a low state of mind—it's a genuine condition that influences someone's physical and emotional feelings Adolescent depression is associated with mood disorders and severe mental illness in adult life Nearly 1.8 million people die from suicide each year and suicide are the fourth leading cause of death in 21–29-year-olds, according to WHO. Amongst the top major diseases causing disability or incapability, five are mental illnesses -depression being the most prominent of these. Hence, the disease burden due to depression is vast. The prevalence of depression in the adult population is approximately 5% across cultures, and 20% in its milder forms (i.e., partial symptoms, mild depression, and probable depression). Among adults, those most at risk are within the middle-aged population. Also, the world-wide occurrence of depression is increasing, with a rise of 28% between 2010 and 2019. However, early professional intervention can improve mental symptoms (e.g., absence of self-confidence and rumination) and resolve somatic problems (e.g., gastrointestinal problems and sleeping disorders) in most of the cases.

Early detection of depressive symptoms followed by assessment and treatment can considerably improve chances for curbing symptoms and the underlying disease; mitigate

negative implications for Ill-being and health as Ill as personal, economic, and social life. However, detection of depressive symptoms is challenging and resource demanding. Current approaches are mainly based on clinical interviews and questionnaire surveys by hospitals or agencies where psychological evaluation tables are utilized to make predictions on mental disorder.

This approach is mostly based on one-to-one questionnaires and can roughly diagnose the psychological disorder for depression. An alternative approach to interview or questionnaire-based predictions of depression is the analysis of informal texts provided by users. Previous studies in clinical psychology have shown that the relationship between the user of a language (e.g., speaker or writer) and their text is meaningful and has potential for the future. A recent study by Havigerova et al. indicate a potential for text-based detection of persons at risk for depression, using a sample of informal text written about a holiday. Hence, online records are increasingly seen as a valuable data source in supporting health care with decision support. The approach to identify depression symptoms from informal texts is promising, as it allows for benefitting from recent advances in natural language processing and Artificial Intelligence (AI). AI applied for natural language processing employs linguistics and computing techniques to help machines to understand underlying phenomena such as sentiments or emotions from texts. In that case, the core intent is to analyze opinions, ideas, and thoughts via the assignment of polarities either negative or positive

### II. LITERATURE SURVEY

With the gradual increase in social media usage and the extensive level of self-disclosure within such platforms, efforts to detect depression from Social Media data have increased. Park et al. indicated that depressed Social Media users tend to post comments containing negative emotional sentiments more than healthy users. In addition, De Choudhury et al. found that depressive signals are noticeable in comments posted by users with major depressive disorder. Thus far, different features have been used to detect depression from Social Media data. De Choudhury et al. collected more than 2 M comments from 476 users who are clinically diagnosed as depressed and had Social Media accounts. They used behavioral attributes related to social engagement, emotion, language and linguistic styles, ego network, and mentions of antidepressant medications to build a classifier that provides estimates of the risk of depression.

They leveraged these distinguishing attributes to build an SVM classifier that can predict the risk of depression with 70% classification accuracy. Tsugawa et al. [13] revealed that frequencies of word usage, along with topic modeling, are useful features for the prediction model. Using the radial kernel SVM classifier, they obtained 69% classification accuracy in predicting depression of 81 participants out of the 209 collected using a questionnaire. In addition, Reece et al. [9] extracted predictive features for measuring the effect, linguistic style, and context from users' comments; built models using these features with supervised learning algorithms, and successfully discriminated between depressed and healthy contents. Their data is collected from 105 out of the 204 depressed users and the CESD scores relied on the identification of depressed users. The best classifier performance was obtained using a 1200-tree random forest classifier, increasing the precision to 0.866, compared to other study results. Nadeem et al. [10] utilized the bag-of-words approach for better depression detection, which uses word occurrence frequencies to quantify the content of a measured on a document level. They employed four types of binary classifiers: linear SVM classifier, decision tree (DT), Naïve Bayes (NB) algorithm, and logistic regressive approach, and found that NB outperformed other classifiers with an accuracy of 81% and precision of 0.86. They used a corpus of more than 2.5 M comments gathered from the Shared Task organizers of CLPsych 2015, online, from users who indicated they are diagnosed as depressed (326) or with PTSD. On the other hand, Nadeem et al. [10], Coppersmith et al. [3], and Moir et al. [2] considered sentiment analysis as a feature to detect depression from Social Media data. Jamil et al. [14] concluded that the use of sentiment analysis, along with percentage of depressed comments, increases the precision and recall of detecting depression. The classifier was trained on 95 users who disclosed their own depression (which was equal to 5% of users participating in the study, while the remaining 95% are healthy users), using SVM, which provided a recall of 0.875 and precision of 0.775. De Choudhury et al. [8] and Jamil et al. [14] used the benefits of depressed people comments for extracting features that helped increase the detection accuracy. De Choudhury et al. [8] built a depression lexicon of terms that are likely to appear in postings from individuals discussing depression or its symptoms in online settings. In contrast, Jamil et al. [17] used the percentage of depressed comments, along with self-indication of depression, to decide whether a user should be removed from the training set and found this feature to increase the model's accuracy.

### III. PROPOSED HYBRID APPROACH

There are many researches going on in this area to help the mankind. There are still many unsolved challenges to online depression detection methods. First, compared with Twitter and Facebook, the detection of depressed users in Chinese communities is far from adequate. Second, the classification model used in many current studies is relatively limited. For different feature sets or data sets, the performance of the classifier is unstable, and there is a lack of research on the classification model. Therefore, this paper constructs a hybrid

classification model suitable for identifying depressed users in social media to achieve better performance effects.

#### A. Dataset

The dataset includes depressed users and other users, among which the depressed users are a sample collected under the microblogging supertopic "Depression," which clusters a large number of active depressed patients under the community. In real life, depressed patients are often isolated or even rejected by others, whereas studies have shown that people with similar intrinsic distress chooses to cluster together on social media. Therefore, our strategy for selecting depressed users is to search for user IDs posted on that topic, which can improve the efficiency of collecting samples of depressed users. For other user IDs, I chose to collect them under other active and living super topics, such as "photography" under the photography section, "everyday" in the daily section, learning "study account" under that section, and "gourmet" in the food section. At this stage, all information from the user's personal interface is crawled once according to the collected user ID, including 315 depressive users and 562 other users. After cleaning the data, I collected enough depressed user IDs and other user IDs and then obtained the personal interface information according to their user ID, including the number of users' fans, the number of followers, the number of posts, the post content, whether it is original content, post time, location information, publishing tools and other information fields. The final data sets of depressed users and other users are 130 and 320, respectively, and the numbers of microblog posts are 42,827 and 293,102, respectively. Every algorithm involves many different Hyper-parameters, I used the method of network searching to determine the best combination of parameters. In the model, 70% of the data is used for training and 30% for testing, and the optimal parameter combination of each model was determined by the method of 10-fold cross-validation.

#### B. Feature Engineering

The feature engineering pipeline is the preprocessing steps that transform raw data into features that can be used in machine learning algorithms, such as predictive models. Predictive models consist of an outcome variable and predictor variables, and it is during the feature engineering process that the most useful predictor variables are created and selected for the predictive model. Previous studies have proven that many functional features are very effective in distinguishing depressed patients from normal users, such as personal pronouns, negative words, and the patient's posting time. Combined with the existing research, microblog platform function and data field structure, feature engineering is divided into two aspects: text function and post behavior. In the part of speech features and personal pronouns section, I have added unique new features. At the same time, I extracted text features in the form of statistics to eliminate the influence caused by the large differences in the total number of posts among users. Specific data processing process is shown in Figure 1.

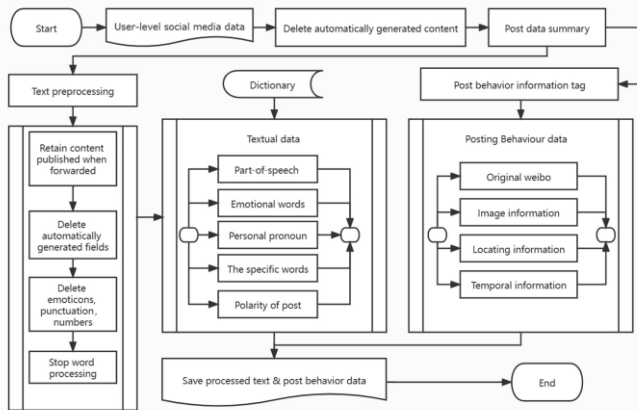


Figure 1

### Text features

As the part of speech is the basic grammatical attribute of words, part of speech, tagging is also the foundation of natural language processing, so the accuracy of tagging sets determines the results of subsequent classification prediction. In order to improve the accuracy of word segmentation, I use Wikipedia based anti-depression ([http://en.wikipedia.org/wiki/List\\_of\\_antidepressants](http://en.wikipedia.org/wiki/List_of_antidepressants)) build a dictionary list of drugs for depression to improve the accuracy of word segmentation. At the same time, some common network terms and abbreviations have also been added to the dictionary. In addition, most studies use the part of speech of LIWC text to mark, while LIWC is mainly oriented to English text information. Abnormal emotional preference is one of the important symptoms of patients with depression. The frequency of emotional words of different categories can be used as important information to distinguish different users. According to the Emotion Dictionary of Dalian University of Technology (Xu et al., 2008), seven kinds of fine-grained emotions that appeared in user posts I counted, namely, happy, like, anger, sad, fear, surprise and disgust. Previous studies have shown that personal pronouns reflect the psychological distance between depressed users and others (Vedula and Parthasarathy, 2017). Here, I added 'others' as a personal pronoun feature, which I believe to be valid based on previous research. Specific words mainly include negative words and the frequency of interrogative words. In previous studies, the number of negative emotional posts was usually calculated to identify the depression tendency of users (Sadeque et al., 2018), while the proportion of posts with different polarities was used to distinguish them in this study. Because the catharsis of negative emotions is not completely the same as depression, only when the negative posts reach a certain proportion can the painful mental state of the users be reflected. Here, I chose to use the text sentiment analysis API of the Baidu Intelligent Cloud Platform to mark the polarity of all original posts.

### Understanding Behavioral Attributes with Posts

A study on Facebook found that non-original posts in users' status updates are strongly associated with depression (Chen et al., 2020). I separately counted the percentage of each user's original posts in the total number of posts and used it

to reflect the user's posting habits. By observing the post behavior of different users on the platform, it was found that patients with depression prefer to express their feelings and mental status through words and less visual information (e.g., pictures). In addition, based on the functional characteristics of the microblog platform, I increased the posting frequency of users to display location information. The posting time is considered to be a reflection of the user's daily schedule, so I will compare the posting rates of users at different times which are every 6 h in a 24-h day and time periods with intervals. Early in the morning is the peak of depression, so late at night, depressed users post more frequently than other users. In addition, depressed users rely heavily on social media to vent their pain, and whether it is a workday or not may significantly affect the frequency of the patients' posts, so I believe this feature will be effective in providing useful information.

### C. Proposed Model

In order to improve the recognition ability of depressed users in online social media, this paper proposes a recognition framework of hybrid feature selection and ensemble learning strategies. In the feature selection stage, a feature selection method combining multiple modes is designed. This method uses recursive elimination method and extremely randomized trees method to get feature importance, combines mutual information method to get the importance score of each feature, and eliminates the feature with lower score after fusion to get the best feature subset. In the classification stage, a subset of the identified features is input into the stacking ensemble model to classify. The stacking which is being used is nothing but an ensemble machine learning algorithm that learns how to best combine the predictions from multiple well-performing machine learning models. Accuracy, f1-measure, precision and recall are used as evaluation criteria in the experimental phase, and the robustness of the model was measured by a 10-fold cross validation. Figure 2 shows the proposed method identification framework, which consists of the following three phases:

Dataset processing stage: First, the collected data is preprocessed to extract functional features from two perspectives of user's text and posting behavior information, and then convert them into data formats suitable for analysis.

Feature selection stage: In order to avoid the impact of high dimensional sparsity of original data on the classification effect of the model, this paper adopts a hybrid feature selection method, which combines the advantages of wrapper, filter and embedded methods to eliminate the feature vectors with low relevance or redundancy in the identification of depressed users.

User classification stage: In order to improve the identification

accuracy of depressed users, I use naive bayes, k-nearest neighbor, regularized logistic regression and support vector machine as basic learners, and use simple logistic regression algorithm as combination strategy to establish an ensemble model.



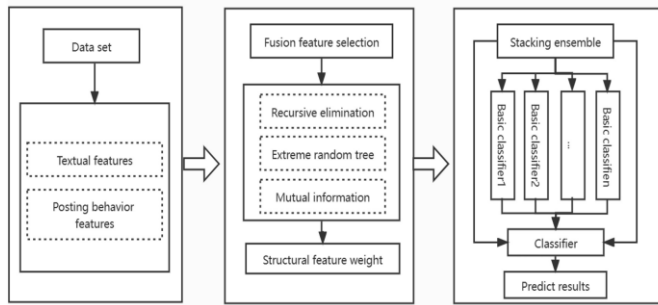


Figure 2

### 1. Mixed Feature Selection

Feature selection methods can apply search techniques to obtain new feature subsets from a given data set and evaluate the scores. Feature selection can remove irrelevant and redundant features, eliminating the disaster caused by the dimension curse. Reducing the functional space not only helps to build more accurate prediction models, but also reduces training time and improves model generalization ability. In practice, I cannot determine the only optimal feature subset, and different feature subsets may produce the best classification effect for different machine learning algorithms. Therefore, the feature selection technology has been the focus of researchers' attention, currently mainly including three types of methods: wrapper, filter and embedded. The wrapper uses a machine learning algorithm to select a subset of features and iterate over all remaining features each time. The filter method uses statistical techniques to evaluate the relationship between features and target values, independent of machine learning algorithms, and classification performance can be used as evaluation criteria for selecting feature subsets for wrapper method. Compared with the wrapper method, the embedded method enables feature selection and algorithm training to be carried out at the same time, and the computational cost is lower, but it is not easy to obtain higher performance. The recursive feature elimination method (RFE) based on random forest classifier is selected for wrapper method, which has the ability to select predictive variables with higher accuracy. The mutual information method is one of the most important methods in the filter method and can be used as a measure of the interdependence of variables. In the field of machine learning, the explanatory nature of tree far exceeds that of complex models such as neural network, so I choose extremely randomized trees to calculate the importance score of characteristics.

A single feature selection method may ignore other potential information contained in the original feature set, thus affecting

the classification effect. I chose to combine the advantages of multiple feature selection methods and apply each feature selection method to all feature sets. By excluding the 20% of features with low scores, I obtained the best feature subset, including a total of 42 features. Figure 3 shows the importance of features as determined by our feature selection strategy. It can be observed from the figure that features with significant influence on depression users' recognition include negative words, first-person singular, second person plural, and interrogative words.

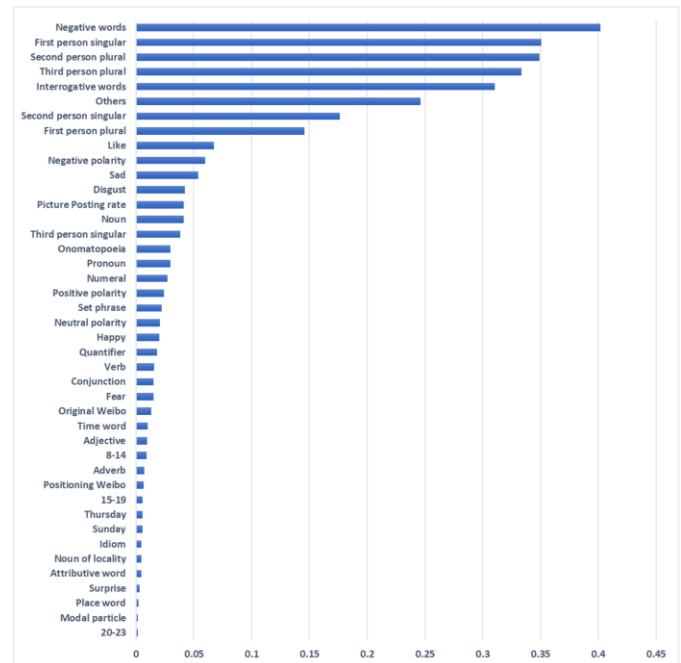


Figure 3

### 2. Ensemble Learning and Stacking Model

The core idea is to build multiple basic learning models, and use the output information of the basic prediction models to combine into a more powerful prediction model for final decision-making. The stack ensemble is a framework of layered combinatorial models. To be more precise, there are usually two stages in stack integration. The first stage consists of several basic models, which are trained separately on the training set, thus establishing the first level of prediction in the stack system. The predicted values for this phase are then collected as a data set for the next phase. In this new data set, the output values of each model in the first stage constitute the new feature items. In the second stage, the new data set is used as the feature input to the secondary model for training, while the prediction results of the test set in the first stage are used as the test set for prediction, and the final learning results are output. In fact, stack ensemble can group the heterogeneity of multiple basic models into one and combine the prediction results of basic models to reduce the generalization error. Therefore, the combination of basic learning models can be effectively deployed to reduce the bias and improve the prediction accuracy.

In the binary logistic regression, the probability of the user being identified as a patient is taken as the prediction index by the five models combined in the first layer, and the final prediction result is determined by the model.

## IV. DISCUSSION

### A. Model Results and Comparison

The proposed approach builds classification models at three different levels: (1) a model with only text data, (2) a model that includes posting behavior information, and (3) a model that includes text characteristics and posting behavior data. The classification prediction problem includes a variety of performance metrics, and different classification tasks are suitable for choosing different indicators to measure. I chose

accuracy, f1-measure, precision and recall as the metrics for performance evaluation. It can be observed in Table 1 that the performance of the single and ensemble methods on only text characteristics is mostly better than the models with only posting behavior characteristics; that is, the accuracy of model (1) is generally better than that of model (2), except for the naive bayes algorithm. At the same time, it can be noted that model (3) to model (1) has a better predictive effect, and the table also reflects that language characteristic information has a more significant contribution in predicting depressed users than the posting behavior models. Several key conclusions can be drawn from the analysis. First, the fusion feature selection and stacking ensemble model I designed further improves the prediction of every single algorithm.

Finally, to further validate the validity of the proposed models, I tested the results of the proposed methods compared with other ensemble experimental methods, including random forest (RF), gradient boosting (GB), bagging (BG) and adaboost (AB), which performed ill in previous depression recognition and text analysis studies (Zhang et al., 2018; Cacheda et al., 2019; Budhi et al., 2018). The default parameters are chosen for each method, and the results are checked using a 10-fold cross-validation method. Figure 4 shows the accuracy values of random forest, gradient boosting, bagging, adaboost and the stacking ensemble model I constructed before and after applying mixed feature selection. The default parameters are chosen for each method,

and the results are checked using a 10-fold cross-validation method. Figure 5 shows the accuracy values of random forest,

gradient boosting, bagging, adaboost and the stacking ensemble model which was constructed before and after applying mixed feature selection. The results show that the stacking method gives better results than other ensemble models. Bagging has an accuracy of 0.9020 as a classifier in identifying depression, while our model achieves the best performance (0.9027). At the same time, I find that the feature fusion method of text design is applied to various integration models and the classification effect is improved to a certain extent. So, this paper fused feature selection and integration method has a certain competitiveness in identifying depression users and can improve the performance of classification model.

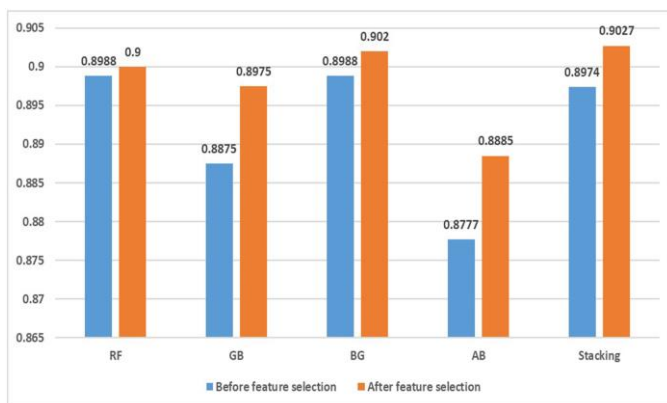


Figure 5

In some cases, the stack ensemble algorithm can complement the advantages of multiple classifiers; that is, the disadvantages of a single classifier on some features will be compensated by other classifiers. Second, both the text characteristics and the posting behavior characteristics have a positive impact on the performance of the system. However, language characteristic information has a stronger predictive ability than posting behavior information, so text characteristics of depression may reflect more information than other features. Experimental results show that compared with other machine learning algorithms, the proposed hybrid method, which integrates feature selection and ensemble, has a higher accuracy of 90.27% in identifying online patients.

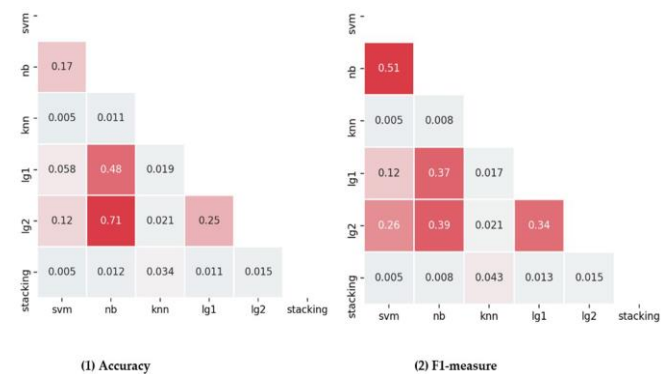


Figure 4

Group	Feature	Description	References
Textual features	The part of speech	Percentage of 20 parts of speech in all posts by users	Islam et al., 2018
	Emotional words	The number of 7 types of emotional words classified according to the Chinese sentiment dictionary	Leis et al., 2019
	Personal pronoun	The frequency of singular/plural first-person pronouns, and the frequency of other pronouns	Vedula and Parthasarathy, 2017
	The specific words	Mainly include the number of negative words and interrogative pronouns	Leis et al., 2019
Posting behavior	Polarity	The emotional orientation of the post, 0 means negative, 1 means neutral, 2 means positive	Sadeque et al., 2018
	Posting habits	Proportion of original posts, posts with pictures and display positioning	Chen et al., 2020
	Posting time	The frequency of users post over a week and over a 24-h period	De Choudhury et al., 2013b

Table 1

## B. LIMITATIONS

This paper draws some valuable conclusions, but there are still some limitations. First, due to the younger age of depressive users, the majority of users participating in the microblog depression community are adolescents, and the sample studied may not represent the general population. At the same time, the users of the control group may also suffer from other mental diseases, so there are some limitations in the promotion of the current findings. Second, this study can provide a preliminary judgment to some users who have difficulty in determining whether they are diseased or not, and it is difficult to avoid mis-identification cases. It is difficult to fully understand other symptoms and reality of patients through social media.

## C. FUTURE WORK

In future studies, I hope to further improve recognition accuracy by combining social media data with clinical data. Finally, due to the limitation of the sample data of depressed users, I could not conduct large-scale deep learning algorithm research. Therefore, in subsequent studies, researchers can enrich the specialization and richness of the data set. In terms of data, future research objectives may provide targeted research for different user types, such as gender, age and family composition. In addition to user differences, different types of depression may also have differences or common factors, such as seasonal emotional disorders, bipolar emotional disorders and postpartum depression. At the method level, future research can integrate more unique learning methods to explore whether different variables or parameters can produce higher predictive performance than previous advanced ensemble methods.

## CONCLUSION

In this paper we have exhibited the capability of using social media as a tool for measuring and detecting major depression among its users. contributing toward depression. The diagnosis of depression is quite complicated and time-consuming, while the detection of psychological diseases through the network community has provided researchers with new ideas. Currently, various single classifiers have been deployed to identify depressed patients in online social media, but most results were unstable or unsatisfactory. This paper proposes a novel online social media depression user detection framework based on feature selection and ensemble learning. Second, in order to better identify depressive patients in online social media an ensemble model was built. The results

conclude that the detection framework that combines feature selection and stacking ensemble method is more suitable for the identification of depressed users and has a strong competitive advantage. It is necessary to identify this issue as soon as possible so that preventive measures can be taken. More Research and advanced

## REFERENCES

- [1] [A. N Hasan, B. Twala, and T. Marwala, Moving Towards Accurate Monitoring and Prediction of Gold Mine Underground Dam Levels, IEEE IJCNN (WCCI) proceedings, Beijing, China, 2014.
- [2] A.K.Jose, N.Bhatia, and S.Krishna, Twitter Sentiment Analysis, National Institute of Technology, Calicut, 2010.
- [3] T.Mitchell, H.McGraw, "Machine Learning, Second Edition, Chapter One", January 2010.
- [4] C.D.Manning, P.Raghavan, H.Schutze, Introduction to Information Retrieval, Cambridge UP, 2008
- [5] 1. Scott J. Social network analysis. Thousand Oaks: Sage; 2017.
- [6] 2. Serrat O. Social network analysis. In: Knowledge solutions. Singapore: Springer; 2017. p. 39–43.
- [7] 3. Mikal J, Hurst S, Conway M. Investigating patient attitudes towards the use of social media data to augment depression diagnosis and treatment: a qualitative study. In: Proceedings of the fourth workshop on computational linguistics and clinical psychology—from linguistic signal to clinical reality. 2017.
- [8] 4. Conway M, O'Connor D. Social media, big data, and mental health: current advances and ethical implications. *Curr Opin Psychol*. 2016;9:77–82.
- [9] 5. Ofek N, et al. Sentiment analysis in transcribed utterances. In: Pacific- Asia conference on knowledge discovery and data mining. 2015. Cham: Springer.
- [10] 6. Yang Y, et al. User interest and social influence based emotion prediction for individuals. In: Proceedings of the 21st ACM international conference on Multimedia. 2013. New York: ACM.
- [11] 7. Tausczik YR, Pennebaker JW. The psychological meaning of words: LIWC and computerized text analysis methods. *J Lang Soc Psychol*. 2010;29(1):24–54.
- [12] 8. Pennebaker JW, Francis ME, Booth RJ. Linguistic inquiry and word count: LIWC 2001, vol. 71. Mahway: Lawrence Erlbaum Associates; 2001. p. 2001.
- [13] 9. Holleran SE. The early detection of depression from social networking sites. Tucson: The University of Arizona; 2010.
- [14] 10. Greenberg LS. Emotion-focused therapy of depression. *Per Centered Exp Psychother*. 2017;16(1):106–17.
- [15] 11. Haberler G. Prosperity and depression: a theoretical analysis of cyclical movements. London: Routledge; 2017.
- [16] 12. Guntuku SC, et al. Detecting depression and mental illness on social media: an integrative review. *Curr Opin Behav Sci*. 2017;18:43–9.
- [17] 13. De Choudhury M, et al. Predicting depression via social Media. In: ICWSM, vol. 13. 2013. p. 1–10.
- [18] 14. De Choudhury M, Counts S, Horvitz E. Predicting postpartum changes in emotion and behavior via social media. In: Proceedings of the SIGCHI conference on human factors in computing systems. New York: ACM;
- [19] 15. O'Dea B, et al. Detecting suicidality on Twitter. *Internet Interv*. P.Taylor, Text-to-Speech Synthesis, Cambridge, U.K.:Cambridge University Press, M. Young, The Technical Writer's Handbook. Mill Valley, CA: University Science, 1989.