REVIEW ARTICLE

Machine Learning Algorithms for Analysis and Prediction of Depression

Mohini Kilaskar[1] · Neha Saindane[1] · Nabeel Ansari[1] · Dhaval Doshi[1] · Mayuri Kulkarni[1]

Abstract

Today, depression is one of the critical mental health problems faced by humans of all ages and gender. In this era of increasing technology, it causes a life of less physical work, continuous pressure on one's life, which creates a risk of intellectual disturbance. The work culture, peer pressure, stressful life, emotional imbalance, family disturbances, and social life are resulting in depression. Depression may also sometimes lead to a heart attack. Depression causes adverse effects and becomes a serious medical problem in how individuals feel and act in everyday life. This psychological state causes feelings of sadness, anxiety, loss of interest in things and jobs, and could barely result in suicide. In this paper, the analysis of different Machine Learning Algorithms has been done and compared them by selecting various parameters and then showing which algorithm is more accurate for predicting depression.

Keywords Machine learning · Depression · XGBoost · SVM · Logistic regression · Random forest

Introduction

Everybody in the world is facing extensive pressure due to work, family, personal life, etc. It leads the stress in life. Due to this, it affects the individual's medical, personal, professional experience. Affecting medical life leads to emotional imbalance. Some persons easily cope with such inequality. But some of them do not easily cope up with the situation.

✉ Nabeel Ansari
    nabeelansari2250@gmail.com

    Mohini Kilaskar
    mohini2001kilaskar1943@gmail.com

    Neha Saindane
    nehasaindane.16@gmail.com

    Dhaval Doshi
    dhavaldoshi100@gmail.com

    Mayuri Kulkarni
    mayuridkulkarni@gmail.com

1   Department of Computer Engineering, Shri Vile Parle
    Kelavani Mandal's Institute of Technology, Affiliated
    to DBATU University, Dhule, India

Because of this, they move toward mental health disease called depression.

This reason makes depression very common around the globe. Worldwide, more than 264 million people of each age have depression [1]. Depression is one of the major contributors to disability to the global burden of disease. The primary and extensive reasons for depression in men and women are as follows:

1. Women are more sensitive in interpersonal relationships, whereas men are more sensitive to their careers and goals.
2. Women tend to contemplate, which may lead to developmental illness. In contrast, men react to hard times by supporting fishy techniques like anger, alcohol, or drug misuse.
3. Women have more emotional quotient in comparison with men, which results in relationship problems.
4. Due to work pressure and household chores, women are less aware of their physical and psychological constraints, due to that she feels more burdened.

From the analysis, it came to know that women get more affected by depression than men, due to this depression the (possible actions) reason why the person thinks of suicide. Depression can result in suicide. Depression is different

from regular mood shifts, and emotions that are for a short time will lead to challenges of the standard of living. Specifically, when it lasts for a long time and with medium or high intensity, depression will become a severe psychological health problem. Nearly 800,000 people commit suicide each year. Suicide is the second-leading cause of death in 15–29 years old [3].

The problem of depression and other psychological state conditions is increasing worldwide. During depression person feels a loss of interest and pleasure, and less energy, resulting in a decreased workflow for weeks. Many people also suffer from anxiety symptoms, less and disturbed sleep and appetite, and lots of have feelings of low self-worth, guilt, less concentration [1]. Nearly two out of three people suffering from depression do not participate or receive proper treatment [3].

Prior detection of psychological disorder helps to start treatment at an early stage and improves the person's life as well as his/her family. The psychological state handles strain in an individual's life and is very important for making choices in each life stage. The psychological condition is critical in every step of life, whether it's childhood or adultism. An excellent psychological state will support a person to give its full potential in life. It also helps the person to cope with stress in life. It shall increase the productivity of the people [8].

As per the prediction of the World Health Organization (WHO), by 2030, depression will be one of the second-leading disabilities [4, 5]. Although very much prepared specialists, clinical and mental medicines are accessible for depression treatment, people or families are hesitant to stand up/arrive at specialists about this problem for different social reasons. Diagnosis of depression disorder includes several visits with patient and patient's family, clinical analysis, questionnaires, which are time-consuming and need to have well-trained specialists. In this present generation of Machine learning, automated depression detection is not complicated and deployed quickly. The automation should use less resources, provide accurate results with more reachability [4].

Literature Review

Different ways to analyze depression are as follows:

Anees Ul Hassan, Jamil Hussain, Musarrat Hussain, Muhammad Sadiq, and Sungyoung Lee have proposed that people use social networking sites to express their thoughts, feelings, and many more so. Here Twitter is a social site that allows users to tweet statements up to 140 characters. The sentimental analysis helps determine the emotion behind a series of words and teaches Machine Learning techniques to analyze those meanings. Four main components are

pre-processing, feature extraction, meta-learning, and training data. In pre-processing, a paragraph gets converted into sentences, and sentences get converted into words, and then from those words, stop words get removed. Stop words are useless words and do not bring enough meaning, for example. And, of, is, from, in, on, over, under, etc. are stop words. The feature detection from text is related to a name or a sequence of words that can be a feature or not [9].

Some of the feature extraction methods are N-grams, Parts of Speech (POS) Tagging, Negation, Sentiment Analyzer. In Meta-Learning, Voting is the easiest way to combine the predictions from multiple machine learning algorithms. Here three classifiers, Support Vector Machine, Naive Bayes, and Maximum Entropy classifiers, are used as learners. Training data includes training datasets, i.e., negative, positive, feature extractor, and word features. There are two datasets included, i.e., the Twitter dataset and 20 news-groups. Figure 1 shows the Machine Learning Algorithms with their accuracy [9].

Anna Maridaki, Anastasia Pampouchidou, Kostas Marias and, Manolis Tsiknakis developed a framework based on the dataset provided by AVEC'14 for depression analysis. The proposed work presents two different motion representations: (1) Gabor Motion History Image (GMHI), and (2) Motion History Image (MHI). Various combinations of low-level features get extracted from both representations. These features get further used for training and testing with different machine learning techniques. The proposed approach got an F1 score of 81.93%, both for MHI and GMHI, with Support Vector Machine. The achieved performance is like state-of-the-art approaches, while manages to outperform several others. Except for accomplishing a competitive performance, the proposed work explores various combinations of the investigated motion representations and classifiers [2].

PCA method decreases the feature dimensionality of the data by casting the data into a lower-dimensional space. The smaller feature set gets used as the extracted features for the classification step. The feature extraction step's feature vector is operated by a selected classifier to evaluate the system in the classification step. The machine learning algorithms Naïve Bayes, k-Nearest Neighbors, Random Forest, and Support Vector Machine get tested. The below graph shows the F1 score results implemented in classifiers for MHI [2] (Fig. 2).

Vanishri Arun, Prajwal V., Murali Krishna, Arun Kumar B. V., Padma S. K, and Shyam V. Proposed a system in which they have validated an utterly unique technique for prediction of depression. They collected clinical information from the on-going MYsore studies of Natal effects on Aging and Health (MYNAH). The participants completed a complete evaluation for psychological feature function, psychological state, and cardiometabolic disorders in South India. This model developed using the XGBoost Machine

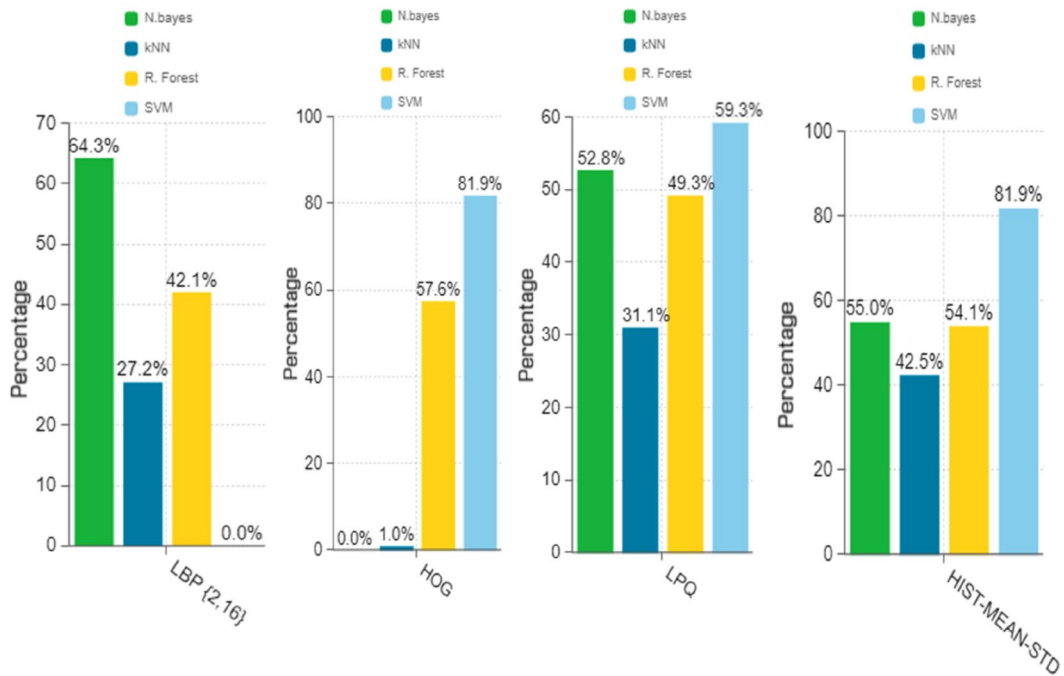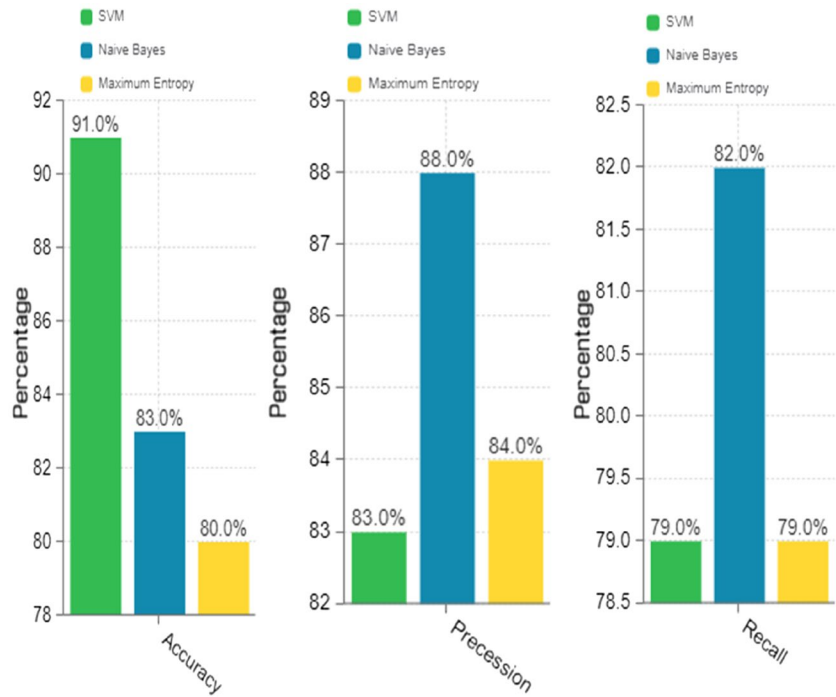**Fig. 1** Performance analysis graph

Fig. 2 Graph for F1 score results implemented in classifiers for MHI

learning algorithm. The dataset's training and learning are quicker in the XGBoost Algorithm than other prediction algorithms [7] (Fig. 3).

Swati Jain, Suraj Prakash Narayan, Rupesh Kumar Dewang, Utkarsh Bhartiya, Nalini Meena, and Varun Kumar Proposed the suicidal acts system based on the depression level. In this system, they have. By filling of questionnaires, they collect real-time data. Their system uses a questionnaire similar to PHQ-9, which is an advanced version of it. It covers all characteristics or components and symptoms, leading to depression. Categorizing machine algorithms is used to train and classify it in five stages depending on
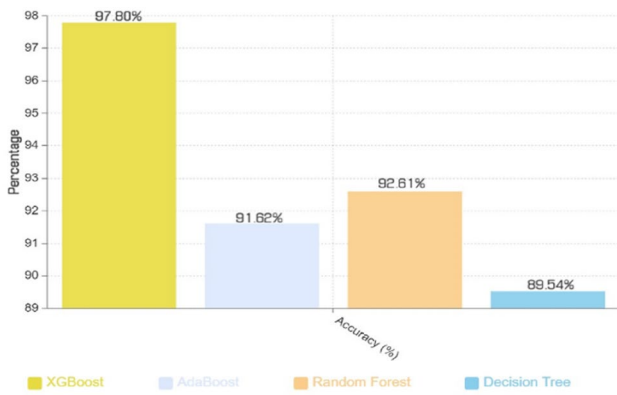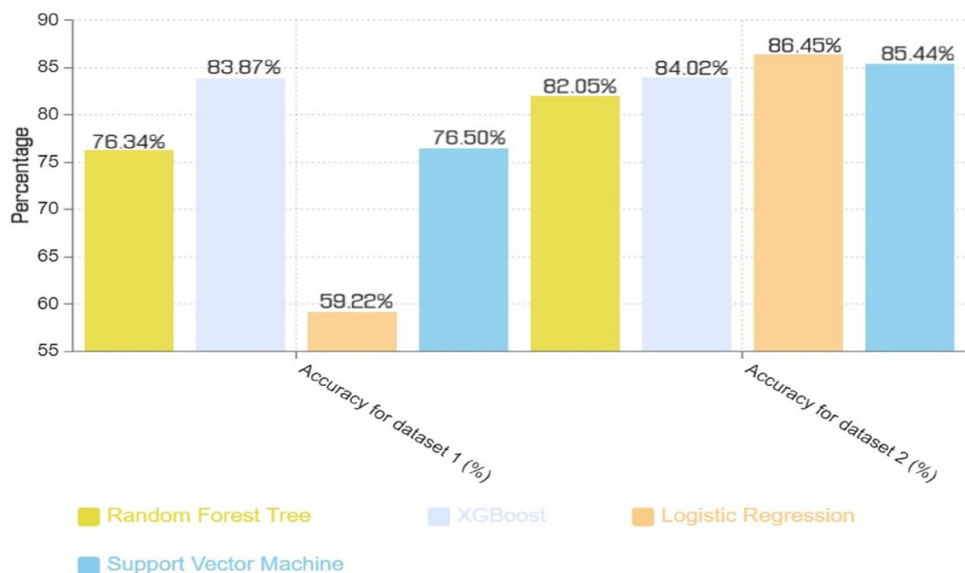
Fig. 3 Graph of results using MYNAH cohort in various machine learning algorithms

classifiers. Voice feature gets extracted using the COVAREP tool. SMOTE analysis gets used to overcoming the class imbalance, and 93% accuracy obtained with the support vector machine algorithm, which results in the Depression Classification Model (DCM). An android application cured Deployed on the WAMP server on Amazon Web Service (AWS) cloud is developed to self-assess depression using DCM and PHQ-8 questionnaire. The application was tested on real-time data of 50 persons in the supervision of a quali-fied psychiatrist, and 90% accuracy got obtained. Three classification techniques, namely Logistic Regression, Random Forest, and SVM, are implemented in the scikit-LEARN toolbox. The evaluation metrics are calculated individually for each classifier— over the DIACWOZ data corpus speech and shown in Fig. 5. Finally, the "Depression Classification
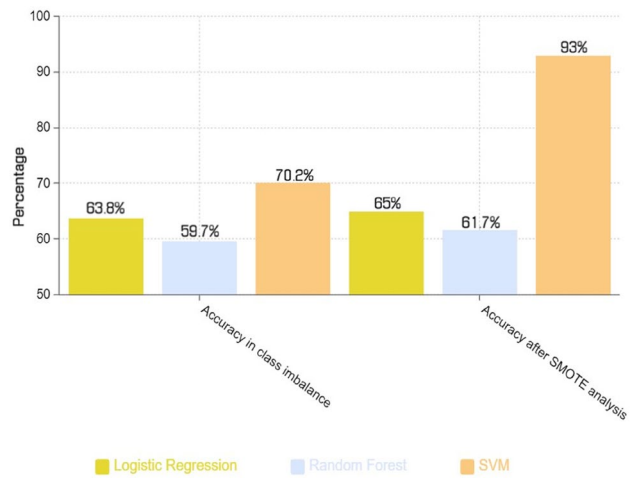
severity—Minimal or none, mild, moderate, moderately, and severe. Two datasets get chosen, dataset: (1) questionnaires, (2) Twitter tweets. For dataset 1, XGBoost classifier gave the highest accuracy, i.e., 83.87%, and Logistic Regression gave the lowest, unlike dataset 2 where Logistic regression showed the highest accuracy of 86.45% because dataset 1 consists of 18 features, which were more than the words on which tweet gets extracted, and model gets trained. Logistic regression starts to falter when there are many features and a good number of lost data. Also, so many categorical vari-ables are a problem for it [6] (Fig. 4).

Bhanusree Yalamanchili, Nikhil Sai Kota, Maruthi Saketh Abbaraju, Venkata Sai Sathwik Nadella and, Sandeep Varma Alluri have proposed a system in which acoustic feature gets used to training a classification model to categorize a human as Depressed or not-Depressed. DIAC-WOZ database avail-able with the AVEC2016 challenge chosen for training the

Fig. 5 Graph for accuracy of classifiers on class imbalance and SMOTE analysis dataset

Fig. 4 Graph for accuracy of algorithms applied to the dataset

Model (DCM)" is generated with 93% accuracy on the validation data set using acoustic features [4].

Proposed System

As you have seen, many (or every) authors described depression and provided the solution with their end. We have also researched and understood the points written by them. Here we are going to provide our proposed solution to this.

First, we will discuss "XGBoost". XGBoost is an open-source library that provides a gradient boosting framework for Java, Python, R, Perl, and Scala. It works on Linux, Windows, and macOS. The project description aims to provide a "Scalable, Portable, and Distributed Gradient Boosting Library" [10]. It took Input in Tabular format and explained in detail, but no such application of XGBoost was mentioned in the papers to predict depression.

This paper shows the smart way to predict depression using Social Media Stories, tweets, voice, searches on the browser, etc. The solution we will describe is it can be an Admin app or Add-on in the browser or an IoT device. The module's primary function is to capture data frequently, so it is necessary to refresh after some consecutive limit of time.

The working of our idea shows in the below Fig. 6.

In the above Fig. 6, the Data sets are Social Media Stories, tweets, voice, searches on the browser, etc. The converter is used to calculate the depressive content and convert it into Tabular format (Input can be Images, Texts, and Voices). The last section that is Calculated Output is the part of the solution in which it processes input from XGBoost and reframes it into human-understandable language (Fig. 7).

A brief description of the converter is as follows (why how).

As the XGBoost is taking input in tabular format so that the dataset's inputs are raw data, we must process it and convert it into a tabular form. Here we need the smart converter. The figure represents the converter's basic structure of converting the raw data into a tabular format. First, the Social media stories that have 24 h are in image format (or a video format, mostly typical depressed person do not put a video on the story), so save it. It is necessary to process and check for depression; now, we use image processing algorithms to identify the facial expressions and check for depression of the person and generate output in tabular format to XGBoost Algorithm. Now the story of social media could also be in text format, so the text would be identified by the text processing algorithm, which checks the depressive words and generates output in tabular form for XGBoost Algorithm as it takes input only in tabular format.

The next one is depressive tweets. Tweets are always in the text format so that input is already in the text format, so check the depressive words in the tweet make it in tabular form so the XGBoost could recognize the data.

Next is Depressed Voice Recognition, Here the inputs are of voice captured by phone microphone is any accessible
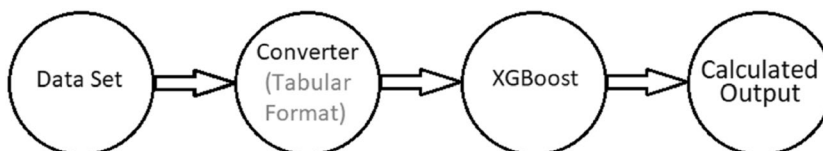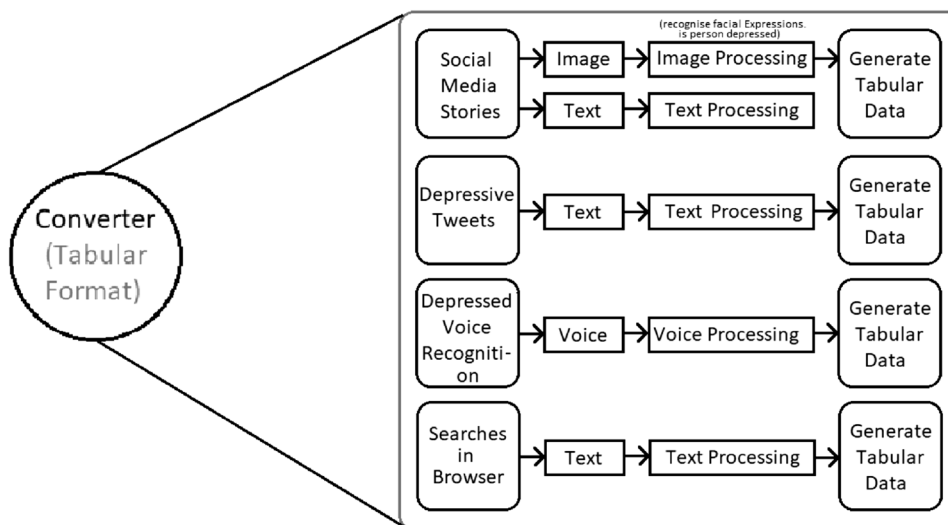

Fig. 6 Basic structure of solution

Fig. 7 Converter section functions (components of converter)

microphone if possible, call recordings, so the algorithm for detecting depressive voice gets used. It will check the depressive voices in the recorded conversation and convert it to tabular format so that XGBoost Algorithm can recognize it.

The last one is the "Searches in Browser". It mostly happens that the depressive person searches the suicidal contents on the Internet so that the system will capture and process it in an algorithm, check depressive words, and generate output, which will be understandable to the XGBoost (tabular).

Future Scope

- As the system will be an Admin Application (so it needs to monitor the camera, stories, voice, etc., of the user), it is necessary to gain user trust. Hence tying-up with the trusted IT companies (i.e., Google, Amazon, etc.) is the best solution.
- Cloud-based data are memorizing for faster access from anywhere.
- Cloud-based parental control will also be the best part for parents to monitor their children smartly.
- Video calling depression detection, detection of suicidal activities.
- Monitoring financial loss in business, less account balance of well-earning employees, fewer marks obtained by the genius student, etc.

Conclusion

Paper reports machine learning algorithms for analysis and prediction of depression. The algorithm is implemented based on the user's social behavior and data. Future research will get conducted toward upgrading the performance of the system. This paper also discussed the methodology of machine learning. This paper studied various machine learning algorithms implemented in past years in the depression prediction combined with the machine learning algorithms and learned how they outshined.

Declarations

Conflict of Interest The authors declare that they have no conflict of interest.

References

1. WHO depression page, https://www.who.int/news-room/fact-sheets/detail/depression, Accessed 10 Oct 2020.
2. Anna Maridaki, Anastasia Pampouchidou, Kostas Marias and, Manolis Tsiknakis "Machine Learning Techniques for Automatic Depression Assessment", 41st IEEE International Conference on Telecommunications and Signal Processing (TSP), August 2018.
3. Nafiz Al Asad, Md. Appel Mahmud Pranto, Sadia Afreen, and Md. Maynul Islam, "Depression Detection by Analyzing Social Media posts of User", IEEE International Conference on Signal Processing, Information, Communication & Systems (SPIC-SCON), November 2019.
4. Bhanusree Yalamanchili, Nikhil Sai Kota, Maruthi Saketh Abbaraju, Venkata Sai Sathwik Nadella and, Sandeep Varma Alluri, "Real-time Acoustic based Depression Detection using Machine Learning Technique", IEEE International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE), April 2020.
5. Prajwal Kharel, Kalpana Sharma, Sunil Dhimal and, Sital Sharma, "Early Detection of Depression and Treatment Response Prediction using Machine Learning: A Review", IEEE Second International Conference on Advanced Computational and Communication Paradigms (ICACCP), October 2019.
6. Swati Jain, Suraj Prakash Narayan, Rupesh Kumar Dewang, Utkarsh Bhartiya, Nalini Meena, and Varun Kumar, "A Machine Learning-based Depression Analysis and Suicidal Ideation Detection System using Questionnaires and Twitter", IEEE students Conference on Engineering and Systems (SCES), February 2020.
7. Vanishri Arun, Prjwal V, Murali Krishna, Arunkumar B. V., Padma S. K. and, Shyam V., "A Boosted Machine Learning Approach for Detection of Depression", IEEE Symposium Series on Computational Intelligence (SSCI), January 2019.
8. Purude Vaishali Narayanrao and, P. Lalitha Surya Kumari, "Analysis of Machine Learning Algorithms for Predicting Depression", International Conference on Computer Science, Engineering and Applications (ICCSEA), July 2020.
9. Anees Ul Hassan, Jamil Hussain, Musarrat Hussain, Muhammad Sadiq and, Sungyoung Lee, "Sentimental analysis of social networking sites (SNS) data using machine learning approach for the measurement of depression", International Conference on Information and Communication Technology Convergence (ICTC), December 2017.
10. Wikipedia XGBoost page, https://en.wikipedia.org/wiki/XGBoost. Accessed 10 Oct 2020.