Effective Machine Learning Based Techniques for Predicting Depression

Mhd Saeed Sharif
Computer Science and DT, ACE, UEL
University Way, London, UK
S.Sharif@uel.ac.uk

Abin Zorto
Computer Science and DT, ACE, UEL
University Way, London, UK
u2091940@uel.ac.uk

Abisha Thayyath Kareem
Computer Science and DT, ACE, UEL
University Way, London, UK
u2093871@uel.ac.uk

Rasha Hafidh
Computer Science and DT, ACE, UEL
University Way, London, UK
r.hafidh2@uel.ac.uk

***Abstract*—Depression is a global disorder with serious consequences. With more depression-related data and improved machine learning, it may be possible to build intelligent systems that can detect depression early on. This research uses the burns depression checklist as the gold standard for diagnosing depression and the support vector machine, decision tree, and light gradient boosting method as algorithms to create models capable of diagnosing depression on a data-set of 604 surveyed participants. This research demonstrates the efficiency of machine learning algorithms within the field of mental health. This paper serves to increase the body of knowledge by training insufficiently researched algorithms on a commonly used depression detection data-set with the goal of reaching or surpassing the level of performance seen in current research. This experimental research has found the decision tree classifier to be the best approach for predicting depression with an accuracy of 95.66% while that of the support vector machine classifier and the light gradient boosting classifier are 91.48% and 94.58%, respectively. The techniques presented in this paper perform better than those being used in current machine learning research. This research study may support the clinicians in determining what attributes are most crucial in diagnosis of depressed individuals as well as improve the health of the general populace.**

***Index Terms*—Depression, BDC, Machine Learning, SMOTE**

I. Introduction

There are many different types of mental illness, but depression is the most commonly seen. Depression may significantly damage functioning in work, school, and family, and may even lead to self-harm. Only a few prospective research works have examined a broad variety of predictors throughout many domains for new-onset (incidental) depression in adulthood. With the growth of data sets that are relevant to depression, as well as the improvement of machine learning, there is the possibility to construct intelligent systems that can recognise signs of depression in early stages. Negative thinking, decreased concentration, and decreased productivity are all symptoms of depression and an early diagnosis of this mental disorder can improve both the patient and his or her family's standard of living. The major aim of this research is to develop an efficient and scalable model to assess if a person is depressed or not, as well as determine efficient machine learning techniques for identifying depressed people.

According to the World Health Organisation (WHO) [1], depression is present in 3.8% of the total global population, with women more likely than men to suffer from various types of depression. Following the Covid-19 outbreak, depression became a severe public health issue, with 322 million people suffering from depression at any given moment. Depression has now been linked to a range of chronic illnesses, including diabetes, heart disease, and other maladies. It is the second most important risk factor for the development of chronic illnesses [1]. Suicidal tendencies can be triggered by severe depression leading to over half of the 0.8 million suicides occurring throughout the world to be caused by depression, according to the statistics [2].

There is a rising demand for Machine Learning algorithms to infer meaningful patterns from data from a variety of diverse industries. Although Machine Learning algorithms have been widely employed in the care and psychiatric industries [3], their application in the mental field is still relatively minimal. Statistical tests have been utilised in psychological profiles and psychological tests for many years now, and with good reason. Following the CA scandal, Machine Learning has received a great deal of interest in the press for its application in personality tests. Researchers in the fields of personality assessments and mental analysis are increasingly moving to Machine Learning from statistics findings due to the prediction validity concerns connected with probabilistic reasoning [4].

II. Literature Review

The research work done by [5] examined six computational intelligence classifiers that use various socio-demographic as well as psychological data to determine whether a person is depressed or not. The SelectKBest feature selection strategy of the AdaBoost classifier has surpassed all other techniques with an accuracy of 92.56%. Synthetic minority oversampling technique (SMOTE) has also been used to lessen the class imbalance of training data to improve the accuracy of the prediction depression. However the research by Jagtap et al. [6] has presented a method to identify depression employing ensembled learning and Natural Language Processing approaches. As a result of this comparative study, an accuracy of 96.35% is obtained using the highest scoring machine learning technique that was used in the paper.

A study by Choudhury et al. [7] aimed to find the best strategy for predicting depression among participants, three algorithms were compared and Random Forest was determined to be the most accurate algorithm with an f-measure of 75% and 60%, with superior precision and recall as well as fewer false negatives than the runner up in the Support Vector Machine.

Hatton et al. [8] found that the frequency of depression in 284 older individuals was predicted using psychometric and demographic data. The accuracy of the Logistic Regression models for predicting the chance of depression reoccurring were compared with Extreme Gradient Boosting. They concluded that Extreme Gradient Boosting Outperformed Logistic Regression in terms of accuracy.

Postpartum Depression is a common kind of depression experienced by new mothers after childbirth. Natarajan et al. [9] designed a model to help detect Postpartum Depression on a data-set of 173 new moms of various ethnicities. They measured the effectiveness of the Functional Gradient Boosting algorithm to other standard neural network models and discovered it was the most effective in diagnosing Postpartum Depression.

Zarandi et al. [10] evaluated the stage of depression through the use of type-2 fuzzy logic. They applied the Mutual Information Feature Selection approach to increase the original study reliability and evaluate the amount of depression in patients. This proposed method had an accuracy of 84.00% and only used 15 inquiries to determine the extent of depression. Meanwhile [11] aimed to predict depression in a person based on his Facebook postings and comments. They created models using several machine learning classifiers, such as Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Decision trees, and Ensemble classifier, after extracting psycholinguistic data from the person's Facebook posts and comments. In this case, Decision Trees outperformed other classifiers.

Early identification of depression has received much less research. Ophir, Asterhan and Schwarz [12] looked for early warning signs of depression in young Facebook, Inc. users with the hope of using their methodology to develop early diagnosis tools, although they make no such suggestions. In this study, the researchers found that when a person is depressed, they post more often and in more negative ways, engage less socially, and concentrate more on themselves, all of which are correlated with the beginning of depression.

There is a great deal of variance in how depression affects each person. A diagnosis of depression based only on symptoms makes it difficult, if not impossible, to objectively quantify psychological as opposed to physiological occurrences. Those who are mentally ill but are unwilling to seek treatment or are unaware of how to diagnose themselves can benefit from this research's success. Its results can help the government and psychologists become more efficient in identifying and treating mental health issues.

III. Methodology

A. Data Description

An investigation was carried out between April and August of 2020 in order to acquire information about the participants who fall into a variety of categories. A survey was carried out with a questionnaire consisting of 55 questions was devised. The very first 30 questions were constructed to capture complicated psycho-social and socio-demographic data from the subjects, and a modified version of the Burns Depression Checklist (BDC), a widely recognized depression rating scale created by David Burns [13] was used for the remaining 25 question in the survey and was used to determine each participant's true emotional status.

The new edition of BDC has 25 questions separated across four sections. the first ten questions are focusing on the present thoughts and emotions of the participants. the focus of the next seven are on their most recent endeavours and relationships. Within the next five questions, the working features and symptoms of the individual are being discussed. The final section of the questionnaire investigates the respondent's propensity toward suicidal behaviour [14].

In order for the BDC to properly assess an individual's level of depression, participants were required to rate the severity of numerous depressive symptoms that they had experienced over the course of the previous days, including the day before the survey. Within the most recent iteration of BDC, the severity of the symptoms can range anywhere from 0 to 4. It is consistent, but it focuses on specific indicators of depression rather than the more general symptoms of mental illness [13]. The intensity of each symptom that the subject has supplied gets added to determine the person's overall BDC score. Unless a person's overall score gets greater than 10, they are termed depressed by BDC. The database includes the contributions of 604 individuals and table I below [5, Tab. 1] shows the variables for predicting depression.

The dataset consists of 397 individuals classed as depressed and 207 individuals classed as not depressed.

B. Data Preprocessing

Pre-processing strategies primarily focus on transforming raw data into a comprehensible format. What this implies is that the computer can readily interpret, anticipate, and analyse what is in the data using different machine learning methods.

1) Feature Selections: Both the classification algorithm and the feature selection strategy have an impact on the accuracy of a classifier. Inaccurate outcomes might be obtained if the classifier is confused by characteristics that are irrelevant or improper. If you want a classifier to be more efficient and accurate, you need to choose the right features. This is the process of minimizing the number of factors in a forecasting model by selecting the most important ones to include. The computational cost of modeling could be decreased, and the model's performance might be increased if the number of input parameters is decreased. The relevance of feature selection can not be overstated in a variety of contexts [15]. It is

TABLE I
VARIABLES FOR PREDICTING DEPRESSION [5].

	Variable	Possible Values
1	Age range	16–20, 21–25, 26–30, 31–35, 36–40, 41–45, 46-50, 51-55, 56-60, 61+
2	Gender	Male, Female
3	Educational qualifications	SSC, HSC, Graduate, Post Graduate
4	Profession	Student,Businessman, Unemployed, Other
5	Marital status	Unmarried, Married, Divorced
6	Type of residence	Village, Town, City
7	Lives with family or not	With Family, Without Family
8	Satisfied with living environment	Yes, No
9	Satisfied with current position/ academic achievements	Yes, No
10	Financial stress	Yes, No
11	Debt	Yes, No
12	Frequency of physical exercises	Never, Sometimes, Regularly
13	Smokes	Yes, No
14	Drinks alcohol	Yes, No
15	Serious illness	Yes, No
16	On prescribed medication	Yes, No
17	Eating disorders	Yes, No
18	Average hours of sleep	$<$5 h, 5 h, 6 h, 7 h, 8 h, $>$8 h
19	Suffers from insomnia	Yes, No
20	Average hours on social network (in a day)	$<$2 h, 2–4 h, 5–7 h, 8-10 h, $>$10 h
21	Current work or study pressure of the participant	Severe, Moderate, Mild, No Pressure
22	Feels anxiety for something or not	Yes, No
23	Recently felt that he/she has been depression	Yes, No
24	Felt abused (physically, sexually, emotionally) or not	Yes, No
25	Felt cheated by someone recently	Yes, No
26	Faced any life-threatening event recently	Yes, No
27	Suicidal thoughts	Yes, No
28	Suffers from inferiority complex	Yes, No
29	Recently engaged in any kind of conflicts with friends or family	Yes, No
30	Recently lost someone close to him	Yes, No
31	Target variable	0 (Not Depressed). 1 (Depressed)

usual practice to pick features sequentially. Greedy search techniques are used to reduce a large feature space into a smaller feature subspace where k is less than or equal to the original d-dimensional feature space [16]: The approach minimises a criterion over all viable feature subsets. Mean squared error and misclassification rate are common criterion (for classification models). It also comprises of a sequential search method that adds or eliminates features while evaluating a criterion. Since comparing the criteria value at all 2n subsets of an n-feature data set is usually impossible based on the scale of n and the cost of objective calls, sequential searches always enlarge or shrink the candidate set. The method has two variants: Sequential forward selection (SFS), where more characteristics are added successively to an empty candidate set until the criteria do not change as a result of future additions.

Sequential backward selection (SBS), in which characteristics are systematically deleted from a whole candidate set until the removal of subsequent characteristics increases the criteria.

2) Data-set Splitting: This study utilised 80% of the data set for training. The remaining 20% of the data-set has been utilised for testing.

3) Applying SMOTE: An uneven data-set might lead to erroneous predictions when training a classifier. As a result of the unbalanced nature of the training data, SMOTE is used on this portion of the data-set. This is done to reduce bias in the prediction performance of the created models. Predictive accuracy for the minority class can only be improved by creating a more evenly distributed data-set. By using feature space, SMOTE creates synthetic minority group data. The manufactured samples are put all along a line connecting each minority class sampling to its own K-next minority class example neighbors. Well before synthetic instance would ever be formed, a random variable between 0 and 1 is multiplied with the difference between a minority class instance's feature space as well as its nearest neighbor. To create a synthetic instance of the minority class, the result of multiplying the feature vector is added to it [17]. Consider the case where f_i is the sample's feature vector, and f_{near} is one of f_i's K-nearest neighbors. Eq. (1) may be used to describe the new synthetic sample, where R is a chance number ranging from 0 to 1.

$$f_{\text{new}} = f_i + (f_i - f_{\text{near}}) \times R \qquad (1)$$

C. Algorithm selection

In this research, Support Vector Machine, Decision tree, and LGBM classifier are used. The specifics of these classifiers are outlined below.

1) SVM Classifier: The SVM method works by locating the hyperplane with the shortest minimum distance to the training instances. Within SVM theory, this gap is referred to as "margin." The optimal hyperplane is the one with the greatest margin of separation between two classes. Hyperplanes are chosen based on how far from each data point on each side they are to each other. If one exists ”maximum margin hyperplane” was what it’s called, as well as the linear classifier it generates is called the ”maximum margin classifier.” Another of the main goals of the SVM approach would be to identify the best boundary or lines for splitting n-dimensional spaces among groups so that we can quickly categorise data points. A hyperplane seems to be a plane which represents the best option [18]. To make the hyperplane, the SVM looks for the most extreme points and vectors in the data-set. A Support Vector Machine (SVM) is an algorithm that looks for patterns even in the most extreme cases. Using a decision boundary or hyperplane, the following graphic, figure 1, classifies two separate categories:

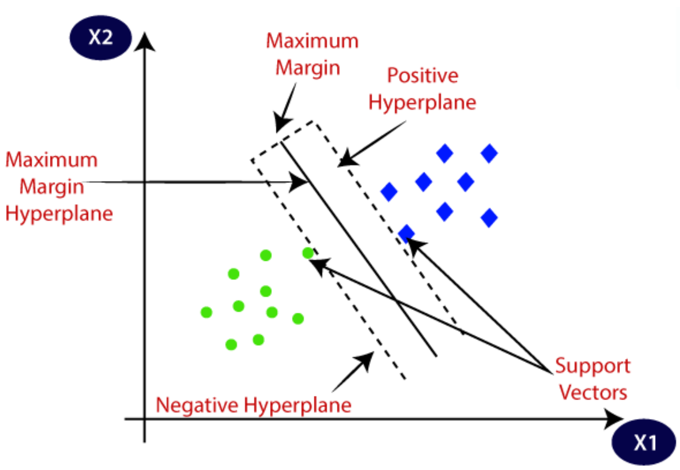

Fig. 1. Structure of Hyperplane [18].

2) Decision Trees: Decision trees learn basic decision rules derived from input features and forecast future value of the target variables inside a non-parametric supervised learning algorithm. To put it another way, decision tree categorisation is the process of identifying decision trees in training records that have been marked with labels. Tree-like structures with internal nodes denoting tests on documents and branches indicating the results of those tests are known as decision trees. Each leaf node is labeled with a class name. It’s a top-down approach that builds a decision tree classifier recursively.

The process begins at the root node of the tree when using a decision tree to forecast the data-set’s class. When the root attribute is compared to the record (actual data-set) attribute, it goes to the next node and follows the branch accordingly, as illustrated in figure 2 [18].

When going onto the next node, the technique compares the parameter of next node to a number of other sub-nodes. It proceeds until it reaches the leaf node of the tree when it comes to a halt.

3) LGBM Classifier: The Gradient Boosting Method, also known as Light GBM or Light Gradient Boosting Method, is a tree-based approach. Its algorithm is based on trees, which develop vertically rather than horizontally, it is referred to as a ”light” classifier because it can process quicker than another classifier. When compared to other techniques, the Light gradient boosting approach is very quick and efficient when dealing with huge data sets [19].

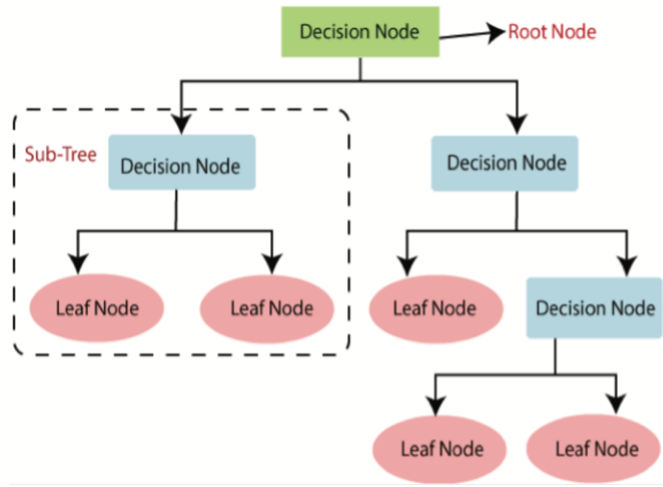

Fig. 2. How does the Decision Tree algorithm Work? [18]

IV. Results

After applying these machine learning classifiers, we have compared them using different performance metrics: accuracy, precision, F1-score, sensitivity, and specificity. Table II shows the experimental findings of the proposed classifiers for depression detection.

TABLE II
Results of Proposed Classifier.

Model/Matrix	Accuracy	Precision	F1-Score	Sensitivity	Specifity
SVC Classifier	91.48	82.21	81.80	95.74	97.10
Decision Tree Classifier	95.66	96.11	94.20	91.94	94.67
LGBM Classifier	94.58	74.10	73.32	82.97	89.47

A. Result Comparison

When assessing the overall research quality, it’s indeed critical to compare the results or accomplishments of such an investigation to other previous studies. The majority of these studies are designed to predict depression among people of a particular age category, employment, or medical problem. The most essential socio-demographic and psychosocial elements that contribute to depression have been identified by a few of them. From the above results, we can see that the decision tree classifier provides good accuracy for depression prediction. This section shows the comparison of machine learning-based models by Zulfiker et al. [4] using a SelectKBest Feature Selection method as recommended by the author, to the proposed models within this paper. These results were chosen to highlight the importance of the algorithms and preprocessing techniques used on the same data-set.

Table III shows the comparison results of base and proposed models for depression prediction.

TABLE III
COMPARISON RESULTS FOR DEPRESSION PREDICTION.

	Base Models (SelectKBest) [5]			Proposed Models (Sequential)		
Metrics / Models	Bagging	Gradient Boosting	Adaboost	SVC	DT	LGBM
Accuracy	90.91	91.74	92.56	91.48	95.66	94.58
Precision	94.37	94.44	95.77	82.21	96.11	74.10
F1-Score	92.41	93.15	93.79	81.80	94.20	73.32
Sensitivity	90.54	91.89	91.89	95.74	91.94	82.97
Specifity	91.49	91.49	93.62	97.10	94.67	89.47

From this comparison, we can say that although both papers and techniques are comparable, the decision tree classifier is the most accurate model for predicting depression with an accuracy of 95.66%. The results also show that while some of the models in the original paper excel in F1-score, the decision tree model performs better than all the models. Accuracy in tandem with F1-score will be considered the optimal metrics for judging performance. Figure 3 illustrate a comparison between bagging, gradient boosting, adaboost, SVC, DT and LGBM.

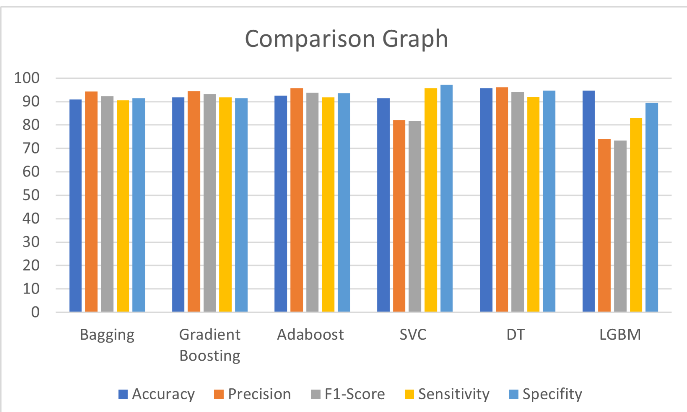

Fig. 3. Performance Comparison Graph of Base and Proposed Model

V. CONCLUSION AND FUTURE WORK

This research study developed an accurate model aiming to support clinicians in determining what attributes are most crucial in diagnosis of depressed individuals at an early stage. The performed experiment showed that the decision tree approach is the most accurate model for predicting depression, as compared to the different algorithms with a 95.66% accuracy, and the other classifiers such as the SVC classifier and LGBM classifier lag behind with 91.48% and 94.58% accuracy respectively. The accuracy obtained in this paper by the decision tree model, is higher than the one achieved in other related work conducted on the same data-set. Which would lead us to conclude the optimal approach for depression detection would be the decision tree model as it possesses a high accuracy of 95.66%. This research investigation used BDC as the gold standard for diagnosing depression without any bio-markers. Hence the future work of this research will be focusing on creating and acquiring a comprehensive data-set which includes clinical bio-markers monitored by expert clinicians. Such data-set will make a perfect input for the machine learning based model to achieve more accurate results within the context of depression treatment. Moreover, it is recommended that an expanded version of the emotional feature data-set be employed. The emotional process component will provide a bigger number of traits, which should be considered for inclusion in our future work in this field. Lastly more depression analysis from a wider range of social media domains is required for improved accuracy and sensitivity, which will be the topic of future research publication on the subject. The significance of such results is that they will open the path for future study into mental health and machine learning as well as show the importance to investigate possible causative elements from social media. If the findings of this research were utilised to investigate further links, it may lead to lowering anxiety and depression rates as well as increase in suicide prevention rate.

REFERENCES

[1] "Depression," World Health Organization, Sep. 13, 2021. https://www.who.int/news-room/fact-sheets/detail/depression (accessed Aug. 15, 2022).

[2] A. Thapar, S. Collishaw, D. S. Pine, and A. K. Thapar, "Depression in adolescence," Lancet, vol. 379, no. 9820, pp. 1056–1067, Mar. 2012.

[3] A. Priya, S. Garg, and N. P. Tigga, "Predicting Anxiety, Depression and Stress in Modern Life using Machine Learning Algorithms," Procedia Computer Science, vol. 167. pp. 1258–1267, 2020. doi: 10.1016/j.procs.2020.03.442.

[4] G. Orrù, M. Monaro, C. Conversano, A. Gemignani, and G. Sartori, "Machine Learning in Psychometrics and Psychological Research," Front. Psychol., vol. 10, p. 2970, 2019.

[5] M. S. Zulfiker, N. Kabir, A. A. Biswas, T. Nazneen, and M. S. Uddin, "An in-depth analysis of machine learning approaches to predict depression," Current Research in Behavioral Sciences, vol. 2, p. 100044, Nov. 2021.

[6] N. Jagtap, H. Shukla, V. Shinde, S. Desai, and V. Kulkarni, "Use of Ensemble Machine Learning to Detect Depression in Social Media Posts," in 2021 Second International Conference on Electronics and Sustainable Communication Systems (ICESC), Aug. 2021, pp. 1396–1400.

[7] A. A. Choudhury, M. R. H. Khan, N. Z. Nahim, S. R. Tulon, S. Islam, and A. Chakrabarty, "Predicting Depression in Bangladeshi Undergraduates using Machine Learning," in 2019 IEEE Region 10 Symposium (TENSYMP), Jun. 2019, pp. 789–794.

[8] C. M. Hatton, L. W. Paton, D. McMillan, J. Cussens, S. Gilbody, and P. A. Tiffin, "Predicting persistent depressive symptoms in older adults: A machine learning approach to personalised mental healthcare," J. Affect. Disord., vol. 246, pp. 857–860, Mar. 2019.

[9] S. Natarajan, A. Prabhakar, N. Ramanan, A. Bagilone, K. Siek, and K. Connelly, "Boosting for Postpartum Depression Prediction," in 2017 IEEE/ACM International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE), Jul. 2017, pp. 232–240.

[10] M. H. Fazel Zarandi, S. Soltanzadeh, A. Mohammadi, and O. Castillo, "Designing a general type-2 fuzzy expert system for diagnosis of depression," Appl. Soft Comput., vol. 80, pp. 329–341, Jul. 2019.

[11] M. R. Islam, M. A. Kabir, A. Ahmed, A. R. M. Kamal, H. Wang, and A. Ulhaq, "Depression detection from social network data using machine learning techniques," Health Inf Sci Syst, vol. 6, no. 1, p. 8, Dec. 2018.

[12] Y. Ophir, C. S. C. Asterhan, and B. B. Schwarz, "Unfolding the notes from the walls: Adolescents' depression manifestations on Facebook," Comput. Human Behav., vol. 72, pp. 96–107, Jul. 2017.

[13] D. Burns, H. Westra, M. Trockel, and A. Fisher, "Motivation and Changes in Depression," Cognitive Therapy and Research, vol. 37, no. 2. pp. 368–379, 2013. doi: 10.1007/s10608-012-9458-3.

[14] S. A. Jabbar and H. I. Zaza, "Post-traumatic Stress and Depression (PSTD) and general anxiety among Iraqi refugee children: a case study from Jordan," Early Child Dev. Care, vol. 189, no. 7, pp. 1114–1134, Jun. 2019.
[15] J. Liu and M. Shi, "A Hybrid Feature Selection and Ensemble Approach to Identify Depressed Users in Online Social Media," Front. Psychol., vol. 12, p. 802821, 2021.
[16] "Sequential Feature Selection - MATLAB & Simulink - MathWorks United Kingdom." https://uk.mathworks.com/help/stats/sequential-feature-selection.html (accessed Jul. 24, 2022).
[17] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic Minority Over-sampling Technique," J. Artif. Intell. Res., vol. 16, pp. 321–357, Jun. 2002.
[18] Kumbhar and Mali, "A survey on feature selection techniques and classification algorithms for efficient text classification," Int. J. Sci. Res. (Raipur), 2016.
[19] S. S. Bama* et al., "Identification of Default Payments of Credit Card Clients using Boosting Techniques," International Journal of Recent Technology and Engineering (IJRTE), vol. 8, no. 6. pp. 4990–4994, 2020. doi: 10.35940/ijrte.f8897.038620.