Nanotechnology Perceptions
ISSN 1660-6795
www.nano-ntp.com

Comparative Analysis of Machine Learning Strategies for Depression Detection in Indian College Students

Rajesh Saxena[1], Dr. Ashish Saini[2]

[1]*Research Scholar, Department of Computer Science & Engineering, Quantum University, Roorkee. Email: mr.rajeshsaxena24@gmail.com*
[2]*Assistant Professor, Department of Computer Science & Engineering, Quantum University, Roorkee. Email: ashishsainigkv7@gmail.com*

Depression is a significant psychological issue that affects college students worldwide. Research shows that 10% to 20% of students face mental health challenges such as stress, anxiety, and depression, which can hinder their development and lead to serious mental disorders if left untreated. This study investigates depression levels among undergraduate students and examines how factors like gender, location (urban or rural), academic discipline (Computer Science vs. Management), social class, year of study, and job satisfaction may influence these levels. To assess depression, BDC (Burns Depression Checklist), developed by David Burns, was used. A total of 572 students from various academic fields participated in the study. In this study, Six Machine Learning classifiers were applied to analyze socio-demographic and psychosocial data to predict depression. Results showed that the Naïve Bayes classifier, without feature selection, achieved the highest TPR (True Positive Rate) of 0.923, meaning it accurately identified most cases of depression. Logistic Regression demonstrated consistent precision (approximately 0.845), while KNN excelled with a TPR of 0.937 using the CfsSubsetEval technique. The study also found that Bagging and KNN had the highest ROC area values, around 0.981 and 0.969 respectively, indicating strong overall performance. This research highlights the potential of machine learning in identifying depression among college students, emphasizing the importance of early detection in addressing mental health concerns effectively.
Keywords: BDC, Depression, Feature Selection, Machine Learning, Mental Health.

1. Introduction

Depression, stress and anxiety are major issues that affect many teenagers and often go unrecognized and untreated. These conditions can have a profoundly impact on their academic achievement, familial relationships and self-perception. The World Health Organization (WHO) estimates that around 1 billion people worldwide suffer from mental disorders, with over 300 million suffering from depression [1]. Depression constitutes a predominant factor contributing to the manifestation of suicidal ideation, with an estimated global incidence of approximately 800,000 suicides annually. This highlights the need for immediate and much-needed attention to mental health issues. Depression affects not only a person's emotional well-being but also their social and economic life, often leading to isolation. Counseling and therapy are effective treatments. In India, one in five people suffers from depression, according to government studies. Despite growing awareness, mental health issues, especially among students, continue to rise. Causes include life events such as neglect, financial problems, relationship issues, and job-related stress, with college students being particularly affected. College students go through a very complex transition as they move from adolescence to adulthood. During this time, they face many new types of stressors, making this stage the most difficult period of their lives [2]. Undergraduate education is a critical period where mental health issues, particularly depression, are prevalent and can hinder students' academic and personal growth. Addressing these challenges necessitates the implementation of robust and well-informed intervention strategies. Machine learning (ML) is increasingly being used to help mental health by analyzing large amounts of data and identifying patterns that can predict depression and other conditions. Machine learning (ML) methodologies facilitate the early detection of mental health disorders by identifying nuanced behavioral patterns and deviations. This predictive ability allows healthcare professionals to intervene early and tailor treatments more effectively.

2. Related Work

The aspects of this kind of research problem are notably complex and need to do an in-depth investigation. This section has gone through several related research articles to find out the tools and techniques used in the existing works and determine the research gaps. Reddy et al. [3] conducted their analyses using data from a 2017 survey of technology employees, applying various models trained on this dataset. The initial dataset comprised 750 responses collected from persons employed across various technical departments, encompassing 68 attributes that pertain to both professional and personal dimensions. After the data cleaning process, 14 parameters remained, which were subsequently transformed into numerical values using one-hot encoding (1-of-n) and label encoding methods. Responses categorized as 'Yes', 'No', and 'Maybe' were assigned numerical values of 1, 0, and 0.5, respectively, with any NaN values replaced by 0. Nominal data were also converted into numerical values using label encoding. They selected models that had been previously validated in classification tasks and implemented them using Python's Scikit-learn library. The models included logistic regression, the k-nearest neighbor method, decision trees, random forest, boosting, and bagging algorithms. Each of these models was used to predict whether an individual required treatment, with model accuracies ranging from 69.43% to 75.13%. The bagging algorithm demonstrated the lowest accuracy, while the boosting algorithm achieved the highest accuracy. The survey results suggested that factors such as the gender of the individual, family history, and the availability of mental health services provided by employers had the most significant impact on stress and mental health. For future research, they recommend the application of deep learning (DL) techniques and the collection of more comprehensive and detailed datasets. They also propose refining the questionnaire to better align with the nature of the responses, increasing the number of attributes considered, and incorporating questionnaires from organizations like the WHO that focus on stress and mental health. Finally, they propose the establishment of a standardized instrument for the precise assessment and quantification of stress levels. Ekong et al. [4] proposed a soft computing-based approach that integrates neural networks, fuzzy logic, and case-based reasoning to assess the intensity of depression. Using various physical and psychological factors, they classified depression into five different categories: Almost nonexistent, Mild, Moderate, Severe, and Very severe. The results indicated that this method demonstrated better efficiency than the Diagnostic and Statistical Manual of Mental Disorders (DSM). (DSM). Hatton et al. [5] studied psychological and demographic data from a group of around 284 aged patients to estimate the prevalence of depression. The primary objective of their study was to forecast the persistence of depression using advanced machine learning techniques. To this end, they utilized the XGBoost (Extreme Gradient Boosting) algorithm, a robust ensemble learning method known for its high predictive accuracy and efficiency. The performance of XGBoost algorithm was evaluated in comparison to that of the Logistic Regression model, a conventional statistical method frequently employed in binary classification tasks. The results showed that XGBoost performed better than logistic regression models in predicting the prevalence of depression in elderly patients. Specifically, XGBoost provided higher predictive accuracy, indicating its greater effectiveness in handling complex, nonlinear relationships within the data. The study findings underline the potential of advanced machine learning algorithms such as Extreme Gradient Boosting to make more accurate and reliable predictions in the context of mental health assessment. This suggests a valuable route to improving the identification and management of depression in the elderly population, thereby contributing to more targeted and effective interventions. The authors [6] conducted analyses based on various demographic factors, including gender, age ranges, geographical location, and ethnicity. Additionally, they examined total annual income, employment status, cohabitation with a partner, and the age of the youngest child in the household. This allowed them to identify groups at risk and those impacted by COVID-19. Observations from years with limited data were excluded, which could have led to less precise estimates. Changes in mental health were evaluated through regression models that exclusively incorporated individuals who had data from both the COVID-19 survey and at least one dataset collected prior to the pandemic. Consequently, participants aged 16 and 17 were excluded from this analysis. The General Health Questionnaire (GHQ-12) index value was constructed during the pandemic and incorporated into a time-variable model, using average scores as the baseline rather than a binary index to avoid reducing the statistical power and generalizability of the results. The final model included the following factors: age, sex, family income, employment status, cohabitation with a partner, and the presence of risk factors.

3. Machine Learning Approach to Detect Depression

Machine learning, a branch of Artificial Intelligence, enables systems to learn from experience without direct programming. It uses algorithms that analyze data and improve over time. In healthcare, where vast data is generated, machine learning is transformative. It creates predictive models that reduce human error and speed up diagnoses. In this study, responses from students of different college of Saharanpur (India) were collected to create a dataset with 24 attributes and 1 target variable. This dataset was used to train machine learning models, which are known for their effectiveness in healthcare tasks, improving both prediction accuracy and efficiency.

Logistic Regression (LR):

LR is a predictive analysis method used for scenarios where a binary outcome is dependent on one or more predictor variables [7]. Logistic regression models the probability P (y=1⁄(X))that a student is depressed (target y=1) based on features X (such as behavioral and emotional attributes). The probability is given by the sigmoid function:

$$P(y = 1/X) = \frac{1}{1 + e^{-(\beta_0+\beta_1 X_1+\beta_2 X_2+\cdots+\beta_n X_n)}}$$

Here, β_0 is the intercept and β_0,……,β_n are the coefficients for the features X_1,…..,X_n. The model outputs a probability score, and a threshold (e.g.0.5) is applied to classify whether a student is likely depressed.

k-Nearest Neighbors (kNN) Classifier:

The k-Nearest Neighbors (kNN) classifier is a supervised learning algorithm that operates on labeled data. It classifies the dependent variable by evaluating the similarity between its independent variables and those of known instances within the dataset.[8]. Mathematically, for a input feature vector x, the objective is to predict its class label (e.g. depressed or not depressed). For each sample x_i in the training dataset, compute the distance between x and x_i commonly using Euclidean distance formula:

$$d(x, x_i) = \sqrt{\sum_{j=1}^{n}(x_j - x_{ij})^2}$$

Here x_j and x_ij are feature values of the input and training sample respectively. After this, identify the k nearest samples based on the distance d(x,x_i). The majority class among the k nearest neighbors determines the predicted class label for depression.

Multi-Layer Perceptron:

A Multi-Layer Perceptron (MLP) is a type of artificial neural network composed of an input layer, one or more hidden layers, and an output layer [9]. The input vector X=[x_1,x_2,....,x_n] represents the features (e.g. responses from questionnaires, demographic data, and academic performance) used for depression detection. Each hidden layer consists of neurons that apply a non-linear activation function to the weighted sum of inputs. For neuron j in a hidden layer, the output is given by:

$$h_j = f\left(\sum_{i=1}^{n} w_{ij}x_i + b_j\right)$$

Where $w_{ij}$are the weights connecting input x_i to neuron j. b_j is the bias term for neuron j. f(x) is the activation function.

The output layer provides the prediction (e.g. depressed or not depressed). For neuron k in the output layer, the output y_k is computed as:

$$y_k = f\left(\sum_{j=1}^{m} w_{jk}h_j + b_k\right)$$

Here w_{jk} are the weights connecting hidden neuron j to output neuron k. b_k is the bias term for the output neuron.

Naïve Bayes Classifier:

The Naive Bayes Classifier is a probabilistic machine learning model based on Bayes' theorem with the assumption of independence between features. In depression detection research, it is used to classify text data, such as social media posts or survey responses, by predicting the likelihood of depression-related content based on word frequencies or other features [10]. For depression detection, the model calculates the probability that a set of symptoms $X = \{x_1, x_2, \dots, x_n\}$ belongs to the class $C_{depressed}$ as follows:

$$P(C_{depressed}/X) = \frac{P(X/C_{depreesed}).P(C_{depressed})}{P(X)}$$

Here:

- $P(C_{depressed}/X)$ is posterior probability that the individual is depressed with given symptoms X.
- $P(X/C_{depressed})$ is likelihood of observing the symptoms X in depressed individuals.
- $P(C_{depressed})$ is prior probability of depression.
- P(X) is the total probability of observing the symptoms X.

The classifier then predicts “depressed” if $P(C_{depressed}/X)$ is higher than the probability for any other class (for example “not depressed”).

AdaBoost:

AdaBoost (Adaptive Boosting) is an ensemble learning method that combines multiple weak classifiers to form a powerful classifier. This technique repeatedly adjusts the weights of repeatedly misclassified samples so that difficult cases are given more attention in each iteration, thereby improving the overall accuracy. In depression detection research, AdaBoost can be utilized to enhance the performance of machine learning models by integrating various weak classifiers, such as decision trees or support vector machines, to better identify patterns and risk factors associated with depression. This results in more accurate detection of depressive symptoms and helps in the early intervention and support for affected individuals [11].

Bagging:

Bootstrap aggregating (or bagging) is a method in which multiple models are built on small parts of the data. Then the predictions of all these models are combined to get the final result. This method helps in increasing the stability and accuracy of the model, as it combines the results obtained from different parts and reduces the probability of error. [12]. Bagging for depression detection means that we create many small data groups using the bootstrapping method. These groups are selected from the main data in such a way that the same data point can be selected multiple times. If D be the original dataset containing. N samples and D_i represent the i^th bootstrapped dataset where |D_i |=N. Each model M_i is trained on its corresponding subset D_i. For a new observation x, each model produces a prediction y_i. The final prediction y is obtained by aggregating the predictions, often through majority voting for classification tasks, expressed mathematically as:

$$y = \mathrm{mode}(y_1, y_2, \dots \dots, y_B)$$

Here B is the total number of models.

Random Forest:

A Random Forest classifier is a technique based on ensemble learning method that builds multiple decision trees from random subsets of the data and features. For depression detection, each tree in the forest is trained on a bootstrapped subset of the dataset, where the input features may include various indicators like mood scores, activity levels, or responses to questionnaires. At each node in a tree, a random subset of features is considered to split the data. The trees output their predicted label (e.g., "depressed" or "not depressed"), and the final classification is determined by majority voting across all trees. Mathematically, if $T_1, T_2, \dots\dots, T_n$ are the trees in the forest and each $T_i(x)$ represents the prediction of the i^{th} tree for inputx, the final prediction y is the mode of all individual predictions:

$$y = \mathrm{mode}\big(T_1(x), T_2(x), \dots\dots, T_n(x)\big)$$

4. Methodology

This section is organized into seven subsections, detailing the complete methodology of the study.

4.1 Data Description:

The study involved Indian undergraduate students from different colleges in the Saharanpur region. Data was collected via a survey conducted from March to June 2024. A 24-question structured questionnaire, based on a modified version of the Burns Depression Checklist (BDC), gathered comprehensive psycho-social and socio-demographic information from participants. [13]. The BDC was used to accurately assess each participant's emotional state. Participants rated the severity of various depressive symptoms they experienced in the days leading up to the survey, including the day immediately before. In this version of the BDC, symptoms were rated on a scale from 0 to 4, focusing on depression-specific indicators rather than general mental health symptoms. The total BDC score was calculated by summing the severity ratings for each symptom. A score greater than 10 indicated that the individual was classified as depressed. A total of 572 students took part in the study, offering insights into depression prevalence among college students. The dataset contains 24 predictor variables and a target variable, derived from the BDC. Predictors include demographic factors (e.g., gender, age), academic factors (e.g., course, year, grades), and social factors (e.g., type of residence, satisfaction with environment and academics). Mental health variables, such as insomnia, anxiety, and suicidal thoughts, were also recorded. Additional factors like social media use, smoking habits, financial stress, and recent conflicts or trauma were included. The target variable indicates whether a participant is considered depressed.

Table 1. Predictive variables for assessing depression

S.No	Variable Name	Variable Type	Variable Description	Domain
1	GEN	Predictor	Gender of the participant	Male, Female
2	AGE	Predictor	Age of participant in years	18 - 24 Years
3	COURSE	Predictor	Current course of participant	B.Sc (Comp. Sci.), B.Com, MBA
4	YEAR	Predictor	Current year of course	1 - 3 Year
5	PERMAR	Predictor	Percentage of marks in last year	55% - 87%
6	TYPRES	Predictor	Type of residing place	Village, Town, City
7	SATENV	Predictor	Satisfaction with living environment	Yes, No
8	SATACA	Predictor	Satisfaction with academic performance	Yes, No
9	FINSTR	Predictor	Facing financial stress	Yes, No
10	SMOKE	Predictor	Smoking habits	Yes, No
11	ILLSER	Predictor	Facing serious illness	Yes, No
12	EATDIS	Predictor	Suffering from eating disorders	Yes, No
13	INSOSUF	Predictor	Suffering from insomnia	Yes, No
14	AVGSLP	Predictor	Average sleep hours at night	3 - 8 hours
15	AVGSOM	Predictor	Average hours on social networks	1 - 6 hours
16	PRESSTD	Predictor	Pressure due to study	Severe, Moderate, Mild, No Pressure
17	FLTANX	Predictor	Recently felt anxiety	Yes, No
18	FLTABU	Predictor	Recently felt abused	Yes, No
19	FLTCHT	Predictor	Recently felt cheated	Yes, No
20	FLTDEP	Predictor	Recently felt depressed	Yes, No
21	SUIDTH	Predictor	Recent suicidal thoughts	Yes, No
22	SUFINF	Predictor	Recently felt inferior	Yes, No
23	CONFSOM	Predictor	Conflicts with friends or family	Yes, No
24	LSTSOM	Predictor	Recently lost someone close	Yes, No
25	DEPRESSED	Target	Whether the participant is depressed	0 = Not Depressed, 1 = Depressed

4.2 Data Analysis

Out of 572 people in the study, 63.81% (365) are depressed. Depression is higher among females (68.53%) compared to males (62.23%). Village residents have the highest depression rate (70.04%), followed by town (61.72%) and city residents (58.54%). Among students, depression rates are 63.88% for MBA, 62.80% for B.Sc (CS), and 64.33% for B.Com programs, highlighting notable variations across demographics and academic disciplines.

Table 2. Distribution of depressed and non-depressed participants in the dataset

Class	Number of Students	Percentage
Depressed	365	63.81%
Not Depressed	207	36.19%

Table 3. Distribution of depressed and non-depressed participants in the dataset based on various criteria

Criteria	Category	Total Participants	Number of Depressed Participants	Number of Not Depressed Participants	Depressed (%)	Not Depressed (%)
Gender	Male	429	267	162	62.24%	37.76%
	Female	143	98	45	68.53%	31.47%
Course	B.Sc (C.S.)	164	103	61	62.80%	37.20%
	B.Com	300	193	107	64.33%	35.67%
	MBA	108	69	39	63.89%	36.11%
Type of Residence	Village	217	152	65	70.05%	29.95%
	Town	162	100	62	61.73%	38.27%
	City	193	113	80	58.55%	41.45%

4.3 Data Preprocessing and Feature Selection

From an initial dataset of 618 responses with 24 attributes, we excluded 46 incomplete responses, leaving 604 valid responses for the final analysis. Feature selection is very important in machine learning because by removing irrelevant attributes, it is possible to enhance the performance of the model. In this study, WEKA [14], a well-known machine learning and data mining software. This software was developed by Waikato University, New Zealand, and can be used for classification to identify hidden patterns within the dataset. WEKA supports various machine learning tasks, including data pre-processing, classification, regression, and prediction [15]. It was the primary tool for this research, enabling the comparison of different methods to identify the most effective approach for predicting depression. Selecting only the most relevant features is crucial, as irrelevant ones can negatively impact a model's performance. In WEKA, feature selection reduces dataset dimensionality, improving overall efficiency. WEKA provides various evaluators to assess the importance of each attribute, ensuring optimal model performance.

4.3.1 Correlation-based Feature Subset Evaluation

CfsSubsetEval (Correlation-based Feature Subset Evaluation) in WEKA evaluates feature subsets by measuring their correlation with the target variable while minimizing redundancy among features. It selects features that are highly predictive of the target but have low inter-correlation, ensuring each contributes unique information. By focusing on informative, non-redundant features, CfsSubsetEval enhances model efficiency and accuracy, improving performance while reducing dimensionality and avoiding overfitting in machine learning models. The mathematical formula for CfsSubsetEval is given by:

$$M_s = \frac{k.\overline{r_{cf}}}{\sqrt{k + k.(k-1).\overline{r_{ff}}}}$$

Where:

- $M_s =$ Merit of the feature subset.
- $k =$ Number of features in the subset
- $\overline{r_{cf}} =$ Average correlation between the features in the subset and the target varibale.
- $\overline{r_{ff}} =$ Average intercorrelation among the features in the subset

4.3.2 GainRatioAttributeEval

GainRatioAttributeEval is a feature selection method in WEKA that assesses the importance of attributes based on information gain, adjusted for the number of distinct values. This method identifies informative features for predicting the target variable while reducing bias towards attributes with many values. Information gain quantifies the reduction in uncertainty about the class label provided by a feature, while the gain ratio accounts for the intrinsic information related to an attribute's ability to split the dataset, enhancing the evaluation process. Following formula is used to calculate the Gain Ratio:

$$\text{Gain Ratio} = \frac{\text{Information Gain}}{\text{Intrinsic Value}}$$

Where:

$$\text{Information Gain} = H(C) - H(C/A)$$

- $H(C) =$ Entropy of the class variable C
- $H(C/A) =$ Conditional entropy of the class given the attribute A

$$\text{Intrinsic Value} = -\sum_{i=1}^{n} \frac{N_i}{N} \log_2\left(\frac{N_i}{N}\right)$$

- $N_i =$ Number of instance in the i^{th} value of
- attritube A
- $N =$ Total number of instances in the dataset

4.3.3 CorrelationAttributeEval

CorrelationAttributeEval is a feature selection method in WEKA that assesses the relevance of attributes by measuring their correlation with the target variable. It utilizes Pearson's correlation coefficient to quantify the linear relationship between each feature and the class, yielding a score that indicates both the strength and direction of this relationship. This method is straightforward and efficient, allowing for the easy identification of features with strong connections to the target variable. By selecting features that significantly enhance predictive accuracy and discarding irrelevant ones, CorrelationAttributeEval improves overall model performance. The Pearson correlation coefficient is calculated as:

$$r = \frac{\mathrm{Cov}(X, Y)}{\sigma_X \sigma_Y}$$

Where:

- r = Pearson correlation cofficient
- $\mathrm{Cov}(X, Y)$ = Covariance between X and Y
- σ_X = Standard deviation of X
- σ_Y = Standard deviation of Y

4.4 Implementation Procedures

This research was conducted using WEKA software, with data from 572 participants. The dataset included one target variable (indicating depression status) and 24 predictor variables covering demographic, academic, and behavioral factors. After data preparation, the dataset was split into 80% training and 20% testing subsets. Both the training and testing data were encoded to convert categorical variables into numerical values, as machine learning models generally perform better with numeric data. To improve model performance, feature selection was applied to remove irrelevant or redundant variables that could reduce classifier efficiency. Three feature selection techniques were used independently to identify the most relevant predictors, ensuring that the model focused on key variables. This approach aimed to enhance accuracy and generalizability by reducing dataset dimensionality.

Table 4. Selected Features Derived from Feature Selection Techniques

Technique	Total Features	Selected Features
CfsSubsetEval	11	SATENV, SATACA, FINSTR, INSOSUF, FLTANX, FLTABU, FLTCHT, FLTDEP, SUFINF, CONFSOM, LSTSOM
CorrelationAttributeEval	9	SATENV, SATACA, FINSTR, FLTANX, FLTABU, FLTCHT, FLTDEP, SUFINF, CONFSOM
GainRatioAttributeEval	12	SATENV, SATACA, FINSTR, INSOSUF, FLTANX, FLTABU, FLTCHT, FLTDEP, SUIDTH, SUFINF, CONFSOM, LSTSOM

In machine learning, evaluating the performance of models is crucial, and several key metrics help us understand how well a model is performing. One important metric is the True Positive Rate (TPR), also known as Recall. This metric measures the proportion of actual positive cases that the model correctly identifies. For instance, if a model correctly predicts 80 out of 100 actual positive cases, the recall would be 0.8 or 80%. Conversely, the False Positive Rate (FPR) assesses how many negative cases are incorrectly classified as positive, and ideally, this value should be as close to 0 as possible. Precision is another vital metric that indicates the accuracy of the positive predictions made by the model. For example, if a model predicts 50 cases as positive and 40 of these predictions are correct, the precision would be 0.8. The F1-Measure is a combined metric that balances both precision and recall, with a good F1 score typically being above 0.7. The Matthews Correlation Coefficient (MCC) is a more comprehensive metric that accounts for all four values in a confusion matrix: True Positives (TP), False Positives (FP), True Negatives (TN), and False Negatives (FN). MCC values range from -1 to +1, with +1 indicating a perfect prediction model. Additionally, the Receiver Operating Characteristic (ROC) curve illustrates the trade-off between TPR and FPR, and a high area under the curve (AUC) is ideal, nearing 1. The Precision-Recall Curve (PRC) focuses specifically on precision and recall, especially important in imbalanced datasets, with a higher AUC indicating better performance. In our analysis, we train multiple classifiers, including Logistic Regression, Naïve Bayes, K-Nearest Neighbors (KNN), Random Forest, AdaBoost, and Bagging, using training datasets. After training, each classifier is tested for its ability to predict depression among participants in test datasets. We calculate various performance metrics—such as accuracy, precision, sensitivity, specificity, F1-score, and AUC—for each model. This comprehensive evaluation helps us determine which classifier is most effective at accurately identifying cases of depression. Ultimately, selecting the optimal model is essential for ensuring reliable predictions and effectively identifying depression among participants.

$$\text{Accuracy } (\%) = \left(\frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}\right) \times 100$$

$$\text{Sensitivity or Recall}(\%) = \left(\frac{\text{TP}}{\text{TP} + \text{FN}}\right) \times 100$$

$$\text{Precision}(\%) = \left(\frac{\text{TP}}{\text{TP} + \text{FP}}\right) \times 100$$

$$\text{F1 Score} = 2 \times \left(\frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}\right)$$

Here are the definitions in simpler terms:

- True Positive (TP): This happens when the model correctly predicts that a participant is depressed.
- True Negative (TN): This occurs when the model correctly predicts that a participant is not depressed.
- False Positive (FP): This is when the model incorrectly predicts that a participant is depressed when they are not.
- False Negative (FN): This happens when the model incorrectly predicts that a participant is not depressed when they actually are.

Table 5. Confusion and Performance Matrices of Classifiers with Various Feature Selection Techniques

Classifier Name	Feature Selection Technique	TP	TN	FP	FN	TP Rate	FP Rate	Precision	F-Measure	MCC	ROC Area
Naïve Bayes	Without using Feature Selection Technique	191	318	47	16	0.923	0.129	0.803	0.858	0.774	0.961
Logistic Regression		186	331	34	21	0.899	0.093	0.845	0.871	0.796	0.969
KNN		189	275	90	18	0.913	0.247	0.677	0.778	0.641	0.87
AdaBoost		169	318	47	38	0.816	0.129	0.782	0.799	0.682	0.93
Random Forest		151	329	36	56	0.729	0.099	0.807	0.766	0.646	0.927
Bagging		174	351	14	33	0.841	0.038	0.926	0.881	0.821	0.97
Naïve Bayes	Feature selection using CfsSubsetEval technique	188	320	45	19	0.908	0.123	0.807	0.855	0.768	0.957
Logistic Regression		176	335	30	31	0.85	0.082	0.854	0.852	0.769	0.959
KNN		194	332	33	13	0.937	0.09	0.855	0.894	0.832	0.981
AdaBoost		169	318	47	38	0.816	0.129	0.782	0.799	0.682	0.93
Random Forest		187	339	26	20	0.903	0.071	0.878	0.89	0.827	0.979
Bagging		171	336	29	36	0.826	0.079	0.855	0.84	0.752	0.956
Naïve Bayes	Feature selection using CorrelationAttributeEval technique	181	319	46	26	0.874	0.126	0.797	0.834	0.735	0.944
Logistic Regression		167	336	29	40	0.807	0.079	0.852	0.829	0.736	0.945
KNN		182	327	38	25	0.879	0.104	0.827	0.852	0.766	0.958
AdaBoost		168	317	48	39	0.812	0.132	0.778	0.794	0.674	0.904
Random Forest		178	331	34	29	0.86	0.093	0.84	0.85	0.763	0.957
Bagging		172	329	36	35	0.831	0.099	0.827	0.829	0.732	0.94
Naïve Bayes	Feature selection using GainRatioAttributeEval technique	190	321	44	17	0.812	0.918	0.121	0.812	0.862	0.779
Logistic Regression		176	336	29	31	0.859	0.85	0.079	0.859	0.854	0.772
KNN		194	334	31	13	0.862	0.937	0.085	0.862	0.898	0.838
AdaBoost		169	318	47	38	0.782	0.816	0.129	0.782	0.799	0.682
Random Forest		188	340	25	19	0.883	0.908	0.068	0.883	0.895	0.835
Bagging		171	337	28	36	0.859	0.826	0.077	0.859	0.842	0.756

5. Results and Discussion

In evaluating the performance of various machine learning classifiers with different feature selection techniques, we can gain valuable insights into their effectiveness using several metrics. These metrics include True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN), along with derived metrics like True Positive Rate (TPR), False Positive Rate (FPR), Precision, F-Measure, Matthews Correlation Coefficient (MCC), and the Area Under the Receiver Operating Characteristic curve (ROC Area).

When no feature selection technique is employed, Logistic Regression stands out as the top performer. It achieves a high Precision of 0.845 and an impressive F-Measure of 0.871, along with an excellent ROC Area of 0.969. This performance indicates a strong balance between sensitivity (the ability to correctly identify positive cases) and specificity (the ability to correctly identify negative cases). In contrast, the Bagging classifier also demonstrates strong performance, boasting an even higher Precision of 0.926 and the lowest False Positive Rate of 0.038. Its F-Measure of 0.881 and an MCC of 0.821 further validate its effectiveness.

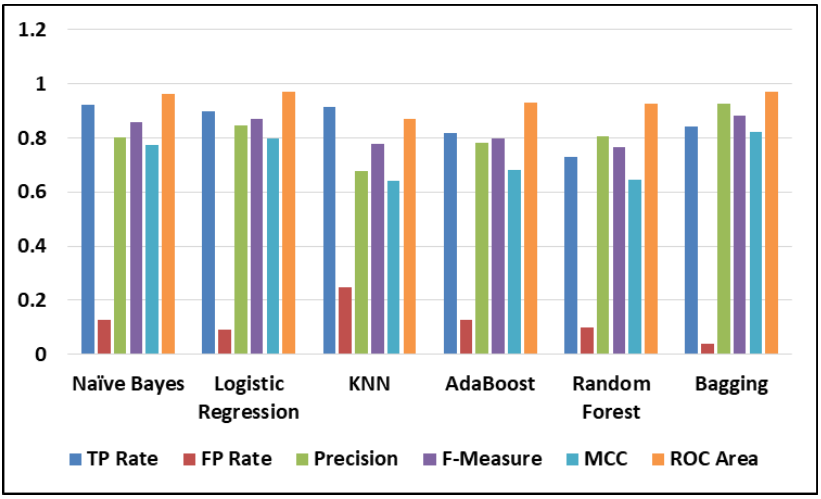

Figure 1. Comparison of ML models without feature selection

Applying the CfsSubsetEval feature selection technique shifts the performance landscape, with K-Nearest Neighbors (KNN) achieving the highest TPR at 0.937, along with a Precision of 0.855 and an F-Measure of 0.894. This shows that KNN effectively captures the relevant features after selection. The Random Forest classifier also shows marked improvement, attaining a robust F-Measure of 0.89 and the lowest False Positive Rate of 0.071 among all classifiers.

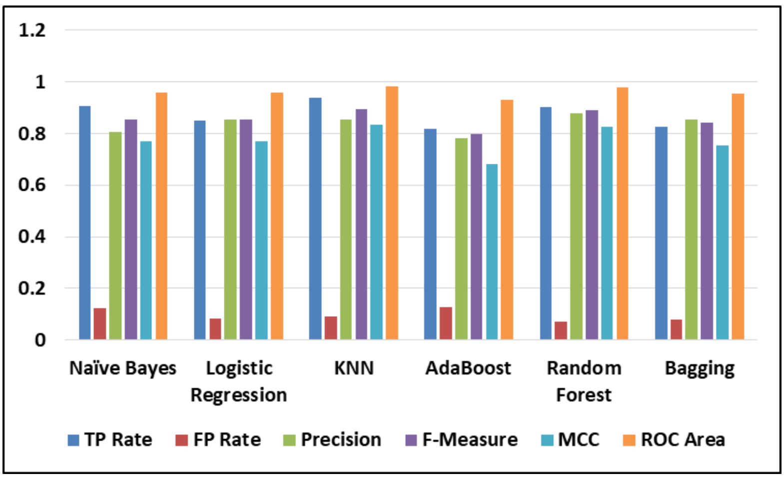

Figure 2. Comparison of ML models with CfsSubsetEval technique

When utilizing the CorrelationAttributeEval technique, KNN maintains its lead, with a high

Precision of 0.827, an F-Measure of 0.852, and a commendable MCC of 0.766. Its ROC Area is also notable at 0.958, further solidifying its status as a leading model. Meanwhile, Random Forest and Bagging deliver comparable results, with F-Measures around 0.85 and ROC Areas near 0.957.

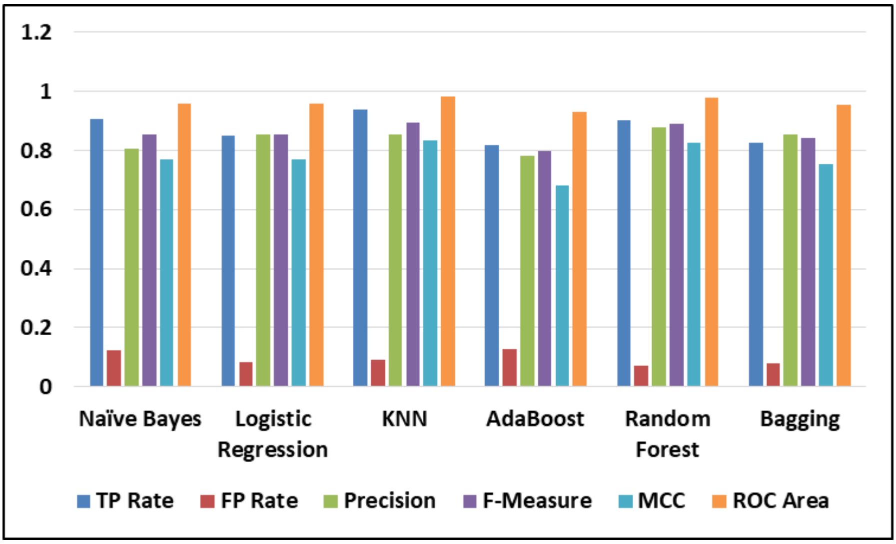

Figure 3. Comparison of ML models with CorrelationAttributeEval technique

With the GainRatioAttributeEval technique, Random Forest excels again, achieving a high TPR of 0.883 and an F-Measure of 0.895, alongside the highest ROC Area of 0.835, highlighting its superior performance. KNN also performs admirably, with an F-Measure of 0.898 and the highest MCC of 0.838. These results mark both models as reliable choices when utilizing this feature selection method.

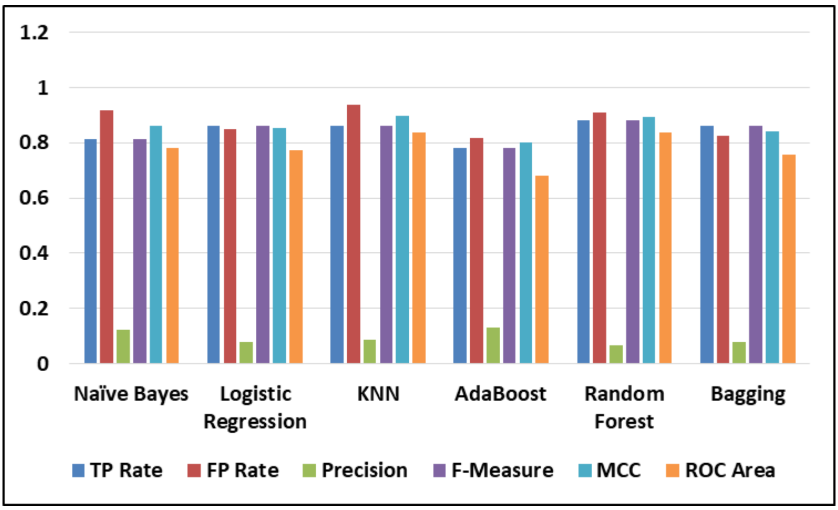

Figure 4. Comparison of ML models with GainRatioAttributeEval technique

Throughout the study, KNN and Random Forest consistently demonstrate superior performance across multiple feature selection techniques. For instance, using the CfsSubsetEval technique, KNN achieves a TPR of 0.937, indicating it correctly identifies a high proportion of actual positive cases. Random Forest, with a TPR of 0.903, also shows effectiveness in minimizing false negatives. This capability is crucial in applications like medical diagnosis or fraud detection, where missing a positive case can have serious implications.

In terms of Precision, KNN's score of 0.855 when utilizing the CfsSubsetEval technique suggests that a significant majority of its positive predictions are accurate. Random Forest also performs well, with a Precision score of 0.878, highlighting its reliability. The F-Measure, which balances Precision and Recall, shows KNN's robust performance at 0.894 compared to Random Forest's 0.890.

The MCC scores reflect the models' comprehensive performance, with KNN achieving 0.832 and Random Forest at 0.827 using the CfsSubsetEval technique, both indicating strong predictive capabilities across all confusion matrix elements. The ROC Area values of KNN at 0.981 and Random Forest at 0.979 confirm their effectiveness in distinguishing between positive and negative classes.

The strong performance of KNN and Random Forest can be attributed to their inherent characteristics. KNN's instance-based learning allows it to adapt well to local patterns, making it effective for complex, non-linear relationships. On the other hand, Random Forest's ensemble learning approach aggregates predictions from multiple decision trees, enhancing robustness against overfitting and improving generalization across diverse datasets.

In conclusion, this analysis indicates that KNN and Random Forest are the best-performing classifiers evaluated, particularly due to their high true positive rates, precision, and balanced performance as indicated by F-Measure and MCC. Their strengths in handling data complexities underscore their suitability for applications that demand accurate and reliable predictions.

6. Conclusion and Limitations

In this study, we assessed the effectiveness of various machine learning classifiers in detecting depression among 572 undergraduate students using the Beck Depression Inventory (BDI) scale. Our primary goal was to evaluate how different feature selection techniques impact the performance of these classifiers, and the results revealed significant variations in their effectiveness. We found that the Naïve Bayes classifier consistently achieved high true positive rates (TPR) and area under the ROC curve (AUC), both with and without feature selection. This performance indicates that Naïve Bayes is a robust method for identifying individuals with depression, making it a reliable choice in mental health diagnostics. In addition, the Logistic Regression model demonstrated promising results, particularly when feature selection techniques were employed. The incorporation of these techniques led to enhancements in the model's precision and F-measure, showcasing the importance of feature selection in improving classifier performance.

Among the classifiers tested, the K-Nearest Neighbors (KNN) classifier stood out when paired with the CfsSubsetEval feature selection method. This combination yielded the highest recall rate and a commendable F-measure, suggesting that KNN is particularly effective in identifying depressed individuals when the right features are selected. In contrast, the Random Forest and AdaBoost models displayed lower overall effectiveness in detecting depression. This finding indicates that these models may require further optimization to enhance their predictive capabilities.

The study's findings emphasize the crucial role of feature selection in boosting the performance of classifiers designed for depression detection among students. Feature selection not only streamlines the data but also enhances the model's ability to identify significant patterns associated with depression. Consequently, for future research, we recommend exploring additional feature selection techniques and integrating deep learning approaches. These methods may yield better performance and more accurate predictions, particularly in larger and more diverse datasets. Moreover, it is essential to consider the inclusion of various demographic variables, such as age, gender, and socioeconomic status, in future studies. Analyzing how these factors impact classifier performance could provide valuable insights into creating effective mental health interventions tailored to specific student populations.

One limitation of our study is that we relied solely on the BDC as the ground truth for diagnosing depression, which may not capture the full spectrum of the condition. The dataset did not include any biological markers, which are significant in predicting depression and could enhance the model's predictive efficiency. Future research could address this gap by incorporating biological factors alongside psychological assessments, allowing for a more comprehensive understanding of depression.

Additionally, while this study focused on predicting the presence of depression, extending the research to determine the severity of depression could provide deeper insights into student mental health. Understanding the nuances of depression severity may lead to more targeted interventions. Several studies suggest that applying different dimensionality reduction algorithms during data preprocessing can further improve model performance. We encourage future researchers to explore these methods and compare their results with those obtained in this study.

Overall, the findings underscore the potential of machine learning techniques in mental health diagnostics. By developing automated screening tools, we can facilitate early intervention and support for students at risk of depression, ultimately improving mental health outcomes in this vulnerable population.

Data Availability Statement: The dataset generated and analyzed during this study are not publicly available duo to privacy restrictions.

Conflict of Interest Statement: There is no conflict of interest.

References

[1] T. Karmakar and S. K. Behera, “Depression among the College Students: An Empirical Study,” Educ. Quest- An Int. J. Educ. Appl. Soc. Sci., vol. 8, no. 1, p. 163, 2017, doi: 10.5958/2230-7311.2017.00024.1.

[2] A. S. Chokshi, P. P. Rangwala, G. H. Dumra, M. R. Thakrar, A. J. Singh, and B. M. Lakdawala,

"Depression, anxiety and stress amongst students in science verses non-science stream: a comparative study," Int. J. Community Med. Public Heal., vol. 8, no. 7, p. 3461, 2021, doi: 10.18203/2394-6040.ijcmph20212602.

[3] U. S. Reddy, A. V. Thota, and A. Dharun, "Machine Learning Techniques for Stress Prediction in Working Employees," 2018 IEEE Int. Conf. Comput. Intell. Comput. Res. ICCIC 2018, no. May, 2018, doi: 10.1109/ICCIC.2018.8782395.

[4] V. E. Ekong and E. A. Onibere, "A Softcomputing Model for Depression Prediction," Egypt. Comput. Sci. J., vol. 39, no. 4, pp. 1–21, 2015, [Online]. Available: https://www.researchgate.net/publication/282960557_A_Softcomputing_Model_for_Depression_Prediction

[5] 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, 2019, https://www.sciencedirect.com/science/article/abs/pii/S0165032718319931?via%3Dihub

[6] M. Pierce et al., "Mental health before and during the COVID-19 pandemic: a longitudinal probability sample survey of the UK population," The Lancet Psychiatry, vol. 7, no. 10, pp. 883–892, 2020, doi: 10.1016/S2215-0366(20)30308-4.

[7] M. Maalouf, "Logistic regression in data analysis: An overview," Int. J. Data Anal. Tech. Strateg., vol. 3, no. 3, pp. 281–299, 2011, doi: 10.1504/IJDATS.2011.041335.

[8] G. Guo, H. Wang, D. Bell, Y. Bi, and K. Greer, "KNN model-based approach in classification," Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 2888, no. November 2012, pp. 986–996, 2003, doi: 10.1007/978-3-540-39964-3_62.

[9] P. Marius-Constantin, V. E. Balas, L. Perescu-Popescu, and N. Mastorakis, "Multilayer perceptron and neural networks," WSEAS Trans. Circuits Syst., vol. 8, no. 7, pp. 579–588, 2009.

[10] Rish I, "An empirical study of the naive bayes classifier," IJCAI 2001 Work. Empir. methods Artif. Intell., no. January 2001, pp. 41–46, 2001, [Online]. Available: https://sites.cc.gatech.edu/home/isbell/classes/reading/papers/Rish.pdf

[11] C. Tu, H. Liu, and B. Xu, "AdaBoost typical Algorithm and its application research," MATEC Web Conf., vol. 139, 2017, doi: 10.1051/matecconf/201713900222.

[12] S. B. Kotsiantis, G. E. Tsekouras, and P. E. Pintelas, "Bagging model trees for classification problems," Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 3746 LNCS, no. December 2013, pp. 328–337, 2005, doi: 10.1007/11573036_31.

[13] D. Burns, H. Westra, M. Trockel, and A. Fisher, "Motivation and changes in depression," Cognit. Ther. Res., vol. 37, no. 2, pp. 368–379, 2013, doi: 10.1007/s10608-012-9458-3.

[14] S. Rai and A. Desai, "Analysis of Machine Learning Algorithms using Weka," Int. J. Comput. Appl., vol. *-No.*, no. February 2013, pp. 975–8887, 2011, [Online]. Available: https://www.researchgate.net/publication/259756674_Analysis_of_Machine_Learning_Algorithms_using_Weka

[15] G. Naga and R. Devi, "Comparative Study on Machine Learning Algorithms using Weka," pp. 102–113, 2014, [Online]. Available: https://www.ijert.org/comparative-study-on-machine-learning-algorithms-using-weka