A Comprehensive Machine Learning Framework for Classification of Depression from Survey-Based Behavioural Data

Supervised By

Mr. Md Rajib Mia

Lecturer (Senior Scale)

Department of Software Engineering

Daffodil International University

Submitted By

MD. Tanvir Hossain

221-35-1013

Department of Software Engineering

Daffodil International University

This thesis report has been submitted in fulfilment of the requirements for the Degree of Bachelor of Science in Software Engineering.

APPROVAL

This thesis titled on **"A Comprehensive Machine Learning Framework for Classification of Depression from Survey-Based Behavioural Data"** submitted by Student Name**: Md. Tanvir Hossain (ID: 221-35-1013)** to the Department of Software Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of Bachelor of Science in Software Engineering and approval as to its style and contents.

BOARD OF EXAMINERS

Fazla Elahe

-- **Chairman**

Dr. Fazla Ealhe
Assistant Professor & Associate Head
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University

Marzia

-- **Internal Examiner 1**

Dr. Marzia Ahmed
Assistant Professor
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University

-- **Internal Examiner 2**

Dr. Shabnom Mustary
Assistant Professor
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University

-- **Internal Examiner 3**

Md. Rajib Mia
Lecturer (Senior Scale)
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University

-- **External Examiner**

Mohammod Abul Kashem, PhD
Professor
Department of Computer Science and Engineering
DUET, Bangladesh

A Comprehensive Machine Learning Framework for Classification of Depression from Survey-Based Behavioural Data

Md. Tanvir Hossain
Id: 221-35-1013

Bachelor of Science

DAFFODIL INTERNATIONAL UNIVERSITY

DAFFOFIL INTERNATIONAL UNIVERSITY

DECLARATION OF THESIS AND COPYRIGHT

Author's Full Name	:	Md. Tanvir Hossain
Date of Birth	:	07-03-2002
Title	:	A Comprehensive Machine Learning Framework for Classification of Depression from Survey-Based Behavioural Data.
Academic Session	:	2022-2025

I declare that this thesis is classified as:

☐	CONFIDENTIAL	(Contains confidential information under the Official Secret Act 1997)*
☐	RESTRICTED	(Contains restricted information as specified by the organization where research was done)*
☐	OPEN ACCESS	I agree that my thesis to be published as online open access (Full Text)

I acknowledge that Daffodil International University reserves the following rights:

1. The Thesis is the Property of Daffodil International University.
2. The Library of Daffodil International University has the right to make copies of the thesis for the purpose of research only.
3. The Library of Daffodil International University has the right to make copies of the thesis for academic exchange.

Certified by:

Tanvir
(Student's Signature)

221-35-1013
Date: 24-12-2025

(Supervisor's Signature)

Mr. Md Rajib Mia
Date: 24.12.2025

THESIS DECLARATION LETTER (OPTIONAL)

Librarian,
Daffodil International University,
Daffodil Smart City,
Ashulia.Dhaka,Bangladesh

Dear Sir,

CLASSIFICATION OF THESIS AS RESTRICTED

Please be informed that the following thesis is classified as RESTRICTED for a period of three (3) years from the date of this letter. The reasons for this classification are as listed below.
Author's Name
Thesis Title

Reasons (i)

(ii)

(iii)

Thank you.

Yours faithfully,

____
(Supervisor's Signature)

Date:

Stamp:

Note: This letter should be written by the supervisor and addressed to the Librarian, *Daffodil International University* with its copy attached to the thesis.

SUPERVISOR'S DECLARATION

I hereby declare that I have checked this Thesis entitled "**A Comprehensive Machine Learning Framework for Classification of Depression from Survey-Based Behavioural Data**", and in my opinion, this thesis is adequate in terms of scope and quality for the award of the degree of Bachelor of Science in Software Engineering.

(Supervisor's Signature)

Full Name : Mr Md Rajib Mia

Position : Lecture (Senior Scale), Department of Software Engineering, Daffodil International University

Date : 24 December 2025

STUDENT'S DECLARATION

I hereby declare that the work in this thesis is based on my original work except for quotations and citations which have been duly acknowledged. I also declare that it has not been previously or concurrently submitted for any other degree at Daffodil International University or any other institution.

Tanvir
(Student's Signature)

Full Name : Md Tanvir Hossain
ID Number : 221-35-1013
Date : 24 December 2025

A Comprehensive ML Framework for Classification of Depression from Survey-Based Behavioural Data

MD. TANVIR HOSSAIN
Id: 221-35-1013

Thesis submitted in fulfillment of the requirements
for the award of the degree of
Bachelor of Science

Department of Software Engineering

DAFFODIL INTERNATIONAL UNIVERSITY

DECEMBER 2025

ACKNOWLEDGEMENTS

This work would not have been possible without the support and contributions of many individuals over the past two semesters. We are deeply grateful to everyone who has assisted us in one way or another.

First, we express our heartfelt thanks and gratefulness to the almighty for His divine blessing making it possible for us to complete the defense successfully.

We really grateful and wish our profound our indebtedness to Mr. Md Rajib Mia, Lecturer (Senior Scale), Department of SWE Daffodil International University, Dhaka. Deep Knowledge & keen interest of our supervisor in the field of "Machine Learning" to carry out this project. His endless patience,s cholarly guidance ,continual encouragement , constant and energetic supervision, constructive criticism, valuable advice, reading many inferior draft and correcting them at all stage have made it possible to complete this project.

We would like to express our heartfelt gratitude to the Head of the Department of Computer Science and Engineering, for his kind help in finishing our project and also to other faculty members and the staff of the Department of Computer Science and Engineering, Daffodil International University.

We would like to thank our entire course mate in Daffodil International University, who took part in this discuss while completing the course work.

Finally, we must acknowledge with due respect the constant support and patients of our parents.

DEDICATION

This thesis is lovingly dedicated to my parents, whose constant support, guidance, and silent sacrifices have shaped every step of my journey. Their belief in my abilities has been my greatest strength, especially throughout the challenges of this research.

I also dedicate this work to my family and well-wishers, whose encouragement and prayers have motivated me to keep moving forward. Their presence has reminded me that determination and hope can turn hard work into meaningful achievement.

ABSTRACT

This paper will anticipate depression among people by analyzing various demographic, social and economic variables that are a combination of their everyday life and economic statuses. The variables provided in the dataset are sex, age, marital status, family size, education level, asset conditions, source of income and spending, and investment behavior, all that were chosen since they have an indirect way of showing the level of emotional and psychological stress. The data was refined properly before the analysis proper, in terms of filling in missing values, coding categorical variables, and scaling the numerical variables to ensure similarity across the variables. There were 14 ML models used on both the original dataset and a SMOTE-balanced version so that there was equal treatment of both the imbalanced and balanced performance. Accuracy, precision, recall, F1-score and AUC-ROC were used to evaluate each model. Random Forest, XGBoost, LightGBM, Stacking, and Voting Classifier algorithms were the most useful and have demonstrated the greatest accuracy at 0.9755 on the original dataset as well as high precision and F1-scores. Ensemble-based models also performed well in the SMOTE dataset, with LightGBM and the Random Forest achieving more than 0.97 accuracy. These results show that prediction of depression is very effective when there is strong ensemble learning, and the use of structured socioeconomic and lifestyle-based data is employed. The general findings show that ML can be used to early detect depression risks, particularly in settings where the psychological assessment resources are scarce.

TABLE OF CONTENT

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ML	Machine Learning
DL	Deep Learning
LR	Logistic Regression
DT	Decision Tree
NB	Naïve Bayes
SVM	Support Vector Machine
KNN	K-Nearest Neighbour
XGBoost	Extreme Gradient Boosting
LightGBM	Light Gradient Boosting Machine
CatBoost	Categorical Boosting

CHAPTER 1

INTRODUCTION

1.1 Introduction

The condition of depression is widespread, multifaceted, and usually hidden by the normal course of actions, which makes it difficult to detect without organized assistance. The proposed project has a more practical and end-to-end approach to constructing such support with the aid of data mining and ML. Our initial point of view is an open source dataset at Kaggle, and the unprocessed records are our starting point - not the solution. Preprocessing must be done carefully: to clean inconsistent records, missing values, to put numbers on a common scale, to encode categories, to prepare text fields when necessary. The goal is a reliable feature space which represents signal but not noise.

Having the dataset ready we develop a clear evaluation loop by dividing data into training, testing and validation sets. The training split drives the learning process through a variety of algorithms selected to have complementary strong suit - probabilistic baselines, linear models, tree-based learners, and (where convenient) neural architectures of sequence patterns or high-dimensional patterns. The test split provides an objective interpretation of the way those learned patterns generalize. The validation split is used to validate the stability, tune hyperparameters and prevent overfitting.

Accuracy can never be used as a criterion in performance. Precision, recall, F1-score, confusion matrices, and calibration allow us not only to know the rate of correct predictions but also when failures occur and how much the confidence matches the reality. In instances where feasible, we discuss model explainability - feature importance, partial dependence or local explanations - to shed light on the factors behind the predictions and to aid in responsible interpretation.

The project is neither a clinical diagnostic instrument nor a replacement of professional assessment. The reason is that it is demonstrative: by using disciplined preprocessing, transparent experimentation, and balanced metrics, it is possible to create a reproducible pipeline screening patters that are correlated with depressive states. In addition to findings, we touch on ethical and practical issues - data quality, bias, privacy and right use - and therefore any revelations are intended with caution. At the end we present a comparative perspective of model performance and a workflow of inference which is ready to run that can take new inputs and provide a reasoned prediction. The pipeline is meant to be both auditable and extensible and this is the way other people can build on it when better data and better methods are available.

1.2 Motivation

- Increasing the burden of depression in the world and the necessity of screening assistance on a large scale.
- Late help-seeking as a result of stigmatization; data-driven instruments can promote prompt addressing.
- There is an ample amount of digital and survey data that is not used to provide mental health insight on a timely basis.
- The comparative modeling explains how effective the approaches are on the characteristics in the real world.
- Clearly defined metrics create responsible deployment and minimize guesswork in assessment.
- An ethically and safely iterable reproducible pipeline can assist researchers and practitioners.

1.3 Objective

- Crawl a Kaggle dataset and apply strict preprocessing to clean and consistent inputs.
- Engineer characteristics which retain signal with reduced bias and information leakage.
- Optimize the hyperparameters and stabilize the generalizable performance of various models.

- Assess using train -test-validation splits and non-accuracy measures.
- Make performance comparisons to trade off interpretability and cost of computation.
- Provide a reproducible, auditable and extensible prediction workflow.

1.4 Project Outcome

- An annotated, end-to-end preprocessing pipeline which can be reused on other similar datasets.
- Compare the performance of algorithms, including accuracy and complementary error analysis.
- Best model chosen due to a reason and evidence supported by trade-off and validation.
- Prediction made on new inputs using deployed inference script/notebook.
- Visual displays that describe the effect of features and point out limitations of models.
- Real-life recommendations on the future data collection, fairness, and ethical use.

1.5 Organization of Report

The report dates are divided into six detailed chapters which discuss an important area of the study.

Chapter 1 - Introduction is the background information that describes the objectives, scope, and problem statement of the study. It further demonstrates the importance of the research, the key research questions, the outcomes that have been expected, and the general outline of the report.

Chapter 2 - Background is the starting point that represents the basic terminologies and concepts of the terms and literature review necessary to get the feel of the nature and the scope of the problem. A gap analysis has also been included in this chapter to establish the gaps in the past research.

Chapter 3 - Research Methodology explains the proposed research method in terms of data collection procedures, data description, preprocessing strategies, and statistical analysis. It also explains the DL models applied in the study.

Chapter 4 - Experimental Results and Discussion shows and discusses the findings regarding the experiments and provides a critical assessment of the performances of the models.

Chapter 5 - Impact on Society, Environment and Sustainability examines the wider impacts of the research such as the societal gains, environmental concerns, the ethical concerns, and the sustainability strategies in the long run.

Chapter 6 - Conclusion and Future Scope provides a conclusion of the findings of the study and possible directions of the future research. At the end of the report, there is a Reference list where all sources mentioned in the course of writing this report are mentioned.

CHAPTER 2

BACKGROUND

2.1 Terminology

Depression is a disease that affects a lot of lives but is usually undetected until it interferes with the health, education or employment. This project investigates the ability of structured data and contemporary analytics to help in the initial screening based on learning patterns associated with depression states. Our initial data source is published records obtained through Kaggle, and a formal pipeline is created to model those records. The pipeline standardizes formats, fixes missing or inconsistent records, and encodes text or categorical variables to allow their use by algorithms.

2.2 Related Work

Walsh et al. based on massive medical data, the group trained artificial intelligence models to identify a risk of suicide attempts. The maximum AUC of the system was 0.84 and the better the prediction window, the higher the accuracy. Their discussion brings out the time-dependent characteristics of risk changes and the most critical predictors at various points of time [1].

The model used by Ware et al. to determine the cognitive and behavioral indicators of depression had a score of 0.86 by combining Wi-Fi and smartphone signals of 182 students. It did not need active user engagement and was better than previous studies that concentrated on the severity in general [2].

Data on smartphone and fitness-trackers by Chikersal et al. was utilized to identify depressed students. They forecasted symptoms 85.7 percent correctly and identified severity changes 85.4 percent correctly - to 15 weeks prior to the end of the semester - to indicate the relevance of previous, targeted interventions [3].

Five ML models were compared by Siddique et al. in one of the studies of Bangladeshi students at the university; the most successful model showed a high level of accuracy. Another outcome of the project was the creation of a mental-health application, which revealed the importance of prediction appliances and simplified referrals to mental health care in necessity [4].

ML emerged as a risk factor of major depression and anxiety in data provided by 3,984 schoolchildren by Qasrawi et al. SVM (92.5) and the Random Forest (76.4) have the highest scores, which draw attention to bullying, family income, and school violence as specific areas of support that can be provided in schools [5].

Smartphone features fed eight models were used in Siraji et al. on the international university students in Bangladesh, after feature selection. CatBoost produced the best accuracy revealing that a multi-phase, stage-based detection can enhance early mental-health screening [6].

Shafiee et al. a review of mental-health challenges in higher education identified financial strain, lack of social support and academic pressure as significant contributors. In supervised models, SVM ranked most frequently, and accuracy levels were between 70-96 percent [7].

The pattern of Ahmed et al. app-usage among 100 Bangladeshi students allowed the quick detection of depression. LightGBM had the highest accuracy of 82.4 percent and among many others, it indicates that low-friction, app-based signals can facilitate the early identification [8].

The smartphone traces provided by Asare et al. out of 629 participants within 22.1 average days produced 22 behavioral markers. ML models with as much as 98.14% accuracy; internet activity and screen time were considered as good predictors of depression risk [9].

The study, conducted by Hong et al. on ecological momentary assessments, was a combination of emotion-regulation factors and demographic factors. Another model that worked better than other models was a Random Forest using 13 key variables and the variables that significantly increased depression prediction were age, anxiety and social-emotional regulation [10].

Xu et al. derived an interpretable model of depression that relies on ecological assessment through the application of Random Forest and 13 significant characteristics. It performed better than conventional baselines by 9.7%, and a secondary dataset proved that it was robust - indicating viable application in screening environments [11].

The risk of depression was connected to the reading habits of university students (Hou et al.). A polynomial NB model was tested after having tested Bayesian, SVM and kNN reaching 82.3% accuracy. LR was also more effective than linear counterparts at psychological prediction in varied conditions [12].

The ecological measures used by Arringoni et al. on young adults were useful in predicting the symptoms of depression. Random Forest pointed out the age, anxiety, and emotional regulation as the key predictors, which implies the possibility of earlier assistance using behavioral analytics [13].

Y. Zhou et al. The 15 ML models were compared to screen depression during the time of quarantine. AdaBoost took the first place with its 0.7917 accuracy (ACC) and an AUC of over 0.83, implying that automated tools have the potential to support mental-health surveillance during time of crisis [14].

The language used in p. jain et al. Language from r/depression and r/SuicideWatch was analyzed to indicate at-risk users. SVM and NB performed well in terms of risk tracking, whereas Random Forest, LR, as well as NB performed well in terms of classification in terms of suicide-related discourse [15].

Sharma et al. Vocal features Vocal features were modeled bi-directionally to an extent of depression with a trained ML pipeline that was trained on a large amount of speech. The model was effective in identifying the symptoms, which demonstrates that voice analysis can be used to estimate the level of depression to be present without the use of the traditional questionnaires [16].

Manikandaprabhu et al. Student texts were modelled to be stressed through word choice and tone. SVM obtained 81.79 percent and LSTM got 70 percent. Simultaneously, NLP on Internet forums was able to detect depression with 80% of accuracy which supports language as an effective signal [17].

Ahmed et al. Out of a total user reviews of 205,581 depression-support chatbot users, text mining (based on Python) has emerged with themes of trust and consultation. Positive and negative feedback - in equal measure - indicates potential in AI helpers and definite opportunity for usability and safety enhancement [18].

The authors A. Musleh et al. Adding a neutral sentiment classification (better) to depression-related classification of Arabic tweets. Random Forest was the best RF, SVM, NB, and AdaBoost with the highest level of 82.39% accuracy to support social-media-based detection [19].

Wang et al. 4,262 people were modeled using KNN, SVM, Random Forest and AdaBoost. SVM was the most accurate model that exhibited consistency in separating mental-health conditions to be used in specific interventions [20].

According to Richter et al. Behavioral tests, the emotional tendencies of undiagnosed people were classified with 71.44% accuracy. Findings claim the use of individualized diagnostic instruments and earlier symptom identification in order to enhance treatment directions [21].

Bauer et al. A speech model that was trained on 550 clinical interviews was able to detect major depressive disorder with the accuracy of 66%. Combined predictions with PHQ-9 and QIDS-C increased accuracy to 73, which indicates that algorithmic tools may be used to supplement existing clinical scales [22].

Kumar et al. SVM, Rand Forest and LR were compared using 80:20 split. Random Forest won with most accuracy of 88 percent so that social-media interaction can succeed in detecting depression-related signals [23].

2.3 Compare between existing work

Table 2.1: Comparative analysis with previous work

SL No	Author Name	Used Algorithm	Best Accuracy with Algorithm
1.	C. G. Walsh, et al.[1]	ML	84%
2.	S. Ware et al [2	ML	86%
3.	P. Chikersal et al [3]	ML	85.7%
4.	R. Qasrawi,et al.,[5]	ML	SVM: 92.5%
5.	Muntequa Imtiaz Siraji et al [6]	ML	44%
6.	Shafiee, Nor Safika Mohd,et al.. [7]	ML	SVM=96%
7.	Y. Hou, J. Xu, et al. [12]	ML	NB= 82.3%

2.4 Scope of the Problem

Depression is pervasive, unreported and often hidden by daily practices. In this project, we discuss scalable screening at an earlier stage where the raw Kaggle data is transformed into structured signals that can be used by the ML. Its scope extends all the way through the pipeline: data source; stringent preprocessing of numeric, categorical, and - where applicable - text data; and responsible training, testing, and validation splits. There are various models that are compared in order to have balance accuracy, stability, and interpretability. Assessment is not only about headline accuracy, but also about patterns of errors and their generalization. The desired result is a workflow that is repeatable and can be fed new data and give rationale predictions - useful in research, aiding in triage, or planning resources. It is not a diagnostic alternative; it is simply an illustration of how responsible data mining can bring patterns associated with depressive conditions to light and lead to additional evaluation.

2.5 Challenges

The derivation of an accurate and dependable prediction model of depression among university students in Bangladesh has a number of challenges:

- Data quality & heterogeneity: Preprocessing and fair learning are complicated by missing values, noise, class imbalance, mixed features.
- Generalization risk: in models that are optimized to a particular cohort or source, they can not be applied to novel populations or new conditions.
- Leakage of features and bias: Predictions are inflated by improper encoding or time leakage; biases in predictions may be due to demographic or behavioral biases.
- Selection of metrics: Accuracy is not misleading, so complementary metrics need to be chosen and reported.
- Interpretability Owing to safety: Responsible adoption requires explanations of predictions, descriptions of assumptions and establishing safe limits of use.

2.6 Gap Analysis

Although a lot of research has been conducted in the field of mental health prediction, there are still a lot of gaps in the detection of depression among the Bangladeshi university students with data mining methods.

- Data representativeness: Narrow cohorts are of frequent use in many studies; in general, culturally diverse samples are still scarce.
- Longitudinal robustness: The number of pipelines that test time or drift stability is fewer; long-term validation is usually not present.
- Real world integration: Benchmarks are almost never concerned with the implementation requirement such as latency, privacy, audit trails, and maintainability.
- Explainability depth: Importance of features is reported, but actionable, end user explanations are not consistently built.

- Ethical protections: The structured guidelines of consent, de-identification, and escalation routes are disproportionate and present research-practice adoption gaps.

CHAPTER 3

METHODOLOGY

3.1 Proposed Method and Components

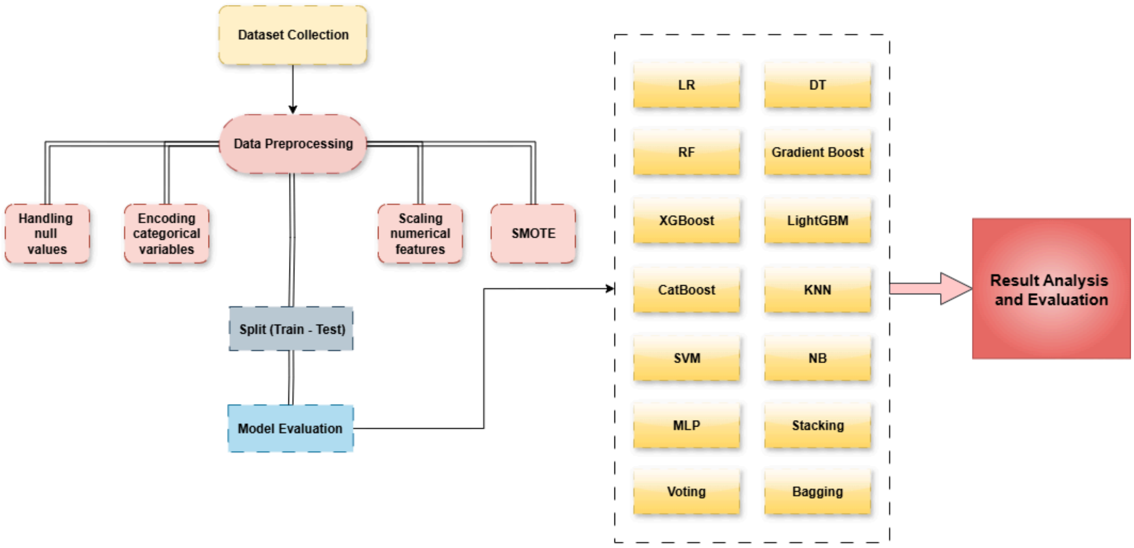


Figure 3.1: Architecture Diagram

The process of work takes place in definite, verifiable steps. We initially collect answers and compile a serviceable dataset. Then there is preprocessing - clean up sloppy entries, close gaps, and convert fields to readable models. The ML models are then trained on this prepared data. Another test stage is used to test the trained model, which gives a binary response: depressive or not-depressive to each student. Predicted Depressive is termed in case of a yes prediction; Predicted Non-Depressive in case of a no prediction. Lastly, findings are given in a structured output. The end to end structure ensures the process is efficient, reproducible and appropriate to student based screening.

3.2 Data Collection

The data was obtained on Kaggle, with 1430 entries that had diverse backgrounds. Its features cut across behavior, lifestyle, emotional cues, social activity and even psychological cues - sufficient breadth to examine depression in populations. To improve the quality of the data, we used the imputation method to fill missing values, coded categorical variables, and eliminated duplicates and inconsistencies. The data was refined, and then divided into training, testing, and validation subsets. The model construction was driven by training, the test sets ensured accuracy and generalization, and validation. This preparation has generated an equally balanced and analysis-ready resource to be forecasted with reliability.

Table 3.1: Analyse the target value

Value counts for Depressed		Percentage of instances in each Depressed	
Not Depressed	Depressed	Not Depressed	Depressed
1172	257	82.015395	17.984605

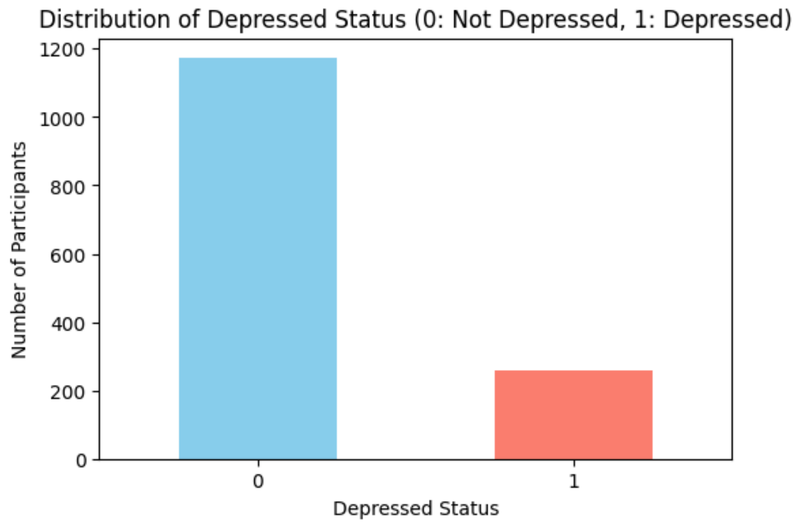


Figure 3.2: Overall Depressed & Not Depressed status

3.3 Dataset Description

A questionnaire survey of ten questions was employed, and the responses obtained were 1430. We also set the ratio of 80 and 20 percent on training and testing. The data set has 21 categorical variables, each of them is associated with a particular question in the survey and details. The columns are self-contained - no item is dependent on another one - the numeric codes are the response options or scores of statistical labor. This clean design retrieves depression trends within the sample of Bangladeshi university students and can be used to predict and forecast their needs and inform early intervention.

Table 3.2: Dataset Description

SN	Attribute	Full Form
1	sex	Gender of the individual
2	Age	Age of the person
3	Married	Marital status
4	Number_children	Number of children
5	education_level	Education level
6	total_members	Total family members
7	gained_asset	Value of gained assets
8	durable_asset	Value of durable assets
9	save_asset	Amount of savings or financial assets
10	living_expenses	Monthly living expenses
11	other_expenses	Other additional expenses
12	incoming_salary	Income from salary or job
13	incoming_own_farm	Income from own farming activities
14	incoming_business	Income from business ownership
15	incoming_no_business	Income from non-business sources
16	incoming_agricultural	Income from agricultural activities
17	farm_expenses	Expenses related to farming
18	labor_primary	Cost of primary labor or workforce
19	lasting_investment	Value of long-term investments
20	no_lasting_investmen	Value of non-lasting or short-term investments
21	depressed	Depression status (Yes/No)

3.4 Statical Analysis

- We have 1,430 total dataset records - that is sufficient to both train and test our pipeline, as well as to stress-check it and be sure that everything is working as intended.
- Our sample is clean and we have 80 per cent training to acquire patterns and 20 per cent testing to generalize on unknown cases.
- It contains 22 columns in the table that presents a wide perspective of the inputs but can be modeled and reviewed easily.
- It is constructed on 10 survey questions, as they reflect fundamental signals related to mood, habits, and context.
- Format: 1 dependent target and 21 independent/input variables, in a simple supervised learning format.
- To model, we executed a stack of 11 models and compared their accuracy, stability and inference behaviour.

3.5 Data Preprocessing

3.5.1. Handling missing/null values

Absence or missing values could decrease the accuracy of a data set, thus, they need to be handled in advance. In this step, blank fields are recognized and filled with either average or median or most frequent values, or deleted. This allows consistency and prevents false model results.

3.5.2. Encoding categorical variables

Categorical features are features that have text or labels that cannot be read by the machine. Encoding transforms these labels into a numerical form to enable the process of models. Depending on the type of feature, different techniques, such as the label encoding or one-hot encoding, are employed. This makes the dataset understandable to the algorithms without distorting the meaning of the original categories.

3.5.3. Scaling numerical features

Numerical values can be on various scales that do not balance model learning. Scaling also normalizes these values within a similar range such that no values overpower other features. Such processes as standardization or normalization are used to make numbers closer. This enhances the accuracy, the stability of the models and their overall performance in the training.

3.5.4. Smote

SMOTE (Synthetic Minority Over-sampling Technique) represents a preprocessing technique which is applied to address the problem of class imbalance in a dataset. It functions by producing artificial samples to the minority group rather than merely copying those that already exist. It does this by picking random nearest neighbors and creating new data points in between them. Subsequently, SMOTE assists in enhancing the performance of a model since it gives balanced training data to achieve better classification performance.

Table 3.3: Statistical analysis on Smote function

Before Smote Function		After Smote Function	
Depressed	Not Depressed	Depressed	Not Depressed
238	1191	953	953

3.6. Proposed Model

3.6.1. Logistic Regression

LR is a straightforward classification technique, which forecasts results on the probability basis. It does not fit a straight line but rather fits a sigmoid curve to map values between 0 and 1. It is best when the associations among features and classes are more or less linear in nature, and can be used to make binary predictions or multi-class predictions.

3.6.2. Decision Tree

A DT divides data into branches according to rules that are most effective to separate classes. The splits are the questions, and the leaves provide a final decision. It is simple to comprehend and illustrate and is therefore handy in description of model behavior. Nonetheless, it may overfit unless it is controlled.

3.6.3. Random Forest

Random Forest constructs numerous DTs and to enhance the precision of prediction, the results are synthesized. All trees are trained with random samples and features, which minimize overfitting and enhance stability. Through averaging, it develops a better model. It is very efficient both in classification and regression with various data sets.

3.6.4. Gradient Boosting

Gradient Boosting develops a model of weak models, with each model rectifying the errors of the earlier one. It aims at reduction of errors in a graduate manner through gradient-based optimization. This slow enhancement creates a powerful last model. It works well over structured data but can undergo overfitting and take excessive time to train.

3.6.5. XGBoost

XGBoost is a faster and more precise gradient boosting. It regulates to minimize overfitting, auto-processes missing values, and works efficiently with data. Its success on structured data sets renders it popular in competitions and in real world applications with concerns in precision and computational efficiency.

3.6.6. LightGBM

LightGBM LightGBM is a boosting algorithm where trees are grown leaf-wisely rather than level-wisely and can be trained faster and more accurately. It can deal with high levels of dimension and relatively large volumes of data. Its graph-based method is very specific to histograms, which makes it an appropriate option in real-time or scale-intensive ML tasks.

3.6.7. CatBoost

CatBoost is a gradient boosting algorithm that is made to operate with categorical features in the natural way. It also encodes automatically so that it does not require heavy preprocessing. Its algorithm avoids over fitting in the case of ordered boosting methods and works in noisy data. It is extensively applied in classification problems with mixed-type characteristics.

3.6.8. K-Nearest Neighbors (KNN)

KNN makes predictions based on the most similar examples in the data set. It makes the assumptions that the similar points have similar labels and thus, decisions are made based on the majority of the immediate neighbors. It is an easy to understand and intuitive algorithm, but one which may be slow with large data sets and is also susceptible to variations in feature scale and distance metrics.

3.6.9. Support Vector Machine (SVM)

SVM identifies the optimum boundary that separates the classes maximizing the margin between them. It is able to manipulate the linear and non-linear data using various kernel functions. SVM has high dimensional capability and can also be effectively tuned, however, it can be computationally costly with large datasets.

3.6.10. Naïve Bayes

NB is a Bayes-theorem classifier that is probabilistic. It is simple, although surprisingly effective, and it assumes that features are independent. The algorithm is suitable in text classification and spam detection, among other things, where probabilities determine the course of action. It is a fast lightweight code and works with large datasets.

3.6.11. MLP (Neural Network)

Multi-layer Perceptron is a multi-layered neural network consisting of linked nodes. It is a learning algorithm that learns patterns by adjusting weights by means of backpropagation. MLP represents a complex relationship among features hence suitable in classification and regression. It can also require special fine tuning and more computing power than simpler models despite being powerful.

3.6.12. Stacking Classifier

A Stacking Classifier is an ensemble of several different models and learns to integrate the capabilities of all of them using a final model (meta-model). Individual predictions of each base model are inputs to the meta-model. This multi-layered strategy has been shown to be very accurate, as it is able to record the different patterns that an individual algorithm may fail to capture.

3.6.13.Voting Classifier

A Voting Classifier combines multiple models together and finally makes a final decision based on their decision. In simple voting, all the models vote on a single class and the one having the majority wins. This will aid in eliminating individual model bias and produces a more balanced and reliable prediction by fusing several viewpoints.

3.6.14. Bagging Classifier

Bagging Classifier is used to create numerous versions of the same model with random samples of the data. Each model is independent and the prediction made by each of them will be pooled together to create the final output. This method makes overfitting less likely, makes performance more stable, and is particularly effective with those algorithms that are prone to changing with input data.

CHAPTER 4

RESULTS AND DISCUSSION

4.1 Introduction

The issue of depression has been a on the rise of the globe that is caused by personal, social and financial factors. By analyzing these trends based on data-driven approaches, it will be possible to identify the at-risk individuals. This paper involved a structured data comprising of demographic information, family statuses, source of income, spending, and asset holdings, and investment behaviors as a way of examining the propensity towards depression. Thirteenteen ML models were used to determine the relationship between various attributes and depressive states after the necessary preprocessing (or missing values, category encoding and numeric scale) was performed. In order to deal with the problem of class imbalance, the data was evaluated in two versions the original data and the balanced version produced through the SMOTE technique. Through the comparison of the two datasets, the research is supposed to identify the algorithms that are consistently good and offer good predictions in the study of depression patterns.

4.2 Comparative Analysis

The comparison was made in a comprehensive manner between fourteen ML models on the original data and the SMOTE-balanced data. Five metrics (Accuracy, Precision, Recall, F1-score, and AUC-ROC) were taken into account during the evaluation in order to know the model reliability in different ways. Random Forest, XGBoost, LightGBM, Bagging and Voting are the most successful ensemble-based models in the original dataset, reaching the accuracy of approximately 0.9755 with high precision and balanced F1-scores. The natural imbalance of the data was better managed in these models and the predictions were stable without large decreases in sensitivity.

Conversely, the SMOTE dataset provided a balanced distribution of depressed and non-depressed cases, and the models can now identify the minority patterns. Models such as LightGBM, random forest and XGBoost demonstrated strong performances once again with the majority reaching an accuracy of above 0.97 and with high recall scores. Interestingly, algorithms like KNN and SVM showed much better recall in the presence of SMOTE and this is an indication of the sensitivity of these algorithms to balanced inputs. Meanwhile, simpler models such as LR and NB recorded moderate improvements but still continued to be not as good as ensemble methods.

Generally, the analysis indicates that ensemble algorithms are always superior to other methods in the two datasets and thus they can effectively deal with complex interactions among features and balanced/unbalanced scenarios.

4.3 Result and Discussion

4.3.1. Logistic Regression

Table 4.1: Model Accuracy of LR

Data	Accuracy	Precision	Recall	F1-score	AUC-ROC
Original	0.8636	0.9615	0.3968	0.5618	0.6962
SMOTE	0.8112	0.5529	0.7460	0.6351	0.7878

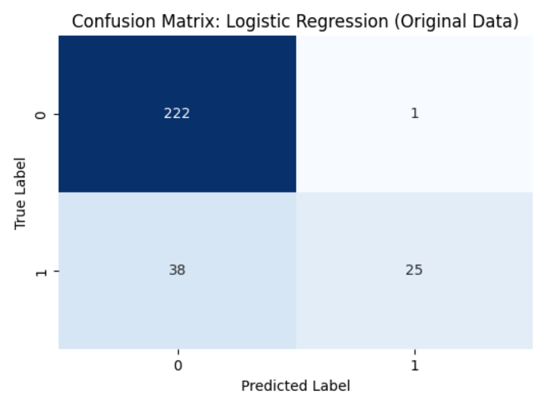
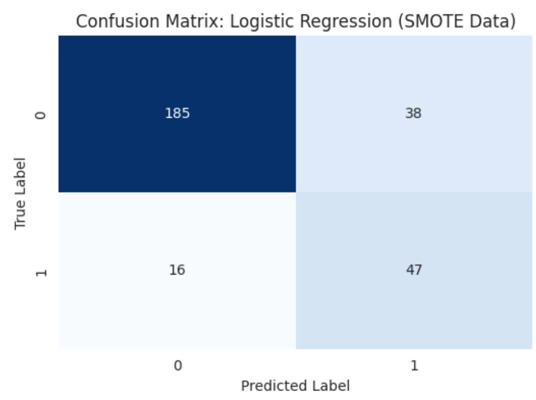
Original Data of LR	SMOTE Data of LR
	

Figure 4.1: Confusion Matrix of LR

4.3.2. Decision Tree

Table 4.2: Model Accuracy of DT

Data	Accuracy	Precision	Recall	F1-score	AUC-ROC
Original	0.9685	0.9655	0.8889	0.9256	0.9400
SMOTE	0.9615	0.9194	0.9048	0.9120	0.9412

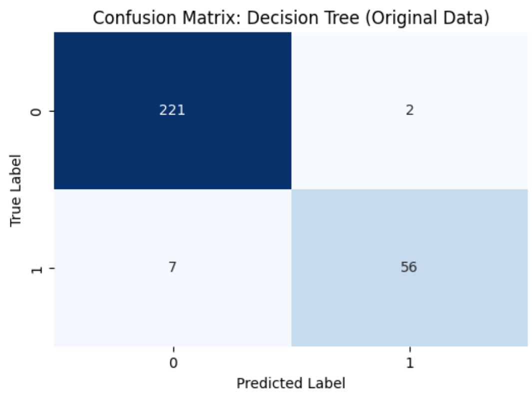
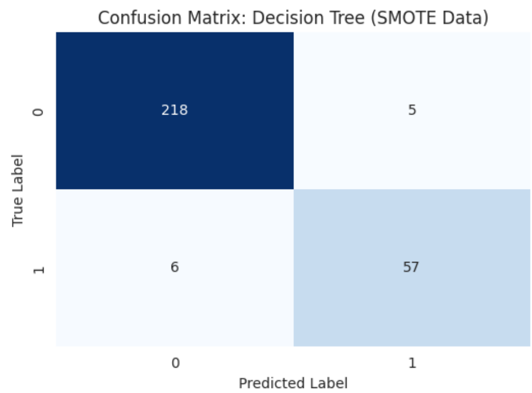
Original Data of DT	SMOTE Data of DT
	

Figure 4.2: Confusion Matrix of DT

4.3.3. Random Forest

Table 4.3: Model Accuracy of Random Forest

Data	Accuracy	Precision	Recall	F1-score	AUC-ROC
Original	0.9755	1.0000	0.8889	0.9412	0.9444
SMOTE	0.9720	1.9825	0.8889	0.9333	0.9422

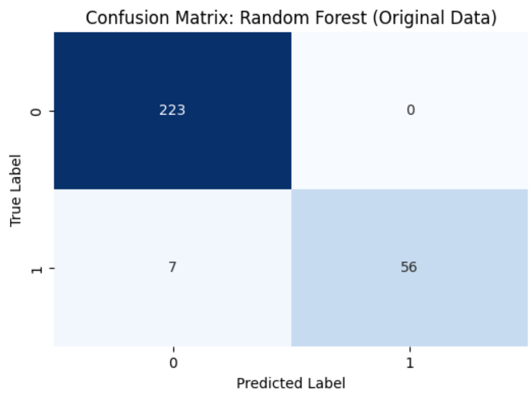
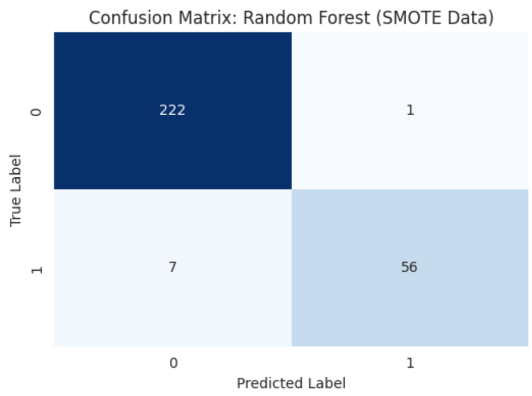
Original Data of Random Forest	SMOTE Data of Random Forest
	

Figure 4.3: Confusion Matrix of Random Forest

4.3.4. Gradient Boosting

Table 4.4: Model Accuracy of Gradient Boosting

Data	Accuracy	Precision	Recall	F1-score	AUC-ROC
Original	0.9126	1.0000	0.6032	0.7525	0.8016
SMOTE	0.8881	0.7541	0.7302	0.7419	0.8314

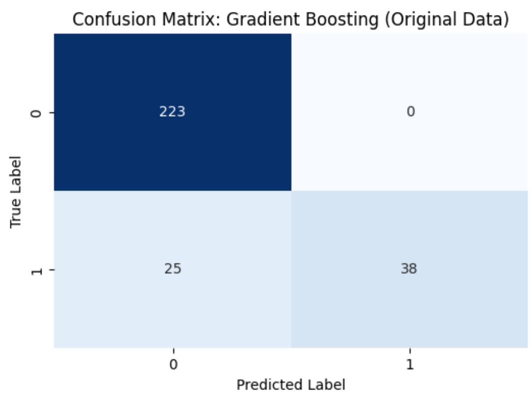
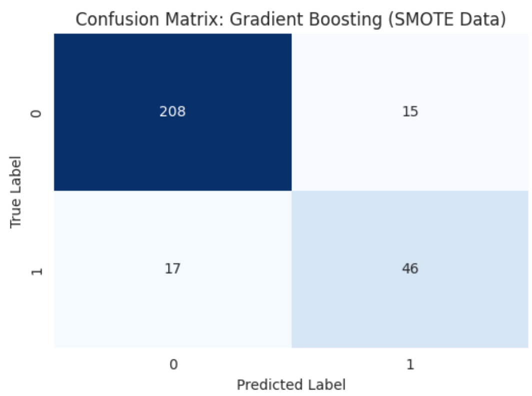
Original Data of Gradient Boosting	SMOTE Data of Gradient Boosting
	

Figure 4.4: Confusion Matrix of Gradient Boosting

4.3.5. XGBoost

Table 4.5: Model Accuracy of XGBoost

Data	Accuracy	Precision	Recall	F1-score	AUC-ROC
Original	0.9755	1.0000	0.8889	0.9412	0.9444
SMOTE	0.9685	0.9655	0.8889	0.9256	0.9400

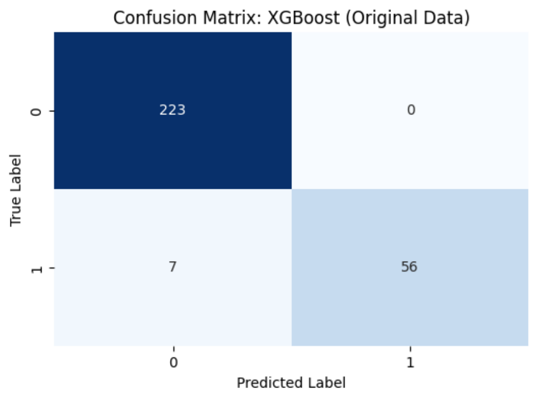
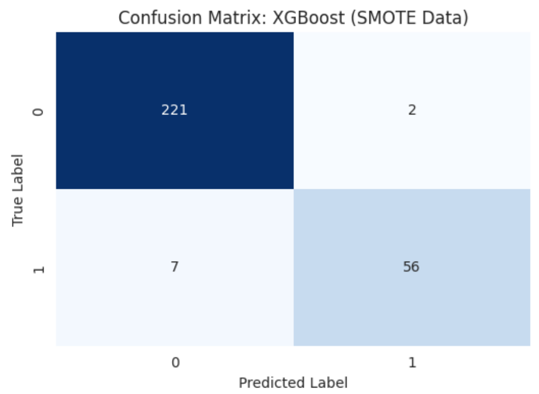
Original Data of XGBoost	SMOTE Data of XGBoost
	

Figure 4.5: Confusion Matrix of XGBoost

4.3.6. LightGBM

Table 4.6: Model Accuracy of LightGBM

Data	Accuracy	Precision	Recall	F1-score	AUC-ROC
Original	0.9755	1.0000	0.8889	0.9412	0.9444
SMOTE	0.9720	0.9825	0.8889	0.9333	0.9422

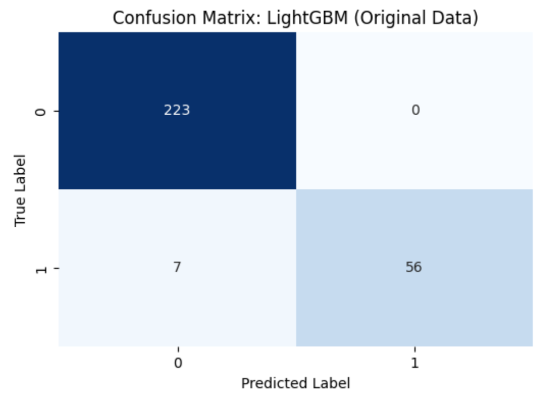
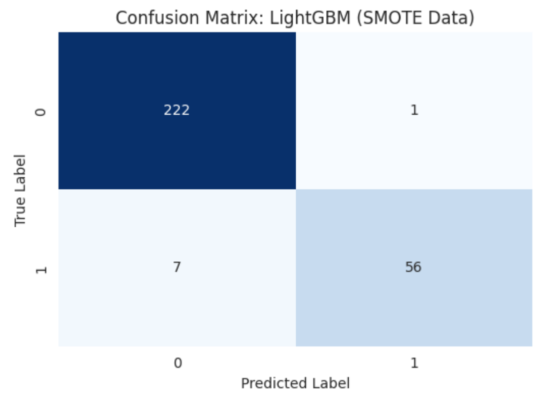
Original Data of LightGBM	SMOTE Data of LightGBM
	

Figure 4.6: Confusion Matrix of LightGBM

4.3.7. CatBoost

Table 4.7: Model Accuracy of CatBoost

Data	Accuracy	Precision	Recall	F1-score	AUC-ROC
Original	0.9441	1.0000	0.7460	0.8545	0.8730
SMOTE	0.9545	0.9032	0.8889	0.8960	0.9310

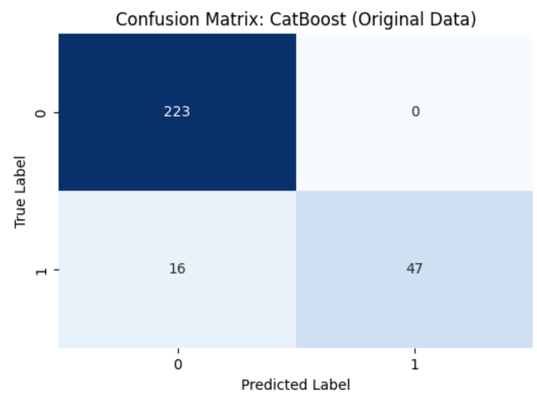
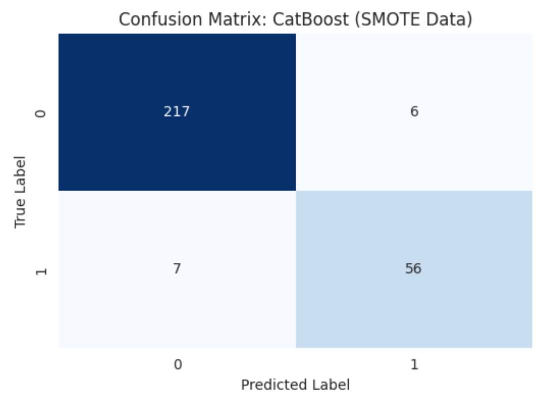
Original Data of CatBoost	SMOTE Data of CatBoost
	

Figure 4.7: Confusion Matrix of CatBoost

4.3.8. K-Nearest Neighbors (KNN)

Table 4.8: Model Accuracy of KNN

Data	Accuracy	Precision	Recall	F1-score	AUC-ROC
Original	0.9056	0.8000	0.7619	0.7805	0.8540
SMOTE	0.9126	0.7500	0.9048	0.8201	0.9098

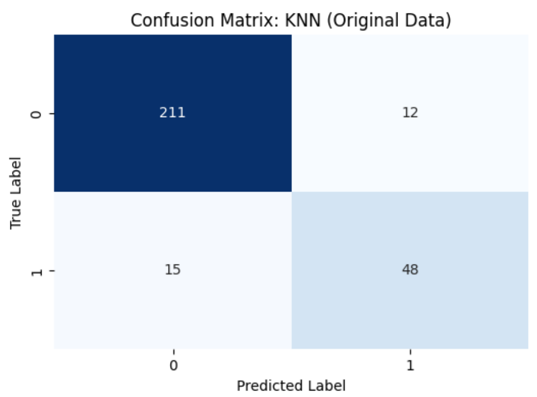
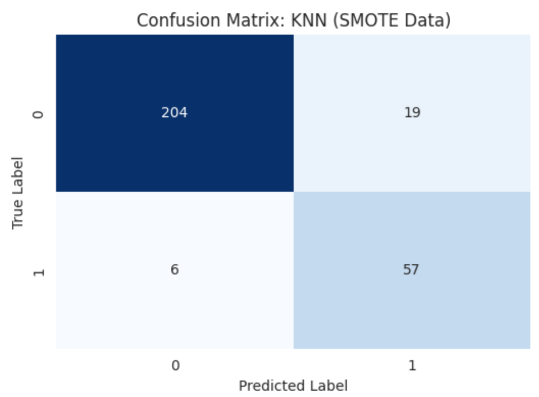
Original Data of KNN	SMOTE Data of KNN
	

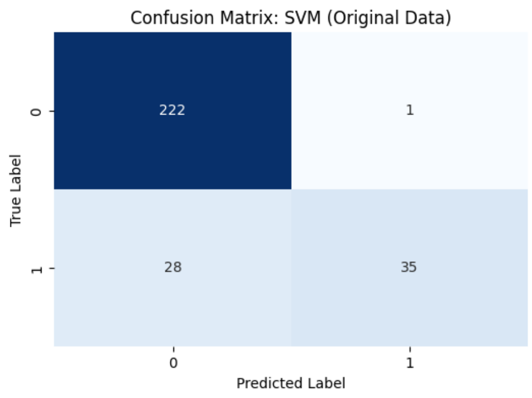
Figure 4.8: Confusion Matrix of KNN

4.3.9. Support Vector Machine (SVM)

Table 4.9: Model Accuracy of SVM

Data	Accuracy	Precision	Recall	F1-score	AUC-ROC
Original	0.8986	0.9722	0.5556	0.7071	0.7755
SMOTE	0.8601	0.6292	0.8889	0.7368	0.8705

Original Data of SVM



SMOTE Data of SVM

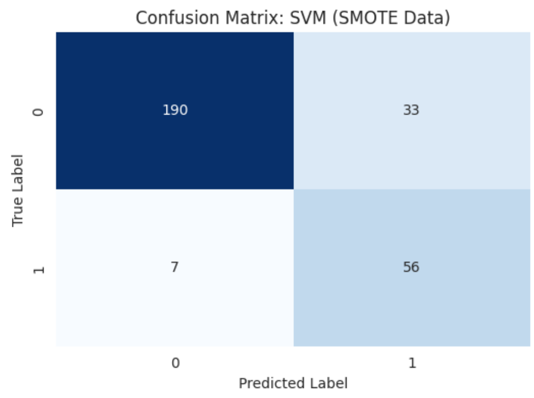


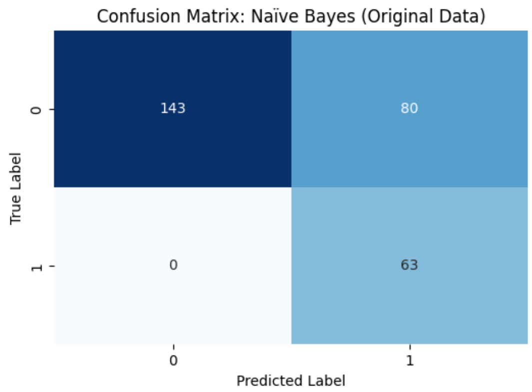
Figure 4.9: Confusion Matrix of SVM

4.3.10. Naïve Bayes

Table 4.10: Model Accuracy of NB

Data	Accuracy	Precision	Recall	F1-score	AUC-ROC
Original	0.7203	0.4406	1.0000	0.6117	0.8206
SMOTE	0.7203	0.4406	1.0000	0.6117	0.8206

Original Data of NB



SMOTE Data of NB

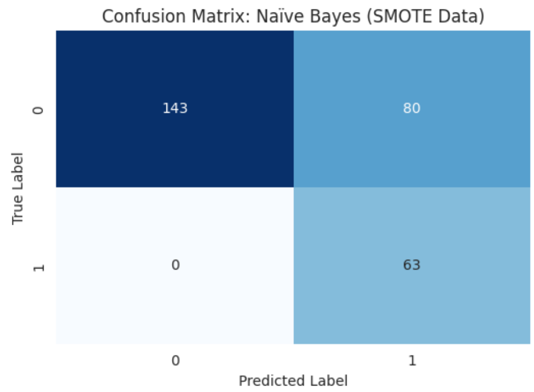


Figure 4.10: Confusion Matrix of NB

4.3.11. MLP (Neural Network)

Table 4.11: Model Accuracy of MLP

Data	Accuracy	Precision	Recall	F1-score	AUC-ROC
Original	0.9685	0.9655	0.8889	0.9256	0.9400
SMOTE	0.9580	0.9180	0.8889	0.9032	0.9332

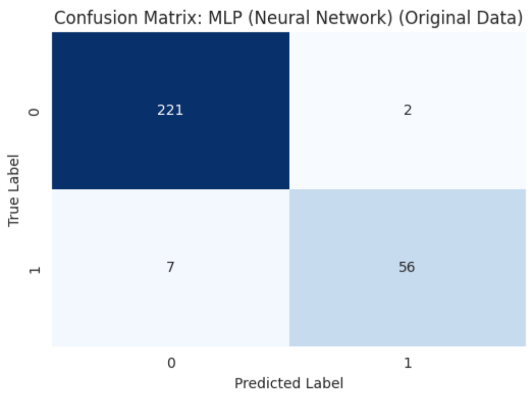
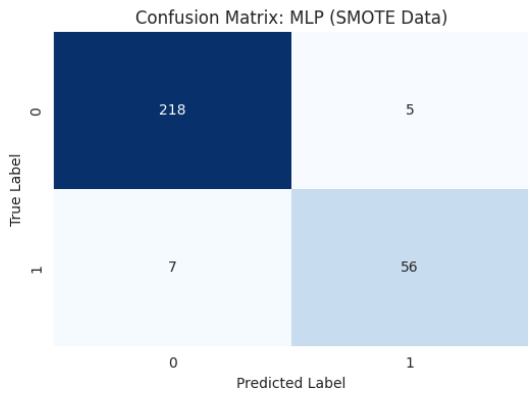
Original Data of MLP	SMOTE Data of MLP
	

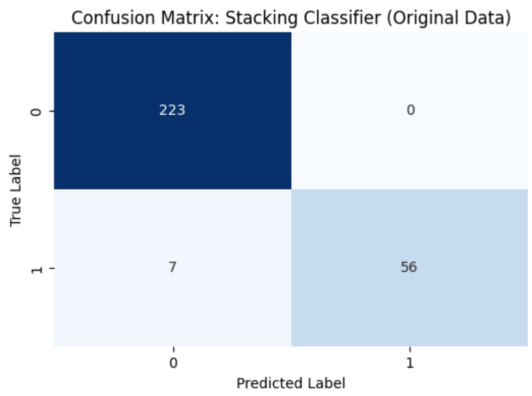
Figure 4.11: Confusion Matrix of MLP

4.3.12. Stacking Classifier

Table 4.12: Model Accuracy of Stacking Classifier

Data	Accuracy	Precision	Recall	F1-score	AUC-ROC
Original	0.9755	1.0000	0.8889	0.9412	0.9444
SMOTE	0.9650	0.9492	0.8889	0.9180	0.9377

Original Data of Stacking Classifier	SMOTE Data of Stacking Classifier



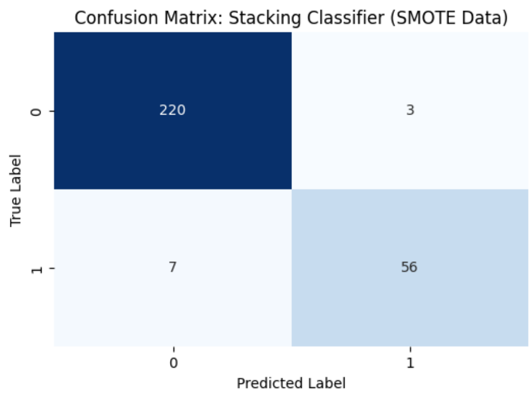


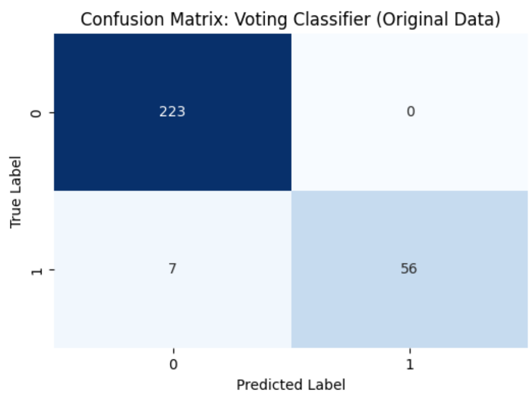
Figure 4.12: Confusion Matrix of Stacking Classifier

4.3.13. Voting Classifier

Table 4.13: Model Accuracy of Voting Classifier

Data	Accuracy	Precision	Recall	F1-score	AUC-ROC
Original	0.9755	1.0000	0.8889	0.9412	0.9444
SMOTE	0.9650	0.9492	0.8889	0.9180	0.9377

Original Data of Voting Classifier	SMOTE Data of Voting Classifier



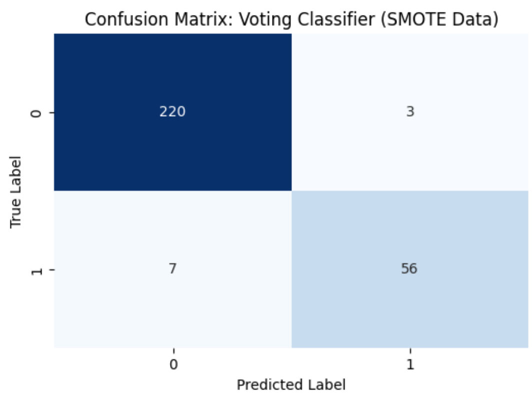


Figure 4.13: Confusion Matrix of Voting Classifier

4.3.14. Bagging Classifier

Table 4.14: Model Accuracy of Bagging Classifier

Data	Accuracy	Precision	Recall	F1-score	AUC-ROC
Original	0.9720	1.0000	0.8730	0.9322	0.9365
SMOTE	0.9650	0.9492	0.8889	0.9180	0.9377

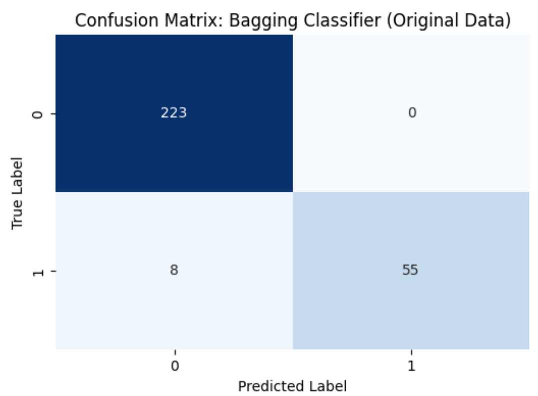
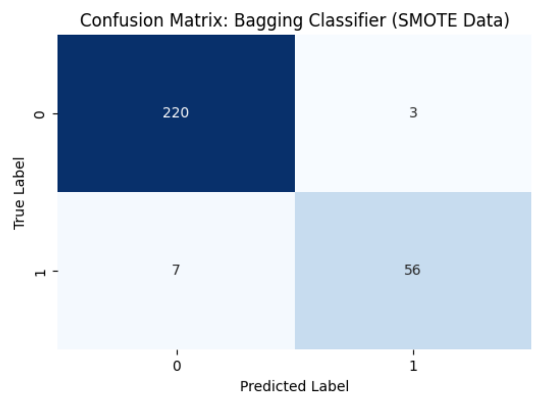
Original Data of Bagging Classifier	SMOTE Data of Bagging Classifier
	

Figure 4.14: Confusion Matrix of Bagging Classifier

4.4 Chapter Summary

This paper investigated the issue of predicting depression with fourteen ML models on original data and SMOTE-balanced data. Data has captured major socioeconomic and demographic aspects of human lives, and thus, it could be feasible to establish patterns that were linked to depressive tendencies. The models were tested with the help of Accuracy, Precision, Recall, F1-score, and AUC-ROC in order to assess the performance of the models in a global way. These Ensemble models - especially Random Forest, XGBoost, LightGBM, Bagging and Voting - were always better than the other models in both datasets and demonstrated the high ability to handle the complexity of features and detect depressive predictors. Approaches such as KNN and SVM also improved significantly with SMOTE as opposed to LR and NB that were relatively weak. The findings indicate that it is essential to manage the problem of the class imbalance and choose powerful algorithms to obtain trustworthy predictions. Altogether, the research proves that there is a distinct relationship between lifestyle patterns and the likelihood of being depressed.

4.4.1. Model Comparison Accuracy with Original Data:

Table 4.15: Model Comparison of Original Data

Model	Accuracy	Precision	Recall	F1-score	AUC-ROC
LR	0.8636	0.9615	0.3968	0.5618	0.6962
DT	0.9685	0.9655	0.8889	0.9256	0.9400
Random Forest	0.9755	1.0000	0.8889	0.9412	0.9444
Gradient Boosting	0.9126	1.0000	0.6032	0.7525	0.8016
XGBoost	0.9755	1.0000	0.8889	0.9412	0.9444
LightGBM	0.9755	1.0000	0.8889	0.9412	0.9444
CatBoost	0.9441	1.0000	0.7460	0.8545	0.8730
KNN	0.9056	0.8000	0.7619	0.7805	0.8540
SVM	0.8986	0.9722	0.5556	0.7071	0.7755
NB	0.7203	0.4406	1.0000	0.6117	0.8206
MLP	0.9685	0.9655	0.8889	0.9256	0.9400
Stacking	0.9755	1.0000	0.8889	0.9412	0.9444
Voting	0.9755	1.0000	0.8889	0.9412	0.9444
Bagging	0.9720	1.0000	0.8730	0.9322	0.9365

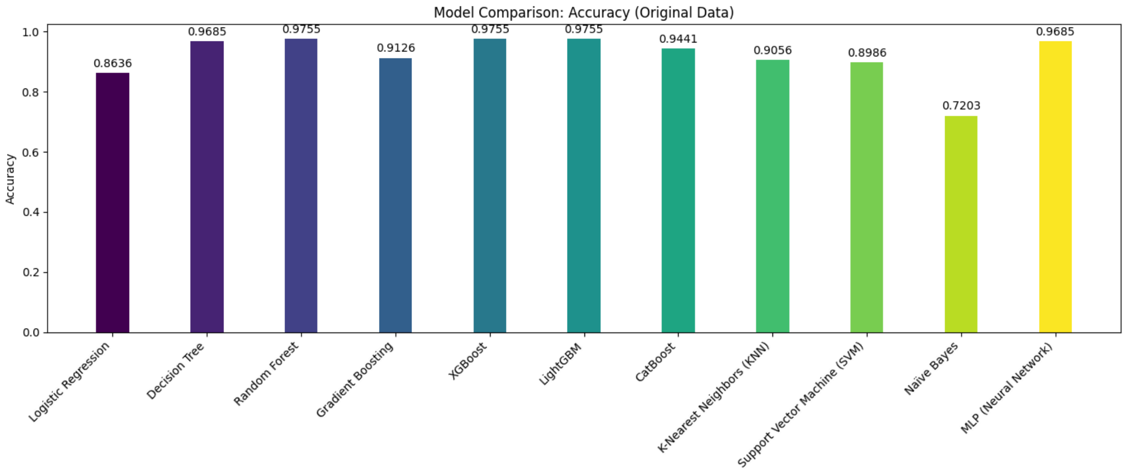


Figure 4.15: Model Comparison of Original Data

4.4.2. Model Comparison Accuracy with SMOTE Data:

Table 4.16: Model Comparison of SMOTE Data

Model	Accuracy	Precision	Recall	F1-score	AUC-ROC
LR	0.8112	0.5529	0.7460	0.6351	0.7878
DT	0.9615	0.9194	0.9048	0.9120	0.9412
Random Forest	0.9720	1.9825	0.8889	0.9333	0.9422
Gradient Boosting	0.8881	0.7541	0.7302	0.7419	0.8314
XGBoost	0.9685	0.9655	0.8889	0.9256	0.9400
LightGBM	0.9720	0.9825	0.8889	0.9333	0.9422
CatBoost	0.9545	0.9032	0.8889	0.8960	0.9310
KNN	0.9126	0.7500	0.9048	0.8201	0.9098
SVM	0.8601	0.6292	0.8889	0.7368	0.8705
NB	0.7203	0.4406	1.0000	0.6117	0.8206

MLP	0.9580	0.9180	0.8889	0.9032	0.9332
Stacking	0.9650	0.9492	0.8889	0.9180	0.9377
Voting	0.9650	0.9492	0.8889	0.9180	0.9377
Bagging	0.9650	0.9492	0.8889	0.9180	0.9377

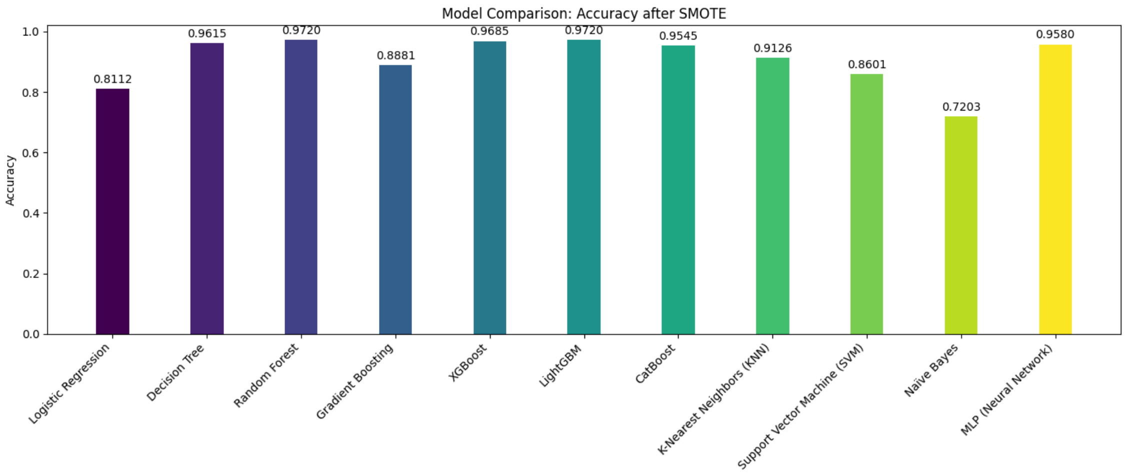


Figure 4.16: Model Comparison of SMOTE Data

CHAPTER 5

IMPACT ON SOCIETY, ENVIRONMENT AND SUSTAINABILITY

5.1. Compliance with the Standards

5.1.1. Software Standard

The software environment used in this project was founded on a set of effective and popular tools to offer uniformity, efficiency and manageable team work. Relevant supporting literature to scientific researches was obtained through research database of Research Gate and Google Scholar. The development and training of models was performed with Google Colab that allowed supporting the use of GPUs and collaborating in a team easily. This was implemented in Python alongside DL models such as TensorFlow and Keras to model, test and optimize. An extra data processing and local experimentations were done using a personal laptop to ensure a confirmation of findings. The documentation and the production of reports were made using Microsoft word so that they were written and presented in an academic and professional manner.

5.1.2. Hardware Standard

The system has minimum requirements to run the system as Windows 7/8/8.1 operating system, latest generation Intel i3 processors, 4 GB RAM, and 4 GB graphics card (DirectX version 11). Therefore, at least 3 GB of free storage is needed to process datasets and project files.

The Windows 10 operating system, Intel i5 processor, 8 GB of memory and a graphic card with 6 GB of memory or higher are recommended to use it at the optimal. At least 3GB of the free space is strongly advised to operate DirectX 12 without performance issues to maintain the stable training and testing and workflow performance.

5.1.3. Communication Standard

The most important thing in this project was communication and collaboration. Regular meetings on Zoom and Google Meet were planned to delegate specific tasks, discuss problems, and update about the progress. Google drive made it easy to access and control the version of project files, model outputs and documentation. Email and WhatsApp were used to do day to day coordination and quick updates to ensure that communication was quick and direct. All the formal documentation and reporting has been undertaken in Microsoft word so as to have a formatted and professional report. Feedback loops were created that would facilitate agile workflow making sure that the team was always on track, issues were labelled during the development path of the project and that joint decisions would be made in the end.

5.2. Impact on Society

- Previous assistance: Flag risk patterns earlier, which encourages students to have check-ins, counseling, or peer support before escalation of problems.
- Less stigma: Making screening a regular analytics is a normal practice and the discussion of mood and stress no longer sounds confrontational but more like practical.
- Better triage: Provides counselors with more focused resources on those with more pressing risk features so that more people can receive services.
- Academic continuity: Relates well-being to performance, allowing proactive accommodations that help students remain involved in course work.
- Equity lens: Brings to light the areas of risk that are concentrated among demographics or contexts, and informs targeted outreach that does not omit less vocal groups.
- Public-health insight: Trends aggregated in a de-identified manner are useful to campus programs and community efforts without providing information on individuals.
- Digital literacy: Educates students and personnel about the operation of data, models, and guardrails, developing healthier expectations of technology in care.

5.3. Impact on Environment

- Compute efficiency: Preferred lean features and right-sized models to restrain training energy and emissions.
- Carbon-conscious fuzzy scheduling: Scheduling of heavier activities when grids are cleaner, or when regions have greener energy compositions.
- Hardware lifetime: Chooses more modest architectures and mixed precision where it can, and lifecycle of devices.
- Data reuse: Promotes management and open datasets to limit replicating data collection and storage.
- Paperless reporting: Paperless digital dashboards will be used instead of printed summaries, decreasing the amount of materials used.
- Observing cost: Tracks calculate and store and experiment sprawl to retire unused artifacts and maintain footprints.

5.4. Ethical Aspects

Privacy and consent are the initial ones: present data usage, provide an option to discontinue, and deanonymize records. Limit scope - it is a tool of screening, not diagnosis and involve a human in any sensitive decisions. Reduce bias by using balanced splits, fairness checks and error analyses between subgroup. Strict access controls, audit trails and documented assumptions are used to prevent leakage and misuse. Provide straightforward descriptions of model behavior and constraints. Assign secure points of escalation to competent support in case of flags. Periodically inspect the pipeline to check the drift, damage, and unwanted effects.

5.5. Sustainability Plan

Make the pipeline stable with point-to-point data, reproducible code, and documented preprocessing. Apply MLOps fundamentals: automated validation, drift notifications, and scheduled light retraining on new, consent-based data. Favor small models to be inferred on conventional hardware. Issue transparent model cards, equity reports, and operating instructions. Create feedback loops with users and counselors to improve features and thresholds. Fund the persistent security and backups and limited cloud storage; delete obsolete experiments. Keep governance alive - not a one-time affair with institutional ethics boards. To the extent anything mentioned here is personal, it is possible to discuss it with a loved one or a local professional.

CHAPTER 6

CONCLUSION

6.1 Conclusion

This project was aimed at making predictions on depression; a dataset of demographic, family, economic and lifestyle-related variables was used. This was initiated by some basic preprocessing steps such as missing values, categorical variables encoding, and scaling of numerical variables to make the dataset appropriate in terms of training a model. Two datasets were created to evaluate the effects of the imbalance in depressive cases the original and a SMOTE-balanced dataset.

Seventeen ML models were used in total, that is, linear, tree-based, ensemble, distance-based, probabilistic, and neural network methods. They were assessed based on five important measures, including Accuracy, Precision, Recall, F1-score, and AUC-ROC. Model ensemble models like Random Forest, XGBoost, LightGBM, Bagging and Voting presented the best and most consistent results in both datasets with high accuracy and balanced results. The improvement in models such as KNN, SVM and MLP when trained on the SMOTE dataset was observable, as the balance in terms of class distributions was provided.

During the procedure, the project showed the correlation between different personal and financial aspects and the depression and the different algorithms that perceive these tendencies. The research presents a thorough comparison that brings out the importance of preprocessing, balancing methods and model selection in mental health prediction exercises.

In this paper, the researcher was able to portray the application of ML in comprehending the patterns of depression through real life personal, financial, and social characteristics. The study presented a profound understanding of how the distribution of classes, interactions between the features, and the nature of the model affect the performance of prediction by measuring the performance of fourteenth algorithms on original and SMOTE-balanced data sets. Ensemble techniques turned out to be the most helpful, as their accuracy and balanced assessment scores are always high. They have the advantage of pooling the efforts of more than one learner, and thus they have a chance to detect more complicated patterns than the simpler models might be able to. In the meantime, the SMOTE data set was useful in improving the recall of various models and indicates that when making predictions on sensitive conditions such as depression, then it is important to take care of the class imbalance. The paper gives an emphasis that depression is a phenomenon that depends on several facets of human life and that can be applied to uncover the trends objectively by means of ML. On the whole, the project preconditions the creation of more sophisticated, efficient, and timely systems directed at the initial identification of depression and mental health assistance..

6.2 Future Work

- Expansion of Features: Add more behavioral, medical, and psychological characteristics to enhance greater understanding of real world indicators of depression by the model.
- DL Models: Use LSTM, CNN or hybrid neural networks to recognize more complex patterns.
- Real-Time Prediction System: Build a web or mobile app giving users a chance to get an estimate of the risk of depression in real time.
- Explainable AI (XAI): Provide a model with SHAP, LIME, or other approaches to explain its decision-making to make it clinically interpretable and enhance trust from its users.
- Bigger and Heterogeneous Dataset: Gather a larger amount of data using broader populations in order to generalize and decrease bias in models.

REFERENCES

[1] C. G. Walsh *et al.*, “Predicting Risk of Suicide Attempts Over Time Through Machine Learning,” *Clinical Psychological Science*, vol. 5, no. 3, pp. 457–469, Apr. 2017.

[2] S. Ware *et al.*, “Predicting depressive symptoms using smartphone data,” *Smart Health*, vol. 15, p. 100093, Mar. 2020.

[3] P. Chikersal *et al.*, “Detecting Depression and Predicting its Onset Using Longitudinal Symptoms Captured by Passive Sensing,” *ACM Trans. Comput.-Hum. Interact.*, vol. 28, no. 1, pp. 1–41, Feb. 2021.

[4] A. B. Siddique *et al.*, “A ML-based approach to predict university students’ depression pattern and mental healthcare assistance scheme using Android application,” *Int. J. Data Anal. Tech. Strateg.*, vol. 14, no. 2, p. 122, 2022.

[5] R. Qasrawi *et al.*, “Schoolchildren’ Depression and Anxiety Risk Factors Assessment and Prediction: ML Techniques Performance Analysis (Preprint),” *JMIR Formative Research*, Aug. 2021.

[6] M. I. Siraji *et al.*, “Impact of mobile connectivity on students’ wellbeing: Detecting learners’ depression using ML algorithms,” *PLOS ONE*, vol. 18, no. 11, pp. e0294803, Nov. 2023.

[7] N. S. M. Shafiee *et al.*, “Prediction of mental health problems among higher education student using ML,” *Int. J. Educ. Manage. Eng.*, vol. 10, no. 6, pp. 1–9, 2020.

[8] M. S. Ahmed *et al.*, “A Minimal and Faster System to Identify Depression Through Smartphone: An Explainable ML-Based Approach (Preprint),” *JMIR Formative Research*, Dec. 2022.

[9] K. O. Asare *et al.*, “Predicting Depression From Smartphone Behavioral Markers Using ML Methods, Hyperparameter Optimization, and Feature Importance Analysis,” *JMIR mHealth and uHealth*, vol. 9, no. 7, p. e26540, Jul. 2021.

[10] J. Hong *et al.*, “Depressive Symptoms Feature-Based ML Approach to Predicting Depression Using Smartphone,” *Healthcare*, vol. 10, no. 7, p. 1189, Jun. 2022.

[11] X. Xu *et al*., "Leveraging Routine Behavior and Contextually-Filtered Features for Depression Detection among College Students," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 3, no. 3, pp. 1–33, Sep. 2019.

[12] Y. Hou, J. Xu *et al*., "A big data application to predict depression in the university based on the reading habits," in *Proc. 3rd Int. Conf. Systems and Informatics (ICSAI)*, Shanghai, China, 2016, pp. 1085–1089.

[13] F. Arringoni De *et al*., "Using ecological momentary assessment and ML techniques to predict depressive symptoms in emerging adults," *Psychiatry Research*, vol. 332, p. 115710, Feb. 2024.

[14] Y. Zhou *et al*., "Construction and validation of ML algorithm for predicting depression among home-quarantined individuals during COVID-19 outbreak: based on Adaboost model," *BMC Psychology*, vol. 12, no. 1, Apr. 2024.

[15] P. Jain, K. R. Srinivas, and A. Vichare, "Depression and Suicide Analysis Using ML and NLP," *J. Phys.: Conf. Ser.*, vol. 2161, no. 1, p. 012034, Jan. 2022.

[16] Y. Sharma and B. K. Singh, "Depression Analysis of Voice Samples Using ML," *J. Univ. Shanghai Sci. Technol.*, vol. 22, no. 11, pp. 429–438, Nov. 2021.

[17] P. Manikandaprabhu and S. Sowmiya, "A Study on Depression Analysis Using ML Techniques among College Students," *Int. J. Sci. Res. Eng. Dev.*, vol. 6, no. 3, pp. 137–141, June 2023.

[18] A. Ahmed *et al*., "Thematic Analysis on User Reviews for Depression and Anxiety Chatbot Apps: ML Approach," *JMIR Formative Research*, vol. 6, no. 3, p. e27654, Mar. 2022.

[19] A. Musleh *et al*., "Twitter Arabic Sentiment Analysis to Detect Depression Using ML," *Comput. Mater. Continua*, vol. 71, no. 2, pp. 3463–3477, 2022.

[20] T. Wang *et al*., "Unraveling the Distinction between Depression and Anxiety: A ML Exploration of Causal Relationships," *Comput. Biol. Med.*, vol. 174, p. 108446, May 2024.

[21] T. Richter *et al*., "Using ML-Based Analysis for Behavioral Differentiation between Anxiety and Depression," *Scientific Reports*, vol. 10, p. 16381, Oct. 2020.

[22] J. F. Bauer *et al*., "Validation of ML-Based Assessment of Major Depressive Disorder from Paralinguistic Speech Characteristics in Routine Care," *Depression and Anxiety*, 2024.

[23] A. Kumar, "Sentiment Analysis of Twitter for Detection of Depression Using ML Algorithms," *J. Propuls. Technol.*, vol. 44, no. 04, pp. 7997–8005, Dec. 2023.