The Impact of Smoking and Alcohol Consumption on University Students' Psychological Wellness with Machine Learning Approach

By
Wahid Tausif Islam
201-15-3244

FINAL YEAR DESIGN PROJECT REPORT

This Report Presented in Partial Fulfillment of the Requirements for the **Degree of Bachelor of Science in Computer Science and Engineering**

Supervised by

Ms. Lamia Rukhsara
Lecturer (Senior Scale)
Department of Computer Science and Engineering Daffodil International University

Co-Supervised by

Mr. Md. Shakhawath Hossain
Lecturer
Department of Computer Science and Engineering Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY
Dhaka, Bangladesh

January 13, 2025

APPROVAL

This Project titled **"The Impact of Smoking and Alcohol Consumption on University Students' Psychological Wellness with Machine Learning Approach"**, submitted by **Wahid Tausif Islam,** ID No: **201-15-3244** to the Department of Computer Science and Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on 13 January, 2025.

<u>BOARD OF EXAMINERS</u>

Dr. S.M Aminul Haque (SMAH) — **Chairman**
Professor and Associate Head
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Md. Abbas Ali Khan (AAK) — **Internal Examiner**
Assistant Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Mr. Md. Aynul Hasan Nahid (AHN) — **Internal Examiner**
Lecturer
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Dr. Md. Zulfiker Mahmud (ZM) — **External Examiner**
Professor
Department of Computer Science and Engineering
Jagannath University

DECLARATION

We hereby declare that this project has been done by us under the supervision of **Ms. Lamia Rukhsara, Lecturer (Senior Scale)**, Department of Computer Science and Engineering, Daffodil International University. We also declare that neither this project nor any part of this project has been submitted elsewhere for the award of any degree or diploma.

Supervised by:

Lamia Rukhsara 13.01.25

Ms. Lamia Rukhsara
Lecturer (Senior Scale)
Department of Computer Science and
EngineeringDaffodil International
University

Co-Supervised by:

____

Mr. Md. Shakhawath Hossain
Lecturer
Department of Computer Science and
EngineeringDaffodil International
University

Submitted by:

Tausif Islam

Wahid Tausif Islam
Student ID: 201-15-3244
Department of Computer Science and
EngineeringDaffodil International
University

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 **Final Year Design Project(FYDP)** successfully.

We are grateful and wish our profound indebtedness to **Ms. Lamia Rukhsara**, **Lecturer (Senior Scale)**, Department of Computer Science and Engineering, Daffodil International University, Dhaka, Bangladesh. Deep knowledge and keen interest of our supervisor in the field of **Machine Learning in Health Informatics** to carry out this project. His endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior drafts, and correcting themat all stages 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-mates at Daffodil International University, who took part in this discussion while completing the coursework.

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

ABSTRACT

This study explores the relationship between smoking, alcohol consumption, and psychological wellness among university students, employing machine learning techniques to uncover patterns and provide predictive insights. A dataset of 1163 responses was analyzed, incorporating demographic, behavioral, and mental health-related features. The study applied six machine learning models Logistic Regression, SVM, KNN, XGBoost, Stacking Classifier, and an ensemble of Deep Neural Networks to predict psychological wellness. Among these, the Stacking Classifier emerged as the most effective, achieving an accuracy of 81%, showcasing the advantages of ensemble learning methods in handling complex data patterns. The findings show a strong link between drinking and smoking and mental health outcomes, emphasizing the necessity of focused treatments to break negative patterns and advance mental health. When developing evidence-based plans to enhance students' well-being, policymakers, healthcare professionals, and educational institutions may all benefit from these results. The research highlights sustainability, ethical issues, and the significance of treating data responsibly. Additionally, it offers a starting point for further study to improve mental health prediction and intervention methods, such as utilizing cutting-edge machine learning techniques, integrating longitudinal data, and investigating other behavioral aspects. Students' better lives and behavioral health analytics are advanced by this research.

Table of Contents

List of Figures

List of Tables

Chapter 1
Introduction

1.1 Introduction

University students often experience high levels of stress due to academic pressures, social challenges, and the transition to independence. These stressors can lead to adverse coping mechanisms, such as smoking and alcohol consumption, which are commonly reported among young adults in higher education. While these behaviors may initially serve as temporary stress relievers, they often contribute to long-term psychological challenges, including anxiety, depression, and decreased overall wellness [1].

The connection between smoking, alcohol consumption, and mental health has been widely studied in behavioral and health sciences. However, the majority of these studies focus on traditional statistical analyses, which may not fully capture the intricate and nonlinear relationships within complex datasets. With advancements in technology, machine learning (ML) has emerged as a powerful tool for analyzing multifaceted interactions and identifying hidden patterns [2]. ML models can provide deeper insights into how smoking and drinking habits impact psychological wellness, enabling more effective interventions and support strategies.

This study focuses on using machine learning techniques to investigate the connection between university students' psychological wellbeing and alcohol and smoking usage. The study examines how lifestyle choices affect mental health outcomes using a dataset of 1163 survey responses and assesses the prediction ability of many machine learning models, such as ensemble methods, K-Nearest Neighbors, XGBoost, Support Vector Machine, and Logistic Regression [3].

By integrating behavioral data with ML analysis, this study aims to address critical gaps in existing research, offering a novel perspective on how universities and health practitioners can better support student well-being. The findings contribute to the growing field of technology-driven health interventions and provide actionable insights for reducing harmful habits among students [4].

1.2 Motivation

University students often face significant stress from academic, social, and financial challenges, leading many to adopt unhealthy coping mechanisms like smoking and alcohol consumption. These behaviors can harm mental health, yet traditional research often fails to fully capture the complex relationships between lifestyle habits and psychological wellness.

Machine learning (ML) offers a powerful approach to uncover hidden patterns and interactions within behavioral health data. By applying ML techniques, this study aims to bridge gaps in existing research, identify risk factors, and enable targeted interventions to promote mental well-being [5]. Ultimately, this work seeks to provide actionable insights and practical tools to enhance the quality of life for university students.

1.3 Objectives

The rising prevalence of mental health challenges among university students has become a critical concern for educators, health professionals, and policymakers. Stress from academic demands, coupled with unhealthy coping mechanisms such as smoking and alcohol consumption, significantly impacts students' psychological well-being. While much research has explored these issues, there is a pressing need for more comprehensive approaches that can analyze the multifaceted interactions between lifestyle behaviors and mental health outcomes.

This study is rooted in the recognition that traditional statistical methods often overlook the complexity of behavioral health data. By leveraging machine learning, it becomes possible to explore intricate patterns and predict psychological wellness with greater accuracy. The integration of advanced computational techniques in this field offers new avenues for identifying risk factors, understanding behavioral impacts, and tailoring effective interventions.

The study fills this knowledge vacuum and offers a data-driven framework for promoting the wellbeing of students [6]. The knowledge acquired may be used by academic institutions and medical professionals to proactively minimize unhealthy behaviors and encourage healthier lives, which will eventually improve mental health outcomes in learning environments.

1.4 Research Questions

This study seeks to explore the relationship between smoking, alcohol consumption, and psychological wellness among university students using machine learning techniques. The following research questions guide this investigation:

1. How do smoking and alcohol consumption behaviors impact the psychological wellness of university students?

2. What patterns and correlations exist between smoking, alcohol use, and mental health indicators such as stress levels and coping mechanisms?

3. Which machine learning model demonstrates the highest accuracy in predicting psychological wellness based on behavioral data?

4. How do ensemble learning techniques compare to individual machine learning models in this context?

5. What behavioral or demographic factors are the strongest predictors of psychological wellness, and how can these insights guide targeted mental health interventions?

1.5 Project Outcome

It is anticipated that this study will offer important new information about the effects of alcohol and tobacco use on college students' psychological health. The study intends to find important trends and connections between these habits and mental health markers like stress levels and coping strategies by using machine learning models to analyze survey data. In predicting psychological well-being, it is expected that ensemble learning methods would perform better than individual models and show maximum accuracy. In order to provide a better understanding of the lifestyle choices that lead to psychological discomfort, the study also seeks to uncover important risk factors and behavioral predictors of mental health issues. In the end, it is anticipated that the results will help mental health professionals and academic institutions create focused treatments that would lessen negative behaviors and assist students in acquiring more constructive coping mechanisms.

1.6 Organization of the Report

This project is organized into a series of structured phases to ensure a logical and systematic progression from problem identification to analysis and conclusion. It begins with an Introduction, which outlines the background, objectives, and significance of the study, providing a clear foundation for the research. A thorough Literature Review follows, summarizing existing studies and identifying gaps that the research aims to address. In order to investigate the connection between smoking, alcohol use, and psychological health, the Methodology section outlines the dataset, preprocessing procedures, machine learning models, and assessment metrics used. Data collection and preprocessing include exploratory analysis, feature engineering, and cleaning to get the dataset ready for machine learning applications. In Model Implementation and Evaluation, the effectiveness of several machine learning models—including ensemble approaches—in forecasting psychological wellbeing is then compared. The Results and Discussion section presents key findings, interprets them in the context of the research questions, and highlights behavioral patterns and model performance. Finally, the project concludes with a Conclusion and Recommendations, summarizing insights and suggesting practical interventions and future research directions. A comprehensive list of References ensures proper acknowledgment of all sources and tools used. This cohesive layout ensures that the study is presented in a clear, concise, and accessible manner.

Chapter 2
Background

2.1 Preliminaries

This study focuses on analyzing the impact of smoking and alcohol consumption on university students' psychological wellness using machine learning techniques. The dataset, comprising 1163 responses, includes demographic, behavioral, and mental health-related features. Data preprocessing, including cleaning, feature engineering, and normalization, was conducted to prepare the data for analysis. Machine learning models such as Logistic Regression, Support Vector Machine, XGBoost, K-Nearest Neighbors, and ensemble approaches were applied to uncover patterns and predict psychological wellness. This study builds on previous research by using machine learning to offer insights that are both descriptive and predictive. These first steps serve as the basis for the study's methodology and analysis.

2.2 Literature Review

M. Rezapour et al. [7] use X-ray and CT scan image analysis to evaluate how well different machine learning algorithms diagnose COVID-19. The findings show that ResNet18 had the best accuracy of 100% for X-ray pictures, with DenseNet201 coming in second with an accuracy of 99.70%. When it came to CT scans, DenseNet169 fared better than other models, with an accuracy of 99.80%, whereas Inception ResNet v2 only managed 99.65%. The efficiency of deep learning models in medical imaging for COVID-19 diagnosis is demonstrated by these results.

A random forest machine learning model is created by Y. Shen et al. [8] to forecast the likelihood of suicide attempts among medical college students, a group that has frequently been disregarded in earlier studies. The model made use of 37 input characteristics that were found to be important predictors, such as suicidal thoughts, suicide intentions, anxiety, depression, and the degree of the participant's connection with their father. 90.1% accuracy, 73.51% sensitivity, and 91.68% specificity were attained by the model. This high accuracy shows how well the model can detect those who are in danger of attempting suicide, allowing for prompt treatment. The study highlights how machine learning may be used to estimate suicide risk on college campuses and recommends including it in regular health examinations for individuals who pose a danger.

The application of machine learning (ML) techniques in mental health research to improve diagnosis, prognosis, treatment, and public health outcomes is the main emphasis of A. B. R. Shatte et al. [9]. In order to increase the precision of diagnoses and customize treatments for illnesses like depression and Alzheimer's disease, it examines research using a variety of data sources, including biological data, clinical evaluations, and social media. The study notes that accuracy varies greatly depending on the approach and datasets utilized, but it does not give a particular model accuracy. It highlights that more study is required to verify these machine learning approaches for use in therapeutic settings.

N. H. Gottlieb et al. [10] examine how different demographic and socioeconomic characteristics relate to lifestyle health habits. The study used regression analyses to assess the effects of physical activity and alcohol intake on health outcomes, as well as discriminant analysis to compare groups according to smoking status and health behaviors. However, because of the cross-sectional nature of the data and the uneven distribution of cases among groups, the discriminant functions only explained a relatively tiny percentage of the variance in lifestyle health behaviors, suggesting low model accuracy. Although discriminant analysis was the particular model employed, the report did not provide the model's exact accuracy metric. Rather, it draws attention to the model's limits in explaining variation, implying that its accuracy is poor.

According to N. C. Atuegwu et al. [11], a nationally representative sample of young adults who do not smoke is used to determine the characteristics linked to e-cigarette usage using machine learning (ML) algorithms. The study was effective in identifying both new and previously documented characteristics associated with e-cigarette usage. By reducing reliance on existing knowledge and exploratory hypotheses, the ML technique minimizes the possibility of missing significant aspects related to the result. The snippets that were supplied did not specifically mention the models that were utilized in the study, nor did they provide information on their correctness. Please see the paper's complete text for more detailed information on the models and their performance measures.

Hanawi S. A et al [12] focus on examining the relationship between healthy lifestyle scores and psychological well-being, specifically stress, anxiety, and depression among undergraduate Biomedical Science students. The study utilized the Simple Lifestyle Indicator Questionnaire (SLIQ) and the Depression Anxiety Stress Scale (DASS-42) to assess participants' lifestyles and mental health. The model used in the analysis is a regression model, which showed a statistically significant relationship with an R^2 value of 0.102 and a p-value of 0.007, indicating that the model explains about 10.2% of the variance in healthy lifestyle scores based on the independent variables of stress, anxiety, depression, age, gender, and type of residence.

Examining the connection between psychological well-being, particularly stress, anxiety, and depression, and healthy lifestyle scores among undergraduate Biomedical Science students is the main goal of Hanawi S. A. et al. To evaluate the lives and mental health of the individuals, the study used the Depression Anxiety Stress Scale (DASS-42) and the Simple Lifestyle Indicator Questionnaire (SLIQ). With an R2 value of 0.102 and a p-value of 0.007, the regression model utilized in the analysis demonstrated a statistically significant relationship, meaning that depending on the independent variables of stress, anxiety, depression, age, gender, and type of residence, the model explains roughly 10.2% of the variance in healthy lifestyle scores.

Confirmatory Factor Analysis (CFA) was used by J. A. Epstein et al. [13] to evaluate the connections between a number of latent components linked to teen alcohol use and psychological well-being. Each of the five latent components identified by the model was assessed using four indicator items. With the following accuracy metrics: chi-square (χ^2) = 508, degrees of freedom (df) = 158, χ^2/df = 3.2, Standardized Root Mean Squared Residual (SRMR) = 0.031, and Comparative Fit Index (CFI) = 0.97, the CFA results showed that the measurement model was a good-to-excellent fit.

A deep-learning algorithm is created by S. Hassanpour et al. [14] with the goal of detecting high substance use risk by analyzing Instagram postings. This model combines text and picture data, using a combination of word2vec and LSTM networks for text representation and a ResNet18 architecture for image feature extraction. With a precision of 68.6%, recall of 76.6%, F-measure of 72.4%, and an Area Under the Receiver Operating Characteristic Curve (AUROC) of 0.00008, the model substantially outperformed random chance in identifying alcohol risk. This study is an innovative attempt to use social media data for drug use screening, and by identifying socio-environmental characteristics associated with substance use risk, it may improve current evaluation techniques.

2.3 Comparative Analysis and Summary

The comparison investigation reveals notable differences in how effectively machine learning models predict psychological wellbeing. The best accuracy (0.81) was obtained by the Stacking Classifier, proving that ensemble approaches are useful for identifying intricate patterns. K-Nearest Neighbors (0.68) outperformed Logistic Regression (0.63) and Support Vector Machine (0.58), while XGBoost (0.55) showed modest results. The Deep Neural Network ensemble achieved 0.64 accuracy, performing below simpler models like KNN.

All things considered, the findings highlight the effectiveness of ensemble methods such as the Stacking Classifier, which use many models to increase prediction accuracy and are therefore useful for behavioral health research.

2.4 Scope of the Problem

The increasing prevalence of mental health challenges among university students is a pressing issue, with smoking and alcohol consumption often contributing to these struggles. As students face academic, social, and personal stressors, many resort to these behaviors as coping mechanisms, potentially exacerbating psychological distress. Despite the growing concern, traditional methods of understanding and addressing these behaviors often fall short in capturing the complex interplay between lifestyle habits and mental health outcomes.

In order to close this gap, this study uses machine learning to investigate the connections between alcohol use, smoking, and psychological well-being. The research seeks to identify trends in behavioral and demographic data that might guide focused treatments and aid in the prediction of mental health outcomes. The scope goes beyond only identifying risk factors; it also includes offering practical advice on how colleges and mental health professionals may enhance the wellbeing of students. Thus, this study advances our knowledge of the behavioral aspects influencing mental health and aids in the creation of evidence-based intervention options.

2.5 Challenges

There were a number of difficulties in carrying out this investigation, mostly pertaining to data gathering, preprocessing, and model optimization. Finding high-quality, self-reported data via online surveys was one of the first challenges since replies might be skewed by biases like exaggeration or underreporting. Another issue was making sure the dataset was big enough and varied enough to capture the behavioral patterns of college students.

Data preprocessing posed its own set of challenges, including handling missing values, encoding categorical variables, and addressing potential outliers that could skew the analysis. Selecting the most suitable machine learning models and fine-tuning their hyperparameters required careful experimentation to achieve optimal performance. Additionally, balancing the interpretability of the models with predictive accuracy was crucial, as overly complex models, such as deep neural networks, were less effective than simpler ensemble approaches in this context [15].

Last but not least, ensuring relevance and practicality requires careful consideration when converting the results into usable insights for real-world applications. The study effectively offers insightful information about the connection between university students' psychological well-being, alcohol use, and smoking despite these obstacles.

Chapter 3
Research Methodology

3.1 Research Subject and Instrumentation

The research subjects for this study consisted of 1163 university students who voluntarily participated in an online survey. These students represented diverse academic backgrounds, years of study, and demographic profiles, ensuring a broad perspective on the relationship between smoking, alcohol consumption, and psychological wellness. The inclusion criteria focused on individuals currently enrolled in a university program and willing to provide self-reported data on their lifestyle habits and mental health. A Google Forms-administered structured online survey served as the instrumentation. The survey was created to gather information on important factors such demographics, how often people smoke and drink, coping strategies, mental health, and behaviors related to getting treatment. In order to obtain a thorough understanding of the respondents' behaviors and psychological well-being, the questionnaire contained both numerical and category items. The survey was pre-tested with a limited sample of participants prior to complete deployment in order to guarantee the quality and reliability of the findings. A strong dataset for machine learning research was produced by combining a variety of subjects with a well-crafted survey instrument. This allowed for the investigation of complex patterns and correlations that guide the goals of this work.

3.1.1 Data Sources

This study's core data was gathered via an online survey that was delivered using Google Forms [16]. The purpose of the survey was to collect thorough self-reported data from college students on their demographics, coping strategies, mental health, and smoking and alcohol use patterns. Since 1163 people freely submitted their responses, the dataset was heterogeneous, reflecting a range of academic fields, study years, and demographic characteristics. Through a combination of numerical and category questions, the questionnaire aimed to capture important characteristics associated with lifestyle choices and psychological wellbeing. In order to improve the survey's clarity and dependability, a pre-testing step was conducted to guarantee the data's quality and relevancy. In order to explore the complex connections between students' behavioral patterns and their mental health outcomes, the machine learning analysis was built around this basic dataset.

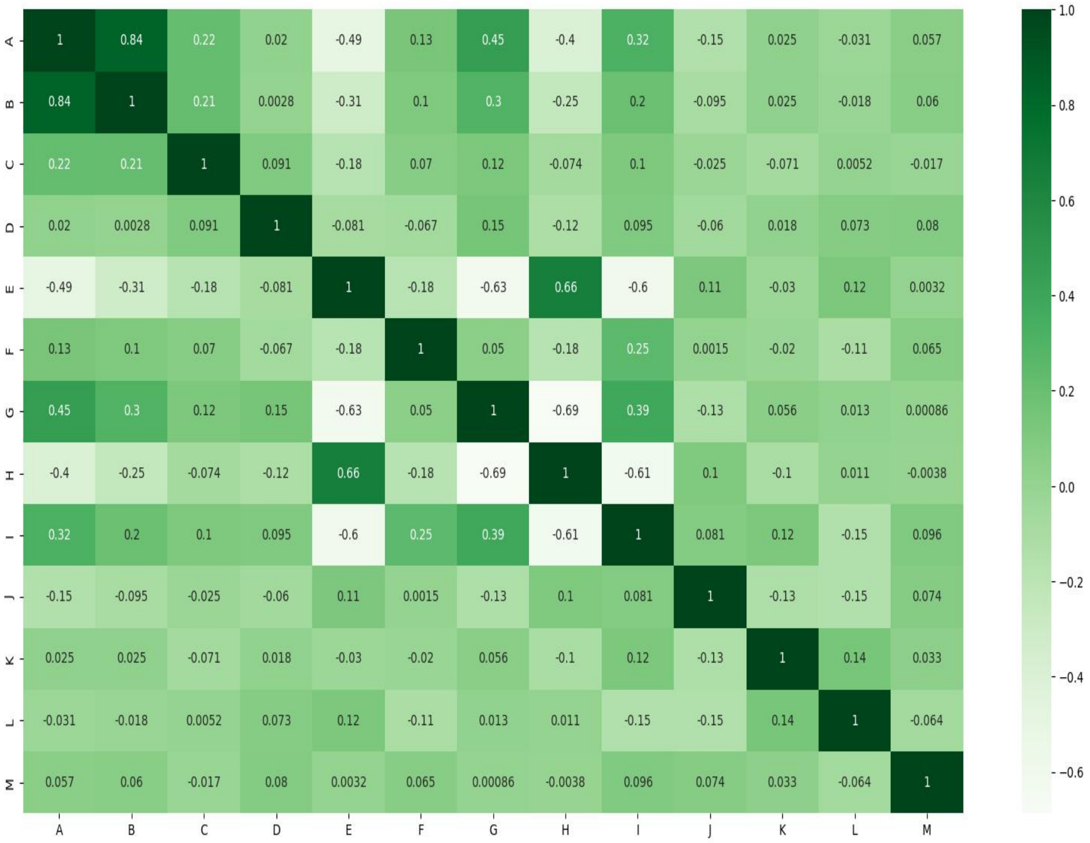

Figure 3.1: Corelation matrix of the dataset

3.1.2 Tools and Software

To make data preparation, analysis, and visualization easier, this study used a variety of tools and applications. Google Forms was used to gather data since it offered a user-friendly and effective platform for conducting the survey and managing replies. With the help of essential modules like scikit-learn for creating and assessing machine learning models, pandas for data manipulation, and NumPy for numerical calculations, Python was the main programming language used for data pretreatment and analysis. Libraries such as XGBoost for gradient boosting and TensorFlow/Keras for deep neural network development and training were used to further improve advanced machine learning approaches. Using Matplotlib and Seaborn, data visualization was accomplished, yielding lucid and perceptive graphical depictions of the results. Coding, debugging, and documenting were all made easy by the integrated development environment (IDE), Jupyter Notebook. The study's reliable and repeatable results were guaranteed by these tools and software, which also made it easier to handle the dataset and build machine learning models.

3.2 Data Collection Procedure

An online survey administered to university students using Google Forms was used to gather data for this study. The poll asked about demographics, alcohol and tobacco use, stress, mental health, and behaviors related to getting treatment. To guarantee thorough data collection, it included open-ended, multiple-choice, and ordinal scale questions. Survey participation was optional and anonymous, and it was disseminated by email, social media, and academic networks. The questionnaire had a pre-testing process to improve its clarity. 1163 full replies were obtained from the procedure, giving machine learning analysts a varied dataset.

3.3 Statistical Analysis

This study's statistical research uses a multipronged approach, concentrating on key performance indicators to thoroughly assess machine learning algorithms' efficacy across many security tiers. To provide a more comprehensive view of algorithmic performance across a wide range of security settings, the metrics Accuracy, Precision, Recall, and F1-Score are employed.

3.3.1 Accuracy

Calculating the accuracy of a model involves dividing the total number of predictions by the number of correctly predicted cases, both positive and negative. This straightforward metric performs well on datasets that are balanced. Due to its inability to differentiate between various types of errors, it might be misleading when the data is not balanced [17].

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad \text{(i)}$$

Where,

TP (True Positives): The number of instances correctly predicted as positive (presence of liver disease).

TN (True Negatives): The number of instances correctly predicted as negative (absence of liver disease).

FP (False Positives): The number of instances incorrectly predicted as positive.

FN (False Negatives): The number of instances incorrectly predicted as negative.

3.3.2 Precision

Precision is concerned with the caliber of accurate forecasts. Out of all the cases that were projected to be positive, it determines the percentage of real positive predictions. When a model has high accuracy, it produces fewer false positives [18].

$$Precision = \frac{TP}{TP+FP} \quad \text{(ii)}$$

3.3.3 Recall

The model's recall gauges how well it can recognize real positive examples. Out of all the real positive cases in the dataset, it determines the percentage of genuine positives. In situations like illness diagnosis or fraud identification, a high recall means fewer false negatives [19].

$$Recall = \frac{TP}{TP+FN} \quad \text{(iii)}$$

3.3.4 F1-Score

A balanced statistic that accounts for both false positives and false negatives is provided by the F1 Score. It is computed by taking the accuracy and recall harmonic means. It is particularly useful when the dataset is imbalanced or when accuracy and recall need to be modified [20].

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad \text{(iv)}$$

3.4 Proposed Methodology

This study uses machine learning approaches to investigate the association between university students' psychological well-being, alcohol intake, and smoking in a methodical manner. 1163 students participated in an online survey that gathered information on their demographics, behaviors, and mental health. To find patterns and correlations, exploratory data analysis (EDA) was carried out after the data had been cleaned, encoded, and normalized. Metrics including accuracy, precision, recall, and F1 Score were used to assess the implementation and performance of many machine learning models, such as ensemble approaches, K-Nearest Neighbors, XGBoost, Support Vector Machine, and Logistic Regression. Through comparison research, the top-performing model was found in order to extract valuable information and offer practical suggestions for improving the well-being of students.

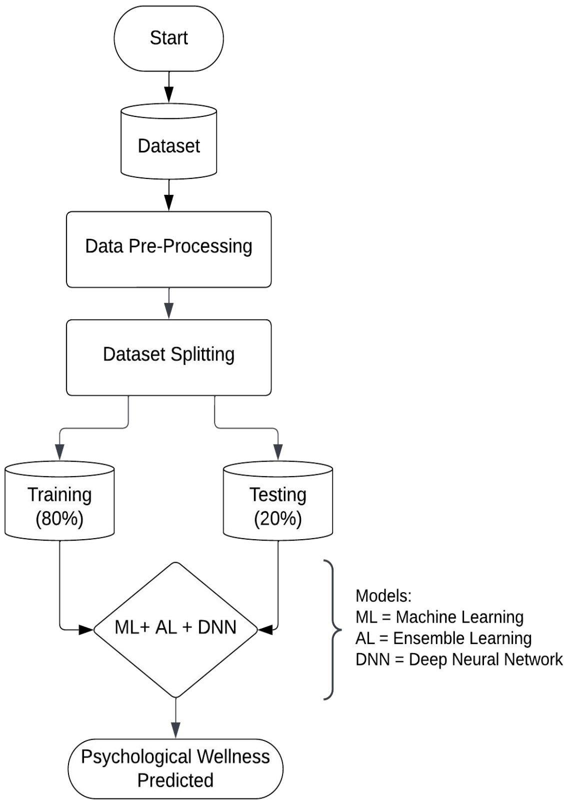

Figure 3.2: Proposed Model

The suggested model will initially acquire information from the dataset. Following that, the data will be preprocessed using the appropriate procedures.

After that, the dataset will be separated into two halves: training and testing. 20% should be maintained for testing, with the remaining 80% for training. The model will be trained with machine learning across four distinct layers. Finally, you should be able to tell whether or not it was located.

3.5 Implementation Requirements

To conduct data preparation, machine learning analysis, and result visualization, this study's implementation calls for a mix of hardware, software, and technical resources. In terms of hardware, a system that can handle the dataset and effectively train machine learning models needs a multi-core CPU and at least 8 GB of RAM. Access to a GPU-accelerated system is advantageous for computationally demanding activities, especially when training deep neural networks.

The Python programming language and libraries like pandas, NumPy, and scikit-learn are needed for data preparation and model implementation. Additionally, deep learning models and sophisticated machine learning methods require specific libraries like TensorFlow/Keras and XGBoost. Matplotlib and Seaborn are used to visualize the findings, while Jupyter Notebook is utilized as the development environment for coding and documentation.

Accurate model implementation and interpretation also depend on technical proficiency in data processing, feature engineering, and machine learning. Together, these prerequisites offer the basis for effectively evaluating the dataset and accomplishing the study's goals.

3.5.1 Machine Learning Models

3.5.1.1 Logistic Regression (LR)

The statistical model known as logistic regression is applied to situations involving binary categorization. The logistic function is used to forecast the likelihood of a class label. It converts any real-valued integer to a value between 0 and 1 [21].

Formula:

$$P(y = 1|X) = \frac{1}{1+e^{-(\beta_0+\beta_1 X_1+\beta_2 X_2+\ldots+\beta_n X_n)}} \quad \text{(v)}$$

Where β_0 is the intercept, $\beta_1, \beta_2, \ldots, \beta_n$ are the coefficients, and $X_1, X_2, \ldots, X_n$ are the features.

3.5.1.2 Support Vector Machine (SVM)

SVM is a supervised learning model that is applied to regression and classification. In feature space, it determines the best hyperplane to optimize the margin between classes [22].

Formula for Decision Boundary:

$$f(X) = \mathrm{sign}((w, X) + b) \quad \text{(vi)}$$

Where w is the weight vector, X is the input vector, b is the biasm and (·) represents the dot product.

Optimization Objective:

$$\min \frac{1}{2} ||w||^2 \text{ subject to } y_i((w, X_i) + b) \geq 1 \text{ for all i} \quad \text{(vii)}$$

3.5.1.3 XGBoost:

A decision-tree-based ensemble machine learning model called XGBoost (Extreme Gradient Boosting) employs boosting to increase prediction accuracy. To avoid overfitting, it optimizes an objective function that consists of a regularization term and a loss function [23].

Objective Function:

$$\mathrm{Obj} = \sum_{i=1}^{n} L(y_i \hat{y}_i) + \sum_{k=1}^{K} \Omega\,(f_k) \quad \text{(viii)}$$

Where L is the loss function, $\hat{y}_i$ is the predicted value, and $\Omega(f_k)$ is the regularization term for tree complexity.

3.5.1.4 K-Nearest Neighbors (KNN):

A non-parametric approach called KNN uses the majority class of its K nearest neighbors to classify a data item [24].

Distance Metric (Euclidean Distance):

$$d(p, q) = \sqrt{\sum_{i=1}^{n} (p_i - q_i)^2} \qquad \text{(ix)}$$

Where p and q are two points in n-dimensional space

Classification Rule:

Assign the label y to a data point based on the majority label among K-Nearest Neighbors.

3.5.1.5 Ensemble Learning (Stacking Classifier):

In order to get final predictions, stacking integrates predictions from many base models and applies a meta-model. The meta-model combines the outputs of the basis models to increase accuracy, while the base models learn complementing characteristics [25].

General Formula:

Let base models be $f_1(X), f_2(X), \dots, f_n(X)$ and the meta-model be g

$$\hat{y} = g(f_1(X), f_2(X), \dots, f_n(X), \dots, f_n(X)) \qquad \text{(x)}$$

Where g takes the outputs of the base models as input and generates the final prediction.

3.5.1.6 Ensemble of fully connected Deep Neural Networks (DNNs):

Higher-order representations of the data are learned by each layer of a DNN, which is made up of layers of neurons coupled by weights. To increase accuracy and robustness, the ensemble integrates many DNNs [26].

Forward propagation in a single year:

$$z^{(l)} = W^{(l)} a^{(l-1)} + b^{(l)} \qquad \text{(xi)}$$

$$a^{(l)} = \sigma(z^{(l)}) \qquad \text{(xii)}$$

Where $z^{(l)}$ is the linear transformation, $W^{(l)}$ *and* $b^{(l)}$ are weights and biases for layer l, $a^{(l)}$ is the activation, and σ is the activation function.

LossFunction: Problem-specific factors determine the DNN's loss function. The use of cross-entropy loss for binary classification is common:

$$L = -\frac{1}{N}\sum_{i=1}^{N}[y_i \log(\hat{y}_i) + (1 - y_i)\log(1 - \hat{y}_i)] \qquad \text{(xiii)}$$

Where y_i is the true label, and $\hat{y}_i$ is the predicted propability

Chapter 4

Implementation and Results

4.1 Experimental Setup

To manage missing values, standardize numerical characteristics, and encode categorical variables, the survey data had to be preprocessed as part of the experimental setup. The equipment used for the trials has a multi-core CPU, 16 GB of RAM, and a GPU for demanding calculations. Python was utilized for data processing, model development, and assessment, utilizing packages such as scikit-learn, pandas, NumPy, and TensorFlow/Keras. Using a training-testing ratio of 80:20, models such as Stacking Classifier, SVM, XGBoost, KNN, Logistic Regression, and a Deep Neural Network ensemble were trained on the dataset [27].

4.2 Experimental Results & Analysis

The experimental findings demonstrate how well different machine learning models predict psychological well-being based on alcohol and smoking behaviors. The best accuracy of 81% was attained by the Stacking Classifier, proving that ensemble approaches are useful for integrating many predictions. Deep Neural Networks and Logistic Regression performed quite well, with accuracies of 63% and 64%, respectively, while KNN came in second with 68%. With accuracy rates of 58% and 55%, respectively, SVM and XGBoost fared worse. These findings underscore the influence of data properties in model performance and the significance of ensemble approaches for reliable predictions. Targeted treatments may be developed with the help of the data, which offer important insights into how smoking and drinking affect students' mental health.

4.2.1 Machine Learning Models

This study used machine learning to predict students' psychological well-being based on smoking and alcohol use. Logistic regression analyzed linear relationships, SVM captured nonlinear patterns, and XGBoost handled complex interactions. KNN provided a non-parametric approach, while ensemble methods like Stacking Classifier and DNNs excelled in accuracy. Models were evaluated with metrics like accuracy, precision, recall, and F1 Score.

4.2.1.1 For Logistic Regression (LR)

Here are the details of the classification metrics report:

Table 4.1: Classification Result for Logistic Regression

Accuracy	Precision	Recall	F1-Score
0.64	0.58	0.64	0.59

Confusion Matrix:

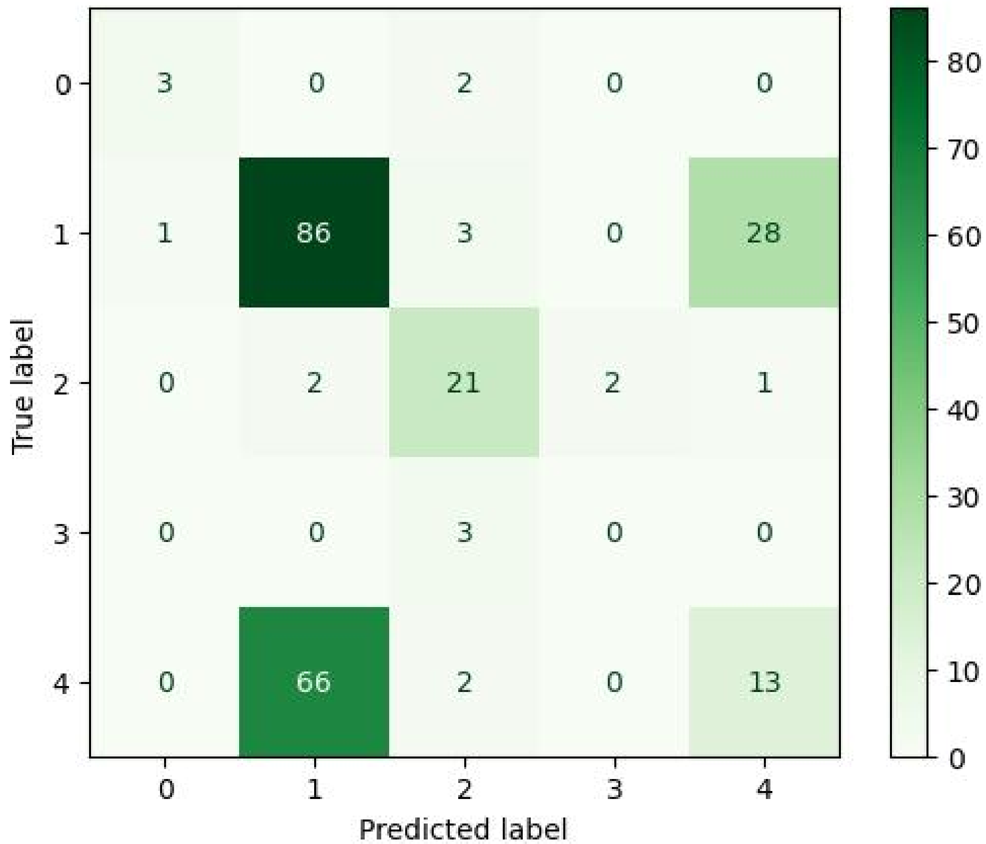

Figure 4.1: Confusion Matrix for Logistic Regression

The distribution of true vs anticipated labels for every class is displayed in the fig. 4.1 confusion matrix. Class 0 and Class 3 are poorly predicted, but Class 1 and Class 2 have high correct predictions. Class 4 is noticeably misclassified as Class

1. To enhance model performance for underrepresented classes, more research is required.

4.2.1.2 For Support Vector Machines (SVM)

Here are the details of the classification metrics report:

Table 4.2: Classification Result for Support Vector Machines

Accuracy	Precision	Recall	F1-Score
0.59	0.58	0.59	0.55

Confusion Matrix:

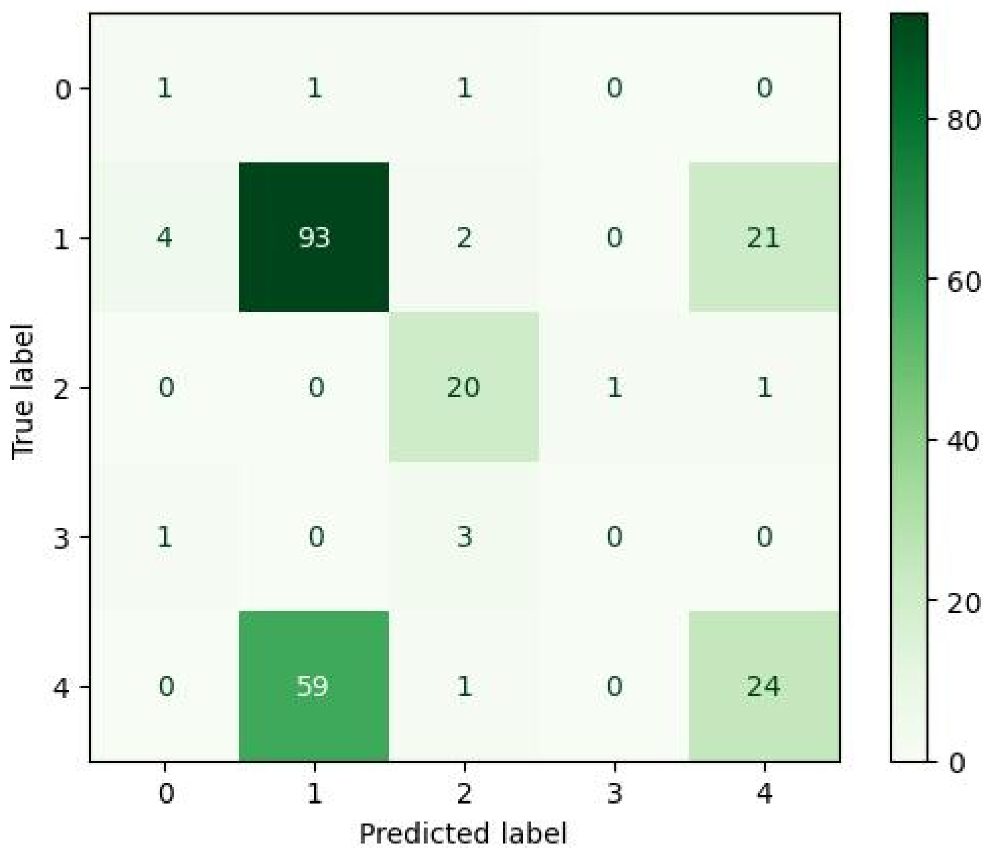

Figure 4.2: Confusion Matrix for Support Vector Machines

According to the fig. 4.2 confusion matrix, Class 1 and Class 2 have a respectable level of accuracy, with the majority of their predictions being accurate. Class 3 provides few correct predictions and Class 4 has significant misclassification into Class 1, suggesting that the model needs to be improved for these classes.

4.2.1.3 For XGBoost

Here are the details of the classification metrics report:

Table 4.3: Classification Result for XGBoost

Accuracy	Precision	Recall	F1-Score
0.56	0.50	0.56	0.51

Confusion Matrix:

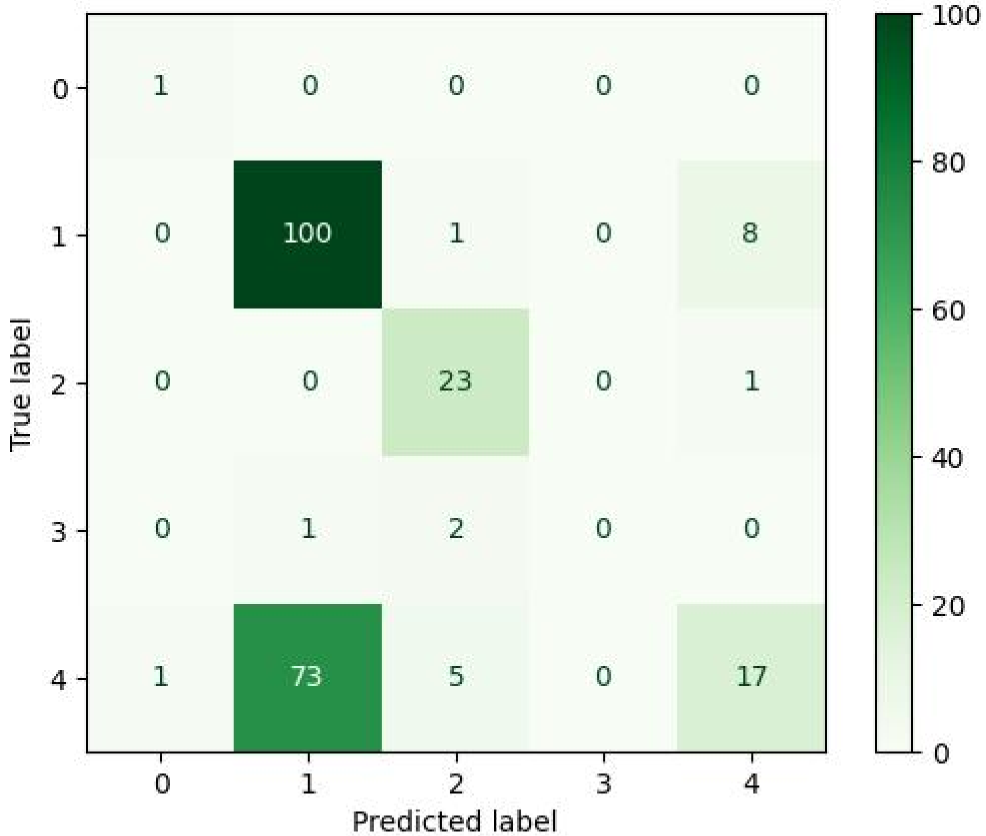

Figure 4.3: Confusion Matrix for XGBoost

The fig. 4.3 confusion matrix in this figure is shown as a heatmap. It displays a classification model's performance over five different classes (0–4). The columns match the expected labels, while the rows match the real labels.

4.2.1.4 For K-Nearest Neighbors (KNN)

Here are the details of the classification metrics report:

Table 4.4: Classification Result for K-Nearest Neighbors (KNN)

Accuracy	Precision	Recall	F1-Score
0.69	0.67	0.69	0.67

Confusion Matrix:

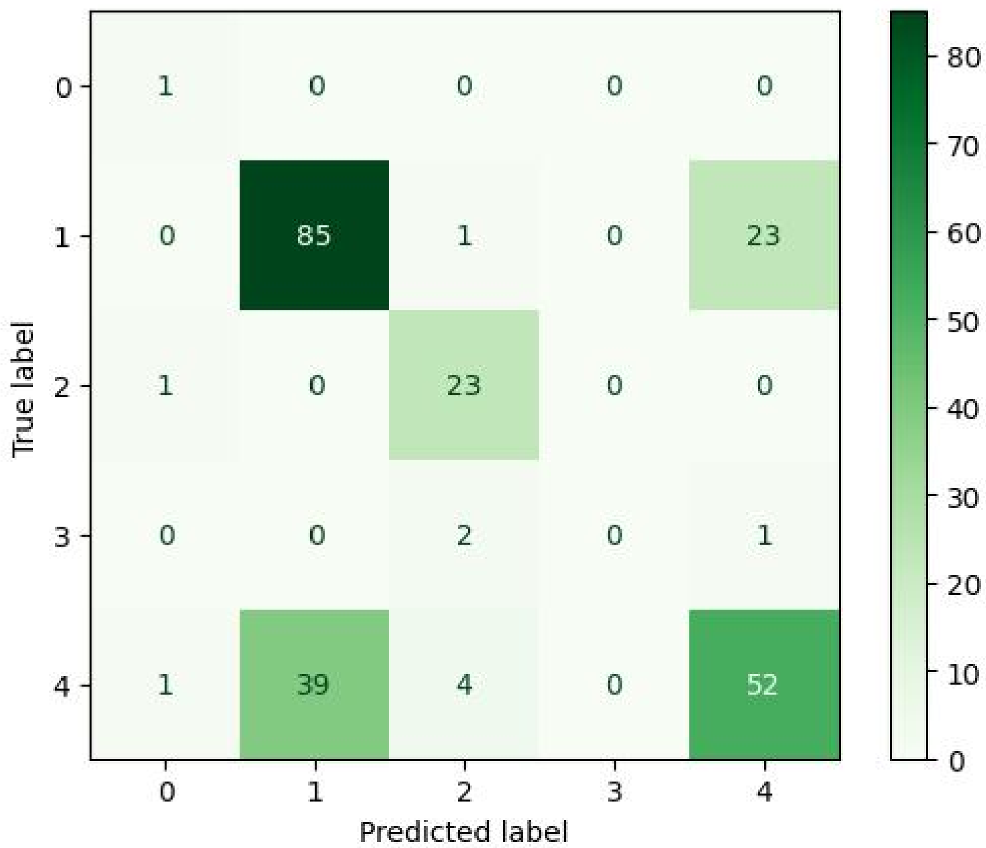

Figure 4.4: Confusion Matrix for KNN

The model's performance over five classes (0–4) is displayed in this fig. 4.4 confusion matrix, with accurate predictions along the diagonal (e.g., 86 for class 1, 57 for class 4). Off-diagonal misclassifications can be seen, such as the 27 class 4 samples that were mistakenly categorized as class 1. With deeper hues for lower values and yellow for greater ones, the color intensity corresponds to prediction counts.

4.2.1.5 For Ensemble learning (Stacking Classifier)

Here are the details of the classification metrics report:

Table 4.5: Classification Result for Ensemble Learning

Accuracy	Precision	Recall	F1-Score
0.81	0.80	0.81	0.79

Confusion Matrix:

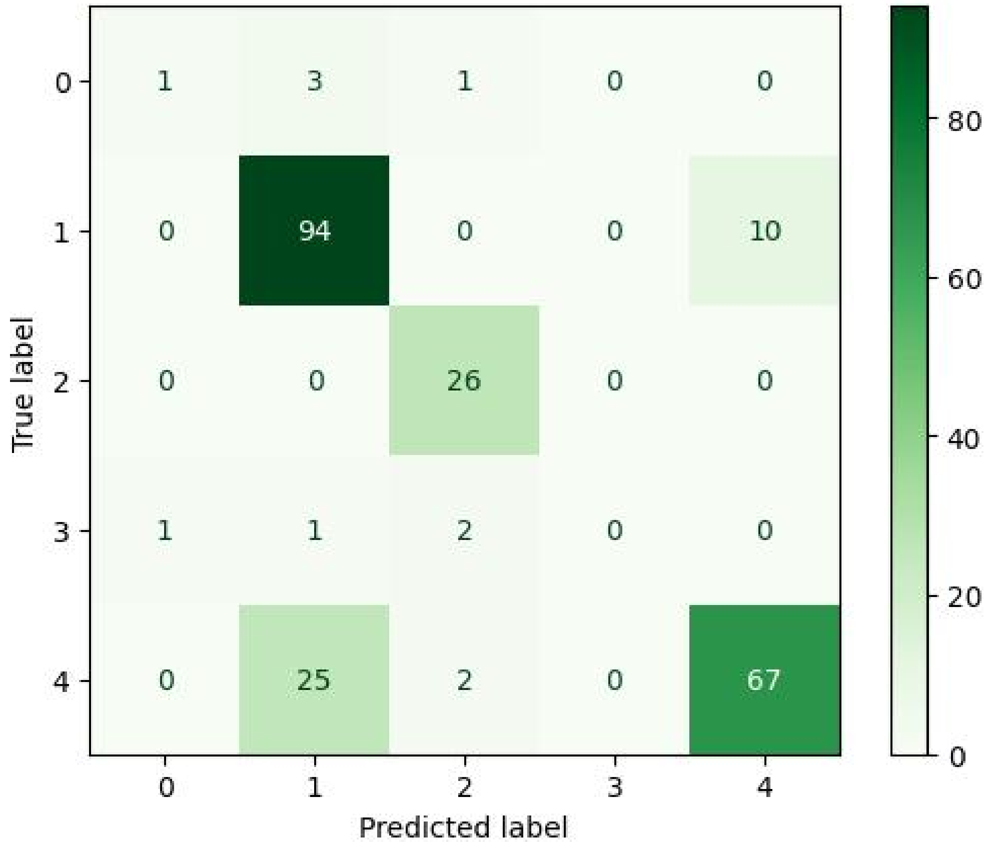

Figure 4.5: Confusion Matrix for Ensemble Learning

The model's classification performance is displayed in the fig. 4.5 confusion matrix, where the x-axis represents predicted labels and the y-axis represents genuine labels. The diagonal represents accurate predictions, whereas the off-diagonal represents misclassifications. Higher values correspond to darker hues.

4.2.1.6 For Ensemble of fully connected DNN

Table 4.6: Classification Result for ensemble fully connected DNN

Accuracy	Precision	Recall	F1-Score
0.61	0.62	0.61	0.60

Confusion Matrix:

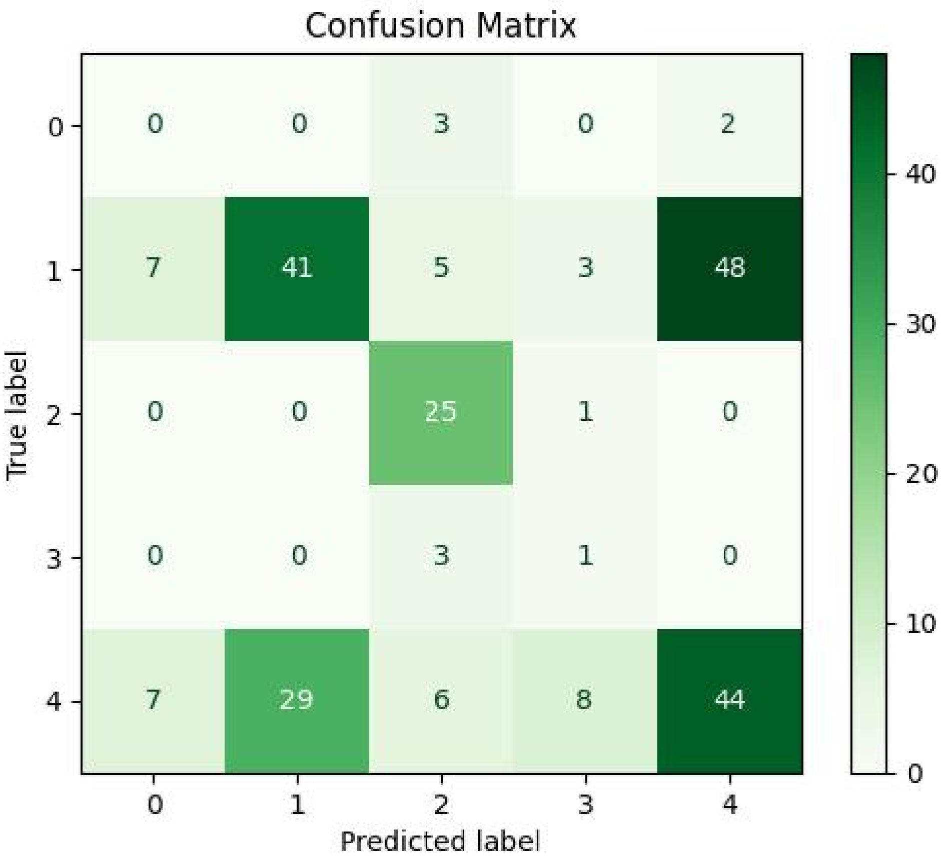

Figure 4.6: Confusion Matrix for Ensemble of fully connected DNN

The model's classification performance is displayed in the fig. 4.6 confusion matrix, where rows represent actual labels and columns represent predictions. On the diagonal, darker cells show accurate classifications, whereas off-diagonal cells show incorrect classifications, such the frequent mistakes between classes 1 and 3 and classes 4 and 3.

Training and Validation Performance:

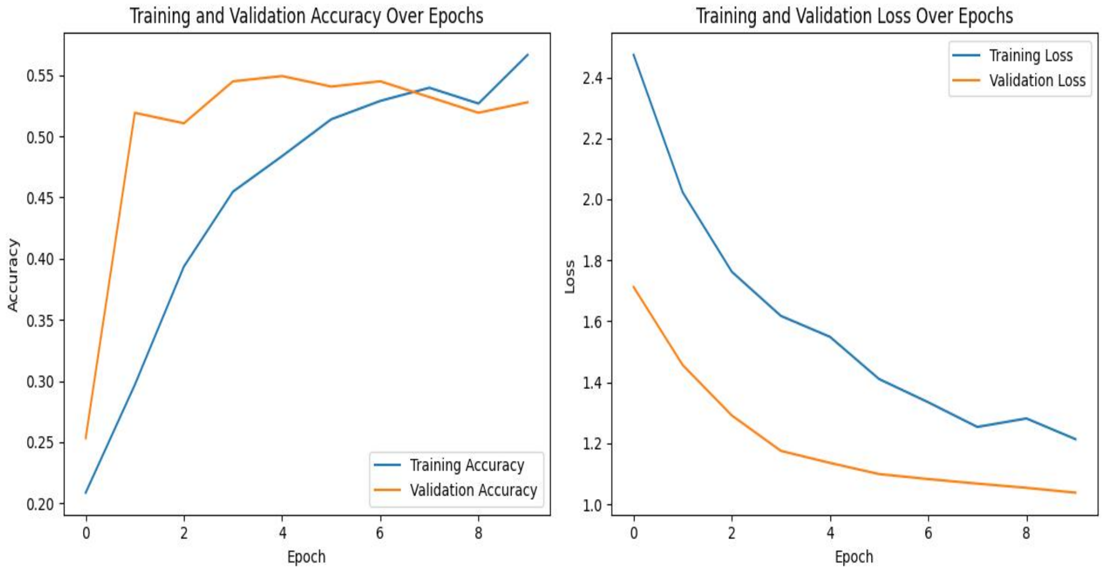

Figure 4.7 Training and Validation Performance Over Epochs

This fig. 4.7 shows the model's performance over 10 epochs. Training accuracy and loss improve consistently, indicating effective learning. Validation accuracy plateaus, and validation loss flattens after initial improvement, suggesting slight overfitting and the need for techniques like regularization or early stopping.

Here the comparison is highlighted among the Machine Learning Models.

Table 4.7: Classification Results of the models

Models	Accuracy	Precision	Recall	F1-Score
LR	0.64	0.58	0.64	0.59
SVM	0.59	0.58	0.59	0.55
XGB	0.56	0.50	0.56	0.51
KNN	0.69	0.67	0.69	0.67
Ensemble Learning	0.81	0.80	0.81	0.79
DNN	0.61	0.62	0.61	0.60

Table 4.7 compares model performance in terms of accuracy, precision, recall, and F1-score. Ensemble Learning had the best accuracy (0.81) and F1-score (0.79), effectively balancing precision and recall. K-Nearest Neighbors (KNN) came in second with 0.69 accuracy and 0.67 F1-score. The results of Logistic Regression (LR) and Support Vector Machine (SVM) were moderate, while XGBoost (XGB) performed poorly. The Deep Neural Network (DNN) produced reasonable results, with an accuracy of 0.61 and an F1-score of 0.60. These findings emphasize the efficacy of ensemble approaches for this purpose.

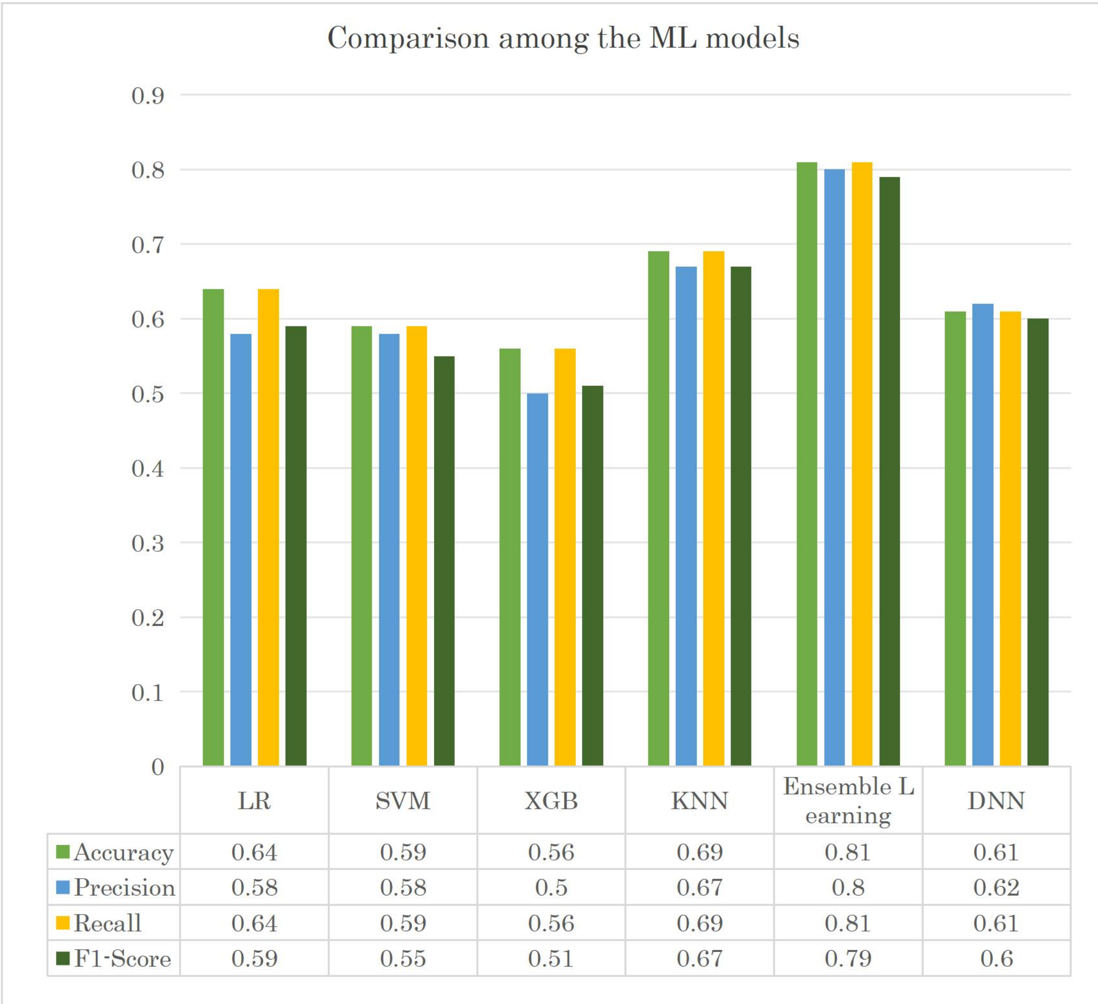

Figure 4.6: Comparison of the models

Figure 4.6 examines the performance of various machine-learning models. Ensemble Learning outperforms all other methods in terms of accuracy (0.81), precision (0.80), recall (0.81), and F1-score. K-Nearest Neighbors (KNN) follows with an accuracy of 0.69 and an F1-score of 0.67. Logistic Regression (LR) and Support Vector Machine (SVM) have similar accuracies of 0.64 and 0.59.

XGBoost (XGB) has the poorest performance, with an accuracy of 0.56 and an F1-score of 0.51. The Deep Neural Network (DNN) yields reasonable results, with an accuracy of 0.61 and an F1-score of 0.60. Ensemble Learning obviously surpasses the other models in every statistic.

4.3 Discussion

This study shows how effectively machine learning can analyze the effects of alcohol and tobacco use on college students' psychological well-being. The Stacking Classifier demonstrated the power of ensemble approaches by outperforming other models with an accuracy of 81%. The intricacy of the dataset presented difficulties for models like SVM and XGBoost, but KNN and Logistic Regression performed rather well. The findings highlight the important role that behavioral variables play in mental health and provide information for early detection of at-risk individuals and focused treatments. Generalizability may be impacted by restrictions such data imbalance and the use of self-reported replies, which emphasizes the need for more improvement in subsequent studies.

Chapter 5

Engineering Standards and Design Challenges

5.1 Impact on Society

This study has important societal ramifications, especially when it comes to university students' psychological well-being. The study offers data-driven insights that can guide public health policies and intervention tactics by using machine learning to examine the impacts of drinking and smoking on mental health. By identifying at-risk kids early on, the predictive models used in this study can provide resources and tailored support to reduce negative behaviors and enhance mental health. The results also highlight the close relationship between lifestyle decisions and mental health, which makes preventative treatment a top priority for politicians, healthcare professionals, and educational institutions. Furthermore, the study emphasizes how technology-driven methods in mental health analytics may promote creativity in tackling pervasive social problems like stress and drug misuse [28]. By encouraging better habits and strengthening support networks, this study helps create a society that is more resilient, healthy, and knowledgeable.

5.2 Impact on Environment

The psychological well-being of college students is the main emphasis of this study, but it also subtly highlights the wider environmental effects of alcohol and tobacco use. Through the garbage produced by cigarette butts, alcohol packaging, and industrial procedures that put pressure on natural resources, these activities contribute to environmental deterioration. This research can help develop methods to lower consumption and the environmental impact it causes by understanding the factors that drive such behaviors and advocating healthier alternatives. The awareness raised by research such as this may promote sustainable actions in institutions and communities, such as promoting ecofriendly goods and putting in place waste reduction initiatives. Resolving these patterns of behavior not only improves the health of the individual but also helps to maintain a better, cleaner environment [29].

5.3 Ethical Aspects

To maintain the integrity and respect of the participants, this study complied with stringent ethical criteria. Everyone who participated voluntarily gave their informed consent for the data to be gathered. To reduce any possible hazards or discomfort related to the study, participants were guaranteed anonymity and the confidentiality of their answers. No compulsion of any kind was used in the study, and participants were free to leave at any time. Responsible data analysis ensured that results were utilized only for scholarly and social purposes, free from abuse or deception. By adhering to these moral guidelines, the research supports the values of equity, openness, and respect for personal freedoms [30].

5.4 Sustainability Plan

A sustainability strategy has been developed with an emphasis on ongoing implementation and distribution of the study's results in order to guarantee its long-term influence. Ongoing insights into the connection between lifestyle choices and psychological well-being are made possible by the ability to update the prediction models created with fresh data on a regular basis, which keeps them accurate and relevant. This flexibility guarantees that the findings of the study will continue to be relevant to changing social trends. Working together with academic institutions, medical professionals, and legislators is essential to putting the study's recommendations into practice. The findings may be used to create digital tools and educational initiatives that encourage students to lead better lifestyles. Furthermore, the study promotes virtual approaches for data collecting and analysis in order to reduce resource waste, which is consistent with ecologically friendly practices. While upholding ecological responsibility, this approach guarantees that the study's conclusions will have a long-lasting positive impact on society [31].

Chapter 6
Conclusion

6.1 Summary

This study used machine learning techniques to investigate the effects of drinking and smoking on university students' psychological well-being. In order to find trends and forecast mental health consequences depending on lifestyle choices, a dataset of 1163 answers was examined. With an accuracy of 81%, the Stacking Classifier outperformed the other machine learning models, which included Logistic Regression, SVM, KNN, XGBoost, Stacking Classifier, and an ensemble of DNNs. Important associations between these activities and mental health were found, providing important information for focused treatments and early at-risk person identification. This study uses machine learning to improve behavioral health analytics, guide public health initiatives, and encourage university students to lead healthier lifestyles. By highlighting the significance of moral behavior, sustainability, and the good of society, the study makes a significant contribution to both scholarly research and real-world mental health solutions.

6.2 Conclusion

This study effectively illustrated how machine learning may be used to comprehend the intricate connection between university students' psychological well-being, alcohol use, and smoking. The study demonstrates how predictive analytics can be used to uncover behavioral patterns that impact mental health by using a variety of models, with the Stacking Classifier doing the best. The knowledge acquired highlights the necessity of focused treatments to address negative habits and advance mental health. The report also emphasizes how crucial it is to use data-driven methods to guide public health initiatives, opening the door to more individualized and efficient solutions. Notwithstanding several drawbacks, such as its dependence on self-reported data, this study offers a solid basis for further research and applications targeted at enhancing student health and promoting beneficial social effects.

6.3 Further Work

This study provides a number of new directions for investigating the connection between lifestyle choices and mental health. Findings can be made more broadly applicable by expanding the dataset to encompass a more varied population from various demographic groups and geographical areas. A more thorough knowledge of the elements influencing mental health might be obtained by conducting more research on other topics, such as eating patterns, exercise regimens, or social support networks. To increase predicted accuracy, more sophisticated machine learning methods might be investigated, such as deep learning with bigger datasets or natural language processing on qualitative replies. Furthermore, using longitudinal data may shed light on how smoking and alcohol use affect psychological well-being over the long run. These avenues have the potential to enhance comprehension of mental health predictors and facilitate the creation of more focused and successful intervention techniques.

References

[1] A. P. Smith, “Smoking, Alcohol, Wellbeing and Academic Attainment,” Journal of Health and Medical Sciences, vol. 2, no. 3, Sep. 2019, doi: https://doi.org/10.31014/aior.1994.02.03.55.

[2] L. Delgado-Lobete et al., “Individual and Environmental Factors Associated with Tobacco Smoking, Alcohol Abuse and Illegal Drug Consumption in University Students: A Mediating Analysis,” International Journal of Environmental Research and Public Health, vol. 17, no. 9, p. 3019, Apr. 2020, doi: https://doi.org/10.3390/ijerph17093019.

[3] C. Tembo, S. Burns, and F. Kalembo, “The association between levels of alcohol consumption and mental health problems and academic performance among young university students,” PLOS ONE, vol. 12, no. 6, p. e0178142, Jun. 2017, doi: https://doi.org/10.1371/journal.pone.0178142.

[4] A. Steptoe, J. Wardle, T. M. Pollard, L. Canaan, and G. Jill. Davies, “Stress, social support and health-related behavior: A study of smoking, alcohol consumption and physical exercise,” Journal of Psychosomatic Research, vol. 41, no. 2, pp. 171–180, Aug. 1996, doi: https://doi.org/10.1016/0022-3999(96)00095-5.

[5] R. A. Krukowski, L. J. Solomon, and S. Naud, “Triggers of Heavier and Lighter Cigarette Smoking in College Students,” Journal of Behavioral Medicine, vol. 28, no. 4, pp. 335–345, Jul. 2005, doi: https://doi.org/10.1007/s10865-005-9003-x.

[6] P. Córdova, Patricia Gasser Gordillo, Hernán Naranjo Mejía, I. La, Alberto Grajeda Chacón, and Alberto Sanjinés Unzueta, “Academic stress as a predictor of mental health in university students,” Cogent Education, vol. 10, no. 2, Jul. 2023, doi: https://doi.org/10.1080/2331186x.2023.2232686.

[7] M. Rezapour, M. K. K. Niazi, and M. N. Gurcan, “Machine learning-based analytics of the impact of the Covid-19 pandemic on alcohol consumption habit changes among United States healthcare workers,” Scientific Reports, vol. 13, no. 1, p. 6003, Apr. 2023, doi: https://doi.org/10.1038/s41598-023-33222-y.

[8] Y. Shen et al., “Detecting risk of suicide attempts among Chinese medical college students using a machine learning algorithm,” Journal of Affective Disorders, vol. 273, pp. 18–23, Aug. 2020, doi: https://doi.org/10.1016/j.jad.2020.04.057.

[9] A. B. R. Shatte, D. M. Hutchinson, and S. J. Teague, “Machine learning in mental health: a scoping review of methods and applications,” Psychological Medicine, vol. 49, no. 09, pp. 1426–1448, Feb. 2019, doi: https://doi.org/10.1017/s0033291719000151.

[10] N. H. Gottlieb and L. W. Green, “Life Events, Social Network, Life-Style, and Health: An Analysis of the 1979 National Survey of Personal Health Practices and Consequences,” Health Education Quarterly, vol. 11, no. 1, pp. 91–105, Mar. 1984, doi: https://doi.org/10.1177/109019818401100105.

[11] N. C. Atuegwu, C. Oncken, R. C. Laubenbacher, M. F. Perez, and E. M. Mortensen, “Factors Associated with E-Cigarette Use in U.S. Young Adult Never Smokers of Conventional Cigarettes: A Machine Learning Approach,” International Journal of

Environmental Research and Public Health, vol. 17, no. 19, p. 7271, Oct. 2020, doi: https://doi.org/10.3390/ijerph17197271.

[12] "Impact of a Healthy Lifestyle on the Psychological Well-being of University Students - Pharmaceutical Research and Allied Sciences," Ijpras.com, 2021. https://ijpras.com/article/impact-of-a-healthy-lifestyle-on-the-psychological-well-being-of-university-students

[13] J. A. Epstein, K. W. Griffin, and G. J. Botvin, Prevention Science, vol. 3, no. 2, pp. 95–104, 2002, doi: https://doi.org/10.1023/a:1015479216401.

[14] S. Hassanpour, N. Tomita, T. DeLise, B. Crosier, and L. A. Marsch, "Identifying substance use risk based on deep neural networks and Instagram social media data," Neuropsychopharmacology, vol. 44, no. 3, pp. 487–494, Oct. 2018, doi: https://doi.org/10.1038/s41386-018-0247-x.

[15] M. S. Chow, S. H. L. Poon, K. L. Lui, C. C. Y. Chan, and W. W. Lam, "Alcohol Consumption and Depression Among University Students and Their Perception of Alcohol Use," East Asian Archives of Psychiatry, vol. 31, no. 4, pp. 87–96, Dec. 2021, doi: https://doi.org/10.12809/eaap20108.

[16] W. T. Islam, "Student Smoking, Alcohol, and Psychological Wellness Data," Mendeley Data, vol. 1, Nov. 2024, doi: https://doi.org/10.17632/7y66xb3gmc.1.

[17] E. Mortaz, "Imbalance accuracy metric for model selection in multi-class imbalance classification problems," Knowledge-Based Systems, p. 106490, Sep. 2020, doi: https://doi.org/10.1016/j.knosys.2020.106490.

[18] M. Chantry, T. Thornes, T. Palmer, and P. Düben, "Scale-Selective Precision for Weather and Climate Forecasting," vol. 147, no. 2, pp. 645–655, Jan. 2019, doi: https://doi.org/10.1175/mwr-d-18-0308.1.

[19] B. F. Arnold et al., "Optimal Recall Period for Caregiver-reported Illness in Risk Factor and Intervention Studies: A Multicountry Study," American Journal of Epidemiology, vol. 177, no. 4, pp. 361–370, Feb. 2013, doi: https://doi.org/10.1093/aje/kws281.

[20] Z. DeVries et al., "Using a national surgical database to predict complications following posterior lumbar surgery and comparing the area under the curve and F1-score for the assessment of prognostic capability," The Spine Journal, Feb. 2021, doi: https://doi.org/10.1016/j.spinee.2021.02.007.

[21] P. Ranganathan, C. S. Pramesh, and R. Aggarwal, "Common pitfalls in statistical analysis: Logistic regression," Perspectives in Clinical Research, vol. 8, no. 3, pp. 148–151, 2017, doi: https://doi.org/10.4103/picr.PICR_87_17.

[22] C. Ding, T.-Y. Bao, and H.-L. Huang, "Quantum-Inspired Support Vector Machine," IEEE Transactions on Neural Networks and Learning Systems, pp. 1–13, 2021, doi: https://doi.org/10.1109/tnnls.2021.3084467.

[23] E. K. Sahin, "Assessing the predictive capability of ensemble tree methods for landslide susceptibility mapping using XGBoost, gradient boosting machine, and random forest," SN Applied Sciences, vol. 2, no. 7, Jun. 2020, doi: https://doi.org/10.1007/s42452-020-3060-1.

[24] Z. Deng, X. Zhu, D. Cheng, M. Zong, and S. Zhang, "Efficient k NN classification algorithm for big data," Neurocomputing, vol. 195, pp. 143–148, Jun. 2016, doi: https://doi.org/10.1016/j.neucom.2015.08.112.

[25] P. Shunmugapriya and S. Kanmani, "Optimization of stacking ensemble configurations through Artificial Bee Colony algorithm," Swarm and Evolutionary Computation, vol. 12, pp. 24–32, Oct. 2013, doi: https://doi.org/10.1016/j.swevo.2013.04.004.

[26] F. Jimenez and M. Katzfuss, "Vecchia Gaussian Process Ensembles on Internal Representations of Deep Neural Networks," arXiv (Cornell University), Jan. 2023, doi: https://doi.org/10.48550/arxiv.2305.17063.

[27] J. Huang, Y.-F. Li, and M. Xie, "An empirical analysis of data preprocessing for machine learning-based software cost estimation," Information and Software Technology, vol. 67, pp. 108–127, Nov. 2015, doi: https://doi.org/10.1016/j.infsof.2015.07.004.

[28] F. Bolinski, N. Boumparis, A. Kleiboer, P. Cuijpers, D. D. Ebert, and H. Riper, "The effect of e-mental health interventions on academic performance in university and college students: A meta-analysis of randomized controlled trials," Internet Interventions, vol. 20, p. 100321, Apr. 2020, doi: https://doi.org/10.1016/j.invent.2020.100321.

[29] S. McKee, R. Hinson, D. Rounsaville, and P. Petrelli, "Survey of subjective effects of smoking while drinking among college students," Nicotine & Tobacco Research, vol. 6, no. 1, pp. 111–117, Feb. 2004, doi: https://doi.org/10.1080/14622200310001656939.

[30] R. Mathur and S. Swaminathan, "National ethical guidelines for biomedical & health research involving human participants, 2017: A commentary," Indian Journal of Medical Research, vol. 148, no. 3, p. 279, 2018, doi: https://doi.org/10.4103/0971-5916.245303.

[31] A. R. Anderson and B. J. Fowers, "Lifestyle Behaviors, Psychological Distress, and Well-Being: A Daily Diary Study," Social Science & Medicine, vol. 263, p. 113263, Aug. 2020, doi: https://doi.org/10.1016/j.socscimed.2020.113263.

The Impact of Smoking and Alcohol Consumption on University Students' Psychological Wellness with Machine Learning Approach

ORIGINALITY REPORT

20%	17%	10%	%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	dspace.daffodilvarsity.edu.bd:8080 Internet Source	8%
2	www.mdpi.com Internet Source	1%
3	internationalpubls.com Internet Source	1%
4	doctorpenguin.com Internet Source	<1%
5	data.mendeley.com Internet Source	<1%
6	Yongkui Li, Yan Liu, Jiansong Zhang, Lingyan Cao, Qinyue Wang. "Automated analysis and assignment of maintenance work orders using natural language processing", Automation in Construction, 2024 Publication	<1%
7	escholarship.org Internet Source	<1%