Predicting Student Stress and Smartphone Addiction using machine learning

By
Sumyia Akter
212-15-4171
&
Sadia Akter
212-15-4225

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

Dr. S M Aminul Haque
Professor & Associate Head

Department of Computer Science and Engineering
Daffodil International
University

Co-Supervised by

Mehadi Hasan
Lecturer

Department of Computer Science and Engineering
Daffodil International
University

DAFFODIL INTERNATIONAL UNIVERSITY
Dhaka, Bangladesh

May 14, 2025

APPROVAL

This Project titled **"Predicting Student Stress And Smartphone Addiction Using Machine Learning"**, submitted by Sadia Akter, ID No: **212-15-4225 &** Sumyia Akter, **ID No: 212-15-4171** 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 **14 May, 2025.**

BOARD OF EXAMINERS

Dr. Arif Mahmud(AM) **Chairman**
Associate Professor and Associate Head
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

14.5.25

Md. Sadekur Rahman(SR) **Internal Examiner**
Assistant Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Tapasy Rabeya(TRA) **Internal Examiner**
Sr. Lecturer
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Dr. Md. Manowarul Islam **External Examiner**
Associate 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 **Dr. S M Aminul Haque, Professor & Associate Head,** Department of Computer Science and Engineering, Daffodil International University. We also declare that neither this projectnor any part of this project has been submitted elsewhere for the award of any degree or diploma.

Supervised by:

Dr. S M Aminul Haque

Professor & Associate Head

Department of Computer Science and Engineering, Daffodil International University

Co-Supervised by:

14.05.2025

Mehadi Hasan

Lecturer

Department of Computer Science and Engineering, Daffodil International University

Submitted by:

Sumyia

Sumyia Akter

Student ID: 212-15-4171

Department of Computer Science and Engineering, Daffodil University

Sadia

Sadia Akter

Student ID: 212-15-4225

Department of Computer Science and Engineering, Daffodil International University

ACKNOWLEDGEMENTS

This journey would not have reached its fulfillment without the heartfelt support, love, and encouragement of many cherished individuals who walked alongside us through this endeavor. We hold deep gratitude in our hearts for everyone who has touched this project with their kindness and generosity.

First, we offer our sincerest thanks to the divine presence that has guided us with grace, filling us with strength and hope to bring this **Final Year Design Project (FYDP)** to completion with joy and purpose.

With all our hearts, we extend profound gratitude to **Dr. S M Aminul Haque, Professor & Associate Head and Mehadi Hasan, Lecturer** Department of Computer Science and Engineering, Daffodil International University, Dhaka, Bangladesh. Their boundless passion for computer science ignited our inspiration, while their unwavering patience, heartfelt mentorship, and uplifting encouragement breathed life into this work. Their thoughtful guidance, warm support, and belief in us—through every draft, every challenge, and every moment of doubt—have been the cornerstone of this project's success.

We are deeply thankful to the Head of the Department of Computer Science and Engineering for their compassionate support and for fostering an environment where our ideas could flourish. Our gratitude also flows to the faculty and staff of the Department of Computer Science and Engineering at Daffodil International University, whose kindness and encouragement helped us find our way.

To our dear course-mates at Daffodil International University, we owe a special thanks for the laughter, shared dreams, and spirited discussions that enriched this journey and reminded us of the joy in collaboration.

Above all, we hold immense love and appreciation for our parents, whose endless patience, unwavering belief, and boundless love have been our anchor and our light throughout this chapter.

ABSTRACT

Student stress and smartphone addiction have emerged as critical issues in contemporary academic environments, affecting mental health, academic performance, and overall well-being. This study explores the intricate relationships among behavioral factors, physiological indicators, and smartphone usage patterns using machine learning (ML) techniques. A total of ten regression models—Linear Regression, Decision Tree, Random Forest, Gradient Boosting, Support Vector Regressor (SVR), K-Nearest Neighbors (KNN), ElasticNet, XGBoost, LightGBM, and CatBoost—were evaluated for their ability to predict self-reported stress and addiction levels among students. Performance was measured using MSE, RMSE, R^2, and computational efficiency. Results revealed that CatBoost demonstrated superior performance for stress prediction, achieving the lowest MSE (1.634) and highest R^2 (0.793), while Linear Regression performed best for addiction prediction with the lowest MSE (0.377) and highest R^2 (0.954). Correlation analysis highlighted strong associations between high stress levels and poor academic performance ($r = 0.85$), reduced sleep duration ($r = –0.69$), and high smartphone dependency. Notably, nighttime phone usage, frequent device unlocks, and high notification counts were found to significantly influence both stress and addiction levels. Beyond model accuracy, this study provides a comprehensive impact analysis across societal, environmental, ethical, and sustainability dimensions. It emphasizes the urgent need for proactive strategies in educational and mental health domains to mitigate digital overdependence and stress. Furthermore, it advocates for sustainable research practices, including energy-efficient computing and privacy-centered ethical frameworks, aligning technological progress with social responsibility. The findings pave the way for targeted interventions through mobile applications and policy initiatives to enhance student well-being in an increasingly digital academic landscape.

Table Of Content

List of Figures

List of Tables

CHAPTER 1

Introduction

With the onset of the digital age, student pressure has been one of the main issues regarding, academic, social and lifestyle issues. The increased usage of a smartphone and mobile phones has been linked with increased stress, anxiety, and poor academic performance due to increased screen time. Stress is a psychological and physiological response to challenging circumstances that often begins with academic pressure, social expectations, and personal problems. As per the World Health Organization (WHO), stress is a natural reaction, However when its elevated, it can have an adverse effect on mental and physical well-being, including inducing anxiety and impairing sleep and cognition [1]. Smartphone addiction is a form of excessive and compulsive mobile phone use that affects daily activities such as academic performance, mental health etc. Research has shown that long periods with a smartphone in hand can lead to stress, emotional exhaustion and trouble with concentration, the latter compounding students' problems at school[2]. With smart phones becoming increasingly prevalent in student life, their overuse has been associated with increased stress, social isolation, and procrastination. As students experience more stress and become increasingly addicted to smartphones, the need for data-based solutions for early warning and intervention becomes more important. Predictive models using machine learning have analyzed students' behavioral patterns, screen time, academic workload, etc. to identify those at risk of high stress and subsequent smartphone addiction[3]. Using machine learning approaches to train predictive models for preventing the adverse impact of stress and smartphone addiction, this study aims to address the correlation between stress and smartphone addiction among the students, as well as use machine learning to mitigate the ill effects.

1.1 Introduction

The rising stress experienced by students has become a global educational issue because of school requirements and time constraints as well as social performance anxiety. The wide availability of smartphones causes students' stress to worsen because these devices function as both performance tools and distractors which lead to anxiety and addiction issues[4], [5]. Student smartphone addiction continues to rise resulting in negative effects that include higher stress along with feelings of despair and anxiety[6]. Machine learning has become a leading approach that detects mental health issues while helping to understand such problems. Sound predictions about student stress and smartphone addiction levels have been achieved through the application of decision trees alongside support vector machines and neural networks and their processing of sleep habits and academic workload and social contact and screen time features [7], [8]. Researchers now investigate one consolidated predictive model that combines stress along with smartphone addiction monitoring through analysis of sleep patterns combined with social media habits and mobile app activities per study[7]. The development of precise stress and smartphone addiction models remains challenging because of data quality issues together with privacy problems and complicated behavioral relationships between stress and smartphones [9]. The problems faced by machine learning do not diminish its hopeful potential to develop personalized treatments for student stress control and smartphone addiction management. [10].

Students handle growing academic responsibilities alongside social challenges resulting in stress that produces serious issues for their school mental health and performance. Recent studies have identified a major problem in mobile phone dependency because young people stay excessively long on social media applications while entertainment applications drive stress levels higher. Research studies have focused on evaluating student outcomes for each variable independently yet there is limited availability of complete forecasting models that unite all elements. The use of machine learning methods successfully determines smartphone addiction levels together with stress prevalence through behavioral measurements that incorporate screen time data and sleep duration as well as study responsibilities. The current modeling approach analyzes individual aspects separately from examining a combined examination of stress together with smartphone usage dynamics leading to predictions and solutions of imprecise outcomes. The research develops an enhanced machine learning framework which unites various elements between subjective stress results and screen time measurements with sleep duration metrics to establish tailored intervention methods for students dealing with stress symptoms and smartphone addiction [5], [7], [11]

1.2 Motivation

Modern life has experienced a fundamental transformation because of smartphones as these devices now dominate human communication and data access relationships. Students use smartphones as essential academic and social instruments yet widespread smartphone usage creates major mental health problems along with stress levels and behavioral addiction. Technological immersion which expands across all environments continues to merge productive behaviors with non-productive activities thus leading to environments that generate stress-related symptoms and decline mental health.

The major distressing effect from extensive technological engagement results in smartphone addiction which produces both compulsive behavior and anxiety during downtimes and withdrawal features. Students face exceptional risk because of their educational demands and developmental phase alongside social performance expectations. Multiple testimonies together with initial research work shows that excessive smartphone usage at night and continual alerts disturb natural sleep patterns and academic performance. Although we lack an advanced system that combines data analysis to measure the effects between device usage and student risks.

Student mental health presents a serious concern because their stress and feelings of burnout together with anxiety levels continue to increase. Multiple student problems remain interrelated because they develop from life patterns together with academic challenges as well as innovative technology usage practices. Traditional student stress evaluation relies primarily on survey-based research methods which limit frequency, scale, and objectivity of results acquisition. The detection and proactive solutions for these problems require comprehensive methods that provide stable performance alongside unbiased assessments.

The purpose behind this study arises from where technology converges with mental health and sustainability. Today's expanding machine learning technology enables the development of predictive models that combine technical competence with ethical responsibility to both healthcare and educational domains as well as human behavior analysis. The research applies ML models to actual behavioral and physiological data to merge data science strategies with student well-being practices. In addition to its other objectives the research tracks sustainable approaches for data utilization and cloud services along with mental health intervention methods.

The study finds its motivation through its goal to help educational institutions make decisions based on evidence. Employing precise prediction insights enables educational institutions to create specific behavioral treatments such as digital rest programs and counseling initiatives with awareness education that suits student requirements. Programs that depend on quality data and models produce timely solutions that achieve better results.

The main purpose behind this research work is to develop an all-encompassing AI solution for stress management and smartphone addiction observation specifically in student populations. This initiative continues its effort beyond being a mere technical implementation because it responds to an emerging public health and social challenge that aims to establish beneficial mobile device usage behaviors and boost educational accomplishments and promote enduring community wellness.

1.3 Objectives

Objectives of the research will be:

- To develop a machine learning-based predictive model for student stress and smartphone addiction, leveraging diverse datasets and real-time smartphone data for improved accuracy and generalizability.
- To evaluate multiple machine learning models and apply Pearson correlation analysis to identify the most effective approach for assessing stress-mobile usage relationships.
- To generate data-driven insights for personalized student well-being interventions, focusing on stress reduction, screen time management, and sleep improvement.
- To bridge predictive analytics with actionable intervention strategies, enhancing student welfare through evidence-based solutions.

1.4 Methodology

This study set out to understand how students' stress levels connect to their smartphone habits. We asked over 1,000 college students to complete an online survey over nine days in January 2025. The questions covered everything from how stressed they felt to how often they checked their phones, how well they slept, and how their phone use affected their schoolwork. Before crunching the numbers, we cleaned up the survey responses by removing duplicate answers, fixing any incomplete entries, and weeding out extreme outliers that might skew the results. We then divided the data, using most of it (80%) to train our computer models and saving the rest (20%) to test how well those models worked.

We tried out several different analysis methods to see which one could best predict the relationship between phone use and stress. Some were simpler approaches like basic trend analysis, while others were more sophisticated techniques that can spot complex patterns. We also looked at how strongly different factors - like late-night scrolling or constant notifications - were linked to students' stress levels. What makes this research particularly useful is that it combines both cold, hard numbers (like exactly how many times someone unlocked their phone) with personal experiences (how anxious or stressed students reported feeling). By tracking when responses came in, we could also see if stress or phone use tended to spike at certain times of day. The findings could help create better support systems for students - whether that's campus mental health resources, apps to promote healthier phone habits, or school policies that take these digital stressors into account. At its core, this was about using real student experiences to find practical ways to make college life a little less stressful in our always-connected world.

1.5 Project Outcome

- Predictive Model Development: An AI model will be created which can predict the level of stress and the tendency to become addicted to the use of smartphones among students based on various data sets and actual data which should increase the level of accuracy and generalization.
- Model Evaluation: Varying machine learning models will be analyzed and the Pearson correlation analysis will be conducted in order to decipher the best approach to understanding the correlation between stress and smartphone usage.
- Personalized Interventions: Data driven insights will be used to inform personalized interventions to reduce stress, manage screen time and improve sleep to promote student well being.
- Actionable Strategies: The use of predictive analytics will develop actionable intervention strategies that will provide evidence-based solutions for reducing stress and increasing smartphone utilization and sleep which will promote student welfare.

1.6 Organization of the Report

The report is laid out that it offers an in depth and structured discourse of research that has been conducted in predicting the stress amongst students and the addiction to smartphones based on machine learning techniques.

With CHAPTER 1, the Introduction, the research is introduced with the problem of student stress and smartphone addiction laid out at the outset. This CHAPTER also offers the key reasoning behind the study, its aims, approaches used and what is expected from this study. Furthermore, it contains a short overview on how the report is structured, which is a map for readers of the content that is presented after.

CHAPTER two, background, explores the existing literature and background of the study. It examines such applications and related research in fields of student stress smartphone addiction, providing an overview of what has been previously conducted and what remains to be further investigated. The CHAPTER ends with a gap analysis indicating where this research sits within the bigger picture, and where it contributes towards moving field forward.

CHAPTER 3; Research Methodology, the report describes the methodology used in the study. This CHAPTER outlines the data collection, but in the form of student surveys, and explains the preprocessing of the data collected to get it into usable form for analysis. It also gives a detailed description of the stipulated machine learning models used to predict stress and smartphone addiction level. The CHAPTER also provides an overview of functional and nonfunctional requirements for the research system and context and data flow diagrams to present the system as a structure with data processing procedures.

CHAPTER 4, Implementation and Results is concerned with the practical application of the research. It covers how the machine learning models were trained and tested, presenting the outcome of each model's prediction ability in forecasting student stress and addiction. A comparison of all the various models is also presented, giving a clear picture of how which models performed best and why. This CHAPTER is benchmark in showing the effectiveness of machine learning in this context.

CHAPTER 5 Engineering Standards and Design Challenges discusses what was followed in the research and the challenges encountered in the design and implementation phases. It touches on the socio-environmental implications of the study especially in relation to what it adds to mental health and wellness. Ethical concerns are also addressed, including responsible use of data and privacy protection.

Finally, CHAPTER 6 Conclusion summarizes the main results of the research, comments on the limitations of the study, and suggests possible directions of further investigation. This CHAPTER thus highlights the relevance of the work and indicates how the finding might be utilized in real-world settings to enhance student well-being. The report ends with an extensive list of references, which back up the research and give additional context on the study´s results.

CHAPTER 2

Background

2.1 Introduction

The excessive use of smartphones continues to grow as a shared problem among students in our current digital era which adds to their stress levels and contributes to addiction complications. Academic performance together with mental state and sleep quality suffer due to excessive screen time according to research reports[12]. Machine learning enables researchers to detect behavioral patterns in students so they can predict both stress and smartphone addiction conditions. This investigation establishes a systematic approach for fast recognition and assistance through a three-factor analysis between perceived emotional pressure and screen time lengths and learning results[13] . Using real-time data machines, learn today how to give insight into real time stress and addiction level. This opens up the possibility for personalised interventions to improve mental health and reduce smartphone addiction.

2.2 Literature Review

Current times present two major contemporary challenges which effect students through negative impacts on their academic achievement as well as their psychological state and routine operations. Experts believe that machine learning represents an effective solution to predicting such problems through data analysis of smartphone use patterns together with academic performance and behavioral behavior tracking. The prediction of student stress and smartphone addiction demonstrates high accuracy according to researchers who apply Support Vector Machines (SVM) along with Random Forests, Deep Neural Networks and Gradient Boosting algorithms. The application of machine learning systems together with real-time tracking and intervention strategies enables improved student mental health support from universities and professionals which prepares better conditions for an academic setting.

Singh et al. (2024)[14], in their study "Machine Learning Algorithms for Detecting Mental Stress in College Students" linked their research to the thesis topic about predicting student stress and smartphone addiction through machine learning methods. A total of 843 college students aged 18 to 21 answered a stress-related question during a survey process which received approval from AIIMS Raipur India specialists. A total of seven machine learning models analyzed the dataset that contained 28 features through Decision Trees (87%), Random Forest (90%), Support Vector Machine (SVM) (95%), AdaBoost (85%) and Naive Bayes (83%), Logistic Regression (88%), k-Nearest Neighbors (KNN) (80%). Support Vector Machine achieved the highest accuracy rate at 95% from all models tested for stress level assessment. The research shows that machine learning functions effectively to quantify stress but points out several constraints including human-originated survey responses as well as restricted multifaceted sensor analysis and difficulties in monitoring brief stress changes. The research dataset contains data from students within a restricted age range of 18-21 so it cannot provide practical benefits to students of different age groups. The considerable difference in both stress levels and stress reactions between age groups prevents researchers from effectively applying generalized conclusions. Adding a broader sample of different ages to the dataset could lead to better performance quality of the model because it will become robust to various age groups.

Arya et al. (2024) [15] conducted a research study named "Predicting the Stress Level of Students Using Supervised Machine Learning and Artificial Neural Networks (ANN)" at Tribhuvan University in Nepal to discover what factors affect student stress. Across the study 12 machine learning models joined by two deep learning models were utilized to analyze student stress levels into low medium or high tiers using Support Vector Machine (SVM), Random Forest (RF), Gradient Boosting (GB), AdaBoost, CatBoost, LightGBM, ExtraTree, XGBoost, Logistic Regression (LR) K-Nearest Neighbors (KNN), Naïve Bayes (NB) and Decision Tree (DT) together with Multi-Layer Perceptron (MLP) and ANN. Among all models Naïve Bayes proved to be the most accurate with 90% while Random Forest maintained 90% accuracy level and LightGBM along with ExtraTree achieved 89.09% accuracy. The test accuracy of SVM was found to be the lowest at 85.45%. The research team applied both hyperparameter optimization and cross-validation strategies to achieve better model effectiveness. Academic periods emerge as the peak period of stress for university students based on research. Constraints in study data collection regarding its sample size of 1,100 students in a particular area often limit the ability to generalize findings. The study depends on survey data collected from participants which may contain biases and the real-time measurement data from wearable devices remains excluded. Research development should use real-time stress detection systems alongside sensor inputs and sophisticated deep learning algorithms to achieve better prediction outcomes.

Research published by Lee and Kim (2021) [16] analyzed how well mobile phone log data could forecast problematic smartphone use through their study named "Prediction of Problematic Smartphone Use: A Machine Learning Approach." The study accessed data from 29,712 respondents who participated in the 2017 Korea Internet and Security Agency (KISA) survey as it contained basic demographics alongside smartphone usage statistics. The researchers utilized Decision Tree, Random Forest and XGBoost machines to learn algorithms to create binary classifications between high risk and normal smartphone users. The system achieved its best accuracy level at 82.59% with Random Forest while XGBoost reached 80.77% and Decision Tree performed at 74.56%. This research demonstrated that basic audience attributes including age and occupational group, and sex did not predict smartphone addiction but actual device use activities together with monetary expenditures proved more effective. A vital study shortcoming stems from using self-reported survey data instead of actual smartphone log data since this data collection method could affect the classification results.

The study conducted by Raj et al. (2024)[17] evaluated "Machine Learning Model for Prediction of Smartphone Addiction" through a machine learning approach to sort individuals into different smartphone addiction categories. A public database containing 5,000 records served as the basis for analysis while 19 variables pertaining to smartphone habits and social interactions and smartphone dependency characteristics were included in the dataset. The research implemented Decision Tree, Logistic Regression and Random Forest as machine learning models to predict addiction levels.. Users were categorized as "addicted," "not addicted," or "may be addicted." Random Forest had the best accuracy (82.59%), followed by XGBoost (80.77%) and Decision Tree (74.56%). The study discovered that users' general demographic characteristics, such as age, job classification, and gender, had little ability to predict for smartphone addiction, where as usage behavior and spending patterns were more significant predictors. However, one major limitation of the study is that it relies on self-reported survey data rather than actual smartphone log data, which may introduce bias into the classification. Furthermore, the study lacks real-time behavioral tracking and physiological markers, which could improve precision. Future studies could include real-time monitoring with mobile app data, advanced deep learning models, and multi-modal approaches to improve prediction of problematic smartphone use.

Shahapur et al. (2024) [18] done a research investigation titled "Decoding Minds: Estimation of Stress Level in Students Using Machine Learning", with the goal of predicting student stress levels using self-reported survey data and machine learning algorithms. The study used a Kaggle dataset with over 6,000 students and 20 different parameters, including anxiety level, depression, sleep quality, study load, peer pressure, and academic performance. To classify stress levels into three categories (low, moderate, and high), the authors used five machine learning models: logistic regression (89.46%), K-Nearest Neighbors (KNN) (92.8%), decision tree (94.5%), random forest (95%), and gradient boosting (90.15%). Within these, Random Forest were having the highest accuracy of 95%, proving its effectiveness in stress prediction. The research investigation recognized essential sources of stress which included substandard sleep quality in combination with both mental health heritage and academic performance. This study reviews several machine learning approaches yet excludes techniques from deep learning such as artificial neural networks (ANN) and recurrent neural networks (RNN) which may enhance prediction results.

Selvaraj et al. (2023) [19] presented "Machine Learning Algorithms for Analysis of Smartphone Addiction" which reviews distinctive machine learning procedures utilized for smartphone addiction detection along with analysis through their review study. The examined research utilizes Decision Tree along with Naïve Bayes and K-Means Clustering models together with Logistic Regression and Support Vector Machine (SVM) and Random Forest in addition to other models. Decision Trees demonstrated 89.7% accuracy levels in addiction identification according to the review however Naïve Bayes and K-Means algorithms showed varying results based on dataset characteristics. The evaluation explores how behavioral tracking along with survey reports and smartphone patterns and psychological inputs serve as predictor elements for diagnosing smartphone addiction. The research faces a major drawback because it depends on subjective survey responses that could introduce errors or bias into the results.

The findings lack broad applicability because most studies work with participant samples between 240 and 500. The review points out that real-time smartphone usage monitoring together with deep learning algorithms (such as neural networks) and merged sensor information lacks investigation in the field thus leading to potential research opportunities. Ongoing research into smartphone addiction needs to refine their use of increased sample sizes of various populations alongside deep learning modeling for achieving higher precision in real-time detection capacities.

Ding et al. (2023) in “A Systematic Hybrid Machine Learning Approach for Stress Prediction” reports a hybrid model for stress prediction.[20] . This study combined Random Forest (RF) and Gradient Boosting Machine (GBM) through soft voting ensemble prediction to compute results by using model probability scores.. Physiological characteristics like snoring range, body temperature, respiration rate, limb movement, blood oxygen levels, eye movement, and heart rate were all included in the dataset, which was acquired from Kaggle. While Random Forest and GBM each achieved 99% accuracy, the suggested hybrid model (HB) overcame individual models with a perfect accuracy of 100%. Support Vector Machine (SVM) and Logistic Regression (LR) performed at 95% and 96%, respectively. The model's stability was validated by a 10-fold cross-validation, and its significance over different approaches was confirmed by a statistical T-test. A small dataset (630 records), reliance on pre-existing research data rather than real-world collected samples, and the lack of deep learning techniques like neural networks—which could boost performance—are some of the study's limitations, despite its remarkable accuracy. By using multimodal data sources like wearable sensor data for ongoing stress monitoring and larger, real-time datasets, future research could improve generalizability.

Giraldo-Jiménez et al.[21] (2022) conducted a study on "Smartphone Dependency Risk Analysis Using Machine Learning Predictive Models", with the aim to assess smartphone addiction using machine learning techniques. The study involved 1,228 university students from a private institution in Cali, Colombia, who completed the Smartphone Dependency Test (SDT), a Risk Factors Questionnaire, and a musculoskeletal symptoms assessment. The study used six machine learning models: Decision Tree (70.5%), Logistic Regression (76.4%), Random Forest (76.6%), SVM with Polynomial Kernel (76.4%), SVM with RBF Kernel (77.2%), Multilayer Perceptron (73.6%), and TabNet (67.4%). —classify students based on their level of smartphone dependency. SVM with RBF Kernel had the highest accuracy (77.2%), followed by Random Forest (76.6%) and Logistic Regression (76.4%), with TabNet bringing the lowest (67.4%). The study found a strong link between smartphone dependency and musculoskeletal discomfort, particularly in the wrists, neck, and shoulders, emphasizing the physiological effects of excessive smartphone use. However, one major limitation is that it is based on self-reported data, which may introduce biases, and there is no real-time smartphone usage tracking. Furthermore, the study did not look at deep learning architectures, which could boost prediction accuracy. Future study should employ larger datasets, real-time behavioral tracking.

2.2.1 Similar Applications

There are different applications developed to track and predict the amount of stress and addiction in people, including students, in the area of mental health and smartphone addiction. These applications harness machine learning approaches and behavior data to generate real-time feedback and actions. Some noteworthy examples of similar applications that concern the issue either of predicting the stress or monitoring smartphone addiction, or both, include:

1. **Moment**: Moment is an app to track your mobile and screen time in your smartphone. It also helps users monitor their screen time and is aimed at reducing handing over to smartphone addiction. Moment tracks how much time users spend on their phones and generates weekly reports on their phone usage patterns: by app, with breakdown on where and when they spent time on their apps. Although it is not directly dealing with stress levels it manages the root cause of phone addiction by enabling the users to manage and minimize their screen time, this can minimize stress caused by excessive mobile consumption.
2. **Headspace**: Headspace is one of the more well-known mindfulness and newspaper apps aims to help uses get to the core of stress, anxiety, and rumpus better. It provides a number of guided meditations, breathing exercises, and practices and mindfulness intended to aid users in managing the tension of a normal day and maintaining a mentally healthy level. In addition to mood tracking and habit tracking, the app also has sleep aids and relaxation exercises to assist students manage their stress better. Although Headspace does not focus on the specific issue with smartphone addiction, by utilizing the data from smartphones, its performance can be increased. Headspace could give users a more all-encompassing approach to get rid of both stress and mobile addiction by combining knowledge on phone behavior with mindfulness techniques.

2.2.2 Related Research

Table 2.1: Summary of the literature review

Authors	**Dataset**	**Methods Used**	**Results**
Singh et al. (2024)	843 students survey data (28 features)	SVM, Random Forest, AdaBoost, Naive Bayes, Decision Tree, KNN	SVM 95%, Random Forest 90%, AdaBoost 85%, Naive Bayes 83%
Arya et al. (2024)	Unknown dataset, student stress predictio	Naïve Bayes, Random Forest, SVM, Logistic Regression	Naïve Bayes 90%, Random Forest 90%, LightGBM 89%

Lee and Kim (2021)	KISA dataset (29,712 participants)	Decision Tree, Logistic Regression, SVM	Random Forest 82.59%, Logistic Regression 80.77%, Decision Tree 74.56%
Raj et al. (2024)	5,000 records of smartphone usage	Decision Tree, Logistic Regression, Random Forest	Random Forest 89%, Decision Tree 86%, Logistic Regression 74%
Shahapur et al. (2024)	Over 6,000 students from Kaggle	Logistic Regression, KNN, Decision Tree, Random Forest, Gradient Boosting	Random Forest 95%, KNN 92%, Decision Tree 94%
Selvaraj et al. (2023)	Varied dataset from multiple studies	Classical and deep learning models: Logistic Regression, Random Forest, SVM	Best performance by Random Forest, SVM, Logistic Regression
Ding et al. (2023)	Kaggle dataset (physiological data)	Gradient Boosting, Random Forest, XGBoost, SVM, AdaBoost, etc.	Best performance by SVM, Random Forest (above 91% sensitivity)
Giraldo-Jiménez et al. (2022)	1,228 university students from Cali, Colombia	Logistic Regression, Random Forest, SVM, Multilayer Perceptron, TabNet	Random Forest 76.6%, SVM (RBF) 77.2%, Logistic Regression 76.4%

2.3 Gap Analysis

Table 2.2: Gap Analysis

Researcher	Singh et al.	Lee and Kim	Arya et al. (2024)	Chung and Teo	Kavosh et al.	Thomée et al.
Integration	No	No	Yes	No	Yes	No
Model Consensus	Yes	Yes	Yes	Yes	Yes	Yes
Behavioral Data	Yes	Yes	Yes	Yes	Yes	Yes
Short-Term vs Long-Term	No	No	Yes	Yes	Yes	No

Usage Patterns	No	No	Yes	Yes	Yes	Yes
Societal Impact	No	No	Yes	Yes	Yes	Yes

2.4 Summary

This thesis reviews rising attention regarding student stress and smartphone dependence which emerged during the digital era. The damaging effects of excessive smartphone usage result in academic underperformance and mental health issues as well as sleep disturbances according to various research that call attention to early identification and therapeutic intervention. Through the analysis of behavioral patterns alongside academic performance records and smartphone utilization specialists have demonstrated the ability to forecast such concerns. Research studies demonstrate different prediction accuracy levels regarding student stress and addiction tendencies while using Decision Trees, Support Vector Machines, and Random Forests algorithms. The thesis adopts previous research methods built on machine learning which helps study and forecast student stress levels and smartphone addiction while providing prospective prevention support.

CHAPTER 3

Research Methodology

3.1 Methodology

3.1.1 Overview

The research sought to determine the link between stress and mobile phone addiction in student populations. Students completed a Google Form that asked about mobile phone addiction symptoms along with stress measurements. The information collection process involved removing unneeded columns and handling outliers to achieve high predictive accuracy in the data. A set of machine learning algorithms assessed the dataset for mobile phone addiction prediction and stress level assessment. We evaluated model effectiveness through preprocessed data testing and established the most accurate prediction method by assessing their performance.

3.1.2 Proposed Methodology

The proposed methodology is trying to find a method to predig student stress and smartphone addiction in using a machine learning. Such processes include a series of key steps, like data collection, training and evaluation. The approach is intended to guarantee that the predictive models are both correct and applicable in different student populations.

1. **Data Collection:** Data will be gathered on-line via an internet-based survey with 1,042 college student members over the course of 9 days. The survey is meant to collect knowledge on stress levels, mobile device utilization, academic workload, sleep cycles, among other behaviors.
2. **Data Preprocessing:** Data cleaning will require dealing with missing values, remove outliers, and discard unrequired fields. Phenotype-genotype correlation using Pearson correlation analysis will enable selection towards the most important features for modeling.
3. **Model Selection:** Supervised machine learning models of Linear Regression, Decision Tree, Random Forest, Gradient Boosting, Support Vector Regressor, KNN, ElasticNet, XGBoost, LightGBM, CatBoost will be implemented to forecast stress and smartphone addiction levels.
4. **Model Training and Evaluation:** The data will be used 80% for the training phase while 20% for testing phase. Models will be ranked based on MSE, R^2, Precision, Recall, and F1-Score for the assessment of prediction accuracy.
5. **Correlation Analysis:** Pearson correlation will examine relationships between smartphone utilization patterns, stress ratings, and addictive behaviors and also give information to essential taking part variables.

6. **Model Optimization:** Cross-validation and hyperparemeters tuning was used to boost the performance of the model or prevent overfitting.

3.1.3 Functional and Nonfunctional Requirements

Functional Requirements: The system should have functional requirements which will enable the collection of data from the student through online survey answering questions to do with smartphone use, stress experience, study and leisure time, sleep patterns as well as any other relevant behavioral factors. After the collection of data, it must be preprocessed to deal with the missing values, to remove the outliers, to eliminate the such columns. Machine learning techniques such as Linear Regression, Decision Trees, Random Forest and Gradient Boosting is going to be used to predict student stress and smartphone addiction. Pearson correlation analysis will be done to determine associations between smartphone use and stress. Further, the system has to assess the model's performance based on metrics like MSE, R^2, Precision, Recall and F1-Score. Must delivered visualizations to assist in suddennessness the divorce, as well as, partly upon the divorce, the scheme inch automatic ought to generate partes a conventional further assisted remembering to alleviate having chez putting across behaviours, applying by judging off online high level of fibre nous later aim for reduced screen inner ordinate et stay dim.

NonFunctional Requirements: Performance, scalability, reliability, security, usability and compatibility are the non-functional requirements of the system. The system must be able to handle big data quite quickly, with fast completion of machine learning model training and evaluation for real-time prediction. It should be able to scale up to support ever-more data and new data feeds. The system has to be robust ensuring good forecasts and steadiness performance under diverse conditions. Data privacy & security matters and system must get along with data protection regulations such as GDPR, encryption & access control. Usability is key and the system must have a very user-friendly interface so it is easy to navigate. It should also be maintainable with modular Subsystem, allowing simple updates or extensions. It must be compatible with popular operating systems and also integrate with other tools for data collection like Google Forms.

3.1.4 Context Diagram

A Context Diagram is a structural image that describes the system's limits and its interactions with external units, such as users, other systems, or database. It presents it as a single process, thus how data travels between the system and these external actors. The diagram emphasizes inputs and outputs, but no inner processes are mentioned. It is to define the system's scope, identify external stakeholders and give a clear understanding of what system interacts with what. This diagram is usually prepared during the initial stages of system design for the sake of clarifying its general composition.

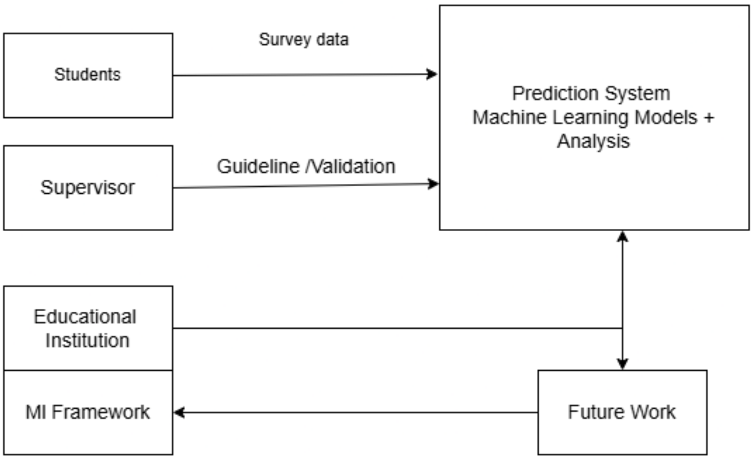

Figure 3.1: Context Diagram for Predicting Student Stress and Smartphone Addiction

3.1.5 Data Flow Diagram Level 1

The Data Flow Diagram (DFD) is used in illustrating the flow of data within a system, presenting a picture of the processes of the handling of information. It usually contain aspects such as data sources, process, data storage and outputs, and arrows to show data flow between these elements. In the given diagram, data begins with Data Collection in which raw data is collected.

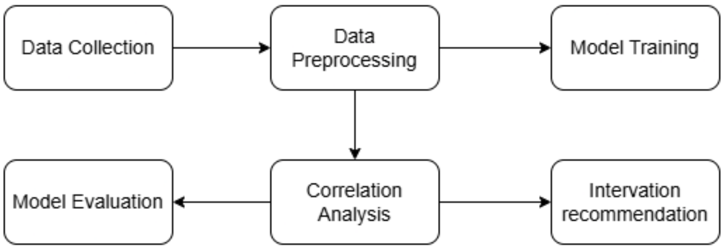

Figure 3.2: Data flow diagram for Predicting Student Stress and Smartphone Addiction

It then goes into Data Preprocessing where the data is cleaned and ready for analysis. Model Training is the next step where algorithms are deployed to develop predictive models. The flow moves to Model Evaluation, where the performance of the trained models is evaluated. After that, the Correlation Analysis deals with the issues of relationships between variables. Lastly, the results result to Intervention Recommendations wherein insights are applied to come up with suggestions for improvement or those that should be done.

3.2 Detailed Methodology and Design

This research established predictions regarding both student-based cell phone addiction and their related stress levels. A thorough cleaning process was applied to the data which involved the elimination of extra columns as well as duplicitous rows and null values before handling outlying data points. Consequently, the dataset became prepared for further analysis.

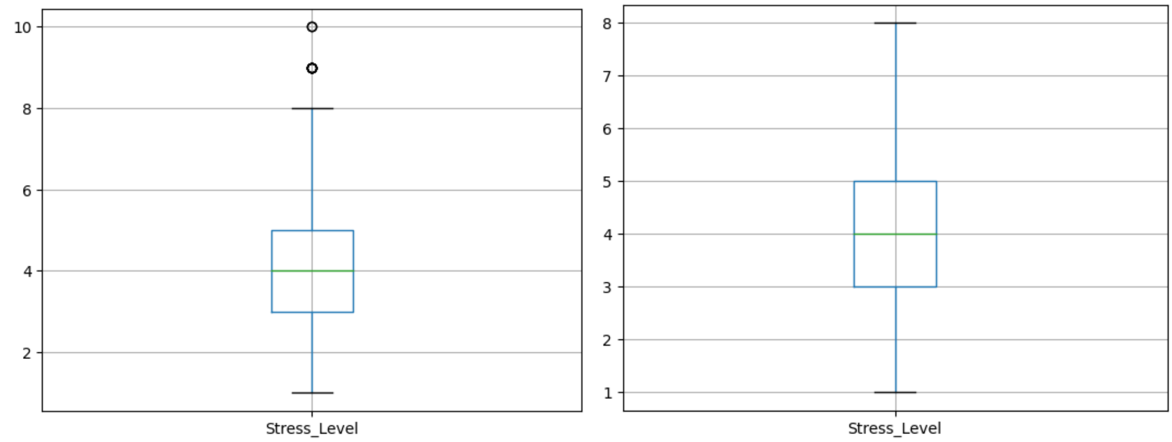

Figure 3.3: Removing Outlier

A division of the data into two parts was performed where 80% went to model training and 20% remained for testing purposes.

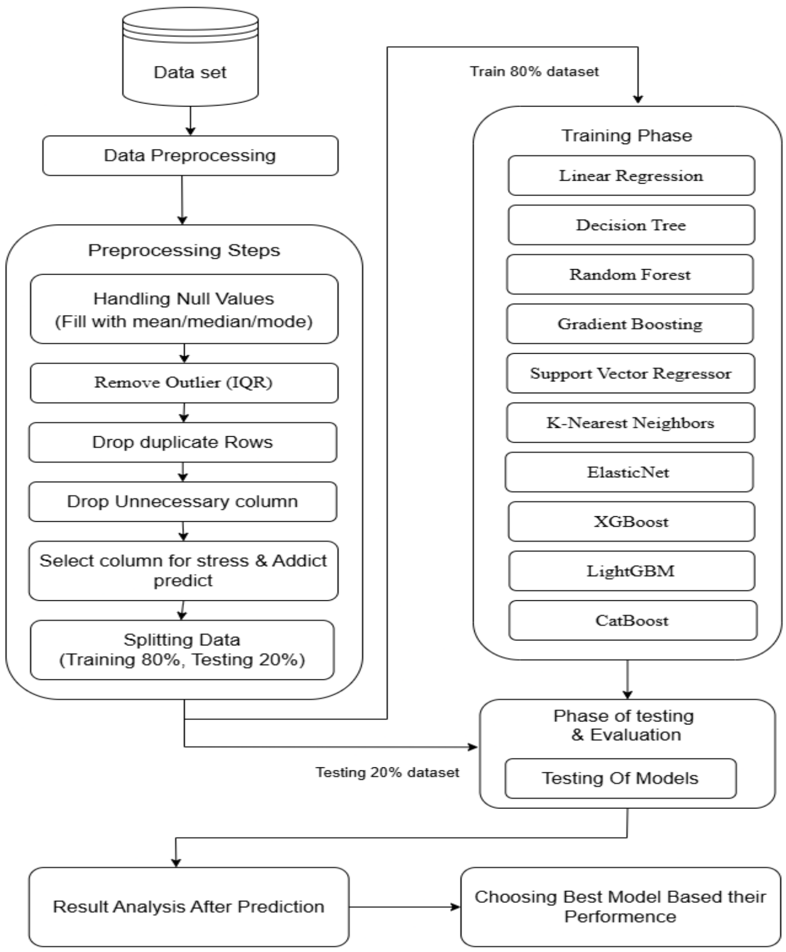

Figure 3.4: Methodology for Predicting Student Stress and Smartphone Addiction

Multiple models from machine learning served as choices such as Linear Regression Mode, Decision Tree Model, Random Forest Model, Gradient Boosting Model, Support Vector Regressor.[22], Mode, K-Nearest Neighbors Model, ElasticNet Regression, Model, XGBoost Regressor Model and LightGBM Regressor Model and CatBoost Regressor Model for analysis.

The models received training through the training dataset before testing occurred through the evaluation of their performance with the testing set. The analysis employed Pearson correlation [23] to study the link between mobile addiction and stress levels in order to identify their relationship patterns. The research outcomes from every model received analysis to validate which model proved superior in mobile phone addiction detection and stress level connections.

Dataset Description

The research collected 1,042 responses from students over 9 days starting on January 1st to January 9th, 2025, to study digital behavior patterns in relation to academic stress and mental health. All data points present three items that let students describe their activities along with their current stress levels and smartphone usage behaviors.

3.2.1.1 Demographic and Background Information:

i. Age (20–27): Participant age
ii. Gender (Male/Female)
iii. Is_Student (Yes/No): Confirms student status

3.2.1.2 Stress & Mental Health Indicators:

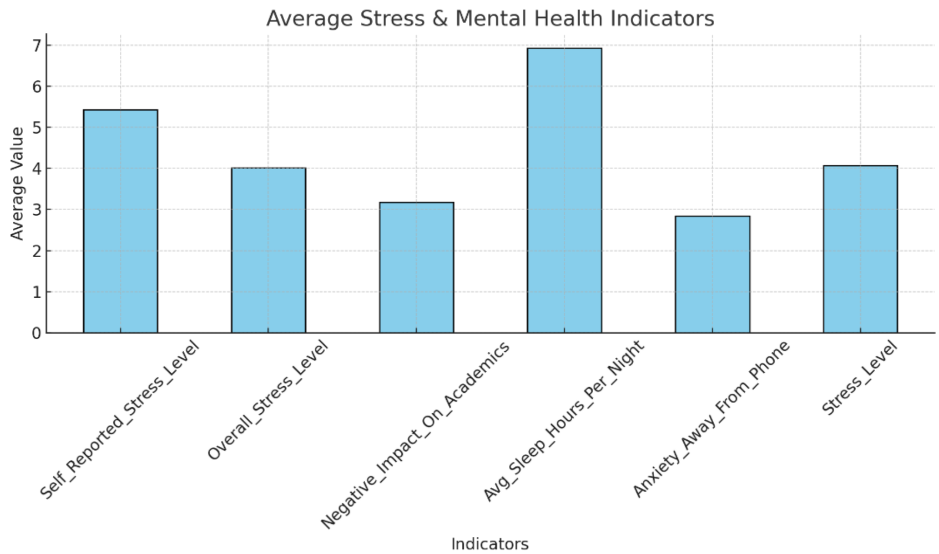

Figure 3.5: Stress & Mental Health Indicators

3.2.1.3 Smartphone Usage & Addiction Metrics

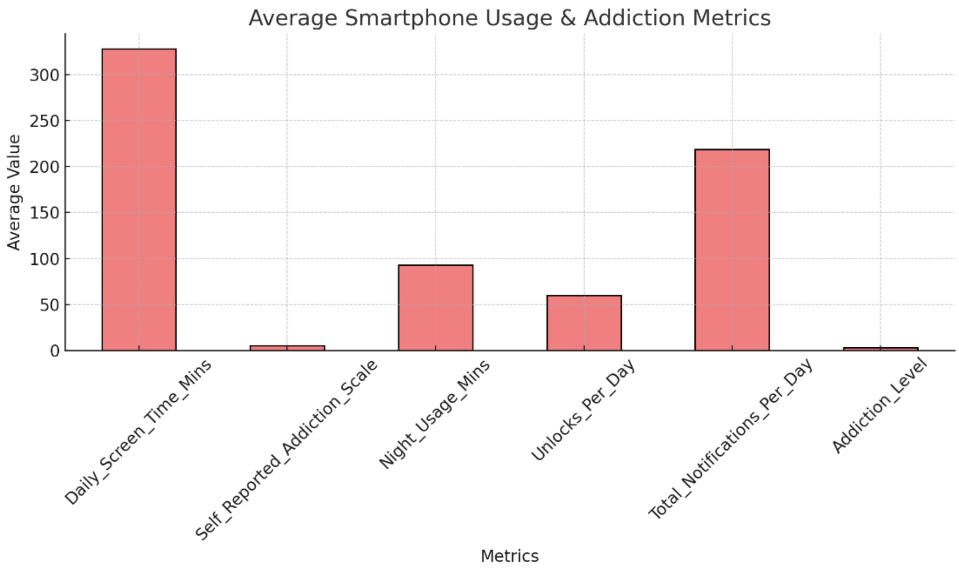

Figure 3.6: Smartphone Usage & Addiction Metrics

3.2.1.4 Behavioral & Contextual Data

i. Attempts_To_Reduce_Usage

ii. Timestamp: Date and time of response

The database holds significant practical value because it records actual student challenges they face regarding their internet usage and academic tension. Time-stamped information enables research analysis to discover specific time patterns between phone utilization and stress measurements at individual hours or daily periods. Students gain a whole view of their academic experience through the dataset which combines objective behavioral logs alongside subjective self-reported metrics. The analyzed data can help develop purpose-specific mental health support programs as well as create digital health tools while serving as a reference point for new academic policies to promote student mental well-being in digital education environments.

Ethical Consent

The study was preceded by the procedure of ethical consent. The dataset and questionnaire were validated by consulting a licensed psychologist. The process involved was in such a way that all ethical aspects were adhered to, and the data garnered conforms to established psychological norms. The research has respected ethical guidelines for the protection of participants' rights and for the credibility of the research.

3.2.1.5 Hardware Or Software Requirement

Hardware Requirements:

1. Graphics Processing Unit (GPU): Kaggle offers free access to NVIDIA Tesla P100 GPUs, which can significantly accelerate model training, especially for deep learning applications [24]

2. Central Processing Unit (CPU**):** Each notebook session is allotted four CPU cores, allowing for efficient data processing and model training operations [25].
3. Random Access Memory(RAM): A standard amount of 16 GB of RAM is offered each notebook session, allowing for the handling of large datasets and sophisticated computations. [25]
4. Storage: Each notebook session gets 20 GB of auto-saved storage space in the /kaggle/working directory, which is ideal for storing datasets, model outputs, and intermediate files [25].

Software Requirements:

1. Python[26] is a common programming language for machine learning, data analysis, and modeling.
2. Python packages like Pandas [27], Numpy [28], Matplotlib [29], and Seaborn [30] are critical for training machine learning models, data manipulation, and visualization.
3. Kaggle notebooks [31] and google collab [32] were used because they provide an easy-to-use platform for data analysis, preprocessing, and machine learning model training, particularly given the availability of cloud computing resources.
4. PyCharm [33] and Vscode [34] increase productivity by providing enhanced code development, debugging features, and a diverse set of plugins and extensions for a more efficient coding experience.

3.3 Project Plan

Phase 1: Initiation of project and collection of data (weeks 1-3). The project will start with the design and distribution of a survey, in order to collect data on students' levels of stress, smartphone use, academic burden, etc. – other behavioral factors. This stage will take three weeks, during which the survey responses will be collected and the ethical matters taken into consideration, e.g. the informed consent.

Phase 2: Data Preprocessing and Exploration During this stage, the data obtained will be preprocessed, where missing data will be accommodated, duplicate data removed, and outliers eliminated. Data exploration and visualization will be carried out to get an idea about the distribution of this dataset and key features. The clean and cleaned data will be ready for use to train models.

Phase 3: Model Development and Training During this stage, various AI models will be installed and trained with the preprocessed data. Such models as Linear Regression, Decision Trees, Random Forest and Gradient Boosting will be choosen for training. The models will be assessed on metrics like MSE, R^2 and F1-Score and improvements will be made through tuning them.

Phase 4: Model Testing and Evaluation This phase requires testing the trained models using another separate test dataset so as to evaluate their performance. To increase the accuracy and reliability of the models, cross-validation and hyperparameter optimization will be carried out. The ending performance metrics will be calculated for the purpose of identifying the best model for predicting a student's stress and smartphone addiction.

Phase 5: Final Review and Reporting The last part will involve the analysis of the models' results, interpreting them, and writing the thesis. The report shall contain the methodology, results and conclusions from the research. After making alterations and being finalized, the finished thesis will be submitted.

3.4 Task Allocation

To make workflow issues efficient, the division of tasks is made as follows:

- Data Collection: Sumyia Akter will design and distribute the survey to gather information regarding stress levels, smartphone usage and academic load.
- Data Preprocessing: Sumyia Akter will clean the dataset, fill missing values, drop outliers, and pick the features relevant to stress and addiction prediction.
- Model Development and Training: Sadia Akter will train and implement several machine learning models ranging from Linear Regression to Random Forest and fine-tune the models.
- Testing and Evaluation: Sadia Akter will try out the models on the 20% test dataset and determine their capacity to perform with relevant metrics such as MSE and R^2.
- Result Analysis: Sadia Akter will analyze the results, choose the best model, and use its findings for drawing conclusions regarding stress and palmaphobia.
- Report Writing: Sumyia Akter will be a coordinator for the preparation of the report by maintaining consistency and appropriate formatting in the final document.

3.5 Summary

The research investigates the connection between student mobile phone addiction and their stress levels. Research used responses from 1,042 students to evaluate stress and addiction levels through Support Vector Regression and Gradient Boosting and Random Forest machine learning models. The research demonstrates through empirical evidence that mobile phone addiction strongly correlates with increased student stress levels thus reinforcing the necessity to establish treatment methods for this problem.

CHAPTER 4

Implementation and Results

In this chapter, the application and assessment of machine learning approaches for predicting mobile phone addiction among students based on their stress levels are examined. A complete list of the analytical models employed in this research establishes Linear Regression, Decision Tree, Random Forest, Gradient Boosting, Support Vector Regressor, K-Nearest Neighbors, ElasticNet, XGBoost, LightGBM and CatBoost. This section completely probes each model as it investigates their unique characteristics and justifies their adoption for the prediction task.

The paper provides detailed information about the execution process which covers both data readiness and training algorithms alongside parameter refinement activities. The model performance improves through the implementation of cross-validation together with grid search protocols. The assessment of models happens through a combination of Mean Squared Error (MSE) alongside R-squared and Precision, Recall and F1 Score to evaluate predictive abilities thoroughly. The result analysis shows clear advantages and disadvantages among the algorithms which help establish their practicality for student cell phone addiction and stress forecasting

4.1 Train Model

The section presents an in-depth review of the execution process for different machine learning models utilized in this research work

Linear Regression: Linear regression stands as a statistical process which uses linear equations to represent how dependent variables relate to their multiple independent variables based on observed data [35] Linear regression defines the two-variable relationship through a single line pattern, as indicated by the equation:

$$\mathbf{y = \beta_0 + \beta_1 x + \varepsilon} \qquad \text{...... (i)}$$

y represents the dependent variable.

x represents the independent variable.

When the independent variable (x) is set to zero the y-intercept term (β_0) shows the predicted outcome value for y.

The change in the y variable corresponds to β_1 for every x variable increase of one unit.

The error term ε shows the portion of y unpredictability which cannot be explained through the linear correlation to x.

Linear regression locates its best β_0 and β_1 parameter values by reducing the difference of squares between actual and estimated output values. Analysis using this method is routinely applied in predictive modeling to determine variable relationships because it depends on linear variable connections.

Decision Tree: The supervised learning algorithm named Decision trees{Citation} enables both classification tasks and regression problems. The prediction process runs on criteria based on features. They partition data recursively into sections that use feature data to generate homogeneous groups that pertain to the target variable. The algorithm selects at every node the best feature and threshold which optimizes data splitting through considerations of Gini impurity and information gain. The calculation of Gini impurity index for classification follows this formula

$$\mathbf{Gini(D) = 1 - \Sigma\ (p_i)^2} \quad \text{...... (ii)}$$

The value p_i represents the probability for assigning an element into a particular class. Decision trees in regression create predictions about continuous values through minimizing the subset variance. Decision trees are simple models with excellent interpretability but show high susceptibility to overfitting especially when working with complex datasets. To address this problem pruning approaches together with tree depth restrictions serve as solutions to enhance model generalization capabilities. The objective of this research is to understand mobile phone addiction patterns by using decision trees which analyze mobile phone addiction relationships with student stress levels.

Random Forest: During training Random Forest utilizes multiple decision trees to build up its prediction capabilities while halting overfitting from taking place [36]. A bootstrapped subset of data trains each forest tree while randomization occurs when selecting subset features for node splitting. Through the bagging procedure models aim to lower their variance levels while avoiding additional bias growth. The regression prediction model calculates its response as the average output between all single-tree forecasts [37]:

$$\mathbf{f(x) = (1/B) * \Sigma\ f\_b(x)} \quad \text{...... (iii)}$$

The model considers B trees while f_b(x) represents the prediction from the b-th tree when processing input x. During classification problems the final prediction outcome emerges from the collective vote of all trees in the ensemble. Random Forest improves model robustness and accuracy by aggregating predictions from several uncorrelated trees [36], [37] .

Gradient Boosting: Gradient Boosting is an ensemble learning strategy that develops predictive models successively, with each new model attempting to repair the faults of the previous one. This strategy combines numerous weak learners, usually decision trees, to produce a strong predictive model [38]. At each iteration, the method applies a new model to the current ensemble's residual errors (the discrepancies between observed and anticipated values). The update rule for the model at iteration m is as follows:

$$\mathbf{F_m(x) = F_{m-1}(x) + \eta * h_m(x)} \quad \text{...... (iv)}$$

Where:

$F_m(x)$ represents the model's forecast at iteration m.$F_{m-1}(x)$ represents the prediction from the preceding iteration [39].The learning rate (η) determines the contribution of new models.The new model, $h_m(x)$, is fitted to the residual data.

This iterative procedure continues until a predetermined number of models have been added or no further improvements are possible [40]. Gradient Boosting is commonly utilized for regression and classification tasks because of its high prediction accuracy. However, it is sensitive to noisy data and can overfit if not correctly tuned.[41]

Support Vector Regression: Support Vector Regression (SVR) is a supervised learning technique based on Support Vector Machines (SVM) that predict continuous values. SVR seeks to identify a function that approximates the relationship between input variables and a continuous target variable while minimizing prediction error[42] [43]. SVR works by mapping input data into a high-dimensional feature space with a kernel function, then generating a hyperplane that best matches the data within a defined margin of tolerance (ε). This margin creates a tube in which errors are permitted [22], [42], [43]. Mathematically, the SVR model aims to minimize the following objective function:

$$\mathbf{(1/2)\ ||w||^2 + C\ \Sigma\ (\xi_i + \xi_i^0)} \qquad \textbf{...... (v)}$$

subject to the constraints:

$$\mathbf{y_i - (w \cdot \varphi(x_i) + b) \leq \varepsilon + \xi_i}$$

$$\mathbf{(w \cdot \varphi(x_i) + b) - y_i \leq \varepsilon + \xi_i^0} \qquad \textbf{...... (vi)}$$

$$\mathbf{\xi_i,\ \xi_i^0 \geq 0}$$

Where:

W and B define the hyperplane.

In scikit-learn, $\varphi(x_i)$ indicates how input x_i is mapped into the feature space.[44]

y_i represents the actual goal value.

ξ_i and ξ_i^0 are slack variables that allow for errors within the ε margin.

C is a regularization parameter that governs the trade-off between attaining low errors on training data and reducing model complexity.

SVR solves this optimization problem by finding a function that generalizes well to previously unknown data while balancing model complexity and training data fit.

K-Nearest Neighbors: K-Nearest Neighbors is a supervised learning technique used for classification and regression. It works on the concept that related data points are close together in the feature space. Given a new data point, K-NN determines the 'k' training samples closest to it, often using distance measures such as Euclidean distance, defined as :

$$d(x, z) = \sqrt{\Sigma}\ (x_i - z_i)^2 \qquad \textbf{...... (vii)}$$

The feature values x_i and z_i correspond to the data points x and z, respectively. [45]

In classification, the algorithm assigns the majority class among these 'k' neighbors to the new data point. In regression, the prediction is the average of the target values of the 'k' nearest neighbors. The choice of 'k' is crucial; a small 'k' can be sensitive to noise, while a large 'k' may smooth over distinctions between classes. Additionally, feature scaling or normalization is important to ensure that all features contribute equally to the distance computation, especially when they have different units or scales[45]. K-NN is non-parametric, which means it does not take a fixed form for the underlying data distribution and generates predictions based on the full training dataset. However, as the dataset grows, the computational cost increases since the system must calculate distances to all training samples for each prediction[46].

Elastic Net**:** Elastic Net, a regularized regression technique, combines L1 (lasso) and L2 (ridge) penalties to improve predictive accuracy and promote group selection in models with linked data. The objective function of Elastic Net is:

$$(1/2n)\ ||y - X\beta||^2_2 + \lambda_1||\beta||_1 + \lambda_2||\beta||^2_2 \qquad \textbf{...... (viii)}$$

Where:

n is the number of observations.

y is the response vector [47]

X is the predictor matrix.

β is the coefficient vector.

The regularization parameters λ_1 and λ_2 determine the influence between L1 and L2 Penalty terms as both parameters are non-negative.

The simultaneous use of variable selection and regularization functions identifies Elastic Net as highly beneficial for processing data with multiple predictive variables and connected elements. The Elastic Net model provides two purpose-based operations where it works as lasso when $\lambda_2 = 0$ and as ridge regression when $\lambda_1 = 0$ thus enabling detailed control for complex dataset modeling.

XGBoost**:** XGBoost represents an open-source machine learning tool which enhances gradient boosting performance for supervised learning operations especially regression and classification. XGBoost constructs its models one after another by adding new layers that try to fix errors from earlier models while making use of decision trees as basic learners. XGBoost includes a combination of loss function and complexity penalty terms in its objective function for achieving accurate predictions while preventing overfitting. A combination of L1 (Lasso) and L2 (Ridge) penalties exists in the regularization term to enhance generalization. The loss function approximation through second-order Taylor expansion in XGBoost achieves quick gradient and Hessian computations when performing optimization. Due to its architectural design XGBoost supports distributed computing processes across multiple computing environments. The efficiency combined with performance qualities make XGBoost the preference for machine learning contests and real-world applications..[48], [49], [50]

LightGBM**:** Supervised learning problems require the gradient boosting framework LightGBM as a superior implementation. The approach creates leaves one at a time through a process that targets the most loss-stricken leaves for precise modeling. LightGBM comprises an objective function that unites predictive accuracy measurement through a convex loss function together with regularization to stop overfitting:

$$\textbf{Objective = Loss Function} + \boldsymbol{\Omega}\textbf{(f)}$$

The term Ω(f) contains both L1 (Lasso) penalty and L2 (Ridge) penalty.

$$\boldsymbol{\Omega(f) = \lambda_1\|w\|_1 + \lambda_2\|w\|^2_2} \qquad \textbf{...... (ix)}$$

The weights of the model use w while the regularization parameters λ_1 and λ_2 control the combination between L1 and L2 penalties. The regularization framework helps users control model complexity and enhance generalization through its implementation.

CatBoost: Machine learning professionals can use CatBoost as an open-source gradient boosting system for dealing with categorical information[51]which Yandex developed. This framework creates decision trees one by one to address remaining mistakes introduced by previous trees through gradient boosting procedures[52]. The algorithm builds its decision trees one after the other to address errors introduced by former trees through gradient boosting[53]. Target statistics provides CatBoost with the capability to transform categorical values into numerical data points according to target variable points thereby decreasing model overfitting and enhancing prediction accuracy [54]. The ordered boosting mechanism in CatBoost uses permuted methods to decrease target leakages preventing prediction errors and producing unbiased gradient estimations. The training process usually chooses Root Mean Squared Error (RMSE) as its loss function to optimize the objective function [55], defined as:

$$\textbf{RMSE} = \textbf{sqrt}\ ((\Sigma_{i=1}^{n} w_i * (a_i - t_i)^2) / \Sigma_{i=1}^{n} w_i) \qquad \textbf{...... (x)}$$

Where:

The symbol n represents the count of observations while w_i represents the weight value and ai indicates the predicted value for observation i which corresponds to the target value ti.

The weight of observation number i is denoted as w_i.

The a_i value denotes the anticipated measurement of the i-th observation.

The actual target value for the i-th observation equals t_i.

The defined formula helps assess the predictive accuracy of the constructed model. The high efficiency and accuracy of CatBoost allows its applications in search engines together with recommendation systems and weather forecasting tasks.

4.2 System Testing and Model Evaluation

The prediction of student stress and smartphone addiction through regression models occurred within this stage by analyzing standardized dataset variables. The analysis incorporated multiple regression models that consisted of linear regression as well as decision tree regression and random forest regression and gradient boosting regression and other methods. The preprocessing applied StandardScaler to the data and used an 80-20 separation between train and test for objective predictions on both metrics through 5-fold cross-validation during training phases. The fit() method served to train the models which were assessed by evaluating MSE and R^2 Score. Evaluation of performance required test set predictions that led to calculating the residual errors. Linear Regression along with Decision

Tree employed default parameters as their hyperparameters yet Random Forest and Gradient Boosting employed GridSearchCV tool to optimize their elements including estimators and learning rates and tree depth. The created function conducted MSE and R^2 score computations for every model against its corresponding target variable. The model achieved better convergence and made more accurate forecasts because of feature scaling techniques. Pearson correlation analysis revealed that students who spent more time on their phones experienced increased stress levels according to the results. Multiple regression models enabled a comprehensive analysis for selecting the most efficient method to estimate student mobile dependency and psychological stress.

4.3 Results

The study found strong links between student stress, phone use, and academic achievement. Higher levels of self-reported stress were consistently associated with more unfavorable academic outcomes, implying that stress has a significant impact on student performance. Sleep patterns appeared as an important moderating element, with irregular sleep lengths, particularly insufficient sleep, exhibiting strong connections with increased stress and academic difficulties. These findings highlight sleep quality as an important factor in student well-being and success.

4.4 Experimental Result

Phone addiction measures showed substantial links to both stress and academic performance. Students who reported higher anxiety when disconnected from their devices had poorer sleep quality and faced more academic issues. Nighttime phone use and frequent device unlocks were particularly strongly associated with these negative outcomes, revealing certain usage patterns that may be most harmful to student well-being.

This research evaluates ten machine learning models—Linear Regression, Decision Tree, Random Forest, Gradient Boosting, Support Vector Regressor (SVR), K-Nearest Neighbors (KNN), ElasticNet, XGBoost, LightGBM, and CatBoost—for predicting student stress levels using behavioral and physiological factors. Model effectiveness was assessed using performance indicators such as R^2, MSE, RMSE, and computational efficiency.

In this study, different regression models were used to predict both stress and smartphone addiction levels in students. The table below presents the mean squared error (MSE) and R^2 scores for each model, emphasize the predicting performance for both stress and addiction targets

Table 4.1: MSE and R^2 for every model

Model	Stress MSE	Stress R^2	Addiction MSE	Addiction R^2
Linear Regression	1.733768	0.779947	0.377224	0.953765
Decision Tree	1.706552	0.783401	1.561420	0.808621
Random Forest	1.645959	0.791092	1.141199	0.860126
Gradient Boosting	1.739346	0.779239	0.731382	0.910356
SVR	1.716774	0.782104	0.475553	0.941713
KNN	1.976469	0.749143	0.788959	0.903299
ElasticNet	1.732152	0.780152	0.486095	0.940421
XGBoost	1.680703	0.786682	0.881356	0.891974
LightGBM	1.667953	0.788300	0.906749	0.888862
CatBoost	1.634175	0.792587	1.078971	0.867753

The experimental results reveal that CatBoost had the best performance for forecasting stress, with the lowest Stress MSE (1.634175) and maximum Stress R^2 (0.792587). The model accurately predicted smartphone addiction (MSE = 1.078971, R^2 = 0.867753). Among the other models, Random Forest and XGBoost performed consistently well, with strong results for both targets. Random Forest had Stress R^2 of 0.791092 and Addiction R^2 of 0.860126, but XGBoost had R^2 of 0.786682 for stress and 0.891974 for addiction.

K-Nearest Neighbors (KNN) had the highest Stress MSE (1.976469) and the lowest Stress R^2 (0.749143), indicating it struggled with stress prediction. However, its performance in addiction prediction was decent (Addiction R^2 = 0.903299). The Decision Tree model, albeit being fast, did not perform as well as other models, especially in addiction prediction (Addiction R^2 = 0.808621), perhaps due to its sensitivity to overfitting.

In all, the results show that ensemble approaches such as Random Forest, Gradient Boosting, and XGBoost, as well as CatBoost, are the most successful models for predicting stress and addiction, with CatBoost outperforming both targets.

Correalation Matrix: The supplied matrix depicts the links between several variables connected to stress and addiction in students, shedding light on how different aspects of phone usage and stress are linked. The matrix displays the correlation coefficients, which represent the strength and direction of the relationships. Strong positive correlations are discovered between Self_Reported_Stress_Level and

Negative_Impact_On_Academics (0.85), implying that students who report higher stress levels face more academic difficulty.

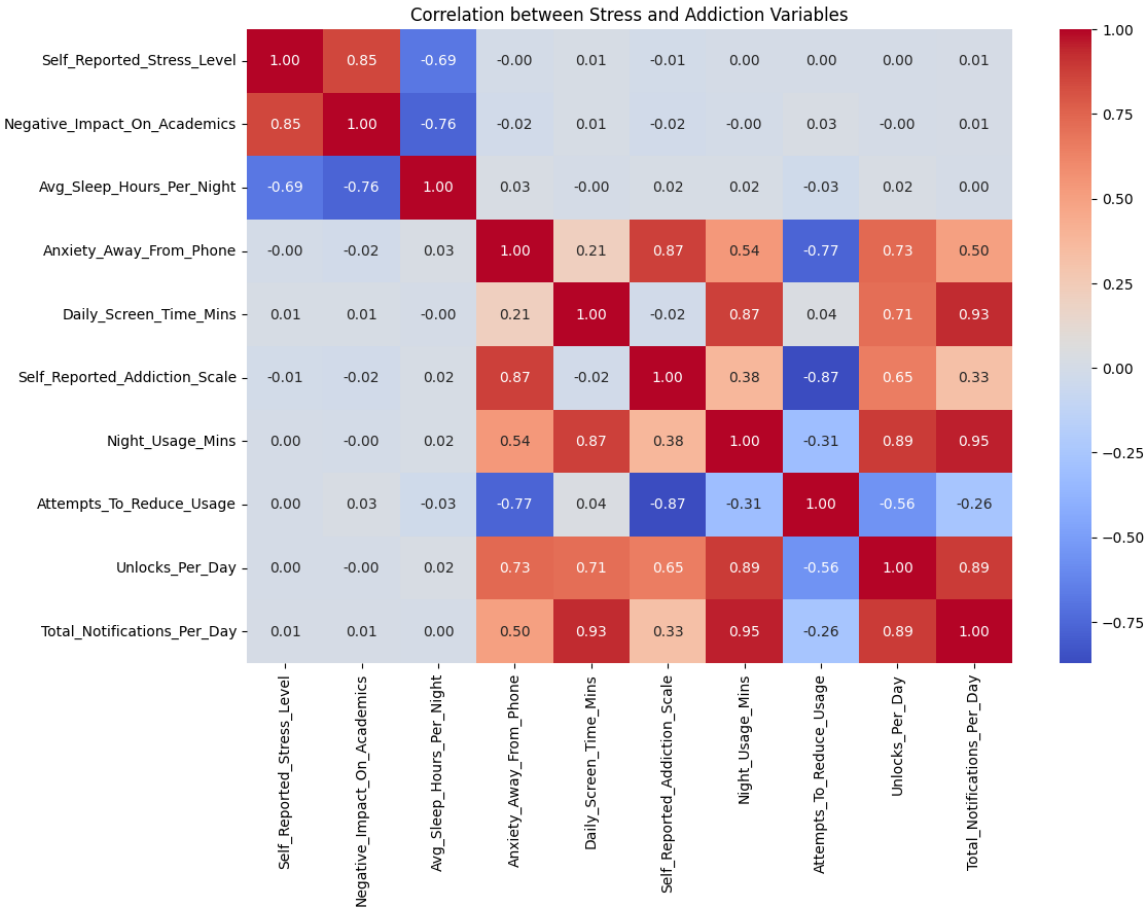

Figure 4.1: Correlation Matrix

There is a significant negative connection (-0.69) between Self_Reported_Stress_Level and Avg_Sleep_Hours_Per_Night, showing that students who sleep less are more likely to report stress. Furthermore, Self_Reported_Addiction_Scale has a substantial positive connection with Anxiety_Away_From_Phone (0.87), implying that higher self-reported addiction is related with increased anxiety when away from the phone. The Daily_Screen_Time_Mins and Night_Usage_Mins have a substantial positive association (0.87), indicating that students who use their phones during the day are more likely to use them at night. Attempts_To_Reduce_Usage, on the other hand, has a substantial negative association with Self_Reported_Addiction_Scale (-0.87), imply that students who make efforts to minimize their phone usage report lower levels of addiction. Furthermore, total_Notifications_Per_Day has a very strong positive connection with Unlocks_Per_Day (0.93), indicating that more notifications per day are directly related to a larger number of phones unlocks. These findings suggest that phone usage patterns, such as the number of unlocks, screen time, and alerts, are inextricably linked to levels of addiction and stress in students.

4.4.1 Performance and Comparative Analysis

To compare model performance, we may break out the Mean Squared Error (MSE) and R^2 (Coefficient of Determination) for Stress and Addiction predictions. These measurements will help us choose which model is most suited to each task.

MSE: Determines the average squared difference between expected and actual values. Lower numbers indicate improved model performance.

R^2: The percentage of variance in the dependent variable that can be predicted based on the independent factors. Higher values (closer to 1) are better.

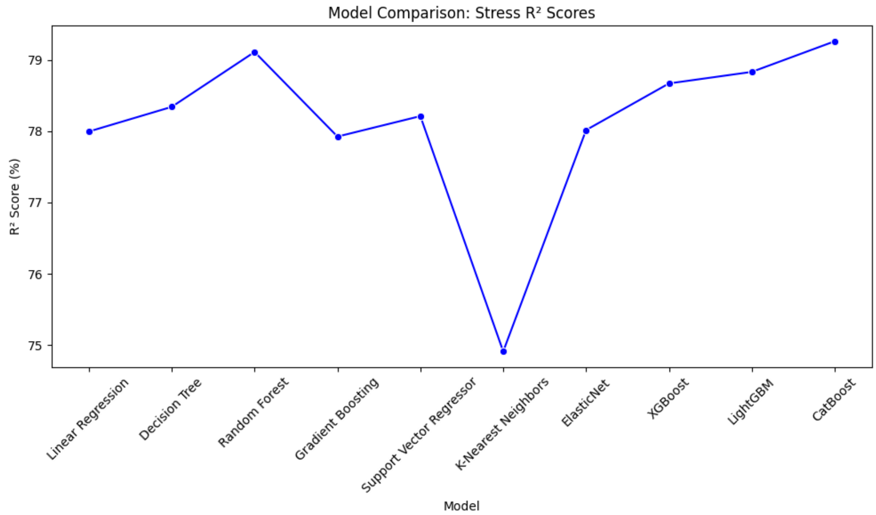

Figure 4.2: Stress R^2 score

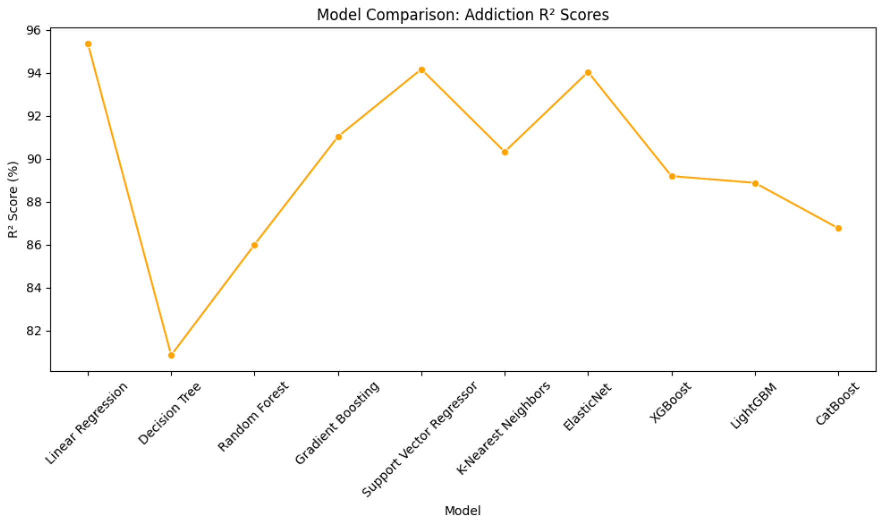

Figure 4.3: Addiction R^2 score

The stress prediction model CatBoost reaches the highest performance levels by showing 0.792587 R^2 value and 1.634175 MSE. CatBoost demonstrates superiority as a stress level prediction model by minimizing errors with a high data variance explanatory capability thus making it the most accurate and explanatory model. The performance of LightGBM and Random Forest results in lower MSE and R^2

values approaching 0.79. The models demonstrate satisfactory results but not to the extent CatBoost achieves. KNN proves to be the least efficient model because it generates maximum MSE (1.976469) alongside minimum R^2 (0.749143). The predictions from KNN reveal its inferior stress forecasting capabilities because it produces higher errors and cannot adequately interpret the dataset variations.
Linear Regression proves to be the best model for addiction prediction since it delivers the minimum MSE (0.377224) alongside the maximum R^2 value (0.953765). Linear Regression proves its excellent performance as an addiction level prediction tool since it shows minimal prediction errors and effective data variance explanation. SVR, ElasticNet and Gradient Boosting provide excellent results through MSE: 0.475553, R^2: 0.941713 while MSE: 0.486095 with R^2: 0.940421 and MSE: 0.731382 plus R^2: 0.910356. Addiction data trends can be accurately tracked by these models because they generate high R^2 values that surpass 0.90. CatBoost demonstrates the lowest success for predicting addiction levels since its MSE reaches 1.078971 while its R^2 falls to 0.867753.

4.5 Summary

The second section explains how multiple machine learning regression algorithms functioned to determine student stress levels and smartphone addiction prevalence. The dataset received a complete preprocessing phase before normalization followed by training-testing set partitioning. Linear Regression joined with Decision Tree and Random Forest together with Gradient Boosting followed by Support Vector Regressor and K-Nearest Neighbors and ElasticNet and XGBoost and LightGBM and CatBoost to form a collection of regression techniques used to evaluate data via Mean Squared Error (MSE) and R^2 Score. The model robustness evaluation used cross-validation while adjustments to hyperparameters delivered the best model performance. The comparison of different model results allowed researchers to determine the most precise and credible indicators for stress and addiction which created a strong foundation for data-based student well-being examination

CHAPTER 5

Engineering Standards and Design Challenges

5.1 Compliance with the Standards

The research on predicting student stress and smartphone addiction based on-ing machine learning follows academic, ethical, and technical standards during the study. Ethical issues were considered top during data collection, the participants were provided with the informed consent to make sure that they realised the purpose of the research and what rights they were having, privacy and confidentiality. The data had been anonymized in adherence to data protection laws such as GDPR. The research adheres to academic standards, over which it ensures that all methodologies models and findings are scientific, and are supported by relevant literatures and referenced properly. The data quality was maintained through the standard preprocessing steps that entailed handling missing values, removing outliers, and feature selection. The performance metrics of recognized Linear Regression and Random Forest models used during this analysis is an example of how machine learning models were assessed not only for their reliability and validity but also for other performance metrics as well. Also, the research is ethical and meets institutional and regulatory standards with all required approvals taken prior to data collection, which strengthens the vigor of the study in both ethics and methodology.

5.1.1 Software Standards

The software tools and technologies employed in this research stick to industry norms for machine learning and data analysis. Python became the first language because of extensive landscape of libraries, Pandas for data arranging, Scikit-learn for machine learning algorithms, and Matplotlib and Seaborn for data visualization. These libraries are also welcome in the research community with regards to their efficiency, reliability, and extensive functionality. These models of machine learning were developed using famous frameworks, including XGBoost, CatBoost, and LightGBM, which are renowned for their supremely strong predictive modeling capabilities. Moreover, Jupyter Notebooks had an interactive coding environment that provided an effective way of exploring the data and testing models. The software setup conforms to the current state of the art of reproducible research practices where all code, models, and data pre-processing steps can be verified independently and replicated by other experts with the necessary knowledge.

5.1.2 Hardware Standards

The hardware used in this research follows the required standards for effective data processing, and model training. The machine learning models were trained and validated in Kaggle's cloud environment, which offers the use of the high-performance equipment such as NVIDIA Tesla P100. These GPUs enhanced the computation tasks significantly, especially if one were training complex models using big data sets. The GPU implementation facilitated a more rapid processing, hence shortening the duration of the entire training process. Further, the environment offered 16 GB of RAM and 20 GB of storage capacity, which was adequate for facilitating the data through and on the study involved. The Kaggle cloud infrastructure supported the hardware prerequisites for machine learning tasks, enabling effective running, and scaling of the models.

5.1.3 Community Standards

This research follows the community norms in respect to ethical research, data privacy and integrity of academic work. The study adheres to set standards for human participants-related research, thus, ensuring all data obtained from the students was protected and the students' privacy maintained. In accordance with expectations of the community, all the participants received informed consent, they were fully aware of the purpose of the study, their rights, including possibility to withdraw at any point. The research adheres to best practices in open science, and the code, the datasets, and the results are available for future use and verification. Commitment to responsible research practices is enhanced in the academic community due to increased transparency, reliability and participation rights that are advanced through adherence to these community standards of the research.

5.2 Impact on Society, Environment and Sustainability

5.2.1 Impact on Life

The mental and physical well-being gets affected substantially by school-related stress as well as personal issues and drug addiction. Sustained stress develops into complications such as depression alongside anxiety and sleep issues and weaken immunity. Stress causes life to become harder by affecting personal connections along with work or educational progress and enjoyment of daily life.

People experience stress which leads to poor mental health when they get addicted to either objects or technology. For example, spending too much time on your phone can make you feel alienated, less productive, and more nervous because you are always connected. It can also cause you to forget to take care of yourself, increasing stress.

To deal with stress and addiction, it is critical to employ healthy coping strategies such as exercise, relaxation techniques, and seeking help. This allows you to lessen the harmful impacts of stress and addiction while also feeling better and improving your general well-being.

5.2.2 Impact on Society and Environment

Stress and addiction have far-reaching consequences for society and the environment, affecting the health of individuals, families, and communities. On the social level, the growth in stress-related illnesses such as anxiety, depression, and burnout has resulted in a greater need for mental health treatments. This not only stresses healthcare systems, but also lowers productivity, resulting in economic losses. In the workplace, stress frequently leads in absenteeism, high turnover rates, and poor performance, hurting corporate outcomes and national productivity[56].

Addiction, namely substance addiction, exacerbates these problems. It causes health difficulties, social isolation, and often a cycle of poverty, putting a strain on social support systems. The costs of addiction treatment and recovery, as well as coping with the repercussions of addiction, such as criminal behavior and emergency healthcare, place a strain on society's resources. Addiction produces immediate effects on families and communities through domestic violence and child neglect along with criminality that causes neighborhoods to become unstable and intensifies societal tensions.

Addiction and stress can result in environmentally harmful activities. People battling with addiction may participate in illicit activities such as drug production or substance disposal that pollute the air, soil, and water. Furthermore, plastics and chemicals included in addiction-related items lead to environmental damage. On a societal level, stress caused by widespread addiction can lead to greater consumption of disposable goods and services, putting additional demand on natural resources and leading to environmental damage. Stressful lifestyles may also contribute to unsustainable behaviors such as excessive waste, overuse of resources, and disdain for environmental conservation [57]. As these problems deepen, society becomes caught in a cycle of chaos, health crises, and harm to the environment. Addressing stress and addiction holistically, through improved mental health care, support systems, and public health efforts, is critical for reducing their negative impact on society and the environment.

5.2.3 Ethical Aspects

Ethical considerations for stress and addiction include personal responsibility, equity, stigma, and privacy. Addiction and stress frequently hinder people's ability to make educated decisions, creating concerns about personal liberty and responsibility. While individuals suffering from addiction may lose control over their behaviors, ethical questions arise regarding the degree to which they should be held accountable for their behavior, given the extrinsic factors including heredity, environment, and cultural influences [58]. Access to care is another important ethical concern, as all individuals, regardless of their circumstances, should have equal access to healthcare, treatment, and support for managing stress and addiction. In this context, ethical conduct entails treating people with dignity and respect, without moral judgment or exclusion. Businesses and educational institutions must accept responsibility for both events that decrease and elevate stress among their members. Businesses alongside educational organizations need to build mental and physical equilibrium within their premises while offering therapy services to their users. [59]. The healthcare system depends on strict privacy implementation for sensitive information because breaches in privacy damage patients while reducing their trust in medical services. A comprehensive method that establishes harmony between personal freedoms and community justice and organizational accountability needs to exist for ethical stress and addiction treatment. Appreciation-centered ethical constructs together with concepts of justice and empathy serve as key tools for helping individuals facing such concerns [60].

5.2.4 Sustainability Plan

The research project "Predicting Student Stress and Smartphone Addiction by Machine Learning" develops both environmental reduction initiatives and community-wide long-term advantages. The study will cut energy waste by studying sustainable methods and using renewable cloud computing platforms for model training and deployment. The project will introduce green data management procedures which combine responsible resource utilization with decreased carbon footprint.

The research will adopt emerging technologies for sustainable behavior promotion thus reducing environmental impact. A positive impact on student mental well-being will be among the social benefits this study seeks to achieve through its management of stress and smartphone addiction. Research findings might create apps or programs that help students tackle the said issues. The implementation of these concepts will become practical through partnership work between educational institutions and mental healthcare organizations. The study adds value to social sustainability and environmental sustainability by using efficient energy methods and delivering tangible benefits to students. The main purpose involves creating ethical sustainable options to enhance student well-being alongside reducing environmental implications.

5.3 Project Management and Financial Analysis

The project budget covers all necessary parts of execution, including appropriate funding for each phase to ensure smooth progress. Allocations exist for hardware and infrastructure needs together with software and tools acquisition as well as data collection and processing and documentation and report writing and more. The project's desired results require all resources and activities which belong to the established categories. Actual expenses might differ depending on project complexity as well as shifting project requirements.

Table 5.1: Financial Estimation of the Project

SN	Components	Estimated Cost
01	Hardware/Infrastructure	1000-1500
02	Software and Tools	2000-4000
03	Data Collection and Processing	4000-5000

04	Documentation and Report Writing	1000-2000
05	Miscellaneous	2000-3000
Total Estimated Cost		10000-15500

5.4 Complex Engineering Problem

The complex engineering problem centers around predicting student stress and smartphone addiction through machine learning, with major concerns regarding data quality, model selection, and ethical standpoints. The task is data cleaning as well as preprocessing to ensure this data is correct, missing values, duplicates, and outliers are dealt with. More than one machine learning model is assessed and tuned to forecast stress and addiction severities where performance is quantified by Mean Squared Error (MSE) and R^2 scores. Correlation analysis is used to detect relationships between the input features such as the screen time, sleep mode and academic pressure, and target variables. In addition to that privacy and moral issues are controlled by following data protection requirements and protecting the identity of the participants. The end goal is to create a predictive model to evaluate stress and addiction but also find real-time intervention strategies for enhancing the student's well-being in the digital age.

5.4.1 Complex Problem Solving

Table 5.2: Mapping with Complex Problem Solving

Ep1 Dept. of knowledge	Ep2 Range of Conflictiing Requirements	Ep3 Depth of Analysis	Ep4 Familiarity Issues	Ep5 Extent of Applicable Codes	Ep6 Extent Of Stakeholder Involvemnet	Ep7 Interdependence
Yes	No	Yes	No	No	No	Yes

EP1 – Depth of Knowledge

The project is an indication of profound knowledge in machine learning methods and behaviour data analysis. The application of multiple algorithms including Random Forest, XGBoost, and Support Vector Machines (SVM) in order to predict student stress & smartphone addiction demonstrates the mastery over complex models. In addition, the combination of data preprocessing methods such as feature selection and missing values enables one to have robust knowledge in the area of data manipulation, which is an essential part of the problem-solving process.

EP3 – Depth of Analysis

The depth of analysis in this project will involve developing a robust machine learning model to predict student stress and smartphone addiction, using diverse datasets and real-time data. It will include evaluating multiple models, applying Pearson correlation analysis to identify the most effective approach, and exploring the relationship between screen time, sleep, and stress. The analysis will provide actionable insights to inform personalized interventions aimed at improving student well-being and academic performance.

EP7 – Interdependence

This project includes several successive layers such as data collection, preprocessing, model training, and evaluation. Every stage is dependent on the success of the preceding stage, which helps to make the predictions made by the model a reality and relevant. The collaboration of various methodologies, machine learning and statistical analysis, evidences a high level of interdependence used in creating the final model for the prediction of student stress and also predicting student smartphone addiction. Combining these elements makes the project's overall success easier because the models are reliable and accurate.

Mapping with Knowledge Profile for EP1

Table 5.3: Mapping with Knowledge Profile

K3 Engineering Fundamentals	K4 Specialist Knowledge	K5 Engineering Design	K6 Engineering Practice	K8 Research Literature
Yes	Yes	Yes	Yes	Yes

K3 – Engineering Fundamentals

Some of the significant engineering principles that are applied in this research are data collection technique, algorithmic design and statistical analysis. The mastery of machine learning and its application to real life data points out the engineering underpinnings implemented in this project. Using systematic problem-solving approaches and validation techniques, this project represents the incorporation of engineering knowledge in the building of predictive models.

K4 – Specialist Knowledge

The project incorporates K4 Specialist Knowledge by using more sophisticated machine learning techniques and behavioural analysis to identify students' stress level and smartphone addiction. The application of the unique sorts of algorithms like Random Forest, XGBoost, and Support Vector Regressor shows the profound knowledge of the fields of predictive models.

K5 – Engineering Design

In case of K5 Engineering Design, the project creates a strong framework for data preprocessing, model training, and testing, guaranteeing effective solution for prediction of the stress and addiction levels.

K6 – Engineering Practice

The research is also congruent with K6 Engineering Practice and applies this investigation of engineering principles in the practical environment of mental health and digital behavior by using the best practices during data analysis, model optimisation and system evaluation.

K8 – Research Literature

The project is based extensively on K8 Research Literature, taking off from previous studies done in domain of machine learning, student well-being and smartphone addiction and adding its own unique knowledge by application of machine learing model for prediction of mental health.

5.4.2 Engineering Activities

Table 5.4: Mapping with Complex Engineering Activities

EA1 Range of resources	EA2 Level of Interaction	EA3 Consequences for innovation environment	EA4 Familarity
Yes	Yes	Yes	Yes

EA1 – Range of Resources

The project required the use of diverse resources such as machine learning, survey data, and cloud-based platforms such as Kaggle GPUs to facilitate efficient training of the models in the use of the data. These resources made possible the creation of a strong model for stress and smartphone addictions among students.

EA2 – Level of Interaction

There was a great need to interact between different phases of model development improving the efficiency in collecting data, preprocessing them, model evaluation and tuning of performance. Members of the team collaborated to improve the models ensuring embedding of domain knowledge at each step.

EA3 – Consequences for Innovation Environment

This research presented new strategies for predicting student stress and smartphone addictions using machine learning models to make live predictions. The application of cloud computing for model training and testing and XAI add to the project’s net effect of furthering technology around mental health applications.

EA4 – Familiarity

The project utilizes well-known machine learning algorithms and techniques that are commonly used both in predicting analytics and behavioral studies. This knowledge of existing methodologies meant that the project remained true to best practices, while creating reliable results.

5.5 Summary

The research's more extensive effects on society along with environmental consequences and sustainability aspects are evaluated in the "Impact on Society and Environment and Sustainability" CHAPTER. The study emphasizes the requirement to use energy-aware methodology together with renewable energized cloud computing to decrease environmental effects when training and deploying models. Environmental data management techniques receive special attention from the study since they are essential for reducing the research's carbon emissions. The research tackles student stress and smartphone addiction issues on a social level and introduces new methods to enhance student welfare. School cooperation with mental health groups ensures that the study implements practical and effective concepts. The CHAPTER demonstrates that researchers should perform ethical and environmentally friendly studies which align technical advances with sustainability targets to produce societal and environmental advantages.

CHAPTER 6

Conclusion

6.1 Conclusion

This thesis presents solutions to student stress and smartphone addiction by focusing on machine learning techniques for detection and improvement of these problems. Various developed models enable researchers to understand better how education and personal life gets affected by these factors. This study demonstrates that sustainable practices should continue to matter because they enable both minimization of environmental damage through energy-efficient computing methods and creation of practical mental health solutions for society. The research demonstrates how technological capabilities reach substantial transformations when uniting ethical components with sustainable procedures. The research findings contribute to both academic understanding of stress and addiction while simultaneously offering applications for student welfare enhancement. The use of machine learning technology in mental health prediction will advance to develop purposeful treatment approaches which serve educational groups alongside individual clients.

6.2 Limitations

The study delivers important knowledge about student stress and smartphone addiction, yet it also contains specific constraints. The study depended on self-reported information that might contain reporting errors. Study participants had the tendency to either interpret questions incorrectly or respond with conventional answers rather than revealing their true feelings. The research accuracy and reliability could be affected by this factor. The investigation only focused on one student group which means its results cannot be applied to students from diverse nationalities or cultural backgrounds or different age brackets.

The data failed to include essential factors such as personality traits, family background and mental health history among its key variables. Multiple elements could produce considerable effects on stress and addiction levels. Machine learning models performed various tests, yet their execution might prove inconsistent with distinct data arrangements in actual usage.

6.3 Future Work

Research into additional factors regarding student stress and smartphone addiction would generate valuable findings. Further investigations should focus on understanding how academic pressure and external activities together with distinct methods of dealing with stress impact these problems. Future research should address the important issue of how smartphone addiction permanently affects a person's academic results together with their mental health status and social bonds. The research results' applicability would improve when the study extended to examine various age ranges combined with geographical locations and cultural backgrounds. To evaluate the effectiveness of smartphone addiction and stress reduction for student's researchers should test methods such as mindfulness programs together with time management tools and digital detox techniques.

The future development of student stress and smartphone addiction research could focus on using advanced machine learning techniques including deep learning to enhance prediction accuracy. Further research should integrate a variety of data types including audio tapes and video recordings with biometric measurements to achieve better understanding of the root factors that lead to student stress and smartphone addiction. A continuous approach needs to handle ethical dimensions regarding personal data usage in machine learning models. Complete confidence from students in predictive system interaction depends on privacy compliance and clear regulations about how their data will be used.

References

[1] "Stress," World Health Organization. [Online]. Available: https://www.who.int/news-room/questions-and-answers/item/stress. [Accessed: Mar. 10, 2025].

[2] "How do smartphones affect mental health?" Therapy Brands. [Online]. Available: https://therapybrands.com/blog/what-is-the-impact-of-smartphone-addiction-on-mental-health. [Accessed: Mar. 10, 2025].

[3] S. Thomée, A. Härenstam, and M. Hagberg, "Mobile phone use and stress, sleep disturbances, and symptoms of depression among young adults - a prospective cohort study," *BMC Public Health*, vol. 11, no. 1, p. 66, Dec. 2011, doi: 10.1186/1471-2458-11-66.

[4] A. K. Przybylski and N. Weinstein, "Can you connect with me now? How the presence of mobile communication technology influences face-to-face conversation quality," *Journal of Social and Personal Relationships*, vol. 30, no. 3, pp. 237–246, May 2013, doi: 10.1177/0265407512453827.

[5] D. J. Kuss and M. D. Griffiths, "Online social networking and addiction—a review of the psychological literature," *Int. J. Environ. Res. Public Health*, vol. 8, no. 9, pp. 3528–3552, Aug. 2011, doi: 10.3390/ijerph8093528..

[6] R. Misra and M. McKean, "College students' academic stress and its relation to their anxiety, time management, and leisure satisfaction," *Am. J. Health Stud.*, vol. 16, no. 1, pp. 41–51, 2000. [Online]. Available: https://www.proquest.com/docview/210480531. [Accessed: Mar. 10, 2025].

[7] S. Arya, A. Anju, and N. Azuana Ramli, "Predicting the stress level of students using supervised machine learning and artificial neural network (ANN)," *Indian J. Eng.*, vol. 21, no. 56, pp. 1–24, Aug. 2024, doi: 10.54905/disssi.v21i55.e9ije1684.

[8] K. Tang, A. Kumar, M. Nadeem, and I. Maaz, "CNN-based smart sleep posture recognition system," *IoT*, vol. 2, no. 1, pp. 119–139, Feb. 2021, doi: 10.3390/iot2010007.

[9] J. Chung and J. Teo, "Mental health prediction using machine learning: taxonomy, applications, and challenges," *Appl. Comput. Intell. Soft Comput.*, vol. 2022, pp. 1–19, Jan. 2022, doi: 10.1155/2022/9970363.

[10] L. C. López Steinmetz, M. Sison, R. Zhumagambetov, J. C. Godoy, and S. Haufe, "Machine learning models predict the emergence of depression in Argentinean college

students during periods of COVID-19 quarantine," *Front. Psychiatry*, vol. 15, p. 1376784, Apr. 2024, doi: 10.3389/fpsyt.2024.1376784.

[11] Kavosh Cognitive Behavior Sciences and Addiction Research Center, School of Medicine, Guilan University of Medical Sciences, Rasht, Iran. *et al.*, "Relationship Between Smartphone Addiction and Stress and Life Satisfaction in Medical Students," *jour guilan uni med sci*, vol. 30, no. 2, pp. 144–155, Jul. 2021, doi: 10.32598/JGUMS.30.2.1742.1.

[12] A. K. Przybylski and N. Weinstein, "Can you connect with me now? How the presence of mobile communication technology influences face-to-face conversation quality," *Journal of Social and Personal Relationships*, vol. 30, no. 3, pp. 237–246, May 2013, doi: 10.1177/0265407512453827.

[13] J. D. Elhai, R. D. Dvorak, J. C. Levine, and B. J. Hall, "Problematic smartphone use: A conceptual overview and systematic review of relations with anxiety and depression psychopathology," *Journal of Affective Disorders*, vol. 207, pp. 251–259, Jan. 2017, doi: 10.1016/j.jad.2016.08.030.

[14] A. Singh, K. Singh, A. Kumar, A. Shrivastava, and S. Kumar, "Machine Learning Algorithms for Detecting Mental Stress in College Students," Dec. 10, 2024. doi: 10.1109/I2CT61223.2024.10544243.

[15] S. Arya, A. Anju, and N. Azuana Ramli, "Predicting the stress level of students using Supervised Machine Learning and Artificial Neural Network (ANN)," *Indian J. Eng.*, vol. 21, no. 56, pp. 1–24, Aug. 2024, doi: 10.54905/disssi.v21i55.e9ije1684.

[16] J. Lee and W. Kim, "Prediction of Problematic Smartphone Use: A Machine Learning Approach," *IJERPH*, vol. 18, no. 12, p. 6458, Jun. 2021, doi: 10.3390/ijerph18126458.

[17] S. Likhith, C. Chitteti, M. Dharani, V. Nivedhitha, N. G. Geethika, and V. Godwin, "Machine Learning Model for Prediction of Smartphone Addiction," in *2024 International Conference on Expert Clouds and Applications (ICOECA)*, Bengaluru, India: IEEE, Apr. 2024, pp. 924–929. doi: 10.1109/ICOECA62351.2024.00163.

[18] Department of Electronics and Communication, Jain College of Engineering, Karnataka, India *et al.*, "Decoding Minds: Estimation of Stress Level in Students using Machine Learning," *IJST*, vol. 17, no. 19, pp. 2002–2012, May 2024, doi: 10.17485/IJST/v17i19.2951.

[19] "(PDF) A Review: Machine Learning Algorithms for Analysis of Smartphone Addiction," ResearchGate. Accessed: Mar. 10, 2025. [Online]. Available: https://www.researchgate.net/publication/373901475_A_Review_Machine_Learning_Algorithms_for_Analysis_of_Smartphone_Addiction

[20] C. Ding, Y. Zhang, and T. Ding, “A systematic hybrid machine learning approach for stress prediction,” *PeerJ Computer Science*, vol. 9, p. e1154, Feb. 2023, doi: 10.7717/peerj-cs.1154.

[21] C. F. Giraldo-Jiménez, J. Gaviria-Chavarro, M. Sarria-Paja, L. A. Bermeo Varón, J. J. Villarejo-Mayor, and A. L. F. Rodacki, “Smartphones dependency risk analysis using machine-learning predictive models,” *Sci Rep*, vol. 12, no. 1, p. 22649, Dec. 2022, doi: 10.1038/s41598-022-26336-2.

[22] V. Jakkula, "Tutorial on Support Vector Machine (SVM)," *School of Electrical Engineering and Computer Science, Washington State University*, 2006. [Online]. Available: https://course.ccs.neu.edu/cs5100f11/resources/jakkula.pdf

[23] “Pearson Correlation - an overview | ScienceDirect Topics.” Accessed: Apr. 08, 2025. [Online]. Available: https://www.sciencedirect.com/topics/computer-science/pearson-correlation

[24] “How much RAM does Kaggle provide? – Fabalabse.” *Fabalabse*, Apr. 8, 2025. [Online]. Available: https://fabalabse.com/vpn/how-much-ram-does-kaggle-provide. [Accessed: Apr. 8, 2025].

[25] “VenkatManda/KaggleQuestions · Datasets at Hugging Face.” Accessed: Apr. 08, 2025. [Online]. Available: https://huggingface.co/datasets/VenkatManda/KaggleQuestions/viewer

[26] G. Van Rossum and F. L. Drake Jr., *Python tutorial* (Vol. 620). Amsterdam, The Netherlands: Centrum voor Wiskunde en Informatica, 1995

[27] W. McKinney, "pandas: a foundational Python library for data analysis and statistics," *Python for High Performance and Scientific Computing*, vol. 14, no. 9, pp. 1–9, 2011. [Online]. Available: https://pandas.pydata.org/

[28] C. R. Harris, K. J. Millman, S. J. van der Walt, R. Gommers, P. Virtanen, and others, "Array programming with NumPy," *Nature*, vol. 585, no. 7825, pp. 357–362, 2020. [Online]. Available: https://numpy.org/

[29] J. D. Hunter, "Matplotlib: A 2D graphics environment," Computing in Science & Engineering, vol. 9, no. 3, pp. 90–95, 2007. [Online]. Available: https://matplotlib.org/

[30] M. Waskom, “seaborn: statistical data visualization,” *JOSS*, vol. 6, no. 60, p. 3021, Apr. 2021, doi: 10.21105/joss.03021.

[31] “Find Open Datasets and Machine Learning Projects | Kaggle.” Accessed: Apr. 08, 2025. [Online]. Available: https://www.kaggle.com/datasets

[32] L. Pereira, "Conceptualizing Python in Google Colab: Hands-on Practical Sessions," *International Journal of Modern Engineering Research*, vol. 2, no. 5, pp. 3482–3496, 2012

[33] P. G. Naik, "Conceptualizing Python in Google Colab: Hands-on Practical Sessions," *International Journal of Modern Engineering Research*, vol. 2, no. 5, pp. 3482–3496, 2012.

[34] P. G. Naik, "Conceptualizing Python in Google Colab: Hands-on Practical Sessions," *International Journal of Modern Engineering Research*, vol. 2, no. 5, pp. 3482–3496, 2012.

[35] A. Madhu, M. Rajani, M. Venkataramana, M. Subbarayudu, and K. N. Sreenivasulu, "Time varying estimation of simple linear regression model using elemental slopes," *South Asian Journal of Marketing & Management Research*, vol. 5, no. 5, pp. 61–69, 2015.

[36] B. Bhoomeshwar and Y. Nagesh, "Random Forest Classifier for Classifying Bird Species using Scikit-learn," *International Journal of Advances in Engineering & Technology*, vol. 10, pp. 1373–1377, 2019.

[37] G. Biau, "Analysis of a random forests model," *Journal of Machine Learning Research*, vol. 13, pp. 1063–1095, 2012.

[38] Anshul, "Gradient Boosting Algorithm: Guide for Beginners," Analytics Vidhya. Accessed: Apr. 08, 2025. [Online]. Available: https://www.analyticsvidhya.com/blog/2021/09/gradient-boosting-algorithm-a-complete-guide-for-beginners/

[39] C. Li, *A Gentle Introduction to Gradient Boosting*, Northeastern University, 2016. [Online]. Available: https://www.chengli.io/tutorials/gradient_boosting.pdf. [Accessed: May 9, 2025]

[40] "What Is Gradient Boosting?," Snowflake. Accessed: Apr. 08, 2025. [Online]. Available: https://www.snowflake.com/guides/what-gradient-boosting/

[41] "A Guide to The Gradient Boosting Algorithm." *DataCamp*, Apr. 8, 2025. [Online]. Available: https://www.datacamp.com/tutorial/guide-to-the-gradient-boosting-algorithm. [Accessed: Apr. 8, 2025].

[42] "Support Vector Regression - an overview | ScienceDirect Topics." *ScienceDirect*, Apr. 8, 2025. [Online]. Available: https://www.sciencedirect.com/topics/computer-science/support-vector-regression. [Accessed: Apr. 8, 2025]

[43] A. Sethi, “Support Vector Regression Tutorial for Machine Learning,” Analytics Vidhya. Accessed: Apr. 08, 2025. [Online]. Available: https://www.analyticsvidhya.com/blog/2020/03/support-vector-regression-tutorial-for-machine-learning/

[44] “Understanding Support Vector Machine Regression.” *MathWorks*, Apr. 8, 2025. [Online]. Available: https://www.mathworks.com/help/stats/understanding-support-vector-machine-regression.html. [Accessed: Apr. 8, 2025]

[45] “*k*-nearest neighbors algorithm,” *Wikipedia*. Mar. 14, 2025. Accessed: Apr. 08, 2025. [Online]. Available: https://en.wikipedia.org/w/index.php?title=K-nearest_neighbors_algorithm&oldid=1280352468

[46] “K-Nearest Neighbor (KNN) Explained | Pinecone.” Accessed: Apr. 08, 2025. [Online]. Available: https://www.pinecone.io/learn/k-nearest-neighbor/

[47] H. Zou and T. Hastie, “Regularization and Variable Selection Via the Elastic Net,” *Journal of the Royal Statistical Society Series B: Statistical Methodology*, vol. 67, no. 2, pp. 301–320, Apr. 2005, doi: 10.1111/j.1467-9868.2005.00503.x.

[48] “XGBoost,” *Wikipedia*. Mar. 25, 2025. Accessed: Apr. 08, 2025. [Online]. Available: https://en.wikipedia.org/w/index.php?title=XGBoost&oldid=1282242745

[49] A. Tyagi, “What is XGBoost Algorithm?,” Analytics Vidhya. Accessed: Apr. 08, 2025. [Online]. Available: https://www.analyticsvidhya.com/blog/2018/09/an-end-to-end-guide-to-understand-the-math-behind-xgboost/

[50] “Unveiling Mathematics Behind XGBoost,” KDnuggets. Accessed: Apr. 08, 2025. [Online]. Available: https://www.kdnuggets.com/unveiling-mathematics-behind-xgboost

[51] “CatBoost,” *Wikipedia*. Feb. 24, 2025. Accessed: Apr. 08, 2025. [Online]. Available: https://en.wikipedia.org/w/index.php?title=CatBoost&oldid=1277472721

[52] “How training is performed | CatBoost.” *CatBoost*, Apr. 8, 2025. [Online]. Available: https://catboost.ai/docs/en/concepts/algorithm-main-stages. [Accessed: Apr. 8, 2025].

[53] “How training is performed |.” *CatBoost*, Apr. 8, 2025. [Online]. Available: https://catboost.ai/docs/en/concepts/algorithm-main-stages. [Accessed: Apr. 8, 2025].

[54] “Understanding CatBoost: The Gradient Boosting Algorithm for Categorical Data | by Aravind Kolli | Medium.” Accessed: Apr. 08, 2025. [Online]. Available: https://aravindkolli.medium.com/understanding-catboost-the-gradient-boosting-algorithm-for-categorical-data-73ddb200895d

[55] A. Kolli, “Understanding CatBoost: The Gradient Boosting Algorithm for Categorical Data,” Medium. Accessed: Apr. 08, 2025. [Online]. Available:

https://aravindkolli.medium.com/understanding-catboost-the-gradient-boosting-algorithm-for-categorical-data-73ddb200895d

[56] N. Schneiderman, G. Ironson, and S. D. Siegel, “Stress and Health: Psychological, Behavioral, and Biological Determinants,” *Annu. Rev. Clin. Psychol.*, vol. 1, no. 1, pp. 607–628, Apr. 2005, doi: 10.1146/annurev.clinpsy.1.102803.144141.

[57] “(PDF) Causes and Health Consequences of Environmental Degradation and Social Injustice,” *ResearchGate*, Oct. 2024, doi: 10.1016/S0277-9536(02)00055-2.

[58] “Combining medically assisted treatment and Twelve-Step programming: a perspective and review - PubMed.” Accessed: Apr. 08, 2025. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/28387530/

[59] C. Maslach, W. B. Schaufeli, and M. P. Leiter, “Job burnout,” *Annu Rev Psychol*, vol. 52, pp. 397–422, 2001, doi: 10.1146/annurev.psych.52.1.397.

[60] “Confidentiality in health care: Conflicting legal and ethical issues,” *ResearchGate*, Oct. 2024, doi: 10.7748/ns2007.01.21.21.42.c4513.

212-15-4171

ORIGINALITY REPORT

21%	16%	12%	14%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Daffodil International University Student Paper	2%
2	dspace.daffodilvarsity.edu.bd:8080 Internet Source	2%
3	Submitted to United International University Student Paper	1%
4	ijphs.iaescore.com Internet Source	1%
5	Mukadder Erdem, Esra Kaya. "Educational management in a digital age: boss phubbing and teacher motivation", International Journal of Evaluation and Research in Education (IJERE), 2025 Publication	<1%
6	kth.diva-portal.org Internet Source	<1%
7	Submitted to University of Nottingham Student Paper	<1%
8	arxiv.org Internet Source	<1%
9	www.techscience.com Internet Source	<1%