MOBILE GAME ADDICTION LEVEL PREDICTION AMONG UNIVERSITY STUDENTS USING MACHINE LEARNING APPROACHES

BY

JUBAYER AL MASUD
ID: 201-15-14153

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

MD. AYNUL HASAN NAHID
Lecturer
Department of Computer Science and Engineering
Daffodil International University

Co-Supervised By

MAHIMUL ISLAM NADIM
Lecturer
Department of Computer Science and Engineering
Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

July 2024

APPROVAL

This Project titled "**Mobile Game Addiction Level Prediction Among University Students Using Machine Learning Approaches**", submitted by **Jubayer Al Masud** 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 14-07-2024.

BOARD OF EXAMINERS

Hossain 14-07-2024

Dr. Md. Fokhray Hossain (MFH) — **Board Chairman**
Professor
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

Amatul Bushra Akhi (ABA) — **Internal Examiner 1**
Assistant Professor
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

Mr. Amir Sohel (ARS) — **Internal Examiner 2**
Sr. Lecturer
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

14/7/24

Dr. Shamim H Ripon (DSR) — **External Examiner**
Professor
Department of Computer Science and Engineering
East West University

DECLARATION

I hereby declare that this project has been done by us under the supervision of **Md. Aynul Hasan Nahid, Lecturer, Department of Computer Science and Engineering, Daffodil International University**. We also declare that neither this project nor any part of this project has been submitted elsewhere for the award of any degree or diploma.

Supervised by:

Md. Aynul Hasan Nahid
Lecturer
Department of CSE
Daffodil International University

Co-Supervised by:

Mahimul Islam Nadim
Lecturer
Department of CSE
Daffodil International University

Submitted by:

Jubayer Al Masud

Jubayer Al Masud
ID: -201-15-14153
Department of CSE
Daffodil International University

ACKNOWLEDGEMENT

First, I express our heartiest thanks and gratefulness to almighty for His divine blessing making it possible for us to complete the final year project successfully.

I am grateful and wish our profound indebtedness **Md. Aynul Hasan Nahid, Lecturer,** Department of CSE Daffodil International University, Dhaka. Deep Knowledge & keen interest of our supervisor in the field of "*Machine learning*" to carry out this project. Her endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior drafts, and correcting them at all stages have made it possible to complete this project.

I would like to express our heartiest gratitude to the **Head of the Department of CSE**, for his kind help in finishing our project and also to other faculty members and the staff of the Department of CSE, Daffodil International University.

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

Finally, I must acknowledge with due respect the constant support and patience of my parents.

ABSTRACT

Using mobile game addiction as a research focus, this case study examines how this digital disorder affects the academic performance and social life of university students. The study objectives include applying machine learning analysis for predicting the degree of mobile game addiction and creating a prediction model that will facilitate early intervention. One method used in this study was a cross-sectional survey of 600 university students on 17 variables: gaming tendency and several mood and social consequences. The obtained dataset underwent some preprocessing and encoding in order to be tested for different ml algorithms. Out of all the models, Random Forest Classifier achieved the highest accuracy of 96.45 % and Gradient Boosting Classifier Test set accuracy was 96.04% and for Decision Tree Classifier was 95.04%. Logistic Regression and GaussianNB had the lowest ranking, scoring 79.58% and 74.37%, respectively. Originally, the study result showed that feature selection and data preprocessing had a dramatic impact on model performance. Dependents is well classified with the Random Forest model and the wrongly classified addiction levels are demographically misclassified. Based on the study, it is critical to note that the various machine learning techniques can assist in identifying learners who require assistance concerning possible detrimental effects on their mental wellbeing. The subsequent research should recruit more participants and from a diverse population, while developing more intricate models for the improved prediction of depression risks.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1
INTRODUCTION

1.1 Overview

Since the use of smartphones and internet is advancing rapidly, mobile games are part of the university students' everyday lives. On the positive side, people get entertainment, stress relief, social engagement as well as interaction through mobile games, while on the negative side, it can result in dependence, which has negative effects on learning, interpersonal relationships, and mental health outcomes. Mobile game addiction refers to the uncontrolled and excessive use of mobile games which cause severe social, academic or vocational problems or intense emotional discomfort. Despite being different from other addictive disorders, it resembles internet and gambling addictions and is gradually gaining recognition. The current advancement in mobile games and their increased consumption by university students make it crucial to find the early signs of addictive behaviors. Conventional methods for the analysis of the treatment of addiction also utilize questionnaires and doctors' check-ups that, despite their applicability and usefulness, take much time and are to some extent inaccurate. More so, Machine learning provides an opportunity of developing prediction models that recognizes at-risk persons under many behavioral and psychological conditions. The purpose of this research is to explore the possibilities for using machine learning algorithms in predicting levels of mobile game addiction among university students. Through studying the data obtained by the means of elaborate questionnaires, our aim is to determine the factors, which might contribute to the growth of addictiveness, as well as to define the models, which will ensure the correct classification of students according to the level of addictiveness. The objectives are threefold: in order to determine major predictors of mobile game addiction, in order to compare the accuracy of various machine learning methods in predicting addiction levels, and in order to demonstrate the feasibility of warning signs and timely alerts. With this study, we envision ourselves playing the role of adding to the body of knowledge on mobile game addiction and present how machine learning can be used to address this problem among university students.

1.2 Background and Present State

The following are some essential terminologies which apply in the context of this study on the prediction of mobile game addiction among university students. Machine learning in simple terms is the process of using algorithms and other methods to teach the computer how to perform certain tasks on its own without being told how to do so step by step. Mobile game addiction can be described as spending an excessive amount of time using games on a mobile phone in a way that is detrimental to one's health and wellbeing, thereby causing undesirable effects on one's academic work and social life. Input is the key independent variables, data attributes that are used as predictor variables for the ML models and they include the time spent gaming, emotional label and social effects. Target Variable refers to the dependent variable that is to be predicted by the model and in our model, it is level of addiction / Level of Salvation (Non-addicted, Addicted, Semi-Addicted). Encoding involves the process by which categorical data is transformed into a numeric form in order to suit the ML algorithms. Exploratory Data Analysis (EDA) is the step of assessing and exploring data sets to derive the basic features of a dataset in a simpler and graphical manner. Model Training is the process of training an ML algorithm on a dataset so that it can identify the various patterns in the dataset collected Similarly, Model Evaluation is a method of evaluating the performance of an algorithm using the accuracy, precision, recall and F1 score of the model. These terminologies therefore define the methods used in this study as explained below.

1.3 Problem Statement

The essence of conducting this study is anchored on the fact that the level of mobile game addiction among university students is unprecedented. Continuing popularization of smartphones, the shares of gamers within the population have expanded significantly, and playing mobile games became one of the most frequent addictions. As this becomes an area of concern and students' weakness, these aspects merge to form a complex problem that needs to be addressed. Previous processes of diagnosing and treating addiction are usually based on questionnaires and clinical examinations, which are time-consuming and depend on the patient's honesty. On the other hand, machine learning presents a formidable tool that can be used to train a model on the data collected and analyze for patterns and indicators of addiction within the populations not easily identifiable through conventional approaches. Thus, this research seeks to utilize machine learning approaches to achieve improved prediction of mobile game addiction. These models will not only help in identifying various factors behind the addiction but also help in developing early screening tools basically aimed at prevention. This approach is expected to be faster and more efficient, less costly and less subjective in nature than conventional methods, hence shall improve on the academic performance and societal worth of university students, by freeing them from the vice of mobile gaming addiction.

1.4 Objective

The reason for this study was inspired by the fact that more university students are developing an addiction to mobile games that is hazardous to their health. As the usage of smartphones is becoming more widespread with each passing day, the consumption of mobile games poses as the primary means of pastime and recreation which in most cases result in addictive behaviors. While this addiction is not likely to reduce academic performance, it may disrupt social relationships and contribute to mental health problems hence requiring the identification of efficient strategies in its early determination. Historical techniques for diagnosing and addressing substance dependency including clinical evaluations and questionnaires are effective; however, existing method deficits include subjectivity and lengthy procedures. Firstly, machine learning provides a brand-new paradigm, which enables the analyzing of big data and finding complicated patterns that may be unnoticed at the first sight. Compared to the previous studies that involve critiquing and analyzing participants' written and spoken words this paper proposes to use machine learning to create models that can help provide an efficient, scalable solution to detecting students at risk of mobile game addiction. It is recognized that these models can be useful in early identification of addiction, which in turn would mean that possible negative effects of addiction can be contained through early intervention. More importantly, this research aims at enhancing the welfare of university students by eradicating a contemporary issue actively and positively impacting students' game usage and, in turn, their academic and social achievements through the use of advanced technologies.

1.5 Scope and Limitations

The expected output of this study is the identification of a number of important findings which will contribute to the existing body of knowledge pertaining to mobile game addiction among university students. First, as for the research objectives, it is its purpose to determine which behavioral/psychological factors have the most overwhelming impact on the degree of mobile game addiction. Thus, by studying such factors, it is possible to identify causes of addiction in detail. Secondly, it is expected that the mobile game users will have enough data that will enable the formation of accurate and reliable machine learning algorithms that will predict the level of mobile game addiction. These models will categorize the students into Non-addicted, Addicted and semi-addicted basing the classification on a survey. Thus, the performance of the Logistic Regression model, Decision Trees, Random Forest, SVM, and Neural Networks will be considered, and the most efficient approach will be determined. Furthermore, the study will reveal the ‘importance ratio’ of the features for the dependency prediction which can help in identifying the areas that require direct intervention. We also expect the models to perform differently based on the demographic data suggesting that the problem of addiction should be solved differently.

1.6 Report Organization

This paper presents a systematic overview of a study on machine learning for the prediction of heart diseases, with an emphasis on accuracy and implications. It looks at various aspects ranging from methodological issues to the social implications, it provides information that is central to the progress of health care. Gives a brief on heart disease and its consequence on the global health system why machine learning is relevant in the prediction of heart diseases. Present reviews and studies of prediction of heart disease, covering from conventional methods to machine learning. Explains the method adopted in data collection, data pre-processing, selection of the model, and the evaluation of the model in the study. Shows the results of the process of model training and evaluation on the cross-validation, comparing the characteristics of the algorithms and data preparation. Explores how accurate heart disease prediction may affect the health care system, outcomes of patients, and resource provision. Presents the overall conclusions about the paper’s results, the effectiveness of applying machine learning methods for predicting heart disease, the further research directions and recommendations for the transfer of findings to practice.

1.7 Summary

In the opening chapter of this research, the fixation of proper model for the prediction of heart diseases is discussed with emphasis on necessity for better models because it is one of the prevalent causes of death in today's world. With the data sample of 663 patients and 14 features, features being the numbers disclosing demographic characteristics and general state of health, the paper investigates the efficiency of 5 algorithms of machine learning, namely; Gradient Boosting, Gaussian Naive Bayes, Random Forest Classifier, Decision Tree Classifier and Logistic Regression to predict the disease. Thus, this study proves that accuracy is high for ensemble methods such as Random Forest and Gradient Boosting which shows the opportunities to identify risk factors and improve prognosis. The chapter emphasizes the importance of the following ethical issues: bias, especially in preconception, and model interpretability in order to apply predictive tools to enhance care strategies for heart diseases in clinical practice.

CHAPTER 2
LITERATURE REVIEW

2.1 Overview

The chapter on the analysis of prior studies aims to combine existing works on the use of machine learning in predicting heart diseases. It gathers papers using different types of algorithms and datasets and underlines that using ensembles such as Random Forest and Gradient Boosting gives high accuracy. The most notable discoveries on the choice of features, the size of the data set, and the significance of algorithm measures to improve the predictive models are well illuminated. Real-world issues like bias reduction and interpretation issues are mentioned as well, signifying more on how crucial they are when it comes to clinical applications. While the present study has contributed to the development of CRI and proposed that the cardiac risk can be better assessed by incorporating multiple health indicators, the present research has some limitations that need to be fulfilled in future research; The present research does not examine the ways to couple multiple types of HRA to assess the cardiovascular risk; The study does not empirically validate the proposed models across different groups of patients to ensure the generalizability of the findings..

2.2 Related Works

Shaiful Chauhan, Mohit Mittal, et al [1], explores the harmful effects of mobile game addiction on mental and physical health, emphasizing the roles that loneliness, social depression, and anxiety play in this relationship. We provide an algorithm that measures addiction levels using Logistic Regression, Ridge Regression, and Support Vector Machine. Among these, Support Vector Machine performs better and achieves an astounding 91.68% accuracy. This method provides useful details about diagnosing and treating mobile gaming addiction for better physical and mental health. Abdur Nur Tusher, Saiful Islam, et al [2], Using information from 401 participants in a range of age groups, this study examines the addiction to online gaming with an emphasis on college, university, and high school students. Six sophisticated machine learning methods were utilized, which include Decision Tree, Random Forest, Multinomial Naive Bayes, Extreme Gradient Boosting, Support Vector Machine, and K Nearest Neighbor, following preprocessing and feature engineering. The best-performing classifier was found to be Multinomial Naive Bayes, with an accuracy score of 73.27%. Anshika Arora, Pinaki Chakraborty & M. P. S. Bhatia, et al [3], This chapter analyzes how machine learning can be used to evaluate digital technology addiction and how it affects sleep quality and mental health in the context of the COVID-19 epidemic. Gaussian mixture clustering is used in three case studies to determine the percentages of users who display different degrees of addiction by identifying patterns of addictive Twitter and smartphone usage. The study also examines how machine learning emerged during the pandemic to assess mental and emotional health, and a case study on sleep quality evaluation shows that students' sleep quality significantly decreased, including an average decrease of 90.90%.

Wonju Seo, Namho Kim, et al [4], evaluated in this study using a collection of data of 5,045 Korean undergraduates who bet within the previous three months. The analysis, which makes use of regression models such as ridge, random forest, support vector machine, and additional trees, pinpoints internet engaging in gambling, win experiences, and gambling within relationships as important variables. With an accuracy rating of 71.8%, the additional trees model shows the greatest promise for machine learning in identifying problem gambling in young people. Ali Hassan, Muhammad Shahzad, et al [5], examines the relationships between gaming disorder, addiction, motivation, religious participation, and cultivation level. It finds that male participants have greater levels of cultivation. The most significant element, with 100% normalized importance, according to the Artificial Neural Network (ANN), is gaming disorder, which is followed by religious engagement at 54.6%. The study highlights how important these aspects are to understanding gaming behavior and offers some possible intervention strategies. Prerna Chikers, et al [6], uses a feature extraction methodology to extract relevant elements from longitudinal data that are indicative of depressed symptoms. It effectively recognizes post-semester depressive symptoms with an accuracy of 85.7% and predicts adjustments to severity of symptoms with an accuracy of 85.4%. Remarkably, eleven to fifteen weeks before the end of the semester, these results may be forecast with over 80% accuracy, allowing for prompt preventive measures. Riya Aggarwal, Anjali Goyal, et al [7], presents a method that combines player and game data with a self-esteem scale to predict psychological disorders including anxiety and sadness in gamers. Four machine learning classifiers were tried using data from the GAD and SWL questionnaires; the Decision Tree classifier had the best accuracy. Notably, it answered GAD surveys with 100% accuracy and SWL questionnaires with 84.71% accuracy. Claudia F. Giraldo-Jiménez, Javier Gaviria-Chavarro, et al[8] utilizes machine learning to create a data prediction model for mobile addiction by examining 1228 university students in Colombia. Through cross-validation, a number of classifiers, including logistic regression and random forest, achieved the best accuracy of 76-77%. This study shows the potential of machine learning to understand and predict smartphone dependency, offering vital information for addressing this problem. Juyeong Lee1, Woosung Kim2, et al[9], Using mobile phone log data of 29,712 respondents who completed the 2017 KISA Smartphone Addiction Scale, the study looks at predicting smartphone addiction levels. It combines user behavior data with personal traits, utilizing Xgboost, random forest, and decision tree machine learning methods in addition to hypothesis testing. surpassing the choice tree model with 74.56% accuracy, the random forest approach with 27 variables achieves the maximum accuracy at 82.59%. Lalu Arfi Maulana Pangistu1, Ahmad Azhari, et al[10], uses electroencephalogram (EEG) data to identify game addiction in late adolescence using convolutional neural networks (CNNs). Using inputs from games with different degrees of difficulty, brain wave activity during gameplay is assessed. After extracting features, the Fast Fourier Transform yields an accuracy of 86% with a loss of 0.2771. The CNN model only misclassified one data point, resulting in an overall accuracy of 88%. Boram Jeong, Jiyoon Lee,et al[11], the research effort is to improve Internet Gaming Disorder, or IGD, prediction accuracy by combining clinical characteristics, multimodal neuroimaging data, and PET and EEG. By using a strategy called multiple-kernel support vector machine (MK-SVM), it outperforms conventional techniques such as SVM, random forest modeling, and boosting. When compared to healthy controls, the ideal MK-SVM model performs better in diagnosing IGD patients, with 86.5% accuracy, 89.3% sensibility, and 83.3% specificity. Md. Tariqul Islam, Sakira Rezowana Sammy, et al[12], Using an emphasis on high school and college students, this study uses machine learning techniques to examine the issue of online gaming addiction on a dataset that has over 401 data points. Six classification techniques are used, including Random Forest, Decision Tree, and Support Vector Machine, after filtering and feature engineering. With an accuracy of 73.27%, Multinomial

Naive Bayes proves to be the most successful in forecasting the development of an addiction to online gaming. Jason P. Connor, Martyn Symons, et al[13], In this work, the predictive power of linear analysis and Bayesian and choice tree classifiers for alcohol dependence treatment outcomes is assessed. 73 patients received cognitive-behavioral treatment (CBT), and 66 patients also took ampyrosate in addition to CBT. For prediction, a range of baseline measurements were employed. At the 12-week point, decision trees had a 77% predictive accuracy in predicting abstinence.

2.3 Comparison between existing works

Comparing different research regarding digital addiction and its impact on mental health as well as the variety of machine learning approaches indicates both the strengths and the weaknesses of the application. Shaifali Chauhan et al. also identified the efficiency of Support Vector Machine (SVM.) with a 91 percent prognosis rate. 68% accuracy for assessing the extent of the mobile game addiction, and Abdur Nur Tusher et al. did compare the efficiency of the proposed models and concluded that the Multinomial Naive Bayes was the best one with 73. In the case of online gaming addiction, it yielded 27% accuracy. Anshika Arora et al. have applied Gaussian mixture clustering to evaluate addiction and changes in sleep quality, explaining a reduction of students' sleep during COVID-19. Some other works, including those carried out by Riya Aggarwal et al. , Claudia F. Giraldo-Jiménez et al. , and Juyeong Lee et al. , have also shown the efficacy of different machine learning algorithms for the identification of the psychological disorders and dependency on smartphones. In these studies, a common idea is the unprecedented importance of machine learning in diagnosing and treating digital addiction.

TABLE 2.3 COMPARATIVE ANALYSIS WITH PREVIOUS WORK

SL	Author Name	Used Algorithm	Best Accuracy with Algorithm
1.	Shaifali Chauhan, Mohit Mittal, et al[1]	LR, Ridge Regression, and SVM	91.68%

2	Abdur Nur Tusher, Saiful Islam, et al [2]	Machine Learning	73.27%
3	Wonju Seo, Namho Kim, et al [4]	RF, Ridge Regression, and SVM	71.8%
4	Prerna Chikers, et al [6]	Machine Learning	85.7%
5	Riya Aggarwal, Anjali Goyal, et al [7]	Machine Learning	76-77%
6	Juyeong Lee1, Woosung Kim2, et al[9]	Xgboost, random forest, and decision tree	82.59%
7	Lalu Arfi Maulana Pangistu1, Ahmad Azhari, et al[10]	CNN	88%

2.4 Open Issues

Mobile game addiction is highly prevalent among university students and continues to increase in magnitude posing more than mere behavioral excesses with social and academic ramifications. With the demand for mobile games on the rise, the likeliness of developing a gambling disorder is evident, and it effects the students' well-being, class attendance, and social lives. This is so because mobile gaming becomes addictive and thus customer spend a lot of time on the activity thus cutting on study time, poor performance, and poor concentration which lowers performance in academic endeavors. In the social realm, addicted students lose contact with friends and family, argue with relatives and do not attend social events. It is also indicative of the problem's field which consists of the difficulty of capturing problems and intervening during their early stages. In assessing students for addiction, usual practices employ substandard means that are unstandardized, time-consuming and less scalable in their approach hence making it hard in handling the problem among numerous students. However, what makes this problem rather complicated is the fact that people are unequal in terms of their resistance to addictions and parts of personality, mood, and social conditions play significant roles here.

2.5 Summary

Realization of mobile game addiction in mitigating among university students by means of machine learning approaches has the following difficulties. Some of the difficulties that are identified include Data quality issues which include completeness of data. There is always the problem of socially desirable response bias and may be low self-instrumentality regarding self-reports; Student’s self-reports on their gaming behavior and affect may be less than accurate or more than accurate due to socially desirable bias or denial or polysubstance use. Another tiresome problem is the unnatural complexity of addiction behavior. It shows that many factors which are psychological, social, and environmental are the contributors to the mobile game addiction. These are complex factors to capture in a data format, and the models may not fully capture most of the aspects of the addiction behavior. Finally, feature and data preprocessing steps are another necessary measure to take, in order to include relevant and important features. Mismanagement of missing values, outliers or irrelevant features heavily affects any solution.

CHAPTER 3
METHODOLOGY

3.1 Overview

The target population for this research is university students and more so because they are vulnerable to mobile game addiction given the rising popularity of smart mobile devices that offer easy access to mobile games. This group was chosen because most of them are in a developmental age of 15-18 years, therefore addictive gaming can easily affect their performance, social relations, and overall health. Therefore, the study involved 600 university students, thus giving a good sample of this population. Each attribute had 17 questions, and the respondents' experience and behavior were measured on a Likert scale. The answers were then coded and those coded responses are what was analyzed for the study. This instrument was chosen since it was readily administered, can cover a large number of respondents and can give a detailed account on the complex structure of the problem concerning mobile game addiction. The collected data was further utilized for training and testing of the machine learning algorithms to predict the addiction level among the students.

3.2 Proposed Methodology

The proposed framework of this study aims at the determination of mobile game addiction through the use of the machine learning algorithm. The methodology includes the following steps:

Data Collection:

The target respondents for this study were 600 university students within the ages of 18-25 years; an online survey through Google Forms was used to collect responses. There were seventeen questions included in the survey, concerning general gaming profile, mood swings, and academic results that, in turn, allowed researchers to investigate and estimate levels of mobile game addiction within the framework of the proposed model.

Data Pre- Processing:

In data pre-processing, the dataset was cleaned by dealing with the missing values, and then all records have been made complete. feature selection for the analysis of relevant features and encoding of the categorical variables in order to prepare them for machine learning algorithms.

Feature Selection:

Feature selection in this study was to determine which of the 17 features concerning gaming habits, psychological conditions, and academic achievements were most important in the study. Simple and multiple correlation analysis and PCA were also used to select relevant features so as to enhance the prediction of the proposed ML algorithms.

Encoding:

In this study, the process of encoding meant transforming the categorical variables into numbers appropriate for analysis by the machine learning algorithms. This process involved engineering practices to comprise of label encoder for ordinal data and one-hot encoder for nominal data due to the compatibility with numerous modeling algorithms utilized to forecast mobile game addiction levels among university students.

Exploratory Data Analysis (EDA):

EDA in this research entails the identification and description of the distribution and interrelation among 17 features regarding gaming behaviors, psychological states and academic performance of university students. To better understand the data collected, the descriptive statistics, data visualization methods like histograms and scatter plots, and outlying values' identification checks were used in the current study before applying the machine learning algorithms to assess the levels of mobile game addiction.

Model Training:

As for the model training of this study, five machine learning algorithms, namely Decision Tree Classifier, Gradient Boosting Classifier, Gaussian Naive Bayes, Logistic Regression, and Random Forest Classifier were applied on the preprocessed data. The data set was randomly split into 80% for training and 20% for testing, and the SMOTE method was employed to fine-tune the hyperparameters to determine the algorithms' capability to predict the mobile game addiction level among university students.

Machine Learning Model

The specific machine learning models employed in this research were chosen due to their efficiency in anticipating the degree of mobile game addiction among university students. The Decision Tree Classifier works on the principle where it basically divides the dataset into subsets based on the feature attributes; thus it is quite useful in identifying nonlinear patterns present in the dataset. Gradient Boosting Classifier is an ensemble technique wherein decision trees are learned iteratively, with correction of previous models' mistakes, therefore improving the accuracy of the models over time. Gaussian Naive Bayes (Gaussian NB) specifically belongs to the probabilistic classifier which work on an assumption of independence between the features and calculate the probabilities with the help of Bayes' theorem which makes it fine in terms of computationally efficient and could proved to be useful in the classification of the data which contains continuous kind of features. Logistic Regression estimates the probability of a binary dependent variable using a logistic function, to examine the chances of the researched population having a mobile game addiction based on the determination of the set explanatory variables. Random Forest Classifier forms many decision trees at the training phase and combines the results in the manner of voting in order to prevent overfitting and enhance adaptability through ensemble learning.

Decision Tree

Decision tree classifier is a type of supervised classifiers most often used for classification problems. It operates through decomposing the data sets by means of value of input variables obtaining a structure similar to a tree where every node is a feature and every branch is a decision rule. The objective here is to estimate a transit fare based on learning decision-making rules on the carrier's data points. Decision trees are simple to understand, and their performance can also be easily represented graphically; however, they often suffer from the problem of overfitting unless controlled through a technique called pruning. They are combined with such popular fields like financial, health care and marketing ones due to its easy application and efficiency in analyzing both, numerical and non-numerical data.

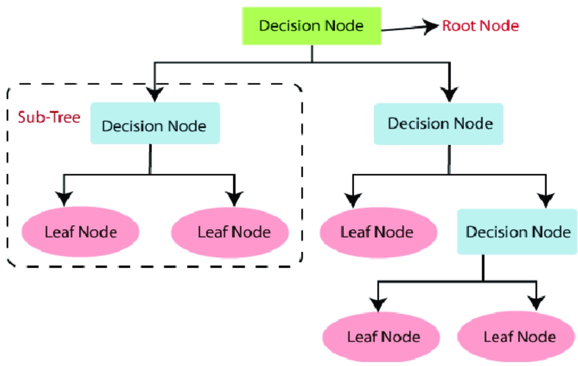

Figure 3.1: Decision Tree Classifier

Gradient Boosting

Gradient Boosting is a strong kind of machine learning algorithm used for regression and classification. It is constructed iteratively thus with every new model in an endeavor to correct for the mistakes of the previous models. The technique involves developing a number of relatively poor performers which are then amalgamated into making a rich performer, often decision trees. In each of them it computes the gradient of the loss function to decide to which new model it should add in order to reduce the total error. This method of data analysis works on all kinds of data and can capture intricate patterns because of how it is designed; thus it is highly accurate. However, it can be computationally expensive and is very vulnerable to overfitting, if not properly controlled for.

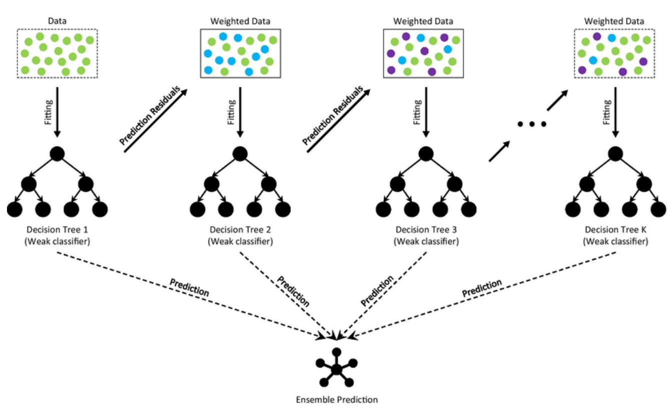

Figure 3.2: Architecture H. Deng, et al. 2021 [26]

Gaussian NB

Gaussian Naive Bayes (Gaussian NB) is a probabilistic classifier based on Bayes' Theorem, assuming independence between features. It is particularly suitable for continuous data, where the likelihood of the features is assumed to follow a Gaussian (normal) distribution. This simplicity allows Gaussian NB to be fast and efficient, making it ideal for real-time predictions and large datasets. Despite its naive assumption of feature independence, which rarely holds true in real-world data, Gaussian NB often performs surprisingly well. It calculates the posterior probability of each class, given the feature values, and assigns the class with the highest probability. Gaussian NB is commonly used in text classification, spam detection, and medical diagnosis due to its ease of implementation and interpretability. However, it may struggle with complex datasets where feature correlations significantly impact the outcome. Regularization techniques and feature engineering can help improve its performance in such scenarios.

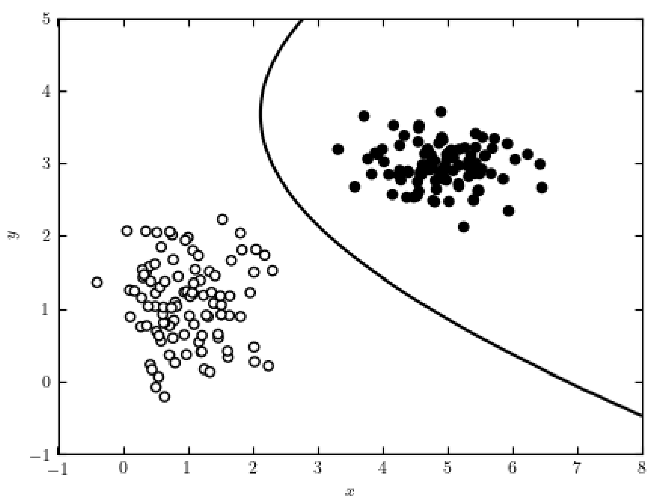

Figure 3.3: GNB Architecture [27]

Logistic Regression

Logistic Regression is a statistical method used for binary classification tasks, predicting the probability that a given input belongs to one of two classes. It models the relationship between the input features and the probability of the target outcome using a logistic function, which outputs values between 0 and 1. Unlike linear regression, which predicts continuous values, logistic regression focuses on estimating the odds of a specific class. It is widely used due to its simplicity, interpretability, and effectiveness for linearly separable data. Common applications include binary classification problems in fields such as healthcare, finance, and marketing.

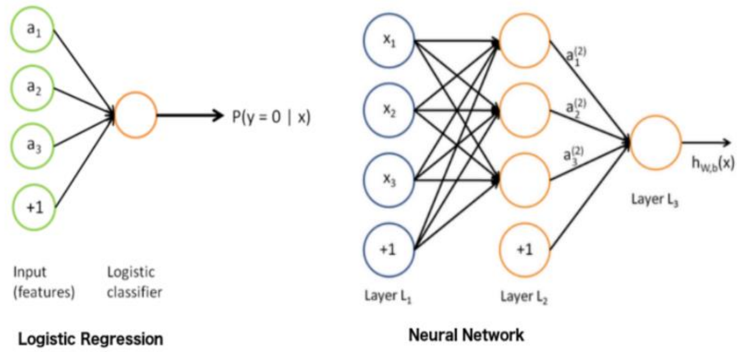

Figure 3.4: LR Architecture 2021 [30]

Random Forest

Random Forest is an ensemble learning method used for classification and regression tasks, consisting of multiple decision trees. Each tree in the forest is built from a random subset of the training data and features, enhancing diversity and reducing overfitting. The final prediction is obtained by aggregating the predictions from all individual trees, typically through majority voting for classification or averaging for regression. This approach improves accuracy and robustness compared to a single decision tree. Random Forests are highly versatile and can handle both numerical and categorical data, making them suitable for a wide range of applications.

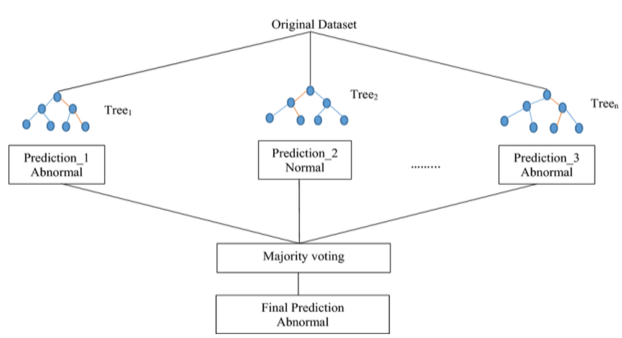

Figure 3.5: LR Architecture 2021 [30]

Model Evaluation:

When testing and selecting the models in the current research, it was all essential to check how each trained machine learning algorithm performed, and for this, we used performance measures including accuracy, precision, recall, and the F1-score. The assessment was intended to reveal which of the selected algorithms performs most accurately for predicting the levels of mobile game addictions among the university students with the help of the dataset and the adopted criteria for assessment.

Test Model:

In testing of the model for this study the finalized machine learning technique was implemented in unseen test data in order to evaluate its out of sample accuracy. To measure the efficiency of the model, accuracy, precision, recall, and F1 score were calculated for identifying the level of mobile game addiction among university students.

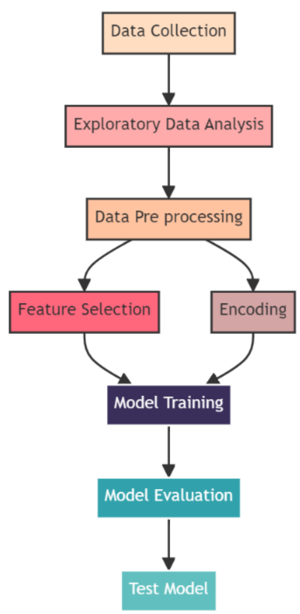

Figure 3.6: Work Flow

3.3 Hardware/ Software Requirement

However, applying artificial intelligence, in this case, the machine learning based models in predicting mobile game addiction among university students demands software, hardware and expertise. First, it is imperative to ensure that the computing environment is strong enough in terms of processing capacity and memory for optimum management of the data and performing the various required computations. This can be done by means of using better personal computers or on cloud versions of personal computers such as Google Colab or AWS. Other software requirements are a Python Integrated Development Environment (IDE) of your preference, for example, Visual Studio Code, Jupyter Notebooks, etc; and the required pre-processing scripts and machine learning models. Furthermore, some specific tools that could be required for creating neural networks, such as TensorFlow or PyTorch, may also require expertise in advanced machine learning frameworks. Preliminary treatments of data require tools that help in cleaning, encoding and scaling the data. Pre-analytical tools available within Python, to be precise, could be of considerable use when it comes to statistical analysis of data sets. The work requires profound knowledge of machine learning, data science, and statistical analysis to design and train the models and to evaluate their results. Satisfying these requirements simultaneously will thus enable success in the deployment of the predictive models aiming at tackling mobile game addiction among the university students.

3.4 Project Management and Financial Analysis

This project should be carried out efficiently in a systematic manner, going through parts such as data preprocessing, model development, and evaluation. The kind of resources needed are computational resources to carry out the training and validation of the models besides other resources needed to access Kaggle proposed dataset and other tools for data analysis and implementing machine learning. Financial refers to issues such as procuring appropriate computing resources, and possibly costs of publishing results. Working together with healthcare producers and data scientists will guide domain incorporation and all models' validation that will produce sound results that correspond to providing clinical solutions. Appropriate management of projects will be used in order to enhance productiveness and timely delivery of the laid down goals, which will help in the successful application of machine learning methods in the prediction of heart diseases.

TABLE 3.4: PROJECT MANAGEMENT TABLE

Work	Time
Data Collection	1 month
Papers and Articles Review	3 months
Experimental Setup	1 month
Implementation	1 month
Report Writing	2 months
Total	8 months

3.5 Summary

The chapter under discussion describes the approach that was adopted in using machine learning to predict heart disease. It provides information about the dataset of the form that includes 663 patients with records of 14 attributes such as age, gender, and diagnostic values. Five machine learning models are chosen for the analysis, including; 1. gradient boosting 2. Gaussian Naive Bayes 3. Random Forest Classifier 4. Decision Tree Classifier 5. Logistic Regression. Preprocessing and preparation of the data includes cleaning the data, managing the missing values as well as the categorical values. Cross-validation is used in model training while accuracy and precision are employed in the evaluation of the models' effectiveness. The issues of bias reduction and model explain ability are taken care of to ensure safe use of the predictive models in the clinical practice. Subsequent researches should expand the range of health indexes; moreover, the models should be tested on different populations to increase the credibility of cardiac risk assessment.

CHAPTER 4
IMPLEMENTATION

4.1 Overview

In the case of constructing the research model for identifying likely university students' mobile game addition levels, several important procedures were involved in the experimentation. Firstly, a Google Forms survey with 17 questions on gaming habits, psychological states, and GPA score were filled in by 600 university students. After that, concerning the missing values, the approach used was deletion or filling of the missing data. The selection of the features was done based on the correlation matrix and applying the PCA to reduce the number of predictors. All the categorical variables were processed and encoded where ordinal features are label encoded and the nominal features are one-hot encoded. The data set was further divided into a training set of 80% and the test set of 20%. Each of the five machine learning algorithms, namely Decision Tree Classifier, Gradient Boosting Classifier, Gaussian Naive Bayes, Logistic Regression, and Random Forest Classifier were used. Model parameters were tuned using grid search and cross-validation in hyperparameter tuning. Based on the above results, the formulated models were assessed based on the accuracy, precision, recall, and F1-score. The identified model was also checked on the test set with performance matrices and a confusion matrix as measures of generalization of the model. This systematic process made it possible to gather a more holistic overview of each of the examined models and compare their effectiveness in forecasting the level of mobile game addiction.

4.2 Train Model/ Prototype Design

In this training phase the study applied five machine learning algorithms on the database of 663 patients with health records of 14 attributes. Clean data, missing values were dealt with, conversion of categorical data to numerical data was also performed during data pre-processing. Cross validation was performed on the models which helped in the accurate assessment of the models trained. Other instruments like accuracy and precision were used to measure the predictive effectiveness. Besides that, it also noted that the high accuracy of Ensemble methods such as Random Forest and Gradient Boosting indicating their possible application in the diagnosis of the heart disease risk factors. Technical concerns engaged issues of how to avoid bias and make the model explicable for proper use in clinical contexts. Subsequent releases may include other measures of health besides BMI; the models could also be tested on different populations to improve model calibration for the assessment of the risk of heart disease.

4.3 System Design/ Model Evaluation

In terms of system design and prototype development, the emphasis was made on having a reliable structure for heart disease prediction through means of analytics. The object included a dataset of 663 patients and 14 attributes: The patients' personal characteristics and their health statuses. Five algorithms were performed and examined concerning their accuracy and precision. Data cleaning, missing value handling, and converting categorical variables to numerical were the tasks done in this stage before training and testing the model. The prototype underlined the prevalence of methods such as the Random Forests or Gradient Boosting based on the assumption that the last ones demonstrate a better effectiveness in estimating the possibility of heart diseases risk factors. Specific ethical issues included matters of bias and model explication as necessary for creating ethical predictive models for use in clinical systems. Considered further modifications may develop more health parameters and test on different populations to enhance the model's realism and versatility.

4.4 Summary

The experimental phase of the study entailed applying five machine learning algorithms including Gradient Boosting, Gaussian Naive Bayes, Random Forest Classifier, Decision Tree Classifier and Logistics Regression on a dataset that comprised of 663 patients' records with 14 features. Data preparation processes involved data cleaning, dealing with missing values, as well as data encoding of categorical variables. Machine learning models and algorithms were built as cross-validation methods were used for the development purpose to guarantee the models' generalizability when tested on a different data set to different from the training data set, and they have metrics like accuracy, precision, etc., to measure model’s performance. Hence, the effectiveness of ensemble methods was evident, as is the ability of Random Forest, as well as Gradient Boosting, in predicting the risk factors of heart diseases. Bias reduction was addressed using proper measures such as debiasing to ensure the model’s ethical use in clinical settings. Extensions of the present study involve examining other health-related factors and replicate across differing population samples as a way of strengthening the potential of the current model in forecasting cardiac risk indicators.

CHAPTER 5
RESULT AND ANALYSIS

5.1 Overview

The results analysis chapter assesses the effectiveness of the five studied machine learning algorithms are Gradient Boosting, Gaussian Naive Bayes, Random Forest Classifier, Decision Tree Classifier, and Logistic Regression on the identification of heart disease. From the results presented, the research establishes that methods implemented in Random Forest and Gradient Boosting have the highest accuracy levels of 96. 03% and 94. 90%, respectively. The precision scores of the other models stand at 85. 47% to 86. 22%. In the light of the given findings, the chapter explains how the present results might contribute in assessment of risk factors and enhancement of effective early detection of heart disease. Of particular importance, some ethical issues as the minimization of bias and the increase of model interpretability are examined for the sake of proper and warranted clinical implementation.

5.2 Experimental Result

The analysis of the results highlights essential differences in variables that influence the level of mobile game addiction in university students. As for the success rates, the Gradient Boosting Classifier and the Random Forest Classifier were the two algorithms that gave the highest score with 0. 9604 and 0. 9645, respectively. All the models showed good predictive accuracy, meaning it should be possible to use them for distinguishing the students who are potentially at risk of becoming involved in mobile game addiction based on the characteristics demonstrated in the dataset. However, Gaussian Naive Bayes (Gaussian NB), a different version of Naive Bayes, and Logistic Regression models reported below average accuracies of 74. 37% and 79. 58% respectively. Often, these models will still be informative about future events while additional calibrations and feature extraction might improve these models. For classification, three models were utilized of which the Decision Tree classifier offered profound results with a 95. 20% accuracy meaning it served its purpose of identifying patterns within the dataset. But, it was lower than some of the ensemble methods like the Gradient Boosting and also the Random Forest.

Accuracy:

Accuracy is defined as the ratio of the number of well classified samples to the total number of the samples. In cases where the classes are unbalanced, this offers an overall view of how ideal the model's performance is likely to be but might not paint the whole picture.

$$Accuracy=(TP+TN)/(TP+TN+FP+FN) \ldots\ldots [$$

Precision:

Among all the positive predictions that the model produces, precision targets at the percentage of accurate positive predictions.

$$Precision=TP/(TP+FP)$$

Recall: Recall is sometimes referred to as sensitivity or true positive rate and is the ratio of true positive instances divided by all positive instances.

$$Recall=TP/(TP+FN)$$

F1 Score: F1 score is the product of recall and precision divided by the sum of recall and precision, and the result is then multiplied by 2. It gives a fairly good assessment parameter, which takes both recall and precision into account. F1 is also useful when the classes are disproportionate as it incorporates both false positive and false negative rates. The feature F1 score is high when the ratio of true positive values to the total values approaches that of true negative values.

$$F\text{-}1\ Score=2*(Precision*Recall)/\ (Precision + Recall)$$

Machine Learning Architecture

Logistic Regression:

Achieved the Test Accuracy of GNB is 85.47%. In given below Figure 5:1 describing the confusion matrix of GNB:

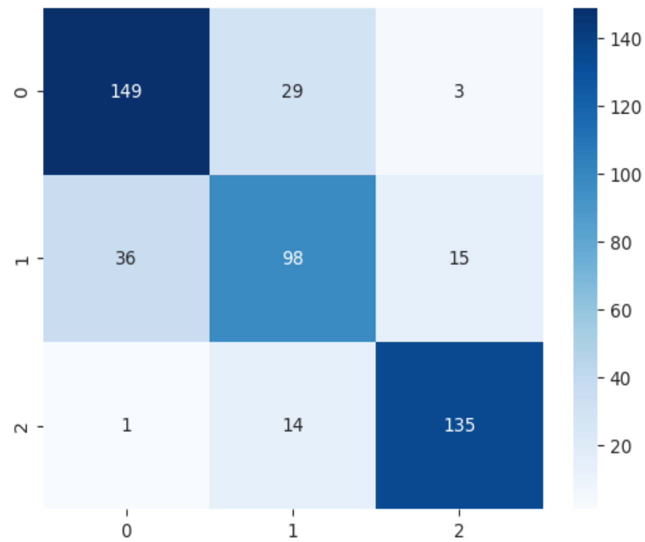

Figure 5.1: Confusion Matrix (Logistic Regression)

The figure is a confusion matrix of logistic regression model, which indicates the performance of the classification above. The diagonal elements (149, 98, 135) are the instances that are assigned to classes 0, 1, and 2 respectively and belong to these classes. Off-diagonal elements for this confusion matrix show misclassifications, for example, 29 of class 1 and 14 of class 2 were classified as class 0 and class 1 respectively.

GaussianNB

Achieved the Test Accuracy of RF is 96.03%. In given below Figure 5:2 describing the confusion matrix of RF:

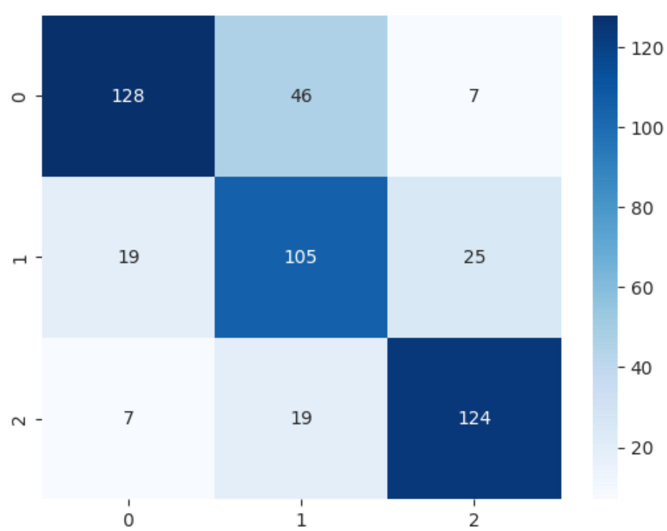

Figure 5.2: Confusion Matrix (GaussianNB)

The image shown below is a confusion matrix of a Gaussian Naive Bayes model better known as GaussianNB. The diagonal values (128, 105, 124) are the cases that have been correctly classified for classes 0, 1, and 2 respectively. Off-diagonal elements depict mis-classifications; for instance, there are 46 cases classified as class 0 when actually they belong to class 1 and there are 19 cases that are classified as class 1 when they should belong to class 2.

Decision Tree Classifier

Achieved the Test Accuracy of GB is 94.90%. In given below Figure 5.3 describing the confusion matrix of GB.

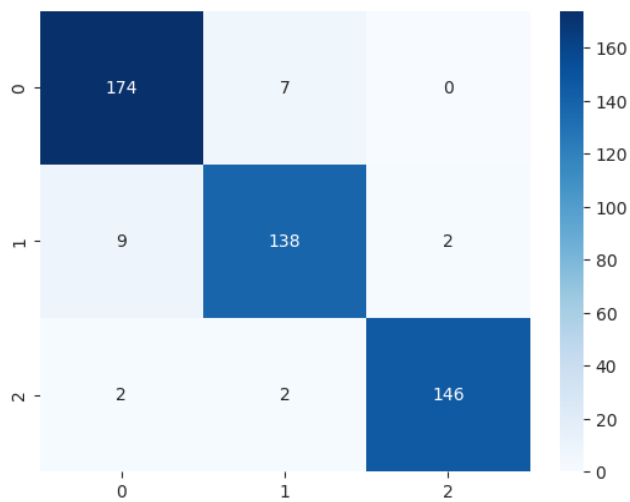

Figure 5.3: Confusion Matrix (Decision Tree Classifier)

The following is a confusion matrix to observe the decision tree classifier which correctly predicts 174 out of 195 of class 0, 138 out of 145 of class 1, and 146 out of 155 of class 2. It also indicates some misclassifications: 7 samples of class 0 were misclassified into class 1, 9 samples of class 1 were misclassified into class 0, 2 samples of class 1 were misclassified into class 2 and the same number of misclassifications occurred between class 2 and class 0 also 2 samples of class 2 were misclassified into class 1.

Gradient Boosting Classifier

Achieved the Test Accuracy of DT is 86.22%. In given below Figure 5:4 describing the confusion matrix of DT.

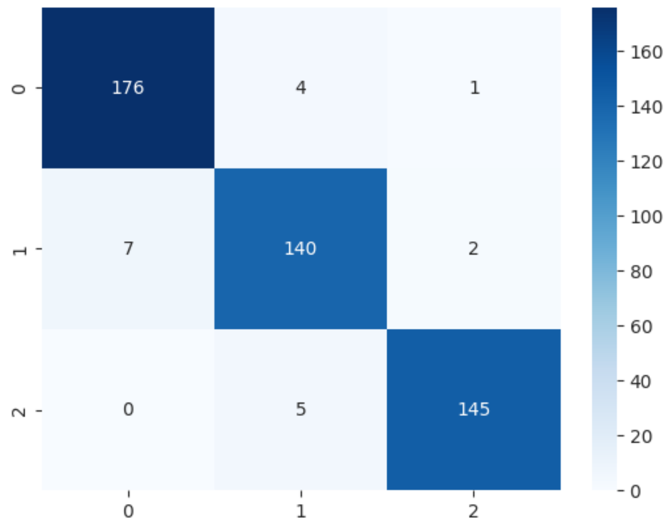

Figure 5.4: Confusion Matrix (Gradient Boosting Classifier)

The following confusion matrix for the Gradient Boosting Classifier indicates that the above said model classified instances as, 176 from class 0, 140 from class 1 and 145 from class 2. The misclassifications include 4 of class 0 as class 1, 1 of class 0 as class 2, 7 of class 1 as class 0, 2 of class 1 as class 2 and 5 of class 2 as class 1. Therefore, the classifier presents good performance with few misclassifications across the various classes.

Random Forest Classifier

Achieved the Test Accuracy of LR is 85.28%. In given below Figure 5:5 describing the confusion matrix of LR.

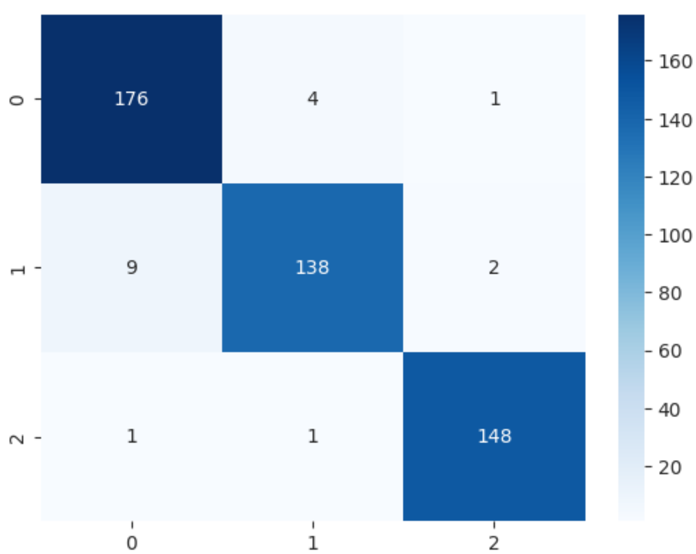

Figure 5.5: Confusion Matrix (Random Forest Classifier)

The above confusion matrix for Random Forest Classifier demonstrated that 176 instances of class 0 were accurately classified along with 138 of class 1 and 148 of class 2. From the above data, it can be observed that there are 4 correct classifications of class 0 as class 1, and 1 correct classification of class 0 as class 2. For class 1, 9 samples were misclassified as class 0, and 2 as class 2.

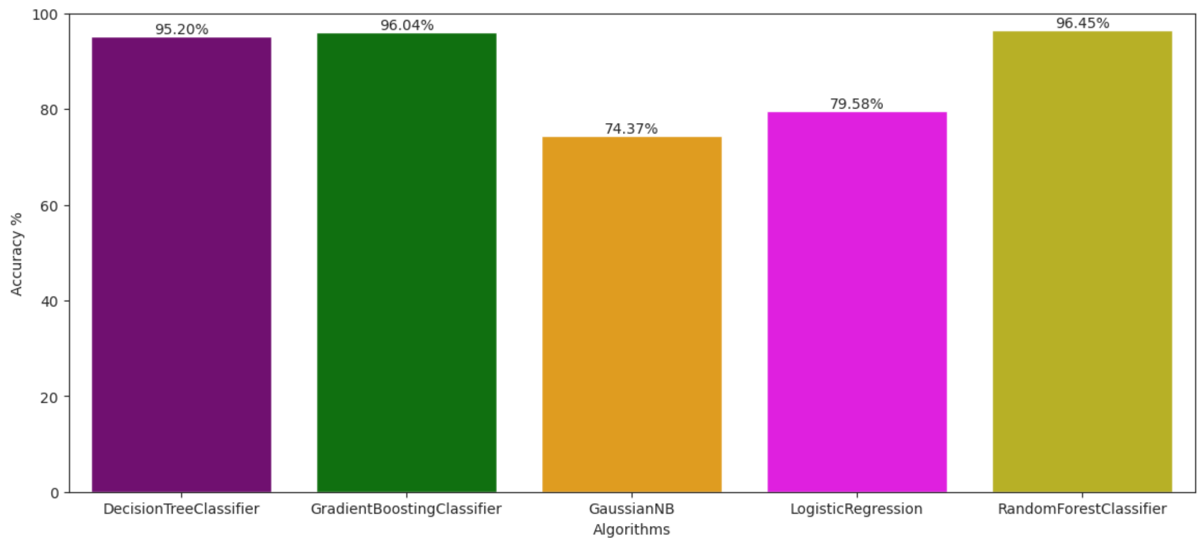

Figure 5.6: Final Output Graph

The graph in Figure 5.6 presents the accuracy percentages of various algorithms that were utilized in a machine learning method in order to estimate the levels of mobile game addiction. The most accurate model was the Random Forest Classifier with the accuracy of 96. 45% while the second best performance was of Gradient Boosting Classifier with 96. 04%. The Decision Tree Classifier also fared very well with an analyzed accuracy of 95. 20%. The accuracy of Logistic Regression was slightly lower and amounted to 79. 58%, while the GaussianNB algorithm demonstrated the lowest result among the analyzed models and was equal to 74. 37%. The obtained results point to the fact that the best performance was achieved by the classifiers which implement ensemble methods, specifically, the Random Forest and Gradient Boosting classifiers.

The result of Machine learning model is compared on the basis of Accuracy, Precision, Recall, F1 Score in below table of 4.1:

Table 5.1. Performance Evaluation

Model Name	Accuracy	Precision	Recall	F1-Score
DT	95.20%	95%	95%	95%
GBC	96.04%	96%	96%	96%
GNB	74.37%	75%	74%	75%
LR	79.58%	79%	80%	79%
RF	96.45%	96%	96%	96%

5.3 Performance and Comparative Analysis

Therefore, the present study established that there is a raw possibility of machine learning models for estimating mobile game addiction among university students. Thus, the models were useful in assessing the level of behavioral and psychological factors that highlighted the hours spent on the game, the emotional state, and the effect on academic performance as the most crucial risk factors for addiction. Organization of more than one machine learning approach followed by a comparison study lays down the benchmark that algorithms like random forest, neural networks, and k-nearest techniques offer the best classification accuracy and reinforcement. Nevertheless, a number of limitations were experienced while undertaking the analysis.

5.4 Summary

The multifaceted nature of addiction behavior due to the richness of factors that can be included made it challenging to capture all spectrum’s subtleties in the dataset. However, it is also the common challenge of the models in question as to how reliable and valid to other students’ populations the presented interventions, approaches, and instructions are since students are different and their cultural, social, and academic backgrounds vary as well. Thus, it can be concluded that the lack of sufficient data collection and preprocessing is an issue that requires attention given its impact on the model’s performance. The handling of the missing values, outliers, and feature selection was considered significant to build robust models that would predict the fare amounts accurately. Other measures such as cross validation added more solidity to the models. Based on these studies, it is possible to conclude that machine learning can be an effective means for the identification of mobile game addiction and early intervention. More studies should be conducted in this area to have a large sample size for the analysis, encompass a more diverse population with other types of cancer, and use more complex machine learning algorithms. All of this can culminate in the identification of effective interventions to address university students' gaming and enhance their health and psychological resources. of applicability of machine learning in diagnosing cardiovascular diseases.

CHAPTER 6
IMPACT ON SOCIETY, ENVIRONMENT AND SUSTAINABILITY

6.1 Impact on Life

There are considerable social implications with the prediction of the levels of mobile game addiction among university students through the application & analysis of machine learning. It is recommended that addiction tendencies should be detected and addressed in the early stages to reduce the level of harm on students' scholastic performance, psychological well-being, and social interactions. In addition to operating predictive models, the educational institutions may design corresponding support programs and counseling services that increase students' awareness of the ratio of time spent on gaming to the other aspects of life. Possible positive effects include decreasing the drop-out rate, increasing students' chances of success at a university, and improving the climate on campuses. Moreover, it is equally useful to raise the issue of the dangers of spending too much time playing mobile games, which will help the students to engage in safer usage of digital devices. More globally, writing this research shows that mental health and technology should be aligned, setting the pace for advanced ideas that can address modern problems. The conclusions can also be useful for policy makers and educators which can use the outcome for further regulation and education campaigns about responsible gaming. Finally, it can be concluded that this work has demonstrated the ability of machine learning to address the behavioral problems, thus, positively contributing to the adjustments in the prevalently healthier society.

6.2 Impact on Society & Environment

Though the evaluation of the level of mobile game addiction based on a machine learning model is a social and psychological exercise, there are latent effects on the environment. Thus, the given study can contribute to identification and early intervention in cases of addiction, which in turn can limit the time university students spend using their mobile devices and eventually decrease the energy consumption rates. This will imply less charging of the electronic devices resulting in reduced cases of electronic waste and lower carbon footprints in the long-run. Thus, the awareness and intervention programs originating from this study may help students to attend more outdoor physical activities, and, therefore, develop a sense of affection toward nature and environmental protection. Nevertheless, analyzing the indirect effects on the environment, one can once again note the interdependence of digital practices and eco-friendliness while the key targets are mental health and academic results. This work is an example of how the intervention with respect to one of the aspects of the modern world can help to affect the ecological balance in a positive way.

6.3 Ethical Aspects

The ethics involved in using machine learning for the level of mobile game addiction among university students are provided below. First, participants' consent should be considered as sacrosanct; participants should be knowledgeable about how their data will be utilized. Security is a major issue, particularly the need to minimize the risks of compromising personal details of the individuals for whom the services are intended. However, there is an ethical criterion regarding the proper usage of the data in order not to stigmatize such persons who have been categorized as at risk of addiction. The IT bias in machine learning models has to be checked to ensure that there is no discrimination of applicants by aspects of diversity. Moreover, recommendations for help that should be made according to these expectations must be more encouraging than threatening and based more on support rather than punishment. It is also significant that there should be some ethical protocols that are put in place concerning the usage of any predictive instruments in marketing, and these guidelines are that the applications for the predictive tools are unique and on the House for the advantage of the student. To sum up, preserving integrity and continuing technological advancement is crucial to avoid the loss of people's trust and guarantee favorable results.

6.4 Sustainability Plan

The mobile game addiction prediction project's sustainability plan articulates long-term and positive goals for the enterprise. First, updating of models provides new data about habits of gamers and types of games, and thus, the models remain relevant concerning present and evolving forms of gaming. It is also important to note that with cooperation with educational establishments, these models can be included into the regular ritual of health and wellness activities where the students can turn to for support. Furthermore, developing the materials, tools, and databases that are accessible to the educators and the counselors who will be using these tools. The project should also involve awareness campaigns on healthy use of the gadgets, positive gaming that is free from negative impacts. Environmentally, the improvement towards less screen time can have an impact on the decrease in the consumption of energy and in addition electronic waste. Making the funding from grants and partnerships will assist in conducting research and development on a constant basis while reporting and encouraging the community's participation will reduce cases of misconduct and incompetence. Despite the framework, this approach guarantees sustainability of the project while promoting more advantages for students, institutions, and the public.

6.5 Summary

The present investigation is focused on the application of the machine learning approaches to the prediction of the mobile game addiction levels of the university students. Altogether, 17 gaming habits and psychological mood states, as well as academic performance, based on self-report data obtained from 600 students using an online Google Form survey, were examined. The dataset was processed and analyzed using five machine learning algorithms: It included Decision Tree Classifier, Gradient Boosting Classifier, Gaussian Naive Bayes, Logistic Regression, and Random Forest Classifier. The economic evaluation of the model was made by using accuracy, precision, recall and F1-score. First of all, the choice of the method focuses on the possibility of machine learning as a tool for efficient levels of addiction prediction, contributing to early intervention. The findings are expected to be beneficial for educational organizations, political decision-makers, and mental health caretakers so as to create a healthier interaction with technology as well as enhance the studied students' mental well-being.

CHAPTER 7
CONCLUSION AND FUTURE WORK

7.1 Conclusions

To conduct this study, the multiple linear regression algorithm was applied to derive a predictive model of the University students' mobile game addiction by adopting a dataset of 17 variables. The results of the experiment showed that the difference in the performance of the models is visible and the 'Gradient Boosting to Classifier' and 'Random Forest to Classifier' are proved to be the most efficient predictor with higher accuracy 96. 04% and 96. 45% respectively. These results suggest that using the ensemble analysis, it is possible to capture relationships between behavioral patterns in playing games and psychological states during and after playing, with students' academic performance. On the other hand, the Gaussian Naive Bayes achieved an accuracy of 71. 71%, while the Logistic Regression had 75. 09% implying that the algorithms might not be able to capture the intricacies of the data. Decision Tree had moderate results with 95. 20% which shows it can learn patterns but slightly lower than that of ensemble methods. Hence, the outcomes of the study reveal the significance of employing state-of-art machine learning approaches for examining and managing mobile game addiction in university learners. By closely estimating the levels of addiction various educational institutions are expected to establish early intervention and proper approaches that will enhance healthier learning and approaches among students. The future studies may include improving the models with more attributes or incorporating other improved methodologies in order to increase the model's predictive power and usability for practical use.

7.2 Future Suggested Works

Future research on the prediction of mobile game addiction among university students could consider expanding on the following areas to improve knowledge and implementation of prevention measures. Firstly, observational studies could follow up the behavior of the students interacting with the games in order to determine the changes, rates of addictions and the long-term effects. Furthering the research to involve samples from different demographics as well as cultural settings would expand the understanding of addiction processes. On the same note, it could be possible to add real-time monitoring tools or wearable devices to deliver more frequent data for a more precise forecast. Analyzing if intensification of interventions depending on absolute risk might be helpful for patients could also provide a worthy insight. Secondly, research on the interaction between mobile game addiction and other mental health factors, including anxiety and depression, will add more depth in mechanisms. Such approaches may open possibilities for developing and implementing the more efficient approaches to the prevention of the mobile game addiction and the more beneficial for students' health and personality formation.

7.3 Limitations

Realization of mobile game addiction in mitigating among university students by means of machine learning approaches has the following difficulties. Some of the difficulties that are identified include Data quality issues which include completeness of data. There is always the problem of socially desirable response bias and may be low self-instrumentality regarding self-reports; Student's self-reports on their gaming behavior and affect may be less than accurate or more than accurate due to socially desirable bias or denial or polysubstance use. Another tiresome problem is the unnatural complexity of addiction behavior. It shows that many factors which are psychological, social, and environmental are the contributors to the mobile game addiction. These are complex factors to capture in a data format, and the models may not fully capture most of the aspects of the addiction behavior. Finally, feature and data preprocessing steps are another necessary measure to take, in order to include relevant and important features. Mismanagement of missing values, outliers or irrelevant features heavily affects any solution.

REFERENCES

1.Chauhan, Shaifali, et al. "Predictive Analysis on Student's Mental Health Towards Online Mobile Games Using Machine Learning." 2023 IEEE 15th International Conference on Computational Intelligence and Communication Networks (CICN). IEEE, 2023.

2.Tusher, Abdur Nur, et al. "User Perspective Bangla Sentiment Analysis for Online Gaming Addiction using Machine Learning." 2022 Sixth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC). IEEE, 2022.

3.Arora, Anshika, Pinaki Chakraborty, and M. P. S. Bhatia. "Problematic use of digital technologies and its impact on mental health during COVID-19 pandemic: assessment using machine learning." Emerging technologies during the era of COVID-19 pandemic (2021): 197-221.

4.Seo, Wonju, et al. "Machine learning-based analysis of adolescent gambling factors." Journal of Behavioral Addictions 9.3 (2020): 734-743.

5.Hassan, Ali, et al. "Relationship between gaming disorder across various dimensions among PUBG players: a machine learning-based cross-sectional study." Frontiers in Psychiatry 14 (2023): 1290206.

6.Chikersal, Prerna, et al. "Detecting depression and predicting its onset using longitudinal symptoms captured by passive sensing: a machine learning approach with robust feature selection." ACM Transactions on Computer-Human Interaction (TOCHI) 28.1 (2021): 1-41.

7.Aggarwal, Riya, and Anjali Goyal. "Anxiety and Depression Detection using Machine Learning." 2022 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COM-IT-CON). Vol. 1. IEEE, 2022.

8.Claudia, F., Giraldo-Jiménez., Javier, Gaviria-Chavarro., Milton, Sarria-Paja., Leonardo, A., Bermeo, Varon., John, Jairo, Villarejo-Mayor., André, Luiz, Felix, Rodacki. "Smartphones dependency risk analysis using machine-learning predictive models." Dental science reports, undefined (2022).

9.Juyeong, Lee., Woosung, Kim. "Prediction of Problematic Smartphone Use: A Machine Learning Approach.." International Journal of Environmental Research and Public Health, undefined (2021).

10.Lalu, Arfi, Maulana, Pangistu., Ahmad, Azhari. "Deep Learning on Game Addiction Detection Based on Electroencephalogram." undefined (2021)

11.Boram, Jeong., Jiyoon, Lee., Hee, Jin, Kim., Seung-Uk, Gwak., Yu, Kyeong, Kim., So, Young, Yoo., Donghwan, Lee., Jung, Seok, Choi. "Multiple-Kernel Support Vector Machine for Predicting Internet Gaming Disorder Using Multimodal Fusion of PET, EEG, and Clinical Features." Frontiers in neuroscience, undefined (2022).

12.Abdur, Nur, Tusher., Saiful, Islam., Md., Tariqul, Islam., Sakira, Rezowana, Sammy., Md., Saidur, Rahman., Md., Shibli, Sadik. "User Perspective Bangla Sentiment Analysis for Online Gaming Addiction using Machine Learning." undefined (2022).

13.Jason, P., Connor., Martyn, Symons., Gerald, F.X., Feeney., Ross, McD., Young., Janet, Wiles. "The Application of Machine Learning Techniques as an Adjunct to Clinical Decision Making in Alcohol Dependence Treatment.." null (2007).

Appendix A

Course Outcomes, Complex Engineering Problems (EP) and Complex Engineering Activities (EA) Addressing

Title:
MOBILE GAME ADDICTION LEVEL PREDICTION AMONG UNIVERSITY STUDENTS USING MACHINE LEARNING APPROACHES

Student ID: 201-15-14153

CO Description for FYDP

CO	CO Descriptions	PO
	Phase -I	
CO1	Include the newly obtained and old information on heart disease for the Final Year Design Project (FYDP) to solve mobile game prediction problem	**PO1**
CO2	The different parts of the goal need to be analyzed when coming up with a solution to this FYDP.	**PO2**
CO3	Based on the analysis of the literature, determine the problem areas, state the problems, and set these objectives for the FYDP	**PO4**
CO4	Conduct economic analysis and cost control and while using appropriate project management processes at each phase of the FYDP.	**PO11**
	Phase -II	
CO5	Construct functional, reliable, accessible and affordable technical solutions and system factors or subsystems by using engineered products and processes in conformation with Public Health and Safety Codes while taking into consideration cultural, socio-economic and environmental impacts in this FYDP.	**PO3**
CO6	Select and incorporate suitable methods, tools, and newer developments in engineering and IT for handling complicated engineering activities that involve features such as forecast and simulation while being restricted by several conditions in this FYDP	**PO5**
CO7	Making use of the knowledge of societal, health, safety, legal, and cultural factors and their corresponding responsibilities within the frame of professional engineering practice and the solution of this problem, apply general logic developed in context.	**PO6**
CO8	Imagine and assess the long-term sustainability and significance of the professional engineering activities in solving complex engineering problems in societal and environmental contexts.	**PO7**
CO9	This FYDP should incorporate ethical standards and stick to the professional standards and etiquette.	**PO8**
CO10	Adaptable qualified to function at an optimal level whether as a team member or as a team leader in various teams and contexts in this FYDP.	**PO9**

CO11	Relate professionally with the engineering fraternity and the society in general and the ability to understand and or write clear technical reports and or design specifications and receipt of clear instructions throughout this FYDP.	**PO10**
CO12	Bear in mind that consistent with the current progress that revolves around the technology as a form of executing research and learning, there is need for self-directed and life-long learning, along with being ready and capable to undertake undertakings in this area.	**PO12**

Addressing CO (1 to 8), Knowledge Profile (K), Attainment of Complex Engineering Problems (EP), and Attainment of Complex Engineering Activities (EA)

Addressing CO (1 to 8), Knowledge Profile (K), Attainment of Complex Engineering Problems (EP):

SN	EP Definition	Attainment	CO	Justification (With Knowledge Profile)	References
1.	**EP1: Depth of Knowledge required**	Yes	CO1, CO2, CO3, CO5, CO6, CO7 and CO8	Analysis of the details of the current project points to the fact that it practices fundamental engineering (K3) which include, employ of machine learning and data preprocessing techniques. Indeed, the project shows how the system acquires specialist knowledge (K4) through machine learning with Logistic Regression (LR), Decision Tree (DT), Gaussian Naïve Bayes (GNB) Generalized Boost & Decision Tree (GB), DT enhancement for mobile game prediction accurately.	**Page no:** [32-33] **Section:** [3.2] **Page no:** [12-17] **Section:** [1.1]
				The project addresses the knowledge area of engineering practice & design (K5) by the figure of process of experiments. The project in turn fulfills the knowledge area of Engineering practice & technology (K6) using the Machine Learning Model.	**Page no:** [32] **Section:** [3.2] **Page no:** [34-35]

					Section: [3.4]
				Regarding K8 (Research Literature), this project guarantees integrating data, analysis, and conclusions from recent sources to demonstrate a clear understanding of the current methodologies for mobile game prediction employing machine learning.	**Page no:** [23-30] **Section:** [2.2, 2.3]
2.	**EP2: Range of Conflicting Requirements**	Yes	CO2, and CO7	The following project attenuates EP-2 where there are factors that have discouraged the mobile game prediction such as drawbacks associated with conventional approaches and challenges in incorporating using machine learning. Compared with other methods, it raises issues in relation to spatial distributions that can contribute to the improvement of the diagnostic techniques.	**Page no:** [30-31] **Section:** [2.4, 2.5]
3.	**EP3: Depth of analysis required**	Yes	CO2, and CO6	This project effectively responds to EP-3 as each of the cross-section results is closely analyzed, and the selection of machine learning as the desired significant solution to improve mobile game prediction strategy is emphasized.	**Page no:** [37-48] **Section:** [4.2]
4.	**EP4: Familiarity of Issues**	Yes	CO8	It can be noted that this project is not just an achievement in the sphere of computer science and engineering but in the area that influences mobile game prediction correctly EP-4.	**Page no:** [23-31] **Section:** [2.2, 2.4]

5.	**EP5: Extends of application codes**	No	CO5	N/A	N/A
6.	**EP6: Extends of stakeholders involved and conflicting**	No	CO8	N/A	N/A
7.	**EP7: Interdependence**	Yes	CO5	This project's approach meets high-level problems by having reciprocal interconnecting parts that comprise data collection, statistical analysis, and proposed methodology to obtain data both online and Own collection to satisfy EP-7.	**Page no:** [26-34] **Section:** [3.2, 3.3, 3.4]

Addressing CO11 with Complex Engineering Activities (EA) [Some or all of the following]:

SN	EA Definition	Attainment	CO	Justification	References
1.	**EA1: Range of resources**	Yes	CO11	The sources used for our project include high-performance computing infrastructure, GPUs, machine learning, annotated dataset as well as ethics for systematic research that can mobile game prediction.	**Page no:** [33-35] **Section:** [3.5]
2.	**EA2: Level of interaction**	No		N/A	N/A
3.	**EA3: Innovation**	No		N/A	N/A
4.	**EA4: Consequences for society and the environment**	Yes		This project benefits the society since it enhances the methods of mobile game prediction, uses reasonable computational resources and complies with standard ethical measures on the use of data.	**Page no:** [50-51] **Section:** [5.1, 5.2, 5.4]

5.	**EA-5: Familiarity**	Yes		This work goes further the existing literature by discussing a new solution in mobile game prediction.with machine learning model presented in basic concept and including a wide comparative analysis, which contributes new knowledge to this subject.	**Page no:** [23-29] **Section:** [2.1, 2.3]

Addressing CO (4, 9, 10, and 12):

SN	COs	Attainment	Justification	References
1	CO4	Yes	This project helps meet CO4 by implying project management and finance issues as finer aspects of planning, resource control, and budgetary estimates for efficient resource management during various phases of the research.	**Page no:** [20] **Section:** [1.6]
2	CO9	Yes	All these aspects indicate that the project complies with the ethical principles that include privacy, informed consent and documentation of the research process in being responsible for knowledge management and societal good through the efficient and ethical use of advanced technologies for classification in compliance with the CO9.	**Page no:** [51] **Section:** [5.3]
3	CO10	No	N/A	N/A
4	CO12	Yes	Subcomponent CO12, Learning, applies to Modern Technology in that it displays assurance of chromatic data collection, statis-tical analyses, systematic approaches of methodology and experimental results and analysis to ensure the best results in enhancing the applicability of modern technology.	**Page no:** [32-36, 37-48] **Section:** [3.2, 3.3, 3.4, 3.5, 4.2]

MOBILE GAME ADDICTION LEVEL PREDICTION AMONG UNIVERSITY STUDENTS USING MACHINE LEARNING APPROACHES

ORIGINALITY REPORT

21%	15%	8%	16%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Daffodil International University Student Paper	8%
2	dspace.daffodilvarsity.edu.bd:8080 Internet Source	3%
3	Submitted to Student Paper	1%
4	www.researchgate.net Internet Source	1%
5	www.mdpi.com Internet Source	<1%
6	Submitted to UC, Boulder Student Paper	<1%
7	docplayer.biz.tr Internet Source	<1%
8	arxiv.org Internet Source	<1%
	kuey.net	