

Research Article

Cite this article: Rahman, M., Sarwar, H., Hasan, M. K., & Tin, T. T. (2026). Analyzing Digital Addiction through Feature Selection and Machine Learning Techniques. *Educational Process: International Journal*, 23, e2026067. <https://doi.org/10.22521/edupij.2026.23.67>

Received May 28, 2025

Accepted September 7, 2025

Keywords: Digital Addiction, Student Academic Performance, Sustainable Development Education, Machine Learning, Public Health

Author for correspondence:

Hasan Sarwar

✉ hsarwar@cse.uui.ac.bd

✉ United International University, Dhaka, Bangladesh

Analyzing Digital Addiction through Feature Selection and Machine Learning Techniques

Mizanur Rahman , Hasan Sarwar , Md Kamrul Hasan , Ting Tin Tin 

Abstract

Background/purpose. Digital addiction has become an increasing concern, particularly among students, as excessive engagement with digital platforms negatively impacts academic performance, social interactions, and mental health. This study primarily aims to evaluate and compare the effectiveness of various machine learning and deep learning models in predicting digital addiction. While we examine key features that contribute to classification performance, our focus is not on discovering new predictors but on selecting features that enhance model accuracy. However, few datasets are publicly available, and, moreover, these datasets do not employ similar feature types. As a result, it is not readily clear which set of features a researcher should consider to obtain the most accurate prediction.

Materials/methods. We utilized two publicly available datasets and conducted feature selection using ANOVA scores and Random Forest. We then performed prediction of digital addiction using four machine learning (ML) models—K-Nearest Neighbors (KNN), Support Vector Machine (SVM), Random Forest (RF), and Gradient Boosting (GB)—as well as one deep learning model, Artificial Neural Network (ANN). These models were evaluated using standard performance metrics, including accuracy, precision, recall, and F1 Score.

Results. Our results show that Random Forest, when used to identify the most important feature, performs better overall, especially for models such as KNN. The Random Forest and Gradient Boosting models achieved perfect scores (100% accuracy, precision, recall, and F1-score) on one dataset, and the ANN model reached up to 99.5% accuracy.

Conclusion. Random Forest-based feature importance is particularly effective, and the overall results suggest that ensemble models such as RF and GB are highly reliable. This research will guide future researchers, as they create their own dataset, on the importance of selecting appropriate features in combination with machine learning algorithms to achieve better predictive accuracy. These insights may support early interventions in educational settings to reduce the academic risks associated with digital addiction.



OPEN ACCESS

© The Author(s), 2025. This is an Open Access article, distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted re-use, distribution, and reproduction, provided the original article is properly cited.

1. Introduction

Digital addiction means addictive behavior toward digital technologies such as PCs, mobile phones, the internet, tablets, etc. Over the past two decades, these technologies have become deeply integrated into daily life, shaping individuals' routines and lifestyles from morning to night. (Karakose et al., 2023). Digital addiction primarily includes three types: smartphone addiction, internet addiction, and social media addiction. (Al-Mamun et al., 2024 ; Dhanusia et al., 2024). The use of digital technologies has rapidly increased in recent decades. Digital technologies, including social networking sites and digital media platforms such as Facebook, Instagram, TikTok, Messenger, X, etc., have become a crucial part of young students' daily activities. (Masrom et al., 2024) (Mehta & Handriana, 2024).

Like many other areas of society and human activity, education has been significantly impacted by technological developments, particularly social networking sites. (Chen et al., 2020; Karakose & Tülübaş, 2023). A study by Karakose & Tülübaş (2023) Mentioned that digital technologies have considerable potential benefits in the 21st-century education system. Several studies have identified the advantages of digital technologies and social networking as significant opportunities. Specifically, (Tower et al., 2014) Found that the digital media platform Facebook serves as a valuable tool for study support. Other studies have mentioned that digital media platforms play a substantial role in information gathering, communication, and collaboration in community-based learning environments. (Kitsantas et al., 2016; Mazman & Usluel, 2010; Mirabolghasemi et al., 2016). An experiment by Junco et al. (2011) It was conducted with two groups of students: one connected via the digital media platform Twitter for group study discussions, and the other did not. The results revealed that the group using Twitter for study discussions experienced higher semester grades and participated more in class discussions than the other group. These findings suggest that digital technologies can enhance learning experiences when used effectively.

Although digital technologies have significant potential in education for children, adolescents, and young students, excessive use is becoming an addiction. (Jahan et al., 2021; Khazaie et al., 2023). After the COVID-19 pandemic, the use of digital media increased drastically worldwide, particularly among women and young students. (Meng et al., 2022; Jahan et al. 2021; Al-Mamun et al. 2024). Several studies have revealed that overuse of digital technologies contributes to physical and mental health issues such as loneliness, depression, anxiety, and burnout. (Balhara et al., 2018; Balta et al., 2020; Donnelly & Kuss, 2016; J Kuss et al., 2014; Marciano et al., 2022; Rumpf et al., 2018; Tin Tin et al., 2024). Digital addiction also diminishes relationships with family and friends and reduces interactions within the social community. (Abbas et al., 2019; Aqeel et al., 2022; Dresch-Langley & Hutt, 2022). A study by Leventhal et al. (2021) Highlighted an increased risk of suicidal behavior due to digital addiction among young generations. American national statistics indicate that a significant percentage of young adults engage excessively with social media, online dating, video games, and other digital platforms. As a result, digital technology has been referred to as a "new drug." (Smith & Anderson, 2018).

One of the most concerning aspects of digital addiction is its impact on academic performance. Notable studies have reported poor academic outcomes due to digital addiction. (Alwagait et al., 2015; Giunchiglia et al., 2018). For instance, Alwagait et al. (2015b) Identified digital addiction as a major factor contributing to declining academic performance. Students struggled with time management when they began using social media platforms. Time passed quickly, yet they remained dissatisfied with how they used it.

Thus, digital addiction has become a significant concern within the research community. The extensive use of the internet has raised widespread concerns about problematic online behavior and associated issues (Chia et al., 2020). Numerous research studies have been conducted to identify the

causes of digital addiction and provide predictive insights, gaining considerable attention. These studies have employed various analytical approaches, including quantitative and cross-sectional analyses (Ferdous et al., 2023); (Mosharrafa et al., 2024); machine learning prediction models (Al-Alawi et al., 2023), deep learning models (Masrom et al., 2023), and hybrid machine learning techniques (Mukta et al., 2022). In machine learning approaches, several algorithms have been applied, including Decision Tree (DT), Random Forest (RF), K-Nearest Neighbor (KNN), Support Vector Machine (SVM), Naive Bayes (NB), Gradient Boosting (GB), and XGBoost (XGB). In the deep learning domain, Artificial Neural Networks (ANN) have been implemented in several research publications (Rodríguez-Hernández et al., 2021).

Many studies have explored digital addiction, but most have focused on how it harms students' academic performance, rather than how to identify or predict addiction itself. This shows a clear gap in research, especially in creating reliable tools to detect addiction early. One major challenge is the lack of open, shared datasets—most past research used private data, making it hard for others to reproduce or compare results. Even when data is available, the wide differences in people's digital habits make it difficult to train machine learning models that work well for everyone. In this study, we address these problems by using two publicly available datasets from GitHub, which improve the reliability and openness of our work. We also use careful data processing and feature selection methods, and test several machine learning and deep learning models to identify the best approach to predicting digital addiction. While many existing studies focus on the negative impacts of digital addiction, fewer have explored how to predict addiction levels using computational models. This lack of predictive focus represents a key research gap, especially given the potential for early detection and intervention. Our study addresses this gap by applying and comparing multiple machine learning and deep learning models to classify and predict the level of digital addiction using well-defined behavioral and demographic features. Therefore, this study aims to (1) identify the behavioral and demographic factors associated with digital addiction that have variable impacts on different predictors and (2) compare the performance of machine learning and deep learning algorithms in predicting addiction severity. Specifically, we investigated the following research questions: RQ1: What behavioral and demographic features are most predictive of digital addiction among students? and RQ2: Which machine learning and deep learning models provide the highest accuracy in classifying digital addiction levels?

The remainder of this paper is organized as follows: Section 2 provides a comprehensive review of related literature, highlighting key findings and research gaps in the study of digital addiction. Section 3 outlines the methodological framework, detailing the data sources, preprocessing steps, feature selection techniques, and the machine learning and deep learning models employed, along with model performance metrics across both datasets. Section 4 presents the experimental results of two datasets. In Section 5, we interpret the results and discuss their implications. Section 6 addresses the study's constraints and proposes directions for future research. Finally, Section 7 summarizes the key contributions and concludes the paper.

2. Literature Review

Several studies have explored the relationship between digital addiction and student academic performance using both quantitative analysis and machine learning prediction methods. Quantitative analysis typically involves surveys, interviews, and questionnaires to gather direct feedback from students, offering insights into their online behaviors and academic outcomes. On the other hand, machine learning models rely on large, structured datasets to predict academic performance based on digital usage patterns. Interestingly, the data collected through quantitative methods often serves as the foundation for machine learning models, helping to refine and enhance prediction accuracy. In our research, we prioritize both approaches to gain a well-rounded understanding of how digital

addiction correlates with student academic success, ensuring a balanced mix of statistical insights and predictive analytics.

Ferdous et al. (Ferdous et al., 2023) have explored the dual role of social media as a tool for new cultural morality and emerging risks, particularly among university students. The study employs a quantitative methodology, primarily using an online survey with 217 participants and 20 case studies selected through non-probability sampling. The research is focused on sociological and behavioral analysis. This paper helps to create a dataset with their survey responses. The dataset can consist of survey responses and case study data, providing insights into students' experiences with social media across various dimensions, such as relationships, cyberstalking, education, journalism, and entrepreneurship. The findings reveal that while social media facilitates communication, education, and entrepreneurship, it also amplifies negative experiences, including cyber harassment, consumer fraud, and misinformation. The study underscores the urgent need for regulations and awareness campaigns to mitigate these risks and enhance the constructive use of social media.

Acharya et al. (Acharya et al., 2023) have investigated the prevalence and associated factors of internet addiction among undergraduate students in Kathmandu, Nepal. The study follows a cross-sectional methodology, utilizing a self-administered questionnaire that includes the Internet Addiction Test (IAT) scale. A sample of 344 students from colleges affiliated with Pokhara University participated in the study. The dataset can consist of survey responses detailing students' internet usage habits, psychological well-being, and interpersonal relationships. The results reveal that nearly 30% of students exhibit signs of internet addiction, with factors such as parental control, perceived loneliness, and time spent online playing significant roles. The study highlights the urgent need for awareness programs to educate students, parents, and educators about the risks of excessive internet use and to advocate for balanced digital habits.

Mosharrafa et al. (Mosharrafa et al., 2024) have explored the impact of social media use and its effects on the academic performance of university students in Bangladesh, emphasizing the mediating role of mental health. The research has followed a quantitative cross-sectional methodology, utilizing survey-based data collection and structural equation modeling (SEM) to analyze relationships between variables. It relies on statistical techniques like confirmatory factor analysis (CFA) and mediation analysis. The dataset can consist of 380 survey responses from university students, assessing social media addiction (using the Bergen Social Media Addiction Scale), mental health (measured by the Patient Health Questionnaire PHQ-9), and self-reported academic performance. Results indicate that social media usage positively influences academic performance when mediated by improved mental health, contradicting common assumptions that excessive social media use hampers academic success. These findings suggest that a balanced approach to social media could enhance students' academic outcomes while supporting mental well-being.

Seema et al. (Seema & Varik-Maasik, 2023) have examined the relationship between digital addiction, screen time, and learning difficulties among teenagers in Estonia, particularly in the context of inclusive education. The research follows a quantitative approach, utilizing a large-scale e-survey with responses from 8,486 students aged 11 to 20. The study relies on linear regression models to assess how digital addiction and screen time influence perceived learning difficulties. The primary outcome for making a dataset of self-reported measures on digital addiction, learning challenges, and screen usage habits. The results reveal a positive correlation between screen time, digital addiction, and learning difficulties, with digital addiction alone explaining 10% of the variability in learning difficulties and 11% when combined with screen time. However, the research also highlights the limitations of survey-based methods in accurately capturing data from students with special needs, emphasizing the need for alternative, more inclusive data collection techniques in future studies.

Al-Alawi et al. (2023) have implemented supervised machine learning techniques to identify factors negatively affecting the academic performance of college students on probation in Oman. Following the Knowledge Discovery in Databases (KDD) methodology, the research analyzes a dataset of 6,514 students spanning 11 years (2009–2019) from a major public university. The study uses the Information Gain (InfoGain) algorithm for feature selection. It employs classification-based machine learning models, including J48 DT, RF, NB, SVM, KNN, and ANN, as well as ensemble methods such as Logit Boost, Vote, and Bagging. These models classify students into categories and predict whether they will be placed on academic probation based on features such as study duration, secondary school performance, gender, estimated graduation year, and academic specialization. The dataset consists of academic records and demographic information, and the results highlight that study duration and previous academic performance are the strongest predictors of probation. The study demonstrates that machine learning models can effectively predict academic struggles, offering valuable insights for educators and policymakers to improve student retention and academic success.

Ahamed et al. (2017) have focused on designing an intelligent system to predict academic performance during adolescence by analyzing socioeconomic, psychological, and academic factors. The research follows a machine-learning-based approach, using ANN, KNN, and SVM to classify students' higher secondary (HSC) grades into performance categories. The methodology involves extensive data preprocessing techniques, including data balancing (SMOTEBoost), dimensionality reduction (PCA, SOM, GHA), and normalization to optimize model performance. The dataset consists of survey responses from 423 students, covering factors such as family background, study habits, psychological well-being, and social influences. The results indicate that ANN achieves the highest accuracy of 86.11% in predicting student performance, outperforming SVM and K-NN. Key findings suggest that parental involvement, previous academic records, and socioeconomic status significantly influence student success. The study underscores the potential of machine learning in educational predictions and highlights the importance of early interventions to support students at risk of academic struggles.

Ahmed et al. (2023) has explored a minimalistic machine learning approach to predict and understand the relationship between app usage and students' academic performance. The research has collected real-time smartphone usage data from 124 Bangladeshi students over seven days using a custom-built data retrieval tool. The methodology involves unsupervised clustering (DBSCAN) to analyze usage behaviour and supervised machine learning models (KNN, SVR, GB, RF, and XGBoost) to predict students' CGPA, with an accuracy of ± 0.36 relative to the actual CGPA. The dataset consists of students' real smartphone usage data, including app categories such as Productivity, Books, Video, and Social Media. Results reveal that higher use of the Productivity and Books apps correlates positively with CGPA, whereas use of the Video app negatively impacts academic performance. Interestingly, social media usage does not show a statistically significant relation with CGPA. The study highlights the potential of instant app usage data for early academic intervention, offering insights for designing tools to help students manage their smartphone usage more effectively.

Xu et al. (2019) explored the relationship between internet usage behaviors and academic performance, using machine learning techniques to predict students' academic outcomes. The research has analyzed internet usage data from 4,000 undergraduate students, extracting features such as online duration, internet traffic volume, and connection frequency. Three supervised machine learning models—DT, NN, and SVM—were used to classify students by academic performance. The results indicate that the frequency of internet connection is positively correlated with academic success, whereas high internet traffic volume negatively impacts performance. Additionally, the study found that increasing the number of behavioral features improves prediction accuracy, highlighting the potential of internet usage data as a strong predictor of academic success. The study underscores the importance of behavioral discipline in achieving better academic

outcomes, suggesting that controlled internet usage can positively influence students' educational achievements.

Carlos et al. (2021) has systematically implemented ANN to predict academic performance in higher education, focusing on classification accuracy and key predictor evaluation. Using a dataset of 162,030 students from Colombian universities, the study applies ANNs to categorize students' academic performance as either high or low, achieving 82% accuracy for high performance and 71% for low performance. The methodology follows a structured approach that includes data collection, preprocessing, model implementation, and evaluation, comparing ANN performance with that of other machine learning models. The research includes a dataset consisting of prior academic achievement records, socioeconomic data, high school characteristics, and students' work status. Results indicate that prior academic achievement, socioeconomic conditions, and high school background are the strongest predictors of university performance. Moreover, ANNs outperform traditional statistical methods in recall and F1 scores, making them a robust tool for educational predictions. The study emphasizes the need for early warning systems to support struggling students and suggests that well-designed ANN models can enhance decision-making in academic institutions.

Masrom et al. (2023) have investigated the impact of excessive Instagram use on students' academic performance using a two-stage methodology that combines structural equation modeling (SEM) and ANN. The research follows the Uses and Gratifications Theory (UGT) framework and collects data from 285 university students to analyze factors such as social gratification, content gratification, entertainment, social presence, and escapism. The dataset comprises survey responses that measure students' Instagram usage habits and academic performance. ANN analysis validates SEM findings and identifies escapism as the most significant predictor of excessive Instagram use, followed by entertainment, social presence, and social gratification. The results show that excessive Instagram use is significantly associated with academic impairment, emphasizing the need for strategies to regulate screen time and promote healthier social media habits among students.

Mukta et al. (2022) have developed a hybrid machine learning model to predict students' academic performance based on their mental health and psychological attributes, inferred from social media interactions, particularly Facebook. The methodology involves a two-level prediction model, where the first level uses Bidirectional Long Short-Term Memory (BiLSTM) regression to predict mental health attributes from Facebook posts, and the second level employs classification models such as KNN, RF, and LightGBM (LgBoost) to predict students' academic performance. The dataset comprises 302 students' Facebook posts (27,889 posts in total) and their self-reported academic performance (GPA). Results show that self-efficacy and conscientiousness have the strongest correlations with academic success, while depression and neuroticism negatively impact GPA. The ensemble classification model achieves an AUC-ROC score of 0.95 and an F1-score of 0.94, outperforming baseline models such as Linguistic Inquiry and Word Count (LIWC) and Empath. These findings highlight how social media activity can serve as a predictive indicator of students' academic outcomes, emphasizing the potential of AI-driven early intervention systems to support struggling students.

Pelima et al. (2024) have conducted a systematic literature review (SLR) on predicting university student graduation using machine learning techniques and academic performance data. The research follows Kitchenham's systematic review methodology, analyzing 70 journal articles published between 2018 and 2023. It does not introduce a new machine learning model but compiles insights from past studies, identifying SVM and RF as the most commonly used algorithms for graduation prediction. The dataset sources primarily come from Learning Management Systems (LMS) and Student Information Systems (SIS), focusing on student retention, academic records, and behavioral data. The review finds that some models achieve up to 90% accuracy in predicting graduation outcomes. However, it also highlights key challenges, including data limitations, model

interpretability, and ethical concerns in applying machine learning to predict student success. The study emphasizes the potential of AI-driven early intervention systems and calls for more transparent, adaptable, and ethically responsible predictive models.

Rather than grounding this study in a formal behavioral theory, we adopt a practical, data-driven behavioral modeling perspective. This approach is based on existing research that consistently links observable behaviors—such as social media screen time, frequency of use, satisfaction, and self-control—with the risk of digital addiction. These indicators, widely used in previous empirical studies, provide a strong foundation for constructing predictive models. By aligning our feature selection with prior findings on user behavior and digital engagement, we ensure that the predictive framework remains grounded in real-world usage patterns.

3. Methodology

3.1. Machine Learning Techniques

3.1.1. KNN

The KNN algorithm is a supervised learning method commonly used for classification and regression tasks. It operates by identifying the k most similar instances in the training data using a chosen distance measure, then using their labels to determine the output for a new input. To apply KNN, a distance function is used to quantify how close the input point is to other data points in the feature space. While Euclidean distance is most widely used, alternatives like Manhattan, Minkowski, or cosine distance can also be applied depending on the nature of the dataset (Rahman et al., 2023). KNN is a non-parametric algorithm, meaning it doesn't assume any specific underlying distribution for the data. This makes it especially effective in cases where the decision boundaries are complex and non-linear. Several hyperparameters can be adjusted to improve KNN performance, including the number of neighbors (k), the choice of distance metric, and the weighting strategy (wherein closer neighbors may be given more influence than distant ones). The algorithm generally performs well on low-dimensional datasets where the relationships among data points can be captured efficiently through proximity-based decisions (Sarker, 2021).

3.1.2. SVM

SVM is a powerful supervised learning algorithm used for both classification and regression tasks. It operates by identifying an optimal hyperplane that separates data points of different classes with the maximum possible margin. SVMs are particularly effective in high-dimensional spaces and can handle non-linear relationships through kernel functions such as linear, polynomial, radial basis function (RBF), sigmoid, and precomputed kernels. Key hyperparameters that influence its performance include the kernel type, the regularization parameter (which controls the trade-off between achieving a low error on training data and minimizing model complexity), and the gamma parameter (which defines the influence of a single training example). SVM is especially useful when the number of features exceeds the number of samples and when a clear margin of separation is required, even in complex or non-linearly separable datasets (Roy & Chakraborty, 2023).

3.1.3. RF

RF is a robust ensemble learning algorithm that builds multiple decision trees on randomly selected subsets of data and features, combining their outputs to improve prediction accuracy. For classification, each tree casts a vote, and the majority vote becomes the final output. Key hyperparameters include the number of trees ($n_estimators$), the maximum depth of each tree (max_depth), and the number of features considered at each split. Random Forest is well-suited for datasets with complex, non-linear relationships and performs reliably even with noisy or partially

missing data. Its versatility has led to successful applications across various fields, such as finance, healthcare, marketing, and bioinformatics (Prakash et al., 2023; Rahman et al., 2024).

3.1.4. Gradient Boosting

Gradient Boosting is a powerful ensemble learning technique used for both regression and classification tasks, which builds a strong predictive model by sequentially combining multiple weak learners, typically decision trees. The algorithm works iteratively, with each new model attempting to correct the errors of previous models by focusing more on misclassified or high-error samples. The process starts by fitting an initial weak model, usually a shallow decision tree, and then computing the residual errors. In each iteration, a new tree is trained to predict these residuals, and the predictions are combined using a learning rate parameter to control each tree's contribution. Key parameters of Gradient Boosting include the number of estimators (trees), learning rate (which controls the step size during updates), max depth (which limits tree complexity), and subsample (which controls the fraction of data used in each boosting step to introduce randomness).

3.2. Deep Learning

3.2.1. ANN

ANN is a computational model inspired by the human brain, consisting of layers of interconnected neurons that process and learn patterns from data. The working procedure of an ANN involves three main layers: the input layer, the hidden layers, and the output layer. Each neuron in a layer applies a weighted sum of inputs, followed by an activation function (e.g., ReLU, Sigmoid, Tanh) to introduce non-linearity. The network learns through backpropagation, in which errors are propagated backward using gradient descent and optimization algorithms such as Adam or RMSprop to update the weights. Key parameters of an ANN include the number of layers (the network's depth), the number of neurons per layer, the learning rate (which controls weight updates), the batch size, and the number of epochs (the number of training iterations). Hyperparameters such as dropout rate (to prevent overfitting) and weight initialization methods (e.g., Xavier, He initialization) also impact performance.

3.3. Dataset Description

To build a solid foundation for our research on digital addiction and academic performance, we carefully selected two publicly available datasets from GitHub that are related to our concern. These datasets cover a wide range of real-world factors, including screen time, social media habits, and demographic details like age and gender. Unlike many previous studies that relied on private or undisclosed data, we wanted our approach to be open, reproducible, and accessible for future research. We also ensured the datasets were complete, well-structured, and required minimal preprocessing, reducing potential bias and ensuring high-quality inputs for our machine learning models. By using trusted, open-access sources, we're not just strengthening our research; we're also contributing to a more transparent and ethical conversation about digital addiction and its impact on students.

Our first dataset (pk673, 2025) is collected from the GitHub public repository. It has 175 students' data and aims to classify whether a student is addicted to social media based on 15 features. The provided dataset details weekly social media usage metrics. It includes fields for various social media platforms, such as WhatsApp, Instagram, Snapchat, Telegram, Facebook/Messenger, BeReal, TikTok, WeChat, Twitter, LinkedIn, and Messages. Each platform's usage is quantified in hours per week. Besides, the dataset captures the total social media screen time and the number of times social media apps are opened in hourly intervals over a day, aggregated over a week. A categorical field, "Social Media Addiction Level," is included to classify users based on their usage frequency. Users are labeled as "Addicted" if they open social media apps 105 times or more in a week, and "Not

Addicted" if the count is below 105. This dataset is likely used to analyze social media habits and their potential impact on users' behavior and well-being.

Our second dataset (Patil, 2025) is also collected from GitHub. It contains 19 attributes and 1,000 instances. This dataset is a comprehensive collection of user data related to social media usage and its impact. It includes various attributes such as demographic information (Age, Gender, Location), financial status (Income, Debt), professional details (Profession), and specific social media usage metrics (Platform, Total Time Spent, Video Length, Importance Score, etc.). The dataset also captures behavioral aspects such as Scroll Rate, Frequency of Use, and reasons for watching videos (Watch Reason). Additionally, it includes psychological and productivity-related metrics such as Productivity Loss, Satisfaction, Self Control, and Addiction Level. The data types used are integers (int), floating-point numbers (float), and categorical variables (category).

Both datasets used in this study were sourced from GitHub and are accompanied by descriptive project documentation and exploration analyses. The first dataset was developed to classify students' social media addiction based on usage metrics, including total screen time and frequency of app openings. The addiction labels were predefined by the dataset creators using clustering and regression models, and users were classified as "Addicted" if they opened social media apps more than 105 times per week. The second dataset contains psychological, behavioral, and demographic variables collected through self-reporting and was similarly designed to investigate predictors of digital addiction. Although the authors did not collect these datasets, their structure, clarity, and use in exploratory research make them suitable for evaluating predictive models in this study.

3.4. Dataset Analysis

The dataset 1 consists of 100 addicted and 75 non-addicted students, resulting in a slight class imbalance (100:75 ratio). This imbalance remains within acceptable limits for classification tasks and does not introduce significant bias. The data was collected from a diverse student population, ensuring fair representation. Figure 1 shows the label distribution ratio. To better understand the dataset, we analyzed attribute correlations and statistical values, which are crucial for feature selection and model training. Table 1 presents statistical metrics (mean, median, minimum, etc.) for each attribute. These statistics help identify attributes that might be indicative of addiction. For instance, the mean and median usage times for social media apps such as WhatsApp, Instagram, and Snapchat are relatively high compared to other apps. This suggests that students spend significant time on these platforms. Similarly, the maximum usage time for these apps is considerably high, potentially indicating overuse. On the other hand, Twitter and LinkedIn indicate lower mean and median usage times, suggesting that students engage with these platforms less frequently.

Also, the correlation matrix (Figure 2) provides insights into strongly correlated features. The correlation coefficient (r) ranges from -1 to +1, where +1 indicates a strong positive correlation (both attributes increase together), -1 indicates a strong negative correlation (one increases while the other decreases), and 0 indicates no relationship between attributes. Two features are considered strongly correlated when the absolute value of their correlation coefficient exceeds 0.7. From the correlation matrix, we identified key relationships that influence decision-making. For example, TikTok and WeChat are highly correlated, with a coefficient of 0.88, indicating similar usage patterns. Since highly correlated features can introduce redundancy, one feature may be selected over the other to improve model efficiency.

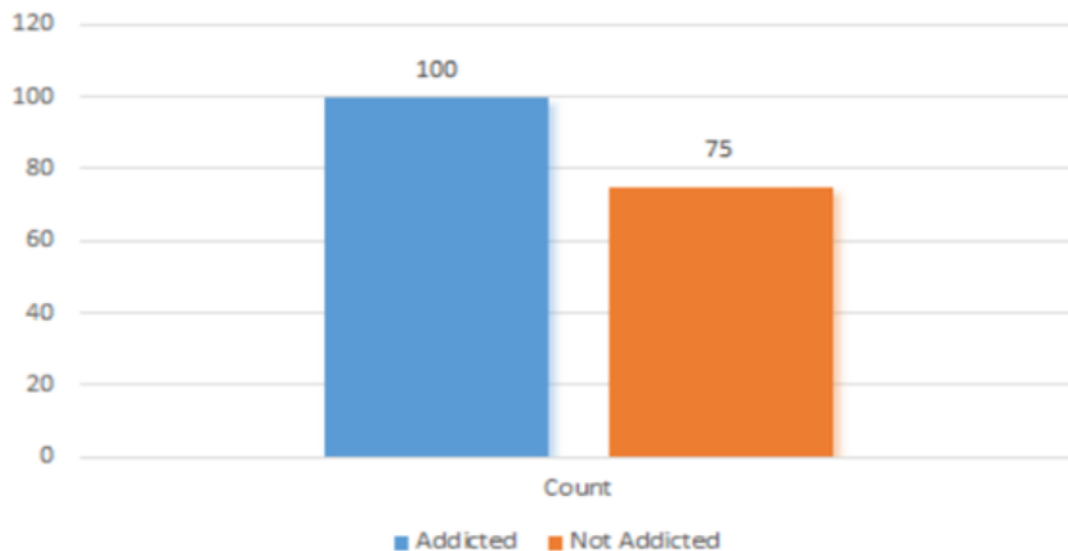


Figure 1. Social Media Addiction Distribution Level of Dataset 1

Table 1. Summary Statistics of Social Media Usage from Dataset 1

Platform	Mean	Std	Min	25%	50% (Median)	75%	Max
WhatsApp	7.88	4.67	0.00	5.06	7.50	10.10	22.50
Instagram	8.25	5.09	0.00	4.75	7.00	11.23	24.00
Snapchat	1.41	2.18	0.00	0.00	0.00	1.53	12.10
Telegram	0.12	0.29	0.00	0.00	0.00	0.06	2.39
Facebook	0.16	0.42	0.00	0.00	0.00	0.00	2.35
BeReal	0.12	0.78	0.00	0.00	0.00	0.00	8.60
TikTok	0.09	0.46	0.00	0.00	0.00	0.00	3.90
WeChat	0.35	1.63	0.00	0.00	0.00	0.00	10.50
Twitter	0.25	1.02	0.00	0.00	0.00	0.00	8.50
LinkedIn	3.26	4.78	0.00	0.42	1.42	4.00	22.80
Messages	0.59	1.66	0.00	0.00	0.06	0.40	10.30
Total Social Media Screen Time	22.47	10.56	0.58	15.68	21.62	28.08	55.00
Times Opened (Hourly)	111.44	37.99	30.00	94.00	110.00	128.00	257.00

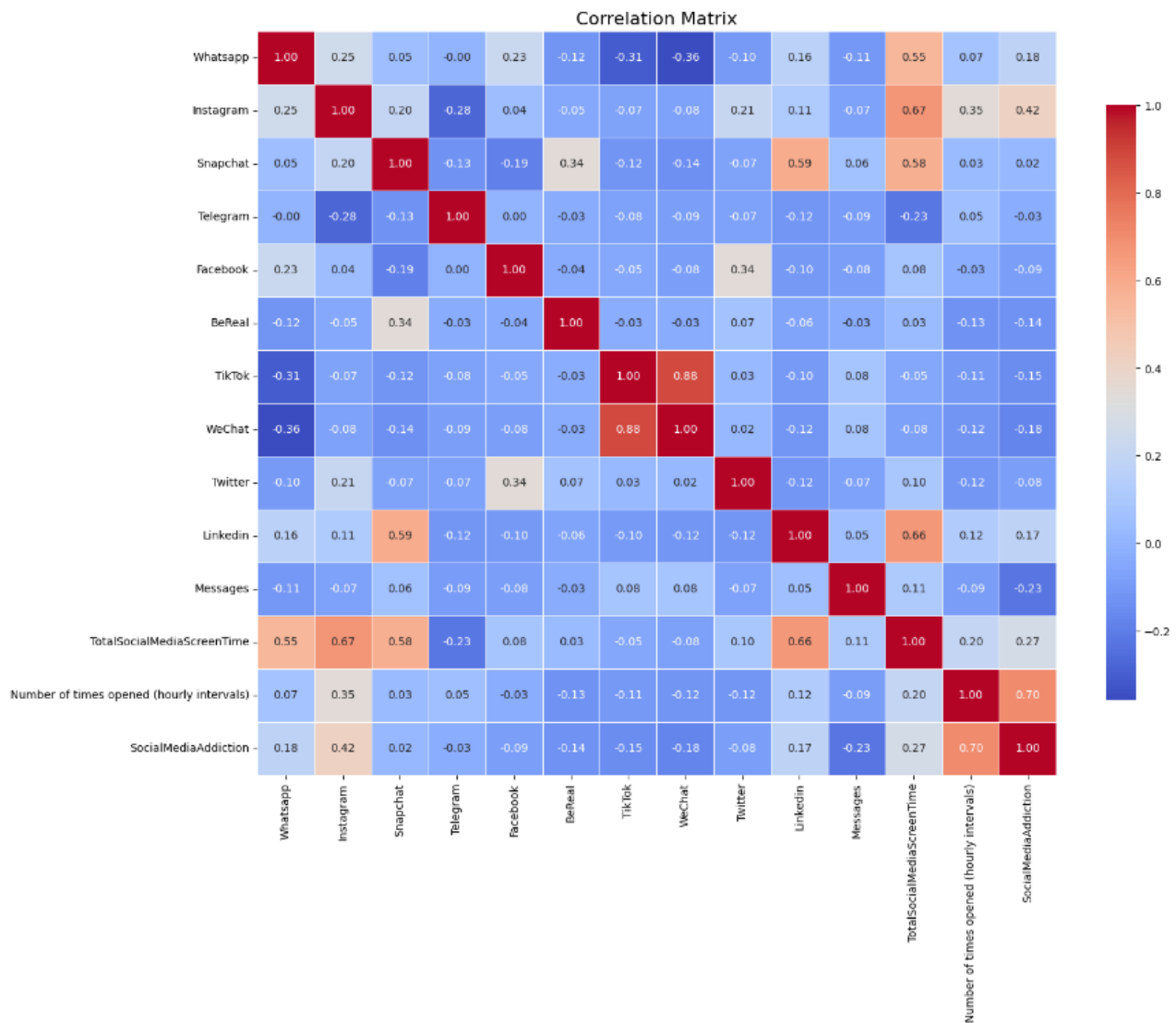
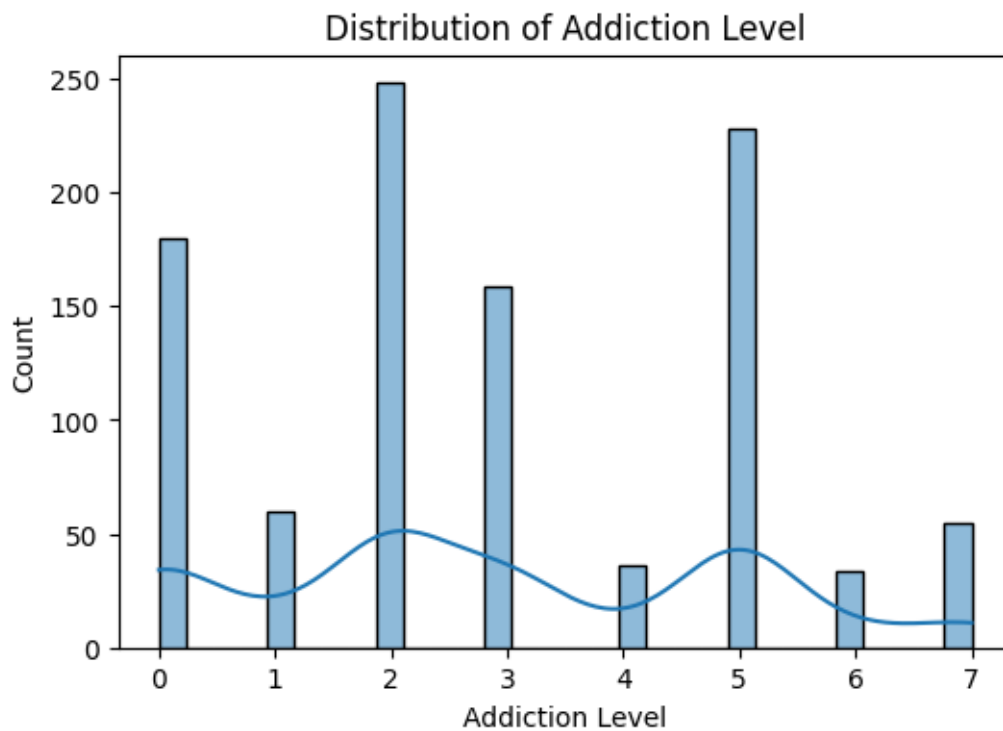


Figure 2. Correlation Matrix with Features of Dataset 1.

The dataset 2 categorizes digital addiction levels into eight tiers, ranging from 0 to 7. These levels serve as a measure of digital addiction among students. Figure 3 shows the distribution of digital addiction levels in the dataset. The x-axis represents addiction severity levels ranging from 0 (lowest addiction) to 7 (highest addiction), as defined by behavioral and psychological indicators in the dataset. The y-axis shows the number of participants assigned to each level. Figure 3 illustrates the distribution of addiction levels, showing that levels 2 and 5 have the highest occurrences compared to other levels. In Figure 4, we present the correlation matrix for the dataset attributes. The matrix reveals a strong correlation between productivity loss and self-control. Additionally, income and debt exhibit the second-highest correlation. Conversely, self-control and satisfaction show a strong negative correlation. In Table 2, we have described the regressive features such as income, debt, platform usage, and total time spent on videos to understand the mean, median, and other statistical values of these key variables. These features provide insight into user behavior, including their interaction with the platform, video preferences, and satisfaction levels. The dataset helps us understand how factors such as self-control, productivity loss, and frequency of use affect user engagement.

Table 2. Summary Statistics of Attributes from Dataset 2

Attributes	Mean	Std	Min	25%	50%	75%	Max
Income in?	59524.21	23736.21	20138	38675.25	58805	79792.25	99676
Debt	0.59	0.49	0	0	1	1	1
Platform	1.55	1.09	1	1	2	2.25	3
Total Time Spent (minutes)	151.41	83.95	10	78	152	223	298
Video Length	15.21	8.22	1	8	15	22	29
Importance Score	5.13	2.58	1	3	5	7	9
Time Spent on Video	14.97	8.20	1	8	15	22	29
Number of Videos Watched	25.24	14.02	1	14	25	37	49
Scroll Rate	49.77	29.19	1	23	50	74	99
Frequency	1.49	1.14	1	1	1	3	3
Productivity Loss	5.13	2.12	1	3	5	6	9
Satisfaction	4.86	2.12	1	4	5	7	9
Watch Reason	1.30	1.00	1	1	2	2	3
Self-Control	7.09	2.85	3	5	7	8	10

**Figure 3.** Level-wise Digital Addiction Distribution of Dataset 2.

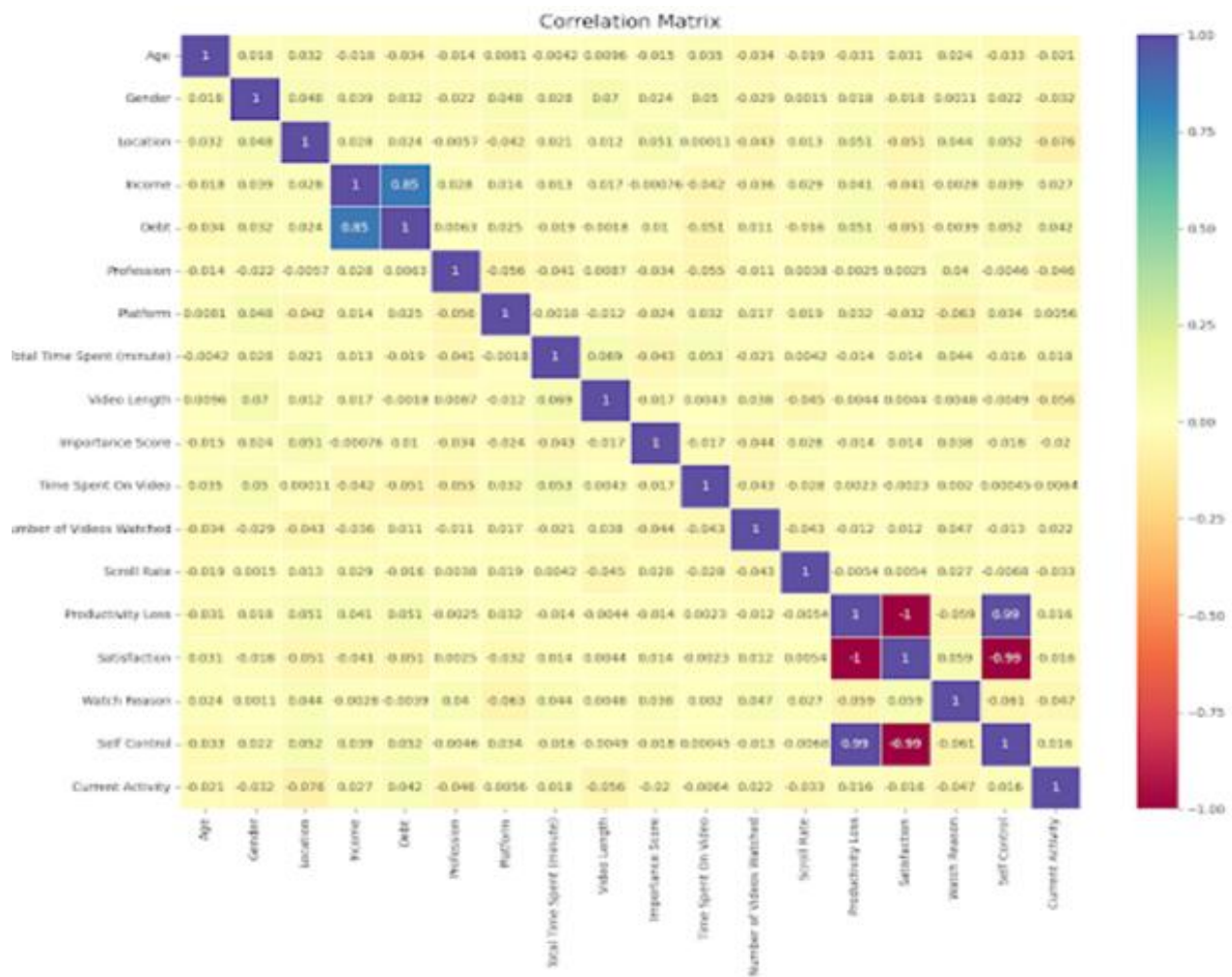


Figure 4. Correlation Matrix with Attributes of Dataset 2.

3.5. Dataset Preprocessing

Before training the models with the dataset, preprocessing steps were performed, including attribute encoding, normalization of numerical values, and handling of missing data. Both datasets were carefully examined for completeness. If missing values were detected, they were either removed or imputed using appropriate methods, such as zero or mean substitution, depending on the context.

3.6. Feature Selection

When training a model, not all features contribute equally to accurate predictions. Using too many features increases the risk of overfitting during training. To mitigate this, we apply feature selection to identify key features that enhance model performance while reducing complexity. To get important features, we have employed the ANOVA (Analysis of Variance) and the Random Forest (RF) feature importance technique.

- ANOVA assesses the statistical significance of each feature by comparing means across different groups. ANOVA provides a measure of statistical significance based on the variance explained by the feature. Higher ANOVA scores indicate stronger differentiation, suggesting that the feature has a substantial impact.
- Random Forest determines feature importance based on the reduction of impurity at each node in the decision trees. One common metric for this impurity reduction is the Gini impurity. The Gini impurity measures how often a randomly selected element in the dataset would be misclassified.

In a Random Forest, each feature is evaluated for its ability to reduce the Gini impurity, thus improving the accuracy of predictions. The feature importance score reflects how much each feature contributes to the reduction in Gini impurity across all decision trees in the forest. Features that reduce impurity more effectively and lead to better splits in the trees are assigned higher importance. This allows the model to identify which features are the most influential in making accurate predictions.

3.6.1. Dataset 1

By analyzing the correlation matrix, we eliminated redundant features, such as WeChat, because it is highly correlated (0.88) with TikTok. This step ensured that we avoided redundancy in our dataset. Following this, we employed both ANOVA and RF feature importance scores to perform feature selection. From the ANOVA score, the most significant feature is Number of times opened (hourly intervals), followed by Instagram usage and Total Social Media Screen Time. This indicates that the frequency of app usage plays a crucial role in influencing user behavior. Similarly, the Random Forest analysis identifies the number of times opened (hourly intervals), Instagram, and LinkedIn as the top contributing features, reinforcing the importance of activity frequency in predictive modeling. These are important values using both techniques, summarized in Table 3.

3.6.2. Dataset 2

The same process was followed for Dataset 2. Table 4 shows the ANOVA and RF importance scores for all attributes in dataset 2. To reduce overfitting and ensure proper model training, we have selected the top 11 features based on feature selection analysis. The procedure has applied here in the same way as for Dataset 1.

Table 3. ANOVA Feature Importance and Random Forest Importance Scores for Dataset 1

Features	ANOVA Score	Features	RF Importance Score
Number of times opened (hourly intervals)	167.117729	Number of times opened (hourly intervals)	0.667219
Instagram	37.083476	Instagram	0.074557
TotalSocialMediaScreenTime	13.651025	Linkedin	0.058705
Messages	9.348525	Messages	0.047742
Whatsapp	5.540970	TotalSocialMediaScreenTime	0.042548
Linkedin	5.294010	Snapchat	0.036233
BeReal	3.232298	Whatsapp	0.033135
Facebook	1.295063	Facebook	0.016717
Twitter	1.221309	Telegram	0.010999
Telegram	0.145319	Twitter	0.007596
Snapchat	0.068852	BeReal	0.004549

Table 4. ANOVA and Random Forest Feature Importance Scores for Dataset 2

Feature	ANOVA Score	Feature	Random Forest Importance
Self Control	∞	Self Control	0.3029
Satisfaction	14924.88	Productivity Loss	0.2979
Productivity Loss	14924.88	Satisfaction	0.2703
Frequency	185.1274	Frequency	0.0601
Location	2.0125	Income	0.0075
Number of Videos Watched	1.4959	Number of Videos Watched	0.0075
Watch Reason	1.4188	Scroll Rate	0.0073
Debt	1.1749	Total Time Spent (minutes)	0.0069
Income	1.1360	Video Length	0.0062
Platform	1.0876	Time Spent On Video	0.0062
Gender	0.9234	Age	0.0061
Video Length	0.8614	Location	0.0048
Age	0.7843	Profession	0.0041
Profession	0.7126	Importance Score	0.0036
Current Activity	0.6929	Platform	0.0023
Total Time Spent (minutes)	0.5996	Watch Reason	0.0023
Importance Score	0.5459	Current Activity	0.0018
Scroll Rate	0.5232	Gender	0.0014
Time Spent On Video	0.5032	Debt	0.0008

3.7. Dataset Splits

For training and evaluation, the datasets were split 80-20, with 80% allocated to training and 20% to testing. Before splitting, the datasets were randomly shuffled to mitigate any potential bias in the model training process.

3.8. Performance Evaluation Matrix

The following four metrics were considered when evaluating our model's performance. Accuracy was the first statistic assessed, as it shows how the model performs across all classes. The ratio of accurate forecasts to all predictions is used to calculate accuracy (Equation 1).

Our second parameter, precision, measures how accurately the model classified a sample as positive. This measure is defined as the ratio of correctly categorized positive samples to all samples correctly identified as positive. Equation 2 shows the mathematical formula.

We also examined the recall value, which evaluates the model's capability to identify positive samples. It is measured as the proportion of positively categorized positive samples to all positively classed positive samples. Equation 3 shows the mathematical formula.

Finally, the F1 score metric was examined. This metric combines precision and recall into a single number. The harmonic mean of recall and precision is calculated to determine this value. Equation 4 shows the mathematical formula F1.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

$$F1 - score = 2 * \frac{Precision * Recall}{Precision + Recall} \quad (4)$$

Table 5. Performance Comparison On Dataset 1

Algorithm	Anova				RF importance			
	Accuracy	Precision	Recall	F1	Accuracy	Precision	Recall	F1
KNN	0.8286	0.82	0.83	0.82	0.8	0.79	0.79	0.79
SVM	0.9714	0.98	0.96	0.97	0.9714	0.98	0.96	0.97
RF	0.9714	0.98	0.96	0.97	0.9714	0.98	0.96	0.97
GB	0.9714	0.98	0.96	0.97	0.9714	0.98	0.96	0.97
ANN	0.9429	0.9388	0.9388	0.938	0.9429	0.9388	0.9388	0.9388



Figure 5. Accuracy Comparison Graph for Dataset 1

4. Results

Our analysis treats both classes equally, and the model evaluation metrics confirm that the dataset does not favour one class over the other. Thus, the dataset is appropriate for studying social media addiction without introducing systematic bias. The results of the model training are analyzed below.

4.1. Dataset 1

After training five models, we have observed significant output. Table 5 presents a comprehensive performance comparison of five machine learning algorithms—KNN, SVM, RF, GB, and ANN—on Dataset 1, using two different feature selection techniques: ANOVA and Random Forest (RF) importance. The evaluation metrics include Accuracy, Precision, Recall, and F1-score. Among all models, SVM, RF, and GB consistently achieved the highest performance, with identical results across both feature selection methods—an accuracy of 97.14%, precision of 0.98, recall of 0.96, and an F1-score of 0.97. The ANN model also demonstrated strong performance, achieving slightly lower but consistent metrics (accuracy: 94.29%, F1-score: 0.9388) for both selection methods. KNN showed comparatively lower performance, especially under the RF importance method, where accuracy dropped to 80%. These results highlight the robustness of SVM, RF, and GB across different feature selection techniques and suggest their suitability for this dataset.

In Figure 5, the classification performance of five machine learning models—KNN, SVM, RF, GB, and ANN—on two different feature selection techniques: Anova and RF features importance applied to Dataset1. Across both methods, SVM, RF, and GB models achieved the highest accuracy of 0.97, indicating strong and consistent performance. ANN also performed well, achieving an accuracy of 0.94 in both cases. KNN had slightly lower performance, with an accuracy of 0.83 for Anova and 0.80 for RF, suggesting it may be less effective for this dataset.

Table 6. Performance Comparison on Dataset 2

Algorithm	Anova				RF importance			
	Accuracy	Precision	Recall	F1	Accuracy	Precision	Recall	F1
KNN	0.695	0.62	0.69	0.64	0.74	0.66	0.74	0.68
SVM	0.975	0.97	0.93	0.95	0.965	0.97	0.96	0.96
RF	1	1	1	1	1	1	1	1
GB	1	1	1	1	1	1	1	1
ANN	0.995	1	0.99	0.995	0.995	1	0.99	0.99



Figure 6. Accuracy Comparison Graph for Dataset 2

4.2. Dataset 2

Table 6 presents the performance of five classification algorithms on Dataset 2 using two feature selection methods: ANOVA and Random Forest importance. Random Forest (RF) and Gradient Boosting (GB) achieved perfect performance across all metrics (Accuracy, Precision, Recall, F1-score = 1.0). The Artificial Neural Network (ANN) also performed exceptionally well, with an accuracy of 0.995, a precision of 1.0, a recall of 0.99, and an F1-score of 0.995 under ANOVA and similar metrics under RF importance. SVM achieved strong results, with F1-scores of 0.95 (ANOVA) and 0.96 (RF importance). In contrast, KNN showed lower performance, especially with ANOVA (accuracy: 0.695, F1-score: 0.64), though it improved with RF-based selection (accuracy: 0.74, F1-score: 0.68). These results highlight the superior effectiveness of ensemble methods and deep learning models on this dataset.

Figure 6 shows a comparison of the performance of five classifiers—KNN, SVM, RF, GB, and ANN—on Dataset2 using two feature selection techniques: Anova and RF feature importance. Random Forest (RF) and Gradient Boosting (GB) achieved perfect accuracy (1.0) across both feature selection methods, indicating exceptional performance and robustness. The Artificial Neural Network (ANN) also performed nearly perfectly with an accuracy of 0.99 in both cases. SVM showed slightly higher accuracy with RF-selected features (0.96) than with Anova (0.94). Notably, KNN had the lowest accuracy, but its performance improved from 0.69 (Anova) to 0.74 (RF), suggesting that RF-based feature selection had a positive impact on this model. Overall, RF and GB were the top-performing models, while feature selection had the greatest impact on the KNN classifier.

5. Discussions

This study investigated the effectiveness of various machine learning and deep learning models for predicting digital addiction, and analyzed two feature selection methods: ANOVA and Random Forest (RF) feature importance. The experimental results across two publicly available datasets highlighted key patterns in model performance and in the impact of feature selection. Ensemble-based models such as Random Forest and Gradient Boosting consistently outperformed other algorithms, achieving perfect classification scores (100% accuracy, precision, recall, and F1-score) across both datasets. This confirms the strength of ensemble techniques in handling complex, nonlinear patterns and mitigating the influence of irrelevant features. The Artificial Neural Network (ANN) model also showed excellent performance, achieving up to 99.5% accuracy, indicating the potential of deep learning for predicting digital behavior when paired with appropriate feature selection. Importantly, the choice of feature selection method significantly impacted model performance. While both ANOVA and RF importance were effective, RF importance yielded better or equal results in most cases, particularly for simpler models such as KNN. These findings emphasize the need to carefully select input features to improve overall model accuracy and reliability.

Table 7. Performance Comparison of Both Datasets Among ML Algorithms.

	KNN	SVM	RF	GB	ANN
Dataset1 ANOVA	0.8286	0.9714	0.9714	0.9714	0.9429
Dataset1 RF Feature importance	0.8	0.9714	0.9714	0.9714	0.9429
Dataset2 ANOVA	0.695	0.975	1.00	1.00	0.995
Dataset2 RF Feature importance	0.74	0.965	1.00	1.00	0.995

Our findings are consistent with prior studies that suggest time spent on social media and frequency of usage are significant predictors of digital addiction among students. Like previous studies that used logistic regression and clustering models, our results affirm that behavioral indicators—such as the number of app opens or total screen time—can reliably predict addiction severity. However, unlike most prior studies that used private or institution-specific datasets, our use of open-source datasets makes the results more generalizable. The application of AI methods not only enhances prediction accuracy but also helps build intelligent systems that support early intervention strategies in academic and digital wellness programs. In summary, the study effectively addresses its research questions. For RQ1, we found that behavioral features such as total screen time, number of app openings, frequency of use, and low self-control are strong predictors of digital addiction among students. These variables consistently rank high in importance across both feature selection methods and models. Tables 3 and 4 show the importance of the features of both ANOVA and RF. For RQ2, Table 7 presents the performance comparison of various machine learning and deep learning algorithms (KNN, SVM, RF, GB, ANN) across two datasets using two feature selection methods: ANOVA and Random Forest importance. The results indicate that ensemble models (RF and GB) consistently achieved perfect classification (1.00) on Dataset 2, regardless of the feature selection technique. The ANN model also performed strongly, reaching up to 0.995 accuracy on dataset 2. However, KNN shows better accuracy, 0.82, on dataset 1 according to ANOVA.

6. Limitations & Future Work

Although the study provides meaningful insights, several limitations should be noted. First, the datasets used were limited in size and scope, potentially reducing the generalizability of the findings to broader populations. Moreover, there is currently a lack of large-scale, standardized datasets specifically focused on digital addiction. Existing datasets often vary in terms of feature sets, labeling criteria, and target populations, making it difficult to benchmark models or compare results across studies. Additionally, while the models performed exceptionally well, especially on Dataset 2, there is a risk of overfitting when perfect classification scores are achieved. The datasets may also lack diversity in behavioral or demographic attributes that could impact digital addiction outcomes. The use of publicly available datasets, while offering transparency, may limit full control over how variables (such as addiction thresholds) were defined; however, we preserved the original structure and labels to maintain reproducibility.

Future research should focus on validating these findings using larger, more diverse, and real-world datasets that reflect a broader range of digital behaviors and user profiles. There is a strong need for the creation and sharing of high-quality, standardized datasets tailored to digital addiction, enabling more robust benchmarking and reproducibility. Incorporating longitudinal or real-time data could help model the progression of digital addiction over time. Additionally, combining hybrid feature selection techniques with behavioral insights from psychology may improve both predictive performance and interpretability. Developing real-time monitoring or early intervention systems, possibly integrated into mobile platforms, also presents an impactful direction for future work.

7. Conclusions

This study sets out to achieve two primary goals: (1) to identify impactful behavioral and demographic predictors of digital addiction, and (2) to evaluate and compare the performance of machine learning and deep learning models for predicting addiction severity among students. Using two publicly available student-focused datasets, we applied ANOVA and Random Forest-based feature selection techniques alongside multiple ML/DL classifiers. The strongest predictors of digital addiction included frequency of app usage, self-control, satisfaction, and time spent on social media. Among the models tested, classifiers such as Random Forest and Gradient Boosting achieved perfect scores across all metrics, while the ANN model reached 99.5% accuracy. These results validate the

importance of robust feature engineering and model selection in building reliable digital behavior prediction systems. Beyond technical contributions, this work holds educational relevance by supporting early detection of addiction patterns that may impact students' academic performance and cognitive focus. By identifying at-risk users through behavioral patterns, institutions can explore proactive interventions to support learning outcomes and digital well-being. Future research can expand on these findings to design real-time monitoring systems and targeted interventions in academic environments.

Declarations

Conflicts of Interest. The authors declare no conflict of interest.

Funding. The work reported in this paper was supported by the Institute for Advanced Research (IAR), United International University (UIU), Bangladesh, under Grant No. UIUIAR-02-2023-SE-10. The authors would also like to thank INTI International University, Malaysia, for collaborating on this publication.

Ethical Approval. No ethical approval is required.

Data Availability Statement. The data used in the paper are publicly available, and the code can be provided by the corresponding author upon request.

References

- Abbas, J., Aqeel, M., Abbas, J., Shaher, B., Jaffar, A., Sundas, J., & Zhang, W. (2019). The moderating role of social support for marital adjustment, depression, anxiety, and stress: Evidence from Pakistani working and nonworking women. *Journal of Affective Disorders, 244*, 231–238. <https://doi.org/10.1016/j.jad.2018.07.071>
- Acharya, S., Adhikari, L., Khadka, S., Paudel, S., & Kaphle, M. (2023). Internet Addiction and Its Associated Factors among Undergraduate Students in Kathmandu, Nepal. *Journal of Addiction, 2023*. <https://doi.org/10.1155/2023/8782527>
- Ahamed, A. T. M. S., Mahmood, N. T., & Rahman, R. M. (2017). An intelligent system to predict academic performance based on different factors during adolescence. *Journal of Information and Telecommunication, 1*(2), 155–175. <https://doi.org/10.1080/24751839.2017.1323488>
- Ahmed, M. S., Rony, R. J., Hadi, M. A., Hossain, E., & Ahmed, N. (2023). A Minimalistic Approach to Predict and Understand the Relation of App Usage with Students' Academic Performance. *Proceedings of the ACM on Human-Computer Interaction, 7*(MHCI), 1–28. <https://doi.org/10.1145/3604240>
- Al-Alawi, L., Al Shaqsi, J., Tarhini, A., & Al-Busaidi, A. S. (2023). Using machine learning to predict factors affecting academic performance: the case of college students on academic probation. *Education and Information Technologies, 28*(10), 12407–12432. <https://doi.org/10.1007/s10639-023-11700-0>
- Al-Mamun, F., Hasan, M. E., Mostofa, N. B., Akther, M., Mashruba, T., Arif, M., Chaahat, A. H., Salam, A. B., Akter, M., Abedin, M. A. A., Bulbul, M. I. A., Adnan, M. S., Islam, M. S., Ahmed, M. S., Shahin, M. S. M., Islam, S., Hussain, M. M., Al Habib, A., ALmerab, M. M., ... Mamun, M. A. (2024). Prevalence and factors associated with digital addiction among students taking university entrance tests: a GIS-based study. *BMC Psychiatry, 24*(1), 322. <https://doi.org/10.1186/s12888-024-05737-9>
- Alwagait, E., Shahzad, B., & Alim, S. (2015a). Impact of social media usage on students' academic performance in Saudi Arabia. *Computers in Human Behavior, 51*, 1092–1097. <https://doi.org/https://doi.org/10.1016/j.chb.2014.09.028>

- Alwagait, E., Shahzad, B., & Alim, S. (2015b). Impact of social media usage on students' academic performance in Saudi Arabia. *Computers in Human Behavior, 51*, 1092–1097. <https://doi.org/https://doi.org/10.1016/j.chb.2014.09.028>
- Aqeel, M., Rehna, T., Shuja, K. H., & Abbas, J. (2022). Comparison of students' mental wellbeing, anxiety, depression, and quality of life during COVID-19's full and partial (smart) lockdowns: a follow-up study at a 5-month interval. *Frontiers in Psychiatry, 13*, 835585. <https://doi.org/10.3389/fpsy.2022.835585>
- Balhara, Y. P. S., Mahapatra, A., Sharma, P., & Bhargava, R. (2018). Problematic internet use among students in South-East Asia: Current state of evidence. *Indian Journal of Public Health, 62*(3), 197–210. https://doi.org/10.4103/ijph.ijph_288_17
- Balta, S., Emirtekin, E., Kircaburun, K., & Griffiths, M. D. (2020). Neuroticism, trait fear of missing out, and phubbing: The mediating role of state fear of missing out and problematic Instagram use. *International Journal of Mental Health and Addiction, 18*, 628–639. <https://doi.org/10.1007/s11469-018-9959-8>
- Chen, X., Zou, D., & Xie, H. (2020). Fifty years of British Journal of Educational Technology: A topic modeling based bibliometric perspective. *British Journal of Educational Technology, 51*(3), 692–708. <https://doi.org/10.1111/bjet.12907>
- Chia, D. X. Y., Ng, C. W. L., Kandasami, G., Seow, M. Y. L., Choo, C. C., Chew, P. K. H., Lee, C., & Zhang, M. W. B. (2020). Prevalence of internet addiction and gaming disorders in Southeast Asia: A meta-analysis. *International Journal of Environmental Research and Public Health, 17*(7), 2582. <https://doi.org/10.3390/ijerph17072582>
- Dhanusia, S., Santhana Lakshmi, S., Kumar, A., Prabhu, R., Srinivasan, V., Suganthirababu, P., Kumar, P., Kumaresan, A., Vishnuram, S., Alagesan, J., & Vasanthi, R. K. (2024). Impact of mobile phone usage on dynamic postural control among South Indian college students. *Work, 78*(2), 441–446. https://doi.org/10.3233/WOR-230161/ASSET/046CC4A7-79A5-43C9-AD51-4328DFB6AFAB/ASSETS/GRAPHIC/10.3233_WOR-230161-FIG3.JPG
- Donnelly, E., & Kuss, D. J. (2016). Depression among users of social networking sites (SNSs): The role of SNS addiction and increased usage. *Journal of Addiction and Preventive Medicine, 1*(2), 107. <https://doi.org/10.19104/japm.2016.107>
- Dresp-Langley, B., & Hutt, A. (2022). Digital addiction and sleep. *International Journal of Environmental Research and Public Health, 19*(11), 6910. <https://doi.org/10.3390/ijerph19116910>
- Ferdous, A., & Huda, Z. (2023). Social media, new cultures, and new threats: impact on university students in Bangladesh. *Human Behavior and Emerging Technologies, 2023*. <https://doi.org/10.1155/2023/2205861>
- Giunchiglia, F., Zeni, M., Gobbi, E., Bignotti, E., & Bison, I. (2018). Mobile social media usage and academic performance. *Computers in Human Behavior, 82*, 177–185. <https://doi.org/10.1016/j.chb.2017.12.041>
- Kuss, D., Griffiths, M., Karila, L., & Billieux, J. (2014). Internet addiction: A systematic review of epidemiological research for the last decade. *Current Pharmaceutical Design, 20*(25), 4026–4052. <https://doi.org/10.2174/13816128113199990617>

- Jahan, I., Hosen, I., Al Mamun, F., Kaggwa, M. M., Griffiths, M. D., & Mamun, M. A. (2021). How has the COVID-19 pandemic impacted internet use behaviors and facilitated problematic internet use? A Bangladeshi study. *Psychology Research and Behavior Management*, 1127–1138. <https://doi.org/10.2147/prbm.s323570>
- Junco, R., Heiberger, G., & Loken, E. (2011). The effect of Twitter on college student engagement and grades. *Journal of Computer Assisted Learning*, 27(2), 119–132. <https://doi.org/10.1111/j.1365-2729.2010.00387.x>
- Karakose, T., Demirkol, M., Yirci, R., Polat, H., Ozdemir, T. Y., & Tülübaşı, T. (2023). A conversation with ChatGPT about digital leadership and technology integration: Comparative analysis based on human–AI collaboration. *Administrative Sciences*, 13(7), 157. <https://doi.org/10.3390/admsci13070157>
- Karakose, T., & Tülübaşı, T. (2023). Digital leadership and sustainable school improvement—A conceptual analysis and implications for future Research. *Educ. Process Int. J*, 12(1), 10–22521. <https://doi.org/10.22521/edupij.2023.121.1>
- Khazaie, H., Lebni, J. Y., Abbas, J., Mahaki, B., Chaboksavar, F., Kianipour, N., Toghroli, R., & Ziapour, A. (2023). Internet addiction status and related factors among medical students: a cross-sectional study in Western Iran. *Community Health Equity Research & Policy*, 43(4), 347–356. <https://doi.org/10.1177/0272684x211025438>
- Kitsantas, A., Dabbagh, N., Chirinos, D. S., & Fake, H. (2016). College students' perceptions of positive and negative effects of social networking. *Social Networking and Education: Global Perspectives*, 225–238. https://doi.org/10.1007/978-3-319-17716-8_14
- Leventhal, A. M., Cho, J., Keyes, K. M., Zink, J., Riehm, K. E., Zhang, Y., & Ketema, E. (2021). Digital media use and suicidal behavior in US adolescents, 2009–2017. *Preventive Medicine Reports*, 23, 101497. <https://doi.org/10.1016/j.pmedr.2021.101497>
- Marciano, L., Ostroumova, M., Schulz, P. J., & Camerini, A.-L. (2022). Digital media use and adolescents' mental health during the COVID-19 pandemic: a systematic review and meta-analysis. *Frontiers in Public Health*, 9, 793868. <https://doi.org/10.3389/fpubh.2021.793868>
- Masrom, M., Busalim, A., Griffiths, M. D., Asadi, S., & Mohd Ali, R. (2023). The impact of excessive Instagram use on students' academic study: a two-stage SEM and artificial neural network approach. *Interactive Learning Environments*, 1–20. <https://doi.org/10.1080/10494820.2023.2184393>
- Masrom, M., Busalim, A., Griffiths, M. D., Asadi, S., & Mohd Ali, R. (2024). The impact of excessive Instagram use on students' academic study: a two-stage SEM and artificial neural network approach. *Interactive Learning Environments*, 32(7), 3546–3565. <https://doi.org/10.1080/10494820.2023.2184393>
- Mazman, S. G., & Usluel, Y. K. (2010). Modeling educational usage of Facebook. *Computers & Education*, 55(2), 444–453. <https://doi.org/10.1016/j.compedu.2010.02.008>
- Mehta, A. M., & Handriana, T. (2024). Analyzing CSR and customer engagement through green banking digitalization: with the mediating effect of perceived environmental value and moderation effect of customer's eco-consciousness. *Cogent Business & Management*, 11(1). <https://doi.org/10.1080/23311975.2024.2332502>
- Meng, S.-Q., Cheng, J.-L., Li, Y.-Y., Yang, X.-Q., Zheng, J.-W., Chang, X.-W., Shi, Y., Chen, Y., Lu, L., Sun, Y., Bao, Y.-P., & Shi, J. (2022). Global prevalence of digital addiction in general population: A systematic review and meta-analysis. *Clinical Psychology Review*, 92, 102128.

- Mirabolghasemi, M., Iahad, N. A., & Rahim, N. Z. A. (2016). Students' perception towards the potential and barriers of social network sites in higher education. *Social Networking and Education: Global Perspectives*, 41–49. https://doi.org/10.1007/978-3-319-17716-8_3
- Mosharrafa, R. A., Akther, T., & Siddique, F. K. (2024). Impact of social media usage on academic performance of university students: Mediating role of mental health under a cross-sectional study in Bangladesh. *Health Science Reports*, 7(1), e1788. <https://doi.org/10.1002/hsr2.1788>
- Mukta, M. S. H., Islam, S., Shatabda, S., Ali, M. E., & Zaman, A. (2022). Predicting academic performance: Analysis of students' mental health condition from social media interactions. *Behavioral Sciences*, 12(4), 87. <https://doi.org/10.3390/bs12040087>
- Patil, M. (2025). *Digital Addiction Dataset*. <https://github.com/data-manavpatil/DigitalAddiction/blob/main/dataset.csv>
- Pelima, L. R., Sukmana, Y., & Rosmansyah, Y. (2024). Predicting University Student Graduation Using Academic Performance and Machine Learning: A Systematic Literature Review. *IEEE Access*. <https://doi.org/10.1109/access.2024.3361479>
- pk673. (2025). *Social Media Addiction Analysis - Class Survey Dataset*. <https://github.com/pk673/Social-Media-Addiction-Analysis/blob/main/ClassSurvey.csv>
- Prakash, P., Bhanu, S., & Bigul, S. D. (2023). Random forest and logistic regression algorithms: A comparison of their performance. *AIP Conference Proceedings*, 2548(1). <https://doi.org/10.1063/5.0118420>
- Rahman, M., Roy, P. P., Ali, M., Gonçalves, T., & Sarwar, H. (2023). Software Effort Estimation using Machine Learning Technique. *International Journal of Advanced Computer Science and Applications*, 14(4), 822–827. <https://doi.org/10.14569/IJACSA.2023.0140491>
- Rahman, M., Sarwar, H., Kader, A., Goncalves, T., & Tin, T. T. (2024). Review and Empirical Analysis of Machine Learning-Based Software Effort Estimation. *IEEE Access*, 12, 85661–85680. <https://doi.org/10.1109/ACCESS.2024.3404879>
- Rodríguez-Hernández, C. F., Musso, M., Kyndt, E., & Cascallar, E. (2021). Artificial neural networks in academic performance prediction: Systematic implementation and predictor evaluation. *Computers and Education: Artificial Intelligence*, 2, 100018. <https://doi.org/10.1016/j.caeai.2021.100018>
- Roy, A., & Chakraborty, S. (2023). Support vector machine in structural reliability analysis: A review. *Reliability Engineering & System Safety*, 109126. <https://doi.org/10.1016/j.res.2023.109126>
- Rumpf, H.-J., Achab, S., Billieux, J., Bowden-Jones, H., Carragher, N., Demetrovics, Z., Higuchi, S., King, D. L., Mann, K., Potenza, M., Saunders, J. B., Abbott, M., Ambekar, A., Aricak, O. T., Assanangkornchai, S., Bahar, N., Borges, G., Brand, M., Chan, E. M.-L., ... Poznyak, V. (2018). Including gaming disorder in the ICD-11: The need to do so from a clinical and public health perspective: Commentary on: A weak scientific basis for gaming disorder: Let us err on the side of caution (van Rooij et al., 2018). *Journal of Behavioral Addictions*, 7(3), 556–561. <https://doi.org/10.1556/2006.7.2018.59>
- Ryan, P. (2018). Technology: The new addiction. *US Naval Institute Publications. Proceedings*, 144, 387.
- Sarker, I. H. (2021). Machine learning: Algorithms, real-world applications and research directions. *SN Computer Science*, 2(3), 160. <https://doi.org/10.1007/s42979-021-00592-x>

- Seema, R., & Varik-Maasik, E. (2023). Students' digital addiction and learning difficulties: shortcomings of surveys in inclusion. *Frontiers in Education*, 8, 1191817. <https://doi.org/10.3389/feduc.2023.1191817>
- Smith, A., & Anderson, M. (2018). *Social Media Use in 2018*. <https://www.pewresearch.org/internet/2018/03/01/social-media-use-in-2018/>
- Tin, T. T., Tiung, L. K., Hong, O. J., Hui, C. J., Shen, C. J., Jian, T. Y., & Ikumapayi, O. M. (2024). Does Different Social Media Platforms Lead to Depression, Anxiety, Stress and Bipolar Disorder? A Cross-Sectional Analysis on Private University. *Pakistan Journal of Life and Social Sciences (PJLSS)*, 22(2). <https://doi.org/10.57239/PJLSS-2024-22.2.00152>
- Tower, M., Latimer, S., & Hewitt, J. (2014). Social networking as a learning tool: Nursing students' perception of efficacy. *Nurse Education Today*, 34(6), 1012–1017. <https://doi.org/10.1016/j.nedt.2013.11.006>
- Xu, X., Wang, J., Peng, H., & Wu, R. (2019). Prediction of academic performance associated with internet usage behaviors using machine learning algorithms. *Computers in Human Behavior*, 98, 166–173. <https://doi.org/10.1016/j.chb.2019.04.015>

About the Contributor(s)

Mizanur Rahman, School of Computer Science, Western Illinois University, Macomb, Illinois, USA
Email: mizancse7462@gmail.com

Hasan Sarwar, Department of Computer Science and Engineering, United International University, Dhaka, Bangladesh
Email: hsarwar@cse.uiu.ac.bd

Md Kamrul Hasan, Department of English, United International University, Dhaka, Bangladesh
Email: kamrul@eli.uiu.ac.bd

Ting Tin Tin, Faculty of Data Science and Information Technology, Inti International University, Malaysia
Email: tintin.ting@newinti.edu.my

Publisher's Note: *The opinions, statements, and data presented in all publications are solely those of the individual author(s) and contributors and do not reflect the views of Universitepark, EDUPIJ, and/or the editor(s). Universitepark, the Journal, and/or the editor(s) accept no responsibility for any harm or damage to persons or property arising from the use of ideas, methods, instructions, or products mentioned in the content.*
