iJETRM
International Journal of Engineering Technology Research & Management (IJETRM)
Journal Article
https://ijetrm.com/issue/

SOCIAL MEDIA ADDICTION PREDICTION USING MACHINE LEARNING TECHNIQUES

Alade, Samuel Mayowa
Adejumo, Samuel Olujimi
Akaukwu, Ifeanyi Godspower
Sonaike, Opeyemi Oluwafemi
Ruphin, Kusinza Byamungu
[1,2]Department of Cyber security, Nnamdi Azikiwe University, Awka, Nigeria
[2]Department of Computer Science, Nnamdi Azikiwe University, Awka, Nigeria
[3]Department of Software Engineering, Nnamdi Azikiwe University, Awka, Nigeria
[4]Department of Computer Science, Abiola Ajimobi Technical University, Ibadan, Nigeria
[5]Department of Computer Science, Olivia University, Bujumbura, Burundi
sm.alade@unizik.edu.ng, So.adejumo@unizik.edu.ng, gi.akawuku@unizik.edu.ng, sonaikeope@gmail.com, rkusinza@oliviauniversity.ed.bi

ABSTRACT
Nowadays, social media is seen as a typical way to kill time. Concerns over social media's effects on productivity, mental health, and general well-being has grown significantly as a result of its explosive expansion. Overuse frequently results in social media addiction, a behavioral disease marked by excessive use and a loss of self-control. Using demographic and behavioral data, this study presents the design, implementation, and evaluation of a machine learning prediction model to forecast the degree of social media addiction. The research employed a dataset obtained from a public accessible online digital repository called Kaggle containing important features like age, frequency of use, happiness level, productivity loss, and self-control. The dataset collected was subjected to a comprehensive preprocessing, including data cleaning, normalization, encoding, feature scaling, outlier analysis, feature engineering and further pre-processed based on the Machine learning design pipeline. Prior to being trained, four (4) machine learning algorithms, including Random Forest (RF), Logistic Regression, Support (LR) Vector Classifier (SVC), and K-Nearest Neighbors (KNN) classifiers, were implemented and comparatively evaluated. Performance metrics such as accuracy, precision, recall, F1-score, ROC_AUC were used for assessment. Results indicated that the K-Nearest Neighbor achieved the best realistic performance than Random Forest, Logistic Regression, and SVC models, which all gave performance figures that were almost flawless in terms of the metrics with an accuracy of 87% and ROC_AUC of 0.97. The biggest predictors of addiction, according to feature importance analysis, were self-control, productivity loss, frequency of use, and satisfaction level. The result also showed that machine learning has the potential to be a useful predictive tool for the early identification of social media addiction. However, the results' generalizability was limited by bias in the sample and the
absence of representation from African consumers. The study suggests developing real-time monitoring tools, integrating psychological and environmental aspects, and using more diversified datasets.

Keywords:
Social Media Addiction, Machine Learning, Behavioral Analysis, Digital Well-being

INTRODUCTION

In the contemporary digital landscape, the introduction of Internet technology has opened up previously unseen avenues for human communication and socialization, particularly with social network. Social networking often refers to as the series of activity involving sharing information and communication with group of people using the internet application through a website that are specifically designed for the purpose. Social networking specifically, has grown in popularity over the last decade [1]. People are linking and conversing in ways that were previously unimaginable using platforms such as MySpace, Face book, Twitter, WhatsApp, Flickr, and Instagram [2].
An increasing need to use the internet and uncontrollable behaviour, which is typified by an increased desire to use the internet again, are characteristics of internet addiction (IA). Teenagers' everyday life has become more convenient due to the internet, but there are still concerns involved. There is considerable variation in the

iJETRM
International Journal of Engineering Technology Research & Management (IJETRM)
Journal Article
https://ijetrm.com/issue/

prevalence estimates for IA; recent data from China indicate that the incidence of IA among adolescents is steadily increasing, with estimates ranging from 2.2 to 21.5% [3]. This is higher than the rates in the US and numerous European nations. Approximately 1.9% of college students in Hong Kong and Macau were judged to meet the requirements for IA. Adolescents' physical and emotional health are significantly impacted by improper internet use, which is becoming a public health concern [14]. Hence, any addiction that involves utilizing electronic devices, such as a computer, laptop, or mobile phone, for entertainment purposes, such as watching, listening, or playing games, can be classified as a digital addiction [4]. According to [4], there are three distinct forms of digital addiction that people can be diagnosed with: computer game and Internet addiction, social media addiction (SMA), and smartphone addiction.

According to [5], social media otherwise refers to as third party internet -based platform that focuses on social interactions, community-based inputs, and content sharing among its community of user giving contents created by users and not that licensed from third party. Social media is now considered a common way to pass the time. In fact, different social media platforms have become integral to daily life as a means of socialization that could be used for sharing information, opinion, learning new ideas etc. [6]. The pervasive use of social media by individuals of all ages generated a massive amount of data for a variety of research themes, including recommender systems, link predictions, visualization, and social network investigation [7]. Facebook, Instagram, Twitter, and TikTok are not only new ways of social interaction but also new types of behavioural patterns that require analysis [8].

Historically, out of about 7.9 billion people in the world, 4.62 billion people are active on social media with an average time of about 6hrs 58mins per day using an average of two (2) social media platform. Despite the outstanding presence, advantages and benefits they offer to users, it has also opened up a new framework for malicious purposes such as spamming, trolling, and online bullying and linked to negative mental issues called social media addiction (SMA) [9].

SMA is defined as the compulsive and excessive use of social media that interferes with an individual's psychological well-being, relationships, academic or professional performance, and overall quality of life (Keles et al., 2020). Nowadays, the sheer volume of people venting their unhappiness on social media makes it impossible to identify those who are underrepresented. These individuals' viewpoints will be progressive and contingent. Their mental health will be exposed through social media posts, outreach, comments, and criticism [10]. Take for instance, American Academy paediatrics in 2011 [11] and United Kingdom royal Society for Public health (2017) in their separate report identified a phenomenon called Facebook depression and a strong relationship between social media use and mental health based on survey.

Spending too much time on social media causes people to get dissatisfied, unhappy, and afflicted with a variety of illnesses. The sheer volume of people venting their unhappiness on social media makes it impossible to identify those who are underrepresented. These individuals' viewpoints will be progressive and contingent. Their mental health will be exposed through social media posts, outreach, comments, and criticism [11]. In addition, the development of social media addiction is multifaceted brought about by psychological (anxiety, depression, and low self-esteem), social (peer pressure and fear of Missing out) and technological factors (infinite scrolling).

In this digital age, social media platforms have become integral to everyday life. Its explosive growth has offered unprecedented connectivity and changed how people engage, communicate, share, and consume digital material and text, and entertainment. Although there are many advantages to these platforms. However, the pervasive or widespread have also added to the rising problem of social media addiction, which has been connected to negative impacts on social connections, scholastic performance, and mental health called social media addiction (SMA). It is defined as the compulsive and excessive use of social media that interferes with an individual's psychological well-being, relationships, academic or professional performance, and overall quality of life. Studies have revealed that approximately 5-10% of social media user’s exhibit addictive behaviours, with adolescents and young adults being particularly vulnerable due to developmental factors and the platforms' algorithmic designs that encourage prolonged engagement through notifications, infinite scrolling, and personalized content feeds.

The growing incidence of this problem, especially in young adults, emphasizes how urgently improved diagnosis and intervention techniques are needed. Conventional techniques for diagnosing social media addiction, such psychological testing, Bergen Social Media Addiction Scale (BSMAS) or the Internet Addiction Test (IAT), and self-report questionnaires, frequently depend on subjective answers which could be bias, may not fully reflect actual behavioural patterns and limited in scale. These techniques often detect addiction only after significant harm has occurred, lacking the predictive capability to enable early intervention. Moreover, the dynamic nature of social media usage encompasses metrics like session duration, frequency of interactions, content preferences, and emotional response generates vast amounts of behavioural data that remain underutilized in conventional diagnostic frameworks.

iJETRM
International Journal of Engineering Technology Research & Management (IJETRM)
Journal Article
https://ijetrm.com/issue/

This drawback has spurred academics to investigate data-driven methods, such machine learning, to identify addiction tendencies and objectively assess user behaviour. The underutilization of behavioural trends and demographic parameters in current addiction prediction models is another important driving force behind this study. Although machine learning has been effectively employed in previous research to identify problematic social media usage, many of these studies have largely concentrated on user engagement measures without properly accounting for the influence of demographic factors like age, gender, and occupation. The need for a data-driven approaches algorithm such as machine learning and deep learning that integrate the demographic characteristics, behavioural data, and psychological variables with online activity patterns (i.e., a longitudinal research) is required to capture the accurate prediction of the dynamic character of social media addictions and offering a more comprehensive understanding of addiction risk factor.

OBJECTIVES

This study is aimed at developing an efficient and adaptable machine learning based social media addiction model. Hence, the objectives of the research are to examine the existing prediction processes employed in social media addiction predictions, design and develop the computational (machine learning) model for the predictive process behaviour using four machine learning algorithms, implement the predictive model using Python programming language tool and its libraries and evaluate the behavioural properties of the model using several performance metrics. Therefore, the paper presents a Machine learning prediction model for social media addiction.

The organization of the paper is as follows. Section 2 provides a review of the literature or related works, discussing prior research on Social Media Addiction prediction techniques and identifying existing gaps. Section 3 outlines the methodology, that is, the materials and methods used to achieve the stated research objectives of the social media addiction process operation model, including data collection, preprocessing, and model development. The experimental or implementation with its results and findings including model performance evaluation using Accuracy and ROC-AUC score metrics are discussed in section 4. Finally, in Section 5, we provided a conclusion and introduce a further improvement of the research topic.

LITERATURE REVIEW

Social media addiction has become a growing concern in recent years, with numerous studies exploring its psychological, behavioural, and technological aspects. The rapid rise in social media usage has led to increased research efforts aimed at understanding its effects on users and developing predictive models for addiction detection. With the advancement of artificial intelligence and machine learning, researchers have begun leveraging data-driven approaches to analyse user behaviour patterns and predict addiction tendencies. This literature review examines existing studies on social media addiction, the role of demographic and behavioural factors in addiction prediction, and the application of machine learning techniques in identifying addictive behaviours. By analysing previous research, this study aims to build upon established findings and identify gaps that can be addressed through the development of a robust machine learning model for predicting social media addiction.

Most of the research that has been carried out has primarily focused on the psychological aspects of SMA, or social media addiction [12]. According to the uses and gratifications theory postulated by Kales [13], people use social media to satisfy basic needs [14]. However, if the user's needs are not met by the content they receive, they attempt to cope with obsessive-compulsive disorder by viewing more social media pages [15]. This theory serves as a foundation for further analysis of the motivations behind social media use and offers a fundamental explanation of why someone might decide to engage in addictive behaviour [8].

To comprehend the pattern of social media addiction, the study applied the social comparison theory. According to this theory, people compare themselves to other people and consequently feel anxious and low in value [8]. Users often create a negative perception of themselves by comparing themselves to the social media representations of the lives of celebrities or other friends [16]. Nonetheless, studies show a significant positive correlation between social media [8] comparison, and the amount of time spent on SNS [17]. According to this, people may use social networking sites to avoid feeling inadequate, only to end up caught in a vicious cycle of addiction [18].

While a lot of research has been done on social media addiction, little attention has been paid to modeling and predicting addiction behavioural patterns using machine learning techniques. Although the analysis of the mental and behavioural tendencies of addictions [19] has previously benefited from the use of traditional techniques, machine learning shows promise in comprehending the patterns found in large volume of data on the internet. Thanks to recent developments in data analysis, researchers can now incorporate psychological aspects into the prediction process in addition to user behaviour parameters.

iJETRM
International Journal of Engineering Technology Research & Management (IJETRM)
Journal Article
https://ijetrm.com/issue/

Many researchers have documented addiction-related behaviours applying primarily machine learning features in different contexts [17] For instance, using quantitative data, researchers have created predictive models to classify substance use. Even though these approaches offer insightful information about the development of machine learning, their use in treating social media addiction is still lacking. As a result, there is now a research gap on how to combine machine learning with behavioural and psychological theories [20].

RELATED WORKS

[3] addressed the challenge of adolescent internet addiction with a view to examine the risk factors associated with the addiction and provides valuable insights into utilizing machine learning approach to predict adolescent internet addiction. The dataset used consists of 8,824 schoolchildren reports obtained from the Chengdu Positive Child Development (CPCD) through a clustering sampling survey method were analyzed, where 33.3% of participants were identified with Young's Internet Addiction (IA) (Age: 10.97 ± 2.31, Male: 51.73%), and pre-processed for further analysis by the machine learning algorithms. The data set was partitioned into 70% and 30% for training and testing. Further, the authors employed six (6) Machine Learning algorithms (Extra Random Forest, XGBoost, Logistic Regression (LR), Bernoulli Naïve Bayes (BNP), Multi-Layer Perceptron (MLP), and Transformer Encoder) and SHAP to interpret risk factor contributions, particularly the 23 key predictors identified, and predict the performance of the system. The performance of the system was evaluated using 10-fold cross-validation and test sets was performed across survey. Similarly, feature selection and SHapley Additive exPlanations (SHAP) analysis were applied to enhance the model and interpretability respectively. Results showed that ExtraRFC achieved the best performance (Test AUC = 0.854, Accuracy = 0.798, F1 = 0.659), outperforming all other models across most metrics and external validations. This research is methodologically important as it aligns with and contributes to the study on social media addiction prediction by highlighting the necessity of including both behavioural and psychosocial variables in machine learning models to enhance precision and facilitate actionable early intervention strategies. The findings underscore the significant influence of psychological factors in predictive modeling while providing a solid framework for understanding and predicting addictive behaviours in youth.

[21] proposed a machine learning framework to detect and classify social media addiction through user behaviour and mental health data analysis with a view of classifying participants numerically into different levels of social media addiction. To achieve this, the authors utilized the machine learning pipeline methodology. The data used for the research was obtained from an online data repository called Kaggle, where the following tasks (data collection, preparation, preprocessing, standardization, label encoding, feature engineering and model development, etc.) were performed on the data. Random Forest Classifier was used to categorize the addiction level. The model was implemented and evaluated using Python Programming Language tool with its libraries and evaluated using performance metrics. Result shows that the Random Forest ML algorithm employed in classifying the addition level achieved an accuracy of 88.6%, precision, recall and F1-scores of 0.83, 0.84, and 0.89 respectively.

[22] investigated the effect of digital addiction among Malaysian university students during the CoVID-19 pandemic period. Because of the change to distant learning and the lack of in-person engagement during lockdowns, the author highlighted that digital addiction encompasses a wide variety of online behaviours, including social media use. To examine the effect, frequency and severity of SMA on the students, the author applied a machine learning-based algorithm (K-Means clustering) to the dataset, which was collected through surveys (i.e., questionnaires) given to 100 participating students. The survey question has five (5) primary addiction criteria recorded in the dataset, including psychological-emotional, addiction intensity, influence on study time, disruption of social life, and physical impacts. Based on behavioural patterns, the model successfully classified students into three groups: low, middle, and high levels of digital addiction. To find the ideal cluster count, the author employed unsupervised learning approaches and validating their findings using the Elbow Plot, Calinski-Harabasz Index, and Silhouette Scores. The findings indicated that excessive screen time was a major contributing factor to the high addiction cluster, which included a sizable majority of students. Additionally, the high addiction group was more likely to be female and to be enrolled in a public university.

[23] examined the use machine learning techniques to find predictors of smartphone addiction. Data for the experiment was obtained by the Korea Internet and Security Agency (KISA) gathered data from 29,712 participants for the study, which included 27 categories pertaining to personal traits and smartphone usage patterns. In the study, hypothesis was formulated and tested using t-test. Further, three (3) machine learning algorithms namely: random Forest, XGBoost and Decision tree were applied to predict the addiction level of students to smartphone usage based on the usage pattern and the general characteristics of users. The performance of the algorithms was also evaluated with RF achieving the highest accuracy of 82.59%, followed by XGBoost (80.77%) and Decision tree

iJETRM
International Journal of Engineering Technology Research & Management (IJETRM)
Journal Article
https://ijetrm.com/issue/

(74.56%). This Study Illustrates how well machine learning can identify complex patterns, which supports our project's use of behavioural data to predict addiction.

[24], in their study investigated the connection between social media addiction, demographic characteristics and perceived stress levels among youths of Indian. To achieve the aim of the study, the authors used machine learning algorithms (Logistic Regression, Decision Tree, Support Vector Machine (SVM), and Naïve Bayes) to analyse the data collected from 1,011 undergraduate and graduate students between the ages of 18 and 35. Males were more addicted to social media and more stressed than females, with the 18–24 age group being the most impacted, according to key findings. Addiction and stress levels were higher among undergraduates than among postgraduates, indicating that maturity may lessen dependency. Based on the evaluation of the learning algorithms applied, findings revealed that SVM outperformed other machine learning algorithms in gender-based predictions, achieving high accuracy and F1-score of 0.92.

[25] investigated the prevalence and behavioural patterns of social media addiction in Bangladesh. Data used in the study was obtained using survey methods to extract information from the participants with emphasis on 23 characteristics that are commonly associated with addiction, including excessive internet use, emotional dependence, and social disengagement. To examine the data, the authors used five machine learning algorithms: Multinomial Naive Bayes (NB), k-Nearest Neighbors (K-NN), Random Forest (RF), Support Vector Classifier (SVC), and Decision Tree (DT). The Random Forest classifier outperformed K-NN in terms of accuracy, with 82% and 80%, respectively. Result indicated that 81% of girls and 97% of males under the age of 18 engaged in addictive behaviours. The study emphasized the substantial negative effects of social media addiction on mental health, especially among the young populations. Additionally, the study highlights how machine learning may be used to detect those who are at risk and recommends customized interventions to lessen addiction. A strong framework for preparing and interpreting survey data is presented in the paper.

[26] applied machine learning (ML) to examine burnout predictors in 1,576 high school teachers in Hungary. 19.7% of individuals had burnout, as measured by the Mini Oldenburg Burnout Inventory (MOLBI) and Maslach Burnout Inventory (MBI). Result from the experiment revealed that out of the three machine learning algorithms (Decision Tree (DT), Random Forest (RF), and Support Vector Machine (SVM)) applied, RF performed best achieving a score of Area Under cover (AUC) to be 0.811, and balanced accuracy of 0.745.

[27] examined the relationship between social media addiction, happiness, and life satisfaction in adults. Data for the analysis was obtained through a survey using descriptive and cross-sectional study, Elastic Net Regression, Structural Equation Modeling, and SHAP for feature importance. Result showed that the higher social media addiction correlates with lower life satisfaction and happiness as life satisfaction mediates the effect of addiction on happiness.

Despite these consistent problems associated with people, especially the young adults getting addicted to social media, emphases on how urgently improved diagnosis and intervention techniques are needed. This is because, conventional techniques for diagnosing social media addiction, such psychological testing and self-report questionnaires frequently depend on subjective answers and could not fully reflect actual behavioural patterns [28]. Similarly, the underutilization of behavioural trends and demographic parameters in current addiction prediction models is another important driving force. Although machine learning has been effectively employed in previous research to identify problematic social media usage, many of these studies have largely concentrated on user engagement measures without properly accounting for the influence of demographic factors like age, gender, and occupation [29]. This drawback has spurred academics to investigate data-driven methods, such machine learning, to identify addiction tendencies and objectively assess user behaviour [30].

Therefore, the need for a data-driven approaches such as machine learning and deep learning that integrate both demographic characteristics, behavioural data, and psychological variables with online activity patterns (i.e.,a longitudinal research) is required to capture the accurate prediction of the dynamic character of social media addictions and offering a more comprehensive understanding of addiction risk factor.

METHODOLOGY

3.1 Data Acquisition

This research utilizes a dataset comprising 2,000 records related to social media addiction, sourced from Kaggle, a public dataset repository platform that provides access to different datasets for researchers and developers. The dataset contains behavioural patterns affecting Social Media Addiction. The dataset contained 28 features and 2000 records namely (i) Age, Social Media Platform, Total Time Spent, Video Category, Scroll Rate, Satisfaction Level and Addiction Level, (ii) Feature Categories, including demographic variables: Age, Gender, Occupation &

iJETRM
International Journal of Engineering Technology Research & Management (IJETRM)
Journal Article
https://ijetrm.com/issue/

Location and Behavioural attributes such as Frequency of use, Satisfaction, Scroll Rate, Total Time Spent, Productivity loss, Self-control; Target variable: Addiction level (Severe, Moderate & Mild).

3.2. Proposed Model

The methodology adopted by the study aims to develop an efficient and adaptable machine learning-based social media addiction prediction model. Figure 1 depicts the Project workflow for the model development.

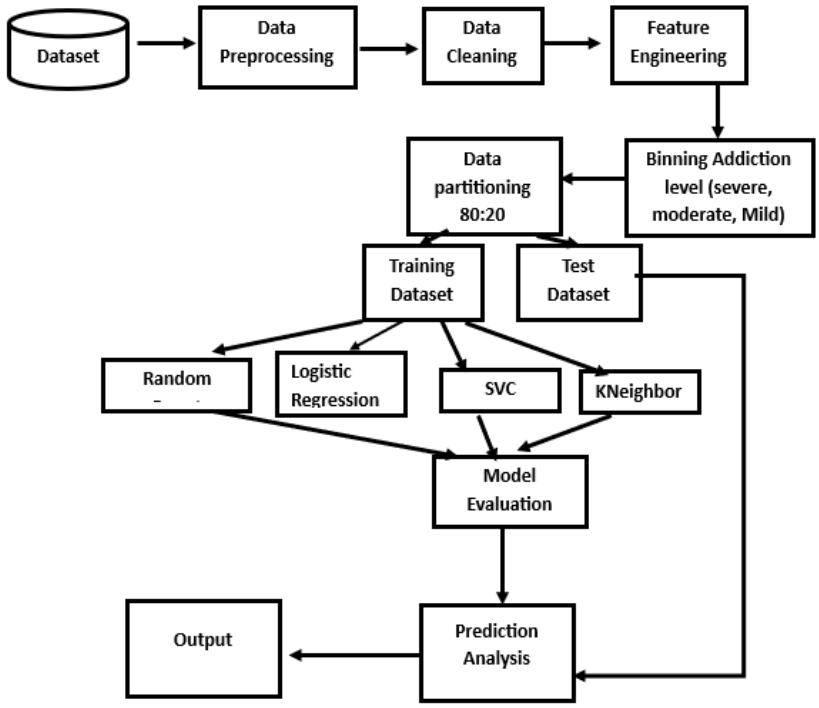

Figure 1. Project Methodology Workflow

3.2.1 Data Preprocessing

After obtaining the dataset, several preprocessing steps were carried out to prepare the dataset for further analysis. This step includes:

1) Data Cleaning: This is the stage where data cleaning pipeline is created and employed to examine the dataset thoroughly for missing and duplicate values. In the dataset to prevent bias in the training process, there is no missing or duplicate records were found.
2) Feature Encoding: At this stage, some unnecessary variables, such as timestamps, were eliminated from the columns to improve the efficiency of the analysis. Gender, marital status, and the use of various social media platforms are examples of categorical variables that were converted into numerical values using label encoding. After checking the data for any missing entries, we applied the appropriate imputation techniques. To minimize bias and maintain the integrity of the dataset, we substituted the median of the corresponding columns for missing values.
3) Feature Scaling: All numerical features were standardized using Scikit -Learn's StandardScaler to improve performance of distance-based models like K-Nearest Neighbors (KNN).
4) Outlier Analysis: Distribution plots were used to detect anomalies. No extreme outlier was found that needed removal or capping.

These preprocessing steps as shown in Figure 2 ensured the data was clean, consistent and optimized for model training.

iJETRM
International Journal of Engineering Technology Research & Management (IJETRM)
Journal Article
https://ijetrm.com/issue/

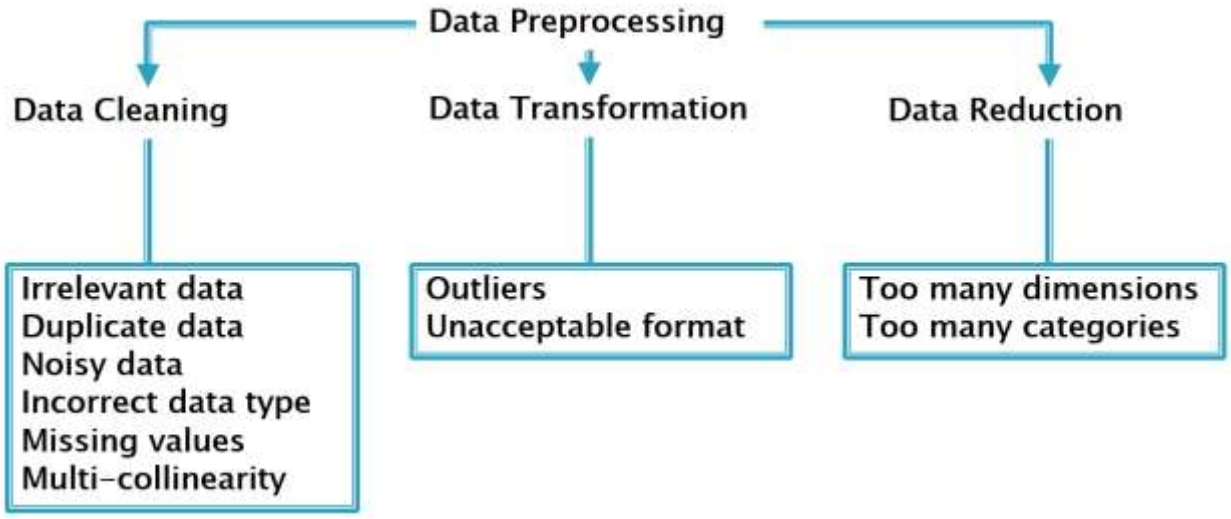

Figure 2 Data Preprocessing Steps.

3.2.2 Feature Engineering

Feature engineering is an important step aimed at improving the performance or sensitivity of the given dataset. The following feature engineering steps were carried out to improve the predictive capability of the machine learning models:

1) Correlation Analysis: A Correlation heatmap was used to identify linear relationships between independent variables and the target variables.
2) Feature Importance Ranking: Tree-based models (Random Forest) feature importance was used to determine the key predictors of addiction.

To increase model efficiency, features that have little to no effect on addiction prediction or show a lot of noise were either eliminated.

3.3 Model Development and Training

In the development process of the model for the prediction of social media addiction, four (4) supervised learning models were developed and compared in this study. They are;

1. Logistic Regression: This is a supervised machine learning algorithm used for classification tasks, where the goal is to predict the probability that an instance belongs to a given class or not, particularly binary classification. The logistic regression model transforms the linear regression function's continuous value output into a categorical value output using a sigmoid function, which maps any real-valued set of independent variables input into a value between 0 and 1. This model is selected as a baseline model for its simplicity and Interpretability.
2. Random Forest Classifier: The RF is selected for its ability to handle overfitting, linear and non-linear data, and feature importance insights.
3. Support Vector Classifier (SVC): This algorithm is included for its effectiveness in high-dimensional spaces.
4. K-Nearest Neighbors (KNN): This algorithm is selected for its simplicity and interpretability in classification tasks. In addition, KNN works better when the dataset is properly scaled, and is not extremely large, capturing similarity-based behavioral patterns and not on interactions. Moreso, owing to its non-parametric nature, non-assumption of linear relation and normal distribution, and not overfittting thus serving as a baseline model which does not require model training but adapts to data changes and multiclass classification.Moreover, the ethical use of machine learning (ML) in mental health prediction requires a careful balance between fostering innovation and respecting human rights. While ML offers the potential to enhance early detection and tailor personalized care, it also presents significant challenges including privacy, data protection, informed consent, and algorithmic bias. To uphold ethical standards, sensitive mental health data collected from the repository with their features (Age, Gender, Occupation and Location, and Behavioural attributes) is securely stored and used only with explicit, fully informed consent to protect individuals' autonomy and confidentiality. Similarly, in the design the ML models are made transparent and fair, actively working to prevent discriminatory outcomes, particularly for marginalized populations.

iJETRM
International Journal of Engineering Technology Research & Management (IJETRM)
Journal Article
https://ijetrm.com/issue/

3.4 Model Evaluation

The performance of the supervised machine learning models was evaluated and compared using Accuracy Score, Classification Report and AUC scores.

1. Accuracy Score: This is the metric that compares the number of accurate predictions to the total number of predictions, providing a clear indicator of the model's overall performance.
2. Receiver Operating Characteristic Curve and Area under Cover (ROC_AUC) Score: This metric evaluates the model's ability to distinguish between classes.
3. Classification Report (Precision, Recall, F1-Score): This metric analyses the model's performance across the addiction levels. The computation of the classification report is represented using the following formula as listed in the equations

$$Precision = \frac{TP}{TP + FP} \quad (1)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (2)$$

$$F1\_Score = 2 * \frac{Prcision * Recall}{Precision + Recall} \quad (3)$$

3.5 Software and Tools

This study used a variety of software tools. Python programming language in Anaconda, environment using Jupyter notebook and several libraries, including Matplotlib and Seaborn, scikit-learn, pandas were used in all aspects of the analysis. These libraries made it possible to create a variety of interactive and static plots that help visualize the relationships in the data, including correlation heatmap, target variable distribution, addiction level, etc. Also, the software editors used in the model development include Visual Studio Code, GitHub and Power BI.

RESULTS AND DISCUSSION

In the experiment, the dataset collected from the online repository was checked for missing and duplicate values, and none is found. Before data was fed into the Machine Learning Models, Exploratory Data Analysis (EDA) was carried out to understand the relationships between each feature. To carry out EDA, the dataset is first split into two (2) parts:80:20 training-testing set, and other analysis like feature importance, etc., were carried out.

4.1 Feature Importance

The determination of the significance of features is a crucial part of Social Media Interaction (SMI) research. In the same vein, evaluating feature importance is another crucial method for determining the causes of SMI. The permutation importance analysis allowed the model to assess the importance of different predictors in the decision-making process as a whole. Similarly, interpreting the feature importance of the best-performing model showed that the best indicators of addiction were frequency of use, self-control, satisfaction, and productivity. The most significant factor was found to be the average amount of time spent on social media, which accounted for the highest frequency of the importance score. The current study confirms previous research that indicates a direct correlation between the amount of time spent on social media platforms and the likelihood of developing addictive behaviors. These observations as shown in Figure 3 give stakeholders (educators, parents, legislators, and mental health professionals) useful information for tackling the negative effects of social media addiction.

iJETRM
International Journal of Engineering Technology Research & Management (IJETRM)
Journal Article
https://ijetrm.com/issue/

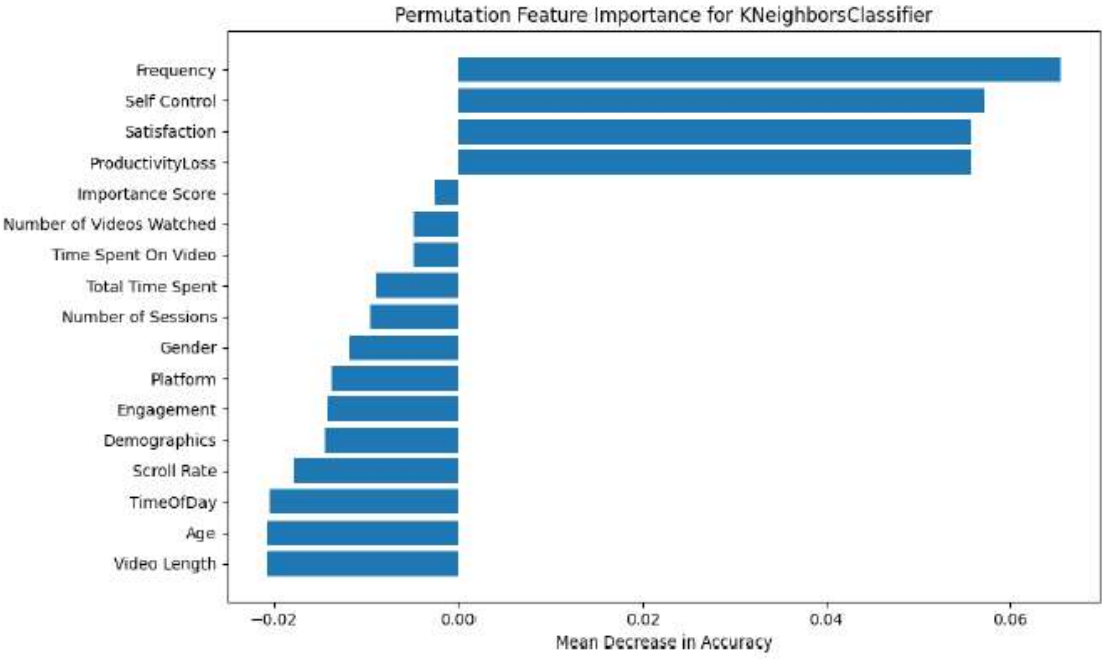

Figure 3 Permutation Feature Importance Using Machine Learning Algorithms

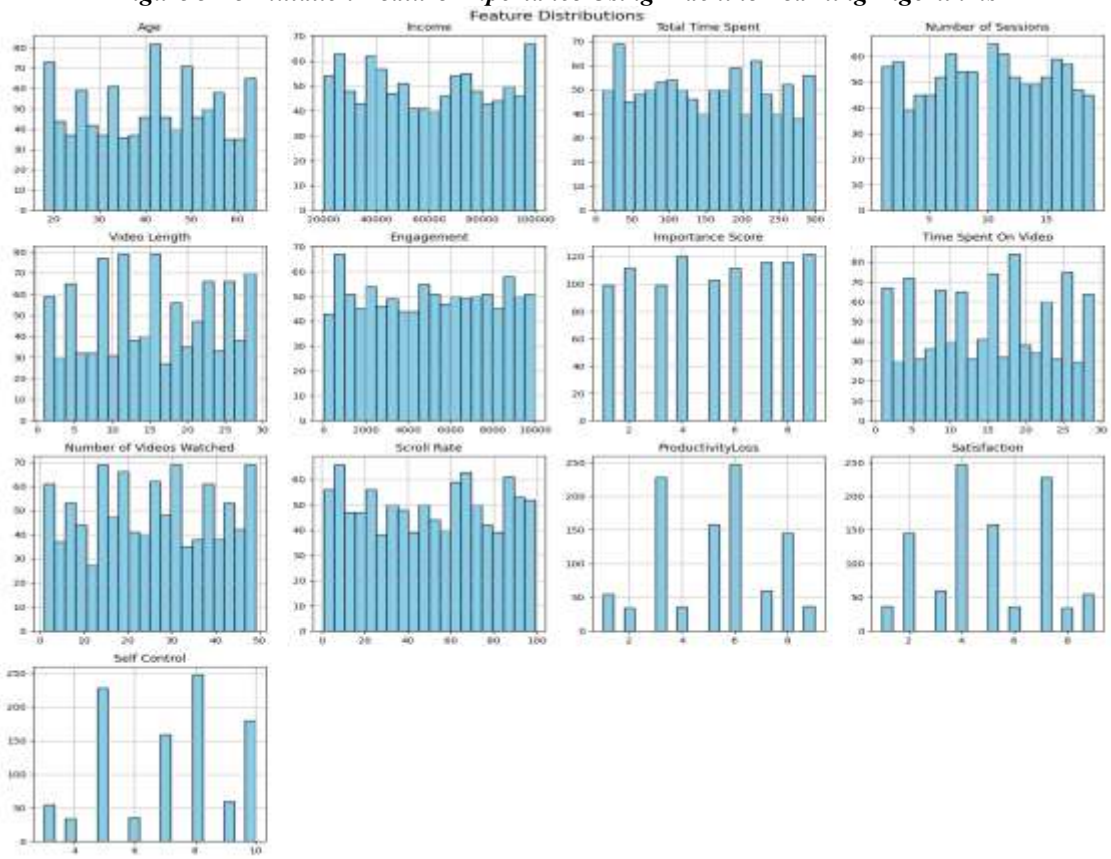

Figure 4. Feature Distributions

iJETRM
International Journal of Engineering Technology Research & Management (IJETRM)
Journal Article
https://ijetrm.com/issue/

4.2 Target variable Distribution and Age Group Analysis

Similarly, an analysis of the participants' addiction levels' percent distribution as shown in Figure 5 revealed a fairly equitable distribution across mild, moderate, and severe levels. The count plot, which shows how frequently study participants experience a particular state, shows that the moderate category had the highest count. Moreover, participants in this study were classified into different age groups. The most represented age group was Middle-age adults (32.8%), followed closely by older adults (31.9%) as shown in Figure 6 (a). In Figure 6 (b), the Older and Middle age adults report higher levels of severe & moderate addictions while younger participants experience more of moderate Levels.

1.3. Demographics & Frequency of Engagement Analysis

Based on demographics & frequency of Engagement Analysis, participants from Rural areas reported more severe and moderate levels of addiction compared to Urban Areas as shown in Figure 6(a) and Figure (b) respectively. Figure 6(c) also reported more severe and moderate levels at night, more moderate levels during the afternoon, and more mild levels during the morning.

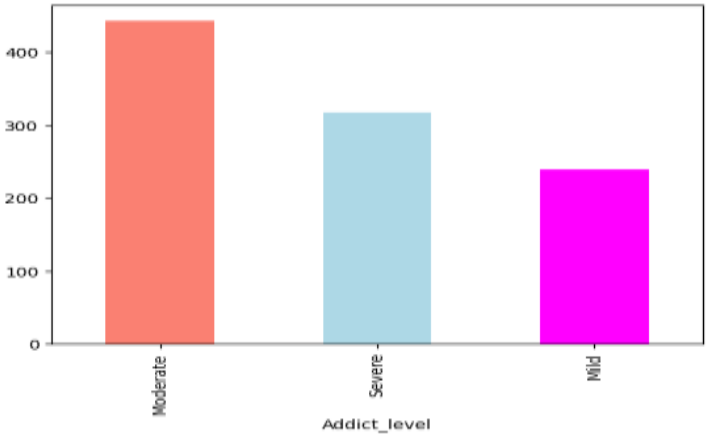

Figure 5 Target variable Distribution

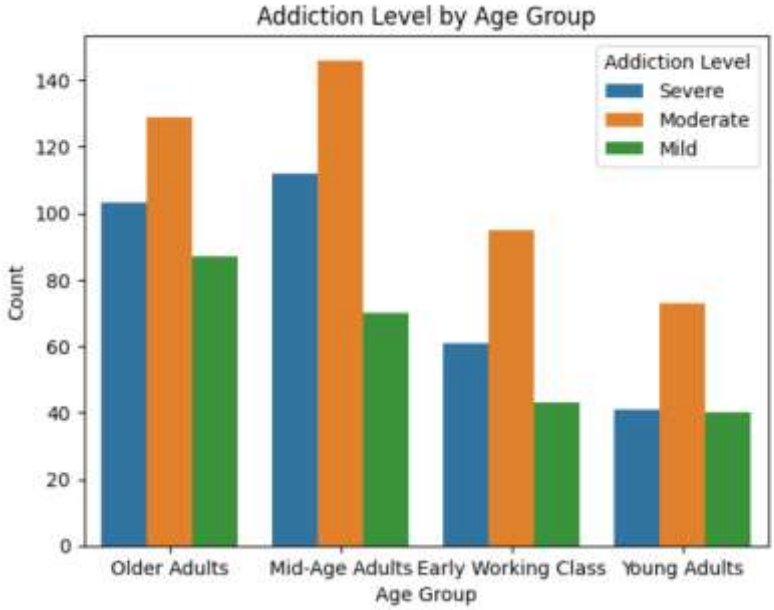

Figure 6 (a). Addiction level by Age Group

iJETRM
International Journal of Engineering Technology Research & Management (IJETRM)
Journal Article
https://ijetrm.com/issue/

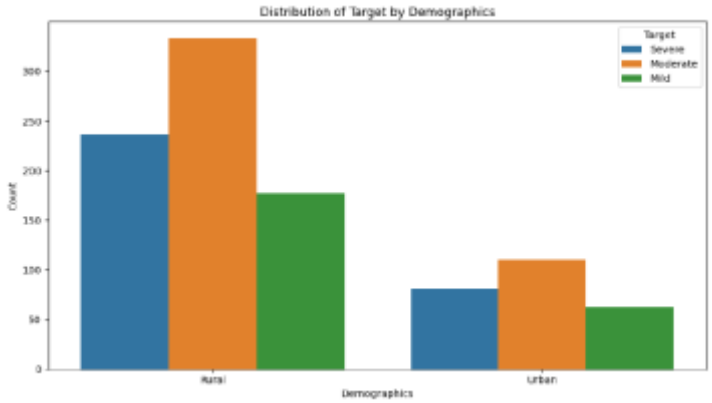

Figure 6 (b). Addiction Level by Demography

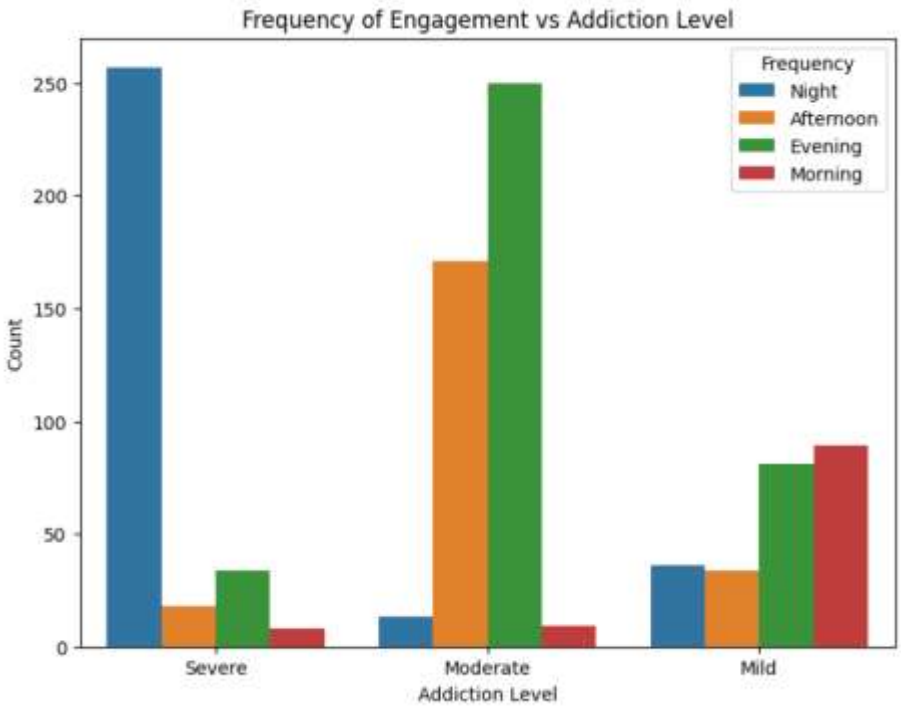

Figure 6(c). Frequency of Engagement Vs Addiction Level

1.4. Correlation Heatmap

A correlation heatmap as shown in Figure 7 was created to investigate the relationship between the dataset's features. Consequently, the chosen graphic depiction showed strong positive correlations between the average amount of time spent on social media and feelings of restlessness and distraction. To be more precise, people who reported completing the main task in less time expressed greater levels of restlessness and distraction when utilizing social media. Similar relationships lend credence to the idea that formal interaction exacerbates emotions and leads to addiction.

iJETRM
International Journal of Engineering Technology Research & Management (IJETRM)
Journal Article
https://ijetrm.com/issue/

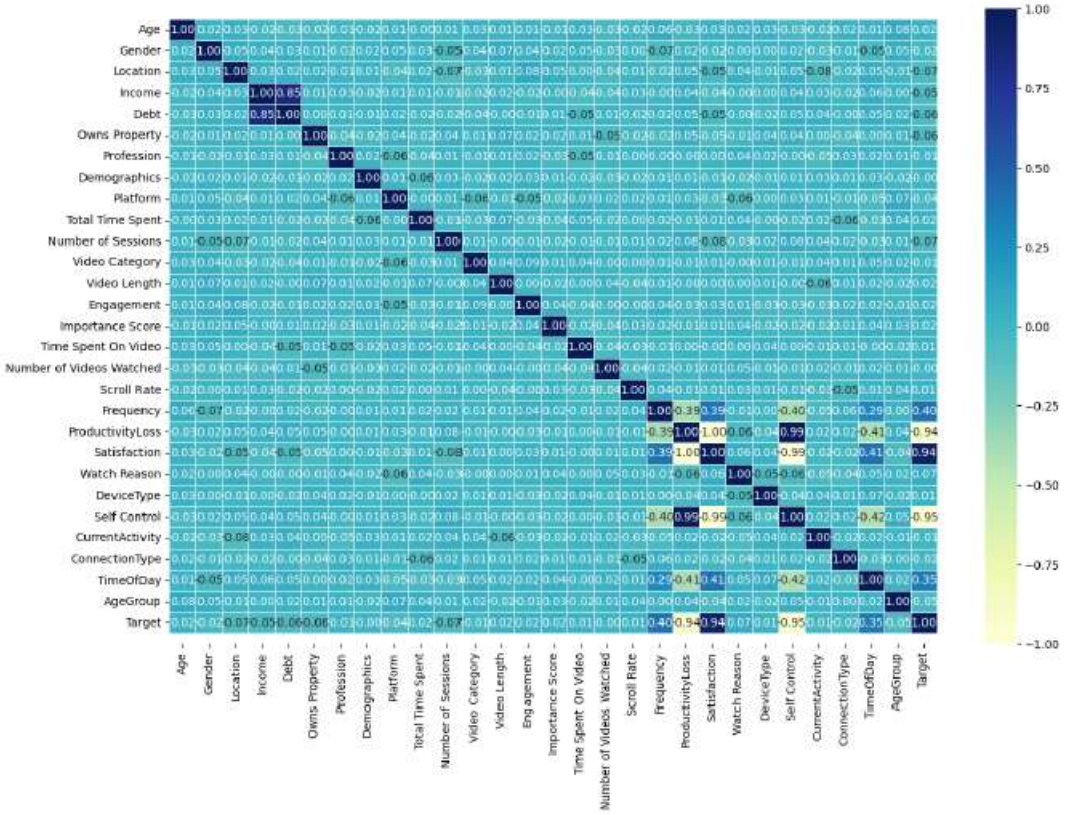

Figure 7. Correlation Heatmap

4.5 Productivity-loss & Self-control

The impact of Productivity-loss & Self-control on participant's addiction levels as illustrated in Figure 8. Figure 8 shows that participants with productivity loss values less than 3 reported severe levels, those within the ranges of 4-6 reported moderate levels and 7+ values reported mild levels. This trend suggests that the higher the productivity loss, the lower the addiction levels tend to get. In Figure 9, participants with low self-control levels reported severe levels, followed closely with average self-control reported medium levels closing out those with extremely high self-control reporting mild levels.

4.6 Model Evaluation Results

This section evaluates and compares the four machine learning models chosen for the study: Random Forest (RF), Logistic Regression (LR), Support Vector Classifier (SVC), and K-Neighbours (KNN) Classifier. After training and testing, the models yielded the following accuracy and ROC_AUC values respectively: Random Forest Classifier achieved (1.0,1.0), Logistic Regression (0.998,1.0), SVC Classifier (0.995,1.0), and K-Neighbours Classifier (0.87,0.97).

After the assessment of the machine learning classifiers (RF, LR, SVC and KNN) results, it is evident that Random Forest, Logistic Regression, and SVC performed exceptionally well with unrealistic results, which is suspicious as depicted in Figure 11.Random Forest, Logistic Regression, and Support Vector Classifier are overfitted as a result of high model complexity in relation to dataset size, multicollinearity among behavioral features, and insufficient regularization. The models captured noise in the training data rather than generalized behavioral patterns exhibited by KNN, thus resulting in reduced performance on unseen test data. However, the model that shows a near realistic performance among all is K-neighbors Classifier, balancing accuracy and generalization, making it more reliable as depicted in Table 1.

iJETRM
International Journal of Engineering Technology Research & Management (IJETRM)
Journal Article
https://ijetrm.com/issue/

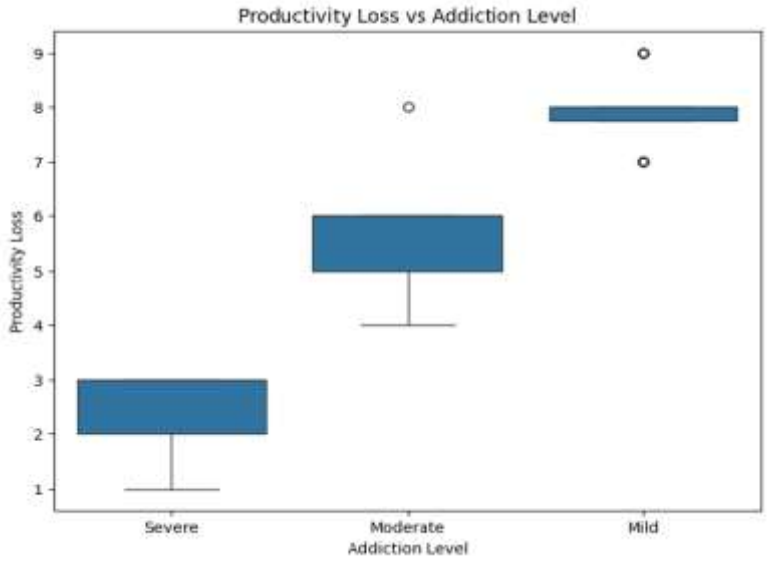

Figure 8 Productivity Loss vs Addiction Level

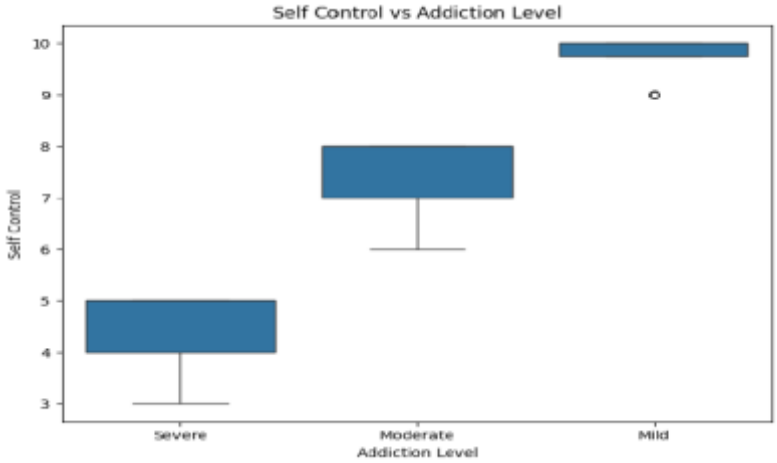

Figure 9. Self Control vs Addiction level

After the assessment of the machine learning classifiers (RF, LR, SVC and KNN) results, it is evident that Random Forest, LogisticRegression, and SVC performed exceptionally well with unrealistic results, which is suspicious as depicted in Figure 10. However, the model that shows a near-realistic performance among all is K-Nearest neighbors Classifier, balancing accuracy and generalization, making it more reliable as depicted in Table 1.

Table 1. Classification Report for K-Nearest neighbors Classifier

	Precision	Recall	F1-score	Support
0	0.82	0.81	0.81	88
1	0.84	0.89	0.86	184
2	0.97	0.88	0.92	128
accuracy			0.87	400
macro avg 0.87		0.86	0.87	400
Weighted avg 0.87		0.63	0.87	400

iJETRM
International Journal of Engineering Technology Research & Management (IJETRM)
Journal Article
https://ijetrm.com/issue/

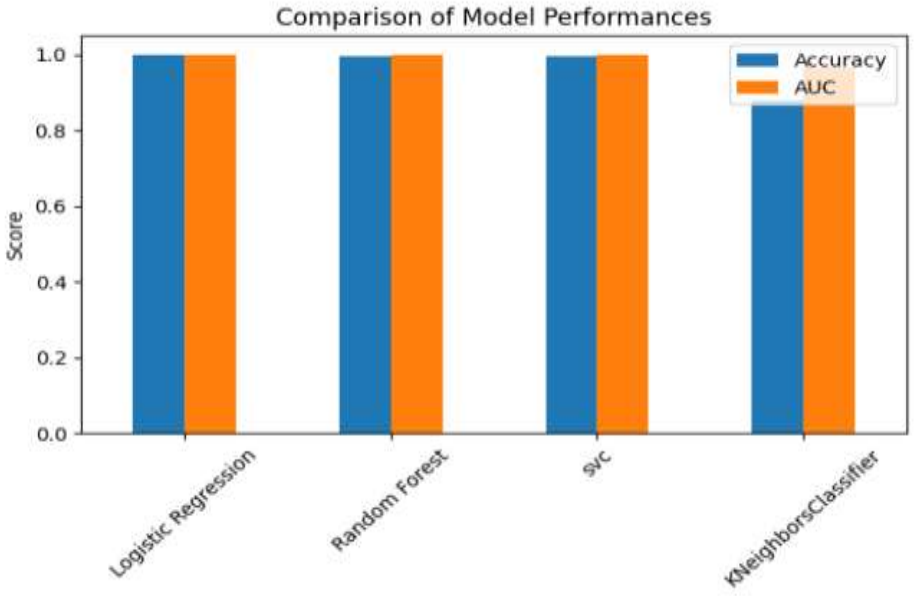

Figure 11 Comparison of Model Performances

4.8 Discussion

The research findings indicate a relationship between social media satisfaction and the degree of addiction, self-control and addiction level, and productivity loss and addiction level. Furthermore, it is essential to know that the result obtained in this study is in tandem with outcomes of previous studies, which reveal that the amount of time expended on social media networks and compulsive engagement predicts the SMA. Previous research has demonstrated that excessive use of social media contributes to poor psychological well-being, including anxiety and depression, among other outcomes, which are forms of illness [31]. The hypothesis that SM addiction is multifaceted and contingent on behavioural patterns and the amount of time spent on SM is also supported by this study. These results are consistent with a 2016 study by Liu Yi Lin et al. (2016), which found that people with higher levels of addiction are more likely to engage in compulsive. However, the present model performed better than existing models developed by[11, 27,21, 23] in that the performance of the model achieved while evaluating using different ML algorithms.

CONCLUSION

The results of this study demonstrate that social media addiction may be accurately predicted using machine learning. Demographic factors were comparatively less important predictors than behavioural measures like productivity loss, satisfaction, and usage frequency. The study concludes that rather than demographic characteristics, behavioural and psychological patterns have a significant impact on social media addiction. When properly evaluated, machine learning models can be useful instruments for identifying those who are at risk. K-Nearest Neighbors produced the most realistic results out of all the models examined, indicating that it can be used for comparable addiction prediction tasks. Furthermore, the predictive models can provide early-warning indicators of social media addiction by demonstration, stating that this study adds to the expanding body of research on AI-driven mental health studies. Nevertheless, the dataset employed in this study, which came from Kaggle, is one of its main limitations. The dataset lacked diversity in participant representation, especially from African regions, despite offering a wealth of information on user behaviour and social media usage patterns. Because of this, the results might not accurately represent the subtleties of African users' behaviour, cultural influences, or digital habits. This regional disparity restricts the results' applicability to other populations, particularly in areas with widely disparate access to digital infrastructure and platform preferences.

Future studies are urged to improve model generalizability and cultural relevance by using more varied and region-specific datasets, particularly from African communities. Model predictions can be validated and addiction-related

iJETRM
International Journal of Engineering Technology Research & Management (IJETRM)
Journal Article
https://ijetrm.com/issue/

behaviours can be better interpreted by including professional advice from behavioural scientists and psychologists. Also, researchers and developers should look on real-time social media usage monitoring apps that employ predictive models to give users who are acting addictively immediate feedback and warnings. Moreso, additional behavioural and environmental factors, such as emotional states, academic achievement, and peer influence, should be included in future research to improve the predictive depth and usefulness of the model.

REFERENCES

[1] Lane, J., & Stuart, F. (2022). How social media use mitigates urban violence: Communication visibility and third-party intervention processes in digital urban contexts. *Qualitative Sociology*, 1-19.
[2] Price-Mitchell, M. (2020, October 02). Teens Discuss Disadvantages of Social Networking. Retrieved from https://www.rootsofaction.com/disadvantages-of-social-networking/.
[3] Liu, J., Chen, L., Chen, Y., Luo, J., Yu, K., Fan, L,,Yong, C., He, H., Liao, S., Ge, Z. and Jiang, L. (2025). Explainable machine learning prediction of internet addiction among Chinese primary and middle school children and adolescents: a longitudinal study based on positive youth development data (2019–2022). Frontiers in Public Health, pp.1-12. https://doi.org/10.3389/fpubh.2025.1590689.
[4] Aziz, N., Nordin, M. J., Abdulkadir, S. J. and Salih, M. M. M. (2021). Digital addiction: systematic review of computer game addiction impact on adolescent physical health. *Electronics*, *10*(9), 996. https://doi.org/10.3390/electronics10090996
[5] Asur, S. and Huberman, B. A. (2010, August). Predicting the future with social media. In *2010 IEEE/WIC/ACM international conference on web intelligence and intelligent agent technology*, 1, pp. 492-499). IEEE..
[6] Malesev, S. and Cherry, M. (2021). Digital and social media marketing-growing market share for construction SMEs. Construction Economic and Building, 21(1):65-82. https://doi.10.1145/3442.3452314.
[7] Ianni, M., Masciari, E., & Sperlí, G. (2021). A survey of Big Data dimensions vs Social Networks analysis. *Journal of Intelligent Information Systems*, 57(1), 73-100. https://doi.org/10.1007/s10844-020-00629-2
[8] Khalaf, A. M., Alubied, A. A., Khalaf, A. M and Rifaey, A. A. (2023). The Impact of Social Media on the Mental Health of Adolescents and Young Adults: A Systematic Review. Cureus.15(8):e42990. https://doi.org/10.7759/cureus.42990.
[9] Nandiraju, A and Verma, A. (2023). Mitigating Social Media-Induced Dopamine Loops through Machine Learning. Available at SSRN. http://dx.doi.org/10.2139/ssrn.4977488.
[10] https://doi.org/10.1542/peds.2011-0054Mim, M. N., Firoz, M., Islam, M. M., Hasan, M and Habib, M. T. (2024). A study on social media addiction analysis on the people of Bangladesh using machine learning algorithms. *Bulletin of Electrical Engineering and Informatics*, *13*(5), 3493–3502. https://doi.org/10.11591/eei.v13i5.5680.
[11] O'Keeffe, G. S. and Pearson, k. S. (2011). Clinical report: The Impact of Social Media on Children, Adolescents, and Families. American Academy of Pediatrics, 1-7.
[12] van den Eijnden, R. J., Lemmens, J. S and Valkenburg, P. M. (2016). The social media disorder scale. *Computers in Human Behavior*, *61*, 478–487. https://doi.org/10.1016/j.chb.2016.03.038.
[13] Keles, B., McCrae, N. and Grealish, A. (2020). A systematic review: The influence of social media on depression, anxiety, and psychological distress in adolescents. International Journal of Adolescence and Youth, Taylor and Francis, 25(1), 79-93. https://doi.org/10.1080/02673843.2019.1590851.
[14] Liu, R., Wang, S., Tarimo, C.S., Feng, Y., Zhao, L., Liu, S., Chen, X., Wu, J, and Zhao, Q. (2025). Building Machine Learning Predictive Models for Adolescent Internet Addiction: Key Findings on Self- Esteem and Resilience Interaction. 1–19.Preprint version 1. Available at Research Square. https://doi.org/10.21203/rs.3.rs-5606509/v1.
[15] Savci, M.., Tekin, A and Elhai, J. D. (2022). Prediction of problematic social media use (PSU) using machine learning approaches, Current Psychology. 41(5): 2755–2764. https://doi.org/10.1007/S12144-020-00794-1
[16] Jin, L., Chen, Y., Wang, T., Hui, P. and Vasilakos, A. V. (2013). Understanding user behavior in online social networks: A survey, IEEE Communication. Magazine. 51(9): 144–150. https://doi.org/ 10.1109/MCOM.2013.6588663.
[17] Choudhury, S., Deb, J. P., Biswas, S. and. Pramanik, A. (2024). Social Media Disorder Scale: Structure, Reliability and Validity in Indian Context, International. *Journal of Experiment Research and Review*, 41, no. Spl 024. 290–304. https://doi.org/10.52756/ijerr.2024.v41spl.024.
[18] Akter, M., Ritu, K. F., Habib, M. T., Rahman, M. S and Ahmed, F. (2022). A Machine Learning Approach To Predict Social Media Addiction During COVID-19 Pandemic. *International Conference on Applied Artificial Intelligence and Computing (ICAAIC)*, Salem, India, pp.401-405. https://doi.org/10.1109/ICAAIC53929.2022.9793193.

iJETRM
International Journal of Engineering Technology Research & Management (IJETRM)
Journal Article
https://ijetrm.com/issue/

[19] Mak, K. K., Lee, K. and Park, C. (2019). Applications of machine learning in addiction studies: A systematic review. *Psychiatry Research*, *275*(March 2019), 53–60. https://doi.org/10.1016/j.psychres.2019.03.001.
[20] Seo, D. B and Ray, S. (2019). Habit and addiction in the use of social networking sites: Their nature, Antecedents, and Consequences, *Computer. Human Behaviour*, vol. 99, pp. 109–125. https://doi.org/10.1016/J.CHB.2019.05.018.
[21] Ehsan, T and Basit, J. (2025). Machine Learning for Detecting Social Media Addiction Patterns: Analyzing User Behavior and Mental Health Data. International Journal of Innovation in Science and Technology (IJIST), 6(4), 1789-1807.https://www.researchgate.net/profile/Tahir-Ehsan
[22] Jamaludin, N. D. (2021). Prediction on Digital Addiction among Undergraduate Students during Pandemic using Machine Learning. . http://www.utp.edu.my/id/eprint/24137.
[23] Lee, J. and Kim, W. (2021). Prediction of problematic smartphone use: A machine learning approach. *International Journal of Environmental Research and Public Health*, *18*(12). 6458-6471. https://doi.org/10.3390ijerph18126458.
[24] Gupta, D. E., Srivastava, D. A., Goyal, D. and Pathak, D. A. (2022). Addiction to Social Media Platforms and Perceived Stress Level Amongst Youth of India: A Machine Learning Approach. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3991411.
[25] Feher, G., Kapus, K., Tibold, A., Banko, Z., Berke, G., Gacs, B., Varadi, I., Nyulas, R. and Matuz, A. (2024). Mental issues, internet addiction and quality of life predict burnout among Hungarian teachers: a machine learning analysis. *BMC Public Health*, *24*(1), 1–15. https://doi.org/10.1186/s12889-024-19797-9.
[26]Çiftci, N. and Yıldız, M. (2023). The Relationship between Social Media Addiction, Happiness, and Life Satisfaction in Adults: Analysis with Machine Learning Approach. *International Journal of Mental Health and Addiction*, *21*(5), 3500–3516. https://doi.org/10.1007/s11469-023-01118-7.
[27] Andreassen, C. S., Torsheim, T., Brunborg, G. S. and Pallesen, S. (2012). Development of a Facebook Addiction Scale. Psychological Reports, 110(2), 501-517. https://journals.sagepub.com/doi/abs/10.2466/02.09.18.pr0.110.2.501-517.
[28] Hussain, Z., Griffiths, M. D. and Sheffield, D. (2020). An investigation into problematic social media use: The role of attachment, fear of missing out, and narcissism. Computers in Human Behavior, 6, 378–386. https://doi.org/10.1556/2006.6.2017.052.
[29] Brailovskaia, J., Margraf, J. and Köllner, V. (2021). Addictive social media use and mental health: A longitudinal study among German Facebook users. International Journal of Environmental Research and Public Health, 18(17), 8976. https://doi.org/10.1016/j.chb.2021.106720.
[30] Kuss, D. J. and Griffiths, M. D. (2012). Internet Gaming Addiction: A Systematic Review of Empirical Research," *International Journal of Mental Health Addiction*. 10 (2): 278–296. https://doi.org/10.1007/S11469-011-9318-5.