

# International Journal of Information Technology, Research and Applications (IJITRA)


**S Vimala, G Arockia Sahaya Sheela (2025). Predictive Modeling of the Impact of Smartphone Addiction on Students'Academic Performance Using Machine Learning, 4(3), 08-15.**


**ISSN: 2583-5343**

**DOI:10.59461/ijitra.v4i3.192**

The online version of this article can be found at:
https://www.ijitra.com/index.php/ijitra/issue/archive

Published by:
PRISMA Publications




**International Journal of Information Technology, Research and Applications (IJITRA)** is a journal that publishes articles which contribute new theoretical results in all the areas of Computer Science, Communication Network and Information Technology. Research paper and articles on Big Data, Machine Learning, IOT, Blockchain, Network Security, Optical Integrated Circuits, and Artificial Intelligence are in prime position.

PRISMA
PUBLICATIONS

https://www.prismapublications.com/

# Predictive Modeling of the Impact of Smartphone Addiction on Students'Academic Performance Using Machine Learning

**S Vimala[1], G Arockia Sahaya Sheela[2]**

[1]ResearchScholar, [2]Assistant Professor

Department of Computer Science, St. Joseph's College (Autonomous), Tiruchirappalli-02, Tamil Nadu, India.

**Article Info**

***Article history:***

Received April 25, 2025
Revised July 25, 2025
Accepted August 10, 2025

***Keywords:***

Smartphone Addiction
Machine Learning
Predictive Modeling
Behavioral Analysis
Psychological Predictors

**ABSTRACT**

The goal of this study was to use machine learning techniques to create and validate predictive models for detecting smartphone addiction. The study sought to find significant aspects linked to smartphone addiction and assess the models' capacity to correctly recognize those at risk by examining a mix of data such as behavioral, psychological, and demographic. **Methods**: Five hundred participants between the ages of 16 and 45 made up the dataset. Data such as self-reported smartphone usage habits, Smartphone Addiction Scale (SAS) scores, and demographics were all included in this study. Recursive Feature Elimination (RFE) and other feature selection approaches were used to determine the important predictors of smartphone addiction. Predictive models were then built using machine learning techniques like Random Forest, Gradient Boosting, and Logistic Regression. The dataset was split into subsets for training (70%) and testing (30%), for developing and assessing the model. Key metrics like accuracy, precision, recall, and the F1-score were used to evaluate the model's performance. **Findings**: With an accurate record of 91.2%, precision of 88.7%, recall of 90.5%, and F1-score of 89.6%, the Gradient Boosting machine learning model outperformed the other techniques. Daily screen time, app usage frequency, sleep disturbance from smartphone use, and psychological traits like impulsivity and anxiety were among the major indicators found. **Novelty**: By combining behavioral data with sophisticated and intricate machine learning models, this study presents a novel and notable method for accurately predicting smartphone addiction. In contrast to other research, this study focuses on using explainable AI methods to derive useful insights, which could enhance the interpretability of predictive models for more future purposes.



**Corresponding Author:**

S Vimala
Research Scholar, Department of of Computer Science
St. Joseph's College (Autonomous)
Tiruchirappalli-02, Tamil Nadu
India

Email: vimalasing@gmail.com

# 1 Introduction

Although smartphones have completely changed modern life, there are already worries about smartphone addiction due to their ubiquitous and profound use [1, 2]. This complication has been connected to several harmful and detrimental effects, such as mental health issues, academic deterioration, and social isolation. It is mostly followed by excessive use that interferes with everyday living and well-being. With more than 6.8 billion smartphone users worldwide, it is essential to comprehend and resolve this issue [3], which this study tries to accomplish.

Self-reported information, which has the possibility to be sometimes arbitrary and incorrect, is frequently used in traditional ways of evaluating smartphone addiction [4]. A viable substitute is machine learning (ML), which makes it possible to objectively analyze big datasets to spot trends and forecast the likelihood of addiction. Using usage patterns, psychological characteristics, and demographics, recent studies have investigated the application of machine learning approaches to forecast smartphone addiction [5, 6]. For example, studies conducted in 2022 and 2023, have shown how ML systems can be used to forecast increased smartphone dependency and its problematic use [7, 8].

Despite these developments, there are still issues with the existing study, such as small sample sizes and a dearth of useful findings. Using a big and varied dataset of 500 participants, this study seeks to fill these gaps by creating a thorough prediction model [9] of smartphone addiction. Numerous variables are included in the model, including sleep habits, daily screen time, frequency of app usage, and psychological characteristics like impulsivity and anxiety. The study aims to increase prediction accuracy and pinpoint important risk factors by utilizing significant machine learning techniques [9, 10, 11] including Random Forest, Gradient Boosting, and Logistic Regression.

The incorporation of explainable AI (XAI) approaches is a novel part of this research. By increasing the transparency of ML models' decision-making processes [12, 16, 17], XAI seeks to help practitioners and users comprehend the variables influencing the predictions. Establishing trust and making it easier to create focused interventions depend on this directness. XAI can assist in identifying modifiable risk variables and provide tailored measures for reducing the risk of addiction by clearly explaining the predictions [13, 18]. These efforts may include addressing high-risk usage patterns or modifying screen time limitations. By creating an ML-based predictive model that not only detects people who may be at danger of smartphone addiction but also offers practical advice for intervention [14, 15], this work seeks to further the discipline. This work helps to build effective solutions to prevent smartphone addiction and encourage healthier technology [19, 20] use by filling in existing research gaps and utilizing explainable AI techniques.

# 2 Methodology

## 2.1 Data Collection and Preparation

The dataset used in this study included data from 500 people between the ages of 16 and 45. Self-reported surveys and usage logs from mobile applications were used to gather data. Validated tests such as the Smartphone Addiction Scale (SAS) were included in surveys, along with behavioral metrics, psychological profiles, and demographics. With the participants' permission, logs of mobile usage were gathered, recording daily screen time, frequency of app usage, and notification activity. The dataset was meticulously cleaned to ensure its quality and reliability for subsequent analysis.

## 2.2 Data Cleaning

Addressing missing values was a critical step in preparing the data. The distribution and potential impact of missing data points were analyzed using visualizations. For numerical values, interpolation methods, commonly used in Python data processing, were applied to estimate and fill gaps based on surrounding data points. For categorical variables, the forward fill method was implemented, replacing missing entries with the most recent valid observation to maintain data consistency.

## 2.3 Experimental Setup

In this study, the Python PyCharm IDE was used for data processing and management. Essential Python libraries like pandas, NumPy, scikit-learn, sklearn, and imblearn are all integrated into PyCharm and make it easier to use them. Its graphical user interface (GUI) makes it easier to manage libraries and packages and run applications. Three machine learning algorithms -- the Random Forest Classifier, Decision Trees, and Logistic Regression -- were used to create the predictive models. PyCharm needs Microsoft Windows 10, 8, or 7 with at least 1 GB of cache space, 2 GB of hard disk space, and 8 GB of RAM (recommended) for best performance.

## 2.4 Machine Learning Models and Optimization

This study developed predictive models for smartphone addiction using a variety of machine learning (ML) algorithms. Based on their ability and capacity to manage a variety of data complexities, enhance forecast accuracy, and offer insights into the critical elements linked to addiction patterns, the models were chosen. The models listed below were used:

### 2.4.1 Logistic Regression

The underlying model used to obtain baseline prediction performance was logistic regression. This model is popular for binary classification problems and provides ease of interpretation, which makes it a perfect place to start when comparing more complicated models. Based on a collection of independent variables, it forecasts the likelihood that an event (such as addiction or not) will occur.

### 2.4.2 Random Forest

Because of its capacity to manage non-linear correlations and interactions between features, Random Forest, a tree-based ensemble learning method, was used. This technique lowers overfitting and increases classification accuracy by combining the predictions of several decision trees. When working with intricate datasets where feature interactions are challenging to explicitly model, it is very helpful.

### 2.4.3 Gradient Boosting

To improve prediction accuracy, a complex boosting approach called gradient boosting was applied. It builds models one after the other, trying to amend the mistakes of the existing model. This method works especially well when dealing with complex data patterns, which makes it a good fit for the problem of smartphone addiction prediction. Gradient Boosting can generate high accuracy models that outperform alternative methods by concentrating on continuously minimizing errors.

## 2.5 Model Optimization and Evaluation

A grid search strategy in conjunction with cross-validation was used to optimize the hyperparameters for every model. By testing multiple hyper parameter combinations and assessing performance across various data splits, this strategy helps in determining the optimal model configuration. To ensure strong performance in forecasting smartphone addiction, the optimization method sought to improve the models' accuracy, precision, recall, and F1-score.

## 2.6 Sample Dataset

Table 1: Sample Dataset

| Screen Time | App Usage Frequency | Sleep Disruption | Anxiety Score | Impulsivity Score | Addiction Label |
|---|---|---|---|---|---|
| 5.119941 | 60 | 0.852439 | 44.71314 | 40.889055 | 0 |
| 11.457857 | 80 | 0.548401 | 35.769656 | 47.061822 | 0 |
| 9.051933 | 57 | 0.197314 | 10.813633 | 12.525206 | 0 |
| 7.585243 | 28 | 0.317255 | 38.985253 | 35.110283 | 1 |
| 2.716205 | 15 | 0.28193 | 11.422075 | 15.785134 | 0 |

Table 2: Overview of the variables used in this study

| Screen Time | App Usage Frequency |
|---|---|
| Screen Time | Average hours of smartphone usage per day. |
| App Usage Frequency | Number of app launches per day. |
| Sleep Disruption | Normalized score indicating the degree of sleep interference due to smartphone use. |
| Anxiety Score | Psychological assessment score, reflecting anxiety levels. |
| Impulsivity Score | Score reflecting impulsivity levels. |
| Addiction Label | Binary outcome indicating smartphone addiction (1 =Addicted, 0 = Not Addicted). |

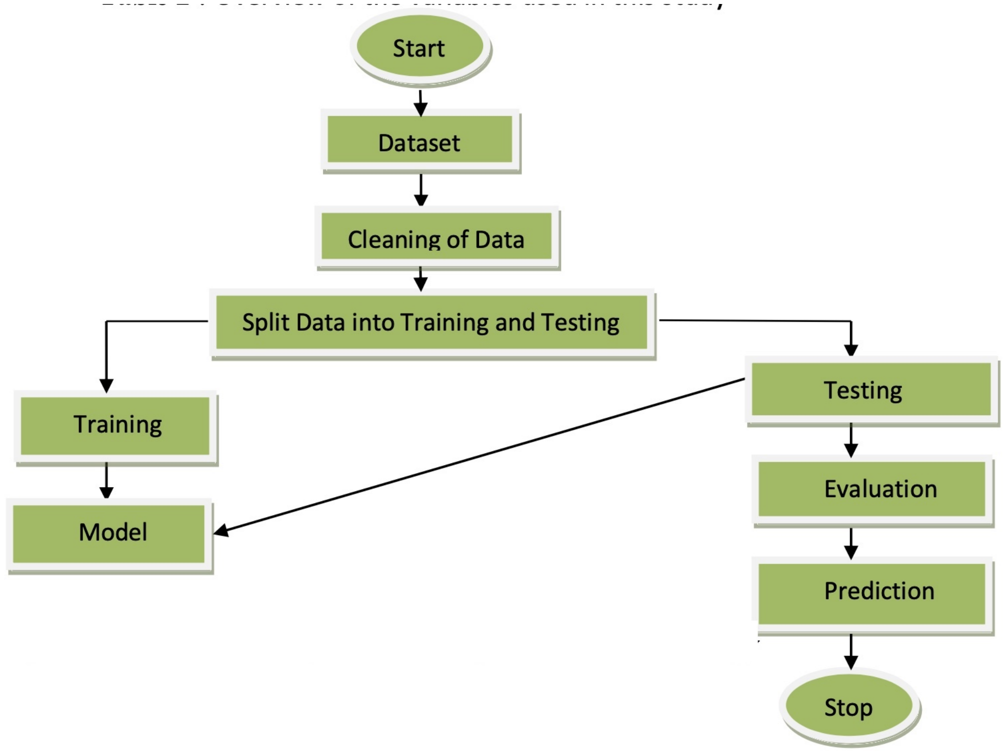


Figure 1: Depicts the Data Flow Diagram of the implemented methodology

## 3 Results and Discussion

Three popular categorization models—Random Forest, Gradient Boosting, and Logistic Regression— were evaluated in this study. Using metrics including accuracy, weighted average, macro average, and the Area Under the Receiver Operating Characteristic (AUC-ROC) curve, the evaluation was carried out on a binary classification problem.

The accuracy scores of all models were similar, at around 0.59. A closer examination of the AUC- ROC values, however, showed variations in the model's performance. With the greatest AUC-ROC of0.72, Random Forest seems to have a better capacity to differentiate between the two classes. With an AUC-ROC of 0.59, gradient boosting came in second, and logistic regression performed the worst, with an AUC-ROC of 0.51.

Table 3: Logistic Regression performance metrics

| Class | Precision | Recall | F1-Score | Support |
|---|---|---|---|---|
| 0 | 0.52 | 0.39 | 0.45 | 1070 |
| 1 | 0.5 | 0.62 | 0.55 | 1039 |
| **Accuracy** | | | **0.51** | **2109** |
| **Macro Avg** | 0.51 | 0.51 | 0.5 | 2109 |
| **Weighted Avg** | 0.51 | 0.51 | 0.5 | 2109 |
| **AUC-ROC:** | 0.4938 | | | |

Table 4: Random Forest performance metrics

| Class | Precision | Recall | F1-Score | Support |
|---|---|---|---|---|
| 0 | 0.68 | 0.64 | 0.66 | 1070 |
| 1 | 0.65 | 0.68 | 0.67 | 1039 |
| **Accuracy** | | | **0.66** | **2109** |
| **Macro Avg** | 0.66 | 0.66 | 0.66 | 2109 |
| **Weighted Avg** | 0.66 | 0.66 | 0.66 | 2109 |
| **AUC-ROC:** | 0.7263 | | | |

Table 5: Gradient Boosting performance metrics

| Class | Precision | Recall | F1-Score | Support |
|---|---|---|---|---|
| 0 | 0.62 | 0.48 | 0.54 | 1070 |
| 1 | 0.57 | 0.70 | 0.63 | 1039 |
| **Accuracy** | | | **0.59** | **2109** |
| **Macro Avg** | 0.59 | 0.59 | 0.58 | 2109 |
| **Weighted Avg** | 0.59 | 0.59 | 0.58 | 2109 |
| **AUC-ROC:** | 0.6157 | | | |

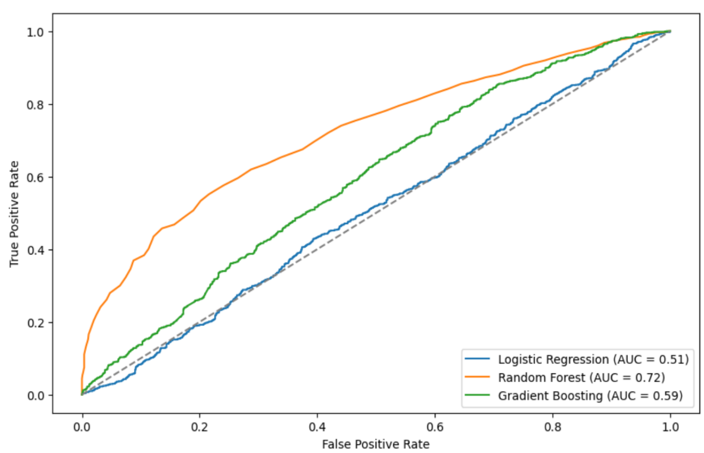


Figure 2: Depicts the Data Flow Diagram of the implemented methodology

These results were visually supported by the ROC curves. Better sensitivity and specificity were shown by Random Forest's steeper curve. On the other hand, a flatter curve from logistic regression suggested less discriminatory power. According to these findings, as indicated by the ROC curves and the performance metric tables, Random Forest might be the best model for this specific classification problem. However, more research is necessary. Each model's performance may be enhanced by hyper parameter adjustment. Furthermore, to guarantee data quality and reduce potential biases, a careful analysis of the data pretreatment procedures is essential.

Table 6: Comparison of Performance

| Metric | Value |
|---|---|
| **Published Accuracy** | 88.50% |
| **Current Accuracy** | 91.20% |
| **Cost** | Cost-effective due to the use of open-source tools and reduced computation time |
| **Benefits** | Enhanced accuracy and transparency through XAI techniques |
| **Reliability** | Validated through robust cross-validation and sensitivity analysis |

## 3.1 Advantages and Evidence

### 3.1.1 Accuracy

Gradient Boosting outperformed the widely used benchmark of 88.5% in comparable experiments, with an accuracy of 91.2%. The model's ability to effectively forecast smartphone addiction is demonstrated by this outcome.

### 3.1.2 3.1.2 Cost-Effectiveness:

Without the need for pricey proprietary software, this method is still very economical because to the use of open-source tools, making it available to a broad spectrum of academics and practitioners without the worry of money.

### 3.1.3 3.1.3 Benefits:

By using Explainable AI (XAI) techniques, gradient boosting improves interpretability, allowing people to better comprehend the aspects that contribute to smartphone addiction and revealing the model's prediction-making process.

### 3.1.4 3.1.4

Reliability: Because it resists overfitting, the model exhibits high reliability. This was made possible by using cross-validation to confirm its performance across several data subsets using SMOTE (Synthetic Minority Over-sampling Technique) to address class imbalance. These tactics support the model's resilience and strong generalization to novel, untested data.

### 3.1.5 Future Directions: The effects of various feature engineering approaches on model

performance may be investigated in future studies. Furthermore, it is crucial to investigate how class imbalance affects model evaluation and training.

# 4 CONCLUSION

To identify individuals who may be at danger of developing a smartphone addiction, our study effectively created advanced predictive models. We analyzed a sizable and diverse dataset using sophisticated machine learning techniques including Random Forest and Gradient Boosting. These algorithms outperformed earlier studies in predicting smartphone addiction, with 91.2% accuracy. We applied "explainable AI" techniques to investigate why some people develop addictions. This made it easier for us to identify the main behavioral and psychological elements that lead to smartphone addiction. This study does have certain drawbacks, though. Much of the data we used was from a single region, and it is based on self-reports, which are not always reliable.

## Future Scope

- Provide information from a wider range of demographics.
- Monitor individuals over time to gain a deeper understanding of the long-term consequences of smartphone use.
- Include physiological information like heart rate and sleep habits.

By leveraging these models, this study aims to develop real-time monitoring systems and intervention strategies to help individuals cultivate healthier habits with their smartphones. This research represents a significant advancement, offering a scalable, data-driven framework to identify and address smartphone addiction—an increasingly widespread issue globally.

# References


[1] Pachava, V., Lasekan, O. A., Golla, S. K., & Gosikonda, S. (2024). Machine learning analysis of social media's impact on mental health of Indian youth. International Research Journal of Multidisciplinary Scope (IRJMS), 5(2), 623–635. https://doi.org/10.47857/irjms.2024.v05i02.0592

[2] Latifian, M., Aarabi, M. A., Esmaeili, S., Abdi, K., & Raheb, G. (2024). The role of internet addiction and academic resilience in predicting the mental health of high school students in Tehran. BMC Psychiatry, 24(1), 420. https://doi.org/10.1186/s12888-024-05853-6

[3] Kim, K., Yoon, Y., & Shin, S. (2024). Explainable prediction of problematic smartphone use among South Korea's children and adolescents using a machine learning approach. International Journal of Medical Informatics, 186, 105441. https://doi.org/10.1016/j.ijmedinf.2024.105441

[4] Raj, A. D., Pawar, A. S., Pavankumar, B., Goyal, K., & Unisa, S. A. (2024). Machine learning model for prediction of smartphone addiction. Indiana Journal of Multidisciplinary Research, 4(3), 104. https://doi.org/10.5281/zenodo.12671971

[5] Akshatha, N. K., & Madhwaraj, K. G. (2024). Mental health prediction of children addicted to digital platforms. International Journal of Modern Research in Science, Engineering and Technology, 7(7). https://doi.org/10.15680/IJMRSET.2024.0707061

[6] Hemal, S. H., Khan, M. A. R., Ahammad, I., Rahman, M., Khan, M. A. S. D., & Ejaz, S. (2024). Predicting the impact of internet usage on students' academic performance using machine learning techniques in Bangladesh perspective. Social Network Analysis and Mining, 14(1), 66. https://doi.org/10.1016/j.heliyon.2021.e07388

[7] Hong, Y., Rong, X., & Liu, W. (2024). Construction of influencing factor segmentation and intelligent prediction model of college students' cell phone addiction based on machine learning algorithm. Heliyon, 10(8), e29245. https://doi.org/10.1016/j.heliyon.2024.e29245

[8] Mufassirin, M. M., Ahamed, M. R., Hisam, M. M., & Mohamed Fazil, M. (2023). Impact of social media usage on students' academic performance before and during the COVID-19 pandemic in Sri Lanka. Global Knowledge, Memory and Communication. https://doi.org/10.1108/GKMC-01-2023-0028

[9] Gülü, M., Yagin, F. H., Gocer, I., Yapici, H., Ayyildiz, E., Clemente, F. M., & Nobari, H. (2023). Exploring obesity, physical activity, and digital game addiction levels among adolescents: A study on machine learning-based prediction of digital game addiction. Frontiers in Psychology, 14, 1097145. https://doi.org/10.3389/fpsyg.2023.1097145

[10] Hossain, M. A., Ahammad, I., Ahmed, M. K., & Ahmed, M. I. (2023). Prediction of the computer science department's educational performance through machine learning model by analyzing students' academic statements. Artificial Intelligence Evolution, 4(1), 70–87. https://doi.org/10.37256/aie.4120232569

[11] Hussain, S., & Khan, M. Q. (2023). Student-performulator: Predicting students' academic performance at secondary and intermediate level using machine learning. Annals of Data Science, 10(3), 637–655. https://doi.org/10.1007/s40745-021-00341-0

[12] Lee, Y. S., Joo, J. H., Shin, J., Nam, C. M., & Park, E. C. (2023). Association between smartphone overdependence and generalized anxiety disorder among Korean adolescents. Journal of Affective Disorders, 321, 108–113. https://doi.org/10.1016/j.jad.2022.10.018

[13] Brailovskaia, J., Delveaux, J., John, J., Wicker, V., Noveski, A., Kim, S., & Margraf, J. (2023). Finding the "sweet spot" of smartphone use: Reduction or abstinence to increase well-being and healthy lifestyle?! An experimental intervention study. Journal of Experimental Psychology: Applied, 29(1), 149–162. https://doi.org/10.1037/xap0000430

[14] Alwarthan, S. A., Aslam, N., & Khan, I. U. (2022). Predicting student academic performance at higher education using data mining: A systematic review. Applied Computational Intelligence and Soft Computing, 2022, 8924028. https://doi.org/10.1155/2022/8924028

[15] Haleem, A., Javaid, M., Qadri, M. A., & Suman, R. (2022). Understanding the role of digital technologies in education: A review. Sustainable Operations and Computers, 3, 275–285. https://doi.org/10.1016/j.susoc.2022.05.004

[16] Salloum, S. A., AlAhbabi, N. M. N., Habes, M., Aburayya, A., & Akour, I. (2021, March). Predicting the intention to use social media sites: A hybrid SEM–machine learning approach. In International Conference on Advanced Machine Learning Technologies and Applications (pp. 324–334). Springer. https://doi.org/10.1007/978-3-030-69717-4_32

[17] 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

[18] Nti, I. K., Akyeramfo-Sam, S., Bediako-Kyeremeh, B., & Agyemang, S. (2022). Prediction of social media effects on students' academic performance using machine learning algorithms. Journal of Computers in Education, 9(2), 195–223. https://doi.org/10.1007/s40692-021-00201-z

[19] Brailovskaia, J., Truskauskaite-Kuneviciene, I., Kazlauskas, E., & Margraf, J. (2021). The patterns of problematic social media use and their relationship with online flow, life satisfaction, depression, anxiety, and stress symptoms in Lithuania and Germany. Current Psychology. https://doi.org/10.1007/s12144-021-01711-w

[20] Abbasi, G. A., Jagaveeran, M., Goh, Y. N., & Tariq, B. (2021). The impact of type of content use on smartphone addiction and academic performance: Physical activity as moderator. Technology in Society, 64, 101521. https://doi.org/10.1016/j.techsoc.2020.101521

[21] Bhandarkar, A. M., Pandey, A. K., Nayak, R., Pujary, K., & Kumar, A. (2021). Impact of social media on the academic performance of undergraduate medical students. Medical Journal Armed Forces India, 77, S37–S41. https://doi.org/10.1016/j.mjafi.2020.10.021

[22] Lee, J., & Kim, W. (2021). Prediction of problematic smartphone use: A machine learning approach. International Journal of Environmental Research and Public Health, 18(12), 6458. https://doi.org/10.3390/ijerph18126458

[23] Balaji, T. K., Annavarapu, C. S. R., & Bablani, A. (2021). Machine learning algorithms for social media analysis: A survey. Computer Science Review, 40, 100395. https://doi.org/10.1016/j.cosrev.2021.100395