E-ISSN: 3025-0994
Halaman 189-198
Volume 4 Nomor 2 Tahun 2026

Leveraging Supervised Machine Learning to Improve Accuracy in Predicting Smartphone Addiction

Muhammad Nur Arafah[1], Imran Iskandar[2*], Rahmat Fuadi Syam[3], Fitri Rahmadani[4], Mario Dendo[5], Serpasius Dappa Sudda[6]

[1]Irmex Digital Akademika, Indonesia
[2,3,4,5,6] Program Studi Ilmu Komputer, Universitas Pancasakti, Indonesia
e-mail: imran.iskandar@unpacti.ac.id

Abstract

The transformation of *smartphones* into essential life companions triggers new challenges in the form of digital addiction that erodes productivity and cognitive focus. In contrast to previous studies that focused on screen duration, this study explored a more critical dimension, namely the degradation of productive hours. The research aims to improve the prediction of addiction using *supervised machine learning* by integrating productivity variables as a core diagnostic feature. Through a database of 7,500 respondents, a comparative analysis was carried out on the *Random Forest, K-Nearest Neighbors (KNN),* and *Naïve Bayes algorithms* in The results of the study proved that the productivity variable significantly increased the predictability of the model. *Random Forest* consistently outperformed other algorithms across scenarios, with the highest accuracy reaching 93.32% at a 70:30 data ratio, surpassing *KNN* (89.66%) and *Naïve Bayes* (87.96%). With a specificity of 92.70%, these findings reveal that workflow disruption has a higher differentiating power than conventional duration metrics. In conclusion, productivity interference is a key indicator in identifying digital behavioral disorders. In practical terms, this research provides guidance for developers and managers in designing precise early warning systems to restore daily efficiency amid the distractions of the digital world.

Keywords : *Classification, Machine Learning, Productivity, Smartphone Addiction*

INTRODUCTION

The phenomenon of smartphone use has transformed into an essential instrument that dominates modern life, but carries the risk of digital addiction that threatens mental health and cognitive function (Fischer-Grote et al., 2019; Sarker, 2021). The main problem in this study is the real degradation of productive hours due to compulsive device use, where many users fail to realize the transition from functional to pathological use (Lee & Kim, 2021; Samaha & Hawi, 2016). As a result, there is a decline in the quality of self-control that triggers persistent procrastination as well as major

obstacles in achieving professional goals (Chaudhury, 2018; Syakinah & Daulay, 2022). In line with the trend of digital transformation in various industrial sectors, the integration of smart technology is now the key to mapping user behavior (Jeffry et al., 2023; Sihite et al., 2025)

Although machine learning has been widely used for behavior detection, analysis of consumer transaction patterns (Yusril et al., 2025) and classification of social problems through interpretable approaches (Arafah, Afifah, et al., 2025), there is a significant research gap in the exploration of productivity variables as a determinant feature in addiction classification (Alosaimi et al., 2016; Bottaro et al., 2024). The majority of studies still focus on screen duration without considering interruptions to the workflow as the main parameter (Jabeen et al., 2023). In addition, the development of human resource competencies in understanding this technology is crucial for mitigating the negative impacts of digitalization (Usman et al., 2024). Therefore, this study aims to optimize the accuracy of predictions through the comparative approach of the Random Forest, K-Nearest Neighbors, and Naïve Bayes algorithms. By utilizing the ensemble learning method, this study is expected to be able to map productivity features with precision to restore productive time efficiency in the modern work environment (Arafah, Aziz, et al., 2025; He & Ar, 2026)

METHOD

This study used a secondary dataset of 7,500 digital behavioral records focusing on the Work/Study Hours and Academic Work Impact variables. The selection of this productivity variable refers to the findings of (AlSaif et al., 2023) which confirm the existence of a significant relationship between digital addiction and decreased academic performance and time management. Before entering the modeling stage, the data goes through a strict pre-processing process including data cleansing from missing values, categorical variable transformation using Label Encoding, and normalization of features through StandardScaler to ensure algorithm accuracy (Arafah, Aziz, et al., 2025). This normalization step is crucial to ensure features with large value ranges do not dominate model calculations, especially on distance-based algorithms.

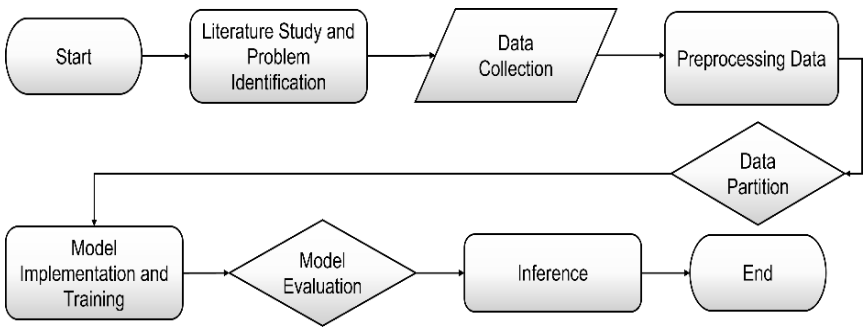


Figure 1. Research Stages

The classification architecture in this study compares three main algorithmic approaches, namely Random Forest (RF), K-Nearest Neighbors (KNN), and Naïve Bayes (NB). RF is applied as an ensemble method to minimize the risk of overfitting and handle high-dimensional data efficiently (Sarker, 2021). For comparison, KNN is used to classify digital behavior patterns based on the similarity of their attributes (Liu et al., 2023). In addition, Naïve Bayes is employed as a probabilistic classification method that assumes feature independence and is effective in handling large datasets with relatively low computational complexity, making it suitable for baseline comparison.

Data analysis was carried out by dividing the dataset using a ratio of 70% training data and 30% test data through stratified sampling techniques to maintain a balance of label distribution. The model's performance is comprehensively evaluated using the Confusion Matrix to generate the Accuracy, Precision, Recall, and F1-Score metrics. Furthermore, a Gini Importance-based Feature Importance analysis was carried out on the RF model to determine the most significant productivity variables influenced by smartphone addiction behavior, in accordance with the methodological framework developed by (Kibriya & et.al., 2022). All stages of this research are presented systematically in the flowchart in Figure 1.

RESULT AND DISCUSSION

RESULTS

Research on 7,500 respondents showed that the average daily screen duration reached 5.52 hours, while productive time was only 4.85 hours. The results of the hypothesis test confirmed that there was a significant negative correlation between the use of gadgets and productivity.

Table 1. Descriptive Statistics of Key Variables

Variable	Mean	Median	Std. Deviasi	Min	Max
Daily Screen Time	5,52	5,40	2,85	1,00	12,00
Work Study Hours	4,85	4,90	2,15	0,00	12,00
Social Media Hours	3,12	3,00	1,50	0,00	10,00
Sleep Hourse	6,21	6,00	1,10	3,00	10,00

Table 2 illustrates the distribution of data across nine systematic partitioning scenarios (S-01 to S-09). The distribution ranges from a 10% training proportion (750 samples) to 90% (6,750 samples). This procedure ensures model sensitivity is tested objectively, preventing data selection bias.

Table 2. Data Partition Sharing

Partitioning	Training	Test
10:90	750	6,750
20:80	1,500	6,000
30:70	2,250	5,250
40:60	3,000	4,500
50:50	3,750	3,750
60:40	4,500	3,000
70:30	**5,250**	**2,250**
80:20	6,000	1,500
90:10	6,750	750

Following the partitioning, three classification algorithms—Random Forest, K-Nearest Neighbors (KNN), and Naïve Bayes were implemented using the Scikit-Learn library. The technical specifications are detailed in Table 3.

Table 3. Technical Specifications and Training Parameters

Training Phase	Parameter
Data Input	*Scenarios S-01 to S-09*
Algoritma I	*Random Forest (100 Trees)*
Algoritma II	*K-Nearest Neighbors (Grid Search)*
Algoritma III	*Naïve Bayes (Gaussian Classifier)*
Fitur Importance	*Screen Time, Productivity, Stress*
Stopping Criteria	*Full Convergence*
Model Output	*Knowledge Base*

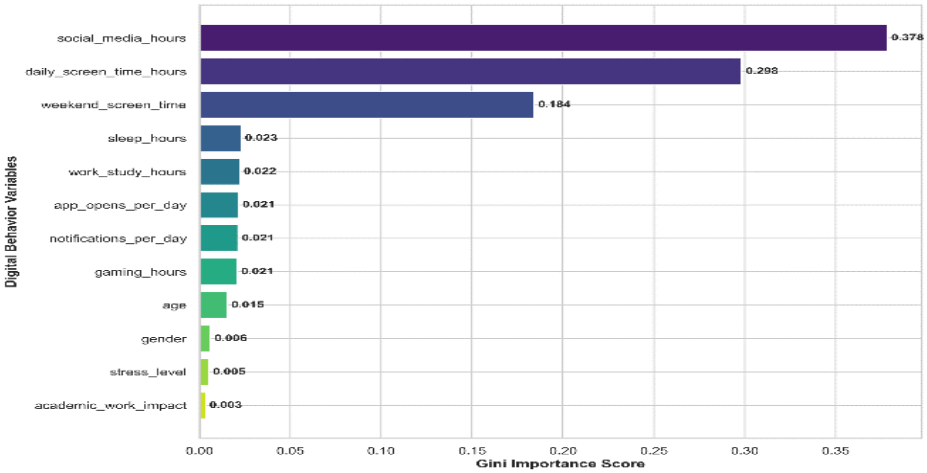


Figure 2. Gini Important Score

The Gini Importance Score was utilized to identify the weight of influential features, which then guided the classification across all scenarios. Table 4 compares the accuracy performance of the algorithms.

Table 4. Model Accuracy Comparison

Scenario	Ratio	Naïve Bayes	KNN	Random Forest
S-01	10:90	0.8853	0.8859	0.9305
S-02	20:80	0.8848	0.8902	0.9316
S-03	30:70	0.8886	0.8931	0.9329
S-04	40:60	0.8845	0.8948	0.9327
S-05	50:50	0.8827	0.8973	0.9317
S-06	60:40	0.8820	0.8954	0.9330
S-07	**70:30**	**0.8796**	**0.8966**	**0.9332**
S-08	80:20	0.8806	0.8961	0.9332
S-09	90:10	0.8815	0.8900	0.9277

Based on Table 4, Random Forest (RF) consistently achieves the highest performance with accuracy exceeding 93% across all scenarios. Figure 3 visualizes this dominance, showcasing RF's stability in mapping the correlation between device usage and productivity loss compared to Naïve Bayes and KNN. The peak accuracy of 93.32% was achieved in scenarios S-07 and S-08.

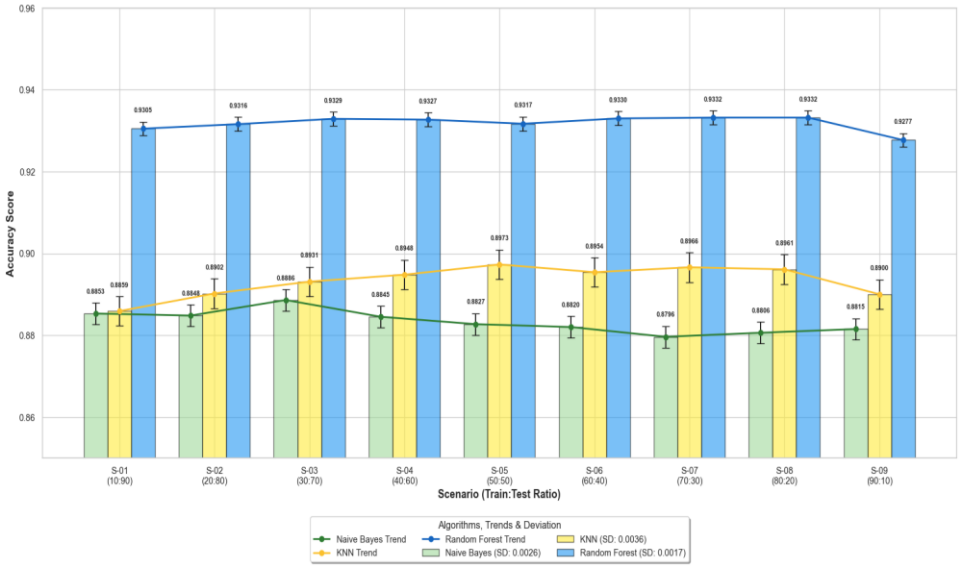


Figure 3. Visualization of Scenario Model Comparison

A detailed evaluation via the Confusion Matrix (Figure 4) for the best-performing scenario (S-07) highlights RF's precision. The model achieved a high specificity rate, accurately identifying non-addicted users in 610 out of 658 cases (92.70%).

Mathematical Calculation Confusion Matrix

- $Random\ Forest: \frac{(1.490 + 610)}{2250} = 93.32\%$ .....*(1)*
- $Naive\ Bayes: \frac{(1.431 + 548)}{2250} = 87{,}96\%$ .....*(2)*
- $KNN: \frac{(1.495 + 522)}{2.250} =$

$89{,}66\%$ ..... *(3)*

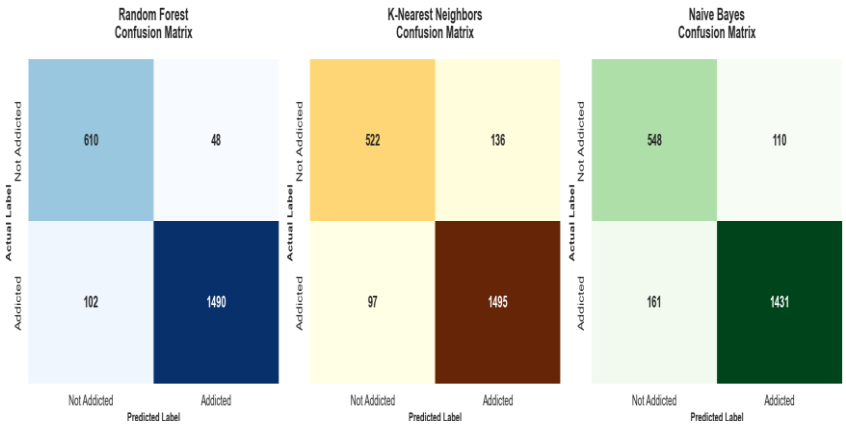


Figure 4. Confusion Matrix Adiction

The results indicate that the Random Forest model minimizes False Alarms while maintaining high sensitivity. The dominance of values along the main diagonal validates the 70:30 partition as a reliable configuration for classification in digital behavior monitoring.

DISCUSSION

The experimental findings reveal a profound relationship between digital behavior and individual performance, where productivity degradation serves as a more precise indicator of addiction than total screen time alone. The consistently superior performance of Random Forest (RF) across all scenarios, particularly with a peak accuracy of 93.32%, indicates that an ensemble-based decision tree approach is highly effective in capturing the non-linear thresholds of behavioral change. Unlike linear or distance-based models, RF's ability to process the interaction between "Work/Study Hours" and "Screen Time" allows for a more nuanced classification that reflects the reality of modern multitasking and digital-dependent work environments.

The high Gini Importance Score attributed to productivity-related features suggests a shift in the diagnostic paradigm for digital addiction. The data implies that the most critical factor in identifying pathological smartphone usage is not necessarily the frequency of interaction, but the extent to which that interaction interupts professional or academic flow. This is further validated by the model's 92.70% Specificity in the optimal S-07 (70:30) scenario. Such a high rate of correctly identifying non-addicted users demonstrates that the integration of productivity metrics effectively filters out "functional high-users" (individuals who use smartphones extensively for work) from those experiencing true addiction. This distinction is vital for ensuring that digital intervention systems are both accurate and credible in professional settings.

The performance stagnation of Naïve Bayes and KNN as training data increased highlights the limitations of their underlying assumptions in complex behavioral datasets. Naïve Bayes' failure to account for the interdependence of screen time and stress, along

with KNN's sensitivity to overlapping feature boundaries in high-density data, makes them less reliable for real-time monitoring. In contrast, the stability of Random Forest—even in minimal data ratios—proves that the "Forest" architecture effectively mitigates noise and provides a robust knowledge base. Ultimately, these results confirm that a classification model prioritizing productivity features provides a superior solution for mitigating performance degradation, offering a scientifically validated framework for restoring efficiency in the digital age.

CONCLUSION

This study concludes that leveraging supervised machine learning by integrating productivity variables significantly improves smartphone addiction prediction accuracy. Random Forest emerged as the superior classification solution with an accuracy of 93.32%, surpassing K-Nearest Neighbors (89.66%) and Naïve Bayes (87.96%). Reduced work and study efficiency were identified as predictors with the highest discriminative power compared to conventional screen duration. Practically, digital health developers and human resource managers should adopt productivity parameters in behavioral monitoring systems for more precise early interventions. Future research should explore productivity patterns across diverse occupational groups to reinforce the model's generalization on a broader scale.

REFERENCES

Alosaimi, F. D., Alyahya, H., Alshahwan, H., Al Mahyijari, N., & Shaik, S. A. (2016). Smartphone addiction among university students in Riyadh, Saudi Arabia. *Saudi Medical Journal*, *37*(6), 675–683. https://doi.org/10.15537/Smj.2016.6.14430

AlSaif, H. I., Alhozaimi, Z. A., Alrashed, A. S., Alanazi, K. S., Alshibani, M. G., Almigbal, T. H., Alsaad, S. M., Alrasheed, A. A., & Alosaimi, F. D. (2023). Is There an Association between Increased Stress and Smartphone Addiction? Insights from a Study on Medical Students from Saudi Arabia during the COVID-19 Pandemic. *Medicina (Kaunas, Lithuania)*, *59*(8). https://doi.org/10.3390/medicina59081501

Arafah, M. N., Afifah, N. A., Ejung, M. A., Berti, M. D., Handayani, R., Syam, R. F., & Iskandar, I. (2025). Explainable Machine Learning Approach to Classifying Poverty Levels Across Indonesia. *Integrated Journal of Artificial Intelligence and Data Science*, *1*(2), 57–66. https://journal.irmexdigika.com/index.php/ijaids/article/view/50

Arafah, M. N., Aziz, F., Usman, S., Syam, R. F., & Rimalia, W. (2025). *Klasifikasi Pengonsumsi Cokelat Menggunakan Algoritma Ensemble Bagging Untuk Meningkatkan Akurasi Prediksi Classification of Chocolate Consumers Using Ensemble Bagging Algorithm to Improve Prediction Accuracys*. *1*(1), 1–6. https://journal.irmexdigika.com/index.php/ijaids/article/view/23/11

Bottaro, R., Valenti, G. D., & Faraci, P. (2024). Internet Addiction and Psychological Distress: Can Social Networking Site Addiction Affect Body Uneasiness Across

Gender? A Mediation Model. *Europe's Journal of Psychology*, *20*(1), 41–62. https://doi.org/10.5964/ejop.10273

Chaudhury, P. (2018). A Study on impact of smartphone addiction on academic performance. *International Journal of Engineering & Technology*, *7*, 50. https://doi.org/10.14419/ijet.v7i2.6.10066

Fischer-Grote, L., Kothgassner, O. D., & Felnhofer, A. (2019). Risk factors for problematic smartphone use in children and adolescents: a review of existing literature. *Neuropsychiatrie : Klinik, Diagnostik, Therapie Und Rehabilitation : Organ Der Gesellschaft Osterreichischer Nervenarzte Und Psychiater*, *33*(4), 179–190. https://doi.org/10.1007/s40211-019-00319-8

Il, B. S. U. N., & Ar, K. U. M. (2026). *PREDICTING SMART PHONE ADDICTION*. *13*(2), 282–287.

Jabeen, F., Luukela-Tandon, A., Azad, N., Islam, A. K. M., & Pereira, V. (2023). The dark side of social media platforms: A situation-organism-behaviour-consequence approach. *Technological Forecasting and Social Change*, *186*, 122104. https://doi.org/10.1016/j.techfore.2022.122104

Jeffry, J., Nurdyansa, N., Usman, S., Sasmita, D., & Arafah, M. N. (2023). Transformasi Digital Dalam Industri Kerajinan Rumahan Melalui Penggunaan Teknologi Machine Learning. *JMM (Jurnal Masyarakat Mandiri)*, *7*(6), 5840–5849.

Kibriya, S., & others. (2022). The impact of smartphone addiction on academic performance: A study of university students. *Journal of Digital Learning*.

Lee, J., & Kim, W. (2021). Prediction of Problematic Smartphone Use: A Machine Learning Approach. *International Journal of Environmental Research and Public Health*, *18*(12). https://doi.org/10.3390/ijerph18126458

Liu, Q., Zhang, Y., & Chen, L. (2023). Machine learning applications in consumer behavior prediction. *Expert Systems with Applications*, *213*, 118926.

Samaha, M., & Hawi, N. S. (2016). Relationships among smartphone addiction, stress, academic performance, and satisfaction with life. *Computers in Human Behavior*, *57*, 321–325. https://doi.org/https://doi.org/10.1016/j.chb.2015.12.045

Sarker, I. H. (2021). Machine learning: Algorithms, real-world applications and research directions. *SN Computer Science*, *2*(3), 160.

Sihite, M., Suparwata, D. O., Arafah, M. N., & Uhai, S. (2025). SWOT Analysis of the Strategy of Digital Transformation of SMEs. *International Journal of Business, Law, and Education*, *6*(2), 1546–1555.

Syakinah, M., & Daulay, N. (2022). The Effect of Smartphone Addiction on Learning Motivation and Well-Being. *Jurnal Basicedu*, *6*, 5735–5743. https://doi.org/10.31004/basicedu.v6i4.3095

Usman, S., Nurdyansa, N., Aziz, F., Arafah, M. N., Wijaya, N. S., & Syam, R. F. (2024). Pengembangan Kompetensi Sdm Mitra Cv Sehat Kerupuk Melalui Pelatihan Machine Learning Dan Ilmu Komunikasi. *jurnal pengabdian kepada masyarakat TEKNO*, *4*(2), 1–8. https://journal.intelekmadani.org/index.php/globalabdimas/article/view/555/445

Yusril, M., Mahendra, K., Arafah, M. N., & Salzabilah, N. (2025). *Implementasi Machine Learning untuk Analisis Pola Perilaku Konsumen E-Commerce Berbasis Data Transaksi Machine Learning Implementation for Analyzing Consumer Behavior Patterns in E-Commerce Transactions*. *1*(2), 36–41. https://journal.irmexdigika.com/index.php/ijaids/article/view/47