

Predicting Smartphone Addiction Levels with K-Nearest Neighbors Using User Behavior Patterns

M. Rhifky Wayahdi^{*1}, Fahmi Ruziq²

^{1,2}Information Systems, Universitas Battuta, Indonesia

Email: ¹muhammadrhifkywayahdi@gmail.com

Received : Jun 18, 2025; Revised : Jul 16, 2025; Accepted : Aug 12, 2025; Published : Oct 16, 2025

Abstract

Smartphones have become an integral part of everyday life, but their ever-increasing popularity has raised growing global concerns about excessive use (nomophobia), which impacts quality of life, mental health, and academic performance. Existing research often relies on subjective questionnaires, limiting scalability and objectivity. This study addresses this gap by developing a machine learning model to predict smartphone addiction levels through an objective analysis of user behavior patterns. This research evaluates the effectiveness of the K-Nearest Neighbor (KNN) algorithm, identifies the most influential behavioral features, and assesses the model's classification performance. Using a dataset of 3,300 user behavior entries with 11 features, a waterfall-based framework was employed for data preprocessing, model design, and evaluation. The KNN model achieved 95% accuracy in classifying addiction levels. Permutation Feature Importance analysis confirmed 'App Usage Time' and 'Battery Drain' as the two most influential predictive features. This study demonstrates that KNN is a powerful and viable method for objectively classifying smartphone addiction. The findings provide a strong foundation for developing scalable, AI-driven early detection and intervention systems, offering significant contributions to the fields of computer science and digital well-being.

Keywords : *K-Nearest Neighbors, Machine Learning, Prediction, Smartphone Addiction, User Behavior Patterns.*

This work is an open access article and licensed under a Creative Commons Attribution-Non Commercial 4.0 International License



1. INTRODUCTION

Smartphones are technological devices that have a huge impact on life [1]. Smartphones play a role in enhancing human-technology interaction with various advantages. However, the growing popularity may lead to the problem of overuse [2]. Smartphone overuse (nomophobia) [1] has become a global problem [2], [3] that impacts lives. Studies show that uncontrolled smartphone use results in decreased quality of life [4], mental health disorders such as depression, anxiety, stress [1], [2], [5]-[7], reduced academic performance [1], [3], [6], [8], [9], disturbed sleep [2], [6], [9], [10], musculoskeletal problems, neurological changes [2], and reduced social interaction [1]. Factors such as loneliness, aggression [11], and the use of social-media and gaming as maladaptive coping strategies [8] exacerbate this condition. Women are reported to be more prone to smartphone addiction than men [1], [2]. This condition shows that smartphone addiction is a multidimensional issue that urgently needs to be studied in more depth with an objective and data-based approach.

Indonesia ranks 4th in the world in the number of smartphone users (187.7 million users and 68.1% penetration), indicating high digitalization. Digital advancement also brings the threat of addiction that damages physical and social health, yet research on the impact of smartphone addiction still has limitations. As shown in Figure 1 and Figure 2, Indonesia's large number of smartphone users underscores the national urgency of addressing this issue. Some studies rely on subjective questionnaires [12], while others use app usage data with limited generalizability [13] on specific samples, such as university students in Malaysia [8] or Serbia [6], as well as Android users only [7]. Objective data-

driven approaches to user behavior patterns analysis have not been widely explored to accurately predict addiction levels. Machine learning offers a potential solution through measurable analysis of smartphone usage patterns. This highlights a critical research gap: the need for an objective, scalable prediction model.

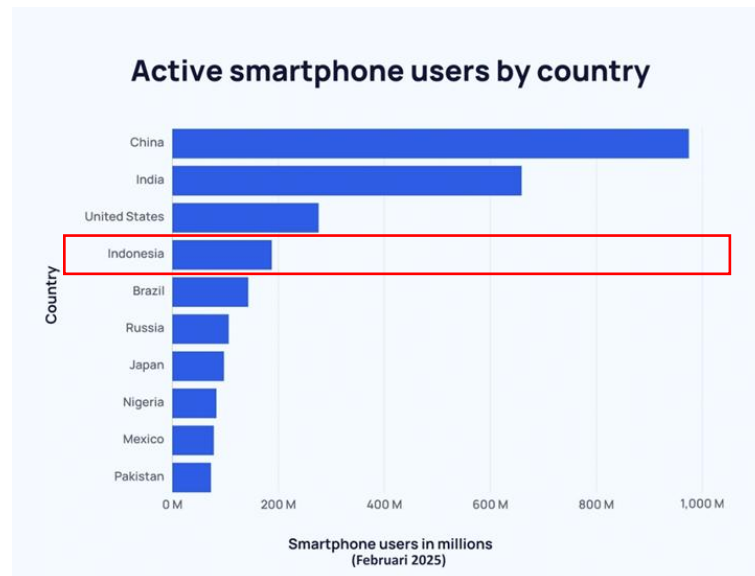


Figure 1. Smartphone User Demographics of Top 10 Countries

Sources: <https://explodingtopics.com/blog/smartphone-stats#number-of-smartphones>

Rank	Country	Smartphone Users	Total Population	Smartphone Penetration
1	China	974.69 million	1.43 billion	68.4%
2	India	659 million	1.42 billion	46.5%
3	United States	276.14 million	338.29 million	81.6%
4	Indonesia	187.7 million	275.5 million	68.1%
5	Brazil	143.43 million	215.31 million	66.6%
6	Russia	106.44 million	144.71 million	73.6%
7	Japan	97.44 million	123.95 million	78.6%
8	Nigeria	83.34 million	218.54 million	38.1%
9	Mexico	78.37 million	127.5 million	61.5%
10	Pakistan	72.99 million	235.82 million	31%

Figure 2. Number of Smartphone Users Top 10 Countries

Sources: <https://explodingtopics.com/blog/smartphone-stats#number-of-smartphones>

Machine learning is a subset of Artificial Intelligence (AI) [14], [15] that is capable of processing big data [16], identifying complex patterns [17], and accelerating decision-making [15]. It works similar to human learning [18], potentially revolutionizing industries with efficiency [19]. One of the popular machine learning algorithms is K-Nearest Neighbor (KNN) [20]. KNN is effective for both large [17] and small dimensional datasets [21], and has shown high accuracy in a variety of problem solving, such as lung cancer prediction [22], heart disease [23], [24], skin disease [21], biodiesel production [25],

social-media sentiment analysis [26], [27], cybercrime detection [28], and image classification [29]-[31]. This makes it suitable for modeling user behavior patterns in the context of smartphone addiction.

Previous studies on smartphone addiction tend to use questionnaire-based approaches that are subjective or experimental data with limited coverage in certain populations. These studies, such as those conducted by [1], [6], [8]-[10], [12], [13], generally utilize application usage data or user perceptions, but have not deeply applied machine learning-based predictive approaches that rely on actual behavioral data. On the other hand, machine learning algorithms such as Random Forest and Support Vector Machine (SVM) are more widely used in digital behavior analysis [7], [13], while the K-Nearest Neighbor (KNN) algorithm, which has proven effective in various fields such as disease detection, sentiment analysis, and image classification [21]-[31], is still rarely applied in the context of predicting the level of smartphone addiction. This research, therefore, offers a novel approach by applying and validating the KNN algorithm on objective behavioral data, contrasting with the prevailing subjective methods and exploring an underutilized but potentially powerful classification technique for this specific problem. In addition to methodological and technical contributions, this research also presents an open dataset for further validation and development. The resulting model has applicative potential in the form of integration into the system.

The formulation of this research problem is to what extent can the K-Nearest Neighbor (KNN) method predict the level of smartphone addiction based on the analysis of user behavior patterns, what factors most influence the prediction, and what are the weaknesses of the KNN method in classifying the level of addiction to smartphone use? This research is important to develop an objective prediction system of smartphone addiction level based on user behavior patterns and overcome the limitations of conventional questionnaire methods that are subjective. The findings can form the basis for the development of AI-based addiction early detection applications. The study also enriches the literature by validating the KNN algorithm for smartphone addiction classification using a data-driven approach, while identifying the behavioral features that most influence prediction. The results of this research are expected to produce a more objective and scalable model than previous studies, while providing a basis for the development of early detection or intervention systems for smartphone addiction.

2. METHOD

The research method used is a process framework approach based on the waterfall model [32], [33] which is adapted to the development flow of the machine learning model.

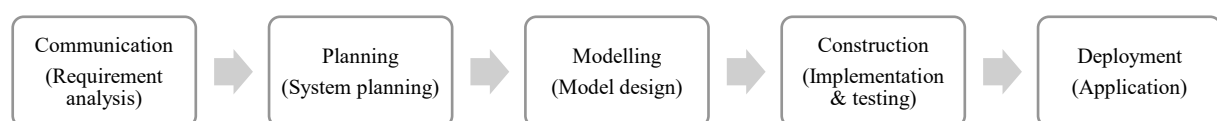


Figure 3. Research Flow

2.1. Communication (Requirement Analysis)

This initial stage focuses on formulating the objective of building a predictive model for smartphone addiction based on user behavior patterns, such as usage duration and screen on time. The research foundation is strengthened through a literature review covering machine learning, the K-Nearest Neighbor (KNN) method, and related studies to ensure the scientific validity of the approach.

2.2. Planning (System Planning)

In this stage, a systematic technical plan is developed. This plan includes the data collection strategy (using public datasets and tools like Python), the design of the data processing workflow

(preprocessing and feature selection), and the establishment of evaluation metrics such as Accuracy, Precision, and F1-Score to measure the model's success.

2.3. Modelling (Model Design)

The Model Design stage is the core of the technical implementation of the research. This process begins with the collection of the actual dataset containing smartphone user behavior data as planned. The amount of data used is 3,300 and the number of features is 11 ('User ID', 'Device Model', 'Operating System', 'App Usage Time (min/day)', 'Screen on Time (hours/day)', 'Battery Drain (mAh/day)', 'Number of Apps Installed', 'Data Usage (MB/day)', 'Age', 'Gender', 'User Behavior Class').

The raw data then goes through a series of rigorous pre-processing processes to ensure its quality and readiness. This includes data cleaning from missing values, transformation of categorical data into numerical format through one-hot encoding, feature selection using correlation analysis to remove redundancy, and data normalization using Min-Max or Z-Score methods so that all features have comparable scales.

Feature Selection (Pearson Correlation Analysis): To measure the linear relationship between two features (variables) and eliminate redundancy, the Pearson Correlation Coefficient (r) is used, as defined in Equation (1).

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} \quad (1)$$

Where:

- r_{xy} : The Pearson correlation coefficient between variables x and y . Its value ranges from -1 (perfect negative linear correlation) to +1 (perfect positive linear correlation). A value near 0 indicates no linear correlation.
- n : The total number of data samples
- x_i : The i -th value of variable x
- y_i : The i -th value of variable y
- $\bar{x}$: The mean value of variable x
- $\bar{y}$: The mean value of variable y

Data Normalization (Min-Max Normalization): This method rescales data to a specific range, typically between 0 and 1, calculated using the formula in Equation (2).

$$x'_{norm} = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (2)$$

Where:

- x'_{norm} : The new value after normalization
- x : The original data value
- x_{min} : The minimum value of all data points in that feature
- x_{max} : The maximum value of all data points in that feature

Z-Score Standardization, shown in Equation (3), transforms data to have a mean of 0 and a standard deviation of 1.

$$z = \frac{x - \mu}{\sigma} \quad (3)$$

Where:

- z : The new value after standardization (the Z-score)
- x : The original data value
- μ : The mean of all data points in that feature
- σ : The standard deviation of all data points in that feature

Both methods are considered standard for preparing data for distance-based algorithms like KNN. For this study, Z-Score standardization was ultimately applied to the features used in the decision boundary visualizations as it effectively handles outliers and centers the data around zero, which is ideal for visualization. Once the data is clean and ready, it is divided into two parts: 70-80% as training data and 20-30% as testing data.

Only then, the KNN model specifically designed to classify addiction levels on a scale of 1 to 5 is built, followed by parameter optimization such as determining the best ' k ' value and distance metric (Euclidean) to achieve maximum performance. The Euclidean distance, formulated in Equation (4), was chosen as it is the most common and intuitive metric for measuring distance in a multi-dimensional feature space, assuming that the space is flat and features are comparable after normalization.

$$d(p, q) = \sqrt{\sum_{i=1}^n (p_i - q_i)^2} \quad (4)$$

Where:

- $d(p, q)$: The Euclidean distance between data points p and q
- n : The number of features (dimensions) in the data
- p_i : The value of the i -th feature of data point p
- q_i : The value of the i -th feature of data point q

2.4. Construction (Implementation & Testing)

In the Implementation and Testing stage, the KNN model is trained on the data and validated using cross-validation to ensure reliability. Its performance is then quantitatively evaluated with metrics such as Accuracy, Precision, Recall, and F1-Score, whose formulas are defined in Equations (5), (6), (7), and (8), while the confusion matrix is analyzed to understand classification errors.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (5)$$

$$Precision = \frac{TP}{TP+FP} \quad (6)$$

$$Recall = \frac{TP}{TP+FN} \quad (7)$$

$$F1 - Score = 2 * \frac{Precision * Recall}{Precision + Recall} \quad (8)$$

The process concludes by interpreting these results to identify the most influential features and visualizing user data patterns for deeper insights.

2.5. Deployment (Application)

As the final stage of the research cycle, the Implementation and Publication phase focuses on dissemination of results and documentation. Finally, based on the findings and limitations of this research, a concrete recommendation for further development was formulated, namely the creation of a real-time monitoring application that can apply this model to provide practical benefits to users.

3. RESULTS

This section presents the empirical findings of the study in a systematic and objective manner. The data presented includes exploratory data analysis, evaluation of predictive model performance, classification error analysis, and identification of the most significant user behavior features in predicting the level of smartphone addiction.

3.1. Exploratory Data Analysis (EDA)

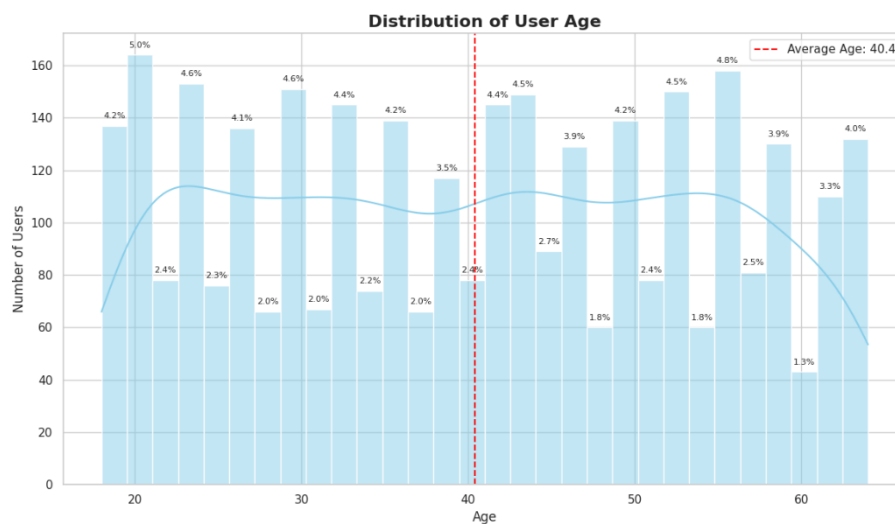


Figure 4. Distribution of User Age

Figure 4 shows the age distribution of smartphone users, which provides a demographic overview of the users whose data was analyzed, showing that users have a wide age range with an average age in their early 40s.

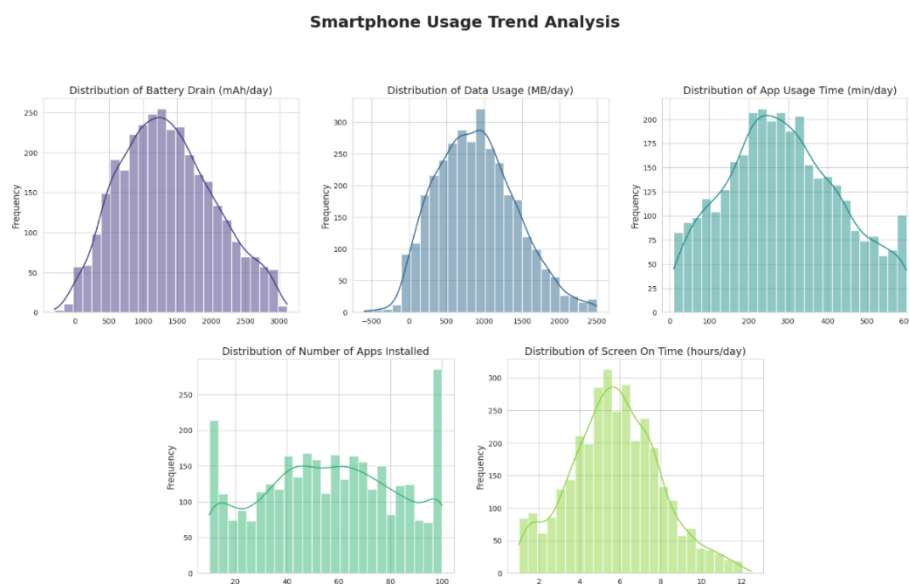


Figure 5. Smartphone Usage Trend Analysis

Figure 5 presents five histograms that summarize various patterns of user behavior in a single view. The graphs show that the majority of users consumed around 1000-1500 mAh of battery power and used 750-1000 MB of internet data per day. In terms of duration, the most common interaction time is around 250-300 minutes (4-5 hours) of app usage, with total screen on time peaking at 5-6 hours a

day. In addition, the data on the number of apps installed shows that there are different groups of users, ranging from those with few apps to those with a large number. Overall, this visualization provides a comprehensive insight into the smartphone usage habits and intensity among the analyzed users.

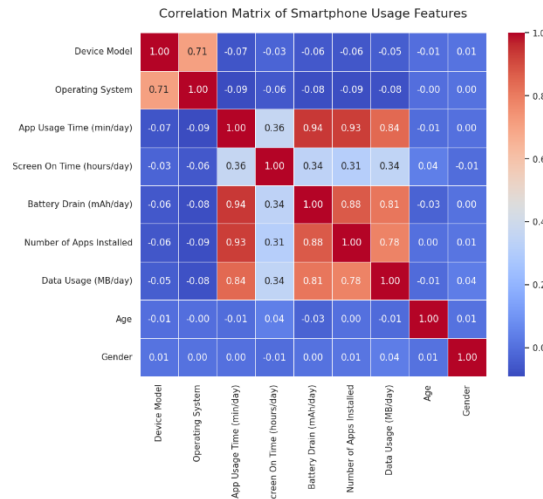


Figure 6. Correlation Matrix of Smartphone Usage Features

Figure 6 displays the correlation matrix which reveals that there is a very strong positive relationship between the core usage metrics, where App Usage Time shows a very high correlation with Battery Drain (with a value of 0.94) and Number of Installed Apps (value of 0.93). Similarly, Battery Drain is strongly related to Number of Apps Installed (0.88) and Data Usage (0.81), indicating that these features are highly interdependent. In contrast, demographic factors such as Age and Gender show almost no relationship at all with usage patterns, with correlation values very close to zero (ranging from -0.03 to 0.04), effectively proving that these factors are insignificant in determining smartphone usage behavior on this dataset.

3.2. Model Building and Performance Evaluation

Following the EDA, the KNN model was trained and tested. The performance was evaluated using a confusion matrix and a classification report. Figure 7 presents the confusion matrix, which provides a visual representation of the model's accuracy across the five addiction level classes. The diagonal values (e.g., 102 for Class 1, 447 for Class 3) show the number of correct predictions, while off-diagonal values represent misclassifications.

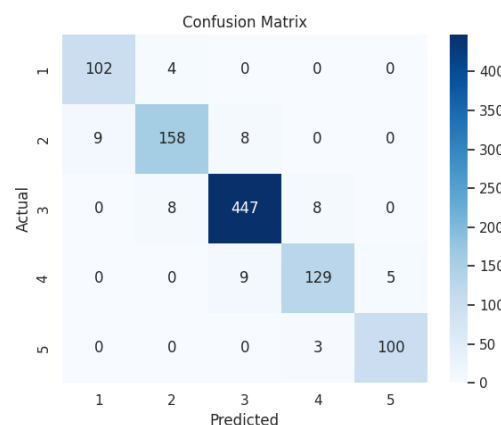


Figure 7. Confusion Matrix

Figure 7 is a strong visual representation of the performance of the classification model with 5 classes (addiction levels). In brief, the model is quite accurate, with major errors occurring when distinguishing between similar or adjacent categories, but it is very reliable in distinguishing very different categories. The detailed performance metrics are summarized in Table 1.

Table 1. Classification Report

	Precision	Recall	F1-Score	Support (Actual Occurrences)
1	0.92	0.96	0.94	106
2	0.93	0.90	0.92	175
3	0.96	0.97	0.96	463
4	0.92	0.90	0.91	143
5	0.95	0.97	0.96	103
Accuracy			0.95	990
Macro avg	0.94	0.94	0.94	990
Weight avg	0.95	0.95	0.95	990

The table confirms the model's high performance, achieving an overall accuracy of 95%. For individual classes, key metrics like Precision, Recall, and F1-Score are consistently high, such as the F1-Score of 0.96 for both Class 3 and Class 5, indicating a balanced and reliable model.

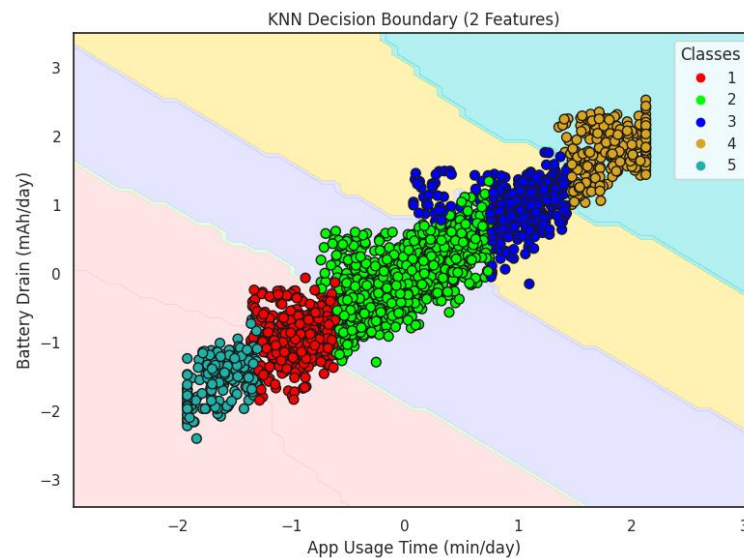


Figure 8. KNN Decision Boundary (2 Features)

Figure 8 explains the relationship between the features, where there is a clear positive correlation between App Usage Time and Battery Drain. The longer the app is used, the higher the battery drain. This can be seen from the data distribution which tends to move from the bottom left to the top right. The KNN model creates decision boundaries that are non-linear and tend to “follow” the shape of the data distribution of each class. The model essentially classifies new data points based on their closest “neighbors” in the training data. This figure effectively shows how the KNN model partitions the feature space into regions corresponding to each class. This helps us understand how the model will make predictions for data that has never been seen before.

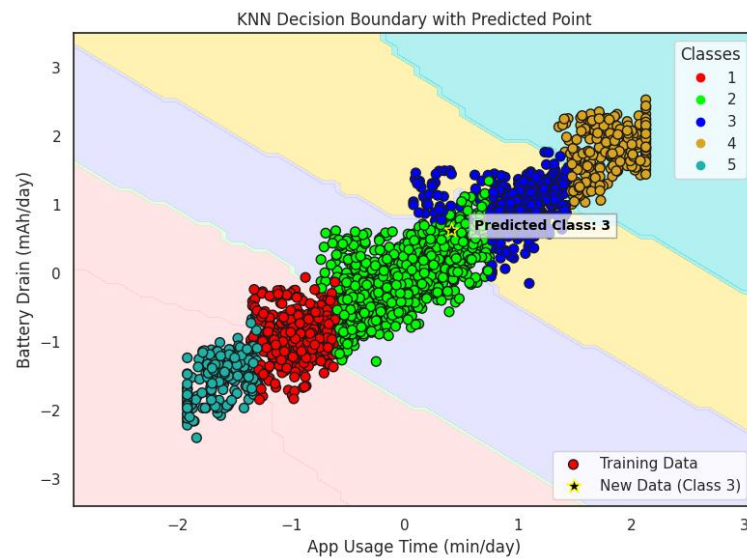


Figure 9. KNN Decision Boundary with Predicted Point

Figure 9 shows a real-life example of the KNN model successfully classifying a new piece of data as Class 3 based on the patterns it has learned from the training data. Figure 10 provides important context for the 2D visualization images we saw earlier. Now we know why App Usage Time and Battery Drain were chosen to be visualized: because they are the two most influential features in determining the model classification. The other two features have a smaller contribution.

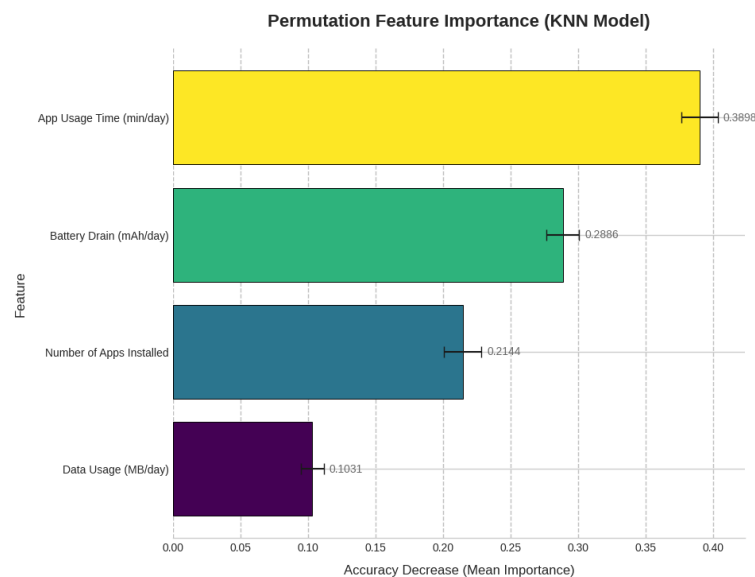


Figure 10. Permutation Feature Importance

Based on the set of analyses presented, a K-Nearest Neighbors (KNN) classification model has been successfully built to classify the data into five classes. The decision boundary visualization shows how the model separates the data and predicts new data points based on its two main features, namely App Usage Time and Battery Drain. Permutation Feature Importance analysis confirmed that these two features were indeed the most influential on the model's predictions. The performance of the model also proves to be very good, as shown by the Confusion Matrix, which displays a high level of accuracy with minimal prediction errors that generally only occur between neighboring classes.

4. DISCUSSION

Based on the results obtained from this research, it can be said that the K-Nearest Neighbor (KNN) model built shows good performance in predicting the level of smartphone addiction based on user behavior patterns, with overall accuracy reaching 95%. This is supported by Precision, Recall, and F1-Score values that are also high in most classes, indicating the model's strong ability in the classification process. These findings highlight the great potential of machine learning-based approaches in providing an objective and data-driven smartphone addiction prediction system, overcoming the limitations of conventional questionnaire methods that tend to be subjective. This is particularly urgent in an era of increasing digitalization, where automated and scalable tools are needed for early mental health intervention. The success of this model demonstrates a significant step toward moving beyond subjective self-reports.

Permutation Feature Importance analysis clearly identified 'App Usage Time (min/day)' and 'Battery Drain (mAh/day)' as the most influential features in determining model predictions, which intuitively makes sense as these two metrics directly reflect the intensity of smartphone usage. Compared to previous studies that tend to rely on subjective questionnaires or app usage data with limited generalization to specific samples, this study offers a more objective and data-driven approach by utilizing actual behavioral data and the KNN algorithm. Although other algorithms such as Random Forest and SVM are more commonly used in digital behavior analysis, the effectiveness of KNN in fields such as disease detection and sentiment analysis has proven to be relevant in this context of smartphone addiction prediction. The high performance of the KNN model can be attributed to its non-parametric nature. Unlike linear models, KNN does not make strong assumptions about the underlying data distribution. As shown by the non-linear decision boundaries in the visualization, the KNN model effectively partitions the feature space to classify addiction levels by capturing the complex, local relationships within the user behavior data, making it well-suited for this classification task.

However, it is important to acknowledge the limitations of this study and the KNN method. The KNN algorithm can be computationally expensive during prediction for very large datasets, as it needs to calculate distances to all training points. It is also sensitive to the "curse of dimensionality," where its performance can degrade as the number of features increases. While this study used a manageable number of highly relevant features, future work with more extensive feature sets might need to consider feature reduction techniques more aggressively. The model's performance also relies heavily on proper feature scaling, a step that was addressed here through normalization but remains a critical consideration for implementation. Despite these limitations, the model generated in this study is expected to be more objective and scalable than previous studies, providing a solid foundation for the development of future AI-based smartphone addiction early detection applications or intervention systems. This research provides a strong proof-of-concept for real-world applications. The resulting model could be integrated into a "digital wellbeing" mobile application that monitors user behavior in real-time (with user consent) and provides personalized feedback or alerts when patterns indicative of addictive behavior are detected.

5. CONCLUSION

The conclusion of this research is that the K-Nearest Neighbor (KNN) method, combined with an objective analysis of user behavior patterns, is a highly effective approach to predicting smartphone addiction levels. With model accuracy reaching 95%, this study successfully demonstrated the KNN's ability to accurately classify addiction levels, with 'App Usage Time (min/day)' and 'Battery Drain (mAh/day)' identified as the most influential features. This data-driven approach overcomes the limitations of conventional methods that are often subjective and have limited generalizability.

These findings make significant methodological and technical contributions to computer science by validating the KNN algorithm for smartphone addiction classification and providing an open dataset

for further research. More importantly, this work paves the way for the development of more objective and scalable AI-based early detection systems. Future research should focus on several key areas: (1) conducting a comparative analysis by benchmarking the KNN model against other advanced algorithms, such as Deep Neural Networks, to explore potential performance gains; (2) developing and deploying a real-time monitoring application to validate the model's practical utility; and (3) expanding the feature set to include more contextual data for even greater predictive accuracy. Ultimately, this research provides a robust foundation for building practical tools that can provide tangible benefits to users and promote digital well-being.

REFERENCES

- [1] C. Osorio-Molina, M. B. Martos-Cabrera, M. J. Membrive-Jiménez, K. Vargas-Roman, N. Suleiman-Martos, E. Ortega-Campos, and J. L. Gómez-Urquiza, "Smartphone addiction, risk factors and its adverse effects in nursing students: A systematic review and meta-analysis," *Nurse Education Today*, vol. 98, p. 104741, Mar. 2021. <https://doi.org/10.1016/j.nedt.2020.104741>
- [2] Z. A. Ratan, A. M. Parrish, S. B. Zaman, M. S. Alotaibi, and H. Hosseinzadeh, "Smartphone addiction and associated health outcomes in adult populations: a systematic review," *International Journal of Environmental Research and Public Health*, vol. 18, no. 22, p. 12257, Nov. 2021. <https://doi.org/10.3390/ijerph182212257>
- [3] O. J. Sunday, O. O. Adesope, and P. L. Maarhuis, "The effects of smartphone addiction on learning: A meta-analysis," *Computers in Human Behavior Reports*, vol. 4, p. 100114, Aug. 2021. <https://doi.org/10.1016/j.chbr.2021.100114>
- [4] A. Sela, N. Rozenboim, and H. C. Ben-Gal, "Smartphone use behavior and quality of life: What is the role of awareness?," *PloS One*, vol. 17, no. 3, p. e0260637, Mar. 2022. <https://doi.org/10.1371/journal.pone.0260637>
- [5] A. L. Sarhan, "The relationship of smartphone addiction with depression, anxiety, and stress among medical students," *SAGE Open Medicine*, vol. 12, p. 20503121241227367, Feb. 2024. <https://doi.org/10.1177/20503121241227367>
- [6] A. Nikolic, B. Bukurov, I. Kocic, M. Vukovic, N. Ladjovic, M. Vrhovac, Z. Pavlović, J. Grujicic, D. Kisic, and S. Sipetic, "Smartphone addiction, sleep quality, depression, anxiety, and stress among medical students," *Frontiers in Public Health*, vol. 11, p. 1252371, Sep. 2023. <https://doi.org/10.3389/fpubh.2023.1252371>
- [7] K. Opoku Asare, Y. Terhorst, J. Vega, E. Peltonen, E. Lagerspetz, and D. Ferreira, "Predicting depression from smartphone behavioral markers using machine learning methods, hyperparameter optimization, and feature importance analysis: exploratory study," *JMIR mHealth and uHealth*, vol. 9, no. 7, p. e26540, Jul. 2021. <https://doi.org/10.2196/26540>
- [8] G. A. Abbasi, M. Jagaveeran, Y. N. Goh, and B. Tariq, "The impact of type of content use on smartphone addiction and academic performance: Physical activity as moderator," *Technology in Society*, vol. 64, p. 101521, Feb. 2021. <https://doi.org/10.1016/j.techsoc.2020.101521>
- [9] B. Rathakrishnan, S. S. Bikar Singh, M. R. Kamaluddin, A. Yahaya, M. A. Mohd Nasir, F. Ibrahim, and Z. Ab Rahman, "Smartphone addiction and sleep quality on academic performance of university students: An exploratory research," *International Journal of Environmental Research and Public Health*, vol. 18, no. 16, p. 8291, Aug. 2021. <https://doi.org/10.3390/ijerph18168291>
- [10] S. Y. Sohn, L. Krasnoff, P. Rees, N. J. Kalk, and B. Carter, "The association between smartphone addiction and sleep: a UK cross-sectional study of young adults," *Frontiers in Psychiatry*, vol. 12, p. 629407, Mar. 2021. <https://doi.org/10.3389/fpsy.2021.629407>
- [11] F. G. Karaoglan Yilmaz, U. Avci, and R. Yilmaz, "The role of loneliness and aggression on smartphone addiction among university students," *Current Psychology*, vol. 42, no. 21, pp. 17909-17917, Jul. 2023. <https://doi.org/10.1007/s12144-022-03018-w>
- [12] R. J. James, G. Dixon, M. G. Dragomir, E. Thirlwell, and L. Hitcham, "Understanding the construction of 'behavior' in smartphone addiction: A scoping review," *Addictive Behaviors*, vol. 137, p. 107503, Feb. 2023. <https://doi.org/10.1016/j.addbeh.2022.107503>
- [13] T. Li, T. Xia, H. Wang, Z. Tu, S. Tarkoma, Z. Han, and P. Hui, "Smartphone app usage analysis:

- datasets, methods, and applications," *IEEE Communications Surveys & Tutorials*, vol. 24, no. 2, pp. 937-966, Mar. 2022. <https://doi.org/10.1109/COMST.2022.3163176>
- [14] M. R. Wayahdi, F. Ruziq, and S. H. N. Ginting, "AI approach to predict student performance (Case study: Battuta University)," *Journal of Science and Social Research*, vol. 7, no. 4, pp. 1800-1807, Nov. 2024. <https://doi.org/10.54314/jssr.v7i4.2332>
- [15] V. Galaz, M. A. Centeno, P. W. Callahan, A. Causevic, T. Patterson, I. Brass *et al.*, "Artificial intelligence, systemic risks, and sustainability," *Technology in Society*, vol. 67, p. 101741, Nov. 2021. <https://doi.org/10.1016/j.techsoc.2021.101741>
- [16] M. R. Wayahdi and M. Zaki, "The Role of AI in Diagnosing Student Learning Needs: Solutions for More Inclusive Education," *International Journal of Educational Insights and Innovations*, vol. 2, no. 1, pp. 1-7, Mar. 2025. <https://ijedins.technolabs.co.id/index.php/ijedins/article/view/6>
- [17] R. K. Halder, M. N. Uddin, M. A. Uddin, S. Aryal, and A. Khraisat, "Enhancing K-nearest neighbor algorithm: a comprehensive review and performance analysis of modifications," *Journal of Big Data*, vol. 11, no. 1, p. 113, Aug. 2024. <https://doi.org/10.1186/s40537-024-00973-y>
- [18] M. M. Taye, "Understanding of machine learning with deep learning: architectures, workflow, applications and future directions," *Computers*, vol. 12, no. 5, p. 91, Apr. 2023. <https://doi.org/10.3390/computers12050091>
- [19] K. Sharifani and M. Amini, "Machine learning and deep learning: A review of methods and applications," *World Information Technology and Engineering Journal*, vol. 10, no. 07, pp. 3897-3904, 2023. Available at SSRN: <https://ssrn.com/abstract=4458723>
- [20] S. Ramadhani and M. R. Wayahdi, "K-Nearest Neighbor and Random Forest Algorithms in Loan Approval Prediction," *Jurnal Minfo Polgan*, vol. 13, no. 1, pp. 1307-1313, Dec. 2024. <https://doi.org/10.33395/jmp.v13i1.14345>
- [21] M. A. Araaf, K. Nugroho, and D. R. Setiadi, "Comprehensive analysis and classification of skin diseases based on image texture features using K-nearest neighbors algorithm," *Journal of Computing Theories and Applications*, vol. 1, no. 1, pp. 31-40, Sep. 2023. <https://doi.org/10.33633/jcta.v1i1.9185>
- [22] M. R. Wayahdi and F. Ruziq, "KNN and XGBoost Algorithms for Lung Cancer Prediction," *Journal of Science Technology (JoSTec)*, vol. 4, no. 1, Jan. 2022. <https://doi.org/10.55299/jostec.v4i1.251>
- [23] M. A. Khan, T. Mazhar, M. M. Yaqoob, M. B. Khan, A. K. J. Saudagar, Y. Y. Ghadi *et al.*, "Optimal feature selection for heart disease prediction using modified Artificial Bee colony (M-ABC) and K-nearest neighbors (KNN)," *Scientific Reports*, vol. 14, no. 1, p. 26241, Oct. 2024. <https://doi.org/10.1038/s41598-024-78021-1>
- [24] A. D. R. Wibisono, S. Hidayat, H. M. T. Ramadhan, and E. Y. Puspaningrum, "Comparison of K-nearest neighbor and decision tree methods using principal component analysis technique in heart disease classification," *Indonesian Journal of Data and Science*, vol. 4, no. 2, pp. 87-96, Jul. 2023. <https://doi.org/10.56705/ijodas.v4i2.70>
- [25] A. Sumayli, "Development of advanced machine learning models for optimization of methyl ester biofuel production from papaya oil: Gaussian process regression (GPR), multilayer perceptron (MLP), and K-nearest neighbor (KNN) regression models," *Arabian Journal of Chemistry*, vol. 16, no. 7, p. 104833, Jul. 2023. <https://doi.org/10.1016/j.arabjc.2023.104833>
- [26] M. I. Hutapea and A. P. Silalahi, "Moderna's Vaccine Using the K-Nearest Neighbor (KNN) Method: An Analysis of Community Sentiment on Twitter," *Jurnal Penelitian Pendidikan IPA*, vol. 9, no. 5, pp. 3808-3814, May 2023. <https://doi.org/10.29303/jppipa.v9i5.3203>
- [27] S. Masturoh, R. L. Pratiwi, M. R. Saelan, and U. Radiyah, "Application of the k-nearest neighbor (KNN) algorithm in sentiment analysis of the Ovo e-wallet application," *JITK (Jurnal Ilmu Pengetahuan dan Teknologi Komputer)*, vol. 8, no. 2, pp. 84-89, Jan. 2023. <https://doi.org/10.33480/jitk.v8i2.3997>
- [28] D. M. Cao, M. A. Sayed, M. T. Islam, M. T. Mia, E. H. Ayon, B. P. Ghosh *et al.*, "Advanced cybercrime detection: A comprehensive study on supervised and unsupervised machine learning approaches using real-world datasets," *Journal of Computer Science and Technology Studies*, vol. 6, no. 1, pp. 40-48, Jan. 2024. <https://doi.org/10.32996/jcsts.2024.6.1.5>

-
- [29] M. R. Wayahdi, D. Syahputra, and S. H. N. Ginting, "Evaluation of the K-Nearest Neighbor Model With K-Fold Cross Validation on Image Classification," *Infokum*, vol. 9, no. 1, pp. 1-6, Dec. 2020. <http://seaninstitute.org/infor/index.php/infokum/article/view/72>
- [30] M. Jagdish, A. M. Guzman, G. F. Sancho, and A. Guerrero-Luzuriaga, "Detection and classification of caterpillar using image processing with K-nearest neighbor classification technique," *Turkish Journal of Computer and Mathematics Education*, vol. 12, no. 5, pp. 719-728, 2021. <https://turcomat.org/index.php/turkbilmat/article/view/1475>
- [31] S. Anraeni, D. Indra, D. Adirahmadi, S. Pomalingo, and S. H. Mansyur, "Strawberry ripeness identification using feature extraction of RGB and K-nearest neighbor," in *2021 3rd East Indonesia Conference on Computer and Information Technology (EIConCIT)*, 2021, pp. 395-398. <https://doi.org/10.1109/EIConCIT50028.2021.9431854>
- [32] M. R. Wayahdi and F. Ruziq, "Designing an Used Goods Donation System to Reduce Waste Accumulation Using the WASPAS Method," *Sinkron: Jurnal dan Penelitian Teknik Informatika*, vol. 8, no. 4, pp. 2325-2334, Oct. 2024. <https://doi.org/10.33395/sinkron.v8i4.14115>
- [33] M. R. Wayahdi and S. Guntur, "Website-Based Village Information System Design (Case Study: Ujung Batu III Village)," *Jurnal Minfo Polgan*, vol. 14, no. 1, pp. 38-44, Mar. 2025. <https://doi.org/10.33395/jmp.v14i1.14621>