Behavioral Segmentation of Smartphone Users in Digital Society Using Unsupervised Clustering Techniques for Digital Engagement Analysis

Mesra betty Yel[1,*], Rodhiyah[2]

[1,2]Information System, Sekolah Tinggi Ilmu Komputer Cipta Karya Informatika, Jakarta Timur,13440, Indonesia

ABSTRACT

The rapid growth of smartphone usage has significantly influenced digital behavior in modern societies. Understanding how individuals interact with smartphones is important for identifying patterns of digital engagement and potential risks associated with excessive usage. This study aims to discover smartphone usage behavior patterns in a digital society using clustering algorithms. A dataset containing 7,500 smartphone users with multiple behavioral variables was analyzed. An automated feature subset search strategy was applied to identify the most informative behavioral indicators, while clustering performance was evaluated using Silhouette Score and Davies–Bouldin Index. The optimal clustering configuration was obtained using the Birch algorithm with two clusters, supported by Principal Component Analysis retaining 85 percent of the dataset variance. The clustering model identified three key behavioral indicators that differentiate smartphone users, namely daily screen time, the proportion of social media usage relative to total screen time, and notification frequency per hour. The results revealed two distinct behavioral groups consisting of moderate engagement users and high digital engagement users. The high engagement cluster demonstrated significantly longer screen time, increased weekend smartphone usage, and a higher proportion of users labeled as addicted. These findings indicate that higher digital engagement intensity is strongly associated with smartphone addiction behavior. The study demonstrates the effectiveness of unsupervised machine learning in uncovering behavioral structures in digital environments and provides insights that may support future research on digital behavior, smartphone addiction, and digital well-being in contemporary digital societies.

Keywords Smartphone Usage Behavior, Digital Engagement, Clustering Algorithms, Smartphone Addiction, Digital Society

Submitted 15 February 2026
Accepted 20 Mei 2026
Published 01 June 2026

Corresponding author
Mesra betty Yel,
mesrabettyyel@stikomcki.ac.id

Additional Information and Declarations can be found on page 127

INTRODUCTION

The rapid development of digital technologies has transformed the way individuals interact with information, communication platforms, and social environments. Among these technologies, smartphones have become one of the most widely used devices in modern societies. Smartphones enable continuous access to digital services such as social media, messaging platforms, online entertainment, and information resources. As a result, smartphones play a central role in shaping everyday digital behavior and have become an essential component of digital society [1]. The increasing integration of smartphones into daily life has significantly influenced communication patterns, social interactions, and individual lifestyles.

Despite the many benefits of smartphone technology, the increasing intensity

How to cite this article: M. B. Yel and Rodhiyah, "Behavioral Segmentation of Smartphone Users in Digital Society Using Unsupervised Clustering Techniques for Digital Engagement Analysis," *J. Digit. Soc.*, vol. 2, no. 2, pp. 112-128, 2026.

of smartphone usage has also raised concerns regarding excessive digital engagement. Prolonged smartphone use has been associated with various behavioral changes such as increased screen time, frequent checking of notifications, and continuous interaction with digital applications [2]. These behaviors may lead to problematic smartphone usage or smartphone addiction, which can negatively affect productivity, sleep quality, and psychological well-being. Understanding how individuals use smartphones and identifying patterns of digital engagement have therefore become important topics in research related to digital behavior and technology use.

Previous studies have attempted to analyze smartphone usage behavior using different analytical approaches. Many studies rely on traditional statistical techniques such as correlation analysis and regression models to examine the relationship between smartphone usage and psychological outcomes. Other studies focus on measuring smartphone addiction through surveys or self-reported behavioral indicators [3]. While these approaches provide valuable insights into the consequences of smartphone use, they often rely on predefined categories of user behavior. Consequently, such approaches may fail to capture hidden patterns of smartphone interaction that naturally emerge from large scale behavioral data.

Recent advances in machine learning have provided new opportunities to analyze digital behavior using data driven methods. In particular, unsupervised learning techniques such as clustering algorithms allow researchers to discover latent behavioral structures without relying on predefined labels. These methods can identify groups of smartphone users who share similar behavioral characteristics based on multiple usage indicators such as screen time, social media interaction, and notification frequency [4]. Several recent studies have applied clustering techniques to segment smartphone users into behavioral groups, demonstrating the potential of machine learning approaches for understanding digital engagement patterns in smartphone environments.

However, several research gaps remain in the existing literature. Many previous studies focus primarily on predicting smartphone addiction using supervised learning models rather than exploring behavioral patterns using unsupervised approaches. In addition, existing studies often use limited behavioral indicators or relatively small datasets, which may restrict the ability to capture the complexity of smartphone usage behavior in digital society. Therefore, this study aims to discover smartphone usage behavior patterns using clustering algorithms applied to large scale smartphone interaction data. By identifying distinct behavioral groups and examining their relationship with smartphone addiction tendencies, this study contributes to a deeper understanding of digital engagement patterns and provides insights into smartphone usage behavior in contemporary digital societies [5].

Literature Review and Related Works

The increasing integration of smartphones into everyday life has significantly influenced digital behavior in modern societies. Smartphones provide constant connectivity to digital services such as social media platforms, communication tools, and online information resources. As a result, smartphones have become central to daily digital interaction and play an important role in shaping communication patterns and information consumption. Previous studies have shown that the widespread use of smartphones has transformed how individuals interact with digital environments and manage daily activities [6], [7]. The rapid growth of smartphone usage has also raised concerns regarding excessive digital engagement and its potential impact on user well-being [8].

A growing body of research has investigated the relationship between smartphone usage patterns and behavioral outcomes. Several studies have reported that prolonged smartphone usage is associated with increased screen time, frequent interaction with notifications, and repeated engagement with digital platforms [9], [10]. These behavioral patterns are often linked to problematic smartphone usage or smartphone addiction. Smartphone addiction is generally characterized by excessive smartphone use, reduced self-control over device interaction, and a strong psychological dependence on digital technologies [11], [12]. Such behaviors may negatively affect productivity, sleep quality, and mental health, highlighting the importance of understanding smartphone usage patterns in digital societies [13].

To analyze smartphone usage behavior, previous studies have applied a variety of analytical approaches. Traditional research in this area often relies on statistical methods such as regression analysis, correlation analysis, and survey-based assessments to explore the relationship between smartphone usage and psychological outcomes [14]. While these approaches provide valuable insights, they are typically based on predefined behavioral categories and theoretical assumptions. As a result, these methods may not fully capture the complex and heterogeneous patterns of smartphone interaction that emerge in large scale behavioral datasets.

Recent advances in machine learning have provided new opportunities for analyzing digital behavior in a data driven manner. In particular, unsupervised learning techniques such as clustering algorithms have been widely used to identify hidden structures in behavioral data. Clustering methods allow researchers to group users based on similarities in their interaction patterns without relying on predefined labels [15], [16]. These techniques have been successfully applied in various domains such as user segmentation, behavioral analytics, and digital engagement analysis.

Several studies have applied clustering algorithms to analyze smartphone usage behavior and identify distinct user groups. These studies demonstrate that smartphone users can often be segmented into different behavioral categories such as high engagement users, social media-oriented users, and moderate usage users [17]. Other research has shown that clustering techniques can reveal meaningful relationships between smartphone usage intensity and behavioral outcomes such as digital engagement and technology dependency [18]. In addition, clustering methods have been used to explore patterns of smartphone interaction based on indicators such as screen time, application usage frequency, and notification activity [19].

Despite these developments, several research gaps remain in the existing literature. Many previous studies focus primarily on predicting smartphone addiction using supervised learning models rather than exploring behavioral patterns using unsupervised approaches [20]. Furthermore, some studies rely on relatively limited behavioral indicators or small-scale datasets, which may limit the ability to capture the complexity of smartphone usage behavior in digital societies. Therefore, there is still a need for research that applies clustering algorithms to large scale smartphone usage datasets in order to discover latent behavioral structures and better understand patterns of digital engagement among smartphone users.

Methodology

This study applies an unsupervised machine learning approach to discover smartphone usage behavior patterns. The methodology consists of several stages including data preprocessing, feature engineering, feature subset selection, dimensionality reduction, clustering, and cluster evaluation. The overall workflow of the proposed methodology is illustrated in figure 1.

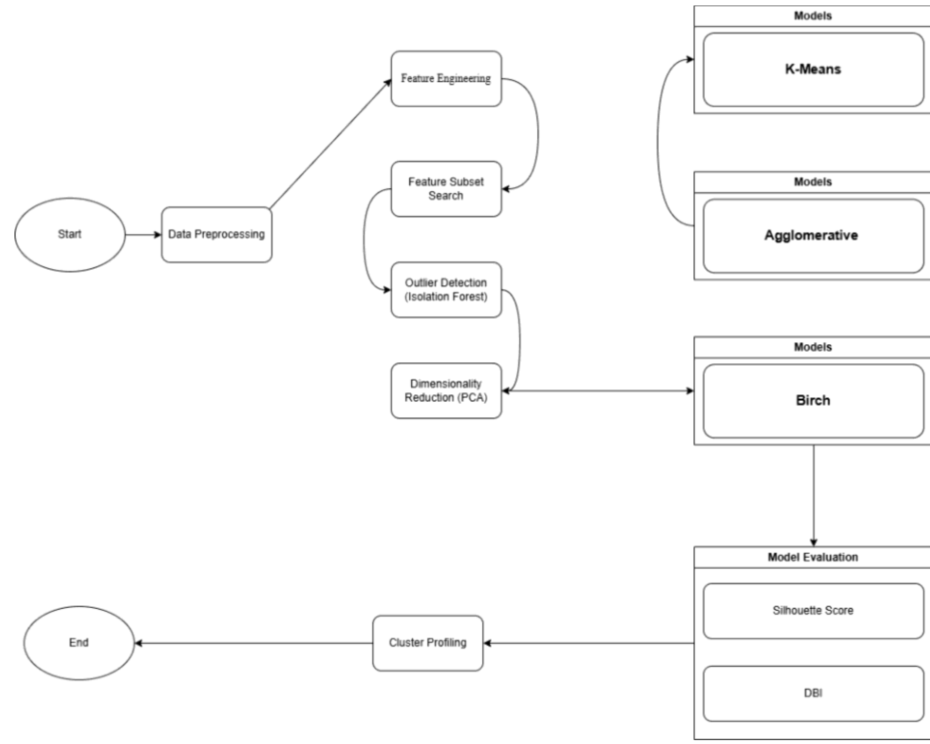

Figure 1 Research Workflow

The workflow begins with dataset preparation and preprocessing, followed by the creation of behavioral indicators through feature engineering. The optimal feature subset is then determined through an automated search strategy. After removing outliers and reducing dimensionality using Principal Component Analysis, clustering algorithms are applied to identify user behavior patterns. Finally, cluster quality is evaluated and behavioral profiles are interpreted.

Dataset Description

The dataset used in this study contains behavioral records of 7,500 smartphone users. Each record represents a user and includes several variables describing smartphone interaction behavior, such as daily screen time, social media usage, gaming activity, notification frequency, and application interaction intensity.

Two identifier attributes, namely transaction_id and user_id, were removed because they do not contribute to behavioral analysis. The remaining attributes represent behavioral indicators used to identify patterns of smartphone interaction.

Data Preprocessing

Data preprocessing was conducted to ensure the dataset is suitable for clustering analysis. Missing values in numerical variables were handled using median imputation, which replaces missing observations with the median value of each variable.

To reduce the influence of extreme observations, winsorization was applied by capping values at the 1st and 99th percentiles. This method prevents extreme behavioral values from dominating the clustering process.

Feature scaling was performed using RobustScaler, which scales the data using the median and interquartile range. The robust scaling transformation is defined as:

$$x_{scaled} = \frac{x - median(x)}{IQR(x)} \quad (1)$$

$IQR(x)$represents the interquartile range of the variable.

Feature Engineering

To capture deeper behavioral characteristics of smartphone interaction, several derived features were constructed from the original variables. These engineered features describe relative usage patterns rather than absolute usage values.

The derived features include:

Weekend Usage Difference:

$$Weekend\_Delta = Weekend\_Screen\_Time - Daily\_Screen\_Time \quad (2)$$

These variable measures the difference between weekend smartphone usage and average daily usage.

Social Media Ratio:

$$Social\_Ratio = \frac{Social\_Media\_Hours}{Daily\_Screen\_Time} \quad (3)$$

This ratio represents the proportion of social media usage relative to total screen time.

Gaming Ratio:

$$Gaming\_Ratio = \frac{Gaming\_Hours}{Daily\_Screen\_Time} \quad (4)$$

This variable describes the relative share of gaming activity within overall smartphone usage.

Notification Frequency per Hour:

$$Notif\_Per\_Hour = \frac{Notifications\_Per\_Day}{Daily\_Screen\_Time} \quad (5)$$

This metric captures the frequency of digital interruptions experienced by the user.

These derived variables provide more informative behavioral indicators that help differentiate smartphone usage patterns.

Feature Subset Selection

To determine the most informative behavioral indicators, an automated feature subset search strategy was applied. Candidate feature combinations were generated from both original and engineered variables.

Each feature subset was evaluated through clustering experiments. Highly correlated variables within a subset were removed to prevent redundancy. A correlation threshold was applied to eliminate features that exhibit strong multicollinearity.

The clustering configurations were then ranked based on clustering quality metrics.

Outlier Detection

Outlier detection was performed using the Isolation Forest algorithm, which isolates anomalous observations through random partitioning of the dataset.

Isolation Forest computes an anomaly score defined as:

$$s(x) = 2^{-\frac{E(h(x))}{c(n)}} \tag{6}$$

$E(h(x))$is the expected path length of observation $x$and $c(n)$represents the average path length of unsuccessful searches in binary trees.

Observations with high anomaly scores were considered potential outliers. In this study, approximately 3 percent of the data was identified as outliers and excluded during the clustering model training process.

Dimensionality Reduction

Principal Component Analysis (PCA) was applied to reduce the dimensionality of the dataset and improve cluster separability.

PCA transforms the original feature space into a new orthogonal coordinate system defined by principal components. The transformation can be expressed as:

$$Z = XW \tag{7}$$

Xrepresents the standardized data matrix and Wrepresents the eigenvector matrix obtained from the covariance matrix of X.

The PCA configuration used in this study retains 85 percent of the total variance, ensuring that the most important behavioral patterns are preserved while reducing dimensional complexity.

Clustering Algorithms

Three clustering algorithms were evaluated in this study, namely K-Means Clustering, Agglomerative Hierarchical Clustering, and Birch Clustering. These algorithms were selected because they represent different clustering strategies and are widely used for behavioral data analysis. K-Means clustering partitions the dataset into a predefined number of clusters by minimizing the within-cluster variance, where each observation is assigned to the nearest cluster centroid. Agglomerative Hierarchical Clustering follows a bottom-up approach that iteratively merges the closest clusters based on similarity measures, forming a hierarchical structure of clusters. Meanwhile, Birch Clustering is designed to efficiently process large datasets by incrementally constructing a clustering feature tree that summarizes the dataset and enables scalable clustering performance. These algorithms were compared to determine the most suitable method for identifying smartphone usage behavior patterns in the dataset.

$$J = \sum_{i=1}^{k} \sum_{x \in C_i} || x - \mu_i ||^2 \qquad (8)$$

$C_i$represents cluster $i$and $\mu_i$represents the centroid of cluster i.

Agglomerative clustering constructs a hierarchical structure by iteratively merging the closest clusters based on similarity measures.

Birch clustering builds a clustering feature tree that incrementally summarizes large datasets and enables efficient clustering for large scale behavioral data.

Cluster Evaluation

To determine the optimal clustering configuration, two internal evaluation metrics were used.

The Silhouette Score measures how similar a data point is to its own cluster compared with other clusters. It is defined as:

$$S(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \qquad (9)$$

$a(i)$is the average distance between observation $i$and other points in the same cluster, while $b(i)$is the smallest average distance between observation $i$and points in another cluster.

The Davies–Bouldin Index (DBI) evaluates cluster compactness and separation and is defined as:

$$DBI = \frac{1}{k} \sum_{i=1}^{k} \max_{j \neq i} \left(\frac{\sigma_i + \sigma_j}{d(c_i, c_j)} \right) \qquad (10)$$

$\sigma_i$represents the average distance of cluster members to the centroid and $d(c_i, c_j)$represents the distance between cluster centroids.

Higher silhouette values and lower DBI values indicate better clustering performance.

Cluster Profiling and Visualization

After identifying the optimal clustering configuration, cluster profiling was performed to interpret the behavioral characteristics of each user group. Mean values of key behavioral variables were calculated for each cluster to describe smartphone interaction patterns.

To visualize cluster separation, Principal Component Analysis visualization was applied by projecting the multidimensional behavioral data into two dimensions. The resulting visualization helps illustrate how the identified clusters are distributed in the behavioral feature space.

Algorithm 1: Smartphone Usage Behavior Clustering Framework

Input: dataset Dcontaining smartphone behavioral features

Output: cluster labels Crepresenting smartphone usage behavior groups

Process:

Start

Load dataset $D = \{x_1, x_2, \ldots, x_n\}$

Remove identifier attributes (transaction_id, user_id)

Handle missing values using median imputation

Scale features using RobustScaler

$$x' = \frac{x - median(x)}{IQR(x)}$$

Detect outliers using Isolation Forest

Remove observations classified as anomalies

Apply dimensionality reduction using PCA

$$Z = XW$$

Wrepresents the eigenvector matrix

Select components such that

$$Var(Z) \geq 0.85$$

Initialize clustering models

M = {KMeans, Agglomerative, Birch}

For each clustering model $m \in M$

For each number of clusters $k \in \{2,3,4\}$

Compute cluster assignments

$$C = m(Z, k)$$

Evaluate clustering performance

Silhouette score

$$S(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))}$$

Davies Bouldin Index

$$DBI = \frac{1}{k} \sum_{i=1}^{k} \max_{j \neq i} \left(\frac{\sigma_i + \sigma_j}{d(c_i, c_j)} \right)$$

Select best clustering configuration with highest silhouette and lowest DBI

Assign final cluster labels to dataset

Compute cluster profiles using mean behavioral features

End

Results

Optimal Clustering Configuration

To identify behavioral patterns in smartphone usage, a comprehensive clustering optimization process was conducted by systematically evaluating multiple clustering configurations using an automated feature subset search strategy. The objective of this procedure was to determine the optimal combination of behavioral features, dimensionality reduction settings, and clustering algorithms capable of producing the most meaningful and well-separated cluster structure. Clustering performance was assessed using two widely adopted internal validation metrics: the Silhouette Score and the Davies–Bouldin Index (DBI). The Silhouette Score measures the degree to which each observation is similar to members of its own cluster compared with observations in other clusters, with higher values indicating better-defined clusters. In contrast, the Davies–Bouldin Index evaluates the average similarity between clusters based on their compactness and separation, where lower values indicate superior clustering quality. Overall, 210 feature subset combinations were evaluated across multiple clustering algorithms, different numbers of clusters, and various dimensionality reduction configurations to identify the model that provided the best representation of smartphone usage behavior.

As summarized in table 1, the optimal clustering configuration was obtained using the Birch algorithm with two clusters (K = 2). The selected model was built using three behavioral indicators that were identified as the most discriminative features: daily screen time hours, the proportion of social media usage relative to total screen time (social ratio), and the number of notifications received per hour (notifications per hour). Before clustering, Principal Component Analysis (PCA) was applied to reduce data dimensionality while preserving 85% of the total variance, thereby minimizing redundancy among correlated variables and reducing the influence of noise while retaining the most informative behavioral characteristics.

Table 1 Optimal Clustering Configuration

Metric	Value
Algorithm	Birch
Number of clusters	2
Silhouette Score	0.4904
Davies–Bouldin Index	0.8764
PCA Variance Retained	0.85
Selected Features	daily_screen_time_hours, social_ratio, notif_per_hour

The final model achieved a Silhouette Score of 0.4904 and a Davies–Bouldin Index of 0.8764, indicating that the identified clusters exhibit a satisfactory balance between within-cluster cohesion and between-cluster separation. Although the Silhouette Score is below the threshold typically associated with highly distinct clusters (e.g., >0.70), a value close to 0.50 is generally considered acceptable in behavioral and human activity datasets, where user behavior naturally overlaps due to variations in daily routines, application preferences, and interaction habits. Furthermore, the relatively low DBI value suggests that the resulting clusters remain sufficiently compact and well separated, supporting the effectiveness of the selected configuration. Consequently, the combination of the Birch clustering algorithm, PCA with 85% retained variance, and the three selected behavioral features provides the most appropriate representation of smartphone usage patterns within the analyzed dataset.

Cluster Distribution

The clustering analysis revealed the presence of two distinct groups of smartphone users within the dataset. The distribution of users across the identified clusters is illustrated in figure 2. The results show that Cluster 0 contains 1,784 users, whereas Cluster 1 contains 5,716 users. This distribution indicates that the clusters are not evenly balanced, with Cluster 1 representing a substantially larger portion of the dataset. The difference in cluster size suggests that a dominant behavioral pattern exists among the smartphone users included in the analysis.

In terms of proportions, approximately 76 percent of the users belong to Cluster 1, while the remaining 24 percent are classified in Cluster 0. The predominance of Cluster 1 indicates that the majority of users demonstrate a similar smartphone usage pattern characterized by higher levels of digital engagement. This pattern reflects users who spend more time interacting with their devices and exhibit stronger behavioral indicators associated with intensive smartphone usage. The presence of a smaller secondary cluster suggests that a subset of users follows a different behavioral pattern with comparatively lower engagement levels. Together, these clusters highlight the existence of distinct smartphone usage behaviors within the digital society represented by the dataset.

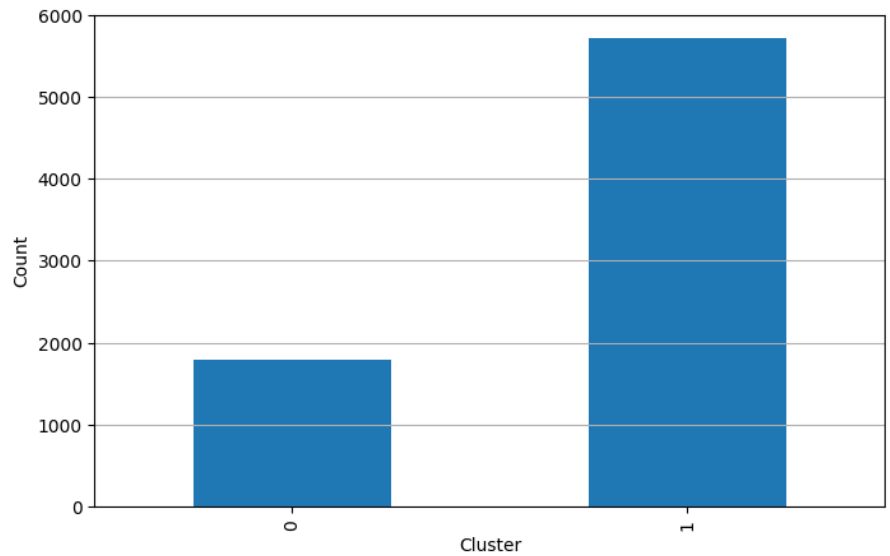

Figure 2 Cluster size distribution of smartphone users.

Cluster Visualization

Principal Component Analysis (PCA) was applied to visualize the separation between the identified clusters in a lower dimensional space. The original dataset contains multiple behavioral variables describing smartphone usage patterns, which makes direct visualization difficult. PCA transforms these variables into a smaller set of components that capture the most important variance in the data. In this study, the multidimensional behavioral space was reduced into two principal components to enable graphical representation of user distribution across clusters. The resulting projection is presented in figure 3, where each point represents an individual user positioned according to the two principal components derived from the behavioral features.

The visualization indicates a noticeable separation between the two clusters, primarily along the first principal component. This pattern suggests that the dominant source of variation among smartphone users is related to differences in digital engagement intensity. Users in one cluster tend to exhibit behavioral characteristics associated with higher levels of smartphone interaction, such as longer screen time and more frequent notification activity, while the other cluster reflects comparatively lower levels of engagement. The separation observed in the PCA projection therefore supports the clustering results by demonstrating that the identified groups occupy distinct regions in the reduced behavioral feature space.

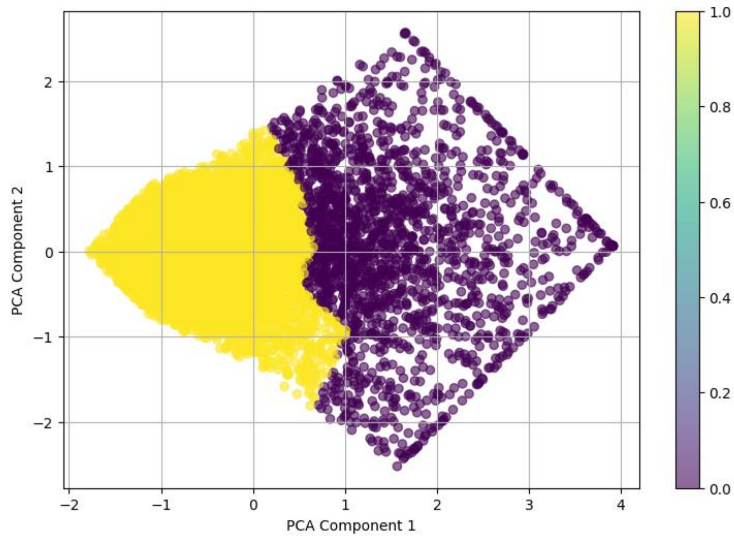

Figure 3 PCA visualization of smartphone usage behavior clusters.

Behavioral Profile of Clusters

To obtain a deeper understanding of the behavioral characteristics represented by each cluster, a comprehensive cluster profiling analysis was performed by calculating the mean value of several smartphone usage variables for every cluster. While the clustering model was constructed using a subset of the most informative features identified during the optimization process, the profiling stage incorporates additional behavioral variables that were not directly used for clustering. This approach provides a more comprehensive interpretation of the resulting clusters by examining a wider range of smartphone usage behaviors. The analyzed variables include daily screen time, social media usage hours, gaming hours, work or study hours, sleep duration, number of notifications received per day, number of application openings, and weekend screen time. These variables collectively describe different dimensions of smartphone engagement, including overall usage intensity, preferred activity types, interaction frequency, and lifestyle-related behaviors. By comparing the average values of these indicators across clusters, it becomes possible to characterize the dominant behavioral patterns associated with each group and provide meaningful interpretations beyond the numerical clustering results.

The summarized behavioral characteristics of the two clusters are presented in table 2. Overall, the descriptive statistics reveal noticeable differences in smartphone usage intensity between the two user groups, although several behavioral variables remain relatively similar. The most prominent distinction is observed in the amount of time users spend interacting with their smartphones on both weekdays and weekends, suggesting that overall device engagement is the primary factor separating the two clusters.

Table 2 Behavioral Characteristics of Smartphone User Clusters

Feature	Cluster 0	Cluster 1
Daily screen time (hours)	4.46	8.45
Social media hours	3.78	3.12
Gaming hours	2.04	2.01
Work/study hours	3.18	3.26
Sleep hours	6.73	6.74
Notifications per day	168.09	123.70
App opens per day	96.62	98.21
Weekend screen time	6.2	10.2

Users assigned to Cluster 0 exhibit relatively moderate smartphone usage behavior, with an average daily screen time of 4.46 hours, which is nearly half that observed in Cluster 1. Despite spending less total time on their smartphones, these users record an average of 3.78 hours of social media usage, indicating that social networking applications account for a substantial proportion of their daily smartphone activity. In addition, Cluster 0 users receive an average of 168.09 notifications per day, considerably higher than users in Cluster 1. This finding suggests that these users interact with their smartphones more frequently through short, notification-driven sessions rather than prolonged continuous usage. The average number of application openings per day (96.62) further supports the interpretation that Cluster 0 users regularly check their devices throughout the day despite having relatively limited overall screen time. Their average weekend screen time of 6.2 hours also indicates only a moderate increase in smartphone usage during non-working days.

In contrast, Cluster 1 represents users with substantially higher levels of smartphone engagement. These users spend an average of 8.45 hours per day on their smartphones, almost double the usage observed in Cluster 0. The difference becomes even more pronounced during weekends, where the average screen time increases to 10.2 hours, suggesting prolonged recreational or entertainment-related smartphone use during leisure periods. Interestingly, although Cluster 1 users spend more total time using their smartphones, they allocate fewer hours specifically to social media (3.12 hours) and receive fewer daily notifications (123.70) compared with Cluster 0. This indicates that their longer screen time may be distributed across a wider variety of applications and activities rather than being dominated by social networking or frequent notification-driven interactions.

Other behavioral variables exhibit relatively minor differences between the two clusters. The average gaming duration remains nearly identical, with users in Cluster 0 and Cluster 1 spending approximately 2.04 and 2.01 hours, respectively. Similarly, work or study activities show comparable averages (3.18 versus 3.26 hours), suggesting that productivity-related smartphone use contributes similarly across both user groups. Average sleep duration is also remarkably consistent between clusters, with both groups reporting approximately 6.7 hours of sleep per day. Furthermore, the average number of application openings differs only slightly (96.62 in Cluster 0 versus 98.21 in Cluster 1), indicating that both groups access applications with similar frequency, even though the duration of smartphone use differs considerably.

Smartphone Addiction Distribution

To further interpret the behavioral significance of the identified clusters, an additional analysis was conducted by examining the distribution of the addicted_label variable across the clustering results. The addicted_label is a binary indicator representing whether a user exhibits behavioral characteristics associated with smartphone addiction, based on the labeling criteria defined in the dataset. Although this variable was not included as one of the features used during the clustering process, it serves as an external reference for evaluating the behavioral relevance of the resulting clusters. By comparing the prevalence of addicted and non-addicted users within each cluster, it becomes possible to assess whether the unsupervised clustering algorithm successfully groups individuals according to meaningful differences in smartphone dependency. Such an analysis also provides additional validation that the clusters capture behavioral patterns associated with problematic smartphone use rather than merely separating users based on arbitrary differences in smartphone activity.

The distribution of smartphone addiction across the two clusters is summarized in table 3. Overall, the results reveal a substantial difference in the proportion of addicted users between the clusters, indicating that the behavioral profiles identified through clustering are closely associated with smartphone addiction status. Although the clustering algorithm was developed without using addiction labels, the resulting groups exhibit markedly different addiction distributions, suggesting that the selected behavioral features effectively capture underlying patterns related to problematic smartphone use.

Table 3 Distribution of Smartphone Addiction Across Clusters

Cluster	Non-Addicted	Addicted
Cluster 0	884	900
Cluster 1	1308	4408

Within Cluster 0, the distribution between addicted and non-addicted users is relatively balanced. Specifically, 900 users are classified as addicted, while 884 users are classified as non-addicted. This corresponds to approximately 50.4% addicted users and 49.6% non-addicted users, indicating that smartphone addiction is not the dominant characteristic of this group. The balanced distribution is consistent with the behavioral profile presented in Table 2, where Cluster 0 is characterized by moderate daily screen time, relatively shorter weekend smartphone use, and a greater proportion of notification-driven interactions. These behavioral characteristics suggest that users in this cluster engage with their smartphones regularly but generally maintain usage patterns that are less indicative of excessive or compulsive smartphone dependence.

In contrast, Cluster 1 exhibits a markedly different distribution. Among the 5,716 users assigned to this cluster, 4,408 users are classified as addicted, whereas only 1,308 users are categorized as non-addicted. Consequently, approximately 77.1% of users in Cluster 1 exhibit smartphone addiction characteristics, compared with only 22.9% who are classified as non-addicted. This considerable imbalance indicates that smartphone addiction is a predominant characteristic of this cluster. The high prevalence of addicted users aligns closely with the behavioral profile previously identified, particularly the substantially higher average daily screen time (8.45 hours) and significantly longer weekend screen time (10.2 hours). These findings suggest that prolonged and sustained smartphone engagement is strongly associated with addictive smartphone behavior.

The comparison between the two clusters demonstrates that overall smartphone usage intensity appears to be one of the strongest behavioral indicators associated with addiction. While both clusters display relatively similar averages for gaming activity, work or study usage, sleep duration, and application opening frequency, the most pronounced differences occur in variables related to total screen exposure. Users in Cluster 1 spend considerably more time interacting with their smartphones throughout both weekdays and weekends, whereas users in Cluster 0 maintain shorter overall usage despite receiving more notifications and allocating a larger proportion of their screen time to social media. This observation suggests that the duration of smartphone use, rather than simply the frequency of interactions or the type of application used, may play a more influential role in distinguishing users with higher levels of smartphone dependency.

Discussion

The clustering analysis provides important insights into the diversity of smartphone usage behaviors within a digital society. The results show that users can be broadly divided into two behavioral groups that differ primarily in terms of their level of digital engagement. The first group represents users who exhibit moderate smartphone usage patterns characterized by lower daily screen time and relatively balanced interaction with their devices. Although these users still engage actively with social media and receive a relatively high number of notifications, their overall interaction with smartphones remains limited compared with the second group. This pattern suggests that smartphones are used mainly as tools for communication and social interaction rather than as a dominant component of daily digital activity. Such behavior reflects a more controlled use of mobile technology where smartphone interaction supports daily routines without significantly extending overall screen exposure.

In contrast, the second cluster represents users with substantially higher levels of digital engagement. Users in this group demonstrate significantly longer daily screen time and increased smartphone usage during weekends, indicating that smartphone interaction extends beyond functional purposes and becomes a central activity during leisure time. The high proportion of addicted users within this cluster further suggests that prolonged exposure to digital environments may contribute to the development of problematic smartphone usage behaviors. Frequent interaction with digital platforms, combined with continuous notification stimuli, may reinforce habitual checking behavior and sustained engagement with mobile applications. These findings highlight how variations in engagement intensity can shape different behavioral patterns among smartphone users and provide evidence that higher levels of digital interaction are closely associated with a greater likelihood of smartphone dependency in contemporary digital environments.

Conclusion

This study aimed to identify smartphone usage behavior patterns within a digital society by applying clustering algorithms to behavioral data derived from smartphone interaction variables. Using an automated feature subset search combined with dimensionality reduction and clustering evaluation metrics, the analysis determined that the Birch algorithm with two clusters provided the most appropriate segmentation of users. The optimal clustering model was formed based on three behavioral indicators consisting of daily screen time, the proportion of social media usage relative to total screen time, and the frequency of notifications received per hour. The results revealed two distinct user groups characterized by different levels of digital engagement. The first group represents users with moderate smartphone interaction patterns, while the second group consists of users with substantially higher levels of digital engagement reflected in longer screen time and increased weekend smartphone usage. The analysis also demonstrated that the cluster characterized by higher digital engagement contains a considerably larger proportion of users labeled as addicted, indicating a strong relationship between prolonged smartphone interaction and problematic usage behavior.

Declarations

Author Contributions

Conceptualization: M.B.Y., R.; Methodology: M.B.Y., R.; Software: R.; Validation: M.B.Y.; Formal Analysis: M.B.Y.; Investigation: R.; Resources: M.B.Y., R.; Data Curation: R.; Writing – Original Draft Preparation: M.B.Y.; Writing – Review and Editing: R.; Visualization: R.; All authors have read and agreed to the published version of the manuscript.

Data Availability Statement

The data presented in this study are available on request from the corresponding author.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] A. Oulasvirta, T. Rattenbury, L. Ma, and E. Raita, "Habits make smartphone use more pervasive," *Personal and Ubiquitous Computing*, vol. 16, no. 1, pp. 105–114, Jun. 2012. doi: 10.1007/s00779-011-0412-2.

[2] K. Demirci, M. Akgönül, and A. Akpinar, "Relationship of smartphone use severity with sleep quality, depression, and anxiety in university students," *Journal of Behavioral Addictions*, vol. 4, no. 2, pp. 85–92, Jun. 2015. doi: 10.1556/2006.4.2015.010.

[3] M. Kwon, D.-J. Kim, H. Cho, and S. Yang, "The smartphone addiction scale: Development and validation of a short version for adolescents," *PLOS ONE*, vol. 8, no. 12, p. e83558, pp. 1-7, Dec. 2013. doi: 10.1371/journal.pone.0083558.

[4] Y. Ding, X. Wan, G. Lu, H. Huang, Y. Liang, J. Yu, and C. Chen, "The associations

between smartphone addiction and self-esteem, self-control, and social support among Chinese adolescents: A meta-analysis," *Frontiers in Psychology*, vol. 13, no. November, p. 1029323, pp. 1-17, Nov. 2022. doi: 10.3389/fpsyg.2022.1029323.

[5] C. Montag, C. Sindermann, and H. Baumeister, "Digital phenotyping in psychological and medical sciences: A reflection about necessary prerequisites to reduce harm and increase benefits," *Current Opinion in Psychology*, vol. 36, no. December, pp. 19–24, Dec. 2020. doi: 10.1016/j.copsyc.2020.03.013.

[6] P. M. Lekgetho, "Mobile technology and social transformations: Access to knowledge in global contexts," *Information, Communication & Society*, vol. 27, no. 1, pp. 198–200, 2024. doi: 10.1080/1369118X.2023.2200310.

[7] Y. Kwon, K. Kang, and C. Bae, "Unsupervised learning for human activity recognition using smartphone sensors," *Expert Systems with Applications*, vol. 41, no. 14, no. October, pp. 6067–6074, Oct. 2014. doi: 10.1016/j.eswa.2014.04.037.

[8] É. Duke and C. Montag, "Smartphone addiction, daily interruptions and self-reported productivity," *Addictive Behaviors Reports*, vol. 6, no. July, pp. 90–95, Jul. 2017. doi: 10.1016/j.abrep.2017.07.002.

[9] J. Harwood, J. M. Gutiérrez Julián, A. J. Scott, and R. Joiner, "Constantly connected – The effects of smart-devices on mental health," *Computers in Human Behavior*, vol. 34, no. May, pp. 267–272, May 2014. doi: 10.1016/j.chb.2014.02.006.

[10] Z. Khalid and N. U. Amin, "Predictive model for analyzing smartphone usage and behavioral patterns using machine learning," *Preprints*, vol. 2025, no. November, pp. 1-9, 2025, doi: 10.20944/preprints202511.1433.v1.

[11] C. Montag, B. Lachmann, M. Herrlich, and K. Zweig, "Addictive features of social media/messenger platforms and freemium games against the background of psychological and economic theories," *International Journal of Environmental Research and Public Health*, vol. 16, no. 14, p. 2612, pp. 1-16, Jul. 2019. doi: 10.3390/ijerph16142612.

[12] C.-R. Poetar, A. Dobrean, and I. S. Florean, "Problematic smartphone use and mental health problems: A network analysis of Romanian adolescents," *Youth & Society*, vol. 56, no. 6, pp. 1477 – 1455, 2023. doi: 10.1177/0044118X231218295.

[13] S. Thomée, "Mobile phone use and mental health: A review of the research that takes a psychological perspective on exposure," *International Journal of Environmental Research and Public Health*, vol. 15, no. 12, p. 2692, pp. 1-25, Nov. 2018. doi: 10.3390/ijerph15122692.

[14] J. A. Roberts, C. Pullig, and C. Manolis, "I need my smartphone: A hierarchical model of personality and cell-phone addiction," *Personality and Individual Differences*, vol. 79, no. June, pp. 13–19, Jun. 2015. doi: 10.1016/j.paid.2015.01.049.

[15] J. B. MacQueen, "Some methods for classification and analysis of multivariate observations," in *Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability*, vol. 1, Berkeley, CA, USA: University of California Press, 1967, pp. 281–297.

[16] G. Gan, C. Ma, and J. Wu, *Data Clustering: Theory, Algorithms, and Applications*. Philadelphia, PA, USA: SIAM, 2007. doi: 10.1137/1.9780898718348.

[17] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*, 3rd ed.

Waltham, MA, USA: Morgan Kaufmann, 2012.

[18] Q. Su, Z. Jia, and Q. Lu, "Research on user behavior clustering algorithm based on mobile application," *Journal of Intelligent & Fuzzy Systems*, vol. 35, no. 1, pp. 1–10, Jun. 2018. doi: 10.3233/JIFS-169673.

[19] S. Zhao, J. A. Ramos Rojas, J. Tao, and Z. Jiang, "Discovering different kinds of smartphone users through their application usage behaviors," in *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp 2016)*, vol. 2016, no. Sep, pp. 498 – 509, 2016. doi: 10.1145/2971648.2971696.

[20] V. Pawar, A. B. Patil, and A. Santra, "Predicting smartphone addiction using behavioral and psychological traits with machine learning," in *2025 Global Conference in Emerging Technology (GINOTECH)*, Pune, India, vol. 2025, no. July, pp. 1–6, 2025, doi: 10.1109/GINOTECH63460.2025.11077064.