ISSN 2579-7069

Identifying Adolescent Behavioral Profiles Through K-Means Clustering Based on Smartphone Usage, Mental Health, and Academic Performance

Dominic Dinand Aristo[1,*], Bhavana Srinivasan[2]

[1] Information System Study Prorgam, Faculty of Computer Science, Amikom Purwokerto University, Indonesia

[2] *Department of Animation and Virtual Reality, JAIN, Bangalore, India*

(Received June 1, 2024; Revised August 5, 2024; Accepted December 2, 2024; Available online January 4, 2025)

Abstract

The pervasive integration of digital devices into students' daily lives has profoundly shaped their learning habits and psychological well-being. As technology becomes increasingly embedded in academic and personal routines, understanding the relationship between digital engagement, mental health, and academic outcomes is vital for developing effective student-support and intervention frameworks in higher education. This study seeks to uncover behavioral patterns among college students by examining the interconnections between smartphone usage, mental health indicators, and academic performance through a data-driven machine learning approach. Utilizing the K-Means clustering algorithm, students were categorized into distinct behavioral profiles derived from eight core features: daily screen time, sleep duration, grade performance, exercise frequency, anxiety level, depression level, self-confidence, and screen exposure before sleep. A dataset comprising 3,000 entries was preprocessed through normalization and analyzed within the Knowledge Discovery in Databases (KDD) framework to ensure structured and reliable data processing. The Elbow Method identified four optimal clusters, each reflecting unique behavioral characteristics. Cluster 1 represented well-balanced students with stable academic and emotional states; Cluster 2 included high-achieving yet anxious individuals; Cluster 3 captured those exhibiting excessive digital engagement and psychological distress; and Cluster 4 comprised moderately engaged students with lower self-confidence. Visual representations, including bar and radar charts, were generated to illustrate inter-cluster variations and enhance interpretability of behavioral distinctions. The findings reveal that digital usage patterns are closely linked to mental health and academic performance, suggesting that excessive or unregulated device use can heighten emotional strain and academic inconsistency. These insights highlight the necessity of personalized mental health initiatives and targeted digital literacy programs grounded in behavioral segmentation. Overall, the study demonstrates the applicability of unsupervised machine learning for behavioral profiling and provides evidence-based recommendations for educators, mental health practitioners, and policymakers seeking to foster balanced and healthy digital habits among students.

Keywords: Machine Learning, Adolescent, Smartphone Usage, Mental Health, K-Means Algorithm, Cluster Analysis, Addiction Patterns

1. Introduction

The use of smartphones has become deeply embedded in the daily lives of adolescents. This technology provides substantial benefits by facilitating communication, instant access to information, online learning, and entertainment. For students in particular, smartphones have become indispensable tools that support educational and social activities. However, alongside these advantages, there is growing evidence of negative consequences associated with excessive or unregulated smartphone use. Numerous studies have indicated that overdependence on smartphones can disrupt adolescents' mental health, reduce sleep quality, impair academic performance, and undermine self-confidence [1].

Adolescence represents a critical developmental stage characterized by emotional fluctuation, identity formation, and social experimentation. During this period, individuals are particularly susceptible to behavioral dependencies, including digital addiction. Typical manifestations include compulsive checking of mobile phones, excessive engagement with social media platforms, prolonged screen exposure, and neglect of physical or academic activities. Such behaviors have been empirically linked to elevated levels of anxiety, depressive symptoms, and diminished cognitive focus [2]. Over time, these tendencies can interfere with emotional regulation, interpersonal relationships, and educational achievement, indicating that smartphone overuse poses both psychological and academic challenges.

Nevertheless, not all adolescents demonstrate uniform behavioral patterns in their digital interactions. Some manage their smartphone use responsibly without significant negative outcomes, while others exhibit maladaptive patterns that

*Corresponding author: Dominic Dinand Aristo (domicdinand@gmail.com)
DOI: https://doi.org/10.47738/ijiis.v8i1.231

ISSN 2579-7069

contribute to mental or academic strain. These variations suggest the presence of complex, multidimensional behavioral dynamics that cannot be fully captured through conventional statistical analyses alone. Traditional methods often assume linear relationships and overlook hidden subgroup structures within heterogeneous populations. Therefore, a more advanced analytical approach is necessary to uncover latent behavioral profiles among adolescents.

In this context, machine learning—particularly unsupervised learning—offers powerful tools for discovering hidden structures within complex data. The K-Means clustering algorithm is one of the most effective and widely used techniques for this purpose. It can automatically group individuals into clusters based on similarities across multiple behavioral and psychological attributes, without the need for predefined labels. This capability makes it especially suitable for identifying subtle and overlapping patterns in adolescent digital behavior and psychosocial data [3].

To ensure systematic and reproducible analysis, this study adopts the Knowledge Discovery in Databases (KDD) framework. KDD provides an end-to-end methodological structure encompassing data selection, preprocessing, transformation, mining, pattern evaluation, and knowledge presentation [4]. This integration not only enhances analytical rigor but also ensures interpretability and replicability of results.

Based on this rationale, the present study aims to classify adolescents according to their smartphone usage behaviors by considering both digital engagement and psychosocial dimensions. Variables analyzed include daily usage duration, sleep hours, academic performance, anxiety level, depression level, and self-confidence. By implementing the K-Means algorithm within the KDD framework, this study seeks to uncover behavioral clusters that can inform targeted digital well-being and mental health intervention strategies tailored to adolescents' specific needs.

2. Literature Review

2.1. Smartphone Usage and Mental Health in Adolescents

Adolescents represent one of the most vulnerable demographics in terms of psychological development and digital exposure. The growing use of smartphones among this group has triggered an increasing number of studies analyzing its correlation with anxiety, depression, and academic disruption [5]. According to Mohd Saat et al [6], excessive screen time among teenagers is significantly associated with depressive symptoms and reduced sleep quality. Similarly, Shin et al [7] conducted a meta-analysis and found that misuse of social media on mobile phones was consistently associated with higher levels of psychological distress. These findings suggest a potential causal pathway between digital behaviors and emotional instability among adolescents. In addition to mental health, educational outcomes are also affected. A longitudinal study by Weerasinghe et al. [8] revealed that students with high daily smartphone usage exhibited poorer academic performance, emphasizing the trade-off between digital engagement and cognitive outcomes. Smartphone usage has become a behavioral marker, not only of communication habits but also of social identity and psychological coping mechanisms in adolescence [9].

2.2. Clustering and Behavioral Profiling with K-Means

The need to group individuals based on complex behavioral data has driven the application of unsupervised learning, particularly clustering. K-Means, as one of the most commonly used clustering algorithms, offers simplicity, scalability, and interpretability in segmenting multidimensional data [10]. According to Jhon and alex [11], clustering is an essential tool in behavioral science, particularly when researchers seek to uncover patterns without predefined labels. In recent years, clustering techniques have gained traction in psychological and educational domains for their ability to uncover latent behavioral traits without predefined labels. For instance, a study by Mohd Talib et al. [12] applied K-Means clustering to identify behavioral patterns among university students based on self-efficacy, learning styles, and learning program outcomes. The study successfully revealed three distinct student profiles low, average, and high performance based on holistic academic and behavioral indicators. The clustering process enabled the researchers to uncover patterns that would likely be overlooked using traditional statistical methods, thus highlighting the power of unsupervised learning approaches in educational data mining contexts

In another context, Mohamed Nafuri et al. [13] employed K-Means and other clustering algorithms to classify student performance among the B40 socioeconomic group in Malaysia, successfully identifying dropout risk and academic groupings. Likewise, Xu et al. [14] demonstrated the applicability of fuzzy clustering for mental health classification, proving effective in uncovering behavioral variance across student populations. These studies further support the relevance of clustering as a tool for both academic and psychological profiling in data-rich environments. The robustness of K-Means in large-scale behavioral datasets lies in its ability to minimize intra-cluster variance and enhance inter-cluster separation. However, one notable limitation is the need to predefine the number of clusters, often addressed using the Elbow Method or Silhouette Score [15]. Despite this limitation, its effectiveness in segmentation has led to widespread adoption in mental health informatics, especially when integrated with domain-specific knowledge.

2.3. KDD Framework in Behavioral Data Mining

Knowledge Discovery in Databases (KDD) provides a systematic pipeline for extracting meaningful patterns from large data repositories. First introduced by Fayyad et al. [16], the KDD process comprises several stages including data selection, preprocessing, transformation, data mining, evaluation, and presentation. In studies involving behavioral data, KDD has proven effective for structuring machine learning workflows while maintaining scientific rigor. When dealing with multidimensional adolescent behavior data, the KDD framework facilitates a clear methodology from identifying relevant features (e.g., sleep duration, anxiety levels, phone checks) to transforming and mining insights using clustering techniques. According to Yu Nie et al [17], KDD is particularly suited to educational and psychological domains due to the variety and complexity of data involved. This aligns with the work of Shu and Ye [18] , who argue that unsupervised learning methods such as cluster analysis and latent class analysis play a vital role in revealing hidden subgroups and behavioral typologies within complex social datasets. By integrating machine learning with classical statistical thinking, the KDD framework supports both inductive pattern recognition and deductive hypothesis testing, thereby enhancing theory development and practical intervention strategies in adolescent behavioral studies. These studies reinforce the utility of Knowledge Discovery in Databases for extracting actionable behavioral archetypes from digital trace data, facilitating targeted educational interventions.

3. Method

This study applies a data-driven approach to segment adolescent smartphone usage behavior using the K-Means clustering algorithm within the Knowledge Discovery in Databases (KDD) framework. The methodological pipeline consists of six major stages, as described below.

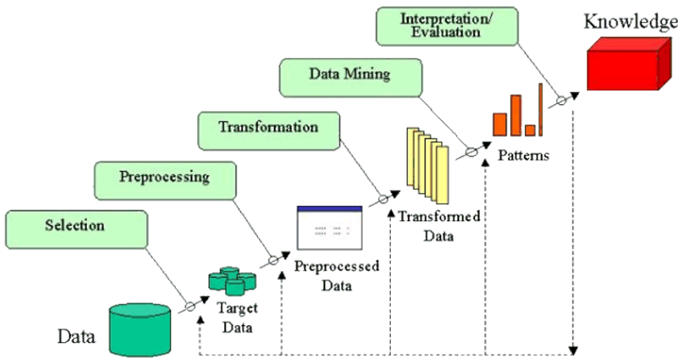

Figure 1. The Knowledge Discovery in Databases (KDD) process

3.1. Data Source and Feature Selection

The dataset used in this study, titled teen_phone_addiction_dataset.csv, comprises 3,000 data entries and 25 attributes representing behavioral, psychosocial, and lifestyle factors associated with adolescent smartphone use. The data were obtained from an open-access behavioral informatics repository focusing on digital addiction patterns among teenagers. Each record corresponds to one adolescent respondent, containing quantitative information related to smartphone habits, academic performance, and emotional indicators. For the purpose of this study, thirteen variables were selected for clustering analysis based on their theoretical relevance and empirical support from previous studies on smartphone usage, mental health, and learning outcomes. The selected features cover three main domains. The first is digital behavior, represented by variables such as daily smartphone usage hours, screen time before bedtime, weekend usage hours, phone checks per day, and the duration of time spent on social media, gaming, and educational applications. The second domain is psychosocial indicators, which include sleep duration, academic performance, anxiety level, depression level, and self-esteem. The final domain encompasses lifestyle factors, specifically daily exercise hours. The inclusion of these variables aims to capture a holistic view of adolescent behavior by linking digital engagement patterns to psychological and academic well-being.

3.2. Data Preprocessing

The preprocessing stage was carried out to ensure that the dataset was clean, consistent, and ready for analysis. A completeness check was first performed to detect missing or null values across all variables. The results indicated that there were no missing entries among the selected features, suggesting good data quality. Outliers were examined using z-score analysis, and although a few extreme values were identified, they were retained to preserve natural behavioral diversity and avoid over-smoothing of the dataset. Feature normalization was then conducted to prevent bias arising from differences in measurement scale. As the K-Means clustering algorithm is highly sensitive to feature magnitude, all numeric variables were standardized using the StandardScaler function from the scikit-learn library, transforming each feature to have a mean of zero and a standard deviation of one. This normalization process ensured that all features contributed equally to the clustering process, preventing variables with larger numerical ranges from dominating the formation of clusters. The outcome of this stage was a clean and standardized dataset suitable for subsequent transformation and clustering analysis.

3.3. Data Transformation

Following the preprocessing stage, all selected features were transformed into a numerical matrix that served as the input for the machine learning algorithm. Categorical and irrelevant attributes, such as gender or location, were excluded to maintain analytical focus on behavioral and psychosocial patterns rather than demographic factors. Each feature was expressed in a standardized numerical form to ensure comparability across dimensions. The final dataset consisted of thirteen continuous and normalized variables, providing a compact yet comprehensive representation of adolescent smartphone usage behavior. This transformation process facilitated efficient computation and reduced the potential influence of noise, enabling the clustering algorithm to operate more effectively in identifying hidden behavioral structures.

3.4. Data Mining

The clustering analysis was performed using the K-Means algorithm to segment adolescents into distinct groups based on their behavioral and psychosocial similarities. The algorithm iteratively assigned each data point to the nearest cluster centroid by minimizing the within-cluster variance using Euclidean distance as the similarity metric. The number of clusters, denoted as k, was determined through the Elbow Method, which involves plotting the Within-Cluster Sum of Squares (WSS) against various k values to identify the point at which additional clusters no longer provide significant improvement. As illustrated in Figure 2, the WSS value dropped sharply between $k = 1$ and $k = 4$ before leveling off, forming a visible “elbow” at $k = 4$. This indicates that four clusters were optimal for representing the structure of the data without overfitting. The resulting model successfully grouped adolescents into four distinct behavioral categories, each reflecting a different combination of smartphone usage intensity, psychological well-being, and academic performance.

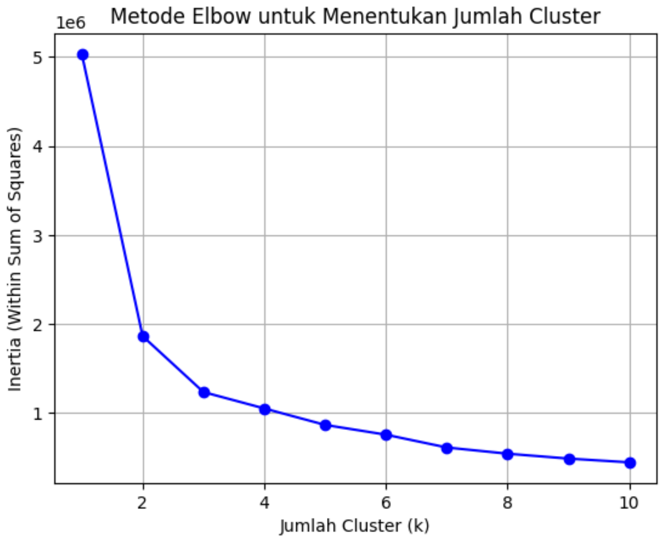

Figure 2. Elbow Method for Determining the Optimal Number of Clusters

This indicates an "elbow" at k = 4, suggesting that four clusters are optimal for capturing the structure of the data without overfitting. Therefore, the number of clusters was set to 4, informed by this exploratory analysis and the interpretability of behavioral patterns. K-Means iteratively assigns individuals to clusters based on the nearest centroid, using Euclidean distance as the similarity metric. The algorithm continues updating cluster centroids until convergence is reached or a predefined iteration limit is met. The algorithm iteratively minimized intra-cluster distance using the Euclidean distance formula:

$$d(x,y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}$$

3.5. Pattern Evaluation

Once clustering was completed, each cluster was analyzed to identify meaningful behavioral patterns. This involved calculating the average value of each feature within each cluster and comparing them across clusters. Based on these characteristics, descriptive labels were assigned to each group, such as “Balanced and Healthy,” “High Anxiety, High Achievers,” or “Heavy Smartphone Users at Risk.” These interpretations were guided by existing psychological theories and digital behavior research.

3.6. Knowledge Presentation

The final stage involved summarizing the findings in the form of tables and visualizations to communicate the differences among clusters effectively. The cluster profiles were presented as evidence of the diversity in adolescent smartphone behavior and its relationship with psychological well-being. These results can provide insights for educators, parents, policymakers, and mental health practitioners in designing targeted interventions [20].

4. Results and Discussion

Based on the KDD framework implementation, the following results present the clustering outcomes and behavioral interpretations derived from the adolescent smartphone usage dataset.

4.1. Cluster Summary and Characteristics

After applying the K-Means algorithm, the dataset was grouped into four distinct clusters. Each cluster represents a unique behavioral and psychological profile of adolescent smartphone users. The number of data points in each cluster and the centroids (mean values) were calculated to describe the central tendencies of each group. Each cluster exhibits

ISSN 2579-7069

distinct combinations of daily usage hours, academic performance, psychological well-being (anxiety, depression, self-esteem), and lifestyle habits (sleep and exercise). These cluster centroids serve as a foundation for interpreting behavioral patterns and labeling user profiles. To visualize the distribution of adolescents across the four clusters, a pie chart was constructed (Figure 3). The chart shows that each cluster holds a relatively similar proportion of the dataset, with Cluster 0 accounting for 26.3%, followed by Cluster 1 (25.2%), Cluster 2 (24.7%), and Cluster 3 (23.7%). This fairly even distribution suggests that no single behavior type dominates the population, which strengthens the generalizability of the analysis.

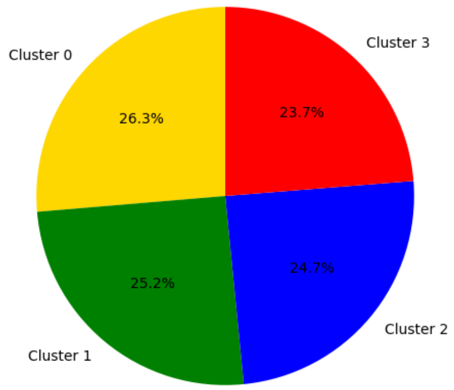

Figure 3. Proportion of Adolescents in Each Cluster

4.2. Representative and Outlier Profiles (Distance-Based Analysis)

To evaluate the cohesiveness of the clustering results, the two nearest and two farthest data points from the cluster centroids were examined using Euclidean distance. These comparisons help validate whether the clusters capture meaningful groupings and whether any behavioral outliers are present.

Table 1 Behavioral profile comparisons between representative and outlier individuals based on distance to cluster centroids.

Original Index	Daily Usage	Sleep	Academic	Exercise	Anxiety	Depression	Self-Esteem	Screen Before Bed	Phone Checks	Social Media	Gaming	Education	Weekend Usage	Pair
210	3.5	7.2	94	0.5	1	6	7	2.0	84	2.7	2.9	0.8	5.2	Nearest
2631	3.0	7.2	94	0.8	1	6	7	1.7	83	1.4	3.6	0.9	4.6	Nearest
263	3.5	6.5	100	2.0	8	4	5	0.9	150	0.2	2.3	1.6	2.7	Farthest
978	5.4	9.0	50	0.6	6	3	9	0.1	20	1.2	2.5	1.0	2.0	Farthest

The nearest data points (210 and 2631) show highly similar values in academic performance and psychological indicators, confirming internal cohesion. In contrast, the farthest points (263 and 978) display extreme behaviors such as excessive phone checks or lower academic performance, indicating potential outliers or subgroup variance within the cluster.

4.3. Behavioral Interpretation of Each Cluster

The clustering process produced four distinct behavioral groups based on adolescents' smartphone usage, psychological indicators, and academic performance. Each cluster presents a unique combination of digital behavior and mental health tendencies that reflect varying levels of risk and balance in their daily lives.

ISSN 2579-7069

Table 2 Cluster-specific behavioral profiles of adolescents based on smartphone usage, psychological indicators, and academic performance.

Cluster	Daily Use	Sleep	Academic	Exercise	Anxiety	Depression	Self-Esteem	ST Before Bed	Phone Checks	Social Media	Gaming	Education	Weekend Use
0	4.98	6.03	79.31	1.08	7.57	6.52	4.34	0.97	57.97	2.66	1.36	1.14	6.24
1	5.05	6.84	75.58	0.87	3.11	5.13	7.26	0.97	67.37	2.60	1.20	1.01	5.46
2	4.80	6.13	73.90	1.15	3.77	5.36	3.54	1.06	104.57	2.34	1.78	0.88	6.72
3	5.27	7.00	70.53	1.07	7.93	4.74	7.15	1.02	105.34	2.39	1.79	1.03	5.62

Cluster 0 is characterized by moderately high academic performance (79.31) and average smartphone usage (4.98 hours per day), yet this group exhibits elevated psychological distress. Specifically, members of this cluster report high anxiety (7.57) and depression levels (6.52), alongside lower self-esteem (4.34) and slightly reduced sleep duration (6.03 hours). Despite not being heavy digital users, the emotional strain and sleep reduction suggest possible academic or personal stressors affecting their well-being. Cluster 1 demonstrates a relatively healthy behavioral profile. This group maintains balanced screen time (5.05 hours) and moderate engagement across social media, gaming, and education-related use. They report the lowest levels of anxiety (3.11) and depression (5.13), and notably high self-esteem (7.26). Additionally, sleep duration (6.84 hours) and physical activity (0.87 hours of exercise) are relatively adequate. These individuals appear emotionally stable and digitally regulated, suggesting a well-managed lifestyle.

Cluster 2 represents adolescents with high frequency of phone checking behavior (104.57 times per day) despite not exhibiting excessive overall screen time (4.80 hours). The group demonstrates lower self-esteem (3.54), moderate academic achievement (73.90), and average scores for anxiety (3.77) and depression (5.36). Their digital habits may reflect compulsive checking behavior or social disengagement, indicating possible latent emotional challenges. Finally, Cluster 3 contains users with the highest total screen time (5.27 hours), gaming usage (1.79 hours), and the most frequent phone checking (105.34 times per day). Although this group shows high anxiety (7.93), it simultaneously reports high self-esteem (7.15), forming a paradox between outward confidence and internal distress. Interestingly, they achieve the lowest academic scores (70.53), and although they sleep the most (7.00 hours), the quality of rest remains questionable due to high digital engagement, particularly before bedtime. To enhance the comparative interpretation of behavioral patterns, a radar chart was constructed to illustrate the normalized values of key indicators across the four clusters (Figure 4). This visualization allows for intuitive cross-dimensional comparison involving smartphone usage, mental health, academic performance, and lifestyle behaviors.

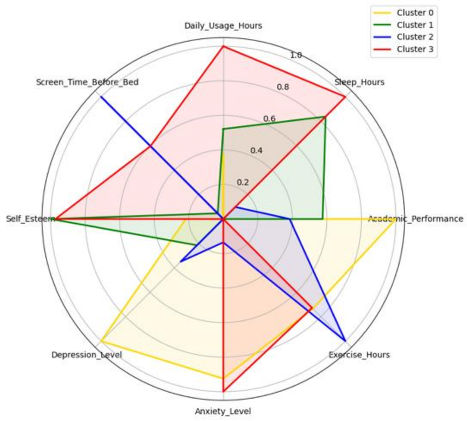

Figure 4 Radar chart

ISSN 2579-7069

These radar patterns reinforce the distinct behavioral orientations of each group and support earlier interpretations drawn from cluster centroids and statistical comparisons. These behavioral profiles provide valuable insights into the digital lifestyles of adolescents, with clear implications for mental health support, academic planning, and digital wellness interventions.

4.4. Cross Cluster Comparasion

To further analyze the behavioral patterns, selected features were compared across clusters using bar charts. These include academic performance, anxiety, depression, self-esteem, daily phone usage, and phone checking frequency.

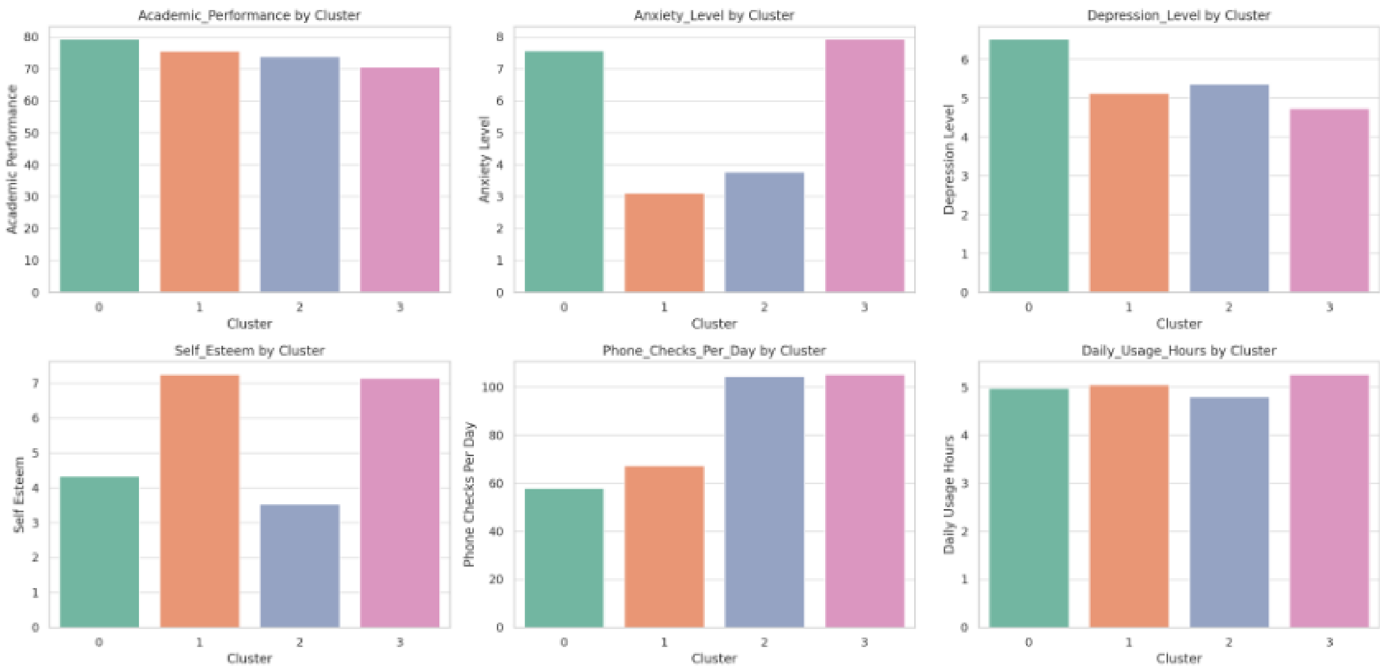

Figure 5. Cluster-wise comparison of behavioral metrics such as academic performance, anxiety, depression, self-esteem, phone check frequency, and daily smartphone usage among adolescents

The comparison reveals that Cluster 0 shows the highest academic performance, but also elevated levels of anxiety and depression. Cluster 1 stands out with the highest self-esteem and lowest anxiety, indicating a balanced psychological and digital lifestyle. On the other hand, Cluster 2 shows low self-esteem and high phone checking behavior, which may reflect compulsive patterns. Meanwhile, Cluster 3 combines heavy phone usage and checking frequency with lower academic scores, yet reports surprisingly high self-esteem, suggesting a disconnect between behavioral indicators and perceived self-image. These visual comparisons support the descriptive interpretation of each cluster and highlight potential risk patterns that may require digital literacy interventions or psychological support in adolescents.

4.5. Implications and Insights

The clustering results provide meaningful insights into adolescent smartphone behavior and its correlation with psychological and academic indicators. Each cluster revealed distinct patterns that could inform interventions in education, mental health, and digital well-being. For instance, Cluster 0, characterized by high academic achievement but elevated anxiety and depression, underscores the possibility that academic excellence may come at the cost of emotional well-being. This suggests a need for balanced academic expectations and mental health support systems even for high-performing students. Cluster 1 represents the most balanced group showing healthy digital habits, high self-esteem, and low psychological distress. This cluster could serve as a benchmark for digital wellness, offering behavioral targets for digital literacy and self-regulation programs. It also illustrates that moderate smartphone usage, combined with adequate sleep and exercise, may contribute to emotional resilience.

 ISSN 2579-7069

On the other hand, Cluster 2 shows high frequency in phone checking and low self-esteem despite average screen time, which may indicate habitual checking behavior or social anxiety. This highlights the importance of addressing screen compulsivity and building digital self-control. Cluster 3 combines high screen time and phone checking frequency with the lowest academic performance and highest anxiety. While self-esteem appears high, this could reflect overconfidence or compensatory behavior rather than actual emotional stability. This group may require the most comprehensive intervention, focusing on digital detox strategies, emotional regulation, and academic support.

From a policy and educational standpoint, these findings emphasize that not all screen time is equal, and the context and behavior behind smartphone use are critical. Tailored interventions can be designed based on behavioral profiles such as encouraging time limits for specific apps, integrating mindfulness training in schools, or supporting parents and educators in monitoring usage patterns without over-surveillance. Moreover, these patterns could inform the development of personalized digital wellness tools and early warning systems to detect at-risk behaviors before they escalate into more serious psychological or academic consequences. The table below summarizes the key characteristics, associated risks, and recommended interventions for each cluster, offering a practical reference for stakeholders in education, mental health, and digital policy

Table 3. Summary of Characteristics, Risks, and Recommended Interventions for Each Adolescent Cluster

Cluster	Key Characteristics	Potential Risks	Recommended Interventions
Cluster 0	High academic performance, but elevated anxiety and depression, low self-esteem.	Academic pressure, emotional burnout risk.	Mental health support for high achievers.
Cluster 1	Balanced digital behavior, high self-esteem, low psychological distress.	Low risk.	Model group for digital literacy and healthy habit promotion.
Cluster 2	Very high phone checking, low self-esteem, moderate screen time.	Compulsive usage, social anxiety.	Digital self-regulation and awareness training.
Cluster 3	Highest screen time and phone checking, lowest academic performance, high anxiety.	Imbalance in digital and psychological well-being, potential overconfidence.	Digital detox, emotional regulation, academic support.

5. Conclusion

This study successfully identified four distinct behavioral profiles of adolescents in their smartphone usage using the K-Means clustering algorithm. The data revealed meaningful groupings: high-achieving but anxious adolescents, balanced and mentally healthy users, individuals with low self-esteem and social withdrawal tendencies, and heavy smartphone users who exhibit emotional distress and poor academic outcomes. The findings highlight that adolescents do not use smartphones in the same way each group shows different usage patterns closely linked to sleep quality, mental health, and academic performance. These insights can be used to design more targeted educational and psychological interventions. Teachers, parents, and school counselors can leverage this information to identify at-risk youth, provide appropriate digital literacy education, and implement supportive programs such as counseling or screen-time management. From a methodological perspective, the study demonstrates that data mining techniques like K-Means are highly effective in uncovering behavioral patterns not only in commercial or technical domains but also in psychological and educational contexts. Future research may explore these patterns longitudinally, evaluate real-world intervention strategies based on cluster types, and incorporate other influential factors such as family background, gender, or social conditions.

ISSN 2579-7069

6. Declarations

6.1. Author Contributions

Author Contributions: Conceptualization, D.D.A. and B.S.; Methodology, D.D.A. and B.S.; Software, D.D.A.; Validation, B.S.; Formal Analysis, D.D.A.; Investigation, D.D.A. and B.S.; Resources, B.S.; Data Curation, D.D.A.; Writing—Original Draft Preparation, D.D.A.; Writing—Review and Editing, B.S.; Visualization, D.D.A. All authors have read and agreed to the published version of the manuscript.

6.2. Data Availability Statement

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

6.3. Funding

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

6.4. Institutional Review Board Statement

Not applicable.

6.5. Informed Consent Statement

Not applicable.

6.6. 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] K. J. Y. L. H. Boon Yew Wong1, "A systematic review on relationship between stress and problematic smartphone use," *International Journal of Public Health Science (IJPHS),* vol. 11, no. 2252, pp. 1133 - 1156, 2022.

[2] P. R. ,. B. W. ,. N. J. K. a. B. C. Sei Yon Sohn, "Prevalence of problematic smartphone usage and associated mental health outcomes amongst children and young people : a systematic review , meta - analysis and Grade of the evidence," *BMC Psychiatry,* vol. 19 (1), no. 356, pp. 1-10, 2020.

[3] J. P. H. T. Jiawei Han, " Fourth Edition," in *Data Mining Concepts and Techniques*, Cambridge, United States, Elsevier, 2023, p. 386.

[4] L. R. Oded Maimon, "Chapter 1," in *Introduction to Knowledge Discovery in Databases*, Tel Aviv, ResearchGate, 2005, p. 3.

[5] J. A. J. W. C. H. Sebastian Hökby, "Adolescents' screen time displaces multiple sleep pathways and elevates depressive symptoms over twelve months," *PLOS GLOBAL PUBLIC HEALTH,* vol. 5, no. 4, pp. 1-21, 2025.

[6] S. A. H. H. H. ,. M. A. N. M. F. F. N. A. A. A. R. Nur Zakiah Mohd Saat, "Relationship of screen time with anxiety, depression, and sleep quality among adolescents: a cross-sectional study," *Frontiers in Public Health,* vol. 12, pp. 1-10, 2024.

[7] M. M. J. Y. ,. E. Shin, "Online media consumption and depression in young people: A systematic review and meta-analysis," *Computers in Human Behavior,* vol. 128, no. 6, pp. 1-12, 2021.

[8] B. V. C. H. S.R. Weerasinghe, "The Extent of Smartphone Addiction and Its Impact on Educational Outcomes Among Sri Lankan Advanced Level Students in the Adolescence Stage," *Journal of Information Systems Engineering and Management,* vol. 10, no. 41s, p. 586, 2025.

[9] L.-C. W. *. a. C. M. Quaiser-Pohl, "Does the Type of Smartphone Usage Behavior Influence Problematic Smartphone Use and the Related Stress Perception?," *Behavioral Sciences,* vol. 12, no. 4, pp. 1-13, 2022.

[10] E. Xiao, "Comprehensive K-Means Clustering," *Journal of Computer and Communications,* vol. 12, no. 3, pp. 146 - 159, 2024.

[11] G. A. T. Gbeminiyi John Oyewole, “Data clustering: application and trends,” *Artifcial Intelligence Review,* vol. 56, p. 6439–6475, 2022.

[12] N. A. A. M. *. a. S. S. Nur Izzati Mohd Talib, “Identification of Student Behavioral Patterns in Higher Education Using K-Means Clustering and Support Vector Machine,” *applied sciences,* vol. 13, no. 5, pp. 1-14, 2023.

[13] N. S. S. N. F. A. Z. H. A. R. M. A. Ahmad Fikri Mohamed Nafuri, “Clustering Analysis for Classifying Student Academic Performance in Higher Education,” *applied sciences,* vol. 12, no. 19, pp. 1-22, 2022.

[14] Z. Xu, “College Students’ Mental Health Support Based on FuzzyClustering Algorithm,” *Contrast Media & Molecular Imaging,* vol. 2022, pp. 1-9, 2022.

[15] B. K. K. M. S. R. D. S. M A Syakur, “Integration K-Means Clustering Method and Elbow Method For Identification of The Best Customer Profile Cluster,” *IOP Conference Series: Materials Science and Engineering,* vol. 336, pp. 1-7, 2018.

[16] G. P.-S. a. P. S. Usama Fayyad, “From Data Mining to Knowledge Discovery in Databases,” *AI Magazine,* vol. 17, no. 3, pp. 1-18, 1996.

[17] X. L. ,. A. Y. Y. YU NIE, “A Data-Driven Knowledge Discovery Framework for Smart Education Management Using Behavioral Characteristics,” *IEEE Access,* vol. 11, pp. 1-13, 2023.

[18] Y. Y. Xiaoling Shu, “Knowledge Discovery: Methods from data mining and machine learning,” *Social Science Research,* vol. 110, pp. 1-16, 2022.

[19] a. B. V. ASHISH P. JOSHI, “Data Preprocessing: the Techniques for Preparing Clean and Quality Data for Data Analytics Process,” *Oriental Journal of Computer Science and Technology,* vol. 13, no. 2, pp. 1-5, 2020.

[20] Y. Liu, “Analysis and Prediction of College Students’ Mental Health Based on K-means Clustering Algorithm,” *Applied Mathematics and Nonlinear Sciences,* vol. 7, no. 1, pp. 1-12, 2021.