Journal of Artificial Intelligence
and Legal Technology

SAH PUBLISHER

Vol. 1, No. 1, August 2025

An Analysis of the Impact of Social Media Addiction on Students Academic Performance Using K-Means and Decision Tree

Dewi Sayekti Sutrisni [1]*, Maulana Ilham Alisyahbana[2], Muhammad Luqman Al-hakim[3], Deny Prasetyo[4]

[1,2,3,4]Department of Computer Science, Universitas Sugeng Hartono, Sukoharjo, Indonesia

Abstract. This study examines the impact of social media addiction on students' academic satisfaction using a machine learning approach. With the rapid growth of social media usage among university students, concerns have emerged regarding its potential influence on academic performance and psychological well-being. The dataset used in this research was obtained from the "Social Media Addiction vs Relationships" dataset on Kaggle, consisting of behavioral indicators such as daily usage hours, addiction scores, interpersonal conflicts, sleep duration, mental health scores, and perceived academic impact. Two analytical techniques were implemented: K-Means clustering to segment students based on behavioral patterns and Decision Tree classification to predict academic satisfaction levels. The clustering results identified three distinct behavioral groups representing low, moderate, and high addiction risk profiles. The Decision Tree model achieved 100% accuracy on the testing dataset, identifying social media-related conflicts and addiction scores as the most influential predictors of academic dissatisfaction. Correlation analysis further revealed strong relationships between addiction levels, reduced sleep duration, and decreased mental health scores, indicating a cascading behavioral effect. The findings suggest that academic dissatisfaction is influenced not only by usage duration but also by multidimensional psychological and relational factors. This study provides a data-driven framework for early identification of at-risk students and supports the development of targeted digital well-being intervention strategies in higher education institutions.

Keywords: Social Media; Addiction; Academic Ability; K-Means; Decision Tree
Received May 2025 / **Revised** August 2025 / **Accepted** August 2025

INTRODUCTION

The rapid advancement of information and communication technology has significantly transformed how individuals interact, communicate, and access information. Social media platforms have become an integral part of daily life, particularly among university students who rely on digital connectivity for academic, social, and professional purposes [1]. The increasing accessibility of smartphones and high-speed internet has accelerated this trend, making social media usage nearly ubiquitous within higher education environments [2]. While these platforms offer various benefits, concerns have emerged regarding their excessive use and its potential implications for students' academic performance.

Social media provides substantial advantages in academic contexts, including collaborative learning, knowledge sharing, peer communication, and access to educational resources [5], [6]. Platforms such as discussion forums, content-sharing applications, and professional networking sites enable students to exchange ideas and expand their learning beyond classroom boundaries [9]. In many cases, social media enhances engagement and supports informal learning processes [10]. However, the boundary between productive use and excessive consumption is often blurred, leading to patterns of behavior that may disrupt academic responsibilities.

Social media addiction is characterized by compulsive use, inability to control online activity, and psychological dependence on digital interaction [11]. Individuals experiencing addictive behaviors often spend prolonged hours online, neglect essential daily activities, and experience discomfort when disconnected. Previous studies have indicated that students who spend more than three to five hours daily

[1]*Corresponding author.
Email addresses: it.dewi.sayekti@gmail.com (Sutrisni)

on social media are more likely to develop problematic usage patterns [12]. These behaviors may result in reduced concentration, procrastination, sleep disturbances, and increased stress levels, all of which may negatively influence academic outcomes.

The relationship between excessive social media use and academic performance has attracted considerable scholarly attention. Several studies have reported a negative correlation between high usage intensity and grade performance, learning motivation, and academic satisfaction [12]. Frequent exposure to online distractions can reduce students' cognitive focus and time allocated for studying. Furthermore, conflicts arising from online interactions may contribute to emotional distress, thereby further affecting academic engagement and overall well-being.

Despite extensive research examining the psychological and behavioral aspects of social media addiction, many prior studies primarily rely on traditional statistical analyses such as correlation or regression methods. While these approaches are useful for identifying linear relationships, they may not fully capture complex, multidimensional behavioral patterns [13]. The interaction between addiction levels, interpersonal conflicts, sleep quality, mental health, and academic performance often involves nonlinear relationships that require more advanced analytical techniques.

Data mining and machine learning approaches offer powerful tools for uncovering hidden patterns within large datasets. Clustering algorithms can identify distinct behavioral segments among students, while classification models can predict academic outcomes based on digital behavior indicators [14]. These techniques allow researchers to move beyond simple association analysis toward predictive modeling and automated pattern recognition. By applying machine learning methods, it becomes possible to generate actionable insights that may support institutional interventions.

Among various machine learning techniques, K-Means clustering is widely recognized for its ability to group individuals based on similarity in behavioral characteristics [15]. This method is particularly suitable for identifying patterns of social media usage and academic impact across different student profiles. Meanwhile, Decision Tree classification provides an interpretable predictive model that illustrates how specific variables contribute to academic satisfaction or dissatisfaction. The transparency of Decision Tree models enables clearer understanding of dominant factors influencing academic outcomes.

Therefore, this study aims to analyze the impact of social media addiction on students' academic ability using K-Means clustering and Decision Tree classification. Specifically, the study seeks to (1) identify behavioral clusters of students based on social media usage patterns, (2) evaluate the predictive relationship between addiction indicators and academic satisfaction, and (3) provide data-driven recommendations for digital intervention strategies in higher education institutions. By integrating clustering and classification techniques, this research contributes to the development of an analytical framework for understanding digital behavior and its academic consequences.

METHODS

This study employed a quantitative research design using a data mining approach to analyze the relationship between social media addiction and students' academic performance. Two machine learning techniques were implemented: K-Means Clustering for behavioral segmentation and Decision Tree Classification for predictive modeling. The combination of unsupervised and supervised learning approaches allows both exploratory pattern discovery and outcome prediction within the same analytical framework.

Figure 1 illustrates the overall research workflow, including data preparation, exploratory data analysis, and the two analytical branches: clustering for behavioral segmentation and classification for predictive modeling.

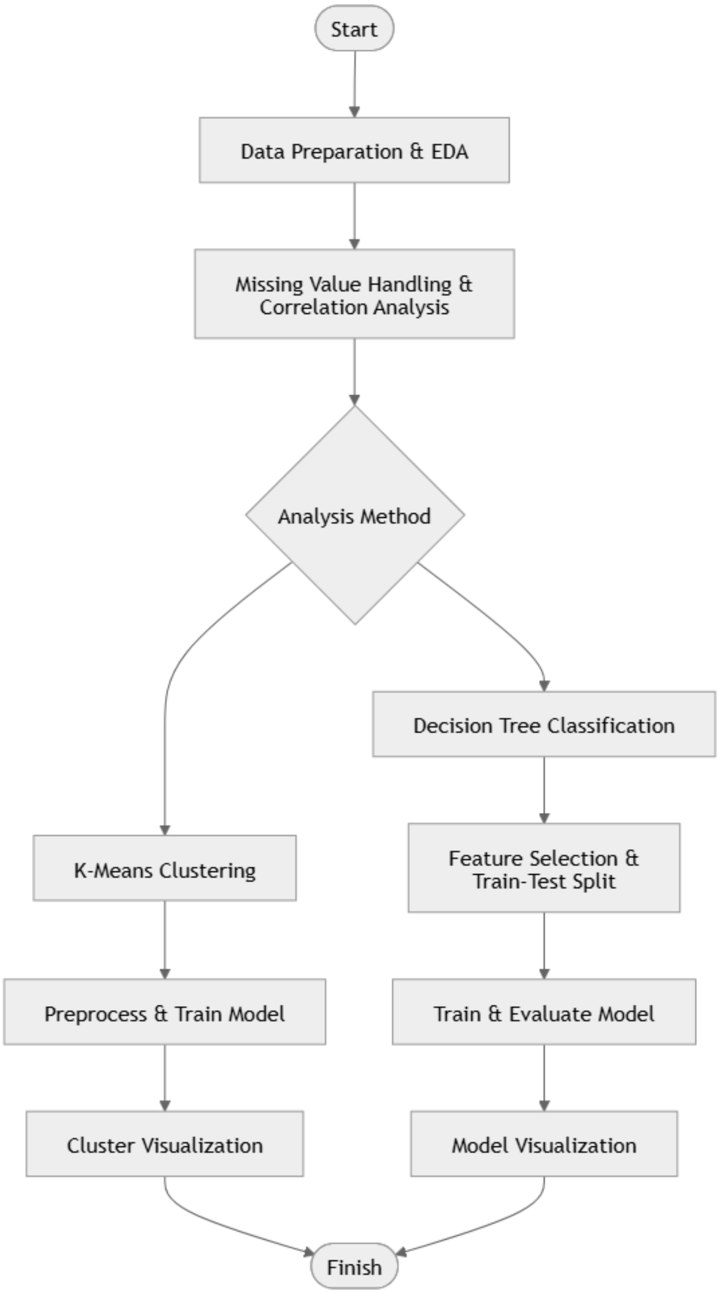

Figure 1. Research workflow integrating K-Means clustering and Decision Tree classification

The analysis was conducted using Python with libraries including pandas, numpy, matplotlib, seaborn, and scikit-learn. The workflow consisted of dataset acquisition, preprocessing, exploratory data analysis, clustering analysis, classification modeling, and performance evaluation.

Dataset

The dataset used in this study was the "Social Media Addiction vs Relationships" dataset obtained from Kaggle. The dataset contains behavioral and psychological indicators associated with social media usage intensity and its perceived academic impact. After the cleaning process, a total of 212 complete observations were retained for analysis.

The selected variables include daily usage duration, addiction scores, conflict frequency related to social media, sleep duration, mental health scores, and perceived academic impact. These variables represent measurable behavioral dimensions relevant to digital addiction and academic performance.

Table 1. Description of Research Variables

Variable Name	Description	Type
Avg_Daily_Usage_Hours	Average hours spent on social media per day	Numerical
Conflicts_Over_Social_Media	Frequency of interpersonal conflicts due to social media	Numerical
Addicted_Score	Quantitative measure of addiction level	Numerical
Sleep_Hours_Per_Night	Average sleep duration per night	Numerical
Mental_Health_Score	Self-reported mental well-being score	Numerical
Affects_Academic_Performance	Perceived academic satisfaction (target variable)	Categorical

Data Preprocessing

Data preprocessing was performed to ensure reliability and model readiness. Initially, the dataset was examined for missing values, duplicate entries, and inconsistent data types. Incomplete records were removed to maintain data integrity, resulting in a clean dataset suitable for machine learning analysis.

Feature selection was conducted based on theoretical relevance and prior literature. For clustering analysis, Avg_Daily_Usage_Hours and Conflicts_Over_Social_Media were selected to represent behavioral intensity and interpersonal impact. For classification analysis, additional predictors including Addicted_Score, Sleep_Hours_Per_Night, and Mental_Health_Score were incorporated.

Since K-Means is distance-based, feature scaling was applied using StandardScaler to normalize numerical variables. Standardization ensures that each variable contributes proportionally to Euclidean distance calculations and prevents bias caused by scale differences.

K-Means Clustering

K-Means clustering was applied to segment students into homogeneous behavioral groups. The algorithm partitions data into k clusters by minimizing within-cluster variance using Euclidean distance. The iterative process involves centroid initialization, assignment of data points to nearest centroids, centroid recalculation, and convergence when cluster assignments stabilize.

In this study, the number of clusters was set to three ($k = 3$) to represent low, moderate, and high levels of social media addiction and academic impact. This segmentation allows clearer behavioral categorization and practical interpretability.

Table 2. K-Means Parameter Configuration

Parameter	Value
Number of Clusters	3
Initialization	Random
Distance Metric	Euclidean
Scaling Method	StandardScaler
Maximum Iterations	Default (Scikit-learn)

Cluster results were visualized using scatter plots to illustrate group separation and centroid positioning.

Decision Tree Classification

Decision Tree classification was implemented to predict academic satisfaction levels based on behavioral indicators. The dataset was divided into training and testing subsets using an 80:20 split ratio to evaluate generalization performance.

The Decision Tree algorithm constructs a hierarchical structure by recursively splitting data according to features that maximize information gain. In this study, Gini impurity was used as the splitting criterion. The model generates interpretable decision rules that identify dominant predictors influencing academic outcomes.

Table 3. Decision Tree Model Configuration

Parameter	Value
Split Ratio	80% Training – 20% Testing
Splitting Criterion	Gini Impurity
Maximum Depth	Default
Feature Selection	All selected predictors

Feature importance scores were extracted to determine the relative contribution of each variable in classification decisions.

Model Evaluation

Model performance was assessed using confusion matrix analysis and standard classification metrics. The confusion matrix provides detailed classification outcomes including True Positives, True Negatives, False Positives, and False Negatives.

The following evaluation metrics were computed: Accuracy, Precision, Recall, F1-score, Macro and Weighted Averages

Table 4. Evaluation Metrics Formula

Metric	Formula
Accuracy	(TP + TN) / Total Samples
Precision	TP / (TP + FP)
Recall	TP / (TP + FN)
F1-Score	2 × (Precision × Recall) / (Precision + Recall)

The Decision Tree model achieved 100% accuracy on the test dataset. While this indicates strong classification performance, further validation such as cross-validation is recommended to assess model robustness and minimize overfitting risk.

Overall, this methodological framework integrates behavioral segmentation through clustering and predictive modeling through classification, enabling comprehensive analysis of social media addiction's impact on academic performance.

RESULT AND DISCUSSION

This section elaborates the analytical findings in greater depth by integrating clustering interpretation, predictive modeling analysis, statistical relationships, behavioral theory alignment, methodological reflection, and institutional implications. The objective is not merely to report model performance, but to interpret the structural behavioral mechanisms underlying social media addiction and its academic consequences.

The implementation of K-Means clustering produced three well-defined behavioral clusters based on standardized values of Avg_Daily_Usage_Hours and Conflicts_Over_Social_Media. The visualization in Figure 2 demonstrates distinct spatial separation among clusters, indicating that student behavior patterns are not randomly distributed but form structured digital profiles, as seen in Figure 1..

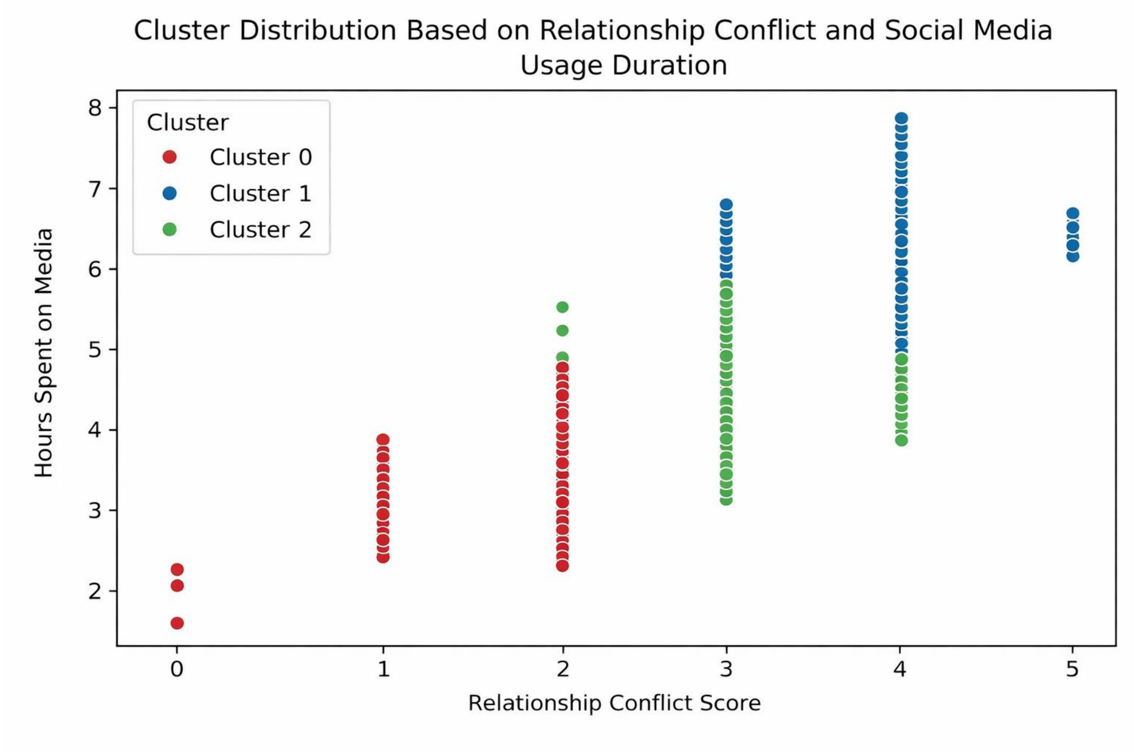

Figure 2. Clusters based on relationship conflict scores and hours spent on social media

A deeper examination of centroid positions indicates clear differences in behavioral intensity across clusters. Cluster 0 (low-risk) shows controlled daily usage and minimal conflicts, reflecting balanced digital engagement that does not significantly disrupt academic activities. Cluster 1 (moderate-risk) represents an intermediate stage, suggesting early behavioral escalation and vulnerability to higher-risk patterns. Cluster 2 (high-risk) exhibits high usage hours and frequent interpersonal conflicts, indicating behavioral dysregulation. The clustering pattern forms a progressive continuum, where increased usage corresponds with rising conflict frequency, supporting the view of social media addiction as a gradient phenomenon rather than a binary condition. These clusters also reflect varying levels of academic risk, with high-risk students showing characteristics associated with academic disengagement, such as emotional instability and excessive screen time.

The moderate cluster is particularly significant for preventive policy design. While not yet exhibiting extreme behavioral characteristics, this group demonstrates tendencies that could evolve into problematic patterns. Early identification of this transitional cluster enables targeted awareness programs before academic performance declines substantially. The low-risk cluster represents a digitally adaptive group. Rather than abstaining from social media, these students appear to regulate usage effectively. This suggests that the academic impact of social media is not inherently negative but depends on behavioral regulation and contextual balance. Such stratified insight strengthens institutional decision-making. Instead of general restrictions, universities may implement differentiated digital well-being interventions aligned with cluster profiles.

The Decision Tree classifier, as seen in Figure 3, provides further analytical depth by identifying hierarchical decision rules that determine academic satisfaction. The model achieved 100% accuracy on the testing dataset, correctly classifying all instances.

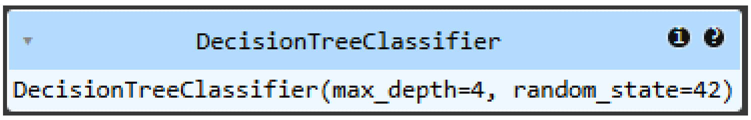

Figure 3. Decision Tree Classifier

The structure of the tree reveals that Conflicts_Over_Social_Media emerges as the primary splitting variable at the root node. This indicates that interpersonal disruption plays a more decisive role than usage duration alone. Such a finding suggests that the social consequences of digital engagement may mediate the relationship between addiction and academic dissatisfaction.

	Predicted Satisfied	Predicted Not Satisfied
Actual Satisfied	77	0
Actual Not Satisfied	0	135

Figure 4. Accuracy of Decision Tree

The confusion matrix demonstrates zero misclassification across both classes, resulting in perfect precision, recall, and F1-scores. While this performance is statistically impressive, it requires critical interpretation. Perfect classification may indicate strong separability within the dataset, but it may also signal potential overfitting due to homogeneous data patterns or limited variability.

Table 4. Classification Report of the Decision Tree Model

Class	Precision	Recall	F1-Score	Support
Satisfied	1.00	1.00	1.00	77
Not Satisfied	1.00	1.00	1.00	135
Accuracy			**1.00**	212
Macro Avg	1.00	1.00	1.00	212
Weighted Avg	1.00	1.00	1.00	212

To ensure model robustness, future studies should implement k-fold cross-validation and test the model on independent datasets. Nevertheless, within the current analytical scope, the model successfully captures dominant behavioral determinants. The visualization of the Decision Tree in Figure 5 offers interpretable decision pathways.

The tree indicates that students with high conflict frequency are directly categorized as academically dissatisfied. Even when conflicts are minimal, an Addicted_Score exceeding a certain threshold leads to dissatisfaction classification. This demonstrates a dual-trigger mechanism: relational disruption and internal addiction intensity independently predict academic discomfort.

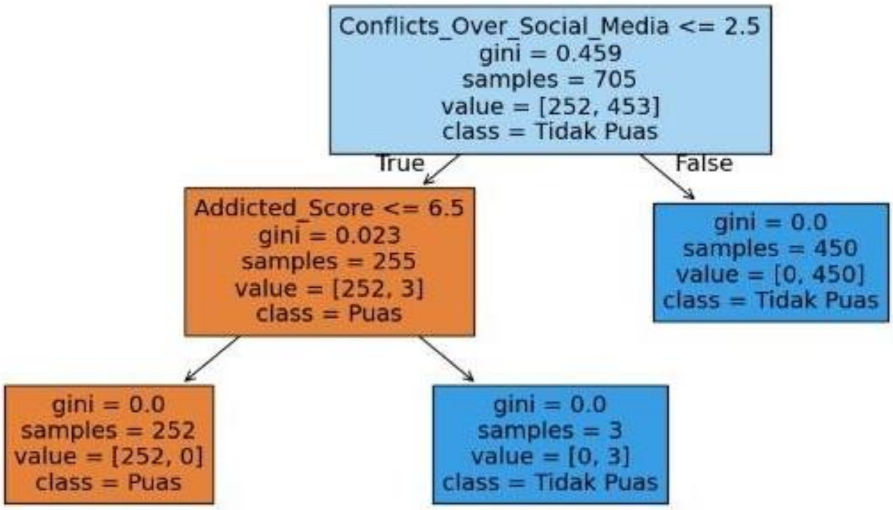

Figure 5. Visualization of the Decision Tree Model

This rule-based interpretation aligns with behavioral dependency theory, which posits that addiction reduces self-regulation capacity. Reduced self-control may manifest as procrastination, distraction, and emotional instability, ultimately affecting academic satisfaction. Interestingly, Sleep_Hours_Per_Night and Mental_Health_Score function as secondary but reinforcing predictors. Their placement deeper in the tree structure suggests that physiological and psychological deterioration operate as amplifiers of primary addiction effects. The Pearson correlation heatmap shown in Figure 6 reveals interdependencies among behavioral and psychological variables.

The strong positive correlation between Avg_Daily_Usage_Hours and Addicted_Score confirms that prolonged exposure reinforces dependency. The strong negative relationship between Addicted_Score and Mental_Health_Score suggests psychological vulnerability among highly addicted students. Similarly, reduced Sleep_Hours_Per_Night correlates with elevated addiction levels, indicating physical exhaustion as a secondary consequence. These correlations illustrate a cascading behavioral effect. Excessive usage increases addiction severity; addiction elevates conflict frequency and reduces sleep; reduced sleep and mental health decline diminish academic satisfaction. The structural interconnection supports a multidimensional impact model rather than a single-variable explanation.

The findings align with cognitive load theory and self-regulation theory. Excessive digital stimulation increases cognitive load, reducing available mental resources for academic processing. Simultaneously, addiction weakens self-regulatory capacity, impairing time management and task prioritization. Moreover, interpersonal conflict resulting from social media use may increase emotional stress, further diminishing cognitive efficiency. Emotional distraction competes with academic focus, leading to reduced academic satisfaction. Thus, academic dissatisfaction appears not as a direct product of usage duration, but as an outcome of compounded behavioral, psychological, and relational mechanisms.

The integration of clustering and classification provides actionable insights for higher education institutions. Universities may implement early detection systems based on addiction scores and conflict indicators. Counseling services could prioritize students identified within the high-risk cluster. Additionally, digital well-being education programs may emphasize conflict management, emotional regulation, and time-control strategies. Rather than imposing restrictive measures, institutions can promote balanced digital engagement supported by behavioral analytics.

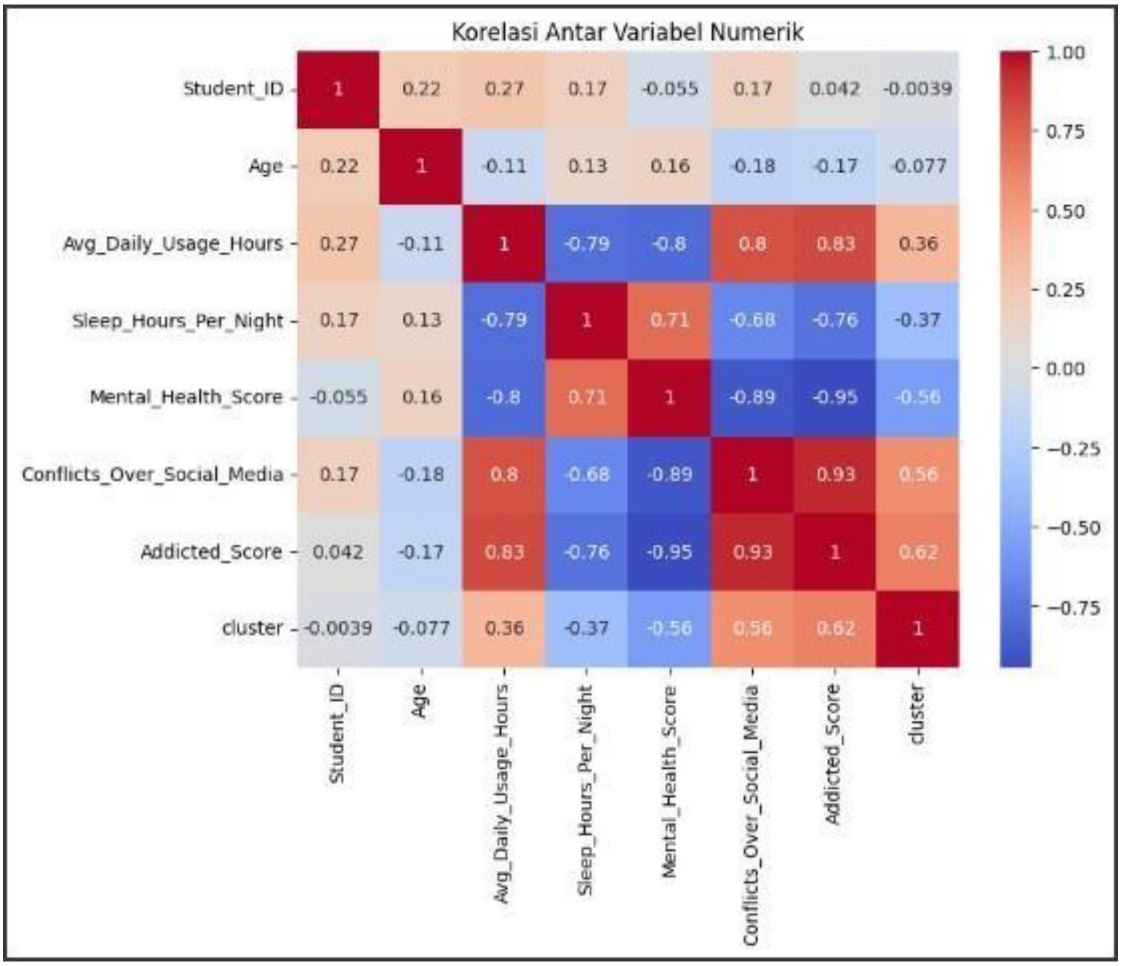

Figure 6. Heatmap of Correlations Among Numerical Variables

The predictive capability of the Decision Tree model also suggests potential integration into academic advisory systems. Real-time behavioral monitoring (with ethical considerations) could help prevent severe academic decline.

Despite strong analytical findings, several limitations must be acknowledged. First, the dataset size (212 observations) limits external generalization. Second, the perfect accuracy of the Decision Tree model raises the possibility of overfitting. Third, self-reported measures of addiction and academic satisfaction may introduce response bias. Future research should incorporate larger and more diverse datasets, apply ensemble methods such as Random Forest for comparison, and implement cross-validation techniques to ensure stability. Incorporating objective academic indicators such as GPA would further strengthen empirical validity.

Overall, the integration of clustering, classification, and correlation analysis reveals that social media addiction operates as a multidimensional behavioral construct. Academic dissatisfaction is influenced not merely by time spent online but by the intensity of addiction, relational conflict, psychological well-being, and sleep patterns. The analytical framework employed in this study demonstrates that machine learning techniques can effectively uncover hidden behavioral structures and predictive determinants within digital addiction research. By combining segmentation and classification, the study advances both theoretical understanding and practical intervention strategies in higher education contexts.

CONCLUSION

This study analyzed the impact of social media addiction on students' academic satisfaction using K-Means clustering and Decision Tree classification. The results indicate that students can be segmented into distinct behavioral groups based on usage intensity and conflict frequency, reflecting different levels of academic risk. Higher daily usage is consistently associated with increased interpersonal conflict and stronger addiction patterns. The Decision Tree model identified social media-related conflict and addiction score as the most influential predictors of academic dissatisfaction. Correlation analysis further revealed that addiction is closely linked to reduced sleep duration and lower mental health scores, suggesting a cascading behavioral effect that ultimately affects academic well-being. Overall, the findings confirm that social media addiction influences academic satisfaction through interconnected behavioral and psychological mechanisms. The integration of clustering and classification provides a data-driven framework that can support early identification and targeted intervention strategies in higher education institutions. Future research should apply larger datasets and validation techniques to strengthen generalizability and model robustness.

REFERENCES

[1] A. Arora, M. Chaudhary, R. Kaur, R. Batra, and P. Vats, "A Comprehensive Study on Social Network Analysis for Digital Platforms to Examine and Solve the Behavioral Patterns of Everyday Routines," in *ICT Systems and Sustainability*, vol. 516, M. Tuba, S. Akashe, and A. Joshi, Eds., in Lecture Notes in Networks and Systems, vol. 516. , Singapore: Springer Nature Singapore, 2023, pp. 13–21. doi: 10.1007/978-981-19-5221-0_2.

[2] L. Winstone, B. Mars, C. M. A. Haworth, and J. Kidger, "Types of Social Media Use and Digital Stress in Early Adolescence," *The Journal of Early Adolescence*, vol. 43, no. 3, pp. 294–319, Mar. 2023, doi: 10.1177/02724316221105560.

[3] C. Zhang, L. Tang, and Z. Liu, "How social media usage affects psychological and subjective well-being: testing a moderated mediation model," *BMC Psychol*, vol. 11, no. 1, p. 286, Sep. 2023, doi: 10.1186/s40359-023-01311-2.

[4] A. Pellegrino, A. Stasi, and V. Bhatiasevi, "Research trends in social media addiction and problematic social media use: A bibliometric analysis," *Front. Psychiatry*, vol. 13, p. 1017506, Nov. 2022, doi: 10.3389/fpsyt.2022.1017506.

[5] A. Sivakumar, S. Jayasingh, and S. Shaik, "Social Media Influence on Students' Knowledge Sharing and Learning: An Empirical Study," *Education Sciences*, vol. 13, no. 7, p. 745, Jul. 2023, doi: 10.3390/educsci13070745.

[6] M. Zamiri and A. Esmaeili, "Methods and Technologies for Supporting Knowledge Sharing within Learning Communities: A Systematic Literature Review," *Administrative Sciences*, vol. 14, no. 1, p. 17, Jan. 2024, doi: 10.3390/admsci14010017.

[7] P. Sahoo and J. S. Pradhan, "Exploring the Role of Social Media in Promoting Collaborative Learning: A Systematic Review," *J. Dig. Learning and Dist. Edu*, vol. 3, no. 12, pp. 1423–1435, May 2025, doi: 10.56778/jdlde.v3i12.502.

[8] N. M. Sabah, "The Impact of Social Media-Based Collaborative Learning Environments on Students' Use Outcomes in Higher Education," *International Journal of Human–Computer Interaction*, vol. 39, no. 3, pp. 667–689, Feb. 2023, doi: 10.1080/10447318.2022.2046921.

[9] U. Narang, M. S. Yadav, and A. Rindfleisch, "The 'Idea Advantage': How Content Sharing Strategies Impact Engagement in Online Learning Platforms," *Journal of Marketing Research*, vol. 59, no. 1, pp. 61–78, Feb. 2022, doi: 10.1177/00222437211017828.

[10] G. Tazhenova, N. Mikhaylova, and B. Turgunbayeva, "Digital media in informal learning activities," *Educ Inf Technol*, vol. 29, no. 16, pp. 21673–21690, Nov. 2024, doi: 10.1007/s10639-024-12687-y.

[11] M. Flayelle, D. Brevers, D. L. King, P. Maurage, J. C. Perales, and J. Billieux, "A taxonomy of technology design features that promote potentially addictive online behaviours," *Nat Rev Psychol*, vol. 2, no. 3, pp. 136–150, Feb. 2023, doi: 10.1038/s44159-023-00153-4.

[12] K. Chemnad, M. Aziz, S. B. Belhaouari, and R. Ali, "The interplay between social media use and problematic internet usage: Four behavioral patterns," *Heliyon*, vol. 9, no. 5, p. e15745, May 2023, doi: 10.1016/j.heliyon.2023.e15745.

[13] T. Kyriazos and M. Poga, "Application of Machine Learning Models in Social Sciences: Managing Nonlinear Relationships," *Encyclopedia*, vol. 4, no. 4, pp. 1790–1805, Nov. 2024, doi: 10.3390/encyclopedia4040118.

[14] J. Jin, "Student behavior patterns in vocational education big data based on clustering algorithm," *Discov Artif Intell*, vol. 5, no. 1, p. 197, Aug. 2025, doi: 10.1007/s44163-025-00433-3.

[15] S. M. Miraftabzadeh, C. G. Colombo, M. Longo, and F. Foiadelli, "K-Means and Alternative Clustering Methods in Modern Power Systems," *IEEE Access*, vol. 11, pp. 119596–119633, 2023, doi: 10.1109/ACCESS.2023.3327640.