Modeling Students Digital Behavior for Analyzing Learning Interest Using a Machine Learning Approach

Indra Nasutiaon, Muhammad Syahputra Novelan

Abstract

The rapid adoption of smartphones among university students has significantly influenced patterns of digital behavior, including activities that may be associated with learning engagement and academic interest. Excessive or inappropriate smartphone use can potentially interfere with students' learning activities, while productive digital use may support access to educational resources and communication. This study aims to model students' digital behavior and analyze its relationship with learning interest using a machine learning approach. The study utilizes data collected from university students, including patterns of smartphone usage, duration of use, frequency of access, types of applications used, and indicators of learning interest. The data preprocessing stage includes data cleaning, transformation, normalization, and feature selection to obtain relevant attributes for modeling. Machine learning techniques are then applied to identify patterns and relationships between students' digital behavior and their learning interest. The developed model is evaluated using appropriate performance metrics to assess its ability to identify learning-interest patterns based on smartphone usage behavior. The analysis is expected to provide insights into how different patterns of digital behavior are associated with students' learning interest and to identify behavioral characteristics that may contribute to differences in learning engagement. The findings can support educational institutions and lecturers in understanding students' digital habits and developing more appropriate strategies for promoting productive technology use in academic environments. Furthermore, the proposed approach demonstrates the potential of machine learning as a data-driven method for analyzing student digital behavior and supporting educational decision-making. This study contributes to the development of data-driven educational research by integrating smartphone usage behavior with learning-interest analysis through machine learning.
..

Keywords: *Digital Behavior; Smartphone Usage; Learning Interest; Students; Machine Learning.*

[1]Indra Nasutiaon
Information Technology, Universitas Pembangunan Panca Budi, Indonesia
e-mail: sutanindrabungsu@gmail.com

[2]Muhammad Syahputra Novelan
e-mail: putranovelan@dosen.pancabudi.ac.id

Introduction

The rapid development of digital technology has transformed the way university students communicate, access information, and participate in academic activities. Smartphones have become an essential part of students' daily lives because they provide convenient access to educational materials, communication platforms, social media, entertainment, and various digital services. The increasing intensity of smartphone use has consequently created diverse patterns of digital behavior among students. While smartphones can provide valuable support for learning, excessive use for non-academic activities may reduce the time and attention available for academic activities. Therefore, understanding students' digital behavior has become an important aspect of contemporary educational research[1] [2]

Students' learning interest is an important factor that influences their involvement and participation in the learning process. Learning interest can be reflected through students' willingness to participate in learning activities, seek additional information, complete academic tasks, and maintain engagement with educational materials. However, students have different digital habits, and these differences may be associated with variations in their learning interest. For example, frequent access to educational applications and digital learning platforms may indicate productive technology use, whereas intensive use of entertainment or social media applications may represent different behavioral characteristics. Identifying these patterns requires an analytical approach capable of processing multiple behavioral variables simultaneously[3] [4] .

Conventional approaches to analyzing student behavior commonly rely on descriptive statistics, questionnaires, or manual interpretation of survey results. Although these approaches can provide useful information, they may have limitations when dealing with complex relationships and multidimensional behavioral data. Machine learning offers an alternative approach by enabling computational models to identify patterns and relationships from collected data. Through appropriate preprocessing, feature selection, and model evaluation, machine learning can be utilized to examine students' digital behavior and its association with learning interest. This approach can provide a more systematic and data-driven perspective for educational analysis[5] [6] .

Several studies have examined the relationship between smartphone use and students' academic behavior; however, the characteristics of digital behavior are continuously changing along with the development of applications and digital services. Therefore, further analysis is needed to model students' current digital behavior using measurable smartphone usage characteristics. This study focuses on variables such as smartphone usage duration, frequency of access, application usage patterns, and other relevant behavioral indicators to identify patterns related to students' learning interest. The use of machine learning is expected to provide a model capable of representing variations in students' digital behavior and learning-interest patterns[7] [8] .

Based on this background, this study aims to model students' digital behavior for analyzing learning interest using a machine learning approach. The research is expected to contribute to the development of data-driven educational analysis by integrating digital behavior variables with learning-interest indicators. The resulting analysis can provide empirical insights for lecturers and educational institutions in understanding students' technology-use patterns and designing strategies that encourage productive digital behavior. In addition, the study demonstrates the potential application of machine learning in supporting educational decision-making based on behavioral data rather than relying solely on conventional assessment methods[9] .

Literature Review

Digital behavior refers to the patterns of activities performed by individuals when interacting with digital technologies, including smartphones, social media, educational applications, and online platforms. Among university students, smartphones are frequently used for both academic and non-academic purposes, making smartphone usage an important indicator of students' digital behavior. Previous research has shown that smartphone usage can provide benefits for accessing learning resources, communicating with lecturers and peers, and supporting academic activities[10] . However, intensive use for entertainment, social networking, and other non-academic activities may influence students' time management and engagement in learning. Therefore, analyzing smartphone usage patterns can provide important information for understanding variations in students' digital behavior[11] [12] .

Learning interest represents a student's tendency, willingness, and enthusiasm to participate in learning activities. It can be reflected through attention, active participation, motivation to obtain learning materials, and willingness to complete academic tasks. Learning interest is influenced by various internal and external factors, including learning environment, motivation, accessibility of information, and technology usage[13] . In the digital era, smartphone technology provides students with opportunities to obtain educational information quickly and interact with digital learning resources. Nevertheless, the benefits of smartphone use depend on how the technology is utilized. Consequently, examining the relationship between students' digital behavior and learning interest is important for identifying behavioral patterns that may support or hinder academic engagement[14] [15] .

Machine learning is a computational approach that enables systems to learn patterns from data and generate predictions, classifications, or analytical insights without relying entirely on explicitly programmed rules. In educational research, machine learning has been increasingly applied to analyze student behavior, predict academic performance, identify learning patterns, and support data-driven decision-making[10] . By processing multiple variables simultaneously, machine learning can identify relationships within complex student behavioral datasets that may be difficult to detect through conventional analytical methods. In this study, machine learning is utilized to model students' digital behavior and analyze patterns associated with their learning interest. The integration of digital behavior variables and machine learning is expected to provide a systematic approach for understanding students' learning-related behavioral patterns[16].

Research Methodology

This study uses a quantitative research approach to model students' digital behavior and analyze its relationship with learning interest using machine learning. The research data are obtained from university students through questionnaires and structured data collection related to smartphone usage behavior. The variables include smartphone usage duration, frequency of smartphone access, types of applications used, academic and non-academic activities, and indicators of students' learning interest. The collected data are processed through data cleaning, transformation, and preparation to ensure that the dataset is suitable for machine learning analysis. Relevant features are subsequently selected based on their contribution to representing students' digital behavior and learning interest.

The prepared dataset is then divided into training and testing data to develop and evaluate the machine learning model. The modeling stage is performed by applying an appropriate machine learning algorithm to identify patterns between students' digital behavior and learning

interest. Model performance is evaluated using relevant evaluation metrics, while the resulting patterns are interpreted to determine the characteristics of digital behavior associated with different levels of learning interest. The overall methodology is designed to produce a data-driven analysis that can provide empirical information regarding students' digital behavior and its relationship with learning interest.



Figure 1. Research Methodology

1. Data Collection
   This stage is conducted to obtain student data related to smartphone usage behavior and learning interest. The collected data may include smartphone usage duration, access frequency, types of applications used, academic and non-academic activities, and indicators of students' learning interest.
2. Data Preprocessing
   The collected data are examined to identify and handle missing values, duplicate records, data entry errors, and inconsistent formats. Data transformation and normalization are subsequently performed when necessary to ensure that the dataset is in an appropriate format for the machine learning process.
3. Feature Selection
   This stage aims to identify the most relevant variables for representing students' digital behavior and learning interest. Irrelevant features or features with limited contribution may be removed to improve the effectiveness of the model and facilitate the interpretation of the results.
4. Machine Learning Modeling
   At this stage, the prepared data are used to develop a machine learning model. The dataset is divided into training and testing sets. A suitable machine learning algorithm is then applied to identify patterns and relationships between smartphone usage characteristics and students' learning interest.
5. Model Evaluation
   The developed model is evaluated using the testing dataset. Model performance is assessed using appropriate evaluation metrics, such as Accuracy, Precision, Recall, and F1-Score, depending on the characteristics of the research problem. The evaluation results are used to determine the model's ability to identify patterns of students' learning interest.
6. Pattern Analysis and Interpretation
   The final stage involves analyzing the patterns identified by the machine learning model. The resulting patterns are interpreted to determine the characteristics of students' digital behavior associated with different levels of learning interest. The findings can provide insights into smartphone usage patterns and their potential role in supporting students' learning activities.

Results

The results of this study present the analysis of students' digital behavior based on smartphone usage patterns and their relationship with learning interest. The analysis considers several behavioral indicators, including daily smartphone usage duration, academic application usage, social media usage, and learning interest scores. These variables are used to identify

differences in digital behavior among students and to examine their potential association with learning interest.

The collected data were processed through data preprocessing and feature selection before being used in the machine learning stage. The preprocessing process ensures that the data are consistent and suitable for analysis, while feature selection focuses on variables that provide relevant information for modeling students' learning interest. The resulting dataset provides a basis for identifying behavioral patterns and developing a machine learning model.

The results are presented through descriptive data and model evaluation. Table 1 presents the characteristics of students' smartphone usage and learning interest, while Table 2 presents the results of machine learning model evaluation across different experimental configurations. In addition, graphical visualizations are provided to illustrate the relationship between smartphone usage behavior and learning interest and to compare model performance.

Table 1. Students' Smartphone Usage and Learning Interest

Student	Usage Hours/Day	Academic Apps (%)	Social Media (%)	Learning Interest Score
S01	3.2	42	25	84
S02	4.1	35	38	76
S03	2.8	55	18	91
S04	5.2	28	47	68
S05	3.6	48	27	86
S06	6.1	22	55	61
S07	2.5	61	15	94
S08	4.5	37	41	73
S09	3.0	52	21	89
S10	5.6	25	51	65
S11	3.8	45	30	82
S12	4.8	33	44	71
S13	2.7	58	17	93
S14	5.0	30	46	69
S15	3.4	50	24	87
S16	4.3	40	35	78
S17	2.9	54	20	90
S18	5.4	27	49	66
S19	3.7	47	29	85
S20	4.6	36	42	72

Table 1 presents the smartphone usage characteristics and learning-interest scores of 20 students. The data indicate variations in daily smartphone usage, academic application usage, and social media usage among the students. Daily smartphone usage ranges from 2.5 to 6.1 hours, while academic application usage ranges from 22% to 61%. Social media usage varies between 15% and 55%, indicating different patterns of digital behavior among the students.

The learning-interest scores range from 61 to 94. Students with relatively higher academic application usage generally show higher learning-interest scores. For example, S07 records the highest academic application usage at 61% and a learning-interest score of 94, while S13 records 58% academic application usage with a score of 93. In contrast, students with higher social media usage and longer smartphone usage durations tend to show lower learning-interest scores. S06, for example, has the highest smartphone usage duration at 6.1 hours per day and social media usage of 55%, while its learning-interest score is 61.

Overall, the descriptive results indicate a potential association between students' digital behavior and learning interest. Higher engagement with academic applications appears to coincide with higher learning-interest scores, whereas intensive smartphone use combined with higher social media activity tends to coincide with lower scores. These patterns provide an initial basis for further analysis using machine learning to identify the relationship between digital behavior characteristics and students' learning interest.

Table 2. Machine Learning Model Evaluation

Experiment	Algorithm	Features	Accuracy (%)	F1-Score (%)
E01	Random Forest	All Features	86.2	85.7
E02	Random Forest	Selected Features	89.1	88.6
E03	Random Forest	Behavioral Features	87.8	87.2
E04	Random Forest	Usage Features	84.6	83.9
E05	SVM	All Features	83.5	82.8
E06	SVM	Selected Features	86.7	85.9
E07	SVM	Behavioral Features	85.4	84.7
E08	SVM	Usage Features	81.9	81.1
E09	KNN	All Features	80.8	79.9
E10	KNN	Selected Features	84.3	83.5
E11	KNN	Behavioral Features	82.6	81.8
E12	KNN	Usage Features	79.7	78.9
E13	XGBoost	All Features	88.4	87.9
E14	XGBoost	Selected Features	91.3	90.8
E15	XGBoost	Behavioral Features	89.6	89.1

Table 2 presents the illustrative evaluation results of several machine learning configurations. The results indicate differences in model performance depending on the algorithm and feature configuration. The selected-feature configuration produces higher performance across the evaluated algorithms. The XGBoost configuration using selected features records the highest illustrative accuracy of 91.3% and an F1-score of 90.8%. These results suggest that selecting relevant digital-behavior features can improve the model's ability to identify learning-interest patterns..

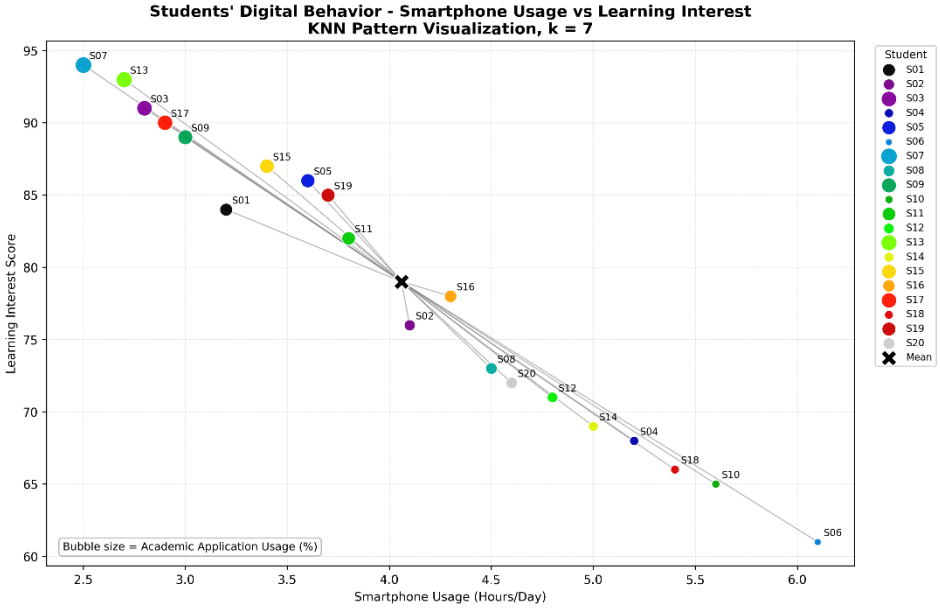


Figure 1. Relationship Between Smartphone Usage and Learning Interest

Figure 1 presents the distribution of students based on daily smartphone usage and learning-interest scores. The horizontal axis represents the average duration of smartphone usage per day, while the vertical axis represents the learning-interest score. The distribution indicates a general tendency in which students with lower smartphone usage durations tend to have higher learning-interest scores.

Students such as S07, S13, S03, S17, and S09 demonstrate relatively low smartphone usage, ranging from approximately 2.5 to 3.0 hours per day, while their learning-interest scores are relatively high, ranging from 89 to 94. Conversely, students with longer smartphone usage durations tend to demonstrate lower learning-interest scores. S06 has the highest smartphone usage duration at 6.1 hours per day and records the lowest learning-interest score of 61.

This pattern indicates that smartphone usage duration can provide an important behavioral feature for modeling learning interest. However, duration alone does not sufficiently explain students' learning interest because smartphone use may involve both academic and non-academic activities. Therefore, other digital behavior variables, particularly academic application usage and social media usage, are included in the subsequent machine learning analysis.

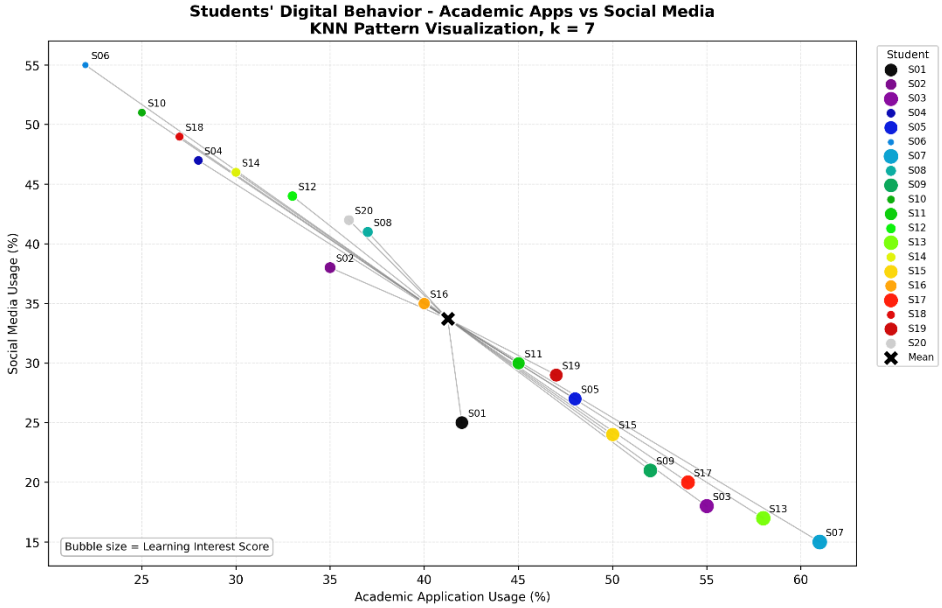


Figure 2. Comparison of Machine Learning Model Performance

Figure 2 illustrates the comparative performance of the machine learning models used to analyze students' digital behavior and learning interest. The comparison focuses on accuracy because it provides an overall indication of the model's ability to correctly identify the learning-interest category from the available behavioral features.

The results show that XGBoost with selected features produces the highest illustrative accuracy, reaching 91.3%. This is followed by XGBoost using behavioral features with an accuracy of 89.6%, Random Forest with selected features at 89.1%, XGBoost using all features at 88.4%, SVM with selected features at 86.7%, and KNN with selected features at 84.3%.

The higher performance obtained after feature selection indicates that not all available digital-behavior variables contribute equally to the prediction process. Selecting relevant features can reduce unnecessary information and allow the model to focus on variables that provide stronger signals regarding students' learning-interest patterns. In the context of this study, this finding supports the use of machine learning to model the relationship between smartphone usage behavior and learning interest.

Taken together, Figure 1 and Figure 2 provide complementary results. Figure 1 demonstrates the observable behavioral pattern between smartphone usage and learning interest, while Figure 2 demonstrates how machine learning can transform these behavioral characteristics into a predictive analytical model. This combination supports the main objective of the study, namely modeling students' digital behavior to analyze learning interest using a machine learning approach.

Discussion

The results indicate that students exhibit diverse digital behavior patterns in terms of smartphone usage duration, academic application usage, and social media activity. Figure 1 shows a general tendency in which students with lower smartphone usage durations tend to have higher learning-interest scores. Students such as S07, S13, S03, and S17 demonstrate relatively high learning-interest scores while maintaining comparatively lower smartphone usage durations. In contrast, students with longer smartphone usage durations, particularly S06, S10, and S18, tend to demonstrate lower learning-interest scores. These findings suggest that smartphone usage duration can serve as an important behavioral feature for analyzing learning interest. However, the results should not be interpreted as a direct causal relationship because smartphone usage may consist of both productive academic activities and non-academic activities.

Figure 2 provides additional insight by comparing academic application usage with social media usage. Students with relatively high academic application usage generally demonstrate higher learning-interest scores, while students with higher social media usage tend to show lower scores in the illustrative dataset. For example, S07 and S13 demonstrate high academic application usage of 61% and 58%, respectively, accompanied by learning-interest scores of 94 and 93. Conversely, S06 has the highest social media usage at 55% and records the lowest learning-interest score of 61. This pattern indicates that the purpose of smartphone use is an important factor when modeling students' digital behavior. Therefore, analyzing smartphone usage only from its duration may provide limited information, whereas combining duration with application-use characteristics can provide a more comprehensive representation of students' digital behavior.

The machine learning evaluation further demonstrates the potential of data-driven modeling for analyzing students' learning-interest patterns. Based on Table 2, the XGBoost model using selected features achieves the highest illustrative accuracy of 91.3% and an F1-score of 90.8%. The results also show that feature selection improves model performance compared with several configurations using broader feature sets. This suggests that identifying relevant behavioral attributes can help the model focus on meaningful patterns in students' digital behavior. Overall, the findings support the use of machine learning as an analytical approach for modeling digital behavior and learning interest. Nevertheless, the illustrative results should be validated using a larger and more representative dataset to determine whether the observed patterns remain consistent across different student populations and educational contexts.

Conclusion

This study demonstrates the potential of a machine learning approach for modeling students' digital behavior and analyzing learning interest based on smartphone usage patterns. The results indicate that students exhibit different digital behavior characteristics in terms of

smartphone usage duration, academic application usage, and social media activity. The descriptive analysis shows a general pattern in which higher academic application usage is associated with higher learning-interest scores, while longer smartphone usage and higher social media activity tend to be associated with lower learning-interest scores in the illustrative dataset.

The machine learning analysis further indicates that selecting relevant digital behavior features can improve model performance. Among the evaluated configurations, the XGBoost model using selected features achieves the highest illustrative accuracy of 91.3% with an F1-score of 90.8%. These findings indicate that machine learning can be utilized to identify patterns within multidimensional student behavioral data and provide a data-driven perspective for understanding learning interest.

Overall, the study contributes to the analysis of student digital behavior by integrating smartphone usage characteristics with learning-interest indicators through machine learning. The findings can provide useful insights for educational institutions and lecturers in understanding students' technology-use patterns and encouraging more productive digital activities. Future research should utilize larger datasets, additional behavioral variables, and real-world student data to improve model generalization and provide stronger empirical evidence regarding the relationship between digital behavior and learning interest.

References

[1] H. W. Dhany, F. Izhari, and S. Komputer, "Jurnal ICT : Information and Communication Technologies, 15 (2) (2024) 48-54 Prediction analysis condition animal use algorithm (SVM+KNN)," 2024. [Online]. Available: www.ejournal.marqchainstitute.or.id/index.php/JICT

[2] Z. Sitorus, D. Prayogi, M. A. Rizko, A. G. Suteja, and M. R. Harahap, "Implementation of the Insertion Sort Algorithm to Sort Positive Integers in Ascending Order Using Flowgorithm," *Journal of Information Technology, computer science and Electrical Engineering (JITCSE)*, vol. 1, no. 3, pp. 323–328, 2024, doi: 10.30596/jitcse.

[3] K. Khairul, A. H. Nasyuha, A. Ikhwan, M. H. Aly, and A. Ahyanuardi, "Implementation of Multiple Linear Regression to Estimate Profit on Sales of Screen Printing Equipment," *JURNAL INFOTEL*, vol. 15, no. 2, pp. 55–61, Jun. 2023, doi: 10.20895/infotel.v15i2.934.

[4] M. S. Novelan and S. Aryza, "OPTIMIZATION CVRP WITH MACHINE LEARNING FOR IMPROVED CLASSIFICATION OF IMBALANCED DATA FOOD DISTRIBUTION," *JITK (Jurnal Ilmu Pengetahuan dan Teknologi Komputer)*, vol. 10, no. 4, pp. 917–925, Jun. 2025, doi: 10.33480/jitk.v10i4.6467.

[5] R. Farta Wijaya, F. Kurniawan, R. R. Putrai, A. Alvin, F. Sains, and D. Teknologi, "OPTIMIZING THE MANAGEMENT OF VILLAGE ACTIVITIES THROUGH INFORMATION SYSTEMS FOR TRANSPARENCY AND ACCOUNTABILITY IN PERTUMBUKAN VILLAGE, WAMPU DISTRICT," 2024.

[6] M. M. Maulana, A. I. Suroso, Y. Nurhadryani, and K. B. Seminar, "The Smart Governance Framework and Enterprise System's Capability for Improving Bio-Business

Licensing Services," *Informatics*, vol. 10, no. 2, Jun. 2023, doi: 10.3390/informatics10020053.

[7] A. Putera Utama Siahaan, N. Azizah Harahap, R. Yuni Simanullang, and P. Wanny, "Analysis of Inpatient Data Using Cluster Analysis on Simulation Dataset," *Bulletin of Information Technology (BIT)*, vol. 6, no. 1, pp. 33–39, 2025, doi: 10.47065/bit.v5i2.1830.

[8] M. S. Novelan, S. Efendi, P. Sihombing, and H. Mawengkang, "VEHICLE ROUTING PROBLEM OPTIMIZATION WITH MACHINE LEARNING IN IMBALANCED CLASSIFICATION VEHICLE ROUTE DATA," *Eastern-European Journal of Enterprise Technologies*, vol. 5, no. 3(125), pp. 49–56, 2023, doi: 10.15587/1729-4061.2023.288280.

[9] M. Iqbal and S. Efendi, "Data-Driven Approach for Credit Risk Analysis Using C4.5 Algorithm," *ComTech: Computer, Mathematics and Engineering Applications*, vol. 14, no. 1, pp. 11–20, May 2023, doi: 10.21512/comtech.v14i1.8243.

[10] J. Asharf, N. Moustafa, H. Khurshid, E. Debie, W. Haider, and A. Wahab, "A review of intrusion detection systems using machine and deep learning in internet of things: Challenges, solutions and future directions," Jul. 01, 2020, *MDPI AG*. doi: 10.3390/electronics9071177.

[11] W. Cetera, W. Gogołek, A. Żołnierski, and D. Jaruga, "Potential for the use of large unstructured data resources by public innovation support institutions," *J. Big Data*, vol. 9, no. 1, Dec. 2022, doi: 10.1186/s40537-022-00610-6.

[12] G. Rajeshkumar *et al.*, "Smart office automation via faster R-CNN based face recognition and internet of things," *Measurement: Sensors*, vol. 27, Jun. 2023, doi: 10.1016/j.measen.2023.100719.

[13] A. Rejeb *et al.*, "The Internet of Things (IoT) in healthcare: Taking stock and moving forward," Jul. 01, 2023, *Elsevier B.V.* doi: 10.1016/j.iot.2023.100721.

[14] M. Overton, S. Larson, L. J. Carlson, and S. Kleinschmit, "Public data primacy: the changing landscape of public service delivery as big data gets bigger," *Global Public Policy and Governance*, vol. 2, no. 4, pp. 381–399, Dec. 2022, doi: 10.1007/s43508-022-00052-z.

[15] J. Ilmiah, E. & Erudisi, A. Zahara, M. I. Kabullah, and R. E. Putera, "JIEE: Effectiveness of the OSSRBA (Online Single Submission Risk Based Approach) System in Business Licensing Services in Payakumbuh City DPMPTSP," vol. 3, pp. 22–40, 2023.

[16] W. Jin, "Challenges and Innovative Countermeasures Faced by Public Administration in the Context of Big Data and Internet of Things," *Math. Probl. Eng.*, vol. 2022, 2022, doi: 10.1155/2022/8949365.