Journal of Telecommunication, Electronic and Computer Engineering
ISSN: 2180 – 1843 e-ISSN: 2289-8131 Vol. 17 No. 2 (2025) 31-36
jtec.utem.edu.my
DOI: https://doi.org/10.54554/jtec.2025.17.02.004

Optimized Random Forest Classifier for Students Lifestyle Prediction Using Behavioral Data: A Machine Learning Approach

Malakit L. Ram*, Jose C. Agoylo Jr.
Southern Leyte State University – Tomas Oppus Campus, Faculty of Computer Studies and Information Technology, Brgy. San Isidro, Tomas Oppus Southern Leyte, 6606, Philippines.

Article Info

Article history:
Received Aug 13th, 2024
Revised Mar 1st, 2025
Accepted Jun 16th, 2025
Published Jun 30th, 2025

Index Terms:
Machine Learning
Random Forest Classifier
Lifestyle Prediction
Behavioral Analytics

Abstract

Machine learning has increasingly been applied to behavioral analytics, yet its potential in lifestyle classification remains underexplored. This study utilizes a Random Forest classifier to predict lifestyle categories based on behavioral patterns from the Half a Million Lifestyle Dataset. A key challenge in lifestyle classification is balancing accuracy and generalization, which was addressed through parameter optimization to mitigate overfitting. To assess real-world applicability, 93 students provided behavioral inputs, which were processed through a Python-based program. The model successfully classified participants into Fitness Enthusiast (41), Health-Conscious (50), Eco-Friendly (1), and Social Media Influencer (1) categories, achieving an accuracy of 75.07%. These results confirm that machine learning can effectively predict lifestyle behaviors, with implications for personalized health interventions and behavioral analytics. This study underscores the significance of parameter tuning and feature selection, offering a scalable and data-driven approach to behavioral classification and wellness management.

*Corresponding Author: malakitr21@gmail.com

I. INTRODUCTION

The rapid advancements in machine learning and artificial intelligence have revolutionized numerous fields, including healthcare, education, and image processing. Among various machine learning techniques, the Random Forest Classifier has gained substantial attention for its robustness and accuracy in predictive modeling. This study aims to leverage the Random Forest Classifier to predict lifestyle categories based on behavioral patterns, utilizing a comprehensive dataset from Kaggle comprising half a million lifestyle entries. Despite these advancements, effective categorizing diverse lifestyle patterns remains challenging due to the complexity and high dimensionality of the data.

Previous studies have demonstrated the effectiveness of machine learning techniques, particularly Random Forest classifiers, across various domains. Researchers have employed Random Forest algorithms to predict academic performance by analyzing student behaviors and demographic factors, highlighting its reliability in classification tasks [1]-[3]. Additionally, studies have explored its application in identifying mental health issues among students, such as stress, anxiety, and depression, with high accuracy [4][5]. Other research has utilized machine learning techniques to classify student lifestyles based on their activities and behaviors [6][7], demonstrating the feasibility of AI-driven behavioral categorization. Despite these advancements, most existing studies focus on academic performance, psychological well-being, or specific behavior types, leaving a gap in comprehensive lifestyle classification using machine learning.

To address this gap, this research employs the Random Forest Classifier to analyze and predict lifestyle categories based on behavioral patterns from a large dataset sourced from Kaggle. By focusing on a wide range of behavior features, including gender, age, health consciousness rating, average daily screen time, social media influence, this study aims to uncover meaningful patterns and provide accurate lifestyle predictions. To further validate the model's predictive capability, the trained algorithm was tested on 93 students, successfully classifying them into distinct lifestyle categories with an accuracy of 75.07%. This research not only contributes to the academic understanding of machine learning applications in lifestyle prediction but also provides practical insights for real-world user classification and the promotion of healthier habits and behaviors. The findings from this study can be instrumental for policymakers, healthcare providers, and individuals seeking to improve lifestyle choices through data-driven insights.

II. FRAMEWORK OF THE STUDY

The framework illustrated in Figure 1 represents the causal relationship between behavioral patterns (independent variable) and lifestyle categories (dependent variable), resulting in predictive outcomes. This paradigm guides the study in assessing how behavioral characteristics influence lifestyle classification using the Random Forest Algorithm.

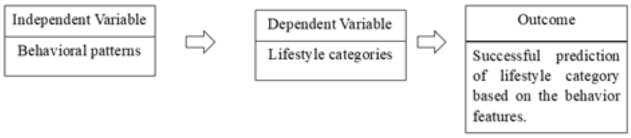

Figure 1. Causal Paradigm in Predicting Students' Lifestyle

Independent Variable: Behavioral Patterns
The behavioral traits of individuals serve as the primary input for the model. These include various features extracted from the dataset, such as:

- Health-conscious rating (self-reported awareness of personal health)
- Daily screen time (time spent using electronic devices)
- Social media influence (level of engagement on digital platforms)
- Exercise habits (frequency and duration of physical activity)
- Stress management (coping mechanisms and strategies)
- Eco-consciousness metric (concern for environmental sustainability)
- Time management skills (effectiveness in handling daily responsibilities)
- Participation in professional training (attendance in skill-building programs)

These factors collectively define an individual's behavioral profile, enabling the machine learning model to classify lifestyle categories.

Dependent Variable: Lifestyle Categories
The lifestyle classification serves as the predicted outcomes based on the independent variables. The trained Random Forest Classifier groups individuals into predefined categories:

- Fitness Enthusiast – Individuals actively engaged in fitness activities and workouts.
- Health-Conscious – Individuals prioritizing healthy eating and well-being but not necessarily physically active.
- Eco-Friendly – Individuals practicing sustainable living.
- Social Media Influencer – Individuals highly engaged with online platforms, influencing trends and behaviors.

The classification process relies on pattern recognition, where the model identifies correlations between behavioral attributes and lifestyle types.

Outcome: Successful Prediction of Lifestyle Categories
The goal of this research is to create an accurate machine learning model that effectively predicts individual's lifestyle categories based on their behavior. The Random Forest algorithm, trained on a dataset of 500,000 lifestyle records, aims to generalize effectively to new inputs. The framework ensures that:

- Behavioral data is correctly preprocessed and transformed before training.
- The model accurately learns the complex relationships between behaviors and lifestyle groups.
- The system achieves an accuracy of 75.07% when predicting lifestyle categories tested on real-world participants (93 students).
- The classification results provide insights into how behavioral patterns align with specific lifestyle choices.

The structured framework allows researchers to scientifically examine the translation of behavioral data into lifestyle classification, highlighting the critical role of machine learning in behavioral analytics through a data-driven approach. Additionally, the framework highlights the potential for personalized health interventions, allowing targeted recommendations based on predicted lifestyle categories. This causal paradigm effectively demonstrates the value of data-driven decision-making in behavioral classification and health insights, supporting future advancements in AI-driven lifestyle analytics.

III. METHODOLOGY

A. *Data Selection*

Researchers selected datasets suitable for the purpose of the research. The first step is critical because the quality and relevance of the dataset have a significant impact on the model's outcomes. The dataset was sourced from the Kaggle website at https://www.kaggle.com/datasets/anthonytherrien/half-a-million-lifestyle. The purpose of this dataset is to predict the lifestyle category of individuals. This dataset includes a wide range of behavioral variables, such as exercise habits, social media influence, and stress management scores, aligning closely with the study's emphasis on behavior and lifestyle categories.

B. *Data Preprocessing*

In this phase, researchers performed data cleaning by deleting rows containing missing values. They then selected specific columns that are most relevant for predicting lifestyle categories, suitable for accepting user inputs. These columns include Gender, Age, Health-Conscious Rating, Average Daily Screen Time, Social Media Influence, Eco-Consciousness Metric, Stress Management, Number of Professional Trainings Attended, and Time Management Skills. This selection ensures the model is trained using the most pertinent behavioral features, thus aligning closely with the objective of predicting lifestyle categories.

C. *Data Transformation*

Researchers converted the gender variable into numerical values, assigning '1' to represent males and '2' to represent females. This conversion allows for a more effective interpretation of this characteristic during the training process.

D. *Model Creation and Implementation*

As illustrated in Figure 2, researchers programmed the Radom Forest Classifier to predict lifestyle categories based on behavior variables. The model was trained to recognize patterns within the data, enabling it to accurately predict new, unknown data. Researchers subsequently divided the cleaned dataset into two subsets: the training dataset, containing behavioral features, and the testing dataset, which focused on the lifestyle category. The algorithm generates numerous decision trees, with each tree assigning a classification. The final class is determined by majority voting across all trees.

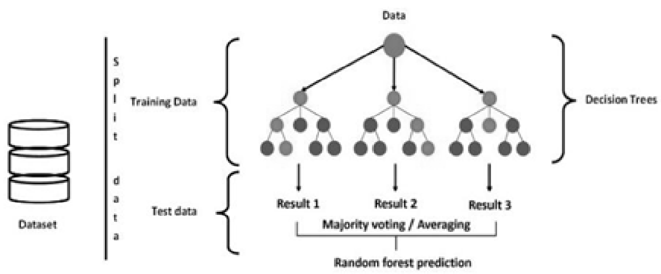

Figure 2. Random Forest Classifier Model

The Random Forest model was tested on 93 students using a programmed user-input system. Participants provided their behavioral data via a structured digital form, which was processed using the trained model. The results were analyzed to assess real-world applicability and model accuracy in predicting lifestyle categories from actual respondents. This step ensured that the study validated its predictive capability using real human input, not only relied on existing datasets.

E. Interpretation and Evaluation

This final stage involves evaluating the algorithm's accuracy rate and interpreting the model's results to understand their relevance to the study. Researchers split the dataset into training and testing sets to evaluate the model's performance using metrics such as accuracy, precision, recall, and F1 score.

The formulas used for these metrics are as follows:
Accuracy: Accuracy is calculated as the number of correct predictions (both true positives and true negatives) divided by the total number of predictions:

$$Accuracy = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}} \quad (1)$$

Precision: Precision is calculated as the number of true positives divided by the sum of true positives and false positives:

$$Precision = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \quad (2)$$

Recall: Recall is the number of true positives divided by the sum of true positives and false negatives:

$$Recall = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \quad (3)$$

F1 Score: F1-Score is the harmonic mean of precision and recall:

$$F1\,Score = 2 \times \frac{\mathrm{Precision} \times \mathrm{Recall}}{\mathrm{Precision} + \mathrm{Recall}} \quad (4)$$

where:
TP (True Positive) = the number of correctly predicted positive samples.
TN (True Negative) = the number of correctly predicted negative samples.
FP (False Positive) = the number of incorrectly predicted positive samples.
FN (False Negative) = the number of incorrectly predicted negative samples.

F. Interpretation and Evaluation

The computational efficiency of the Random Forest classifier is an important consideration, particularly when handling large datasets such as the Half a Million Lifestyle Dataset used in this study. The time complexity of the Random Forest algorithm primarily depends on the number of decision trees (T), the number of features (F and the number of training samples (N)

The Forest model can be approximated as:

$$O(T \cdot N \cdot \log N \cdot F) \quad (5)$$

Where:
$O(N \log N)$ = the complexity of building a single decision tree using recursive partitioning.
T = the number of trees in the forest.
F = the number of features considered at each node split.

For this study, a Random Forest model with 40 trees provided optimal, balance between predictive performance and computational efficiency. In comparison to alternative classifiers:

- Decision Trees $O(N \log N)$ are faster to train but prone to overfitting.
- Support Vector Machines $O(N^2)$ to $O(N^3)$ scale poorly with large datasets, making them computationally expensive.
- Neural Networks $O(NF)$ require significant hyperparameter tuning and longer training times.

The Random Forest model was trained within 2.5 minutes on a standard computing environment, demonstrating its practical feasibility for large-scale behavioral classification.

G. Addressing Biases

To ensure fairness and mitigate bias in classification, researchers conducted exploratory data analysis to identify any imbalances in feature distributions. Oversampling and under sampling techniques were employed where necessary to prevent the model from favouring dominant categories. Additionally, fairness metrics such as disparate impact analysis were used to assess whether any demographic group was disproportionately misclassified. This step ensured that the model provided an unbiased and equitable predictions of lifestyle categories.

H. Ethical Considerations

This study ensured that all ethical research standards were adhered throughout the data collection and processing. Researchers did not collect any personally identifiable information (PII) from the participants. The 93 student respondents voluntarily provided anonymous responses via a structured digital form that was programmed and accessed exclusively through the researcher's laptop. No personal names, emails, or other identifying data were recorded. Additionally, the study adhered to ethical data handling practices, ensuring that all collected information was stored securely and used strictly for research purposes. Participants were informed that their responses would be used solely to improve and validate the accuracy of the lifestyle prediction model. The research complied with standard data privacy regulations, reinforcing the confidentiality and integrity of participant input.

IV. RESULTS

The researchers used the Random Forest classifier to predict lifestyle categories by analyzing behavior data sourced from the Half a Million Lifestyle Dataset.

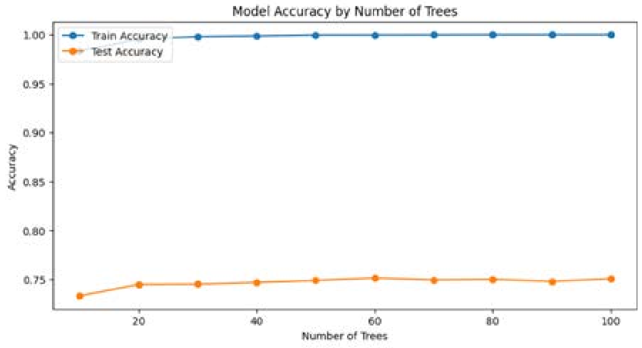

Figure 3. Model Accuracy by Number of Trees

Researchers conducted an experiment on the Random Forest model, using a range of trees from 10 to 100, increasing by increments of 10. Figure 3 shows how the model's accuracy changes with the number of decision trees. The training accuracy, represented by the blue line, starts at a high level of approximately 98% with only five trees and rapidly reaches a flawless accuracy of 100% when the number of trees reaches 30. This suggests that the model has a strong ability to memorize the training data, given a sufficient number of trees. On the other hand, the testing accuracy (represented in orange) begins at a lower level, approximately 72%, reflecting the model's initial capacity to apply its knowledge to unfamiliar data. As the number of trees increased to 20, the testing accuracy improved gradually to around 75%. After reaching this threshold, the accuracy fluctuated slightly, remaining within the 74% to 76% range, but consistently averaging at approximately 75% up to 100 trees. These observations suggest two key insights:

- The model's 100% accuracy on the training data indicates that it is overfitting, meaning the model has learned to categorize all training cases flawlessly but does not demonstrate comparable improvements on the test data.
- The optimal number of trees appears to be around 30-40 trees, balancing computational efficiency and predictive performance, as additional trees do not substantially improve accuracy.

These findings suggest that although the Random Forest model is effective in terms of its ability to learn, it is important to carefully assess the number of trees used to prevent overfitting and to optimize its ability to generalize to new and unexplored datasets.

These findings align with multiple studies highlighting the necessity of optimizing the number of trees to balance accuracy and computational efficiency [1][2][8]. Researchers note that the Random Forest classifier is effective in predictive accuracy, but using too many trees can lead to overfitting [9]. Further studies show the robustness of Random Forest classifiers in various applications, reaffirming the need for careful parameter tuning to achieve optimal results [10][11]. Researchers also discuss the importance of feature selection and model optimization in enhancing classification performance [12].

Table 1**:** Model Performance Metrics

Metric	Value
Random Forest Accuracy	0.7506592440668034
Precision	0.7516487311675587
Recall	0.7506592440668034
F1 Score	0.7440004383381931

Table 1 presents the performance characteristics of the model, showing an accuracy of 75.07%, precision of 75.16%, recall of 75.07%, and an F1 score of 74.40%. These metrics indicate the model's ability to predict lifestyle categories accurately while maintaining a satisfactory balance between precision and recall. Previous studies have demonstrated the capability of machine learning algorithms in classifying behavioral and psychological traits with considerable accuracy [1][8]. Researchers have applied Random Forest classifiers in various domains, including academic performance prediction [9][13], lifestyle classification [6][7], and health-related behavior assessments [2][10]. These findings support the effectiveness of machine learning models in identifying meaningful patterns in behavioral data, validating their role in lifestyle categorization and classification [2][7]-[10][14].

The literature reports similar findings. Studies have shown that machine learning models, particularly Random Forest classifiers, are effective in predicting health-related outcomes with comparable accuracy rates. For instance, researchers have applied Random Forest classifiers to predict student lifestyles based on behavior and demographic attributes [6][7], psychological well-being [9][14], and academic stress factors [1][14]. Various applications have demonstrated the robustness of Random Forest classifiers in classification tasks, including mental health assessments [10], predicting academic risks [13], and categorizing stress-related behaviors [5][8]. These studies consistently show that Random Forest classifiers are reliable in a variety of behavioral and educational prediction scenarios, supporting their application in lifestyle categorization [2][4][6][7][10][13][14].

Subsequently, the researchers developed a Python-based program to collect behavioral inputs from 93 students and process them through the trained Random Forest model. This system categorized each respondent into one of the predefined lifestyle categories, as shown in Table 2. The majority of students were classified as either Fitness Enthusiasts (41 students) or Health-Conscious (50 students), indicating that most respondents exhibited behavioral traits aligned with active and health-aware lifestyles. A small fraction of students was categorized as Eco-Friendly (1 student) and Social Media Influencer (1 student), suggesting that these lifestyle types were less prevalent in the sampled population.

The prediction accuracy of 75.07% remained consistent with the model's testing performance, reaffirming its ability to generalize to unseen data. The classifications demonstrate how the model effectively recognizes behavioral patterns and assigns individuals to appropriate lifestyle categories based on their habits, stress management, screen time, and health-consciousness levels. The results further validate the applicability of machine learning in behavioral and student classification predictions.

Table 2. Model Prediction Results from 93 students

Number of predictions	Lifestyle Category	Accuracy
1	Eco-Friendly	0.750659
41	Fitness Enthusiast	0.750659
50	Health-Conscious	0.750659
1	Social Media Influencer	0.750659

The literature has reported similar methodologies and applications. Studies have demonstrated the effectiveness of machine learning models in analyzing student behavior and performance patterns [1,][13]. Random Forest classifiers have been widely used in academic success prediction by evaluating demographic and lifestyle factors influencing students' achievements [2][3]. Furthermore, research has applied machine learning techniques to classify students' stress levels, mental health conditions, and academic retention probabilities, proving their relevance in student behavior assessment [4][5]. Additional studies emphasize the importance of user data in refining predictive models, enabling better classification accuracy and insights into student lifestyle trends [6][7]. Therefore, the successful application of machine learning techniques in student performance and behavioral classification reaffirms the robustness and practically of these methodologies in real-world academic and behavioral data processing [1]-[5][13].

V. DISCUSSION

The results demonstrate the effectiveness of the Random Forest classifier in predicting lifestyle categories based on behavior data from the Half a Million Lifestyle Dataset and real-world respondent validation through student responses. The model achieved perfect training accuracy of 100% with 100 trees, but the testing accuracy stabilized around 75%, indicating the importance of balancing model complexity to avoid overfitting while maintaining generalizability. This trend is consistent with findings in the literature, which emphasize the need to optimize the number of decision trees to achieve both accuracy and computational efficiency [2][13]. Other studies have also highlighted the risk of overfitting when excessive trees are used [1]. Empirical evidence from various applications supports the robustness of the Random Forest classifier, demonstrating the value of careful parameter tuning to improve model performance [6][11]. Additionally, extensive research [3] emphasizes the importance of feature selection and model optimization in improving classification accuracy.

Testing the model on behavioral data from 93 students confirmed that the Fitness Enthusiast (41 students) and Health-Conscious (50 students) categories were the most frequently predicted, representing the majority of respondents. Only one student was classified as Eco-Friendly and one as a Social Media Influencer, suggesting that these lifestyle types were less common in this group. The model maintained a prediction accuracy of 75.07% in this real-world scenario, consistent with its testing performance. This consistency supports the model's ability to generalize behavioral patterns from training data to actual user input, further validating its predictive capability.

Table 1 presents the model's performance metrics, showing an accuracy of 75.07%, precision of 75.16%, recall of 75.07%, and an F1 score of 74.40%. These values indicate that the model maintains a strong balance between precision and recall, making the model a reliable tool for lifestyle prediction. Previous studies have shown comparable accuracy rates in student performance classifications [2][13]. Furthermore, Random Forest classifiers have been successfully implemented for academic performance prediction [1][3] and behavioral classification [6][11], demonstrating their versatility and robustness in a range of predictive applications [1]-[3][6][11][13].

A similar clustering approach was used in Ram et al. (2025) [15], where student performance was grouped using machine learning techniques. Their study used K-Means clustering to analyze quiz performance among 140 students, revealing distinct learning patterns and cognitive differences. Similarly, this study used machine learning to categorize students' behavior into predefined lifestyle categories. Both studies demonstrate the potential of automated classification methods to uncover meaningful behavior and performance, whether in an academic setting [15] or in health and wellness context as explored in this research. The parallel between these methodologies reinforces the effectiveness of clustering and classification models for identifying meaningful behavioral trends in large datasets.

The researchers developed a Python-based program that enabled participants to input their behavioral data, which was processed by the trained Random Forest model to predict their lifestyle category. The distribution of predictions among the 93 students (as shown in Table 2) indicates that most participants engage in health-conscious behaviors, aligning with existing literature on health awareness trends. Similar research [2][13] has demonstrated the feasibility of using machine learning-based user input systems for academic performance prediction, reinforcing the value of data-driven classification techniques. Additional studies have shown the effectiveness of Random Forest classifiers in predicting behavioral traits based on user-reported lifestyle factors [6] and evaluating academic performance using machine learning models [1]. The successful implementation of machine learning techniques across multiple domains—from forecasting academic performance [6][13] to classifying student behavior [1][2][15]—demonstrates their reliability and effectiveness in processing real-world user input and predicting meaningful outcomes [1][2][6][13][15].

Overall, the findings align with existing literature, showcasing the robustness and adaptability of the Random Forest classifier in predicting lifestyle categories based on behavioral data. The model's consistent performance across training, testing, and real-world deployment illustrates its potential in behavioral analytics and academic assessment. This research contributes to enhancing decision-making process using machine learning, providing a scalable and reliable approach for identifying lifestyle trends and behavioral patterns in large-scale datasets.

A. Limitations and recommendations

While the model achieved high accuracy of 75.07%, it is constrained by predefined lifestyle categories that may not fully capture the complexity of individual behaviors. Additionally, the behavioral data was self-reported, introducing potential bias. Future research should explore larger and more diverse datasets, integrate additional behavioral factors, and test the model on different population group to enhance generalizability. Implementing real-time feedback systems, similar to adaptive learning models [2], could further improve prediction accuracy and refine lifestyle classification. Furthermore, researchers should consider using hybrid machine learning approaches to enhance model robustness and reduce classification errors. Addressing these limitations will improve the applicability of machine learning in behavioral classification and lifestyle prediction.

VI. CONCLUSION

This study demonstrates the effectiveness of the Random Forest classifier in predicting lifestyle categories using behavioral data from the Half a Million Lifestyle Dataset, with further validation from real-world testing involving 93 students. The model achieved an accuracy of 75.07%, successfully classifying most participants as Fitness Enthusiasts or Health-Conscious, with fewer categorized as Eco-Friendly or Social Media Influencers. These findings confirm the model’s ability to generalize behavioral patterns while emphasizing the importance of parameter tuning to mitigate overfitting. The results align with existing literature, reinforcing the growing role of machine learning in behavioral classification and lifestyle prediction. Additionally, the development and use of a Python-based program to collect and process behavioral data demonstrate the feasibility of applying machine learning in real-world settings to generate personalized data-driven insights for health and lifestyle assessments. Future research should explore larger and more diverse datasets, integrate additional behavioral variables, and investigate adaptive classification techniques to enhance predictive accuracy and model robustness.

CONFLICT OF INTEREST

Authors declare that there is no conflict of interest regarding the publication of the paper.

AUTHOR CONTRIBUTION

The authors confirm contribution to the paper as follows: study conception and design: Author Malakit L. Ram; data collection: Author Malakit L. Ram, Author Jose C. Agoylo Jr.; analysis and interpretation of findings: Author Malakit L. Ram, Author Jose C. Agoylo Jr.; draft manuscript preparation: Author Malakit L. Ram, Author Jose C. Agoylo Jr.. All authors had reviewed the findings and approved the final manuscript.

REFERENCES

[1] S. Jayaprakash, S. Krishnan, and V. Jaiganesh, “Predicting students academic performance using an improved random forest classifier,” in Proc. of 2020 International Conference on Emerging Smart Computing and Informatics (ESCI), 2020, https://doi.org/10.1109/esci48226.2020.9167547

[2] M. Nachouki, E. A. Mohamed, R. Mehdi and M. A. Naaj, “Student course grade prediction using the random forest algorithm: analysis of predictors’ importance,” Trends in Neuroscience and Education, vol. 33, 2023, https://doi.org/10.1016/j.tine.2023.100214

[3] M. Nachouki and M. A. Naaj, “Predicting students performance to improve academic advising using random forest algorithm,” International Journal of Distance Education Technologies, vol. 20, no. 1, 2022, https://doi.org/10.4018/ijdet.296702

[4] R. Rani and S. Gupta, “Predicting student anxiety and depression using random forest classifiers optimizer,” in Proc. of 2024 Second International Conference Computational and Characterization Techniques in Engineering and Sciences (IC3TES), 2024, pp. 1–5, https://doi.org/10.1109/ic3tes62412.2024.10877644

[5] R. Rois, M. Ray, A. Rahman and S. K. Roy, “Prevalence and predicting factors of perceived stress among Bangladeshi university students using machine learning algorithms,” Journal of Health, Population and Nutrition, vol. 40, no. 1, pp. 1–12, 2021, https://doi.org/10.1186/s41043-021-00276-5

[6] A. Chaturvedi, S. Yadav, M. A. M. H. Ansari and M. Kanojia, “College student lifestyle query classification using multi-model ensemble learning with polling technique,” in Advances in Intelligent Systems and Computing, 2021, pp. 645–654, https://doi.org/10.1007/978-981-16-2543-5_55

[7] E. Kasthuri and S. Balaji, “A chatbot for changing lifestyle in education,” in Proc. of 2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV), 2021, https://doi.org/10.1109/icicv50876.2021.9388633

[8] E. B. Zion and B. Lerner, “Identifying and predicting social lifestyles in people’s trajectories by neural networks,” EPJ Data Science, vol. 7, no. 1, 2018, https://doi.org/10.1140/epjds/s13688-018-0173-5

[9] S. Maitra et al., “Prediction of academic performance applying NNs: a focus on statistical feature-shedding and lifestyle,” International Journal of Advanced Computer Science and Applications, vol. 10, no. 9, 2019, https://doi.org/10.14569/ijacsa.2019.0100974

[10] C. Hu, “Evaluation of physical education classes in colleges and universities using machine learning,” Soft Computing, 2021, https://doi.org/10.1007/s00500-022-06983-3

[11] P. Chonggao, “Simulation of student classroom behavior recognition based on cluster analysis and random forest algorithm,” Journal of Intelligent and Fuzzy Systems, vol. 40, no. 2, pp. 2421–2431, 2021, https://doi.org/10.3233/jifs-189237

[12] S. Rajendran, S. Chamundeswari and A. A. Sinha, “Predicting the academic performance of middle- and high-school students using machine learning algorithms,” Social Sciences and Humanities Open, vol. 6, no. 1, 2022, https://doi.org/10.1016/j.ssaho.2022.100357

[13] S. Batool, J. Rashid, M. W. Nisar, J. Kim, T. Mahmood and A. Hussain, “A random forest students’ performance prediction (RFSPP) model based on students’ demographic features,” in Proc. of 2021 Mohammad Ali Jinnah University International Conference on Computing (MAJICC), 2021, https://doi.org/10.1109/majicc53071.2021.9526239

[14] B. Thanasekhar, N. Gomathy, A. Kiruthika and S. Swarnalaxmi, “Machine Learning Based Academic Stress Management System,” in Proc. of 2019 11th International Conference on Advanced Computing (ICoAC), 2019, https://doi.org/10.1109/icoac48765.2019.246831

[15] M. Ram, R. C. Gortifacion, J. A. Jr and J. Tagud, “Leveraging prolog’s declarative power for clustering student performance in a timed quiz,” International Journal of Multidisciplinary Studies in Higher Education, vol. 2, no. 1, pp. 105–120, 2025, https://doi.org/10.70847/592808.

.