CLASSIFICATION OF UNDERGRADUATE STUDENTS' ADAPTABILITY LEVELS IN HYBRID LEARNING USING MACHINE LEARNING TECHNIQUES

Chutikarn Sriviboon[1], Nutthapat Kaewrattanapat[2],
Martusorn Khaengkhan[3], Jarumon Nookhong[4], Sittichai Pintuma[5]
Suan Sunandha Rajabhat University, Thailand
Email: chutikarn.sr@ssru.ac.th[1]; nutthapat.ke@ssru.ac.th[2];
martusorn.kh@ssru.ac.th[3]; jarumon.no@ssru.ac.th[4]; sittichai.pi@ssru.ac.th[5]

ABSTRACT

This research investigated the adaptability levels of undergraduate students in hybrid learning environments using machine learning techniques. This research study was aimed to identify crucial factors influencing adaptability and assess the effectiveness of three distinct algorithms: Decision Trees, k-Nearest Neighbors (k-NN, k=3), and Naive Bayes. Comprehensive student data encompassing behaviors, adaptivity levels, and demographic information was analyzed to predict adaptability levels. The Decision Tree algorithm provided a foundational understanding but exhibited limitations in predicting higher adaptability, likely due to overfitting. The k-NN algorithm surpassed others, achieving the highest overall accuracy of 74.80% and demonstrating particular strengths in identifying moderate adaptability levels. This success can be attributed to its ability to recognize subtle similarities among data points, a crucial feature for analyzing nuanced hybrid learning experiences. However, k-NN faced challenges with imbalanced data, as evidenced by a lower recall for high adaptability levels. The Naive Bayes algorithm, despite its lower overall performance, offered valuable insights into the role of feature interdependencies. The study concludes that k-N's localized pattern recognition provides a more accurate reflection of student adaptability in hybrid learning contexts, highlighting the importance of contextual and relational data analysis.

Keywords: Classification Analysis, Educational Data Mining, Hybrid Learning, Digital University

INTRODUCTION

The rise of hybrid learning models, which blend traditional classroom instruction with online elements, has become a defining feature of contemporary higher education (Bates, 2015). As the educational landscape undergoes a dynamic shift towards digitalization, the adaptability of students to such models becomes increasingly crucial (Chiu & Lo, 2020). This research study is driven by the need to understand and categorize the adaptability levels of undergraduate students within hybrid learning environments (Nuanmeesri, 2019).

The significance of this endeavor is underscored by the growing reliance on technology in education and the imperative for educational institutions to tailor their teaching methodologies to student capabilities and learning preferences (Ally, 2009). In essence, this study addresses the gap in comprehensive analysis and classification of student adaptability, which is vital for the development of effective hybrid learning strategies (Chang & Chen, 2012).

In the pursuit of this objective, machine learning (ML) techniques emerge as powerful analytical tools. These techniques possess the potential to decipher complex patterns within educational data, offering insights into student behavior and adaptability that often elude traditional analysis (Alshareef et al., 2020). This study focuses on employing and comparing three ML algorithms – Decision Trees, k-Nearest Neighbors (k-NN), and Naive Bayes – to identify and predict the adaptability levels of students. By harnessing these algorithms, the study delves into the multifaceted data, uncovering associations and predictors that are instrumental in comprehending the student experience within hybrid learning models (Alshareef et al., 2020). The analysis through these ML lenses aspires to transcend mere data crunching, endeavoring to translate data patterns into meaningful narratives about student learning journeys.

The implications of this research are far-reaching. From an educational planning perspective, the insights gleaned from ML analysis have the potential to inform pedagogical approaches, curricular designs, and student support systems, aligning them more closely with student needs and the realities of hybrid learning (Wannapiroon et al., 2021). At a policy level, the findings can influence the strategic direction of higher education institutions as they evolve into digital universities (Bates, 2015). The ability to classify and predict student adaptability levels equips educators and administrators with the knowledge to create more inclusive, personalized, and effective learning environments. This research not only contributes to the academic discourse but also serves as a cornerstone for educational institutions to foster a resilient and adaptive student body, prepared to flourish in an increasingly digital world (UNESCO, 2015).

RESEARCH OBJECTIVES

This research study was aimed to

1) To collect and analyze relevant data on student behaviors, adaptivity level, and demographic information to identify key factors affecting adaptability levels.
2) To investigate and analyze the effectiveness of Decision Trees, k-Nearest Neighbors (k-NN), and Naive Bayes algorithms in predicting students’ adaptability levels using RapidMiner.
3) To compare the performance of Decision Trees, k-Nearest Neighbors (k-NN), and Naive Bayes algorithms in predicting students’ adaptability levels and determine which one provides the best results.

CONCEPTUAL FRAMEWORK

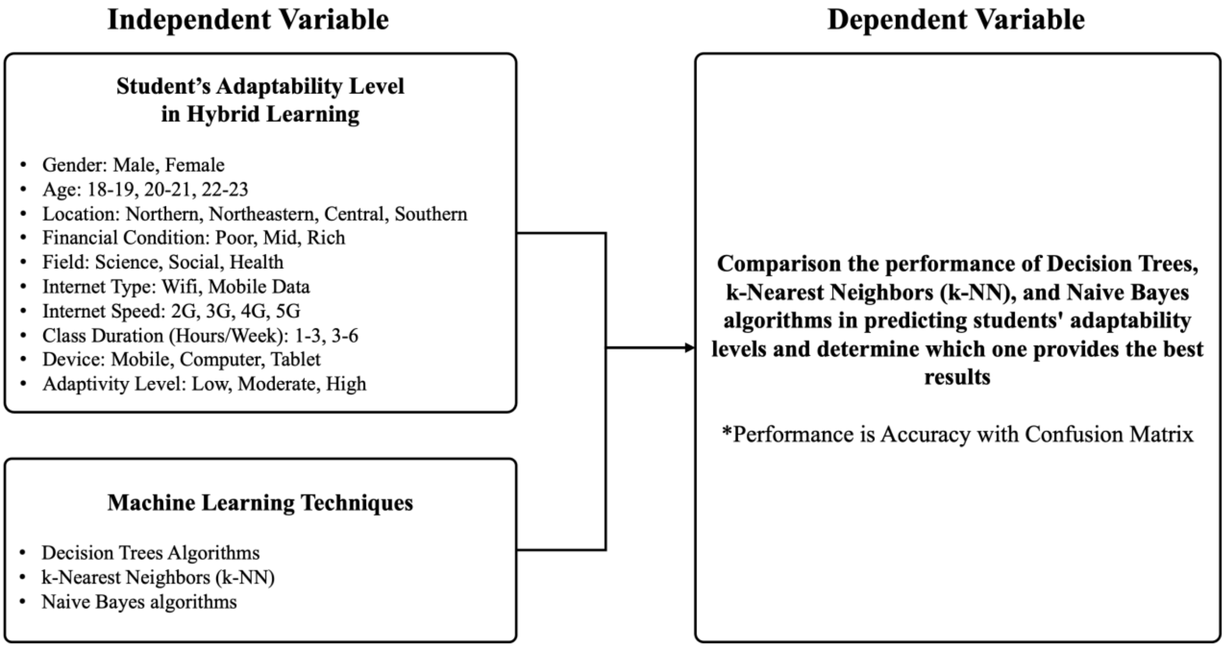

Figure 1. Conceptual Framework for Analyzing
Undergraduate Students' Adaptability Levels in Hybrid Learning

As depicted in Figure 1, the conceptual framework provides a structured approach for evaluating the adaptability levels of 1,000 Thai undergraduate students in hybrid learning environments. Data were collected between August 2022 and September 2023, encompassing various independent variables such as gender, age ranges of 18-19, 20-21, and 22-23, geographic locations across Thailand, financial status, fields of study, internet connectivity details, weekly class hours, and the types of learning devices used. The framework facilitates the examination of these variables against the dependent variable — the adaptability level, categorized as low, moderate, or high. The framework also outlines the use of Decision Trees, k-Nearest Neighbors (k-NN), and Naive Bayes algorithms for predictive analysis, with performance metrics derived from a Confusion Matrix. Figure 1 serves as a foundational model for gaining data-driven understanding of student adaptability in a dynamically changing educational landscape.

METHODOLOGY

In proceeding to elucidate the methodologies underpinning this study, we aim to delineate the systematic and rigorous approach adopted for examining the adaptability levels of undergraduate students in hybrid learning settings. This section is dedicated to detailing the comprehensive methodological framework, encompassing the processes of data collection, analysis, and model implementation.

The methodology outlined here is integral to the scientific rigor of our research. It has been meticulously designed to address the multifaceted dimensions of hybrid learning and to capture the nuanced variations in student adaptability. The ensuing exposition will provide a thorough account of our procedural strategies, including participant selection, data gathering techniques, variable measurements, and the application of sophisticated machine learning algorithms.

The methodological design is underpinned by a commitment to empirical precision and analytical depth. It reflects an endeavor not only to fulfill the specific objectives set forth in this study but also to contribute substantively to the evolving discourse on technology-enhanced education. As such, the following narrative seeks to offer clarity and comprehensive insight into the methodological rigor that characterizes our investigation, underscoring its contribution to the broader field of educational research.

The study conducted a cross-sectional analysis of 1,000 Thai undergraduate students to classify adaptability levels in hybrid learning environments. Data were collected over a 14-month period from August 2022 to September 2023, reflecting a wide spectrum of individual, technological, and educational variables. Participants were Thai undergraduate students across various universities, selected using a stratified sampling technique to ensure representation across genders, age groups, locations, and financial backgrounds. Data were collected through online surveys and institutional records. Surveys captured demographic details, internet usage patterns, class engagement, and perceived adaptability to hybrid learning. Institutional records provided corroborative data for cross-verification. Participants provided informed consent, and data anonymization was employed to ensure privacy.

Independent variables included demographic information (gender, age, location, financial condition), educational context (field of study, class duration), and technological access (internet type, speed, device used). The dependent variable was the adaptability level to hybrid learning, categorized into low, moderate, and high based on a validated adaptability scale.

Data cleansing involved removing duplicates, handling missing values, and ensuring consistency. Normalization procedures were applied to quantitative variables, and categorical variables were encoded appropriately for algorithmic processing.

Three machine learning models—Decision Trees, k-Nearest Neighbors (k-NN), and Naive Bayes—were developed. Model performance was evaluated based on accuracy, derived from the Confusion Matrix. The best-performing model was selected based on these metrics.

Accuracy is a fundamental metric in evaluating the performance of classification models. It is defined as the ratio of correctly predicted observations to the total observations. Mathematically, accuracy can be expressed as:

$$Accuracy = \frac{Number\ of\ Correct\ Predictions}{Total\ Number\ of\ Predictions}$$

This calculation provides a straightforward measure of how often the model correctly predicts the adaptability levels of students. Higher accuracy indicates a better performance of the model in classifying the students accurately.

A Confusion Matrix is a table used to describe the performance of a classification model on a set of test data for which the true values are known. It allows for the visualization of the performance of an algorithm. The matrix compares the actual target values with those predicted by the model, providing a comprehensive view of the correctness and types of errors made by the model.

Table 1: Confusion Matrix for Classification Model Performance

Confusion Matrix	Predicted: Positive	Predicted: Negative
Actual: Positive	*True Positives (TP)*	*False Negatives (FN)*
Actual: Negative	*False Positives (FP)*	*True Negatives (TN)*

This table, referred to as a Confusion Matrix, is utilized to assess the performance of the classification models applied in the study. It consists of four components: *True Positives (TP), True Negatives (TN), False Positives (FP),* and *False Negatives (FN)*. The matrix cross-tabulates the actual versus predicted classifications to provide insights into the accuracy and potential biases of the model. True Positives and True Negatives represent the instances where the model has correctly predicted the adaptability levels, while False Positives and False Negatives signify incorrect predictions. The effectiveness of the Decision Trees, k-Nearest Neighbors (k-NN), and Naive Bayes algorithms in categorizing students' adaptability levels in hybrid learning environments is quantified through this matrix.

RESULT

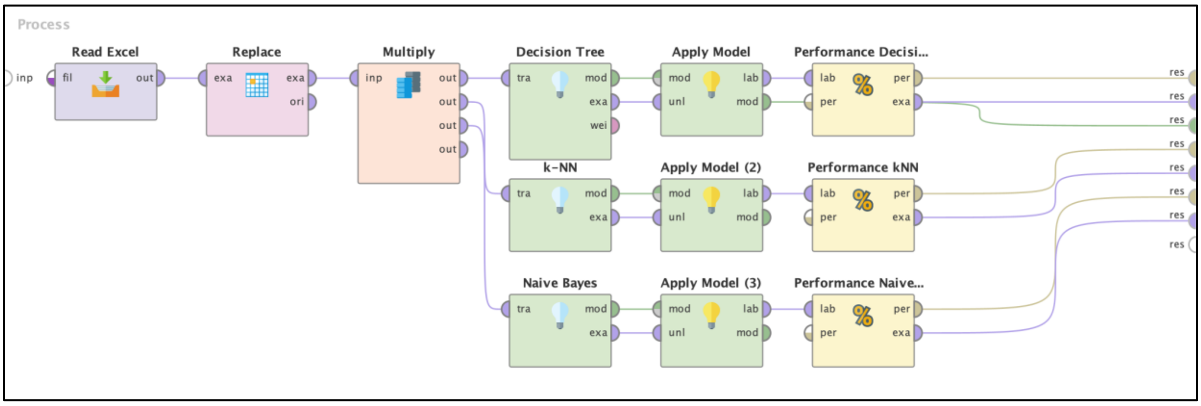

Figure 2. Process Flow for Machine Learning Analysis in RapidMiner

From Figure 2 represents the analytical process flow executed using RapidMiner to classify adaptability levels among undergraduate students in a hybrid learning environment. The initial step involves reading the input data from an Excel file, followed by data preprocessing, where specific values are replaced as needed. The "Multiply" block indicates the branching out of the dataset into three distinct streams for application of different machine learning models: Decision Tree, k-Nearest Neighbors (k-NN, k=3), and Naive Bayes. Each model is applied to the dataset, followed by an assessment of its performance. The performance evaluation for each algorithm is visualized separately, with results derived from accuracy metrics and Confusion Matrix analyses. This flowchart is pivotal in illustrating the systematic approach to applying and evaluating multiple classification techniques to determine the most effective algorithm for predicting student adaptability levels.

Table 2. Confusion Matrix for Decision Tree Algorithm Performance

Classification	true Moderate	true Low	true High	class precision
pred. Moderate	343	142	65	62.36%
pred. Low	173	260	17	57.78%
pred. High	0	0	0	0.00%
class recall	66.47%	64.68%	0.00%	**accuracy: 60.30%**

Table 2 presents the Confusion Matrix derived from employing the Decision Tree algorithm to determine adaptability levels among students. This matrix cross-references the predicted adaptability categories—Moderate, Low, and High—against the true classifications. Predictions made by the algorithm are delineated in rows, while the actual categories are arrayed in columns. Correct predictions are evident along the principal diagonal (343 for

Moderate, 260 for Low, and 0 for High). Precision and recall percentages are calculated for each category, reflecting the model's precision and recall in predicting each class accurately. The Decision Tree model exhibits an overall accuracy of 60.30%, with the precision peaking for the Moderate class at 62.36% and recall for the Low class at 64.68%. A notable observation is the absence of predictions for the High adaptability category, indicating a potential area for model refinement or a deeper examination of data and feature selection that may influence the High adaptability predictions.

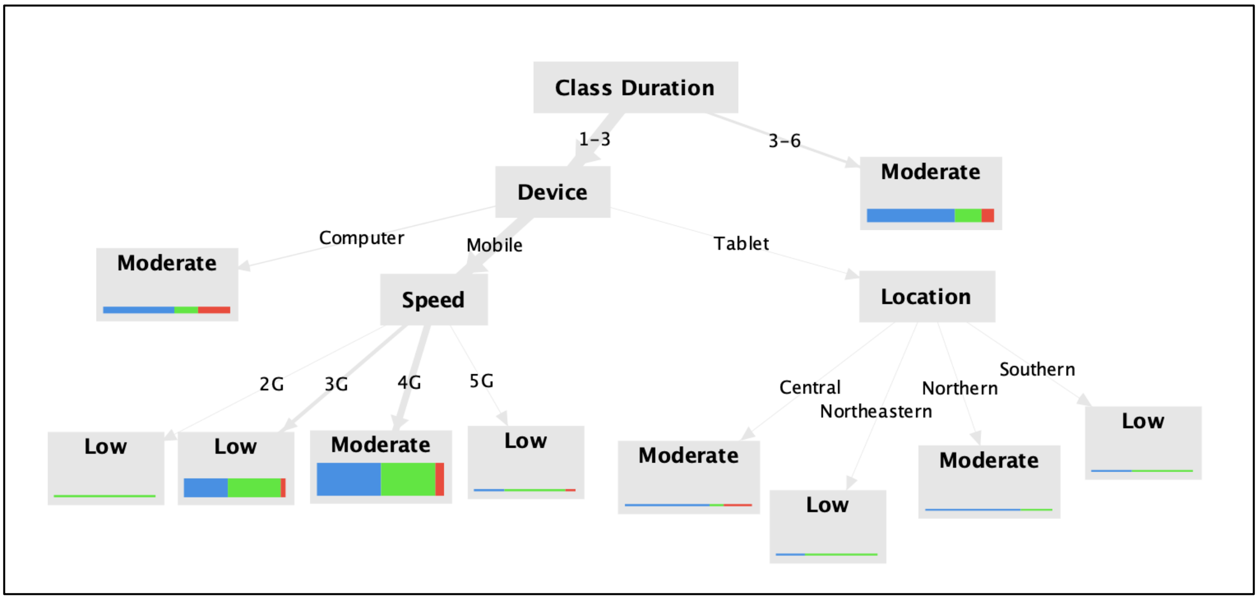

Figure 3. Decision Tree Results on Student's Adaptability Level Attribute

Figure 3 illustrates the results of a Decision Tree analysis showing the influence of various factors on student adaptability levels in a hybrid learning environment. The tree branches represent the decision points based on class duration, device type, internet speed, and geographic location. Each leaf node is color-coded to reflect the adaptability level: Moderate (blue), Low (green), or High (red). The analysis suggests that students using computers with higher internet speeds tend to have moderate adaptability levels, whereas those with lower speeds or using mobile devices are classified as having low adaptability. The class duration also appears to be a significant factor, with longer durations correlating with moderate adaptability, while location shows a mixed impact. For instance, students from the Northern and Central regions exhibit moderate adaptability, whereas those from the Southern region are classified as having low adaptability. This visualization aids in understanding the complex interplay of factors affecting adaptability and informs targeted interventions to enhance student learning experiences.

Table 3. Confusion Matrix for k-Nearest Neighbors (k-NN, k=3) Performance

Classification	true Moderate	true Low	true High	class precision
pred. Moderate	462	141	35	72.41%
pred. Low	53	260	21	77.84%
pred. High	1	1	26	92.86%
class recall	89.53%	64.68%	31.71%	**accuracy: 74.80%**

Table 3 presents the Confusion Matrix for the k-Nearest Neighbors algorithm (with k set to 3) applied in classifying student adaptability levels in a hybrid learning context. The matrix provides a detailed account of the algorithm's classification accuracy by comparing the predicted adaptability levels against the true labels. Here, the classification precision and recall for each adaptability level—Moderate, Low, and High—are reported. The algorithm demonstrates a high precision rate of 92.86% for predicting High adaptability, although the recall for this class is notably lower at 31.71%, indicating a conservative prediction for this category. In contrast, Moderate adaptability predictions show the highest recall at 89.53%, suggesting a strong ability of the algorithm to identify true Moderate cases. The overall accuracy of the k-NN model in this study is calculated at 74.80%, reflecting a substantial predictive capability but also highlighting areas for potential improvement in classifying adaptability levels with greater balance across the classes.

Table 4. Confusion Matrix for Naïve Bayes Performance

Classification	true Moderate	true Low	true High	class precision
pred. Moderate	364	212	63	56.96%
pred. Low	151	190	18	52.92%
pred. High	1	0	1	50.00%
class recall	70.54%	47.26%	1.22%	**accuracy: 55.50%**

Table 4 presents the performance of the Naïve Bayes algorithm in classifying the adaptability levels among students into Moderate, Low, and High categories. The matrix quantifies the algorithm's predictions against the actual classifications, providing insights into the precision and recall for each adaptability level. The model displays a moderate precision of 56.96% for correctly identifying Moderate adaptability and a lower precision of 52.92% and 50.00% for Low and High adaptability, respectively. The recall rates indicate that the model is more effective at identifying Moderate adaptability (70.54%) than Low (47.26%) and High (1.22%) levels. Overall, the Naive Bayes algorithm achieves an accuracy of 55.50%, suggesting that while the model has some predictive capability, there is substantial room for improvement, particularly in identifying the true High adaptability cases, which is reflected in the low recall rate for that class.

CONCLUSION

In this study, our objective to assess the adaptability levels of undergraduate students in hybrid learning environments through machine learning techniques has yielded significant findings. We compared three algorithms: Decision Trees, k-Nearest Neighbors (k-NN, k=3), and Naive Bayes, each with its distinct approach to classification.

The Decision Tree algorithm provided a foundational understanding of the data, yet it fell short in identifying students with high adaptability levels. This might be due to the algorithm's sensitivity to the training data and a potential overfitting to certain classes, leaving it less adept at generalizing to less represented categories.

The k-Nearest Neighbors algorithm emerged as the superior performer with the highest overall accuracy of 74.80%. Its effectiveness in classifying moderate adaptability levels was particularly notable. The strength of k-NN in this context could stem from its ability to capture the nuanced similarities among data points, which is crucial in the varied landscape of hybrid learning environments. However, the algorithm's lower recall for high adaptability levels suggests a difficulty in dealing with imbalanced datasets, which is common in real-world scenarios.

Naive Bayes showed the lowest accuracy but still provided valuable insights. Its assumption of feature independence may have limited its performance, considering the complex interdependencies in student data related to adaptability. The superior performance of the k-NN algorithm suggests that adaptability levels in hybrid learning may be more effectively captured through localized patterns rather than global models, as the algorithm considers the proximity of data points to make predictions. This insight points to the importance of context and the relational aspects of data points in educational data mining.

In conclusion, while k-NN was the most effective algorithm in our study, it is important to note that the choice of algorithm should be aligned with the specific characteristics of the dataset and the research objectives. The findings advocate for a nuanced approach to the application of machine learning in educational settings, where a deep understanding of the data and the underlying educational processes is as crucial as the algorithmic choices made. Future research could explore hybrid models that combine the strengths of different algorithms to enhance prediction accuracy and further our understanding of student adaptability in hybrid learning environments.

ACKNOWLEDGMENTS

The authors would like to thank Suan Sunandha Rajabhat University, Bangkok, Thailand (http://www.ssru.ac.th/) to provide funding support to attend the dissemination of research on this and thank family, friends, colleagues, students in the field of Digital Innovation Management and Content, Digital Technology for Education and The Office of General Education and Innovative e-Learning for cooperation and provide the dataset in research, all of you.

REFFERENCES

Ally, Mohamed, ed. (2009) Mobile Learning Transforming the Delivery of Education and Training. Issues in Distance Education. AU Press. ISBN 9781897425442

Alshareef, F., Alhakami, H., Alsubait, T., & Baz, A. (2020). Educational Data Mining Applications and Techniques. International Journal of Advanced Computer Science and Applications (IJACSA), 11(4). http://dx.doi.org/10.14569/IJACSA.2020.0110494

Bates, T. (2015). Teaching in a digital age: Guidelines for designing teaching and learning. BCcampus Open Education.

Chang, C.-K., & Chen, W.-C. (2012). Factors affecting students' acceptance of e-learning environments. International Journal of Human-Computer Studies, 70(6), 405-414.

Chiu, C.-M., & Lo, Y.-C. (2020). The impacts of digital literacy and self-regulated learning on adaptability in blended learning environments. International Journal of Distance Education, 39(3), 119-137.

Wannapiroon, P., Nilsook, P., Kaewrattanapat, N., Wannapiroon, N., & Supa, W. (2021). The Virtual Learning Resource Center for the Digital Manpower. International Education Studies, 14(9), 28-43.

Nuanmeesri, S. (2019). Extended study of undergraduate students' usage of mobile pplication for individual differentiation learning support of lecture-based general education subjects. International Association of Online Engineering, 99-112.

UNESCO. (2015). Rethinking Education: Towards a Global Common Good? (84 p.). https://doi.org/10.54675/MDZL5552