WIP: Machine learning models for predicting student performance in IoT-enhanced education

Tomohiro Hayashida
Graduate School of Advanced Science and Engineering
Hiroshima University
Higashihiroshima, Japan
ORCID: 0000-0003-4476-7468

Shin Wakitani
Graduate School of Advanced Science and Engineering
Hiroshima University
Higashihiroshima, Japan
ORCID: 0000-0002-3850-3864

Kento Tsutsumi
Faculty of Education
Yamaguchi University
Yamaguchi, Japan
ORCID: 0009-0005-6108-9349

Takuya Kinoshita
Graduate School of Advanced Science and Engineering
Hiroshima University
Higashihiroshima, Japan
ORCID: 0000-0003-4871-5171

Ichiro Nishizaki
Graduate School of Advanced Science and Engineering
Hiroshima University
Higashihiroshima, Japan
ORCID: 0000-0002-0060-4360

Shinya Sekizaki
Graduate School of Advanced Science and Engineering
Hiroshima University
Higashihiroshima, Japan
ORCID: 0000-0002-8493-7700

***Abstract*—This work in progress resaerch-to-practice mainly contributes the development of a multistage classification method to distinguish between learners who are easily predictable and those who exhibit more complex patterns. This methodological innovation not only supports the efficient allocation of educational resources but also aids in customizing learning materials for learning management systems, enhancing the relevance and effectiveness of educational content. In Japan, the GIGA School Concept, which involves the use of one internet-connected electronic device per student, has been introduced in all elementary and junior high schools since 2020. The integration of IoT (Internet of Things) technology supported by such initiatives represents a significant shift towards creating more interactive and tailored learning environments. This evolution aims not only to enhance educational outcomes by providing individual electronic devices to learners but also to optimize learning results and operational efficiency. These technological tools enable learners to progress through their educational journey at their own pace, promoting a more learner-centered approach known as "Personalized Learning." Simultaneously, educators can more effectively monitor and manage the learning process, enabling a more responsive and adaptive educational experience. At the central of this transformation is the role of machine learning models (such as neural networks and random forests) used to predict learners' performance based on past academic achievements. This predictive capability allows for the grouping of learners based on the predictability of their performance, further refining educational strategies to meet diverse learning needs. Furthermore, this study explores the practical application of these models through numerical experiments using actual learner data. The results highlight the potential of combining IoT technology and machine learning to revolutionize the educational domain by providing personalized learning experiences. Such an approach promises not only to improve learning outcomes by aligning educational content and strategies with individual learners' profiles but also to streamline the educational process, reduce the burden on educators, and optimize the overall educational environment. In conclusion, the integration of IoT and machine learning into education presents a visionary method for addressing the unique needs and potential of individual learners.**

***Index Terms*—Learning management systems, elementary school, learning objectives**

This work was supported by JSPS KAKENHI Grant Number JP23K02685.

I. Introduction

In recent years, the introduction of electronic devices and digitalization in educational settings has progressed, backed by the GIGA School Concept [1]. Traditionally, all students learned the same material simultaneously, making it difficult to tailor learning to each individual's understanding. By establishing an environment where each student has their own device, it is now possible for each to learn different content simultaneously while recording their learning history. This allows for individualized learning tailored to each student's needs and progress. Additionally, educators are required to implement suitable educational support to optimize individual learning. To provide proper educational support, it is necessary to understand each learner's characteristics and select and provide learning materials based on those characteristics. Inadequate educational support could lead to a decline in learning motivation and unstable learning habits. If educators can understand a learner's weaknesses in advance based on past performance, they can set and implement educational strategies accordingly.

For predicting learners' proficiency and progress in educational content, Latrellis et al. [2] proposed regression and classification models using the K-means algorithm and Random Forest (RF) [7]. They used regression models to predict the number of years from enrollment to graduation and binary classification models for the number of graduate students. In both models, learners were clustered into several groups using the K-means algorithm, and regression or classification was performed using RF within those clusters. The experimental results showed a 12% reduction in Relative Absolute Error (RAE) for the regression models compared to unclustered cases, and the classification models achieved about 79% accuracy. Keser et al. [3] proposed a hybrid model to predict the performance of middle school students in mathematics and Portuguese in Portugal. They conducted two types of experiments using Gradient Boosting (GB), eXtreme Gradient Boosting (XGBoost) [8], [9], and Light Gradient Boosting Machine (LightGBM) for binary classification of pass/fail and five-level classification based on final grades. Both experiments showed the highest accuracy with the proposed method, achieving about 96% accuracy in binary classification for mathematics, about 91% for Portuguese, and about 80% accuracy for both subjects in five-level classification. Furthermore, to capture learners' characteristics, Hayashida et al. [4] conducted prediction experiments using Feedforward Neural Networks (FNNs) with survey responses indicating each learner's understanding of the material and test results evaluating proficiency. The input for the FNN was the survey results, and the output was the predicted test results, proposing a method to classify learners into three classes: High (H), Middle (M), and Low (L). This method demonstrated high accuracy in classifying learners, especially identifying those who might require individualized learning support in the future with about 70% accuracy. Qu et al. [5] conducted experiments to predict learners' grades using a five-layer Multilayer Perceptron (MLP) by analyzing learners' behavioral patterns. They also used a feature selection algorithm to select essential features for predicting academic performance and conducted predictions using the MLP, achieving 70% accuracy in predicting students' grades. Hirotani [6] proposed a method to predict future proficiency for each learner cluster using Self-Organizing Maps (SOMs) and FNNs. SOM is a type of neural network model that maps high-dimensional observational datasets to a low-dimensional space while preserving the topological structure of the data distribution, making it easier to visualize the relationships between data. By initially using SOM for feature extraction and clustering, and then predicting future proficiency using FNN for each learner cluster, they classified learners into seven groups based on the prediction results. Experiments with actual middle school learner data showed that at least three groups were successfully classified with at least 72% accuracy. While some clusters showed improved prediction accuracy using SOM, others had lower similarity and worse prediction accuracy compared to other clusters.

From the results of introducing prior classification using SOM [6], it can be inferred that pre-classifying learners based on their learning characteristics allows for the identification of groups of similar learners from past learning data, ultimately improving prediction accuracy. This paper proposes a method to classify learners based on their learning characteristics over a certain period, using past records of similar learners and machine learning models, including RF, to conduct numerical experiments on predicting academic performance based on past learner data and proposing a learner classification method to improve prediction accuracy.

The structure of this paper is as follows. Chapter 2 introduces previous studies on learner classification using neural networks, while Chapter 3 covers previous studies on learner classification using decision trees. Chapter 4 explains the procedures of the proposed method. Chapter 5 describes the methodology and results of numerical experiments using actual learner data and discusses these results. Finally, Chapter 6 summarizes the paper and outlines future challenges.

II. Learner Classification Using Neural Networks

Hirotani et al. [6] conducted experiments using learner data from real middle school students with Self-Organizing Maps (SOM) and Feedforward Neural Networks (FNN), successfully classifying them into three groups with at least 72% accuracy. Hayashida et al. [10] used an extended Multi-Context Recurrent Neural Network (exMCRNN), suitable for handling time-series data, to classify learners into five categories: High (H), Middle (M), Low1 (L1), Low2 (L2), and Low3 (L3). The growth characteristics of learners in the low-performing class L are differentiated based on the speed of proficiency growth and the time it takes from the start of learning to the onset of actual proficiency growth, further subdividing class L into three categories. Numerical experiments have successfully raised the accuracy of classes L2 and L3 to 70%, and the accuracy for class L as a whole was over 90%.

Jain et al. [11] aimed to classify learners based on their learning levels using four machine learning models: decision trees, Random Forest (RF), gradient boosting, and XGBoost. The experimental data involved 395 learners, and feature selection was conducted from 32 features including grades, parents' educational levels, and number of past failures, narrowing down to nine key features. The optimized RF model, after feature selection and parameter tuning, showed the best results with 95% accuracy.

III. Predicting Proficiency with Decision Tree-baesd Models

Decision trees, Random Forest (RF) [7], and eXtreme Gradient Boosting (XGBoost) [8], [9] are introduced as means to predict learners' proficiency using learner data [11].

A. Decision Trees

When d-dimensional input data is given to a decision tree, it initially enters the root node. At each node, a branching condition is specified using a feature and a threshold value, and depending on the result of the branching condition, the data is directed to one of the child nodes. Branching conditions that minimize impurity at the node receiving the input data are created, and data is directed to child nodes accordingly. Branching conditions are recursively created at each child node until the leaf nodes are reached, where the process ends and a result corresponding to the input is outputted.

An example of a decision tree is shown in Figure 1.

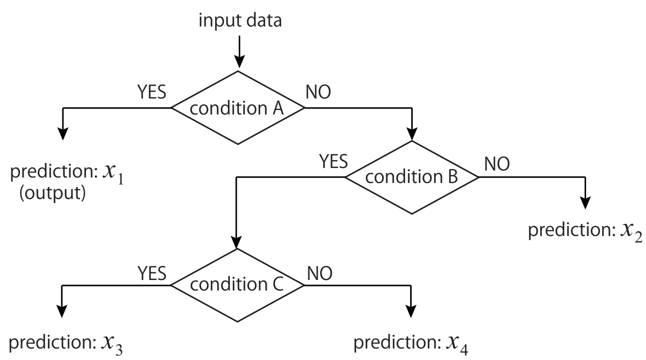

Fig. 1. An example of a decision tree

In the example shown in Figure 1, if the input data satisfies condition A, the predicted value (output) is set to x_1. If not, it checks whether condition B is satisfied. If condition B is satisfied, it then checks if condition C is met. If condition C is met, the predicted value is set to x_3; if not, it is set to x_4. If condition B is not satisfied, the predicted value is set to x_2. This system determines predicted values based on several conditions applied to the input data, which are the explanatory variables. It is useful for data where it is difficult to construct a prediction or classification model using a unified mathematical formula, such as in regression analysis.

B. Random Forest

Random Forest (RF) is composed of multiple decision trees, each learning independently. A RF uses bagging to select the training data for learning the decision trees, which involves randomly sampling data from the original dataset with replacement, allowing each decision tree to learn from different subsets of data. This approach enables individual trees to capture different aspects of the data, improving the overall classification accuracy's stability. Unlike decision trees, which use all features for branching, RF uses only a subset of randomly selected features, allowing each tree to learn based on different features, thus enhancing the model's overall generalization performance. After the learning is complete, the output values of each decision tree determine the final prediction or classification by RF, using averages in regression or majority voting in classification.

C. XGBoost

XGBoost, like RF, is a type of ensemble learning method that uses decision trees as weak learners. However, it is based on a technique called gradient boosting. Unlike RF where each decision tree learns independently, gradient boosting sequentially trains multiple decision trees, with each new tree correcting the errors of the previous one. Furthermore, XGBoost incorporates a regularization term into the loss function to constrain the model's complexity, helping to prevent overfitting.

IV. Proposed Method

This study aims to predict and classify learners' proficiency based on machine learning models. Let the learning data be D-dimensional, and $t(t = 1, 2, \ldots, T)$ represents a period in the data series. When $m(m = 1, 2, \ldots, M)$ are the evaluation criteria, learner i has data $\mathbf{X}_i \equiv (\mathbf{x}_i(1)\mathbf{x}_i(2), \ldots, \mathbf{x}_i(T))$, with $\mathbf{x}_i(t) \equiv (x_{i1}(t), x_{i2}(t), \ldots, x_{iM}(t))$. When predicting the proficiency at period t, $\mathbf{x}_i(t)$, the input is $\mathbf{I}_i(t) = (\mathbf{x}_i(1), \mathbf{x}_i(2), \ldots, \mathbf{x}_i(t-1))$, and the output is $\hat{x}_{im}(t), (m = 1, 2, \ldots, M)$. In other words, the machine learning model predicts $x_{im}(t)(m = 1, 2, \ldots, M)$ based on $\mathbf{I}_i(t)$.

Previous studies used a single neural network model for prediction, which was unable to capture the complex characteristics of learners and did not achieve sufficient accuracy. Therefore, this study classifies learners into those who can be predicted and those who are difficult to predict using machine learning models. For learners deemed difficult to predict, a second stage of classification is performed using a different machine learning model. Those still difficult to predict after this stage are deemed to require individual attention by educators. Figure 2 uses multiple regression analysis or RF for the first stage of classification as "Machine Learning Model (1st stage)". Next, the second stage of classification, as shown in Figure 2 as "Machine Learning Model (2nd stage)", considers using FNN, decision trees, RF, or XGBoost.

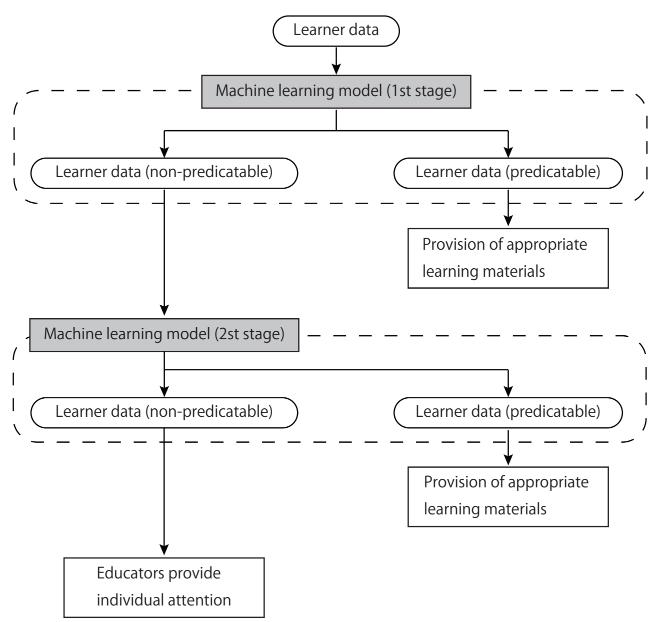

Fig. 2. Procedure of the proposed method.

A. First Stage of Classification Using Machine Learning Models

If the contribution rate of the prediction results of the learning data using machine learning models is less than θ, the learners with the largest prediction errors are removed as outliers. Prediction is again performed using the remaining learner data, and the contribution rate is recalculated. This process is repeated until the contribution rate exceeds θ, and the same procedure is applied to validation and test data, thereby classifying learners into those who can be predicted and those who are difficult to predict. This classification allows for the formulation of learning strategies based on predictions from machine learning models and the provision of suitable materials for predictable learners. For those difficult to predict, attempts are made to predict using a second-stage machine learning model. Moreover, if learners can be classified as either predictable or difficult to predict in any term, and this classification remains consistent in future terms, it suggests that their learning characteristics are similar, which can lead to appropriate educational support. Multiple regression analysis or RF is used for the first stage of the machine learning model.

B. Second Stage of Classification Using Machine Learning Models

Learners classified in the second stage were outliers in the first stage, thus having a lower contribution rate, making it difficult to classify them by contribution rate as in the first stage. Therefore, a threshold for prediction error is considered, classifying learners as predictable if below the threshold and as difficult to predict if above it. Learners defined as predictable are then subjected to prediction using one of the second-stage machine learning models: FNN, decision trees, RF, or XGBoost. On the other hand, learners classified as difficult to predict in the second stage are considered challenging to represent using machine learning models.

V. Numerical Experiments

A. Experimental Data

This study uses evaluation data for five cohorts of middle school students, from the 2013 to 2017 school years, totaling 394 students. Each student's data comprises three years of school, with three terms per year, defining one period as one school term, resulting in nine periods of data per student. The data evaluated are the results of tests conducted each term, with evaluation criteria consisting of four items: creativity and innovation, skills, knowledge, and motivation/interest, referred to sequentially as aspects 1, 2, 3, and 4. Thus, each student's data consists of 36 values (9×4).

B. Experimental Overview

For students enrolled from 2013 to 2016, 80% are randomly selected as the training data, with the remaining 20% used as validation data. The 2017 enrollees' data serve as test data to evaluate the performance of the proposed model. Including students with missing data, the number of students per enrollment year is 80, 78, 77, 80, and 79, respectively.

In this study, we predict the proficiency for aspect 3 from the second term of the first year to the third term of the third year using the learner data presented in the previous section. The output of the machine learning model is for aspect 3. All input and output data are scaled to the range [0,1], and data for learners with missing values are excluded.

C. Experimental Results

D. First Stage of Classification Using Machine Learning Models

The multiple regression model uses thresholds $\theta = 0.6, 0.7, 0.8$; RF uses $\theta = 0.9$, with 10 decision trees, and tree depth set to 6. After predicting the classification for the third term of the third year, we perform predictive classification from the second term of the first year to the second term of the third year, calculating the percentage of learners who are consistently classified as either predictable or difficult to predict compared to the third term of the third year. The results are shown in Tables I and II, here presenting the results for the third term of the second year. Rows 3 and 4 of the tables show the number of learners classified as difficult to predict or predictable for each term in the training, validation, and test data, respectively; row 5 shows the number of learners who were consistently classified in the same group as in the third term of the third year, and row 6 shows this number as a percentage (row 5 / row 3). RF, which shows a high match rate for difficult-to-predict learners in validation and test data, is considered superior.

E. Second Stage of Classification Using Machine Learning Models

In the second stage of classification, we predict for the third term of the second year using machine learning models and classify learners based on prediction errors. FNN settings include a learning rate of 0.0001 and a batch size of 2; decision trees have a depth of 7; RF uses 13 trees with a depth of 2; and XGBoost uses 6 trees with a depth of 5. Results are displayed in Tables III–VI, showing the mean squared error and the number of learners with prediction errors above 0.1 and 0.2 for training, validation, and test data, respectively. The experimental results indicate that RF, having the fewest learners with prediction errors above 0.1 in validation and test data, was the most suitable for modeling learners difficult to predict in the first stage of classification.

References

[1] Ministry of Education, Japan, *The image of the transformation of learning brought by "1 device for 1 student with a high-speed network"*. `https://www.mext.go.jp/en/content/20200716-mxt _ kokusai-000005414 _ 04.pdf` (access 2024/5/12)

[2] O. Latrellis, L. K.Savvas, P. Fitsilis and V. C .Gerogiannis: "A two-phase machine learning approach for predicting student outcomes," *Education and Information Technologies,* 26, pp.69–88, 2021.

[3] S. B. Keser and S. Aghalarova: "HERA: A novel hybrid ensemble learning algorithm for predicting academic performance of students," *Education and Information Technologies,* 27, pp.4521–4552, 2022.

[4] T. Hayashida, T. Yamamoto, S. Wakitani, T. Kinoshita, I. Nishizaki, S. Sekizaki and Y. Tanimoto: "Feature extraction and classification of learners using neural networks," *Proceedings of IEEE Frontiers In Education Conference,* 2018.

TABLE I
$\theta = 0.8$, PREDICTION CLASSIFICATION FOR 2G3T(*) (1ST STAGE: MULTIPLE REGRESSION MODEL)

	Training data		Validation data		Test data	
Predictable	NO	YES	NO	YES	NO	YES
Classification (2Y3G)	87	161	30	33	50	28
Classification (3Y3G)	55	193	31	32	45	33
Number of data of 2Y3T matching 3Y3T	23	129	17	19	30	13
Match rate	26.4%	80.1%	56.7%	57.6%	60.0%	46.4%

(*)The 3rd term of the 2nd year

TABLE II
$\theta = 0.9$, PREDICTION CLASSIFICATION FOR 2G3T (1ST STAGE: RF MODEL)

	Training data		Validation data		Test data	
Predictable	NO	YES	NO	YES	NO	YES
Classification (2Y3G)	47	201	32	31	56	22
Classification (3Y3G)	50	198	35	28	59	19
Number of data of 2Y3T matching 3Y3T	11	162	19	15	44	7
Match rate	23.4%	80.6%	59.4%	48.4%	78.6%	31.8%

TABLE III
PREDICTION RESULTS FOR ASPECT 3 USING FNN (2ND STAGE)

	Training data	Validation data	Test data
Mean squared error	0.0737	0.0428	0.0361
Learners difficult to predict (prediction error above 0.1)	41	21	44
Learners difficult to predict (prediction error above 0.2)	29	13	13

TABLE IV
PREDICTION RESULTS FOR ASPECT 3 USING DECISION TREE (2ND STAGE)

	Training data	Validation data	Test data
Mean squared error	0.0000	0.1352	0.0913
Learners difficult to predict (prediction error above 0.1)	0	26	36
Learners difficult to predict (prediction error above 0.2)	0	21	27

TABLE V
PREDICTION RESULTS FOR ASPECT 3 USING RF (2ND STAGE)

	Training data	Validation data	Test data
Mean squared error	0.0422	0.0449	0.0347
Learners difficult to predict (prediction error above 0.1)	38	18	34
Learners difficult to predict (prediction error above 0.2)	19	14	14

TABLE VI
PREDICTION RESULTS FOR ASPECT 3 USING XGBOOST (2ND STAGE)

	Training data	Validation data	Test data
Mean squared error	0.0072	0.0595	0.0685
Learners difficult to predict (prediction error above 0.1)	8	24	44
Learners difficult to predict (prediction error above 0.2)	2	15	29

[5] S. Qu, K. Li, Z. Han, S. Wu, X. Liu, Z. Huang and B. Wu, "Behavior pattern and compiled information based performance prediction in MOOCs," *Proceedings of IEEE Frontiers In Education Conference,* 2019.

[6] K. Hirotani, T. Hayashida, K. Tsutsumi, S. Wakitani, T. Kinoshita, I. Nishizaki, and S. Sekizaki, "Learner classification method using SOM and FNN," *The Proceedings of 2022 Annual Conference on Electronics, Information and Systems Institute of Electrical Engineers of Japan (IEEJ)* , TC1-3, pp. 10–15, 2022. (in Japanese)

[7] L. Breiman, "Random Forests," *Machine Learning,* 45, pp. 5–32, 2001.

[8] S. Archana and K. Elangovan, "Survey of Classification Techniques in Data Mining," *International Journal of Computer Science and Mobile Applications,* 2, pp. 65–71, 2009.

[9] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," *In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining,* pp. 785–794, 2016.

[10] T. Hayashida, T. Kinoshita, S. Wakitani, T. Yamamoto, I. Nishizaki, S. Sekizaki, and Y. Tanimoto, "Construction of a learner growth model and learner classification method using exMCRNNs," *Journal of Institute of Electrical Engineers of Japan Transactions on Electronics, Information and Systems,* 141, pp. 273–280, 2021.(in Japanese)

[11] A. Jain, and S. Solanki, "An Efficient Approach for Multiclass Student Performance Prediction based upon Machine Learning," *Proceedings of International Conference on Communication and Electronics Systems,* 978-1-7281-1261-9, 2019.