การศึกษาปัจจัยและการพัฒนาระบบพยากรณ์ด้วยการใช้เทคนิคการเรียนรู้ของเครื่องเพื่อวิเคราะห์การเรียนเขียนโปรแกรมคอมพิวเตอร์สำหรับนักเรียนมัธยมศึกษาตอนต้น*

วุฒิชัย พิลึก[1] นนทชัย สามงามจันทร์[2]

(วันที่รับบทความ: 31 มกราคม 2567; วันแก้ไขบทความ: 13 มีนาคม 2567; วันตอบรับบทความ: 2 เมษายน 2567)

บทคัดย่อ

การวิจัยครั้งนี้มีวัตถุประสงค์ 1) เพื่อศึกษาปัจจัยที่ส่งผลต่อการเรียนเขียนโปรแกรมคอมพิวเตอร์ของนักเรียนระดับมัธยมศึกษาตอนต้น 2) เพื่อศึกษาเทคนิคการเรียนรู้ของเครื่องในการวิเคราะห์ปัจจัยในการเรียนเขียนโปรแกรมคอมพิวเตอร์ และ 3) เพื่อพัฒนาระบบพยากรณ์การเรียนเขียนโปรแกรมคอมพิวเตอร์ ข้อมูลที่ใช้แบ่งเป็นการเก็บรวบรวมข้อมูล 2 ครั้ง ครั้งแรกได้มาจากนักเรียน จำนวน 411 คน เพื่อสำรวจปัจจัยเชิงสำรวจและครั้งที่ 2 ได้มาจากการสำรวจจากนักเรียน จำนวน 1,225 คน เพื่อนำมาทดสอบเทคนิคการเรียนรู้ของเครื่อง วิธีการวิเคราะห์ข้อมูลเบื้องต้นกระทำโดยการศึกษา วิเคราะห์เอกสาร และการคัดกรองโดยผู้เชี่ยวชาญ 1 รอบ และนำไปวิเคราะห์องค์ประกอบเชิงสำรวจ (EFA) และนำไปวิเคราะห์ด้วยเทคนิคการเรียนรู้ของเครื่อง 9 เทคนิค ได้แก่ Naïve Bay (NB), Support Vector Machine (SVM), Decision Tree (DT), Random Forest (RF) (RF), Logistics Regression (LR), Gradient Boosting (GB), Extreme Gradient Boosting (XGB), and Artificial Neural Network (ANN) หลังจากนั้นนำมาออกแบบระบบพยากรณ์การเรียนเขียนโปรแกรมคอมพิวเตอร์ สถิติที่ใช้ในการวิเคราะห์ ได้แก่ ค่าเฉลี่ย ส่วนเบี่ยงเบนมาตรฐาน ค่าสัมประสิทธิ์สหสัมพันธ์และอัลกอริทึมการเรียนรู้ของเครื่อง

ผลการวิจัยพบว่า

1. การวิเคราะห์องค์ประกอบเชิงสำรวจและค่าสัมประสิทธิ์สหสัมพันธ์ พบว่าปัจจัยที่มีความสัมพันธ์ต่อการเรียนเขียนโปรแกรม ประกอบด้วย 4 ปัจจัย ได้แก่ 1) ปัจจัยด้านผู้สอน 2) ปัจจัยด้านสภาพแวดล้อมทางการเรียนรู้ 3) ปัจจัยด้านสื่อการจัดการเรียนรู้ และ 4) ปัจจัยด้านผู้เรียน

2. ผลวิเคราะห์เทคนิคการเรียนรู้ของเครื่อง 9 เทคนิคในการพยากรณ์ผลการเรียนเขียนโปรแกรมจากปัจจัย 4 ปัจจัย พบว่า เทคนิค Random Forest (RF) ให้ค่า Accuracy 96.00 % และ Macro F1 ที่ 96.00 %

3. การออกแบบระบบพยากรณ์การเรียนเขียนโปรแกรมคอมพิวเตอร์ มีการออกแบบ Use Case, Data Flow Diagram, ER Diagram, Class Diagram และใช้โมเดลวิเคราะห์พยากรณ์ด้วยเทคนิค Random Forest (RF) โดยมีโปรแกรมที่ใช้ ได้แก่ Visual Studio Code, Xmapp การใช้งานระบบมีเมนูสำหรับผู้ใช้งาน 4 กลุ่ม

คำสำคัญ: เทคนิคการเรียนรู้ของเครื่อง, การเรียนเขียนโปรแกรม, นักเรียนมัธยมศึกษาตอนต้น

---

* บทความวิจัย สาขาวิชาคอมพิวเตอร์ คณะครุศาสตร์ มหาวิทยาลัยราชภัฏนครสวรรค์, 2566

[1] ผู้ช่วยศาสตราจารย์, สาขาวิชาคอมพิวเตอร์ศึกษา คณะครุศาสตร์ มหาวิทยาลัยราชภัฏนครสวรรค์, E-mail: wudhijaya@nsru.ac.th

[2] นักศึกษาบัณฑิตศึกษา, สาขาวิชา STEM Education, International College, National Pingtung University, ประเทศไต้หวัน, E-mail: bib111003@nptu.edu.tw

A Study on Factors and Development of Prediction System by Using Machine Learning for Learning Analytics in Computer Programming of Junior High-school Students*

***Wudhijaya Philuek*[1] *Nontachai Samngamjan*[2]**

(Received: January 31, 2024; Revised: March 13, 2024; Accepted: April 2, 2024)

Abstract

The objectives of this research were 1) to study factors that affect computer programming learning among middle school students, 2) to study machine learning techniques to analyze factors in learning computer programming, and 3) to develop a system for predicting computer programming learning. The data used was divided into 2 times. The first time was from 411 students to explore exploratory factors and the second time was from a survey of students with total of 1,225 students which used to test machine learning techniques. The primary data analysis method was done by study, analyze documents and 1 round of screening by experts and used for Exploratory Factor Analysis (EFA) and analyzed with 9 machine learning techniques including Naïve Bay (NB), Support Vector Machine (SVM), Decision Tree (DT), Random Forest (RF), Logistics Regression (LR), Gradient Boosting (GB), Extreme Gradient Boosting (XGB), and Artificial Neural Network (ANN), then analyze by machine learning techniques to design a prediction system for learning computer programming. Statistics used in the analysis included Mean, Standard Deviation, Correlation Coefficients and Machine Learning Algorithms.

The research results found that:

1. Exploratory factor analysis (EFA) and correlation coefficient showed 4 components: 1) Teacher factors, 2) Learning environment, 3) Learning management media factors and 4) Learner factors.

2. Results of the analysis of 9 machine learning techniques for predicting programming learning results from 4 factors. It was found that the Random Forest (RF) technique gave an Accuracy of 96.00 % and Macro F1 of 96.00 %.

3. Design of a computer programming learning prediction system showed a design process such as Use Case, Data Flow Diagram, ER Diagram, Class Diagram and uses the Random Forest (RF) technique in prediction, with program used including Visual Studio Code, Xmapp system usage had menus for 4 user groups.

Keywords: Machine Learning Techniques, Learning Computer Programming, Junior High-School Students

---

* Research Article, Computer Education Division, the Faculty of Education, Nakhon Sawan Rajabhat University, 2023

[1] Assistant Professor, Computer Education Division, the Faculty of Education, Nakhon Sawan Rajabhat University, E-mail: wudhijaya@nsru.ac.th

[2] Graduate Student, STEM Education Division, International College, National Pingtung University, The Republic of China (ROC), Taiwan, E-mail: bib111003@nptu.edu.tw

Introduction and Problem Statement

Computing Science in Thailand involves learning thinking processes consisting of 4 skills: 1) Problem Decomposition, 2) Pattern Recognition, 3) Abstract Thinking, and 4) Algorithmic Thinking. Each skill can be done from a young age by laying down a simple foundation before becoming more complex as one grows up. Computing Science can be learned from the kindergarten level because Computational Thinking in young children does not require the use of computers, but rather teaches children to use step-by-step planning (Institute for the Promotion of Teaching Science and Technology, 2018). In an era filled with information, more information gives rise to a concept called Big Data and one of them is a technique that can automatically find patterns from large amounts of data, called "Data Mining" (Phongsuk, 2017). In addition, focusing on Education, Sakdulyatham (2010) presented the application of Data mining techniques to predict student academic. The results can be used as guidelines for improving student learning such as developing teaching and learning arrangements. Using techniques for predicting academic performance in various fields of study to see which academic results in required majors are good or mediocre. Data mining is the process of finding all the relationships and patterns that actually exist in a database but are hidden within large amounts of data. Data mining has many different techniques used to solve problems, each of which has specific properties and the selection of techniques is therefore necessary to lead to the best solution. of data mining (Linoff and Berry, 2004) and Giudici (2003) defined data mining as the process of selecting and exploring data as well as creating a model of data to find patterns and find relationships from large amounts of data in order to obtain useful results. There are many techniques for classifying data. It is used such as K-Nearest Neighbors, Logistic Regression, Decision Tree, Support Vector Machine, and Naïve Bayes (Mitchell, 1997; Kotsiantis, 2007; 1997; Bastanlar and Ozuysal, 2014; Sara, 2007) etc. and the tools used to measure the efficiency of data classification are various methods, each of which shows the effectiveness of the technique for measuring the overall classification accuracy of all layers in that technique cannot be used to determine the true performance of techniques with unbalanced data. From the importance of the above, it is necessary to train learners to become more clearly visible in their practical competencies through learning programming language.

Research Objectives

1. to study factors that affect computer programming learning
2. to study machine learning techniques to analyze factors in learning computer programming
3. to develop a system for predicting computer programming learning

Concepts and Related Theories

Machine learning algorithms are models that can learn several methods and automatically finds similarities between data points and outputs the same type of data in groups such as Supervised Learning (Kotsiantis, 2007), Unsupervised Learning (Bastanlar and Ozuysal, 2014) and Reinforcement Learning (RL) (Kittiphon, 2019) which models or techniques in machine learning such as Decision Tree (Quinlan, 1992). Decision Tree algorithms have been used to predict student academic performance in various educational context (Philuek et. al. 2022; Philuek et al, 2021). Wang, Yao and Li (2022) proposed a method for analyzing student academic performance using algorithms and showed that Decision Tree methodology yields good performance analysis which Wang (2022) also found positive results in terms of improved running time and accuracy. Naive Bayes (NB) is a probabilistic clustering technique based on Bayes' Theorem and assumptions that determine the occurrence of events used in clustering. The independent classification learning method of Naive Bay is widely used in machine learning research. In many studies, Naive Bayes has been used to predict student academic performance. It has been found to be up to 78% accurate in predicting admission of students to universities is based on basic and pre-entry information (Naidoo, Jadhav, Sixhaxa, and Ajoodha, 2023). Another study used Naive Bayes to analyze and predict student achievement with an accuracy of 60% (Pandiangan, Lintang and Priyudahari, 2022). Naive Bayes is also used to classify students' academic data and predict the duration of their studies. As a result, the accuracy increased from 90% to 92.94% after feature selection (Nuraeni, Agustin, Rahayu, Kurniadi, Septiana and Lestari, 2021). K-Nearest Neighbor (KNN) is a method that classifies data based on the k closest neighbors, which is a method of grouping new observations that are closest to the learning data set (Duda et al. 2001). Also, Logistic Regression (LR) is used in predicting student performance (Sagala et al, 2022; Nurmalitasari and Purwanto, 2022). It is a mathematical modeling method that helps determine the relationship between a binary dependent variable (student performance) and one or more independent variables (such as student demographics, achievement data, and activity data) (Pande, 2023). Random Forest (RF) is a machine learning algorithm that has been applied in the context of educational data mining (EDM) to assess student academic achievement (Begum and Padmannavar, 2023). In one study, Random Forest was compared with other classification algorithms, the results showed that Random Forest performed well in the classification task (Chenghao, 2022). Random Forest was used as one of the machine learning algorithms, and it achieved in predicting attrition (Adane, Deku, and Asare, 2023). Additionally, while deep learning found better features,

Random Forest achieved higher accuracy in predicting students' performance (Sani, Oladipo, Ogbuju and Agbo, 2022). Gradient Boosting (GB) algorithms have been used in predicting student performance (Mashagba, Al-Saqqar and Al-Shatnawi, 2023). The proposed approach in one study implemented XGBoost, CaBoost, and LightGBM algorithms to analyze student academic performance (Chen and Zhai, 2022). The results showed that the CaBoost algorithm achieved the best prediction results for student final status and student appreciation models, with a prediction accuracy of 92.16% and 86.89% respectively (Agrawal and Narain, 2022). Another study utilized XGBoost for student performance prediction and found that it outperformed five other classic machine learning models (Agrawal and Narain, 2022). Extreme Gradient Boosting (XGB) has been used in several studies to predict student performance. Niels de Winter proposed a Modified XGBoost (MXGB) model that incorporates stream-based analysis to enhance its performance on real-time data (Nadar, 2023). Jiyoung and Meounggun (2023) mentions that the XGboost algorithm was applied to predict student dropout status, not student performance. The paper states that the Extreme Gradient Boosting (XGBoost) model can predict student academic performance with an accuracy of 97.12%. (Ojajuni et al, 2021). Artificial Neural Network (ANN) have been used in predicting student performance (Chavez et al, 2023).

Research Methodology

Population, Sample Group and Instrument: due to the changed of core curriculum in teaching Computer which transform to Computing Science subject the Junior High-school Students (Grade 7-9) was used as population in this research. The sample consisted of Junior High-school Students (Grade 7-9) who studied in schools under the Secondary Education Service Area of Nakhon Sawan, Uthai Thani, Chainat and Phichit Province in Thailand which used to answer research objectives (411 students for the first research objective and 1,225 students for second research objective) and 10 experts is used for the first research objective as well. The Step One was to studied factors which affecting learning computer programming. The researcher used documents, research, and academic papers which conducted in many countries including Thailand (Aunthadet, 2010; Mad-adam, 2016; Kornkao, 2015; Na Thaluatong et. al, 1997), Australia (Donna, Christine McDaniel. 2015), United States (Abdunabi, Hbaei and Yu Ku. 2019), Cyprus (O¨zden and Tezer, 2018), Finland (Lahtinen, Ala Mutka and Ja¨rvinen, 2005), and Malaysia (MD Derss and Mohamad Ali, 2012) to determined factors which affected student learning the programming which can be divided into 9 groups of factors: 1) teacher factors (43 items), 2) learning media factors (50 items),

3) teaching and learning factors (9 items), 4) programming skills factors (80 items), 5) teaching techniques factors (14 items), 6) learning environment factors (35 items), 7) learner factors (102 items), 8) programming factors (16 items), and 9) feeling about programming factors (5 items), totaling 354 items, then collected and interpreted and asked English language experts to check the accuracy of the meaning of each question. Bring corrections according to experts' suggestions in questionnaire and sent to 10 experts to reviewed and cut off. Experts have eliminated extraneous questions and improved them. There were 4 groups remaining: teacher factors (17 items), learning environment factors (9 items), learning media factors (9 items), and learner factors (15 items), totaling 50 questions, the researcher used factor analysis by using Exploratory Factor Analysis (EFA) and Pearson correlation coefficient.

Data Collection and Analysis: the researcher operates in accordance with the data mining development process standards. Cross-Industry Standard Process for Data Mining (CRISP-DM), which provides a research process. The work for each step is as follows; 1) Understanding the problem, 2) Data Understanding, 3) Data Preparation, the information is divided into three sub-steps [3.1) Data Selection, 3.2) Data Cleaning, 3.3) Data Transformation], 4) Modeling Phase [At this stage, the researcher selected the following machine learning techniques and algorithms: Naïve Bay (NB), Support Vector Machine (SVM), Decision Tree (DT), Random Forest (RF), Logistics Regression (LR), Gradient Boosting (GB), Extreme Gradient Boosting (XGB), and Artificial. Neural Network (ANN) with 80% for training and 20% for testing], 5) Evaluation Phase, and 6) Deployment Phase. In this research, the researcher has studied the techniques or processes of system development and summarizes the results of using the SDLC Model (Atiwitthayaporn, 2011) with 7 steps:

1. Requirement Definition; presenting the factors that contribute to the learning process in learning computer programming (Samngamjan et al, 2021) and designing Use Case, Data Flow Diagram, ER Diagram, Class Diagram for Web Application.

2. System Analysis, the researcher analyzed the requirements of the web application system using a diagram showing the user action of the system (Use Case), a Data Flow Diagram, and ER Diagram (Entity-Relationship Diagrams).

3. System Design, database design of database structure with data mining methods for junior high school students and data relations

4. System Development; 4.1) programming learning environment survey including various factors, 4.2) the tools used are: 4.2.1) computer and hardware - CPU Intel® Core™ i5-6300HQ Processor

2.30 GHz (3M Cache), 8 GB RAM, 500 GB Hard Disk, 14" display, 4.2.2) data analysis techniques-Random Forest Techniques, 4.2.3) program or software - XAMPP, Visual Studio Code, Visio, Microsoft Windows 10 operating system, MySQL database management system, Google Chrome Browser used to test the system after uploading to the server, Web Application development tools such as Visual Studio Code. XAMPP program Notepad++, Hosting Digital Ocean for real server testing, writing and system development based on the analysis and design of the system. Students use XAMPP to display web applications and learning outcomes in Various parts which require periodic testing of the system to find faults and develop the system to meet the most demanding requirements.

5. System Testing.

6. System Implement, installing the system using PHP HTML5 programming language and Bootstrap as a Frontend Framework together as a tool to develop the system with MySQL as a database management. Use Visual Studio Code to write code that is used to connect to MySQL.

7. System Maintenance, using web applications, system users will always encounter problems that arise from never using the system, so there is a solution if there is a problem with the web application.

Research Results

The results showed that, 1) The results of investigation the factors that affected learning computer programming of Junior High-school Students by using Exploratory Factor Analysis (EFA) and Pearson correlation coefficient showed that, there were 4 groups remaining: teacher factors (11 items), learning environment factors (4 items), learning media factors (5 items), and learner factors (11 items), totaling 31 items.

2) The results of a study of machine learning techniques to analyze factors in Junior High-school Students learning computer programming are shown in Table 1.

Table 1 Test results of LR, DT, RF, NB, SVM, KNN, GB, XGB, and ANN Techniques.

Model	Results	Class					Accuracy	Macro Avg
		0	1	2	3	4		
Logistic Regression (LR)	**Precision**	0.80	1.00	1.00	0.89	0.77	0.90	0.89
	Recall	0.76	1.00	1.00	0.96	0.76		0.90
Decision Tree (DT)	**Precision**	0.85	0.98	0.98	1.00	0.85	0.93	0.93
	Recall	0.88	1.00	0.98	0.98	0.81		0.93
Random Forest (RF)*	**Precision**	0.92	1.00	1.00	1.00	0.89	**0.96***	0.96
	Recall	0.92	1.00	1.00	0.98	0.91		0.96

Table 1 (Continued)

Model	Results	Class					Accuracy	Macro Avg
		0	1	2	3	4		
Naïve Bayes (NB)	Precision	0.63	0.97	0.71	0.81	0.73	0.77	0.77
	Recall	0.54	0.98	0.84	0.81	0.89		0.77
Support Vector Machine (SVM)	Precision	0.91	1.00	1.00	1.00	0.82	0.94	0.94
	Recall	0.83	1.00	1.00	0.98	0.91		0.94
K Nearest Neighbors (KNN)	Precision	0.85	0.89	0.96	0.96	0.84	0.90	0.90
	Recall	0.85	1.00	0.91	0.96	0.78		0.90
Gradient Boosting (GB)	Precision	0.93	1.00	1.00	1.00	0.85	0.95	0.95
	Recall	0.86	1.00	1.00	0.98	0.93		0.95
Extreme Gradient Boosting (XGB)*	Precision	0.94	1.00	1.00	1.00	0.85	**0.96***	0.96
	Recall	0.86	1.00	1.00	0.98	0.94		0.96
Artificial Neural Network (ANN)	Precision	0.89	1.00	0.98	1.00	0.83	0.94	0.94
	Recall	0.85	1.00	1.00	0.96	0.89		0.94

Table 1 shows test results of Logistic Regression (LR) shows accuracy at 90%, Decision Tree (DT) shows accuracy at 93%, Random Forest (RF) show accuracy at 96%, Naïve Bayes (NB) shows accuracy at 77%, Support Vector Machine (SVM) shows accuracy at 94%, K-Nearest Neighbors (K-NN) shows accuracy at 90%, Gradient Boosting (GB) shows accuracy at 95%, Extreme Gradient Boosting (XGB) shows accuracy at 96%, and Artificial Neural Network (ANN) shows accuracy at 94%, this can be summarized that the most accurate technique in data analysis is Random Forest (RF) and Extreme Gradient Boosting (XGB) while the weak technique is Naïve Bayes (NB), so the Precision, Recall, f1-score, and Support show in the Table above.

3) The development of system for predicting learning in Computer Programming for Junior High-school Students. Once logged in to the website and clicked on "Student Information" Select the grade level. The system will use the results of computer programming to calculate with data analysis of Random Forest Classifier technique to display the results of students in each grade, room, learner, respectively.

Table 2 Data taken into the Random Forest model

List	Variable	Description
1	name	name
2	Last name	surname

Table 2 (Continued)

List	Variable	Description
3	score	score
4	class	class
5	computer	computer grades
6	mathematics	Mathematics
7	Problem teacher	Problems in learning computer programs: instructor factor
8	Problem Learning Environment	Problems in learning computer programs: learning environment factors
9	Problem Learning materials	Problems in learning computer programs: learning media factors
10	Problem Student	Problems in learning computer programs: student factor

Table 2 shows the data created and stored in the survey database in MySQL.

Table 3 Results of Random Forest Technique Test

Class	precision	recall	f1-score	support
Level_1	1.00	1.00	1.00	12
Level_2	1.00	50	1.00	53
Level_3	1.00	50	1.00	18
accuracy	0	0	1.00	83
Macro avg	1.00	1.00	1.00	83
Weighted avg	1.00	1.00	1.00	83

From Table 3, it was found that the value from a group of factors related to learning to program found that the accuracy level was good at 1.00%, the memory at the good level was at 1.00% and the F1-score at the good level was at 1.00%. The accuracy was 1.00% and the support was moderate at 54%, and the accuracy was 83%.

Discussions and Recommendations

This study investigated factors and development of prediction system by using machine learning for learning analytics in Computer Programming of Junior High-school students, the results are showed in following.

1. The factors which affecting learning computer programming, the researcher used documents, research, and academic papers which conducted in many countries to determined factors which affected student learning the programming which at first round can be divided into 9 groups of factors: 1) teacher factors (43 items), 2) learning media factors (50 items), 3) teaching and learning factors (9 items), 4) programming skills factors (80 items), 5) teaching techniques factors (14 items), 6) learning environment factors (35 items), 7) learner factors (102 items), 8) programming factors (16 items), and 9) feeling about programming factors (5 items), totaling 354 items. Then, the second round, researcher collected and interpreted and asked English language experts to check the accuracy of the meaning of each question and sent to 10 experts to reviewed and eliminated extraneous questions and improved them. There were 4 groups remaining: teacher factors (17 items), learning environment factors (9 items), learning media factors (9 items), and learner factors (15 items), totaling 50 questions. The third round, researcher used factor analysis by using Exploratory Factor Analysis (EFA) and Pearson correlation coefficient, there were 4 groups remaining: teacher factors (11 items), learning environment factors (4 items), learning media factors (5 items), and learner factors (11 items), totaling 31 items. In teacher factors such as “a1 The instructor has a teaching method that makes the content easily understandable, clear, and creates an enjoyable atmosphere for learning”; “a2 The instructor has the ability to transfer knowledge in clear steps”, “a3 The instructor comes to teach on time”, “a4 Teachers have the ability to make students develop their thinking”. In learning environment factors such as “b1 The number of computers is sufficient and ready for use”, “b2 The computer classroom has appropriate lighting, such as not too bright or too dark, and no glare affecting the eyes”, and “b3 The school provides appropriate computer rooms for students”. In learning media factors such as “c1 The language used in the media is easy to understand”, “c2 has an easy-to-understand media user manual”, c3 learning media helps students have more knowledge and programming skills”. Learner factors such as “d1 I can write code blocks correctly”, “d2 I can understand the structure of programming” and “d3 I can debug complex programming errors”.

2. The results of machine learning techniques to analyze factors in Junior High-school Students learning computer programming shows that the most accurate technique in data analysis is Random Forest (RF) showed accuracy at 96% and Extreme Gradient Boosting (XGB) showed accuracy at 96% while the weak technique is Naïve Bayes (NB) showed accuracy at 77%, while the other techniques showed results such as Logistic Regression (LR) showed accuracy at 90%, Decision Tree (DT) showed accuracy at 93%, Support Vector Machine (SVM) showed accuracy at 94%, K-Nearest Neighbors (K-NN) showed accuracy at 90%, Gradient Boosting (GB) showed accuracy at 95%, Extreme Gradient Boosting (XGB) showed

accuracy at 96%, and Artificial Neural Network (ANN) showed accuracy at 94%. This may be because Random Forest is a model that takes several Decision Trees to train together (from 10 trees to more than 1000 trees), where each tree receives features and data that are a random subset of all features and data. When making predictions, each Decision Tree makes everyone's predictions and selects the final prediction such a technique is called Bagging or Boostrapping. Making predictions will become more and more accurate when the tree is learned continuously until it has enough depth, the model stops learning when there are no error patterns left from the previous tree to learn. Moreover, Random Forest algorithm has been widely used in various domains for learner analytics. It has proven to be effective in capturing complex relationships between features and target variables, providing valuable insights into different fields. In microbiology research, Chrobak, et. el. (2023) demonstrated the effectiveness of Random Forest in analyzing microbiological data, contributing to the understanding of microbial processes and interactions. In the context of science achievement prediction, Hong et al. applied Random Forest to extract factors associated with different science achievement groups, highlighting the importance of science activities, learning methods, and environmental awareness (Hong, et. al, 2022). In the field of digital forensics, Akanji et al. employed Random Forest as part of a predictive analytics system for image steganalysis, achieving accurate results in pattern recognition (Akanji et. al, 2022). Additionally, Random Forest has been utilized in the domain of active distribution networks, where it was used as a meta learner to improve the accuracy of pseudo-measurements for state estimation calculations (Radhoush, et al, 2023).

3. The develop a system for predicting learning in Computer Programming for Junior High-school Students showed that the systems in predicting the learning ability in learning computer programming using data mining methods for lower secondary school students. There is a design process such as designing a use case to show the user's work in the system. Relationships with subsystems within the system include users, students, teachers, directors, administrators. The student section can only use the information form via the link. Teachers can use the system to fill out surveys by log in. Director can use the system in the form of filling out surveys into the system. Administrators can use, modify, modify all parts of the system, including filling out surveys, logging in, maintaining system operation, designing Data Flow Diagram. showing the process of the system by the workflow of each individual user has access to the system in different sections. Students complete the survey which survey data will be sent to organize in database. Teachers must log in to the system before use which must be registered via the application web page when logging in. It will retrieve information from the user database when the teacher views the results of the form.

The system survey will retrieve data from survey database to represent. The director also use the system for everything like the teacher's, but will add a menu to view the results of teachers teaching students in each class system. School information and information sharing request must be filled out form and the system will send the information to the request base, use the system and notify the information to the administrator. Every database in the system contains a survey database, school database, user database. The linked database and the requested database use the system with data sharing and authorization to use. ER Diagram design shows how the system works by connecting the data in the web application as connected from the database in each database. The Class Diagram design shows all the data of the web application being stored. Each database has a primary key and a secondary key in each reference database, thus making the system for predicting the results of computer programming with data mining methods in the most efficient way. The development of computer programming learning forecasting for junior high school students using data mining methods to understand the problems and factors affecting learning programming. computer that each student has a problem in any aspect. The factors that affect learning programming can be seen from the results of each aspect of the survey through the web application. The system development uses Visual Studio Code by writing PHP language to create components to use the web app. The home page of the web application consists of the home page menu, survey menu, student information menu, teacher information menu and system request menu. Before using the system, it must be logged in before using the system every time. For example, when using it, the user must enter the survey menu, which will find a survey containing various factors affecting the results. Learning to write a computer program which, upon completion of the survey, will find a table showing student data and survey results for teachers, director and administrators, able to retrieve all student data or select individual students. This will be used to help analyze learners in finding an appropriate teaching-learning style.

References

Abdunabi, R., Hbaci, I, & Ku, H-Y. (2019). Towards Enhancing Programming Self-Efficacy Perceptions Among Undergraduate Information Systems Students. *Journal of Information Technology Education: Research*, 18, 185206. https://doi.org/10.28945/4308.

Adane, M. D., Deku, J. K. and Asare, E. K. (2023). Performance Analysis of Machine Learning Algorithms in Prediction of Student Academic Performance. *Journal of Advances in Mathematics and Computer Science*, 38, 74-86. 10.9734/jamcs/2023/v38i51762.

Agrawal, K. and Narain, B. (2022). A Bagging Algorithmic Approach for Prediction of School Student Performance. in *2022 4th International Conference on Advances in Computing, Communication Control and Networking (ICAC3N)*, 56-59.

Ali, M. B. and Zamzuri, A. T. (2012). Difficulties in Learning Programming: Views of Students. Retrieved March 24, 2023, from https://doi.org/10.13140/2.1.1055.7441.

Akanji, W. et al. (2022). A blind steganalysis-based predictive analytics of numeric image descriptors for digital forensics with Random Forest & SqueezeNet. in *2022 5th Information Technology for Education and Development (ITED),* Abuja, Nigeria, pp. 1-7.

Alnoman, A. (2023) Will the Student Get an A Grade? Machine Learning-based Student Performance Prediction in Smart Campus. in *2023 Advances in Science and Engineering Technology International Conferences (ASET)*, Dubai, United Arab Emirates, 2023, pp. 1-6.

Apriyadi, M. R., Ermatita, E. and Palupi, R.D. (2023). Hyperparameter Optimization of Support Vector Regression Algorithm using Metaheuristic Algorithm for Student Performance Prediction. *International Journal of Advanced Computer Science and Applications*, 14, 144-150.

Atiwitthayaporn, J. (2011). *Knowledge Management for Managing Information Systems Using Databases SQL in Educational Institutions*. Thaksin University: Songkhla. (In Thai).

Aunthadet, N. (2010). Development of a Virtual Classroom on the use of Computers in Business According to Constructivist Principles for Undergraduate Students in the Faculty of Business Administration Rajamangala University of Technology Isan. in *Academic conference presenting work 23rd National Graduate Research Conference, Rajamangala University of Technology Isan*, 23-24 December 2011, pp. 935-940. (In Thai).

Chavez, H., Chavez-Arias, B., Contreras-Rosas, S. and Alvarez-Rodríguez, J.M. (2023). *Artificial Neural Network Model to Predict Student Performance Using Nonpersonal Information*. Frontiers in Education. https://doi.org/8.10.3389/feduc.2023.1106679.

Chen, Y. H., and Zhai, L. (2022). XGBoost-Based Student Performance Prediction in Tiered Instruction. *Fourth International Conference on Computer Science and Educational Informatization (CSEI 2022)*.

Chenghao, Y. (2022). Research on Student Academic Performance Prediction Methods. *Highlights in Science, Engineering and Technology*, 24:257-263. https://doi.org/10.54097/hset.v24i.3940

Chrobak, D., Kołodzieczak, M., Kozlovska, P., Krzemińska, A., & Miller, T. (2023). Leveraging Random Forest Techniques for Enhanced Microbiological Analysis: A Machine Learning Approach to Investigating Microbial Communities and Their Interactions. *Scientific Collection «InterConf+»,* (32(151), 386-398.

Giudici, P. (2003). *Applied Data Mining: Statistical Methods for Business and Industry*. Hoboken: Wiley.

Hong, J., Kim, H., & Hong, H. (2022). Random Forest Analysis of Factors Predicting Science Achievement Groups: Focusing on Science Activities and Learning in School. *Asia-Pacific Science Education*, 8(2), 424-451.

Jiyoung, M. and Meounggun, J. (2023). Applying Machine Learning-Based Models to Prevent University Student Dropouts. *Gyo'yug pyeong'ga yeon'gu*, https://doi.org/10.31158/jeev.2023.36.2.289

Kornkao, N. (2015). *A Study of the Learning Environment of Sattha Samut School Students under the Jurisdiction of the Secondary Educational Service Area Office Area 10*. Master of Education Thesis, Burapha University. (In Thai).

Kotsiantis, S. B. (2007) Supervised Machine Learning: A Review of Classification Techniques. *Informatica (Ljubljana)*, 31: 249-268

Lahtinen, E., Ala-Mutka, K. and Järvinen, H. (2005). A Study of the Difficulties of Novice Programmers. *Annual Conference on Innovation and Technology in Computer Science Education. ACM SIGCSE Bulletin*. https://doi.org/37. 14-18. 10.1145/1067445.1067453.

Linoff, G. and Berry, M. (2011). *Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management*. (3rd ed.). Hoboken, New Jersey: Wiley.

Mad-adam, U. (2016). Developing Computer Programming Skills Using Skill Training Kits Computer Programming Using Cooperative Learning Methods for Second-Year Software Engineering Students. Walailak University Library. *Journal Thaksin University*. (In Thai).

Mashagba, E., Al-Saqqar, F. and Al-Shatnawi, A. (2023). Using Gradient Boosting Algorithms in Predicting Student Academic Performance. in *2023 International Conference on Business Analytics for Technology and Security (ICBATS)*, 1-7.

McDaniel, D.C. (2015). *Effectiveness of Integrating Technology Across the Curriculum: Classroom Learning Environments Among Middle-School Students in the USA*. Doctoral Dissertation of Curtin University, Western Australia, Australia.

Nadar, N. (2023). Enhancing Student Performance Prediction Through Stream Analysis Dataset Using Modified XGBoost Algorithm. *International Journal on Information Technologies and Security*, 15, 75-86. https://doi.org/10.59035/KNUG1085.

Naidoo, J. T., Jadhav, A., Sixhaxa, K. and Ajoodha, R. (2023). Using Student Characteristics to Promote Student Success at Higher-education Institutions. In: Kumar, S., Hiranwal, S., Purohit, S.D., Prasad, M. (eds.), *Proceedings of International Conference on Communication and Computational Technologies. Algorithms for Intelligent Systems*. Singapore: Springer.

Na Thaluatong, et. al. (1997). *Guidelines For Developing the Teaching of Computer Subjects in Secondary Schools*. Chulalongkorn University, Bangkok. (In Thai).

Nuraeni, F., Agustin, Y.H., Rahayu, S., Kurniadi, D., Septiana, Y. and Lestari, S.M. (2021) Student Study Timeline Prediction Model Using Naïve Bayes Based Forward Selection Feature. in *2021 International Conference on ICT for Smart Society (ICISS), Bandung, Indonesia, 2021*, pp. 1-5, https://doi.org/ 10.1109/ICISS53185.2021.9532502

Nurmalitasari, N. and Purwanto, E. (2022). Prediksi Performa Mahasiswa Menggunakan Model Regresi Logistik. *Jurnal Derivat: Jurnal Matematika & Pendidikan Matematika*, https://doi.org/ 10.31316/jderivat.v9i2.2639

Ojajuni, O. P., Ayeni, F., Akodu, O., Ekanoye, F., Adewole, S., Ayo, T., Misra, S. and Mbarika, V. W. (2021). Predicting Student Academic Performance Using Machine Learning. *Computational Science and Its Applications-ICCSA 2021,* (pp. 481-491).

Özden, C. and Tezer, M. (2018). The Effect of Coding Teaching on Students' Self-Efficacy Perceptions of Technology and Design Courses. *Sustainability*, 10(10), 3822. https://doi.org/10.3390/su10103822

Pande, S. M. (2023). Machine Learning Models for Student Performance Prediction, in *2023 International Conference on Innovative Data Communication Technologies and Application (ICIDCA)*, Uttarakhand, India, pp. 27-32, https://doi.org/10.1109/ICIDCA56705.2023.10099503.

Pandiangan, N., Lintang, M. and Priyudahari, B. (2022). Naïve Bayes for Analysis of Student Learning Achievement. *SHS Web of Conferences*. 149, 2022.

Philuek, W., Pongsuk, T., and Panawong, N. (2022). Machine Learning Techniques for Classifying Self-Regulated Learning of Secondary Students in Thailand. *Journal of Positive School Psychology*, 6(4), 6429-6436.

Philuek, W., Sornchai, S., Raksarikorn, T., and Janyarat, S. (2021). Machine Learning Techniques for the Ordinary National Educational Test (O-NET) Prediction: Case of Small Sized Schools in Nakhon Sawan Province, Thailand. *Natural Volatiles and Essential Oils*, 8(6), 2821-2833.

Phongsuk, T. (2017). *A Study of Data Mining Techniques to Predict Academic Performance and Learning Behavior of Students in the Computer Studies Department at Northern Universities*. Undergraduate Thesis (Unpublished) Nakhon Sawan Rajabhat University. (In Thai).

Quinlan, J. R. (1992). *C4.5 Programs for Machine Learning*. San Mateo, CA: Morgan Kaufmann.

Radhoush, S., Vannoy, T., Whitaker, B. M. and Nehrir, H. (2023). Random Forest Meta Learner for Generating Pseudo-Measurements in Active Distribution Power Networks. in *2023 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT)*, (pp. 1-5). Washington, D.C., https://doi.org/10.1109/ISGT51731.2023.10066389.

Safira, B. and Padmannavar, S. S. (2023). Student Performance Analysis using Bayesian Optimized Random Forest Classifier and KNN. *International Journal of Engineering Trends and Technology,* 71(5), 132-140.

Sagala, T. M. N., Permai, S. D., Gunawan, A. A. S., Barus, R. O. and Meriko, C. (2022) Predicting Computer Science Student's Performance using Logistic Regression, in *2022 5th International Seminar on Research of Information Technology and Intelligent Systems (ISRITI), Yogyakarta, Indonesia, 2022,* (pp. 817-821), https://doi.org/10.1109/ISRITI56927.2022.10052968.

Sakdulyatham, R. (2010) *Using Data Mining Techniques to Build a Knowledge Base to Predict Academic Achievement of Rajapruek College Students*. Rajapruek College, NontaBuri. (In Thai).

Samngamjan, N., Philuek, W., Malangpoo, P. and Janyarat, S. (2021) A Study on Factors and Using Machine Learning for Learning Analytics in Computer Programming of Junior High-School Students. *Natural Volatiles and Essential Oils*, 8(6), 1224-1233.

Sani, G., Oladipo, F. O., Ogbuju, E. and Agbo, F. (2022). Development of a Predictive Model of Student Attrition Rate. *Journal of Applied Artificial Intelligence*. 3, 1-12.

Sara, C. (2007). *Classification of Graduation Status Groups Using Decision Tree Modeling*. Bangkok: King Mongkut's University of Technology North Bangkok.

Sunarko, B. et al. (2022) Prediction of Student Satisfaction with Academic Services Using Naive Bayes Classifier," 2022 6th International Conference on Information Technology, in *Information Systems and Electrical Engineering (ICITISEE), Yogyakarta, Indonesia*, (pp. 228-233), https://doi.org/10.1109/ICITISEE57756.2022.10057736.

Wang, P. (2022). A Study of Student Performance Under English Teaching Using a Decision Tree Algorithm. *Journal of Control and Decision*, 10(3), 417-422.

Wang, N., Yao, M. and Li, J. (2022). Research on Student Achievement Analysis Method Based on Decision Tree Algorithm. in *2022 International Symposium on Advances in Informatics, Electronics and Education (ISAIEE)*, 352-355.

****