

Machine Learning Techniques for Classifying Self-Regulated Learning of Secondary Students in Thailand

¹Wudhijaya Philuek, ²Tannakon Pongsuk, ³Naruepon Panawong

¹Computer Education Program, Faculty of Education, Nakhon Sawan Rajabhat University, Mueang Nakhon Sawan District, Nakhon Sawan Province, 60000, wudhijaya@nsru.ac.th

²Computer Education Program, Faculty of Education, Nakhon Sawan Rajabhat University, Mueang Nakhon Sawan District, Nakhon Sawan Province, 60000

³Department of Applied Science, Faculty of Science and Technology, Nakhon Sawan Rajabhat University, Mueang Nakhon Sawan District, Nakhon Sawan Province, 60000

Abstract

The objectives of this research were 1) to proposed the Machine Learning (ML) techniques for classifying students' Self-Regulated Learning characteristics, and 2) to test Machine Learning (ML) techniques based on students' Self-Regulated Learning characteristics. The research conducted by analyzing factors on Self-Regulated Learning theory, synthesize relevant Machine Learning (ML) techniques, and analyze data by using Frequency distribution and Machine Learning (ML) techniques. This research has given Human Ethics Approval (HE-RDI-NRRU.046/2565) in research conduction. The finding showed that, 1) the Decision Tree, Artificial Neuron Network, Naive Bayes, Logistic Regression, and K-Nearest Neighbors were the appropriated techniques to use as data analytics techniques for analyzing and enhancing learning based on students' Self-Regulated Learning characteristics, 2) the most technique which gave the high efficacy is Decision Tree which researchers will use in development platform to predict Self-Regulated Learning of students in the future.

Keywords: Self-Regulated Learning, Machine Learning, Thai Secondary School Students.

INTRODUCTION

The Ministry of Education has set a policy to bring digital technology to help administration and education to support the digital government. In this regard, the balancing and development of the management system and the development of the educational database system as well as the development of the learners to have critical thinking skills that can effectively solve the immediate situation by organizing the learning process. Proactive from real experience or simulation will help to open the world and the perspective of the learner and the teacher more. It is imperative to develop learners to be knowledgeable and have life skills to be tools for living and creating careers, such as the use of digital technology for health.

Having a good attitude towards health care with the expectation that learners of all ages will be developed in all dimensions to be good people, competent people with quality and readiness to drive the country's development. towards stability, prosperity and sustainability. The educational data mining process is a technology that is gaining attention nowadays that expands the ability of information systems in traditional educational institutions to obtain information or knowledge that is used to explain or predict trends by processing phenomena. from the large amount of data stored in the system. Today, research in this field is becoming increasingly popular as it is clearly beneficial both for the administration of educational institutions and for providing academic services

to students, which are key corporate clients (Srisatchalertwacha and Prasertthitipong, 2014)

Self-Regulation can be defined as the ability to learn by self-direction, is that the individual has set goals in learning, has the ability to control himself to act according to plans, has an effort to observe his own behavior and has a reaction to himself, such as self-admiration when able achieving set goals, including arranging or adjusting the learning environment to suit. Therefore, those capable of self-directed learning will not only succeed in their present learning but will also make efforts to improve their learning in the future (Zimmerman, 1989).

In addition, Self-Regulation Learning (SRL) is also an important goal of education in the 21st century, an era of rapid technological change that seeks to develop individuals into lifelong learners capable of learning all the time and continually improving one's self, it is imperative that a person has the ability to learn self-directed in order to be able to learn effectively as Self-Regulation is an important aspect of the entire learning curve in education and self-development after graduation (Zimmerman, 2000)

In educational institutions, both public and private organizations take part in the management of policies and projects in the student analysis system, such as the Individualized Education Program: IEP, is an educational management plan for children with special needs. The components of the individual education plan include an individual learner analysis process that leads to a participatory planning process in educational management. consistent and appropriate to the special needs of learners.

Educational institutions can design their own analytical tools by analyzing data in areas such as family background data analysis, learning achievement data analysis, developmental data analysis, interests and talent data analysis, data analysis. Such data is an analysis of fundamental data and relevant factors for the purpose and promotion of the main activities of the project. Most of the analysis was done by

importing the data at the discretion of the teachers. The analytical system used to file documents with word processors, table programs, and use basic statistics to analyze them, thus creating a bias in data import, inconvenience in use, and may result in unintended analysis results. In addition, the introduction of fundamental data factors makes it impossible to analyze learners in depth due to lack of correlation of the data in the analysis of learners.

Therefore, the researcher studies and presents Machine Learning (ML) techniques for analyzing and promoting learning according to learners' learning supervision attributes. As a results, the output of this finding will be use in digital platform design for use in analyzing and promoting character-based learning, directing learner learning is a framework for developing a digital platform to analyze and promote character-based learning. Supervising the learning of learners for teachers and educational personnel to use as a tool for further analysis of individual learners.

Research Methodology

This study uses an analytical process. and synthesize related concepts, theories and research, as well as collecting data for analysis by Machine Learning (ML) techniques as follows:

1. Self-Regulated Learning characteristics

The researcher studied the theory of self-directed learning with the main theory. The 4 theories are conditional theory, cognitive development theory, social intelligence theory and information processing theory. The self-regulation process, which indicates the factors leading to the learning of each theory, was analyzed in Table 1 (Man-Chih, 2006; Philuek, 2017).

Table 1 *Summary table of the theory of Self-Regulated Learning characteristics (Man-Chih, 2006; Philuek, 2017).*

Characteristics	Operant Conditioning of Behaviorism	Cognitive Development of Cognitivism	Social-Cognitive Theory	Information Processing Theory
<i>Key to learning</i>	Reinforcement (Learners will link their learning with the stimuli of environment)	Self-Regulation (Learners will control over their thoughts, feelings, and actions)	Modeling and Observation (Learners will link their learning motivation, cognition, and environment)	Mental Processing (Learners will use metacognition: planning, monitoring, and evaluating to explain SRL and learning strategies)
<i>Regulating process</i>	Self-Monitoring Self-Instruction Self-Reinforcement	Observation Simulation Self-Control Self-Regulation	Self-Observation Self-Judgment Self-reaction	Select Organize Rehearsal Map

2. Measure the students' Self-Regulated Learning

To measure learning characteristics, the researcher applied the concept of self-regulated learning as a concept in the study because self-regulated learning affects learners' learning. For example, learners with low levels of self-directed learning have low chances of success in learning (Zimmerman, 1989; Dabbagh and Kitsantas, 2005; Philuek, 2017) in Motivated Strategies for Learning Questionnaire: MSLQ (Pintrich, 2004). It is a self-report model designed to assess learners' learning motivations and strategies. There are two parts to the questionnaire, the Motivation Questionnaire consists of 31 items to assess learners' goals and beliefs. The Learning Strategies section includes 31 questions assessing learners' use of learning strategies and metacognition, and 19 assessing learners' management of their time, resources, and tasks. The 7-level rating scale was then interpreted using Likert's mean, with a scale of 1 meaning not true, and level 7 being very true, respectively. The Learning Strategies and Motivation Questionnaire used in this research has been used for research in several countries in the study of learning motivation and strategies, including Australia (Fuller, 1999), China (Rao, Moely and Sachs, 2000), Japan (Yamauchi, Kumagai, and Kawasaki, 1999),

Taiwan (Lee, 1997) and Thailand (Booget, 2011; Philuek, 2017; Philuek, Puttasem, Janyarat, and Inchai, 2020). Research included student motivation and performance (Lin and McKeachie, 1999), motivation and learning strategies for academic achievement (DeKeyrel, Dernivish, Epperly and McKay, 2000; Pokay and Blumenfeld-Phyllis, 1990). Self-Learning and Web-Based Learning (McManus, 2000) and Teacher Education (McClendon, 1996; Edgar) and Chen, 2002)

3. Data Analytic Techniques

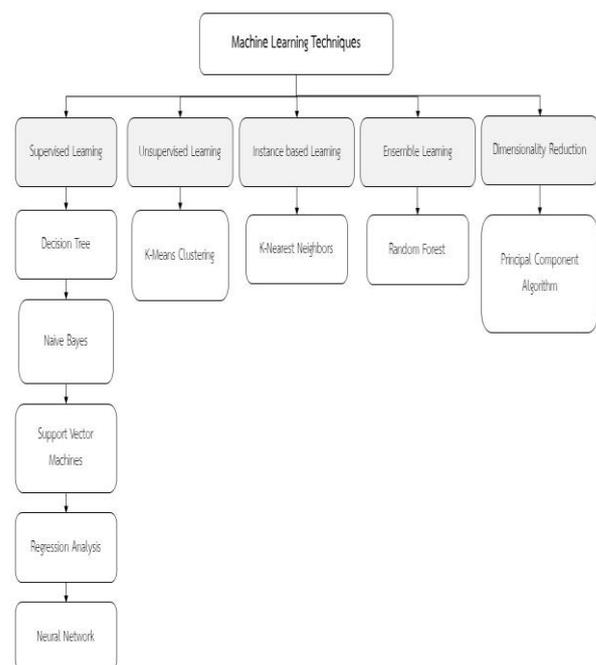


Figure 1. *Machine Learning techniques*

3.1 In data analysis with data mining techniques, there are generally several types of data analysis techniques, which categorizes techniques and techniques as shows in Figure 1.

3.2 In order to apply the technique to develop a model in accordance with the measurement factors of learning supervision

characteristics used in the study of data in the research, the researcher reviewed the literature related to data analysis techniques with data mining techniques for self-development. The model involved in the analysis of learning supervision data and learners' learning achievement was analyzed frequency according to the synthesis table 2 as follows (□ refers to the technique used in the study).

Table 2 *Synthesis of Data Analytic techniques with Machine Learning techniques.*

Authors	Data Analytic techniques with Machine Learning techniques						
	LR	DT	ANN	SVM	NB	ARL	K-NN
Sudhir Deshmukh (2020)	•		•	•			
Abdulsalam Sulaiman (2015)		•					
Yadav (2012)	•	•					
Pal (2013)		•	•			•	•
Steve Essinger (2011)							•
Boonprasom (2018)		•				•	•
Sakdulyatham (2009)		•	•				
Wilailuck (2015)		•	•	•			
Total	2	6	4	2	1	1	3

From Table 2, the synthesis of data analysis techniques using data mining techniques, it was found that the frequency of using the technique was as follows: two Decision Tree (DT) techniques, six Artificial Neuron Network (ANN) techniques, four Naive Bayes (NB) learning, two Support Vector Machines (SVMs), one Logistic Regression (LR), one

Association Rule Learning (ARL), and three K-Nearest Neighbors (K-NN).

3.3 Selection of techniques used to present the most appropriate and effective data analysis techniques based on performance testing. Cross-validation with the frequency distribution according to Table 3 as follows.

Table 3. *The most efficient Machine Learning techniques frequency distribution*

Authors	Data Analytic techniques with Machine Learning techniques						
	LR	DT	ANN	SVM	NB	ARL	K-NN
Sudhir Deshmukh (2020)	✓						
Abdulsalam Sulaiman (2015)		✓					
Yadav (2012)		✓					
Pal (2013)							✓
Steve Essinger (2011)							✓
Boonprasom (2018)						✓	
Sakdulyatham (2009)		✓	✓				

Authors	Data Analytic techniques with Machine Learning techniques						
	LR	DT	ANN	SVM	NB	ARL	K-NN
Wilailuck (2015)			✓				
Total	1	3	2	0	1	0	2

Table 3 shows the most efficient technique frequency distribution table. It was found that the five most effective techniques were Decision Tree (DT), Artificial Neuron Network (ANN), K -Nearest Neighbors (K-NN), Naive Bayes (NB) learning, and Logistic Regression (LR) which the researcher will use as a model for data analysis in the next step.

1. From the synthesis of data analysis techniques with data mining techniques suitable for model development by frequency analysis, there are 5 most effective techniques: Decision Tree (DT); Artificial Neuron Network (ANN), Naive Bayes (NB), Logistic Regression (LR), and K -Nearest Neighbors (K-NN). These techniques can be categorized into 2 types of techniques: Supervised Learning and Uneducated Learning. (Unsupervised Learning) according to Table 4 as follows.

Results

Table 4. *Types of Machine Learning Techniques*

Techniques	
Supervised Learning	Unsupervised Learning
1. Decision Tree (DT)	1. K -Nearest Neighbors (K-NN)
2. Artificial Neuron Network (ANN)	
3. Naive Bayes (NB)	
4. Logistic Regression (LR)	

2. Testing the performance of Machine Learning techniques in analytics Self-Regulated Learning

Table 5 *Test results of Machine Learning techniques in analytics of Self-Regulated Learning*

Model	Class	A*	B+**	B***	C+*** *	C**** *	D+*** ***	accuracy	Macro avg
K-Nearest Neighbors (KNN)	Precision	0.17	0.33	0.33	0.00	1.00	0.00	0.32	0.37
	Recall	0.33	0.25	0.38	0.00	1.00	0.00		0.39
	f1-score	0.22	0.29	0.35	0.00	1.00	0.00		0.37
	Support	3	8	8	2	1	0.00	22	22
Logistic Regression (LR)	Precision	0.00	0.46	1.00	0.33	1.00	0.00	0.55	0.56
	Recall	0.00	0.75	0.50	0.50	1.00	0.00		0.55
	f1-score	0.00	0.57	0.67	0.40	1.00	0.00		0.53
	Support	3	8	8	2	1	0.00	22	22
Artificial Neural Network (ANN)	Precision	0.14	0.25	0.25	0.33	0.00	0.00	0.23	0.20
	Recall	0.33	0.25	0.12	0.50	0.00	0.00		0.24
	f1-score	0.20	0.25	0.17	0.40	0.00	0.00		0.20
	Support	3	8	8	2	1	0.00	22	22
Naïve	Precision	0.33	0.43	0.67	1.00	1.00	0.00	0.50	0.69

Bayes (NB)	Recall	0.33	0.75	0.25	0.50	1.00	0.00	22	0.57
	f1-score	0.33	0.55	0.36	0.67	1.00	0.00		0.58
	Support	3	8	8	2	1	0.00		22
Decision Tree (DT)	Precision	1.00	0.89	1.00	1.00	0.00	0.00	0.91	0.65
	Recall	1.00	1.00	0.88	1.00	0.00	0.00		0.65
	f1-score	1.00	0.94	0.93	1.00	0.00	0.00		0.65
	Support	3	8	8	2	1	0		22

A* = Score above 80%

B**+ = Score between 75 -79%

B*** = Score between 70 – 74%

C+**** = Score between 65 – 69%

D+***** = Score between 55 – 59%

Table 5 shows test results of Artificial Neural Network Technique (ANN), K-Nearest Neighbors Technique (K-NN), Logistic Regression Technique, Naïve Bayes Technique (NB), Decision Tree Technique can be summarized as follows. Testing for the K-Nearest Neighbors (KNN) technique showed accuracy at 0.32, Logistic Regression (LR) technique, accuracy at 0.55, Artificial Neural Network Technique (ANN), accuracy at 0.23 , Naïve Bayes technique (NB) accuracy is found to be 0.50, and Decision Tree (DT) technique shows that accuracy is 0.91. It can be concluded that from the testing of the 5 techniques used for analysis, there are Technique Support Vector Machine Technique, K- Nearest Neighbors Technique (K-NN), Logistic Regression Technique, Naïve Bayes Technique (NB), Decision Tree Technique. There are different analysis results but all techniques have the same support value. The most accurate technique in data analysis is Decision Tree technique.

Conclusions

The research results from the synthesis of data analytic techniques with Machine learning. There are 5 techniques that are most effective: Artificial Neural Network Technique (ANN), K-Nearest Neighbors Technique (K-NN), Logistic. Regression Technique, Naïve Bayes Technique (NB), Decision Tree Technique There are two types of techniques: 1)

Supervised Learning Logistic Regression Technique, Naïve Bayes Technique (NB), Decision Tree Technique, and Artificial Neural Network Technique (ANN), which is consistent with Samngamjan, Philuek, Malangpoo, and Janyarat (2022) who studied predictive techniques for factors affecting computer programming learning. logistics for maximum accuracy and 2) unattended learning. (Unsupervised Learning) is the K-Nearest Neighbors Technique (K-NN). Both techniques can be used to group data to find the analysis relationship between groups and individual factors. For the Machine Learning techniques, the result shows that Decision Tree (DT) technique shows that accuracy is 0.91 which is higher that other four techniques. This result will lead to the next phase that researchers will use in development of digital platform in prediction of Self-Regulated Learning of secondary students.

Reference

- [1] Abdulsalam S., Babatunde A., Hambali M., Babatunde R. (2015). Journal of advances in scientific research & its application (JASRA) 2, 79-92
- [2] Berry J.A., and Gordon S. L. (2004). Data Mining Techniques: for Marketing, Sales, and Customer Relationship Management. 2"d edition. Indiana: Wiley Publishing Inc.
- [3] Booget, N. (2011) Effects of Regulated Learning Methods on Web and Tutoring Methods in Blended Learning upon Learning Achievement and Self-Regulated Learning Skills of Undergraduate Students. Doctoral Dissertation of Chulalongkorn University Graduate School.
- [4] Boonprasom, C. (2018). An Analysis of Predicting Dropouts of Undergraduate Students. using data mining techniques.

- Journal of Industrial Education King Mongkut, North Bangkok, Year 9, No. 1.
- [5] Dabbagh, N. and Kitsantas, A. (2005) Using web-based pedagogical tools as scaffold for self-regulated Learning. *Instructional Science* 33: 513-540
- [6] Daniel T. L. (2005). *Discovering Knowledge in Data: An Introduction to Data Mining* New Jersey: John Wiley & Sons Inc.
- [7] DeKeyrel, A., Dernovish, J., Epperly, A., and McKay, V. (2000). Using motivational strategies to improve academic achievement of middle school students. (ERIC Document Reproduction Service No. ED443550)
- [8] Deshmukh, S. (2020). A Machine Learning Approach to Predict Retention of Computer Science Students at University of Nevada, Las Vegas. UNLV Theses, Dissertations, Professional Papers, and Capstones. 3887.
- [9] Essinger, S. D., and Gail L. R. (2011). An introduction to machine learning for students in secondary education. 2011 Digital Signal Processing and Signal Processing Education Meeting (DSP/SPE). IEEE.
- [10] Fuller, R. (1999). Do university students' conceptions of learning really influence their learning? <http://www.herdsa.org.au/vic/cornerstones/pdf/fuller.pdf>
- [11] Lee, L. H. (1997). Goal orientation, goal setting, and academic performance in college students: An integrated model of achievement motivation in school settings. *Dissertation Abstracts International*, 59(06), 1905A.
- [12] Lin, Y. G., and McKeachie, W. J. (1999). College student intrinsic and/or extrinsic motivation and learning. (ERIC Document Reproduction Service No. ED435954)
- [13] Man-Chih, A. (2006). The effect of the use of self-regulated learning strategies on college students' performance and satisfaction in physical education. Australian Catholic University.
- [14] McClendon, R. C. (1996). Motivation and cognition of preservice teachers: MSLQ. *Journal of Instructional Psychology*, 23(3), 216-220.
- [15] McManus, T. F. (2000). Individualizing instruction in a web-based hypermedia learning environment: Nonlinearity, advance organizers, and self-regulated learners. *Journal of Interactive Learning Research*, 11(2), 219-251.
- [16] Pal, A. and Pal, S. (2013). Data mining techniques in EDM for predicting the performance of students. *International Journal of Computer and Information Technology*. Volume 02 – Issue 06, November.
- [17] Philuek, W. (2017) A study on the effectiveness of anonymous agents in Computer-Supported Collaborative Learning (CSCL) for the facilitation of student learning and execution of team projects. PhD thesis, Murdoch University, Western Australia, Australia
- [18] Philuek, W., Puttasem, D., Janyarat, S. and Inchai, W. (2020) A Study on Self-Regulated Learning (SRL) and Learning Outcome of Computer Preservice Teachers: Case of Nakhon Sawan Rajabhat University, Thailand. *Proceedings of 2020 International Conference on Technological and Vocational Education*, January 4th 2020, National Taiwan Normal University, Taipei, Taiwan (P. 139-148)
- [19] Pintrich, P. R. (2004). A conceptual framework for assessing motivation and self-regulated learning in college students. *Educational Psychology Review*, 16(4), 385-407.
- [20] Pokay, P., and Blumenfeld-Phyllis, C. (1990). Predicting achievement early and late in the semester: The role of motivation and the use of learning strategies. *Journal of Educational Psychology*, 82(1), 41-50
- [21] Rao, N., Moely, B., and Sachs, J. (2000). Motivational beliefs, study strategies, and mathematics attainment in high- and low-achieving Chinese secondary school students. *Contemporary Educational Psychology*, 25(3), 287-316.
- [22] Samngamjan, N., Philuek, W., Malangpoo, P., and Janyarat, S. (2022) A Study on Factors and Using Machine Learning for Learning Analytics in Computer Programming of Junior High-school Students. *Natural Volatiles and Essential Oils*, Vol. 8 Issue 6. (P. 1224-1233).
- [23] Srisatchalertwacha, W. and Prasertthitiphong, D. (2014). *Educational Data Mining (Educational Data Mining)*. Faculty of Science, Chiang Mai University.

- [24] Yadav, S. K., and Saurabh, P. (2012). Data mining: A prediction for performance improvement of engineering students using classification. arXiv preprint arXiv:1203.3832.
- [25] Yamauchi, J., Kumagai, Y., and Kawasaki, Y. (1999). Perceived control, autonomy, and self-regulated learning strategies among Japanese high school students. *Psychological Reports*, 85(3), 779-798.
- [26] Zimmerman, B.J. (1989). A social cognitive view of self-regulated academic learning. *Journal of Educational Psychology*, 81, 329-339.
- [27] Zimmerman, B.J. (2000). Attaining self-regulation: A social cognitive perspective.
- [28] In M. Boekaerts, P. Pintrich, & M. Zeidner (Ed.), *Handbook of self-regulation* (pp. 13-39). San Diego: Academic Press.