OPEN ACCESS

Volume: 12

Special Issue: 1

Month: June

Year: 2024

E-ISSN: 2582-1334

Received: 30.10.2023

Accepted: 25.01.2024

Published: 29.06.2024

Citation:
Philuek, W. (2024). A Study on the Factors on SEEEM of Secondary Education Students during Thailand's COVID-19 Situation: Using Machine Learning in Analytics. *Shanlax International Journal of Education, 12*(S1), 63–69.

DOI:
https://doi.org/10.34293/education.v12iS1-June.6807

A Study on the Factors on SEEEM of Secondary Education Students during Thailand's COVID-19 Situation: Using Machine Learning in Analytics

Wudhijaya Philuek
Nakhon Sawan Rajabhat University, Thailand
https://orcid.org/0000-0002-1931-6633

Abstract
This research purposed to test the accuracy of Machine Learning techniques for learner analytics based on SEEEM factors of secondary education students in Thailand's COVID-19. Research volunteer came from secondary education students in Thailand who invited by researcher. The research questionnaire adapted from Computational Thinking Assessment by Korkmaz et al. (2017), Science Process Skills by Pruekpramool (2014), Environmental Literacy Instrument for Adolescents by EPA (2018), Test of Economic Literacy by Walstad et al. (2013), and Technology and Engineering Literacy Student Questionnaire by NAEP (2018). This research employed the statistics in analysis of Mean and Standard Deviation, and Machine Learning Techniques such as 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. The results of this research as it shown techniques used in data analytics in this paper may benefit to educators, teachers, or students in Thailand.

Keywords: Seem, Data Analytics, Machine Learning Techniques, Secondary Students in Thailand

Introduction

Education is an important infrastructure for human resource development which was unexpectedly affected by COVID crisis. Almost all educational institutions had to announce the closure of teaching and learning without any preparation. Phromwong et al. (2021) mentions the UNESCO report that countries around the world have closed schools across the country which affecting more than 1.5 billion students, or more than 90% of the population. All learners around the world due to the urgent measures taken by governments in almost every country to prevent the spread of the virus, whether semi-closed measures semi-lockdowns and social distancing measures have forced schools to close in response to government policies and to prevent and reduce the potential transmission of the virus. Panto (2020) mentioned the situation in Thailand during the outbreak of the COVID-19 that educational institutions and schools to announce the closure of teaching and learning for several months or in some schools, closing almost all the time and academic year due to being in an area with a severe COVID-19. In Thailand, SEC (2022) mentions education management in the past 2021 academic year that the Office of the Commission Basic Education (OBEC) has established learning styles to replace normal teaching and learning, including ON-AIR, ONLINE, ON-DEMAND, ON-HAND and ON-SITE. The implementation of online learning in education allows students to study and complete assignments whenever and wherever they want (Schwieren et al., 2006) which shows that learning in an online learning environment requires highly autonomous learning skills. The study of Dabbagh and Kitsantas (2005) also states that it is difficult for learners if they have low self-direction skills. Additionally, to foster student learning in an eLearning environment, teachers should inform students of the essential learning factors that lead to success so that students will be effective at learning, have a belief in self-efficacy and control motivation for success (Pintrich, 2000; Zimmermann, 2000).

A study in the United States found that children who were able to do well online during the COVID-19 situation were already doing well in the classroom but the middle group or lower levels in the classroom when studying online will not be able to learn or there will be a disruption in learning, known as learning loss or learning recession that is a problem after the outbreak of the COVID-19. Saifa (2021) explained “Learning Loss” is the loss of the opportunity to learn skills that should be acquired at each age due to various factors that affect language development and communication, relationship skills, socialization, lack of motivation to return to school. Students may stop learning in academic reading and writing or what students should learn according to grade level and age ability but did not develop or build on those skills according to the age criteria they should have received. Chanchalia-Seributt (2021) gave an example of a learning loss in Grade 4 students where the research found that students were unable to learn according to the standards in class level which may be something that has been learned, but after the period has not been reviewed, students cannot remember or in early childhood children who have helped themselves, but when the school has closed the teaching for a period of time until the school was opened again, those students were unable to help themselves as before.

Machine learning techniques have been applied to predict student performance in various educational contexts. Studies have utilized traditional machine learning classifiers (TMLCs) such as Gaussian Naïve Bayes, Support Vector Machine, Decision Tree, Multi-Layer Perceptron, Random Forest, Linear Discriminant Analysis, and Quadratic Discriminant Analysis to estimate early student performance (Mohammad et al., 2023). These classifiers achieved high accuracy, with the best results obtained using the Multi-Layer Perceptron classifier. Additionally, Convolutional Neural Networks (CNNs) have been used to predict student performance on different datasets, including House, Western Ontario University, Experience Application Programming Interface, University of California-Irvine, and Analytics Vidhya (Alalawi et al., 2023). The CNN based method outperformed conventional methods and demonstrated state-of-the-art performance. Machine learning has also been applied to predict student performance based on data from Student-Led Tutorials (SLTs) in an Electromagnetics Course, achieving accuracy rates of over 65% (Vertegaal et al, 2023). Overall, machine learning has shown great potential in predicting student performance and has been used to support decision-making and enhance educational outcomes (Kaur et al., 2023; Minghui and Zhijun, 2023).There are many different techniques used in data analytics. None of them can solve every problem of data mining, so a variety of techniques are necessary to lead to data mining best solution (Berry and Linoff, 2004). Hand et al. (2001) defined data mining as an evolution in storing and interpreting data from simple data storage to a database that can be retrieved. Information data is used until data mining can discover the knowledge hidden in the data. Giudici (2003) defined data mining as the process of selecting and exploring data as well as the modeling of data to find patterns and correlation from large amounts of data to produce useful results such as Unsupervised and Supervised in Machine Learning.

SEEEM is a transformation of STEM that uses Economics and Ecology instead of Technology because it breeds wisdom, believing in the goal of education that is more of a scientific culture than a technological culture. Scientific culture is to understand the real causes and effects of nature. It’s a culture that wants to know why to create an understanding of the phenomenon of effect, to raise awareness that if there is any effect, the causes and factors must be created in accordance with the desired effect. Technological culture only wants to share the effect, ignoring the cause of that effect. Wisdom is higher when education builds a scientific culture in the paradigm of the poor “Scientific

Mind". Development is a coercive view of nature to manipulate nature for the enjoyment of the modified nature. In Thailand, aiming to create a scientific mind that integrates ideas based on the Philosophy of Sufficiency Economy combining Economics and Ecology with Engineering, as a process of life by managing to allow Engineering to be under the supervision of Economics with the balance of Ecology as a condition in building Technology in STEM.

Principles of learning from Research-Based Learning (RBL) Phase 1 has been developed to integrate STEM Education with the Philosophy of Sufficiency Economy which consists of 5 dimensions, namely

- **S (Scientific aspect)** training to develop attitudes and understanding of basic science subjects that can be used to explain phenomena gained from research experience. Science skills are observation, questioning, searching, and discriminating until a culture of thinking is created. The science is embedded into the scientific mind.
- **E (Economic aspect)** is practicing to understand economics in terms of real life arising from reciprocity and reciprocity in nature as an economic resource.
- **E (Ecological aspect)** trains to view ecological phenomena as everything that arises and exists systematically, interrelated and coexisting in a balanced manner.
- **E (Engineering aspect)** is practicing a thinking process that understands the system in which development and evolution must coexist in a balanced manner in order to achieve sustainability based on the real world of a competitive human society.
- **M (Mathematical aspect)** is not a subject to solve equations, but a skill to create equations to solve computational problems (computational thinking), namely: (1) Decomposition skills such as breaking down problems, processes into smaller parts to be able to manage, (2) Pattern Recognition skills to see the similarities and differences of the pattern of change this makes it possible to know the trend in order to predict the future, (3) Abstraction skills helps to understand the general picture (generalization) to be able to think about linking concrete and abstract worlds together resulting in knowledge or theory, and (4) Algorithm design skills in problem solving Know what to do before and after what resources do you use?

Research Methodology

The research of analytics SEEEM of secondary school students in Thailand's COVID-19 situation by Machine Learning Techniques in analytics has research step as follow.

- Study and analyze factors used to promote learning in active teaching with the SEEEM process for secondary school students.
- To test the relationship of factors used to promote learning in active teaching and learning with the SEEEM process for secondary school students.
- To test the Machine Learning techniques in analytics SEEEM process for secondary school students.

Scope of Research

The sample consisted of 500 students in the Secondary Education Service Area of Nakhon Sawan, Uthai Thani, Chainat and Phichit schools.

Content scope, there were five types of assessments in this form: (1) Computational Thinking Assessment by Korkmaz et al. (2017), (2) Science Process Skills by Pruekpramool (2014), (3) Environmental Literacy Instrument for Adolescents by EPA (2018), (4) Test of Economic Literacy by Walstad et al. (2013), and (5) Technology and Engineering Literacy Student Questionnaire by The National Assessment of Educational Progress NAEP (2018).

Method of Collecting Data

Step 1: Select a questionnaire to match the objectives of this research.

Step 2: Bring a memo letter asking permission to use the sample group and send it to all 5 schools.

Step 3: Receive assessment responses from all schools contacted back.

Step 4: Select the program to be analyzed.

Data 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

- Understanding the problem
- Data Understanding
- Data Preparation, the information is divided into three sub-steps: (1) Data Selection by select relevant data for data mining. It is taken from multiple columns. (2) Do data cleaning (Data Cleaning) is the process of cleaning the data in order to format the data in an appropriate format, and eliminate missing data. (3) Data Transformation is the transformation of data into a format that is ready for use in analysis by specifying the quality variables such as Very Good, Good, Moderate, Fair, and Adjust.
- Modeling Phase

This phase was the selection of suitable models to improve the variables for the best results. At this stage, the researcher selected the following data mining techniques and algorithms: Support Vector Machine (SVM), K-Nearest Neighbors (K-NN), Random Forest (RF), Logistic Regression (LR), Decision Tree (DT), and Naïve Bayes (NB), Gradient Boosting (GB), Extreme Gradient Boosting (XGB), and Artificial Neural Network (ANN)with 80% for training and 20% for testing.

- Evaluation Phase
- Deployment Phase

Results

The results of the study on SEEM of students found that a total of 500 students were divided into 3 distribution which were Brilliant 27%, Medium 39.2%, and Weak 33.8% and researcher use this distribution to classified students as label for training data as it shown in Figure 1.

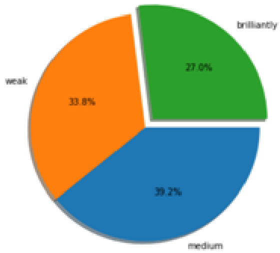

Figure 1 Target Variable Distribution

The results of a study of machine learning techniques for analyzing the relationship between 1) Computational Thinking, 2) Science Process Skills, 3) Environmental Literacy, 4) Economic Literacy, and 5) Technology and Engineering Literacy are shown in Figure 2, and Figure 3 show the relation of data distribution in each ML techniques. Table 1 and Figure 4 show the result of model testing.

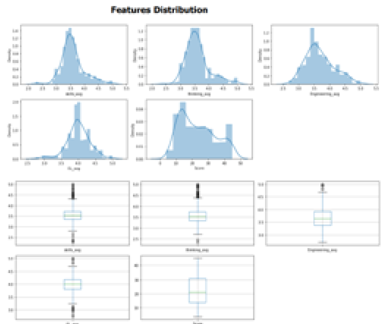

Figure 2 The Mean Each SEEM (Computational Thinking, Science Process Skills, Environmental Literacy, Economic Literacy, and Technology and Engineering Literacy)

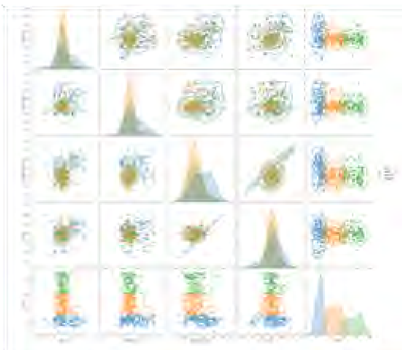

Figure 3 Relation of Data Distribution in each ML Techniques

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

Model	Class	'weak': 0	'medium': 1	'brilliantly': 2	accuracy	Macro avg
Logistic Regression (LR)	Precision	1.00	1.00	0.97	99.1	0.99
	Recall	1.00	0.97	1.00		0.99
	f1-score	1.00	0.99	0.99		0.99
	Support	35	36	38	109	109
Decision Tree Classifier (DT)	Precision	1.00	0.97	1.00	99.11	0.99
	Recall	1.00	1.00	0.97		0.99
	f1-score	1.00	0.99	0.99		0.99
	Support	35	36	38	109	109
Random Forest Classifier (RF)	Precision	1.00	1.00	1.00	100.0	1.00
	Recall	1.00	1.00	1.00		1.00
	f1-score	1.00	1.00	1.00		1.00
	Support	35	36	38	109	109
Naïve Bayes Classifier (NB)	Precision	0.71	0.38	0.67	62.4	0.59
	Recall	0.63	0.22	0.28		0.62
	f1-score	0.67	1.00	0.80		0.58
	Support	35	36	38	109	109
Support Vector Machine (SVM)	Precision	0.92	1.00	1.00	97.2	0.97
	Recall	1.00	0.92	0.96		0.97
	f1-score	1.00	1.00	1.00		0.97
	Support	35	36	38	109	109
K-Nearest Neighbours (KNN)	Precision	0.85	1.00	0.93	91.7	35
	Recall	1.00	0.75	1.00		36
	f1-score	0.92	0.86	0.96		38
	Support	0.93	0.92	0.91	109	109
Gradient Boosting (GB)	Precision	1.00	1.00	1.00	100.0	1.00
	Recall	1.00	1.00	1.00		1.00
	f1-score	1.00	1.00	1.00		1.00
	Support	35	36	38	109	109
Extreme Gradient Boosting (XGB)	Precision	1.00	1.00	1.00	100.0	1.00
	Recall	1.00	1.00	1.00		1.00
	f1-score	1.00	1.00	1.00		1.00
	Support	35	36	38	109	109
Artificial Neural Network (ANN)	Precision	0.97	0.97	1.00	98.2	35
	Recall	0.97	0.97	1.00		36
	f1-score	1.00	1.00	1.00		38
	Support	0.98	0.98	0.98	109	109

Table 1 and Figure 4 shows test results of Logistic Regression (LR) shows accuracy at 99.1, Decision Tree (DT) shows accuracy at 99.11, Random Forest (RF) show accuracy at 100.0, Naïve Bayes (NB) shows accuracy at 62.4, Support Vector Machine (SVM) shows accuracy at 97.2, K-Nearest Neighbors (K-NN) shows accuracy at 91.7, Gradient Boosting (GB) shows accuracy at 100.0, Extreme Gradient Boosting (XGB) shows accuracy at 100.0, and Artificial Neural Network (ANN) shows accuracy at

98.2, this can be summarized that the most accurate technique in data analysis is Random Forest (RF), Gradient Boosting (GB), 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.

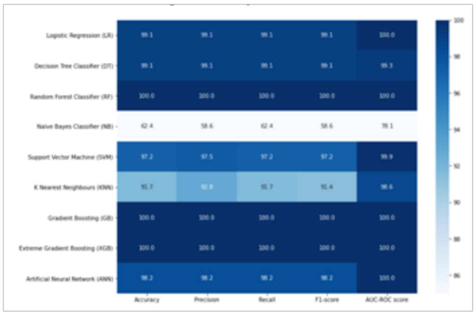

Figure 4 Test results of ML Algorithms Comparison: LR, DT, RF, NB, SVM, KNN, GB, XGB, and ANN

Conclusions

Conclusions of the results of this research consisted of machine learning techniques in analytics students' Computational Thinking, Science Process Skills, Environmental Literacy, Economic Literacy, and Technology and Engineering Literacy which 500 students were divided into 3 distribution which were Brilliant 27%, Medium 39.2%, and Weak 33.8%. The results of a study the accuracy of Machine Learning techniques for learner analytics based on SEEEM factors of secondary education students in Thailand's COVID-19, Machine learning techniques have been applied to predict student performance in various educational contexts. Studies have utilized traditional machine learning classifiers (TMLCs) such as Gaussian Naïve Bayes, Support Vector Machine, Decision Tree, Multi-Layer Perceptron, Random Forest, Linear Discriminant Analysis, and Quadratic Discriminant Analysis to estimate early student performance (Mohammad, et. al 2023). These classifiers achieved high accuracy, with the best results obtained using the Multi-Layer Perceptron classifier. Additionally, Convolutional Neural Networks (CNNs) have been used to predict student performance on different datasets, including House, Western Ontario University, Experience Application Programming Interface, University of California-Irvine, and Analytics Vidhya (Alalawi, et. al, 2023). The CNN-based method outperformed conventional methods and demonstrated state-of-the-art performance. Machine learning has also been applied to predict student performance based on data from Student-Led Tutorials (SLTs) in an Electromagnetics Course, achieving accuracy rates of over 65% (Vertegaal, et. al, 2023). Overall, machine learning has shown great potential in predicting student performance and has been used to support decision-making and enhance educational outcomes (Kaur, et. al, 2023; Minghui and Zhijun, 2023). The result shows that Random Forest (RF), Gradient Boosting (GB), and Extreme Gradient Boosting (XGB) gave the most accuracy (100) in analytics the SEEM of students while the weak technique which is Naïve Bayes (NB) gave the lower accuracy (62.4). From this result, researcher may use this technique in development of digital platform to predict the students' SEEM in the further.

Acknowledgement

This research was funded by Fundamental Fund (2022) under Thailand Science Research and Innovation which organized by Nakhon Sawan Rajabhat University.

References

Alalawi, K., Athauda, R., & Chiong, R. (2023). Contextualizing the current state of research on the use of machine learning for student performance prediction: A systematic literature review. *Engineering Reports*, *5*(12).

Beck, J. S. (2011). *Cognitive Behavior Therapy: Basics and Beyond*. New York: Guilford Press.

Berry, M. J. A., & Linoff, G. S. (2004). *Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management*. John Wiley & Sons.

Chanchalia-Seribut, N. (2021). Recession of knowledge (learning loss) that must be urgently resolved from the perspective of CEO Starfish Education. *Starfish Labz*.

Dabbagh, N., & Kitsantas, A. (2005). Using web-based pedagogical tools as Scaffold for self-regulated learning. *Instructional Science, 33*, 513-540.

EPA. (2018). What is environmental education?. United States Environmental Protection Agency.

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

Gladstone, T. R., Beardslee, W. R., & O'Connor, E. E. (2011). The prevention of adolescent depression. *Psychiatric Clinic of North America*, *34*(1), 35-52.

Hand, D., Mannila, H., & Smyth, P. (2001). *Principles of Data Mining*. MIT Press.

Kaur, H., Kaur, T., & Garg, R. (2023). A prediction model for online student academic performance using machine learning. *Informatica An International Journal of Computing and Informatics, 47*(1), 97-108.

Korkmaz, Ö., Cakir, R., & Özden, M. Y. (2017). A validity and reliability study of the computational thinking scales (CTS). *Computers in Human Behavior*, *72*, 558-569.

Minghui, Z., & Zhijun, L. (2023). Enhancing education performance through machine learning: A study of student learning outcomes prediction using GANs and ANNs. *IEEE International Conference on Control, Electronics and Computer Technology*.

Mohammad, A. S., Al-Kaltakchi, M. T. S., Al-Ani, J. A., & Chambers, J. A. (2023). Comprehensive evaluations of student performance estimation via machine learning. *Mathematics, 11*(14).

Panto, P. (2020). *Teaching and Learning Management in Thailand under the Epidemic Situation of the Coronavirus Disease 2019 (COVID–19)*. Office of Academic Affairs.

Phromwong, W., Prasantri, T., & Sriputtarin, S. (2021). Current situations, problems and solutions to learning management problems the Corona Virus Pandemic (COVID-19) of schools under the Nakhon Phanom primary educational service area office 1. *Rajapark Journal. 15*(40), 202-203.

Pintrich, P. R. (2000). The role of goal orientation in self-regulated learning. In M. Boekaerts, P. R. Pintrich & M. Zeidner (Eds.), *Handbook of Self-Regulation*. Academic Press.

Pongsuk, T. (2017). *A Study of Data Mining Techniques in Prediction of Academic Results and Learning Behavior of Students of Computer Studies Program at Northern University*. Nakhon Sawan Rajabhat University.

Pruekpramool, C. (2014). Assessment of scientific process skills. *Suthiparitat Journal*, *28*.

Saifa, Y. (2021). *Learning Loss Learning Recession*.

Schwieren, J., Vossen, G., & Westerkamp, P. (2006). Using Software Testing Techniques for Efficient Handling of Programming Exercises in an eLearning Platform. *The Electronic Journal of e-Learning, 4*(1), 87-94.

Shearer, C. (2000). The CRISP-DM Model: The New Blueprint for Data Mining. *Journal of Data Warehousing, 5*(4), 13-22.

NAEP. (2018). *Technology and Engineering Literacy Student Questionnaire*. The National Assessment of Educational Progress.

SEC. (2022). *Manual for using the Basic Education Curriculum Framework*. The Secretariat of the Education Council.

Vertegaal, C. J. C., Sundaramoorthy, P., Martinez, C., Serra, R., & Bentum, M. J. (2023). Exploring the potential of machine learning to predict student performance in an EM course. *32nd Annual Conference of the European Association for Education in Electrical and Information Engineering.*

Walstad, W. B., Rebeck, K., & Butters, R. B. (2013). *Test of Economic Literacy Examiner's Manual.* Council for Economic Education.

Zimmerman, B. J. (2000). Attaining self-regulation. A social cognitive perspective. *Handbook of Self-Regulation* (pp. 13-39), Academic Press.

Author Details

Wudhijaya Philuek, *Nakhon Sawan Rajabhat University, Thailand,* ***Email ID****: wudhijaya@nsru.ac.th*