Implementation of Data Analytics and Machine Learning in Thailand Education Sector

https://doi.org/10.3991/ijet.v18i05.36871

Pratya Nuankaew[1(✉)], Patchara Nasa-Ngium[2], Tinnakorn Kunasit[2],
Wongpanya Sararat Nuankaew[2]
[1]University of Phayao, Phayao, Thailand
[2]Rajabhat Maha Sarakham University, Maha Sarakham, Thailand
pratya.nu@up.ac.th

Abstract—Since the global epidemic of the coronavirus disease 2019 (COVID-19) over the past few years, Thailand education sector has been affected by the requisites for a digitization system and distance education. This sudden change has affected the quality of learning and statistical evaluations in the long term. Consequently, data analysis and categorization in learning quality assessment are critical for predicting the number of future students and learning performance after the COVID-19 outbreak. However, vast data analytics might be applied to the education sector in many aspects. In addition, machine learning can influence the categorization of students that are useful for analyzing the performance of different educational systems. Therefore, this study reviews the perspective and usability of data analytics and machine learning that influences current situations in Thailand education sector.

Keywords—data analytics, education sector, machine learning, Thailand

1 Introduction

In general, the development of an educational system depends on various factors and methods related to the current economic situation and the social environment [1]. This heterogeneous data structure is stored and managed as local data for easy access and compliance with the requirements of the active education system [2]. Paper-based data manipulation and reference from academics or administrative staff can be carried out offline [3]. Besides, teachers or lecturers can evaluate students' abilities and performance without relying on any special tools or online technologies [4]. However, changes in the education system and learning methods since the outbreak of coronavirus disease 2019 (COVID-19) have resulted in rapid database expansion [5]. As the report of UNICEF and UNESCO in Figure 1, 140 million students in Southeast Asia education, including Thailand was disrupted by the COVID-19 pandemic and waited several years to return to local classrooms in 2020 [6]. Also, 27 million students were unable to go to school during the COVID-19 pandemic. Although e-learning has been in widespread use in recent years, a large number of students rely on online education more than ever.

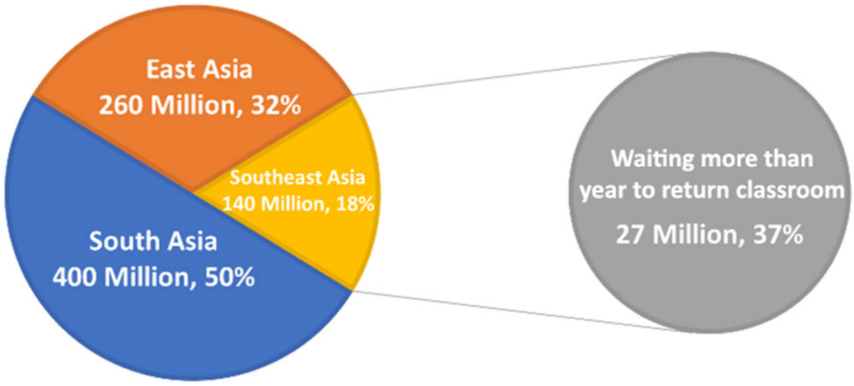

Fig. 1. The portion of students who affected by COVID-19 in the Asia education sector [5]

Differences in database structure, organizational resource constraints, and large student data records, it is likely to be captured and analyzed by various methods [7–8]. Also, error analysis of data leads to erroneous actions in the future. Nevertheless, the general education sectors in Thailand rely on predictions made by information technologies such as artificial intelligence and data mining [8]. Consequently, data analytics and machine are considered tools that can be adopted to improve the educational system.

In the case of data direct access, data analytics is a technology-based approach that can be easily applied to statistical prediction [3]. General information about students, such as grades along with class hierarchy can be used to strategically build a large database. Thus, big data analytics in sophisticated educational programs influence the transformation of institutions either in finance or administration [9–10]. Hence, institutions have attempted to analyze the available data sources and anticipate the following possible scenarios.

1.1 Educational analytics

Educational analytics can strengthen teacher effectiveness in decision-making circumstances [11]. This statistical method supports rational decision-making and leads to successful outcomes.

1.2 Decision making support

The efficiency of data analytics depends on the same technological infrastructure to store, organize, and search for the desired data [12]. Since the data is stored in the same storage, it is not necessary to search for huge numbers of files and folders from other sources. Searching from a single source increases the efficiency of checking and extracting data in a limited time.

It indicates that big data analytics can enhance the efficiency of education sectors that rely on statistical data due to the requirements of digitalization during the COVID-19 pandemic. In addition to predicting the performance of an educational system using data analytics, categorizing datasets by specific groups of students is essential to optimizing a personalized learning scheme [13]. Machine learning (ML) transforms learning and research patterns in problem identification and group-specific performance [14]. Researchers engage in problem identification with ML to uncover new insights by classifying data clusters. Moreover, ML was adopted as an online learning platform through pattern recognition [15], speech transcription [16], bioinformatics classification [17], and structured data analysis [18]. Basically, ML offers a student performance tracking and evaluation mechanism. It also provides issues addressing and solution discovery [19]. Furthermore, the personalized learning-based approaches throughout ML implementation encourage educators in defining proper learning pathways according to specific groups or individual students [20]. A suitable learning scheme improves the self-service capabilities and learning experiences of students. In addition, this method also enhances the prediction of outcomes and ensures the students' success each semester.

Therefore, due to the variety of applications and benefits that vary from system structure, this study reviews various aspects of research related to data analytics and machine learning regarding Thailand education section.

2 Data analytics influences in Thailand education sector

With advances in technology, data analysis methods have been greatly improved. Entrepreneurs have recognized the wealth of information and positive results in supporting business growth. However, the vast amount of collected data cannot be processed through traditional software [21]. Regular users, as well as investors, need a cloud-based platform for data mining. In many business cases, the COVID-19 pandemic is an aggressive warning of possible future risks. Information technology is not just about keeping big business anymore; education sectors also realize this circumstance. They have adopted data analysis and predictive models to predict a significant impact on institutional marketing methods and course structure as well as a student supporting and tracking. A wealth of information about students and faculty insights can be gathered and analyzed to improve learning performance. Data analytics is a method of analyzing various data from big data to assist in assessing the initial performance of a business or education system that relies on large data sets for the desired outcomes.

Typically, pre-processing educational data is raw data that is difficult to interpret so the collected need to be organized into a format that is ready to be processed by technology, algorithms, and models that create the knowledge that educators need to serve their students [22]. Educators in Thailand can use data analytics to assess their data to find patterns that positively impact strategy changes and increase student GPA [23].

Table 1. Different implementation of data analytics in Thailand education sector

Methods	Settings	Objectives	Outcomes
Extra-trees Classifier [24]	1. 28272 anonymous enrollment records of students from the Computer Science department, Thammasat University (CAMPUS_ID, CURRICULUM_ID, TERM, COURSE_ID, STUDENT_ID, GRADE) 2. Classifier comparison of the decision tree, random forest, extra-tree, and support vector machine (5-fold cross-validation) with Scikit-learn and Weka	1. Discover a new dependence between courses to enhance the arrangement orders of prerequisites 2. Discovering the several impacts of the additional curriculum revisions	There were improvements in education management and curriculum planning.
Document Analysis [25]	1. The personal characteristics of undergraduate students (GPA, gender, field of study, and year of study) 2. Learning environment (ICT facilities, faculty staff, library usage, and information literacy course) 3. Self-directed learning (self-managing, self-monitoring, and self-modifying)	1. Identify the influencing factors on the information literacy of Thai undergraduate students 2. Develop the questionnaire using the information literacy factors for undergraduate students	The factors of the study can be categorized into personal characteristics, learning environment, and self-directed learning.
The questionnaire, Pearson Product Moment Correlation Coefficient, and Multiple Linear Regression Analysis [27]	1. 26 teachers from Ban Lat Reur Border Patrol Police School, Ban Rak Thai Border Patrol Police School, Athorn Uthit Border Patrol Police School, and Ban Nuch Thian Border Patrol Police School 2. Ordinary National Educational Test (O-NET) in Mathematics, Thai language, Sciences, and English 3. Four Basic Rules, five Management Strategies, and six Teacher Implementation Problem-Solving Capacity Applying Life Skills Capacity, Thinking Capacity, Communication Capacity, and Technological Application Capacity	Analyzing the strategy that affects the performance of students in O-NET tests using linear multiple regression analysis and questionnaire	1. The four Basic Rules can predict the student's performance in the O-NET test. 2. The five Management Strategies can predict student performance in the key competencies for learning.
Descriptive Analysis and Survey [28]	1. 56 English as a Foreign Language (EFL) Teachers (Filipinos) in Bangkok, Thailand 2. English-language questionnaire based on Frederiksen's theory	Analyzing the perceptions of the specific problems, conditions, and challenges of Filipino EFL teachers in Bangkok, Thailand.	Filipino teachers declared some personal and professional issues in the findings.
Questionnaire and Status Quo Analysis [29]	1. 50 students from the university in Bangkok, Thailand 2. 17 teachers/professors from a university in Bangkok, Thailand 3. 27 Thai industries in Thailand	Develop ETAT Training Centers with ETAT Smart Labs for Automation 4.0 in Thailand	Analysis results can be used to build ETAT Training Centers with ETAT Smart Labs.
Snowball Sampling and Cause and Effect Analysis [30]	1. 28 educators in Nakhon Pathom, Thailand 2. Demographic characteristics of respondents (gender, age, education, teaching experience, workplace, level, and student number)	Defining the educational obstacles, challenges, and opportunities of educators, teachers, and lecturers in Nakhon Pathom, Thailand regarding the effect of the COVID-19 pandemic.	Challenges and opportunities after the COVID-19 pandemic were defined among the demographic characteristic of respondents.

Behavior, outcomes, strategic developments, and educational needs may depend on changing educational requirements. These changes are processed through statistical analysis, and data analysis, which pave the way for a better way of learning for students in several situations as shown in Table 1. A fundamental aspect of the Extra-trees classifier from Scikit-learn was used as a data analysis tool for detecting the dependence of unsorted student records [24]. Additionally, educators can compare different types of classifiers to predict the cumulative scores (grade) for new enrollment based on previous studies with anonymous enrollment records. Grade predictions are designed to recommend the most optimal study plan for individual abilities. This study discovered a new dependence between courses' effect on future curriculum revision to enhance the arrangement orders of prerequisites. Besides collectible records, in terms of document analysis, analyzing the variables of the research works and synthesizing the factors that affected the students' information literacy (IL) skills were conducted [25]. It indicated that the analysis of contents was necessary to categorize the set of factors of IL to personal characteristics, learning environment, and self-directed learning. These factors can be used to predict success in terms of undergraduate students and achievements in IL with many studies in documentary research. Grade predictions are designed to recommend the most optimal study plan for individual abilities. According to Table 1, a fundamental aspect of the Extra-Trees classifier from Scikit-learn was used as a data analysis tool via the Weka classifier for detecting the dependence of 28272 unsorted student records from Thammasat University [24]. This study discovered a new dependence between courses' effect on future curriculum revision to enhance the arrangement orders of prerequisites. A grade character (A-F, W, S, and U) was mapped to an integer to construct a classification model through the 20 most popular courses. The student records were sorted by temporal index (TERM) and filled into the subsequent record as a feature vector in the training process. TERM was used to compare the Scikit-learn with several classifiers including Decision Tree (10 folds validation), Random Forest (10 folds validation), Extra-trees (10 folds validation), and Support Vector Machine (k=5). The Gini importance was implemented for extracting the dependency of courses and acquiring the impurity reduction of feature importance score (FIS) during space partitioning processes. As a result, it showed that the dependence detection results were reasonably correct since the proposed method can predict the dependence signal from the enrollment history and the suggestion in the current curriculum. In addition, it found that some courses had no influent on a subsequent subject that was not a prerequisite. Nevertheless, the suggestion on dependence detection may be unordered in a specific arrangement based on the current curriculum due to the different conditions in student backgrounds, course syllabi, teaching methods, etc.

Besides collectible records, in terms of document analysis, the framework for analyzing variables of the research works and synthesizing the factors that affected the students' information literacy (IL) skills was conducted [25]. It indicated that the analysis of contents was necessary to categorize the set of factors of IL to personal characteristics (gender, years, GPA, and majors), learning environment (courses, staff, library, and facilities), and self-directed learning (self-managing, self-modifying, and self-monitoring). These factors were used to develop the questionnaire for predicting success in terms of undergraduate students and achievements in IL with many studies in documentary research. Although user access information was acclaimed to be effective, there were no evaluation criteria or assessments presented in this study. Due to the popularity of the questionnaire in conjunction with data analytics, hierarchical linear model (HLM) analysis was presented within the framework of appropriate internet behavior among Thai student teachers [26]. The sample contained 1800 student teachers in 30 Thai government institutions through the statistical method. A questionnaire was adopted as a research tool with content validity and reliability at 0.83–0.92 and statistical methods in internet behavioral usage pre-test via two parts of questionnaires. The first part of the questionnaire contained the learning Internet and ethical Internet usability. Later, the second part comprised general information, family information, friend information, and educational information. In addition, the proper HLM for the internet use behavior for Thai student teachers consisted of predictive variables in two levels. In the first level, the hypothesized variables had a positive effect on inappropriate internet use behavior, ability, affective, family situation, institution situation, professors, and friends, and a negative effect on income. The second level offered the positive effect in friends and educational institution situations where friends had a direct effect on internet behavior usage. The rating scale of the content validity and reliability was examined by the six-behavior level based on the value criteria: Very appropriate (5.50–6.00), Appropriate (4.50–5.49), Fairly appropriate (3.50–4.49), Less appropriate (2.50–3.49), Least Appropriate (1.50–2.49), and Inappropriate (1.00–1.49). The results showed that the appropriate internet usage behavior of Thai students' teachers was at a high level. Results showed that Internet usage of students and teachers at home during the research was 65.5% and 41.22% of students' GPAs ranged between 3.00–3.49 through the results of the HLM 7 software program with unconditional and hypothetical models. The statistical methods proposed in this study encouraged results consistency. Due to statistical reliability, the Pearson Product Moment Correlation Coefficient was used in data analytics to predict variables while the multiple linear regression analysis was used to design strategies for using the satellite distance education system that affect student quality in Thailand [27]. Also, the proposed theoretical framework and strategies include four basic rules (school environment organization, visual equipment suitability, teacher supervision, and student participation), five management strategies (learning management, activity cooperation, visual engagement, learning promotion, and learning evaluation), and six teacher implementation (classroom arrangement, activity preparation, learning management, content summarization, learning evaluation, and schedule arrangement) were presented as for predicting student performance in Ordinary National Educational Test (O-NET). Learners' Key Competencies as basic education core curriculum was used to compose 30 teachers in the Northern region of Thailand throughout the questionnaire in 5 level rating scales: 5 denotes most practice, 4 denotes very practical, 3 denotes moderate practice, 2 denotes Less practice, and 1 denotes minimal practice acquiring in the proposed strategies. As a result, the proposed strategies can predict student performance through O-NET tests and Multiple Linear Regression Analysis Stepwise by using three predictor variables for contributing to the development of satellite distance education systems for 36.90 percent statistical significance at the level of 0.05 and 57.90 percent statistical significance at the level of 0.01 along with 0.399 units of the Learners' Key Competencies in standard core format.

From the point of view of teacher circumstances, a survey was conducted on 56 Filipino teachers' perceptions of the specific challenges and conditions of becoming an EFL teacher in Bangkok, Thailand [28]. The questionnaires used in this study were designed based on Frederiksen's theory. Besides, follow-up interviews were used for collecting personal information along with the modified questionnaires. The distribution of reasons for teaching in Bangkok in percentage was responded as strongly agree, somewhat agree, neutral, somewhat disagree, and strongly disagree. Nonetheless, a qualitative content analysis was conducted depending on the data interpretation. It found that 70.83% of Filipino teachers experienced some personal and professional problems such as low salaries, though the experiences of 70.83% of respondents were positive in general cases where. It included problems caused by an unreliable state called Non-Native Teacher (NNEST) of English. However, with the Thailand 4.0 strategy, Thailand is pursuing its main goal to increase the trend of long-term educational system development. This strategy has made great advancements in automation 4.0 in both the industrial and manufacturing sectors. The ETAT project was previously proposed to create a training center called Smart Labs for Automation 4.0 in Thailand [29]. This training center aimed to train young and qualified professionals. The survey results involved three target groups including 50 students (76% bachelor, 18% master, and 6% doctorate degrees), 17 teachers, and 27 industry professionals. The questionnaire design was carried out in several categories such as objective explanation, participation background, knowledge level, technical requirement, and training format. Current knowledge and Status Quo analysis of qualification in Automation 4.0 was adopted to explain the level of satisfaction toward the level of qualification where teachers and industry professionals had a homogeneous view in the most practical skills and critical problems while students had a different problematic aspect. Although a survey of students, teachers, and industry professionals, requires further prevail in Thailand Automation 4.0, an appropriate direction through descriptive analysis along with possible teaching styles can attract new applications in the future. Meanwhile, defining challenges and opportunities through Snowball Sampling is how educators or teachers in Nakhon Pathom, Thailand, reflect the impact of the COVID-19 outbreak [30]. The data were collected through in-depth interviews and interpreted along with the cause-and-effect analysis to determine the challenges and opportunities of 28 experienced educational respondents.

Despite existing studies manipulated in data analytics, it shows that derived data and statistics can assist educators in preparing technology-driven initiatives such as online classrooms, e-books, and online exam platforms in a limited scope. The knowledge gained from data analysis can be effectively used to determine the potential costs of purchasing new software licenses for educational institution technology development. With the advent of data analytics in the education sector, institutions reach better prediction in the characteristics of analysis and applicants the factors according to the application process. This in-depth knowledge allows an educational institution to adapt its strategies for recruiting applicants and allocating funds to international students. Incorporating data analysis methods can result in better institutional incomes and sustainable student productivity. However, massive amounts of data such as big data remain a barrier in several cases in educational data analysis. Although big data analytics is not popular in education research in Thailand, it is considered a very important method that should be presented in the next session.

2.1 Big data influence in Thailand education sector

Big data analytics refers to the vast amount of data generated and collected from each operation [30]. Enormous data has become so problematic in analytics that traditional data management and document-keeping software cannot manipulate it. As the volume of data has grown exponentially due to the digital revolution after the COVID-19 pandemic, the requirements for efficient data analytics programs are a strategic necessity. Additionally, educators can adopt data analysis to record and analyze the following sample datasets:

Student: name, age, weight, height, gender, and religion.
Instructor: name, age, ethnicity, gender, salary, and responsible subjects.
Course: Course ID, number of registered students, list of subjects, academic year, and grade.
Others: Classroom number, number of hours, and number of seats.

In the education sector, a large amount of data is collected from the early level to high school. In higher education, data storage by leveraging technology can be used to track student development over the long term. The data analytics approach allows educators to delve into the performance of their students. Additionally, the big data analytics approach supports discovering the way how students learn and experience their classes [31]. It also affects policy deployment for the course syllabus arrangement each semester. Previously, the National Academy of Education demonstrated that the United States divides education data into two categories including Education Administration and the Learning Process [32]. Education Administration involves administrative data for the education sector, which is information on student behavior and school achievements, such as student scores, standardized tests, and awards. Public and private schools rely on this big data to evaluate the feasibility of achievement each semester. Learning Process is the learning information that is continuously collected. This information on the learning process is considered large data because it is collected from many learners. The data is broad because there are differences in variables and variations in individual learning patterns, called personalized learning.

The UNESCO Institute for Statistics reported that the international corporate large-scale data collection for educational monitoring consistent with the Sustainable Development Goals (SDGs) suggests that collected data requires coordination between organizations and a standardized framework [33]. However, due to budget and technology constraints in some countries, formal data has not yet been collected. Although big data analytics is widely implemented by Thai researchers in several areas, it is not popularly applied in the Thai education sector as a few recent research works are represented in Table 1.

Recently, intelligent real-time series analysis for intelligent universities has been introduced within the framework of big data analytics [34]. The proposed system was used to increase student learning efficiency and support the school administration's planning. To accommodate the unexpected changes during the COVID-19 pandemic, the proposed conceptual framework aimed to improve educational standards and develop a sector strategy. Besides, the adoption of big data analytics engages in saving time, increasing class arrangements, and increasing the efficiency of school management without relying on communication systems.

The ability to analyze big data within the framework of the education sector's requirements enables educators to improve educational outcomes. However, due to the need for continuous collection of educational resources, big data analytics is not prevalent in the Thailand education sector currently. Nonetheless, to encourage improvement in personalized learning of students, a prediction and analysis system is far better than relying on big data analytics in the future.

3 Machine learning adoption in Thailand education sector

From the perspective of general higher education institutions, improving student performance is of the utmost importance. However, it is necessary to classify student groups according to personal abilities or problems to be consistent with personalized learning before designing performance improvement models. Machine learning differs from traditional computing techniques where it relies on a set of model principles that have been trained to explain or solve problems [35]. Data inputs for scientific research can be provided using a computer-based learning approach, called the training set. The implementation and analysis process in a built-in machine learning model does not require manual processes throughout the operation process as with data analytics [36]. Consequently, this study reviews machine learning schemes deployed in the Thailand education sector.

According to Table 2, our previous research proposed a suitable method for grouping students' academic achievement (GPA) [37]. This proposed method supported the individual learning styles and attitudes levels toward different academic achievements among college and university students of Phayao, Thailand. The conceptual framework consisted of five important components including target group design, tool development, target group selection and data collection, data preparation, and model development and performance.

The target group design shared the knowledge in developing educational quality and patterns of different learning styles of students at the University of Phayao via the three sections of the questionnaire. The first section explained the data collection while the second section indicated the general information and backgrounds of respondents. In addition, the third section demonstrated the level of attitude in factor analysis. The essence of 195 questionnaires contained several perceptions in visual, kinesthetic, and auditory learning styles. The grade point average of students (GPA) (x is equal to 2.90) was used for grouping the cluster density through k-Means and x-Means. Two optimization results from the quantification of the machine learning technology process for clustering included k-optimization for k-Means and x-Means clustering. In addition, each dataset was extracted and analyzed by the decision tree through cross-validation and Euclidean distance. It found that the proposed decision Tree model provided the maximum efficiency at a depth of 7 with a 10-fold cross-check method, which matched different students' attitudes and learning styles in suitable educational programs. However, there were some aspects of research, such as dynamic responses, intent coverage, dataset size, and quality of training phrases that can be considered in further studies due to the limitation of interactive activity and workshop. Hence, interactive learning through machine learning, gamification, and social context was presented for STEM education in Thailand [38].

Table 2. Comparison of methods using machine learning in Thailand education sector

Algorithms	Settings	Factors/Inputs	Objectives
K-Means, X-Means, Decision Tree, Cross-Validation, and Euclidean Distance [37]	1. Cluster (K) = 5 2. Priorities Perceptions of Learning Styles (Visual, Auditory, and Kinesthetic) 3. Cross-Validation = 5-Fold, 10-Fold, and Leave-one-out 4. Schools and colleges at the University of Phayao, Thailand	1. GPA (Maximum GPA = 3.90 and Minimum GPA = 1.90) 2. Members (195 students)	Design the different clustering models matched with student's attitudes and learning styles in appropriate educational programs
Linear Regression [38]	1. STEM education is based on machine learning, gamification, and Social Context 2. Four P's of Creative Learning (Projects, Passion, Play, and Peers) 3. RapidMiner 4. Score-base challenge	1. 84 middle school students 2. Mangoes (model testing)	Support the middle school students learning in machine learning with a playful environment through the proposed interactive learning
Deep Learning (Fast.ai) [39]	1. Classroom time (30 mins introduction and 10 mins assessment) 2. AI curriculum design (AI Knowledge, AI Skill, and AI Attitude) 3. Lesson plan with 4 modules 4. Lesson guidelines (modules, lessons, times, lesson goals, and main points) 5. Group project (smart plant projects) 6. AI pictures, crayons, cards, and PictoBlox for class materials	K3–5 students (5–7 years old)	Provide the online machine learning class for K-12 Thai students through the Fast.ai platform and in-class Kaggle competition
Neural Network, Face Recognition, Knowledge-Base System, and Feature-Based Approach [40]	1. Nine-week class with eight lessons 2. Mak Pin Lom dataset 3. In-class Kaggle competition 4. Fast.ai platform 5. Evaluation criteria (Problem statement (15%), Metrics and baselines (15%), Data collection and cleaning (15%), Exploratory data analysis (20%), Modeling and error analysis (20%), and Prototype deployment (15%))	1. 25 Thai K-12 students 2. 21 auditors	1. Develop a proper AI curriculum framework for kindergarten, K3–5 students (5–7 years old) education 2. Define the suitable learning activities using the AI curriculum framework

Google Dialogflow and Cross-Validation [41]	1. Facebook Page (Messenger) 2. Google Dialogflow as Natural Language Understanding (NLU) platform 3. Cross-Validation = 5-Fold 4. Facebook Messenger API	807 sentences from 125 users	Support the availability of online support in higher education in Thailand using the chatbot Dialogflow
Natural Language Processing, questionnaire, descriptive statistics, and 5-level Likert scale [42]	1. Satisfaction level (Very low = 1, Low = 2, Medium = 3, High = 4, and Very high = 5) 2. Interpretation (minimal agreement, somewhat agree, moderate agreement, high agreement, and highest agreement) 3. Measurement process (speech identification, vocabulary correctness, sentence patterns and grammar aspects, and pronunciation fluency)	40 Thai high school students	Improve the English skills of Thai students in high school students via an AIT algorithm model
K-Nearest Neighbor, Logistic Regression, Random Forest, Decision Tree, Random Forest [43]	1. 3,969 samples from the MHESI dataset 2. R software version 3.6.2	1. University name (BUU, CU, CMU, TU, TSU) 2. Year of graduate (2013–2017) 3. Education level (Bachelor, Master, Ph.D.) 4. Study Major (Apply Math, Apply Stat, Statistic) 5. Type of university (Government, National university) 6. Ranking (In THE ranking, Not in THE ranking) 7. National research university (Research, Not research university) 8. Region (North, North East, Central, South)	Define the significant variables and visualization for predicting the employability of Thai graduates using K-NN, logistic regression, random forest decision tree, and random forest

Linear Regression and AI challenge based on agriculture was applied in RapidMiner application to engage 84 middle school students (grade 7 to 9) in 3 days of learning in the process and create machine learning models in the form of games with four P's of Creative Learning (projects, passion, play, and peers) principles. There were three phases of prediction workshops including sweetness, quality, and market challenges. Students were assigned to build a machine learning model within 6 hours using mangoes as a prediction assessment throughout these phases. The accurate prediction yielded different scores that indicated that the middle school level students performed better in machine learning model building in the second phase as the average accuracy was 66.67% compared to 60.89% in the first phase of the workshop.

Meanwhile, another project also proposed an online curriculum to improve the English skills of Thai students via machine learning concepts [39]. With global accessibility in current online classes, performing AI builders online summer school was considered a meaningful project that satisfies K-12 education in Southeast Asia during the COVID-19 pandemic.

A practical deep learning called Fast.ai was applied in coders to create the nine-weeks curriculum with 4 modules including the introduction of AI, machine learning, artificial intelligence techniques, and AI ethics. The in-class Kaggle competition along with the Mak Pin Lom dataset was used for student project correlation. The designed in-class competition encouraged students in learning to build the datasets relying on an open-source license or local machine learning datasets and submit the group project to a GitHub repository for further project explanation and evaluation. Besides learning through AI workshops in a middle school, an artificial intelligence curriculum or AI course with three competencies comprising AI Knowledge, AI Skill, and AI Attitude was designed to enhance kindergarten, K3–5 students (5–7 years old) education in AI knowledge [40]. The proposed AI curriculum framework consisted of four key components including achievements, contents, methods, and evaluations where students can learn about machine learning with study cases and face recognition via PictoBlox. This research work attempted to solve the problem of lacking AI curriculum standardization. Moreover, the social robot was used as a learning material to assist students in AI principles. It can be concluded that this AI teaching method can be performed as an effective learning method. Although the structure and application of machine learning or artificial intelligence technology are quite complex, they can be adopted by young students. Machine learning might not only enhance the education sector system or student performance but also can be applied as the main lesson in general education administration and academic services. To support the availability of online support in higher education in Thailand, Facebook Page was applied to evaluate the benefits of chatbot implementation for the end-to-end development process [41]. The chatbot was designed using Google Dialogflow as Natural Language Understanding (NLU) platform and Facebook Messenger for chatbot Dialogflow with 807 sentences from 125 online users. The proposed chatbots solved the problems faced by page administrators and end-user with 5-Fold cross-validation and satisfying results that achieved 0.984, 0.884, and 0.897 for precision, recall, and F1-score in 33 intents. Furthermore, the AI technology (AIT) model via Natural Language Processing was designed for 40 Thai high school students who studied English in 2020 [42]. The designed AIT model engaged students in understanding the sentence structure and vocabulary in English throughout the English training program development process. There are 40 items of quizzes assisted with AI technology algorithms including speech identification, vocabulary correctness, sentence patterns and grammar aspects, and pronunciation fluency. It revealed that the AIT algorithm prototype improved English skills with high satisfaction of students where IOC consistency ranged between 0.60–1.00, a difficulty between 0.26–0.75, and discriminant power was 0.74. Though analyzing results showed that the English knowledge of students was improved through the implementation of the AIT algorithm, the size of a dataset of students and the educational record is limited to the scope of specific affiliation use. On the other hand, the data might occasionally be collected continuously until reached significantly high dimensional data record, especially in graduate employment data, machine learning was proposed to solve the problem of the larger dataset for Thai graduates by predicting employability in 703777 records collected from 2013 to 2017 [43]. Several algorithms comprising K-NN, logistic regression, random forest decision tree, and random forest were considered and performed as reliable and comparative results based on the prediction quality in various variables including university name, year of graduate, education level, study major, type of university, ranking of the university, national research university, and region. In the case that the training set and the testing set were separated by a 70:30 ratio with 10 k-folds cross-validations and parameter tuning in R programming software, it showed that K-NN offered the most suitable models for employability prediction with an accuracy of 70.554% and 70.158%, 0.746, and 0.762% in AUC while the random forest model provided the best prediction in variables. As a result, it can be concluded that machine learning was applied in different ways in Thailand education sector.

4 Conclusion and discussion

According to the existing methods in several applications of data analytics in the Thailand education sector, it found that most of the research relied heavily on data collection, questionnaires, and surveys combined with statistical processes [25–29], while other academic works preferred different methods, such as classification [24] and snowball sampling [30]. Even if statistical methods provided satisfactory effectiveness for data analysis in different situations where the size is not very large [44], massive data volumes induced by the economic conditions during the COVID-19 outbreak have affected the accuracy of data assessment and analytics, whether it is a study plan [37–40] or an employment projection after each academic year [43]. Currently, though big data analytics is significantly applied to the Thailand education sector, this method is possibly considered a very suitable method for the development of the Thailand education sector in the future.

Additionally, machine learning is another interesting method for industrial, business, and several technological-applied areas in Thailand excluding the education sector. There were just a few numbers of research works existing in the field of Thailand education sector along with distinctive statistical methodologies as the latest studies explained in Tables 1 and 2. Due to the need for technicians and experts, various education sectors in Thailand focused on providing training sessions to increase AI or machine learning knowledge for different groups of students [37–40]. In addition, an adaptation of machine learning in various forms of learning systems such as providing training [37–40], specific development of students' skills [42–43], as well as providing online academic support [41]. Furthermore, implementing machine learning techniques can accurately predict performance and employability through continuously increasing amounts of data [45]. Therefore, even if data analytics and machine learning techniques are not prevalent in the Thailand education sector at present, they can be expected to gain sustained popularity in the future due to reliability and usability in many areas.

5 Acknowledgment

This research project was supported by the Thailand Science Research and Innovation Fund and the University of Phayao (Grant No. FF65-UoE006). In addition, this research was supported by many advisors, academics, researchers, students, and academic staff from the School of Information and Communication Technology, University of Phayao, Phayao, Thailand. The authors would like to thank all of them for their support and collaboration.

6 Conflict of interest

The authors declare no conflict of interest.

7 References

[1] R. Budiman. (2013). Utilizing Skype for providing learning support for Indonesian dis U. Sivarajah et al., “Critical analysis of big data challenges and analytical methods,” Journal of Business Research, vol. 70, pp. 263–286, 2017. https://doi.org/10.1016/j.jbusres.2016.08.001

[2] P. Limna, S. Jakwatanatham, S. Siripipattanakul, P. Kaewpuang, and P. Sriboonruang, “A review of artificial intelligence (AI) in education during the digital era,” Advance Knowledge for Executives, vol. 1, no. 1, pp. 1–9, 2022.

[3] K. Vassakis, E. Petrakis, and I. Kopanakis, “Big data analytics: Applications, prospects and challenges,” Mobile Big Data, pp. 3–20, 2018. https://doi.org/10.1007/978-3-319-67925-9_1

[4] A. G. Picciano, “The evolution of big data and learning analytics in American higher education,” Journal of Asynchronous Learning Networks, vol. 16, no. 3, pp. 9–20, 2012. https://doi.org/10.24059/olj.v16i3.267

[5] B. Daniel, “Big data and analytics in higher education: Opportunities and challenges,” British Journal of Educational Technology, vol. 46, no. 5, pp. 904–920, 2015. https://doi.org/10.1111/bjet.12230

[6] UNESCO Bangkok. The future of 800 million children across Asia at risk as their education has been severely affected by the COVID-19 pandemic, 2021. [Online]. Available: https://bangkok.unesco.org/content/future-800-million-children-across-asia-risk-their-education-has-been-severely-affected. Accessed on: August 1, 2022.

[7] R. Besbes, R. Tourki, and P. Gorce, “Engineering analysis for educational concepts' assessment,” in 2013 3rd Interdisciplinary Engineering Design Education Conference. IEEE, 2013. pp. 121–128. https://doi.org/10.1109/IEDEC.2013.6526772

[8] S. F. Wamba et al., “Big data analytics and firm performance: Effects of dynamic capabilities,” Journal of Business Research, vol. 70, pp. 356–365. https://doi.org/10.1016/j.jbusres.2016.08.009

[9] A. L. Martin and F. A. Thawabieh, "The role of big data management and analytics in higher education," Business, Management and Economics Research, vol. 3, no. 7, pp. 85–91, 2017.
[10] G. F. Marchena Sekli and I. De La Vega, "Adoption of big data analytics and its impact on organizational performance in higher education mediated by knowledge management," Journal of Open Innovation: Technology, Market, and Complexity, vol. 7, no. 4, pp. 221, 2021. https://doi.org/10.3390/joitmc7040221
[11] T. Agasisti and J. B. Alex, "Data analytics and decision making in education: Towards the educational data scientist as a key actor in schools and higher education institutions," Handbook of contemporary education economics. Cheltenham: Edward Elgar Publishing, 2017.
[12] K. Vassakis, E. Petrakis, and I. Kopanakis, "Big data analytics: Applications, prospects and challenges," Mobile Big Data, pp. 3–20, 2018. https://doi.org/10.1007/978-3-319-67925-9_1
[13] A. Klašnja-Milićević, M. Ivanović, and B. Stantić, "Designing personalized learning environments—the role of learning analytics," Vietnam Journal of Computer Science, vol. 7, no. 3, pp. 231–250, 2020. https://doi.org/10.1142/S219688882050013X
[14] C. Janiesch, P. Zschech, and K. Heinrich, "Machine learning and deep learning," Electronic Markets, vol. 31, no. 3, pp. 685–695, 2021. https://doi.org/10.1007/s12525-021-00475-2
[15] Y. Anzai, "Pattern recognition and machine learning," Elsevier, pp. 407, 2012.
[16] L. Deng and X. Li, "Machine learning paradigms for speech recognition: An overview," IEEE Transactions on Audio, Speech, and Language Processing, vol. 21, no. 5, pp. 1060–1089, 2013. https://doi.org/10.1109/TASL.2013.2244083
[17] W. Srimaharaj and R. Chaisricharoen, "A novel processing model for P300 brainwaves detection," Journal of Web Engineering, pp. 2545–2570, 2021. https://doi.org/10.13052/jwe1540-9589.20815
[18] G. BakIr et al., "Predicting structured data," London: MIT press, 2007. https://doi.org/10.7551/mitpress/7443.001.0001
[19] I. Stoica, "Solving system problems with machine learning," Studies in Informatics and Control, vol. 28, no. 2, pp. 119–132, 2019. https://doi.org/10.24846/v28i2y201901
[20] A. Jain et al., "Overview and importance of data quality for machine learning tasks," in Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2020. pp. 3561–3562. https://doi.org/10.1145/3394486.3406477
[21] C. Romero, J. R. Romero, and S. Ventura, "A survey on pre-processing educational data," in Educational data mining. Springer, Cham, 2014, pp. 29–64. https://doi.org/10.1007/978-3-319-02738-8_2
[22] S. Kaisler, F. Armour, J. A. Espinosa, and W. Money, "Big data: Issues and challenges moving forward," in 2013 46th Hawaii international conference on system sciences, 2013, pp. 995–1004. https://doi.org/10.1109/HICSS.2013.645
[23] P. Leekitchwatana and P. Pimdee, "An analysis of Thai student teacher appropriate internet use behaviour," International Journal of Emerging Technologies in Learning (iJET), vol. 16, no. 2, pp. 254–271, 2021. https://doi.org/10.3991/ijet.v16i02.13747
[24] W. Limprasert and S. Kosolsombat, "A case study of data analysis for educational management," in 2016 13th International Joint Conference on Computer Science and Software Engineering (JCSSE). July, 2016, pp. 1–5. https://doi.org/10.1109/JCSSE.2016.7748897
[25] S. Bhornchareon, S. Prapinpongsakorn, and W. Techataweewan, "Documentary analysis for identifying the factors influencing information literacy among Thai undergraduate students," in ERPA International Congresses on Education. 2019, pp. 99–103.
[26] P. Nuankaew, "Self-regulated learning model in educational data mining," International Journal of Emerging Technologies in Learning (iJET), vol. 17, no. 17, pp. 4–27, Sep. 2022, https://doi.org/10.3991/ijet.v17i17.23623

[27] S. Petsuwan, P. Pimdee, and P. Pupat, "Strategies for using the satellite distance education system in Thailand that affect student quality of marginal schools in the lower North region," Revista ESPACIOS, vol. 40, no. 41, 2019. https://doi.org/10.2478/mjss-2019-0040
[28] M. B. Ulla, "Filipinos as EFL teachers in Bangkok, Thailand: Implications for language education policy in the ASEAN region," RELC Journal, vol. 52, no. 3, pp. 588–602, 2021. https://doi.org/10.1177/0033688219879775
[29] E. Rauch, M. De Marchi, P. Jitngernmadan, and F. M. Martin, "A descriptive analysis for education and training on Automation 4.0 in Thailand," in Proceedings of the 11th Annual International Conference on Industrial Engineering and Operations Management, Singapore. March 2021, pp. 7–11.
[30] Y. Vanpetch and M. Sattayathamrongthian, "The challenge and opportunities of Thailand education due to the covid-19 pandemic: Case study of Nakhon Pathom, Thailand," in E3S Web of Conferences. 2020, pp. 18058. https://doi.org/10.1051/e3sconf/202021018058
[31] F. Thompson, M. L. Hoang, and T. H. Ng, "Harnessing the potential of big data in post-pandemic Southeast Asia," Manila: The Asian Development Bank (ADB), pp. 1–6, 2021. https://doi.org/10.22617/BRF210100-2
[32] National Academy of Education, "Big data in education: Balancing the benefits of educational research and student privacy. A workshop summary," DC: ERIC Clearinghouse, 2017.
[33] UNESCO Institute for Statistics, "The data revolution in education," UNESCO, 2017 [Online]. Available: https://www.perlego.com/book/1671402/the-data-revolution-in-education-pdf, Accessed on: Aug. 1, 2022.
[34] K. S. Kyaw et al., "Towards intelligent system framework for smart university using big data analysis," in The 12th Hatyai National and International Conference, pp. 1718–1736, 2021.
[35] N. O'Mahony et al., "Machine learning algorithms for process analytical technology," in 2016 World Congress on Industrial Control Systems Security (WCICSS), pp. 1–7, 2016. https://doi.org/10.1109/WCICSS.2016.7882607
[36] N. O'Mahony et al., "Deep learning vs. traditional computer vision," in Science and information conference. Springer, Cham, pp. 128–144, 2019. https://doi.org/10.1007/978-3-030-17795-9_10
[37] P. Nuankaew, W. Nuankaew, K. Phanniphong, S. Imwut, and S. Bussaman, "Students model in different learning styles of academic achievement at the university of Phayao, Thailand," International Journal of Emerging Technologies in Learning, vol. 14, no. 12, 2019. https://doi.org/10.3991/ijet.v14i12.10352
[38] B. Sakulkueakulsuk, S. Witoon, P. Ngarmkajornwiwat, P. Pataranutaporn, Surareungchai, W., P. Pataranutaporn, and P. Subsoontorn, "Kids making AI: Integrating machine learning, gamification, and social context in STEM education," in 2018 IEEE international conference on teaching, assessment, and learning for engineering (TALE), December 2018, pp. 1005–1010. https://doi.org/10.1109/TALE.2018.8615249
[39] C. Polpanumas et al., "AI builders: Teaching Thai students to build end-to-end machine learning projects online," in 2021 IEEE International Conference on Engineering, Technology & Education (TALE), December 2021, pp. 565–572. https://doi.org/10.1109/TALE52509.2021.9678620
[40] J. Su and Y. Zhong, "Artificial Intelligence (AI) in early childhood education: Curriculum design and future directions," Computers and Education: Artificial Intelligence, vol. 3, pp. 100072, 2022. https://doi.org/10.1016/j.caeai.2022.100072
[41] N. Tavichaiyuth and E. Rattagan, "Developing chatbots in higher education: A case study of academic program chatbot in Thailand," pp. 1–12, 2021. [Online] https://as.nida.ac.th/wp-content/uploads/2021/11/6210422022_Nicha_Paper_PDF.pdf, Accessed on: Aug. 1, 2022.

[42] C. Visaltanachoti and T. Chantana Viriyavejakul, "Teaching english to Thai students using an artificial intelligence technology algorithmic model: A prototype analysis," Turkish Journal of Computer and Mathematics Education (TURCOMAT), vol. 12, no. 14, pp. 5623–5630, 2021.
[43] T. Panityakul, W. Suriyaamorn, R. Chinram, and N. Kongchouy, "Classification models for employability of statistics and related fields graduates from Thailand universities," Computer Science, vol. 17, no. 1, pp. 365–375, 2022.
[44] J. F. Pinto da Costa and M. Cabral, "Statistical methods with applications in data mining: A review of the most recent works," Mathematics, vol. 10, no. 6, pp. 993, 2022. https://doi.org/10.3390/math10060993
[45] C. Jayasree and K. K. Baseer, "Predicitng student performance to improve their employability by applying data mining and machine learning techniques," JCSE, vol. 6, pp. 1292–1308, 2018. https://doi.org/10.26438/ijcse/v6i7.12921308

8 Authors

Pratya Nuankaew is currently an instructor at the School of Information and Communication Technology, University of Phayao, Phayao, 56000, Thailand. (Email: pratya.nu@up.ac.th) He is the corresponding author of this research. His research interests are applied informatics technologies, behavioral sciences analysis with technologies, computer-supported collaborative learning, data science in education, educational data mining, learning analytics, and learning styles, learning strategies for lifelong learning, self-regulated learning, social network analysis, and ubiquitous computing.

Patchara Nasa-Ngium is currently a lecturer at the Faculty of Science and Technology, Rajabhat Maha Sarakham University, Maha Sarakham, 44000, Thailand. (Email: patchara@cs.rmu.ac.th) His research interests include artificial intelligence, machine learning, evolutionary computing, and data mining.

Tinnakorn Kunasit is currently a lecturer at the Faculty of Science and Technology, Rajabhat Maha Sarakham University, Maha Sarakham, 44000, Thailand. (Email: tinnakorn@cs.rmu.ac.th) His research interests include artificial intelligence, machine learning, evolutionary computing, and data mining.

Wongpanya Sararat Nuankaew is currently an assistant professor at the Faculty of Information Technology, Rajabhat Maha Sarakham University, Maha Sarakham, 44000, Thailand. (Email: wongpanya.nu@rmu.ac.th) Her research interests are digital education, innovation and knowledge management, data science, and big data and information technology management.

Article submitted 2022-11-16. Resubmitted 2023-02-01. Final acceptance 2023-02-02. Final version published as submitted by the authors.