



Machine Learning Predictive Model of Academic Achievement Efficiency based on Data Envelopment Analysis

Nor Faezah Mohamad Razi

Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, Perak Branch Tapah Campus, Perak, Malaysia
norfa122@uitm.edu.my

Norhayati Baharun

Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, Perak Branch Tapah Campus, Perak, Malaysia
norha603@uitm.edu.my

Nasiroh Omar

Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, Shah Alam, Selangor, Malaysia
nasiroh@uitm.edu.my

Article Info

Article history:

Received Feb 10, 2022
Revised March 18, 2022
Accepted Apr 27, 2022

Keywords:

Data Envelopment Analysis
Predictive model
Academic achievement
Machine learning
Auto-ML

ABSTRACT

Along the way with the changes in the education landscape nowadays, the grade is not the only determinant to predict the students' success. In the context of a student's academic performance, it is better to focus on measuring the efficiency of academic achievements that used multiple determinants of holistic outcome rather than just focus on the student grade. Data Analysis Envelopment (DEA) is a nonparametric method that widely used in many fields to measure performances efficiency but limited research has been reported on DEA in education domain. Acknowledging DEA time consuming issue when involving a huge size of data, recent research on deploying machine learning in DEA keeps on rapid progressing. This paper presents a new research framework of DEA and Auto-ML predictive model for the academic achievement efficiency. The framework includes variety options of machine learning to be compared from the conventional manual setting into the recent Auto-ML technique. The research framework will provide new insights into the decision-making process particularly in the education context.

Corresponding Author:

Nor Faezah Mohamad Razi

Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, Perak Branch Tapah Campus, Perak, Malaysia
Email: norfa122@uitm.edu.my

1. Introduction

Academic achievement is the extent to which a student or institution has achieved either short-term or long-term educational goals. It emerged in the knowledge, skills and behaviours acquired by any students in education environments[1]. Traditionally, in higher learning institutions, classifying whether students had succeeded in their academic or not was determined by their grade performance during the final examination. They acknowledge the students' achievement by considering the higher their Cumulative Grade Performance Average (CGPA), the more quality the



student-produced by higher education institutions. Overall, the current evaluation process uses grade as a predictor to represent the overall academic achievement of a student.

Along the way with the changes in the education landscape nowadays, the grade is not the only determinant to predict the students' success. More than the students' grade, student attitude [2]–[4] are fundamental and also a part of intelligent. In this 20th century, digital skills are highly valued in the future to accommodate technological advancements. [5] also enhance the important of digital skills in forming a good quality of a student holistically. Previous studies include these two factor in a separate setting using predictive models [6], descriptive statistics [7] and factor analysis [8]. Another promising way to effectively measure the student achievement is to combine all the determinants of student attitude, digital skills and the academic grade. To evaluate all the resources from these determinants will be a very complex measurement and focused on the performance's efficiency from the different groups of determinant appears to be more reliable. As a non-parametric method, Data Analysis Envelopment(DEA) has a better dispersion of result than the parametric method in measuring the students' performances efficiency[9].

DEA has been broadly used for evaluating the efficiency in many areas such as financial institution, manufacturing companies, hospitals, airlines and government agencies. Even though several research works have provided insights into the richness of DEA application in education literature to measure achievement efficiency, many aspects of efficiency in education still need to be explored particularly in determining academic achievement. This is because due to the changes in education landscape. Thus, with the DEA, it will help higher institution management and educator to evaluate the resources provided to students during their learning process and in turn can improve the quality of an academic achievement.

Nevertheless, DEA becomes critical when dealing with the appropriate selection of the variables. The selection input and output used to perform the analysis is essential for applying the method. In fact [10] stated that the most critical part in evaluating efficiency value is preparing the input and output. Failure to ensure effective selection of inputs and outputs will lead to ambiguity to decision-makers [11]. In the education field, the selection of input and output in achieving the student's academic achievement efficiency is volatile and complex [12], often overlooking the available input to support the academic achievement efficiency. Thus, the essential part that needs to be exploring in on the characterization of input and output selection. Therefore, cautious steps should be taken into account in determining input and output in this study. Fail to characterize input and output in DEA will lead to misspecification of the model.

Traditionally DEA cannot determine the optimal output [13]. DEA produces a single comprehensive measure of performance (efficiency score) for each Decision-Making Unit (DMU) based on a given set input and output variables. Efficiency score is treated as an indicator for performance evaluation of DMUs. Currently the process involves repetitive loop of a complex process of recalculation of efficiency score. When one DMUs added, efficiency score needs to be recalculated, which make the process very tedious and time consuming. In line with current data driven developments, this analysis is likely to become more complex. To resolve this issue, the use of machine learning in DEA has been very promising recently but the rapid progress in machine learning demands a more research to be conducted on the methods. This paper provides the basis knowledges on designing the framework of research on DEA and automated machine learning (Auto-ML) in the education domains.

2. Literature Review

2.1 Academic achievement

Academic achievement refers to the amount to which a person has achieved specified goals that were the focus of activities in educational settings, such as school, college, and university. Cognitive goals such as critical thinking or the development of information and comprehension in a single intellectual subject had been evaluated to the students in order to understand their intellectual domain (for instance numeracy, literacy, science and history).

The definition of academic achievement depends on the indicators used to measure it. This is due to the fact that academic achievement is very wide ranging and covers broad variety of educational outcomes. Many variables are believed to have an impact on academic achievement. The determination of the antecedents of academic achievement is based on the outcomes of the educational institutions. Questions that normally arise during the process of identifying the variables that affect academic achievement are:

-
- What is/are the factor(s) that affect academic achievement?
 - How are we going to help the student enhance their academic achievement?

The way questions been arisen with a different objective will lead to different goal in determining academic achievement. Thus, the selection of variables that can predict academic achievement is always ambiguous. Many approaches have been done previously to measure academic achievement and had widely been discussed particularly on the advantages and disadvantages in educational context. Previous studies found students' satisfaction [1], [14], students' motivation [15]–[17], their study habits and their abilities [18] and students' intellectual (IQ) level [18], [19] were the factors that contribute to student achievement. Other than that, Pal (2020) identified that demographic profile also contributed to the academic achievement which included parents' qualifications, parent occupations and parents' supports.

Many research used predicting approach to measure students' achievement [20]. The findings often considered as determinants of a student's excellence which are based on their academic achievement. However, rather than focusing on academic achievement, another measure that can be used is by determining the academic achievement efficiency. There is a difference between measuring academic achievement and academic achievement efficiency in summarizing the success of the students. Thus, academic achievement efficiency can be defined as the student's ability to solve a wide range of problems in helping education institutions to do decision making since it takes into account all the available resources that have been provided by the education institutions as to equip the students with a complete set of academic quality, while on the other hand academic achievement are mainly focus on identifying the determinant to students' achievement.

2.2 Academic achievement efficiency

Another way to measure the academic achievement is by using the efficiency score. Instead of focusing on academic achievement per se, there is a potential measure to be explored holistically on academic achievement efficiency. This is because the evaluation involved is not only focus on the predicted variable; it refers to the available allocation of its benefits and resources provided by the educational institutions.

Therefore, through efficiency values obtain we are able to identify whether the resources provided are at the optimal level or not, does the resources have been fully utilized by the students or not. The educational area is a very wide field to be covered since it covers a range of sectors from kindergarten, primary and secondary schooling, to post-compulsory and higher education [21]. This institution, however, can be seen to be a multi-product organization. Dealing with multiple outputs and input makes the process of decision-making process complex. Thus, the Multiple Criterion Decision Analysis [21] is a suitable tool to handle this situation. Early works explore Canonical Regression Analysis [22] to examine the production of multiple education outputs. However, this method does not provide measures of efficiency [21]. Therefore, an alternative multi-criteria decision analysis technique [23] able handle a production situation with multiple inputs and outputs. Yet, it does not require a prior specification of a functional form. It is known as Data Envelopment Analysis (DEA).

Over the years, DEA techniques have been widely used in many fields such as manufacturing [24], [25], banking [26], [27], transportation [28] and healthcare [29], [30] and many more. Initially, the DEA techniques were only applied to profit organizations, but it was much slower to use their application to non-profit organizations such as education areas [31], [32]. The main reason is that DEA has opened up possibilities for use in cases that have been resistant to other approaches because of the complex (often unknown) nature of the relations between the multiple inputs and multiple outputs.

Thus, to accommodate the entire practice of the development of DEA is nearly impossible. Other than that, only a few DEA studies acknowledge that the presence of endogeneity (such as due to omitted variable bias, measurement errors or selection bias) results in internal validity problems. Other than that, [33] agreed that any specific data may have different fits from different data mining techniques. Therefore, it suggested exploring further the application of the DEA model that would bridge the gap between the DEA efficiency in education literature and the parametric efficiency in education literature. There is some consideration on the application procedure when applying DEA analysis as the following.

- **Selection of DEA input and output**

DEA is generally introduced as a mathematical programming approach for measuring relative efficiencies of Decision-Making Unit (DMU) when multiple inputs and outputs are present. The important step that has to first look into is the selection of input and output variables [34], [35]. An important feature of DEA is its capability to provide efficiency scores while taking account of both multiple inputs and multiple outputs [11]. Traditionally in DEA model, it assumes that the status of each performance factor has a predetermined whether or not as an input or output. Sometimes it is known but sometimes, not known. The unclear known factor can be a flexible measure [36].

Generally, DEA minimizes input and maximize outputs. In other words, smaller levels of the former and larger levels of the latter represent better performance or efficiency. Therefore, one would have to classify these factors efficiently into input and output for use in DEA. Previous studies found that even though some researchers understand the concept of inputs and outputs well, it does not guide the input and output variables [37]. It is often the case that researchers take the notion for granted and little attention tends to be paid to ensuring that the selected measures properly reflect the process under study to the greatest extent possible. An example of the study conducted by [38], [39] involved the analysis of school districts in Texas. They developed in a ratio form of input and output; however, they provide little rationalization about appropriate variables (inputs and outputs) for studying student performance. Another example of unclear status (whether classified as input or output) of the variables is mentioned by [40] that in the efficiency measurement of university departments, the status of research income factor is unclear.

- **Selection of DMUs to be compared**

Two factors influence the selection of DMUs for a study. First factor is homogeneity where the DMUs should perform the same tasks and should have similar objectives. Second factor is the numbers of DMUs that must depend upon the objectives of the DEA study and on the number of homogenous units whose performance in practice has to be compared.

- **Choice of the DEA model**

Three models have been widely used in input maximizing or output maximizing, multiplier or envelopment, and constant or variable return to scale. The output-based formulation would be more appropriate when applications involve inflexible inputs (not fully under control). However, when the management's goals decide output rather than extracting the best possible performance of the DMUs, input-based DEA formulation may be appropriate. Multiplier versions are used when inputs and outputs are emphasized in an application, while envelopment versions are used when the relation among the DMUs are emphasized.

The choice of constant or variable returns to scale depends on the specific application. When the performance of DMUs is not normally expected to rely on the scale of operation (such as comparison of the performance of several large monopolies), constant returns scale (CSR) seems appropriate. Other than that, the variable return scale (VRS) may be a fair assumption.

2.3 Integration of DEA with machine learning

The drawback in DEA analysis also was found since it is deterministic approach. Dealing with huge size of data might cause a trouble to DEA method since its limited capabilities to handle producing efficiency score in a short time. Traditionally, the DEA method had to re-calculation and re-run the efficiency of all DMUs if a new DMU's was added. Since we face numerous datasets growing quickly, re-calculation or re-run of the process obtaining efficiency of the DMUs will become a tedious and never-ending story. To rectify this issue, [41] predicted the DEA efficiency of new DMUs by combining the DEA model with the machine learning algorithm.

With the rapid development of big data, datasets are growing rapidly and the field of education is also no exception in the face of a very rapid increase in its data management. These become a great challenge for the researchers because they have to face with the complex and big dataset. To handle on the complex dataset, automated machine learning or Auto-ML has been developed and keeps on progressing to be improved since its capabilities to handle with massive development of data better than the conventional machine learning. There are many approaches have been introduced for Auto-ML and the recent attention is given on the use of meta-heuristics optimization such as Genetic Programming. Meta-heuristics is a promising method in optimization

problems that widely used in many kinds of application domains. In the Auto Model, meta-heuristics can be used to optimize the optimal hyper-parameters of the machine learning algorithms based on the dataset specifications and needs as to achieve the highest possible of the algorithm efficiency.

3. Research Method

This part describes a new research framework on the academic achievement predictive model that used DEA and machine learning as illustrated in Figure 1.

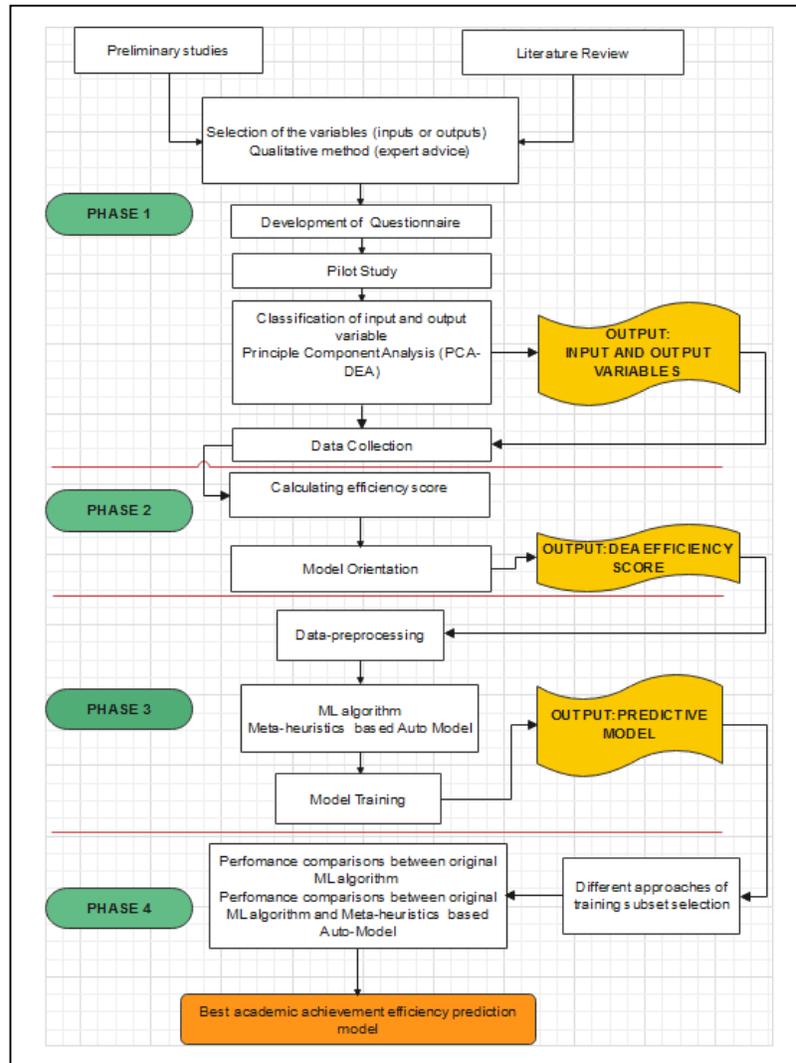


Figure 1. Overall research framework

3.1 Research framework

As seen in Figure 1, a four-phase method was proposed that assess the effect that large universities have on sector's overall efficiency performance which are variable selection, efficiency analysis using DEA, and the machine learning model evaluation. The variable selection methods are important because DEA is a non-parametric approach and loses discriminatory power as the dimensionality of the production space increases. Since there is no common rules and none fill gaps inherent regarding inputs and outputs selection [34], [42] specifically in determining academic

achievement efficiency. Therefore, the process of identification input and output variables will be done thoroughly to risk down misspecification issues in DEA model. Once this study able to identify the suitability of input and output through literature review and verified by the expert review to determine academic achievement efficiency, the process will proceed with the development of the Questionnaire. Further, a preliminary study will be conducted to obtain at least 50 samples to assess the reliability and validity of the questionnaire. Next the selections of the variable go through classification of input and output variable process. Then, the outcome that expected to be obtained in this phase is a set of input and output (Grades, Digital Skills and Student Attitude). Next, process of data collection process will be implemented.

DEA is known as the evaluation of the efficiency of the decision-making units (DMUs) that interact within a competition and development sector. DEA has become standard for the development of processes for comparing, measuring and evaluating efficiency in productive organizations. In this phase, the calculation of efficiency score by using DEA will be conducted by using DEA solver. Next process is to determine model orientation. The outcome in this phase is the efficiency score value. There are two choices of orientation in DEA that are input orientation and output orientation. The aim of input orientation is to minimize the inputs at given output level and the aim of the output orientation is to maximize the output given at the input level. This study will employ input orientation because it is assumed that the inputs regarding to academic achievement are controllable compared to outputs (academic achievement). The same orientation is used by [43]. The outcome in this phase is the efficiency score together with input variables.

Phase 3 will involve machine learning to build a predictive model based on the DEA efficiency score. This task is referred to a supervised because the model is constructed from data where the target/label/output is known (Efficiency based on Grade, Digital Skill and Attitude). Meta-heuristics-based Auto-ML and the empirical research findings of the approach is the new contribution in the research framework.

Last phase involves the process evaluation of predictive model for academic achievement efficiency. In this process, evaluating all the different machine learning algorithms from the conventional machine learning algorithms and the Auto-ML will be empirically conducted in predicting academic achievement efficiency and the compared with the Meta-Heuristic based an Auto-Model. Four performance measures that will be used to identify the best model are R squared, Root Mean Square Error (RMSE), and Mean Absolute Error (MEA). Figure 1 shows the research framework.

3.2 Data analysis procedure

Based on the research framework in Figure 1, the following describes a more details on the data analysis procedure of this research.

Phase 1: Variable selection

The variable selection techniques will be used to confirm the selection of input and output variables. By default, the output selection had been determined which are Grade, Digital skill and Students Attitude. However, it still needs to go analysis methods to confirm it. This stage implemented to avoid from having model misspecification in DEA efficiency estimates [37]. Four most widely used approaches to guide variable specification in DEA are Efficiency Contribution measure (ECM), Principal Component Analysis (PCA), regression test and bootstrapping for variable selection via Monte Carlo simulation.

The evaluation of the Questionnaire will involve Exploratory Factor Analysis (EFA). Exploratory Factor Analysis (EFA) will be conducted to explore the elements (questions) and the dimensions between variables (factors) and respondents. EFA is used to discover the underlying structure of a relatively large set of variables. The reliability analysis was used to measure scale reliability and provides information on the relationship between the individual items of the scale. Intra-class correlation coefficients may be used to compute inter-rater reliability estimates. This process was carried out once the EFA was completed.

Phase 2: Calculating efficiency score using DEA

The DEA model for the learning evaluation is used to identify efficient students and to investigate the performance efficiencies of students' academic achievement. Thus, the output variables of the DEA model are grade, digital skill and student's attitude, which is measured, individually, by the ratio between the actual performance objective value achieved and the expected performance objective value. Theoretically, the basic efficiency measure used in DEA is

the ratio of total outputs to total inputs. Once the efficiency score obtained, it will be determined by using benchmarked as presented in Table 1.

Table 1. Efficiency score

Categorize	Efficiency score	Slack variable
Strong Efficiency	1	All the slack variable are 0
Marginal Efficiency	1	At least 1 slack variable not being 0
Marginal Inefficiency	0.9	
Inefficiency	< 0.75	

Phase 3: Design and develop the academic achievement efficiency prediction models

In this stage the housekeeping process (cleaning) of the data will be conducted to ensure the data were turn into the format that's computer readable and understandable. In addition, this phase will check whether the data complete and free from bias. In a nutshell, data preparation is a set of procedures that helps make your dataset more suitable to be processed in machine learning. This process will start with depositing data into warehouses. These storages are usually created for structured records, meaning they fit into standard table formats. This approach is called Extract, Transform, and Load (ETL). Next process is checking on the quality of the data. Even through machine learning algorithm is a very powerful tool to analyse the data, having a poor data will harm the performance of the predictive model. Table 2 summarizes the important steps of the research in phase 3.

Table 2. Phase 3 activities

Step	Description
1	Data collection and ETL.
2	Machine learning algorithms design and development for the student academic efficiency prediction (conventional machine learning algorithms, meta-heuristics Auto-ML. Identify the most possible inputs/outputs of the DEA for the machine learning features selection.
3	Model training with the training dataset (commonly 70 percent of the dataset used for training).

Phase 4: Evaluating the academic achievement efficiency predictive model

This phase involves observing the model performances by evaluating the machine learning algorithms with the testing dataset. Compare the performances can be divided into two groups. First evaluation is to observe the performances between the conventional and Auto-ML. Second evaluation can be focused on the different setting of the meta-heuristics' parameters of the Auto-ML. Then, the best machine learning algorithm that identified based on the evaluation steps will be selected to predict students' achievement efficiency on the hold-out samples (new student dataset). In order to compare the performance of each classifier, three tools from which various accuracy measures are derived include: R squared, Root Mean Square Error (RMSE), Mean Absolute Error (MEA) and Kappa Statistics.

4. Results and Discussion

This section presents the new contributions of the proposed research framework presented in the Section 3. To the best of our knowledge based on the literature studies, the research framework contributes new knowledge to the following research gaps.

4.1 Determinants for measuring academic achievement

In educational context, there was a vast literature obtained on measuring the determinants of academic achievement and researchers in [16] found that the cognitive and metacognitive strategies in student learning skills and motivation are the factors contributed to academic performance variables. Even though studies on measuring academic achievement is seen continuously gained attention, the search for the determinant of academic achievement is also still

vague and need to do further study. Hence based on the record of recent studies presented in Table 3, it shows some potential areas for exploration since the result revealed only the characteristics of academic achievement (grades) has been used as a single outcome. Among the listed research framework in Table 3, no single study that measure digital competencies.

Table 3. Results of variables involved in measuring academic achievement

Research framework	[16]	[11]	[19]	[18]	[44]	[45]	[46]	[11]	[16]	[47]	[15]	[48]	[49]	[50]	[17]	New
Independent Variable																
Motivation	√								√		√				√	√
Learning Strategies	√															√
Outcome Expectation	√															√
Readiness for e-learning		√														√
Self-Regulated Skills		√						√				√			√	√
Satisfaction		√						√						√		√
Intellectual (I.Q.)			√	√												√
Sensor-aesthetic			√													√
Openness			√													√
Short term memory																
Study habits and their ability				√												√
Gender					√		√		√	√						√
Student Category					√											√
Discussion at home					√											√
Learning facility					√											√
Attendance					√											√
Parents Occupation					√											
Parents Qualification					√											
Parents Support												√				
Online Activities						√										√
Students Readiness								√								√
Learning Strategies									√							√
Emotional Intelligence										√			√			√
Collaboration														√		√
Self-Efficacy										√						√
Outcome Expectancy									√							√
Digital competencies																√
Belief, Values, Ethics											√					√
Social Competencies													√			√
Dependent Variable																
Grades	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
Academic achievement efficiency (student attitude, digital competencies, grades)																√

The model for measuring the students' achievement efficiency that needs to be designed with multiple students' outcome such as to combine multi-determinants from the students' attitude, digital competencies and grades will be more complex to be studied. Some variables from the listed research frameworks can be considered but more literatures are still needed. A well-designed inputs and outputs for the achievement efficiency model are highly critical to be a noteworthy literature contribution from the new research framework. Extensive research from the literatures, data collection and analysis will be involved to achieve the objective. Furthermore, the next research issue is to look into details the methodology gap of the DEA integration with Auto-ML.

4.2 DEA and machine learning

Refer to Table 4, different machine learning algorithms have been successfully used in predicting efficiency of DEA. The literatures revealed that different algorithms have different performance effects within the problem circumstances. Therefore, it is important to note that the process of choosing an appropriate machine learning algorithm is of great significance for improving the accuracy of prediction. Despite the listed machine learning models, there are options for Auto-ML, which has a great potential to be used in a complex prediction model with multi-

objectives of the academic achievement efficiency. Nevertheless, Auto-ML for machine learning in DEA is difficult to be found in the existing literature.

Table 4. Related studies machine learning DEA prediction

Research Framework	ML algorithm									
	DT	NN	SVM	RF	KNN	BPNN	GANN	ISVM	LR	Meta-heuristics Auto-ML
[41]		√	√			√	√	√		
[51]				√						
[13]		√				√				
[52]				√						
[53]	√			√						
[54]		√				√				
[55]		√								
[30]	√			√					√	
[56]	√	√		√						
[57]					√					
[58]			√		√				√	
[59]		√								
[60]		√								
[61]	√			√						
[62]		√				√				
New										√

The listed machine learning found in the literatures are Decision Tree (DT), Neural Network (NN), Support Vector Machine (SVM), Random Forest (RF), K-Nearest Neighbour (KNN), Genetic Algorithm Neural Network (GANN), Back Propagation Neural Network (BPNN), Incremental SVM (ISVM) and Logistic Regression (LR). As seen in Table 4, there is a gap of research on Auto-ML in the DEA prediction.

Evaluating different machine learning algorithms and identifying the appropriate hyper-parameters configuration for each of the algorithm is considerable complex and time-consuming tasks. Recent research on machine learning currently has directed towards Auto-ML to resolve this issue. Research on Auto-ML has provided a promising finding for accelerating the model design and implementation by optimizing the machine learning pipelines, including the algorithm selection and the hyper-parameters setting. Additionally, the interest of this research has been coined towards meta-heuristics in Auto-ML that will open a more possible of research directions.

5. Conclusion

The new research framework is expected to contribute fundamental knowledges to the theories, methodology and empirical gaps in the field of decision science generally and education performance measurement specifically. The selection of appropriate input and output to determine academic achievement efficiency will help the higher education institutions to produce excellent graduates holistically. In addition, with the changes in the education landscape recently, this study will be able to characterize the explicit input and output so as to determine the factor of academic achievement. The educators or researchers could add new knowledge in applying the Auto-ML to predict students' academic achievement efficiency.

Acknowledgements

The authors gratefully acknowledge Universiti Teknologi MARA for the support of this research.

Conflict of Interest

The authors declare no conflict of interest in the subject matter or materials discussed in this manuscript.

References

- [1] N. Yavuzalp, "A structural equation modeling analysis of relationships among university students' readiness for e-learning, self-regulation skills, satisfaction, and academic achievement," 2021.
- [2] K. Gjicali and A. A. Lipnevich, "Got Math Attitude? (In)direct Effects of Student Mathematics Attitudes on Intentions, Behavioral Engagement, and Mathematics Performance in the U.S. PISA," *Contemp. Educ. Psychol.*, vol. 67, no. September, p. 102019, 2021, doi: 10.1016/j.cedpsych.2021.102019.
- [3] W. A. Alamoudi *et al.*, "Correction to: Why do students skip classroom lectures: A single dental school report (BMC Medical Education, (2021), 21, 1, (388), 10.1186/s12909-021-02824-3)," *BMC Med. Educ.*, vol. 21, no. 1, pp. 1–11, 2021, doi: 10.1186/s12909-021-02866-7.
- [4] S. F. Kong and M. M. Mohd Effendi, "Sikap Pelajar Terhadap Implementasi Sains, Teknologi, Kejuruteraan dan Matematik (STEM) dalam Pembelajaran," *J. Dunia Pendidik.*, vol. 2, no. 3, pp. 72–81, 2020.
- [5] University Teknologi Mara Academic Affair Division, "Education 5 at UiTM - UiTM Academic Compass (3.5.2019).pdf." p. 95, 2019.
- [6] M. Al Fanah and M. A. Ansari, "Understanding E-learners' behaviour using data mining techniques," *ACM Int. Conf. Proceeding Ser.*, pp. 59–65, 2019, doi: 10.1145/3322134.3322145.
- [7] O. Kuzminska, M. Mazorchuk, N. Morze, V. Pavlenko, and A. Prokhorov, "Digital competency of the students and teachers in Ukraine: Measurement, analysis, development prospects," *CEUR Workshop Proc.*, vol. 2104, pp. 366–379, 2018.
- [8] X. Wang, Z. Wang, Q. Wang, W. Chen, and Z. Pi, "Supporting digitally enhanced learning through measurement in higher education: Development and validation of a university students' digital competence scale," *J. Comput. Assist. Learn.*, vol. 37, no. 4, pp. 1063–1076, 2021, doi: 10.1111/jcal.12546.
- [9] G. I. Farantos, "The Data Envelopment Analysis Method and the influence of a phenomenon in organizational Efficiency: A literature review and the Data Envelopment Contrast Analysis new application," *Data Envel. Anal. Decis. Sci.*, vol. 2015, no. 2, pp. 101–117, 2015, doi: 10.5899/2015/dea-00098.
- [10] H. Scheel, "EMS: Efficiency Measurement System User's Manual," pp. 1–12, 2000, [Online]. Available: <http://www.holger-scheel.de/ems/ems.pdf>.
- [11] W. D. Cook and J. Zhu, "Classifying inputs and outputs in data envelopment analysis," *Eur. J. Oper. Res.*, vol. 180, no. 2, pp. 692–699, 2007, doi: 10.1016/j.ejor.2006.03.048.
- [12] W. Lio and B. Liu, "Uncertain data envelopment analysis with imprecisely observed inputs and outputs," *Fuzzy Optim. Decis. Mak.*, vol. 17, no. 3, pp. 357–373, 2018, doi: 10.1007/s10700-017-9276-x.
- [13] K. Zhong, Y. Wang, J. Pei, S. Tang, and Z. Han, "Super efficiency SBM-DEA and neural network for performance evaluation," *Inf. Process. Manag.*, vol. 58, no. 6, p. 102728, 2021, doi: 10.1016/j.ipm.2021.102728.
- [14] H. Waheed, S. U. Hassan, N. R. Aljohani, J. Hardman, S. Alelyani, and R. Nawaz, "Predicting academic performance of students from VLE big data using deep learning models," *Comput. Human Behav.*, vol. 104, no. October 2019, p. 106189, 2020, doi: 10.1016/j.chb.2019.106189.
- [15] A. Habók, A. Magyar, M. B. Németh, and B. Csapó, "Motivation and self-related beliefs as predictors of academic achievement in reading and mathematics: Structural equation models of longitudinal data," *Int. J. Educ. Res.*, vol. 103, no. June, p. 101634, 2020, doi: 10.1016/j.ijer.2020.101634.
- [16] S. Nabizadeh, S. Hajian, Z. Sheikhan, and F. Rafiei, "Prediction of academic achievement based on learning strategies and outcome expectations among medical students," *BMC Med. Educ.*, vol. 19, no. 1, pp. 1–11, 2019, doi: 10.1186/s12909-019-1527-9.
- [17] C. R. Wibrowski, W. K. Matthews, and A. Kitsantas, "The Role of a Skills Learning Support Program on First-Generation College Students' Self-Regulation, Motivation, and Academic Achievement: A Longitudinal Study," *J. Coll. Student Retent. Res. Theory Pract.*, vol. 19, no. 3, pp. 317–332, 2017, doi: 10.1177/1521025116629152.
- [18] A. Quilez-Robres, A. González-Andrade, Z. Ortega, and S. Santiago-Ramajo, "Intelligence

- quotient, short-term memory and study habits as academic achievement predictors of elementary school: A follow-up study," *Stud. Educ. Eval.*, vol. 70, 2021, doi: 10.1016/j.stueduc.2021.101020.
- [19] T. Gatzka, "Aspects of openness as predictors of academic achievement," *Pers. Individ. Dif.*, vol. 170, no. September 2020, p. 110422, 2021, doi: 10.1016/j.paid.2020.110422.
- [20] A. Petersen, V. V. Ajanovski, and C. Messom, "Predicting Academic Performance: A Systematic Literature Review," pp. 175–199, 2018.
- [21] J. Johnes, "Operational research in education," *Eur. J. Oper. Res.*, vol. 243, no. 3, pp. 683–696, 2015, doi: 10.1016/j.ejor.2014.10.043.
- [22] J. F. Chizmar and T. A. Zak, "Canonical estimation of joint educational production functions," *Econ. Educ. Rev.*, vol. 3, no. 1, pp. 37–43, 1984, doi: 10.1016/0272-7757(84)90006-2.
- [23] J. Doyle and R. Green, "Data envelopment analysis and multiple criteria decision making," *Omega*, vol. 21, no. 6, pp. 713–715, 1993, doi: 10.1016/0305-0483(93)90013-B.
- [24] J. Alcaraz, L. Anton-Sanchez, J. Aparicio, J. F. Monge, and N. Ramón, "Russell Graph efficiency measures in Data Envelopment Analysis: The multiplicative approach," *Eur. J. Oper. Res.*, vol. 292, no. 2, pp. 663–674, 2021, doi: 10.1016/j.ejor.2020.11.001.
- [25] K. Pliakos, S. H. Joo, J. Y. Park, F. Cornillie, C. Vens, and W. Van den Noortgate, "Integrating machine learning into item response theory for addressing the cold start problem in adaptive learning systems," *Comput. Educ.*, vol. 137, no. June 2018, pp. 91–103, 2019, doi: 10.1016/j.compedu.2019.04.009.
- [26] Y. Liu and M. Schumann, "Data mining feature selection for credit scoring models," *J. Oper. Res. Soc.*, vol. 56, no. 9, pp. 1099–1108, 2005, doi: 10.1057/palgrave.jors.2601976.
- [27] A. Nandy and P. K. Singh, "Application of fuzzy DEA and machine learning algorithms in efficiency estimation of paddy producers of rural Eastern India," *Benchmarking*, vol. 28, no. 1, pp. 229–248, 2021, doi: 10.1108/BIJ-01-2020-0012.
- [28] V. S. Özsoy and H. H. Örkücü, "Structural and operational management of Turkish airports: a bootstrap data envelopment analysis of efficiency," *Util. Policy*, vol. 69, no. December 2019, 2021, doi: 10.1016/j.jup.2021.101180.
- [29] A. Tayal, U. Kose, A. Solanki, A. Nayyar, and J. A. M. Saucedo, "Efficiency analysis for stochastic dynamic facility layout problem using meta-heuristic, data envelopment analysis and machine learning," *Comput. Intell.*, vol. 36, no. 1, pp. 172–202, 2020, doi: 10.1111/coin.12251.
- [30] Y. Xu, Y. S. Park, and J. D. Park, "Measuring the Response Performance of U.S. States against COVID-19 Using an Integrated DEA, CART, and Logistic Regression Approach," 2021.
- [31] B. Montoneri, "Teaching Improvement Model Designed with DEA Method and Management Matrix," *IAFOR J. Educ.*, vol. 2, no. 1, pp. 125–155, 2014, doi: 10.22492/ije.2.1.05.
- [32] S. K. Parahoo, M. I. Santally, Y. Rajabalee, and H. L. Harvey, "Designing a predictive model of student satisfaction in online learning," *J. Mark. High. Educ.*, vol. 26, no. 1, pp. 1–19, 2016, doi: 10.1080/08841241.2015.1083511.
- [33] A. L. M. Anouze and I. Bou-Hamad, "Data envelopment analysis and data mining to efficiency estimation and evaluation," *Int. J. Islam. Middle East. Financ. Manag.*, vol. 12, no. 2, pp. 169–190, 2019, doi: 10.1108/IMEFM-11-2017-0302.
- [34] N. A. A. Aziz, R. M. Janor, and R. Mahadi, "Comparative Departmental Efficiency Analysis within a University: A DEA Approach," *Procedia - Soc. Behav. Sci.*, vol. 90, no. InCULT 2012, pp. 540–548, 2013, doi: 10.1016/j.sbspro.2013.07.124.
- [35] Y. Xiaoming, C. J. Shieh, and W. C. Wu, "Measuring distance learning performance with data envelopment analysis," *Eurasia J. Math. Sci. Technol. Educ.*, vol. 10, no. 6, pp. 559–564, 2014, doi: 10.12973/eurasia.2014.1217a.
- [36] M. Toloo, B. Ebrahimi, and G. R. Amin, "New data envelopment analysis models for classifying flexible measures: The role of non-Archimedean epsilon," *Eur. J. Oper. Res.*, vol. 292, no. 3, pp. 1037–1050, 2021, doi: 10.1016/j.ejor.2020.11.029.
- [37] N. R. Nataraja and A. L. Johnson, "Guidelines for using variable selection techniques in data envelopment analysis," *Eur. J. Oper. Res.*, vol. 215, no. 3, pp. 662–669, 2011, doi: 10.1016/j.ejor.2011.06.045.
- [38] A. Charnes, W. W. Cooper, and E. Rhodes, "Measuring the efficiency of decision making units," *Eur. J. Oper. Res.*, vol. 2, no. 6, pp. 429–444, 1978, doi: 10.1016/0377-2217(78)90138-8.

- [39] A. Charnes, W. W. Cooper, and E. Rhodes, "Evaluating Program and Managerial Efficiency: An Application of Data Envelopment Analysis to Program Follow Through," *Manage. Sci.*, vol. 27, no. 6, pp. 668–697, 1981, doi: 10.1287/mnsc.27.6.668.
- [40] J. E. Beasley, "Comparing university departments," *Omega*, vol. 18, no. 2, pp. 171–183, 1990, doi: 10.1016/0305-0483(90)90064-G.
- [41] N. Zhu, C. Zhu, and A. Emrouznejad, "A combined machine learning algorithms and DEA method for measuring and predicting the efficiency of Chinese manufacturing listed companies," *J. Manag. Sci. Eng.*, no. xxxx, pp. 1–14, 2021, doi: 10.1016/j.jmse.2020.10.001.
- [42] M. Gabriela, M. Peixoto, M. Andreotti, and M. C. Ang, "Performance management in hospital organizations from the perspective of Principle Component Analysis and Data Envelopment Analysis: the case of Federal University Hospitals in Brazil," *Comput. Ind. Eng.*, vol. 150, no. May, 2020, doi: 10.1016/j.cie.2020.106873.
- [43] E. M. Vallespin, "An application of the Data Envelopment Analysis Methodology in the performance assessment of the Zaragoza University Departments," *Doc. Trab. (Universidad Zaragoza. Fac. Econ. y Empres., no. 6, pp. 1–19, 2003, [Online]. Available: <https://dialnet.unirioja.es/servlet/articulo?codigo=1370473%0Ahttps://dialnet.unirioja.es/servlet/citart?info=link&codigo=1370473&orden=57286>*.
- [44] S. Pal, "<Application_of_Machine_Learning_Algorithms_to_Flow.pdf>," no. June, 2020.
- [45] R. Hasan, S. Palaniappan, S. Mahmood, A. Abbas, K. U. Sarker, and M. U. Sattar, "Predicting student performance in higher educational institutions using video learning analytics and data mining techniques," *Appl. Sci.*, vol. 10, no. 11, 2020, doi: 10.3390/app10113894.
- [46] G. Ramaswami, T. Susnjak, A. Mathrani, J. Lim, and P. Garcia, "Using educational data mining techniques to increase the prediction accuracy of student academic performance," *Inf. Learn. Sci.*, vol. 120, no. 7–8, pp. 451–467, 2019, doi: 10.1108/ILS-03-2019-0017.
- [47] H. Yazici, S. Seyis, and F. Altun, "Emotional intelligence and self-efficacy beliefs as predictors of academic achievement among high school students," *Procedia - Soc. Behav. Sci.*, vol. 15, pp. 2319–2323, 2011, doi: 10.1016/j.sbspro.2011.04.100.
- [48] D. Choe, "Parents' and adolescents' perceptions of parental support as predictors of adolescents' academic achievement and self-regulated learning," *Child. Youth Serv. Rev.*, vol. 116, no. January, p. 105172, 2020, doi: 10.1016/j.childyouth.2020.105172.
- [49] Z. Ahmed, M. Asim, and J. Pellitteri, "Emotional intelligence predicts academic achievement in Pakistani management students," *Int. J. Manag. Educ.*, vol. 17, no. 2, pp. 286–293, 2019, doi: 10.1016/j.ijme.2019.04.003.
- [50] H. Abuhassna, W. M. Al-Rahmi, N. Yahya, M. A. Z. M. Zakaria, A. B. M. Kosnin, and M. Darwish, "Development of a new model on utilizing online learning platforms to improve students' academic achievements and satisfaction," *Int. J. Educ. Technol. High. Educ.*, vol. 17, no. 1, 2020, doi: 10.1186/s41239-020-00216-z.
- [51] K. Thaker, V. Charles, A. Pant, and T. Gherman, "A DEA and random forest regression approach to studying bank efficiency and corporate governance," *J. Oper. Res. Soc.*, vol. 0, no. 0, pp. 1–28, 2021, doi: 10.1080/01605682.2021.1907239.
- [52] E. D. La Hoz, R. Zuluaga, and A. Mendoza, "Assessing and classification of academic efficiency in engineering teaching programs," *J. Effic. Responsib. Educ. Sci.*, vol. 14, no. 1, pp. 41–52, 2021, doi: 10.7160/ERIESJ.2021.140104.
- [53] N. Aydin and G. Yurdakul, "Assessing countries' performances against COVID-19 via WSIDEA and machine learning algorithms," *Appl. Soft Comput. J.*, vol. 97, p. 106792, 2020, doi: 10.1016/j.asoc.2020.106792.
- [54] B. Singpai and D. Wu, "Using a DEA–Auto-ML approach to track SDG achievements," *Sustain.*, vol. 12, no. 23, pp. 1–26, 2020, doi: 10.3390/su122310124.
- [55] N. Almaskati, R. Bird, D. Yeung, and Y. Lu, "A horse race of models and estimation methods for predicting bankruptcy," *Adv. Account.*, vol. 52, p. 100513, 2021, doi: 10.1016/j.adiac.2021.100513.
- [56] P. Appiahene, Y. M. Missah, and U. Najim, "Predicting Bank Operational Efficiency Using Machine Learning Algorithm: Comparative Study of Decision Tree, Random Forest, and Neural Networks," *Adv. Fuzzy Syst.*, vol. 2020, 2020, doi: 10.1155/2020/8581202.
- [57] M. Rahmanidoust and J. Zheng, "Evaluation of Factors Affecting Employees' Performance Using Artificial Neural Networks Algorithm: The Case Study of Fajr Jam," *Int. Bus. Res.*, vol. 12, no. 10, p. 86, 2019, doi: 10.5539/ibr.v12n10p86.

-
- [58] A. Gupta, M. Kohli, and N. Malhotra, "Classification based on Data Envelopment Analysis and supervised learning: A case study on energy performance of residential buildings," *1st IEEE Int. Conf. Power Electron. Intell. Control Energy Syst. ICPEICES 2016*, pp. 1–5, 2017, doi: 10.1109/ICPEICES.2016.7853706.
- [59] N. Misiunas, A. Oztekin, Y. Chen, and K. Chandra, "DEANN: A healthcare analytic methodology of data envelopment analysis and artificial neural networks for the prediction of organ recipient functional status," *Omega (United Kingdom)*, vol. 58, pp. 46–54, 2016, doi: 10.1016/j.omega.2015.03.010.
- [60] A. Alinezhad, "An Integrated DEA and Data Mining Approach for Performance Assessment," *Iran. J. Optim. Iran. J. Optim. Iran. J. Optim.*, vol. 8, no. 82, pp. 59–69, 2016.
- [61] H. Y. Kao, T. K. Chang, and Y. C. Chang, "Classification of hospital web security efficiency using data envelopment analysis and support vector machine," *Math. Probl. Eng.*, vol. 2013, 2013, doi: 10.1155/2013/542314.
- [62] A. Emrouznejad and E. Shale, "A combined neural network and DEA for measuring efficiency of large scale datasets," *Comput. Ind. Eng.*, vol. 56, no. 1, pp. 249–254, 2009, doi: 10.1016/j.cie.2008.05.012.

Biography of all authors

Picture	Biography	Authorship contribution
	<p>Nor Faezah Mohamad Razi is a lecturer in University Teknologi MARA (UiTM). She is the Resource Person for Fundamentals of Research. Currently she pursuing her Phd study in Decision Science area. Her research interest in Structural Equation Modeling (SEM), Data Envelopment Analysis and Educational technology. Email: norfa122@uitm.edu.my</p>	<p>Drafting, literature review, methodology and final checking the article</p>
	<p>Dr. Norhayati Baharun is an Associate Professor of Statistics, Universiti Teknologi MARA Perak Branch, Tapah Campus. She received her PhD in Statistics Education from the University of Wollongong Australia in 2012. Her career started as an academic from January 2000 to date at the Universiti Teknologi MARA that specialized in statistics. Other academic qualifications include both Master Degree and Bachelor Degree in Statistics from Universiti Sains Malaysia and Diploma in Statistics from Institute Teknologi MARA. She is also a certified Professional Technologist of the Malaysia Board of Technologist (MBOT), a Fellow Member of the Royal Statistical Society (RSS), London, United Kingdom, a Professional Member of the Association for Computing Machinery (ACM), New York, USA, and a Certified Neuro Linguistic Program (NLP) Coach of the Malaysia Neuro Linguistic Program Academy. Email: norha603@uitm.edu.my</p>	<p>Literature review and analysis</p>
	<p>Assoc. Prof Dr Nasiroh Omar is a lecturer and researcher at Universiti Teknologi MARA, Shah Alam. She graduated with PhD from Nottingham University, United Kingdom, in computer science. Currently she is working on computer-based behavioral analytics in humanities. Her research interests are data visualization, data analytics and scientific storytelling. Email: nasiroh@tmsk.uitm.edu.my.</p>	<p>Methodology</p>