MJFAS
Malaysian Journal of Fundamental and Applied Sciences

RESEARCH ARTICLE

Performance Measurement: Machine Learning as a Complement to DEA for Continuous Efficiency Estimation

Yousef Khoubrane[a], Noor Asiah Ramli[b]* Siti Shaliza Mohd Khairi[b]

[a]EMINES - School of Industrial Management, University Mohammed VI Polytechnic, Hay Moulay Rachid Ben Guerir, 43150, Morocco; [b]School of Mathematical Sciences, College of Computing, Informatics and Mathematics, Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia

Abstract Data Envelopment Analysis (DEA) is a well-established non-parametric technique for performance measurement to assess the efficiency of Decision-Making Units (DMUs). However, its inability to predict the efficiency values of new DMUs without re-conducting the analysis on the entire dataset has led to the integration of Machine Learning (ML) in previous studies to address this limitation. Yet, such integration often lacks a thorough evaluation of ML's adaptability in replacing the current DEA process. This paper presents the results of an empirical study that employed eight ML models, two DEA variants, and a dataset of S&P500 companies. The findings demonstrated ML's remarkable precision in predicting efficiency values derived from a single DEA run and comparable performance in predicting the efficiency of new DMUs, thus eliminating the need for repeated DEA. This discovery highlights ML's robustness as a complementary tool for DEA in continuous efficiency estimation, rendering the practice of re-conducting DEA unnecessary. Notably, boosting models within the Ensemble Learning category consistently outperformed other models, highlighting their effectiveness in the context of DEA efficiency prediction. Particularly, CatBoost demonstrated its superiority as the top-performing model, followed by LightGBM in the second position in most cases. When extended to five enlarged datasets, it shows that the model exhibits superior R^2 values in the CRS scenario.

Keywords: Data Envelopment Analysis (DEA), Machine Learning (ML), Performance Measurement.

Introduction

Performance measurement has experienced a continuous evolution from straightforward single-input and single-output ratio analysis to the more sophisticated framework of relative efficiency evaluation using DEA [1, 2]. In the present day, the domain of performance measurement has transitioned into a new era characterized by the incorporation of ML in conjunction with DEA.

DEA is a non-parametric linear programming-based technique that holds extensive application in assessing the efficiency of DMUs across diverse domains encompassing businesses [3], financial institutions [4], educational establishments [5, 6], healthcare facilities [7, 8], and software projects [9]. Moreover, by correlating efficiency to other relevant indicators, the relevance of DEA extends to fields such as risk management and bankruptcy prediction. Emrouznejad et. al [10] have done a survey for a comprehensive bibliography of DEA where the authors reported a total of 10,300 articles published on the subject until the end of 2016.

Nevertheless, DEA is constrained by certain inherent limitations. Specifically, because the assessment of a DMU's efficiency is conducted with respect to other units within the dataset, DEA needs to be reconducted each time a new DMU is introduced into the dataset. Moreover, the methodology entails solving as many minimization problems as there are DMUs under consideration. This requires substantial computational resources and results in extended processing time, particularly when handling extensive datasets [3].

***For correspondence:**
asiah2206@uitm.edu.my

Received: 30 Nov. 2023
Accepted: 4 Feb. 2024

In response to these scarcities, ML is employed to enhance the DEA process. Machine learning (ML) methods are the combination of computational and mathematical tools that enables the selected algorithm to find the patterns of data efficiently for further assessments. In efficiency assessment, ML is widely used with DEA due to the current research interest on the prediction of efficiency measure for improvement in various applications [11, 12, 13]. The process involved two-stages analysis where in the first stage, the efficiency scores were generated from the DEA models. Next, the scores were trained with ML models for prediction which disclosed the patterns that connect inputs and outputs for efficiency. Through this learning process, the model generalizes these patterns, giving it the capacity to predict the efficiencies of newly added DMUs. Consequently, this eliminates the necessity to repeatedly conduct DEA analyses whenever new units are introduced, simplifying the overall assessment process, and addressing the computational challenges associated with DEA.

Previous literature that merged DEA with ML was typically limited to evaluating the combined model on a single dataset where efficiency values result from a single DEA run [3]. However, to validate the capacity of ML to replace the conventional practice of re-executing DEA when assessing the efficiency of new DMUs, it is imperative to undertake a thorough evaluation by directly comparing with the DEA re-execution scenario.

This paper has two primary objectives. The first empirical study seeks to determine whether the ML model trained on the training set of a DEA efficiency dataset can accurately estimate the efficiency values obtained from the same DEA run. The second empirical study aims to ascertain whether the ML model trained on the efficiency values derived from one DEA run can effectively predict the newly updated values resulting from a subsequent DEA run conducted upon the introduction of additional DMUs to the dataset.

To achieve these objectives, this paper utilized eight ML models and evaluated their performance against two DEA models: Constant Return to Scale (CRS) and Variable Return to Scale (VRS). These two DEA models were utilized to determine which one is more suitable for making predictions using ML, thereby contributing to the existing body of literature. All the results presented in this paper were derived from the empirical analysis conducted on the proposed dataset.

The incorporation of ML in this study is substantiated by existing literature, which consistently affirms its effectiveness in DEA. The selection of eight ML models is deliberate, encompassing widely recognized models within the field. Support Vector Regression (SVR) is specifically included as a benchmarking model and the remaining seven models are chosen for their proven high performance as ensemble learning models. It is worth highlighting that six out of the eight models were also used in the paper of Zhang *et al*. [14]. By opting for a diverse set of high-performing models rather than relying on a single model, the aim is to enable a comparative analysis of the models under consideration, thus improving the overall performance of ML in the study. Remarkably, this paper excludes Neural Networks due to the predominant emphasis on these models in prior research, as stated by Zhu *et al*. [3], and the relative scarcity of exploration into alternative ML models.

The remainder of this paper is structured as follows. Section 2 provides a comprehensive review of the relevant prior literature. In Section 3, the methodology employed in this study is described in detail. Section 4 is dedicated to the presentation and discussion of the results obtained from the empirical analysis. Finally, Section 5 serves as the conclusion, summarizing the key findings and their implications.

Literature Review

This paper builds upon prior efforts to integrate DEA and ML. The earliest attempt to combine DEA and regression was by Bowlin *et al*. [15]. Their primary objective was to compare the effectiveness of the two techniques as analytical tools. To facilitate this comparison, they required a benchmark criterion. The benchmark was created by adopting an approach involving the generation of a dataset. This dataset was constructed based on the relationships between inputs and outputs in an efficient scenario. Subsequently, certain input values were adjusted downwards, while others were maintained at their optimal levels. This manipulation ensured that the dataset included both known efficient and inefficient DMUs. The results of their study indicated that DEA outperformed linear regression. However, it is important to note that this is a rare instance where regression was found to be the less effective method. This outcome can be attributed, at least in part, to the relatively small dataset used in their study, comprising only 15 DMUs. Linear regression relies on the least squares estimator, which becomes more accurate with a larger dataset. Moreover, linear regression is considered one of the weaker ML models, particularly in non-linear contexts.

Following this initial attempt with regression, various studies combining DEA and ML emerged. A significant number of these studies focused on neural networks. Athanassopoulos and Curram [16] were the pioneers in this domain. They regarded DEA and Artificial Neural Networks (ANNs) as comparable non-parametric techniques for evaluating performance. Their findings demonstrated that, despite the disparities between both methods, both provided valuable insights into performance assessment.

The subsequent works placed a stronger emphasis on the integration between DEA and ML to address the inherent limitations of DEA. For example, Salehi *et al*. [17] examined various factors affecting adaptive capacity in a petrochemical plant. They first used DEA and then applied Multilayer Perceptron (MLP), an Artificial Neural Network (ANN) based model, for future estimations. This approach allowed them to create a model capable of predicting abnormal conditions. Later, Jomthanachai *et al*. [18] introduced an integrated method combining DEA and ML to assess risk levels based on DMU efficiency. Their use of the DEA cross-efficiency model has enhanced the Failure Mode and Effect Analysis (FMEA) technique in the field, with the integrated ML component providing DEA with predictive capabilities, thus addressing one of DEA's limitations.

Nishtha *et al*. [19] and Guerrero *et al*. [20] have both explored enhancing DEA by incorporating the predictive capabilities of Support Vector Machines (SVMs). Nishtha's team opted for an empirical approach. They discovered that the combination of DEA with Support Vector Regressor, which they termed DEA-SVR, exhibited remarkable efficacy in estimating efficiency, especially when dealing with imprecise data, as demonstrated in the context of Indian Banks. Meanwhile, Guerrero *et al*. [20] pursued a theoretical approach, introducing what they referred to as a DEA Machine (DEAM) model. They conducted a comparative analysis, contrasting DEAM against traditional DEA while employing Cobb–Douglas production functions. These functions are frequently used in econometrics to establish the relationship between input and output efficiency. Their findings showed that DEAM surpassed DEA in estimating production functions.

Further research, Zhong *et al*. [21] and Zhang *et al*. [14] highlighted a limitation of DEA, which is that it constructs a relative efficient frontier. This means that the efficiency values of DMUs are relative to the specific dataset that underwent DEA and are not benchmarked against a theoretical maximum. To address this limitation, both research teams employed the Slack Based Method (SBM) model of DEA to establish an absolute efficient frontier. They achieved this by utilizing and comparing 15 [21] and 11 [14] ML models, respectively. Remarkably, in both studies, the standout performer among the ML models was the Back Propagation Neural Networks (BPNN) model. Zhong *et al*. [21] incorporated the concept of Super Efficiency in DEA and found that the constructed absolute frontier outperformed previous alternatives in the literature. Taking their research a step further, Zhang *et al*. [14] delved into the optimal partitioning of the train-test dataset. They classified the derived efficiency values into four datasets using the quartile method and examined various combinations of three sets to determine the most effective training set. Their results showed that BPNN demonstrated superior performance when the training set excluded efficiency values falling between 0.35 and 0.43.

Another research that has been done by Zhu *et al*. [3] performs an association between the DEA method and ML algorithms and proposes an option that can combine between DEA and ML (ML-DEA) algorithms to determine the efficiency scores of DMUs. Four ML-DEA algorithms are discussed, which are DEA-CCR model combined with Back-Propagation Neural Network (BPNN-DEA), with Genetic Algorithm (GA) integrated with Back-Propagation Neural Network (GANN-DEA), with Support Vector Machines (SVM-DEA), and with Improved Support Vector Machines (ISVM-DEA). The findings show that the average accuracy of the predicted efficiency of DMUs is about 94%, and the comprehensive performance order of the four ML-DEA algorithms ranked from good to poor is GANN-DEA, BPNN-DEA, ISVM-DEA, and SVM-DEA.

While DEA provides continuous efficiency values from 0 to 1, not all research that combines DEA with ML uses regression. Hong *et al*. [22], for instance, divided the efficiency values in their dataset into four tiers by placing the efficient DMUs in the first tier, reconducting DEA on the inefficient values and placing the new efficient DMUs in the second tier, and repeating the process until all four tiers were established. The C4.5 algorithm, a supervised learning technique for creating decision trees, was employed to classify DMUs into these tiers. Similarly, Visbal-Cadavid *et al*. [23] employed a classification approach. They integrated DEA with ANNs to predict efficiency among Colombian higher education institutions. This two-stage approach provided DEA with predictive capabilities, enhancing its evaluative performance in the context of the study.

Numerous papers which addressed the performance measurement using the DEA model and integrating it with MLhave become popular nowadays. Further research includes Babaei Keshteli and Rostamy [24], Appiahene *et al*. [25], Wei *et al*. [26] and Thaker *et al*. [27].

Based on our review of the literature that combines DEA with ML, it is evident that (1) previous studies have acknowledged the value of enhancing DEA with ML to address its limitations. This reinforces the credibility of the approach adopted in this paper. Furthermore, (2) while prior research has confirmed the effectiveness of this combined approach in predicting the efficiency values for new DMUs, the evaluation has typically been limited to testing on a single dataset where efficiency values result from a single DEA run. This paper extends the evaluation process by conducting an empirical study to assess the use of a combined DEA-ML model for continuous efficiency estimation, comparing it to the practice of re-running DEA. Moreover, (3) in contrast to a significant portion of the classical literature that often focuses on a single DEA model or a single ML model, this paper stands out by not only comparing the performance of eight ML models but also examining them in relation to two DEA models: CRS and VRS, thus enriching the existing body of literature on the subject.

Methodology

The dataset [28] used in this study was obtained from the Kaggle platform. The provider of the dataset scrapped the data from Yahoo Finance API. The dataset included the financial reports of S&P500 companies for the four quarters of the year 2020/2021 (i.e. December 2020, March 2021, June 2021 and September 2021). S&P500 companies consist of 502 largest companies on the United States (U.S.) stock market whose performance is tracked by the S&P500 Index, namely the Standard and Poor's 500 Index. This index is widely considered a representation of the stock market performance in the U.S. After performing data cleaning and studying the correlation between the original columns of the dataset, the study dropped some features due to their strong correlation with the remaining features. Repetitive and null rows were also dropped.

The features utilized as inputs for the DEA, comprise Research Development, Selling General Administrative, Interest Expense, Income Tax Expense, and Cost of Revenue. On the other hand, the output features for DEA, namely, Net Income, Gross Profit, Operating Income, Earnings Before Interest and Tax (Ebit), Total Revenue, and Total Other Income Expense Net. In total, there are about 2000 DMUs used based on four quarters of the year 2020/2021. All values within the dataset that presented in Table 1 are expressed in USD.

Table 1. Distribution of Numerical Variables Within the Dataset (in Millions USD)

	Feature Name	Average Value	Standard Deviation	Minimum Value	Maximum Value
Input	Research Development	199.45	893.92	0.00	16466.00
	Selling General Administrative	1078.12	2487.83	0.00	30331.00
	Interest Expense	103.65	167.74	0.00	1972.00
	Income Tax Expense	203.75	424.23	0.00	6010.00
	Cost of Revenue	3979.29	9302.36	0.00	115000.00
Output	Net Income	889.91	2043.95	0.00	28755.00
	Gross Profit	2589.00	5039.29	0.00	48904.00
	Operating Income	1098.85	2240.18	0.80	33534.00
	Ebit	952.59	2075.50	0.00	33534.00
	Total Revenue	6546.78	12963.76	0.00	152000.00
	Total Other Income Expense Net	263.27	940.65	0.00	25437.00

Data Envelopment Analysis

DEA is a linear programming technique used to measure the relative efficiency of DMUs based on their use of multiple inputs to produce multiple outputs. After distinguishing a subset of DMUs as "best practice", DEA assesses the efficiency of other DMUs in comparison to this frontier. Efficiency scores range from 0 to 100 percent relative to the best performer, with 100 percent indicating the best efficiency. DEA's utility extends beyond merely measuring efficiency; it also provides information on inefficient DMUs and their efficient peers. It serves as a benchmarking tool, allowing decision makers to identify and understand inefficiencies by comparing DMUs with similar profiles.

Having its origins in the seminal work of Charnes *et al*. [1], DEA is a technique rooted in mathematical programming that finds extensive application in the field of Operations Research and Management Science.

The following equation provides the ratio form of the basic DEA model with an output orientation [29]. It is a maximization problem which, when solved, will provide the values of weights a_i and b_j, which will maximize the efficiency of DMU m.

$$\text{Max} \quad \frac{\sum_{j=1}^{J} b_{jm} y_{jm}}{\sum_{i=1}^{I} a_{im} x_{im}} \tag{1}$$

Subject to

$$0 \le \frac{\sum_{j=1}^{J} b_{jm}\ y_{jn}}{\sum_{i=1}^{I} a_{im}\ x_{in}} \le 1\ ; n = 1,2,\dots,\mathrm{N}$$

$$b_{jm}, a_{im} \ge 0\ ; i = 1,2,\dots,\mathrm{I}\ ;\ \ j = 1,2,\dots,\mathrm{J}$$

Where:
b_{jm} = weight of j^{th} output
y_{jm} = j^{th} output of the m^{th} DMU
a_{im} = weight of i^{th} input
x_{im} = i^{th} input of the m^{th} DMU
y_{jn} = j^{th} output of the n^{th} DMU
x_{in} = i^{th} input of the n^{th} DMU

The maximization problem can be converted into a linear problem by normalizing either the numerator or the denominator. Should the weighted sum of inputs be normalized, the problem adopts the form of an Output Maximization DEA program. Conversely, if the weighted sum of outputs equals unity, it constitutes an Input Minimization DEA program.

Table 2. Primal Models for the Input and Output Orientations of CRS and VRS

	Input Orientation	Output Orientation
CRS	$\text{Max} \sum_{j=1}^{J} b_{jm} y_{jm} \quad s.t.$ $\sum_{i=1}^{I} a_{im} x_{im} = 1$ $\sum_{j=1}^{J} b_{jm} y_{jn} - \sum_{i=1}^{I} a_{im} x_{in} \le 0\ ;$ $n = 1,2,\dots,N$ $b_{jm}, a_{im} \ge 0\ ;\ i = 1,2,\dots,I\ ;\ j = 1,2,\dots,J$	$\text{Min} \sum_{i=1}^{I} a'_{im} x_{im} \quad s.t.$ $\sum_{j=1}^{J} b'_{jm} y_{jm} = 1$ $\sum_{j=1}^{J} b'_{jm} y_{jn} - \sum_{i=1}^{I} a'_{im} x_{in} \le 0\ ;$ $n = 1,2,\dots,N$ $b'_{jm}, a'_{im} \ge 0\ ;\ i = 1,2,\dots,I\ ;\ j = 1,2,\dots,J$
VRS	$\text{Max} \sum_{j=1}^{J} b_{jm} y_{jm} - \rho_m \quad s.t.$ $\sum_{i=1}^{I} a_{im} x_{im} = 1$ $\sum_{j=1}^{J} b_{jm} y_{jn} - \sum_{i=1}^{I} a_{im} x_{in} - \rho_m \le 0\ ;$ $n = 1,2,\dots,N$ $b_{jm}, a_{im} \ge 0\ ;\ i = 1,2,\dots,I\ ;\ \ j = 1,2,\dots,J$	$\text{Min} \sum_{i=1}^{I} a'_{im} x_{im} - \rho_m \quad s.t.$ $\sum_{j=1}^{J} b'_{jm} y_{jm} = 1$ $\sum_{j=1}^{J} b'_{jm} y_{jn} - \sum_{i=1}^{I} a'_{im} x_{in} - \rho_m \le 0\ ;$ $n = 1,2,,N$ $b'_{jm}, a'_{im} \ge 0\ ;\ i = 1,2,\dots,I\ ;\ \ j = 1,2,\dots,J$

There are two classical DEA models, CCR [1] and BCC [30]. Each of the models can have two distinct orientations: output maximization or input minimization. The CCR model is commonly referred to as the Constant Return to Scale (CRS) model, while the BCC model is known as the Variable Return to Scale (VRS) model.

The primary distinction between the CRS and VRS models lies in their treatment of production scale. CRS assumes a constant scale of production across all DMUs. Consequently, if a DMU alters its inputs and outputs proportionally, its CRS efficiency score remains unchanged. In contrast, the VRS model considers variations in production scale among DMUs. When a DMU adjusts its inputs and outputs proportionally, the VRS efficiency score can change to reflect the shift in scale. This difference highlights how VRS acknowledges that scale adjustments have the potential to affect efficiency, distinguishing it from the CRS model. Table 2 displays the primal models used to assess efficiency for both models and for both orientations, with m representing the index of the DMU in question.

Machine Learning Models

ML is a technique that improves system performance by learning from experience via computational methods. In computer systems, experience exists in the form of data, and the main task of ML is to develop learning algorithms that build models from data. By feeding the learning algorithm with experience data, one obtains a model that can make predictions (e.g., the watermelon is ripe) on new observations (e.g., an uncut watermelon). If one considers computer science as the subject of algorithms, then ML is the subject of learning algorithms [31].

This paper employed eight different regression ML models. These models are Adaptive Boosting (AdaBoost), Categorical Boosting (CatBoost), Gradient Boosting Machine (Gradient Boosting), Histogram-based Gradient Boosting (Hist Gradient Boosting), Light Gradient Boosting Machine (LightGBM), Random Forest, Support Vector Regression (SVR), and Extreme Gradient Boosting (XGBoost). Regression, a predictive technique for continuous numeric outcomes based on input data, was employed in this context, aligning well with the analysis of DEA efficiency values, which are both continuous and numeric. The remaining seven models apply ensemble learning techniques. Ensemble learning is a strategy that combines the predictions of several individual models to optimize overall accuracy through collaborative decision-making. Among these, Random Forest stands out as the sole bagging model. Bagging is a technique that aims to reduce variance and enhance model stability by training multiple copies of the same model on different subsets of the training data. The other six models fall into the category of Boosting models, which operate differently. Boosting enhances a model's accuracy by assigning greater weight to previously misclassified data points. This approach involves sequentially training multiple weak learners, with each new model rectifying the errors made by its predecessors.

Evaluation Metrics

To evaluate the performance of the ML models, this paper employed three widely used metrics: Coefficient of Determination (R^2), Mean Absolute Error (MAE), and Root Mean Squared Error (RMSE). R^2 assesses how effectively the regression model fits the data by indicating the proportion of variance in the target variable explained by the independent variables. The formula for R^2 is:

$$R^2 = 1 - \frac{SSR}{SST} \tag{2}$$

Where: $SSR = \sum_i (y_i - \hat{y}_i)^2$ is the Sum of Squared Residuals, $SST = \sum_i (y_i - \bar{y})^2$ is the Total Sum of Squares, $\hat{y}_i$ is the i^{th} predicted value, and $\bar{y}$ is the mean of the observed data.

MAE evaluates the average absolute difference between the predicted values $(\hat{y})$ and the actual values (y) in a regression model. The formula is as follows:

$$MAE = \frac{1}{n}\sum_i |y_i - \hat{y}_i| \tag{3}$$

Meanwhile, RMSE assigns more weight to larger errors by taking the square root of the average of squared differences between predicted and actual values. The RMSE formula is:

$$RMSE = \sqrt{\frac{1}{n}\sum_i (y_i - \hat{y}_i)^2} \tag{4}$$

These metrics serve as essential tools for assessing the accuracy and performance of ML regression models from various angles. R^2 indicates the goodness of fit, MAE measures the average accuracy of the model's predictions, and RMSE serves a similar purpose while giving greater weight to larger errors.

Software and Tools

This study used pyDEA to perform the DEA [32]. PyDEA is a Python software package with a graphical user interface. Through this interface, users can input data, designate features as inputs or outputs, select between Constant and Variable Return to Scale, opt for input or output orientation, execute the analysis, and ultimately visualize or export the results of the DEA.

Exploring the Feasibility of ML as an Alternative to DEA for Continuous Efficiency Estimation

This study executed an experimental study to assess the viability of utilizing ML to replace the need for repeated DEA runs. This experiment aimed to determine whether ML can effectively supplant the necessity of further DEA iterations. Should the outcome of this experiment prove promising, it would imply that the requirement for subsequent DEA analyses upon introducing new DMUs to the original dataset could be eliminated. Training an ML model on the initial DEA-generated efficiency scores gave the model the capability to predict the efficiencies of newly introduced DMUs.

The process of this study involved two distinct datasets: one for training and another for prediction. Training Phase: A training dataset was compiled, comprising 1000 rows of data chosen randomly from the original dataset. The target value for the training dataset was the efficiency calculated using DEA applied to the 1000 rows. An ML model was trained on the training dataset.

Prediction Phase: An additional dataset was prepared, encompassing more than 1000 rows. Importantly, this dataset included all the 1000 rows from the training dataset. However, the efficiency values for these overlapping rows were derived from a different DEA analysis conducted on the entire new dataset. The trained ML model was deployed to predict the efficiency values for all DMUs in the expanded dataset. The performance assessment was done by calculating the differences between the true and the predicted efficiency values—specifically, the values derived from the new DEA and the predictions of the model that was trained on the original dataset.

Results and Discussion

DEA

DEA was executed using an output-oriented approach and two distinct models, namely CRS and VRS. In total, 134 DMUs were found fully efficient by the CRS model, and Public Storage was the only company that maintained its full efficiency across the four quarters. On the other hand, 241 DMUs were deemed fully efficient by the VRS model, and eight companies exhibited full efficiency across the four quarters. These companies are Public Storage, Microsoft, Amazon, Facebook, Expeditors, ExxonMobil, AT&T, and Moderna. In this paper, a DMU refers to a row of the dataset, which is a combination of a company and a quarter. Further statistics about the DEA results are shown in Table 3.

Table 3. Summary of DEA Results

Model	Average efficiency	Minimum efficiency	Fully efficient DMUs	Firms fully efficient for 2 quarters or more	Firms fully efficient across the 4 quarters
CRS	59.25%	23.81%	134	31	1
VRS	70.75%	25.55%	241	56	8

Evaluating the Performance of ML Models on Known Datasets

The setup in this sub-section is referred to as a “known dataset,” wherein training and testing were performed on subsets of the same dataset. In contrast, in the following sub-section, the paper introduces the concept of “expanded/new datasets,” where training was conducted on one dataset, and testing was performed on a larger one containing newly incorporated DMUs and the efficiency scores were obtained through a distinct DEA.

The efficiency values obtained from the CRS and VRS models were used as the target features in two separate datasets. These latter datasets comprised identical features and values, differing solely in the Efficiency feature values determined by the respective model.

The eight ML models were trained on the training set of both CRS and VRS datasets. The train and the test sets had a ratio of 70% for training and 30% for testing. Subsequently, the models underwent a hyperparameter tuning process, a systematic approach in ML which aims to identify the optimal combination of hyperparameters that results in the best performance of the model. In this paper, the tuning process was executed via GridSearchCV, integrating a 5-fold cross-validation and adopting Negative Mean Squared Error as the chosen evaluation strategy. The hyperparameters that were optimized included the learning rate, the number of estimators, the maximum depth, and the maximum features. However, variations might arise based on specific scenarios, where certain models could involve the tuning of distinct parameters such as epsilon and regularization, particularly in the context of the SVR model.

The tuned models were compared by evaluating their performance on the test set using three metrics: R^2, RMSE, and MAE. Initially, the ranking of models, ordered from the highest R^2 to the lowest on the CRS dataset, was as follows: CatBoost, LightGBM, Hist Gradient Boosting, XGBoost, Gradient Boosting, AdaBoost, Random Forest and lastly SVR. Table 4 displays the corresponding R^2 scores for the top three models were 92.84%, 91.99%, and 91.41%. The same model sequence persisted when arranged from the lowest RMSE to the highest. The only change that occurred when arranged using MAE was between Gradient Boosting and AdaBoost, where they switched positions. For the VRS dataset, the models maintained the same order when evaluated using each of the three metrics. They followed the previous sequence with only XGBoost and Gradient Boosting swapping positions. The R^2 scores for the best three models, namely CatBoost, LightGBM, and Hist Gradient Boosting, were 94.42%, 94.37%, and 94.02%, respectively. These higher coefficient of determination scores indicate a better fit for the model.

Within both datasets, every model exhibited an R^2 value exceeding 85%, with the only exceptions being SVR whose R^2 scores was 41.74% on the CRS dataset and 56.08% on the VRS dataset. This might be because SVR is based on single kernel which may contribute to poor performance [33]. In addition, the nature of ensemble methods also makes SVR less performed as compared to boosting methods. This finding can also be supported by the work of Zhu *et al*. [3] whereby they also found that SVM-DEA performs less effectively in their analysis.

Moreover, regardless of the metric used, all the ML models considered performed better on the VRS dataset than on the CRS dataset. Table 4 presents the models' scores compare on both models using CRS and VRS. For better visualization, Figure 1 and Figure 2 provide a representation of the scores across the three metrics on both models based on results presented in Table 4. It can be seen that CatBoost, LightGBM, and Hist Gradient Boosting are the top performers, while highlighting SVR as the least effective model on both models.

Table 4. Models' Performance on the Test Set of the CRS and VRS Datasets

Model	CRS			VRS		
	R^2	RMSE	MAE	R^2	RMSE	MAE
CatBoost	92.84	5.60	3.26	94.42	4.50	2.76
LightGBM	91.99	5.92	3.70	94.37	4.52	2.86
Hist Gradient Boosting	91.41	6.13	3.77	94.02	4.66	3.01
Gradient Boosting	88.27	7.17	4.35	90.64	5.83	3.69
XGBoost	89.37	6.82	4.23	90.24	5.95	3.79
AdaBoost	87.37	7.44	4.31	88.86	6.36	3.91
Random Forest	85.77	7.90	4.90	87.01	6.86	4.80
SVR	41.74	15.97	11.96	56.08	12.62	9.25

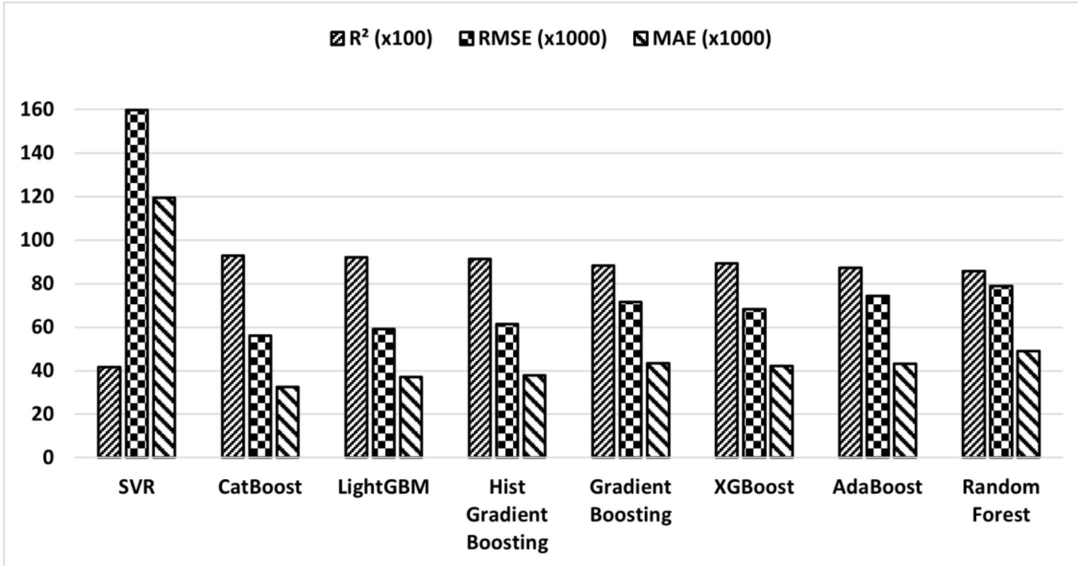

Figure 1. Comparison of models' performance on the CRS dataset

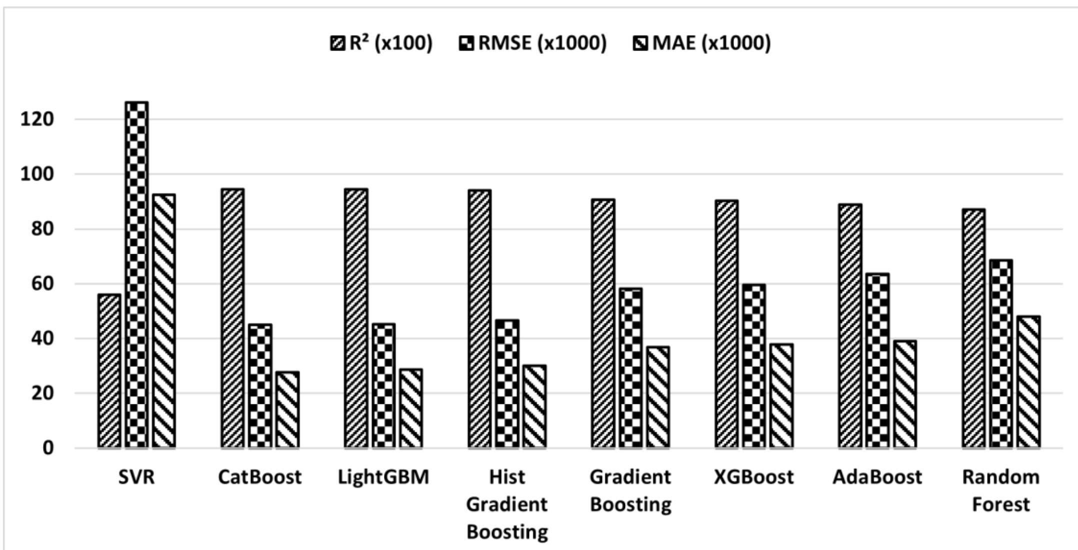

Figure 2. Comparison of models' performance on the VRS dataset

Since CatBoost emerged as the top-performing model across both datasets, we selected it to illustrate the distribution of residuals. Residuals depict the errors, or the differences between the actual efficiency values and the predicted counterparts. Predictions were generated for both training and testing sets. Out of 2007 predictions, only 8 had errors exceeding 0.2 on the CRS dataset and only 4 on the VRS dataset. Notably, there were 39 errors surpassing 0.1 on the CRS dataset and 24 on the VRS dataset. In other terms, 98.06% of CRS efficiency predictions and 98.80% of VRS efficiency predictions deviated by less than 0.1 from the actual values. When we tightened the threshold to 0.02, these percentages became 85.80% for CRS and 86.40% for VRS. Figures 3 and 4 depict the errors' cumulative distribution. The *x*-axis represents error values, and the *y*-axis represents the cumulative count of predictions with absolute errors greater than or equal to each respective error value (*x*).

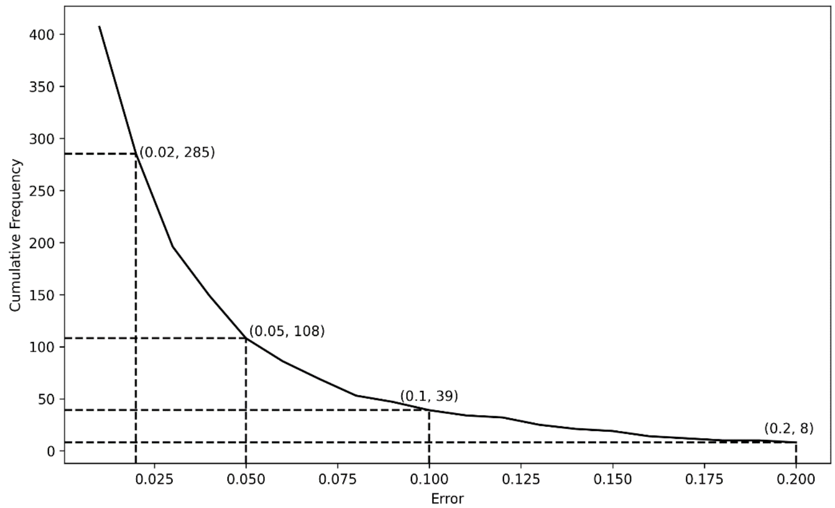

Figure 3. Cumulative distribution of CatBoost absolute errors on the CRS dataset

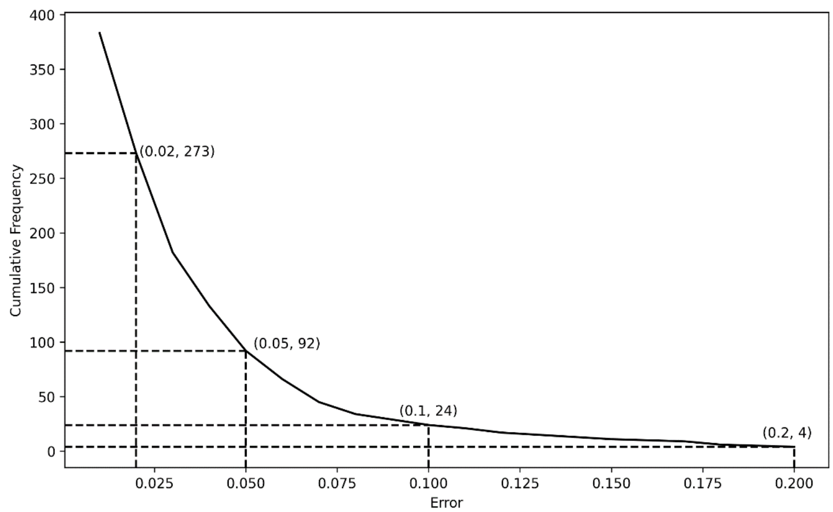

Figure 4. Cumulative distribution of CatBoost absolute errors on the VRS dataset

Ultimately, CatBoost, LightGBM, and Hist Gradient Boosting stood out as the top-performing ML models for predicting efficiency values on a familiar dataset. Notably, the top seven performers were all ensemble learning models, which combine the predictions of multiple individual learners to enhance the overall predictive accuracy. It is intriguing to note that Random Forest, a Bagging model, ranked seventh, underscoring the effectiveness of Gradient Boosting models in the context of DEA efficiency prediction. This observation aligns with prior research findings. For example, in Zhang *et al*.'s [14] study, XGBoost, CatBoost, and LightGBM, along with the BPNN model, were identified as the top performers.

Results of Exploring the Feasibility of ML as an Alternative to DEA for Continuous Efficiency Estimation

The methods employed in this feasibility study were elaborated on in the Methods section. Among the models considered, the top three performers, specifically CatBoost, LightGBM, and Hist Gradient Boosting, were chosen for evaluation. Each of these models underwent training using a dataset comprising 1000 rows and was subsequently applied to predict the efficiency values across five expanded datasets, with sizes ranging from 1200 to 2000 in increments of 200 rows.

As the size of the expanded dataset increased, the R^2 scores of the trained models decreased. However, despite this decrease, all scores of the three models across the five expanded CRS and VRS datasets remained above the notable threshold of 89.82%. Once again, CatBoost demonstrated its superiority as the top-performing model, followed by LightGBM in the second position in most cases. Figures 5 and 6 provide a visual representation of the score evolution as the size of the expanded dataset increases for both the CRS and VRS scenarios. It is worth reiterating that the scores are exceptionally high for the 1000-row dataset, primarily because the models were trained on this dataset.

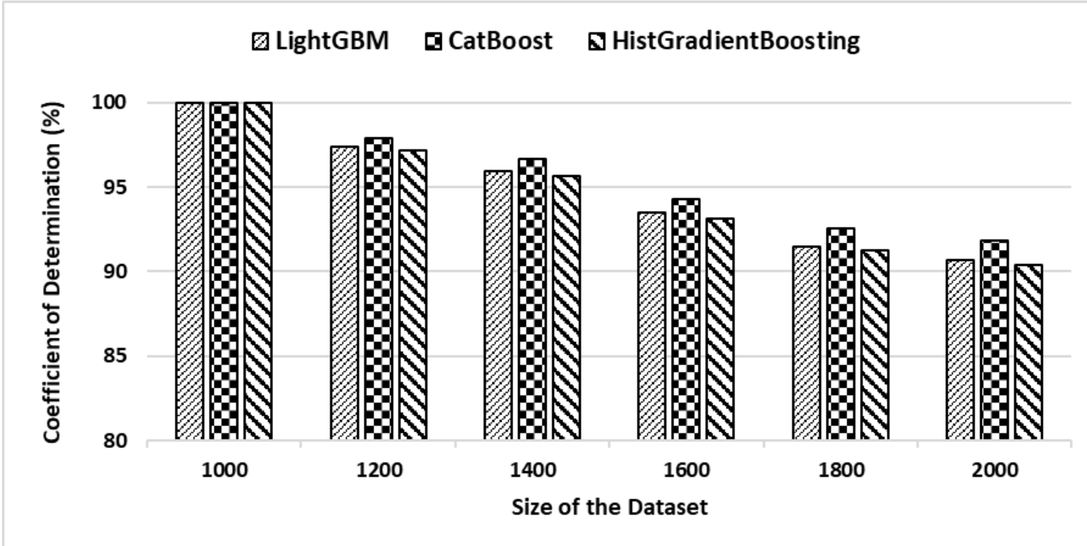

Figure 5. Performance of the three models on expanded CRS datasets

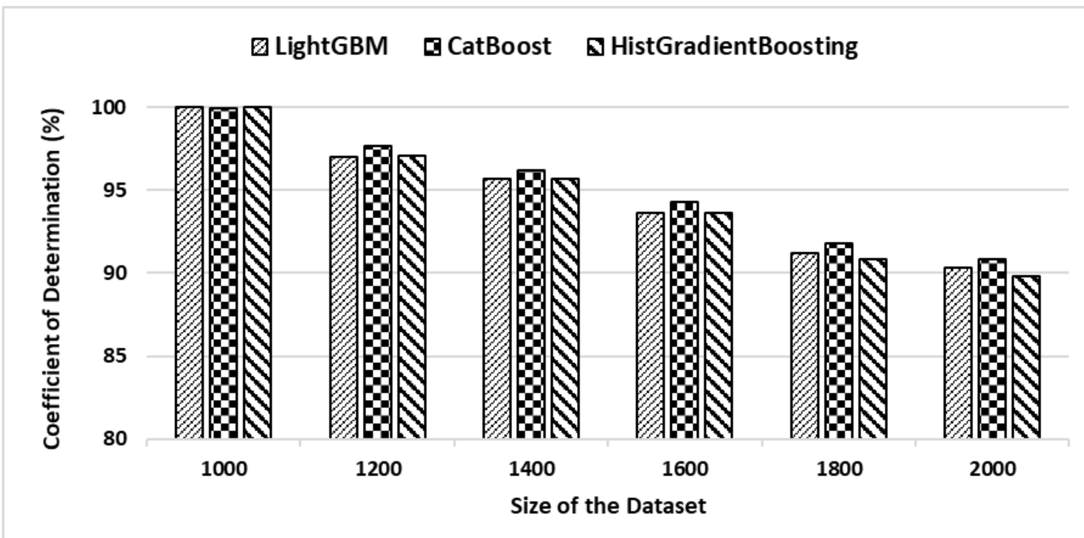

Figure 6. Performance of the three models on expanded VRS datasets

Subsequently, for the sake of simplicity, this paper narrowed its focus to CatBoost, the leading model, to assess its performance across the three metrics. When evaluated using R^2, the model demonstrated superior performance on expanded datasets in the CRS scenario compared to the VRS scenario. Conversely, it excelled more on expanded VRS datasets when assessed using MAE and RMSE. This difference in performance can be attributed to variations in the data distributions of the Efficiency feature in each dataset. Consequently, this divergence in performance across different metrics suggests two key implications: firstly, the model is a strong fit for the CRS dataset since the R^2 value for CRS model higher than VRS model, and secondly, on the VRS dataset, the model excels in making precise predictions since the RMSE and MAE scores lower, even if it does not explain variance as effectively. Table 5 shows the results of the CatBoost model performance on expended datasets. Notably, it shows that the model exhibits superior R^2 values in the CRS scenario, while excelling in making precise predictions with lower RMSE and MAE scores in the VRS scenario.

Table 5. CatBoost Model Performance on Expanded Datasets (All Values Multiplied by 100)

Size of the dataset	CRS			VRS		
	R^2	RMSE	MAE	R^2	RMSE	MAE
1000	99.99	0.17	0.13	99.99	0.18	0.14
1200	97.85	3.05	1.53	97.67	2.94	1.24
1400	96.65	3.77	2.08	96.23	3.71	1.85
1600	94.31	4.92	3.18	94.33	4.56	2.59
1800	92.56	5.59	3.69	91.79	5.46	3.41
2000	91.86	5.84	3.92	90.86	5.77	3.67

Interestingly, in the context of the paper's feasibility study exploring the viability of ML as an alternative to DEA for continuous efficiency estimation, the performance of ML on expanded datasets (with newly introduced DMUs) was comparable to its performance on known datasets. Specifically, when considering the CatBoost model's performance, we noted that the MAE values for datasets of up to 1600 rows consistently outperformed the MAE values calculated from a CatBoost model trained and tested on a known dataset. Remarkably, this trend extended to larger datasets comprising 1800 and 2000 rows, where the model's performance, while slightly inferior, remained remarkably competitive. These findings emphasize the adaptability and robustness of ML for continuous efficiency estimation, even when applied to significantly expanded datasets.

Conclusions

This paper conducted an empirical study to validate the adaptability of ML in enhancing DEA and overcoming its limitations. To this end, it assessed and compared the performance of eight ML models with respect to the CRS and VRS variants of the DEA base model using a dataset of S&P500 companies' financial data.

The assessment was initially conducted on familiar datasets where efficiency values originated from the same DEA run. The results were highly satisfactory and affirmed the adaptability of ML for estimating the DEA findings. Notably, boosting models within the Ensemble Learning category consistently outperformed other models, highlighting their effectiveness in the context of DEA efficiency prediction. The second assessment extended to five enlarged datasets, each incorporating new DMUs and efficiency values obtained from re-conducting DEA. The results exceeded our expectations, demonstrating ML's consistent performance even with significantly expanded datasets. Additionally, we observed that ML predicted VRS efficiencies more accurately than CRS efficiencies on known and expanded datasets, as indicated by MAE and RMSE on Table 5.

These findings align with previous research on the validity of ML as a substitute for DEA re-runs when evaluating the efficiency of new DMUs. Moreover, this study contributes to a more comprehensive validation of this assumption. Nonetheless, this study has limitations, as the analysis was exclusively based on the dataset under consideration. We recommend that future research explores similar investigations using statistical approaches, different ML and Deep Learning models, or further validates these findings on diverse datasets.

Conflicts of Interest

The author(s) declare(s) that there is no conflict of interest regarding the publication of this paper.

Acknowledgment

We are grateful for the financial support provided by the Universiti Teknologi MARA (UiTM) under Pembiayaan Yuran Penerbitan Artikel (PYPA) scheme for providing financial support that enabled us to carry out this study effectively.

References

[1] Charnes, A., Cooper, W. W., & Rhodes, E. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research*, *2*(6), 429-444.

[2] Yu, X., & Lou, W. (2023). An exploration of prediction performance based on projection pursuit regression in conjunction with data envelopment analysis: a comparison with artificial neural networks and support vector regression. *Mathematics*, *11*(23), 4775-4803.

[3] Zhu, N., Zhu, C., & Emrouznejad, A. (2021). A combined machine learning algorithms and DEA method for measuring and predicting the efficiency of Chinese manufacturing listed companies. *Journal of Management Science and Engineering*, *6*(4), 435-448.

[4] Ramli, N. A., Khairi, S. S. M., & Razlan, N. A. (2018). Performance measurement of islamic and conventional banking in Malaysia using two-stage analysis of DEA model. *International Journal of Academic Research in Business and Social Sciences*, *8*(4), 1185-1197.

[5] Yang, Y., & Guo, L. (2021). Research on Diagnostic Test and Treatment for Higher Education System. *2021 IEEE 3rd International Conference on Frontiers Technology of Information and Computer (ICFTIC),* 291-300. IEEE.

[6] Zuluaga, R., Camelo-Guarín, A., & De La Hoz, E. (2023). Assessing the relative impact of Colombian higher education institutions using fuzzy data envelopment analysis (fuzzy-DEA) in state evaluations. *Journal on Efficiency and Responsibility in Education and Science*, *16*(4), 299-312.

[7] Katharakis, G., Katharaki, M., & Katostaras, T. (2014). An empirical study of comparing DEA and SFA methods to measure hospital units' efficiency. *International Journal of Operational Research*, *21*(3), 341-364.

[8] Antunes, J., Hadi-Vencheh, A., Jamshidi, A., Tan, Y., & Wanke, P. (2023). TEA-IS: A hybrid DEA-TOPSIS approach for assessing performance and synergy in Chinese health care. *Decision Support Systems*, 113916-113929.

[9] Mirmozaffari, M., & Kamal, N. (2023). The application of data envelopment analysis to emergency departments and management of emergency conditions: A narrative review. *Healthcare, 11*(18), 2541-2568.

[10] Emrouznejad, A., & Yang, G. L. (2018). A survey and analysis of the first 40 years of scholarly literature in DEA: 1978–2016. *Socio-economic Planning Sciences*, *61*, 4-8.

[11] Yang, G., Ren, X., Khoveyni, M., & Eslami, R. (2020). Directional congestion in the framework of data envelopment analysis. *Journal of Management Science and Engineering*, *5*(1), 57-75.

[12] Song, Y. Y., Yang, G. L., Yang, J. B., Khoveyni, M., & Xu, D. L. (2018). Using two-layer minimax optimization and DEA to determine attribute weights. *Journal of Management Science and Engineering*, *3*(2), 76-100.

[13] Anouze, A. L. M., & Bou-Hamad, I. (2019). Data envelopment analysis and data mining to efficiency estimation and evaluation. *International Journal of Islamic and Middle Eastern Finance and Management*, *12*(2), 169-190.

[14] Zhang, Z., Xiao, Y., & Niu, H. (2022). DEA and machine learning for performance prediction. *Mathematics*, *10*(10), 1776-1798.

[15] Bowlin, W. F., Charnes, A., Cooper, W. W., & Sherman, H. D. (1984). Data envelopment analysis and regression approaches to efficiency estimation and evaluation. *Ann. Oper. Res.*, *2*(1), 113-138.

[16] Athanassopoulos, A. D., & Curram, S. P. (1996). A comparison of data envelopment analysis and artificial neural networks as tools for assessing the efficiency of decision making units. *Journal of the Operational Research Society*, *47*, 1000-1016.

[17] Salehi, V., Veitch, B., & Musharraf, M. (2020). Measuring and improving adaptive capacity in resilient systems by means of an integrated DEA-Machine learning approach. *Applied Ergonomics*, *82*, 102975-102984.

[18] Jomthanachai, S., Wong, W. P., & Lim, C. P. (2021). An application of data envelopment analysis and machine learning approach to risk management. *IEEE Access*, *9*, 85978-85994.

[19] Nishtha, Puri, J., & Setia, G. (2023). Performance prediction of DMUs using integrated DEA-SVR approach with imprecise data: application on Indian banks. *Soft Computing*, *27*(9), 5325-5355.

[20] Guerrero, N. M., Aparicio, J., & Valero-Carreras, D. (2022). Combining data envelopment analysis and machine learning. *Mathematics*, *10*(6), 909-930.

[21] Zhong, K., Wang, Y., Pei, J., Tang, S., & Han, Z. (2021). Super efficiency SBM-DEA and neural network for performance evaluation. *Information Processing & Management*, *58*(6), 102728-102741.

[22] Hong, H. K., Ha, S. H., Shin, C. K., Park, S. C., & Kim, S. H. (1999). Evaluating the efficiency of system integration projects using data envelopment analysis (DEA) and machine learning. *Expert Systems with Applications*, *16*(3), 283-296.

[23] Visbal-Cadavid, D., Mendoza, A. M., & Hoyos, I. Q. (2019). Prediction of efficiency in Colombian higher education institutions with data envelopment analysis and neural networks. *Pesquisa Operacional*, *39*, 261-275.

[24] Babaei Keshteli, H., & Rostamy-Malkhalifeh, M. (2022). A combined machine learning algorithms and Interval DEA method for measuring predicting the efficiency. *International Journal of Data Envelopment Analysis*, *10*(3), 57-64.

[25] Appiahene, P., Missah, Y. M., & Najim, U. (2020). Predicting bank operational efficiency using machine learning algorithm: comparative study of decision tree, random forest, and neural networks. *Advances In Fuzzy Systems*, *2020*, 1-12.

[26] Wei, J., Ye, T., & Zhang, Z. (2021). A machine learning approach to evaluate the performance of rural bank. *Complexity*, *2021*, 1-10.

[27] Thaker, K., Charles, V., Pant, A., & Gherman, T. (2022). A DEA and random forest regression approach to studying bank efficiency and corporate governance. *Journal of the Operational Research Society*, *73*(6), 1258-1277.

[28] Pierre-Louis Danieau. (2021). *Financial Data S&P500 companies*. Kaggle. Retrieved October 15, 2023 from https://www.kaggle.com/datasets/pierrelouisdanieau/financial-data-sp500-companies.

[29] Ramanathan, R. (2003). *An introduction to data envelopment analysis: A tool for performance measurement*. Sage.
[30] Banker, R. D., Charnes, A., & Cooper, W. W. (1984). Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Management Science*, *30*(9), 1078-1092.
[31] Zhou, Z. H. (2021). *Machine Learning.* Springer Nature, Berlin.
[32] Andrea Raith, Fariza Fauzi and Olga Perederieieva. (2016). *pyDEA Documentation*. Retrieved October 31, 2023 from https://araith.github.io/pyDEA/.
[33] Che, J., & Wang, J. (2014). Short-term load forecasting using a kernel-based support vector regression combination model. *Applied Energy*, *132*, 602-609.