Efficiency analysis trees as a tool to analyze the quality of university education

Rohemi Zuluaga-Ortiz[1], Alicia Camelo-Guarín[2], Enrique Delahoz-Domínguez[3]
[1]Faculty of Engineering, Universidad del Sinú, Cartagena, Colombia
[2]Faculty of Military Sciences, Escuela Militar de Cadetes General José María Córdova, Bogotá, Colombia
[3]Faculty of Engineering, Universidad Tecnológica de Bolívar, Cartagena, Colombia

Article Info

Article history:

Received Jul 21, 2022
Revised Oct 6, 2022
Accepted Nov 21, 2022

Keywords:

Data enveloping analysis
Decision trees
Education management
Efficiency
Learning analytics

ABSTRACT

This research aims to design a quality management tool in education. The methodology comprises two stages: first, the construction of the decision tree model, and second, the efficiency evaluation. For the validation and development of this research, the data modelled corresponds to the standardized exams for higher education in Colombia of ninety industrial engineering degrees. Among the results, the citizenship skills (CC_PRO) generate the most significant contribution to the model. On the other hand, the written communication competence (CE_PRO) generates a minor contribution to the model. In addition, the most relevant result of the research is the design and validation of a tool to estimate educational efficiency using the efficiency analysis tree (EAT) and data enveloping analysis (DEA) models. The proposed tool allows the generation of specific targets to increase the level of efficiency of universities through the nodes of the decision tree, which contributes to the spectrum of knowledge on models for educational management. In conclusion, this research presents a tool for the management of educational processes through the analysis of efficiency using EAT, estimating the efficiency of universities and setting the foundations for forecasting future efficiency scenarios.

Corresponding Author:

Enrique Delahoz-Dominguez
Faculty of Engineering, Universidad Tecnológica de Bolivar
Cartagena, Colombia
Email: edelahoz@utb.edu.co

1. INTRODUCTION

The knowledge acquired through education quickly becomes a powerful tool that provides benefits, either as a personal competitive advantage or for the development of organizations [1]. However, a problem currently arising for organizations can be identified as an opportunity or a threat. As is well known, knowledge evolves, and it is necessary to be on the frontier of knowledge, so if organizations are not aware of the new waves of knowledge, this will become a threat; in contrast, if the organizations are aware of the new research, this will be an opportunity [2].

For their part, it is the duty of higher education institutions (HEI) to provide organizations with professionals with training framed in the frontier of knowledge; therefore, it is of great relevance to generate tools that support the management and evaluation of educational processes. In the case of Colombia, the Colombian Institute for the Evaluation of Education Instituto Colombiano para la evaluación de la educación (ICFES) has standardized tests to evaluate and monitor students' academic performance in secondary education, Saber 11, and university education, Saber PRO. Thus, ICFES uses the added value approach, considering that a student's capacities at the end of a university program are not only a consequence of their passage through the HEI but also due to their previous knowledge foundations. Since 2012, the ICFES has developed a method called relative contribution (RC), evolving the concept of added value to measure the quality of education in Colombia.

Usually, methods for estimating technical efficiency in production systems use an efficiency frontier scheme to compare productive units' performance and the deviations associated with inefficient units. However, data envelopment analysis (DEA) models suffer from overfitting problems when estimating efficiency parameters, resulting in conservative models that underestimate the technical efficiency of decision-making units (DMUs). Therefore, becoming models that describe very well the current situation but do not allow a generalized adjustment for the whole system. Consequently, with the rise of machine learning techniques, econometric models are increasingly required to predict future scenarios so that the results of an efficiency model serve as support for objective decision-making and not only for the description of a past situation. Therefore, the approach proposed in this research for analyzing the academic performance of universities relies on using the efficiency analysis trees (EAT) approach to demonstrate the advantages of creating an efficiency frontier through the decision trees technique and will also allow determining the efficiency for out-of-sample DMUs.

Consequently, measuring the quality and efficiency of education today is a challenge for many researchers [3]–[5] due to the number of variables associated with the economic, psychosocial, sociodemographic, environmental, institutional, ethical, spiritual and cultural contexts of each individual [6]–[9]. Therefore, it is essential to manage educational processes in the most effective way possible through reliable techniques, tools and methodologies, which have been studied in Colombia [10], [11]. Unlike models previously developed in this field, the present research proposes to measure the efficiency and productivity of universities in Colombia in engineering careers from the EAT. This computational library was created by [12], and to date, it has no real implementations in the literature related to efficiency analysis issues. Consequently, this research seeks to validate the new EAT model and compare the findings with the classic efficiency model DEA.

DEA is a non-parametric method to measure the efficiency and productivity of decision-making units. For instance, [13] considers that the purpose of the DEA technique is to analyze the level of efficiency of the study units (also known as DMUs, decision-making units). The tool's core focuses on analyzing various inputs to generate desired outputs, as long as they are under equal conditions (ensuring that the assumption of homogeneity is met). In summary, the DEA technique estimates the observations' efficiency levels, taking into account the deviations in the production frontier (isoquant curve formed by the inputs and outputs of the system). The DEA model is non-parametric and, as has been mentioned, by means of the production frontier analysis, it estimates the efficiency of the DMUs [14], [15]; then, using the DMUs, the efficient frontier is constructed taking into account the estimated efficiency levels. On the other hand, several authors claim that the DEA tool is suitable for estimating the performance of DMUs in the public and private sectors [16].

Therefore, statistical inference is possible based on current point estimates resulting from the DEA. However, this model suffers from an overfitting problem since it underestimates the technical inefficiency of the observations, generating estimated frontiers always located below the theoretical frontiers (underlying) [17]. Therefore, DEA can correctly describe the situation from the point of view of efficiency evaluation, but it cannot provide adequate generalization. Thus, DEA determines the efficiency scores but cannot give details of the factors related to inefficiency; therefore, this research seeks to compare the analysis carried out by the DEA against a more recent and little-studied model, the EAT.

Decision tree (DT) models belong to the family of supervised machine learning models and their structure is similar to that of a tree, in this model the leaves correspond to the classification of the output and the branches are the ramifications of the input variable that defines the classification or regression response [18], [19]. In Figure 1, a decision tree model is illustrated that starts with predictor X; if predictor X has the characteristic t, it goes to predictor Y, and if X has the characteristic f, it goes to W. In the same way, each of the new predictors is branched until finding the response of the model that will be either C1 or C2.

The EAT is a new technique proposed by and based on the adaptation of the classification and regression trees (CART) proposed by [20] for the estimation of production frontiers. This new technique allows to calculate the production frontier taking into account the common assumptions for efficiency analysis, using an approach that does not require a specific distribution over the behavior of the data and results in a step function as a predictor.

The new specifications of the EAT model are related to the free disposal hull (FDH) technique. However, EAT makes use of cross-validation to avoid overfitting that may occur at the time of model construction [20]. Additionally, the construction of the EAT model is performed taking into account the mean square error and the use of stopping rules taking into account the number of individuals within the node, thus avoiding the generation of empty nodes, otherwise a response without inputs is obtained.

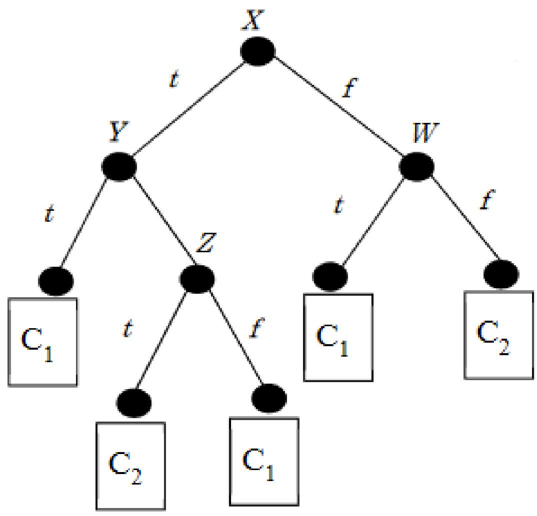

Figure 1. Graphic representation of a decision tree note. Taken from [21]

Unlike CART, EAT by estimating maximum trends instead of averages, guarantees the FDH model criterion and thus succeeds in calculating the production frontier. In EAT, the MSE minimization is the criteria for selecting the predictors in each node, generating binary partitions recursively (the training sample) until a significant partitioning is no longer possible or by a stopping rule. However, similar to the graphical representation of decision tree models, the visual representation of the model is a tree starting at a root node, branching to intermediate nodes and ending at its leaves.

The new EAT technique divides the inputs of the model into two binary responses, it should be noted that each new response is constant. In this sense, the inputs' evolution behaves like a step function. Then, FDH and EAT generate the efficiency estimation similarly by using a production frontier. However, the EAT model avoids the overfitting presented in the FDH model by using cross-validation and overfitting [20]. Thus, the fusion between the assumption of the FDH model (free disposal) and the construction performed by the EAT technique contributes to the data analysis by widening the spectrum of knowledge for efficiency analysis.

On the other hand, the research of [12] through the mean square error, bias and absolute bias, affirms that the EAT model performs better than the FDH model. For the mean squared error of the EAT model, they present performances that outperformed the FDH model between 13% and 70% in the simulations. Additionally, the authors show that as the size of the individuals increased, the mean squared error measure decreased. Also, an interesting advantage of the EAT model is that it allows graphically representing the production frontier generated by the trees. Thus, this tool becomes a proposal for the visual analysis of efficiency. Finally, the evaluation of the input variables is generated on the basis of the predictive importance of the variables of interest.

2. METHOD

The methodology research seeks to estimate the efficiency of universities that offer industrial engineering programs. The research methodology comprises two parts as shown in Figure 2: constructing the decision tree model and evaluating the efficiency and, additionally, a comparison with the classical model of efficiency analysis. The data used in this research were taken from the Mendeley research repository by [22]. It should be noted that from this database, only the results of the industrial engineering program were selected as shown in Table 1. Initially, the information was reviewed and pre-processed to obtain helpful information for the system. Consequently, eliminating the categorical variables that did not add valuable information and homogenizing the useful numerical variables.

Additionally, the EAT methodology proposes training a prediction model (decision tree) before constructing the efficient frontier. In this order of ideas, for this research, it is established that the predictor variables are QR, CS, ENG, WC, and CR, while the output variables are FEP, MSST, and DPLS. However, it is important to highlight that the production frontier has the same configuration as the prediction model (decision tree). Finally, for the data analysis and the construction of the models, R software [23] was used and the EAT package [11] for constructing the efficiency model employing a decision tree.

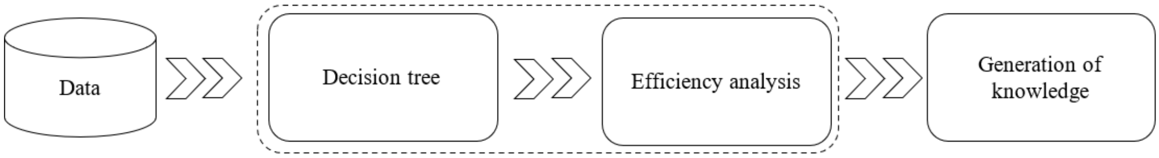

Figure 2. Research methodology

Table 1. Information from the research data

Variable	Full Name	Mean	Standard Deviation
QR	Quantitative reasoning	77.42	22.67
CS	Citizenship skills	62.20	27.67
ENG	English	59.19	28.99
WC	Writing communication	67.50	25.49
CR	Critical reading	53.70	30.00
FEP	Formulation of engineering projects	145.48	40.12
MSST	Mathematical and statistical scientific thinking	133.71	12.99
DPLS	Design of production and logistics systems	147.80	16.50

3. RESULTS

3.1. Stage 1: decision tree

The decision tree model uses parameters numStop (number of observations-DMUs-in a node to perform a partition) and fold (number of partitions of the dataset to make cross-validation during pruning). This model configuration comprises nine leaf nodes (leaves), providing a lesser complex model than a traditional efficiency model. Consequently, creating a training dataset (training: 70%) and, subsequently, a test dataset (test: 30%). Thus, Table 2 presents the cross-validation results, evidencing that the numStop and fold that minimize the root mean squares error (RMSE) are 6 and 5, respectively. Thus, the selected model must be branched, and within its configuration, it is found that it has 12 internal nodes, which are partitioned as shown in Table 3.

Table 2. Cross-validation of the decision tree model

numStop	Fold	RMSE	Leaves
6	5	38.79	9
3	5	39.05	13
3	6	39.05	13
4	4	39.06	11
5	4	39.06	10
4	5	39.06	11
5	5	39.06	10
4	6	39.06	11
3	4	39.10	12
6	4	39.10	8

Table 3. Levels at the frontiers of the leaf nodes

Leaf node	MSST	DPLS	FEP
1	174.12	192.94	187.24
2	154.80	160.33	170.50
3	120.25	132.88	134.88
4	119.25	147.67	142.00
5	123.59	147.67	142.00
6	128.00	147.67	142.00
7	128.00	147.67	142.29
8	128.00	147.67	145.11
9	130.39	150.02	149.34
10	128.00	147.67	147.00
11	128.00	147.70	145.10
12	128.00	147.70	145.10

Therefore, assuming that the decision tree model considers the predictors (independent variables) to predict the variables of interest (dependent variables), it is crucial to observe each academic entry's contribution to the academic skills output as shown in Figure 3. The input-level citizenship skills (CS_PRO) generate the most significant contribution to the model, while the written communication skill (WC_PRO) generate the most negligible contribution to the model. From an academic approach, the results evidence the importance of basic reading and citizenship skills to generate integral learning in engineering students.

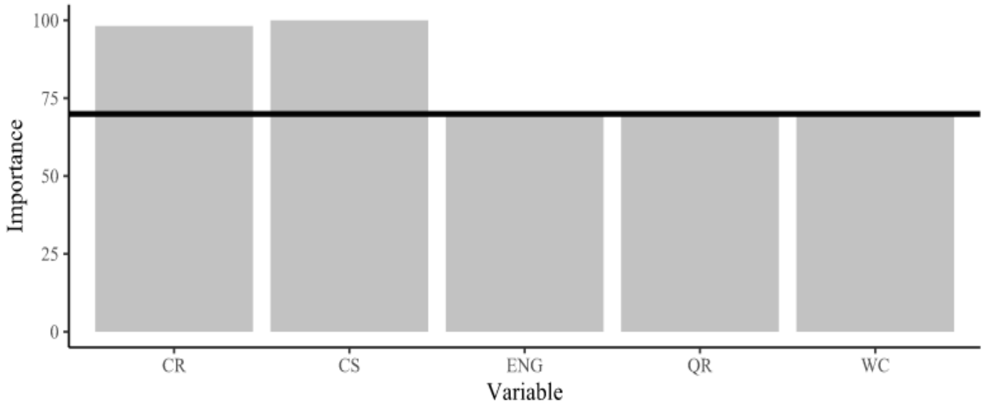

Figure 1. Importance of the model variables

3.2. Stage 2: efficiency analysis

Consequently, Table 4 compares the efficiency model's results using decision trees EAT and the classical efficiency model DEA. However, in general, it is observed that the results of the classical model have a higher efficiency level than the decision tree model for both the constant scale constant return to scale (CRS) and the variable scale variable return to scale (VRS).

Table 4. Results of the efficiency models

Metric	CRS		VRS		Scale performance	
	EAT	DEA	EAT	DEA	EAT	DEA
Efficient number	2 (2.17%)	8 (8.70%)	11 (11.95%)	82 (89.13%)	6 (6.52%)	8 (8.69%)
Mean	0.59	0.80	0.91	0.99	0.66	0.82
Deviation	0.11	0.11	0.08	0.05	0.14	0.11
Minimum	0.44	0.61	0.68	0.72	0.47	0.62
Quartile 2	0.56	0.79	0.91	1.00	0.64	0.80
Quartile 3	0.64	0.88	0.97	1.00	0.74	0.90

On the other hand, Figure 4 shows the concentration of the distribution of the efficiency data for the CEAT and DEA models, evidencing that for the CEAT model, the data distribution concentrates on the left side. In contrast, the data distribution is concentrated on the right side of the DEA model. Thus, indicating that the DEA model has a higher efficiency level than the CEAT model.

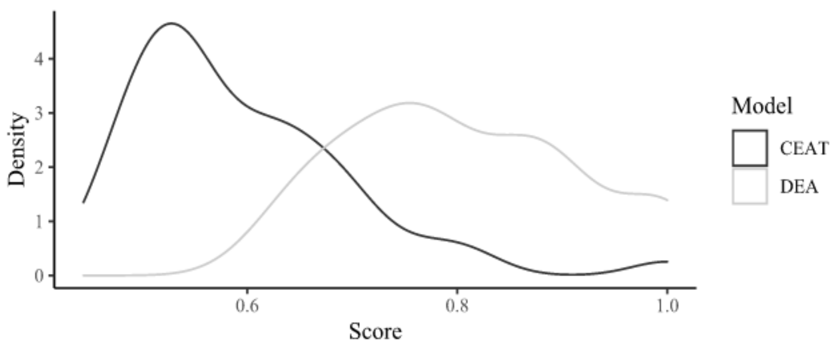

Figure 4. Density of the initial model efficiency

Similarly, Figure 5 shows the concentration of the distribution of the efficiency data for the EAT and FDH models, evidencing a high concentration on the right side for the FDH model. In contrast, for the EAT model, the efficiency data tends to be distributed between 0.8 and 1.0, indicating that the FDH model has a higher average number of efficient units compared to the EAT model. The difference in the model's distribution suggests a trend of the CEAT to underestimate the efficiency scores. Additionally, it is vital to understand the information provided by the decision tree model of Figure 6; for this, Table 5 shows evidence that of the 23 nodes of the model, there are only eight nodes that contain efficient units for the CRS model and 13 nodes with efficient units for the VRS model. In addition, the average of the study variables per node is presented, taking into account only the DMUs found in that node. This efficiency structure provides a broader perspective of the efficiency results by characterizing each node as an efficiency cluster in which it is possible to determine which DMUs shares similar performances.

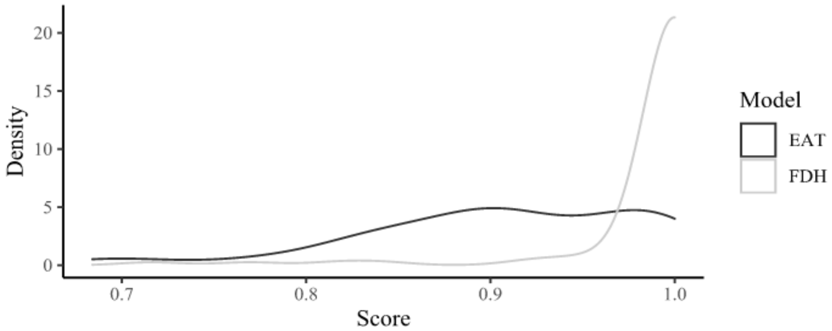

Figure 5. EAT vs FDH efficiency-density

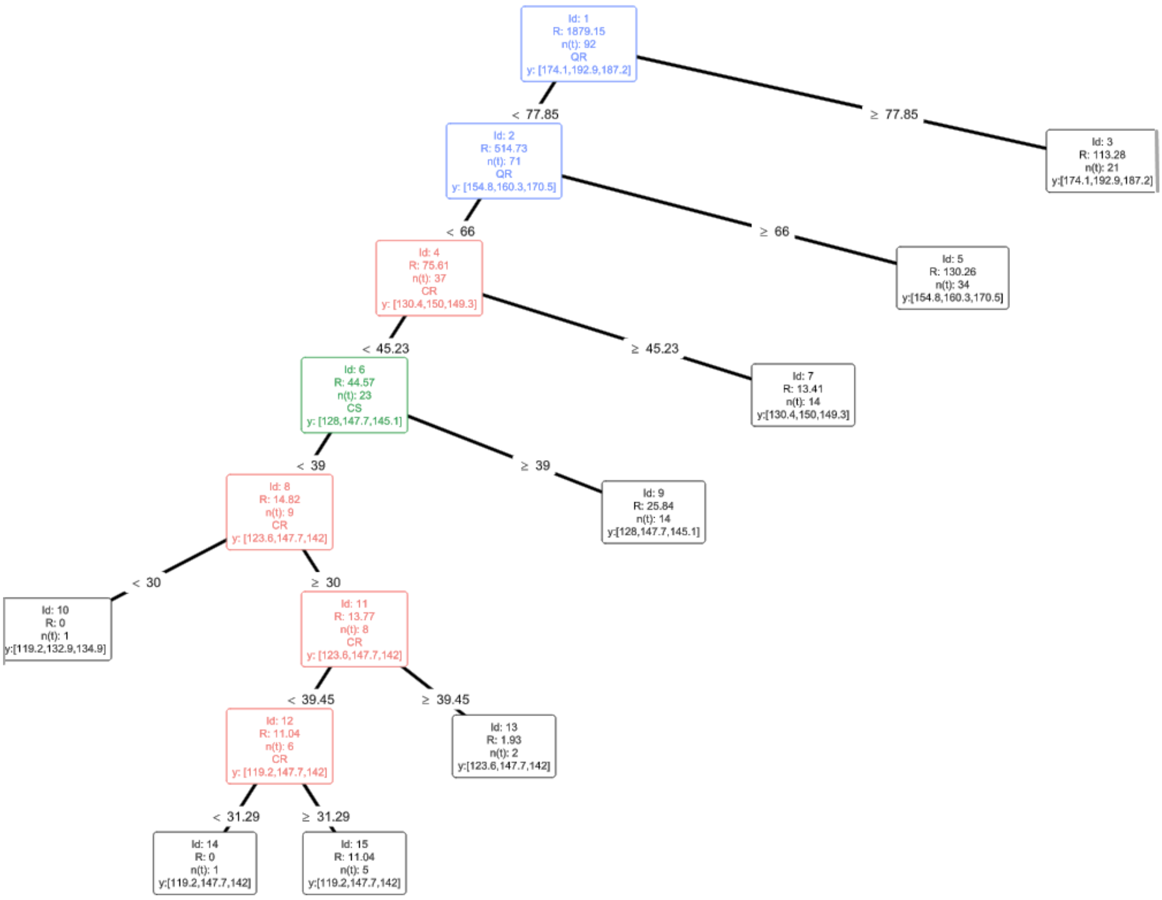

Figure 2. Decision tree for the construction of the DEA model notation: Node id (*Id*), node error (*R*), number of node observations (*n* (*t*)), node predictor variable and node predictions (*y*)

Table 5. Characterization of the decision tree nodes

Node	CR	CS	ENG	WC	QR	MSST	DPLS	FEP	DMUs	EFF (CRS)	EFF (VRS)
1	53.94	52.52	60.08	52.48	67.79	133.71	147.80	150.83	92	2 (2%)	6 (7%)
2	48.34	47.82	54.44	49.73	62.46	128.33	140.72	145.24	71	2 (3%)	5 (7%)
3	72.85	68.41	79.17	61.77	85.79	151.89	171.71	169.73	21	0 (0%)	1 (5%)
4	42.89	42.17	48.21	47.52	54.93	122.30	134.67	138.88	37	2 (5%)	3 (8%)
5	54.28	53.97	61.21	52.14	70.66	134.89	147.31	152.16	34	0 (0%)	2 (6%)
6	39.87	40.37	45.59	46.89	52.55	121.24	132.40	136.41	23	2 (9%)	2 (9%)
7	49.07	45.74	53.52	48.19	59.08	124.48	138.90	143.31	14	0 (0%)	1 (7%)
8	35.12	33.11	41.52	39.75	48.44	118.88	130.92	134.12	8	2 (25%)	1 (13%)
9	41.43	44.03	47.64	51.45	55.11	122.49	132.90	137.33	14	0 (0%)	1 (7%)
10	25.13	17.63	35.63	40.50	59.13	119.25	132.88	134.88	1	1 (100%)	0 (0%)
11	36.86	35.72	41.50	40.34	46.81	118.55	130.58	134.36	8	1 (13%)	1 (13%)
12	35.16	35.29	40.39	40.07	43.74	117.18	130.11	133.16	6	1 (17%)	0 (0%)
13	41.96	36.98	44.83	41.14	56.02	122.67	131.99	137.96	2	0 (0%)	1 (100%)
14	40.99	40.23	44.45	48.13	55.19	121.93	127.47	132.42	6	0 (0%)	1 (17%)
15	41.75	46.88	50.04	53.94	55.05	122.91	136.98	141.01	8	0 (0%)	0 (0%)
16	40.12	47.57	49.69	55.58	49.12	120.60	138.32	138.25	4	0 (0%)	0 (0%)
17	43.39	46.20	50.39	52.30	60.99	125.22	135.64	143.77	4	0 (0%)	0 (0%)
18	49.24	45.16	55.16	49.24	56.98	122.38	136.74	142.56	8	0 (0%)	0 (0%)
19	48.84	46.52	51.34	46.78	61.87	127.27	141.79	144.30	6	0 (0%)	1 (17%)
20	48.10	44.88	55.72	47.89	57.21	122.54	135.94	141.93	7	0 (0%)	0 (0%)
21	57.24	47.12	51.24	58.71	55.41	121.24	142.29	147.00	1	0 (0%)	0 (0%)
22	46.33	43.10	43.45	51.19	52.74	123.31	132.36	136.06	1	0 (0%)	0 (0%)
23	48.39	45.17	57.76	47.34	57.95	122.42	136.54	142.91	6	0 (0%)	0 (0%)

In addition, one of the advantages of this methodology for the efficiency analysis is the prediction of future efficiency scores; therefore, it is possible to have complete management of the process and take preventive actions in the face of various scenarios. Therefore, the predictive model built has five predictor variables (QR, CS, ENG, WC, and CR) and three variables of interest (FEP, MSST, and DPLS). Additionally, a random forest model is constructed to compare the results of the tree model. However, the results show that the tree model has a higher error level due to the simplicity of the model construction. In contrast, the error level of the random forest model is much lower due to its robust construction. Table 6 presents the performance metrics of the decision tree model used to build the efficiency frontier of the system.

Table 6. Performance of the models in the evaluation

Metric	Decision Tree model			Random Forest model		
	MSST	DPLS	FEP	MSST	DPLS	FEP
RMSE	18.16	17.47	16.15	5.62	9.12	5.69
RSquared	0.70	0.66	0.67	0.93	0.87	0.90
MAE	14.21	13.99	12.31	4.82	7.31	4.40

4. DISCUSSION

Education is a series of sequential and evolutionary activities of great relevance for the development of societies in political, economic, technological and knowledge matters. However, objective and reproducible tools to estimate efficiency in educational processes are scarce in the literature. Consequently, this research presents a tool for managing educational processes through efficiency analysis using EAT and DEA and comparing the EAT model with the FDH model.

Consequently, the literature indicates that there will be overfitting problems for estimating efficiency through the DEA and FDH methods [12], [24]. Conversely, the EAT model does not have problems with overfitting by implementing cross-validation and pruning. In addition, the estimation of the production frontier and the relative position of the DMUs in the EAT model is much closer to the theoretical production function (frontier) than in the FDH model. However, the above is reachable because the tree's growth occurs evolutionarily. Besides, the fitting process is achieved by minimizing the MSE and using stopping rules linked to the size of the database, avoiding empty leaf nodes [25].

For its part, it is necessary to mention that the importance of the tool lies in being useful for decision-making in educational environments. This tool helps to identify through two powerful techniques (machine learning and decision tree) the universities' efficiency level. The result of this tool is a tree structure consisting of a number of nodes representing a level of efficiency calculated from the specific characteristics of the universities that are part of the node. In this vein, the nodes can be called the goals that the universities must achieve to increase efficiency, considering that the nodes generate information about the skills each university must improve.

The results presented in Table 4 show that in each EAT model, the amount and level of efficiency of the DMUs is lower than in the classic DEA models. In addition, DEA models traditionally reach efficiency analysis, and the EAT model generates a greater contribution by providing a prediction tool for a more complete management of educational processes. In addition, an important finding is a piece of objective evidence that EAT models are stricter than traditional DEA models.

Besides, the variables with greater relevance for predicting the response variables FEP, MSST and DPLS; QR, CR, CS and ENG; in contrast, the WC according to the model has no relevance as a predictor. On the other hand, this research identified that universities with the highest levels (≥77.85) in QR could achieve a high level of efficiency with results in their specific skills MSST, DPLS and FEP of 174.1, 192.9 and 187.2, respectively. While those universities with a level lower than 77.85 in QR should strengthen their level of CR skill as a second option, if the performance exceeds 45.23 points, their performance in the MSST, DPLS and FEP skills likely is 130.4, 150, 149.3. On the other hand, if the score in their CR skill is less than 45.23, universities should strengthen their level in CS in such a way that by generating a score higher than 39, the performance in the MSST, DPLS and FEP skills is 128, 147.7 and 145.1, respectively. In summary, this research proposes a tool that eliminates the risk of overfitting by articulating machine learning techniques and DEA, estimating the efficiency and productivity of engineering degrees. Thus, setting a reproducible and replicable procedure to evaluate efficiency in other sectors different than education.

5. CONCLUSION

The research objective was to create a tool that contributes to the spectrum of knowledge about models for educational management. Consequently, the result of the research is a tool to estimate educational efficiency through the EAT and DEA models. Thus, considering the EAT model as an adaptation of the decision trees for estimating production frontiers, DEA is the classic linear programming model for efficiency analysis. Consequently, the decision tree model with the best performance considers a minimum number of six observations to perform a partition in a node; the number of partitions of the dataset to perform cross-validation during pruning is five; the number of nodes is nine, and the value of the RMSE is 38.79 (lesser than outputs standard deviation). It is essential to highlight that this model, compared with the traditional DEA model, has less complexity and provides better forecasting results.

Regarding the efficiency models, the production frontier of the EAT model in its constant scale constructed by the decision tree model presents an efficiency level of 2.17% (2 efficient DMUs). For the variable scale, the EAT model has an efficiency level of 11.95% (11 efficient DMUs). In contrast, the classic DEA model in its constant scale provides an efficiency level of 8.70% (8 efficient DMUs), and in its variable scale, its efficiency level is 89.13% (82 efficient DMUs). In addition, it is observed that the scale performance levels for the EAT and DEA models are 6.52% (6 efficient DMUs) and 8.69% (8 efficient DMUs), respectively. It should be noted that this tool is much stricter than conventional efficiency models because in the first instance, variables relevant to the problem must be selected; in addition, the set of observed data must be correctly listed in a way that avoids the problem of instability in the decision trees due to noise factors and predictors not relevant to the problem. Additionally, the robustness of the model is generated by performing the partition of the nodes to construct the efficiency frontier so that only partitions of predictors will be made a whose contribution to the explanation of the response variable is significant. Finally, the EAT model for this research has a lower efficiency level than the traditional DEA model due to the model's rigour to avoid overfitting.

REFERENCES

[1] Y. L. Zhukovskiy, B. U. Vasilev, and N. I. Koteleva, "Quality estimation of continuing professional education of technical specialists," in *2017 International Conference "Quality Management, Transport and Information Security, Information Technologies" (IT&QM&IS)*, sep. 2017, pp. 704–707, doi: 10.1109/ITMQIS.2017.8085921.

[2] A. Y. Noaman, A. H. M. Ragab, A. I. Madbouly, A. M. Khedra, and A. G. Fayoumi, "Higher education quality assessment model: towards achieving educational quality standard," *Studies in Higher Education*, vol. 42, no. 1, pp. 23–46, 2017, doi: 10.1080/03075079.2015.1034262.

[3] T. Agasisti, G. Munda, and R. Hippe, "Measuring the efficiency of European education systems by combining data envelopment analysis and multiple-criteria evaluation," *Journal of Productivity Analysis*, vol. 51, no. 2–3, pp. 105–124, Jun. 2019, doi: 10.1007/s11123-019-00549-6.

[4] J. Aparicio, J. M. Cordero, and L. Ortiz, “Measuring efficiency in education: The influence of imprecision and variability in data on DEA estimates,” *Socio-Economic Planning Sciences*, vol. 68, dic. 2019, doi: 10.1016/j.seps.2019.03.004.

[5] D. Visbal-Cadavid, A. M. Mendoza, and I. Q. Hoyos, “Prediction of efficiency in Colombian higher education institutions with data envelopment analysis and neural networks,” *Pesquisa Operacional*, vol. 39, no. 2, pp. 261–275, ago. 2019, doi: 10.1590/0101-7438.2019.039.02.0261.

[6] M. Biggeri y V. Mauro, “Towards a more ‘sustainable’ human development index: integrating the environment and freedom,” *Ecological Indicators*, vol. 91, pp. 220–231, ago. 2018, doi: 10.1016/j.ecolind.2018.03.045.

[7] S. Chansarn, “The evaluation of the sustainable human development: a cross-country analysis employing slack-based DEA,” *Procedia Environmental Sciences*, vol. 20, pp. 3–11, 2014, doi: 10.1016/j.proenv.2014.03.003.

[8] P. C. Mero, G. H. Franco, J. D. B. Bitar, and C. S. Padilla, “Geodiversity, a component of sustainable development,” (in Spanish) *Journal of Science and Research*, vol. 3, pp. 36–42, 2018.

[9] A. S. Gómez Tabares and M. Narvaez Marin, “Personality dimensions and their relationship with prosocial tendencies and empathy in children and adolescents in psychosocial vulnerability,” (in Spanish) *Revista de Psicología*, vol. 40, no. 1, pp. 37–72, ene. 2022, doi: 10.18800/psico.202201.002.

[10] E. J. Delahoz-Dominguez, T. Fontalvo, and R. Zuluaga, “Evaluation of academic productivity of citizen competencies in the teaching of engineering by using the Malmquist index,” (in Spanish) *Formación universitaria*, vol. 13, no. 5, pp. 27–34, oct. 2020, doi: 10.4067/S0718-50062020000500027.

[11] E. J. Delahoz-Dominguez, T. J. Fontalvo-Herrera, and R. A. Zuluaga-Ortiz, “Partial leasth squares - path modelling for efficiency assessment in the Colombian Professional Football League,” *Pesquisa Operacional* vol. 42, Jun. 2022, doi: 10.1590/0101-7438.2022.042.00254144.

[12] M. Esteve, J. Aparicio, A. Rabasa, and J. J. Rodriguez-Sala, “Efficiency analysis trees: A new methodology for estimating production frontiers through decision trees,” *Expert Systems with Applications*, vol. 162, Dec. 2020, doi: 10.1016/j.eswa.2020.113783.

[13] A. Emrouznejad, I. H. Osman, and A. L. Anouze, “Performance management and measurement with data envelopment analysis,” in *Proceedings of the 8th International Conference of DEA*, 2011.

[14] A. Charnes, W. W. Cooper, and E. Rhodes, “Measuring the efficiency of decision-making units,” *European Journal of Operational Research*, vol. 2, no. 6, pp. 429–444, Nov. 1978, doi: 10.1016/0377-2217(78)90138-8.

[15] E. De La Hoz, R. Zuluaga, and A. Mendoza, “Assessing and classification of academic efficiency in engineering teaching programs,” *Journal on Efficiency and Responsibility in Education and Science*, vol. 14, no. 1, pp. 41–52, mar. 2021, doi: 10.7160/eriesj.2021.140104.

[16] A. Emrouznejad, B. R. Parker, and G. Tavares, “Evaluation of research in efficiency and productivity: A survey and analysis of the first 30 years of scholarly literature in DEA,” *Socio-Economic Planning Sciences*, vol. 42, no. 3, pp. 151–157, Sep. 2008, doi: 10.1016/j.seps.2007.07.002.

[17] A. Mardani, E. K. Zavadskas, D. Streimikiene, A. Jusoh, and M. Khoshnoudi, “A comprehensive review of data envelopment analysis (DEA) approach in energy efficiency,” *Renewable and Sustainable Energy Reviews*, vol. 70, pp. 1298–1322, abr. 2017, doi: 10.1016/j.rser.2016.12.030.

[18] R. N. Rithesh, R. Vignesh, M. R. Anala, “Autonomous traffic signal control using decision tree,” *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 8, no. 3, pp. 1522–1529, jun. 2018, doi: 10.11591/ijece.v8i3.pp1522-1529.

[19] T. A. Assegie y P. S. Nair, “Handwritten digits recognition with decision tree classification: a machine learning approach,” *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 9, no. 5, pp. 4446–4451, Oct. 2019, doi: 10.11591/ijece.v9i5.pp4446-4451.

[20] L. Breiman y R. Ihaka, “Nonlinear discriminant analysis via scaling and ACE,” Tech. Rep., Department of Statistics, University of California, 1984.

[21] G. Feretzakis, D. Kalles, and V. S. Verykios, “On using linear Diophantine equations for in-parallel hiding of decision tree rules,” *Entropy*, vol. 21, no. 1, ene. 2019, doi: 10.3390/e21010066.

[22] E. Delahoz-Dominguez, R. Zuluaga, and T. Fontalvo-Herrera, “Dataset of academic performance evolution for engineering students,” *Data in Brief*, vol. 30, Jun. 2020, doi: 10.1016/j.dib.2020.105537.

[23] R Core Team, “R: A language and environment for statistical computing,” *R Foundation for Statistical Computing*, Vienna, 2013. Accessed Apr 15, 2020. [Online] Available: https://www.r-project.org/

[24] D. Valero-Carreras, J. Aparicio, and N. M. Guerrero, “Support vector frontiers: a new approach for estimating production functions through support vector machines,” *Omega*, vol. 104, oct. 2021, doi: 10.1016/j.omega.2021.102490.

[25] M. Esteve, J. J. Rodriguez-Sala, J. Juan Lopez-Espin, and J. Aparicio, “Heuristic and backtracking algorithms for improving the performance of efficiency analysis trees,” *IEEE Access*, vol. 9, pp. 17421–17428, 2021, doi: 10.1109/ACCESS.2021.3054006.

BIOGRAPHIES OF AUTHORS

Rohemi A. Zuluaga Ortiz has a master’s degree in engineering from the Universidad Tecnológica de Bolívar (UTB), and an Industrial Engineer from the Universidad Tecnológica de Bolívar (UTB). He is currently a professor at the University of Sinú. His areas of research are efficiency and learning analytics. He can be contacted at email: rohemi.zuluaga@unisinu.edu.co.

Alicia Camelo Guarín has a Ph.D. in Social Sciences, Social Communicator and Master in Communication Management from the Universidad del Norte. With experience in research of educational, Military Sciences, entertainment and communication strategies for social change. University professor and methodological consultant in the design of experiments, conducting scientific studies, validation and dissemination of results. She can be contacted at email: aliciacameloguarin@hotmail.com.

Enrique De La Hoz is a Ph.D. candidate in Networks and information systems from the Universitat Oberta de Catalunya (UOC), and received a master's degree in Operations Research from the Universitat de Barcelona and the UPC. He is currently a professor at the Technological University of Bolívar. His research areas are learning analytics, recommendation systems and large-scale data mining. He can be contacted at email: edelahoz@utb.edu.co.