A novel approach to efficiency evaluation through the integration of standard Machine Learning classification models and Data Envelopment Analysis

Ricardo González-Moyano[1], Juan Aparicio[1,2*], Víctor J. España[1] and José L. Zofío[3,4]

[1] Center of Operations Research (CIO). Miguel Hernandez University, Elche, Spain.

[2] ValgrAI - Valencian Graduate School and Research Network of Artificial Intelligence, Valencia, Spain.

[3] Department of Economics, Universidad Autónoma de Madrid, Madrid, Spain.

[4] Erasmus Research Institute of Management, Erasmus University, Rotterdam, The Netherlands.

[*] Corresponding author: j.aparicio@umh.es.

Abstract

In recent decades, the field of efficiency analysis has advanced significantly, especially in evaluating decision-making units (DMUs) across diverse sectors like finance, healthcare, education, and manufacturing. Data Envelopment Analysis (DEA) offers a non-parametric approach to assess the relative efficiency of DMUs by comparing their input-output relationships. Despite its widespread adoption, traditional DEA approaches face challenges in capturing intricate patterns and structures in complex datasets, such as overfitting and handling nonlinearity. With the emergence of machine learning (ML) techniques, there is an opportunity to enhance DEA's capabilities by leveraging ML's computational power and flexibility. This integration can potentially improve the accuracy, robustness, and interpretability of efficiency assessments, advancing performance analysis. By combining ML algorithms with DEA, researchers can develop innovative methodologies to address existing limitations and explore new avenues for efficiency analysis. Our paper contributes to this endeavor by introducing a hybrid methodological framework that integrates DEA with ML classification techniques, specifically Support Vector Machines and Neural Networks. We demonstrate the practical implications of this integration through an empirical example grounded on PISA (Programme for International Student Assessment) data. This new synergy between DEA and ML holds promise to further transform efficiency evaluation and enhancing our understanding of complex systems in production.

Keywords: Data Envelopment Analysis, Machine Learning, Classification models, robustness, variable importance.

1. Introduction

In recent decades, the field of efficiency analysis has witnessed significant advancements, particularly in the evaluation of decision-making units (DMUs) across various sectors such as finance, healthcare, education, and manufacturing. One prominent methodology that has garnered substantial attention is Data Envelopment Analysis (DEA), initially introduced by Charnes, Cooper, and Rhodes in the late 1970s (Charnes et al., 1978). DEA offers a non-parametric approach to assess the relative efficiency of DMUs by comparing their input-output relationships. The fundamental premise of DEA lies in its ability to evaluate the efficiency of DMUs that operate under multiple inputs and outputs, without imposing restrictive assumptions about functional forms or underlying distributions. This characteristic makes DEA particularly appealing for analyzing complex real-world systems where the relationships between inputs and outputs may be nonlinear or unknown. Over the years, DEA has been applied to diverse domains, including banking (Seiford & Zhu, 2002), healthcare (Olesen et al., 2007), and environmental performance assessment (Zhou et al., 2008), among others.

However, despite its widespread adoption and commendable performance, traditional DEA approaches may encounter limitations in capturing the intricate patterns and structures inherent in complex datasets. One notable challenge lies in the potential for overfitting, wherein the model captures noise or idiosyncratic features in the data rather than true underlying relationships (Esteve et al., 2020). This issue is particularly pronounced in DEA when dealing with high-dimensional datasets or when the number of DMUs is relatively small compared to the number of inputs and outputs (Charles et al., 2019). Overfitting in DEA can lead to inflated efficiency scores for certain DMUs, thereby distorting the assessment of relative efficiency and potentially misleading decision-makers. Moreover, traditional DEA models rely on linear programming techniques to estimate efficiency scores, which may not adequately capture nonlinear relationships or interactions among inputs and outputs. As a result, the model may overlook nuanced patterns in the data, leading to biased efficiency estimates. Another significant limitation of traditional DEA is its deterministic nature. Traditional DEA models produce a single efficiency score for each DMU based on the observed input-output data, without accounting for uncertainties or variability inherent in real-world systems. This deterministic approach fails to acknowledge the stochastic nature of many decision-making processes. The deterministic nature of DEA overlooks inherent variability in input and output data, which may arise due to measurement errors, random fluctuations in production processes, or external factors beyond the control of the DMU. Moreover, the deterministic framework of traditional DEA precludes the incorporation of risk considerations into efficiency analysis. Decision-makers in practical settings often face

uncertain environments where outcomes are subject to randomness or unpredictability. By neglecting uncertainty, traditional DEA models provide a narrow perspective on efficiency that fails to account for the associated risks and trade-offs inherent in decision-making.

With the advent of machine learning techniques, there exists a compelling opportunity to enhance the capabilities of DEA by leveraging the computational power and flexibility offered by these methods. By integrating machine learning algorithms with DEA, researchers can potentially improve the accuracy, robustness, and interpretability of efficiency assessments, thereby advancing the state-of-the-art in performance analysis. In the era of Artificial Intelligence (AI), where machine learning algorithms permeate various aspects of our lives, there arises an imperative for scientific inquiry to bridge disciplinary boundaries and foster interdisciplinary collaborations. However, despite its remarkable achievements, the full potential of machine learning remains untapped unless integrated synergistically with other domains of knowledge.

In this context, it becomes almost a scientific duty to create the necessary bridges between machine learning and other fields, such as Data Envelopment Analysis. The combination of machine learning and DEA holds immense promise for enhancing our understanding of efficiency dynamics in real-world settings. Machine learning algorithms can complement DEA by providing advanced techniques for, for example, data preprocessing (Chen et al., 2014), variable importance measurement (Valero-Carreras et al., 2024), and the treatment of the curse of dimensionality (Esteve et al., 2023), thereby facilitating more accurate and comprehensive efficiency assessments. Moreover, machine learning models can capture nonlinear relationships and interactions among inputs and outputs, addressing one of the key limitations of traditional DEA approaches.

In the literature, several bridges between machine learning (ML) and Data Envelopment Analysis (DEA) have already been established. However, we have identified certain gaps that we believe our approach introduced in this paper can address. Before mentioning these gaps, we briefly review the main contributions related to ML and DEA. As we are aware, in the literature, there are two predominant streams of research that explore the integration of machine learning with Data Envelopment Analysis[1]. The first stream focuses on adapting existing ML techniques to

[1]A third line of research in the literature employs Data Envelopment Analysis (DEA) as an alternative method to conventional Machine Learning (ML) classification techniques such as Support Vector Machines (SVM), decision trees, and neural networks. In that line, DEA is utilized to classify observations based on

ensure that the predictive function, typically representing a production function in our context, complies with various shape constraints such as monotonicity or concavity. Researchers in this stream leverage techniques from ML, such as support vector machines (SVM), neural networks, or decision trees, to develop models that capture the underlying relationships between inputs and outputs by imposing shape constraints on the predictive function. Some of these contributions are the following: Kuosmanen and Johnson (2010) demonstrated a novel connection between DEA and least-squares regression, introducing Stochastic Non-smooth Envelopment of Data (StoNED). Parmeter and Racine (2013) proposed innovative smooth constrained nonparametric frontier estimators, incorporating production theory axioms. Daouia et al. (2016) introduced a method using constrained polynomial spline smoothing for data envelope fitting, enhancing precision and smoothness. Esteve et al. (2020) developed Efficiency Analysis Trees (EAT), improving production frontier estimation through decision trees. Valero-Carreras et al. (2021) introduced Support Vector Frontiers (SVF), adapting Support Vector Regression for production function estimation. Aparicio et al. (2021) provided an overview of EAT for estimating production frontiers using ML techniques. Olesen and Ruggiero (2022) proposed hinging hyperplanes as a nonparametric estimator for production functions. Guerrero et al. (2022) introduced Data Envelopment Analysis-based Machines (DEAM) for estimating polyhedral technologies. Valero-Carreras et al. (2022) adapted SVF for multi-output scenarios, improving efficiency measurement. Guillen et al. (2023a, 2023b, 2023c) introduced boosting techniques for efficiency estimation in different scenarios. Tsionas et al. (2023) proposed a Bayesian Artificial Neural Network approach for frontier efficiency analysis. Liao et al. (2024) proposed Convex Support Vector Regression (CSVR) to improve predictive accuracy and robustness in nonparametric regression. On the other hand, the second stream of literature adopts a two-stage approach to integrate DEA with ML techniques. In the first stage, researchers apply a pre-existing DEA model, such as the output-oriented radial model, to compute efficiency scores for each observation in the sample (DMUs). In the second stage, the efficiency scores obtained from DEA are treated as the response variable in a regression model based on standard ML techniques (without shape constraints). The original inputs and outputs, along with potentially additional environmental variables, serve as predictor variables in the regression model. By incorporating ML techniques to the performance evaluation framework, researchers aim to develop more robust and accurate predictive models for assessing efficiency. Some of these contributions are the following: Emrouznejad & Shale (2009) explored a novel approach by combining a neural network with Data Envelopment Analysis (DEA) to address the computational challenges posed by large datasets. Liu et al. (2013) compared standard DEA, three-stage DEA, and neural network approaches to measure the technical efficiency of 29 semi-conductor firms in Taiwan. Fallahpour et al. (2016) presented an integrated model for green

their features. For example, it is applied to identify individuals as carriers of a rare genetic disorder from age and four blood measurements. A recent example of this type of contributions is Jin et al. (2024).

supplier selection under a fuzzy environment, combining DEA with genetic programming to address the shortcomings of previous DEA models in supplier evaluation. Kwon et al. (2016) explored a novel method of performance measurement and prediction by integrating DEA and neural networks. The study used longitudinal data from Japanese electronics manufacturing firms to show the effectiveness of this combined approach. Aydin & Yurdakul (2020) introduced a three-staged framework utilizing Weighted Stochastic Imprecise Data Envelopment Analysis and ML algorithms to assess the performance of 142 countries against the COVID-19 pandemic. Tayal et al. (2020) presented an integrated framework for identifying sustainable manufacturing layouts using Big Data Analytics, Machine Learning, Hybrid Meta-heuristic and DEA. The paper by Nandy & Singh (2020) presented a hybrid approach utilizing DEA and Machine Learning, specifically the Random Forest (RF) algorithm, to evaluate and predict farm efficiency among paddy producers in rural eastern India. Zhu et al. (2021) proposed a novel approach that combines DEA with ML algorithms to measure and predict the efficiency of Chinese manufacturing companies. Jomthanachai et al. (2021) proposed an integrated method combining Data Envelopment Analysis and Machine Learning for risk management. Boubaker et al. (2023) proposed a novel method for estimating a common set of weights based on regression analysis (such as Tobit, LASSO, and Random Forest regression) for DEA to predict the performance of over 5400 Vietnamese micro, small and medium enterprises. Amirteimoori et al. (2023) introduced a novel modified Fuzzy Undesirable Non-discretionary DEA model combined with artificial intelligence algorithms to analyze environmental efficiency and predict optimal values for inefficient Decision-Making Units (DMUs), focusing on CO_2 emissions in forest management systems. Lin & Lu (2024) presented a novel analytical framework utilizing inverse Data Envelopment Analysis and ML algorithms to evaluate and predict suppliers' performance in a sustainable supply chain context.

Both streams of research have contributed valuable insights and methodologies for integrating ML with DEA. However, despite these advancements, there remain certain gaps and limitations that we aim to address in this paper. Specifically, the methodological innovations introduced in this article align more closely with the second stream of the literature than the first one. Techniques within this second group take a smart approach by using the DEA score obtained in the first stage as the response variable in the second stage. However, this strategy poses significant challenges in uncertain, indeterminate, and noisy contexts, where distinguishing between 0.9 and 1.0 regarding efficiency score is difficult. Moreover, techniques in this second group use the same DEA efficiency score determined for each DMU in the first stage as the final evaluation for efficiency of the observations. Therefore, the efficiency evaluation of the data sample is not 'improved' by incorporating ML techniques in the second stage and, therefore, the corresponding

ranking of DMUs remains the same as the original one. These are the two gaps we identify and aim to address in this paper. In this sense, and for the first time in the literature, we will use a classification model rather than a regression model in the second stage of the approach that combines DEA and ML. In fact, we will employ a standard DEA model in the first stage to identify, through Pareto-dominance efficiency evaluation, a labelling that distinguishes between efficient and inefficient units. And, in the second stage, we will attempt to predict this label using all variables of the problem. Additionally, our approach will allow us to modify the measurement of the degree of efficiency of observations, as the efficiency score will be calculated using an eXplainable Artificial Intelligence (XAI) method based on the use of a counterfactual: technical inefficiency will be defined for an inefficient DMU as the minimum changes required in inputs and outputs (or in a certain direction depending on the model orientation and other factors) to change from the inefficient label to the efficient label. Additionally, we aim to demonstrate that DEA can be viewed as a particular case of our approach in the sense that the DEA frontier could be interpreted as the separating surface in the input-output space of the two existing classes (labels): efficient units vs. inefficient units; with the peculiarity of having all efficient DMUs located on the separating surface (the efficient frontier). Therefore, the conceptual foundation motivating the formulation of our counterfactual method aligns with the principles underpinning the conventional approach for quantifying inefficiency in DEA. This entails projecting inefficient units onto the DEA technology frontier until reaching a state where they no longer deviate from the production possibility set (achieving efficiency).

Moreover, the determination of variable (inputs and outputs) importance within DEA models has been pivotal in the literature. As highlighted by Banker and Morey (1986), comprehending the significant contributing factors to relative efficiency empowers organizations to channel efforts towards areas where substantial improvements can be achieved. Moreover, as suggested by Thanassoulis et al. (2015), identifying the most relevant variables not only facilitates strategic decision-making but also provides valuable insights for optimal resource allocation and the implementation of continuous improvement measures. Hence, the assessment of variable importance in the production process is fundamental for maximizing efficiency and productivity across various industries. In this line, another of our objectives is to enhance the methodological framework for determining variable importance in DEA models. While existing studies have provided valuable insights into the significance of variables (Pastor et al., 2002), there is still room for refinement and advancement. Specifically, by incorporating advanced machine learning algorithms, we seek to provide more robust and accurate assessments of variable importance, thereby enabling organizations to make informed decisions and drive continuous improvement initiatives effectively.

This paper aims to explore the synergies between DEA and machine learning techniques, elucidating the potential benefits of their integration in the context of efficiency evaluation. Specifically, we discuss various approaches for combining DEA with machine learning within the category of classification models, introducing a new hybrid framework that integrates both techniques. The paper is structured as follows: In Section 2, we provide background information on Data Envelopment Analysis (DEA) and the two machine learning techniques we will utilize, namely Support Vector Machines (SVM) and Neural Networks. Section 3 introduces our novel approach, which integrates DEA with these two classification techniques, aiming to enhance efficiency assessment for decision-making units (DMUs). We demonstrate the practical implications of this integration and its implications for decision-making and policy formulations through an empirical example based on PISA (Programme for International Student Assessment) in Section 4. Section 5 concludes and points out further research lines.

2. Background

This background section provides a concise yet comprehensive overview of DEA and the main ML techniques that we will apply in this paper (Support Vector Machines and Neural Networks).

2.1. Data Envelopment Analysis

Data Envelopment Analysis (DEA) is a non-parametric method widely used for evaluating the relative efficiency of decision-making units (DMUs) in various fields, including economics, finance, and operations research. It was first introduced by Charnes et al. (1978). DEA offers a powerful framework for assessing the efficiency of DMUs that convert multiple inputs into multiple outputs. DEA operates under the assumption of constant returns to scale (CRS) or variable returns to scale (VRS). VRS is particularly suitable for analyzing real-world production processes, where economies of scale may vary across different units.

In this paper, we are going to explore the evaluation of n units regarding their technical efficiency. These units, which could be firms or organizations, are referred to as Decision-Making Units (DMUs). They utilize various inputs $\mathbf{x}_j = \left(x_{1j},\ldots,x_{mj}\right) \in R_+^m$, such as resources, to generate outputs $\mathbf{y}_j = \left(y_{1j},\ldots,y_{sj}\right) \in R_+^s$, like goods or services; where vectors are represented by letters that are in bold typeface. In a conceptual framework, the term ‘technology’ (also called production

possibility set) encompasses all feasible input-output combinations. This concept is typically represented as:

$$T = \left\{ (\mathbf{x}, \mathbf{y}) \in R_+^{m+s} : \mathbf{x} \text{ can produce } \mathbf{y} \right\}. \tag{1}$$

Among the non-parametric methodologies utilized to ascertain the set T, Data Envelopment Analysis (DEA) stands out as one of the most commonly employed approaches in practical applications. Under VRS, Banker et al. (1984) shows how to estimate T through DEA as:

$$T_{DEA} = \left\{ (\mathbf{x}, \mathbf{y}) \in R_+^{m+s} : y_r \le \sum_{j=1}^{n} \lambda_j y_{rj}, \forall r, x_i \ge \sum_{j=1}^{n} \lambda_j x_{ij}, \forall i, \sum_{j=1}^{n} \lambda_j = 1, \lambda_j \ge 0, \forall j \right\}. \tag{2}$$

Within the existing body of literature, numerous technical efficiency measures are available for application through the utilization of DEA (refer, for instance, to Pastor et al., 2012). In light of this, our focus is directed towards a prevalent measure, namely, the output-oriented radial model:

$$\begin{aligned}
\phi_{DEA}(\boldsymbol{x}_o, \boldsymbol{y}_o) = \quad & \max \quad \phi_o & & (3.0)\\
& s.t. \quad \sum_{j=1}^{n} \lambda_{jo} x_{ij} \le x_{io}, & i = 1,\ldots,m & (3.1)\\
& \sum_{j=1}^{n} \lambda_{jo} y_{rj} \ge \phi_o y_{ro}, & r = 1,\ldots,s & (3.2)\\
& \sum_{j=1}^{n} \lambda_{jo} = 1, & & (3.3)\\
& \lambda_{jo} \ge 0, & j = 1,\ldots,n & (3.4)
\end{aligned} \tag{3}$$

Under the output-oriented radial model, a DMU with a score of one is considered fully efficient, indicating that it operates on the efficiency frontier. Conversely, a radial measure greater than one implies inefficiency relative to the reference technology, with a bigger value indicating a worse degree of efficiency. The radial measure provides valuable insights into the performance of individual DMUs and can guide decision-makers in identifying opportunities for improvement.

2.2. Two well-known Machine Learning Techniques for Classification

In this subsection, we briefly outline the fundamentals of the two machine learning techniques that will be employed throughout the article (Support Vector Machines, SVM, and Neural Networks, NN), as well as eXplainable Artificial Intelligence (XAI). SVM is a powerful supervised learning algorithm used for classification and regression tasks. It works by finding the hyperplane that best separates the data points into different classes while maximizing the margin between classes. On the other hand, NN are a class of deep learning algorithms inspired by the structure and function of the human brain. They consist of interconnected layers of neurons that process input data through nonlinear transformations to learn complex patterns and relationships.

By understanding the underlying principles of SVM and NN, we can try to harness their capabilities to complement the DEA framework.

2.2.1. Support Vector Machines

Support Vector Machines stand as a stalwart within the machine learning arsenal, revered for their versatility and robust performance, particularly in classification tasks. A classification problem is distinguished from a regression problem by the nature of the target variable. In classification, the target variable is categorical and represents membership in a discrete class or category, whereas in regression, the target variable is continuous and represents a numerical quantity. For example, consider a classification problem where we aim to predict whether an email is spam or not spam based on various message features such as the frequency of certain keywords, text length, and the presence of hyperlinks. Here, the target variable would be binary: spam or not spam. Conversely, in a regression problem, we might want to predict the price of a house based on features like size, location, and number of bedrooms, where the target variable would be the price, a continuous quantity. In this paper, we will focus our attention on the context of binary classification: efficient units vs inefficient units. In this regard, this section offers a brief exploration of the main elements of SVM when it is used for classification tasks.

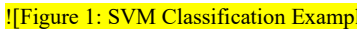
At its core, SVM operates on the principle of identifying an optimal hyperplane that effectively separates data points belonging to distinct classes (usually two classes or labels) in the feature space. This hyperplane is strategically positioned to maximize the margin, representing the perpendicular distance between the hyperplane and the closest data points from each class, known as support vectors. The seminal work of Vapnik and Cortes (1995) in the early 1990s laid the theoretical groundwork for SVM, emphasizing the importance of maximizing the margin to enhance generalization performance and avoids overfitting problems.

A pivotal aspect of SVM lies in its ability to leverage kernels for achieving non-linear transformations in the feature space (the space defined from, for instance, the frequency of certain keywords, text length, and the presence of hyperlinks in the example mentioned above). Kernels serve as a mechanism to map the input data into a higher-dimensional space, where linear separation of the two classes under study becomes feasible. Common kernel functions include the linear kernel, polynomial kernel, radial basis function (RBF) kernel, and sigmoidal kernel. Each kernel induces a specific transformation, altering the shape of the decision boundary and enabling SVM to capture complex relationships within the data. This transformative power of kernels enhances SVM's flexibility and enables it to tackle diverse classification tasks with varying

degrees of complexity. However, in practice, the performance of SVM models heavily depends on the selection of hyperparameters, such as the regularization parameter (C), the margin (ε) and the choice of kernel function (which contains several kernel-specific parameters). To optimize model performance and prevent overfitting, cross-validation emerges as a valuable technique. Cross-validation involves partitioning the dataset into multiple subsets, training the SVM model on a subset, and evaluating its performance on the remaining data. By systematically varying hyperparameters and evaluating model performance across different subsets, cross-validation enables the selection of optimal hyperparameters that generalize well to unseen data.

Furthermore, SVM offers a means to assess the importance of predictors in predicting the response variable. By analyzing the coefficients associated with each predictor in the trained SVM model, one can gauge the relative influence of different features on the classification outcome. This feature importance analysis provides valuable insights into the underlying data dynamics, guiding feature selection and model interpretation efforts.

To illustrate the practical application of SVM in classification, consider a dataset comprising two classes, depicted by red and blue points in a two-dimensional feature space. Figure 1 showcases this scenario, where the circles denote the support vectors, the solid line represents the decision boundary, and the dashed lines delineate the margins.



In Figure 1, the decision boundary (solid line) adeptly separates the red and blue points, with the margins (dashed lines) maximized through strategic placement of support vectors (circles). This example illustrates SVM's capability to discern intricate decision boundaries in high-dimensional feature spaces; evidence to its robustness in handling complex classification tasks.

2.2.2. Neural Networks

Neural Networks represent a cornerstone in the field of machine learning, heralded for their ability to learn complex patterns and relationships from data (LeCun et al., 2015; Goodfellow et al., 2016). In this subsection, we briefly delve into the application of Neural Networks in the context of classification tasks, highlighting their versatility, theoretical foundations, and practical implications.

Neural Networks are inspired by the structure and function of the human brain, comprising interconnected layers of artificial neurons or nodes. The core principle underlying Neural Networks is the process of forward propagation, where input data is sequentially passed through multiple layers of neurons, each layer applying a set of weights and activation functions to produce an output. Through an iterative process known as backpropagation, Neural Networks adjust the weights of connections between neurons based on the error between predicted and actual outputs, thereby minimizing a certain loss function and improving predictive accuracy. In this sense, activation functions play a crucial role in Neural Networks by introducing non-linearity into the model, enabling it to capture complex relationships within the data. Common activation functions include the sigmoidal function, hyperbolic tangent (tanh) function, and rectified linear unit (ReLU) function. Each activation function introduces different properties to the model, influencing its ability to learn and generalize from data.

Similar to SVM, the performance of Neural Networks hinges on the selection of hyperparameters such as the number of layers, the number of neurons per layer, learning rate, and regularization parameters. Hyperparameter tuning is essential to optimize model performance and prevent issues like overfitting or underfitting. Techniques such as grid search, random search, and Bayesian optimization are commonly employed to systematically explore the hyperparameter space and identify optimal configurations.

Despite their remarkable predictive capabilities, one challenge of Neural Networks lies in their black-box nature, which hinders interpretability and understanding of model decisions. However, techniques such as layer-wise relevance propagation (LRP) and gradient-based attribution methods can provide insights into feature importance and highlight the contribution of input features to model predictions. This feature importance analysis aids in model interpretation and decision-making processes.

METER EJEMPLO GRAFICO Y COMENTARIOS AL ESTILO DE LA SECCION ANTERIOR.

![Figure 2: Neural Network Classification Example](neural_network_classification_example.png)

2.3. eXplainable Artificial Intelligence

eXplainable Artificial Intelligence (XAI) has emerged as a critical area of research aimed at enhancing the transparency, interpretability, and trustworthiness of machine learning models (see, for example, Wachter et al., 2017). In this section, we provide an overview of XAI principles and delve into the concept of counterfactual methods, a subset of XAI techniques that facilitate insightful explanations of model predictions.

Overall, XAI encompasses a diverse set of methodologies and techniques designed to elucidate the decision-making process of machine learning models. As AI (Artificial Intelligence) systems become increasingly complex and ubiquitous, there is a growing need for transparency and interpretability to foster trust and facilitate human understanding of model behavior. XAI aims to address this need by providing explanations that are understandable, intuitive, and actionable for end-users, stakeholders, and domain experts.

In particular, counterfactual methods represent a prominent approach within the realm of XAI, focusing on the generation of alternative scenarios or 'counterfactuals' to explain model predictions. The fundamental concept underlying counterfactual methods is the creation of hypothetical instances that are similar to the observed data but differ in one or more attributes. By systematically altering the features of a given instance and observing the corresponding changes in model predictions, counterfactual methods provide valuable insights into the factors driving model decisions and predictions. Moreover, counterfactual explanations offer intuitive and interpretable insights into machine learning models by highlighting the causal relationships between features and model outcomes. These explanations typically take the form of "what-if" scenarios, where adjustments are made to features to generate counterfactual instances that lead to desired outcomes. By identifying the minimal changes required to alter a model prediction, counterfactual explanations shed light on the underlying decision-making process and enable decision-makers to understand the model's behavior in specific contexts (for example, in our production context the question could be "What is the minimum amount of adjustment in inputs and/or outputs that a technically inefficient DMU would need to undertake to transition into being considered efficient?"). Thus, the counterfactual method involves projecting an observation from one class onto the separating surface of the two classes, meaning the projection stops just before a change in label occurs. This 'projection' strategy will be incorporated to our approach in this paper to measure technical efficiency in the context of machine learning and efficiency analysis (see Section 3).

3. Integrating ML techniques for classification and Data Envelopment Analysis

In this section, we explore the integration of any machine learning technique for classification tasks with Data Envelopment Analysis (DEA) to enhance the measurement of technical efficiency. By combining the strengths of both methodologies, we aim to provide robust and insightful efficiency assessments of a set of DMUs.

Before introducing our approach, we aim to elucidate the reinterpretation of DEA, through a graphical toy example (Figure 3), as a classification method that also resorts to counterfactual analysis. DEA can be conceptualized as a classification model wherein the two classes represent feasible and infeasible units of production, with the boundary delineating the separating surface and efficient units positioned precisely onto this surface. Furthermore, this reinterpretation implies that typical efficiency measures utilized in DEA stem from the application of eXplainable Artificial Intelligence (XAI) principles, particularly involving the notion of a counterfactual scenario. Specifically, the movement of an inefficient DMU, by improving its observed inputs and/or outputs in accordance with the orientation and type of efficiency measure, signifies its transition away from its original class label (feasible) through its projection onto the efficient frontier (the separating surface). This movement quantifies the level of technical inefficiency within the DEA framework, thus highlighting the conceptual linkage between DEA and XAI principles.

XXXXFigure 3. Grafico de DEA típico con una proyección de tipo radial output.

After drawing a parallel between standard DEA approaches and classification ML methods, and, most importantly for us, demonstrating that DEA efficiency measures can be seen as emerging from the concept of XAI, particularly from a counterfactual approach, we now proceed to introduce our approach. The core concept underlying our model is a multi-stage methodology aimed at enhancing efficiency assessment through the fusion of DEA and ML techniques. Our method operates in three distinct phases: Firstly, we employ standard DEA to categorize decision-making units into efficient and inefficient categories. Subsequently, in a second phase, we employ a classification ML model, wherein the response variable is the efficiency class (efficient vs. inefficient), and the features encompass both inputs and outputs. Finally, in the third phase of our approach, we ascertain a robust measure of technical inefficiency through the application of XAI principles. Specifically, given a model measuring technical efficiency (such as the output-oriented radial model), we determine the minimum increase required in the output of each inefficient DMU

to transition its class from inefficient to efficient. This structured approach not only facilitates the identification of inefficiencies but also provides actionable insights for decision-makers to enhance performance. For instance, a similar concept can be extended to the efficient units within the framework of DEA. By doing so, we can ascertain a measure indicative of super-efficiency, thereby discerning among the subset of efficient DMUs. Andersen and Petersen (1993) introduced the notion of super-efficiency in DEA. This concept revolves around assessing each observation in relation to all other units within the dataset, wherein the evaluated observation is deliberately omitted from the analysis. Essentially, super-efficiency gauges the efficiency of a DMU by excluding the evaluated observation from the reference technology.

Next, we introduce our approach as an algorithm with different steps to be carried out.

Step 1: Utilize the additive DEA model (Charnes et al., 1985), model (4), to partition decision-making units into categories of efficiency and inefficiency based on the optimal value of the optimization program. A value of zero indicates that the evaluated unit is not Pareto-dominated by any technically feasible input-output combination within the standard DEA production possibility set. This condition underscores the exceptional efficiency of the evaluated unit, demonstrating that there is no room for enhancing any input and/or output without compromising the feasibility of the unit under assessment.

$$
\begin{aligned}
A_{DEA}\left(\boldsymbol{x}_o, \boldsymbol{y}_o\right) = \quad & \max \quad \sum_{i=1}^{m} s_{io}^{-} + \sum_{r=1}^{s} s_{ro}^{+} && && (4.0)\\
& s.t. \quad \sum_{j=1}^{n} \lambda_{jo} x_{ij} = x_{io} - s_{io}^{-}, && i = 1,\ldots,m && (4.1)\\
& \qquad \sum_{j=1}^{n} \lambda_{jo} y_{rj} = y_{ro} + s_{ro}^{+}, && r = 1,\ldots,s && (4.2)\\
& \qquad \sum_{j=1}^{n} \lambda_{jo} = 1, && && (4.3)\\
& \qquad \lambda_{jo} \geq 0, && j = 1,\ldots,n && (4.4)\\
& \qquad s_{io}^{-}, s_{ro}^{+} \geq 0, && \forall i, \forall r && (4.5)
\end{aligned}
\tag{4}
$$

If $A_{DEA}\left(\boldsymbol{x}_o, \boldsymbol{y}_o\right) > 0$, then DMU $\left(\boldsymbol{x}_o, \boldsymbol{y}_o\right)$ is (technically) inefficient. The set of all inefficient DMUs is denotes as $\boldsymbol{I}$. Otherwise, that is, if $A_{DEA}\left(\boldsymbol{x}_o, \boldsymbol{y}_o\right) = 0$, then DMU $\left(\boldsymbol{x}_o, \boldsymbol{y}_o\right)$ is (technically) efficient. The set of all efficient DMUs is denotes as E.

Step 2: Addressing the challenge of class imbalance is crucial for prediction by means of ML techniques (see, for example, XXXX). In particular, in our production context, datasets typically exhibit a higher proportion of inefficient units, which can skew model outcomes and adversely affect the accuracy of predictions. To overcome this hurdle, we propose balancing the sample of data. This step involves adjusting the class distribution to achieve parity between efficient and inefficient units. The selected technique for achieving this balance is synthetic data generation. In practice, this method is primarily applied to augment the representation of efficient units, which are often less prevalent in real datasets. This enrichment of the dataset contributes to more effective generalization by mitigating the bias introduced by the original class imbalance. Next, we talk about the process that we implement in practice to generate the synthetic units.

CONTAR AQUI COMO SE GENERAN LAS UNIDADES EFICIENTES PARA BALANCEAR LAS CLASES.

Step 3: Implement a classification ML model in this phase, where the dependent variable denotes the efficiency status (efficient [class +1] vs inefficient [class -1]), while the independent variables (features) comprise the input and output vectors. In this step, the parameters of the ML model will also be fine-tuned through cross-validation, ensuring the determination of an optimal parameter configuration and a final classification model $\Gamma(\boldsymbol{x},\boldsymbol{y})$. $\Gamma(\boldsymbol{x},\boldsymbol{y})$ predicts the classification of input-output bundle $(\boldsymbol{x},\boldsymbol{y})$ as (technically) efficient (+1) or inefficient (-1).

Step 4: Select a standard technical efficiency measure (for example, the output-oriented radial model). Then, calculate the minimum changes required in inputs and outputs (following the projection strategy marked by the chosen efficiency measure) of each inefficient DMU to transition its classification from inefficient to efficient. In this way, we are applying a counterfactual analysis. The optimization program to be solved is the following one in the case of resorting to the output-oriented radial model for evaluating unit $(\boldsymbol{x}_o,\boldsymbol{y}_o)\in I$:

$$\min\{\tau_o:(\boldsymbol{x}_o,\tau_o\boldsymbol{y}_o)\in E,\tau_o\geq 1\}=\min\{\tau_o:\Gamma(\boldsymbol{x}_o,\tau_o\boldsymbol{y}_o)=+1,\tau_o\geq 1\}. \quad (5)$$

In particular, to solve model (5), we will employ an approximate strategy outlined as follows (inspirated by the line search algorithm without using derivatives by Bazaraa et al., 2006). CONTAR.

Additionally, it is possible to extend the Step 4 above to efficient units to measure super-efficiency, thereby distinguishing among the subset of Pareto-efficient DMUs in the data sample. To do that, we must solve the following optimization program for each observation $(\boldsymbol{x}_o, \boldsymbol{y}_o) \in E$:

$$\max\{\tau_o : (\boldsymbol{x}_o, \tau_o \boldsymbol{y}_o) \in I, \tau_o < 1\} = \max\{\tau_o : \Gamma(\boldsymbol{x}_o, \tau_o \boldsymbol{y}_o) = -1, \tau_o < 1\}. \quad (6)$$

In comparison to model (5), in model (6), we have replaced 'min' with 'max'. This adjustment is made because, in this scenario, we aim to identify the first value of τ_o, with $\tau_o < 1$, for which the output-oriented radial projection of $(\boldsymbol{x}_o, \boldsymbol{y}_o)$, representing an efficient unit, transitions to being considered inefficient according to the classification model, that is, the first value of τ_o such as $(\boldsymbol{x}_o, \tau_o \boldsymbol{y}_o) \in I \Leftrightarrow \Gamma(\boldsymbol{x}_o, \tau_o \boldsymbol{y}_o) = -1$.

Furthermore, we also use ML techniques, specifically Support Vector Machine (SVM) and Neural Networks (NN), to elucidate the significance of variables within our model. ML methods offer a robust framework for feature importance analysis, allowing us to discern the most influential factors driving the efficiency classification of decision-making units (DMUs). For SVM models, variable importance is typically inferred through examining the weights assigned to support vectors, where larger weights correspond to greater importance in separating different classes or categories. Additionally, techniques such as Recursive Feature Elimination (RFE) can be employed to iteratively identify and remove less relevant variables, thereby emphasizing the ones contributing most significantly to model performance. On the other hand, NN employ diverse strategies for assessing variable importance, including sensitivity analysis, gradient-based methods, and layer-wise relevance propagation. Sensitivity analysis involves perturbing individual input variables and observing the resulting changes in model output, providing insights into their relative impact. Gradient-based methods leverage the gradients of loss functions with respect to input variables to quantify their contribution to model predictions. Layer-wise relevance propagation decomposes prediction scores across network layers, attributing relevance to input features based on their influence on subsequent layers' activations. By harnessing these sophisticated techniques within our SVM and NN frameworks, we aim to unravel the nuanced

interplay between input-output variables and efficiency outcomes, thus enhancing the interpretability and utility of our DEA-ML integration approach.

Next, we will illustrate our method through a toy numerical example, complemented by several figures. For the classification ML model, we will employ Support Vector Machines (SVM).

XXXX Aquí debe mostrarse además con mucho detalle qué se hace con los hiperparametros, cómo se mide la importancia de las variables, etc, etc. Que se modifica el ranking de ineficiencia. Que se genera un ranking para las eficientes (interpretado como supereficiencia). Comparar con la eficiencia que da DEA y la supereficiencia que da DEA. Etc.

In the following section, we will demonstrate the merits of our method through its application to an empirical example based on data from the Programme for International Student Assessment (PISA) report. This empirical application will serve to showcase the practical effectiveness and utility of our approach in real-world scenarios, particularly in the context of educational performance evaluation and policy formulation.

4. An empirical application: the efficiency assessment of the Spanish educational sector

In this section, we will exemplify the application of our novel algorithm to a genuine dataset sourced from a public service, thereby evaluating its efficacy in estimating the Data Generating Process (DGP), responsible for generating the sample, and making predictions for unobserved data outputs. To illustrate our methodology, we will utilize data obtained from the Programme for International Student Assessment (PISA), administered by the Organization for Economic Co-operation and Development (OECD). PISA evaluates the competencies of students nearing the end of compulsory education, assessing their aptitude in essential academic skills necessary for effective participation in contemporary societies. Our empirical investigation focuses on analyzing schools as the fundamental unit, consistent with prevailing practices in educational efficiency evaluations (Johnes, 2015; Witte and López-Torres, 2017). This selection ensures alignment with prior research and relevance to ongoing discussions concerning educational institutions and their operational effectiveness. The dataset utilized encompasses data from the year 2018, comprising anonymized records from 1047 Spanish schools randomly selected by the OECD.

Spain's educational system is decentralized, organized into autonomous communities, each with distinct educational policies and practices. This decentralized structure adds complexity to our analysis, as variations across regions can significantly influence overall educational performance in PISA assessments. Understanding these regional nuances is essential for accurate interpretation and targeted interventions within Spain's diverse educational landscape.

Assessing efficiency in the education sector involves examining input variables such as educational resource quality (EDUQUAL), reflecting available physical resources; the socioeconomic status index of students (ESCS), and the teacher-student ratio (TSRATIO), representing human resources within each school. Output variables considered are standardized test scores in mathematics (PVMATH), reading (PVREAD), and science (PVSCIE). Table 1 presents average values for each variable, along with standard deviations (in parentheses) and sample sizes for each autonomous community.

The observed variability in input and output variables across regions underscores significant disparities in educational resources and outcomes, emphasizing the need to investigate regional differences comprehensively. Given that the PISA dataset represents only a subset of the total population, our objective is not to calculate precise technical efficiencies of observed schools. Instead, we aim to leverage the estimated education production function to predict outcomes for schools beyond the observed sample. Consequently, a compelling scenario for educational decision-makers involves optimizing the allocation of educational and human resources to enable schools to attain or surpass certain thresholds in mathematics, reading, and science scores. Notably, modifying the socioeconomic status of students (ESCS), primarily determined by school location, may not be readily feasible for this purpose.

TABLA DE DESCRIPTIVOS

Building upon this production framework, we will employ the technique described in this paper, which combines ML techniques for classification and DEA, to determine a robust technical efficiency analysis. This approach allows us to capture the complex intricacies and idiosyncrasies of the educational sector in Spain, providing a more accurate and contextualized perspective efficiency.

RESULTADOS Y TODOS SUS COMENTARIOS (SIN hablar de hiperparametros, etc, etc con mucho detalle, solo se mencionan de pasada)

Finally, it is worth mentioning that our integration of Machine Learning with Data Envelopment Analysis may be also used to extrapolate efficiency assessments to unseen data, such as schools not included in the initial PISA sample. This capability is particularly valuable in educational policy making, where decision-makers need to predict and evaluate the efficiency of organizations that were not part of the (random) data sample that was used in the original study. In particular, our method utilizes classification models trained on known PISA data to establish a predictive framework that can assess whether an unseen school would likely operate efficiently or not based on its inputs, outputs and context variables. In cases where a school is predicted to be inefficient, our model not only quantifies the level of inefficiency but also provides specific output targets that the school needs to achieve to be considered efficient. Moreover, this predictive ability enhances the practical utility of standard DEA by extending its applicability beyond the traditional analysis of existing units to include even potential future or hypothetical units. Such a predictive model is instrumental for educational authorities as it allows for proactive rather than reactive measures in resource allocation and policy planning. By enabling the evaluation of schools outside the observed dataset, our approach offers a robust tool for continuous improvement and strategic planning in education systems.

5. Conclusions and future work

After examining existing literature, it is clear that a growing number of researchers are focusing on the combined use of ML-DEA methodologies to predict organizational efficiency across various sectors. Although many of these studies focus on utilizing these methodologies to explore the interplay between machine learning enhancements and traditional DEA approaches, our research introduces a new dimension by integrating classification models with DEA. This fusion is not merely theoretical but also practically applicable, as demonstrated through our empirical study using PISA data. Our findings underscore that integrating ML classifiers with DEA not only helps in predicting the efficiency status of decision-making units (or even unseen data) but also in refining the evaluation process of observations by introducing new judgment elements into the nature of traditional DEA assessments.

The advantages of our integrated approach extend beyond just analytical improvements. They also offer practical benefits in terms of scalability and adaptability. The model's ability to handle large datasets efficiently makes it especially relevant in the era of big data, where organizations across sectors are looking to leverage vast amounts of information for enhanced decision-making. Additionally, the flexibility of the ML-DEA framework means it can be tailored to specific sector needs, whether it be healthcare, education, or finance, providing customized efficiency assessments that are both insightful and actionable.

The integration of Machine Learning models with Data Envelopment Analysis represents a compelling advancement in the realm of efficiency analysis, offering a more nuanced understanding and interpretability of the results through variable importance ranking. This synthesis not only enhances traditional DEA by addressing its limitations—such as handling nonlinearity and model overfitting—but also leverages the computational prowess of ML to uncover intricate patterns and relationships within data that are otherwise not discernible. By employing ML techniques, particularly classification models, alongside DEA, we can effectively rank inputs, outputs, and environmental variables in terms of their impact on efficiency scores. This ranking is crucial for decision-makers as it identifies key performance drivers, enabling targeted improvements and resource allocation. The incorporation of ML thus empowers organizations to not only measure efficiency but also to understand the underlying factors contributing to inefficiency, facilitating strategic interventions that are both precise and impactful.

Compared to other methods, the integrated ML-DEA approach brings several distinct advantages:

1. Improved Accuracy and Robustness: The integration of ML algorithms enhances the robustness of the DEA model by enabling it to handle outliers and noise effectively.

2. Enhanced Interpretability: By employing explainable AI techniques, particularly the use of counterfactual explanations within the ML-DEA framework, our method not only quantifies efficiency but also explains it.

3. Flexibility and Customization: The modular nature of our approach allows for the integration of any classification ML technique, depending on the specific characteristics of the dataset and

analytical needs. This adaptability ensures that the model remains relevant across different applications and evolves alongside advancements in machine learning.

In conclusion, the new integration of ML with DEA models could represent a significant advancement in the field of efficiency analysis. Its ability to provide detailed, reliable, and actionable efficiency assessments could make it a valuable tool for researchers and practitioners alike. Ultimately, the true value and relevance of our contribution in the field of efficiency evaluation will be determined by its future application across diverse datasets and contexts, which will validate or challenge the robustness and adaptability of our approach.

Looking forward, several research avenues appear promising. First, the exploration of other machine learning techniques, such as ensemble methods (e.g., Random Forest or Boosting), could provide further improvements in the robustness and accuracy of efficiency predictions. These techniques, known for their effectiveness in capturing nonlinear relationships and high-dimensional data interactions, could be tailored to complement DEA's framework, potentially leading to more nuanced and detailed efficiency analyses. Secondly, the application of our integrated ML-DEA model to other domains, such as environmental sustainability and other public sector performance, could be highly beneficial. These areas, where efficiency and resource optimization are critical, may significantly benefit from the enhanced analytical capabilities that our model offers. Additionally, extending our model to handle real-time data could transform operational efficiency monitoring, allowing organizations to make immediate adjustments based on current performance metrics. Lastly, further research should also focus on the development of more sophisticated counterfactual methods within the ML-DEA framework. These methods would not only enhance the interpretability of the model outcomes but also allow decision-makers to perform scenario analysis and policy testing effectively. Such developments could make ML-DEA an indispensable tool in strategic planning and resource management, especially in sectors where efficiency gains translate directly into improved outcomes for stakeholders and the environment.

Acknowledgments

The authors thank the grant PID2022-136383NB-I00 funded by MCIN/AEI/ 10.13039/501100011033 and by ERDF A way of making Europe. R. Gonzalez-Moyano XXXX. V. España thanks the PhD scholarship ACIF/2021/135 supported by the Conselleria d'Educació,

Universitats i Ocupació (Generalitat Valenciana). Additionally, J. Aparicio thanks the grant PROMETEO/2021/063 funded by the Valencian Community (Spain).

References

Esteve, M., Aparicio, J., Rabasa, A., & Rodriguez-Sala, J. J. (2020). Efficiency analysis trees: A new methodology for estimating production frontiers through decision trees. Expert Systems with Applications, 162, 113783.

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

Seiford, L. M., & Zhu, J. (2002). Modeling undesirable factors in efficiency evaluation. European Journal of Operational Research, 142(1), 16-20.

Olesen, O. B., Petersen, N. C., & Podinovski, V. V. (2007). Staff assessment and productivity measurement in public administration: an application of data envelopment analysis. Omega, 35(3), 297-307.

Zhou, P., Ang, B. W., & Poh, K. L. (2008). A survey of data envelopment analysis in energy and environmental studies. European Journal of Operational Research, 189(1), 1-18

Charles, V., Aparicio, J., & Zhu, J. (2019). The curse of dimensionality of decision-making units: A simple approach to increase the discriminatory power of data envelopment analysis. European Journal of Operational Research, 279(3), 929-940.

Banker, R. D., & Morey, R. C. (1986). Efficiency analysis for exogenously fixed inputs and outputs. Operations Research, 34(4), 513-521.

Thanassoulis, E., Boussofiane, A., & Dyson, R. G. (2015). Applied data envelopment analysis. Springer.

Vapnik, V., & Cortes, C. (1995). Support-vector networks. Machine learning, 20(3), 273-297.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436-444.

Comentado [JA1]: Mete todas las referencias que faltan por citar aquí y ordena por orden alfabético.

Wachter, S., Mittelstadt, B., & Russell, C. (2017). "Counterfactual Explanations Without Opening the Black Box: Automated Decisions and the GDPR." Harvard Journal of Law & Technology, 31(2), 841-887.

Chen, Y., Li, Y., Xie, Q., An, Q., & Liang, L. (2014). Data envelopment analysis with missing data: a multiple imputation approach. International Journal of Information and Decision Sciences, 6(4), 315-337.

Valero-Carreras, D., Moragues, R., Aparicio, J., & Guerrero, N. M. (2024). Evaluating different methods for ranking inputs in the context of the performance assessment of decision making units: A machine learning approach. Computers & Operations Research, 163, 106485.

Esteve, M., Aparicio, J., Rodriguez-Sala, J. J., & Zhu, J. (2023). Random Forests and the measurement of super-efficiency in the context of Free Disposal Hull. European Journal of Operational Research, 304(2), 729-744.

Pastor, J. T., Ruiz, J. L., & Sirvent, I. (2002). A statistical test for nested radial DEA models. Operations Research, 50(4), 728-735.

Jin, Q., Kerstens, K., & Van de Woestyne, I. (2024). Convex and nonconvex nonparametric frontier-based classification methods for anomaly detection. OR Spectrum, 1-27.