



Selección de Entradas y Salidas en DEA mediante Algoritmos Genéticos Multiobjetivo

Nelson Fabián Ramírez Upegui

Artículo de investigación presentado para optar al título de Especialista en Analítica y
Ciencia de Datos

Asesor

Juan Guillermo Villegas Ramirez, Doctor (PhD) en Ingeniería

Universidad de Antioquia
Facultad de Ingeniería
Especialización en Analítica y Ciencia de Datos
Medellín, Antioquia, Colombia
2024

Cita numérica

1

Cita nota al pie

¹ Nelson Fabián Ramírez Upegui, “Selección de Entradas y Salidas en DEA mediante Algoritmos Genéticos Multiobjetivo” (Trabajo de grado especialización, Universidad de Antioquia, 2024).

**Fuentes primarias /
Bibliografía**

Ramírez Upegui Nelson. “Selección de Entradas y Salidas en DEA mediante Algoritmos Genéticos Multiobjetivo”. Trabajo de grado especialización, Universidad de Antioquia, 2024.

Estilo: Chicago 17 (2017) y adaptación de Trashumante. Revista Americana de Historia Social UdeA.



Especialización en Analítica y Ciencia de Datos, Cohorte VII.

Centro de Investigación Ambientales y de Ingeniería (CIA).



Centro de Documentación Ingeniería (CENDOI)

Repositorio Institucional: <http://bibliotecadigital.udea.edu.co>

Universidad de Antioquia - www.udea.edu.co

El contenido de esta obra corresponde al derecho de expresión de los autores y no compromete el pensamiento institucional de la Universidad de Antioquia ni desata su responsabilidad frente a terceros. Los autores asumen la responsabilidad por los derechos de autor y conexos.

Selección de Entradas y Salidas en DEA mediante Algoritmos Genéticos Multiobjetivo

Nelson F. Ramírez Upegui
Departamento de Ingeniería de Sistemas
Universidad de Antioquia
Medellín, Antioquia, Colombia
nelson.ramirez1@udea.edu.co

Juan G. Villegas
Análítica e Investigación para la Toma de Decisiones-ALIADO
Departamento de Ingeniería Industrial
Universidad de Antioquia
Medellín, Antioquia, Colombia 

Abstract— This study presents a method for feature selection in datasets that suffer from high dimensionality, aimed at evaluating relative efficiency using the non-parametric technique known as Data Envelopment Analysis (DEA). This application is implemented using genetic algorithms for multi-objective optimization, applying a basic binary structure to select the variables from the dataset that will enter the DEA evaluation. The selection process is carried out through crossover and mutation of an initially randomly selected population from the solution space. Four different well-known multi-objective genetic algorithms are used and compared using data sets from the DEA literature.

Keywords—Feature Selection, Data Envelopment Analysis, Multi-objective Genetic Algorithms.

I. INTRODUCCIÓN

La técnica no paramétrica conocida como Análisis Envoltente de Datos (DEA) permite evaluar la eficiencia relativa de diferentes unidades de decisión llamadas DMUs [2], el DEA cuenta con diferentes orientaciones para la evaluación de eficiencia; las más comunes son la orientación a las entradas o a las salidas, lo que indica cual es el sentido de la optimización que se desea evaluar, bien sea minimizar los insumos a utilizar o maximizar los productos a generar [3]. En el presente estudio se implementó DEA para la evaluación de eficiencia relativa mediante la librería `amplpy` [4] bajo la cual se puede seleccionar entre tres (3) tipos de modelo a evaluar: en el espacio de los multiplicadores con retornos constantes a escala (CCR), en el espacio de la envoltente (ENV) y en el espacio de la envoltente con retornos a escala variable (BCC), adicionalmente se puede elegir la orientación del modelo, a las entradas (I), o a las salidas (O).

Para la ejecución del DEA suele tenerse en cuenta una regla empírica que se fundamenta en tener al menos entre dos (2) y tres (3) veces la cantidad de DMUs frente a las variables de entrada (insumos) y salida (productos) para que el poder discriminatorio en la evaluación de eficiencia relativa de DEA sea adecuado [5], [6].

Para el caso del modelo DEA como se mencionó previamente, no cumplir la regla empírica para la cantidad de

variables (entradas y salidas) respecto a la cantidad de DMUs genera dentro del modelo un “problema” de “alta dimensionalidad” en el conjunto de datos [5]. Siendo este el inicio para postular la necesidad de aplicar la selección de características, para la elección de subconjuntos de entradas y salidas más representativos que den origen a un nuevo submodelo de DEA que cuente con un resultado que brinde una “mejor” discriminación de eficiencia relativa en el submodelo en comparación con un modelo original que use todas las entradas y salidas disponibles [5], [7].

Para la selección de características sobre el conjunto de datos que no cumple con la regla empírica en DEA, se realizó la implementación de metaheurísticas para la optimización multiobjetivo basadas en algoritmos genéticos. Las cuales mediante operadores de cruce, mutación y selección sobre una representación binaria generan subconjuntos de variables [8] para la evaluación del modelo DEA. Los algoritmos genéticos se codificaron mediante la librería `pymoo` [9] y los algoritmos implementados en el presente estudio fueron NSGA-II [10], NSGA-III [11], SMS-EMOA [12] y AGE-MOEA [13].

Lo mencionado anteriormente se realizó para evaluar tres (3) casos ya revisados en la literatura que no cumplen con la regla empírica para DEA con el fin de validar los resultados de la aplicación en la selección de características realizada para el presente artículo. El primero proviene de un método constructivo para la selección de características en conjuntos de datos para DEA que fue utilizado para evaluar la eficiencia de algunas EPS en Colombia [10], en el segundo se obtuvo un conjunto de datos en los que se evaluó la eficiencia relativa sobre refinerías de petróleo iraquís [30] y por último se evaluó el conjunto de datos sobre la calidad de vida en algunas ciudades de EE.UU. evaluado en [5].

II. REVISIÓN DE LA LITERATURA

A. Análisis Envoltente de Datos (Data envelopment Analysis, DEA)

El Análisis Envoltente de Datos (DEA por sus siglas en inglés) es una técnica no paramétrica (que no se encuentra asociada a una distribución estadística [14]) desarrollada por Charnes, Cooper y Rhodes. Que permite evaluar la eficiencia relativa de diferentes unidades para la toma de decisiones (DMU), la cual se obtiene del máximo de la relación entre las

múltiples entradas y salidas ponderadas que componen las DMU, sujeta a la condición de que las eficiencias para cada DMU deben ser menores o iguales a uno (1), de dónde una DMU se considera eficiente si su eficiencia relativa es igual a uno (1) e ineficiente para cualquier valor menor a uno (1). Como se puede ver en el modelo original de DEA formulado por Charnes, Cooper y Rhodes [1], [2]

$$\max: h_0 = \frac{\sum_{r=1}^s u_r y_{r0}}{\sum_{i=1}^m v_i x_{i0}} \quad (1)$$

Sujeto a:

$$\frac{\sum_{r=1}^s u_r y_{rj}}{\sum_{i=1}^m v_i x_{ij}} \leq 1; \quad j = 1, \dots, n, \quad (2)$$

$$u_r, v_i \geq 0; \quad r = 1, \dots, s; \quad i = 1, \dots, m.$$

El modelo (1)-(2) tiene una formulación no lineal que se conoce como el modelo CCR (por las siglas de sus creadores Charnes, Cooper y Rhodes). Este modelo de DEA es el más básico y se compone de las entradas x_{ij} , las salidas y_{rj} , los pesos de las entradas v_i y los pesos de las salidas u_r [1], [2].

Asimismo, Charnes, Cooper y Rhodes postularon una transformación del modelo fraccional con el fin de obtener un problema de programación lineal ordinario [1], [2] planteado en la siguiente transformación:

$$\max: h_0 = \sum_{r=1}^s u_r y_{r0} \quad (3)$$

Sujeto a:

$$\sum_{i=1}^m v_i x_{i0} = 1; \quad (4)$$

$$\sum_{r=1}^s u_r y_{rj} \leq \sum_{i=1}^m v_i x_{ij}; \quad j = 1, \dots, n, \quad (5)$$

$$\text{Con, } u_r, v_i \geq 0; \quad r = 1, \dots, s; \quad i = 1, \dots, m.$$

A este modelo se le conoce como CCR en el espacio de los multiplicadores, ya que las variables de decisión v_i y u_r son los pesos de las entradas y salidas en la evaluación de cada DMU. Con lo anterior en mente y aplicando el problema dual de la programación lineal, se muestra a continuación:

$$\min: \theta \quad (6)$$

Sujeto a:

$$\sum_{j=1}^n \lambda_j x_{ij} \leq \theta x_{i0}, \quad i = 1, \dots, m \quad (7)$$

$$\sum_{j=1}^n \lambda_j y_{rj} \geq y_{r0}, \quad r = 1, \dots, s \quad (8)$$

$$\text{Con, } \lambda \geq 0, \quad j = 1, \dots, n$$

El anterior modelo es llamado CCR-I en el espacio de la envolvente [3], la I es la indicación de la orientación a las entradas, que busca minimizar el valor de las entradas

obteniendo al menos los valores de salida dados inicialmente por la DMU que se está evaluando.

Los modelos mostrados hasta este punto se asumen bajo un comportamiento de *retornos constantes a escala* (CRS), pero posteriores estudios dieron origen al modelo BCC (Banker-Charnes-Cooper) [3] el cual comprende los *retornos a escala variable* (VRS). Para el modelo postulado en las ecuaciones (6)-(8) se agregó la siguiente restricción:

Sujeto a:

$$\sum_{j=1}^n \lambda_j = 1, \quad (9)$$

Para el modelo de eficiencia relativa en el espacio de los multiplicadores (3)-(5) los pesos óptimos u, v , dan cuenta de la importancia de las entradas y salidas para la eficiencia de cada DMU. Mientras que los pesos λ evaluados en el modelo en el espacio de la envolvente (6)-(8) están relacionados con la combinación lineal de las DMU eficientes para formar la frontera eficiente con las que se miden todas las DMUs [3].

El estudio de los modelos de DEA en el presente artículo se motiva por su gran uso para la evaluación de eficiencia relativa en diversos sectores económicos como agencias del sector público, empresas del sector salud, educativo, transporte, banca, manufactura, energía, etc. [15], [3].

Es importante tener presente que el DEA se ve afectado en su capacidad de discriminación entre DMUs eficientes y no eficientes, cuando existen muchas variables para las entradas y/o para las salidas. En el DEA existen algunas reglas empíricas que indican que, para evitar los problemas de evaluación de eficiencia relativa, la cantidad de DMUs que se evalúen deben ser entre dos y tres veces la cantidad de variables utilizadas (entradas y salidas), Pero cuando obtener la cantidad de DMUs necesarias para “cumplir” con la regla empírica no es posible o es demasiado costoso; es necesario evaluar si todas las variables tanto entradas como salidas que se están evaluando son necesarias en el modelo. Lo anterior es útil para evitar la llamada maldición de la dimensionalidad [5], [6] (la cual se ampliará posteriormente) que afecta a los modelos que requieren un conjunto de datos para realizar su evaluación (lo anterior aplica tanto para DEA como para otros modelos utilizados en la ciencia de datos). Por lo que en el presente artículo se presentan los algoritmos genéticos multiobjetivo como alternativa para la selección de las características representativas (entradas y salidas) con las cuales evaluar la discriminación de eficiencia relativa en los modelos de DEA.

B. Selección de Características (Feature Selection)

Varios autores muestran que la selección de características (en adelante FS por sus siglas en inglés), es una técnica muy utilizada por la ciencia de datos y la minería de datos, con el fin de elegir el *subconjunto más representativo* del total de las *variables o características* de los datos [16]. Lo anterior se muestra dado que en un conjunto que adolece de alta

dimensionalidad (gran número de variables), se generan problemáticas que se traducen en *alto costo computacional* o en el requerimiento del *uso de memoria* [6]. Así mismo, en DEA también aparece la problemática conocida como *maldición de la dimensionalidad*, que afecta la evaluación de eficiencia relativa. Como se define en [17] a *“medida que aumenta el número de variables explicativas, el problema de descubrir la estructura se vuelve más difícil”*. Por lo anterior, los datos que presentan alta dimensionalidad en su estructura pueden llegar a tener variables o características que no son relevantes, que pueden ser contradictorias o también ser redundantes, denotando esto que el espacio de búsqueda sea más grande [6] con lo que puede generar mayor tiempo en la ejecución de modelos o que no se obtengan resultados válidos.

Para el caso del modelo DEA como se mencionó previamente, no cumplir la regla empírica para la cantidad de variables (entradas y salidas) respecto a la cantidad de DMUs genera dentro del modelo un *“problema”* de *“alta dimensionalidad”* en el conjunto de datos [5]. Siendo este el inicio para postular la necesidad de aplicar FS, para la selección de los subconjuntos de entradas y salidas más representativos que den origen a un nuevo submodelo de DEA que mantenga la variabilidad dentro de los datos originales y adicionalmente que cuente con un resultado que brinde una *“mejor”* discriminación de eficiencia relativa en el submodelo en comparación con un modelo original que use todas las entradas y salidas disponibles [5], [7].

C. Algoritmos Genéticos (GA)

La optimización multiobjetivo (MOO por sus siglas en inglés) busca solucionar problemas de optimización con dos o más objetivos que *“compiten”* entre sí, cumpliendo con las restricciones que tenga el espacio de solución de dicho problema [18]. La cuestión principal para la MOO se presenta al momento de evaluar la calidad de las soluciones generadas; como se menciona en [19] al *“evaluar los resultados para este tipo de funciones se debe lidiar con un conjunto de soluciones en un espacio objetivo multidimensional”*. Lo cual hace en muchas ocasiones sea imposible brindar una única solución que sea óptima para el problema que se desea valorar.

Inicialmente para contrastar las soluciones en un ámbito multiobjetivo, es importante conocer el término *“dominancia”* de las soluciones, en la cual se deben cumplir dos condiciones para asegurar que una solución A domina a B:

- i. La solución A es *“mejor”* o *“igual”* que B en todos sus objetivos $f_i(\cdot)$
 - a. En maximización: $f_i(A) \geq f_i(B) \forall i$
 - b. En minimización: $f_i(A) \leq f_i(B) \forall i$
- ii. La solución A es estrictamente mejor que B, en al menos un objetivo del espacio de soluciones.

Para lo anterior se dice entonces que A domina a B ($A \preceq B$) si cumple las dos condiciones anteriores [18], [8] y si una

solución domina a otra se puede decir *“mejor”* en el conjunto de soluciones.

Por otro lado, el evaluar las soluciones bajo el concepto de la *“optimalidad de Pareto”* dará luces sobre las soluciones *“eficientes”* en el subconjunto evaluado. La optimalidad de Pareto se refiere a que, dentro de un espacio de búsqueda, un punto es óptimo si no existe otro que mejore el valor de al menos una función objetivo, sin degradar la de otra. Estos puntos que cumplen la optimalidad de Pareto se denominan *“soluciones no dominadas o eficientes”* [18], [8].

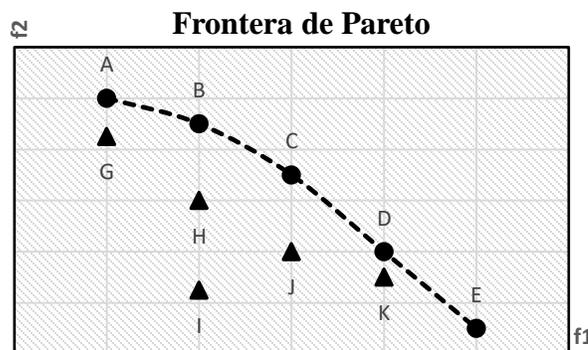


Fig. 1. Ilustración de puntos que satisfacen la optimalidad de Pareto para un problema de maximización de dos funciones objetivo

Tal como se menciona en [18], existen dos tipos de metaheurísticas principales para la MOO, los algoritmos evolutivos/genéticos y los algoritmos basados en enjambre. Para el presente artículo se empleará la técnica evolutiva, la cual incorpora conceptos de la *“evolución natural”*, y a una población inicial de *“individuos”*, mediante el uso de los operadores de *cruce* (generación de descendientes) y *mutación* (cambios aleatorios de los individuos que se van dando a lo largo de las generaciones), va generando nuevas posibles soluciones dentro del espacio de solución del problema de optimización [9].

Existen diversos algoritmos evolutivos planteados para MOO, en el presente artículo se evaluarán los siguientes algoritmos genéticos (en adelante AGs): NSGA-II [10], NSGA-III [11], SMS-EMOA [12] y AGE-MOEA [13].

Según K. Deb [8], para definir la estructura básica de un algoritmo genético, se requieren los siguientes cinco elementos [8], [9]:

- i. Formar una población inicial dentro del espacio muestral de las posibles soluciones,
- ii. Establecer cómo evaluar el valor de aptitud (fitness) de las soluciones objetivo (¿qué métricas de desempeño se usarán?),
- iii. Generar un patrón de reproducción de los individuos, normalmente duplicar las mejores soluciones (los *“más aptos”*), con el fin de eliminar *“malas soluciones”* en la población

- iv. Incluir un operador de cruce, el cual se encarga de tomar varias soluciones (usualmente llamados padres) para combinar “*características*” de estos y crear un nuevo “*descendiente*” o hijo que genere diversidad a la población (para intentar que se puedan evaluar puntos no vistos en el espacio muestral).
- v. Introducir un patrón de mutación en la población, que genere cambios “*inesperados*” o aleatorios en las descendencias, para guiar al algoritmo a regiones del espacio de solución que no se han tenido en cuenta y así diversificar la búsqueda para que el algoritmo pueda salir de óptimos locales.

En la siguiente sección se describirán los elementos necesarios para adaptar los algoritmos genéticos multiobjetivo antes mencionados al FS en DEA. En particular, para realizar la selección de características o variables representativas dentro del conjunto de datos para la evaluación de eficiencia relativa mediante un modelo DEA que no cumple con la regla empírica para la cantidad de DMUs vs la cantidad variables de entrada y salida (por lo que se pueden presentar problemas en la adecuada discriminación de unidades eficientes), Para adaptar un algoritmo genético es necesario tomar algunas decisiones de diseño que detallamos a continuación.

III. ALGORITMOS GENETICOS MULTI OBJETIVO PARA FS EN DEA: DISEÑO E IMPLEMENTACIÓN COMPUTACIONAL

A lo largo de esta sección, se tomará a modo de ejemplo un conjunto de datos ya revisado en la literatura [20], [21], con el ánimo de mostrar más claramente la implementación realizada para el estudio.

A. Evaluación de eficiencia relativa mediante DEA y representación de las soluciones en los AGs

El conjunto de datos en el modelo DEA viene dado por dos tipos de variables, entradas y salidas, que representan los insumos necesarios (entradas), para obtener unos productos específicos (salidas) [1], [2]. Como se mencionó antes, el método de cálculo de la eficiencia relativa aplica para cualquier tipo de negocio [15], dando esto la posibilidad de obtener un conjunto de datos de “*insumos*” y “*productos*” prácticamente para cualquier tipo de organización.

Dentro del presente artículo, se plantea una representación simbólica de las “*Columns*” del conjunto de datos, mediante un vector binario de longitud igual a la cantidad de variables de entrada y salida del conjunto de datos a evaluar, en el que un valor de uno (1) en el vector binario, representa que la variable se toma para el modelo y cero (0) indica que esa variable no será tomada en cuenta dentro de la evaluación de eficiencia en DEA.

Los datos que se tomarán para ilustrar el proceso realizado en el presente artículo se muestran en la Tabla 1 y pertenecen al estudio realizado en [20] y [21].

TABLA 1. CONJUNTO DE DATOS PARA ILUSTRAR EL USO DE FS EN DEA.

DMUs	Entradas			Salidas		
	Empleados Docentes	Salarios Docentes	Salarios Empleados de Apoyo	Estudiantes de Pregrado	Estudiantes de Posgrado	Numero de Artículos de Investigación
1	12	400	20	60	35	17
2	19	750	70	139	41	40
3	42	1500	70	225	68	75
4	15	600	100	90	12	17
5	45	2000	250	253	145	130
6	19	730	50	132	45	45
7	41	2350	600	305	159	97

La colección de datos para el modelo DEA, consta de tres (3) entradas y tres (3) salidas, con siete (7) DMUs y con un resultado de eficiencia como se muestra en la Tabla 2 si se tomaran todas las entradas y salidas.

TABLA 2. RESULTADO DE EFICIENCIA MODELO DEA CCR-I.

DMU	Eficiencia
1	1
2	1
3	1
4	0.820
5	1
6	1
7	1

B. Representación de las variables

Como se mencionó arriba, la *representación simbólica* de las variables viene expresada por un vector binario, que para incorporar el uso de todas las variables del ejemplo de la Tabla 1, quedaría de la siguiente forma. (para mayor facilidad se reemplazan los nombres de las variables de entrada por I_i y las de salida por O_j), donde i, j representan la i -ésima entrada y la j -ésima salida de la colección de datos, de la siguiente manera:

$$\begin{array}{c|c|c|c|c|c|c} I_1 & I_2 & I_3 & O_1 & O_2 & O_3 \\ \hline 1 & 1 & 1 & 1 & 1 & 1 \end{array}$$

Si por el contrario en el modelo se realizara la evaluación con las entradas I_2 e I_3 y las salidas O_1 y O_2 , el vector de las variables quedaría así:

$$\begin{array}{c|c|c|c|c|c|c} I_1 & I_2 & I_3 & O_1 & O_2 & O_3 \\ \hline 0 & 1 & 1 & 1 & 1 & 0 \end{array}$$

Esta es la forma como los distintos AGs para MOO implementados reciben las variables a evaluar en cada individuo de la población.

Cabe aclarar que por ser un proceso de optimización de “*insumos*” y “*productos*”, siempre debe existir por lo menos una *entrada* para obtener al menos una *salida*, o como se menciona en el principio de eficiencia de Pareto “*los recursos no pueden ser redistribuidos de manera que mejore el bienestar de un agente económico sin perjudicar el bienestar de otro agente*”[22], en otras palabras y abstrayendo al caso de eficiencia que atañe al presente artículo, no se puede obtener un producto o servicio sin emplear algún insumo a cambio. Conforme a lo anterior, una de las necesidades para mantener la factibilidad de las soluciones dentro del modelo DEA, es que se

debe garantizar en el conjunto de variables la existencia de al menos una *entrada* y una *salida*, esto se logra al incluir en el código de la evaluación del algoritmo una “*penalización*” con un valor “*extremo*” si la cantidad de entradas o salidas se vuelven cero (0), lo que genera que dicha solución no sea deseable al evaluar la función objetivo.

C. Métricas para la evaluación de los modelos DEA (individuos de la población)

Con el fin de no aumentar la complejidad en el cálculo de las funciones objetivo de los AGs para MOO se configuraron las métricas de tal forma que se pueda mantener un mismo sentido en la optimización (maximización o minimización), buscando con esto, que el algoritmo no tenga dificultades de convergencia en la búsqueda de soluciones óptimas debido a la competencia entre objetivos que entran en conflicto debido al enfoque de la optimización.

Para la evaluación del desempeño de los AGs implementados en el presente estudio, se utilizaron métricas que tengan en cuenta los siguientes aspectos que se desean mantener en un modelo DEA.

- Conservar la discriminación de la eficiencia relativa para las DMUs, y
- Mantener una alta retención de la varianza en los datos, con el fin de reducir la pérdida de la información contenida en la colección de datos utilizada.

Al revisar las métricas que ayudan a cumplir los criterios mencionados arriba, para conservar el sentido en la optimización, se eligió mantener la maximización como orientación de las métricas, por lo cual se explicará para cada una cómo se logra esto.

1) Relación de DMUs ineficientes.

Como se mencionó arriba, se desea mantener la evaluación para maximización, es por esto que obtener una métrica para la eficiencia de las DMUS en dicho sentido, se deben buscar mayores valores de eficiencia relativa que sean menores a uno (1), es decir más DMUs ineficientes [1], [2]. Es por ello por lo que se mide la proporción de DMUs ineficientes sobre la cantidad total de DMUs evaluadas como se muestra en (10).

$$\% DMUs Ineff = \frac{DMUs Ineficientes}{Total DMUs Evaluadas} \quad (10)$$

2) Índice Gini Ponderado para Eficiencia e Ineficiencia.

Como se indica en [23], el índice Gini, se suele utilizar para cuantificar el nivel de desigualdad que existe en la distribución de las variables, en este caso sobre las eficiencias e ineficiencias obtenidas.

Para la aplicación del Coeficiente de desigualdad de Gini (CG), también conocido como índice Gini, se evaluará conforme a lo expuesto en [24], en el cual el índice se construye mediante la *distribución empírica que siguen los datos*. Para este caso como se mencionó, las eficiencias e ineficiencias obtenidas en el modelo de DEA y la línea de distribución teórica procedente de la *curva de Lorenz*.

La ecuación para el índice de Gini utilizado en el presente artículo es la que se indica en [24] para datos no agrupados.

$$CG = \frac{\sum_{i=1}^{n-1} (P_i - Y_i)}{\sum_{i=1}^{n-1} P_i} \quad (11)$$

En dónde P_i es la proporción acumulada de los datos de eficiencias e ineficiencias de DEA y Y_i es la proporción acumulada teórica fundamentada en la *curva de Lorenz*.

A continuación, se muestra la representación gráfica del índice Gini para los resultados de eficiencia e ineficiencia para el ejemplo del presente artículo.

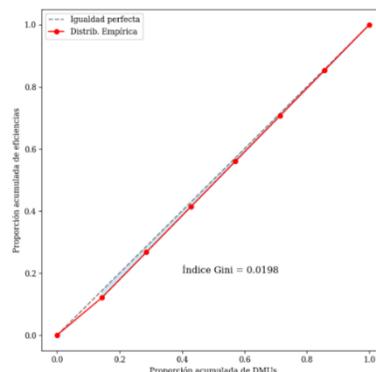


Fig. 2. Representación gráfica del índice Gini.

La Figura 2. Muestra que la distribución de eficiencias e ineficiencias se encuentra muy cerca de la distribución de igualdad perfecta (*Curva de Lorenz*) dado que seis (6) de las siete (7) DMUs disponibles en el ejemplo son eficientes, es por lo anterior, que la distribución de desigualdad se acerca al valor cero (0) con un resultado de $CG = 0.0198$

Si por el contrario se tienen seis (6) DMUs ineficientes y una (1) sola eficiente, al contrastarla frente a la *curva de Lorenz*, la distribución de desigualdad se vería como se muestra en la Figura 3.

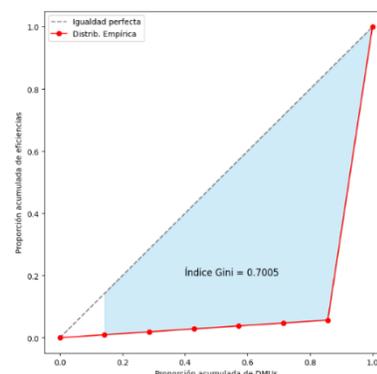


Fig. 3 Gráfica del índice Gini con más DMUs ineficientes que eficientes

Por su parte, como se mencionó anteriormente, una DMU es ineficiente para cualquier valor menor a uno (1). Teniendo esto en mente y al efectuar una comprobación con

resultados de ineficiencia cercanos a uno (1), la distribución de desigualdad de Gini resulta muy similar al obtenido en [21], con el vector de eficiencias/ineficiencias de la Tabla 2. Para ilustrar los anteriormente mencionado, mediante el siguiente vector de eficiencias/ineficiencias [0.90, 0.91, 0.92, 0.93, 0.94, 0.95, 1] que muestra una sola DMU eficiente y las demás ineficientes con valores cercanos a uno (1) da un resultado de $CG = 0.0153$, como puede observarse en la Figura 4.

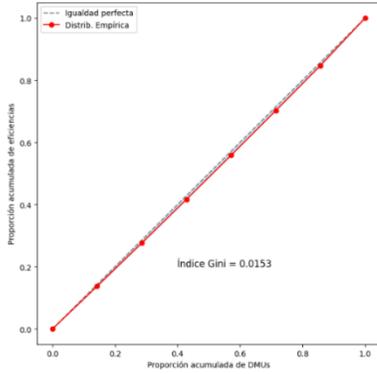


Fig. 4. Gráfica del índice Gini con valores de Ineficiencia cercanos a uno (1)

Para evitar valores de CG muy cercanos para resultados de eficiencia/ineficiencia relativa en DEA, se opta por utilizar la métrica del Gini ponderado según se muestra en [7]. La cual se representa de la siguiente forma.

$$\alpha(\text{Gini eficiencia}) + (1 - \alpha)\text{Gini ineficiencia} \quad (12)$$

De donde al evaluar un valor de $\alpha = 0.5$, se obtiene la ecuación del valor medio para el Gini de las eficiencias y el Gini de las ineficiencias como se observa en (13), evitando así que con resultados de ineficiencia cercanos a uno (1) no se tome en cuenta por su similitud con un resultado de eficiencia relativa que no discrimina adecuadamente las DMUs eficientes.

$$\text{Gini Pond} = \frac{\text{Gini eficiencia} + \text{Gini ineficiencia}}{2} \quad (13)$$

3) Varianza Retenida en el Conjunto de Datos.

Mantener una alta retención de la variabilidad original de los datos, es uno de los principales objetivos del estudio al realizar FS a un conjunto de datos en el que se desea evaluar la eficiencia relativa de las DMUs mediante DEA. Lo anterior se efectúa mediante la evaluación de variables altamente correlacionadas que estén generando redundancia en la información contenida en los datos por lo que solo sería necesario tener alguna de las variables correlacionadas en cuenta para la evaluación de eficiencia.

Es por lo anterior que se aplica el acercamiento estadístico para la reducción de variables en DEA mediante el uso del análisis de covarianza parcial aplicado en [25], [7]. Si el conjunto de datos a revisar se toma como una matriz, el procedimiento para evaluar la covarianza parcial sería el siguiente:

- i. El primer paso con el fin de tratar todas las variables de la misma manera es normalizar los datos de cada columna de la “matriz” original para obtener varianzas = 1 y media = 0 (la matriz con los datos normalizados se denotará como \mathbf{X}).
- ii. El paso siguiente es calcular la matriz de varianza-covarianza derivada de las m columnas de la matriz \mathbf{X} .

$$\text{var}(\mathbf{X}) = \mathbf{V} = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1m} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{m1} & \sigma_{m2} & \cdots & \sigma_{mm} \end{bmatrix} \quad (14)$$

La covarianza entre dos vectores de variables X_i , $X_{i'}$ viene dada por:

$$\text{COV}(X_i, X_{i'}) = E\{[X_i - E(X_i)][X_{i'} - E(X_{i'})]\} \quad (15)$$

$$= E(X_i X_{i'}) - E(X_i)E(X_{i'}) \quad (16)$$

$$= \sigma_{ii'} \quad (17)$$

Dado lo anterior es importante anotar, qué la covarianza parcial entre dos variables i e i' y eliminando el efecto de una tercera variable i'' , viene dado por:

$$\sigma_{ii':i''} = (\sigma_{ii'} - \sigma_{ii''}\sigma_{i'i''}/\sigma_{i''i''}^2) \quad (18)$$

- iii. Realizar una separación de las variables de la matriz \mathbf{X} en dos conjuntos. En los que $i = 1, \dots, p$ son las variables por omitir, y $i = p + 1, \dots, m$ (con m igual al número de variables de entrada y de salida), son las variables por mantener, que representarían la mayor cantidad de información de las m variables del conjunto total. Ahora la matriz de varianza-covarianza denotada como \mathbf{V} sería como se muestra a continuación:

$$\mathbf{V} = \begin{bmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{bmatrix} \quad (19)$$

Dónde V_{11} representa la matriz de varianza-covarianza de las variables $i = 1, \dots, p$ y V_{22} , representa la matriz de varianza-covarianza de las variables $i = p + 1, \dots, m$

- iv. Calcular la matriz de varianza-covarianza parcial $\mathbf{V}_{11.2}$ como se muestra en (20):
- v.
$$\mathbf{V}_{11.2} = \mathbf{V}_{11} - \mathbf{V}_{12}\mathbf{V}_{22}^{-1}\mathbf{V}_{21} \quad (20)$$
- vi. Por último, del valor total de la varianza del conjunto de datos, se resta el resultado obtenido para la matriz $\mathbf{V}_{11.2}$ y se calcula el peso sobre la varianza total como se observa en (21):

$$(m - \mathbf{V}_{11.2})/m \quad (21)$$

Lo anterior se puede interpretar como la varianza que queda de la variable i cuando se elimina el efecto de la variable i'

La métrica de la varianza retenida implementada en el presente estudio, genera un valor entre $[0, 1]$, que comprende la relación de la varianza explicada por las variables que se toman para la evaluación del modelo ($i = p + 1, \dots, m$) y el total de variables, dónde un valor de V cercano a uno (1) indica que las variables elegidas en el modelo tienen una variabilidad similar a tomar un modelo con todas las variables.

4) Indicador de Hipervolumen para evaluar las soluciones no dominadas de los AGs.

El indicador de hipervolumen (HV) como se menciona en [26] es uno de los más utilizados para evaluar la calidad de un conjunto de optimizadores multiobjetivo. Y es una medida de calidad para soluciones no dominadas de las funciones de un algoritmo de MOO [27].

En [27] se muestra como el indicador de HV, consiste en medir la región que está simultáneamente dominada por P y acotada en su parte superior por un punto de referencia $r \in \mathbb{R}^d$, tal que $r \geq (\max_p p_1, \dots, \max_p p_d)$, donde $p = (p_1, \dots, p_d) \in P \subset \mathbb{R}^d$ y la relación \geq se aplica componente a componente.

Asimismo, en [27] se ejemplifica visualmente el HV, la cual consiste en una región de un polígono ortogonal unida mediante n hiper rectángulos con el vértice común en el punto de referencia r .

Para un plano con dos (2) objetivos y tres (3) hiper rectángulos con puntos de soluciones no dominadas $P^{(1)}, P^{(2)}, P^{(3)}$ como se observa en la Figura 5.

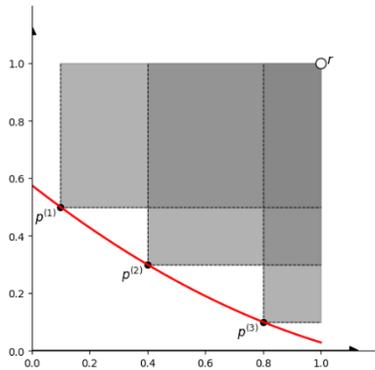


Fig. 5. Indicador de hipervolumen para un caso de dos objetivos

Conforme a lo evaluado en esta sección, para la selección del “mejor” AG dentro de los evaluados en el presente artículo, se utiliza el indicador de HV implementado mediante la librería pymoo [9]. Por otra parte, el HV cumple con la “dominancia” de las soluciones de Pareto por lo cual al valorar el HV, la solución que tenga un resultado mayor a los demás se considera que está más cercana al frente óptimo

de Pareto [28], siendo esta la “mejor” opción del conjunto de soluciones.

D. Implementación en Python

Los modelos DEA fueron implementados haciendo uso del software de modelación algebraica AMPL en Python con la librería amplpy [4]. Por su parte, los algoritmos genéticos se implementaron haciendo uso de la librería de Python pymoo [9].

1) Evaluación de eficiencia relativa en DEA (librería amplpy).

Como se mencionó anteriormente, para comparar los resultados frente del caso ilustrativo analizado en [20] y [21], se realiza en la librería amplpy [4] el código para evaluar la eficiencia relativa mediante DEA del conjunto de datos mostrado en [21], obteniendo los mismos resultados de eficiencia, los cuales se muestran en la Tabla 2 de la Sección III-A.

2) Evaluación de los AGs para MOO (librería pymoo).

Inicialmente para identificar los subconjuntos de las variables seleccionadas al realizar la ejecución de los AGs estructurados para la MOO de las métricas propuestas en la Sección III-C del presente estudio, se muestran los resultados mediante el siguiente formato $[I_1, I_2, I_3, O_1, O_2, O_3]$. Para las entradas se muestra cómo I_j y las salidas como O_j donde $j = 1, \dots, m$.

A continuación, se muestran los resultados obtenidos para cada AG considerado, en el cual se presentan las soluciones no dominadas generadas y una gráfica con el comportamiento de las métricas de desempeño para dichas soluciones.

- NSGA-II.

TABLA 3. SOLUCIONES NO DOMINADAS PARA NSGA-II

Subset Evaluado	% DMUs Ineficientes.	Gini Ponderado	Varianza Retenida Entradas	Varianza Retenida Salidas	Varianza Retenida Total
$[I_2, O_1, O_3]$	57.14%	0.2426	86.45	96.31	95.99
$[I_1, I_2, I_3, O_2, O_3]$	28.57%	0.3879	100	95.65	99.99
$[I_1, I_2, O_1, O_2, O_3]$	14.29%	0.3851	99.92	100	100
$[I_2, O_1, O_2, O_3]$	42.86%	0.2786	86.45	100	98.69

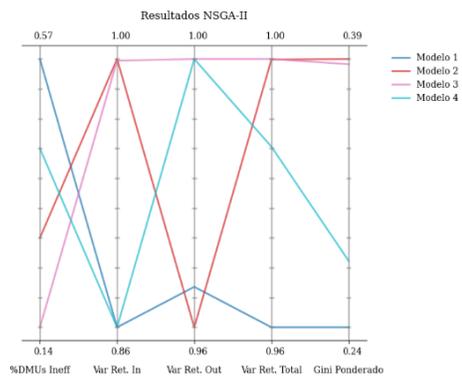


Fig. 6. Comportamiento de las soluciones para NSGA-II

- *NSGA-III*.

TABLA 4. SOLUCIONES NO DOMINADAS PARA NSGA-III

Subset Evaluado	% DMUs Ineficientes.	Gini Ponderado	Varianza Retenida Entradas	Varianza Retenida Salidas	Varianza Retenida Total
$[I_1, I_2, O_2, O_3]$	57.14%	0.3331	99.58	95.65	99.59
$[I_1, I_3, O_1, O_2, O_3]$	14.29%	0.3851	99.92	100	100

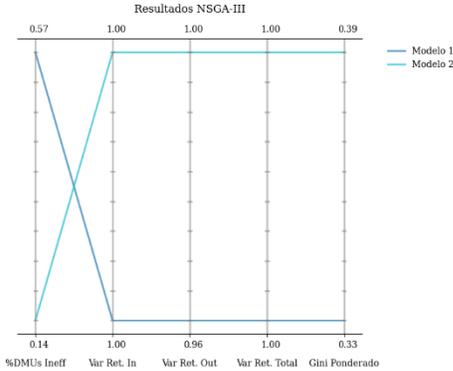


Fig. 7. Comportamiento de las soluciones para NSGA-III

- *AGE-MOEA*.

TABLA 5. SOLUCIONES NO DOMINADAS PARA AGE-MOEA

Subset Evaluado	% DMUs Ineficientes.	Gini Ponderado	Varianza Retenida Entradas	Varianza Retenida Salidas	Varianza Retenida Total
$[I_1, I_3, O_1, O_2, O_3]$	14.29%	0.3851	99.92	100	100
$[I_1, I_2, I_3, O_1, O_3]$	28.57%	0.3680	100	96.31	99.98
$[I_1, I_2, O_1, O_2, O_3]$	28.57%	0.3879	100	95.65	99.99
$[I_2, I_3, O_2, O_3]$	42.86%	0.3521	99.83	95.65	99.60

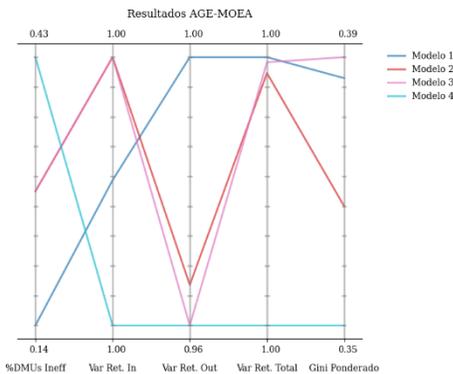


Fig. 8. Comportamiento de las soluciones para AGE-MOEA

- *SMS-EMOA*.

TABLA 6. SOLUCIONES NO DOMINADAS PARA SMS-EMOA

Subset Evaluado	% DMUs Ineficientes.	Gini Ponderado	Varianza Retenida Entradas	Varianza Retenida Salidas	Varianza Retenida Total
$[I_1, I_2, O_2, O_3]$	57.14%	0.3331	99.58	95.65	99.59
$[I_1, O_2]$	85.71%	0.2593	75.15	89.98	93.59
$[I_1, I_3, O_1, O_2, O_3]$	14.29%	0.3851	99.92	100	100
$[I_1, I_3, O_2, O_3]$	28.57%	0.3906	99.92	95.65	99.54

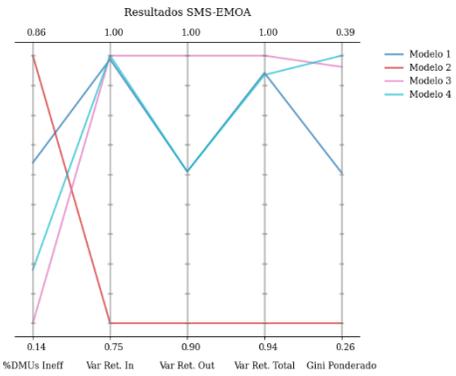


Fig. 9. Comportamiento de las soluciones para SMS-EMOA

3) Evaluación del hipervolumen en los AGs para MOO (librería pymoo).

Para la evaluación del hipervolumen en el presente ejemplo como se puede observar en el punto 4 de la Sección III-C, se obtuvieron los siguientes resultados para los diferentes AGs.

TABLA 7. RESULTADOS DEL HV PARA LOS AG EVALUADOS EN EL CASO DE EJEMPLO

AG Evaluado	Resultado HV
NSGA-II	0.171536
NSGA-III	0.190388
AGE-MOEA	0.156581
SMS-EMOA	0.246001

Conforme a los resultados de HV obtenidos para la evaluación del conjunto de datos del ejemplo muestra que el AG con mayor valor para el indicador de HV es el *SMS – EMOA* = 0.246001. Lo que quiere decir que los modelos resultantes del AG SMS-EMOA tienen una mejor aproximación a la frontera de Pareto.

IV. CASO DE ESTUDIO

Como se ha mostrado a lo largo del artículo, la intención del presente estudio ha sido comprobar la aplicabilidad de implementar la selección de características o FS [6] a un conjunto de datos con variables de entrada y de salida para evaluar eficiencia mediante DEA, que no cumple con la regla empírica de al menos dos (2) y tres (3) veces más DMUs respecto a la cantidad de variables (entradas y salidas)

Por lo anterior se seleccionaron 3 conjuntos de datos evaluados en la literatura [5], [29], [30], con resultados ya demostrados para eficiencia mediante la técnica DEA, que no cumplen la regla empírica [5] (sufren de alta dimensionalidad) y se aplicó la metodología elaborada en el presente estudio con el ánimo de obtener modelos con menos variables para mejorar la discriminación de eficiencia relativa de las DMUs en los conjuntos evaluados.

A. Evaluación de EPS en Colombia

El siguiente conjunto de datos [29] es el primero que se desea evaluar dado que ya existe un proceso para aplicar la selección de características mediante otros métodos de aplicación estadística como PCA, Lasso e ISA-DEA (*iterative selection algorithm*) para evaluar la discriminación de eficiencia en las DMUs mediante DEA [7]. Este trabajo podría considerarse una profundización y un acercamiento alternativo a lo visto en [7].

El conjunto de datos para el caso de estudio de las EPS Colombianas está representado por cuatro (4) entradas, tres (3) salidas y diecisiete (17) DMUs, como se observa a continuación:

TABLA 8. VALORES DE ENTRADAS Y SALIDAS PARA EL MODELO DEA DE EPS COLOMBIA.

Entradas				
EPS/DMUs	Activo Total (I1)	Inversiones (I2)	Total cuentas por cobrar FOSYGA	Activo fijo propiedad planta y equipo
Aliansalud	74.659.289	19.712.425	43.573.089	506.726
Comfenalco Valle	73.192.236	810.002	36.472.629	15.382.635
Compensar	147.183.633	15.127.388	75.399.989	4.243.481
Cooameva	552.127.504	12.757.631	403.192.739	12.549.111
EPM	11.497.792	8.997.307	0.00	0.00
Comfenalco Antioquia	55.003.905	4.500.000	43.468.103	1.489.166
Sura	247.430.381	78.407.570	101.337.602	17.131.377
Famisanar	254.676.457	21.000	176.767.704	8.508.024
Fondo de Pasivo Ferrocarriles	39.370.561	16.297.183	3.739.882	229.932
Golden Group	8.873.885	0.00	1.003.682	686.940
Nueva EPS	1.058.955.212	3.191.139	725.870.151	2.942.152
Colpatria	29.626.063	18.584.149	982.376	6.142
Salud Total	365.396.570	198.238.511	75.315.458	35.242.132
Saludvida	68.974.504	4.785.992	7.520.665	7.904.296
Sanitas	379.950.832	25.376.938	242.987.341	3.000.112
S.O.S.	118.171.938	1.319.225	86.320.782	3.688.966
Solsalud	106.869.758	2.613.716	46.573.124	3.457.365

Salidas			
EPS/DMUs	Ingresos UPC (O1)	Ingresos Recobro al FOSYGA (O2)	Ingresos Operacionales (O3)
Aliansalud	196.384.549	81.430.596	193.387.216
Comfenalco Valle	154.675.269	39.994.397	138.684.343
Compensar	408.946.118	108.062.571	425.911.666
Cooameva	1.412.872.026	384.858.018	1.464.245.760
EPM	11.851.157	0.00	8.712.201
Comfenalco Antioquia	177.047.201	50.863.177	177.934.783
Sura	696.932.272	198.384.339	713.300.668
Famisanar	669.227.215	124.616.077	666.839.628
Fondo de Pasivo Ferrocarriles	54.289.171	0.00	104.784.504
Golden Group	24.376.955	522.175	25.567.417
Nueva EPS	1.903.254.182	788.690.698	1.811.061.005
Colpatria	44.524.895	6.509.171	43.522.988
Salud Total	787.461.401	109.833.981	817.409.281
Saludvida	42.480.528	6.890.123	29.513.895
Sanitas	589.479.187	227.092.530	582.135.254
S.O.S.	376.769.004	89.411.705	359.358.087
Solsalud	87.208.246	17.644.963	87.656.995

Al evaluar la eficiencia relativa del caso de estudio de las EPS Colombianas con el modelo completo (todas las entradas y salidas) y bajo un enfoque en DEA CCR-I, se observa que tiene el siguiente comportamiento:

TABLA 9. RESULTADOS DE EFICIENCIA PARA EPS COLOMBIA MEDIANTE MODELO DEA CCR-I

Eficiencia Relativa con Modelo DEA CCR-I		
DMU	EPS/DMUs	Resultados de Eficiencia
DMU1	Aliansalud	1.0
DMU2	Comfenalco Valle	1.0

DMU3	Compensar	1.0
DMU4	Cooameva	0.974535
DMU5	EPM	1.0
DMU6	Comfenalco Antioquia	1.0
DMU7	Sura	1.0
DMU8	Famisanar	1.0
DMU9	Fondo de Pasivo Ferrocarriles	1.0
DMU10	Golden Group	1.0
DMU11	Nueva EPS	1.0
DMU12	Colpatria	1.0
DMU13	Salud Total	0.848140
DMU14	Saludvida	0.545995
DMU15	Sanitas	0.740990
DMU16	S.O.S.	1.0
DMU17	Solsalud	0.345906

Con el ánimo de comparar los resultados obtenidos, se muestran a continuación los modelos de soluciones optimas no dominadas del frente de Parteo y los resultados de eficiencia relativa de las DMUs que se muestran en [7].

Dado que las métricas “% DMUs Ineficientes” y “Gini Ponderado” utilizados en el presente estudio tienen una pequeña diferencia de enfoque comparado con [7], ya que se utilizaron “% DMUs Eficientes” y para el Gini se realizaron cambios incrementales del α en la ecuación (12), mientras que para esta ocasión se utilizó siempre un $\alpha = 0.5$ en todas las ejecuciones de la metodología del artículo. Dado lo anterior y para llevar la misma lógica se aplicarán las siguientes transformaciones sobre los resultados presentados en [7] así:

$$\%DMUs\ Ineficientes = 1 - \%DMUs\ Eficientes \tag{22}$$

Y para el Gini Ponderado se utilizará el de la ecuación (13) dado que se tienen reportados los resultados de ambas métricas *Gini Eficiencia* y *Gini Ineficiencia*.

TABLA 10. MODELOS SELECCIONADOS POR EL HEURÍSTICO ISA-DEA EN [7]

Subset Evaluado	% DMUs Ineficientes.	Gini Ponderado	Varianza Retenida Entradas	Varianza Retenida Salidas	Varianza Retenida Total
$[I_1, I_2, I_3, I_4, O_2, O_3]$	29.4%	0.42955	100	99.96	99.99
$[I_1, I_2, I_3, O_2, O_3]$	29.4%	0.42345	92.79	99.96	97.73
$[I_1, I_2, I_3, I_4, O_2]$	52.9%	0.44120	100	92.45	99.39
$[I_1, I_2, I_3, O_2]$	52.9%	0.44120	92.79	92.45	95.01
$[I_4, O_2]$	94.1%	0.42735	43.80	92.45	93.50
$[I_2, I_4, O_1]$	70.6%	0.4045	52.49	96.59	98.03
$[I_4, O_1, O_2]$	94.1%	0.42735	43.80	99.95	95.09

TABLA 11. PUNTAJES DE EFICIENCIA PARA LOS MODELOS CON ISA-DEA EN [7]

Modelo	Modelo 1.	Modelo 2	Modelo 3	Modelo 4	Modelo 5	Modelo 6	Modelo 7
DMU1	1.000	1.000	1.000	1.000	0.152	0.531	0.152

DMU2	1.000	1.000	0.981	0.981	0.003	0.096	0.003
DMU3	1.000	1.000	1.000	1.000	0.024	0.148	0.024
DMU4	0.975	0.894	0.882	0.882	0.029	0.184	0.029
DMU5	1.000	1.000	0.000	0.000	0.000	1.000	0.000
DMU6	1.000	1.000	1.000	1.000	0.032	0.183	0.032
DMU7	1.000	1.000	1.000	1.000	0.011	0.062	0.011
DMU8	1.000	1.000	1.000	1.000	0.014	1.000	0.014
DMU9	1.000	1.000	0.000	0.000	0.000	0.295	0.000
DMU10	1.000	1.000	1.000	1.000	0.001	1.000	0.001
DMU11	1.000	1.000	1.000	1.000	0.253	1.000	0.253
DMU12	1.000	1.000	1.000	1.000	1.000	1.000	1.000
DMU13	0.848	0.848	0.609	0.609	0.003	0.034	0.003
DMU14	0.502	0.502	0.477	0.477	0.001	0.014	0.001
DMU15	0.741	0.741	0.741	0.741	0.071	0.296	0.071
DMU16	1.000	1.000	0.991	0.991	0.023	0.385	0.023
DMU17	0.346	0.338	0.322	0.322	0.005	0.053	0.005

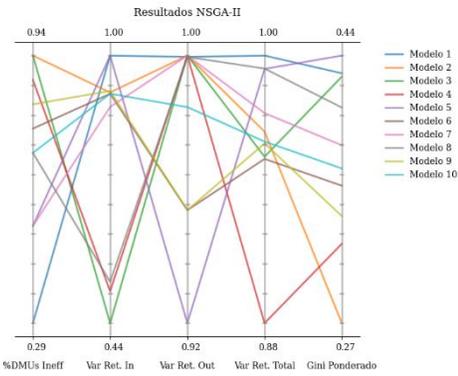


Fig. 10. Comportamiento de las soluciones para NSGA-II en el caso EPS Colombianas

Al realizar la ejecución de la metodología utilizada en el presente artículo, se obtuvieron los siguientes resultados, con resultados satisfactorios tanto en reducción de la dimensionalidad, como en eficiencias relativas para las DMUs. Se muestran los resultados para cada AG

• *NSGA-II.*

TABLA 12. MODELOS SELECCIONADOS POR EL AG NSGA-II

Subset Evaluado	% DMUs Ineficientes.	Gini Ponderado	Varianza Retenida Entradas	Varianza Retenida Salidas	Varianza Retenida Total
$[I_1, I_2, I_3, I_4, O_2, O_3]$	29.41%	0.42955	100	99.96	99.99
$[I_3, I_4, O_1, O_2, O_3]$	94.12%	0.26518	92.25	100	96.56
$[I_4, O_1, O_2, O_3]$	94.12%	0.42731	43.80	100	95.43
$[I_1, O_1, O_2, O_3]$	88.24%	0.31735	50.51	100	87.92
$[I_1, I_2, I_3, I_4, O_2]$	52.94%	0.44118	100	92.45	99.39
$[I_1, I_4, O_3]$	76.47%	0.35559	91.99	95.64	95.33
$[I_1, I_3, O_1, O_2, O_3]$	52.94%	0.38217	89.21	100	97.40
$[I_1, I_4, O_2, O_3]$	70.59%	0.40702	52.49	99.96	99.40
$[I_1, I_2, O_3]$	82.35%	0.33560	92.57	95.64	96.03
$[I_1, I_4, O_1, O_3]$	70.59%	0.36695	91.99	98.55	96.11

TABLA 13. PUNTAJES DE EFICIENCIA PARA LOS MODELOS CON NSGA-II

	Modelo 1.	Modelo 2	Modelo 3	Modelo 4	Modelo 5	Modelo 6	Modelo 7
DMU1	1.000	1.000	1.000	1.000	0.152	0.531	0.152
DMU2	1.000	1.000	0.981	0.981	0.003	0.096	0.003
DMU3	1.000	1.000	1.000	1.000	0.024	0.148	0.024
DMU4	0.975	0.894	0.882	0.882	0.029	0.184	0.029
DMU5	1.000	1.000	0.000	0.000	0.000	1.000	0.000
DMU6	1.000	1.000	1.000	1.000	0.032	0.183	0.032
DMU7	1.000	1.000	1.000	1.000	0.011	0.062	0.011
DMU8	1.000	1.000	1.000	1.000	0.014	1.000	0.014
DMU9	1.000	1.000	0.000	0.000	0.000	0.295	0.000
DMU10	1.000	1.000	1.000	1.000	0.001	1.000	0.001
DMU11	1.000	1.000	1.000	1.000	0.253	1.000	0.253
DMU12	1.000	1.000	1.000	1.000	1.000	1.000	1.000
DMU13	0.848	0.848	0.609	0.609	0.003	0.034	0.003
DMU14	0.502	0.502	0.477	0.477	0.001	0.014	0.001
DMU15	0.741	0.741	0.741	0.741	0.071	0.296	0.071
DMU16	1.000	1.000	0.991	0.991	0.023	0.385	0.023
DMU17	0.346	0.338	0.322	0.322	0.005	0.053	0.005

• *NSGA-III.*

TABLA 14. MODELOS SELECCIONADOS POR EL AG NSGA-III

Subset Evaluado	% DMUs Ineficientes.	Gini Ponderado	Varianza Retenida Entradas	Varianza Retenida Salidas	Varianza Retenida Total
$[I_1, I_4, O_1, O_2]$	70.59%	0.35218	91.99	99.95	95.91
$[I_1, I_2, I_3, I_4, O_2, O_3]$	29.41%	0.42955	100	99.96	99.99
$[I_2, I_4, O_1, O_2, O_3]$	94.12%	0.26518	92.25	100	96.56
$[I_1, I_3, I_4, O_2]$	82.35%	0.36627	97.81	92.45	97.72
$[I_1, I_2, I_3, I_4, O_3]$	41.18%	0.42528	100	95.64	99.67

TABLA 15. PUNTAJES DE EFICIENCIA PARA LOS MODELOS CON NSGA-III

	Modelo 1.	Modelo 2	Modelo 3	Modelo 4	Modelo 5
DMU1	1.000	1.000	0.578	1.000	1.000
DMU2	0.657	1.000	0.098	0.572	0.633
DMU3	0.863	1.000	0.162	0.751	1.000
DMU4	0.827	0.975	0.201	0.639	0.975
DMU5	1.000	1.000	1.000	0.000	1.000
DMU6	1.000	1.000	0.193	0.848	1.000
DMU7	0.875	1.000	0.067	0.983	0.950
DMU8	0.816	1.000	1.000	0.449	1.000
DMU9	0.559	1.000	0.596	0.000	1.000
DMU10	0.853	1.000	1.000	0.170	1.000
DMU11	1.000	1.000	1.000	1.000	1.000
DMU12	1.000	1.000	1.000	1.000	1.000
DMU13	0.670	0.848	0.037	0.609	0.764
DMU14	0.191	0.502	0.014	0.294	0.152
DMU15	0.583	0.741	0.307	0.548	0.722
DMU16	0.991	1.000	0.385	0.694	1.000
DMU17	0.254	0.346	0.056	0.193	0.346

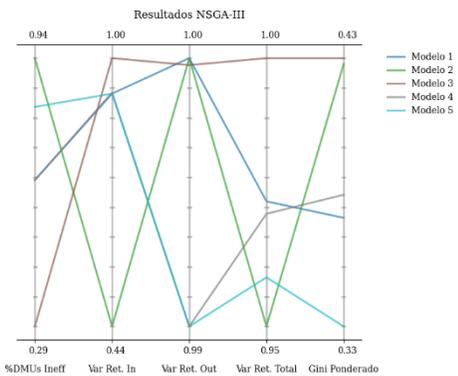


Fig. 11. Comportamiento de las soluciones para NSGA-III en el caso EPS Colombianas

• AGE-MOEA.

TABLA 16. MODELOS SELECCIONADOS POR EL AG AGE-MOEA

Subset Evaluado	% DMUs Ineficientes.	Gini Ponderado	Varianza Retenida Entradas	Varianza Retenida Salidas	Varianza Retenida Total
$[I_1, I_2, O_1, O_3]$	82.35%	0.33023	92.57	98.55	96.26
$[I_1, I_2, O_3]$	82.35%	0.33560	92.57	95.64	96.03
$[I_1, O_1, O_2, O_3]$	88.24%	0.31735	50.51	100	87.93
$[I_1, I_2, I_3, I_4, O_2, O_3]$	29.41%	0.42955	100	99.96	99.99
$[I_3, O_2, O_3]$	94.12%	0.26380	48.78	99.96	89.36
$[I_1, I_2, I_3, I_4, O_2]$	52.94%	0.44118	100	92.45	99.39
$[I_2, I_4, O_1, O_2, O_3]$	70.59%	0.40660	52.49	100	99.46
$[I_1, I_2, I_3, O_1]$	76.47%	0.35446	92.79	96.59	97.25
$[I_2, I_3, O_1, O_3]$	82.35%	0.34121	92.55	98.55	96.79
$[I_1, I_3, I_4, O_1, O_3]$	52.94%	0.39444	97.81	98.55	99.08

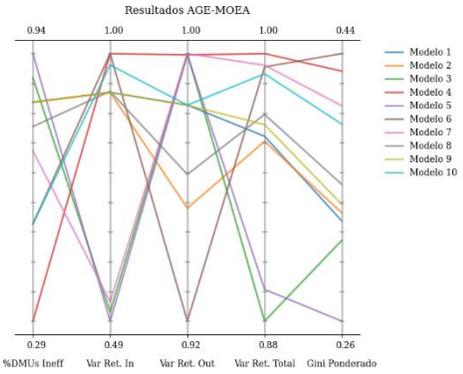


Fig. 12. Comportamiento de las soluciones para AGE-MOEA en el caso EPS Colombianas

TABLA 17. PUNTAJES DE EFICIENCIA PARA LOS MODELOS CON AGE-MOEA

	Modelo 1	Modelo 2	Modelo 3	Modelo 4	Modelo 5
DMU1	0.817	0.801	1.000	1.000	0.282
DMU2	0.664	0.623	0.657	1.000	0.165
DMU3	0.895	0.895	0.895	1.000	0.216
DMU4	0.863	0.863	0.820	0.975	0.144
DMU5	0.320	0.234	0.320	1.000	0.000
DMU6	1.000	1.000	1.000	1.000	0.177
DMU7	0.891	0.891	0.891	1.000	0.295
DMU8	0.955	0.908	0.816	1.000	0.106
DMU9	0.823	0.823	0.823	1.000	0.000
DMU10	1.000	1.000	0.891	1.000	0.079
DMU11	0.627	0.585	0.683	1.000	0.164
DMU12	0.467	0.454	0.467	1.000	1.000
DMU13	0.692	0.692	0.692	0.848	0.220
DMU14	0.192	0.134	0.191	0.502	0.138
DMU15	0.483	0.480	0.565	0.741	0.141
DMU16	1.000	1.000	0.991	1.000	0.156
DMU17	0.267	0.267	0.254	0.346	0.057

	Modelo 6	Modelo 7	Modelo 8	Modelo 9	Modelo 10
DMU1	1.000	0.578	0.853	0.181	1.000
DMU2	0.981	0.098	0.699	0.174	0.699
DMU3	1.000	0.148	0.916	0.221	1.000
DMU4	0.882	0.184	0.803	0.144	0.854
DMU5	0.000	1.000	1.000	1.000	1.000
DMU6	1.000	0.183	1.000	0.167	1.000
DMU7	1.000	0.062	0.952	0.272	0.953
DMU8	1.000	1.000	0.955	0.156	0.835
DMU9	0.000	0.295	0.558	0.940	1.000
DMU10	1.000	1.000	1.000	1.000	1.000
DMU11	1.000	1.000	0.627	0.108	0.925
DMU12	1.000	1.000	0.980	1.000	1.000
DMU13	0.609	0.034	0.766	0.386	0.766
DMU14	0.477	0.014	0.229	0.225	0.229
DMU15	0.741	0.296	0.497	0.099	0.583
DMU16	0.991	0.385	1.000	0.180	1.000
DMU17	0.322	0.053	0.274	0.077	0.301

• SMS-EMOA.

TABLA 18. MODELOS SELECCIONADOS POR EL AG SMS-EMOA

Subest Evaluado	% DMUs Ineficientes.	Gini Ponderado	Varianza Retenida Entradas	Varianza Retenida Salidas	Varianza Retenida Total
$[I_1, I_2, I_3, I_4, O_2, O_3]$	29.41%	0.42955	100	99.96	99.99
$[I_2, I_4, O_1, O_2, O_3]$	70.59%	0.40660	52.49	100	99.46
$[I_1, I_3, O_2]$	88.24%	0.34475	89.21	92.45	91.92
$[I_1, I_2, I_4, O_1, O_2]$	52.94%	0.44118	100	92.45	99.39
$[I_1, I_2, I_3, I_4, O_2]$	52.94%	0.38852	99.86	99.95	99.95
$[I_3, I_4, O_1, O_2, O_3]$	94.12%	0.26518	92.25	100	96.56
$[I_2, I_3, I_4, O_2, O_3]$	35.29%	0.41991	99.86	99.96	99.94
$[I_1, I_2, I_3, O_1, O_2, O_3]$	29.41%	0.42200	92.79	100	98.16
$[I_1, I_2, I_4, O_2]$	64.71%	0.42568	99.86	92.45	98.53
$[I_1, I_3, I_4, O_2]$	82.35%	0.36627	97.81	92.45	97.72

TABLA 19. PUNTAJES DE EFICIENCIA PARA LOS MODELOS CON SMS-EMOA

	Modelo 1	Modelo 2	Modelo 3	Modelo 4	Modelo 5
DMU1	1.000	0.578	1.000	1.000	1.000
DMU2	1.000	0.098	0.572	0.721	0.981
DMU3	1.000	0.162	0.751	0.863	1.000
DMU4	0.975	0.201	0.639	0.908	0.882
DMU5	1.000	1.000	0.000	1.000	0.000
DMU6	1.000	0.193	0.848	1.000	1.000
DMU7	1.000	0.067	0.983	0.875	1.000
DMU8	1.000	1.000	0.449	1.000	1.000
DMU9	1.000	0.596	0.000	0.559	0.000
DMU10	1.000	1.000	0.170	1.000	1.000
DMU11	1.000	1.000	0.683	1.000	1.000
DMU12	1.000	1.000	1.000	1.000	1.000
DMU13	0.848	0.037	0.609	0.670	0.609
DMU14	0.502	0.014	0.294	0.192	0.477
DMU15	0.741	0.307	0.548	0.709	0.741
DMU16	1.000	0.385	0.694	1.000	0.991
DMU17	0.346	0.056	0.193	0.255	0.322

	Modelo 6	Modelo 7	Modelo 8	Modelo 9	Modelo 10
DMU1	0.282	1.000	1.000	1.000	1.000
DMU2	0.165	1.000	1.000	0.716	0.572
DMU3	0.216	1.000	1.000	0.778	0.751
DMU4	0.144	0.964	0.894	0.882	0.639
DMU5	0.000	1.000	1.000	0.000	0.000
DMU6	0.177	0.963	1.000	1.000	0.848
DMU7	0.295	1.000	1.000	0.735	0.983
DMU8	0.106	1.000	1.000	1.000	0.449
DMU9	0.000	1.000	1.000	0.000	0.000

DMU10	0.079	1.000	1.000	1.000	0.170
DMU11	0.253	1.000	1.000	1.000	1.000
DMU12	1.000	1.000	1.000	1.000	1.000
DMU13	0.220	0.685	0.848	0.276	0.609
DMU14	0.138	0.502	0.546	0.111	0.294
DMU15	0.141	0.741	0.741	0.709	0.548
DMU16	0.156	1.000	1.000	0.991	0.694
DMU17	0.057	0.346	0.338	0.208	0.193

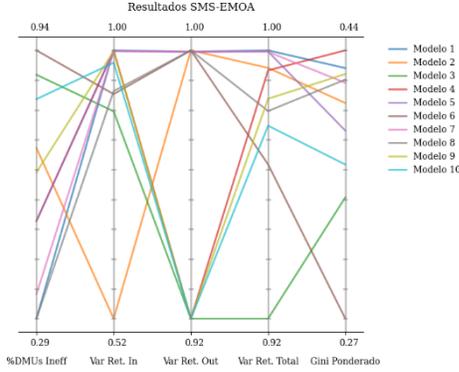


Fig. 13. Comportamiento de las soluciones para SMS-EMOA en el caso EPS Colombianas

- *Indicador de Hipervolumen (HV).*

El resultado obtenido para el HV en el presente caso de análisis de las EPS Colombianas muestra que el AG que se encuentra más cercano al frente óptimo de Pareto entre los cuatro analizados es el SMS-EMOA con un valor de HV = 0.37431.

Al comparar los resultados obtenidos en [7] y los generados por los AGs en este estudio, se puede observar que existen dos (2) soluciones que se destacan como posibles modelos para la reducción de dimensionalidad del conjunto de datos evaluado en este caso ya que se repiten como soluciones no dominadas, los cuales son $[I_1, I_2, I_3, I_4, O_2, O_3]$ y $[I_1, I_2, I_3, I_4, O_2]$, lo cual sugiere que la salida O_1 y posiblemente la O_3 no son significativas en el análisis de DEA. Adicionalmente los anteriores modelos, son los resultados de los AGs que más se repiten (cinco (5) y cuatro (4) veces respectivamente) y al revisar también entre los resultados que se duplicaron, el AG SMS-EMOA fue el que más apareció en la mayoría de las soluciones óptimas no dominadas, esta puede ser la razón por la cual el SMS-EMOA presentó el mejor valor en el indicador de HV, frente a los demás evaluados.

TABLA 20. RESULTADOS DEL HV PARA LOS AG EVALUADOS EN EL CASO DE LAS EPS COLOMBIANAS

AG Evaluado	Resultado HV
NSGA-II	0.371308
NSGA-III	0.345704
AGE-MOEA	0.344415
SMS-EMOA	0.374312

B. Evaluación de Refinerías de Petróleo

Para el segundo caso de estudio se implementó el conjunto de datos de refinerías de petróleo iraquíes del año 2009 mostrado en [30]. El cual está compuesto por cuatro (4) entradas, cuatro (4) salidas y doce (12) DMUs como se muestra en la Tabla 21.

TABLA 21. VALORES DE ENTRADAS Y SALIDAS PARA EL MODELO DEA DE REFINERÍAS DE PETRÓLEO.

DMUs	Entradas			
	Petróleo Crudo (I1)	Trabajadores (I2)	Electricidad (Kwh) (I3)	Hectáreas de Tierra (I4)
DMU1	3.733.935	4.249	77.900.000	855
DMU2	617.740	270	6.048	1.110
DMU3	482.216	549	5.135	323
DMU4	225.340	182	2.300	600
DMU5	40.359.283	2.980	12.042.000	2.000
DMU6	27.231.760	2.220	57.864.000	1.350
DMU7	347.393	204	225.000	80
DMU8	10.112.105	247	360.000	423
DMU9	1.127.080	160	221.000	128
DMU10	8.019.877	4.510	100.174	8.000
DMU11	7.136.817	845	4.712	4.000
DMU12	4.000	375	2.700	600
DMUs	Salidas			
	Nafta (O1)	Gasolina (O2)	Keroseno (O3)	Acete de Petróleo (O4)
DMU1	368.854	452.196	218.503	1.921.998
DMU2	111.188	101.833	64.845	320.098
DMU3	81.893	84.833	48.741	240.454
DMU4	40.100	38.674	20.764	119.262
DMU5	1.195.126	1.433.560	823.529	2.646.866
DMU6	761.253	828.516	569.705	1.898.627
DMU7	4.663	9.268	3.330	22.573
DMU8	0.00	201.068	211.824	859.847
DMU9	2.167	37.063	0.00	89.404
DMU10	642.100	1.446.603	591.398	3.492.130
DMU11	50.849	168.141	86.832	673.677
DMU12	648	180	107	197

Para la evaluación de las refinerías de petróleo de [30], se obtuvieron los siguientes resultados para cada AG.

- *NSGA-II.*

TABLA 22. MODELOS SELECCIONADOS POR EL AG NSGA-II PARA CASO DE LAS REFINERÍAS DE PETRÓLEO.

Subest Evaluado	% DMUs Ineficientes.	Gini Ponderado	Varianza Retenida Entradas	Varianza Retenida Salidas	Varianza Retenida Total
$[I_1, O_1]$	91.67%	0.41274	33.06	87.82	73.51
$[I_1, I_1, O_4]$	75.00%	0.37230	68.88	86.48	91.54
$[I_1, I_2, I_3, I_4, O_1, O_2, O_4]$	25.00%	0.42415	100	99.14	99.98
$[I_1, I_2, I_3, I_4, O_1, O_3, O_4]$	25.00%	0.42750	100	99.33	99.98
$[I_3, O_1, O_2, O_3]$	83.33%	0.45031	36.51	98.34	97.10
$[I_2, O_2, O_3, O_4]$	91.67%	0.28521	48.49	98.46	96.80
$[I_2, I_3, I_4, O_1, O_2, O_3, O_4]$	33.33%	0.44592	80.44	100	99.72
$[I_4, O_3]$	91.67%	0.39149	35.73	92.04	81.94
$[I_1, I_4, O_1]$	66.67%	0.44219	65.48	87.82	87.09
$[I_1, I_2, I_4, O_2, O_3]$	50.00%	0.40732	93.26	96.45	98.57
$[I_2, I_4, O_3, O_4]$	83.33%	0.34586	73.53	97.67	95.83
$[I_1, I_1, O_1, O_3]$	75.00%	0.40814	68.88	93.99	91.15
$[I_1, I_3, O_1, O_3]$	83.33%	0.40064	62.36	93.99	94.78
$[I_2, O_1]$	91.67%	0.38987	48.49	87.82	80.03
$[I_2, O_1, O_2, O_3, O_4]$	75.00%	0.36107	48.49	100	97.12
$[I_2, I_4, O_1, O_2, O_4]$	58.33%	0.37573	73.53	99.14	98.48
$[I_4, O_2]$	91.67%	0.35230	35.73	92.82	81.52
$[I_2, I_4, O_2, O_3, O_4]$	75.00%	0.35292	73.53	98.46	98.65
$[I_2, I_3, I_4, O_1, O_2, O_3]$	41.67%	0.43434	80.44	98.34	99.67
$[I_2, I_3, I_4, O_2, O_3, O_4]$	41.67%	0.43917	80.44	98.46	99.62
$[I_1, I_3, O_1, O_3, O_4]$	75.00%	0.39982	62.36	99.33	97.88

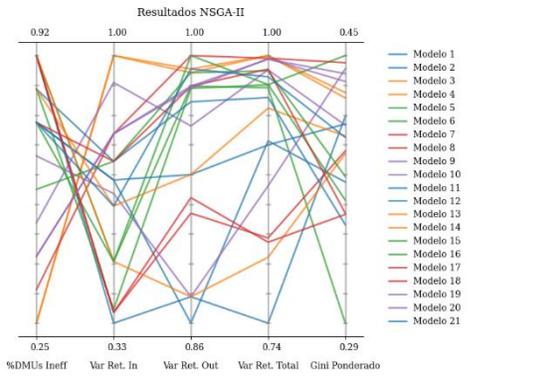


Fig. 14. Comportamiento de las soluciones para NSGA-II en el caso Refinerías de Petróleo

• *NSGA-III.*

TABLA 23. MODELOS SELECCIONADOS POR EL AG NSGA-III PARA CASO DE LAS REFINERÍAS DE PETRÓLEO.

Subset Evaluado	% DMUs Ineficientes.	Gini Ponderado	Varianza Retenida Entradas	Varianza Retenida Salidas	Varianza Retenida Total
$[I_1, O_1]$	66.67%	0.41937	80.44	92.04	98.47
$[I_1, I_1, O_4]$	75.00%	0.38328	73.52	96.45	95.46
$[I_1, I_2, I_3, I_4, O_1, O_2, O_4]$	66.67%	0.39246	35.73	100.00	98.64
$[I_1, I_2, I_3, I_4, O_1, O_3, O_4]$	75.00%	0.38064	93.66	92.04	96.08
$[I_3, O_1, O_2, O_3]$	91.67%	0.40022	93.66	87.82	96.03
$[I_2, O_2, O_3, O_4]$	41.67%	0.43819	74.23	100.00	99.58
$[I_2, I_3, I_4, O_1, O_2, O_3, O_4]$	83.33%	0.45031	36.51	100.00	97.48
$[I_4, O_3]$	91.67%	0.28521	48.49	98.46	96.80
$[I_1, I_4, O_1]$	66.67%	0.39771	93.66	98.14	97.63
$[I_1, I_2, I_4, O_2, O_3]$	58.33%	0.40783	93.66	99.33	98.39
$[I_2, I_4, O_3, O_4]$	41.67%	0.44263	100.00	93.99	99.95
$[I_1, I_1, O_1, O_3]$	25.00%	0.42750	100.00	99.33	99.98
$[I_1, I_3, O_1, O_3]$	83.33%	0.45031	36.51	98.34	97.10

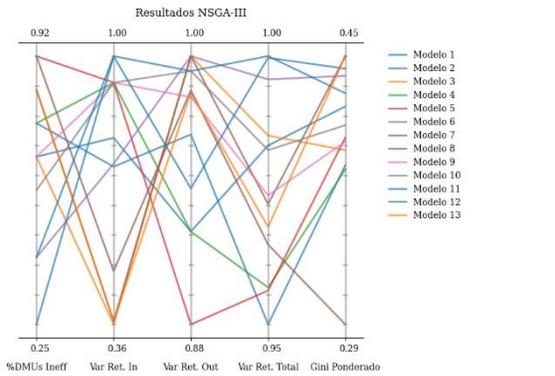


Fig. 15. Comportamiento de las soluciones para NSGA-III en el caso Refinerías de Petróleo

• *AGE-MOEA.*

TABLA 24. MODELOS SELECCIONADOS POR EL AG AGE-MOEA PARA CASO DE LAS REFINERÍAS DE PETRÓLEO.

Subset Evaluado	% DMUs Ineficientes.	Gini Ponderado	Varianza Retenida Entradas	Varianza Retenida Salidas	Varianza Retenida Total
$[I_1, O_1]$	91.67%	0.41274	33.06	87.82	73.51
$[I_1, I_2, I_3, I_4, O_1]$	66.67%	0.42538	100.00	87.82	99.11
$[I_1, I_2, I_3, I_4, O_3, O_4]$	41.67%	0.41429	100.00	97.67	99.83
$[I_2, I_3, O_3]$	75.00%	0.40401	73.52	92.04	95.01
$[I_2, O_2, O_3, O_4]$	91.67%	0.28521	48.49	98.46	96.80
$[I_1, I_2, I_3, I_4, O_1, O_3, O_4]$	25.00%	0.42750	100.00	99.33	99.98

$[I_1, I_2, I_3, O_1, O_3]$	66.67%	0.41777	93.66	93.99	97.23
$[I_2, I_3, I_4, O_1, O_2, O_3, O_4]$	33.33%	0.44592	80.44	100.00	99.72
$[I_1, I_2, I_3, O_3, O_4]$	66.67%	0.38806	93.66	97.67	97.86
$[I_1, I_2, I_4, O_1]$	66.67%	0.43511	93.26	87.82	96.04
$[I_1, I_2, I_3, O_1, O_4]$	66.67%	0.39771	93.66	98.14	97.63
$[I_2, I_3, I_4, O_1, O_3]$	50.00%	0.44295	80.44	93.99	99.34
$[I_1, I_2, I_3, I_4, O_1, O_3]$	41.67%	0.44263	100.00	93.99	99.95
$[I_1, I_3, O_1, O_3, O_4]$	75.00%	0.39982	62.36	99.33	97.88
$[I_1, I_2, I_3, I_4, O_3]$	58.33%	0.42400	100.00	92.04	99.30
$[I_2, I_3, O_1, O_2, O_3]$	66.67%	0.42131	73.52	98.34	97.46
$[I_1, I_4, O_1, O_2, O_4]$	50.00%	0.41823	65.48	99.14	98.76
$[I_1, O_4]$	91.67%	0.38203	33.06	86.48	82.07
$[I_1, I_2, I_3, O_1, O_2, O_3, O_4]$	41.67%	0.42346	93.66	100.00	98.64
$[I_2, I_4, O_3, O_4]$	83.33%	0.34586	73.53	97.67	95.83
$[I_1, I_2, I_3, O_2, O_4]$	50.00%	0.41811	93.66	95.51	97.54

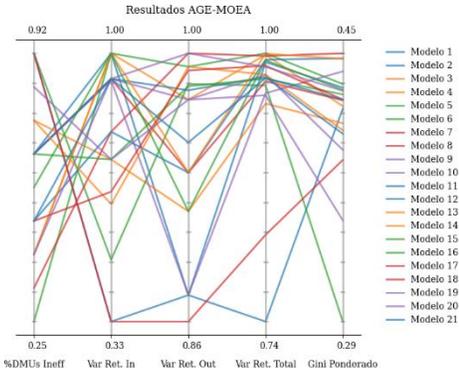


Fig. 16. Comportamiento de las soluciones para NSGA-II en el caso Refinerías de Petróleo

• *SMS-EMOA.*

TABLA 25. MODELOS SELECCIONADOS POR EL AG SMS-EMOA PARA CASO DE LAS REFINERÍAS DE PETRÓLEO.

Subset Evaluado	% DMUs Ineficientes.	Gini Ponderado	Varianza Retenida Entradas	Varianza Retenida Salidas	Varianza Retenida Total
$[I_3, O_2, O_3, O_4]$	91.67%	0.39849	36.51	98.46	97.17
$[I_1, I_2, I_3, I_4, O_1, O_3]$	41.67%	0.44263	100.00	93.99	99.95
$[I_3, O_1, O_2, O_3, O_4]$	83.33%	0.45031	36.51	100.00	97.48
$[I_1, I_2, I_3, I_4, O_1]$	66.67%	0.42538	100.00	87.82	99.11
$[I_1, I_2, I_3, O_1]$	91.67%	0.40022	93.66	87.82	96.03
$[I_2, I_4, O_3, O_4]$	83.33%	0.34586	73.53	97.67	95.83
$[I_2, I_3, I_4, O_1, O_3]$	50.00%	0.44295	80.44	93.99	99.34
$[I_2, I_3, O_1, O_4]$	75.00%	0.41548	73.52	98.14	96.83
$[I_1, I_2, I_3, O_1, O_3, O_4]$	58.33%	0.40783	93.66	99.33	98.39
$[I_1, I_2, I_3, I_4, O_1, O_3, O_4]$	25.00%	0.42750	100.00	99.33	99.98
$[I_1, I_3, I_4, O_1, O_2, O_3, O_4]$	25.00%	0.42395	96.17	100.00	99.91
$[I_1, I_2, I_4, O_1, O_2, O_3, O_4]$	33.33%	0.41042	93.26	100.00	99.31
$[I_2, I_4, O_1, O_2, O_3, O_4]$	58.33%	0.37573	73.53	100.00	98.88
$[I_1, I_2, I_4, O_2, O_3, O_4]$	41.67%	0.40962	93.26	98.46	99.06
$[I_1, I_3, O_1, O_2, O_3, O_4]$	58.33%	0.40595	62.36	100.00	98.06
$[I_2, I_3, I_4, O_1, O_2]$	50.00%	0.42913	80.44	97.48	98.87
$[I_1, I_2, I_3, O_1, O_4]$	66.67%	0.39771	93.66	98.14	97.63
$[I_2, I_3, O_1, O_2, O_3, O_4]$	66.67%	0.41918	73.52	100.00	97.97
$[I_1, I_3, O_1, O_4]$	75.00%	0.40274	62.36	98.14	97.22
$[I_1, I_3, O_1, O_3, O_4]$	75.00%	0.39982	62.36	99.33	97.88
$[I_2, I_3, I_4, O_1, O_3, O_4]$	33.33%	0.44634	80.44	99.33	99.52

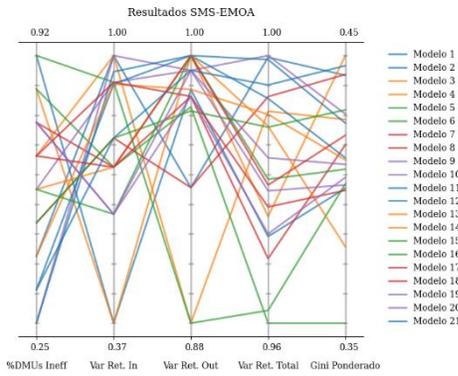


Fig. 17. Comportamiento de las soluciones para SMS-EMOA en el caso Refinerías de Petróleo

Para el caso de las refinerías de petróleo [30] las soluciones que más se repitieron entre los cuatro modelos son los que se muestran en la Tabla 26:

TABLA 26. SOLUCIONES MÁS SELECCIONADAS POR LOS AGs EN EL CASO DE LAS REFINERÍAS DE PETRÓLEO

Subset Evaluado	% DMUs Ineficientes.	Gini Ponderado	Varianza Retenida Entradas	Varianza Retenida Salidas	Varianza Retenida Total
$[I_1, I_2, I_3, I_4, O_1, O_3, O_4]$	25.00%	0.42750	100.00	99.33	99.98
$[I_1, I_2, I_3, I_4, O_1, O_3]$	41.67%	0.44263	100.00	93.99	99.95
$[I_1, I_2, I_3, O_1, O_4]$	66.67%	0.39771	93.66	98.14	97.63
$[I_1, I_3, O_1, O_3, O_4]$	75.00%	0.39982	62.36	99.33	97.88
$[I_2, I_4, O_3, O_4]$	83.33%	0.34586	73.53	97.67	95.83
$[I_2, O_2, O_3, O_4]$	91.67%	0.28521	48.49	98.46	96.80

Al validar el indicador de HV para los AGs en el caso de las refinerías de petróleo se puede observar nuevamente que el SMS-EMOA obtiene el mayor valor con un $HV = 0.38214$ y al evaluar los modelos propuestos por cada AG revisado el SMS-EMOA es que el apareció más veces en las soluciones que concordaban con otros AGs evaluados.

TABLA 27. RESULTADOS DEL HV PARA LOS AG EVALUADOS EN EL CASO DE LAS REFINERÍAS DE PETRÓLEO

AG Evaluado	Resultado HV
NSGA-II	0.327974
NSGA-III	0.371118
AGE-MOEA	0.349818
SMS-EMOA	0.382140

C. Evaluación de Calidad de Vida.

Para el tercer y último caso de estudio se implementó el conjunto de datos referente a la calidad de vida de algunas ciudades de EE.UU. presentado en [5]. El cual está compuesto por seis (6) entradas, seis (6) salidas y quince (15) DMUs como se muestra en la Tabla 26.

Las variables para las entradas y salidas son las siguientes, Entradas: precio de las viviendas de lujo (I1), alquiler mensual de viviendas económicas (I2), costo de una barra de pan francés (I3), costo de un Martini (I4), alquiler de oficinas clase A (I5) y número de crímenes violentos (I6).

Salidas: ingreso mediano del hogar (O1), número de personas con título de licenciatura (O2), número de doctores (O3), número de museos (O4), número de bibliotecas (O5) y número de campos de golf de 18 hoyos (O6).

TABLA 28. VALORES DE ENTRADAS Y SALIDAS PARA EL MODELO DEA DE CALIDAD DE VIDA.

DMUs	Entradas					
	(I1)	(I2)	(I3)	(I4)	(I5)	(I6)
DMU1/ Seattle	586,000	581	1.45	4.50	21	542.3
DMU2/ Denver	475,000	558	0.97	4.00	14	595.6
DMU3/ Philadelphia	201,000	600	1.50	4.75	21	693.6
DMU4/ Minneapolis	299,000	609	1.49	4.00	24	496.5
DMU5/ Ral-Durham	318,000	613	0.99	4.50	18	634.7
DMU6/ St. Louis	265,000	558	0.89	3.00	18	263.0
DMU7/ Cincinnati	467,000	580	1.25	3.75	20	551.5
DMU8/ Washington	583,000	625	1.29	3.75	33	714.5
DMU9/ Pittsburgh	347,000	535	0.99	3.75	17	382.1
DMU10/ Dallas-FW	296,000	650	1.50	5.00	18	825.4
DMU11/ Atlanta	600,000	740	1.19	6.75	20	846.6
DMU12/ Baltimore	575,000	775	0.99	3.99	18	1,296.3
DMU13/ Boston	351,000	888	1.09	4.25	34	686.6
DMU14/ Milwaukee	283,000	727	1.53	3.50	26	518.9
DMU15/ Nashville	431,000	695	1.19	4	26	1,132.5
DMUs	Salidas					
	(O1)	(O2)	(O3)	(O4)	(O5)	(O6)
DMU1	46,928	0.297	4.49	7	117	22
DMU2	42,879	0.291	2.79	5	60	71
DMU3	43,576	0.227	3.64	25	216	166
DMU4	45,673	0.27	2.67	6	131	125
DMU5	4,099	0.319	4.94	7	33	47
DMU6	39,079	0.206	3.4	10	104	62
DMU7	38,455	0.199	2.8	4	71	94
DMU8	54,291	0.373	3.35	30	148	105
DMU9	34,534	0.188	3.66	8	124	112
DMU10	41,984	0.271	1.96	3	98	77
DMU11	43,249	0.263	2.23	9	118	102
DMU12	43,291	0.233	4.02	8	102	45
DMU13	46,444	0.325	5.69	25	240	55
DMU14	41,841	0.214	3.11	6	52	50
DMU15	40,221	0.215	3.25	4	37	37

Dada la alta dimensionalidad del conjunto de datos para este caso, y la cantidad de modelos con soluciones no dominadas generadas, se mostrarán como modelos posibles los que tienen mayores coincidencias entre los AGs. Los cuales se muestran en la Tabla 29.

TABLA 29. MODELOS CON MAYOR APARICIÓN EN LOS AG EVALUADOS PARA EL CASO DE CALIDAD DE VIDA.

Subset Evaluado	% DMUs Ineficientes	Gini Ponderado	Varianza Retenida Entradas	Varianza Retenida Salidas	Varianza Retenida Total
Modelo 1	26,67	0,39846	100	97,54	99,51
Modelo 2	40,00	0,41501	100	86,83	95,94
Modelo 3	46,67	0,41145	100	79,02	92,82
Modelo 4	46,67	0,41145	100	79,02	92,82
Modelo 5	46,67	0,41928	100	55,11	89,51
Modelo 6	40,00	0,38709	100	95,68	99,55
Modelo 7	53,33	0,34713	100	84,95	95,86
Modelo 8	73,33	0,31016	100	58,17	88,52
Modelo 9	46,67	0,37659	93,32	97,54	98,09

Modelo 10	66,67	0,31354	83,33	85,18	92,04
Modelo 11	46,67	0,40984	89,23	86,83	92,79
Modelo 12	53,33	0,34900	75,84	97,54	92,95
Modelo 13	86,67	0,27172	48,70	86,83	77,80
Modelo 14	60,00	0,37994	94,68	65,81	88,39
Modelo 15	53,33	0,39396	94,68	79,02	90,49
Modelo 16	73,33	0,31683	79,28	74,89	84,21
Modelo 17	40,00	0,39963	81,30	95,68	92,49
Modelo 18	60,00	0,40604	100	24,79	81,01

En la Tabla 30, se muestran la combinación de entradas y salidas del caso para la calidad de vida en ciudades de EE.UU.

TABLA 30. MODELOS RESULTADO DE AGS PARA EL CASO DE CALIDAD DE VIDA

Nº Modelo	Combinación Entradas y Salidas
Modelo 1	[I ₁ , I ₂ , I ₃ , I ₄ , I ₅ , I ₆ , O ₁ , O ₂ , O ₄ , O ₅ , O ₆]
Modelo 2	[I ₁ , I ₂ , I ₃ , I ₄ , I ₅ , I ₆ , O ₁ , O ₄ , O ₅ , O ₆]
Modelo 3	[I ₁ , I ₂ , I ₃ , I ₄ , I ₅ , I ₆ , O ₁ , O ₄ , O ₆]
Modelo 4	[I ₁ , I ₂ , I ₃ , I ₄ , I ₅ , I ₆ , O ₁ , O ₅ , O ₆]
Modelo 5	[I ₁ , I ₂ , I ₃ , I ₄ , I ₅ , I ₆ , O ₁ , O ₅]
Modelo 6	[I ₁ , I ₂ , I ₃ , I ₄ , I ₅ , I ₆ , O ₂ , O ₃ , O ₄ , O ₅ , O ₆]
Modelo 7	[I ₁ , I ₂ , I ₃ , I ₄ , I ₅ , I ₆ , O ₃ , O ₄ , O ₅ , O ₆]
Modelo 8	[I ₁ , I ₂ , I ₃ , I ₄ , I ₅ , I ₆ , O ₅ , O ₆]
Modelo 9	[I ₁ , I ₂ , I ₃ , I ₄ , I ₆ , O ₁ , O ₂ , O ₄ , O ₅ , O ₆]
Modelo 10	[I ₁ , I ₂ , I ₃ , I ₄ , O ₂ , O ₃ , O ₄ , O ₅]
Modelo 11	[I ₁ , I ₂ , I ₄ , I ₅ , I ₆ , O ₁ , O ₄ , O ₅ , O ₆]
Modelo 12	[I ₁ , I ₂ , I ₄ , I ₆ , O ₁ , O ₂ , O ₄ , O ₅ , O ₆]
Modelo 13	[I ₁ , I ₂ , O ₁ , O ₄ , O ₅ , O ₆]
Modelo 14	[I ₁ , I ₃ , I ₄ , I ₅ , I ₆ , O ₁ , O ₄ , O ₅]
Modelo 15	[I ₁ , I ₃ , I ₄ , I ₅ , I ₆ , O ₁ , O ₄ , O ₅]
Modelo 16	[I ₁ , I ₃ , I ₄ , I ₅ , I ₆ , O ₃ , O ₅ , O ₆]
Modelo 17	[I ₁ , I ₃ , I ₅ , I ₆ , O ₂ , O ₃ , O ₄ , O ₅ , O ₆]
Modelo 18	[I ₂ , I ₃ , I ₄ , I ₅ , I ₆ , O ₁]

Con el ánimo de ilustrar el comportamiento de los modelos se presentarán los primeros cinco (5) que fueron elegidos por cada AG.

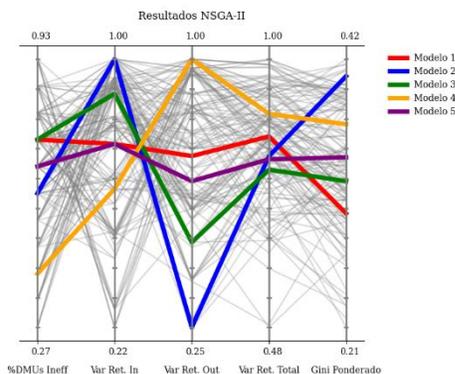


Fig. 18. Primeros cinco modelos en la evaluación de NSGA-II para el caso de calidad de vida

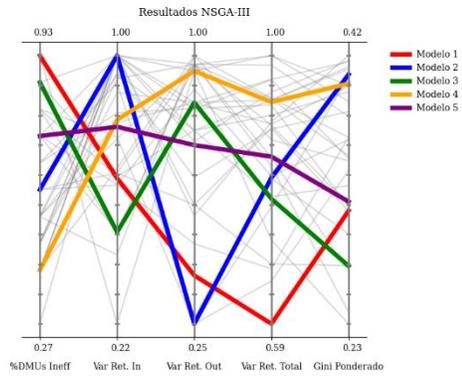


Fig. 19. Primeros cinco modelos en la evaluación de NSGA-III para el caso de calidad de vida

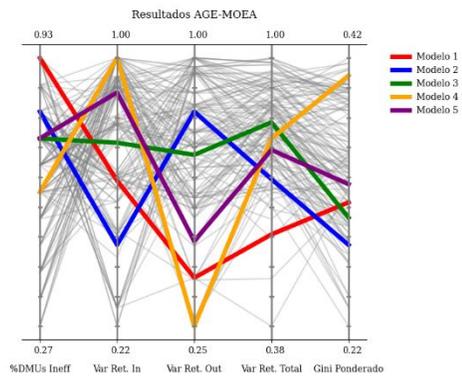


Fig. 20. Primeros cinco modelos en la evaluación de AGE-MOEA para el caso de calidad de vida

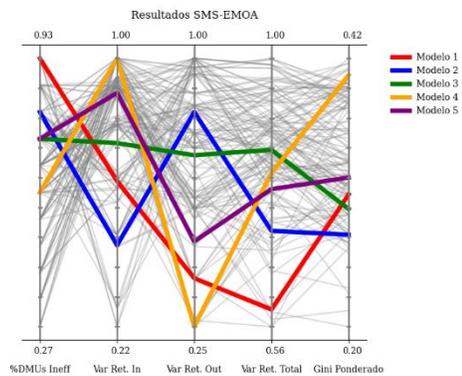


Fig. 21. Primeros cinco modelos en la evaluación de SMS-EMOA para el caso de calidad de vida

Al validar el indicador de HV para los AGs en el caso de la calidad de vida en ciudades de EE. UU., se puede observar que en este caso es el NSGA-II quien obtiene el mayor valor con un $HV = 0.30788$ y al evaluar los modelos propuestos por cada AG revisado el NSGA-II es que el aparece más veces en las soluciones que concordaban con otros los AGs evaluados.

TABLA 31. RESULTADOS DEL HV PARA LOS AG EVALUADOS EN EL CASO DE LA CALIDAD DE VIDA EN EE. UU.

AG Evaluado	Resultado HV
NSGA-II	0.307878
NSGA-III	0.293101
AGE-MOEA	0.306507
SMS-EMOA	0.305793

En la Tabla 32 se presentan los resultados del indicador de HV para cada caso evaluado en el presente estudio, con el fin de obtener una comparación de cada uno.

TABLA 32. RESULTADOS DE HV PARA LOS AGS Y LOS CASOS EVALUADOS

Resultados de HV				
AG Evaluado	Deptos. Médicos	ESP Colombia	Refinerías de Petróleo	Calidad de Vida
NSGA-II	0.171536	0.371308	0.327974	0.307878
NSGA-III	0.190388	0.345704	0.371118	0.293101
AGE-MOEA	0.156581	0.344415	0.349818	0.306507
SMS-EMOA	0.246001	0.374312	0.382140	0.305793

V. CONCLUSIONES

El problema de discriminación de eficiencia relativa en DEA ocasionado por conjuntos de datos que adolecen de alta dimensionalidad al no cumplir con la regla empírica presentada por varios autores como se mostró en este estudio, ha generado interés dado la alta usabilidad de la técnica desarrollada por Charnes, Cooper y Rhodes [2] en diversos sectores económicos [15]. Es por lo anterior que el desarrollo de técnicas que ayuden a mejorar el poder discriminatorio en DEA puede favorecer a generar información más certera y confiable.

El abordar el problema de dimensionalidad en DEA como se menciona en [5] tiene dos formas para afrontarse. La primera es aumentar la cantidad de DMUs (lo que en ciertos casos no es viable o podría resultar en un alto costo) o la forma en la que se planteó en este artículo, mediante el desarrollo de un método para la selección de características representativas. Dicho método se implementó mediante Python a través del uso de algoritmos genéticos multiobjetivo que evalúan subconjuntos de las variables originales del problema buscando mantener la variabilidad de los datos y aumentar la discriminación de eficiencia.

Al comparar los algoritmos implementados el que mayor desempeño mostró a través de los casos evaluados fue el SMS-EMOA para 3 de los 4 casos y el NSGA-II en el último caso con mayor dimensionalidad. El criterio para la evaluación del desempeño se realizó mediante la obtención de un mayor resultado del indicador de hipervolumen. lo que invita a evaluar

el AG SMS-EMOA como una buena opción en cuanto a la metodología de FS para reducción de dimensionalidad en DEA.

El presente trabajo deja diversas oportunidades en las que puede ser reforzado, mediante el uso de otras técnicas de optimización multiobjetivo. Así mismo, generalizar la aplicación de FS a cualquier tipo de conjuntos de datos variando el tipo de medidas de desempeño a las utilizadas aquí. Un punto importante para tener en cuenta para mejorar el presente trabajo es el de la búsqueda de un método diferente para la aleatorización de la población inicial de los AGs, dado el alto costo computacional que se puede tener con una gran cantidad de variables en el conjunto de datos.

REFERENCIAS

- [1] A. Charnes, W. W. Cooper, y E. Rhodes, "Measuring the efficiency of decision making units", *Eur. J. Oper. Res.*, vol. 2, núm. 6, pp. 429–444, nov. 1978, doi: 10.1016/0377-2217(78)90138-8.
- [2] A. Charnes, W. W. Cooper, y R. M. Thrall, "Classifying and characterizing efficiencies and inefficiencies in Data Development Analysis", *Elsevier Sci. Publ.*, vol. 5, núm. 3, pp. 105–110, 1986.
- [3] W. W. Cooper, L. M. Seiford, y K. Tone, *Data envelopment analysis: a comprehensive text with models, applications, references and DEA-solver software*, 2nd ed. New York: Springer, 2007.
- [4] R. Fourer, D. M. Gay, y B. W. Kernighan, "AMPL: A Mathematical Programming Language", *Manag. Sci.*, núm. 36, pp. 519–554, 1990.
- [5] V. Charles, J. Aparicio, y J. Zhu, "The curse of dimensionality of decision-making units: A simple approach to increase the discriminatory power of data envelopment analysis", *Eur. J. Oper. Res.*, vol. 279, núm. 3, pp. 929–940, dic. 2019, doi: 10.1016/j.ejor.2019.06.025.
- [6] S. Khalid, T. Khalil, y S. Nasreen, "A survey of feature selection and feature extraction techniques in machine learning", en *2014 Science and Information Conference*, London, UK: IEEE, ago. 2014, pp. 372–378. doi: 10.1109/SAI.2014.6918213.
- [7] A. M. Gómez Ardila, "Evaluación de eficiencia en Instituciones Prestadoras de Salud en Colombia: un estudio con Análisis Envolvente de Datos", Universidad Nacional de Colombia, 2021.
- [8] K. Deb, *Multi-objective optimization using evolutionary algorithms*, Repr. en Wiley-Interscience series in systems and optimization. Chichester Weinheim: Wiley, 2004.
- [9] J. Blank y K. Deb, "Pymoo: Multi-Objective Optimization in Python", *IEEE Access*, vol. 8, pp. 89497–89509, 2020, doi: 10.1109/ACCESS.2020.2990567.
- [10] K. Deb, A. Pratap, S. Agarwal, y T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: NSGA-II",

- IEEE Trans. Evol. Comput.*, vol. 6, núm. 2, pp. 182–197, abr. 2002, doi: 10.1109/4235.996017.
- [11] K. Deb y H. Jain, “An Evolutionary Many-Objective Optimization Algorithm Using Reference-Point-Based Nondominated Sorting Approach, Part I: Solving Problems With Box Constraints”, *IEEE Trans. Evol. Comput.*, vol. 18, núm. 4, pp. 577–601, ago. 2014, doi: 10.1109/TEVC.2013.2281535.
- [12] N. Beume, B. Naujoks, y M. Emmerich, “SMS-EMOA: Multiobjective selection based on dominated hypervolume”, *Eur. J. Oper. Res.*, vol. 181, núm. 3, pp. 1653–1669, sep. 2007, doi: 10.1016/j.ejor.2006.08.008.
- [13] A. Panichella, “An adaptive evolutionary algorithm based on non-euclidean geometry for many-objective optimization”, en *Proceedings of the Genetic and Evolutionary Computation Conference*, Prague Czech Republic: ACM, jul. 2019, pp. 595–603. doi: 10.1145/3321707.3321839.
- [14] J. L. Devore, “15. Procedimientos libres de distribución”, en *Probabilidad y estadística*, 7a ed., Cengage Learning, 2008, pp. 599–607.
- [15] J. S. Liu, L. Y. Y. Lu, W.-M. Lu, y B. J. Y. Lin, “A survey of DEA applications”, *Omega*, vol. 41, núm. 5, pp. 893–902, oct. 2013, doi: 10.1016/j.omega.2012.11.004.
- [16] J. Li, T. Luo, B. Zhang, M. Chen, y J. Zhou, “IMOABC: An efficient multi-objective filter-wrapper hybrid approach for high-dimensional feature selection”, *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 36, núm. 9, p. 102205, nov. 2024, doi: 10.1016/j.jksuci.2024.102205.
- [17] D. L. Banks y S. E. Fienberg, “Data Mining, Statistics”, en *Encyclopedia of Physical Science and Technology*, 3a ed., Elsevier, 2003, pp. 247–261. doi: 10.1016/B0-12-227410-5/00019-3.
- [18] S. Sharma y V. Kumar, “A Comprehensive Review on Multi-objective Optimization Techniques: Past, Present and Future”, *Arch. Comput. Methods Eng.*, vol. 29, núm. 7, pp. 5605–5633, nov. 2022, doi: 10.1007/s11831-022-09778-9.
- [19] A. Jaskiewicz, “Evaluation of Multiple Objective Metaheuristics”, en *Metaheuristics for Multiobjective Optimisation*, vol. 535, X. Gandibleux, M. Sevaux, K. Sörensen, y V. T’kindt, Eds., en *Lecture Notes in Economics and Mathematical Systems*, vol. 535., Berlin, Heidelberg: Springer Berlin Heidelberg, 2004, pp. 65–89. doi: 10.1007/978-3-642-17144-4_3.
- [20] N. Adler y B. Golany, “Including principal component weights to improve discrimination in data envelopment analysis”, *J. Oper. Res. Soc.*, vol. 53, núm. 9, pp. 985–991, sep. 2002, doi: 10.1057/palgrave.jors.2601400.
- [21] Y.-H. B. Wong y J. E. Beasley, “Restricting Weight Flexibility in Data Envelopment Analysis”, *J. Oper. Res. Soc.*, vol. 41, núm. 9, pp. 829–835, 1990.
- [22] D. M. Mandy, “Pareto Efficiency”, en *Producers, consumers, and partial equilibrium*, First edition., Amsterdam: Elsevier, 2017, pp. 291–295.
- [23] R. Nisbet, J. F. Elder, y G. Miner, “Feature Selection”, en *Handbook of statistical analysis and data mining applications*, Amsterdam Boston: Academic Press/Elsevier, 2009, pp. 77–97.
- [24] F. Medina H. y NU. CEPAL. División de Estadística y Proyecciones Económicas, *Consideraciones sobre el índice de Gini para medir la concentración del ingreso*. en Serie Estudios Estadísticos y Prospectivos, no. 9. s.l: CEPAL, 2001.
- [25] L. Jenkins y M. Anderson, “A multivariate statistical approach to reducing the number of variables in data envelopment analysis”, *Eur. J. Oper. Res.*, vol. 147, núm. 1, pp. 51–61, may 2003, doi: 10.1016/S0377-2217(02)00243-6.
- [26] A. P. Guerreiro, C. M. Fonseca, y L. Paquete, “The Hypervolume Indicator: Computational Problems and Algorithms”, *ACM Comput. Surv.*, vol. 54, núm. 6, pp. 1–42, jul. 2022, doi: 10.1145/3453474.
- [27] C. M. Fonseca, L. Paquete, y M. Lopez-Ibanez, “An Improved Dimension-Sweep Algorithm for the Hypervolume Indicator”, en *2006 IEEE International Conference on Evolutionary Computation*, Vancouver, BC, Canada: IEEE, 2006, pp. 1157–1163. doi: 10.1109/CEC.2006.1688440.
- [28] K. Shang, H. Ishibuchi, L. He, y L. M. Pang, “A Survey on the Hypervolume Indicator in Evolutionary Multiobjective Optimization”, *IEEE Trans. Evol. Comput.*, vol. 25, núm. 1, pp. 1–20, feb. 2021, doi: 10.1109/TEVC.2020.3013290.
- [29] T. J. Fontalvo Herrera, “Eficiencia de las entidades prestadoras de salud (EPS) en Colombia por medio de análisis envolvente de datos”, *Ingeniare Rev. Chil. Ing.*, vol. 25, núm. 4, pp. 681–692, dic. 2017, doi: 10.4067/S0718-33052017000400681.
- [30] S. M. Al-Najjar y M. A. Al-Jaybajy, “Application of Data Envelopment Analysis to Measure the Technical Efficiency of Oil Refineries: A Case Study”, *Int. J. Bus. Adm.*, vol. 3, núm. 5, p. p64, sep. 2012, doi: 10.5430/ijba.v3n5p64.