Asignación de Docentes a Establecimientos Educativos: Un Enfoque Multi-objetivo

Horacio Villalba-Martí[1], Fabio López-Pires[2]
y Eustaquio A Martínez[1].

[1] Facultad Politécnica, Universidad Nacional del Este,
Campus Km 8 Acaray, Ciudad del Este, Paraguay
{vimartih,amartinez}@fpune.edu.py
http://www.fpune.edu.py/
[2] Universidad Internacional Tres Fronteras,
Ciudad del Este, Paraguay
fabio.lopez@uninter.edu.py
http://uninter.edu.py/

Resumen Contar con una adecuada planificación logística contribuye a mejorar el funcionamiento del sistema educativo, impactando positivamente las condiciones asociadas al aprendizaje. Este trabajo propone una nueva formulación matemática del problema de Asignación de Docentes a Establecimientos Educativos (ADEE), con un enfoque multi-objetivo para: (1) minimizar la distancia entre la residencia del docente y el establecimiento educativo, (2) maximizar la cantidad de docentes asignados al mismo establecimiento educativo y (3) maximizar la cantidad de clases dictadas por un docente en diferentes turnos. Para resolver la formulación propuesta se presenta un Algoritmo Evolutivo Multi-Objetivo (MOEA) basado en el NSGA-II. Resultados experimentales con datos reales del Departamento de Alto Paraná del Ministerio de Educación y Ciencias (MEC) de Paraguay con 457 establecimientos educativos, 2995 clases y 1808 docentes, indican mejoras significativas en la asignación.

Keywords: Optimización Multi-Objetivo, Asignación de Docentes, Computación Evolutiva, Logística Educativa, Mejora de Condiciones de Aprendizaje.

1. Introducción

Un creciente número de investigaciones confirma la importancia de las políticas dirigidas a proteger la nutrición, la salud, el desarrollo cognitivo y socio-emocional de los niños en los primeros años de vida. Sin embargo, nuevas evidencias reunidas en investigaciones recientes indican que, una vez que los niños ingresan a la escuela, ningún otro factor es tan importante como la calidad de los docentes [1].

En un análisis del sistema educativo de Paraguay como parte del proyecto "Diseño de la Estrategia de Transformación Educativa del Paraguay 2030" [2], se indica que las condiciones laborales de los docentes en las instituciones educativas distan de ser óptimas, ya que un gran número de ellos trabaja en múltiples turnos, instituciones, jornadas y niveles. Además, los mismos cumplen con un gran número de tareas fuera de lo instruccional sin tener horas de contrato para cumplirlas, lo que imposibilita el trabajo colegiado y el aprendizaje organizacional al interior de las escuelas [3].

En ese contexto, el problema abordado en este trabajo es la Asignación de Docentes a Establecimientos Educativos (ADEE). Cabe mencionar que en el sistema educativo de Paraguay, la educación formal se divide en: la Educación Inicial (EI), la Educación Escolar Básica (EEB) en tres ciclos, la Educación Media (EM) y la Educación Superior (ES) [3].

Este trabajo se enfoca en la EEB del primer y segundo ciclo (1° Grado a 6° Grado). En estos niveles el docente imparte todas las materias de la clase, es decir, una clase (e.g. 1° Grado, Sección A) cuenta con un único docente asignado.

Algunos de los inconvenientes mencionados anteriormente pueden ser reducidos con un esquema de asignación o reasignación. Por ejemplo, minimizar la cantidad de instituciones donde el docente se desenvuelva profesionalmente, puede impactar positivamente en el trabajo de colegiado y el aprendizaje organizacional.

Otro punto clave en línea con los inconvenientes mencionados anteriormente, es el tiempo que el docente dedica al traslado. Minimizar la distancia de residencia del docente con el establecimiento educativo impactaría positivamente en los siguientes puntos:

- Ahorro en gastos de transporte, con impacto financiero positivo y cuidado del ambiente al reducir las emisiones de CO_2.
- Ahorro en tiempo, con el volumen automotor en crecimiento se puede perder hasta varias horas en el tráfico. En zonas rurales, reducir la distancia podría tener un impacto aún mayor con respecto al tiempo.

Las propuestas de nuevas formulaciones matemáticas del problema ADEE que contemplen diversos aspectos mencionados para mejorar la logística en el sistema educativo representa un tema de investigación con alto impacto para países en desarrollo, donde la educación es una de las principales alternativas para lograr los objetivos clave de crecimiento. Por lo tanto, este trabajo se enfoca en proponer una nueva formulación matemática que permita mejorar algunos de los puntos citados, y de esta manera ser una herramienta potencial para mejorar la logística del sistema educativo, mejorando las condiciones asociadas al aprendizaje.

Formalmente, se puede definir el problema ADEE como:

"*Dado un conjunto de docentes con sus respectivos datos asociados y un conjunto de establecimientos educativos con sus respectivos datos asociados, asignar a los docentes en los establecimientos educativos, considerando las restricciones operativas y de recursos y optimizando simultáneamente las funciones objetivo definidas.*"

El resto de este trabajo se encuentra estructurado de la siguiente manera: En la Sección 2 se resume una revisión sistemática de la literatura asociada al problema ADEE. En la Sección 3 se presenta la nueva formulación matemática propuesta para la optimización simultánea de 3 funciones objetivo, mientras que la Sección 4 presenta el Algoritmo Evolutivo Multi-objetivo (Multi-Objective Evolutionary Algorithm, MOEA) propuesto para la resolución de la formulación propuesta. Los resultados experimentales son resumidos en la Sección 5. Las principales conclusiones se detallan en la Sección 6.

2. Revisión de la Literatura

En [4] se propone la utilización de dos algoritmos, Simulated Annealing (SA) y Tabu Search (TS), para resolver el problema de asignación de profesores. En la formulación del problema se busca minimizar la varianza total de la carga docente en una primera fase, luego se toma como entrada para una segunda fase donde se busca minimizar la varianza ponderada total de la carga de docentes. Para evaluar el rendimiento de los algoritmos propuestos, han realizado experimentos con dos conjuntos de datos reales obtenidos de una Universidad de Indonesia y algunos conjuntos de datos generados aleatoriamente. Los resultados experimentales muestran que los algoritmos considerados encuentran mejores soluciones en comparación con asignaciones manuales y un trabajo anterior que utilizó un Algoritmo Genético (AG).

En [5] se aborda el problema de la asignación de clases a profesores en universidades mediante el método Beam Search (BS) y se añaden las preferencias del profesor. Además, se desarrolla una herramienta para realizar simulaciones. Se define una función objetivo para minimizar el costo total de las asignaciones según: las preferencias individuales de los docentes y la similitud entre diferentes disciplinas. Para resolver el problema, desarrollan una herramienta que utiliza una heurística mediante búsqueda en árboles y un algoritmo voraz. Se mostró un desempeño satisfactorio para una prueba con datos reales, donde fue necesario realizar la asignación de 63 clases a 11 profesores, mejorando así la asignación manual.

En [6] se proponen varios algoritmos de aproximación para el problema de asignación de profesores en formación a las escuelas, utilizando como punto de partida el sistema educativo eslovaco y checo, donde cada profesor en formación se especializa en dos materias. En la formulación del problema tienen en cuenta la capacidad de la escuela en cada materia, donde cada profesor indica la materia y la lista de escuela aceptables. Demuestran que es relativamente sencillo proponer algoritmos de aproximación para el problema de encontrar una coincidencia con la cardinalidad máxima para una instancia determinada del problema de asignación de profesores. Concluyen con las siguientes dos preguntas abiertas: (1) ¿podrían los algoritmos refinarse para obtener un mejor límite de aproximación? y (2) ¿es posible que el problema sea APX-Completo?.

En [7] se propone utilizar AGs y Asynchronous Cooperative Parallel Search (ACoPGA) para la resolución del problema de asignación de profesores con el fin de mejorar el rendimiento, la escalabilidad y reducir el largo tiempo de ejecución asociados. Como función objetivo se normalizan y combinan 5 funciones, dando a cada una de ellas un peso para realizar las asignaciones de los docentes. Conforme una revisión de la literatura, concluyen que los AGs demostraron ser uno de los mejores algoritmos meta-heurísticos para resolver problemas de asignación de profesores. Para mejorar el tiempo de resolución realizan diversos experimentos, utilizando un esquema paralelo contra secuencial, luego paralelo asíncrono contra paralelo síncrono. Finalmente encuentran que el enfoque Asynchronous Elite CoPGA (AECoPGA) es muy eficiente para resolver casos de óptimos locales y eleva la precisión de la solución y su velocidad de búsqueda.

En [8] se desarrolla un modelo de programación lineal mixto para balancear la carga de profesores mientras se maximiza las preferencias de los profesores por clase. Para validar el modelo realizaron dos experimentos: (1) utilizando los datos del Departamento de Gestión en la Escuela de Ingeniería de Barcelona de la Universidad Politécnica de Cataluña, uno de los departamentos con mayor número de clases y profesores, y (2) con 750 instancias generadas aleatorias con patrones reales. Los resultados que obtuvieron muestran que el modelo puede resolver escenarios de hasta 40 profesores, logrando soluciones aceptables en un tiempo de cálculo reducido.

En [9] se desarrolla un sistema de recomendación basado en técnicas de minería de datos para la asignación de clases a profesores en la Educación Superior (ES). En la propuesta se difiere de los modelos de optimización tradicionales y se basa en un sistema de recomendación basado en el histórico de asignaturas dictadas por los docentes, en la evaluación del desempeño docente y el perfil de este. Se utilizó una base de datos de una Universidad de Ecuador, donde se tomaron 133000 registros acerca del perfil docente y 3000 registros correspondientes a evaluaciones estudiantiles e históricos académicos de los últimos 10 ciclos (5 años). Las recomendaciones sugeridas por el sistema fueron valoradas por 5 expertos del ámbito académico, considerando los criterios de pertinencia, coherencia y rendimiento de la recomendación obteniendo, en una escala del 1 al 5, resultados de 4.2, 4.0 y 4.2 respectivamente en promedio en cada criterio.

En [10] se realiza una revisión de la literatura sobre problemas de gestión de competencias, en particular, el Problema de Asignación de Docentes (PAD). Determinan que normalmente el PAD se resuelve antes que el problema de organizar el horario del curso. Las investigaciones relacionadas con el PAD tienen un área sin desarrollar ya que sus enfoques no permiten la síntesis de la estructura de competencias. Adicionalmente, la falta de soluciones potenciales al problema y su naturaleza NP-completo requieren la búsqueda de condiciones suficientes, cuyo cumplimiento garantice la existencia del plan de asignación docente. La búsqueda de esas condiciones se vuelve importante: determina la finalidad del trabajo, como el tiempo que consumen estas búsquedas.

En [11] desarrollan una ontología educativa para modelar la semántica de los cursos y los perfiles académicos en las universidades, con el fin de asignar al docente más adecuado para impartir un curso específico. En esta propuesta, se busca realizar el emparejamiento de los cursos con los docentes mediante el grado de afinidad de los tópicos del curso con el perfil del profesor utilizando la ontología, una propuesta diferente a los algoritmos de asignación o a las heurísticas utilizadas en la mayoría de los trabajos estudiados. Se aplica a la Universidad King Abdulaziz (KAU) en Arabia Saudita y se concentra en la Facultad de Tecnologías de la Información y la Computación (FCIT), incluidos sus tres departamentos que siguen las reglas de la Comisión Nacional de Acreditación y Evaluación Académica (NCAAA) para documentar sus datos. Aunque en la prueba no se ha asignado un número significativo de cursos y profesores, el sistema da resultados precisos.

En [12] se propone un enfoque de dos pasos para predecir las preferencias de los profesores utilizando técnicas de minería de datos y realizar la asignación de los profesores utilizando una programación lineal entera. Se vuelve a formular el problema para optimizar la preferencia de los profesores y la varianza de la carga de trabajo. Debido a la dificultad de obtener las preferencias, se realiza una estimación de estas mediante datos históricos. El modelo propuesto se valida mediante experimentos y validación manual. Tienen una desviación menor de la carga de trabajo real de los profesores respecto de la carga de trabajo objetivo y las soluciones se generaron en un tiempo significativamente menor que el proceso manual. La mayoría de las asignaciones de la sección del curso del maestro y el número correspondiente asignado de secciones son satisfactorias.

En [13] se diseñan mecanismos sin las limitaciones de la versión modificada de la Aceptación Diferida (DA), manteniendo las buenas propiedades de incentivo de este, es decir, a prueba de estrategias (lo que significa que los profesores tienen incentivos directos para informar sus preferencias con sinceridad). En esta propuesta los maestros tienen una lista estricta de preferencias y de la misma manera las escuelas cuentan con una lista de preferencia de profesores. Estos mecanismos buscan emparejar de una forma más equitativa que los métodos tradicionales. Muestran que el mecanismo de la versión modificada de DA no es justo y eficiente tanto para los maestros como para las escuelas. Confirman el desempeño de estos mecanismos alternativos, cuando el tamaño crece, se desempeñan mucho mejor en términos de eficiencia utilitaria y equidad. Por último, utilizando un conjunto de datos sobre las solicitudes de transferencia de maestros en Francia para medir las ganancias relevantes, los mecanismos alternativos generan ganancias significativas en eficiencia y equidad.

Cabe destacar que ninguno de los trabajos estudiados incluye la optimización de la distancia entre la residencia del docente y el establecimiento educativo, uno de los principales aportes de este trabajo. Adicionalmente, se destaca que un aspecto relevante identificado por el MEC de Paraguay, representa la optimización de la cantidad de docentes asignados, lo que puede ser enfocado con la optimización de docentes asignados al mismo establecimiento educativo y la cantidad de clases dictadas por un docente en diferentes turnos, siendo también esto un relevante aporte de este trabajo.

3. Formulación Matemática del Problema

En esta sección se presenta la formulación matemática con un enfoque multi-objetivo propuesta para el problema ADEE. Primeramente, se presenta una introducción a la optimización multi-objetivo, continuando con definiciones conceptuales, para posteriormente presentar la formulación compuestas por los datos de entrada, los datos de salida, el conjunto de restricciones, las funciones objetivo propuestas y se termina con un ejemplo básico.

3.1. Optimización Multi-Objetivo

Un Problema de Optimización Multi-Objetivo (POM) puede ser definido como el problema de encontrar un vector de variables de decisión que satisface las restricciones y optimiza una función vectorial, cuyos elementos representan las funciones objetivo. Estas funciones forman una descripción matemática de los criterios de rendimiento que suelen estar en conflicto entre sí. Por tanto, el término ”optimizar” significa encontrar una solución que dé valores de todas las funciones objetivo aceptables para el tomador de decisiones [14].

En [14] se expresa un problema general de optimización multi-objetivo como:

Un conjunto q de objetivos a optimizar:

$$y = f(X) = [f_1(X), f_2(X), \ldots, f_q(X)] \tag{1}$$

Sujeto a r restricciones:

$$e(X) = [e_1(X), e_2(X), \ldots, e_r(X)] \geq 0 \tag{2}$$

Con un conjunto de p variables de decisión:

$$X = [X_1, X_2, \ldots, X_p] \in \mathit{Espacio\ de\ decisión} \tag{3}$$

y el vector objetivo y:

$$y = [y_1, y_2, \ldots, y_q] \in \mathit{Espacio\ objetivo} \tag{4}$$

Un POM por lo general no tiene una única solución óptima, sino un conjunto de soluciones óptimas comprometidas, a diferencia de un problema de optimización mono-objetivo [15]. Mediante la dominancia de Pareto se puede comparar dos soluciones y determinar si una solución es mejor a otra en un contexto multi-objetivo. Esto ocurre cuando una solución es mejor o igual en cada función objetivo y estrictamente mejor en al menos uno [14].

3.2. Definiciones Conceptuales

Para comprender la formulación matemática propuesta, se define cada uno de los conceptos considerados en ésta.

- **Establecimiento:** Lugar físico que alberga una o más instituciones.
- **Institución:** Entidad habilitada para desarrollar las clases. Podría tener sede en más de un establecimiento.
- **Grado:** Corresponde al grado de enseñanza, e.g. 1° Grado, 2° Grado, entre otros.
- **Turno:** Corresponde al turno donde se imparte la clase. e.g. Turno Mañana, Turno Tarde, Turno Noche.
- **Sección:** En caso de que la misma institución, en el mismo establecimiento, tenga más de un grupo para el mismo grado y turno, es utilizada la sección para diferenciar, e.g. 1° Grado - Turno Mañana - Sección A y 1° Grado - Turno Mañana - Sección B.
- **Clase:** Compuesto por un grado, turno, sección, institución y establecimiento. Es la unidad de enseñanza donde debe ser asignado un docente para enseñar a un grupo de alumnos, e.g. 1° Grado – Turno mañana – Sección A – Institución 1 – Establecimiento 1.
- **Docente:** Profesional habilitado para impartir una clase.

3.3. Datos de Entrada

El conjunto de grados disponibles que se representa como un vector G de dimensión n_g:

$$G = \begin{bmatrix} 1 \ 2 \ldots n_g \end{bmatrix} \tag{5}$$

donde:

n_g: es la cantidad de grados.

El conjunto de turnos se representa como un vector T de dimensión n_t:

$$T = \begin{bmatrix} 1 \ 2 \ldots n_t \end{bmatrix} \tag{6}$$

donde:

n_t: es la cantidad de turnos.

El conjunto de secciones se representa como un vector S de dimensión n_s:

$$S = \begin{bmatrix} 1 \ 2 \ldots n_s \end{bmatrix} \tag{7}$$

donde:

n_s: es la cantidad de secciones.

El conjunto de instituciones se representa como un vector I de dimensión n_i:

$$I = \begin{bmatrix} 1 \; 2 \ldots n_i \end{bmatrix} \tag{8}$$

donde:
n_i: es la cantidad de instituciones.

El conjunto de establecimientos se representa como una matriz E de dimensión $(n_e \times 3)$:

$$E = \begin{bmatrix} N_1 & X_1 & Y_1 \\ \vdots & \vdots & \vdots \\ N_{n_e} & X_{n_e} & Y_{n_e} \end{bmatrix} \tag{9}$$

Cada E_k es representado por el número del establecimiento, su coordenada geográfica X_k y su coordenada geográfica Y_k:

$$E_k = \begin{bmatrix} N_k \; X_k \; Y_k \end{bmatrix} \forall k \in \{1, \ldots, n_e\}$$

donde:
N_k: número del establecimiento E_k;
X_k: coordenada geográfica X de la ubicación del establecimiento E_k;
Y_k: coordenada geográfica Y de la ubicación del establecimiento E_k;
n_e: cantidad de establecimientos.

El conjunto de docentes se representa como una matriz D de dimensión $(n_d \times 3)$:

$$D = \begin{bmatrix} N_1 & X_1 & Y_1 \\ \vdots & \vdots & \vdots \\ N_{n_d} & X_{n_d} & Y_{n_d} \end{bmatrix} \tag{10}$$

Cada D_i es representado por el número del docente, su coordenada geográfica X_i y su coordenada geográfica Y_i:

$$D_i = \begin{bmatrix} N_i \; X_i \; Y_i \end{bmatrix} \forall i \in \{1, \ldots, n_d\}$$

donde:
N_i: número del docente D_i;
X_i: coordenada geográfica X de la ubicación de la residencia del docente D_i;
Y_i: coordenada geográfica Y de la ubicación de la residencia del docente D_i;
n_d: cantidad de docentes.

El conjunto de clases activas se representa como una matriz C de dimensión $(n_c \times 5)$:

$$C = \begin{bmatrix} G_1 & T_1 & S_1 & I_1 & E_1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ G_{n_c} & T_{n_c} & S_{n_c} & I_{n_c} & E_{n_c} \end{bmatrix} \tag{11}$$

Cada C_j es representado por el Grado, Turno, Sección, Institución y Establecimiento como:

$$C_j = \begin{bmatrix} G_j \; T_j \; S_j \; I_j \; E_j \end{bmatrix} \forall j \in \{1, \ldots, n_c\}$$

donde:
G_j: Grado de la clase C_j, entonces $G_j \in G$;
T_j: Turno de la clase C_j, entonces $T_j \in T$;
S_j: Sección de la clase C_j, entonces $S_j \in S$;
I_j: Institución de la clase C_j, entonces $I_j \in I$;
E_j: Número de establecimiento de C_j, entonces $E_j \in E$;
n_c: cantidad de clases disponibles.

La matriz calculada U de distancias entre la residencia del docente y los establecimientos de dimensión $(n_d \times n_e)$ se representa como:

$$U = \begin{bmatrix} U_{1,1} & U_{1,2} & U_{1,3} & \ldots & U_{1,n_e} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ U_{n_d,1} & U_{n_d,2} & U_{n_d,3} & \ldots & U_{n_d,n_e} \end{bmatrix} \tag{12}$$

Cada posición $U_{i,k}$ representa la distancia entre la residencia de un docente i y un establecimiento k.

$$\forall i \in \{1, \ldots, n_d\} \wedge \forall k \in \{1, \ldots, n_e\}$$

donde:
$U_{i,k}$: Distancia entre la residencia del docente D_i y el establecimiento E_k.
La matriz calculada V de distancias entre establecimientos de dimensión $(ne \times ne)$ se representa como:

$$V = \begin{bmatrix} V_{1,1} & V_{1,2} & V_{1,3} & \dots & V_{1,n_e} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ V_{n_e,1} & V_{n_e,2} & V_{n_e,3} & \dots & V_{n_e,n_e} \end{bmatrix} \tag{13}$$

Cada posición $V_{k,m}$ representa la distancia entre establecimiento k y el establecimiento m.

$$\forall k \in \{1, \dots, n_e\} \wedge \forall m \in \{1, \dots, n_e\}$$

donde:
$V_{k,m}$: Distancia entre el establecimiento E_k y el establecimiento E_m.

3.4. Datos de Salida

Una solución al problema se representa por la matriz X de dimensión $(n_d \times n_c)$:

$$X = \begin{bmatrix} X_{1,1} & X_{1,2} & X_{1,3} & \dots & X_{1,n_c} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ X_{n_d,1} & X_{n_d,2} & U_{n_d,3} & \dots & X_{n_d,n_c} \end{bmatrix} \tag{14}$$

Cada posición $X_{i,j}$ representa el estado de asignado o no del docente i en la clase j.

$$\forall i \in \{1, \dots, n_d\} \wedge \forall j \in \{1, \dots, n_c\}$$

donde: $X_{i,j}$: es una variable binaria, donde 1 indica que el docente i fue asignado a la clase j, 0(*cero*) en caso contrario;

3.5. Restricciones

En esta sección se definen las restricciones que deben cumplir las soluciones factibles.

(a) Un docente asignado hasta en dos clases:
Un docente debería a lo sumo estar asignado a dos clases, que es el tiempo máximo de jornada laboral posible. Esta restricción se expresa como:

$$\sum_{j=1}^{n_c} X_{i,j} \leq 2 \; \forall i \in \{1, \dots, n_d\} \tag{15}$$

(b) Cada clase debe tener asignado un único docente:
Cada clase solo debe tener un único docente asignado y no puede quedar sin docente, esta restricción se expresa como:

$$\sum_{i=1}^{n_d} X_{i,j} = 1 \; \forall j \in \{1, \dots, n_c\} \tag{16}$$

(c) Turnos diferentes para las asignaciones de un docente:
Para los docentes con dos asignaciones, los turnos de las clases asignadas deben ser diferentes para que el docente pueda cumplir con la impartición de la clase, esta restricción se expresa como:

$$C_{j,2} \neq C_{l,2} \tag{17}$$

$$\forall i \in \{1, \dots, n_d\} \wedge \forall j \in \{1, \dots, n_c\}$$

$$\wedge \forall l \in \{1, \dots, n_c\} : j \neq l \wedge X_{i,j} = 1 \wedge X_{i,l} = 1$$

donde:
$C_{j,2}$: Turno de la clase asignada al docente;
$C_{l,2}$: El otro turno de la clase asignada al docente;
$X_{i,j}$: asignación del docente i en la clase j;
$X_{i,l}$: asignación del docente i en la clase l.

(d) Distancia máxima entre establecimientos asignados a un docente:
Los establecimientos asignados no deben distar más allá de una distancia prudencial que pueda permitir el traslado del docente en el cambio de turno.

$$P_{i,j,l} \leq D_{max} \tag{18}$$
$$\forall i \in \{1,\ldots,n_d\} \wedge \forall j \in \{1,\ldots,n_c\} \wedge \forall l \in \{1,\ldots,n_c\}$$
$$: j \neq l \wedge X_{i,j} = 1 \wedge X_{i,l} = 1$$

donde:
$C_{j,5}$: Establecimiento de la asignación uno del docente i;
$C_{l,5}$: Establecimiento de la asignación dos del docente i;
$P_{i,j,l}$: distancia $V_{C_{j,5},C_{l,5}}$ entre establecimientos asignados al docente i;
D_{max}: distancia máxima permitida entre establecimientos;
$X_{i,j}$: asignación del docente i en la clase j;
$X_{i,l}$: asignación del docente i en la clase l.

3.6. Funciones Objetivo

(a) Minimizar la distancia promedio entre la residencia del docente y el establecimiento educativo de su clase:
Este objetivo permite evaluar la afinidad de las soluciones basándose en las distancias (en KMs) menores entre la residencia del docente y el establecimiento educativo, la misma se expresa como:

$$f_1(X) = \frac{\sum_{i=1}^{n_d} \sum_{j=1}^{n_c} U_{i,C_{j,5}} * X_{i,j}}{n_c} \tag{19}$$

donde:
$U_{i,k}$: Distancia entre la residencia del docente i y el establecimiento k.

(b) Maximizar la cantidad de docentes con dos turnos dentro del mismo establecimiento:
Este objetivo permite evaluar la afinidad de las soluciones basándose en asignaciones dentro de la misma institución para el docente, la función se expresa como:

$$f_2(X) = \frac{\sum_{i=1}^{n_d} Y_i}{\sum_{i=1}^{n_d} R_i} \tag{20}$$

$$Y_i = \begin{cases} 1, & Si\ \exists X_{i,j} = 1 \wedge X_{i,l} = 1 : j \neq l \wedge C_{j,5} = C_{l,5} \\ 0, & caso\ contrario \end{cases}$$
$$R_i = \begin{cases} 1, & Si\ \exists X_{i,j} = 1 \\ 0, & caso\ contrario \end{cases}$$
$$\forall j \in \{1,\ldots,n_c\} \wedge \forall l \in \{1,\ldots,n_c\} \wedge$$

donde:
Y_i: variable binaria que indica si el docente i tiene dos asignaciones en el mismo establecimiento;
R_i: variable binaria que indica si el docente i cuenta con al menos 1 asignación, 0 (*cero*) en caso contrario.

(c) Maximizar la cantidad de aulas asignados a un docente:
Este objetivo permite evaluar la afinidad de las soluciones basándose en las asignaciones para ambos turnos disponibles del docente.

$$f_3(X) = \frac{\sum_{i=1}^{n_d} Z_i}{\sum_{i=1}^{n_d} R_i} \tag{21}$$

$$Z_i = \begin{cases} 2, & Si\ \exists X_{i,j} = 1 \wedge X_{i,l} = 1 : j \neq l \\ 1, & Si\ \exists X_{i,j} = 1 \\ 0, & caso\ contrario \end{cases}$$
$$R_i = \begin{cases} 1, & Si\ \exists X_{i,j} = 1 \\ 0, & caso\ contrario \end{cases}$$
$$\forall j \in \{1,\ldots,n_c\} \wedge \forall l \in \{1,\ldots,n_c\}$$

donde:
Z_i: variable que representa la cantidad de clases asignadas al docente i;
R_i: variable binaria que indica si el docente (i) cuenta con al menos 1 asignación, 0 (*cero*) en caso contrario.

3.7. Ejemplo Básico

En la Fig. 1 se presenta una instancia simple del problema ADEE ubicada en un plano cartesiano. La instancia esta compuesta por 4 docentes y 3 establecimientos educativos con 6 clases en total.

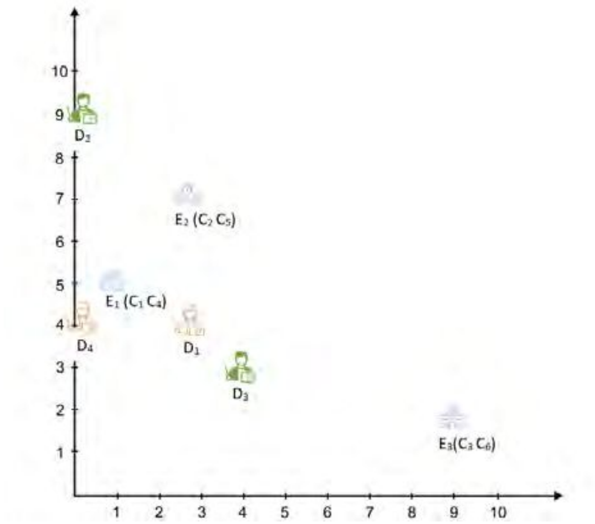

Figura 1. Establecimientos y Docentes ubicados en un plano cartesiano.

Los datos de entrada de la instancia de ejemplo se presentan a continuación G, T, S, I, E, $D\ y\ C$:

$$G = \begin{bmatrix} 1\ 2 \end{bmatrix} : 1 = Primer\ Grado \wedge 2 = Segundo\ Grado$$

$$T = \begin{bmatrix} 1\ 2 \end{bmatrix} : 1 = Turno\ Mañana \wedge 2 = Turno\ Tarde$$

$$S = \begin{bmatrix} 1\ 2 \end{bmatrix} : 1 = A \wedge 2 = B$$

$$I = \begin{bmatrix} 1\ 2\ 3 \end{bmatrix}$$

$$E = \begin{bmatrix} 1\ 1\ 5 \\ 2\ 3\ 7 \\ 3\ 9\ 2 \end{bmatrix} \quad D = \begin{bmatrix} 1\ 3\ 4 \\ 2\ 0\ 9 \\ 3\ 4\ 3 \\ 4\ 0\ 4 \end{bmatrix}$$

$$C = \begin{bmatrix} 1\ 1\ 1\ 1\ 1 \\ 1\ 1\ 1\ 2\ 2 \\ 1\ 2\ 1\ 3\ 3 \\ 2\ 2\ 1\ 1\ 1 \\ 2\ 2\ 1\ 2\ 2 \\ 2\ 1\ 1\ 3\ 3 \end{bmatrix}$$

La Clase 1 C_1: 1° Grado, Turno Mañana, Sección A de la Institución A y del Establecimiento A.
Una solución factible X, que cumple con todas las restricciones, se muestra a continuación:

$$X = \begin{bmatrix} 0\ 1\ 0\ 0\ 1\ 0 \\ 0\ 0\ 0\ 0\ 0\ 0 \\ 0\ 0\ 1\ 0\ 0\ 1 \\ 1\ 0\ 0\ 1\ 0\ 0 \end{bmatrix}$$

En esta solución factible X, al Docente 1 fue asignado a la Clase 2 del Establecimiento 2 y la Clase 5 del Establecimiento 2, mientras que al Docente 2 no se le ha asignado ninguna clase. Por otro lado, al Docente 3 se le ha asignado la Clase 3 del Establecimiento 3 y la Clase 6 del Establecimiento 3; y por último, al Docente 4 se le ha asignado la Clase 1 del Establecimiento 1 y la Clase 4 del Establecimiento 1.

4. Algoritmo Propuesto

En esta sección se presenta el Algoritmo Evolutivo Multi-Objetivo propuesto para la resolución de la formulación matemática del problema ADEE (ver Sección 3).

Los Algoritmos Evolutivos (AE) son métodos de búsqueda que se inspiran en la selección natural y la supervivencia del más apto en el mundo biológico. Los AE difieren de las técnicas de optimización más tradicionales porque involucran una búsqueda de una población de soluciones. Los AEs son particularmente adecuados para resolver problemas de optimización multi-objetivo, ya que consideran un conjunto de posibles soluciones (población) [16].

En este trabajo, se propone un MOEA basado en el NSGA-II propuesto por Kalyanmoy Deb en [15]. Para la utilización de un MOEA es necesario representar una solución al problema en una estructura de cadena denominada cromosoma. Para el problema ADEE, se propone una cadena de números enteros, donde cada gen (posición en la cadena) es una clase a ser asignada y cada alelo (valor que puede tomar el gen) representa al docente asignado a dicha clase.

En la Fig. 2 se muestra el cromosoma para la solución X del Ejemplo Básico (ver Sección 3).

Figura 2. Cromosoma propuesto para la solución X (ver Sección 3).

Cada cromosoma (representación de una solución) puede ser evaluado mediante las funciones objetivo definidas. Al valor de esta evaluación se le denomina fitness (aptitud) y es utilizado por los algoritmos evolutivos para determinar que individuos (soluciones) continúan en las siguientes generaciones.

A continuación, se presenta el Algoritmo 1 con el pseudo-código del MOEA propuesto para la resolución de la formulación matemática propuesta para el problema ADEE.

Algoritmo 1 - MOEA propuesto para resolver la formulación matemática para el problema ADEE (ver Sección 3).

Entrada: $G, T, S, I, E, C, D, U, V, N_{gen}$
Salida: Conjunto Pareto (Soluciones No Dominadas)
1: $P \leftarrow$ Población Aleatoria Factible
2: Evaluar P
3: **mientras** $N_{gen} \neq 0$ **hacer**
4: $\quad Q \leftarrow$ Seleccionar individuos de P según NSGA-II
5: $\quad Q \leftarrow$ Aplicar operador de cruzamiento
6: $\quad Q \leftarrow$ Aplicar operador de mutación
7: $\quad Q \leftarrow$ Reparar individuos
8: $\quad P \leftarrow$ Seleccionar individuos de $P + Q$ según NSGA-II
9: $\quad N_{gen} \leftarrow N_{gen} - 1$
10: **fin mientras**

Con la representación del cromosoma y la evaluación de su fitness es posible generar una población inicial de individuos (Paso 1-2 de Algoritmo 1) y mediante los operadores genéticos de selección, cruzamiento y mutación es posible ir mejorando dichas soluciones en cada generación.

El operador de selección es la estrategia utilizada para seleccionar que individuos sobreviven en la siguiente generación (Paso 4 y 8 de Algoritmo 1). El operador de cruzamiento (Paso 5 de Algoritmo 1) genera nuevos individuos denominados descendencia en base a combinaciones de los individuos actuales denominados padres.

Por último, el operador de mutación (Paso 6 de Algoritmo 1) suele intercambiar aleatoriamente un gen o más de uno en base a una probabilidad. Estos operadores podrían producir soluciones no factibles, por lo que normalmente se utiliza un operador de reparación (Paso 7 de Algoritmo 1) que ajusta la descendencia a soluciones factibles.

5. Solución Propuesta y Resultados Experimentales

En esta sección se presenta la solución propuesta para la aplicación del algoritmo propuesto (ver Sección 4) al caso del sistema educativo de Paraguay, el conjunto de datos considerado, detalles sobre la implementación de la solución y los resultados experimentales.

5.1. Diseño de la Solución Propuesta

Para la aplicación del algoritmo propuesto (ver Sección 4) al caso del sistema educativo de Paraguay, se propone la solución detallada en la Fig. 3. Para las fuentes de datos, se han obtenidos 3 orígenes heterogéneos con una variedad de formatos (característica de variedad de Big Data): (1) el Sistema de Gestión del MEC (SIGMEC), para datos **anonimizados** de los docentes, (2) archivos en formato *JSON* para las clases y (3) el Portal de Datos Abiertos del MEC [17] para establecimientos e instituciones. A continuación, estas fuentes de datos fueron consolidadas en un almacén de datos, mediante un proceso de extracción, transformación y carga (ETL). Por último, el algoritmo propuesto utilizó el almacén de datos para calcular y obtener los resultados de las asignaciones de los docentes a los establecimientos educativos en un Conjunto Pareto P, compuesto por soluciones no dominadas para soporte a la toma de decisión correspondiente.

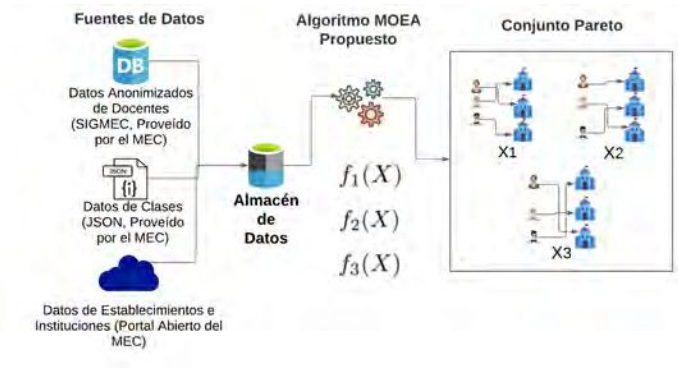

Figura 3. Diseño de la solución propuesta para Paraguay.

5.2. Conjunto de Datos Considerado

Se han utilizado datos de 2020 proveídos por el Ministerio de Educación y Ciencias del Paraguay (MEC). Si bien se ha diseñado una solución para el sistema educativo de Paraguay, los experimentos de este trabajo se han enfocado a un sub-conjunto de datos del Departamento de Alto Paraná del Paraguay. Características de los datos considerados en los experimentos están resumidas en la Tabla 1.

Tabla 1. Características del Conjunto de Datos de Alto Paraná.

Característica	Cantidad
Establecimientos	457
Docentes	1808
Clases	2995
Grados	6 (1° a 6°)
Turnos	2 (Mañana y Tarde)

Para calcular las coordenadas de la residencia de los docentes se ha utilizado el servicio de Geocoding API de Google. Se proveyó los datos en el siguiente formato: "*País, Departamento, Localidad, Distrito y Dirección*retornando las coordenadas en el siguiente formato: "*latitud,longitud*".

Para verificar la precisión del cálculo de las coordenadas se ha utilizado el servicio de Maps JavaScript API de Google. Se ubicaron por cada establecimiento y docente, un marcador dentro del mapa. Los docentes que no se encontraban dentro de los límites del departamento Alto Paraná fueron ajustados en un proceso manual. En la Fig. 4 se puede visualizar como los establecimientos y docentes quedaron dentro de los límites del Departamento Alto Paraná del Paraguay.

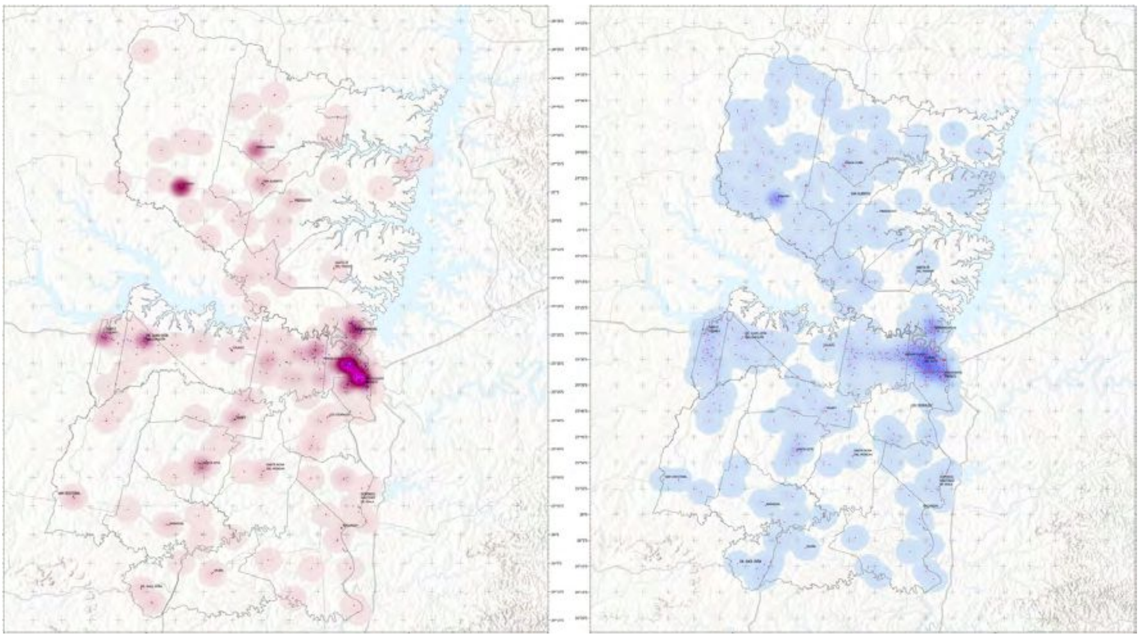

Figura 4. Docentes (izquierda) y Establecimientos (derecha) en Alto Paraná.

Para calcular la distancia entre establecimientos y la residencia de los docentes se utilizó el servicio geopy API, que utiliza la distancia geodésica entre dos coordenadas.

5.3. Implementación de la Solución

Para la implementación de la solución propuesta y para la ejecución de los experimentos se utilizó el siguiente conjunto de herramientas: Lenguaje de Programación Python[1] 3.9.11, Base de Datos PostgreSQL[2] 9.16, Framework de Optimización Multi-Objetivo pymoo[3] 0.5.0, librería geopy[4] 2.2.0 y el entorno de desarrollo Visual Studio Code[5] 1.66.2.

La selección de pymoo se basó en el estudio comparativo de frameworks resumido en la Tabla 2 presentado en [18].

Tabla 2. Comparativa de Frameworks de Optimización Multi-Objetivo presentada en [18].

Nombre	Licencia	Foco en Multi-Objetivo	Python Puro	Visualización	Toma de Decisión
jMetalPY	MIT	✓	✓	✓	–
PyGMO	GPL-3.0	✓	–	–	–
Platypus	GPL-3.0	✓	✓	–	–
DEAP	LGPL-3.0	–	✓	–	–
Inspyred	MIT	–	✓	–	–
pymoo	Apache 2.0	✓	✓	✓	✓

Con respecto al entorno utilizado en los experimentos, fue utilizado una computadora de escritorio con las siguientes características: Sistema Operativo Windows 10 Pro for Workstations, Procesador Intel(R) Xeon(R) W-2145 CPU @ 3.70GHz de 8 Núcleos y Memoria RAM de 32GB.

Con el objetivo de dar reproducibilidad al trabajo, el código fuente, conjunto de datos y resultados experimentales se encuentran disponibles en línea[6].

[1] https://www.python.org/
[2] https://www.postgresql.org/
[3] https://pymoo.org/
[4] https://geopy.readthedocs.io/en/stable/
[5] https://code.visualstudio.com/
[6] https://github.com/horaciov/assign-teacher

5.4. Resultados Experimentales

Por la naturaleza del MOEA propuesto, se han realizado 10 ejecuciones del algoritmo para el conjunto de datos descrito (ver Sección 5.2) con una población inicial de 100 individuos, tamaño de población de 100 individuos y 100 generaciones.

Datos relacionados a la cantidad de soluciones no dominadas obtenidas en cada ejecución y el tiempo de ejecución del algoritmo son resumidos en la Tabla 3.

Tabla 3. Cantidad de Soluciones No Dominadas y Tiempos de Ejecución.

Corrida	Cantidad de Soluciones	Tiempo de Ejecución (HH:mm)
1	95	13:35
2	99	16:20
3	95	15:30
4	92	14:24
5	80	13:51
6	50	18:00
7	100	16:28
8	88	13:55
9	68	15:34
10	100	12:38

Considerando los resultados de las mencionadas ejecuciones del MOEA propuesto, las soluciones encontradas fueron combinadas en un único Conjunto Pareto P. De manera a tener un valor de solución de referencia del mencionado Conjunto Pareto P, se calculó el promedio de los valores de las funciones objetivo, representado por $y(\overline{X})$. Adicionalmente, y a modo de tener una comparativa con las solución actual del sistema educativo de Paraguay, se calculó el valor de las funciones objetivo de la solución proveída por el MEC, representado por $y(X_{MEC})$. Finalmente, para tener una idea del valor óptimo que las soluciones podrían llegar a tener, y solo a modo de referencia, se presenta también los valores de las funciones objetivo para una solución óptima teórica, representado por $y(X_{OPT})$. Los mencionados valores son resumidos a continuación en la Tabla 4.

Tabla 4. Vectores Objetivo: $y(X_{OPT})$, $y(X_{MEC})$ e $y(\overline{X})$

Vector Objetivo	Min $f_1(X)$	Max $f_2(X)$	Max $f_3(X)$
$y(X_{OPT})$	0.0000	1.0000	2.0000
$y(X_{MEC})$	7.6679	0.4850	1.5453
$y(\overline{X})$	6.3462	0.4533	1.9007

Como uno de los principales hallazgos de los experimentos realizados, se puede destacar que la asignación actual de docentes a establecimientos educativos en el sistema educativo de Paraguay X_{MEC} es una solución no dominada con respecto a las soluciones del Conjunto Pareto P. Sin embargo, considerando que una muy aceptada estrategia de selección entre soluciones no dominadas desde un Conjunto Pareto es utilizar el operador de preferencia, se destaca que las soluciones del Conjunto Pareto P, encontradas por el algoritmo propuesto, son preferidas en un 100 % a la solución X_{MEC}, es decir, siempre son mejores en 2 objetivos ($f_1(X)$ y $f_3(X)$. En ese contexto, es importante recordar que una solución no dominada A es preferida a otra B ($A \succ_p B$), si A es mejor en más objetivos B [19].

En la Fig. 5 se muestran los vectores objetivo de las soluciones no dominadas del Conjunto Pareto P, consolidadas de las 10 corridas del algoritmo propuesto. Además, se muestran los vectores objetivo de la solución actual del sistema educativo de Paraguay X_{MEC} y el vector objetivo $y(\overline{X})$. Se visualiza que todas las soluciones encontradas son mejores en las funciones objetivo $f_1(X)$ y $f_2(X)$, con respecto a la solución X_{MEC}.

Considerando los resultados experimentales, y con una comparativa de los valores de $y(X_{MEC})$ con $y(\overline{X})$, a continuación se presentan hallazgos relacionados a cada función objetivo propuesta (ver Sección 3):

- Con respecto a $f_1(X)$: se observa una mejora en promedio del 17.23 % de $y(\overline{X})$ sobre $y(X_{MEC})$, lo que implica una reducción significativa de la distancia promedio entre la residencia de los docentes y los establecimientos educativos.

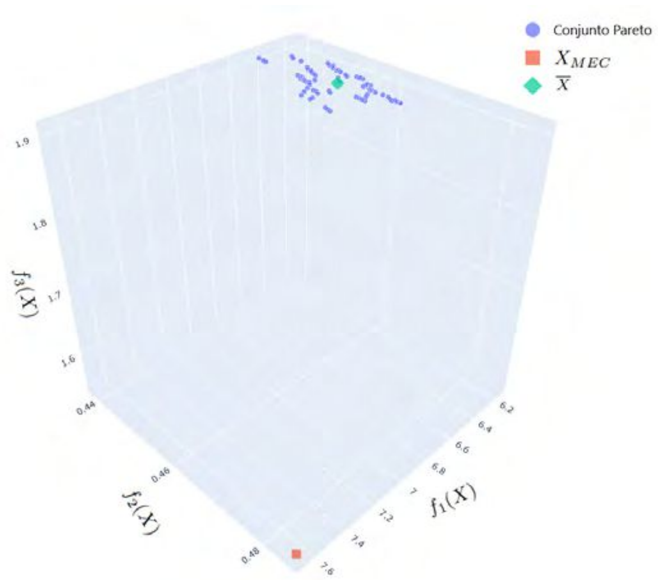

Figura 5. Resumen de Vectores Objetivo de Resultados Experimentales.

- Con respecto a $f_2(X)$: se observa que $y(X_{MEC})$ es mejor que $y(\overline{X})$ en promedio en un 6.53 %. Esto se encuentra relacionado con la mejora que se obtuvo en $f_3(X)$. Al aumentar la cantidad de docentes con ambos turnos, se ve una dificultad de encontrar o asignar docentes en el mismo establecimiento ($f_2(X)$).
- Con respecto a $f_3(X)$: se observa una mejora en promedio del 22.99 % de $y(\overline{X})$ sobre $y(X_{MEC})$, lo que permite que con una cantidad significativamente menor de docentes, se pueda atender la misma cantidad de clases ofertadas por los establecimientos educativos. Aproximadamente, se puede asignar con un 13 % menos de docentes la misma cantidad de clases, esto representa 235 docentes menos para nuestro caso. Esto resulta en un mejor aprovechamiento de los recursos.

6. Conclusiones y Trabajos Futuros

En este trabajo se ha propuesto una nueva formulación matemática para el problema ADEE, que permite estudiarlo en un contexto de optimización multi-objetivo (ver Sección 3). Se ha propuesto un Algoritmo Evolutivo Multi-Objetivo (MOEA) para validar la formulación propuesta y resolver el problema en tiempos razonables (ver Sección 4).

Resultados experimentales obtenidos con datos reales del sistema educativo de Paraguay, específicamente para el Departamento de Alto Paraná, muestran que que la propuesta representa una alternativa válida para resolver el problema ADEE, con mejoras significativas que podrían mejorar las condiciones de aprendizaje. En ese sentido, se redujo la distancia entre la residencia del docente y el establecimiento educativo, así como se mejoró la asignación para que el docente tenga asignados dos cursos, impactando así positivamente en su calidad de vida, reduciendo el tiempo de traslado y los gastos en el medio de transporte. Esto, por lo tanto, podrá contribuir a que puedan dar una clase de calidad a sus alumnos. Se consiguió optimizar la utilización de recursos, reduciendo la cantidad de docentes necesarios para atender la cantidad de clases ofertadas, en promedio en un 13 %.

En el contexto de este trabajo, se realizarán otros experimentos con los datos de los demás departamentos del Paraguay. Además, se proponen como trabajos futuros:

- Comparar experimentalmente el MOEA propuesto con otros Algoritmos Evolutivos Multi-Objetivo.
- Extender la formulación matemática propuesta para incluir no solo la distancia entre la residencia del docente al establecimiento educativo, sino también el tiempo de traslado como un criterio significativo.
- Proponer otras funciones objetivo que permitan mejorar la formulación matemática propuesta a otros escenarios de asignación de docentes, ya en un contexto de optimización de muchos objetivos (*Many-Objective Optimization*).

Referencias

[1] B. Bruns and J. Luque, "Foro sobre desarrollo de américa latina profesores excelentes cómo mejorar el aprendizaje en américa latina y el caribe," p. 76, 2014.

[2] "Diseño de la estrategia de transformación educativa del paraguay 2030," http://www.feei.gov.py/, accedido: 11/05/2022.

[3] G. N. del Paraguay, "Análisis del sistema educativo nacional," p. 64, 2021.

[4] A. Gunawan and K. Ng, "Solving the teacher assignment problem by two metaheuristics," *International Journal of Information and Management Sciences*, vol. 22, pp. 73–86, 03 2011.

[5] A. T. D. Azevedo, A. F. S. M. Ohata, J. A. Amorim, and P. M. Gustavsson, "Assigning classes to teachers in universities via mathematical modelling: Using Beam Search method and simulation in Java," pp. 577–586, 2013.

[6] K. Cechlárová, P. Eirinakis, T. Fleiner, D. Magos, I. Mourtos, E. Oce, and I. M. Preprint, "Approximation algorithms for the teachers assignment problem," 2014.

[7] E. Turki, "Solving teacher assignment problem by asynchronous cooperative parallel genetic algorithm," in *ICFCCS*, 05 2014.

[8] B. Domenech and A. Lusa, "A MILP model for the teacher assignment problem considering teachers' preferences," *European Journal of Operational Research*, vol. 249, no. 3, pp. 1153–1160, 2016.

[9] F. Pesántez-Avilés, D. Calle-López, V. Robles-Bykbaev, M. Rodas-Tobar, and C. Vásquez-Vásquez, "A recommender system based on data mining techniques to support the automatic assignment of courses to teachers in higher education," in *2017 International Conference on Information Systems and Computer Science (INCISCOS)*, 2017, pp. 231–236.

[10] E. Szwarc, I. Bach-Dąbrowska, and G. Bocewicz, "Competence management in teacher assignment planning," in *Information and Software Technologies*, R. Damaševičius and G. Vasiljevienė, Eds. Cham: Springer International Publishing, 2018, pp. 449–460.

[11] G. Ashour, A. Al-Dubai, and I. Romdhani, "Ontology-based course teacher assignment within universities," *International Journal of Advanced Computer Science and Applications*, vol. 11, 01 2020.

[12] R. Tejada and I. A. Martinez, "A two-step approach involving forecasting preferences integrating curriculum, rank, educational attainment and interest, and assignment to shorten teacher-course assignment process," in *2020 IEEE World Conference on Engineering Education (EDUNINE)*, 2020, pp. 1–6.

[13] J. Combe, O. Tercieux, and C. Terrier, "The Design of Teacher Assignment: Theory and Evidence," *The Review of Economic Studies*, 02 2022.

[14] C. A. C. Coello, G. B. Lamont, and D. A. V. Veldhuisen, *Evolutionary algorithms for solving multi-objective problems.* Springer, 2007.

[15] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: NSGA-II," *IEEE Transactions on Evolutionary Computation*, vol. 6, no. 2, pp. 182–197, 2002.

[16] M. C. Bhuvaneswari, "Application of evolutionary algorithms for multi-objective optimization in vlsi and embedded systems," *Application of Evolutionary Algorithms for Multi-Objective Optimization in VLSI and Embedded Systems*, pp. 1–174, 1 2015.

[17] "Portal de datos abiertos del mec," https://datos.mec.gov.py/, accedido: 01/03/2022.

[18] J. Blank and K. Deb, "pymoo: Multi-objective optimization in Python," *IEEE Access*, vol. 8, pp. 89 497–89 509, 2020.

[19] F. Talavera, J. Crichigno, and B. Barán, "Policies for dynamical multiobjective environment of multicast traffic engineering," in *IEEE ICT*, 2005.