



Universidad de Castilla-La Mancha
Escuela Superior de Ingeniería Informática

Trabajo Fin de Grado
Grado en Ingeniería Informática
Computación

**Optimización Multiobjetivo de la asignación
docente en el Grado de Ingeniería Informática
para evitar superposiciones y satisfacer
preferencias de horarios de los estudiantes**

José Ángel Serrano Pardo

Julio, 2025



TRABAJO FIN DE GRADO

Grado en Ingeniería Informática

Computación

Optimización Multiobjetivo de la asignación docente en el Grado de Ingeniería Informática para evitar superposiciones y satisfacer preferencias de horarios de los estudiantes

Autor: José Ángel Serrano Pardo

Tutor: José Miguel Puerta Callejón

Julio, 2025

*A mis padres,
por su dedicación silenciosa
y su confianza infinita en mí.*

Declaración de autoría

Yo, José Ángel Serrano Pardo , con DNI 49800390P, declaro que soy el único autor del trabajo fin de grado titulado “Optimización Multiobjetivo de la asignación docente en el Grado de Ingeniería Informática para evitar superposiciones y satisfacer preferencias de horarios de los estudiantes”, que el citado trabajo no infringe las leyes en vigor sobre propiedad intelectual, y que todo el material no original contenido en dicho trabajo está apropiadamente atribuido a sus legítimos autores.

Albacete, a 9 de Julio de 2025

Fdo.: José Ángel Serrano Pardo

Resumen

Los algoritmos genéticos son una rama de la inteligencia artificial donde se simula el proceso de la evolución biológica. Se usan métodos de búsqueda heurística que utilizan procesos de selección, cruzamiento y mutación para generar soluciones óptimas o cercanas a las óptimas a problemas complejos que son difíciles de resolver mediante técnicas convencionales. A través de iteraciones sucesivas, una población inicial de soluciones potenciales evoluciona hacia una solución óptima o cercana, adaptándose al entorno definido por una función fitness específica. En este estudio, se adapta un algoritmo genético para abordar la optimización multiobjetivo en la asignación de grupos para las matrículas en el Grado de Ingeniería Informática de la Universidad de Castilla-La Mancha. El desafío incluye múltiples objetivos, como maximizar la satisfacción de los alumnos, minimizar el número de solapes entre los horarios de las matrículas de un alumno, maximizar la cohesión de grupos de teoría y práctica para asignaturas del mismo curso en los alumnos, y maximizar el equilibrio del número de alumnos en los grupos de teoría y prácticas en cada asignatura. El algoritmo genético diseñado utiliza una codificación específica para representar las variables del problema y operadores genéticos adaptados para mantener la diversidad genética en la población, permitiendo así una exploración efectiva del espacio de soluciones.

Agradecimientos

En primer lugar, quiero expresar mi más profundo agradecimiento a mis padres y al resto de mi familia por todo el apoyo que me han brindado a lo largo de este camino. Su paciencia, comprensión y ayuda en los momentos más difíciles han sido fundamentales para llegar hasta aquí.

También quiero dar las gracias a mis amigos, tanto a los de siempre como a aquellos que he tenido la suerte de conocer durante este proceso. Gracias por acompañarme, por ser compañeros de batalla y por hacer que esta etapa fuera más llevadera y, sobre todo, inolvidable. Sin vosotros, nada de esto habría sido igual.

Por último, me gustaría agradecer sinceramente a mi tutor, José Miguel Puerta Callejón, por su dedicación, orientación y valiosos consejos durante la elaboración de este trabajo. Su guía ha sido esencial en todo el desarrollo del proyecto.

Índice general

1	Introducción	1
1.1	Motivación	1
1.2	Objetivos	2
1.3	Competencias	2
1.4	Estructura del documento	3
2	Estado del arte	5
2.1	Escuela Superior de Ingeniería Informática de Albacete	5
2.2	Plan de Estudios del Grado en Ingeniería Informática	6
2.3	Metaheurísticas evolutivas	11
2.4	Algoritmos Genéticos	12
2.5	Algoritmos Genéticos Multiobjetivo	15
2.5.1	NSGA-III	17
3	Desarrollo	21
3.1	Modelización del problema	21
3.1.1	Representación de los individuos	21
3.1.2	Objetivos	23
3.2	Algoritmo genético	27
3.2.1	Función fitness	27
3.2.2	Población inicial	28
3.2.3	Selección	29
3.2.4	Cruce	30
3.2.5	Mutación	32

3.2.6	<i>Sustitución de la población</i>	32
3.2.7	<i>Condición de parada</i>	33
3.2.8	<i>Resultados Ejecución</i>	34
3.3	NSGA-III	35
3.3.1	<i>Población Inicial</i>	35
3.3.2	<i>Puntos de referencia</i>	36
3.3.3	<i>Clasificación no dominada</i>	37
3.3.4	<i>Asignación a puntos de referencia</i>	37
3.3.5	<i>Selección</i>	38
3.3.6	<i>Cruce y Mutación</i>	39
3.3.7	<i>Reemplazo</i>	40
3.3.8	<i>Condición de parada</i>	40
3.3.9	<i>Resultados de la Ejecución</i>	41
4	Experimentación	43
4.1	Introducción	43
4.2	Entorno de ejecución	44
4.3	Evaluación de métodos del algoritmo genético	44
4.4	Evaluación de hiperparámetros del algoritmo genético	49
4.5	Evaluación del algoritmo NSGA-III	53
5	Desarrollo de la aplicación web	63
5.1	Introducción	63
5.2	Diseño de la aplicación	63
5.3	Flask	64
5.4	Arquitectura del proyecto	64
5.5	Aplicación web	65
6	Conclusiones	73
6.1	Conclusiones	73
6.2	Trabajo futuro	74
A	Anexos	75
A.1	Repositorio del código y scripts	75
	Referencia bibliográfica	78

Índice de figuras

2.1	Horario 1º cuatrimestre Grupo 1ºA	9
2.2	Horario 1º cuatrimestre Grupo 2ºA	9
2.3	Flujo básico de un algoritmo genético. Fuente: https://www.cs.us.es/~fsancho/Blog/posts/Algoritmos_Geneticos.md.html	14
2.4	Visualización del frente de Pareto en el espacio objetivo.	16
2.5	Flujo básico de NSGA-III. Fuente: https://www.researchgate.net/figure/Flowchart-of-NSGA-III_fig3_329986923	19
3.1	Representación del genotipo de dos estudiantes. Cada pareja de valores representa el grupo de teoría y subgrupo de prácticas de una matrícula variable.	22
4.1	Evolución del mejor fitness promedio para la combinación Torneo Uniforme Reemplazo	47
4.2	Evolución del mejor fitness promedio para la combinación Fitness Uniforme Reemplazo	47
5.1	Estructura del proyecto	65
5.2	Página principal de la aplicación web	67
5.3	Pantalla de carga de la aplicación web	68
5.4	Parte superior de la pantalla de resultados del algoritmo clásico	69
5.5	Parte inferior de la pantalla de resultados del algoritmo clásico	69
5.6	Parte superior de la pantalla de resultados NSGA-III	70
5.7	Parte inferior de la pantalla de resultados NSGA-III	71

Índice de tablas

4.1	Resumen de resultados promediados tras 10 ejecuciones por combinación	46
4.2	Resultados segundo script de experimentación	52
4.3	Primeras 10 combinaciones del tercer script de experimentación	56
4.4	Elementos frente de pareto de una combinación	57
4.5	Primeras 10 combinaciones del tercer script de experimentación con mayor diversidad	57
4.6	Mejor combinación encontrada para NSGA-III	61

1. Introducción

1.1. Motivación

La gestión de matrículas para asignar grupos es uno de los procesos esenciales y más complejos en cualquier universidad, ya que actualmente existe una gran cantidad de alumnos y no todos los grupos son iguales, ya que tienen distintos horarios, siendo algunos por la mañana y otros por la tarde, además de que cada alumno tiene preferencias distintas, ya sea por qué prefiere un horario temprano o por qué le complementa mejor con las asignaturas que tiene u otras actividades.

Actualmente, los estudiantes analizan muy bien los horarios de las asignaturas antes del proceso de matriculación para calcular cuál sería su configuración óptima y tener el menor número de solapes de horarios y menor pérdida de tiempo entre clases, aunque no todos consiguen matricularse en sus grupos ideales debido a que los alumnos no realizan la matrícula a la vez, sino que se hace en distintos días y primero la realizan los estudiantes con mayor nota. Esto puede causar frustración en los alumnos que no consiguen matricularse en el grupo que querían debido a que no quedaban plazas, pudiendo afectar a su rendimiento académico u organización personal.

Después de realizar las matrículas, se tiene que seleccionar el grupo de prácticas. En este caso, sí se realiza al mismo tiempo para todos los alumnos, por lo que suele ser complicado obtener el grupo que se desea, añadiendo más dificultad a los alumnos en su tarea de conseguir un horario óptimo, teniendo en cuenta que los grupos tienen que tener una cantidad de alumnos limitada y es imposible satisfacer todas las preferencias de los alumnos de forma individual.

Por eso nace la idea de desarrollar un sistema con inteligencia artificial que, recibiendo las matrículas que han realizado los alumnos y usando procesos de optimización, haga una asignación de grupos de forma que se intente favorecer a los alumnos, teniendo el menor número de solapes de horarios en asignaturas de un alumno y reduciendo el tiempo muerto entre clases para cada alumno. Todo esto manteniendo el equilibrio de alumnos por grupo en las asignaturas e intentando mantener cierto parecido con la configuración original.

1.2. Objetivos

1. Analizar el problema de la asignación docente:
 - Identificar y analizar los principales factores y restricciones que afectan a la asignación docente, cómo las preferencias de los alumnos, las restricciones horarias, las plazas limitadas en los grupos de teoría y prácticas.
2. Diseñar un modelo de optimización multiobjetivo:
 - Desarrollar un modelo matemático que represente el problema de la asignación docente, donde estén explicados todos los parámetros a optimizar y sus restricciones.
3. Implementar un algoritmo genético adaptado:
 - Configurar e implementar un algoritmo genético específico para resolver el modelo de optimización, incluyendo la codificación de soluciones, la función de fitness, y los operadores de selección, cruzamiento y mutación.
4. Realizar experimentos y evaluar el rendimiento del algoritmo:
 - Diseñar y llevar a cabo una batería de experimentos para probar el algoritmo genético con distintas configuraciones y en escenarios diferentes, comparando el rendimiento y la evolución del tiempo de ejecución.
5. Analizar los resultados y validar el modelo:
 - Evaluar los resultados obtenidos, analizando la eficacia y eficiencia del algoritmo genético en la optimización de la asignación docente, usando varias métricas para determinar cuál es la mejor configuración que puede tener este algoritmo genético para obtener la mejor solución.
6. Desarrollar una aplicación web para la ejecución de los algoritmos genéticos:
 - Implementar una interfaz gráfica sencilla que permita a los usuarios cargar los ficheros de entrada, usar los algoritmos genéticos y poder escoger entre una configuración manual o la óptima determinada durante la fase de experimentación, y visualizar los resultados finales, incluyendo la evolución del algoritmo y la ubicación de los ficheros generados.

1.3. Competencias

En este proyecto se han trabajado las siguientes competencias.

- Comunes a la rama de la informática

- **[CO6]** Conocimiento y aplicación de los procedimientos algorítmicos básicos de las tecnologías informáticas para diseñar soluciones a problemas, analizando la idoneidad y complejidad de los algoritmos propuestos.
 - **[CO7]** Conocimiento, diseño y utilización de forma eficiente de los tipos y estructuras de datos más adecuados para la resolución de un problema.
 - **[CO14]** Conocimiento y aplicación de los principios fundamentales y técnicas básicas de la programación paralela, concurrente, distribuida y de tiempo real.
 - **[CO15]** Conocimiento y aplicación de los principios fundamentales y técnicas básicas de los sistemas inteligentes y su aplicación práctica.
- Tecnología específica. Computación.
- **[CM3]** Capacidad para evaluar la complejidad computacional de un problema, conocer estrategias algorítmicas que puedan conducir a su resolución y recomendar, desarrollar e implementar aquella que garantice el mejor rendimiento de acuerdo con los requisitos establecidos.
 - **[CM4]** Capacidad para conocer los fundamentos, paradigmas y técnicas propias de los sistemas inteligentes y analizar, diseñar y construir sistemas, servicios y aplicaciones informáticas que utilicen dichas técnicas en cualquier ámbito de aplicación.
 - **[CM5]** Capacidad para adquirir, obtener, formalizar y representar el conocimiento humano en una forma computable para la resolución de problemas mediante un sistema informático en cualquier ámbito de aplicación, particularmente los relacionados con aspectos de computación, percepción y actuación en ambientes o entornos inteligentes.

1.4. Estructura del documento

La estructura de los capítulos de la memoria es la siguiente.

- Introducción
- Presenta el contexto y la motivación del estudio, describe los desafíos de la asignación docente y establece los objetivos generales y específicos del trabajo.
- Estado del arte
- Revisa el contexto del Grado Universitario en Ingeniería Informática, el funcionamiento del sistema de matrículas y la organización de los grupos. Proporciona una visión general de las metaheurísticas evolutivas y una revisión detallada de los algoritmos genéticos y su aplicación en problemas similares.
- Desarrollo

-
- Detalla el diseño del algoritmo genético, incluyendo la representación de soluciones, la función de fitness y los operadores genéticos utilizados. Describe la implementación del algoritmo en un entorno de desarrollo específico.
 - Experimentación
 - Describe la configuración experimental, los escenarios de prueba, los métodos de evaluación y presenta los resultados obtenidos. Incluye una discusión sobre la eficacia del algoritmo
 - Desarrollo de la aplicación web
 - Explica el diseño, arquitectura e implementación de una interfaz web desarrollada para facilitar la interacción con los algoritmos genéticos implementados. Describe las tecnologías utilizadas, la estructura del proyecto, y cómo los usuarios pueden configurar y ejecutar los algoritmos desde la web.
 - Conclusiones
 - Resume los hallazgos principales del estudio, discute la contribución del trabajo y su impacto en la mejora de la asignación docente. Propone posibles mejoras al algoritmo y sugiere direcciones para futuras investigaciones.

2. Estado del arte

En este capítulo se describe el funcionamiento de las matrículas en la Escuela Superior de Ingeniería Informática de Albacete y el tipo de algoritmos genéticos que se pueden usar para optimizar la asignación de grupos en las matrículas de los estudiantes.

2.1. Escuela Superior de Ingeniería Informática de Albacete

Las titulaciones en la Escuela Superior de Ingeniería Informática de Albacete se inician en el año 1985 con los estudios de Diplomatura en Informática. Posteriormente, en el curso académico 1992-93, se implantan los títulos de Ingeniería Técnica en Sistemas e Ingeniería Técnica en Gestión, ambos con una duración de tres años. En el curso 1999-2000 se incorpora la titulación de Ingeniería Informática de cinco años. Estas tres titulaciones permanecen vigentes hasta el curso 2010-11, momento en el que se introduce el actual Grado en Ingeniería Informática, adaptado al Espacio Europeo de Educación Superior [of Education, 1999], sustituyendo así a las titulaciones anteriores.

Actualmente, además del Grado en Ingeniería Informática, la Escuela ha ampliado su oferta formativa con nuevas titulaciones. Entre ellas se encuentra el doble grado en Ingeniería Informática y Administración y Dirección de Empresas, que combina competencias en el ámbito de la informática y la gestión empresarial. En cuanto a estudios de posgrado, se imparten el Máster Universitario en Ingeniería Informática y el programa de Doctorado en Tecnologías Informáticas Avanzadas, junto con diversos títulos propios como el Máster en Sistemas Informáticos Embarcados y el Máster en Ciberseguridad y Seguridad de la Información. A esta oferta se suman recientemente nuevos másteres oficiales como el Máster en Big Data y Computación en la Nube y el Máster en Inteligencia Artificial, los cuales refuerzan la formación en áreas estratégicas con gran proyección profesional.

El presente trabajo se desarrolla en el contexto del actual Grado en Ingeniería Informática. Por ello, se describe la organización del plan de estudios, los tipos de asignaturas que lo componen, así como el funcionamiento del sistema de matrícula y su relación con el problema planteado. Toda la información relativa a este grado se encuentra disponible en la página web oficial de la Escuela

2.2. Plan de Estudios del Grado en Ingeniería Informática

Dentro del sitio web oficial de la Escuela Superior de Ingeniería Informática de Albacete, la información relevante sobre el Grado en Ingeniería Informática se encuentra principalmente en los apartados dedicados al plan de estudios y los horarios. El primero ofrece un desglose detallado de los módulos que componen el grado, así como los créditos ECTS asignados a cada uno de ellos.

El grado se estructura en los siguientes módulos:

- **Módulo I – Formación Básica (60 ECTS):** Compuesto por 10 asignaturas obligatorias de 6 ECTS, la mayoría impartidas durante el primer y segundo curso.
- **Módulo II – Común a la Rama de Informática (96 ECTS):** Incluye asignaturas obligatorias para todos los estudiantes, repartidas entre el segundo y tercer curso.
- **Módulo III – Tecnologías Específicas (48 ECTS):** Se ofrecen cuatro intensificaciones, de las cuales el estudiante debe elegir una.
- **Módulo IV – Trabajo Fin de Grado (12 ECTS):** Corresponde al proyecto final obligatorio del grado.
- **Módulo V – Optativas (24 ECTS cursados de una oferta de 72 ECTS):** Permite al estudiante personalizar su itinerario académico según sus intereses.

Los módulos I y II son especialmente relevantes para este trabajo, ya que agrupan asignaturas con múltiples grupos de teoría y prácticas, lo que permite aplicar técnicas de optimización en la asignación de matrículas. Las asignaturas con grupo único, en cambio, no ofrecen posibilidades de optimización en este sentido, aunque sí afectan al conjunto de restricciones del problema.

A continuación, se detallan las asignaturas correspondientes a los dos primeros módulos, los cuales agrupan el conjunto de materias obligatorias impartidas durante los primeros cursos del grado:

Asignaturas del Módulo I – Formación Básica (60 ECTS):

- | | |
|--------------------------------------|------------|
| ■ Álgebra y Matemática Discreta | (1º curso) |
| ■ Cálculo y Métodos Numéricos | (1º curso) |
| ■ Estadística | (2º curso) |
| ■ Estructura de Computadores | (1º curso) |
| ■ Fundamentos de Gestión Empresarial | (1º curso) |
| ■ Fundamentos de Programación I | (1º curso) |
| ■ Fundamentos de Programación II | (1º curso) |

- Fundamentos Físicos de la Informática (1º curso)
- Lógica (2º curso)
- Tecnología de Computadores (1º curso)

Asignaturas del Módulo II – Común a la Rama de Informática (96 ECTS):

- Arquitectura de Computadores (3º curso)
- Aspectos Profesionales de la Informática (3º curso)
- Bases de Datos (2º curso)
- Estructura de Datos (2º curso)
- Ingeniería del Software I (2º curso)
- Ingeniería del Software II (3º curso)
- Interacción Persona-Ordenador I (3º curso)
- Metodología de la Programación (2º curso)
- Organización de Computadores (2º curso)
- Programación Concurrente y Tiempo Real (2º curso)
- Redes de Computadores I (1º curso)
- Redes de Computadores II (2º curso)
- Sistemas de Información (1º curso)
- Sistemas Distribuidos (3º curso)
- Sistemas Inteligentes (3º curso)
- Sistemas Operativos I (2º curso)

Estos dos módulos cubren todas las asignaturas del primer y segundo curso, así como parte del tercer curso, incluyendo el 1º cuatrimestre entero y la asignatura del 2º cuatrimestre “Aspectos Profesionales de la Informática”. Esta organización modular permite una estructuración clara de los itinerarios y facilita el análisis automatizado de los posibles solapamientos y distribuciones de alumnos que se aborda en este trabajo.

Por otro lado, la sección de horarios disponible en la web oficial de la Escuela [UCLM, 2024b]

proporciona información esencial para el desarrollo de este trabajo. En dicha sección, los horarios están organizados por cuatrimestres y muestran las franjas horarias asignadas a cada asignatura y grupo, lo que resulta crucial para la detección de solapamientos entre clases, un aspecto que se abordará en secciones posteriores.

Antes de profundizar en el funcionamiento del sistema de matrícula y su relevancia para este estudio, es importante comprender la estructura de grupos y horarios de los cursos implicados. En los cursos de primero, segundo y en el primer cuatrimestre de tercero, se ofrece una modalidad adicional, el grupo bilingüe. Sin embargo, este grupo y sus matrículas

no se han tenido en cuenta para este trabajo, ya que al existir únicamente un grupo bilingüe por asignatura, no permite ningún tipo de optimización ni elección alternativa.

Para el primer curso, existen tres grupos de teoría disponibles: grupo A, B y C, que en la documentación interna y datos de matrícula también se identifican como grupo 10, 11 y 12, respectivamente. Cada uno de estos grupos incluye, a su vez, dos subgrupos de prácticas diferenciados para cada asignatura.

En segundo curso, hay dos grupos disponibles: grupo A y B (grupos 10 y 11), también con dos subgrupos de prácticas por grupo de teoría. Esta misma estructura se mantiene para el primer cuatrimestre del tercer curso, en el que también se dispone de dos grupos (A y B, equivalentes a 10 y 11) para las asignaturas impartidas en ese periodo. Además, la asignatura “Aspectos Profesionales de la Informática”, también se ofrece en dos grupos.

Para todas estas asignaturas con varios grupos, se mantiene un patrón común en su distribución horaria, cada grupo de teoría imparte dos sesiones semanales de teoría y una sesión de prácticas. Por otro lado, el resto de asignaturas correspondientes a cursos más avanzados (segundo cuatrimestre de tercero y cuarto curso) se imparten en grupo único, tanto para la parte teórica como para las prácticas, por lo que no ofrecen posibilidad de optimización en términos de asignación de estudiantes.

El sistema de matriculación de la universidad asigna a cada estudiante una fecha y hora específicas para realizar su inscripción a través de la plataforma de Secretaría Virtual [UCLM, 2024c]. Durante este proceso, el alumnado selecciona las asignaturas en las que desea matricularse. En el caso de asignaturas con varios grupos, es necesario elegir el grupo de teoría deseado, la asignación a subgrupos de prácticas se realiza posteriormente mediante la página correspondiente de cada asignatura en el campus virtual [UCLM, 2024a].

Cuando un estudiante se matricula únicamente en asignaturas de un mismo curso, la selección de grupos no suele presentar dificultades. Sin embargo, cuando se combinan asignaturas de diferentes cursos en el mismo cuatrimestre, la complejidad del proceso aumenta notablemente. Un diseño inadecuado de la matrícula puede provocar conflictos de horarios que impidan la asistencia completa a todas las clases.

Por ejemplo, un estudiante que desee cursar simultáneamente “Fundamentos de Programación I” (primer curso) y “Organización de Computadores” (segundo curso) deberá tener en cuenta tanto los grupos de teoría como los subgrupos de prácticas para evitar solapamientos. En este caso concreto, seleccionar el mismo grupo (por ejemplo, grupo A) para ambas asignaturas podría resultar inviable si se desea asegurar la asistencia regular a todas las sesiones.

1º GRUPO A (10) - 1 Semestre 2023-24 Primer Cuatrimestre					
	Lunes	Martes	Miércoles	Jueves	Viernes
8:15 9:45	Fundamentos de Gestión Empresarial AULA 1.11	Fundamentos de Programación I AULA 1.11	Fundamentos Físicos de la Informática AULA 1.11	Tecnología de Computadores AULA 1.11	Cálculo y Métodos Numéricos AULA 1.11
9:45 11:15	Cálculo y Métodos Numéricos AULA 1.11	Fundamentos de Gestión Empresarial AULA 1.11	Fundamentos de Programación I AULA 1.11	Fundamentos Físicos de la Informática AULA 1.11	Tecnología de Computadores AULA 1.11
11:35 13:05	Fundamentos Físicos de la Informática Lab. Física 2	Tecnología de Computadores Elect.Digit.1	Cálculo y Métodos Numéricos SOFTW. 4	Fundamentos de Gestión Empresarial Aula Multiusos	Fundamentos de Programación I SOFTW. 7
16:00 17:30	Fundamentos Físicos de la Informática Lab. Física 2	Tecnología de Computadores Elect.Digit.1	Cálculo y Métodos Numéricos SOFTW. 4	Fundamentos de Gestión Empresarial Aula Multiusos	Fundamentos de Programación I SOFTW. 7

Figura 2.1: Horario 1º cuatrimestre Grupo 1ºA

2º GRUPO A (10) - 1 Semestre 2023-24 Primer Cuatrimestre					
	Lunes	Martes	Miércoles	Jueves	Viernes
9:45 11:15	Estructura de Datos SOFTW. 4	Lógica Aula Multiusos	Organización de Computadores SOFTW. 7	Ingeniería del Software I SOFTW. 7	Sistemas Operativos I SOFTW. 4
11:35 13:05	Ingeniería del Software I AULA 1.11	Sistemas Operativos I AULA 1.11	Estructura de Datos AULA 1.11	Lógica AULA 1.11	Organización de Computadores AULA 1.11
13:05 14:35	Organización de Computadores AULA 1.11	Ingeniería del Software I AULA 1.11	Sistemas Operativos I AULA 1.11	Estructura de Datos AULA 1.11	Lógica AULA 1.11
17:30 19:00	Estructura de Datos SOFTW. 4	Lógica Aula Multiusos	Organización de Computadores AULA 1.11	Ingeniería del Software I SOFTW. 7	Sistemas Operativos I SOFTW. 4

Figura 2.2: Horario 1º cuatrimestre Grupo 2ºA

Una vez completado el proceso de matrícula por parte de todos los estudiantes, y considerando tanto sus preferencias como las plazas disponibles en cada grupo, se obtiene una configuración inicial. Esta configuración servirá como punto de partida para los algoritmos evolutivos desarrollados en este trabajo, cuya finalidad es optimizar dicha asignación inicial.

A partir de la información obtenida de la web oficial de la escuela, se han generado tres

ficheros clave que servirán como base para el desarrollo del algoritmo genético. El primero es *asignaturas*, que tienen un listado completo de asignaturas del grado y su clasificación, incluyendo:

- **COD.ASIG:** Identificador único de la asignatura.
- **NOMBRE ASIGNATURA:** Nombre completo de la asignatura.
- **TECNOLOGÍA:** Indica que tipo de asignatura es, obligatoria, optativa o si pertenece a alguna intensificación.
- **CUATRIMESTRE:** Cuatrimestre en el que se imparte la asignatura.

El segundo fichero, *horarios*, tiene todos los horarios de los grupos de teoría y práctica de las asignaturas, extraídos de la sección de horarios. Este fichero incluye la siguiente información sobre los horarios de cada asignatura:

- **CUATRIMESTRE:** Cuatrimestre en el que se imparte la asignatura.
- **CURSO:** Curso en el que se imparte la asignatura.
- **GRUPO:** Grupo al que pertenece este horario.
- **ID GRUPO:** Identificador numérico del curso.
- **DÍA:** Día en el que se imparte la asignatura.
- **HORARIO:** Horario en el que se imparte la asignatura.
- **BLOQUE:** Indica al bloque al que pertenece la asignatura.
- **ASIGNATURA:** Nombre completo de la asignatura.
- **CODIGO:** Identificador único de la asignatura.
- **TEORÍA/PRÁCTICA:** Indica si el horario es de un grupo de teoría o prácticas.

Por último está el fichero *matrículas*, que tiene los registros anonimizados de las matrículas de los estudiantes del curso 2023-2024. La información que hay sobre cada matrícula es:

- **AÑO:** Año lectivo de la matrícula.
- **PLAN:** Plan docente al que pertenece la matrícula.
- **CODIGO:** Identificador único de la asignatura.
- **ASIGNATURA:** Nombre completo de la asignatura.
- **DNI:** Dni anonimizado del estudiante al que pertenece la matrícula.
- **ALUMNO:** Nombre anonimizado del estudiante al que pertenece la matrícula.
- **GRUPO:** Grupo de teoría al que pertenece la matrícula.
- **GP:** Subgrupo de prácticas al que pertenece la matrícula.

2.3. Metaheurísticas evolutivas

En el ámbito de la optimización, las metaheurísticas son consideradas herramientas fundamentales para resolver problemas complejos, especialmente aquellos que no pueden abordarse mediante métodos exactos debido a su alta dimensionalidad, no linealidad o la falta de modelos matemáticos explícitos. Una metaheurística es un enfoque general de optimización que se basa en heurísticas flexibles y adaptativas, diseñadas para explorar eficientemente grandes espacios de búsqueda en busca de soluciones cercanas al óptimo global.

El término "metaheurística" proviene del griego "meta", que significa "más allá", y "heurística", que se refiere a un método de descubrimiento o aprendizaje. Esto destaca su objetivo de ir más allá de las técnicas heurísticas convencionales para encontrar soluciones de calidad en un marco más amplio y general.

Las principales características que poseen las metaheurísticas son:

- **Flexibilidad:** Son aplicables a una amplia gama de problemas, desde la planificación logística hasta la inteligencia artificial.
- **Efectividad:** Debe ser capaz de encontrar soluciones de alta calidad, cercanas al óptimo, incluso en problemas complejos y de gran escala.
- **Eficiencia:** Es crucial que haga un uso equilibrado de los recursos disponibles, minimizando el tiempo de ejecución y el consumo de memoria.
- **Generalidad:** La metaheurística debe ser versátil y funcionar bien en distintos tipos de problemas, manteniendo un rendimiento consistente.
- **Robustez:** Su desempeño debe mantenerse estable incluso cuando el problema o el entorno presentan pequeñas variaciones.
- **Autonomía:** La metaheurística debe requerir la menor parametrización posible, o ser capaz de ajustar sus parámetros automáticamente para facilitar su uso.

Las metaheurísticas evolutivas son una categoría dentro de las metaheurísticas, se caracterizan por su inspiración en procesos biológicos y evolutivos. Estas técnicas se fundamentan en principios de selección natural, reproducción, mutación y supervivencia del más apto, adaptando estos conceptos para tratar problemas más complejos de optimización. Imitan la evolución natural como un modelo de búsqueda y optimización. En lugar de centrarse en un único candidato a solución, trabajan con una población de posibles soluciones, que evoluciona a través de iteraciones o generaciones. Cada generación se construye aplicando operadores inspirados en la biología, como la selección, el cruce y la mutación. Estos mecanismos permiten la exploración de nuevos puntos en el espacio de búsqueda y el refinamiento progresivo de las soluciones.

Las metaheurísticas evolutivas poseen todas las propiedades generales de las Metaheurísticas, pero también presentan características propias que aumentan su aplicabilidad en problemas específicos:

- **Exploración y explotación equilibradas:** Gracias al uso de operadores de cruce y mutación, las metaheurísticas evolutivas logran un balance entre explorar nuevas áreas

del espacio de búsqueda (exploración) y refinar las soluciones existentes (explotación). Esto ayuda a evitar óptimos locales y mejorar la calidad de las soluciones.

- **Adaptabilidad:** Se pueden adaptar a múltiples dominios, ya que los operadores genéticos pueden ser modificados o personalizados según las características del problema a resolver.
- **Soluciones poblacionales:** El uso de una población de soluciones permite trabajar en paralelo con múltiples candidatos, esto incrementa la diversidad y reduce una convergencia prematura en óptimos locales.
- **Robustez frente a incertidumbre:** Las metaheurísticas evolutivas son menos susceptibles a variaciones en las condiciones del problema debido a su enfoque probabilístico, lo que las hace robustas en escenarios inciertos o dinámicos.

2.4. Algoritmos Genéticos

Los algoritmos genéticos (AG) son una de las implementaciones más consolidadas y estudiadas dentro del conjunto de metaheurísticas evolutivas. Estas técnicas se inspiran en los principios de la evolución natural formulados por Charles Darwin [Darwin, 1859], aplicando mecanismos biológicos como la selección, la reproducción y la mutación para abordar problemas complejos de optimización. Introducidos por John Holland en la década de 1970 [Holland, 1975], los AG han demostrado ser herramientas eficaces en campos diversos, desde la ingeniería hasta la biología computacional.

El funcionamiento de un algoritmo genético se basa en una población de soluciones potenciales, conocidas como individuos. A lo largo de múltiples iteraciones o generaciones, estos individuos evolucionan mediante operadores genéticos como el cruce y la mutación. Cada individuo es evaluado a través de una función fitness, que determina su adecuación con respecto al problema planteado.

Uno de los retos característicos en el diseño de los algoritmos genéticos es evitar la llamada convergencia prematura. Este fenómeno ocurre cuando la población se vuelve homogénea demasiado pronto, lo que implica que las soluciones tienden rápidamente a una misma zona del espacio de búsqueda, reduciendo drásticamente la diversidad genética. Como consecuencia, el algoritmo puede estancarse en óptimos locales y dejar de explorar otras regiones potencialmente mejores del espacio de soluciones. La convergencia prematura suele estar asociada a una presión selectiva excesiva, un tamaño de población reducido o una mala configuración de los operadores genéticos.

La representación de los individuos es un aspecto clave en el comportamiento del algoritmo, ya que influye directamente en la eficiencia de los operadores de búsqueda. Las representaciones más utilizadas son:

- **Codificación binaria:** Utiliza cadenas de bits (0 y 1) para representar las soluciones. Es especialmente adecuada para problemas combinatorios [Goldberg, 1989].
- **Codificación real:** Utiliza valores numéricos reales, lo que resulta más eficiente en problemas de optimización continua [Deb, 2001].

- **Codificación por permutación:** Diseñada para problemas en los que el orden de los elementos es importante, cómo el problema del viajante de comercio (*TSP*) [Deb, 2001].

El flujo básico de un algoritmo genético se estructura en los siguientes pasos:
[Goldberg, 1989]

1. **Inicialización:** Se genera una población inicial de individuos de forma aleatoria o utilizando heurísticas específicas. Esta población representa un conjunto diverso de posibles soluciones al problema.
2. **Evaluación:** Cada individuo es evaluado mediante la función fitness, que proporciona una medida de su calidad. Esta función es clave para guiar la evolución hacia mejores soluciones.
3. **Selección:** Los individuos más aptos tienen una mayor probabilidad de ser seleccionados para el cruce. Existen distintos métodos.
4. **Cruce:** Se combinan pares de individuos seleccionados para producir descendencia. Este operador permite explorar nuevas regiones del espacio de búsqueda al intercambiar información genética entre los individuos.
5. **Mutación:** Se introducen pequeñas modificaciones en los individuos para mantener la diversidad genética y evitar la convergencia prematura.
6. **Reemplazo:** La nueva población resultante reemplaza total o parcialmente a la población anterior, dependiendo de la estrategia de reemplazo utilizada.
7. **Condición de parada:** El algoritmo genético continúa hasta que se cumple una condición, cómo alcanzar un número máximo de generaciones, encontrar una solución con un valor de fitness suficiente o ver que la población no ha cambiado en varias generaciones (convergencia).

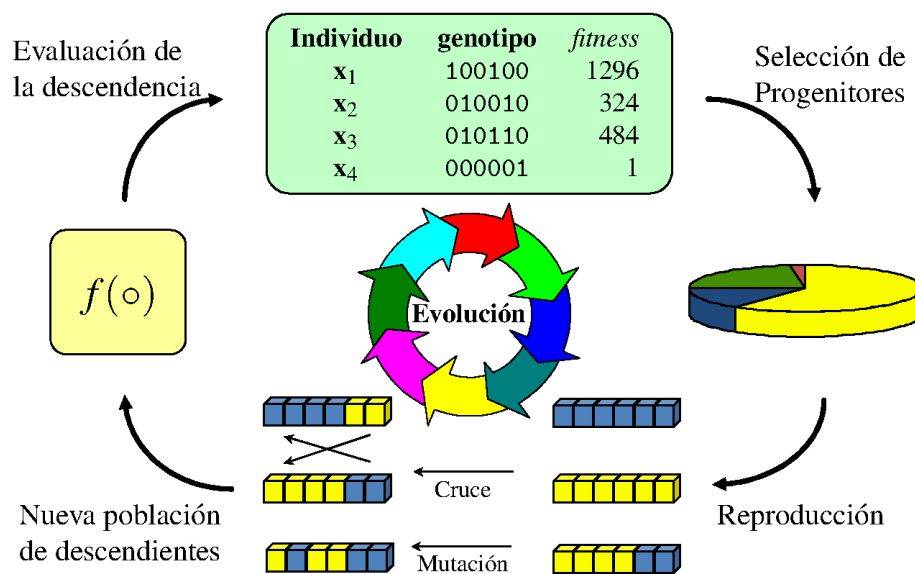


Figura 2.3: Flujo básico de un algoritmo genético. Fuente: https://www.cs.us.es/~fsancho/Blog/posts/Algoritmos_Geneticos.md.html.

Aunque los algoritmos genéticos son herramientas poderosas y muy utilizadas, tienen algunas desventajas a considerar. La primera es que estos algoritmos pueden requerir un número elevado de generaciones para converger a soluciones óptimas, sobre todo en problemas de alta dimensionalidad o cuando el espacio de búsqueda contiene numerosos óptimos locales. Este comportamiento puede traducirse en un tiempo de ejecución considerablemente alto en escenarios complejos.

Los algoritmos genéticos son altamente sensibles a la configuración de sus parámetros. La correcta elección de estos hiperparámetros es esencial para garantizar un buen equilibrio entre exploración del espacio de búsqueda y explotación de las soluciones prometedoras. Una parametrización inadecuada puede provocar una exploración insuficiente o, por el contrario, un consumo excesivo de recursos computacionales sin mejoras significativas en la calidad de las soluciones.

A continuación, se describen los principales hiperparámetros que deben ser definidos al implementar un algoritmo genético:

- **Tamaño de la población:** Determina cuántas soluciones se mantienen en cada generación.
- **Número de generaciones:** Indica cuántas iteraciones del ciclo evolutivo se llevarán a cabo.
- **Método de selección:** Define la estrategia mediante la cuál se eligen los individuos que participarán en la reproducción.
- **Método de cruce:** Establece cómo se combinan los genes de los progenitores para generar descendencia.

- **Método de sustitución:** Determina cómo se forma la nueva población a partir de la actual y de los descendientes generados.
- **Probabilidad de cruce:** Representa la probabilidad de que dos individuos seleccionados se crucen para generar descendencia.
- **Probabilidad de mutación:** Controla la probabilidad de alterar aleatoriamente uno o más genes de un individuo.
- **Condición de parada:** Define cuándo debe finalizar la ejecución del algoritmo.

Otra limitación importante recae en el coste computacional asociado a la evaluación de la función objetivo. En cada generación la calidad de los individuos debe ser calculada, lo que puede ser muy costoso en problemas donde la evaluación conlleva cálculos complejos.

Por último, los algoritmos genéticos presentan desafíos específicos en el contexto de problemas multiobjetivo. En estos casos, se busca optimizar simultáneamente múltiples objetivos conflictivos, lo que introduce una mayor complejidad en la búsqueda de soluciones. Sin un diseño adecuado, puede llevar a soluciones que favorecen únicamente algunos objetivos, perdiendo la diversidad. Estos desafíos hacen necesario el desarrollo de estrategias específicas para abordar eficazmente problemas multiobjetivo.

2.5. Algoritmos Genéticos Multiobjetivo

En la mayoría de problemas de optimización del mundo real, es común enfrentarse a múltiples objetivos conflictivos que tienen que ser optimizados simultáneamente. Los algoritmos genéticos multiobjetivo (MOGA, por sus siglas en inglés) se han desarrollado como una extensión de los algoritmos genéticos convencionales para tratar este tipo de problemas. En los algoritmos genéticos clásicos se tiende a combinar los objetivos en una única función escalar, en cambio, en los MOGA se busca un conjunto de soluciones conocidas como frente de Pareto, que representa el equilibrio entre los distintos objetivos, mostrando soluciones en las que mejorar un objetivo implica empeorar otro, esto proporciona una variedad de opciones igual de buenas para elegir según las necesidades del problema.

La optimización multiobjetivo se basa en el concepto de Pareto-optimalidad, introducido por el economista y sociólogo Vilfredo Pareto en el contexto de la economía [Pareto, 1971]. Esto parte de la idea de que, en muchos problemas reales, no existe una única solución que optimice simultáneamente todos los objetivos. En su lugar, se busca un conjunto de soluciones eficientes que representen distintos compromisos entre dichos objetivos.

Desde el punto de vista de la teoría de la optimización, una solución es considerada óptima en el sentido de Pareto si no puede mejorarse un objetivo sin empeorar al menos uno de los otros. Este tipo de soluciones conforman lo que se conoce como el frente de Pareto, que constituye el conjunto de todas las soluciones no dominadas. Estas soluciones no son mejores ni peores entre sí, sino que ofrecen distintas ventajas relativas, proporcionando al tomador de decisiones alternativas entre las que elegir según sus prioridades específicas.

Formalmente, dado un problema en el que se deben minimizar m funciones objetivo, se dice que una solución $x_1 \in S$ domina a otra solución $x_2 \in S$ si se cumplen las siguientes

condiciones:

- $f_i(x_1) \leq f_i(x_2) \quad \forall i \in 1, \dots, m$
- Existe al menos un $j \in 1, \dots, m$ tal que $f_j(x_1) < f_j(x_2)$

El conjunto de todas las soluciones que no son dominadas por ninguna otra conforman el llamado frente de Pareto. El objetivo de los algoritmos de optimización multiobjetivo es aproximar este frente de forma eficiente, generando soluciones que presenten:

- **Alta convergencia:** es decir, cercanía al frente de Pareto verdadero, lo que indica soluciones de alta calidad.
- **Buena diversidad:** variedad de soluciones en distintas zonas del frente, permitiendo un abanico amplio de opciones para la toma de decisiones.

Este enfoque tiene una gran ventaja, separa el proceso de búsqueda de soluciones del proceso de toma de decisiones, ya que permite presentar al usuario final un conjunto de soluciones eficientes sin imponer un criterio único de selección.

Para ilustrar gráficamente este concepto, se muestra en la Figura 2.4 un ejemplo en el que se minimizan dos objetivos. Los puntos marcados en rojo representan el frente de Pareto, mientras que los puntos en negro corresponden a soluciones dominadas.

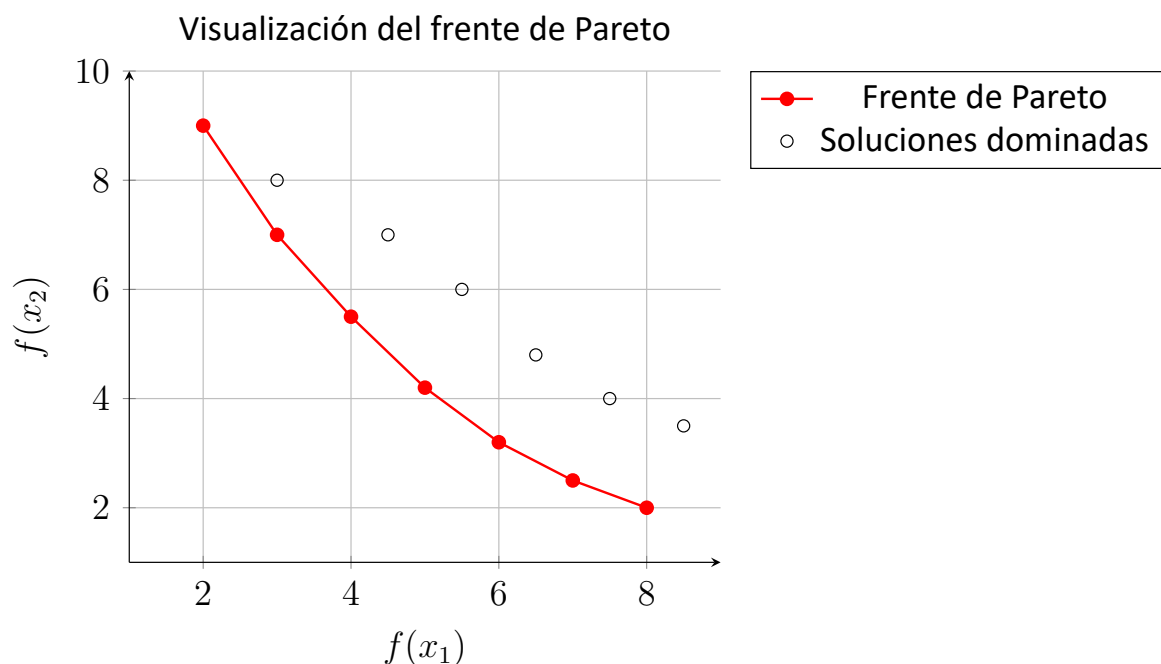


Figura 2.4: Visualización del frente de Pareto en el espacio objetivo.

Para cumplir estos objetivos, los MOGA usan técnicas que extienden las ideas clásicas de los algoritmos genéticos. En la evaluación una solución no se basa en un fitness calculado

mediante una función escalar, sino que se tiene en cuenta el nivel de no-dominancia de cada solución respecto a las demás. Esto se realiza clasificando las soluciones en frentes jerárquicos, donde las soluciones no dominadas forman el primer nivel, el siguiente frente lo formarán las soluciones dominadas sólo por el frente anterior y así sucesivamente. Con esto se asegura que las mejores soluciones tengan una mayor prioridad.

Para mantener la diversidad entre las soluciones, se emplean estrategias como el cálculo de distancias entre las soluciones del frente. Este tipo de estrategias se usan para evitar que las soluciones converjan hacia una única región en el frente de Pareto. Los MOGA son especialmente relevantes en aplicaciones prácticas gracias a su capacidad para manejar objetivos conflictivos de manera simultánea y eficiente. Al contrario que los algoritmos genéticos clásicos que ofrecen una única solución óptima, los MOGA proporcionan un conjunto diverso de soluciones que pueden adaptarse a distintas necesidades y restricciones.

2.5.1. NSGA-III

Existen diversos algoritmos evolutivos para optimización multiobjetivo, entre los que destacan NSGA-II [Deb et al., 2002], SPEA2 [Zitzler et al., 2001] o MOEA/D [Zhang and Li, 2007], los cuales han demostrado un buen rendimiento en problemas con dos o tres objetivos. Sin embargo, cuando el número de objetivos crece como ocurre en este trabajo, que plantea un problema con cinco objetivos a optimizar, muchos de estos algoritmos pierden eficacia debido a la degradación de los mecanismos de selección basados en dominancia y a la dificultad para mantener diversidad en el conjunto de soluciones.

Por esta razón, se ha optado por utilizar el algoritmo NSGA-III (*Non-dominated Sorting Genetic Algorithm III*), una evolución del conocido NSGA-II, específicamente diseñado para afrontar problemas de optimización con un alto número de objetivos. NSGA-III fue propuesto por Deb y Jain en 2014 [Deb and Jain, 2014] como una mejora sustancial sobre su predecesor, NSGA-II [Deb et al., 2002], que también fue desarrollado por Deb junto a Pratap, Agarwal y Meyarivan.

La principal aportación de NSGA-III está en la incorporación de un conjunto predefinido de puntos de referencia en el espacio de objetivos, los cuales guían el proceso evolutivo y permiten lograr una distribución uniforme de las soluciones a lo largo del frente de Pareto. Este mecanismo resulta esencial en problemas con muchos objetivos, donde los enfoques clásicos tienden a generar soluciones muy concentradas o incluso degeneradas. Gracias a este sistema, NSGA-III mantiene una buena diversidad en el frente y favorece la exploración de regiones representativas del espacio objetivo.

El flujo básico del algoritmo NSGA-III es el siguiente [Deb and Jain, 2014]:

1. **Inicialización:** Se genera una población inicial aleatoria de soluciones, a partir de esa población usando cruce y mutación se genera otra población, la unión de esas dos servirá como población inicial. Se define un conjunto de puntos de referencia equidistantes en el espacio objetivo. Estos puntos sirven como guía para asegurar una distribución uniforme de las soluciones en el frente de Pareto.
2. **Evaluación:** Cada individuo es evaluado en todos los objetivos del problema. No existe una función fitness para valorar la solución.

-
3. **Clasificación no dominada:** Se clasifica la población en frentes según el principio de Pareto. El primer frente contiene las soluciones no dominadas, el segundo incluye aquellas dominadas sólo por las del primer frente, y así sucesivamente.
 4. **Asignación a puntos de referencia:** Las soluciones de los frentes se asignan a uno de los puntos de referencia. Esta asignación permite medir qué tan representativa es cada solución respecto a la diversidad deseada en el frente.
 5. **Selección:** Cuando el último frente no cabe completamente en la nueva población, se seleccionan las soluciones más cercanas (o menos representadas) en relación con los puntos de referencia. Este proceso asegura un equilibrio entre convergencia y diversidad.
 6. **Cruce y mutación:** Se generan nuevos individuos mediante los operadores genéticos de cruce y mutación.
 7. **Reemplazo:** Se crea una nueva población combinando padres e hijos. A continuación, se repite la clasificación no dominada y la selección basada en los puntos de referencia.
 8. **Condición de parada:** El algoritmo continúa iterando hasta alcanzar una condición de parada, como un número máximo de generaciones o una estabilidad en el frente de soluciones obtenidas.

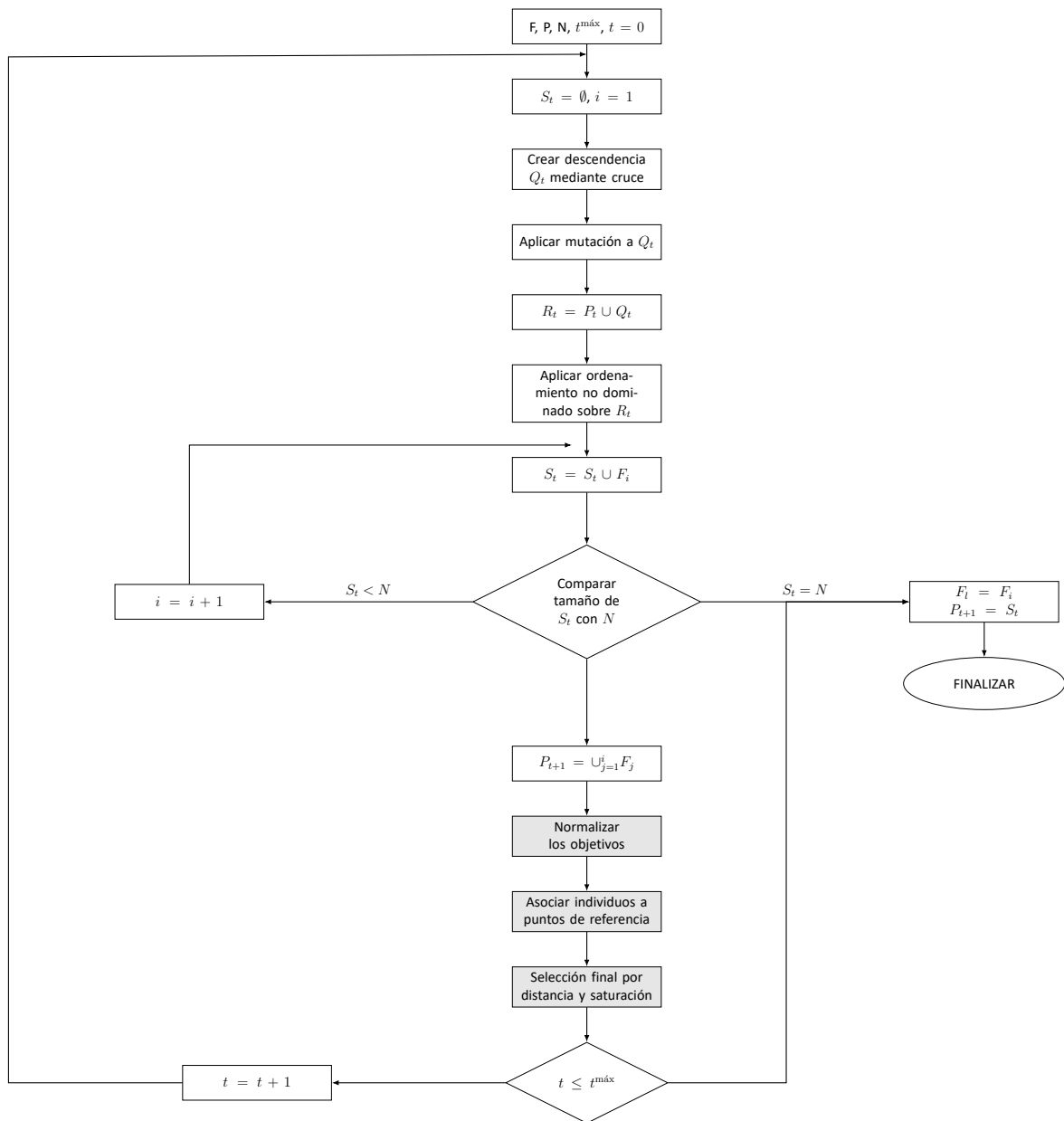


Figura 2.5: Flujo básico de NSGA-III. Fuente: https://www.researchgate.net/figure/Flowchart-of-NSGA-III_fig3_329986923

3. Desarrollo

En este capítulo se explica el proceso de desarrollo del algoritmo genético, la modelización matemática del problema, los ficheros de salida y las distintas implementaciones que se han realizado.

Actualmente, existen diversas librerías especializadas que permiten ejecutar algoritmos genéticos clásicos y multiobjetivo con relativa facilidad, cómo es el caso de *DEAP* para algoritmos evolutivos clásicos [Fortin et al., 2012] o *pymoo* para algoritmos cómo NSGA-II y NSGA-III [Blank and Deb, 2020]. Estas herramientas ofrecen implementaciones optimizadas y flexibles de múltiples algoritmos, junto con funcionalidades de visualización y evaluación.

Sin embargo, en este trabajo se ha optado por desarrollar desde cero toda la lógica del algoritmo utilizando exclusivamente el lenguaje de programación Python. Esta decisión ha permitido un mayor control sobre el funcionamiento interno del algoritmo, facilitando su adaptación a las particularidades del problema de asignación de matrículas planteado, así cómo una mejor comprensión de su estructura y comportamiento.

3.1. Modelización del problema

Una vez planteado el problema, es necesario establecer una modelización adecuada que permita su resolución mediante un algoritmo genético. Para ello, deben definirse de forma precisa los componentes clave del modelo evolutivo cómo la representación de los individuos que conforman cada población, los objetivos que se desean optimizar, también los métodos para su cálculo e integración dentro de una función fitness.

Además, es imprescindible especificar el funcionamiento de los principales operadores genéticos, cómo la selección, el cruce y la mutación, junto con la condición de parada que determinará el fin del proceso evolutivo.

3.1.1. Representación de los individuos

En este problema, cada individuo representa una solución completa para la asignación de matrículas de todos los alumnos. Para modelar esta representación se utiliza una codificación real en la que cada matrícula está descrita por su grupo de teoría y subgrupo de

prácticas. Esta codificación es la que se ve afectada directamente por los operadores genéticos de cruce y mutación y constituye lo que se conoce como el genotipo del individuo.

El genotipo se estructura como una cadena lineal en la que se concatenan, para cada alumno, los valores correspondientes a las asignaturas matriculadas con múltiples grupos disponibles. Cada par de valores representa el grupo de teoría y el subgrupo de prácticas asignado a una de esas asignaturas. En la Figura 3.1, por ejemplo, se muestra el genotipo de una solución que contempla dos estudiantes: el primero con tres asignaturas variables (seis valores) y el segundo con otras tres (seis valores).

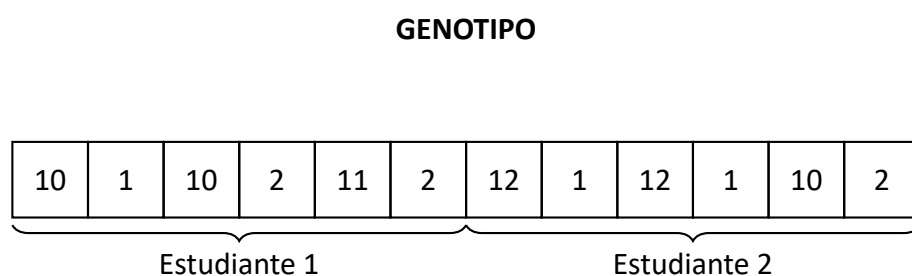


Figura 3.1: Representación del genotipo de dos estudiantes. Cada pareja de valores representa el grupo de teoría y subgrupo de prácticas de una matrícula variable.

Tal y como se ha explicado anteriormente, los grupos de teoría podrán tomar distintos valores según el curso al que pertenezca la asignatura:

- **10, 11 y 12** para asignaturas de primer curso.
- **10 y 11** para asignaturas de segundo curso y del primer cuatrimestre del tercer curso, incluyendo la asignatura “Aspectos Profesionales de la Informática”.

En todos estos casos, cada grupo de teoría cuenta con dos subgrupos de prácticas, numerados como 1 y 2. Durante la ejecución del algoritmo, estos valores podrán ser modificados mediante los operadores genéticos con el objetivo de encontrar una configuración óptima.

En cambio, las asignaturas que pertenecen a otros cursos o módulos tienen un único grupo de teoría y un único grupo de prácticas, por lo que no forman parte del genotipo y su asignación se mantiene fija.

La interpretación completa de la solución, conocida como fenotipo, se obtiene al complementar el genotipo con las asignaturas de grupo único para cada estudiante. Esta representación es la que se emplea para la evaluación de la solución durante la ejecución del algoritmo, permitiendo calcular los objetivos definidos a partir de la configuración real que cada individuo representa.

3.1.2. Objetivos

Una vez definida la representación de los individuos, es necesario establecer los objetivos que guiarán el proceso de optimización. Para este problema se consideran cinco objetivos, seleccionados por su relevancia. A continuación, se describe cada uno de ellos junto con el criterio utilizado para su evaluación.

1. Número de solapes (minimizar):

Este objetivo cuantifica la suma total de conflictos de horarios entre asignaturas para cada estudiante, siendo un objetivo a minimizar. Se define un *solape* como la coincidencia del horario entre dos asignaturas en las que un mismo alumno está matriculado, lo que le imposibilita asistir simultáneamente a ambas. Para el cálculo de este valor, se analiza individualmente a cada alumno, identificando los solapes exclusivamente entre asignaturas del mismo cuatrimestre. Esta distinción es relevante, ya que los solapes entre asignaturas de distintos cuatrimestres no afectan a la asistencia efectiva y, por tanto, no se consideran en la evaluación de este objetivo.

$$\text{Solapes}_{\text{cuatrimestre}} = \sum_{\substack{i=1 \\ c_i \in Q}}^{n-1} \sum_{\substack{j=i+1 \\ c_j \in Q}}^n \text{intersección}(C_i, C_j)$$

Donde:

- C_i y C_j son asignaturas específicas en las que está matriculado el estudiante.
- Q representa el conjunto de asignaturas de ese cuatrimestre en las que está matriculado el estudiante.
- n es el número total de asignaturas en las que está matriculado el estudiante en ese cuatrimestre.
- La función $\text{intersección}(C_i, C_j)$ devuelve 1 si las asignaturas C_i y C_j se solapan en el tiempo y 0 en caso contrario.

Una vez que se tiene la definición para calcular los solapes totales por cuatrimestre de cada estudiante, se realiza la suma de ambos para obtener el total de solapes de un estudiante:

$$\text{Solapes}_{\text{estudiante}} = \text{Solapes}_{\text{cuatrimestre}_1} + \text{Solapes}_{\text{cuatrimestre}_2}$$

El número total de solapes en un individuo se obtiene sumando los solapes de cada uno de los estudiantes:

$$\text{Solapes Totales} = \sum_{i=1}^n \text{Solapes}_{\text{estudiante}_i}$$

Donde:

- n es el número de estudiantes presentes en la solución representada por el individuo.

2. Equilibrio de alumnos por clase (maximizar):

Este objetivo refleja el grado de equidad en la distribución de estudiantes entre los diferentes grupos de teoría y subgrupos de prácticas de las asignaturas. Se considera un criterio especialmente relevante, ya que una mala distribución puede afectar directamente a la viabilidad de la docencia debido a la capacidad limitada de las aulas. Además, una distribución desequilibrada puede dificultar la atención personalizada al alumnado y generar problemas logísticos.

Mientras que otros objetivos están orientados a mejorar el horario del estudiante, cómo evitar solapamientos o facilitar horarios cómodos, este objetivo se vincula con la factibilidad real de implementar la solución. Por tanto, su cumplimiento se considera prioritario.

Para calcular esta métrica, se evaluará el equilibrio alcanzado en cada asignatura obligatoria. A continuación, se describen los pasos seguidos para su cálculo.

2.1 Número ideal de estudiantes por grupo de teoría:

Para la asignatura C_i con N_i estudiantes matriculados y G_i grupos, el número ideal de estudiantes por grupo es $N_{ideal_i} = N_i/G_i$

2.2 Proporción de equilibrio por grupo principal:

Para cada grupo principal G_i de la asignatura C_i , el equilibrio se calcula restando la cantidad ideal de estudiantes N_{ideal_i} al número real de estudiantes n_{G_i} en ese grupo. Esta diferencia se normaliza dividiéndola por N_{ideal_i} y el resultado se toma en valor absoluto para evitar resultados negativos. Finalmente, se resta esta proporción de 1 para obtener una medida del equilibrio, donde valores cercanos a 1 significa que el número de estudiantes en ese grupo está cerca del valor ideal y valores menores que 1 indica mayor desequilibrio:

$$E_{G_i} = 1 - \left| \frac{n_{G_i} - N_{ideal_i}}{N_{ideal_i}} \right|$$

2.3 Número ideal de estudiantes por subgrupo de prácticas:

Será el número ideal de alumnos en el grupo principal N_{ideal_i} entre 2

$$N_{ideal_{s_{G_i}}} = \frac{N_{ideal_i}}{2}$$

2.4 Proporción de equilibrio por subgrupo de prácticas:

Para cada subgrupo de prácticas s_{G_i} de la asignatura C_i , el equilibrio se calcula restando la cantidad ideal de estudiantes $N_{ideal_{s_{G_i}}}$ al número real de estudiantes $n_{s_{G_i}}$ en ese grupo. Esta diferencia se normaliza dividiéndola por $N_{ideal_{s_{G_i}}}$ y el resultado se toma en valor absoluto para evitar resultados negativos. Finalmente, se resta esta proporción de 1 para obtener

una medida del equilibrio, donde valores cercanos a 1 significa que el número de estudiantes en ese subgrupo está cerca del valor ideal y valores menores que existe mayor desequilibrio:

$$E_{s_{G_i}} = 1 - \left| \frac{n_{s_{G_i}} - N_{ideal_{s_{G_i}}}}{N_{ideal_{s_{G_i}}}} \right|$$

2.5 Promedio de equilibrio en la asignatura:

Para calcular el equilibrio total de cada asignatura C_i se calcula con el promedio de las proporciones de todos los grupos de teoría y subgrupos de prácticas.

$$E_{C_i} = \frac{\sum_{g=1}^G E_{G_{ig}} + \sum_{s=1}^{2G} E_{s_{G_{is}}}}{G + 2G}$$

Donde:

- G es el número de grupos principales de teoría de la asignatura.
- $2G$ es el número de subgrupos de prácticas de la asignatura.

Una vez calculado el equilibrio de cada asignatura obligatoria se puede calcular el equilibrio total del individuo de la siguiente manera.

$$E_{total} = \frac{1}{A_o} \sum_{a=1}^{A_o} E_{C_{ia}}$$

Donde:

- A_o es el número de asignaturas obligatorias (con más de un grupo).

3. Tasa de cohesión de teoría por curso para un estudiante (maximizar):

Con este objetivo se representa que un alumno este matriculado en el mismo grupo de teoría para asignaturas del mismo curso, con esto se trata de evitar los tiempos muertos, además también puede favorecer a reducir los solapes entre horarios de asignaturas del mismo curso, se calcula de la siguiente manera. Sea S el conjunto total de estudiantes, y para cada estudiante $s \in S$, tenemos C_s como el conjunto de cursos en los que está matriculado un estudiante s . Para cada curso $c \in C_s$, tenemos un conjunto de asignaturas $A_{s,c}$ en las que el estudiante s está matriculado dentro de ese curso. Además, para cada asignatura $a \in A_{s,c}$, el estudiante está asignado a un grupo de teoría $g_{s,a}$. La tasa de cohesión por curso para un estudiante s en un curso c , denotada como $T_{s,c}$ se calcula como la proporción del grupo más frecuente en el que el estudiante está matriculado dentro de ese curso, respecto al total de asignaturas del curso en las que está matriculado. Esto es:

$$T_{s,c} = \frac{\max(\text{freq}(g_{s,a}))}{|A_{s,c}|} \quad \forall a \in A_{s,c}$$

La tasa de cohesión promedio para un estudiante s , que toma en cuenta todos los cursos en los que está matriculado, se calcula cómo el promedio de las tasas de cohesión por curso de ese estudiante

$$T_s = \frac{1}{|C_s|} \sum_{c \in C_s} T_{s,c}$$

Finalmente, la tasa de cohesión general para todos los estudiantes de un individuo se obtiene cómo el promedio de las tasas de cohesión de todos los estudiantes:

$$T = \frac{1}{|S|} \sum_{s \in S} T_s$$

4. **Tasa de cohesión prácticas por curso para un estudiante** (maximizar): Este objetivo representa lo mismo que el anterior pero esta vez con los grupos de prácticas. Sea S el conjunto total de estudiantes, y para cada estudiante $s \in S$, tenemos C_s cómo el conjunto de cursos en los que está matriculado. Para cada curso $c \in C_s$, tenemos un conjunto de asignaturas $A_{s,c}$ en las que el estudiante está matriculado dentro de ese curso. Además, para cada asignatura $a \in A_{s,c}$, el estudiante está asignado a un subgrupo de prácticas $sg_{s,a}$. La tasa de cohesión por curso para un estudiante s en un curso c , denotada cómo $T_{s,c}$ se calcula cómo la proporción del grupo de prácticas más frecuente en el que el estudiante está matriculado dentro de ese curso, respecto al total de asignaturas del curso en las que está matriculado. Esto es:

$$TP_{s,c} = \frac{\max(\text{freq}(sg_{s,a}))}{|A_{s,c}|} \quad \forall a \in A_{s,c}$$

La tasa de cohesión promedio para un estudiante s , que toma en cuenta todos los cursos en los que está matriculado, se calcula cómo el promedio de las tasas de cohesión por curso de ese estudiante

$$TP_s = \frac{1}{|C_s|} \sum_{c \in C_s} T_{s,c}$$

Finalmente, la tasa de cohesión general para todos los estudiantes de un individuo se obtiene cómo el promedio de las tasas de cohesión de todos los estudiantes:

$$TP = \frac{1}{|S|} \sum_{s \in S} TP_s$$

5. **Tasa coincidencias configuración inicial** (maximizar): Uno de los ficheros iniciales contiene una configuración de matrículas realizada previamente por los estudiantes. Este objetivo busca maximizar la similitud entre la solución generada por el algoritmo y dicha configuración inicial, respetando en la medida de lo posible las preferencias del alumnado.

Para calcular este valor, se comparan las asignaciones de grupo de teoría y subgrupo de prácticas de las matrículas en las asignaturas que disponen de más de un grupo.

Se contabiliza cómo coincidencia cada vez que el grupo de teoría o el subgrupo de prácticas de una matrícula en el individuo coincide con el correspondiente en la configuración inicial.

La tasa total de coincidencias se calcula cómo:

$$TC = \frac{\text{Número total de coincidencias}}{\text{Número total de asignaciones comparables}}$$

Donde:

- El número total de coincidencias corresponde a la suma, para todos los alumnos, de coincidencias en grupo de teoría y subgrupo de prácticas respecto a la configuración original, únicamente en las asignaturas con más de un grupo.
- El número total de asignaciones comparables se calcula cómo el número total de matrículas de asignaturas con más de un grupo, multiplicado por 2 (ya que se comparan dos valores por matrícula: grupo de teoría y subgrupo de prácticas).

3.2. Algoritmo genético

3.2.1. Función fitness

Una vez definidos los objetivos a optimizar y la forma de medirlos, es necesario establecer una función fitness que permita cuantificar la calidad de una solución. Para ello, es fundamental que todos los objetivos se encuentren en una escala comparable, facilitando su combinación en una única métrica global.

En este caso, tres de los cuatro objetivos definidos se encuentran naturalmente en una escala entre 0 y 1. Sin embargo, el número total de solapes es una variable entera positiva no acotada, lo que requiere su normalización. Una estrategia habitual consiste en dividir el valor observado por un valor máximo teórico, pero dado que es difícil estimar el número máximo posible de solapes en este contexto, se opta por una aproximación alternativa.

En este caso, se toma cómo referencia la configuración inicial de matrículas proporcionada, calculando el número total de solapes que presenta. Este valor se utilizará cómo cota superior de referencia para normalizar el número de solapes en cada individuo, de modo que:

$$\text{Solapes Normalizados} = \frac{\text{Solapes}_{\text{individuo}}}{\text{Solapes}_{\text{configuración inicial}}}$$

Este enfoque garantiza una escala relativa adecuada para integrar esta métrica en la función de evaluación, es cierto que, en caso de que algún individuo supere el número de solapes de la configuración inicial, su valor normalizado podrá ser superior a 1. No obstante, este comportamiento es deseable, ya que penaliza aun más a las soluciones con un alto grado de solapamiento.

Una vez normalizadas todas las métricas, se define la función fitness cómo una combinación lineal ponderada de los objetivos. Dado que se trata de una función que se desea maximizar, los objetivos formulados cómo métricas a maximizar se suman, mientras que los objetivos a minimizar se restan. De esta manera, se priorizan las soluciones que presenten mejores valores globales, teniendo en cuenta la importancia relativa de cada criterio a través de los pesos asignados en la ponderación.

$$f(i) = -w_1 * \text{Solapes Totales} + w_2 * E_{total} + w_3 * T + w_4 * TP + w_5 * TC$$

La asignación de pesos en la función de evaluación tiene cómo finalidad reflejar la importancia relativa de cada uno de los objetivos definidos. Tal cómo se ha argumentado anteriormente, el objetivo relacionado con el *equilibrio del número de alumnos por grupo* es fundamental para garantizar la viabilidad de las soluciones en un contexto real, por lo que su peso w_3 debe ser considerablemente más elevado que el del resto.

En este trabajo, los valores de los pesos se mantendrán fijos a lo largo de todas las pruebas experimentales. Se ha asignado el mismo peso a los objetivos de *solapes*, *cohesión de grupos de teoría* y *cohesión de grupos de prácticas*, al considerar que todos ellos contribuyen de manera equilibrada al diseño de un horario funcional y coherente.

Por otro lado, el objetivo relacionado con las preferencias del alumnado (tasa de coincidencias con la configuración inicial) ha recibido un peso ligeramente superior al de los tres mencionados anteriormente, con el fin de priorizar, en cierta medida, la satisfacción de las elecciones realizadas por los estudiantes en su matrícula original.

- $w_1 = 0.1$
- $w_2 = 0.1$
- $w_3 = 0.55$
- $w_4 = 0.1$
- $w_5 = 0.15$

3.2.2. Población inicial

Una vez definida la representación del problema, el siguiente paso consiste en generar la población inicial. El tamaño de esta población estará determinado por un hiperparámetro previamente establecido. Para construir cada individuo, se parte de un clon de la configuración inicial de matrículas, ya que esta contiene la información sobre las asignaturas en las que cada alumno desea matricularse.

Es importante destacar que en todos los individuos de la población se mantiene invariable tanto el orden de los estudiantes cómo el orden de sus matrículas. Esto garantiza una correspondencia directa entre las posiciones del genotipo y las asignaciones específicas de cada alumno, facilitando la aplicación de los operadores genéticos y la evaluación del individuo.

A partir de esta base común, se procede a asignar de forma aleatoria los grupos de teoría y los subgrupos de prácticas. La asignación de grupos de teoría se realiza por curso, de

modo que todas las asignaturas de un mismo curso dentro de un alumno en un individuo pertenecen al mismo grupo de teoría. En cambio, los subgrupos de prácticas se asignan completamente al azar para cada asignatura, sin considerar el curso al que pertenecen.

Esta estrategia de inicialización se adopta con el objetivo de evitar configuraciones excesivamente desestructuradas que puedan dificultar el proceso de evolución. Si cada asignatura tuviera un grupo completamente aleatorio, se generarían soluciones caóticas que requerirían muchas generaciones para converger, o incluso podrían inducir una convergencia prematura hacia óptimos locales de baja calidad. Por tanto, proporcionar una población inicial con cierta coherencia facilita que, a través de pequeñas modificaciones sucesivas, el algoritmo explore soluciones progresivamente mejores.

3.2.3. Selección

El proceso de selección juega un papel fundamental en los algoritmos genéticos, ya que determina qué individuos de la población actual serán seleccionados para reproducirse y formar parte de la próxima generación, ya sea directamente o como padres de nuevos descendientes [Goldberg, 1989]. Este proceso influye directamente en la *presión selectiva*, un concepto que hace referencia al grado de preferencia que se otorga a los individuos con mejor fitness durante la selección.

Existen numerosos métodos de selección, cada uno con sus propias características y efectos sobre la presión selectiva. En este trabajo se han implementado tres de los métodos más representativos, sobre los cuales se realizarán experimentos para determinar cuál ofrece un mejor rendimiento en el problema abordado. Los métodos considerados son:

1. **Basada en fitness:** En este método cada individuo tendrá una probabilidad de ser elegido directamente proporcional al fitness, para calcular la probabilidad que tiene cada individuo se puede crear una distribución de probabilidad usando la fórmula referenciada en [Goldberg, 1989], de forma que la probabilidad de cada individuo es:

$$p_s(i) = \frac{f(i)}{\sum_{i=1}^N f(i)}$$

donde:

- $f(i)$ representa el valor de *fitness* del individuo i ,
- N es el número total de individuos en la población.

El problema de este método es que si existe un individuo muy bueno va a tener muchas probabilidades de ser seleccionado, acaparando la población y evitando la diversidad genética lo que supondrá no hacer una exploración de espacio de soluciones efectiva y pudiendo ocasionar una convergencia prematura y teniendo como resultado un óptimo local.

2. **Basada en rango:** Con este método se trata de reducir la presión selectiva del método anterior. La idea es crear un ranking de individuos según su fitness, donde mayor

fitness hace que ocupen mejores posiciones en el ranking, y con una formula asignar nuevas probabilidades sin tener en cuenta el fitness y sólo teniendo en cuenta la posición en el ranking, teniendo una población de N individuos y que cada individuo ocupa una posición i en el ranking, podemos calcular las nuevas probabilidades selección con la siguiente formula [Baker, 1985]

$$p_s(i) = \frac{2(N - i + 1)}{N^2 + N}$$

Con esta formula los elementos con mejor fitness todavía tendrán mayor probabilidad de ser seleccionados pero se ha reducido bastante, de esta forma los elementos con peor fitness también tendrán posibilidades de ser seleccionados con lo que se tiene mayor diversidad genética y ofrece una mejor exploración del espacio de soluciones.

3. **Selección por torneo:** Este método de selección se basa en el mecanismo de torneo, donde la presión selectiva puede ajustarse mediante un parámetro k , que representa el tamaño del torneo. El procedimiento consiste en seleccionar aleatoriamente k individuos de la población y elegir aquel con mayor valor de fitness como candidato para la siguiente generación. Con este método cada individuo participa en k torneos [Miller and Goldberg, 1995].

La presión selectiva aumenta con el valor de k , ya que se incrementa la probabilidad de seleccionar individuos con fitness elevado. Por el contrario, valores bajos de k favorecen una mayor diversidad genética al ofrecer más oportunidades de selección a individuos con menor fitness. Esta característica permite una exploración más amplia del espacio de soluciones, reduciendo la probabilidad de estancamiento en óptimos locales, aunque puede ralentizar la convergencia del algoritmo.

Un valor intermedio de k puede ofrecer un equilibrio adecuado entre exploración y explotación, permitiendo mantener una presión selectiva suficiente para mejorar progresivamente la población, sin comprometer la diversidad genética.

3.2.4. Cruce

Es el primer operador de variación que se aplica, a partir de dos padres seleccionados a través de alguno de los métodos anteriores se cruzaran para generar hijos nuevos, normalmente el cruce se suele realizar bajo una probabilidad bastante grande, en este problema se hará siempre, con este operador se mezcla los estudiantes de dos individuos, nunca se introduce material genético nuevo, por lo que al hacer esto se está explorando el espacio de soluciones para intentar encontrar una solución mejor a partir de otras soluciones buenas.

Es importante recalcar que el cruce se hará a nivel de estudiante, cada individuo en la población tiene una lista de estudiantes, y cada estudiante tiene a su vez una lista de matrículas variables, el objetivo es optimizar los grupos de teoría y subgrupos de prácticas de cada matrícula variable. Existen dos enfoques de cómo realizar el cruce para este problema:

1. **Cruzar las matrículas variables individualmente:** Para cada estudiante, se pueden intercambiar las matrículas variables entre los padres, lo que significa que algunas provengan del padre 1 y otras del padre 2.

2. **Cruzar la configuración completa de cada estudiante:** De esta manera se selecciona la configuración completa de matrículas de un padre u otro para cada estudiante de forma independiente. Es decir, para cada estudiante Se decide si hereda toda la configuración del padre 1 o del padre 2.

La segunda opción, donde se hereda la configuración completa del estudiante, se considera más adecuada en este problema. Al cruzar individualmente las matrículas se corre el riesgo de desordenar demasiado el material genético, lo que podría ocasionar dificultades en la convergencia de una buena solución. Al heredar la configuración completa de cada estudiante, se realiza una exploración más controlada del espacio de soluciones, lo que facilita modificaciones graduales a través del operador de mutación.

Una vez explicado el esquema del cruce, hay que explicar su implementación. En los algoritmos genéticos existen distintos tipos de operadores de cruce, cada uno tiene unas características y ventajas. En este trabajo se utilizarán tres variantes que se explican a continuación:

1. **Un punto:** En este caso en cada generación se seleccionará un punto de manera aleatoria en la lista de alumnos para cada cruce, de forma que los segmentos de los padres antes de ese punto se mantienen, mientras que los segmentos posteriores se intercambian. Este es uno de los métodos más sencillos, su principal desventaja es que las combinaciones genéticas están limitadas por la ubicación del único punto de cruce.
2. **Varios puntos:** Este método es una evolución del anterior, ahora se eligen dos puntos de corte en cada cruce, las secciones de la lista de alumnos ubicadas entre estos dos puntos se intercambian entre los padres, el resto de segmentos se mantienen igual. Con esto se gana ventaja de aumentar la diversidad genética en comparación con el método anterior, lo que permite explorar un espacio de búsqueda más amplio, aunque puede ser insuficiente si los genes importantes están muy dispersos.
3. **Uniforme:** Este cruce tiene una forma más aleatoria, ya que se basa en una máscara de cruce formada por bits que indica para cada alumno de la lista si proviene del padre 1 o del padre 2, esta máscara será aleatoria en cada cruce. Este método es el que mayor diversidad genética proporciona ya que cada alumno puede provenir de cualquiera de los padres sin seguir una estructura predeterminada. Este método es muy recomendable cuando se tiene que explorar un espacio de soluciones muy amplio, además combina muy bien con mutaciones de baja probabilidad ya que al combinar configuraciones completas de alumnos se harán pequeños cambios que hagan llegar a una buena solución.

Es importante señalar que este operador garantiza la validez de los individuos generados. Al heredarse configuraciones completas de estudiantes y mantenerse la estructura original del genotipo, se evita la creación de individuos inconsistentes o inválidos, lo cual es esencial para mantener la integridad del proceso evolutivo.

3.2.5. Mutación

Este será el segundo operador de variación que se use, se aplicará a los hijos generados en el proceso de cruce, consiste en introducir pequeñas modificaciones aleatorias en el individuo, en este caso la mutación cambiará el grupo de teoría o prácticas en las matrículas variables de los alumnos, al igual que el operador de cruce, el de mutación también ayuda a explorar el espacio de soluciones, pero de manera diferente, ya que el operador de cruce combina el material genético existente entre los individuos, el operador de mutación introduce nuevo material genético que no estaba presente en los padres. Esto da la capacidad al algoritmo genético de explorar áreas del espacio de búsqueda que no hubieran sido accesibles a través del cruce. Antes se ha mencionado que el cruce se aplicará siempre, la mutación es todo lo contrario, se suele aplicar con probabilidades muy pequeñas ya que se puede desorganizar la información genética ya adquirida. A continuación se va a explicar que probabilidades de mutación existen en este problema y cómo funcionan:

- *p_mutacion_alumno*: Este será el primer filtro, para cada alumno de cada individuo se generará un número real aleatorio del 0 al 1 y si es menor que la probabilidad significa que podría mutar y se pasarán al siguiente paso.
- *p_mutacion_teoría*: Para los alumnos que han pasado el primer filtro se realizará el mismo proceso en sus matrículas variables, si se cumple la probabilidad se cambiará el grupo de teoría de esa matrícula por uno distinto al actual
- *p_mutacion_práctica*: Funciona igual que el anterior, se aplicará en las matrículas variables de los alumnos que pasaron el primer filtro, si se cumple la probabilidad se cambiará el grupo de práctica de esa matrícula.

3.2.6. Sustitución de la población

Es una fase crucial en los algoritmos genéticos, en esta fase se eligen que individuos de una generación se mantendrán o reemplazarán en la siguiente. Esto afecta tanto a la convergencia del algoritmo hacia una buena solución como a la diversidad genética de la población a lo largo de las generaciones. Existen muchos tipos de sustitución, a continuación se explicará cuáles se han implementado para este problema y que pros y contras tiene cada uno. En la sustitución está la población con la que se empezó la generación, P_t y existirá otra población que ha sido la resultante de aplicar la selección, cruce y mutación P'

1. **Reemplazo**: Este es el método más sencillo, P' se convertirá en la nueva generación P_{t+1} , de esta manera se hace una exploración intensa del espacio de soluciones, ya que introduce una población nueva en cada generación, las principales desventajas son que puede llevar a una convergencia más lenta ya que no se conserva ninguna información de los mejores individuos de la generación anterior, esto también lleva a perder soluciones de alta calidad si los descendientes no logran alcanzar el mismo nivel de fitness.
2. **Elitismo**: Este método de sustitución tiene como objetivo conservar la mejor solución encontrada hasta el momento, corrigiendo una de las principales limitaciones del método anterior. En este enfoque, la nueva población P_{t+1} se construye seleccionando

todos los individuos generados en la población nueva P' , excepto aquel con peor valor de fitness, y reemplazándolo por el mejor individuo de la población actual P_t . Esta estrategia garantiza que el mejor individuo no se pierda durante el proceso evolutivo, lo que favorece una convergencia más rápida hacia soluciones de alta calidad. Sin embargo, la inclusión excesiva de individuos de generaciones anteriores puede reducir la diversidad genética y aumentar el riesgo de estancamiento en óptimos locales. Por este motivo, en esta implementación se conserva únicamente el mejor individuo de P_t , lo que representa un compromiso entre preservación de calidad y mantenimiento de diversidad en la población.

3. **Truncamiento:** Este método de sustitución combina la población actual P_t y la generada P' para formar la nueva población P_{t+1} . El procedimiento consiste en unir ambas poblaciones y ordenarlas de forma descendente según su valor de fitness. A continuación, se seleccionan los individuos con mayor fitness hasta alcanzar el tamaño de población original. En caso de empate entre individuos de P_t y P' , se prioriza la inclusión del individuo perteneciente a P' . Esta estrategia garantiza que la siguiente generación esté compuesta por los individuos más destacados en términos de calidad, lo que puede acelerar la convergencia del algoritmo hacia soluciones óptimas. Además, permite una preservación selectiva de características favorables presentes tanto en los progenitores como en los descendientes, facilitando un equilibrio entre explotación y exploración del espacio de búsqueda. Este método tiene el problema de que al eliminar por completo a los individuos con menor fitness, puede reducir considerablemente la diversidad genética de la población. En consecuencia, existe un mayor riesgo de estancamiento en óptimos locales si no se combina adecuadamente con operadores de mutación y cruce que favorezcan la introducción de variabilidad.

3.2.7. Condición de parada

La condición de parada define el criterio mediante el cuál el algoritmo genético finaliza su ejecución. Dado que se trata de un proceso iterativo orientado a mejorar progresivamente la calidad de las soluciones a lo largo de múltiples generaciones, establecer un criterio adecuado de finalización resulta fundamental para alcanzar un equilibrio entre la calidad de la solución obtenida y la eficiencia computacional.

Existen diversos enfoques para definir esta condición. Una opción es detener la ejecución cuando se alcanza una solución óptima. Sin embargo, este enfoque presenta limitaciones importantes, especialmente en problemas complejos en los que no se conoce a priori el valor óptimo, lo que podría implicar un coste computacional elevado e innecesario.

Otra alternativa común es establecer la parada en función de la convergencia de la población, es decir, cuando las soluciones generadas en sucesivas generaciones presentan escasa variabilidad. No obstante, este método puede inducir una convergencia prematura hacia óptimos locales, y su detección requiere un seguimiento continuo de la diversidad poblacional, lo cuál incrementa la complejidad del algoritmo.

En esta implementación se ha optado por utilizar un enfoque basado en un número máximo de generaciones como criterio de parada. Este método presenta la ventaja de ser sencillo de implementar y garantiza un control directo sobre el tiempo de ejecución. La principal

desventaja radica en que una cantidad insuficiente de generaciones puede impedir alcanzar soluciones de calidad, mientras que un número excesivo puede suponer un desperdicio de recursos computacionales una vez alcanzada la convergencia. Por este motivo, el número de generaciones se establecerá a partir de los resultados obtenidos en la fase de experimentación, buscando un compromiso adecuado entre eficiencia y rendimiento.

3.2.8. Resultados Ejecución

Cómo resultado de la ejecución del algoritmo genético, se generan dos ficheros de salida principales que documentan los resultados finales de la asignación de alumnos a los grupos de teoría y subgrupos de práctica. Estos ficheros contienen información sobre la configuración de las matrículas para cada alumno y las distribución total de alumnos en cada grupo de teoría y subgrupo de práctica para cada asignatura.

La generación del fichero que contiene las matrículas de los alumnos implica realizar la decodificación del individuo, transformando su representación interna (genotipo) en una solución completa y comprensible (fenotipo). Esta decodificación consiste en reconstruir la configuración de matrícula de cada estudiante, combinando las asignaciones variables (optimizadas por el algoritmo) con las asignaciones fijas de asignaturas con grupo único.

El objetivo de estos ficheros es permitir un análisis detallado de la solución obtenida y facilitar la interpretación de los resultados finales.

- *Fichero de configuraciones de matrículas:*

Este fichero almacena la configuración de matrículas de la solución con mayor fitness del algoritmo genético, la estructura sera igual que la del fichero de matrículas que se ha explicado en el capítulo anterior.

- *Fichero de distribución de alumnos por grupo:*

Este fichero constituye un elemento clave dentro del conjunto de resultados generados, ya que permite analizar el grado de cumplimiento del objetivo de equilibrio de alumnos entre los distintos grupos de teoría y subgrupos de prácticas de cada asignatura. Este objetivo es especialmente relevante, ya que una distribución desequilibrada puede derivar en asignaciones inviables desde el punto de vista logístico, debido a las limitaciones físicas de las aulas. Por ello, este archivo resulta esencial para evaluar la viabilidad real de las soluciones propuestas por el algoritmo.

El fichero se presenta en formato Excel y contiene dos hojas diferenciadas:

- La primera hoja recoge la distribución de alumnos correspondiente a la configuración inicial, es decir, aquella proporcionada como punto de partida para la ejecución del algoritmo.
- La segunda hoja muestra la distribución resultante de la mejor solución obtenida por el algoritmo genético tras su ejecución.

En ambas hojas, las columnas del fichero, ordenadas de izquierda a derecha, son las siguientes:

- **NOMBRE:** Nombre completo de la asignatura
- **ALUMNOS:** Número total de alumnos matriculados en la asignatura
- **ESPERADOS POR GRUPO TEORÍA:** Número ideal de alumnos en cada grupo de teoría para esa asignatura
- **ESPERADOS POR GRUPO PRÁCTICAS:** Número ideal de alumnos en cada subgrupo de prácticas para esa asignatura
- **GRUPO 10:** Número de alumnos asignados al grupo 10
- **GP1:** Número de alumnos en el primer subgrupo de prácticas del grupo 10
- **GP2:** Número de alumnos en el segundo subgrupo de prácticas del grupo 10
- **GRUPO 11:** Número de alumnos asignados al grupo 11
- **GP1:** Número de alumnos en el primer subgrupo de prácticas del grupo 11
- **GP2:** Número de alumnos en el segundo subgrupo de prácticas del grupo 11
- **GRUPO 12:** Número de alumnos asignados al grupo 12
- **GP1:** Número de alumnos en el primer subgrupo de prácticas del grupo 12
- **GP2:** Número de alumnos en el segundo subgrupo de prácticas del grupo 12

Además, para facilitar la identificación visual de posibles desequilibrios, junto a cada columna que indica el número de alumnos asignados a un grupo o subgrupo, se incluye una columna adicional que refleja el grado de desviación respecto al valor ideal. Cuando esta desviación es igual o superior al 15 %, la celda correspondiente se resalta con un fondo rojo. Esta visualización permite detectar rápidamente situaciones problemáticas y valorar de forma más precisa la viabilidad de la solución desde el punto de vista del equilibrio de carga entre grupos.

3.3. NSGA-III

3.3.1. Población Inicial

Igual que en el algoritmo genético, el primer paso consiste en generar una población inicial de soluciones. El número de individuos de la población inicial lo marca un hiperparámetro, se usará la misma lógica descrita anteriormente 3.2.2, cada uno de ellos se creará clonando la configuración inicial que contiene las asignaturas en las que desea matricularse cada alumno. Posteriormente, se asignarán aleatoriamente los grupos de teoría y prácticas para cada asignatura de forma coherente, un grupo de teoría común por curso y grupos de prácticas aleatorios para cada asignatura, con el objetivo de evitar soluciones excesivamente caóticas y favorecer una convergencia más eficiente.

En NSGA-III esto es la primera parte, una vez generada esta población inicial P_t , aplicando los operadores de cruce y mutación se creará una segunda población Q_t del mismo tamaño. A continuación, ambas poblaciones se combinan en una única población extendida $R_t = P_t \cup Q_t$, esta unión será la que se usará como población inicial.

Esta combinación es necesaria ya que en los siguientes pasos del algoritmo, hay que seleccionar N individuos para formar la nueva generación. si sólo se crea la única población

inicial P_t , no existiría un proceso de competición real, ya que todos los individuos pasarían automáticamente a la siguiente generación.

3.3.2. Puntos de referencia

La principal innovación del algoritmo NSGA-III frente a sus predecesores es el uso de puntos de referencia para guiar el proceso de selección y mantener la diversidad de las soluciones en problemas de optimización con múltiples objetivos.

Estos puntos de referencia permiten distribuir las soluciones de forma equilibrada a lo largo del espacio de los objetivos, facilitando así una representación variada del conjunto de soluciones no dominadas. Para lograr esto, se proyectan las soluciones generadas por el algoritmo sobre un conjunto de direcciones predefinidas en dicho espacio.

La técnica empleada para generar estos puntos fue propuesta por Das y Dennis [Das and Dennis, 1998]. Su método consiste en construir una malla regular de puntos en el espacio de objetivos utilizando un parámetro H , que representa el número de divisiones por objetivo.

Se generan todas las combinaciones posibles de números enteros no negativos (a_1, a_2, \dots, a_m) , donde m es el número total de objetivos, que cumplan la siguiente condición:

$$\sum_{i=1}^m a_i = H, \quad a_i \in \mathbb{Z}_{\geq 0}$$

Cada a_i representa el peso relativo que se asigna al objetivo i dentro de una combinación determinada. Estas combinaciones se convierten en puntos de referencia al normalizarse, es decir, al dividir cada componente entre H :

$$\mathbf{p} = \left(\frac{a_1}{H}, \frac{a_2}{H}, \dots, \frac{a_m}{H} \right), \quad \sum_{i=1}^m p_i = 1$$

El resultado es un conjunto de puntos distribuidos uniformemente que guían la selección de soluciones en cada generación del algoritmo. Gracias a este enfoque, NSGA-III logra mantener una buena diversidad incluso cuando se optimizan muchos objetivos, donde otros algoritmos tienden a fallar.

El parámetro H determina el nivel de granularidad o resolución de los puntos generados, esto significa que cuanto mayor sea H , mayor será el número de puntos, por lo que la cobertura del espacio objetivo será mejor. Sin embargo, esto también aumenta el coste computacional.

El número total de puntos generados mediante este proceso se obtiene aplicando la fórmula de combinaciones con repetición:

$$N = \binom{H + m - 1}{m - 1}$$

3.3.3. Clasificación no dominada

Una vez evaluados todos los individuos de la población, el siguiente paso consiste en clasificarlos en distintos frentes de no dominancia. Esta clasificación permite organizar las soluciones según su grado de optimalidad relativa, y es fundamental para guiar el proceso evolutivo en algoritmos como NSGA-III.

El primer frente estará formado por aquellos individuos que no son dominados por ningún otro, es decir, soluciones que no presentan otra que sea mejor en todos los objetivos simultáneamente. El segundo frente contendrá individuos que sólo son dominados por los del primer frente, el tercero por los dos anteriores, y así sucesivamente.

Para llevar a cabo esta clasificación, se emplea el siguiente procedimiento:

1. Para cada individuo de la población se calcula:
 - El número de veces que es dominado por otros individuos (contador de dominancia).
 - Una lista con los individuos a los que domina.
2. Los individuos cuyo contador de dominancia sea cero (es decir, no dominados por ningún otro) se asignan al primer frente.
3. A continuación, para cada individuo en el frente actual, se recorren los elementos a los que domina y se decrementa en una unidad su contador de dominancia. Si tras esta operación el contador de un individuo llega a cero, se incluye en el siguiente frente.
4. Este proceso se repite iterativamente hasta que todos los individuos hayan sido asignados a un frente.

Este método garantiza una separación jerárquica de la población basada en su calidad relativa respecto a los objetivos del problema, lo que permite aplicar una presión selectiva ordenada y progresiva durante el proceso evolutivo.

3.3.4. Asignación a puntos de referencia

Una vez que los individuos han sido organizados en frentes de no dominancia, el siguiente paso en NSGA-III consiste en asignarlos a puntos de referencia.

Para realizar esta asignación correctamente, es necesario normalizar previamente los valores de los objetivos de cada individuo. Aunque estos ya se encuentren en el rango $[0, 1]$, la normalización garantiza que todas las soluciones estén representadas de forma coherente con los puntos de referencia definidos. Esto es fundamental para evitar sesgos en las distancias entre soluciones y puntos de referencia, lo que podría llevar a asignaciones incorrectas durante el proceso evolutivo.

Dado un individuo \mathbf{x} con objetivos $f_1(\mathbf{x}), f_2(\mathbf{x}), \dots, f_m(\mathbf{x})$, la normalización se realiza cómo:

$$f_i^{\text{norm}}(\mathbf{x}) = \frac{f_i(\mathbf{x}) - z_i^{\text{mín}}}{z_i^{\text{máx}} - z_i^{\text{mín}} + \varepsilon}$$

donde:

-
- z_i^{\min} y z_i^{\max} son los valores mínimo y máximo observados del objetivo i en la población actual,
 - ε es una constante pequeña (por ejemplo, 10^{-10}) que evita divisiones por cero,
 - y los objetivos a maximizar deben invertirse previamente, convirtiéndolos en problemas de minimización.

Una vez normalizados, los individuos se proyectan sobre todos los puntos de referencia para identificar el más cercano. En NSGA-III esta cercanía se mide mediante la *distancia perpendicular* a la dirección que define cada punto de referencia, no mediante la distancia euclidiana directa. Esta elección mejora la asignación al identificar mejor la región del espacio objetivo que representa cada punto.

Sea $\mathbf{x} \in \mathbb{R}^m$ el vector de objetivos normalizado del individuo y $\mathbf{r} \in \mathbb{R}^m$ un punto de referencia, la distancia perpendicular se define cómo:

$$d_{\perp}(\mathbf{x}, \mathbf{r}) = \left\| \mathbf{x} - \left(\frac{\mathbf{x} \cdot \mathbf{r}}{\|\mathbf{r}\|^2} \right) \mathbf{r} \right\|$$

donde:

- $\mathbf{x} \cdot \mathbf{r}$ es el producto escalar,
- $\|\mathbf{r}\|^2$ es la norma cuadrada de \mathbf{r} ,
- y $\|\cdot\|$ indica la norma euclidiana (L2).

Cada individuo se asocia finalmente al punto de referencia con menor distancia perpendicular. Este mecanismo permite detectar regiones del frente de Pareto que están escasamente representadas, facilitando una selección más equitativa y mejorando la diversidad global. Este aspecto es especialmente importante en problemas con muchos objetivos, donde las soluciones tienden a agruparse en ciertas zonas densas del espacio objetivo.

3.3.5. Selección

Una vez que los individuos han sido clasificados en frentes de no dominancia y asignados a su punto de referencia más cercano, se procede al proceso de selección para determinar cuáles pasarán a la siguiente generación. El número total de individuos a seleccionar está definido por un hiperparámetro que representa el tamaño de la población.

El procedimiento comienza añadiendo directamente todos los individuos de los primeros frentes, mientras no se supere el tamaño máximo. Sin embargo, cuando se alcanza un frente cuyo número de individuos excede la capacidad restante, se requiere aplicar un criterio adicional para decidir cuáles deben ser seleccionados.

En este paso, cobran importancia los puntos de referencia, ya que permiten preservar la diversidad en el espacio de soluciones. Para ello, se consideran todos los puntos de referencia y se ordenan en orden ascendente en función del número de individuos que ya han sido seleccionados y están asociados a cada punto.

A continuación, se revisan los puntos de referencia ordenados. si un punto tiene actualmente cero individuos asociados (es decir, ningún individuo seleccionado hasta ese momento lo tiene cómo punto más cercano), se distinguen dos casos:

- *sí ningún individuo del frente actual* tiene ese punto como más cercano, se pasa al siguiente punto.
- *sí uno o más individuos del frente actual* tienen ese punto como más cercano, se selecciona:
 - directamente *sí hay un sólo individuo*;
 - el individuo con menor distancia perpendicular al punto, en caso de que haya varios.

En caso de que el punto de referencia ya tenga uno o más individuos asociados, se comprueba *sí alguno de los individuos del frente actual* tiene ese punto como más cercano. *sí es así*, se selecciona:

- el único individuo, *sí sólo hay uno*;
- o el más cercano al punto, *sí hay varios*.

Este proceso se repite iterativamente hasta completar el número exacto de individuos requeridos para la siguiente generación, garantizando de este modo un compromiso adecuado entre calidad de las soluciones (según su nivel de no dominancia) y su dispersión en el espacio objetivo (diversidad).

Una vez seleccionados los individuos que formarán la siguiente generación, el siguiente paso consiste en aplicar los operadores genéticos de cruce y mutación para generar la nueva descendencia.

Antes de aplicar el cruce, se realizan $\frac{N}{2}$ iteraciones (siendo N el tamaño de la población), en cada una de las cuales se seleccionan dos padres sobre los que se aplicará el operador de cruce, generando así dos hijos. A continuación, a los hijos resultantes se les aplicará la mutación.

En el desarrollo del algoritmo genético clásico se implementaron distintos métodos de selección basados en el valor de la función fitness. Sin embargo, en NSGA-III, al tratarse de un algoritmo evolutivo multiobjetivo, no se dispone de una única función de fitness, por lo que ha sido necesario adaptar el método de *selección por torneo*.

En este caso, también se introduce un hiperparámetro k que indica el tamaño del torneo. Para seleccionar un padre, se eligen aleatoriamente k individuos de la población, y se escoge aquel que pertenezca al mejor frente de no dominancia. En caso de que varios candidatos estén en el mismo frente, se selecciona el individuo cuya distancia al punto de referencia asignado sea menor, fomentando así la diversidad.

Al igual que ocurre en el algoritmo clásico, un valor más alto de k implica una mayor precisión selectiva, ya que al tener más candidatos en el torneo, es más probable que se escojan individuos pertenecientes a frentes de mejor calidad.

3.3.6. Cruce y Mutación

Una vez completado el proceso de selección descrito anteriormente, se procede a aplicar los operadores genéticos de cruce y mutación para generar la nueva población. En esta

implementación del algoritmo NSGA-III, se utilizarán los mismos operadores que los definidos previamente en el desarrollo del algoritmo genético clásico.

Se mantienen los tres tipos de cruce implementados (cruce de un punto, de varios puntos y uniforme), así como la estructura del operador de mutación con sus respectivos parámetros de probabilidad: $p_{mutacion_alumno}$, $p_{mutacion_teoria}$ y $p_{mutacion_practica}$. Esta elección permite conservar la lógica de exploración y explotación del espacio de soluciones ya diseñada, aplicándola ahora en un contexto multiobjetivo.

3.3.7. Reemplazo

A diferencia de lo que ocurre en el algoritmo genético clásico, donde se aplican distintos métodos de reemplazo en función del valor de la función fitness, en NSGA-III se realiza una sustitución completa de la generación. Este enfoque se alinea con la lógica evolutiva del algoritmo, centrada en la clasificación por frentes de no dominancia y el mantenimiento de la diversidad mediante puntos de referencia.

Después de aplicar el proceso de selección, cruce y mutación, se ha obtenido una población de N individuos. Sin embargo, como se ha explicado al inicio, el algoritmo requiere una población extendida de tamaño $2N$ para realizar el proceso evolutivo completo. Por tanto, es necesario generar otros N individuos adicionales a partir de los ya existentes de nuevo.

Para ello, se evalúa a los individuos de la población y se realiza nuevamente la ordenación de la población actual en frentes de no dominancia, seguida de la asignación de los individuos a sus puntos de referencia más cercanos. Con esta estructura, se lleva a cabo una nueva selección de padres, sobre los cuales se aplican los operadores de cruce y mutación. De esta forma, se obtienen nuevos descendientes, que se añaden a la población existente para formar la población extendida de individuos.

Una vez formada esta población extendida, se vuelve a empezar el ciclo. Esto se repite de forma iterativa hasta que se cumpla la condición de parada definida.

3.3.8. Condición de parada

El último paso consiste en definir la condición de parada del algoritmo. Al igual que en el desarrollo del algoritmo genético clásico, se establece un número máximo de generaciones como límite principal para la ejecución. Sin embargo, debido a la mayor complejidad computacional que implica NSGA-III, tanto por la evaluación multiobjetivo como por la clasificación en frentes y el uso de puntos de referencia, se introduce una condición adicional con el objetivo de agilizar las pruebas y optimizar el uso de recursos.

Una de las características clave de NSGA-III es su capacidad para mantener la diversidad entre las soluciones a lo largo de las generaciones. Por tanto, si se detecta que el mejor frente no ha variado entre generaciones consecutivas, se considerará que el algoritmo ha entrado en un estado de estancamiento. Esta falta de evolución sugiere que el algoritmo ha dejado de explorar nuevas regiones del espacio de soluciones, lo cual justifica su detención anticipada. Esta estrategia permite ahorrar tiempo de cómputo sin comprometer la calidad del resultado final, especialmente en casos donde el algoritmo ya ha convergido o ha dejado de mostrar mejoras significativas.

3.3.9. Resultados de la Ejecución

En el caso del algoritmo NSGA-III, los ficheros de salida generados siguen la misma estructura descrita en la sección 3.2.8, correspondiente al algoritmo genético clásico. La principal diferencia es el criterio de selección de las soluciones exportadas, mientras que en el caso clásico se guarda únicamente la solución con mayor valor de *fitness*, en NSGA-III se consideran todos los individuos que forman parte del frente de Pareto final.

Para cada uno de estos individuos, se crea una carpeta independiente dentro de la ruta de resultados. En cada carpeta se generan los mismos dos archivos explicados anteriormente:

- *Fichero de configuraciones de matrículas*: contiene la asignación concreta de grupos para el individuo correspondiente.
- *Fichero de distribución de alumnos por grupo*: muestra el reparto de estudiantes en grupos de teoría y subgrupos de prácticas, comparado con la distribución inicial, permitiendo evaluar el equilibrio alcanzado.

Esta estrategia permite un análisis individualizado de cada solución no dominada generada por el algoritmo, facilitando la comparación entre ellas y ofreciendo al usuario final varias opciones que representan diferentes compromisos entre los objetivos definidos.

4. Experimentación

4.1. Introducción

En el capítulo anterior se ha descrito el desarrollo de los dos algoritmos genéticos propuestos para resolver el problema, el algoritmo genético clásico y el algoritmo NSGA-III. Ambos algoritmos cuentan con diferentes mecanismos en algunos pasos como el cruce, así como una serie de hiperparámetros que influyen directamente en su comportamiento y en la calidad de las soluciones obtenidas.

Dado que no existe una configuración única que garantice los mejores resultados para todos los casos, en este capítulo se presentará una serie de experimentos con el objetivo de identificar las combinaciones más eficaces de métodos e hiperparámetros. Para ello se ha diseñado una metodología en tres fases:

- En primer lugar, se evaluarán todas las combinaciones posibles de métodos de selección, cruce y sustitución del algoritmo genético clásico, utilizando una configuración fija de hiperparámetros.
- A partir de los mejores resultados obtenidos en la fase anterior, se llevará a cabo un segundo bloque de experimentación para estudiar distintas combinaciones de hiperparámetros, usando las mejores combinaciones de métodos obtenidas en la primera fase.
- Finalmente, se aplicará el conocimiento obtenido en las fases anteriores para definir y ejecutar una serie de experimentos con el algoritmo NSGA-III, evaluando el impacto de los hiperparámetros en su rendimiento.

Este enfoque progresivo permite analizar de forma estructurada el comportamiento de ambos algoritmos y facilita la identificación de configuraciones óptimas para el problema planteado.

4.2. Entorno de ejecución

Todos los experimentos se han realizado en una maquina con un procesador AMD Ryzen 9 7900X 12-Core Processor a 4.70 GHz y 64 GB de memoria ram, en cuanto al software se ha usado python 3.9. Esta configuración asegura que los resultados obtenidos son reproducibles y permite comparar el rendimiento de distintas configuraciones de manera consistente.

4.3. Evaluación de métodos del algoritmo genético

En esta fase el objetivo es evaluar distintas combinaciones de métodos de selección, cruce y sustitución usando una configuración estática de hiperparámetros. Con esto se podrá identificar la configuración de métodos que maximiza la calidad de la solución y la estabilidad de la convergencia, estableciendo una base sólida para la fase posterior de optimización de hiperparámetros. La idea es probar todas las combinaciones que existen de los métodos que se han implementado. Antes se han descrito los ficheros de salida que se crean en la ejecución de este algoritmo genético 3.2.8.

Esos ficheros sirven para ver la configuración resultante de las matrículas de los alumnos y el equilibrio en los grupos de cada asignatura, pero no se pueden usar para analizar la evolución completa del algoritmo.

Además, debido a que los algoritmos genéticos incorporan un alto grado de aleatoriedad en procesos clave como la generación de la población inicial, la selección, la mutación o el cruce, una única ejecución puede no ser representativa del comportamiento general del algoritmo. Por ello, cada combinación será ejecutada un total de 10 veces. Los resultados mostrados en el análisis corresponderán al promedio de estas ejecuciones, lo que permite obtener una visión más fiable y robusta del rendimiento real de cada configuración, evitando así la posibilidad de que un buen resultado se deba únicamente al azar.

Dado el elevado número de combinaciones a analizar, no sería práctico ejecutar manualmente cada una por separado. Para ello, se ha desarrollado un *script* automatizado que ejecuta todas las combinaciones y genera ficheros con los resultados necesarios para su análisis posterior. Dado que la ejecución de cada combinación es independiente y se dispone de una máquina con un procesador potente, el *script* hace uso de concurrencia para aprovechar al máximo los recursos disponibles, permitiendo que las combinaciones se ejecuten en paralelo. Este enfoque reduce significativamente el tiempo total de ejecución en comparación con una ejecución secuencial, aunque cada combinación individual pueda requerir ligeramente más tiempo debido a la carga compartida.

El código completo del *script* se encuentra disponible en el repositorio enlazado en el Anexo A.1.

Los ficheros que se crean con este script que ayudarán a identificar las mejores combinaciones son los explicados a continuación.

- *resumen.csv*: Este fichero contiene una vista agregada de cada combinación de métodos de selección, cruce y sustitución tras 10 ejecuciones. Las columnas principales son:

- **(Selección, Cruce, Sustitución):** Identifica la combinación de métodos utilizada.
 - **Generación(μ):** Generación promedio en la que se ha obtenido el mejor individuo.
 - **Solapes(μ), Cohesión Teoría(μ), Equilibrio(μ), Cohesión Prácticas(μ), Preferencias(μ):** Valores medios de la mejor ejecución de cada objetivo a lo largo de las ejecuciones.
 - **Fitness(μ):** Valor medio del mejor fitness obtenido.
 - **Fitness(σ):** Desviación estándar del mejor fitness, útil para evaluar la estabilidad de la combinación.
 - **Tiempo(μ):** Tiempo medio total de ejecución por combinación.
- *seleccion-cruce-sustitucion.csv:*
Este fichero contiene el seguimiento generación a generación para una combinación específica. Las columnas incluyen:
- **Generación:** Número de la generación actual.
 - **Fitness(μ):** Valor medio del mejor fitness en esa generación.
 - **Solapes(μ), Cohesión Teoría(μ), Equilibrio(μ), Cohesión Prácticas(μ), Preferencias(μ):** Valores medios de los objetivos del mejor individuo en la generación.
 - **Tiempo(μ):** Tiempo que ha tardado la generación en ejecutarse.

Con estos ficheros se puede analizar de una forma sencilla que combinaciones son las que mejor funcionan, con el primer fichero se detectará rápidamente cuales son las que mejores resultados están dando, y con el segundo fichero se puede analizar mejor cada combinación y ver si ha convergido o cómo ha ido evolucionando el fitness.

La configuración fija de hiperparámetros que se ha usado para esta prueba es la siguiente:

- *poblacion* = 500
- *generaciones* = 50
- *mutacion_elemento* = 0.00009
- *mutacion_teoría* = 0.0005
- *mutacion_practica* = 0.0005
- *t_torneo* = 10

Después de ejecutar el script de experimentación de la primera fase se procede al análisis del contenido del fichero de salida *resumen.csv*, a continuación se muestra ordenado de mayor a menor fitness:

Selección, Cruce, Sustitución	Generación(μ)	Solapes(μ)	Cohesión Teoría(μ)	Equilibrio(μ)	Cohesión Prácticas(μ)	Preferencias(μ)	Fitness(μ)	Fitness(σ)	Tiempo(μ)
(torneo uniforme reemplazo)	50	148	0.9997	0.9590	0.7310	0.5888	0.7577	0.0022	9.42 min
(torneo uniforme elitismo)	50	165	0.9997	0.9675	0.7353	0.5796	0.7577	0.0018	18.74 min
(torneo uniforme truncamiento)	50	199	0.9998	0.9711	0.7220	0.5557	0.7485	0.0041	18.74 min
(rango varios_puntos truncamiento)	50	241	0.9995	0.9504	0.7169	0.5014	0.7186	0.0021	18.01 min
(rango varios_puntos elitismo)	50	230	0.9997	0.9407	0.7190	0.5142	0.7176	0.0036	18.13 min
(rango punto reemplazo)	48	231	0.9997	0.9421	0.7148	0.5035	0.7165	0.0026	9.12 min
(rango punto elitismo)	49	240	0.9997	0.9429	0.7186	0.5032	0.7156	0.0029	18.01 min
(rango varios_puntos reemplazo)	49	228	0.9998	0.9353	0.7189	0.5131	0.7153	0.0032	9.18 min
(rango uniforme truncamiento)	48	213	1.0000	0.9247	0.7165	0.5308	0.7150	0.0023	18.82 min
(torneo varios_puntos elitismo)	27	250	0.9998	0.9468	0.7201	0.4925	0.7140	0.0057	17.82 min
(rango punto truncamiento)	44	258	0.9997	0.9484	0.7158	0.4984	0.7137	0.0033	19.86 min
(rango uniforme elitismo)	46	210	0.9999	0.9236	0.7174	0.5181	0.7134	0.0021	18.90 min
(torneo varios_puntos reemplazo)	26	254	0.9996	0.9479	0.7166	0.4914	0.7132	0.0030	8.90 min
(fitness punto truncamiento)	49	242	0.9997	0.9408	0.7151	0.4970	0.7127	0.0020	18.06 min
(fitness varios_puntos truncamiento)	50	246	0.9999	0.9397	0.7146	0.5047	0.7127	0.0020	18.17 min
(rango uniforme reemplazo)	48	216	0.9999	0.9225	0.7202	0.5230	0.7124	0.0018	9.58 min
(torneo varios_puntos truncamiento)	22	254	0.9997	0.9461	0.7136	0.4855	0.7109	0.0048	17.76 min
(torneo punto elitismo)	25	254	0.9998	0.9440	0.7129	0.4917	0.7103	0.0053	17.81 min
(torneo punto truncamiento)	24	250	1.0000	0.9397	0.7133	0.4840	0.7089	0.0032	17.87 min
(torneo punto reemplazo)	18	254	0.9998	0.9407	0.7130	0.4882	0.7087	0.0025	8.92 min
(fitness uniforme truncamiento)	48	229	1.0000	0.9232	0.7164	0.5055	0.7065	0.0029	19.06 min
(fitness punto elitismo)	34	267	1.0000	0.9229	0.7111	0.4701	0.6935	0.0015	18.19 min
(fitness varios_puntos elitismo)	40	273	1.0000	0.9233	0.7086	0.4813	0.6933	0.0023	18.41 min
(fitness varios_puntos reemplazo)	34	272	0.9999	0.9209	0.7128	0.4748	0.6917	0.0027	9.15 min
(fitness uniforme elitismo)	23	270	0.9999	0.9191	0.7115	0.4746	0.6914	0.0024	19.02 min
(fitness punto reemplazo)	22	275	1.0000	0.9184	0.7124	0.4783	0.6910	0.0018	9.03 min
(fitness uniforme reemplazo)	22	277	1.0000	0.9217	0.7105	0.4713	0.6903	0.0014	9.55 min
Tiempo Total									3.16 hr

Tabla 4.1: Resumen de resultados promediados tras 10 ejecuciones por combinación

A primera vista se puede observar que los mejores métodos de selección y cruce son torneo y uniforme respectivamente, ya que esa combinación con los distintos métodos de sustitución forman las tres mejores combinaciones, también se puede ver que en las tres mejores la mejor generación ha sido la última, esto es buena señal ya que significa que la evolución no se ha quedado estancada y ha ido mejorando a lo largo de las generaciones haciendo una búsqueda en el espacio de soluciones de manera óptima, por ejemplo, tomando la última combinación, la media de la mejor generación ha sido 22, lo que significa que se ha tenido una convergencia prematura y no ha podido evolucionar más, en las siguientes figuras se puede observar como ha ido evolucionando el mejor fitness medio en cada generación en las ejecuciones de la mejor y peor combinación.

También se puede observar que la desviación estandar es muy baja en todas las combinaciones, señal que indica que el funcionamiento del algoritmo es estable.

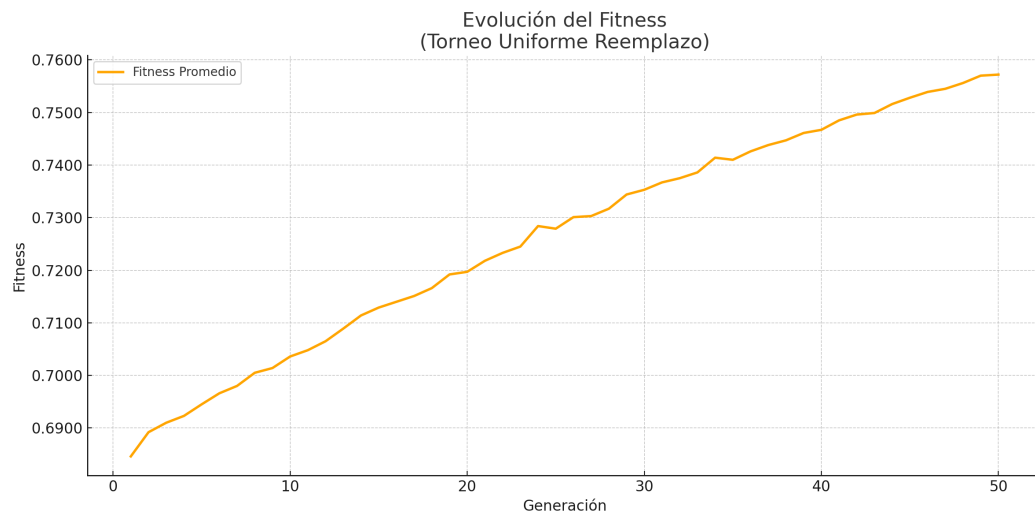


Figura 4.1: Evolución del mejor fitness promedio para la combinación Torneo Uniforme Reemplazo

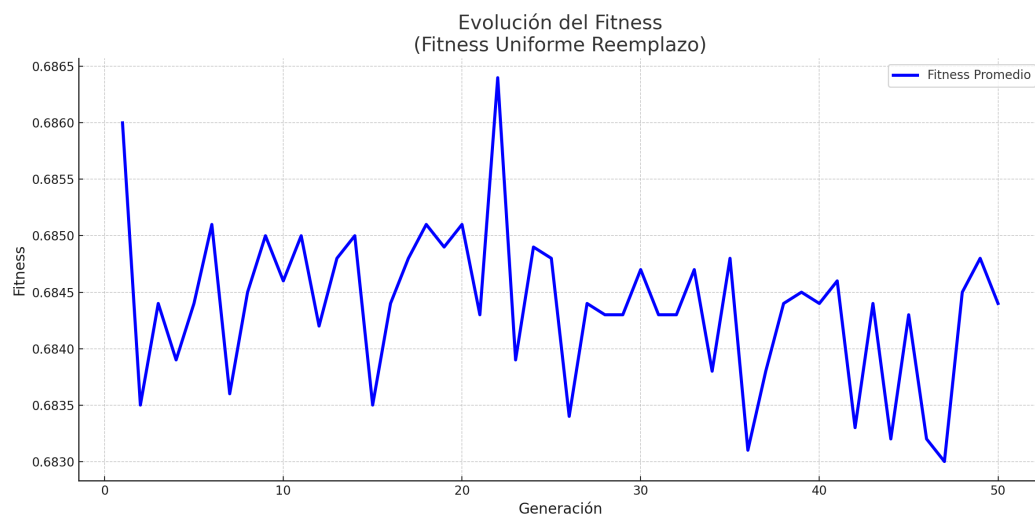


Figura 4.2: Evolución del mejor fitness promedio para la combinación Fitness Uniforme Reemplazo

Al analizar las gráficas de evolución del fitness para ambas combinaciones de métodos, se aprecian diferencias claras en cuanto al comportamiento del algoritmo y su capacidad para generar soluciones de calidad.

En el caso de la combinación que utiliza fitness-uniforme-reemplazo, se observa una evolución del fitness inestable, con múltiples oscilaciones a lo largo de las generaciones. No

se aprecia una tendencia clara de mejora, lo que sugiere que el algoritmo tiene dificultades para converger hacia soluciones óptimas. Esta inestabilidad puede deberse a que, al no aplicarse ningún tipo de elitismo, los mejores individuos pueden perderse en cada generación. Además, al usar la selección basada en fitness hay una gran presión selectiva, por lo que no se usara mucha diversidad para hacer el cruce. Como consecuencia, el algoritmo tiende a explorar de forma ineficiente el espacio de soluciones, sin una dirección clara de mejora.

En cambio, la combinación que aplica selección por torneo junto con cruce uniforme y sustitución por reemplazo muestra un comportamiento mucho más estable y efectivo. En este caso, se observa una mejora progresiva del fitness durante todas las generaciones, lo que indica una buena convergencia hacia soluciones de calidad. La selección por torneo con este tamaño introduce una menor presión selectiva, favoreciendo a los individuos con menor rendimiento a poder participar en el cruce, lo que guía el proceso evolutivo permitiendo explorar mejor el espacio de soluciones, esto es importante ya que el proceso de cruce entre individuos con fitness medio o incluso bajo puede generar combinaciones más prometedoras.

Volviendo a la figura 4.1, en las peores combinaciones, aunque la mayoría de los objetivos de la solución final presentan valores aceptables, el factor determinante que compromete su viabilidad es el objetivo de equilibrio de alumnos entre los distintos grupos. Un análisis más detallado del fichero de distribución de alumnos revelaría un notable desequilibrio en varias asignaturas, con asignaciones que superan claramente la capacidad real de las aulas. Este aspecto es especialmente crítico, ya que convierte la solución en inviable desde el punto de vista logístico, independientemente del rendimiento alcanzado en los demás objetivos.

Si se sigue analizando, se observa que los métodos de selección y sustitución de la mejor combinación (torneo y reemplazo), usados con el cruce de varios puntos ofrece también malos resultados. Esto resalta la relevancia del operador de cruce en el rendimiento del algoritmo, especialmente cuando se combina con selección por torneo y el método de reemplazo. En contraste, la tercera combinación posible con esa misma selección y sustitución pero con el cruce restante, obtiene una de las peores posiciones en el ranking de resultados, presentando además una evolución del fitness similar a la combinación ineficiente ya analizada.

En estos casos, el uso de un torneo con tamaño reducido genera una presión selectiva muy baja, lo que limita la capacidad del algoritmo para favorecer a los mejores individuos. Al combinarse con el método de reemplazo, que no garantiza la preservación de los mejores individuos entre generaciones, el papel del cruce se vuelve crítico. El análisis confirma que los operadores de cruce de un punto y de varios puntos no son adecuados para este problema, ya que no permiten una exploración eficaz del espacio de soluciones.

Los resultados muestran que en la mayoría de combinaciones analizadas no se ha producido un estancamiento temprano, por el contrario, el algoritmo ha continuado mejorando su rendimiento hasta las últimas generaciones. Dado que existen tres opciones por cada uno de los métodos de selección, cruce y sustitución, el número total de combinaciones posibles es $3^3 = 27$. Si además se consideran seis hiperparámetros y se evaluarán tres opciones para cada uno, se obtendrían $3^6 = 729$ combinaciones adicionales. La evaluación exhaustiva de todas estas combinaciones implicaría un coste computacional elevado.

Como conclusión, dado que las tres mejores combinaciones presentan una diferencia notable respecto al resto y comparten los mismos métodos de selección y cruce, se ha decidido reducir el espacio de búsqueda y centrarse exclusivamente en dichas combinaciones para la fase posterior de optimización de hiperparámetros.

4.4. Evaluación de hiperparámetros del algoritmo genético

Tras la primera fase de experimentación, se ha observado que las mejores combinaciones de métodos corresponden a la pareja formada por selección por torneo y cruce uniforme, independientemente del método de sustitución empleado. Esta combinación ha ofrecido resultados consistentemente superiores, por lo que se utilizará como base para esta segunda fase.

En esta etapa se evaluarán de forma sistemática los distintos métodos de sustitución en combinación con la pareja fija de selección-cruce (torneo-uniforme). Además, se explorarán diferentes configuraciones de los hiperparámetros clave del algoritmo (cómo el tamaño de población, número de generaciones, tamaño del torneo, y probabilidades de mutación), con el objetivo de identificar una configuración óptima que maximice la calidad de las soluciones para el problema planteado.

Al igual que en la fase anterior, se ha desarrollado un script automatizado de experimentación que está disponible en el repositorio A.1, que permite ejecutar múltiples combinaciones de forma concurrente. Esto no sólo optimiza el uso de los recursos computacionales disponibles, sino que también permite reducir significativamente el tiempo necesario para completar el conjunto de experimentos.

Dado el carácter estocástico de los algoritmos genéticos, cada combinación será ejecutada 5 veces para asegurar la validez estadística de los resultados obtenidos.

Así, los ficheros generados por este segundo script reflejan el promedio de las métricas a lo largo de todas las ejecuciones. Esta estrategia permite analizar de forma más robusta y fiable el comportamiento global e individual del algoritmo bajo diferentes configuraciones.

■ *resumen.csv*:

En este fichero se recoge información agregada de todas las combinaciones probadas. Las columnas que contiene, de izquierda a derecha, son:

- **Hiperparámetros**: Representa la configuración de hiperparámetros utilizada, se mostrarán en el siguiente orden: (Población, Generaciones, Mutación_Elemento, Mutación_Teoría, Mutación_Práctica, Torneo).
- **(Selección, Cruce, Sustitución)**: Indica la combinación de métodos utilizada.
- **Mejor Generación(μ)**: Generación promedio en la que se ha obtenido el mejor individuo.
- **Mejor Fitness(μ)**: Valor promedio del fitness y los objetivos del mejor individuo, después del fitness se mostrarán el valor de los objetivos en el siguiente orden: solapes, cohesión teoría, equilibrio, cohesión prácticas y tasa preferencias.

-
- **Fitness(σ)**: Desviación estándar del mejor fitness, útil para evaluar la estabilidad de la combinación.
 - **Tiempo**: Tiempo promedio que ha tardado en ejecutarse la combinación.
- *Carpetas por combinación de hiperparámetros*:
Para organizar los resultados, se genera una carpeta por cada combinación de hiperparámetros evaluada. Dentro de cada carpeta se encuentra un fichero con nombre *selección-cruce-sustitución.csv*, con el mismo formato que en la fase anterior. Este fichero contiene información detallada por generación:
- **Generación**: Número de generación.
 - **Fitness**: Valor del mejor fitness en esa generación.
 - **Solapes, Cohesión Teoría, Equilibrio, Cohesión Prácticas, Preferencias**: Valores de los objetivos del mejor individuo en la generación.
 - **Fitness(μ)**: Valor medio del fitness en esa generación.
 - **Fitness(σ)**: Desviación estándar del fitness en la generación.
 - **Tiempo**: Tiempo de ejecución de esa generación.

Con estos ficheros, al igual que en la fase anterior, se pueden identificar rápidamente las configuraciones más prometedoras y analizar en detalle la evolución de una combinación concreta a lo largo de las generaciones.

cómo se ha explicado previamente, en esta fase los métodos de selección y cruce se mantienen fijos, empleando la combinación identificada como más prometedora en la fase anterior. A partir de esta configuración base, se evaluarán los distintos métodos de sustitución junto con múltiples combinaciones de hiperparámetros, con el objetivo de encontrar la configuración más adecuada para el problema planteado.

Los valores considerados para cada hiperparámetro durante esta fase de experimentación son los siguientes:

- *población* = [100, 300, 500]
- *generaciones* = [50, 80, 100]
- *mutacion_elemento* = [0.00009]
- *mutacion_teoría* = [0.0005, 0.005, 0.05]
- *mutacion_práctica* = [0.0005, 0.005, 0.05]
- *t_torneo* = [10, 20, 50]

En la primera fase se evaluaron un total de 27 combinaciones, resultado de aplicar las tres variantes implementadas para cada uno de los métodos de selección, cruce y sustitución. En esta segunda fase, al mantener fijos los métodos de selección y cruce, el número de combinaciones posibles depende de los seis hiperparámetros restantes junto con los tres métodos de sustitución. Si se probaran tres valores por cada uno de ellos, el número total de combinaciones ascendería a $3^6 \times 3 = 2187$, lo cual supondría una carga computacional muy elevada.

Para reducir significativamente el número de ejecuciones sin perder representatividad en los resultados, se ha optado por fijar un único valor para el hiperparámetro *mutacion_elemento*, dado que la variabilidad ya está suficientemente cubierta mediante los valores asignados a *mutacion_teoría* y *mutacion_práctica*. Con esta decisión, el número total de combinaciones se reduce a $3^5 \times 3 = 729$, lo que permite una exploración amplia y eficiente del espacio de configuraciones con un coste computacional razonable.

Después de ejecutar el script correspondiente a la segunda fase de experimentación, se analiza el archivo `resumen.csv`, el cuál recoge los resultados promedios obtenidos para cada combinación de selección, cruce y método de sustitución evaluada. A partir de este fichero, es posible analizar comparativamente el rendimiento de las distintas configuraciones en función de su capacidad de optimización.

En la Tabla 4.2 se muestran los primeros 50 registros, ordenados de forma descendente según su valor de fitness, lo que permite identificar rápidamente qué combinaciones han ofrecido los mejores resultados en términos de calidad de solución.

Hiperparámetros	Selección Cruce Sustitución	Mejor Gen.(μ)	Mejor Fitness(μ)	Mejor Fitness(σ)	Tiempo(μ)
500-100-9e-05-0.0005-0.0005-10	(torneo uniforme reemplazo)	79	0.7726 (141 0.9991 0.9769 0.7379 0.6098)	0.0028	18.79 min
500-100-9e-05-0.0005-0.005-10	(torneo uniforme reemplazo)	77	0.7722 (151 0.9994 0.9774 0.7387 0.6155)	0.0018	18.66 min
500-80-9e-05-0.0005-0.005-10	(torneo uniforme reemplazo)	77	0.7721 (148 0.9995 0.9768 0.7382 0.6174)	0.0035	15.04 min
500-80-9e-05-0.0005-0.0005-10	(torneo uniforme reemplazo)	75	0.7688 (147 0.9997 0.9783 0.7323 0.5913)	0.0042	14.91 min
500-80-9e-05-0.0005-0.0005-10	(torneo uniforme elitismo)	73	0.7686 (149 0.9994 0.9763 0.7394 0.5915)	0.0035	29.95 min
500-80-9e-05-0.0005-0.005-10	(torneo uniforme elitismo)	69	0.7672 (155 0.9995 0.9765 0.7349 0.5951)	0.0018	30.44 min
500-100-9e-05-0.005-0.005-10	(torneo uniforme reemplazo)	73	0.7660 (161 0.9956 0.9736 0.7367 0.6090)	0.0016	18.69 min
500-100-9e-05-0.0005-0.005-10	(torneo uniforme elitismo)	70	0.7659 (155 0.9996 0.9776 0.7315 0.5847)	0.0053	37.22 min
500-100-9e-05-0.0005-0.05-10	(torneo uniforme reemplazo)	70	0.7658 (156 0.9997 0.9692 0.7298 0.6145)	0.0015	18.62 min
500-80-9e-05-0.005-0.0005-10	(torneo uniforme reemplazo)	73	0.7654 (167 0.9958 0.9767 0.7349 0.6036)	0.0032	15.11 min
500-100-9e-05-0.005-0.0005-10	(torneo uniforme reemplazo)	74	0.7651 (174 0.9949 0.9758 0.7361 0.6115)	0.0027	18.63 min
500-80-9e-05-0.0005-0.05-10	(torneo uniforme reemplazo)	73	0.7649 (154 0.9995 0.9710 0.7326 0.6039)	0.0022	15.12 min
500-100-9e-05-0.0005-0.0005-10	(torneo uniforme elitismo)	72	0.7641 (161 0.9995 0.9775 0.7338 0.5816)	0.0051	37.52 min
500-80-9e-05-0.005-0.005-10	(torneo uniforme reemplazo)	67	0.7633 (162 0.9955 0.9713 0.7336 0.6024)	0.0036	15.04 min
500-80-9e-05-0.005-0.0005-10	(torneo uniforme elitismo)	70	0.7628 (168 0.9954 0.9757 0.7334 0.5901)	0.0047	30.42 min
500-100-9e-05-0.0005-0.05-10	(torneo uniforme reemplazo)	66	0.7613 (170 0.9964 0.9675 0.7353 0.6086)	0.0017	18.72 min
500-80-9e-05-0.0005-0.05-10	(torneo uniforme elitismo)	65	0.7611 (164 0.9996 0.9703 0.7259 0.5939)	0.0023	29.90 min
500-100-9e-05-0.0005-0.05-10	(torneo uniforme elitismo)	64	0.7606 (167 0.9998 0.9703 0.7329 0.5920)	0.0036	37.23 min
500-80-9e-05-0.005-0.05-10	(torneo uniforme reemplazo)	63	0.7595 (169 0.9965 0.9650 0.7318 0.6108)	0.0045	15.05 min
500-50-9e-05-0.0005-0.0005-10	(torneo uniforme reemplazo)	50	0.7594 (149 0.9999 0.9584 0.7357 0.5995)	0.0023	9.44 min
300-100-9e-05-0.0005-0.0005-10	(torneo uniforme reemplazo)	68	0.7593 (171 0.9998 0.9744 0.7320 0.5764)	0.0029	11.19 min
500-100-9e-05-0.0005-0.0005-10	(torneo uniforme elitismo)	67	0.7588 (180 0.9960 0.9751 0.7326 0.5860)	0.0027	37.21 min
300-80-9e-05-0.0005-0.005-10	(torneo uniforme reemplazo)	63	0.7568 (176 0.9997 0.9745 0.7347 0.5645)	0.0019	9.01 min
500-50-9e-05-0.0005-0.005-10	(torneo uniforme reemplazo)	50	0.7568 (157 0.9998 0.9587 0.7296 0.5972)	0.0011	9.48 min
300-100-9e-05-0.0005-0.005-10	(torneo uniforme reemplazo)	59	0.7566 (180 0.9997 0.9741 0.7277 0.5728)	0.0046	11.20 min
500-50-9e-05-0.0005-0.005-10	(torneo uniforme elitismo)	50	0.7564 (166 0.9997 0.9601 0.7348 0.5939)	0.0024	18.71 min
300-80-9e-05-0.0005-0.0005-10	(torneo uniforme reemplazo)	60	0.7561 (183 0.9999 0.9747 0.7287 0.5703)	0.0024	9.06 min
500-50-9e-05-0.0005-0.0005-10	(torneo uniforme reemplazo)	50	0.7560 (153 0.9988 0.9563 0.7332 0.5932)	0.0014	9.46 min
500-80-9e-05-0.005-0.005-10	(torneo uniforme elitismo)	60	0.7560 (181 0.9957 0.9714 0.7344 0.5770)	0.0066	30.00 min
500-100-9e-05-0.005-0.005-10	(torneo uniforme elitismo)	61	0.7557 (183 0.9948 0.9715 0.7345 0.5776)	0.0054	37.29 min
500-100-9e-05-0.0005-0.0005-20	(torneo uniforme reemplazo)	48	0.7553 (189 0.9999 0.9751 0.7278 0.5690)	0.0031	18.54 min
500-50-9e-05-0.0005-0.0005-10	(torneo uniforme elitismo)	50	0.7553 (168 0.9997 0.9656 0.7305 0.5751)	0.0029	18.80 min
300-50-9e-05-0.0005-0.005-10	(torneo uniforme reemplazo)	50	0.7553 (175 1.0000 0.9709 0.7279 0.5706)	0.0019	5.75 min
500-100-9e-05-0.005-0.05-10	(torneo uniforme elitismo)	62	0.7550 (174 0.9967 0.9648 0.7328 0.5901)	0.0036	37.42 min
500-50-9e-05-0.005-0.005-10	(torneo uniforme reemplazo)	50	0.7548 (159 0.9978 0.9567 0.7360 0.5895)	0.0039	9.42 min
300-50-9e-05-0.0005-0.0005-10	(torneo uniforme reemplazo)	50	0.7548 (172 0.9998 0.9697 0.7264 0.5648)	0.003	5.73 min
500-50-9e-05-0.005-0.005-10	(torneo uniforme elitismo)	48	0.7545 (169 0.9977 0.9630 0.7329 0.5799)	0.0039	18.76 min
300-100-9e-05-0.005-0.0005-10	(torneo uniforme reemplazo)	58	0.7545 (190 0.9969 0.9750 0.7276 0.5703)	0.0026	11.21 min
500-50-9e-05-0.0005-0.05-10	(torneo uniforme elitismo)	50	0.7542 (166 1.0000 0.9583 0.7301 0.5920)	0.0008	18.99 min
500-50-9e-05-0.0005-0.0005-10	(torneo uniforme elitismo)	49	0.7540 (171 0.9965 0.9637 0.7274 0.5832)	0.001	18.75 min
500-80-9e-05-0.005-0.05-10	(torneo uniforme elitismo)	61	0.7539 (182 0.9964 0.9669 0.7289 0.5868)	0.0053	29.97 min
300-80-9e-05-0.005-0.005-10	(torneo uniforme reemplazo)	59	0.7537 (184 0.9976 0.9740 0.7227 0.5651)	0.0043	9.10 min
300-80-9e-05-0.005-0.0005-10	(torneo uniforme reemplazo)	59	0.7535 (184 0.9970 0.9730 0.7268 0.5676)	0.0043	8.98 min
500-50-9e-05-0.0005-0.05-10	(torneo uniforme reemplazo)	49	0.7535 (160 0.9999 0.9543 0.7345 0.5907)	0.0017	9.47 min
500-50-9e-05-0.0005-0.0005-20	(torneo uniforme reemplazo)	45	0.7534 (183 0.9997 0.9731 0.7269 0.5625)	0.0029	9.37 min
300-80-9e-05-0.0005-0.05-10	(torneo uniforme reemplazo)	56	0.7530 (178 0.9997 0.9696 0.7292 0.5616)	0.0022	9.13 min
500-80-9e-05-0.0005-0.005-20	(torneo uniforme reemplazo)	45	0.7526 (184 0.9998 0.9732 0.7259 0.5571)	0.0015	14.81 min
300-50-9e-05-0.005-0.0005-10	(torneo uniforme reemplazo)	50	0.7526 (182 0.9983 0.9683 0.7308 0.5676)	0.0033	5.75 min
500-80-9e-05-0.0005-0.0005-10	(torneo uniforme truncamiento)	59	0.7521 (182 0.9996 0.9728 0.7223 0.5544)	0.003	29.84 min
500-50-9e-05-0.005-0.05-10	(torneo uniforme reemplazo)	49	0.7520 (164 0.9973 0.9556 0.7321 0.5863)	0.003	9.47 min

Tabla 4.2: Resultados segundo script de experimentación

Analizando los resultados obtenidos en esta fase de experimentación, se observa que las mejores combinaciones utilizan el método de sustitución por reemplazo, un detalle importante es que al igual que la fase anterior la desviación estandar es muy baja por lo que el comportamiento del algoritmo genético es consistente. La primera combinación que emplea elitismo aparece en la quinta posición del ranking, mientras que la primera con truncamiento no está en las mejores 50 combinaciones. Estos resultados refuerzan la idea de que, para este problema en particular, los métodos que favorecen una alta presión selectiva tienden a explorar menos el espacio de búsqueda, lo cuál limita la capacidad del algoritmo para alcanzar soluciones de calidad.

En la fase anterior, el método de reemplazo ya se posicionó como el más eficaz, mientras que truncamiento, a pesar de estar en tercera posición, mostró diferencias de fitness no sig-

nificativas respecto a las mejores combinaciones. En esta segunda fase, el comportamiento se confirma, el reemplazo permite una renovación completa de la población sin priorizar exclusivamente a los individuos con mayor fitness. Esto favorece la conservación de diversidad genética, un factor clave para alcanzar buenas soluciones, ya que estas no siempre surgen de los mejores individuos, sino de otros con potencial genético complementario.

En lo que respecta a las combinaciones de hiperparámetros, se identifica una tendencia clara entre las configuraciones más destacadas. La mayoría utiliza una población de 500 individuos, el valor más alto considerado, lo cuál incrementa la diversidad de la población. También se observa un uso mayoritario de un tamaño de torneo igual a 10, el valor más bajo, lo que reduce la presión selectiva y permite que un mayor número de individuos tenga opciones de contribuir a la siguiente generación.

En cuanto a las probabilidades de mutación, las configuraciones más efectivas emplean principalmente los dos valores más bajos tanto en la mutación de teoría cómo en la de práctica. Esto sugiere que una intensidad de mutación baja favorece una evolución más estable, en contraposición a mutaciones más agresivas que pueden romper estructuras genéticas útiles.

Respecto al número de generaciones, la mayoría de las mejores combinaciones utilizan 100 u 80 generaciones. Sin embargo, al revisar en qué generación se obtuvo el mejor fitness, se observa que en la mayoría de los casos no se alcanzó el máximo. Esto indica que el algoritmo tiende a converger antes, por lo que un número elevado de generaciones no siempre es necesario. Por ejemplo, la combinación en primera posición usó 100 generaciones, mientras que la misma configuración con 80 generaciones aparece en cuarta posición, habiendo convergido de media en la generación 75. Esta diferencia puede atribuirse a la variabilidad introducida en la generación inicial, lo que sugiere que esta tercera combinación podría considerarse la más eficiente para este problema, al ofrecer un resultado competitivo con menor coste computacional.

4.5. Evaluación del algoritmo NSGA-III

Esta será la última fase del proceso de experimentación. En las fases anteriores se evaluaron todas las combinaciones posibles del algoritmo genético clásico, identificando las configuraciones de métodos e hiperparámetros más efectivas. En esta última fase se aplicará todo lo aprendido para estudiar el comportamiento del algoritmo NSGA-III desarrollado en este trabajo.

A diferencia del algoritmo clásico, NSGA-III dispone de un único método para la selección y sustitución de individuos. Basándose en los resultados obtenidos anteriormente, se utilizará el método de cruce *uniforme*, que ha demostrado ofrecer los mejores resultados en las fases previas.

El resto de hiperparámetros utilizados serán los mismos que en la fase anterior, con la inclusión de un nuevo parámetro propio de NSGA-III, el número de particiones, que representa el parámetro H que se usa en la fórmula para determinar el número de puntos de referencia generados que se explica en 3.3.2. A continuación se detallan los valores considerados:

-
- *población* = [100, 300, 500]
 - *generaciones* = [50, 80, 100]
 - *mutacion_elemento* = [0.00009]
 - *mutacion_teoría* = [0.0005, 0.005, 0.05]
 - *mutacion_práctica* = [0.0005, 0.005, 0.05]
 - *t_torneo* = [10, 20, 50]
 - *particiones* = [6, 7, 8]

Al igual que en las fases anteriores, se ha implementado un script de experimentación que ejecuta de forma concurrente todas las combinaciones posibles de estos hiperparámetros. El script se encuentra disponible en el repositorio referenciado en el Anexo A.1.

El script generará diversos ficheros que facilitarán el análisis posterior de los resultados. Estos archivos permiten evaluar tanto la diversidad del frente de Pareto como la calidad de las soluciones encontradas.

- *resumen.xlsx*: Este archivo recopila los resultados globales de todas las combinaciones de hiperparámetros evaluadas. Contiene dos hojas:
 - **resumen**: Cada fila representa una combinación distinta de hiperparámetros. Las columnas son:
 - **Población**: Tamaño de la población utilizada.
 - **Generaciones**: Número total de generaciones ejecutadas.
 - **m_elemento**: Tasa de mutación para los elementos.
 - **m_teoría**: Tasa de mutación aplicada a la parte teórica.
 - **m_práctica**: Tasa de mutación aplicada a la parte práctica.
 - **t_torneo**: Tamaño del torneo en la selección.
 - **Particiones**: Número de particiones empleadas para generar los puntos de referencia.
 - **Número de elementos**: Número de individuos distintos en el frente de Pareto final.
 - **Tiempo**: Tiempo que ha tardado la combianción.

El número de elemntos permite evaluar la diversidad de las soluciones obtenidas. Por ejemplo, si una combinación ha generado un frente de Pareto formado por soluciones idénticas, se registrará como un único elemento, evidenciando pérdida de diversidad. Por el contrario, combinaciones con múltiples soluciones únicas indican una mayor exploración del espacio de búsqueda.

- **Datos**: Presenta el detalle de los individuos únicos del frente de Pareto final para cada combinación, organizados en bloques con el siguiente formato:
 - **Fila 1**: Cabecera con los nombres de los hiperparámetros usados.
 - **Fila 2**: Valores concretos para esa configuración.
 - **Fila 3**: Indicador del número de soluciones únicas (ej. "Elementos: 5").

- **Fila 4:** Cabecera con las métricas de evaluación:
 - **Solapes**
 - **Cohesión Teoría**
 - **Equilibrio**
 - **Cohesión Prácticas**
 - **Preferencias**
 - **Filas siguientes:** Métricas de cada solución única.
- **Carpetas por combinación de hiperparámetros:** Por cada combinación específica se genera una carpeta con nombre basado en los valores utilizados. Dentro de cada carpeta se encuentra un archivo:
- **resumen.xlsx:** Este archivo tiene el mismo formato que el anterior, pero cada bloque corresponde a una generación concreta dentro de la combinación. La estructura es:
 - **resumen:**
 - **Generación:** Número de generación.
 - **Número de elementos:** Cantidad de soluciones únicas del frente en esa generación.
 - **Tiempo:** Tiempo que ha tardado la combinación.
 - **Datos:** Organizada por bloques, uno por generación. Cada bloque contiene:
 - **Fila 1:** Título de la generación.
 - **Fila 2:** Tiempo que ha tardado cada generación.
 - **Fila 3:** Número de elementos únicos (ej. "Elementos: 7").
 - **Fila 4:** Cabecera con las métricas:
 - **Solapes**
 - **Cohesión Teoría**
 - **Equilibrio**
 - **Cohesión Prácticas**
 - **Preferencias**
 - **Filas siguientes:** Métricas correspondientes a cada solución única del frente de la generación.

Este archivo permite analizar la evolución de la diversidad y la calidad del frente de Pareto a lo largo del tiempo, detectando posibles convergencias prematuras o pérdidas de diversidad en determinadas fases de la ejecución.

Con todos estos ficheros, se puede realizar un análisis exhaustivo sobre el comportamiento del algoritmo NSGA-III en términos de diversidad, calidad de las soluciones, y estabilidad del proceso evolutivo en función de los diferentes valores de los hiperparámetros considerados.

Una vez ejecutado el script correspondiente, se procede al análisis de los resultados obtenidos. En la Tabla 4.3 se muestran las diez primeras combinaciones evaluadas del fichero *resumen.xlsx*.

Población	Generaciones	m_elemento	m_teoria	m_practica	t_torneo	particiones	Nº Elementos	Tiempo
100	50	0.00009	0.0005	0.0005	10	6	1	4.55 min
100	50	0.00009	0.0005	0.0005	10	7	6	6.00 min
100	50	0.00009	0.0005	0.0005	10	8	1	6.17 min
100	50	0.00009	0.0005	0.0005	20	6	1	3.23 min
100	50	0.00009	0.0005	0.0005	20	7	1	2.45 min
100	50	0.00009	0.0005	0.0005	20	8	1	3.13 min
100	50	0.00009	0.0005	0.0005	50	6	1	1.50 min
100	50	0.00009	0.0005	0.0005	50	7	1	1.43 min
100	50	0.00009	0.0005	0.0005	50	8	1	1.47 min
100	50	0.00009	0.0005	0.005	10	6	1	4.13 min

Tabla 4.3: Primeras 10 combinaciones del tercer script de experimentación

Al observar los datos recogidos, se constata que ninguna de estas combinaciones ha logrado aproximarse a un frente de Pareto de calidad. En concreto, la combinación más destacada apenas alcanza seis soluciones únicas en su frente final. Esta escasa diversidad se explica principalmente por el uso de valores bajos en los hiperparámetros relativos al tamaño de población y número de generaciones. Aunque se han combinado diferentes valores de tamaño de torneo y número de particiones, la limitada capacidad de exploración asociada a poblaciones pequeñas y a un número reducido de generaciones ha impedido que el algoritmo explore de forma efectiva el espacio de soluciones.

Este resultado indica que, en el contexto de NSGA-III, ciertos hiperparámetros tienen un impacto más determinante en la calidad del frente obtenido. En particular, la falta de diversidad en el frente de Pareto no puede ser compensada únicamente ajustando el tamaño del torneo o el número de particiones utilizadas para generar los puntos de referencia. Por tanto, asegurar una población suficientemente grande y un número adecuado de generaciones se vuelve esencial para garantizar una buena exploración del espacio y la generación de soluciones diversas y de calidad.

Haciendo un análisis detallado de la combinación que ha generado un frente de Pareto con seis soluciones únicas, que es la combinación con mayor diversidad obtenida en diez primeras combinaciones. En la Tabla 4.4, se muestran los valores de los cinco objetivos para cada uno de los individuos pertenecientes al frente final.

poblacion	generaciones	m_elemento	m_teoría	m_practica	t_torneo	particiones
100	50	0,00009	0,0005	0,0005	10	7
Elementos				6		
Solapes	Cohesión Teoría	Equilibrio	Cohesión Práctica	Preferencias		
184	0,9998	0,8975	0,7447	0,5703		
183	0,9998	0,897	0,7447	0,5696		
185	0,9998	0,8975	0,7447	0,5710		
185	0,9998	0,897	0,7445	0,5719		
183	0,9998	0,8965	0,7445	0,5705		
184	0,9998	0,897	0,7445	0,5713		

Tabla 4.4: Elementos frente de pareto de una combinación

Al examinar los resultados, se observa que varios de los individuos presentan valores favorables en algunos de los objetivos respecto a los alcanzados por el algoritmo genético clásico en las mejores combinaciones. Sin embargo, también se identifican casos en los que los valores obtenidos son inferiores. En concreto, el objetivo relativo al equilibrio de alumnos entre grupos, identificado a lo largo del trabajo como uno de los más relevantes, dado su impacto directo en la viabilidad real de las soluciones, presenta valores notablemente más bajos que los obtenidos con el enfoque clásico.

Esta deficiencia en el equilibrio sugiere que, aunque se han conseguido soluciones con diversidad en términos de otros objetivos, las configuraciones resultantes tendrían un grado elevado de desequilibrio en ciertas asignaturas, comprometiendo así su aplicabilidad práctica debido a la limitación física del espacio en las aulas y a la distribución desigual de estudiantes.

En la siguiente tabla 4.5 se presentan las diez combinaciones de hiperparámetros que han generado un mayor número de soluciones únicas en el frente de Pareto final. El análisis de estas combinaciones revela varios patrones relevantes.

Población	Generaciones	m_elemento	m_teoría	m_practica	t_torneo	particiones	Nº Elementos	Tiempo
500	100	0.00009	0.05	0.0005	10	8	393	57.03 min
500	80	0.00009	0.05	0.005	20	7	391	44.07 min
500	80	0.00009	0.005	0.0005	10	8	386	46.85 min
500	80	0.00009	0.0005	0.005	10	6	384	43.23 min
500	100	0.00009	0.05	0.005	10	7	381	52.90 min
500	80	0.00009	0.005	0.0005	10	6	380	43.20 min
500	100	0.00009	0.05	0.05	10	8	380	51.82 min
500	80	0.00009	0.005	0.005	10	6	374	43.43 min
500	80	0.00009	0.005	0.0005	10	7	373	44.52 min
500	100	0.00009	0.05	0.005	10	8	369	55.42 min

Tabla 4.5: Primeras 10 combinaciones del tercer script de experimentación con mayor diversidad

En primer lugar, todas las combinaciones con mayor diversidad utilizan el valor más alto establecido para el tamaño de la población, así como un número elevado de generaciones.

Además, se observa que, en general, estas configuraciones emplean un tamaño de torneo más pequeño. En cuanto al número de particiones utilizadas, se encuentran representadas todas las opciones evaluadas, al igual que sucede con las probabilidades de mutación, lo que indica que estos parámetros, aunque influyentes, no presentan una tendencia clara en relación directa con la diversidad.

Estos resultados refuerzan la conclusión de que valores elevados en el número de individuos por generación y en el número total de generaciones favorecen significativamente la diversidad de soluciones encontradas. Esto se puede contrastar con los datos previamente analizados en la Tabla 4.3, donde las primeras diez combinaciones evaluadas, caracterizadas por tamaños reducidos de población y generaciones, no lograron generar prácticamente diversidad en el frente de Pareto.

Otro punto importantes es que, en las combinaciones que aparecen en la Tabla 4.5, se alcanza una media de casi 400 soluciones únicas por frente, lo cuál resulta especialmente notable si se considera que cada generación contiene 500 individuos. Dado el funcionamiento del algoritmo NSGA-III, donde se combinan la población actual y la descendencia antes de la selección (es decir, se evalúan un total de 1000 individuos por generación), este resultado implica que aproximadamente un 40 % de los individuos finales acaban formando parte del frente de Pareto. Este porcentaje refleja una alta diversidad y sugiere que las configuraciones con mayor tamaño poblacional y número de generaciones permiten una exploración más amplia del espacio de soluciones.

Sin embargo, cómo se ha comentado anteriormente, el hecho de que una combinación produzca un número elevado de soluciones en el frente de Pareto no implica necesariamente que dichas soluciones sean de alta calidad. Por esta razón, se ha optado por enfocar el análisis en las combinaciones que han obtenido los mejores resultados en el objetivo de equilibrio de alumnos por grupo, el cuál, cómo se ha argumentado a lo largo del trabajo, representa el criterio más relevante para evaluar la viabilidad de las soluciones generadas.

En la siguiente tabla 4.6 se muestra la combinación de hiperparámetros que ha conseguido el mejor valor para dicho objetivo. Es importante destacar que esta combinación no coincide con las que previamente se presentaron cómo las que ofrecen mayor diversidad, lo que ya sugiere un compromiso entre diversidad y calidad en ciertos objetivos.

poblacion	generaciones	m_elemento	m_teoria	m_practica	t_torneo	particiones
300	50	9e-05	0.0005	0.0005	10	6
Elementos				164		
Solapes	Cohesion Teoria	Equilibrio	Cohesion prácticas	Preferencias		
145	1	0.9097	0.7865	0.6366		
145	1	0.9203	0.7883	0.6307		
154	1	0.9233	0.7898	0.6408		
147	1	0.9182	0.7912	0.6349		
152	1	0.9271	0.7917	0.6324		
152	1	0.9284	0.7913	0.628		
146	1	0.9033	0.7887	0.6385		
153	1	0.9333	0.7913	0.625		
149	1	0.9233	0.7888	0.6377		
150	1	0.924	0.791	0.6341		
157	1	0.9222	0.7908	0.6422		
150	1	0.9197	0.7889	0.64		
152	1	0.9232	0.7935	0.6251		
146	1	0.9278	0.7893	0.6276		
149	1	0.9171	0.7892	0.6422		
150	1	0.9175	0.7894	0.6372		

152	1	0.9136	0.7924	0.6348
147	1	0.9172	0.7856	0.6426
147	1	0.9303	0.7894	0.6296
153	1	0.9151	0.7914	0.6362
145	1	0.9125	0.7903	0.6303
146	1	0.9193	0.7911	0.6255
146	1	0.9178	0.7899	0.6304
151	1	0.9293	0.7901	0.6276
153	1	0.9205	0.7946	0.6315
152	1	0.928	0.7908	0.6347
146	1	0.9127	0.7907	0.6357
156	1	0.9251	0.7918	0.6337
152	1	0.9332	0.7908	0.6366
149	1	0.9026	0.7887	0.6455
150	1	0.9159	0.794	0.628
151	1	0.9107	0.7943	0.6308
151	1	0.9183	0.794	0.6237
146	1	0.9124	0.7902	0.6372
155	1	0.9219	0.7884	0.6423
149	1	0.9187	0.788	0.6387
153	1	0.915	0.7939	0.6321
155	1	0.9291	0.7926	0.6273
156	1	0.9119	0.795	0.6344
151	1	0.9158	0.7916	0.631
153	1	0.9262	0.7927	0.6326
148	1	0.9289	0.7891	0.6314
152	1	0.9159	0.7923	0.6319
148	1	0.9186	0.7883	0.6407
150	1	0.915	0.7904	0.6367
147	1	0.9172	0.788	0.6385
150	1	0.9061	0.7938	0.6302
153	1	0.9288	0.7925	0.632
149	1	0.9027	0.7933	0.6314
148	1	0.9234	0.7918	0.6225
150	1	0.9264	0.7924	0.6297
151	1	0.9126	0.7895	0.6389
148	1	0.919	0.7875	0.6419
154	1	0.9263	0.7893	0.6397
147	1	0.9192	0.7918	0.6225
151	1	0.9268	0.7903	0.6347
150	1	0.9167	0.7916	0.6311
150	1	0.9095	0.7915	0.6329
151	1	0.93	0.7868	0.6365
153	1	0.9061	0.7886	0.6427
155	1	0.9173	0.7966	0.624
152	1	0.9214	0.7957	0.6255
155	1	0.9078	0.7924	0.6396
148	1	0.9195	0.7899	0.6312
143	1	0.9142	0.7884	0.6306
154	1	0.916	0.7887	0.6454
152	1	0.9149	0.7919	0.6352
156	1	0.9276	0.7884	0.6405
149	1	0.9211	0.785	0.6402
160	1	0.9245	0.7918	0.642
151	1	0.9123	0.7928	0.6328
159	1	0.9078	0.7922	0.642
149	1	0.9022	0.7849	0.6471
149	1	0.9268	0.7888	0.6365
150	1	0.9197	0.7904	0.6348
147	1	0.9189	0.7924	0.6245
157	1	0.9242	0.7944	0.6258
154	1	0.9253	0.7934	0.6272
145	1	0.9148	0.79	0.6335
148	1	0.9207	0.7871	0.6335
145	1	0.9147	0.7928	0.6255
154	1	0.9326	0.79	0.6373
152	1	0.9144	0.7934	0.6322

153	1	0.9098	0.7895	0.6415
151	1	0.9077	0.7912	0.6394
150	1	0.9255	0.7904	0.6364
151	1	0.9089	0.7915	0.6348
150	1	0.9116	0.7935	0.6292
151	1	0.9262	0.7905	0.6337
154	1	0.9087	0.7906	0.6435
153	1	0.9195	0.7894	0.6404
146	1	0.9153	0.7923	0.6275
150	1	0.929	0.79	0.6329
150	1	0.9194	0.7938	0.623
150	1	0.919	0.7935	0.6255
147	1	0.9226	0.7917	0.6291
150	1	0.9269	0.791	0.6314
146	1	0.9129	0.792	0.63
143	1	0.9078	0.7888	0.6332
147	1	0.9083	0.7888	0.638
144	1	0.9128	0.7897	0.6356
153	1	0.9003	0.7919	0.6399
154	1	0.9225	0.7919	0.642
158	1	0.9238	0.7954	0.6324
149	1	0.9122	0.7933	0.6324
146	1	0.9097	0.7923	0.6323
150	1	0.9223	0.7932	0.6302
154	1	0.927	0.7914	0.6374
150	1	0.9187	0.7879	0.6409
149	1	0.9168	0.7859	0.6431
143	1	0.9123	0.7925	0.6223
149	1	0.9135	0.7908	0.6384
150	1	0.9352	0.7905	0.6241
158	1	0.9221	0.7937	0.6379
152	1	0.9171	0.7912	0.6356
151	1	0.8987	0.7918	0.6352
154	1	0.919	0.794	0.6338
155	1	0.9144	0.7929	0.6385
147	1	0.9078	0.7918	0.6334
155	1	0.9212	0.7929	0.6347
148	1	0.9135	0.7941	0.6217
146	1	0.9141	0.7905	0.6356
154	1	0.9096	0.7958	0.631
151	1	0.9159	0.7928	0.6306
151	1	0.9147	0.7952	0.6262
149	1	0.9166	0.7939	0.6212
148	1	0.9185	0.7926	0.629
146	1	0.923	0.7896	0.6299
151	1	0.9228	0.788	0.6447
149	1	0.9228	0.7905	0.6315
152	1	0.9257	0.7925	0.6291
155	1	0.9222	0.7874	0.6461
145	1	0.9091	0.7908	0.633
152	1	0.918	0.7912	0.6374
145	1	0.9076	0.7879	0.6356
151	1	0.9289	0.7917	0.6246
152	1	0.9216	0.7915	0.6333
149	1	0.9147	0.7901	0.6362
157	1	0.9321	0.7935	0.6309
154	1	0.9103	0.792	0.6364
157	1	0.9099	0.7874	0.6473
155	1	0.9107	0.7892	0.6427
149	1	0.9215	0.7936	0.6225
147	1	0.9182	0.7913	0.6339
152	1	0.9082	0.7914	0.6356
150	1	0.906	0.7907	0.6388
153	1	0.899	0.7927	0.6352
156	1	0.9077	0.7869	0.6474
150	1	0.9093	0.7896	0.6399

148	1	0.9079	0.7891	0.6385
153	1	0.9254	0.7875	0.6423
148	1	0.923	0.7916	0.6267
152	1	0.9209	0.7897	0.6393
152	1	0.925	0.7908	0.6379
149	1	0.9257	0.7881	0.6379
155	1	0.9194	0.7919	0.6371
146	1	0.9172	0.7909	0.6256
147	1	0.9102	0.7922	0.6306
152	1	0.9177	0.7946	0.6286
151	1	0.9255	0.7898	0.6367
155	1	0.9237	0.7957	0.6236
156	1	0.9002	0.7894	0.6441
155	1	0.9267	0.7928	0.6333
151	1	0.9281	0.7891	0.6374

Tabla 4.6: Mejor combinación encontrada para NSGA-III

En este caso concreto, se ha utilizado una configuración con tamaño de población intermedio y un número de generaciones relativamente bajo. Las probabilidades de mutación han sido también las más reducidas, al igual que los valores del tamaño del torneo y del número de particiones. Con estas condiciones, el número total de individuos evaluados por generación es de 600, de los cuales únicamente un 27 % terminan formando parte del frente de Pareto final. Al analizar los individuos que conforman ese frente, se observan buenos valores en varios de los objetivos definidos, comparables incluso a los obtenidos con el algoritmo genético clásico. No obstante, el valor más alto alcanzado para el objetivo de equilibrio ha sido 0.9352, lo cuál sigue reflejando un grado de desequilibrio considerable en la distribución de alumnos, lo que significa que en algunas asignaturas, se está comprometiendo la viabilidad práctica de las soluciones encontradas.

Este resultado pone de manifiesto una limitación inherente a los algoritmos genéticos multiobjetivo como NSGA-III, al utilizar una estrategia basada en la no dominancia para la comparación entre individuos, todos los objetivos se consideran con la misma importancia relativa. Esto significa que, a menos que un individuo sea estrictamente dominado por otro en todos los objetivos, puede formar parte del frente de Pareto aunque presente deficiencias significativas en alguno de ellos. En consecuencia, el algoritmo favorece la diversidad y la representación equilibrada de soluciones en distintas zonas del espacio objetivo, pero puede perder eficacia a la hora de optimizar un único objetivo crítico de forma prioritaria.

En contraste, en el enfoque del algoritmo genético clásico se ha empleado una función fitness agregada, en la cuál se asignan pesos a cada objetivo en función de su importancia relativa. Esta ponderación permite guiar más eficazmente la búsqueda hacia soluciones que prioricen explícitamente aquellos objetivos considerados más relevantes, como es el caso del equilibrio entre grupos. Por ello, el algoritmo clásico ha demostrado una mayor capacidad para generar soluciones factibles desde el punto de vista organizativo, al penalizar de forma más intensa aquellas configuraciones con desequilibrios excesivos.

5. Desarrollo de la aplicación web

5.1. Introducción

Una vez completado el desarrollo de los dos algoritmos genéticos y tras realizar un análisis exhaustivo de sus hiperparámetros con el objetivo de identificar las configuraciones más eficientes, se ha considerado oportuno implementar una aplicación web que permita utilizar estos algoritmos de manera sencilla y accesible.

Esta aplicación no ha tenido un estudio previo de requisitos ni una necesidad funcional concreta, sino que se plantea cómo una herramienta práctica que facilite la ejecución y visualización de los resultados obtenidos por ambos algoritmos. De este modo, se busca acercar la solución al usuario final, eliminando la necesidad de interactuar directamente con el código o scripts de experimentación.

5.2. Diseño de la aplicación

El objetivo principal de esta aplicación web es proporcionar a los usuarios una interfaz sencilla e intuitiva que les permita ejecutar los algoritmos genéticos desarrollados. A través de esta interfaz, el usuario podrá seleccionar qué algoritmo desea ejecutar, adjuntar los ficheros de entrada necesarios y definir la configuración de los hiperparámetros de forma personalizada.

Además de la configuración manual, también se ofrece la posibilidad de utilizar la combinación de hiperparámetros óptima que ha sido determinada previamente en el capítulo de experimentación. Esto permite al usuario beneficiarse directamente del análisis realizado sin necesidad de profundizar en la configuración técnica.

Una vez realizada la ejecución del algoritmo, la aplicación mostrará la evolución obtenida a lo largo de las generaciones, permitiendo consultar los valores alcanzados por los objetivos en las mejores soluciones. Asimismo, se indicará la ruta en la que se han generado los ficheros de salida, los cuales incluyen las configuraciones de matrículas correspondientes a las soluciones finales.

5.3. Flask

Dado que la implementación de ambos algoritmos ha sido realizada íntegramente en *Python*, y considerando que la aplicación web se plantea cómo una capa de interacción directa con esta lógica, se ha optado por utilizar un framework web basado en el mismo lenguaje. Esta elección garantiza una integración fluida entre la interfaz y el núcleo computacional del sistema.

El framework seleccionado ha sido Flask, un microframework ligero y flexible que permite desarrollar aplicaciones web de forma sencilla y rápida. Flask ofrece una arquitectura minimalista que facilita la incorporación de rutas, plantillas HTML y gestión de formularios, sin necesidad de estructuras complejas. Gracias a su simplicidad y compatibilidad con bibliotecas de terceros, resulta especialmente adecuado para proyectos académicos o de carácter experimental como este, donde se prioriza la funcionalidad frente a una infraestructura compleja.

Para más información sobre Flask, se puede consultar la documentación oficial [Projects, 2025] o el artículo introductorio en GeeksforGeeks[for Geeks, 2025].

5.4. Arquitectura del proyecto

En esta sección se describe la estructura del proyecto y el funcionamiento de los distintos archivos que lo componen, los cuales pueden verse en la Figura 5.1. A continuación, se explica el propósito de cada uno de ellos:

- **app.py**: Archivo principal desde el cuál se desarrolla la aplicación web. Aquí se reciben las peticiones de los usuarios, se redirigen a la lógica correspondiente y se devuelve una respuesta adecuada a la interfaz.
- **Carpeta model**: Contiene los archivos relacionados con la implementación de los algoritmos genéticos. En particular:
 - `alumno.py`, `matricula.py` y `solucion.py`: Estos archivos definen la estructura de los individuos en la población. Un individuo (o solución) está compuesto por una lista de alumnos, y cada alumno tiene una lista de matrículas.
 - `genetico.py` y `nsga3.py`: Implementan la lógica completa de los algoritmos genéticos desarrollados: el algoritmo clásico y NSGA-III respectivamente.
- **Carpeta utils**: Contiene archivos de utilidades empleadas en el desarrollo de los algoritmos:
 - `export.py`: Incluye funciones relacionadas con la exportación de resultados y generación de ficheros de salida.
 - `utils.py`: Agrupa diversas funciones de ayuda utilizadas a lo largo del código.
- **Carpeta static**: Contiene los recursos estáticos de la aplicación web. En particular, el archivo `style.css` que define los estilos visuales utilizados en las distintas pantallas.

- **Carpeta templates:** Incluye todas las plantillas HTML que conforman la interfaz web. Aquí se encuentran las vistas que permiten al usuario subir los ficheros de entrada, seleccionar configuraciones y visualizar las soluciones obtenidas tras la ejecución de los algoritmos.

```
C:
├─ app.py
├─ model
│  ├─ alumno.py
│  ├─ genetico.py
│  ├─ matricula.py
│  ├─ nsga3.py
│  └─ solucion.py
├─ utils
│  ├─ export.py
│  └─ utils.py
├─ static
│  └─ style.css
└─ templates
   ├─ index.html
   ├─ loading.html
   ├─ result_genetico.html
   └─ result_nsga3.html
```

Figura 5.1: Estructura del proyecto

5.5. Aplicación web

Una vez explicada la estructura del proyecto, se procede a detallar el funcionamiento de la aplicación web desarrollada. El diseño de la interfaz gráfica es intencionadamente sencillo e intuitivo, con el objetivo de facilitar el uso de los algoritmos sin necesidad de conocimientos técnicos.

La pantalla principal presenta un formulario en el que el usuario puede interactuar con los siguientes elementos:

- **Carga de ficheros de entrada:** se proporcionan dos botones para subir los archivos necesarios para la ejecución del algoritmo, el fichero de horarios y la configuración inicial de matrículas, los cuales servirán como referencia.
- **Selección del algoritmo:** mediante un desplegable, el usuario puede elegir entre ejecutar el algoritmo genético clásico o el algoritmo NSGA-III.
- **Configuración del algoritmo:** el usuario debe seleccionar entre dos modos:
 - **Configuración óptima:** aplica automáticamente la combinación de hiperparámetros que ha sido determinada como la mejor durante la fase de experimen-

tación. Al seleccionar esta opción, todos los campos relacionados con los hiperparámetros quedan deshabilitados, mostrando sus valores fijados únicamente a modo informativo.

- **Configuración manual:** permite al usuario definir libremente los valores de los hiperparámetros. Los parámetros categóricos como los métodos de selección, cruce y sustitución se presentan mediante menús desplegables. En el caso del algoritmo NSGA-III, dado que sólo dispone de un método de selección y sustitución, estos campos aparecen deshabilitados con el valor correspondiente preestablecido. Los hiperparámetros numéricos deben ser introducidos directamente por el usuario.

■ **Campos condicionales:** algunos campos sólo se mostrarán cuando sean relevantes:

- *Tamaño del torneo:* sólo aparecerá si el método de selección elegido es por torneo.
- *Número de particiones:* únicamente visible si se selecciona el algoritmo NSGA-III, ya que este parámetro es necesario para la generación de puntos de referencia.

■ **Ejecución del algoritmo:** una vez completado el formulario, el usuario puede iniciar el proceso pulsando el botón *Ejecutar*. El sistema comenzará la ejecución del algoritmo con los parámetros proporcionados y mostrará el resultado al finalizar.

Antes de iniciar la ejecución, se lleva a cabo una validación completa de los campos del formulario. Se comprueba que los archivos hayan sido cargados correctamente y que todos los hiperparámetros tengan valores válidos y dentro del formato requerido (por ejemplo, que sean números positivos o que no falten valores). En caso de que se detecte algún error de validación, se mostrará un mensaje informativo al usuario indicando claramente qué campo está generando el problema y qué se espera de él.

Además, si durante la ejecución del algoritmo ocurre un error inesperado (por ejemplo, por errores en los datos de entrada o excepciones internas), el usuario será redirigido automáticamente a la pantalla de inicio. En ese caso, se mostrará un mensaje genérico indicando que ha ocurrido un error durante el proceso. Para facilitar el diagnóstico, en la consola desde la que se ejecutó el servidor se mostrará información técnica más detallada sobre el error.

Algoritmo Genético - TFG

Archivo de horarios:
 No se ha seleccionado ningún archivo.

Archivo de matrículas:
 No se ha seleccionado ningún archivo.

Algoritmo:
NSGA-III

Configuración:
Manual

Selección:
Torneo

Tamaño Torneo:

Cruce:
Uniforme

Sustitución:
Reemplazo

Particiones:

Población:

Generaciones:

Mutación Elemento:

Mutación Teoría:

Mutación Práctica:

Figura 5.2: Pagina principal de la aplicación web

Tras pulsar el botón de *Ejecutar*, se inicia la ejecución del algoritmo genético seleccionado con la configuración indicada. La aplicación no muestra una vista en tiempo real de la evolución del algoritmo, en su lugar, se presentará una pantalla de carga mientras se ejecuta el proceso.

Dado que el tiempo de ejecución puede variar significativamente en función de la combinación de hiperparámetros elegida, esta pantalla de carga incluirá un botón que permitirá cancelar la ejecución en cualquier momento. En caso de cancelación, el usuario será redirigido nuevamente a la pantalla de inicio para realizar una nueva configuración o cargar otros datos.

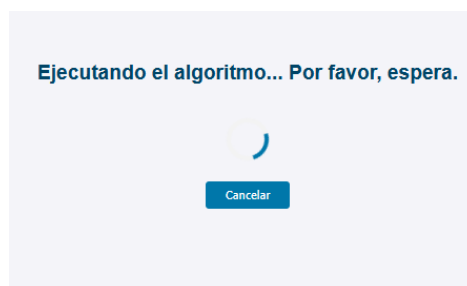


Figura 5.3: Pantalla de carga de la aplicación web

Una vez finaliza el proceso de ejecución, la aplicación carga una nueva pantalla en la que se muestran los resultados obtenidos por el algoritmo genético. Esta pantalla tiene como objetivo proporcionar al usuario una visión general de la evolución del algoritmo, los detalles de la mejor solución encontrada y la localización de los ficheros generados, que contienen la configuración final de matrículas.

La interfaz que se presenta dependerá del tipo de algoritmo ejecutado. En el caso del algoritmo genético clásico, la estructura de la pantalla será la siguiente:

- **Hiperparámetros:** Sección en la que se muestra la configuración completa de hiperparámetros utilizada en la ejecución.
- **Ruta de ficheros de salida:** Se indica la ruta donde se han almacenado los ficheros de salida generados, los cuales contienen tanto la configuración de matrículas como la distribución de alumnos (véase la sección 3.2.8).
- **Mejor solución:** Se presenta información detallada del mejor individuo encontrado durante la ejecución. Se indica en qué generación fue hallado, así como los valores alcanzados en cada uno de los objetivos.
- **Evolución del algoritmo:** Se muestra una tabla resumen con estadísticas por cada generación. Las columnas incluyen:
 - Mejor fitness
 - Fitness promedio
 - Peor fitness
 - Desviación estándar
 - Tiempo de ejecución por generación
- **Navegación:** Al final de la pantalla se incluye un botón que permite regresar a la pantalla principal para iniciar una nueva ejecución si se desea.



Figura 5.4: Parte superior de la pantalla de resultados del algoritmo clásico

Resumen por generación

Generación	Máx	Media	Min	Dev. Estándar	Tiempo (s)
1	0.6827 (294 1.0000 0.9182 0.7098 0.4598)	0.6592	0.5991 (288 1.0000 0.7630 0.7080 0.4644)	0.0108	5.65 seg
2	0.6925 (263 1.0000 0.9224 0.7039 0.4701)	0.6667	0.6324 (287 1.0000 0.8314 0.7060 0.4359)	0.0085	7.30 seg

[Volver al inicio](#)

Figura 5.5: Parte inferior de la pantalla de resultados del algoritmo clásico

En el caso del algoritmo NSGA-III, la pantalla de resultados mantiene una estructura muy similar a la del algoritmo genético clásico, aunque adaptada a las particularidades de un enfoque multiobjetivo. La estructura que se presenta es la siguiente:

- **Hiperparámetros:** Sección en la que se muestra la configuración completa de hiperparámetros utilizada durante la ejecución, tanto los fijados por el usuario cómo los establecidos automáticamente en el caso de haber seleccionado la configuración óptima.
- **Ruta de ficheros de salida:** Se indica la ruta del directorio donde se han almacenado los resultados. A diferencia del algoritmo clásico, en este caso se genera una carpeta

por cada individuo perteneciente al frente de Pareto final. Dentro de cada una de estas carpetas se encuentran los dos ficheros de salida habituales, que contienen la configuración de matrículas correspondiente a esa solución y su distribución de alumnos (véase la sección 3.3.9).

- **Frente de Pareto:** Se presenta una tabla con los individuos que forman parte del frente de Pareto obtenido en la última generación. Para cada uno de ellos se muestran los valores alcanzados en cada uno de los objetivos. Dado que no existe un único mejor individuo, esta sección permite comparar las distintas soluciones no dominadas que ha encontrado el algoritmo.
- **Evolución del algoritmo:** Se ofrece una tabla resumen que muestra cómo ha evolucionado el frente de Pareto a lo largo de las generaciones. Las columnas de la tabla incluyen:
 - Número de individuos en el primer frente
 - Tiempo de ejecución de la generación
- **Navegación:** Al final de la pantalla se incluye un botón que permite regresar a la pantalla principal para realizar una nueva ejecución o modificar la configuración actual.



Figura 5.6: Parte superior de la pantalla de resultados NSGA-III

Frente de Pareto

Estas son las soluciones no dominadas del frente final:

#	Solapes	Cohesión teoría	Equilibrio grupos	Cohesión prácticas	Preferencias
1	270	1.0000	0.8796	0.7077	0.4847
2	279	1.0000	0.8812	0.6987	0.4810
3	290	1.0000	0.8671	0.7117	0.4442
4	254	1.0000	0.8640	0.7123	0.4684

Resumen por generación

Generación	Tamaño del frente	Tiempo
1	4	0.47 seg
2	4	2.10 seg

[Volver al inicio](#)

Figura 5.7: Parte inferior de la pantalla de resultados NSGA-III

6. Conclusiones

6.1. Conclusiones

A lo largo de este trabajo se ha abordado la problemática asociada al proceso de matriculación de los estudiantes, centrada principalmente en la dificultad de seleccionar grupos compatibles y equilibrados. Tras un análisis detallado del problema, se ha revisado la teoría relacionada con los algoritmos genéticos clásicos y los algoritmos evolutivos multiobjetivo, identificando sus fundamentos y características principales.

Se ha representado formalmente el problema dentro del marco de la computación evolutiva, desarrollando dos implementaciones, una basada en un algoritmo genético clásico y otra en el algoritmo multiobjetivo NSGA-III. Para evaluar el comportamiento y rendimiento de ambas propuestas, se ha diseñado una fase de experimentación compuesta por múltiples combinaciones de métodos y valores de hiperparámetros. Esto ha permitido identificar configuraciones óptimas o prometedoras.

En el caso del algoritmo genético clásico, se ha logrado encontrar una configuración que ofrece buenos resultados, generando soluciones que pueden considerarse factibles en un entorno real. Por el contrario, aunque el algoritmo NSGA-III ha sido capaz de producir soluciones que cumplen satisfactoriamente la mayoría de los objetivos planteados, ha presentado dificultades en el cumplimiento del objetivo relativo al equilibrio de alumnos entre grupos. Este aspecto es especialmente crítico, dado que las limitaciones físicas de las aulas requieren una distribución de estudiantes lo más equitativa posible. Esta diferencia en el comportamiento se explica porque en el algoritmo clásico se ha utilizado una función de fitness ponderada, lo que ha permitido otorgar mayor importancia a dicho objetivo, mientras que en NSGA-III no se realiza dicha ponderación explícita.

Además de los algoritmos, se ha desarrollado una aplicación web que proporciona una interfaz gráfica sencilla e intuitiva que permite al usuario ejecutar los algoritmos implementados sin necesidad de conocimientos técnicos. Gracias a esta herramienta, es posible cargar los datos de entrada, seleccionar el tipo de algoritmo, configurar los hiperparámetros y ejecutar el proceso completo desde un entorno accesible. La aplicación, tras la ejecución, muestra un resumen visual del comportamiento del algoritmo y proporciona acceso a los ficheros generados con las soluciones.

6.2. Trabajo futuro

Como posibles líneas de trabajo futuro, una primera mejora consistiría en refinar la definición formal del problema incorporando restricciones adicionales que reflejen con mayor precisión las condiciones reales del proceso de matriculación. Esta ampliación permitiría realizar nuevas pruebas más específicas y profundizar en el comportamiento del algoritmo *NSGA-III*, con el objetivo de obtener soluciones que no sólo optimicen múltiples objetivos, sino que también sean factibles y aplicables en un entorno real.

También, sería interesante explorar otros algoritmos genéticos multiobjetivo más allá de *NSGA-III*. Aunque este ha sido elegido por su adecuación a problemas con más de dos objetivos, existen alternativas como *SPEA2*, *MOEA/D* o *NSGA-II* que podrían ofrecer diferentes equilibrios entre diversidad y convergencia, y cuya evaluación podría resultar beneficiosa para mejorar la calidad de las soluciones obtenidas.

Respecto a la aplicación web desarrollada, se identifican varias mejoras que podrían implementarse para ampliar su funcionalidad y mejorar la experiencia del usuario:

- **Implementación como servicio web accesible online:** Actualmente, la aplicación se ejecuta de forma local. Desplegarla en un servidor permitiría a múltiples usuarios acceder desde cualquier dispositivo sin necesidad de instalación previa.
- **Visualización directa de la configuración de matrículas:** Una posible mejora sería mostrar directamente en la interfaz web los resultados de la asignación de matrículas, sin necesidad de acceder al fichero generado en el sistema de archivos. Esto facilitaría el análisis inmediato de la solución obtenida.
- **Mejora del sistema de seguimiento en tiempo real:** Aunque actualmente se muestra una pantalla de carga, una posible extensión sería permitir la visualización del progreso del algoritmo en tiempo real, incluyendo métricas y evolución de los objetivos generación a generación.
- **Validación avanzada de entradas:** Incluir mecanismos para verificar la validez y consistencia de los ficheros de entrada antes de la ejecución del algoritmo ayudaría a evitar errores en tiempo de ejecución.

Estas mejoras permitirían evolucionar la herramienta desde un prototipo funcional hacia una solución más robusta, accesible y completa, integrando más funcionalidades y ofreciendo mayor usabilidad.

A. Anexos

A.1. Repositorio del código y scripts

Todo el código fuente desarrollado para este trabajo, incluidos los algoritmos genéticos, los scripts de experimentación, los módulos de apoyo y la aplicación web, se encuentra disponible en el siguiente repositorio de GitHub:

`https://github.com/Mega57/TFG`

El repositorio está organizado de forma estructurada y contiene:

- Implementaciones completas del algoritmo genético clásico y del algoritmo NSGA-III.
- Scripts utilizados para la experimentación y análisis de hiperparámetros.
- Código de la aplicación web desarrollada con Flask.
- Ficheros de entrada y salida utilizados durante el proceso de evaluación.
- Documentación adicional en el archivo `README.md`, que explica cómo ejecutar los distintos componentes del proyecto.

Cualquier referencia en esta memoria a scripts concretos o fragmentos de código puede consultarse directamente en el repositorio.

Referencia bibliográfica

- [Baker, 1985] Baker, J. E. (1985). Adaptive selection methods for genetic algorithms. *Proceedings of the 1st International Conference on Genetic Algorithms*, pages 101–111.
- [Blank and Deb, 2020] Blank, J. and Deb, K. (2020). Pymoo: Multi-objective optimization in python.
- [Darwin, 1859] Darwin, C. (1859). *On the Origin of Species by Means of Natural Selection*. John Murray, London.
- [Das and Dennis, 1998] Das, I. and Dennis, J. E. (1998). Normal-boundary intersection: A new method for generating the pareto surface in nonlinear multicriteria optimization problems. *SIAM Journal on Optimization*, 8(3):631–657.
- [Deb, 2001] Deb, K. (2001). *Multi-Objective Optimization using Evolutionary Algorithms*. John Wiley & Sons.
- [Deb et al., 2002] Deb, K., Agrawal, S., Pratap, A., and Meyarivan, T. (2002). A fast and elitist multiobjective genetic algorithm: Nsga-ii. *IEEE Transactions on Evolutionary Computation*, 6(2):182–197.
- [Deb and Jain, 2014] Deb, K. and Jain, H. (2014). An evolutionary many-objective optimization algorithm using reference-point-based nondominated sorting approach, part i: Solving problems with box constraints. *IEEE Transactions on Evolutionary Computation*, 18(4):577–601.
- [for Geeks, 2025] for Geeks, G. (2025). Introduction to flask. Accessed: 2025-07-06.
- [Fortin et al., 2012] Fortin, F.-A., De Rainville, F.-M., Gardner, M.-A., Parizeau, M., and Gagné, C. (2012). Deap: Evolutionary algorithms made easy. *Journal of Machine Learning Research*, 13(Jul):2171–2175.
- [Goldberg, 1989] Goldberg, D. E. (1989). *Genetic Algorithms in Search, Optimization, and Machine Learning*. Addison-Wesley.

-
- [Holland, 1975] Holland, J. H. (1975). *Adaptation in Natural and Artificial Systems*. University of Michigan Press.
- [Miller and Goldberg, 1995] Miller, B. L. and Goldberg, D. E. (1995). Genetic algorithms, tournament selection, and the effects of noise. *Complex Systems*, 9(3):193–212.
- [of Education, 1999] of Education, E. M. (1999). The bologna process: Setting up the european higher education area. Accessed: 2025-07-06.
- [Pareto, 1971] Pareto, V. (1971). *Manuale di Economia Politica*. UTET.
- [Projects, 2025] Projects, P. (2025). Flask documentation. Accessed: 2025-07-06.
- [UCLM, 2024a] UCLM (2024a). Campus virtual de la uclm. Accessed: 2025-07-06.
- [UCLM, 2024b] UCLM (2024b). Portal web de la escuela superior de informática. Accessed: 2025-07-06.
- [UCLM, 2024c] UCLM (2024c). Secretaría virtual de la uclm. Accessed: 2025-07-06.
- [Zhang and Li, 2007] Zhang, Q. and Li, H. (2007). Moea/d: A multiobjective evolutionary algorithm based on decomposition. *IEEE Transactions on Evolutionary Computation*, 11(6):712–731.
- [Zitzler et al., 2001] Zitzler, E., Laumanns, M., and Thiele, L. (2001). Spea2: Improving the strength pareto evolutionary algorithm. *TIK-Report*, 103.