



Evidencia 1. Proyecto Integrador

A00573182 Barush Caliel Copado Luna¹ and A01741659 Pedro Manuel Montes Valle²

¹ *Tecnológico de Monterrey, Campus Guadalajara*

Abstract— Este trabajo presenta la reproducción y el análisis del proceso de optimización multiobjetivo del grado de desacetilación del quitosano, basado en el estudio de Dinculescu, D.D., et al. (2024). Se implementó el algoritmo metaheurístico NSGA-II en Python utilizando la librería pymoo para optimizar dos variables de respuesta: el grado de desacetilación (DD) y la masa molar (MM). Se utilizaron los modelos de regresión obtenidos a partir del diseño experimental como funciones objetivo, y los frentes de Pareto obtenidos se compararon con los reportados en el artículo original. Se aplicaron métricas de desempeño como Hypervolume e IGD para evaluar la precisión y la convergencia de los resultados. La implementación logró reproducir las principales tendencias de optimización reportadas en la literatura, validando el uso de NSGA-II como una alternativa a la implementación de GA en MATLAB para la optimización multiobjetivo de procesos químicos.

Keywords— NSGA-II, optimización multiobjetivo, quitosano, desacetilación, pymoo, Hypervolume, IGD

I. INTRODUCCIÓN

En este reporte veremos la reproducción y análisis crítico del artículo "Simultaneous Optimization of Deacetylation Degree and Molar Mass of Chitosan from Shrimp Waste" (Polymers, 2024). El estudio original investiga la optimización del proceso de desacetilación del quitosano obtenido a partir de residuos de camarón con el objetivo de controlar dos propiedades clave del biopolímero; el grado de desacetilación (DD) y la masa molar media (MM). Estas dos propiedades determinan las aplicaciones del quitosano, que van desde formulaciones farmacéuticas hasta agentes materiales de embalaje.

La razón por la que seleccionamos este artículo es porque ya teníamos previo conocimiento acerca del tema, por lo que decidimos utilizar eso a nuestro favor, además de que nos pareció un artículo interesante.

Nuestro objetivo con esta reproducción es

- Implementar los modelos de respuesta (RSM) para MM y DD
- Utilizar una versión funcional de la optimización multiobjetivo usando algoritmos genéticos
- Comparar los frentes obtenidos con los reportados

a. Marco teórico

El algoritmo genético que utilizamos para la recreación es NSGA-II, que es una extensión de los algoritmos genéticos diseñada para problemas de optimización multiobjetivo. A diferencia de los métodos tradicionales que combinan todos los objetivos en una sola función, NSGA-II mantiene una población diversa de soluciones y busca aproximar el frente de Pareto. El algoritmo emplea mecanismos de selección basada en dominancia no dominada, crowding distance para mantener diversidad, y elitismo, lo que garantiza la preservación de las mejores soluciones encontradas durante las generaciones.

El objetivo o problema a optimizar son dos propiedades del quitosano:

- El grado de desacetilación (DD), que mide el porcentaje de grupos acetilo eliminados del quitosano.
- La masa molar (MM) relacionada con el peso molecular del polímero, esto influye en su viscosidad, etc.

La manera en que se aborda este problema es que, por un lado, tratamos de maximizar DD y minimizar MM, y por el otro buscamos maximizar ambos. Esto para poder producir quitosano con baja masa molar pero alta desacetilación, y otro caso en el que no se tenga que sacrificar su masa molar.

Para analizar y comparar los resultados utilizamos métricas de desempeño comunes en optimización multiobjetivo.

- IGD: mide que tan cerca está el frente obtenido del frente del paper.
- Hypervolume: Evalúa la convergencia y la diversidad del frente de Pareto. Una HV alta significa un mejor desempeño del algoritmo.
- Visualización del frente: A manera subjetiva, observar las gráficas de MM y DD nos sirve para comparar nuestra versión con la original.

b. Implementación

La implementación se hizo en Python debido a su gran cantidad de bibliotecas de optimización y porque fue la manera que lo abordamos en clase. Las principales librerías que se utilizaron fueron pandas, numpy, statsmodels, pymoo, matplotlib.

La manera en que se organizó el código fue en varias etapas claramente separadas:

1. Carga de los datos: se creó el dataframe con los valores de las tablas dados en el paper original.
2. Modelado estadístico: Se aplicó el algoritmo elegido que en nuestro caso fue NSGA-II. Se replicó el experimento.
3. Evaluación de resultados: Se calcularon métricas como IGD e Hypervolume y se compararon los resultados obtenidos con los de referencia del paper.

Las principales dificultades a las que nos enfrentamos recreando este artículo fueron que, primero que nada, no incluía un repositorio de código como tal, por lo que la realización del código fue totalmente manual. Además de que el artículo no incluía los datos numéricos exactos del frente, únicamente una imagen de ellos y, en general, faltaba especificación de los parámetros en el artículo.

II. SUPUESTOS Y ADAPTACIONES

Se realizaron 4 supuestos, con el fin de acercarse a los resultados y seguir una metodología apegada a lo largo de la replicación del artículo. Además, se hizo una pequeña adaptación como propuesta de mejora.

Al no haber especificaciones del algoritmo bioinspirado, librería o fuente de información, se supuso que, por las características mencionadas de algoritmos genéticos y por el resultado del frente de Pareto, que se trabajó con un NSGA-II. Dentro del artículo, se mencionó un tamaño de población de 100, siendo el único parámetro concreto. Fuera de este, se asumió que se trabajó con los parámetros por defecto: pop size = 100, SBX(eta=15, prob=0.9) y mutation: PM = PM(eta=20).

Los datos experimentales dentro del artículo no se replicaron por causas de acceso a laboratorio, costos de reactivos y más limitantes que no repercuten en la actividad de este trabajo. Por ello, se asumieron como reales, con buenas prácticas de laboratorio y capturados correctamente. Asimismo, se tomó como verdadero los rangos codificados de los valores reales de la concentración de NaOH y razón de líquido y sólido, y la duración de la extracción.

Al inicio de las pruebas estadísticas, en las pruebas ANOVA del modelo cuadrático, no se estableció un nivel de significancia para el trabajo. Por ello, a partir de cómo fue el uso de las pruebas estadísticas, se inferió que los investigadores utilizaron un p-value en el rango de 0.35 a 0.40. Este valor tan alto y flexibilidad la justificaron, y por ello, se trabajó igualmente en el presente trabajo.

Además, la cuarta suposición fue que una muestra pequeña de la frontera de Pareto realizada en el artículo es suficiente para probar métricas de comparación como el *Inverted Generational Distance* (IGD) y el *Hyperbolic Volume* (HV) y comparar variaciones del algoritmo implementado en este trabajo.

a. Variantes y mejoras

Se realizó una mejora adaptada a los dos procesos, de maximización de DD y minimización de MM así como la maximización de ambas variables. Para ello se importaron los métodos *PM* y *SBX* de la librería pymoo, relacionados con mutación y *crossover* respectivamente.

Las fronteras de Pareto, al verse muy parecidas gráficamente (única forma de comparar con la obtenida en el artículo), se hipotizó que en vez de aumentar la exploración, era mejor aumentar la explotación por parte del algoritmo NSGA-II. Es decir, las aproximaciones con los valores por defecto cualitativamente se observaron razonables; no obstante, podrían mejorar con ligeras variaciones.

Así, los parámetros que cambiaron fueron:

- pop size = 200
- crossover prob = SBX(eta = 30, prob = 0.9)
- mutation = PM(eta=60)

De este modo, se mantuvo un porcentaje alto de los padres que se iban a combinar en el *crossover* y una mutación muy pequeña (mientras más alto *eta* menor es la tasa de mutación). Asimismo, se aumentó el tamaño de la población a 200, pues en implementación del algoritmo con pop size = 100, a la hora de maximizar DD y minimizar MM, se obtuvieron una cantidad limitada de puntos en la frontera de Pareto. En sí, esta adaptación buscó mejorar la cantidad de soluciones óptimas y la proximidad al frente con la explotación.

III. RESULTADOS EXPERIMENTALES

Al proponerse, en el artículo estudiado, un modelo cuadrático del tipo:

$$y = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_{12}x_1x_2 + b_{13}x_1x_3 + b_{23}x_2x_3 + b_{11}x_1^2 + b_{22}x_2^2 + b_{33}x_3^2 \quad (1)$$

donde y se utiliza tanto para DD y para MM. Es decir, este modelo es el propuesto para capturar la complejidad y variabilidad de los dos objetivos que se quieren maximizar y minimizar.

a. Prueba ANOVA sobre los modelos cuadráticos DD y MM

Se aplicó una prueba ANOVA para determinar los coeficientes con mayor nivel de significancia. Estos fueron los resultados para los modelos de DD y de MM.

TABLE 1: RESULTADOS DE LA REGRESIÓN OLS PARA EL GRADO DE DESACETILACIÓN (DD)

Variable	Coef	Error Std	t	P> t	[0.025, 0.975]
const	90.3432	4.976	18.154	0.000	[78.576, 102.111]
x1	16.0728	2.338	6.874	0.000	[10.543, 21.602]
x2	-2.0665	2.338	-0.884	0.406	[-7.596, 3.463]
x3	5.6821	2.338	2.430	0.045	[0.153, 11.211]
x1x2	1.1187	3.054	0.366	0.725	[-6.103, 8.340]
x1x3	-3.3688	3.054	-1.103	0.306	[-10.590, 3.853]
x2x3	2.1312	3.054	0.698	0.508	[-5.090, 9.353]
x1_2	-10.6993	2.576	-4.153	0.004	[-16.791, -4.608]
x2_2	0.4207	2.576	0.163	0.875	[-5.671, 6.512]
x3_2	-3.4749	2.576	-1.349	0.219	[-9.566, 2.617]

TABLE 2: RESULTADOS DE LA REGRESIÓN OLS PARA LA MASA MOLAR (MM)

Variable	Coef	Error Std	t	P> t	[0.025, 0.975]
const	761.5597	61.784	12.326	0.000	[615.464, 907.655]
x1	229.7855	29.031	7.915	0.000	[161.138, 298.433]
x2	19.9224	29.031	0.686	0.515	[-48.725, 88.570]
x3	41.4807	29.031	1.429	0.196	[-27.167, 110.128]
x1x2	37.7313	37.914	0.995	0.353	[-51.922, 127.384]
x1x3	78.3112	37.914	2.065	0.078	[-11.342, 167.964]
x2x3	3.4162	37.914	0.090	0.931	[-86.237, 93.069]
x1_2	-146.0852	31.983	-4.568	0.003	[-221.713, -70.458]
x2_2	-103.1412	31.983	-3.225	0.015	[-178.769, -27.514]
x3_2	-202.8135	31.983	-6.341	0.000	[-278.441, -127.186]

TABLE 3: ESTADÍSTICOS GENERALES DE LOS MODELOS DD Y MM

Estadístico	DD	MM
R-cuadrado	0.915	0.945
R-cuadrado ajustado	0.806	0.873
F-estadístico	8.372	13.25
p-valor (F)	0.00526	0.00128
AIC	126.5	212.1
BIC	134.8	220.4
Número de observaciones	17	17
Grados de libertad del modelo	9	9
Grados de libertad del residuo	7	7

b. Gráficas de Superficie de Respuesta para $x_2 = 0$

Para respaldar los datos anteriores, se graficaron las superficies de respuesta. A continuación se ven las gráficas para x_1 y x_3 :

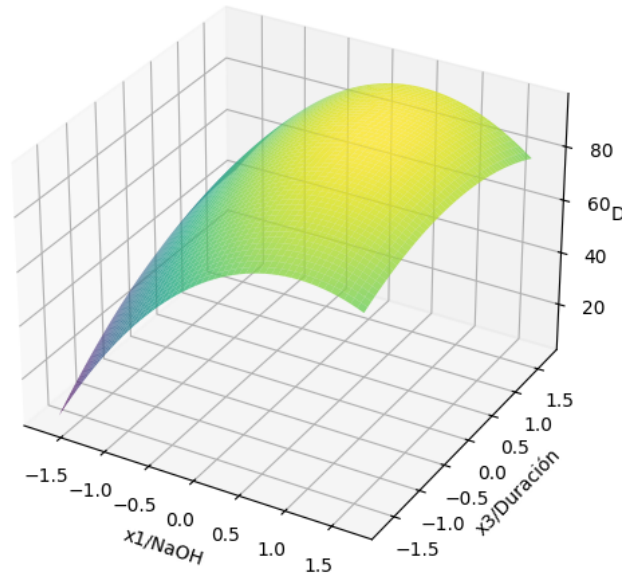


Fig. 1: Superficie de Respuesta DD ($x_2 = 0$)

El mismo procedimiento se efectuó para visualizar el comportamiento de optimización simple (no optimización multiobjetivo) para MM.

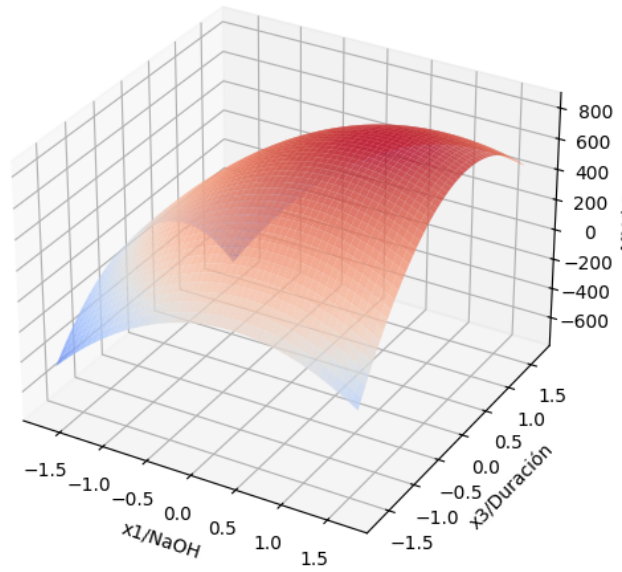


Fig. 2: Superficie de Respuesta MM ($x_2 = 0$)

c. Resultados de NSGA-II con parámetros por defecto

Como se mencionaron en los supuestos, se trabajó con un p-value muy alto a lo usual. Al revisar los valores, se tomaron con mayor flexibilidad el criterio del p-value y, dentro del artículo, se estipuló que se agregarían las variables y sus coeficientes con p-values no tan grandes (con un criterio relativo al rango de los resultados del ANOVA) a las funciones objetivo del algoritmo. De este modo, se utilizaron como funciones objetivo:

$$DD = 90.897 + 16.072x_1 - 2.066x_2 + 5.682x_3 - 3.369x_1x_3 - 10.823x_1^2 - 3.599x_3^2 \quad (2)$$

$$MM = 761.559 + 229.786x_1 + 41.480x_3 + 78.312x_1x_3 - 146.086x_1^2 - 103.142x_2^2 - 202.813x_3^2 \quad (3)$$

Aplicando el algoritmo NSGA-II con una población de 100 y parámetros por defecto, se obtuvieron las siguientes gráficas en los problemas de maximización y minimización:

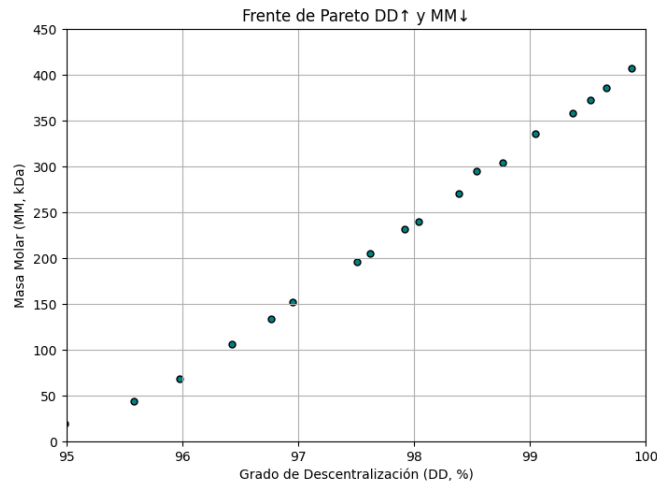


Fig. 3: Frente de Pareto DD↑ y MM↓

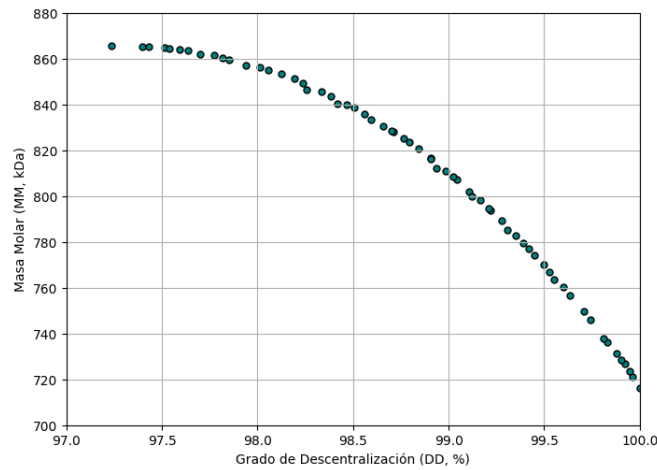


Fig. 4: Frontera de Pareto para DD↑ y MM↑

d. Algoritmo NSGA-II con Parámetros para Explotación

Para las modificaciones y adaptaciones, se utilizó la misma metodología, cambiando los parámetros y capturando los datos. Utilizando los datos presentados en el artículo (tomándolo como el frente de Pareto real), se comparó la muestra de la frontera de Pareto de los autores con la generada con el algoritmo de este trabajo. Se hizo uso del IGD y el HV, métricas relativas, para conocer si existe una mejora si se amplifica la explotación.

1. Resultados Maximizar DD y Minimizar MM

TABLE 4: RESULTADOS DE LA OPTIMIZACIÓN (MAXIMIZAR DD Y MINIMIZAR MM)

Configuración	IGD respecto al paper	Hypervolume
Parámetros por defecto	6.6325	179327.8979
Parámetros modificados	1.5766	179581.6402

Con estos datos, se traspusieron en una sola gráfica para visualizar cómo es el comportamiento de los dos algoritmos respecto a la muestra del frente de Pareto del artículo.

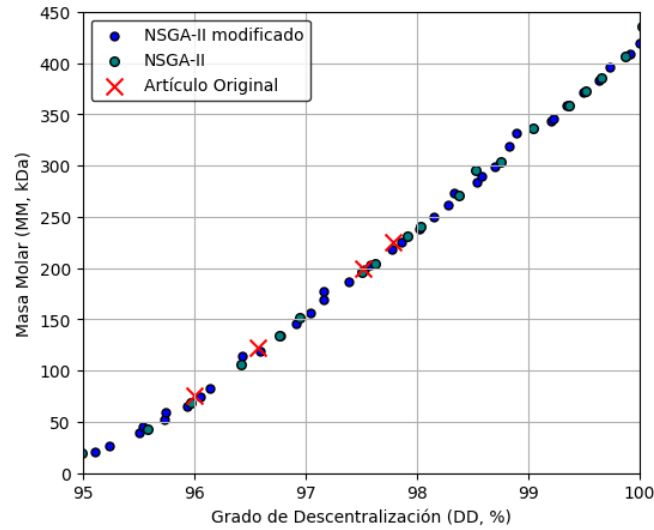


Fig. 5: Comparación del frente de Pareto (DD↑, MM↓)

2. Resultados Maximizar Ambos DD y MM

Mismo procedimiento para la obtención de los datos con IGD e HV.

TABLE 5: RESULTADOS DE LA OPTIMIZACIÓN (MAXIMIZAR DD Y MM)

Configuración	IGD respecto al paper	Hypervolume
Parámetros por defecto	0.1535	309776.9876
Parámetros modificados	0.1519	309784.5023

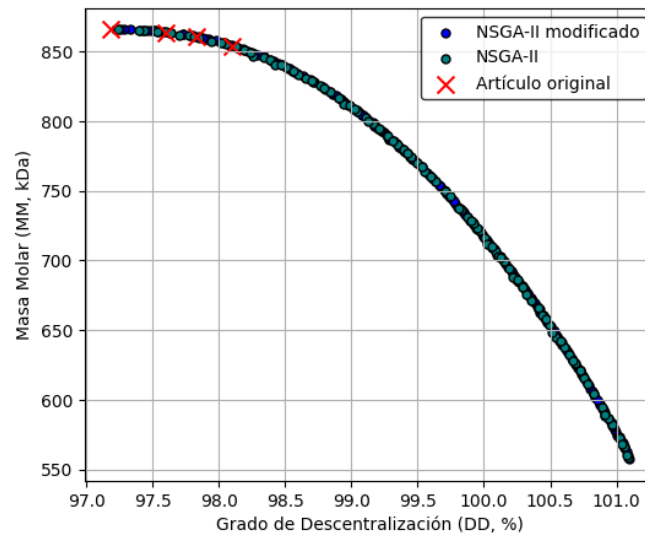


Fig. 6: Comparación del frente de Pareto (DD↑, MM↑)

IV. DISCUSIÓN

En las pruebas ANOVA que se realizaron en el artículo original, no se estableció un nivel de significancia. El criterio que utilizaron para descartar o no la hipótesis nula fue que el valor del p-value no fuera suficientemente alto con referencia a los demás. En este trabajo se comprobó con los mismos valores y se observó que el umbral estuvo dentro de un p-value $\in [0.35, 0.40]$ en promedio. Asimismo, la forma de corroborar que el modelo captaba de forma correcta la variabilidad de los datos fue por medio del criterio del R^2 , en donde los valores fueron $R^2_{DD} = 0.91$ y $R^2_{MM} = 0.95$, siendo valores muy altos y respaldados por una prueba de p-value de Fisher de 0.005 y 0.001 para DD y MM respectivamente.

En términos generales, si se logró reproducir las tendencias principales de optimización, los valores de DD y MM obtenidos mostraron una buena correspondencia, tanto visual como cuantitativa, con los resultados del paper. Al comparar los frentes de Pareto, se puede observar que las soluciones obtenidas presentan ligeras desviaciones respecto a la muestra del artículo. Esta discrepancia puede deberse a varios factores:

- El artículo no proporcionó los datos numéricos exactos del frente de Pareto, solo gráficas, por lo que se tuvo que trabajar con la muestra de soluciones óptima en el frente de Pareto del artículo.
- Las funciones objetivo se construyeron a partir de modelos ajustados a los datos experimentales disponibles; cualquier diferencia en los coeficientes puede alterar ligeramente los resultados. Concretamente, el software que se utilizó en el artículo original fue MATLAB, y en este trabajo se replicó con Python. Estos dos lenguajes, a la hora de manejar decimales y cálculos, pudieron haber hecho redondeos o estar configurados a un límite de decimales (cosa que pudo haber añadido ruido).
- Los parámetros del algoritmo NSGA-II no estaban explícitamente detallados, por lo que se asumieron valores por defecto.

Al realizar las adaptaciones y mejoras, se observó una mejor convergencia y diversidad en los frentes de Pareto, lo cual se vio reflejado en las métricas IGD e HV más cercanas a los valores del artículo. Esto reforzó la idea de que, para este problema, las funciones objetivo modelaban bien la complejidad de los datos y el planteamiento del algoritmo NSGA-II fue bueno para aproximarse correctamente al frente de Pareto. En sí, el aumentar la explotación dentro del algoritmo, con los parámetros de crossover y mutation, afectó ligera y positivamente a encontrar soluciones más acercadas al frente de Pareto. En general, los resultados originales del artículo parecen válidos y consistentes con los principios de optimización multiobjetivo.

Por último, el proceso de reproducción permitió aprender varios aspectos clave:

- La importancia de documentar todos los parámetros y datos experimentales para asegurar reproducibilidad.
- Cómo interpretar y aplicar modelos de regresión como funciones objetivo en optimización multiobjetivo.
- La utilidad de métricas como IGD e HV para cuantificar proximidad y diversidad en frentes de Pareto.
- Cómo los supuestos pueden afectar a los resultados.

V. CONCLUSIONES

La reproducción del artículo “Simultaneous Optimization of Deacetylation Degree and Molar Mass of Chitosan from Shrimp Waste” nos ayudó a confirmar la efectividad del algoritmo NSGA-II para abordar problemas de optimización multiobjetivo, específicamente la maximización del grado de desacetilación y la optimización de la masa molar de quitosano.

Los resultados obtenidos mostraron alta similitud con los resultados del estudio original, reproduciendo correctamente la forma del frente de Pareto y la relación inversa entre ambos objetivos. El comportamiento global del modelo confirma la validez de los resultados publicados en el artículo.

En cuanto a la reproducibilidad, nos dimos cuenta de que faltaba documentación detallada, lo cual representó una dificultad importante para replicar los resultados con más precisión. Esto resalta la importancia de compartir datos, código y configuraciones completas para fortalecer la transparencia y credibilidad de la investigación.

Como posibles trabajos futuros, podríamos extender aún más el estudio original aplicando otro tipo de metaheurísticas multiobjetivo, como MOEA/D, y comparar el desempeño entre algoritmos.

REFERENCES

- Dinculescu, D. D., Apostol, L., Ungureanu, E., & Ghiurea, M. (2024). *Simultaneous optimization of deacetylation degree and molar mass of chitosan from shrimp waste*. *Polymers*, 16(2), 170. <https://doi.org/10.3390/polym16020170>
- Matica, M. A., Aachmann, F. L., Tøndervik, A., Sletta, H., & Ostafe, V. (2019). *Chitosan as a Wound Dressing Starting Material: Antimicrobial Properties and Mode of Action*. *International Journal of Molecular Sciences*, 20(23), 5889. <https://doi.org/10.3390/ijms20235889>
- pymoo - NSGA-II: Non-dominated Sorting Genetic Algorithm. (s. f.). <https://pymoo.org/algorithms/moo/nsga2.html>
- Román-Doval, R., Torres-Arellanes, S. P., Tenorio-Barajas, A. Y., Gómez-Sánchez, A., & Valencia-Lazcano, A. A. (2023). *Chitosan: Properties and Its Application in Agriculture in Context of Molecular Weight*. *Polymers*, 15(13), 2867. <https://doi.org/10.3390/polym15132867>