

Estudios de inteligencia artificial para mejorar el desempeño de modelos estadísticos en codigestión anaerobia

Artificial intelligence studies to improve the performance of statistical models in anaerobic codigestion

Recepción del artículo: 20/06/2024 • Aceptación para publicación: 07/08/2024 • Publicación: 01/09/2024

● <https://doi.org/10.32870/e-cucba.vi23.359>

Juan Pablo Morales Rivera*

ORCID: <https://orcid.org/0000-0003-0682-5102>

Jean Michelle Flores Gómez

Francisco González Torres

Universidad de Guadalajara. Centro Universitario de Tonalá. División de Ingenierías e Innovación Tecnológica. Tonalá, Jalisco, México.

*Autor para correspondencia:

juan.morales9853@academicos.udg.mx

Resumen

Este trabajo busca aplicar inteligencia artificial para mejorar los resultados obtenidos por modelos basados en sistemas estadísticos como la metodología de superficie de respuesta (MSR) en la literatura, empleando técnicas computacionales basadas en IA, que han demostrado ser valiosas para el diseño y predicción del comportamiento de los sistemas de digestión anaerobios, se empleó una matriz de 3 por 3 para encontrar la mejor arquitectura para la red neuronal, logrando mejorar el rendimiento del modelo propuesto en la literatura que logra de R^2 de 0.90 y con esta metodología empleada en esta investigación se logra un ajuste R^2 de 0.99.

Palabras clave: Análisis estadístico, optimización, red neuronal artificial.

Abstract

This paper seeks to apply artificial intelligence to improve the results by models based on statistical systems such as response surface methodology (RMS) in literature, using AI -based computational techniques, which have proven to be valuable for the design and prediction of the behavior of anaerobic digestion systems, a matrix of 3 by 3 was used to find the best architecture for the neuronal network, improving the performance of the model proposed in the literature that achieves R^2 of 0.90 and with this methodology used in this research an R^2 adjustment of 0.99 is achieved.

Keywords: Statistical analysis, optimization, artificial neuronal network.

En el fenómeno de codigestión anaerobia se integran diversas mezclas de sustratos orgánicos biodegradables, el proceso de cogestión anaeróbica aplica una serie de diversos residuos orgánicos cómo lo son, residuos de origen industrial aguas residuales y residuos de origen de granjas mezcladas en un digestor anaerobio. Una vez estos diversos microorganismos empiezan a degradar toda la materia orgánica el proceso culmina con la obtención de biogás y sustratos orgánicos que pueden ser utilizados cómo abono, este tipo de tecnologías ofrece amplias ventajas para la utilización de residuos biológicos ya que sus costos operacionales son muy bajos y prácticamente no generan residuos, más sin embargo también presenta varios desafíos que se necesitan superar para lograr aumentar sus eficiencia y viabilidad por lo que es necesaria ampliar las investigaciones con el objetivo de maximizar sus rendimientos ambientales y económicos (Almomani *et al.*, 2020; Karki *et al.*, 2021; Chow *et al.*, 2020; Ma *et al.*, 2020) por kilogramo de mezcla degradada.

Un modelo matemático puede entenderse como una serie de relaciones matemáticas que intentan describir el comportamiento de las variables dentro de un sistema y que son obtenidas al estudiar a fondo las leyes físicas, por tanto, un modelo matemático puede llegar a ser tan sencillo como una sola relación que describe todo un sistema o por el contrario todo un grupo de funciones que se relacionan entre sí para simular el comportamiento global de un sistema (Hatata *et al.*, 2021; Mohamadou *et al.*, 2020; Zeb *et al.*, 2020).

Las redes neuronales artificiales son una técnica computacional que imita la habilidad biológica de procesamiento del cerebro humano; además, pueden manejar el modelado y la simulación de sistemas no lineales extremadamente complejos comprobado su validez para simular y modelar una amplia variedad de sistemas los cuales principalmente son sistemas altamente no lineales y por tanto muy complejos de modelar con técnicas basadas en ecuaciones diferenciales ordinarias, otras metodologías basados en análisis estadísticos como lo es la metodología de superficie de respuesta (Iweba *et al.*, 2021; Ahmad *et al.*, 2023).

Las redes neuronales artificiales están basadas en el comportamiento de los enlaces neuronales biológicos que hay dentro del cerebro humano. Estas neuronas incluyen varios parámetros de peso y funciones de activación coma así mismo están estructuradas en 3 capas: entrada, oculta y salida, para que la red neuronal artificial sea más efectiva depende en gran medida del volumen de datos y no es recomendable emplear bases de datos demasiado pequeñas ya que esto podría provocar falta de información haciendo que las predicciones sean muy inexactas.

Dentro de su funcionamiento es necesario fraccionar los datos en el proceso de modelado así como seleccionar el algoritmo de entrenamiento más apto para calibrar los parámetros del modelo (pesos de conexión), asegurando así que la red puede ajustarse a las diversas relaciones no lineales y complejas. Con el objetivo de determinar la arquitectura ideal para los modelos basados en redes neuronales artificiales es determinante definir el flujo de información que va a pasar por la red así como la magnitud de los pesos de conexión. Para esta investigación se utilizó una red neuronal tipo fit forward ya que se trata de una de las variantes más simples y que no requiere altos costos computacionales en este tipo de sistemas la base de datos fluye en una sola dirección primero por la capa de entrada, después por la capa oculta y termina en la capa de salida (Che *et al.*, 2020; Cruz *et al.*, 2022). En la literatura reciente siguen empleando de manera muy significativa los estudios basados en inteligencia artificial en este caso redes neuronales artificiales para modelar y comprender de mejor manera diversos sistemas basados en digestión anaerobia, para estudio de antibióticos nuevos, análisis de una diversidad de aguas residuales tanto industriales como locales (Mougari *et al.*, 2021; Haffiez *et al.*, 2022; Neta *et al.*, 2021; Zaied *et al.*, 2020; Abdelhay *et al.*, 2016).

Materiales y métodos

Este trabajo toma como base los resultados obtenidos en (Abdelhay *et al.*, 2016), que son optimización y modelado de producción de biogás a partir de la codigestión de residuos verdes/residuos biológicos utilizando lixiviados y lodos. Quienes realizaron una mezcla 60:40 de residuos orgánicos municipales y residuos orgánicos de poda (residuos verdes). Obtuvieron una mejora del 66% respecto a la mono digestión en obtención de biogás con una concentración al 78% de metano, dichos resultados obtenidos gracias a una mezcla al 37% de lixiviados y 9.5 g/L de sólidos volátiles. A partir de ese modelo realizaron una metodología de superficie de respuesta con un R2 de 0.90 y 0.72 para volumen de biogás y concentración de metano. En este trabajo se realizó una arquitectura de red neuronal empleando la herramienta de RNA, dentro del software Matlab R2021A (Math-Works Inc., Natick, MA, USA), como lo muestra la figura 1, usando un modelo experimental basado en una matriz de 3 por 3, con el objetivo de encontrar la mejor configuración para predecir el comportamiento de la codigestión anaerobia de residuos de poda y residuos orgánicos municipales. Dicha matriz consta de los entrenamientos: Levenberg-Marquardt (trainlm), Regularización Bayesiana (trainbr) y Scaled Conjugated Gradient (trainscg).

Además, se utilizaron 8, 12 y 15 neuronas en la capa oculta de la RNA y se utilizaron las configuraciones de 70/15/15, 80/10/10 y 90/5/5 para los datos de entrenamiento, validación y pruebas como lo muestra el Cuadro 1.

Cuadro 1. Diseño de matriz experimental.

RNA	Algoritmo de entrenamiento	Neuronas en capa oculta	Entrenamiento/ Validación/ Pruebas
1	trainlm	8	70/15/15
2	trainlm	8	80/10/10
3	trainlm	8	90/5/5
4	trainlm	12	70/15/15
5	trainlm	12	80/10/10
6	trainlm	12	90/5/5
7	trainlm	15	70/15/15
8	trainlm	15	80/10/10
9	trainlm	15	90/5/5
10	trainbr	8	70/15/15
11	trainbr	8	80/10/10
12	trainbr	8	90/5/5
13	trainbr	12	70/15/15
14	trainbr	12	80/10/10
15	trainbr	12	90/5/5
16	trainbr	15	70/15/15
17	trainbr	15	80/10/10
18	trainbr	15	90/5/5
19	trainscg	8	70/15/15
20	trainscg	8	80/10/10
21	trainscg	8	90/5/5
22	trainscg	12	70/15/15
23	trainscg	12	80/10/10
24	trainscg	12	90/5/5
25	trainscg	15	70/15/15
26	trainscg	15	80/10/10
27	trainscg	15	90/5/5

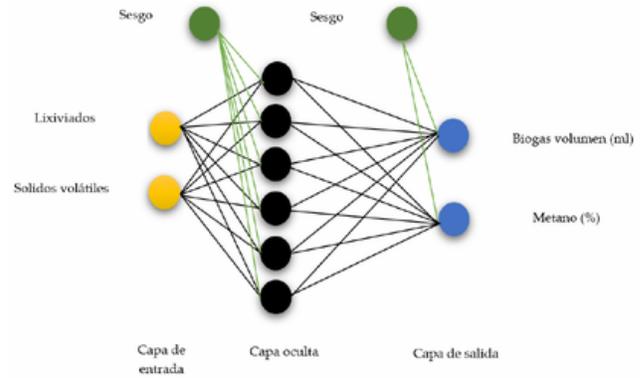


Figura 1. Arquitectura de la RNA.

Análisis estadístico

Para validar de manera completa los modelos RNA encontrados se tomaron ocho índices estadísticos los cuales nos permitieron diferir cuál arquitectura es la más idónea para optimizar el modelo de codigestión anaerobia según Cuadro 2.

Cuadro 2. Índices estadísticos.

No	Equations
1	$R = \frac{\sum_{i=1}^n (y_{p,i} - y_{p,ave})(y_{a,i} - y_{a,ave})}{\sqrt{[\sum_{i=1}^n (y_{p,i} - y_{p,ave})^2][\sum_{i=1}^n (y_{a,i} - y_{a,ave})^2]}}$
2	$R^2 = 1 - \frac{\sum_{i=1}^n (y_{a,i} - y_{p,i})^2}{\sum_{i=1}^n (y_{p,i} - y_{a,ave})^2}$
3	$Adjusted R^2 = 1 - \left[(1 - R^2) \times \frac{n-1}{n-k-1} \right]$
4	$MSE = \frac{1}{n} \sum_{i=1}^n (y_{p,i} - y_{a,i})^2$
5	$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_{p,i} - y_{a,i})^2}$
6	$MAE = \frac{1}{n} \sum_{i=1}^n (y_{a,i} - y_{p,i}) $
7	$SEP = \frac{RMSE}{y_{a,ave}} \times 100$
8	$AAD = \frac{1}{n} \left(\sum_{i=1}^n \left(\frac{ (y_{a,i} - y_{p,i}) }{y_{a,i}} \right) \right) \times 100$

Resultados

Con el objetivo de tener una RNA adecuada para esta investigación fue que se propuso trabajar en base a una matriz experimental de 33 para dar un total de 27 experimentos por triplicado, al ejecutar todas estas simulaciones se logró definir la mejor arquitectura, cuya configuración fue la que arrojó un error medio cuadrado menor de entre todas las 27 experimentaciones planteadas, es representada por la experimentación número 18. Para esta investigación los resultados de predicción en la obtención de biogás y concentración de metano, son presentados en la figura 2 que presenta el gráfico de las predicciones valores de salida contra los objetivos valores experimentales, cuyos resultados

encuentran valores de R de 1, entrenamiento de 1 y para todos los datos de 0.999, lo cual es un indicativo del excelente rendimiento de la red neuronal propuesta por este trabajo.

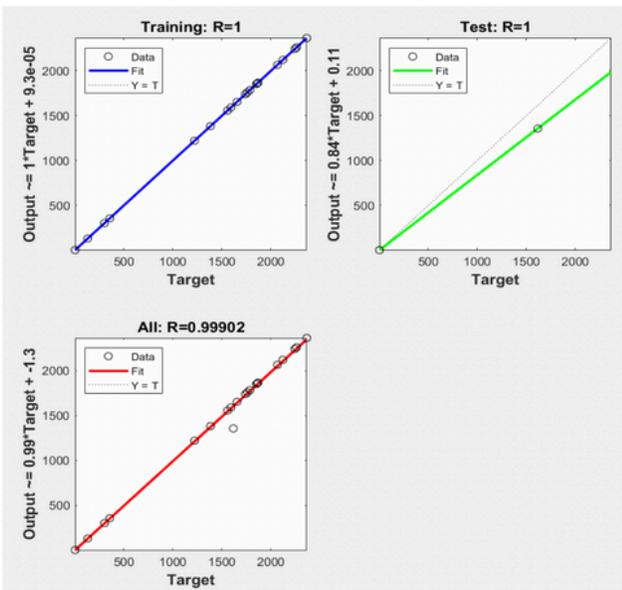


Figura 2. Gráficos de correlación RNA.

Resultados para el modelado de la producción de biogás

Se indica la Figura 3 el desempeño de la RNA para predecir con exactitud los valores experimentales en la generación de biogás son muy destacados y demuestran una superioridad al momento de emplear esta metodología contra la que se propuso que es MSR, indicando que aunque el modelo sea altamente no lineal es posible modelarlo de manera precisa.

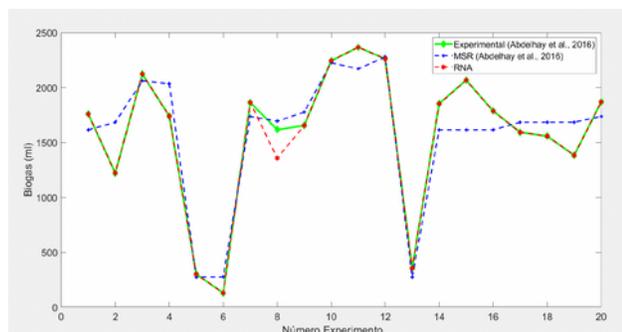


Figura 3. Gráfico comparativo de resultados Experimental, MSR de la literatura (Abdelhay *et al.*, 2016) y RNA (este trabajo) para producción de biogás (ml).

El cuadro 3 nos presenta las evaluaciones de los índices estadísticos tanto para el modelo MSR como para el modelo RNA del cual podemos obtener los resultados, un valor de r mayor cuando se utiliza la RNA que empleándola MSR lo que nos demuestra que el primer método de una mejor correlación entre los datos experimentales y los datos predichos.

Los valores R^2 confirman esta tendencia de un mejor rendimiento del modelo RNA así como el cálculo del ajustado R^2 el cual se obtiene para evitar la sobreestimación del coeficiente de correlación R^2 , demostrando así la significancia del modelo RNA. Para confirmar la cercanía de la línea ajustada a los puntos de datos se calcularon El EMC REMC y estos resultados al ser bastante pequeños demuestran una buena cercanía entre este conjunto de datos.

Cuadro 3. Índices estadísticos modelado de producción de biogás.

Índices Estadísticos	Biogás acumulado MSR (Abdelhay <i>et al.</i> , 2016)	Biogás acumulado RNA
Coefficiente de Correlación (R)	0.944	0.996
Coefficiente de Determinación (R^2)	0.891	0.992
Error Cuadrado Ajustado ($AdjR^2$)	0.885	0.991
Error Medio Cuadrado (MSE)	22.538	8.195
Raíz de Error Medio Cuadrado (RMSE)	4.747	2.86
Error Medio Absoluto (MAE)	64.478	12.958
Error Estándar de Predicción (SEP)	13.019	3.651
Desviación Media Absoluta (DMA)	9.476	1.580

Resultados para el modelado concentración de metano

La Figura 4 nos presenta la comparación de los resultados experimentales contra los obtenidos por los modelados MSR y RNA respectivamente logrando destacar un seguimiento casi exacto de la red neuronal en comparación con lo presentado por la otra metodología lo cual es de esperarse ya que la metodología MSR al estar basada en métodos estadísticos resulta tener varios inconvenientes a la hora de simular sistemas tan altamente no lineales como lo son los basados en sistemas biológicos como es en este caso con la co digestión anaerobia.

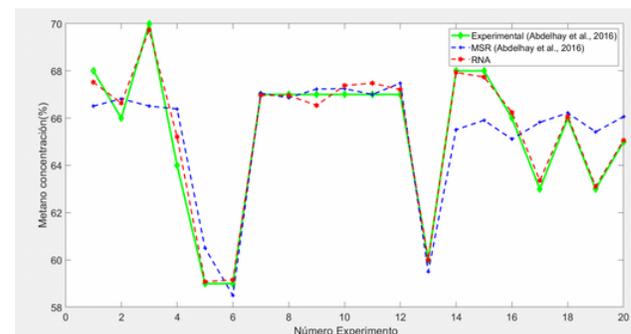


Figura 4. Gráfico comparativo de resultados Experimental, MSR de la literatura (Abdelhay *et al.*, 2016) y RNA (este trabajo) para concentración de metano (%).

Finalmente el Cuadro 4, nos indica el alcance de los índices estadísticos al evaluar los resultados de los modelos para simular la concentración de metano en el sistema, obteniéndose un índice R de 0.851 para MSR y de 0.962 para RNA al evaluar el coeficiente r cuadrado demuestra

esta tendencia de un mejor desempeño para un modelo basado en inteligencia artificial así mismo se convalidaron estos resultados al calcular el error cuadrado ajustado logrando obtener resultados muy superiores de la red neuronal contra los demostrados por los métodos estadísticos en este caso MSR, asimismo los errores medio cuadrado y la raíz de este demuestran que este modelo fue más preciso para simular los datos experimentales del sistema de digestión anaerobia cómo lo reafirman los demás índices estadísticos.

de 0.90 y 0.72 para biogás acumulado y concentración de metano. Mientras que con RNA se obtuvieron valores de R^2 de 0.991 y 0.956 respectivamente. En general, se espera que este trabajo sirva de aportación a futuros estudios que busquen desarrollar metodologías para la implementación de proyectos de codigestión, invitando a los ingenieros a realizar experimentación suficiente para minar los datos necesarios para entrenar una RNA.

Cuadro 4. Índices estadísticos modelado en concentración de metano.

Índices Estadísticos	Concentración de Metano por MSR	Concentración de Metano por RNA
Coefficiente de Correlación (R)	0.851	0.962
Coefficiente de Determinación (R^2)	0.725	0.963
Error Cuadrado Ajustado (Adj R^2)	0.709	0.956
Error Medio Cuadrado (MSE)	31.23	0.1825
Raíz de Error Medio Cuadrado (RMSE)	5.58	0.4271
Error Medio Absoluto (MAE)	9.143	0.3159
Error Estándar de Predicción (SEP)	37.43	11.91
Desviación Media Absoluta (DMA)	4.637	0.477

Conclusiones

Los modelos son herramientas que aportan soluciones en el diseño, operación, control y mejora de, en este caso, sistemas biológicos complejos altamente no lineales, como la codigestión anaerobia. El objetivo del trabajo fue aplicar IA para aumentar, de manera significativa, el rendimiento de los modelos que describen la producción de biogás y el contenido de metano, de una experimentación extraída de la literatura. Esto permitió descubrir que con la medición adecuada de algunas variables en el sistema se pueden generar datos suficientes para alimentar, entrenar y probar una RNA que modele adecuadamente éste fenómeno. A través del análisis estadístico se demostró la mejora y viabilidad de emplear IA para mejorar los modelos de este tipo de sistemas biológicos. El punto fuerte de este trabajo es que con una RNA se lograron buenos resultados para ambas variables de salida, mejorando incluso los resultados obtenidos a través de la MSR. El Modelo experimental de la matriz de 3 por 3 para obtener 27 configuraciones diferentes, para descubrir la arquitectura de la RNA adecuada fue el camino correcto para modelar codigestión anaerobia. Emplear el algoritmo de regularización bayesiana para evitar el sobre entrenamiento de la red, encontrar el número adecuado de neuronas en la capa oculta y distribuir los datos de distintas maneras para obtener la RNA que se acercara más a los valores experimentales. Finalmente, en este trabajo se concluye que, es posible mejorar los modelos empleando IA, a través de la literatura y empleando MSR se obtuvieron valores de R^2

Literatura citada

- Abdelhay, A., Albsoul, A., Hadidi, F. y Abuothman, A. (2016). Optimization and Modeling of Biogas Production From Green Waste/Biowaste Co-Digestion Using Leachate and Sludge. *CLEAN–Soil, Air, Water*, 44(11), 1557-1563. <https://doi.org/10.1002/clen.201500514>
- Ahmad, A., Yadav, A. K., Singh, A. y Singh, D. K. (2023). Optimisation of biogas yield from anaerobic co-digestion of dual waste for environmental sustainability: ANN, RSM and GA approach. *International Journal of Oil, Gas and Coal Technology*, 33(1), 75-101. <https://doi.org/10.1504/IJOGCT.2023.130377>
- Almomani, F. y Bhosale, R. R. (2020). Enhancing the production of biogas through anaerobic co-digestion of agricultural waste and chemical pre-treatments. *Chemosphere*, 255, 126805. <https://doi.org/10.1016/j.chemosphere.2020.126805>
- Chen, Y., Song, L., Liu, Y., Yang, L. y Li, D. (2020). A review of the artificial neural network models for water quality prediction. *Applied Sciences*, 10(17), 5776. <https://doi.org/10.3390/app10175776>
- Chow, W. L., Chong, S., Lim, J. W., Chan, Y. J., Chong, M. F., Tiong, T. J., ... y Pan, G. T. (2020). Anaerobic co-digestion of wastewater sludge: A review of potential co-substrates and operating factors for improved methane yield. *Processes*, 8(1), 39. <https://doi.org/10.3390/pr8010039>
- Cruz, I. A., Chuenchart, W., Long, F., Surendra, K. C., Andrade, L. R. S., Bilal, M., ... y Ferreira, L. F. R. (2022). Application of machine learning in anaerobic digestion: Perspectives and challenges. *Bioresource Technology*, 345, 126433. <https://doi.org/10.1016/j.biortech.2021.126433>
- Haffiez, N., Chung, T. H., Zakaria, B. S., Shahidi, M., Mezbahuddin, S., Maal-Bared, R. y Dhar, B. R. (2022). Exploration of machine learning algorithms for predicting the changes in abundance of antibiotic resistance genes in anaerobic digestion. *Science of The Total Environment*, 839, 156211. <https://doi.org/10.1016/j.scitotenv.2022.156211>
- Hatata, A., Galal, O. H., Said, N. y Ahmed, D. (2021). Prediction of biogas production from anaerobic co-digestion of waste activated sludge and wheat straw using two-dimensional mathematical models and an artificial neural network. *Renewable Energy*, 178, 226-240. <https://doi.org/10.1016/j.renene.2021.06.050>
- Iweka, S. C., Owuama, K. C., Chukwunke, J. L. y Falowo, O. A. (2021). Optimization of biogas yield from anaerobic co-digestion of corn-chaff and cow dung digestate: RSM and python approach. *Heliyon*, 7(11). [https://www.cell.com/heliyon/pdf/S2405-8440\(21\)02358-6.pdf](https://www.cell.com/heliyon/pdf/S2405-8440(21)02358-6.pdf)
- Karki, R., Chuenchart, W., Surendra, K. C., Shrestha, S., Raskin, L., Sung, S., ... y Khanal, S. K. (2021). Anaerobic co-digestion: Current status and perspectives. *Bioresource Technology*, 330, 125001. <https://doi.org/10.1016/j.biortech.2021.125001>
- Ma, G., Ndegwa, P., Harrison, J. H. y Chen, Y. (2020). Methane yields during anaerobic co-digestion of animal manure with other feedstocks: A meta-analysis. *Science of the Total Environment*, 728, 138224. <https://doi.org/10.1016/j.scitotenv.2020.138224>
- Mohamadou, Y., Halidou, A. y Kapen, P. T. (2020). A review of mathematical modeling, artificial intelligence and datasets used in the study, prediction and management of COVID-19. *Applied Intelligence*, 50(11), 3913-3925. <https://doi.org/10.1007/s10489-020-01770-9>
- Mougari, N. E., Largeau, J. F., Himrane, N., Hachemi, M. y Tazerout, M. (2021). Application of artificial neural network and kinetic modeling for the prediction of biogas and methane production in anaerobic digestion of several organic wastes. *International Journal of Green Energy*, 18(15), 1584-1596. <https://doi.org/10.1080/15435075.2021.1914630>
- Neto, J. G., Ozorio, L. V., de Abreu, T. C. C., Dos Santos, B. F. y Pradelle, F. (2021). Modeling of biogas production from food, fruits and vegetables wastes using artificial neural network (ANN). *Fuel*, 285, 119081. <https://doi.org/10.1016/j.fuel.2020.119081>
- Zeb, A., Alzahrani, E., Erturk, V. S. y Zaman, G. (2020). Mathematical model for coronavirus disease 2019 (COVID-19) containing isolation class. *BioMed research international*, 2020(1), 3452402. <https://doi.org/10.1155/2020/3452402>
- Zaied, B. K., Rashid, M., Nasrullah, M., Bari, B. S., Zularisam, A. W., Singh, L., ... y Krishnan, S. (2020). Prediction and optimization of biogas production from POME co-digestion in solar bioreactor using artificial neural network coupled with particle swarm optimization (ANN-PSO). *Biomass Conversion and Biorefinery*, 1-16. <https://link.springer.com/article/10.1007/s13399-020-01057-6>