Revista Internacional de Investigación e Innovación Tecnológica

Página principal: www.riiit.com.mx

Análisis comparativo de modelos de regresión, una alternativa para la predicción de la generación de residuos sólidos urbanos a partir de su composición

Comparative analysis of regression models: an alternative for predicting municipal solid waste generation based on composition

Hernández-Cruz, A.[a], Castañeda-Gutiérrez, J.L.[b], Hernández-Hernández, J.C.[b]*, Sánchez-Olarte, J.[c]

[a] Doctorado en Ciencias de la Ingeniería; Tecnológico Nacional de México / Instituto Tecnológico de Apizaco; CP 90491, Tzompantepec, Tlaxcala. ORCID: https://orcid.org/0000-0001-8704-5460.
[b] División de Estudios de Posgrado e Investigación; Tecnológico Nacional de México / Instituto Tecnológico de Apizaco; CP 90491, Tzompantepec, Tlaxcala. ORCID: https://orcid.org/0009-0004-3971-6935, https://orcid.org/0000-0001-7245-9564*.
[c] División de Ingeniería en Gestión empresarial; Tecnológico Nacional de México / Instituto Tecnológico Superior de San Martín Texmelucan; CP 74120, San Martín Texmelucan, Puebla. ORCID: https://orcid.org/0000-0003-4469-2726.
d95370469@apizaco.tecnm.mx; jorge.cg@apizaco.tecnm.mx; crispin.hh@apizaco.tecnm.mx*; josset.sanchez@smartin.tecnm.mx

Innovación tecnológica: Comparativa experimental de modelos de aprendizaje automático para la optimización de predicciones en la generación de Residuos Sólidos Urbanos.

Área de aplicación industrial: Gestión de Residuos sólidos Urbanos y modelos de predicción basados en aprendizaje automático.

Recibido: 31 julio 2026
Aceptado: 22 septiembre 2026

Abstract

The growing generation of municipal solid waste (MSW) and the operational closure of landfills pose critical challenges for municipal management. This study aimed to evaluate and compare the predictive performance of three regression architectures—second-order polynomial regression (PR, d=2), support vector regression (SVR), and extreme gradient boosting (XGBoost)—to estimate weekly MSW generation based on the waste types generated in the previous week (t-1). Methodologically, a historical time series (2010–2024) from a municipality in Tlaxcala was processed; a chronological split was performed, allocating 80% of the data for training and 20% for testing, using a time-series cross-validation scheme (TimeSeriesSplit with k=5). Empirical results demonstrated that polynomial regression achieved optimal performance (R^2 = 0.9972),

outperforming more complex non-linear models such as XGBoost (R^2 = 0.9926) and SVR (R^2 = 1.0585). These results indicate that increasing algorithmic complexity does not guarantee superior predictive capability when the time series retains a strong autoregressive structure and well-defined compositional trends. It is concluded that lightweight parametric architectures offer an efficient, low-computational-cost alternative, serving as a quantitative tool with the potential to support logistics planning and decision-making in municipal waste management.

Keywords: machine learning, polynomial regression, municipal solid waste.

Resumen

La creciente generación de residuos sólidos urbanos (RSU) y el cierre operativo de vertederos plantean desafíos críticos para la gestión municipal. El objetivo de este estudio fue evaluar y comparar el desempeño predictivo de tres arquitecturas de regresión: Regresión Polinómica de segundo orden (RP, d=2), Regresión de Vectores de Soporte (RVS) y Extreme Gradient Boosting (XGBoost), para estimar la generación semanal de RSU a partir de los tipos de residuos generados de la semana previa (*t-1*). Metodológicamente, se procesó una serie temporal histórica del 2010 a 2024 de un municipio de Tlaxcala, se realizó una división cronológica asignando el 80% para entrenamiento y 20% para prueba, bajo un esquema de validación cruzada temporal (TimeSeriesSplit y k=5). Los resultados empíricos demostraron que la Regresión Polinómica obtuvo un desempeño óptimo con R^2 = 0.9972, superando a modelos no lineales de mayor complejidad como XGBoost con R^2 = 0.9926 y SVR (R^2= -1.0585). Estos resultados muestran que incrementar la complejidad algorítmica no garantiza una mayor capacidad predictiva cuando la serie conserva una fuerte estructura autorregresiva y tendencias de composición bien definidas. Se concluye que las arquitecturas paramétricas ligeras son una alternativa eficiente y de bajo costo computacional, proyectándose como un insumo cuantitativo con el potencial de apoyar la planificación logística y la toma de decisiones en la gestión municipal de residuos.

Palabras clave: aprendizaje automático, regresión polinómica, residuos sólidos urbanos.

I. Introducción

En aumento en la producción de residuos sólidos urbanos (RSU) representa importantes retos para la gestión municipal a nivel global, ya que a nivel mundial se supera los 2.01 mil millones de toneladas anuales, proyectando 3.4 mil millones para los próximos 30 años (Cook et al., 2026). En América Latina y el Caribe se generan 541,000 t/día, destinado el 90% a disposición final y reciclando apenas el 10% (Programa, 2018), lo que absorbe hasta el 50% del presupuesto operativo en ciudades medianas (ONU-Habitat, s.f.). En México la cifra alcanza 139,000 t/día con solo 5% de tratamiento adecuado (Naturales, s.f.), panorama agravado por el cierre acelerado de rellenos sanitarios y la proliferación de vertederos clandestinos (Gobierno del Estado de Tlaxcala, 2025; Procuraduría Federal de Protección al Ambiente, 2026).

Frente a esta complejidad sistémica, la optimización operacional ha abordado rutas de recolección (Alshaikh & Abdelfatah, 2024; Hannan et al., 2020), modelos

biobjetivo (Alzate Montoya et al., 2024; Aydemir-Karadag, 2022) y enfoques multicriterio (Barma & Modibbo, 2022; Li et al., 2025). Estrategias aplicadas en Shanghái redujeron 15.3% los costos y 12.7% de CO_2 (Zhou et al., 2022), mientras que en Monterrey se lograron reducciones de hasta 20% en distancias recorridas (Benítez-Bravo et al., 2021). En este contexto, las ciudades inteligentes han integrado Tecnologías de la Información, Internet de las Cosas y Aprendizaje Automático (Machine Learning, ML) para responder a estos desafíos urbanos (Aktemurer al., 2020; Szpilko et al., 2023).

La previsión de RSU ha evolucionado de estimaciones per cápita con modelos matemáticos de programación lineal (Koushik et. al, 2020) hacia modelos predictivos avanzados Support Vector Regression, Random Forest, XGBoost abarcando transporte, tratamiento y disposición final (Abdallah et al., 2020; Arroyo et al., 2023; Subedi et al., 2025; Zhang et al., 2024). Algoritmos como SVR, Regresión Polinómica (PR) y XGBoost destacan por modelar dinámicas no lineales complejas y adaptarse a datos con preprocesamiento estructurado (Khoirunisa et al., 2025; Bhattacharya et al., 2021; Soegianto et al., 2024; Xia et al., 2022). No obstante, existen vacíos críticos al omitir la integración de depuración temporal, clasificación detallada por tipología de residuo y modelado comparativo cuantitativo en América Latina (Zou et al., 2025). Específicamente, existe una limitación recurrente en la literatura al no incorporar la composición fraccionada semanal en *t-1* como variable dinámica de entrada para predecir la masa total en el periodo futuro (t).

Para resolver esta limitación, este estudio evalúa y compara el desempeño predictivo de tres arquitecturas de aprendizaje supervisado: Regresión Polinómica (PR), Regresión de Vectores de Soporte (SVR) y Extreme Gradient Boosting (XGBoost). Utilizando doce categorías de residuos desfasadas en *t-1* como variables independientes, se evalúa el rendimiento mediante R^2, MAE y RMSE. El análisis reconoce metodológicamente la influencia de la estructura por composición y la dependencia autorregresiva en los márgenes de ajuste de modelos complejos. Los resultados se proyectan como un insumo analítico para la toma de decisiones en la planificación municipal de recolección y la reducción del confinamiento en vertederos, alineándose con las metas de sustentabilidad de la Agenda 2030 (CEPAL, s.f.). El artículo se estructura de la siguiente manera: la Sección II presenta los materiales y equipo; la Sección III detalla los métodos experimentales y la formulación matemática; la Sección IV aborda los resultados y la discusión; y la Sección V expone las conclusiones de forma sintetizada.

II. Materiales y equipos

Todos los experimentos computacionales y las arquitecturas de los modelos se implementaron utilizando Python 3.10, empleando bibliotecas especializadas para el análisis de datos y aprendizaje automático de alto rendimiento (por ejemplo, scikit-learn v1.2, pandas v2.0 y xgboost v1.7).

Los datos analizados provienen de la Secretaría de Medio Ambiente del Estado de Tlaxcala (2025). Las 12 categorías de residuos se definieron como variables independientes (X) en el periodo ($t-1$) detalladas en la Tabla 1, mientras que la variable dependiente (Y) representa la masa total generada en el periodo futuro (t).

Tabla 1. Clasificación de RSU.

No.	Categoría de residuos	Subproducto asociado
1	Orgánicos	Orgánico
2	Papel y cartón	Papel cerrado, cartón, cartón cerrado, cartón multicapa
3	Plástico	LDPE (bolsas), PP, PET, PVC, PS, HDPE, bolsas metalizadas, caucho
4	Sanitarios	Residuos sanitarios (toallas, tampones, pañales), papel higiénico y servilletas.
5	Textiles	Textiles
6	Otros	Otros residuos / inertes / finos
7	Metales	Materiales ferrosos, materiales no ferrosos (latas de aluminio)
8	Madera	Madera
9	Vidrio	Vidrio transparente, vidrio de color
10	Construcción	Materiales de construcción
11	Peligroso	Residuos potencialmente peligrosos
12	Electrónicos	Residuos electrónicos

El conjunto de datos históricos abarca registros semanales desde 2010 hasta 2024, consolidando un total de 700 observaciones. El área de estudio corresponde a un municipio del estado de Tlaxcala. Cabe destacar que, por motivos de confidencialidad institucional solicitados por las autoridades proveedoras de la información, la ubicación geográfica exacta y el nombre del municipio se mantienen en anonimato. Sin embargo, esto no afecta la reproducibilidad cuantitativa del análisis.

III. Métodos experimentales

Se adoptó un enfoque cuantitativo y comparativo mediante aprendizaje automático supervisado para la estimación de RSU. El análisis utiliza una serie temporal retrospectiva de 700 observaciones semanales recolectadas entre 2010 y 2024. Se evalúan tres arquitecturas de regresión con distintos fundamentos matemáticos: RP, SVR y XGBoost. Cada algoritmo se sometió a un protocolo estandarizado de entrenamiento, basado en las doce categorías fraccionadas de la semana previa (*t-1*) para predecir la masa generada en el periodo futuro (*t*), aplicando una división temporal estricta (80% entrenamiento, 20% prueba) y una evaluación estadística mediante las métricas R^2, MAE y RMSE.

La Figura 1 sintetiza el flujo metodológico implementado, desde la adquisición y preprocesamiento de los datos por tipo de RSU y el ajuste de hiperparámetros hasta la comparación final del desempeño predictivo entre las arquitecturas evaluadas TimeSeriesSplit, (k=5), hasta la comparación final del desempeño predictivo entre las arquitecturas evaluadas.

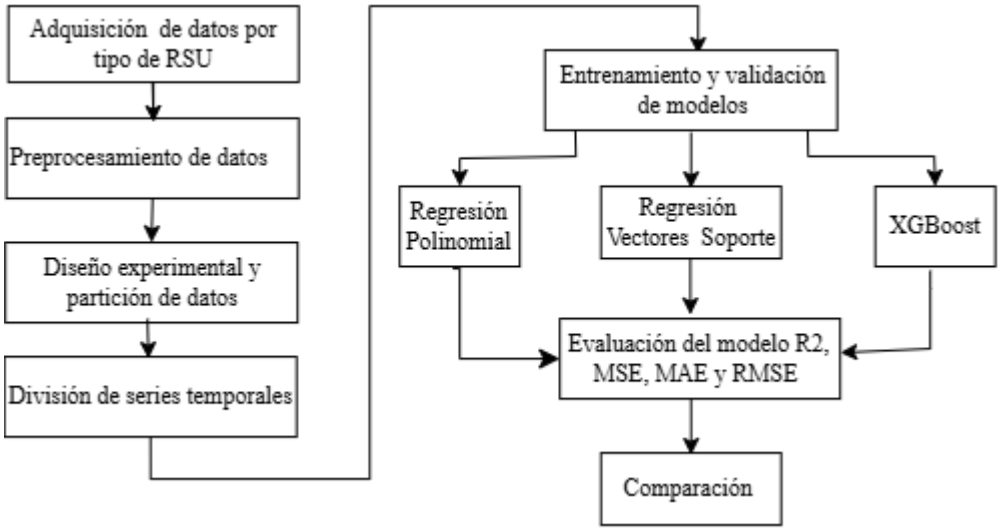


Figura 1. Esquema general del flujo metodológico de la investigación.

A. Adquisición y preprocesamiento de datos

La fase inicial consistió en identificar y extraer los registros brutos procedentes de las fuentes primarias institucionales, consolidándolos en una matriz de datos (formato CSV). Posteriormente, se ejecutó un protocolo de control de calidad de los datos, lo que incluye la limpieza de datos, la imputación de valores faltantes, la detección de valores atípicos y la normalización de características para garantizar la integridad y la estabilidad numérica del conjunto de datos antes de la etapa de modelado.

Modelos de Regresión e hiperparámetros

Para estimar la generación semanal de Residuos Sólidos Urbanos (RSU), se evaluaron tres arquitecturas con supuestos matemáticos diferenciados: RP, SVR y XGBoost. Todos los experimentos se ejecutaron sobre una partición temporal de la serie histórica, 80% entrenamiento y 20% para la prueba, sin reordenamiento aleatorio (shuffle=False), con el fin de preservar la secuencia cronológica y evitar la fuga de información futura (data leakage) (Cerqueira et al., 2021; Hyndman & Athanasopoulos, 2021).

Las 12 categorías de residuos en *t-1* se ingresaron como matriz explicativa *(X)*, mientras que la masa total semanal en *t* se definió como variable objetivo *(Y)*. Previo al ajuste de modelos sensibles a la escala como SVR, las características se estandarizaron mediante escalado Z-score (StandardScaler), calculando la media y varianza únicamente sobre el conjunto de entrenamiento. La optimización de hiperparámetros se realizó sobre el conjunto de entrenamiento mediante búsqueda en malla (GridSearchCV) en combinación con un esquema de validación cruzada para series temporales (TimeSeriesSplit con $k=5$). Para la RP (d=*2*) se aplicó un ajuste directo por Mínimos Cuadrados Ordinarios (OLS). Los métodos de preprocesamiento, esquemas de selección y la configuración de hiperparámetros óptimos seleccionados para cada arquitectura se sintetizan en la Tabla 2.

B. Formulación matemática de los modelos

Para estimar la generación semanal de RSU a partir de las categorías independientes, se precisaron diferentes arquitecturas de regresión con distintos supuestos matemáticos:

1. Regresión polinómica (PR)

Inicialmente se consideró como referencia inicial la regresión lineal múltiple que asume una relación lineal entre la variable de respuesta. *Y* y un conjunto de predictores (Yang et al., 2021). El modelo se expresa mediante:

$$Y = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p + \varepsilon \quad \textbf{(Ec.1)}$$

Sin embargo, para capturar interacciones complejas y curvaturas sin hacer de lado la estimación eficiente por Mínimos Cuadrados Ordinarios (OLS), se derivó la Regresión Polinómica de segundo orden (d=2) como su extensión directa (Chanchí Golondrino et al., 2020):

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \beta_3 X^3 \cdots + \beta_p X^p + \varepsilon \quad \textbf{(Ec. 2)}$$

Donde p=12 representa el número de categorías fraccionadas en *t-1*, y ε es el término de error.

2. Regresión de Vectores de Soporte (SVR)

Para evaluar las posibles relaciones no lineales mediante un enfoque no paramétrico, se utilizó el modelo de Regresión de Vectores de Soporte (Abdallah et al., 2020), este modelo busca una función $f(x)$ que no se desvíe más de un umbral ε de los resultados reales y_i:

$$f(x) = (w^\top, \phi(x)) + b \quad \textbf{(Ec. 3)}$$

Donde *w* equivale al vector de pesos y *b* se refiere al término de sesgo y , $\phi(x)$ representa la transformación del espacio de entrada mediante la función de núcleo RBF.

3. Potenciación de Gradiente XGBoost

Con un enfoque de aprendizaje en ensamble, XGBoost construye iterativamente una serie de árboles de decisión f_k mediante descenso de gradiente (Wiens et al., 2025). La función objetivo a optimizar en la iteración *t* se define como:

$$L^{(t)} = \sum_{i=1}^{n} l(y_i, \hat{y}_i^{(t-1)} + \ + f_t(x_i)) + \Omega(f_t) \quad \textbf{(Ec. 4)}$$

Donde l es una función de pérdida diferenciable que mide la discrepancia entre la etiqueta real y_i la predicción $\hat{y}_i$, aunado a ello $\Omega(f_t) = \gamma T + \frac{1}{2}\lambda \sum_{j=1}^{T} \omega_j^2$ es un término de regularización para penalizar la complejidad del árbol donde *T* es el número de hojas y *w* son sus pesos. Para controlar la velocidad de aprendizaje y prevenir el sobreajuste, se establece una tasa de aprendizaje (η) al incorporar un nuevo árbol.

C. Diseño experimental y división temporal de datos

Para evaluar los modelos y garantizar la validez metodológica, la serie temporal se dividió mediante una representación cronológica de entrenamiento y prueba (*Train/Test Split*). Se asignó el 80% inicial de los datos históricas al conjunto de entrenamiento para la optimización de parámetros, reservando el 20% restante que corresponde al periodo más reciente como conjunto de prueba independiente. Esta separación secuencial sin reordenamiento aleatorio (*shuffle=False*) evita la fuga de información (*data leakage*) y simula un escenario real de pronóstico temporal (Cerqueira et al., 2021).

Las variables de entrada corresponden a la matriz de composición fraccionada de la semana previa (*t-1*), mientras que la variable objetivo es la masa total de RSU generada en la semana actual (*t*). Previo al entrenamiento de modelos sensibles a la escala como SVR, las características se estandarizaron mediante escalado Z-score (*Standard*

Scaler), calculando los parámetros exclusivamente sobre la partición de entrenamiento.

- **Configuración e hiperparámetros de los modelos**

La optimización de los hiperparámetros se realizó sobre la partición de entrenamiento mediante búsqueda exhaustiva en malla (*GridSearchCV*) en combinación con validación cruzada para series temporales (*TimeSeriesSplit con k=5*), garantizando el respeto al orden cronológico del flujo de datos para prevenir la fuga de información (Cerqueira et al., 2020).

La Regresión Polinómica aplicó una transformación de grado *d=2* sobre las variables en t-1 con ajuste por Mínimos Cuadrados Ordinarios (OLS). Para SVR (núcleo RBF) y XGBoost, los rangos del espacio de búsqueda y las configuraciones óptimas que minimizan el error de validación, consolidándose en la Tabla 2. Todas las rutinas se implementaron en el entorno Python 3.10.

Tabla 2. Resumen de configuración e hiperparámetros de los modelos evaluados.

Modelo	Preprocesamiento	Método de Selección	Hiperparámetros Óptimos
Regresión Polinómica	Polynomial Features (*d=2*)	Ajuste directo (OLS)	Grado *d=2*
SVR (RBF)	Standard Scaler (Z-score)	Grid Search CV (*Time Series Split, k=5*)	*C=100, ϵ=0.01, γ='scale'*
XGBoost	Ninguno (basado en árboles)	Grid Search CV (*Time Series Split, k=5*)	*n_estimators=100, max_depth=4, learning_rate=0.05, subsample=0.8*

Fuente: Elaboración propia. Estrategia de optimización e hiperparámetros adaptada de las metodologías propuestas por Alqahtani y El-Khatib (2024) y Zou et al. (2025) para series temporales de RSU.

IV. Resultados y discusión

La evaluación inicial de la calidad de los datos reveló una ausencia total de valores faltantes, eliminando la necesidad de técnicas de imputación. El análisis de la variable objetivo (*Y*) mediante la Estimación de Densidad de Kernel (KDE) mostró una distribución asimétrica y no normal, con una concentración de observaciones en niveles bajos y un rango operativo entre 49,700.0 kg y 77,232.79 kg, con una media de 61,140.09 kg, una mediana de 60,055.12 kg y una desviación estándar de 8,117.0 kg como se muestra en la Figura 2.

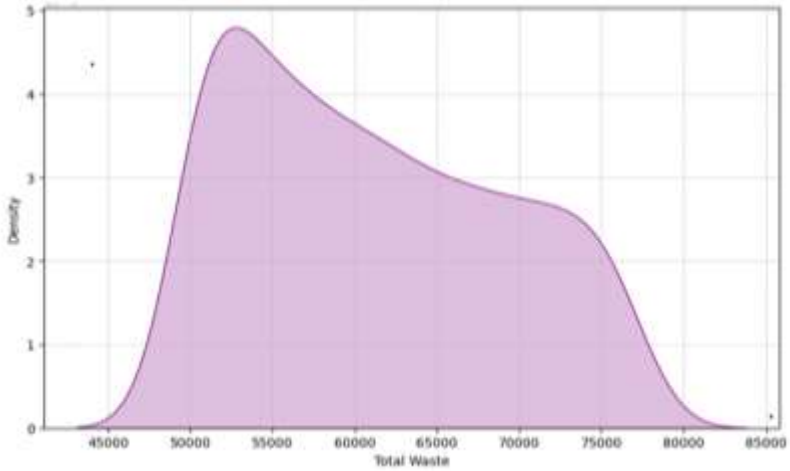


Figura 2. Distribución de probabilidad de la variable objetivo mediante la estimación de densidad de Kernel (KDE).

Para caracterizar la estructura interna de la serie, se calculó la matriz de correlación de Pearson que se muestra en la Figura 3.

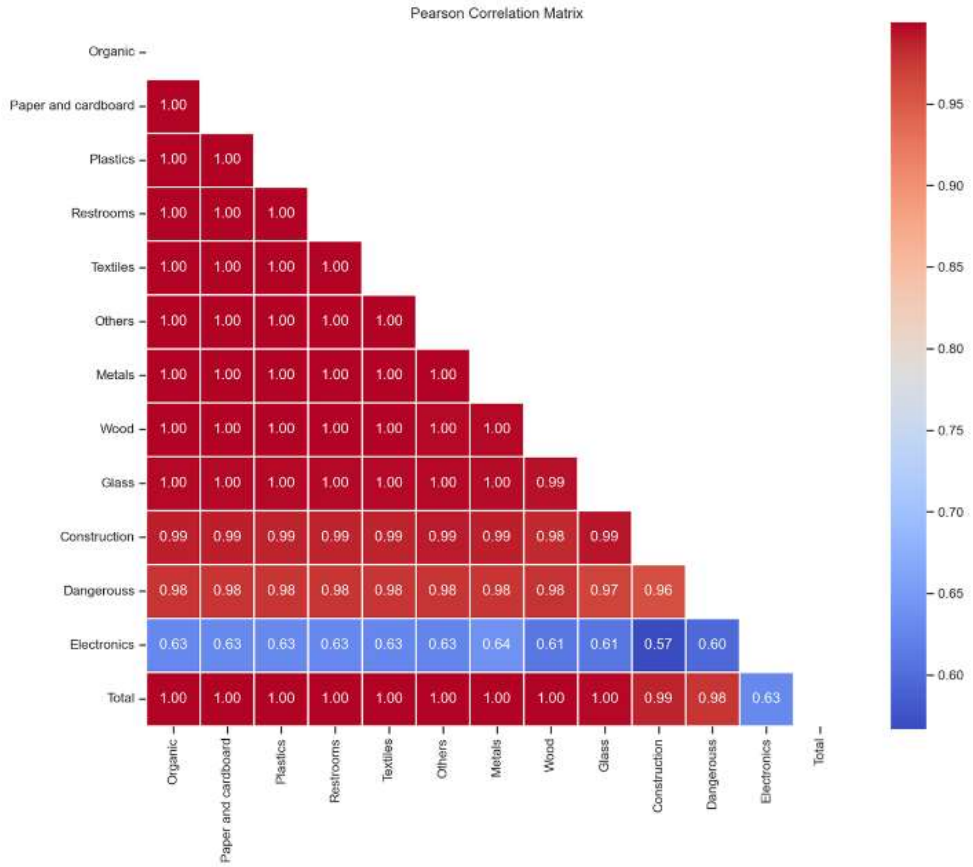


Figura 3. Matriz de correlación de Pearson entre los tipos de residuos y la masa total.

La matriz de correlación muestra coeficientes altos ($r > 0.95$) entre múltiples variables predictoras y la variable objetivo. Si bien esta condición introduce una elevada multicolinealidad, lo que restringe la interpretación causal o inferencial de los coeficientes individuales, el estudio se aborda desde una perspectiva puramente predictiva.

A. Evaluación del pronóstico temporal

El conjunto de datos se particionó cronológicamente (shuffle=False), asignando las semanas 1 a 560 (80%) para entrenamiento y las semanas 561 a 700 (20%) como ventana de prueba futura. Esta estructura evaluó directamente la capacidad de generalización en horizontes no vistos.

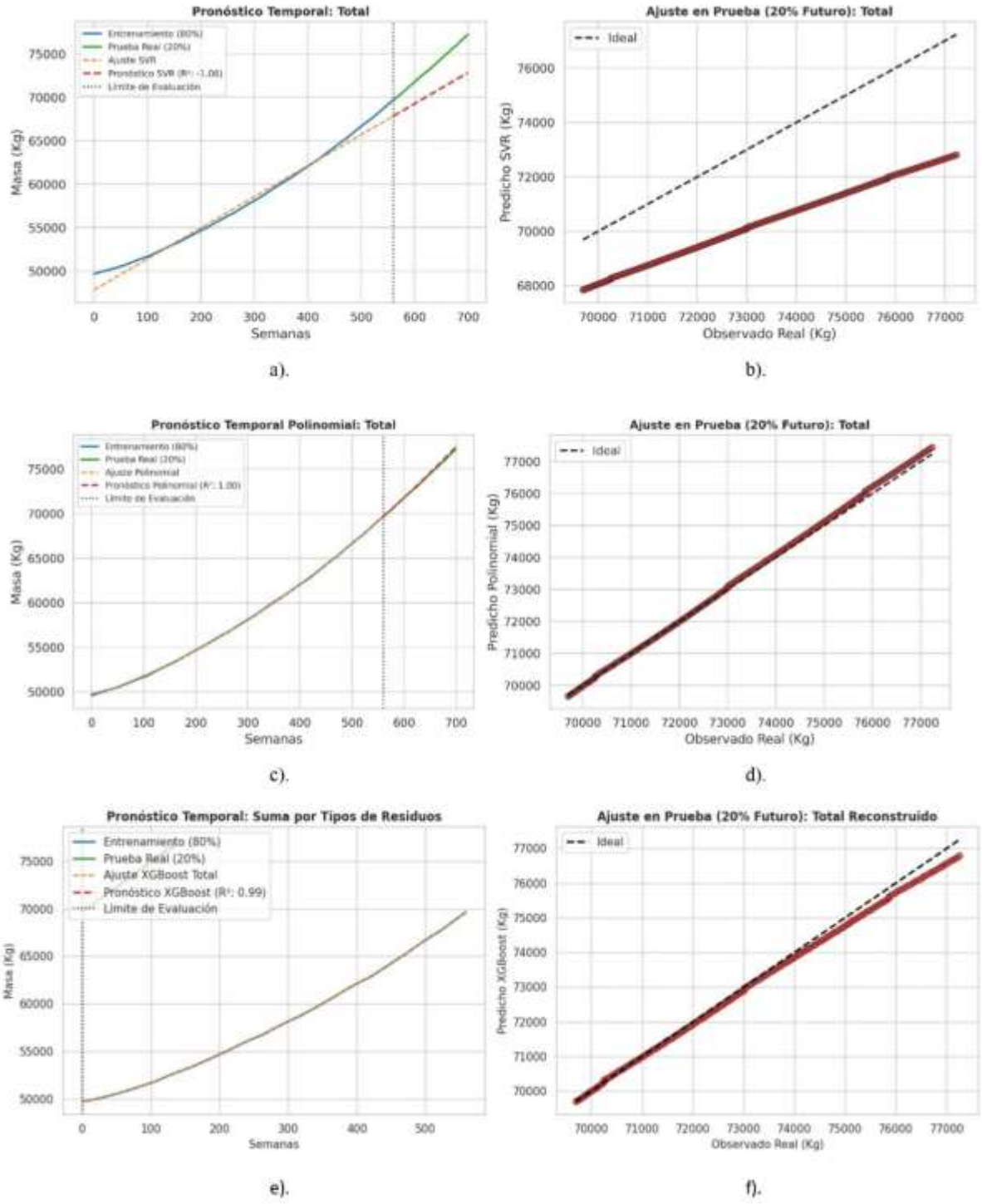


Figura 4. Pronóstico temporal y desempeño en prueba (20% futuro) para la masa total de residuos condicionado por la composición de residuos: a) y b) SVR, c) y d) Regresión Polinomial, y e) y f) XGBoost.

La evaluación en el conjunto de prueba se muestra en la Figura 4, donde se observan comportamientos divergentes. En el caso del modelo SVR obtuvo un R^2 = -1.0585, en la (Figura 4 a-b) se muestra como colapsó tras la semana 560 por la incapacidad del núcleo RBF para extrapolar la tendencia temporal al acoplar los retardos de las distintas

categorías de residuos. Sin embargo, con la RP d=2 se obtuvo un R^2 = 0.9972, en la (Figuras 4c–d) se muestra que se ajustó la trayectoria real gracias a la naturaleza continua de la curva cuadrática. Por otra parte, XGBoost: R^2 = 0.9926, (Figuras 4e–f) logró una precisión equivalente al capturar las dependencias autorregresivas multivariadas en t-1. Ambos modelos mostraron una alineación homogénea sobre la diagonal ideal, confirmando que considerar la estructura de la serie garantiza la estabilidad de la predicción futura.

B. Evaluación del desempeño e inferencia predictiva

Los resultados numéricos demuestran que la Regresión Polinómica de segundo grado alcanzó el rendimiento óptimo, registrando un R^2 =0.9972 con los menores márgenes de error MAE=87.6302 y RMSE=115.3839. Estos hallazgos muestran que un incremento en la complejidad algorítmica no garantiza una mayor precisión predictiva sobre series temporales. XGBoost, aunque mantuvo una elevada precisión R^2 =0.9926, presentó errores marginalmente superiores MAE=142.2482 y RMSE=188.6473. Por su parte, el modelo SVR sufrió un colapso en su capacidad de aproximación en el horizonte futuro R^2=−1.0585, RMSE=3,047.3023.

C. Análisis e interpretación de hallazgos

Al comparar la precisión predictiva de las arquitecturas evaluadas frente a los antecedentes reportados en la literatura especializada (Tabla 3), se observan aspectos metodológicos clave respecto a la dinámica territorial y la complejidad de los modelos. En primer lugar, los resultados obtenidos para la serie semanal del municipio de Tlaxcala muestran la efectividad de la RP d=2, la cual registró el desempeño global más alto del estudio R^2 = 0.9972, RMSE = 115.3839, MAE = 87.6302$ y MAPE = 0.12. Este nivel de ajuste supera marginalmente al coeficiente alcanzado por Chanchí Golondrino et al. (2020) R^2 = 0.99662, demostrando que la estructura parabólica ajustada por Mínimos Cuadrados Ordinarios abstrae fidedignamente la tendencia continua y no lineal de las series temporales de residuos sin incurrir en sobreajuste. Por otra parte, la aproximación mediante ensambles no lineales mostró matices relevantes. El modelo XGBoost exhibió una capacidad predictiva sobresaliente R^2 = 0.9926, MAPE = 0.19, ubicándose muy cerca del ajuste polinómico y superando los coeficientes reportados por Yang et al. (2021) para el mismo algoritmo R^2 = 0.90 y por Khoirunisa et al. (2025) RMSE = 0.147. No obstante, la ligera ventaja en precisión de la RP frente a XGBoost resalta un principio de parsimonia en la escala municipal: mientras que los árboles de decisión segmentan el espacio de búsqueda mediante funciones a partes sensibles a fluctuaciones locales, la formulación cuadrática absorbe suavemente la variabilidad estacional de las 12 categorías en t-1.

Tabla 3. Comparación del rendimiento del modelo para la predicción de la generación de RSU.

Referencia (APA)	Ámbito y Territorio de Estudio	Modelo Evaluado	MAE	RMSE	MAPE (%)	R^2
Yang et al. (2021)	Recolección RSU municipal (China)	DNN	861.03	1288.80	N/A	0.97
Yang et al. (2021)	Recolección RSU municipal (China)	XGBoost	1219.91	1706.78	N/A	0.90
Yang et al. (2021)	Recolección RSU municipal (China)	RF	774.30	1348.63	N/A	0.91
Khoirunisa et al. (2025)	Residuos con imputación (Java, Indonesia)	XGBoost	N/A	0.147 t	N/A	N/A
Chanchí Golondrino et al. (2020)	Serie temporal continúa acumulada	Regresión Polinómica (d=2)	N/A	150.04 u	N/A	0.99662
Este estudio	**Serie semanal (Municipio de Tlaxcala)**	**Regresión Polinómica ($d=2)**	**87.6302**	**115.3839**	**0.12**	**0.9972**
Este estudio	Serie semanal (Municipio de Tlaxcala)	SVR	3136.5751	3047.3023	4.13	-1.0585
Este estudio	Serie semanal (Municipio de Tlaxcala)	XGBoost	142.2482	188.6473	0.19	0.9926

Un hallazgo crítico reside en el colapso operativo del modelo SVR R^2 = -1.0585, MAPE = 4.13. Un coeficiente R^2 negativo evidencia que la proyección del hiperplano mediante el kernel ajustado resultó menos precisa que la media aritmética simple de la serie. Este comportamiento dista de la estabilidad habitual de SVR en la literatura y se atribuye a la alta dimensionalidad relativa al ajustar los retardos temporales de las distintas tipologías de residuos, lo que provocó una saturación en las funciones de margen. En términos de aplicabilidad tecnológica para la gestión pública local, la clara superioridad de la Regresión Polinómica representa una ventaja sustancial. Al no depender del ajuste hiperparamétrico ni de la opacidad algorítmica de los modelos de ensamble o redes profundas como el DNN de Yang et al., 2021, R^2 = 0.97, el modelo cuadrático proporciona una formulación matemática transparente, determinista y de bajísimo costo computacional, ideal para su integración inmediata en herramientas de planificación operativa municipal

Desde la perspectiva de la ingeniería y la gestión municipal, estos hallazgos confirman la viabilidad de implementar sistemas predictivos de alta fidelidad, computacionalmente ligeros y claros para estimar la generación semanal de RSU. La capacidad de anticipar volúmenes mediante el uso de datos históricos (t-1) brinda un insumo cuantitativo con el potencial de apoyar la planificación logística, la programación de rutas y la toma de decisiones operativas. De tal modo que, las predicciones resultantes se posicionan básicamente como una herramienta para el diseño futuro de políticas públicas y estrategias de minimización de residuos orientadas a la sostenibilidad ambiental.

Conclusiones

En este estudio se evaluó el desempeño de tres arquitecturas de regresión, Regresión Polinómica, XGBoost y SVR para la estimación de la generación semanal de

residuos sólidos urbanos (RSU), empleando una serie histórica comprendida entre los años 2010 y 2024 bajo una partición temporal estricta (shuffle=False). Los resultados empíricos demostraron que la Regresión Polinómica de segundo orden ofrece la mayor precisión y estabilidad predictiva en el horizonte futuro con un R^2 = 0.9972, superando a modelos no lineales de mayor complejidad algorítmica como XGBoost y SVR. Este comportamiento confirma que un aumento en la complejidad del modelo no garantiza una mejora en la capacidad de extrapolación cuando la serie temporal conserva patrones autorregresivos y tendencias de composición bien definidas. Dada su alta precisión predictiva y su reducido costo computacional, la regresión polinómica se consolida como un insumo cuantitativo con el potencial de apoyar la planificación logística y la toma de decisiones en la gestión municipal de RSU. Finalmente, se sugiere incorporar en investigaciones futuras variables exógenas (como indicadores socioeconómicos, demográficos o climáticos) que permitan evaluar la robustez del modelo frente a perturbaciones exógenas que alteren la dinámica histórica de generación de los RSU.

Bibliografía

1 Cook, E., Ionkova, K., Bhada-Tata, P., Yadav, S., & van Woerden, F. (s.f.). *Panorama global de la gestión de residuos sólidos hacia la circularidad hasta 2050. https://doi.org/10.1596/978-1-4648-2309-1*

2 Programa de las Naciones Unidas para el Medio Ambiente (PNUMA/UNEP). (2018). *Perspectiva de la Gestión de Residuos en América Latina y el Caribe*. s.l.: PNUMA. Recuperado de https://wedocs.unep.org/handle/20.500.11822/26448

3 ONU-Hábitat. (s.f.). *Recolectar y eliminar residuos de manera eficiente*. s.l.: ONU-Hábitat. Recuperado de https://onu-habitat.org/index.php/recolectar-y-eliminar-residuos-de-manera-eficiente

4 Secretaría de Medio Ambiente y Recursos Naturales (SEMARNAT). (s.f.). *Semarnat presenta diagnóstico nacional para transformar la gestión de residuos y avanzar hacia la economía circular*. Ciudad de México: Gob.mx. Recuperado de http://www.gob.mx/semarnat/prensa/semarnat-presenta-diagnostico-nacional-para-transformar-gestion-de-residuos-y-avanzar-hacia-la-economia-circular

5 Secretaría de Medio Ambiente del Estado de Tlaxcala. (2025). *Diagnóstico básico y programa para la prevención, gestión integral y economía circular de los residuos del estado de Tlaxcala 2025–2027*. Tlaxcala: Gobierno del Estado de Tlaxcala. Recuperado de https://cjuridica.tlaxcala.gob.mx/images/Diagnostico/DIAGNOSTICO%20BASICO%20Y%20PROGRAMA%20PARA%20LA%20PREVENCION%20GESTION%20INTEGRAL%20Y%20ECONOMIA%20CIRCULAR%20DE%20LOS%20RESIDUOS%202.12.2025.pdf

6 Gobierno del Estado de Tlaxcala. (2025, 3 de noviembre). *Gobierno del Estado inicia clausura de tiraderos de Panotla y Huamantla*. Tlaxcala:

Coordinación de Comunicación. Recuperado de https://comunicacion.tlaxcala.gob.mx/index.php?view=article&id=20136:tlaxcala-avanza-como-referente-nacional-en-el-cierre-de-sitios-de-disposicion-final-semarnat

7 Procuraduría Federal de Protección al Ambiente (PROFEPA). (2026, 9 de enero). *Profepa impone multa y clausura definitiva a relleno sanitario de Huehuetla, Puebla*. Ciudad de México: Gob.mx. Recuperado de https://www.gob.mx/profepa/prensa/profepa-impone-multa-y-clausura-definitiva-a-relleno-sanitario-de-huehuetla-puebla

8 Alshaikh, R., & Abdelfatah, A. (2024). "Técnicas de optimización en la gestión de residuos sólidos municipales: una revisión sistemática". *Sustainability*, vol. 16, no. 15, julio de 2024, art. 6585. https://doi.org/10.3390/su16156585

9 Hannan, M. A., Akhtar, M., & Begum, R. A. (2020). "Objetivos, limitaciones, enfoques de modelado y desafíos de la optimización de la recolección de residuos sólidos para el logro de los objetivos de desarrollo sostenible". *Journal of Cleaner Production*, vol. 277, art. 123557, diciembre de 2020, pp. 1-15. https://doi.org/10.1016/j.jclepro.2020.123557

10 Alzate, Montoya, P. M., Bustamante Gutiérrez, L. C., & Alzate Álvarez, A. M. (2024). "Un modelo de rutas para la recolección de residuos sólidos domiciliarios: Un caso de estudio". *Revista de Métodos Cuantitativos para la Economía y la Empresa*, vol. 38, no. 1, junio de 2024, pp. 1-14. https://doi.org/10.46661/rev.metodoscuant.econ.empresa.7820

11 Aydemir-Karadag, A. (2022). "Algoritmo de búsqueda adaptativa de vecindario amplio con dos objetivos para el problema de enrutamiento de inventario de ubicación periódica de residuos sanitarios". *Arabian Journal for Science and Engineering*, vol. 47, no. 3, marzo de 2022, pp. 3861-3876. https://doi.org/10.1007/s13369-021-06106-4

12 Barma, M., & Modibbo, U. M. (2022). "Modelo de optimización matemática multiobjetivo para la gestión de residuos sólidos municipales con análisis económico de la reutilización/reciclaje de materiales de desecho recuperados". *Journal of Computational and Cognitive Engineering*, vol. 1, no. 3, septiembre de 2022, pp. 122-137. https://doi.org/10.47852/bonviewJCCE149145

13 Li, W., Wang, P., Xu, Y., Pan, L., Nie, C., & Yang, B. (2025). "Optimización multiobjetivo de la recolección de residuos sólidos municipales basada en la búsqueda adaptativa de vecindarios grandes". *Electronics*, vol. 14, no. 1, enero de 2025, art. 103. https://doi.org/10.3390/electronics14010103

14 Zhou, J., Zhang, M., & Wu, S. (2022). "Problema de enrutamiento

vehicular multiobjetivo para la clasificación y recolección de residuos con consideraciones de sostenibilidad: el caso de la ciudad de Shanghái". *Sustainability*, vol. 14, no. 18, septiembre de 2022, art. 11498. https://doi.org/10.3390/su141811498

15 Benítez-Bravo, R., Gomez-González, R., Rivas-García, P., Botello-Álvarez, J. E., Huerta-Guevara, O. F., García-León, A. M., & Rueda-Avellaneda, J. F. (2021). "Optimización de rutas de recolección de residuos sólidos municipales en un contexto latinoamericano". *Journal of the Air & Waste Management Association*, vol. 71, no. 11, noviembre de 2021, pp. 1415-1427. https://doi.org/10.1080/10962247.2021.1957040

16 Aktemur, I., Erensoy, K. & Kocyigit, E. "Optimización de la recogida de residuos en ciudades inteligentes mediante el uso de algoritmos evolutivos", 2020 International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA), Ankara, Turkey, 2020, pp. 1-8, https://doi.org/10.1109/HORA49412.2020.9152865

17 Szpilko, D., de la Torre Gallegos, A., Jiménez Naharro, F., Rzepka, A., & Remiszewska, A. (2023). "Gestión de residuos en la ciudad inteligente: prácticas actuales y direcciones futuras". *Recursos*, vol. 12, no. 10, octubre de 2023, art. 115. https://doi.org/10.3390/resources12100115

18 P. Koushik, S. Chattopadhyay, A. Dutta, A. P. Krishna, S. Ray, "A comprehensive optimization model for integrated solid waste management system: A case study", Revista de investigación de Ingeniería Ambiental, Volumen 24, no.2, pp 220-237, 2019. https://doi.org/10.4491/eer.2018.132

19 Abdallah, M., Abu Talib, M., Feroz, S., Nasir, Q., Abdalla, H., & Mahfood, B. (2020). "Aplicaciones de inteligencia artificial en la gestión de residuos sólidos: una revisión sistemática de la investigación". *Waste Management*, vol. 109, mayo de 2020, pp. 231-246. https://doi.org/10.1016/j.wasman.2020.04.057

20 Arroyo, L. A. B., Nicola, R. J. V., Moreno, W. V., & Baque, G. D. L. (2023). "La intervención de la logística en la gestión de residuos sólidos en Latinoamérica". *RECIAMUC*, vol. 7, no. 2, abril de 2023, pp. 328-350. https://doi.org/10.26820/reciamuc/7.(2).abril.2023.328-350

21 Subedi, A., Shrestha, S., & Paudel, S. R. (2025). "Aplicación del aprendizaje automático para la predicción de la generación de residuos sólidos". *Natural Built Social Environment Health*, vol. 1, no. 2, enero de 2025, pp. 14-33. https://doi.org/10.63095/NBSEH.25.831599

22 Zhang, J., Wen, Z., Hu, Y., Fei, F., Wang, Y., & Xie, Y. (2024). "Simulación de sistemas y metodología de optimización

multiobjetivo para la gestión sostenible de la clasificación de residuos sólidos municipales: un estudio de caso en China". *Producción y consumo sostenibles*, vol. 50, noviembre de 2024, pp. 475-485. https://doi.org/10.1016/j.spc.2024.08.014

23 Khoirunisa, R., Sani, A. F., Riatma, D. L., Masbahah, M., & Rachman, Y. F. (2025).” Imputing Data and Predicting Waste with Machine Learning in East Java. Brilliance: Research of Artificial Intelligence, 5(1), 521–528. https://doi.org/10.47709/brilliance.v5i1.6461

24 Bhattacharya, S., Kalita, K., Čep, R., & Chakraborty, S. (2021). "Análisis comparativo del rendimiento predictivo de modelos de regresión durante el mecanizado de materiales compuestos". *Materials*, vol. 14, no. 21, noviembre de 2021, art. 6689. https://doi.org/10.3390/ma14216689

25 Soegianto, L. M., Hinandra, A. T., Suri, P. A., & Fajar, M. (2024). "Comparación del rendimiento de modelos en el sector inmobiliario utilizando regresión lineal, regresor de bosque aleatorio, SVR y redes neuronales". *Procedia Computer Science*, vol. 245, octubre de 2024, pp. 1139-1145. https://doi.org/10.1016/j.procs.2024.10.343

26 Xia, W., Jiang, Y., Chen, X., & Zhao, R. (2022). "Aplicación de algoritmos de aprendizaje automático en la gestión de residuos sólidos municipales: una mini revisión". *Waste Management & Research*, vol. 40, no. 6, junio de 2022, pp. 609-624. https://doi.org/10.1177/0734242X211033716

27 Zou, Q., Duan, H., Yang, Z., Wang, X., Tian, Y., Hou, H., & Yang, J. (2025). "Gestión del ciclo de vida de residuos sólidos asistida por aprendizaje automático: aplicaciones, limitaciones y oportunidades futuras". *Resources, Conservation and Recycling*, vol. 219, agosto de 2025, art. 108320. https://doi.org/10.1016/j.resconrec.2025.108320

28 Comisión Económica para América Latina y el Caribe (CEPAL). (s.f.). *Objetivos de Desarrollo Sostenible (ODS). Agenda 2030 en América Latina y el Caribe*. Santiago de Chile: CEPAL. Recuperado de https://agenda2030lac.org/es/objetivos-de-desarrollo-sostenible-ods

29 Wiens, M., Verone-Boyle, A., Henscheid, N., Podichetty, J. T., & Burton, J. (2025). "Un tutorial y un ejemplo práctico del algoritmo de inteligencia artificial eXtreme Gradient Boosting (XGBoost) para aplicaciones en el desarrollo de fármacos". *Clinical and Translational Science*, vol. 18, no. 3, marzo de 2025, art. e70172. https://doi.org/10.1111/cts.70172

30 Cerqueira, V., Torgo, L. y Mozetič, I. “Evaluación de modelos de pronóstico de series temporales: un estudio empírico sobre métodos de estimación del rendimiento”. Mach Learn 109, 1997–2028 (2020).

https://doi.org/10.1007/s10994-020-05910-7

31 Yang, L., Zhao, Y., Niu, X., Song, Z., Gao, Q., & Wu, J. (2021). "Pronóstico de residuos sólidos municipales en China basado en modelos de aprendizaje automático". *Fronteras en la Investigación Energética*, vol. 9, diciembre de 2021, art. 763977. https://doi.org/10.3389/fenrg.2021.763977

32 Chanchí Golondrino, G. E., Campo Muñoz, W. Y., & Sierra Martínez, L. M. (2020). "Aplicación de la regresión polinomial para la caracterización de la curva del COVID-19, mediante técnicas de machine learning". *Investigación e Innovación en Ingenierías*, vol. 8, no. 2, julio-diciembre de 2020, pp. 87-105. https://doi.org/10.17081/invinno.8.2.4103