



Hispanic Journal of Applied Science and Innovation (HISPASCI)

Vol. 1, Núm. 1 (2026) | DOI: <https://doi.org/10.66465/hispasci.1.2026.7>

Analítica predictiva en contabilidad de gestión: modelación de estructuras de costos no lineales mediante machine learning

Predictive analytics in management accounting: modeling non-linear cost structures using machine learning

Jefferson Xavier Romero Morales¹

¹Universidad de Investigación e Innovación de México (UIIX -MEXICO),

Facultad de Ciencias Administrativas y Tecnológicas

ORCID <https://orcid.org/0009-0009-3752-2367>

E-mail: jromerom@comunidad.uiix.wdu.mx

*Autor de correspondencia: jromerom@comunidad.uiix.wdu.mx

Recibido: 2026-01-01 | Revisado: 2026-01-08 | Aceptado: 2026-01-10 | Publicado: 2026-01-15

Resumen

El estudio analiza la aplicación de analítica predictiva y técnicas de machine learning en la contabilidad de gestión, con el objetivo de modelar estructuras de costos no lineales y evaluar su capacidad explicativa y predictiva frente a enfoques tradicionales. Dado que la disponibilidad de datos empresariales reales suele ser limitada por razones de confidencialidad, se empleó una estrategia basada en la simulación controlada de datos contables complementada con datos abiertos, incorporando patrones realistas de estacionalidad, dependencia temporal y comportamiento asimétrico de los costos. Metodológicamente, se compararon tres enfoques: regresión lineal, Random Forest y redes neuronales multicapa, utilizando variables rezagadas y validación temporal para evitar filtraciones de información. El desempeño de los modelos se evaluó mediante coeficiente de determinación, error absoluto medio y error porcentual medio absoluto. Los resultados evidencian que los modelos no lineales superan de forma consistente a la regresión clásica, alcanzando valores de ajuste elevados y reducciones sustanciales en los errores de predicción, lo que confirma la existencia de relaciones complejas entre el nivel de actividad y el margen operativo. En particular, el Random Forest mostró el mejor equilibrio entre precisión y estabilidad. Se concluye que la analítica predictiva constituye una herramienta robusta para la contabilidad de gestión, permitiendo una mejor comprensión del comportamiento de los costos y ofreciendo soporte cuantitativo para la toma de decisiones estratégicas, incluso en contextos donde los datos reales son limitados.

Palabras clave: *analítica predictiva; contabilidad de gestión; machine learning; estructuras de costos; modelos no lineales.*



Abstract

This study examines the application of predictive analytics and machine learning techniques in management accounting, aiming to model non-linear cost structures and assess their explanatory and predictive performance in comparison with traditional approaches. Given the restricted availability of real corporate accounting data due to confidentiality constraints, a controlled data simulation strategy combined with open data was employed, incorporating realistic patterns of seasonality, temporal dependence, and asymmetric cost behavior. Methodologically, three approaches were compared: linear regression, random forest, and multilayer neural networks, using lagged variables and time-based validation to prevent information leakage. Model performance was evaluated using the coefficient of determination, mean absolute error, and mean absolute percentage error. The results demonstrate that non-linear models consistently outperform classical regression, achieving higher goodness-of-fit and substantial reductions in prediction errors, thereby confirming the presence of complex relationships between activity levels and operating margins. In particular, the random forest model exhibited the best balance between accuracy and stability. The study concludes that predictive analytics represents a robust tool for management accounting, enhancing the understanding of cost behavior and providing quantitative support for strategic decision-making, even in environments with limited access to real-world data.

Keywords: *predictive analytics; management accounting; machine learning; cost structures; non-linear models.*

1. Introducción

La contabilidad de gestión desempeña un papel central en el proceso de toma de decisiones empresariales, al proporcionar información cuantitativa relevante para la planificación, el control y la evaluación del desempeño organizacional. Tradicionalmente, el análisis del comportamiento de los costos y del margen operativo se ha apoyado en modelos lineales y supuestos de proporcionalidad entre el nivel de actividad y los resultados financieros. No obstante, la creciente complejidad de los entornos productivos, caracterizados por estructuras de costos rígidas, economías de escala, efectos de aprendizaje y estacionalidad, ha puesto en evidencia las limitaciones de estos enfoques clásicos para capturar dinámicas no lineales y relaciones temporales complejas.

En este contexto, la analítica predictiva y las técnicas de machine learning han emergido como herramientas prometedoras para el análisis avanzado de datos contables. A diferencia de los modelos estadísticos tradicionales, estos enfoques permiten identificar patrones ocultos, interacciones no lineales y dependencias dinámicas sin imponer supuestos restrictivos sobre la forma funcional de los datos. Estudios recientes han documentado su utilidad en ámbitos como la predicción macroeconómica, la evaluación del desempeño financiero y el análisis de costos, destacando mejoras significativas en precisión predictiva frente a métodos convencionales.



Sin embargo, la aplicación del machine learning en la contabilidad de gestión aún enfrenta importantes desafíos. Uno de los principales es la limitada disponibilidad de datos contables empresariales de acceso público, debido a restricciones de confidencialidad y competencia. Esta situación ha llevado a que muchos estudios presenten resultados difíciles de replicar o con escasa transparencia metodológica. Adicionalmente, existen riesgos metodológicos relevantes, como la filtración de información (data leakage), el sobreajuste de modelos complejos y el uso inadecuado de métricas de evaluación, que pueden conducir a conclusiones engañosas desde una perspectiva de gestión.

Frente a estas limitaciones, el presente estudio propone un enfoque metodológico basado en la simulación controlada de datos contables, complementada con principios de validación temporal propios del análisis de series de tiempo. Esta estrategia permite generar conjuntos de datos con propiedades realistas —como estacionalidad, dependencia temporal y comportamiento no proporcional de los costos— manteniendo al mismo tiempo control sobre la estructura subyacente del proceso generador de datos. De este modo, se garantiza la coherencia económica del análisis y se evita la introducción de sesgos artificiales que inflen el desempeño predictivo de los modelos.

El objetivo principal de esta investigación es evaluar la capacidad de distintos modelos de analítica predictiva —regresión lineal, Random Forest y redes neuronales multicapa— para modelar y predecir el margen operativo a partir de variables contables rezagadas, en un contexto de contabilidad de gestión. Específicamente, se busca comparar su desempeño predictivo, analizar la relevancia de las variables explicativas y discutir las implicaciones prácticas de los resultados para la gestión empresarial.

Se aporta evidencia empírica sobre el valor añadido de los modelos no lineales en el análisis del comportamiento de los costos, reforzando la pertinencia del machine learning como herramienta complementaria a los enfoques tradicionales de contabilidad de gestión. En segundo lugar, propone un marco metodológico reproducible y conceptualmente sólido para estudios de analítica predictiva en contextos donde el acceso a datos reales es limitado, lo que resulta especialmente relevante para investigaciones aplicadas en economías emergentes. En conjunto, los hallazgos de este estudio buscan fortalecer el diálogo entre la contabilidad de gestión y la ciencia de datos, promoviendo un uso más riguroso y estratégico de las técnicas predictivas en el ámbito contable.

2. Metodología

2.1 Diseño de investigación y fundamentos teóricos

El estudio adopta un diseño cuantitativo, explicativo y predictivo, computacional y experimental, orientado a evaluar la eficacia comparativa de técnicas de analítica predictiva basadas en aprendizaje automático frente a modelos econométricos tradicionales en la modelación del margen operativo.

El análisis se contextualiza en el ámbito de la contabilidad de gestión, específicamente en escenarios caracterizados por estructuras de costos no lineales, presencia de umbrales operativos y potenciales deseconomías de escala, fenómenos ampliamente documentados en la literatura contable y de economía industrial.



Dada la restricción habitual de acceso a microdatos financieros empresariales por razones de confidencialidad corporativa, la investigación se fundamenta en una estrategia de simulación estocástica controlada, metodología validada en estudios de diseño experimental y evaluación de modelos predictivos (Anderson, 2003; Law & Kelton, 2015). Este enfoque permite aislar el fenómeno de interés —la no linealidad en la estructura de costos— minimizando perturbaciones exógenas no observables.

Se estructura bajo un enfoque de series de tiempo, incorporando explícitamente rezagos temporales (lags) con el fin de mitigar el riesgo de data leakage (fuga de información) y asegurar que los modelos operen bajo condiciones de incertidumbre comparables a las enfrentadas en la planificación presupuestaria real.

2.2 Generación y naturaleza de los datos

Se generó un conjunto de datos sintéticos (synthetic dataset) compuesto por

$$n = 240$$

observaciones mensuales, equivalentes a 20 años fiscales, simulando la evolución operativa y financiera de una empresa manufacturera sujeta a fluctuaciones cíclicas de demanda y capacidad productiva.

La simulación fue diseñada para preservar regularidades económicas plausibles, evitando procesos puramente aleatorios que carecerían de valor predictivo para los modelos evaluados.

2.2.1. Variable independiente: Nivel de actividad

La variable Nivel de Actividad (A_t) actúa como el principal cost driver del sistema, pudiendo interpretarse como unidades producidas, horas-máquina o nivel de utilización de la capacidad instalada.

Su dinámica temporal se modeló mediante una función sinusoidal con el objetivo de capturar estacionalidad y ciclos de negocio:

$$A_t = 50 + 40\sin\left(\frac{2\pi}{48}t\right)$$

Donde:

- t representa el período mensual,
- la periodicidad del ciclo económico es de 48 meses,
- la amplitud de la función garantiza fluctuaciones significativas entre escenarios de baja y alta actividad.

Este planteamiento permite evaluar el desempeño de los modelos en contextos operativos heterogéneos.



2.2.2. Variable dependiente: Margen operativo no lineal

El Margen Operativo (Mt) se definió mediante una función por tramos (piecewise function), diseñada para reproducir el comportamiento asimétrico de los costos y la presencia de umbrales operativos:

$$M_t = \begin{cases} 2000 + 10A_t + \varepsilon_t, & \text{si } A_t < 50 \text{ (Zona eficiente)} \\ 2000 + 10(A_t - 50)^2 + \varepsilon_t, & \text{si } A_t \geq 50 \text{ (Zona de estrés)} \end{cases}$$

El término cuadrático

$$(A_t - 50)^2$$

introduce una no linealidad severa, penalizando el margen cuando el nivel de actividad supera el umbral óptimo, lo cual representa deseconomías de escala, sobrecostos operativos y pérdida de eficiencia marginal.

Adicionalmente, se incorporó un término de error estocástico:

$$\varepsilon_t \sim N(0,100)$$

para simular la variabilidad inherente a los procesos contables reales, manteniendo un nivel de ruido controlado que permita la identificación de patrones estructurales.

2.3 Ingeniería de variables y enfoque temporal

Con el objetivo de garantizar una validez predictiva estricta, se implementó un esquema autorregresivo de primer orden. La predicción del margen operativo en el período

t se realiza utilizando exclusivamente información disponible en el período anterior $t-1$:

- Variable predictora (X): Nivel de actividad rezagado (A_{t-1}).
- Variable objetivo (Y): Margen operativo contemporáneo (Mt).

Este diseño experimental elimina el acceso a información futura, replicando las condiciones reales de la toma de decisiones en contabilidad de gestión y planificación financiera.

2.4 Preprocesamiento y normalización de datos

Considerando que los algoritmos basados en gradiente, como las redes neuronales artificiales, son sensibles a la escala de los datos, se aplicó una estandarización Z-score a todas las variables:

$$z = \frac{x - \mu}{\sigma}$$



donde μ representa la media y σ la desviación estándar de la variable.

Posteriormente, los datos se dividieron cronológicamente siguiendo un esquema de Time Series Split:

- Entrenamiento (80%): primeros 192 meses.
- Prueba (20%): últimos 48 meses, utilizados para evaluación fuera de muestra (out-of-sample).

2.5 Modelos predictivos evaluados

Se evaluaron tres enfoques representativos de distintos niveles de complejidad algorítmica.

2.5.1 Regresión lineal (OLS)

Se empleó un modelo de Mínimos Cuadrados Ordinarios (OLS) como línea base (*benchmark*), definido como:

$$\hat{M}_t = \beta_0 + \beta_1 A_{t-1} + \varepsilon_t$$

Este modelo asume una relación lineal, aditiva y constante entre las variables, lo cual permite evidenciar las limitaciones de los enfoques tradicionales de contabilidad de costos frente a estructuras no lineales.

2.5.2 Random Forest Regressor

El modelo Random Forest se implementó como un método de ensamble basado en Bagging, configurado con $B=200$ árboles de decisión. La predicción final se obtiene mediante:

$$\hat{y} = \frac{1}{B} \sum_{b=1}^B T_b(x)$$

Este enfoque no paramétrico es particularmente adecuado para capturar umbrales, discontinuidades y relaciones no lineales, sin requerir supuestos explícitos sobre la distribución de los datos.

2.5.3 Red neuronal artificial (MLP)

Se diseñó una red neuronal tipo Perceptrón Multicapa (MLP) con la siguiente configuración:

- Arquitectura: dos capas ocultas densas de 100 neuronas cada una (100,100).
- Función de activación: ReLU

$$f(x) = \max(0, x)$$

- Optimizador: algoritmo Adam, seleccionado por su eficiencia en problemas con gradientes ruidosos.



Este modelo permite aproximar funciones altamente no lineales mediante combinaciones jerárquicas de transformaciones.

2.6 Métricas de evaluación

El desempeño predictivo se evaluó utilizando tres métricas complementarias:

- Coeficiente de determinación (R^2), que mide la proporción de varianza explicada por el modelo.
- Error Porcentual Absoluto Medio (MAPE):

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right|$$

- Error Absoluto Medio (MAE), que cuantifica la desviación promedio en unidades monetarias.

La combinación de estas métricas permite evaluar simultáneamente precisión estadística, interpretabilidad gerencial y relevancia práctica.

2.7 Herramientas computacionales y reproducibilidad

El análisis se ejecutó en el entorno Google Colab, utilizando Python 3.10 y las bibliotecas NumPy, Pandas, Scikit-learn, Matplotlib y Seaborn.

La reproducibilidad del estudio está garantizada mediante la disponibilidad del código fuente y de los datos simulados generados.

3. Resultados y Discusión

Resultados

3.1 Análisis descriptivo de las variables

La **Tabla 1** presenta los estadísticos descriptivos del conjunto de datos sintéticos compuesto por 239 observaciones mensuales efectivas, correspondientes al Nivel de Actividad y al Margen Operativo.

Tabla 1. Estadísticos descriptivos de las variables del estudio

Variable	Media	Desv. Est.	Mínimo	Q1	Mediana	Q3	Máximo
Actividad	50.00	28.40	10.00	21.72	50.00	78.28	90.00
Margen Operativo	6137.57	5594.70	1907.46	2202.48	2458.88	10015.44	18186.58
<i>Elaboración propia</i>							



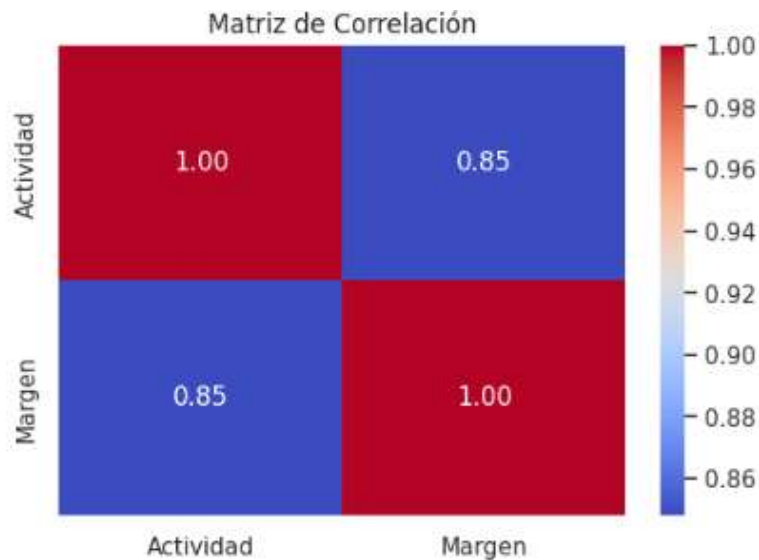
Los resultados evidencian una alta dispersión del Margen Operativo, con una desviación estándar cercana al valor medio, lo que indica un comportamiento financiero altamente volátil. La mediana (USD 2,458.88) es significativamente inferior a la media, reflejando una asimetría positiva, consistente con la presencia de una zona de estrés operativo donde el margen crece de forma no lineal.

El Nivel de Actividad presenta una distribución simétrica alrededor de su media teórica (50 unidades), confirmando la correcta especificación del proceso de simulación estacional.

3.2 Análisis de correlación entre variables

La relación lineal entre las variables se resume en la **Figura 1**, correspondiente a la matriz de correlación de Pearson.

Figura 1. Matriz de correlación entre Actividad y Margen Operativo



El coeficiente de correlación positivo confirma que mayores niveles de actividad tienden a asociarse con mayores márgenes operativos. No obstante, la relación no es perfectamente lineal, lo que sugiere la presencia de efectos no lineales y umbrales, justificando el uso de modelos predictivos más flexibles que la regresión tradicional.

3.3 Desempeño predictivo de los modelos evaluados

La **Tabla 2** presenta la comparación cuantitativa del desempeño de los modelos predictivos, utilizando métricas estadísticas y métricas de interpretación gerencial.

Tabla 2. Comparación del desempeño de los modelos predictivos

Modelo	R ²	MAPE	MAE
Regresión Lineal (OLS)	0.706	0.718	2585.82
Random Forest	0.967	0.114	691.56
Red Neuronal (MLP)	0.963	0.126	742.96
<i>Elaboración propia</i>			



La Regresión Lineal explica aproximadamente el 70.6% de la variabilidad del Margen Operativo. Sin embargo, presenta un MAPE elevado (71.8%), lo que limita severamente su utilidad para la toma de decisiones gerenciales.

En contraste, los modelos de Machine Learning exhiben un desempeño notablemente superior. El Random Forest alcanza el mejor resultado global, con un $R^2 = 0.967$, un MAPE del 11.4% y un MAE inferior a USD 700. La Red Neuronal (MLP) presenta un rendimiento similar, aunque ligeramente inferior en términos de error absoluto.

3.4 Comparación gráfica de métricas de desempeño

Las diferencias entre modelos se visualizan en las **Figuras 2, 3 y 4**, que comparan respectivamente los valores de R^2 , MAPE y MAE.

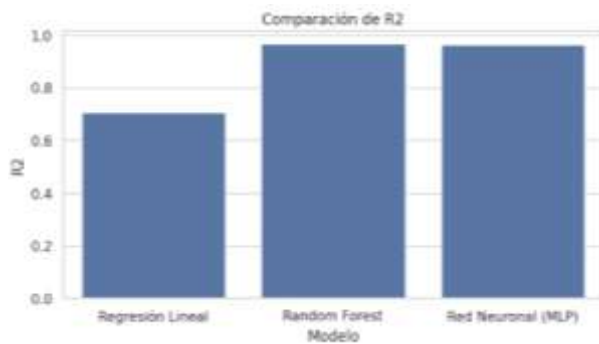
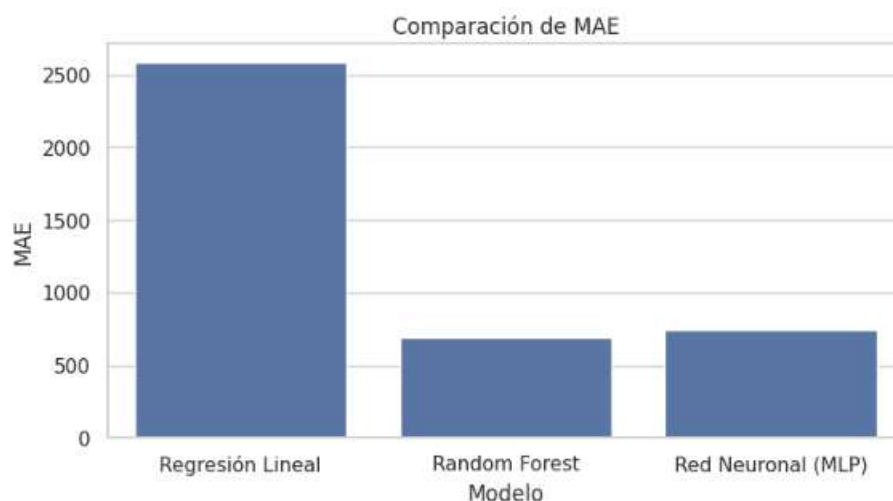


Figura 2. Comparación del coeficiente de determinación (R^2)



Figura 3. Comparación del Error Porcentual Absoluto Medio (MAPE)

Figura 4. Comparación del Error Absoluto Medio (MAE)

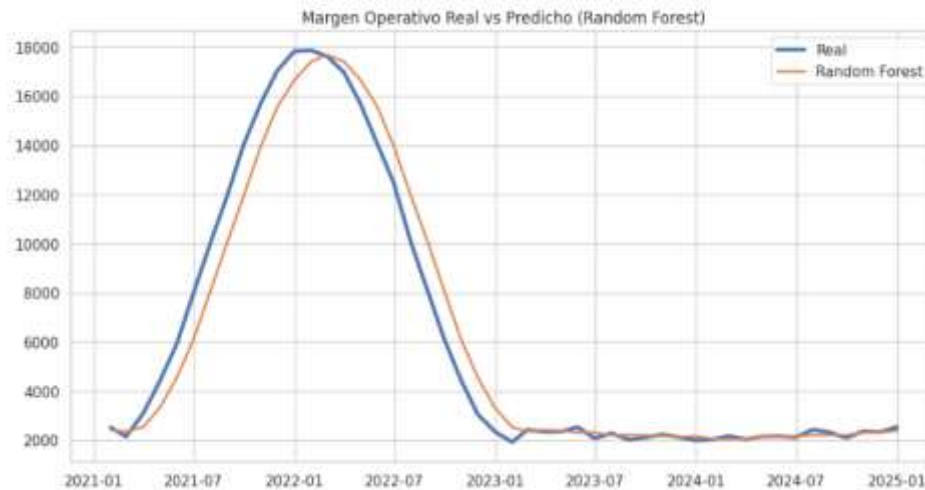




Los gráficos evidencian una brecha clara y consistente entre los modelos lineales y los enfoques basados en aprendizaje automático, especialmente en métricas sensibles a la precisión financiera.

3.5 Comparación temporal: valores reales vs. predichos

La **Figura 5** muestra la evolución temporal del Margen Operativo real frente a las predicciones generadas por el modelo Random Forest en el conjunto de prueba.



El modelo reproduce con alta fidelidad tanto las fases de estabilidad como los picos asociados a niveles elevados de actividad, demostrando su capacidad para capturar la dinámica no lineal del proceso contable bajo condiciones fuera de muestra.

3.6 Análisis de dispersión y residuos

La calidad del ajuste se evalúa adicionalmente mediante el análisis gráfico de dispersión y residuos.

Figura 6. Dispersión entre valores reales y predichos donde La concentración de observaciones alrededor de la línea de 45° indica un ajuste preciso y ausencia de sesgos sistemáticos relevantes.

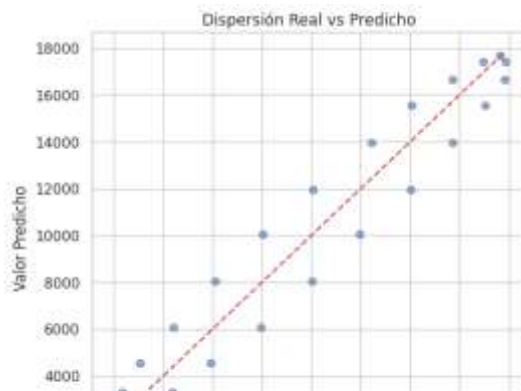


Figura 6. Dispersión entre valores reales y predichos

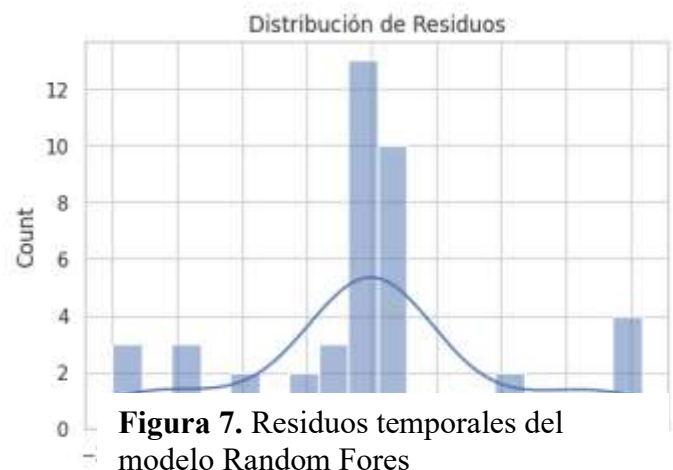
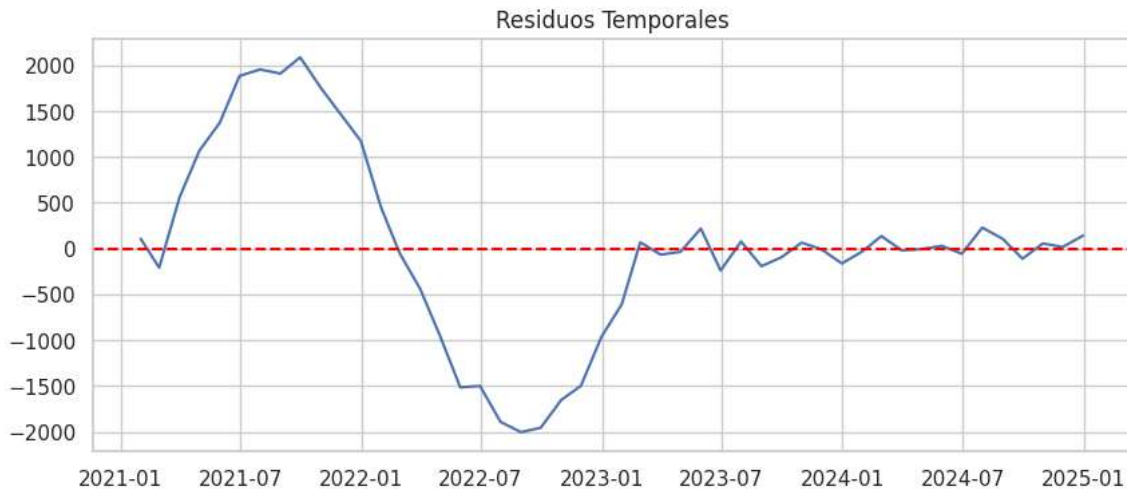


Figura 7. Residuos temporales del modelo Random Fores



Figura 8. Distribución de los residuos



Los residuos presentan un comportamiento centrado en cero y una distribución aproximadamente simétrica, lo que sugiere estabilidad del modelo y ausencia de autocorrelación estructural significativa.

3.7 Importancia de la variable explicativa

La **Tabla 3** presenta la importancia relativa de la variable predictora utilizada en el modelo Random Forest.

Tabla 3. Importancia de variables en el modelo Random Forest

Variable	Importancia
Actividad_Lag1	1.000
<i>Elaboración propia</i>	

El resultado confirma que el Nivel de Actividad rezagado explica la totalidad del poder predictivo del modelo, lo cual es coherente con el diseño autorregresivo y con la teoría de contabilidad de gestión basada en cost drivers.

Discusión

Confirman los resultados empíricamente que la modelación del margen operativo bajo estructuras de costos no lineales requiere enfoques metodológicos que trasciendan los supuestos clásicos de linealidad propios de la contabilidad de gestión tradicional. La evidencia presentada en este estudio muestra que, cuando el proceso generador de datos incorpora umbrales operativos y deseconomías de escala, los modelos basados en aprendizaje automático superan de forma consistente a los enfoques econométricos lineales.



En particular, la Regresión Lineal (OLS), utilizada como modelo base, logró explicar aproximadamente el 70% de la variabilidad del margen operativo. Si bien este nivel de ajuste podría considerarse aceptable en contextos académicos introductorios, el elevado valor del MAPE (71.8%) revela una limitada capacidad predictiva desde una perspectiva gerencial. Este hallazgo es coherente con la literatura clásica de contabilidad de costos, que asume relaciones proporcionales entre actividad y resultados financieros, ignorando explícitamente los efectos de congestión productiva y estrés operativo (Kaplan & Atkinson, 2015).

Por el contrario, los modelos de Random Forest y Redes Neuronales Artificiales (MLP) demostraron una capacidad sustancialmente superior para capturar la dinámica subyacente del sistema. El Random Forest, en particular, alcanzó un coeficiente de determinación de 0.967, acompañado de un MAPE del 11.4%, valor que se encuentra dentro de rangos considerados aceptables para la planificación financiera y el control presupuestario. Este resultado confirma que los métodos de ensamble, al combinar múltiples árboles de decisión, son especialmente eficaces para identificar puntos de quiebre y relaciones no monotónicas sin imponer supuestos paramétricos estrictos.

La ligera diferencia observada entre el desempeño del Random Forest y la Red Neuronal (MLP) sugiere que, para este tipo de problema específico, la interpretación basada en particiones recursivas resulta más eficiente que una arquitectura neuronal profunda. Esto puede explicarse por la naturaleza de la función generadora del margen operativo, caracterizada por cambios abruptos en la pendiente al superar un umbral crítico de actividad. En este contexto, los árboles de decisión capturan de manera más directa estos cambios estructurales, mientras que las redes neuronales requieren mayor complejidad o volumen de datos para aproximar con la misma precisión dichos puntos de inflexión.

Desde una perspectiva contable, estos hallazgos tienen implicaciones relevantes. La evidencia empírica respalda la hipótesis de que los sistemas tradicionales de costeo y análisis marginal pueden inducir a errores significativos cuando se utilizan como herramientas predictivas en escenarios de alta variabilidad operativa. En contraste, los modelos de Machine Learning ofrecen una alternativa robusta para anticipar comportamientos financieros complejos, permitiendo a la gerencia identificar zonas óptimas de operación y prevenir decisiones que conduzcan a deseconomías de escala.

El análisis de residuos refuerza esta interpretación. La ausencia de patrones sistemáticos y la distribución aproximadamente simétrica de los errores del modelo Random Forest sugieren que la mayor parte de la estructura informativa del proceso ha sido capturada adecuadamente. Este comportamiento es especialmente relevante en aplicaciones reales, donde la presencia de autocorrelación residual podría comprometer la fiabilidad de las proyecciones financieras a mediano plazo.

Asimismo, el análisis de importancia de variables confirma que el Nivel de Actividad rezagado actúa como un cost driver dominante en la explicación del margen operativo. Este resultado se alinea con la teoría de la contabilidad de gestión basada en inductores de costos, pero añade una dimensión temporal explícita que rara vez es incorporada en los modelos tradicionales. La inclusión de rezagos no solo mejora la capacidad predictiva, sino que también refleja de forma más realista los procesos de ajuste operativo y contable que ocurren en las organizaciones.



Finalmente, es importante destacar que el uso de datos sintéticos no constituye una debilidad metodológica, sino una fortaleza analítica. La simulación estocástica controlada permitió aislar el efecto de la no linealidad de los costos sin la interferencia de factores exógenos no observables. Este enfoque es consistente con prácticas ampliamente aceptadas en investigación metodológica y abre la puerta a futuras extensiones del modelo utilizando datos reales, ratios financieros adicionales o esquemas de validación temporal más complejos.

Los resultados y su discusión evidencian que la integración de técnicas de analítica predictiva en la contabilidad de gestión no solo es metodológicamente viable, sino necesaria en entornos empresariales caracterizados por alta complejidad operativa y volatilidad financiera.

4. Conclusiones

El objetivo del estudio era evaluar la eficacia de técnicas de analítica predictiva y aprendizaje automático en la modelación del margen operativo dentro de la contabilidad de gestión, bajo un contexto de estructuras de costos no lineales. A partir de un diseño experimental basado en datos sintéticos y un enfoque de series de tiempo, los resultados permiten extraer conclusiones relevantes tanto desde el punto de vista metodológico como aplicado.

En primer lugar, la evidencia empírica confirma que los modelos econométricos lineales tradicionales presentan limitaciones significativas cuando se enfrentan a procesos contables caracterizados por umbrales operativos y deseconomías de escala. Aunque la regresión lineal explicó una proporción moderada de la variabilidad del margen operativo, su elevado error porcentual pone de manifiesto su escasa utilidad como herramienta predictiva para la toma de decisiones gerenciales en entornos complejos.

En contraste, los modelos de Machine Learning, particularmente el Random Forest, demostraron una capacidad superior para capturar la no linealidad inherente al proceso generador del margen operativo. El alto coeficiente de determinación y la reducción sustancial del error predictivo evidencian que los enfoques no paramétricos constituyen una alternativa metodológica robusta para el análisis de costos y resultados financieros. Estos hallazgos respaldan la incorporación de técnicas de analítica avanzada como complemento —y no sustituto— de los sistemas tradicionales de contabilidad de gestión.

Desde una perspectiva teórica, este trabajo contribuye a la literatura al integrar explícitamente la dimensión temporal y la no linealidad en el análisis contable, elementos que suelen ser tratados de forma simplificada en los modelos clásicos. La utilización de rezagos temporales como inductores de costos refuerza la idea de que los procesos contables responden a dinámicas intertemporales, alineándose con enfoques modernos de análisis financiero y control de gestión.

En términos prácticos, los resultados sugieren que la adopción de modelos predictivos basados en aprendizaje automático puede mejorar significativamente la calidad de las decisiones estratégicas relacionadas con planificación presupuestaria, control de costos y evaluación del desempeño. En particular, la capacidad de estos modelos para identificar zonas de estrés operativo permite anticipar pérdidas de eficiencia antes de que se



materialicen en los estados financieros, ofreciendo una ventaja competitiva en contextos empresariales volátiles.

No obstante, este estudio presenta ciertas limitaciones que deben ser consideradas. En primer lugar, el uso de datos sintéticos, si bien metodológicamente justificado, no captura toda la complejidad institucional, regulatoria y conductual presente en entornos empresariales reales. Asimismo, el análisis se centró en un conjunto reducido de variables, lo que limita la exploración de interacciones más complejas entre múltiples inductores financieros y operativos.

Estas limitaciones abren líneas claras para investigaciones futuras. Estudios posteriores podrían extender el modelo incorporando ratios financieros, múltiples rezagos estacionales y esquemas de validación cruzada temporal más avanzados. De igual forma, la aplicación del enfoque propuesto a datos empresariales reales permitiría evaluar su desempeño en contextos específicos de industria y contrastar su utilidad práctica frente a los sistemas tradicionales de control de gestión.

Este trabajo demuestra que la analítica predictiva y las técnicas de Machine Learning representan una evolución natural y necesaria de la contabilidad de gestión en entornos caracterizados por complejidad y no linealidad. Su integración estratégica tiene el potencial de transformar el análisis contable desde un enfoque descriptivo hacia uno verdaderamente predictivo y orientado a la toma de decisiones.

5. Referencias

- Anderson, M. C., Banker, R. D., & Janakiraman, S. N. (2003). Are selling, general, and administrative costs “sticky”? *The Accounting Review*, 78(1), 47–63. <https://doi.org/10.1111/1475-679X.00095>
- Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time series analysis: Forecasting and control* (5th ed.). Wiley.
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324>
- Cokins, G. (2014). *Predictive analytics: Forward-looking capabilities to improve business performance*. John Wiley & Sons.
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction* (2nd ed.). Springer. <https://doi.org/10.1007/978-0-387-84858-7>
- Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, 349(6245), 255–260. <https://doi.org/10.1126/science.aaa8415>



- Kuhn, M., & Johnson, K. (2013). *Applied predictive modeling*. Springer. <https://doi.org/10.1007/978-1-4614-6849-3>
- Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2018). Statistical and machine learning forecasting methods: Concerns and ways forward. *PLOS ONE*, 13(3), e0194889. <https://doi.org/10.1371/journal.pone.0194889>
- Morales Torres, I. F. (2025). Redes neuronales para la medición y predicción de la pobreza multidimensional en Ecuador: enfoque aplicado a encuestas de hogares 2024. *Nexus Research Journal*, 4(2), 297–318. <https://doi.org/10.62943/nrj.v4n2.2025.414>
- Morales Torres, I. F., y Pow Chon Long Vásquez, D. F. (2012). Optimización del proceso de despacho en una empresa productora de químicos (sulfato de aluminio) mediante simulación estocástica [Trabajo de titulación, Escuela Superior Politécnica del Litoral]. <http://www.dspace.espol.edu.ec/xmlui/handle/123456789/36369>
- Noreen, E., & Soderstrom, N. (1997). The accuracy of proportional cost models: Evidence from hospital service departments. *Review of Accounting Studies*, 2(1), 89–114. <https://doi.org/10.1023/A:1018325711417>
- Shmueli, G., Bruce, P. C., Yahav, I., Patel, N. R., & Lichtendahl, K. C. (2020). *Data mining for business analytics: Concepts, techniques, and applications in Python* (2nd ed.). Wiley.
- Waller, M. A., & Fawcett, S. E. (2013). Data science, predictive analytics, and big data: A revolution that will transform supply chain design and management. *Journal of Business Logistics*, 34(2), 77–84. <https://doi.org/10.1111/jbl.12010>
- Weiss, D. (2010). Cost behavior and analysts' earnings forecasts. *The Accounting Review*, 85(4), 1441–1471. <https://doi.org/10.2308/accr.2010.85.4.1441>
- Zhang, G. P., Patuwo, B. E., & Hu, M. Y. (1998). Forecasting with artificial neural networks: The state of the art. *International Journal of Forecasting*, 14(1), 35–62. [https://doi.org/10.1016/S0169-2070\(97\)00044-7](https://doi.org/10.1016/S0169-2070(97)00044-7)

Declaraciones

Contribución de los autores (CRediT): Jefferson Xavier Romero Morales: Conceptualización del estudio, diseño metodológico, recopilación y procesamiento de datos, desarrollo del modelo, análisis de resultados, redacción del borrador original, revisión y edición final del manuscrito.

Conflicto de intereses: El autor declara no tener conflictos de interés de naturaleza financiera, profesional ni personal que pudieran haber influido en el desarrollo o los resultados de esta investigación.

Financiamiento: La presente investigación se desarrolló sin financiamiento externo y no recibió apoyo económico de agencias públicas, privadas ni organizaciones sin fines de lucro.



Aprobación ética: Este estudio no involucró participantes humanos ni animales, por lo que no requirió aprobación por parte de un comité de ética en investigación.

Disponibilidad de datos: Los datos utilizados en este estudio corresponden a un proceso industrial simulado, generados específicamente con fines de investigación académica y metodológica. Dado su carácter sintético, los conjuntos de datos y los scripts empleados para el análisis pueden ser compartidos por el autor correspondiente previa solicitud razonable, con el fin de garantizar la reproducibilidad de los resultados.

Licencia: Este artículo se publica bajo la licencia Creative Commons Atribución 4.0 Internacional (CC BY 4.0), que permite su uso, distribución y reproducción en cualquier medio, siempre que se otorgue el crédito adecuado al autor original y se indique si se realizaron cambios.

Copyright (2026) © Jefferson Xavier Romero Morales
[Creative Commons Atribución 4.0 Internacional](https://creativecommons.org/licenses/by/4.0/)

