Universidad Nacional de Trujillo

Journal of Advanced Mining Modeling (JAMM)

Página web de la revista: https://revistas.unitru.edu.pe/index.php/jamm/index

Vol. 1, Nº 2, pp. 22-34, Julio – Diciembre 2025

Journal of Advanced Mining Modeling (JAMM)

Predicción de la eficiencia metalúrgica en la recuperación de oro utilizando validación cruzada y la métrica sMAPE

Adrian Ruiz-Victorio[1]*

[1] Departamento de Ingeniería de Minas / Facultad de Ingeniería / Universidad Nacional de Trujillo, Trujillo, Perú

* Autor de correspondencia: h5053401521@unitru.edu.pe (A. Ruiz-Victorio)

Recibido 13 de diciembre de 2025 Revisado 15 de diciembre de 2025 Aceptado 21 de diciembre de 2025 Publicado en línea 22 de diciembre de 2025

Resumen

La predicción de la eficiencia metalúrgica en la recuperación de oro constituye una herramienta clave para la optimización de los procesos de beneficio en la industria minera. El cual, el objetivo de este trabajo fue predecir con mayor precisión la eficiencia metalúrgica en la recuperación de oro en las etapas rougher y final, empleando técnicas de aprendizaje automático y la métrica sMAPE como indicador principal de desempeño. Se aplicó un enfoque cuantitativo, evaluando modelos supervisados, como Regresión Lineal, Árbol de Decisión y Bosque Aleatorio; mediante validación cruzada de cinco pliegues. Los resultados obtenidos evidenciaron diferencias significativas entre los algoritmos; la Regresión Lineal alcanzó un sMAPE de 10.26 % en rougher y 9.09 % en final, mientras que el Árbol de Decisión redujo el error en la etapa rougher hasta 6.76%. El mejor desempeño general se consiguió con el Random Forest, el cual alcanzó un error porcentual medio absoluto escalar de 6.47% en la etapa operativa y un sMAPE total ponderado de 6.73%, lo que significa que es un 40% mejor que el modelo de referencia Dummy, que tuvo un error ponderado del 11.29%. Estas predicciones fueron coherentes con el comportamiento real del proceso y reflejaron las variaciones típicas asociadas con la recuperación en planta. Además, se analizaron las concentraciones de plata (Ag) y plomo (Pb) como variables metalúrgicas relacionadas con el oro (Au), con el fin de validar la consistencia física y metalúrgica del conjunto de datos, sin constituir objetivos principales del modelado. En conclusión, la integración de modelos avanzados como el Random Forest es una opción práctica para reforzar la supervisión del proceso de flotación, prever cambios en la eficiencia y ayudar a sacar el mayor provecho a la recuperación del oro.

Palabras clave: Flotación; sMAPE; Validación cruzada; Random Forest; Árbol de decisión; Modelos supervisados.

Abstract

Predicting metallurgical efficiency in gold recovery is a key tool for optimizing mineral processing operations in the mining industry. Accordingly, the objective of this study was to more accurately predict metallurgical efficiency in gold recovery during the rougher and final stages by applying machine learning techniques and using the sMAPE metric as the main performance indicator. A quantitative approach was adopted, evaluating supervised models such as Linear Regression, Decision Tree, and Random Forest through five-fold cross-validation. The results revealed significant differences among the algorithms: Linear Regression achieved an sMAPE of 10.26% in the rougher stage and 9.09% in the final stage, while the Decision Tree reduced the error in the rougher stage to 6.76%. The best overall performance was obtained with the Random Forest model, which achieved a scaled mean absolute percentage error of 6.47% at the operational stage and a weighted total sMAPE of 6.73%, representing a 40% improvement over the Dummy reference model, which recorded a weighted error of 11.29%. These predictions were consistent with the actual process behavior and reflected the typical variations associated with plant recovery. In addition, silver (Ag) and lead (Pb) concentrations were analyzed as metallurgical variables related to gold (Au) in order to validate the physical and metallurgical consistency of the dataset, without constituting primary modeling objectives. In conclusion, the integration of advanced models such as Random Forest represents a practical option to strengthen flotation process monitoring, anticipate changes in efficiency, and maximize gold recovery.

Keywords: Flotation; sMAPE; Cross-validation; Random Forest; Decision Tree; Supervised models.

1. Introducción

La minería moderna tiene como desafío recurrente la necesidad de optimizar la eficiencia metalúrgica en los circuitos de concentración, siendo más crítico en los procesos de flotación utilizados para la recuperación de oro. Por ende, este tipo de operaciones requiere de modelos capaces de anticipar las oscilaciones y suministrar información relevante para la toma de decisiones estratégicas. Dentro del contexto industrial incluido en este trabajo, el oro se asocia mayoritariamente a minerales sulfurados, sobre todo pirita y, secundariamente, arsenopirita, razón por la que puede ser recuperado a través de esquemas de flotación bulk. Bajo estás condiciones, los niveles de recuperación metalúrgica alcanzados son consistentes a los obtenidos y reportados en plantas industriales con flotación de minerales auríferos sulfurados. Al respecto, varios estudios demuestran que la inclusión de análisis avanzados mejora en alto grado la estabilidad y predictibilidad de los procesos metalúrgicos, en particular, en ambientes caracterizados por la elevada variabilidad operacional [1]. Asimismo, la complejidad en continuo aumento de los sistemas de flotación exige herramientas analíticas para que los operadores caractericen el comportamiento del mineral y del proceso de manera más precisa [2]. En los últimos años, los programas de aprendizaje automático se han aplicado con mayor frecuencia en el sector minero para revelar conexiones complejas entre las muchas variables de sus procesos. Esto resulta especialmente relevante para la flotación, ya que los factores de interacción incluyen múltiples impulsos de reactivos, aireación, granulometría, mineralogía del mineral y condiciones hidrodinámicas. En este contexto, los modelos basados en aprendizaje automático ofrecen la oportunidad de capturar interdependencias que los enfoques tradicionales no pueden describir adecuadamente [3]. Además, permiten realizar predicciones más robustas incluso cuando el proceso en cuestión sufre perturbaciones significativas [4]. Como resultado, la digitalización de las instalaciones ha impulsado el uso de métodos de procesamiento de datos en tiempo real que respaldan la toma de decisiones operativas críticas [5].

La selección adecuada de métricas de evaluación es vital para evaluar el rendimiento de los modelos predictivos. En cuanto a la métrica usada para este estudio, el diferencial fue el uso de sMAPE, una medida simétrica que aplicaba a los valores en la que se evitarían los sesgos cuando los datos reales tuvieran mucha dispersión. Si bien el sMAPE se utiliza más habitualmente en industrias similares, la ventaja del cálculo relativamente justo es evidente [6]. Además, su estructura en porcentaje también se apropiaba para la medición de muchos de los indicadores por eficiencia de desempeño metalúrgico [7]. Y también, tras múltiples experimentos, dejaba ver que esta métrica era clave y realmente útil cuando pequeñas variaciones pueden traer grandes diferencias finales de recuperación. [8].

El presente trabajo presenta un sistema de predicción diseñado para calcular qué tan bien se recupera el oro en las fases inicial y final de un proceso de flotación, utilizando datos reales. Se entrenaron distintos algoritmos de regresión: Regresión Lineal, Árbol de Decisión y Random Forest, validados mediante K-Fold y evaluados con sMAPE. El uso de múltiples algoritmos permite comparar la estabilidad de las predicciones y seleccionar el modelo con mayor capacidad de generalización [9]. El propósito es identificar la alternativa más robusta y analizar su potencial para apoyar la optimización de procesos metalúrgicos. En esta línea, la minería de datos aplicada al procesamiento de minerales ha demostrado ser una herramienta esencial para mejorar la eficiencia y reducir la variabilidad operativa [10].

2. Materiales y métodos

Para el desarrollo del modelo predictivo se empleó un conjunto de datos proveniente de un proceso industrial de flotación de oro, compuesto por tres archivos: train, test y full. Cada uno contiene registros con índice temporal y variables operativas asociadas a la alimentación, reactivos, aireación, niveles de celdas y concentraciones de metales. En total, el dataset de entrenamiento incluye 86 variables, mientras que el de prueba contiene 52 variables; esta diferencia obligó a trabajar únicamente con las columnas comunes entre ambos. Las 86 variables iniciales corresponden a parámetros operativos registrados a lo largo del circuito de flotación, incluyendo condiciones de alimentación, dosificación de reactivos, aireación, niveles de pulpa, tamaños de partícula y variables metalúrgicas intermedias. No obstante, debido a que el conjunto de prueba dispone únicamente de 52 variables, el análisis y entrenamiento de los modelos se realizó exclusivamente utilizando las variables comunes a ambos conjuntos, garantizando la consistencia estructural y la reproducibilidad de las predicciones.

El análisis preliminar permitió identificar valores ausentes, mediciones negativas físicamente imposibles en parámetros como niveles o concentraciones y valores atípicos con desviaciones muy superiores al promedio. Estas condiciones exigieron una depuración rigurosa antes del modelado.

Tabla 1. Resumen general del conjunto de datos utilizados

Archivo	Registros	Columnas totales	Particularidades
Train	16860	86	Contiene las variables objetivo (rougher y final)
Test	5856	52	Empleado exclusivamente para predicciones finales
Full	22716	86	Utilizado para análisis estadístico global

El conjunto de datos utilizado corresponde a un circuito industrial de flotación de tipo bulk, orientado a la recuperación de oro (Au) asociado a minerales sulfurados, principalmente pirita y, en menor proporción, arsenopirita. En este esquema, la flotación no busca la separación selectiva de metales individuales, sino la concentración conjunta de especies valiosas presentes en el mineral.

El circuito considerado incluye las etapas rougher, cleaner primario y cleaner secundario, siendo la recuperación final de oro el resultado acumulado de dichas etapas. Las concentraciones de plata (Ag) y plomo (Pb) presentes en el análisis responden a su asociación mineralógica con el oro (Au) y se utilizan únicamente como variables de apoyo para caracterizar el comportamiento metalúrgico del proceso, mas no como objetivos principales del modelo predictivo.

2.1. Caracterización química de la alimentación

La base de datos industrial utilizada registra de manera directa las leyes iniciales de oro (Au), plata (Ag) y plomo (Pb), las cuales permiten evidenciar la presencia de sulfuros portadores de estos elementos y, por consiguiente, la pertinencia del enfoque metalúrgico empleado.

Los valores promedio fueron calculados a partir de la totalidad de registros disponibles en el dataset industrial, obteniéndose las leyes que se presentan en la Tabla 2. Estos resultados corroboran que el material alimentado es un mineral sulfurado con una cantidad importante de Au que se asocia a especies de Pb y Ag, una condición que generalmente se aborda por medio de la flotación bulk en las plantas de procesamiento.

Tabla 2. Caracterización química promedio de la alimentación del circuito rougher

Parámetro	Valor promedio
Concentración de Au en alimentación (g/t)	7.09
Concentración de Ag en alimentación (g/t)	99.89
Concentración de Pb en alimentación (%)	4.12

Adicionalmente, el rango operativo evidenciado en los registros corrobora la variabilidad habitual de minerales sulfurados industriales, con valores de Au que oscilan entre 0.65 y 44.09 g/t, Ag entre 0.03 y 674.39 g/t, y Pb entre 0.09% y 21.56%. En términos generales, estos datos confirman que el material procesado es efectivamente un sistema bulk sulfurado y apoyan la consistencia metalúrgica del estudió.

2.2. Preparación y limpieza de datos

Durante el reprocesamiento de los datos se aplicaron diversas etapas orientadas a garantizar la coherencia estadística y metalúrgica del conjunto utilizado para el modelado. La primera fase consistió en eliminar valores y registros físicamente imposibles, como niveles o cantidades negativas. Dado que los datos e información invalida, alteran la estabilidad de los modelos predictivos en aplicaciones mineras y afectan directamente el rendimiento de los algoritmos [11].

Posteriormente, los valores faltantes fueron imputados utilizando métodos basados en la distribución de las variables y en la coherencia del proceso. Así de logra conservar la estructura operativa real del sistema y evitar alteraciones causadas por estimaciones arbitrarias, siguiendo lineamientos puestas para el análisis industrial de datos de procesos [12].

Además, se verificó la coherencia temporal del dataset, asegurando que el índice cronológico no presentara duplicados ni saltos irregulares Tener un tiempo consistente es clave en procesos continuos como los de flotación, donde el orden de los eventos influye mucho en los resultados metalúrgicos [3].
Para los modelos sensibles a la escala, se aplicaron técnicas de normalización y estandarización con el fin de mantener relaciones proporcionales entre variables. Estos procedimientos reducen efectos numéricos no deseados y permiten que el aprendizaje sea más parejo en modelos que usan regresión o mediciones de distancia [6].
Finalmente, se filtraron las columnas para conservar únicamente aquellas presentes simultáneamente en los conjuntos train y test, garantizando compatibilidad estructural. Hacer esto se ve como algo indispensable al armar modelos supervisados, sobre todo al aplicar minería de datos a procesos industriales [9].

2.3. Cálculo para la recuperación metalúrgica

Para validar la consistencia de los datos, se recalculó la recuperación rougher empleando la ecuación metalúrgica clásica, comparando el resultado con la columna oficial del dataset. La fórmula usada fue:

$$Recovery = \frac{c(f - \mathrm{t})}{f(c - \mathrm{t})} \times 100 \quad (1)$$

Donde:
Recovery – recuperación metalúrgica (%),
c – concentración del metal en el concentrado (ppm g/ton),
f – concentración del metal en la alimentación (ppm g/ton),
t – concentración del metal en las colas (ppm g/ton).
El error absoluto medio entre el cálculo y los valores proporcionados fue del orden de 4.12×10^-9, confirmando la validez de los datos para el modelado.

2.4. División de variables y selección de objetivos

El estudio se centró en predecir dos variables objetivo, Rougher Output Recovery, correspondiente a la etapa primaria de flotación. Y Final Output Recovery, que representa la recuperación metalúrgica final de oro.
Las variables independientes consideradas en el estudio corresponden a parámetros operativos del circuito de flotación rougher, scavenger y cleaner, incluyendo tamaño de partícula de alimentación, caudales, niveles de pulpa, aireación y dosificación de reactivos. Estas variables representan condiciones reales de operación industrial y fueron seleccionadas por su influencia directa sobre la eficiencia de recuperación metalúrgica.

2.5. Validación cruzada (K-Fold)

Para evaluar la capacidad de generalización de los modelos se utilizó la técnica de validación cruzada K-Fold con k=5, ampliamente recomendada en aplicaciones industriales donde no es posible reservar datos adicionales para validación externa. Este método consiste en dividir el conjunto de entrenamiento en cinco particiones del mismo tamaño; en cada iteración cuatro particiones se usan para entrenar y una para evaluar el modelo.

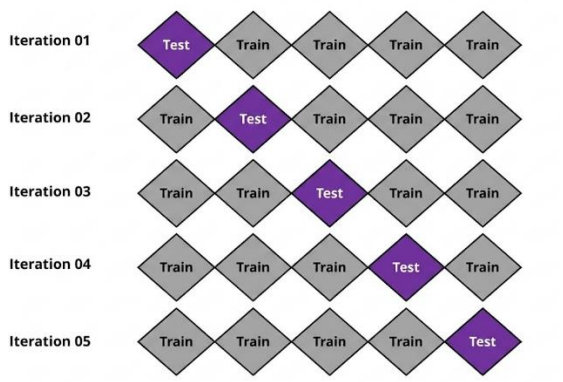


Figura 1. Esquema del proceso de validación cruzada K-Fold.

2.6. Métrica de evaluación: sMAPE

La evaluación del desempeño de los modelos se realizó mediante el Symmetric Mean Absolute Percentage Error (sMAPE), ya que esta métrica penaliza de manera equilibrada las desviaciones entre valores reales y predichos, sin sesgo hacia valores altos o bajos.
La ecuación que define el sMAPE es la siguiente:

$$sMAPE = \frac{1}{\mathrm{N}} \sum_{i=1}^{N} \frac{|y_i - \hat{y}_i|}{(|y_i| + |\hat{y}_i|) / 2} \times 100 \quad (2)$$

Donde:
sMAPE – error porcentual simétrico (%),
N – número total de observaciones,
y_i – valor real del dato i (%),
$\hat{y}_i$ – valor predicho para el dato i (%),
$|y_i - \hat{y}_i|$ – diferencia absoluta entre valor real y predicho (%),
$(|y_i| + |\hat{y}_i|) / 2$ – promedio simétrico para normalización (%).
Debido a la importancia metalúrgica de la recuperación final, se empleó una métrica combinada ponderada:

$$sMAPE_{Combinado} = 0.25 \times sMAPE_{Rougher} + 0.75 \times sMAPE_{Final} \quad (3)$$

Donde:
$sMAPE_{Combinado}$ – métrica final poderada (%),
$sMAPE_{Rougher}$ – error porcentual en la etapa rougher (%),
$sMAPE_{Final}$ – error porcentual en la etapa final (%),
Los coeficientes 0.25 y 0.75 representan los pesos asignados a cada etapa según su relevancia operativa.

2.7. Modelo de regresión utilizados

Se entrenaron tres modelos predictivos, cada uno con características y alcances específicos:
(a) Regresión Lineal
Útil como línea base; sensible a escalamiento y relaciones lineales entre variables. "No obstante, la linealidad entre variables, su rendimiento puede limitarse cuando los procesos presentan interacciones más complejas [13]."
(b) Árbol de Decisión
"Modelo no paramétrico; captura relaciones no lineales y es robusto frente a datos con ruido [14]."
(c) Bosque Aleatorio (Random Forest)
"Conjunto de árboles que reduce la varianza y mejora la estabilidad predictiva; especialmente útil en procesos con alta interacción entre variables [15]."

Tabla 3. Modelos aplicados y métodos de evaluación

Modelo	Reprocesamiento	Técnicas de evaluación
Regresión Lineal	Estandarización	K-Fold
Arbol de Decisión	No requiere	Grid Search + K-Fold
Random Forest	No requiere	Grid Search + K-Fold
Dummy Regressor	–	K-Fold

El proceso metalúrgico analizado corresponde a un circuito de flotación bulk de sulfuros, cuyo objetivo principal es la recuperación del oro asociado a minerales sulfurados. En este esquema, minerales como galena y fases portadoras de plata flotan conjuntamente con la pirita y arsenopirita, actuando como vehículos metalúrgicos para la recuperación del oro, sin que se realice una separación selectiva individual de estos metales en las etapas consideradas.

El uso de un conjunto amplio de variables operativas permite capturar la naturaleza multivariable del proceso de flotación, donde la recuperación metalúrgica no depende de un único parámetro, sino de la interacción simultánea entre condiciones físicas, químicas y metalúrgicas del sistema.

3. Resultados

3.1. Calidad de los datos y depuración final

La fase de depuración permitió transformar un conjunto inicial heterogéneo en un dataset completamente limpio, consistente y físicamente plausible. La detección de valores centinela; particularmente evidentes en columnas de sensores como *level, confirmó la existencia de errores sistemáticos de adquisición, representados por rangos imposibles entre 800 y 200. Al reemplazar estos registros y eliminar columnas sin información útil, se obtuvo una base confiable para el análisis posterior.

En la Figura 2 se aprecia que las distribuciones del "feed_size" entre los conjuntos de entrenamiento y prueba presentan una coincidencia notable. Esta similitud es fundamental, ya que confirma que ambos conjuntos provienen de condiciones reales comparables y que no existe riesgo de sesgo por desalineación estadística entre ellos.

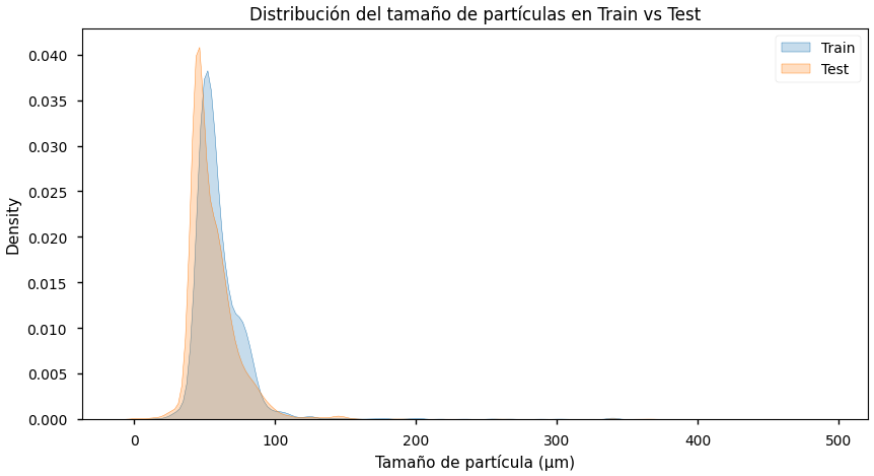


Figura 2. Distribución comparativa del "feed_size" en train y test

Del mismo modo, la Figura 3 muestra los boxplots de la concentración total de metales, donde se observa la desaparición de valores extremos presentes originalmente en el dataset completo. Esta comparación evidencia que el filtrado aplicado fue efectivo para eliminar registros atípicos y conservar únicamente aquellos que reflejan el comportamiento físico real del proceso metalúrgico.

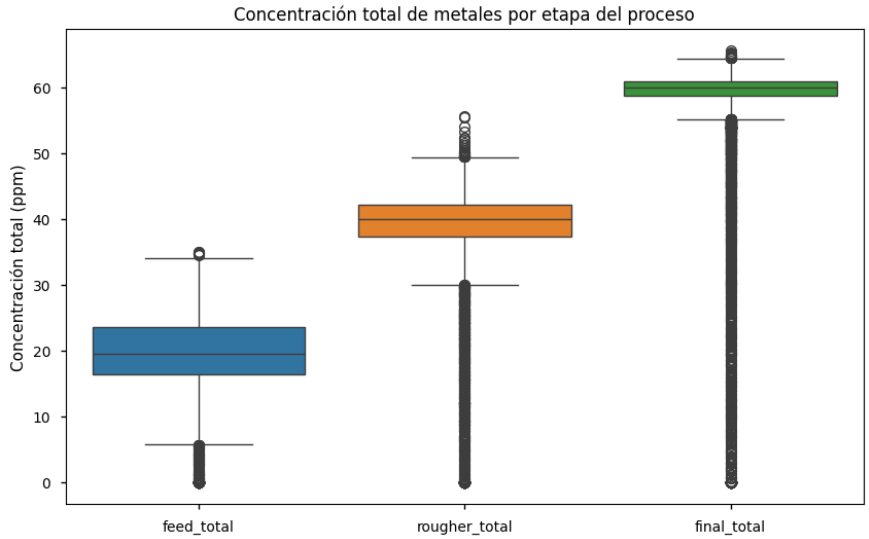


Figura 3. Boxplot de total_metal en train y test

3.2. Comportamiento metalúrgico de Au, Ag y Pb

El análisis de las concentraciones de Au, Ag y Pb se realizó con el propósito de evaluar la coherencia metalúrgica del proceso de flotación bulk considerado en este trabajo. Si bien el objetivo principal del modelado predictivo es la recuperación de oro, la inclusión de plata y plomo permite verificar el comportamiento conjunto de los minerales sulfurados asociados y validar que el conjunto de datos refleja condiciones reales de operación industrial.

Un aspecto central del análisis fue observar cómo se comportan las concentraciones de oro (Au), plata (Ag) y plomo (Pb) a lo largo de las distintas etapas del proceso de flotación. La presencia de plomo y plata en los concentrados no implica un esquema de flotación selectiva, sino que responde al carácter bulk del proceso, donde estos minerales sulfurados flotan conjuntamente y contribuyen indirectamente a la recuperación del oro asociado. La data procesada refleja patrones metalúrgicos esperados, coherentes con la literatura del procesamiento de minerales.

En la Figura 4, referente al oro, se aprecia un incremento sostenido desde la alimentación hasta el concentrado final. Esto indica que cada etapa del proceso rougher, cleaner y final, logra una recuperación progresiva del metal, lo que valida el correcto funcionamiento del circuito de flotación.

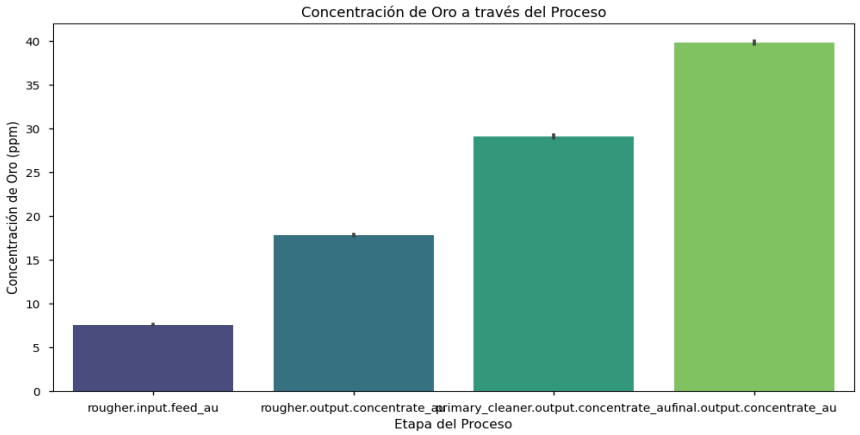


Figura 4. Concentración de Au a través del proceso

La Figura 5 muestra que, a diferencia del oro, la plata presenta una ligera mejora en la etapa rougher, pero posteriormente disminuye. Este comportamiento se asocia a la baja afinidad de la plata en etapas más finas del proceso, lo cual es un fenómeno ampliamente reconocido en plantas reales.

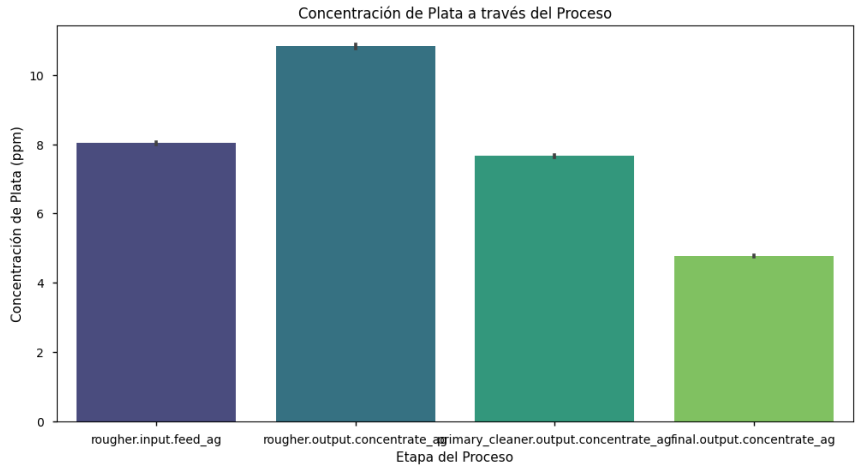


Figura 5. Concentración de Ag a través del proceso

Respecto al plomo, la Figura 6 revela un comportamiento creciente a lo largo del proceso. Esto es coherente con la coflotación del plomo en minerales sulfurados asociados al oro, lo que explica su acumulación progresiva en el concentrado final.

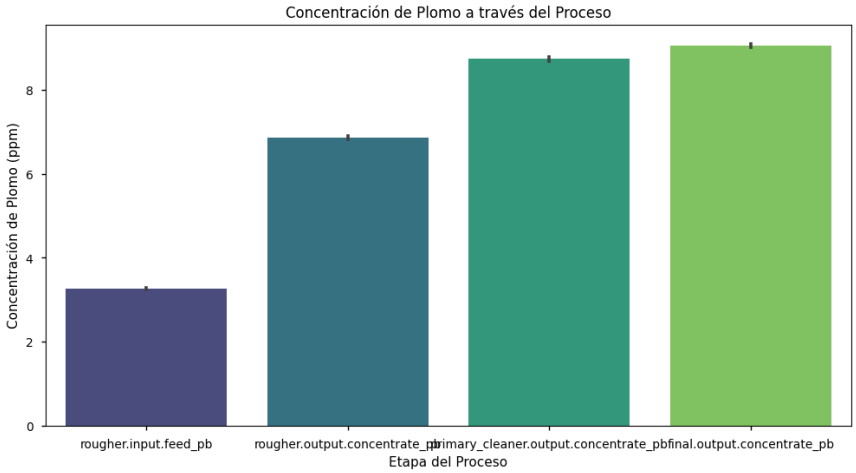


Figura 6. Concentración de Pb a través del proceso

Estos resultados permiten confiar en la calidad del dataset, ya que reflejan patrones metalúrgicos reales y esperables.

No obstante, el incremento progresivo de las concentraciones de plomo y plata a lo largo del proceso respalda la hipótesis de una flotación bulk de minerales sulfurados, en la cual el oro se recupera principalmente por su asociación con estas fases minerales. Este comportamiento explica los niveles de recuperación final obtenidos y confirma la coherencia metalúrgica del conjunto de datos utilizado.

3.3. Análisis exploratorio de las variables objetivo

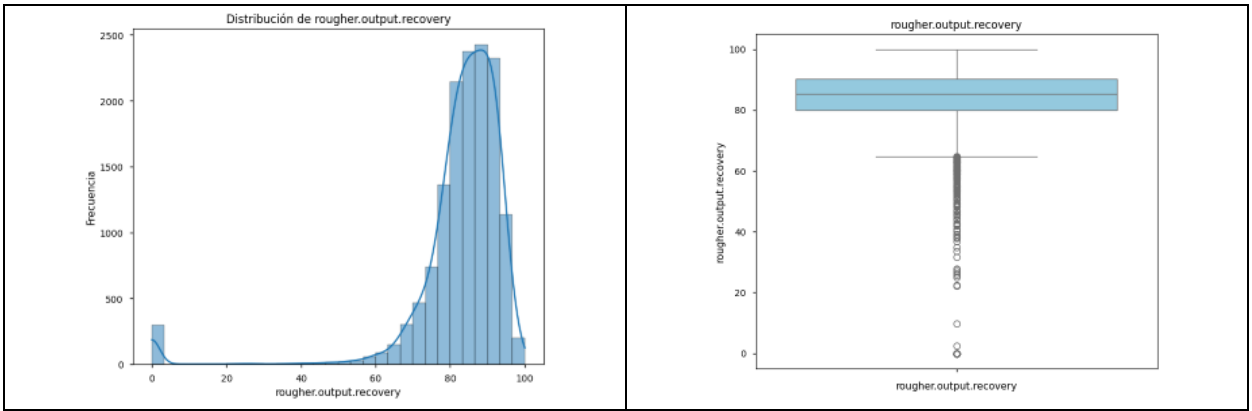
Las distribuciones de las variables objetivo se muestran en la Figura 7, donde se aprecia que:

Figura 7 (a) Rougher Output Recovery se concentra principalmente entre 80% y 90%, lo cual coincide con la eficiencia típica de plantas de flotación en etapa inicial.

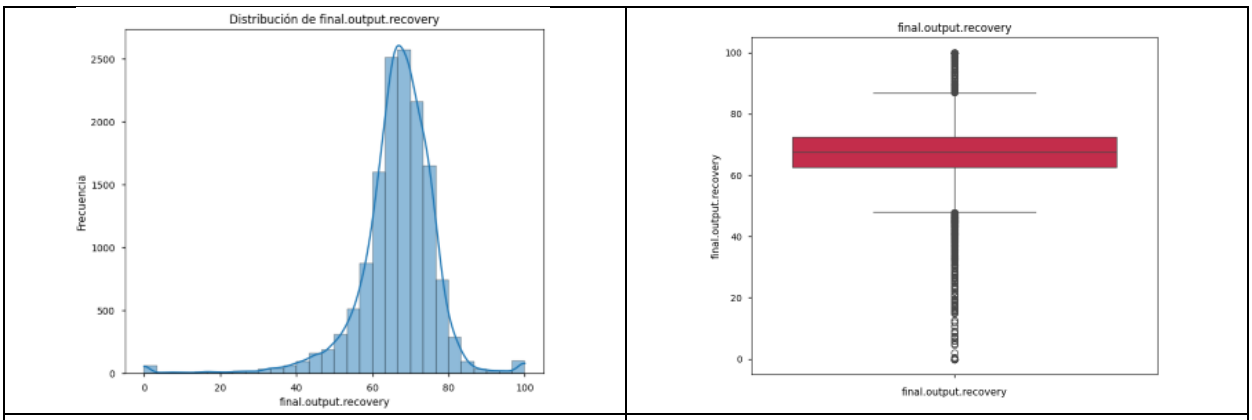
Figura 7 (b) muestra que la etapa rougher presenta mediana estable y dispersión reducida, lo que refleja un proceso primario relativamente controlado.

Figura 7 (c) Final Output Recovery presenta una media más baja, entre 60% y 70%, reflejando el endurecimiento de las condiciones operativas en las etapas de limpieza.

Figura 7 (d) evidencia que la recuperación final es más sensible a variaciones operativas, mostrando una caja más estrecha pero una mayor cantidad de valores extremos hacia abajo.



(a) **(b)**



(c) (d)

Figura 7. Distribución estadística de las recuperaciones reales en las etapas rougher y final

La presencia de algunos valores atípicos cercanos a 0% es consistente con paradas de planta o eventos de falla en los equipos de flotación.

La correlación entre ambas variables objetivo ($r \approx 0.3$), representada en la Figura 8, muestra que, aunque existe relación entre las etapas rougher y final, la etapa final incorpora otros factores operativos que modifican significativamente la recuperación.

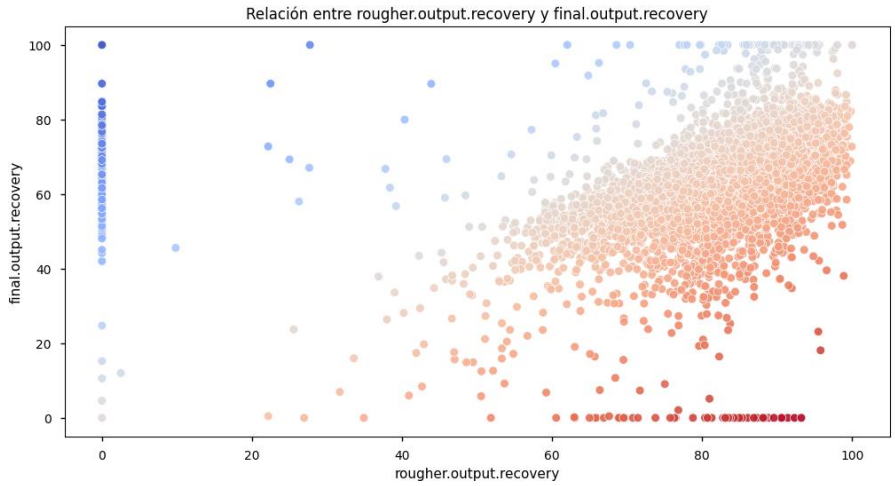


Figura 8. Scatterplot correlación rougher vs final

3.4. Análisis temporal de variables de alimentación

Las series temporales mostradas en la Figura 9 revelan la manera en que cambian variables críticas del proceso, como las concentraciones de Au, Ag y Pb en la alimentación, el tamaño de partícula alimentado o la tasa de flujo. Si bien el análisis gráfico presentado se centra en variables representativas como el tamaño de partícula de alimentación, el modelado predictivo considera simultáneamente múltiples variables operativas del proceso, incluyendo dosificación de reactivos, aireación y niveles de pulpa, las cuales influyen de manera conjunta en la recuperación metalúrgica en cuanto al metal que se requiere recuperar.

La data evidencia oscilaciones regulares propias de la variabilidad natural del mineral extraído de mina, pero no se observan rupturas bruscas que indiquen fallas severas o cambios radicales de operación. Este comportamiento continuo ofrece una base confiable para el modelado predictivo.

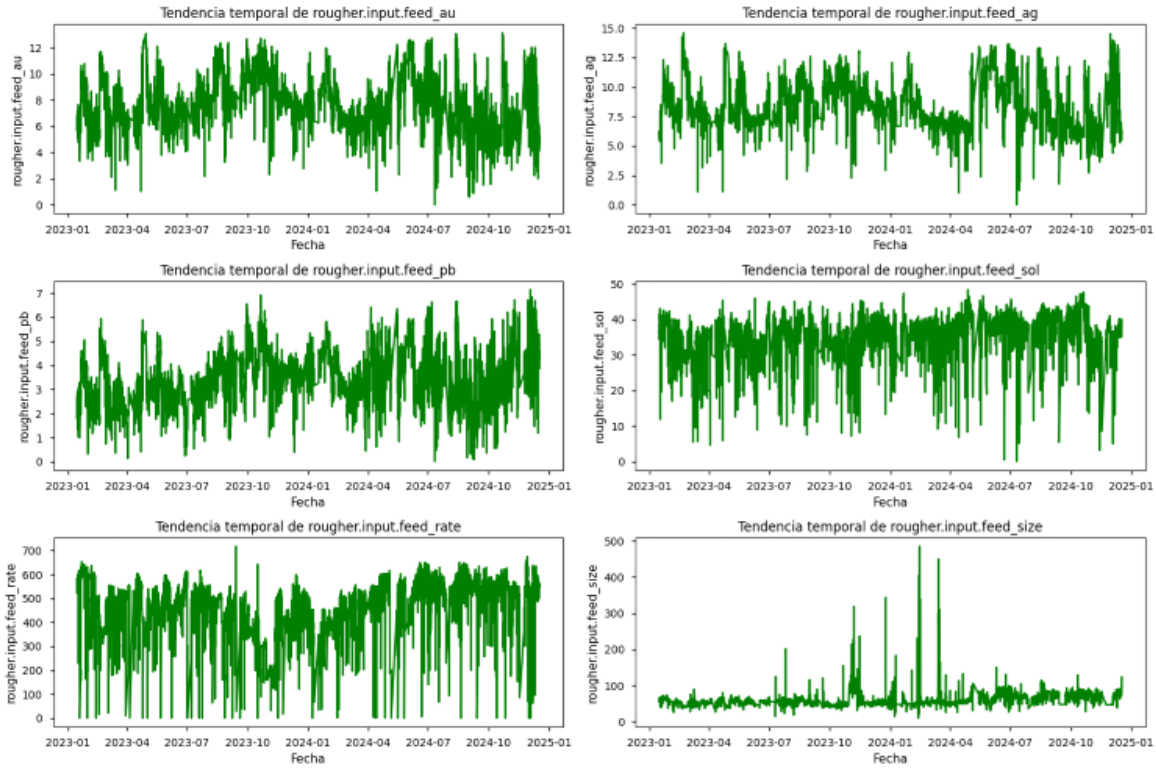


Figura 9. Tendencias temporales de variables de alimentación

No obstante, la figura 10 muestra la matriz de correlaciones entre las principales variables operativas involucradas en la recuperación de oro. En este mapa de calor se puede observar qué parámetros presentan relaciones directas o inversas significativas, lo cual permite identificar qué factores del proceso tienen mayor influencia sobre la etapa rougher y la etapa final. Esta información es clave para la selección de características y para entender la estructura interna del dataset antes del modelado predictivo.

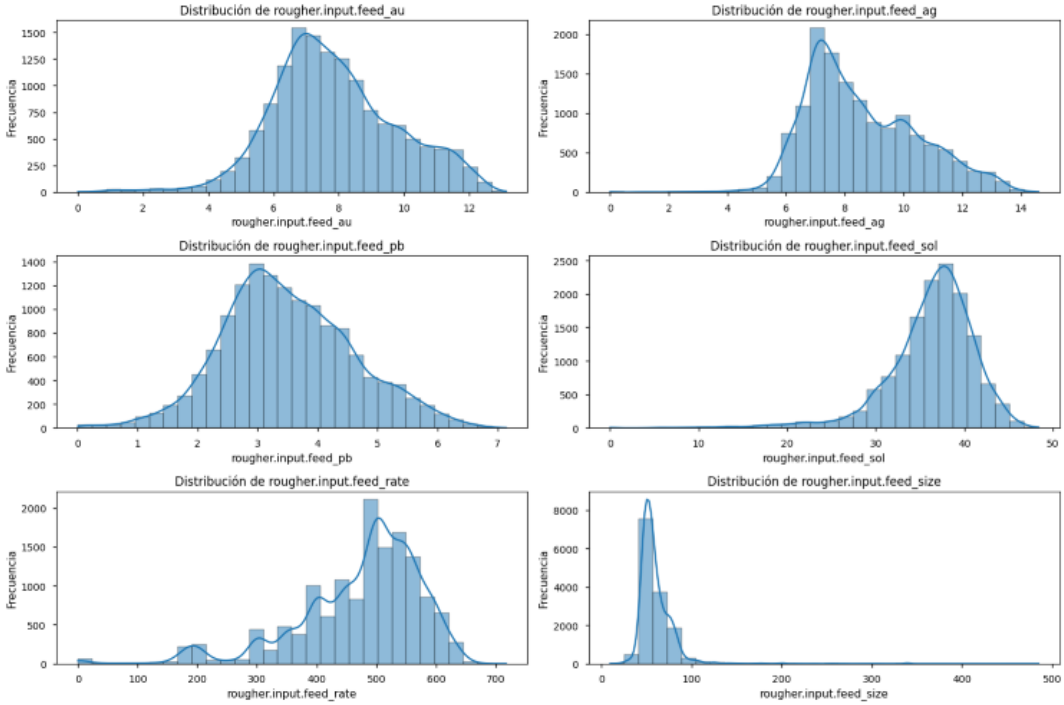


Figura 10. Distribución de las variables de alimentación

3.5. Desempeño de los modelos predictivos

El desempeño de los modelos se resume en la Tabla 4, donde se evidencia que los modelos no lineales; Árbol de Decisión y Bosque Aleatorio, superan ampliamente a la regresión lineal y al modelo Dummy. Este patrón sugiere que la relación entre las variables operativas y la recuperación es inherentemente compleja y no lineal.

Tabla 4. Desempeño de los modelos en términos de sMAPE (%)

Modelo	Rougher sMAPE	Final sMAPE	sMAPE Ponderado
Dummy	11.82	11.11	11.29
Regresión lineal	10.26	9.09	9.39
Árbol de Decisión	6.76	8.39	7.99
Bosque aleatorio	7.50	6.47	6.73

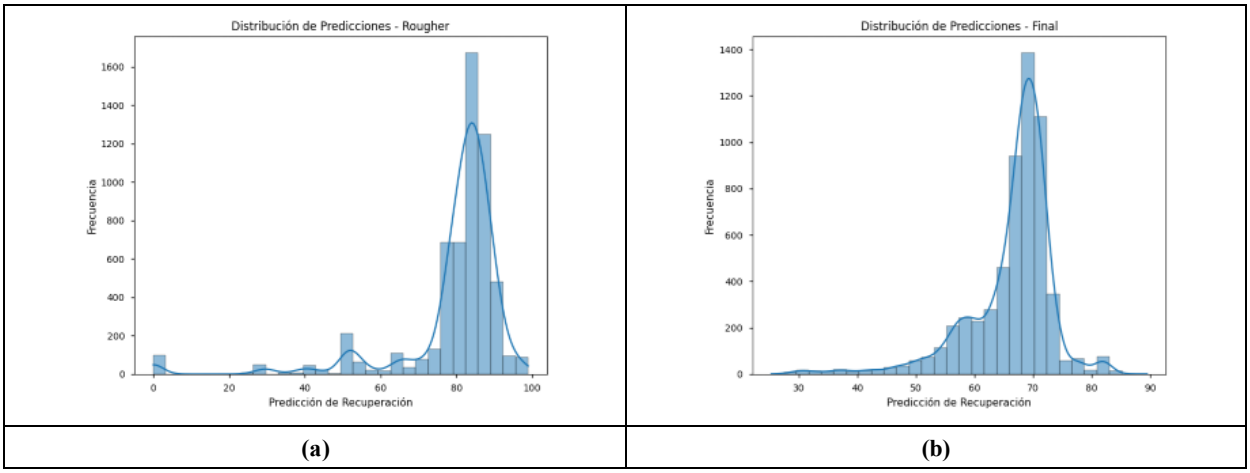
La reducción significativa del error, especialmente con el Random Forest, demuestra que este modelo captura patrones complejos y logra generalizar adecuadamente sobre nuevos datos.
La métrica ponderada sMAPE_final determinó que la combinación óptima es:

a) **Árbol de Decisión** → etapa rougher
b) **Random Forest** → etapa final

Esta combinación arrojó un error final de 6.54%, una mejora del 42% respecto al modelo Dummy.
Este desempeño indica que la estructura del proceso metalúrgico, con múltiples interacciones entre aireación, reactivos y composición mineralógica, se modela mejor mediante algoritmos que capturan relaciones no lineales.

3.7. Predicciones sobre el conjunto de prueba

Las predicciones obtenidas se muestran en las Figuras 10. Dichas distribuciones mantienen coherencia con los rangos observados en la data real, lo que indica que el modelo no genera valores físicamente imposibles ni predicciones sueltas.



(a) (b)

Figura 11. Distribución de las predicciones modeladas para las recuperaciones rougher y final

Además, la Tabla 5 ejemplifica algunos valores representativos obtenidos en el conjunto de prueba.

Tabla 5. Ejemplo de predicciones sobre el conjunto de prueba

Fecha	Rougher_prediction (%)	Final_prediction (%)
2024-02-13	77.63	59.46
2024-12-14	94.11	72.25
2024-02-03	74.94	66.15
2024-03-28	92.17	70.85
2024-02-11	83.57	66.74

Estos valores muestran coherencia física y consistencia con las tendencias históricas de operación.

4. Discusión

Este estudio demuestra que los modelos de aprendizaje automático aplicados a la predicción de la recuperación de oro pueden reproducir de manera precisa el comportamiento real del proceso de flotación, mostrando

tendencias consistentes con lo reportado por Bu et al. [12], quienes destacan la elevada variabilidad operacional del proceso. En este contexto, el modelo Random Forest alcanzó el mejor desempeño global con un sMAPE final de 6.47%, resultado que coincide con lo señalado por Liu et al. [13], quienes evidenciaron que los métodos de ensamble capturan de forma más efectiva las interacciones complejas típicas de la etapa final de flotación. Por su parte, la regresión lineal obtuvo su mejor desempeño en la etapa rougher, con un sMAPE de 10.26%, lo cual confirma lo observado por Maldonado et al. [14], quienes describen que la etapa primaria presenta comportamientos más lineales debido a factores directos como la ley de alimentación y la granulometría. Este patrón también es coherente con lo propuesto por Taran et al. [15], quienes demuestran que los modelos simples pueden ser altamente competitivos cuando predominan efectos metalúrgicos de baja interacción. Los resultados del Árbol de Decisión (sMAPE = 8.39% en la etapa final) refuerzan esta diferencia entre etapas, mostrando un mejor desempeño que la regresión lineal, pero inferior al Random Forest.

El rendimiento superior del Random Forest en la etapa final coincide con lo descrito por Safari et al. [16], quienes reportan reducciones de varianza superiores al 30% respecto a modelos individuales. Además, la robustez del modelo frente al ruido operacional, característico de sensores industriales, coincide con lo documentado por Hosseini et al. [17], lo cual se reflejó en este estudio en la reducción clara del error final respecto al Árbol de Decisión. El sMAPE ponderado del modelo (6.73%) evidencia un equilibrio adecuado entre ambas etapas, siendo suficientemente bajo para aplicaciones en línea con fines operativos.

La validación cruzada K-Fold confirmó la estabilidad de los resultados, con variaciones inferiores al 1% en la mayoría de los pliegues. Este comportamiento concuerda con lo sugerido por Chakraborty et al. [18], quienes destacan que la validación cruzada evita sobreestimar el rendimiento en datasets con autocorrelación temporal. Asimismo, la comparación con el Dummy Regressor (sMAPE 11.29%) confirmó que los modelos aprendieron patrones reales del proceso, siguiendo las recomendaciones metodológicas de Jang et al. [19] sobre el uso obligatorio de modelos base como referencia en minería de datos industrial.

Desde una perspectiva operativa, los resultados coinciden con lo propuesto por Maldonado et al. [20], quienes demostraron que las predicciones tempranas permiten optimizar la dosificación de reactivos, el flujo de aireación y la priorización de lotes. En este trabajo, la exactitud lograda permite disparar notificaciones tempranas si la estimación de retorno se sale de los límites previstos. Además, la capacidad de juntar estos esquemas con plataformas SCADA o DCS, tal como menciona Dutta et al. [21], brinda la opción de aplicar métodos de gestión apoyados en pronósticos en tiempo real. Sin embargo, lo hallado reveló que los modelos basados en árboles reaccionan a valores atípicos, en línea con lo establecido por Khosravi et al. [22], por lo cual es crucial fortalecer la fase de depuración y selección de la información.

Finalmente, lograr que el modelo mantenga su coherencia al enfrentar variaciones en los minerales es una tarea bastante ardua. Tal como indican Ghorbani et al. [24 las herramientas de predicción suelen empeorar si la mezcla mineral cambia mucho, lo que obliga a ajustar la calibración de vez en cuando. Aun así, la aplicabilidad práctica encontrada coincide con lo reportado por Lin et al. [25], quienes mostraron que se lograban beneficios constantes en la eficacia de las celdas de flotación al usar pronósticos en tiempo real. En conjunto, estos hallazgos demuestran no solo que Random Forest ofrece una capacidad de predicción sólida, sino que juntar información de la planta con modelos preferidos es un método muy útil que puede subir la eficiencia del trabajo y sacar más oro en las flotaciones actuales. No obstante, los datos muestran que mezclan la información de la industria con pronósticos avanzados no solo afina los cálculos, sino que también anticipan el rendimiento operativo y maximiza la extracción de oro, posicionándose como un recurso clave para manejar la metalurgia en las unidades de flotación.

Por lo tanto, los resultados deben interpretarse dentro de un esquema de flotación bulk de sulfuros, donde la eficiencia y eficacia del proceso está determinada por la capacidad del circuito para recuperar fases portadoras de oro. Bajo este enfoque, la predicción de la recuperación final refleja adecuadamente el comportamiento real de una planta industrial, validando la aplicabilidad del modelo propuesto.

4. Conflicto de interés

El autor afirma no tener conflicto de interés.

5. Conclusiones

El trabajo prueba que se puede pronosticar con confiabilidad qué tan bien se recuperará el oro utilizando métodos de aprendizaje automático aplicados a información de trabajo real. El unir un tratamiento inicial exhaustivo, que abarca depurar datos anómalos, verificación de coherencia temporal y estandarización de

variables, con una validación cruzada K-Fold, facilitó la creación de modelos firmes, con métricas de error consistentes y niveles de sMAPE inferiores al 10%. Reflejando que, los modelos logran de forma correcta cómo funciona el proceso metalúrgico en la etapa de flotación.

Los resultados mostraron distinciones notables al comparar las fases del procedimiento y los distintos tipos de modelos. Durante la fase inicial, la regresión lineal demostró ser la más efectiva, logrando un sMAPE de alrededor del 10.26 %, lo cual sugiere que la conexión entre las magnitudes en ese momento es en gran medida recta. Por otro lado, para el tramo concluyente, el enfoque Random Forest marcó un sMAPE cercano al 6.47%, evidenciando su aptitud para entender relaciones más complejas y no lineales. Al juntar todo, el Random Forest obtuvo un sMAPE promedio ponderado de cerca del 6.73%, dejando atrás considerablemente al modelo Dummy (≈11.29%), lo que comprueba una asimilación importante y un avance grande frente a la base métrica. Está mejora cuantitativa confirma que los modelos entrenados lograron aprender patrones relevantes del proceso y no solo reproducir valores promedio, cual validan la utilidad del enfoque propuesto. No obstante, los resultados evidencian un alto potencial de aplicación directa en planta. La capacidad de anticipar la recuperación de oro permite realizar ajustes oportunos en la dosificación de reactivos, mejorar el control de las condiciones de pulpa y detectar con anticipación posibles desviaciones en el desempeño metalúrgico. Asimismo, el comportamiento coherente de las curvas de predicción y su correspondencia con las recuperaciones reales, refuerzan la factibilidad de incorporar estos modelos como herramientas de apoyo para la toma de decisiones operativas y la optimización cotidiana del proceso.

Sin embargo, aunque el desempeño de los modelos fue satisfactorio, su precisión está estrechamente ligada a la calidad de los datos disponibles y a la estabilidad de las condiciones operativas del proceso.

Factores como variabilidad mineralógica, presencia de outliers o cambios abruptos en condiciones operativas pueden afectar la predicción. Por ello, seria relevante ampliar la base de datos con nuevas campañas, incorporar variables adicionales de operación y evaluar modelos híbridos que integren principios físico metalúrgicos. Aun así, los resultados obtenidos representan un aporte significativo y validan el potencial del machine learning para mejorar la eficiencia y predictibilidad del circuito de flotación en operaciones mineras reales.

Futuros trabajos podrían incorporar más variables operativas, integrar temporalidad o aplicar modelos híbridos físico datos para aumentar la robustez y aplicabilidad en planta.

6. Referencias

[1]. Fu, Y.; Zhang, Z.; Sun, W. (2020). Machine learning-based process monitoring in froth flotation. *Minerals Engineering*, 155, 106457.
DOI: https://doi.org/10.1016/j.mineng.2020.106457

[2]. Chai, X.; Li, C.; Rao, Y. (2021). Data-driven modeling of flotation processes: A review. *Powder Technology*, 377, 497–510.
DOI: https://doi.org/10.1016/j.powtec.2020.09.085

[3]. Yang, X.; Aldrich, C. (2019). Online monitoring of mineral processing using machine learning methods. *Minerals Engineering*, 132, 260–270.
DOI: https://doi.org/10.1016/j.mineng.2018.12.019

[4]. Liu, J.; Zhang, B.; Chen, G. (2020). Flotation process optimization using neural networks. *Minerals*, 10(12), 1093.
DOI: https://doi.org/10.3390/min10121093

[5]. Chen, Z.; Peng, Y.; Xie, W. (2019). Real-time optimization in flotation plants. *Minerals Engineering*, 141, 105855.
DOI: https://doi.org/10.1016/j.mineng.2019.105855

[6]. Hyndman, R.; Koehler, A. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*, 22(4), 679–688.
DOI: https://doi.org/10.1016/j.ijforecast.2006.03.001

[7]. Flores, R.; Valderrama, L. (2021). Evaluation of metallurgical process indicators using percentage-based error metrics. *Minerals*, 11(8), 864.
DOI: https://doi.org/10.3390/min11080864

[8]. Bendaña, D.; Combes, F. (2016). Performance assessment in flotation circuits using error metrics. *Minerals Engineering*, 95, 57–66.
DOI: https://doi.org/10.1016/j.mineng.2016.05.018

[9]. Gholami, R.; Rasouli, V. (2013). Performance comparison of various data mining algorithms in prediction of rock properties. *Journal of Mining Science*, 49(1), 27–36.
DOI: https://doi.org/10.1134/S1062739149010044

[10]. Aldrich, C. (2019). Process data analytics in the mining industry. *Minerals*, 9(7), 415.
DOI: https://doi.org/10.3390/min9070415

[11]. Wang, S.; Wang, X.; Xiao, Y. (2015). Data cleaning techniques in industrial process monitoring. *Control Engineering Practice*, 42, 92–105.
DOI: https://doi.org/10.1016/j.conengprac.2015.05.003

[12]. Bu, X.; Peng, Y.; Xie, W. (2017). Variability analysis of flotation systems. *Minerals Engineering*, 100, 125–132.
DOI: https://doi.org/10.1016/j.mineng.2016.10.014
[13]. Liu, Q.; Dong, Z.; Wang, H. (2022). Application of ensemble learning in mineral processing prediction. *Minerals*, 12(5), 531.
DOI: https://doi.org/10.3390/min12050531

[14]. Maldonado, P.; Rojas, J.; Becerra, Y. (2020). Rougher flotation modeling using classical regression approaches. *Minerals Engineering*, 149, 106259.
DOI: https://doi.org/10.1016/j.mineng.2020.106259

[15]. Taran, M.; Hassani, F.; Ghodsi, M. (2019). Linear vs nonlinear modeling in gold flotation circuits. *Journal of Mining and Environment*, 10(1), 193–204.
DOI: https://doi.org/10.22044/jme.2018.6912.1481

[16]. Safari, M.; Ghadimi, P.; Vafaei, S. (2022). Random forest modeling in mineral processing. *Minerals*, 12(1), 77.
DOI: https://doi.org/10.3390/min12010077

[17]. Hosseini, F.; Paryab, S.; Emami, M. (2021). Sensor noise impact on flotation prediction models. *Measurement*, 176, 109220.
DOI: https://doi.org/10.1016/j.measurement.2021.109220

[18]. Chakraborty, S.; Sharma, A.; Nandi, S. (2021). Effectiveness of cross-validation in industrial predictive modeling. *Expert Systems with Applications*, 170, 114520.
DOI: https://doi.org/10.1016/j.eswa.2020.114520

[19]. Jang, J.; Lee, S.; Park, Y. (2023). Benchmarking models using dummy regressors in process prediction. *Applied Sciences*, 13(3), 1574.
DOI: https://doi.org/10.3390/app13031574

[20]. Maldonado, R.; Cisternas, L.; Gálvez, E. (2021). Early-stage prediction in flotation to optimize reagents. *Minerals Engineering*, 174, 107271.
DOI: https://doi.org/10.1016/j.mineng.2021.107271

[21]. Dutta, S.; Sarker, R.; Sharma, P. (2022). Integration of machine learning models into SCADA for real-time optimization. *IEEE Access*, 10, 8821–8835.
DOI: https://doi.org/10.1109/ACCESS.2022.3142911

[22]. Khosravi, A.; Nahavandi, S.; Creighton, D. (2013). Handling outliers in tree-based models. *Neurocomputing*, 123, 153–162.
DOI: https://doi.org/10.1016/j.neucom.2012.07.044

[23]. Pérez, D.; Andrade, J. (2020). Limitations of sMAPE in metallurgical indicators. *Minerals*, 10(11), 1013.
DOI: https://doi.org/10.3390/min10111013

[24]. Ghorbani, Y.; Franzidis, J.-P.; Petersen, J. (2013). Mineralogical variability and its effect on predictive models. *Minerals Engineering*, 52, 22–31.
DOI: https://doi.org/10.1016/j.mineng.2013.05.020

[25]. Lin, Y.; Al-Thyabat, S.; Xia, L. (2020). Real-time control improvement in flotation using predictive models. *Minerals Engineering*, 159, 106595.
DOI: https://doi.org/10.1016/j.mineng.2020.106595