

**UNIVERSIDAD TECNICA FEDERICO SANTA MARIA
DEPARTAMENTO DE INGENIERIA METALURGICA
Y DE MATERIALES**



**Optimización del Rendimiento en la Molienda SAG a
través del Control de la Velocidad de los Feeders
Aplicando Analítica Avanzada Codelco División
Chuquicamata**

JOSEFINA IGNACIA CHACÓN BEALE

Trabajo de propuesta de memoria de titulación presentado en
conformidad a los requisitos y reglamentos requeridos para obtener el
título de Ingeniera Civil Metalúrgica

Profesor Patrocinante: Dr. Juan Patricio Ibáñez Rivera



CONSTANCIA DE VALIDACIÓN Y CONFIDENCIALIDAD DE MONOGRAFÍA A REPOSITORIO ACADÉMICO

1.- IDENTIFICACIÓN DEL TRABAJO ACADÉMICO

Tipo de monografía (marcar una opción): Memoria o trabajo de título Tesis de Postgrado

Título del trabajo: Optimización del Rendimiento en la Molienda SAG a través del Control de la Velocidad de los Feeders Aplicando Analítica Avanzada Codelco División Chuquicamata

Nombre del candidato(a): Josefina Ignacia Chacón Beale

Carrera / Grado: Ingeniería Civil Metalúrgica

Campus: Casa Central Departamento: Ingeniería Minas, Metalúrgicas y Materiales.

2.- VALIDACIÓN DEL PROFESOR GUÍA/DIRECTOR DE TESIS

Yo, Juan Patricio Ibáñez Rivera, en mi calidad de profesor patrocinante del trabajo académico mencionado anteriormente **DEJO CONSTANCIA** que:

- He revisado esta versión del documento y corresponde a la versión final aprobada del trabajo.
- El trabajo cumple con los requisitos académicos y de formato establecidos por la institución.

3.- EVALUACIÓN DE CONFIDENCIALIDAD POR PROPIEDAD INDUSTRIAL (marcar una opción)

El trabajo **NO contiene** información que amerite confidencialidad y puede ser publicado de inmediato en repositorio con acceso abierto.

El trabajo **CONTIENE** información con potenciales implicancias de propiedad industrial o intelectual y requiere un periodo de confidencialidad (**embargo**) por (**marcar una opción**):

6 meses 12 meses 2 años 3 años 5 años 10 años

Fundamentación de la necesidad de confidencialidad (obligatorio si se solicita embargo):

4.- FIRMAS

Profesor(a) guía o director(a) de memoria o tesis:

Fecha: _____ 02-04-2026 _____ Firma: _____

Estudiante o Candidato(a):

Fecha: _____ 01-04-2026 _____ Firma: _____

Este formulario debe ser insertado como página 2 de la memoria o tesis, completado y firmado por estudiante y profesor(a) antes de la entrega en portal PRISMA de Biblioteca USM.

DEDICATORIA

A mi madre.

AGRADECIMIENTOS

A mi familia. A mi madre, Edna, por quedarse leyendo hasta tarde aquellos libros que yo no alcanzaba a terminar en el colegio, permitiéndome dormir con la tranquilidad de que estaban completos. Estoy segura de que aún lo haría. Gracias por la educación, los valores y el cariño que me formaron en la persona que soy hoy. Aunque ahora esté un poco lejos, siempre la he sentido cerca.

A mi padre, Bernardo, quien me motivó a elegir esta carrera y me llevó en aquellos paseos a los que no podía negarme. Incluso cuando tenía entregas pendientes, pensaba: “eres reemplazable en cualquier otro lugar, menos en tu familia”.

A mi hermana, Dominique, con quien compartí gran parte de mi vida universitaria. Gracias por acompañarme en noches de trabajo, ayudarme cuando lo necesité y cuidar de nuestros gatos mientras yo estaba fuera de la ciudad.

A Pía, mi mejor amiga, compañera en cada celebración y testigo de cada rincón de mi camino universitario. Gracias por recibirme cada vez que iba a Santiago y recordarme, con cariño, mi deber de titularme.

A quienes compartían conmigo en la oficina del CEEMET, un espacio que nos permitió crear un ambiente agradable, donde incluso cuando no avanzábamos en nuestros quehaceres, encontrábamos un respiro en los crucigramas. Especialmente a la generación 2017, que me acogió en uno de mis momentos más difíciles; y a la generación 2019 —Rafael, Francisca S. y Antonia— quienes me mantuvieron al tanto de lo que ocurría mientras yo estaba en Calama y escucharon mis historias por esas tierras. A Diego R., mi fiel compañero de trabajos, con quien solo bajo presión lográbamos avanzar, razón por la cual recién ahora nos estamos titulando. Gracias por estar presente en cada etapa.

A Fernanda, de la generación 2018, a quien veo apenas dos veces al año —en los cumpleaños de cada una— pero que aun así siempre ha estado.

A los memoristas de la UdeC con quienes compartí la experiencia en Chuquicamata, especialmente José, con quien viví momentos gratos y quien incluso me motivó a ir al gimnasio para no caer en la locura.

A mi profesor patrocinador, JPI, por aceptarme como memorista incluso cuando cambié de tema más de una vez.

A los apoyos académicos —Claudia, Carolina, Macarena, Linda, Leticia, Luis, Felipe y Johnny— quienes fueron fundamentales en el desarrollo de cada proyecto de laboratorio. Siempre los percibí con disposición y buena voluntad, tanto durante mi carrera como en mi rol de ayudante.

Finalmente, a quienes ya no están, con la firme convicción de que cada persona llega a tu vida en el momento preciso, para enseñarte algo.

RESUMEN

La optimización de circuitos de molienda semi-autógena (SAG) constituye uno de los principales desafíos operacionales en plantas concentradoras, debido a la alta sensibilidad del proceso frente a variaciones en las características del mineral y en las condiciones de alimentación. En este contexto, el presente estudio propone una metodología basada en analítica avanzada para apoyar la toma de decisiones operacionales mediante la integración de modelos de aprendizaje automático y técnicas de optimización evolutiva.

Se desarrolló un modelo supervisado utilizando XGBoost para predecir el rendimiento del circuito de molienda en función de variables operacionales y características de la alimentación. Sobre este modelo predictivo se implementó un algoritmo de optimización Differential Evolution, cuyo objetivo es determinar las proporciones granulométricas de mineral en la alimentación que maximizan el throughput del circuito SAG. Las configuraciones granulométricas óptimas obtenidas por el algoritmo son posteriormente traducidas en acciones operacionales mediante el ajuste de la velocidad de los feeders, los cuales constituyen la principal variable manipulable para controlar las proporciones granulométricas que ingresan al molino.

Los resultados del estudio demuestran que la integración de modelos de aprendizaje automático con técnicas de optimización evolutiva permite capturar de manera efectiva la relación entre las condiciones de alimentación y el desempeño del circuito de molienda, identificando configuraciones operacionales que maximizan la capacidad de tratamiento. Bajo las condiciones evaluadas, la aplicación de la estrategia propuesta permitiría incrementar el throughput del circuito en aproximadamente un 3,7%, lo que se traduce en un aumento potencial cercano a 10 ktpa de cobre fino.

Estos resultados evidencian el valor de la analítica avanzada como herramienta complementaria al criterio experto en la operación de plantas concentradoras, permitiendo transformar datos históricos de operación en recomendaciones accionables para mejorar la eficiencia operativa y la productividad del proceso. Se recomienda evaluar su implementación en sistemas de apoyo operacional o control avanzado, con el fin de validar su impacto en condiciones de operación industrial.

ABSTRACT

Optimizing semi-autogenous grinding (SAG) circuits remains one of the main operational challenges in concentrator plants due to the high sensitivity of the process to variations in ore characteristics and feed conditions. In this context, this study proposes a methodology based on advanced analytics to support operational decision-making through the integration of machine learning models and evolutionary optimization techniques.

A supervised model based on XGBoost was developed to predict grinding circuit throughput as a function of operational variables and feed characteristics. Based on this predictive model, a Differential Evolution optimization algorithm was implemented to determine the feed size distribution proportions that maximize SAG circuit throughput. The optimal granulometric configurations identified by the algorithm are subsequently translated into operational actions by adjusting the speed of the feeders, which constitute the main manipulable variable for controlling the proportions of each size fraction entering the mill.

The results of this study demonstrate that integrating machine learning models with evolutionary optimization techniques effectively captures the relationship between feed conditions and grinding circuit performance, enabling the identification of operational configurations that maximize processing capacity. Under the evaluated conditions, the proposed strategy could increase circuit throughput by approximately 3.7%, which translates into a potential increase of nearly 10 ktpa of fine copper production.

These findings highlight the value of advanced analytics as a complementary tool to expert judgment in concentrator plant operations, enabling the transformation of historical operational data into actionable recommendations to improve operational efficiency and process productivity. It is recommended to evaluate its implementation within operational decision-support systems or advanced process control frameworks in order to validate its impact under industrial operating conditions.

TABLA DE CONTENIDOS

1. DEDICATORIA.....	i
2. AGRADECIMIENTOS	ii
3. RESUMEN	iv
4. ABSTRACT	v
5. TABLA DE CONTENIDOS	vi
6. CAPÍTULO I: INTRODUCCIÓN	1
7. CAPÍTULO II: ANTECEDENTES, MARCO TEÓRICO Y ESTADO DEL ARTE.....	3
8. 1. ANTECEDENTES DE LA EMPRESA	3
9. 1.1 Desempeño y Producción Global	3
10. 1.2 Productos y Divisiones	4
11. 1.3 División Chuquicamata: Historia y Transformación.....	4
12. 1.3.1 Descripción del Procesos Productivo de la División.....	5
13. 2. MARCO TEÓRICO.....	7
14. 2.1 Conceptos de molienda SAG y alimentación por granulometría	7
15. 2.2 Mapa de Calor de Correlación de Pearson	7
16. 2.3 Bases de modelado predictivo con XGBoost	8
17. 2.4 Indicadores de Ajuste del Modelo Predictivo	9
18. 2.4.1 Error Cuadrático Medio de la Raíz (RMSE)	9
19. 2.4.2 Coeficiente de Determinación (R^2).....	10
20. 2.4.3 Error Absoluto Medio (MAE).....	10
21. 2.4.4 Error Porcentual Absoluto Medio (MAPE).....	11
22. 2.5 Fundamento de interpretación mediante SHAP	11
23. 2.6 Optimización Basada en el Modelo con Differential Evolution.....	12
24. 3. ESTADO DEL ARTE	17
25. CAPÍTULO III: PROBLEMÁTICA, HIPÓTESIS, ALCANCE, OBJETIVOS.....	21
26. 1. PROBLEMÁTICA	21
27. 2. HIPÓTESIS	22
28. 3. ALCANCE	22
29. 4.1 Objetivo General.....	23

30. 4.2 Objetivos Específicos	23
31. CAPÍTULO IV: METODOLOGÍA DE TRABAJO.....	24
32. 1.1 Disponibilización de Información	24
33. CAPÍTULO V: RESULTADOS Y DISCUSIÓN	31
34. CAPÍTULO VI: CONCLUSIONES Y RECOMENDACIONES	48
35. REFERENCIAS.....	51
36. ANEXOS	54
Figura 1. Mapa de acceso a Mina Chuquicamata[6].	4
Figura 2. <i>Shapley</i> para un modelo <i>XGBoost</i> aplicado en altura de <i>lifter</i> (modificado)[10].	12
Figura 3. Diagrama de caja de la distribución granulométrica alimentada al SAG 16.....	31
Figura 4. Diagrama de caja de la distribución granulométrica alimentada al SAG 17.....	32
Figura 5. Mapa de calor de correlaciones de Pearson para modelamiento del molino SAG 16.	35
Figura 6. Gráfica de dispersión de rendimiento Real vs Predicho del SAG 16.	38
Figura 7. Gráfica de comparación temporal del rendimiento real vs predicho del SAG 16.	39
Figura 8. Gráfica de residuos en el tiempo del SAG 16.....	39
Figura 9. Gráfica de dispersión de rendimiento Real vs Predicho del SAG 17.	40
Figura 10. Gráfica de comparación temporal del rendimiento real vs predicho del SAG 17.	41
Figura 11. Gráfica de residuos en el tiempo del SAG 17.....	41
Figura 12. Características más importantes. a) SHAP de SAG 16. b) SHAP de SAG 17.	42
Figura 13. Resultados de modelo de optimización DE. a) gráfica de alimentación predicha vs actual. b) gráfica de distribución de alimentación total.....	44
Tabla 2. Hiperparámetros para el modelo <i>XGBoost</i> de los molinos SAG 16 y 17.	36
Tabla 3. Métricas de rendimiento promedio del modelo <i>XGBoost</i> en los conjuntos de prueba de los molinos SAG.	37
Tabla 4. Impacto productivo de la optimización del rendimiento SAG mediante Machine Learning.....	47

CAPÍTULO I: INTRODUCCIÓN

El procesamiento de minerales, particularmente la etapa de conminución es el proceso de mayor intensidad energética en la minería. Dentro de esta etapa, el Molino Semi-Autógeno de Molienda (SAG) es la pieza de equipo que representa la mayor complejidad y el desafío más significativo para la eficiencia operacional. La alta variabilidad en las propiedades del mineral, especialmente en la granulometría de alimentación, genera una inestabilidad que impacta directamente en la potencia consumida y el rendimiento (t/h) [1].

La aplicación de técnicas avanzadas de Machine Learning (ML) y Optimización Evolutiva se ha consolidado como una estrategia robusta para modelar la dinámica no lineal de estos circuitos y, consecuentemente, optimizar los parámetros operacionales clave, como la composición granulométrica de la carga, con el objetivo primordial de maximizar el tonelaje procesado [2], [3]

El escenario global actual en la industria del cobre está marcado por un alza sostenida en la demanda de este metal, impulsada fundamentalmente por la transición energética y el desarrollo de tecnologías verdes, el crecimiento explosivo de la inteligencia artificial, la automatización, y los procesos de industrialización en economías emergentes. En este contexto, Codelco, líder en la producción nacional de este metal, enfrenta la presión de satisfacer esta demanda creciente.

Pese a esta exigencia, la División Chuquicamata opera con infraestructura de la planta concentradora que data de 1952, lo que plantea un desafío estructural para lograr los aumentos de producción necesarios a menores costos. Específicamente, el *stockpile* que alimenta la molienda SAG presenta una problemática crítica: la segregación de carga por granulometría, que provoca una alimentación no homogénea en los distintos *feeders* y obliga a una preferencia operativa por la alimentación de partículas finas para no comprometer el rendimiento.

Si bien se han explorado soluciones de intervención física en la infraestructura para modificar la trayectoria de deposición del mineral, favoreciendo una acumulación más vertical y, por ende, una distribución granulométrica más uniforme, la alta velocidad del mineral transportado compromete la integridad estructural del sistema de descarga. Las soluciones de refuerzo estructural son, en muchos casos, inviables económicamente.

Frente a la inviabilidad o alto costo de las modificaciones físicas, la analítica avanzada emerge como una alternativa eficiente. Actualmente, la operación de los *feeders* en

Chuquicamata carece de un proceso estandarizado y depende del criterio experto, lo que genera una alta variabilidad y dificulta el establecimiento de una estrategia operativa general. Además, existe una falta de consenso sobre las variables de entrada más pertinentes para construir un modelo predictivo del rendimiento de molienda. Este trabajo de memoria busca erradicar esta brecha de conocimiento y operativa al desarrollar una estrategia basada en análisis de datos que permita estandarizar y optimizar la alimentación.

El alcance de este estudio se limita a la construcción de un modelo predictivo utilizando *XGBoost*, el cual será interpretado mediante SHAP y optimizado con *Differential Evolution*, basándose en datos históricos de molienda SAG de un periodo de dos años. Los resultados, por tanto, serán válidos dentro de los rangos operacionales observados en dicho periodo.

El presente trabajo de memoria tiene como fin proponer una metodología de control operativo más eficiente para la molienda SAG de Codelco División Chuquicamata a través de analítica avanzada.

En síntesis, la revisión teórica evidencia que el proceso de molienda y alimentación de minerales (fino, intermedio, grueso) puede ser sustentado mediante los fundamentos de la conminución, que proveen una base para comprender cómo el tamaño de alimentación interviene en el rendimiento de molienda. Asimismo, las técnicas de predicción, interpretación mediante SHAP y optimización con *Evolución Diferencial* ofrecen un marco robusto para el desarrollo del modelo propuesto: predicción de condiciones operacionales, interpretar variables clave y recomendar la configuración granulométrica optimizada a alimentar para maximizar el rendimiento. Esta integración teórico-metodológica permite erradicar la brecha detectada entre el entendimiento clásico de molienda y la implementación avanzada de modelos de recomendación operativa en molinos SAG.

CAPÍTULO II: ANTECEDENTES, MARCO TEÓRICO Y ESTADO DEL ARTE

1. ANTECEDENTES DE LA EMPRESA

CODELCO (Corporación Nacional del Cobre de Chile) es una empresa autónoma propiedad del Estado chileno, formalizada el 1 de abril de 1976. Como empresa estatal, su misión es maximizar su valor económico, ambiental y social, contribuyendo de manera sostenible al Estado a través de la minería del cobre y sus subproductos, al cual dispone del 100% de los excedentes. La Casa Matriz, ubicada en Santiago, coordina la estrategia corporativa, que es desarrollada por un directorio formado por nueve integrantes y el presidente ejecutivo de la empresa. Las instalaciones operacionales se encuentran estratégicamente alejadas de la urbanización, con acceso restringido únicamente a su personal.

1.1 Desempeño y Producción Global

A nivel global, CODELCO es un actor clave, conteniendo aproximadamente el 6% de las reservas mundiales de cobre. En el año 2024, la empresa se mantuvo en el tercer puesto de la producción mundial de cobre, detrás de Freeport-McMoRan (primer lugar) y BHP (segundo lugar) [4].

Durante 2024, la corporación produjo 1.328.425 toneladas de cobre fino, lo que representó el 5,8% de la producción mundial de cobre de mina y el 24,4% de la producción nacional (sin incluir su participación en Minera El Abra y Anglo American Sur). La producción de molibdeno alcanzó las 16.200 tmf. El desempeño económico resultó en un EBITDA de US\$ 5.439 millones, con un aporte al Fisco de US\$ 1.534 millones (un 8% más que en 2023). La empresa logró una reducción del 2% tanto en los costos directos (C1) como en el neto a cátodo (C3), debido principalmente a un tipo de cambio favorable y menores costos de insumos [5].

Para el presente año, se espera que la producción de cobre se ubique entre 1.370 a 1.400 Ktmf, con un costo directo (C1) proyectado entre 195 a 198 c/lb y una inversión de capital (Cápex) entre US\$4.600 a US\$5.600 millones [5].

1.2 Productos y Divisiones

CODELCO produce y comercializa una amplia gama de productos. Los productos refinados corresponden a cátodos de cobre con 99,99% de pureza obtenidos por electrorrefinación y electroobtención. Los no refinados incluyen concentrado de cobre, concentrado de cobre tostados, ánodos y cobre blíster. Entre los subproductos se destacan el molibdeno, ácido sulfúrico, oro, plata y renio. Además, la empresa manufactura alambroón en Alemania a partir de sus cátodos de cobre.

La producción se obtiene a través de ocho divisiones operativas: Radomiro Tomic, Chuquicamata, Ministro Hales, Gabriela Mistral, Salvador, Ventanas, Andina y El Teniente.

1.3 División Chuquicamata: Historia y Transformación

La División Chuquicamata, ubicada a 15 km al norte de Calama en la Región de Antofagasta a 2.870 m s. n. m. donde se puede revisar el acceso en la Figura 1, es una de las operaciones más históricas de CODELCO. Su producción de cobre representa cerca del 21,8% del total de las divisiones de la Corporación (revisar Anexo 1).

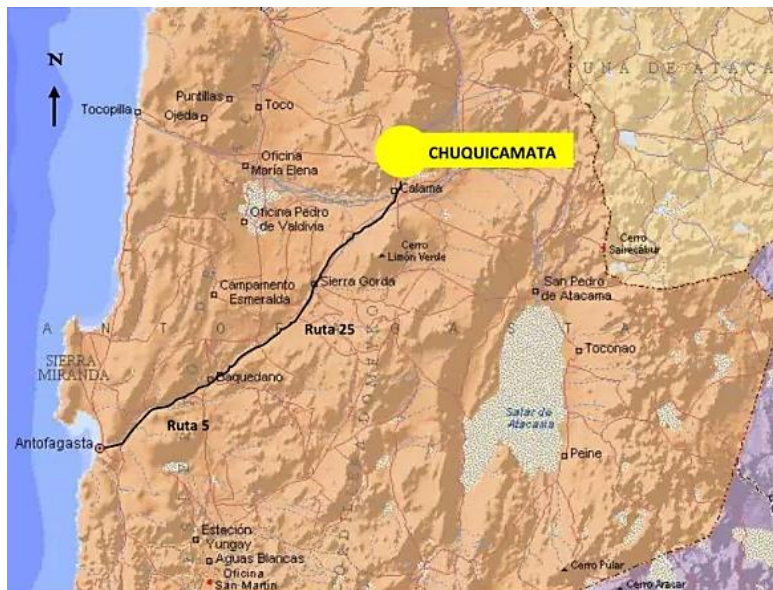


Figura 1. Mapa de acceso a Mina Chuquicamata[6].

La minería a gran escala comenzó en 1915 bajo la Chile *Exploration Company*, propiedad de la familia Guggenheim, quienes aplicaron el novedoso método de explotación a rajo abierto, convirtiendo a Chuquicamata en un referente mundial. Posteriormente, la propiedad

pasó a la *Anaconda Copper Mining Company*, consolidándose como la mina de cobre a cielo abierto más grande del mundo en dimensiones.

Actualmente, Chuquicamata se encuentra en medio de un proceso de transformación estructural sin precedentes para asegurar su continuidad por al menos 40 años más. Este cambio es impulsado por el agotamiento de las reservas superficiales y la baja ley del mineral en el fondo del rajo. Tras más de 100 años de explotación a cielo abierto, el yacimiento inició su conversión a una operación subterránea mediante el Proyecto Chuquicamata Subterránea (CChS). La operación subterránea comenzó sus obras y la extracción de las primeras toneladas de mineral en 2019, marcando el fin gradual de la explotación a rajo abierto. La división produce principalmente cátodos electro – refinados, concentrado de cobre y molibdeno.

1.3.1 Descripción del Procesos Productivo de la División

En el marco de su transformación estructural, Chuquicamata está impulsando proyectos clave para optimizar su producción. Específicamente, se ha contemplado un proyecto de aumento de producción para los próximos años de 210 ktpd, con un foco en el 2026 para consolidar y acelerar la curva de producción de la mina subterránea. Este proyecto busca asegurar la continuidad operacional de las plantas concentradoras, aprovechando las reservas de alta ley del nuevo yacimiento subterráneo para impactar positivamente la producción total de la División.

El área de molienda de la gerencia concentradora consta de 4 plantas de circuito cerrado, 2 de ellas en las que se emplean molienda convencional, mientras las otras corresponden a molienda semi – autógena. La concentradora es alimentada por mineral propio extraído de la mina rajo abierto y de la mina subterránea, además de recibir desde la División Radomiro Tomic mediante correas o, en su defecto, CAEX.

El mineral reducido en chancado primario (P100 < 8”) se transporta hacia una de las plantas de molienda SAG y otra proporción vuelve a reducir su granulometría en los chancadores secundarios y terciarios con una tolerancia de un 20% de masa retenida $\pm \frac{1}{2}$ ” en el producto hacia las plantas de moliendas convencionales. Aprovechando la proximidad entre la concentradora y la planta de fundición de ánodos, se transporta escoria hacia la otra molienda SAG.

La primera planta de molienda en la división fue inaugurada en 1952, se compone de 13 secciones cada una con un molino de barras y 2 de bolas a excepción de la primera que posee uno de cada molino, con un molino de bolas de mayor tamaño. La siguiente planta de molienda convencional se divide en 3 secciones con un molino de cada tipo y una batería de hidrociclones (BHC) por sección al igual que la primera planta. Por su parte, la molienda de mineral mediante tecnología SAG se configura como un circuito SABC-A, correspondiente a un arreglo compuesto por un molino SAG seguido de una etapa secundaria de molienda en molinos de bolas, con recirculación de *pebbles* chancados hacia el molino SAG. Este circuito está conformado por dos secciones paralelas, cada una de las cuales dispone de un molino SAG, dos molinos de bolas y dos BHC, mientras que, la molienda SAG asignada exclusivamente para la escoria, cumple el mismo circuito en una sola sección con una BHC. Actualmente, las plantas de molienda están implementando estrategias para incrementar su rendimiento en 192 mil toneladas por día. Para lograr este objetivo, es fundamental garantizar la continuidad operacional.

Mediante canaletas o tuberías se traslada la pulpa desde la salida de los molinos hacia una cuba en la cual se bombea hacia la BHC que la clasifica cumpliendo el “40/30”, es decir, el producto que recibe la flotación debe tener 40% de sólido con un 30% de masa retenida sobre la malla 65 de la marca Tyler, de no ser así se recircula hacia los molinos de bolas.

En el caso de la molienda de mineral SAG se añade un nuevo molino de bolas con mayor capacidad, con el principal objetivo de disminuir la carga circulante en los molinos de bolas y la recirculación proveniente de los chancadores de *pebbles* hacia los molinos SAG, logrando aumentar el rendimiento de la planta.

La pulpa ambientada con cal, que se extrae por el *overflow* del hidrociclón es transportada hacia la flotación primaria de cada planta obteniendo un concentrado con alto contenido en cobre y molibdeno. El concentrado colectivo se transporta a la planta de molibdeno en la cual se realiza la depresión del cobre para flotar el molibdeno con reactivos (depresor, espumante y modificadores de pH), se filtra, seca y recupera como subproducto. Se recupera alrededor de 85% de cobre desde una segunda etapa de flotación, alimentada de concentrado selectivo, luego se dirige a la planta de filtros donde disminuye su humedad entre 9 a 12% y se almacena. Es transportado hacia la planta de fundición para la producción de ánodos y posterior refinación hasta la obtención de cátodos de un 99,99% de pureza (ver Anexo 2).

2. MARCO TEÓRICO

En este capítulo se presentan los principales conceptos, teorías y modelos relevantes en la aplicación de aprendizaje autónomo para el proceso de molienda en molinos semi-autógenos. El propósito es fundamentar teóricamente el desarrollo del modelo predictivo y optimización, que busca recomendar el tipo de mineral a alimentar según condiciones operacionales, mediante una comprensión de la conminución, el modelado de variables operacionales y los algoritmos de optimización empleados.

2.1 Conceptos de molienda SAG y alimentación por granulometría

En plantas de procesamiento de minerales, la alimentación al molino SAG puede clasificarse según la granulometría: mineral fino, intermedio o grueso. El tipo de mineral y su tamaño influyen la dinámica de molienda (por ejemplo, carga de bolas/pebble, impacto, abrasión, autógena) y en consecuencia el rendimiento (t/h), consumo de energía y distribución del tamaño del producto. Aunque no existe una “ley” universal específica para SAG en la misma forma que para molienda de bolas, los conceptos clásicos de conminución permiten ilustrar que alimentaciones más gruesas requieren mayor trabajo de fractura y pueden generar mayores porcentajes de carga autógena, mientras que alimentaciones muy finas pueden provocar sobre molienda o desgaste excesivo.

Asimismo, es esencial definir las variables operacionales clave que se usan como entradas en modelos predictivos: granulometría de alimentación (por ejemplo, F_{80}), porcentaje de bolas, velocidad del molino, carga del molino, porcentaje de sólidos, etc. Estas variables permiten explicar y modelar la relación entre condiciones de alimentación y resultado del proceso.

2.2 Mapa de Calor de Correlación de Pearson

El Mapa de Calor de Correlación de Pearson es una herramienta gráfica fundamental en el análisis exploratorio de datos (EDA) y la minería de procesos, permitiendo visualizar de manera concisa y colorimétrica las relaciones lineales bivariadas entre un conjunto de variables. El coeficiente de correlación de Pearson, denotado como r , mide la fuerza y la dirección de la relación lineal entre dos variables continuas. Sus valores oscilan entre -1 (correlación negativa perfecta, donde al aumentar una variable, la otra disminuye) y +1

(correlación positiva perfecta, donde ambas variables aumentan o disminuyen conjuntamente), siendo 0 indicativo de la ausencia de una relación lineal. En el contexto de los procesos industriales, esta visualización es invaluable para identificar variables con alta redundancia (multicolinealidad), lo cual puede impactar la estabilidad de los modelos predictivos, y para descubrir variables latentes o *drivers* del proceso que, aunque no son la variable de salida, presentan una fuerte relación con otros parámetros operativos [7].

2.3 Bases de modelado predictivo con XGBoost

El modelo propuesto se sustenta en técnicas de aprendizaje supervisado, específicamente en el método *XGBoost* (*eXtreme Gradient Boosting*). *XGBoost* es un algoritmo de potenciación de gradiente altamente escalable y eficiente. Opera construyendo una serie de árboles de decisión de forma aditiva, donde cada nuevo árbol intenta corregir los errores cometidos por la suma de todos los árboles anteriores [8].

La arquitectura de *XGBoost* se basa en el concepto de modelos aditivos, donde el resultado final es la suma de las predicciones de K árboles de decisión (f_k).

$$\hat{y}_i = \sum_{k=1}^k f_k(x_i), \quad f_k \in F \quad (1)$$

Donde \hat{y}_i es la predicción del rendimiento (tonelaje) para la i-ésima muestra de entrada x_i , y F es el espacio de las funciones de regresión de árboles.

La clave de *XGBoost* reside en la minimización de una función objetivo regularizada $L(\phi)$:

$$\mathcal{L}(\phi) = \sum_i l(y_i, \hat{y}_i) + \sum_k \Omega(f_k) \quad (2)$$

El primer término, $\sum_i l(y_i, \hat{y}_i)$, es la función de pérdida (que mide el error de predicción, como RMSE), y el segundo término, $\sum_k \Omega(f_k)$, es el término de regularización que penaliza la complejidad del modelo (evitando el sobreajuste, o *overfitting*). El término de regularización $\Omega(f_k)$ se define típicamente como:

$$\Omega(f_k) = \gamma T + \frac{1}{2} \lambda \sum_{j=1}^T w_j^2 \quad (3)$$

Donde T es el número de hojas del árbol, y controla el número de nodos (poda), λ es el coeficiente de la regularización L2, y w_j es el puntaje de la hoja j [8].

Su principal ventaja es la alta capacidad de generalización, el manejo eficiente de datos con ruido o correlaciones, y su capacidad de capturar relaciones no lineales y complejas entre las variables operacionales de entrada y salida. Su aplicabilidad en molienda se debe a que las variables operacionales presentan interacciones no lineales y multivariadas, difíciles de capturar con modelos estadísticos tradicionales.

Además, *XGBoost* incorpora técnicas de regularización que evitan el sobreajuste, permitiendo un mejor equilibrio entre precisión y generalización.

En el contexto de molienda SAG, este enfoque permite predecir el rendimiento esperado del molino (*throughput*) a partir de las condiciones operacionales y la granulometría de alimentación.

De esta manera, el modelo no solo estima valores futuros, sino que también puede ser utilizado como herramienta de recomendación para determinar qué tipo de mineral (fino, intermedio o grueso) se debería alimentar para alcanzar el máximo rendimiento o estabilidad operativa.

2.4 Indicadores de Ajuste del Modelo Predictivo

Para validar la solidez del modelo *XGBoost*, se emplean métricas de evaluación de regresión que cuantifican la magnitud y la dirección de los errores de predicción.

2.4.1 Error Cuadrático Medio de la Raíz (RMSE)

Indicador que mide la magnitud promedio de los errores de predicción, representando la desviación estándar de los residuos o diferencias entre los valores observados y los valores predichos por el modelo. Esta métrica penaliza de manera más severa los errores grandes debido al término cuadrático en su cálculo, lo que la hace especialmente útil cuando se busca evitar predicciones con desviaciones extremas. Su unidad es la misma que la variable objetivo (por ejemplo, toneladas), y un valor de RMSE menor indica un mejor ajuste del modelo a los datos observados.

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad (4)$$

Donde y_i es el valor real, \hat{y}_i es el valor predicho, y N es el número de muestras.

2.4.2 Coeficiente de Determinación (R^2)

Evalúa la proporción de la variabilidad de la variable dependiente (por ejemplo, el rendimiento) que puede ser explicada por las variables independientes, como la granulometría. En otras palabras, mide qué tan bien el modelo logra representar la relación entre las variables de entrada y la variable objetivo. Su principal función es indicar la bondad de ajuste del modelo, mostrando qué tan cerca están las predicciones de los valores reales observados. El valor de R^2 varía entre 0 y 1, donde valores cercanos a 1 indican un mejor desempeño del modelo. Por ejemplo, un $R^2 = 0.95$ significa que el modelo explica el 95 % de la variabilidad del rendimiento, lo que a primera vista sugiere una excelente capacidad predictiva.

Sin embargo, un valor de R^2 superior a 0.90 puede ser una señal de sobreajuste (*overfitting*), especialmente en modelos de aprendizaje automático o con muchas variables explicativas. El sobreajuste ocurre cuando el modelo se ajusta excesivamente a los datos de entrenamiento, capturando no solo las relaciones reales sino también el ruido o las fluctuaciones aleatorias del conjunto de datos. Esto provoca que el modelo tenga un desempeño excelente en los datos con los que fue entrenado, pero pierda capacidad de generalización al aplicarse a nuevos datos. Por ello, aunque un R^2 alto parece deseable, es fundamental evaluar otras métricas de validación y comprobar el rendimiento del modelo en datos no vistos para asegurar que el ajuste sea representativo y no producto del sobreajuste.

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2} \quad (5)$$

Donde \bar{y} es la media del valor real del rendimiento.

2.4.3 Error Absoluto Medio (MAE)

Mide el promedio de las magnitudes absolutas de los errores entre los valores observados y los predichos por el modelo. A diferencia del RMSE, esta métrica no eleva los errores al cuadrado, por lo que no penaliza de forma exagerada los errores grandes, siendo así una medida más robusta frente a valores atípicos. Su interpretación es sencilla e intuitiva, ya que representa el error promedio en las mismas unidades que la variable predicha (por

ejemplo, toneladas). Un MAE menor indica un mejor desempeño del modelo, reflejando predicciones más precisas y consistentes con los datos reales.

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i| \quad (6)$$

2.4.4 Error Porcentual Absoluto Medio (MAPE)

mide el error de predicción en términos porcentuales respecto al valor real, proporcionando una visión relativa del desempeño del modelo. Esta métrica resulta especialmente útil en contextos operacionales o de negocio, ya que permite interpretar fácilmente el nivel de error en relación con la magnitud de los valores observados. Su función principal es ofrecer una medida intuitiva y comparable del error, independientemente de las unidades de la variable analizada. El MAPE se expresa en porcentaje (%), y valores más bajos indican un modelo con mayor precisión y mejor capacidad de predicción.

$$MAPE = \frac{100\%}{N} \sum_{i=1}^N \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (7)$$

2.5 Fundamento de interpretación mediante SHAP

Una vez entrenado el modelo predictivo para estimar el rendimiento o la condición operativa del proceso, la aplicación del método *SHAP* (*SHapley Additive exPlanations*) permite cuantificar de manera precisa la contribución individual de cada variable a la predicción generada. Este enfoque, basado en la teoría de los valores de Shapley, descompone la salida del modelo en efectos aditivos atribuibles a cada característica, ofreciendo una interpretación consistente y matemáticamente fundamentada de su influencia. De este modo, *SHAP* proporciona transparencia y explicabilidad al modelo, permitiendo identificar qué variables ejercen mayor impacto sobre las predicciones y cómo lo hacen, de manera positiva o negativa (Figura 2). Esta información resulta especialmente útil para la toma de decisiones operacionales, ya que facilita comprender las relaciones entre las condiciones del proceso y el resultado estimado, respaldando así la recomendación de acciones más eficientes y fundamentadas [9].

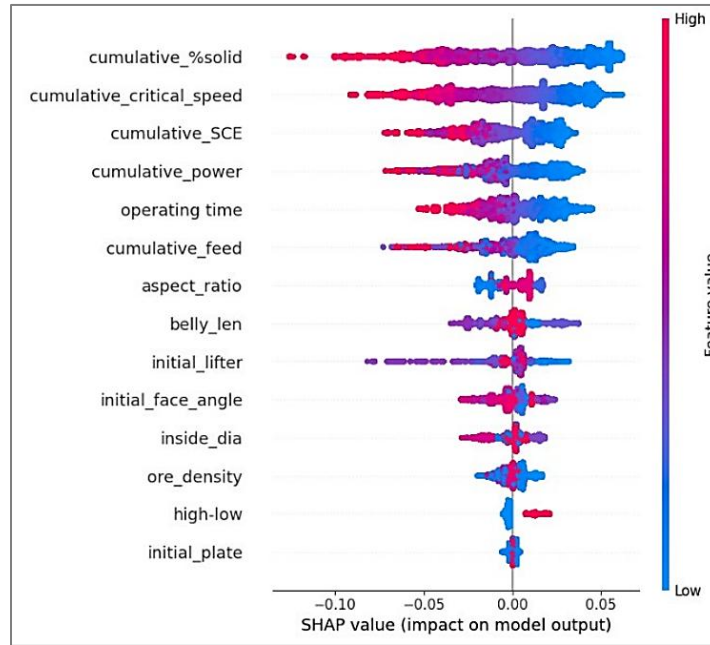


Figura 2. *Shapley* para un modelo *XGBoost* aplicado en altura de *lifter* (modificado)[10].

2.6 Optimización Basada en el Modelo con Differential Evolution

En cuanto a algoritmos de optimización, destaca el método *Differential Evolution* (Evolución Diferencial), propuesto por Rainer Storn y Kenneth Price en 1997 [11]. Este método se basa en una población de vectores-solución que se mutan mediante la diferencia de otros vectores y luego se cruzan y seleccionan, sin uso de derivadas, lo que lo hace robusto para problemas de optimización continua no lineal. Su ventaja es la simplicidad de parámetros, buena convergencia, y aplicabilidad en procesos industriales complejos [12].

Differential Evolution (DE) es un método de Optimización Evolutiva (metaheurística) que se caracteriza por su simplicidad, eficiencia y robustez para encontrar óptimos globales en problemas no lineales, multimodales y con múltiples restricciones. A diferencia de otros métodos de optimización basados en gradientes, DE no requiere información de la derivada de la función objetivo, lo que lo hace ideal para su uso con modelos de ML complejos como *XGBoost* [11].

La función objetivo $f(x)$ a maximizar es la predicción del rendimiento (t/h) generada por el modelo *XGBoost*, siendo x el vector de granulometría a optimizar:

$$x = (\%fino, \%intermedio, \%grueso) \quad (8)$$

$$\text{Maximizar: } f(x) = \hat{y}_{XGBoost}(x)$$

El algoritmo busca la composición granulométrica x^* que produce el máximo tonelaje predicho, $\hat{y}_{XGBoost}$, dentro de los límites operacionales definidos.

1.1 Proceso Evolutivo (Mutación, Cruce, Selección):

1.1.1 Mutación

La etapa de Mutación en el algoritmo de DE tiene como objetivo crear un vector mutante (v_i) para cada vector objetivo x_i de la población actual. Este proceso introduce la diversidad necesaria para la exploración del espacio de búsqueda.

El procedimiento se basa en la manipulación de vectores existentes. Para un vector objetivo x_i , el algoritmo DE selecciona tres vectores aleatorios distintos (x_{r1} , x_{r2} , x_{r3}) de la población actual. Es fundamental que los índices de estos vectores sean diferentes entre sí ($r1 \neq r2 \neq r3$) y diferentes del índice del vector objetivo ($r1, r2, r3 \neq i$). El vector mutante v_i se genera combinando estos vectores mediante la diferencia escalada de dos de ellos, la cual se suma al tercer vector.

La expresión matemática que define la operación de mutación es:

$$v_i = x_{r1} + F(x_{r2} - x_{r3}) \quad (9)$$

Donde, x_{r1} corresponde al vector base aleatorio, el término $(x_{r2} - x_{r3})$ es la diferencia vectorial entre dos individuos aleatorios, que proporciona la dirección y magnitud de la perturbación. El Factor de Mutación, F , es un parámetro de control en el rango $[0, 1]$, que escala la diferencia vectorial y determina la intensidad de la mutación.

Este esquema de mutación aprovecha la información de la población (las diferencias entre los individuos) para generar nuevos puntos en el espacio de búsqueda. El término $F \cdot (x_{r2} - x_{r3})$ actúa como un vector de perturbación, añadiendo una componente diferencial a un vector base x_{r1} . Este mecanismo es el núcleo de la capacidad de DE para la autoadaptación y la exploración robusta del espacio.

1.1.2 Cruzamiento

La etapa de Cruce (*Crossover*) sigue a la Mutación y tiene como objetivo generar un vector de prueba (u_i) al combinar elementos del vector objetivo (x_i) y el vector mutante (v_i). Este proceso se realiza utilizando una probabilidad de cruce. El Cruce se aplica componente a componente para el vector x_i .

El esquema de cruce más comúnmente utilizado es el Cruce Binomial, también conocido como *Uniform Crossover*. El elemento j del vector de prueba $u_{i,j}$ toma su valor del vector mutante $v_{i,j}$ si un número aleatorio generado uniformemente ($rand_j$) es menor o igual a la probabilidad de cruce, o si j es igual a un índice aleatorio j_{rand} (para garantizar al menos una transferencia). En caso contrario, el elemento se toma del vector objetivo $x_{i,j}$.

La formulación matemática es la siguiente:

$$u_{i,j} = \begin{cases} v_{i,j}, & \text{si } rand_j \leq CR \text{ o } j = j_{rand} \\ x_{i,j}, & \text{en otro caso} \end{cases} \quad (10)$$

Donde $v_{i,j}$ es el elemento j del vector mutante, $x_{i,j}$ representa el elemento j del vector objetivo. CR es la Probabilidad de Cruce, un parámetro de control en el rango $[0, 1]$; $rand_j$ es un número aleatorio uniforme generado para la componente j , comprendido en el rango $[0,1]$ y j_{rand} representa un índice aleatorio entero seleccionado una vez por cada vector x_i perteneciente a un conjunto $\{1, 2, \dots, D\}$.

La inclusión de la condición " $j = j_{rand}$ " es un aspecto crucial y de diseño fundamental en el Cruce Binomial de la Evolución Diferencial, ya que garantiza la diversidad en el vector de prueba u_i . Específicamente, esta condición asegura que al menos una componente del vector de prueba sea heredada obligatoriamente del vector mutante v_i , incluso si la Probabilidad de Cruce es muy baja o si el número aleatorio generado es mayor que CR para todas las componentes; de esta manera, se evita la inutilidad de la operación, impidiendo que el vector de prueba u_i sea idéntico al vector objetivo x_i , lo que garantiza que se presente una solución potencialmente nueva a la etapa de Selección y se mantenga la exploración activa del espacio de búsqueda [11].

1.1.3 Selección

La etapa de Selección es el paso final en la generación de la nueva población en la Evolución Diferencial. Su propósito es decidir qué vector (el objetivo original x_i o el nuevo vector de prueba u_i) sobrevivirá para la siguiente generación.

La decisión se basa en una simple comparación de sus respectivos valores de aptitud. Para un problema de minimización, el vector de prueba u_i reemplaza al vector objetivo x_i en la población si su valor de aptitud es igual o mejor (igual o menor). Si el vector objetivo x_i tiene un valor de aptitud mejor (menor) que el vector de prueba u_i , x_i se mantiene en la población para la siguiente generación.

Matemáticamente, la selección se expresa como:

$$x_i^{next} = \begin{cases} u_i, & \text{si } f(u_i) \leq f(x_i) \\ x_i, & \text{en otro caso} \end{cases} \quad (11)$$

Al aplicar estos métodos al proceso de molienda SAG, se considera que la granulometría de alimentación, la carga y velocidad del molino, entre otras variables, pueden optimizarse para maximizar el rendimiento o minimizar el consumo de energía, recomendando el tipo de mineral según la agrupación operativa[13].

La granulometría manejada en la empresa se clasifica en cuatro categorías según el tamaño de partícula del mineral. El mineral fino corresponde a la fracción del material con un tamaño inferior a 0,5". El mineral intermedio comprende partículas con granulometría entre 0,5" y 2". El mineral grueso abarca tamaños de partícula entre 2" y 8", mientras que el supergrueso se define por partículas con dimensiones superiores a 8". Esta clasificación permite caracterizar la distribución granulométrica del mineral y constituye un elemento esencial para evaluar su comportamiento dentro del proceso de molienda.

El rendimiento representa la tasa de procesamiento del molino, expresada en toneladas por hora (t/h), e indica la cantidad de mineral tratado en un determinado periodo de tiempo. Este parámetro depende de variables operacionales como la carga del molino, que corresponde al porcentaje del volumen interno ocupado por el mineral y/o el medio de molienda (bolas o *pebbles*), así como de la granulometría de alimentación (F80) y la granulometría de producto (P80). El F80 describe el tamaño de partícula al cual pasa el 80 % en peso del mineral de alimentación, mientras que el P80 indica el tamaño al cual pasa el 80 % en peso del producto resultante del proceso de molienda.

El rendimiento también depende las características del mineral. El *Work Index* (Índice de trabajo) es un parámetro energético, expresado en kWh/ton, que mide la energía necesaria para reducir el tamaño del mineral desde un tamaño teóricamente infinito hasta 100 μm , según la teoría de Bond [14]. Este índice permite comparar la dureza relativa de diferentes minerales y estimar los requerimientos energéticos de los procesos de conminución.

3. ESTADO DEL ARTE

Con el aumento del uso de tecnologías digitales en la industria minera, la cantidad de datos de producción almacenados en un sistema centralizado está en continuo crecimiento. Los conjuntos de datos en las minas y las plantas de procesamiento se utilizan para establecer vínculos entre los minerales extraídos y el rendimiento de las plantas metalúrgicas.

Considerando que la etapa de molienda representa aproximadamente el 50% de los costos totales asociados al procesamiento de minerales [1], y en el contexto de una creciente complejidad mineralógica, la volatilidad de los precios de los *commodities* y de la necesidad de eficiencia energética, la implementación de enfoques basados en *machine learning* (ML) para la predicción del rendimiento del molino ha sido objeto de múltiples investigaciones previas ([2], [7], [15], [16], [17], [18]).

Los enfoques tradicionales, basados en regresiones lineales o ecuaciones mecánicas, presentan limitaciones para modelar la dinámica no lineal inherente al proceso de molienda semiautógena (SAG). En este contexto, los modelos de ML se posicionan como herramientas robustas para capturar relaciones complejas entre variables operacionales y variables de diseño, con el fin de mejorar la eficiencia del proceso, predecir el comportamiento de componentes críticos y optimizar parámetros operacionales. Saldaña *et al.* [3] exploran diversos algoritmos, incluyendo árboles de decisión (*Random Forest*, *Gradient Boosting Machine*, *XGBoost*) y redes neuronales artificiales (ANN) aplicados en mejorar los indicadores productivos del proceso, como la producción y el consumo de energía. Entre estos, los modelos basados en ANN y *XGBoost* destacaron por su mejor desempeño, evaluado mediante indicadores de ajuste como el error absoluto medio (MAE), error cuadrático medio (RMSE) y R^2 (coeficiente de determinación).

En su investigación, Plural *et al.* [10] demostraron que una predicción precisa del desgaste de los *liners* resulta esencial para optimizar el rendimiento, reducir costos operativos y mejorar la planificación del mantenimiento. El estudio destaca el potencial del ML para estimar el desgaste en múltiples molinos con geometrías distintas, incluso en escenarios de extrapolación. Se compararon distintos enfoques (regresión lineal simple, regresión múltiple, árboles de decisión y ANN) identificando a *XGBoost* como el modelo más exitoso en un enfoque genérico, utilizando el error absoluto medio porcentual (MAPE) como métrica principal de validación, logrando un 5,27%. Un aspecto que incluye esta investigación es la incorporación de SHAP (*SHapley Additive exPlanations*) para interpretar la influencia de

variables como el tiempo de operación, la energía específica, la velocidad crítica y el diseño del *liner*. No obstante, la precisión del modelo se ve limitada por la ausencia de variables relevantes, por lo que se propone ampliar el conjunto de datos e incorporar variables fisicoquímicas para mejorar la caracterización del mecanismo de desgaste.

La metodología utilizada por Ghasemi *et al.* [19] será adaptada en el presente estudio. En su investigación, los autores desarrollaron un marco inteligente que combina ML, conocimiento experto y algoritmos evolutivos para identificar condiciones óptimas de parámetros operacionales dentro de límites físicos y técnicos, con el objetivo de maximizar el rendimiento de la molienda SAG. Se evaluaron 17 modelos predictivos, incluyendo enfoques clásicos, modelos basados en árboles de decisión, redes neuronales y métodos de ensamble. Para mejorar la precisión, se implementaron técnicas de selección de variables mediante *Recursive Feature Elimination* (RFE) y detección de valores atípicos mediante *Local Outlier Factor* (LOF). El mejor modelo fue posteriormente integrado con tres algoritmos evolutivos —*Differential Evolution* (DE), *Genetic Algorithm* (GA) y *Particle Swarm Optimization* (PSO)— destacándose DE por lograr predicciones de rendimiento con baja dispersión entre ejecuciones. La validación del modelo incluyó validación cruzada y métricas como RMSE, MAE, MAPE, SWAPE y R^2 , además de una etapa de postvalidación con expertos industriales para asegurar la viabilidad operativa de las configuraciones óptimas propuestas. Asimismo, se propuso actualizar el modelo periódicamente con datos recientes para mantener su vigencia, así como incorporar variables adicionales, como la dureza del mineral, las propiedades del medio de molienda y las características del *liner*, con el fin de mejorar su capacidad predictiva.

Los algoritmos evolutivos, como *Differential Evolution*, ofrecen una estrategia eficaz para optimizar variables de decisión relevantes, entre ellas la granulometría de alimentación. En particular, DE ha demostrado ser un algoritmo de optimización efectivo, por su capacidad de explorar espacios de búsqueda complejos sin requerir derivadas ni continuidad en la función objetivo, alcanzando predicciones de mayor rendimiento del molino con una robustez superior en comparación con otros métodos analizados [15], [19].

Según Wang *et al.* [20] este método presenta ventajas significativas en problemas de molienda y clasificación, logrando convergencia eficiente hacia soluciones óptimas. En el estudio de Mhlanga *et al.* [21], se implementa un algoritmo evolutivo multiobjetivo para el diseño de circuitos de conminución, demostrando mejoras significativas en la productividad y en la eficiencia energética del proceso. En este contexto, el circuito de molienda fue

optimizado en función del rendimiento esperado de la etapa de trituración y cribado, considerando un margen de conciliación del 5%.

Por otro lado, Hingston *et al.* [22] aplicaron algoritmos evolutivos para diseñar plantas de procesamiento, obteniendo configuraciones superiores a las prácticas convencionales. Asimismo, Sun *et al.*[23] propusieron un enfoque adaptativo de DE basado en aprendizaje por refuerzo, mejorando la eficiencia en entornos industriales. Una limitación recurrente en la literatura es que la mayoría de los estudios priorizan la maximización del rendimiento del molino y/o la disminución del consumo energético, desatendiendo la granulometría de alimentación, que constituye un indicador clave para evaluar la eficiencia y la estabilidad del proceso.

Las hipótesis estudiadas por Eltaeib y Mahmood [13] indican un mejoramiento del rendimiento de la molienda SAG entre un 4% y un 9%, dependiendo de la configuración inicial y de la calidad de los datos. El enfoque permite simular más de 10.000 configuraciones en menos de dos horas en escenarios operacionales, lo que evidencia una alta eficiencia computacional. *Differential Evolution* supera a algoritmos clásicos como GA y PSO en problemas con restricciones complejas, como las proporciones granulométricas que deben sumar el 100%, condición presente en el caso de estudio. Una de las ventajas de DE es que puede emplear como función objetivo modelos *Boosting* entrenados con datos reales de planta, lo que posibilita una optimización indirecta sin necesidad de simulaciones físicas. Asimismo, presenta gran adaptabilidad, ajustándose dinámicamente a cambios mineralógicos, condiciones de alimentación o fallas operativas. Sin embargo, su eficiencia depende de la adecuada selección de parámetros como el *scale factor* (F) y la probabilidad de cruce (CR). Además, la presencia de datos inconsistentes o poco representativos puede afectar la convergencia del algoritmo.

Uno de los principales desafíos del modelado basado en datos en procesos industriales complejos proviene de la naturaleza misma de los datos de proceso, que con frecuencia son de alta dimensión, no seguir una distribución normal y presentar no estacionariedad, además de incluir relaciones no lineales y multimodales, valores faltantes o defectuosos, ruido, valores atípicos, correlaciones y autocorrelaciones significativas, así como tasas variables y retrasos en el muestreo y en el propio proceso ([24],[25]). La disponibilidad de datos con suficiente variabilidad para que un modelo capture las verdaderas relaciones entre variables constituye también una dificultad, al igual que la disponibilidad de datos etiquetados, particularmente crítica para las técnicas de aprendizaje supervisado. Para una

descripción detallada de las técnicas de aprendizaje automático en enfoques supervisados y no supervisados existe una amplia variedad de fuentes de referencia [26], [27].

Dado el reconocido desempeño del *machine learning* en el ajuste de modelos complejos, su aplicación en la industria de procesamiento de minerales presenta un alto potencial para optimizar procesos como la molienda SAG, ya sea a través de la mejora de los indicadores de producción o de la reducción del consumo energético. En esta investigación, y considerando el reconocimiento del *boosting* como una de las técnicas de aprendizaje más efectivas [28], así como su desempeño superior demostrado en estudios previos, se emplea *XGBoost* como modelo de predicción.

La sinergia entre modelos predictivos, como *XGBoost*, y algoritmos evolutivos, como *Differential Evolution* —cuya capacidad para manejar restricciones, explorar espacios complejos y adaptarse a condiciones dinámicas lo convierte en un candidato idóneo para aplicaciones industriales— constituye una herramienta robusta y flexible para la optimización de circuitos de molienda SAG. Este enfoque híbrido permite maximizar el rendimiento global de los molinos, manteniendo la adaptabilidad frente a condiciones variables y restricciones técnicas. La evidencia empírica y académica respalda su aplicabilidad y eficacia en contextos industriales.

En conjunto, estos enfoques reflejan una evolución metodológica hacia sistemas inteligentes, escalables y adaptativos para la gestión avanzada del proceso SAG. El marco inteligente desarrollado en este estudio se extiende más allá de la mera aplicación de métodos de ML e IA. La participación de un experto en la selección de las características relevantes y la evaluación de las soluciones propuestas garantiza un proceso de optimización informado.

CAPÍTULO III: PROBLEMÁTICA, HIPÓTESIS, ALCANCE, OBJETIVOS

1. PROBLEMÁTICA

El escenario global evidencia un alza sostenida en la demanda del cobre impulsada por la transición energética y tecnologías verdes; el crecimiento en sectores de inteligencia artificial y automatización; además de la industrialización y urbanización en economías emergentes. Codelco lidera la producción nacional del metal rojo, sin embargo, la empresa enfrenta algunos desafíos en proveer la exigente demanda del producto. Aun así, no se han efectuados mayores modificaciones en cuanto a infraestructura, manteniendo en uso los equipos que desde 1952 dieron inicio a la operación de la planta concentradora. La necesidad de aumentar la producción a menores costos hace fundamental el estudio y aplicación de herramientas estadísticas junto con técnicas de aprendizaje automático que permitan modelar las dinámicas de estos procesos complejos. En particular, el *stockpile* que alimenta la molienda SAG de la división Chuquicamata presenta segregación de carga por granulometría, lo que afecta la homogeneidad en la distribución granulométrica del mineral entre los distintos *feeders*. Esto establece una preferencia operativa por la liberación de partículas de menor tamaño, con el fin de no comprometer el rendimiento del proceso de molienda.

Una alternativa abordada ha sido la intervención en la infraestructura del sistema de alimentación hacia la pila, con el objetivo de modificar la trayectoria de proyección del mineral y favorecer una depositación más vertical. Esta disposición permite que la fracción gruesa del mineral se acumule preferentemente en el contorno de la pila, contribuyendo así a una distribución granulométrica más homogénea entre los distintos *feeders* y, en consecuencia, optimizando la uniformidad del material alimentado al proceso de molienda.

No obstante, el mineral transportado por las correas es impulsado a altas velocidades, lo que compromete la integridad estructural del sistema de descarga. Si bien las soluciones basadas en refuerzos estructurales representan una alternativa, su implementación no resulta económicamente viable en muchos casos. En contraste, la aplicación de técnicas de analítica avanzada para el modelado del proceso de molienda SAG ha emergido como una alternativa más eficiente y reciente. Actualmente, no existe un proceso estandarizado para la alimentación a través de los *feeders*, ya que su operación depende principalmente del criterio experto de los operadores en sala de control. Esta variabilidad ha impedido establecer una estrategia operativa general, dado que cada condición operacional presenta

desafíos específicos, los cuales están determinados por factores como la capacidad de los equipos, la disponibilidad de recursos e insumos, y la dotación de fuerza laboral. Asimismo, no se ha establecido una estandarización respecto a las variables de entrada que controlan un modelo predictivo del rendimiento de molienda. Aunque se han desarrollado estudios orientados a la identificación de factores relevantes para anticipar su comportamiento, persiste la falta de consenso sobre cuáles resultan más pertinentes.

2. HIPÓTESIS

Se plantea que la aplicación de técnicas de *Machine Learning* sobre datos operacionales del proceso de molienda permite estabilizar y optimizar la curva de distribución granulométrica al interior del molino. De esta forma, al ajustar dicha distribución hacia una condición óptima, es posible aumentar el rendimiento del proceso y mejorar su estabilidad operacional.

3. ALCANCE

Se limita a la construcción de un modelo predictivo con *XGBoost*, interpretado mediante SHAP y optimizado con *Differential Evolution* en base a datos históricos de molienda SAG, siendo sus resultados dependientes de la calidad y representatividad de dichos datos en un periodo de dos años, válidos solo en los rangos observados.

4. OBJETIVOS

4.1 Objetivo General

Optimizar el rendimiento en la molienda SAG a través del control de la velocidad de los feeder aplicando analítica avanzada en la División Chuquicamata de Codelco-Chile.

4.2 Objetivos Específicos

- Relacionar variables operacionales con el rendimiento de la molienda SAG a partir de un levantamiento de data histórica.
- Desarrollar un modelo predictivo que permita proyectar el comportamiento en el rendimiento en la planta de molienda SAG, en función de las variables operacionales de entrada del mineral.
- Establecer un modelamiento de optimización para definir la alimentación a la molienda SAG a partir de los modelos predictivos.
- Evaluar el impacto económico que tendría la implementación de analítica avanzada en la molienda SAG.

CAPÍTULO IV: METODOLOGÍA DE TRABAJO

El proyecto implica la aplicación de *XGBoost* como modelo predictivo supervisado, haciendo hincapié en la visualización del impacto de las variables a través de los valores SHAP. Luego, se implementa un modelo prescriptivo de optimización, el cual emplea el algoritmo *Differential Evolution*, para determinar las combinaciones óptimas de variables operacionales que maximicen dicho rendimiento.

Todo el modelado y análisis se realiza utilizando el lenguaje de programación *Python* 3.13.3, empleando la biblioteca *Scikit-learn* para algoritmos de aprendizaje automático, la biblioteca *XGBoost* para el aumento de gradiente y la biblioteca SHAP para la interpretabilidad del modelo. El preprocesamiento y la visualización de los datos se llevaron a cabo mediante las librerías *Pandas*, *Matplotlib* y *Seaborn*, respectivamente. Para el proceso de optimización, se utilizó el método DE implementado en la biblioteca *SciPy*, la cual permite la búsqueda global de soluciones óptimas en espacios complejos y no lineales. Adicionalmente, se incorporaron librerías auxiliares como *NumPy* para el manejo de matrices y operaciones numéricas, y *Warnings* para la gestión de mensajes de advertencia durante la ejecución del modelo.

1. Fase I: Preparación de Datos

1.1 Disponibilización de Información

La generación de una fuente de datos requiere del levantamiento de información. La recolección de datos históricos y tags de operación del molino debe abarcar un periodo representativo, en particular, desde que los orígenes de alimentación de minerales sean los tres actuales, es decir, mina rajo Radomiro Tomic, mina rajo y subterránea Chuquicamata. El conjunto de datos utilizado en este estudio se recopiló a partir de una infraestructura de datos del sistema AVEVA PI *System* y la granularidad de los datos se determina en base a la interpolación en intervalos de un minuto.

Previo al entrenamiento del modelo, se realiza un proceso de selección de variables con el objetivo de identificar aquellos parámetros operacionales con mayor influencia potencial sobre el rendimiento del circuito SAG. En una primera etapa, se propone un conjunto inicial de variables a partir de criterio experto, considerando conocimiento operativo del proceso de molienda y experiencia en plantas concentradoras. Posteriormente, la selección final se valida mediante un análisis estadístico basado en un mapa de correlación de Pearson, el

cual permite cuantificar el grado de relación lineal entre las variables candidatas y el rendimiento del molino, así como identificar redundancias o colinealidad entre variables explicativas.

1.2 Preprocesamiento de Datos

Una vez recopilados, la fuente de datos requiere una limpieza exhaustiva y un procesamiento; esto incluye el manejo de los valores nulos, atípicos y ruidos. Dado que el modelo considera variables numéricas con diferentes escalas y magnitudes, es necesario realizar la estandarización de las variables de entrada con el fin de garantizar una correcta interpretación y equilibrio en el aprendizaje del modelo *XGBoost*, así como asegurar un desempeño estable durante el proceso de optimización mediante *Differential Evolution*. Asimismo, las variables categóricas deben ser codificadas mediante su transformación en variables binarias, aplicando la técnica de *one-hot encoding*.

El *one-hot encoding* genera tantas variables binarias como categorías tenga cada variable categórica. Cada nueva variable representa una categoría y toma el valor 1 si la observación pertenece a ella, o 0 en caso contrario. Esto permite incorporar información nominal en el modelo sin introducir un orden artificial entre categorías ni afectar la geometría del espacio de características. En este caso, se dispone de variables categóricas que representan el estado operativo (*On/Off*) de los molinos. Para su incorporación al modelo, estas variables se transforman en binarias, asignando el valor 1 al estado 'Funciona' y 0 al estado 'No funciona'.

Para el tratamiento de los valores numéricos, se aplicó un filtro de *outliers* mediante el método *Z-score*, el cual se recomienda por su capacidad para identificar valores atípicos en función de la desviación estándar respecto a la media, permitiendo mantener la consistencia estadística de las variables sin distorsionar su distribución original. Para el manejo de los valores faltantes (NaN), se calculó el porcentaje de predominancia en cada variable, definiendo que aquellas con más del 15% de datos faltantes sean imputados por valores representativos.

Posteriormente, se aplicaron filtros específicos basados en los rangos de operación estándar de la planta para cada variable de entrada y para la variable objetivo, con el fin de asegurar que la interpretación de los resultados sea consistente con las condiciones reales del proceso.

Se implementó una estrategia de ingeniería de características mediante la incorporación de *lags* temporales en el conjunto de variables predictoras y en la variable objetivo. Esta metodología busca explotar la capacidad del algoritmo *XGBoost* para modelar la dependencia serial y los patrones dinámicos inherentes al proceso de molienda. Los *lags* definen la memoria histórica y el horizonte temporal de influencia del modelo. La resolución temporal de la serie corresponde a la frecuencia de muestreo de los datos. Específicamente, el valor de cada *lag* se determinó a partir del promedio móvil de la hora anterior, estableciendo un horizonte temporal de cuatro horas. Esto permite al modelo inferir la causalidad entre las condiciones operacionales precedentes y el rendimiento actual del proceso.

2. Fase II: Modelado Predictivo Supervisado de Rendimiento con XGBoost

El primer paso para la optimización de la molienda SAG es la creación de un modelo predictivo que pueda estimar con precisión el rendimiento (TPH) a partir de las variables de control y perturbación. Dada la naturaleza compleja y no lineal de la molienda, los algoritmos ensamblados basados en árboles han demostrado ser superiores a los modelos empíricos o de regresión lineal.

Una vez generadas las variables con *lags* y definidos los conjuntos de entrada y salida, se procedió al entrenamiento y validación del modelo predictivo basado en *XGBoost*. Previamente, se realizó un control para evitar el *leakage* (filtración de información del futuro hacia el pasado) eliminando de las variables explicativas todos los *lags* asociados directamente al *target*. Este paso es fundamental para garantizar la validez temporal del aprendizaje, evitando que el modelo obtenga un desempeño artificialmente elevado por disponer de información que no estaría disponible en un escenario real de predicción.

2.1 Entrenamiento y Validación del Modelo

En este estudio, se utiliza un conjunto de datos industriales extensivo que comprende 460.761 y 472.687 registros para la molienda del SAG 16 y SAG 17, respectivamente. Para la evaluación del desempeño del modelo se utilizó una validación cruzada temporal (*TimeSeriesSplit*) con cinco divisiones, lo que permite respetar la secuencia cronológica de los datos y simular un entorno de predicción en línea, donde el modelo se entrena con información pasada y se evalúa sobre datos futuros. En cada iteración se entrenó el modelo *XGBRegressor* con una configuración ajustada de hiperparámetros, buscando un equilibrio entre precisión y capacidad de generalización. La función objetivo del modelo define qué se

intenta minimizar durante el entrenamiento, siendo en este caso el error cuadrático entre las predicciones del rendimiento de la planta de molienda (t/h) y los valores reales, con un término adicional de regularización (Ω) que penaliza la complejidad de los árboles mediante α y λ , reduciendo así el riesgo de sobreajuste. Los principales parámetros definidos fueron: una profundidad máxima de 6 niveles por árbol y una tasa de aprendizaje moderada de 0.03.

$$Obj_{XGB} = \sum_{i=1}^n (y_i - \hat{y}_i)^2 + \Omega(f)$$

Donde, y_i representa el valor real del rendimiento (t/h) en la observación i , \hat{y}_i es la predicción del modelo *XGBoost*, $\Omega(f)$ es el término de regularización de los árboles (α y λ) y n es el número de observaciones.

2.2 Evaluación del Modelo

Durante cada partición, se calcularon las métricas de desempeño MSE, RMSE, MAE, MAPE y R^2 , las cuales permiten cuantificar el error medio, la precisión relativa y el grado de ajuste del modelo a los datos observados. Estas métricas se promediaron al final del proceso de validación para obtener una estimación robusta del rendimiento general del modelo.

Además, se incorporó un análisis de interpretabilidad mediante SHAP aplicado al último conjunto de validación. Este análisis permite identificar la contribución individual de cada variable al resultado del modelo, facilitando el diagnóstico de posibles fuentes de *leakage* o la detección de variables con influencia dominante en la predicción del rendimiento. Se reportaron las cinco variables con mayor importancia media absoluta de SHAP, lo que permite validar la consistencia del modelo y su dependencia respecto de las variables operacionales más relevantes.

Finalmente, se revisaron los valores promedio de las métricas para verificar la ausencia de sobreajuste. En caso de obtener valores de R^2 excesivamente altos (superiores a 0,95) o errores porcentuales muy bajos (MAPE < 1%), se considera la posibilidad de revisar las variables incluidas o los *lags* generados, ya que tales resultados podrían indicar un posible *leakage* residual. Este procedimiento asegura la robustez y confiabilidad del modelo *XGBoost*, que posteriormente sirve como módulo de predicción del rendimiento de molienda dentro del modelo global de optimización mediante *Differential Evolution*.

3. Fase III: Implementación del Optimizador de Molinos SAG mediante Differential Evolution

Una vez que se dispone de un modelo *XGBoost* preciso, este se utiliza como función objetivo dentro de un algoritmo de optimización para determinar el conjunto ideal de variables de entrada que maximice el rendimiento. La granulometría de alimentación es la variable a optimizar, ya que es el factor que puede ser modificado para adaptarse a la variabilidad del mineral.

Conformemente, para complementar las predicciones generadas por el modelo *XGBoost*, se desarrolló un optimizador de molinos SAG basado en el algoritmo de *Differential Evolution*. Este módulo tiene como objetivo determinar las proporciones óptimas de granulometría de la alimentación de ambos molinos (SAG 16 y SAG 17) que maximicen el rendimiento total de la planta, medido en toneladas por hora (t/h).

3.1 Carga y Preparación de Modelos Predictivos

El optimizador se inicializa cargando los modelos *XGBoost* entrenados previamente para SAG 16 y SAG 17, junto con la información de las variables de entrada y los valores promedio de estas. Esta información permite preparar los datos de entrada del optimizador, asegurando que las variables que no se ajustan mantengan valores coherentes con el comportamiento histórico del proceso.

3.2 Definición de Variables de Control y Restricciones

Se definieron seis variables de control, correspondientes a las proporciones de mineral fino, intermedio y grueso para cada molino. El optimizador incorpora restricciones explícitas, garantizando que la suma de las proporciones de cada molino sea igual a 100%, penalizando soluciones que violen esta condición para mantener viabilidad operativa.

3.3 Función Objetivo

La función objetivo del optimizador evalúa la combinación de proporciones de granulometría utilizando los modelos *XGBoost* para predecir el rendimiento de cada molino. La función se define como la suma ponderada de los rendimientos predichos, donde se pueden asignar diferentes pesos a SAG 16 y SAG 17 según criterios estratégicos. Se aplica una penalización adicional si las proporciones no cumplen las restricciones de suma, garantizando que la solución final sea válida y factible.

Formalmente:

$$Obj_{DE}(x) = -(w_{16} \cdot \hat{y}_{16}(x) + w_{17} \cdot \hat{y}_{17}(x) - P(x))$$

Donde, x corresponden al conjunto de proporciones granulométricas de alimentación de los molinos SAG (fino, intermedio y grueso), $\hat{y}(x)$ indica las predicciones de *XGBoost* para SAG 16 y SAG 17 usando las proporciones x , w representa los pesos para cada molino y $P(x)$ es la penalización por violar restricciones.

3.4 Ejecución de la Optimización

El algoritmo DE explora de manera global el espacio de soluciones, ajustando iterativamente las proporciones hasta maximizar el rendimiento total ponderado. El proceso es robusto frente a óptimos locales, y permite manejar múltiples restricciones y límites de operación de manera natural. La optimización genera como salida:

- Las proporciones óptimas de granulometría para cada molino.
- Las predicciones de rendimiento individuales y totales.
- Información sobre el desempeño del algoritmo, incluyendo número de iteraciones evaluaciones de función y tiempo de ejecución.

4. Fase IV: Evaluación económica de implementación de ML

La evaluación económica del proyecto se desarrolla con el objetivo de cuantificar el impacto operacional y financiero asociado a la implementación de un sistema de aprendizaje automático orientado a la optimización conjunta de la alimentación a los molinos SAG del circuito de molienda. Dado que la molienda constituye el cuello de botella operacional de la planta concentradora, cualquier incremento sostenido en el rendimiento del circuito se traduce directamente en un aumento de la capacidad de procesamiento global.

En primer lugar, se establece un escenario base representativo de la operación actual, considerando el rendimiento conjunto de los dos molinos SAG. Posteriormente, se define un escenario optimizado correspondiente a la operación asistida por el sistema de ML, el cual permite proyectar un incremento del rendimiento.

Para ambos escenarios se estima la producción anual de cobre fino a partir de variables metalúrgicas representativas de la operación, incluyendo ley promedio de alimentación, recuperación metalúrgica global y un régimen operacional de horas efectivas por año. La producción de cobre recuperado se calcula mediante el balance metalúrgico, considerando el contenido metálico del mineral tratado y la eficiencia global de recuperación del proceso concentrador.

El beneficio económico asociado a la optimización es evaluado desde la perspectiva del costo unitario de producción, utilizando el indicador C1, definido como el costo operacional directo por libra de cobre fino producida. El análisis se estructura mediante un enfoque incremental, evaluando la variación del C1 producto del aumento de producción generado por la optimización.

Para ello, los costos operacionales son desagregados conceptualmente en componentes fijos y variables. Los costos fijos incluyen mano de obra, servicios permanentes, contratos base y gastos administrativos, mientras que los costos variables consideran principalmente consumo energético, medios de molienda y reactivos. Se adoptó una estructura de costos informada por la empresa, considerando una distribución operacional de 70% correspondiente a costos fijos y 30% a costos variables.

Finalmente, el impacto económico anual es calculado considerando tanto el incremento de producción de cobre fino como la reducción del costo unitario de producción, utilizando como referencia el precio promedio del cobre correspondiente al instante en que se identifica el mayor rendimiento con las proporciones granulométricas óptimas.

CAPÍTULO V: RESULTADOS Y DISCUSIÓN

Este estudio introduce un marco inteligente que aprovecha el conocimiento de expertos, técnicas de aprendizaje automático y algoritmos evolutivos para abordar esta necesidad de investigación.

1. Análisis de la Distribución Granulométrica de Alimentación a Molinos SAG

El análisis exploratorio de datos revela una marcada asimetría en la distribución granulométrica de los alimentadores hacia los molinos SAG 16 y 17, lo que sugiere una tendencia operativa divergente en el uso del recurso mineral fino.

En la Figura 3 la operación del molino SAG 16 demostró una clara predominancia en la alimentación de mineral fino (fracción con menor tamaño de partícula).

- El rango total de los datos para la proporción de mineral fino alimentado se extendió desde un mínimo de aproximadamente el 18% hasta un máximo cercano al 80%.
- La mediana (P50) de esta distribución se ubicó aproximadamente en el 50%.

Esta característica operacional vislumbra que el molino SAG 16 procesó una carga con una proporción de material fino significativamente mayor al porcentaje óptimo necesario para maximizar la eficiencia de la molienda semi – autógena. Esta sobreutilización de la fracción fina puede interpretarse preliminarmente como una utilización subóptima del recurso mineral fino, ya que, en lugar de aumentar el rendimiento, limitaría la capacidad efectiva de impacto.

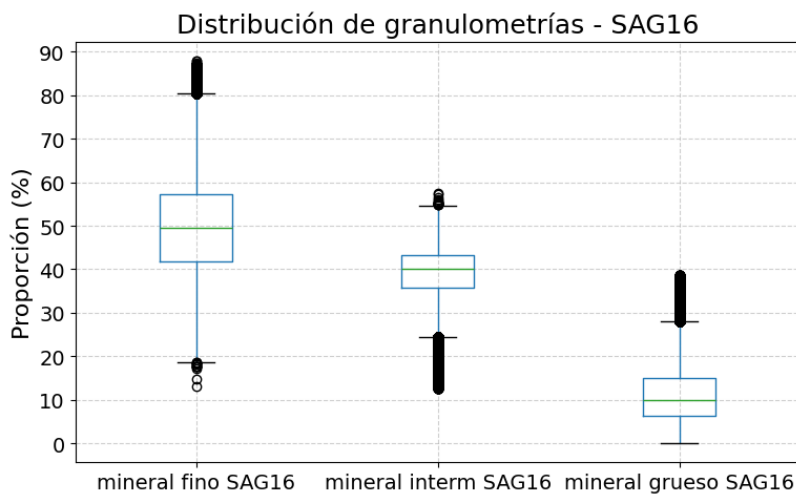


Figura 3. Diagrama de caja de la distribución granulométrica alimentada al SAG 16.

En contraste, en la Figura 4 el molino SAG 17 exhibe una menor participación del mineral fino en su carga.

- La proporción de material fino alimentado varió en un rango más estrecho, desde aproximadamente el 10% hasta el 42%.
- La mediana (P50) de la distribución se situó en un valor significativamente menor, cercano al 25%.

Esta diferencia se atribuye directamente a la heterogeneidad del *stockpile*, donde se identifica una mayor concentración de material grueso destinado al molino SAG 17. Esta condición de almacenamiento dificulta la distribución uniforme de la fracción fina.

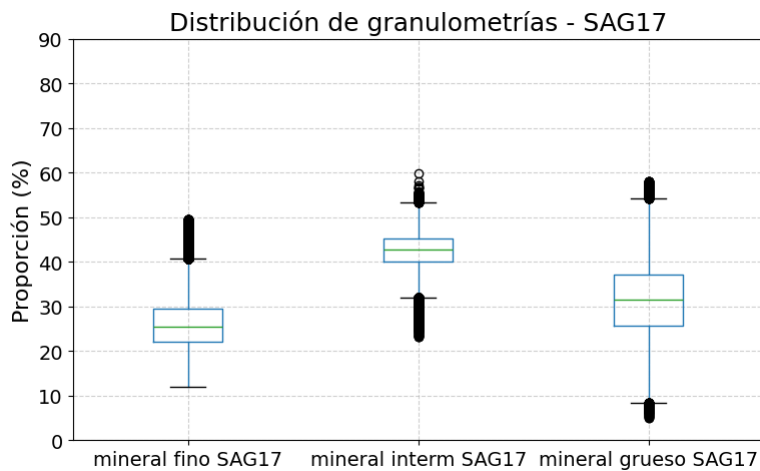


Figura 4. Diagrama de caja de la distribución granulométrica alimentada al SAG 17.

2. Variables Seleccionadas para el Análisis de Rendimiento

La selección de un conjunto de 25 variables y características operacionales fue determinada con base en el criterio experto para el análisis de rendimiento de los molinos semi – autógenos (16 y 17).

Dicho conjunto incluye parámetros fundamentales de alimentación y control de caudal, tales como la velocidad y ganancia de los *feeders* (desde el 3 hasta el 8), la velocidad de los alimentadores de sobre tamaño, y las proporciones granulométricas de la carga (fracciones fino, intermedio y grueso).

Adicionalmente, se integraron variables que describen las condiciones mecánicas y operativas (velocidad de rotación de los molinos, presión de lubricación de los descansos y el ciclo de la campaña del revestimiento), junto con las variables hidrológicas y de

recirculación, como el porcentaje de sólidos, los flujos de agua (manual y automático), y el flujo de *pebbles* recirculados. Finalmente, se incorporaron indicadores contextuales, incluyendo el nivel del *stockpile* A2, el indicador de dureza SPI (*SAG Power Index*), y el estado operacional de los molinos (encendido o apagado). El nombre de cada variable puede consultarse en el Anexo 4 y Anexo 5.

Asimismo, se incorporó la variable de rendimiento de la planta de molienda SAG como variable objetivo (*target*) en el modelo de predicción basado en *XGBoost*. En este caso, el rendimiento se define como la tasa de procesamiento combinada en toneladas por hora (TPH) de ambos molinos SAG. Los rendimientos predichos por este modelo constituyen la variable de entrada del modelo de optimización.

Este conjunto exhaustivo de variables permite el modelado y la evaluación detallada de la influencia de los parámetros operacionales sobre las tendencias identificadas en la granulometría de alimentación, siendo crucial para caracterizar la dinámica de operación de cada SAG.

Los datos recopilados corresponden a un período de dos años. Este intervalo fue definido considerando que, tras la caída del domo de la pila mina el 03 de agosto de 2022, la toma de datos de las variables seleccionadas se vio temporalmente afectada. La estabilización y restablecimiento del sistema de medición se alcanzó meses después del evento, momento a partir del cual se dispone de registros consistentes y confiables para el análisis.

2.1 Evaluación de Multicolinealidad y Redundancia de Variables

El análisis exploratorio de la matriz de correlación de Pearson de la Figura 5 revela un predominio de coeficientes de magnitud baja a moderada, visualizados mediante una baja saturación de color en el mapa de calor (*heatmap*). En el contexto de la modelación estadística, la multicolinealidad severa —que justificaría la reducción de dimensionalidad o la eliminación de variables predictoras— se identifica habitualmente cuando los coeficientes de correlación absoluta ($|r|$) superan los umbrales de 0,80 o 0,90.

2.2 Análisis de Correlaciones Positivas y Dinámica Operativa

Los coeficientes de correlación positivos ($\rho > 0$) denotan relaciones de proporcionalidad directa entre las variables del proceso. Un ejemplo destacado es la asociación entre las velocidades de los alimentadores (*feeders*) 103 y 104, cuantificada en 0,65. Este valor refleja una lógica operativa de balance de carga. Si bien existe una tendencia a sincronizar

el aumento de velocidad en ambas líneas, el coeficiente alejado de la unidad (1.0) indica que operan con grados de libertad independientes, aportando información diferenciada que el algoritmo debe aprender.

Asimismo, se observa una correlación de 0,55 entre la “alimentación al SAG16” y el “flujo de agua manual”. Este comportamiento evidencia una respuesta de control coherente, donde los operadores incrementan la inyección hídrica en respuesta a mayores tasas de alimentación de mineral, manteniendo así la reología adecuada dentro del molino.

2.3 Análisis de Correlaciones Negativas y Restricciones Físicas

Las correlaciones negativas ($p < 0$) indican relaciones inversamente proporcionales, fundamentadas en las restricciones físicas y de balance de masa del sistema. La relación inversa más fuerte se presenta entre las fracciones de “mineral fino” y “mineral grueso” ($\rho = -0,75$). Este comportamiento es intrínseco a la caracterización granulométrica. Dado que las fracciones son porcentajes de una totalidad (100%), el incremento de una implica necesariamente la disminución de la otra. No obstante, la existencia de una fracción de “mineral intermedio” impide que esta correlación sea perfectamente lineal (-1.0), validando la necesidad de incluir las tres variables para una descripción completa del perfil granulométrico.

Adicionalmente, la correlación de -0,57 entre la “presión de lubricación en descansos” y la “velocidad del molino SAG16” describe un fenómeno tribológico y mecánico relevante. La variación inversa entre estas variables sugiere un patrón de comportamiento físico del equipo que resulta crítico para la capacidad del modelo de detectar anomalías operativas o desviaciones en el estado de la maquinaria.

2.4 Robustez del Espacio de Características

La inspección de la matriz confirma la ausencia de bloques de alta colinealidad (representados por colores intensos fuera de la diagonal principal), lo cual constituye un escenario favorable para el entrenamiento de modelos predictivos. Esta diversidad en las relaciones bivariadas indica que el *dataset* posee una alta densidad de información. La eliminación arbitraria de variables aparentemente similares, como la “velocidad del feeder 104”, implicaría una pérdida significativa de varianza explicada única (aproximadamente un 35% de independencia respecto a su par), degradando la capacidad predictiva. En suma, el conjunto de variables seleccionado es estadísticamente robusto y adecuado para proceder a las fases de entrenamiento.

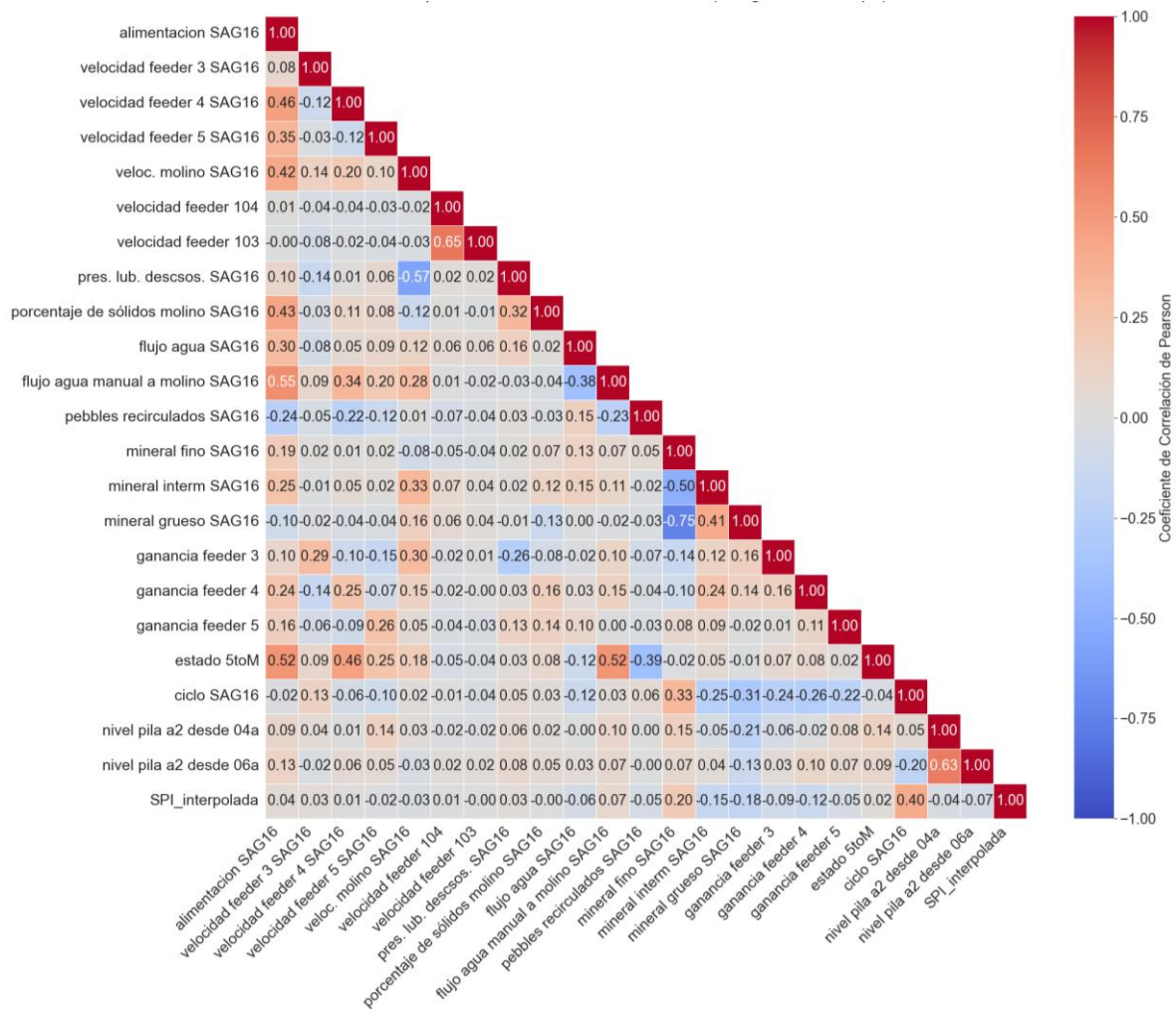


Figura 5. Mapa de calor de correlaciones de Pearson para modelamiento del molino SAG 16.

De manera análoga, se procedió a evaluar la matriz de correlación correspondiente a las variables del modelo predictivo para el SAG17, replicando los criterios de inspección de multicolinealidad y coherencia operativa previamente descritos. Los resultados obtenidos ratifican la robustez estadística de este segundo conjunto de datos, evidenciando una estructura de coeficientes moderados que descarta la redundancia entre predictores y valida la consistencia de las relaciones físicas, tales como la compensación inversa observada entre los alimentadores de la correa. El desglose pormenorizado de estos coeficientes se presenta en detalle en el Anexo 3.

3. Modelamiento Predictivo con XGBoost

3.1 Hiper – parámetros

Los hiperparámetros de los modelos XGBoost con mejor desempeño se presentan en la Tabla 1.

Tabla 1. Hiperparámetros para el modelo XGBoost de los molinos SAG 16 y 17.

Hiperparámetros	Valor
n estimators	1200
max depth	6
learning rate	0,03
subsample	0,8
colsample bytree	0,8
reg alpha	0,5
reg lamda	1
min child weight	3
random state	42

La configuración óptima del algoritmo *XGBoost* fue determinada mediante una búsqueda sistemática dentro de intervalos acotados, destacando la selección de 1200 estimadores en conjunción con una tasa de aprendizaje reducida (*learning rate*) de 0,03. Esta combinación específica sugiere una estrategia de modelado robusta que prioriza la generalización mediante pasos de gradiente pequeños pero numerosos, evitando la convergencia prematura en mínimos locales. Asimismo, la restricción de la profundidad máxima de los árboles a 6 actúa como un mecanismo de regularización estructural crítico, limitando la complejidad del modelo para mitigar el riesgo de sobreajuste (*overfitting*) que inducirían árboles más profundos. Este balance entre sesgo y varianza, reforzado por técnicas de submuestreo estocástico (*subsample* y *colsample* al 0,8), fundamenta matemáticamente la maximización de la función objetivo, logrando explicar un 88,74% de la variabilidad para el Molino SAG 16 y un 80,28% para el SAG 17, validando que la estabilidad predictiva se alcanza mediante una arquitectura conservadora y no simplemente incrementando la complejidad computacional.

Complementariamente a la estructura del árbol, la sintonización fina del algoritmo incorporó mecanismos de regularización rígidos para gestionar la naturaleza ruidosa inherente a los datos de conminución. La selección de un “min_child_weight” de **3** —dentro del intervalo iterado [1-10]— establece un umbral conservador que obliga al algoritmo a descartar nodos

terminales con baja representatividad estadística (pocos datos), evitando así que el modelo aprenda de anomalías transitorias en la señal de los sensores.

Asimismo, la configuración de los términos de penalización “reg_alpha” (L1) en 0,5 y “reg_lambda” (L2) en 1 introduce una restricción dual sobre los pesos de las hojas. Al fijar “reg_alpha” en el límite inferior del rango explorado [0.5-1], se aplica una dispersión moderada (*sparsity*) que ayuda a la selección de características relevantes sin ser excesivamente agresiva, mientras que el valor unitario de “reg_lambda” asegura una suavización gaussiana estándar de los pesos. En conjunto, esta arquitectura de hiperparámetros prioriza la parsimonia del modelo, garantizando que la alta precisión reportada (R^2 aprox. 0,89) sea producto de capturar patrones estructurales genuinos y no artefactos del ruido de proceso.

3.2 Indicadores de Ajuste

La evaluación de las métricas de desempeño del modelo *XGBoost*, presentadas en la Tabla 2, demuestra una capacidad predictiva robusta, aunque heterogénea entre las unidades de conminución. Específicamente para el molino SAG 16, el coeficiente de determinación (R^2) promedio de 0,887 indica que el modelo logra explicar aproximadamente el 89% de la varianza de los datos reales, manteniendo un Error Porcentual Absoluto Medio (MAPE) acotado al 3.21%. En contraste, la modelación para el molino SAG 17 exhibe un ajuste ligeramente inferior, reportando un R^2 de 0,803 y un incremento en el error promedio a un 4,04%, lo que sugiere una mayor complejidad intrínseca en la dinámica operativa de esta segunda unidad.

Tabla 2. Métricas de rendimiento promedio del modelo *XGBoost* en los conjuntos de prueba de los molinos SAG.

Output	RMSE		MAE		MAPE		R ²	
	Prom.	Desv. Est.	Prom.	Desv. Est.	Prom.	Desv. Est.	Prom.	Desv. Est.
Rendimiento SAG 16	113,0	19,85	83,64	14,62	3,21	0,60	0,887	0,04
Rendimiento SAG 17	95,50	20,47	69,83	15,11	4,04	0,79	0,803	0,06

Aunque las métricas globales validan la aptitud del algoritmo para capturar la variabilidad central del proceso, se identifica una limitación en la capacidad de generalización del modelo frente a valores atípicos o extremos. Esta restricción predictiva, sumada a la divergencia en los coeficientes de determinación entre ambos equipos, es atribuible a la naturaleza estocástica de las variables de entrada particulares de cada molino. El

comportamiento más errático o menos lineal en la alimentación del SAG 17 desafía la capacidad de inferencia del modelo en los límites del dominio operativo, resultando en una precisión menor en comparación con la estabilidad observada en el SAG 16.

La inspección visual de las gráficas de dispersión corrobora la jerarquía de desempeño establecida en las métricas cuantitativas. Para el molino SAG 16 (Figura 6), la nube de puntos exhibe una colinealidad estricta en torno a la bisectriz ideal, con una densidad compacta que valida visualmente el coeficiente de determinación de 0,887.

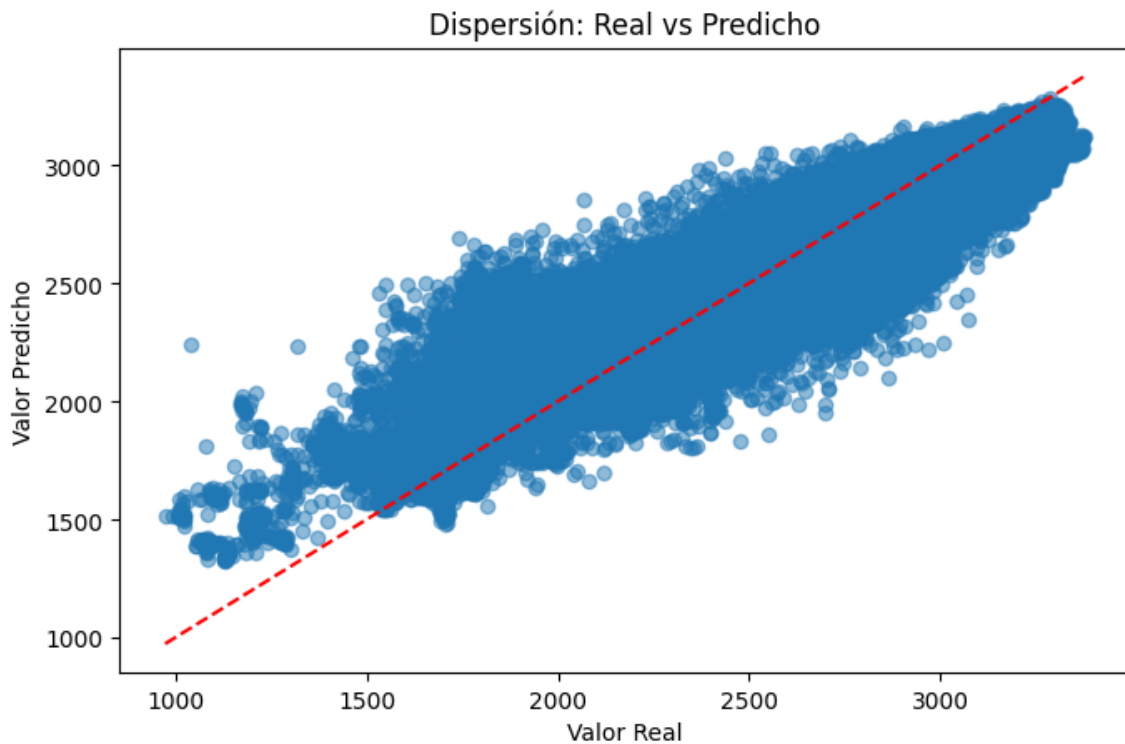


Figura 6. Gráfica de dispersión de rendimiento Real vs Predicho del SAG 16.

La gráfica de comparación temporal (Figura 7) refuerza esta consistencia, evidenciando que la curva predicha (naranja) replica con alta fidelidad la varianza de alta frecuencia y la amplitud de la señal real (azul), lo que justifica el bajo MAPE de 3,21%.

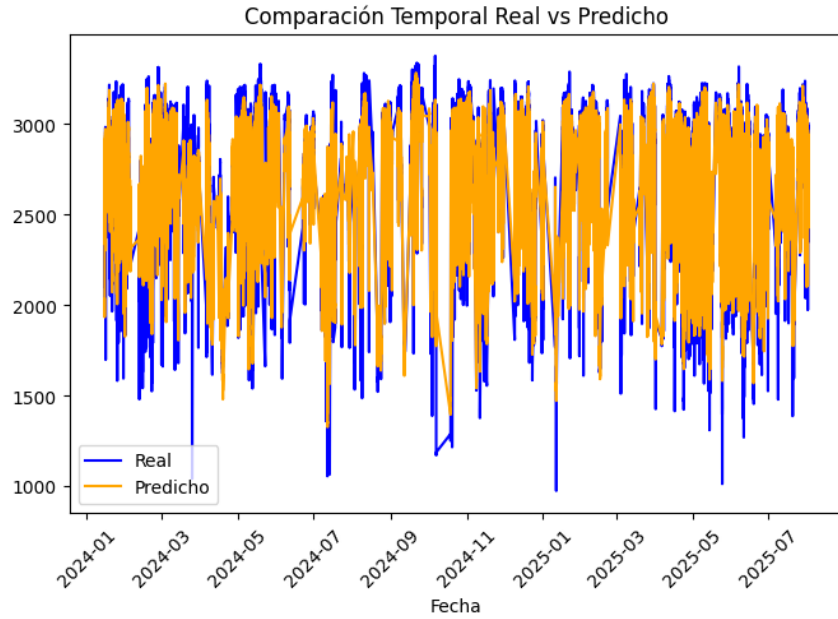


Figura 7. Gráfica de comparación temporal del rendimiento real vs predicho del SAG 16.

Los residuos asociados mantienen una distribución mayormente simétrica y acotada, indicando que el modelo captura eficazmente la dinámica del sistema sin sesgos sistemáticos significativos (Figura 8).

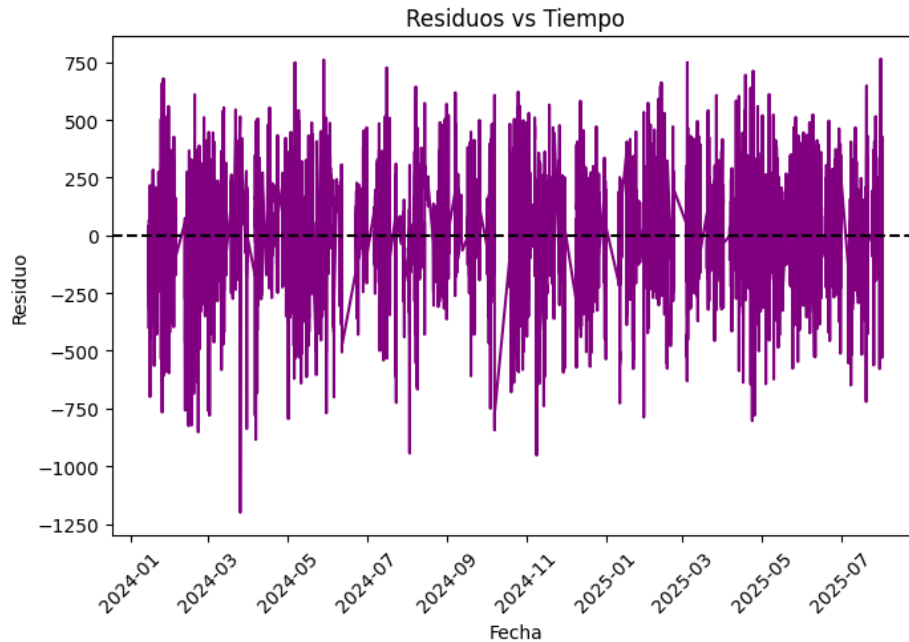


Figura 8. Gráfica de residuos en el tiempo del SAG 16.

Por el contrario, el análisis gráfico del molino SAG 17 evidencia la mayor complejidad estocástica sugerida por su R^2 inferior de 0,803. En el gráfico de dispersión, Figura 9, se observa una mayor excentricidad y holgura de los datos respecto a la línea de ajuste, especialmente en los rangos superiores de alimentación, lo que es congruente con el incremento del MAPE al 4,04% y un RMSE más elevado.

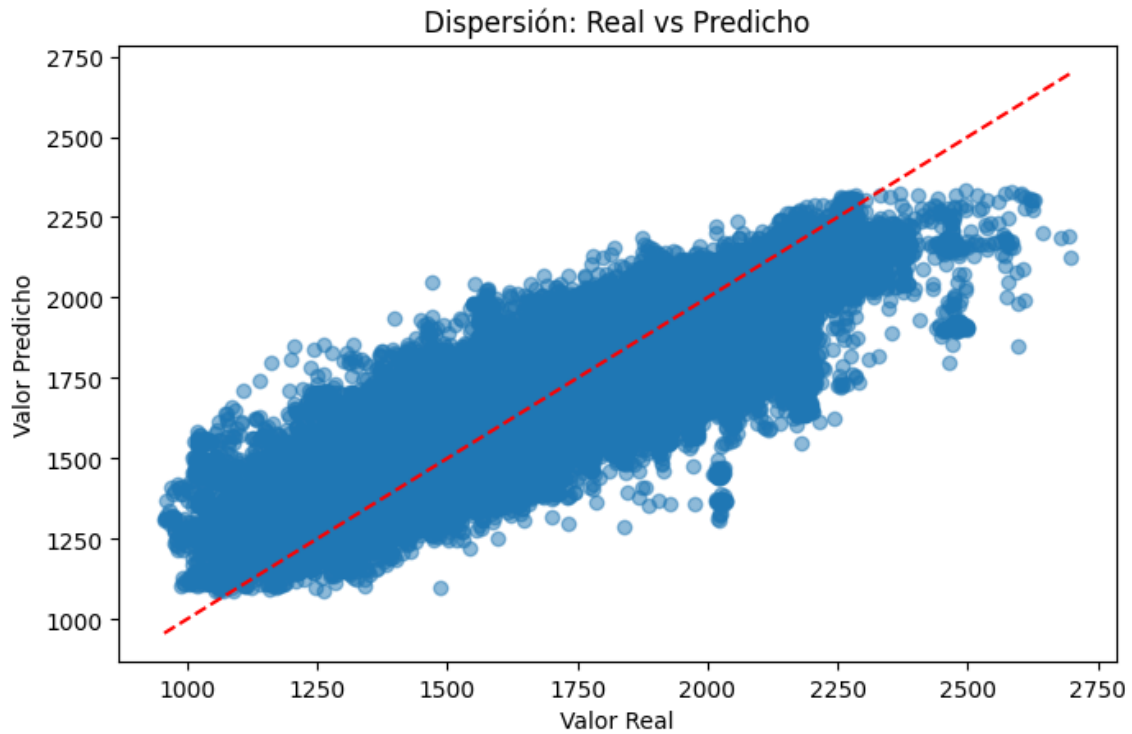


Figura 9. Gráfica de dispersión de rendimiento Real vs Predicho del SAG 17.

Si bien la serie temporal (Figura 10) demuestra que el modelo reproduce adecuadamente la tendencia macroscópica del proceso, evidenciando una correcta captura del comportamiento global de la señal.

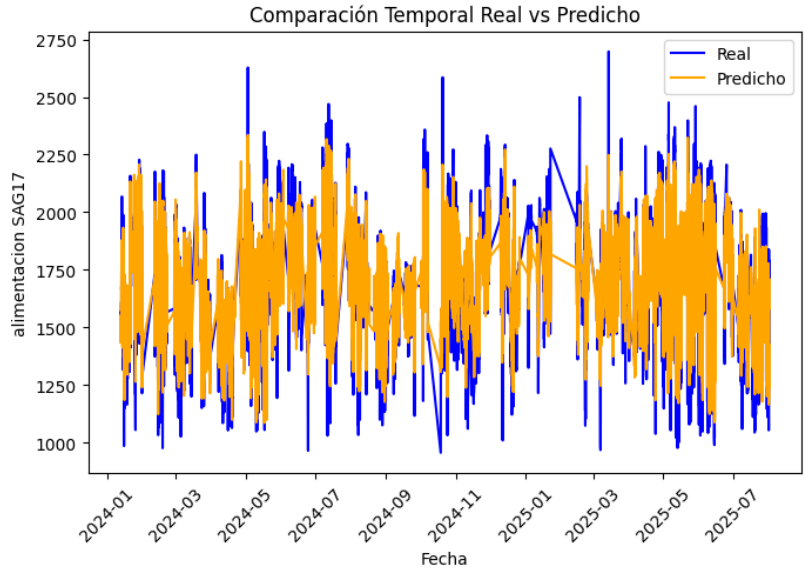


Figura 10. Gráfica de comparación temporal del rendimiento real vs predicho del SAG 17.

En tanto, las excursiones de mayor magnitud observadas en la gráfica de residuos (Figura 11) pueden atribuirse a una menor capacidad del modelo para capturar dinámicas transitorias propias de la unidad 16. Este comportamiento suele asociarse a variaciones operacionales abruptas —como cambios en la dureza del mineral, fluctuaciones en la alimentación o ajustes operacionales— que introducen componentes no lineales difíciles de representar mediante el modelo entrenado.

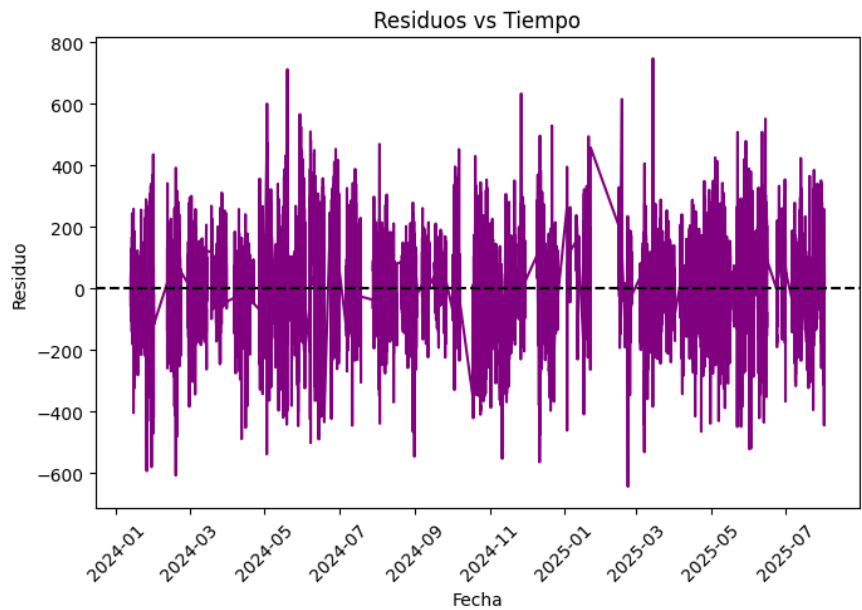


Figura 11. Gráfica de residuos en el tiempo del SAG 17.

3.3 Shapley Additive Explanations Values

El análisis de interpretabilidad de los modelos mediante valores SHAP revela una convergencia estructural en los predictores críticos para ambas unidades de conminución, identificando un núcleo común de cuatro variables determinantes: el porcentaje de sólidos, los flujos de agua (tanto de adición manual como automática) y la tasa de alimentación histórica. Si bien este conjunto de parámetros hidrodinámicos y másicos gobierna la predicción del rendimiento en ambos casos, la jerarquía de su impacto relativo difiere sustancialmente, denotando sensibilidades operativas distintas para cada molino.

Específicamente, en el modelo del SAG 16 (Figura 12a), la influencia es predominantemente reológica, posicionando al porcentaje de sólidos como la característica de mayor peso predictivo (con un valor SHAP superior a 120), seguida inmediatamente por el flujo de agua, lo que indica que el rendimiento de esta unidad es altamente dependiente de la densidad de la pulpa. Por el contrario, en el SAG 17, Figura 12b, la dinámica inercial cobra primacía; la variable dominante es la alimentación promediada con un rezago de una hora (“alimentación SAG17_lag1h_avg”), desplazando a los factores hidráulicos a un segundo plano. Esta distinción sugiere que, mientras la eficiencia del SAG 16 se maximiza controlando la viscosidad interna, el SAG 17 responde más fuertemente a la estabilidad y magnitud de su carga másica antecedente.

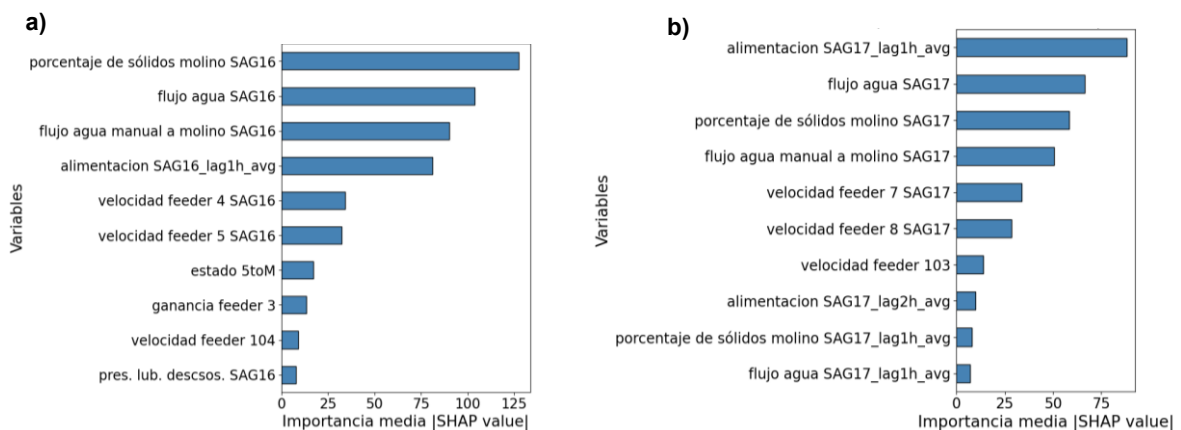


Figura 12. Características más importantes. a) SHAP de SAG 16. b) SHAP de SAG 17.

4. Modelamiento de Optimización *Differential Evolution*

El algoritmo de Evolución Diferencial se aplicó para determinar las proporciones granulométricas óptimas de alimentación en los molinos semi-autógenos, con el objetivo de maximizar el rendimiento sin comprometer la operación conjunta del circuito.

Para el SAG 16, la configuración óptima corresponde a 51,6% de finos, 28,6% de intermedios y 19,8% de gruesos (Anexo 6a). Este resultado demuestra que el rendimiento no crece proporcionalmente con el aumento de finos, contradiciendo el supuesto operacional tradicional. El óptimo se sitúa cercano a la tendencia central histórica ($Q2 \approx 50\%$), evidenciando que operar cerca del máximo histórico de finos (82%) resulta innecesario e ineficiente. En consecuencia, el modelo permite reducir la fracción intermedia hacia el cuartil inferior sin afectar el tonelaje procesado, estableciendo un punto operativo eficiente basado en equilibrio granulométrico y no en maximización de finos.

En contraste, el SAG 17 presenta una estrategia operativa distinta. La solución óptima obtenida de 38,8% de finos, 15,6% de intermedios y 45,6% de gruesos (Anexo 6b) implica un cambio de régimen respecto de su comportamiento histórico. El nivel de finos sitúa al molino cercano a su límite máximo registrado (42%), muy por sobre su rango habitual ($Q3 \approx 30\%$), mientras que la fracción intermedia propuesta se ubica incluso bajo el mínimo histórico observado (32%). Esta condición justifica la redistribución de material fino acumulado en los feeders 3, 4 y 5 hacia los feeders 6, 7 y 8, permitiendo desplazar al SAG 17 desde su zona recurrente de bajo desempeño (mediana ≈ 1800 TPH) hacia una nueva zona de eficiencia operacional.

En conjunto, el modelo de Evolución Diferencial identifica composiciones de carga diferenciadas para cada molino, revelando una estrategia dicotómica de operación que incrementa el rendimiento global optimizando cada unidad de manera independiente y sin generar desmedro entre ellas.

Con las proporciones granulométricas óptimas definidas, se proyecta un rendimiento de 3274,5 TPH para el SAG 16 y 2324,4 TPH para el SAG 17, alcanzando un rendimiento conjunto estimado de 5600,9 TPH. Estos resultados representan un incremento aproximado de 3,72% respecto de los límites actuales de operación, establecidos en 3200 TPH para el SAG 16 y 2200 TPH para el SAG 17, evidenciando el potencial del esquema optimizado para expandir la capacidad efectiva de procesamiento del circuito (Figura 13a).

El modelo reconoce además la mayor contribución relativa del SAG 16 dentro del sistema, como se muestra en la Figura 13b, concentrando aproximadamente el 58,5% del procesamiento total, mientras que el SAG 17 asume un rol complementario dentro de la estrategia global de maximización del *throughput*, permitiendo aumentar el rendimiento conjunto sin generar desbalance operacional entre ambas unidades.

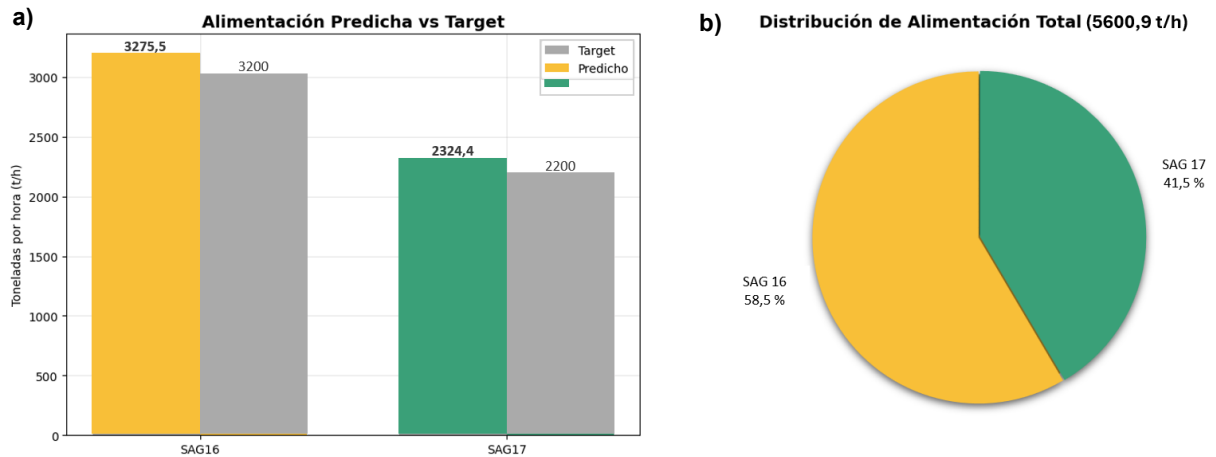


Figura 13. Resultados de modelo de optimización DE. a) gráfica de alimentación predicha vs actual. b) gráfica de distribución de alimentación total.

5. Generación de Recomendaciones

A partir de la configuración óptima definida por el algoritmo de optimización, el sistema transforma los resultados numéricos en recomendaciones operacionales aplicables al circuito SAG, integrando las etapas previas en un sistema orientado a optimizar el rendimiento de la molienda.

El proceso iterativo del algoritmo *Differential Evolution* genera un conjunto de soluciones cercanas al óptimo global, las cuales son posteriormente filtradas según criterios operacionales adicionales, tales como estabilidad del proceso, coherencia metalúrgica y factibilidad de implementación en planta. De esta manera, las recomendaciones finales no corresponden únicamente al máximo matemático del modelo, sino a configuraciones operables bajo condiciones reales.

Las proporciones granulométricas óptimas obtenidas por el algoritmo se traducen operacionalmente en ajustes en las velocidades de los feeders de alimentación, variable mediante la cual se controla la participación relativa de cada fracción granulométrica en la carga del molino. En la operación del circuito, los feeders regulan el flujo proveniente de los distintos orígenes de mineral, determinando el blend de alimentación que ingresa a los molinos SAG.

Dado que la variable objetivo del modelo corresponde al rendimiento de la molienda (t/h), las modificaciones en el blend de alimentación impactan directamente en las condiciones de conminución dentro del molino, influyendo en la eficiencia de fractura del mineral y, en consecuencia, en la capacidad de tratamiento del circuito. De esta forma, el ajuste de las velocidades de los feeders constituye el mecanismo operacional mediante el cual las configuraciones optimizadas del modelo se traducen en incrementos potenciales del throughput del sistema.

En este contexto, el sistema no actúa como un controlador automático, sino como una herramienta de apoyo a la toma de decisiones operacionales, entregando escenarios comparativos que cuantifican el beneficio potencial asociado a cada modificación propuesta.

Con esta etapa se establece el vínculo entre el modelamiento predictivo-optimizado y la aplicación práctica en operación, habilitando un esquema continuo donde nuevas condiciones de proceso pueden ser evaluadas, optimizadas y reinterpretadas de forma sistemática dentro del mismo marco analítico.

6. Evaluación Económica de la implementación de ML en el rendimiento de los SAG

La implementación del sistema de optimización basado en *Machine Learning* permite proyectar un aumento del rendimiento conjunto del circuito de molienda SAG desde 5400 tph hasta 5600,9 tph, equivalente a un incremento operacional de 3,72%. Considerando un régimen de operación anual de 8000 horas, este aumento representa un incremento en el mineral procesado desde 43,2 Mt/año a 44,8 Mt/año, equivalente a 1,6 Mt adicionales tratadas anualmente.

Bajo las condiciones metalúrgicas definidas —correspondientes a una ley de cabeza promedio de 0,77% y una recuperación global promedio de 80,8%, determinadas en el instante operacional en que el modelo de optimización identifica el mayor rendimiento del circuito— la producción anual de cobre fino en el escenario base se estimó en aproximadamente 270 ktpa. Con la implementación del sistema optimizado, la producción proyectada alcanza aproximadamente 279,9 ktpa, lo que representa un incremento cercano a 9,9 ktpa de cobre fino adicional, equivalente a aproximadamente 21,8 millones de libras de cobre.

Utilizando un precio de referencia del cobre de 4,30 USD/lb, el aumento de producción genera un ingreso incremental estimado del orden de 93,7 millones de dólares anuales, atribuible exclusivamente al desbloqueo de capacidad existente del circuito de molienda sin incorporación de nuevo equipamiento.

Desde la perspectiva de costos, el incremento de producción produce una reducción estructural del costo C1 debido a la dilución de costos fijos sobre un mayor volumen de cobre producido. Bajo indicación de una estructura operacional compuesta por 70% de costos fijos y 30% de costos variables, el análisis revela una disminución aproximada del costo C1 del orden de 2,6% respecto del escenario base.

Para una operación con un costo C1 inicial de 1,50 USD/lb, la optimización permitiría reducir el costo unitario hasta aproximadamente 1,46 USD/lb. Considerando una producción anual cercana a 617 millones de libras de cobre fino, esta reducción equivale a un beneficio económico adicional aproximado de 24 millones de dólares por año asociado a eficiencia operacional. La Tabla 3 presenta un resumen del impacto generado por la aplicación de aprendizaje automático en comparación con la condición operacional actual.

Tabla 3. Impacto productivo de la optimización del rendimiento SAG mediante Machine Learning.

Variable	Actual	Con ML	Impacto	Beneficio
Rendimiento SAG	5400 tph	5600,9 tph	+3,72%	+24,1MUSD/año
Mineral tratado	43,2 Mtpa	44,8 Mtpa	+1,6 Mt	-
Producción Cu fino	270 ktpa	280 ktpa	+9,9 kt	-
Producción	595 Mlb	617 Mlb	+21,8 Mlb	+93,7 MUSD/año

En conjunto, los resultados demuestran que la optimización inteligente de la alimentación a los molinos SAG genera valor económico significativo mediante dos mecanismos complementarios: incremento directo de producción y reducción del costo unitario de operación. Ambos efectos se originan en el aumento sostenido del *throughput* del circuito de molienda, confirmando que la aplicación de técnicas de aprendizaje automático constituye una estrategia efectiva para desbloquear capacidad productiva utilizando infraestructura existente.

CAPÍTULO VI: CONCLUSIONES Y RECOMENDACIONES

1. Conclusiones

La aplicación de analítica avanzada en el circuito de molienda SAG permitió demostrar que el análisis de datos históricos puede utilizarse de manera efectiva para apoyar la toma de decisiones operacionales en plantas concentradoras. A partir del levantamiento y análisis de información histórica de operación, fue posible identificar relaciones entre variables operacionales y el rendimiento del circuito, evidenciando la influencia de la composición granulométrica de la alimentación y de las condiciones de operación del molino en la capacidad de tratamiento. El análisis de interpretabilidad mediante valores SHAP permitió cuantificar la contribución de cada variable en el modelo, lo que permitió establecer relaciones consistentes entre las variables operacionales y el rendimiento de la molienda SAG, dando cumplimiento al primer objetivo específico del estudio.

El modelo predictivo desarrollado mediante *XGBoost* logró representar adecuadamente las relaciones no lineales presentes en el proceso de conminución, permitiendo proyectar el comportamiento del rendimiento del circuito en función de las condiciones de alimentación del mineral y de las variables operacionales asociadas al control de la velocidad de los *feeders*. El desempeño del modelo evidenció un buen nivel de ajuste, alcanzando coeficientes de determinación promedio de $R^2 = 0,887$ para SAG 16 y $R^2 = 0,803$ para SAG 17, junto con errores porcentuales medios (MAPE) de 3,21% y 4,04%, respectivamente, lo que indica una adecuada capacidad para explicar la variabilidad del rendimiento del circuito y mantener errores relativos acotados. De esta forma, el modelo constituye una herramienta capaz de anticipar el desempeño del circuito bajo distintos escenarios operacionales, dando cumplimiento al segundo objetivo específico de desarrollar un modelo predictivo para la molienda SAG.

La integración del modelo predictivo con el algoritmo de optimización *Differential Evolution* permitió identificar configuraciones operacionales que maximizan el rendimiento del circuito a partir de la definición de proporciones granulométricas óptimas en la alimentación al molino SAG. Bajo estas condiciones, el rendimiento del circuito podría incrementarse desde aproximadamente 5400 tph hasta 5600,9 tph, lo que representa un aumento cercano al 3,7% en la capacidad de tratamiento. Estos resultados permiten validar la hipótesis planteada en esta investigación, la cual establecía que la aplicación de técnicas de *Machine Learning* sobre datos operacionales del proceso de molienda permite estabilizar y optimizar

la curva de distribución granulométrica al interior del molino. De esta forma, al ajustar dicha distribución hacia una condición óptima, es posible aumentar el rendimiento del proceso y mejorar su estabilidad operacional. Asimismo, el desarrollo de este esquema de optimización permite cumplir el tercer objetivo específico del estudio.

El incremento proyectado en la capacidad de procesamiento se traduce en un aumento del mineral tratado desde aproximadamente 43,2 Mtpa a 44,8 Mtpa. Bajo las condiciones metalúrgicas definidas para el análisis —correspondientes a una ley de cabeza promedio de 0,77% Cu y una recuperación global de 80,8%— este incremento permitiría aumentar la producción anual de cobre fino desde aproximadamente 270 ktpa hasta 280 ktpa, equivalente a cerca de 9,9 kt adicionales de cobre fino por año.

Desde el punto de vista económico, este aumento productivo representa aproximadamente 21,8 millones de libras adicionales de cobre fino anuales, lo que equivale a un beneficio estimado del orden de 93,7 millones de dólares por año. A ello se suma un beneficio económico asociado a la reducción del costo C1 cercano a 24,1 millones de dólares anuales. Estos resultados evidencian que la incorporación de herramientas de analítica avanzada en la operación de plantas concentradoras puede generar un impacto económico significativo mediante mejoras relativamente acotadas en el rendimiento del proceso, dando cumplimiento al cuarto objetivo específico del estudio.

En conjunto, los resultados obtenidos demuestran que los modelos basados en datos pueden complementar el criterio experto en la operación de circuitos de molienda, permitiendo identificar oportunidades de optimización que contribuyen a mejorar la eficiencia operativa y reducir la variabilidad del proceso. En este contexto, la metodología desarrollada constituye una base para la incorporación de sistemas de apoyo a la decisión basados en analítica avanzada en la operación de plantas concentradoras de gran escala.

2. Recomendaciones

Evaluar la implementación de la metodología propuesta en sistemas de apoyo a la operación o plataformas de analítica operacional que permitan entregar recomendaciones en tiempo cercano al real para la gestión del circuito de molienda.

Desarrollar mecanismos de integración entre el modelo de optimización y las variables manipulables de la operación, particularmente mediante el ajuste de la velocidad de los *feeders*, con el fin de aproximar en operación las proporciones granulométricas recomendadas por el modelo.

Ampliar el análisis incorporando periodos de operación más extensos y diferentes condiciones mineralógicas, con el objetivo de mejorar la robustez y generalización del modelo predictivo.

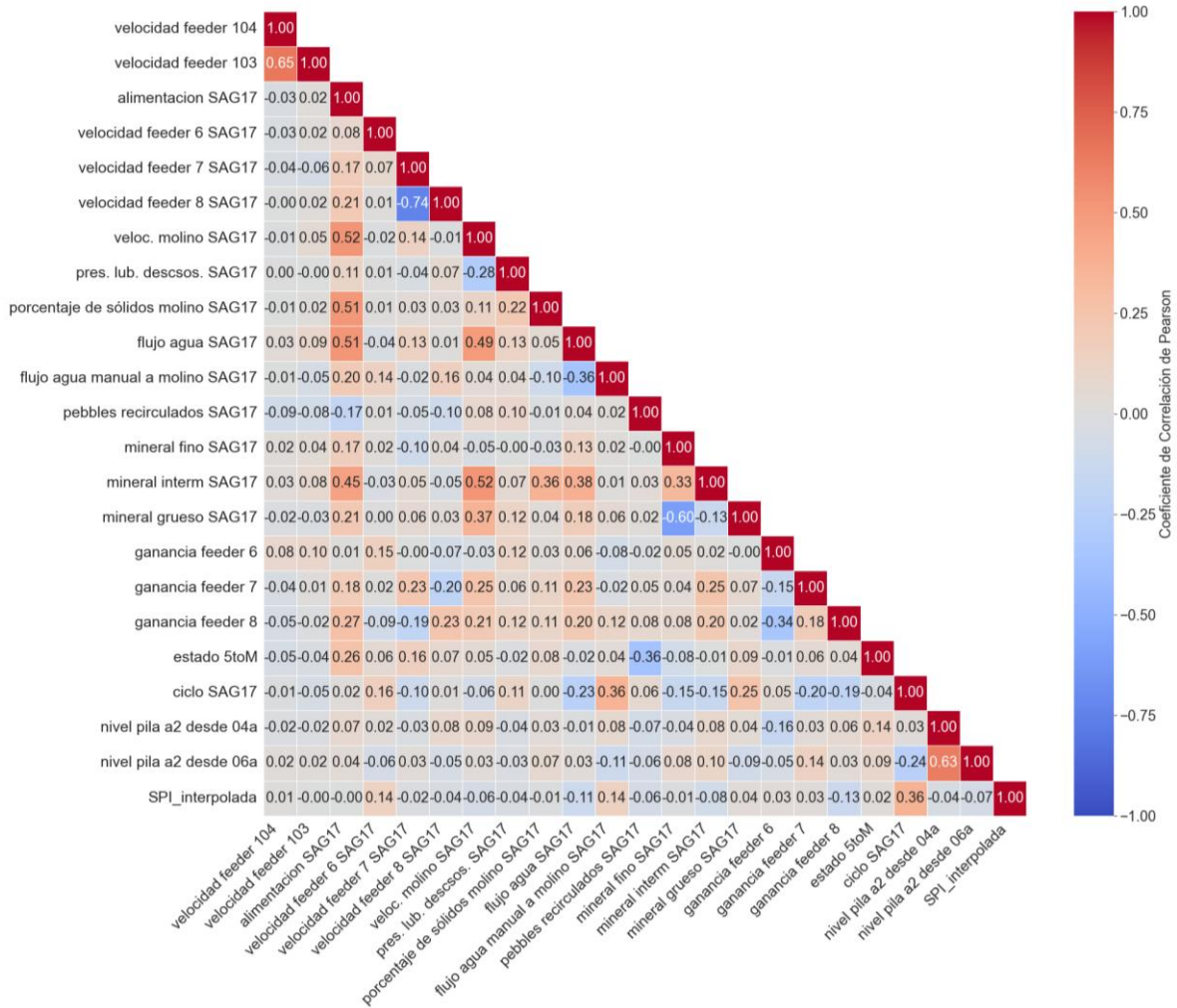
Evaluar la incorporación de variables adicionales relacionadas con características geo – metalúrgicas del mineral, que puedan mejorar la capacidad predictiva del modelo y su aplicabilidad en escenarios operacionales más complejos.

REFERENCIAS

- [1] Z. Ferrofin and D. Martin, "Evaluación de la Influencia de las Variables en la Distribución Granulométrica del Producto de Molienda por Bolas de Minerales Mediante Diseños Experimentales," Universidad Nacional de San Agustín, Facultad de Ingeniería de Procesos, Escuela Profesional de Ingeniería Metalúrgica, Arequipa, Perú, 2017.
- [2] J. T. McCoy and L. Auret, "Machine learning applications in minerals processing: A review," Mar. 01, 2019, *Elsevier Ltd*. doi: 10.1016/j.mineng.2018.12.004.
- [3] M. Saldaña, E. Gálvez, A. Navarra, N. Toro, and L. A. Cisternas, "Optimization of the SAG Grinding Process Using Statistical Analysis and Machine Learning: A Case Study of the Chilean Copper Mining Industry," *Materials*, vol. 16, no. 8, Apr. 2023, doi: 10.3390/ma16083220.
- [4] M. Pistilli, "Global Trends Today," Top 10 Copper-producing Companies (Updated 2024).
- [5] CODELCO, "CODELCO," Codelco cierra 2024 con Ebitda de US\$ 5.439 millones y un aporte al Fisco de US\$ 1.534 millones.
- [6] Y. Alpaca, "CARACTERIZACION GEOTECNICA, ESTRUCTURAL DEL MACIZO ROCOSO DE LOS DESARROLLOS SUBTERRANEOS DE LA FASE V, CHUQUICAMATA-CHILE, Tesis Para optar el Título Profesional de: Ingeniero Geólogo," UNIVERSIDAD NACIONAL DE SAN AGUSTÍN DE AREQUIPA FACULTAD DE GEOLOGÍA, GEOFÍSICA Y MINAS ESCUELA PROFESIONAL DE INGENIERÍA GEOLÓGICA, Perú, 2016.
- [7] Z. Ghasemi, F. Neumann, M. Zanin, J. Karageorgos, and L. Chen, "A comparative study of prediction methods for semi-autogenous grinding mill throughput," *Miner Eng*, vol. 205, Jan. 2024, doi: 10.1016/j.mineng.2023.108458.
- [8] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Association for Computing Machinery, Aug. 2016, pp. 785–794. doi: 10.1145/2939672.2939785.

- [9] S. Lundberg and S.-I. Lee, "A Unified Approach to Interpreting Model Predictions," Nov. 2017, [Online]. Available: <http://arxiv.org/abs/1705.07874>
- [10] Y. E. Pural, T. Ledezma, M. Hilden, G. Forbes, F. Boylu, and M. Yahyaei, "Application of Machine Learning for Generic Mill Liner Wear Prediction in Semi-Autogenous Grinding (SAG) Mills," *Minerals*, vol. 14, no. 12, Dec. 2024, doi: 10.3390/min14121200.
- [11] R. Storn and K. Price, "Differential Evolution-A Simple and Efficient Heuristic for Global Optimization over Continuous Spaces," Kluwer Academic Publishers, 1997.
- [12] D. Craft, "Differential Evolution: a stochastic nonlinear optimization algorithm by Storn and Price, 1996 This presentation is based on: Storn, Rainer, and Kenneth Price. Differential Evolution-A Simple and Efficient Heuristic for Global Optimization over Continuous Spaces," 2003.
- [13] T. Eltaeib and A. Mahmood, "Differential Evolution: A Survey and Analysis," *Applied Sciences (Switzerland)*, vol. 8, no. 10, Oct. 2018, doi: 10.3390/app8101945.
- [14] F. C. Bond, *The Third Theory Of Comminution*. The American Institute of Mining, Metallurgical, and Petroleum Engineers, 1952.
- [15] Z. Ghasemi, M. Neshat, C. Aldrich, M. Zanin, and L. Chen, "Optimising SAG mill throughput and circulating load using machine learning models: A multi-objective approach for identifying optimal process parameters," *Miner Eng*, vol. 232, Oct. 2025, doi: 10.1016/j.mineng.2025.109551.
- [16] T. Ou, J. Liu, F. Liu, W. Chen, and J. Qin, "Coupling of XGBoost ensemble methods and discrete element modelling in predicting autogenous grinding mill throughput," *Powder Technol*, vol. 422, May 2023, doi: 10.1016/j.powtec.2023.118480.
- [17] M. Silva Balocchi, "Variables Operacionales y su Influencia en la Operación, Pro-Optimización del Consumo Energético en la Molienda SAG," Universidad de Chile, Facultad de Ciencias Físicas y Matemáticas, Departamento de Ingeniería de Minas, Santiago de Chile, 2013.
- [18] N. Adhikari, *Machine Learning for Identification of Operating States in SABC Circuits*. The University of Queensland, Australia, 2023.

- [19] Z. Ghasemi *et al.*, “An integrated intelligent framework for maximising SAG mill throughput: Incorporating expert knowledge, machine learning and evolutionary algorithms for parameter optimisation,” *Miner Eng*, vol. 212, Jul. 2024, doi: 10.1016/j.mineng.2024.108733.
- [20] Y. Wang, X. Chen, W. Gui, C. Yang, L. Caccetta, and H. Xu, “A hybrid multiobjective differential evolution algorithm and its application to the optimization of grinding and classification,” *J Appl Math*, vol. 2013, 2013, doi: 10.1155/2013/841780.
- [21] S. Mhlanga, J. Ndlovu, C. Mbohwa, and M. Mutingi, “Design of comminution circuits for improved productivity using a multi-objective evolutionary algorithm (MOEA),” in *2011 IEEE International Conference on Industrial Engineering and Engineering Management*, South Africa, Zimbabwe, Singapore: IEEE, Dec. 2011, pp. 1680–1684. doi: 10.1109/IEEM.2011.6118202.
- [22] S. Huband, L. Barone, P. Hingston, L. While, D. Tuppurainen, and R. Bearman, “Designing Comminution Circuits with a Multi-Objective Evolutionary Algorithm,” in *2005 IEEE Congress on Evolutionary Computation*, Australia: IEEE, 2005, pp. 1815–1822. doi: 10.1109/CEC.2005.1554908.
- [23] J. Sun, X. Liu, T. Bäck, and Z. Xu, “Learning adaptive differential evolution algorithm from optimization experiences by policy gradient,” Feb. 2021, [Online]. Available: <http://arxiv.org/abs/2102.03572>
- [24] Z. Ge, Z. Song, and F. Gao, “Review of recent research on data-based process monitoring,” Mar. 13, 2013. doi: 10.1021/ie302069q.
- [25] P. Kadlec, B. Gabrys, and S. Strandt, “Data-driven Soft Sensors in the process industry,” Apr. 21, 2009. doi: 10.1016/j.compchemeng.2008.12.012.
- [26] C. M. Bishop, *Pattern Recognition and Machine Learning*. Cambridge, U.K.: Springer Science & Business Media, 2006.
- [27] Kevin P. Murphy, *Machine Learning: A Probabilistic Perspective*. Cambridge, Massachusetts: Massachusetts Institute of Technology, 2012.
- [28] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning Data Mining, Inference, and Prediction*, vol. 2. Standford, California: Springer, 2009.



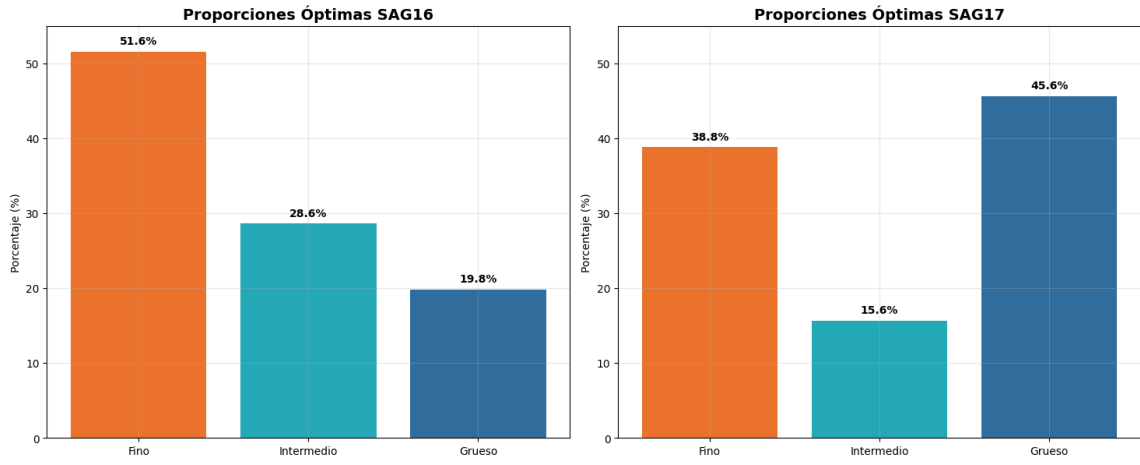
Anexo 3. Mapa de calor de correlaciones de Pearson para modelamiento del molino SAG 17.

Anexo 4. Variables seleccionadas para modelamiento de molino SAG 16.

N°	Descripción	Unidad	Límite	
			Inferior	Superior
1	velocidad feeder 3 sag-16	%	0	100
2	velocidad feeder 4 sag-16	%	0	100
3	velocidad feeder 5 sag-16	%	0	100
4	veloc. molino sag16	rpm	7	10,8
5	velocidad feeder 104	%	0	100
6	velocidad feeder 103	%	0	100
7	pres. lub. descsos. Sag1	kg/cm2	59	67,7
8	porcentaje de sólidos molino sag-16	%	72	80
9	flujo agua a molino sag-16	m3/h	0	600
10	flujo agua manual a molino sag-16	m3/h	0	220
11	pebbles recirculados sag 16	TPH	0	550
12	mineral fino sag16	%	0	100
13	mineral interm sag16	%	0	100
14	mineral grueso sag16	%	0	100
15	nivel pila a2 desde 04A	%	70	100
16	nivel pila a2 desde 06A	%	70	100
17	ganancia feeder 3	%	0	100
18	ganancia feeder 4	%	0	100
19	ganancia feeder 5	%	0	100
20	ciclo campaña revestimientos S16	Ton	0	9770000
21	SPI output	-	3	4,6
22	estado molino sag16	-	OFF	ON
23	estado molino bolas 16a	-	OFF	ON
24	estado molino bolas 16b	-	OFF	ON
25	estado 5to molino	-	OFF	ON

Anexo 5. Variables seleccionadas para modelamiento de molino SAG 17.

N°	Description	Unit	Límite	
			Inferior	Superior
1	velocidad feeder 6 sag-17	%	0	100
2	velocidad feeder 7 sag-17	%	0	100
3	velocidad feeder 8 sag-17	%	0	100
4	veloc. molino sag17	rpm	7	10,8
5	velocidad feeder 104	%	0	100
6	velocidad feeder 103	%	0	100
7	pres. lub. descsos. Sag2	kg/cm2	58	71,7
8	porcentaje de sólidos molino sag-17	%	72	80
9	flujo agua a molino sag-17	m3/h	0	600
10	flujo agua manual a molino sag-17	m3/h	0	220
11	pebbles recirculados sag 17	TPH	0	550
12	mineral fino sag17	%	0	100
13	mineral interm sag17	%	0	100
14	mineral grueso sag17	%	0	100
15	nivel pila a2 desde 04A	%	70	100
16	nivel pila a2 desde 06A	%	70	100
17	ganancia feeder 6	%	0	100
18	ganancia feeder 7	%	0	100
19	ganancia feeder 8	%	0	100
20	ciclo campaña revestimientos S17	Ton	0	5900000
21	SPI output	-	3	4,6
22	estado molino sag17	-	OFF	ON
23	estado molino bolas 17a	-	OFF	ON
24	estado molino bolas 17b	-	OFF	ON
25	estado 5to molino	-	OFF	ON



Anexo 6. Resultados de proporciones granulométricas óptimas de alimentación de los SAG. a) Proporciones SAG 16. b) Proporciones SAG 17.