

UNIVERSIDAD DE CONCEPCIÓN
Facultad de Ingeniería
Departamento de Ingeniería Metalúrgica

Profesor Patrocinante
Dr. Luver Echeverry

Ingeniero Supervisor
Sebastián Gallegos
Lilian Ávila

**“EVALUACIÓN DE UN MODELO PREDICTIVO DE LA PLANTA
DE MOLIENDA CONVENCIONAL “A1” DIVISIÓN
CHUQUICAMATA, CODELCO”**

JOSÉ DAVID VALENZUELA ALARCÓN

Informe de memoria de título

Para optar al Título de

Ingeniero Civil Metalúrgico

Marzo, 2026

AGRADECIMIENTOS

Quiero expresar mi más profundo agradecimiento a mi madre, Alicia Alarcón, por su apoyo incondicional a lo largo de toda mi vida. Su fortaleza para enfrentar las dificultades, su entrega diaria y su cariño constante han sido un pilar fundamental. La admiro con todo mi corazón y no concibo haber podido llegar hasta aquí sin su acompañamiento y sus enseñanzas.

También me agradezco a mí mismo por la disciplina que me ha definido y que, en gran medida, heredé y aprendí de mi madre. El entrenamiento y el deporte han sido fundamentales en mi formación personal, la cual me han enseñado a superar barreras, a mantener la constancia en los momentos difíciles y a cuidarme como persona. Esa rutina de esfuerzo y perseverancia no solo forjó mi condición física, sino que fortaleció mi carácter y mi capacidad para asumir retos académicos y personales.

Agradecer a Chuquicamata por la oportunidad de desarrollar este trabajo en sus instalaciones y por el trato recibido. En particular, agradezco al gerente de Suministros y Servicios, Alexis Rivas, por sus consejos y apoyo. Mi reconocimiento también para el equipo de la gerencia concentradora, en especial a Guillermo Riffo, Lilian Ávila, Carlos Terraza y Sebastián Gallegos, por su profesionalismo cordialidad y buena onda durante todo el proceso.

Agradecer al área de Logística de la molienda, con quienes compartí oficina y vivencias durante bastantes meses. Agradezco especialmente a Norman Arqueros por su gran carisma, sus consejos su buena disposición y las muy buenas anécdotas, y a Luis Gonzales conocido como “El Chivato” por su experiencia en terreno y su apoyo constante en el área. Valoro profundamente cada salida a terreno que permitieron comprender cómo funciona una planta en la práctica. Asimismo, agradezco a los operadores de la planta A1, en particular a Luis Vásquez y Eduardo Flores, por compartir su experiencia y enseñanzas en el día a día.

Finalmente, agradezco al profesor Luver Echeverry, quien fue mi profesor guía en la etapa final de la carrera. Su orientación, observaciones y acompañamiento académico fueron clave para culminar esta memoria.

RESUMEN

La División Chuquicamata procesa mineral por distintas vías de conminución, siendo molienda convencional (plantas “A0” y “A1”) y SAG (planta “A2”). El control operativo de la molienda convencional se apoya en el sistema experto PROFIT; sin embargo, la necesidad de contar con herramientas complementarias de diagnóstico y predicción motivó el desarrollo y evaluación de un modelo predictivo del rendimiento (tph) aplicado a la planta A1 (secciones 13, 14 y 15). El objetivo principal fue evaluar la factibilidad de un modelo predictivo útil para la toma de decisiones y cuantificar la influencia de las principales señales operacionales sobre el tonelaje procesado.

Para caracterizar la alimentación se realizó un análisis detallado del *blending mineral* (fuentes de alimentación: rajo, subterráneo y RT), la granulometría respecto al estándar de 12.700 micrones (P80) y el *Work Index* (WI) horario. Se observó una clara relación entre mayor participación del rajo y aumentos en WI; el WI promedio del período analizado fue de 13,3 kWh/t y el 81 % de los registros se concentró en rangos de 12.6 a 14.0 kWh/t. La base de datos utilizada para el modelo está compuesta por variables operacionales registradas por minuto del año 2024 (527040 registros). El preprocesado incluyó: filtro IQR inicial, filtrado manual y un modo “estable”. Además, se estudiaron desfases temporales (*lags*) para identificar las variables adelantadas o rezagadas respecto al rendimiento. Tras la limpieza, las filas válidas por sección fueron 49.61%, 62.64% y 69.27% del total de datos para las secciones 13, 14 y 15 respectivamente.

Se implementó una estrategia comparativa entre un modelo lineal (OLS) y no lineal (Random Forest). Las métricas usadas para la evaluación fueron R^2 , MAE, RMSE y MAPE; por requerimiento operacional se estableció $R^2 \geq 0.70$ como umbral mínimo de utilidad práctica. El modelo Random Forest demostró una superioridad estadística significativa, alcanzando valores de R^2 cercanos al 90% en la etapa de prueba para las tres secciones, superando ampliamente al modelo OLS, el cual presentó un sobreajuste severo. A pesar del alto desempeño del modelo Random Forest, el análisis global reveló que solo la sección 15 posee una representatividad operativa integral. En esta sección, el modelo logró integrar coherentemente variables de conminución como el *Work Index* y la potencia del molino de bolas, otorgando al sistema una capacidad predictiva de corto plazo. Por el contrario, las otras dos secciones se descartaron, debido a predicciones casi exclusivamente de variables mecánicas y reactivas como la velocidad de bombas y flujos de agua.

Finalmente, se estableció que el modelo desarrollado para la sección 15 cumple una función explicativa y de diagnóstico predictivo, actuando como un complemento estratégico al sistema de control PROFIT.

ABSTRACT

The Chuquicamata Division processes ore through various comminution routes, including conventional grinding (plants “A0” and “A1”) and SAG milling (plant “A2”). Operational control of conventional grinding is supported by the PROFIT expert system; however, the need for complementary diagnostic and predictive tools motivated the development and evaluation of a predictive throughput (tph) model applied to Plant A1 (sections 13, 14, and 15). The main objective was to evaluate the feasibility of a predictive model useful for decision-making and to quantify the influence of key operational signals on processed tonnage.

To characterize the feed, a detailed analysis of mineral blending (feed sources: open pit, underground, and RT), granulometry relative to the 12,700-micron standard (P80), and hourly Work Index (WI) was conducted. A clear relationship was observed between higher open-pit participation and increases in WI; the average WI for the analyzed period was 13.3 kWh/t, with 81% of records concentrated in the 12.6 to 14.0 kWh/t range. The database used for the model consisted of operational variables recorded per minute during 2024 (527,040 records). Preprocessing included an initial IQR filter, manual filtering, and a stable mode. Additionally, time lags were studied to identify leading or lagging variables relative to throughput. After cleaning, the valid rows per section were 49.61%, 62.64%, and 69.27% of the total data for sections 13, 14, and 15, respectively.

A comparative strategy between a linear model (OLS) and a non-linear model (Random Forest) was implemented. The metrics used for evaluation were R^2 , MAE, RMSE, and MAPE; an operational threshold of $R^2 \geq 0.70$ was established as the minimum for practical utility. The Random Forest model demonstrated significant statistical superiority, reaching R^2 values near 90% in the testing phase for all three sections, significantly outperforming the OLS model, which exhibited severe overfitting. Despite the high performance of the Random Forest model, global analysis revealed that only section 15 possesses comprehensive operational representativeness. In this section, the model successfully integrated comminution variables such as the Work Index and ball mill power, providing the system with short-term predictive capacity. Conversely, the other two sections were discarded due to predictions based almost exclusively on mechanical and reactive variables, such as pump speed and water flows.

Finally, it is established that the model developed for section 15 fulfills an explanatory and predictive diagnostic function, acting as a strategic complement to the PROFIT control system.

ÍNDICE

1 INTRODUCCIÓN.....	1
2 OBJETIVOS	2
2.1 Objetivo General.....	2
2.2 Objetivos Específicos.....	2
3 ANTECEDENTES	3
3.1 CODELCO.....	3
3.2 CODELCO CHUQUICAMATA	3
3.2.1 Ubicación y acceso	4
3.2.2 Geología del yacimiento.....	5
3.2.3 Litología.....	5
3.3 Estructuras	7
3.3.1 Falla Oeste	7
3.3.2 Sistema Mesabi.....	8
3.3.3 Sistema de Fallas Estanques Blancos	8
3.3.4 Fallas de orientación NNW.....	8
3.4 Antecedentes generales del proceso.....	9
3.4.1 Distribución mineral	10
3.4.2 Etapa de chancado.....	11
3.4.3 Molienda convencional.....	11
3.4.4 Flotación	14
4 MARCO TEÓRICO	15
4.1 Teoría de conminución.....	15
4.2 Molienda de minerales.....	16
4.2.1 Equipo en molienda convencional de minerales	16
4.3 Configuraciones de Circuitos de Molienda – Clasificación	17
4.4 Modelos de Machine Learning	19
4.4.1 Tipos de Machine Learning.....	20
4.4.2 Algoritmos de <i>Machine Learning</i>	22
4.5 Aplicaciones de Machine Learning en el Procesamiento de mineral	27
4.5.1 Aplicación de ML para la predicción de balance de masa y recuperación en plantas de flotación de fosfato.	27

4.5.2 Aplicación de computación blanda y ML en la optimización de circuitos de conminución..	28
5 METODOLOGÍA	29
5.1 Recopilación de Datos.	29
5.1.1 Variables operacionales	29
5.2 Evaluación y comparación de datos operacionales del 2024.....	31
5.2.1 Análisis del “ <i>blending</i> mineral”	31
5.2.2 Comportamiento de la granulometría de alimentación y del rendimiento en base al Work Index	32
5.2.3 Comportamiento del rendimiento en base a la granulometría	33
5.3 Modelo estadístico predictivo	33
5.3.1 Filtro IQR	34
5.3.2 Filtrado manual basado en el estándar operacional	34
5.3.3 Modo estable	35
5.3.4 Correlaciones operacionales	35
5.3.5 Determinación y aplicación de desfases operacionales (<i>Lags</i>).....	36
5.3.6 Desarrollo y evaluación de modelo predictivo del rendimiento	37
6 RESULTADOS	40
6.1 Evaluación de las variables de alimentación a la molienda convencional.....	40
6.1.1 Resultados de la evaluación de <i>blending</i> mineral y Work Índice.	40
6.1.2 Evaluación de la granulometría y rendimiento en base al Work Índice.....	41
6.1.3 Evaluación rendimiento y <i>Work Index</i> en base a la granulometría.	43
6.2 Implementación del modelo predictivo (filtros y entrenamiento)	44
6.2.1 Aplicación de Filtro IQR.....	44
6.2.2 Filtrado manual de variables fuera de rango operacional	45
6.2.3 Filtrado mediante modo estable	46
6.2.4 Evaluación de correlaciones operacionales	47
6.2.5 Desfases temporales (<i>Lags</i>)	49
6.2.6 Comparación modelos	52
7 CONCLUSIONES Y RECOMENDACIONES	64
7.1 Conclusiones	64
7.2 Recomendaciones.....	65
8 BIBLIOGRAFIA	66
9 ANEXOS	68
9.1 Anexo A. PI ProcessBook	68

9.2 Anexo B. PI DataLink	69
9.3 Anexo C. Tag de variables	69
9.4 Anexo D. Extracción de datos de las variables operacionales	70
9.5 Anexo E. Informe de detenciones “Base de datos molienda”	71
9.6 Anexo F. Extracción de datos del Work Index	71
9.7 Anexo G. Distribución mineral diaria del año 2024	72
9.8 Anexo H. Tabla resumen de <i>Outliers</i> al aplicar filtro IQR.....	73
9.8.1 Variables sección 13.....	73
9.8.2 Variables sección 14.....	74
9.8.3 Variables sección 15.....	75
9.9 Anexo I. Sistema de control PROFIT.....	76
9.10 Anexo J. Tabla resumen de los desfases temporales	77
9.11 Anexo K. Variables operacionales escogidas por Lasso.....	80

ÍNDICE DE FIGURAS

Figura 1. Vista aérea del rajo Chuquicamata y áreas industriales adyacentes. [5].....	4
Figura 2. Unidades Litológicas de la mina Chuquicamata. Se observan tipos litológicos y estructuras con una continuidad superior a tres bancos. [7]	6
Figura 3. Principales fallas de la mina Chuquicamata de Codelco. [6]	9
Figura 4. Diagrama esquemático del procesamiento de minerales por molienda convencional. [9]	10
Figura 5. Circuito de una sección de la planta A1. Elaboración propia.....	12
Figura 6. Tipos de molinos de barras en base a su descarga. (a) Descarga periférica por el extremo final; (b) Descarga por rebalse (b); (c) Descarga periférica central. [14]	17
Figura 7. Tipos de molinos de bolas en base a su descarga. (a) Descarga por rebalse; (b) Descarga por parrillas. [14]	17
Figura 8. Circuito cerrado directo, Molienda – Clasificación.....	18
Figura 9. Circuito cerrado Inverso, Molienda – Clasificación.....	18
Figura 10. Esquema del circuito de molienda con ubicación de sensores de variables operacionales obtenidas desde <i>PI ProcessBook</i>	31
Figura 11. Distribución mineral que ingresa a la tolva gruesa por rangos de Work Index. Elaboración propia	40
Figura 12. Distribución diaria de mineral que ingresa a la tolva gruesa y Work Index. Elaboración propia	41
Figura 13. Comportamiento de la granulometría y rendimiento en base al Work Index. Elaboración propia	42
Figura 14. Comportamiento del rendimiento en base a rangos granulométricos. Elaboración propia	43
Figura 15. Diagrama de correlaciones de Pearson.	48
Figura 16. Serie temporal del rendimiento real vs predicho por el modelo OLS, sección 13.	55
Figura 17. Serie temporal del rendimiento real vs predicho por el modelo OLS, sección 14.	55
Figura 18. Serie temporal del rendimiento real vs predicho por el modelo OLS, sección 15.	55
Figura 19. Desempeño modelo predictivo en la sección 13. a) Serie temporal del rendimiento real y el predicho por el Modelo RF. (b) Importancia relativa de las variables operacionales en el modelo RF.	58
Figura 20. Desempeño modelo predictivo en la sección 14. a) Serie temporal del rendimiento real y el predicho por el Modelo RF. (b) Importancia relativa de las variables operacionales en el modelo RF.	59

Figura 21. Desempeño modelo predictivo en la sección 15. a) Serie temporal del rendimiento real y el predicho por el Modelo RF. (b) Importancia relativa de las variables operacionales en el modelo RF.	60
Figura 22. Diagrama general de la planta concentradora entrega por PI ProcessBook. ...	68
Figura 23. Diagrama de la molienda convencional “A1” entregada por PI ProcessBook. ...	68
Figura 24. Interfaz de la extracción de datos de PI DataLink.	69
Figura 25. Recopilación de datos de las variables operacionales de la sección “A1”.	70
Figura 26. Detenciones mensuales correspondiente a la planta “A1”.	71
Figura 27. Dashboards de la planta concentradora.....	71
Figura 28. Dashboard correspondiente al Work Index que entra a la tolva gruesa.	72
Figura 29. Distribución diaria de mineral en la molienda convencional.	72
Figura 30. Interfaz del controlador experto PROFIT de la planta molienda convencional.	76

ÍNDICE DE TABLAS

Tabla 1. Distribución mezcla de alimentación a la molienda convencional.	10
Tabla 2. Estándares operacionales de la planta “A1”.	14
Tabla 3. Cantidad de Outliers de las secciones de la planta “A1”	44
Tabla 4. Filtrado manual de variables para la sección 13.	45
Tabla 5. Filtrado manual de variables para la sección 14.	45
Tabla 6. Filtrado manual de variables para la sección 15.	46
Tabla 7. Resumen del porcentaje de filas válidas para la aplicación del modelo.	47
Tabla 8. Resumen de variables con desfases positivos (capacidad predictiva).	50
Tabla 9. Resumen del entreno modelo OLS.	52
Tabla 10. Resumen del entreno modelo OLS.	53
Tabla 11. Resumen de la prueba modelo OLS.	53
Tabla 12. Resumen del entreno modelo Random Forest.....	57
Tabla 13. Resumen de la prueba modelo Random Forest.	57
Tabla 14. Ejemplo de Tag utilizados para la extracción de datos.	69
Tabla 15. Límites y Outliers de la sección 13 (MC).	73
Tabla 16. Límites y Outliers de la sección 14 (MC).	74
Tabla 17. Límites y Outliers de la sección 15 (MC).	75
Tabla 18. Desfases temporales en base al rendimiento de la sección 13 de la planta MC.	77
Tabla 19. Desfases temporales en base al rendimiento de la sección 14 de la planta MC.	78
Tabla 20. Desfases temporales en base al rendimiento de la sección 15 de la planta MC.	79
Tabla 21. Variables operacionales seleccionadas mediante Lasso de cada una de las secciones.	80

NOMENCLATURA

tpd: Toneladas por día

tph: Toneladas por hora

kWh: Kilovatio Hora

WI: Work Index

MC: Molienda convencional

BHC: Batería de Hidrociclones

DCH: División Chuquicamata

DRT: División Radomiro Tomic

CH2: Chancado secundario

CH3: Chancado terciario

CT: Carros tripper

P80: Tamaño del 80% acumulado pasante en el producto (μm)

F80: Tamaño del 80% acumulado pasante en la alimentación (μm)

K80: Tamaño de transferencia descarga

RM: Rod mill (Molino barra)

BM: Ball mill (Molino bolas)

CV: Coeficiente de variación

ML: Machine Learning

1 INTRODUCCIÓN

La molienda convencional constituye una de las etapas más sensibles y determinantes dentro del procesamiento de minerales en la División Chuquicamata, debido a que su desempeño impacta directamente tanto en el tonelaje tratado como en la granulometría de alimentación hacia flotación. A pesar de que cada sección del circuito está diseñada bajo el mismo principio operativo, en la práctica su comportamiento dinámico no es homogéneo, ya que, las variaciones en condiciones minerales, distribución granulométrica, carga circulante, disponibilidad de agua y respuestas hidráulicas generan diferencias significativas entre secciones, dificultando la predicción y el control del rendimiento en tiempo real.

Para enfrentar esta complejidad, la planta opera bajo un sistema de control automático basado en inteligencia operacional (PROFIT), el cual ajusta variables manipuladas como el tonelaje, el flujo de agua hacia el cajón y la velocidad de la bomba con el fin de estabilizar el proceso y mantenerlo dentro de los rangos operacionales definidos. No obstante, PROFIT trabaja principalmente sobre la base de reglas y modelos empíricos, por lo que su desempeño depende de la correcta interpretación de cómo las variables del proceso influyen entre sí.

En este contexto, y considerando que en operación real no existe una claridad completa del efecto cuantitativo que tiene cada parámetro operacional sobre el rendimiento, se vuelve necesario complementar el sistema de control existente mediante herramientas de análisis estadístico avanzadas. El uso de metodologías multivariantes y técnicas de aprendizaje automático ha cobrado relevancia en minería, ya que permiten analizar grandes volúmenes de datos, identificar patrones no triviales y describir la contribución relativa de cada variable en el comportamiento del proceso.

Bajo esta perspectiva, el presente trabajo busca cuantificar el efecto real que ejercen las principales variables operacionales sobre el rendimiento del circuito de molienda convencional, utilizando para ello una base de datos minuto a minuto y metodologías estadísticas y de *machine learning* aplicadas de manera rigurosa. Además de caracterizar estas relaciones, el estudio pretende desarrollar un modelo capaz de predecir el rendimiento (tph) bajo distintas condiciones operativas, entregando una herramienta sencilla, interpretativa y coherente con la realidad del proceso, aportando una base cuantitativa robusta para mejorar la estabilidad operacional y orientar decisiones técnicas con mayor fundamento.

2 OBJETIVOS

2.1 Objetivo General

Evaluar la factibilidad y el desempeño potencial de un modelo predictivo del rendimiento de la Planta de molienda convencional A1 de la División Chuquicamata de CODELCO, mediante análisis de datos operacionales históricos y experimentación con técnicas de modelamiento.

2.2 Objetivos Específicos

- Analizar y evaluar la relación del Wok Index, la granulometría y la distribución mineral de la alimentación con el rendimiento de la Planta A1.
- Explorar distintos enfoques de modelamiento (p. ej., regresión multivariantes/Lasso-Ridge, Random Forest) y cuantificar su desempeño con MAE, RMSE y R^2 .
- Determinar las principales brechas que limitan el modelado y proponer recomendaciones operacionales, independiente del éxito predictivo alcanzando.

3 ANTECEDENTES

3.1 CODELCO

La corporación nacional del Cobre, más conocida como Codelco, es una empresa autónoma, de propiedad del estado chileno, la cual fue formalizada oficialmente por decreto el 1 de abril de 1976 como se le conoce en la actualidad, siendo un pilar económico para el país. Dedicada a la exploración, desarrollo y explotación de recursos cupríferos, así como al procesamiento y comercialización de cobre refinado y subproductos, tales como molibdeno, ácido sulfúrico, oro, plata y renio [1].

Codelco actualmente opera 7 divisiones, en las cuales se encuentra, Chuquicamata, Radomiro Tomic, Ministro Hales, Gabriela Mistral, Salvador, Andina y El teniente, además de la refinería Ventanas, con cada matriz en Santiago, contando con una dotación personal de aproximadamente 72000 personas, siendo 16000 trabajadores propios [2]. Codelco posee el 6% de las reservas globales de cobre, que se encuentran en yacimientos de clase mundial. Además, ocupa el segundo lugar en la producción de molibdeno con una participación del 10%. Esto subraya la importancia de la empresa en el mercado de los metales.

En el año 2024, la producción total alcanzo 1.441.886 toneladas de cobre, considerando estos datos con la participación de El Abra, Anglo American Sur y Quebrada Blanca, aportando USD 1534 millones al Fisco, un 8% más que en el año 2023 [3].

3.2 CODELCO CHUQUICAMATA

La División Chuquicamata de CODELCO, una de las unidades más antiguas de la empresa estatal, se encuentra en el norte de Chile a 15 km de la ciudad de Calama, segunda región de Antofagasta. Iniciando operaciones en 1915, esta división tiene una larga historia, ya que sus yacimientos eran conocidos desde épocas prehispánicas. Su operación comprende la mina de rajo abierto y una mina subterránea. La producción del año 2024 fue de 289 mil toneladas métricas de cobre fino, con una dotación propia a finales de ese año (31 de diciembre del 2024) de 3939 [4]. En la **Figura 1** se muestra una vista aérea del rajo Chuquicamata.



Figura 1. Vista aérea del rajo Chuquicamata y áreas industriales adyacentes. [5]

3.2.1 Ubicación y acceso

El complejo minero de Chuquicamata está ubicado a 1.650 km al Norte de Santiago, en la Región de Antofagasta, Provincia El Loa, a 240 km al noroeste de la ciudad de Antofagasta ($68^{\circ}54'$ longitud Oeste y $22^{\circ}17'$ latitud Sur), a 2.870 m sobre el nivel del mar. El acceso terrestre principal corresponde a la Carretera Panamericana (ruta 5) y la ruta 25, que une la localidad de Carmen Alto con la ciudad de Calama. Desde ésta se llega al yacimiento por un camino pavimentado de doble vía [6].

3.2.2 Geología del yacimiento

El yacimiento Chuquicamata destaca por ser un megadepósito, relacionado a un magnetismo intrusivo de edad eocena-oligocena, cuyo emplazamiento ocurrió dentro del sistema de Falla de Domeyko. Es uno de los mayores pórfidos cupríferos a nivel mundial y es el más grande de la región y del país. El entorno geológico del yacimiento Chuquicamata se caracteriza por la presencia dominante de rocas intrusivas y fallas regionales, además de rocas volcánicas y sedimentarias. El sistema de falla Domeyko (ZFD), incluyendo la Falla Oeste de Chuquicamata, divide el distrito en dos dominios: uno occidental donde afloran unidades del Mesozoico al Terciario principalmente; y uno oriental en el que predominan rocas del Paleozoico [6].

3.2.3 Litología

3.2.3.1 Complejo intrusivo Chuquicamata

El Complejo Intrusivo Chuquicamata, o “Pórfido Chuqui”, se formó entre el Eoceno Superior y el Oligoceno Inferior (hace aproximadamente 41–31 Ma) como un dique subvertical de cuerpos alargados con rumbo NE. Esta unidad constituye la roca anfitriona de la mineralización de cobre y molibdeno en Chuquicamata, estando delimitada por fallas como la Falla Mesabi y la Falla Oeste y controlada por el sistema de la Falla de Domeyko. En su seno se diferencian tres variedades principales, Pórfido Este, Pórfido Oeste y Pórfido Banco relacionadas con la culminación de la Fase Incaica de compresión regional [7]. En la **Figura 2** se presentan las unidades litológicas de la mina Chuquicamata.

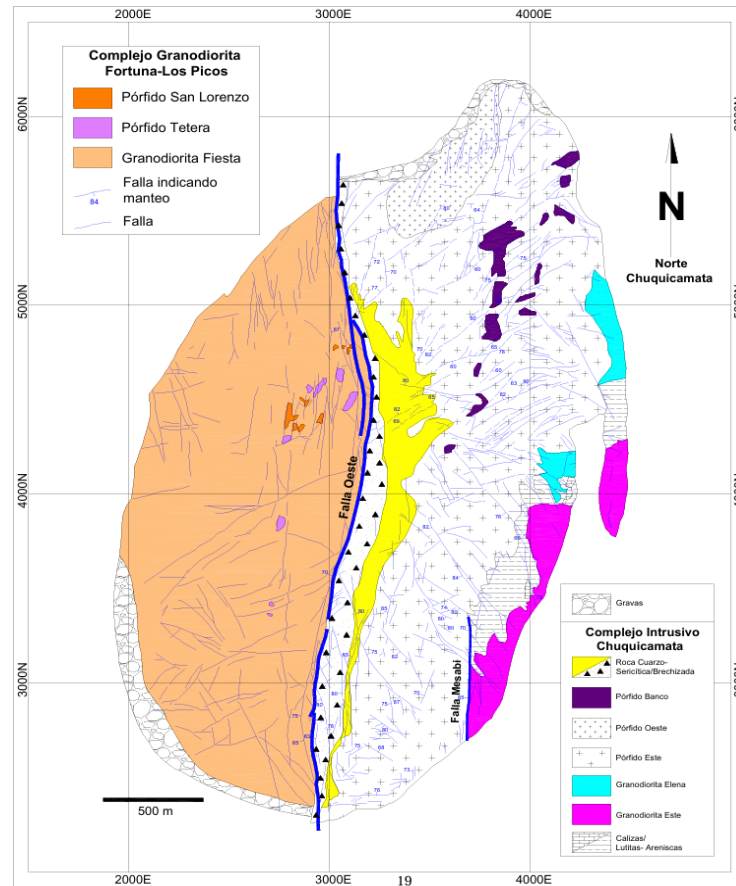


Figura 2. Unidades Litológicas de la mina Chuquicamata. Se observan tipos litológicos y estructuras con una continuidad superior a tres bancos. [7]

3.2.3.1.1 Pórfido Este

Unidad de mayor distribución en el yacimiento. Es un cuerpo elongado de orientación NNE y esta intruído por el Pórfido Oeste y el Pórfido Banco. Su composición varía de granodiorita a monzodiorita cuarcífera de biotita, con textura fanerítica inequigranular a incipientemente porfídica, de grano medio. Se compone de fenocristales de feldespato potásico, plagioclasas subhedrales a euhedrales y fenocristales de biotita con escasa hornblenda. Los minerales accesorios presentes son; magnetita, esfeno y zircón. Las dataciones a través del método U – Pb en zircón, arrojaron la edad de cristalización del Pórfido Este en 34.6 Ma [6].

3.2.3.1.2 Pórfido Oeste

Unidad litológica diferenciada del Pórfido Chuqui que se encuentra en el extremo norte del depósito. Su litología corresponde a un pórfido monzogranítico con textura porfídica fina que corresponden a plagioclasas, feldespato potásico, ojos de cuarzo y biotita en una matriz de microgranito. Se disponen como cuerpos de formas irregulares con orientación NNE y los contactos con el pórfido Este son gradacionales y difusos. Las dataciones a través del método U-Pb en Zircón arrojaron una edad de 33.5 Ma [6].

3.2.3.1.3 Pórfido Banco

Pórfido monzodiorítico que se dispone como filón que se emplaza principalmente en el área norte del yacimiento y corresponde a la unidad de menor tamaño del Pórfido Chuqui. Su composición es muy similar al Pórfido Este, pero se distingue por presentar plagioclasas en su masa fundamental, siendo sus contactos nítidos y abruptos. Las dataciones con el método U-Pb en zircón arrojaron una edad de 33.3 Ma [6].

3.3 Estructuras

Dentro del yacimiento de Chuquicamata, los principales sistemas estructurales y fallas que ayudaron al emplazamiento del pórfido, que controlaron la alteración y distribución de la mineralización son los siguientes:

3.3.1 Falla Oeste

De orientación NS a NNE y manteo de 75° a 80° al Oeste, que decrece levemente en profundidad, es la estructura mejor reconocida del depósito y la más sistemática y consistentemente interpretada en el modelo geológico. Al interior del depósito aparece principalmente como una falla post-mineral, con desplazamiento sinistral-inverso y que limita abruptamente el yacimiento por el Oeste, truncando al Pórfido Chuqui y a la mineralización temprana y poniéndolos en contacto con la Granodiorita Fortuna estéril. La evidencia en sondajes profundos de simetrías mineralógicas, de ley y de alteración, entre los bordes Este y Oeste del depósito, indican que esta falla post-mineral no se emplaza en el eje del depósito, sino que está desplazada hacia el Oeste respecto al centro de la mineralización hipógena y, por lo tanto, la franja de mineral disectada y “perdida” por el desplazamiento de la falla sería muy menor. Por otra parte, la Falla Oeste presenta una divergencia angular con el eje principal de alteración y mineralización del depósito, que es aproximadamente Nor-Este, lo que indicaría que solo controla eventos muy tardíos de mineralización

(sericita/pirita/enargita) que siguen aproximadamente su posición y, en cambio, disecta los eventos principales, más tempranos [6].

3.3.2 Sistema Mesabi

Sistema de fallas de orientación NNE en el que se incluyen la Zona de Deformación Este asociado a la falla Mesabi, la falla americana, la falla C-2 y la falla Zaragoza. Estas estructuras son jóvenes con respecto a la evolución del depósito y probablemente controla el emplazamiento del Pórfido Chuqui y las etapas tempranas de la mineralización que son relevantes en cuanto al aporte de cobre y molibdeno en el depósito [6].

3.3.3 Sistema de Fallas Estanques Blancos

Este conjunto de fracturas de rumbo NE y buzamiento casi vertical atraviesa todo el depósito, pero en el sector Centro-Norte adquiere especial relevancia la Falla Estanques Blancos, reconocida por su destacada penetración y un claro desplazamiento dextral que incluso alcanza el borde del tajo de Chuquicamata. En esta zona, las vetas y halos de alteración tanto hipogénica como supergénica se orientan siguiendo el trazo de la falla, lo que demuestra que ha sido un canal preferencial para los fluidos hidrotermales a lo largo de toda la evolución del depósito.

Tras la fase principal de mineralización, diversos trabajos documentan una reactivación con sentido de movimiento derecho que generó desplazamientos de entre 200 y 300 m, de acuerdo con los mapas de Renzetti (1955). Además, el sistema Estanques Blancos–Portezuelo exhibe un componente significativo de falla normal, donde el bloque sur se hunde respecto al norte, formando bloques escalonados en dirección sur que explican la variación en la profundidad de la mineralización desde Chuquicamata hasta la cuenca del río Loa [8].

3.3.4 Fallas de orientación NNW

Corresponden a un set de fallas de actitud NNW a NW con inclinación sub-verticales presentes en todo el depósito, pero es particularmente frecuente en el sector central y Sur. Según estudios estructurales, este set de fallas ha sido considerado un sistema post- mineral y tardío en la evolución del depósito (véase Figura 3). También se ha evidenciado que estas fallas NNW cortan a la Falla Oeste, desplazándola algunos metros en sentido sinistral. Sin embargo, al interior del yacimiento existen también vetas y vetillas mineralizadas preferentemente tardías (pirita-enargita), que están controladas por estas estructuras [6].

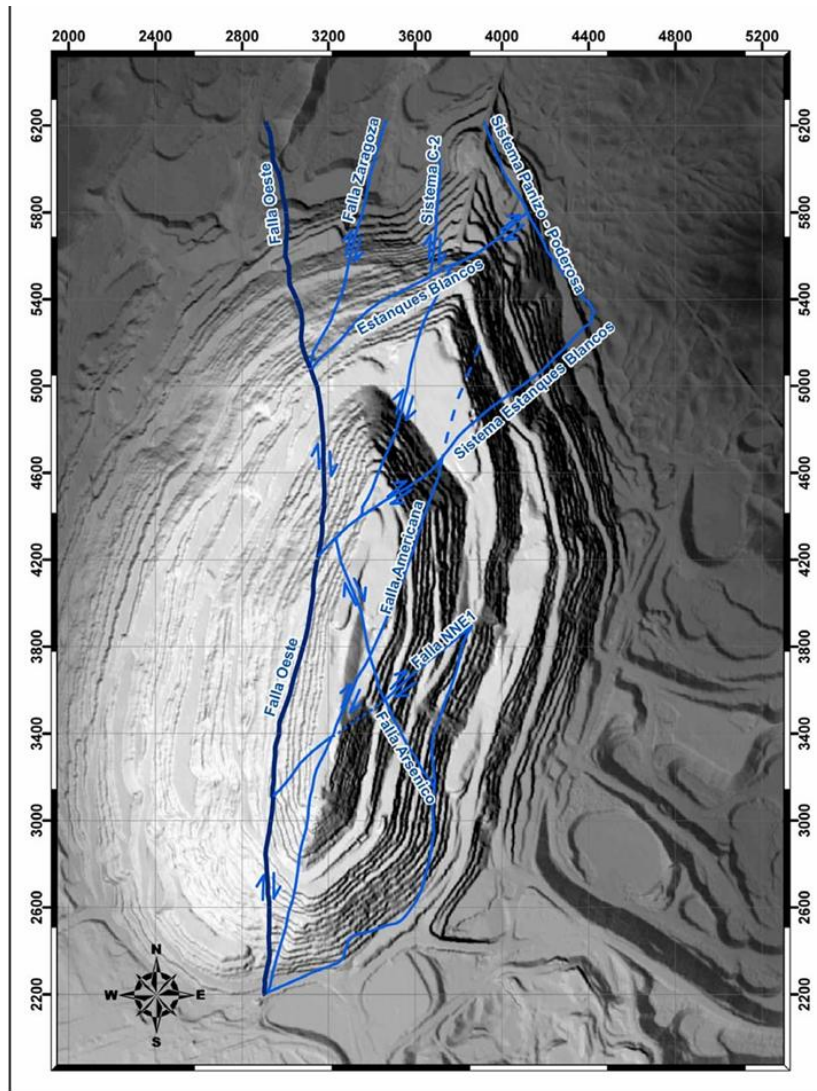


Figura 3. Principales fallas de la mina Chuquicamata de Codelco. [6]

3.4 Antecedentes generales del proceso

La División Chuquicamata cuenta con dos rutas de procesamiento mineral dentro de su línea de conminución: **molienda SAG** y **molienda convencional**. Ambas vías cumplen el mismo propósito, la cual es, reducir el tamaño del mineral para permitir la liberación de especies valiosas, pero operan con configuraciones y dinámicas distintas.

La molienda SAG procesa mineral directamente desde chancado primario, mientras que la molienda convencional requiere etapas sucesivas de chancado previo y clasificación para alcanzar la granulometría necesaria.

Dado que el presente estudio se enfoca en el **circuito de molienda convencional**, en la **Figura 4** se muestra el diagrama de flujo general de procesamiento de minerales de la División Chuquicamata por esta vía.

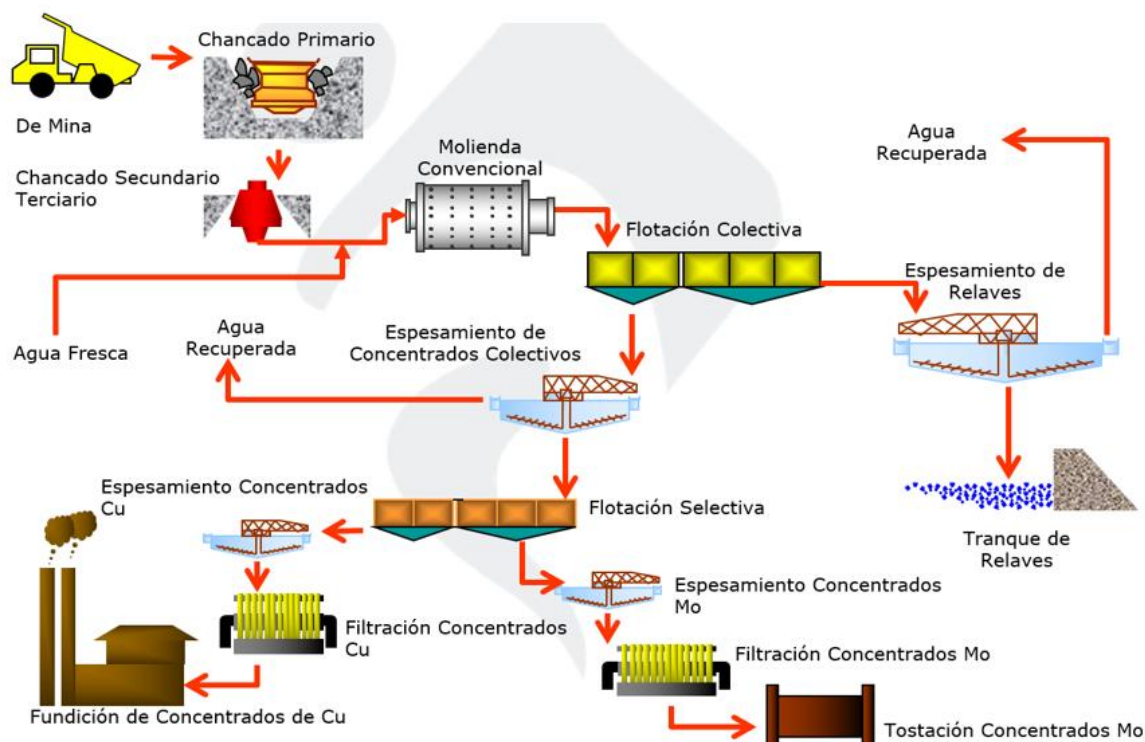


Figura 4. Diagrama esquemático del procesamiento de minerales por molienda convencional. [9]

3.4.1 Distribución mineral

La concentradora de la división Chuquicamata implementa una estrategia de usar “*blending*” de minerales, es decir, mezclas de distintas fuentes, la cual incluye minerales provenientes del rajo, mina subterránea y minerales sulfurados provenientes de RT (División Radomiro Tomic) [10], esto nace con el fin de la necesidad de satisfacer la capacidad de las plantas de molienda SAG y convencional, la cual tiene una capacidad teórica de 210000 toneladas. El porcentaje de alimentación y leyes promedio de cobre correspondiente a cada fuente se evidencia en la **Tabla 1**.

Tabla 1. Distribución mezcla de alimentación a la molienda convencional.

Fuente	% Alimentación	Ley
DRT	48.4%	0.6%
DCH (Rajo)	36.9%	0.4 %
DCH (subterránea)	14.7%	0.9 %

Las especies mineralógicas de sulfuros que se encuentran en la mezcla mineral que entra a la tolva gruesa, contribuyen de manera única a las propiedades y características de esta, afectando principalmente al “Work Index”, la cual este índice de dureza permite estimar en forma simple las variaciones que se pueden esperar al variar las características del mineral de acuerdo con el frente de explotación de la mina.

3.4.2 Etapa de chancado

El mineral extraído de la mina presenta una granulometría muy amplia, con partículas que van desde menos de 1 mm hasta fragmentos de más de 1 m de diámetro. Para homogeneizar el tamaño y garantizar un tamaño de partícula máximo de ½ pulgada (1,27 cm o 12700 μ m), el chancado se realiza en serie mediante tres o cuatro equipos dispuestos en línea [10].

En la etapa primaria, el chancador reduce el diámetro máximo de los trozos hasta 8 pulgadas (20.32 cm), luego todo el producto de ese proceso es depositado en una tolva gruesa, la correa 134 se divide en 5 correas (A, B, C, D, E), la cual tienen por objetivo transportar el material de la tolva gruesa al chancado secundario y terciario (CH2 y CH3). A continuación, en la etapa secundaria, el material se reduce hasta aproximadamente 3 pulgadas (7,62 cm). Finalmente, posterior a la etapa terciaria el producto es transportado mediante carros tripper (CT-137 Y CT-137A), suministrando a cada uno de los silos correspondientes a cada sección de la planta y se logra un tamaño final de ½ pulgada, De este modo, cada fase va disminuyendo gradualmente el tamaño de las partículas, asegurando la uniformidad requerida para los procesos posteriores.

3.4.3 Molienda convencional

La planta MC cuenta con dos líneas de procesamiento, la planta “A0” y la planta “A1”. La primera (A0) está conformada por 13 secciones numeradas desde la 0 a la 12; mientras que la planta A1, que es el foco de este estudio, opera con tres secciones correspondientes a la 13, 14 y 15.

Ambas plantas funcionan bajo un circuito cerrado e inverso de molienda convencional. En esta configuración, el mineral ingresa primero a los molinos de barras, los cuales en la planta A1 presentan dimensiones de 18 pies de largo por 13.5 pies de diámetro, con una potencia nominal de 1750 HP. El producto descargado se dirige posteriormente al cajón de descarga, desde donde es impulsado por bombas hacia la batería de hidrociclones, cuyos equipos poseen un vortex de 13.4 in y un apex de 6.7 in [9].

El material fino (overflow) proveniente de los hidrociclones alimenta la etapa de flotación primaria, mientras que el material grueso (underflow) retorna al molino de bolas. Estos molinos de bolas presentan dimensiones de 16.6 pies de largo por 21 pies de diámetro, con una potencia instalada

de 3500 HP [9], completando así el circuito de reducción de tamaño. En la **Figura 5** se muestra un diagrama representativo del circuito de molienda convencional, aplicable a las tres secciones operativas de la planta A1.

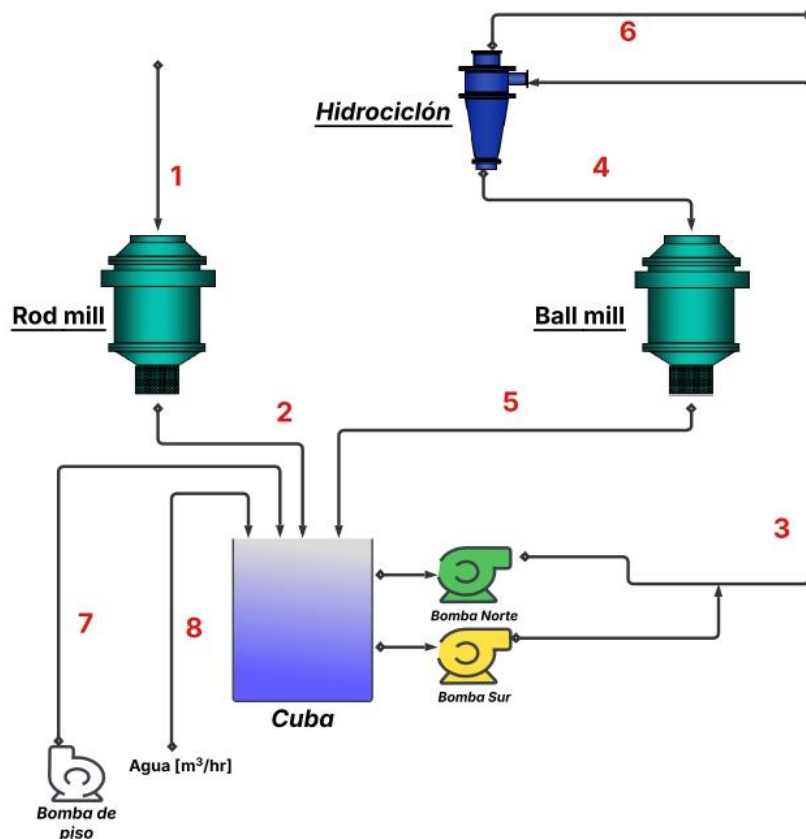


Figura 5. Circuito de una sección de la planta A1. Elaboración propia

Donde la numeración corresponde a los principales flujos del circuito:

1. **Tonelaje de alimentación (rendimiento):** Alimentación seca que ingresa al molino de barras (Rod mill).
2. **Descarga de molino de barras:** pulpa proveniente del RM hacia cajón de descarga.
3. **Flujo de pulpa hacia batería de hidrociclones:** flujo del material desde el cajón de descarga.
4. **Spigot (Underflow):** flujo de material grueso que retorna al Ball mill para su remolienda.
5. **Descarga del molino de bolas:** producto del BM que retorna al cajón de descarga.
6. **Overflow del hidrociclón:** flujo fino clasificado que se dirige hacia la planta de flotación.
7. **Flujo de bombas de piso:** recirculación de pulpa desde el piso, la activación de estas bombas es a causa de algún rebalse producido en el cajón de descarga.
8. **Flujo de agua de proceso:** aporte de agua de proceso hacia el cajón de descarga.

3.4.3.1 Estado actual del proceso

Actualmente las secciones de la molienda convencional son controladas mediante un sistema de inteligencia denominado **PROFIT**, desarrollado principalmente por la Empresa KAIROS Mining en conjunto con la UCAP (Unidad de control automático de proceso), el cual mediante el control de parámetros operacionales busca como objetivo:

- Estabilizar la operación, disminuyendo la variabilidad del proceso.
- Mantener un rango optimo de la granulometría hacia flotación.
- Aumentar el rendimiento de la planta.
- Optimizar el uso de energía en los equipos de molienda.

El sistema se organiza en torno a dos grupos de variables; las variables manipuladas (aquellas que el controlador puede ajustar directamente) y las variables de controladas.

a) Variables manipuladas.

- Tonelaje de alimentación (tph).
- Flujo de agua hacia el cajón (m^3/h).
- Velocidad de la bomba (%).

b) Variables controladas.

- Presión de la batería de hidrociclones (Kg/cm^2).
- Nivel de cajón (%).
- Malla +65 (%).
- Potencia del molino de bolas (kW).
- Densidad hacia la batería de hidrociclones (gr/cm^3).

Este conjunto de variables constituye la base del control automático que actualmente regula la operación. Si bien el control automático mediante **PROFIT** guía gran parte de la operación, la planta también se rige por **estándares operacionales definidos** que establecen los rangos normales de funcionamiento para las variables clave del proceso. Estos se resumen en la **Tabla 2**.

Tabla 2. Estándares operacionales de la planta “A1”.

Parámetros	Normal Low Limit	Normal High Limit
Presión BHC (PSI)	7	12
Potencia RM (kW)	1100	1300
Potencia BM (kW)	2000	2600
Nivel Pozo (%)	45	65
% Sólido RM	76	79
% Sólido descarga	38	40
% Sólido BHC	55	60
Malla + #65		30%
Rendimiento (tph)	≥482	≤650
Nivel de llenado Barras RM (%)		38
Nivel de llenado bolas BM (%)		32

3.4.4 Flotación

La División Chuquicamata opera tres plantas de flotación vinculadas al tratamiento de mineral: A0, A1 y A2. Las plantas A0 y A1 reciben la pulpa proveniente de sus respectivos circuitos de molienda convencional, mientras que A2 procesa la pulpa generada en el circuito SAG. En estas instalaciones se desarrolla la flotación de minerales sulfurados, cuyo propósito es recuperar las especies valiosas presentes en la pulpa mediante su separación selectiva.

El proceso se desarrolla en celdas mecánicas dispuestas en etapas rougher, cleaner y scavenger, donde las partículas hidrofóbicas se adhieren a burbujas de aire y son recuperadas como concentrado. Su desempeño depende de variables como porcentaje de sólidos, aireación, granulometría, pH y dosificación de reactivos (colectores, espumantes y modificadores).

4 MARCO TEÓRICO

4.1 Teoría de conminución

La conminución corresponde al proceso de reducción de tamaño del mineral, etapa fundamental del procesamiento, ya que, determina la liberación de especies valiosas y condiciona el rendimiento de la molienda y flotación. Diversos autores han propuestos modelos energéticos para describir la relación entre el tamaño final del mineral y la energía requerida para su reducción, siendo las tres leyes clásicas, las formuladas por Rittinger, Kick y Bond [11].

Mientras que el postulado de Rittinger (1867) se asocia habitualmente a la fractura de partículas gruesas en etapas de chancado, y el de Kick (1885) se aplica mejor en la molienda de partículas finas, es la Tercera ley de la conminución o Postulado de Bond (1952) la que mejor se ajusta a los requerimientos técnicos de la molienda convencional, la cual se evidencia en la ecuación 1.

- **Postulado de BOND (1952) (Tercera Ley de la Conminución)**

La energía consumida para reducir el tamaño 80% de un material, es inversamente proporcional a la raíz cuadrada del tamaño 80%; siendo este último igual a la abertura del tamiz (en micrones) que deja pasar el 80% en peso de las partículas.

Bond definió el parámetro KB en función del Work Index (índice de trabajo del material), que corresponde al trabajo total (expresado en [kWh/ton. corta]) necesario para reducir una tonelada corta de material desde un tamaño teóricamente infinito hasta partículas que en un 80% sean inferiores a 100 [μm] [11].

$$E_B = 10W_i * \left[\left(\frac{1}{\sqrt{P_{80}}} \right) - \left(\frac{1}{\sqrt{F_{80}}} \right) \right] \quad (1)$$

Donde:

E_B : Energía específica de conminución [kWh/ton]

W_i : Índice de trabajo

El parámetro W_i depende tanto del material (resistencia a la conminución) como del equipo de conminución utilizado, debiendo ser determinado experimentalmente para cada aplicación requerida. También representa la dureza del material y la eficiencia mecánica del equipo.

4.2 Molienda de minerales

La molienda, siendo la fase final de la fragmentación industrial, sucede después de la trituración y se caracteriza por obtener productos con tamaños que van desde algunos milímetros hasta algunas micras. Su objetivo principal es lograr la liberación de especies mineralógicas, permitiendo concentración diferenciada o alcanzando el tamaño de grano requerido para el siguiente método de utilización del producto. Esta operación se lleva a cabo mediante molinos, los cuales pueden trabajar en condiciones secas o húmedas [9].

La distinción entre molienda en seco y en húmedo se basa comúnmente en un contenido de humedad alrededor del 8%, marcando un límite en el comportamiento de los equipos de molienda. Por debajo de este valor se considera trabajo en seco, siendo deseable que la humedad superficial no supere el 3%-4%. El trabajo en húmedo implica rangos de humedad más elevados, normalmente entre el 40% y el 70%, lo que se traduce en valores de relación de sólidos en peso (Sp) entre 55% y 80% para la mayoría de los minerales.

En condiciones de humedad entre el 5% y el 10%, puede generarse una pasta pegajosa que reduce significativamente el rendimiento del molino. Para humedades mayores, típicas del trabajo en húmedo, se recupera un desempeño adecuado mientras la dilución no sea excesiva; por ejemplo, valores de relación de sólidos por debajo del 50 % tienden a disminuir la eficiencia, ya que las bolas pierden capacidad de impacto efectivo sobre los granos [12].

4.2.1 Equipo en molienda convencional de minerales

- **Molino de barras:** Los molinos de barras generalmente se utilizan en aplicaciones industriales de molienda húmeda. Para los rangos de aplicación de molienda más fina en estos molinos (P_{80} variando entre 2 y 0.5 mm), se acostumbra a emplear los “*overflow rod mills*” (**Figura 6 (b)**); ósea los que descargan por rebalse. Para el rango de molienda más gruesa en molino de barra ($P_{80} > 2$ mm), se emplean normalmente los “*center peripheral discharge roll mills*” (**Figura 6 (c)**); que descargan por el centro de la carcasa. Esta última aplicación se efectúa cuando el producto de molienda requerido no puede contener más de que una cantidad ínfima de material extremadamente fino (tal como sucede en las especificaciones de arenas). La molienda seca en molinos de barras no es aconsejable, ya que material seco fluye con dificultad, lo que origina trabamiento de las barras, su posterior ruptura. De cualquier forma, si la necesidad del caso lo requiere (planta de molienda de cemento Clinker, por ejemplo), se acostumbra a diseñar los molinos de barras, para la molienda seca, con descarga periférica por el extremo final (“*End Peripheral Discharge Rod Mills*”), según se ilustra en la **Figura 6 (a)** [13].

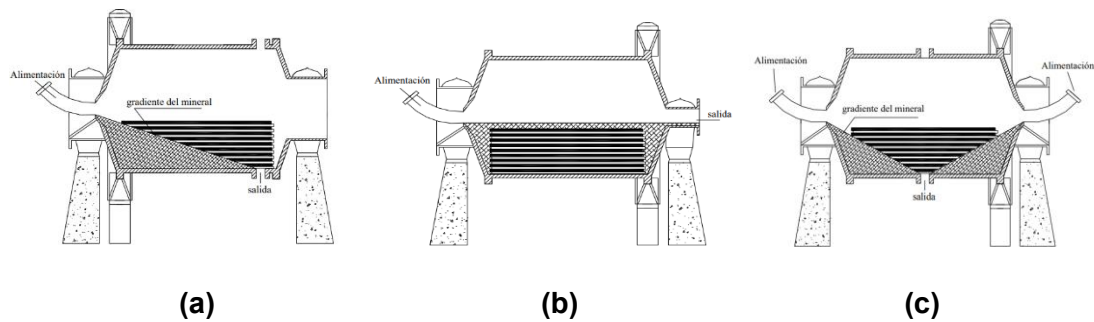


Figura 6. Tipos de molinos de barras en base a su descarga. (a) Descarga periférica por el extremo final; (b) Descarga por rebalse (b); (c) Descarga periférica central. [14]

- **Molinos de bolas:** Los molinos de bolas se utilizan generalmente, en todas aquellas aplicaciones industriales en que se requiera obtener ya sea un producto con granulometría intermedia (P_{80} variado entre 0,5 mm y 75 micrones), o un producto más fino ($P_{80} < 75 \mu m$). Dependiendo de las características propias de material a moler y de granulometría final requerida, existen diversos diseños de molinos y mecanismos de descarga en los mismos, dos de los cuales se esquematizan en la **Figura 7**.

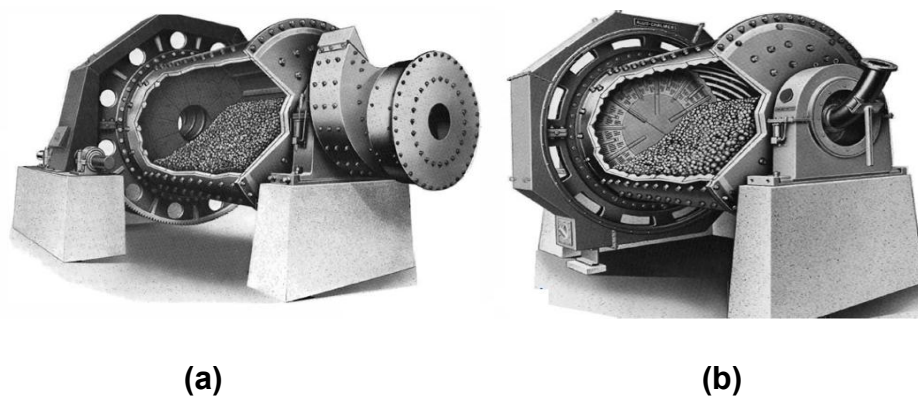


Figura 7. Tipos de molinos de bolas en base a su descarga. (a) Descarga por rebalse; (b) Descarga por parrillas. [14]

4.3 Configuraciones de Circuitos de Molienda – Clasificación

En general los circuitos de molienda y clasificación pueden ser configurados de 2 formas distintas principalmente, los cuales son circuito directo (**Figura 8**) e inversos (**Figura 9**).

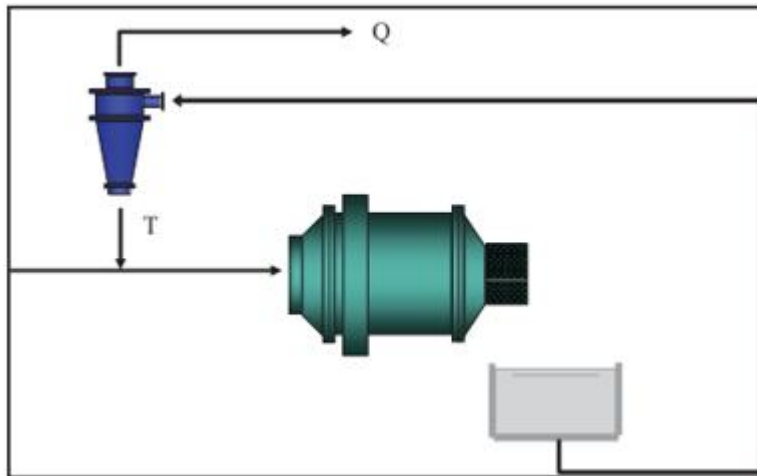


Figura 8. Circuito cerrado directo, Molienda – Clasificación.

En el caso de los circuitos directos, la descarga del hidrociclón (Underflow) más la alimentación fresca ingresan directamente al molino para la posterior clasificación.

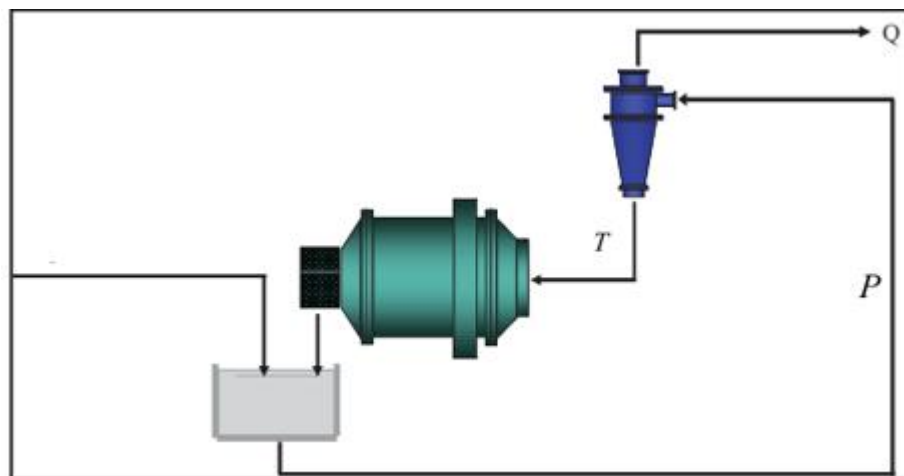


Figura 9. Circuito cerrado Inverso, Molienda – Clasificación.

Donde P , Q y T corresponden a los flujos másicos de alimentación, rebalse y descarga al hidrociclón [15].

En los circuitos inversos, la alimentación fresca se mezcla con la descarga del molino para alimentar el clasificador, aunque solo el material que sale del hidrociclón retorna al molino. Sea cual sea la configuración, es fundamental evitar la sobremolienda de partículas finas, ya que un exceso de finos no solo degrada la eficiencia del proceso, sino que también eleva el consumo energético al reprocesar material que ya cumple con la granulometría requerida.

Para mitigar esto, se incorpora un clasificador que, por un lado, asegura que solo pase al siguiente proceso el tamaño de partícula deseado y, por otro, genera la carga circulante del circuito. Dicha

carga recirculante actúa como una “escoba” dentro del molino: al ingresar arrastra los finos hacia el clasificador, favoreciendo una población interna de partículas más gruesas. Existen 2 conceptos importantes a definir, y se representan con las ecuaciones 2 y 3.

- Carga Circulante:

$$CC = \frac{Q+T}{Q} \quad (2)$$

- Razón de recirculación:

$$C = \frac{T}{Q} \quad (3)$$

El conocimiento de la razón de recirculación C es importante en la descripción de la operación de un circuito cerrado de molienda-clasificación. Ella no se calcula generalmente a partir de los flujos máxicos, de acuerdo con su definición, sino que se utiliza para ello los análisis granulométricos p_i, q_i y t_i , estos términos corresponden a las fracciones en masa de partículas en el intervalo de tamaño i , respectivamente [15]. De la ecuación anterior se puede deducir que la razón de recirculación se puede definir como se evidencia en la ecuación 4.

$$C = \frac{T}{Q} = \frac{p_i - q_i}{t_i - p_i}, \quad 1 \leq i \leq n \quad (4)$$

En termino de C , la carga circulante queda descrita por $(C + 1)$.

4.4 Modelos de Machine Learning

El *machine learning* ha adquirido una relevancia creciente en el ámbito tecnológico en los últimos años. Como subcampo de la Inteligencia Artificial, se ha convertido en un componente central de numerosos sistemas y servicios que empleamos cotidianamente [16].

En términos sencillos, el aprendizaje automático es la disciplina que permite a los computadores “aprender” a partir de datos y experiencias, de manera análoga a cómo lo hacen las personas. Mediante algoritmos y modelos matemáticos, los sistemas procesan grandes volúmenes de información para identificar patrones, extraer estructura y generar inferencias [16].

Así, ¿qué es exactamente *machine learning*? También llamado aprendizaje automático en español es la rama de la Inteligencia Artificial que utiliza algoritmos para reconocer regularidades en conjuntos de datos masivos y construir modelos capaces de realizar predicciones (análisis predictivo). Gracias a este aprendizaje, las máquinas pueden ejecutar tareas específicas de forma

autónoma, es decir, sin que cada acción haya sido codificada explícitamente por un programador, adaptándose además cuando reciben nuevos datos.

En la actualidad los modelos de *machine learning* se han convertido en un recurso tecnológico implementado en herramientas de uso diario como filtros antispam para correos electrónicos, conducción automática de coches o softwares de reconocimiento de voz. El ML presenta una forma muy interesante de generación de aprendizaje tomando como base la información que se extrae de los datos analizados por los computadores emulando el comportamiento humano en cierta medida. Tomando los datos y su comportamiento estos modelos de aprendizaje podemos crear modelos predictivos de sucesos o módulos gerenciales pensados para la toma de decisiones con un nivel de eficiencia considerablemente alto [17].

4.4.1 Tipos de Machine Learning

Dentro de este campo de desarrollo tecnológico en plena expansión existen diferentes tipos de *machine learning*. El desarrollo de aprendizaje automático se centra en tres categorías principales que se conocen como aprendizaje supervisado, aprendizaje no supervisado y aprendizaje profundo [17].

4.4.1.1 Aprendizaje supervisado

El aprendizaje supervisado es una modalidad de *machine learning* en la que un sistema adquiere conocimiento a partir de ejemplos etiquetados: conjuntos de datos donde cada entrada va acompañada de su resultado conocido. A partir de esos pares entrada-salida, el modelo ajusta sus parámetros internos para capturar la relación subyacente y así poder generalizar y entregar predicciones sobre casos nuevos que no ha visto antes [17].

Para que este aprendizaje sea fiable es clave que los datos etiquetados sean representativos y de buena calidad: el volumen de datos, la cobertura de condiciones operativas, la precisión del etiquetado y el equilibrio entre clases influyen directamente en la capacidad del modelo para predecir correctamente. En la práctica se suele separar el conjunto en bloques (entrenamiento / validación / prueba) y se evalúa el desempeño con métricas adecuadas (por ejemplo, precisión, error medio) para medir la capacidad de generalización y detectar fallos en la estimación.

Las aplicaciones son muy variadas: desde filtros antispam y reconocimiento de imágenes o de voz, hasta usos industriales como la detección temprana de fallas, el monitoreo de condiciones de equipos, la predicción de caudales o rendimientos y la optimización de parámetros de planta. En todos los casos, el aprendizaje supervisado facilita la automatización de decisiones y la asistencia en la toma de decisiones operativas al transformar datos históricos en predicciones accionables.

No obstante, tiene límites: el rendimiento depende totalmente de la calidad y representatividad de las etiquetas y puede degradarse si cambian las condiciones del proceso (por eso requiere actualización y monitoreo). Además, si los datos contienen sesgos o errores, el modelo tenderá a reproducirlos, por lo que parte importante del trabajo es diseñar buenas estrategias de muestreo, limpieza y validación de los datos antes de confiar en las predicciones [16].

4.4.1.2 Aprendizaje no supervisado

El aprendizaje no supervisado trabaja con conjuntos de datos sin etiquetas, es decir, sin una variable objetivo conocida de antemano. Su propósito es explorar la estructura intrínseca de la información para descubrir patrones, relaciones y regularidades que no eran evidentes antes del análisis. En lugar de predecir una salida concreta, este enfoque busca extraer conocimiento descriptivo y resumido a partir de los datos mismos [16].

Dentro de este campo hay dos usos habituales:

- Agrupamiento (*clustering*). Aquí se organiza la información en grupos o “*clústeres*” de observaciones que comparten características semejantes. El objetivo es identificar subgrupos homogéneos dentro del conjunto de datos, por ejemplo, segmentos de clientes con comportamientos parecidos, modos operacionales de una planta, o familias de eventos de falla para facilitar la interpretación y la toma de decisiones [17].
- Reducción dimensional. En casos con muchas variables (por ejemplo, largas series de sensores o atributos extraídos de procesos complejos) se busca condensar la información eliminando redundancias y conservando lo más relevante. Esto simplifica el análisis, acelera el procesamiento y mejora la visualización, permitiendo identificar factores dominantes y facilitar pasos posteriores como la monitorización o la ingeniería de características [17].

Además de estos dos enfoques, el aprendizaje no supervisado aporta otras ventajas prácticas: ayuda a detectar valores atípicos o anomalías, guía la limpieza y el pre-procesamiento de datos, y sirve como etapa exploratoria previa a modelos supervisados. En contextos industriales, por ejemplo, permite encontrar patrones operativos recurrentes, agrupar condiciones de alimentación similares, o reducir cientos de lecturas de sensores a unos pocos indicadores representativos que resumen el comportamiento del sistema.

4.4.1.3 Aprendizaje Reforzado

El aprendizaje reforzado es un enfoque de aprendizaje automático orientado a enseñar a un agente a tomar decisiones a partir de la interacción con un entorno. En lugar de aprender a partir de ejemplos etiquetados, el agente actúa, observa las consecuencias de sus acciones y recibe una

recompensa (o penalización). Esa señal de recompensa guía la adaptación del comportamiento: las acciones que aumentan la recompensa tienden a repetirse, mientras que las que la reducen se descartan [17].

Conceptualmente intervienen tres elementos: el agente (quien actúa), el entorno (donde ocurren las acciones) y la señal de recompensa (el criterio que cuantifica qué tan buena fue una acción). El ciclo acción → respuesta → recompensa se repite, permitiendo que el agente mejore su política de decisión con el objetivo de maximizar la recompensa acumulada a lo largo del tiempo.

Aunque a veces se asocia con el “aprendizaje profundo” cuando se usan redes neuronales para representar las políticas o las funciones de valor, la esencia del paradigma es la retroalimentación basada en consecuencias: aprender por ensayo y error, no, por ejemplo, respuesta predefinidos. Dos conceptos operativos importantes son la exploración (probar acciones nuevas para descubrir si son mejores) y la explotación (usar las acciones conocidas que dan buena recompensa), y su equilibrio es clave para un aprendizaje efectivo.

En la práctica industrial, el aprendizaje reforzado se ha usado para problemas donde resulta difícil especificar una regla fija, pero sí es posible definir objetivos y medir resultados: ajuste automático de control para maximizar rendimiento y minimizar energía, optimización de secuencias de mantenimiento, gestión adaptativa de recursos en tiempo real, control de robots y vehículos autónomos, o estrategias de operación que deben adaptarse a condiciones variables de planta. Un ejemplo concreto podría ser entrenar un agente para ajustar “*setpoints*” de molienda y alimentación con el fin de maximizar la producción neta mientras se minimiza el consumo energético y se respetan límites de seguridad; el agente aprende a través de iteraciones cuáles combinaciones de “*setpoints*” producen mejores recompensas operacionales [16].

No obstante, este enfoque tiene limitaciones prácticas: necesita muchas interacciones (o simulaciones realistas) para aprender bien, la definición de la función de recompensa es crítica, si se diseña incorrectamente puede inducir comportamientos indeseados, y requiere medidas de seguridad para evitar acciones que pongan en riesgo procesos reales mientras el agente explora.

4.4.2 Algoritmos de *Machine Learning*

Un algoritmo de “*Machine Learning*” es una fórmula que permite a una máquina aprender de datos en lugar de recibir instrucciones explícitas. Se usa para predecir (supervisado), descubrir estructura (no supervisado) o aprender por ensayo y error (reforzado). Para funcionar bien necesita datos limpios, una métrica de desempeño y vigilancia continua en operación para evitar degradación del modelo [16].

4.4.2.1 Algoritmos de regresión

Los algoritmos de regresión son herramientas diseñadas para estimar y modelar la relación entre una variable dependiente (lo que queremos predecir) y una o más variables independientes (las entradas). Su propósito es encontrar la función matemática que mejor aproxima los datos observados, de modo que pueda usarse para predecir valores futuros o comprender cómo influyen las variables explicativas sobre la variable objetivo [18].

A continuación, se hace una revisión de los modelos de regresión más utilizados y comunes en algoritmos de “*machine learning*”.

- **Regresión lineal:** Busca encontrar la mejor línea recta que se ajuste a los datos para predecir una variable de salida continua. En su forma más simple, para una variable independiente x y una variable dependiente y , la regresión lineal se puede expresar mediante la ecuación 5.

$$y = mx + b \quad (5)$$

Donde “ y ” es la variable dependiente, “ x ” la independiente, “ m ” la pendiente (efecto por unidad de “ x ”) y “ b ” la intersección o sesgo [18].

El objetivo del entrenamiento es encontrar los valores óptimos de m y b para minimizar la diferencia entre las predicciones y los valores reales observados.

- **Regresión lineal multivariable:** Cuando tenemos múltiples variables independiente $x_1, x_2, \dots, x_n$, la regresión lineal múltiple se expresa como (ecuación 6):

$$y = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} + \varepsilon, \quad i = 1, \dots, n \quad (6)$$

En forma matricial (ecuación 7):

$$y = X\beta + \varepsilon \quad (7)$$

Donde n es el número de observaciones, p el número de regresores (sin intercepto), $p + 1$ el número total de parámetros (incluido β_0), X tiene dimensión $n \times (p + 1)$, β es $1 \times (p + 1)$ y ε es $n \times 1$ [18].

El criterio más habitual para ajustar estos coeficientes es el método de los mínimos cuadrados ordinarios (OLS), busca el vector $\hat{\beta}$ que minimiza la suma de los cuadrados de las diferencias entre los valores observados y los predichos por el modelo [18]; así se obtiene la mejor aproximación lineal en sentido cuadrático y su expresión cerrada es (ecuación 8):

$$\hat{\beta} = (X^T X)^{-1} X^T y \quad (8)$$

El valor predicho es $\hat{y} = X\hat{\beta}$ y los residuos estimados son $\hat{\varepsilon} = y - \hat{y}$.

La varianza del estimador, bajo la homocedasticidad $Var(\varepsilon) = \sigma^2 I$ es (ecuación 9):

$$Var(\hat{\beta}) = \sigma^2 (X^T X)^{-1}, \quad \text{con } \hat{\sigma}^2 = \frac{\sum_{i=1}^n \hat{\varepsilon}_i^2}{n-(p+1)} \quad (9)$$

Además, el modelo de regresión lineal por mínimos cuadrados ordinarios (OLS) se sustenta en una serie de supuestos estadísticos que aseguran la validez y confiabilidad de sus estimaciones. Entre ellos se encuentran: la normalidad de los residuos, que garantiza la correcta aplicación de pruebas de significancia; la homocedasticidad, que implica que los errores presentan una varianza constante a lo largo de los valores predichos; y la independencia o baja colinealidad entre las variables explicativas, condición necesaria para obtener coeficientes estables y sin redundancia entre predictores.

Para evaluar esta colinealidad en la práctica se utiliza el Factor de Inflación de Varianza (VIF), el cual cuantifica cuánto aumenta la varianza de un coeficiente debido a la correlación lineal con el resto de las variables del modelo. Valores de VIF superiores a 10 suelen considerarse indicativos de multicolinealidad problemática, por lo que este indicador sirve como criterio para identificar predictores redundantes o altamente correlacionados.

El cumplimiento de estos supuestos permite interpretar los resultados del modelo OLS con un mayor grado de confianza estadística y compararlos de manera consistente con modelos regularizados o no lineales [18].

- **Regresión Lasso:** La regresión LASSO (*Least Absolute Shrinkage and Selection Operator*) es una técnica de regresión lineal que incorpora un término de penalización basado en la suma de los valores absolutos de los coeficientes. Al añadir esta restricción, el método no solo reduce la varianza del modelo evitando el sobreajuste, sino que además tiende a forzar muchos coeficientes a cero, lo que resulta en un modelo más parsimonioso y con selección automática de variables. Esta característica la hace especialmente útil en estudios aplicados

y memorias técnicas donde se dispone de muchas variables y se busca identificar las entradas más relevantes sin sacrificar interpretabilidad [19].

Desde el punto de vista operativo, LASSO equilibra el ajuste a los datos y la complejidad del modelo mediante un parámetro de regularización que controla la magnitud del castigo sobre los coeficientes. En escenarios con predictores altamente correlacionados, LASSO puede volverse inestable al elegir entre variables redundantes; por ello, en la práctica se recomienda evaluar alternativas (por ejemplo, Elastic Net) o realizar análisis complementarios para interpretar la selección de variables. La elección del parámetro de penalización suele realizarse mediante validación cruzada para obtener un balance adecuado entre sesgo y varianza, la función de pérdida es evidenciada en la ecuación 10.

$$\text{Funcion de perdida} = \sum_{i=1}^n (y_i - y'_i)^2 + \alpha \sum_{j=1}^p |w_j| \quad (10)$$

Donde y'_i son las predicciones, y_i los valores reales, w_j los coeficientes y α controla la intensidad de la penalización.

4.4.2.2 Algoritmos Bayesianos

Los algoritmos bayesianos aplican el razonamiento probabilístico para clasificar y tomar decisiones: en lugar de afirmar con certeza, trabajan con grados de probabilidad que se actualizan cuando llega nueva información. Parten del Teorema de Bayes, que permite ajustar la creencia inicial sobre un suceso a la luz de evidencia reciente. Esto los hace especialmente útiles cuando los datos son inciertos o parciales, porque entregan una medida explícita de confianza en las predicciones [17].

En términos sencillos: si ya teníamos una idea (una probabilidad *a priori*) sobre que ocurra un evento A, y observamos el evento B, el teorema nos dice cómo combinar ambas piezas de información para obtener la probabilidad *a posteriori* de A dado B. Es un proceso de aprendizaje incremental, ósea “ajustar creencias” cada vez que aparece nueva evidencia, muy intuitivo para problemas como diagnóstico médico, detección de spam o clasificación con información incompleta. La fórmula básica es representada mediante la ecuación 11:

$$P(A|B) = \frac{P(B|A) P(A)}{P(B)} \quad (11)$$

Donde:

- $P(A|B)$ es la probabilidad de que ocurra A sabiendo que B ocurrió (posterior).
- $P(B|A)$ es la probabilidad de que ocurra B sabiendo que A ocurrió (posterior).
- $P(A)$ es la probabilidad *a priori* del evento A.

- $P(B)$ es la probabilidad *a priori* del evento B .

Una implementación práctica muy difundida es el clasificador Naive Bayes, que asume independencia entre las características (una simplificación que suele funcionar sorprendentemente bien en la práctica). En entornos industriales o de planta, los enfoques bayesianos permiten, por ejemplo, combinar mediciones de sensores con conocimientos previos sobre fallas para estimar la probabilidad de un evento de interés y así priorizar inspecciones o acciones preventivas [20].

4.4.2.3 Algoritmos de árbol de decisión

Un árbol de decisión es una herramienta estructurada muy útil para seleccionar alternativas basándose en criterios gerenciales previamente definidos. De forma análoga a un diagrama de flujo, representa mediante bifurcaciones los posibles resultados que surgen al tomar una decisión. En su estructura aparecen nodos que simbolizan variables o preguntas concretas, y en las ramas se visualizan las salidas derivadas de cada prueba [17].

Funcionamiento Básico:

1. El proceso arranca en un nodo raíz, que engloba a todo el conjunto de datos.
2. En cada nodo se formula una pregunta o se evalúa una característica del conjunto.
3. Según la respuesta obtenida, el conjunto se divide en subconjuntos más homogéneos.
4. Los pasos 2 y 3 se repiten en cada subconjunto, creando nodos hijos que continúan la partición.
5. La construcción del árbol finaliza al cumplirse criterios de paro, como alcanzar una profundidad prefijada o tener menos de un número mínimo de instancias en un nodo.

4.4.2.4 Algoritmos de redes neuronales

Una red neuronal artificial está formada por muchas unidades (neuronas) organizadas en capas conectadas entre sí, de manera similar a cómo se interconectan las neuronas en el cerebro. Estas capas se comunican entre ellas: la salida de una capa sirve como entrada de la siguiente, y en conjunto procesan la información para resolver tareas concretas. Las redes permiten modelar relaciones complejas en los datos gracias a la combinación y transformación sucesiva de señales a lo largo de la arquitectura [17].

Neurona Artificial:

Una neurona artificial toma entradas ponderadas, les aplica una función de activación y produce una salida (ecuación 12).

$$z = \sum_{i=1}^n w_i x_i + b \quad , \quad a = \sigma(z) \quad (12)$$

Donde:

- n es el número de entradas.
- x_i son las entradas.
- w_i son los pesos asociados a las respectivas entradas.
- b es el sesgo.
- z es la suma ponderada de las entradas y el sesgo.
- a es la salida.
- σ es la función de activación.

Funciones de activación comunes:

- Función Sigmoide: $\sigma(z) = \frac{1}{1+e^{-z}}$
- Función ReLU (Rectified Linear Unit): $\sigma(z) = \max(0, z)$
- Función Tanh (Tangente hiperbólica): $\sigma(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}}$

4.5 Aplicaciones de Machine Learning en el Procesamiento de mineral

En el procesamiento de minerales las herramientas de *machine learning* (ML) permiten modelar relaciones complejas, identificar patrones difíciles de detectar mediante métodos tradicionales y generar modelos predictivos o interpretativos que apoyan la toma de decisiones operacionales. Su aplicación aborda desde caracterización mineralógica y la estimación de propiedades del mineral, hasta la optimización de procesos como la conminución y la flotación.

A continuación, se revisan dos estudios representativos que muestran la utilidad del *machine learning* en distintos contextos del procesamiento mineral.

4.5.1 Aplicación de ML para la predicción de balance de masa y recuperación en plantas de flotación de fosfato.

Los modelos geometalúrgicos son esenciales para la industria minera. Estos modelos combinan geología, geoquímica, mineralogía, planificación minera, mecánica de rocas y metalurgia para predecir el rendimiento del procesamiento de minerales. Este estudio aborda el desafío de predecir el comportamiento metalúrgico en una planta de fosfato, con el objetivo principal de desarrollar un modelo capaz de predecir, de manera simultánea, 7 variables críticas de respuesta (rendimientos de masa y recuperaciones de P_2O_5) a partir de variables de entrada basadas en la composición mineral (ROM) y la concentración del colector utilizado.

Se implementaron dos modelos de redes neuronales artificiales (ANN) para pronosticar las variables: uno utilizando la totalidad de las variables recolectadas y otro optimizando mediante la técnica de eliminación recursiva de características (RFE). Esta técnica permitió identificar que solo 6 variables (las leyes de P_2O_5 , Fe_2O_3 , Al_2O_3 , MgO , CaO y la dosis de colector) eran suficientes para describir el proceso, descartando elementos como la sílice o el titanio. Se compararon los pronósticos obtenidos en ambos escenarios y los resultados demostraron la robustez de la inteligencia artificial, el modelo de red neuronal entrenado con el conjunto reducido de variables, logro predecir las 7 variables metalúrgicas con una precisión muy similar al modelo completo, registrando un incremento en la raíz del error cuadrático medio (RMSE) de 2.6%. Esto prueba que es posible simplificar el monitoreo en planta sin perder capacidad predictiva [21].

4.5.2 Aplicación de computación blanda y ML en la optimización de circuitos de conminución.

Este estudio aborda el desafío de mejorar la eficiencia en la etapa de conminución, la cual representa hasta el 50% del consumo energético de una planta concentradora mediante el uso de técnicas de computación blanda (*Soft Computing*). El objetivo principal del estudio es desarrollar modelos predictivos que permitan el tamaño de partícula óptimo y estabilizar la operación frente a las variaciones del mineral bruto (ROM). Para esto, se usaron varias técnicas de ML; por un lado, aplicaron redes neuronales convolucionales (CNN) para que, mediante cámaras, el sistema “vea” y clasifique el tamaño de las rocas automáticamente, y, por otro lado, se usaron redes neuronales artificiales (ANN), la cual estos funcionan como “sensores virtuales” para estimar la distribución de tamaño de partícula.

La metodología se basa en la instalación de cámaras en las correas transportadoras para que la CNN analice las fotos en vivo, cruzando esa información con datos técnicos, tales como la potencia del molino, presión de descansos y el flujo de agua para entrenar los modelos. Como resultados destacados, el sistema de visión alcanza una precisión del 96.4% en la clasificación de fragmentos y el error (RMSE) en la predicción del tamaño de partícula final fue de 6.11%, permitiendo que el circuito reaccione de forma autónoma a cambios en la dureza de mineral. En conclusión, la aplicación de estos modelos permite una reducción significativa del consumo energético al evitar el sobre-procesamiento, demuestra que los sensores virtuales son una alternativa confiable y económica, marcando el cambio hacia una minería mucho más automatizada [22].

5 METODOLOGÍA

En este apartado, se entrega una descripción del procedimiento que se ha realizado con el propósito de realizar el modelo predictivo de la MC de la planta “A1” con datos del año 2024, este se desarrolló íntegramente en Python, integrando adquisición y limpieza de datos, filtrado y agregación temporal, selección de variables, ajuste y validación del modelo y análisis de escenarios. Para seleccionar la sección de estudio se aplicaron los mismos filtros y criterios de calidad a las 3 secciones que pertenecen a la planta, la sección elegida fue la que arrojó el mayor volumen de datos útiles tras el proceso de depuración (es decir, la mejor calidad/representatividad de registros bajo los criterios definidos). Esta decisión fue además validada en terreno con los operadores, incorporando su conocimiento operativo, lo que se denomina **juicio experto**, asegurando que los datos seleccionados fuesen representativos y operativamente relevantes.

5.1 Recopilación de Datos.

Con el objetivo de entender y analizar los parámetros operacionales ya mencionados de la planta “A1” para la correcta utilización del modelo correspondiente. se ha implementado un análisis detallado de estos datos. Este análisis se realiza mediante la utilización de *PI ProcessBook*, como se detalla en el **Anexo A**, en conjunto con la herramienta *PI DataLink* de Excel (Anexo B).

La elección de *PI ProcessBook* como plataforma de análisis se basa en su capacidad para visualizar datos en tiempo real y su interfaz intuitiva que facilita la interpretación de patrones y tendencias, además de que la empresa cuenta con la licencia correspondiente para su utilización. La información fue extraída a intervalos de un minuto como se observa en el **Anexo D**. Se recopila la información a partir del año 2024, ya que este periodo refleja con mayor precisión las condiciones actuales de la Planta. Esto se debe a que recién hace un par de años comenzó a recibir alimentación desde tres fuentes distintas. Por lo mismo, utilizar datos muy antiguos no permitiría representar de forma adecuada la situación real en la que se encuentra la Planta hoy.

5.1.1 Variables operacionales

A continuación, se detalla la extracción de los datos operacionales utilizados en el estudio, en la cual se implementó la herramienta *PI datalink* de Excel, utilizando los Tag correspondientes a cada variable. La ubicación de los sensores que registran los datos de las variables operacionales dentro del circuito se muestra en la **Figura 10**, donde cada punto de medición se encuentra identificado mediante letras. La lista de los Tag utilizados para la extracción de estas variables se especifican en el **Anexo C**, cabe mencionar nuevamente que toda la data, fue extraída por intervalos de 1 minutos durante el año 2024:

- a) Porcentaje granulométrico en la tolva gruesa: Indica la fracción de material con tamaño mayor al estándar (12700 micras) que alimenta a la molienda convencional, expresado en porcentaje sobre el valor especificado. Un alto porcentaje eleva la carga de conminución y tiende a disminuir el flujo.
- b) Alimentación seca: Corresponde al flujo masico de mineral solido que ingresa al molino de barras, expresado en toneladas por hora (tph). Determina directamente la carga de tratamiento del circuito.
- c) Flujo de agua al molino de barras y al cajón: Indica el caudal volumétrico de agua de proceso que ingresa al molino y al cajón, expresado en metros cúbicos por hora (m^3/h).
- d) Porcentaje de sólidos (in y out): Mide la proporción de mineral solido respecto al agua en la pulpa, expresada en porcentaje (%). Este parámetro se registra tanto en la alimentación al molino de barras como en la descarga del circuito, correspondiendo esta última al **overflow** proveniente de la batería de hidrociclones que se dirige posteriormente a la planta de flotación.
- e) Potencia RM y BM: Refleja el consumo energético de los equipos, medida en kilowatts (kW). Es un indicador del esfuerzo mecánico ejercido sobre la carga interna.
- f) Nivel de cajón: Corresponde al nivel de pulpa en el cajón de descarga del circuito, medido principalmente en porcentaje de altura total (%). Un nivel estable garantiza alimentación continua hacia las bombas e hidrociclones.
- g) Velocidad de bomba: Indica la velocidad de rotación del impulsor de la bomba que alimenta los hidrociclones, expresada como porcentaje (%) de su velocidad nominal o en revoluciones por minuto (rpm). Un aumento de la velocidad incrementa el caudal y la presión de alimentación, pero excesos pueden generar cavitación o daños en la bomba.
- h) Flujo pulpa a BHC y su densidad: El flujo de pulpa corresponde al caudal que alimenta las baterías de hidrociclones, medido en metros cúbicos por hora (m^3/h), mientras que la densidad se expresa en kilogramo por metro cubico (kg/m^3).
- i) Presión BHC: Corresponde a la presión de operación en la batería de hidrociclones, registrada en el sistema *PI ProcessBook* en unidades de kg/cm^2 , aunque en la práctica es habitual convertirla y manejarla en psi (libras por pulgada cuadrada). Este parámetro es el factor clave que condiciona la eficacia del proceso de clasificación.
- j) Malla +65: Indica el porcentaje de material retenido en la malla de referencia #65. Esta variable representa la granulometría del producto de la molienda. Un mayor porcentaje de malla +65 implica un producto grueso, lo que puede afectar negativamente la eficiencia de la flotación posterior.

composición y propiedades físicas distintas, es necesario cuantificar cómo se mezclan esos aportes antes de entrar al circuito: el “*blending*” condiciona fuertemente la respuesta operativa.

Para abordar esto se realizó un análisis estadístico mediante scripts en Python que calculan, por fuente y de forma agregada, estadísticas descriptivas (media, mediana, desviación estándar y percentiles) del *Work Index* y de otras variables relevantes. A partir de esos resultados se definieron rangos representativos y un rango objetivo u óptimo de *Work Index* para la tolva gruesa. Conocer estos rangos permite detectar alimentaciones que podrían provocar un desempeño subóptimo, por ejemplo, entradas con *Work Index* excesivamente altos o bajos, y fundamenta criterios operacionales para estrategias de mezcla o control que estabilicen la planta.

Los datos utilizados provinieron del *dashboard* de planta y fueron extraídos con resolución horaria para todo el periodo de estudio; la serie horaria y el procedimiento de extracción están evidenciados en el **Anexo F** (Figura 27). Los procedimientos de procesamiento están disponibles en el mismo Anexo, lo que facilita la reproducibilidad del análisis y la trazabilidad de las conclusiones.

5.2.2 Comportamiento de la granulometría de alimentación y del rendimiento en base al *Work Index*

Para estudiar la relación entre la granulometría de alimentación, el rendimiento de cada una de las secciones (13, 14 y 15) y el *Work Index* (WI) se trabajó con tres series temporales distintas: el *Work Index* (resolución horaria), y la granulometría y el rendimiento (resolución minutar). La granulometría se expresa como porcentaje respecto al estándar de 12700 μm (% sobre el estándar); las series y el procedimiento de extracción están documentados en **Anexo F**.

El procedimiento fue el siguiente:

- **Filtrado del rendimiento:** En Excel se filtraron los registros de rendimiento (tph) conservando únicamente valores en el rango 350–570 tph. Este umbral se estableció para excluir eventos de transferencia desde otras secciones (valores >570 tph) y para definir el mínimo operativo representativo (≤ 350 tph se marca como procesamiento bajo).
- **Ajuste de resolución temporal:** Dado que el WI está registrado por hora y las otras variables por minuto, los registros de rendimiento filtrados se agregaron (promediaron) por hora para alinearlos temporalmente con el WI. El mismo procedimiento de agregación se aplicó a la granulometría.
- **Filtrado de granulometría:** Antes de promediar, la granulometría se limitó al rango 1%-34% sobre el estándar para concentrar el análisis en el intervalo operativo relevante; posteriormente estos valores se promediaron por hora.

- **Análisis y visualización:** Con las tres series emparejadas por hora (WI, granulometría media horaria y rendimiento medio horaria) se generaron gráficos temporales y de dispersión que muestran la evolución de la granulometría y del rendimiento a medida que aumenta el *Work Index*. Estos gráficos permiten identificar tendencias, picos y posibles zonas críticas donde la mezcla afecta el desempeño.

La agregación horaria permite comparar directamente con el WI, pero atenúa variaciones de alta frecuencia; por eso es importante complementar el análisis con inspecciones puntuales si se detectan anomalías. Los umbrales escogidos (350 y 570 tph; 1%–34% de granulometría) responden a criterios operativos para excluir transferencias y valores no representativos; sin embargo, son parámetros que pueden ajustarse en análisis de sensibilidad para evaluar su impacto en los resultados.

5.2.3 Comportamiento del rendimiento en base a la granulometría

Usando las series ya promediadas a resolución horaria (método descrito en 5.2.2), se investigó la relación entre el porcentaje granulométrico de la alimentación y el rendimiento (tph). Para ello se construyeron gráficos de dispersión y series temporales y se calcularon estadísticas descriptivas (media, mediana, percentiles) por tramos de granulometría, con el objetivo de evaluar cómo responde el rendimiento al aumentar la fracción granulométrica en la tolva. Este análisis permite detectar tendencias, sensibilidad operativa y posibles umbrales donde el rendimiento comienza a degradarse o mejorar, información útil para ajustar criterios de mezcla y control de alimentación.

5.3 Modelo estadístico predictivo

El desarrollo del modelo estadístico predictivo del rendimiento de la planta se apoyó mayoritariamente en scripts implementados en Python. Aunque la sección recoge el modelo descriptivo/final, su obtención requirió una secuencia ordenada de procesos y verificaciones previas para garantizar calidad de datos, robustez y reproducibilidad. En términos generales, se procedió primero a depurar la base (filtrado IQR), aplicar filtros de rango operativos, alinear resoluciones temporales, realizar ingeniería y selección de variables, y finalmente particionar los datos para entrenamiento, validación y prueba antes del ajuste del modelo y la evaluación con métricas estándar.

Adicionalmente, durante la etapa de filtrado se incorporó un criterio explícito para el manejo de datos faltantes, la cual considera en conservar únicamente aquellas filas que presentaran a lo más dos valores ausentes, descartando toda fila que exhibiera tres o más. Esta decisión buscó asegurar una señal operativa suficientemente representativa sin reducir en exceso el tamaño muestral, evitando

dos riesgos contrapuestos, por un lado, trabajar sólo con escenarios “perfectos” que no reflejan la variabilidad real de la planta; y por otro, incluir observaciones con demasiada incompletitud, lo que podría introducir ruido o sesgos en la estimación. El umbral de dos faltantes se definió tras revisar la densidad temporal de variables críticas y en coordinación con personal de planta, quienes indicaron que bajo condiciones reales es habitual que ciertos sensores presenten breves pérdidas de señal sin afectar la operación global, por lo que mantener observaciones con hasta dos ausencias resulta operativamente razonable y estadísticamente aceptable para preservar estabilidad, representatividad y trazabilidad del modelo.

5.3.1 Filtro IQR

Para garantizar la calidad de las series temporales se aplicó un filtrado por rango intercuartílico (IQR) a cada variable operacional antes de la agregación y el modelado. Se calculó utilizando la ecuación 13.

$$IQR = Q3 - Q1 \quad (13)$$

Donde $Q3$ y $Q1$ son cuartiles (75% y 25%, respectivamente) y se definieron límites (superior e inferior) como $[Q1 - 1,5 * IQR, Q3 + 1,5 * IQR]$ (factor $k = 1,5$). Los registros que quedaron fuera de ese intervalo fueron considerados atípicos y se eliminaron del conjunto de trabajo para evitar que picos espurios (errores de registro, ruidos instrumentales o transferencias interseccionales) sesgaran las estimaciones del modelo. Este preprocesamiento redujo la presencia de valores extremos y mejoró la robustez de las estadísticas calculadas posteriormente [23].

El procedimiento se ejecutó de forma reproducible mediante scripts en Python; los listados y la cuantificación de registros eliminados (o *Outliers*) por variable se encuentran documentados en el **Anexo H** (ver Anexo H: Tabla resumen de *Outliers*)

5.3.2 Filtrado manual basado en el estándar operacional

El ajuste final de los umbrales operativos se realizó tras comprobaciones en terreno con los operadores de planta, integrando su conocimiento operativo a la evaluación estadística previa. Aunque se tomó el estándar como referencia inicial, los límites se adaptaron empíricamente buscando dos objetivos: (1) reflejar fielmente las condiciones reales de operación del periodo estudiado y (2) reducir la presencia de registros espurios que generaban un exceso de “*Outliers*” y que podían sesgar el modelado. Estas modificaciones se aplicaron siempre conservando la trazabilidad y manteniendo los criterios operativos acordados con respecto al estándar operacional de la planta evidenciados en la **Tabla 2**. Además de la validación con operadores, la cual asegura

que el filtrado sea robusto y alineado con la práctica operativa, sin eliminar eventos reales relevantes para el análisis.

5.3.3 Modo estable

Para asegurar que el conjunto de entrenamiento contenga registros operacionales homogéneos, se implementó un “modo estable” que filtra filas en base al **coeficiente de variación (CV)** calculado en ventanas móviles de 15 minutos. El procedimiento consistió en calcular, para cada instante t , el promedio y la desviación estándar de la variable dentro de la ventana móvil y a partir de ello el CV, como se evidencia en la ecuación 14.

$$CV_t = \frac{Desviacion\ estandar_t}{|promedio_t| + \epsilon} \quad (14)$$

Donde ϵ es un valor pequeño adimensional usado para evitar división por cero.

Se considera una fila **estable** cuando el CV de cada variable fijada quedo por debajo de un umbral, que denominamos CV_{max} , se considera un umbral del 5%, lo que, en términos de desviación estándar, equivale a lo mostrado en la ecuación 15.

$$Desviacion\ estandar_t < 0,05 * Promedio_t \quad (15)$$

Las variables sobre las que se aplicó este criterio fueron la potencia del molino de bolas y el porcentaje de sólidos en el molino de barras. Estas 2 variables fueron escogidas debido a que en planta no poseen cambios muy bruscos a medida transcurre el tiempo, por eso un cambio muy radical en la data indica una mala lectura de los sensores correspondientes produciendo ruido al sistema. El cálculo se realizó en las ventanas móviles de 15 minutos sobre la serie de alta resolución; las filas que no cumplieron la condición para ambas variables fueron excluidas del *dataset* de entrenamiento.

5.3.4 Correlaciones operacionales

Este apartado es de gran importancia debido a que es posible visualizar de manera clara las correlaciones de cada variable operacional, usando el coeficiente de Pearson la cual mide la intensidad y dirección de la relación lineal entre dos variables cuantitativas. Se define como la covarianza entre X e Y dividida por el producto de sus desviaciones estándar [8], la cual se evidencia en la ecuación 16.

$$r = \frac{cov(X,Y)}{\sigma_X \sigma_Y}, \quad -1 \leq r \leq 1 \quad (16)$$

Si:

- $r = +1.00$ → Correlación positiva perfecta (ambas suben y bajan al unísono).
- $r = 0.00$ → Ninguna relación lineal aparente.
- $r = -1.00$ → Correlación negativa perfecta (cuando una sube, la otra baja).

Este análisis se realizó con un fin exploratorio y visual al generar mapas de correlación y gráficos que faciliten la identificación intuitiva de variables potencialmente relevantes para el modelo predictivo. Es decir, se empleó Pearson como herramienta de cribado visual para priorizar las variables candidatas al modelo y entender como dependen unas de otras dentro del circuito de molienda.

Este tipo de enfoque se optó porque, si bien la teoría proporcione un marco de referencia para entender el comportamiento de la planta, el análisis de correlaciones permite descubrir interacciones operacionales reales que pueden quedar fuera de las ecuaciones tradicionales. Este enfoque empírico se utiliza para 2 principales razones:

- Priorizar variables claves: Al identificar los coeficientes de correlación más elevados, se sabe dónde enfocar ajustes, y quitar variables muy correlacionadas entre sí.
- Validar supuestos: Cuando una relación esperada no aparece (o es débil), se cuestiona la vigencia de los modelos teóricos, lo cual puede dar pie a investigar factores ocultos o condiciones de operación cambiantes.

5.3.5 Determinación y aplicación de desfases operacionales (*Lags*)

Con el objetivo de capturar la respuesta retardada de las variables operacionales frente al flujo mineral (rendimiento), se analizó el comportamiento del sistema tras un cambio de 5 tph en el rendimiento, este procedimiento de desfases temporales (*lags*) se implementó en Python mediante el cálculo de correlaciones cruzadas entre la variable objetivo y las variables de entrada. Este método permitió estimar el tiempo que tarda cada parámetro operacional en impactar el rendimiento de la planta, considerando la dinámica hidráulica y mecánica propia del circuito de molienda convencional. A partir de los datos recopilados, se evaluaron desfases temporales dentro de una ventana de **25 minutos**, considerando que ese lapso corresponde al tiempo total que tarda el proceso de molienda convencional en completar su ciclo hasta la etapa de flotación. Este periodo incluye tanto el tiempo de residencia del mineral en los molinos, como el trayecto del flujo a través del cajón y las bombas, además del tiempo de permanencia en la batería de hidrociclones.

La correlación entre las variables desplazadas y el rendimiento (variable objetivo) se determinó mediante el coeficiente de Pearson (ecuación 16). De este modo, se identificó el *lag* óptimo para cada variable, definido como el desplazamiento que maximiza el valor absoluto de dicha correlación. Para asegurar la validez del modelo, se aplicó una condición de causalidad que limitó el análisis

exclusivamente a desfases positivos (≥ 0), de manera que la variable predictora siempre preceda temporalmente al rendimiento. Los *lags* negativos se descartaron, ya que, indicarían que los cambios operacionales ocurren después de la variación en el flujo mineral (rendimiento), lo que implicaría predecir el pasado a partir del futuro. Esta situación carece de sentido físico en un proceso en línea y, por tanto, no resulta válida para el desarrollo de un modelo predictivo real.

Finalmente, se estableció un umbral mínimo de correlación de $|r| \geq 0.05$ para recomendar el uso de un *lag* en el modelado; en caso contrario, se asignó un desfase de 0 minutos. El resultado del procedimiento fue una tabla ordenada de mayor a menor correlación, indicando para cada variable el *lag* más representativo y su correlación asociada.

5.3.6 Desarrollo y evaluación de modelo predictivo del rendimiento

5.3.6.1 Preparación del conjunto de datos

Se trabajó con registros operacionales minuto a minuto del circuito, convirtiendo todas las variables a formato numérico y conservando únicamente los parámetros de proceso. Se excluyeron los indicadores de calidad, como **Malla +65** y **porcentaje de sólidos a flotación**, por no corresponder a variables de control y para evitar dependencias no causales con el rendimiento (flujo mineral, variable objetivo). Posteriormente, se incorporaron los desfases operacionales (*lags*) definidos en el análisis previo, desplazando cada predictor según su tiempo de respuesta característico.

Finalmente, se optó por una distribución aleatoria de los datos, asignando un 70% de los registros al entrenamiento y el 30% restante para la etapa de prueba. Esta selección al azar busca que el modelo sea evaluado con una muestra representativa de la variabilidad operativa registrada durante todo el año, permitiendo validar su desempeño frente a diversas condiciones de la planta.

5.3.6.2 Modelos aplicados

Con el objetivo de analizar el comportamiento del flujo mineral y evaluar la factibilidad de construir un modelo predictivo, se aplicaron dos enfoques de modelamiento basados en aprendizaje supervisado: **regresión lineal múltiple (OLS) con selección de variables mediante Lasso** y **bosques aleatorios (Random Forest)**. Cada modelo fue entrenado utilizando el mismo conjunto de datos preprocesados.

El primer enfoque de modelamiento corresponde a una regresión lineal con regularización por Lasso (*Least Absolute Shrinkage and Selection Operator*), la cual integra los principios de los mínimos cuadrados ordinarios (OLS) con una técnica de selección automática de variables. Mientras que el modelo OLS convencional (ecuaciones 6 a 9) estima los coeficientes que minimizan la suma de los

errores cuadráticos para aproximar el comportamiento promedio del proceso, la incorporación de la penalización Lasso (ecuación 10) permite reducir el sobreajuste y mejorar la interpretabilidad del modelo. De esta forma, el modelo retiene solo los predictores más relevantes para explicar la variación del flujo mineral, eliminando aquellos que presentan menor contribución o alta redundancia. La selección de la penalización óptima (α) se realizó mediante validación cruzada (*cross-validation*) aplicando la regla del 1-SE, que consiste en elegir el valor más grande de α cuyo error medio se mantiene dentro de una desviación estándar del mínimo obtenido en la validación cruzada. Este criterio favorece un modelo más simple y estable, evitando el sobreajuste y mejorando su capacidad de generalización.

Finalmente, se aplicó el modelo **Random Forest**, para capturar las relaciones no lineales y posibles efectos de saturación característicos del proceso de molienda. Al combinar múltiples árboles de decisión, el algoritmo facilitó la identificación de interacciones complejas entre variables operacionales que un modelo lineal podría omitir. Además de su evaluación predictiva, se utilizó este método para jerarquizar los parámetros según su impacto real en la variabilidad del rendimiento, empleando el análisis de importancia por permutación como herramienta de diagnóstico complementaria.

5.3.6.3 Evaluación comparativa de desempeño

Para evaluar el desempeño de los modelos aplicados, se utilizaron métricas estadísticas que permiten cuantificar la capacidad de ajuste y la precisión predictiva de cada enfoque. En todos los casos, la evaluación se realizó sobre el conjunto de prueba, manteniendo los mismos datos de entrenamiento y validación para garantizar la comparabilidad de los resultados.

Las métricas empleadas fueron el **coeficiente de determinación (R^2)**, el **error absoluto medio (MAE)**, la **raíz del error cuadrático medio (RMSE)** y el **error porcentual medio absoluto (MAPE)**. El R^2 expresa el grado en que las variables operacionales explican la variabilidad del flujo mineral, mientras que las métricas de error permiten estimar la magnitud promedio de las desviaciones entre los valores observados y los predichos. El MAE mide el error medio en las mismas unidades del proceso, las cuales son toneladas por hora de mineral (tph), el RMSE penaliza de manera más significativa los errores grandes, y el MAPE entrega el error relativo en porcentaje, facilitando la interpretación del desempeño en términos proporcionales. Por requerimientos operacionales de la empresa se estableció un umbral mínimo aceptable de $R^2 \geq 0.70$ para considerar un modelo adecuado para toma de decisiones.

Cada modelo empleado, regresión lineal OLS con selección de variables mediante Lasso y Random Forest fue evaluado bajo los mismos criterios, permitiendo comparar su comportamiento de forma

objetiva a través de las métricas mencionadas. Esta evaluación se complementó con la verificación de supuestos estadísticos en el modelo lineal y con el análisis de importancia de variables en los modelos regularizados y no lineales. Los valores obtenidos de cada métrica se presentan en el capítulo de resultados, junto con su interpretación y análisis comparativo.

6 RESULTADOS

6.1 Evaluación de las variables de alimentación a la molienda convencional.

6.1.1 Resultados de la evaluación de blending mineral y Work Index.

El análisis del *blending* aporta información relevante sobre cómo se distribuye el mineral que ingresa a la tolva gruesa. Dicha tolva recibe material de tres fuentes, rajo, subterránea y mineral de RT que alimentan la molienda convencional. En la **Figura 12** se muestra la proporción porcentual del mineral que entra a la tolva, organizada por rangos crecientes de *Work Index*; adicionalmente, en la **Figura 11** se presenta la cantidad (distribución) de registros contenidos en cada uno de esos rangos.

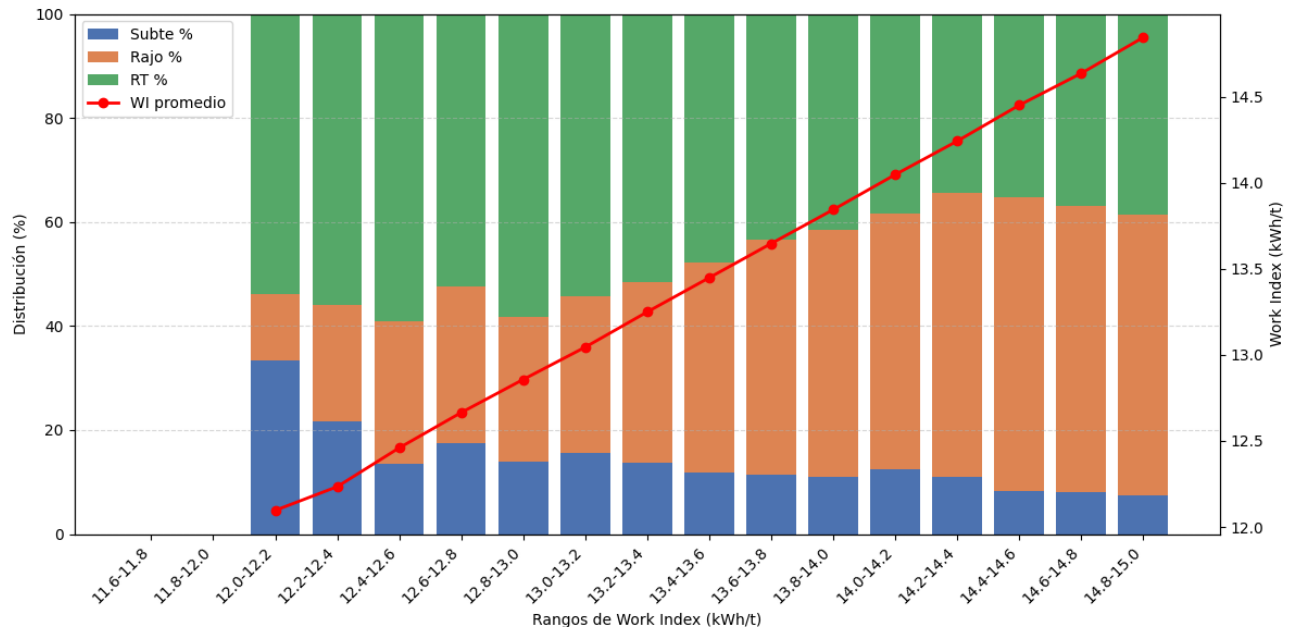


Figura 11. Distribución mineral que ingresa a la tolva gruesa por rangos de Work Index. Elaboración propia

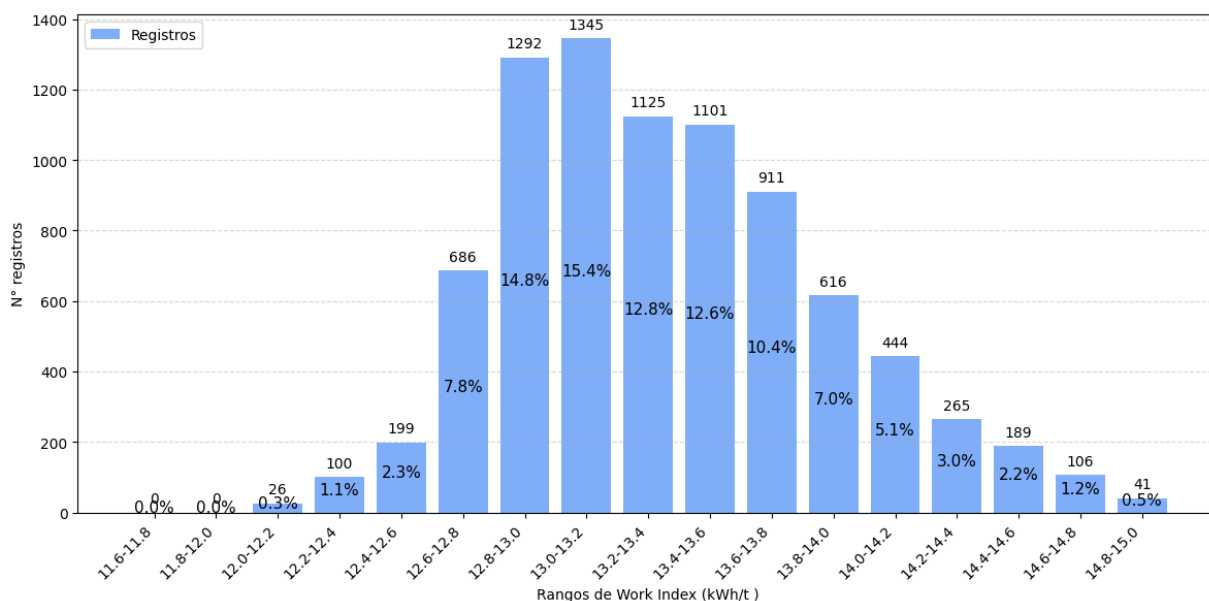


Figura 12. Distribución diaria de mineral que ingresa a la tolva gruesa y Work Index. Elaboración propia

En total se analizaron 8500 registros correspondientes al periodo 2024. Las **Figuras 11 y 12** muestran de manera complementaria la relación entre la composición de la tolva gruesa y el *Work Index* (WI). Se aprecia una tendencia clara, al incrementarse la participación del material procedente del rajo, el WI se eleva notablemente; en contraste, cuando predomina el aporte subterráneo, el WI se mantiene en los valores más bajos. El 81 % de los registros se concentra en el intervalo de 12.6 a 14.0 kWh/t, con un WI promedio de 13.3 kWh/t. Además, como complemento, en la **Figura 29** (Anexo G) se muestra la distribución diaria del mineral que ingresa a la tolva gruesa; visualmente queda claro que la mayor parte del tiempo la tolva está dominada por material RT y rajo, mientras que la contribución del mineral subterráneo es mucho menor. Esta composición predominante tiene implicaciones directas para el modelado, debido a que la distribución de esta afecta directamente al *Work Index*.

6.1.2 Evaluación de la granulometría y rendimiento en base al Work Índice

El análisis del comportamiento del flujo mineral permite evaluar la estabilidad operacional de las secciones de molienda convencional y su relación con variables de resistencia del mineral, como el *Work Index*, que influye directamente en la capacidad de tratamiento y la granulometría del producto. En la **Figura 13** se observa la tendencia de los flujos de mineral correspondientes a las secciones 13, 14 y 15 durante el periodo analizado (Año 2024), junto con la variación de la granulometría asociada.

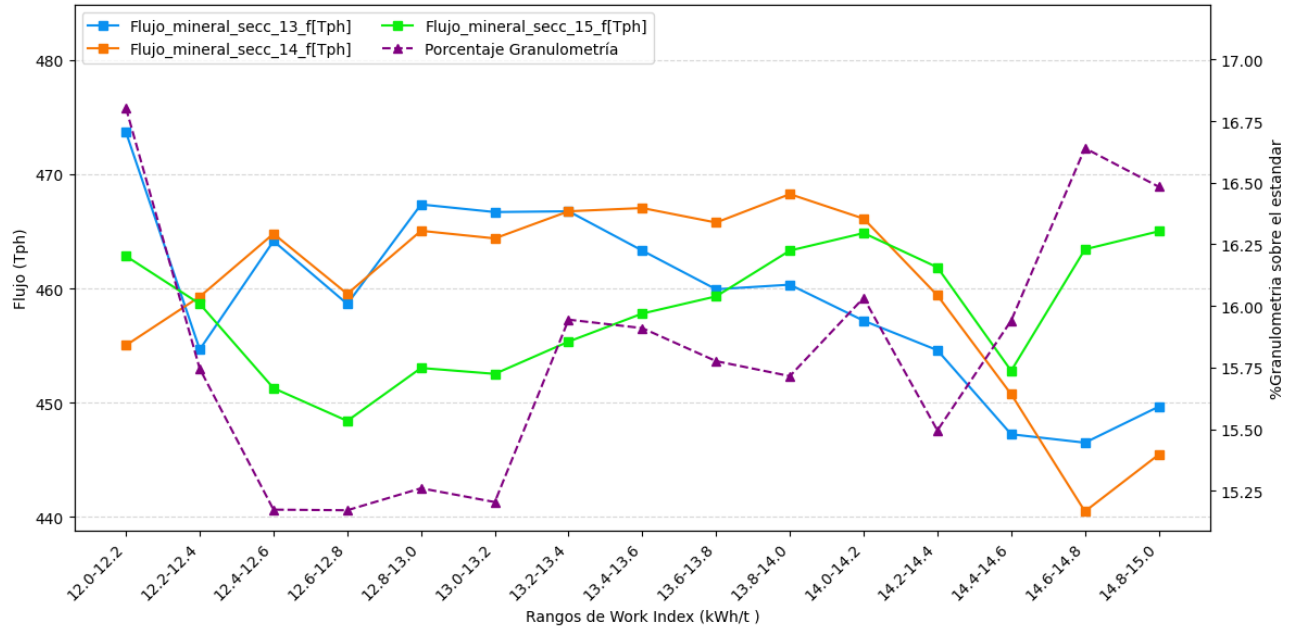


Figura 13. Comportamiento de la granulometría y rendimiento en base al Work Index. Elaboración propia

Al analizar la **Figura 13** se aprecia que las tres secciones muestran comportamientos diferenciados, con rendimientos que oscilan aproximadamente entre 445 y 470 tph. La sección 14 exhibe, en buena parte del rango, los mayores valores de producción, mientras que la sección 15 registra los niveles más bajos, pero con una respuesta uniforme. Se identifica además una relación inversa entre la granulometría y el rendimiento, y una relación directa entre la granulometría y el *Work Index*. A partir de valores de WI sobre 14 kWh/t se observa una disminución marcada del rendimiento en las tres secciones acompañada de un aumento de la fracción gruesa, lo que evidencia el impacto de la dureza del mineral sobre la tasa de procesamiento y la estabilidad del flujo.

Se observa un comportamiento no regular, en los tramos finales de la **Figura 13**. Cuando el WI supera los 14.6 kWh/t, se evidencia un incremento en el flujo mineral (rendimiento), especialmente en la sección 15, alcanza un rendimiento promedio de 465 tph. Este fenómeno no concuerda con los principios de la conminución, ya que, según Ley de Bond evidenciada en la ecuación 1, el WI es una medida directa de la resistencia del mineral, por lo tanto, a mayor WI, mayor es la energía requerida para alcanzar la reducción de tamaño deseada en la planta. Entonces un mineral más duro debería forzar una reducción de rendimiento para permitir que el molino entregue la energía necesaria. De lo contrario, se corre el riesgo de una liberación deficiente de las partículas que impacta negativamente procesos posteriores, o incluso fallas mecánicas severas, como lo es el desplazamiento de las barras en los molinos. Sin embargo, al observar la **Figura 12**, los registros de WI superiores a 14.6 kWh/t solo representan el 3.9% de la totalidad de los datos, por lo tanto, al ser un segmento con tan baja representatividad, no es posible discriminar con certeza el

comportamiento real a valores muy altos de WI, ya que, la escasez de muestra provoca que cualquier valor atípico desplace excesivamente el promedio, distorsionando la tendencia esperada de la operación.

6.1.3 Evaluación rendimiento y *Work Index* en base a la granulometría.

Comparar el impacto del rendimiento en función de la granulometría permite identificar qué factor afecta en mayor medida a la molienda convencional: si lo determinante es la distribución de tamaño de alimentación o el *Work Index*. El estándar de referencia para la granulometría de alimentación es 12.700 micrones, pero en la operación dicho estándar se refiere al **P80**. Por tanto, este análisis no solo evalúa cómo responde el rendimiento ante variaciones en la fracción gruesa, sino que además cuantifica el “sobretamaño” que llega al circuito de molienda convencional. En la **Figura 14** se muestran los registros y los promedios agrupados por rangos de granulometría desde un 3% hasta un 31% para facilitar esta comparación.

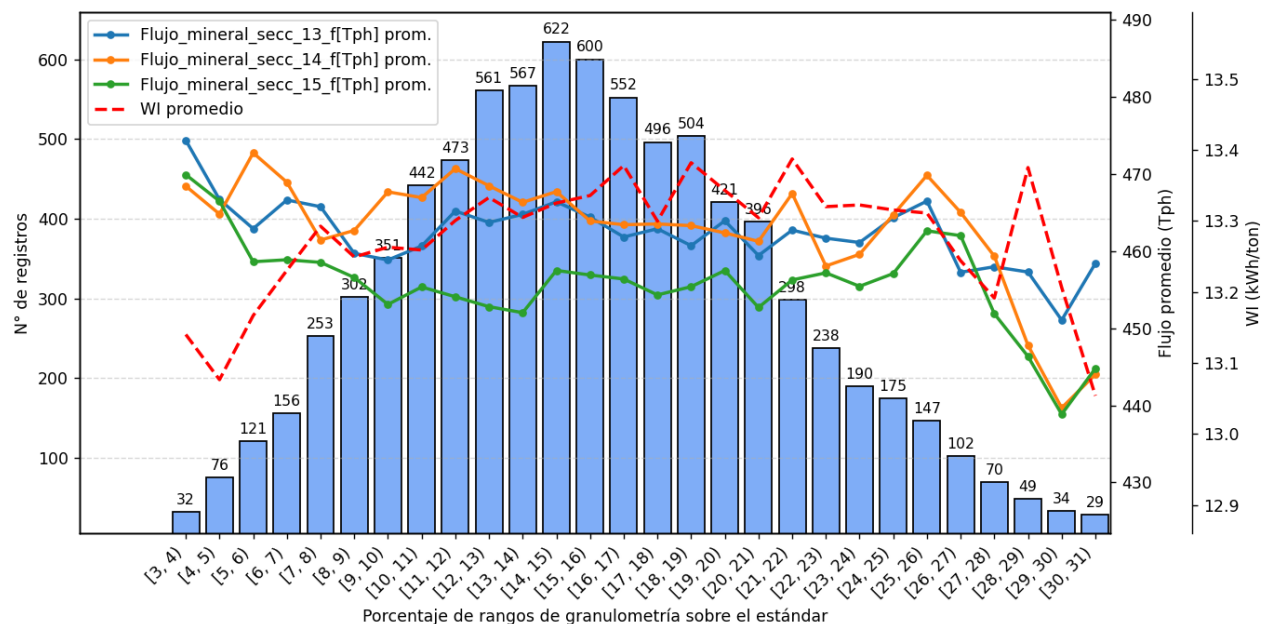


Figura 14. Comportamiento del rendimiento en base a rangos granulométricos. Elaboración propia

La evaluación reúne un total de 8257 registros y muestra que la mayor parte de las observaciones (el 72%) corresponden a alimentaciones con, como máximo, un 20% de sobretamaño respecto al estándar granulométrico; en esos rangos la alimentación hacia molienda convencional cumple con la referencia operativa. Se aprecia un comportamiento óptimo de la planta cuando el porcentaje de sobretamaño es bajo (aproximadamente 3 - 8%), donde las tres secciones alcanzan los picos más altos de rendimiento. En contraste, al superar el 25% de sobretamaño todas las secciones experimentan una caída significativa de la producción: en ningún caso promediaron más de 465 tph

en esos rangos, lo que identifica dicho umbral como crítico. Para rangos intermedios el efecto sobre el rendimiento no resulta tan claro a simple vista. Además, los datos confirman la relación entre granulometría y *Work Index*, los valores de WI bajos tienden a asociarse con menores porcentajes de sobretamaño.

6.2 Implementación del modelo predictivo (filtros y entrenamiento)

6.2.1 Aplicación de Filtro IQR

La aplicación de este filtro permite establecer límites estadísticos y eliminar aquellos valores que queden fuera de dichos umbrales. Antes de detallar el procedimiento conviene recordar que las series se obtuvieron con resolución de un minuto para el año 2024; por tanto, el conjunto original contenía **527.040** registros antes de aplicar el filtro correspondiente. Las tablas de los límites de las variables arrojadas por el filtro y las observaciones detectadas como atípicas se presentan en el **Anexo H**. Los datos resumidos por sección, que muestran la cantidad de *Outliers* detectados tras aplicar el IQR y el porcentaje de filas válidas resultante, se muestran en la **Tabla 3**.

Tabla 3. Cantidad de *Outliers* de las secciones de la planta “A1”.

Sección	Outliers	% Filas Válidas
13	150818	71.38
14	122866	76.69
15	130300	75.28

Al analizar la **Tabla 3** se observa que la aplicación del filtro IQR produjo una reducción considerable del volumen de datos disponibles en todas las secciones, eliminando del orden del 25 % de los registros. No obstante, en la sección 13 la reducción fue un poco más severa, alcanzando un 28.62% de filas eliminadas, lo que evidencia una mayor presencia de valores atípicos o inconsistentes. Este resultado es especialmente relevante si se considera que una fila no es descartada automáticamente por tener hasta dos datos faltantes, lo que implica que la reducción observada no se debe únicamente al criterio de incompletitud, sino a la existencia de un número significativo de filas que simultáneamente presentan valores fuera de rango y múltiples variables sin información. En términos prácticos, este comportamiento sugiere que la sección 13 concentra episodios de menor calidad de datos, probablemente asociados a pérdidas de señal, ruidos de instrumentación o variaciones operacionales abruptas. En conjunto, la tabla refleja que, aun con un criterio de tolerancia moderada a la falta de datos, la depuración mediante IQR permitió estabilizar la base y excluir registros que no aportaban información confiable para el entrenamiento del modelo

La efectividad y coherencia del filtrado queda corroborada al comparar los límites extremos que aparecen en las tablas de cada sección (Anexo H) con el estándar operacional presentado en la **Tabla 2**, además se recalca que los valores destacados (marcados en rojo en el Anexo) se encuentran fuera del rango operativo habitual de planta. No obstante, se enfatiza que estos parámetros extremos no deben eliminarse de forma indiscriminada, cualquier ajuste debe realizarse con criterio operativo y sin suprimir en exceso observaciones representativas. En consecuencia, las decisiones finales sobre el filtrado manual de estos registros fueron tomadas siguiendo el **juicio experto** de los operadores de planta.

6.2.2 Filtrado manual de variables fuera de rango operacional

Tras la aplicación del filtro IQR (Sección 6.2.1) se detectaron algunas variables cuyos límites quedaban alejados del rango operativo habitual de planta. Por ello, se procedió a un ajuste manual de ciertas variables, corrigiendo sus límites inferiores y superiores para que reflejen la realidad del proceso. En este paso, resulto fundamental fijar el límite superior de *Work index* en 14.6 kWh/t, una decisión basada en los análisis de dureza (Sección 6.1.2), debido a que, el rendimiento aumentaba sobre ese valor de WI, por lo tanto, al ser un comportamiento no regular y además poco representativo para el modelo se procedió a restringir el parámetro. Se presentan las variables consideradas para cada una de las secciones y los límites aplicados en este filtro manual en las **Tablas 4-6**.

Tabla 4. Filtrado manual de variables para la sección 13.

Variables	Límite bajo	Límite alto
<i>Rendimiento [tph]</i>	350	540
<i>% Solidos RM</i>	68.82	90
<i>Flujo pulpa a BHC [m³/h]</i>	300	3500
<i>Flujo agua a cajón [m³/h]</i>	189.34	600
<i>Flujo agua a RM [m³/h]</i>	60	212.24

Tabla 5. Filtrado manual de variables para la sección 14.

Variables	Límite bajo	Límite alto
<i>Presión BHC [Kg/cm²]</i>	0.2	0.85
<i>Flujo pulpa a BHC [m³/h]</i>	300	3500
<i>Flujo agua a cajón [m³/h]</i>	152.53	600

Tabla 6. Filtrado manual de variables para la sección 15.

Variables	Límite bajo	Límite alto
<i>Presión BHC [Kg/cm²]</i>	0.2	0.84
<i>Flujo agua a cajón [m³/h]</i>	170.98	600
<i>Flujo agua a RM [m³/h]</i>	80	209.98

Al revisar la **Tabla 4** (sección 13) se aprecia que ésta requiere muchos más ajustes manuales de variables operacionales que las otras dos secciones, que solamente presentaron tres variables cada una para corregir. La mayor parte de las correcciones correspondieron a flujos, tanto de agua como de pulpa, este último con dirección hacia la batería de hidrociclones, ya que esos canales mostraron lecturas claramente fuera del rango operativo.

Para fijar algunos límites (por ejemplo, el del rendimiento y el del flujo de agua al cajón) se utilizó el indicador *Profit* (ver Anexo I) como referencia para el límite superior; el resto de los umbrales se definieron mediante juicio experto de operación. Al comparar los resultados del filtro IQR (Anexo H: tablas resumen de *Outliers*) con los límites ajustados manualmente se observa que, en general, los cambios aplicados no fueron drásticos; sin embargo, hay dos excepciones notables: el rendimiento en la sección 13 (Tabla 4) y el flujo de pulpa hacia la batería de hidrociclones en la sección 14 (Tabla 5).

En el primer caso, el IQR arrojó un límite inferior irreal (por encima de 950 t/h), mientras que un límite demasiado conservador como 0 tph tampoco es útil porque no refleja condiciones normales de operación; por esa razón se optó por fijar el rango operativo manual en **[350 – 540] tph**, considerando que valores muy bajos indican detenciones o paradas de equipo. En la sección 14, el IQR propuso un límite superior de **5245 m³/h** para el flujo de pulpa hacia los hidrociclones, valor que igualmente resulta excesivo para la capacidad real de la batería.

6.2.3 Filtrado mediante modo estable

Al aplicar el **modo estable** como último paso de depuración para obtener el “*Dataframe*” definitivo destinado al modelo predictivo, en la **Tabla 7** se resumen los resultados del filtrado: cantidad de filas eliminadas y porcentaje de filas válidas que permanecen. Estos valores permiten cuantificar el efecto del filtro y verificar que el conjunto final sea representativo y apto para el entrenamiento del modelo.

Tabla 7. Resumen del porcentaje de filas válidas para la aplicación del modelo.

<i>Sección</i>	<i>Outliers IQR</i>	<i>Outliers</i>	<i>Outliers Modo</i>	<i>% Filas Válidas</i>
		<i>Manual</i>	<i>estable</i>	
13	150818	102218	12539	49.61
14	122866	44657	29354	62.64
15	130300	12817	18867	69.27

En cuanto a la cantidad de *Outliers*, se aprecia que la mayor eliminación de registros la realiza el filtro IQR, el ajuste manual descarta una porción adicional en las secciones 14 y 15, pero en cambio la sección 13 descarta más de 100000 *Outliers* al aplicar los filtros manuales y el modo estable hace una limpieza fina final (valores menores en *Outliers* Modo estable).

De forma particular para cada una de las secciones:

- **Sección 13** es la que sufrió la mayor depuración: el IQR detectó **150818** *Outliers* y, tras los ajustes manuales y el modo estable, la proporción de filas válidas quedó en **49.61 %**. Esto indica que los datos de la sección 13 son menos representativos y/o más ruidosos bajo los criterios definidos, por lo que deben ser tratados con cautela en el modelado.
- **Sección 14** muestra una reducción relevante también (IQR: **122866**; manual: **44657**; modo estable: **12817**) y conserva **62.64 %** de filas válidas.
- **Sección 15** presenta el mejor balance, aunque el IQR detectó **130300** *Outliers*, los filtros adicionales descartaron menos registros y, finalmente, quedó con **69.27 %** de filas válidas, lo que sugiere una mayor estabilidad operativa y una mayor cantidad de datos útiles para entrenamiento del modelo.

6.2.4 Evaluación de correlaciones operacionales

El análisis mediante el coeficiente de Pearson muestra que las mismas parejas de variables se relacionan de forma distinta según la sección considerada. En la **Figura 15** se presentan los mapas de correlación de cada sección, donde se puede apreciar visualmente cómo varían en intensidad y signo las asociaciones entre variables en cada línea de molienda.

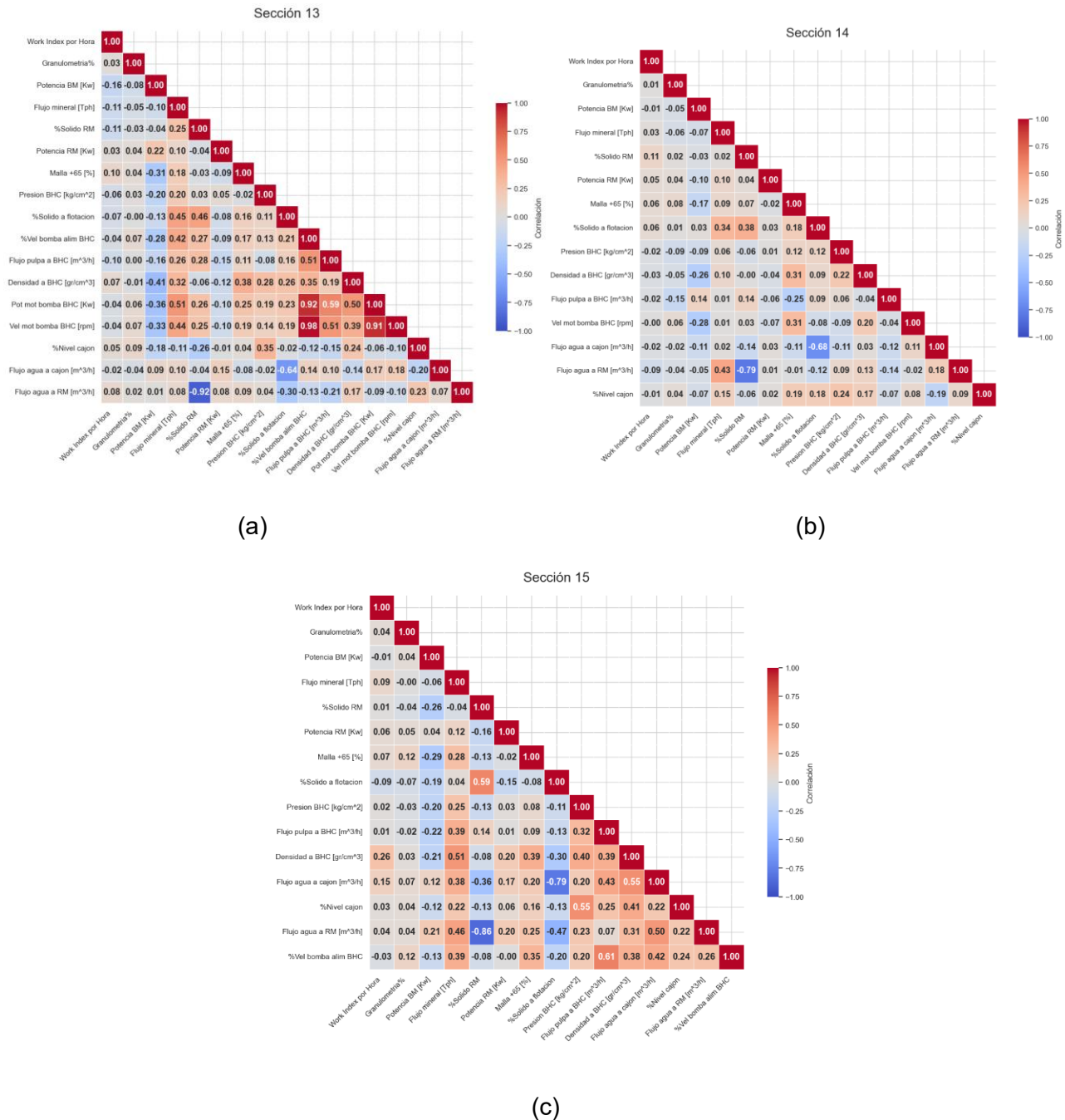


Figura 15. Diagrama de correlaciones de Pearson.

A pesar de que el análisis de correlaciones evidencia comportamiento operativos distintos para cada sección de molienda, se identificaron dos asociaciones clave (con un coeficiente de Pearson bastante elevado) presentes en las tres líneas: (i) el porcentaje de sólidos que va a flotación se relaciona **negativamente** con el flujo de agua hacia el cajón de descarga, y (ii) el flujo de agua hacia el molino de barras también presenta una relación inversa con el porcentaje de sólidos en molino de barras (**Figura 15**). Esto tiene sentido físico, pero no operacional, ya que, a mayor porcentaje de

sólidos, hay menos agua en la mezcla, pero al tener un alto porcentaje de sólido se debe aumentar el caudal de agua para lograr el porcentaje de sólido adecuado, indicando que la física de dilución predomina sobre la lógica de control. En la **sección 14 (Figura 15 (b))** existe una correlación positiva notable ($r = 0.38$) entre el porcentaje de sólidos en el molino de barras y el porcentaje de sólidos que va a flotación, la cual esta correlación está presente en las tres secciones evaluadas y siendo una correlación lógica operacionalmente, debido a que verse reflejado un aumento en el porcentaje de sólidos al principio del circuito, esta alza se arrastrara hasta la salida y otra relación directa es entre el rendimiento de la planta y el flujo de agua al molino de barras, esto indica una conexión operativa clave, pues a mayores tasas de alimentación se requiere más agua para sostener el proceso.

En las **secciones 13 y 15 (Figuras 15 (a) y (c))** el mapa de calor muestra una mayor heterogeneidad de relaciones. Muchas de las asociaciones observadas en la sección 14 se mantienen, pero con distinta intensidad. Por ejemplo, la sección 13 no presenta una correlación significativa entre el rendimiento (tph) y el flujo de agua al molino de barras ($r \approx 0.08$); esta debilidad en la relación se atribuye a una restricción de suministro hídrico en esa sección, confirmada por los operadores (juicio experto).

La **sección 15** presenta las asociaciones más coherentes desde el punto de vista operacional: el rendimiento se correlaciona positivamente con la velocidad de la bomba, con los flujos de agua hacia las barras y hacia el cajón, con la densidad y con la presión de la batería de hidrociclones. En cuanto a variables de calidad, la fracción gruesa (+65 μm) muestra relación positiva con el rendimiento ($r = 0.28$), mientras que la potencia del molino de bolas tiende a relacionarse de forma inversa con la malla +65 ($r = -0.29$), la presión de los hidrociclones y la densidad.

En conjunto, la evidencia empírica y la disponibilidad de datos indican que **la sección 15** es la candidata más adecuada para desarrollar el modelo predictivo del flujo mineral (tph): no presenta restricciones de agua como la sección 13, además cuenta con un mayor volumen de registros válidos y mantiene relaciones operacionales consistentes. La sección 14, con escasa representación y pocas correlaciones significativas, es menos apropiada como base de entrenamiento sin un tratamiento adicional.

6.2.5 Desfases temporales (Lags)

En el **Anexo J** (Tablas 18-20) se presentan a detalle cada uno de los desfases operacionales de cada sección respecto del rendimiento. En la mayoría de los casos, los *lags* obtenidos son **negativos**, lo que indica que las variables operacionales se modifican **después** de que cambia el

flujo mineral. Desde el punto de vista del modelamiento predictivo esto no es utilizable, ya que el objetivo es anticipar el rendimiento a partir de las variables de entrada y no al revés.

Sin embargo, se han identificado variables específicas que presentan desfases positivos, las cuales actúan como verdaderos predictores al anteceder los cambios en el rendimiento. En la **Tabla 8** presenta un resumen de estas variables para las 3 secciones evaluadas.

Tabla 8. Resumen de variables con desfases positivos (capacidad predictiva).

<i>Variable</i>	<i>Sección 13 [min]</i>	<i>Sección 14 [min]</i>	<i>Sección 15 [min]</i>
<i>Work Index</i>	25	-	3
<i>Granulometría</i>	-	19	2
<i>Flujo agua cajón</i>	15	-	10
<i>Presión BHC</i>	-	25	-
<i>Sólidos RM</i>	-	25	-
<i>Potencia BM</i>	-	-	6
<i>Potencia RM</i>	24	-	14
<i>Flujo pulpa a BHC</i>	-	23	-
<i>Nivel de cajón</i>	25	-	-

En la sección 15 (**Tabla 20**) se observan desfases que resultan muy coherentes desde el punto de vista operacional, aunque también son mayoritariamente negativos. Por ejemplo, ante un cambio en el rendimiento, el flujo de agua hacia el molino de barras presenta un desfase de -1 minuto, es decir, el ajuste de agua ocurre un minuto después del cambio en el flujo mineral y además posee un coeficiente de Pearson de 0.412. De forma similar, se aprecia un cambio en el porcentaje de sólidos en el molino de barras alrededor de los 9 minutos, y un cambio en el porcentaje de sólidos hacia flotación aproximadamente a los 10 minutos. Esto se relaciona con lo observado en el diagrama de correlaciones de la **Figura 15 (c)**, donde la única correlación directa importante del porcentaje de sólidos hacia flotación es precisamente con el porcentaje de sólidos en el molino de barras. En otras palabras, primero se modifica la concentración de sólidos en el molino de barras y, más tarde, se refleja en el porcentaje de sólidos enviado a flotación.

Algo similar ocurre con el bloque formado por densidad hacia la batería de hidrociclones, flujo de pulpa y presión en la batería. Los cambios en estas tres variables se observan aproximadamente 9, 7 y 7 minutos después de la variación en el rendimiento, lo que indica que están fuertemente relacionadas entre sí, en línea con la lógica operacional del circuito. Considerando únicamente los **desfases positivos** de esta sección, es decir, aquellos compatibles con un uso predictivo, se

identifican cinco variables (Tabla 8), la cual es el flujo de agua hacia el cajón, la potencia del molino de barras, potencia del molino de bolas, el WI y la granulometría, con *lags* de 10, 14, 6, 3 y 2 minutos, respectivamente, la cual conforman una ventana de anticipación técnica coherente con la física del proceso.

En la sección 14 (Tabla 19) se observan desfases con tiempos más largos, alcanzando incluso el límite máximo de evaluación de 25 minutos. Aun así, se mantiene cierta coherencia operacional, especialmente en el caso del flujo de agua hacia el molino de barras, donde un cambio en el rendimiento se refleja en el flujo de agua con un desfase de negativo de 1 minuto y un coeficiente de Pearson cercano a 0.422, lo que indica una relación importante. Sin embargo, al comparar estos resultados con los de la sección 15, si bien los desfases también son coherentes desde el punto de vista operativo, los coeficientes de correlación tienden a ser mucho más bajos, lo que los vuelve poco representativos para efectos de modelamiento. Un ejemplo claro es el WI, que presenta un *lag* positivo de 3 minutos en la **Tabla 19**, lo que en teoría permitiría anticipar el rendimiento con esta variable; no obstante, debido a que su correlación es solo de 0.036, el sistema lo deja automáticamente con *lag* 0 al no superar el umbral mínimo establecido de $|r| \geq 0.05$. Por esta razón, la variable no se considera un predictor robusto y no se ve reflejada en la **Tabla 8**.

Por su parte, la sección 13 (Tabla 18) también presenta un desfase de -1 minuto en el flujo de agua hacia el molino de barras, pero asociado a una correlación baja ($r \approx 0.071$). Este comportamiento era esperable debido a la **restricción de agua** comentada previamente en el apartado **6.2.4 Evaluación de correlaciones operacionales**, lo que limita la variación de esta variable. En general, la sección 13 muestra un patrón similar al de la sección 15, con correlaciones comparables, aunque con respuestas más rápidas en algunas variables al cambio del rendimiento. Por ejemplo, la presión hacia la batería de hidrociclones y el flujo de pulpa hacia la BHC presentan desfases de aproximadamente 5 y 4 minutos, respectivamente, posteriores al cambio en el rendimiento, mientras que el porcentaje de sólidos y el porcentaje de sólidos hacia flotación cambian alrededor de los 10 minutos, manteniendo entre ellos una correlación positiva relevante. Al observar la **Tabla 8**, en esta sección se identifican cuatro variables con *lag* positivo, el nivel de cajón, WI, potencia del molino de barras y flujo de agua al cajón. No obstante, en el caso del flujo de agua al cajón, la correlación es baja ($r \approx 0.067$), y el *Work Index* presenta un desfase positivo de 25 minutos, lo que corresponde al límite máximo considerado en el análisis y resulta un tiempo excesivo para tratarlo como una variable de entrada útil en un modelo predictivo.

6.2.6 Comparación modelos

En esta etapa se compararon los distintos modelos evaluados para estimar el rendimiento del circuito.

6.2.6.1 OLS con selección de variables mediante Lasso

En cada sección se entrenó un modelo OLS con el total de variables disponibles, excluyendo parámetros de calidad como porcentaje de sólidos a flotación y Malla +65, y, en paralelo, un modelo OLS reducido según la selección automática de variables realizada por **Lasso** (ecuación 10).

La **Tabla 9** resume los valores del parámetro de penalización α obtenidos mediante Lasso, tanto el que minimiza directamente el error cuadrático medio de validación cruzada (MSE mínimo), como el valor seleccionado bajo la **regla 1-SE**, que escoge el α más grande cuyo MSE se encuentra dentro del intervalo de la ecuación 17.

$$(MSE \text{ mínimo}) + (1 * desviación \text{ estándar}) \quad (17)$$

Al utilizar esto el modelo se ajusta casi igual de bien, pero con menos variables y más robustez, evitando el sobreajuste del modelo. Por otro lado, en el **Anexo K** se presentan las variables escogidas por Lasso de cada una de las secciones de la planta.

Tabla 9. Resumen del entreno modelo OLS.

Sección	$\alpha(\text{mínimo MSE})$	$\alpha(1 - SE)$
13	0.1060	0.5490
14	0.1894	1.3110
15	0.0366	0.6661

Al analizar los resultados por sección, se observa lo siguiente:

Para la sección 13, Lasso selecciona 9 variables, la cual es un número elevado de variables, y por tanto, con un riesgo alto de sobreajuste. Sin embargo, los valores de $\alpha(\text{mínimo MSE})$ y $\alpha(1 - SE)$ se mantienen en un rango razonable, indicando un balance adecuado entre ajuste y complejidad. Aun así, tener demasiadas variables implica una menor interpretabilidad y una mayor sensibilidad frente a cambios operacionales. Por otra parte, los valores de los parámetros α para la sección 14 son bastantes altos, señalando que el algoritmo necesita una penalización más fuerte para evitar un modelo inestable. Esto puede indicar que las variables del conjunto presentan correlaciones más débiles o más ruido operacional, lo cual hace menos confiable la selección automática.

Finalmente, para sección 15, Lasso selecciona 6 variables, un numero adecuado para mantener un equilibrio entre simplicidad, interpretabilidad y capacidad predictiva, además es la sección con el comportamiento más favorable desde el punto de vista predictivo, ya que, evidencia el valor más pequeño de α (*mínimo MSE*) entre todas las secciones, lo que refleja que el ajuste del modelo se logra sin necesidad de una penalización intensa. Estos resultados sugieren que la sección 14 posee una estructura operacional más consistente y con relaciones más claras entre variables, lo que facilita obtener un modelo más estable y robusto.

A continuación, en las **Tablas 10 y 11** se presenta el desempeño del modelo OLS para cada sección de la planta, reportando las principales métricas de evaluación tanto en el conjunto de entrenamiento como en el conjunto de prueba.

Tabla 10. Resumen del entreno modelo OLS.

<i>Sección</i>	<i>R²</i>	<i>MAE</i>	<i>RMSE</i>	<i>%MAPE</i>
13	0.7621	10.29	14.99	2.28
14	0.7296	10.52	15.97	2.26
15	0.8050	9.385	14.77	2.07

Tabla 11. Resumen de la prueba modelo OLS.

<i>Sección</i>	<i>R²</i>	<i>MAE</i>	<i>RMSE</i>	<i>MAPE%</i>
13	0.4254	18.79	25.11	8.08
14	0.6791	11.85	17.25	5.80
15	0.7820	10.77	15.50	2.50

Al analizar en conjunto los resultados del entrenamiento y de la prueba, se observan diferencias importantes en el desempeño del modelo entre las tres secciones evaluadas. Estas diferencias se manifiestan tanto en la capacidad predictiva como en la estabilidad del ajuste frente a los datos conocidos.

En primer lugar, durante el entrenamiento (Tabla 10), la **sección 15** presenta el mejor comportamiento durante el entrenamiento. Su R^2 es el más alto entre las secciones, teniendo un valor de 0.8050, lo que indica que una mayor proporción de la variabilidad del rendimiento logra ser explicada por las variables operacionales consideradas. Además, es la sección con los errores más bajos, tanto en MAE como en RMSE, destacando un RMSE de 14.77 tph, el cual, recordado, este representa el error cuadrático medio del modelo penalizando de forma más severa las desviaciones

grandes. En procesos industriales, este indicador es especialmente relevante porque permite identificar si existen errores puntuales significativos que puedan tener impacto operacional.

La **sección 13** muestra un desempeño similar al de la sección 15, manteniendo un R^2 sobre 0.75 y un RMSE de 14.99. Esto sugiere que, en esta etapa, el modelo logra capturar razonablemente la dinámica de la sección, aunque con una ligera menor precisión en comparación con la sección 15. Por otra parte, la **sección 14** evidencia el ajuste más débil durante el entrenamiento con un $R^2 = 0.7296$, esto indica que el modelo tiene la menor capacidad para representar la variabilidad del rendimiento.

En cuanto a la otra métrica de desempeño **%MAPE**, las tres secciones presentan valores similares, todos inferiores al 2.5% en el entrenamiento. Dado que, en general un MAPE menor al 5% suele considerarse un desempeño “**adecuado**” en procesos industriales, estos resultados sugieren que, en términos relativos, el modelo mantiene un comportamiento consistente al medir el error porcentual promedio.

Al analizar el desempeño en prueba (**Tabla 11**), se observa que la única sección donde el modelo OLS mantiene resultados satisfactorios con datos no utilizados en el entrenamiento es la sección 15, con un $R^2 = 0.7820$, levemente inferior al de entrenamiento, lo que es esperable al trabajar con datos no conocidos. Las métricas de error aumentan de forma moderada y el %MAPE alcanza un 2.5 %, lo que sigue siendo muy bueno para un contexto industrial. En contraste, la sección 13 muestra una caída importante de desempeño ($R^2 = 0.4254$ y $RMSE = 25.11$ tph), evidenciando problemas de generalización del modelo en esta sección. La sección 14 posee un valor de $R^2 = 0.6791$, con errores de magnitud comparable; sin embargo, su capacidad explicativa sigue siendo limitada, debido a que arroja un valor menor a 0.7 en el ajuste mediante R^2 , por lo que el modelo OLS no resulta del todo recomendable para esta sección. Finalmente, en las **Figuras 16,17 y 18** se presentan las curvas del rendimiento real versus el predicho por el modelo, graficando únicamente **800 registros** de cada caso con el fin de facilitar la visualización dada la gran cantidad de registros disponibles.

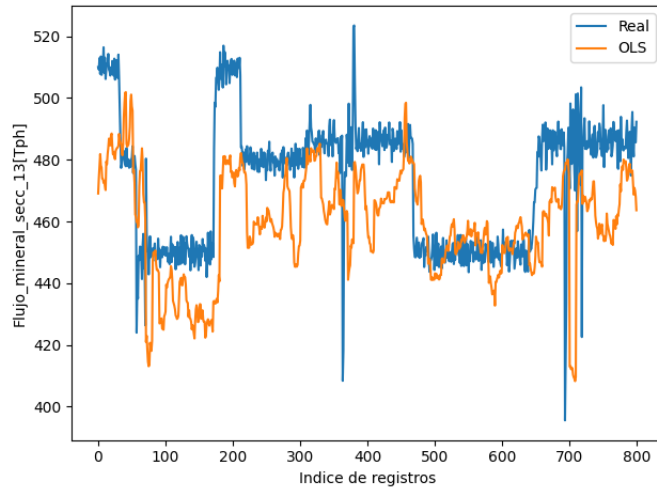


Figura 16. Serie temporal del rendimiento real vs predicho por el modelo OLS, sección 13.

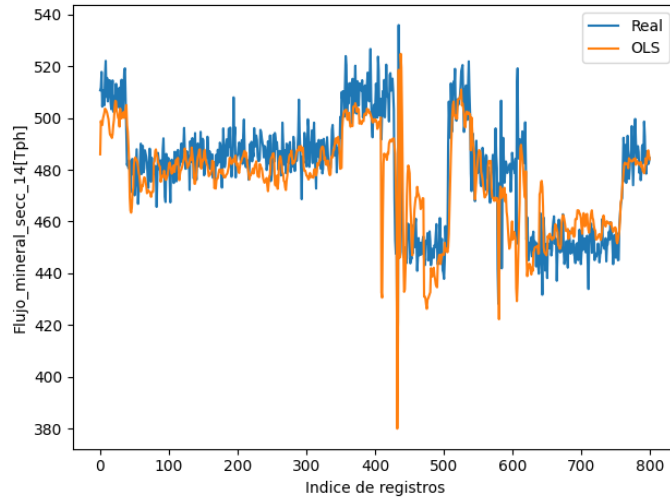


Figura 17. Serie temporal del rendimiento real vs predicho por el modelo OLS, sección 14.

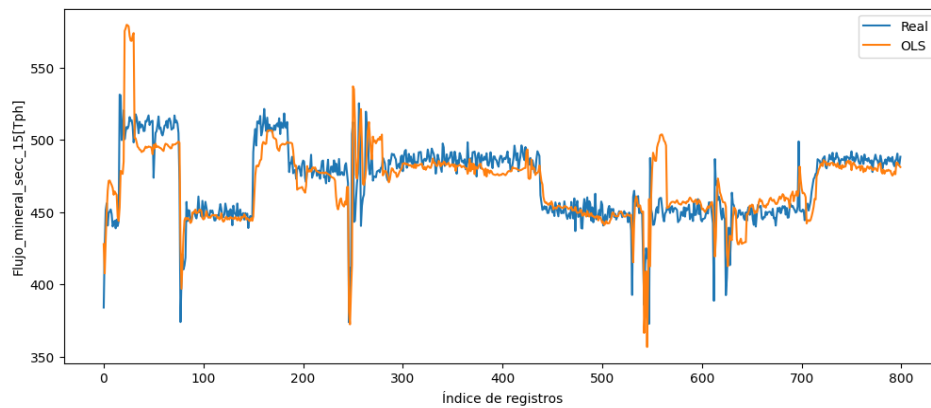


Figura 18. Serie temporal del rendimiento real vs predicho por el modelo OLS, sección 15.

Debido que la **sección 15** es la única que presenta un desempeño predictivo adecuado en el modelo OLS, tanto en entrenamiento como en prueba, se decidió incluir exclusivamente la ecuación final de esta sección, la cual se evidencia en la ecuación 18.

$$\begin{aligned} \text{rendimiento} = & 465.656 + 48.980x_{\% \text{Sólidos RM}} + 2.552x_{\text{Flujo pulpa a BHC}} + 2.576x_{\text{Densidad a BHC}} + \\ & 56.047x_{\text{Flujo agua a RM}} - 0.856x_{\text{Potencia BM}} + 1.421x_{\% \text{Vel Bomba alimentación}} - 0.641x_{\% \text{Granulometría lag2}} \end{aligned} \quad (18)$$

Al observar la ecuación 18, obtenida mediante la regresión lineal con selección de variables mediante Lasso, se evidencia una dependencia crítica hacia dos variables, la cual es el porcentaje de sólidos y el flujo de agua hacia el molino de barras. Al contrastar estos resultados con el diagrama de correlaciones de la sección 15 (**Figura 15 (c)**), se observa un coeficiente de Pearson de -0.86 entre ambas variables, lo que indica que la física de dilución predomina sobre la lógica de control del circuito.

Esta correlación sugiere que el agua se ajusta de forma manual, permitiendo que el porcentaje de sólidos varíe libremente como respuesta directa a la adición de agua. A pesar de que Lasso tiene como objetivo seleccionar predictores con baja redundancia, la inclusión simultánea de estas dos variables muy correlacionadas entre sí provoca un sobreajuste que compromete la representatividad del modelo. En términos prácticos, el modelo termina representado el rendimiento como un fenómeno de transporte hidráulico, ignorando variables fundamentales de la conminución. Esto se confirma al observar que el *Work Index* fue excluido como predictor, dejando a la granulometría como el único indicador de las características del mineral entrante, lo que simplifica excesivamente la complejidad metalúrgica del proceso en la sección 15.

6.2.6.2 Random Forest

A diferencia del modelo lineal OLS, el algoritmo de Random Forest no requiere estrictamente un proceso previo de selección de variables, dado que es un método no lineal y no paramétrico que utiliza subconjuntos aleatorios de predictores en cada partición. Sin embargo, para efectos de este modelamiento, se decidió excluir las variables con alta correlación entre sí. Esta medida se adoptó con el fin de evitar el sobreajuste (*overfitting*) y prevenir la dilución de la importancia de los parámetros, asegurando que la jerarquización de variables sea técnicamente interpretable y representativa del proceso. Al igual que en los casos anteriores, se mantuvieron excluidas las variables de calidad mencionadas previamente. A continuación, en las Tablas 12 y 13, se presentan los resultados del entrenamiento y prueba, reportando las métricas de desempeño correspondientes.

Tabla 12. Resumen del entreno modelo Random Forest.

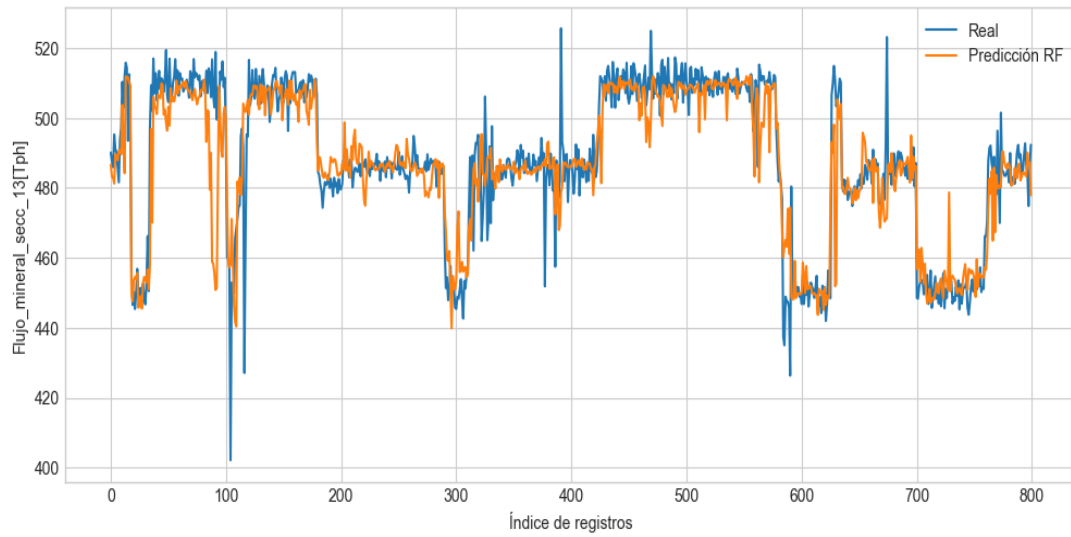
<i>Sección</i>	<i>R²</i>	<i>MAE</i>	<i>RMSE</i>	<i>MAPE%</i>
13	0.9839	1.355	4.021	0.39
14	0.9781	2.582	4.534	0.55
15	0.9823	2.122	4.582	0.47

Tabla 13. Resumen de la prueba modelo Random Forest.

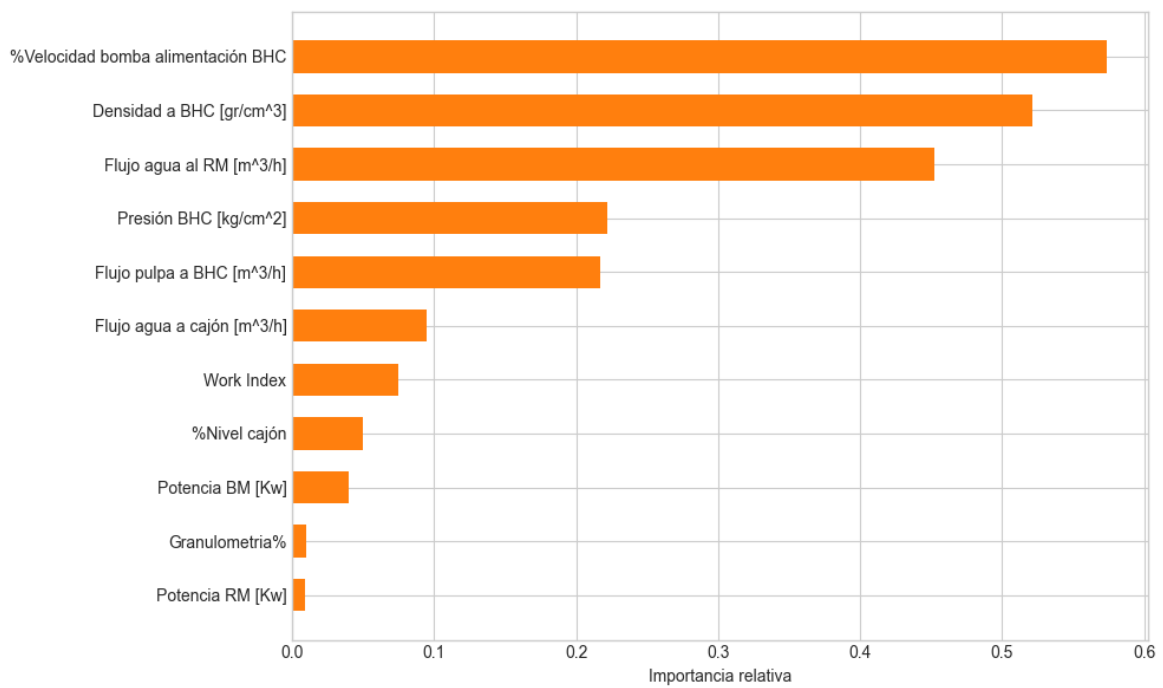
<i>Sección</i>	<i>R²</i>	<i>MAE</i>	<i>RMSE</i>	<i>MAPE%</i>
13	0.9361	3.886	6.934	1.52
14	0.8692	8.964	10.73	6.87
15	0.9056	5.729	8.523	3.52

Durante el entrenamiento (Tabla 12), el modelo muestra valores extremadamente altos de ajuste en todas las secciones, con **R² superiores a 0.97**, MAE y RMSE muy bajos y un MAPE inferior al 1%. Esto indica que el modelo es capaz de aprender muy bien los patrones presentes con datos conocidos, sin embargo, en primera instancia este rendimiento tan elevado del modelo es característico de un fenómeno de **sobreajuste**, especialmente común en modelos basados en arboles cuando se trabaja con datos ruidosos o muy variables.

Al evaluar el modelo con datos no conocidos (Tabla 13), se evidencia una caída en la capacidad predictiva, pero es algo común, debido a que se evalúan datos no conocidos. En general las 3 secciones arrojan buenas métricas en el entrenamiento, la sección 14 posee el R² más bajo de las tres secciones con un valor de 0.8692, pero sigue siendo un valor excelente como aproximación del modelo Random Forest al rendimiento de la planta. En las **Figuras 19,20 y 21** se evidencian las curvas del rendimiento predicho por el modelo Random Forest y la importancia relativa de cada una de las variables en el modelo.

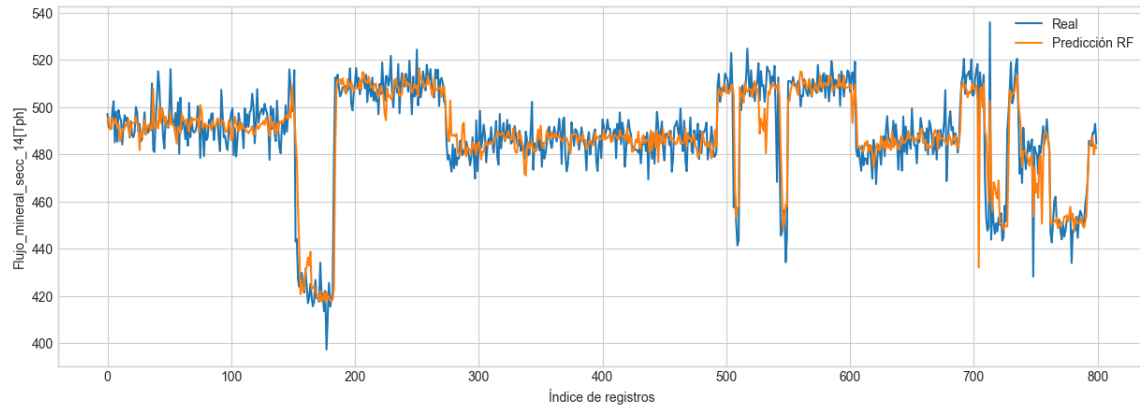


(a)

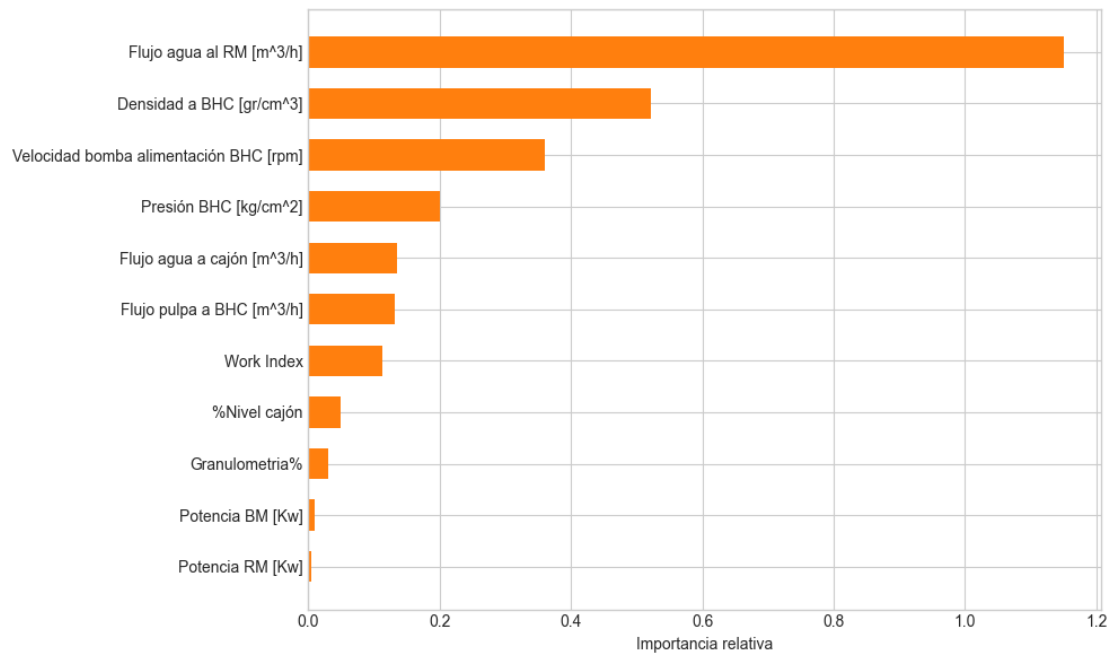


(b)

Figura 19. Desempeño modelo predictivo en la sección 13. a) Serie temporal del rendimiento real y el predicho por el Modelo RF. (b) Importancia relativa de las variables operacionales en el modelo RF.

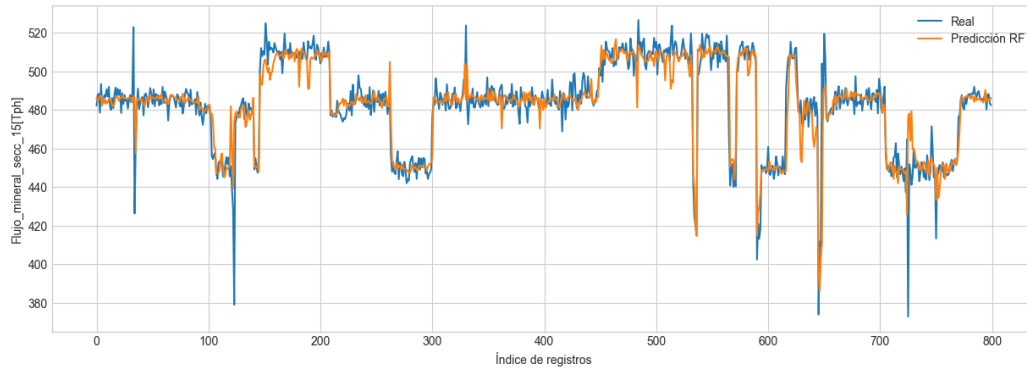


(a)

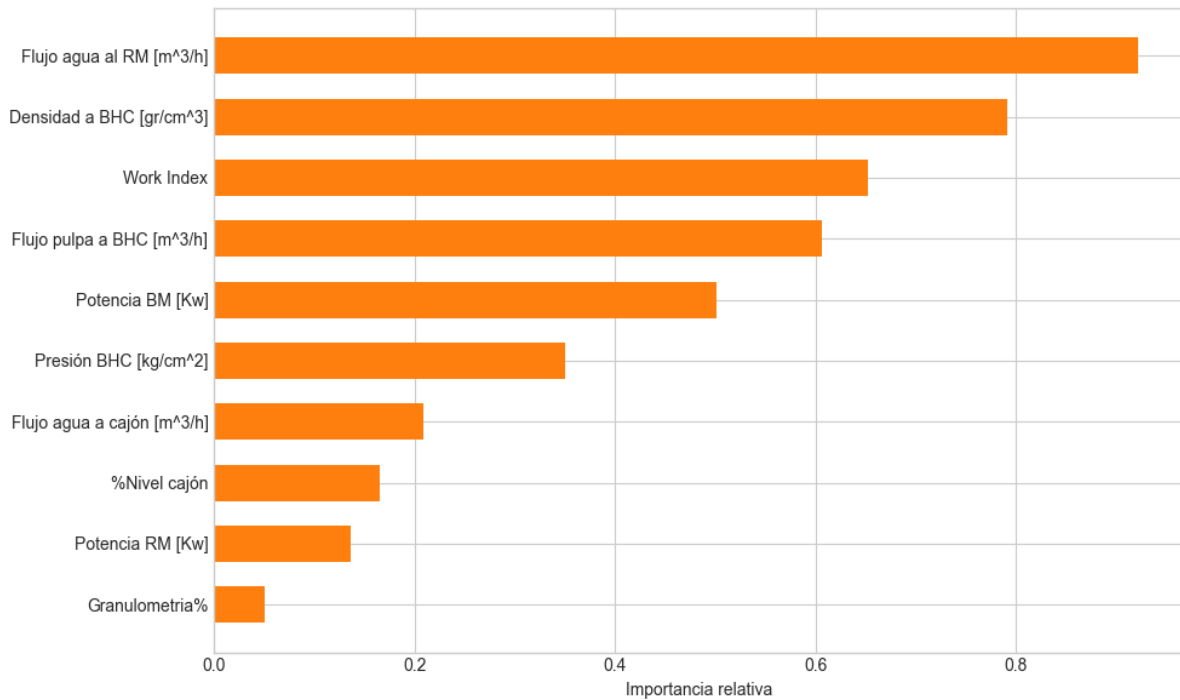


(b)

Figura 20. Desempeño modelo predictivo en la sección 14. a) Serie temporal del rendimiento real y el predicho por el Modelo RF. (b) Importancia relativa de las variables operacionales en el modelo RF.



(a)



(b)

Figura 21. Desempeño modelo predictivo en la sección 15. a) Serie temporal del rendimiento real y el predicho por el Modelo RF. (b) Importancia relativa de las variables operacionales en el modelo RF.

Al analizar el desempeño del modelo en la sección 13 (**Figura 19 (a)**), se observa un ajuste preciso respecto a los datos de rendimientos reales de la planta, presentado únicamente ciertas dificultades menores en predecir picos muy bajos o altos de rendimiento. Al observar la **Figura 19 (b)**, que detalla la relevancia de cada parámetro en el modelo RF, se reitera que la naturaleza no lineal y el fundamento basado en múltiples arboles de decisión de este algoritmo impiden la obtención de una

ecuación algebraica explícita. No obstante, este método permite cuantificar y visualizar la importancia relativa de las variables evaluadas mediante gráfico de barras.

En esta sección se identifica que la variable más importante para la predicción es la velocidad de la bomba de alimentación hacia la batería de hidrociclones, con un valor de importancia aproximado del 0.55. Este resultado sugiere que el modelo se fundamenta principalmente en el transporte mecánico del flujo. Sin embargo, esto no significa que el rendimiento de la planta se explique simplemente porque la bomba gire más rápido, más bien, el modelo utiliza esta variable como un atajo estadístico, ya que la bomba simplemente reacciona al volumen de pulpa que recibe. Como segunda variable de importancia se encuentra la densidad, la cual complementa el análisis anterior. Al observar el diagrama de correlación (Figura 15 (a)), se nota que la densidad y la velocidad de la bomba presentan una relación directamente proporcional de 0.35, por ende, al estar correlacionadas y destacar como principales variables, el modelo RF, tiende a “arrastrar” estas decisiones, dado que ambos parámetros son directamente proporcionales al flujo mineral (rendimiento). Además, al analizar los *lag* de estas dos variables (Tabla 18) respecto al rendimiento, se comprueba que sus cambios ocurren después de que lo haga el rendimiento, por lo tanto, se clasifican como variables no predictoras para el modelo.

Respecto a las potencias de los molinos, se evidencia que su importancia en el modelo es casi nula. Este fenómeno encuentra una explicación metalúrgica al observar las correlaciones de la **Figura 15 (a)**, donde existe una correlación negativa de -0.41 entre la densidad y la potencia del molino de bolas, siendo la correlación más alta registrada entre estas dos variables entre todas las secciones evaluadas. Este comportamiento sugiere la ocurrencia del efecto “colchon”, el cual se produce cuando existe una densidad alta en el molino provocando una pérdida de potencia en estos, entonces al tratarse de un comportamiento errático, el modelo RF lo clasifica como “ruido” y opta por descartarlo en su proceso de decisión. En consecuencia, al ignorar la conminución mineral, el modelo se convierte principalmente en una representación del transporte de cañerías. Esta conclusión se confirma al notar variables de entrada críticas como el WI y la granulometría, presentan una importancia extremadamente baja en el modelo, con valores de importancia menores a 0.1.

En la **Figura 20 (b)**, correspondiente a la sección 14, se evidencia un comportamiento significativamente distinto al observado en la sección 13. En este caso la variable con mayor relevancia es la fujo del agua hacia el molino de barras, con un valor de importancia superior a 1, esta reacción era esperada dado que, en el diagrama de correlaciones (Figura 15 (b)), esta sección no presenta asociaciones de alto impacto entre variables operativas entre sí, ni respecto al

rendimiento. Ante esta falta de relaciones claras, el modelo identifica el flujo de agua como la única variable que “mueve” el rendimiento.

Inmediatamente, como segunda variable de importancia es la densidad, con un valor aproximado de 0.52. La diferencia extrema entre la primera y segunda posición sugiere una alta dependencia del algoritmo hacia el control hídrico del sistema. Además, debido a la baja correlación entre los predictores y la variable objetivo, no es posible representar de manera precisa el rendimiento en esta sección, limitando la capacidad del modelo para capturar eventos complejos.

Al igual que en la sección 13, se observa que las variables de entrada críticas y las potencias de los molinos se sitúan al final de la **Figura 20 (b)**, con una muy baja importancia. Este patrón reafirma que el modelo RF, al no encontrar lógica fuerte entre la conminución y el rendimiento, opta por ignorar la metalurgia del proceso.

Finalmente, al evaluar la sección 15 (**Figura 21**) se evidencia una superioridad del modelo en comparación con las otras secciones, reflejada en una distribución de importancia de variables mucho más balanceada y representativa (**Figura 21 (b)**).

Como variable principal, al igual que en la sección previa, se sitúa el flujo de agua al molino de barras, aunque con un impacto relativo menor de 0.92 aproximadamente. Desde una perspectiva operacional, esta relevancia es esperada, ya que el incremento de tratamiento mineral exige necesariamente un aumento en el flujo de agua, lo cual se valida con las correlaciones obtenida en los diagramas de Pearson (**Figura 15 (c)**). En segunda instancia, como variable clave tenemos la densidad con un valor de 0.79 de importancia relativa. Este hallazgo confirma que el modelo integra correctamente la relación masa-volumen, ya que a mayor carga mineral la densidad del circuito tiende a aumentar. Además, en el análisis de desfases (*lags*), la densidad refleja variaciones a los 9 minutos después de los cambios del rendimiento, actuando como un valioso indicador de la estabilidad del flujo.

Un punto diferenciador en esta sección es la aparición del *Work Index* en el tercer puesto de importancia (0.65). Es la única sección donde la dureza del mineral adquiere un peso significativo dentro del modelo, lo que otorga una alta validez técnica y representatividad operativa. Luego, en un nivel medio de importancia, se identifican variables como el flujo de pulpa hacia BHC, potencia BM y presión BHC. El flujo de pulpa identifica claramente el contorno operativo de la sección y siendo fundamental para explicar la carga circulante del sistema, mientras que la potencia del BM refleja el trabajo mecánico de molienda secundaria, el hecho de que la energía consumida posea este peso indica que el modelo reconoce el consumo energético como un factor determinante para sostener el flujo mineral. Por su parte, la presión en el BHC, que presenta una correlación de 0.25 respecto al

rendimiento (**Figura 15 (c)**) y con una importancia relativa de 0.37, esta actúa como indicador sustituto (*proxy*) ante la ausencia de variables de calidad directa, tales como el porcentaje de malla +65 y porcentaje de sólidos hacia flotación.

Las variables con menor peso se clasifican como reactiva o de soporte del sistema. El flujo de agua hacia el cajón y el nivel de cajón, debido a su alta correlación con la densidad con un valor de 0.55, cumplen funciones de compensación de carga para estabilizar el bombeo, por lo que su variabilidad no es un controlador primario del rendimiento, y el modelo logra identificarlo correctamente. Asimismo, se observa que la operación del molino de barras es significativamente más estable que el molino de bolas, por ello, su potencia no destaca como un factor determinante, posicionando a la molienda secundaria (BM) como el verdadero indicador operacional para regular y predecir el rendimiento.

Finalmente, aunque la sección 15 logra mejor consistencia operativa, su capacidad predictiva se considera de corto plazo. Esto se explica porque, si bien el modelo identifica variables con desfases positivos como la potencia BM (6 minutos), el *Work Index* (3 min) y el flujo de agua al cajón (10 min), los parámetros con mayor importancia relativa en el algoritmo, el flujo de agua al molino de barras y la densidad, poseen un comportamiento reactivo (desfases negativos) en base al rendimiento. En consecuencia, el modelo cumple primordialmente una función explicativa del rendimiento con una predicción acotada (corto plazo).

7 CONCLUSIONES Y RECOMENDACIONES

7.1 Conclusiones

El análisis comparativo realizado entre los modelos permitió establecer diferencias claras en su capacidad predictiva y la coherencia metalúrgica del sistema en las tres secciones de la planta de molienda convencional. En primer lugar, el modelo OLS no resultó adecuado para este proceso. Aunque durante el entrenamiento alcanzó valores de R^2 aceptables, en la prueba su desempeño cayó drásticamente, especialmente en las secciones 13 y 14. Esta discrepancia evidencia un sobreajuste severo, por lo que el modelo no logra generalizar y queda descartado para fines operacionales. Por el contrario, en la sección 15, el modelo OLS mostró un desempeño significativamente más estable con un R^2 superior 0.75 y los errores más bajos del estudio (MAE, RMSE Y %MAPE), sin embargo, su validez operativa es limitada, ya que la selección automática de variables (Lasso) aplicadas para este modelo priorizó parámetros altamente correlacionados entre sí, como el flujo de agua y el porcentaje de sólidos en el molino de barras, reduciendo la ecuación a un balance hídrico simplificado. En la práctica el modelo depende casi exclusivamente de estas dos variables, ignorando la complejidad metalúrgica de la molienda.

El modelo Random Forest demostró una superioridad estadística clara en las tres secciones, con coeficientes de R^2 superiores a 95% en entrenamiento y cercanos al 90% en las pruebas. No obstante, el análisis de importancia de variables obligo a descartar las secciones 13 y 14 por falta de sustento metalúrgico, en el caso de la sección 13, el algoritmo se fundamentó excesivamente en el transporte mecánico (velocidad de bomba con importancia de 0.55), ignorando variables críticas como la potencia de los molinos y el *Work Index*, por otra parte la sección 14 a pesar de obtener los errores más bajos, el rendimiento dependió de forma casi exclusiva del flujo de agua hacia el molino de barras (importancia de 1.15). Debido a que el diagrama de calor no mostró correlaciones claras en esta línea, el modelo utilizó el agua como un indicador sustituto. Esto confirma que el algoritmo se limitó a seguir la respuesta hídrica del sistema ante la falta de patrones metalúrgicos claros que permitieran una predicción basada en la conminución, además de que esta sección al igual que la anterior ignora el *Work Index*.

La sección 15 destaca por alcanzar la mayor coherencia metalúrgica y operativa en los tres casos de estudio. En este escenario, el modelo no solo integra métricas de alta precisión, sino que logra integrar variables de conminución de forma lógica. Un hecho relevante es la validación de los desfases temporales (*lags*) positivos, que permiten una ventana predicción real basada tanto en la física del proceso como en la dinámica del flujo: el *Work Index* con 3 minutos desfase, la potencia del molino de bolas con 6 minutos, la granulometría con 2 minutos y el flujo de agua a cajón un

desfase de 10 minutos. Estos parámetros permiten anticipar a corto plazo el comportamiento del rendimiento, a diferencia de variables reactivas como el flujo de agua al molino de barras o la densidad, cuyos *lags* eran negativo debido a que responden en base a la condición actual del sistema.

La calidad de la información fue un factor crítico en estos resultados. Tras aplicar filtros robustos, la sección 15 conservó aproximadamente un 70 % de los registros, mientras que las secciones 13 y 14 redujeron su información válida a 49.61 % y 62.64 %, respectivamente. Esta mayor disponibilidad de datos válidos en la línea 15 facilitó un entrenamiento más profundo y una mejor captura de variabilidad del proceso.

Finalmente, se establece que el modelo desarrollado para la sección 15 posee una capacidad predictiva de corto plazo, sustentada en la validación de variables con desfases positivo que anteceden los cambios en el rendimiento. No obstante, debido a que los parámetros de mayor peso estadístico en el modelo operan de forma reactiva, el sistema cumple fundamentalmente una función explicativa y de diagnóstico. Al cuantificar con precisión cuanto contribuye cada variable al rendimiento en régimen operativo, el modelo permite identificar condiciones asociadas a aumentos o caídas de tonelaje. Esto lo convierte en un complemento útil al sistema de control PROFIT, ya que, facilita interpretar el efecto de las acciones de control y priorizar intervenciones operativas basadas en evidencia.

7.2 Recomendaciones

- 1) Realizar análisis de sensibilidad del modelo de la sección 15, con el fin de evaluar como varía el rendimiento ante cambios controlados en las variables operacionales seleccionadas. Esto permitiría comprender mejor la magnitud del aporte de cada predictor.
- 2) Replicar el estudio en otras secciones de la planta cuando se disponga de datos operacionales de mejor calidad, especialmente en las secciones 13 y 14, que actualmente presentan mayor dispersión y menor confiabilidad de registros. En estos casos, además de los modelos lineales, se recomienda evaluar enfoques híbridos (p. ej., componente lineal + componente no lineal), de modo de capturar relaciones aproximadamente lineales y, a la vez, comportamientos no lineales asociados a cambios de régimen operacional.

8 BIBLIOGRAFIA

1. CODELCO. (s.f.). *Historia*. Obtenido de CODELCO:
<https://www.codelco.com/nosotros/la-corporacion/historia>
2. CODELCO. (2023). *Memoria Anual 2023 CODELCO*. Santiago, Chile: CODELCO.
3. Logística, A. (2 de Mayo de 2025). Obtenido de Codelco informó que producción de cobre propia en 2024 alcanzó las 1.441.886 toneladas:
<https://www.agendalogistica.cl/actualidad-cobre-codelco/codelco-informo-que-produccion-de-cobre-propia-en-2024-alcanzo-las-1441886-toneladas/1931680>
4. CODELCO. (s.f.). *División Chuquicamata*. Obtenido de CODELCO:
<https://www.codelco.com/division-chuquicamata>
5. Observatory, N. E. (4 de enero de 2024). Obtenido de Copper mining at Chuquicamata:
<https://earthobservatory.nasa.gov/images/152368/copper-mining-at-chuquicamata>
6. Aguilar Aguilera, J. (2008). *Caracterización geotécnica y estructural de la rampa de exploración y del túnel de drenaje, Mina Chuquicamata*. Universidad de Chile.
7. Ballard, J., Palin, J., Williams, I., Campbell, I., & Faunes, A. (2001). Two ages of porphyry intrusion resolved for the supergiant Chuquicamata copper deposit of northern Chile by ELA-ICP-MS and SHRIMP. *Geology*, 29(5), 383-386.
8. Faunes, A., Hintze, F., Siña, A., Véliz, H., & Vivanco, M. (2005). Chuquicamata, core of a planetary scale Cu–Mo anomaly. En T. Porter, *Super Porphyry Copper & Gold Deposits: A Global Perspective* (Vol. 1, págs. 151-174). PGC Publishing.
9. Acevedo, H., Busto, P., & Hernández, D. (2021). *Curso avanzado de molienda SAG: Gerencia concentradora, División Chuquicamata, CODELCO Chile*. CODELCO.
10. SIP. (2025). *Datos operacionales de procesos*. Calama, Chile: Superintendencia de Ingeniería de Procesos, División Chuquicamata, CODELCO.
11. Álvares Rodríguez, B. (2010). *Análisis de la influencia de los modelos de distribución de tamaños de partículas en la determinación de consumos energéticos en molienda mediante el método de Bond*. Universidad de Oviedo.
12. Blanco, E. A. (2013). Capítulo 8. Molienda. Universidad de Cantabria.
13. Gutiérrez, L., & Sepúlveda, J. (1986). *Dimensionamiento y optimización de plantas concentradoras mediante técnicas de modelación matemática*. Centro de investigación Minera y Metalúrgica.

14. Universidad Politécnica de Cartagena. (s.f.). *Tecnología mineralúrgica: Tema 5: Reducción de tamaño – molienda. Molienda II [Presentación en PDF]*. Obtenido de OpenCourseWare UPCT:
https://ocw.bib.upct.es/pluginfile.php/14070/mod_resource/content/2/Molienda%20%28Parte%20II%292.pdf
15. Austin, L. G., & Concha, F. A. (1994). *Diseño y simulación de circuitos de molienda y clasificación*. CYTED: Programa Iberoamericano de Ciencia y Tecnología para el desarrollo.
16. Goldfarb Casas, D. (2021). *Introducción al machine learning: Automatización e inteligencia artificial [Presentación en PowerPoint]*. Pontificia Universidad Católica de Chile.
17. Figueroa Matamala, F. A. (2024). *Modelación de restricciones de la molienda primaria para optimizar el tratamiento de la planta concentradora en Minera Los Pelambres*. Universidad de Chile, Facultad de Ciencias Físicas y Matemáticas, Departamento de Ingeniería Industrial.
18. Wooldridge, J. (2016). *Introductory Econometrics: A Modern Approach* (6th ed. ed.). Cengage Learning.
19. Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1), 267-288.
20. Bishop, C. (2006). *Pattern Recognition and Machine Learning*. Springer. Obtenido de <https://www.microsoft.com/en-us/research/wp-content/uploads/2006/01/Bishop-Pattern-Recognition-and-Machine-Learning-2006.pdf>
21. Niquini, F. G. F., Branches, A. M. B., Leite Costa, J. F. C., Moreira, G. C., Schneider, C. L., Araújo, F. C., & Capponi, L. N. (2023). Recursive feature elimination and neural networks applied to the forecast of mass and metallurgical recoveries in a Brazilian phosphate mine. *Minerals*, 13(748).
22. Herrera, N., Sinche Gonzalez, M., Okkonen, J., & Mollehuara, R. (2023). Soft computing application in mining, mineral processing and metallurgy with an approach to using it in mineral waste disposal. *Minerals*, 13(1450).
23. Tukey, J. (1977). *Exploratory Data Analysis*. Addison-Wesley.

9 ANEXOS

9.1 Anexo A. PI ProcessBook

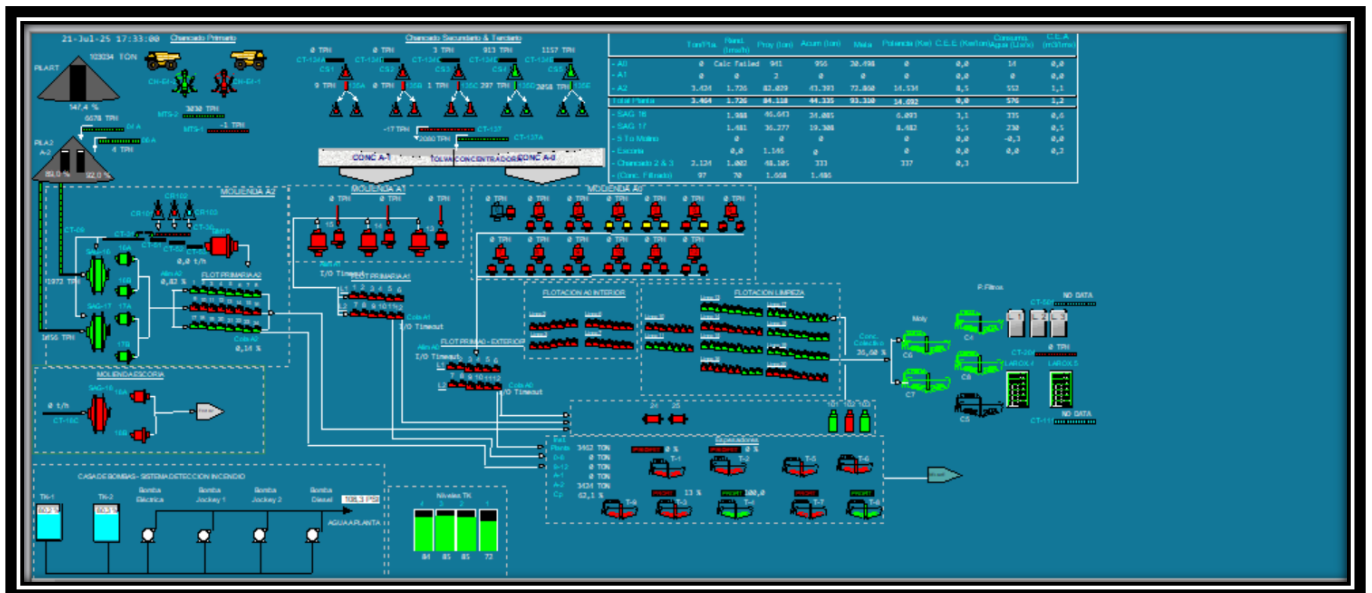


Figura 22. Diagrama general de la planta concentradora entregada por PI ProcessBook.

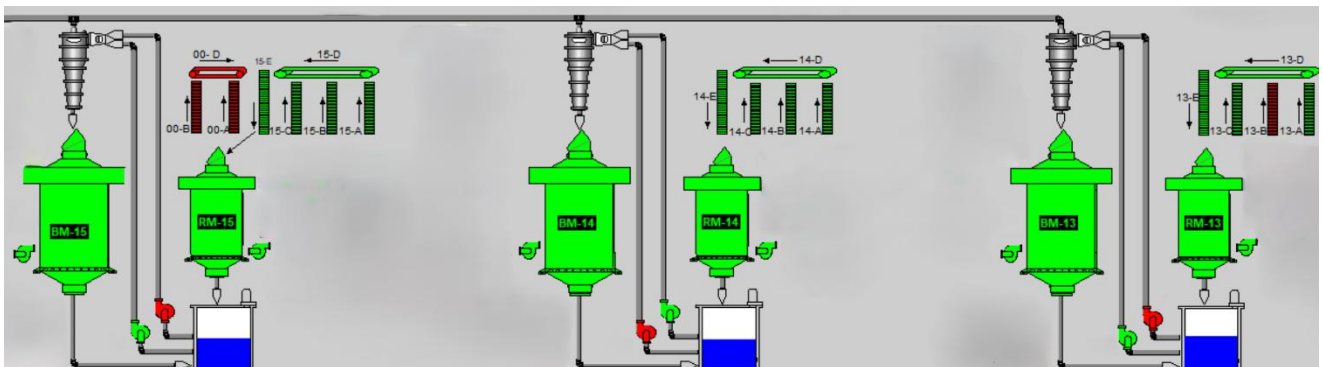


Figura 23. Diagrama de la molienda convencional "A1" entregada por PI ProcessBook.

En las Figuras evidenciadas en el **Anexo A** (Figura 22 y 23), se presenta el diagrama general de la concentradora, junto con el área de interés del análisis que es la molienda convencional "A1", por razones de confiabilidad no se presenta los parámetros operacionales en tiempo real, los cuales son proporcionados por PI ProcessBook.

9.2 Anexo B. PI DataLink

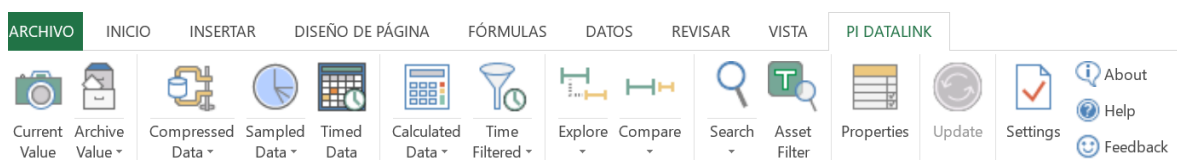


Figura 24. Interfaz de la extracción de datos de PI DataLink.

En la **Figura 24** se exhiben las diversas funciones disponibles en PI-DataLink. Este software permite la incorporación de datos operativos directamente en hojas de cálculo. Dentro del alcance de esta memoria de título, se emplearon las funciones clave de Calculated Data, diseñada para calcular y extraer los valores totales del parámetro especificado; Time Filtered, utilizada para obtener datos en intervalos temporales válidos, es decir, dentro de expresiones temporales pertinentes que representen fechas y horas específicas.

9.3 Anexo C. Tag de variables

Tabla 14. Ejemplo de Tag utilizados para la extracción de datos.

Tag Utilizados	
Variable	Tag
<i>Presión BHC [kg/cm²]</i>	<i>MA1_XPI1312</i>
<i>% Velocidad bomba alimentación BHC</i>	<i>MA1_XSY1302_OP</i>
<i>Flujo pulpa a BHC [m³/h]</i>	<i>MA1_XFI1310</i>
<i>% solido a flotación</i>	<i>MA1_XPS1313</i>
<i>Densidad a BHC [gr/cm³]</i>	<i>MA1_XDC1305</i>
<i>% Malla +65</i>	<i>MA1_XAI1313V</i>
<i>Flujo agua a Cajón [m³/h]</i>	<i>MA1_XFC1308</i>
<i>Flujo mineral [tph]</i>	<i>MA1_XWC1304</i>
<i>Potencia m. bolas [kW]</i>	<i>MA1_XJI1318</i>
<i>Potencia motor-bomba BHC norte [kW]</i>	<i>MA1_7JI1302</i>
<i>Velocidad motor-bomba BHC norte [rpm]</i>	<i>MA1_7SI1302</i>
<i>Velocidad motor-bomba BHC sur [rpm]</i>	<i>MA1_7SI1301</i>
<i>% Nivel cajón</i>	<i>MA1_XLC1309</i>
<i>Flujo agua a RM [m³/h]</i>	<i>MA1_XFC1306</i>
<i>% solido m. barra</i>	<i>MA1_XPSC1325</i>
<i>Potencia m. barras [kW]</i>	<i>MA1_XJI1317</i>

En la **Tabla 14** se evidencia ejemplos de los tags correspondientes que se utilizaron para una sección específica de la planta. Cabe mencionar que cada una de las variables presentadas en la Tabla fueron extraídas para las 3 secciones correspondiente de la planta. Cada uno de los tags fue empleado individualmente, es decir, no se aplicó ningún filtro, ya que, no era necesario, puesto que los datos extraídos corresponden al promedio por minuto.

9.4 Anexo D. Extracción de datos de las variables operacionales

The screenshot shows a software interface for data extraction. The main window displays a spreadsheet with columns for time intervals and various tags. A sidebar on the right, titled 'Calculated Data', contains fields for 'Data item', 'Root path (optional)', 'Data item(s)', 'Start time', 'End time', 'Time interval (optional)', and 'Filter expression (optional)'. The 'Data item' field is set to 'Data item', and the 'Root path' is set to '\\cnchpi02'. The 'Data item(s)' field is set to 'Hoja1'!\$H\$3. The 'Start time' field is set to 'Hoja1'!\$B\$2. The 'End time' field is set to 'Hoja1'!\$B\$3. The 'Time interval' field is set to 'Hoja1'!\$B\$4. The 'Filter expression' field is empty.

Figura 25. Recopilación de datos de las variables operacionales de la sección “A1”.

En la **Figura 25**, se observa la manera de extraer los datos, la cual se debe indicar el “Root”, la casilla del tag correspondiente además de la hora de inicio y final, y el intervalo de tiempo deseado.

9.5 Anexo E. Informe de detenciones “Base de datos molienda”

PLANTA	SECCIÓN	EQUIPO	ACONTECIMIENTOS	CAUSALIDAD	Inicio	Fin	OPER	MEC	ELEC	ELECT	Pérdida TMR	Parada*	Sem	Día	Mes*	Año	Factor**	Hrs. Eq.
A-1	Sección-15	RM-15	Carguo de barras.	Carguo de Barras	13:38:01	13:49:31	0:11:30					94		22	1 JUNIO	2024	1	0.19
A-1	Sección-15	RM-15	Atollo chute de alimentación.	Elemento Extraño	9:54:16	9:59:16						41		22	1 JUNIO	2024	1	0.08
A-1	Sección-15	RM-15	Carguo de barras.	Carguo de Barras	13:10:30	13:21:31	0:11:01					90		23	3 JUNIO	2024	1	0.18
A-1	Sección-15	RM-15	Atollo en chute alimentación	Elemento Extraño	5:16:16	5:30:01						113		23	3 JUNIO	2024	1	0.23
A-1	Sección-15	RM-15	Atollo chute de alimentación.	Elemento Extraño	8:12:16	8:24:16						98		23	3 JUNIO	2024	1	0.20
A-1	Sección-15	RM-15	Atollo chute alimentación	Elemento Extraño	23:34:46	23:41:30						55		23	3 JUNIO	2024	1	0.11
A-1	Sección-15	RM-15	Limpieza de parrillas de descarga	Limpieza de Circuito	4:27:00	4:39:15	0:12:15					100		23	3 JUNIO	2024	1	0.20
A-1	Sección-15	RM-15	Mantenimiento general.	Mantenimiento Programada	7:55:46	5:00:00		21:04:14				10352		23	4 JUNIO	2024	1	21.07
A-1	Sección-15	RM-15	Mantenimiento general.	Mantenimiento Programada	5:00:00	5:00:00		0:00:00				11791	NO	23	5 JUNIO	2024	1	24.00
A-1	Sección-15	RM-15	Mantenimiento general.	Mantenimiento Programada	5:00:00	1:14:00		20:14:00				10413		23	6 JUNIO	2024	1	20.23
A-1	Sección-15	RM-15	Bajo flujo descanso de alimentación.	Ajuste de Terreno	3:34:16	3:40:46	0:06:30					56		23	6 JUNIO	2024	1	0.11
A-1	Sección-15	CT-15E	calibración completa	Mantenimiento Programada	9:16:46	9:23:16				0:06:30		59		23	7 JUNIO	2024	1	0.11
A-1	Sección-15	RM-15	Carguo de barras	Carguo de Barras	10:36:01	10:46:01	0:10:00					90		23	7 JUNIO	2024	1	0.17
A-1	Sección-15	RM-15	Carguo de barras	Carguo de Barras	10:29:01	10:45:16	0:16:15					136		23	9 JUNIO	2024	1	0.27
A-1	Sección-15	RM-15	Carguo de barras	Carguo de Barras	9:45:16	9:58:46	0:13:30					113		24	11 JUNIO	2024	1	0.23
A-1	Sección-15	RM-15	Bajo flujo de agua sello bba norte.	Ajuste de Terreno	21:45:15	21:49:30	0:04:15					36		24	12 JUNIO	2024	1	0.07
A-1	Sección-15	RM-15	Carguo de barras.	Carguo de Barras	17:59:01	18:08:31	0:09:30					79		24	13 JUNIO	2024	1	0.16
A-1	Sección-15	RM-15	Carguo de barras	Carguo de Barras	8:42:30	8:50:30	0:08:00					67		24	13 JUNIO	2024	1	0.13
A-1	Sección-15	RM-15	Pérdida de aire embrague	Embrague Dañado	9:03:16	9:09:16		0:06:00				50		24	13 JUNIO	2024	1	0.10
A-1	Sección-15	BM-15	Filtración de aire por rotorseal	Línea Aire Embrague Rota	11:00:30	12:00:00		0:59:30				497		24	13 JUNIO	2024	1	0.99
A-1	Sección-15	RM-15	Atollo chute de alimentación.	Elemento Extraño	21:16:16	21:32:15						134		24	14 JUNIO	2024	1	0.27
A-1	Sección-15	RM-15	Carguo de barras.	Carguo de Barras	17:54:45	18:06:00	0:11:15					94		24	15 JUNIO	2024	1	0.19
A-1	Sección-15	RM-15	Cortocircuito en cable alimentación agil	Configuración errónea de las pro	8:28:15	14:50:15			6:22:00			3,192		25	17 JUNIO	2024	1	6.37
A-1	Sección-15	BM-15	Sin integración de flujo a la cuba, se rei	Otro Instrumentación	14:50:15	15:09:30				0:19:15		161		25	17 JUNIO	2024	1	0.32
A-1	Sección-15	RM-15	Carguo de barras	Carguo de Barras	10:45:31	10:55:01	0:09:30					79		25	18 JUNIO	2024	1	0.16
A-1	Sección-15	RM-15	Se abre interruptor principal en maniob	Otro Eléctrico	10:19:31	10:43:31			0:24:00			201		25	18 JUNIO	2024	1	0.40
A-1	Sección-15	RM-15	Mto CT140-141-142	Falta de Mineral (CH23)	12:12:16	5:00:00						8,420		25	18 JUNIO	2024	1	16.80
A-1	Sección-15	RM-15	Mto CT140-141-142	Falta de Mineral (CH23)	5:00:00	8:24:00						1,704		25	19 JUNIO	2024	1	3.40
A-1	Sección-15	RM-15	Respiete de pernos sueltos en cilindro.	Perno Suelto	17:36:00	18:03:00		0:27:00				226		25	19 JUNIO	2024	1	0.45
A-1	Sección-15	RM-15	Carguo de barras.	Carguo de Barras	13:59:30	14:16:45	0:17:15					144		25	20 JUNIO	2024	1	0.29
A-1	Sección-15	RM-15	Carguo de barras.	Carguo de Barras	13:18:30	13:29:30	0:11:00					92		25	22 JUNIO	2024	1	0.18
A-1	Sección-15	RM-15	Carguo de barras.	Carguo de Barras	12:27:00	12:30:30	0:03:30					29		26	24 JUNIO	2024	1	0.06
A-1	Sección-15	RM-15	falla estructural espesador T-5	Detención por Disp. Espesadores	20:24:30	5:00:00						4,307		26	25 JUNIO	2024	1	8.59

Figura 26. Detenciones mensuales correspondiente a la planta “A1”.

9.6 Anexo F. Extracción de datos del Work Index

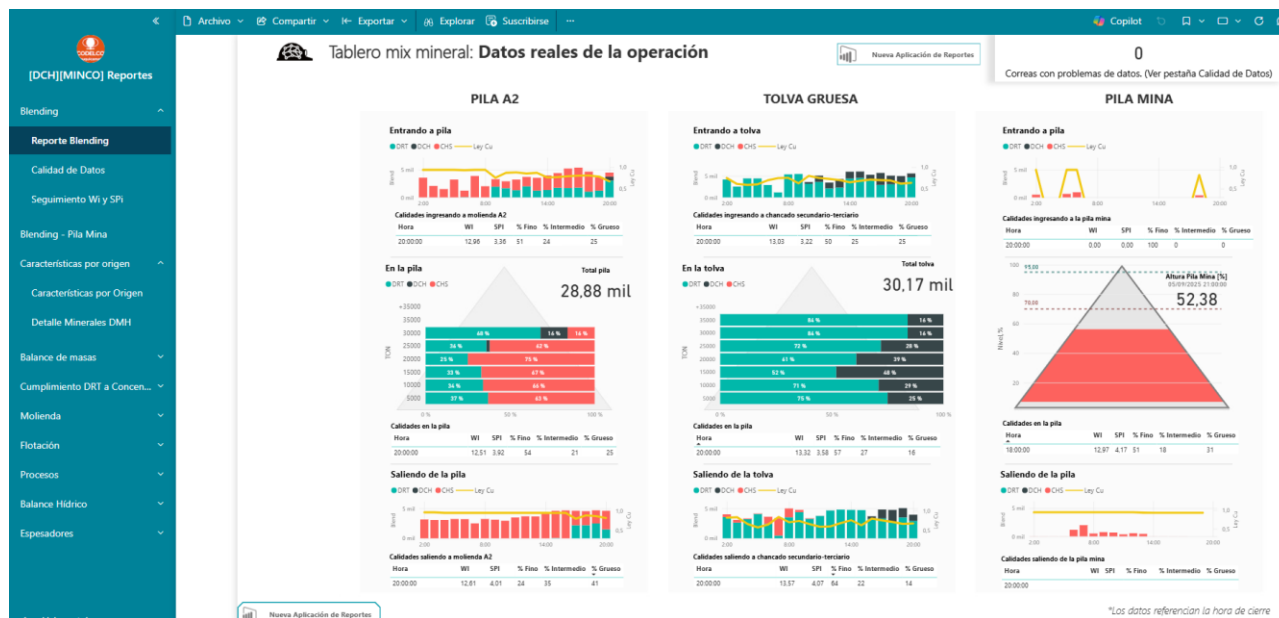


Figura 27. Dashboards de la planta concentradora.



Figura 28. Dashboard correspondiente al Work Index que entra a la tolva gruesa.

En la **Figura 27** se muestran los registros de distinta información de la empresa y en la **Figura 28** se evidencia los registros del *Work Index* que entra a la tolva gruesa que extraje directamente del *dashboard* de planta, correspondientes a la serie temporal con resolución de una hora durante el año 2024.

9.7 Anexo G. Distribución mineral diaria del año 2024

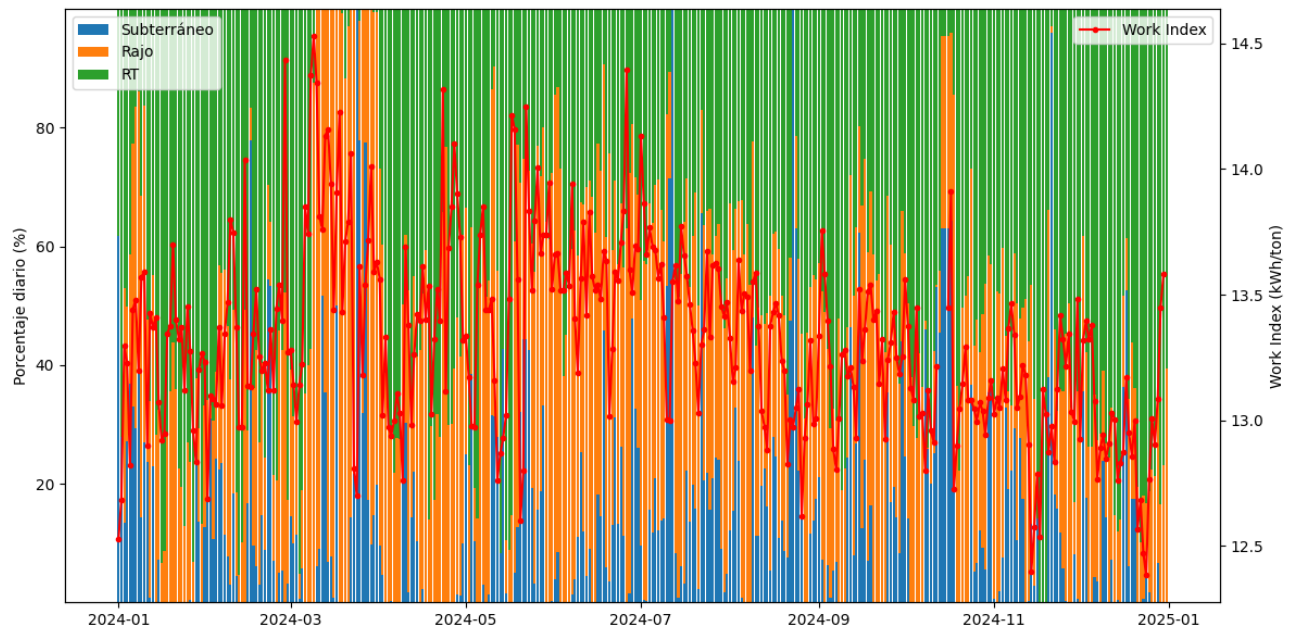


Figura 29. Distribución diaria de mineral en la molienda convencional.

9.8 Anexo H. Tabla resumen de *Outliers* al aplicar filtro IQR.

9.8.1 Variables sección 13

Tabla 15. Límites y Outliers de la sección 13 (MC).

Variables	Limite bajo	Limite alto	Outliers
<i>Potencia BM [Kw]</i>	2100.36	2552.65	5271
<i>Rendimiento [tph]</i>	0.00	976.65	0
<i>% Solidos RM</i>	68.82	98.90	11240
<i>Potencia Rm [Kw]</i>	745.05	1420.95	17045
<i>% Malla + 65</i>	11.85	48.84	94013
<i>Presión BHC [kg/cm²]</i>	0.29	0.75	20484
<i>% Solidos a flotación</i>	36.02	56.30	76769
<i>% Velocidad bomba</i>	63.00	83.80	16678
<i>Flujo pulpa a BHC [m³/h]</i>	0.00	3739.17	0
<i>Densidad a BHC [gr/cm³]</i>	1.32	1.71	123680
<i>Potencia motor bomba [Kw]</i>	84.89	269.55	7458
<i>Velocidad motor bomba [rpm]</i>	200.81	263.74	1831
<i>% Nivel cajón</i>	24.35	79.40	45367
<i>Flujo agua a cajón[m³/h]</i>	189.34	626.61	82223
<i>Flujo agua a RM[m³/h]</i>	0.00	212.24	0

9.8.2 Variables sección 14

Tabla 16. Límites y Outliers de la sección 14 (MC).

Variables	Limite bajo	Limite alto	Outliers
<i>Potencia BM [Kw]</i>	2084.38	2536.67	9802
<i>Rendimiento [tph]</i>	377.23	557.93	29557
<i>% Solidos RM</i>	70.61	84.97	20960
<i>Potencia Rm [Kw]</i>	670.42	1474.56	14636
<i>% Malla + 65</i>	12.03	47.56	50053
<i>% Solidos a flotación</i>	34.67	53.71	61067
<i>Presión BHC [kg/cm²]</i>	0.04	0.85	11138
<i>Densidad a BHC [gr/cm³]</i>	1.45	1.66	174202
<i>Flujo pulpa a BHC [m³/h]</i>	0.00	5245.17	0
<i>Velocidad motor bomba [rpm]</i>	208.48	284.41	29830
<i>Flujo agua a cajón[m³/h]</i>	152.54	681.01	73964
<i>Flujo agua a RM[m³/h]</i>	30.63	211.89	91519
<i>% Nivel cajón</i>	25.33	85.36	83860

9.8.3 Variables sección 15

Tabla 17. Límites y Outliers de la sección 15 (MC).

Variables	Limite bajo	Limite alto	Outliers
<i>Potencia BM [Kw]</i>	2064.60	2516.36	9747
<i>Rendimiento [tph]</i>	340.79	567.93	8285
<i>% Sólidos RM</i>	70.31	88.39	11952
<i>Potencia Rm [Kw]</i>	820.01	1339.95	15001
<i>% Malla + 65</i>	10.29	44.46	51253
<i>% Sólidos a flotación</i>	33.97	57.81	40961
<i>Presión BHC [kg/cm²]</i>	0.11	0.84	10201
<i>Flujo pulpa a BHC [m³/h]</i>	328.84	3185.28	91372
<i>Densidad a BHC [gr/cm³]</i>	1.31	2.11	38114
<i>Flujo agua a cajón[m³/h]</i>	170.98	623.20	28838
<i>% Nivel cajón</i>	22.82	82.83	64375
<i>Flujo agua a RM[m³/h]</i>	6.75	209.98	109390
<i>% Velocidad bomba</i>	64.1	80.1	15241

En las **Tablas 15, 16 y 17** se detallan los límites calculados por el filtro IQR para cada variable operacional, junto con el número de *Outliers* detectados en las tres secciones de la molienda convencional. Aquellas variables resaltadas en rojo indican límites que se alejan de la realidad operativa registrada en planta y, por tanto, requieren un ajuste o verificación adicional antes de ser consideradas válidas.

9.9 Anexo I. Sistema de control PROFIT

MODO CONTROLADOR PROFIT ON				ACCION CONTROLADOR PROFIT OPTIMIZANDO			
MANIPULADAS				CONTROLADAS			
TONELAJE	AGUA POZO	VELOCIDAD BOMBA	PRESION BHC	NIVEL CAJON	MALLA +65	POTENCIA BM	DENSIDAD

Figura 30. Interfaz del controlador experto PROFIT de la planta molienda convencional.

La **Figura 30** presenta la pantalla del controlador PROFIT que opera en la planta convencional. En ella se pueden ver, qué señales el regulador modifica (variables manipuladas) y cuáles vigila para mantener la planta estable (variables controladas). En la práctica esto significa que PROFIT ajusta parámetros como el caudal de agua o la velocidad de bombas para mantener variables críticas, tales como, presión en la batería de hidrociclones, nivel del cajón, potencia de molino, calidad de producto, etc. dentro de los objetivos operacionales. Por motivos de confiabilidad y seguridad no se muestran en la imagen los límites numéricos ni las alarmas asociadas.

9.10 Anexo J. Tabla resumen de los desfases temporales

Tabla 18. Desfases temporales en base al rendimiento de la sección 13 de la planta MC.

Variables	Lag [min]	Mejor correlación	Mejor Lag [min]
<i>Potencia BM [Kw]</i>	-7	-0.170	0
<i>% Sólidos RM</i>	-10	0.288	0
<i>Potencia Rm [Kw]</i>	24	0.104	24
<i>% Malla + 65</i>	-21	0.206	0
<i>Presión BHC [kg/cm²]</i>	-5	0.280	0
<i>% Sólidos a flotación</i>	-10	0.513	0
<i>% Velocidad bomba</i>	-5	0.507	0
<i>Flujo pulpa a BHC [m³/h]</i>	-4	0.251	0
<i>Densidad a BHC [gr/cm³]</i>	-10	0.364	0
<i>Potencia motor bomba [Kw]</i>	-3	0.568	0
<i>Velocidad motor bomba [rpm]</i>	-9	0.506	0
<i>% Nivel cajón</i>	25	-0.136	25
<i>Flujo agua a cajón [m³/h]</i>	15	0.067	15
<i>Flujo agua a RM [m³/h]</i>	-1	0.071	0
<i>Work index</i>	25	-0.116	25
<i>% Granulometría</i>	8	-0.034	0

Tabla 19. Desfases temporales en base al rendimiento de la sección 14 de la planta MC.

Variables	Lag [min]	Mejor correlación	Mejor Lag [min]
<i>Potencia BM [Kw]</i>	-9	-0.088	0
<i>% Solidos RM</i>	25	-0.065	25
<i>Potencia Rm [Kw]</i>	23	0.048	0
<i>% Malla + 65</i>	-12	0.160	0
<i>% Solidos a flotación</i>	-10	0.306	0
<i>Presión BHC [kg/cm²]</i>	25	-0.054	25
<i>Densidad a BHC [gr/cm³]</i>	-25	0.042	0
<i>Flujo pulpa a BHC [m³/h]</i>	23	0.054	23
<i>Velocidad motor bomba [rpm]</i>	25	-0.024	0
<i>Flujo agua a cajón[m³/h]</i>	25	0.021	0
<i>Flujo agua a RM[m³/h]</i>	-1	0.422	0
<i>% Nivel cajón</i>	-20	0.119	0
<i>Work index</i>	3	0.036	0
<i>% Granulometria</i>	19	-0.086	19

Tabla 20. Desfases temporales en base al rendimiento de la sección 15 de la planta MC.

Variables	Lag [min]	Mejor correlación	Mejor Lag [min]
<i>Potencia BM [Kw]</i>	6	-0.152	6
<i>% Sólidos RM</i>	-9	0.050	0
<i>Potencia Rm [Kw]</i>	14	0.113	14
<i>% Malla + 65</i>	-10	0.318	0
<i>% Sólidos a flotación</i>	-10	0.167	0
<i>Presión BHC [kg/cm²]</i>	-7	0.283	0
<i>Flujo pulpa a BHC [m³/h]</i>	-7	0.385	0
<i>Densidad a BHC [gr/cm³]</i>	-9	0.507	0
<i>Flujo agua a cajón[m³/h]</i>	10	0.339	10
<i>% Nivel cajón</i>	-21	0.183	0
<i>Flujo agua a RM[m³/h]</i>	-1	0.412	0
<i>% Velocidad bomba</i>	-5	0.331	0
<i>Work index</i>	3	0.236	3
<i>% Granulometría</i>	2	-0.104	2

En las **Tablas 18, 19 y 20** se presentan los *lags* operacionales de cada sección respecto del rendimiento. La primera columna indica el desfase evaluado (minutos) y la segunda muestra la correlación de Pearson asociada a ese desfase. La última columna, “Mejor *lag*”, recoge el desfase que cumple simultáneamente los dos criterios definidos en Metodología: causalidad (*lag* > 0) y fuerza de asociación ($r \geq 0.05$). Si alguna variable no satisface ambas condiciones, se asigna 0 como lag recomendado para el modelo.

9.11 Anexo K. Variables operacionales escogidas por Lasso

Tabla 21. Variables operacionales seleccionadas mediante Lasso de cada una de las secciones.

Sección 13	Sección 14	Sección 15
<i>Flujo agua a RM</i> [m^3/h]	<i>Flujo agua a cajón</i> [m^3/h]	<i>Flujo agua a RM</i> [m^3/h]
% <i>Sólidos RM</i>	% <i>Sólidos RM</i>	% <i>Sólidos RM</i>
<i>Potencia motor bomba</i> [Kw]	<i>Work index</i>	<i>Flujo pulpa a BHC</i> [m^3/h]
<i>Potencia Rm</i> [Kw]	% <i>Nivel cajón</i>	% <i>Velocidad bomba</i>
<i>Presión BHC</i> [kg/cm^2]	<i>Potencia Rm</i> [Kw]	<i>Densidad a BHC</i> [gr/cm^3]
<i>Densidad a BHC</i> [gr/cm^3]	% <i>Granulometría</i>	<i>Potencia BM</i> [Kw]
<i>Velocidad motor bomba</i> [rpm]		% <i>Granulometría</i>
% <i>Granulometría</i>		
<i>Potencia BM</i> [Kw]		

UNIVERSIDAD DE CONCEPCIÓN – FACULTAD DE INGENIERÍA
Departamento de Ingeniería Metalúrgica
Hoja Resumen Memoria de Título

Título: EVALUACIÓN DE UN MODELO PREDICTIVO DE LA PLANTA DE MOLIENDA CONVENCIONAL "A1"
DIVISIÓN CHUQUICAMATA, CODELCO

Nombre Memorista: José David Valenzuela Alarcón

Modalidad	Investigación	Profesor(es) Guía (s) Prof. Luver Echeverry Vargas
Concepto	Muy Bueno	
Calificación	6.7	
Fecha	19-03-2026	
Prof. Froilán Vergara G.		Ingeniero Supervisor: Lilian Ávila / Sebastián Gallegos
		Institución: CODELCO DCH/ UdeC.

Comisión (Nombre y Firma)

Prof. Fernando Betancourt	Prof. Andrés Ramírez
---------------------------	----------------------

Resumen

La División Chuquicamata procesa mineral por distintas vías de conminución, siendo molienda convencional (plantas "A0" y "A1") y SAG (planta "A2"). El control operativo de la molienda convencional se apoya en el sistema experto PROFIT; sin embargo, la necesidad de contar con herramientas complementarias de diagnóstico y predicción motivó el desarrollo y evaluación de un modelo predictivo del rendimiento (tph) aplicado a la planta A1 (secciones 13, 14 y 15). El objetivo principal fue evaluar la factibilidad de un modelo predictivo útil para la toma de decisiones y cuantificar la influencia de las principales señales operacionales sobre el tonelaje procesado.

Para caracterizar la alimentación se realizó un análisis detallado del *blending mineral* (fuentes de alimentación: rajo, subterráneo y RT), la granulometría respecto al estándar de 12.700 micrones (P80) y el *Work Index* (WI) horario. Se observó una clara relación entre mayor participación del rajo y aumentos en WI; el WI promedio del período analizado fue de 13,3 kWh/t y el 81 % de los registros se concentró en rangos de 12.6 a 14.0 kWh/t. La base de datos utilizada para el modelo está compuesta por variables operacionales registradas por minuto del año 2024 (527040 registros). El preprocesado incluyó: filtro IQR inicial, filtrado manual y un modo "estable". Además, se estudiaron desfases temporales (*lags*) para identificar las variables adelantadas o rezagadas respecto al rendimiento. Tras la limpieza, las filas válidas por sección fueron 49.61%, 62.64% y 69.27% del total de datos para las secciones 13, 14 y 15 respectivamente.

Se implementó una estrategia comparativa entre un modelo lineal (OLS) y no lineal (Random Forest). Las métricas usadas para la evaluación fueron R^2 , MAE, RMSE y MAPE; por requerimiento operacional se estableció $R^2 \geq 0.70$ como umbral mínimo de utilidad práctica. El modelo Random Forest demostró una superioridad estadística significativa, alcanzando valores de R^2 cercanos al 90% en la etapa de prueba para las tres secciones, superando ampliamente al modelo OLS, el cual presentó un sobreajuste severo. A pesar del alto del alto desempeño del modelo Random Forest, el análisis global reveló que solo la sección 15 posee una representatividad operativa integral. En esta sección, el modelo logró integrar coherentemente variables de conminución como el *Work Index* y la potencia del molino de bolas, otorgando al sistema una capacidad predictiva de corto plazo. Por el contrario, las otras dos secciones se descartaron, debido a predicciones casi exclusivamente de variables mecánicas y reactivas como la velocidad de bombas y flujos de agua.

Finalmente, se estable que el modelo desarrollado para la sección 15 cumple una función explicativa y de diagnóstico predictivo, actuando como un complemento estratégico al sistema de control PROFIT.