



Universidad Nacional Mayor de San Marcos

Universidad del Perú. Decana de América

Facultad de Ingeniería Geológica, Minera, Metalúrgica y Geográfica

Escuela Profesional de Ingeniería Metalúrgica

**Optimización del Control Operativo del Circuito de
Molienda-Clasificación usando Machine Learning en la
UM Cerro Corona – Gold Fields La Cima**

TESIS

Para optar el Título Profesional de Ingeniero Metalúrgico

AUTOR

Jesús Junior ALVARADO PEÑA

ASESOR

José Luis VIDARTE MERIZALDE

Lima, Perú

2025



Reconocimiento - No Comercial - Compartir Igual - Sin restricciones adicionales

<https://creativecommons.org/licenses/by-nc-sa/4.0/>

Usted puede distribuir, remezclar, retocar, y crear a partir del documento original de modo no comercial, siempre y cuando se dé crédito al autor del documento y se licencien las nuevas creaciones bajo las mismas condiciones. No se permite aplicar términos legales o medidas tecnológicas que restrinjan legalmente a otros a hacer cualquier cosa que permita esta licencia.

Referencia bibliográfica

Alvarado, J. (2025). *Optimización del Control Operativo del Circuito de Molienda-Clasificación usando Machine Learning en la UM Cerro Corona – Gold Fields La Cima*. [Tesis de pregrado, Universidad Nacional Mayor de San Marcos, Facultad de Ingeniería Geológica, Minera, Metalúrgica y Geográfica, Escuela Profesional de Ingeniería Metalúrgica]. Repositorio institucional Cybertesis UNMSM.

Metadatos complementarios

Datos de autor	
Nombres y apellidos	Jesús Junior Alvarado Peña
Tipo de documento de identidad	DNI
Número de documento de identidad	71336927
URL de ORCID	NO APLICA
Datos de asesor	
Nombres y apellidos	José Luis Vidarte Merizalde
Tipo de documento de identidad	DNI
Número de documento de identidad	07397539
URL de ORCID	https://orcid.org/0000-0002-3353-2551
Datos del jurado	
Presidente del jurado	
Nombres y apellidos	Vidal Sixto Aramburú Rojas
Tipo de documento	DNI
Número de documento de identidad	08608364
Miembro del jurado 1	
Nombres y apellidos	Luis Alfredo Orihuela Salazar
Tipo de documento	DNI
Número de documento de identidad	08004136
Miembro del jurado 2	
Nombres y apellidos	Manuel Prudencio Caballero Rios

Tipo de documento	DNI
Número de documento de identidad	09136425
Datos de investigación	
Línea de investigación	Diseño y aplicación de nuevas tecnologías
Grupo de investigación	No aplica
Agencia de financiamiento	Sin financiamiento
Ubicación geográfica de la investigación	Edificio: Unidad Minera Cerro Corona País: Perú Departamento: Cajamarca Provincia: Hualgayoc Distrito: Hualgayoc Latitud: -6.74406 Longitud: -78.56917
Año o rango de años en que se realizó la investigación	2023 - 2024
URL de disciplinas OCDE	Minería, Procesamiento de minerales https://purl.org/pe-repo/ocde/ford#2.07.05 Ciencias de la computación https://purl.org/pe-repo/ocde/ford#1.02.01



«Año de la recuperación y consolidación de la economía peruana»

**ACTA DE SUSTENTACIÓN DE TESIS PARA OPTAR TÍTULO PROFESIONAL DE
INGENIERO METALÚRGICO**

En el salón de Grados y Títulos de la Escuela Profesional de Ingeniería Metalúrgica de la Facultad de Ingeniería Geológica, Minera, Metalúrgica y Geográfica de la Universidad Nacional Mayor de San Marcos, el miércoles 04, del mes de junio del 2025 siendo las 10:00 horas, en presencia de los Señores Docentes designados como Miembros del Jurado Calificador:

Dr. VIDAL SIXTO, Aramburú Rojas	(PRESIDENTE)
Mg. LUIS ALFREDO, Orihuela Salazar	(MIEMBRO)
Ing. MANUEL PRUDENCIO, Caballero Rios	(MIEMBRO)

Reunidos para el Acto Académico Público de la Sustentación de la TESIS Titulada: «Optimización del Control Operativo del Circuito de Molienda-Clasificación usando Machine Learning en la UM Cerro Corona – Gold Fields La Cima» presentada por Don: JESÚS JUNIOR ALVARADO PEÑA, Bachiller en Ingeniería Metalúrgica, para la obtención del Título Profesional de Ingeniero: **METALÚRGICO**.

Los miembros del Jurado Calificador, escuchada la sustentación respectiva, plantearon al graduando las observaciones pertinentes, que fueron absueltas a:

.....**SATISFACCIÓN**.....

Concluida la Sustentación de Tesis, el Jurado procedió a evaluar y calificar según el reglamento vigente, cuyo resultado fue la nota de:

.....**QUINCE (15) APROBADO**.....

Habiendo sido aprobada la Sustentación de la Tesis por el Jurado Calificador, el Presidente del Jurado, recomienda que la Facultad de Ingeniería Geológica, Minera, Metalúrgica y Geográfica, otorgue el **TÍTULO PROFESIONAL DE INGENIERO METALÚRGICO**, a Don: **Jesús Junior Alvarado Peña**.

Siendo las **11:05** horas, se dió por concluido el acto académico, expidiéndose cinco (5) Actas Originales de la Sustentación de Tesis.

Lima, 4 de junio del 2025


Dr. VIDAL SIXTO ARAMBURÚ ROJAS
Presidente


Mg. LUIS ALFREDO ORIHUELA SALAZAR
Miembro


Ing. MANUEL PRUDENCIO CABALLERO RIOS
Miembro


Ing. JOSÉ LUIS VIDARTE MERIZALDE
ASESOR DE TESIS



Universidad Nacional Mayor de San Marcos
Universidad del Perú. Decana de América
FACULTAD DE INGENIERÍA GEOLÓGICA, MINERA, METALÚRGICA y GEOGRÁFICA
ESCUELA PROFESIONAL DE INGENIERÍA METALÚRGICA

CERTIFICADO DE SIMILITUD

Yo, JOSÉ LUIS VIDARTE MERIZALDE en mi condición de asesor acreditado con la Resolución Decanal N.º000278-2024-D-FIGMMG/UNMSM de la tesis, cuyo título es “Optimización del Control Operativo del Circuito de Molienda-Clasificación usando Machine Learning en la UM Cerro Corona – Gold Fields La Cima” presentado por el bachiller **JESÚS JUNIOR ALVARADO PEÑA** para optar el título profesional de Ingeniero Metalúrgico.

CERTIFICO que se ha cumplido con lo establecido en la Directiva de Originalidad y de Similitud de Trabajos Académicos, de Investigación y Producción Intelectual. Según la revisión, análisis y evaluación mediante la plataforma PIDS, el documento evaluado cuenta con el **14%** de similitud, nivel **PERMITIDO** para continuar con los trámites correspondientes y para su **publicación en el repositorio institucional**.

Se emite el presente certificado en cumplimiento de lo establecido en las normas vigentes, como uno de los requisitos para la obtención del título correspondiente.

Lima, 31 de octubre de 2024

Firma del Asesor _____

Ing. JOSÉ LUIS VIDARTE MERIZALDE

DNI: 07397539

DEDICATORIA

A mi familia, que es y siempre será mi mayor fortaleza. A mis padres, María Peña y Jesús Alvarado, por ser un ejemplo de esfuerzo y dedicación, y por brindarme las herramientas necesarias para alcanzar mis metas. A mi abuela, Herminia Cataño, por su cariño inquebrantable, siempre presente en cada etapa de mi vida. A mis hermanos, Mayra y Harol, por su apoyo constante y sus palabras de aliento en todo momento.

AGRADECIMIENTO

En primer lugar, quiero expresar mi gratitud a Gold Fields La Cima, por permitirme realizar este trabajo y por brindarme el apoyo necesario para su desarrollo. Agradezco especialmente a los compañeros del Centro Integrado de Operaciones, por su orientación y apoyo constante en el proyecto desarrollado. A la Universidad Nacional Mayor de San Marcos, mi más sincero agradecimiento por la formación académica que me proporcionó y por fomentar en mí el interés por la investigación y la excelencia profesional. A mis profesores y compañeros, quienes me acompañaron en este proceso con sus enseñanzas y consejos, les estaré siempre agradecido.

Tabla de contenido

RESUMEN	11
1 CAPITULO I: INTRODUCCIÓN	1
1.1 Introducción	1
1.2 Planteamiento del Problema	3
1.2.1 Descripción de la problemática	3
1.2.2 Formulación del problema	5
1.3 Objetivos de la investigación	5
1.3.1 Objetivo General	5
1.3.2 Objetivos Específicos	5
1.4 Importancia y alcance de la investigación	6
1.5 Limitaciones de la Investigación	8
2 CAPITULO II: MARCO TEÓRICO	9
2.1 Antecedentes de la investigación	9
2.2 Bases teóricas	12
2.2.1 Ciencia de Datos	12
2.2.2 Machine Learning	19
2.2.3 Big Data	27
2.2.4 Integración de Ciencia de Datos, Machine Learning y Big Data	28
2.2.4 PI System	30
2.2.5 Gold Fields	33
2.2.6 Minería del Cobre	44

2.2.7	Mineralogía del Cobre	45
2.2.8	Chancado.....	47
2.2.9	Molienda y Clasificación	48
2.2.10	Flotación de Cobre	50
2.3	Definición de términos.....	53
3	CAPITULO III: HIPÓTESIS Y VARIABLES	61
3.1	La hipótesis.....	61
3.2	Las Variables	61
3.2.1	Variables Independientes	61
3.2.2	Variable Dependiente.....	62
3.3	Operacionalización de variables.....	63
4	CAPITULO IV: MATERIALES Y MÉTODOS.....	65
4.1	Área de estudio.....	65
4.2	Diseño de investigación	65
4.3	Población y Muestra.....	65
4.4	Procedimiento, técnicas e instrumentos de recolección de información	65
4.4.1	Gestión de la Data	68
4.5	Análisis estadístico.....	70
4.5.1	Variabilidad del P80	72
4.5.2	Efecto del P80 en la Recuperación de Cobre.....	73
4.5.3	Análisis Exploratorio de datos	75

4.5.4	Selección del Modelo	86
4.5.5	Sistema de Recomendaciones	87
4.5.6	Conectividad	89
4.5.7	Puesta en Línea	90
5	CAPITULO V: RESULTADOS	93
5.1	Métricas del Modelo	93
5.1.1	Fórmula del modelo.....	95
5.2	Sistema de Recomendaciones.....	97
5.3	Simulación Histórica.....	101
5.4	Puesta en línea	103
6	CAPITULO VI: DISCUSIÓN DE RESULTADOS	105
7	CONCLUSIONES	107
8	RECOMENDACIONES	109
9	REFERENCIAS BIBLIOGRÁFICAS	110
10	ANEXOS.....	113

Nomenclatura

Unidad	Descripción	Magnitud
μm	Micrómetro (micrón)	Longitud
tn	Tonelada	Masa
%	Porcentaje	Proporción
m^3/h	Metros cúbicos por hora	Flujo volumétrico
kPa	Kilopascal	Presión
ton/h	Toneladas por hora	Flujo másico
inch	Pulgada	Longitud
kW	Kilovatio	Potencia
kg/cm^2	Kilogramo por centímetro cuadrado	Presión
rpm	Revoluciones por minuto	Velocidad angular
kWh/tn	Kilovatio-hora por tonelada	Energía específica
kg/m^3	Kilogramo por metro cúbico	Densidad

Índice de tablas

Tabla 1: Lista de Variables Independientes.....	61
Tabla 2: Lista de Variable Dependiente	62
Tabla 3: Operacionalización de variables.....	63
Tabla 4: Tags del Circuito Molienda Clasificación	68
Tabla 5: Fechas de Mantenimientos en el Circuito Molienda - Clasificación.....	75
Tabla 6: Cantidad de Valores numéricos	77
Tabla 7: Límites Operativos	81
Tabla 8: Correlaciones entre Variables Predictoras	82
Tabla 9: Correlaciones entre Variables Predictoras y Variables a Predecir	83
Tabla 10: VIF de las variables predictoras	84
Tabla 11: Evaluación de modelos	86
Tabla 12: Variables consideradas en el Sistema de Recomendaciones.....	88
Tabla 13: Métricas de valores históricos y recomendados de variables operativas..	100
Tabla 14: Matriz de Consistencia	113

Índice de figuras

Figura 1: Mapa de Ubicación de la unidad minera Cerro Corona.....	33
Figura 2: Producción Acumulada durante el año 2022	34
Figura 3: Organigrama de CIO - Gold Fields	40
Figura 4: Diagrama de Flujo Proceso de Producción.....	43
Figura 5: Flowsheet Proceso de Chancado	66
Figura 6: Flowsheet Proceso de Molienda.....	67
Figura 7: Diagrama de Variables Operativas	69
Figura 8: Tonelaje Procesado por turno por alteración mineralógica	71
Figura 9: P80 por Tonelaje Procesado	72
Figura 10: Desviaciones Estándar históricos de P80	73
Figura 11: Recuperación de Cobre y Consumo Específico por rangos de P80.....	74
Figura 12: Recuperación de Cobre por Rangos de Desviación Estándar de P80	74
Figura 13: Variables con valores no numéricos	78
Figura 14: Gráficos de Violin de variables del circuito Molienda - Clasificación	80
Figura 15: Estandarización de variables independientes	85
Figura 16: Cuadro de Recomendaciones de Variables Operativas.....	87
Figura 17: Esquema de Conectividad del Modelo.....	89
Figura 18: Extracción de datos del PI System	90
Figura 19: Envío de datos al PI System	91
Figura 20: Tarea de Windows para ejecución del Modelo.....	91
Figura 21: Diagrama de Flujo del Sistema de Recomendaciones.....	92
Figura 22: Resultados del Modelo	93
Figura 23: Simulación del Modelo.....	94

Figura 24: Estructura del Modelo.....	95
Figura 25: Representación Gráfica de la Estructura del Modelo	96
Figura 26: Pareto de Variables con mayores apariciones en las Recomendaciones	97
Figura 27: Gráfico de barras con mayores apariciones en las Recomendaciones	98
Figura 28: Distribuciones de valores de variables según histórico y Recomendaciones	99
Figura 29: Gráfico de Distribución de P80 Simulado y Predicho	101
Figura 30: Beneficio de Recuperación de Cobre Banco 1 por Disminución de Variabilidad	102
Figura 31: Beneficio de Recuperación Final de Cobre por Disminución de Variabilidad	103
Figura 32: Visualización del Modelo en Línea en el PI Vision	104

RESUMEN

La presente tesis tiene como principal propósito desarrollar un sistema de recomendaciones en línea que optimice el control operativo del circuito de Molienda-Clasificación en la planta concentradora de la Unidad Minera Cerro Corona. La investigación busca abordar la variabilidad en el tamaño de las partículas del producto final, que afecta directamente a la recuperación de minerales. A través de la implementación de técnicas de Machine Learning, se pretende reducir esta variabilidad y, por ende, mejorar la eficiencia operativa de la planta.

El enfoque metodológico de la investigación incluye la recopilación y análisis de los datos históricos de la planta concentradora. A partir de estos datos, se desarrolló un modelo predictivo utilizando algoritmos de Machine Learning que permite anticipar el comportamiento del P80 (Tamaño de partícula en el que el 80% del material molido pasa a través de un tamiz de malla determinada) en función de diferentes variables operativas. Este modelo sirve como base para un sistema de recomendaciones situacional de variables controlables en línea del Circuito Molienda - Clasificación.

Los resultados obtenidos demostraron que el sistema de recomendaciones es capaz de reducir la desviación estándar del P80 en 2.84 μm , según las simulaciones realizadas, lo que representa una disminución significativa en la variabilidad del proceso. Este avance mejora la recuperación de minerales teniendo un impacto positivo en la sostenibilidad de la operación minera.

En conclusión, la investigación presenta una contribución innovadora al aplicar Machine Learning para la optimización en tiempo real de procesos. Donde proporciona una herramienta valiosa para las plantas concentradoras, permitiendo un control más preciso y eficiente.

Palabras Claves: Machine Learning; Molienda; Clasificación; P80

ABSTRACT

The purpose of this thesis is to develop an online recommendation system that optimizes the operational control of the Grinding-Classification circuit in a concentrator plant. The research aims to address the variability in the particle size of the final product, which directly affects mineral recovery. By implementing Machine Learning techniques, the goal is to reduce this variability and, consequently, improve the plant's operational efficiency.

The research methodology includes collecting and analyzing historical data from grinding operations at the Cerro Corona Mining Unit. Based on this data, a predictive model was developed using Machine Learning algorithms to anticipate the behavior of P80 (the particle size at which 80% of the milled material passes through a specified mesh screen, expressed in microns) based on different operational variables. This model serves as the foundation for a recommendation system that suggests controllable variables for the Grinding-Classification Circuit.

The results showed that the recommendation system significantly reduces P80 variability, contributing to greater stability and efficiency in the grinding process. This improvement enhances mineral recovery, having a positive impact on the sustainability of mining operations.

In conclusion, this research presents an original contribution by applying Machine Learning for real-time optimization of critical processes in the mining industry. The development of this recommendation system provides a valuable tool for concentrator plants, enabling more precise and efficient control.

Keywords: Machine Learning; Grinding; Classification; P80

1 CAPÍTULO I: INTRODUCCIÓN

1.1 Introducción

La industria minera enfrenta desafíos cada vez más complejos y exigentes en su búsqueda por optimizar la extracción de recursos de manera eficiente y sostenible. En este contexto las nuevas tecnologías emergentes, especialmente los modelos de Machine Learning, han surgido como herramientas cruciales para abordar estos desafíos.

El Machine Learning ofrece la capacidad de analizar grandes volúmenes de datos generados por las operaciones mineras, permitiendo la identificación de patrones y la predicción de variables o eventos críticos. Estas capacidades no solo mejoran la eficiencia operativa de las plantas concentradoras, sino que también contribuyen a la toma de decisiones informadas, lo que resulta en un uso más sostenible y seguro de los recursos.

Las plantas concentradoras juegan un papel crucial en la industria minera, ya que son responsables de procesar grandes cantidades de mineral y aumentar su concentración antes de su posterior refinamiento. El circuito de Molienda – Clasificación de una planta concentradora es de vital importancia en el procesamiento de minerales, pues la eficiencia de esta etapa tiene un impacto directo en el control del tamaño de las partículas, la recuperación del mineral y la reducción de costos. Este proceso puede sufrir diversas perturbaciones las cuales afectan directamente al tamaño de partícula del producto final, lo cual influye posteriormente en el proceso de flotación.

El propósito de esta investigación es brindar un sistema de recomendaciones que permite generar recomendaciones situacionales en tiempo real sobre las variables operativas más influyentes, para responder de forma dinámica a las perturbaciones que afectan al circuito, con el objetivo de reducir la variabilidad del tamaño del producto final, ya que debido a la complejidad

del circuito, encontrar la variable más influyente que está afectando al tamaño de partícula del circuito en un determinado momento, demanda un mayor tiempo de análisis y el proceso no se sigue desarrollando de una manera óptima.

Se propone iniciar el trabajo de investigación con el entendimiento del proceso, para identificar y analizar las variables del proceso que influyen el tamaño de las partículas en el circuito de Molienda - Clasificación de la planta concentradora. A través de la recopilación y análisis de datos históricos, se desarrollará y entrenará un algoritmo de Machine Learning capaz de predecir el tamaño de las partículas del circuito basado en estas variables. Posteriormente, a raíz de la aplicabilidad y explicabilidad del modelo, se desarrollará el sistema de recomendaciones que proporcionará los valores del proceso de las variables más importantes del circuito para la reducción de variabilidad del P80.

1.2 Planteamiento del Problema

1.2.1 Descripción de la problemática

El circuito de Molienda - Clasificación es clave en el procesamiento de minerales, ya que reduce el tamaño de partícula para facilitar la liberación y recuperación de los minerales. Dado su alto consumo energético, cualquier ineficiencia incrementa los costos y reduce la productividad. Además, la variabilidad en el tamaño de partícula compromete la eficiencia del proceso y la calidad del producto final.

A pesar de los avances tecnológicos en la automatización y el control de procesos, muchas plantas concentradoras continúan enfrentando dificultades para mantener un tamaño de partícula consistentemente óptimo en el circuito de Molienda - Clasificación. En el caso de estudio presente, en el periodo de tiempo planteado para la investigación se tiene una desviación estándar de 6.49 μm del P80. Esta variabilidad puede ser causada por varios factores, como la heterogeneidad del mineral alimentado, las condiciones fluctuantes del proceso y las limitaciones en la capacidad de respuesta de los sistemas de control actuales. La falta de recomendaciones precisas y adaptativas para ajustar las variables del proceso en tiempo real, impide la optimización continua del circuito de molienda, lo que resulta en un rendimiento subóptimo y mayores costos.

Abordar este problema es fundamental para mejorar la eficiencia operativa y la rentabilidad de las plantas concentradoras. La capacidad de predecir y ajustar las variables del proceso de molienda en tiempo real puede reducir significativamente la variabilidad del tamaño de las partículas, optimizando así la recuperación de minerales. La efectividad de la flotación de partículas finas en comparación con las partículas más gruesas es ampliamente reconocida, y una disminución del P80 en 10 micrones puede aumentar las tasas de recuperación del mineral hasta en un 5% (Fuerstenau, Jameson, & Yoon, 2007). El desarrollo de un algoritmo de Machine Learning que pueda aprender de los datos operativos y predecir el P80 con cierta exactitud, para

luego generar un sistema de recomendaciones en base al modelo desarrollado, puede proporcionar ajustes precisos y adaptativos para las variables del proceso, mejorando la eficiencia y la estabilidad del circuito de Molienda – Clasificación.

1.2.2 Formulación del problema

¿Se podría optimizar el control operativo del circuito de Molienda - Clasificación mediante un sistema de recomendaciones basado en un algoritmo de Machine Learning, con el fin de reducir la variabilidad del P80 y mitigar su impacto negativo en la recuperación de cobre?

1.3 Objetivos de la investigación

1.3.1 Objetivo General

- Generar un sistema de recomendaciones en línea, utilizando un modelo de Machine Learning que pueda minimizar la variabilidad del P80 del circuito de Molienda – Clasificación.

1.3.2 Objetivos Específicos

- Obtener un modelo predictivo óptimo para el P80 utilizando algoritmos y técnicas de Machine Learning.
- Generar un sistema de recomendaciones de las variables operativas a partir de la aplicabilidad y explicabilidad del modelo de Machine Learning, con capacidad de respuesta situacional en línea .
- Implementar en el PI System, un modelo de Machine Learning que permita predecir el P80 y tener el sistema de recomendaciones en línea.

1.4 Importancia y alcance de la investigación

La presente investigación desarrolla un algoritmo de Machine Learning y un sistema de recomendaciones para optimizar el circuito de molienda en una planta concentradora, lo cual resulta relevante tanto para la industria minera como para el ámbito académico.

La Molienda de Minerales es uno de los procesos más costosos y energéticamente intensivos en las operaciones mineras. Mejorar la eficiencia del circuito de molienda puede llevar a reducciones significativas en los costos operativos y la reducción de variabilidad en el tamaño de las partículas teniendo un impacto directo en la recuperación de minerales valiosos.

La aplicación de Machine Learning en el control y optimización de procesos industriales representa un avance significativo en el uso de tecnologías emergentes. El desarrollo de un algoritmo que pueda adaptarse y aprender de los datos en tiempo real, tiene el potencial de revolucionar la forma en que se gestionan los procesos mineros, proporcionando un ejemplo de cómo la digitalización y la inteligencia artificial pueden integrarse en las operaciones mineras.

Desde una perspectiva académica, esta investigación es un aporte a la creciente literatura sobre la aplicación de Machine Learning en el procesamiento de minerales. Proporciona un caso de estudio detallado sobre cómo los métodos de aprendizaje automático pueden ser utilizados para resolver problemas complejos en la industria, ofreciendo una base para futuras investigaciones y desarrollos en la industria.

Esta investigación se centra en mejorar el control operativo del circuito de Molienda-Clasificación de una planta concentradora tipo SABC-B. El estudio se llevó a cabo en la planta concentradora de la Unidad Minera Cerro Corona, ubicada en el distrito y provincia de Hualgayoc, departamento de Cajamarca. En el período comprendido entre Enero y Agosto del año 2023. La población de estudio estará constituida por los datos en línea del mineral procesado en la planta

de procesos en el período mencionado, seleccionadas mediante un muestreo con intervalo de cada cinco minutos obtenidos del PI System.

Se utilizarán técnicas estadísticas y simulaciones históricas para evaluar la variabilidad del P80 y su impacto en la recuperación del mineral. No se incluirán en el estudio los efectos de variaciones en la composición química del mineral ni los procesos posteriores al circuito de flotación.

1.5 Limitaciones de la Investigación

- La investigación se realiza con en el circuito Molienda – Clasificación de la Unidad Minera Cerro Corona, considerando las variables operativas en línea que se cuentan específicamente en dicha planta concentradora. Lo cual limita la generalización de los resultados obtenidos a otras plantas concentradoras.
- El periodo considerado en esta investigación corresponde a una campaña específica de liners del Molino SAG, por lo que los resultados y su interpretación podrían variar si se presentan cambios significativos en los liners o en las características del mineral procesado en la planta concentradora.
- El modelo de Machine Learning seleccionado resulta adecuado para los objetivos de este estudio, ofreciendo un buen equilibrio entre precisión, interpretabilidad y eficiencia computacional. Si bien existen alternativas más complejas, como redes neuronales profundas, no fueron consideradas ya que sus requerimientos de recursos y su menor interpretabilidad no se alineaban con el enfoque práctico y explicativo de esta investigación.
- La evaluación de los objetivos operativos se llevó a cabo principalmente mediante técnicas estadísticas, una decisión que permitió analizar de forma segura y eficiente los datos sin interferir con el funcionamiento de la planta. Esta metodología fue la más adecuada, dado el alto riesgo asociado a los experimentos controlados en la planta concentradora y la necesidad de considerar, en una futura implementación, los objetivos y estrategias a nivel de toda la operación minera.

2 CAPITULO II: MARCO TEÓRICO

2.1 Antecedentes de la investigación

Mamani Sacsí, J. (2022). En su tesis titulada, “APLICACIÓN DE REDES NEURONALES COMO ALGORITMOS DE MACHINE LEARNING PARA LA PREDICCIÓN DE RECUPERACIONES AURÍFERAS MEDIANTE CIANURACIÓN POR AGITACIÓN EN YAUCA - CARAVELÍ”. Desarrolla una de las aplicaciones de las redes neuronales como algoritmos de predicción de Machine Learning para predecir recuperaciones de oro (Au) en pruebas metalúrgicas de lixiviación cianurada en celdas de agitación a nivel laboratorio. Estas redes neuronales trabajan como un método de Inteligencia Artificial que procesan información como lo harían las redes neuronales biológicas, inspiradas en el proceso de aprendizaje del cerebro humano.

Palma Villanueva, R. (2021). En su tesis titulada, “INTELIGENCIA ARTIFICIAL Y MACHINE LEARNING PARA EL DESARROLLO E IMPLEMENTACIÓN DE SOFTSENSORS EN LA PREDICCIÓN DEL P80 PARA LA MOLIENDA SAG (SABC-A) - MINERA LAS BAMBAS”. Se desarrolló un softsensor evaluando distintos algoritmos de Machine Learning e Inteligencia Artificial para la planta concentradora de Cobre de Minera Las Bambas SA, ubicada en la cordillera de los andes al Sur de Perú departamento de Apurímac. La medición del tamaño de particular en el molino es problemática a pesar de la existencia de instrumentos en línea que muchas veces no tienen la disponibilidad adecuada. Sin embargo, este problema se puede abordar y resolver mediante el uso de softsensors, por lo que el tamaño de partícula se puede inferir a partir de las mediciones de otras variables del proceso.

Gallegos Cornejo, Y. (2022). En su tesis titulada, “ELABORACIÓN DE UN MODELO DE REDES NEURONALES RECURRENTES PARA LA PREDICCIÓN DE LA ALIMENTACIÓN A UNA CHANCADORA PRIMARIA”. Tiene como objetivo obtener un modelo de redes neuronales recurrentes para la predicción de la alimentación de una chancadora primaria, utilizando redes del tipo LSTM desarrolladas con Tensorflow. Como resultado se obtuvo un modelo que tuvo como RMSE de 26844.85 tn, dicho modelo consta de una arquitectura de una capa densa de entrada de 32 neuronas, una capa de filtro Dropout del 25%, una capa LSTM de 64 neuronas, obteniéndose una mejora del 13% en comparación del modelo preliminar.

Carrión Osnayo, C. (2018). en su tesis, “APLICACIÓN DE REDES NEURONALES ARTIFICIALES PARA LA PREDICCIÓN DE LA RECUPERACIÓN DE PLANTA CONCENTRADORA EN MINSUR S.A. – UNIDAD SAN RAFAEL”. Desarrolló un trabajo donde contiene la aplicación de inteligencia artificial o computacional (redes neuronales artificiales) en la predicción de la recuperación del mineral de Planta Concentradora mediante el uso de redes neuronales y para comparar los resultados obtenidos con los reales y establecer los entornos más adecuados para su uso, se utilizó el software NeuralTools, el cual utiliza la Red Neuronal más eficiente de acuerdo al tipo de salida que se desea (Categoría o Numérico Dependiente) pudiendo ser: Red Neuronal Probabilística y Red Neuronal de Regresión Generalizada. Las Redes Neuronales Artificiales (RNA), imita la funcionalidad del cerebro humano y se orientan como herramientas para la resolución de problemas prácticos, son capaces de aprender de la experiencia a partir de las señales o datos provenientes del exterior.

Ghasemi et al. (2023) desarrollaron un marco de inteligencia híbrida que combina conocimiento experto, técnicas de aprendizaje automático y algoritmos evolutivos para mejorar la eficiencia de los molinos SAG. Su estudio utilizó un extenso conjunto de datos industriales con 36,743 registros y evaluó el rendimiento de 17 modelos de Machine Learning, identificando a CatBoost como el modelo más preciso. Además, emplearon algoritmos evolutivos para optimizar parámetros clave, logrando mejorar las predicciones del rendimiento del molino y superar métodos tradicionales. Estos hallazgos resaltan el potencial del Machine Learning para optimizar la molienda y reducir la variabilidad del producto final, como el P80, lo que justifica la exploración de enfoques similares en otros contextos de molienda.

Zhang et al. (2024) llevaron a cabo un estudio en el circuito de molienda de Canadian Royalties Inc., donde utilizaron modelos de aprendizaje automático para analizar el comportamiento del proceso y proponer estrategias de optimización. Su investigación reveló que estas metodologías pueden identificar patrones complejos en los datos operativos, proporcionando información valiosa para mejorar la toma de decisiones. Además, los modelos de Machine Learning permitieron detectar correlaciones entre múltiples variables del proceso, facilitando la implementación de estrategias de control más precisas. Estos hallazgos refuerzan la importancia de incorporar inteligencia artificial en la gestión del circuito de molienda, permitiendo mejorar la estabilidad del proceso, reducir costos operativos y maximizar la eficiencia en plantas concentradoras.

2.2 Bases teóricas

El presente capítulo desarrolla los fundamentos teóricos que sustentan esta investigación, seleccionando temas clave que permiten comprender el contexto y la base científica del estudio. Se abordan conceptos esenciales como Ciencia de Datos, Machine Learning y Big Data, así como su integración en entornos industriales, lo cual es fundamental para el desarrollo del modelo propuesto. Asimismo, se incluye una revisión del sistema PI System y de la empresa minera Gold Fields, como elementos representativos del entorno operativo. Finalmente, se analizan aspectos técnicos del proceso minero, como la minería del cobre, su mineralogía, el chancado, la molienda y la flotación, los cuales son cruciales para entender las variables que influyen en el rendimiento del circuito de molienda. Esta estructura teórica busca proporcionar un marco sólido para la comprensión integral del problema abordado y la propuesta de solución basada en ciencia de datos.

2.2.1 Ciencia de Datos

La ciencia de datos se define como un campo interdisciplinario que combina la recopilación, limpieza, análisis y visualización de grandes conjuntos de datos con el objetivo de obtener información útil y conocimiento. Se basa en la aplicación de técnicas estadísticas, computacionales y de aprendizaje automático para descubrir patrones, tendencias y relaciones en los datos (Gandomi & Haider, 2015).

Actualmente en el contexto del Aprendizaje Automático o Machine Learning, la ciencia de datos juega un papel fundamental al proporcionar los datos necesarios para entrenar y evaluar los modelos de aprendizaje automático. La ciencia de datos también se encarga de la limpieza y preparación de los datos, lo cual es crucial para el éxito de cualquier modelo de aprendizaje automático (Jordan & Mitchell, 2015).

2.2.1.1 Etapas principales en los proyectos de Ciencia de Datos

Para la realización de proyectos de Ciencia de Datos, se tienen 5 etapas principales las cuales son conocidas también como el ciclo de vida de Proyectos de Ciencia de datos (Brownlee, 2023).

1. **Negocios:** El primer paso es comprender el problema de negocio que se busca resolver mediante la ciencia de datos. Esto implica definir los objetivos, las métricas de éxito y las limitaciones del proyecto.
2. **Datos:** El segundo componente es la recopilación y preparación de los datos. Esto incluye la limpieza, transformación y selección de las variables relevantes para el análisis.
3. **Análisis:** El tercer componente es el análisis de los datos. Esto implica la aplicación de técnicas estadísticas, computacionales y de aprendizaje automático para extraer conocimiento y generar información útil.
4. **Visualización:** El cuarto componente es la visualización de los datos. Esto implica la creación de gráficos, tablas y otros elementos visuales para comunicar los resultados del análisis de manera clara y efectiva.
5. **Comunicación:** El quinto y último componente es la comunicación de los resultados a las partes interesadas. Esto implica presentar los hallazgos de manera que sean comprensibles para los diferentes públicos objetivo.

2.2.1.2 Análisis Exploratorio de Datos

El análisis exploratorio de datos (EDA) se ha establecido como una etapa esencial en el proceso de análisis de datos, desempeñando un papel crucial en la comprensión inicial de los datos antes de aplicar técnicas analíticas más avanzadas. Según Tukey y Wilk (2021), el EDA utiliza una combinación de técnicas visuales y estadísticas que permiten a los analistas detectar patrones, identificar relaciones y evaluar la estructura de los datos de manera intuitiva. Estas herramientas iniciales no solo ayudan a entender las características fundamentales de los datos,

sino que también ofrecen una base sólida para la formulación de hipótesis y el diseño de modelos predictivos, minimizando así los riesgos de cometer errores en etapas posteriores del análisis.

Una característica destacada del EDA es su capacidad para adaptarse a la naturaleza iterativa del proceso analítico, donde las suposiciones iniciales pueden cambiar conforme se obtiene una comprensión más profunda de los datos. Behrens y Yu (2020) subrayan la importancia del EDA en la identificación de nuevas tendencias y la reevaluación de hipótesis conforme se exploran diferentes dimensiones del conjunto de datos. Esta flexibilidad es particularmente valiosa en contextos de datos complejos o no estructurados, donde los analistas deben estar preparados para ajustar su enfoque y redirigir su análisis basado en los hallazgos preliminares que emergen durante el EDA.

Además, el EDA no solo facilita la exploración inicial de los datos, sino que también es fundamental para la validación de los supuestos estadísticos que subyacen a muchos modelos analíticos más avanzados. Zhang y Rao (2019) argumentan que, antes de proceder con técnicas inferenciales como la regresión o el análisis de varianza, es crucial que los investigadores validen la conformidad de los datos con ciertos supuestos estadísticos, como la normalidad y la homogeneidad de varianzas. El EDA proporciona las herramientas necesarias para esta validación, permitiendo la detección de datos atípicos y la evaluación de la distribución de los datos, lo que, en última instancia, fortalece la robustez y la validez de las conclusiones obtenidas en las etapas posteriores del análisis.

2.2.1.3 Medidas de Tendencia Central

Las medidas de tendencia central son valores que representan el centro o punto típico de un conjunto de datos. Estas medidas son fundamentales en la estadística descriptiva, ya que

permiten resumir un conjunto de datos con un único valor que representa una tendencia general.

Las tres medidas principales de tendencia central son la media, la mediana y la moda.

1. **Media:** La media aritmética es la medida de tendencia central más común y se calcula sumando todos los valores del conjunto de datos y dividiéndolos por el número total de observaciones. La media es muy útil cuando los datos están distribuidos de manera simétrica, pero puede verse afectada por valores atípicos, que pueden sesgarla (Liao & Wu, 2021).
2. **Mediana:** La mediana es el valor que divide al conjunto de datos en dos mitades iguales cuando los datos están ordenados. Si el número de observaciones es impar, la mediana es el valor central; si es par, es el promedio de los dos valores centrales. La mediana es particularmente útil para datos asimétricos o cuando existen valores extremos o anómalos, ya que no se ve afectada por estos (Kang & Ding, 2020).
3. **Moda:** La moda es el valor que aparece con mayor frecuencia en un conjunto de datos. Un conjunto de datos puede tener una sola moda (unimodal), más de una moda (multimodal) o ninguna moda si todos los valores son diferentes. La moda es útil para describir datos categóricos o cuando se quiere identificar el valor más común en un conjunto de datos (Chung & Zhang, 2019).

2.2.1.4 Medidas de Dispersión

Las medidas de dispersión describen cuán dispersos o extendidos están los datos en torno a las medidas de tendencia central, estas medidas son cruciales para entender la variabilidad de los datos y proporcionan contexto adicional sobre la distribución de los datos. Las principales medidas de dispersión incluyen el rango, la varianza, la desviación estándar, el rango intercuartílico y los percentiles.

1. **Rango:** El rango es la diferencia entre el valor máximo y el valor mínimo de un conjunto de datos. Es la medida de dispersión más simple, pero también la más limitada, ya que

solo tiene en cuenta los valores extremos y no refleja la distribución completa de los datos (Zhou & Gao, 2021).

2. **Varianza:** La varianza mide la dispersión de los datos respecto a la media. Se calcula como el promedio de las diferencias al cuadrado entre cada valor y la media del conjunto de datos. La varianza da una idea de cuán dispersos están los datos, pero debido a que se expresa en unidades cuadradas, a menudo se prefiere utilizar la desviación estándar (Lee & Kim, 2020).
3. **Desviación Estándar:** La desviación estándar es la raíz cuadrada de la varianza y proporciona una medida de dispersión en las mismas unidades que los datos originales. Es ampliamente utilizada porque es fácil de interpretar y compara la dispersión relativa entre diferentes conjuntos de datos (Wang & Chen, 2020).
4. **Rango Intercuartílico (IQR):** El rango intercuartílico es la diferencia entre el tercer cuartil (Q3) y el primer cuartil (Q1), y mide la dispersión de la mitad central de los datos. Es especialmente útil para describir la dispersión en datos asimétricos o cuando hay valores atípicos, ya que se centra en la mediana y la distribución central (Zhang & Liu, 2020).
5. **Percentiles:** Los percentiles son medidas que dividen un conjunto de datos ordenados en cien partes iguales, proporcionando información sobre la posición relativa de un valor dentro de la distribución. Los percentiles son útiles para interpretar la distribución de los datos y comparar diferentes conjuntos de datos en términos de sus posiciones relativas (Chen & Huang, 2021).

2.2.1.5 Estandarización de Datos

La estandarización de datos es un proceso fundamental en el análisis estadístico y la ciencia de datos, que consiste en transformar los datos para que sigan una escala o distribución común. Este proceso es crucial cuando los datos provienen de diferentes fuentes o tienen unidades de medida distintas, lo que podría afectar la precisión de los análisis posteriores. La

estandarización permite comparar diferentes conjuntos de datos de manera coherente y facilita la aplicación de métodos estadísticos que requieren datos normalizados. Existen varios métodos de estandarización, entre los que destacan la normalización (min-max scaling), la estandarización z-score y la escalado robusto.

1. **Normalización (Min-Max Scaling):** La normalización, también conocida como min-max scaling, es un método que transforma los datos para que caigan dentro de un rango específico, generalmente entre 0 y 1. Este método es especialmente útil cuando los datos tienen diferentes unidades o rangos de valores. La fórmula básica para la normalización es:

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}}$$

Donde X es el valor original, X_{min} es el valor mínimo del conjunto de datos, y X_{max} es el valor máximo. Este tipo de estandarización es útil en algoritmos de aprendizaje automático que se basan en distancias, como el k-nearest neighbors, ya que garantiza que todas las características contribuyan de manera equitativa al modelo (Hastie, Tibshirani, & Friedman, 2020).

2. **Estandarización Z-Score:** La estandarización z-score transforma los datos para que tengan una media de 0 y una desviación estándar de 1. Este método es ideal cuando se necesita comparar datos que siguen una distribución normal o cuando los algoritmos de análisis asumen que los datos están distribuidos de esta manera. La fórmula para el z-score es:

$$Z = \frac{X - \mu}{\sigma}$$

Donde X es el valor original, μ es la media del conjunto de datos, y σ es la desviación estándar. Este tipo de estandarización es particularmente útil en técnicas estadísticas como la regresión lineal o en algoritmos de aprendizaje automático que requieren datos centrados y escalados (Gelman et al., 2019).

3. **Escalado Robusto:** El escalado robusto es un método que transforma los datos de manera que la mediana es 0 y el rango intercuartílico (IQR) es 1. Este enfoque es útil cuando los datos contienen valores atípicos, ya que estos pueden distorsionar las medidas de tendencia central y dispersión en métodos como la normalización o la estandarización z-score. El escalado robusto se calcula como:

$$X_{robust} = \frac{X - Mediana}{IQR}$$

Este método es particularmente efectivo en situaciones donde los datos están sesgados o tienen datos anómalos significativos, ya que minimiza su impacto en el análisis (Iglewicz & Hoaglin, 2020).

2.2.1.7 Tamaño de Muestra para Poblaciones Finitas

El cálculo del tamaño de muestra es un aspecto crítico en la planificación de estudios estadísticos, ya que determina cuántos elementos deben ser seleccionados de una población para que los resultados sean representativos. Cuando se trabaja con poblaciones finitas, es esencial ajustar el tamaño de muestra para reflejar el tamaño total de la población. Esto se hace

utilizando una fórmula que considera tanto el tamaño de la población como el nivel de confianza y el margen de error deseado.

La fórmula general para calcular el tamaño de muestra n en una población finita es:

$$n = \frac{N \cdot Z^2 \cdot p \cdot (1 - p)}{e^2 \cdot (N - 1) + Z^2 \cdot p \cdot (1 - p)}$$

Donde:

- **N** es el tamaño de la población.
- **Z** es el valor z correspondiente al nivel de confianza deseado.
- **p** es la proporción esperada de la población que tiene la característica de interés (si no se conoce, se asume 0.5 para maximizar la variabilidad).
- **e** es el margen de error deseado.

Esta fórmula ajusta el tamaño de muestra calculado para una población infinita utilizando el factor de corrección para poblaciones finitas. Esto es necesario porque, en poblaciones pequeñas, la variabilidad dentro de la muestra puede ser menor, lo que significa que se necesita una muestra más pequeña para obtener resultados precisos (Cochran, 2020).

2.2.2 Machine Learning

2.2.2.1 Principios Básicos

El Aprendizaje Automático (Machine Learning o ML) es un campo de la Inteligencia Artificial que se centra en el desarrollo de sistemas que pueden aprender y mejorar su rendimiento de forma autónoma a partir de la experiencia. Estos sistemas se entrenan con grandes conjuntos de datos, lo que les permite identificar patrones y realizar predicciones sin ser

programados explícitamente. Los principios básicos del ML se basan en los siguientes componentes clave (Géron, 2017):

- **Datos:** El elemento fundamental del ML son los datos. Los sistemas de ML se entrenan con grandes cantidades de datos, lo que les permite aprender y mejorar su rendimiento. Estos datos pueden ser estructurados o no estructurados.
- **Algoritmos:** Los algoritmos de ML son los responsables de analizar los datos y aprender de ellos. Existen diversos tipos de algoritmos de ML, cada uno con sus propias ventajas y desventajas.
- **Modelos:** Los modelos de ML son representaciones aprendidas de los datos. Estos modelos se utilizan para realizar predicciones, tomar decisiones y generar información útil.
- **Evaluación:** La evaluación del rendimiento de los modelos de ML es un paso crucial para garantizar su eficacia y confiabilidad. La evaluación se realiza utilizando métricas específicas que dependen del tipo de problema que se busca resolver.

2.2.2.2 Scikit-learn

Scikit-learn es una biblioteca de Python ampliamente utilizada en el ámbito del aprendizaje automático, reconocida por su capacidad para facilitar el análisis de datos y la creación de modelos predictivos. Ofrece una variedad de algoritmos que abarcan desde el aprendizaje supervisado hasta el no supervisado, incluyendo técnicas como regresión, clasificación, clustering y reducción de dimensionalidad. Esta versatilidad, junto con una interfaz intuitiva, permite a usuarios de distintos niveles de experiencia implementar modelos complejos sin necesidad de una programación avanzada, lo que la convierte en una herramienta accesible para una amplia gama de aplicaciones en ciencia de datos.

Además, Scikit-learn proporciona un conjunto de herramientas esenciales para el preprocesamiento de datos, como la normalización, estandarización, imputación de valores

faltantes y selección de características. Estas funcionalidades son fundamentales para preparar los datos de manera adecuada antes de la creación de modelos predictivos. Su integración con otras bibliotecas de Python, como NumPy y pandas, facilita la manipulación eficiente de datos, mientras que su compatibilidad con matplotlib permite la visualización clara de los resultados obtenidos, consolidándose así como una pieza clave en el flujo de trabajo de la ciencia de datos.

2.2.2.3 Tipos de Algoritmos

Los algoritmos de Machine Learning (ML) se clasifican en tres categorías principales: aprendizaje supervisado, aprendizaje no supervisado y aprendizaje por refuerzo, donde cada categoría se basa en un enfoque diferente para el aprendizaje y la predicción (Géron, 2017).

Tipos de algoritmos:

- **Aprendizaje supervisado:** En el aprendizaje supervisado, el algoritmo se entrena con un conjunto de datos que incluye ejemplos de entrada y salida. El objetivo es que el algoritmo aprenda la relación entre las entradas y las salidas para que pueda realizar predicciones precisas para nuevos datos.
- **Aprendizaje no supervisado:** En el aprendizaje no supervisado, el algoritmo se entrena con un conjunto de datos sin etiquetas de salida. El objetivo es que el algoritmo encuentre patrones y estructuras en los datos sin ninguna guía explícita.
- **Aprendizaje por refuerzo:** En el aprendizaje por refuerzo, el algoritmo aprende a tomar decisiones en un entorno interactivo. El algoritmo recibe recompensas o penalizaciones por sus acciones, y con el tiempo aprende a tomar decisiones que maximizan su recompensa a largo plazo.

El sobreajuste y el subajuste son dos problemas comunes en el aprendizaje automático que afectan la capacidad de generalización de un modelo. El sobreajuste ocurre cuando un modelo se ajusta demasiado a los datos de entrenamiento, capturando tanto las relaciones

verdaderas como el ruido o las irregularidades específicas de esos datos. Como resultado, aunque el modelo puede mostrar un alto rendimiento en el conjunto de entrenamiento, su capacidad para generalizar y hacer predicciones precisas en datos nuevos e inéditos se ve comprometida. Por otro lado, el subajuste se presenta cuando un modelo es demasiado simple para capturar las relaciones subyacentes en los datos, lo que lleva a un rendimiento deficiente tanto en el conjunto de entrenamiento como en el de prueba. En este caso, el modelo no ha aprendido lo suficiente sobre la estructura de los datos y, por lo tanto, falla al hacer predicciones precisas. Encontrar el equilibrio entre sobreajuste y subajuste es crucial para desarrollar modelos de aprendizaje automático robustos y generalizables.

2.2.2.4 Algoritmos de Árboles de Decisión

Los algoritmos de árboles de decisión son herramientas fundamentales en el ámbito del Aprendizaje Automático debido a su capacidad para tomar decisiones basadas en datos de manera jerárquica y estructurada. Un árbol de decisión es una estructura que se utiliza para representar decisiones y sus posibles consecuencias, incluyendo eventos de azar, resultados de las decisiones y costos o ganancias asociados. Cada nodo interno del árbol representa una prueba o condición sobre un atributo, cada rama representa el resultado de la prueba, y cada hoja del árbol representa una clase o valor de salida. Este tipo de algoritmo es apreciado por su simplicidad y capacidad para manejar tanto datos categóricos como numéricos, lo que lo hace útil en diversas aplicaciones, desde la clasificación y regresión hasta la toma de decisiones en negocios (Quinlan, 1986).

Una de las principales ventajas de los árboles de decisión es su interpretabilidad; es fácil para los humanos entender y visualizar las decisiones que se están tomando en cada nodo del árbol. Sin embargo, los árboles de decisión tienen limitaciones, como su tendencia a sobreajustar los datos de entrenamiento, especialmente en árboles muy profundos. Para mitigar este problema, se suelen emplear técnicas como la poda, que consiste en eliminar ramas que tienen

poca importancia o que no contribuyen significativamente a la precisión del modelo. Además, los árboles de decisión son la base de modelos más avanzados como los bosques aleatorios y el boosting, que combinan múltiples árboles para mejorar la precisión y robustez de las predicciones (Breiman, 2001).

Además de los algoritmos de árboles de decisión tradicionales, existen varios modelos avanzados que se basan en la combinación de múltiples árboles para mejorar la precisión y robustez de las predicciones. Entre estos modelos destacan `ExtraTreesRegressor`, `RandomForestRegressor`, `CatBoostRegressor`, `LGBMRegressor`, `XGBRegressor` y `GradientBoostingRegressor`, cada uno con características específicas que los hacen adecuados para diferentes tipos de problemas de regresión.

- **RandomForestRegressor** es una extensión del algoritmo de árboles de decisión que crea un conjunto de árboles de decisión (un "bosque") y los combina para mejorar la precisión de la predicción y controlar el sobreajuste. Cada árbol en el bosque se entrena en una muestra aleatoria del conjunto de datos con reemplazo, y el modelo final predice el valor promedio de todas las predicciones de los árboles individuales. Este enfoque robusto es especialmente útil para datos con mucha variabilidad o cuando se necesita un modelo generalizable (Khan & Zubair, 2020).
- **ExtraTreesRegressor (Extremely Randomized Trees Regressor)** es un modelo que, al igual que los bosques aleatorios, construye múltiples árboles de decisión. Sin embargo, a diferencia de `RandomForestRegressor`, `ExtraTreesRegressor` introduce una mayor aleatoriedad en el proceso de construcción de los árboles, lo que puede reducir la varianza del modelo y, en algunos casos, mejorar la precisión. Este modelo selecciona los puntos de corte de forma aleatoria para cada nodo, en lugar de elegir el mejor punto basado en una métrica como la reducción de la varianza (Geurts et al., 2006).

- **CatBoostRegressor** es un modelo basado en árboles de decisión que utiliza un enfoque de boosting, optimizando la predicción mediante el entrenamiento secuencial de múltiples árboles, donde cada árbol trata de corregir los errores de los anteriores. Una característica distintiva de CatBoost es su manejo eficiente de datos categóricos sin necesidad de preprocesarlos extensivamente, lo que puede llevar a mejores resultados con menos ajuste manual (Prokhorenkova et al., 2018).
- **LGBMRegressor (LightGBM Regressor)** es otro algoritmo de boosting basado en árboles, conocido por su alta velocidad y eficiencia, especialmente en conjuntos de datos grandes. LGBMRegressor utiliza una técnica de aprendizaje de gradiente para construir los árboles de forma iterativa, optimizando continuamente el modelo. Es particularmente efectivo en problemas de clasificación y regresión en grandes volúmenes de datos, gracias a su capacidad para reducir la memoria y el tiempo de entrenamiento (Ke, 2017).
- **XGBRegressor (Extreme Gradient Boosting Regressor)** es un modelo popular de boosting que se ha destacado en competiciones de ciencia de datos. XGBRegressor optimiza la predicción al minimizar la función de costo a través del método de boosting, lo que resulta en un modelo muy preciso y eficiente. Es conocido por su flexibilidad y la capacidad de manejar datos con gran cantidad de características, así como por su implementación de técnicas como la regularización para prevenir el sobreajuste (Chen, 2016).
- **GradientBoostingRegressor** es un modelo que también utiliza la técnica de boosting para mejorar las predicciones. Este modelo construye árboles de decisión secuencialmente, cada uno tratando de corregir los errores del anterior mediante el ajuste de la predicción basada en los residuos. Es especialmente útil cuando se busca un balance entre la capacidad de generalización y la precisión del modelo, aunque tiende a ser más sensible al sobreajuste si no se controla adecuadamente (Natekin & Knoll, 2013).

2.2.2.2 Evaluación y Selección de Modelos

La evaluación y selección de modelos de Machine Learning (ML) es un paso crucial para determinar el modelo más adecuado para un problema específico. Es importante considerar diversos factores para garantizar que el modelo elegido sea preciso, confiable y generalizable a nuevos datos.

Es fundamental seleccionar las métricas de evaluación adecuadas para el tipo de problema que se busca resolver. Es importante comparar diferentes modelos de ML para determinar el que mejor se ajusta al problema específico. Además de las métricas de evaluación y la comparación de modelos, se debe considerar otros factores como la complejidad del modelo, la interpretabilidad y la eficiencia computacional (Brownlee, 2023).

Entre las métricas de evaluación más comunes para problemas de regresión se encuentran el coeficiente de determinación (R^2), el error cuadrático medio (MSE), la raíz del error cuadrático medio (RMSE) y el error absoluto porcentual medio (MAPE).

- **R^2** , o coeficiente de determinación, es una métrica que indica la proporción de la varianza en la variable dependiente que es explicada por el modelo. Un valor de R^2 cercano a 1 sugiere que el modelo explica bien los datos, mientras que un valor cercano a 0 indica que el modelo no captura adecuadamente la variabilidad de los datos. Por ejemplo, si se utiliza R^2 para evaluar un modelo de predicción de precios de viviendas, un R^2 de 0.85 significaría que el 85% de la variabilidad en los precios puede ser explicada por el modelo (Chicco, 2021).
- **MSE** (Mean Squared Error) es una métrica que calcula el promedio de los errores al cuadrado entre los valores predichos y los valores reales. Esta métrica penaliza los errores grandes más que los pequeños, debido al cuadrado de las diferencias, y es útil para identificar si un modelo tiene predicciones con grandes desviaciones. Por ejemplo,

en un modelo que predice el consumo de energía, un MSE bajo indicaría que las predicciones están, en promedio, cerca de los valores reales (Hastie, 2017).

- **RMSE** (Root Mean Squared Error) es la raíz cuadrada del MSE y se utiliza para devolver el error en las mismas unidades que la variable de salida, lo que facilita la interpretación del error medio. Por ejemplo, si se utiliza RMSE para evaluar un modelo de predicción de temperaturas, un RMSE de 2 grados indica que, en promedio, las predicciones del modelo se desvían en 2 grados de los valores reales (Hastie, 2017).
- **MAPE** (Mean Absolute Percentage Error) es una métrica que expresa el error absoluto promedio en términos porcentuales, lo que facilita la interpretación y comparación entre diferentes modelos y escalas de datos. Es especialmente útil cuando se desea comprender el error relativo respecto al tamaño de los valores predichos. Por ejemplo, en un modelo de ventas, un MAPE del 5% indicaría que, en promedio, las predicciones del modelo tienen un error del 5% en relación con los valores reales de ventas (Makridakis, 2018).

2.2.2.3 SHAP

SHAP (SHapley Additive exPlanations) es un enfoque de interpretación de modelos de aprendizaje automático basado en la teoría de juegos, específicamente en el concepto de valores de Shapley. Desarrollado por Lloyd Shapley en 1953, el valor de Shapley es una solución para distribuir de manera justa la ganancia o el costo entre los participantes en un juego cooperativo. En el contexto del Aprendizaje Automático, SHAP se utiliza para explicar la contribución de cada característica individual en la predicción realizada por un modelo. Este enfoque proporciona una medida consistente y equitativa del impacto que cada característica tiene en la predicción, permitiendo una comprensión más profunda de cómo y por qué un modelo toma determinadas decisiones (Lundberg & Lee, 2017).

Una de las ventajas clave de SHAP es su capacidad para proporcionar interpretaciones locales y globales de las predicciones. A nivel local, SHAP asigna un valor a cada característica para una observación específica, mostrando cómo cada característica influye en la predicción. A nivel global, estos valores pueden agregarse para entender la importancia relativa de las características en todo el conjunto de datos. Esta dualidad hace de SHAP una herramienta poderosa en el ámbito de la interpretabilidad de modelos, especialmente en modelos de caja negra como los basados en árboles de decisión o redes neuronales. Además, SHAP es coherente con varias propiedades deseables, como la simetría y la eficiencia, lo que asegura que las explicaciones generadas sean racionales y justas (Lundberg et al., 2020).

2.2.3 Big Data

2.2.3.1 Definición y contexto

Big Data o "datos masivos" se refiere a conjuntos de datos que son demasiado grandes y complejos para ser procesados por herramientas tradicionales de análisis de datos (Gandomi & Haider, 2015). Se caracterizan por las 3 V: volumen (gran cantidad de datos), velocidad (alta velocidad de generación de datos) y variedad (diversidad de tipos de datos).

El Big Data ha revolucionado diversos sectores, incluyendo la salud, las finanzas, el marketing y la industria. Permite a las empresas y organizaciones obtener información valiosa de sus datos para tomar decisiones más inteligentes, mejorar la eficiencia operativa y crear nuevos productos y servicios (Gandomi & Haider, 2015).

2.2.3.2 Características del Big Data

El Big Data se caracteriza por un conjunto de propiedades que lo diferencian de los datos tradicionales. Estas características, conocidas como las "V del Big Data", son esenciales para comprender la naturaleza y el impacto de este fenómeno (Gandomi & Haider, 2015).

- **Volumen:** Se refiere a la gran cantidad de datos que se generan y almacenan. Los conjuntos de Big Data pueden llegar a terabytes, petabytes o incluso exabytes.
- **Velocidad:** La velocidad a la que se generan y procesan los datos es otra característica importante del Big Data. Los datos pueden fluir en tiempo real, lo que exige herramientas y técnicas de análisis especiales.
- **Variedad:** El Big Data se caracteriza por la diversidad de tipos de datos que abarca. Puede incluir datos estructurados, semiestructurados y no estructurados, como textos, imágenes, vídeos, sensores.

2.2.4 Integración de Ciencia de Datos, Machine Learning y Big Data

2.2.4.1 Puntos de Convergencia y Sinergia entre estos Campos

La Ciencia de Datos, el Machine Learning (ML) y el Big Data son tres disciplinas que se complementan y potencian mutuamente. La convergencia de estas áreas permite abordar problemas complejos y obtener información valiosa de grandes conjuntos de datos (Davenport & Patil, 2012).

Las tres áreas comparten el objetivo de obtener conocimiento a partir de los datos. La Ciencia de Datos se encarga de la extracción, limpieza, análisis y visualización de datos; el ML se centra en el desarrollo de algoritmos para aprender de los datos y hacer predicciones; y el Big Data proporciona la infraestructura y las herramientas para gestionar y procesar grandes conjuntos de datos. Las tres áreas trabajan juntas en el ciclo de vida del análisis de datos, desde la definición del problema hasta la implementación y evaluación de las soluciones (Davenport & Patil, 2012).

2.2.4.2 Beneficios de la Implementación Conjunta

La convergencia entre la Ciencia de Datos, el Machine Learning (ML) y el Big Data genera una poderosa sinergia que ofrece diversos beneficios para las empresas y organizaciones (Gandomi & Haider, 2015).

Mejora en la toma de decisiones, la capacidad de analizar grandes conjuntos de datos permite a las empresas tomar decisiones más informadas y estratégicas. El ML puede identificar patrones y tendencias en los datos que serían difíciles de detectar manualmente, lo que permite a las empresas anticiparse a los cambios del mercado y tomar decisiones más proactivas (Gandomi & Haider, 2015).

Optimización de procesos, la Ciencia de Datos y el ML pueden identificar áreas de ineficiencia en los procesos y recomendar soluciones para optimizarlos. Esto puede traducirse en una reducción de costos, un aumento de la productividad y una mejora en la calidad del servicio (Gandomi & Haider, 2015).

2.2.4 PI System

PI System es una plataforma integral de gestión de datos en tiempo real desarrollada por OSIsoft, que se utiliza para recopilar, almacenar, analizar y visualizar grandes volúmenes de datos industriales. Este sistema es ampliamente empleado en diversas industrias, incluyendo energía, manufactura, petróleo y gas, y servicios públicos, entre otras, debido a su capacidad para integrar datos de múltiples fuentes y proporcionar una visión unificada del rendimiento operativo. PI System facilita la toma de decisiones basadas en datos al ofrecer acceso inmediato a información crítica y al permitir la monitorización continua de procesos, lo que ayuda a mejorar la eficiencia, reducir costos y optimizar la producción (OSIsoft, 2020).

Una de las características clave de PI System es su capacidad para manejar datos históricos y en tiempo real simultáneamente. Esto es posible gracias a su arquitectura escalable, que incluye el PI Data Archive, encargado de almacenar grandes cantidades de datos de manera eficiente, y el PI Asset Framework, que organiza los datos en estructuras lógicas y relaciona la información con los activos físicos de la organización. Además, PI System ofrece herramientas avanzadas de análisis, como PI Vision, que permiten a los usuarios crear visualizaciones personalizadas, realizar análisis detallados y generar informes, todo en tiempo real. Estas capacidades hacen de PI System una herramienta esencial para la transformación digital en entornos industriales (Wang et al., 2021).

Como se mencionó en el párrafo anterior, existen componentes específicos que amplían su funcionalidad y facilitan la integración y análisis de datos en tiempo real.

- **PI Data Link** es una herramienta que permite a los usuarios integrar datos de PI System directamente en hojas de cálculo de Microsoft Excel. Esta integración es útil para realizar análisis avanzados y generar informes personalizados utilizando las capacidades de cálculo de Excel, mientras se mantiene la conexión a los datos en tiempo real proporcionados por PI System. PI Data Link facilita a los usuarios el acceso y

manipulación de datos históricos y en tiempo real sin necesidad de conocimientos avanzados de programación (OSIsoft, 2020).

- **PI Vision** es una herramienta de visualización basada en web que permite a los usuarios crear dashboards interactivos y personalizados para monitorear el rendimiento de activos y procesos en tiempo real. Con PI Vision, los usuarios pueden arrastrar y soltar elementos para construir visualizaciones que muestren datos de múltiples fuentes de manera coherente y fácil de entender. Esta herramienta es esencial para la toma de decisiones rápidas y basadas en datos, ya que proporciona una vista clara y unificada del estado operativo de la organización (Wang et al., 2021).
- **PI Web API** es una interfaz de programación de aplicaciones que permite a los desarrolladores acceder a los datos de PI System a través de servicios web. Esta API es fundamental para la integración de PI System con otras aplicaciones empresariales, plataformas en la nube, o sistemas personalizados, facilitando el intercambio de datos y la creación de soluciones personalizadas que se adaptan a las necesidades específicas de la organización. PI Web API soporta tanto el acceso a datos históricos como en tiempo real, lo que permite a las empresas aprovechar al máximo su infraestructura de datos (OSIsoft, 2020).
- **PI Data Archive** es el núcleo de PI System, encargado de almacenar y gestionar grandes volúmenes de datos de series temporales. Este componente es responsable de la recolección y almacenamiento de datos procedentes de diversas fuentes, incluyendo sensores, sistemas SCADA, y otros dispositivos industriales. El PI Data Archive está diseñado para ser altamente escalable y eficiente, lo que permite a las organizaciones manejar cantidades masivas de datos mientras mantienen un rendimiento óptimo en la consulta y recuperación de información (Bhandari & Injeti, 2018).

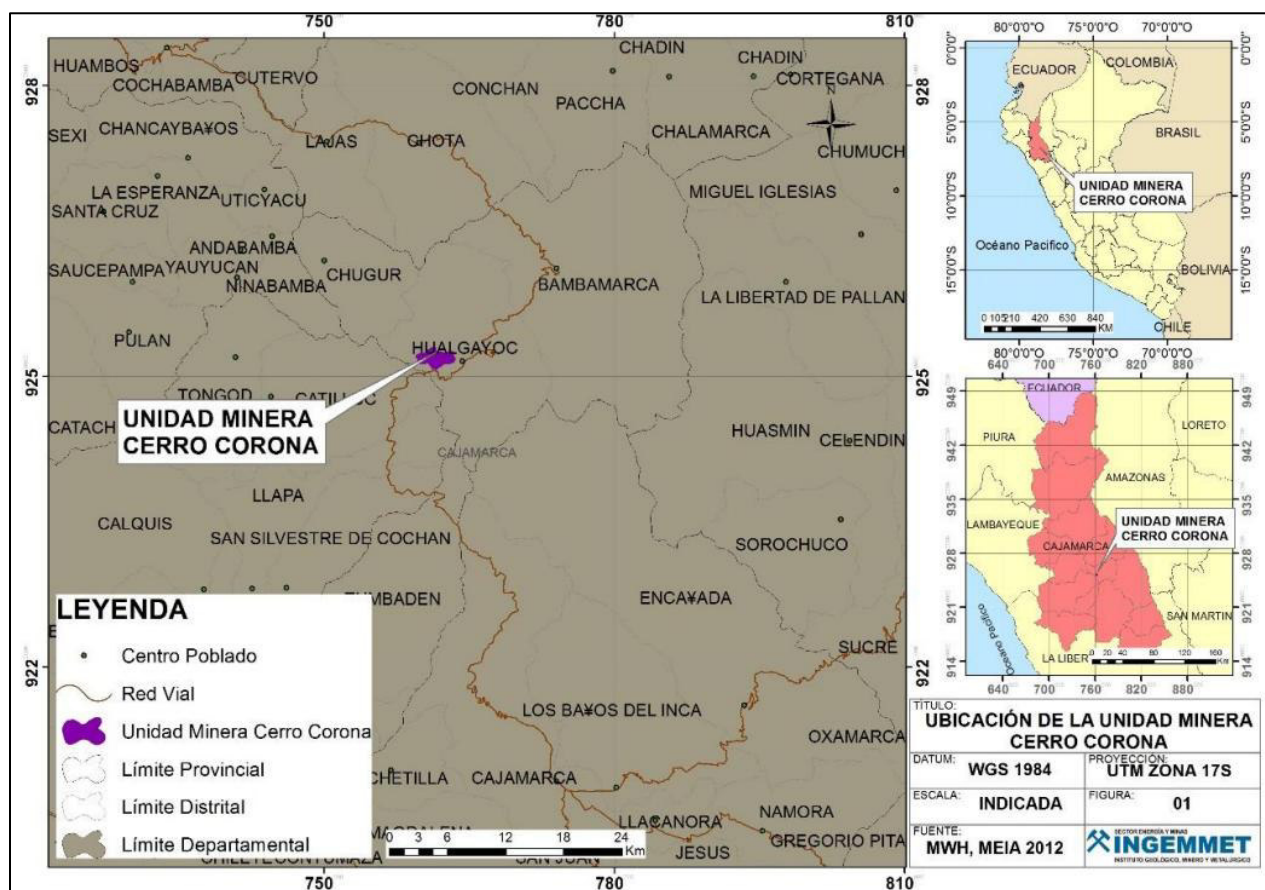
- **PI Asset Framework (AF)** es un modelo de datos orientado a objetos que organiza y contextualiza los datos almacenados en PI System. AF permite a las organizaciones definir estructuras lógicas que reflejan la organización física y operacional de sus activos, facilitando el acceso a información relevante y la realización de análisis más complejos. Además, AF permite la creación de plantillas que estandarizan la forma en que se capturan y presentan los datos, lo que ayuda a asegurar la consistencia y precisión en los análisis a lo largo del tiempo (Zhou et al., 2019).
- **PI Event Frames (EF)** es una extensión de PI Asset Framework que se utiliza para capturar y analizar eventos en tiempo real. PI Event Frames permite a los usuarios definir eventos específicos basados en condiciones de proceso, como paradas no programadas, cambios de estado o variaciones significativas en los datos. Estos eventos se pueden analizar para identificar patrones, causas raíz y oportunidades de mejora, lo que es fundamental para la implementación de estrategias de mantenimiento predictivo y mejora continua (OSsoft, 2020).

2.2.5 Gold Fields

Gold Fields es una empresa minera perteneciente al grupo sudafricano Gold Fields Ltd., corporación con más de 130 años de experiencia y reconocida como una de las productoras de oro más importantes en el mundo. Gold Fields Ltd. cuenta con operaciones mineras en Sudáfrica, Ghana, Australia y Perú, así como con proyectos en Chile y Filipinas.

En el Perú, Gold Fields inicio operaciones a mediados del año 2008 en la unidad Cerro Corona, ubicada aproximadamente a 80 km al norte de la ciudad de Cajamarca, provincia y distrito de Hualgayoc, en la comunidad campesina El Tingo, anexo predio La Jalca, caseríos Coymolache y Pilancones.

Figura 1: Mapa de Ubicación de la unidad minera Cerro Corona



Fuente: INGENMET

Unidad Minera Cerro Corona produce concentrados ricos en cobre y con las características necesarias (concentración, leyes, etc.) para salir al mercado.

La producción acumulada de la operación alcanzó un nivel de 28,636 toneladas de las 28,300 presupuestadas para el 2021. En el 2022 se logró alcanzar un total de 29.357 kilotoneladas minadas, lo que significó un aumento de 3,12% más de las kilotoneladas minadas presupuestadas (28.468 ktn). Esta cifra hace referencia tanto al mineral como al desmonte que luego se movilizan hacia sus respectivos depósitos finales. Alcanzando una producción de 129.267 de onzas de oro (Au) y 26. 995 toneladas de cobre (Cu) totalizando 260. 455 onzas equivalentes de oro (Au). Con relación a la producción de minerales, en 2022, procesamos un total de 6.721 kilotoneladas, lo que representó una disminución del 1,41 % en comparación con 2021, y el cumplimiento del 99,94% del volumen presupuestado para el año, como se muestra en la siguiente figura:

Figura 2: Producción Acumulada durante el año 2022

Producción real vs producción presupuestada de minerales en kilotoneladas (kt)			
AÑO	2020	2021	2022
REAL	6.796	6.796	6.796
PRESUPUESTADA	6.749	6.749	6.749
Producción real vs producción presupuestada de oro en onzas			
AÑO	2020	2021	2022
REAL	119.393	113.278	129.267
PRESUPUESTADA	153.342	129.807	120.212
Producción real vs producción presupuestada de cobre en onzas			
AÑO	2020	2021	2022
REAL	24.857	25.948	26.955
PRESUPUESTADA	27.493	24.704	27.013

Fuente: Gold Fields

2.2.5.1 Propósito

El propósito de Gold Fields es: “Crear valor que perdure más allá de la minería”, implicando esto que la empresa busca generar beneficios sostenibles y duraderos no solo para sí misma, sino también para las comunidades locales, el medio ambiente y otras partes interesadas involucradas en la operación minera. En lugar de centrarse únicamente en la extracción de recursos minerales, con este propósito, busca maximizar los beneficios a largo plazo para todas las partes involucradas, incluso después de que se agoten los recursos minerales. Esto implica adoptar prácticas y enfoques responsables que promuevan la sostenibilidad, la mitigación de impactos ambientales, la creación de empleo, el desarrollo comunitario y la generación de valor económico y social en general.

2.2.5.2 Visión

La visión de Gold Fields es: “Ser la empresa minera aurífera más reconocida por generar un valor superior sostenible”, con esta visión la empresa busca:

- **Reconocimiento:** La empresa aspira a ser reconocida tanto a nivel nacional como internacional como líder en la industria aurífera.
- **Valor superior:** La empresa se propone generar un valor superior en comparación con otras empresas mineras. Esto incluye maximizar la rentabilidad, eficiencia operativa y la creación de valor económico a largo plazo.
- **Sostenibilidad:** Esto implica operar de manera responsable y consciente del medio ambiente, mitigando los impactos negativos y promoviendo prácticas ambientales y sociales responsables.
- **Generación de valor sostenible:** La empresa busca no solo generar beneficios económicos a corto plazo, sino también crear valor sostenible a largo plazo. invirtiendo en el desarrollo de las comunidades locales, generando empleo y promoviendo prácticas empresariales responsables.

2.2.5.2 Valores

Los valores de una de Gold Fields son: la seguridad, integridad, respeto, responsabilidad, innovación y cumplimiento. Estos valores guían el comportamiento y las acciones de la empresa y entienden de la siguiente manera:

- **Seguridad:** La seguridad es un valor crucial en la industria minera, ya que implica garantizar la seguridad y el bienestar de los empleados mediante la implementación de políticas y prácticas de seguridad, el uso de equipos de protección personal, la identificación y mitigación de riesgos y la promoción de una cultura de seguridad.
- **Integridad:** La integridad se refiere a la honestidad, la ética y la transparencia en todas las actividades de la empresa. Implica actuar con honestidad, cumplir con los más altos estándares éticos y legales, y ser transparente en las operaciones y decisiones comerciales.
- **Respeto:** El valor del respeto implica tratar a todas las personas con consideración, dignidad y equidad, independientemente de su posición o antecedentes.
- **Responsabilidad:** La responsabilidad implica asumir las consecuencias de las acciones y decisiones de la empresa. Esto involucra el tomar medidas para minimizar los impactos negativos en el medio ambiente, comprometerse con el desarrollo sostenible, cumplir con las regulaciones, normas aplicables, etc.
- **Innovación:** La innovación implica fomentar la creatividad y la mejora continua en todas las áreas de la empresa, desarrollando tecnologías y prácticas mineras más eficientes y sostenibles.
- **Cumplimiento:** El cumplimiento se refiere al apego a las regulaciones, normas y políticas aplicables. Esto implica cumplir con las leyes ambientales, laborales, de seguridad y otras regulaciones relevantes, así como con las políticas internas de la empresa.

2.2.5.3 Objetivos de la empresa

Los objetivos de Gold Fields se plantean como 3 pilares estratégicos para el año 2023, siendo estos los siguientes:

a. Maximizar el potencial de nuestros activos actuales por medio de las personas y la innovación:

El objetivo estratégico de maximizar el potencial de los activos actuales mediante las personas e innovación implica una gestión centrada en obtener el máximo valor de los activos, mediante el desarrollo del talento humano, el fomento de la innovación y la adopción de tecnologías y prácticas más eficientes y sostenibles. Esto busca impulsar la productividad, la rentabilidad y la competitividad de la empresa minera en el mercado.

- **Maximizar el potencial de los activos:** La empresa busca aprovechar al máximo el valor y el rendimiento de sus activos existentes, como las minas, las instalaciones de procesamiento y los recursos minerales. Esto significa optimizar la eficiencia y la productividad de las operaciones, garantizar la utilización óptima de los recursos y maximizar los retornos financieros.
- **Personas:** Reconoce la importancia del capital humano en el éxito de la empresa. Implica atraer, desarrollar y retener talento altamente capacitado, motivado y comprometido. Esto se logra mediante programas de capacitación y desarrollo, promoción de un ambiente de trabajo positivo y colaborativo, y fomento de la participación y la excelencia en todas las áreas de la organización.
- **Innovación:** La empresa busca impulsar la innovación en sus operaciones y procesos. Esto involucra fomentar un enfoque creativo y receptivo al cambio, promover la adopción de nuevas tecnologías y prácticas más eficientes y sostenibles, e identificar oportunidades para mejorar la productividad, la seguridad y la rentabilidad mediante la innovación continua.

b. Consolidar nuestro compromiso de primer orden en materia de ESG

El objetivo hace referencia a que Gold Fields minera se compromete a operar de manera responsable y sostenible, minimizando los impactos negativos y generando un impacto positivo en el medio ambiente, las comunidades y las prácticas de gobernanza. Esto refleja el compromiso de la empresa con la responsabilidad corporativa y el desarrollo sostenible en todas sus operaciones.

- Medio ambiente (Environmental): Se busca minimizar los impactos ambientales de sus operaciones, cumplir con las regulaciones ambientales, adoptar prácticas de gestión sostenible de recursos naturales y promover la conservación y protección del medio ambiente. Esto implica implementar medidas para reducir la huella ambiental, gestionar adecuadamente los residuos, mitigar los riesgos ambientales y promover la biodiversidad.
- Aspectos sociales (Social): La empresa se compromete a tener un impacto positivo en las comunidades locales y otras partes interesadas. Esto involucra respetar los derechos humanos, promover relaciones laborales justas y seguras, contribuir al desarrollo socioeconómico de las comunidades locales, apoyar la educación y la salud, y mantener un diálogo abierto y transparente con las partes interesadas.
- Gobernanza (Governance): La empresa busca establecer prácticas de gobernanza sólidas y transparentes. Esto significa tener una estructura organizativa clara, contar con políticas y procedimientos éticos, asegurar la gestión responsable de los recursos financieros, promover la diversidad y la inclusión en la toma de decisiones, y garantizar la transparencia en la divulgación de información financiera y no financiera.

c. Hacer crecer el valor y la calidad de nuestra cartera de activos

El objetivo estratégico de Gold Fields para hacer crecer el valor y la calidad de su cartera de activos toma en cuenta diversas consideraciones:

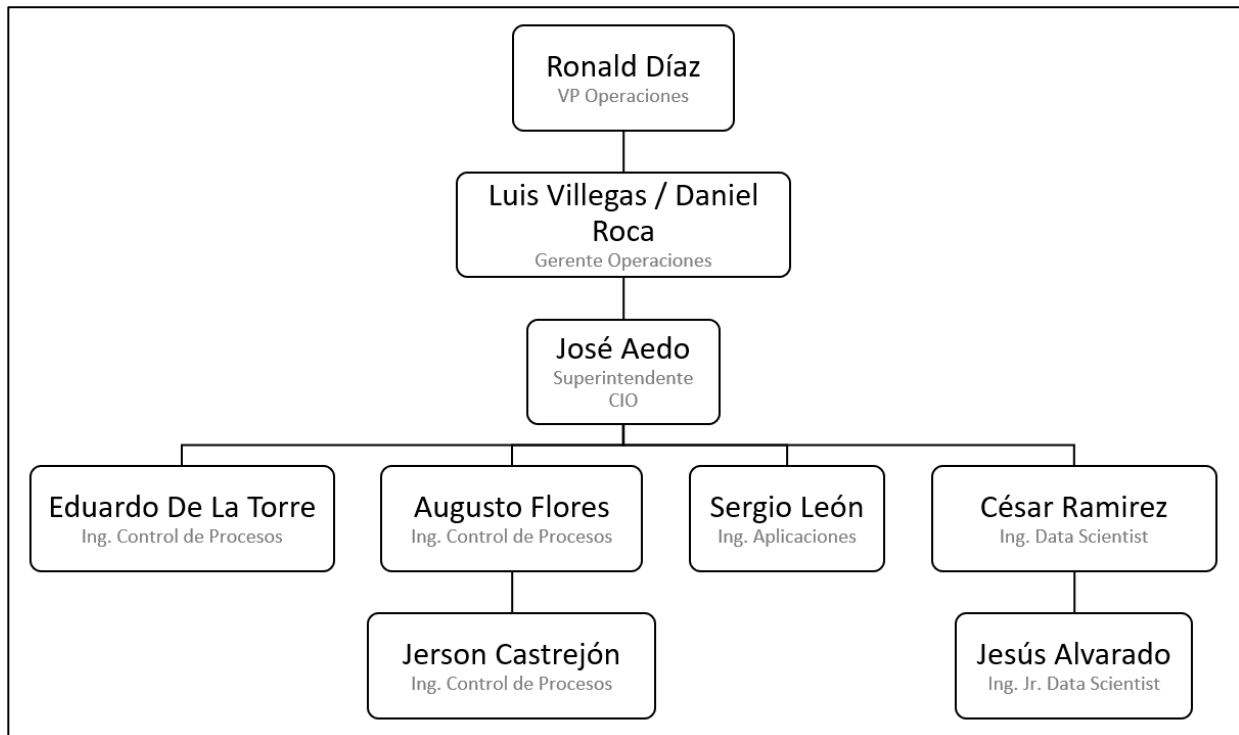
- **Crecer el valor:** La empresa busca aumentar el valor económico y financiero de su cartera de activos. Esto se resuelve con la adquisición de nuevos activos mineros con alto potencial de rentabilidad, la expansión de operaciones existentes, el desarrollo de nuevos proyectos mineros o la diversificación hacia otros recursos minerales o sectores relacionados.
- **Mejorar la calidad de los activos:** La empresa se propone mejorar la calidad de su cartera de activos, lo cual puede implicar optimizar la calidad de los minerales extraídos, aumentar las reservas minerales, mejorar la eficiencia y productividad de las operaciones, y adoptar tecnologías y prácticas más avanzadas y rentables.
- **Evaluación y gestión de riesgos:** La empresa realiza una evaluación continua de los riesgos asociados a su cartera de activos, incluyendo factores geológicos, económicos, políticos y regulatorios.
- **Generación de valor a largo plazo:** La empresa busca generar valor sostenible y a largo plazo en su cartera de activos. Esto significa la adopción de prácticas de gestión sólidas, la optimización de la rentabilidad a largo plazo y la adaptación a los cambios en el mercado y la industria.

En resumen, el objetivo estratégico de hacer crecer el valor y la calidad de la cartera de activos implica que la empresa minera busca aumentar el valor económico y financiero de sus activos, mejorar su calidad y rendimiento, gestionar los riesgos de manera efectiva y generar valor sostenible a largo plazo. Esto contribuye al crecimiento y la rentabilidad de la empresa en el tiempo.

2.2.5.3 Organigrama de la Empresa

El área sponsor del proyecto es el Centro Integrado de Operaciones, siendo esta parte de la vicepresidencia de operaciones, está conformado según la siguiente figura:

Figura 3: Organigrama de CIO - Gold Fields



Fuente: Elaboración Propia

La Vicepresidencia de Operaciones está liderada por el ingeniero Ronald Díaz, quien dirige la gestión operativa apoyado por los gerentes de operaciones, ingenieros Luis Villegas y Daniel Roca. A su vez, esta estructura da paso al Centro Integrado de Operaciones (CIO), encabezado por el superintendente ingeniero José Aedo y un equipo de profesionales especializados.

El Centro Integrado de Operaciones (CIO) de la unidad minera Cerro Corona, perteneciente a Gold Fields, constituye un área estratégica encargada del monitoreo y control centralizado de las operaciones mineras. Su principal objetivo es supervisar y coordinar de forma eficiente las actividades relacionadas con la producción, la seguridad y el rendimiento operativo.

El CIO cuenta con tecnología avanzada que permite la recopilación de datos en tiempo real desde distintas áreas de la mina, tales como extracción, transporte, procesamiento y seguridad, facilitando una toma de decisiones más oportuna y precisa.

2.2.5.4 Proceso de Producción

Nuestra unidad minera Cerro Corona produce un concentrado de cobre que contiene partículas de oro mediante métodos convencionales de explotación a tajo abierto y tratamiento de minerales de sulfuros y la extracción por flotación de concentrado. Posteriormente, el mineral extraído se transporta en camiones, aproximadamente unos 380 kilómetros hasta el puerto de Salaverry (en el norte del Perú), para enviarlos a las fundiciones de Asia y Europa por vía marítima. A continuación, se describe de manera general el proceso de producción:

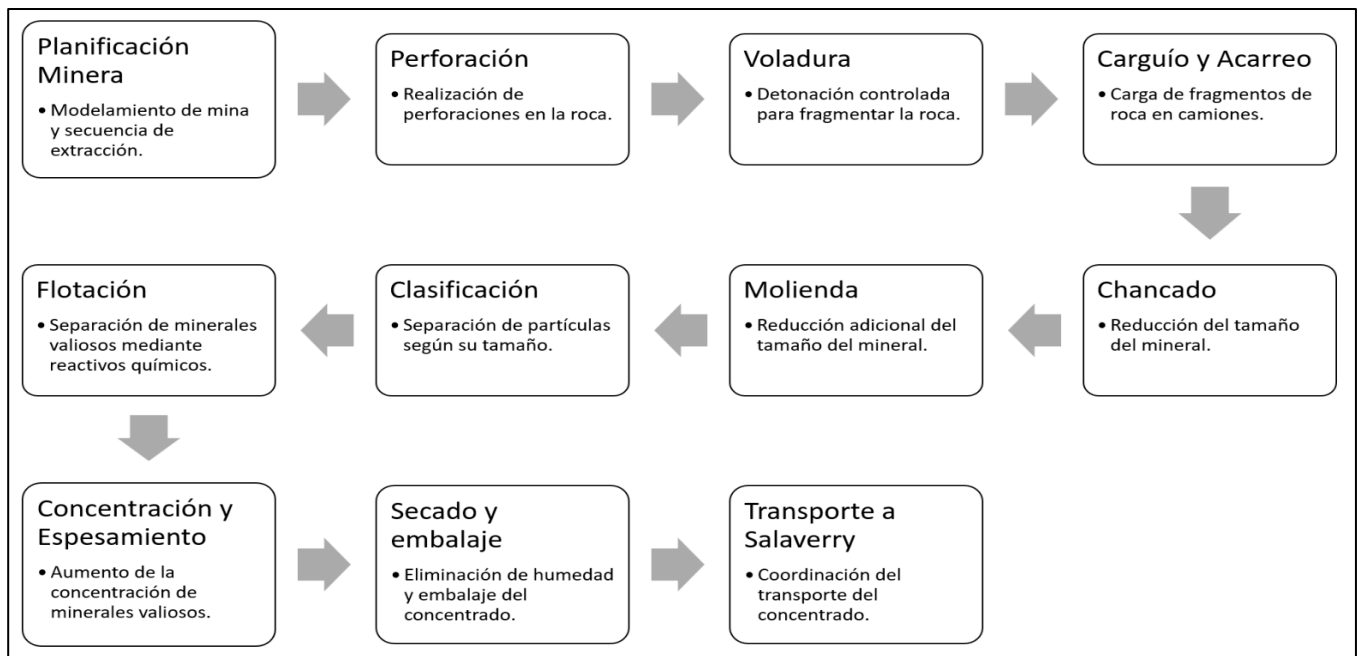
- **Planificación minera:** Antes de iniciar la extracción, se realiza una exhaustiva planificación minera que incluye el diseño de la mina, la definición de los bancos y la secuencia de extracción. Esto implica considerar factores como la calidad del mineral, la estabilidad de las paredes de la mina y la optimización de los recursos.
- **Perforación:** En esta etapa, se utilizan equipos de perforación para realizar perforaciones en la roca del banco. Estas perforaciones se realizan de acuerdo con un patrón establecido.
- **Voladura:** Una vez que se han realizado las perforaciones, se coloca explosivo. Posteriormente, se realiza la detonación controlada para fragmentar la roca y permitir la posterior carga y transporte del material.
- **Carga y transporte:** Después de la voladura, los fragmentos de roca son cargados en camiones o palas de gran capacidad para su transporte hacia el área de chancado. Estos equipos de carga y transporte pueden variar dependiendo del tamaño de la mina y de la cantidad de material a mover.

- **Chancado primario:** El mineral extraído de la mina es transportado hacia el área de chancado, donde grandes equipos de chancado, como trituradoras giratorias o de mandíbulas, se encargan de reducir el tamaño del mineral. En esta etapa, el mineral se tritura a un tamaño más manejable para el proceso de molienda posterior.
- **Chancado secundario y terciario:** En algunos casos, el mineral chancado primario puede pasar por etapas adicionales de chancado secundario y terciario. Estas etapas utilizan equipos de chancado más pequeños, como trituradoras de cono, para reducir aún más el tamaño del mineral.
- **Molienda:** Una vez que el mineral ha sido chancado, pasa a la etapa de molienda. En esta etapa, el mineral chancado se introduce en molinos de bolas o molinos SAG (Semi-Autogenous Grinding) que contienen bolas de acero u otros medios de molienda. El mineral y los medios de molienda se agitan dentro del molino, generando fuerzas de impacto y fricción que reducen aún más su tamaño.
- **Clasificación:** Después de la molienda, el material se somete a un proceso de clasificación. Esto permite separar las partículas de tamaño adecuado para continuar con el proceso de beneficio, mientras que las partículas más gruesas o finas son redirigidas para su posterior tratamiento.
- **Flotación:** Después de la etapa de molienda y clasificación, se utiliza el proceso de flotación para separar selectivamente los minerales de cobre y oro de los otros minerales y materiales presentes. En este proceso, se utilizan reactivos químicos que se adhieren a los minerales valiosos y los hacen flotar en la superficie, mientras que los minerales no deseados se hunden.
- **Concentración y Espesamiento:** Los minerales de cobre y oro flotados se recogen y se concentran mediante etapas adicionales de flotación y espesamiento. Esto permite aumentar la concentración de los minerales valiosos en el concentrado final.

- **Secado y embalaje:** El concentrado de cobre y oro obtenido se somete a un proceso de secado para eliminar la humedad antes de ser embalado en contenedores para su transporte y posterior comercialización.
- **Transporte:** Nuestro programa de transporte y despacho de concentrado desde Cerro Corona hacia el almacén de Salaverry se coordina semanalmente en función a la producción estimada y permite manejar un stock mínimo de concentrado en el hangar de la mina. El transporte se realiza diariamente y se inicia con la salida de tres convoyes cargados con concentrado, mientras que otros tres convoyes vacíos retornan a Cerro Corona. Contamos con camiones de tolvas metálicas cubiertas con tapas de fibra de vidrio que aseguran que el concentrado no quede expuesto en absoluto al medio ambiente.

Es importante tener en cuenta que pueden existir otras etapas y procesos auxiliares específicos en la producción de concentrado en la unidad minera Cerro Corona.

Figura 4: Diagrama de Flujo Proceso de Producción



Fuente: Elaboración Propia

2.2.6 Minería del Cobre

La minería del cobre en el Perú se refiere a la extracción, procesamiento y refinación del mineral de cobre. Este proceso se lleva a cabo en diferentes etapas, desde la exploración y explotación del yacimiento hasta la obtención del producto final, que puede ser cátodos de cobre o concentrado de cobre.

La minería del cobre en el Perú es una actividad económica de gran importancia, tanto por su contribución al Producto Bruto Interno (PBI) como por su generación de empleos y divisas. El país posee una de las mayores reservas de cobre del mundo, lo que lo convierte en un actor clave en el mercado internacional de este metal (Ministerio de Energía y Minas, 2021).

El Perú tiene una larga tradición minera que se remonta a la época preincaica. La explotación del cobre se intensificó a partir del siglo XIX con la llegada de empresas extranjeras. En la actualidad, el sector minero es uno de los pilares de la economía peruana, representando alrededor del 10% del PBI y el 60% de las exportaciones del país (Sociedad Nacional de Minería, Petróleo y Energía, 2023).

2.2.6.1 Importancia Industrial

El cobre es un metal versátil y vital para la industria moderna. Su amplia gama de propiedades lo convierte en un componente esencial en una gran variedad de sectores industriales. La minería del cobre, como actividad que provee este metal, juega un rol fundamental en el desarrollo y funcionamiento de la industria a nivel global.

La industria del cobre, a pesar de su importancia económica e industrial, enfrenta diversos retos y desafíos que ponen en riesgo su sostenibilidad a largo plazo. Estos desafíos abarcan desde la escasez de recursos y la volatilidad del mercado hasta el impacto ambiental y social de la actividad minera.

Además, la transición hacia una economía verde ha incrementado significativamente la demanda global de cobre, ya que es un insumo esencial para el desarrollo de energías renovables, vehículos eléctricos y sistemas de almacenamiento de energía. Esto refuerza la necesidad de adoptar prácticas mineras más eficientes, responsables y sostenibles que aseguren el suministro continuo de este recurso estratégico.

2.2.7 Mineralogía del Cobre

La mineralogía del cobre abarca el estudio de los diversos minerales que contienen este metal esencial. La comprensión de sus propiedades físicas, químicas y cristalográficas es crucial para la exploración, extracción y procesamiento del cobre.

Principales características de los minerales de cobre:

Química: El cobre se encuentra principalmente en minerales sulfurados, como la calcopirita, bornita y calcosina. También se presenta en minerales oxidados (malaquita, azurita), carbonatados (malaquita, azurita) y nativos (cobre nativo).

Cristalografía: Los minerales de cobre exhiben una variedad de estructuras cristalinas, incluyendo la tetragonal (calcopirita), cúbica (bornita) y ortorrómbica (calcosina).

Propiedades físicas: El color, brillo, dureza y densidad son propiedades físicas importantes que ayudan a identificar los minerales de cobre.

2.2.8 Yacimiento de Cerro Corona

El yacimiento de Cobre y Oro Cerro Corona se encuentra en un pórfido de diorita de forma cilíndrica y orientación subvertical, con un diámetro aproximado de entre 600 y 700 metros. Este pórfido está emplazado sobre formaciones de rocas calcáreas, margas y siliclásticas que datan del Cretáceo medio. La mineralización de Cobre y Oro dentro del pórfido está principalmente relacionada con zonas que presentan un sistema de vetillas de cuarzo. A diferencia de otros yacimientos similares, Cerro Corona destaca por su inusualmente alto contenido de Oro.

En su interior se han identificado tres zonas mineralizadas distintas: la Zona del Anillo, la Zona Norte y la Zona Sur, las cuales son tratadas de forma independiente en los modelos geológicos y de recursos.

Además de estas zonas, el yacimiento se caracteriza por la presencia de diversos dominios definidos según el grado de oxidación e intemperismo. Los procesos de lixiviación y oxidación supergénica han dado lugar al desarrollo de un manto de enriquecimiento de Cobre, que varía de débil a moderado. Esto ha permitido dividir el yacimiento, desde la superficie hacia niveles más profundos, en varias zonas: una zona de óxidos, una zona mixta de óxidos y sulfuros, una zona de sulfuros secundarios enriquecidos (supergénica), y una zona de sulfuros primarios (hipogénica).

Los materiales que se alimentan al proceso son clasificados según el tipo de alteración hidrotermal presente en el yacimiento Cerro Corona, los cuales se agrupan de la siguiente manera:

- **Alteración Argílica:**

Este tipo de alteración argílico se caracteriza por la siguiente mineralogía: Esmeclita/montmorillonita – sericita – clorita - óxidos de fierro, de coloración verdoso a naranja con venillas de cuarzo mineralizadas. El tipo de arcilla que predomina en este tipo de alteración, son arcillas del grupo de las esmeclitas.

- **Alteración Potásica:**

Los principales minerales que están presente en este tipo de alteración son la biotita secundaria - feldespato potásico - plagioclasas (albita o anortita) y magnetita. La mineralización presente en este tipo de alteración es por sulfuros primarios (calcopirita, bornita y pirita). El material que predomina en este tipo de alteración es el hipógeno, con presencia de CuOx asociado a fracturas y estructuras.

- **Alteración Silicificada:**

Este tipo de alteración se caracteriza por ser pervasiva donde la roca original queda convertida en una roca silícea, por lo general la mineralogía está dada por calcopirita, bornita y pirita, con presencia esporádica de fracturas rellenas de óxidos de cobre. También se caracteriza por abundantes vetas de cuarzo.

Cada una de las alteraciones hidrotermales antes mencionadas, tienen un comportamiento metalúrgico diferente en tonelaje, recuperaciones y concentraciones.

2.2.9 Chancado

El chancado de minerales, una etapa crucial en la metalurgia extractiva, reduce el tamaño de las rocas y minerales extraídos para facilitar su procesamiento posterior, este proceso implica la aplicación de fuerzas de compresión, impacto o corte para romper las partículas. Su objetivo principal es reducir el tamaño de las partículas a un tamaño adecuado para la molienda, la clasificación y la flotación, liberando los minerales valiosos contenidos en las rocas y facilitando su separación y recuperación. La eficiencia energética es un factor importante en el chancado, ya que se busca minimizar el consumo de energía para reducir costos y emisiones (Fuerstenau & Han, 2020).

Existen diferentes tipos de chancadores, como los de mandíbulas, giratorios y de cono, cada uno con sus ventajas y desventajas. Los factores que afectan el chancado incluyen las propiedades del material (dureza, abrasividad, forma), el tamaño de alimentación y el tamaño de producto deseado. La selección del tipo de chancador y la configuración del proceso dependen de estos factores (Wills & Finch, 2016).

2.2.10 Molienda y Clasificación

2.2.10.1 Principios Básicos Molienda

La molienda es un proceso crucial en la metalurgia extractiva que se lleva a cabo después del chancado para reducir aún más el tamaño de las partículas de mineral. Este proceso facilita la liberación de los minerales valiosos contenidos en las rocas, permitiendo su posterior separación y recuperación (Wills & Finch, 2016).

Principios de la molienda:

- **Reducción de tamaño:** La molienda reduce el tamaño de las partículas mediante la aplicación de diferentes fuerzas, como la compresión, el impacto o la abrasión. Diversos mecanismos de fragmentación, como la rotura por compresión, la abrasión y el astillado, actúan sobre las partículas, dependiendo del tipo de molino y las características del mineral (Napier-Munn & Morrell, 2019).
- **Liberación de minerales:** La molienda no solo reduce el tamaño, sino que también libera los minerales valiosos de la ganga, posibilitando su procesamiento posterior. El grado de liberación depende de la dureza del mineral, la granulometría del material y la energía aplicada en la molienda (Fuerstenau & Han, 2020).
- **Tipos de molinos:** Existen diversos tipos de molinos, como los de bolas, de barras, rodillos y SAG, cada uno con sus ventajas y desventajas. La elección del tipo de molino depende de las características del mineral, la capacidad de producción deseada y el tamaño de producto final (Wills & Finch, 2016).

La eficiencia de la molienda se ve afectada por diversos factores, como la dureza del mineral, el tamaño de alimentación, la velocidad de molienda, el tipo de molino utilizado, la cantidad de carga y el medio de molienda. La optimización de estos factores es crucial para

obtener un producto final con la granulometría deseada y minimizar el consumo de energía (Fuerstenau & Han, 2020).

2.2.10.2 Principios Básicos Clasificación

La clasificación de minerales es un proceso fundamental en la metalurgia extractiva que se lleva a cabo después de la molienda para separar las partículas de mineral en diferentes rangos de tamaño. Esta separación facilita el procesamiento posterior de los minerales, como la concentración y la recuperación de los minerales valiosos (Wills & Finch, 2016).

Adicionalmente, la clasificación se basa en la diferencia de tamaño de las partículas. Diversos métodos de clasificación se emplean, como la tamización, la sedimentación y la centrifugación. La elección del método depende del tamaño de las partículas, la densidad del mineral y la viscosidad del fluido utilizado (Napier-Munn & Morrell, 2019).

También, la eficiencia de la clasificación se ve afectada por diversos factores, como el tamaño de las partículas, la densidad del mineral, la viscosidad del fluido, el tipo de clasificador utilizado y la velocidad de flujo. La optimización de estos factores es crucial para obtener una clasificación precisa y eficiente (Fuerstenau & Han, 2020).

2.2.10.3 Importancia del P80

El P80, o tamaño de partícula al que el 80% de la cantidad del material molido es inferior, es un parámetro fundamental en la molienda de minerales, este valor tiene un impacto significativo en la eficiencia de los procesos posteriores. (Fuerstenau & Han, 2020)

Efectos del P80 en la Metalurgia Extractiva:

- **Liberación de Minerales:** Un P80 adecuado libera los minerales valiosos de la ganga, permitiendo su posterior procesamiento. Si el P80 es demasiado grande, algunos minerales valiosos podrían quedar atrapados en la ganga, mientras que si es demasiado

pequeño, se podría generar un exceso de finos que dificultan la recuperación (Napier-Munn & Morrell, 2019).

- **Eficiencia de la Concentración:** Un P80 adecuado optimiza la eficiencia de los procesos de concentración, como la flotación y la lixiviación. Un tamaño de partícula adecuado facilita la separación de los minerales valiosos de la ganga, mejorando la recuperación y la calidad del producto final (Xiao & Fuerstenau, 2021).
- **Consumo de Energía:** Un P80 demasiado fino puede aumentar el consumo de energía en la molienda, ya que se requiere más energía para reducir las partículas a un tamaño tan pequeño. La optimización del P80 puede ayudar a minimizar el consumo de energía y los costos operativos (Fuerstenau & Han, 2020).

El P80 óptimo depende de varios factores, como las características del mineral, el tipo de proceso de concentración utilizado y las especificaciones del producto final. La determinación del P80 ideal requiere un análisis cuidadoso de estos factores y la realización de pruebas de molienda (Wills & Finch, 2016).

2.2.11 Flotación de Cobre

2.2.11.1 Principios Básicos de la Flotación

La flotación es un proceso fundamental en la metalurgia extractiva para la recuperación de minerales valiosos de sus menas. Este proceso se basa en la diferencia de hidrofobicidad entre las partículas minerales y la ganga. Las partículas minerales se vuelven hidrofóbicas mediante la adición de reactivos químicos, mientras que la ganga permanece hidrofílica (Wills & Finch, 2016).

Adherencia de Minerales a Burbujas de Aire:

Las partículas minerales hidrofóbicas se adhieren a las burbujas de aire introducidas en la suspensión mineral. La formación de burbujas estables es crucial para la eficiencia del

proceso. La adición de agentes espumantes ayuda a estabilizar las burbujas y a aumentar la superficie disponible para la adhesión de las partículas minerales (Napier-Munn & Morrell, 2019).

Separación de Minerales y Ganga:

La diferencia de hidrofobicidad entre las partículas minerales y la ganga permite su separación. Las partículas minerales hidrofóbicas se adhieren a las burbujas de aire y ascienden a la superficie de la suspensión, mientras que la ganga hidrofílica se hunde en el fondo. La espuma que se forma en la superficie contiene los minerales valiosos, mientras que la ganga se descarta como residuo (Xiao & Fuerstenau, 2021).

2.2.11.2 Reactivos Utilizados y su Funcionamiento

Los reactivos son sustancias químicas que se añaden a la suspensión mineral durante la flotación para modificar las propiedades superficiales de las partículas y mejorar la eficiencia del proceso (Fuerstenau & Han, 2020). Su funcionamiento se basa en tres tipos principales: colectores, depresores y modificadores.

1. Colectores:

Los colectores son los reactivos más importantes en la flotación. Son moléculas con una parte hidrofóbica y otra hidrofílica. La parte hidrofóbica se adhiere a la superficie de las partículas minerales, haciéndolas hidrofóbicas y permitiéndoles unirse a las burbujas de aire (Wills & Finch, 2016). La parte hidrofílica se disuelve en el agua, lo que facilita la interacción del colector con las partículas minerales.

Tipos de Colectores:

- **Colectores de Xantato:** Son los más utilizados para la flotación de sulfuros metálicos (Napier-Munn & Morrell, 2019).

- **Colectores Ditionofosfatos:** Son más selectivos que los xantatos y se utilizan para la flotación de minerales como el cobre y el plomo.
- **Colectores de Amina:** Se utilizan para la flotación de minerales oxidados y carbonatados

2. Depresores:

Los depresores son reactivos que se añaden para hacer que las partículas de ganga sean más hidrofílicas, evitando que se adhieran a las burbujas de aire y mejorando la selectividad del proceso (Xiao & Fuerstenau, 2021).

Tipos de Depresores:

- **Cianuro:** Se utiliza para deprimir la pirita y otros sulfuros durante la flotación de minerales de oro.
- **Cal:** Se utiliza para deprimir la ganga de carbonato durante la flotación de minerales de sulfuro.
- **Silicato de sodio:** Se utiliza para deprimir la ganga de arcilla durante la flotación de minerales de óxido.

3. Modificadores:

Los modificadores son reactivos que se añaden para ajustar las propiedades del agua y mejorar la eficiencia del proceso de flotación.

Tipos de Modificadores:

- **Reguladores de pH:** Se utilizan para ajustar el pH del agua a un valor óptimo para la flotación.
- **Agentes espumantes:** Se utilizan para crear y estabilizar las burbujas de aire.
- **Floculantes:** Se utilizan para aglomerar las partículas finas y mejorar su recuperación.

2.3 Definición de términos

Técnicas Estadísticas: Son un conjunto de herramientas y métodos matemáticos que se utilizan para recopilar, organizar, analizar e interpretar datos.

Modelos: Un modelo es una representación simplificada de un objeto, sistema, proceso o idea. Su objetivo es facilitar la comprensión, el análisis o la predicción del comportamiento del objeto real.

Python: Es un lenguaje de programación de alto nivel, interpretado y de propósito general, conocido por su sintaxis sencilla y legible.

Pandas: Es una biblioteca de código abierto ampliamente utilizada para el análisis y manipulación de datos

Numpy: Es una biblioteca fundamental para la computación científica en Python. Proporciona soporte para arreglos multidimensionales y funciones matemáticas de alto rendimiento para operar sobre ellos.

Nodo: Es uno de los componentes fundamentales de la estructura del árbol de decisión, y representa un punto de decisión o un resultado dentro del modelo.

Poda: Es una técnica utilizada para reducir el tamaño del árbol de decisión, eliminando ramas que tienen poca importancia o que pueden llevar al modelo a sobre ajustarse a los datos de entrenamiento.

MSE: Significa Error Cuadrático Medio (Mean Squared Error, en inglés) y es una métrica de evaluación muy común en problemas de regresión.

MAPE: El MAPE (por sus siglas en inglés: Mean Absolute Percentage Error) o Error Porcentual Absoluto Medio es una métrica comúnmente utilizada para evaluar la precisión de

modelos de regresión o predicción de series temporales, especialmente cuando se desea expresar el error en términos porcentuales.

Tags: Son palabras clave o términos asociados a un contenido que sirven para clasificar, organizar y facilitar su búsqueda o identificación. Su uso es común en informática, desarrollo web, bases de datos, y muchas aplicaciones digitales.

INGEMMET: Es el acrónimo de Instituto Geológico, Minero y Metalúrgico del Perú. Es una institución pública adscrita al Ministerio de Energía y Minas (MINEM) y su principal función es realizar investigaciones geológicas, mineras y metalúrgicas para promover el desarrollo sostenible del país.

Entrenamiento de modelos: El entrenamiento de modelos es un proceso de aprendizaje automático en el que se ajusta un modelo para que pueda realizar una tarea específica.

Datos Estructurados: Los datos estructurados son un tipo de información organizada en un formato específico y predecible. Se asemeja a una tabla con filas y columnas, donde cada fila representa un registro individual y cada columna contiene un tipo específico de información sobre ese registro.

Datos no estructurados: Los datos no estructurados son información que no reside en un formato predefinido o un modelo de datos. Es decir, no se organizan en filas, columnas o campos como en una base de datos tradicional.

Datos semi estructurados: Los datos semi estructurados son aquellos que no se ajustan a un formato de tabla tradicional como los datos estructurados, pero que tampoco son completamente aleatorios como los datos no estructurados. Estos datos poseen cierta organización interna, pero no siguen un esquema rígido predefinido.

Algoritmo: Un algoritmo es un conjunto de pasos ordenados y lógicos que se siguen para resolver un problema en general.

Métricas: Las métricas son medidas cuantitativas que se utilizan para evaluar el rendimiento de un proceso, sistema o actividad.

Bytes: Un byte es una unidad de información digital que equivale a 8 bits. Se utiliza para medir la capacidad de almacenamiento de dispositivos como discos duros, memorias USB y memorias RAM. También se utiliza para medir el tamaño de archivos y programas informáticos.

Insights: Es un descubrimiento significativo sobre un comportamiento, necesidad o deseo de los consumidores que no era evidente anteriormente

Productividad: Es la eficiencia con la que se transforman los insumos (mano de obra, materiales, energía) en productos.

Yacimiento: Un yacimiento es un lugar donde se encuentra una concentración natural de un material de interés económico, como minerales, rocas, fósiles o combustibles.

Pórfido: Es un tipo de depósito mineral formado a partir de intrusiones ígneas, generalmente de composición intermedia (como diorita o granodiorita), que contienen minerales diseminados en grandes volúmenes de roca.

Diorita: Es una roca ígnea intrusiva de composición intermedia, formada por el enfriamiento lento del magma en el interior de la corteza terrestre.

Rocas Calcáreas: Son un tipo de roca sedimentaria compuesta principalmente por carbonato de calcio.

Rocas Margas: Son rocas sedimentarias detríticas compuestas por una mezcla de caliza y arcilla, en proporciones variables.

Rocas Siliclásticas: Son un tipo de rocas sedimentarias formadas a partir de fragmentos (clastos) de minerales y rocas preexistentes, principalmente de origen silíceo (como el cuarzo y los feldespatos). Se generan por procesos de meteorización, transporte, sedimentación y compactación.

Lixiviación: Es un proceso natural o inducido mediante el cual los componentes solubles de una roca o mineral son disueltos y transportados por un fluido, generalmente agua, hacia otras zonas del perfil geológico o hacia la superficie.

Oxidación: Es un proceso químico en el cual un elemento pierde electrones, generalmente al reaccionar con oxígeno. En geología y minería, este proceso ocurre comúnmente cuando los minerales entran en contacto con agua y oxígeno atmosférico, generando cambios en su composición química.

Supergénica: Se refiere a los procesos geológicos y minerales formados cerca de la superficie terrestre, como resultado de la alteración, oxidación y enriquecimiento de minerales primarios debido a la acción de agua, oxígeno y otros agentes atmosféricos.

Hipogénica: Se refiere a los procesos geológicos y a los minerales formados en condiciones profundas bajo la superficie terrestre, generalmente a partir de fluidos hidrotermales a altas temperaturas y presiones.

Alteración Hidrotermal: Es un proceso geológico en el que los minerales de una roca original se transforman químicamente debido a la interacción con fluidos calientes (hidrotermales), que circulan a través de la roca a altas temperaturas y presiones.

Cátodos de Cobre: Un cátodo de cobre es una lámina de metal puro obtenida mediante la electrorrefinación del cobre

Concentrado de Cobre: El concentrado de cobre es un producto intermedio en la producción de cobre. Se obtiene mediante la molienda y la flotación del mineral de cobre, y se caracteriza por un contenido de cobre significativamente mayor que el mineral original.

Producto Bruto Interno: El Producto Bruto Interno (PBI) es el valor total de todos los bienes y servicios finales producidos en un país durante un período determinado, generalmente un año.

Mineral: Un mineral es una sustancia natural, inorgánica, de composición química definida y estructura cristalina. Se encuentra en la Tierra en forma de cristales, granos o masas compactas.

Mineral Sulfurado: Un mineral sulfurado es un mineral que contiene azufre en su composición química. Estos minerales son generalmente opacos, con brillo metálico y dureza variable.

Mineral Oxidado: Un mineral oxidado es un mineral que ha experimentado una reacción química con oxígeno, generalmente a partir del aire o del agua. Esta reacción cambia la composición química del mineral original, creando un nuevo mineral con diferentes propiedades.

Estructura Cristalina: La estructura cristalina es la disposición ordenada y repetitiva de átomos, iones o moléculas en un sólido cristalino. Esta organización interna define las propiedades físicas y químicas del material, como su dureza, conductividad eléctrica y punto de fusión.

Eficiencia Energética: La eficiencia energética se define como la optimización del consumo de energía para realizar una actividad o servicio, sin afectar su calidad. Se trata de utilizar la menor cantidad de energía posible para obtener los mismos resultados.

Recuperación: La recuperación en metalurgia se refiere al porcentaje de metal valioso que se extrae de un mineral durante el proceso de procesamiento. Es un indicador clave de la eficiencia de la operación minera y metalúrgica.

Dureza: La dureza es la resistencia que opone un material a ser penetrado, rayado o deformado por otro material más duro.

Abrasión: Abrasión se refiere al desgaste o roce superficial de un material debido a la fricción con otro material.

Compresión: La fuerza de compresión es la fuerza que actúa perpendicularmente a la superficie de un objeto, empujándolo hacia adentro.

Impacto: Se refiere a la fuerza neta que actúa sobre un objeto durante un intervalo de tiempo.

Grado de Liberación: Se refiere proporción o porcentaje de un mineral valioso que ha sido separado físicamente del material no valioso luego de la molienda.

Granulometría: La granulometría es la medición y análisis de la distribución del tamaño de las partículas en un material granular.

Chancadora de Mandíbula: Una chancadora de mandíbula es una máquina empleada para triturar grandes bloques de roca en fragmentos más pequeños. Funciona mediante la acción de dos mandíbulas, una fija y otra móvil, que comprimen el material hasta fracturarlo.

Chancadora de Rodillos: Una chancadora de rodillos es un equipo de trituración que utiliza dos rodillos cilíndricos que giran en direcciones opuestas para aplastar y reducir el tamaño del material que pasa entre ellos.

Chancadora Cónica: Una chancadora cónica es un equipo de trituración que reduce el tamaño de rocas y minerales mediante la acción de un cono giratorio dentro de una cámara

cónica fija. El material se tritura a medida que es comprimido entre el cono y la pared de la cámara.

Molino Bolas: Un molino de bolas es un cilindro metálico que gira sobre su eje y que contiene bolas de acero, cerámica o goma en su interior. Estas bolas trituran el material que se introduce en el molino por impacto y fricción, reduciendo su tamaño a partículas finas.

Molino Barras: Un molino de barras es una máquina que utiliza barras de acero para moler material rocoso hasta un tamaño fino. Las barras se encuentran dentro de un cilindro giratorio y al girar, impactan y muelen el material.

Molino SAG: Es un tipo de molino utilizado en la industria minera para la molienda gruesa de minerales. Se caracteriza por su gran tamaño y por utilizar como medio de molienda tanto bolas de acero como el propio mineral.

PSI 300: Es un equipo de analizador de partículas en línea utilizado para medir de manera continua la distribución del tamaño de partículas en la descarga de un molino o clasificador. Este sistema utiliza tecnología láser para proporcionar datos precisos y en tiempo real sobre el tamaño de las partículas.

Tamización: La tamización, también conocida como cribado, es un proceso físico de separación de mezclas de sólidos heterogéneos que utiliza un tamiz o cedazo. El tamiz posee una malla con aberturas de tamaño específico que permite el paso de partículas de menor tamaño, mientras que las partículas de mayor tamaño son retenidas.

Sedimentación: Proceso por el cual las partículas en suspensión en un líquido se depositan en el fondo debido a la gravedad.

Centrifugación: La centrifugación es una técnica que utiliza la fuerza centrífuga para separar partículas de diferentes densidades en una suspensión.

Ganga: Ganga es el término que se utiliza para referirse a los minerales que no tienen valor económico en un yacimiento mineral.

Mena: Se refiere al mineral o conjunto de minerales que se extraen de un yacimiento por ser económicamente explotables.

Pebbles: Los pebbles son fragmentos de roca de tamaño intermedio que resultan del proceso de molienda en molinos SAG. Estos fragmentos son suficientemente grandes para ser rechazados por el tamiz del molino y, a menudo, se reciclan de nuevo al molino para una molienda adicional o se procesan en una trituradora secundaria. Son esenciales para mantener la eficiencia en la molienda semiautógena.

Hidrofobicidad: Es la propiedad de una sustancia de ser repelida por el agua. Las sustancias hidrofóbicas no tienen afinidad por el agua y no se disuelven en ella.

Hidrofilicidad: Se refiere a la propiedad de una sustancia de ser atraída por el agua. Las sustancias hidrofílicas tienen una afinidad por el agua y se disuelven fácilmente en ella.

Blending: Es el proceso de mezclar diferentes tipos de minerales o materiales de diferentes calidades para obtener una mezcla homogénea que cumpla con especificaciones de calidad específicas antes de su procesamiento o venta.

Tareas de Windows: Las tareas de Windows son procesos o aplicaciones que se ejecutan en segundo plano o de forma automática en el sistema operativo Windows, estas tareas pueden ser programado por el usuario o por el sistema operativo.

3 CAPITULO III: HIPÓTESIS Y VARIABLES

3.1 La hipótesis

La implementación de un sistema de recomendaciones para el circuito de Molienda–Clasificación, basado en un modelo de Machine Learning, permitirá optimizar el control operativo, reduciendo la variabilidad del P80 y, con ello, mitigando su impacto negativo en la recuperación de cobre.

3.2 Las Variables

3.2.1 Variables Independientes

Para esta investigación, se dispone de 22 variables independientes correspondientes al circuito de Molienda - Clasificación. Estas variables pueden clasificarse según el equipo asociado y el tipo de control operativo que presentan:

- Control operativo directo: El operador de la sala de control puede manipular directamente la variable.
- Control operativo indirecto: La variable no se puede modificar directamente, pero su valor puede cambiar como consecuencia de ajustar una variable de control directo.
- Variable no manipulable: Su valor depende exclusivamente de las características del mineral y no puede ser alterado por el operador.

Esta clasificación permite entender mejor el grado de influencia que cada variable tiene sobre el proceso.

Tabla 1: Lista de Variables Independientes

Etapas	Equipo	Descripción	um	Control Operativo
MOLIENDA PRIMARIA	MOLINO SAG	Flujo Agua Molino SAG	m3/h	Directo
		Velocidad Molino SAG	RPM	Directo
		Presión Lado Alimentación Molino SAG	kPa	Indirecto

		Presión Lado Descarga Molino SAG	kPa	Indirecto
		Tonelaje Molino SAG	ton/h	Directo
		F80 Molino SAG	inch	No Manipulable
		Potencia Molino SAG	KW	Indirecto
CHANCADO PEBBLES	CHANCADORA PEBBLES	Presión Clamplimg Chancadora de Pebbles	kg/cm2	Indirecto
		Tonelaje Chancadora de Pebbles	tmh/h	Indirecto
		P80 Chancadora de Pebbles	mm	Indirecto
MOLIENDA SECUNDARIA	MOLINO DE BOLAS	Flujo Agua chute Alimentación Molino de Bolas	m3/h	Directo
		Flujo Agua Trommel Molino de Bolas	m3/h	Directo
		Potencia Molino de Bolas	KW	Indirecto
		Presión Lado Alimentación 1 Molino de Bolas	kPa	Indirecto
		Presión Lado Alimentación 2 Molino de Bolas	kPa	Indirecto
		Presión Lado Descarga 1 Molino de Bolas	kPa	Indirecto
		Presión Lado Descarga 2 Molino de Bolas	kPa	Indirecto
		Velocidad Motor 2 Molino de Bolas	RPM	Indirecto
		WorkIndex Molienda	kwh/tm	Indirecto
CLASIFICACIÓN	CAJÓN 01	Nivel Cajón de Bombeo	%	Directo
		Flujo Agua alimento Cajón	m3/h	Directo
	NIDO HIDROCICLONES	Presión Hidrociclones	kPa	Directo

Fuente: Elaboración Propia

En la Tabla 1 se presentan las variables en línea que se tienen en el circuito de Molienda – Clasificación, estas variables son controladas mediante el sistema de control y están asociadas a los diferentes equipos que se tienen en el circuito. Dichas variables son las que se usarán como variables predictoras o independientes para el desarrollo del modelo de Machine Learning.

3.2.2 Variable Dependiente

Para la investigación se cuentan con una sola variable dependiente.

Tabla 2: Lista de Variable Dependiente

Etapas	Equipo	Descripción	um	Control Operativo
CLASIFICACIÓN	HIDROCICLONES	Tamaño de Partícula P80	μm	Indirecto

Fuente: Elaboración Propia

En la Tabla 2 se resume la variable a predecir que es el P80 del Circuito Molienda – Clasificación, esta es una variable que se cuenta en línea y se mide en el overflow de los hidrociclones mediante el equipo analizador de partículas en línea “PSI 300”, el cual se calibra periódicamente para contar con la confiabilidad de los valores obtenidos.

3.3 Operacionalización de variables

La sección de operacionalización de variables se define cómo se miden y analizan las variables del estudio.

Tabla 3: Operacionalización de variables

EQUIPO	Descripción	Dimensión	um
SAG	Control Flujo Agua Proceso Alimentación Molino SAG	Flujo líquido	m3/h
	Control de Velocidad Molino SAG	Velocidad	RPM
	Presión Descanso Principal Lado Alimentación Molino SAG	Presión	KPa
	Presión Descanso Principal Lado Descarga Molino SAG	Presión	KPa
	Control Tonelaje CV003	Flujo másico	tn/h
	Split on Line F80	Tamaño partícula	inch
	Potencia SAG	Potencia	Kw
CHANCADORA PEBBLES	Medición Presión de Clamping Chancadora de Pebbles	Presión	kg/cm2
	Tonelaje Faja Pebbles	Flujo Másico	tn/h
	Split on Line Faja #7 F80	Tamaño partícula	mm
BOLAS	Agua chute de alimentación molino de bolas	Flujo líquido	m3/h
	Agua de limpieza al trommel molino de bolas	Flujo líquido	m3/h
	Potencia Ball Mill	Potencia	Kw

	Medición Presión Descanso Principal Lado Alimentación Molino Bolas	Presión	KPa
	Medición Presión Descanso Principal Lado Alimentación Molino Bolas	Presión	KPa
	Medición Presión Descanso Principal Lado Descarga Molino Bolas	Presión	KPa
	Medición Presión Descanso Principal Lado Descarga Molino Bolas	Presión	KPa
	Velocidad Motor 2 - Molino Bolas	Velocidad	RPM
	WorkIndex de Molienda	Indice de Dureza	kwh/tn
CAJÓN 01	Control Nivel Cajón Bombeo 310-PB-001	Porcentaje de nivel	%
	Agua que se alimenta al cajón 01	Flujo líquido	m3/h
NIDO	Medición Presión Batería Hidrociclones	Presión	KPa
HIDROCICLONES	PSI 300 Molienda - Tamaño de Particula P80	Tamaño particula	µm

Fuente: Elaboración Propia

4 CAPITULO IV: MATERIALES Y MÉTODOS

4.1 Área de estudio

La presente investigación se realizó en las oficinas de la Superintendencia del Centro Integrado de Operaciones del Área de Procesos de la unidad minera Cerro Corona, ubicada en el distrito y provincia de Hualgayoc, departamento de Cajamarca.

4.2 Diseño de investigación

Diseño no experimental, transversal – relacional ya que describe la relación de variables operativas del circuito de Molienda - Clasificación en intervalos de tiempos determinados para la predicción del P80 y así generar recomendaciones en línea de las variables operativas más influyentes.

4.3 Población y Muestra

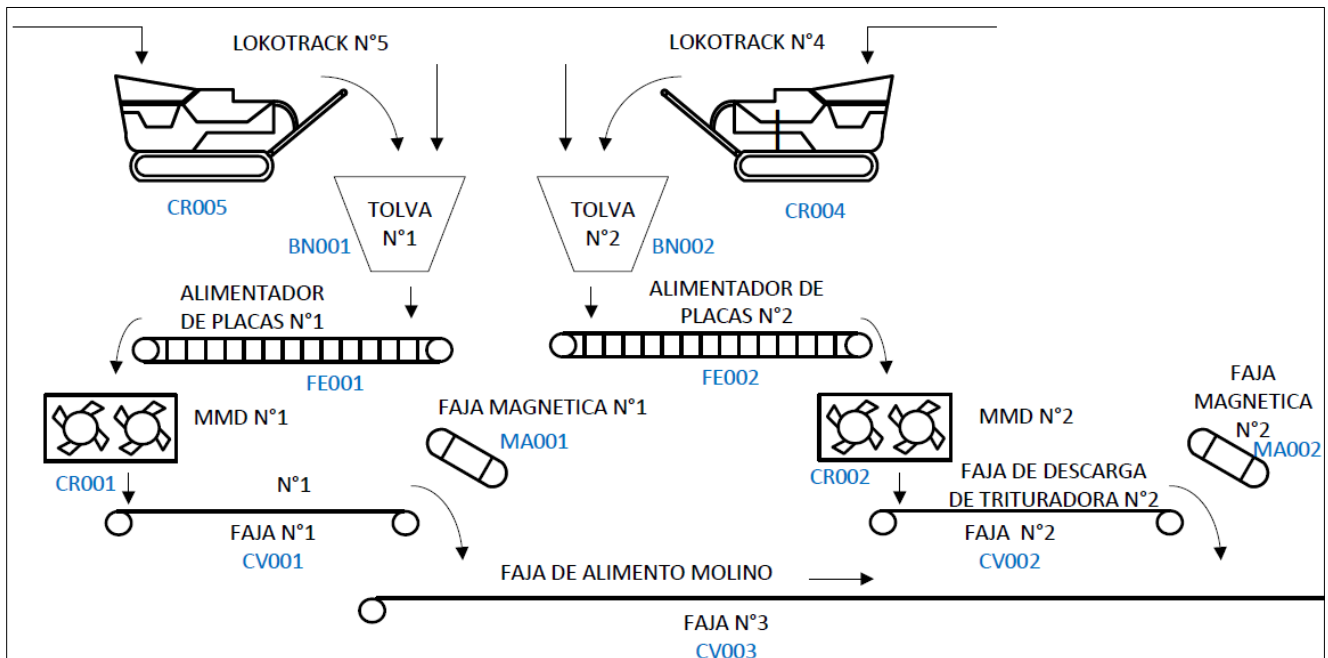
La población que aborda esta investigación comprende las variables en línea que se tienen en el circuito de Molienda – Clasificación, para realizar el desarrollo del modelo se considera data de Enero a Agosto del 2023. La puesta en línea del modelo y sistema de recomendaciones se realizarán con data en línea cada 5 minutos.

4.4 Procedimiento, técnicas e instrumentos de recolección de información

En el circuito de Molienda Clasificación se identifican todas las variables en línea que se tienen en el PI System con el fin de poder contar con la totalidad de información y luego evaluar cuales son las más importantes que nos ayudan a obtener el modelo.

También es necesario tener un conocimiento pleno del circuito de la planta concentradora, para entender cómo se relacionan las variables de control y como estas afectan al producto final.

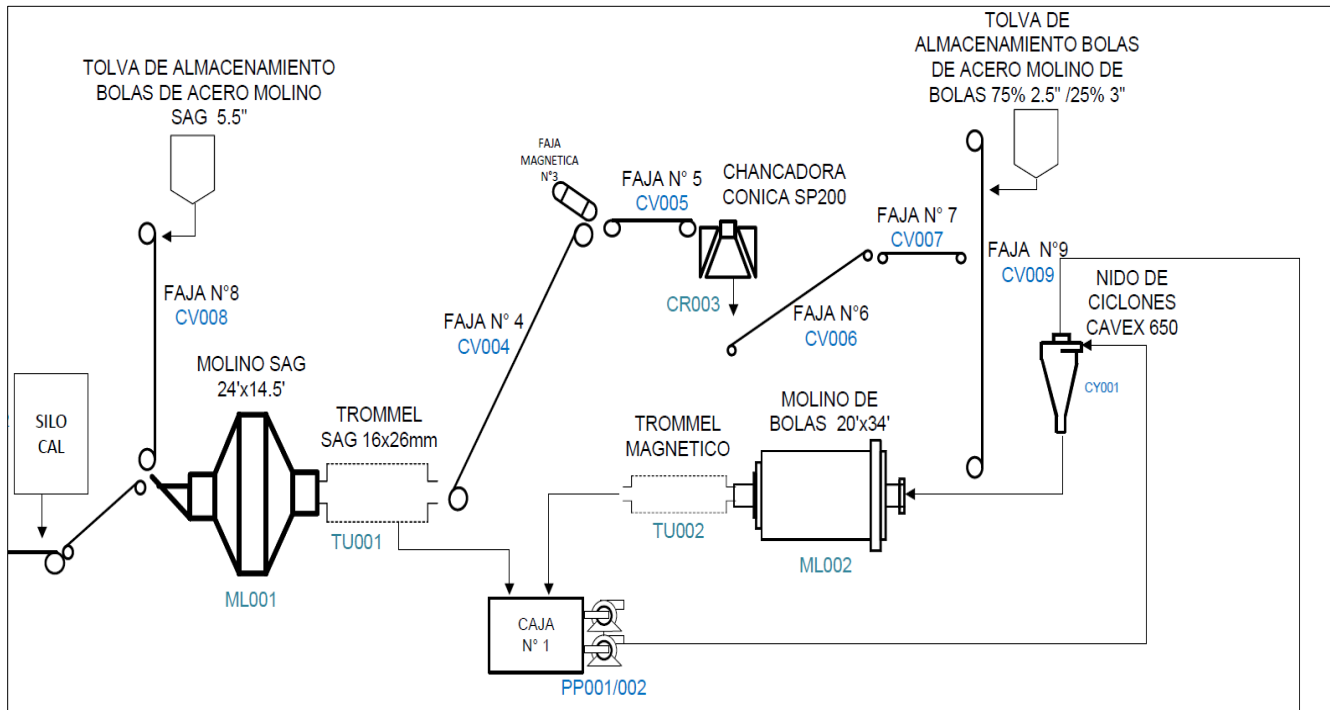
Figura 5: Flowsheet Proceso de Chancado



Fuente: Gold Fields La Cima

En la Figura 5, se muestra el proceso de Chancado, en el cual el mineral que ingresa a la planta proviene de cinco stocks que se alimenta a la planta en diferentes proporciones según el blending del turno. La planta opera con dos líneas que se alternan dependiendo de las condiciones operativas. Para el chancado primario, se utilizan chancadoras móviles de mandíbula "Lokotrack", y para el chancado secundario, se emplean chancadoras de rodillos. El mineral reducido en esta etapa alimenta la faja transportadora N°3, que alimenta el mineral al molino SAG.

Figura 6: Flowsheet Proceso de Molienda



Fuente: Gold Fields La Cima

El circuito de molienda es del tipo SABC-B el cual se muestra en la Figura 6, donde el primer equipo es un molino SAG. En este molino, el Trommel clasifica los Pebbles, enviándolos a la chancadora cónica SP200 para reducir su tamaño antes de ser alimentados al Molino de Bolas. La pulpa descargada del molino SAG se alimenta al Cajón de Bombeo N°1, que también recibe la pulpa del Molino de Bolas. Desde el Cajón de Bombeo N°1, la pulpa es bombeada al Nido de Hidrociclones, donde se clasifica por tamaño. Dependiendo de la clasificación, la pulpa se envía al Circuito de Flotación o vuelve al Molino de Bolas para una etapa de molienda adicional.

4.4.1 Gestión de la Data

Se empieza identificando todos los tags del circuito de Molienda - Clasificación que se tienen en línea en el PI System. Se identifican 31 variables en línea las cuales están clasificadas al equipo asociado.

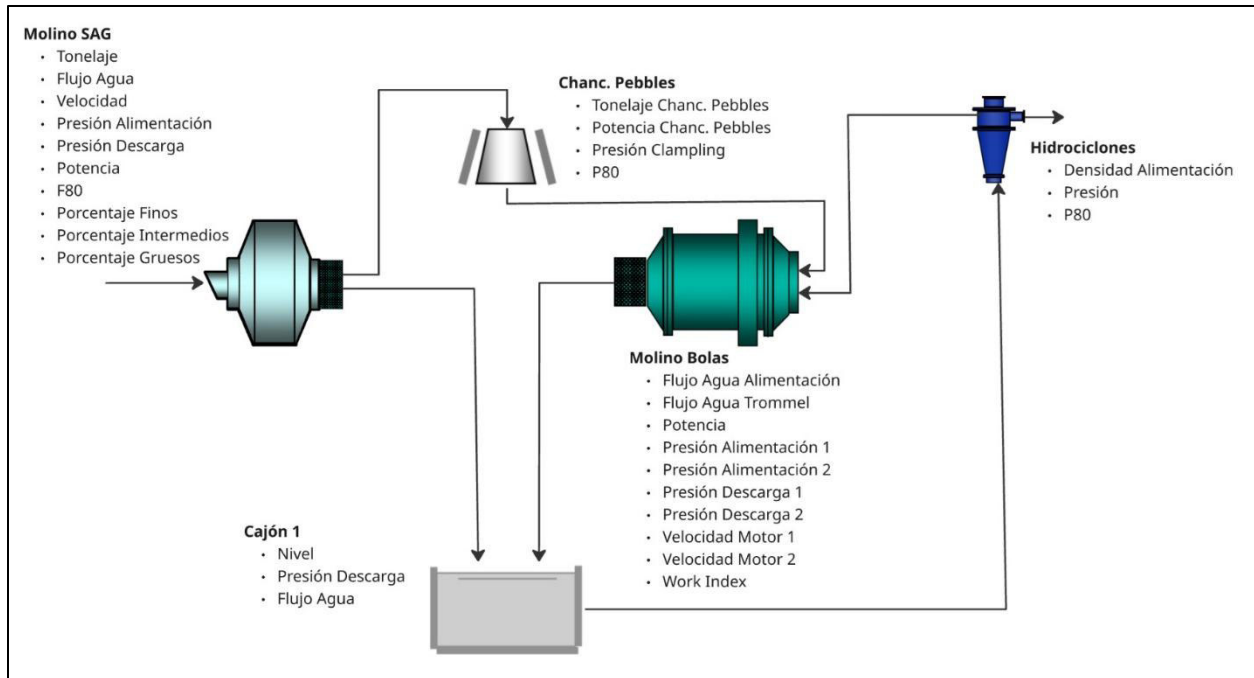
Tabla 4: Tags del Circuito Molienda Clasificación

Equipo	Descripción	um
MOLINO SAG	Porcentaje Solidos Molino SAG	%
	Flujo Agua Molino SAG	m3/h
	Velocidad Molino SAG	RPM
	Presión Lado Alimentación Molino SAG	kPa
	Presión Lado Descarga Molino SAG	kPa
	Tonelaje Molino SAG	ton/h
	F80 Molino SAG	inch
	Porcentaje Material Fino	%
	Porcentaje Material Intermedio	%
	Porcentaje Material Grueso	%
	Potencia Molino SAG	KW
CHANCADORA PEBBLES	Presión Clamplng Chancadora de Pebbles	kg/cm2
	Potencia Chancadora de Pebbles	Kw
	Tonelaje Chancadora de Pebbles	tmh/h
	P80 Chancadora de Pebbles	mm
MOLINO DE BOLAS	Flujo Agua chute Alimentación Molino de Bolas	m3/h
	Flujo Agua Trommel Molino de Bolas	m3/h
	Potencia Molino de Bolas	KW
	Presión Lado Alimentación 1 Molino de Bolas	kPa
	Presión Lado Alimentación 2 Molino de Bolas	kPa
	Presión Lado Descarga 1 Molino de Bolas	kPa
	Presión Lado Descarga 2 Molino de Bolas	kPa
	Velocidad Motor 2 Molino de Bolas	RPM
	Velocidad Motor 1 Molino de Bolas	RPM
	WorkIndex Molienda	kwh/tm
CAJÓN 01	Nivel Cajón de Bombeo	%
	Presión Descarga Cajón	kPa
	Flujo Agua alimento Cajón	m3/h
HIDROCICLONES	Densidad Alimentación Hidrociclones	Kg/m3
	Presión Hidrociclones	kPa
	P80	µm

Fuente: Elaboración Propia

La Figura 7, corresponde a un esquema operativo en el que se identifican las principales variables asociadas a cada uno de los equipos involucrados en el proceso. Estas variables operativas son registradas en tiempo real mediante el sistema de control de planta, permitiendo una supervisión continua del comportamiento del proceso.

Figura 7: Diagrama de Variables Operativas



Fuente: Elaboración Propia

Posteriormente, se crearon plantillas de Excel para la extracción de datos mediante el PI Data Link:

- Tipo de Extracción: Sample.
- Intervalo de Extracción: 5 minutos.

Obteniendo 230936 filas en un periodo de tiempo desde 16 de Enero hasta el 16 de Agosto del 2023.

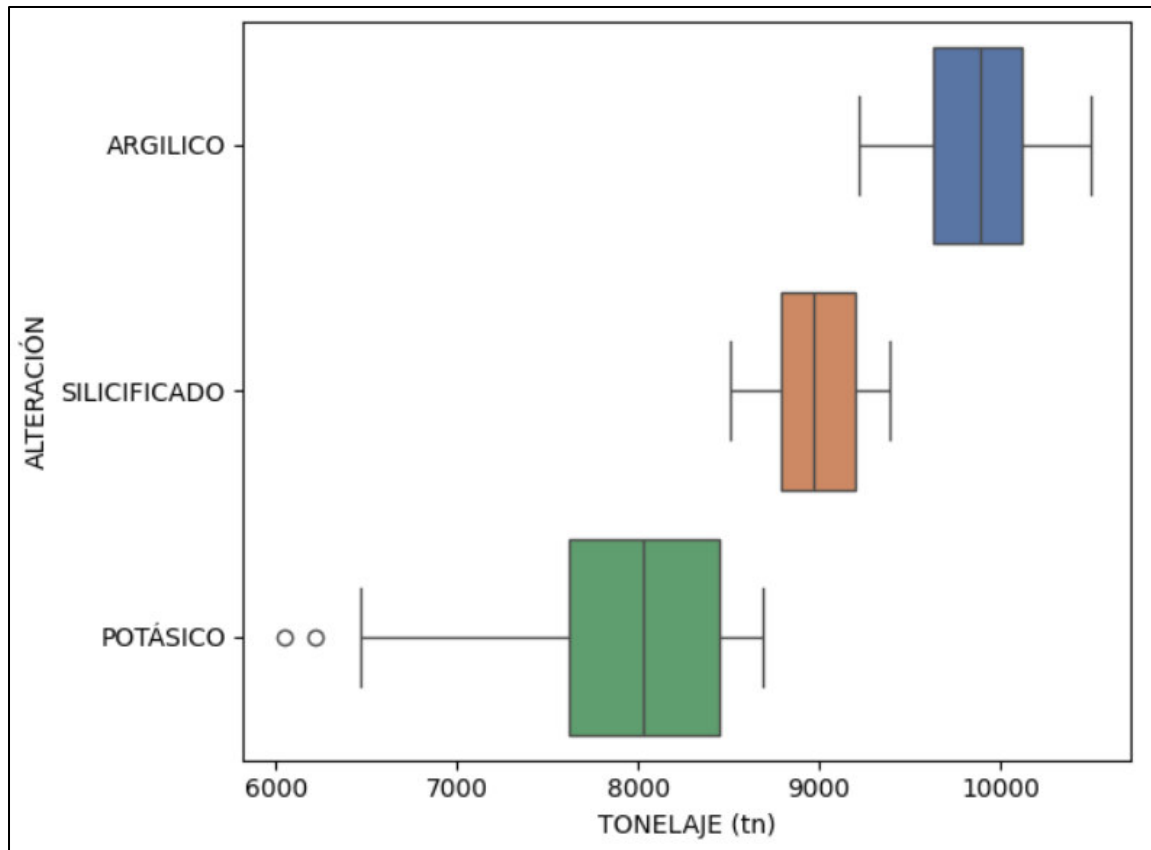
4.5 Análisis estadístico

Se realiza un análisis de las alteraciones, el P80 y Recuperación de Cobre, luego se realiza el análisis exploratorio de datos para el modelamiento, la búsqueda del mejor modelo, el desarrollo del sistema de recomendaciones y la puesta en línea del sistema.

4.6 Análisis por Alteraciones Mineralógicas

Cada tipo de alteración hidrotermal presenta un comportamiento metalúrgico distinto al ser procesado en la planta concentradora. Tal como se observa en la Figura 8, existen diferencias notables en el tonelaje procesado por turno según el tipo de alteración mineralógica. La alteración Argílica, al ser la de menor competencia, permite procesar en promedio 10,000 toneladas por turno (12 horas). En cambio, la alteración Silicificada, de competencia media, alcanza un procesamiento promedio de 9,000 toneladas por turno, mientras que la alteración Potásica, caracterizada por una mayor competencia, permite procesar alrededor de 8,000 toneladas por turno. Estas diferencias reflejan cómo la naturaleza mineralógica influye directamente en la capacidad operativa de la planta.

Figura 8: Tonelaje Procesado por turno por alteración mineralógica

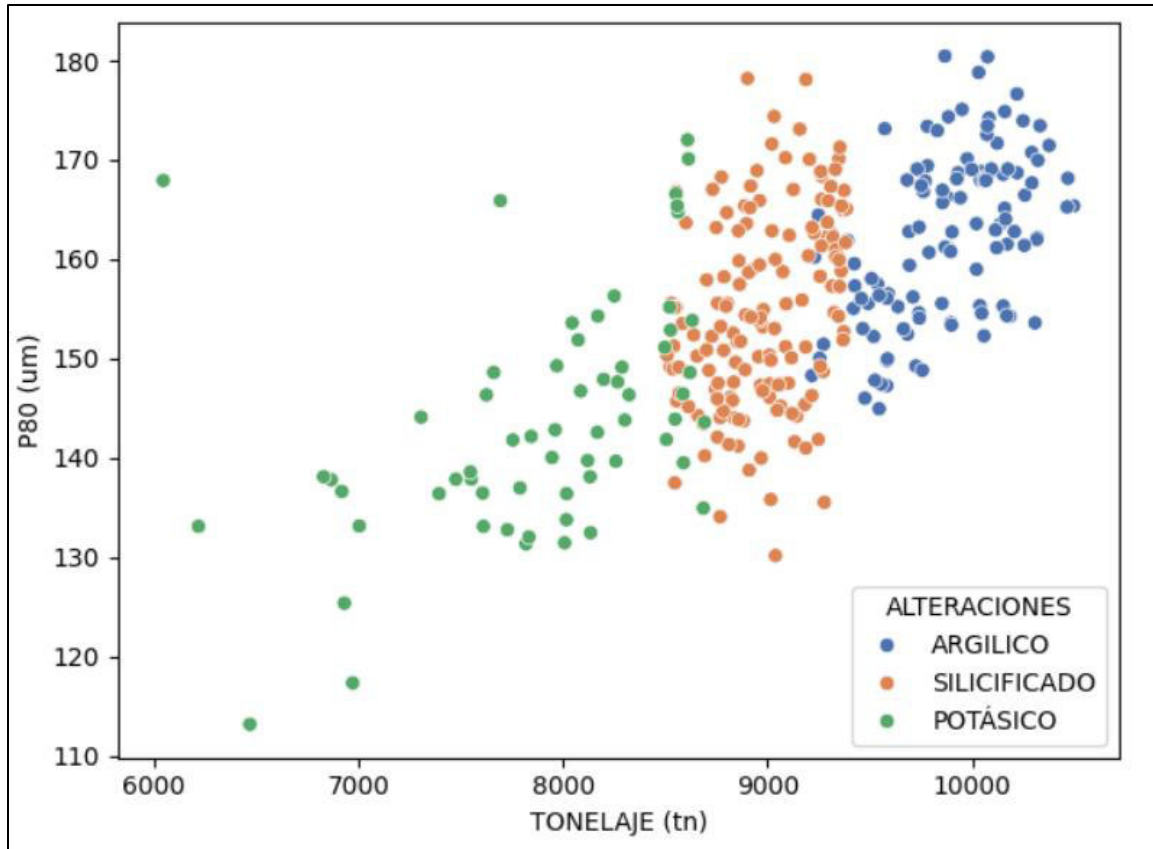


Fuente: Elaboración Propia

La información sobre las alteraciones hidrotermales se obtiene por turno de operación, en función de la caracterización de los blendings realizados durante ese periodo. A diferencia de otras variables operativas del proceso, esta no se registra en línea ni con alta frecuencia (por ejemplo, cada 5 minutos). Sin embargo, es posible inferir el tipo de alteración mineralógica a partir del tonelaje procesado, ya que esta variable sí se encuentra disponible en línea y en tiempo real. Esta inferencia resulta valiosa, ya que el tonelaje procesado está estrechamente relacionado con el tipo de alteración y, por lo tanto, constituye una variable relevante para la predicción del

P80 mediante modelos de Machine Learning. Donde existe una relación directa entre el tonelaje procesado y el P80 se aprecia claramente en la Figura 9.

Figura 9: P80 por Tonelaje Procesado

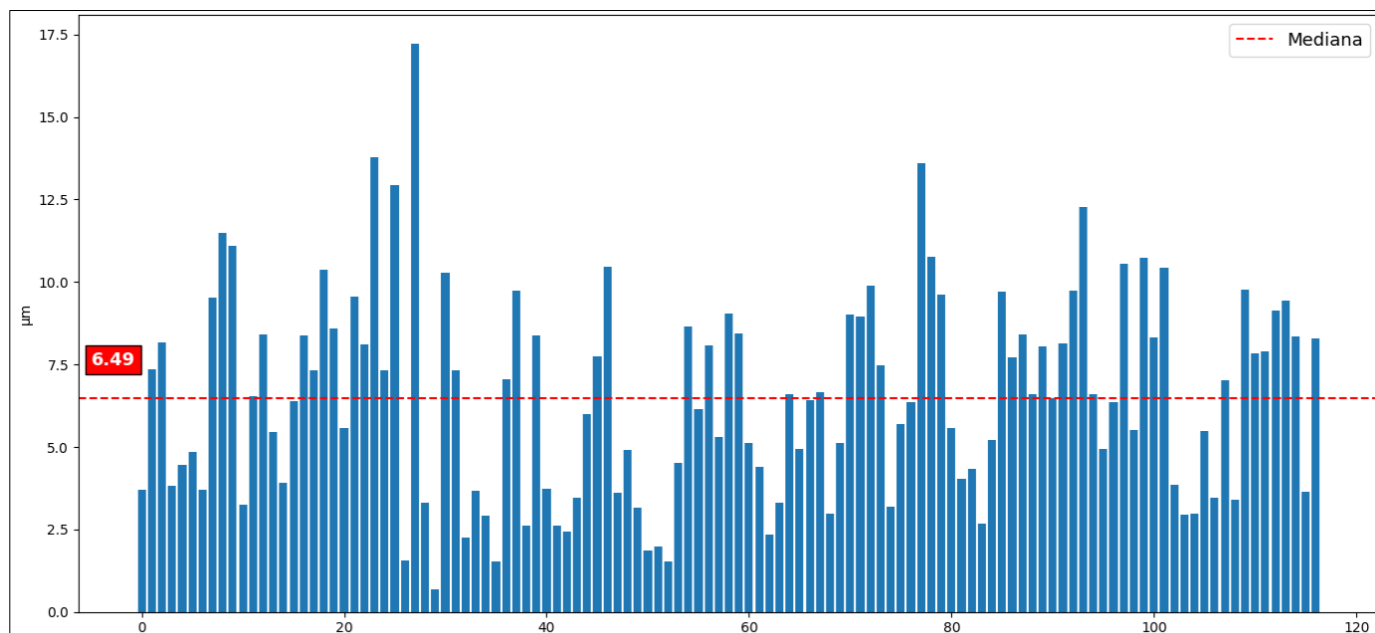


Fuente: Elaboración Propia

4.6.1 Variabilidad del P80

Dentro del periodo de análisis de la investigación se evalúa cual ha sido la variabilidad del P80. Donde la métrica usada es la desviación estándar. Según el gráfico adjunto se evidencia una desviación estándar en promedio de P80 de 6.49 μm .

Figura 10: Desviaciones Estándar históricos de P80



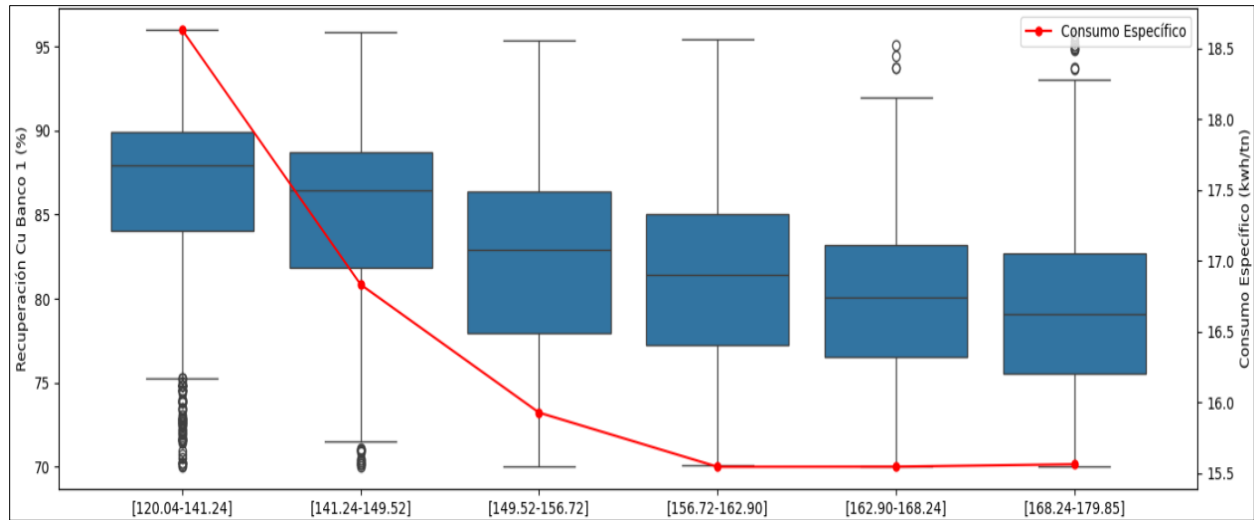
Fuente: Elaboración Propia

En la Figura 10, se presenta los valores de las desviaciones estándar por día del P80, esto con el objetivo de tener una línea base de la variabilidad del P80 para luego poder hacer las comparaciones de los resultados de la optimización del circuito.

4.6.2 Efecto del P80 en la Recuperación de Cobre

Se realiza un análisis histórico del efecto del P80 en la recuperación de Cobre del Banco 1 del circuito Rougher. En la Figura 11, se evidencia que a rangos de P80 más pequeños es más beneficioso para la Recuperación de Cobre, sin embargo, esto requiere de un mayor Consumo Específico del circuito de Molienda. Se propone un rango de P80 óptimo entre 148.42 μm y 156.72 μm que representa un aumento en la recuperación de Cobre con un Consumo Específico del circuito de Molienda - Clasificación adecuado.

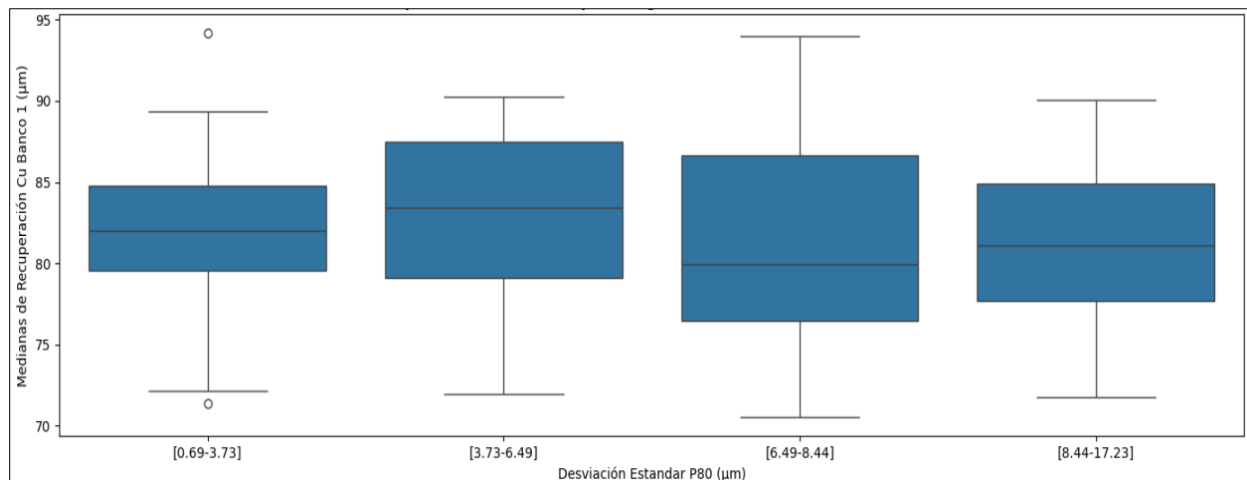
Figura 11: Recuperación de Cobre y Consumo Específico por rangos de P80



Fuente: Elaboración Propia

De la misma manera, en la Figura 12, se evidencia un rango de desviación estándar del P80, donde favorezca la recuperación de Cobre según la data histórica. Se propone un rango de desviación estándar del P80 entre 3.73 μm y 6.49 μm como óptimo para que no afecte considerablemente a la recuperación de Cobre.

Figura 12: Recuperación de Cobre por Rangos de Desviación Estándar de P80



Fuente: Elaboración Propia

4.6.3 Análisis Exploratorio de datos

Con el análisis exploratorio de datos (EDA) se busca para comprender y familiarizarse con los datos antes de aplicar técnicas estadísticas más avanzadas. El objetivo principal del EDA es descubrir patrones, tendencias y relaciones en los datos mediante la visualización, la limpieza y el resumen de los datos.

4.6.3.1 Fechas de Paradas de Planta

Durante el Periodo de análisis de extracción de la data se considera las paradas de planta por mantenimientos preventivos y correctivos.

Tabla 5: Fechas de Mantenimientos en el Circuito Molienda - Clasificación

FECHA INICIAL	CONDICION INICIO	FECHA FINAL	CONDICION FINAL	DURACION	TIPO DE PARADA
31-ene-2023	Detenido	31-ene-2023	Funcionando	00 : 21 : 16	Mantenimiento Correctivo
13-feb-2023	Detenido	15-feb-2023	Funcionando	60 : 07 : 41	Mantenimiento Preventivo
15-feb-2023	Detenido	15-feb-2023	Funcionando	00 : 30 : 35	Mantenimiento Correctivo
17-feb-2023	Detenido	17-feb-2023	Detenido	00 : 18 : 33	Mantenimiento Correctivo
17-feb-2023	Detenido	17-feb-2023	Funcionando	00 : 18 : 33	Mantenimiento Correctivo
25-feb-2023	Detenido	25-feb-2023	Funcionando	04 : 15 : 59	Mantenimiento Correctivo
26-feb-2023	Detenido	26-feb-2023	Funcionando	00 : 48 : 13	Mantenimiento Correctivo

08-mar-2023	Detenido	08-mar-2023	Funcionando	00 : 52 : 48	Mantenimiento Correctivo
19-mar-2023	Detenido	19-mar-2023	Detenido	00 : 17 : 25	Mantenimiento Correctivo
19-mar-2023	Detenido	19-mar-2023	Funcionando	00 : 52 : 13	Mantenimiento Correctivo
31-mar-2023	Detenido	01-abr-2023	Funcionando	05 : 13 : 11	Mantenimiento Correctivo
10-abr-2023	Detenido	10-abr-2023	Funcionando	16 : 44 : 38	Mantenimiento Preventivo
20-abr-2023	Detenido	20-abr-2023	Funcionando	01 : 13 : 07	Mantenimiento Correctivo
30-abr-2023	Detenido	30-abr-2023	Funcionando	00 : 55 : 05	Mantenimiento Correctivo
13-may-2023	Detenido	14-may-2023	Funcionando	29 : 27 : 04	Mantenimiento Preventivo
16-may-2023	Detenido	16-may-2023	Funcionando	00 : 36 : 27	Mantenimiento Correctivo
18-may-2023	Detenido	18-may-2023	Funcionando	01 : 10 : 36	Mantenimiento Correctivo
07-jun-2023	Detenido	07-jun-2023	Funcionando	00 : 45 : 42	Mantenimiento Correctivo
11-jun-2023	Detenido	11-jun-2023	Funcionando	02 : 08 : 14	Mantenimiento Correctivo
18-jun-2023	Detenido	18-jun-2023	Funcionando	00 : 39 : 19	Mantenimiento Correctivo

22-jun-2023	Detenido	22-jun-2023	Funcionando	00 : 30 : 37	Mantenimiento Correctivo
22-jun-2023	Detenido	22-jun-2023	Funcionando	02 : 52 : 49	Mantenimiento Correctivo
23-jun-2023	Detenido	23-jun-2023	Funcionando	02 : 38 : 07	Mantenimiento Correctivo
26-jun-2023	Detenido	28-jun-2023	Funcionando	51 : 09 : 05	Mantenimiento Preventivo
26-jul-2023	Detenido	26-jul-2023	Funcionando	00 : 47 : 51	Mantenimiento Correctivo
16-ago-2023	Detenido	16-ago-2023	Funcionando	15 : 08 : 47	Mantenimiento Preventivo

Fuente: Elaboración Propia

En la Tabla 5, se muestra las fechas de paradas de planta, las cuales generan data no representativa, por lo tanto, de la data extraída se filtrará estos tiempos detallados en la tabla anterior.

4.6.3.2 Valores no numéricos

En la Tabla 6 se muestra la cantidad de datos con valores no numéricos y su porcentaje.

Tabla 6: Cantidad de Valores numéricos

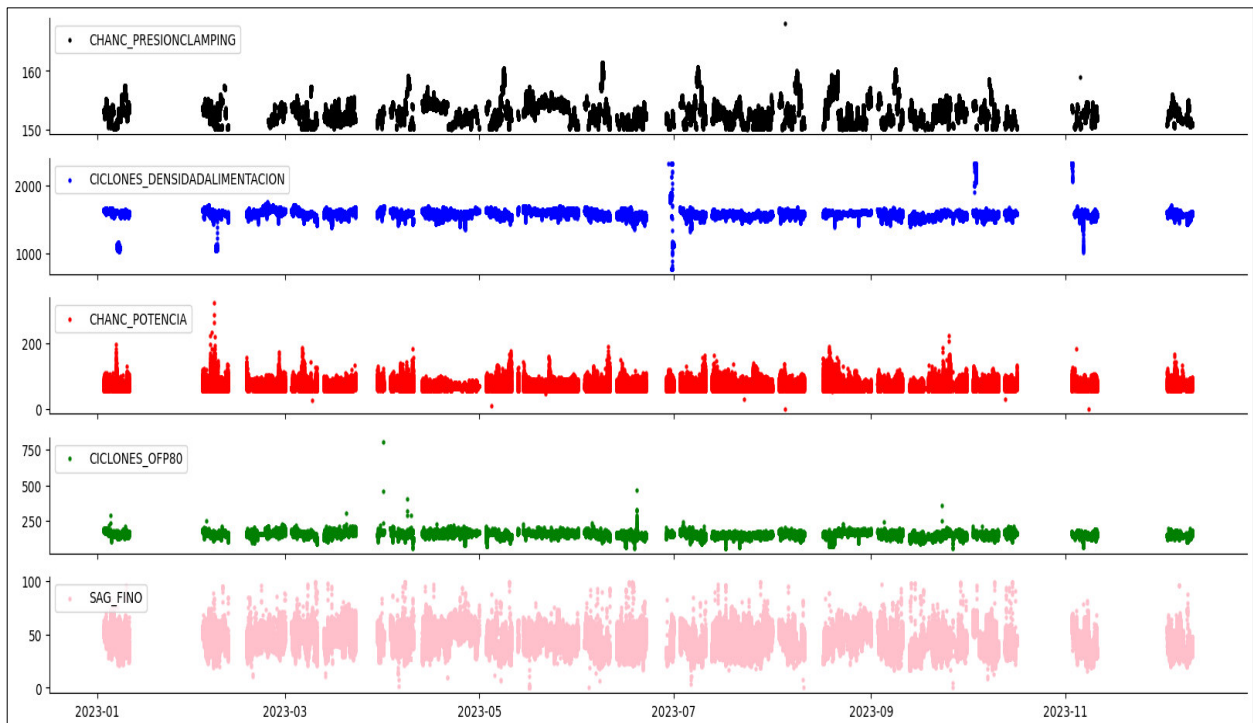
Variables	Cantidad	%
CHANC_PRESIONCLAMPING	8208	3.57
CICLONES_DENSIDADALIMENTACION	1560	0.68
CHANC_POTENCIA	386	0.17
CICLONES_OFP80	386	0.17
SAG_FINO	386	0.17
SAG_GRUESO	386	0.17
SAG_INTER	386	0.17

CHANC P80	386	0.17
BALL WI	381	0.17
CAJON NIVEL	381	0.17
CAJON PRESIONDESCARGA	381	0.17
CHANC TON	381	0.17
BOMBA VELOCIDAD2	381	0.17
CICLONES PRESIONPSI	381	0.17
BOMBA VELOCIDAD1	381	0.17
CICLONES PRESIONKPA	125	0.05
BALL FLUJOAGUA	64	0.03
SAG PRESIONCARGA	11	0.00
SAG F80	5	0.00

Fuente: Elaboración Propia

Considerando un gráfico de serie de tiempo de las variables con mayor contenido de valores no numéricos se visualiza en la Figura 13, que los valores no numéricos es decir los que pierden la continuidad de los puntos son en las mismas fechas.

Figura 13: Variables con valores no numéricos



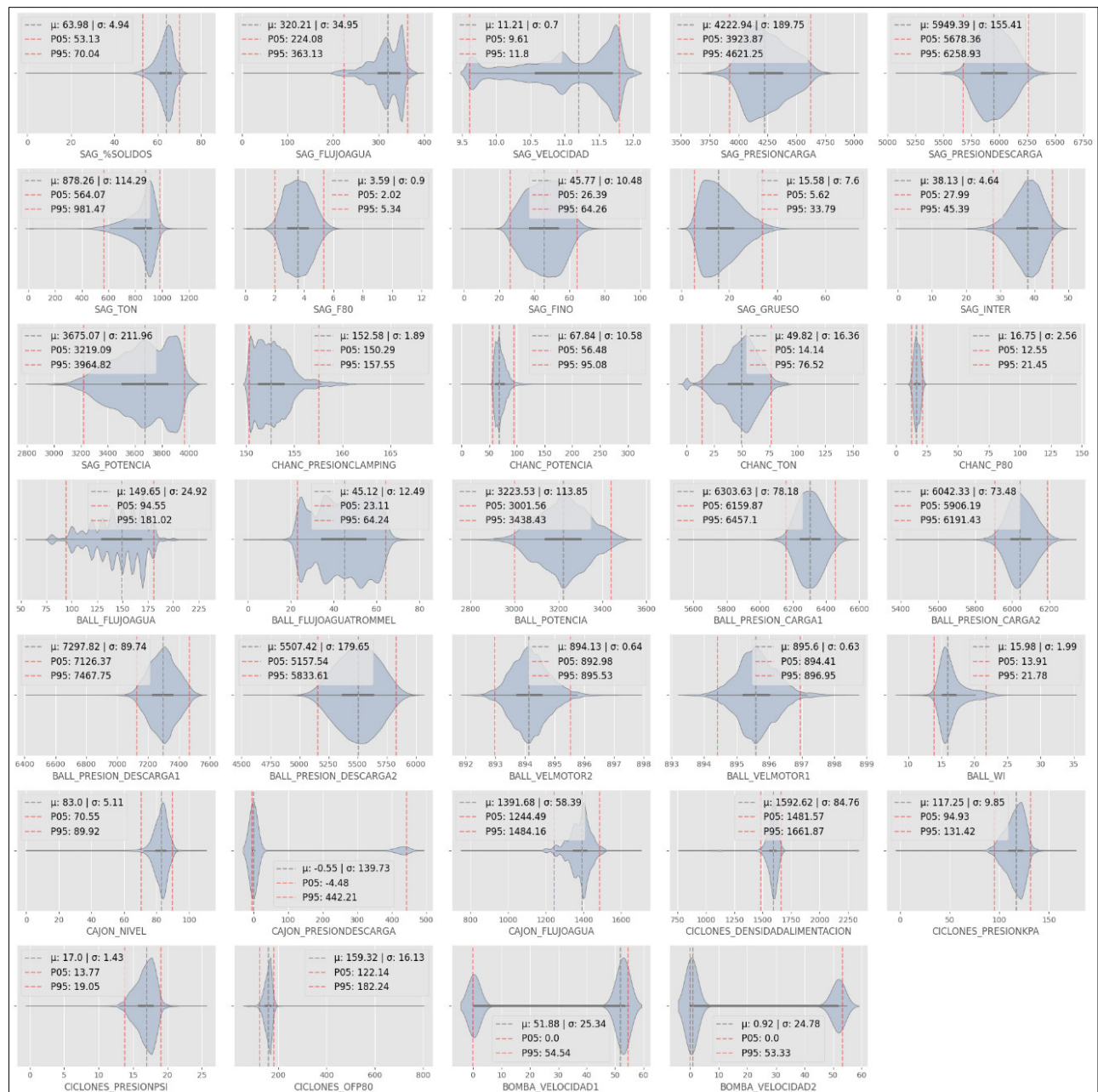
Fuente: Elaboración Propia

Considerando que los valores no numéricos se presentan en el mismo periodo de tiempo y que la variable con mayor cantidad de valores no numéricos representan el 3.57% de la data total, se decide eliminar las filas de estos valores ya que representan valores no representativos en condiciones de operación normal.

4.6.3.3 Imputación de Datos

Para realizar la imputación de datos es necesario visualizar la distribución de datos por cada variable, con el fin de definir una técnica de imputación en caso de que se encuentren valores atípicos. Para ello en la Figura 14, se muestran gráficos de violín por cada variable, mostrando la mediana, percentil 2.5 y percentil 97.5.

Figura 14: Gráficos de Violín de variables del circuito Molienda - Clasificación



Fuente: Elaboración Propia

Entendiendo la distribución de los datos de cada variable se hace una limpieza por límites operacionales, los cuales son valores que se pueden tener en condiciones operativas según el conocimiento del proceso, cuyos límites por cada variable son mostrados en la Tabla 7.

Tabla 7: Límites Operativos

Descripción	um	Límite Mínimo	Límite Máximo
Porcentaje Solidos Molino SAG	%	40	80
Flujo Agua Molino SAG	m3/h	200	400
Velocidad Molino SAG	RPM	9.5	12
Presión Lado Alimentación Molino SAG	KPa	3700	4750
Presión Lado Descarga Molino SAG	KPa	5400	6500
Tonelaje Molino SAG	tn/h	400	1000
F80 Molino SAG	inch	0	8
Porcentaje Material Fino	%	0	80
Porcentaje Material Intermedio	%	20	50
Porcentaje Material Grueso	%	0	60
Potencia Molino SAG	Kw	3000	4200
Presión Clamplimg Chancadora de Pebbles	kg/cm2	150	160
Potencia Chancadora de Pebbles	Kw	50	120
Tonelaje Chancadora de Pebbles	tnh/h	0	100
P80 Chancadora de Pebbles	mm	0	50
Flujo Agua chute Alimentación Molino de Bolas	m3/h	75	225
Flujo Agua Trommel Molino de Bolas	m3/h	10	70
Potencia Molino de Bolas	Kw	2900	3550
Presión Lado Alimentación 1 Molino de Bolas	KPa	5800	6600
Presión Lado Alimentación 2 Molino de Bolas	KPa	5800	6300
Presión Lado Descarga 1 Molino de Bolas	KPa	7000	7600
Presión Lado Descarga 2 Molino de Bolas	KPa	5000	6000
Velocidad Motor 2 Molino de Bolas	RPM	892	897
Velocidad Motor 1 Molino de Bolas	RPM	892	900
WorkIndex Molienda	Kwh/tn	10	30
Nivel Cajón de Bombeo	%	40	100
Presión Descarga Cajón	KPa	0	100
Flujo Agua alimento Cajón	m3/h	1000	1600
Densidad Alimentación Hidrociclones	Kg/m3	1250	2000
Presión Hidrociclones	KPa	50	150
P80	µm	0	300

Fuente: Elaboración Propia

Luego de filtrar los datos por los límites operativos se hace una limpieza de datos por los percentiles 2.5 y 97.5, lo cual se busca eliminar los valores más extremos de su distribución ya que son valores poco frecuentes y pueden ser considerados como valores anómalos a una condición operativa normal.

4.6.3.4 Análisis de Correlación

Es una medida que indica el grado de relación lineal entre dos variables, esta relación puede ser positiva, negativa o nula, esta medida se mide entre -1 y 1, valores cercanos a 1 significa una relación positiva, en cambio valores cercanos a -1 significa una relación negativa.

4.6.3.4.1 Correlación entre Variables Predictoras

Tabla 8: Correlaciones entre Variables Predictoras

Variable 1	Variable 2	Correlación
BOMBA_VELOCIDAD1	BOMBA_VELOCIDAD2	-0.98
SAG_F80	SAG_GRUESO	0.94
SAG_FINO	SAG_GRUESO	-0.91
SAG_F80	SAG_FINO	-0.90
SAG_TON	BALL_WI	-0.89
BALL_PRESION_CARGA1	BALL_PRESION_CARGA2	0.87
BALL_PRESION_CARGA2	BALL_PRESION_DESCARGA1	0.86
BALL_PRESION_CARGA1	BALL_PRESION_DESCARGA1	0.86
SAG_%SOLIDOS	SAG_TON	0.79
SAG_PRESIONCARGA	SAG_PRESIONDESCARGA	0.76
SAG_FINO	SAG_INTER	-0.74
SAG_%SOLIDOS	BALL_WI	-0.63
SAG_TON	CICLONES_OFP80	0.61
CICLONES_PRESIONKPA	CICLONES_OFP80	0.56
SAG_VELOCIDAD	SAG_PRESIONDESCARGA	0.55
SAG_TON	CICLONES_PRESIONKPA	0.55
BALL_POTENCIA	CAJON_FLUJOAGUA	-0.53
BALL_POTENCIA	CICLONES_OFP80	-0.52
SAG_F80	SAG_INTER	0.51
SAG_%SOLIDOS	CICLONES_OFP80	0.51
BALL_POTENCIA	CICLONES_PRESIONKPA	-0.50
SAG_PRESIONDESCARGA	BALL_WI	0.50
BALL_FLUJOAGUA	CAJON_FLUJOAGUA	0.50
BALL_FLUJOAGUA	BALL_POTENCIA	-0.50
SAG_PRESIONDESCARGA	SAG_TON	-0.50

Fuente: Elaboración Propia

4.6.3.4.2 Correlaciones de Variable a Predecir

Tabla 9: Correlaciones entre Variables Predictoras y Variables a Predecir

Variables Predictoras	Correlación
SAG_TON	0.61
CICLONES_PRESIONKPA	0.56
BALL_POTENCIA	-0.52
SAG_%SOLIDOS	0.51
SAG_VELOCIDAD	-0.43
CAJON_FLUJOAGUA	0.41
SAG_PRESIONDESCARGA	-0.39
CICLONES_DENSIDADALIMENTACION	0.38
SAG_GRUESO	-0.38
BALL_WI	-0.35
SAG_F80	-0.35
BALL_FLUJOAGUA	0.33
CAJON_NIVEL	0.32
SAG_FINO	0.32
SAG_PRESIONCARGA	-0.28
SAG_POTENCIA	-0.25
BALL_PRESION_CARGA2	0.23
BALL_PRESION_DESCARGA2	0.22
CAJON_PRESIONDESCARGA	0.2
SAG_FLUJOAGUA	-0.18
BALL_PRESION_DESCARGA1	0.16
BALL_PRESION_CARGA1	0.16
SAG_INTER	-0.11
BALL_VELMOTOR1	0.08
CHANC_POTENCIA	-0.06
CHANC_TON	0.06
BOMBA_VELOCIDAD1	0.05
BALL_FLUJOAGUATROMMEL	0.05
BALL_VELMOTOR2	0.03
CHANC_P80	-0.02
BOMBA_VELOCIDAD2	-0.01
CHANC_PRESIONCLAMPING	0

Fuente: Elaboración Propia

En la Tabla 8 y 9, se da a conocer una primera aproximación de la dependencia entre variables predictoras y la variable a predecir. Lo cual indica que tan correlacionadas están las variables predictoras y cuáles son las variables que pueden explicar la variable a predecir.

4.6.3.5 Selección de Variables

Se utilizó el VIF (Factor de inflación de la varianza) con el cual podemos identificar multicolinealidad entre las variables que se van a utilizar en el modelo para predecir la recuperación de Cobre (que una variable pueda ser explicada por una o por otras variables).

Si el valor del estadístico VIF es superior o igual a 10, sugiere que existe una multicolinealidad fuerte y la recomendación es eliminar esa variable dentro del análisis.

Según el análisis se dejan considerar las siguientes variables: BALL_VELMOTOR1, SAG_FINO,SAG_GRUESO,SAG_INTER,CICLONES_OFP80,CICLONES_DENSIDADALIMENTACION.

Tabla 10: VIF de las variables predictoras

Variables	VIF
BALL_PRESION_CARGA2	7.11
BALL_WI	7.00
BALL_PRESION_DESCARGA1	6.20
SAG_PRESIONDESCARGA	5.75
CICLONES_PRESIONKPA	4.86
CHANC_PRESIONCLAMPING	4.82
SAG_TON	4.76
SAG_PRESIONCARGA	4.58
SAG_FLUJOAGUA	4.42
BALL_PRESION_CARGA1	3.60
BALL_FLUJOAGUA	3.46
SAG_VELOCIDAD	3.41
CAJON_FLUJOAGUA	3.04
SAG_POTENCIA	2.90
BALL_POTENCIA	2.62
BALL_VELMOTOR2	1.63
CHANC_P80	1.62
CHANC_TON	1.57

BALL_FLUJOAGUATROMMEL	1.35
SAG_F80	1.24
BALL_PRESION_DESCARGA2	1.20
CHANC_POTENCIA	1.20
CAJON_NIVEL	1.18

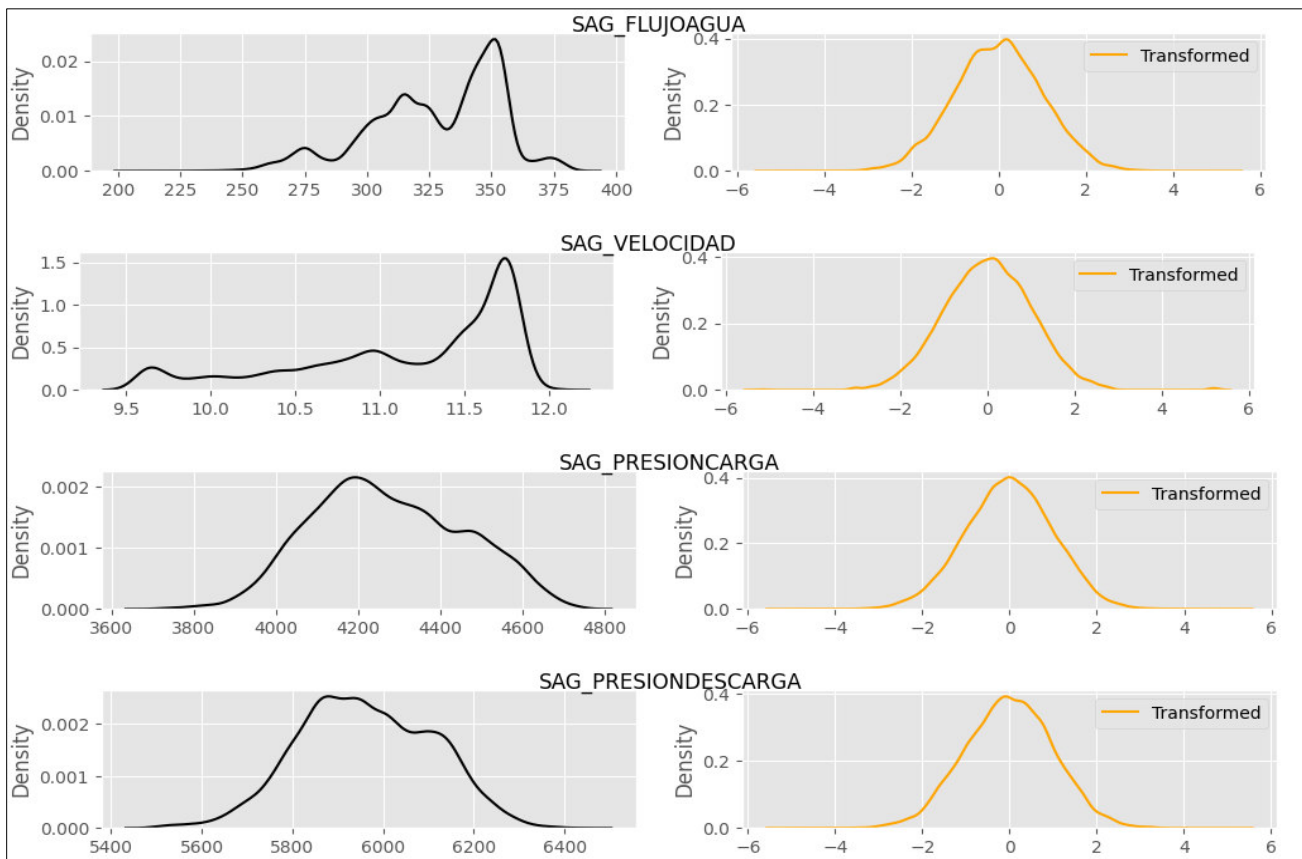
Fuente: Elaboración Propia

En la Tabla 10, se muestra los valores VIF de las variables predictoras que no muestran multicolinealidad y son consideradas para el desarrollo del modelo según criterios de correlaciones entre variables y Multicolinealidad.

4.6.3.6 Estandarización de los datos

Para hacer a las variables comparables se estandarizan los datos, se usa la técnica de “Escala Robusta” la cual transforma los valores de las variables para seguir una distribución normal.

Figura 15: Estandarización de variables independientes



Fuente: Elaboración Propia

En la Figura 15 se muestra los gráficos de densidad antes y después de ser transformados, los cuales se comprueban visualmente que siguen una distribución normal teniendo todas las variables la misma media lo que le facilita al modelo de Machine Learning hacer estas variables comparables y evitar el sesgo de darle más importancia a una variable por su unidad de medida.

4.6.4 Selección del Modelo

Para la creación del modelo, se utilizó la biblioteca Scikit-learn, que proporciona una amplia variedad de algoritmos de Machine Learning, búsqueda de hiper-parámetros y evaluación de modelos. Luego del entrenamiento del modelo con la data histórica se utilizaron las métricas MSE, R Cuadrado y MAPE para su comparación como se muestra en la Tabla 11.

Tabla 11: Evaluación de modelos

MODELO	MSE	R2	MAPE
ExtratreesRegressor	8.83	0.93	1.4
RandomForestRegressor	10.97	0.92	1.54
CatBoostRegressor	10.88	0.92	1.58
LGBMRegressor	11.25	0.92	1.60
XGBRegressor	11.45	0.92	1.62
GradientBoostingRegressor	13.44	0.90	1.75
SVR	16.95	0.88	1.89

Fuente: Elaboración Propia

Para todos los modelos se dividió la base de datos en entrenamiento y testeo, esta es una buena práctica para crear modelos ya que se evita el overfitting (sobreajuste) y el underfitting (subajuste) ya que son dos problemas muy comunes y hace que sus predicciones no sean confiables. Se buscó la mejor combinación de hiper-parámetros con el fin de mejorar las métricas de medición de los modelos encontrados. Los entrenamientos se realizaron de forma dinámica

considerando tiempos de entrenamiento de 6 horas para la predicción de 5 minutos, esto con el objetivo de que el modelo se entrene y por ende se adapte a las últimas condiciones que se tiene en la planta de Molienda – Clasificación.

4.6.5 Sistema de Recomendaciones

Para interpretar el modelo obtenido y entender como es afectado por las variables predictoras, se usa la librería SHAP la cual descompone las predicciones del modelo de Machine Learning en contribuciones individuales de cada variable predictora y calcula cuánto cada variable predictora afecta a la predicción en particular. Mediante estos valores SHAP se puede saber cuál es la importancia de cada variable predictora y mediante el lenguaje Python se crea una lógica para evaluar si cada variable independiente aumenta o disminuye el valor de la variable a predecir que es el P80, con el fin de reducir su variabilidad.

Figura 16: Cuadro de Recomendaciones de Variables Operativas

				Valor Actual	Target	Orden
Inicio	Final	Variables				
2023-04-06 00:10:00	2023-04-06 00:15:00	CAJON_NIVEL		84.123291	79.23	Disminuir
		CICLONES_PRESIONKPA		129.517905	120.96	Disminuir
2023-04-06 00:15:00	2023-04-06 00:20:00	CAJON_NIVEL		82.856931	79.23	Disminuir
		CICLONES_PRESIONKPA		128.050632	120.96	Disminuir
		CHANC_P80		19.620137	20.57	Aumentar
...	...	...		...	...	...
2023-05-03 23:30:00	2023-05-03 23:35:00	BALL_FLUJOAGUA		139.746085	149.87	Aumentar
2023-05-03 23:35:00	2023-05-03 23:40:00	CAJON_FLUJOAGUA		1383.437207	1398.87	Aumentar
		CAJON_NIVEL		79.190402	84.17	Aumentar
2023-05-03 23:40:00	2023-05-03 23:45:00	CAJON_FLUJOAGUA		1381.937085	1398.87	Aumentar
		CAJON_NIVEL		78.912868	83.69	Aumentar

Fuente: Elaboración Propia

En la Figura 16 se muestra un ejemplo de las recomendaciones de los valores del proceso, el valor actual de esa variable y el orden de disminuir o aumentar.

Para construir el modelo, se utilizan todas las variables disponibles en línea, pero para el sistema de recomendaciones no se consideran las variables que no son operativamente manipulables. Además, no se incluyen las variables operativas del Molino SAG, ya que las recomendaciones para estas podrían interferir con las estrategias de optimización del tonelaje tratado por dicho molino. Por esta razón, en la Tabla 12 se especifican las variables que se considerarán en el sistema de recomendaciones las cuales están enfocadas en el Circuito de Chancado de Pebbles, Molienda de Bolas y Clasificación.

Tabla 12: Variables consideradas en el Sistema de Recomendaciones

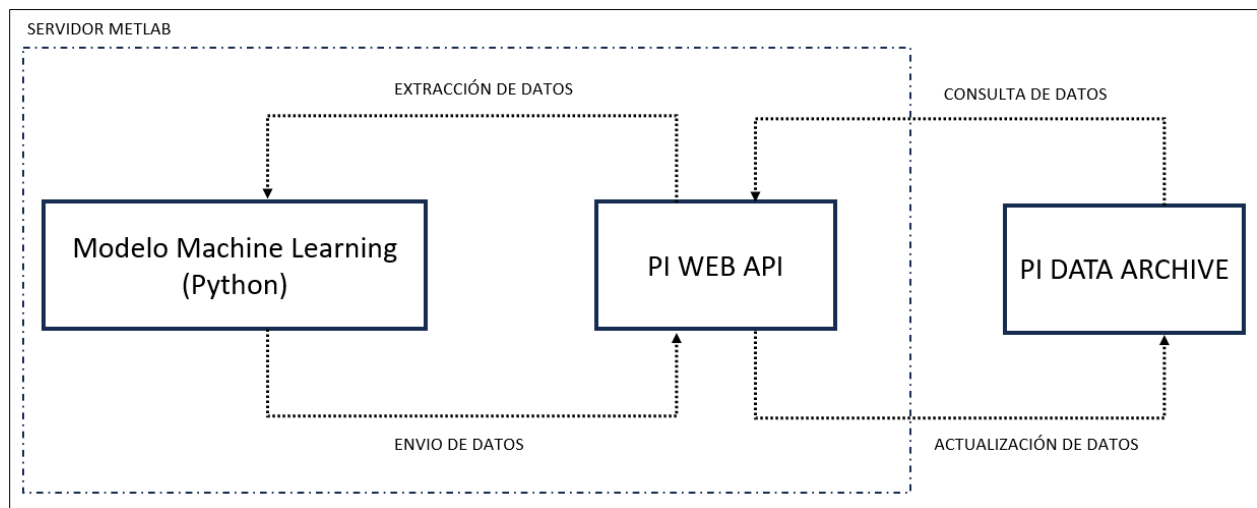
TAG	Descripción
CHANC_P80	P80 Chancadora de Pebbles
BALL_FLUJOAGUA	Flujo Agua chute Alimentación Molino de Bolas
BALL_FLUJOAGUATROMMEL	Flujo Agua Trommel Molino de Bolas
CAJON_NIVEL	Nivel Cajón de Bombeo
CAJON_FLUJOAGUA	Flujo Agua alimento Cajón
CICLONES_PRESIONKPA	Densidad Alimentación Hidrociclones

Fuente: Elaboración Propia

4.6.6 Conectividad

La conectividad del modelo se realiza mediante “PI Web API”. Que es una interfaz de programación de aplicaciones (API) que proporciona acceso programático al sistema PI. Permite interacción con datos de PI System (PI Data Archive) y el entorno de Python en el cual se desarrolla el Modelo de Machine Learning y el Sistema de Recomendaciones.

Figura 17: Esquema de Conectividad del Modelo



Fuente: Elaboración Propia

La Figura 17 muestra el esquema de conectividad del modelo con el PI System. En el servidor llamado "METLAB", que es donde se hace el desarrollo del proyecto, se realiza la consulta y el envío de datos utilizando la PI WEB API hacia el servidor del PI Data Archive.

4.6.7 Puesta en Línea

Se crean hacen las conexiones con PIWebAPI y modificación del código fuente para que se pueda extraer los datos en línea y en tiempo real, hacia el script de Python.

Figura 18: Extracción de datos del PI System

```
paths = {"pi:\\\\10.112.0.171\\220WI1210_DRY.DCS": "SAG_TON",
        "pi:\\\\10.112.0.171\\310ML001_SPEED_SP.DCS": "SAG_VELOCIDAD",
        "pi:\\\\10.112.0.171\\310PI2245.DCS": "SAG_PRESIONCARGA",
        "pi:\\\\10.112.0.171\\310PI2246.DCS": "SAG_PRESIONDESCARGA",
        "pi:\\\\10.112.0.171\\310LIC2301.PV.DCS": "NIVEL_CAJON",
        "pi:\\\\10.112.0.171\\SPLIT_F80.DCS": "SAG_F80",
        "pi:\\\\10.112.0.171\\310FIC2100.PV.DCS": "SAG_FLUJOAGUA",
        "pi:\\\\10.112.0.171\\310JI2200.DCS": "SAG_POTENCIA",
        "pi:\\\\10.112.0.171\\SPLIT_MATERIAL_FINO.DCS": "SAG_FINO",
        "pi:\\\\10.112.0.171\\SPLIT_MATERIAL_INTER.DCS": "SAG_INTER",
        "pi:\\\\10.112.0.171\\SPLIT_MATERIAL_GRUESO.DCS": "SAG_GRUESO",
        "pi:\\\\10.112.0.171\\320WI2560.DCS": "CHANC_TON",
        "pi:\\\\10.112.0.171\\310DIC2100.PV.DCS": "CHANC_PRESIONCLAMPING",
        "pi:\\\\10.112.0.171\\320CR006.POTENCIA": "CHANC_POTENCIA",
        "pi:\\\\10.112.0.171\\SPLIT2_F80.DCS": "CHANC_P80",
        "pi:\\\\10.112.0.171\\310FIC2304.PV.DCS": "BALL_FLUJOAGUA",
        "pi:\\\\10.112.0.171\\310FI2307.DCS": "BALL_FLUJOAGUATROMMEL",
        "pi:\\\\10.112.0.171\\310JI2500.DCS": "BALL_POTENCIA",
        "pi:\\\\10.112.0.171\\310PI2544.DCS": "BALL_PRESION_CARGA1",
        "pi:\\\\10.112.0.171\\310PI2545.DCS": "BALL_PRESION_CARGA2",
        "pi:\\\\10.112.0.171\\310PI2546.DCS": "BALL_PRESION_DESCARGA1",
        "pi:\\\\10.112.0.171\\310PI2547.DCS": "BALL_PRESION_DESCARGA2",
        "pi:\\\\10.112.0.171\\310SE2500.DCS": "BALL_VELMOTOR2",
        "pi:\\\\10.112.0.171\\310SE2510.DCS": "BALL_VELMOTOR1",
        "pi:\\\\10.112.0.171\\WorkIndex_Oper.DCS": "BALL_WI",
        "pi:\\\\10.112.0.171\\310LIC2301.PV.DCS": "CAJON_NIVEL",
        "pi:\\\\10.112.0.171\\310FIC2305.PV.DCS": "CAJON_FLUJOAGUA",
        "pi:\\\\10.112.0.171\\310DI2330.DCS": "CICLONES_DENSIDADALIMENTACION",
        "pi:\\\\10.112.0.171\\310PI2342.DCS": "CICLONES_PRESIONKPA",
        "pi:\\\\10.112.0.171\\310UY3811A.DCS": "CICLONES_OFFP80"
    }

filter = "" 'Downtimes.Faja CV003.Funcionando'=1 and '320HV2910.ST.DCS'='Cerrada Confirmada' and '320CR006MD1.DCS'='Funcionando' and '210CR007M2D1.DCS'

data = None

for i in paths:
    df = client.data.get_interpolated_values(i,
                                            start_time = "t-30m", #t-4d
                                            end_time   = "",
                                            interval   = "1m",
                                            selected_fields = "items.timestamp;items.value"
                                            )
```

Fuente: Elaboración Propia

Se crean los tags necesarios historizar las predicciones y métricas necesarias, se envían los datos al PI mediante PIWebAPI.

Figura 19: Envío de datos al PI System

ENVIAR PREDICCIONES A PI

```
df.index = df.index + timedelta(hours = 5)
df.reset_index(inplace = True)
df['PRED']=df['PRED'].round(3)
df
```

	Timestamp	PRED
0	2024-02-29 00:10:00	149.94

```
#Creación de objetos
streamValuesItems = PIItemsStreamValues()
streamValue1 = PIStreamValues()
value1 = PITimedValue()

#Añadir valores y tiempos en los atributos de los valores creados
value1.value = df['PRED'][0]
value1.timestamp = (df['Timestamp'])[0]

#Obtener Tag a alimentar datos
point = client.point.get_by_path("\\\\10.112.0.171\\310UY3811A.Test")
```

Fuente: Elaboración Propia

En las figuras 18 y 19, se muestra el código utilizado para la extracción y envío de datos al PI System. Teniendo conectividad se programa para la ejecución del modelo cada cinco minutos mediante tareas de Windows como se muestra en la Figura 20, cuyo objetivo es que se tenga una comunicación constante entre el modelo y los datos operativos de la planta logrando así tener el Modelo de Machine Learning y el Sistema de Recomendaciones en línea.

Figura 20: Tarea de Windows para ejecución del Modelo

 P80_Modelamiento

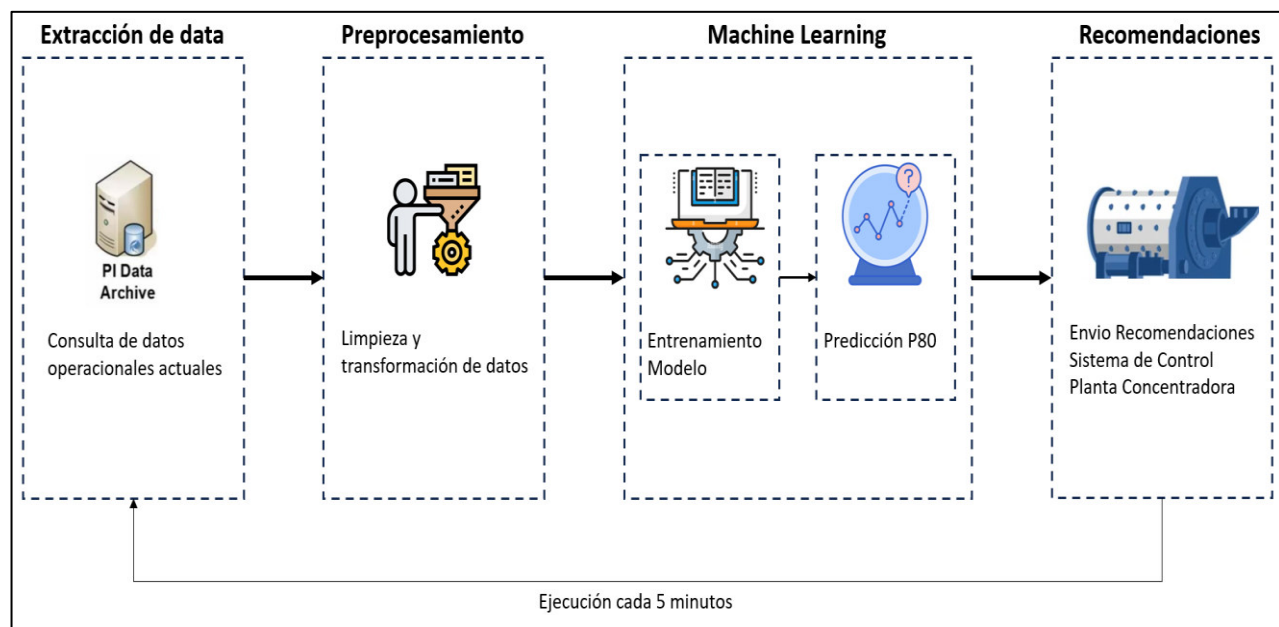
Ready At 00:00 on 14/2/2024 - After triggered, repeat every 00:01:00 indefinitely.

Fuente: Elaboración Propia

La figura 21, muestra el flujo de trabajo del sistema de recomendaciones basado en el modelo de Machine Learning, aplicado a la planta concentradora. El proceso inicia con la extracción de datos operacionales actuales desde el PI Data Archive, seguido de una etapa de

preprocesamiento, donde los datos se limpian y transforman para su análisis. A continuación, en la fase de Machine Learning, se entrena un modelo predictivo que estima el valor del P80. Con base en esta predicción, se generan recomendaciones que son enviadas al sistema de control de la planta concentradora para optimizar la operación. Todo este ciclo se ejecuta automáticamente cada cinco minutos, asegurando una actualización continua y en tiempo real del sistema.

Figura 21: Diagrama de Flujo del Sistema de Recomendaciones



Fuente: Elaboración Propia

5 CAPITULO V: RESULTADOS

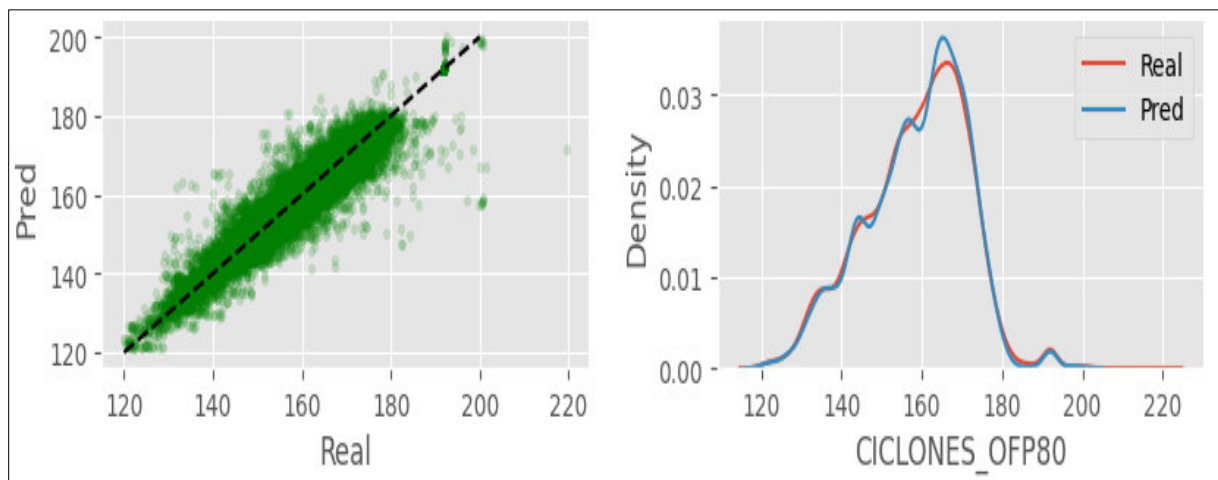
Los resultados de la tesis desarrollada se dividen en las siguientes secciones: Métricas del Modelo, Sistema de Recomendaciones, Simulación Histórica y Puesta en Línea.

5.1 Métricas del Modelo

Luego de evaluar con todos los modelos mencionados en la sección anterior, el modelo con el mejor rendimiento es el ExtraTreesRegressor con las siguientes métricas:

- $R^2 = 0.93$
- $RMSE = 8.83$
- $MAPE = 1.4$

Figura 22: Resultados del Modelo



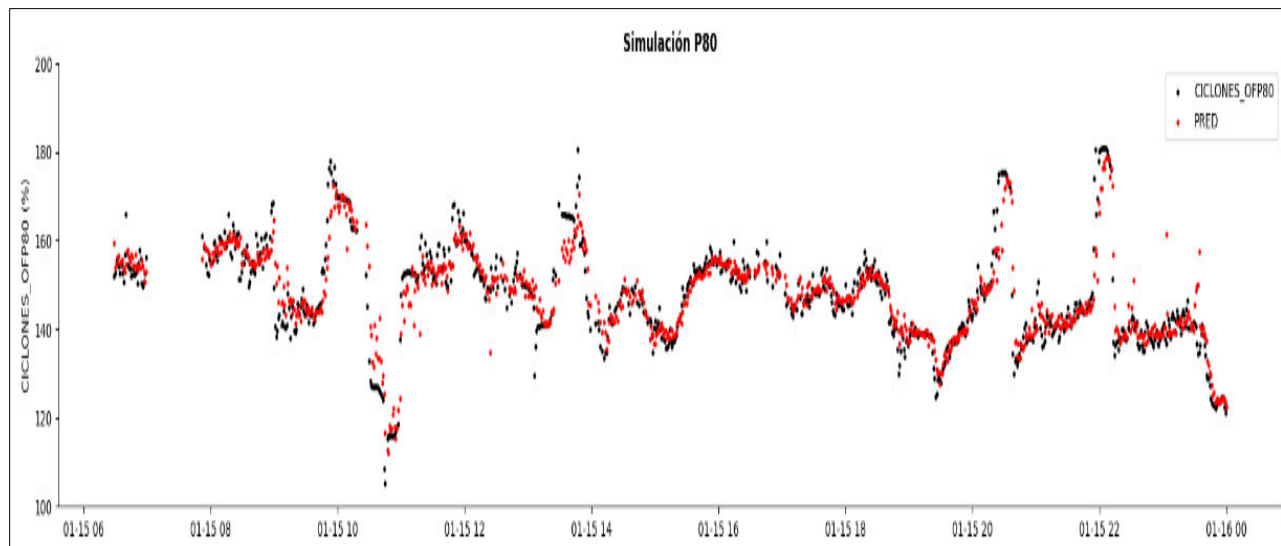
Fuente: Elaboración Propia

En la Figura 22, en el gráfico de la izquierda se muestra un gráfico de dispersión del valor real y el valor predicho, donde se observa que los puntos convergen en la línea diagonal de los ejes que representa la exactitud de las predicciones del modelo. En el gráfico de la derecha se

muestra las distribuciones de densidad donde se muestra que las concentraciones de los datos de los valores reales y predichos son similares.

En la Figura 23, se hace una simulación con la data real y la data predicha por el Modelo de Machine Learning, en donde se nota visualmente que las predicciones corresponden a las tendencias de los valores reales indicando que el modelo es acertado con la realidad y que es adaptativo de acuerdo a las diversas condiciones operativas que se presentan en la planta de Molienda - Clasificación.

Figura 23: Simulación del Modelo



Fuente: Elaboración Propia

5.1.1 Fórmula del modelo

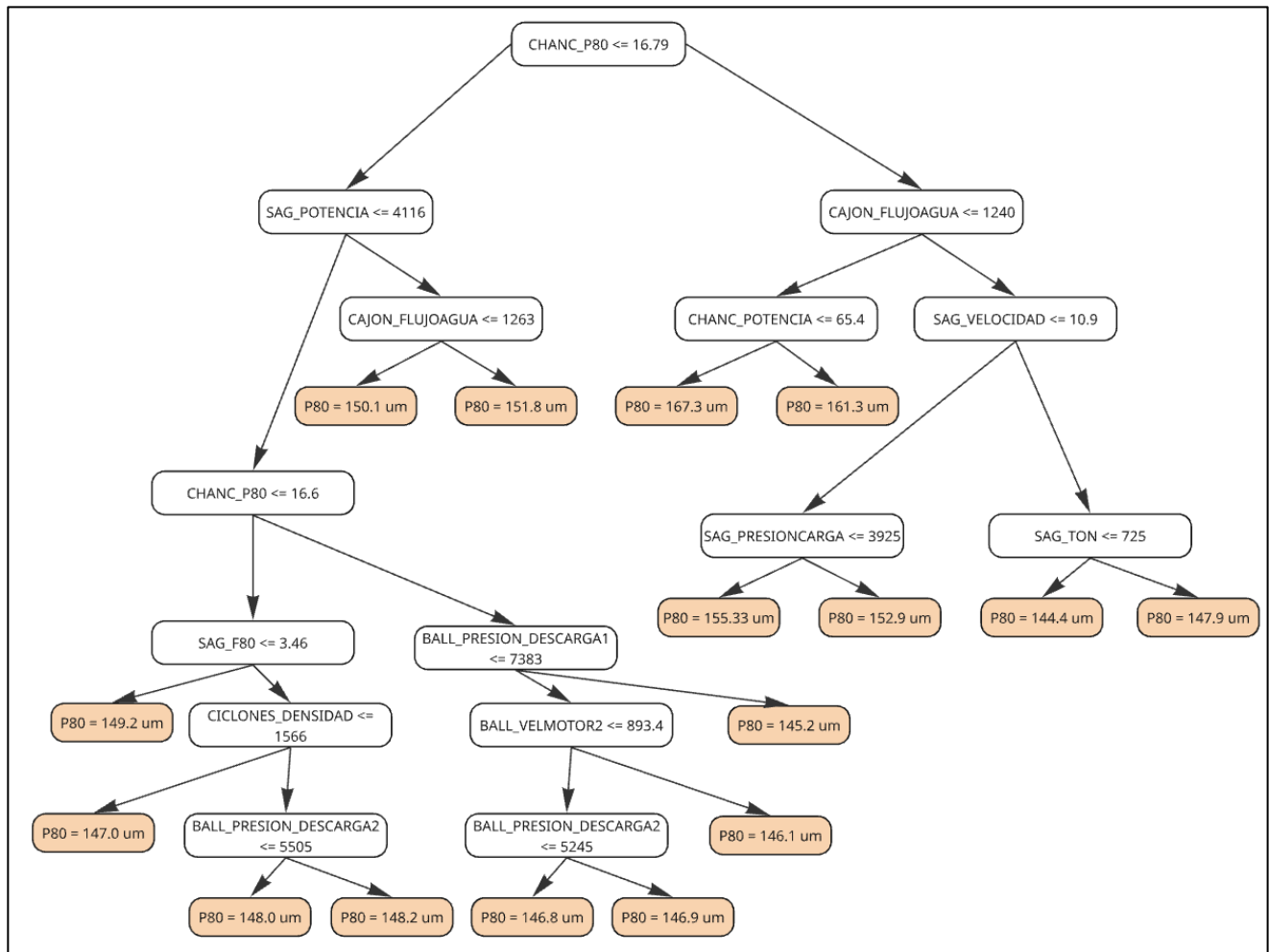
La fórmula del modelo ExtraTreesRegressor es una representación textual del árbol de decisión. Cada línea de la fórmula representa una regla de decisión. La cual según sea el caso evaluado se va clasificando según las características entrenadas para obtener una predicción.

Figura 24: Estructura del Modelo

```
|--- CHANC_P80 <= 16.79
|   |--- SAG_POTENCIA <= 4116.36
|   |   |--- CHANC_P80 <= 16.62
|   |   |   |--- SAG_F80 <= 3.45
|   |   |   |   |--- value: [149.20]
|   |   |   |--- SAG_F80 > 3.45
|   |   |   |   |--- CICLONES_DENSIDADALIMENTACION <= 1566.91
|   |   |   |   |   |--- value: [147.45]
|   |   |   |   |--- CICLONES_DENSIDADALIMENTACION > 1566.91
|   |   |   |   |   |--- BALL_PRESION_DESCARGA2 <= 5505.29
|   |   |   |   |   |   |--- value: [148.01]
|   |   |   |   |   |--- BALL_PRESION_DESCARGA2 > 5505.29
|   |   |   |   |   |   |--- value: [148.25]
|   |   |   |--- CHANC_P80 > 16.62
|   |   |   |   |--- BALL_PRESION_DESCARGA1 <= 7383.40
|   |   |   |   |   |--- BALL_VELMOTOR2 <= 893.43
|   |   |   |   |   |   |--- BALL_PRESION_DESCARGA2 <= 5245.94
|   |   |   |   |   |   |   |--- value: [146.88]
|   |   |   |   |   |   |--- BALL_PRESION_DESCARGA2 > 5245.94
|   |   |   |   |   |   |   |--- value: [146.98]
|   |   |   |   |   |--- BALL_VELMOTOR2 > 893.43
|   |   |   |   |   |   |--- value: [146.18]
|   |   |   |   |--- BALL_PRESION_DESCARGA1 > 7383.40
|   |   |   |   |   |--- value: [145.23]
|   |   |--- SAG_POTENCIA > 4116.36
|   |   |   |--- CAJON_FLUJOAGUA <= 1263.94
|   |   |   |   |--- value: [150.18]
|   |   |   |--- CAJON_FLUJOAGUA > 1263.94
|   |   |   |   |--- value: [150.89]
|   |--- CHANC_P80 > 16.79
|   |   |--- CAJON_FLUJOAGUA <= 1240.42
|   |   |   |--- CHANC_POTENCIA <= 65.38
|   |   |   |   |--- value: [167.30]
|   |   |   |--- CHANC_POTENCIA > 65.38
|   |   |   |   |--- value: [161.35]
|   |   |--- CAJON_FLUJOAGUA > 1240.42
|   |   |   |--- SAG_VELOCIDAD <= 10.91
|   |   |   |   |--- SAG_PRESIONCARGA <= 3925.83
|   |   |   |   |   |--- value: [155.33]
|   |   |   |   |--- SAG_PRESIONCARGA > 3925.83
|   |   |   |   |   |--- value: [152.97]
|   |   |   |--- SAG_VELOCIDAD > 10.91
|   |   |   |   |--- SAG_TON <= 725.94
|   |   |   |   |   |--- value: [144.47]
|   |   |   |   |--- SAG_TON > 725.94
|   |   |   |   |   |--- value: [148.00]
```

Fuente: Elaboración Propia

Figura 25: Representación Gráfica de la Estructura del Modelo



Fuente: Elaboración Propia

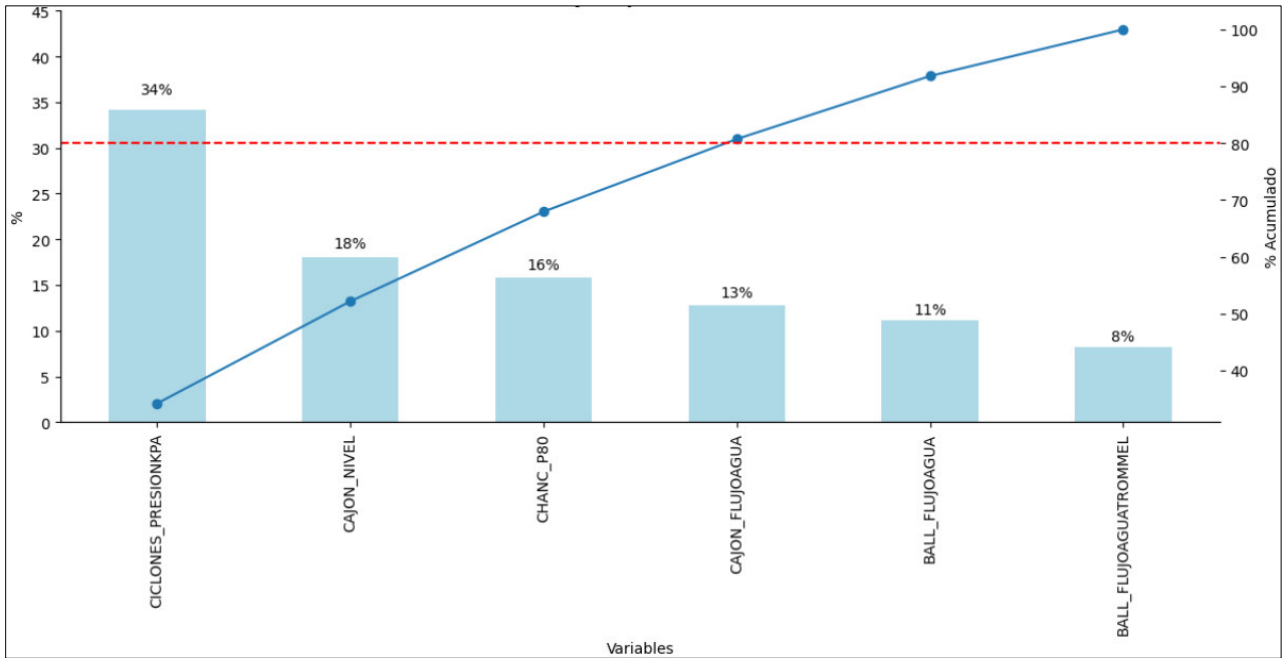
Las Figuras 24 y 25 ilustran la estructura del modelo de árboles de decisión y su representación esquemática. En estas figuras, se aprecia cómo el modelo, construido a partir de los datos de entrenamiento, organiza su proceso de toma de decisiones en una estructura jerárquica similar a un árbol. Cada nodo interno del árbol representa una prueba o condición sobre una variable predictora; esto se refiere a una evaluación que el modelo realiza para determinar si un determinado valor de la variable cumple con un criterio específico, como ser

mayor o menor que un umbral dado. Las ramas resultantes de ese nodo indican los caminos posibles según el resultado de dicha prueba. De esta manera, el modelo define sus reglas de decisión de manera secuencial, dividiendo repetidamente los datos en subconjuntos más pequeños en función de los valores de las variables predictoras. Este proceso permite al modelo diferenciar entre distintas clases o valores objetivo, optimizando su capacidad para realizar predicciones precisas.

5.2 Sistema de Recomendaciones

Se realiza una Simulación de las recomendaciones brindadas en el periodo del análisis de la presente investigación, con el objetivo de evaluar cuáles son las variables con mayores apariciones y si su orden fue aumentar o disminuir respecto a su valor real en cada estampa de tiempo de predicción respectiva.

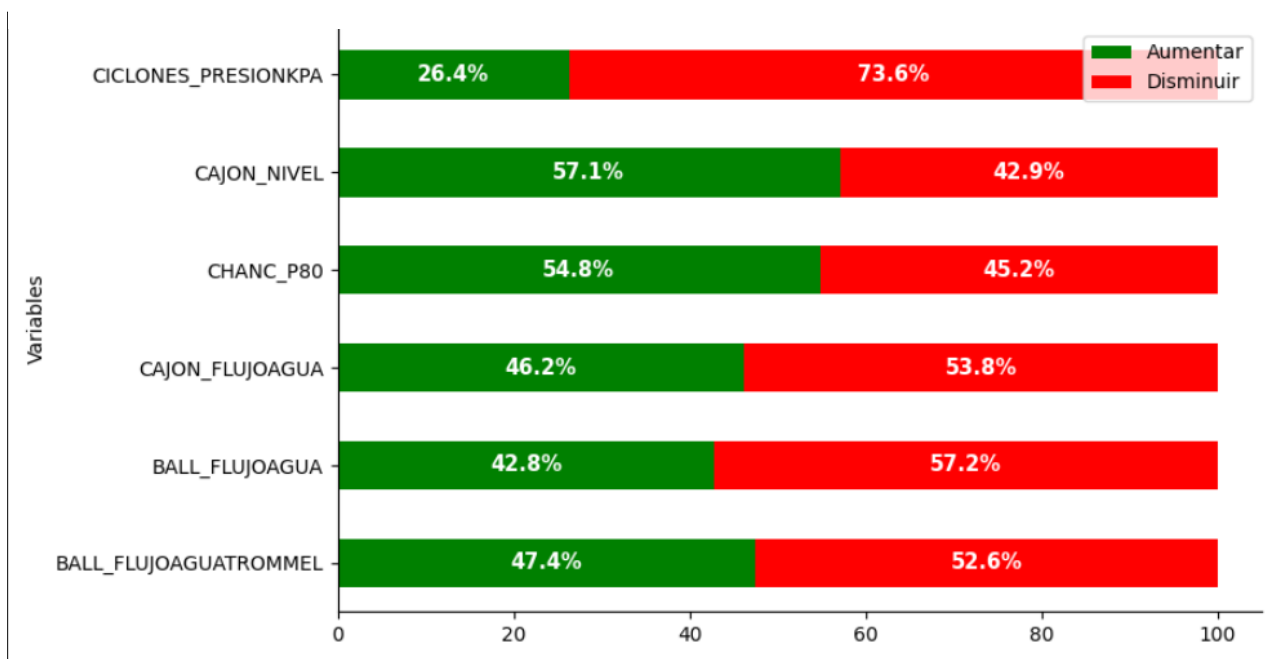
Figura 26: Pareto de Variables con mayores apariciones en las Recomendaciones



Fuente: Elaboración Propia

La Figura 26, muestra el gráfico de Pareto donde muestra la importancia relativa de las diferentes variables en las recomendaciones simuladas. Las variables se ordenan de mayor a menor importancia, según su porcentaje de aparición en las recomendaciones. Se observa que la variable CICLONES_PRESIONKPA es la que mayor frecuencia muestra, con un 34%. Las otras variables contribuyen con un porcentaje menor, siendo CAJON_NIVEL la segunda variable más importante con un 18%. La línea roja indica el 80% de las apariciones de todas las recomendaciones por las variables, lo que indica que las primeras cuatro variables explican el 80% de la variabilidad del P80 del circuito Molienda - Clasificación.

Figura 27: Gráfico de barras con mayores apariciones en las Recomendaciones

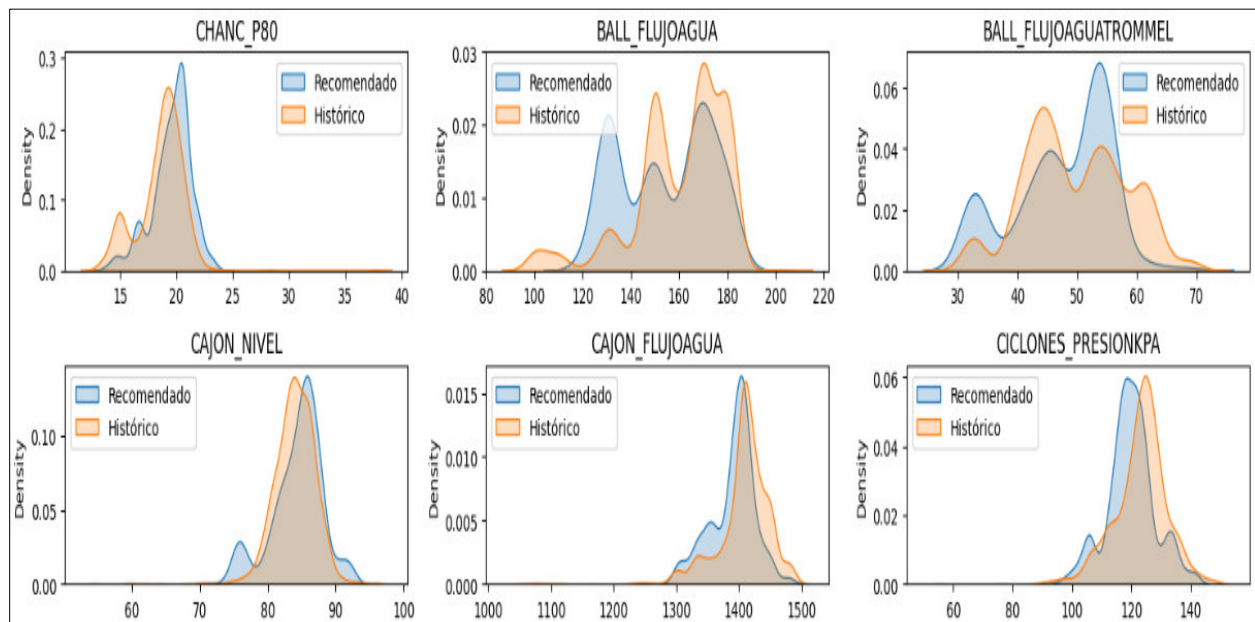


Fuente: Elaboración Propia

La Figura 27, muestra el porcentaje de veces que se recomienda aumentar o disminuir cada variable. Se puede observar que, para la mayoría de las variables, se recomienda disminuirlas más que aumentarlas. Para la variable "CICLONES_PRESIONKPA", se ha recomendado disminuirla un 73.6% de todas veces que se recomendó esta variable y para la

variable "CAJON_NIVEL" se ha recomendado aumentar un 57.1%. Esto indica que para motivos de reducir la variabilidad del P80 del circuito Molienda – Clasificación, existe algunas tendencias de disminuir o aumentar los valores que usualmente se operan de las variables en recomendación.

Figura 28: Distribuciones de valores de variables según histórico y Recomendaciones



Fuente: Elaboración Propia

En la Figura 28, se muestra la comparación entre las distribuciones de los valores históricos y recomendados de las variables consideradas para el Sistema de Recomendaciones. Se observa distribuciones similares pero desfasadas lo que indica la tendencia de aumento o disminución según cada variable. En el caso de las variables de “BALL_FLUJOAGUA” y “BALL_FLUJOAGUATROMMEL” se observa que su distribución varía a ser trimodal, lo cual indica que en las recomendaciones se sugieren valores que históricamente no eran frecuentes.

Tabla 13: Métricas de valores históricos y recomendados de variables operativas.

Tag	UM	Histórico		Recomendado		Reducción Variabilidad
		Mediana	Desv. Estandar	Mediana	Desv. Estandar	
CAJON_FLUJOAGUA	m3/h	1410.85	44.67	1398.66	35.38	21%
CHANC_P80	mm	18.97	2.05	19.87	1.75	15%
CICLONES_PRESSIONKPA	kPa	123.9	9	119.78	8.16	9%
BALL_FLUJOAGUATROMMEL	m3/h	48.4	8.42	51.11	8.06	4%
CAJON_NIVEL	%	84.1	3.18	85.06	3.08	3%
BALL_FLUJOAGUA	m3/h	167.41	19.19	149.88	18.75	2%

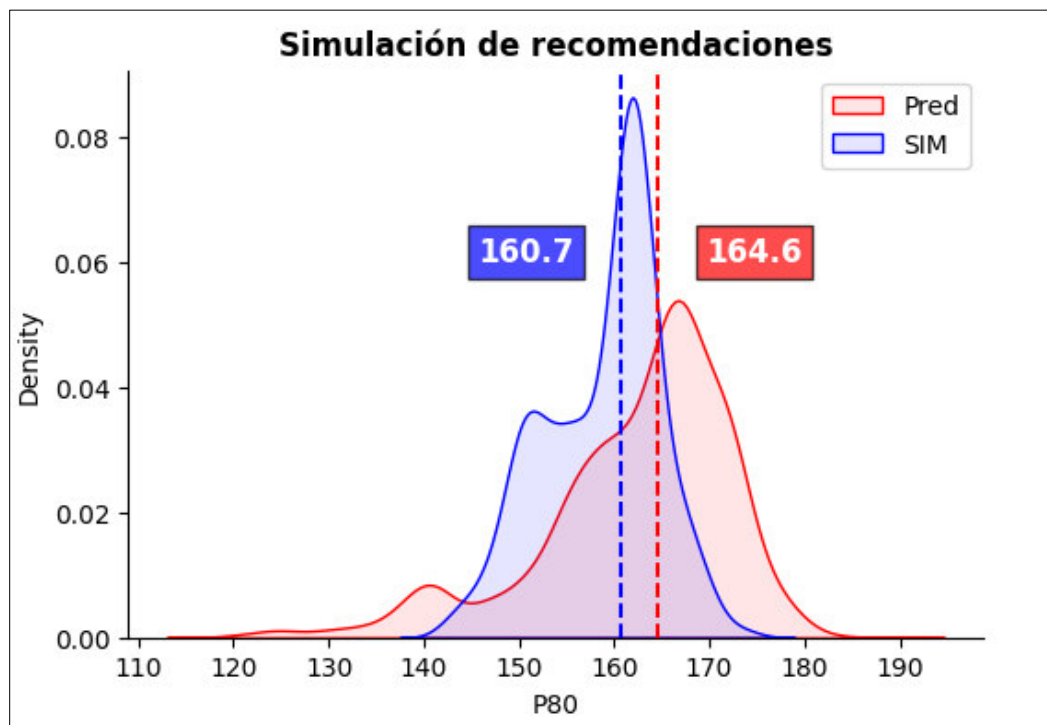
Fuente: Elaboración Propia

En la Tabla 13, Se muestra las métricas de los valores históricos y recomendados de las variables consideradas en el Sistema de Recomendaciones. En todas las variables se evidencia una disminución de variabilidad, siendo las más significativas las variables de “CAJON_FLUJOAGUA” con una reducción de variabilidad del 21% y “CHANC_P80” con una reducción de variabilidad del 15%. Esto evidencia que se tendría mayor control sobre los valores de las variables operativas y esto genera mayor estabilidad en el circuito Molienda – Clasificación.

5.3 Simulación Histórica

Se realizan las simulaciones del impacto en el P80 en el caso se hubieran modificado las variables operativas de acuerdo lo recomendado por el modelo de Machine Learning para reducir la variabilidad del P80.

Figura 29: Gráfico de Distribución de P80 Simulado y Predicho

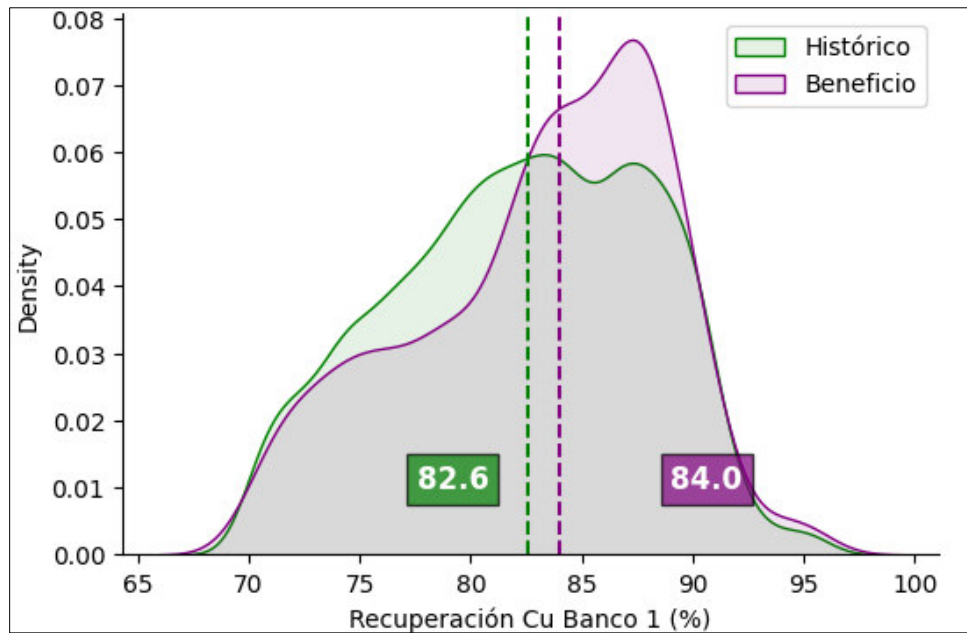


Fuente: Elaboración Propia

Como se observa en la Figura 29, en el caso del P80 simulado se tiene una media de 160.7 μm y una variación estándar de 4.48 en comparación con el P80 predicho el cual tiene una mediana de 164.67 y una desviación estándar de 7.32, esto significa que las recomendaciones que se toman reducen la variabilidad del P80 siendo así beneficioso para la flotación.

Según los valores obtenidos por la simulación de reducción de variabilidad del P80, se calcula mediante valores históricos que este impacto representa un aumento del 1.4% de Recuperación en Cu del Banco 1 del circuito Rougher, como se muestra en la Figura 30.

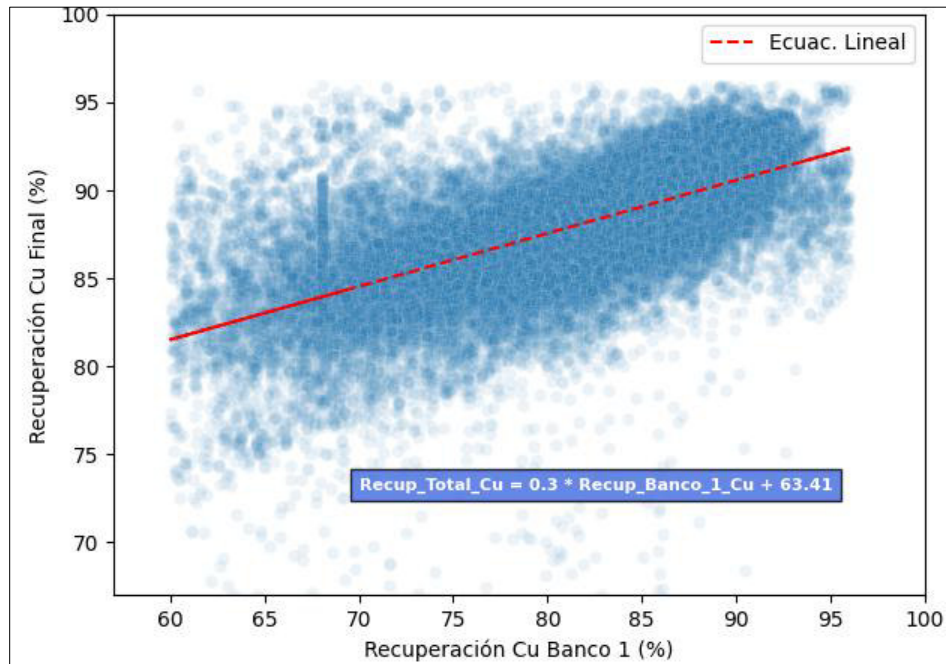
Figura 30: Beneficio de Recuperación de Cobre Banco 1 por Disminución de Variabilidad



Fuente: Elaboración Propia

De igual manera se calcula el beneficio del 0.42% en la Recuperación Final de Cobre, debido a que la Recuperación de Cobre del Banco 1 tiene una relación directa con la Recuperación de Cobre Total como se evidencia en la Figura 31.

Figura 31: Beneficio de Recuperación Final de Cobre por Disminución de Variabilidad

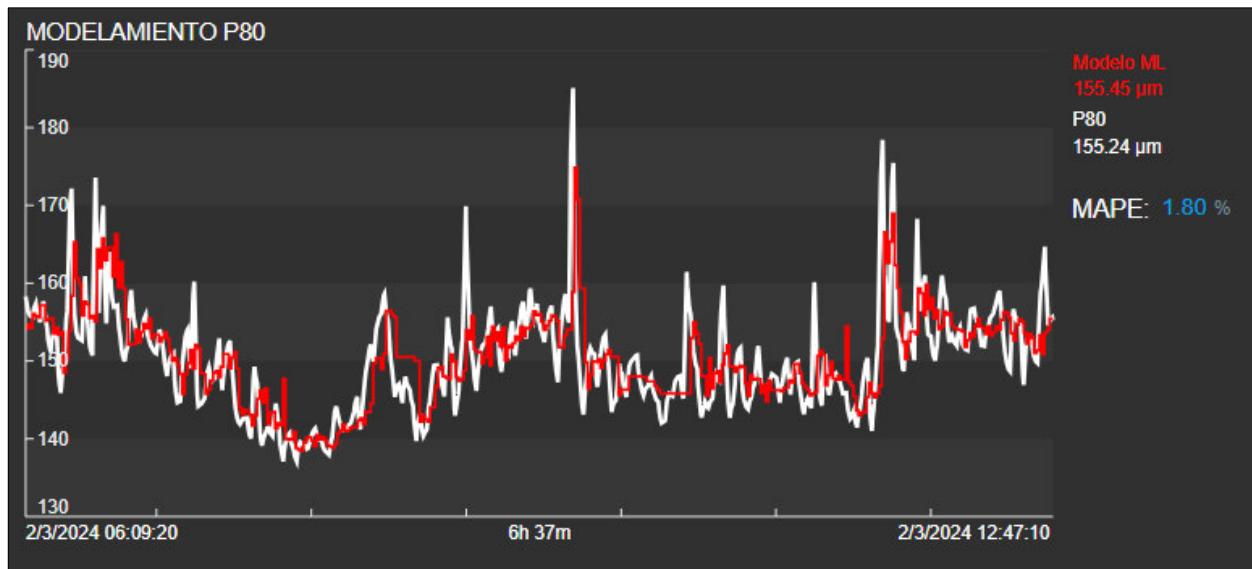


Fuente: Elaboración Propia

5.4 Puesta en línea

Teniendo el modelo ejecutándose cada 5 minutos y enviando datos al PI Data Archive, mediante el PI Vision se crean visualizaciones de las tendencias de las predicciones del modelo y las variables a recomendar para cada estampa de tiempo y así poder ver el desempeño del modelo y las recomendaciones brindadas en tiempo real acorde de la operación de la Planta de Molienda – Clasificación.

Figura 32: Visualización del Modelo en Línea en el PI Vision



Fuente: Elaboración Propia

En la Figura 32, se muestra el despliegue del modelo en el PI Vision en el cual se le puede dar un seguimiento operativo en línea frente a las diversas situaciones operacionales que se presentan.

6 CAPITULO VI: DISCUSIÓN DE RESULTADOS

En esta sección se discuten los resultados de la presente investigación, se resaltan los hallazgos más importantes de los resultados obtenidos. Los resultados se discuten en relación con los objetivos planteados inicialmente. A través de esta investigación se pretende ofrecer una comprensión más profunda de la aplicación de técnicas de Machine Learning en el circuito de Molienda - Clasificación y su impacto en el procesamiento de minerales.

- En el desarrollo del modelo de Machine Learning tiene como variable a predecir el P80 del circuito de Molienda – Clasificación, se logró obtener un modelo que es asertivo a la realidad teniendo un error porcentual medio de 1.4 %. Además, al ser un modelo basado en árboles de decisión, permite conocer cual son las reglas de decisión que se usan para realizar una predicción.
- El sistema de recomendaciones desarrollado muestra que para cada predicción hay diferentes variables que tienen más influencia en la predicción del P80. Lo cual es acorde con la realidad debido a la complejidad del proceso el cual tiene diversas variables y es difícil saber cuál es la mayor perturbación que se tiene en un determinado escenario operativo. Sin embargo, en el periodo de tiempo utilizado para la investigación se evidencia que las variables recomendadas con mayores apariciones son la Presión de los Hidrociclones y el Nivel del Cajón de Bombeo, las cuales muestran tendencia a disminuir con un 73.6% y aumentar con un 57.1% respectivamente en todos los casos que se recomiendan.

Además, se evidencia una reducción en la variabilidad de las variables consideradas en el Sistema de Recomendaciones, donde las reducciones de variabilidad más significativas son en las variables de Flujo de Agua alimentado al Cajón de Bombeo

con 21% y el P80 de la Chancadora de Pebbles con 15% de reducción de variabilidad respecto a los valores históricos.

- Mediante las simulaciones históricas y comparando los resultados de las recomendaciones brindadas y la data histórica del periodo de tiempo en evaluación, se muestra una reducción de variabilidad del P80, teniendo una reducción de la desviación estándar del 2.84 μm . Esta reducción de variabilidad tiene un beneficio histórico del 0.42% de aumento en la Recuperación Final de Cu.
- El despliegue del modelo y sistema de recomendaciones mediante el PI System evidencian la adaptabilidad del modelo a las diversas situaciones operativas que se presentan, esto se demuestra con la estabilidad de las métricas del modelo que son comparadas con los valores reales del P80, lo que da confiabilidad del estudio realizado y las recomendaciones brindadas.

7 CONCLUSIONES

- Según las simulaciones históricas realizadas por el Sistema de Recomendaciones en el tiempo de análisis de la presente investigación, se obtuvo una reducción de variabilidad del P80 del 2.84 μm que representa un aumento en la Recuperación Final de Cu del 0.42%.
- Se logró desarrollar un modelo de Machine Learning para el P80 del circuito Molienda – Clasificación con un error absoluto porcentual medio de 1.4%. Lo cual evidencia que el modelo es acertado con la realidad y logra adaptarse a las diversas situaciones operacionales, dando confiabilidad a las recomendaciones que se dan a partir de la explicabilidad del modelo.
- Según los resultados de las simulaciones, el 80% de la variabilidad del P80 en el Circuito de Molienda y Clasificación fue controlada principalmente por las siguientes variables: Presión de los Hidrociclones, Nivel del Cajón de Bombeo, P80 de la Chancadora de Pebbles, y Flujo de Agua al Cajón de Bombeo. Entre las recomendaciones generadas por el modelo de Machine Learning, se sugirió reducir la presión de los Hidrociclones en un 73.6% de las veces, y aumentar el nivel del Cajón de Bombeo en un 57.1% de las veces. Esto indica que los valores históricos de estas variables no son óptimos, lo que explica las tendencias de aumentar o disminuir los valores de las variables.
- El Sistema de Recomendaciones logró reducir la variabilidad en el circuito Molienda - Clasificación, destacándose una disminución del 21% en el flujo de agua al Cajón de Bombeo y del 15% en el P80 de la Chancadora de Pebbles. Estos resultados evidencian un mayor control y estabilidad de las variables operativas, al tiempo que permiten una toma de decisiones más oportuna, precisa y adaptada a las condiciones reales del proceso en tiempo real.

- La interfaz PIWebAPI permite conectar el PI System con el entorno de Machine Learning, desplegando el modelo en línea y el Sistema de Recomendaciones con un tiempo de ejecución de 5 minutos. Esto mejora significativamente el control del P80, que antes solo se ajustaba tras aumentos notables y según criterios operativos variados.

8 RECOMENDACIONES

- Se recomienda realizar pruebas durante 74 días, considerando que una campaña de liners del Molino SAG dura 120 días. Este periodo se calculó con un 95% de confianza y un margen de error del 5%, utilizando muestreo para poblaciones finitas.
- Se recomienda utilizar mayor cantidad de variables mineralógicas y metalúrgicas que se cuenten en línea siempre y cuando sean confiables. Esto aportaría más información de los blendings que se hacen en el día y podría mejorar las predicciones del modelo de Machine Learning.
- Para lograr una mayor automatización, es posible implementar el despliegue de las recomendaciones generadas por el modelo de Machine Learning directamente en el Sistema de Control de la planta concentradora. De esta manera, se permite el ajuste automático de las variables operativas según las condiciones del proceso, optimizando el rendimiento en tiempo real. Esta integración debe realizarse considerando estrictamente las medidas de seguridad operacionales, con el fin de garantizar la continuidad y estabilidad de la operación.

9 REFERENCIAS BIBLIOGRÁFICAS

Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*.

Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and challenges. *Science*, 349(6245).

Brownlee, J. (2023). *Machine learning mastery with Python*. Jason Brownlee.

Géron, A. (2017). *Hands-on machine learning with Scikit-Learn and TensorFlow*. O'Reilly Media.

Ministerio de Energía y Minas (2021). *Anuario de la Minería 2020*. Lima: Ministerio de Energía y Minas.

Sociedad Nacional de Minería, Petróleo y Energía (2023). *Perú: Perfil Minero 2023*. Lima: Sociedad Nacional de Minería, Petróleo y Energía.

International Copper Association. (2023). *Copper in the modern world*. International Copper Association.

Fuerstenau, M. C., & Han, K. N. (2020). *Principles of mineral processing*. SME.

Wills, B. A., & Finch, J. A. (2016). *Wills' mineral processing technology: An introduction to the practical aspects of ore treatment and mineral recovery*.

Napier-Munn, T. J., & Morrell, S. (2019). *Mineral processing technology: An introduction*. Elsevier.

Xiao, Z., & Fuerstenau, M. C. (2021). *Mineral processing technology: Current trends and future directions*. SME.

Behrens, J. T., & Yu, C. H. (2020). Flexible Exploratory Data Analysis for Big Data. *Journal of Computational and Graphical Statistics*

Tukey, J. W., & Wilk, M. B. (2021). *Modern Approaches to Exploratory Data Analysis*. Data Science Journal

Zhang, H., & Rao, A. R. (2019). *Assessing Data Assumptions in Exploratory Data Analysis*. Statistical Analysis and Data Mining

Chen, Y., & Huang, J. (2021). Understanding Percentiles in Data Analysis. *Journal of Statistical Methods*

Chung, R., & Zhang, W. (2019). The Role of Mode in Data Analysis. *Statistics and Probability Letters*

Kang, H., & Ding, Y. (2020). Median as a Robust Measure of Central Tendency. *International Journal of Data Science*

Lee, S., & Kim, M. (2020). Variance and its Importance in Data Interpretation. *Applied Statistical Analysis*

Liao, Y., & Wu, Z. (2021). The Arithmetic Mean: Applications and Limitations. *Data Science Journal*

Wang, J., & Chen, X. (2020). Standard Deviation: A Comprehensive Guide. *Journal of Statistical Research*

Zhang, H., & Liu, P. (2020). Interquartile Range as a Measure of Statistical Dispersion. *Journal of Applied Statistics*

Zhou, Q., & Gao, F. (2021). Range and Its Applications in Descriptive Statistics. *Journal of Mathematics and Statistics*

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2019). *Bayesian Data Analysis* (3rd ed.)

Hastie, T., Tibshirani, R., & Friedman, J. (2020). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (2nd ed.)

Iglewicz, B., & Hoaglin, D. C. (2020). How to Detect and Handle Outliers. *Journal of Statistical Methods*

Cochran, W. G. (2020). *Sampling Techniques* (3rd ed.). Wiley.

Quinlan, J. R. (1986). Induction of decision trees. *Machine Learning*.

Breiman, L. (2001). Random forests. *Machine Learning*.

Chen, T., He, T., Benesty, M., Khotilovich, V., & Tang, Y. (2016). *XGBoost: Extreme Gradient Boosting*.

Geurts, P., Ernst, D., & Wehenkel, L. (2006). Extremely randomized trees. *Machine Learning*, 63(1).

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., ... & Liu, T. Y. (2017). *LightGBM: A highly efficient gradient boosting decision tree. In Advances in neural information processing systems*.

Khan, S., & Zubair, S. (2020). *Random Forests: A Comparative Study of Random Forest Algorithm with the C4.5 and SVM Algorithms*. *International Journal of Computer Applications*.

Natekin, A., & Knoll, A. (2013). *Gradient boosting machines, a tutorial*. *Frontiers in Neurorobotic*.

Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V., & Gulin, A. (2018). *CatBoost: unbiased boosting with categorical features*. In *Advances in neural information processing systems*.

Chicco, D., Warrens, M. J., & Jurman, G. (2021). *The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation*. *PeerJ Computer Science*.

Hastie, T., Tibshirani, R., & Friedman, J. (2017). *The elements of statistical learning: Data mining, inference, and prediction (2nd ed.)*. Springer.

Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2018). *Statistical and Machine Learning forecasting methods: Concerns and ways forward*. *PloS one*.

Lundberg, S. M., & Lee, S.-I. (2017). *A unified approach to interpreting model predictions*. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*.

Lundberg, S. M., Erion, G. G., & Lee, S.-I. (2020). *Consistent individualized feature attribution for tree ensembles*.

OSIsoft. (2020). *PI System: Overview and Capabilities*. OSIsoft, LLC.

Wang, J., Zhang, H., & Li, Y. (2021). *Data-driven predictive maintenance for smart manufacturing systems: A PI System approach*. *Journal of Manufacturing Systems*.

Bhandari, V., & Injeti, K. (2018). *Real-time data management and analytics using PI System*. *International Journal of Engineering & Technology*.

Zhou, H., Li, G., & Cheng, Z. (2019). *Data integration framework based on PI System for digital twin in manufacturing*. *Procedia CIRP*.

Ghasemi, Z., Neshat, M., Aldrich, C., Karageorgos, J., Zanin, M., Neumann, F., & Chen, L. (2023). *A hybrid intelligent framework for maximising SAG mill throughput: An integration of expert knowledge, machine learning and evolutionary algorithms for parameter optimisation*. *Advanced Engineering Informatics*.

Zhang, X., Wang, J., Li, Y., & Chen, M. (2024). *Machine learning-based performance evaluation and optimization of the grinding circuit at Canadian Royalties Inc. Minerals*.

10 ANEXOS

Tabla 14: Matriz de Consistencia

PROBLEMA	OBJETIVOS	JUSTIFICACIÓN	HIPOTESIS	VARIABLES
PROBLEMA GENERAL	OBJETIVO GENERAL	La presente investigación surge a partir de la complejidad del control operativo del circuito de Molienda-Clasificación lo cual se evidencia en una alta variabilidad del P80 influyendo negativamente en la recuperación de Cobre.	HIPÓTESIS PRINCIPAL	VARIABLES INDEPENDIENTES
Inadecuado Control Operacional de las variables operativas del circuito de Molienda - Clasificación lo cual se ve reflejado en una alta variabilidad del P80 afectando la recuperación de Cobre.	Generar un sistema de recomendaciones en línea, utilizando un modelo de Machine Learning que pueda minimizar la variabilidad del P80 del circuito de Molienda – Clasificación		La implementación de un sistema de recomendaciones para el circuito de Molienda-Clasificación, basado en un modelo de Machine Learning, permitirá optimizar el control operativo, reduciendo la variabilidad del P80 y, con ello, mitigando su impacto negativo en la recuperación de cobre.	Flujo Agua Molino SAG (m3/h) Velocidad Molino SAG (RPM) Presión Lado Alimentación Molino SAG (kPa) Presión Lado Descarga Molino SAG (kPa) Tonelaje Molino SAG (ton/h) F80 Molino SAG (pulg) Potencia Molino SAG (kw) Presión Clampling Chancadora de Pebbles (kg/cm2) Tonelaje Chancadora de Pebbles (ton/h) P80 Chancadora de Pebbles (mm) Flujo Agua chute Alimentación Molino de Bolas (m3/h) Flujo Agua Trommel Molino de Bolas (m3/h) Potencia Molino de Bolas (kw) Presión Lado Alimentación 1 Molino de Bolas (kPa) Presión Lado Alimentación 2 Molino de Bolas (kPa) Presión Lado Descarga 1 Molino de Bolas (kPa) Presión Lado Descarga 2 Molino de Bolas (kPa) Velocidad Motor 2 Molino de Bolas (RPM) WorkIndex Molienda (kwh/tm) Nivel Cajón de Bombeo (%) Flujo Agua alimento Cajón (m3/h) Presión Hidrociclones (kPa)
PROBLEMAS ESPECIFICOS	OBJETIVOS ESPECIFICOS		HIPÓTESIS ESPECIFICAS	VARIABLE DEPENDIENTE
¿Se puede predecir el P80 del circuito Molienda - Clasificación?	Obtener un modelo predictivo óptimo del P80 utilizando algoritmos y técnicas de Machine Learning.		Se puede obtener un modelo de Machine Learning predictivo del P80 el cual sea acertado a la realidad.	Tamaño de Partícula P80 (µm)
¿Se puede mejorar el control operativo de las variables que impactan al P80?	Generar un sistema de recomendación de las variables operativas a partir de la explicabilidad del modelo de Machine Learning.		El Sistema de recomendación generado por el modelo de Machine Learning brindará información acerca de las variables operativas que son más influyentes en la variabilidad del P80.	
¿Es posible tener las recomendaciones en línea?	Implementar en el PI System, un modelo de Machine Learning que permita predecir el P80 y tener el sistema de recomendación en línea.		Se puede brindar recomendaciones en tiempo real del Modelo de Machine Learning mediante el PI System, de acuerdo a las condiciones que se presenten en el proceso.	

Fuente: Elaboración Propia