

**UNIVERSIDAD TÉCNICA FEDERICO SANTA MARÍA  
DEPARTAMENTO DE INDUSTRIAS**

**ANALÍTICA AVANZADA PARA LA GESTIÓN DE ACTIVOS  
EN EL SECTOR ENERGÉTICO: EVALUACIÓN Y  
VALIDACIÓN DE SOLUCIONES ANALÍTICAS DE  
MANTENIMIENTO EN COLBÚN S.A.**

**MEMORIA PARA OPTAR AL TÍTULO DE INGENIERA CIVIL INDUSTRIAL**

**AUTOR  
GABRIELA BLANCO CORNEJO**

**PROFESOR GUÍA: DAVID GODOY RAMOS**

**CORREFERENTE: RODRIGO POYANCO**

**SANTIAGO, 2026**



## CONSTANCIA DE VALIDACIÓN Y CONFIDENCIALIDAD DE MONOGRAFÍA A REPOSITORIO ACADÉMICO

### 1.- IDENTIFICACIÓN DEL TRABAJO ACADÉMICO

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

Título del trabajo: Analítica avanzada para la gestión de activos en el sector energético de mantenimiento en Colbún S.A

Nombre del candidato(a): Gabriela Eugenia Blanco Cornejo

Carrera / Grado: Ingeniería Civil Industrial

Campus: Vitacura

Departamento: Industrias

### 2.- VALIDACIÓN DEL PROFESOR GUÍA/DIRECTOR DE TESIS

Yo, David Godoy R., en mi calidad de profesor(a) guía/director(a) del trabajo académico mencionado anteriormente **DEJO CONSTANCIA** que:

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

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

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

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

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

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

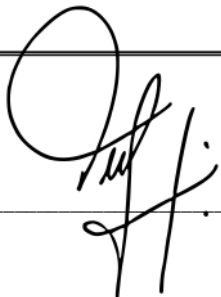
---

---


---

### 4.- FIRMAS

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

Fecha: 23-04-2026 Firma: 

Estudiante o Candidato(a):

Fecha: 16/04/2026 Firma: 

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

## Tabla de contenido

1	Resumen .....	4
2	Contexto.....	6
3	Problema de Investigación.....	8
4	Objetivos.....	9
4.1	Objetivo General .....	9
4.2	Objetivos Específicos.....	9
5	Marco Teórico .....	10
5.1	Centrales de Generación Eléctrica .....	10
5.1.1	Centrales Termoeléctricas.....	11
5.1.2	Centrales Hidroeléctricas .....	12
5.2	Mantenimiento .....	13
5.2.1	Maduración del Mantenimiento.....	13
5.2.2	Mantenimiento Basado en la Condición.....	14
5.2.3	Mantenimiento Predictivo.....	15
5.3	Modelos Predictivos.....	16
5.3.1	Metodología Backtesting .....	17
5.3.2	Algoritmos Utilizados.....	17
5.3.3	Lenguaje de Programación .....	21
5.3.4	Plataformas y Tecnologías para el Análisis de Datos.....	22
5.4	Metodología Cross-Industry Standard Process for Data Mining .....	24
6	Metodología.....	27
6.1	Business Understanding.....	27
6.2	Data Understanding.....	27
6.3	Data Preparation.....	30
6.4	Modeling .....	32
6.5	Evaluation.....	37
7	Resultados.....	39
7.1	Data Understanding.....	39
7.2	Data Preparation.....	40
7.3	Modeling .....	42
7.4	Evaluation.....	47
8	Alcance y Limitaciones .....	57
8.1	Alcance.....	57
8.2	Limitaciones .....	64
9	Conclusiones.....	66
10	Trabajos Futuros .....	68
11	Referencias .....	69

## 1 Resumen

En este trabajo se evalúa la aplicabilidad y escalabilidad de un enfoque predictivo interno, previamente implementado en activos de generación hidráulica, hacia un entorno de generación térmica, utilizando como caso de estudio la Central Nehuenco I de Colbún S.A. En particular, el trabajo se orienta a analizar la capacidad de modelación y predicción de variables críticas asociadas a fenómenos de degradación y falla, mediante el uso de datos históricos provenientes de sistemas de adquisición industrial y su procesamiento en una arquitectura analítica basada en Databricks. Con el propósito de asegurar trazabilidad y reproducibilidad, la investigación se estructura bajo el marco metodológico CRISP-DM, abarcando las etapas de comprensión del negocio, entendimiento de los datos, preparación, modelado y evaluación. El proceso incorpora una clasificación jerárquica de señales por macrosistema, sistema y subsistema, junto con un análisis de redundancia orientado a identificar sensores con comportamiento altamente correlacionado. Para este fin, se emplean métricas de asociación tales como la correlación de Pearson para determinar redundancia y un coeficiente de correlación global para seleccionar señales representativas dentro de cada grupo, reduciendo dimensionalidad y favoreciendo la interpretabilidad. Posteriormente, se construyen modelos predictivos basados en Random Forest, considerando configuraciones que comparan escenarios con señales redundantes y no redundantes, tanto a nivel de grupo como a nivel de sistema y macrosistema. La validación se desarrolla mediante backtesting, definiendo ventanas temporales de entrenamiento y predicción que permiten simular el desempeño del enfoque en condiciones operacionales comparables a un despliegue real.

Los resultados evidencian que, bajo una preparación adecuada de datos y una selección consistente de variables representativas, el enfoque presenta desempeño satisfactorio en la

predicción de variables objetivo asociadas a condiciones anómalas, permitiendo aproximar tendencias y anticipar desviaciones relevantes. En consecuencia, se concluye que la metodología analítica evaluada es transferible al contexto térmico de Nehuenco I, aportando una base para su eventual extensión hacia otros sistemas, unidades y centrales.

## 2 Contexto

En el escenario actual de la transformación digital que experimenta la industria energética, las estrategias de mantenimiento predictivo y monitoreo de condición se han consolidado como herramientas clave para la gestión eficiente de activos físicos. Estas prácticas se sustentan en la captura y análisis de datos operacionales tanto en tiempo real como históricos, con el propósito de anticipar fallas y optimizar las intervenciones. El mantenimiento predictivo se entiende como aquel que realiza un seguimiento de cada una de las variables relacionadas con el funcionamiento de las máquinas para predecir posibles fallas y tomar las acciones correctivas más apropiadas en el momento oportuno (William Olarte C, 2010). Por otro lado, el monitoreo de condición es un enfoque del mantenimiento predictivo que se basa en la recopilación en tiempo real de datos operacionales para supervisar la salud de activos y sistemas (McMiller, 2024).

Este tipo de mantenimiento es crucial para las empresas de generación eléctrica, ya que, en caso de que una generadora no pueda salir al mercado o deje de generar energía debido a una falla inesperada, podría desencadenar perjuicios económicos y/o legales (Mortera de Iruarrizaga, 2024).

Colbún S.A., a través de su Gerencia de Generación, ha demostrado un compromiso concreto con el fortalecimiento de estas capacidades, buscando maximizar la disponibilidad, confiabilidad y eficiencia operativa de todas sus plantas, de diversas tecnologías.

Bajo este contexto, la compañía ha desarrollado modelos predictivos internos basados en regresiones mediante ensambles de árboles de decisiones, los cuales han sido aplicados exitosamente como prueba de conceptos en la Central Hidroeléctrica Colbún. Siendo esta última

una central de embalse que produce energía a través de la liberación controlada de agua almacenada, destinada al accionamiento de turbinas. En paralelo, las centrales de tecnología térmica operan bajo esquemas de monitoreo de condición tercerizados, utilizando técnicas de clustering sobre datos operacionales, presentando buenos resultados en la detección de fallas críticas. Un ejemplo de la aplicación de estos esquemas se da en el Complejo Nehuenco, compuesto por las unidades de generación Nehuenco I, II y III, que utilizan principalmente gas natural como fuente de combustible, donde el primero corresponde a un ciclo combinado, que su tecnología consiste en el aprovechamiento energético de un combustible en una turbina de gas y la conversión en energía eléctrica de la energía potencial contenida en una masa de vapor a una presión y temperatura determinada (García, 2013).

En el mercado actual, el monitoreo de condición asume un rol central en la operación en línea de los activos, al posibilitar la observación continua de variables críticas. La integración de estas mediciones en plataformas de supervisión, registradores históricos de proceso y sistemas analíticos especializados permite la detección temprana de patrones anómalos, la generación de alertas basadas en umbrales dinámicos y la elaboración de indicadores de salud de equipos, los cuales sirven de insumo para la planificación del mantenimiento y la toma de decisiones en tiempo real.

Asimismo, la disponibilidad de grandes volúmenes de datos operacionales históricos favorece el desarrollo de modelos analíticos más robustos, capaces de identificar desviaciones sutiles respecto del comportamiento normal de la planta, fortaleciendo con ello la capacidad de anticipar fallas de carácter incipiente. En este contexto, el monitoreo de condición trasciende la de la observación pasiva de variables y se integra con herramientas de analítica avanzada, como lo son los modelos predictivos.

### 3 Problema de Investigación

Considerando el contexto de la industria de la generación eléctrica y específicamente el de la empresa Colbún S.A, surge la siguiente interrogante ¿Es el modelo predictivo desarrollado internamente por Colbún escalable y aplicable con precisión en un entorno térmico como Nehuenco I? Por lo tanto, se propone una evaluación exhaustiva de la metodología interna aplicada en el entorno hidroeléctrico, con el objetivo de determinar su pertinencia y escalabilidad al contexto térmico, específicamente en la Central Térmica Nehuenco I. Para ello, se desarrollan nuevos modelos predictivos basados en datos operacionales históricos provenientes de AVEVA PI, una plataforma integrada que permite la recopilación, depuración y almacenamiento de datos en tiempo real de las operaciones (AVEVA PI, 2020), los cuales serán procesados e integrados en la plataforma Databricks. La validación de estos modelos se lleva a cabo mediante backtesting, comparando sus predicciones con fallas previamente conocidas del activo, con el fin de evaluar su precisión, robustez y aplicabilidad práctica.

Desde la perspectiva de la ingeniería civil industrial, este estudio se inserta en el ámbito del mantenimiento inteligente y la analítica avanzada para la gestión de activos físicos, áreas de creciente importancia tanto en la literatura académica como en la práctica empresarial. Mediante la comparación entre metodologías internas y externas, esta investigación contribuye al cuerpo de conocimiento en la toma de decisiones estratégicas sobre la internalización de capacidades analíticas, evaluando no solo la viabilidad técnica sino también el valor económico potencial para la compañía. Surge así una segunda interrogante: ¿Los resultados obtenidos respaldan el levantamiento de un caso de negocio para la implementación permanente de los modelos? Ya que, en caso de obtener resultados favorables, la implementación abriría el camino hacia una operación más autónoma, eficiente y fundamentada en datos.

## 4 Objetivos

### 4.1 Objetivo General

Desarrollar un modelo predictivo para la Central Térmica Nehuenco I, a partir de la evaluación y adaptación de los modelos previamente implementados en Colbún en otras centrales, con el fin de determinar su viabilidad técnica y estratégica como herramienta de mantenimiento predictivo y monitoreo de condición en plantas térmicas.

### 4.2 Objetivos Específicos

Desarrollar modelos predictivos utilizando datos históricos de la central térmica Nehuenco I, integrando y procesando la información, con el fin de replicar la metodología utilizada en las centrales hidroeléctricas de Colbún S.A.

Validar el desempeño de los modelos predictivos a través de un proceso de backtesting, comparando las predicciones de fallas con las incidencias históricas conocidas en la central térmica Nehuenco I, evaluando su capacidad para capturar fallas críticas.

Determinar la viabilidad técnica de implementar los modelos predictivos de manera permanente en la central térmica Nehuenco I, evaluando su escalabilidad y la potencial mejora en la eficiencia operativa.

## 5 Marco Teórico

Con el fin de comprender el problema planteado, resulta fundamental abordar con precisión ciertos conceptos clave, tales como la composición de las centrales de generación termoeléctricas y hidroeléctricas, el mantenimiento predictivo y los modelos que lo sustentan, el monitoreo de condición, así como el funcionamiento de las plataformas de gestión de datos disponibles en el mercado.

### 5.1 Centrales de Generación Eléctrica

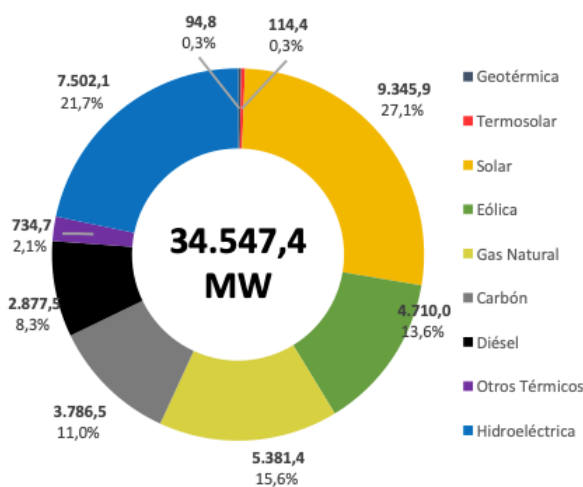
El sistema de suministro eléctrico está conformado por tres actores principales: las empresas generadoras, transmisoras y distribuidoras. Las centrales de generación tienen la capacidad de transformar la energía mecánica, obtenida a partir de la conversión de la energía potencial de distintas fuentes primarias, en energía eléctrica. Estas centrales pueden definirse como un conjunto de equipos destinados a la producción de energía eléctrica, cuyo componente principal es el generador, siendo su tipo el que determina la naturaleza de la planta. Las transmisoras se dedican al transporte de la energía eléctrica en alta tensión desde las plantas de generación hasta las subestaciones, con el propósito garantizar la continuidad y seguridad del suministro eléctrico. Por su parte, las distribuidoras se encargan de transportar la energía en niveles de media y baja tensión desde las subestaciones hasta los usuarios finales.

Actualmente, el Sistema Eléctrico Nacional (SEN) posee 34.547,4 MW de capacidad instalada (potencia máxima bruta) para generar energía eléctrica, de los cuales 15.516,8 MW corresponden a centrales de Energía Renovable No Convencional. (Coordinador Eléctrico Nacional, 2024)

En la Figura 1, se puede apreciar la composición del Sistema Eléctrico Nacional por tecnología:

## Figura 1

Capacidad según tecnología



Nota. Tomado de Reporte Coordinador Eléctrico Nacional (2024)

### 5.1.1 Centrales Termoeléctricas

Una central termoeléctrica es una instalación donde la energía mecánica que se necesita para mover el rotor del generador, y por tanto obtener la energía eléctrica, se obtiene a partir del vapor formado al hervir el agua en una caldera. El vapor generado tiene una gran presión, y se hace llegar a las turbinas para que en su expansión sea capaz de mover los álabes de las mismas (Foro nuclear, 2025).

Las centrales termoeléctricas generan energía eléctrica a partir de la transformación de combustibles fósiles, donde se distinguen dos tipos principales: las de ciclo simple y las de ciclo combinado. Las primeras operan mediante cuatro componentes fundamentales: la caldera, donde el agua se convierte en vapor mediante la quema del combustible; los serpentines, que corresponden a las cañerías por las cuales circula el agua en proceso de vaporización; la turbina de vapor, que se acciona por el impulso del vapor de agua; y el generador, encargado de transformar la energía mecánica del eje de la turbina en energía eléctrica. Las centrales de ciclo

combinado incorporan los mismos elementos del ciclo simple, pero además integran una turbina de combustión interna, compuesta por un compresor que introduce aire a presión para la combustión del gas, una cámara de combustión donde se mezcla el aire con el gas, una turbina que se activa mediante la expansión de los gases resultantes y una caldera de recuperación, donde ocurre un ciclo agua-vapor.

### 5.1.2 Centrales Hidroeléctricas

Una central hidroeléctrica es una instalación cuya misión es convertir la energía potencial y cinética del agua en energía eléctrica disponible (Sanz, 2016).

Existen tres tipos de centrales hidroeléctricas según la forma en que aprovechan las masas de agua que fluyen por los ríos. Las centrales de embalse almacenan agua y la liberan de manera controlada para hacer girar las turbinas. Las centrales de pasada aprovechan el caudal natural del río. Las centrales de bombeo reversible operan mediante dos embalses a distinta altura, bombeando el agua al embalse superior y dejándola descender posteriormente hacia el embalse inferior.

Respecto a sus componentes, las centrales del primer tipo cuentan con una presa, estructura encargada de contener y almacenar el agua. En tanto, elementos como los rebosaderos, que permiten la liberación de agua sin que esta pase por la sala de máquinas; los destructores de energía, que reducen la energía del agua; la turbina, que transforma la energía cinética en energía mecánica; el alternador, que convierte la energía mecánica en energía eléctrica; las conducciones, que son los canales o tuberías por donde circula el agua; las válvulas, que regulan su flujo; y las chimeneas de equilibrio, que estabilizan la presión en el sistema, están presentes en los distintos tipos de centrales hidroeléctricas.

## 5.2 Mantenimiento

El mantenimiento es el conjunto de actividades técnicas y administrativas que se realizan para garantizar el correcto funcionamiento de los equipos e instalaciones.

### 5.2.1 Maduración del Mantenimiento

Para comprender de mejor manera la evolución del mantenimiento en una empresa, se utiliza la siguiente pirámide. En ella se señala la maduración del mantenimiento, que es la evolución progresiva que se caracteriza por pasar de un enfoque reactivo a uno proactivo, predictivo y más estratégico.

**Figura 2**

*Maduración del Mantenimiento*



*Nota.* Tomado de Grupo ORS (2023)

El primer nivel corresponde al mantenimiento reactivo, también conocido como mantenimiento correctivo no planificado. Su enfoque se basa en intervenir únicamente cuando se produce una falla, actuando una vez que el problema ya ha ocurrido, sin medidas preventivas previas. Debido a su naturaleza es, una actitud pasiva, frente a la evolución del estado de los equipos, a la espera de la avería o fallo (Gómez de León, 1998).

El mantenimiento preventivo es una estrategia de mantenimiento planificada que establece intervenciones periódicas a los activos antes de una falla. Pretende disminuir o evitar la reparación mediante una rutina de inspección periódica y la renovación de los elementos deteriorados, lo que se conoce como “las tres erres del mantenimiento” (Gómez de León, 1998).

Posteriormente se encuentra el mantenimiento basado en la condición, que se basa en la evaluación del estado real de un activo mediante el seguimiento de ciertas variables, con el fin de detectar desviaciones de manera temprana.

El penúltimo nivel de maduración es el mantenimiento predictivo, que busca anteponerse a una falla mediante el análisis del estado real de funcionamiento, donde las intervenciones a los activos solo ocurren cuando una falla se considera inminente o probable.

Finalmente, está el mantenimiento basado en el riesgo, que se caracteriza por evaluar la probabilidad de ocurrencia de las fallas y el impacto de dichas fallas, de manera tal que se priorizan las tareas y los recursos hacia los activos que presentan un mayor riesgo con consecuencias mayores. Su procedimiento se entiende como, primero se determinan los posibles modos de fallo del sistema. Se establece que, se trata de optimizar tanto la gestión como la propia labor de mantenimiento de una instalación, minimizando todos los riesgos y abordando todas las contingencias (Facility Management Services, 2023).

### 5.2.2 Mantenimiento Basado en la Condición

El monitoreo de condición corresponde a la medición de una variable física que se considera representativa de la condición del equipo y su comparación con valores que indican si el equipo está en buen estado o deteriorado (Amendola, 2025).

El monitoreo de condición también puede entenderse como una técnica previa al mantenimiento predictivo que consiste en la supervisión continua de parámetros específicos de la maquinaria, como vibraciones y temperatura, con el fin de identificar cambios que puedan

indicar el desarrollo de una posible falla. Este proceso fundamental permite programar las intervenciones de mantenimiento de manera oportuna y tomar medidas preventivas.

En las centrales de generación eléctrica, esta técnica se utiliza para evaluar el estado de generadores, turbinas, motores, transformadores y otros equipos críticos del proceso de generación. Entre los análisis más utilizados en la industria se encuentran el análisis de vibraciones para detectar desbalances, el monitoreo de temperatura para identificar puntos calientes, el análisis de la lubricación para evaluar desgastes internos y el ultrasonido para detectar fugas o problemas eléctricos.

Para llevar a cabo el monitoreo, se emplean sensores especializados y sistemas de adquisición de datos que favorecen la recolección y análisis de información en tiempo real. Para optimizar el análisis, estos sistemas suelen integrarse con plataformas de gestión de mantenimiento.

### 5.2.3 Mantenimiento Predictivo

Dentro de las estrategias de mantenimiento, se distinguen principalmente dos tipos: el mantenimiento preventivo, que se centra en la realización de intervenciones programadas para prevenir fallas, y el mantenimiento predictivo, que se basa en la supervisión continua del estado de los activos. A diferencia del mantenimiento preventivo, el mantenimiento predictivo utiliza técnicas avanzadas como análisis de vibraciones, termografía, ultrasonido y monitoreo en tiempo real para detectar anomalías tempranas y anticipar posibles fallas. Esto permite optimizar la programación de las intervenciones, reducir costos operativos, minimizar tiempos de inactividad no planificados y aumentar la vida útil de los equipos. Así, el mantenimiento predictivo se posiciona como una herramienta clave para mejorar la confiabilidad, seguridad y eficiencia operativa en diversas industrias.

El mantenimiento predictivo se puede definir como el seguimiento organizado con

medición periódica o continua de variables de estado del sistema y su comparación con unos patrones preestablecidos, para la determinación del instante en que se debe producir la intervención de mantenimiento. Por tanto, consiste en subsanar el fallo cuando éste aún se encuentra en estado incipiente (Manés, 1998).

#### *5.2.3.1 Ventajas del Mantenimiento Predictivo:*

Existen múltiples ventajas asociadas a la implementación del mantenimiento predictivo, especialmente relevantes para el sector energético. En primer lugar, permite reducir los costos operativos, ya que al anticipar fallas se evitan reparaciones mayores y detenciones no planificadas que pueden generar sanciones económicas significativas. Además, incrementa la vida útil de los activos, al mantenerlos en condiciones óptimas mediante intervenciones oportunas basadas en datos reales y análisis continuos. Otra ventaja es la optimización de la planificación de las actividades de mantenimiento, dado que el conocimiento preciso del estado de los equipos posibilita programar intervenciones en momentos estratégicos que minimizan el impacto sobre la generación y el suministro eléctrico. Finalmente, favorece una mayor confiabilidad operativa, al reducir la frecuencia y gravedad de fallas inesperadas, aspecto crucial para mantener la estabilidad y continuidad del sistema eléctrico.

### **5.3 Modelos Predictivos**

Corresponde a una herramienta que utiliza técnicas de análisis de datos y estadísticas para analizar datos históricos y en base a estos, realizar predicciones del futuro. En la industria, específicamente en el mantenimiento predictivo, se emplean dichos modelos para prever el comportamiento de un equipo y sus componentes, de manera tal de contribuir en la optimización de la planificación del mantenimiento y reducir los costos asociados a detenciones de la producción imprevistas. Dicha explicación, se sustenta con la siguiente definición, el análisis predictivo es una rama de análisis avanzados que hace predicciones sobre resultados futuros

utilizando datos históricos combinados con modelado estadístico, técnicas de extracción de datos y aprendizaje automático (IBM, 2025).

### 5.3.1 Metodología Backtesting

Es una técnica que busca evaluar la efectividad, robustez y desempeño de un modelo, sistema o estrategia mediante la comparación con los datos históricos. El backtesting implica probar una estrategia frente a datos históricos para ver como funciona en diferentes condiciones de mercado. Este proceso le permite identificar patrones de comportamiento, validar parámetros y ajustar configuraciones para mejorar la consistencia de los resultados (Souza, 2025).

Por lo que, se entiende el Backtesting como una técnica de validación empírica que evalúa el rendimiento de un modelo. A través de esta metodología, se simula el comportamiento de una variable bajo condiciones pasadas, con el propósito de ajustar y optimizar el modelo, permitiendo así una mejora en su capacidad predictiva.

Aunque esta validación no asegura el éxito futuro del modelo, proporciona una base sólida para la toma de decisiones. Su utilidad radica en la posibilidad de contrastar distintos modelos predictivos, transformando los datos históricos en una herramienta de validación cuantitativa que permite mitigar riesgos antes de implementar los modelos en escenarios reales.

### 5.3.2 Algoritmos Utilizados

Los modelos predictivos utilizan diversos algoritmos que permiten analizar datos históricos y presentes, para así anticiparse a los eventos futuros más probables.

Los algoritmos pueden clasificarse según la supervisión de su aprendizaje, es decir, algoritmos de aprendizaje supervisado, algoritmos de aprendizaje no supervisado y algoritmos de aprendizaje por esfuerzo. Respecto a los primeros, estos buscan identificar las entradas y salidas utilizando un conjunto de datos etiquetado, algunos ejemplos de estos son las regresiones lineales, regresiones logísticas, máquinas de soporte vectorial, k-vecinos más cercanos, random

forest y los árboles de decisión.

En el aprendizaje supervisado, los algoritmos trabajan con datos “etiquetados” (*labeled data*), intentado encontrar una función que, dadas las variables de entrada (*input data*), les asigne la etiqueta de salida adecuada. El algoritmo se entrena con un “histórico” de datos y así “aprende” a asignar la etiqueta de salida adecuada a un nuevo valor, es decir, predice el valor de salida (Telefónica Tech, 2021).

Respecto a los algoritmos de aprendizaje no supervisado, estos no requieren de etiquetar los datos, son comúnmente utilizados para identificar patrones y estructuras ocultas en los datos sin aplicar supervisiones explícitas, y algunos de estos son los análisis de componentes principales, K-Means, DBSCAN y Autoencoders.

El aprendizaje no supervisado tiene lugar cuando no se dispone de datos “etiquetados” para el entrenamiento. Sólo conocemos los datos de entrada, pero no existen datos de salida que correspondan a un determinado input. Por tanto, sólo podemos describir la estructura de los datos, para intentar encontrar algún tipo de organización que simplifique el análisis. Por ello, tienen un carácter exploratorio (Telefónica Tech, 2021).

Finalmente, en relación con los algoritmos de aprendizaje por esfuerzos, estos aprenden a través de la interacción con el entorno, por lo que se aplica un sistema de recompensas según su desempeño en el entorno. Algunos de los tipos más comunes son el Q-Learning y las Deep Q-Learning.

Este tipo de aprendizaje se basa en mejorar la respuesta del modelo usando un proceso de retroalimentación. El algoritmo aprende observando el mundo que le rodea. Su información de entrada es el feedback o retroalimentación que obtiene del mundo exterior como respuesta a sus acciones. Por lo tanto, el sistema aprende a base de ensayo-error (Telefonica Tech, 2021).

### 5.3.2.1 *Análisis Exploratorio de Datos*

El análisis exploratorio de datos, conocido como EDA por sus siglas en inglés (Exploratory Data Analysis), puede considerarse la versión moderna de la Estadística Descriptiva, teniendo como objetivo la organización, representación y descripción de los datos, como pasos necesarios para extraer la información que contienen (Vilar, 2005).

Es importante señalar que no corresponde a un algoritmo de machine learning, sino más bien es la herramienta que permite comprender las relaciones, comportamientos y estructuras presentes en un conjunto de datos antes de aplicar modelos más complejos. Su enfoque se basa en la formulación de preguntas orientadas a identificar patrones relevantes, sin buscar confirmar hipótesis, sino más bien interpretar y organizar la información disponible. A partir de este proceso, es posible depurar los datos, entender su composición y orientar la selección del modelo más adecuado para un análisis posterior. Por lo que, se puede entender como una herramienta que prepara la información para volverla una base de datos sin errores, datos faltantes y/o inconsistencias, de manera tal que se puedan construir análisis y modelos más complejos, precisos y confiables.

### 5.3.2.2 *Árboles de Decisión*

El aprendizaje mediante árboles de decisión es un método de aproximación de una función objetivo de valores discretos en el cual la función objetivo es representada mediante un árbol de decisión (Alonso, 2025).

Además, los árboles de decisión son un algoritmo de aprendizaje supervisado no paramétrico utilizado para la clasificación y la regresión. El objetivo es crear un modelo que prediga el valor de una variable objetivo mediante la aplicación de reglas de decisión simples provenientes del análisis de las características de los datos y su comportamiento, destaca por su facilidad de comprensión e interpretación, ya que representa las decisiones en una estructura

que se asemeja a un árbol, donde la raíz representa una condición, para luego nacer ramas donde cada nodo interno representa una decisión sucesiva, hasta llegar a las hojas que corresponden a la predicción.

#### 5.3.2.3 *Random Forest*

Es un algoritmo de aprendizaje supervisado basado en los árboles de decisión, ya que su construcción se basa en la elaboración de múltiples árboles de decisión que se combinan con el objetivo de realizar una predicción más certera, donde se destaca su resistencia al sobreajuste. Basado en la literatura, los árboles aleatorios reciben este nombre debido a que cada árbol es entrenado a partir de un subconjunto de muestras de datos seleccionadas aleatoriamente y cada nodo toma un subconjunto aleatorio de atributos al determinar la mejor separación (Pineda, 2022).

Para lograr una mejor comprensión, resulta clave considerar el concepto de Bagging, abreviación de Bootstrap Aggregating, el cual corresponde a una técnica de ensamblado orientada a mejorar la precisión y estabilidad de los modelos predictivos. Esta metodología entrena múltiples modelos de forma independiente, utilizando distintos subconjuntos extraídos del conjunto de datos original, con reposición, y luego combina sus resultados. Este enfoque adquiere especial relevancia al considerar que el modelo Random Forest se construye sobre la lógica del Bagging, debido a su selección aleatoria de datos y subconjunto aleatorio de atributos.

#### 5.3.2.4 *Coefficiente de Correlación Global (GCC):*

Se entiende como una medida de dependencia global que cuantifica en qué grado una variable puede ser explicada, de manera conjunta, por el resto de las variables que componen un sistema multivariado. En términos generales, permite evaluar la relación global de una variable con todas las demás, capturando la proporción de variabilidad que dicha variable comparte con la estructura completa del sistema.

Desde esta perspectiva, puede interpretarse como un indicador de redundancia y colinealidad, donde valores elevados del GCC sugieren que la variable presenta una fuerte correlación con el conjunto y, en consecuencia, puede considerarse un representante válido de un espacio de información compartido. Por el contrario, valores bajos indicarían una mayor singularidad o aporte específico de información.

En la práctica, el GCC suele utilizarse como herramienta complementaria al análisis tradicional de covarianzas y correlaciones de Pearson, especialmente en contextos donde existe multicolinealidad entre variables. En este tipo de estructuras, su aplicación resulta útil para el diagnóstico y la selección de variables en modelos estadísticos y de aprendizaje automático, contribuyendo a reducir redundancias, mejorar la estabilidad de las estimaciones y facilitar la interpretación de los resultados.

#### 5.3.2.5 *Métrica de Importancia*

Medida que cuantifica la relevancia relativa de cada variable explicativa en la construcción de un modelo, en función del aporte que realiza a la calidad de la explicación o predicción de la variable objetivo. Bajo esta lógica, una variable posee mayor importancia en la medida que su inclusión permite reducir el error, mejorar el ajuste o aumentar la capacidad discriminante del modelo frente a las alternativas disponibles.

Por ejemplo, en los modelos de Random Forest, la importance se calcula a partir del efecto que tiene cada variable sobre la reducción de impureza o de varianza del error en los nodos donde es utilizada para particionar los datos. De este modo, una importancia elevada indica que la variable contribuye de manera sistemática a mejorar el desempeño predictivo, mientras que valores cercanos a cero reflejan un aporte marginal.

#### 5.3.3 *Lenguaje de Programación*

En el contexto de la ciencia de datos, la programación cumple un rol esencial en la manipulación de la información, la implementación de modelos y la automatización de procesos

analíticos. Diversos lenguajes son utilizados en la industria, entre los cuales destacan R, Python, JavaScript y SQL, cada uno con características que responden a distintos requerimientos dentro del ciclo de análisis de datos.

Para el desarrollo del presente informe se emplea el lenguaje de programación Python, ya que corresponde al lenguaje base de las plataformas utilizadas. Este lenguaje se caracteriza por su simplicidad y claridad, lo que facilita su comprensión y aplicación. Python es considerado un lenguaje completo, con una amplia variedad de librerías que permiten acceder a bases de datos, procesar archivos y construir interfaces gráficas. Una de sus principales ventajas radica en la posibilidad de lograr resultados eficientes con pocas líneas de código, lo que contribuye a una mayor productividad y a una menor probabilidad de errores por parte del programador.

Una gran ventaja de Python es la legibilidad de los programas escritos en ese lenguaje. Otros lenguajes de programación utilizan innumerables marcaciones, como punto (.) o punto y coma (;) en el final de cada línea, además de los marcadores de comienzo y fin de bloque como llaves ({} ) o palabras especiales (begin/end). Esos marcadores vuelven a los programas un tanto más difíciles de leer y por suerte no son usados en Python (Menezes, 2017).

#### 5.3.4 Plataformas y Tecnologías para el Análisis de Datos

Las plataformas de análisis de datos son herramientas diseñadas para procesar, analizar y visualizar grandes volúmenes de información, permitiendo transformar datos crudos de las mediciones en entradas valiosas para los modelos predictivos, favoreciendo la toma de decisiones de manera informada. En el ámbito industrial, esto se traduce en la optimización de procesos, mantenimiento predictivo, mejora en la calidad del producto, y en la toma de decisiones basadas en datos.

Dentro de las principales funciones de dichas plataformas, se encuentra la recopilación de datos, almacenamiento de información, procesamiento de información, análisis avanzado,

visualización de datos y toma de decisiones.

Las plataformas de Machine Learning y Data Science, son utilizadas para crear modelos predictivos y análisis avanzados usando técnicas de machine learning e inteligencia artificial. Estas plataformas permiten entrenar modelos con datos históricos para hacer predicciones sobre comportamientos futuros, como fallas de equipos o patrones de consumo.

#### *5.3.4.1 AVEVA PI*

Existen también plataformas de apoyo en el análisis predictivo, como lo es AVEVA PI System, que es capaz de recoger datos de forma continua, identificándolos como tags únicos, desde una amplia variedad de dispositivos industriales, permitiendo que los modelos predictivos sean alimentados con información actualizada de manera constante. Concretamente se define como una cartera integrada de soluciones que permite a las operaciones industriales recopilar, limpiar, almacenar, enriquecer y visualizar datos operativos en tiempo real, con total conformidad en materia de TI. AVEVA PI System facilita a los operadores, gerentes de planta, analistas de datos y ejecutivos optimizar la eficiencia, garantizar la resiliencia e impulsar la sostenibilidad (AVEVA, 2025).

#### *5.3.4.2 Databricks*

Se trata de una plataforma de análisis de datos en la nube que ofrece un entorno unificado para procesar información de forma rápida y escalable. Permite construir, entrenar y desplegar modelos de machine learning, así como analizar datos en tiempo real y colaborar mediante notebooks de programación. Se conoce que, una plataforma de Inteligencia de Datos de Databricks permite a toda su organización utilizar datos e IA. Está construida sobre un sistema de almacenamiento en la nube (lakehouse) para proporcionar una base abierta y unificada para todos los datos y su gobernanza, y se basa en un Motor de Inteligencia de Datos que comprende la singularidad de sus datos (Databricks, 2025).

La plataforma busca entregar una experiencia inteligente, combinando inteligencia

artificial generativa con las capacidades de la nube para conformar un motor de inteligencia de datos eficiente, lo que permite optimizar el rendimiento en los procesos analíticos. Al mismo tiempo, se orienta a la simplicidad, ya que el motor es capaz de interpretar el lenguaje utilizado dentro de la organización y ofrece asistencia en la escritura de código, además de facilitar la detección y corrección de errores.

Una de las herramientas más útiles de esta plataforma para la elaboración de modelos predictivos corresponde a los experimentos, los cuales, en el entorno de Databricks, constituyen la unidad básica de organización del proceso de modelamiento. Cada experimento agrupa las distintas ejecuciones asociadas a un mismo problema, registrando para cada run los parámetros utilizados, las métricas de desempeño obtenidas y los artefactos generados, tales como tablas de resultados o modelos entrenados.

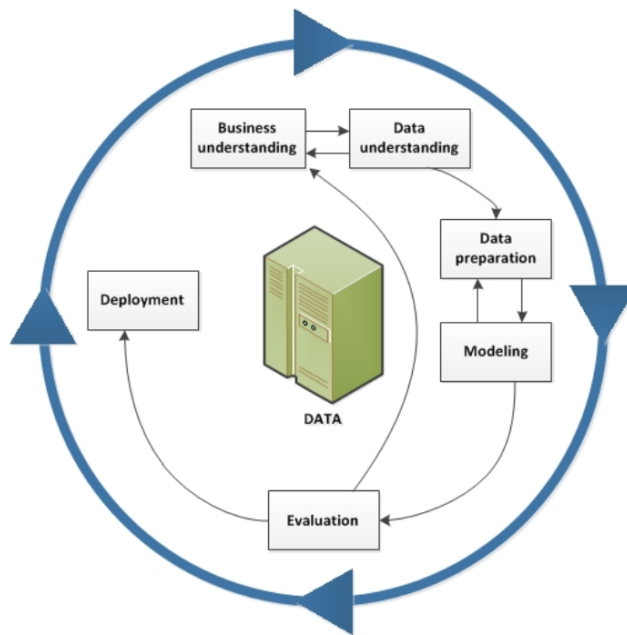
Esta estructura permite asegurar la trazabilidad y reproducibilidad de los modelos, facilitar la comparación sistemática entre configuraciones alternativas y respaldar la selección del modelo final en función de criterios cuantitativos. Adicionalmente, los experimentos se integran con el registro de modelos, lo que habilita una transición controlada desde la etapa de análisis hacia el despliegue de los modelos en ambientes operacionales.

## 5.4 Metodología Cross-Industry Standard Process for Data Mining

La metodología Crisp DM busca gestionar los trabajos de minería de datos, de manera tal que exista una estructura definida de como abordar los problemas y trabajar la información en distintas fases.

### **Figura 3**

*Fases Metodología Crisp-DM*



*Nota. Tomado de IBM SPSS Modeler*

La fase inicial, denominada Business Understanding, corresponde a la comprensión del negocio. En ella se realiza un análisis exhaustivo para entender la problemática en su totalidad, identificar oportunidades, objetivos estratégicos e indicadores de éxito, y culminar con una formulación clara del problema de investigación y de los objetivos que permiten abordarlo.

La fase siguiente, denominada Data Understanding, corresponde a la comprensión de los datos. En esta etapa se analiza el tipo de información disponible, se accede a las fuentes y se explora el conjunto de datos para identificar su volumen, su organización, la naturaleza de la información y su calidad, con el propósito de detectar patrones, inconsistencias y formular hipótesis.

Posteriormente, la fase Data Preparation corresponde a la transformación de la información en datos utilizables mediante procesos de limpieza, normalización y estandarización, ingeniería de variables, integración de fuentes e identificación de periodos para entrenamiento, validación y prueba. El resultado es un conjunto de datos lo más apropiado posible para la etapa de modelado.

La fase de Modeling, se basa el entrenamiento de los modelos según el tipo de problema que se está enfrentando, donde utilizando la data limpia, se seleccionan los algoritmos candidatos, se ajustan los hiperparámetros y se comparan con métodos de evaluación.

Finalmente, la fase Evaluation comprueba el desempeño de los modelos y verifica el cumplimiento de los criterios definidos, analizando los errores y sus impactos, para resolver la viabilidad de la implementación. A continuación, la fase Deployment establece la arquitectura de inferencia, define el monitoreo del desempeño y prepara la documentación oficial que certifica el procedimiento desarrollado y sus resultados.

## 6 Metodología

Para poder alcanzar los objetivos de la presente memoria, se elabora una metodología concisa que se compone de una serie de pasos que toman como referencia las etapas de la metodología Crisp-DM.

### 6.1 Business Understanding

Se orienta a alcanzar una comprensión del negocio de tal nivel que posibilite identificar problemáticas y deficiencias existentes, con el fin de seleccionar correctamente la problemática a abordar y establecer objetivos adecuados para enfrentarla. Este paso se desarrolla principalmente al iniciar con el desarrollo de la memoria, ya que la problemática seleccionada proviene de una problemática que ya se encontraba siendo abordada por Colbún S.A para las centrales de generación hidroeléctrica, donde el éxito de dicho proyecto genera la interrogante de la viabilidad de un proyecto similar en centrales de mayor complejidad como lo son las centrales de generación termoeléctricas. Por lo que, en base a esto, se selecciona el problema de investigación y tras un análisis exhaustivo se elaboran los objetivos detallados en el inciso objetivos.

### 6.2 Data Understanding

Esta fase de la metodología se compone de una serie de tareas orientadas a comprender la información disponible, con el fin de identificar carencias y determinar las modificaciones necesarias para su adecuado tratamiento.

#### 1. Comprender la Central Nahuenco I.

Es necesario estudiar los distintos procesos que componen el gran proceso de la

generación eléctrica en una central térmica de ciclo combinado como lo es Nehuenco I, con el objetivo de entender el flujo de energía/material donde se localizan los sensores.

Entender la forma en que se almacena la información asociada a los sensores de la planta, la cual se registra mediante distintos tags que se almacenan en AVEVA PI, plataforma en la que es posible visualizar toda la información relacionada con cada tag, como la familia de árboles a la que pertenece, el tipo de valor que entrega y la medición propiamente tal.

## 2. Analizar la información disponible sobre los distintos tags de sensores.

Recopilar la información disponible sobre los tags, se analiza su comportamiento en distintas ventanas de tiempo y se verifica su conexión actual en AVEVA PI. Dicha información se encuentra almacenada tanto en AVEVA PI como en diversas planillas gestionadas por la central, las cuales registran la cantidad de tags asociados a Nehuenco I, junto con antecedentes relevantes vinculados a estos. En relación con el comportamiento en distintas ventanas de tiempo, sus mediciones se deben realizar empleando la plataforma Databricks.

## 3. Clasificar los Tags.

Para trabajar de manera más efectiva con los distintos tags, resulta fundamental clasificarlos en tres niveles de información, de acuerdo con su ubicación en la central, el proceso al que pertenecen y la naturaleza de la señal que entregan.

En particular, los criterios corresponden a Macrosistema, Sistema y Subsistema, proceso que se desarrolla bajo la supervisión de especialistas de la central. El criterio de Macrosistema representa grandes ensambles de componentes que, en su conjunto, conforman un gran componente reconocido en la planta. El criterio de Sistema se orienta a identificar de manera más específica los componentes y sistemas que integran dicho macrosistema. Finalmente, el criterio de Subsistema busca describir con mayor detalle la información que entrega el tag,

indicando de forma explícita qué se está midiendo en la posición correspondiente. Estos criterios se conciben como un desglose jerárquico, donde la lógica establece que dentro de cada macrosistema existe un conjunto de sistemas y, dentro de cada uno existen distintos subsistemas.

4. Identificar fechas de detención en los últimos 2 años.

Verificar la existencia de períodos asociados a detenciones de la central, tales como mantenimientos mayores, mantenimientos menores e intervenciones de emergencia. Para ello, se contrasta el registro de fechas oficiales de mantenimiento con el comportamiento de la potencia, identificando tramos en los que esta se reduce por debajo de la potencia mínima técnica de 120 MV o se mantiene en valores cercanos a cero durante intervalos continuos de tiempo, lo que permite confirmar y caracterizar las detenciones efectivamente ocurridas.

5. Seleccionar falla a modelar desde el sistema ERP de la compañía.

De manera específica, desde SAP se realiza un análisis detallado de las fallas ocurridas en los últimos dos años, revisando las órdenes y notificaciones asociadas a cada evento. En este proceso, se evalúa el nivel de información disponible (descripción de la falla, causa raíz registrada, actividades ejecutadas y tiempos de intervención), su recurrencia en el tiempo y el impacto que genera en la operación de la central, tanto en términos de disponibilidad como de potencia no suministrada. A partir de estos criterios, se procede a priorizar y seleccionar la falla que presenta mayor relevancia técnica y operativa para ser utilizada como caso de estudio en el modelo a desarrollar.

6. Seleccionar periodo de entrenamiento.

En función de la falla seleccionada y de los períodos de detención previamente identificados, se define el período de entrenamiento del modelo, estableciendo una frecuencia de extracción de datos de 10 minutos. Para ello, se selecciona un horizonte temporal que incluya

suficiente operación normal previa a la falla, de manera que el modelo pueda captar tanto el comportamiento estándar de la central como las variaciones asociadas al deterioro o a condiciones anómalas. Adicionalmente, se procura que el período no cuente con tramos de datos faltantes, inconsistentes o afectados por intervenciones operativas no relacionadas con la falla en estudio, siendo así como se establece el periodo sano para entrenamiento.

#### 7. Seleccionar periodo de predicción:

Resulta fundamental definir con precisión el periodo en el cual se evalúa el desempeño de la predicción, el cual se divide en dos tramos. En una primera instancia, se busca predecir en un periodo posterior al entrenamiento, pero cercano a la falla, con el propósito de analizar el comportamiento de las predicciones en un contexto previo al evento. Posteriormente, se considera un segundo periodo de predicción que ya incorpora el intervalo en el que ocurre la falla, lo que permite evaluar la capacidad del modelo para anticipar y representar adecuadamente dicho evento.

### 6.3 Data Preparation

Se entiende como la fase de transformación de la información en datos utilizables.

#### 1. Extraer datos de los tags relacionados a la falla.

Se elabora un código en Databricks que extrae los datos del período de entrenamiento, con una resolución de 10 minutos en la fecha previamente identificada como sana, para todos los tags que comparten el mismo nivel de macrosistema que el tag asociado a la falla. La información se filtra considerando únicamente los períodos en que la potencia es superior a la potencia técnica mínima y se eliminan los valores nulos presentes en el conjunto de datos.

#### 2. Identificar Redundancias.

Existen sensores ubicados en posiciones muy cercanas entre sí, lo que da lugar a la existencia de tags que entregan información prácticamente equivalente y, por tanto, redundante.

Para su identificación, se elaboran pairplots en Databricks, agrupando los tags que coinciden en todos sus niveles de clasificación (macrosistema, sistema y subsistema) y se calcula el coeficiente de correlación de Pearson con el fin de determinar los tags redundantes.

Como criterio, se considera redundante todo tag cuyo coeficiente de correlación presenta un valor superior a 0,98

### 3. Identificar tags representativos en grupos de redundantes.

Para los tags que resultan redundantes entre sí, se selecciona aquel que presenta mayor representatividad mediante la aplicación del coeficiente de correlación global (GCC). Este indicador permite evaluar, para cada tag, el grado de asociación que mantiene con el resto de los tags del grupo redundante, considerando de manera conjunta sus correlaciones. De este modo, se identifica el tag que mejor sintetiza la información compartida por el conjunto, el cual se debe identificar para el modelamiento.

### 4. Eliminación de variables.

Se excluye todo tag cuya naturaleza corresponda a una variable categórica binaria, con valores exclusivamente 0 y 1, debido a que este tipo de señal representa estados lógicos (por ejemplo, encendido/apagado o abierto/cerrado) más que magnitudes físicas continuas. En consecuencia, su variabilidad es limitada y su aporte al análisis estadístico y a los modelos de regresión orientados a variables continuas resulta reducido, pudiendo incluso introducir sesgos en la estimación de correlaciones y en la interpretación de la importancia de variables. Adicionalmente, la inclusión de estas variables binarias puede dificultar la identificación de relaciones finas entre las variables de proceso, por lo que se opta por concentrar el análisis en

tags que entregan mediciones continuas, más coherentes con los objetivos del modelo.

## 6.4 Modeling

Consiste en la fase de entrenamiento y predicción de los modelos con la data resultante de la fase anterior para el macrosistema Turbina a Gas y sistema Compresor, utilizando algoritmos específicos con sus respectivos hiperparámetros.

### 1. Seleccionar algoritmo a utilizar.

Se selecciona el algoritmo Random Forest debido a las características propias de los datos utilizados y por los objetivos del modelamiento. En primer lugar, los registros corresponden a datos tabulares multivariados provenientes de múltiples tags de sensores, con relaciones no lineales e interacciones entre variables de proceso, lo que hace poco apropiados los modelos lineales tradicionales. Random Forest, al estar basado en ensambles de árboles de decisión, permite capturar de forma flexible dichas relaciones sin requerir supuestos estrictos de linealidad o normalidad. En segundo lugar, este algoritmo presenta un desempeño robusto frente a ruido y valores atípicos, condición habitual en datos operacionales de plantas de generación. Adicionalmente, no exige un preprocesamiento complejo en términos de escalado de variables, a pesar de la existencia de distintas unidades físicas.

Considerando la cantidad de tags asociados al macrosistema turbina a gas y el volumen de datos válidos disponibles para cada uno de ellos, se establecen los hiperparámetros del modelo de Random Forest de manera de equilibrar capacidad predictiva y robustez.

### 2. Seleccionar métricas para las predicciones:

Para evaluar el rendimiento de los modelos, es necesario definir métricas que permitan cuantificar su capacidad predictiva. En este contexto se utilizan el MAE, el MAPE y el

coeficiente de determinación  $R^2$ , los cuales permiten medir, en distintas escalas, el error de predicción y el grado en que el modelo explica la variabilidad de la variable objetivo.

### 3. Modelos:

- Modelos redundantes: Consiste en modelar los grupos de tags redundantes identificados dentro del mismo sistema asociado a la falla, considerando que la redundancia se establece a nivel de subsistema. En una primera etapa, se desarrolla un modelo en el que cada uno de los tags pertenecientes a un determinado grupo de redundancia se utiliza como variable dependiente, mientras que el resto de los tags del mismo grupo se emplean como variables independientes en un esquema de modelamiento multivariable. Posteriormente, para cada uno de estos tags se construye un segundo modelo en el que su comportamiento se explica exclusivamente en función del tag de referencia previamente identificado como mejor dentro del grupo redundante.
- Modelo sistema: Los modelos se estructuran en función del grupo asociado al sistema en el que se origina la falla. En primer lugar, se desarrolla un modelo en el que cada uno de los tags no redundantes se modela en función del conjunto de tags no redundantes restantes y del tag definido como mejor dentro de los redundantes. En segundo lugar, se construye un modelo complementario en el que cada uno de los tags redundantes se modela respecto del conjunto de tags no redundantes, con el fin de evaluar la coherencia y el aporte de estas señales en la explicación del comportamiento del sistema.
- Modelo macrosistema: Consiste en modelar cada una de las variables asociadas al sistema en el que se origina la falla respecto de todas las variables pertenecientes al macrosistema correspondiente. En particular, se desarrolla un modelo en el que se

consideran como variables dependientes todos los tags clasificados bajo el mismo sistema de la falla, mientras que como variables independientes se utilizan todos los tags asociados al macrosistema de la falla, aplicando previamente un proceso de depuración para eliminar redundancias, de modo de conservar únicamente el mejor de los tags redundantes y la totalidad de los tags no redundantes.

#### 4. Selección de modelos:

Se opta por utilizar aquellos modelos que aportan una mayor cantidad de información al realizar una predicción, es decir, cuyos resultados son significativos para la anticipación de la falla. En particular, se selecciona el segundo modelo redundante, en el cual cada uno de los tags redundantes es modelado en función del tag identificado como “mejor”; adicionalmente, se consideran los modelos correspondientes a nivel de sistema y de macrosistema.

Es importante señalar que resulta más conveniente modelar el comportamiento de cada tag redundante respecto del tag definido como mejor, y el de este último respecto del segundo mejor, debido a que dichas relaciones permiten capturar de manera directa las desviaciones relevantes entre señales que debieran ser coherentes entre sí. En cambio, el modelo que incorpora simultáneamente a todos los tags redundantes de un mismo grupo introduce un alto grado de colinealidad, ya que las señales presentan comportamientos muy similares, lo que reduce el aporte marginal de cada variable, dificulta la interpretación del modelo y no incrementa de forma significativa su poder predictivo.

#### 5. Elaboración de modelos:

Utilizando los datos del período de entrenamiento, se elaboran cuatro modelos en lenguaje Python sobre la plataforma Databricks, los cuales se registran y gestionan como experimentos en la herramienta, quedando asociadas a cada uno sus respectivas métricas de desempeño y artefactos generados. En este contexto, los artefactos se entienden como el

conjunto de tags utilizados como variables de entrada para el desarrollo de los modelos.

- Para el caso del modelo macrosistema, es necesario realizar un nuevo proceso de análisis de datos y limpieza de datos, ya que aumenta el número de variables considerablemente. Esto debido a que, al aumentar el número de variables, crece también la probabilidad de que el modelo encuentre asociaciones que parecen estadísticamente significativas en el conjunto de entrenamiento, pero que no responden a una relación causal. Además, un número elevado de variables incrementa la complejidad del modelo, lo que se traduce en mayor varianza en las predicciones, mayor sensibilidad ante pequeñas variaciones en los datos y mayor dificultad para interpretar los resultados.

Considerando esto, se recurre al coeficiente de correlación lineal de Pearson para depurar el conjunto inicial de variables, eliminando aquellos tags que presentan una baja relación con los tags que se desean explicar. Como criterio de selección, se conserva aproximadamente el 60% de los tags con mayor valor absoluto de dicho coeficiente.

Posteriormente, se emplea el algoritmo Random Forest para modelar cada tag asociado al sistema de descanso en función de los tags del macrosistema, utilizando un mismo periodo de entrenamiento. A partir de la métrica de importancia de variables entregada por el modelo, se cuantifica la contribución relativa de cada predictor al desempeño del modelo predictivo en contraste con el resto de las variables, aplicando como criterio la selección de las 10 variables con mayor peso predictivo.

Ya habiendo realizado la selección de variables, es posible continuar con el modelamiento en fase de entrenamiento para este modelo.

## 6. Desarrollo de predicciones:

Considerando el período seleccionado para la elaboración de predicciones, se invocan los experimentos previamente generados para cada modelo y se desarrollan las correspondientes predicciones. De este modo, se obtiene un valor estimado para cada instante de tiempo en que se dispone de datos, el cual se contrasta con el valor real conocido para esa fecha y hora específica. Esta comparación permite calcular el MAE, el MAPE y el  $R^2$  asociados a las predicciones realizadas por cada modelo.

## 7. Umbrales:

Para identificar de manera oportuna desviaciones significativas que permitan reconocer la ocurrencia o proximidad de una falla, se establece un umbral para cada modelo predictivo, de modo que la salida continua del modelo se traduzca en un criterio de decisión operativo. En particular, el umbral se define con el propósito de habilitar el levantamiento de alertas que sustenten acciones preventivas, manteniendo un equilibrio entre detección temprana y activaciones innecesarias, y contribuyendo así a reducir el riesgo de consecuencias críticas.

Estos umbrales permiten verificar el cumplimiento de los objetivos, al establecer de manera objetiva si el modelo predictivo logra anticiparse a la falla mediante la detección oportuna de desviaciones significativas.

Para su determinación, se analizan los residuos del modelo al predecir en un periodo previo a la falla y posterior al entrenamiento, definidos como la diferencia entre el valor observado y el valor predicho, considerando su magnitud con el fin de cuantificar las desviaciones con independencia del signo. A partir de la distribución de dichos residuos, se evalúa una familia de umbrales parametrizada por un factor  $k$ , de modo de probar distintos valores de  $k$  y construir alternativas de umbral basadas en el comportamiento típico del error y en su variabilidad. Para cada  $k$ , se calcula el número de observaciones que excede el umbral

correspondiente y se expresa como un porcentaje. Con base en este criterio, se selecciona el  $k$  asociado al 99%, en consecuencia, se define el umbral de decisión para cada tag en falla según el modelo considerado.

## 6.5 Evaluation

Evaluación del desempeño de los modelos y del cumplimiento de los criterios establecidos.

### 1. Comparación entre valor real y predicción mediante gráficas:

Mediante el análisis de las gráficas resultantes, que muestran tanto el comportamiento de las predicciones en el horizonte temporal como la evolución real de los datos en el período considerado, es posible apreciar la magnitud de la discrepancia entre las predicciones de los distintos modelos y los valores observados en la realidad.

### 2. Comparación entre valor real y predicción mediante indicadores:

Los modelos predictivos elaborados se evalúan mediante los indicadores MAE, MAPE y  $R^2$ , con el objetivo de cuantificar, en distintas escalas, el error de predicción y la variabilidad asociada a las variables objetivo.

### 3. Backtesting:

Se contrasta el comportamiento reportado por el tag asociado a la falla en la plataforma de la empresa, con el propósito de verificar el cumplimiento de la predicción en el periodo analizado. Para ello, se emplea el umbral definido como criterio de decisión, identificando los instantes en que la desviación del modelo supera dicho valor y, en consecuencia, se activa la condición de alerta. De este modo, se determina la capacidad predictiva del modelo, el intervalo

temporal en que las predicciones se anticipan a la ocurrencia de la falla y el comportamiento de las variables relacionadas durante dicho periodo.

## 7 Resultados

A continuación, se presentan y analizan los resultados obtenidos en cada uno de los pasos de la metodología establecida.

### 7.1 Data Understanding

- Analizar la información disponible sobre los distintos tags de sensores: Colbún en Aveva PI cuenta con 1118 tags asociados a la Central Nehuenco I, los cuales actualmente entregan información de manera constante.
- Clasificar los Tags: El proceso de clasificación da como resultado 11 macrosistemas, 87 sistemas y 136 subsistemas distintos.
- Identificar fechas de detención en los últimos 2 años: Según la información oficial respecto al mantenimiento a la central, se efectuaron los siguientes mantenimientos mayores.

#### Figura 4

*Mantenimientos Mayores Históricos.*

Intervención	Fecha Inicio	Fecha Fin	Duración
Mantenimiento Mayor 2023	03-08-2023	17-01-2024	28 semanas
Mantenimiento Mayor 2024	02-12-2024	16-12-2024	2 semanas

- Seleccionar falla a modelar desde el sistema ERP de la compañía:

#### Figura 5

*Falla seleccionada para modelar*

Nombre de la Falla	Tag	Fecha de detección
Falla en la temperatura de los cojinetes	11MBDCT101C_XQ01	02-2025

La falla seleccionada se detecta en febrero de 2025 y presenta su cierre administrativo en Agosto del mismo año. Esta falla se elige debido al elevado nivel de

información disponible, que incluye detalles del impacto, descripción de la desviación y tags asociados, además de la posibilidad de cuantificar su efecto sobre Nehuenco I. En síntesis, la selección de esta falla responde a la combinación de su relevancia en la operación de la central y la riqueza de antecedentes técnicos asociados.

- Seleccionar periodo de entrenamiento: Debido a que la falla seleccionada ocurre a inicio del 2025, el periodo de entrenamiento debe ser antes de dicha fecha. Considerando la extenuante duración del mantenimiento mayor 2023, se considera un evento muy relevante e influyente en el desempeño de la central, por lo que se selecciona como periodo de entrenamiento el primer trimestre del 2024, con fechas específicas que van desde el 17/01/2024 al 17/04/2024.
- Seleccionar periodo de predicción: Considerando que la falla seleccionada es identificada en Febrero del 2025, se decide tener como primer periodo de predicción el segundo trimestre del 2024, específicamente del 17/05/2024 al 17/08/2024.

Según la información disponible en la empresa, la falla se presenta desde el 11/02/2025 y se logra identificar el 12/02/2025, por lo que el segundo periodo de predicción va del 08/02/2025 al 12/02/2025.

## 7.2 Data Preparation

Para esta sección es relevante considerar que la falla seleccionada es clasificada bajo el macrosistema Turbina a Gas, sistema Descanso y subsistema Temperatura, específicamente en el tag *OPC UA.Siemens.2.11MBD11CT101C\_\_XQ01*.

Asimismo, para la construcción de los modelos, se requiere identificar los tags que comparten esta misma clasificación.

## Figura 6

*Variables que comparten clasificación con el tag en falla*

Tag	Macrosistemas	Sistemas	Subsistema
OPC UA.Siemens.2.11MBD11CT101A XQ01	Turbina a Gas	Descanso	Temperaturas
OPC UA.Siemens.2.11MBD11CT101B XQ01	Turbina a Gas	Descanso	Temperaturas
OPC UA.Siemens.2.11MBD11CT101C XQ01	Turbina a Gas	Descanso	Temperaturas
OPC UA.Siemens.2.11MBD11CY101 XQ01	Turbina a Gas	Descanso	Vibraciones
OPC UA.Siemens.2.11MBD11CY102 XQ01	Turbina a Gas	Descanso	Vibraciones
OPC UA.Siemens.2.11MBD11CY103 XQ01	Turbina a Gas	Descanso	Vibraciones
OPC UA.Siemens.2.11MBD11CY940 XQ01	Turbina a Gas	Descanso	Vibraciones

- Extraer datos de los tags relacionados a la falla: Al extraer todos los tags bajo el macrosistema Turbina a Gas y aplicar todos los filtros correspondientes, para el periodo de entrenamiento seleccionado con toma de datos cada 10 minutos, se cuenta con un universo de 2.324.996 datos, donde hay 7.003 timestamp válidos.

Además, se extraen mediciones cada 10 minutos para todos los tags asociados al mismo macrosistema de la falla y se generan tablas de datos depuradas para cada uno de los sistemas que conforman el macrosistema Turbina a Gas.

- Identificar Redundancias.

El macrosistema Turbina a gas cuenta con 333 tags distintos, de los cuales se identifica que 117 presentan algún grado de redundancia con al menos otro tag, conformando 50 subgrupos diferenciados. En conjunto, este universo de tags da lugar a 819.351 mediciones.

Dentro del mismo sistema asociado a la falla se identifican dos subsistemas, de los cuales solo en el subsistema de Temperatura se observan redundancias, donde los tres tags del subsistema son redundantes entre sí.

- Identificar tags representativos en grupos de redundantes.

Dado que se identifican 50 grupos de tags redundantes en el macrosistema turbina a gas, al desarrollar el proceso de selección del tag más representativo en cada grupo, se obtiene un total de 50 tags representativos.

En relación con la falla, la redundancia se presenta en el subsistema temperaturas y se identifica como mejor al tag *OPC UA.Siemens.2.11MBD11CT101B\_\_XQ01* y como segundo mejor *OPC UA.Siemens.2.11MBD11CT101C\_\_XQ01*.

### 7.3 Modeling

#### 1. Seleccionar algoritmo a utilizar.

Al efectuar la extracción y el filtrado de los datos de entrenamiento, se identifica que cada tag dispone de 7.003 mediciones válidas para su modelamiento, por lo que se seleccionan los siguientes hiperparámetros para el algoritmo Random Forest.

#### **Figura 7**

*Hiperparámetros para Random Forest*

Hiperparámetro	Valor
Número de árboles	500
Profundidad	12
Mínimo de observaciones	20
VARIABLES	sqrt
Tasa de muestreo	0,8
Número máximo de intervalos	64
Semilla	42

#### 2. Modelamiento Redundantes:

Se evalúa cada tag que compone el grupo de redundantes respecto del tag identificado como “mejor” y, en el caso de este último, se lo evalúa respecto del segundo mejor. A partir de este análisis se obtienen las siguientes métricas.

## Figura 8

*Resultados Modelo en Fase Entrenamiento*

Variable Dependiente	MAE	MAPE (%)	R2
OPC UA Siemens 2 11MBD11CT101A XQ01	0.06720	0.06996	0.96601
OPC UA Siemens 2 11MBD11CT101B XQ01	0.06509	0.06797	0.96679
OPC UA Siemens 2 11MBD11CT101C XQ01	0.06297	0.06585	0.96712

Los resultados muestran que el modelo reproduce adecuadamente el comportamiento de las tres variables dependientes. Los valores de MAE y de MAPE indican errores absolutos y relativos acotados. Asimismo, los  $R^2$  entre 0,96 y 0,97 evidencian una alta capacidad explicativa del modelo.

### 3. Modelamiento Sistema:

Modelo No Redundantes con la Mejor:

Se modela cada uno de los tags no redundantes en función del conjunto formado por los demás tags, incorporando además el tag identificado como “mejor” dentro del grupo de redundantes, obteniéndose como resultado las siguientes métricas.

## Figura 9

*Resultados Modelo en Fase Entrenamiento*

Variable Dependiente	MAE	MAPE (%)	R2
OPC_UA_Siemens_2_11MBD11CY101_XQ01	0.06746	3,36768E+15	0.93595
OPC_UA_Siemens_2_11MBD11CY102_XQ01	0.06855	3,32071E+16	0.93969
OPC_UA_Siemens_2_11MBD11CY103_XQ01	0.06578	6,5671E+15	0.15534
OPC_UA_Siemens_2_11MBD11CY940_XQ01	39253E+13	11291E+12	0.21907

En este conjunto de resultados el desempeño dispar del modelo se considera coherente con la naturaleza de las variables, ya que, en su mayoría, no existe una relación fuerte entre ellas y la variable explicada. Solo algunas señales muestran un  $R^2$  aceptable, mientras que para el resto los bajos valores de  $R^2$  y los errores más altos reflejan precisamente esa falta de dependencia, por lo que es esperable que el modelo no logre

predecirlas adecuadamente.

#### Modelo Redundantes respecto No Redundantes:

Se utilizan como variables dependientes los tags pertenecientes al grupo de redundantes temperaturas, que se buscan explicar por el resto de los tags que componen el sistema descanso y no son redundantes.

### **Figura 10**

#### *Resultados Modelo en Fase Entrenamiento*

<b>Variable Dependiente</b>	<b>MAE</b>	<b>MAPE (%)</b>	<b>R2</b>
OPC_UA_Siemens_2_11MBD11CT101A__XQ01	0.36898	0.38391	0.15135
OPC_UA_Siemens_2_11MBD11CT101B__XQ01	0.36699	0.38297	0.15145
OPC_UA_Siemens_2_11MBD11CT101C__XQ01	0.35345	0.36943	0.15909

Se aprecia un MAE alto, lo que se puede explicar ya que el indicador está expresado en las unidades originales de cada variable, por lo que su magnitud puede verse inflada si las señales tienen escalas distintas o rangos amplios. En cambio, el MAPE presenta valores que indican que el error relativo porcentual es moderado y más comparable entre variables. No obstante, los coeficientes  $R^2$  cercanos a 0,15 muestran que el modelo solo logra explicar una fracción limitada de la variabilidad de las señales, lo que resulta coherente con la baja relación estructural esperada entre estas variables y el resto de los predictores

#### 4. Modelo Macrosistema:

Se modela cada uno de los tags pertenecientes al sistema asociado a la falla en función del conjunto de tags que conforman el macrosistema Turbina a Gas. Dado que en este enfoque se considera un número elevado de variables independientes, se hace necesario implementar un nuevo proceso de limpieza y depuración de datos que permita mitigar el riesgo de sobreajuste del modelo. Este procedimiento se adapta a la naturaleza

de la información disponible y se desarrolla de la siguiente manera.

Por la naturaleza del modelo, se eliminan los tags redundantes que no corresponden al mejor de su grupo y, adicionalmente, se separan los tags del sistema descanso del resto. Como resultado de este filtrado inicial, se dispone de 123 tags distintos para su utilización como variables explicativas, los cuales corresponden a tags del macrosistema que buscan explicar los 7 tags del sistema descanso.

Con el propósito de reducir este conjunto de 123 tags, se calcula el coeficiente de correlación de Pearson de cada tag del macrosistema respecto de los tags del sistema Descanso, seleccionando el 60 % con mayor coeficiente. Se observa que, bajo este umbral de corte, los coeficientes de Pearson se sitúan aproximadamente en torno a 0.5, lo cual valida que el corte no elimina variables con aporte significativo en una predicción. Dicho proceso reduce el número de tags del macrosistema a 67.

Posteriormente, se recurre al algoritmo Random Forest como herramienta de selección de variables. El modelo se entrena utilizando como predictores los tags previamente filtrados y, a partir de su estructura interna de árboles de decisión, se estima la importancia de cada variable en términos de su contribución a la reducción del error de predicción. Este enfoque permite evaluar el aporte de los tags en un contexto multivariable, capturando interacciones y redundancias que no quedan plenamente reflejadas en un análisis basado únicamente en correlaciones individuales.

De este modo, es posible identificar y descartar aquellas variables cuya importancia es marginal o cercana a cero, concentrando el modelamiento en un subconjunto reducido de tags que contribuye de manera significativa a la explicación del comportamiento del sistema, disminuyendo el riesgo de sobreajuste y facilitando la interpretación de los resultados. Como resultado de este proceso, se identifican los 10 tags con mayor peso en

la explicación del modelo.

En consecuencia, el modelo a nivel de macrosistema consiste en explicar el comportamiento de cada uno de los 7 tags del sistema descanso a partir de los 10 tags más explicativos del macrosistema turbina a gas, obteniéndose del correspondiente modelamiento las siguientes métricas.

### Figura 11

*Resultados Modelo en Fase Entrenamiento*

Variable Dependiente	MAE	MAPE (%)	R2
OPC_UA_Siemens_2_11MBD11CT101A__XQ01	0.08309	0.08652	0.95689
OPC_UA_Siemens_2_11MBD11CT101B__XQ01	0.07952	0.08303	0.95972
OPC_UA_Siemens_2_11MBD11CT101C__XQ01	0.07813	0.08171	0.95791
OPC_UA_Siemens_2_11MBD11CY101__XQ01	0.13745	67599E+11	0.72387
OPC_UA_Siemens_2_11MBD11CY102__XQ01	0.13690	64702E+11	0.74520
OPC_UA_Siemens_2_11MBD11CY103__XQ01	0.05420	54421E+11	0.37262
OPC_UA_Siemens_2_11MBD11CY940__XQ01	22758E+11	64272E+12	0.71379

Para las variables del subsistema Temperaturas se aprecia un desempeño muy adecuado, el MAE se mantiene entorno a 0,08 y el MAPE cercano al 8 %, lo que indica errores relativos moderados, mientras que los  $R^2$  superiores a 0,95 evidencian que la mayor parte de la variabilidad de estas señales se explica correctamente.

En cambio, para las variables del subsistema Vibraciones, el ajuste es menos consistente. Aunque los  $R^2$  se sitúan en valores intermedios. En estas señales resulta más razonable interpretar el desempeño a partir del  $R^2$  y del MAE, recordando que este último se expresa en las unidades propias de cada variable y puede verse influido por diferencias de escala.

#### 5. Umbrales:

Recordando que la falla se reporta en el tag OPC

UA.Siemens.2.11MBD11CT101C\_\_XQ01, se obtienen los umbrales para dicho tag en cada uno de los modelos, para el caso de los modelos de sistema se toman las predicciones de ambos modelos en dicho nivel.

**Figura 12**

*Umbrales obtenidos de predicciones en periodo posterior al entrenamiento y previo a la falla.*

<b>Modelo</b>	<b>MAE</b>	<b>K</b>	<b>Umbral</b>
Redundantes	0.08954448952164293	6	0.53722669371298
Sistema	0.52358591405306371	5	2.61792957026531
Macrosistema	0.15189105346020185	5	0.75945526730100

**7.4 Evaluation**

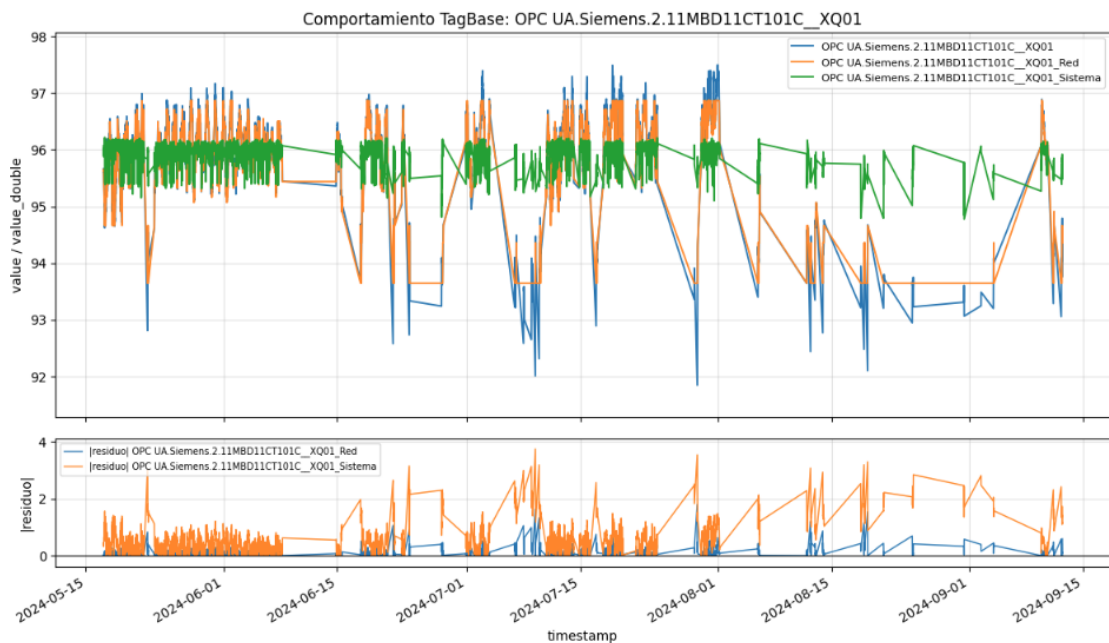
1. Predicciones Primer Periodo:

Se entiende que las predicciones realizadas por el modelo de redundantes generan un nuevo tag identificado con el prefijo \_Red, mientras que las predicciones provenientes de los modelos de sistema generan un nuevo tag identificado con el prefijo \_Sistema.

Considerando lo anterior, se interpreta la siguiente figura, en la cual se presenta el comportamiento de las predicciones en contraste con el comportamiento real del tag asociado a la falla, OPC UA.Siemens.2.11MBD11CT101C\_\_XQ01. Adicionalmente, la gráfica incorpora el valor residual en términos absolutos, con el propósito de entregar una mejor comprensión de la magnitud de la diferencia entre lo predicho y lo real para cada modelo.

**Figura 13**

*Comportamiento Predicciones de OPC UA.Siemens.2.11MBD11CT101C\_\_XQ01*



Se aprecia un comportamiento consistente entre la señal original y sus versiones \_Red y \_Sistema. A lo largo del periodo analizado, las tres series siguen trayectorias muy similares, se mantienen en un rango de valores comparable y reproducen de manera coherente las variaciones de la variable de referencia. Esto confirma que las mediciones capturan el mismo fenómeno, pero es importante destacar que las predicciones provenientes del modelo de redundantes presentan un comportamiento muy similar al original, donde sus diferencias con la realidad son solo en situaciones donde la variación original es más extrema.

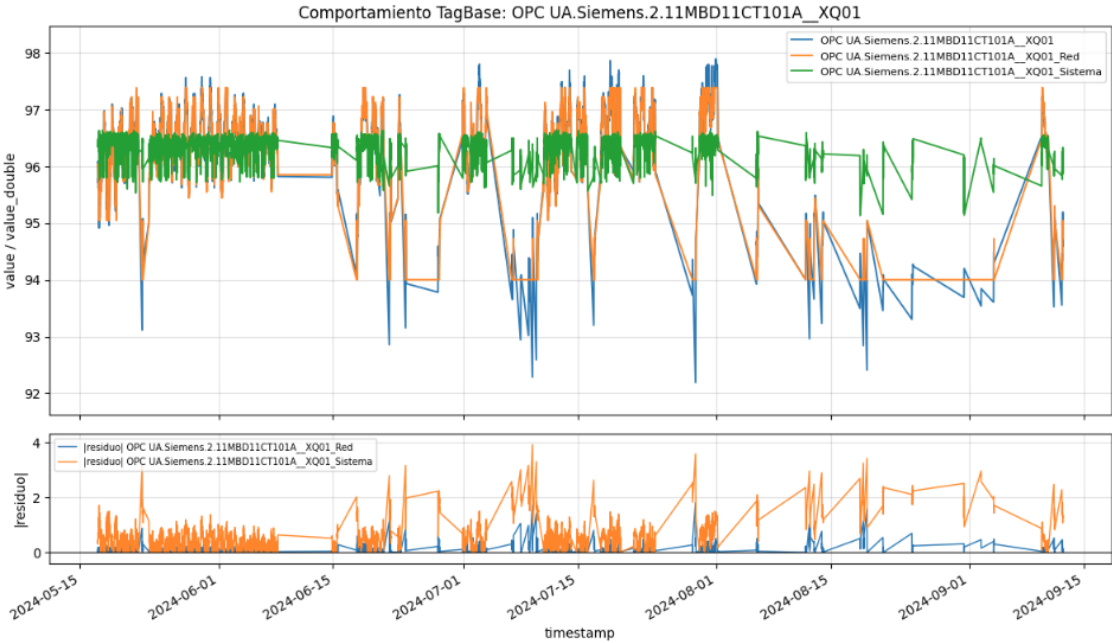
En el apartado de residuos, la mayor parte de las discrepancias se concentra en valores cercanos a cero y solo se identifican desviaciones puntuales de mayor magnitud. Desde una perspectiva operativa, ello resulta favorable, ya que la mayor parte del tiempo las señales son prácticamente intercambiables y permiten disponer de mediciones alternativas muy próximas entre sí, lo que refuerza la confiabilidad del monitoreo y aporta robustez frente a eventuales fallas o pérdidas de información en alguno de los canales.

Además, considerando la información disponible en el reporte de la empresa, se

incorpora para el análisis el comportamiento de las predicciones de las demás variables pertenecientes al mismo subsistema que la falla. En consecuencia, en las siguientes figuras es posible apreciar el comportamiento de las otras variables de temperatura asociadas.

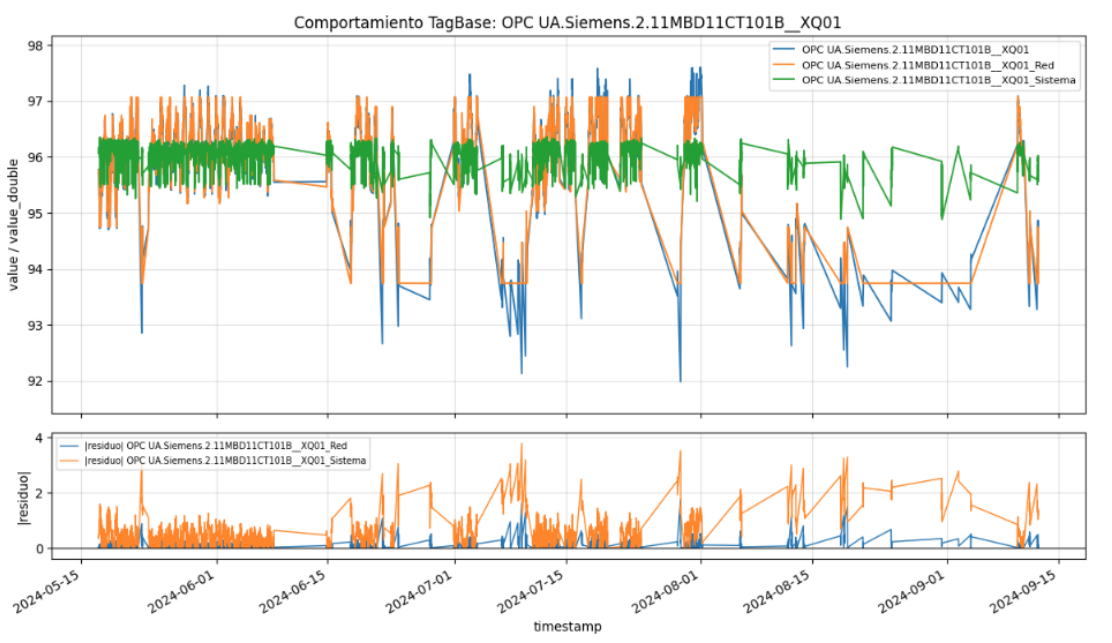
**Figura 14**

*Comportamiento Predicciones de OPC UA.Siemens.2.11MBD11CT101A\_XQ01*



**Figura 15**

*Comportamiento Predicciones de OPC UA.Siemens.2.11MBD11CT101B\_XQ01*

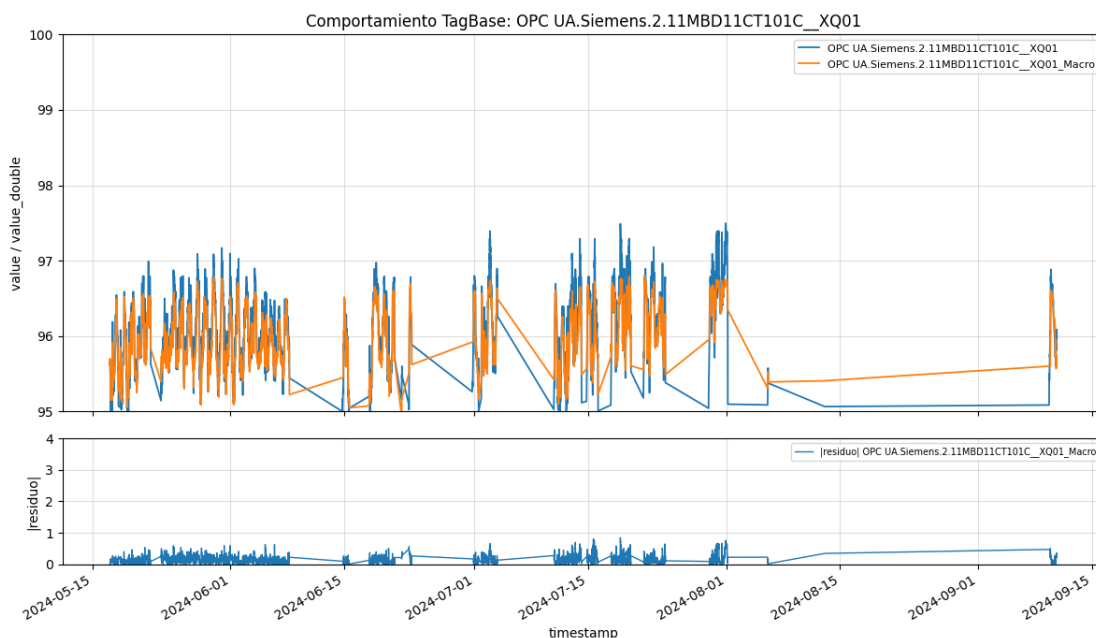


Las predicciones obtenidas por el modelo de macrosistema generan un nuevo tag identificado con el prefijo `_Macro`. Al igual que para el caso anterior, primero se muestra la figura que permite observar el comportamiento de la predicción sobre la falla y posteriormente las predicciones para las otras temperaturas del subsistema.

Respecto al tag asociado a la falla `OPC UA.Siemens.2.11MBD11CT101C__XQ01` se puede apreciar el comportamiento del modelo en contraste con los valores reales y también el valor residual en términos absolutos que se generan en dicho contraste.

### Figura 16

*Comportamiento Predicciones de OPC UA.Siemens.2.11MBD11CT101C\_\_XQ01*



La predicción `_Macro` para el tag `OPC UA.Siemens.2.11MBD11CT101C__XQ01` sigue de forma consistente la tendencia y el rango de la señal real durante la mayor parte del periodo, capturando el comportamiento típico sin introducir variaciones artificiales. Las diferencias más notorias aparecen en cambios bruscos o transitorios, donde el modelo responde de manera más suavizada y no replica completamente los extremos, pero sigue

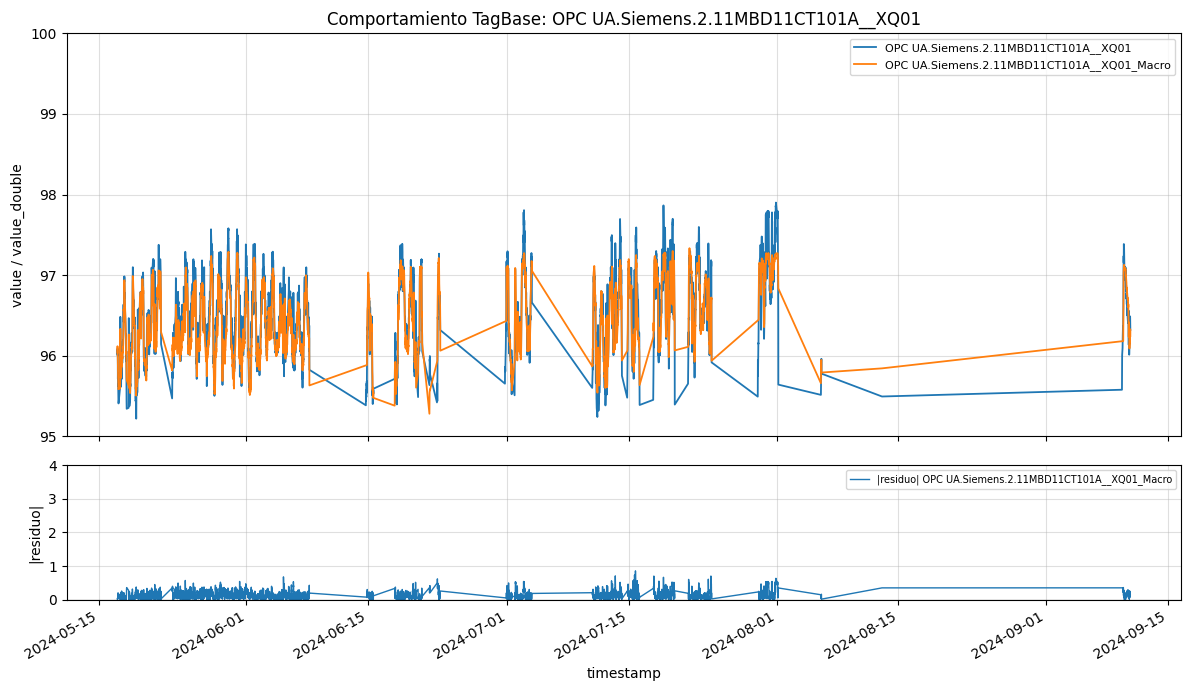
demostrando un comportamiento consistente y favorable.

En el panel de residuo absoluto, la magnitud del error se mantiene baja y cercana a cero la mayor parte del tiempo, y solo aumenta en ventanas puntuales asociadas a cambios extremos del comportamiento.

Además, al igual que con los modelos anteriores, se incorpora para el análisis el comportamiento de las predicciones de las demás variables pertenecientes al mismo subsistema que la falla.

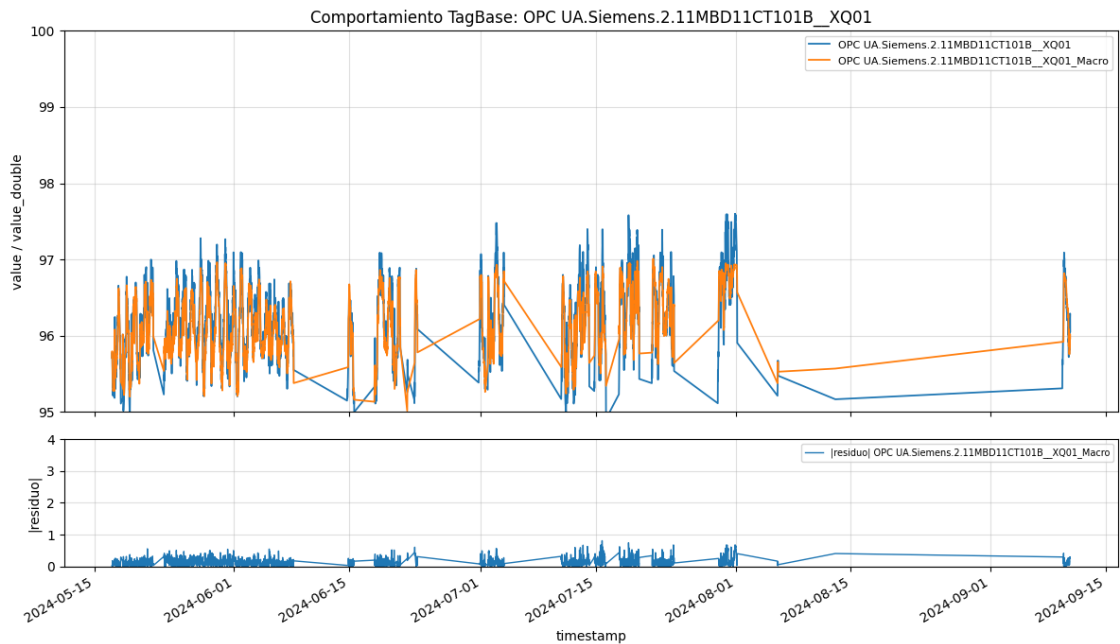
**Figura 17**

*Comportamiento Predicciones de OPC UA.Siemens.2.11MBD11CT101A\_\_XQ01*



**Figura 18**

*Comportamiento Predicciones de OPC UA.Siemens.2.11MBD11CT101B\_\_XQ01*



Se aprecia un comportamiento globalmente consistente entre las señales originales OPC UA.Siemens.2.11MBD11CT101A\_\_XQ01 y OPC UA.Siemens.2.11MBD11CT101B\_\_XQ01 y sus respectivas predicciones \_Macro. A lo largo del periodo analizado, en ambos casos la predicción se mantiene en un rango comparable y reproduce de manera coherente la tendencia general. Esto sugiere que el modelo captura adecuadamente el comportamiento típico de la variable y entrega una referencia estable, especialmente en los tramos con mayor continuidad de datos.

Las diferencias más visibles se concentran en instantes donde las señales reales presentan variaciones más bruscas. En dichas situaciones, la predicción tiende a responder de forma suavizada, conectando los cambios mediante transiciones graduales y sin perseguir completamente los extremos.

En la gráfica de residuo absoluto, la magnitud del error se mantiene baja y cercana a cero durante la mayor parte del periodo para ambos tags, mientras que los incrementos se observan de forma puntual y asociados a los cambios más abruptos de la señal. Desde una perspectiva operativa, este patrón resulta favorable, ya que confirma un buen ajuste

bajo condiciones normales.

## 2. Predicciones Segundo Periodo:

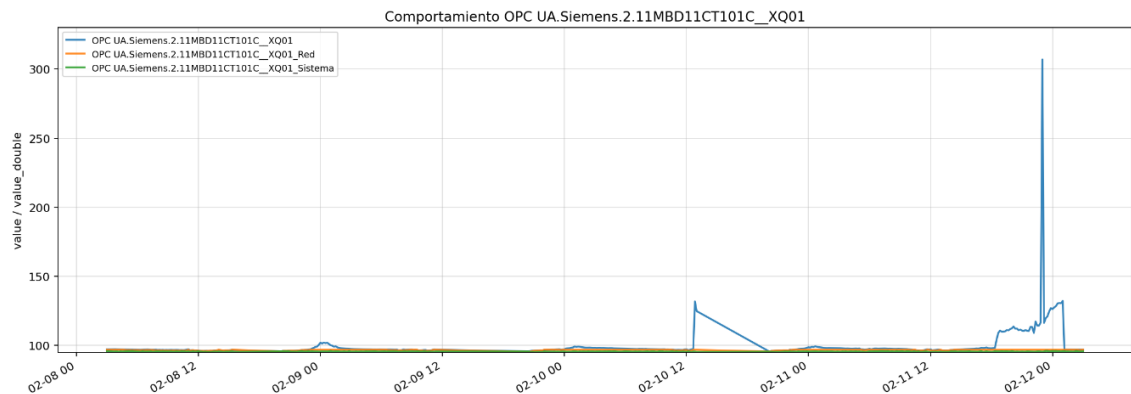
En este periodo de predicción se busca capturar la falla estudiada para poder establecer si los modelos elaborados mediante la metodología propuesta son lo suficientemente precisos, es por esto que se consideran los resultados del tag que reporta la falla OPC UA.Siemens.2.11MBD11CT101C\_\_XQ01.

Es importante señalar que se siguen utilizando para el análisis los mismos prefijos asociados a cada modelo predictivo mencionados anteriormente.

Se aprecia un comportamiento consistente entre la señal original OPC UA.Siemens.2.11MBD11CT101C\_\_XQ01 y las predicciones \_Red y \_Sistema durante gran parte del periodo analizado. En el régimen normal, ambas predicciones se mantienen muy próximas al valor real, en un rango comparable y siguiendo adecuadamente la dinámica base, lo que sugiere que el modelo captura el comportamiento típico de la variable y entrega una referencia estable para monitoreo.

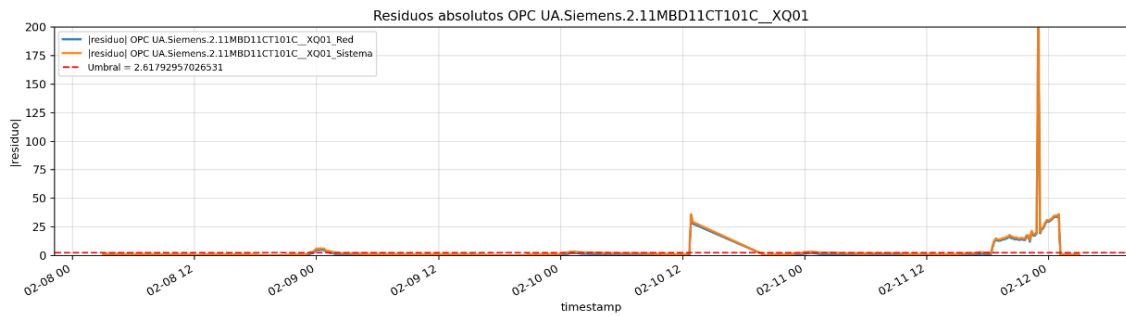
### **Figura 19**

*Comportamiento Predicciones de OPC UA.Siemens.2.11MBD11CT101C\_\_XQ01*



**Figura 20**

*Residuos Absolutos de OPC UA.Siemens.2.11MBD11CT101C\_\_XQ01*



La gráfica de residuos absolutos contiene el umbral seleccionado de 2.61792957026531 que establece un criterio objetivo para distinguir entre discrepancias esperables del modelo y desviaciones operacionalmente relevantes.

Durante la mayor parte del periodo analizado, los residuos asociados a las predicciones `_Red` y `_Sistema` se mantienen por debajo de dicho umbral, lo que indica que la diferencia entre el valor real y el valor predicho permanece acotada y consistente con el comportamiento normal de la variable.

Hacia el tramo final, el residuo absoluto comienza a superar el umbral de manera evidente y con mayor frecuencia. Este sobrepaso progresivo sugiere un cambio de régimen, en el cual la señal real deja de ajustarse al patrón aprendido por el modelo bajo condiciones típicas de operación. En consecuencia, el cruce del umbral permite delimitar una ventana temporal previa en la que la desviación se vuelve significativa, interpretándose como un indicio de aproximación a la falla, al reflejar una separación sostenida entre el comportamiento observado y el comportamiento esperado.

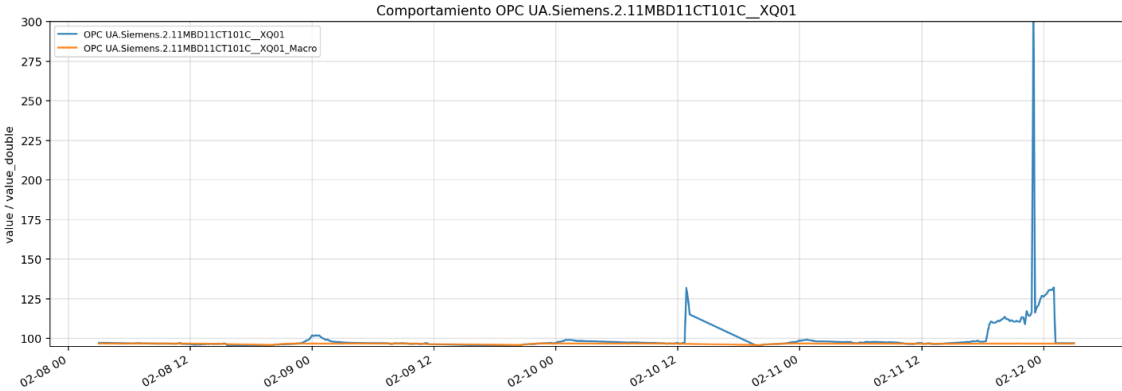
Adicionalmente, la similitud entre las curvas de residuo correspondientes a `_Red` y `_Sistema` refuerza la robustez de esta interpretación, dado que el incumplimiento del umbral no se atribuye a un comportamiento particular de un único modelo, sino a una desviación consistente del proceso que ambos enfoques reconocen como anómala.

Considerando lo expuesto previamente, se evidencia la capacidad del modelo de anticiparse a la falla al constatar que el 10 de febrero, el residuo absoluto experimenta un incremento abrupto y significativo que supera el umbral definido, lo que señala una desviación relevante respecto del comportamiento esperado. Posteriormente, el 12 de febrero se materializa la falla y en consecuencia, el aumento del residuo detectado con anterioridad se constituye como un indicio temprano de degradación, permitiendo delimitar una ventana temporal de alerta en la cual el modelo logra anticiparse al evento.

Respecto al comportamiento de las predicciones elaboradas por el modelo de macrosistema, se aprecia un comportamiento globalmente consistente entre la señal original OPC UA.Siemens.2.11MBD11CT101C\_\_XQ01 y la predicción \_Macro durante gran parte del periodo analizado. En el régimen normal, la predicción se mantiene próxima al valor real, en un rango comparable y reproduciendo adecuadamente la tendencia base, lo que sugiere que el modelo captura el comportamiento típico de la variable y entrega una referencia estable para monitoreo.

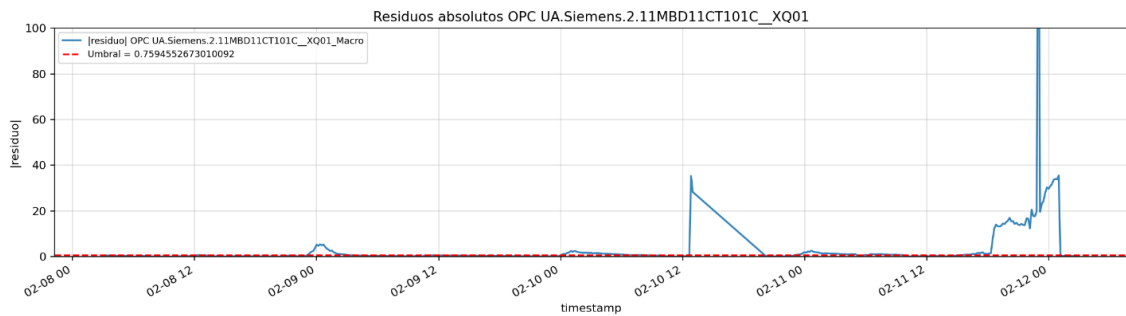
**Figura 21**

*Comportamiento Predicciones de OPC UA.Siemens.2.11MBD11CT101C\_\_XQ01*



**Figura 22**

*Residuos Absolutos de OPC UA.Siemens.2.11MBD11CT101C\_\_XQ01 con umbral*



La gráfica de residuos absolutos incorpora el umbral seleccionado de 0.7594552673010092, donde se observa que, durante la mayor parte del periodo analizado, el residuo absoluto asociado a la predicción \_Macro se mantiene por debajo de dicho umbral, lo que indica que la diferencia entre el valor real y el valor predicho permanece acotada y coherente con el comportamiento normal. Sin embargo, el 10 de Febrero el residuo absoluto supera el umbral de manera evidente, para posteriormente capturar la ocurrencia de la falla el 12 de Febrero donde la señal real deja de ajustarse al patrón aprendido por el modelo.

En consecuencia, el aumento detectado con anterioridad se interpreta como un indicio temprano de degradación, permitiendo delimitar una ventana temporal de alerta en la cual el modelo logra anticiparse al evento.

## 8 Alcance y Limitaciones

### 8.1 Alcance

Con el propósito de establecer el alcance real de la metodología y evaluar su viabilidad técnica de implementación en la central termoeléctrica Nehuenco I, se selecciona una falla adicional del sistema, asociada a un macrosistema distinto.

En particular, la falla a estudiar corresponde a una bajada de intensidad del ventilador 4 en las torres de refrigeración, identificada a través del tag OPC UA.Siemens.2.10PAD12CE002\_\_XQ01, el cual se clasifica bajo el macrosistema Torres de Refrigeración, el sistema Ventiladores y el subsistema Corriente.

En base a la información disponible, en el arranque del 18 de junio de 2024 se observa una disminución de la corriente asociada al ventilador desde aproximadamente 200 A hasta 100 A, lo que evidencia una reducción relevante en su condición operativa.

Como referencia operacional, un evento previo con un comportamiento similar se asocia a la rotura de un álabe, por lo que dicho mecanismo se considera una hipótesis técnica consistente para la condición registrada que posteriormente es confirmada mediante una inspección visual.

Siguiendo la metodología propuesta, se selecciona el periodo de entrenamiento considerando la fecha de ocurrencia de la falla y los periodos de mantenimiento prolongado, por lo cual se decide entrenar en el intervalo comprendido entre el 20/05/2023 y el 30/06/2023. Si bien este periodo es previo al mantenimiento mayor de mayor duración, se considera relevante capturar el comportamiento operativo con aproximadamente un año de anticipación respecto de la falla.

Respecto a los periodos de predicción, se decide testear los modelos en dos intervalos

distintos considerando el mantenimiento mayor. En consecuencia, el primer periodo de predicción seleccionado abarca del 17/01/2024 al 17/04/2024, mientras que el segundo periodo de predicción se orienta a representar la ocurrencia de la falla, por lo que se define entre el 15/06/2024 y el 19/06/2024.

Como ya se menciona anteriormente, la falla se clasifica bajo el macrosistema Torres de Refrigeración, el sistema Ventiladores y el subsistema Corriente. Asimismo, para comprender los modelos aplicados, se requiere identificar los tags que comparten esta misma clasificación.

**Figura 23**

*Variables bajo la misma clasificación de Sistema que el tag en falla*

<b>TAG</b>	<b>Macrosistemas</b>	<b>Sistemas</b>	<b>Subsistema</b>
OPC UA.Siemens.2.10PAD11CE001 XQ01	Torres de Refrigeración	Ventiladores	Corriente
OPC UA.Siemens.2.10PAD11CE002 XQ01	Torres de Refrigeración	Ventiladores	Corriente
OPC UA.Siemens.2.10PAD12CE001 XQ01	Torres de Refrigeración	Ventiladores	Corriente
OPC UA.Siemens.2.10PAD13CE001 XQ01	Torres de Refrigeración	Ventiladores	Corriente
OPC UA.Siemens.2.10PAD13CE002 XQ01	Torres de Refrigeración	Ventiladores	Corriente
OPC UA.Siemens.2.10PAD14CE001 XQ01	Torres de Refrigeración	Ventiladores	Corriente

Esta falla resulta de interés para su estudio debido a que no se identifican redundancias en ninguno de los niveles de clasificación asociados. En consecuencia, no se aplica el modelo de redundancia ni el primer modelo de sistema, en el cual se modelan todos los tags no redundantes junto con la mejor variable representativa de cada grupo de redundancia.

Bajo este escenario, se aplica el último de los modelos de sistema, incorporando una leve modificación para dejar fijado el tag asociado a la falla como variable a explicar, utilizando como predictores el resto de los tags no redundantes que comparten la misma clasificación del sistema. Adicionalmente, se aplica el modelo de macrosistema.

Respecto al rendimiento de los modelos en fase de entrenamiento, para el modelo de sistema aplicado se obtienen las siguientes métricas.

**Figura 24**

*Comportamiento modelo de predicción sistema*

Variable Dependiente	MAE	MAPE (%)	R2
UA.Siemens.2.10PAD12CE002__XQ01	0.04395	0.02197	0.91343

Mientras que para el modelo de macrosistema, las métricas obtenidas son las siguientes.

**Figura 25**

*Comportamiento modelo de predicción macrosistema*

Variable Dependiente	MAE	MAPE (%)	R2
UA.Siemens.2.10PAD12CE002__XQ01	0.08309	0.08652	0.95690

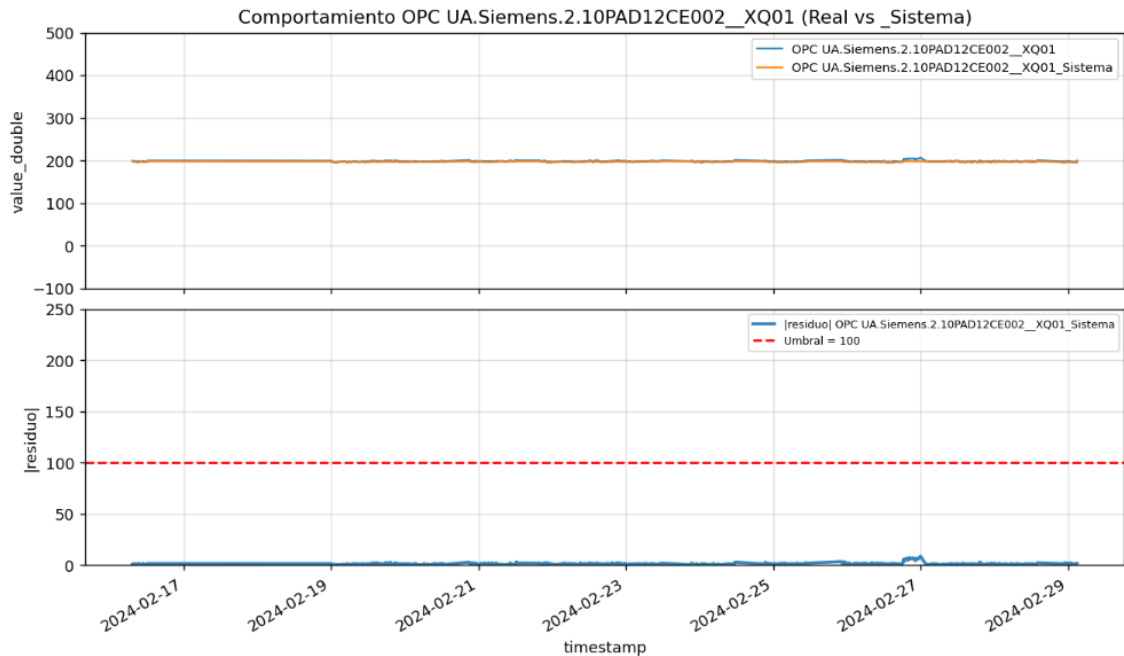
En ambas tablas se aprecia un desempeño satisfactorio de los modelos para la variable analizada, ya que las métricas reflejan un ajuste global consistente y un nivel de error acotado, lo que indica que las predicciones logran representar adecuadamente el comportamiento de la señal. En conjunto, los resultados respaldan que ambos enfoques son útiles para el monitoreo y análisis, con diferencias principalmente asociadas al nivel de agregación de las variables explicativas.

Respecto al umbral, en este caso se conoce históricamente que una variación de corriente de 100 A es consistente con la rotura de un álabe de los ventiladores, condición que, si bien no impide su funcionamiento, sí disminuye su rendimiento; por este motivo, se establece dicho valor como umbral de referencia para la identificación de este evento.

Respecto al primer periodo de predicción, se puede decir que el modelo a nivel sistema reproduce de manera adecuada el comportamiento de la señal según como se puede apreciar en el siguiente gráfico.

## Figura 26

*Comportamiento modelo de sistema en el primer periodo de testeo.*

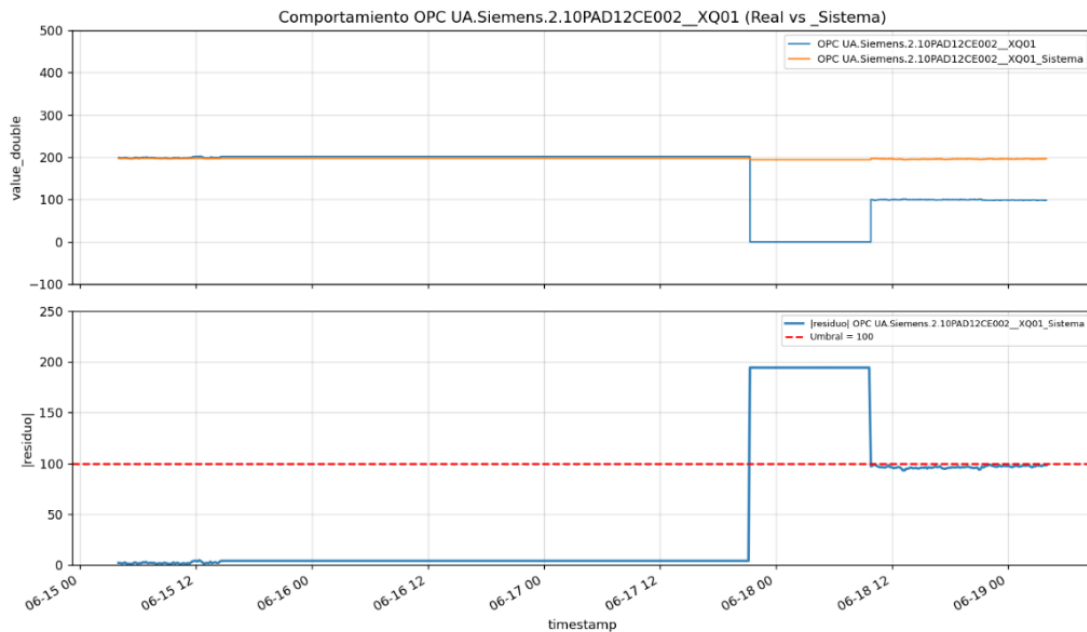


Según lo que se observa, la trayectoria se mantiene muy cercana a la medición real. La superposición entre ambas curvas evidencia una buena capacidad de ajuste y sugiere que el modelo captura correctamente el patrón normal de operación durante el periodo analizado. Complementariamente, el residuo se mantiene acotado y estable, lo que indica que las discrepancias entre lo predicho y lo observado corresponden principalmente a variaciones menores esperables.

Respecto al segundo periodo de predicción del mismo modelo que corresponde al periodo donde ocurre la falla que se logra capturar satisfactoriamente.

## Figura 27

*Comportamiento modelo de sistema en el segundo periodo de testeo.*



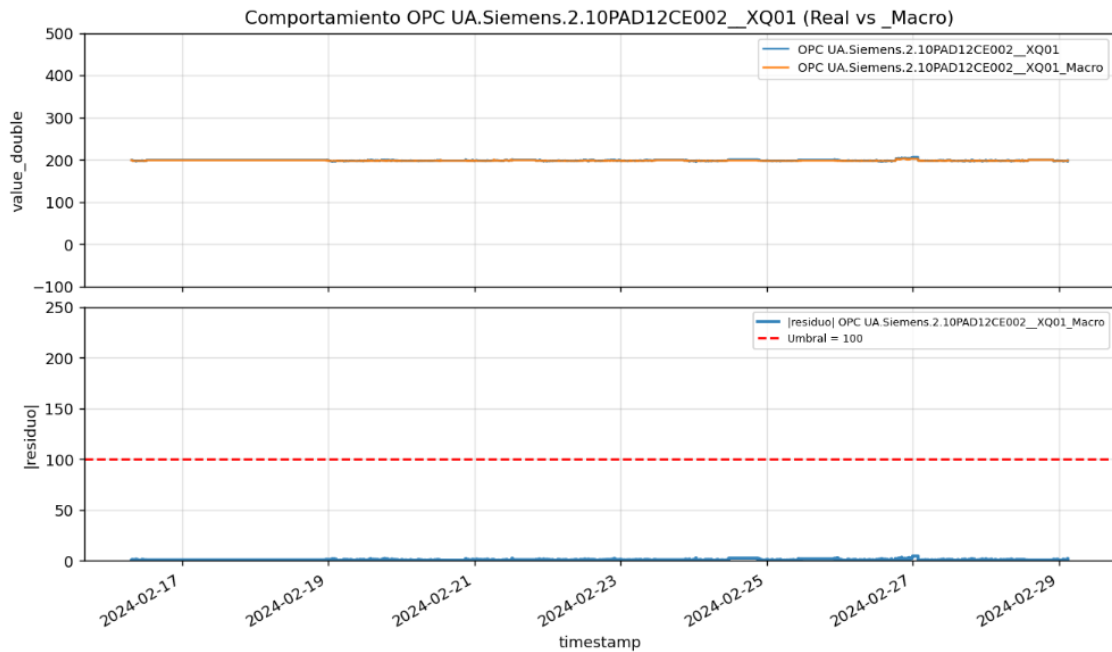
En la figura se aprecia el comportamiento de la corriente del tag analizado junto con su predicción a nivel sistema, observándose que tanto la serie predicha como el residuo presentan un tramo horizontal constante, lo cual se explica por la forma de representar el periodo de inactividad de la planta, manteniendo fijo el último valor predicho correspondiente a la condición en operación.

Posteriormente, al momento del arranque, se observa un cambio brusco consistente con la rotura del álabe de los ventiladores. En este escenario, el modelo conserva la referencia del comportamiento esperado y no reproduce el salto repentino, por lo que se genera una diferencia clara entre el valor real y el predicho; dicha diferencia se refleja en un aumento notorio del residuo, lo que permite identificar el evento de manera objetiva mediante el umbral definido.

Para el modelo a nivel de macrosistema, el primer período de predicción presenta un rendimiento destacado según los indicadores evaluados y un comportamiento satisfactorio al analizar la siguiente figura.

**Figura 28**

*Comportamiento modelo macrosistema en el primer periodo de testeo*

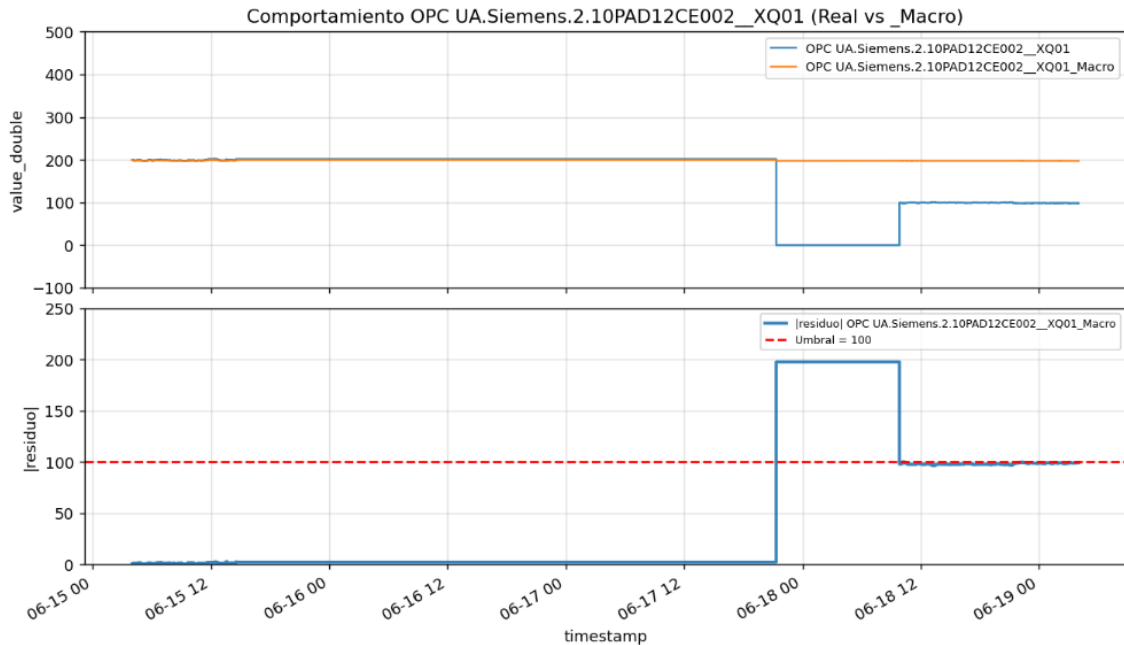


En la figura se observa que la predicción a nivel macrosistema reproduce de manera adecuada el comportamiento de la corriente del tag analizado, manteniéndose cercana a la señal real y siguiendo su tendencia general durante el periodo evaluado. La similitud entre ambas curvas sugiere que el modelo captura correctamente el patrón típico de operación sin introducir desviaciones relevantes. Esto se refleja también en el residuo, el cual permanece bajo y estable, sin superar el umbral definido, lo que indica que en el intervalo analizado no se presentan variaciones compatibles con el evento de falla y que el modelo entrega una referencia confiable para el monitoreo del comportamiento normal.

En el segundo periodo de predicción, se logra capturar la falla de manera consistente al ocurrir la puesta en marcha de la unidad.

**Figura 29**

*Comportamiento modelo macrosistema en el segundo periodo de testeo*



Se aprecia el comportamiento de la corriente del tag analizado junto con su predicción a nivel macrosistema, donde el tramo horizontal de la predicción responde a la representación del periodo sin operación manteniendo constante el último valor estimado en condición operativa. Al momento del arranque se observa un cambio abrupto en la señal real, consistente con el evento de rotura del álabe, mientras el modelo conserva la referencia del comportamiento esperado y no reproduce el salto, generando una diferencia marcada entre lo observado y lo predicho. Esta discrepancia se evidencia en el residuo, el cual aumenta de forma significativa y supera el umbral definido, permitiendo identificar el evento de manera objetiva y destacando la utilidad del modelo como línea base para la detección de desviaciones relevantes.

Considerando el desempeño positivo de los modelos para esta falla, se establece que se logra el objetivo de modelar la falla al seleccionar un evento representativo en un macrosistema distinto y estructurar su análisis bajo la misma metodológica. Definiendo un periodo de entrenamiento previo que caracteriza el comportamiento normal de la corriente del ventilador y,

posteriormente, evaluando el desempeño en ventanas de predicción antes y durante la ocurrencia del evento.

Dado que no se identifican redundancias en los niveles de clasificación asociados, la modelación se aborda mediante un modelo de sistema ajustado para explicar específicamente el tag de la falla a partir de los demás tags no redundantes del mismo sistema, y en paralelo mediante un modelo de macrosistema, logrando en ambos casos una representación consistente del régimen normal y una referencia estable de comportamiento esperado.

Al producirse el arranque donde se presenta la rotura del álabe, la señal real experimenta un cambio abrupto que no es replicado por el modelo, generando una separación clara entre lo observado y lo predicho; esta diferencia se refleja en un incremento marcado del residuo, el cual supera el umbral operacional definido con base histórica, permitiendo identificar el evento de manera objetiva. En conjunto, los resultados evidencian que la metodología es capaz de caracterizar la operación normal y resaltar desviaciones compatibles con la falla, validando su aplicabilidad para monitoreo y detección de este tipo de condiciones en la central.

## 8.2 Limitaciones

Dentro de las limitaciones de la metodología se reconoce que su aplicación puede requerir ajustes en función de las características del sistema y de la disponibilidad de variables asociadas. En primer lugar, cuando no existe redundancia en los niveles de clasificación considerados, se hace necesaria una leve modificación en la implementación del segundo modelo de sistemas, fijando explícitamente el tag de interés como variable dependiente y utilizando como predictores el resto de los tags no redundantes del mismo sistema, lo que implica que el procedimiento no se ejecuta de manera completamente uniforme en todos los casos. En segundo lugar, para determinadas fallas recurrentes, la definición del umbral no se establece únicamente a partir de

criterios estadísticos del error, sino que se fija con base en incidencias históricas y conocimiento operacional, lo cual exige comprender el mecanismo de falla y su forma de manifestación en la variable monitoreada. En consecuencia, la metodología presenta una dependencia parcial de la interpretación técnica del fenómeno, por lo que quien modela debe comprender la naturaleza de la información disponible y reconocer el tipo de falla a evaluar para aplicar correctamente los criterios de modelación y detección.

Pese a ello, considerando que a futuro se busca implementar un monitoreo continuo de las variables estudiadas, el modelamiento de las fallas no requiere modificaciones en los códigos de los modelos una vez establecidos, ya que su reutilización se mantiene válida bajo condiciones operativas comparables. Únicamente puede resultar necesario actualizar los periodos de entrenamiento cuando ocurren mantenimientos mayores o situaciones excepcionales que alteran el comportamiento habitual de las máquinas y, por tanto, modifican el patrón de referencia sobre el cual se construyen las predicciones y la detección de desviaciones.

## 9 Conclusiones

En función del desarrollo metodológico y de los resultados obtenidos, se establece que el estudio da respuesta a las interrogantes del problema de investigación y cumple el objetivo general y los objetivos específicos definidos, al demostrar que la metodología predictiva interna de Colbún es escalable y aplicable en un entorno térmico como la Central Térmica Nehuenco I, y que los resultados respaldan técnicamente el levantamiento de un caso de negocio para su implementación, en la medida en que se formalicen criterios de despliegue, mantenimiento del modelo y gobernanza de datos.

En primer lugar, se verifica el cumplimiento del objetivo asociado a replicar y adaptar la metodología en Nehuenco I mediante la construcción de un flujo completo desde la obtención de información operacional hasta la validación por backtesting. Donde la implementación se sustenta en la disponibilidad de datos en AVEVA PI y en su integración en Databricks, lo que permite estructurar y depurar un universo representativo para modelamiento. En particular, la caracterización inicial evidencia una base instrumental amplia y una clasificación operacional que habilita el análisis jerárquico, lo cual resulta determinante para organizar el modelamiento en niveles comparables y trazables.

En segundo lugar, se constata el cumplimiento del objetivo de validar el desempeño mediante backtesting, al aplicarse el enfoque sobre una falla principal y una falla adicional en un macrosistema distinto, lo que permite evaluar robustez y transferibilidad dentro del mismo entorno térmico y comparar con la información disponible en los sistemas propios de la compañía con resultados satisfactorios.

En tercer lugar, se verifica el cumplimiento del objetivo orientado a determinar viabilidad técnica de implementación permanente, en tanto el estudio demuestra que el método entrega una referencia estable del comportamiento esperado y que la detección de desviaciones se materializa

mediante un criterio cuantificable. Para la falla principal, la evidencia se consolida al observar que el análisis de residuos absolutos y el umbral definido permiten distinguir entre el error típico del modelo y desviaciones operacionales relevantes. Adicionalmente, la aplicación a una falla perteneciente a un macrosistema distinto refuerza la consistencia del enfoque y su capacidad de adaptarse a escenarios con distinta estructura de datos.

Finalmente, respecto de la segunda interrogante asociada al levantamiento de un caso de negocio, el estudio entrega fundamentos técnicos concretos; integración con las fuentes corporativas existentes, capacidad de operar con grandes volúmenes de datos, estructuración por clasificación operacional, reglas explícitas de tratamiento de redundancia y selección de variables, y un mecanismo cuantitativo de alerta temprana basado en residuos y umbrales. En este marco, la viabilidad estratégica se sustenta en que el método aporta una herramienta estandarizada para fortalecer el monitoreo de condición, aumentar la objetividad en la identificación de desviaciones y habilitar decisiones más oportunas de mantenimiento. A la vez, las limitaciones identificadas delimitan condiciones de implementación: la necesidad de ajustes cuando no existe redundancia, y la dependencia parcial de conocimiento operacional para fijar umbrales en ciertos fenómenos. Aun con estas consideraciones, el desempeño verificado en los casos analizados permite concluir que la metodología constituye una base sólida para un despliegue progresivo, siempre que se mantenga la actualización de periodos de entrenamiento ante cambios mayores en operación y se definan criterios de gobernanza para su uso continuo en planta.

## 10 Trabajos Futuros

Se identifican dos líneas de trabajo futuro cuya implementación aporta valor directo al fortalecimiento y escalabilidad de la metodología. En primer lugar, resulta relevante desarrollar un enfoque que permita construir modelos incorporando variables pertenecientes a macrosistemas distintos, de modo que sea posible explicar el comportamiento de una variable objetivo a partir de señales asociadas a otras clasificaciones a nivel de macrosistema. Esta extensión habilita el análisis de relaciones cruzadas entre procesos, capturando dependencias operacionales que no quedan representadas cuando el modelamiento se restringe a una única clasificación, y ampliando la capacidad de caracterización del comportamiento esperado bajo condiciones reales de operación.

En segundo lugar, se considera pertinente replicar y validar la metodología en una central de generación térmica distinta, con el propósito de evaluar su robustez frente a cambios en configuración, instrumentación y patrones de operación. En este contexto, la aplicación en las unidades Nehuenco II y Nehuenco III permite contrastar el desempeño del enfoque en escenarios comparables, identificar eventuales ajustes requeridos para su transferencia y consolidar criterios de generalización para su despliegue a mayor escala dentro de la organización.

## 11 Referencias

Amendola, L. (2011). *Ensayo de monitoreo por condición*. Obtenido de <https://ingkatherinerincon.wordpress.com/wp-content/uploads/2011/04/ensayo-de-monitoreo-por-condicic3b3n.pdf>

Alonso, J. (2025). *Tema 12: Técnicas de aprendizaje supervisado*. Obtenido de <https://www.cs.us.es/~jalonso/cursos/ra-00/temas/tema-12.pdf>

Aveva. (2025). *What is AVEVA PI System?*. Obtenido de <https://www.aveva.com/en/products/aveva-pi-system/>

Aveva. (2025 ). *AVEVA PI System*. Obtenido de <https://www.aveva.com/es-es/products/aveva-pi-system/>

Comisión para el Mercado Financiero (CMF). (2025). *Estadísticas del Mercado de Valores*. Obtenido de [https://www.cmfchile.cl/portal/estadisticas/617/articles-40018\\_doc\\_pdf.pdf](https://www.cmfchile.cl/portal/estadisticas/617/articles-40018_doc_pdf.pdf)

Coordinador Eléctrico Nacional. (2024). *Reporte energético SEN – Febrero 2024*. Obtenido de [https://www.coordinador.cl/wp-content/uploads/2024/01/CEN\\_Reporte\\_Energetico\\_SEN\\_Feb24.pdf](https://www.coordinador.cl/wp-content/uploads/2024/01/CEN_Reporte_Energetico_SEN_Feb24.pdf)

Chris McMillen. (2025) *What is Condition Monitoring?*. Obtenido de <https://www.bakerhughes.com/bently-nevada/blog/what-condition-monitoring>

Databricks. (2025). *Data Intelligence Platform*. Obtenido de <https://www.databricks.com/product/data-intelligence-platform>

Facility Management Services. (2023). *Mantenimiento basado en riesgos: qué es, cómo funciona y quién lo necesita*. Obtenido de

[https://www.facilitymanagementservices.es/actualidad/mantenimiento-basado-riesgos-que-es-como-funciona-quien-necesita\\_20230929.html](https://www.facilitymanagementservices.es/actualidad/mantenimiento-basado-riesgos-que-es-como-funciona-quien-necesita_20230929.html)

Fernández Cabanas, M., (1998). Técnicas para el mantenimiento diagnóstico de máquinas eléctricas rotativas. Obtenido de [https://www.google.cl/books/edition/Técnicas\\_para\\_el\\_mantenimiento\\_diagnós/X3p4bZfoqgEC?hl=es&gbpv=1](https://www.google.cl/books/edition/Técnicas_para_el_mantenimiento_diagnós/X3p4bZfoqgEC?hl=es&gbpv=1)

Foro Nuclear. (2025). *¿Qué es una central eléctrica?* Obtenido de <https://www.foronuclear.org/descubre-la-energia-nuclear/preguntas-y-respuestas/sobre-distintas-fuentes-de-energia/que-es-una-central-electrica/>

Foro Nuclear. (2025). *¿Qué es una central termoeléctrica?* Obtenido de <https://www.foronuclear.org/descubre-la-energia-nuclear/preguntas-y-respuestas/sobre-distintas-fuentes-de-energia/que-es-una-central-termoelectrica/>

Gómez de León, F. C. (1998). *Tecnología del mantenimiento industrial*. Obtenido de [https://www.google.cl/books/edition/Tecnolog%C3%ADa\\_del\\_mantenimiento\\_industrial/bOrFC3532MEC?hl=es&gbpv=0](https://www.google.cl/books/edition/Tecnolog%C3%ADa_del_mantenimiento_industrial/bOrFC3532MEC?hl=es&gbpv=0)

Grupo ORS. (2023). *¿Cuál es el nivel de madurez del mantenimiento que aplica en su industria?* Obtenido de <https://grupoors.com.mx/2023/04/29/cual-es-el-nivel-de-madurez-del-mantenimiento-que-aplica-en-su-industria/>

IBM. (2025). *Análisis predictivo*. Obtenido de <https://www.ibm.com/es-es/topics/predictive-analytics>

IBM. (2025). *CRISP-DM: descripción general de la ayuda*. Obtenido de <https://www.ibm.com/docs/es/spss-modeler/saas?topic=dm-crisp-help-overview>

Menezes, N. N. C. (2017). *Introducción a la programación con Python: Algoritmos y*

*lógica de programación para principiantes.* Obtenido de [https://www.google.cl/books/edition/Introducci%C3%B3n\\_a\\_la\\_programaci%C3%B3n\\_con\\_Pyth/p7M3DwAAQBAJ?hl=es](https://www.google.cl/books/edition/Introducci%C3%B3n_a_la_programaci%C3%B3n_con_Pyth/p7M3DwAAQBAJ?hl=es)

Mortera de Iruarrizaga, M. S. (2024, Agosto). *Marco regulatorio en materia eléctrica.* Biblioteca del Congreso Nacional de Chile. Obtenido de <https://obtienearchivo.bcn.cl/obtienearchivo?id=repositorio/10221/36331/2/SancionesyMultasformatoinforme.pdf>

Olarte C., W., Botero Arbeláez, M., & Cañon Zabaleta, B. (2010). *Técnicas de mantenimiento predictivo utilizadas en la industria.* *Scientia Et Technica*, 2(45). Obtenido de <https://doi.org/10.22517/23447214.355>

Ortega Candel, J. M. (2022). *Big data, machine learning y data science en Python.* Obtenido de [https://www.google.cl/books/edition/Big\\_Data\\_Machine\\_Learning\\_y\\_Data\\_Science/L7VoEA\\_AAQBAJ?hl=es](https://www.google.cl/books/edition/Big_Data_Machine_Learning_y_Data_Science/L7VoEA_AAQBAJ?hl=es)

Pineda Pertuz, C. (2022). *Aprendizaje automático y profundo en Python: Una mirada hacia la inteligencia artificial.* Obtenido de [https://www.google.cl/books/edition/Aprendizaje\\_Autom%C3%A1tico\\_y\\_Profundo\\_en\\_Pyt/hZoQEAAAQBAJ?hl=es](https://www.google.cl/books/edition/Aprendizaje_Autom%C3%A1tico_y_Profundo_en_Pyt/hZoQEAAAQBAJ?hl=es)

Sabugal García, S. (2013). *Centrales térmicas de ciclo combinado.* Obtenido de <https://books.google.es/books?id=KCG5BgAAQBAJ&lpg=PA1&ots=KTA1kyqdoi&dq=central%20termicas%20de%20ciclo%20combinado&lr&hl=es&pg=PA1#v=onepage&q=central%20termicas%20de%20ciclo%20combinado&f=false>

Sanz Osorio, J. F. (Comp.). (2016). *Energía hidroeléctrica: Serie Energías renovables*

(Vol. 139). Prensas de la Universidad de Zaragoza. Obtenido de <https://books.google.com/books?id=8416933316>

Souza, D. (2025). *Guía avanzada sobre la acción del precio en los mercados financieros*. Obtenido de [https://www.google.cl/books/edition/Gu%C3%ADa\\_Avanzada\\_sobre\\_la\\_Acci%C3%B3n\\_del\\_Preci/S\\_YgEAAAQBAJ?hl=es](https://www.google.cl/books/edition/Gu%C3%ADa_Avanzada_sobre_la_Acci%C3%B3n_del_Preci/S_YgEAAAQBAJ?hl=es)

Telefonica Tech. (2021). *¿Qué algoritmo elegir en machine learning (aprendizaje automático)?* . <https://telefonicatech.com/blog/que-algoritmo-elegir-en-ml-aprendizaje>

Vilar Barrio, J. (2005). *Control estadístico de los procesos (SPC)*. Obtenido de [https://www.google.cl/books/edition/Control\\_estad%C3%ADstico\\_de\\_los\\_procesos\\_SPC/jc1qaKMBYvIC?hl=es&gbpv=0&kptab=overview](https://www.google.cl/books/edition/Control_estad%C3%ADstico_de_los_procesos_SPC/jc1qaKMBYvIC?hl=es&gbpv=0&kptab=overview)