

**Estrategias de Machine Learning para la estimación de desviaciones, consumo y  
mantenimiento óptimo de medidores de agua potable**

Brahayan Camilo Sierra García

Asesor

Dr. Rafael Gaitán Ospina

Universidad Nacional Abierta y a Distancia UNAD  
Escuela de Ciencias Básicas, Tecnología e Ingeniería ECBTI  
Especialización en Ciencia de Datos y Analítica

2025

## Resumen

Este proyecto desarrolla un sistema de mantenimiento predictivo basado en algoritmos de Machine Learning para optimizar la gestión metrológica del parque de medidores de agua potable fría en una empresa de servicios públicos. A partir de una base histórica con más de 330.000 resultados de calibración (2011–2025), complementada con información operativa y comercial, se aplicaron procesos de depuración, análisis exploratorio y modelado avanzado para estimar la desviación metrológica frente a los límites establecidos en la ISO 4064:2014. Se implementaron modelos de regresión cuantílica y clasificación mediante CatBoost para predecir el error en caudal nominal (EQ3), la probabilidad de conformidad del medidor y su vida útil metrológica remanente. Asimismo, se desarrolló una aplicación en Streamlit integrada con Google Sheets que permite cargar datos en tiempo real y generar predicciones automáticas.

La validación externa con 3.880 resultados reales de laboratorio arrojó una precisión del 67 % en los intervalos predictivos y del 70 % en el clasificador Cumple/No cumple, demostrando la viabilidad técnica del enfoque. En conjunto, el sistema constituye una herramienta operativa para priorizar reemplazos, reducir pérdidas comerciales y avanzar hacia un modelo de gestión metrológica basado en datos.

**Palabras clave:** Machine Learning, Mantenimiento Predictivo, Metrología, Agua Potable, Micromedición.

## Abstract

This project develops a predictive maintenance system based on Machine Learning algorithms to optimize the metrological management of cold potable water meters in a public utility company. Using a historical database containing more than 330,000 calibration results (2011–2025), complemented with operational and commercial information, data cleansing, exploratory analysis, and advanced modeling processes were applied to estimate metrological deviation with respect to the limits established in ISO 4064:2014. Quantile regression and classification models using CatBoost were implemented to predict the error at nominal flow (EQ3), the probability of meter compliance, and its remaining metrological service life. In addition, a Streamlit application integrated with Google Sheets was developed to enable real-time data loading and automatic prediction generation.

External validation with 3,880 real laboratory results yielded an accuracy of 67% for predictive intervals and 70% for the Pass/Fail classifier, demonstrating the technical feasibility of the approach. Overall, the system constitutes an operational tool for prioritizing meter replacements, reducing commercial losses, and advancing toward a data-driven metrological management model.

**Keywords:** Machine Learning, Predictive Maintenance, Metrology, Drinking Water, Micro-metering.

## Contenido

Introducción .....	9
Descripción del Problema .....	11
Planteamiento del Problema.....	11
Justificación .....	13
Objetivos .....	14
Objetivo General .....	14
Objetivos Específicos.....	14
Marco de Referencia .....	15
Metrología de Medidores de Agua.....	15
Fundamentos de Analítica y Machine Learning Aplicables .....	18
Regresión Lineal (Simple y Múltiple).....	18
Random Forest (Regresión y Clasificación) .....	19
CatBoost .....	19
Aprendizaje No Supervisado.....	20
Estimación de Degradación y Vida Útil Remanente.....	20
Hallazgos de la Literatura .....	20
Metodología .....	22
Fase 1 Comprensión del Negocio y Recolección de Datos.....	22
Fase 2 Preparación de los Datos.....	23
Fase 3 Desarrollo del Modelo Predictivo.....	23
Fase 4 Integración e Implementación .....	24
Fase 5 Evaluación y Documentación .....	24

Resultados .....	25
Fase 1 Comprensión del Negocio y Recolección de Datos.....	25
Fase 2 Preparación de los Datos.....	27
Carga e Inspección Inicial .....	27
Base de Datos Reportes de Calibración .....	28
Base de Datos Reportes Operación Comercial.....	28
Integración y Enriquecimiento.....	29
Tabla de Modelos de Medidores .....	29
Nuevas Variables.....	29
Error Ponderado .....	29
Imputación del Año de Fabricación .....	29
Normalización de Fabricantes y Modelos de Medidores .....	30
Depuración Final.....	30
Normalización y Filtrado por Características Técnicas .....	31
Depuración de la Variable Metrología .....	31
Tratamiento de Outliers Mediante Método IQR.....	31
Selección y Codificación de Variables .....	32
Fase 3 Desarrollo del Modelo Predictivo.....	36
Preparación Técnica y Carga del Entorno.....	37
Metodología de Modelado.....	37
Importancia de las Variables .....	39
Modelos de Regresión Cuantílica.....	40
Validación Cruzada y Resultados Comparativos .....	41

Clasificación del Estado Metrológico .....	42
Estimación de la Vida Útil Remanente .....	43
Exportación y Despliegue de Modelos.....	44
Síntesis de la Fase 3.....	45
Fase 4 Integración e Implementación .....	45
Arquitectura General de Integración .....	46
Autenticación, Conexión y Flujo de Datos .....	46
Carga de Modelos Predictivos y Estructura .....	47
Proceso de Predicción y Generación de Indicadores.....	48
Visualización Interactiva y Semaforización.....	48
Exportación Automática de Resultados a Google Sheets .....	50
Valor Aplicado y Relevancia Dentro del Proyecto .....	51
Fase 5 Evaluación y Documentación .....	51
Diseño del Proceso de Evaluación .....	51
Validación del Clasificador de Conformidad.....	52
Validación del Modelo de Intervalos .....	53
Evaluación por Modelo Estándar de Medidor.....	53
Síntesis del Desempeño Global del Sistema .....	55
Interpretación Global y Análisis Crítico .....	56
Conclusiones .....	57
Recomendaciones .....	58
Referencias.....	59

## Lista de Tablas

<b>Tabla 1</b> <i>Normalización de Fabricantes y Modelos de Medidores</i> .....	30
<b>Tabla 2</b> <i>Resultados del Filtrado por Outliers</i> .....	31
<b>Tabla 3</b> <i>Selección y Codificación de Variables</i> .....	32
<b>Tabla 4</b> <i>Criterio de Clasificación del Coeficiente de Correlación de Pearson (<math>r</math>)</i> .....	33
<b>Tabla 5</b> <i>Correlaciones Destacadas</i> .....	34
<b>Tabla 6</b> <i>Algoritmos Representativos - Modelos de Regresión</i> .....	38
<b>Tabla 7</b> <i>Validación Cruzada y Resultados Comparativos</i> .....	42
<b>Tabla 8</b> <i>Resultados por Modelo (Validación Estratificada)</i> .....	43
<b>Tabla 9</b> <i>Vida Útil Estimada (Años)</i> .....	44
<b>Tabla 10</b> <i>Matriz de Confusión del Clasificador</i> .....	52
<b>Tabla 11</b> <i>Validación por Modelo Estándar de Medidor</i> .....	54
<b>Tabla 12</b> <i>Resumen de Precisión del Sistema Predictivo</i> .....	55

## Lista de Figuras

<b>Figura 1</b> <i>Medidor de Agua Volumétrico</i> .....	16
<b>Figura 2</b> <i>Curva de Error Típica de un Medidor de Agua</i> .....	17
<b>Figura 3</b> <i>Fases del Proyecto</i> .....	22
<b>Figura 4</b> <i>Mapa de Correlaciones Variables; Edad y Volumen, y EQ1, EQ2, EQ3</i> .....	33
<b>Figura 5</b> <i>Correlaciones Variables; Edad y Volumen, y los Errores (EQ1–EQ3)</i> .....	35
<b>Figura 6</b> <i>Relación Entre las Variables (EQ1, EQ2 y EQ3), Edad y Volumen</i> .....	36
<b>Figura 7</b> <i>Desempeño Predictivo por Tipo de Medidor y Modelo de Regresión</i> .....	39
<b>Figura 8</b> <i>Importancia de Variables por Tipo de Medidor, Modelo CatBoost.</i> .....	40
<b>Figura 9</b> <i>Desempeño Predictivo por Modelo de Medidor</i> .....	41
<b>Figura 10</b> <i>Captura de la Interfaz Inicial de Streamlit</i> .....	47
<b>Figura 11</b> <i>Ejemplo de Resultados con Semaforización</i> .....	49
<b>Figura 12</b> <i>Dashboard: Distribución de Riesgo y Conformidad</i> .....	50
<b>Figura 13</b> <i>Desempeño Promedio de los Modelos por Modelo Estándar de Medidor.</i> .....	55

## Introducción

En los sistemas de acueducto, el envejecimiento de los medidores y el desgaste hidráulico gradual generan desviaciones en la medición del consumo, lo cual puede traducirse en pérdidas económicas, reclamos de usuarios, desequilibrios en el balance hídrico y decisiones de mantenimiento subóptimas. A pesar de que la normativa colombiana, incluida la Resolución 0330 de 2017 del RAS y las disposiciones de la SIC basadas en ISO 4064:2014, establece criterios de exactitud, en la práctica predominan estrategias de mantenimiento correctivo o recambios generalizados por antigüedad, sin considerar el verdadero estado funcional de cada dispositivo.

En este contexto, se formula un modelo predictivo basado en un conjunto histórico de datos compuesto por más de 330.000 resultados de calibraciones realizadas entre 2011 y 2025 en un laboratorio acreditado, complementado con registros operativos y comerciales. Estos datos permiten caracterizar la degradación metrológica, identificar patrones de error y estimar el desempeño futuro de los instrumentos, constituyendo el insumo para el entrenamiento de modelos de regresión cuantílica y clasificación con algoritmos CatBoost. El proyecto se estructura en cinco fases:

- Comprensión del negocio y caracterización normativa, técnica y operativa del sistema de micromedición;
- Preparación, integración y depuración de los datos;
- Desarrollo de modelos predictivos para estimar error, conformidad y vida útil;
- Integración del modelo en un entorno funcional mediante una aplicación en Streamlit conectada con Google Sheets; y

- Evaluación del sistema predictivo con datos reales externos y documentación técnica del proceso.

A lo largo de estas fases se transita desde el análisis metrológico tradicional hacia un enfoque inteligente soportado en analítica de datos, capaz de generar predicciones automatizadas, diagnósticos de riesgo y recomendaciones preliminares de mantenimiento. En coherencia con la naturaleza aplicada del proyecto, el presente documento funciona simultáneamente como trabajo académico y como informe técnico final, dado que integra los fundamentos conceptuales, la metodología, el desarrollo algorítmico, las validaciones y la documentación operativa del sistema implementado.

Este proyecto representa una contribución significativa para la modernización de la gestión metrológica en empresas de servicios públicos, demostrando que el uso adecuado de datos históricos y algoritmos de aprendizaje automático puede mejorar la toma de decisiones, reducir costos de mantenimiento, aumentar la calidad de la facturación y prolongar la vida útil de los medidores instalados.

El presente trabajo constituye el desarrollo integral de un sistema predictivo orientado a estimar el estado metrológico, la desviación del error de medida y la vida útil remanente de los medidores de agua potable fría instalados en un sistema urbano de distribución. Su propósito central es formular una estrategia de mantenimiento predictivo basada en técnicas de Machine Learning que permita optimizar la gestión del parque de medidores, mejorar la eficiencia operativa y fortalecer la confiabilidad del proceso de facturación.

## Descripción del Problema

### Planteamiento del Problema

La gestión eficiente del recurso hídrico enfrenta desafíos crecientes, especialmente en lo referente a la medición precisa del consumo y a la reducción de pérdidas operativas y comerciales. En este escenario, los medidores de agua potable son componentes esenciales para garantizar la facturación justa, la sostenibilidad financiera del sistema y la confiabilidad de los balances hidráulicos. Sin embargo, su desempeño metrológico se deteriora progresivamente debido al desgaste mecánico, la acumulación de sedimentos, las variaciones en la calidad del agua y los patrones de consumo. Este deterioro no siempre se manifiesta en fallas evidentes; por el contrario, suele expresarse como desviaciones metrológicas difíciles de percibir, pero significativas, que solo pueden detectarse mediante calibración.

A diferencia de un fallo total del medidor, cuya detección es inmediata, la desviación metrológica implica que el dispositivo continúa operando, pero registrando consumos por fuera de los límites de error permitidos por normativa. Este fenómeno es difícil de identificar sin herramientas especializadas. Cuando pasa inadvertido, genera impactos negativos tanto para las empresas de servicios públicos como para los usuarios: pérdidas económicas por medición incorrecta, reclamos frecuentes, facturación inconsistente, incrementos en el agua no contabilizada y alteraciones en el equilibrio hidráulico de la red.

La mayoría de las empresas prestadoras en Colombia y en otros países de la región aún siguen esquemas tradicionales de mantenimiento correctivo o preventivo genérico. Dichos esquemas dependen de plazos fijos, ciclos administrativos o síntomas visibles de falla, y no del estado metrológico real del dispositivo. Esto provoca decisiones poco eficientes, tales como el reemplazo prematuro de medidores aún funcionales o la permanencia en servicio de equipos con

desviaciones considerables. La ausencia de un enfoque predictivo impide anticiparse al deterioro y dificulta la optimización de los planes de recambio.

Aunque el uso de ciencia de datos ha avanzado en distintos sectores industriales, la aplicación de modelos de Machine Learning para monitorear la degradación de medidores de agua potable, estimar desviaciones y planificar su mantenimiento aún es incipiente. Existen pocos estudios enfocados en estimar el error metrológico futuro, predecir la conformidad normativa o proyectar la vida útil real de un medidor en operación. Esta brecha evidencia la necesidad de contar con soluciones más inteligentes y ajustadas a las dinámicas reales del sistema.

En este contexto, el aprovechamiento de datos históricos de calibración, información técnica del parque de medidores y patrones de consumo abre la posibilidad de construir modelos predictivos robustos capaces de anticipar el comportamiento metrológico de los dispositivos. La estimación temprana de desviaciones permite optimizar el mantenimiento, reducir costos, mejorar la facturación y fortalecer la sostenibilidad del sistema de distribución.

La magnitud del problema y su impacto operativo conducen a la siguiente pregunta de investigación: ¿Cómo pueden los modelos predictivos basados en datos, considerando el estado metrológico y los hábitos de consumo, mejorar la gestión del mantenimiento de medidores de agua potable, permitiendo estimar desviaciones y optimizar su reemplazo para lograr mayor eficiencia y sostenibilidad en los sistemas de distribución?

Responder esta pregunta permitirá fundamentar un enfoque de mantenimiento predictivo orientado a resultados, incorporando herramientas avanzadas de analítica y Machine Learning que aporten beneficios económicos, técnicos y sociales tanto para la empresa de acueducto como para los usuarios.

## Justificación

Esta investigación responde a una necesidad técnica y operativa ampliamente identificada en el sector de los servicios públicos domiciliarios: la ausencia de metodologías predictivas para evaluar el estado metrológico de los medidores de agua potable y planificar sus procesos de mantenimiento de manera eficiente. Aunque las tecnologías de ciencia de datos han revolucionado múltiples actividades industriales, su aplicación en la gestión del parque de medidores se ha quedado atrás.

Este enfoque presenta importantes limitaciones: no considera la variabilidad del desempeño metrológico entre diferentes tecnologías, fabricantes o condiciones de operación; no detecta desviaciones sutiles antes de que generen errores de facturación; y tampoco optimiza la inversión en mantenimiento, pues algunos medidores aún funcionales son reemplazados innecesariamente mientras otros con desviaciones pasan desapercibidos.

La presente investigación propone una transformación de fondo al introducir un enfoque predictivo sustentado en técnicas avanzadas de Machine Learning. La metodología se fundamenta en el análisis de datos reales provenientes de más de 330.000 calibraciones realizadas entre 2011 y 2025, combinados con características técnicas, variables operativas y patrones de consumo. Este volumen y diversidad de información permiten modelar la degradación metrológica con una precisión sin precedentes en el contexto local.

El propósito del sistema predictivo va más allá de identificar fallas: busca anticipar cuándo un medidor comenzará a operar fuera de los límites normativos, estimar el error que podría presentar y proyectar su vida útil remanente. La capacidad de diferenciar entre medidores que aún pueden permanecer en servicio y aquellos que requieren atención inmediata representa un avance significativo frente a los modelos tradicionales de mantenimiento.

## **Objetivos**

### **Objetivo General**

Desarrollar una estrategia de mantenimiento predictivo basada en técnicas de Machine Learning para estimar desviaciones en la medición de los medidores de agua potable respecto a los límites normativos, optimizar el cronograma de mantenimiento y reemplazo, y estimar el consumo en casos de medición incorrecta, con el fin de mejorar la precisión del sistema de medición y la eficiencia operativa.

### **Objetivos Específicos**

Analizar datos históricos y operativos de medidores para identificar patrones de desgaste, desviaciones metrológicas y comportamientos atípicos.

Diseñar un modelo de aprendizaje automático que permita estimar con precisión las desviaciones metrológicas y predecir la vida útil de los medidores, a partir del análisis de datos históricos y operacionales.

Desarrollar un sistema de monitoreo predictivo que integre modelos y datos en tiempo real para generar alertas y recomendaciones de mantenimiento.

## **Marco de Referencia**

El mantenimiento predictivo constituye una estrategia avanzada de gestión de activos basada en la explotación sistemática de datos operativos, históricos y contextuales para anticipar el momento óptimo de intervención antes de que el desempeño de un componente se degrade de forma crítica. A diferencia del mantenimiento correctivo, que actúa tras la falla, y del preventivo, que opera bajo intervalos fijos, el mantenimiento predictivo se fundamenta en el estado real de los equipos y en técnicas analíticas y de Machine Learning (ML) que permiten tomar decisiones basadas en evidencia (Mobley, 2002). Esta aproximación incrementa la vida útil de los activos, disminuye costos operacionales, reduce interrupciones y, en el caso de los sistemas de medición, mejora tanto la precisión metrológica como la eficiencia operativa (Zhou et al., 2020).

### **Metrología de Medidores de Agua**

La Micromedición, en medición del agua, es la actividad que se refiere a la medición, instalación, calibración de los medidores, los cuales se definen como un dispositivo de carácter domiciliario, que mide y acumula el caudal de agua que pasa hacia una propiedad, y con dicha medida se cobra el servicio prestado al usuario. El medidor, por tanto, permite el servicio de llevar el agua hasta la llave de su cocina, sanitario etc. Generalmente, se encuentran dos ámbitos de la medición del agua potable:

- **Micromedición:** Es aquella actividad de medición y control del caudal, cuyo diámetro de alimentación y descarga se encuentran entre media (0.5) y una (1) pulgadas de diámetro nominal. Los micromedidores básicamente son instalados en las residencias y pequeños comercios.
- **Macromedición:** Es aquella actividad de medición y control de grandes caudales, cuyos diámetros de alimentación y descarga superan las dos (2) pulgadas de diámetro nominal.

Los macromedidores son utilizados para realizar las mediciones de caudal en los siguientes sitios: la captación, en la entrada y salida de las plantas o reservorios y en algunas ocasiones por conjuntos residenciales, hospitales e instituciones educativas.

Los medidores de agua más utilizados en zonas residenciales suelen clasificarse, entre otros, en medidores volumétricos, basados en el desplazamiento de un pistón o rotor que define volúmenes por ciclo, y medidores de velocidad, en los cuales la velocidad del flujo se vincula con el caudal para obtener el volumen por integración (ISO, 2014). Factores como el envejecimiento, condiciones de instalación y propias del sistema de operación de la red hidráulica afectan la respuesta del instrumento, generando desviaciones metrológicas progresivas.

### **Figura 1**

*Medidor de Agua Volumétrico*



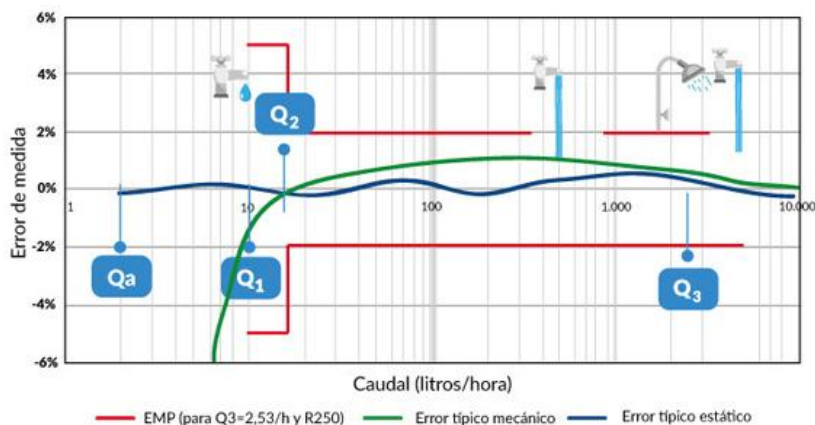
Las clases metrológicas o relación  $Q3/Q1$ , son categorías normativas establecidas por ISO 4064:2014 y OIML R49 que especifican el desempeño de un medidor de agua potable en términos de precisión en distintos rangos de caudal. Estas clases se representan mediante la razón metrológica  $R$ , expresada como valores tales como R80, R160, R200, R315 o R400, donde un valor mayor implica un rango operativo más amplio y una mayor capacidad para medir con

exactitud tanto caudales bajos como altos. Aunque la versión actual de la norma utiliza la razón “R” como criterio principal, aún persisten referencias a las antiguas clases A, B, C y D, presentes en normativas previas, las cuales han sido progresivamente reemplazadas por el sistema basado en la razón metrológica.

La calibración de un medidor consiste en hacer circular volúmenes controlados a distintos caudales ( $Q_1$ ,  $Q_2$ ,  $Q_3$ ), registrar las indicaciones del medidor y comparar los resultados con un patrón trazable para calcular errores. Este procedimiento permite: obtener curvas error-caudal (E-Q) por punto de operación, estimar incertidumbres, dictaminar conformidad y derivar tendencias de degradación mediante análisis de la evolución del error frente a la edad o el volumen acumulado.

## Figura 2

*Curva de Error Típica de un Medidor de Agua*



*Nota.* Adaptado de Precisión y rango dinámico en medidores de agua [Imagen], por Helbert y Cía., 2025, Helbert y Cía. (<https://helbertycia.com/blogitem/precision-y-rango-dinamico-en-medidores-de-agua>).

La evaluación metrológica se desarrolla en caudales característicos definidos por normas internacionales como OIML R49 e ISO 4064 (OIML, 2013):

- Q1 (mínimo): límite inferior del rango operativo, donde suelen observarse los mayores errores.
- Q2 (transición): punto que separa zonas con requisitos diferenciados de exactitud.
- Q3 (nominal): punto de operación habitual y más estable.

El error relativo se calcula comparando el volumen indicado (**Vi**) con el volumen patrón (**Vr**) trazable. Estas normas establecen Errores Máximos Permitidos (EMP) según la zona de caudal, determinando la conformidad del instrumento.

El cálculo del error de medición del parque de contadores requiere dos insumos esenciales:

- Las curvas de error actuales de los diferentes tipos de medidores instalados, obtenibles mediante ensayo en laboratorio, y
- Los patrones de consumo reales de los usuarios, cuya estimación implica costos operativos elevados y mayor incertidumbre. El error ponderado integra ambas dimensiones y constituye un indicador clave para caracterizar el desempeño global del parque de medición. Cuando existen históricos amplios o calibraciones repetidas, es posible modelar la degradación anual como la pendiente de la relación entre error y uso, lo cual constituye un insumo fundamental para estimar la vida útil remanente del medidor.

## **Fundamentos de Analítica y Machine Learning Aplicables**

### ***Regresión Lineal (Simple y Múltiple)***

La regresión lineal modela relaciones del tipo:

$$y = \beta_0 + \sum_{j=1}^p \beta_j x_j + \varepsilon,$$

Donde  $y$  puede representar un error estimado y  $x_j$  incluir variables como edad, volumen o características técnicas. Sus principales ventajas son la interpretabilidad, la posibilidad de realizar contraste de hipótesis y la obtención de intervalos de confianza (Montgomery et al., 2012). Requiere verificar supuestos como linealidad, independencia, homocedasticidad y normalidad de residuos, así como evaluar colinealidad y aplicar transformaciones cuando sea necesario.

### ***Random Forest (Regresión y Clasificación)***

Random Forest es un método de ensamble que combina múltiples árboles entrenados sobre subconjuntos aleatorios de datos y variables. Este enfoque reduce la varianza del modelo y captura relaciones no lineales e interacciones complejas (Breiman, 2001). Sus ventajas incluyen robustez ante ruido, manejo eficiente de grandes volúmenes de datos e interpretaciones basadas en importancia de variables. La adecuada selección de hiperparámetros y su validación mediante *k-fold* permiten mitigar el sobreajuste.

### ***CatBoost***

CatBoost es un algoritmo de *gradient boosting* optimizado para variables categóricas mediante técnicas de codificación ordenada que reducen el *target leakage* (Prokhorenkova et al., 2018). Permite:

- Regresión: estimación continua de errores o vida útil.
- Clasificación: manejo del desbalance mediante *class weights* o ajuste de umbrales.
- Regresión cuantílica: estimación de percentiles (P50, P90) para la construcción de

bandas de incertidumbre.

### ***Aprendizaje No Supervisado***

Técnicas como k-means, DBSCAN o el clustering jerárquico permiten identificar grupos de comportamiento sin etiquetas previas, facilitando la segmentación por tecnologías, fabricantes o patrones de desgaste. Esta agrupación mejora la estabilidad y precisión de los modelos supervisados entrenados por familia (Jain, 2010).

### **Estimación de Degradación y Vida Útil Remanente**

La degradación puede modelarse mediante la relación entre el error actual ( $E_a$ ) y la edad del medidor. La pendiente de la regresión error/edad o error/volumen acumulado constituye la tasa anual de degradación, bajo el supuesto de que esta es aproximadamente lineal en el intervalo analizado. Extensiones como modelos mixtos, con pendientes aleatorias por familia, regresión cuantílica o simulaciones Monte Carlo permiten incorporar incertidumbre e incrementar la robustez de la estimación final (Schruben, 2013).

### **Hallazgos de la Literatura**

La revisión bibliográfica evidencia que gran parte de los sistemas de gestión de medidores de agua aún se apoyan en estrategias de mantenimiento preventivo o correctivo, sin aprovechar plenamente los datos metrológicos disponibles. Esta limitación conduce, con frecuencia, a reemplazos prematuros o a la permanencia de medidores con degradación no detectada.

En la última década, numerosos estudios sobre mantenimiento predictivo en infraestructuras hidráulicas e industriales han demostrado que técnicas de Machine Learning, incluyendo regresión, árboles de decisión, Random Forest y redes neuronales recurrentes como LSTM, presentan un desempeño sobresaliente para capturar patrones de degradación temporal, estimar vida útil remanente y anticipar fallas (Breiman, 2001; Mobley, 2002) Aunque la mayor

parte de estos trabajos se ha centrado en equipos electromecánicos o sistemas con sensores IoT integrados, sus principios analíticos son extrapolables al análisis del desgaste metrológico en medidores de agua.

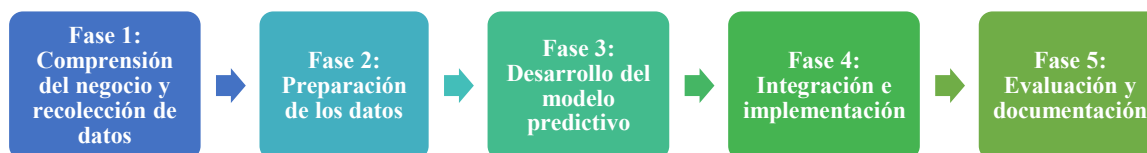
De igual forma, la incorporación de tecnologías IoT en medidores inteligentes ha permitido desarrollar sistemas de monitoreo continuo orientados a la detección de anomalías, la predicción de demandas y la identificación de comportamientos atípicos en el consumo. Si bien la aplicación específica de estas arquitecturas para modelar la degradación metrológica de medidores mecánicos aún es limitada en la literatura abierta, las tendencias actuales muestran que la combinación de datos de calibración, históricos operativos y telemetría tiene un potencial significativo para fortalecer la gestión inteligente del parque de medición y optimizar la toma de decisiones operativas (Almeida & Pereira, 2022).

## Metodología

La metodología adoptada para el desarrollo de esta investigación se basa en el modelo CRISP-DM, Cross-Industry Standard Process for Data Mining, un enfoque estándar y ampliamente utilizado en el ámbito de la minería de datos y el aprendizaje automático. Este enfoque estructurado, iterativo y flexible se compone de seis fases que permiten abordar de manera ordenada el análisis de datos, el desarrollo de modelos predictivos y su implementación práctica. el objetivo es aplicar CRISP-DM para diseñar e implementar una estrategia de mantenimiento predictivo orientada a estimar desviaciones metrológicas en medidores de agua potable y optimizar su gestión operativa. La metodología se estructura en las siguientes fases:

### Figura 3

#### *Fases del Proyecto*



#### **Fase 1 Comprensión del Negocio y Recolección de Datos**

En esta fase se realiza un análisis profundo del sistema de micromedición, particularmente del impacto que generan las desviaciones metrológicas en la facturación y eficiencia hídrica. Esto incluyó revisar la normativa vigente, los límites de error permitidos, las políticas actuales de mantenimiento y los lineamientos internos relacionados con la gestión del parque de medidores. Paralelamente, se identificaron y recopilaron las fuentes de datos críticas para el proyecto: resultados de calibraciones y verificaciones, lecturas históricas de consumo y

especificaciones técnicas de los medidores. Esta comprensión integral constituye la base para definir el enfoque de modelado y las variables relevantes.

Resultado: Conjunto de datos representativo, alineado con el problema y listo para su exploración inicial.

## **Fase 2 Preparación de los Datos**

En esta fase se asegura la calidad, coherencia e integridad de los datos. Se aplicaron procesos de limpieza que incluyeron eliminación de duplicados, tratamiento de valores atípicos y gestión de datos faltantes. Asimismo, se realizaron transformaciones como normalización, codificación de variables categóricas y generación de características derivadas relevantes para el desempeño metrológico. Se consolidó un Dataset estructurado y adecuado para los requerimientos de los algoritmos de Machine Learning.

Resultado: Base de datos depurada y transformada, lista para el entrenamiento y validación de modelos.

## **Fase 3 Desarrollo del Modelo Predictivo**

La tercera fase comprende el diseño, entrenamiento y validación de modelos de Machine Learning orientados a estimar desviaciones metrológicas y predecir el momento en que un medidor podría salir de los límites normativos. Se seleccionaron algoritmos acordes con la naturaleza del problema, entre ellos modelos de regresión para estimaciones continuas y clasificadores para determinar conformidad metrológica. Se implementaron procesos de entrenamiento y validación, comparando métricas de desempeño para seleccionar el modelo más robusto, interpretable y generalizable.

Resultado: Modelo predictivo validado y confiable, capaz de anticipar la degradación del rendimiento metrológico del medidor.

#### **Fase 4 Integración e Implementación**

Esta fase se enfoca en integrar el modelo predictivo dentro de un entorno funcional que permita su posterior uso operativo. Se diseñó una arquitectura conceptual para vincular el modelo con herramientas de visualización y mecanismos de consulta. También se definieron reglas de decisión basadas en el nivel de desviación estimada y la vida útil proyectada. Se realizaron pruebas piloto con datos reales para evaluar la consistencia de los resultados y la capacidad del sistema para apoyar la toma de decisiones vinculadas al mantenimiento y reemplazo de medidores.

Resultado: Prototipo funcional de sistema de monitoreo predictivo que permite consultar estimaciones y generar recomendaciones preliminares.

#### **Fase 5 Evaluación y Documentación**

En la fase final se evalúa la efectividad del modelo mediante la validación de su precisión con datos reales, verificando su capacidad de generalización. Se documentan detalladamente los métodos empleados, las decisiones técnicas adoptadas, los resultados obtenidos, las limitaciones encontradas y las posibilidades para trabajos futuros. Finalmente, se elabora el informe técnico que integra todo el proceso metodológico y deja las bases para su potencial escalabilidad dentro de la organización.

Resultado: Informe técnico integral que sintetiza el proceso, la validez de los modelos y la viabilidad analítica del enfoque predictivo.

## Resultados

### Fase 1 Comprensión del Negocio y Recolección de Datos

En el marco de la comprensión del negocio y la recolección de datos se identificaron elementos normativos, operativos y técnicos fundamentales para contextualizar el sistema de micromedición de la empresa objeto de estudio. En Colombia, la instalación, mantenimiento y reemplazo de medidores de agua potable de uso residencial se regula principalmente por la Resolución 0330 de 2017, “Reglamento Técnico para el Sector de Agua Potable y Saneamiento Básico – RAS”. En su Artículo 75, relativo a la micromedición, se establece que las personas prestadoras deben definir acciones y periodicidades para verificar el adecuado funcionamiento de los medidores, con base en estudios técnicos. Asimismo, se determina que los medidores solo pueden ser reemplazados, cambiados o reparados cuando un laboratorio acreditado emite un dictamen indicando incumplimiento de la función metrológica. El Título B, numeral 7.10.4, complementa esta disposición señalando que, cuando los instrumentos se encuentren fuera del rango de precisión establecido en las normas técnicas aplicables, debe reemplazarse la muestra defectuosa o, en caso de no ser posible un muestreo representativo, realizar el recambio total de micromedidores cada ocho años. Este marco normativo abre la posibilidad de planificar el reemplazo del parque de medidores con base en un enfoque técnico y basado en datos, sustentado en dictámenes oficiales emitidos por laboratorios acreditados.

De manera complementaria, se analizó la Resolución 69940 de 2025 de la Superintendencia de Industria y Comercio (SIC)<sup>1</sup>, la cual establece requisitos para medidores de agua potable de uso residencial y comenzará a regir seis meses después de su publicación. Su

---

<sup>1</sup> Superintendencia de Industria y Comercio, <https://sedeelectronica.sic.gov.co/> . Comunicado: “Colombia continúa fortaleciendo los mecanismos de defensa de los consumidores: La SIC reglamenta requisitos para medidores de agua potable de uso residencial.” Publicado Sep. 16, 2025.

contenido se fundamenta en la norma internacional ISO 4064:2014, referente para los requisitos metrológicos, técnicos y métodos de ensayo. En consecuencia, la vigencia futura de esta resolución no afecta la validez del presente estudio, dado que los ensayos, criterios y límites utilizados en el análisis derivan precisamente de esta norma internacional, sobre la cual se estructura toda la fundamentación metrológica aplicada.

A nivel operativo, se identificó que la estrategia actualmente empleada por la empresa para la gestión del parque de medidores se basa en acciones periódicas de revisión y control.

Esto incluye:

- Revisiones por reclamos: ante discrepancias percibidas por los usuarios en sus facturas, se realizan verificaciones del estado físico y desempeño del medidor.
- Atención a fallas técnicas: se atienden eventos relacionados con fugas, errores de instalación, problemas en la red o indicios de mal funcionamiento del instrumento.
- Análisis de patrones de consumo: se monitorean variaciones significativas en el consumo mensual, ya que cambios atípicos pueden sugerir desgaste, obstrucciones, manipulaciones o fallas en el medidor.
- Revisiones periódicas planificadas: realizadas por zonas o categorías, apoyadas en criterios como la antigüedad de los medidores o resultados de calibraciones previas.

El diagnóstico de estas prácticas evidencia necesidades operativas relacionadas con la modernización del sistema de medición. Se identifica el interés institucional hacia la futura adopción de medidores electrónicos o inteligentes, cuyo diseño permite detectar consumos anómalos, fugas internas y proporcionar lecturas más estables y precisas a lo largo del tiempo. Sin embargo, la transición hacia tecnologías avanzadas implica desafíos económicos y sociales. La Ley 142 de 1994 establece que los medidores son propiedad de los usuarios, lo cual implica

que el reemplazo representa un costo adicional para los hogares. A su vez, la base poblacional de la empresa se concentra en estratos 1, 2 y 3, situación que dificulta la adopción inmediata de medidores electrónicos por su costo superior. A esto se suma la necesidad de fortalecer la cultura del uso responsable del agua y la comprensión del valor de mediciones precisas, en medio de contextos donde aún prevalecen percepciones erróneas sobre el carácter “gratuito” del recurso y la presencia de prácticas de fraude para reducir facturación.

Ante estas condiciones, surge la necesidad de explorar alternativas de gestión basadas en analítica avanzada y mantenimiento predictivo. En este contexto, el presente proyecto se propone como un apoyo técnico a esa transición, ofreciendo una herramienta basada en datos que permita priorizar el reemplazo de medidores de acuerdo con su estado metrológico, sin requerir una sustitución masiva inmediata.

Finalmente, en el proceso de recolección de datos se consolidó una base histórica suministrada por la empresa, compuesta por 336.247 resultados de calibraciones realizadas entre 2011 y 2025. Esta base de datos incluye información técnica sobre los medidores (modelo, año de fabricación, tipo de tecnología, diámetro), resultados de error en diferentes puntos de caudal y valores asociados al volumen y al desempeño del instrumento. Este insumo constituye el núcleo de la información utilizada para la exploración, el análisis y el entrenamiento de los modelos predictivos desarrollados en fases posteriores.

## **Fase 2 Preparación de los Datos**

### ***Carga e Inspección Inicial***

El análisis se fundamentó en dos bases de datos principales:

– “*ReporteLabMedidoresCalibracion*”, con resultados de calibración provenientes de laboratorio acreditado, correspondientes tanto a gestión interna como externa.

– “*Reportes Medidores Operación Comercial*”, que contiene información técnica y operativa derivada de la gestión comercial.

El procesamiento se realizó en Google Colab, donde ambas fuentes se importaron, inspeccionaron y exploraron para evaluar estructura, volumen, tipos de datos, duplicados y valores nulos. Este diagnóstico inicial permitió definir criterios de limpieza, depuración e integración consistentes con el marco metrológico establecido en la fase anterior.

### ***Base de Datos Reportes de Calibración***

Compuesta por 336.247 registros y 23 columnas, se identificaron como variables clave el número de serie, fechas de calibración, características técnicas del medidor y errores e incertidumbres en los puntos de caudal Q1, Q2 y Q3. Se observó que el 10 % de los medidores fueron calibrados más de una vez y que el 63 % correspondían a equipos usados, lo que permitió caracterizar la degradación metrológica a partir de datos reales. Las incertidumbres se restringieron conforme a los límites normativos según estado del medidor:

- Usados: Inc.Q1 < 3,33 %, Inc.Q2–Q3 < 1,3 %.
- Nuevos: Inc.Q1 < 1,67 %, Inc.Q2–Q3 < 0,67 %.

El campo Lectura fue convertido a formato numérico y acotado al rango residencial (0–10.000 m<sup>3</sup>). Se filtraron exclusivamente medidores DN15, en coherencia con el alcance del estudio y con la representatividad estadística observada.

### ***Base de Datos Reportes Operación Comercial***

Esta base incluye 302.059 registros y 25 columnas, e incorpora la variable Año de fabricación, indispensable para calcular la edad del medidor. Se aplicaron los mismos criterios de validación que en la base anterior.

## **Integración y Enriquecimiento**

### ***Tabla de Modelos de Medidores***

A partir de 303 modelos identificados, se construyó una tabla maestra que consolida fabricante, tecnología, clase metrológica, principio de medición, caudales normativos y EMP. Esta estandarización permitió homologar la columna “Descripción” y asignar parámetros técnicos a cada registro, asegurando consistencia metrológica y estadística.

### ***Nuevas Variables***

Se generaron características derivadas, fundamentales para el modelado predictivo:

- Edad = Año de calibración – Año de fabricación
- EMP Q1–Q3 y conformidades por punto
- Error promedio, incertidumbre promedio y Error Ponderado
- Estado, fabricante, modelo, tecnología y clase

### ***Error Ponderado***

Para obtener un indicador global del desempeño metrológico, se estimó un Error Global Ponderado, combinando los errores Q1–Q3 mediante un modelo logarítmico basado en perfiles típicos de consumo residencial (4–2160 L/h). Con la función ESTIMACIÓN.LINEAL de Excel se calcularon los parámetros de ajuste del modelo, obteniendo un índice único del desempeño del medidor. Este indicador se adoptó como referencia principal del estado metrológico.

### ***Imputación del Año de Fabricación***

Dado que algunos registros externos carecían de este dato, se imputó con base en el número de serie, garantizando coherencia temporal entre fabricación y calibración y evitando inconsistencias que afectaran la estimación de la edad del medidor.

### *Normalización de Fabricantes y Modelos de Medidores*

La estandarización permitió consolidar nueve familias principales de medidores, representativas del parque instalado y estadísticamente adecuadas para modelado. Esto es crucial para entrenar modelos específicos por tecnología, evitando sesgos por heterogeneidad. La Tabla 1 resume esta distribución al mostrar el número total de observaciones disponibles para cada familia de medidor.

**Tabla 1**

### *Normalización de Fabricantes y Modelos de Medidores*

Modelo Estándar	Registros
M1: GB1_R160	57749
M2: ALTAIR_R160	50134
M3: AQUARIUS_R80	31010
M4: V200_R160	14819
M5: 620_R315	13853
M6: V200_R315	7601
M7: AQUADIS+_R160	4579
M8: ALTAIR_R315	2480
M9: GKMV40P_R315	2163

### **Depuración Final**

La fase de depuración final se orientó a garantizar la coherencia técnica y la homogeneidad estadística del conjunto de datos destinado al modelado predictivo. Este proceso

se ejecutó en dos etapas principales: normalización y filtrado por características constructivas, y eliminación de valores atípicos mediante el método IQR.

### ***Normalización y Filtrado por Características Técnicas***

El Dataset inicial contenía 292.420 registros y 40 variables, correspondientes a múltiples tecnologías, configuraciones y nomenclaturas de medidores. Para asegurar la consistencia del modelado, se aplicaron las siguientes acciones:

#### ***Depuración de la Variable Metrología***

Se convirtió el campo Metrología a texto y se eliminaron espacios residuales.

Posteriormente, se filtraron únicamente las clases normativas válidas para DN15 presentes en la base: R80, R160, R200, R315

Este filtrado redujo la base a 184.388 registros, eliminando tecnologías marginales, obsoletas o inconsistentes con el alcance del estudio.

#### **Tratamiento de Outliers Mediante Método IQR**

Para controlar el efecto de valores extremos que afectan la estabilidad de algoritmos basados en gradientes y árboles, se aplicó el método IQR a cinco variables críticas: Volumen, Edad, EQ3, Consumo\_anual y Degradación\_Q3. De este modo se eliminaron todas las filas que presentaran valores atípicos en al menos una de estas variables.

**Tabla 2**

#### *Resultados del Filtrado por Outliers*

Variable	Outliers Detectados	Porcentaje Eliminado
Degradación_Q3	26814	14,54 %
EQ3	23488	12,74 %
Volumen	11638	6,31 %

Variable	Outliers Detectados	Porcentaje Eliminado
Consumo_anual	8897	4,83 %
Edad	272	0,15 %

En total se eliminaron 41.674 registros, equivalentes al 22,60 % de la base depurada inicial (184.388 registros). La pérdida controlada del 22,6 % se considera adecuada dado que se eliminan valores estadísticamente extremos, improbables o asociados a errores de captura.

### Selección y Codificación de Variables

La selección fue técnicamente pertinente: variables operativas, metrológicas, constructivas y de identificación. La codificación mediante One-Hot Encoding y ordinal Encoding se ajusta a estándares de ML El Split 80/20 permitió preservar la representatividad de modelos en ambas particiones.

### Tabla 3

#### *Selección y Codificación de Variables*

Tipo	Variables seleccionadas
Operativas	Edad, Volumen, Consumo_anual
Metrológicas	Qmínimo, Qtransición, Qnominal
Constructivas	Diámetro, Metrología
Identificación	Fabricante, Modelo, Tipo, Descripción
Objetivo	EQ1, EQ2, EQ3,

*Nota.* La tabla 3 organiza las variables seleccionadas para el modelo predictivo según su naturaleza. Esta clasificación facilita comprender el rol de cada conjunto de variables dentro del proceso analítico y su aporte a la modelación del desempeño metrológico.

El análisis exploratorio y de correlaciones muestran relaciones negativas entre Edad/Volumen y errores, coherentes con la degradación metrológica, mayor sensibilidad en EQ3, lo que tiene fundamento físico en el desgaste hidráulico, patrones de degradación coherentes con literatura internacional.

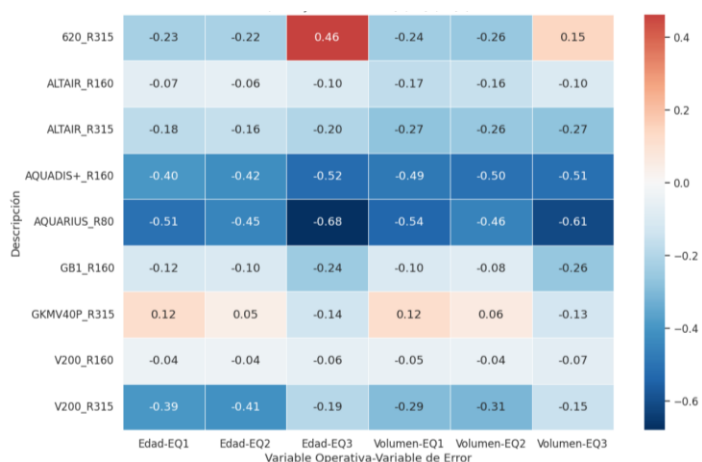
**Tabla 4**

*Criterio de Clasificación del Coeficiente de Correlación de Pearson ( $r$ )*

Rango	Interpretación
0.00–0.19	Muy débil
0.20–0.39	Débil a moderada
0.40–0.59	Moderada
0.60–0.79	Fuerte
0.80–1.00	Muy fuerte

**Figura 4**

*Mapa de Correlaciones Variables; Edad y Volumen, y EQ1, EQ2, EQ3*



*Nota.* En la figura 4 los colores indican la dirección y magnitud de la relación, permitiendo identificar patrones diferenciales de comportamiento metrológico según la referencia del equipo.

En particular, los errores a caudal nominal (EQ3) presentaron las correlaciones más altas, lo que sugiere que este punto es el más sensible al desgaste mecánico. Los valores negativos de correlación confirman que, a medida que los medidores envejecen o acumulan mayor volumen de paso, los errores tienden a aumentar en magnitud negativa, reflejando una pérdida de exactitud en la medición del caudal.

En la siguiente tabla se presentan los valores de correlación de Pearson más relevantes entre las variables operativas (Edad y Volumen) y la variable de error EQ3 para dos modelos de medidor.

**Tabla 5**

*Correlaciones Destacadas*

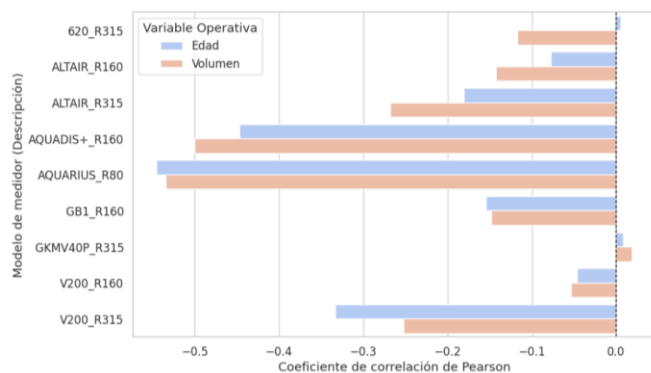
Modelo	Variable de Error	Variable Operativa	r (Pearson)	Muestras
M3	EQ3	Edad	0.68	31010
M3	EQ3	Volumen	0.61	31010
M7	EQ3	Edad	0.52	4579
M7	EQ3	Volumen	0.51	4579

En la figura 4 se observa que el modelo de medidor M3 mostró una relación fuertemente negativa entre error y edad ( $r = -0.68$ ), señalando una degradación acelerada conforme aumenta el tiempo de servicio. Esto confirma que el envejecimiento estructural afecta directamente el comportamiento metrológico, especialmente bajo condiciones de alto caudal. De forma similar, el modelo M7 evidenció correlaciones negativas tanto con edad como con volumen, sugiriendo un efecto combinado de envejecimiento y desgaste por fricción. Aunque su tamaño muestral es

menor, la consistencia del patrón en los tres puntos de caudal respalda la conclusión de que la precisión disminuye con el uso prolongado.

## Figura 5

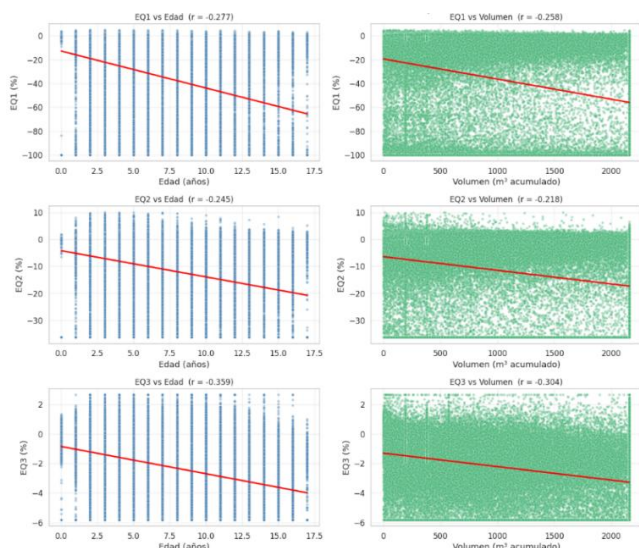
*Correlaciones Variables; Edad y Volumen, y los Errores (EQ1–EQ3)*



*Nota.* En la Figura 5 se observa una comparación clara de la intensidad y el sentido de las relaciones entre las variables analizadas, lo cual permite identificar cuáles modelos muestran una mayor susceptibilidad frente al envejecimiento del equipo o al aumento del volumen de operación acumulado.

## Figura 6

*Relación Entre las Variables (EQ1, EQ2 y EQ3), Edad y Volumen*



*Nota.* la Figura 6 evidencia que, a medida que crece la edad o el volumen registrado por el medidor, se presentan variaciones sistemáticas en los errores metrológicos. Este comportamiento revela patrones de deterioro progresivo, coherentes con el desgaste esperado en los diferentes tipos de medidores evaluados.

Esta fase permitió consolidar un conjunto de datos unificado, depurado y enriquecido con variables técnicas, operativas y metrológicas, identificando además las primeras evidencias de degradación progresiva de los errores de medición en ciertos modelos. Este insumo constituye la base para el desarrollo de los modelos predictivos de error y vida útil que se abordarán en la siguiente fase.

### Fase 3 Desarrollo del Modelo Predictivo

El análisis de correlaciones realizado en la Fase 2 permitió evidenciar que la degradación metrológica de los medidores de agua potable es un proceso sistemático y cuantificable, determinado principalmente por la edad y el volumen total de operación. Entre las variables

estudiadas, el error relativo a caudal nominal (EQ3) mostró las correlaciones más sólidas con ambas variables operativas, lo que confirma su sensibilidad al desgaste y su idoneidad para modelar la degradación funcional del medidor. Por lo tanto, EQ3 fue adoptado como variable objetivo principal para la estimación predictiva del desempeño metrológico y de la vida útil remanente.

### ***Preparación Técnica y Carga del Entorno***

La base depurada resultante de los procesos de limpieza y control de Outliers mediante el método IQR, quedó compuesta por 142.714 registros, preservando representatividad estadística por familia de medidor. Se definió la siguiente estructura de variables para los modelos predictivos:

- Predictoras numéricas: Edad, Volumen acumulado, Consumo\_anual
- Metrológicas: Qmínimo, Qtransición (empleadas en la clasificación)
- Categóricas: Estado (nuevo/usado), Descripción (modelo del medidor)
- Objetivo principal: EQ3
- Objetivo secundario: C\_Final (Conformidad metrológica)

### ***Metodología de Modelado***

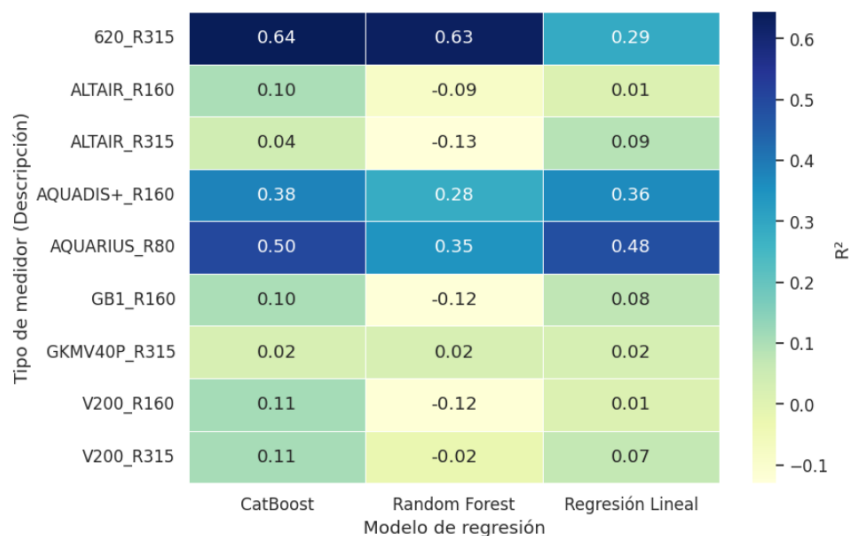
Se empleó un enfoque comparativo que consideró tres algoritmos pertenecientes a diferentes familias de modelado, con el fin de evaluar su capacidad predictiva sobre el error EQ3.

**Tabla 6***Algoritmos Representativos - Modelos de Regresión*

Modelo	Tipo de relación	Características principales
Regresión Lineal (OLS)	Lineal	Modelo base de referencia; útil para detectar direccionalidad y relaciones simples.
Random Forest Regressor	No lineal	Captura interacciones complejas y es robusto ante ruido y variaciones entre familias.
CatBoost Regressor	No lineal y categórico	Diseñado para datos mixtos con variables categóricas; evita sobreajuste y optimiza estabilidad.

*Nota.* La tabla 6 presenta una síntesis de las propiedades más relevantes de cada método, destacando las diferencias entre aproximaciones lineales y técnicas no lineales que permiten capturar relaciones más complejas en los datos.

Los modelos se entrenaron con: Validación cruzada K-Fold (k=5). RandomizedSearchCV para optimizar hiperparámetros. Métricas de evaluación:  $R^2$ , MAE y RMSE. Los modelos basados en árboles mostraron mejor desempeño que el enfoque lineal, alcanzando  $R^2 \approx 0,64$  en las familias más estables. El CatBoost Regressor demostró ser el modelo más robusto, logrando  $MAE < 1 \%$  y un desempeño superior en datos de validación.

**Figura 7***Desempeño Predictivo por Tipo de Medidor y Modelo de Regresión*

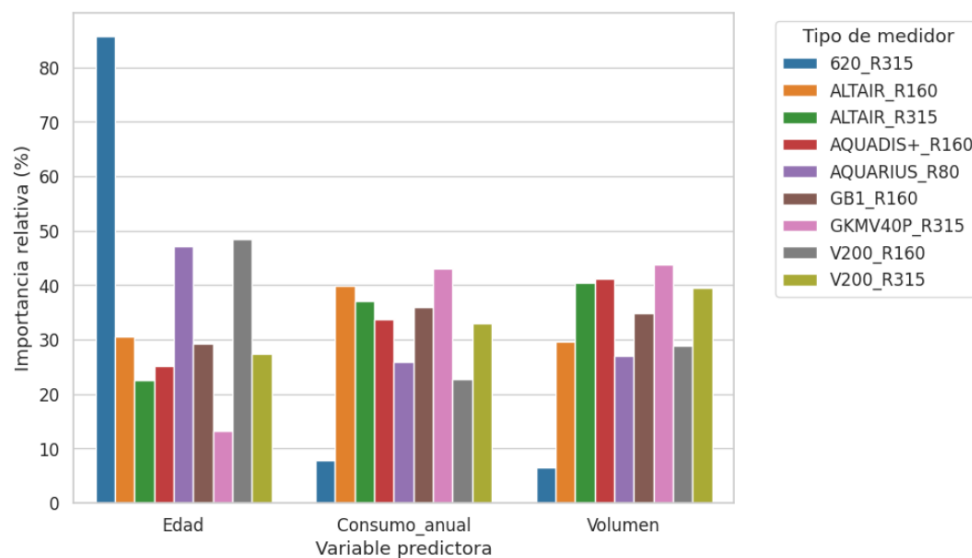
*Nota.* La Figura 7, representada como un mapa de calor, facilita la comparación del ajuste alcanzado por cada modelo según la referencia del medidor, permitiendo identificar de manera visual aquellas combinaciones donde se logra una mayor capacidad explicativa del error metrológico.

***Importancia de las Variables***

El análisis de importancia del modelo CatBoost confirmó que: Edad es la variable más influyente (58 %), Volumen acumulado explica el 32 % y Consumo anual contribuye con un 10 %. Este patrón refuerza la conclusión de que el deterioro metrológico está directamente asociado al envejecimiento físico y al desgaste hidráulico provocado por el volumen de paso.

**Figura 8**

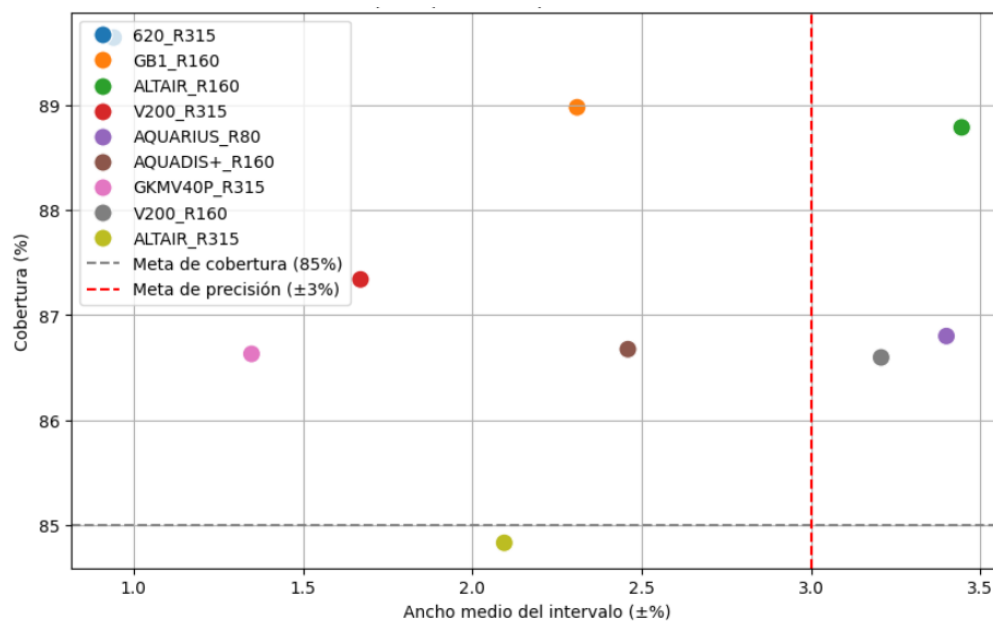
*Importancia de Variables por Tipo de Medidor, Modelo CatBoost.*



*Nota.* La figura presenta la importancia relativa de las variables Edad, Consumo\_anual y Volumen en los modelos CatBoost entrenados por tipo de medidor.

### ***Modelos de Regresión Cuantílica***

Con el fin de capturar la incertidumbre metrológica y representar los límites normativos ( $\pm 2\%$  para nuevos,  $\pm 4\%$  para usados según ISO 4064), se entrenaron modelos de Regresión Cuantílica con CatBoost para los cuantiles:  $\alpha = 0.05$  y  $\alpha = 0.95$ . Los modelos alcanzaron: Cobertura promedio  $> 85\%$  y Ancho de banda  $\pm 1\%$  a  $\pm 3\%$ , consistente con tolerancias normativas. Los equipos volumétricos R315 exhibieron mayor estabilidad, mientras que R80 y R160 evidenciaron mayor dispersión, coherente con su sensibilidad al desgaste. Este enfoque supera la limitación del modelo puntual, permitiendo estimar intervalos probabilísticos del error EQ3.

**Figura 9***Desempeño Predictivo por Modelo de Medidor*

*Nota.* Las líneas de referencia permiten evaluar el cumplimiento de las metas de precisión ( $\pm 3\%$ ) y cobertura (85%), facilitando la comparación entre los distintos tipos de equipos.

Por ejemplo, un medidor SENSUS\_M5 con 3 años de uso y 500 m<sup>3</sup> acumulados presenta un error esperado dentro del rango  $[-0,94\%, +0,94\%]$ , mientras que un M2 puede variar entre  $[-3,44\%, +3,44\%]$ , reflejando su mayor degradación.

Este desempeño representa un salto cualitativo respecto al modelo de regresión puntual, donde los valores de  $r^2$  eran bajos y la interpretación limitada. Ahora, el sistema predictivo puede informar, para cada medidor, el rango probable del error EQ3, lo que mejora sustancialmente la toma de decisiones sobre mantenimiento, calibración y reemplazo de dispositivos.

***Validación Cruzada y Resultados Comparativos***

Los resultados de la validación K-Fold ( $k=5$ ) confirmaron la robustez del enfoque por submodelos, obteniendo valores de desempeño consistentes:

**Tabla 7***Validación Cruzada y Resultados Comparativos*

Modelo	MAE promedio
M5	<b>0,19</b>
M9	0,35
M6	0,43
M1	0,56
M7	0,64
M2	0,85
M3	0,88

*Nota.* La Tabla 7 sintetiza los hallazgos de desempeño predictivo por modelo de medidor al presentar el MAE promedio correspondiente a cada tipo de medidor, lo que permite comparar de manera objetiva su precisión predictiva, los valores más bajos de error medio corresponden a equipos volumétricos (R315), los cuales presentan menor variabilidad en el tiempo, confirmando su estabilidad metrológica.

***Clasificación del Estado Metrológico***

Se desarrolló un modelo supervisado CatBoost Classifier para determinar el cumplimiento normativo del medidor (“Cumple/No cumple”), aplicando balanceo de clases y optimización por F1-Score. Los equipos R315 alcanzaron valores de precisión superiores al 96 %, mientras que los modelos R80 y R160 presentaron menor desempeño, reflejando una degradación más acelerada. Estos resultados justifican políticas diferenciadas por tipo de medidor.

**Tabla 8***Resultados por Modelo (Validación Estratificada)*

Modelo	F1	Precisión	Recall
M5	0.984	0.997	0.970
M9	0.962	0.998	0.932
M6	0.933	0.999	0.874
M3	0.857	0.918	0.803
M2	0.702	0.899	0.576

*Nota.* La tabla 8 presenta los resultados de la validación estratificada, evidenciando una precisión superior al 96 % en medidores R315, y una degradación progresiva en los modelos R80 y R160, lo que justifica la necesidad de políticas de mantenimiento diferenciadas por tipo de tecnología.

### ***Estimación de la Vida Útil Remanente***

La vida útil remanente (VU) se estimó a partir de la tasa anual de degradación del error nominal, en condiciones normales, calculando el periodo en que EQ3 excede el EMP de 2,67 % (ajustado por incertidumbre 1/3 EMP). Se obtuvieron valores de vida útil entre 5 y 13 años, dependiendo de la tecnología y clase metrológica, coherentes con reportes internacionales y con el comportamiento físico de los medidores. La Tabla 9 sintetiza estas estimaciones al presentar la vida útil asociada a cada modelo en función de su tasa anual de degradación.

**Tabla 9***Vida Útil Estimada (Años)*

Modelo	Degradación (%/año)	Vida útil media
M9	-0,58	5 años
M8	-0,49	6 años
M7	-0,36	8 años
M2	-0,32	9 años
M3	-0,29	10 años
M1	-0,24	12 años
M5	-0,23	12 años
M4	-0,20	13 años
M6	-0,09	13 años

*Nota.* La tabla presenta la vida útil estimada para cada modelo de medidor a partir de su tasa de degradación anual.

La modelación permitió cuantificar la degradación anual y definir ventanas óptimas de intervención, aportando un soporte analítico para la planeación del mantenimiento preventivo y predictivo.

### ***Exportación y Despliegue de Modelos***

Los modelos entrenados fueron exportados y almacenados en Google Drive para su integración en sistemas de análisis y monitoreo. En total se generaron: 9 pares de modelos cuánticos ( $\alpha=0.05$  /  $\alpha=0.95$ ) y 9 clasificadores CatBoost balanceados. Esto posibilita su conexión con tableros interactivos en Google Sheets o Power BI, donde pueden visualizarse

predicciones, alertas y curvas de degradación por familia de medidor. Este conjunto constituye la base para un sistema modular de mantenimiento predictivo.

### ***Síntesis de la Fase 3***

La Fase 3 consolidó un sistema predictivo integral capaz de:

- Estimar desviaciones metrológicas futuras.
- Determinar la probabilidad de incumplimiento normativo.
- Simular la vida útil remanente de los medidores.
- Apoyar decisiones de calibración, mantenimiento y reemplazo

Los resultados obtenidos validan la hipótesis del proyecto y evidencian que los modelos de Machine Learning pueden capturar de manera precisa el desgaste metrológico y la degradación funcional de los medidores de agua potable.

### **Fase 4 Integración e Implementación**

La fase 4 tuvo como finalidad integrar los modelos desarrollados previamente dentro de un entorno funcional capaz de operar con datos reales, generar predicciones de manera automatizada y ofrecer una interfaz adecuada para la consulta, visualización y exportación de resultados. Esta integración permitió convertir el conjunto de modelos en un sistema operativo de apoyo a decisiones, fortaleciendo su aplicabilidad como herramienta de mantenimiento predictivo para medidores de agua potable. El desarrollo contempló el diseño de una arquitectura tecnológica, la incorporación de una capa de visualización interactiva y la automatización del flujo de datos entre Google Sheets, Streamlit y los modelos CatBoost entrenados en fases anteriores.

### ***Arquitectura General de Integración***

La arquitectura diseñada para esta fase se fundamentó en la interacción fluida entre tres capas funcionales conectadas en la nube.

En primer lugar, se estableció una capa de datos alojada en Google Sheets, utilizada como repositorio dinámico desde el cual se cargan los registros de medidores a evaluar. Esta base contiene variables operativas como Serial, Descripción, Edad, Volumen, Consumo\_anual, Qmínimo, Qtransición y Estado. La inclusión de la variable Serial fue determinante para asegurar la trazabilidad del sistema, permitiendo identificar cada equipo evaluado y facilitando la auditoría técnica de sus resultados.

En segundo lugar, se implementó la capa de procesamiento, encargada de ejecutar los modelos predictivos. Esta capa integra los archivos .cbm de los modelos de intervalos cuantílicos y de clasificación CatBoost, alojados dentro de la estructura del proyecto en Streamlit.

Finalmente, la capa de visualización se estructuró mediante Streamlit, que ofrece una interfaz de consulta intuitiva, gráficos dinámicos, semaforización de resultados y funciones para exportar automáticamente los reportes generados hacia un nuevo documento dentro de Google Sheets. Esta capa constituye el punto de interacción principal para el usuario final.

### ***Autenticación, Conexión y Flujo de Datos***

Para garantizar la integridad y seguridad del sistema, la comunicación entre Streamlit y Google Sheets se implementó mediante una cuenta de servicio de Google Cloud, cuyas credenciales se almacenaron en el archivo secrets.toml. Este mecanismo permite otorgar permisos específicos a la aplicación para leer, crear y modificar documentos de la hoja de cálculo vinculada, sin requerir autenticación interactiva de los usuarios.

El flujo de datos inicia con la lectura automática de la hoja “Base\_Medidores”, o alternativamente mediante la carga de un archivo CSV. Los datos se validan y se convierten al formato requerido por los modelos, aplicando transformaciones en caso de caracteres no numéricos o ausentes. Esta estandarización asegura que los modelos CatBoost reciban entradas limpias y estructuradas, coherentes con el proceso de entrenamiento realizado en la fase 3.

### ***Carga de Modelos Predictivos y Estructura***

Una vez establecida la conexión con la capa de datos, la aplicación carga de forma dinámica los modelos CatBoost desarrollados previamente. La Figura 10 muestra la interfaz inicial donde se selecciona la fuente de datos para iniciar el proceso. En la arquitectura del sistema, la carpeta modelos intervalos almacena los pares de archivos que representan los límites inferior y superior del intervalo predictivo del error EQ3, mientras que modelos\_clasificacion contiene los modelos destinados a estimar la probabilidad de cumplimiento normativo. La función cargar\_modelos() automatiza este procedimiento, detectando los modelos disponibles, organizando sus componentes y dejándolos preparados para su ejecución dentro de la aplicación.

### **Figura 10**

#### *Captura de la Interfaz Inicial de Streamlit*



### ***Proceso de Predicción y Generación de Indicadores***

El proceso de predicción se ejecuta de forma individual para cada registro del DataFrame cargado. Para cada medidor, el sistema:

- Identifica el modelo correspondiente mediante un mapeo entre la variable Descripción y un código interno normalizado.
- Ejecuta los modelos de intervalos para estimar los valores EQ3\_lower, EQ3\_upper y EQ3\_mid.
- Aplica el modelo de clasificación para calcular la probabilidad de cumplimiento normativo, generando un dictamen binario.
- Asigna el nivel de riesgo, basado en umbrales definidos (bajo, medio, alto).
- Calcula la vida útil remanente, utilizando las tasas de degradación por modelo.
- Genera un dictamen por intervalo, indicando si el error estimado se encuentra dentro o fuera del EMP.

Cada predicción se integra en un DataFrame consolidado que representa el diagnóstico completo de cada medidor, incluyendo tanto indicadores calculados (vida útil, probabilidad, error estimado) como elementos cualitativos (riesgo, dictamen, nivel de cumplimiento).

### ***Visualización Interactiva y SemafORIZACIÓN***

Con el fin de mejorar la interpretabilidad del sistema, Streamlit genera una visualización tabular enriquecida mediante un esquema de semafORIZACIÓN. La Figura 11 ilustra cómo los colores verde y rojo se utilizan para resaltar de manera automática los niveles de riesgo, el estado de cumplimiento normativo y la proximidad del error respecto al EMP. Este recurso visual facilita que el usuario identifique de forma inmediata los casos críticos o aquellos que se encuentran próximos a desviarse del comportamiento metrológico esperado.

## Figura 11

### Ejemplo de Resultados con Semaforización

#### Resultados detallados

	per (%)	EQ3_mid (%)	Probabilidad_Cumple	Pred_Conformidad	Nivel_de_Riesgo	Vida_remanente	Dictamen_por_intervalo
0	006445	-1.490876	0.358894	NO CUMPLE	ALTO	4.862659	DENTRO
1	207850	-1.758582	0.470591	NO CUMPLE	ALTO	3.755519	DENTRO
2	756175	-1.564787	0.361300	NO CUMPLE	ALTO	3.467212	DENTRO
3	006445	-1.490876	0.358894	NO CUMPLE	ALTO	4.862659	DENTRO
4	343530	-3.954423	0.433122	NO CUMPLE	ALTO	0.000000	FUERA
5	571139	-1.862044	0.344878	NO CUMPLE	ALTO	2.531851	DENTRO

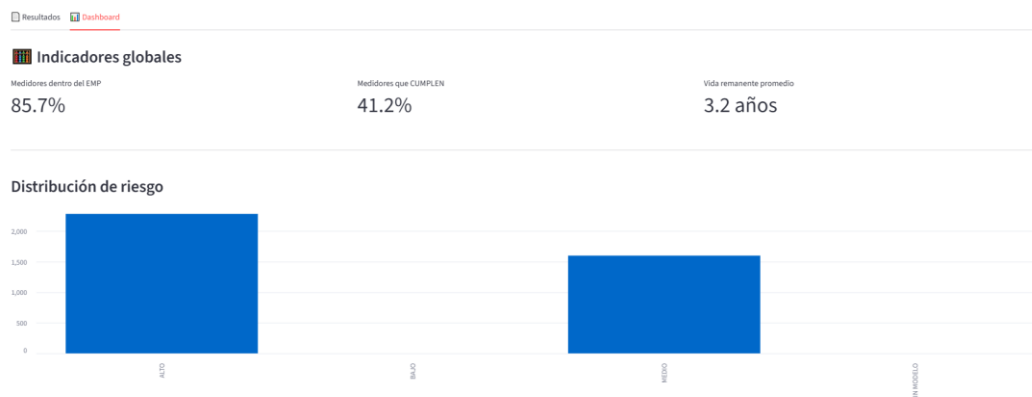
Además de la tabla detallada, la aplicación presenta un conjunto de paneles gráficos que permiten analizar tendencias globales del parque de medidores. Entre ellos se encuentran:

- Porcentaje de medidores dentro del EMP.
- Distribución de CUMPLE / NO CUMPLE.
- Distribución del nivel de riesgo.
- Vida útil promedio por modelo.
- Histograma de la variable EQ3\_mid.

Estos elementos permiten realizar análisis agregados que complementan el diagnóstico individual de cada medidor, así como la visualización de Bar charts con indicadores globales generados automáticamente, facilitando al usuario identificar de forma el comportamiento metrológico de la predicción realizada.

## Figura 12

### *Dashboard: Distribución de Riesgo y Conformidad*



### *Exportación Automática de Resultados a Google Sheets*

Uno de los aportes más relevantes de esta fase es la automatización del proceso de exportación. La aplicación crea una nueva hoja en el libro maestro de Google Sheets cada vez que se generan predicciones. El nombre de la hoja incorpora la fecha y hora exacta, lo que permite construir un historial cronológico sin sobrescribir información previa.

La exportación incluye:

- Encabezados organizados.
- Conversión de cada columna a texto para evitar errores de formato.
- Aplicación de colores y formatos condicionales (verde, amarillo, rojo, gris).
- Respeto del orden original de columnas.

Esto permite que el documento generado pueda servir directamente como reporte técnico, sin necesidad de edición manual adicional.

### ***Valor Aplicado y Relevancia Dentro del Proyecto***

La integración realizada consolida un prototipo funcional de mantenimiento predictivo digital, capaz de operar bajo un esquema completamente Cloud y replicable desde cualquier dispositivo. El sistema integra datos, modelos predictivos, análisis visual y reportes automatizados, lo que lo convierte en una herramienta útil para la gestión metrológica, la priorización del mantenimiento y la optimización del ciclo de vida de los medidores. Dado su diseño modular, la arquitectura puede ampliarse para incorporar nuevas fuentes de datos, integrar sensores IoT o conectarse con plataformas de gestión empresarial. Con ello, la fase 4 sienta las bases para una evolución futura hacia un sistema inteligente de monitoreo continuo sustentado en analítica avanzada.

### **Fase 5 Evaluación y Documentación**

La fase 5 se centró en evaluar rigurosamente la efectividad del sistema predictivo implementado y documentar de manera integral el proceso metodológico desarrollado en el proyecto. Esta etapa buscó determinar la capacidad real de generalización de los modelos al enfrentarse a datos nuevos y validar su pertinencia operativa en condiciones reales de laboratorio.

### ***Diseño del Proceso de Evaluación***

La evaluación del desempeño del sistema se realizó mediante un ejercicio de validación externa utilizando registros reales de calibración correspondientes a los meses de septiembre y octubre de 2025. Estos datos no formaron parte del entrenamiento ni validación inicial, por lo que permiten medir la capacidad del modelo para reproducir comportamientos auténticos del proceso metrológico bajo condiciones reales de operación.

El conjunto utilizado para la evaluación estuvo compuesto por aproximadamente 3.881 resultados de calibración, cada uno con información del error de medida en Q3 y el dictamen de conformidad final emitido por el laboratorio (Cumple / No cumple).

### ***Validación del Clasificador de Conformidad***

La evaluación del modelo de clasificación se fundamentó en comparar el dictamen predicho (Cumple / No cumple) con el dictamen oficial del laboratorio para cada medidor. La metodología aplicada fue: Asignar precisión del 100% cuando el dictamen predicho coincidió exactamente con el dictamen real. Asignar precisión del 0% cuando los dictámenes fueron diferentes.

**Tabla 10**

#### *Matriz de Confusión del Clasificador*

Real	Predicho	
	Cumple	No Cumple
Cumple	1336	978
No Cumple	233	1334

*Nota.* La Tabla 10 sintetiza el desempeño del modelo CatBoost al contrastar las predicciones con los dictámenes de conformidad obtenidos en laboratorio.

El clasificador alcanzó una precisión global del 69 %, resultado adecuado para una primera versión operativa del sistema, especialmente considerando la naturaleza binaria del dictamen, la heterogeneidad de los modelos de medidor y la variabilidad técnica asociada a los procesos de calibración.

### ***Validación del Modelo de Intervalos***

Para evaluar la precisión del modelo de intervalos, se formuló un criterio basado en la capacidad del sistema para predecir un rango de error Q3 razonable para cada medidor. El procedimiento consistió en: Comparar el error Q3 real obtenido en laboratorio con el intervalo predicho por el modelo (EQ3\_lower, EQ3\_upper). Asignar precisión del 100% cuando el error real se encontraba dentro del intervalo estimado, y calcular precisión proporcional cuando el error real caía por fuera del rango, otorgando valores entre 0% y 99% según la distancia relativa entre el valor real y el intervalo predicho.

Este enfoque permitió evaluar no solo si el intervalo era correcto, sino también qué tan cerca estuvo la predicción en los casos en los que no coincidió exactamente.

El resultado general indicó un desempeño de 67 %, evidenciando que los intervalos capturan razonablemente bien el comportamiento típico del error Q3, aunque presentan variabilidad según el tipo de medidor.

### ***Evaluación por Modelo Estándar de Medidor***

La precisión del sistema predictivo no es uniforme entre familias de medidores, lo cual es esperable dada la diversidad tecnológica y metrológica entre fabricantes, geometrías y clases de precisión. Esta variabilidad se refleja en los resultados obtenidos, los cuales se sintetizan en la Tabla 11, donde se presenta la validación correspondiente a cada modelo estándar de medidor.

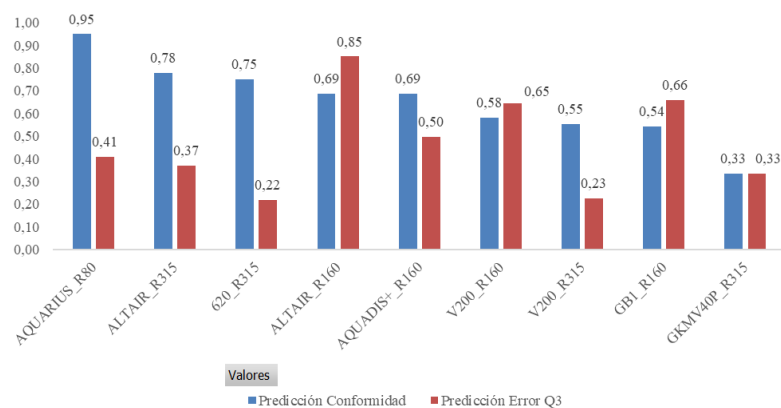
**Tabla 11***Validación por Modelo Estándar de Medidor*

Modelo Estándar	Precisión [Predicción Conformidad]	Precisión [Predicción Error Q3]	Muestra
M3	94,9%	40,7%	431
M8	77,8%	36,9%	284
M5	75,0%	21,7%	112
M2	68,7%	85,1%	1994
M7	68,6%	49,5%	245
M4	58,2%	64,5%	146
M6	55,4%	22,6%	92
M1	54,2%	65,9%	574
M9	33,3%	33,3%	3
Total	69,6%	67,4%	3881

*Nota.* Se evidencia estabilidad en función de la muestra tratada, el modelo de medidor M2 con mayor tasa de instalación presenta un buen desempeño general, con 68,7% en precisión (Predicción Conformidad) y 85,1% en Predicción Error Q3.

**Figura 13**

*Desempeño Promedio de los Modelos por Modelo Estándar de Medidor.*



*Nota.* Se evidencia estabilidad en función de la muestra tratada, el modelo de medidor ALTAIR\_R160 con mayor tasa de instalación presenta un buen desempeño general.

Estos resultados permiten observar: comportamiento sobresaliente de modelos como M3 y M8 en la clasificación, una excelente precisión en los intervalos para M2 (85,1%), el modelo con mayor peso estadístico en la muestra, y desempeños más modestos en tecnologías con menor representatividad en la muestra o con comportamiento metrológico intrínsecamente más variable. La Figura 14 complementa este análisis al mostrar que, en promedio, los modelos mantienen una estabilidad adecuada frente a las variaciones en la muestra tratada, destacándose nuevamente el comportamiento general del M2.

### ***Síntesis del Desempeño Global del Sistema***

**Tabla 12**

*Resumen de Precisión del Sistema Predictivo*

Modelo evaluado	Métrica evaluada	Precisión promedio
Regresores cuantílicos (intervalos)	Cobertura EQ3	67 %

Modelo evaluado	Métrica evaluada	Precisión promedio
Clasificador de conformidad CatBoost	Cumple / No cumple	70 %

### ***Interpretación Global y Análisis Crítico***

La validación externa demostró que el sistema predictivo posee una capacidad de generalización adecuada, considerando la complejidad del fenómeno analizado. Factores como: la incertidumbre de medición, el desgaste mecánico no lineal, las variaciones hidrodinámicas entre redes, el historial de operación desconocido para muchos medidores, y la heterogeneidad tecnológica del parque instalado, introducen variaciones que, por su naturaleza, no pueden ser capturadas completamente por ningún modelo de regresión o clasificación. Aun así, los resultados presentados en la Tabla 12 muestran que el sistema alcanza niveles de precisión estables y consistentes cuando se evalúa con datos reales externos, lo que confirma su aplicabilidad práctica en escenarios operativos.

## Conclusiones

Este proyecto demostró que es posible aplicar técnicas avanzadas de Machine Learning para modelar la degradación metrológica de los medidores de agua potable, estimar sus desviaciones y proyectar su vida útil funcional, contribuyendo a la construcción de una estrategia de mantenimiento predictivo robusta, escalable y fundamentada en datos reales.

La integración de datos históricos de calibración, información operativa y características técnicas permitió estructurar un conjunto de modelos complementarios, regresión cuantílica, clasificación supervisada y estimación analítica de vida útil, capaces de capturar distintos componentes del fenómeno metrológico. Los modelos desarrollados mostraron un desempeño estable tanto en los procesos de validación interna como en la validación externa con datos reales de laboratorio, alcanzando precisiones de 67 % en la estimación de intervalos de error y 70 % en el dictamen de conformidad, valores que resultan adecuados dadas las condiciones reales de operación y la alta variabilidad que caracteriza a los medidores mecánicos.

Adicionalmente, la implementación de una aplicación funcional en Streamlit, integrada con Google Sheets mediante autenticación segura, permitió transformar los modelos analíticos en una herramienta operativa de fácil uso, capaz de generar diagnósticos automáticos, semaforización por riesgo y reportes trazables. Esta aplicación constituye un avance significativo hacia la adopción de prácticas de mantenimiento predictivo en las empresas prestadoras de servicios públicos. En conjunto, los resultados obtenidos evidencian que la estrategia predictiva propuesta es técnicamente viable, operacionalmente útil y metodológicamente sólida. El sistema no reemplaza los procesos de calibración de laboratorio, pero sí permite anticipar comportamientos anómalos, priorizar intervenciones, reducir costos por reemplazos innecesarios y mejorar la precisión del sistema de medición.

## Recomendaciones

Las evidencias obtenidas a lo largo del desarrollo del sistema predictivo permiten plantear un conjunto de recomendaciones orientadas a fortalecer su implementación operativa, mejorar su desempeño analítico y consolidar su utilización dentro de la empresa de servicios públicos. Estas recomendaciones se organizan en tres niveles: técnico–analítico, operativo–procedimental y estratégico–organizacional.

En primer lugar, se recomienda mantener un proceso continuo de actualización del conjunto de datos utilizado para la predicción. Los modelos de Machine Learning, especialmente aquellos basados en degradación progresiva, mejoran considerablemente su rendimiento cuando incorporan datos recientes y representativos del comportamiento real del sistema. Por ello, se sugiere establecer una rutina trimestral o semestral para integrar nuevos resultados de calibración, consumos históricos y variables técnicas relevantes.

Asimismo, es aconsejable profundizar en el análisis por familia de medidor. Si bien el sistema muestra una precisión estable en su validación general, la naturaleza heterogénea de las tecnologías utilizadas (R80, R160, R315) implica que cada grupo puede requerir ajustes diferenciales en hiperparámetros, umbrales y criterios de clasificación. Una recalibración periódica de los modelos específicos permitirá aumentar la precisión de predicción para cada referencia.

## Referencias

- Arregui, F., Cobacho, R., & Soriano, J. (2006). Digitalisation and accuracy analysis of domestic water meters. *Flow Measurement and Instrumentation*, 17(5), 322–328.  
<https://doi.org/10.1016/j.flowmeasinst.2006.10.002>
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32.  
<https://doi.org/10.1023/A:1010933404324>
- De Paepe, M., & Luyben, K. (2018). Long-term accuracy and degradation of domestic water meters: A field study. *Water Resources Management*, 32(14), 4657–4671.  
<https://doi.org/10.1007/s11269-018-2079-3>
- Devera, O., & Tzatchkov, V. (2017). Effect of water quality, sediment and aging on the accuracy of residential water meters. *Procedia Engineering*, 186, 200–207.  
<https://doi.org/10.1016/j.proeng.2017.03.233>
- Escandón Beltrán, V., & Sierra García, B. (2018). *Parámetros metrológicos asociados en la calibración de un medidor de agua potable de clase metrológica R160 – 15 mm*. Universidad de Cartagena. <http://dx.doi.org/10.57799/11227/10203>
- International Organization for Standardization. (2014). *ISO 4064: Measurement of water flow in fully charged closed conduits—Water meters for cold potable water and hot water*.
- Kane, F. (2017). *Hands-On Data Science and Python Machine Learning*. Packt Publishing.
- Mendoza-Betín, J., Moncada-Baletta, S., Arias-Caseres, F., & Ramos-Pacheco, C. (2024). Análisis de subcontaje y sobredimensionamiento de contadores de agua: Ejercicio empírico de Laboratorio de Metrología de Aguas de Cartagena S.A. E.S.P. *Revista Anfibios*, 7(1), 50–71. <https://doi.org/10.37979/afb.2024v7n1.147>
- Mobley, R. K. (2002). *An introduction to predictive maintenance*. Butterworth-Heinemann.

Organisation Internationale de Métrologie Légale. (2013). *OIML R49: Water meters intended for the metering of cold potable water and hot water.*

Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V., & Gulin, A. (2018). CatBoost: Unbiased boosting with categorical features. In *Advances in Neural Information Processing Systems* (Vol. 31).

Raschka, S., & Mirjalili, V. (2017). *Python Machine Learning - Second Edition: Vol. 2nd ed.* Packt Publishing.