

**Sistema inteligente de monitoreo y control de calidad para los días de consumos facturables  
de los servicios públicos EGA en Antioquia**

Julián Andrés Henao Bustamante

Asesor

Andrea Ximena Robles Uriza

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

2025

### **Dedicatoria**

Este trabajo está dedicado a todas las personas que de forma directa e indirectamente me apoyaron con su paciencia y entendimiento aportando su granito de arena e instruyéndome en focalizar y formar las capacidades necesarias para poderlo realizar.

### **Agradecimientos**

Quiero expresar mi más sincero agradecimiento a todos los que ayudaron a terminar este trabajo: a mi directora de tesis Andrea Ximena Robles Uriza, por sus asesorías expertas y correcciones a tiempo. A mis compañeros, por su colaboración en los inicios del proyecto. A la UNAD, por ser mi institución mentora.

## Resumen

Este proyecto tiene como objetivo desarrollar un Sistema inteligente de monitoreo y control de calidad para los días de consumos facturables de los servicios públicos EGA en el departamento de Antioquia. El sistema integrará múltiples fuentes de datos, identificará inconsistencias y proporcionará información en tiempo real para garantizar una facturación precisa y la satisfacción del cliente. La metodología del proyecto implica la extracción, estandarización y análisis de datos, así como el desarrollo de algoritmos para detectar inconsistencias. La efectividad del sistema se evaluará a través de un estudio longitudinal y se desarrollará un mecanismo para garantizar el cumplimiento de los estándares regulatorios como lo establecido por la superintendencia de servicios públicos domiciliarios.

***Palabras claves:*** Base de datos, Control de calidad, Desempeño, Eficiencia, Indicadores, Optimización.

### **Abstract**

This project aims to develop a Quality Monitoring and Control System for the analysis of consumption and billing days in a public utility company. The system will integrate multiple data sources, identify inconsistencies, and provide real-time information to ensure accurate billing and customer satisfaction. The project's methodology involves data extraction, standardization, and analysis, as well as the development of algorithms to detect inconsistencies. The effectiveness of the system will be evaluated through a longitudinal study, and a mechanism will be developed to ensure compliance with regulatory standards as established by the Superintendence of Public Utilities.

***Keywords:*** Database, Quality control, Performance, Efficiency, Indicators, ptimization

## Tabla de Contenido

Introducción .....	12
Planteamiento del Problema .....	14
Justificación .....	15
Objetivos .....	16
Objetivo General .....	16
Objetivos Específicos.....	16
Marco Legal y Teórico.....	17
La Ley 142 de 1994 Regula los Servicios Públicos Domiciliarios en Colombia. ....	17
Normativa Sobre Lectura de Medidores. ....	17
Regulación Específica.....	17
Derechos de los Usuarios.....	18
La Gestión de Datos .....	18
Almacenamiento y Organización de Datos.....	18
Mejora de la Calidad de los Datos .....	18
Transformación de Datos .....	19
Metodología .....	20
Paso 1 Recopilación de Información a Partir de los Datos.....	20
Paso 2 Selección de Fuentes de Datos .....	20
Paso 3 Carga y Validación de Datos.....	20
Paso 4 Transformación y Limpieza de Datos .....	21
Paso 5 Identificación de Inconsistencias.....	21
Paso 6 Análisis y Segmentación .....	21

Paso 7 Creación de Variables Adicionales .....	22
Paso 8 Visualización y Validación .....	22
Paso 9 Modelo Predictivo .....	22
Paso 10 Interpretación de Resultados .....	23
Desarrollo de Alertas Tempranas .....	23
Provisión de Recomendaciones Específicas .....	23
Prevención de Inconsistencias y Mejora Continua .....	24
Clasificación, Limpieza y Análisis de los Datos Obtenidos .....	24
Limpieza de Datos .....	27
Parámetros y Métricas Descriptivas de los Datos .....	28
Ecosistema .....	32
Modelo de Datos .....	35
Modelo de Random Forest.....	36
Verdaderos Positivos (TP) .....	37
Verdaderos Negativos (TN).....	37
Falsos Positivos (FP) .....	37
Falsos Negativos (FN) .....	37
Sistema de Monitoreo y Control de Calidad para Empresas de Servicios Públicos en Antioquia .....	37
Resultados Obtenidos.....	39
Potencial a Futuro .....	41
Conclusiones .....	42
Recomendaciones .....	43

Glosario.....	44
Referencias Bibliográficas .....	47
Apéndices.....	47

**Lista de Tablas**

<b>Tabla 1</b> <i>Datos Obtenidos</i> .....	27
<b>Tabla 2</b> <i>Tamaño Nueva Bases de Datos</i> .....	30
<b>Tabla 3</b> <i>Evaluación del Modelo</i> .....	36

## Lista de Figuras

<b>Figura 1</b> <i>Flujo de la Extracción de Datos</i> .....	25
<b>Figura 2</b> <i>Tabla Relacional de la Fuente de Datos</i> .....	28
<b>Figura 3</b> <i>Verificar los Tipos de Datos</i> .....	29
<b>Figura 4</b> <i>Comportamiento Según Cada Distribución</i> .....	31
<b>Figura 5</b> <i>Diseño del Ecosistema del Modelo</i> .....	32
<b>Figura 6</b> <i>Comparación del Consumo de Energía Leído vs Consumo Real según la Categoría.</i>	33
<b>Figura 7</b> <i>Inconsistencias dentro de Cada una de las Regiones en ANTIOQUIA</i> .....	34
<b>Figura 8</b> <i>Diagrama de Flujo Modelo</i> .....	35
<b>Figura 9</b> <i>Matriz de Confusión</i> .....	36
<b>Figura 10</b> <i>Enfoque Geografico del Resultado del Modelo</i> .....	38
<b>Figura 11</b> <i>Piramide de Resultados Obtenidos</i> .....	39
<b>Figura 12</b> <i>Potencial a Futuro</i> .....	41

**Lista de Apendices**

<b>Apéndice A</b> <i>Exploración de los Datos</i> .....	50
<b>Apéndice B</b> <i>Caracterización y Clasificación</i> .....	52
<b>Apéndice C</b> <i>Video de Socialización del Trabajo de Grado</i> .....	53

## Introducción

En el mundo actual, donde la confianza y la transparencia son pilares fundamentales en la relación entre las empresas y sus clientes, la gestión de la información se convierte en un aspecto crítico para el éxito de cualquier organización. En el contexto de una empresa de servicios públicos, esta realidad se manifiesta de manera palpable a través de la experiencia de sus usuarios. Muchos de ellos, al recibir sus facturas, se ven confrontados con errores e inconsistencias en los días que fueron tomados los registros de sus medidores y que generan no solo dudas sobre la veracidad de los cobros, sino también una frustración que puede erosionar la confianza en la compañía. Esta situación no es simplemente un inconveniente; es un desafío que afecta la satisfacción del cliente y, en última instancia, la reputación de la empresa.

La complejidad de gestionar el rango de días de consumo y facturación de múltiples productos y servicios, con diferentes periodicidades y fuentes de información, hace que la tarea de garantizar la precisión y coherencia de estos datos sea monumental. Sin un sistema robusto que supervise y controle la calidad de la información, la empresa se enfrenta a una serie de consecuencias negativas: facturación inexacta, quejas recurrentes de los clientes y un daño significativo a su credibilidad. En este sentido, la falta de un enfoque proactivo para identificar y abordar inconsistencias se asemeja a buscar una aguja en un pajar, donde la ineficiencia en la resolución de problemas no solo afecta la satisfacción del cliente, sino que también limita la capacidad de la empresa para tomar decisiones informadas basadas en datos.

Por lo tanto, es imperativo que la empresa desarrolle un sistema confiable y eficiente que permita monitorear y controlar la calidad del rango de los días de consumo y facturación. Este sistema debe ser capaz de integrar diversas fuentes de información, identificar inconsistencias de manera oportuna y proporcionar información en tiempo real. Al implementar una solución de

este tipo, la empresa no solo podrá recuperar la confianza de sus clientes, sino también reducir las pérdidas de ingresos y mejorar su reputación general en el mercado. Este proyecto de grado se propone abordar esta necesidad crítica, ofreciendo un enfoque innovador que prioriza la satisfacción del cliente y la integridad de la información enfocados a los días que se tomó la lectura del producto o medidor.

## Planteamiento del Problema

Se ha identificado productos de usuarios registrados en la empresa de servicios públicos, que al recibir su factura han encontrado errores o inconsistencias que hacen dudar de la veracidad de los cobros y muchos clientes se enfrentan a esta frustración, lo que genera desconfianza e insatisfacción con la compañía por ende la empresa se esfuerza por gestionar y hacer un seguimiento de los datos de consumo y facturación de múltiples productos y servicios. Con distintas periodicidades (mensuales, bimensuales, trimestrales) y diversas fuentes de datos, para garantizar la precisión y la coherencia en la información ya que es una tarea compleja. La falta de un sistema sólido para supervisar y controlar la calidad conduce a;

- Facturación inexacta.
- Quejas e insatisfacción de los clientes.
- Daño a la reputación y credibilidad de la empresa.

Es como intentar encontrar una aguja en un pajar. Es difícil identificar y abordar las inconsistencias de manera oportuna, lo que lleva a un enfoque reactivo para la resolución de problemas. Esto no solo afecta la satisfacción del cliente, sino que también obstaculiza la capacidad de la empresa para tomar decisiones basadas en datos por lo cual la empresa necesita un sistema confiable y eficiente para monitorear y controlar la calidad de los días de consumo y facturación. Este sistema debe ser capaz de integrar múltiples fuentes de datos, identificar inconsistencias y brindar información en tiempo real para garantizar una facturación precisa y la satisfacción del cliente. Al desarrollar un sistema de este tipo, la empresa puede recuperar la confianza de los clientes, reducir las pérdidas de ingresos y mejorar su reputación general.

## Justificación

Este proyecto es relevante y necesario porque aborda un problema crítico en la industria de servicios públicos: la facturación incorrecta y las pérdidas de ingresos debido a inconsistencias en los datos de consumo, facturación y los días que se visitan en campo ya que según [Ley 142 de 1994 - Gestor Normativo. (n.d.). Gov.co.], los usuarios tienen derecho a obtener de las empresas la medición de sus consumos reales mediante instrumentos tecnológicos apropiados, dentro de plazos y términos que para los efectos fije la comisión reguladora, con atención a la capacidad técnica y financiera de las empresas o las categorías de los municipios establecida por la ley.

La falta de un sistema sólido para monitorear y controlar la calidad genera insatisfacción de los clientes y según [de Energía y Gas-CREG, C. de R. (n.d.)], el usuario tienen derecho de conocer su promedio de consumo, en unidades correspondientes, del servicio de los últimos seis (6) meses, y en ocasiones no se hace correctamente ocasionando un daño a la reputación de la empresa e incumplimiento de las normas regulatorias.

Mediante el desarrollo de un Sistema de Control y Seguimiento de la Calidad, este proyecto pretende mejorar la satisfacción del cliente, reducir las pérdidas de ingresos y mejorar la reputación de la empresa. La relevancia académica del proyecto radica en su contribución al campo del análisis de datos y el control de calidad, mientras que su relevancia social radica en su potencial para mejorar la vida de los clientes y la eficiencia general de la industria de los servicios públicos.

## **Objetivos**

### **Objetivo General**

Crear un sistema para monitorear y controlar la calidad de los días de consumo y facturación en una empresa de servicios públicos en Antioquia, garantizando el cumplimiento de los estándares regulatorios y reduciendo inconsistencias.

### **Objetivos Específicos**

Desarrollar un sistema de monitoreo y control de calidad, con varias integraciones a bases de datos para analizar e identificar con mejor precisión y eficiencia los días de consumos en la facturación de acuerdo con la visita en campo en Antioquia.

Identificar y analizar las inconsistencias, por medio de algoritmos de Python en la detección de anomalías.

Recopilar y analizar datos para identificar áreas de mejora y oportunidades dentro de la compañía.

## **Marco Legal y Teórico**

El desarrollo de un Sistema de Monitoreo y Control de Calidad para analizar los rangos de días y consumo diario para determinar la cantidad a facturar en una empresa de servicios públicos es un proceso complejo. Esta tarea demanda un sólido entendimiento del marco regulatorio aplicable, así como habilidades en la gestión de datos y en la implementación de principios de control de calidad.

### **La Ley 142 de 1994 Regula los Servicios Públicos Domiciliarios en Colombia.**

Estableciendo que la periodicidad de lectura de medidores debe ser adecuada para garantizar una facturación precisa. Generalmente, se espera que la lectura se realice mensualmente, aunque puede variar según el tipo de servicio y las disposiciones específicas de cada prestador. [Ley 142 de 1994 - Gestor Normativo. (n.d.). Gov.co.].

### **Normativa Sobre Lectura de Medidores.**

La Ley 142 de 1994 establece que los prestadores de servicios públicos deben realizar lecturas de medidores de manera regular, generalmente cada mes, para asegurar que la facturación refleje el consumo real.

La periodicidad puede variar dependiendo del tipo de servicio (agua, electricidad, gas) y de las condiciones específicas acordadas en el contrato de prestación del servicio.

### **Regulación Específica**

La Comisión de Regulación de Agua Potable y Saneamiento Básico (CRA) y otras entidades reguladoras pueden emitir resoluciones que especifiquen la frecuencia de las lecturas y los procedimientos a seguir.

Es importante que los usuarios estén informados sobre la periodicidad de las lecturas y cualquier cambio que pueda ocurrir en la normativa o en las políticas de facturación de su prestador de servicios[de Energía y Gas-CREG, C. de R. (n.d.)].

### **Derechos de los Usuarios**

Los usuarios tienen derecho a ser informados sobre cómo se determinan y valoran sus consumos, así como a recibir facturas que reflejen lecturas precisas y oportunas.

En caso de que no se realicen lecturas en el periodo estipulado, los usuarios pueden solicitar aclaraciones y ajustes a su facturación.

### **La Gestión de Datos**

La gestión de datos implica la recopilación, almacenamiento, organización, mantenimiento y uso de datos de manera eficiente y segura. Es fundamental para garantizar que la información sea precisa, coherente y accesible cuando se necesite.

### **Almacenamiento y Organización de Datos**

El almacenamiento adecuado de datos es esencial para su gestión eficaz. Esto incluye la utilización de bases de datos relacionales y no relacionales, sistemas de almacenamiento en la nube y tecnologías de big data. La organización de datos implica clasificarlos y estructurarlos de manera que sean fácilmente accesibles y utilizables.

### **Mejora de la Calidad de los Datos**

La calidad de los datos se refiere a su precisión, coherencia, integridad y relevancia. Las técnicas de mejora de la calidad incluyen la limpieza de datos, la validación y la normalización. La limpieza de datos elimina errores y duplicados, mientras que la validación asegura que los datos sean correctos y completos.

## **Transformación de Datos**

La transformación de datos implica convertirlos de un formato a otro para que sean más útiles y accesibles. Esto puede incluir la agregación de datos, la creación de nuevas variables y la aplicación de algoritmos de machine learning para extraer información valiosa.

## **Metodología**

### **Paso 1 Recopilación de Información a Partir de los Datos**

La recopilación de información en este proyecto fue un proceso meticuloso y cuidadosamente diseñado para garantizar la precisión y relevancia de los datos utilizados. Este proceso comenzó con la identificación de las fuentes de datos más pertinentes y la selección de las variables clave que serían fundamentales para el análisis.

### **Paso 2 Selección de Fuentes de Datos**

Se eligieron bases de datos que contenían información detallada sobre el consumo y la facturación de servicios públicos en Antioquia. Estas bases de datos fueron integradas utilizando consultas SQL para extraer las variables necesarias, asegurando que cada dato recopilado tuviera un propósito claro y específico en el contexto del proyecto.

### **Paso 3 Carga y Validación de Datos**

Los registros fueron cargados en un archivo QVD, un formato de almacenamiento de datos de la aplicación Qlik Sense, que permite manejar grandes volúmenes de información de manera eficiente. Se procedió a la carga de datos correspondientes a cuatro meses, lo que permitió tener una visión completa de las periodicidades y variaciones en los días del consumo.

Se utilizó un análisis estadístico descriptivo para comprender la distribución de las variables. Utilizando dos clases de variables (categóricas y numéricas), se caracterizaron los datos obtenidos. Se identificaron las variables que pueden usarse para análisis posteriores y la selección de un modelo predictivo o clasificación de algoritmos supervisados[Qvd. (s/f)].

#### **Paso 4 Transformación y Limpieza de Datos**

Para limpiar adecuadamente los datos, se utilizó Jupyter Notebook y el lenguaje de programación Python. Se codificaron los datos en el DataFrame producido en Jupyter Notebook de acuerdo con la siguiente secuencia: nulo, vacío, repetido, sin formato y reemplazable.

Para garantizar la calidad de los datos, se realizaron varias transformaciones. Las columnas de fechas fueron convertidas al formato adecuado, y se crearon nuevas variables como "DIAS\_TRANSURRIDOS", que representa la diferencia entre la fecha de ejecución final y la fecha de lectura anterior. Además, se llenaron los valores nulos y se ajustaron los tipos de datos para asegurar la consistencia y precisión.

#### **Paso 5 Identificación de Inconsistencias**

Se establecieron condiciones específicas para identificar productos que cumplieran con las normativas establecidas para cada periodicidad de consumo. Los productos que no cumplieran con estas condiciones fueron marcados como "No Cumple", lo que permitió una clasificación clara y precisa de los datos.

#### **Paso 6 Análisis y Segmentación**

Se implementaron modelos de Machine Learning, incluyendo Regresión Lineal, Regresión Logística y Random Forest, utilizando un 80% en datos de entrenamiento y 20% en datos de prueba, para predecir las inconsistencias en los días facturados y factores clasificatorios que podrían afectar la calidad de los días de consumo.

Los datos fueron analizados para identificar áreas de mejora y oportunidades dentro de la compañía. Se crearon variables adicionales como "Etiqueta" y "PRIORIDAD", que ayudaron a segmentar los datos y a identificar patrones y anomalías en el consumo.

### **Paso 7 Creación de Variables Adicionales**

Se crearon varias variables adicionales para enriquecer el análisis, como ANOMALIA\_DIAS, que calcula la diferencia en días entre el consumo real y el máximo permitido para cada periodicidad. También se crearon las variables SERVICIO, CATEGORIA, y ESTRATO para segmentar los datos de acuerdo con diferentes criterios.

### **Paso 8 Visualización y Validación**

Se utilizaron gráficos y mapas para visualizar la distribución de los datos y las inconsistencias. Estas visualizaciones permitieron una comprensión más profunda de los datos y facilitaron la identificación de áreas críticas que requerían atención.

### **Paso 9 Modelo Predictivo**

Finalmente, se desarrolló un modelo predictivo utilizando algoritmos de aprendizaje automático para anticipar posibles inconsistencias en el futuro. Este modelo fue entrenado y evaluado, mostrando un rendimiento excepcional en la clasificación de productos con anomalías.

En donde se definió como variable objetivo en este modelo la inconsistencia en los días de consumo. Se creó una nueva columna llamada INCONSISTENCIA que marcaba los casos donde la prioridad era alta y los días de anomalía eran mayores o iguales a uno, se seleccionaron varias características relevantes para el modelo y se realizó la codificación de variables categóricas estas fueron codificadas utilizando la técnica de One-Hot Encoding (una técnica para convertir variables categóricas en formato numérico binario. Crea columnas para cada categoría de la variable original, asignando un valor de 1 a la columna correspondiente si la categoría está presente en la observación, y 0 en caso contrario), para convertirlas en un formato adecuado para el modelo de machine learning, los datos fueron divididos en conjuntos de entrenamiento y prueba utilizando la función `train_test_split` de la biblioteca `sklearn`. Se utilizó un 80% de los

datos para el entrenamiento y un 20% para la prueba. Se entrenó un modelo de Random Forest con 80 árboles ( $n\_estimators=80$ ) y una semilla aleatoria ( $random\_state=42$ ) para asegurar la reproducibilidad de los resultados.

El modelo fue evaluado utilizando métricas de clasificación como la precisión, el recall y el F1-score. Los resultados mostraron un rendimiento excepcional, con una precisión y recall de 1.00 para ambas clases (0 y 1), indicando que el modelo no cometió errores en sus predicciones. Finalmente, se utilizaron las predicciones del modelo para filtrar los productos que cumplían con las condiciones de inconsistencia. Estos productos fueron identificados y se les asignaron coordenadas para su análisis geográfico [Emersion Systems. (2024, November 25)].

### **Paso 10 Interpretación de Resultados**

Una vez realizado el análisis de las variables que afectan la calidad en los días de consumo y la facturación en la empresa de servicios públicos, utilizando el mejor modelo predictivo y clasificador de machine learning, se procede a interpretar los resultados obtenidos por el algoritmo seleccionado. Este análisis nos lleva a cumplir los tres últimos objetivos específicos del proyecto de grado como ;

#### ***Desarrollo de Alertas Tempranas***

A partir de los resultados obtenidos, se desarrollan alertas tempranas para identificar y prevenir inconsistencias en rango de días de consumo y la facturación. Estas alertas permiten a la empresa actuar de manera proactiva, evitando problemas antes de que se conviertan en situaciones críticas.

#### ***Provisión de Recomendaciones Específicas***

Se proporcionan recomendaciones específicas basadas en los datos analizados para mejorar la precisión y eficiencia en los días facturados. Estas recomendaciones están diseñadas

para mitigar las inconsistencias y asegurar que los estándares regulatorios se cumplan de manera efectiva.

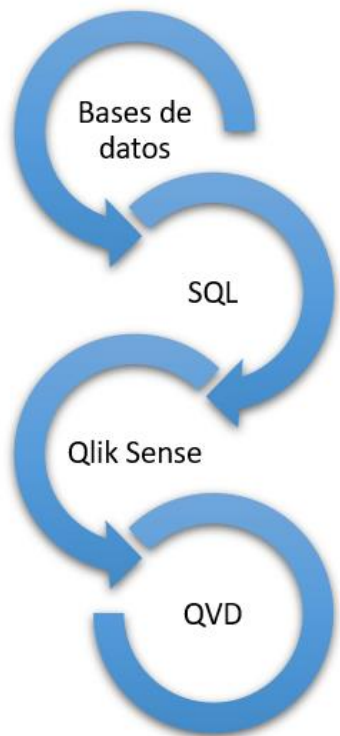
### ***Prevención de Inconsistencias y Mejora Continua***

Finalmente, se implementan estrategias para evitar la recurrencia de inconsistencias en el futuro. Estas estrategias incluyen la mejora continua de los procesos de monitoreo y control de calidad, así como la capacitación del personal para manejar y resolver problemas de manera eficiente.

### ***Clasificación, Limpieza y Análisis de los Datos Obtenidos***

Es importante destacar que la siguiente hoja de datos se generó a partir de un modelo relacional, utilizando una consulta SQL para extraer la información más relevante y las variables clave necesarias para el proyecto a continuación, se comparte la estructura de datos:

Luego de la estructura de datos se realiza un mapa de los flujos que involucran el proceso de los tiempos o días de consumo en una lectura.

**Figura 1***Flujo de la Extracción de Datos*

Se diseñó un sistema que integre múltiples fuentes de datos presentando como una necesidad imperante. Este sistema no solo busca la recopilación de datos, sino que también se enfoca en la estandarización de la información proveniente de diversas tablas del facturador de la empresa. La estandarización es un proceso crítico que garantiza que los datos sean consistentes, precisos y comparables, lo que a su vez facilita la toma de decisiones informadas y la implementación efectiva de los objetivos propuestos en el presente proyecto.

Identificador del producto 'PRODUCT\_ID'

Numero de orden de la atención 'ORDER\_ID'

Numero del servicio 'SESUSERV'

Categoría del servicio 'SESUCATE'

Sub categoría del servicio 'SESUSUCA'

Ciclo de facturación 'CICLO'

Tarifa 'SESUPLFA'

Fecha que se cerro la orden 'LEGALIZATION\_DATE'

Fecha que se asigno la orden 'ASSIGNED\_DATE'

Fecha que se atendio laordne en campo 'EXECUTION\_FINAL\_DATE'

Perfil del fraude 'PERFIL\_FRAUDES'

Estado de la cuenta sobre el producto 'ESTADO\_CUENTA'

Estado de la Mora sobre el producto 'ESTADO\_MORA'

Cantidad de kwh ó  $m^3$  leidos 'CONSUMO\_LEIDO'

Codigo de causa de no lectura 'CAUSANL'

Descripción de causa de no lectura 'DESC\_CAUSANL'

Codigo observación adicional 'OBS\_ADIC'

Descripción de la observación adicional 'DESC\_OBS\_ADIC'

Frecuencia en la que se va tomar la lectura 'PERIODICIDAD'

Fecha que se atendio la orden el mes anterior 'F\_LECT\_ANTERIOR'

Descripción de causa de no lectura del mes anterior 'DESC\_CAUSANL\_ANTERIOR'

Descripción de la observación adicional del mes anterior  
'DESC\_OBS\_ADIC\_ANTERIOR'

Descripción de la novedad mes anterior 'NOVED\_LECT\_ANT'

Ruta del recorrido 'ROUTE\_ID'

Número de días recorridos de lectura 'DIAS\_TRANSCURRIDOS'

Descripción de la novedad 'Etiqueta'

Número de días 'ANOMALIA\_DIAS'

Descripción del servicio 'SERVICIO'

Descripción de las categoría 'CATEGORIA'

Número de estrato 'ESTRATO',

Cantidad de kwh ó  $m^3$  leídos que se deben cobrar 'CONSUMO\_REAL'

Diferencia de consumos 'DIFERENCIA\_CONSUMO'

Prioridad o importancia para ser atendida 'PRIORIDAD'

### **Tabla 1**

#### *Datos Obtenidos*

Data Base	Productos leídos 2024-2025
Columnas	27
Registros	3.116.069

*Nota.* Datos Obtenidos.

### **Limpieza de Datos**

El primer paso en este proceso fue la identificación de las diferentes variables que influyen en los puntos clave de la implementación. Estas variables incluyen, entre otras, datos de clientes, productos y tiempos de facturación. Al comprender cómo cada una de estas variables interactúa y contribuye a los resultados generales, logramos establecer un marco de referencia que nos permitió abordar los desafíos específicos que enfrenta la organización en su proceso de los rangos en los días facturables del consumo, realizando un filtro.

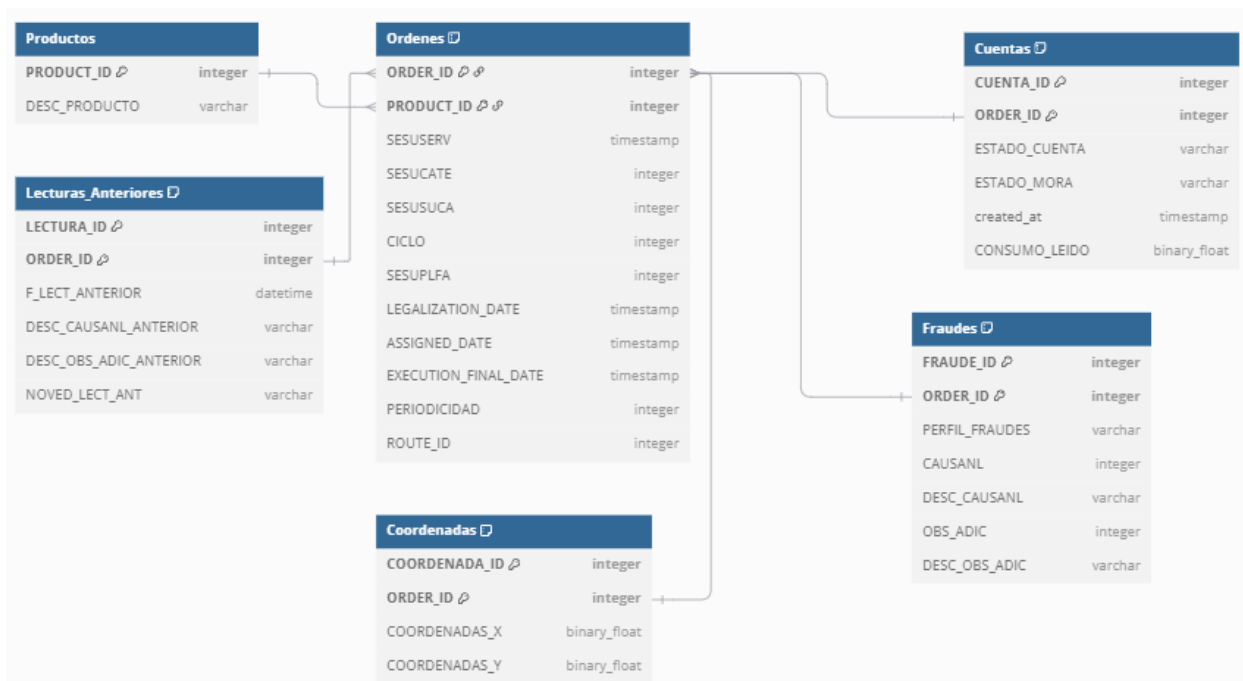
## Parámetros y Métricas Descriptivas de los Datos

Se realizó un análisis descriptivo de las variables numéricas y categóricas para entender su comportamiento. Se seleccionaron las variables principales para utilizar en los algoritmos supervisados, como el consumo leído, días transcurridos y periodicidad.

Una vez que se identificó las variables relevantes, el siguiente paso fue la creación de un sistema que permitió la integración de estas fuentes de datos. Este sistema fue capaz de extraer información de las bases de datos, sistemas de gestión y aplicaciones utilizadas por la empresa. La integración de los datos no solo implicó la recopilación de la información, sino también la transformación de estos datos en un formato que fue útil y accesible. Para lograr esto, se implementaron procesos de extracción, transformación y carga (ETL) que asegurarán que los datos fueran procesados de manera eficiente y efectiva.

### Figura 2

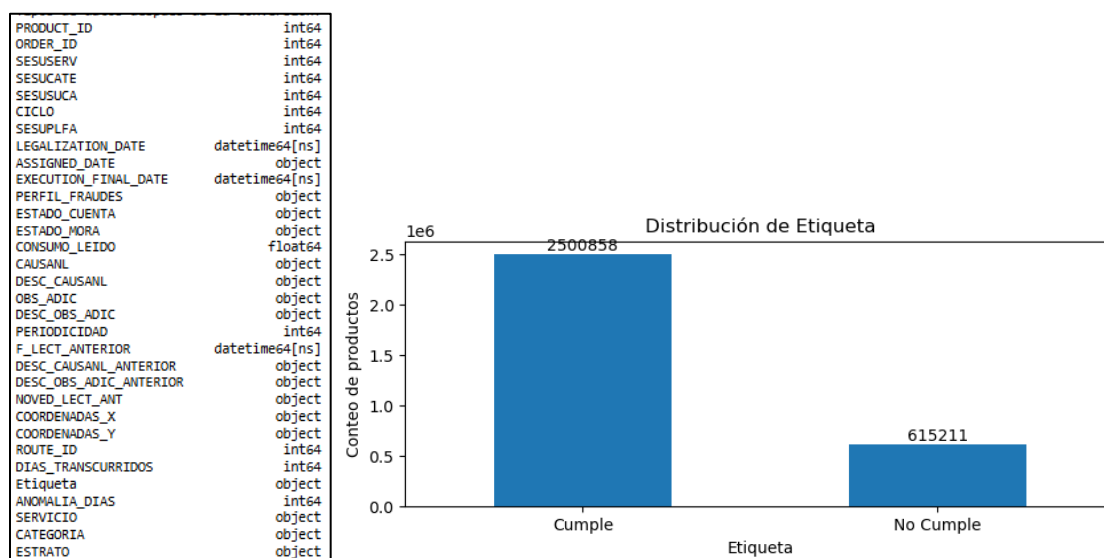
Tabla Relacional de la Fuente de Datos



La limpieza de datos fue otro componente esencial en este proceso. A menudo, los datos recopilados tenían errores, duplicados o información irrelevante que lograba afectar la calidad de los análisis posteriormente. Por lo tanto, se establecieron procedimientos rigurosos de limpieza de datos que incluyeron la identificación y eliminación de duplicados, la corrección de errores geográficos y la validación de la información. Este proceso no solo mejoró la calidad de los datos, sino que también aumentó la confianza.

### Figura 3

*Verificar los Tipos de Datos*



El gobierno de datos fue un aspecto fundamental que no debe ser pasado por alto. Un marco de gobierno de datos bien definido aseguró que la información sea gestionada de manera adecuada, cumpliendo con las normativas y políticas establecidas por la organización. Esto incluyó la definición y responsabilidades, así como la implementación de controles de acceso y seguridad para proteger la integridad de los datos. Además, se establecieron políticas de

retención de datos para que de forma segura la información se almacene y se elimine de acuerdo con las regulaciones aplicables.

**Tabla 2**

*Tamaño Nueva Bases de Datos*

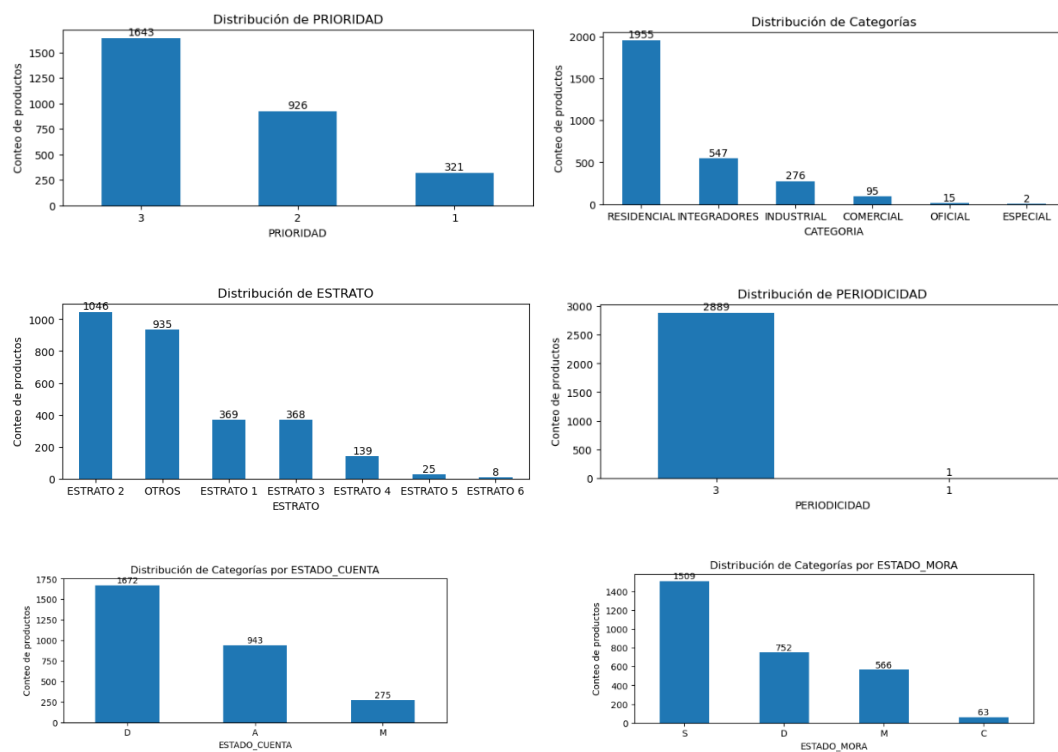
Data Base	$df\_filtrado = df[(df['Etiqueta'] == 'No Cumple') \& (df['NOVED\_LECT\_ANT'] == 'SIN NOVEDAD') \& (df['ANOMALIA\_DIAS'] >= 1)]$
Registros iniciales	3.116.069
Registros finales	2.890

*Nota.*Nueva bases de datos

Una vez que los datos han sido estandarizados, limpiados y gobernados, se logró aplicar una serie de filtros que permitieron estructurar la información de manera fácil, accesible y comprensible. Estos filtros incluyeron la segmentación de datos por categorías, la creación de dashboards interactivos y la generación de informes personalizados que facilitaron la visualización de la información. La capacidad de filtrar y estructurar los datos de esta manera no solo mejoro la eficiencia operativa, sino que también permitió a los tomadores de decisiones identificar tendencias, patrones y oportunidades de mejora en el proceso teniendo en cuenta los días de consumo.

## Figura 4

### Comportamiento Según Cada Distribución



Estado de la cuenta: A-Al día; D-Con deuda; C- Castigado; M-En mora.

Estado de mora: S-Sospecha de mora; M-En mora; C-Mora cobrada; D-Al día.

## Ecosistema

El ecosistema de la información inicia desde una base de datos en donde se extrae la información por medio de una consulta parcialmente construida y verificada desde la aplicación Qlik Sense que tiene un motor de bases de datos muy robusto y logra almacenar en archivos tipo QVDs, donde la información es más liviana, luego desde Jupyter Notebook se corre todo el modelo por medio de lenguaje de Python en donde se logra integrar la extracción de los datos con el análisis de la información el resultado de el modelo, se puede ver el antes y el después de forma gráfica y finalmente se puede tener la información en archivos de Excel para poder compartir los datos al ente encargado del seguimiento y programación.

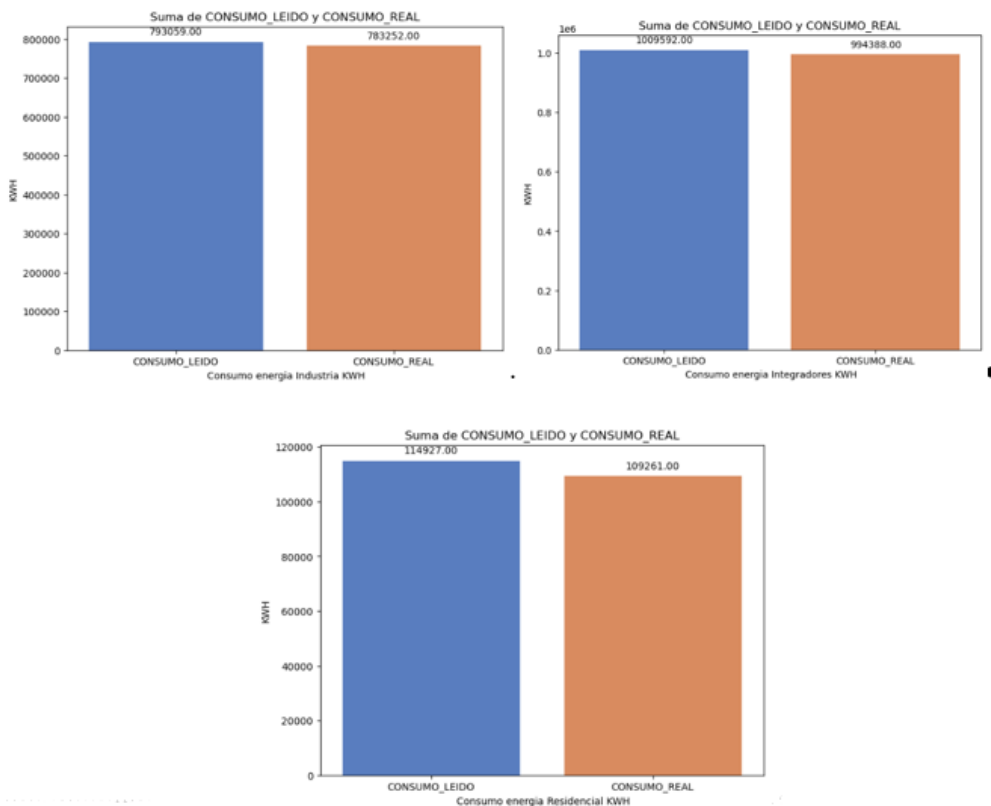
### Figura 5

*Diseño del Ecosistema del Modelo*



**Figura 6**

*Comparación del Consumo de Energía Leído vs Consumo Real según la Categoría*

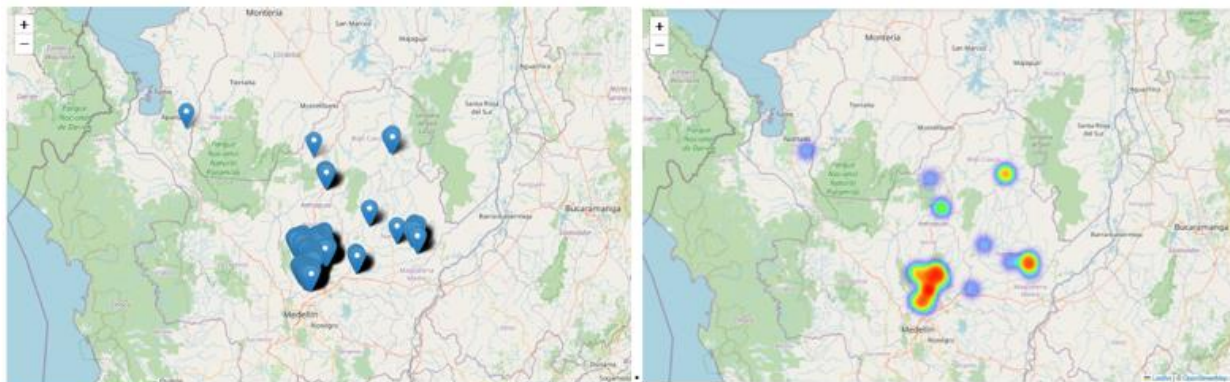


Además, la implementación de este sistema integrado de gestión de datos para el análisis de días de consumo no solo beneficiará al departamento de facturación, sino que también tendrá un impacto positivo en otras áreas de la organización tanto al proceso de lectura de los medidores como también el mejoramiento y reducción continua de PQR por consumos incorrectos.

Implementar herramientas de visualización de datos que faciliten la interpretación de los consumos y facturación, por medio de capas geográficas es algo indispensable como geodatabases por eso se implemento dentro del proyecto un mapa geografico para identificar esas zonas que tiene dicha afectación .

**Figura 7**

*Inconsistencias dentro de Cada una de las Regiones en ANTIOQUIA*



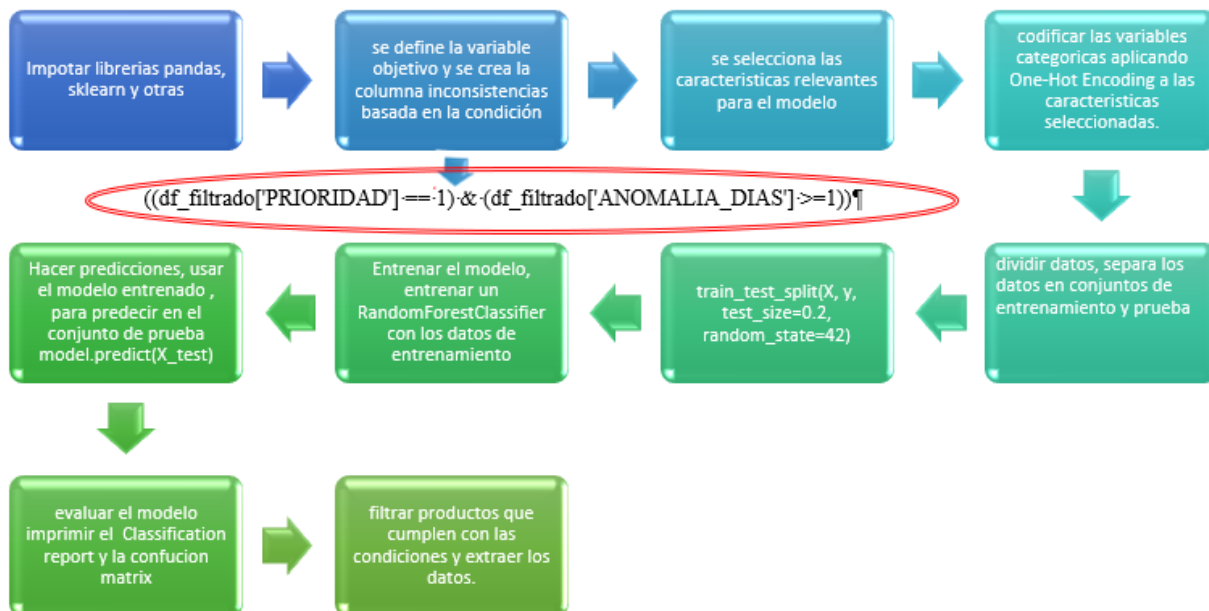
Establecer protocolos de actualización y mantenimiento del sistema para asegurar su eficacia a largo plazo.

## Modelo de Datos

Se establece la creación de un modelo predictivo para poder tener segmentado posibles inconsistencias a futuro y poderlas corregir a tiempo ya que la variable objetivo se tiene contemplado validar los productos que van a tener anomalías en los rangos de los días del consumo identificando la prioridad que sea igual a uno y que tengan los días de anomalías mayores a cero (0).

**Figura 8**

*Diagrama de Flujo Modelo*



### Modelo de Random Forest

El modelo de Random Forest se entrenó con las características seleccionadas y mostró un rendimiento perfecto en la clasificación de inconsistencias. La precisión, recall y F1-score fueron 1.00 para ambas clases (0 y 1), indicando que el modelo no cometió errores en sus predicciones.

**Tabla 3**

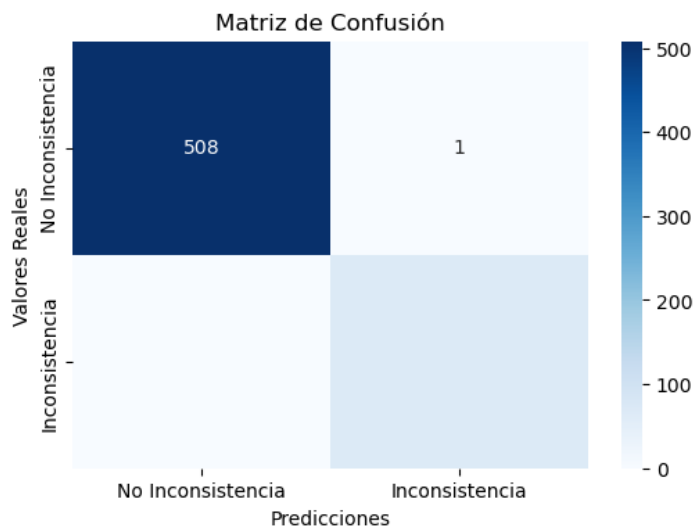
*Evaluación del Modelo*

Data Base	Precisión	recall	f1-score	support
0	1.00	1.00	1.00	509
1	0.99	1.00	0.99	69
accuracy			1.00	578
macro avg	0.99	1.00	1.00	578
weighted avg	1.00	1.00	1.00	578

*Nota.* Evaluación del modelo.

**Figura 9**

*Matriz de Confusión*



**Verdaderos Positivos (TP)**

Casos en los que el modelo predijo correctamente la clase positiva de productos que no tienen inconsistencias.

**Verdaderos Negativos (TN)**

Casos en los que el modelo predijo correctamente la clase negativa de los productos que tienen inconsistencias.

**Falsos Positivos (FP)**

Casos en los que el modelo predijo la clase positiva incorrectamente por debajo del promedio, pero son predicciones falsas de productos que no tienen inconsistencias.

**Falsos Negativos (FN)**

Casos en los que el modelo predijo la clase negativa incorrectamente de los productos que tienen inconsistencias.

**Sistema de Monitoreo y Control de Calidad para Empresas de Servicios Públicos en Antioquia**

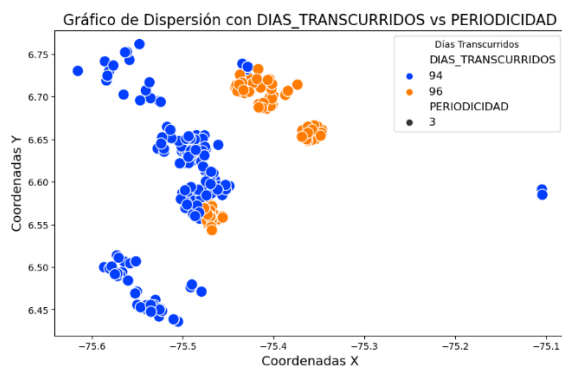
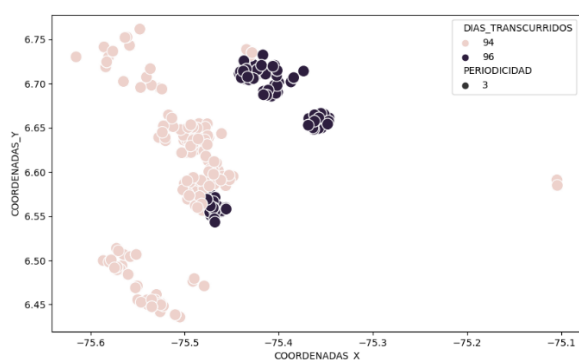
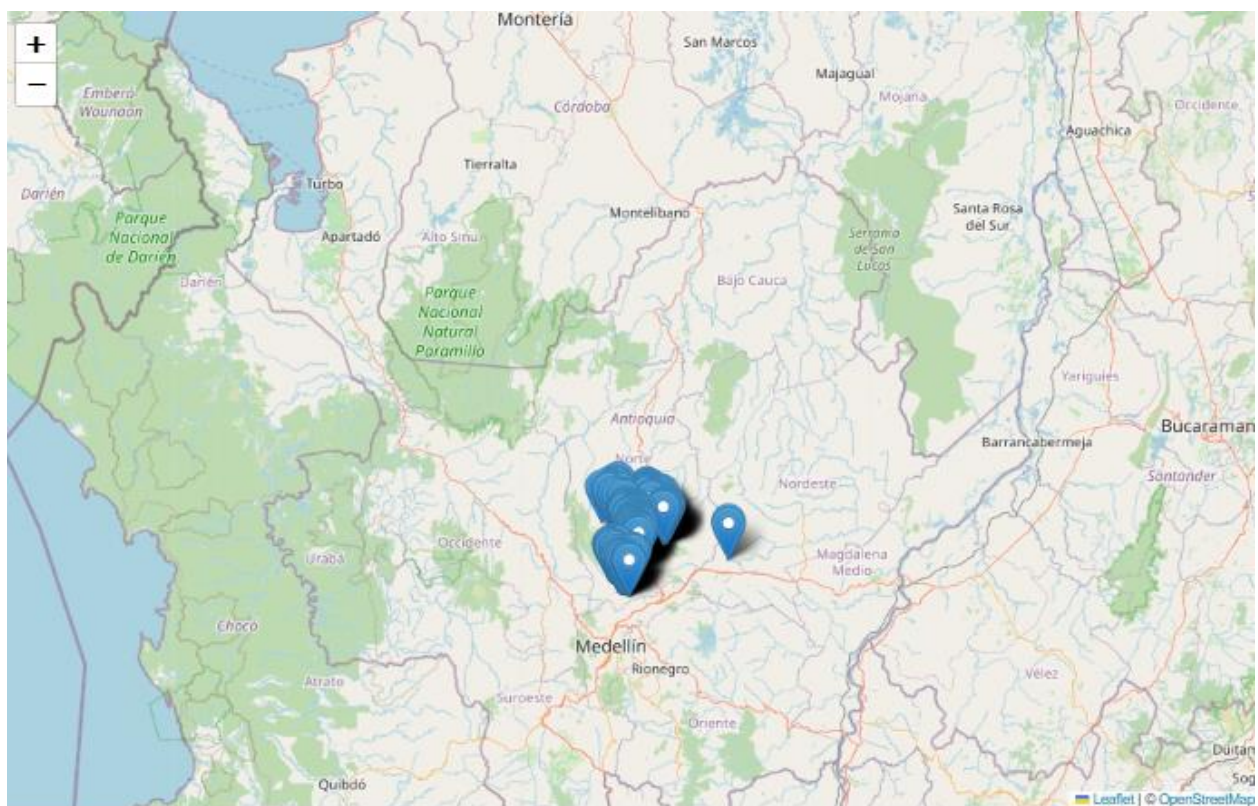
Podemos generar recomendaciones sobre la calidad de los días de consumo y facturación en una empresa de servicios públicos en Antioquia mediante el uso del análisis predictivo y siguiendo con el umbral definido en la variable objetivo días transcurridos, con Random Forest Classifier podríamos clasificar qué productos estarían en riesgo de acuerdo con las diferentes características arrojadas por el modelo. Este algoritmo nos arroja la mayor precisión, el modelo aporta información valiosa para desarrollar un sistema de monitoreo y control de calidad. En este caso, el modelo nos dio una precisión del 99%, y analizando la matriz de confusión, la variable objetivo tiene en la diagonal de verdaderos los valores o instancias con mayor número y

equilibrado, por este motivo las métricas del modelo son altas y confiables por encima del 80% y tiene buena capacidad para predecir.

También podemos ver una capa geográfica en donde podemos enfocar la mayor parte del esfuerzo operativo como también extraer reportes y ver por medio de un gráfico de dispersión la concentración en los días transcurridos y el enfoque de la periodicidad.

**Figura 10**

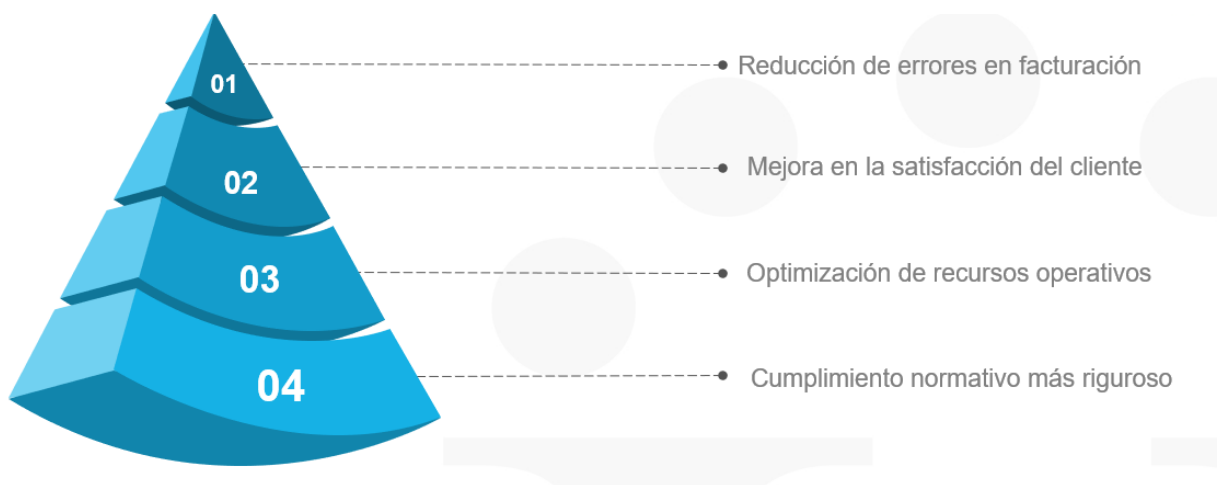
*Enfoque Geográfico del Resultado del Modelo*



## Resultados Obtenidos

**Figura 11**

*Piramide de Resultados Obtenidos*



A partir del análisis detallado de los resultados operativos, se logró una focalización precisa de las zonas con mayor impacto en la prestación de los servicios públicos EGA. Esta estrategia permitió optimizar la cobertura sin comprometer la capacidad operativa ni exceder los márgenes establecidos de intervención, garantizando así una gestión eficiente de los recursos disponibles.

La implementación de este enfoque no solo redujo significativamente los errores asociados al cálculo de los días de consumo facturable, sino que también fortaleció la precisión en la facturación, permitiendo cobrar al usuario final por un consumo real y justo. Este avance representa un hito importante en la mejora de la experiencia del cliente, al brindar mayor transparencia, confianza y satisfacción en la relación con el prestador del servicio.

Adicionalmente, esta mejora operativa asegura el cumplimiento riguroso de la normatividad vigente establecida por la Superintendencia de Servicios Públicos, lo que refuerza el compromiso institucional con la legalidad, la eficiencia y la calidad del servicio. En conjunto,

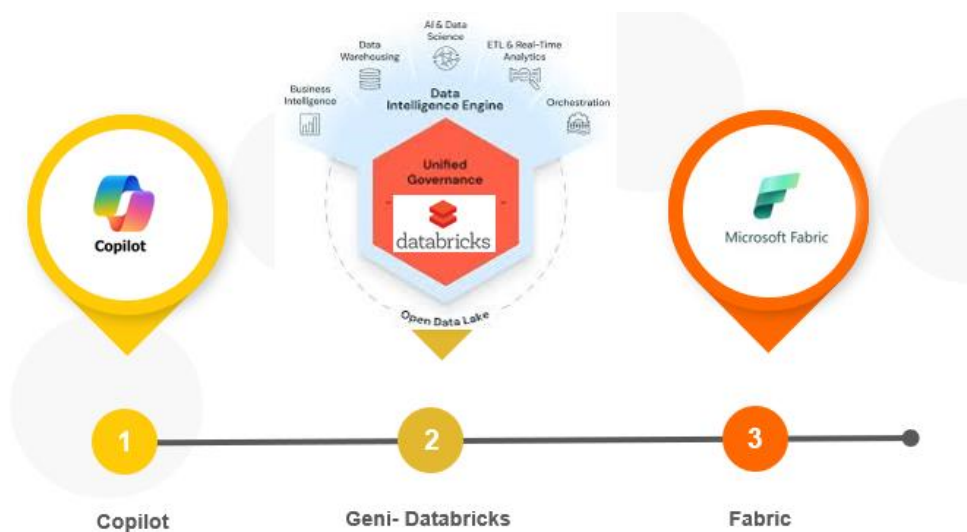
estas acciones consolidan una gestión orientada a resultados, centrada en el usuario y alineada con los principios de sostenibilidad y mejora continua.

## Potencial a Futuro

A futuro, este modelo tiene un alto potencial de evolución mediante la integración de tecnologías de inteligencia artificial en diversas plataformas, como Copilot, Geni de Databricks y la infraestructura unificada de Microsoft Fabric. Esta sinergia tecnológica abre un universo de posibilidades, permitiendo una toma de decisiones más eficiente, ágil y basada en datos a gran escala, superando ampliamente la capacidad de análisis humano. La incorporación de estas herramientas avanzadas no solo potenciará la automatización y la inteligencia operativa, sino que también consolidará un ecosistema digital robusto, preparado para enfrentar los desafíos de la analítica moderna y la transformación digital.

### Figura 12

#### *Potencial a Futuro*



## Conclusiones

El sistema desarrollado permite monitorear y controlar la calidad de los días de consumo y facturación, identificar y analizar inconsistencias, y recopilar datos para mejorar la eficiencia operativa de la empresa de servicios públicos en Antioquia. El uso de algoritmos de Python y técnicas de análisis de datos ha sido crucial para lograr estos objetivos.

Este enfoque en la exploración de los datos permitió una comprensión profunda y detallada de la información, facilitando la identificación de patrones y anomalías que son cruciales para mejorar la calidad del servicio y la satisfacción del cliente.

Los modelos de machine learning mostraron una alta precisión en la detección de anomalías, con un rendimiento perfecto en la clasificación de inconsistencias. Este sistema de alerta temprana puede mejorar la eficiencia y precisión en la facturación de servicios públicos.

## **Recomendaciones**

Se podría profundizar en cómo el modelo de machine learning contribuye a la precisión de las predicciones y justificar la elección del algoritmo con comparaciones frente a otras metodologías. También sería valioso incluir un apartado sobre los posibles desafíos en la implementación y cómo podrían abordarse, lo que daría una visión más completa del impacto del proyecto.

Proponer un plan piloto de implementación en una zona específica de Antioquia, con cronograma, recursos requeridos y métricas de éxito.

Diseñar un dashboard interactivo (por ejemplo, en Power BI o Qlik Sense) que permita visualizar en tiempo real las alertas de inconsistencias, zonas críticas y evolución de indicadores.

## Glosario

**Rango días de consumo** Son los días transcurridos entre la última fecha en que se visitó el inmueble y la fecha actual .

**Machine Learning (ML)** Se ocupa de la creación de modelos y algoritmos que permitan a las computadoras aprender de los datos y realizar tareas particulares sin necesidad de ser programadas explícitamente.

**Predicción** Proceso de estimar o pronosticar un resultado futuro o desconocido basado en datos históricos o información disponible en el presente.

**Regresión Lineal** Una técnica estadística que supone una relación lineal entre una variable dependiente y una o más variables independientes para simular esa relación

**Acuerdo de niveles de servicio (ANS)** Son las promesas de venta o servicio que la Empresa como proveedor hace a sus clientes para buscar la mejor calidad posible en los productos y en las transacciones, desde el punto de vista técnico, operativo y Comercial.

**Calidad** Grado en el que un conjunto de características inherentes cumple con los requisitos.

**Cliente** Persona natural o jurídica, quien se beneficia del servicio público de energía eléctrica, acueducto, agua residual o gas por red, bien como propietario del inmueble en donde se prestan los servicios, o como receptor directo del servicio. A este último usuario se denomina también consumidor.

**Instalación tipo rural (R)** Son los predios o instalaciones que se caracterizan por estar ubicados en zonas de alta dispersión, lo que significa que presentan una mayor presencia de terrenos no urbanizados, como áreas agrícolas, bosques, montañas y áreas protegidas.

Los criterios utilizados para definir una correría y/o predio como “Tipo R” pueden incluir aspectos como los tiempos de desplazamiento entre las distintas instalaciones que conforman el sector operativo, teniendo en cuenta tanto el desplazamiento desde un punto específico (por ejemplo, la sede o la Cabecera Municipal) hacia la zona de operación, como el tiempo de desplazamiento interno entre las instalaciones. Además, que los predios se encuentren en zonas que carecen de un trazado o nomenclatura de calles, carreras, avenidas y otros elementos viales.

Instalación tipo urbana (U) Son los predios o instalaciones que se caracterizan por estar ubicados en zonas con una baja dispersión, es decir, áreas que cuentan con conjuntos de edificaciones y estructuras contiguas agrupadas en manzanas destinadas a diversos usos, como viviendas, comercios, industrias y otras actividades.

Los criterios utilizados para definir una correría y/o un predio como “Tipo U”, pueden incluir aspectos como el tiempo de desplazamiento entre las distintas instalaciones del sector operativo, excluyendo el desplazamiento desde un punto específico (por ejemplo, la sede o la Cabecera Municipal) hacia la zona de operación. Además, los predios están delimitados por calles, carreras o avenidas y otros elementos viales, que se encuentran en una concentración mínima de veinte (20) viviendas contiguas, vecinas o adosadas entre sí, ubicadas en el área rural de un municipio o en un Corregimiento Departamental (Centros Poblados).

Contrato de Servicios Públicos Es un contrato consensual, de suministro, en virtud del cual un Prestador de Servicios Públicos presta el servicio público a un suscriptor y/o usuario a cambio de un precio en dinero. El contrato podrá contener, cláusulas escritas previamente definidas por aquel para ofrecerlas a muchos suscriptores y/o usuarios no determinados, todas las condiciones, actividades, usos, costumbres, prácticas, procedimientos que se aplican de manera

uniforme en la prestación del servicio y cláusulas pactadas especialmente con uno o algunos suscriptores y/o usuarios.

**Desviación significativa** Se entenderá por desviaciones significativas, en el período de facturación correspondiente, los aumentos o reducciones en los consumos, que, comparados con los promedios de los últimos tres períodos, si la facturación es bimestral, o de los últimos seis períodos, si la facturación es mensual, sean mayores a los porcentajes que señala la Comisión Reguladora.

**Lectura de medidores** Conjunto de actividades que se realizan para determinar los consumos facturables a los clientes y/o usuarios e incluirlos en la factura, durante el período de facturación mensual, bimestral o trimestral, rigiéndose por el calendario de facturación definido para este efecto.

**Queja** Es el medio por el cual el usuario o suscriptor pone de manifiesto su inconformidad a determinados funcionarios, o su inconformidad con la forma y condiciones en que se ha prestado el servicio. (Art. 45 decreto 1842 de 1991).

**Reclamación** Es una solicitud del suscriptor o usuario con el objeto de que una Entidad (pública o privada) prestadora de servicios públicos revise, mediante una actuación preliminar, la facturación de los servicios públicos y tome una decisión final o definitiva del asunto, en un todo de conformidad con los procedimientos previstos en el presente contrato, en la Ley 142 de 1994 y en el Código Contencioso Administrativo. (Art. 46 Decreto 1842 de 1991).

## Referencias Bibliográficas

- Alberca, A. S. (n.d.). \*La librería Pandas\*. Aprende con Alf.  
<https://aprendeconalf.es/docencia/python/manual/pandas/>
- Annual vs monthly billing cycles. (2024, November 19). \*Binary Stream\*.  
<https://binarystream.com/annual-vs-monthly-billing-cycles/>
- Castiglione, C. (2022, mayo 20). The best way to learn SQL. \*Learn To Code In 30 Days!\*  
<https://learn.onemonth.com/the-best-way-to-learn-sql/>
- Creg, P. (n.d.). \*Resoluciones\*. Portal CREG.  
<https://creg.gov.co/loader.php?lServicio=Documentos&lFuncion=infoCategoriaConsumo&tipo=RE&idDirectorio=366024>
- datetime — Basic date and time types. (n.d.). \*Python Documentation\*.  
<https://docs.python.org/3/library/datetime.html>
- Comisión de Regulación de Energía y Gas - CREG. (n.d.). \*Alejandría - Derechos, deberes y obligaciones\*. [https://gestornormativo.creg.gov.co/gestor/entorno/docs/ddo\\_1\\_6.htm](https://gestornormativo.creg.gov.co/gestor/entorno/docs/ddo_1_6.htm)
- Does how often you pay for it matter? The impacts of billing frequency. (n.d.). \*Unc.edu\*.  
<https://efc.web.unc.edu/2020/05/07/does-how-often-you-pay-for-it-matter-the-impacts-of-billing-frequency/>
- Emersion Systems. (2024, noviembre 25). The future of utility billing: Trends and predictions – emersion. \*Emersion.com\*. <https://www.emersion.com/blog/utility-billing-services/the-future-of-utility-billing-trends-and-predictions/>
- Etiquetado de recursos, grupos de recursos y suscripciones con Python - Azure Resource Manager. (n.d.). \*Microsoft.com\*. <https://learn.microsoft.com/es-es/azure/azure-resource-manager/management/tag-resources-python?tabs=windows>

Follow, A. (2018, diciembre 19). Classmethod() in Python. \*GeeksforGeeks\*.

<https://www.geeksforgeeks.org/classmethod-in-python/>

Henao, J. [@julianhenao2205]. (n.d.). Grabando 69 2. \*YouTube\*.

<https://www.youtube.com/watch?v=wP7iI8DIll0>

if ... elif ... else. (n.d.). \*Mclibre.org\*.

<https://www.mclibre.org/consultar/python/lecciones/python-if-else.html>

Introduction to Python | DCNC. (n.d.). \*DCNC Sciences\*.

<https://www.dcnsciences.com/introduction-to-python/>

j2logo. (2020, junio 1). Crear archivo Excel en Python con openpyxl. \*J2LOGO\*.

<https://j2logo.com/python/crear-archivo-excel-con-openpyxl/>

La CRA reitera compromiso en defensa de tarifas justas para cerca de 11 millones de usuarios.

(n.d.). \*Gov.co\*. <https://www.cra.gov.co/prensa/noticias/cra-reitera-compromiso-defensa-tarifas-justas-cerca-11-millones-usuarios>

Leer archivos Excel con Python. (2020, abril 1). \*Análisis y Decisión - Transformando datos en decisiones\*. <https://analisisydecision.es/leer-archivos-excel-con-python/>

Ley 142 de 1994 - Gestor Normativo. (n.d.). \*Gov.co\*.

<https://www.funcionpublica.gov.co/eva/gestornormativo/norma.php?i=2752>

NumPy. (n.d.). \*Numpy.org\*. <https://numpy.org/>

openpyxl - A Python library to read/write Excel 2010 xlsx/xlsm files — openpyxl 3.1.2

documentation. (n.d.). \*Readthedocs.io\*. <https://openpyxl.readthedocs.io/en/stable/>

os — Interfaces misceláneas del sistema operativo — documentación de Python - 3.10.13. (n.d.).

\*Python.org\*. <https://docs.python.org/es/3.10/library/os.html>

os.path — Common pathname manipulations. (n.d.). \*Python documentation\*.

<https://docs.python.org/es/3/library/os.path.html>

Peset, F., & Millán, L. (2017). \*Ciencia abierta y gestión de datos de investigación (RDM)\* (pp. 55–72). Ediciones Trea.

Pineda, R. (2022). \*Aplicación de la Ciencia de Datos - Líneas de Investigación (ECBTI) [OVI]\*. <https://repository.unad.edu.co/handle/10596/50443>

Qvd. (n.d.). \*PyPI\*. <https://pypi.org/project/qvd/>

Tesaurus de la UNESCO. (n.d.). \*Unesco.org\*.

<https://vocabularies.unesco.org/browser/thesaurus/es/>

TOPdesk. (n.d.). qlik-sense-logo - \*TOPdesk Marketplace\*.

<https://marketplace.topdesk.com/qlik-sense-logo/>

Truety.com. (n.d.). \*Truety.com\*. <https://www.truety.com/submetering-vs-rubs-fair-utility-billing-explained/>

¿Cómo crear un DataFrame? (2023, junio 28). \*4Geeks\*. <https://4geeks.com/es/how-to/crear-dataframe>

Ebooks7.24.com. (n.d.). \*Ebooks7.24.com\*.

<https://www.ebooks7.24.com/stage.aspx?il=6443&pg=&ed=>

Org.Mx. (n.d.). \*Scielo.org.mx\*.

[https://www.scielo.org.mx/scielo.php?script=sci\\_arttext&pid=S0185-16672003000300015&lang=es](https://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S0185-16672003000300015&lang=es)

## Apéndices

### Apéndice A

#### *Exploración de los Datos*

Data Base	Productos leídos 2024-2025	
	Columnas	27
	Registros	3.116.069
<hr/> 1 ORDER_ID 2 SESUSERV 3 SESUCATE 4 SESUSUCA 5 CICLO 6 SESUPLFA 7 LEGALIZATION_DATE 8 ASSIGNED_DATE 9 EXECUTION_FINAL_DATE 10 PERFIL_FRAUDES 11 ESTADO_CUENTA 12 ESTADO_MORA 13 CONSUMO_LEIDO 14 CAUSANL 15 DESC_CAUSANL 16 OBS_ADIC 17 DESC_OBS_ADIC 18 PERIODICIDAD 19 F_LECT_ANTERIOR 20 DESC_CAUSANL_ANTERIOR 21 DESC_OBS_ADIC_ANTERIOR 22 NOVED_LECT_ANT <hr/>		

---

23 COORDENADAS\_X

24 COORDENADAS\_Y

25 ROUTE\_ID

---

## Apéndice B

### *Caracterización y Clasificación*

Data Base	Productos leídos 2024-2025	
	Columnas	27
	Registros	2.890
ORDER_ID	object	
SESUSERV	object	
SESUCATE	object	
SESUSUCA	object	
CICLO	object	
SESUPLFA	object	
LEGALIZATION_DATE	datetime64[ns]	
ASSIGNED_DATE	object	
EXECUTION_FINAL_DATE	datetime64[ns]	
PERFIL_FRAUDES	object	
ESTADO_CUENTA	object	
ESTADO_MORA	object	
CONSUMO_LEIDO	object	
CAUSANL	object	
DESC_CAUSANL	object	
OBS_ADIC	object	
DESC_OBS_ADIC	object	
PERIODICIDAD	object	
F_LECT_ANTERIOR	datetime64[ns]	
DESC_CAUSANL_ANTERIOR	object	
DESC_OBS_ADIC_ANTERIOR	object	
NOVED_LECT_ANT	object	
COORDENADAS_X	object	
COORDENADAS_Y	object	

---

ROUTE_ID	object
----------	--------

---

## Apéndice C

### *Video de Socialización del Trabajo de Grado*

---

<b>Nombre del video</b>	<b>Enlace</b>
Sistema inteligente de monitoreo y control de calidad para los días de consumos facturables de los servicios públicos EGA en Antioquia	<a href="https://youtu.be/wP7iI8DIllo?si=LnWVwOBm8F6vksrf">https://youtu.be/wP7iI8DIllo?si=LnWVwOBm8F6vksrf</a>

---