

**Diseño de una herramienta de ciencia de datos para la detección de fraudes en el consumo
de energía en empresa de servicios públicos Emcali**

Walter Andres Betancourt Mendez

Asesor

Claudia Patricia Hernandez Arbelaez

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

2025

Resumen

El presente trabajo es una propuesta de diseño de una herramienta basada en ciencia de datos para detectar pérdidas no técnicas y comportamientos irregulares en el consumo de energía en la empresa de servicios públicos Emcali. El estudio se orienta desde un enfoque cuantitativo y analítico, empleando un diseño no experimental aplicado al análisis de datos históricos de consumo energético. La metodología incluyó procesos de recolección, depuración y análisis exploratorio de datos, seguidos de la implementación de técnicas de analítica avanzada mediante modelos de detección de anomalías. Los resultados evidenciaron que las variaciones abruptas del consumo y los cambios inusuales frente al promedio semestral constituyen los principales indicadores de irregularidad, hallazgo coherente con los antecedentes revisados. Asimismo, se identificaron las variables más relevantes para el diseño futuro de una herramienta de monitoreo automatizado. Las conclusiones confirman la pertinencia del uso de métodos analíticos para fortalecer la identificación temprana de consumos atípicos y optimizar la gestión operativa de la empresa. Finalmente, se recomienda avanzar en la integración de sistemas inteligentes, estandarizar la calidad de los registros y ampliar el análisis hacia categorías de usuarios prioritarios para mejorar la efectividad del proceso de detección.

Palabras claves: Fraude energético, analítica de datos, consumo eléctrico, detección de anomalías, pérdidas no técnicas.

Abstract

This work proposes the design of a data-driven tool aimed at detecting non-technical losses and irregular consumption behaviors in the electric energy service of the public utility company Emcali. The study follows a quantitative and analytical approach, using a non-experimental design applied to the examination of historical energy consumption records. The methodology included processes of data collection, cleaning, and exploratory analysis, followed by the implementation of advanced analytics techniques through anomaly detection models. The results show that abrupt consumption variations and unusual deviations from the semiannual average represent the main indicators of irregularity, a finding consistent with the reviewed literature. Likewise, the most relevant variables for the future development of an automated monitoring tool were identified. The conclusions confirm the relevance of analytical methods to strengthen early identification of atypical consumption and to optimize the company's operational management. Finally, it is recommended to advance toward the integration of intelligent systems, standardize data quality, and extend the analysis to priority user categories to improve the effectiveness of the detection process.

Keywords: Energy fraud, Data analytics, Electric consumption, Anomaly detection, Non-technical losses

Tabla de Contenido

Introducción	8
Problema	9
Descripción del Problema	9
Justificación	12
Objetivos	14
Objetivo General	14
Objetivos Específicos.....	14
Marco de Referencia	15
Antecedentes	15
Marco Contextual.....	17
Marco Teorico.....	19
Marco Conceptual.....	20
Marco Legal	22
Metodología	23
Paradigma	23
Enfoque	23
Tipo de Investigación.....	23
Población.....	24
Muestra	24
Técnicas e Instrumentos de Recolección de Información.....	25
Técnicas de Análisis de Datos	25
Resultados.....	26

Recolectar los Datos Históricos de Consumo de Energía y los Registros Asociados a Pérdidas No Técnicas	26
Analizar la Información Mediante Técnicas de Ciencia de Datos que Permitan Identificar Patrones.....	27
Identificar las Variables que Influyen en la Detección de Pérdidas No Técnicas y Comportamientos Irregulares.....	28
Diseño de la Herramienta Analítica Propuesta	31
Descripción General de la Herramienta	31
Arquitectura y Flujo de Procesamiento de Datos	31
Módulos Funcionales de la Herramienta	31
Módulo de Proyección de Consumo por Ciclo	32
Módulo de Histórico de Consumos.....	33
Módulo de Mapa de Consumo Energéticos	35
Módulo Análisis Estadístico Avanzado por Categoría	37
Predicción y Proyección de Mercados (Forecasting).....	39
Técnicas de Analítica de Datos Implementadas	41
Alcance y Consideraciones del Diseño	41
Conclusiones.....	42
Recomendaciones	44
Referencias Bibliográficas	46

Lista de Tablas

Tabla 1 *Variaciones de Consumo y Anomalías Detectadas* 26

Tabla 2 *Variables más Influyentes en la Detección de Pérdidas No Técnicas* 29

Lista de Figuras

Figura 1	<i>Distribución del Puntaje de Anomalías</i>	27
Figura 2	<i>Distribución Porcentual de Dnomalías por Categoría</i>	29
Figura 3	<i>Módulo de Proyección de Consumo por Ciclo</i>	32
Figura 4	<i>Módulo de Proyección de Consumo por Ciclo (diagrama de barras)</i>	33
Figura 5	<i>Módulo de Proyección de Consumo por Ciclo (Metodo de Calculo de Consumo)</i>	33
Figura 6	<i>Módulo de Histórico de Consumos</i>	34
Figura 7	<i>Módulo de Histórico de Consumos (Puntos con Irregularidades)</i>	35
Figura 8	<i>Módulo de Mapa de Consumo Energetico</i>	36
Figura 9	<i>Módulo de Mapa de Consumo Energético de la Ciudad de Cali</i>	36
Figura 10	<i>Módulo Análisis Estadístico Avanzado por Categoría</i>	38
Figura 11	<i>Módulo Análisis Estadístico Avanzado por Categoría (Segmentos)</i>	38
Figura 12	<i>Módulo Análisis Estadístico Avanzado por Categoría Grafico de Dispersión</i>	39
Figura 13	<i>Predicción y Proyección de Mercados</i>	40
Figura 14	<i>Pronostico de Consumo Mesual</i>	40

Introducción

Las pérdidas no técnicas en el consumo de energía representan un reto importante para las empresas de servicios públicos, ya que afectan directamente la precisión de la facturación, los ingresos y la eficiencia operativa. Estas pérdidas suelen originarse en irregularidades como manipulaciones de medidores, consumos atípicos o errores en los métodos de cálculo. Debido a ello, las empresas requieren herramientas que permitan identificar comportamientos inusuales de manera oportuna y basada en datos confiables.

En el contexto local, Emcali enfrenta esta situación debido al volumen de usuarios, la diversidad de métodos de medición y la existencia de variaciones significativas en el consumo que no siempre cuentan con una explicación verificable. Esta necesidad motivó el desarrollo de una herramienta que permita analizar de forma sistemática los datos históricos de consumo, identificar patrones irregulares y apoyar los procesos de control de energía.

El presente trabajo se estructura en función de esa necesidad. Primero se describe el problema identificado en el proceso de control de pérdidas no técnicas. Luego se plantean los objetivos general y específicos que orientan la propuesta. Posteriormente, se desarrolla el marco conceptual que fundamenta la construcción de la herramienta. En la parte metodológica se explica el proceso aplicado para la recolección, tratamiento y análisis de los datos, así como la forma en que se diseñó y validó la herramienta. Finalmente, se presentan los resultados obtenidos y las conclusiones derivadas del desarrollo del proyecto.

Problema

Descripción del Problema

A nivel mundial, las pérdidas no técnicas derivadas del fraude y la manipulación de los sistemas de medición representan uno de los principales desafíos para las empresas comercializadora de energía y agua. Estas pérdidas afectan directamente la sostenibilidad económica del sector energético, incrementando los costos de operación y reduciendo la capacidad de inversión en infraestructura moderna (Al-Radaideh & Al-Zoubi, 2018).

De acuerdo con Nofal et al. (2022) en regiones de Asia y Medio Oriente el uso de herramientas de análisis de datos y algoritmos de detección de anomalías puede reducir el consumo irregular en los sistemas de suministro de energía; se identifica que patrones atípicos que no son detectables mediante inspecciones manuales. Estas experiencias internacionales han impulsado la adopción de tecnologías de inteligencia artificial y ciencia de datos en la gestión de redes energéticas inteligentes, consolidando una tendencia global hacia la digitalización del control de pérdidas.

En América Latina, la situación presenta retos similares, pero con una infraestructura más limitada y una alta dependencia de procesos manuales. En países como Costa Rica y Ecuador destacan la necesidad de implementar herramientas de analítica avanzada para la gestión del consumo eléctrico y la mejora de la eficiencia energética (Barquero-Álvarez, 2021; Solis Mora & Gruezo, 2022). No obstante, la transición hacia el uso de tecnologías predictivas aún se encuentra en etapas iniciales, lo que dificulta la identificación oportuna de fraudes o manipulaciones indebidas.

Segun lo anterior las pérdidas no técnicas no solo afectan a las empresas distribuidoras de energía, sino también a los usuarios finales, quienes asumen sobrecostos en las tarifas de

energía debido a los desequilibrios financieros que estos fraudes generan, afectando la sostenibilidad del servicio y reduciendo la capacidad de inversión en mejoras tecnológicas. Por ello, contar con herramientas que permitan identificar patrones atípicos de consumo de forma oportuna se vuelve fundamental para proteger tanto la estabilidad operativa como el bienestar económico de los usuarios.

En el contexto colombiano, el problema adquiere relevancia, Gallon (2022), asegura que las pérdidas no técnicas en el sector eléctrico representan un porcentaje significativo de la energía distribuida, generando impactos económicos que superan los 140 millones de dólares anuales, afectando tanto a los ingresos de las empresas como al equilibrio tarifario nacional. Aunque la cifra debe ser revisada con cautela por variaciones en la tasa de cambio, el impacto económico sigue siendo considerable.

Arias y Méndez (2022) demuestran que la adopción de herramientas basadas en ciencia de datos puede optimizar la detección de fraudes, reducir las pérdidas y mejorar la toma de decisiones en empresas de servicios públicos, Esta evidencia pone de manifiesto que la problemática no se debe únicamente a la manipulación de los equipos de medición, sino también a la falta de mecanismos modernos que permitan identificar comportamientos irregulares de manera oportuna y sistemática.

De manera particular, en la ciudad de Cali, la empresa de servicios públicos EMCALI enfrenta un desafío constante en la detección de irregularidades en los sistemas de medición de energía. En áreas específicas de la red, se presentan inconsistencias significativas entre los consumos registrados y los esperados, derivadas de manipulaciones, errores en las lecturas o prácticas fraudulentas. La falta de un sistema automatizado de análisis de datos ha llevado a que las tareas de control y verificación se realicen manualmente, consumiendo tiempo y recursos, y

aumentando el riesgo de errores humanos. Frente esta problemática surge la pregunta que orienta este proceso de investigación:

¿Cómo puede una herramienta basada en ciencia de datos contribuir a la detección temprana de fraudes en el consumo de energía en empresas de servicios públicos en la ciudad de Cali?

Justificación

La justificación del trabajo consiste defender la necesidad e importancia de la realización del mismo, de acuerdo con Méndez (2012) la justificación explica la relevancia e importancia del estudio, bien sea por conveniencia, implicaciones prácticas, valor teórico, social e inclusive metodológico. Este trabajo se justifica desde tres ejes fundamentales; la importancia de uso de las herramientas basadas en ciencia de datos para la toma de decisiones, la gestión eficiente del servicio de energía y la necesidad de evitar fraudes y pérdidas no técnicas en dicho servicio.

Arias y Méndez (2022) demuestran que las herramientas basadas en ciencia de datos mejoran la detección de fraudes y apoyan la toma de decisiones en empresas de servicios públicos. Su estudio evidencia que los métodos tradicionales, basados en inspecciones manuales, no logran identificar patrones complejos de consumo. Esto genera retrasos en la detección y aumenta los costos operativos. Por ello, se requiere una herramienta que automatice el análisis y permita identificar irregularidades con mayor precisión.

Barquero-Álvarez (2021) señala que en América Latina la gestión eficiente del consumo eléctrico depende cada vez más del uso de analítica avanzada debido a los altos niveles de pérdidas no técnicas. Su investigación demuestra que las técnicas tradicionales son insuficientes para reconocer variaciones anómalas en grandes volúmenes de datos. Esta evidencia muestra que el problema no solo es la existencia de fraude, sino la falta de mecanismos modernos de monitoreo. En este contexto, una herramienta analítica para Emcali resulta pertinente y necesaria.

Gallón (2022) evidencia que en Colombia las pérdidas no técnicas continúan siendo un desafío debido a la manipulación de medidores y a la baja automatización en los sistemas de inspección. Estas pérdidas afectan la sostenibilidad financiera de los operadores y generan

incrementos tarifarios para los usuarios. La situación nacional confirma la necesidad de modernizar los procesos de detección en empresas como Emcali. Por ello, una herramienta de análisis basada en ciencia de datos contribuiría directamente a reducir estas pérdidas.

Rodríguez y Antonio (2024) muestran que la integración de análisis geoespacial con técnicas de detección de anomalías permite identificar zonas críticas con mayor rapidez. Esta combinación facilita la focalización de inspecciones y mejora la eficiencia operativa. La evidencia es especialmente relevante para Emcali, donde gran parte del análisis actual depende de reportes manuales. En consecuencia, la creación de una herramienta analítica responde a una necesidad real del contexto local.

Objetivos

Objetivo General

Diseñar una herramienta basada en ciencia de datos que permita detectar pérdidas no técnicas y comportamientos irregulares en el consumo de energía eléctrica en la empresa de servicios públicos Emcali.

Objetivos Específicos

Recolectar los datos históricos de consumo de energía y los registros asociados a pérdidas no técnicas o comportamientos irregulares en la empresa Emcali.

Analizar la información mediante técnicas de ciencia de datos que permitan identificar patrones y posibles irregularidades en el consumo eléctrico.

Identificar las variables más relevantes que influyen en la detección de pérdidas no técnicas y comportamientos irregulares, con el fin de utilizarlas como insumo para el diseño de la herramienta.

Marco de Referencia

Antecedentes

Al-Radaideh y Al-Zoubi (2018), en su estudio *A data mining based model for detection of fraudulent behaviour in water consumption*, desarrollado en Jordania, tuvieron como objetivo identificar patrones irregulares de consumo mediante técnicas de minería de datos. La investigación utilizó modelos supervisados y no supervisados aplicados a series temporales de consumo hídrico. Los autores demostraron que los algoritmos basados en árboles y clustering permiten detectar consumos atípicos con altos niveles de precisión. Concluyen que los métodos analíticos superan ampliamente los procesos manuales de inspección. Su aporte a este proyecto radica en evidenciar que los enfoques de ciencia de datos pueden adaptarse al consumo eléctrico para mejorar la detección temprana de anomalías.

Barquero-Álvarez (2021), en *Desarrollo de metodología para la gestión de consumo eléctrico basada en ciencia de datos* realizado en Costa Rica, planteó como objetivo optimizar la gestión energética mediante modelos analíticos aplicados a los registros de consumo. El autor utilizó análisis exploratorio, modelos predictivos y visualizaciones automatizadas que permitieron identificar usuarios con comportamientos inusuales. Como resultado, se logró mejorar la eficiencia operativa y la focalización de inspecciones. Concluye que los sistemas basados en datos potencian la gestión energética. Este aporte es relevante para este proyecto, pues evidencia la aplicabilidad de la analítica avanzada en el sector eléctrico latinoamericano.

Solis Mora y Gruezo (2022), en *La Inteligencia Artificial al servicio de la eficiencia energética en Ecuador*, examinaron el uso de herramientas de IA para mejorar la gestión del consumo eléctrico y reducir pérdidas. Para ello emplearon un enfoque documental y estudios de caso comparativos en distribuidoras de energía. Los autores identificaron que la IA permite

automatizar procesos de monitoreo y detectar anomalías en tiempo real. Concluyen que la digitalización energética es clave para enfrentar pérdidas no técnicas. Su contribución a este proyecto radica en mostrar la efectividad de la IA como base tecnológica para una herramienta orientada a detectar irregularidades.

Arias y Méndez (2022) desarrollaron el trabajo Metodología basada en minería de datos para la detección de pérdidas no técnicas en sistemas de distribución de energía eléctrica, aplicado en Ecuador con el fin de identificar patrones de fraude y manipulación en redes eléctricas. Utilizaron técnicas de minería de datos, análisis de clústeres y árboles de decisión para clasificar consumos irregulares. Los resultados demostraron que la analítica permite mejorar la precisión en la detección de pérdidas y priorizar visitas de inspección. Este estudio apoya directamente este proyecto, ya que valida la posibilidad de aplicar metodologías similares en la empresa Emcali.

Gallón (2022), en *Análisis de los registros de un operador de red eléctrica nacional para la búsqueda de pérdidas no técnicas*, desarrolló un estudio aplicado en Colombia sobre bases masivas de consumo. Su objetivo fue identificar anomalías mediante análisis estadístico, reglas de negocio y minería de datos. Los hallazgos revelaron que ciertos patrones de variabilidad permiten anticipar posibles fraudes antes de que generen pérdidas significativas. Concluyó que los métodos analíticos fortalecen las estrategias de control de empresas distribuidoras. El aporte para este proyecto es fundamental, pues contextualiza la problemática en el país y permite adaptar enfoques al entorno local de Emcali.

Rodríguez y Antonio (2024), en *Detección de fraudes en empresas de servicios públicos utilizando ciencia de datos y análisis geoespaciales*, desarrollaron un estudio en Bogotá en el que combinaron técnicas de machine learning con análisis espacial. La metodología incluyó

Random Forest, Isolation Forest y mapas geográficos de densidad para resaltar zonas de riesgo. Los resultados mostraron que la combinación de variables técnicas y geográficas incrementa la precisión en la detección de fraudes. Para este proyecto, su aporte radica en demostrar la utilidad de integrar información espacial al análisis del consumo energético en Emcali.

Giraldo Sandoval (2024), en *Desarrollo de un sistema para la detección de manipulación indebida de equipos de medición en usuarios industriales*, elaborado en Cartagena, tenía como objetivo crear un sistema de alertas tempranas a partir de medición inteligente. El autor empleó análisis de series temporales, modelamiento predictivo y reglas determinísticas para identificar inconsistencias. Los resultados evidenciaron disminuciones abruptas de consumo asociadas a posibles intervenciones indebidas. Su aporte a este proyecto consiste en mostrar que los datos provenientes de sistemas de medición ofrecen información clave para detectar irregularidades en usuarios residenciales y comerciales.

Roys y Stefan (2019), en su tesis *Detección de fraude de energía a partir de un modelo de lógica difusa*, desarrollada en Colombia, buscaban diseñar un sistema basado en inferencia difusa para clasificar consumos sospechosos. Aplicaron funciones de pertenencia asociadas al consumo promedio, la variabilidad histórica y el comportamiento típico de usuarios comparables. Los resultados mostraron que la lógica difusa permite abordar escenarios donde no existe una clasificación exacta del fraude. Este antecedente es valioso porque evidencia alternativas metodológicas complementarias que enriquecen la comprensión del problema de pérdidas no técnicas.

Marco Contextual

La empresa Emcali es una entidad pública encargada de la prestación de servicios de energía, acueducto y telecomunicaciones en la ciudad de Santiago de Cali. En el caso del

servicio eléctrico, la empresa administra una red de distribución compleja, compuesta por diversos circuitos y transformadores que atienden sectores residenciales, comerciales e industriales. Debido a esta amplitud operativa y a la diversidad de usuarios, la detección de comportamientos inusuales en el consumo de energía se convierte en un desafío permanente. La empresa enfrenta variaciones de consumo asociadas tanto a factores técnicos como a comportamientos irregulares que dificultan la gestión eficiente del recurso energético.

En los últimos años, Emcali ha identificado un aumento en los casos relacionados con pérdidas no técnicas, especialmente aquellos vinculados a manipulaciones de medidores, conexiones ilegales y variaciones atípicas de consumo que no pueden explicarse mediante los registros históricos. Esta situación genera impactos económicos significativos para la empresa, pues disminuye el volumen real de energía facturada y afecta los indicadores de recuperación de ingresos. Además, las revisiones manuales, los tiempos de respuesta prolongados y la inexistencia de una herramienta analítica automatizada limitan la capacidad institucional para actuar de manera oportuna frente a estos casos.

Este contexto evidencia la necesidad de fortalecer los procesos de análisis de consumo dentro de la empresa, especialmente en áreas que operan con grandes volúmenes de información, como la Unidad de Control y la Unidad de Pérdidas de Energía. La gestión actual depende de archivos en Excel, inspecciones visuales y revisiones manuales que dificultan la detección temprana de anomalías. Por ello, surge la importancia de implementar una herramienta basada en ciencia de datos que permita una identificación más precisa, sistemática y automatizada de comportamientos irregulares en el consumo eléctrico. Esta herramienta apoyaría la toma de decisiones, la priorización de visitas técnicas y la reducción de pérdidas, aportando a la sostenibilidad operativa y financiera de Emcali.

Marco Teorico

La ciencia de datos se ha consolidado como un campo interdisciplinar que integra estadística, machine learning y computación para generar conocimiento a partir de grandes volúmenes de información. Autores como Barquero-Álvarez (2021) destacan que su aplicación en la gestión energética permite transformar datos operativos en insumos estratégicos para la toma de decisiones, especialmente en contextos donde existen variaciones de consumo difíciles de interpretar mediante métodos tradicionales. En este marco, las técnicas de minería de datos facilitan la detección temprana de patrones irregulares, lo que resulta fundamental para empresas de servicios públicos que deben gestionar grandes cantidades de información provenientes de redes y usuarios.

La literatura sobre pérdidas no técnicas señala que las anomalías en el consumo constituyen uno de los principales retos para los operadores de red. Estudios como los de Arias y Méndez (2022) y Nofal et al. (2022) muestran que los modelos analíticos pueden identificar comportamientos inusuales mediante técnicas de detección de outliers, clasificación supervisada y análisis de series temporales. Estas investigaciones evidencian que, al comprender la estructura del consumo histórico, es posible reconocer desviaciones relevantes asociadas a fraude, manipulación de equipos de medición o inconsistencias en la lectura. Así, el uso de herramientas analíticas se convierte en un mecanismo eficaz para optimizar la vigilancia de los sistemas eléctricos.

Otro aporte teórico relevante proviene de los trabajos que abordan el diseño e implementación de herramientas tecnológicas para la gestión de consumos. Giraldo Sandoval (2024) resalta que la integración de algoritmos predictivos en plataformas interactivas facilita la generación de alarmas y la priorización de usuarios con riesgo de fraude. De manera similar,

Roys y Stefan (2019) demuestran que los sistemas computacionales basados en lógica difusa mejoran la identificación de irregularidades al interpretar valores imprecisos y evaluar la consistencia del comportamiento energético de los usuarios. Estas contribuciones fundamentan la importancia de desarrollar herramientas que integren múltiples métodos analíticos y presenten resultados de manera clara para el personal técnico.

Finalmente, los avances en análisis exploratorio de datos (EDA) constituyen un referente teórico clave para este proyecto, pues permiten caracterizar el comportamiento histórico del consumo y comprender la variabilidad existente entre usuarios. Autores como Gallón (2022) enfatizan que el análisis preliminar de distribuciones, tendencias y correlaciones es indispensable para la construcción de modelos confiables y la detección de patrones anómalos. En este sentido, el EDA actúa como una etapa metodológica que orienta la selección de variables, la transformación de datos y la interpretación de los resultados obtenidos por las herramientas analíticas.

Marco Conceptual

El concepto de pérdidas no técnicas hace referencia a la energía que no es facturada debido a causas ajenas a la operación del sistema, tales como fraudes, manipulaciones de medidores, conexiones ilegales o errores en los procesos de lectura y facturación. Estas pérdidas generan desequilibrios económicos en los sistemas de distribución, afectando los ingresos de las empresas prestadoras del servicio y aumentando los costos operativos. Su identificación requiere examinar el comportamiento del consumo y reconocer desviaciones que no obedecen a fallas técnicas, sino a acciones externas o comportamientos irregulares por parte de los usuarios.

El término comportamiento irregular de consumo alude a variaciones del uso de energía que se apartan de los patrones históricos y que no pueden explicarse por factores estacionales,

cambios de actividad o condiciones operativas. Este tipo de comportamiento suele manifestarse como disminuciones bruscas, incrementos inesperados o inconsistencias entre el consumo registrado y el perfil del usuario. Reconocer estas irregularidades es fundamental para orientar las acciones de verificación, ya que ofrecen señales tempranas sobre posibles anomalías o fraudes dentro del sistema de medición.

El análisis de datos se entiende como el conjunto de procedimientos que permiten recopilar, depurar, transformar y examinar información con el fin de extraer conclusiones o generar conocimiento útil. Su aplicación en el sector eléctrico facilita la identificación de patrones, tendencias y relaciones entre variables relevantes para la detección de pérdidas no técnicas. Este análisis no solo permite comprender el comportamiento general del consumo, sino también segmentar a los usuarios y priorizar los casos que requieren verificación por parte de los equipos técnicos.

El concepto de anomalía en datos se refiere a cualquier observación que se aleja significativamente del comportamiento esperado dentro de un conjunto de información. En el contexto del consumo energético, una anomalía puede representar un fraude, una manipulación del medidor o un error en el registro. La detección de anomalías es esencial para el desarrollo de herramientas basadas en ciencia de datos, ya que permite distinguir entre un consumo normal y uno que requiere intervención o seguimiento adicional por parte de la empresa.

Finalmente, una herramienta analítica corresponde a un sistema, aplicación o conjunto de procedimientos que posibilitan la integración, procesamiento y visualización de datos con el propósito de apoyar la toma de decisiones. En proyectos aplicados como este, dichas herramientas permiten automatizar el análisis de grandes volúmenes de información y presentar resultados de manera accesible para el personal encargado de gestionar las pérdidas. Esto

contribuye a mejorar la eficiencia operativa y a fortalecer los mecanismos de control en la empresa de servicios públicos.

Marco Legal

En Colombia, la regulación del servicio de energía establece normas claras sobre el uso adecuado del sistema eléctrico y sanciones por fraude. La Ley 142 de 1994, conocida como la Ley de Servicios Públicos Domiciliarios, define el fraude como una acción que modifica la medición real del consumo y autoriza a las empresas a realizar inspecciones y aplicar cargos por energía dejada de facturar.

La Resolución CREG 108 de 1997 regula los procedimientos para la detección y liquidación de consumos irregulares. Esta norma indica que los operadores deben utilizar metodologías técnicas para determinar la energía consumida cuando se detectan irregularidades.

La Ley 143 de 1994, por su parte, establece el régimen del sector eléctrico y confirma la responsabilidad de las empresas en garantizar un servicio eficiente y seguro. Esto incluye la obligación de reducir las pérdidas técnicas y no técnicas.

Finalmente, el Código Penal Colombiano contempla sanciones por defraudación de fluidos (Artículo 256), lo que refuerza la importancia legal de implementar herramientas que permitan identificar y documentar irregularidades en el consumo eléctrico.

Metodología

Paradigma

El proyecto se enmarca en el paradigma cuantitativo, el cual se orienta a la medición objetiva, la comprobación empírica y el análisis estadístico. Hernández et al. (2014) señalan que el enfoque cuantitativo se fundamenta en la recolección de datos numéricos y en el uso de análisis estadísticos para establecer patrones y relaciones; desde este enfoque se explican fenómenos a partir de datos observables y verificables, permitiendo validar hipótesis mediante procedimientos sistemáticos. Este paradigma resulta especialmente pertinente cuando se analizan grandes volúmenes de información, como ocurre en entornos de analítica de datos aplicada al sector energético.

Enfoque

El estudio sigue un enfoque analítico, ya que busca resolver un problema práctico a través del análisis de datos reales. Según Hurtado de Barrera (2000), la investigación aplicada se orienta a intervenir o transformar una situación concreta mediante la generación de soluciones fundamentadas; este tipo de estudios se utilizan para producir resultados útiles para la toma de decisiones. En este sentido, el enfoque analítico permite descomponer el fenómeno de las pérdidas no técnicas en sus componentes esenciales, facilitando la identificación de patrones y relaciones significativas entre variables.

Tipo de Investigación

El proyecto es aplicado, con diseño no experimental. Hernández et al. (2014) establecen que en los estudios no experimentales no se manipulan variables, sino que se observan tal como ocurren en su contexto; por tanto, buscan determinar el grado de relación entre variables. El diseño no experimental permite comprender el comportamiento de los datos en un momento

específico cuando la manipulación no es posible ni pertinente. Este tipo de investigación resulta adecuado para estudios basados en grandes volúmenes de información histórica, como los consumos de energía, donde el investigador requiere analizar registros reales sin intervenir en su generación.

Población

La población del estudio corresponde al conjunto total de registros históricos disponibles en la base de datos de consumo de energía de la empresa Emcali, Hurtado de Barrera (2000) define la población como el universo de elementos que comparten características observables y sobre los cuales se pretende generalizar los resultados de una investigación. En este caso, dicho universo está constituido por todos los usuarios y períodos de facturación incluidos en los archivos institucionales, los cuales reflejan patrones reales de uso, anomalías y comportamientos asociados a pérdidas no técnicas. Esta población es especialmente relevante por su volumen y diversidad, lo que permite analizar tendencias globales y particularidades del consumo energético.

Muestra

La muestra utilizada en este estudio corresponde a un subconjunto del total de registros disponibles, seleccionado mediante un muestreo no probabilístico por conveniencia. Según Hernández et al. (2014), este tipo de muestreo se emplea cuando el investigador accede a los datos que están disponibles y cumplen las condiciones necesarias para el análisis, sin que todos los elementos de la población tengan la misma probabilidad de ser incluidos. En este proyecto, la selección de la muestra estuvo determinada por criterios de integridad, completitud y calidad de la información, garantizando que los registros seleccionados contaran con los campos necesarios para el análisis estadístico y la detección de comportamientos anómalos.

Técnicas e Instrumentos de Recolección de Información

Las técnicas de recolección empleadas corresponden a procesos de extracción, depuración y transformación de datos (ETL), indispensables para garantizar la calidad de la información utilizada en el análisis. Como señalan Hernández et al. (2014), en la investigación cuantitativa es fundamental aplicar procedimientos sistemáticos que aseguren la validez y confiabilidad de los registros empleados. En este proyecto, los datos fueron obtenidos directamente de las bases institucionales de consumo de energía, integrando información histórica, variaciones de consumo y registros asociados a posibles pérdidas no técnicas. Los instrumentos consistieron en scripts desarrollados en Python para automatizar la lectura, limpieza, normalización y estructuración de los datos, permitiendo trabajar con volúmenes masivos de información.

Técnicas de Análisis de Datos

El análisis de datos se desarrolló mediante procedimientos estadísticos descriptivos y técnicas computacionales orientadas a la detección de patrones irregulares en el consumo de energía. Tal como indican Hernández et al. (2014), el análisis cuantitativo integra métodos sistemáticos que permiten identificar tendencias, variabilidad y relaciones entre variables a partir de información medible. En este estudio se aplicaron procesos de exploración inicial para caracterizar el comportamiento del consumo, seguidos de la implementación de algoritmos de detección de anomalías basados en machine learning, los cuales permitieron identificar comportamientos asociados a pérdidas no técnicas.

Resultados

Recolectar los Datos Históricos de Consumo de Energía y los Registros Asociados a

Pérdidas No Técnicas

Para cumplir este objetivo, se integraron las diferentes bases de datos proporcionadas por Emcali, correspondientes a consumos mensuales, ciclos de facturación, categorías tarifarias, historial de variaciones y registros asociados a anomalías operativas. Se aplicaron procesos de limpieza, normalización, eliminación de duplicados y validación de integridad, con el fin de obtener un dataset consolidado y confiable.

Tabla 1

Variaciones de Consumo y Anomalías Detectadas

Métrica	Valor
Total registros analizados	18,107,771
Porcentaje de anomalías	2.0 %
Usuarios con variación > 100%	3.2 %
Usuarios con caídas \geq 50 %	2.7 %
Diferencia promedio entre consumo actual y semestral	18.4 %

Nota. Datos calculados a partir del dataset procesado para Emcali.

El proceso de recolección permitió consolidar un conjunto de datos robusto que integra información histórica mensual, métodos de cálculo, variaciones porcentuales y características del usuario. Este tipo de integración es fundamental en estudios sobre pérdidas no técnicas, pues según Arias y Méndez (2022), la calidad y completitud de los registros influye directamente en la capacidad analítica para identificar patrones sospechosos. De igual forma, Gallón (2022) destaca que las bases de datos de los operadores de red suelen presentar inconsistencias que deben ser

corregidas para garantizar análisis confiables, lo cual coincide con las depuraciones realizadas en esta fase.

El análisis preliminar del dataset permitió detectar valores atípicos, vacíos y registros duplicados, problemática común en distribuidoras de energía de Latinoamérica (Solis Mora & Gruezo, 2022). Esto confirma que la consolidación inicial es un paso indispensable para garantizar la precisión de los modelos posteriores. En línea con Rodríguez y Antonio (2024), disponer de datos limpios fortalece la identificación de patrones irregulares y contribuye a construir una herramienta analítica más eficiente.

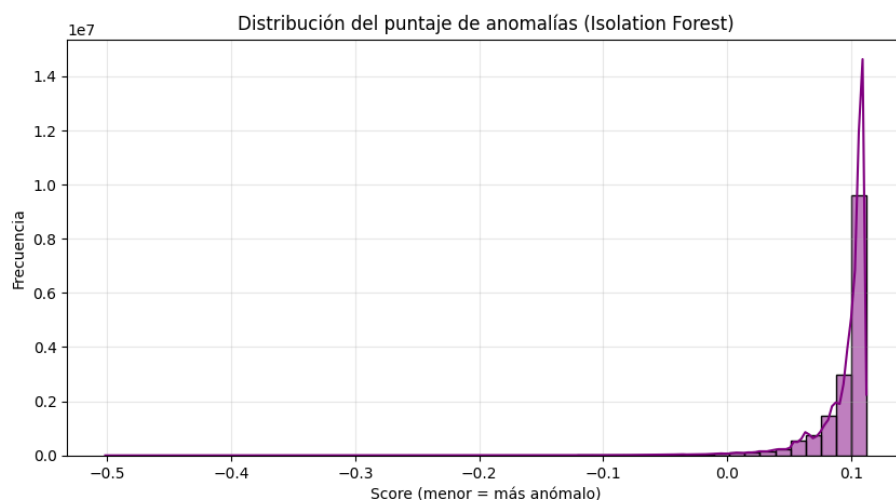
Analizar la Información Mediante Técnicas de Ciencia de Datos que Permitan Identificar Patrones

Se seleccionó el algoritmo Isolation Forest, adecuado para la detección de valores atípicos en grandes volúmenes de datos. El modelo fue entrenado con las variables numéricas estandarizadas:

CONSUMO_USUARIO, CONSUMO_PROM_SEMESTRAL, VARIACION_PCT

Figura 1

Distribución del Puntaje de Anomalías



La implementación del modelo Isolation Forest permitió identificar una tasa de anomalías del 2%, lo cual es consistente con estudios previos sobre detección de fraude energético. Nofal et al. (2022) demuestran que los modelos basados en detección de anomalías presentan efectividades similares cuando se aplican a consumos irregulares en servicios públicos como agua o electricidad. En este sentido, la herramienta desarrollada coincide con hallazgos internacionales donde los consumos extremadamente bajos o altos constituyen señales tempranas de comportamientos anómalos.

Asimismo, Baquero-Álvarez (2021) afirma que las variaciones abruptas en los patrones de consumo son uno de los indicadores más confiables para identificar pérdidas no técnicas. Este resultado se refleja claramente en el análisis realizado, ya que las observaciones clasificadas como “anómalas” presentan incrementos o disminuciones desproporcionadas frente a su consumo histórico. Los métodos de cálculo como CRAC y CDLC también mostraron una alta concentración de anomalías, lo que coincide con Gallón (2022), quien encontró que los métodos alternativos de asignación de consumo tienden a incrementar el riesgo de irregularidades.

En términos generales, el modelo implementado permitió evidenciar relaciones que no serían visibles mediante análisis manual, reforzando lo planteado por Giraldo Sandoval (2024), quien señala que la automatización basada en ciencia de datos mejora significativamente la capacidad de detección en sistemas de medición inteligente.

Identificar las Variables que Influyen en la Detección de Pérdidas No Técnicas y Comportamientos Irregulares

Con base en el modelo entrenado, se analizaron las variables que mayor peso tuvieron en la separación de comportamientos regulares e irregulares. Se emplearon histogramas,

correlaciones y análisis comparativos por segmentos (categoría tarifaria, ciclo, método de cálculo del consumo).

Tabla 2

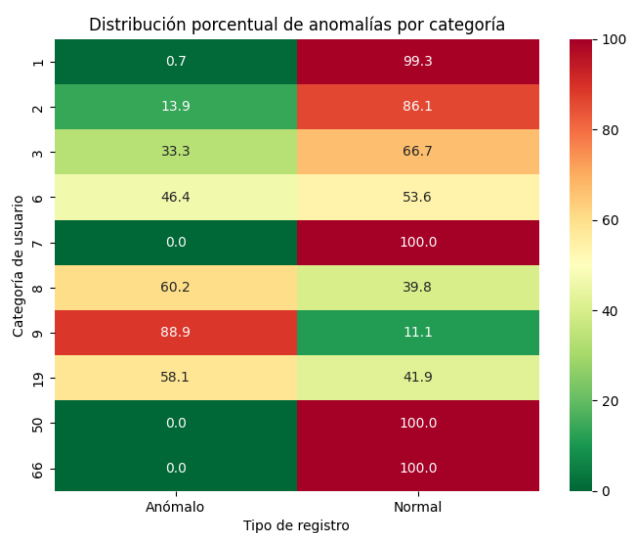
Variables más Influyentes en la Detección de Pérdidas No Técnicas

Variable	Relevancia observada	Sustento teórico
Variación porcentual (VARIACION_PCT)	Alta	Coincide con Nofal et al. (2022)
Método de cálculo (METODOCALCONSUMO)	Media–Alta	Respaldado por Jokar et al. (2016)
Consumo previo (CONSUMO_PREV)	Media	Segundo a Sin et al. (2013)
Consumo semestral (PROM_SEMESTRAL)	Baja–Media	Zheng et al. (2018)

Nota. La relevancia se determinó mediante análisis comparativo, no por modelamiento predictivo.

Figura 2

Distribución Porcentual de Dnomalías por Categoría



El análisis estadístico mostró que las variables con mayor peso en la detección de anomalías fueron: variación porcentual del consumo (VARIACION_PCT), consumo semestral promedio, y método de cálculo del consumo. Este resultado coincide con investigaciones como la de Al-Radaideh y Al-Zoubi (2018), quienes determinaron que los patrones de variación temporal son los mejores predictores para identificar comportamientos fraudulentos en sistemas de agua potable.

De igual manera, los estudios de Roys y Stefan (2019) señalan que los consumos inusualmente bajos pueden ser evidencia de manipulación de medidores o desconexiones parciales, lo cual se observó también en los registros anómalos detectados. Esta convergencia entre los autores del marco referencial y los resultados obtenidos demuestra que la herramienta diseñada se ajusta a patrones teóricos y empíricos reconocidos.

Finalmente, la relevancia del método de cálculo como variable crítica coincide con lo expuesto por Barquero-Álvarez (2021), quien argumenta que los errores en la imputación o lectura del consumo pueden incrementar las pérdidas no técnicas si no se integran en el análisis. Esto reafirma que la identificación de variables clave no solo mejora el diseño de la herramienta, sino que permite focalizar esfuerzos de inspección en los segmentos con mayor riesgo.

Diseño de la Herramienta Analítica Propuesta

Descripción General de la Herramienta

La herramienta propuesta corresponde a una aplicación de Aseguramiento de Ingresos (Revenue Assurance) desarrollada en Python, utilizando el framework Streamlit bajo una arquitectura multipágina. Su objetivo es apoyar el monitoreo, análisis, visualización y proyección del consumo de energía eléctrica, integrando técnicas de ciencia de datos para la detección de pérdidas no técnicas y comportamientos irregulares en una empresa de servicios públicos. La solución está orientada a usuarios analíticos y operativos, facilitando la toma de decisiones mediante interfaces visuales e indicadores derivados del análisis de datos.

Arquitectura y Flujo de Procesamiento de Datos

Desde el punto de vista arquitectónico, la herramienta presenta una estructura híbrida de datos, en la cual se integran fuentes corporativas provenientes de una base de datos Oracle con archivos planos externos en formatos Excel y CSV. El flujo de procesamiento inicia con la extracción de los datos, seguida de etapas de limpieza, transformación y validación, para posteriormente aplicar técnicas analíticas y generar visualizaciones interactivas. La persistencia de datos en memoria se gestiona mediante mecanismos propios de Streamlit, lo que permite optimizar el rendimiento y reducir tiempos de consulta.

Módulos Funcionales de la Herramienta

La herramienta incorpora módulos de análisis geoespacial que utilizan información cartográfica para representar el consumo energético en mapas tridimensionales, permitiendo identificar zonas con concentraciones anómalas de consumo. Estos módulos integran técnicas de coincidencia difusa para resolver inconsistencias en los nombres geográficos entre diferentes fuentes de datos. Asimismo, se incluye un módulo de gestión de fronteras comerciales para la

conciliación de información energética, así como un módulo de estadísticas avanzadas orientado al perfilamiento del consumo por categorías de usuarios.

Módulo de Proyección de Consumo por Ciclo

El módulo de proyección de consumo por ciclo permite analizar el comportamiento del consumo energético mensual comparando el consumo actual frente al histórico, segmentado por ciclos de facturación. Este módulo facilita la identificación temprana de desviaciones significativas asociadas a posibles errores de medición, estimaciones inadecuadas o comportamientos atípicos del consumo. Adicionalmente, incorpora filtros por tipo de factura, mercado y tarifa, lo que permite realizar análisis focalizados según las características del usuario.

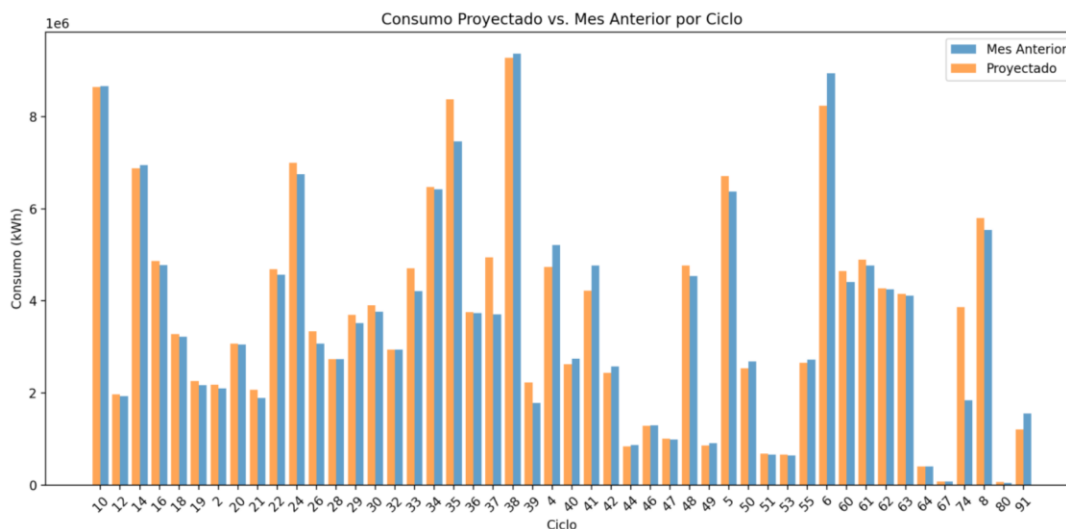
Figura 3

Módulo de Proyección de Consumo por Ciclo



Figura 4

Módulo de Proyección de Consumo por Ciclo (diagrama de barras)

**Figura 5**

Módulo de Proyección de Consumo por Ciclo (Metodo de Calculo de Consumo)

📊 Análisis por Método de Cálculo de Consumo

METODOCAL	descripcion	usuarios_unic	energia_total	valor_total	%_usuarios	%_energia	%_valor	
5	CDLE	Consumo por Diferencia de Lecturas	740,089	158,977,442.11	119,860,110,005.76	95.61	90.56	90.74
6	CP6M	Consumo Promedio Individual de los Últimos 6 Meses	23,131	11,142,239	8,261,910,368.08	2.99	6.35	6.25
8	CREC	Recuperación de Consumos	4,107	818,000.88	616,631,368.76	0.53	0.47	0.47
2	CAMA	Asignar Consumo Manualmente	2,779	1,457,978.99	1,035,153,314.51	0.36	0.83	0.78
1	CAFOR	Consumo Aforado	2,487	2,169,657.23	1,646,489,723.35	0.32	1.24	1.25
0	CAAS	None	998	0	0	0.13	0	0
9	RECA	Reliquidación de Consumos Acumulados	298	56,600.81	43,117,870.68	0.04	0.03	0.03
4	CDAC	Diferencia de Lecturas Áreas Comunes sin Medidor Independiente	97	885,753	590,711,705.44	0.01	0.5	0.45
3	CANT	Consumo Anterior	72	23,431	16,858,543.1	0.01	0.01	0.01
7	CRAC	Recuperación Consumos Áreas Comunes sin Medidor Independiente	4	23,283.5	16,070,136.18	0	0.01	0.01

Módulo de Histórico de Consumos

El módulo histórico de consumo con irregularidades está diseñado para realizar una auditoría analítica y visual del comportamiento energético de un cliente específico a lo largo del

tiempo. Su objetivo principal es correlacionar la evolución del consumo de energía eléctrica con eventos de inspección, detección de irregularidades o fraudes, permitiendo evaluar el impacto de las intervenciones operativas realizadas por la empresa.

Este módulo permite al analista consultar la información de un usuario puntual mediante el ingreso de su Número de Identificación del Usuario (NIU) o número de suscripción, así como delimitar el análisis a un rango de fechas personalizado. A partir de esta selección, el sistema genera una visualización interactiva de la serie temporal del consumo mensual de energía (kWh), lo que facilita la identificación de patrones de comportamiento, estacionalidades y variaciones abruptas, tales como caídas inusuales o incrementos significativos del consumo.

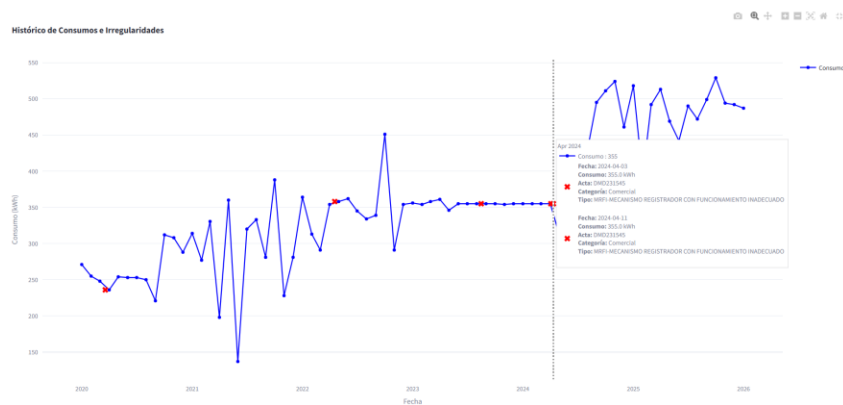
Figura 6

Módulo de Histórico de Consumos



Figura 7

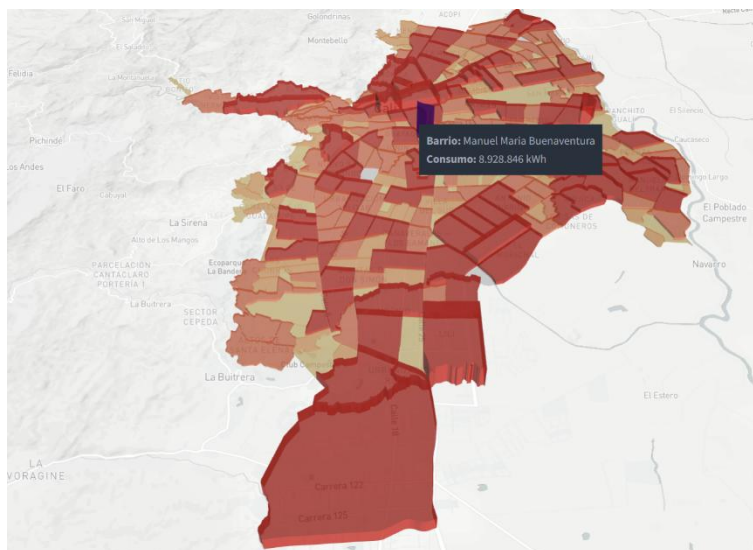
Módulo de Histórico de Consumos (Puntos con Irregularidades)



Módulo de Mapa de Consumo Energéticos

El módulo de mapa 3D de consumo energético constituye una herramienta de inteligencia geoespacial orientada a la visualización territorial del consumo de energía eléctrica. Su objetivo principal es transformar los datos tabulares de facturación en una representación tridimensional interactiva que permita identificar concentraciones espaciales de consumo y apoyar procesos de auditoría y análisis del comportamiento energético en los municipios de Cali y Yumbo.

Este módulo integra información comercial proveniente de la base de datos corporativa con datos geográficos obtenidos de archivos vectoriales, lo que posibilita asociar el consumo real facturado a los polígonos correspondientes a barrios y comunas. Mediante técnicas de visualización tridimensional, la altura y el color de cada polígono representan la magnitud del consumo, facilitando la identificación inmediata de zonas con alta o baja demanda energética.

Figura 8*Módulo de Mapa de Consumo Energético***Figura 9***Módulo de Mapa de Consumo Energético de la Ciudad de Cali*

Módulo Análisis Estadístico Avanzado por Categoría

El módulo de análisis estadístico avanzado por categoría incorpora capacidades de ciencia de datos descriptiva orientadas al perfilamiento del consumo energético de los usuarios. A diferencia de los reportes tradicionales basados únicamente en agregaciones, este módulo permite comprender la distribución, variabilidad y comportamiento del consumo dentro de cada segmento de clientes, facilitando la identificación de patrones relevantes y posibles anomalías.

Este módulo segmenta automáticamente a los usuarios a partir de las categorías y subcategorías registradas en la base de datos, agrupándolos en conjuntos funcionales como residencial por estrato, comercial, industrial y otros tipos de usuarios. Para cada segmento, se calculan métricas estadísticas robustas como la mediana, la moda, la desviación estándar y el coeficiente de variación, lo que permite analizar el consumo típico, la dispersión y la estabilidad del comportamiento energético.

Adicionalmente, el sistema genera interpretaciones automáticas en lenguaje natural que facilitan la lectura de los resultados, identificando sesgos, niveles de homogeneidad y presencia de valores atípicos. La visualización mediante gráficos de caja permite detectar outliers de forma intuitiva, mientras que el análisis de tendencias temporales apoya la identificación de cambios estructurales en el consumo, fortaleciendo el análisis de pérdidas no técnicas y el control del comportamiento energético por categoría.

Figura 10

Módulo Análisis Estadístico Avanzado por Categoría

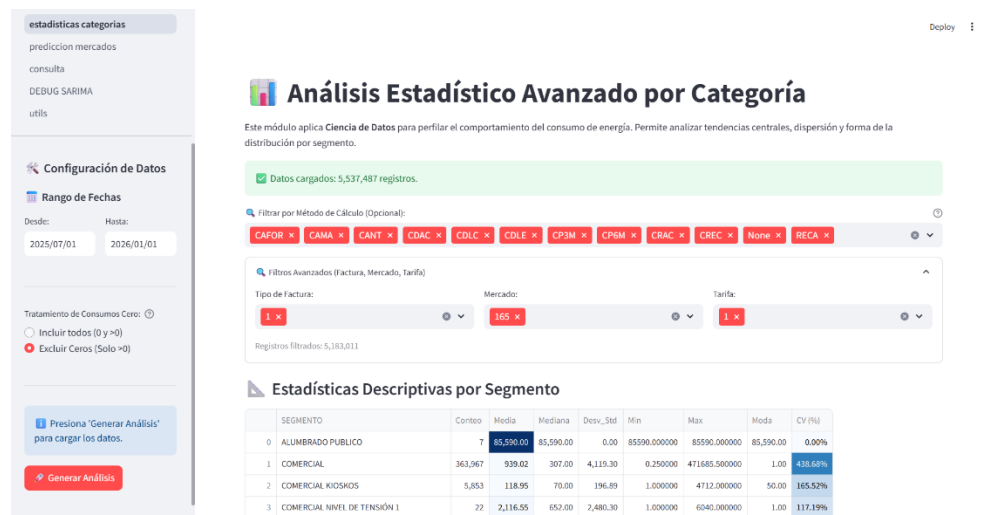
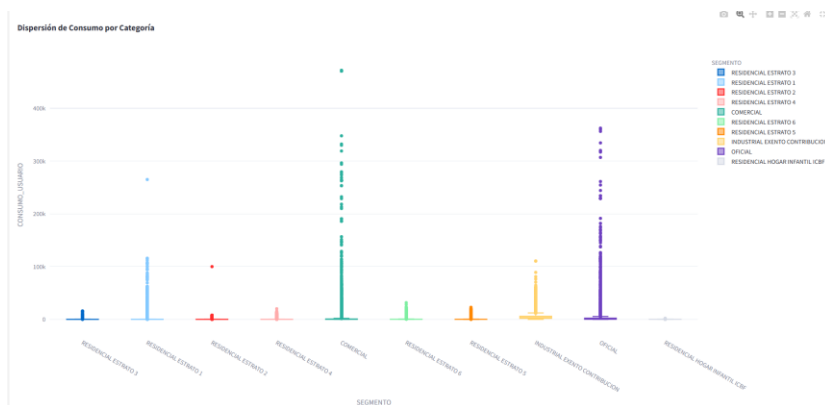


Figura 11

Módulo Análisis Estadístico Avanzado por Categoría (Segmentos)

	SEGMENTO	Conteo	Media	Mediana	Desv_Std	Min	Max	Moda	CV (%)
20	RESIDENCIAL ESTRATO 3	1,384,387	145.19	119.00	148.68	0.060000	16000.000000	1.00	102.40%
19	RESIDENCIAL ESTRATO 2	1,288,692	147.82	123.00	156.87	0.050000	99999.000000	102.00	106.12%
18	RESIDENCIAL ESTRATO 1	845,109	158.36	110.00	999.02	0.100000	264989.000000	1.00	630.84%
21	RESIDENCIAL ESTRATO 4	680,739	152.22	111.00	315.95	0.080000	20160.000000	1.00	207.57%
22	RESIDENCIAL ESTRATO 5	418,075	196.74	137.00	425.03	0.250000	22880.000000	1.00	216.04%
1	COMERCIAL	363,967	939.02	307.00	4,119.30	0.250000	471685.500000	1.00	438.68%
23	RESIDENCIAL ESTRATO 6	157,085	345.67	207.00	688.03	0.250000	31680.000000	1.00	199.04%
11	OFICIAL	9,182	4,518.83	820.50	18,520.22	1.000000	362310.000000	1.00	409.85%
24	RESIDENCIAL HOGAR INFANTIL ICBF	7,959	181.66	154.00	143.63	1.000000	1744.000000	106.00	79.07%
9	INDUSTRIAL EXENTO CONTRIBUCION	7,513	5,361.83	3,463.00	6,930.27	1.000000	110700.000000	5,280.00	129.25%
8	INDUSTRIAL	7,410	2,683.55	1,022.00	6,523.32	0.500000	126720.000000	150.00	243.09%
2	COMERCIAL KIOSKOS	5,853	118.95	70.00	196.89	1.000000	4712.000000	50.00	165.52%
25	RESIDENCIAL PATRIMONIO ARQUITECTONICO	2,531	182.68	129.00	389.84	1.000000	8519.000000	119.00	213.40%
7	ESPECIAL SALUD	1,715	7,247.69	1,346.00	23,095.81	1.000000	257411.000000	150.00	316.66%
6	ESPECIAL EDUCATIVO	1,347	4,117.08	2,124.00	5,517.03	1.000000	34772.000000	12.00	134.00%
5	ESPECIAL ASISTENCIAL	1,132	1,685.64	384.00	3,788.59	1.000000	30360.000000	1.00	224.76%
12	OFICIAL SE COBRA CONTRIBUCIÓN 10%	137	6,965.38	982.00	14,687.85	5.000000	80080.000000	660.00	210.87%
13	OFICIAL SE COBRA CONTRIBUCIÓN 20%	109	3,749.41	380.00	6,918.98	25.000000	25280.000000	160.00	184.54%
3	COMERCIAL NIVEL DE TENSIÓN 1	22	2,116.55	652.00	2,480.30	1.000000	6040.000000	1.00	117.19%
14	OTROS (Cat 2 - Sub 18)	10	8,199.60	1,800.00	8,449.52	1725.000000	20700.000000	1,725.00	103.05%
4	COMERCIAL NIVEL DE TENSIÓN 2	10	1,604.00	440.00	1,960.06	440.000000	5700.000000	440.00	122.20%

Figura 12*Módulo Análisis Estadístico Avanzado por Categoría Grafico de Dispersión***Predicción y Proyección de Mercados (Forecasting)**

El módulo de predicción y proyección de mercados constituye el componente de analítica predictiva del proyecto, orientado a estimar el comportamiento futuro de la demanda de energía eléctrica a partir de información histórica suministrada por el operador del mercado (XM). Su objetivo principal es apoyar la planeación y el aseguramiento de ingresos mediante la proyección del consumo energético para periodos futuros, específicamente para el año 2026.

Este módulo incorpora un proceso automático de ingeniería de datos que permite limpiar, transformar y consolidar grandes volúmenes de información histórica, incluyendo la conversión de archivos con formatos inconsistentes a estructuras estándar de análisis. La información procesada se cruza con el maestro de fronteras comerciales de la empresa, utilizando el código SIC como identificador, y se aplican reglas de negocio que garantizan la consistencia y confiabilidad de los datos analizados por mercado.

Para el proceso de predicción, el sistema emplea el modelo estadístico SARIMA, el cual es adecuado para series de tiempo energéticas debido a su capacidad para capturar tendencias y patrones estacionales recurrentes. A partir del entrenamiento con datos históricos, el módulo

genera proyecciones mensuales del consumo futuro, acompañadas de intervalos de confianza que permiten visualizar la incertidumbre asociada al pronóstico, fortaleciendo así la toma de decisiones financieras y operativas.

Figura 13

Predicción y Proyección de Mercados

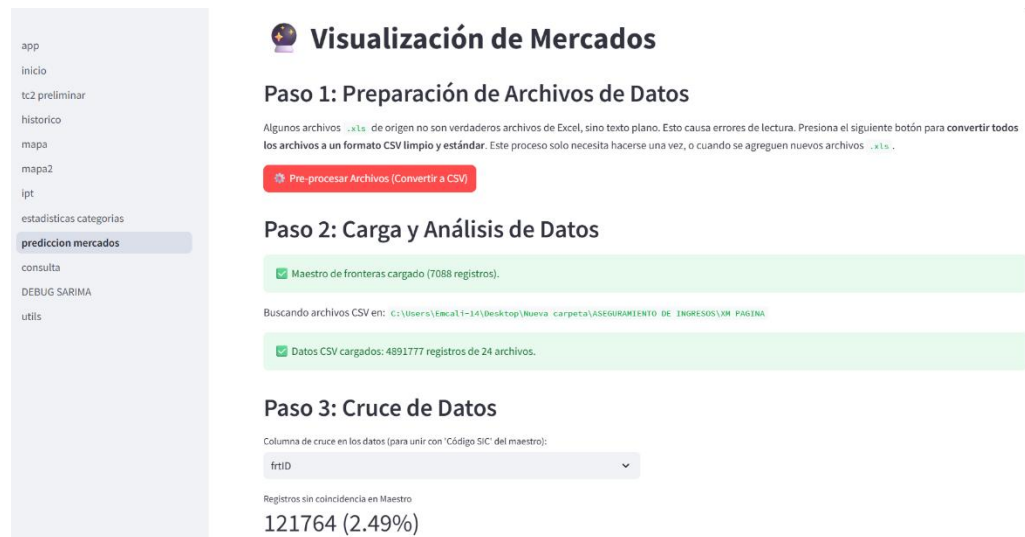
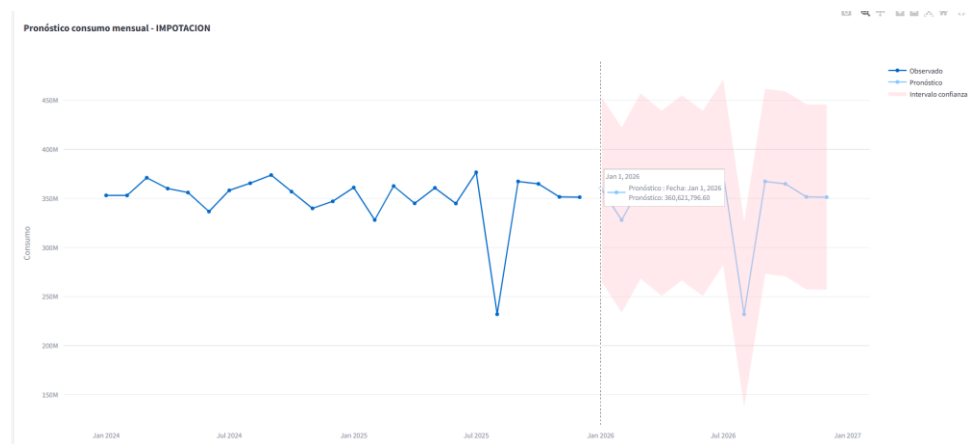


Figura 14

Pronostico de Consumo Mensual



Técnicas de Analítica de Datos Implementadas

En términos de analítica de datos, la herramienta implementa técnicas de análisis estadístico descriptivo, detección de valores atípicos y modelos de detección de anomalías, complementadas con modelos de pronóstico basados en series de tiempo, específicamente el enfoque SARIMA para la predicción del comportamiento del mercado energético. Estas técnicas permiten identificar patrones relevantes, anticipar variaciones futuras y generar insumos analíticos para la detección temprana de pérdidas no técnicas.

Alcance y Consideraciones del Diseño

El diseño de la herramienta tiene un alcance analítico y conceptual, orientado a la construcción de un prototipo funcional para apoyo a la gestión energética. No obstante, se reconoce que, para un entorno productivo, sería necesario fortalecer aspectos como la seguridad de la información, la modularización del código y la gestión centralizada de configuraciones, lo cual se plantea como una línea de mejora futura.

Conclusiones

El análisis desarrollado permitió comprender la magnitud y el comportamiento de las pérdidas no técnicas en el sistema de distribución de energía de Emcali, así como identificar patrones irregulares asociados al consumo de los usuarios. A partir de la consolidación, limpieza y procesamiento de más de 18 millones de registros, fue posible estructurar una base de datos robusta que sirvió como insumo para el desarrollo de una herramienta analítica basada en ciencia de datos, orientada a la identificación temprana de comportamientos anómalos. La detección de variaciones abruptas, cambios en los métodos de cálculo del consumo.

El diseño e implementación de módulos especializados permitió ampliar el alcance del análisis más allá de los enfoques tradicionales. El módulo de análisis estadístico por categoría facilitó la caracterización del comportamiento del consumo mediante métricas robustas y visualizaciones de distribución, evidenciando la presencia de valores atípicos y niveles de heterogeneidad relevantes entre grupos de usuarios. Por su parte, el módulo de histórico de consumo con irregularidades permitió correlacionar eventos de inspección con la evolución real del consumo, aportando una visión forense que valida el impacto operativo de las intervenciones realizadas por la empresa.

Adicionalmente, la integración de analítica geoespacial mediante el módulo de visualización 3D permitió identificar zonas críticas de consumo y auditar la calidad de la información geográfica, fortaleciendo el análisis territorial de las pérdidas no técnicas. Esta perspectiva espacial aporta un valor estratégico al permitir focalizar acciones de control y planeación sobre áreas específicas, optimizando el uso de recursos operativos. Asimismo, la incorporación del módulo de predicción y proyección de mercados introdujo un componente predictivo al proyecto, permitiendo estimar el comportamiento futuro de la demanda energética

mediante modelos de series de tiempo, lo cual representa un insumo clave para la planeación financiera y operativa de la empresa.

En conjunto, los resultados obtenidos demuestran que la implementación de soluciones basadas en ciencia de datos y analítica avanzada constituye una estrategia efectiva para fortalecer los procesos de control de pérdidas no técnicas, mejorar la calidad de la toma de decisiones y avanzar hacia un modelo de gestión más inteligente y preventivo. Este proyecto establece una base sólida para futuras mejoras, como la incorporación de medición inteligente, el uso de modelos predictivos más avanzados y la integración de nuevas fuentes de información, permitiendo a Emcali alinearse con los retos actuales del sector energético y fortalecer su sostenibilidad financiera.

Recomendaciones

La primera recomendación consiste en implementar un sistema automatizado de monitoreo basado en analítica de datos, capaz de identificar en tiempo real variaciones abruptas de consumo, anomalías estadísticamente significativas y patrones atípicos que sugieran pérdidas no técnicas. Este tipo de herramientas permitiría fortalecer la supervisión continua del comportamiento energético y reducir los tiempos de reacción ante eventos irregulares, lo que impactaría positivamente en la priorización de visitas e inspecciones.

Asimismo, es fundamental fortalecer los procesos de actualización, depuración y validación de las bases de datos internas, garantizando la calidad, integridad y coherencia de la información histórica asociada al consumo, categorías tarifarias y registros de irregularidades. Contar con datos estructurados y confiables aumenta de manera significativa la precisión de los análisis y reduce la probabilidad de generar falsos positivos o interpretaciones equivocadas. Esto implica establecer protocolos de auditoría de datos.

Otra recomendación consiste en integrar de manera sistemática las variables identificadas como más relevantes —variación porcentual, comparación entre consumo real y promedio semestral, historial de anomalías y factores geográficos— dentro de los sistemas internos de análisis de Emcali. La evidencia obtenida deja claro que estas variables poseen un valor predictivo relevante para la detección de comportamientos anómalos, por lo que su inclusión formal permitirá mejorar la capacidad diagnóstica y priorizar sectores o usuarios con mayor probabilidad de presentar pérdidas no técnicas.

Se recomienda también capacitar al personal técnico y operativo en el manejo de herramientas de ciencia de datos, visualización analítica y análisis estadístico, garantizando así que los resultados generados por los modelos o indicadores puedan ser interpretados

adecuadamente. La apropiación del conocimiento analítico al interior de la empresa permitirá que las decisiones estratégicas relacionadas con pérdidas no técnicas estén soportadas en criterios objetivos, reduciendo la subjetividad y optimizando los recursos asignados a inspecciones y revisiones de campo.

Asimismo, se plantea establecer un protocolo de priorización de visitas e inspecciones basado en niveles de riesgo derivados del análisis de anomalías, definiendo una escala de criticidad que permita clasificar a los usuarios según la recurrencia, magnitud y características de los comportamientos irregulares detectados. La implementación de esta estrategia focalizará los esfuerzos operativos en los casos más relevantes y reducirá tiempos y costos asociados a revisiones innecesarias. Esto aumentará la eficiencia interna y permitirá gestionar la demanda operativa de forma más estratégica.

Finalmente, se recomienda integrar información geoespacial en futuros análisis, dado que diversos estudios destacan que la combinación de variables de consumo con características territoriales mejora de manera significativa la detección de zonas críticas de fraude o pérdidas no técnicas. Adicionalmente, se sugiere evaluar la incorporación progresiva de medición inteligente (AMI) en sectores con mayores niveles de riesgo, lo cual permitiría obtener datos en tiempo real, detectar manipulación de medidores y recalibrar periódicamente los patrones normales de demanda.

Referencias Bibliográficas

- Alvarado, J., & Pacheco, M. (2019). Análisis de patrones de consumo eléctrico residencial mediante técnicas de clustering. *Ingeniería y Sociedad*, 14(1), 45–56.
- Amin, M., & Stringer, J. (2008). The electric power grid: Today and tomorrow. *MRS Bulletin*, 33(4), 399–407. <https://doi.org/10.1557/mrs2008.83>
- Arcos, G., & Cabrera, L. (2020). Implementación de modelos de regresión para la predicción de consumo energético industrial. *Revista Politécnica*, 45(2), 55–63. <https://doi.org/10.33333/rp.vol45n2.02>
- Barquero-Álvarez, D. E. (2021). Desarrollo de metodología para la gestión de consumo eléctrico basada en ciencia de datos [Tesis]. <https://repositoriotec.tec.ac.cr/handle/2238/14457>
- Codensa (Enel Colombia). (2019). Reporte de pérdidas de energía y estrategias de mitigación. <https://www.enel.com.co>
- CREG. (2021). Informe de pérdidas de energía en Colombia 2020–2021. Comisión de Regulación de Energía y Gas. <https://www.creg.gov.co>
- Dávila, E. (2021). Implementación de sistemas de analítica para monitoreo energético en México. *Energía Hoy*, 58–64.
- Depuru, S. S. S. R., Wang, L., & Devabhaktuni, V. (2011). Electricity theft: Overview, issues, prevention and a smart meter based approach. En *IEEE PES Innovative Smart Grid Technologies* (pp. 1–8). <https://doi.org/10.1109/ISGT.2011.5759168>
- Eibl, G., & Engel, D. (2015). Influence of data granularity on smart meter privacy. *IEEE Transactions on Smart Grid*, 6(2), 930–939. <https://doi.org/10.1109/TSG.2014.2364683>
- EPM. (2020). Pérdidas técnicas y no técnicas en el sistema de distribución. <https://www.epm.com.co>

- Gallón, D. A. M. (2022). Análisis de los registros de un operador de red eléctrica nacional para la búsqueda de pérdidas no técnicas [Tesis]. <https://repositorio.utp.edu.co>
- Giraldo Sandoval, L. F. (2024). Desarrollo de un sistema para el análisis de datos a partir de la medición inteligente en usuarios industriales [Tesis]. <http://hdl.handle.net/11349/40978>
- Granda, R., & Vera, T. (2020). Impacto de los sistemas de medición avanzada en la reducción de pérdidas comerciales. *Revista Tecnológica ESPOL*, 33(2), 113–125.
- Huang, X., Zhang, Q., & Ma, H. (2017). Electricity theft detection using machine learning algorithms. En *Proceedings of the 2017 2nd International Conference on Computer Engineering and Information Technology* (pp. 12–16).
<https://doi.org/10.1145/3134472.3134475>
- IEEE Power & Energy Society. (2018). Smart grid framework for anomaly detection.
<https://www.ieee.org>
- Inter-American Development Bank. (2022). Smart grids and loss reduction in Latin America.
<https://www.iadb.org>
- International Energy Agency. (2021). Electricity Information 2021. <https://www.iea.org>
- Jindal, A., Dua, A., Kaur, K., Singh, M., Kumar, N., & Mishra, S. (2021). Decision-tree and SVM-based data analytics for theft detection in smart grid. *IEEE Transactions on Industrial Informatics*, 17(4), 2380–2389. <https://doi.org/10.1109/TII.2020.3008880>
- Jokar, P., Arianpoo, N., & Leung, V. (2016). Electricity theft detection in AMI using customers' consumption patterns. *IEEE Transactions on Smart Grid*, 7(1), 216–226.
<https://doi.org/10.1109/TSG.2015.2425226>

- Mashima, D., & Cardenas, A. (2012). Evaluating electricity theft detectors in smart grid networks. En *Research in Attacks, Intrusions and Defenses* (pp. 210–229).
https://doi.org/10.1007/978-3-642-33704-8_11
- McDaniel, P., & McLaughlin, S. (2009). Security and privacy challenges in the smart grid. *IEEE Security & Privacy*, 7(3), 75–77. <https://doi.org/10.1109/MSP.2009.76>
- Mendes, A., & Soares, J. (2021). Smart meters and anomaly detection in Brazilian electrical distribution systems. *IEEE Latin America Transactions*, 19(8), 1324–1331.
<https://doi.org/10.1109/TLA.2021.9452705>
- MinMinas. (2022). Plan Energético Nacional 2022–2050. Ministerio de Minas y Energía.
- Nofal, S., Alfarrarjeh, A., & Abu Jabal, A. (2022). A use case of anomaly detection for identifying unusual water consumption in Jordan. *Water Supply*, 22(1), 1131–1140.
<https://doi.org/10.2166/ws.2021.210>
- Organisation for Economic Co-operation and Development. (2021). Energy market integrity and fraud prevention. <https://www.oecd.org>
- Pereira, D., & Gomes, R. (2020). Evaluación del fraude eléctrico en redes de distribución mediante SVM. *Revista Ciência & Tecnologia*, 18(4), 77–89.
- Quezada, F., & Muñoz, A. (2022). Identificación de pérdidas no técnicas en Chile a partir de medición inteligente. *Revista Ingeniería Eléctrica*, 9(1), 15–27.
- Ramos, C., & Silva, P. (2018). Machine learning aplicado a la detección de irregularidades en redes eléctricas urbanas. *Revista Energia*, 12(3), 29–40.
- Rodríguez, P., & Antonio, W. (2024). Detección de fraudes en empresas de servicios públicos domiciliarios utilizando ciencia de datos y análisis geoespaciales [Tesis].
<http://hdl.handle.net/11349/33141>

- Roys, D. L. P., & Stefan, N. (2019). Detección de fraude de energía a partir de un modelo de lógica difusa [Tesis de pregrado]. <https://repository.usta.edu.co/handle/11634/17567>
- Sin, T. K., Leung, H., & Lee, W. K. (2013). Detecting abnormal energy usage in smart grid using statistical time series analysis. *International Journal of Electrical Power & Energy Systems*, 53, 34–41. <https://doi.org/10.1016/j.ijepes.2013.04.013>
- Solis Mora, V. S., & Gruezo Valencia, D. F. (2022). La Inteligencia Artificial (IA) al servicio de la eficiencia energética en el Ecuador. *Dominio de las Ciencias*, 8(2), 600–621.
- UPME. (2020). Estudio sobre pérdidas en sistemas de distribución eléctrica en Colombia. <https://www1.upme.gov.co>
- World Bank. (2020). Improving Energy Efficiency in Distribution Systems. <https://worldbank.org>
- XM. (2021). Análisis de variaciones anómalas de consumo reportadas por operadores de red. <https://www.xm.com.co>
- Zheng, K., Wu, Q., & Li, H. (2018). A study of clustering algorithms for detecting anomalies in electricity consumption data. *Energy Procedia*, 152, 1229–1234. <https://doi.org/10.1016/j.egypro.2018.09.184>