

**Estudio de la aplicación de técnicas de machine learning para la optimización y gestión de pérdidas no técnicas en el sector eléctrico**

Ana Daniela Caicedo Salazar

Asesor

Elvis Orlando Rodríguez Contreras

Universidad Nacional Abierta y a Distancia UNAD

Escuela de Ciencias Básicas, Tecnología e Ingeniería ECBTI

Especialización en Ciencia de datos y Analítica

2025

---

Elvis Orlando Rodriguez Contreras

---

Jurado

---

Jurado

## **Dedicatoria**

Este trabajo va dedicado a mis padres y hermanos.

## **Agradecimientos**

A Dios y a la virgen por permitirme culminar esta etapa profesional, a mi familia por su apoyo incondicional

Este trabajo de grado ha sido el resultado de mi tiempo y dedicación a lo largo de todo el proceso de formación en la especialidad.

Agradezco sinceramente a todo el personal administrativo de la Escuela de Ciencias Básicas, Tecnología e Ingeniería – ECBTI de la UNAD (Universidad Nacional Abierta y a Distancia), así como al cuerpo docente de la especialización en Ciencia de Datos y Analítica, por su apoyo, seguimiento y valiosas enseñanzas.

Mi agradecimiento especial al director de este trabajo de grado, el ingeniero Elvis Orlando Rodriguez, por su constante apoyo, enseñanzas, acompañamiento y, sobre todo, por ser un excelente líder durante todo el proceso de desarrollo del presente trabajo.

## Resumen

En el presente documento, se analizan los diferentes proyectos nacionales como internacionales, desarrollados en el sector eléctrico, examinando la aplicación de las diferentes técnicas de *machine learning* para la optimización y gestión de las PNT ( Pérdidas No Técnicas) de energía en usuarios finales. Su objetivo principal es analizar proyectos nacionales e internacionales que han implementado estas técnicas para detectar condiciones anómalas asociadas a PNT, como el hurto de energía, fallas en los sistemas de medición y errores en la facturación.

El enfoque metodológico incluye el uso de métodos estadísticos, métricas, la identificación de variables relevantes y la aplicación de técnicas de *machine learning*, con el propósito de documentar, analizar y evaluar las técnicas utilizadas en otros países y su aplicabilidad en el contexto colombiano. A partir de este análisis, se busca identificar las estrategias más efectivas para la detección y control de pérdidas no técnicas, contribuyendo a la optimización de su gestión y facilitando la toma de decisiones en el sector eléctrico.

Este estudio también permite proponer estrategias y recomendaciones basadas en los hallazgos obtenidos, fortaleciendo las acciones preventivas y correctivas que ayuden a reducir las pérdidas no técnicas y mejorar la eficiencia del sistema eléctrico en Colombia.

**Palabras clave:** anomalía, consumo, detección, fraude, *machine learning*, pérdidas de energía.

## Abstract

This document analyzes various national and international projects developed in the electricity sector, examining the application of different machine learning techniques for the optimization and management of non-technical energy losses among end-users. Its main objective is to analyze national and international projects that have implemented these techniques to detect anomalous conditions associated with non-technical energy losses, such as energy theft, measurement system failures, and billing errors.

The methodological approach includes the use of statistical methods, metrics, the identification of relevant variables, and the application of machine learning techniques. The purpose of this study is to document, analyze, and evaluate the techniques used in other countries and their applicability in the Colombian context. Based on this analysis, the study seeks to identify the most effective strategies for the detection and control of non-technical losses, contributing to the optimization of their management and facilitating decision-making in the electricity sector. This study also allows for the proposal of strategies and recommendations based on the findings, strengthening preventive and corrective actions that help reduce non-technical losses and improve the efficiency of the Colombian electrical system.

**Keywords:** anomaly, consumption, detection, energy losses, fraud, machine learning.

## Contenido

Introducción .....	10
Descripción del Problema .....	11
Planteamiento del Problema.....	12
Sistematización del Problema .....	13
Justificación.....	16
Objetivos .....	17
Objetivo General.....	17
Objetivos Específicos .....	17
Marco de Referencia.....	18
Estado del Arte.....	18
Marco Contextual .....	20
Marco Teórico.....	21
Marco Conceptual.....	24
Marco Normativo.....	26
Metodología .....	32
Conclusiones .....	69
Recomendaciones .....	72
Referencias Bibliográficas .....	75

## Lista de Tablas

<b>Tabla 1</b> <i>Estudios Previos Sobre Aplicación de Técnicas de Machine Learning</i> .....	32
<b>Tabla 2</b> <i>Identificación de Variables Sobre Estudios Realizados</i> .....	37
<b>Tabla 3</b> <i>Comparación Técnicas de Machine Learning</i> .....	43
<b>Tabla 4</b> <i>Definición de Métricas Proyecto1</i> .....	58
<b>Tabla 5</b> <i>Resultados Obtenidos en Métricas de Evaluación Proyecto1</i> .....	59
<b>Tabla 6</b> <i>Resultados Matriz de Confusión</i> .....	60
<b>Tabla 7</b> <i>Resultados de las Métricas del Proyecto2</i> .....	61
<b>Tabla 8</b> <i>Definición Métricas Proyecto2</i> .....	62
<b>Tabla 9</b> <i>Resultados Métricas Proyecto3</i> .....	63

## Lista de Figuras

<b>Figura 1</b> <i>Representación de Campo de la Inteligencia Artificial (IA)</i> .....	52
<b>Figura 2</b> <i>Diagrama Metodologia CRISP-DM</i> .....	55

## Introducción

En el sector eléctrico colombiano, las pérdidas no técnicas (PNT) en usuarios representan un desafío significativo para las empresas de distribución y comercialización de energía. Estas pérdidas, derivadas de fraudes eléctricos, conexiones ilegales, errores en la medición y fallas en los sistemas de facturación, generan impactos económicos negativos y comprometen la sostenibilidad del sistema eléctrico.

La identificación y mitigación de estas pérdidas es una tarea compleja debido a la naturaleza dinámica del consumo eléctrico y la diversidad de factores como ambientales de acuerdo a la zona o región de localización de los usuarios, que influyen en su detección.

En este contexto, con el uso de técnicas de *machine learning* se busca encontrar una solución innovadora que contribuya a mejorar el control y análisis de las pérdidas no técnicas en el sector eléctrico Colombiano. Algoritmos de aprendizaje supervisado y no supervisado permiten analizar grandes volúmenes de datos históricos, identificar patrones de consumo irregulares y predecir posibles casos de fraude o anomalías en el suministro de energía, a fin de optimizar la detección de pérdidas, permitiendo a las empresas tomar decisiones más informadas y eficientes.

El siguiente trabajo tiene como objetivo analizar proyectos internacionales en empresas del sector eléctrico enfocados en la aplicación de diferentes técnicas de *machine learning* para detección de PNT en usuarios finales, a fin de tomarlos como referentes para optimizar y gestionar las pérdidas no técnicas en usuarios del sector eléctrico colombiano.

## **Descripción del Problema**

Las PNT en usuarios del sector eléctrico colombiano representan un problema de gran escala a nivel nacional, generando desbalances en la energía suministrada y afectando la estabilidad del sistema eléctrico; estas pérdidas, derivadas de fraudes, conexiones ilegales, errores en la medición y fallos en la facturación, impactan negativamente el indicador de pérdidas en consumo por kWh, reduciendo la eficiencia del sector y aumentando los costos operativos de las empresas del sector.

Además de las pérdidas económicas significativas para las compañías eléctricas y el estado, este fenómeno afecta la confiabilidad del sistema comercial, incrementa las tarifas para los usuarios regulares y puede comprometer la inversión en infraestructura y mantenimiento. La identificación y control de estas pérdidas se ven limitados por la dificultad de detectar patrones irregulares en el consumo, lo que hace necesario el desarrollo de estrategias avanzadas basadas en análisis de datos y tecnologías inteligentes para mitigar su impacto.

La ubicación geográfica de los usuarios en las diferentes zonas del país infuye en el crecimiento de las PNT, además de las normas de construcción entre otros aspectos.

## Planteamiento del Problema

En empresas del sector eléctrico colombiano, las PNT en usuarios finales, originadas por fraudes, conexiones ilegales, fallas en el sistema de medición y errores en la facturación, generan un impacto económico considerable. Estas pérdidas disminuyen los ingresos, incrementan los costos operativos y elevan los indicadores de pérdidas de energía. Además, afectan la sostenibilidad del sistema eléctrico, fomentan la inequidad en el pago del servicio y pueden derivar en aumentos tarifarios que impactan a los usuarios que cumplen con la normatividad.

Hoy en día, en la mayoría de empresas que prestan el servicio de energía eléctrica los análisis se realizan de forma manual o utilizando herramientas con infraestructura tecnológica desactualizada y al detectar irregularidades en el consumo, las empresas deben asignar personal para investigar las causas de dichas variaciones o irregularidades. Sin embargo, debido al gran volumen de datos y la cantidad de usuarios, este proceso resulta ineficiente y propenso a errores, dificultando aún más la gestión de extensas bases de información. Por esta razón, esta investigación contribuye como una fuente documental que sirva como referencia para la implementación y mejora de un modelo de *machine learning* que permita identificar de manera rápida y precisa las causas de estas anomalías en usuarios finales y poder desarrollarlas en un campo real aplicado al sector eléctrico colombiano.

## Sistematización del Problema

Las PNT en usuarios finales, de empresas prestadoras del servicio del sector eléctrico colombiano generan desbalances energéticos, afectan los indicadores de consumo y provocan pérdidas económicas significativas. Estas pérdidas se originan principalmente por fraude eléctrico, errores administrativos, fallas en la medida y supervisión inadecuada del consumo, lo que impacta la estabilidad financiera de las empresas distribuidoras, eleva el indicador de pérdidas y deteriora la infraestructura eléctrica, entre los actores involucrados se encuentran las empresas de energía, el gobierno, y los usuarios. Para mitigar este problema, se requiere de un buen manejo de información mediante el uso de herramientas tecnológicas adecuadas y la implementación de tecnologías avanzadas como la aplicación de técnicas de *machine learning*, auditorías rigurosas, sanciones estrictas, campañas de concienciación, seguimiento y control continuo para promover el consumo legal de energía.

Las pérdidas de energía son una problemática significativa en varias regiones de Colombia. Por ejemplo, en algunas zonas como Atlántico, Magdalena y La Guajira, según lo publicado en la revista LR (La República, 2022), las empresas prestadoras del servicio, Aire y Afinia, fueron las que presentaron el mayor índice de pérdidas: Aire con un 32% y Afinia con un 28.29% en las zonas de Bolívar, Cesar, Córdoba, Sucre y Magdalena.

Por otro lado Celsia, la compañía que opera en Valle del Cauca y Tolima, informó en el Valle, el índice de pérdidas llega a 9,39%, es decir, cerca de 39 gigavatios hora por año (GWh/año). Destacaron que desde 2019 han invertido para mejorar ese número, con lo que se han recuperado 7,7 GWh en este departamento. En el caso de Tolima, esta cifra es mayor, siendo de 13,42%. Esto, representa 92,72 GWh/año. (La Republica, 2022 ).

En cambio para el caso de Enel-Codensa el operador de Bogotá y Cundinamarca es la compañía con menor incidencia por este hecho. Su cifra fue de 7,53% a cierre de 2021, explicó Francesco Bertoli, gerente general de la compañía. (La República, 2022 ).

Además según la noticia publicada en el periódico El Tiempo (2022), Enel Colombia informó que durante 2022, el robo para ese año, fue de 230 GWh/año, lo que equivale al consumo mínimo vital de energía eléctrica de Chía, Zipaquirá y Cajicá durante un año, lo que se traduce a aproximadamente 145.000 hogares.

De acuerdo con la noticia publicada en el medio de comunicación Seguimiento.co (2021), en total en los departamentos de Atlántico, Magdalena y La Guajira se han realizado en los meses de octubre y noviembre 20 mil inspecciones, de las cuales se han detectado 12 mil casos de irregularidades. Esto implica que de cada 100 inspecciones se detectan 62 casos de presuntas manipulaciones de los equipos de medida o conexiones fraudulentas.

En 2025 según El Tiempo (2025), los operativos recientes entre la empresa prestadora del servicio Air-e y las autoridades para combatir las conexiones eléctricas ilegales en establecimientos comerciales y residenciales en los departamentos de Atlántico, Magdalena y La Guajira, han detectado que la serie de irregularidades continúan y aumentan en varios barrios y sectores encontrando conexiones fraudulentas, alteraciones en los equipos de medida entre otras irregularidades que representan un grave perjuicio económico para la empresa y un riesgo para la seguridad de la comunidad. Por otro lado acorde al comunicado en el diario digital para Colombia Infobae (2023), conforme con el más reciente análisis de Enel Colombia, desde enero 1 hasta el 31 de diciembre de 2023, se reportaron 28.800 casos de hurto de energía, de los cuales cerca de 12.000 corresponden a conexiones directas a la red de forma no autorizada, además de

intervenciones para alterar las conexiones y elementos del medidor, con el fin de evitar que se registre el consumo de energía de manera adecuada (Infobae,2025).

En este contexto con base en la publicación en el medio de comunicación El Pereirano el robo de energía ha crecido en Pereira y Cartago, en la empresa llamada Elite Vehículos, un centro técnico, se detectó una conexión fraudulenta que causaba pérdidas millonarias.El robo de energía en Elite Vehículos afectó la economía local, causando pérdidas anuales de 24 millones de pesos por 2,178 kWh mensuales no pagados, por tanto la empresa Energia de Pereira tomara acciones legales del caso la conexión ilegal ya que no solo representaba una evasión del pago, sino que también comprometía la seguridad y la eficiencia del sistema de energía en la zona. (El Pereirano, 2025)

De acuerdo a lo anterior, las condiciones fraudulentas incluyen varios tipos como manipulación de medidores, conexiones ilegales, redes clandestinas, alteración de medidores, conexiones directas a la red. Las empresas distribuidoras de energía en cada región suelen implementar estrategias para detectar y reducir estas pérdidas, colaborando con las autoridades locales y promoviendo campañas de concienciación entre los usuarios.

## Justificación

Para las empresas del sector eléctrico en Colombia, es fundamental llevar a cabo un monitoreo continuo de los datos de consumo de los usuarios con el fin de identificar patrones irregulares y detectar anomalías asociadas a pérdidas no técnicas, como fraudes y errores en la facturación. Una gestión eficiente en este aspecto permite diseñar estrategias más efectivas para mitigar estos riesgos y reducir las PNT.

En cuanto al uso de *machine learning* para la detección de PNT, Colombia aún tiene un margen significativo de avance; aunque existen iniciativas locales, la mayoría de los desarrollos en este campo han sido impulsados en otros países, particularmente en regiones con sistemas eléctricos más digitalizados y una mayor adopción de tecnologías de análisis de datos. A diferencia de empresas internacionales, en Colombia aún hay pocos estudios ampliamente difundidos sobre la implementación de *machine learning* para abordar esta problemática.

En este contexto, resulta necesario analizar proyectos que se enfoquen en la aplicación de técnicas de *machine learning* en otros países para la detección y control de pérdidas no técnicas, con el propósito de evaluar su viabilidad y aplicabilidad en el sector eléctrico colombiano. Este estudio permitirá seleccionar los enfoques más eficaces, proponer mejoras adaptadas a la realidad del país y contribuir a la sostenibilidad del sistema eléctrico, además de servir como referencia para futuras aplicaciones en otros sectores.

## Objetivos

### Objetivo General

Analizar la aplicación de diferentes técnicas de *machine learning* para optimizar y gestionar las pérdidas no técnicas en usuarios finales y su aplicación en el sector eléctrico colombiano.

### Objetivos Específicos

Analizar proyectos nacionales e internacionales que utilicen técnicas de *machine learning* para la detección de pérdidas no técnicas en el sector eléctrico.

Identificar las variables clave empleadas en la detección de pérdidas no técnicas en el sector eléctrico.

Comparar diferentes técnicas de *machine learning* aplicadas a la identificación de pérdidas no técnicas en el sector eléctrico y determinar las más efectivas para el contexto colombiano.

Documentar los resultados del análisis de los modelos de *machine learning*, destacando las mejores prácticas para optimizar la gestión de pérdidas no técnicas y evaluar su aplicabilidad en Colombia.

Proponer estrategias y recomendaciones basadas en los hallazgos del estudio, con el fin de mejorar la toma de decisiones y fortalecer las acciones preventivas y correctivas en la gestión de pérdidas no técnicas en Colombia

## Marco de Referencia

### Estado del Arte

Se han realizado investigaciones a nivel nacional como internacional enfocadas en el sector eléctrico, sin embargo no todas se centran en el caso específico de la presente investigación cuyo enfoque se centra en la aplicación de técnicas de *machine learning* para detección de las pérdidas no técnicas en usuarios finales, pero de alguna manera contribuyen con la investigación planteada.

Ahora bien, según la investigación realizada por Lezama Gonzales, Cachay Boza y Zevallos León (2024), su estudio se centra en la predicción y detección del robo de electricidad y consumos anómalos en la empresa Distriluz en Perú, mediante técnicas avanzadas de análisis de datos y aprendizaje automático. Estos autores desarrollan un modelo de detección proactiva mediante el empleo de *machine learning* y *deep learning*, que permite a las compañías eléctricas identificar patrones sospechosos y comportamientos inusuales en tiempo real, con el fin de prevenir pérdidas financieras y salvaguardar la integridad del sistema.

Lezama Gonzales, Cachay Boza y Zevallos León (2024), realiza 4 etapas durante su investigación que son: Preprocesamiento de datos, lógica de proceso, extracción de características y validación.

Según Ríos Villegas y Uribe Aguirre (2013), proponer e implementar una metodología para el estudio de clientes con alta probabilidad de fraude, en sistemas de distribución por medio de la minería de datos, aplicando técnicas inteligentes como herramienta de clasificación y extracción de características, lo anterior no tiene el enfoque en las pérdidas no técnicas en usuarios finales. Sin embargo, estos aportan a esta investigación dado que realiza la identificación de parámetros importantes como perfiles de carga anómalos y permite ampliar los

conocimientos enfocados en el sector eléctrico colombiano. Además realiza diferentes métodos para el análisis de consumos como críticas de consumo, consumo estimado, consumo proyectado, por capacidad del transformador de distribución, por caracterización de consumo y a partir del factor multiplicador, en ese orden de ideas plantean las etapas de medir, seleccionar, detectar, normalizar y controlar las pérdidas no técnicas. Igualmente, aplica algoritmos de aprendizaje supervisado para estimar la precisión.

De acuerdo con la investigación realizada en Colombia por Moreno Molina (2023), utiliza el algoritmo de *machine learning random forest* para analizar las características de los datos y las métricas de desempeño aplicando métodos de balanceo. Sumando a esto, propone estrategias para inversión de recursos en zonas de Bogotá y Cundinamarca con más riesgos de pérdidas no técnicas .

Conforme a Giraldo et al. (2022) el proyecto para detección de robo de energía en Latinoamérica y Caribe consta de dos módulos: un algoritmo de *machine learning* y una interfaz de usuario, el algoritmo realiza la combinación de 3 modelos 1 supervisado (*boosting*), uno semisupervisado (*autoencoders*) y finalmente un modelado de reglas analíticas (modelos estadísticos). Lo anterior ayuda a la presente investigación para comprender y evaluar el funcionamiento efectivo en la detección de pérdidas no técnicas aplicado al sector Colombiano.

Conforme a lo mencionado, los estudios previos realizados mediante la aplicación de técnicas de *machine learning* proporcionan una solución innovadora a la hora de manipular gran cantidad de datos, permitiendo tomar decisiones acertadas y efectivas y reduciendo costos operativos para las empresas del sector eléctrico en general; para el caso colombiano el uso de técnicas de *machine learning* es un proceso de mejora continua que día a día seguirá avanzando al margen de la tecnología.

Además, en Colombia, diversas empresas han avanzado tecnológicamente mediante la conformación de una alianza estratégica denominada *Colombia Eficiente*, cuyo objetivo es aplicar la Inteligencia Artificial (IA) en distintas áreas del sector eléctrico. Se menciona este contexto porque el *machine learning* es solo una pequeña parte de la IA. De acuerdo con el documento Retos y oportunidades de la IA en el sistema eléctrico colombiano (Juan D. Molina, 2024), se analizan varios aspectos del sector eléctrico. Sin embargo, un punto clave de esta investigación es la experiencia aplicada a nivel internacional en eficiencia operacional, con el propósito de fortalecer la analítica de datos para mejorar la predicción y el pronóstico de variables como el consumo energético y las condiciones ambientales, además de optimizar la eficiencia energética.

### **Marco Contextual**

Las PNT en el sector eléctrico representan un problema de gran relevancia tanto a nivel nacional como internacional. Estas pérdidas se generan por factores externos a la operación del sistema, como el fraude, conexiones ilegales, errores en la medición y facturación, o deficiencias en la gestión comercial de las empresas prestadoras del servicio.

En Colombia, el fenómeno de las pérdidas no técnicas tiene un impacto significativo en la sostenibilidad económica del sector eléctrico. Diversas empresas distribuidoras de energía enfrentan altos niveles de estas pérdidas, lo que afecta su estabilidad financiera y la calidad del servicio prestado a los usuarios. Particularmente, los consumidores con altos niveles de demanda energética pueden ser un foco de fraude, ya que los costos elevados por consumo pueden incentivar la manipulación ilícita del suministro eléctrico. Para hacer frente a esta problemática, cada operador de red ha desarrollado estrategias internas para mitigar estos riesgos, implementando programas de monitoreo, control y sanción de prácticas fraudulentas.

A nivel global, el desarrollo de tecnologías avanzadas ha permitido mejorar la gestión de las PNT. La aplicación de *machine learning* en el sector eléctrico ha demostrado ser una herramienta clave para la identificación y prevención del fraude energético. Países con infraestructura tecnológica avanzada han logrado optimizar sus sistemas de detección de anomalías en el consumo mediante el análisis de grandes volúmenes de datos en tiempo real.

En el caso de Colombia, si bien han surgido proyectos e iniciativas que buscan incorporar tecnologías como *big data*, *machine learning* e inteligencia artificial en la gestión de PNT, su implementación aún enfrenta barreras tecnológicas, económicas y normativas. La transición hacia un modelo más tecnológico y automatizado en la detección y prevención del fraude eléctrico requiere de inversiones significativas en infraestructura y capacitación, así como del fortalecimiento de políticas regulatorias que faciliten la adopción de estas herramientas.

El análisis de estudios previos y la experiencia internacional en la aplicación de estas tecnologías pueden servir como referencia para mejorar la eficiencia y confiabilidad del sistema eléctrico colombiano. En este contexto, la integración de soluciones basadas en datos y algoritmos inteligentes se presenta como una alternativa prometedora para reducir las PNT y optimizar la gestión del servicio eléctrico en el país.

### **Marco Teórico**

El sector eléctrico es fundamental para el desarrollo económico y social de cualquier país, su correcto funcionamiento depende de una serie de sistemas, dispositivos y regulaciones que garantizan la generación, transmisión, distribución y consumo eficiente de la energía. En este contexto, a continuación, se presentan algunos de los elementos más relevantes dentro de este sector, así como las herramientas tecnológicas utilizadas para mejorar su eficiencia y sostenibilidad.

Las pérdidas no técnicas (PNT), corresponde a energía que se pierde en un Mercado de Comercialización por motivos diferentes al transporte y transformación de la energía eléctrica.(CREG 172, 2011). Es decir la diferencia entre la energía entregada y la energía facturada por ejemplo el robo hace referencia a la energía tomada por los usuarios tomada a través de conexiones ilegales , fraude que se refiere a la alteración en los elementos del sistema de medida o conexiones por fuera de medida, otra caso de PNT son la energía no facturada o mal medida así mismo debido a problemas de gestión es decir errores humanos en la contabilidad ó errores en la toma de lecturas ó de sistemas comerciales obsoletos que arrojan cálculos erróneos al momento de la facturación respectiva. Así mismo las PNT son una parte fundamental para el cálculo del IPT (Indicador de pérdidas Totales ) dentro del plan de regulación de pérdidas establecido por la CREG como ente regulador.(CREG 172, 2011).

El equipo de medida o medidor es un dispositivo fundamental el cual sirve para medir y registrar el consumo o las transferencias de energía eléctrica, el cual incluye la medición de la energía activa o reactiva, mediante la integración de la potencia activa o reactiva con respecto al tiempo. (Resolución CREG 038 de 2014, 2014).

El sistema eléctrico está compuesto por centrales eléctricas, redes de transmisión y distribución, además de interconexiones y equipos relacionados. En este contexto, XM (s. f.) señala que “el sistema eléctrico colombiano es el conjunto de participantes del Mercado de Energía Mayorista colombiano que hacen parte de la cadena productiva, así: generadores, transmisores, distribuidores y comercializadores”.

Por otro lado, la energía facturada corresponde a la cantidad de energía o consumo que las empresas comercializadoras facturan a los clientes, medida por el equipo de medida en kilovatios-hora (kWh) para energía activa y kilovoltiamperios reactivos por hora (kVARh) para

energía reactiva. Ahora bien para comprender el comportamiento de los datos registrados en el equipo de medida, se analizan mediante *softwares* de acuerdo al tipo, marca y forma de registro del medidor los datos convirtiendolos en lo que se conoce como el perfil de carga o de consumo, el cual refleja el patrón de demanda de electricidad de un usuario o un grupo de usuarios durante un período determinado. La demanda, en este contexto, XM (s. f.) la define como “carga en kWh solicitada a la fuente de suministro en el punto de recepción durante un periodo de tiempo determinado”.

En Colombia, varias entidades regulan y supervisan el sector eléctrico. XM (s. f.) señala que “la Unidad de Planeación Minero Energética (UPME) es una unidad administrativa especial, adscrita al Ministerio de Minas y Energía, encargada de la planeación integral del sector minero energético, creada por el decreto 2119 de 1992 y organizada según lo previsto en el artículo 15 de la Ley 143 de 1994”.

Por otro lado, XM (s.f.) define que “la Comisión de Regulación de Energía y Gas (CREG) es el Organismo creado mediante el artículo 68 y siguientes de la Ley 142 de 1994, como unidad administrativa especial, con independencia administrativa, técnica y patrimonial, adscrita al Ministerio de Minas y Energía, encargada de emitir la regulación del sector eléctrico y de gas combustible. En el mismo contexto XM (s.f.) señala que “ la Superintendencia de Servicios Públicos Domiciliarios (SSPD) es la entidad encargada de la inspección y vigilancia de las entidades que presten los servicios públicos domiciliarios, y los demás servicios públicos a los que se aplica la Ley 142 de 1994, creada por el artículo 75 y siguientes de dicha Ley”.

En el ámbito tecnológico, a nivel global, la inteligencia artificial (IA) y el *machine learning* (aprendizaje automático) han cobrado gran importancia en la optimización del sistema eléctrico. En ese orden de ideas *machine learning* es una rama de la inteligencia artificial que

permite a los sistemas aprender y mejorar su rendimiento mediante el análisis de datos sin necesidad de ser programados explícitamente. Para ello, se utilizan herramientas como la librería Pandas, una de las más populares en el desarrollo con *Python*, especialmente en el ámbito de *data science* y *machine learning*. Pandas facilita la manipulación y tratamiento de datos, permitiendo cargar, modelar, analizar y preparar información para mejorar la eficiencia en la toma de decisiones dentro del sector eléctrico y , en este contexto hay evidencia de estudios realizados los cuales utilizan técnicas de *machine learning* y han sido eficientes a la hora de gestionar las pérdidas y para acciones para reducir y controlar estas pérdidas las cuales pueden ser las técnicas, que representan la energía disipada como calor en los elementos de conexión entre la fuente en la red de suministro eléctrico y los puntos de entrega o usuarios y se deben a las leyes físicas, efecto *Joule*, eficiencia de los componentes del sistema, y a la falta de calibración de los equipos de medición; por otro lado existen las pérdidas no técnicas las cuales se definen como la diferencia entre las pérdidas totales y las técnicas y corresponden a las generadas por robos o conexiones ilegales a la red eléctrica, modificación ilegal de los equipos de medida, errores de medición y de facturación, entre otros. (Alvarez y Román, 2024).

### **Marco Conceptual**

Las pérdidas no técnicas (PNT) en el sector eléctrico representan un desafío significativo, especialmente en el caso de usuarios con altos consumos. Estos usuarios, debido a la gran cantidad de energía que demandan, deben asumir costos elevados en sus facturas mensuales, lo que puede aumentar la probabilidad de que algunos recurran a prácticas ilegales para reducir sus pagos, como el fraude eléctrico o manipulaciones en los medidores.

En Colombia, existen múltiples empresas encargadas de la prestación del servicio de energía eléctrica, muchas de las cuales enfrentan pérdidas de energía considerables cada año.

Para mitigar este impacto, cada operador implementa estrategias internas dirigidas a la detección y reducción de estas pérdidas, mediante el uso de auditorías, tecnología de medición avanzada y planes de control de fraude.

A nivel internacional, la implementación de técnicas avanzadas de *machine learning* en el sector eléctrico ha permitido la detección y reducción eficiente de pérdidas no técnicas. Gracias al desarrollo de infraestructuras tecnológicas sofisticadas, se han logrado avances significativos en la identificación de patrones de consumo anómalos y en la automatización de procesos de monitoreo. En Colombia, si bien existen iniciativas y proyectos enfocados en la aplicación de estas tecnologías, aún se requiere mayor inversión, apoyo y mejoras en la infraestructura tecnológica para alcanzar el nivel de desarrollo observado en otros países.

De acuerdo a lo mencionando por Yépez-García, R. A., & Jiménez Mori, R. (Eds.). (2024). En República Dominicana y Brasil, por ejemplo, se han identificado porcentajes significativos de usuarios irregulares cuyas conexiones han sido realizadas por los propios hogares y/o forman parte de redes que requieren mejoras y modernización. En cierta medida, esta situación se explica por el crecimiento desordenado de las ciudades latinoamericanas, las cuales habrían sobrepasado las capacidades de inversión de las empresas del sector. Esta situación ha devenido en la persistencia de un suministro deficiente de energía y en la presencia de amplios grupos de usuarios que no pagan por los servicios. Además, las fuentes de pérdidas mencionadas evidencian un cierto grado de imposibilidad de las empresas para medir la electricidad suministrada a usuarios finales. Es decir, las empresas pierden ingresos por el consumo de electricidad que no pueden atribuir o identificar.

En los países en desarrollo, las PNT pueden variar de entre el 10% y el 40% de la generación total (Smith, 2004). Incluso en Estados Unidos (EEUU) y Europa se pueden

encontrar estados y empresas eléctricas con pérdidas que superan de los niveles de eficiencia. Por ejemplo, en EEUU hay estados donde las pérdidas se ubican entre el 8% y 13.3% (Wirfs Brock, 2015). En Europa, un grupo de países (por ejemplo, Kosovo, Montenegro, Macedonia del Norte y Serbia) presentaron niveles de pérdidas superiores al 12% en 2018 (CEER, 2020). Yépez-García, R. A., & Jiménez Mori, R. (Eds.). (2024).

El análisis de las PNT mediante estudios basados en *big data* y *machine learning* facilita la optimización de los procesos de gestión y análisis de grandes volúmenes de datos. Además, estos estudios sirven como referencia para la toma de decisiones estratégicas que permitan mejorar la eficiencia del sistema eléctrico y reducir el impacto económico de PNT en el país.

### **Marco Normativo**

El marco normativo para el sector eléctrico a nivel mundial es amplio dependiendo de las normas de construcción de cada País y las condiciones de prestaciones de los servicios públicos, ahora bien en Colombia existen entes que regulan aspectos importantes relacionados con la Generación, Transmisión, Distribución y Comercialización de energía eléctrica como la CREG(Comisión de Regulación de Energía y Gas), la SSPD(Superintendencia de Servicios Públicos Domiciliarios), MinEnergía(Ministerio de Minas y Energía), UPME (Unidad de Planeación Minero Energética), ASIC(Administrador de Servicios de Intercambios Comerciales), CND(Centro nacional de Despacho), CNO(Concejo Nacional de Operación). Los entes mencionados anteriormente han consolidado a través de leyes, decretos, circulares, acuerdos y resoluciones que en conjunto estas instituciones trabajan para mantener la sostenibilidad, calidad y acceso equitativo a la energía en el país, asegurando que el servicio eléctrico cumpla con los estándares técnicos y normativos requeridos.

### ***Defraudación de Fluidos***

La Ley 599 de 2000, en su artículo 256, penaliza el hurto de energía, agua, gas natural o señal de telecomunicaciones mediante mecanismos clandestinos o alteraciones en medidores. El que mediante cualquier mecanismo clandestino o alterando los sistemas de control o aparatos contadores, se apropie de energía eléctrica, agua, gas natural, o señal de telecomunicaciones, en perjuicio ajeno, incurrirá en prisión de dieciséis (16) a setenta y dos (72) meses y en multa de uno punto treinta y tres (1.33) a ciento cincuenta (150) salarios mínimos legales mensuales vigentes., siendo un referente clave para el análisis de las pérdidas no técnicas en el presente estudio. (Departamento Administrativo de la Función Pública, s.f.)

### ***Ley 142 de 1994***

La ley 142 de 1994 regula el régimen de los servicios públicos domiciliarios en Colombia, ya que establece derechos y deberes de los usuarios y las empresas prestadoras del servicio, además permite la suspensión del servicio en casos de fraude y el cobro de consumos no facturados. Con forme a lo mencionado, esta normativa es fundamental para la realización de recuperación o procesos administrativos de recuperación de energía en el caso de identificar condiciones anormales o fraudulentas que permitan facturar la energía recuperada. (Departamento Administrativo de la Función Pública, s.f.)

### ***Resolución CREG 174 de 2011.***

La Resolución CREG 174 de 2011 establece criterios para la reducción de pérdidas no técnicas y define incentivos para su control, promoviendo la eficiencia en la prestación del servicio de energía. Esta resolución es clave, ya que obliga a las empresas distribuidoras a implementar estrategias efectivas para la detección y mitigación del fraude eléctrico, garantizando así un servicio más confiable y equitativo para los usuarios. Su impacto en la

regulación del sector eléctrico contribuye a reducir los costos asociados a estas pérdidas, beneficiando tanto a las empresas como a los usuarios finales. (Comisión de Regulación de Energía y Gas [CREG], 2011).

#### ***Resolución CREG 24 de 2021***

La Resolución CREG 24 de 2021 establece los índices de pérdidas eficientes, define los costos de reposición de referencia y determina las sendas de reducción de pérdidas, contribuyendo así a una gestión más eficiente y transparente en el sector eléctrico. (Comisión de Regulación de Energía y Gas [CREG], 2021).

#### ***Resolución CREG 038 de 2014***

La Resolución CREG 038 de 2014 establece las características técnicas, así como los requisitos de instalación, calibración y mantenimiento de los equipos de medición, garantizando su precisión y confiabilidad en la medida. (Comisión de Regulación de Energía y Gas [CREG], 2014).

#### ***Resolución CREG 108 de 1997***

En esta resolución se señalan criterios generales sobre protección de los derechos de los usuarios de los servicios públicos domiciliarios de energía eléctrica y gas combustible por red física, en relación con la facturación, comercialización y demás asuntos relativos a la relación entre la empresa y el usuario. Es importante conocer esta resolución para garantizar el beneficio del usuario y también de la empresa en cualquier proceso que se realice. (Comisión de Regulación de Energía y Gas [CREG], 1997).

#### ***Reglamento Técnico de Instalaciones Eléctricas (RETIE)***

El Reglamento Técnico de Instalaciones Eléctricas (RETIE) es una normativa en Colombia que establece los requisitos técnicos y de seguridad que deben cumplir las

instalaciones eléctricas en el país. Su principal objetivo es proteger la vida y la salud de las personas, la preservación del medio ambiente y la seguridad de la infraestructura eléctrica, minimizando los riesgos asociados al uso de la electricidad. Es fundamental conocer y hacer cumplir este reglamento a fin de garantizar la seguridad de la vida de cualquier individuo.

(Ministerio de Minas y Energía, s.f.)

### ***Decreto 3735 de 2003***

Este decreto reglamenta artículos de la Ley 812 de 2003 en relación con el programa de normalización de redes eléctricas y los esquemas diferenciales de prestación del servicio público domiciliario de energía eléctrica. Busca, entre otros objetivos, reducir las pérdidas no técnicas asociadas al fraude y conexiones ilegales. (Ministerio de Minas y Energía, s.f.)

### ***Normativa Internacional***

En países como España la normativa para la defraudación de Fluidos esta contenida en el artículo 255 del Código Penal Ley Orgánica 10/1995, el cual Penaliza la manipulación fraudulenta de contadores de servicios como electricidad y agua, las sanciones pueden incluir multas económicas significativas, dependiendo del valor defraudado. (Boletín Oficial del Estado [BOE], 1995). Además, mediante el Real Decreto 1110/2007 (Boletín Oficial del Estado [BOE], 2007), se aprueba, el Reglamento Unificado de Puntos de Medida del Sistema Eléctrico, que establece las condiciones técnicas y funcionales de los equipos de medida y su gestión.

Por otro lado, en Ecuador la Ley Orgánica del Servicio Público de Energía Eléctrica (LOSPEE). La presente ley tiene por objeto garantizar que el servicio público de energía eléctrica cumpla los principios constitucionales de obligatoriedad, generalidad, uniformidad, responsabilidad, universalidad,, accesibilidad, regularidad, continuidad, calidad, sostenibilidad ambiental, precaución, prevención y eficiencia, para lo cual, corresponde a través del presente

instrumento, normar el ejercicio de la responsabilidad del Estado de planificar, ejecutar, regular, controlar y administrar el servicio público de energía eléctrica. (Ley Orgánica del Servicio Público de Energía Eléctrica, 2015).

Por otro lado en Ecuador también se implementan directrices para la implementación de sistemas de medición y estrategias para la reducción de pérdidas en la distribución eléctrica.

Ahora bien, en Chile la Defraudación de Energía esta regulada por el Código Penal de su País, en el cual sanciona el fraude de energía eléctrica, agua o gas. (Codigo Penal de Chile, 1874). Así mismo, cumplen con las normas técnicas de electricidad establecidas por la Comisión Nacional de Energía (CNE) (Comisión Nacional de Energía, s.f.).

Por otro lado según la Ley Orgánica del Sistema y Servicio Eléctrico (República Bolivariana de Venezuela, 2010), regula la prestación del servicio eléctrico, estableciendo derechos y deberes de los usuarios y las obligaciones de las empresas prestadoras, como también según la Norma COVENIN 200-2004 (FONDONORMA, 2004) , las cuales regulan las instalaciones eléctricas en el país, con el fin de garantizar la seguridad y eficiencia.

En ese orden de ideas, en países como Perú, la defraudación de servicios públicos está regulada por el artículo 186 del Código Penal del Perú (Congreso de la República del Perú, 1991/2016), que tipifica como delito la manipulación o alteración de los sistemas de medición de servicios públicos, incluyendo la energía eléctrica, con sanciones que pueden incluir penas privativas de libertad. Igualmente, Perú aplica el Código Nacional de Electricidad (Ministerio de Energía y Minas del Perú, 2004), el cual establece las condiciones técnicas y de seguridad que deben cumplir las instalaciones eléctricas en el país. Además, mediante la Política Energética Nacional 2050 incluye objetivos para fortalecer los estándares de calidad y eficiencia en el sector energético. (Ministerio de minas y energía del Perú, 2024).

Finalmente Perú aplica las políticas Regulatorias al Sector Energía y Minería.

(Osinergmin, 2017).

De acuerdo a lo anterior, cada país supervisa e implementa medidas normativas de acuerdo con los requerimientos específicos de su sector eléctrico, estableciendo criterios claros para garantizar la transparencia en la prestación del servicio de energía. Estas regulaciones permiten asegurar una medición confiable del consumo, controlar las pérdidas no técnicas en los usuarios finales y aplicar una facturación precisa que no perjudique ni a los consumidores ni a las empresas prestadoras del servicio. De esta manera, se busca equilibrar la sostenibilidad del sistema eléctrico con la protección de los derechos de los usuarios.

## Metodología

La metodología para analizar la aplicación de diferentes técnicas de *machine learning* para optimizar y gestionar las PNT en usuarios finales y su aplicación en el sector eléctrico colombiano se desarrolla en 5 etapas clave, utilizando un enfoque mixto de tipo descriptivo, exploratorio y observación experimental utilizando tanto técnicas cuantitativas como cualitativas. El enfoque principal es el análisis de datos en estudios previos nacionales e internacionales para identificar variables que aporten en la detección de PNT en usuarios finales y las técnicas de Machine Learning aplicadas que se pueden aplicar al sector eléctrico colombiano.

### Etapa 1

En esta etapa se analiza proyectos internacionales y documentos que hagan uso de técnicas de Machine Learning para la detección de PNT en el sector eléctrico, además de proyectos nacionales, y que aportan conceptos fundamentales para la presente investigación conforme a la siguiente tabla:

**Tabla 1**

*Estudios Previos Sobre Aplicación Técnicas de Machine Learning*

Año	Título del Estudio	País
2024	Molina, J. D. (2024). <i>Retos y oportunidades de la IA en el sistema eléctrico colombiano.</i>	Colombia
2024	Melgar García, L., Torres Maldonado, J. F., Troncoso, A., & Riquelme Santos, J. C. (2024). <i>Técnicas Big Data para la predicción de la demanda y precio eléctrico.</i>	España
2016	Glauner, P., Meira, J. A., & Dolberg, L. (2016). <i>Neighborhood Features Help Detecting Non-Technical Losses in Big Data Sets.</i>	Brazil
2007	Quiñonez Pinilla, F. A., Sánchez Ramírez, F. A., & Avendaño Sepúlveda, J. V. (2007). <i>Estudio de viabilidad para el control y disminución sostenible del índice de pérdidas técnicas y no técnicas en el sistema eléctrico de Casanare.</i>	Colombia

Año	Título del Estudio	País
2024	Lezama Gonzales, P. M., Cachay Boza, O., & Zevallos León, M. (2024). <i>Modelo predictivo para la detección del hurto de la energía eléctrica.</i>	Perú
2023	Deloitte. (2023). <i>Deep Dive: Digitalización.</i>	Chile
2022	Solis Mora, V. S., & Gruezo Valencia, D. F. (2022). <i>La Inteligencia Artificial (IA) al servicio de la eficiencia energética en Ecuador.</i>	Ecuador
2013	Lima Pérez, L., & Vásquez Stanescu, C. L. (2013). <i>Variables determinadas para la detección de fraude en empresas de servicio eléctrico</i>	Venezuela
2024	Tibaquirá Giraldo, J. E., Restrepo Victoria, Á. H., Caicedo Bravo, E. F., Alfonso Morales, W., Ospina Castro, S., Valencia, J. C., Barrera Zapata, M., Duque, M., Tabares Giraldo, J. J., Zapata Mina, M., & Valencia Duque, M. (2024). <i>Integra: Herramienta para la gestión y visualización de indicadores de eficiencia energética en el sector eléctrico colombiano.</i>	Colombia
2024	Biblioteca Felipe Herrera del Banco Interamericano de Desarrollo. (2024). <i>Economía de las pérdidas de electricidad en América Latina y el Caribe</i>	América Latina
2022	Biblioteca Felipe Herrera del Banco Interamericano de Desarrollo, Giraldo, M. C., Ríos, C., Alarcón, A., Snyder, V., Echevarría, C., Riobo, A., Hallack, M., & Irigoyen, J. L. (2022). <i>Energizados: Beneficios de una herramienta basada en Machine Learning para la detección de robo eléctrico</i>	América Latina
2024	Ortiz-Torres, L. F., Gómez-Luna, E., & Marlés-Sáenz, E. (2024). <i>Estudio del uso y contribución de la inteligencia artificial para la operación en redes eléctricas</i>	Colombia
2014	Flores Coaguila, J. D. (2014). <i>Propuesta de modelo de detección de fraudes de energía eléctrica en clientes residenciales de Lima Metropolitana aplicando minería de datos</i>	Perú
2021	Crespo García, D., & Monzo Sánchez, C. (2021). <i>Aplicación de Machine Learning al consumo eléctrico de edificios inteligentes</i>	España
2023	Ortega Díaz, L., Cárdenas Rangel, J., & Osma Pinto, G. (2023). <i>Estrategias de predicción de consumo energético en edificaciones: una revisión</i>	Colombia
2013	Ríos Villega, A. F., & Uribe Aguirre, K. A. (2013). <i>Minería de Datos Aplicada a la Detección de clientes con alta probabilidad de fraudes en sistemas de distribución.</i>	Colombia
2023	Moreno Molina, H. E. (2023). <i>Estrategia para seguimiento de agresividad por pérdidas no técnicas utilizando balances de energía.</i>	Colombia

Año	Título del Estudio	País
2021	Alfonso Ortiz, J. E., & Bogoya Mora, D. J. (2021). <i>Predicción del Recurso Solar diario mediante técnicas de Machine Learning para la proyección de generación de Energía Eléctrica</i>	Colombia
2024	Álvarez Castillo, S. M., & Román Pereira, L. M. (2024). <i>Detección de instalaciones fraudulentas utilizando datos históricos de consumo energéticos</i>	Colombia
2023	Llagua Arévalo, J. L. (2023). <i>Detección de Pérdidas No Técnicas en clientes especiales con telemedición basada en inteligencia artificial con aplicación en la empresa Ambato.</i>	Ecuador
2022	López Cardona, J. E. (2022). <i>Desarrollo de Plataforma de Monitoreo para activos críticos del Sector Eléctrico en Transformadores.</i>	Colombia
2020	Loor Sánchez, R. C. (2020). <i>Modelo de detección de fraudes eléctricos en clientes residenciales en una empresa distribuidora de electricidad en la ciudad de Guayaquil.</i>	Ecuador
2023	Pachón Rodríguez, W. A. (2023). <i>Detección de fraudes en empresas de servicios públicos domiciliarios utilizando ciencia de datos y análisis geoespaciales.</i>	Colombia
2023	Polanco, G. A. (2023). <i>Análisis de información para la reducción de pérdidas no técnicas utilizando herramientas tecnológicas en el sector.</i>	Colombia
2022	Messina, D., Contreras Lisperguer, R., & Salgado Pavez, R. (2022). <i>Tendencias en materia de digitalización del sector eléctrico.</i>	América Latina
2021	Santos Páez, C. D. (2021). <i>Algoritmo para la identificación de pérdidas no técnicas en baja tensión.</i>	Colombia
2019	Gonzales Reyes, J. D., & Rodríguez Moreno, L. E. (2019). <i>Sistema conceptual para implementar un algoritmo de detección y ubicación de fraudes eléctricos en un sistema de distribución.</i>	Colombia
2023	Macía Fiteni, A. (2023). <i>Smart University, Deteccion de Anomalias en consumo energético.</i>	Colombia
2024	Robert, A., Potter, K., & Frank, L. (2024). <i>Machine Learning Applications in Electric Power Systems: Enhancing Efficiency, Reliability, and Sustainability.</i>	NA
2023	Burgos Díaz, W. (2023). <i>Identificación del origen de fallas en el sistema de distribución eléctrica del Ecuador mediante la aplicación de técnicas de machine learning.</i>	Ecuador

Año	Título del Estudio	País
2022	Massaferro Saquieres, P. (2022). <i>Detección de pérdidas no técnicas en redes eléctricas en un contexto de migración tecnológica y maximizando el retorno económico.</i>	Uruguay
2021	Bustos, T. (2021). <i>Estrategia para la conceptualización de modelos de ia en el contexto de gestión de la energía.</i>	España
2017	Quezada, F. M. (2017). <i>Modelo basado en minería de datos para la detección de pérdidas no técnicas de redes de distribución.</i>	España
2023	Torres Sánchez, A. S. (2023). <i>Predicción de la Demanda de Energía Eléctrica Usando Modelos de Inteligencia Artificial para Series Temporales.</i>	Colombia
2022	Sánchez Richar, M. A., & Cuasapaz, P. E. J. (2022). <i>Predicción del hurto de energía eléctrica a través del uso de la inteligencia artificial mediante algoritmos de machine learning para cnel ep unidad de negocios santo domingo.</i>	Ecuador
2021	Celi Peña, C. E., & Guartán Fiel, F. E. (2021). <i>Aplicación de tecnicas de machine learning para desagregacion y pronostico de perfil dec carga en el sector industrial.</i>	Ecuador
2024	CODELECTRA - Comité de Electricidad de Venezuela. (1997). <i>Código Electrico Nacional FONDONORMA</i>	Venezuela
2024	Ramos de Lavarrello, M., & Marie, R. (2024). <i>Politica Energetica nacional del Perú.</i>	Perú
1999	Gobierno de Venezuela. (1999). Capítulo de normas del código eléctrico. Universidad Metropolitana de Venezuela. <i>Norma venezolana covenin código eléctrico nacional.</i>	Venezuela
2006	Ministerio de Minas y Energía, República de Perú. (2006). <i>Código Nacional de Electricidad.</i>	Perú
2003	Córdoba Angulo, M. (2003). <i>La defraudacion de fluidos en la legislación penal colombiana.</i>	Colombia
2015	Rivadeneira Burbano, G., & Rivas Ordoñez, L. (2015). <i>Ley organica del servicio publico de energia eléctrica.</i>	Ecuador

Año	Título del Estudio	País
2024	Gobierno de Honduras, Empresa Nacional de Energía. (2024). <i>Plan de inversión pública.</i>	Honduras
2024	Barrera Devia, J. S., & Carvajal García, J. (2024). <i>Análisis de modelos de aprendizaje automático para la segmentación y predicción de la demanda de clientes no regulados en el mercado eléctrico colombiano.</i>	Colombia
2014	Porras, J. A. (2014). <i>Identificación de pérdidas en sistemas de energía mediante aplicación de técnicas de análisis y visualización de información.</i>	Colombia
2010	Vivas Velasco, D. (2010). <i>Ley orgánica del sistema y servicio eléctrico.</i>	Venezuela
2017	Organismo Supervisor de la Inversión en Energía y Minería (Osinergmin). (2017). <i>Políticas regulatorias aplicadas a los sectores de energía y minería.</i>	Perú
2010	Comisión de Regulación de Energía y Gas (CREG). (2011). <i>Planes de reducción de pérdidas no técnicas de energía eléctrica.</i>	Colombia

Nota. Se analizaron 48 documentos incluyendo estudios previos en varios países y normativa.

## Etapa 2

El consumo eléctrico de un usuario final puede verse influenciado por diversos factores, tanto externos como internos, lo que puede generar que se presenten las PNT. Entre los factores externos se encuentran condiciones climáticas, eventos socioeconómicos, festividades, variaciones en los precios de la energía o patrones de consumo del usuario, conexiones ilegales, los cuales están fuera del control del operador de la red eléctrica. Por otro lado, los factores internos abarcan aspectos técnicos relacionados con el punto de conexión a la red de distribución, daños en los equipos y sistemas de medición, fallas en los diseños. Estos aspectos entre otros, son elementos a considerar ya que pueden generar PNT de tipo administrativas, errores en la medida, usuarios consumos ceros, fraudes ocasionados por condiciones anormales o conexiones ilegales.

En esta etapa se presenta las principales variables fundamentales que influyen en la detección de PNT en el sector eléctrico, cabe resaltar que la selección de las variables está fundamentada en el análisis exploratorio realizado a los estudios nacionales e internacionales previos

referentes a la detección de pérdidas en el sector eléctrico, ya que las investigaciones muestran diferentes perspectivas según el contexto. Este análisis se ajusta a las particularidades de la problemática planteada.

## Tabla 2

### *Identificación de Variables Sobre Estudios Realizados*

Categoría de la variable	Descripción de variables	Tipo
Variables de Consumo Eléctrico	<p>Historial de consumo (kWh).</p> <p>Curva de demanda diaria, mensual , anual.</p> <p>Demanda de energía eléctrica total.</p> <p>Consumo de energía diario, mensual, semestral, anual.</p> <p>Potencia activa por hora.</p> <p>Variación de consumo.</p> <p>Consumo atípico o anomalías.</p> <p>Potencia promedio mensual.</p> <p>Potencia pico.</p> <p>Consumo de energía pico.</p> <p>Perfil de carga.</p> <p>Matriz de consumo.</p> <p>Vaor de energía de entrada.</p> <p>Valor de energía de salida.</p> <p>Consumo por actividad.</p> <p>Calculo de energía.</p> <p>Consumo total.</p> <p>Lectura de medidor por periodo.</p> <p>Valor de consumo de energía no registrado.</p>	Cuantitativas

Categoría de la variable	Descripción de variables	Tipo
	Cantidad de energía acumulada.	
	Lectura energía activa.	
	Lectura energía reactiva.	
	Valor de perdidas de energía por subestacion.	
	Cantidad de fallas.	
	Consumo macro.	
	Consumo por carga instalada	
	Rango crítico.	
	Rango intermedio.	
	Rango corriente.	
	Rango de tensión .	
	Corriente de línea.	
	Angulo de tensión.	
	Tensión de fase.	
	Potencia real.	
	Potencia aparente .	
	Corriente total demandada .	
	Consumo legal en el nodo.	
	Tensión de fase .	
	Intervalo de tiempo.	
	Valor máximo de curva diaria.	
	Valor mínimo de curva diaria.	
	Valor de energia promedio anual, semestral, trimestral.	
	Valor energía facturada.	

Categoría de la variable	Descripción de variables	Tipo
	Energía no facturada anual y mensual.	
	Porcentaje de energía no registrada .	
	Porcentaje usuarios consumo cero.	
	Energía activa acumulada.	
	Lectura activa acumulada.	
	Lectura reactiva acumulada.	
	Potencia total acumulada.	
Variables Espaciales y Geográficas	Ubicación del cliente (urbano/rural)	Cualitativas y cuantitativas
	Nivel socioeconómico por zona (Estrato).	
	Zonas con alta incidencia de fraude.	
	Coordenadas (Longitud, Latitud).	
Variables del usuario	Tipo de uso (residencial, industrial, comercial)	Cualitativas, cuantitativas y binarias
	Historial de pagos/morosidad	
	Reclamaciones frecuentes	
	Cambios de comportamiento de consumo.	
	Ocupación	
	Nombre	
	Dirección	
	Apellidos	
	Usuario reincidente.	
	Usuario PARE.(Usuario con Procesos Administrativos de Recuperación de Energía)	
	Código usuario	
	Número de revisiones	
	Identificación de usuarios especiales.	

Categoría de la variable	Descripción de variables	Tipo
	Estado del usuario (Activo, inactivo, suspendido)	
	Actividad comercial	
	Cantidad de inspecciones.	
	Cantidad de suspensiones.	
	Jornadas laborales.	
Variables Técnicas	Intervenciones técnicas o mantenimientos.	Cualitativas y Cuantitativas
	Tamaño de los usuarios comerciales.	
	Tipo de conexión.	
	Topología del Circuito.	
	Nivel de tensión.	
	CT (Transformador de corriente) de Medida.	
	PT (Transformador de tensión) de Medida.	
	Medidor local.	
	Tipo de macromedidor.	
	Carga instalada.	
	Tipo de red eléctrica	
	Numero de transformador.	
	Estado medidor.	
	Factor.	
	Registro histórico de fallas.	
	Diagrama fasorial.	
	Capacidad instalada	
	Número de transformadores.	
	Numero de circuitos	
Variables Temporales	Estacionalidad o ciclos diarios/semanales/mensuales	Cuantitativas y cualitativas.

Categoría de la variable	Descripción de variables	Tipo
	Cambios bruscos de consumo en períodos específicos.	
	Registro de eventos.	
	Patrones de consumo.	
	Precio de bolsa.	
	Lista de agentes.	
	Volumen de los embalses.	
	Mantenimientos.	
	Daños.	
Variables de Infraestructura y Red	Pérdidas por circuito o alimentador	Cualitativas y cuantitativas.
	Calidad del servicio eléctrico	
	Gestión remota de datos.	
	Tipo de comunicación	
	Cantidad de conexiones	
Variables Derivadas de Técnicas Analíticas	Clustering de perfiles de consumo normales.	Cuantitativas
	Clustering de perfiles de consumo anormales.	
	Indicador de pérdidas.	
	Valores fuera de rango.	
	Comparacion de consumos entre periodos.	
	Balances de Energía.	
	Variable fraude.	
	Perfil de carga diario, mensual.	
Variables meteorológicas	Precipitación diaria .	Cuantitativas
	Velocidad del viento.	
	Temperatura.	
	Indice de claridad.	

Categoría de la variable	Descripción de variables	Tipo
	Humedad relativa. Presión atmosférica. Dispersión . Absorción . Reflexión. Radiacion .	
Variables de Tiempo	Año. Mes. Dia. Hora. Tipo dia. Tiempo de funcionamiento. Tasa de abandono. Fecha de revisión.	Cuantitativas y cualitativas

### **Etapas 3**

Esta etapa tiene como objetivo explorar y comparar las diferentes técnicas de *machine learning* aplicadas a la detección de (PNT), analizando su desempeño reportado en estudios previos y valorando su adecuación a las particularidades del sector eléctrico colombiano.

A continuación, se presenta una tabla que recopila las principales técnicas de *machine learning* identificadas en los estudios analizados, las cuales incluyen aprendizaje supervisado y no supervisado:

**Tabla 3***Comparación Técnicas de Machine Learning*

Técnica de <i>machine learning</i>	Tipo	Descripción Breve	Aplicabilidad a PNT	Relevancia para la detección de PNT
GMM(Gaussian Mixture Model)	No Supervisado	Usado principalmente para <i>clustering</i> o agrupamiento de datos, así como para modelado de densidad. No utiliza etiquetas para clasificar datos, asume que los datos provienen de una mezcla de distribuciones gaussianas (normales) y busca estimar los parámetros de estas distribuciones, como medias, varianzas y probabilidades de cada componente.	Si	Este modelo permite agrupar los datos de consumo de electricidad en diferentes " <i>clusters</i> " o grupos según sus características, permite identificar <i>clusters</i> de usuarios con patrones de consumo similares, y es útil cuando los <i>clusters</i> de fraude no siguen una estructura simple. Además de detectar patrones de consumo anómalos sin etiquetas previas.
K-MEANS	No Supervisado	K-Means agrupa los datos en K <i>clústeres</i> basados en la similitud de las características de los datos	Si	Ideal para la detección de valores anómalos, y para segmentar datos en grupos similares, como agrupar usuarios en base a patrones de consumo. También para localizar zonas con alta densidad de usuarios, además de observar usuarios que se mantienen por debajo del

Técnica de <i>machine learning</i>	Tipo	Descripción Breve	Aplicabilidad a PNT	Relevancia para la detección de PNT
				consumo promedio y como están distribuidos.
<i>Clustering Jerárquico</i>	No Supervisado	Analiza los datos identificando grupos homogéneos de acuerdo a características seleccionadas, agrupandolos en un árbol de <i>clústeres</i> , buscando construir una jerarquía de grupos.	Si	Segmenta usuarios sin necesidad de especificar número de grupos.
Modelo <i>Prophet</i>	Semi- <i>machine learning</i>	Modelo aditivo con base estadística, pero automatizado.	Parcial	Ideal para el manejo de series temporales de consumos de energía.
<i>Redes Neuronales LSTM((Long Short-Term Memory)</i>	Supervisado	Red neuronal recurrente (RNN).	Si	Ideal para grandes volúmenes de datos , series temporales y de gran utilidad para modelado de datos complejos no lineales, multivariados.
VGG16( <i>Visual Geometry Group 16</i> )	Red neuronal	Convolutiva, profunda diseñada para clasificar imágenes 13 capas convolucionales y 3 capas completamente conectadas ( <i>fully connected</i> )	Si	VGG16 puede procesar directamente las imágenes originales como entradas. Durante el entrenamiento, la red aprende automáticamente las características relevantes a partir de los datos de consumo.
<i>Random Forest</i>	Supervisado	Combina varios árboles de decisión buscando mejorar la	Si	Alta precisión en la clasificación robusta de los

Técnica de <i>machine learning</i>	Tipo	Descripción Breve	Aplicabilidad a PNT	Relevancia para la detección de PNT
XGBoost ( <i>Extreme Gradient Boosting</i> )	Supervisado	precisión y reducir el sobreajuste. (Lezama Gonzales, Cachay Boza, & Zevallos León, s.f.)	Si	datos y ayuda a identificar características de consumo asociadas a comportamientos sospechosos, o manipulación de equipos de medida, a partir de variables de consumo como datos históricos, consumo promedio, horarios de producción entre otras.
<i>Logistic Regresión</i>	Supervisado	Basado en árboles de decisión que se enfoca en mejorar la velocidad y el rendimiento del modelo. XGBoost utiliza árboles de decisión débiles de forma secuencial, donde cada nuevo árbol se ajusta a los errores cometidos por los árboles anteriores. (Lezama Gonzales, Cachay Boza, & Zevallos León, 2024).	Si	Adecuado para detectar fraude en datos de consumo de energía, donde las características de fraude pueden ser sutiles.
<i>Logistic Regresión</i>	Supervisado	Utiliza la función logística para estimar la probabilidad de que una instancia pertenezca a una clase particular (Lezama Gonzales, Cachay Boza, & Zevallos León, s.f.)	Si	Ideal para validación cruzada y clasificación variable fraude o No fraude

Técnica de <i>machine learning</i>	Tipo	Descripción Breve	Aplicabilidad a PNT	Relevancia para la detección de PNT
CatBoost	Supervisado	basado en árboles que está diseñado para manejar de manera eficiente variables categóricas sin requerir una codificación previa (Lezama Gonzales, Cachay Boza, & Zevallos León, s.f.)	Si	Previene el sobreajuste, gran capacidad para manejar variables categóricas como tipo de uso, actividad comercial entre otras.
<i>Fruzy Rule</i>	Supervisado No supervisado Depende el tipo de algoritmo	Es un enfoque de <i>machine learning</i> basado en lógica difusa ( <i>Fuzzy Logic</i> ).	Si	Son adecuados para modelar incertidumbre y variabilidad en datos donde los usuarios están entre los comportamientos normales y sospechosos y no son claras. Aplica categorías de clasificación de usuarios considerando la incertidumbre y la imprecisión en los datos.
<i>Tree Ensemble Learner</i>	Supervisado	Grupo de algoritmos de <i>machine learning</i> que combinan muchos árboles de decisión.	Si	Ideal para realizar predicciones de consumo en usuarios más robustas y precisas combinando varios arboles de decisión.
SMV ( <i>Support Vector Machine</i> )	Supervisado	Utilizado para clasificación, regresión y detección de valores atípicos. Efectividad en espacios de alta dimensionalidad.	Si	Ideal para la detección de fraudes en el consumo de energía eléctrica, mediante clasificación

Técnica de <i>machine learning</i>	Tipo	Descripción Breve	Aplicabilidad a PNT	Relevancia para la detección de PNT
<i>K-Nearest Neighbors</i> (KNN)	Supervisado	clasifica un punto en función de las etiquetas de sus vecinos más cercanos (con menor distancia) realizando una votación por mayoría entre los k puntos seleccionados.	Si	binaria e identificación de <i>outliers</i> . Ideal para el manejo de datos etiquetados de histórico de usuarios etiquetados (casos confirmados de fraude y no fraude).
<i>Modelo Decisión Tree Classifier</i> (DTC)	Supervisado	Utilizado para clasificación y regresión . Clasifica datos siguiendo reglas de decisiones.	Si	Ideal para a través del consumo clasificar si un usuario es normal o usuario fraude
<i>Multi-layer Perceptron classifier</i> (MLP)	Supervisado	Tipo de red neuronal artificial que esta compuesto de multiples capas de nodos (entrada, salida, capas ocultas)	Si	Ideal para identificar patrones como procesar combinaciones complejas a apartir de datos de consumo promedio, ubicación , historial para obtener como salida: 0 = cliente normal, 1 = cliente sospechoso de pérdida no técnica.
<i>GBosting (Gradient Boosting)</i>	Supervisado	Se usa tanto en clasificación como en regresión basados en la combinación de modelos predictivos débiles como arboles de decisión para construir un modelo fuerte	Si	Eficaz para fraudes complejos y optimización de predicciones.

Técnica de <i>machine learning</i>	Tipo	Descripción Breve	Aplicabilidad a PNT	Relevancia para la detección de PNT
<i>Autoencoders</i>	Semi supervisado	Tipo de red neuronal	Si	Apartir de sólo conocer los usuarios normales en su entrada. Este tipo de red neuronal aprende y detecta anomalías en su salida final.
RNA <i>Keras (Red Neuronal Artificial Keras)</i>	Supervisado	Implementación práctica de una red neuronal usando <i>Keras</i> , una de las librerías más populares de <i>Python</i> para construir y entrenar modelos de deep learning.	Si	Ideal para aprender relaciones complejas y no lineales entre muchas variables por ejemplo (consumo, ubicación, historial de cortes, tipo de cliente, etc.).
Redes neuronales artificiales (ANN)( <i>Artificial Neural Networks</i> )	Supervisado	Basado en el funcionamiento del cerebro humano.Su finalidad es aprender a reconocer patrones complejos a partir de datos, ajustando pesos entre nodos (neuronas) conectados en una estructura de capas.	Si	Eficaces para encontrar patrones complejos de consumo en grandes volúmenes de datos teniendo variables como consumo, históricos de consumo ,numero de interrupciones, datos socioeconomicos entre otras.
Detección de NTL basada en ensambles de Clasificadores ( <i>Bagging, Boosting</i> )	Supervisado No supervisado Depende de las combinaciones entre modelos.	Un ensamble es una técnica donde varios clasificadores (modelos) son combinados para producir una mejor predicción que cualquiera de los clasificadores individuales por separado.	Si	De acuerdo al resultado de la mayoría de los modelos entrenados, el resultado final se obtiene a partir de la decisión colectiva de los distintos modelos entrenados.Ejempl

Técnica de <i>machine learning</i>	Tipo	Descripción Breve	Aplicabilidad a PNT	Relevancia para la detección de PNT
Detección de NTL basada en redes neuronales y aprendizaje profundo (Redes neuronales de convolución – CNN), (Redes neuronales recurrentes – RNN)	Supervisado No supervisado	Las técnicas de aprendizaje profundo como las CNN y RNN permiten manejar tipos específicos de datos con más eficacia, como imágenes (para análisis visual de medidores) o series temporales (para patrones de consumo a lo largo del tiempo).	Parcial	o resultado Final: fraude o no fraude CNN adecuado para patrones espaciales o cuando los datos se pueden representar como imágenes, además posee capacidad de detectar patrones jerárquicos y complejos. Además pueden ayudar a detectar patrones en los datos de consumo que podrían ser indicativos de fraude (como variaciones inusuales de consumo o patrones espaciales relacionados con el fraude), también pueden ser útiles para representar datos como matrices o secuencias, Por otro lado, las RNN pueden capturar dependencias temporales, puede recordar comportamientos pasados (por ejemplo, si un cliente ha tenido un comportamiento de consumo anómalo en los días

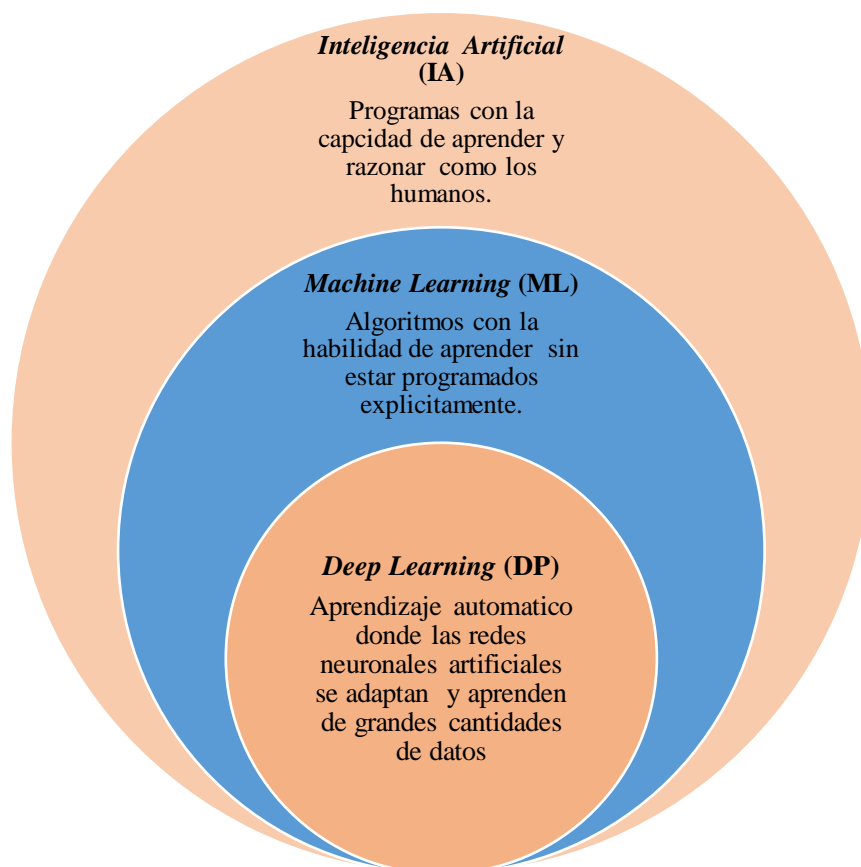
Técnica de <i>machine learning</i>	Tipo	Descripción Breve	Aplicabilidad a PNT	Relevancia para la detección de PNT
Redes Neuronales Profundas (DNN)( <i>Deep Neural Networks</i> )	Supervisado	Éstas, tienen múltiples capas ocultas, lo que las hace capaces de aprender representaciones de alto nivel a partir de los datos de entrada. Estas redes son ideales para identificar patrones complejos que no se pueden capturar fácilmente con métodos tradicionales.	Si	<p>anteriores), lo que es crucial para predecir futuros comportamientos fraudulentos.</p> <p>Son eficientes para detectar fraudes que no siguen patrones fáciles de identificar.</p> <p>Ideal para trabajar con datos etiquetados .</p> <p>Manejan grandes volúmenes de datos, lo que es importante cuando ya que las bases de datos de usuarios del sector eléctrico por lo general son robustas.</p>
<i>Naive Bayes</i>	Supervisado	Enfoque de clasificación basado en el principio de la independencia condicional(Teorema de <i>Bayes</i> )	Si	Puede usarse para clasificar si un usuario está involucrado en fraude o no, en función de características como: Historial de consumo eléctrico, ubicación, tipo de contrato, patrones de consumo temporal.
<i>Automated machine learning</i> (Auto ML)	Supervisado No supervisado	Es un enfoque de <i>machine learning</i> que permite utilizar los conocimientos incorporados de los científicos de datos	Si	Mediante el uso de herramientas de AutoML como <i>Google Cloud AutoML</i> , TPOT, o <i>datarobot</i> , se

Técnica de <i>machine learning</i>	Tipo	Descripción Breve	Aplicabilidad a PNT	Relevancia para la detección de PNT
		sin gastar tiempo y dinero para desarrollar capacidades por si mismos, mejorando al mismo tiempo del ROI.		puede automatizar muchas de las tareas de construcción de modelos basandonos en datos etiquetados como fraude o no fraude y series temporales en muestras aleatorias.
Modelo oculto de Markov(HMM)	No supervisado  Supervisado	Es una técnica de <i>machine learning</i> , específicamente un tipo de modelo probabilístico que se utiliza principalmente para trabajar con secuencias de datos o series temporales.	Parcial	En detección de fraudes, puede ayudar a modelar comportamientos normales y detectar desviaciones de consumo o comportamientos de usuarios sospechosos.

De las técnicas mencionadas previamente, para elegir la más adecuada para su aplicación en el sector eléctrico colombiano, es importante tener en cuenta que el campo de la inteligencia artificial es muy amplio, e incluye tanto técnicas de *machine learning* como de *deep learning*, como se puede observar en la siguiente imagen:

**Figura 1**

*Representación de Campo de la Inteligencia Artificial (IA)*



Siguiendo con lo anterior, para la elección de la técnica adecuada, se debe implementar una combinación de modelos supervisados y no supervisados, dependiendo del enfoque y la naturaleza de los datos disponibles. En este sentido, los modelos supervisados son útiles cuando los datos están etiquetados, es decir, cuando se conoce qué usuarios están involucrados en fraude o manipulación. Por el contrario, los modelos no supervisados son adecuados cuando los datos no tienen etiquetas o cuando se desea explorar patrones de consumo sin asumir previamente qué es "fraude" o "no fraude". Además de comparar y evaluar los resultados de las métricas de rendimiento obtenidas en estudios previos, y según los análisis realizados, se estima que las técnicas más adecuadas para la detección de PNT en el sector eléctrico colombiano son aquellas

basadas en árboles de decisión (*Random Forest*, *Gradient Boosting* y *XGBoost*), en algoritmos de agrupamiento como *K-Means*, y en redes neuronales recurrentes como *LSTM*. Estas metodologías se destacan por su capacidad para modelar datos de alta dimensionalidad, adaptarse a patrones complejos de consumo y ofrecer altos niveles de precisión y sensibilidad en la identificación de irregularidades. Además, su implementación puede contribuir a optimizar los procesos de supervisión, reducir las pérdidas económicas y fortalecer la toma de decisiones en la gestión de redes eléctricas. Asimismo, se sugiere considerar el uso de *Support Vector Machine* (*SVM*), un algoritmo de clasificación supervisada que ha demostrado ser altamente efectivo en la identificación de patrones complejos, especialmente en contextos donde los datos presentan alta dimensionalidad y desbalance entre clases, como ocurre en los problemas de detección de pérdidas no técnicas. Además, se recomienda utilizar técnicas de clasificación y agrupación para el pronóstico del hurto de electricidad, evaluando su desempeño en un entorno real o de campo para asegurar que sean efectivas y puedan ser implementadas con éxito en la detección de hurto de electricidad en la operación diaria de las empresas del sector eléctrico colombiano.

Es crucial también monitorear constantemente los modelos y realizar ajustes cuando sea necesario.

#### **Etapas 4**

En esta fase se presentan algunos de los resultados obtenidos en base a las métricas derivadas del análisis de diversas técnicas de machine learning aplicadas a la detección de pérdidas no técnicas de energía. Para ello, se utilizó información proveniente de estudios previos en los que se implementaron múltiples enfoques supervisados y no supervisados, tales como: *Gaussian Mixture Model* (*GMM*), *K-Means*, *Clustering* jerárquico, modelo *Prophet*, Redes Neuronales *LSTM* (*Long Short-Term Memory*), *VGG16*, *Random Forest*, *XGBoost* (*Extreme*

*Gradient Boosting*), Regresión Logística, CatBoost, *Fuzzy Rule*, *Tree Ensemble Learner*, *Support Vector Machine* (SVM), *K-Nearest Neighbors* (KNN), *Decision Tree Classifier* (DTC), *Multi-Layer Perceptron Classifier* (MLPC), *Gradient Boosting* (*GBoosting*), *Autoencoders*, Red Neuronal Artificial en Keras (RNA Keras), Redes Neuronales Artificiales (ANN), así como métodos de detección basados en ensambles de clasificadores (Bagging y Boosting), y técnicas de aprendizaje profundo como Redes Neuronales Convolucionales (CNN), Redes Neuronales Recurrentes (RNN), Redes Neuronales Profundas (DNN), Naive Bayes, AutoML y el Modelo Oculto de Markov (HMM).

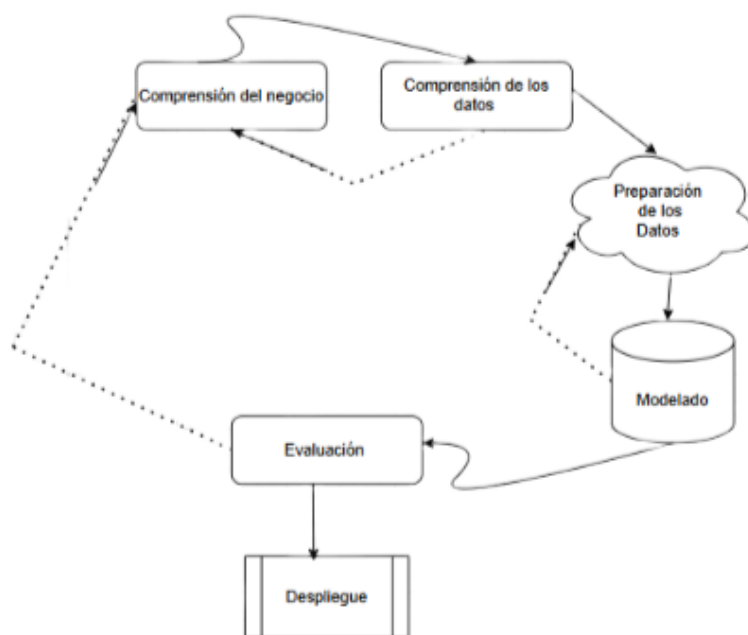
Adicionalmente, se tuvo en cuenta la metodología empleada en cada estudio, reconociendo su enfoque empírico-analítico, y destacando que los resultados obtenidos son el producto de procesos específicos de diseño, implementación y validación definidos por cada autor. La evaluación de desempeño reportada en los estudios analizados se basó principalmente en métricas como *Accuracy*, *Precision*, *Recall*, *F1-Score* y el área bajo la curva ROC (AUC-ROC), aplicando técnicas de validación cruzada para asegurar la robustez de los modelos.

En primera instancia, para el desarrollo de los estudios previos, se encontró que, al elegir cada una de las técnicas de *machine learning*, se consideraron aspectos como: el tipo de modelo a utilizar, la complejidad de implementación, el volumen de datos requeridos, la predicción a largo plazo, los recursos económicos y tecnológicos, además de seguir una metodología similar. Entre las metodologías utilizadas, destaca la metodología CRISP-DM (*Cross Industry Standard Process for Data Mining*), la cual es la más adecuada para proyectos de ciencia de datos, *machine learning* y minería de datos en cualquier país. Esta metodología es particularmente adecuada para aplicarse en proyectos de detección de PNT en el sector eléctrico colombiano, ya

que orienta todo el proceso de desarrollo de un proyecto de análisis de datos de manera organizada. Esta contempla las siguientes etapas:

## Figura 2

Diagrama Metodología CRISP-DM



Dentro de las etapas de preparación de los datos y modelado de la metodología CRISP-DM, los autores de los diferentes estudios llevaron a cabo la limpieza y preparación de los datos de consumo de electricidad antes de proceder con el análisis y modelado. El objetivo principal de estas etapas es asegurar que el conjunto de datos esté libre de errores, valores faltantes y registros inválidos, además de aplicar diversos métodos estadísticos, como el análisis de Pareto, histogramas, muestreo, análisis descriptivo de variables, correlación de variables, técnicas de balanceo de clases y la distribución t de Student, entre otros. Estas técnicas se utilizaron con el fin de comprender mejor las características de los datos y preparar adecuadamente el conjunto de entrenamiento antes de aplicar los algoritmos de *machine learning* seleccionados.

Posteriormente, los modelos fueron evaluados mediante la comparación de métricas de rendimiento como *Accuracy*, *Precision*, *Recall*, *F1-Score* y el área bajo la curva ROC (AUC-ROC), utilizando técnicas de validación cruzada para garantizar resultados más robustos.

Para el desarrollo de estos procesos, se evidenció el uso de diversas herramientas de programación y plataformas de procesamiento de datos, entre las que destacan: *Python* (con librerías como *Scikit-Learn*, *TensorFlow*, *Keras* y *TSFEL*), *AWS SageMaker*, *Apache Spark*, así como tecnologías web como *HTML* y *JavaScript* empleadas en el despliegue de soluciones o visualización de resultados. En algunos casos, también se utilizaron servicios de computación en la nube para el entrenamiento y la implementación de los modelos.

Continuando con el análisis realizado en los estudios previos, se identificó un gran número de variables que influyen en la detección de PNT en el sector eléctrico. Sin embargo, las más relevantes, y que deben ser consideradas en la implementación de los modelos, son: el consumo diario, mensual, trimestral, semestral o anual, así como los datos históricos asociados, el tipo de uso del usuario, número de revisiones, factor  $>1$ , zona, estado, fecha de revisión. A partir de estas variables, mediante técnicas de procesamiento de datos, es posible calcular medidas de resumen y generar nuevas variables derivadas, como el perfil de carga, la detección de valores atípicos, coeficiente de variación, consumo promedio, curva típica y la construcción de la variable objetivo denominada fraude.

Asimismo, es fundamental considerar variables como el tipo de uso del servicio eléctrico por parte del usuario, clasificándolo como residencial, comercial o industrial. En el contexto colombiano, esta variable adquiere una relevancia particular, ya que se ha evidenciado una mayor probabilidad de fraude en usuarios del sector industrial, especialmente en aquellos catalogados como "superusuarios" debido a su elevado nivel de consumo energético.

En el mismo contexto, resulta especialmente pertinentes las variables de tipo socioeconómico y demográfico, dado que existen zonas con una mayor incidencia de fraude o manipulación del servicio eléctrico. Estas situaciones suelen estar asociadas a factores como la alta demanda de energía, la ubicación en áreas de difícil acceso, las condiciones climáticas adversas, los bajos niveles de ingresos económicos y los problemas de orden público. Todos estos elementos influyen directamente en los patrones de consumo y en la probabilidad de ocurrencia de PNT

Adicionalmente, debe tenerse en cuenta las variables de tipo temporal ya que Colombia es un país caracterizado por una marcada diversidad cultural y regional y en muchas zonas se celebran eventos y fiestas patronales en espacios públicos, por tanto debe haber un control de parte de las empresas del servicio de energía ya que esto genera una demanda temporal de energía eléctrica. En numerosos casos, esto conlleva la realización de conexiones ilegales improvisadas para alimentar equipos de sonido, iluminación y otros dispositivos, lo que incrementa considerablemente el riesgo de PNT y se vería reflejado en el aumento de la demanda total en la zona y reducción en la facturación. Resulta fundamental considerar este tipo de situaciones dentro de los modelos predictivos, así como incorporar información sobre las suspensiones programadas del servicio eléctrico realizadas por las empresas prestadoras, ya que estas interrupciones pueden afectar el patrón normal de consumo y, en consecuencia, influir en la detección de anomalías o fraudes.

Por otro lado es importante realizar un control riguroso de los sistemas de medición, garantizando su óptimo funcionamiento tanto a nivel de *hardware* como de *software*.

Basados en lo anterior, a continuación se presenta los resultados obtenidos en las métricas de evaluación de algunos de los proyectos objeto de estudio de esta investigación:

## Proyecto 1 Detección de Instalaciones Fraudulentas Utilizando Datos Históricos de Consumo Energético

Este proyecto, desarrollado por Álvarez Castillo y Román Pereira (2024), se centró en la detección de instalaciones fraudulentas mediante el análisis de datos históricos de consumo energético. Para el entrenamiento del modelo, se utilizó el 80 % de los datos disponibles, reservando el 20 % restante para la fase de prueba. Las métricas de desempeño del modelo se calcularon empleando la matriz de confusión provista por la biblioteca scikit-learn. A partir de este análisis, se presentan a continuación los principales conceptos aplicados y los resultados obtenidos para lo cual el autor utilizó el 80 % de los datos para el entrenamiento del modelo y el 20 % restante para la fase de prueba y se calcularon las métricas de desempeño utilizando la función de matriz de confusión de scikit-learn. A continuación, se presentan los siguientes conceptos y resultados obtenidos:

**Tabla 4**

### Definición de Métricas Proyecto1

Métrica	Descripción	Formula
<i>Accuracy</i>	Mide la proporción de las predicciones correctas (verdaderos positivos, verdaderos negativos, falsos negativos)	$Accuracy = \frac{tp + tn}{tp + tn + fp + fn}$
<i>Precision</i>	Indica la proporción de casos predichos como positivos que son verdaderamente positivos (fraude).	$Precision = \frac{tp}{tp + fp}$
<i>Recall</i>	Muestra la proporción de casos positivos (Fraude) que fueron correctamente identificados con el modelo.	$Recall = \frac{tp}{tp + fn}$
<i>F1Score</i>	Es una medida de precisión balanceada entre <i>Precisión</i> y <i>Recall</i>	$F1Score = 2 * \left( \frac{precision * Recall}{Precision + Recall} \right)$

*Nota.* Adaptado de Detección de instalaciones fraudulentas utilizando datos históricos de consumo energéticos, por Álvarez Castillo y Román Pereira (s.f.).

Donde  $tp$ =trues positives,  $tn$ =trues negatives,  $fp$ =falses positivives,  $fn$ =falses negatives.

Después de definir los anteriores conceptos, a continuación se presenta en la siguiente tabla los resultados obtenidos para cada una de las técnicas de machine learning y las métricas evaluadas

**Tabla 5**

*Resultados Obtenidos en Métricas de Evaluación Proyecto1*

Metrica	SVC	LogisticRegresión	Kneighbors	RandomForest	Decision Tree	MLPClassifier
Accuracy	0.7001	0.7292	0.7559	0.8884	0.8694	0.6182
Precision	0.5497	0.6171	0.742	0.9193	0.8737	0.4775
Recall	0.8212	0.6088	0.4705	0.7487	0.7352	0.8954
F1 score	0.6586	0.6129	0.5759	0.8253	0.7985	0.6229

*Nota.* Adaptado de Detección de instalaciones fraudulentas utilizando datos históricos de consumo energéticos, por Álvarez Castillo y Román Pereira (s.f.).

De acuerdo al resultado de las métricas mostradas en la Tabla 5, *Random Forest* y *Decision Tree* según el criterio del autor tienen el mejor desempeño general (altos valores en *Accuracy*, *Precision*, *Recall*, *F1score*).

*MLPClassifier* tiene el mejor *Recall* (0.8954), lo que indica que detecta la mayoría de los fraudes (aunque su *Precision* es baja, también da falsos positivos).

*K-Neighbors* tiene buen *Precision* pero bajo *Recall* (identifica bien los fraudes detectados, pero deja pasar muchos).

*SVC* y *Logistic Regression* tienen desempeño aceptable, pero no sobresaliente.

Por tanto *Random Forest* obtuvo mejores resultados. Con exactitud en el entrenamiento de 0.84 y exactitud en la prueba de 0.86.

Además según los resultados de la matriz de confusión el modelo *RandomForestClassifier* clasificó correctamente 1037 casos de no fraude y 541 casos de fraude, además, clasificó erróneamente 53 casos como fraudulentos cuando en realidad eran no

fraudulentos, es decir los falsos positivos y 203 casos donde el modelo no logró identificar instalaciones fraudulentas, es decir los falsos negativos. (Álvarez Castillo & Román Pereira, 2024). Como se observa en la siguiente tabla:

**Tabla 6**

*Resultados Matriz de Confusión*

Variable Fraude	0	1
0 (Negativo)	1037 (tn)	53(fp)
1 (Positivo)	203 (fn)	541(tp)

*Nota.* Adaptado de Detección de instalaciones fraudulentas utilizando datos históricos de consumo energéticos, por Álvarez Castillo y Román Pereira (s.f.).

Según los resultados de las métricas de exactitud y F1-Score presentados en la Tabla 5 anterior se observa que los modelos *Support Vector Classifier* (SVC), K vecinos más cercanos (KNN), y red neuronal MLP, muestran signos de posible sobreajuste debido a diferencias notables entre las métricas de entrenamiento y prueba. Por otro lado, *Random Forest* y *Decision Tree Classifier* muestran un mejor equilibrio entre el rendimiento en entrenamiento y prueba, indicando una mejor capacidad de generalización. (Álvarez Castillo & Román Pereira, 2024.)

En concordancia con lo expuesto, se ha identificado los modelos *Random Forest* y *Decision Tree Classifier* como los más efectivos, para la detección de fraudes en el sistema eléctrico, se han identificado los modelos *Random Forest* y *Decision Tree Classifier* como los más efectivos, destacándose por su capacidad para equilibrar la precisión en la identificación de fraudes con la minimización de falsos positivos. Este equilibrio es fundamental en la gestión de fraudes, ya que el objetivo no solo es detectar transacciones fraudulentas, sino también reducir las alarmas. (Álvarez Castillo & Román Pereira, 2024).

## Proyecto 2 Detección de Faudes en Empresas de Servicios Públicos Domiciliarios

### Utilizando Ciencia de Datos y Análisis Geoespaciales

Este proyecto, desarrollado por Pachón Rodríguez (2023), empleó técnicas de ciencia de datos combinadas con análisis geoespacial para identificar patrones de fraude en empresas de servicios públicos domiciliarios. A partir de este enfoque, se obtuvieron los siguientes resultados, los cuales se presentan en la Tabla 7.

**Tabla 7**

*Resultados de las Métricas del Proyecto2*

Técnica de Machine Learning	Best Accuracy Hiperparam	F1 Score	Recall
Random Forest	58.1%	54.3%	57.6%
XGBost	57.1%	58.0%	60.6%
Regresión Logística	59.6%	58.7%	74.2%
K-Nearest Neighbors	56.9%	55.4%	56.1%
SMV(SVC)	61.1%	50.0%	50.8.%
Naive Bayes	51.6%	14.5%	8.3%
Red neuronal MLP	60.1%	57.8%	60.6%
RNA Keras	58.4%	58.5%	72.7%

*Nota.* Adaptado de Detección de fraudes en empresas de servicios públicos domiciliarios utilizando ciencia de datos y análisis geoespaciales, por M.Sc. (c) Willian Antonio Pachón Rodríguez (2023).

Con base en los resultados anteriores el autor del proyecto seleccionó el algoritmo *Support Vector Classification* (SVC) debido a que brindó mejores niveles de *F1 Score* y *Recall*, así como el menor indicador de falsos negativos. Además, no se recomienda utilizar como métrica de evaluación la exactitud porque para problemas de clasificación es necesario evaluar otras métricas calculadas a partir de la matriz de confusión evitando sobrecostos en la operación de campo (Pachón Rodríguez, 2023). Por tanto las técnicas de *machine learning* más adecuadas son SVC ya que brinda mejores resultados de *FIScore* y *Recall*.

### Proyecto 3 Detección de Pérdidas No Técnicas en Redes Eléctricas en un Contexto de Migración Tecnológica y Maximizando el Retorno Económico.

El proyecto, desarrollado por Massafarro Saquieres (2022), presenta los resultados obtenidos al entrenar los siguientes modelos: MLP (Perceptrón Multicapa), CNN (Redes Neuronales Convolucionales), LSTM (Long Short-Term Memory), LR (Regresión Logística), SVM (Máquinas de Vectores de Soporte), RF (Random Forest), GB (Gradient Boosting) y XGB (XGBoost). Los resultados de estos modelos fueron evaluados mediante las siguientes métricas:

**Tabla 8**

*Definición Métricas Proyecto3*

Metrica	Definicion
AUC-PR (Área bajo la curva <i>Precision-Recall</i> )	Mide qué tan bien el modelo mantiene alta precisión y alta recuperación al mismo tiempo. Útil cuando las clases están desbalanceadas Más alto = mejor.
AUC-ROC (Área bajo la curva ROC)	Evalúa la capacidad del modelo para distinguir entre clases (fraude vs no fraude). Más alto = mejor.
MCC ( <i>Matthews Correlation Coefficient</i> )	Métrica que toma en cuenta todos los tipos de errores (tp, tn, fp, fn). Muy buena para evaluar modelos en datos desbalanceados. Va de -1 a +1 (más cerca de 1 = mejor).
Fm: ( <i>F-measure</i> o <i>F1 Score</i> )	Promedio armónico entre <i>precisión</i> y <i>recall</i> . Se usa para balancear los falsos positivos y negativos. Más alto = mejor.

*Nota.* Adaptado de Detección de pérdidas no técnicas en redes eléctricas en un contexto de migración tecnológica y maximizando el retorno económico, por Massafarro Saquieres (2022).

Después de presentar los conceptos de las métricas evaluadas dentro del proyecto, en la Tabla 9 se presenta los resultados del proyecto sobre *Detección de pérdidas no técnicas en redes eléctricas en un contexto de migración tecnológica y maximizando el retorno económico*, aplicado a las diferentes técnicas empleadas.

**Tabla 9***Resultados Métricas Proyecto3*

Técnica de <i>machine learning</i>	PR_AUC	ROC_AUC	Fm	MCC	Precision	Recall
MLP	25.8	65.5	32.0	0.202	29.5	35.1
CNN	28.6	67.9	34.1	0.232	33.0	35.3
LSTM	28.2	66.3	34.2	0.241	35.0	32.3
LR	34.1	71.3	37.9	0.268	33.5	43.6
SVM	34.2	71.4	38.2	0.284	38.6	37.9
RF	38.8	75.2	41.3	0.310	36.1	48.3
GB	40.3	75.9	42.3	0.328	41.0	43.6
XGB	40.5	76.2	42.0	0.318	37.3	48.1
MLP	37.0	73.2	39.8	0.299	38.7	40.9
CNN	38.2	73.8	40.0	0.299	37.9	42.4
LSTM	38.4	73.6	40.4	0.301	37.3	44.2

*Nota.* Adaptado de Detección de pérdidas no técnicas en redes eléctricas en un contexto de migración tecnológica y maximizando el retorno económico, por Massaferrero Saquieres (2022).

Asimismo, aunque Gradient Boosting (GB) y XGBoost (XGB) ofrecieron mejores resultados en ciertos escenarios, en el contexto descrito por Massaferrero Saquieres (2022), se sugiere que las redes neuronales profundas DNN, como MLP, CNN y LSTM, podrían ser una mejor opción para conjuntos de datos de alta dimensión, como los provenientes de medidores inteligentes. Esto se debe a que los algoritmos basados en árboles de decisión (*Random Forest*, GB, XGB) funcionan de manera muy eficiente con datos tabulares clásicos (es decir, con pocas variables o variables claramente definidas). En contraste, las DNN tienden a sobresalir cuando los datos son más extensos, complejos o de alta dimensionalidad. Además, estas redes superan a métodos clásicos como SVM, y presentan un desempeño competitivo frente a alternativas modernas como XGB.

Por otra parte, en el estudio desarrollado por Lezama Gonzales, Cachay Boza y Zevallos León (2024), se implementaron diversos modelos para la detección de pérdidas no técnicas, cada uno con un enfoque particular en la extracción y clasificación de características. En dicho

estudio, se utilizó el modelo VGG16 para la extracción de características, seguido por algoritmos de clasificación como XGBoost, *Random Forest*, *Logistic Regression* y CatBoost. La validación del rendimiento de estos modelos se realizó mediante la evaluación de métricas clave como precisión (precision), exhaustividad (*recall*), F1-Score, Coeficiente de Correlación de Matthews (MCC), el Área bajo la Curva ROC (AUC-ROC) y exactitud (*accuracy*). Estas métricas brindan una visión integral sobre la capacidad predictiva de los modelos, particularmente en la identificación de casos de hurto de electricidad y la diferenciación entre clases.

En el contexto de esta comparación, el análisis evidencia que para el autor, *Random Forest* se posiciona como la alternativa más destacada, gracias a sus sólidas métricas de rendimiento. Con una exactitud del 95.0%, este modelo alcanza un elevado porcentaje de predicciones correctas, lo que refleja su eficacia global. El MCC alcanza un valor de 90.9%, lo que indica un rendimiento equilibrado y robusto en tareas de clasificación. Por su parte, el AUC del 95.1% subraya la capacidad del modelo para discriminar entre clases, lo cual resulta fundamental en problemas de clasificación binaria. Finalmente, el F1-Score de 94.3% evidencia un balance notable entre precisión y recall, fortaleciendo la confianza en la validez de las predicciones realizadas por el modelo ( Lezama Gonzales, Cachay Boza y Zevallos León ,2024).

En virtud de lo expuesto, es relevante señalar que, aunque se revisaron diversos estudios relacionados con la aplicación de técnicas de machine learning para la detección de pérdidas no técnicas, en este trabajo se presentan los resultados de la evaluación basada en métricas de tres de ellos. Estos estudios fueron seleccionados debido a su efectividad y pertinencia en el análisis, con el objetivo de resaltar las técnicas más eficaces y aplicables al sector eléctrico colombiano. La selección exclusiva de estos tres estudios permite preservar la concisión del trabajo, evitando la extensión innecesaria del contenido, y subraya la importancia de realizar una evaluación

basada en datos reales y métricas, lo cual es crucial para la adecuada selección de la técnica más efectiva.

## **Etapas 5**

En esta fase se proponen las siguientes estrategias para fortalecer las acciones preventivas y correctivas en la gestión de PNT en Colombia.

### **Implementación de un Área Especializada para Control y Gestión de Datos**

Crear una unidad interna dedicada exclusivamente a la detección, análisis y control de PNT, dotada de personal altamente capacitado en análisis de datos, inteligencia artificial y procesos de auditoría técnica.

Desarrollar proyectos de reducción y gestión de PNT enfocados en I+D+I (Investigación, desarrollo e innovación), con comunidades colombianas.

Definir e implementar la metodología CRISP-DM de llevar un orden en el desarrollo de proyectos de ciencia de datos y garantizar la calidad de los datos.

Aplicar técnicas de agrupamiento como *K-Means* para identificar grupos de usuarios con patrones de consumo similares, lo que facilita la detección de desviaciones o irregularidades.

Implementación de técnicas de machine learning supervisadas y no supervisadas para el análisis de los datos. Esto facilita a las empresas identificar desviaciones inusuales que podrían estar relacionadas con pérdidas no técnicas, especialmente en regiones con alta incidencia de fraude.

Adoptar herramientas avanzadas de análisis y visualización de datos como *Power Bi*, *Plotly*, *Tableau*, *Databricks*, entre otras.

Diseñar e implementar sistemas de monitoreo que permitan desarrollar modelos de *machine learning* para analizar en tiempo real los patrones de consumo eléctrico y detectar

comportamientos anómalos que puedan indicar fraude o manipulación del servicio, permitiendo a las empresas del sector eléctrico colombiano identificar en tiempo real posibles fraudes o manipulaciones del servicio.

### **Diseño e Implementación Plan Integral de Reducción de Pérdidas**

Elaborar un plan estratégico que contemple metas específicas de reducción a corto, mediano y largo plazo, priorizando zonas de Colombia de alta incidencia de pérdidas.

Integrar acciones como campañas de socialización de usuarios, programas de concientización, mejoras tecnológicas en los sistemas de medición, y fortalecimiento de los procesos de facturación y cobro. Incluir dentro del plan un sistema de monitoreo y evaluación de resultados para realizar ajustes continuos en la estrategia.

### **Reforzamiento y Modernización de la Infraestructura Tecnológica y los Sistemas de Información y Almacenamiento de Datos**

Invertir en la actualización y expansión de redes de medición inteligente (AMI), sistemas SCADA (*Supervisory Control and Data Acquisition*), telemedida y plataformas de análisis de *big data*.

Mejorar los sistemas existentes de almacenamiento y procesamiento de datos para soportar modelos de *machine learning* de forma eficiente.

Asegurar el buen funcionamiento de los sistemas de información comercial de los sistemas de medición, facturación, control operativo y administrativo garantizando una gestión integrada, y mitigar los posibles errores humanos que puedan ocasionar PNT administrativas .

### **Reformulación de Políticas Internas de la Empresa**

Proponer actualizaciones a los reglamentos internos para reforzar la identificación y sanción del fraude eléctrico en Colombia.

Impulsar la creación de unidades especializadas de análisis de pérdidas dentro de las empresas eléctricas del sector eléctrico Colombiano.

### **Destinar Recursos e Invertir en el Área Operativa y Personal Técnico**

Contratar personal especializado en análisis de datos, inteligencia artificial, *machine learning* y auditoría de consumos eléctricos.

Capacitar continuamente a los equipos técnicos en nuevas tecnologías, normativas y metodologías de detección de pérdidas no técnicas.

Dotar al área operativa de equipos modernos como dispositivos portátiles de lectura remota, equipos de comunicación de telemedida, equipos con planes de datos , y software especializado en gestión de pérdidas.

Reforzar la coordinación entre el área operativa y el área de analítica de datos, asegurando que la información de campo alimente los modelos predictivos y los sistemas de monitoreo en tiempo real.

Establecer metas e indicadores de desempeño para medir la eficiencia de las brigadas operativas en la detección y recuperación de energía no facturada.

Al integrar todas las estrategias previamente mencionadas, se logra un fortalecimiento significativo de las empresas del sector eléctrico colombiano, particularmente en lo que respecta a la gestión y control de pérdidas no técnicas. La combinación de enfoques tecnológicos avanzados, procesos operativos optimizados y herramientas de monitoreo contribuyen a una mayor eficiencia en la identificación, prevención y reducción de estas pérdidas. Esto no solo mejora la rentabilidad de las empresas, sino que también incrementa la confianza de los usuarios y fortalece la sostenibilidad del sistema eléctrico en el país. La colaboración entre actores clave del sector, así como la adopción de soluciones innovadoras, son esenciales para enfrentar los

retos actuales y futuros, garantizando un suministro más seguro y accesible para toda la población.

## Conclusiones

Es fundamental realizar una selección adecuada de variables altamente correlacionadas para la predicción y detección de pérdidas no técnicas de energía, ya que una correcta elección impacta directamente en la precisión y efectividad de los modelos.

En los estudios previos de carácter internacional se evidencia una mayor innovación tecnológica en ciencia de datos y analítica aplicada al sector eléctrico, en comparación con Colombia, donde aún se encuentra en fases iniciales de implementación de proyectos de investigación, desarrollo e innovación (I+D+i).

Los estudios analizados presentan una estructura metodológica similar en la implementación de proyectos de detección de pérdidas no técnicas de energía. Se destaca el uso de la metodología CRISP-DM, considerada la más adecuada por su enfoque ordenado y adaptable, la cual es totalmente aplicable a empresas del sector eléctrico colombiano.

Se identificó que en algunos estudios, los datos correspondientes a consumos cero fueron eliminados por considerarse irrelevantes; sin embargo, esta práctica puede ser un error, ya que estos registros podrían contener patrones indicativos de pérdidas no técnicas y deberían ser objeto de análisis detallado.

Se evidenció que el uso de herramientas de análisis de datos y *machine learning* permitió optimizar y reducir los tiempos de revisión de pérdidas no técnicas en grandes volúmenes de información, aumentando así la efectividad de los procesos de detección.

A través del uso de técnicas estadísticas, métodos de *machine learning* y validaciones adicionales, fue posible identificar y corregir inconsistencias en los conjuntos de datos, asegurando una base sólida para realizar análisis de alta calidad.

La revisión de los modelos descritos e implementados en estudios anteriores permitió ampliar el conocimiento sobre las técnicas de *machine learning*, identificando aquellas que presentan mayor efectividad para la detección de PNT.

El análisis de diversos proyectos nacionales e internacionales permitió extraer información relevante y buenas prácticas específicas para el fortalecimiento de procesos de detección de PNT en el sector eléctrico colombiano.

Aunque algunas variables no fueron relevantes para la construcción final de los modelos predictivos por su baja correlación con el fenómeno de interés, sirvieron como elementos de apoyo en la visualización, caracterización y segmentación de los usuarios, aportando valor en etapas exploratorias del análisis.

El *machine learning* representa solo una rama de la inteligencia artificial, dentro de la cual el *deep learning* también juega un papel clave. El uso combinado de ambos enfoques constituye una excelente opción para el manejo de grandes volúmenes de datos, extracción de características relevantes y obtención de resultados efectivos en la detección de PNT, siendo completamente viable su aplicación en el contexto colombiano.

Los resultados obtenidos en los estudios previos muestran que técnicas como árboles de decisión, métodos de agrupamiento (K-Means), máquinas de vectores de soporte (SVM) y redes neuronales LSTM son especialmente efectivas para la gestión de grandes volúmenes de datos, optimizando los tiempos de análisis y ofreciendo resultados precisos y robustos en la detección de pérdidas no técnicas.

El análisis exhaustivo de las técnicas de *machine learning* aplicadas a la detección de PNT en el sector eléctrico demuestra la importancia de una adecuada preparación de los datos, la selección de variables relevantes y el uso de metodologías estructuradas como CRISP-DM. Los

resultados de estudios nacionales e internacionales confirman que la innovación tecnológica, el fortalecimiento de proyectos de I+D+i y la adaptación de estrategias de análisis avanzado son factores clave para enfrentar el desafío de las pérdidas no técnicas en Colombia. La integración de tecnologías de *machine learning* y *deep Learning* no solo optimiza los procesos de detección y gestión de pérdidas, sino que también ofrece una oportunidad para transformar la eficiencia operativa del sector eléctrico colombiano, promoviendo la modernización, la sostenibilidad y la equidad en el acceso al servicio de energía. La implementación de estas prácticas, acompañada de políticas de apoyo institucional, permitirá avanzar hacia un sistema eléctrico más seguro, confiable y justo

## Recomendaciones

Conocer en profundidad el sector eléctrico donde se enfocará el proyecto, incluyendo aspectos técnicos, normativos y sociales.

Contar con fuentes de datos adecuadas, asegurando que las variables disponibles presenten un alto nivel de correlación con la detección de PNT.

Realizar un tratamiento y preparación adecuada de los conjuntos de datos, para mitigar errores en los análisis posteriores.

Optimizar y mejorar las herramientas y plataformas de análisis de datos, garantizando resultados más efectivos y confiables.

Diseñar proyectos viables y adaptados a las condiciones culturales, climáticas y socioeconómicas específicas de cada región de Colombia.

Fortalecer programas de socialización y concientización en zonas de difícil acceso o con problemas de orden público, con apoyo gubernamental.

Implementar proyectos de I+D+i en el sector eléctrico, enfocados en la detección y reducción de PNT, aplicando las técnicas de *machine learning* más efectivas supervisadas y no supervisadas, analizadas en el presente estudio.

Aumentar el despliegue de tecnologías como la medición inteligente (AMI), sistemas de supervisión remota y plataformas avanzadas de bases de datos.

Utilizar herramientas de inteligencia de negocios y plataformas de analítica avanzada para el control y monitoreo de pérdidas no técnicas.

Invertir en la capacitación de talento humano en inteligencia artificial, machine learning y redes neuronales aplicadas al sector eléctrico.

Evaluar las capacidades de generación de energía y de inversión económica en proyectos de reducción y control de pérdidas no técnicas.

Fortalecer el papel de las autoridades gubernamentales en la lucha contra las PNT.

Reformar el código penal para endurecer las penas contra el hurto de energía y agilizar los procesos administrativos y judiciales.

Mejorar los esquemas tarifarios y de contratación para facilitar la legalización de usuarios informales.

Formular indicadores de control de pérdidas, esquemas de supervisión y multas efectivas para quienes infrinjan la normatividad.

Establecer unidades independientes especializadas en la gestión y monitoreo de pérdidas no técnicas.

Reformular los planes de mantenimiento de los sistemas de medición para garantizar el cumplimiento de la normativa CREG 038 de 2014.

Realizar auditorías internas eficientes que permitan detectar fallos en componentes de los sistemas de medida, mitigando así las pérdidas derivadas de errores técnicos.

Gestionar integralmente los procesos de medición, facturación, cobro, desconexión/reconexión e inspección de medidores.

Impulsar proyectos de transición energética en zonas con alta incidencia de conexiones ilegales, fomentando el uso de energías renovables.

Estudiar el mercado de los superusuarios para clasificarlos como usuarios no regulados, permitiéndoles tarifas más accesibles y promoviendo el consumo legal de energía.

Aumentar el personal operativo en campo para la detección y control de pérdidas.

Implementar brigadas especializadas en la detección de PNT y equipos de trabajo en terreno altamente capacitados.

Crear plataformas internas en las empresas de energía para que los usuarios puedan reportar eventos ilegales.

Elaborar planes de reducción de pérdidas a largo plazo, estableciendo metas claras de corto, mediano y largo plazo que involucren a todos los actores del sector.

Se recomienda utilizar técnicas de *machine learning* supervisadas cuando se trabaje con datos etiquetados, y técnicas no supervisadas cuando los datos no contengan etiquetas.

Implementar la aplicación de las técnicas de *machine learning* analizadas y evaluadas en el presente trabajo con el fin de que las empresas prestadoras de energía del sector eléctrico colombiano reduzcan el tiempo de análisis de grandes volúmenes de datos para la detección de pérdidas no técnicas mejorando su efectividad en las revisiones.

Es fundamental asegurar la privacidad y el uso ético de los datos en Colombia, de acuerdo con lo establecido en la Ley 1581 de 2012 y las normativas emitidas por los entes regulatorios del sector eléctrico, como la CREG, la SSPD y el Ministerio de Minas y Energía (MME).

### Referencias Bibliográficas

- Alex, M. S., & Javier, P. C. (2022). *Predicción del hurto de energía eléctrica a través del uso de la inteligencia artificial mediante algoritmos de machine learning para CNEL EP unidad de negocios santo domingo*. Ecuador.
- Angulo, M. C. (s.f.). *La defraudación de fluidos en la legislación penal Colombiana*. Colombia.
- Arevalo, J. L. (2023). *Detección de pérdidas no técnicas en clientes especiales con telemedición, basada en inteligencia artificial, con aplicación en la empresa electrica Ambato*. Ecuador .
- Arévalo, N. M., Perilla, N. J., & Echeverri, L. A. (2022). *Proceso de formulación y gestión de proyectos de investigación, desarrollo e innovación (i+d+i) de acuerdo con los requisitos de la norma técnica colombiana 5802 del 2008*. UNAD.
- Bonaccorso, G. (2018). *Machine learning algorithms second edition*. Packt Publishing Ltd.
- Bustos, T. (2021). *Estrategia para la conceptualización de modelos de ia en el contexto de gestión de la energía*. España.
- Castillo, S. M., & Pereira, L. M. (2024). *Detección de instalaciones fraudulentas utilizando datos históricos de consumo*. Medellin.
- Coaguila, J. D. (2014). *Propuesta de modelo de detección de fraudes de energía eléctrica en clientes residenciales de Lima metropolitana aplicando minería de datos* . Perú.
- Colombia, G. d. (2003). *Función Pública*. Obtenido de Ley 812 de 2003:  
<https://www.funcionpublica.gov.co/eva/gestornormativo/norma.php?i=8795>
- Colombia, G. d. (2025). *Superintendencia de Servicios Públicos Domiciliarios superservicios*. Obtenido de <https://www.superservicios.gov.co/>

- Colombia, G. d. (s.f.). *Ministerio de Minas y Energia MinEnergíaCo*. Obtenido de <https://www.minenergia.gov.co/es/>
- Colombia, G. d., & Bogotá, A. m. (s.f.). *La ciencia de datos: una opción para el sector publico*. Obtenido de <https://tic.bogota.gov.co/node/450>
- Colombia, R. d. (s.f.). *DAFP, Departamento Administrativo de la Funcion Pública*. Obtenido de <http://www.funcionpublica.gov.co>
- CREG. (2011). *Planes de reducción de pérdidas no técnicas de energía eléctrica*. Bogotá.
- CREG. (2013). *Documento CREG 128*.
- CREG, C. d. (s.f.). *Gobierno de Colombia CREG*. Obtenido de <https://creg.gov.co/>
- Deloitte. (2023). *El futuro de la energía Deep Dive Digitalización*. Chile.
- Desarrollo, B. F., Rios, C., Alarcón, A., Sneyder, V., Echevarria, C., Riobo, A., . . . Irigoyen, J. (2022). *Energizados los beneficios de una herramienta basada en las tecnologías de machine learning para facilitar la detección de robo electrico* .
- Desarrollo., B. F. (2024). *Economía de las pérdidas de electricidad en América Latina y el Caribe* .
- Devia, J. S., & García, J. C. (2024). *Análisis de modelos de aprendizaje automático para la segmentación y predicción de la demanda de clientes no regulados en el mercado eléctrico colombiano*.
- Díaz, W. J. (2023). *Identificación del origen de fallas en el sistema de distribución eléctrica del Ecuador mediante la aplicación de técnicas de machine learning*. Ecuador.
- Díez, M. M., Crespo, D., & Sanchez, C. (2021). *Aplicación de machine learning al consumo electrico de edificios inteligentes*. España.

- Ecuador, R. d. (2017). *Reglamento a la ley organica del servicio público de energía eléctrica*. Ecuador.
- Ecuador, R. d. (s.f.). *Ley organica del servicio público de energía eléctrica* . Obtenido de <https://www.gob.ec/regulaciones/ley-organica-servicio-publico-energia-electrica>
- Fernando Crespo, D. e., & Tiago Alves Ferreira, ., U. (2021). *Ciencia de datos, inteligencia artificial y sus impactos sobre la sociedad*. Chile.
- Fiteni, À. M. (2023). *Detección de anomalías en consumo energético*. España.
- García, L. M., Maldonado, J. F., Troncoso, A., & Santos, J. C. (2024). *Técnicas big data para la predicción de demanda y precio eléctrico*.
- Giraldo, T. (2024). *Herramienta para la gestión y visualización de indicadores de eficiencia energética en el sector eléctrico colombiano*.
- Glauner, P., Meira, J. A., & Dolberg, L. (2017). *Neighborhood features help detecting non-technical losses in big data sets*.
- Gobierno de Colombia Gov.co, M. d. (2024). *Reglamento Técnico de Instalaciones Eléctricas - RETIE*. Obtenido de <https://www.minenergia.gov.co/es/misional/energia-electrica-2/reglamentos-tecnicos/reglamento-t%C3%A9cnico-de-instalaciones-el%C3%A9ctricas-rietie/>
- Gonzales, P. M., Boza, O. C., & León, M. Z. (2024). *Modelo predictivo para la detección del hurto de la energía eléctrica*. Perú.
- Gov.co, G. d. (1997). *Gestor Normativo Alejandría 2.0*. Obtenido de Resolución CREG 108 de 1997: [https://gestornormativo.creg.gov.co/gestor/entorno/docs/resolucion\\_creg\\_0108\\_1997.htm](https://gestornormativo.creg.gov.co/gestor/entorno/docs/resolucion_creg_0108_1997.htm)



- Mateo, F. M. (2017). *Modelo basado en minería de datos para la detección de pérdidas no técnicas de redes de distribución*. España.
- Messina, D., Lisperguer, R. C., & Pavez, R. S. (2022). *Tendencias en materia de digitalización del sector eléctrico*.
- Molina, H. E. (2023). *Estrategia para seguimiento de agresividad por pérdidas no técnicas utilizando balances de energía*. Bogotá.
- Molina, J. D. (2024). *Retos y oportunidades de la IA en el sistema eléctrico colombiano*.
- Ortega, L., Cardenas, J., & Osman, G. (2023). *Estrategias de predicción de consumo energético en edificaciones: una revisión*. Bucaramanga.
- Ortiz, J. E., & Mora, D. J. (2021). *Predicción del recurso solar diario mediante técnicas de machine learning para la proyección de generación de energía eléctrica*. Bogotá.
- Ortiz-Torres, L. F., Gómez-Luna, E., & Marlés-Sáenz, E. (2024). *Estudio del uso y contribución de la inteligencia artificial para la operación en redes eléctricas*. Cali.
- Paez, C. D. (2021). *Algoritmo para la identificación de pérdidas no técnicas en baja tensión*. Bogotá.
- Peñañiel, C. E., & Guartan, F. E. (2021). *Aplicación de técnicas de machine learning para la desagregación y pronóstico del perfil de carga en el sector industrial*. Ecuador.
- Pereirano, E. (2025). *El Pereirano noticias hoy*. Obtenido de <https://elpereirano.com/>
- Pérez, L. L., & Stanescu, C. L. (2012). *Variables determinadas para la detección de fraude en empresas de servicio eléctrico*. Venezuela.
- Perú, G. d., & Energía, M. d. (2006). *Código Nacional de Electricidad*. Perú.
- Perúano, E. (s.f.). *Osinerghmin organismo supervisor de la Inversión en energía y minería*. Obtenido de <https://www.gob.pe/osinerghmin>

- Porras, J. A. (2014). *Identificación de pérdidas en sistemas de energía mediante aplicación de técnicas de análisis y visualización de información.*
- Pública, G. d. (1994). *Ley 142 de 1994.* Obtenido de <https://www.funcionpublica.gov.co/eva/gestornormativo/norma.php?i=2752>
- Pública, G. d. (2000). *Ley 599 de 2000.* Obtenido de <https://www.funcionpublica.gov.co/eva/gestornormativo/norma.php?i=6388>
- Quiñonez, F., Sanchez, F., & Avendaño, J. (2014). *Estudio de viabilidad para el control y disminución sostenible del índice de pérdidas técnicas y no técnicas en el sistema eléctrico del Casanare . Bucaramanga .*
- Ramírez, C. Y. (2022). Uso de algoritmos de aprendizaje automático para analizar datos de energía eléctrica facturada. Caso: Chile 2015 – 2021. *Revista de I+D Tecnológico.*
- Reyes, J. D., & Moreno, L. E. (2019). *Sistema conceptual para implementar un algoritmo de detección y ubicación de fraudes eléctricos en un sistema de distribución.* Bogotá.
- RIF, F. (2025). *Cultura de la calidad en Venezuela Fondonorma.* Obtenido de <https://www.fondonorma.org.ve/index.php/es/>
- Robert, A., & Kaledio Potter, L. F. (2024). *Machine learning applications in electric power systems: enhancing efficiency, reliability, and sustainability.*
- Rodriguez, W. A. (2023). *Detección de fraudes en empresas de servicios públicos domiciliarios utilizando ciencia de datos y análisis geoespaciales.* Bogotá.
- S.A, X. (s.f.). *XM Colombia.* Obtenido de <https://www.xm.com.co/proveedores/informaci%C3%B3n-de-inter%C3%A9s/glosario>
- S.A.S., L. R. (2025). *LR.* Obtenido de <https://www.larepublica.co/>

- Sánchez, A. S. (2023). *Predicción de la demanda de energía eléctrica usando modelos de inteligencia artificial para series temporales*. Medellín.
- Saquires, P. M. (2022). *Detección de pérdidas no técnicas en redes electricas en un contexto de migración tecnológica y maximizando el retorno económico*. Motevideo .
- SebastianRaschkaVahi. (2017). *PythonMachineLearning-SecondEditionUnloc*.
- Solis-Mora, V. S., & Gruezo-Valencia, D. F. (2022). *La inteligencia artificial al servicio de la eficiencia energética en Ecuador*. Ecuador.
- Tibaquirá, J., Restrepo, A., Caicedo, E., Morales, W., Ospina, S., Duque, M., . . . Ospina, M. (2024). *Integra herramienta para la gestión y visualización de indicadores de eficiencia energética en el sector electrico colombiano*. Colombia.
- Tiempo, E. (2025). *El Tiempo*. Obtenido de <https://www.eltiempo.com/>
- UNAD. (2024). *Universidad Abierta y a Distancia*. Obtenido de <https://repository.unad.edu.co/handle/10596/67258>
- Venezuela, C. C. (1997). *Código eléctrico nacional Fonfonorma 200:2004*. Venezuela.
- Vidarte, G. A. (2023). *Análisis de información para la reducción de pérdidas no técnicas utilizando herramientas tecnológicas en el sector eléctrico*. Bogotá.
- Villegas, A. F., & Uribe, K. A. (2013). *Minería de datos aplicada a la detección de clientes con alta probabilidad de fraudes en sistemas de distribución*. Pereira.
- Willey, J. (s.f.). *Data sciene & big data analytics*. Obtenido de El libro: <https://elibro-net.bibliotecavirtual.unad.edu.co/es/ereader/unadenglish/182718?page=20>
- XM, C. C. (s.f.). *Colombia Inteligente*. Obtenido de <https://colombiainteligente.org/>