

REVISIÓN DE ESTRATEGIAS DE REDUCCIÓN DEL ÍNDICE DE PÉRDIDAS DE LA  
ELECTRIFICADORA DE SANTANDER MEDIANTE ANÁLISIS DE DATOS DE  
INFRAESTRUCTURA DE MEDICIÓN AVANZADA (AMI)

Ing. Laura Marcela Delgado Blanco

Ing. Paula Angelica Torres Pabón



Universidad Industrial de Santander

Facultad de Ingenierías Físico-mecánicas

Escuela de Ingeniería Eléctrica, Electrónica y de Telecomunicaciones

Especialización en Sistemas de Distribución de Energía Eléctrica

Bucaramanga

2025

REVISIÓN DE ESTRATEGIAS DE REDUCCIÓN DEL ÍNDICE DE PÉRDIDAS DE LA  
ELECTRIFICADORA DE SANTANDER MEDIANTE ANÁLISIS DE DATOS DE  
INFRAESTRUCTURA DE MEDICIÓN AVANZADA (AMI)

Ing. Laura Marcela Delgado Blanco

Ing. Paula Angelica Torres Pabón

Monografía presentada como requisito para optar al título de Especialista en Sistemas de  
Distribución de Energía Eléctrica

Directora:

PhD. Yulieth Jiménez Manjarrés

Universidad Industrial de Santander

Facultad de Ingenierías Físico-mecánicas

Escuela de Ingeniería Eléctrica, Electrónica y de Telecomunicaciones

Especialización en Sistemas de Distribución de Energía Eléctrica

Bucaramanga

2025

## **DEDICATORIA**

A Dios, por otorgarme salud y sabiduría para superar cada paso de este proceso.

A mis padres, quienes me enseñaron el valor de la perseverancia y la dedicación, impulsándome a seguir adelante sin rendirme.

A mi pareja, por su amor incondicional, su apoyo constante y su comprensión en los momentos difíciles. Gracias por estar siempre a mi lado, ofreciéndome su compañía y fortaleza.

*Laura Marcela Delgado Blanco*

A Dios, fuente de fortaleza y sabiduría, quien ha sido mi guía constante y luz en mi vida y particularmente, en este camino académico. A mi familia, por su amor incondicional y apoyo, que me ha sostenido en los momentos de mayor desafío. Gracias por creer en mí y motivarme a alcanzar mis sueños.

*Paula Angélica Torres Pabón*

## **AGRADECIMIENTOS**

Con la culminación de esta tesis, quiero expresar mi más sincero agradecimiento a todas las personas que, de una u otra forma, contribuyeron a que este trabajo fuera posible.

A mi compañera de tesis, por su apoyo, su ánimo constante y colaboración durante todo el proceso. Gracias por las ideas, el tiempo y el esfuerzo para que este trabajo sea posible.

A la directora de proyecto, por su guía y valiosas sugerencias, que me permitieron fortalecer este trabajo. Agradezco su dedicación y confianza en mi capacidad para llevar a cabo este proyecto.

### ***Laura Marcela Delgado Blanco***

A todos quienes hicieron parte de este proceso, un agradecimiento especial por su tiempo y apoyo.

A mi compañera, cuya dedicación, disciplina y excelente actitud fueron fundamentales para la realización de esta monografía. Su colaboración y entusiasmo siempre estuvieron presentes.

A nuestra directora de proyecto, por su liderazgo, su valioso conocimiento y su paciencia. Su guía no solo enriquecía nuestro trabajo, sino también fortalecía nuestra confianza para afrontar los retos académicos.

A la Universidad que, con sus recursos, sus profesores y su compromiso con la excelencia académica, nos brindó los conocimientos y la formación necesarios para culminar con éxito esta etapa de especialización.

### ***Paula Angélica Torres Pabón***

## TABLA DE CONTENIDO

INTRODUCCIÓN .....	13
<b>1 INTRODUCCION A LA INFRAESTRUCTURA DE MEDICIÓN AVANZADA (AMI).....</b>	<b>18</b>
1.1 Antecedentes regulatorios en Colombia .....	18
1.2 Estructura de un sistema AMI.....	21
1.2.1 Unidad de medida (UM) .....	21
1.2.2 Unidad concentradora (UC).....	22
1.2.3 Sistema de Gestión y Operación (SGO) .....	22
1.2.4 Comunicaciones .....	22
1.2.5 Seguridad .....	23
1.2.6 Interoperabilidad .....	23
1.3 Componentes y proveedores en Colombia .....	24
1.3.1 Medidores Inteligentes (Smart Meters) .....	24
1.3.2 Concentradores de Datos (Data Concentrators) .....	26
1.3.3 Tecnologías de Comunicación .....	26
1.3.4 Sistemas de Gestión AMI (Head-End Systems).....	28
1.3.5 Equipos Complementarios: .....	28
<b>2 TRATAMIENTO DE DATOS ENTREGADOS POR AMI .....</b>	<b>28</b>
2.1 Recopilación de datos AMI .....	29

2.1.1	<i>Frecuencia de recopilación</i> .....	30
2.2	Almacenamiento de Datos .....	31
2.3	Tipos de datos recopilados.....	33
2.3.1	<i>Datos de Consumo Energético</i> .....	33
2.3.2	<i>Datos de Calidad de la Energía</i> .....	33
2.3.3	<i>Datos de Estado del Medidor</i> .....	34
2.3.4	<i>Datos de Análisis de Energía Bidireccional</i> .....	34
2.3.5	<i>Datos de Comunicación</i> .....	34
2.3.6	<i>Datos de Configuración y Diagnóstico</i> .....	35
2.4	Métodos de limpieza y preprocesamiento.....	38
2.4.1	<i>Eliminación de valores atípicos</i> .....	38
2.4.2	<i>Interpolación de valores faltantes</i> .....	39
2.4.3	<i>Normalización y estandarización</i> .....	39
2.4.4	<i>Corrección de duplicados</i> .....	40
2.4.5	<i>Conversión de formatos</i> .....	40
2.4.6	<i>Validación cruzada de fuentes</i> .....	40
2.4.7	<i>Enriquecimiento de datos</i> .....	41
<b>3</b>	<b>ESTRATEGIAS DE ANALISIS DE DATOS PARA MEJORAR LOS INDICADORES DE PÉRDIDAS NO TECNICAS</b> .....	<b>41</b>
3.1	Métodos orientados a datos.....	43
3.1.1	<i>Métodos supervisados</i> .....	43

3.1.1.1	<i>Support Vector Machine (SVM)</i> .....	44
3.1.1.2	Redes neuronales artificiales (ANN).....	44
3.1.1.3	<i>Optimum Path Forrest (OPF)</i> . ....	45
3.1.1.4	Árbol de decisión. ....	46
3.1.1.5	Vecino más cercano (K-NN).....	46
3.1.1.6	Redes bayesianas.....	47
3.1.1.7	<i>Deep learning</i> . ....	48
3.1.2	<i>Métodos no supervisados</i> .....	48
3.1.2.1	Mapas Autoorganizativos (SOM).....	49
3.1.2.2	Algoritmos de Agrupamiento ( <i>Clustering</i> ). ....	49
3.1.2.3	Sistemas Expertos. ....	50
3.1.2.4	Control estadístico.....	51
3.1.2.5	Modelos de regresión.....	52
3.1.2.6	Detección de valores atípicos ( <i>Outliers</i> ).....	52
3.1.2.7	Enfoques Basados en Teoría de Juegos. ....	53
3.2	Métodos orientados a redes.....	53
3.2.1	<i>Método de flujo de carga</i> .....	54
3.2.2	<i>Método de estimación de estado</i> .....	54
3.2.3	<i>Método de red de sensores</i> .....	54
3.3	Casos de éxito en la implementación de AMI .....	55
<b>4</b>	<b>ESTRATEGIAS CON MAYOR POTENCIAL APLICABLES EN ESSA</b> .....	<b>59</b>
	CONCLUSIONES .....	67
	REFERENCIAS BIBLIOGRÁFICAS.....	69

## LISTA DE FIGURAS

<b>Figura 1</b> <i>Indicador de reducción y control de energía ESSA</i> .....	14
<b>Figura 2</b> <i>Índice de pérdidas ESSA</i> .....	15
<b>Figura 3</b> <i>Arquitectura del sistema AMI</i> .....	21
<b>Figura 4</b> <i>Proceso del tratamiento de datos entregados por AMI</i> .....	29

## LISTA DE TABLAS

Tabla 1 <i>Familias de características principales utilizadas para la detección de NTL.</i> .....	35
Tabla 2 <i>Ponderación de los criterios tenidos en cuenta para la selección.</i> .....	62
Tabla 3 <i>Comparativa entre los métodos supervisados</i> .....	63
Tabla 4 <i>Comparativa entre los métodos no supervisados</i> .....	65

## LISTA DE ABREVIATURAS

<b>AMI</b>	<i>Advanced Metering Infrastructure</i>
<b>ANN</b>	<i>Artificial Neural Networks</i>
<b>CREG</b>	Comisión de Regulación de Energía y Gas
<b>ESSA</b>	Electrificadora de Santander
<b>K-NN</b>	<i>K-Nearest Neighbors</i>
<b>LOF</b>	<i>Local Outlier Factor</i>
<b>MGD</b>	<i>Multivariate Gaussian Distribution</i>
<b>OPF</b>	<i>Optimum Path Forrest</i>
<b>OR</b>	Operador de Red
<b>PCA</b>	<i>Principal Component Analysis</i>
<b>SGO</b>	Sistema de Gestión y Operación
<b>SOM</b>	<i>Self Organizing Map</i>
<b>SVM</b>	<i>Support Vector Machine</i>
<b>UC</b>	Unidad Concentradora
<b>UM</b>	Unidad de Medida

## RESUMEN

**TÍTULO:** Revisión de estrategias de reducción del índice de pérdidas de la Electrificadora de Santander mediante análisis de datos de Infraestructura de medición avanzada (AMI)

**AUTOR:** Ing. Laura Marcela Delgado Blanco, Ing. Paula Angelica Torres Pabón

**PALABRAS CLAVE:** Medición inteligente, Infraestructura de medición avanzada, pérdidas no técnicas, *Big data*, estrategias, análisis de datos, métodos supervisados, métodos no supervisados.

### DESCRIPCIÓN:

Actualmente, las empresas de distribución de energía enfrentan pérdidas no técnicas derivadas de fraudes cometidos por usuarios en las redes eléctricas, manipulación de los medidores y/o errores administrativos en las lecturas de consumo. Estas pérdidas representan un impacto significativo para las empresas, ya que la energía consumida no se factura y, además, pueden comprometer el suministro normal del fluido eléctrico. La implementación de nuevas tecnologías al proceso de medición permite a las empresas ajustar correctamente sus políticas y estrategias para optimizar sus operaciones. En este sentido, considerando la gran cantidad de datos generados con la implementación de la Infraestructura de Medición Avanzada (AMI) en el sistema eléctrico, en esta monografía se revisan las estrategias de análisis de datos con el fin de mejorar los indicadores de pérdidas en el sector eléctrico y beneficiar a la Electrificadora de Santander. Como resultado de la revisión bibliográfica de los métodos y su aplicación para disminuir pérdidas, se ha observado que al integrar dos metodologías (supervisadas y no supervisadas) se obtienen mayores rendimientos en la detección de anomalías. Por tanto, se propone implementar las técnicas de *Clustering* y Máquinas de soporte vectorial (SVM), funcionando la primera como entrenador de la SVM y la segunda como método para la detección de pérdidas.

## **ABSTRACT**

**TITLE:** Review of loss rate reductions strategies for Electrificadora de Santander through data analysis of Advanced Metering Infrastructure (AMI)

**AUTHOR:** Eng. Laura Marcela Delgado Blanco, Eng. Paula Angelica Torres Pabón

**KEYWORDS:** Smart metering, Advanced Metering Infrastructure, non-technical losses, big data, strategies, data analysis, Supervised methods, Unsupervised methods.

### **DESCRIPTION:**

Currently, energy distribution companies face non-technical losses due to fraud committed by users in the electrical networks, meter tampering and/or administrative errors in consumption readings. These losses represent a significant impact for the companies, since the energy consumed is not billed and can also compromise the normal supply of electricity. The implementation of new technologies to the metering process allows companies to correctly adjust their policies and strategies to optimize their operations. In this sense, considering the large amount of data generated with the implementation of the Advanced Metering Infrastructure (AMI) in the electric system, this monograph reviews data analysis strategies to improve the loss indicators in the electric sector and benefit Electrificadora de Santander. As a result of the bibliographic review of the methods and their application to reduce losses, it has been observed that by integrating two methodologies (supervised and unsupervised) higher yields are obtained in the detection of anomalies. Therefore, the Clustering and Support Vector Machine (SVM) techniques are proposed, the first one as a trainer of the SVM and the second one as a method for loss detection.

## INTRODUCCIÓN

La Electrificadora de Santander S.A. E.S.P. (ESSA) es una empresa de servicios públicos dedicada a los negocios de generación, transmisión, comercialización y principalmente a la distribución de energía eléctrica, comprometida con la mejora continua de los procesos en pro de la satisfacción de los usuarios. Sin embargo, en todo el proceso energético es inevitable que en su funcionamiento se presenten pérdidas técnicas, producidas por el aumento de temperatura de los equipos durante el flujo continuo de energía eléctrica. Por ejemplo, la energía consumida en las líneas de transmisión, en los circuitos de distribución y en el proceso de transformación para la entrega del servicio eléctrico al consumidor final.

Por otra parte, existen también las pérdidas no técnicas, las cuales hacen referencia a aquella energía que no es facturada debida a las conexiones ilegales del consumidor o a la manipulación indebida de los medidores para alterar el registro de consumo, obteniendo así una cifra menor al consumo real. Las pérdidas no técnicas también se originan por operación deficiente de los equipos como lo es una mala calibración en los medidores y por deficiencias administrativas, como es el caso de realización de lecturas erróneas [1].

Para los operadores de red, las pérdidas de energía eléctrica, en especial las pérdidas no técnicas, son un aspecto importante de atender debido a que anualmente esta defraudación de fluidos genera pérdidas millonarias, afectando negativamente su rentabilidad financiera y deteriorando la calidad del servicio prestado a los usuarios.

En ESSA, en el año 2023, el Indicador de Pérdidas se ubicó en el 12,86%, incrementando 0,74 puntos respecto a 2022 y llegando a un 88% de la meta propuesta que correspondía a 11,49% como se muestra en la figura 1 [2].

### Figura 1

*Indicador de reducción y control de energía ESSA*



*Nota.* Tomada de Informe de sostenibilidad ESSA 2023.

En los últimos tres años, ESSA ha registrado los siguientes índices de pérdidas técnicas y no técnicas (ver figura 2), los cuales reflejan el esfuerzo continuo de la empresa por reducir estas cifras y lo fundamental que es tomarlo como foco de atención para los próximos años [2]:

**Figura 2**

*Índice de pérdidas ESSA.*

Índices pérdidas técnicas (%)	2021	2022	2023	Variación
IP técnicas en la red de transmisión	1,09%	1,14%	0,82%	-28%
IP técnicas de distribución	7,64%	8,69%	7,50%	-13,69%
IP no técnicas de distribución	5,05%	3,43%	5,36%	56,27%

*Nota.* Tomada de: Informe de sostenibilidad ESSA 2023.

Ahora bien, la Resolución CREG 015 de 2018, establece la metodología para la remuneración de la actividad de distribución de energía eléctrica en el Sistema Interconectado Nacional y define en el capítulo 7 la metodología para la determinación de los índices de pérdidas por nivel de tensión y los índices de pérdidas reconocidos, los cuales son porcentajes máximos que pueden ser remunerados. Además, define la metodología para los planes de gestión de pérdidas con el fin de que los OR presenten su plan de control y reducción de pérdidas [3,4].

Por tal razón, actualmente, la Electrificadora de Santander ha intensificado sus esfuerzos para implementar estrategias que reduzcan las pérdidas en el sistema, destinando aproximadamente 45 mil millones de pesos anuales para llevar a cabo estas acciones. La compañía se ha enfocado en la implementación de planes de recuperación de energía y la legalidad del suministro de energía, mediante programas y acciones de inversión que permiten identificar las pérdidas, mejorar la calidad de vida de los hogares y reducir los riesgos eléctricos.

Dentro de las soluciones que ESSA ha implementado como estrategias para reducción de pérdidas de energía se encuentran las siguientes [2]:

- **Programa medida prepago:** este programa permite el pago por adelantado del consumo de energía permitiendo a los clientes disfrutar del servicio acorde con sus capacidades económicas. En el año 2023 se intervinieron 10 asentamientos, vinculando 138 suscriptores prepago, normalizando 930 viviendas de clientes con dificultad de pago, para una gestión total de 85 MWh-mes.
- **Programa Habilitación de Viviendas:** ofrece la normalización del servicio, mejorando la calidad de vida y garantizando que sea un cliente legal de la empresa. En el año 2023 se normalizaron 4.144 viviendas, beneficiando a más de doce mil personas con una inversión de 10.145 MCOP.
- **Seguimiento a clientes industriales:** se ha instalado macromedida y telemedida en estos clientes para tener balances de energía o registros hora a hora de los consumos.
- **Intervención en transformadores con altas pérdidas:** se ha instalado medida telegestionada en transformadores de altas pérdidas y donde se ha detectado reincidencia en fraudes.

Con la gestión que ESSA ha realizado a lo largo de los años, en el año 2023 se logró recuperar 145,18 GWh-año impactando a 96.000 clientes, con un beneficio estimado de \$58.469 MCOP. Aunque lo anterior representa un buen resultado a los esfuerzos realizados, el reto continúa y se espera aumentar la eficiencia en el proceso de distribución de energía eléctrica, mejorar la calidad de la energía y reducir pérdidas al gestionar los datos de forma frecuente y detallada mediante las nuevas tecnologías de medición inteligente.

En línea con lo anterior, la implementación de la infraestructura de medición avanzada (AMI) en el sistema eléctrico, generará una gran cantidad de datos derivados de la lectura de las variables eléctricas. La obtención de estos datos permitirá a la empresa de energía mejorar su gestión, obtener información en tiempo real, identificar patrones, anomalías y áreas críticas, mantener un control más acertado de su distribución y, por consiguiente, reducir las pérdidas que ocasionan costos más elevados [5].

En este sentido, con esta monografía se busca revisar e identificar estrategias de análisis de datos generados por AMI para mejorar los indicadores de pérdidas en el sector eléctrico y proporcionar beneficios a la Electrificadora de Santander, para ello se plantearon los siguientes objetivos:

- Explorar estrategias para reducir pérdidas no técnicas mediante el análisis de datos de medición inteligente a partir de literatura científica y experiencias de otras organizaciones y proveedores.
- Identificar aquellas estrategias con mayor potencial para reducir el índice de pérdidas de la Electrificadora de Santander.

# **1 INTRODUCCION A LA INFRAESTRUCTURA DE MEDICIÓN AVANZADA (AMI)**

De acuerdo con el Ministerio de Minas y Energía, la infraestructura de medición avanzada (AMI) es “la infraestructura que permite la comunicación bidireccional con los usuarios del servicio de energía eléctrica”, que integra hardware (medidores avanzados, centros de gestión de medida, enrutadores, concentradores, antenas, entre otros), software y arquitecturas y redes de comunicaciones, que permiten la operación de la infraestructura y la gestión de los datos del sistema de distribución de energía eléctrica y de los sistemas de medida” [6].

En esencia, siendo AMI un sistema integrado con conexión bidireccional entre los consumidores y un centro de gestión de la medida, posibilita la medición, recolección y análisis en tiempo real del comportamiento de los usuarios con relación al consumo de energía eléctrica. Fundamentalmente, AMI crea un vínculo crucial entre la red eléctrica y los clientes y sus cargas, permitiendo una gestión más eficiente y mayor calidad de servicio.

## **1.1 Antecedentes regulatorios en Colombia**

En Colombia, la implementación de la infraestructura de medición avanzada (AMI) mediante un despliegue masivo requiere de una normativa sólida y clara que permita una participación equitativa de todos los interesados y promueva la inversión y el desarrollo de nuevos negocios en este ámbito. Un primer antecedente regulatorio en cuanto al sistema de medida es la Resolución CREG 025 de 1995 llamada Código de Redes, dentro del cual se establece un capítulo

con el reglamento del Código de Medida. Luego, en el año 2014, este Código de Medida fue ampliado y sustituido por la Resolución CREG 038 donde se establecen los requisitos para garantizar que las medidas sean exactas y confiables.

En 2018, el Ministerio de Minas y Energía expide la resolución 40072, modificada por las Resoluciones 40483 de 2019 y 40142 de 2020 con el fin de establecer los diferentes lineamientos, políticas y mecanismos para la implementación de la infraestructura de medición avanzada en el servicio público domiciliario de energía eléctrica.

En el año 2020, la Comisión de Regulación de Energía y Gas (CREG) publicó los proyectos de Resolución 131 y 219 en los cuales se da un mayor alcance en cuanto a las condiciones para la implementación de AMI en el Sistema Interconectado Nacional, SIN, se establecen los responsables de cada una de sus etapas y los lineamientos para garantizar un adecuado funcionamiento de esta tecnología en la prestación del servicio de energía eléctrica. Luego de la publicación de estos documentos de consulta pública y teniendo en cuenta las observaciones presentadas por cada uno de los interesados, la CREG responde de manera general los estudios y análisis realizados mediante el Documento 019 de 2021.

En 2021, se actualiza la norma técnica NTC 6079, la cual establece los requisitos mínimos que deben cumplir los sistemas AMI para su implementación en las redes de distribución [7]. Adicional, en este año, se expide la Ley de Transición Energética, cuyo objetivo es modernizar la legislación vigente y promover la transición hacia fuentes más sostenibles, es relevante destacar que la Medición Avanzada de Infraestructura (AMI) desempeña un papel fundamental, ya que

permite una medición más precisa y detallada del consumo de energía, lo que fomenta su uso eficiente. Además, facilita la integración de energías renovables no convencionales en el sistema energético nacional, proporcionando una gestión eficiente que contribuye al monitoreo y la administración de la demanda de energía en tiempo real [8,9].

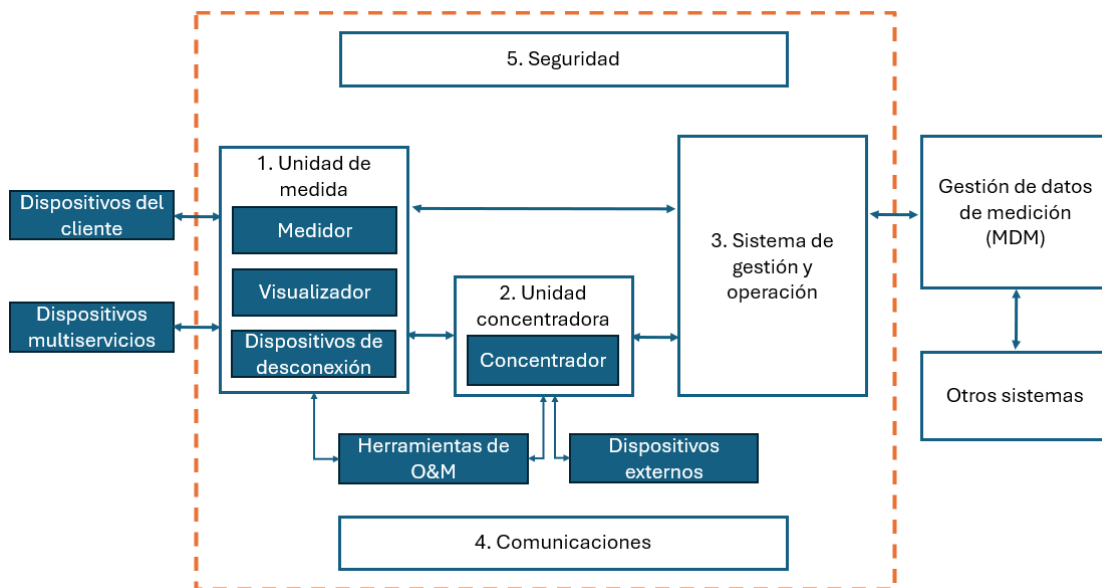
Siguiendo la línea de tiempo, en 2022 la CREG publicó la Resolución 101 001, en el cual se dictan las disposiciones generales para la implementación de la infraestructura AMI en el SIN, las responsabilidades de los prestadores del servicio y los derechos y deberes de los usuarios. Por otro lado, esta Resolución plasma los requisitos técnicos que deben tener estos sistemas de medición tomando como referencia la norma técnica NTC 6079 expedida por ICONTEC o aquella que la modifique o sustituya. Adicional y con la finalidad de iniciar un despliegue masivo de AMI, la Resolución 101 001 de 2022 establece que los OR deben plantear unos planes de despliegue constituidos por dos fases donde se detallan aspectos de costos relevantes, teniendo en cuenta que cada Operador de Red debe realizar un análisis de eficiencia y demostrar un resultado de beneficio/costo positivo para la implementación de la tecnología en la zona identificada. Por último, la Resolución presenta un marco regulatorio de la gestión de datos para garantizar la interoperabilidad y ofrecer un buen manejo, uso y protección de datos bajo condiciones de confidencialidad y ciberseguridad [10]. No obstante, la corte constitucional declaró inexecutable el inciso 2° del artículo 56 de la ley 2099 de 2021 por tanto, es conveniente que la Resolución 101 001 de 2022 sea ajustada. En línea con lo anterior, la CREG expide a consulta pública el proyecto de resolución 701 011 de 2022, donde se modifican los títulos II, V, VI y VII de la resolución CREG 101 001 de 2022 [11].

## 1.2 Estructura de un sistema AMI

Según la norma NTC 6079 del 2021, la infraestructura de medición avanzada se estructura en los siguientes cinco módulos: unidades de medida, unidad concentradora, sistema de gestión y operación, comunicaciones y seguridad (ver Figura 3).

**Figura 3**

*Arquitectura del sistema AMI.*



*Nota.* Adaptada de: Norma NTC 6079 de 2021.

### 1.2.1 Unidad de medida (UM)

La unidad de medida se compone de un medidor de energía eléctrica, el cual permite realizar registros de las variables eléctricas en el suministro de energía; un visualizador el cual muestra la energía registrada por el medidor; el dispositivo de conexión o desconexión, el cual es

capaz de interrumpir o restablecer el suministro de energía eléctrica mediante un comando enviado por el sistema de gestión y operación; dispositivos del cliente final y multiservicios, sin embargo, estos últimos no se encuentran dentro del alcance de implementación en Colombia.

### ***1.2.2 Unidad concentradora (UC)***

La unidad concentradora actúa como intermediario entre la unidad de medida y el sistema de gestión y operación (SGO). Su función principal es administrar la red para la comunicación y el intercambio de datos entre el SGO y el medidor. Además, permite habilitar la conexión de dispositivos para la operación y el mantenimiento local del concentrador.

### ***1.2.3 Sistema de Gestión y Operación (SGO)***

El sistema de gestión y operación es un software que permite realizar la gestión y el control de la información de las unidades de medida y la unidad concentradora, además dentro de su funcionalidad también permite el control, administración y operación de todos los módulos que componen la infraestructura AMI.

### ***1.2.4 Comunicaciones***

La función primordial del sistema de comunicaciones consiste en asegurar un flujo de información claro y preciso entre todos los componentes, dado que se trata de un sistema transversal. Para un correcto funcionamiento, la comunicación puede llevarse a cabo de dos maneras principales:

**Comunicación cableada:** las conexiones físicas garantizan una transmisión de información confiable utilizando tecnologías como PLC (*Power Line Communication*), puertos eléctricos o fibra óptica.

**Comunicación inalámbrica:** Esta opción brinda flexibilidad y movilidad en la comunicación, es posible realizarla a través de puertos ópticos o radiofrecuencia.

### ***1.2.5 Seguridad***

Aunque la implementación de esta nueva tecnología conlleva beneficios significativos tanto para los usuarios como para los operadores de red, no se debe dejar de lado la seguridad del sistema, ya que también introduce riesgos relacionados con la confidencialidad de datos y la vulnerabilidad del sistema. Por tanto, es fundamental considerar la seguridad transversal en todos los componentes del sistema de infraestructura AMI.

Para garantizar la integridad y confidencialidad de la información, es crucial llevar a cabo una gestión de riesgos efectiva, lo cual implica identificar y controlar las posibles alteraciones o amenazas al sistema [7].

### ***1.2.6 Interoperabilidad***

Dentro de los aspectos claves de un sistema AMI se debe mencionar la interoperabilidad, que se refiere a la capacidad del sistema para permitir que diferentes dispositivos, aplicaciones,

plataformas y tecnologías de diversos fabricantes trabajen juntos de manera efectiva y segura. Esto implica que los componentes del sistema AMI puedan intercambiar información, procesarla y utilizarla sin importar diferencias en su origen o especificaciones técnicas [12].

### **1.3 Componentes y proveedores en Colombia**

Con la implementación de AMI, el Gobierno Nacional de Colombia busca modernizar la infraestructura eléctrica, mejorar la eficiencia energética y cumplir con los lineamientos regulatorios emitidos por la Comisión de Regulación de Energía y Gas (CREG). Los equipos utilizados en estas iniciativas incluyen una variedad de dispositivos y tecnologías avanzadas que, de acuerdo con su configuración, entregan una cuantiosa cantidad de datos (*Big Data*) que requieren de un tratamiento adecuado a fin de cumplir con las iniciativas en mención, así como la detección de pérdidas no técnicas, tema que motiva la presente monografía.

A continuación, se describen algunos de los principales componentes y los proveedores más implementados a nivel país:

#### ***1.3.1 Medidores Inteligentes (Smart Meters)***

Estos dispositivos permiten medir el consumo de energía en tiempo real, recopilar datos de forma remota y enviar esta información a los operadores del sistema. Algunos de los modelos y fabricantes utilizados en Colombia incluyen:

- **Landis+Gyr:** Con opciones para medidores monofásicos y trifásicos que cumplen con estándares internacionales de comunicación. Este fabricante destaca por su solución *Gridstream*, que permite recopilar datos en intervalos tan pequeños como 5 minutos y utiliza múltiples tecnologías de comunicación como RF Mesh y PLC. Sus sistemas están diseñados para la interoperabilidad y la integración en redes inteligentes avanzadas, además de ofrecer beneficios en detección de fallas y gestión de carga [13].
- **Siemens Energy Management** ofrece soluciones integrales para la infraestructura de medición avanzada (AMI), destacando su sistema EnergyIP. Este software permite gestionar grandes volúmenes de datos de medidores inteligentes en tiempo real, habilitando funciones como la detección de pérdidas no técnicas, predicción de demandas y gestión eficiente de la red. Sus tecnologías son altamente escalables y compatibles con diversas plataformas, utilizando estándares abiertos para facilitar la interoperabilidad en redes inteligentes avanzadas [14].
- **Itron:** se especializa en soluciones de medición y gestión de datos con su plataforma *OpenWay Riva*. Este sistema combina comunicación bidireccional y análisis en tiempo real directamente en el dispositivo para optimizar la detección de anomalías, reducir pérdidas no técnicas y mejorar la eficiencia de la red. Sus soluciones aprovechan tecnologías como RF Mesh y PLC, además de integrar inteligencia distribuida para una toma de decisiones más rápida y precisa en el nivel del medidor [15].

- **Elster (ahora parte de Honeywell)**, se enfoca en soluciones avanzadas de medición con su serie de medidores inteligentes y software de gestión como EnergyAxis. Este sistema permite la recopilación de datos en intervalos cortos, utilizando tecnologías de comunicación flexibles como RF y PLC. Además, EnergyAxis está diseñado para proporcionar herramientas avanzadas de análisis de datos, lo que facilita la detección de fraudes y la optimización operativa en redes de distribución eléctrica [16].

### ***1.3.2 Concentradores de Datos (Data Concentrators)***

Estos dispositivos recopilan información de varios medidores inteligentes y la transmiten a los centros de control de las empresas de distribución eléctrica. Empresas como Schneider Electric y ABB fabrican concentradores con tecnologías de comunicación adaptadas al entorno propio de Colombia.

### ***1.3.3 Tecnologías de Comunicación***

Para garantizar la conectividad entre los medidores, concentradores y centros de control, se emplean diferentes tecnologías:

- **PLC (*Power Line Communication*)**: Se basa en la comunicación a través de líneas eléctricas existentes, es decir, los datos viajan desde los medidores inteligentes hasta concentradores o centros de control a través de la red eléctrica [17].

- RF Mesh (Red de Radiofrecuencia Mallada): Usa dispositivos de radiofrecuencia para crear una red de malla (mesh) donde cada medidor actúa como un nodo. Los datos pueden "saltar" de un nodo a otro, hasta llegar al concentrador o punto de acceso, si un nodo está fuera de servicio, la red redirige automáticamente el tráfico a través de otros nodos disponibles. Esta tecnología es de alta confiabilidad para áreas urbanas densamente pobladas [18].
- Redes celulares (4G/5G): Se emplean redes de telecomunicaciones comerciales para la transmisión de datos, ideales para áreas donde otras tecnologías son inviables como en zonas rurales, no obstante, pueden tener costos de operación más altos debido a tarifas de datos [19].
- LoRa (*Long Range*): Es una tecnología de comunicación diseñada para transmitir datos a largas distancias utilizando baja potencia, están ganando popularidad por su bajo consumo de energía y largo alcance, por ejemplo, en un sistema AMI, LoRa se utiliza para conectar medidores inteligentes distribuidos en áreas rurales o de baja densidad de población. Los medidores envían datos de consumo y alertas de anomalías al *gateway* más cercano. Este *gateway* transmite la información a los servidores centrales, donde se analizan para detectar fraudes u optimizar el suministro [20].

En proyectos de AMI, las tecnologías suelen combinarse para maximizar la cobertura y adaptarse a las características del área (por ejemplo, RF Mesh en ciudades y PLC en áreas rurales).

Esto permite recopilar datos con mayor precisión y continuidad, crucial para análisis avanzados como la detección de pérdidas no técnicas.

#### ***1.3.4 Sistemas de Gestión AMI (Head-End Systems)***

Son plataformas de software que centralizan y analizan los datos provenientes de los medidores. Algunos ejemplos de software usados son: MDM (*Meter Data Management*): Soluciones de empresas como Oracle y Siemens y SCADA adaptados a AMI: Supervisión avanzada para la operación remota.

#### ***1.3.5 Equipos Complementarios:***

- Repetidores de Señal: Para extender el alcance de la comunicación en áreas geográficamente desafiantes.
- Transformadores con monitoreo inteligente: Permiten integrar datos de calidad de la energía directamente a los sistemas AMI.

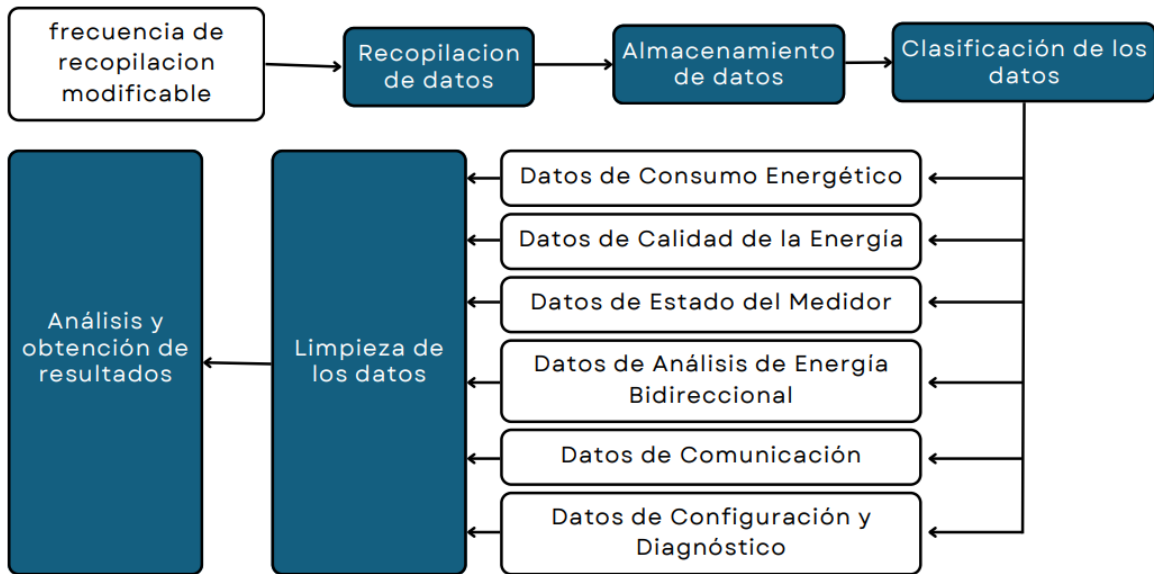
## **2 TRATAMIENTO DE DATOS ENTREGADOS POR AMI**

Uno de los beneficios de los sistemas AMI es el aprovechamiento de los datos de medida para generar información útil para los diferentes interesados. En este sentido, es fundamental garantizar la calidad, confiabilidad y utilidad de los datos derivados de los sistemas AMI mediante

los procesos de recopilación, almacenamiento, limpieza y preprocesamiento, los cuales se describen en la figura 4.

**Figura 4**

*Proceso del tratamiento de datos entregados por AMI*



## 2.1 Recopilación de datos AMI

La implementación de Infraestructura de Medición Avanzada (AMI) en redes eléctricas genera una cantidad masiva de datos en intervalos regulares, como cada 5, 15 o 30 minutos. Por ejemplo, un solo medidor inteligente configurado para recopilar datos cada 15 minutos genera 96 registros diarios. Si una red cuenta con 10,000 medidores, se estarían procesando casi un millón de registros al día. Este volumen de información representa una oportunidad sin precedentes para mejorar la eficiencia de las operaciones, detectar anomalías y reducir pérdidas no técnicas.

Sin embargo, la utilidad de estos datos depende en gran medida de su calidad y organización. Los datos recopilados pueden contener errores, duplicados, valores faltantes o inconsistencias que dificultan su análisis efectivo. Por ello, los métodos de limpieza y preprocesamiento son esenciales. Estas etapas iniciales garantizan que la información sea precisa, fiable y utilizable, facilitando la detección de fallas no técnicas como fraudes, errores de facturación o conexiones ilegales. El preprocesamiento incluye la eliminación de valores atípicos, la normalización de datos para análisis comparables y la interpolación para corregir valores faltantes. Estos pasos son clave para entrenar modelos predictivos y sistemas de detección de anomalías que permitan a los operadores identificar patrones sospechosos y priorizar intervenciones [21,22].

### ***2.1.1 Frecuencia de recopilación***

- Intervalos de recopilación estándar

**Intervalos de 15 minutos a 1 hora:** Esta es la frecuencia más común en los sistemas AMI para medir el consumo energético. Permite un equilibrio entre detalle en los datos y eficiencia en la transmisión.

**Datos en tiempo real (segundos o minutos):** En casos donde se requiere monitoreo detallado, como para la calidad de la energía o gestión de la generación distribuida, los datos pueden recogerse con mayor frecuencia.

- Frecuencia configurable

Los operadores de red pueden ajustar los intervalos según los requerimientos específicos de análisis o supervisión. Por ejemplo, en situaciones de contingencia (como fallas), los sistemas AMI pueden aumentar temporalmente la frecuencia para proporcionar datos más detallados.

- Frecuencia de transmisión de datos

Aunque los datos se recopilan con alta frecuencia, su transmisión a los sistemas centrales puede realizarse en intervalos más largos (por ejemplo, cada 6 horas o diariamente) para optimizar el uso de la red de comunicación [21,22].

## 2.2 Almacenamiento de Datos

El almacenamiento de datos recopilados por sistemas de medición avanzada (AMI) es indispensable para garantizar la efectividad de la gestión energética. Como se vio en el apartado anterior, los datos generados por los medidores inteligentes son voluminosos y requieren soluciones de almacenamiento escalables y confiables. Bajo este entender, las tecnologías más utilizadas incluyen bases de datos NoSQL, lagos de datos (*data lakes*) y almacenamiento en la nube [23,24,25].

- Bases de datos NoSQL: Estas bases de datos son ideales para manejar grandes volúmenes de datos no estructurados o semiestructurados generados por los medidores. Entendiendo que los estructurados son datos organizados en un formato definido, como tablas, los

semiestructurados tienen una estructura parcialmente definida, pero no siguen estrictamente un esquema fijo como los datos estructurados. Utilizan etiquetas o delimitadores para definir jerarquías o relaciones.

Permiten una rápida inserción y consulta de datos, lo que las hace adecuadas para aplicaciones en tiempo real como la detección de anomalías y el análisis de tendencias.

- **Lagos de datos:** son sistemas de almacenamiento diseñados para guardar grandes cantidades de datos en su formato original o en formatos semiestructurados. Los lagos de datos ofrecen un almacenamiento centralizado y flexible para grandes volúmenes de datos en su formato original. Esto es especialmente útil en el caso de AMI, ya que facilita la integración de datos de diferentes fuentes para análisis más profundos.
- **Almacenamiento en la nube:** La nube proporciona elasticidad y escalabilidad, permitiendo a las empresas de servicios públicos manejar el crecimiento exponencial de datos. Además, la nube soporta herramientas avanzadas de análisis de datos y aprendizaje automático para extraer información valiosa.

Estos enfoques de almacenamiento no solo garantizan la seguridad y la integridad de los datos, sino que también facilitan el cumplimiento de estándares de accesibilidad y eficiencia energética en el sector eléctrico [23,24,25].

## **2.3 Tipos de datos recopilados**

A continuación, se presentan y categorizan los datos recolectados por los medidores de acuerdo con su aplicación y/o enfoque.

### **2.3.1 *Datos de Consumo Energético***

- Consumo acumulado: Total de energía consumida (kWh) durante un período específico.
- Consumo en tiempo real: Energía consumida en intervalos de minutos o horas.
- Perfil de carga: Datos históricos del consumo por franjas horarias (útil para tarifas dinámicas o diferenciadas).
- Consumo máximo (demanda pico): Identificación de los momentos de mayor consumo.

### **2.3.2 *Datos de Calidad de la Energía***

- Voltaje: Valores instantáneos, promedio, mínimo y máximo registrados en el punto de medición.
- Frecuencia: Desviaciones de la frecuencia estándar (normalmente 50 Hz o 60 Hz según la red).
- Factores de potencia: Información sobre si la carga es reactiva o inductiva, lo que afecta la eficiencia de la red.
- Distorsión armónica: Detección de perturbaciones en la señal de corriente y voltaje.

### **2.3.3 *Datos de Estado del Medidor***

- Estado de la conexión: Indica si el medidor está energizado o desconectado.
- Interrupciones del servicio: Registro de eventos de apagones, su duración y frecuencia.
- Alarmas de manipulación: Alertas en caso de intentos de fraude o manipulación del medidor.
- Estado de la batería interna: En caso de tener una para respaldo.

### **2.3.4 *Datos de Análisis de Energía Bidireccional***

- Energía consumida vs. generada: En instalaciones con generación distribuida (como paneles solares), estos medidores detectan energía exportada a la red y energía importada.
- Balance energético: Datos precisos sobre flujos de energía en ambos sentidos.

### **2.3.5 *Datos de Comunicación***

- Intensidad de la señal: Calidad de la conexión entre el medidor y el sistema de comunicación (red PLC, RF Mesh, celular, etc.).
- Última comunicación exitosa: Registro del último envío exitoso de datos al concentrador o centro de control.

### 2.3.6 Datos de Configuración y Diagnóstico

- Tarifas configuradas: Información sobre las estructuras tarifarias aplicadas (pico, valle, horario nocturno).
- Identificación del medidor: Número de serie, ubicación geográfica, y datos del cliente asociado.
- Log de eventos: Historial de eventos técnicos (reinicios, actualizaciones de firmware, cambios de configuración).

De esta cantidad de información recopilada, en la tabla 1 se presenta las características principales para la detección de pérdidas no técnicas de acuerdo con Messinis y Hatziargyriou [26].

**Tabla 1**

*Familias de características principales utilizadas para la detección de NTL.*

<b>Nombre de la función</b>	<b>Descripción</b>	<b>Cómo ayuda a identificar pérdidas no técnicas</b>
<b>Promedio, Máx./Mín., Desviación estándar</b>	Estadísticas estándar calculadas para un período de tiempo específico.	Permite identificar consumos atípicos y detectar desviaciones en el consumo esperado que podrían ser indicativas de fraude o errores en la facturación.
<b>Factor de potencia/energía</b>	Relación entre el consumo de potencia activa (kW) y potencia reactiva (kVAr). Utiliza datos de alta resolución para estimaciones precisas.	Un bajo factor de potencia puede indicar consumo anómalo o manipulaciones en el medidor, sugiriendo pérdidas no técnicas.

<b>Factor de carga</b>	Relación entre consumo medio y consumo máximo de energía activa.	Permite detectar picos de consumo que no se alinean con la demanda general, sugiriendo posible manipulación o errores en la facturación.
<b>Rachas</b>	Número de veces que la curva de consumo pasa por encima y por debajo de una línea promedio.	Un alto número de fluctuaciones puede indicar irregularidades en el consumo, potencialmente señalando fraude.
<b>Consumo diario de potencia contratada</b>	Suma del consumo de energía activa en un período más la potencia contratada.	Puede ayudar a verificar que la potencia contratada esté siendo utilizada de manera coherente, detectando errores en el registro del consumo.
<b>Coefficiente de Pearson</b>	Mide la precisión de una ecuación lineal en la relación entre consumo de energía activa y tiempo.	Ayuda a identificar patrones anómalos en el consumo que no siguen la tendencia general, indicando posibles errores o manipulaciones.
<b>Coefficiente de energía facturada-consumida</b>	Diferencia de energía facturada entre energía activa consumida, dividida por la potencia contratada.	Permite verificar si la energía facturada es coherente con el consumo real. Desviaciones pueden indicar errores en la medición o fraude.
<b>kWh previstos</b>	Predicción del consumo de energía activa y comparación con el consumo medido.	Ayuda a identificar diferencias significativas entre consumo real y estimado, lo cual podría indicar fallas en el sistema o manipulaciones.
<b>Coefficientes wavelet</b>	Diferencias en coeficientes calculados a partir de la curva de consumo de años anteriores.	Detecta patrones de consumo inusuales al compararlos con datos históricos, revelando posibles errores o fraude.
<b>Coefficientes de Fourier</b>	Similar a las wavelets, utiliza la frecuencia de los coeficientes para identificar diferencias en el consumo.	Útil para analizar la periodicidad del consumo y detectar anomalías que podrían indicar errores de medición o manipulación.
<b>Coefficientes de ajuste polinomial</b>	Diferencia de los coeficientes de los polinomios que ajustan la curva de consumo comparada con años anteriores.	Permite comparar el ajuste de la curva de consumo actual con datos históricos y detectar desviaciones.

<b>Distancia euclidiana al cliente medio</b>	Mide la distancia entre la curva de consumo y la curva media de todos los consumidores.	Ayuda a identificar clientes con comportamientos de consumo atípicos, lo cual podría indicar fraudes o errores en la facturación.
<b>Pendiente de la curva de consumo</b>	Representa la inclinación de la curva de consumo ajustada con una línea recta.	Puede señalar cambios inusuales en la tendencia de consumo que podrían estar relacionados con errores de medición o manipulación.
<b>Componentes de PCA</b>	Componentes calculados de un análisis de componentes principales (PCA) que reducen la dimensionalidad de las curvas de consumo.	Facilita la identificación de patrones complejos en el consumo que podrían señalar fraudes o errores.
<b>Errores dinámicos de orden fraccionario</b>	Características que expresan la diferencia entre el uso de un medidor perfilado y el consumo en tiempo real.	Permite detectar diferencias en el uso de medidores y en el consumo registrado, revelando posibles manipulaciones.
<b>Relación de desajuste</b>	Diferencia entre el consumo medido y la suma de las mediciones de medidores inteligentes.	Identifica discrepancias que podrían ser indicadores de errores en la medición o posibles fraudes.
<b>Tasas de consumo estacionales</b>	Comparación del consumo en diferentes estaciones del año.	Ayuda a identificar cambios inusuales en el patrón de consumo, lo que podría indicar fraude o errores en la medición.
<b>Coefficientes de la transformada discreta del coseno</b>	Utiliza los coeficientes de la transformación para analizar la periodicidad del consumo.	Puede revelar patrones de consumo inusuales o anómalos.
<b>Disminución del consumo respecto al período anterior</b>	Reducción de porcentaje en el consumo comparado con períodos anteriores.	Detecta una disminución que podría relacionarse con el uso de técnicas de manipulación o errores en la lectura.
<b>Lecturas estimadas</b>	Número de lecturas que no se realizaron físicamente y que se estimaron debido a la	Ayuda a identificar patrones en las lecturas estimadas que pueden indicar errores de estimación o posible manipulación.

---

imposibilidad de acceso al  
medidor.

---

*Nota.* Reproducida de " Review of non-technical loss detection methods.", por Messinis, G. M;  
Hatziaargyriou, N. D, 2018, Electric Power Systems Research, Volumen(158), p. 250-266.

## **2.4 Métodos de limpieza y preprocesamiento.**

Los métodos de limpieza y preprocesamiento de datos derivados de AMI son fundamentales para garantizar la calidad, precisión y utilidad de los datos en la detección de pérdidas no técnicas. Los principales métodos incluyen:

### **2.4.1 *Eliminación de valores atípicos***

La eliminación de valores atípicos asegura que los datos que se encuentran significativamente fuera del rango esperado no interfieran con el análisis. Estos valores pueden ser causados por errores en los medidores o fallos de comunicación. Eliminar los valores atípicos permite que los modelos de detección sean más robustos y fiables al eliminar interferencias que puedan distorsionar el análisis.

Algunas técnicas son las siguientes: Métodos estadísticos, como percentiles o desviación estándar, los cuales permiten identificar datos atípicos. Alternativamente, algoritmos avanzados como Isolation Forest o técnicas basadas en aprendizaje automático ayudan a detectar valores anómalos en grandes conjuntos de datos [27].

### **2.4.2 Interpolación de valores faltantes**

En sistemas AMI, las lagunas de datos pueden deberse a interrupciones en la comunicación, fallas en el hardware o errores en el registro. Este método se enfoca en rellenar esos vacíos para evitar inconsistencias, garantizar la continuidad de las series temporales y mejorar la calidad del análisis. Algunas técnicas son las siguientes: Métodos simples como la interpolación lineal o el promedio, estos pueden ser útiles en casos básicos y para escenarios más complejos, las redes neuronales o métodos estadísticos avanzados como imputación múltiple son efectivos [28].

### **2.4.3 Normalización y estandarización**

Estos procesos convierten los datos a un rango común, especialmente importante para algoritmos sensibles a la escala de los valores. Mejora el desempeño de modelos de aprendizaje automático y asegura que las métricas sean comparables entre diferentes conjuntos de datos [29].

Dentro de las técnicas se encuentran las siguientes:

- *Normalización Min-Max*: Ajusta los datos para que se encuentren dentro de un rango específico, generalmente entre 0 y 1.
- *Z-score (Estandarización)*: Ajusta los datos para que tengan una media de 0 y una desviación estándar de 1.

#### **2.4.4 Corrección de duplicados**

Los registros duplicados pueden surgir por errores en el sistema o redundancias en la recopilación de datos, por tanto, esto evita que los duplicados distorsionen las métricas de consumo o generen interpretaciones incorrectas. Se utilizan técnicas como la comparación directa de registros en bases de datos pequeñas y algoritmos basados en similitud, como *fuzzy matching*, para conjuntos de datos más grandes y complejos [30].

#### **2.4.5 Conversión de formatos**

Los datos recopilados en sistemas AMI suelen estar en múltiples formatos. Por tanto, transformar estos datos en formatos estándar facilita su análisis e integración, mejora la interoperabilidad entre sistemas de análisis y asegura la consistencia. Se usan técnicas como la conversión de tipos de datos (por ejemplo, de texto a numérico) y la codificación de variables categóricas utilizando enfoques como one-hot encoding [31].

#### **2.4.6 Validación cruzada de fuentes**

Comparar datos AMI con otras fuentes asegura la precisión y consistencia, evitando análisis basados en información incorrecta, incrementando la confianza en los datos analizados y reduciendo la probabilidad de errores críticos. Se realizan comparaciones manuales para pequeños conjuntos de datos y correlación automatizada para conjuntos masivos, identificando patrones inconsistentes [32].

#### **2.4.7 *Enriquecimiento de datos***

Agregar datos contextuales, como condiciones climáticas o tarifas, ayuda a entender mejor los patrones de consumo, además, mejora la precisión de los modelos predictivos y permite un análisis más integral. En el enriquecimiento de datos se usa la integración con bases de datos externas y sistemas de sensores adicionales para recopilar datos relevantes [33].

### **3 ESTRATEGIAS DE ANALISIS DE DATOS PARA MEJORAR LOS INDICADORES DE PÉRDIDAS NO TECNICAS**

En este capítulo se analizan los diferentes enfoques para la detección de pérdidas no técnicas. Estos métodos se agrupan en tres grandes categorías: orientados a datos, orientados a redes e híbridos.

Los métodos orientados a datos son aquellos que nos permiten extraer información útil a partir de las técnicas de análisis de datos. Estos métodos se clasifican en dos categorías: los supervisados y los no supervisados. Los métodos supervisados tienen como objetivo predecir o clasificar en función de un mapeo preestablecido de las entradas, mientras que los métodos no supervisados se centran en identificar patrones o estructuras subyacentes dentro de los datos. Algunos algoritmos comunes dentro de los métodos supervisados son las redes neuronales y los

k-vecinos más cercanos, en cuanto a los métodos no supervisados se encuentran los modelos de regresión y detección de valores atípicos [26].

Los métodos orientados a redes aprovechan los datos recopilados por medidores inteligentes y consideran diversas variables, como la topología de la red y su conectividad. Estos enfoques proponen la identificación de las pérdidas no técnicas mediante técnicas como el análisis de balances energéticos, la verificación del estado de la red, la detección de errores y la optimización de la ubicación de los medidores inteligentes dentro de la red [26].

Los métodos híbridos combinan las técnicas mencionadas previamente para aprovechar las fortalezas de ambos enfoques, lo que les permite lograr una mayor precisión en la detección de pérdidas no técnicas. Estos métodos son particularmente útiles cuando se deben abordar problemas complejos o cuando la cantidad de datos disponibles es limitada [26].

Dentro del análisis de los métodos se tienen dos conceptos importantes que serán tratados a lo largo de la monografía, estos conceptos son: falso positivo y falso negativo, los cuales hacen referencia a los errores en la clasificación al evaluar los resultados de los modelos supervisados o no supervisados utilizados para identificar anomalías o fraudes [34,35].

- **Falso Positivo (FP).**

Un falso positivo ocurre cuando el modelo clasifica un caso legítimo (sin anomalías ni fraude) como si fuera una pérdida no técnica o un consumo anómalo. Estos FP pueden generar

inspecciones innecesarias, aumentando costos operativos y afectando la relación con clientes legítimos. Ejemplo en AMI: Si un cliente tiene un patrón de consumo inusual debido a un evento legítimo (como una fiesta o cambios temporales en su rutina), pero el sistema lo marca erróneamente como un caso de posible fraude [34,35].

- **Falso Negativo (FN).**

Un falso negativo ocurre cuando el modelo no detecta un caso real de anomalía o fraude, clasificándolo como un consumo legítimo, lo cual representaría pérdidas económicas directas para la empresa debido a la falta de detección de irregularidades. Ejemplo en AMI: Un cliente manipula su medidor para reducir su consumo registrado, pero el modelo no identifica este comportamiento fraudulento [34,35].

### **3.1 Métodos orientados a datos**

Como su nombre lo indica, estos métodos se basan en el análisis de datos recopilados de los usuarios con el fin de evaluar los perfiles de los clientes y ubicar los que tienen más probabilidad de ser fraudulentos [36].

#### **3.1.1 *Métodos supervisados***

Los métodos supervisados son aquellos que para su desarrollo necesitan un entrenamiento previo a partir de datos clasificados o etiquetados, estos métodos son ampliamente usados en la detección de usuarios sospechosos [36]. A continuación, se presentan algunas técnicas:

### **3.1.1.1 *Support Vector Machine (SVM).***

Las Máquinas de Vectores de Soporte (SVM) son algoritmos de aprendizaje supervisado que se basan en la idea de separar los datos mediante hiperplanos. Originalmente, las SVM se limitaban a resolver problemas de clasificación binaria o de una sola clase. Sin embargo, con el tiempo han ido evolucionando, permitiendo la clasificación multiclase y la asignación de pesos diferenciados a cada clase, lo que facilita la detección de datos atípicos [37,38].

Las SVM son métodos popularmente los más usados para la clasificación de datos y suelen presentar métricas de rendimiento altas siempre y cuando se ajuste bien a las necesidades. En [38] se utiliza el SVM para clasificar usuarios con perfiles de consumo normal y consumo sospechoso a partir de los datos de energía obtenidos de medidores inteligentes. Estos se agrupan de acuerdo con ciertas reglas como localización geográfica y tipo de cliente; el algoritmo y el análisis realizado presenta un buen resultado en la detección de los usuarios fraudulentos.

### **3.1.1.2 *Redes neuronales artificiales (ANN).***

Las redes neuronales artificiales son algoritmos diseñados para simular el funcionamiento del sistema nervioso, utilizando modelos matemáticos y computacionales. Estas redes se organizan en capas: la capa de entrada, que recibe la información del entorno; la capa oculta, encargada del procesamiento de los datos; y la capa de salida, que proporciona la respuesta generada por el

modelo. Cada capa está formada por neuronas interconectadas por pesos, que determinan el comportamiento y la influencia de cada neurona en el proceso [37,38].

Este modelo se caracteriza por su capacidad para aprender y generalizar a partir de ejemplos reales, lo que le permite adaptarse y mejorar su rendimiento en tareas complejas. Por ejemplo, en [39] se realiza el análisis para la detección de pérdidas no técnicas mediante la técnica de máquina de aprendizaje extremo o ELM por sus siglas en inglés, esta técnica usa el histórico de consumo de una red para determinar patrones de comportamiento de los clientes y producir clases que faciliten la detección de comportamientos inusuales en el consumo.

### **3.1.1.3 *Optimum Path Forrest (OPF).***

El bosque de caminos óptimos es un método que, a diferencia del SVM y las ANN, no se basa en la optimización mediante hiperplanos, sino en trayectorias. Estas trayectorias se generan a partir de muestras, también conocidas como prototipos, que actúan como las raíces del árbol de rutas óptimas. Este algoritmo destaca por su bajo tiempo de entrenamiento y es adecuado para aplicaciones de agrupamiento y clasificación, como la detección de fraudes [37,38].

En [40] se realiza una aplicación del algoritmo OPF con datos de una empresa brasilera, que los clasificaron en dos conjuntos: uno que contiene perfiles comerciales y los otros perfiles industriales; para cada conjunto se tienen datos de potencia instalada, demanda contratada, factor de carga y demanda máxima. Como resultado, muestran mayor rendimiento preciso y rápido mediante el uso del algoritmo de aprendizaje.

#### **3.1.1.4 Árbol de decisión.**

Este método consiste en una estructura que organiza condiciones y acciones de forma secuencial mediante un árbol jerárquico. Su objetivo es alcanzar un valor objetivo a partir de valores de entrada, que van formando ramas según las posibles respuestas a cada pregunta planteada en el proceso. Gracias a su simplicidad, este algoritmo es ampliamente usado en modelos de reconocimiento de patrones [37,38].

Con el fin de detectar los comportamientos anormales en los usuarios, en [41] construyen un modelo de árbol de decisión basándose en las tendencias de consumo, tasa de pérdidas y el número de alarmas al detectar un consumo inusual, sin embargo, durante la investigación se concluye buscar criterios de evaluación que ayuden en la clasificación para que su desempeño no sea fluctuante.

#### **3.1.1.5 Vecino más cercano (K-NN).**

Es un método de aprendizaje ampliamente utilizado para detectar pérdidas no técnicas. Este algoritmo consiste en incluir una muestra dentro de un conjunto con determinadas características, y asignarle la clase mayoritaria entre sus vecinos más cercanos. La asignación de la clase depende del valor de  $k$ , que determina cuántos vecinos se consideran para votar en la clasificación [37,38].

Un caso de análisis de esta técnica se presenta en [42], allí usan conjuntos de datos desequilibrados para construir modelos que puedan predecir las pérdidas no técnicas. El procesamiento de datos consiste en dividir el vecindario en cuadrículas de diferentes tamaños para calcular la proporción de clientes y pérdidas encontradas entre los usuarios inspeccionados. Al finalizar se analizan las distribuciones de las características generadas las cuales muestran utilidad para predecir pérdidas no técnicas.

### **3.1.1.6 Redes bayesianas.**

Las redes bayesianas son un método gráfico de probabilidad conjunta que se basa en las relaciones de independencia o dependencia entre las variables de un conjunto. Este método es fácil de interpretar, lo que permite identificar las características que influyen en las pérdidas no técnicas. De esta manera, se puede predecir el comportamiento de los usuarios fraudulentos de forma más precisa [37,38].

En [43] se diseña un modelo bayesiano a partir de un análisis de causa-consecuencia, en este modelo se representan gráficamente los nodos con eventos o perturbaciones del sistema y a partir de estos, el principal énfasis es encontrar la causa de cada evento para que mediante la actualización y propagación de las variables se puedan determinar las ocurrencias potenciales futuras.

### **3.1.1.7 *Deep learning*.**

El *Deep Learning* o aprendizaje profundo son algoritmos inspirados en la funcionalidad del cerebro, consiste básicamente en un algoritmo de múltiples capas conectadas y agrupadas entre sí, cuyas salidas corresponde a las entradas de la siguiente capa [44].

En [45] se implementa *Deep learning* como un método que aborda problemas de desequilibrio de los datos para la detección de las pérdidas no técnicas. Los datos usados se obtuvieron de *Korea Electric Power Corporation* (KEPCO) y *Irish Social Science Data Archive*, como resultado se obtiene un buen comportamiento y se concluye como un método robusto para la clasificación de usuarios anormales.

### **3.1.2 *Métodos no supervisados***

En el entorno del aprendizaje automático, los algoritmos de aprendizaje no supervisado trabajan directamente con datos sin etiquetas o bien, que hacen uso de una sola etiqueta, lo que significa que los datos tienen una clasificación básica, por ejemplo, clasificar un conjunto de datos de consumo eléctrico en normales o anómalos.

Aunque los métodos no supervisados no requieren etiquetas durante el entrenamiento, la "etiqueta única" puede ser un punto de referencia en fases posteriores, como al validar los resultados del modelo con métodos supervisados. Por ejemplo, detectar patrones anómalos y compararlos con casos etiquetados como fraude para confirmar la efectividad del método.

Los métodos no supervisados son útiles cuando no se dispone de datos etiquetados, como es común en escenarios donde no hay un historial completo de comportamientos fraudulentos identificados. Estos métodos se centran en identificar patrones o anomalías inherentes en los datos y son clave para detectar consumos irregulares que puedan ser indicativos de pérdidas no técnicas. A continuación, se describen los principales enfoques:

### **3.1.2.1 Mapas Autoorganizativos (SOM).**

Los SOM son redes neuronales que representan datos en un espacio de menor dimensión (normalmente 2D) para facilitar su análisis. Este enfoque permite identificar grupos de consumidores con comportamientos similares, como los que presentan patrones normales, anómalos o potencialmente fraudulentos [26].

Dentro de las ventajas que contiene el SOM se encuentran: reducir la complejidad de los datos haciéndolos visualmente interpretables y agrupar comportamientos similares, ayudando a priorizar inspecciones. Sin embargo, este es un método con una precisión baja en comparación con otros métodos y, además, los resultados requieren evaluación adicional por expertos o integración con lógica secundaria para determinar las inspecciones a realizar [26].

### **3.1.2.2 Algoritmos de Agrupamiento (*Clustering*).**

Estos algoritmos dividen a los consumidores en grupos según patrones de consumo similares. Primero, detecta consumidores con comportamientos atípicos, aunque no

necesariamente fraudulentos, también entrena clasificadores supervisados sobre grupos específicos, mejorando la precisión y reduciendo falsos positivos con el fin de generar perfiles de referencia o prototipos que ayudan a identificar desviaciones significativas en nuevos datos.

Algunas técnicas comunes de *clustering* se presentan a continuación:

- **Clustering difuso (*Fuzzy C-means*):** Asigna probabilidades de pertenencia a cada grupo, en lugar de una clasificación estricta, lo que permite mayor flexibilidad en la detección de fraudes.
- **DBSCAN (*Density-Based Spatial Clustering*):** Separa datos según su densidad y detecta comportamientos fuera de lo normal. Combinado con técnicas como PCA, permite una representación visual eficiente de los datos.

Por ejemplo, esta técnica se implementó en Brasil permitiendo identificar patrones irregulares en el consumo que indicaban posibles pérdidas no técnicas en sistemas de distribución eléctrica [46].

### **3.1.2.3 Sistemas Expertos.**

Se basan en reglas simples o complejas definidas por expertos técnicos, como inspeccionar medidores con caídas bruscas de consumo mayores al 50% o verificar medidores donde la potencia reactiva supera a la activa. Una ventaja que posee este método es que son simples y fáciles de implementar además no requieren aprendizaje previo ni entrenamiento del sistema, sin embargo,

dependen mucho de la experiencia humana y su efectividad puede ser limitada si las reglas no se actualizan según los patrones emergentes [47].

La implementación de este método en Alemania en colaboración con varias empresas europeas de distribución eléctrica condujo a mejoras significativas en la precisión de detección de pérdidas no técnicas, probadas en sistemas eléctricos europeos [47].

#### **3.1.2.4 Control estadístico.**

El análisis estadístico se basa en series temporales para detectar anomalías en el consumo eléctrico. Algunas técnicas para el control estadístico se presentan a continuación:

- **Gráficos XMR:** Monitorean la variabilidad y detectan cambios significativos en el consumo.
- **Gráficos EWMA y CUSUM:** Ideales para detecciones rápidas, aunque pueden producir falsos positivos.
- **Bandas de Bollinger:** Comparan el consumo actual con límites calculados a partir de desviaciones estándar, identificando irregularidades.

Este enfoque presenta algunas desventajas en comparación con los otros y es que no detectan fraudes si estos ocurren de manera constante desde el inicio del período monitoreado y pueden confundir cambios legítimos con anomalías, aumentando los falsos positivos [26].

### 3.1.2.5 Modelos de regresión.

Modelos como ARMA y ARIMA predicen valores esperados basándose en datos históricos no fraudulentos, comparando estas predicciones con valores medidos para identificar discrepancias. Este modelo es bien establecido, ampliamente usado, y bastante útil para series temporales, no obstante, es menos efectivo en casos de consumo irregular no representado en los datos históricos.

En Estados Unidos, se logró la implementación de sistemas de detección de fraude en redes inteligentes de distribución, reduciendo las pérdidas no técnicas y recuperando ingresos [48].

### 3.1.2.6 Detección de valores atípicos (*Outliers*).

Son técnicas de detección de valores atípicos los cuales buscan identificar datos que se desvían significativamente de patrones normales. Los métodos comunes se encuentran a continuación:

- **Distribución Gaussiana Multivariante (MGD):** Modela los datos normales y evalúa si una nueva muestra pertenece al patrón esperado.
- **Local Outlier Factor (LOF):** Evalúa la densidad local de una muestra y la compara con sus vecinos para detectar anomalías.

Este enfoque no necesita datos completamente limpios para funcionar además son muy flexibles permitiendo así ajustes personalizados según necesidades. No obstante, un LOF alto no siempre indica fraude, por lo que requieren reglas adicionales para mejorar su precisión [26].

### **3.1.2.7 Enfoques Basados en Teoría de Juegos.**

Estos enfoques modelan el fraude como un juego entre el atacante (fraudulento) y el defensor (sistema de detección). Su aplicabilidad principalmente es establece para diseñar estrategias que maximicen la eficacia del sistema frente a usuarios malintencionados. Una de las limitaciones que posee este enfoque es que, aunque parece prometedor, su implementación práctica aún está en etapas iniciales.

En Taiwán, se identificó de forma precisa clientes con consumo anómalo en micro redes de distribución [49].

## **3.2 Métodos orientados a redes**

Estos métodos aprovechan los parámetros de la red recopilados a partir de medidores instalados en puntos estratégicos, sin embargo, para las empresas distribuidoras esta implementación puede generar altos costos de inversión [36]. A continuación, se presentan algunas técnicas:

### ***3.2.1 Método de flujo de carga***

El objetivo de este método es realizar un balance energético entre las lecturas de un medidor llamado medidor observador y la suma de las mediciones de cada usuario. En [50] se analiza un modelo de flujo de carga linealizado que permite localizar el robo de energía en redes donde se desconoce la longitud de las acometidas.

### ***3.2.2 Método de estimación de estado***

Es una técnica basada en dos fases: la predicción y la actualización, la primera se realiza a partir de observaciones o mediciones del sistema y en la segunda, el estimador actualiza el estado de acuerdo con las nuevas medidas para que se ajusten a la realidad [36].

### ***3.2.3 Método de red de sensores***

El objetivo de este método es determinar la ubicación óptima y la cantidad de medidores que deben instalarse en una red, con el fin de reducir costos. Este enfoque busca mejorar la observabilidad de la red, para lo cual es necesario seleccionar medidores que, al cruzar información con los medidores inteligentes de los usuarios, permitan detectar posibles anomalías [36].

### 3.3 Casos de éxito en la implementación de AMI

Algunos casos de éxito en la implementación de estrategias basadas en análisis de datos de medición inteligente para reducir pérdidas no técnicas incluyen ejemplos de empresas tanto en Colombia como a nivel internacional:

**Electro Oriente en Perú:** Esta empresa implementó un sistema de monitoreo basado en AMI que combinaba análisis predictivo y auditorías en tiempo real para detectar conexiones ilegales y manipulación de medidores. Como resultado, lograron una reducción significativa de pérdidas no técnicas en redes de baja tensión [51].

- Reducción de pérdidas: Se logró una reducción del 3.2% en las pérdidas no técnicas.
- Tiempo de implementación: Dos años.
- Estrategia utilizada: Implementación de tecnologías de medición avanzada y detección de anomalías mediante sistemas inteligentes para identificar conexiones fraudulentas y pérdidas administrativas mediante:
  - Históricos: Comparación de datos de consumo en diferentes periodos para identificar irregularidades.
  - *Clustering*: Identificación de patrones anómalos agrupando clientes con características similares en consumo.
  - Tendencias: Uso de modelos para detectar desviaciones en las series temporales de consumo energético.
- Métodos de limpieza y preprocesamiento:

- Eliminación de duplicados en registros de medición.
- Interpolación de datos faltantes debido a errores de transmisión en medidores inteligentes.
- Normalización de datos de consumo y tarifas energéticas para garantizar consistencia en el análisis [52].

**Comisión Federal de Electricidad (CFE) en México:** La CFE implementó un programa que incluía sistemas de medición avanzada y análisis de datos para identificar fraudes y conexiones ilegales. Este programa también optimizó la lectura remota y permitió mejoras en la detección de anomalías en facturación y consumo, ayudando a reducir las pérdidas no técnicas a través de acciones específicas y oportunas [53].

- Reducción de pérdidas: Alcanzaron una disminución del 4% en las pérdidas no técnicas.
- Tiempo de implementación: Tres años.
- Estrategia utilizada: Uso de análisis avanzado de datos provenientes de medidores inteligentes, junto con campañas de regularización y estrategias legales para disminuir el fraude energético
  - Supervisadas como algoritmos de clasificación para diferenciar consumos legítimos y fraudes energéticos.
  - No supervisadas como la detección de anomalías mediante análisis de *outliers* en patrones de consumo.
- Enfoques predictivos: Modelos de regresión y series temporales para prever comportamientos atípicos.

- Métodos de limpieza y preprocesamiento:
  - Validación cruzada entre datos de AMI y lecturas manuales para corregir inconsistencias.
  - Segmentación de usuarios por categorías de consumo y características demográficas para análisis focalizados.
  - Filtrado de ruido en los datos mediante técnicas estadísticas como la media móvil [54].

**Codensa en Colombia:** Esta empresa adoptó tecnologías AMI para mejorar la detección de pérdidas no técnicas, enfocándose en áreas de alta incidencia de fraudes eléctricos. Integraron análisis de datos con sistemas de alerta temprana que facilitaron la implementación de acciones correctivas rápidas. El resultado fue una disminución considerable en el índice de pérdidas no técnicas en su zona de operación [55].

- Reducción de pérdidas: Logró una reducción del 2.5%.
- Tiempo de implementación: Un año y medio.
- Estrategia utilizada: Integración de sistemas AMI con herramientas de *big data* para detectar patrones inusuales y realizar intervenciones focalizadas mediante:
  - Históricos: Revisión y comparación de series temporales de consumo histórico.
  - Tendencias: Uso de análisis de regresión para identificar cambios en el consumo.
  - Algoritmos de *clustering* como K-means para agrupar consumos sospechosos.
- Métodos de limpieza y preprocesamiento:
  - Eliminación de valores atípicos mediante técnicas estadísticas como percentiles.

- Enriquecimiento de datos con información climática y demográfica para mejorar los análisis.
- Interpolación para rellenar lagunas en datos por fallos de comunicación de los medidores [56].

**Tata Power en India:** Con la implementación de medidores inteligentes y sistemas de análisis de datos, *Tata Power* pudo identificar y abordar patrones de uso sospechosos en redes urbanas y rurales. Sus estrategias incluyeron la segmentación de clientes y la inspección dirigida basada en datos, lo que permitió reducir significativamente las pérdidas no técnicas en áreas críticas [57].

- Reducción de pérdidas: Alcanzaron una reducción del 5.2% en las áreas intervenidas.
- Tiempo de implementación: Cuatro años.
- Estrategia utilizada: Desarrollo de algoritmos predictivos personalizados para identificar zonas de alto riesgo de pérdidas, combinado con la modernización de la infraestructura eléctrica y campañas de concienciación mediante:
  - Modelos de clasificación como *Random Forest* y *Support Vector Machines (SVM)* para identificar fraudes.
  - Algoritmos de detección de anomalías como *Isolation Forest*.
  - *Clustering*: Segmentación de zonas de alto riesgo para priorizar intervenciones.
- Métodos de limpieza y preprocesamiento:
  - Normalización y estandarización de datos para algoritmos sensibles a la escala.

- Uso de técnicas avanzadas de imputación, como redes neuronales, para datos faltantes.
- Integración de múltiples fuentes de datos, como medición AMI y SCADA, para análisis más completos [58].

#### **4 ESTRATEGIAS CON MAYOR POTENCIAL APLICABLES EN ESSA**

ESSA intenta combatir el fraude mediante inspecciones *in situ*, en las que generalmente se comprueban los aparatos de medición. Este proceso requiere una buena planificación debido a su elevado costo, que incluye la movilización, la formación y el pago del equipo de expertos, normalmente subcontratado. Además, la formación debe impartirse de forma continua.

Actualmente, ESSA está planeando llevar a cabo un proyecto piloto de infraestructura de medición avanzada, dirigido aproximadamente a 3.000 usuarios en las ciudades de Bucaramanga y Barrancabermeja. El objetivo principal de este proyecto es generar pruebas que faciliten la modernización y optimización de los sistemas de medición existentes. Para ello, se consideran elementos clave como la red de comunicación, la cual debe ser robusta y garantizar una transmisión segura y eficiente de los datos al centro de operaciones, este último, encargado de la gestión, supervisión y análisis de la información recopilada. Se espera que los datos de estos sistemas de medida puedan contribuir a esta lucha contra el fraude.

La detección temprana de pérdidas a partir de la eficiencia de la medición es importante para evitar pérdidas económicas en la compañía. Es por esta razón, que en el capítulo 3 se exploraron las diferentes estrategias de análisis de datos y cómo estas han sido aplicadas para la detección de pérdidas no técnicas en las redes de distribución de energía eléctrica en otras geografías y compañías. Sin embargo, es necesario recordar que para obtener un buen desempeño en la aplicación de los algoritmos se debe tener presente una selección óptima de las características y reducir al máximo los errores en los datos de entrada y las etiquetas de las clases.

En la literatura [26] se mencionan los métodos de aprendizaje no supervisados como técnicas adecuadas para tratar los errores en los datos y en las etiquetas de las clases, ya que estos se encargan de descubrir y agrupar los datos con patrones similares. Sin embargo, los errores en los datos son más propensos a encontrarse cuando la lectura de consumos se realiza mediante el personal técnico de la compañía. Con la implementación de AMI en las redes de la Electrificadora de Santander, estos errores disminuirían ya que los medidores inteligentes proporcionarían datos más confiables de acuerdo con el monitoreo en tiempo real de la red.

En la revisión se encuentra una gran variedad de algoritmos estudiados en diferentes países que arrojan un buen rendimiento para cumplir el objetivo propuesto. Además, se determina que la elección del método depende de varios factores, como la naturaleza de los datos, la complejidad del problema y las características específicas del entorno en el que se implementará la solución [37].

Para la elección del método más adecuado en el análisis de datos de infraestructura de medición avanzada (AMI) se evalúan criterios clave que permiten comparar el desempeño de cada técnica en tareas típicas de identificación y agrupación. A continuación, se describen las métricas principales consideradas [59,60]:

- **Precisión promedio (%):** Representa el desempeño general de un método en la identificación o agrupación correcta de patrones.
- **Tolerancia al ruido (%):** Evalúa la capacidad del método para manejar datos con variabilidad o perturbaciones no deseadas.
- **Tamaño de datos soportado:** Indica la escalabilidad del método en relación con el volumen de datos procesables.
- **Tiempo de entrenamiento (segundos):** Tiempo requerido para procesar un conjunto de datos estándar (~50,000 registros).
- **Memoria requerida (MB):** Recurso necesario para almacenar y ejecutar el modelo de análisis.
- **Interpretabilidad (%):** Mide qué tan comprensibles y explicables son los resultados obtenidos.

Para realizar una comparación efectiva, se emplean valores aproximados basados en estudios reportados en la literatura científica sobre el desempeño típico de estos métodos [61,62,63,64,65,66,67,68]. Los criterios evaluados se categorizan según rangos porcentuales, aplicables a todos los indicadores, con el fin de facilitar la ponderación y selección de la técnica más adecuada:

- Los porcentajes **superior al 90%**, reflejan alta efectividad o capacidad en el criterio evaluado bajo condiciones ideales y con datos bien preparados.
- Un rendimiento entre el **70% y el 90%**, correspondiente a un desempeño confiable en la mayoría de los escenarios típicos.
- Porcentajes **inferiores al 70%**, sugieren una efectividad limitada o una fuerte dependencia de las características específicas del conjunto de datos.

Así mismo, se asigna un peso porcentual a cada criterio con el fin de ponderar los métodos como se presenta en la Tabla 2, teniendo en cuenta la precisión y tolerancia al ruido como criterios clave para la selección.

**Tabla 2**

*Ponderación de los criterios tenidos en cuenta para la selección.*

<b>Criterio</b>	<b>Peso (%)</b>
<b>Precisión promedio</b>	30%
<b>Tolerancia al ruido</b>	20%
<b>Tamaño de datos soportado</b>	15%
<b>Tiempo de entrenamiento</b>	15%
<b>Memoria requerida</b>	10%
<b>Interpretabilidad</b>	10%

A continuación, la Tabla 3 presenta una comparación entre los métodos supervisados vistos en el capítulo 3, según los criterios detallados anteriormente.

**Tabla 3***Comparativa entre los métodos supervisados*

<b>Método</b>	<b>Precisión promedio (%)</b>	<b>Tolerancia al ruido (%)</b>	<b>Tamaño de datos soportado</b>	<b>Tiempo de entrenamiento (seg.)</b>	<b>Memoria requerida (MB)</b>	<b>Interpretabilidad (%)</b>	<b>Ponderación Total (%)</b>
<b><i>Support Vector Machine (SVM)</i></b>	90	70	Medio (10k–100k)	30	100	50	<b>76</b>
<b>Redes Neuronales (ANN)</b>	85	80	Alto (50k–1M)	120	200	40	<b>72.5</b>
<b><i>Optimum Path Forest</i></b>	88	75	Medio (10k–500k)	60	150	60	<b>75.5</b>
<b>Árbol de Decisión</b>	75	60	Bajo a Medio (1k–50k)	15	50	90	<b>70.5</b>
<b>Vecino más cercano (K-NN)</b>	80	65	Medio a Alto (10k–500k)	20	100	50	<b>69</b>
<b>Redes Bayesianas</b>	70	50	Bajo a Medio (1k–50k)	25	50	80	<b>65.5</b>
<b><i>Deep Learning</i></b>	92	85	Muy Alto (>1M)	300	500	30	<b>79.5</b>

De acuerdo con los datos presentados en la Tabla 3, se identificó que las estrategias con mayor potencial para reducir el índice de pérdidas son *Deep Learning* y *Support Vector Machine* (SVM), destacándose ambas por su alta precisión y robustez frente al ruido. Sin embargo, en el caso específico de ESSA, es crucial contar con un método que facilite la toma de decisiones y permita interpretar fácilmente los resultados. Por esta razón, se selecciona SVM, ya que ofrece un mejor desempeño en términos de interpretabilidad en comparación con *Deep Learning*, lo que lo convierte en una opción más adecuada para explicar la toma de decisiones.

Ahora bien, los SVM son algoritmos fuertes en clasificación de datos no lineales lo cual es ventajoso para la detección de pérdidas ya que estas pueden estar asociadas con patrones complejos en los datos de consumo [37]. Adicional, los sistemas AMI generan una gran cantidad de datos con muchas variables, como el consumo, hora del día, condiciones meteorológicas, entre otras. Por tanto, se necesita de un método con un buen manejo de datos de alta dimensionalidad como lo es el SVM.

En cuanto a los métodos no supervisados tratados en el numeral 3.1.2, son útiles cuando no se dispone de datos etiquetados, lo cual es común en escenarios donde no hay un historial completo de comportamientos fraudulentos identificados, ya que estos métodos se centran en identificar patrones o anomalías inherentes en los datos y son clave para detectar consumos irregulares que puedan ser indicativos de pérdidas no técnicas.

ESSA tiene dentro de sus análisis previos una base de referencia que contiene el comportamiento de usuarios fraudulentos encontrados en las filiales del grupo EPM, por tanto, de esta manera es posible alimentar el enfoque SVM con dichas muestras para su entrenamiento. Sin embargo, en caso de que estas referencias no sean lo suficientemente valiosas para la clasificación, debido a que las filiales del grupo EPM se encuentran ubicadas geográficamente en zonas donde las culturas de los usuarios son diferentes a la cultura del área de cobertura de ESSA, sería necesario implementar un enfoque no supervisado para realizar el muestreo de comportamientos anómalos propios de la geografía de Santander y así incluir mediante técnicas especiales de clasificación el SVM para reducir los indicadores de pérdidas en ESSA al detectar las pérdidas no técnicas.

La Tabla 4 presenta una comparación entre los métodos no supervisados usando la metodología de ponderación descrita anteriormente [69,70,71,72,73,74,75,76]:

**Tabla 4**

*Comparativa entre los métodos no supervisados.*

<b>Método</b>	<b>Precisión promedio (%)</b>	<b>Tolerancia al ruido (%)</b>	<b>Tamaño de datos soportado</b>	<b>Tiempo de entrenamiento (seg.)</b>	<b>Memoria requerida (MB)</b>	<b>Interpretabilidad (%)</b>	<b>Ponderación Total (%)</b>
<b>Mapas Autoorganizados (SOM)</b>	80–85	70–75	Medio (10k–100k)	30–120	100–500	50–60	<b>72</b>
<b>Clustering (K-Means, etc.)</b>	70–90	60–80	Medio a Alto (10k–500k)	10–300	50–300	60–70	<b>73.5</b>
<b>Sistemas Expertos</b>	85–90	40–60	Bajo a Medio (1k–50k)	50–200	50–100	80–90	<b>70</b>
<b>Control Estadístico (XMR, etc.)</b>	75–85	60–70	Bajo (1k–10k)	5–30	20–50	70–80	<b>72</b>
<b>Modelos de Regresión</b>	80–90	50–70	Medio a Alto (10k–100k)	10–60	50–150	50–60	<b>74</b>
<b>Detección de valores atípicos</b>	65–80	70–80	Bajo a Medio (1k–50k)	5–20	20–100	60–70	<b>70.5</b>
<b>Teoría de Juegos</b>	70–85	50–60	Bajo a Medio (1k–50k)	100–500	200–400	50–60	<b>68</b>

Como resultado de la Tabla 4, se obtiene que los métodos con mejor puntuación en su rendimiento son *Clustering* y Modelos de Regresión. Sin embargo, teniendo en cuenta que los patrones de comportamiento en ESSA son complejos y no lineales se selecciona el *Clustering*

como el modelo que permita agrupar a los usuarios de acuerdo con patrones de consumo similares, para luego ser llevado al SVM procesando los datos a partir de grupos específicos y de esta manera reducir el porcentaje de falsos positivos que se pueden presentar en el procesamiento de los datos. Se descarta el Modelo de Regresión ya que estos solo son bastante efectivos cuando las variables son lineales o casi lineales.

En [77,78,79,80,81] se realizan estudios de los diferentes enfoques para la detección de anomalías, de los cuales se indica que el *Clustering* es frecuentemente utilizado como un paso inicial para agrupar usuarios en categorías antes de aplicar técnicas supervisadas como SVM para aumentar la eficiencia del sistema. Estas combinaciones han mostrado mejoras en el rendimiento.

Por ejemplo:

- Los estudios que usan *Clustering* seguido de SVM, mejoran la precisión global hasta en un 15% respecto al uso de SVM solo [82].
- El uso de *Clustering* redujo la cantidad de datos irrelevantes, optimizando el proceso de entrenamiento del SVM [83].

Por lo anterior, se seleccionó el uso conjunto de los algoritmos *Clustering* y SVM ya que facilitan la identificación de comportamientos anómalos y mejoran la eficiencia en la detección de fraudes, lo cual resultaría una metodología robusta y efectiva para la toma de decisiones con el fin de disminuir los indicadores de pérdidas en la Electrificadora de Santander.

## CONCLUSIONES

Esta monografía permitió realizar una revisión de las diferentes estrategias de análisis de datos existentes que contribuyen positivamente en la gestión de los datos recopilados por la medición inteligente, teniendo como base la literatura científica y la experiencia de otros autores en la aplicación de estas técnicas para la detección de las pérdidas no técnicas en los sistemas de energía eléctrica.

Se identificó que las estrategias con mayor potencial para reducir el índice de pérdidas en ESSA son SVM y *Clustering*, de manera complementaria, con el fin de obtener mayor rendimiento en la detección de anomalías. El *clustering*, se selecciona como un método para la detección de patrones e identificación de comportamientos anómalos sin necesidad de datos etiquetados, además, al ser un método simple y rápido, hace fácil su implementación y a su vez posibilita la segmentación de los usuarios antes de aplicar el modelo supervisado. Por su parte, el modelo supervisado SVM es generalmente la opción preferida para la detección de pérdidas no técnicas a partir de datos de AMI, debido a su capacidad para manejar relaciones complejas en los datos (naturaleza de los datos de AMI por ser datos multidimensionales), su robustez en problemas no lineales y su buena precisión en clasificación, destacando su desempeño en términos de interpretabilidad.

La adopción de estas metodologías podría contribuir significativamente a la disminución del índice de pérdidas no técnicas en la Electrificadora de Santander, optimizando los recursos y mejorando la eficiencia operativa, lo cual resultaría en beneficios tanto económicos como sociales

para la empresa y sus usuarios. En este contexto, resulta fundamental continuar con la investigación y la implementación de soluciones tecnológicas innovadoras, para lograr una gestión más eficiente y sostenible.

## REFERENCIAS BIBLIOGRÁFICAS

- [1] Gomez Lopez, J.M; Castan Luna, R; Montero Cervantes, J.C; Meneses Ruiz, J; Garcia Hernandez, J. (2015). Aplicación de tecnologías de medición avanzada (AMI) como instrumento para reducción de pérdidas. Boletín IIE. Octubre-diciembre. <https://www.researchgate.net/publication/297280966> Aplicacion de tecnologias de medicion avanzada AMI como instrumento para reduccion de pérdidas
- [2] Electrificadora de Santander [ESSA]. (2023). Informe de sostenibilidad 2023. Electrificadora de Santander [ESSA]. <https://www.essa.com.co/site/transparencia-ita/transparencia-essa/informes-de-sostenibilidad>
- [3] Polanco Vidarte, G. A. (2023) *análisis de información para la reducción de pérdidas no técnicas utilizando herramientas tecnológicas en el sector eléctrico*. [Tesis de especialista en alta gerencia]. Universidad Militar Nueva Granada.
- [4] Resolución CREG 015 de 2018 [Comisión de Regulación de Energía y Gas]. Por la cual se establece la metodología para la remuneración de la actividad de distribución de energía eléctrica en el Sistema Interconectado Nacional. 29 de enero de 2018.
- [5] Pérez, A. (2021). Medición avanzada AMI ¿qué significa para Colombia? Bismark. <https://bismark.net.co/medicion-avanzada-ami-que-significa-para-colombia/>
- [6] Resolución 40072 de 2018 [Ministerio de Minas y Energía]. Por la cual se establecen los mecanismos para implementar la Infraestructura de Medición Avanzada en el servicio público de energía eléctrica. 30 de enero de 2018.
- [7] NTC6079, del 19 de mayo de 2021, relativo a Requisitos para sistemas de infraestructura de medición avanzada (AMI) en redes de distribución de energía eléctrica. Icontec. <https://wdn2.ipublishcentral.com//hipertexto500148/viewinsidehtml/501710891072677>

- [8] CREG (2020). Taller Infraestructura de Medición Avanzada Resolución CREG 131 de 2020. [Diapositiva de PowerPoint]. [https://creg.gov.co/public\\_html/info/creg/media/tmp/pdf66046.pdf](https://creg.gov.co/public_html/info/creg/media/tmp/pdf66046.pdf)
- [9] Ley 2099 de 2021. Por medio del cual se dictan las disposiciones para la transición energética, la dinamización del mercado energético, la reactivación económica del país y se dictan otras disposiciones. 10 de julio de 2021.
- [10] Bedoya, S. (2022). *Estrategias Técnico – Regulatorias para la implementación de la infraestructura AMI en el Horizonte 2030 en Colombia*. [Tesis de Magister en Ingeniería Eléctrica]. Universidad Nacional de Colombia.
- [11] Proyecto de Resolución CREG 701 011 de 2022 [Comisión de Regulación de Energía y Gas]. Por la cual se modifican los Títulos II, V, VI y VII de la Resolución CREG 101 001 de 2022. 30 de junio de 2022.
- [12] IEEE Standards Association. IEEE Standard for Smart Grid Interoperability of Energy Technology and Information Technology Operation with the Electric Power System (EPS) and End-Use Applications and Loads. IEEE Std 2030-2011, IEEE, 2011, doi:10.1109/IEEESTD.2011.6018239
- [13] Landis+Gyr. (n.d.). Multi-Commodity AMI. Recuperado el 30 de noviembre de 2024, de <https://www.landisgyr.eu>
- [14] Siemens. (2023). EnergyIP Grid Edge Software. En: Siemens Energy. Recuperado de: <https://www.siemens.com>
- [15] Itron. (2023). OpenWay Riva: Redes inteligentes avanzadas. En: Itron Solutions. Recuperado de: <https://www.itron.com>
- [16] Honeywell (Elster). (2023). EnergyAxis: Plataforma avanzada de medición. En: Honeywell Smart Energy. Recuperado de: <https://www.honeywell.com>

[17] International Energy Agency. (2023). *Smart grids and communication technologies: PLC and its applications*. Recuperado de <https://www.iea.org>.

[18] IEEE Smart Grid. (2023). Wireless communication technologies for smart grids. Recuperado de <https://smartgrid.ieee.org>.

[19] GSMA Intelligence. (2023). *Mobile IoT: Transforming utilities with smart metering*. Recuperado de <https://www.gsma.com>.

[20] Semtech Corporation. (2023). *Introduction to LoRa and its role in IoT for smart cities*. Recuperado de <https://www.semtech.com>.

[21] IEEE Smart Grid. (2023). Advanced Metering Infrastructure and data collection intervals. Recuperado de <https://smartgrid.ieee.org>.

[22] International Energy Agency. (2023). Smart Metering Data: Collection and Utilization. Recuperado de <https://www.iea.org>.

[23] U.S. Department of Energy. (2020). Advanced Metering Infrastructure and Customer Systems: Results from the Smart Grid Investment Grant Program. Recuperado de <https://www.energy.gov>

[24] International Energy Agency (IEA). (2023). Smart Metering Data: Collection, Storage, and Analysis for Modern Grids. Recuperado de <https://www.iea.org>

[25] MIT News. (2023). Powering the energy transition with better storage. Recuperado de <https://news.mit.edu>

[26] Messinis, G. M; Hatziargyriou, N. D (2018). Review of non-technical loss detection methods. Electric Power Systems Research (Vol. 158, pp 250-266).

- [27] Dos Angelos, E.W.S., Saavedra, O.R., Cortés, O.A.C., & De Souza, A.N. (2011). Detection and identification of abnormalities in customer consumptions in power distribution systems. *IEEE Transactions on Power Delivery*, 26(4), 2436-2442.
- [28] Chen, C., Wang, J., & Liu, C.C. (2014). Data quality improvement for load monitoring in AMI systems. *IEEE Transactions on Smart Grid*, 5(3), 1395-1404.
- [29] Zeng, B., Shuai, X., & Su, D. (2019). Outlier detection and imputation of power consumption data in smart grids. *Electric Power Systems Research*, 171, 148-156.
- [30] Yan, Y., Qian, Y., Sharif, H., & Tipper, D. (2013). A survey on smart grid communication infrastructures: Motivations, requirements, and challenges. *IEEE Communications Surveys & Tutorials*, 15(1), 5-20.
- [31] Grilo, A., & Jardim-Gonçalves, R. (2010). Challenging the interoperability of data exchange in the Smart Grid. *Computers in Industry*, 61(9), 863-872.
- [32] Smith, T., & Brown, L. (2018). Cross-validation techniques in power systems. *Journal of Energy Engineering*, 144(2), 04018015.
- [33] Liu, Z., & Yu, Y. (2019). Enhancing smart grid analytics with data enrichment. *Smart Grid and Renewable Energy*, 10(3), 45-56.
- [34] BMC Software. (s.f.). Anomaly Detection: What You Need To Know. Recuperado de <https://www.bmc.com/learn/anomaly-detection.html>
- [35] ResearchGate. (2021). A Review of Anomaly Detection Techniques in Advanced Metering Infrastructure. Recuperado de [https://www.researchgate.net/publication/348938237\\_A\\_Review\\_of\\_anomaly\\_detection\\_techniques\\_in\\_advanced\\_metering\\_infrastructure](https://www.researchgate.net/publication/348938237_A_Review_of_anomaly_detection_techniques_in_advanced_metering_infrastructure)

- [36] Arias, C. Mendez, G. (2022) Metodología basada en minería de datos para la detección de pérdidas no técnicas en sistemas de distribución de energía eléctrica. [Tesis de maestría]. Universidad Politécnica Salesiana de Ecuador.
- [37] Quezada Mateo, F. N. (2017) *Modelo basado en minería de datos para la detección de pérdidas no técnicas de redes de distribución*. [Tesis de Máster en Sistemas de Energía Eléctrica]. Universidad de Sevilla.
- [38] S. S. S. R. Depuru, L. Wang and V. Devabhaktuni, "Support vector machine based data classification for detection of electricity theft," 2011 IEEE/PES Power Systems Conference and Exposition, Phoenix, AZ, USA, 2011, pp. 1-8.
- [39] A. H. Nizar, Z. Y. Dong and Y. Wang, "Power Utility Nontechnical Loss Analysis With Extreme Learning Machine Method," in *IEEE Transactions on Power Systems*, vol. 23, no. 3, pp. 946-955, Aug. 2008.
- [40] C. C. O. Ramos, A. N. de Sousa, J. P. Papa and A. X. Falcão, "A New Approach for Nontechnical Losses Detection Based on Optimum-Path Forest," in *IEEE Transactions on Power Systems*, vol. 26, no. 1, pp. 181-189, Feb. 2011.
- [41] W. Dian-Gang, D. Jin-Chen, H. Lin and G. Yan, "Anomaly Behavior Detection Based on Ensemble Decision Tree in Power Distribution Network," 2018 4th Annual International Conference on Network and Information Systems for Computers (ICNISC), Wuhan, China, 2018, pp. 312-316.
- [42] P. Glauner, J. A. Meira, L. Dolberg, R. State, F. Bettinger and Y. Rangoni, "Neighborhood Features Help Detecting Non-Technical Losses in Big Data Sets," 2016 *IEEE/ACM 3rd International Conference on Big Data Computing Applications and Technologies (BDCAT)*, Shanghai, China, 2016, pp. 253-261.

- [43] Y. Hernández, G. Arroyo-Figueroa, G. Rodríguez, M. Santos and H. Escobedo, "Towards a Framework to Detect and Prevent Non-technical Losses in Power Distribution Based on Data-Mining Techniques and Bayesian Networks," 2015 Fourteenth Mexican International Conference on Artificial Intelligence (MICAI), Cuernavaca, Mexico, 2015, pp. 157-161.
- [44] Iglesias, I. (2022). Analisis y detección de anomalías usando técnicas de Deep learning: caso de estudio en cotizaciones de Pfizer y Moderna. Universidad Pontificia de Comillas
- [45] J. Lee, Y. G. Sun, I. Sim, S. H. Kim, D. I. Kim and J. Y. Kim, "Non-Technical Loss Detection Using Deep Reinforcement Learning for Feature Cost Efficiency and Imbalanced Dataset," in *IEEE Access*, vol. 10, pp. 27084-27095, 2022
- [46] Dos Angelos, E.W.S., Saavedra, O.R., Cortés, O.A.C., & De Souza, A.N. (2011). Detection and identification of abnormalities in customer consumptions in power distribution systems. *IEEE Transactions on Power Delivery*, 26(4), 2436–2442. <https://doi.org/10.1109/TPWRD.2011.2161621>
- [47] Glauner, P., Boechat, A., Dolberg, L., State, R., Bettinger, F., Rangoni, Y., & Duarte, D. (2016). Large-scale detection of non-technical losses in imbalanced data sets. 2016 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT), 1–5. <https://doi.org/10.1109/ISGT.2016.7781159>
- [48] Mashima, D., & Cárdenas, A.A. (2012). Evaluating Electricity Theft Detectors in Smart Grid Networks. En Springer (pp. 210–229). [https://doi.org/10.1007/978-3-642-33338-5\\_11](https://doi.org/10.1007/978-3-642-33338-5_11)
- [49] Lin, C., Chen, S.-J., Kuo, C., & Chen, J. (2014). Non-cooperative game model applied to an advanced metering infrastructure for non-technical loss screening in micro-distribution systems. *IEEE Trans. Smart Grid*, 5(5), 2468–2469. <https://doi.org/10.1109/TSG.2014.2327809>

[50] Weckx, S; Gonzalez, C; Tant, J; De Rybel, T; Driesen, J. "Parameter identification of unknown radial grids for theft detection," 2012 3rd IEEE PES Innovative Smart Grid Technologies Europe (ISGT Europe), Berlin, Germany, 2012, pp. 1-6

[51] Electro Oriente S.A. (2020). Implementación de infraestructura avanzada para reducción de pérdidas. Recuperado de: <https://www.electrooriente.com.pe>

[52] Comité de Energía Eléctrica de CIER. (2023). Reducción de pérdidas eléctricas en América Latina: Experiencias y estrategias. Recuperado de <https://www.cecacier.org>

[53] Comisión Federal de Electricidad (CFE). (2014). Programa de reducción de pérdidas técnicas y no técnicas en redes de distribución. Recuperado de ResearchGate: [https://www.researchgate.net/publication/312881997\\_Programa\\_de\\_reduccion\\_de\\_p%C3%A9rdidas\\_CFE](https://www.researchgate.net/publication/312881997_Programa_de_reduccion_de_p%C3%A9rdidas_CFE)

[54] CFE. (2023). Avances en la reducción de pérdidas no técnicas mediante medición inteligente. Recuperado de <https://www.cfe.mx>

[55] Codensa S.A. (2023). Reducción de pérdidas no técnicas mediante medición avanzada. En: Enel Colombia. Recuperado de: <https://www.enel.com.co>

[56] Enel Codensa. (2023). Estrategias avanzadas de reducción de pérdidas en el sistema eléctrico colombiano. Recuperado de <https://www.enel.com.co>

[57] Tata Power. (2019). Estrategias avanzadas para la gestión de pérdidas no técnicas. En: Tata Power Case Studies. Recuperado de: <https://www.tatapower.com>

[58] Tata Power. (2023). *Innovaciones en detección de pérdidas mediante análisis de datos en redes inteligentes*. Recuperado de <https://www.tatapower.com>

- [59] Rabinovich, L. (n.d.). *Evaluación de tiempo y memoria en modelos de aprendizaje automático*. Tesis, Universidad de Buenos Aires. Recuperado de [https://bibliotecadigital.exactas.uba.ar/download/tesis/tesis\\_n7470\\_Rabinovich.pdf](https://bibliotecadigital.exactas.uba.ar/download/tesis/tesis_n7470_Rabinovich.pdf)
- [60]datasource.ai. (n.d.). *Métricas de evaluación de modelos en el aprendizaje automático*. Recuperado de <https://www.datasource.ai/es/data-science-articles/metricas-de-evaluacion-de-modelos-en-el-aprendizaje-automatico>
- [61] Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20(3), 273–297. <https://doi.org/10.1007/BF00994018>
- [62] LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436–444. <https://doi.org/10.1038/nature14539>
- [63] Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324>
- [64] Bishop, C. M. (2006). *Pattern recognition and machine learning*. Springer.
- [65] Altman, N. S. (1992). An introduction to kernel and nearest-neighbor nonparametric regression. *The American Statistician*, 46(3), 175–185. <https://doi.org/10.1080/00031305.1992.10475879>
- [66] Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.
- [67] Papa, J. P., Falcão, A. X., & Suzuki, C. T. N. (2009). Optimum-Path Forest: A novel and effective algorithm for supervised learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(2), 215–227. <https://doi.org/10.1109/TPAMI.2008.85>

- [68] Rudin, C. (2019). Stop explaining black box machine learning models for high-stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, 1(5), 206–215. <https://doi.org/10.1038/s42256-019-0048-x>
- [69] Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (2nd ed.). Springer.
- [70] Kohonen, T. (2001). *Self-Organizing Maps* (3rd ed.). Springer.
- [71] Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly detection: A survey. *ACM Computing Surveys*, 41(3), 1–58.
- [72] Montgomery, D. C. (2012). *Introduction to Statistical Quality Control* (7th ed.). Wiley.
- [73] Jackson, P. (1998). *Introduction to Expert Systems* (3rd ed.). Addison-Wesley.
- [74] Myerson, R. B. (1991). *Game Theory: Analysis of Conflict*. Harvard University Press.
- [75] Jain, A. K., Murty, M. N., & Flynn, P. J. (1999). Data clustering: A review. *ACM Computing Surveys*, 31(3), 264–323.
- [76] Benchmarking Methodologies International Group. (2010). *Benchmarking Handbook: A Practical Guide to Best Practice*. Cambridge University Press.
- [77] Najeh, L., & Jarray, F. (2018). Enhanced Energy Theft Detection Using Support Vector Machines Combined With Data Clustering Techniques. *International Journal of Electrical Power & Energy Systems*, 101, 413-420. <https://doi.org/10.1016/j.ijepes.2018.03.034>
- [78] Kulkarni, P., & Thansekhar, M. R. (2016). Anomaly Detection in Smart Grid Using Machine Learning Techniques. *Procedia Computer Science*, 92, 190-195. <https://doi.org/10.1016/j.procs.2016.07.347>

- [79] Jiang, J., He, H., & Yan, J. (2017). A Novel Data-Driven Framework for Power Theft Detection With Smart Meter Data. *IEEE Transactions on Industrial Informatics*, 13(2), 559-570. <https://doi.org/10.1109/TII.2016.2607179>
- [80] Passos Júnior, L. A., et al. (2016). Unsupervised Non-Technical Losses Identification Through Optimum-Path Forest. *Electrical Power Systems Research*, 140, 413-423. <https://doi.org/10.1016/j.epsr.2016.05.036>
- [81] Mashima, D., & Cárdenas, A. A. (2012). Evaluating Electricity Theft Detectors in Smart Grid Networks. En *International Conference on Security and Privacy in Communication Systems* (pp. 210-229). Springer. [https://doi.org/10.1007/978-3-642-33338-5\\_11](https://doi.org/10.1007/978-3-642-33338-5_11)
- [82] T. Siriteerakul and V. Boonjing, "Support Vector Machine accuracy improvement with k-means clustering," 2013 International Computer Science and Engineering Conference (ICSEC), Nakhonpathom, Thailand, 2013, pp. 218-221.
- [83] U. Maulik and S. Bandyopadhyay, "Performance evaluation of some clustering algorithms and validity indices," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 12, pp. 1650-1654, Dec. 2002, doi: 10.1109/TPAMI.2002.1114856.
- [84] A. E. -S. Ezugwu, M. B. Agbaje, N. Aljojo, R. Els, H. Chiroma and M. A. Elaziz, "A Comparative Performance Study of Hybrid Firefly Algorithms for Automatic Data Clustering," in *IEEE Access*, vol. 8, pp. 121089-121118, 2020, doi: 10.1109/ACCESS.2020.3006173.