

Revisión de métodos de caracterización de usuarios del servicio de energía a partir de datos de infraestructura de medición avanzada (AMI)

Jerimberg González Monroy

Trabajo de Grado para Optar el título de:
Especialista en Sistemas de Distribución de Energía Eléctrica

Directora

Yulieth Jiménez Manjarrés

Doctora en Ingeniería

Universidad Industrial de Santander

Facultad de Ingenierías Físico - Mecánicas

Escuela de Ingeniería Eléctrica, Electrónica y de Telecomunicaciones

Bucaramanga

2025

Dedicatoria

Dedico el logro de la culminación de mi posgrado a mis padres, mi hermano y mi prometida junto con mi hijo que viene en camino quienes, con dedicación y amor incondicional, me brindaron todas las herramientas que necesité para lograr esta meta, y este proyecto escrito es su representación.

-Jerimberg González Monrroy

Agradecimientos

Gracias a Dios por todas las bendiciones que ha derramado sobre mi vida.

A mis padres, Jerimberg y Martha que son mi mayor fortaleza, los grandes amores de mi vida, mi motor, mis ganas de salir adelante, gracias por su esfuerzo, por su amor sin medida, por su guía, por su compañía y enseñanzas, a quienes les debo todo.

A mi hermano Greidy, por ser mi amigo, mi ejemplo, mi apoyo incondicional y procurar siempre mi bienestar.

A María Fernanda por ser mi alegría, mi apoyo, mi compañera en la construcción de este amor bonito que nació en el campus universitario, gracias por hacer más felices mis días en este camino llamado vida.

A mi hijo que viene en camino, gracias por ser mi motor y embellecer mis días futuros, expectante a tenerte pronto en mis brazos.

A la ingeniera Yulieth Jiménez, gracias por su apoyo profesional, por compartirnos su conocimiento, enseñanzas y su guía en la realización de este proyecto.

A la Universidad Industrial de Santander por la educación de calidad que me brindó, los amigos que me regaló y todos los recuerdos que siempre permanecerán con sonrisas, cariño y orgullo.

-Jerimberg González Monrroy

Tabla de contenido

Introducción	13
1. Planteamiento del problema.....	14
2. Justificación	15
3. Objetivos	17
3.1. Objetivo general.....	17
3.2. Objetivos específicos	18
4. Marco de referencia	19
4.1. Marco de antecedentes	19
4.2. Casos de implementación de AMI y métodos de caracterización de usuarios aplicados en el mundo.....	20
4.3. Marco teórico	22
4.3.1. Análisis de datos	22
4.3.2. Caracterización de usuarios de energía eléctrica	22
4.3.3. Eficiencia energética.....	23
4.3.4. Infraestructura de medición avanzada (AMI)	23
4.3.5. Medidor inteligente.....	24
4.3.6. Operador de Red	24
4.3.7. Programas de respuesta a la demanda.....	24
4.3.8. Recursos energéticos distribuidos.....	25
4.3.9. Redes inteligentes.	25
4.3.10. Transición energética.	25
5. Marco metodológico	26
5.1. Establecimiento de la ecuación de búsqueda.....	26
5.2. Búsqueda preliminar y refinación de la ecuación	26
5.3. Revisión de artículos seleccionados.....	27
5.4. Depuración de la información.....	27

5.5.	Resumen de la información de valor recolectada	27
5.6.	Clasificación de los métodos seleccionados	27
5.7.	Análisis de resultados	28
6.	Resultados	28
6.1.	Establecimiento de la ecuación de búsqueda	28
6.2.	Búsqueda preliminar y refinación de la ecuación	31
6.3.	Revisión de los artículos seleccionados	34
6.4.	Depuración de la información	36
6.4.1.	Análisis bibliométrico por factores	37
6.5.	Resumen de la información de valor recolectada	40
6.6.	Clasificación de los métodos seleccionados	57
6.7.	Análisis de resultados	60
6.8.	Retos y oportunidades del análisis de datos en AMI	62
6.9.	Aplicaciones y tendencias basadas en datos AMI	65
6.9.1.	Eficiencia energética	65
6.9.2.	Pérdidas de energía	65
6.9.3.	Pronóstico, previsión y respuesta a la demanda	66
6.9.4.	Segmentación de usuarios	67
6.9.5.	Internet de energía e integración con energías renovables	67
6.9.6.	Inteligencia artificial	68
7.	Conclusiones	69
8.	Referencias Bibliográficas	71

Lista de Tablas

Tabla 1. Cumplimiento de objetivos	18
Tabla 2. Palabras clave y relacionadas.....	29
Tabla 3. Ecuaciones de búsqueda utilizadas	32
Tabla 4. Métodos de caracterización de usuarios encontrados en la literatura	43
Tabla 5. Clasificación de métodos de caracterización de usuarios	59

Lista de Figuras

Figura 1. Metodología de la investigación.....	26
Figura 2. Coincidencias de documentos entre bases de datos	35
Figura 3. Artículos escogidos después de la depuración	37
Figura 4. Producción científica anual	38
Figura 5. Clasificación de documentos por autores	38
Figura 6. Clasificación de artículos por relación de palabras clave de los autores.....	39
Figura 7. Clasificación de documentos por países.....	40
Figura 8. Procedimiento básico de clasificación de datos AMI.	42

Glosario

- AERO:** Extracción de Actividad con Observación Racional (*Activity Extraction with Rational Observation*)
- ALR:** Tasa de carga promedio durante el pico (*Average load rate during peak1*)
- AMI:** Infraestructura de Medición Avanzada (*Advanced Metering Infrastructure*)
- ARIMA:** Modelo Autorregresivo Integrado de Media Móvil (*Autoregressive Integrated Moving Average*)
- CDI:** Índice de demanda crítico (*Critical Demand Index*)
- CFSFDP:** Agrupación mediante búsqueda rápida y búsqueda de picos de densidad (*Clustering by Fast Search and Find of Density Peaks*)
- CREG:** Comisión de Regulación de Energía y Gas.
- CV:** Coeficiente de variación (*Coefficient of Variation*)
- CVI's:** Varios Índices de validez de agrupamiento (*Various clustering validity indices*)
- DBI:** Índice de Davies-Bouldin (*Davies-Bouldin index*)
- DBSCAN:** Agrupamiento Basado en Densidad para Aplicaciones Espaciales con Ruido (*Density-Based Spatial Clustering of Applications with Noise*)
- DFT:** Transformada discreta de Fourier (*Transformada Discreta de Fourier*)
- DI:** Índice de Dunn (*Dunn index*)
- DPP:** Distancia Ponderada de Pearson (*Weighted Pearson Distance*)
- DTW:** Alineación Temporal Dinámica (*Dynamic Time Warping*)
- DWT:** Transformada discreta de Wavelet (*Discrete Wavelet Transformation*)
- ESSA:** Electrificadora de Santander S.A E.S.P
- EVI:** Índice de varianza de energía (*Energy Variance Index*)
- FCA:** Adecuación de agrupamiento difuso (*Fuzzy Clustering Adequacy*)
- FCM:** Algoritmo difuso c-means (*Fuzzy c-means*)
- GMM:** Modelo mixto Gaussiano (*Gaussian mixed model*)
- GRU:** Unidad Recurrente con Puertas (*Gated Recurrent Unit*)
- GSOM:** Mapas auto-organizados en crecimiento (*Growing Self-Organizing Map*)
- HHM:** Modelos de Markov ocultos (*Hidden Markov Models*)
- IoC:** Índice de Cambio (*Index of Change*)

- ISI:** Instituto para la Información Científica (*Institute for Scientific Information*)
- LAR:** Regresión de ángulo mínimo (*Least Angle Regression*)
- LSTM:** Memoria a largo plazo (*Long Short Term Memory*)
- MAPE:** Error Porcentual Absoluto Medio (*Mean Absolute Percent- Age Error*)
- MCIAT:** Interpretación multicriterio de la tabla de pruebas aplicadas (*Multi-Criteria Interpretation of Applied Test Table*)
- MIA:** Índice de Adecuación Promedio (*Mean Index Adequacy*)
- MME:** Ministerio de Minas y Energías
- MSE:** Error cuadrático Medio (*Mean Square Error*)
- OLAP:** Procesamiento Analítico en Línea (*Online Analytical Processing*)
- OR:** Operador de Red
- OT:** Hilos de Operación (*Operation threads*)
- PAA:** Aproximación Agregada por Partes (*Piecewise Aggregate Approximation*)
- PCA:** Análisis de Componentes Principales (*Principle Component Analysis*)
- PCC:** Coeficiente de correlación de Pearson (*Pearson Correlation Coefficient*)
- PCP:** Explorador de patrones de consumo de energía (*Power Consumption Pattern Explorer*)
- RAPV:** Proporción de la carga total durante el pico respecto al valle (*Ratio of total load during peak to valley*)
- RI:** Índice de Rand (*Rand index*)
- RI:** Redes Inteligentes
- RP:** Proporción de la carga total durante el pico (*Ratio of total load during peak*)
- RST:** Teoría de los Conjuntos Aproximados (*Rough Sets Theory*)
- RTPV:** Proporción de la carga promedio durante el pico respecto al valle (*Ratio of average load during pea to valley*)
- SAICC:** Evaluación estadística para identificar cambios en el consumo (*Statistical Assessment for Identifying Changes in Consumption*)
- SAX:** Aproximación Simbólica Agregada (*Symbolic Aggregate Approximation*)
- SBD:** Distancia basada en forma (*Shape-based distance*)
- SC:** Coeficiente de silueta (*silhouette coefficient*)
- SGEM:** Redes inteligentes y Mercados de Energía (*Smart Grids and Energy Markets*)
- SOM:** Mapas auto-organizados (*Self-organizing Map*)
- SSE:** Suma de Errores al Cuadrado (*Sum of Square Error*)

STLF: Pronósticos de carga a corto plazo (*Short-Term Load Forecasting*)

SVD: Descomposición en Valores Singulares (*Singular Value Decomposition*)

SVM: Máquina de Vectores de Soporte (*Support Vector Machine*)

TPL: Perfil de carga típico (*Typical Load Profile*)

UIS: Universidad Industrial de Santander

UPME: Unidad de Planeación Minero-Energética

WOS: *Web of Science*

XM: Administradores del Mercado Eléctrico

Resumen

Título: Revisión de métodos de caracterización de usuarios del servicio de energía a partir de datos de infraestructura de medición avanzada (AMI)

Autor: Jerimberg González Monrroy

Palabras Clave: Infraestructura de medición avanzada (AMI), Operadores de red, Caracterización de usuarios, Big data, Métodos de caracterización, Técnicas avanzadas de análisis, Medidores avanzados, Medición inteligente, Eficiencia energética.

Descripción: Para los diferentes operadores de red es de gran valor comprender el comportamiento de los usuarios de energía eléctrica y de esta manera poder establecer patrones de consumo, predecir demandas futuras a corto plazo e identificar oportunidades de ahorro de energía. En el ámbito regional, la infraestructura eléctrica se encuentra en transición, ya que se están diseñando e implementando pilotos de medición avanzada por parte de los Operadores de Red (OR), que permitan en un futuro la captura de datos en gran volumen. Esta Infraestructura de medición avanzada (AMI) permitirá el acceso a datos detallados, que serán generados con mayor frecuencia, y con los cuales se podrá analizar dicho comportamiento de consumo obteniendo por ejemplo el consumo de energía por hora, por dispositivo, por zona horaria y la incorporación de otros tipos de datos nuevos como los parámetros de calidad del servicio, eficiencia energética, vehículos eléctricos, almacenamiento y generación distribuida, entre otros. Por ello, la revisión de métodos de análisis y caracterización de usuarios del servicio de energía eléctrica a partir de datos de AMI es esencial. Esta monografía proporcionará a las empresas proveedoras de energía herramientas y conocimientos para implementar estrategias en pro de mejorar la gestión operativa, la satisfacción del cliente y el cumplimiento de objetivos de sostenibilidad energética mediante la identificación de dichos métodos.

Abstract

Title: Review of energy service user's characterization methods based on advanced metering infrastructure (AMI) data.

Author(s): Jerimberg Gonzalez Monrroy

Key Words: Advanced metering infrastructure (AMI), Distribution System Operators (DSO), User characterization, Big data, Characterization methods, Advanced analysis techniques, Advanced meters, Smart metering, Energy efficiency.

Description: The comprehension of the behavior of electricity users is key for the distribution system operators (DSO), in order to establish consumption patterns, predict future short-term demands and identify energy saving opportunities. In the regional context, these companies are designing and implementing smart metering pilot projects which will enable large-scale data capture in the future, under the undergoing energy transition. This Advanced Metering Infrastructure (AMI) will enable access to more frequent and detailed data, facilitating the analysis of consumption patterns. For instance, it allows for the measurement of energy consumption by hour, device and time zone, as well as the incorporation of novel data types such as service quality parameters, energy efficiency metrics, electric vehicle integration, energy storage, and distributed generation, among others. Consequently, the examination of analytical methods and user characterization techniques based on AMI data is of paramount importance. This study aims to equip energy service providers with advanced tools and insights to implement strategies that enhance operational efficiency, customer satisfaction, and the achievement of energy sustainability goals, leveraging the potential of these methodologies.

Introducción

La transición hacia sistemas energéticos más eficientes y sostenibles requiere una comprensión profunda del comportamiento de los usuarios de energía eléctrica (UPME, 2023). La implementación de Infraestructuras de Medición Avanzada (AMI) se presenta como una herramienta clave para modernizar la gestión de los recursos energéticos mediante el uso de datos detallados y frecuentes (ESSA, 2024). Este enfoque permite optimizar la operación de las redes eléctricas y fomentar un uso más racional de la energía (XM, 2024).

El uso de AMI facilita la recolección y análisis de datos que antes eran inaccesibles, como patrones de consumo por hora y dispositivos, así como indicadores de calidad del servicio y eficiencia energética (MME, 2019). Estos datos permiten caracterizar a los usuarios de manera precisa, promoviendo la creación de programas personalizados de eficiencia energética, respuesta a la demanda y la integración de nuevas tecnologías (CREG, 2022). Sin embargo, el aprovechamiento de estas ventajas implica superar retos técnicos, como el manejo y análisis de grandes volúmenes de datos.

Esta monografía aborda estos desafíos mediante una revisión de métodos y herramientas disponibles para la caracterización de usuarios a partir de datos generados por AMI. El objetivo es proporcionar conocimientos y herramientas que contribuyan a mejorar la gestión operativa de las empresas proveedoras de energía, aumentar la satisfacción de los usuarios y promover el cumplimiento de objetivos de sostenibilidad energética.

1. Planteamiento del problema

La infraestructura de medición avanzada (AMI) se conforma por hardware, software y redes que permiten una comunicación bidireccional entre los operadores de red y los usuarios. Se espera que mediante AMI se tenga acceso a un gran volumen de datos detallados, que serán generados con mayor frecuencia, y con los cuales se podrá analizar el comportamiento de consumo de energía eléctrica de los diferentes usuarios. De hecho, AMI permitirá adquirir una amplia variedad de datos a los cuales no se tenían acceso anteriormente para analizar información interesante como el consumo de energía por hora, por dispositivo, por zona horaria y la incorporación de otros tipos de datos nuevos como los parámetros de calidad del servicio, eficiencia energética, vehículos eléctricos, almacenamiento y generación distribuida, entre otros.

Se entiende por caracterizar usuarios del servicio de energía cómo identificar, organizar y analizar datos demográficos, económicos, sociales, culturales y de consumo energético para comprender de manera más profunda sus patrones de comportamiento (Hurtado et al., 2009). Esto no solo permite clasificar o categorizar a los usuarios, sino también sistematizar de forma crítica la información para identificar oportunidades estratégicas para la creación de programas de eficiencia energética, esquema de demanda flexible y toma de decisiones acertadas en la optimización y operación de la red. Esta tarea de caracterización es fundamental para avanzar hacia un sistema energético más eficiente, sostenible y adaptado a las necesidades individuales de los usuarios y de la red.

Aunque en la literatura se encuentra una variedad de métodos convencionales que han sido utilizados durante muchos años, esta tarea sigue representando un desafío significativo para las empresas prestadoras del servicio de energía con la implementación de AMI, por aspectos como el gran volumen de datos o “Big Data”. En este orden de ideas, en esta

monografía se propone revisar métodos de caracterización de usuarios y técnicas avanzadas de análisis de datos que aprovechen al máximo los datos provenientes de AMI.

2. Justificación

La Unidad de Planeación Minero-Energética (UPME) ha estimado en la proyección de demanda de energía eléctrica 2023-2037 un crecimiento del 3,31% (UPME, 2023). De igual forma, los administradores del mercado eléctrico (XM), en un comunicado de prensa informaron que, en 2023, la demanda de energía en Colombia fue de 79.985 GMh, habiendo aumentado un 4,45% en comparación con el año anterior (XM, 2024). En el contexto regional, la Electrificadora de Santander (ESSA) obtuvo una demanda anual de energía como OR de 3.271 GWh. Actualmente del 100% de energía eléctrica que comercializa ESSA, un 4,6% corresponde a pérdidas de energía por conexiones ilegales o provenientes de fraudes, de las cuales recuperó 32 GWh durante el primer periodo del 2023 implicando una inversión de más de 47 mil millones de pesos anuales para implementar acciones que le permitan gestionar las pérdidas de energía (ESSA, 2023).

Una de las iniciativas en las que ESSA está incursionando con este fin es precisamente el proceso de implementación de AMI, a través de la integración de redes de comunicación, software y hardware como los medidores avanzados, que son equipos automatizados que envían remotamente información sobre diferentes parámetros y medidas de la red eléctrica a la compañía y reciben órdenes remotas, permitiendo realizar toma de lectura, suspensión y reconexión en tiempo real, entre otras funcionalidades. ESSA actualmente cuenta con la reglamentación para realizar esta actividad. Dicho documento presenta los lineamientos a seguir para la integración y puesta en operación de sistemas de medición avanzados (*Smart*

Metering) o sistemas de infraestructura de medición avanzada (AMI) (ESSA, 2017).

El Ministerio de Minas y Energía expidió la Resolución número 40483 de 2019, en la cual se establecen los mecanismos para implementar la Infraestructura de Medición Avanzada-AMI en el servicio público de energía eléctrica, con un tiempo de transición hasta el año 2030 (Meta: 75% de los usuarios urbanos y el 50% de los usuarios rurales del país cuenten con la nueva tecnología (AMI) (MME, 2019). Además, la CREG definió la Resolución 101 001 de 2022 con las condiciones para la implementación de la infraestructura en medición avanzada en la prestación del servicio público domiciliario, sin definir aún el esquema de remuneración de AMI¹ (CREG, 2022). Esta Resolución debe ser modificada dada la modificación que sufrió la Ley 2099 de Transición Energética de 2021 en relación con los costos del medidor avanzado (Ley 2099, 2021, art. 56).

ESSA realizará una inversión de 5.316 millones para la instalación y puesta en marcha del Piloto Infraestructura de Medición Avanzada (AMI) entre el 2024-2027. Se formulará el proyecto con un escenario preliminar para la implementación de AMI en zonas de altas pérdidas y se irá ajustando a medida que el Ministerio de Minas y Energía vaya publicando nuevas condiciones (ESSA, 2024).

De la implementación de AMI se esperan beneficios tanto para las empresas como para los usuarios. La medición avanzada tiene como beneficios mejorar la calidad del suministro de energía eléctrica, la reducción de las pérdidas técnicas y no técnicas, así como proporcionar nuevas alternativas para que los usuarios dispongan de información confiable en tiempo real, lo que permite hacer un uso eficiente y racional de la energía eléctrica. Además, presenta como

¹ Esta Resolución debe ser modificada debido a la declaración de inexecutable de un inciso del Artículo 56 de la Ley 2099 de Transición Energética de 2021

beneficio lograr un mejor conocimiento de los usuarios al disponer de información más detallada y frecuente que con medidores convencionales. Para materializar estos beneficios de AMI se requiere un análisis de datos que permita aprovechar dicha información de forma eficiente y precisa, adaptada al contexto regional. Por su parte, una adecuada caracterización de los usuarios trae consigo muchos beneficios, dentro de los cuales se encuentran: comprender mejor los patrones y perfiles de consumo para brindar ofertas comerciales y servicios personalizados, identificar tendencias y anticipar demandas futuras que permitan la optimización de recursos energéticos, el desarrollo de programas de eficiencia energética, la realización de una gestión eficaz de la demanda de energía y la identificación de comportamientos anómalos, y facilitar la integración de las fuentes no convencionales de energía en pro de la transición energética.

En conclusión, la revisión de métodos de caracterización de usuarios del servicio de energía a partir de datos de AMI es esencial para comprender el comportamiento de los usuarios, establecer patrones de consumo, predecir demandas futuras a corto plazo e identificar oportunidades de ahorro de energía. Esta monografía proporcionará a las empresas proveedoras de energía herramientas y conocimientos para implementar estrategias en pro de mejorar la gestión operativa, la satisfacción del cliente y el cumplimiento de objetivos de sostenibilidad.

3. Objetivos

Para el desarrollo de esta monografía se planteó el siguiente objetivo general, que se cumplió a través de los objetivos específicos a continuación.

3.1. Objetivo general

Identificar métodos de caracterización de usuarios de electricidad a partir de datos provenientes de Infraestructura de Medición Avanzada (AMI).

3.2. Objetivos específicos

- Revisar información disponible en la literatura científica y del mercado sobre los diferentes métodos para caracterizar usuarios a partir de datos de medidores avanzados.
- Identificar métodos de caracterización de usuarios a partir de datos de medidores avanzados que presenten mayor potencial para el contexto regional.
- Describir los retos y oportunidades que implica el análisis de los datos de AMI para la caracterización de usuarios y otras aplicaciones.

A continuación, la Tabla 1 señala en qué parte del documento se evidencia el cumplimiento de estos objetivos.

Tabla 1. Cumplimiento de objetivos

Objetivo	Cumplimiento
<ul style="list-style-type: none"> • Revisar información disponible en la literatura científica y del mercado sobre los diferentes métodos para caracterizar usuarios a partir de datos de medidores avanzados. 	Sección 6.5
<ul style="list-style-type: none"> • Identificar métodos de caracterización de usuarios a partir de datos de medidores avanzados que presenten mayor potencial para el contexto regional. 	Sección 6.6
<ul style="list-style-type: none"> • Describir los retos y oportunidades que implica el análisis de los datos de AMI para la caracterización de usuarios y otras aplicaciones. 	Sección 6.8

4. Marco de referencia

En este capítulo se identifican las leyes, normas y regulaciones que promueven la integración de AMI en Colombia. Además, se presenta un contexto internacional y nacional de casos de estudios de caracterización de usuarios mediante AMI, con el fin de mostrar la relevancia de la investigación y sentar las bases de esta.

4.1. Marco de antecedentes

En Colombia, se evidencia una creciente demanda de energía permitiendo que se desarrollen políticas, leyes y regulaciones en pro de buscar alternativas y soluciones que permitan optimizar el uso de los recursos disponibles en la red. La UPME solicitó un estudio “Smart Grids Colombia Visión 30” (Grupo técnico proyecto BID, 2016), para definir la ruta a seguir para la implementación de Redes Inteligentes (RI) en Colombia, identificando AMI como una de las tecnologías habilitadoras para las demás tecnologías de RI.

Además de la UPME, se han integrado otros actores importantes del sector como el Ministerio de Minas y Energías (MME) que mediante la Resolución MME 40072 establece como objetivo para el 2030 un despliegue masivo de medidores avanzados con cobertura del 95% de usuarios (MME, 2018). Pero esta fue modificada respectivamente en el 2019 y 2020 por la Resolución MME 40483, la cual estableció una cobertura del 75% de usuarios con AMI para el 2030 (MME, 2019) y la Resolución MME 40142 mediante la cual se prorrogó el plazo del 15 de abril al 30 de noviembre del 2020 para que la CREG estableciera las condiciones para la implementación de la Infraestructura de Medición Avanzada en la prestación del servicio público domiciliario de energía eléctrica en el Sistema Interconectado Nacional (SIN) (MME, 2020).

La CREG mediante la propuesta de Resolución 131 de 2020 hizo público el proyecto en el cual se establecieron las condiciones de implementación de AMI en el SIN en una primera versión para comentarios (CREG, 2020). Posteriormente, a través de la Resolución 101 001 de 2022 se establece de una forma más clara y específica las condiciones para la implementación de la infraestructura de medición avanzada en el SIN (CREG, 2022), teniendo en cuenta la modificación realizada en la ley 2099 con respecto a quien asume los costos de AMI (Ley 2099, 2021, art. 56). Además, esa resolución definió las responsabilidades, deberes y derechos de operadores de red, comercializadores, usuarios y el gestor independiente de datos e información (GIDI), y también estableció que el OR deberá implementar un plan de despliegue piloto de AMI teniendo en cuenta las fases de despliegue descritas en la resolución. Finalmente, se espera que la CREG defina pronto la forma en que se remunerará la implementación de AMI.

4.2. Casos de implementación de AMI y métodos de caracterización de usuarios aplicados en el mundo.

Con base en el panorama regulatorio colombiano en materia de AMI, cada OR deberá presentar un plan de despliegue de AMI ante la CREG y MME con el fin de identificar la tecnología que mejor se ajuste a su solución particular. En Colombia existen proyectos piloto de medidores avanzados implementados por empresas como EMCALI, EPSA, ELECTRICARIBE (AFINIA Y AIR-e) y CODENSA (Grupo Técnico Proyecto BID, 2016). En la ESSA también se realizó una prueba de concepto de medición avanzada con el sistema SITELRED: sistema diseñado para apoyar la gestión de pérdidas de energía eléctrica del sistema de distribución de energía eléctrica en la Electrificadora de Santander (Ordoñez et al., 2017) y se está planeando un Piloto AMI (ESSA, 2024).

En materia nacional se han implementado métodos de caracterización de usuarios como regresión lineal o polinomial ajustado como la presentada en la tesis “Impacto en la gestión de la demanda de usuarios residenciales que incorporan sistemas de medición con medidores avanzados” (Flórez Duarte, 2021). Además, en Chile se implementó un “Análisis de clúster para identificar similitudes en el comportamiento de los perfiles de carga en clientes residenciales” (Marrero et al., 2021). Por último, a nivel internacional más específicamente en España se evidencia el estudio de “La caracterización de la demanda de energía mediante patrones estocásticos en REI (Serrano Guerrero, 2020).

Por otro lado, se han implementado metodologías de caracterización de usuarios de datos provenientes de AMI para otras aplicaciones como los programas de respuesta a la demanda de los cuales se encuentran: “Detección de cambios en los patrones de consumo de energía eléctrica mediante información adaptativa usando algoritmos teóricos” (Kojury-Naftchali et al., 2021a) y “Mejorar la implementación de la respuesta a la demanda en red de distribución por minería de datos en AMI” mediante la utilización de series temporales sobre las que se aplican procedimientos de reconocimiento de patrones (Kojury-Naftchali & Fereidunian, 2019).

Otras aplicaciones, son la de predicción de carga a corto plazo utilizando datos de AMI mediante un método llamado pronóstico utilizando factorización matricial (Mansoor et al., 2023) y la identificación de patrones de consumo de electricidad mediante agrupamiento (Qiu et al., 2016a). Por lo anterior, es importante la revisión aplicada al análisis de datos en medidores avanzados identificando las técnicas y metodología adoptadas o desarrolladas para abordar cada aplicación en particular como análisis de carga, previsión de carga y gestión de carga (Wang et al., 2019). Se hará una descripción más detallada de los métodos encontrados

en el desarrollo de la monografía en el numeral 6.6, seleccionando los que mejor se adapten al contexto regional.

4.3. Marco teórico

El marco teórico, que se desarrolla a continuación, permite conocer los conceptos básicos necesarios para el entendimiento del desarrollo de esta monografía.

4.3.1. Análisis de datos

El análisis de datos es muy importante para los sistemas energéticos dado que permite realizar diversos estudios para pronosticar la demanda, prevenir fallos, optimizar el consumo energético y analizar el comportamiento de productores y consumidores. La toma de decisiones basada en el análisis de datos maximiza el costo-beneficio y garantiza una planificación y operación energética exitosa. Con el desarrollo de medidores avanzados, sistemas de control en los sistemas eléctricos, así como la recopilación y evaluación de datos de generación y consumo, se busca la flexibilidad operativa y la rentabilidad económica mediante programas de gestión de la demanda y respuesta a la demanda para garantizar la planificación y el funcionamiento fuera de la red (Çetinkaya et al., 2020) .

4.3.2. Caracterización de usuarios de energía eléctrica

La caracterización es un tipo de descripción cualitativa que puede basarse en datos o en aspectos cuantitativos, con el fin de profundizar el conocimiento sobre algo. Para cualificar ese algo previamente, se deben identificar y organizar los datos; y a partir de ellos, describir (caracterizar) de una forma estructurada; y posteriormente, establecer su significado (sistematizar de forma crítica) (Hurtado et al., 2009).

La caracterización de los usuarios tiene como intención principal la identificación de los elementos que determinan los rasgos demográficos, económicos, sociales, culturales y de consumo de los diferentes estratos socioeconómicos de los usuarios. Entre los aspectos demográficos se encuentran: la distribución por edad, sexo, estado civil, estructura familiar; entre las características económicas, se consideran los niveles de ingreso, los patrones de consumo y la actividad laboral (Carlos & Devia Tabares, 2017).

4.3.3. Eficiencia energética

Es la relación entre la energía aprovechada y la total utilizada en cualquier proceso de la cadena energética, que busca ser maximizada a través de buenas prácticas de reconversión tecnológica o sustitución de combustibles. A través de la eficiencia energética, se busca obtener el mayor provecho de la energía, bien sea a partir del uso de una forma primaria de energía o durante cualquier actividad de producción, transformación, transporte, distribución y consumo de las diferentes formas de energía, dentro del marco del desarrollo sostenible y respetando la normatividad vigente sobre el ambiente y los recursos naturales renovables (CREG, 2020).

4.3.4. Infraestructura de medición avanzada (AMI)

Según la definición de la Resolución número 40072 de 2018 expedida por el Ministerio de Minas y Energía, es la infraestructura que permite la comunicación bidireccional con los usuarios del servicio de energía eléctrica incluyendo hardware (medidores avanzados, centros de gestión de medida, enrutadores, concentradores, antenas, entre otros), software y redes de comunicaciones que, en conjunto, 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 (CREG, 2020)

4.3.5. Medidor inteligente.

También llamado medidor avanzado de energía eléctrica, medidor avanzado: Según la definición de la Resolución número 40072 de 2018 expedida por el Ministerio de Minas y Energía, es el dispositivo que mide y registra datos de uso de energía eléctrica de los usuarios, en intervalos máximos de una hora, con capacidad de almacenar y transmitir dichos datos, por lo menos, con una frecuencia diaria (CREG, 2020).

4.3.6. Operador de Red

Persona jurídica encargada de la planeación de la expansión, las inversiones, la operación y el mantenimiento de todo o parte de un Sistema de Transmisión Regional o Sistema Distribución Local, incluidas sus conexiones al Sistema de Transmisión Nacional. El OR siempre debe ser una empresa de servicios públicos domiciliarios (CREG, 2020).

4.3.7. Programas de respuesta a la demanda

Los Programas de Respuesta de la Demanda (RD) son el pago de incentivos destinados a alterar el consumo de energía del usuario final en respuesta a las fluctuaciones del precio de la energía ya sea para inducir un menor consumo de energía cuando los precios del mercado son altos o para mejorar la confiabilidad del sistema si está comprometida (Cappers et al., 2010).

4.3.8. Recursos energéticos distribuidos

Los Recursos Energéticos Distribuidos, DER por sus siglas en inglés, son recursos a pequeña escala que pueden ser gestionados de forma automática o manual, instalados cerca de los centros de consumo, conectados a la red de distribución, con posibilidad de inyectar energía, consumir energía o proveer servicios complementarios a la red de forma dinámica. Dentro de los DER se incluyen la respuesta de la demanda, los vehículos eléctricos, la generación distribuida, los sistemas de almacenamiento de energía eléctrica y la autogeneración a pequeña y gran escala conectados a la red de distribución, entre otros. (MME, 2021).

4.3.9. Redes inteligentes.

La red inteligente se puede definir como la red de energía eléctrica que se caracteriza por una infraestructura eficiente y confiable con el uso de técnicas sofisticadas y modernas de control, comunicación, detección y medición, así como por tener la capacidad de controlar todas las partes de la red y abordar los problemas si ocurren, cosa que no sucede en las de redes tradicionales (Zakaria et al., 2020).

4.3.10. Transición energética.

La transición energética se refiere a la transición del sector energético global hacia fuentes de energía renovables y eficiencia energética, reemplazando la producción de energía basada en combustibles fósiles y sistemas de consumo (petróleo, gas y carbón). Los cuatro pilares principales hacia esta transición energética, también conocida como 4D, será presentado de la siguiente manera: Descarbonización, Descentralización, Digitalización y Democratización (Doukas, 2022).

5. Marco metodológico

La metodología empleada en esta monografía para la revisión de la literatura sigue las directrices establecidas en el texto "Metodología de la investigación" (HernándezHernández Sampieri & Mendoza Torres, 2018). De acuerdo con el enfoque propuesto por los autores, en la Figura 1 se detalla el procedimiento a seguir para la ejecución de la revisión sistemática.

Figura 1. Metodología de la investigación



Fuente: Adaptada de Hernández, R. & Mendoza, C. (2018). *Metodología de la investigación : las rutas cuantitativa, cualitativa y mixta*. McGraw-Hill Education.

5.1. Establecimiento de la ecuación de búsqueda

En esta etapa, se procuró establecer una ecuación de búsqueda mediante palabras clave, conceptos y terminología relacionada a la investigación, que, al ser introducida en las diferentes bases de datos IEEE, *Scopus* y *Web of Science* (WoS), permitiera encontrar los artículos que representaran información relevante respecto a la temática a profundizar.

5.2. Búsqueda preliminar y refinación de la ecuación

Siguiente al establecimiento de la ecuación de búsqueda inicial, se realizó una búsqueda y lectura preliminar para analizar los artículos encontrados inicialmente con el fin de plantear

criterios de inclusión y exclusión de manera que la ecuación de búsqueda final omitiera artículos que no eran de interés y de esta manera fuese refinada.

5.3. Revisión de artículos seleccionados

Una vez refinada la ecuación, de los artículos y publicaciones generados se fueron escogiendo aquellos identificados como relevantes y a su vez omitiendo aquellos identificados como no relevantes para la revisión de información mediante la lectura de títulos, resúmenes y palabras claves.

5.4. Depuración de la información

Después de la revisión anterior, se prosiguió con la lectura de los artículos escogidos, que permitió a su vez seguir depurando la información y conservar únicamente la relevante.

5.5. Resumen de la información de valor recolectada

Una vez depurada la información se procedió a realizar un resumen de cada método encontrado, procurando en la medida de lo posible recopilar la información de manera uniforme para todos los métodos, según criterios establecidos, que serán abordados posteriormente en el siguiente documento.

5.6. Clasificación de los métodos seleccionados

Después de resumir la información se realizó un cuadro comparativo de los diferentes métodos, con el propósito de recopilar y exponer el funcionamiento de cada método, sus usos, ventajas, desventajas, posibles hallazgos, casos de estudio y aplicación.

5.7. Análisis de resultados

Finalmente, se concluyó la revisión de la información, se incluyeron recomendaciones y posibles líneas futuras de investigación generadas a partir de la revisión.

6. Resultados

En este capítulo se detalla lo obtenido luego de la aplicación de la metodología, por ejemplo, los metadatos y ecuaciones de búsqueda. Seguidamente, se presentan los métodos de caracterización de usuarios y se realiza un cuadro comparativo mostrando una breve descripción, aplicación, caso de estudio, ventajas y desventajas. Después, se clasifican los métodos de acuerdo con los criterios establecidos y se analizan los resultados obtenidos. Finalmente, se describen los retos y oportunidades del análisis de datos en AMI y se presentan las futuras líneas de investigación identificadas.

6.1. Establecimiento de la ecuación de búsqueda

Con el propósito de abordar la temática de investigación, se estableció una ecuación de búsqueda centrada en identificar los principales métodos existentes y utilizados actualmente para la caracterización de usuarios del servicio de energía eléctrica a partir de datos generados por la infraestructura de medición avanzada (AMI), de forma que esta revisión sea una herramienta de ayuda para los operadores de red a nivel regional.

A continuación, se expone el planteamiento de la ecuación de búsqueda para las bases de datos IEEE, Scopus y Web of Science (WoS), las cuales pertenecen al Institute for Scientific Information (ISI) y que fueron seleccionadas debido su facilidad de acceso para el investigador, siendo estas proporcionadas por la Universidad Industrial de Santander y a su vez por tener gran reconocimiento y confiabilidad en el ámbito científico.

Se definió que la búsqueda se realizaría en los idiomas inglés y español y se identificaron los siguientes conceptos y/o palabras claves para la temática de investigación:

- Métodos
- Caracterización
- Usuarios
- Servicio de energía eléctrica
- Infraestructura de medición avanzada (AMI)
- Análisis de datos

En la Tabla 2 se observan las palabras clave seleccionadas y posteriormente la ecuación de búsqueda establecida inicialmente.

Tabla 2. *Palabras clave y relacionadas*

Concepto	Palabras asociadas
Métodos de caracterización / <i>Characterization methods</i>	<i>Methods, Tools, Measurement, Metric, System, Technique, Classification, Model, Ranking, Characterization</i>
Comportamiento Usuarios / <i>Users</i>	<i>User, consumer, beneficiary, client, usufructuary, buyer, acquirer, customer, consumption behavior, consumption pattern</i>
Servicio de energía eléctrica / <i>Electric energy service</i>	<i>Electric power, energy, electrical service, electricity service, electricity supply, electrical power, energy supply, power supply, public service, energy public service</i>
Infraestructura de medición avanzada (AMI) / <i>Metering Infrastructure (AMI)</i>	<i>Advanced Metering Infrastructure (AMI), Grid system, fixed grid, Advanced integrated system, smart system, smart metering, smart meter, Metering Infrastructure (AMI)</i>

watthorimeter, electric meter, electricity meter, grid integrated system, bidirectional metering, bidirectional meter, smart networks

Análisis de datos / *Data Analysis* *Data exploration, data investigation, data observation, data classification, data measurement, data characterization, data processing, data observation, information characterization, information classification, information processing, information observation*

Después de establecer las palabras clave y relacionadas a la temática de investigación, se elaboró la ecuación de búsqueda para las bases de datos mencionadas con tipo de búsqueda: *Article Title, Abstract y Keywords*. Luego a esto, se correlacionaron los diferentes términos mediante truncadores y operadores booleanos para refinar la búsqueda de información. A continuación, se evidencian la ecuación de búsqueda establecida inicialmente:

TITLE-ABS-KEY (((*method* OR "Measurement*" OR "Metric" OR "System*" OR "Technique*" OR "Model*" OR "Ranking" OR "Characterization" OR "Classification") AND (*user* OR "Consumer*" OR "User* classification" OR "Consumer* classification" OR "User* Characterization" OR "Consumer* characterization") AND (*consumption AND behavior* OR "consumption pattern*" OR "Consumption") AND (*electric AND power OR "energy" OR "electrical service" OR "electricity service" OR "electricity supply" OR "electrical power" OR "energy supply" OR "power supply" OR "energy public service") AND (*advanced AND metering AND infrastructure OR "Advanced measurement infrastructure" OR "AMI" OR "Grid system" OR "fixed grid integrated system*" OR "smart system*" OR "smart meter*" OR "wattmeter" OR "electric meter*" OR "electric grid integrated system*" OR "bidirectional meter*" OR "smart grid") AND (*data AND exploration OR " data investigation" OR "data observation" OR "data classification" OR "data measurement*" OR "data characterization" OR "data processing" OR "data observation" OR "information characterization" OR "information classification" OR "information processing" OR "information observation"))))******

6.2. Búsqueda preliminar y refinación de la ecuación

Se observó que al utilizar el anterior planteamiento de la ecuación de búsqueda en las bases de datos IEEE y Web Of Science, no fue posible obtener ningún documento como resultado, dado que se excede el número de palabras claves que es permitido por cada una. Sin embargo, al ingresar la ecuación en la base de datos Scopus, se obtuvo un total de 8 documentos, de los cuales 2 eran artículos y 6 papers de conferencia. Mediante una lectura rápida de títulos, resúmenes y palabras clave, se identificó que solo 2 de estos se relacionaban con la temática de investigación. Por lo tanto, se decidió refinar la ecuación de búsqueda, simplificando y generalizando la misma de manera que se pudiera obtener una mayor cantidad de resultados y de mejor calidad.

Después de varios intentos, lecturas rápidas de títulos y múltiples iteraciones en las 3 bases de datos, se establecieron las siguientes ecuaciones de búsqueda obteniendo resultados presentados en la Tabla 3.

Tabla 3. Ecuaciones de búsqueda utilizadas

Base de datos	Ecuación de búsqueda utilizada	Resultados	Idioma	Tipo de documento
IEEE	<i>((Method* OR "Measurement*" OR "Metric" OR "Technique*" OR "Model*" OR "Characterization" OR "Classification") AND (User* OR "Consumer*" OR "Customer*") AND (Advanced AND measurement AND infrastructure OR "AMI") AND (electri* OR "energy") AND (data AND analysis) AND (smart AND meter*) AND (consumption AND behavior OR "consumption pattern*"))</i>	(32)	English (32)	Conferences (23) Journals (8) Book (1)
Scopus	<i>((Method* OR "Measurement*" OR "Metric" OR "System*" OR "Technique*" OR "Model*" OR "Ranking" OR "Characterization" OR "Classification") AND (User* OR "Consumer*" OR "User* classification" OR "Consumer* classification" OR "User* Characterization" OR "Consumer* characterization") AND (Consumption behavior* OR "consumption pattern*" OR "Consumption") AND (Electric power OR "energy" OR "electrical service" OR "electricity service" OR "electricity supply" OR "electrical power" OR "energy supply" OR "power supply" OR "energy public service") AND (Grid system OR "fixed grid integrated system*" OR "smart system*" OR "smart meter*" OR "wattmeter" OR "electric meter*" OR "electric grid integrated system*" OR "bidirectional meter*" Or "smart grid") AND (Data exploration OR " data investigation" OR "data observation" OR "data classification" OR "data</i>	(76)	English (73) Chinese (3)	Conference paper (48) Article (23) Conference review (2) Review (1) Book chapter (1) Book (1)

measurement" OR "data characterization" OR "data processing" OR "data observation" OR "information characterization" OR "information classification" OR "information processing" OR "information observation"))*

<p>Scopus / Web of Science</p>	<p><i>((Method* OR "Measurement*" OR "Metric" OR "System*" OR "Technique*" OR "Model*" OR "Ranking" OR "Characterization" OR "Classification") AND (Consumption behavior* OR "consumption pattern*") AND (Electric power OR "energy" OR "electrical service" OR "electricity service" OR "electricity supply" OR "electrical power" OR "energy supply" OR "power supply" OR "energy public service") AND (User* OR "Consumer*") AND (Grid system OR "fixed grid integrated system*" OR "smart system*" OR "smart meter*" OR "wattmeter" OR "electric meter*" OR "electric grid integrated system*" OR "bidirectional meter*" Or "smart grid") AND (Data exploration OR " data investigation" OR "data observation" OR "data classification" OR "data measurement*" OR "data characterization" OR "data processing" OR "data observation" OR "information characterization" OR "information classification" OR "information processing" OR "information observation"))</i></p>	<p>(25) Scopus (18) Web of Science</p>	<p>Scopus English (24) Chinese (1) Web of Science English(18)</p>	<p>Scopus Conference paper (24) Article (7) Conference review (1) Review (1) Web of Science Article (18)</p>
<p>Scopus / Web of Science</p>	<p><i>((Advanced AND measurement AND infrastructure OR "AMI") AND (electri* AND (data AND analysis) AND (smart AND meter*) AND (consumption AND behavior OR "consumption pattern*"))</i></p>	<p>(14) Scopus (16) Web of Science</p>	<p>Scopus English (14) Web of Science English (16)</p>	<p>Scopus Article (15) Early Access (1) Review Article (1) Web of Science Article (18)</p>

Hasta este momento no se habían agregado filtros adicionales en las bases de datos, sin embargo, se observó que los documentos obtenidos se concentran mayormente desde los años 2019 a 2024, quizás argumentando que es una temática relativamente nueva. Aun así, se encontraron documentos relacionados a la temática de investigación desde el año 2010 hasta la actualidad. A su vez se evidencia que la mayoría de los documentos obtenidos en esta búsqueda preliminar son resultado de conferencias o investigaciones realizadas en su mayoría de los siguientes países en ese respectivo orden: China, India, Arabia Saudí, Reino Unido, Estados Unidos, y gran cantidad de países europeos. Así mismo se identificaron dos documentos relacionados a la temática con autores colombianos. Posteriormente, después de una lectura a mayor profundidad de la documentación, se establecerá cuáles documentos son de real valor para la monografía, pudiendo concluir los años o el lapso al que pertenecen, los países de proveniencia, tipos de documentos, sus idiomas, entre otros datos de interés.

6.3. Revisión de los artículos seleccionados

Para la revisión de los documentos obtenidos en la anterior búsqueda se establecieron 3 grupos de documentos, uno por cada base de datos utilizada. Inicialmente el principal objetivo de la revisión consistió en identificar artículos que estuvieran duplicados entre las diferentes bases de datos y entre las diferentes ecuaciones de búsqueda utilizadas. La Figura 2, explica simplificada los hallazgos.

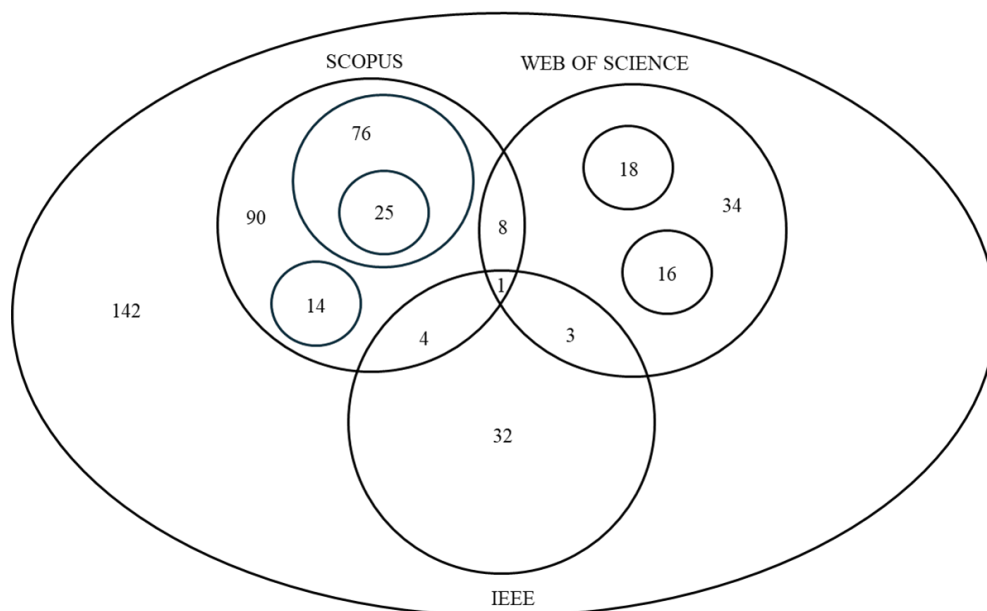
Mediante lectura rápida de títulos se pudo identificar que para la base de datos Scopus, los 25 documentos hallados con la segunda ecuación de búsqueda estaban ya incluidos dentro de los 76 documentos encontrados inicialmente; sin embargo, los 14 artículos obtenidos en esta

misma base de datos con la última ecuación de búsqueda eran completamente diferentes a los encontrados en el grupo de 25 y del grupo de 76.

Para la base de datos *Web Of Science* se evidenció que los 18 documentos generados con la tercera ecuación de búsqueda son completamente diferentes a los 16 generados con la cuarta ecuación de búsqueda. A su vez se presentaron ocho coincidencias con los documentos de Scopus, cuatro en el grupo de los 18 y cuatro en el de los 16.

Para la base de datos IEEE solo se presentaron cuatro coincidencias respecto a los documentos de Scopus y tres coincidencias respecto a los de *Web Of Science*.

Figura 2. *Coincidencias de documentos entre bases de datos*



Fuente: Elaboración propia

6.4. Depuración de la información

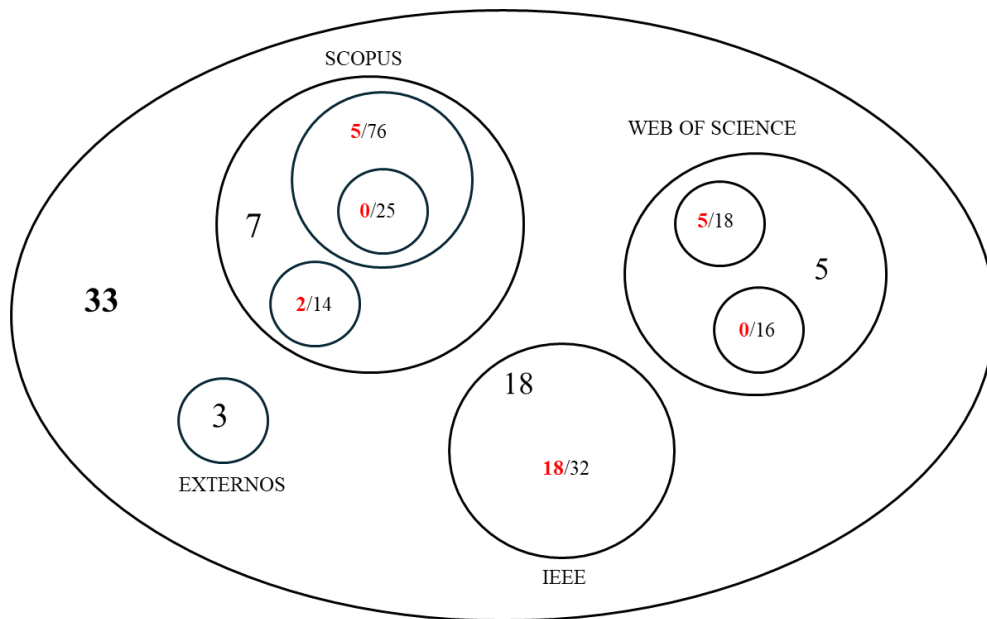
Después de identificar y eliminar los duplicados, se obtuvieron 142 documentos; sin embargo, esto no significó que la totalidad de estos fuesen claves para la investigación, por lo cual, se descargaron (de las diferentes bases de datos) y agruparon en un documento Excel la lista de resúmenes y palabras clave de estos, clasificándolos después como Relevantes o No Relevantes para la investigación. En esta etapa se obtuvo como resultado un total 45 documentos que se consideraron relevantes para la temática a investigar hasta ese momento.

Después, se inició con una lectura más completa y profunda de estos 45 documentos añadiendo introducción, contenido general, análisis de resultados y conclusiones. En esa etapa se descartaron 12 documentos adicionales, dando como resultado un total de 33 documentos. Los 12 artículos fueron descartados, dado que no se enfocaban en la temática a tratar, o se relacionaban, pero no aportaban la información de valor requerida.

En la Figura 3 se evidencia la proveniencia de los 33 artículos escogidos como relevantes para la investigación, especificando cuántos artículos fueron finalmente escogidos por cada grupo de las bases de datos.

Esta etapa concluyó con que, de las cuatro ecuaciones de búsqueda utilizadas, y de los documentos obtenidos en las tres bases de datos, se lograron recopilar 33 documentos que realmente aportan valor a esta investigación, ya que poseen contenido clave que permitirá el cumplimiento de los objetivos propuestos, o temáticas relacionadas que sirven de apoyo a la misma.

Finalmente, se realizó la lectura de cada uno de estos 33 documentos entre dos a cinco veces según fue necesario, para la captura de la información requerida para desarrollar la monografía.

Figura 3. Artículos escogidos después de la depuración

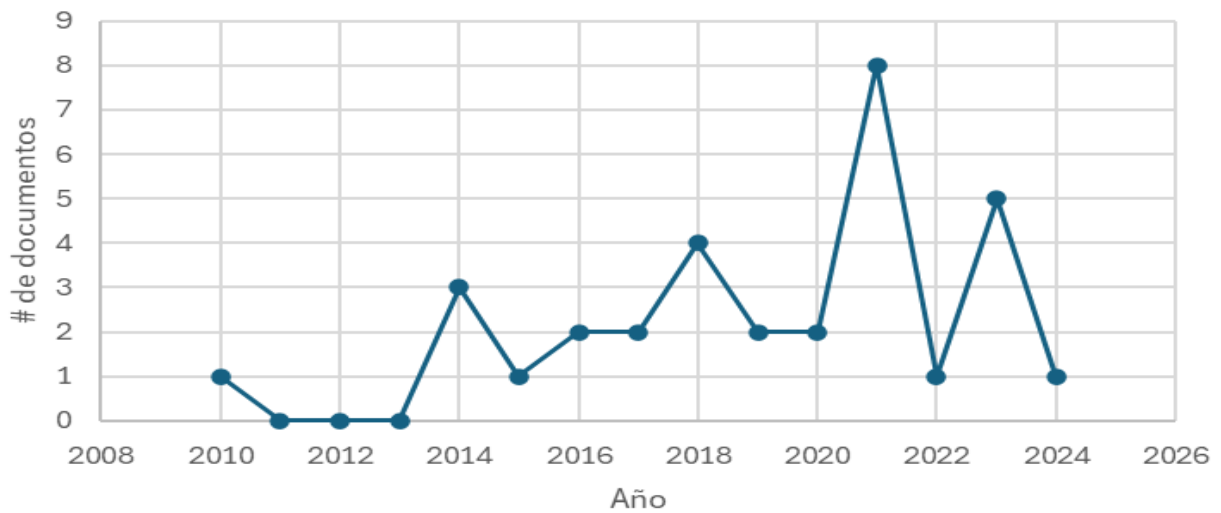
Fuente: Elaboración propia

6.4.1. Análisis bibliométrico por factores

El análisis bibliométrico se realizó para los siguientes factores: Años de publicación autores relevantes, palabras claves de los diferentes autores y países.

Para empezar, se observó que la temática de interés ya iniciaba a tener un crecimiento gradual dentro de las bases de datos desde el año 2010 hasta el año 2018. Sin embargo, los picos más altos de investigación acerca del tema han sido en los años 2021 y 2023 (ver Figura 4).

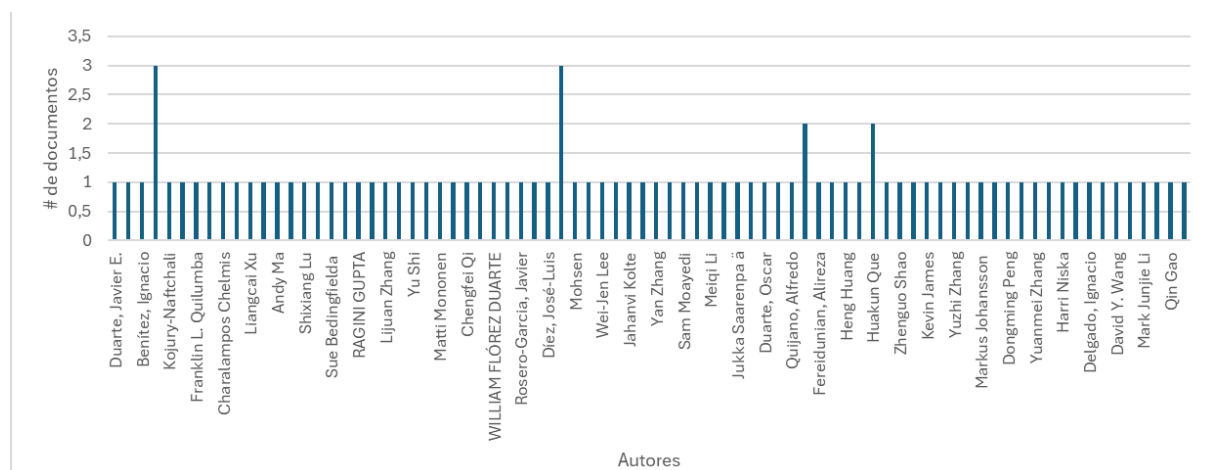
Figura 4. Producción científica anual



Fuente: Elaboración propia

En cuanto a los autores de la literatura encontrada, se observa que, figuran dos autores que cuentan con tres publicaciones cada uno, y otros 2 autores que cuentan cada uno con dos publicaciones acerca de la temática en estudio. Los autores más relevantes encontrados son: Hamid Lesani de Irán, Huakun Que de China, Mohsen Kojury-Naftchali de Irán y Alireza Fereidunian de Irán (ver Figura 5).

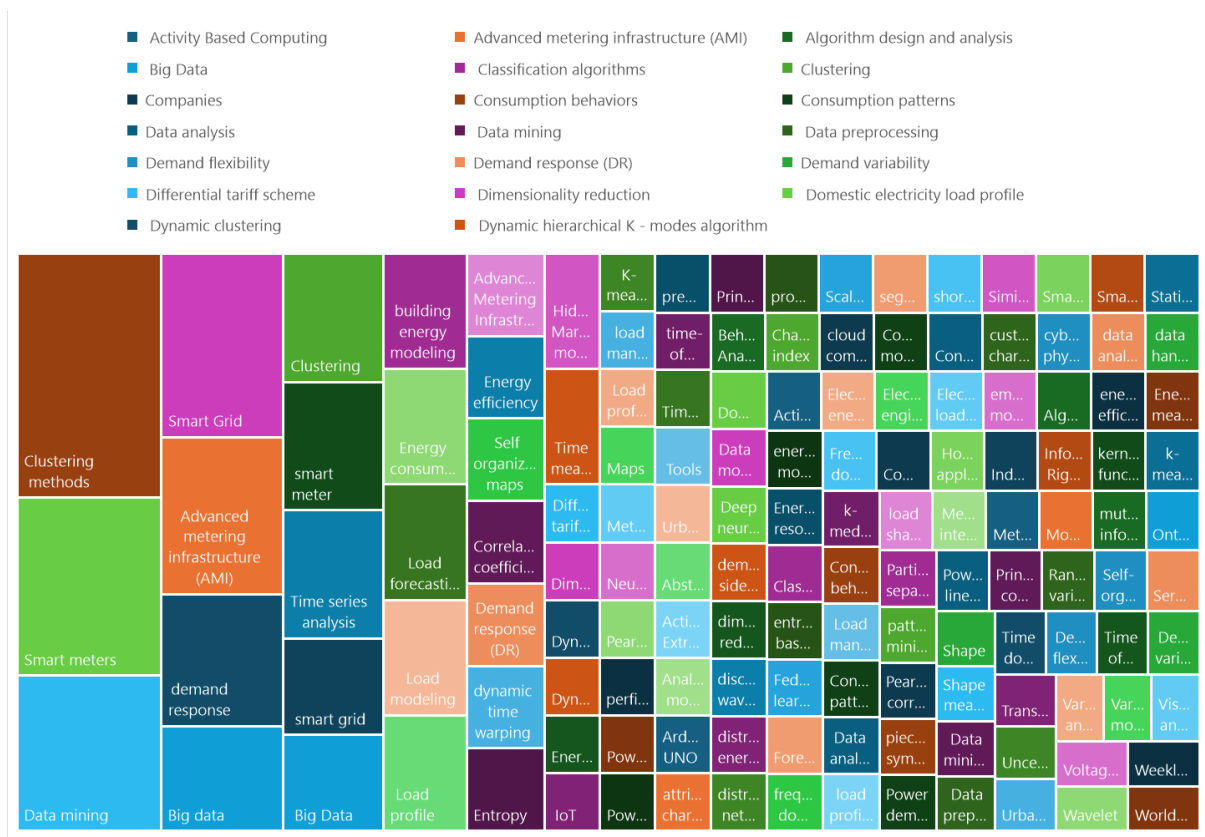
Figura 5. Clasificación de documentos por autores



Fuente: Elaboración propia

Por otra parte, se analizaron las palabras claves empleadas por los autores, con el fin de identificar aquellas que se utilizan con más frecuencia dentro de la temática de investigación, dando como resultado aquellas que inicialmente se definieron en las ecuaciones de búsqueda, tales como: Clustering methods, advanced infrastructure measurement (AMI), smart metering, smart grids, data mining, demand response, big data, energy consumption, consumption patterns, entre las más relevantes (ver Figura 6).

Figura 6. Clasificación de artículos por relación de palabras clave de los autores

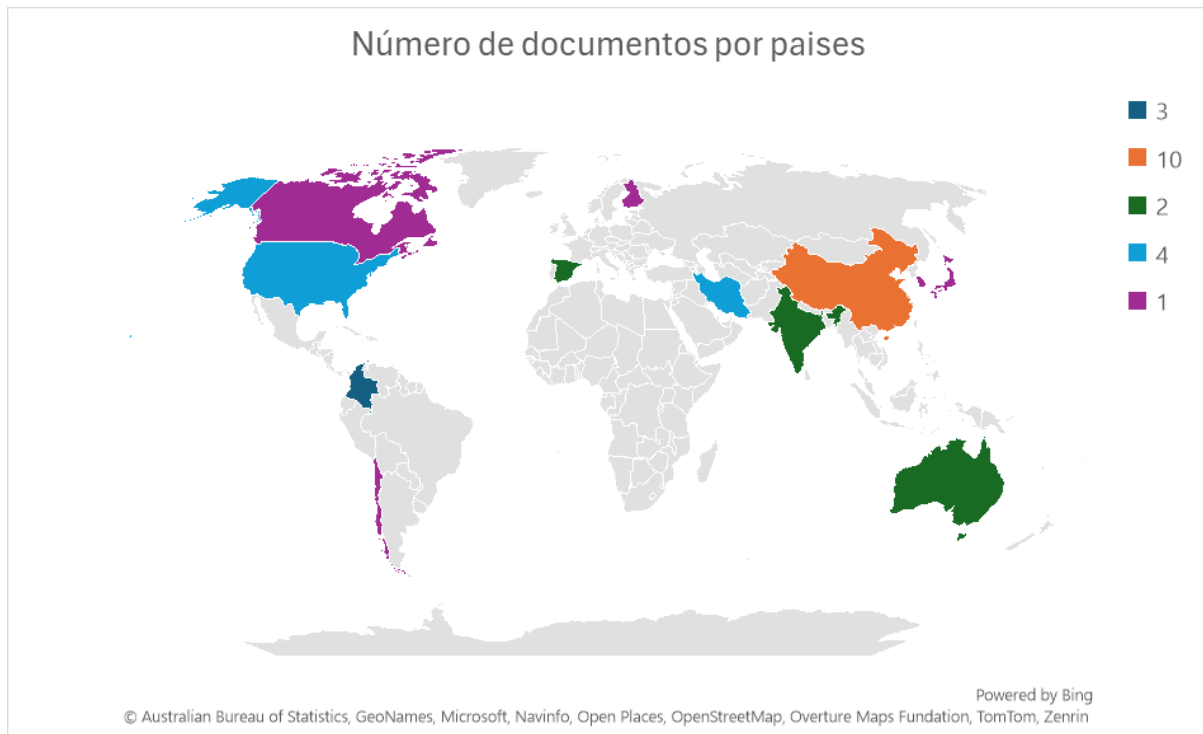


Fuente: Elaboración propia

Se observó también que la mayor cantidad de publicaciones acerca de la temática de interés se han realizado en China (10), Irán (4) y Estados Unidos (4), seguido de países como Colombia (3), España (2), India (2), Australia (2), Canadá (1), Corea del Sur (1), Japón (1),

Finlandia (1), Chile (1). Además, se debe tener en cuenta, que el 98% del total de publicaciones se encuentra en el idioma inglés y el restante en chino. En la figura 7 se puede observar la distribución geográfica de las publicaciones consolidadas.

Figura 7. Clasificación de documentos por países



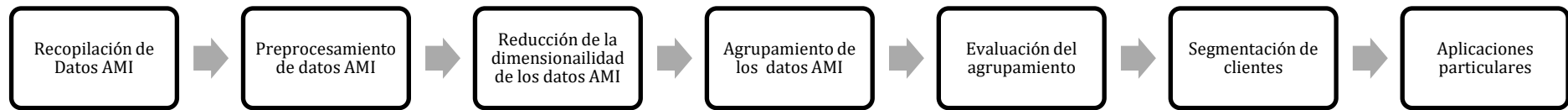
Fuente: Elaboración propia

6.5. Resumen de la información de valor recolectada

De acuerdo con la literatura estudiada, las metodologías más usadas para la clasificación de usuarios de energía eléctrica son las basadas en clústeres o agrupamiento, series de tiempo, métodos determinísticos, métodos estadísticos y nuevas técnicas basadas en redes neuronales y manejo de grandes volúmenes de datos. Para los métodos basados en clústeres se realiza una subdivisión de acuerdo con el algoritmo empleado en su análisis, dentro de los cuales se encuentran: los basados en particiones, niveles jerárquicos, densidad, cuadrícula y modelos (Si et al., 2021).

Los métodos presentados tienen un algoritmo de base, pero estos generalmente son modificados, mejorados y adaptados para cada caso en particular dependiendo de diferentes factores como el volumen de datos, aplicación, restricciones de costos, restricciones de infraestructura, entre otros. En consecuencia, los algoritmos son adaptados con nuevas técnicas de programación y análisis de datos para la reducción de la dimensionalidad, segmentación y clasificación de estos. La Figura 8 presenta un flujograma con el proceso base para la recopilación, procesamiento, clasificación y aplicación de la información obtenida a través de los medidores avanzados, este esquema es adaptado y mejorado del propuesto para la elaboración de perfiles de carga (Wang et al., 2015).

Figura 8. *Procedimiento básico de clasificación de datos AMI.*



Fuente: Adaptada de Wang, Y (2015). Load Profiling and Its Application to Demand Response: A Review. *IEEE*, 20(2).

Los métodos basados en agrupamiento (*Clúster*), se caracterizan principalmente por realizar cálculos de similitud, reducir la dimensionalidad de los datos, encontrar el número de óptimo de clústeres, evaluar mediante índices los clústeres hallados y su flexibilidad por adaptarse a diferentes aplicaciones, uno de los algoritmos más usado es el *k-means*. Las series temporales se enfocan en la reducción de la dimensionalidad, es muy usada para datos de consumo de energía eléctrica. Además, aparecen métodos convencionales como métodos estadísticos mediante histogramas y métodos determinísticos empleando regresión polinomial con pruebas de bondad y ajuste. Finalmente, se encuentran técnicas particulares de visualización y procesamiento de datos en línea.

En la tabla 4, se presenta un resumen de los métodos de caracterización de usuarios encontrados en la literatura presentando una breve descripción, aplicaciones en diferentes campos de la energía eléctrica y análisis de datos, casos de estudios reales realizados en diferentes partes del mundo y finalmente las posibles ventajas y desventajas que presentan estos métodos en su implementación.

Tabla 4. *Métodos de caracterización de usuarios encontrados en la literatura*

Método	Descripción	Aplicación	Casos de estudio	Ventajas	Desventajas
Clúster Basado en Particiones	<p><i>Algoritmo k-means</i> <i>Cálculo de centroides</i></p> <ul style="list-style-type: none"> • Elbow method • Media de miembros <p><i>Medición de similitud</i></p> <ul style="list-style-type: none"> • Euclidiano <p><i>Índices de evaluación</i></p>	<p>El método emplea algoritmos de agrupamiento no supervisado, como el <i>k-means</i>, para segmentar perfiles de carga según similitudes en comportamientos de consumo aplicados a mejorar la precisión de la previsión de carga.</p> <p>(Quilumba et al., 2014)</p>	<p>Caso de estudio aplicado a 819 medidores inteligentes clientes residenciales en New York, USA. Perfiles de cargo promedio durante 15 de julio de 2013 al 21 de julio de 2013</p>	<p>Reducción de errores al trabajar con datos segmentados. Identificación de patrones de consumo Posibilidad de trabajar con grandes volúmenes de datos gracias a técnicas de minería.</p>	<p>Dependencia de la calidad de los datos. Complejidad computacional, evitar mínimos locales. Interpretación de resultados.</p>

Método	Descripción	Aplicación	Casos de estudio	Ventajas	Desventajas
	<ul style="list-style-type: none"> Índice Silueta (SC) Davies Bouldin (DB) Índice de entropía MSE RST 	<p>El método consiste en un enfoque de dos niveles que segmenta y clasifica a los usuarios en grupos homogéneos basados en sus patrones de consumo, características de sus viviendas y datos socioeconómicos. Este método emplea algoritmos de <i>clustering</i> ponderado k-means para crear grupos relevantes y comparar el consumo de electricidad con grupos similares.(Mononen et al., 2014)</p>	<p>El método está en fase de desarrollo y pruebas como parte del proyecto finlandés SGEM. Donde, se estudia un hogar seleccionado al azar en dos casos: a) evaluación comparativa con hogares similares, y b) evaluación comparativa con hogares similares que tienen un sistema de calefacción diferente</p>	<p>Considera múltiples factores (carga eléctrica, características del hogar, clima). Admite diferentes variables según la necesidad. Promueve eficiencia energética y reduce emisiones de CO2.</p>	<p>Requiere integración de múltiples fuentes de datos y procesos de normalización. Dependencia de calidad de datos Implementación inicial y mantenimiento pueden ser altos.</p>
Clusters Basado en particiones	<p><i>Algoritmo</i></p> <ul style="list-style-type: none"> k-means <p><i>Cálculo de centroides</i></p> <ul style="list-style-type: none"> Elbow method Media de miembros <p><i>Medición de similitud</i></p> <ul style="list-style-type: none"> Euclidiano <p><i>Índices de evaluación</i></p> <ul style="list-style-type: none"> Índice Silueta (SC) Davies Bouldin (DB) Índice de entropía MSE RST 	<p>Este marco combina técnicas de aprendizaje automático para pronósticos de carga a corto plazo (STLF), análisis de patrones de consumo usando algoritmo k-means, y estadísticas no paramétricas para estimar la reducción del consumo durante eventos de respuesta a la demanda (DR). (Zhou et al., 2016)</p>	<p>Estudio realizado en el Oeste de Estados Unidos a 500 usuarios residenciales entre 2012 y 2014 para programas de DR teniendo en cuenta la relación Temperatura - Consumo.</p>	<p>Uso eficiente de datos AMI granulares para personalizar estrategias de DR. Aplicación de técnicas robustas de aprendizaje automático y estadística. Identificación proactiva de clientes con alta flexibilidad de consumo.</p>	<p>Dependencia de modelos predictivos que pueden ser sensibles a errores de especificación. Necesidad de datos históricos extensos y de alta calidad. Complejidad computacional en métodos como regresión de vectores de soporte (SVR) y clustering.</p>
		<p>El método utiliza medidores avanzados junto con la infraestructura AMI para recolectar datos sobre el consumo eléctrico de los usuarios mediante la comunicación por la línea de energía (PLC, <i>Power Line</i></p>	<p>Estudio realizado en India. Su objetivo era monitorear en tiempo real de redes eléctricas en zonas urbanas y rurales. Identificación y mitigación de hurto eléctrico en redes distribuidas. Como trabajo futuro se propone el método</p>	<p>Facilidad operativa Convergencia rápida Altamente explicativo Resultado de agrupación de carga preferible</p>	<p>Sensible a los valores atípicos Fácil de caer en un óptimo local Adquisición de software como Matlab.</p>

Método	Descripción	Aplicación	Casos de estudio	Ventajas	Desventajas
		<p><i>Communication</i>). Este enfoque permite el monitoreo remoto del consumo energético y proporciona una visión detallada de los patrones de uso. El análisis se realiza utilizando simulaciones en MATLAB y algoritmos de agrupamiento (clustering) como k-means. (Elakshum & Ponraj, 2017)</p>	<p>de agrupación k-means con el enfoque de transformación de wavelet en aplicaciones de DR e integración de servicios en la nube por medio de IoT.</p>		
Clusters Basado en particiones	<p><i>Algoritmo</i></p> <ul style="list-style-type: none"> ● k-means ● <i>Cálculo de centroides</i> Elbow method ● Media de miembros <p><i>Medición de similitud</i></p> <ul style="list-style-type: none"> ● Euclidiano <p><i>Índices de evaluación</i></p> <ul style="list-style-type: none"> ● Índice Silueta (SC) ● Davies Bouldin (DB) ● Índice de entropía ● MSE ● RST 	<p>El método propuesto utiliza técnicas de clustering, específicamente el algoritmo K-means y Elbow method, para segmentar datos de consumo eléctrico obtenidos de medidores inteligentes. Se desarrolló una biblioteca de código abierto basada en el entorno de software R para estandarizar los datos de medición inteligente, realizar un análisis estadístico y correlacionar los patrones de consumo con la información geográfica. (Cely Callejas, 2018)</p>	<p>Proyecto piloto en Bogotá, Colombia, con datos de 6060 medidores inteligentes distribuidos en 14 localidades a lo largo de un periodo de 4 meses. Segmentación por localidad y estrato socioeconómico para analizar el consumo eléctrico en distintas zonas y niveles económicos.</p>	<p>Elimina registros anómalos y reduce el ruido en los datos. Adaptable a distintas geografías y niveles socioeconómicos. Permite un análisis detallado de grandes volúmenes de datos en poco tiempo.</p>	<p>Altas tasas de datos no válidos pueden influir en la representatividad del análisis. Requiere herramientas específicas como R. Los resultados dependen de una adecuada selección del número de clusters.</p>
		<p>El método utiliza datos de series temporales provenientes de medidores inteligentes (AMI) para caracterizar y predecir patrones de consumo de energía. Combina técnicas de minería de datos (como</p>	<p>La primera recopilación de datos real incluye más de 5 casas con una liquidación oportuna de más de 400 millones de documentos de uso de energía bruta en India. Durante el período de</p>	<p>Alta precisión en predicciones (hasta 99.6 con datos bien entrenados). Personalización de análisis para usuarios y grupos.</p>	<p>Complejidad computacional elevada debido al uso de redes neuronales profundas. Requiere grandes volúmenes de datos y recursos computacionales.</p>

Método	Descripción	Aplicación	Casos de estudio	Ventajas	Desventajas
		agrupamiento por k-means y minería de patrones frecuentes) y modelos de aprendizaje profundo (LSTM y GRU) para analizar y clasificar el comportamiento energético de los usuarios.(Balachander & Paulraj, 2021)	preprocesamiento, se redujo a 20 millones. Del mismo modo, de más de 21 millones de registros sin procesar, el segundo conjunto de datos real AMPds2 se redujo a cuatro millones de registros, originalmente de 1 minuto de duración.	Capacidad de actualización incremental para mantener la relevancia del modelo.	Dependencia de la calidad y granularidad de los datos recolectados.
Clústeres Basado en particiones	<p><i>Algoritmo</i></p> <ul style="list-style-type: none"> ● k-means ● <i>Cálculo de centroides</i> Elbow method ● Media de miembros <p><i>Medición de similitud</i></p> <ul style="list-style-type: none"> ● Euclidiano <p><i>Índices de evaluación</i></p> <ul style="list-style-type: none"> ● Índice Silueta (SC) ● Davies Bouldin (DB) ● Índice de entropía ● MSE ● RST 	Se utiliza una metodología basada en el algoritmo K-means para identificar patrones de comportamiento en el consumo de sus clientes. En particular, se caracterizan grupos conformados por medio de sus centroides, posibilitando la conversión de los grandes volúmenes de datos.(Marrero et al., 2021)	El estudio consideró los consumos de energía obtenidos de 1179 medidores inteligentes asociados a los clientes residenciales del sur de Chile en 2017. Se efectúan mediciones en intervalos de 15 minutos y corresponden al periodo comprendido entre el 4 de marzo y el 4 de abril de 2019.	Proporciona una visión clara de patrones de consumo para empresas distribuidoras. Mejora la planificación energética y fomenta la eficiencia. Metodología replicable con diferentes conjuntos de datos.	Sensible a la selección de parámetros iniciales y número de clústeres. Puede necesitar técnicas adicionales para manejar datos incompletos o ruido significativo.
		El algoritmo red neuronal basado en la optimización de enjambres de partículas y k-means mejorado con computación en la nube. Además, de tecnologías como SparkR en lenguaje R optimiza el análisis de grandes volúmenes de datos. (Li et al., 2022)	Caso de análisis de grandes volúmenes de datos en empresas industriales, residentes urbanos y rurales de China.	Eficiencia en el análisis gracias a la computación en la nube. Adaptabilidad a diferentes tipos de datos y usuarios. Posibilidad de detectar anomalías rápidamente.	Sensibilidad a los valores iniciales. Complejidad en el manejo de grandes volúmenes de datos. Adquisición de plataformas de análisis de big data.
Clústeres Basado en particiones	<p><i>Algoritmo</i></p> <ul style="list-style-type: none"> ● k-means <p><i>Cálculo de centroides</i></p>	Se propone un método basado en el algoritmo K-means para clasificar y analizar patrones de consumo eléctrico de los	Este trabajo se realizó en China, donde se ejecuta el algoritmo de servicio jerárquico propuesto	Mejora la asignación de recursos energéticos y reduce costos operativos. Permite desarrollar	Puede ser costoso en términos de procesamiento, especialmente con grandes

Método	Descripción	Aplicación	Casos de estudio	Ventajas	Desventajas
	<ul style="list-style-type: none"> Media de miembros Elbow method <p>Medición de similitud</p> <ul style="list-style-type: none"> Euclidiano <p>Índices de evaluación</p> <ul style="list-style-type: none"> Índice Silueta (SC) Davies Bouldin (DB) Índice de entropía MSE RST 	<p>usuarios utilizando datos de AMI. Este método agrupa a los usuarios según similitudes en sus perfiles de consumo eléctrico, identificando patrones específicos y facilitando una gestión optimizada de la demanda energética. (Guo, 2023)</p> <hr/> <p>El método evalúa la variabilidad mediante el coeficiente de variación (CV), calculado como la relación entre la desviación estándar y el promedio del consumo horario. Se emplean técnicas de análisis de datos como el clustering (k-means) y el algoritmo a priori para identificar patrones de consumo y segmentar usuarios en grupos con características similares. (Duarte et al., 2024)</p>	<p>(Iyapunov) y el algoritmo codicioso (<i>Benchmark</i>) en tres escenarios eólicos, y se calcula el costo de tiempo promedio de cada intervalo de tiempo de acuerdo con el método anterior.</p> <hr/> <p>Estudio comprende 2.470.092.041 registros con un tamaño de 157 GB reflejando el consumo de 73.794 usuarios del sistema eléctrico colombiano entre el 2018 y el 2021. Datos suministrados por el OR de Bogotá, además el análisis está segmentando datos según días laborables, sábados y domingos/feriados. Se redujo la muestra de 73.794 usuarios iniciales a 44.574, 44.529 y 44.516 para cada tipo de día. Se redujo el almacenamiento de 19 a 4,3 GB.</p>	<p>estrategias específicas adaptadas a cada grupo de usuarios. Se adapta a diferentes contextos y sistemas energéticos.</p> <hr/> <p>Identifica intervalos de tarifas diferenciadas adaptados a patrones específicos de consumo. Eficiencia energética. Puede aplicarse a datos AMI en diferentes contextos geográficos.</p>	<p>conjuntos de datos. Los resultados dependen fuertemente del número de clústeres seleccionados.</p> <hr/> <p>Requiere grandes volúmenes de datos horarios limpios y estandarizados. Las etapas iniciales de procesamiento y análisis son intensivas en recursos. Riesgo potencial de exposición de datos sensibles de consumo si no se toman medidas adecuadas.</p>
Clústeres Basado en particiones	<p>Algoritmo</p> <ul style="list-style-type: none"> K-medoids <p>Cálculo de centroides</p> <ul style="list-style-type: none"> El miembro menos diferente a los demás en general <p>Medición de similitud</p> <ul style="list-style-type: none"> Varias medidas 	<p>Método utiliza componentes principales (PCA) y técnicas de clustering como k-medoids para analizar los datos de AMI. Se desarrolló un algoritmo novedoso para la agrupación de series temporales basado en la distancia de Hausdorff que</p>	<p>Los datos se obtuvieron de la microrred del campus de la Universidad del Sur de California. Estos comprenden una colección de valores de consumo de electricidad observados (medidos en kWh cada 15 minutos) de 115</p>	<p>Procesa grandes volúmenes de datos sin perder eficiencia. Aplicable a cualquier conjunto de datos temporales de alta resolución. Segmenta consumidores</p>	<p>Dependencia de la calidad de datos. La elección del número de clústeres y semillas iniciales puede influir significativamente. Requiere preprocesamiento intensivo para manejar</p>

Método	Descripción	Aplicación	Casos de estudio	Ventajas	Desventajas
	<ul style="list-style-type: none"> Distancia de Hausdorff DTW (Dynamic Time Warping) <p><i>Índice de evaluación</i></p> <ul style="list-style-type: none"> Índice de Dunn Índice de Calinski Harabasz EVI (Índice de varianza de energía) Índice COP <p><i>Reducción de dimensionalidad</i></p> <ul style="list-style-type: none"> Análisis de componentes principales (PCA) 	<p>agrupa edificios de manera eficiente bajo nuestra métrica de distancia y técnica de almacenamiento de datos.(Chelmis et al., 2015)</p> <hr/> <p>El método propuesto emplea la técnica de agrupamiento k-medoid con medidas de distancia basadas en DTW para analizar la similitud de las curvas de consumo. El índice COP se utiliza para validar y determinar el número óptimo de clústeres.(Oyedokun et al., 2018)</p>	<p>edificios, recopilados durante un período de cinco años (del 1 de enero de 2009 al 30 de diciembre de 2013), con un total de 18.127.680 puntos de datos en todos los medidores inteligentes.</p> <hr/> <p>El conjunto de datos considerado en este estudio comprende 99 registros de medidores inteligentes de clientes, que fueron proporcionados por Electric Ireland y Sustainable Energy Authority of Ireland. Se empleó el paquete dtwclust, desarrollado para el software estadístico R para implementar la agrupación de datos de medidores inteligentes. Se seleccionó un perfil de carga diaria representativo.</p>	<p>para enfocar esfuerzos en usuarios clave.</p> <hr/> <p>Alta precisión en la identificación de patrones de consumo individuales. Puede manejar grandes volúmenes de datos. Se adapta a diferentes granularidades de datos temporales.</p>	<p>datos complejos.</p> <hr/> <p>Complejidad computacional alta debido al uso de DTW. Sensibilidad al ruido en los datos del medidor inteligente. Requiere ajustes manuales para definir parámetros como el tamaño de ventana en DTW.</p>
Clústeres Basado en particiones	<p><i>Algoritmo</i></p> <ul style="list-style-type: none"> Dynamic Hierarchical K-Modes <p><i>Medición de similitud</i></p> <ul style="list-style-type: none"> Método simple de coincidencia <p><i>Métricas de distancia</i></p> <ul style="list-style-type: none"> Distancia de Hamming Método de enlace promedio 	<p>El método propuesto utiliza el algoritmo de Dynamic Hierarchical K-Modes siendo una extensión del algoritmo K-Modes tradicional, diseñado para datos categóricos, pero mejorado con un enfoque jerárquico dinámico que permite determinar el número óptimo de clústeres (kutilizando un análisis de dendrogramas. Además, se incluye un proceso de</p>	<p>Estudio realizado en China evalúa las ventajas de los modos K jerárquicos dinámicos en comparación con los modos K tradicionales. Se simula por computadora un conjunto de datos de curva con 10 dimensiones, 3 clases y 500, 200 y 50 muestras en cada clase. En el que el número de muestras bimodales es 500, el número de muestras unimodales es 200 y el</p>	<p>Automatización del número de clústeres. Maneja datos categóricos y continuos. Refleja mejor las formas de las curvas de consumo eléctrico.</p>	<p>Complejidad computacional La selección de umbrales y métodos de distancia puede influir en los resultados. Fenómenos de clasificación errónea Limitación en la interpretación</p>

Método	Descripción	Aplicación	Casos de estudio	Ventajas	Desventajas
		extracción de características de datos para representar las curvas de consumo eléctrico.(Shi et al., 2022)	número de muestras de un solo valle es 50.		
Clústeres Basado en particiones	<p><i>Algoritmo</i></p> <ul style="list-style-type: none"> • K-means adaptativo <p><i>Cálculo de centroides</i></p> <p>1) Media de los conglomerados</p> <p>2)Durante la formación del conglomerado</p> <p><i>Medición de similitud</i></p> <ul style="list-style-type: none"> • Euclidiano <p><i>Métricas de distancia</i></p> <ul style="list-style-type: none"> • DTW <p>DFT</p>	<p>Se proponen dos métodos de caracterización de usuarios en el dominio del tiempo y otro en el dominio de la frecuencia. Un algoritmo de clustering basado en K-means adaptativos y métricas de distancia DTW (Dynamic Time Warping) para el dominio del tiempo. El método en el dominio de la frecuencia utiliza la Transformada Discreta de Fourier (DFT) para analizar características periódicas en las series temporales.(Ma, 2023)</p>	Datos AMI recopilados de un importante proveedor de servicios públicos de Arizona que atiende a más de un millón de clientes en el área metropolitana de Phoenix, USA. Dado que se recopila un solo punto de datos para cada cliente cada hora, los datos comprenden más de 35 mil millones de entradas, con un volumen total de aproximadamente 2920 terabytes, por año.	<p>Número de conglomerados ajustado adaptativamente</p> <p>Capacidad para realizar múltiples tareas al mismo tiempo.</p>	<p>Sensible a los valores atípicos</p> <p>Propenso a caer en un óptimo local</p>
	<p><i>Algoritmo</i></p> <ul style="list-style-type: none"> • K-shape • <i>Cálculo de centroides</i> <p>Correlación cruzada</p> <ul style="list-style-type: none"> • Elbow method <p><i>Medición de similitud</i></p> <ul style="list-style-type: none"> • Distancias basado en la curva (SBD) <p><i>Índices de evaluación</i></p> <ul style="list-style-type: none"> • SSE 	<p>Este documento propone un método de análisis de similitud basado en k-shape para el comportamiento de consumo de electricidad de AMI, que puede realizar una rápida extracción de características y una clasificación final del comportamiento de consumo de electricidad.(Qi et al., 2023)</p>	En nuestro experimento, se seleccionaron 42.373 clientes en una región de China y se utilizó un monitoreo no intrusivo para obtener datos diarios de consumo de electricidad. Los datos seleccionados corresponden al período 2014-2016, con un total de 1034 entradas en kWh. Por último, se conservan 16.554 datos válidos	<p>Capaz de explorar funciones secuenciales</p>	<p>Difícil de extender a grandes conjuntos de datos</p> <p>Límites de validez a conjuntos de datos específicos</p>
Clústeres Basado en particiones	<p><i>Algoritmo</i></p> <ul style="list-style-type: none"> • Fuzzy c-means (FCM) 	El método presentado utiliza una combinación de técnicas de clasificación difusa para	Caso de estudio tomando como referencia el pronóstico del 10 de noviembre de 2017,	Reducción de errores al incorporar corrección de predicciones.	Entrenamiento de redes y cálculo de pesos pueden ser costosos en tiempo y

Método	Descripción	Aplicación	Casos de estudio	Ventajas	Desventajas
	<p><i>Cálculo de centroides</i></p> <p>1) Media de miembros</p> <p>2) Una instancia no puede ser clasificada</p> <p><i>Medición de similitud</i></p> <ul style="list-style-type: none"> • Euclidiano <p><i>Índices de evaluación</i></p> <ul style="list-style-type: none"> • Periodic Index • Dunn Index DBI 	<p>caracterizar usuarios y la red neuronal Elman para predecir y corregir el consumo de electricidad a corto plazo. (Lijuan, 2023)</p>	<p>con muestras de consumo en intervalos de 15 minutos para usuarios residenciales urbanos en China.</p>	<p>Sensibilidad a datos históricos y condiciones cambiantes.</p>	<p>recursos.</p> <p>Dependencia de datos de calidad</p> <p>El modelo puede necesitar ajustes para cada aplicación particular.</p> <p>Uso de técnicas avanzadas de inteligencia artificial</p>
		<p>El método propuesto es Federated Fuzzy k-Means (FFKM), diseñado para el análisis de comportamiento de usuarios del servicio eléctrico en redes inteligentes, preservando la privacidad de los datos. Combina técnicas de clustering difuso (fuzzy k-means) con aprendizaje federado, lo que permite un análisis distribuido sin centralizar los datos sensibles.(Wang et al., 2023)</p>	<p>Datos de medidores inteligentes abiertos de Irlanda y Londres. El primer conjunto de datos consta de 6.435 consumidores y el segundo consta de 5.567 hogares. Antes de entrar en la fase de agrupación, todos los perfiles de carga media diaria son: Irlanda contiene 1072 perfiles de carga y, en el conjunto de datos de Londres, cada cliente contiene 928 perfiles de carga.</p>	<p>No requiere compartir datos sensibles; los cálculos se realizan localmente. Admite grandes volúmenes de datos distribuidos. Más fácil de alcanzar el óptimo global Robusto a valores atípicos</p>	<p>Desafíos con datos no balanceados. Incremento en cálculos debido al enfoque distribuido y difuso.</p>
Clústeres Jerárquico	<p><i>Algoritmo</i></p> <ul style="list-style-type: none"> • Aglomerativo <p>La primera categoría:</p> <p><i>Método de vinculación</i></p> <ul style="list-style-type: none"> • Enlace completo • Enlace promedio <p>La segunda categoría:</p> <p>Algoritmos de centroides</p> <ul style="list-style-type: none"> • Enlace de Ward <p><i>Índices</i></p> <ul style="list-style-type: none"> • DPP • SVM 	<p>Método de medición de similitud mediante algoritmo de agrupamiento jerárquico ponderado basado en Pearson, adaptada para datos de curvas de carga eléctrica. Este enfoque introduce el concepto de "puntos de alteración de tendencia" para identificar cambios significativos en las curvas de carga. La distancia ponderada de Pearson (DPP) evalúa la similitud general.(Lin</p>	<p>Los casos de estudio se basan en dos conjuntos de datos de carga diaria de una subestación europea, estos fueron proporcionados por la Compañía de Distribución de Energía de Eslovaquia Oriental. Puntos de carga diario (48 MW), Datos de entrenamiento (730 días), Datos de prueba (31 días) The daily load data of a substation in Nanjing, China</p>	<p><i>Enlace completo</i></p> <p>Clústeres de forma más compacta</p> <p><i>Enlace promedio</i></p> <p>Compromiso entre enlaces simples y completos</p> <p>Resistente al ruido</p>	<p><i>Enlace completo</i></p> <p>Sensible a los valores atípicos</p> <p><i>Enlace promedio</i></p> <p>Computacionalmente costoso, especialmente para grandes conjuntos de datos</p>

Método	Descripción	Aplicación	Casos de estudio	Ventajas	Desventajas
		et al., 2018)	Puntos de carga diario (96 MW), Datos de entrenamiento (274 días), Datos de prueba (15 días)		
Clústeres Basado en la Densidad	<p><i>Algoritmo</i></p> <ul style="list-style-type: none"> • DBSCAN <p>Basado en vecindad Densidad alcanzable</p> <p><i>Medición</i></p> <ul style="list-style-type: none"> • Euclidean distance <p><i>Índice de evaluación:</i></p> <ul style="list-style-type: none"> • ARL • RP12 • RAPV • RTPV 	<p>En primer lugar, se propone un método para extraer el perfil típico de uso, en el que el algoritmo DBSCAN basado en densidad con una estrategia de corrección que apunta a eliminar la distorsión significativa en los resultados. En segundo lugar, se realiza un agrupamiento de K-medias para agrupar los perfiles de carga típicos.(Qiu et al., 2016b)</p>	<p>Los datos de consumo de electricidad real de más de 100 clientes industriales y comerciales en China, adquiridos en intervalos de 15 minutos de AMI se utilizan para verificar la eficacia de los métodos propuestos.</p>	<p>No es necesario especificar el número de clúster por adelantado Capaz de descubrir pistas de formas arbitrarias Insensible a las anomalías Identificación automática de anomalías</p>	<p>Peor calidad de agrupación para conjuntos de datos de carga con densidad no uniforme o con grandes diferencias en la distribución entre clústeres Alta complejidad temporal Ajuste de parámetros complicado</p>
Clúster Basado en Cuadrícula	<p><i>Algoritmo</i></p> <ul style="list-style-type: none"> • Mapas Auto-Organizados (SOM) <p>Red neuronal no supervisada</p> <p><i>Índices de evaluación</i></p> <ul style="list-style-type: none"> • MIA • CDI • GSOM 	<p>El método aplica los SOM de Kohonen, una técnica de redes neuronales no supervisada utilizada para el agrupamiento y la reducción de dimensionalidad de datos.(Kojury-Naftchali et al., 2018)</p>	<p>Datos reales proporcionados por el Archivo Irlandés de Datos de Ciencias Sociales (ISSDA) provenientes de AMI.</p>	<p>Identificación automática de centros de clústeres. Reducción de dimensionalidad en grandes volúmenes de datos. SOM permite ajustes en tiempo real con nuevos datos.</p>	<p>Alta Capacidad Computacional Dependencia en Datos de Calidad Ajuste Complejo de Parámetros.</p>
		El método utiliza un algoritmo escalable de mapas autoorganizados en crecimiento (GSOM, por sus siglas en inglés) para analizar y caracterizar el consumo eléctrico de hogares a partir de datos recopilados de medidores inteligentes. Este enfoque	Los datos de muestra fueron proporcionados por una gran empresa australiana de servicios públicos. Los datos comprendieron lecturas de media hora durante un año completo (del 1 de abril al 31 de marzo) para 10.000	Capacidad para procesar grandes volúmenes de datos en paralelo. Permite análisis detallado y de alto nivel simultáneamente. Reducción del tiempo de procesamiento en un 95% frente a otros métodos.	Requiere infraestructura de computación distribuida para grandes volúmenes de datos. Precisa registros completos y detallados para un análisis efectivo.

Método	Descripción	Aplicación	Casos de estudio	Ventajas	Desventajas
		identifica patrones de consumo a niveles múltiples de granularidad (diario, semanal, anual) y permite el análisis masivo de datos gracias a su capacidad paralela y escalable.(Bedingfield et al., 2018)	clientes. El conjunto de datos sin procesar tenía un tamaño aproximado de 12 gigabytes con más de 175 millones de registros relacionados con los 10.000 clientes.		
Clúster Basado en Cuadrícula	Algoritmo <ul style="list-style-type: none"> • Wavecluster Basado en cuadrículas con principio de transformación de ondículas (Wavelet) índices	El método se basa en un enfoque de agrupamiento (clustering) utilizando transformada wavelet discreta (DWT) y el algoritmo de k-means para caracterizar perfiles de carga de consumidores. (Xiao et al., 2014)	El conjunto de datos incluye más de 18000 usuarios, que pertenecen a los 5 principales tipos de fabricación en DongGuan. La red de recopilación de datos generará 80 GB de datos por mes desde DongGuan, que es una importante ciudad manufacturera en esta área de China. Estudio realizado en 2012.	Reduce la dimensionalidad sin perder información clave. Captura patrones en diferentes niveles de granularidad. Mejora los resultados del clustering mediante el uso de datos refinados en múltiples etapas.	Complejidad computacional. Dependencia de parámetros Implementar DWT y clustering de manera óptima requiere conocimientos avanzados.
	<ul style="list-style-type: none"> • DWT • PAA 	El método propuesto se basa en la extracción y análisis de la variabilidad de los perfiles de carga eléctrica de edificios mediante técnicas de descomposición de señales (DWT), utilizando datos de (AMI). Este análisis identifica patrones de variabilidad en el consumo energético a resoluciones de tiempo finas, como datos de un minuto o menos.(Parker et al., 2021)	Estudio de caso de un conjunto de datos de resolución temporal real de menos de un minuto, recopilado en 2019, para dos edificios en un estado del Medio Oeste de los EE. UU. Edificios energéticamente eficientes (GEB)	Proporciona detalles que permiten identificar picos de demanda y variaciones rápidas. Aplicable a cualquier conjunto de datos de series temporales, no limitado a perfiles de carga eléctrica. Ayuda a generar perfiles sintéticos que reflejan variabilidad real.	La descomposición de señales y el análisis a alta resolución requieren recursos significativos. La falta de datos disponibles a resoluciones inferiores a 15 minutos puede dificultar el análisis.
Algoritmo Basado en	Algoritmo <ul style="list-style-type: none"> • GMM 	El método combina segmentación basada en	La evaluación de los procedimientos propuestos se	Capaz de modelar la probabilidad de que se	Iteración computacionalmente

Método	Descripción	Aplicación	Casos de estudio	Ventajas	Desventajas
Modelos	<p>Modelo paramétrico que se puede ver como una combinación de k modelos gaussianos simples</p> <p><i>Índice</i></p> <p>FCA (Adecuación de agrupamiento difuso)</p>	<p>entropía ponderada localmente y selección adaptativa de características usando información mutua (MI). Este método permite optimizar el número de clústeres de patrones de consumo y seleccionar las características temporales más influyentes usando algoritmo de agrupamiento difuso con un único clúster. (Kojury-Naftchali et al., 2021b)</p>	<p>realiza utilizando un conjunto de datos real relacionado con 1200 hogares irlandeses anónimos. Estos datos de consumo de los clientes se registran en intervalos de 30 minutos. Datos reales proporcionado por el Archivo Irlandés de Datos de Ciencias Sociales (ISSDA)</p>	<p>asignen muestras de carga eléctrica a cada clúster</p>	<p>intensiva</p> <p>Posibilidad de caer en un óptimo local</p>
	<p><i>Algoritmo</i></p> <ul style="list-style-type: none"> Modelo de Markov oculto (HMM) <p>Modelo de probabilidad de series temporales</p> <p><i>Índices</i></p> <ul style="list-style-type: none"> DBI PAA 	<p>El método utiliza la técnica de aproximación agregada a trozos para reducir las dimensiones de los datos. Además, se propone un algoritmo k-means adaptativo basado en el índice de Davies-Bouldin para agrupar a los usuarios. Finalmente, se establece un modelo de MMarkov oculto que describe las transiciones probabilísticas de diferentes niveles de carga para cada clúster. (Lu et al., 2019)</p>	<p>Se analizaron 200 grandes usuarios en Foshan, provincia china de Guangdong, en 2016 (366 días), y el intervalo de muestreo es de 15 minutos. Se seleccionan los datos de carga de 165 usuarios grandes durante 250 días consecutivos como el conjunto de datos estudiado, y hay un total de 5.892 perfiles semanales.</p>	<p>Sin memoria</p>	<p>Suponiendo que el estado actual solo está relacionado con el estado anterior</p>
Clúster Series temporales	<p><i>Algoritmo</i></p> <ul style="list-style-type: none"> Series temporales <p><i>Reducción de dimensionalidad</i></p> <ul style="list-style-type: none"> PCA SVD <p><i>Índices de evaluación</i></p>	<p>El método combina el análisis de series temporales de consumo eléctrico con el índice de entropía para identificar patrones de comportamiento en los consumidores susceptibles a</p>	<p>Datos reales de 100 consumidores residenciales durante seis meses proporcionados por el Archivo Irlandés de Datos de Ciencias Sociales (ISSDA) provenientes de (AMI). Estos</p>	<p>Identifica patrones con alta precisión mediante datos en tiempo real. Reduce costos operativos para las empresas eléctricas. Proporciona insumos</p>	<p>La implementación de algoritmos de minería de datos puede ser computacionalmente intensiva. La calidad del análisis</p>

Método	Descripción	Aplicación	Casos de estudio	Ventajas	Desventajas
	<ul style="list-style-type: none"> Davies Bouldin (DB) <p>Índice de entropía</p>	<p>programas DR o políticas.(Kojury-Naftchali et al., 2016)</p>	<p>datos se registran con intervalos de 30 minutos. Es decir, por cada cliente, por cada 24 horas, tenemos 48 registros de consumo.</p>	<p>valiosos para la personalización de tarifas y planificación.</p>	<p>depende de la limpieza y representatividad de los datos.</p>
		<p>El método propuesto utiliza datos provenientes de (AMI) para analizar patrones de consumo de electricidad a través de series temporales. Los datos se agrupan en clústeres basados en similitudes de comportamiento utilizando algoritmos de minería de datos y reconocimiento de patrones. (Kojury-Naftchali & Fereidunian, 2019)</p>	<p>Se estudiaron los datos de consumo de 1200 hogares anónimos para un horizonte temporal de 12 meses en la red de distribución irlandesa. Estos datos se registran en intervalos de 30 minutos. Por lo tanto, para cada cliente en cada 24 horas, hay 48 registros. Los datos provienen del Archivo Irlandés de Datos de Ciencias Sociales (ISSDA).</p>	<p>Permite decisiones más precisas para operadores de red. Incrementa la eficiencia energética y reduce picos de demanda. Facilita la adaptación dinámica a cambios en los patrones de consumo</p>	<p>Requiere datos de alta resolución y sistemas de almacenamiento robustos. La variabilidad en los comportamientos de consumo puede afectar la precisión. Dependencia de herramientas avanzadas de minería de datos.</p>
<p>Método Basado Agregación simbólica por segmentos</p>	<p><i>Algoritmo</i></p> <ul style="list-style-type: none"> Perfil de carga típico (TLP) <p><i>Reducción dimensional</i></p> <ul style="list-style-type: none"> PCC SAX <p><i>Índice de evaluación</i></p> <ul style="list-style-type: none"> MSE DBI 	<p>El método propuesto utiliza agregación simbólica por segmentos para agrupar perfiles de carga típicos (TPL).(Xu et al., 2021)</p>	<p>Caso de estudio con medidores inteligentes irlandeses con perfiles de carga diarios de 6332 clientes durante 6 meses con muestras en intervalos de 15 min, se seleccionan los perfiles de carga de todos los clientes el 14 de julio 2009.</p>	<p>Manejo eficaz de perfiles irregulares. Mejora la precisión en la representación de tendencias. Reducción significativa de dimensionalidad.</p>	<p>Complejidad computacional adicional en comparación con métodos tradicionales. Posible sobresegmentación, lo que lleva a la división de perfiles similares en diferentes grupos.</p>
<p>El método AERO</p>	<p><i>Algoritmo</i></p> <ul style="list-style-type: none"> AERO (Activity Extraction with Rational Observation) <p>OT (Operation threads)</p> <p>Subproceso de operación</p> <p>Ontología</p>	<p>El método AERO propone extraer las actividades de los usuarios mediante el análisis de datos de consumo eléctrico proporcionados por medidores inteligentes y la infraestructura AMI. Combina el uso de ontologías, contextos</p>	<p>Caso de estudio real en una casa de muestra en Corea, se enfoca en la creación de tareas a partir de observaciones de actividades, asignándoles metainformación semántica.</p>	<p>Proceso automático que no depende de sensores adicionales. Permite operar en modo local sin compartir datos sensibles con terceros. Se puede aplicar a múltiples contextos más</p>	<p>Dependencia de datos iniciales Limitación en detección de actividades complejas Costos de implementación</p>

Método	Descripción	Aplicación	Casos de estudio	Ventajas	Desventajas
		temporales, y datos de ubicación de electrodomésticos para identificar actividades específicas en el hogar. (Cho et al., 2010)		allá del hogar"	
Métodos de Análisis Visual	<p>Análisis de patrones espaciales y temporales</p> <p><i>Reducción de dimensionalidad</i></p> <ul style="list-style-type: none"> t-SNE <p><i>Herramienta visual</i></p> <ul style="list-style-type: none"> Power Consumption Pattern Explorer (PCP) 	<p>El método propone un marco de análisis visual espacio temporal combinando técnicas de reducción de dimensionalidad (t-SNE). Además, se implementó una herramienta interactiva, (PCP), que permite a los usuarios explorar y comprender patrones de comportamiento energético de manera visual e intuitiva.(Niu et al., 2021)</p>	<p>En este estudio se utilizan datos de consumo de electricidad de resolución diaria del distrito de Pudong en Shanghai, China. Colaboramos con la Shanghai National Grid Company y recopilamos los datos de 2015 a 2018. Los datos comprenden 10.000 clientes de energía domésticos.</p>	<p>Visualización intuitiva</p> <p>Los usuarios pueden explorar patrones en detalle y ajustar el análisis según sus necesidades.</p> <p>Aplicable a diversos contextos geográficos y temporales.</p> <p>Eficiencia energética:</p>	<p>Resultados sensibles a datos incompletos o ruidosos.</p> <p>Mayor precisión requiere datos con alta resolución temporal.</p> <p>Necesidad de familiarizarse con herramientas como PCP.</p>
Método OLAP (Online Analytical Processing)	<p><i>Algoritmo de modelado de datos</i></p> <ul style="list-style-type: none"> OLAP <p><i>Algoritmo de predicción</i></p> <ul style="list-style-type: none"> ARIMA 	<p>El método utiliza técnicas como el modelado de datos OLAP (Online Analytical Processing) y procesamiento distribuido, el sistema organiza la información en un cubo multidimensional que permite analizar consumo por ubicación, tiempo y tipo de dispositivo. Se apoya en algoritmos de predicción como ARIMA para generar datos sintéticos y realizar análisis a gran escala IoT.(Gupta et al., 2020)</p>	<p>Se sintetizan un millón de datos de medidores inteligentes de energía de áreas residenciales para un millón de medidores inteligentes domésticos utilizando un conjunto de datos real de un año. Los datos de los medidores se distribuyeron geográficamente en las siete áreas residenciales de los Emiratos Árabes Unidos en función del último porcentaje de población en cada Emirato.</p>	<p>Escalabilidad para analizar millones de hogares.</p> <p>Respuesta rápida (menos de 15 minutos) para grandes volúmenes de datos.</p> <p>Reducción de costos al usar hardware y software de código abierto.</p> <p>Incrementa la transparencia entre usuarios y proveedores de energía.</p>	<p>Mayor complejidad en la configuración inicial.</p> <p>Dependencia en técnicas de modelado de datos, lo cual puede introducir errores si no se calibra correctamente.</p>
Modelo determinista	<p>Modelo Determinístico</p> <ul style="list-style-type: none"> Modelo de 	<p>Este método permite clasificar a los usuarios residenciales según sus patrones de demanda</p>	<p>Caso de estudio realizado en Colombia, con datos de consumo promedio diario de</p>	<p>Mejora la eficiencia en la gestión de la demanda.</p> <p>Personaliza el servicio</p>	<p>Dependencia de la calidad y disponibilidad de datos.</p> <p>Posible resistencia de</p>

Método	Descripción	Aplicación	Casos de estudio	Ventajas	Desventajas
	regresión polinomial Pruebas de bondad y ajuste <i>Índices</i> <ul style="list-style-type: none"> • SSE • LAR 	de energía eléctrica, ajustándose a sus características socioeconómicas (estrato) y preferencias de consumo. Se basa en herramientas estadísticas como la regresión polinomial y software como MATLAB.(Flórez, 2021)	energía eléctrica de 418 clientes residenciales del OR Celsia S.A E.S.P entre los meses de agosto, septiembre y octubre del año 2019. Estos clientes cuentan con un medidor avanzado instalado en el hogar y están diferenciados por estrato social.	según necesidades específicas. Reduce los costos operativos y emisiones de gases de efecto invernadero. Facilita la integración de generación distribuida y energías renovables	usuarios al cambio tecnológico
Modelo Estadístico	Metodología SAICC <i>(Statistical Assessment for Identifying Changes in Consumption)</i> Histogramas Método SAEC <i>Índices</i> <ul style="list-style-type: none"> • IoC • MCIAT 	La metodología SAICC es una herramienta estadística diseñada para caracterizar el comportamiento de la demanda de electricidad. Se enfoca en identificar patrones estocásticos y evaluar cambios en los perfiles de carga eléctrica, permitiendo detectar anomalías y asociarlas con posibles eventos en el sistema eléctrico.(Serrano, 2020)	La metodología se aplica al edificio 5E de la UPV en Valencia, España. Seguidamente, se estudia el consumo de electricidad de un conjunto de 14 edificios, al analizar los datos del consumo de la UPS Sede Cuenca en Ecuador. Los PCE diarios están conformados por datos de potencia activa media en intervalos por lo general de 1 hora o cada 15 minutos.	Requiere menos recursos que métodos basados en inteligencia artificial. Los resultados estadísticos son más fácilmente comprensibles y aplicables. Puede aplicarse a diferentes niveles de la red eléctrica y tipos de usuarios.	Dependencia de la calidad de los datos Limitaciones en escenarios dinámicos Requiere conocimientos estadísticos avanzados para su implementación y análisis.

6.6. Clasificación de los métodos seleccionados

Con el fin de realizar una adecuada clasificación de los métodos seleccionados se tuvieron en cuenta las siguientes características:

- **Datos:** es una de las variables más importantes en este análisis, es necesario entender si el método es capaz de manejar, almacenar y procesar grandes volúmenes de datos.
- **Complejidad del software o algoritmo:** se evalúa la necesidad de realizar procesos de alta complejidad como redes neuronales, aplicaciones de IA o si se maneja mediante métodos convencionales como las series de tiempo, los modelos determinísticos y estadísticos.
- **Precisión:** Esta variable es importante al momento de saber qué factores influyen en realizar una caracterización adecuada, es decir, se tiene en cuenta su tolerancia al ruido, su dependencia de valores iniciales y su adaptabilidad a factores externos que puedan alterar la medición.
- **Casos de estudio:** Se identifican los casos de estudios realizados a nivel internacional y nacional dado que las condiciones para ambos ambientes son diferentes en cuanto a infraestructura, tecnología, recursos, variables geográficas y políticas socioeconómicas.
- **Costos:** Esta variable relaciona los costos asociados a licencias y/o suscripciones de software, herramientas o gestores de bases de datos y almacenamiento en la nube. Además, de los costos asociados a la adquisición de hardware como computadoras capaces de procesar grandes volúmenes de datos.
- **Escalabilidad:** Se tiene en cuenta su capacidad de crecimiento, adaptabilidad e integración con nuevas técnicas que permitan desarrollar futuras aplicaciones como programas de respuesta a la demanda, eficiencia energética, previsión de carga, pérdidas de energía, ciberseguridad, entre otros.

Se presentan las abreviaciones a considerar para poder comprender la siguiente clasificación: Bajo (B), Medio (M), Alto(A).En la tabla 5, se presentan la clasificación de los métodos de caracterización de usuarios.

Tabla 5. Clasificación de métodos de caracterización de usuarios

Métodos	Algoritmo	Criterios de clasificación															
		Datos			Complejidad software			Precisión			Casos de estudio		Costos			Escalabilidad	
		B	M	A	B	M	A	B	M	A	Internacional	Nacional	B	M	A	No	Si
Basados en particiones	K-Means			x		x				x	x	x		x			x
	K-Medoids			x		x			x		x			x			x
	K-Means Adaptativo			x		x			x		x			x			x
	FCM			x			x		x		x				x		x
	K-Shape		x			x			x		x			x		x	
	K-Modes			x			x	x			x			x		x	
Jerárquicos	Aglomerativo			x			x			x	x				x		x
Basado en la densidad	DBSCAN		x			x		x			x			x			x
Basado en cuadrículas	SOM			x			x			x	x				x		x
	Wavecluster			x			x			x	x				x		x
Basado en modelos	GMM			x		x			x		x				x		x
	HMM			x		x			x		x				x		x
Series temporales	PCA-SVD			x		x				x	x			x			x
Agregación simbólica	TPL			x			x		x		x				x		x
AERO	Ontología	x					x		x		x					x	x
Análisis Visual	PCP			x	x			x			x			x			x
OLAP	ARIMA			x		x			x		x			x			x
Modelos determinísticos	Regresión polinomial			x	x				x				x	x			x
Métodos Estadísticos	Metodología SAICC		x		x				x		x			x			x

6.7. Análisis de resultados

La Tabla 5 permite identificar cuales métodos de caracterización de usuarios se pueden adaptar a las necesidades particulares de estudio teniendo en cuenta variables como cantidad de datos, complejidad computacional, precisión del método, casos de estudio, costos y escalabilidad. Además, permite establecer tendencias tales como:

- La mayoría de los métodos son implementados internacionalmente donde se presenta un mayor despliegue de AMI
- Los costos más altos se presentan con los métodos de clústeres que implementa redes neuronales, lógica difusa y computación en la nube, que a su vez requieren mayor complejidad computacional para el manejo de grandes volúmenes de datos, pero no se relacionan directamente con la precisión de los mismos y su capacidad para escalar a otras aplicaciones.

Luego de la clasificación de métodos de caracterización de usuarios presentados en la Tabla 5, se obtienen las siguientes observaciones:

Los métodos de clústeres o agrupamiento basados en particiones se destacan los que implementan algoritmos *K-means* y Fuzzy C-Means (FCM). Los algoritmos *K-means* no presentan una alta complejidad en su implementación, aunque su precisión depende de la calidad de los datos de entrada y la elección del número de clústeres óptimos. Además, estos algoritmos son usados para aplicaciones de respuesta a la demanda, pronóstico de carga, eficiencia energética entre otros, integrándose con diferentes lenguajes de programación como R, redes neuronales y nuevas técnicas de análisis de datos. Finalmente, son los más usados tanto internacionalmente como nacionalmente para aplicaciones de AMI sin representar un costo significativo dado que reduce la dimensionalidad de los datos. Un ejemplo de esto son los casos de estudio implementados por Enel (Cely Callejas, 2018). En cuanto a los Fuzzy C-

means, son más complejos de implementar y requerirán una alta complejidad computacional reflejándose en el incremento de recursos y costos.

Entre los otros métodos de clústeres o agrupamiento, de igual modo se destacan los basados en cuadrícula con los algoritmos SOM y Wavelet, aunque presentan una alta complejidad en su implementación y costos son ampliamente usados en casos de estudios internacionales dado su adaptabilidad y respuesta a los grandes volúmenes de datos, estos presentan desafíos técnicos económicos. Por otro lado, también se encuentran los basados en modelos como los GMM y HMM, respectivamente usan métodos probabilísticos como Gaussianos y cadenas de Markov, los cuales requieren menores recursos computacionales y son capaces de manejar Big Data.

Los métodos convencionales como las series de tiempo, los modelos determinísticos y estadísticos ofrecen la ventaja de contar con información fácilmente disponible y de sencilla comprensión. Además, no requieren técnicas complejas, lo que contribuye a reducir tanto los costos operativos como los de implementación. Sumado a esto, son adaptados para la medición avanzada, un ejemplo es el modelo determinístico implementado en Celsia para la representación de grandes datos de consumo de energía en programas de respuesta a la demanda (Flórez Duarte, 2021).

Seguidamente, se analizan los métodos novedosos dentro de los cuales se destaca el método OLAP (*Online Analytical Processing*) por su reducción de costos al usar métodos de código abierto e integrar métodos de pronósticos de carga como ARIMA, ser escalable a grandes volúmenes de datos y sus posibles aplicaciones.

Finalmente, en el contexto regional se plantea un escenario de despliegue de AMI progresivo, empezando con pilotos de prueba en áreas de influencia del OR ESSA y proyectando su escalabilidad masiva para el año 2030. Teniendo en cuenta lo mencionado

anteriormente y la Tabla 5, se tienen como referencia dos posibles métodos para iniciar la implementación, los cuales destacan con los mejores desempeños de relación costo/beneficio, facilidad de implementación, precisión en resultados, casos de estudios en pilotos AMI en Colombia y escalabilidad a diferentes aplicaciones: En primer lugar, el modelo determinístico implementado en Celsia para programas de respuesta a la demanda, el cual cumple con todos los parámetros exceptuando su implementación a nivel internacional. En segundo lugar, se encuentra el método basado en particiones empleado por Enel, basándose en algoritmos K-means y redes neuronales mediante software gratis y de código abierto como R. Este caso, cumple con todas las características de clasificación siendo más difícil de implementar, pero con casos de referencia a nivel nacional e internacional y una gran escalabilidad a la gestión de grandes volúmenes de datos.

6.8. Retos y oportunidades del análisis de datos en AMI

Teniendo en cuenta la literatura revisada, se puede concluir que algunos de los mayores retos que implica el análisis de datos de AMI para la caracterización de usuarios y otras aplicaciones son:

- **Big data:** La escala y complejidad de los datos generados por la red eléctrica han aumentado, lo que genera un escenario retador en términos de Big data, debido a que su recopilación frecuente, análisis, almacenamiento y transmisión deben considerarse cuidadosamente (Si et al., 2021). En aplicaciones prácticas, puede haber millones de clientes para analizar para poder realizar perfiles de consumo. El reto se encuentra en la capacidad de captación de data de calidad, capacidad de análisis y tratamiento de grandes cantidades de datos mediante técnicas eficientes y precisas, permitiendo arrojar resultados y transmisión de la información en tiempos adecuados para la buena toma

de decisiones, teniendo en cuenta la seguridad y privacidad de la información (Wang et al., 2015).

- Nuevas tecnologías: El análisis de datos de medidores avanzados es un campo interdisciplinario que involucra ingeniería eléctrica e informática, particularmente el aprendizaje automático. El desarrollo del aprendizaje automático ha tenido un gran impacto en el análisis de datos de los medidores avanzados. La aplicación de nuevas tecnologías de aprendizaje automático es un aspecto importante de la analítica de medidores avanzados (Wang et al., 2019). Además, desarrollos como las API de Spark proporcionan algoritmos de aprendizaje automático (clasificación y regresión, filtrado colaborativo, agrupación, reducción de dimensionalidad, análisis de componentes principales, etc.) para el procesamiento distribuido, pero también soporte para Scala, Java, R y Python para implementar algoritmos personalizados (Oprea et al., 2019). De acuerdo con lo descrito anteriormente, estas tecnologías avanzan de manera vertiginosa requiriendo profesionales especializados en estas áreas específicas y en constante actualización. Además, de la adquisición de software y hardware específico que está impactando los costos de su implementación.
- Seguridad y privacidad: El acceso a todos los datos de los medidores avanzados puede generar problemas de privacidad tanto para los consumidores de energía eléctrica como para los proveedores del servicio de energía. En el caso de la seguridad de datos, se han estudiado ampliamente los trabajos sobre seguridad cibernética física (CPS) en la red inteligente, sin embargo, se deben estudiar más a fondo los diferentes tipos de ataques cibernéticos a los datos de consumo de electricidad (Wang et al., 2019). Adicionalmente, es esencial invertir, ahondar e investigar sobre la seguridad de la información, así como sistemas eficientes de almacenamiento y transmisión de datos y

su capacidad computacional requerida (Si et al., 2021). De esta manera, se podrán establecer sistemas de seguridad de la información y controles de riesgos en seguridad informática.

Adicionalmente, la obtención de datos provenientes de AMI abre la puerta a grandes oportunidades como las presentadas a continuación:

- **Caracterización precisa de usuarios:** Los datos de AMI permiten obtener una visión precisa del comportamiento de consumo energético de los usuarios. A través del análisis de patrones de consumo, es posible identificar segmentos de usuarios, sus necesidades específicas, y predecir su comportamiento a futuro. Esto facilita la personalización de los servicios ofrecidos, como tarifas dinámicas o recomendaciones para mejorar la eficiencia energética (Quilumba et al., 2014).
- **Optimización de la gestión energética:** La información en tiempo real generada por los sistemas AMI abre la puerta para optimizar la distribución y consumo de energía eléctrica. Al analizar estos datos, las empresas pueden predecir mejor la demanda, reducir las pérdidas de energía y mejorar la respuesta ante picos de consumo. Esto se traduce en una gestión más eficiente de los recursos, lo que resulta en beneficios tanto económicos como medioambientales (Si et al., 2021).
- **Nuevos modelos de negocios:** Una mayor desregulación de los mercados minoristas, la caracterización y segmentación de usuarios con datos AMI, la integración de la energía renovable distribuida y el progreso de las tecnologías de la información acelerarán varios modelos de negocio en el lado de la demanda (Wang et al., 2019).

6.9. Aplicaciones y tendencias basadas en datos AMI

A continuación, se describen las principales aplicaciones y tendencias basadas en datos AMI que se están investigando e implementando a nivel mundial.

6.9.1. Eficiencia energética

Los sistemas de medición requieren el uso de datos de modelos, datos históricos de series temporales de carga y datos de sistemas de generación fotovoltaica. Sin embargo, los datos de carga suelen estar disponibles en resoluciones temporales bajas (15, 30 o 60 minutos), lo que puede ser insuficiente para simulaciones que requieren alta resolución de tiempo. Gracias a tecnologías recientes como la infraestructura de medición avanzada (AMI), sistemas de control de edificios avanzados, IoT y redes de sensores, ahora es posible obtener datos de alta resolución casi en tiempo real sobre consumo energético, comportamiento de usuarios y rendimiento de edificios. Esto permite mejorar la eficiencia energética, optimizar servicios y activos de los edificios en el dominio de la red avanzada y acercar a los edificios a convertirse en edificios energéticamente eficientes e interactivos con la red proporcionando un análisis más detallado del impacto de la demanda eléctrica en el sistema de distribución, por lo que se recomienda ahondar en los sistemas de análisis y medición avanzados, así como los diferentes recursos y capacidad computacional requerida para su funcionamiento (Parker et al., 2021).

6.9.2. Pérdidas de energía

A nivel mundial, el robo de energía y fraude causan gastos de servicios públicos por alrededor de \$96 miles de millones de dolares en pérdidas no técnicas al año. Estas pérdidas pueden tener consecuencias graves, como subir el nivel del precio que deben pagar

los clientes y evitar que los proveedores de servicios públicos obtengan los recursos necesarios para futuras inversiones de capital (Si et al., 2021). Por lo anterior, una de las grandes líneas de investigación que requieren de atención constante es la detección de anomalías en la red eléctrica, lecturas falsas, fraudes y robo de energía, que generalmente se realiza mediante clasificadores de detección de robos para detectar datos desequilibrados y maliciosos activando alarmas después de un cierto número de detecciones anormales. Sin embargo, también se requiere identificar los tipos de robo de energía representando un amplio campo en el cual profundizar (Wang et al., 2015).

6.9.3. Pronóstico, previsión y respuesta a la demanda

La estimación del perfil de carga es un ejercicio conveniente, ya que las formas de carga, así como la carga pico diaria, son factores vitales para planificar la producción y la fijación de precios de la electricidad. En el corto plazo, es esencial saber con la mayor precisión posible cuál será la demanda total y local del sistema en los próximos minutos, horas y días, de modo que se puedan programar generadores con diversas restricciones para optimizar la eficiencia total. Los programas tradicionales de previsión de carga se basan en datos de consumo de energía agregados a nivel de sistema (corporativo) con poca o ninguna información sobre los perfiles de consumo de energía de diferentes clases de clientes. Con la implementación de la Infraestructura de medición avanzada (AMI), se introduce una nueva perspectiva sobre la forma en que se realiza la medición de carga. Por lo tanto, se recomienda investigar acerca de métodos, técnicas y metodologías con capacidad de procesar y analizar datos de medidores avanzados anuales para predecir y poder hacer estimaciones de patrones de consumo (Quilumba et al., 2014).

6.9.4. Segmentación de usuarios

El agrupamiento tiene una larga y rica historia. Se han desarrollado muchos algoritmos de agrupamiento diferentes a lo largo del tiempo. Sin embargo, no hay una respuesta única sobre cuál es el mejor, ya que depende en gran medida de la naturaleza del conjunto de datos y de lo que constituye clústeres significativos en una aplicación para la segmentación de usuarios (Quilumba et al., 2014). Por lo anterior, una posible oportunidad de investigación y estudio es la comparación, caracterización, clasificación y aplicación de los diferentes clústeres encontrados en la literatura a conjunto específico de datos para lograr conclusiones de posibles usos, ventajas, desventajas y alcance de cada uno.

6.9.5. Internet de energía e integración con energías renovables

Yi Wang (2015) propone en el documento “Perfiles de carga y su aplicación a la respuesta a la demanda: una revisión” el concepto de Internet de la energía en el que se integran la energía renovable con las tecnologías de internet, creando nuevos desafíos para la elaboración de perfiles de carga de los clientes, ya que pueden utilizar la electricidad generada por su propia energía renovable o almacenamiento, así como la integración de las redes eléctricas, las redes de tráfico, las redes de gas natural y las redes de suministro de calor, lo que cambia y dificulta medir los patrones tradicionales de consumo de energía eléctrica de los clientes. En este sentido, Yi Wang indica que la minería del comportamiento de los clientes ya no se limita al consumo eléctrico, sino que debe incluir un análisis exhaustivo del consumo energético en lugar del consumo eléctrico.

El desarrollo de las energías renovables, las plantas de generación fotovoltaica distribuidas que se pueden instalar en los hogares se están convirtiendo en un punto crítico para el consumo de energía eléctrica residencial ya que convierten la energía solar en energía para la autogeneración y el uso, y el resto se inyecta a la red eléctrica. Esta fluctuación aleatoria y la intermitencia de la energía renovable transformó los nodos de carga tradicionales en nodos de carga generalizados, por lo tanto, la recomendación es investigar y realizar estudios de agrupamiento de carga para estos últimos (Si et al., 2021).

El consumidor con sistema fotovoltaico en su tejado se convierte en prosumidor y puede comerciar con electricidad con otros prosumidores. Siendo que las aplicaciones existentes de los análisis de datos de medidores avanzados se estudian principalmente desde la perspectiva de consumidores individuales y minoristas (Wang et al., 2019), por lo que otra recomendación sería investigar y hacer estudios acerca de cómo y cuantos datos de los medidores avanzados se deben analizar en el micro mercado eléctrico para integrar adecuadamente el consumo de energía eléctrica con las energías renovables.

6.9.6. *Inteligencia artificial*

En los últimos años, además de la agrupación, varias tecnologías de IA como la lógica difusa, el aprendizaje profundo y el aprendizaje adversario han logrado avances continuos en la resolución de datos, el aprendizaje y la capacidad computacional, que tendrán amplias perspectivas de aplicación en la investigación de la agrupación de carga eléctrica, por lo que se recomienda realizar investigación y estudios aplicando la percepción inteligente en tiempo real combinada con el estado físico, modelos basados en datos combinados con simulación y toma de decisiones asistida combinada con control de operación en datos de carga y consumo de los usuarios (Si et al., 2021).

7. Conclusiones

Los métodos de caracterización de usuarios están siendo investigados en diferentes partes del mundo como USA, China, Australia, Irán, España, Chile, Colombia entre otros. En esta monografía se identificaron 142 potenciales investigaciones relacionadas con AMI y la caracterización de usuarios; de las cuales 33 aportaban de manera implícita y explícita a esta revisión, mostrando la importancia de entender, clasificar, segmentar y caracterizar el comportamiento de los consumos de los clientes en las futuras líneas de investigación como programas de respuesta a la demanda, eficiencia energética, pérdidas de energía, pronóstico de carga entre otros.

De los métodos más implementados se encuentran los clústeres o agrupamientos basados en particiones que tienen como base los algoritmos de minería de datos como k-means; sobre estos existen varias investigaciones, aplicaciones y casos de estudios como los realizados en USA a 819 usuarios para programas de respuesta a la demanda. Además, de los aplicados en Colombia por Enel para 73 mil usuarios representando 157 GB teniendo en cuenta diferentes variables como geografía y estratos socioeconómicos. Dentro de sus principales características se encuentra su fácil ejecución y flexibilidad a la hora de integrarse con inteligencia artificial, nuevas técnicas de análisis de datos y software como R. Aun así, sigue presentando retos como su dependencia a los datos de entrada y el número óptimo de clústeres.

En la actualidad, las metodologías más convencionales incluyen las series temporales, debido a su facilidad para interpretar los datos de consumo; los métodos determinísticos, como la regresión polinomial, utilizados por Celsia para integrar programas de respuesta a la demanda; y los métodos estadísticos presentados en España para caracterizar la demanda mediante patrones estocásticos. Estos métodos son conocidos, su implementación es menos compleja y tienen un potencial para integrarse con métodos de optimización mediante IA y

análisis de datos. Por otro lado, también se encuentran técnicas como: redes neuronales, minería de datos y computación en la nube mediante diferentes algoritmos como los de lógica difusa, mapas auto-organizados, wavecluster, jerárquicos entre otros. Estos presentan un gran potencial en la implementación de grandes volúmenes de datos, pero son más difíciles de entender, ejecutar y presentan una alta complejidad computacional.

Se identificaron los retos y oportunidades del análisis de datos en AMI para la caracterización de usuarios y otras aplicaciones. Dentro de los retos, se destacan el Big Data en cuanto al volumen, variedad y velocidad de procesamiento en su ejecución, aunque paralelamente se han desarrollado nuevas herramientas para gestionar el flujo de datos como las API de Spark. Además, de los retos de privacidad y seguridad en transmisión, recopilación y tratamiento de la información de los usuarios. A esto se agrega tener personal capacitado en análisis de datos, programación y ciberseguridad. Por otro lado, se presentan oportunidades como nuevos mercados minoristas de energía eléctrica, programas de respuesta a la demanda siendo los usuarios prosumidores, programas de eficiencia energética, previsión de carga a corto plazo, integración de fuentes de energía renovables, vehículos eléctricos, almacenamiento de energía y la digitalización de la energía.

Finalmente, la caracterización de usuarios mediante datos AMI presenta un gran potencial de investigación dado que la caracterización de usuarios depende de factores externos como el clima, la geografía, los estratos, los datos, las políticas de cada país, la economía y es particular a la aplicación que se esté orientando. Como trabajo futuro se propone implementar alguno de los métodos de caracterización de usuarios presentados en esta monografía, en un proyecto piloto con datos AMI reales que se adapte a las aplicaciones y necesidades particulares del OR ESSA.

8. Referencias Bibliográficas

- ADMINISTRADOR DEL MERCADO ELÉCTRICO (XM). (2024, January 31). *DEMANDA DE ENERGÍA EN COLOMBIA 2023*. <https://www.xm.com.co/noticias/6584-en-2023-la-demanda-de-energia-en-colombia-aumento-un-445-en-comparacion-con-el-ano>
- Balachander, K., & Paulraj, D. (2021). Building energy time series data mining for behavior analytics and forecasting energy consumption. *KSII Transactions on Internet and Information Systems*, 15(6), 1957–1980. <https://doi.org/10.3837/tiis.2021.06.001>
- Bedingfield, S., Alahakoon, D., Genegedera, H., & Chilamkurti, N. (2018). Multi-granular electricity consumer load profiling for smart homes using a scalable big data algorithm. *Sustainable Cities and Society*, 40, 611–624. <https://doi.org/10.1016/j.scs.2018.04.006>
- Cappers, P., Goldman, C., & Kathan, D. (2010). Demand response in U.S. electricity markets: Empirical evidence. *Energy*, 35(4), 1526–1535. <https://doi.org/10.1016/j.energy.2009.06.029>
- Carlos, S. A. E. S. P., & Devia Tabares, A. (2017). *Caracterización de los usuarios del sector residencial como herramienta estratégica para la gestión comercial de la empresa de energía de Pereira S.A. E.S.P.* EAFIT.
- Cely Callejas, J. David. (2018). A Clustering Approach for Domestic Smart Metering Data Preprocessing. *2018 IEEE ANDESCON : Conference Proceedings : 22nd Aug - 24th Aug 2018, Cali, Colombia*.
- Çetinkaya, Ü., Avcı, E., & Bayindir, R. (2020). Time Series Clustering Analysis of Energy Consumption Data. *2020 9th International Conference on Renewable Energy Research and Application (ICRERA)*.
- Chelmis, C., Kolte, J., & Prasanna, V. (2015). Big Data Analytics for Demand Response: Clustering Over Space and Time. *2015 IEEE International Conference on Big Data : Oct 29-Nov 01, 2015, Santa Clara, CA, USA*, 2959.
- Cho, H. S., Yamazaki, T., & Hahn, M. (2010). AERO: Extraction of User's Activities from Electric Power Consumption Data. *IEEE*.
- COMISIÓN DE REGULACIÓN DE ENERGÍA Y GAS. (2020). *Resolución CREG 131 de 2020. Por la cual se ordena hacer público el proyecto de resolución: Por la cual se establecen las condiciones para la implementación de la infraestructura de medición avanzada en el SIN.*

[https://gestornormativo.creg.gov.co/Publicac.nsf/1c09d18d2d5ffb5b05256eee00709c02/33d0b0fa08574678052585a50075e521/\\$FILE/Creg131-2020.pdf](https://gestornormativo.creg.gov.co/Publicac.nsf/1c09d18d2d5ffb5b05256eee00709c02/33d0b0fa08574678052585a50075e521/$FILE/Creg131-2020.pdf)

COMISIÓN DE REGULACIÓN DE ENERGÍA Y GAS. (2022). *Resolución CREG 101 001 de 2022. Por la cual se establecen las condiciones para la implementación de la infraestructura de medición avanzada en el SIN.*
[https://gestornormativo.creg.gov.co/Publicac.nsf/1c09d18d2d5ffb5b05256eee00709c02/5942c864340e83fb052587fb0080940b/\\$FILE/Creg101%20001.pdf](https://gestornormativo.creg.gov.co/Publicac.nsf/1c09d18d2d5ffb5b05256eee00709c02/5942c864340e83fb052587fb0080940b/$FILE/Creg101%20001.pdf)

Doukas, H. (2022). ENERGY TRANSITIONS, INTELLIGENCE AND BIG DATA: Towards a prosumer concept with energy autonomy. *13th International Conference on Information, Intelligence, Systems and Applications, IISA 2022.*
<https://doi.org/10.1109/IISA56318.2022.9904415>

Duarte, J. E., Rosero-Garcia, J., & Duarte, O. (2024). Analysis of Variability in Electric Power Consumption: A Methodology for Setting Time-Differentiated Tariffs. *Energies, 17(4).*
<https://doi.org/10.3390/en17040842>

El Congreso de la República. (2021). *Ley 2099 de 2021. Por medio de la cual se dictan disposiciones para transición energética, la dinamización del mercado energético, la reactivación económica del país y se dictan otras disposiciones.10 de julio de 2021.*
Función Pública Gestor Normativo.
<https://www.funcionpublica.gov.co/eva/gestornormativo/norma.php?i=166326>

Elakshum, S., & Ponraj, A. (2017). A server based load analysis of smart meter systems. *2017 IEEE International Conference on Nextgen Electronic Technologies: Silicon to Software : ICNETS2 2017 : 23-25 March 2017, VIT University, Chennai Campus, India.*

ELECTRIFICADORA DE SANTANDER S.A. E.S.P. (2017). *NORMA TÉCNICA DE MEDICIÓN CENTRALIZADA E INFRAESTRUCTURA DE MEDICIÓN AVANZADA-AMI.*

ELECTRIFICADORA DE SANTANDER S.A E.S.P. (2024). *Plan Empresarial ESSA 2024-2027.*

ELECTRIFICADORA DE SANTANDER S.A E.S.P - (ESSA). (2023). *Las conexiones ilegales afectan la calidad del servicio de la región.*
<https://www.essa.com.co/site/proveedores/detalle-articulo/las-conexiones-ilegales-afectan-la-calidad-del-servicio-de-energia-en-la->

- Big Data Management. *SGC : 2017 Smart Grid Conference : 20-21 December 2017, Tehran, Iran*, 382.
- Kojury-Naftchali, M., Fereidunian, A., Savaghebi, M., & Akhbari, B. (2021a). Change Detection in Electricity Consumption Patterns Utilizing Adaptive Information Theoretic Algorithms. *IEEE Systems Journal*, 15(2), 2369–2377. <https://doi.org/10.1109/JSYST.2020.3011313>
- Kojury-Naftchali, M., Fereidunian, A., Savaghebi, M., & Akhbari, B. (2021b). Change Detection in Electricity Consumption Patterns Utilizing Adaptive Information Theoretic Algorithms. *IEEE Systems Journal*, 15(2), 2369–2377. <https://doi.org/10.1109/JSYST.2020.3011313>
- Li, J., Kong, L., Cheng, C., & Dong, J. (2022). Research on key technologies of net metering for smart grid AMI and data mining of user electricity consumption. *IEEE Proceedings - 2022 International Conference on Electronics and Devices, Computational Science, ICEDCS 2022*, 95–99. <https://doi.org/10.1109/ICEDCS57360.2022.00028>
- Lijuan, Z. (2023). Research on Algorithm and Modification of Elman Network Based on Fuzzy Classification and Short-Term Electricity Consumption Prediction. *2023 IEEE International Conference on Electrical, Automation and Computer Engineering, ICEACE 2023*, 523–526. <https://doi.org/10.1109/ICEACE60673.2023.10442689>
- Lin, R., Wu, B., & Su, Y. (2018). An adaptive weighted pearson similarity measurement method for load curve clustering. *Energies*, 11(9). <https://doi.org/10.3390/en11092466>
- Lu, S., Lin, G., Liu, H., Ye, C., Que, H., & Ding, Y. (2019). A Weekly Load Data Mining Approach Based on Hidden Markov Model. *IEEE Access*, 7, 34609–34619. <https://doi.org/10.1109/ACCESS.2019.2901197>
- Ma, A. (2023). Addressing Big Data Issues in Load Shape Profiling for Utility Customers. *Proceedings - 2023 International Conference on Artificial Intelligence and Power Engineering, AIPE 2023*, 31–41. <https://doi.org/10.1109/AIPE58786.2023.00014>
- Mansoor, H., Ali, S., Khan, I. U., Arshad, N., Khan, M. A., & Faizullah, S. (2023). Short-Term Load Forecasting Using AMI Data. *IEEE Internet of Things Journal*, 10(24), 22040–22050. <https://doi.org/10.1109/JIOT.2023.3295617>
- Marrero, L., Carrizo, D., García, L., & Ulloa, F. (2021). Uso de algoritmo K-means para clasificar perfiles de clientes con datos de medidores inteligentes de consumo eléctrico: Un caso de estudio. *Revista Chilena de Ingeniería*, 29(4), 778–787.

- MINISTERIO DE MINAS Y ENERGÍAS. (2019). *Resolución MME 40483 de 2019. Por la cual se modifica la Resolución número 40072 de 2018.* https://gestornormativo.creg.gov.co/gestor/entorno/docs/resolucion_minminas_40483_2019.htm
- MINISTERIO DE MINAS Y ENERGÍAS MME. (2021). *LINEAMIENTOS DE POLÍTICA DE RECURSOS ENERGÉTICOS DISTRIBUIDOS Y ARENERAS REGULATORIAS.*
- Mononen, M., Saarenpaa, J., Johansson, M., & Niska, H. (2014). Data-driven Method for Providing Feedback to Households on Electricity Consumption. *ISSNIP : 2014 IEEE Ninth International Conference on Intelligent Sensors, Sensor Networks and Information Processing : 21-24 April 2014, Singapore*, 441.
- Niu, Z., Wu, J., Liu, X., Huang, L., & Nielsen, P. S. (2021). Understanding energy demand behaviors through spatio-temporal smart meter data analysis. *Energy*, 226. <https://doi.org/10.1016/j.energy.2021.120493>
- Oprea, S.-V., Bâra, A., & Diaconita, V. (2019). Big Data solutions - data ingestion and stream processing for demand response management. *2019 23rd International Conference on System Theory, Control and Computing (ICSTCC)*.
- Ordoñez, G. O., Duarte, C. A., Ramon, J. H., Cadena, R. H., Jaimes, W. Y., Moreno, J. F., & Carrillo, D. C. (2017). *Sistema de medición centralizada en redes de distribución de baja tensión para la reducción de pérdidas eléctricas. Universidad Industrial de Santander (UIS) y Electrificadora de Santander (ESSA S.A E.SP).*
- Oyedokun, J., Bu, S., & Xiao, Y. (2018). Smart Meter Data Characterization and Clustering for Peak Demand Targeting in Smart Grids. *2018 IEEE PES Innovative Smart Grid Technologies Conference Europe (ISGT-Europe) : Sarajevo, Bosnia and Herzegovina, October 21-25, 2018 : Conference Proceedings*.
- Parker, A., Moayedi, S., James, K., Peng, D., & Alahmad, M. A. (2021). A Case Study to Quantify Variability in Building Load Profiles. *IEEE Access*, 9, 127799–127813. <https://doi.org/10.1109/ACCESS.2021.3112103>
- Qi, C., Liu, Y., Yang, X., & Zhong, J. (2023). SAEBKA: Similarity Analysis of Electricity Behavior based on K-shape Algorithm in AMI System. *2023 3rd International Conference on Electrical Engineering and Mechatronics Technology, ICEEMT 2023*, 420–424. <https://doi.org/10.1109/ICEEMT59522.2023.10262908>

- Qiu, W., Zhai, F., Bao, Z., Li, B., Yang, Q., & Cao, Y. (2016a). *Clustering Approach and Characteristic Indices for Load Profiles of Customers Using Data from AMI* .
- Qiu, W., Zhai, F., Bao, Z., Li, B., Yang, Q., & Cao, Y. (2016b). Clustering Approach and Characteristic Indices for Load Profiles of Customers Using Data from AMI . *IEEE*, 10–13.
- Quilumba, F., Jen Lee, W., & Huan, H. (2014). An overview of AMI data preprocessing to enhance the performance of load forecasting. *2014 IEEE Industry Application Society Annual Meeting*.
- Serrano, X. (2020). *Caracterización de la demanda de energía mediante patrones estocásticos en las Redes Eléctricas Inteligentes*.
<https://riunet.upv.es/bitstream/handle/10251/153810/Serrano%20-%20Caracterizaci%C3%B3n%20de%20la%20demanda%20de%20energ%C3%ADa%20mediante%20patrones%20estoc%C3%A1sticos%20en%20las%20Redes%20El...pdf?squence=1>
- Shi, Y., Li, M., Zhang, Y., & Zhang, Y. (2022). Clustering Analysis of Power Grid Data Based on Dynamic Hierarchical K-Modes. *Lecture Notes in Electrical Engineering*, 827 LNEE, 1608–1614. https://doi.org/10.1007/978-981-16-8052-6_235
- Si, C., Xu, S., Wan, C., Chen, D., Cui, W., & Zhao, J. (2021). Electric Load Clustering in Smart Grid: Methodologies, Applications, and Future Trends. *Journal of Modern Power Systems and Clean Energy*, 9(2), 237–252. <https://doi.org/10.35833/MPCE.2020.000472>
- UNIDAD DE PLANEACIÓN MINERO ENERGÉTICA (UPME). (2023, December). *PROYECCIÓN DE LA DEMANDA DE ENERGÍA ELÉCTRICA Y POTENCIA MÁXIMA 2023-2037*.
<https://www1.upme.gov.co/DemandayEficiencia#:~:text=Los%20resultados%20a%20mediano%20plazo,a%C3%B1o%20del%203%2C1%25>.
- Wang, Y., Chen, Q., Hong, T., & Kang, C. (2019). Review of Smart Meter Data Analytics: Applications, Methodologies, and Challenges. *IEEE Transactions on Smart Grid*, 10(3), 3125–3148. <https://doi.org/10.1109/TSG.2018.2818167>
- Wang, Y., Chen, Q., Kang, C., Zhang, M., Wang, K., & Zhao, Y. (2015). Load Profiling and Its Application to Demand Response: A Review. *IEEE*, 20(2).

- Wang, Y., Ma, J., Gao, N., Wen, Q., Sun, L., & Guo, H. (2023). Federated fuzzy k-means for privacy-preserving behavior analysis in smart grids. *Applied Energy*, 331. <https://doi.org/10.1016/j.apenergy.2022.120396>
- Xiao, Y., Yang, J., Que, H., Li, M., & Gao, Q. (2014). Application of Wavelet-based clustering approach to load profiling on AMI measurements. *2014 China International Conference on Electricity Distribution (CICED 2014)*.
- Xu, L., Zhang, Y., & Shao, Z. (2021). An Approach to Cluster Electrical Load Profiles Based on Piecewise Symbolic Aggregation. *Proceedings - 2021 6th Asia Conference on Power and Electrical Engineering, ACPEE 2021*, 1000–1004. <https://doi.org/10.1109/ACPEE51499.2021.9437105>
- Zakaria, B. M., Eldien, A. S. T., & Elkholy, M. S. (2020, December 15). Development of Smart Grid System. *Proceedings of ICCES 2020 - 2020 15th International Conference on Computer Engineering and Systems*. <https://doi.org/10.1109/ICCES51560.2020.9334620>
- Zhou, D., Balandat, M., & Tomlin, C. (2016). Residential Demand Response Targeting Using Machine Learning with Observational Data. *2016 IEEE 55th Conference on Decision and Control (CDC)*.