Análisis de modelos de aprendizaje automático para la segmentación y predicción de la demanda de clientes no regulados en el mercado eléctrico colombiano

Johan Sebastián Barrera Devia

Jonathan Carvajal García

Asesor

Felipe Alexander Pipicano Guzmán

Universidad Nacional Abierta y a Distancia UNAD

Escuela de Ciencias Básicas Tecnologías e Ingeniería ECBTI

Especialización en Ciencia de Datos y Analítica

2024

Dedicatoria

Dedicatoria, Jonathan Carvajal García: A la memoria de mi madre, al apoyo de mi padre, esposa y familia.

Agradecimientos

Queremos extender un especial agradecimiento al Ph. D. Felipe Alexander Pipicano Guzmán, asesor de la monografía, quien nos guio de la mejor forma y permitió llevar a final término esta investigación.

También queremos hacer extensivo nuestro agradecimiento al Ph. D. Rafael Gaitán Ospina, jurado de la monografía quien con sus recomendaciones no permitió fortalecer la investigación.

Resumen

La cadena de valor del sector energético colombiano incluye cuatro actividades principales: generación, transmisión, distribución y comercialización, y en cada uno se realizan procesos de pronóstico de la demanda. Esta investigación se centra en la comercialización, específicamente en los usuarios no regulados, quienes pactan el costo de la energía con comercializadoras y deben contar con un sistema de medición remota para registrar y almacenar su consumo. Estos datos se utilizan para elaborar informes regulatorios, pero pueden surgir discrepancias debido a fallas en los equipos, errores informáticos o humanos. La falta de un modelo automático para detectar anomalías implica una revisión manual de grandes volúmenes de información. Para mejorar la eficiencia, se propone el uso de aprendizaje automático, que permite predecir la demanda y detectar inconsistencias.

Se revisan estudios previos sobre predicción de demanda en la cadena de valor eléctrica, analizando variables y modelos como K-Vecinos, K-Means y Clustering jerárquico para segmentar los usuarios no regulados. Además, se evalúan modelos de predicción como Prophet, ARIMA y LSTM, seleccionando el más adecuado para la predicción de la demanda de este tipo de usuarios. Finalmente, se presenta el modelo CRISP-DM, que proporciona una estructura eficiente para implementar proyectos de ciencia de datos.

***Palabras clave*:** no regulado, consumo, demanda, algoritmo, segmentación, predicción.

Abstract

The value chain of the Colombian energy sector includes four main actors: generation, transmission, distribution, and commercialization, each of which involves demand forecasting processes. This research focuses on the commercialization stage, specifically on non-regulated users, who negotiate energy costs with retailers and are required to have a remote metering system to record and store their consumption data. This data are used to generate regulatory reports, but discrepancies may arise due to equipment failures, software errors, or human mistakes. The absence of an automated model to detect anomalies necessitates the manual review of large volumes of information. To improve efficiency, the use of machine learning is proposed, enabling demand prediction and inconsistency detection.

The review focuses on previous studies on demand forecasting in the electrical value chain, analyzing variables and models such as K-Nearest Neighbors, K-Means, and Hierarchical Clustering to segment unregulated users. Additionally, predictive models like Prophet, ARIMA, and LSTM are evaluated, selecting the most suitable one for forecasting the demand of this type of user. Finally, the CRISP-DM model is presented, providing an efficient framework for implementing data science projects.

***Keywords*:** unregulated, consumption, demand, segmentation, prediction.

Tabla de Contenido

Lista de Tablas

Lista de Figuras

Introducción

El sector energético es un pilar esencial del desarrollo moderno y un componente clave para las industrias y la sociedad. Por ello, es crucial asegurar que el sistema eléctrico se adapte continuamente a los cambios del mercado de cada país. Esta flexibilidad se alcanza garantizando que los recursos energéticos actuales satisfagan la demanda de los usuarios y desarrollando planes de expansión orientados a cubrir las necesidades futuras.

En el contexto del sector energético colombiano, el Ministerio de Minas y Energía, junto con las entidades responsables de la planeación, regulación y administración del sistema eléctrico, desempeña un papel fundamental en garantizar un suministro confiable y continuo. Estas instituciones implementan estrategias dirigidas a satisfacer la demanda energética actual y a preparar los recursos necesarios para afrontar las demandas futuras en generación, transmisión y distribución, asegurando así la estabilidad del sistema. Una de las principales herramientas para garantizar la confiabilidad y proyección a futuro del sistema eléctrico son los pronósticos de la demanda, para lo cual las entidades regulatorias recopilan información que es reportada por los diferentes actores del sistema eléctrico colombiano.

Una de las principales fuentes de información para garantizar el pronóstico de la demanda proviene de las empresas comercializadoras de energía, que reportan a las entidades regulatorias el consumo horario de los usuarios no regulados. Estos usuarios, que incluyen pequeñas, medianas y grandes empresas, son grandes consumidores de energía. A diferencia de los usuarios regulados, cuyo costo de energía es determinado por el Estado, los usuarios no regulados establecen contratos directamente con las compañías comercializadoras para pactar dicho costo. El proceso de recolección del consumo horario de estos usuarios debe llevarse a cabo de forma remota, almacenarse en una base de datos y ser reportado a la entidad regulatoria.

Un desafío clave al enviar los reportes regulatorios es la calidad de los datos, que puede verse afectada por diversos factores, como fallas en los equipos de medición, errores en los sistemas informáticos o fallos humanos. Esto puede generar predicciones incorrectas de la demanda del mercado, que no reflejan los consumos reales, lo que, a su vez, puede ocasionar pérdidas de energía en las empresas distribuidoras o problemas en los procesos de facturación de los usuarios.

En el contexto de la problemática generada por los errores en los reportes regulatorios, este estudio tiene como objetivo realizar una revisión bibliográfica sobre los aspectos del mercado eléctrico colombiano relacionados con los usuarios no regulados. Además, se busca explorar métodos de aprendizaje automático que permitan clasificar a los usuarios según características determinadas y predecir sus consumos. La predicción de los consumos se presenta como un insumo clave para realizar una comparación entre los valores medidos y los valores predichos, con el fin de establecer si existe una diferencia significativa. Este análisis permitirá identificar posibles errores en los reportes regulatorios y mejorar la calidad de los datos utilizados para la toma de decisiones en el sector energético.

La identificación de una diferencia significativa puede ser de gran utilidad para la empresa comercializadora, ya que permitirá determinar si la discrepancia se debe a un error en el reporte o a un daño en el sistema de medición. Esto contribuirá a mejorar la calidad de los datos en los reportes regulatorios, lo que, a su vez, permitirá que las estimaciones de la demanda realizadas por las entidades gubernamentales estén más ajustadas a la realidad. Con datos más precisos, será posible establecer las políticas necesarias para garantizar la estabilidad de la red eléctrica tanto a corto como a largo plazo.

Contexto y Justificación

El mercado eléctrico colombiano permite la celebración de contratos entre comercializadores de energía y usuarios, como pequeñas, medianas o grandes empresas, en los cuales se establece un costo fijo por kilovatio hora (kWh). Para que el usuario pueda acceder al modelo no regulado, deben superar un consumo mensual promedio de 0.1 MW o 55 MWh durante seis meses continuos. Una vez establecido el contrato, el usuario debe contar con un sistema de medición que registre su consumo hora a hora, cuyos consumos son almacenados por las empresas comercializadoras y reportados al Administrador del Sistema de Intercambios Comerciales (ASIC) dentro de las 48 horas posteriores al día de consumo. Estos reportes diarios, con 24 periodos horarios, son fundamentales tanto para los procesos de facturación como para el reconocimiento económico a las empresas del sistema eléctrico colombiano. Sin embargo, estos registros pueden verse afectados por errores derivados de diversos factores como fallas en los sistemas de medición, problemas en los sistemas informáticos o errores humanos. Estos errores pueden generar reportes inexactos al ASIC o facturaciones incorrectas a los clientes, lo que puede resultar en pérdidas económicas tanto para los usuarios como para la empresa prestadora de servicio. Además, puede afectar el análisis de la demanda eléctrica del mercado no regulado realizada por el Administrador del Sistema de Intercambios Comerciales.

Actualmente, cuando se identifican anomalías en el consumo, los comercializadores deben destinar recursos humanos para analizar las causas de las variaciones, este proceso, debido al alto volumen de datos y la cantidad de usuarios, es ineficiente y propenso a errores, complicando aún más la gestión de grandes bases de datos de información. Por tanto, esta investigación busca desarrollar una base documental que sirva de referencia para la implementación de un modelo de aprendizaje automático que facilite la identificación rápida y precisa de la causa de las anomalías

en el consumo de los usuarios, optimizando los tiempos de análisis y garantizando la exactitud de los reportes regulatorios, en beneficio tanto de las empresas comercializadoras como de los usuarios finales.

Objetivos

Objetivo General

Analizar diversos modelos de segmentación y predicción basados en técnicas de aprendizaje automático, con el propósito de clasificar a los clientes no regulados y anticipar su demanda en el mercado eléctrico colombiano. Este enfoque busca garantizar la calidad de los datos en los reportes regulatorios y detectar de manera eficiente anomalías en los registros de consumo eléctrico, contribuyendo a la mejora de la gestión y la toma de decisiones en el sector.

Objetivos Específicos

Identificar las variables que influyen significativamente en la demanda eléctrica de los clientes no regulados del mercado eléctrico colombiano.

Comparar diferentes algoritmos que pueden ser aplicados para la segmentación de los usuarios no regulados.

Contrastar diferentes algoritmos de aprendizaje automático que pueden ser utilizados para la predicción de la demanda eléctrica, de los usuarios no regulados del mercado eléctrico colombiano.

Marco Conceptual y Teórico

Marco Conceptual

Transformadores de corriente (TC): Según lo establecido por el Código de Medida en la Resolución 038 de 2014 (CREG, 2014), un transformador de corriente (TC), es un dispositivo diseñado para medir la corriente primaria en líneas eléctricas a través de una corriente secundaria. Esta transformación permite ajustar los niveles de corriente a rangos aceptables para su medición y registro.

Transformadores de potencial (TP): De acuerdo con la Resolución 038 de 2014 (CREG, 2014), un transformador de potencial (TP) es un instrumento de medición que reduce el nivel de tensión de las líneas eléctricas a un nivel secundario, facilitando así la medición precisa de valores de tensión.

Usuario No Regulados: La Resolución 131 de 1998 (CREG, 1998b) define como usuario no regulado a aquella persona natural o jurídica que demanda una potencia máxima superior a 0.1 MW o que consuma en promedio al menos 55 MWh-mes durante los últimos 6 meses.

Comisión de Regulación de Energía y Gas (CREG): La CREG fue establecida bajo la Ley 142 de 1994 (Congreso de la República de Colombia), la cual regula los servicios públicos en Colombia. Su misión incluye la regulación del monopolio en la prestación de servicios públicos, fomentar la competencia, garantizar la calidad del servicio y prevenir abusos por parte de entidades en posición dominante.

Administradora del Sistema de Intercambios Comerciales (ASIC): En el contexto del mercado no regulado, la ASIC actúa como la entidad responsable de la gestión del mercado de energía mayorista (MEN). Su función incluye la gestión de la información del MEN, las

transacciones en la bolsa de energía, así como el registro y liquidación de contratos a largo plazo (CREG, 2013b).

Operador de red: De acuerdo con la Resolución 24 de 2013 (CREG, 2013a), el Operador de Red (OR) es un actor fundamental en la distribución y comercialización de energía. Su responsabilidad abarca la planificación de inversiones para la expansión de activos en sistemas de distribución, así como el mantenimiento y operación de dichos activos y las conexiones con el sistema de transmisión.

Niveles de tensión: La Resolución 082 de 2022 (CREG, 2002) establece la clasificación de los niveles de tensión nominal, que se detallan a continuación:

Tabla 1

Niveles de Tensión Nominal

Nivel de tensión	Rango
Nivel 4	Mayor o igual a 57.5 kV y menor a 220 kV
Nivel 3	Mayor o igual a 30 kV y menor a 57.5 kV
Nivel 2	Mayor o igual a 1 kV y menor a 30 kV
Nivel 1	Menor a 1 kV

Nota. Tomados de (Resolución 082, 2022)

Sistema de distribución local (SDL): La Resolución 082 de 2022 (CREG, 2002) define el Sistema de Distribución Local (SDL) como el sistema encargado de transportar energía eléctrica para el servicio de uno o varios Mercados de Comercialización en niveles de tensión 3, 2 y 1.

Contador: Según el Código de Medida Resolución CREG 038 (CREG, 2014), un contador es un dispositivo que registra el consumo de energía en un punto determinado del sistema eléctrico.

Frontera comercial: La Resolución 070 de 1998 (CREG, 1998a) define una frontera comercial como un "punto de medición donde las transferencias de energía permiten determinar la demanda de un comercializador". Estas fronteras pueden clasificarse en fronteras de comercialización entre agentes y fronteras de comercialización para agentes y usuarios.

Consumo eléctrico: La Resolución 108 de 1997 (CREG, 1997) establece que el consumo eléctrico se refiere a la cantidad de kilovatios-hora de energía activa recibida por el suscriptor o usuario en un período determinado, ya sea a través de equipos de medición o mediante cálculos de acuerdo con la metodología establecida en la resolución.

Aprendizaje automático: De (Mitchell, 1997) el aprendizaje automático es la capacidad que tiene un programa informático de aprender gracias a la experiencia de elaborar una tarea y evaluar su rendimiento con base en los resultados de la tarea, para reiniciar el ciclo.

Series de tiempo, Estacionalidad y Tendencia: Según Jansen, S. (2018) una serie de tiempo es un conjunto de datos que tiene una secuencia temporal y para la cual es posible llevar a cabo análisis que permitan entender el pasado o realizar predicciones con base en datos históricos. La estacionalidad es la ocurrencia de un patrón que se repite en intervalos regulares de tiempo. Es habitual encontrar estacionalidad en el clima, la producción de alimentos, la demanda eléctrica, entre otros. La estacionalidad puede manifestarse en periodos de tiempo cuatrimestrales, semestrales, anuales, entre otros. Por otro lado, una tendencia es un patrón observado durante un período de tiempo determinado, que puede ser positiva (alcista) o negativa (bajista). La tendencia no se refiere a una línea recta, sino a un comportamiento que se mantiene a lo largo del tiempo.

Predicción de demanda: De acuerdo con (Armstrong, 2001), la predicción de la demanda es el proceso de estimar la cantidad de un producto o servicio que los consumidores estarán

dispuestos a comprar en un futuro determinado, utilizando técnicas estadísticas y modelos de aprendizaje automático.

Modelos predictivos: Se puede definir un modelo predictivo como un" Proceso de desarrollo de una herramienta o modelo matemático que genere una predicción precisa." (Kuhn & Johnson, 2013, p. 2)

Ciencia de datos: Según (Provost & Fawcett, 2013) describe a la ciencia de datos como un campo interdisciplinario que utiliza métodos, procesos, algoritmos y sistemas para extraer conocimientos y perspectivas de datos estructurados y no estructurados. Combina estadística, informática y conocimiento del dominio para transformar datos en información útil.

Análisis Exploratorio: Para (Tukey, 1977) el análisis exploratorio es una metodología para analizar conjuntos de datos a fin de resumir sus principales características, a menudo utilizando visualizaciones, para identificar patrones y anomalías.

Marco Teórico

El mercado eléctrico colombiano se compone de clientes tanto regulados como no regulados. Los clientes no regulados son aquellos que no están sujetos a los precios fijados por la Comisión de Regulación de Energía y Gas (CREG), lo que les otorga libertad para negociar tarifas directamente con los comercializadores de energía. Este grupo de clientes está compuesto principalmente por grandes empresas e industrias, que muestran patrones de consumo más complejos y menos predecibles en comparación con los clientes regulados. Estos últimos, en su mayoría usuarios residenciales con bajo consumo, cuentan con tarifas de servicio eléctrico reguladas por el Estado.

La segmentación de los clientes no regulados es crucial para entender sus patrones de consumo y anticipar la demanda de energía. La segmentación permite agrupar a los clientes por

características de consumo similares, lo que facilita tanto la predicción de demanda como la optimización de la gestión de la red eléctrica. Para lograrlo, las técnicas de aprendizaje automático se han convertido en herramientas poderosas, ya que permiten clasificar a los clientes de manera eficiente y predecir su comportamiento de consumo futuro a partir de grandes volúmenes de datos históricos y características relacionadas (CREG, 2022).

La segmentación es una técnica de aprendizaje automático empleada para explorar conjuntos de datos y descubrir patrones o relaciones entre diversas variables (Dangeti, 2017). En el contexto de la predicción de la demanda, segmentar a los usuarios no regulados resulta especialmente útil, ya que el consumo histórico del usuario constituye uno de los principales insumos para los pronósticos. No obstante, para mejorar la precisión de las predicciones, es esencial que los datos históricos utilizados en el modelo presenten características similares, lo que ayuda a evitar sesgos y garantiza que los resultados sean más precisos y ajustados a las particularidades de cada segmento de usuarios.

Desde, la promulgación de la Ley 143 de 1994, que redefine el marco legal del sector eléctrico en Colombia, se ha puesto el énfasis en la importancia de atender la demanda del mercado de manera confiable. En la actualidad, la ciencia de datos permite prever y analizar el comportamiento del consumo de los usuarios. Esto facilita la planificación y la toma de decisiones respecto a los recursos de la red eléctrica, asegurando así la confiabilidad del sistema eléctrico. (Moreno & Gutiérrez, 2019).

La Ley 143 establece las actividades que componen la cadena de valor del sector eléctrico, que incluyen generación, transmisión, distribución y comercialización. Esta última está directamente vinculada a la investigación, ya que implica la interacción directa de la empresa comercializadora de energía con los usuarios del mercado no regulado y abarca tareas de

medición, cobro y cumplimiento de normativas regulatorias (Salazar & Panchi, 2014), tareas que se deben realizar con la mayor eficiencia por lo que las herramientas tecnológicas se han convertido en un componente fundamental para el cumplimiento que tienen las compañías prestadoras de servicios de energía con las entidades regulatorias y con la sociedad.

A nivel global, el sector energético está experimentando una evolución hacia fuentes de energía más sostenibles y una mayor participación de usuarios no regulados, quienes desempeñan un papel crucial en la dinámica del mercado eléctrico regional (García et al., 2017). El análisis del comportamiento de los usuarios no regulados es esencial, y este estudio se centrará en la aplicación de técnicas de análisis de características y de predicción para anticipar la demanda de los usuarios y la identificación de anomalías en los sistemas de medición, optimizando así la gestión de recursos en un entorno en constante cambio.

Este marco teórico explorará diversas dimensiones que contextualizan la investigación, desde teorías del comportamiento del consumidor hasta tecnologías de análisis predictivo aplicadas en mercados similares (Herrera & Natán, 2021). Esta comprensión integral será fundamental para diseñar estrategias que satisfagan las necesidades actuales de los clientes no regulados y anticipen sus futuras demandas.

Para ello, existen diversas metodologías para predecir la demanda eléctrica, desde enfoques estadísticos clásicos hasta modelos predictivos de series temporales más complejos como modelos aditivos y modelos de redes neuronales recurrentes (RNN). Por lo que la elección del modelo más adecuado para la predicción de series de tiempo en el consumo o la demanda eléctrica depende de la naturaleza de los datos, las características del usuario que influyen en la variable principal y los objetivos específicos del análisis. Incluso, se ha demostrado que el uso de técnicas de aprendizaje automático en el sector eléctrico mejora la precisión de las predicciones

de demanda, optimiza la gestión de recursos y minimiza las pérdidas operativas (Caicedo & Alfonso, 2023).

Predicción de la Demanda Eléctrica

Estudios previos han evidenciado la necesidad del sector eléctrico en la predicción de la demanda eléctrica, lo anterior por su importancia para la planificación y el cumplimiento de normativas gubernamentales, como es el ejemplo de esta investigación, que plantea un problema en torno a los reportes regulatorios que se deben realizar con los consumos de los clientes no regulados del mercado eléctrico colombiano. Es importante señalar que las investigaciones en la comunidad académica y científica abordan el problema desde una perspectiva más general del sistema eléctrico, que incluye aspectos como la generación, transmisión y distribución. Sin embargo, estas investigaciones se enfocan específicamente en los usuarios clasificados como no regulados. A pesar de ello, el objetivo principal sigue siendo el mismo. Identificar características que puedan influir en la predicción del consumo eléctrico de los usuarios, utilizando herramientas computacionales.

Por lo antes descrito, se trae a colación un caso de estudio del mercado eléctrico colombiano donde se establece que la predicción de la demanda eléctrica en el Valle del Cauca es crucial para la planificación efectiva de la generación, transmisión y distribución de electricidad. Esta capacidad anticipatoria permite optimizar la operación del sistema eléctrico y garantizar una oferta adecuada para los consumidores (Andrade & Castellanos, 2022). La predicción debe considerar el comportamiento de los clientes no regulados, quienes pueden tener preferencias específicas respecto a fuentes de energía y condiciones contractuales.

La operación eficiente de la frontera comercial como se refiere al punto de conexión entre el cliente y la red eléctrica del operador de red requiere considerar diversos factores, como la

oferta y demanda de energía en tiempo real, la capacidad de infraestructura de transmisión y las condiciones económicas y regulatorias. Por lo tanto, la participación de los clientes no regulados en este proceso añade complejidad, ya que su comportamiento influye en la dinámica comercial (Jaimes, 2019). Por ello, es crucial garantizar la fiabilidad de la medición en los puntos de conexión a la red eléctrica. Esto se puede lograr mediante la predicción del consumo del cliente y su comparación con el consumo real, lo que permite identificar anomalías que puedan afectar tanto al usuario final (por cobros injustificados), como a la empresa prestadora del servicio (por pérdidas) o a la entidad regulatoria (por el cálculo erróneo de la demanda del sistema). En consecuencia, se reafirma la necesidad de contar con mecanismos tecnológicos que permitan agrupar a los clientes según sus características y realizar predicciones, asegurando así la fiabilidad de la información.

Antecedentes de Modelos de Predicción en el Sector Eléctrico

La predicción de la demanda es un aspecto crucial para la operación del sistema eléctrico en Colombia. Generalmente, este proceso se enfoca en el mercado de generación y la demanda de los diferentes departamentos del país. En el estudio de (Rojas, 2024), el autor destaca la importancia de proyectar la demanda de un mercado, específicamente el departamento de Antioquia, debido a la necesidad de cumplir con los compromisos regulatorios nacionales. Estos compromisos están orientados a estimar los recursos de generación necesarios para satisfacer la demanda horaria de todo el sistema eléctrico.

Además, el autor subraya que la predicción de la demanda por mercado debe ser reportada a la entidad regulatoria, y que la desviación respecto a la demanda real no debe superar el 4 %. Por lo tanto, la precisión en las proyecciones es un componente esencial para garantizar la estabilidad del sistema. El autor para abordar esta necesidad propone realizar el pronóstico

utilizando diversos modelos, como regresiones lineales, ARIMA y redes neuronales recurrentes, con el objetivo de evaluar la eficiencia de cada uno de los modelos y determinar cuál ajusta con mayor precisión la predicción con respecto a los valores reales.

La predicción de la demanda eléctrica sigue destacándose como un elemento fundamental en las investigaciones consultadas, donde se subraya la necesidad de incluir variables como las condiciones climáticas, factores sociales, económicos y otros aspectos relevantes que contribuyen a robustecer los modelos predictivos. Asimismo, se enfatiza la importancia de prever la demanda no solo desde una perspectiva técnica, sino también estratégica, para garantizar la disponibilidad de recursos de generación en función de las estimaciones realizadas. Este enfoque es respaldado por lo expuesto en el trabajo de (Gil & Ignacio, 2024) donde se presenta una comparación de la demanda del departamento de Antioquia en un periodo de 7 días comprendidos entre el 25 al 31 de octubre de 2023, entre diferentes modelos utilizado para la predicción de series temporales como XGBoost GAM, MLR, LASSO, MARS y LSTM y realiza una comparación de estos modelos con respecto al modelo de la entidad reguladora (XM) con el fin de tener una referencia. Tras los análisis realizados, determina que incluso el modelo LSTM tiene mejores resultados que el modelo predictivo de XM, expone que la inclusión de fechas especiales como el 31 de octubre podría explicar parte de la desviación de la entidad gubernamental. Esta investigación también demuestra la importancia de la segmentación de los usuarios no regulados con base en sus características y la necesidad de la predicción.

Esta investigación se centra en definir la importancia y en presentar algunos métodos para predecir la demanda de los clientes no regulados del mercado eléctrico colombiano, con el objetivo de garantizar la precisión de los reportes regulatorios. Sin embargo, este análisis también está vinculado a otros factores relevantes, como la planificación de la expansión de la

infraestructura del sistema eléctrico y la programación de la generación, necesarios para asegurar una respuesta eficiente ante una demanda fluctuante.

Adicional a los aspectos ya mencionados el autor (Mystakidis et al., 2024) también resalta la importancia de la predicción de la demanda eléctrica para contribuir a los factores ambientales, puesto que una predicción de la demanda más precisa podría evitar la utilización de fuentes de generación de combustibles fósiles, reduciendo la emisión de partículas contaminantes. El autor también hace una revisión del creciente interés y número de investigaciones sobre la predicción de la demanda, así como la revisión de diferentes métodos para el pronóstico a corto, mediano y largo plazo. La investigación del autor busca dar un amplio panorama de los algoritmos utilizados para previsión de la demanda eléctrica, como pueden ser métodos estadísticos, algoritmos de aprendizaje automático, aprendizaje profundo y métodos de ensamble.

En las investigaciones del contexto colombiano, en general se encuentran enfoques a la generación y distribución de la energía en determinados sectores del país como lo hace el autor (Torres, 2023) donde se aborda la necesidad de la predicción de la demanda para el departamento de Antioquia, en la investigación el autor expone las necesidades de generales para el pronóstico de la demanda en el sector eléctrico y expone los tres métodos más utilizados para las previsiones de la demanda como son los métodos estadísticos clásicos, por ejemplo, ARIMA (*Autoregressive Integrated Moving Average*), métodos de aprendizaje automático (*Machine Learning)* como RF-AR (*Random Forest Autoregressive*) o de aprendizaje profundo (*Deep Learning)* como LSTM (Long Short-Term Memory). El autor en sus conclusiones y con base en la experimentación que realizo, expone que el modelo LSTM fue el de mejores resultados. El análisis del autor, respaldado por esta y otras investigaciones, indica que los algoritmos de

aprendizaje profundo han ganado una fuerte acogida en los últimos años para la predicción de series de tiempo en el sector eléctrico, mostrando resultados satisfactorios en comparación con los métodos de análisis clásicos.

Después de realizar la revisión bibliográfica de múltiples investigaciones relacionadas con la predicción de la demanda eléctrica donde se tienen en cuenta múltiples aspectos del sector como son la generación, transmisión, distribución, comercialización y consumo final, se resalta el propósito de la presente investigación que busca abordar una problemática de uno de los actores más importantes del sector eléctrico colombiano como son los usuarios no regulados y los reportes que deben realizar las empresas comercializadoras de energía.

Según cifras del administrador del mercado eléctrico colombiano (XM, 2024) los usuarios no regulados para agosto de 2024 representaron el 30.47% del consumo de energía de todo Colombia, una cifra muy significativa si se tiene en cuenta que este mercado agrupa a las empresas con altos consumos y no incluye los usuarios de bajo consumo en su mayoría residenciales.

Un aspecto clave para la predicción de la demanda en las diversas fases de la cadena de valor del sector eléctrico, generación, transmisión, distribución y comercialización en los modelos de *machine learning* y *Deep learning* son las variables exógenas que se integran para complementar los datos históricos de consumo y mejorar la precisión de las previsiones. En este contexto, el autor (Aziz et al., 2024) realiza una revisión de las variables más utilizadas para predecir la demanda de energía a largo plazo en el sistema eléctrico de Pakistán, enfocándose en las variables económicas y demográficas. El autor señala que estas variables por sí solas pueden no generar la exactitud esperada en la predicción, y destaca que tales errores en la previsión de la demanda han causado cortes en el suministro eléctrico del país.

Para mejorar la precisión, el autor resalta la necesidad de incluir variables adicionales que permitan una mejor determinación de la demanda máxima, tales como la carga pico mensual, factores estacionales, climáticos, consumos históricos y la distinción entre los diferentes días de la semana. Esta revisión subraya la importancia de las variables complementarias al consumo de los clientes para predecir la demanda, y cómo su selección varía según las particularidades de cada país, fortaleciendo así los modelos predictivos en diferentes contextos internacionales.

Es importante exponer como distintos contextos demográficos presentan diferentes perspectivas de los tipos de variables que son útiles para mejorar la precisión de los modelos, por lo que en el autor (Chung & Jang, 2022) presenta el panorama desde el punto de vista de una de las primeras económicas del mundo y con un alto desarrollo tecnológico, como lo es Corea del Sur. En el estudio, el autor destaca la importancia de la previsión de la demanda eléctrica para mantener el crecimiento económico y el estilo de vida de las naciones, de una forma que pueda ser sostenible. Para lo cual son primordiales los pronósticos de la demanda, con el fin de planear la expansión y sostenibilidad de los recursos necesarios para la operación del sistema eléctrico. Precisa que es necesaria la implementación de datos multivariados con el fin de realizar una estimación precisa de la demanda y el consumo energético, ajustada a las necedades de cada región. De igual forma, se destacan que los modelos LSTM (Long Short-Term Memory) y CNN (Red Neuronal Convolucional) se encuentran presentes en múltiples estudios, lo que los posiciona como una buena opción para su incorporación en la predicción del consumo eléctrico. La investigación presenta diversas variables que son utilizadas en diferentes investigaciones y que resaltan su importancia. Variables tales como las cargas eléctricas, la meteorología, o métricas relacionadas con la macroeconomía y la demografía.

La revisión de los antecedentes históricos resalta la importancia y actualidad de la investigación sobre la predicción de la demanda eléctrica, tanto para la industria como para la sociedad. Además, ofrece un panorama de algunos de los modelos más utilizados y las variables que, según diversas investigaciones, tienen un mayor impacto en la precisión de los modelos, las cuales varían según el contexto de la región de aplicación.

Análisis de Variables Clave para Predecir la Demanda en Clientes No Regulados

El consumo eléctrico de un usuario no regulado puede verse influenciado por diversos factores, tanto externos como internos. Entre los factores externos se encuentran condiciones climáticas, eventos socioeconómicos, festividades, variaciones en los precios de la energía o patrones de consumo del usuario, los cuales están fuera del control del operador de la red eléctrica. Por otro lado, los factores internos abarcan aspectos técnicos relacionados con el punto de conexión a la red de distribución, como daños en los equipos de medición, fallas en los diseños eléctricos o eventos que afecten la red de distribución. Dado que estos factores pueden ser gestionados por el operador de red, una predicción que facilite la identificación de anomalías no solo contribuye a mejorar la precisión del pronóstico, sino que también se convierte en una herramienta valiosa para optimizar los indicadores de calidad del servicio.

Dada esta diversidad de factores, resulta crucial identificar y priorizar aquellos que sean previsibles y relevantes para la empresa comercializadora de energía. Esto permitirá mejorar la precisión de los modelos de predicción, optimizar los pronósticos de demanda en el mercado y detectar anomalías en los reportes regulatorios, garantizando una mayor fidelidad y confiabilidad en los mismos.

Por consiguiente, en esta sección se presentarán las principales variables que afectan la demanda eléctrica, acompañadas de un análisis sobre su influencia en el proceso de predicción. Cabe resaltar que la selección de las variables está fundamentada en dos aspectos, en primer lugar, con base en las variables identificadas en diferentes investigaciones, como lo presenta el autor (Román et al., 2021) quien realiza una revisión de diferentes investigaciones sobre técnicas de aprendizaje y su utilización en problemas específicos de demanda eléctrica. El autor en la investigación también incluye diferentes variables de entrada que son utilizadas para mejorar la

precisión de los algoritmos y ajustarlos a diferentes circunstancias. El autor resalta la presencia de diferentes variables en múltiples investigaciones, destacándose el histórico de la carga como la base a partir de la cual se incluyen otras variables como condiciones climáticas, variables económicas o patrones de consumos de los usuarios.

En segundo lugar, se lleva a cabo un análisis enfocado en los usuarios no regulados del sector eléctrico colombiano, ya que las investigaciones muestran diferentes perspectivas según el contexto. Este análisis se ajusta a las particularidades de la problemática planteada, directamente relacionadas con el modelo del sistema eléctrico en Colombia. Además, el estudio de las variables en el contexto del mercado colombiano se complementa con la realización de una encuesta a personas que desempeñan tareas que tiene alguna relación con la demanda de los usuarios del mercado eléctrico no regulado, lo que valida la selección de las variables propuestas y proporciona un panorama más amplio basado en la experiencia profesional de los encuestados.

Variables Externas

Condiciones climáticas: El factor climático puede ser anticipado, en cierta medida, mediante los pronósticos meteorológicos proporcionados por las entidades gubernamentales especializadas en esta tarea. Entre las variables más relevantes y que pueden afectar la predicción de la demanda se encuentran la temperatura, la humedad y las precipitaciones, ya que son las que más pueden influir en el consumo de los usuarios.

Las condiciones climáticas influyen en los patrones de consumo de los usuarios, en mayor o menor medida, dependiendo de la actividad comercial. Por ejemplo, los pozos de riego muestran una tendencia claramente definida durante la temporada de lluvias, su consumo disminuye significativamente, mientras que en épocas de sequía se observa un aumento drástico en la energía demandada de la red.

Días festivos y jornadas laborales: Las jornadas laborales y los días festivos tienen un impacto directo en la forma de la curva de demanda de los usuarios. Esta puede variar según el tipo de empresa, aquellas con horarios laborales diurnos presentan patrones diferentes a las compañías cuya actividad depende de procesos de producción continuos, caracterizados por un consumo más uniforme a lo largo del día. Por otro lado, las festividades también pueden alterar significativamente el comportamiento de la curva de demanda, ya que muchas empresas suspenden actividades en estos días. Por ejemplo, un miércoles festivo mostraría un comportamiento completamente atípico en comparación con un miércoles laborable, donde la actividad y el consumo eléctrico se desarrollan con normalidad.

Tipo del punto de medición: Esta variable está relacionada con los niveles de consumo de los clientes, ya que se refiere a la clasificación por capacidad instalada del usuario, permitiendo agruparlos en cinco tipos de puntos de medición, del 1 al 5. Es una variable relevante porque facilita la segmentación de los usuarios en distintos rangos de consumo, lo que la convierte en un factor clave al realizar la predicción.Factores sociales: Los factores sociales pueden afectar la demanda del usuario de forma significativa por ejemplo por salud pública, protestas sociales o huelgas laborales, donde en muchos sectores productivos de la sociedad las actividades económicas se detienen o se ven fuertemente afectadas por falta de demanda o problemas de desplazamiento para los trabajadores o las líneas de abastecimiento.Políticas regulatorias: Las entidades gubernamentales que regulan el sistema eléctrico poseen la autoridad legal para emitir resoluciones que pueden modificar los patrones de consumo de los clientes. Un ejemplo de esto es la autorización para que las compañías instalen y operen sus propias unidades de generación de energía. Esto reduce su dependencia de la red de distribución, ya que solo recurren a ella cuando su capacidad de generación no es suficiente para satisfacer su demanda

interna.Tendencias en energías renovables: La instalación de sistemas de energía renovable por parte de los usuarios, como los paneles solares, impacta la demanda de energía, ya que en ciertos momentos del día se reduce la necesidad de consumir electricidad del sistema eléctrico del operador de red. Por lo que, por ejemplo, se puede ver modificada su curva de demanda en horarios de mayor radiación solar.Actividad económica del cliente:La categorización de la actividad comercial del cliente puede dar un panorama de sus patrones de consumo. Por ejemplo, es diferente el patrón de consumo de una institución educativa a un establecimiento de entretenimiento nocturno.Consumo Histórico de Energía: Esta variable se clasifica como interna, ya que depende de las dinámicas de consumo del cliente y se refiere a los registros históricos de consumo del usuario.

Variables Internas (Operativas y Comerciales)

Interrupción del servicio: Los cortes en el suministro eléctrico, ya sea por mantenimiento programado o por fallas en la red de distribución, constituyen un factor controlable por el operador de la red. Estos eventos pueden ser planificados, como parte de un cronograma de mantenimiento, o imprevistos, originados por incidentes en la red. En ambos casos, el operador suele tener información detallada sobre las causas y la duración estimada de los cortes, lo que permite incorporar este parámetro en los modelos de predicción para mejorar su precisión y adaptabilidad a estas situaciones.

Contratos vigentes: Por otro lado, los contratos o acuerdos comerciales entre la compañía comercializadora de energía y el usuario final se negocian para definir tanto el costo como las condiciones del servicio. Como resultado, un cliente puede optar por reducir su consumo si, por ejemplo, en la renovación del contrato, el costo por kilovatio-hora incrementa significativamente.

Encuesta de Identificación de Variables Relevantes

En el marco de las argumentaciones relacionadas con las variables seleccionadas, se realizó una revisión exhaustiva de investigaciones que resaltan la importancia de dichas variables para adaptar los modelos a las necesidades específicas de cada problemática y mejorar la precisión en los resultados de predicción. Este análisis se complementó con una evaluación detallada de cada variable en el contexto de la problemática planteada. Como resultado, se identificaron las variables clave, lo que permitió diseñar una encuesta que incluyó las siguientes consideraciones.

La población objetó de estudio pertenece a CELSIA S.A., una compañía del sector eléctrico colombiano con participación en los componentes de generación, distribución y comercialización dentro del sistema eléctrico nacional. La encuesta se dirigió a una muestra de trabajadores de la empresa, específicamente aquellos involucrados en áreas como reportes regulatorios, análisis de demanda y supervisión de medidas. Debido a sus funciones de recopilar, almacenar, analizar y garantizar la integridad de la información regulatoria, esta muestra cumple con los criterios necesarios para identificar las variables relevantes en la predicción de la demanda de usuarios del mercado no regulado colombiano.

La encuesta solicitó la siguiente información

1. Rol dentro del sector eléctrico
2. Asigne un nivel de importancia donde “Sin relevancia” es el nivel más bajo y “Muy Relevante” es el nivel más alto
3. Se pregunta al encuestado si considera que alguna variable adicional es relevante para la segmentación y predicción de la demanda de los usuarios no regulados.

En la Figura 1 se muestra la estructura de la encuesta.

Figura 1

Encuesta para Identificar las Variables de Segmentación y Predicción

1. ¿Cuál es su rol dentro del sector eléctrico? *

Escriba su respuesta

2. La encuesta tiene como objetivo evaluar el nivel de relevancia de las variables listadas, con el fin de determinar su importancia en el proceso de segmentación y predicción de la demanda de los usuarios no regulados, según sus características.

Asigne un nivel de importancia donde "Sin relevancia" es el nivel más bajo y "Muy Relevante" es el nivel más alto. *

	Sin Relevancia	Relevancia Baja	Relevancia Moderada	Relevancia alta	Muy Relevante
Factores climáticos (temperatura, humedad, etc.)	○	○	○	○	○
Días festivos y jornadas laborales	○	○	○	○	○
Factores sociales (economía, seguridad, salud pública)	○	○	○	○	○
Políticas Regulatorias	○	○	○	○	○
Incorporación de energías renovables (Autogeneradores)	○	○	○	○	○
Actividad económica del usuario	○	○	○	○	○
Consumo histórico de energía del usuario	○	○	○	○	○
Contratos vigentes con el usuario	○	○	○	○	○
Interrupciones del servicio (cortes programado, fallas en la red, etc.)	○	○	○	○	○
Tipo de punto de medición (consumo o capacidad instalada)	○	○	○	○	○

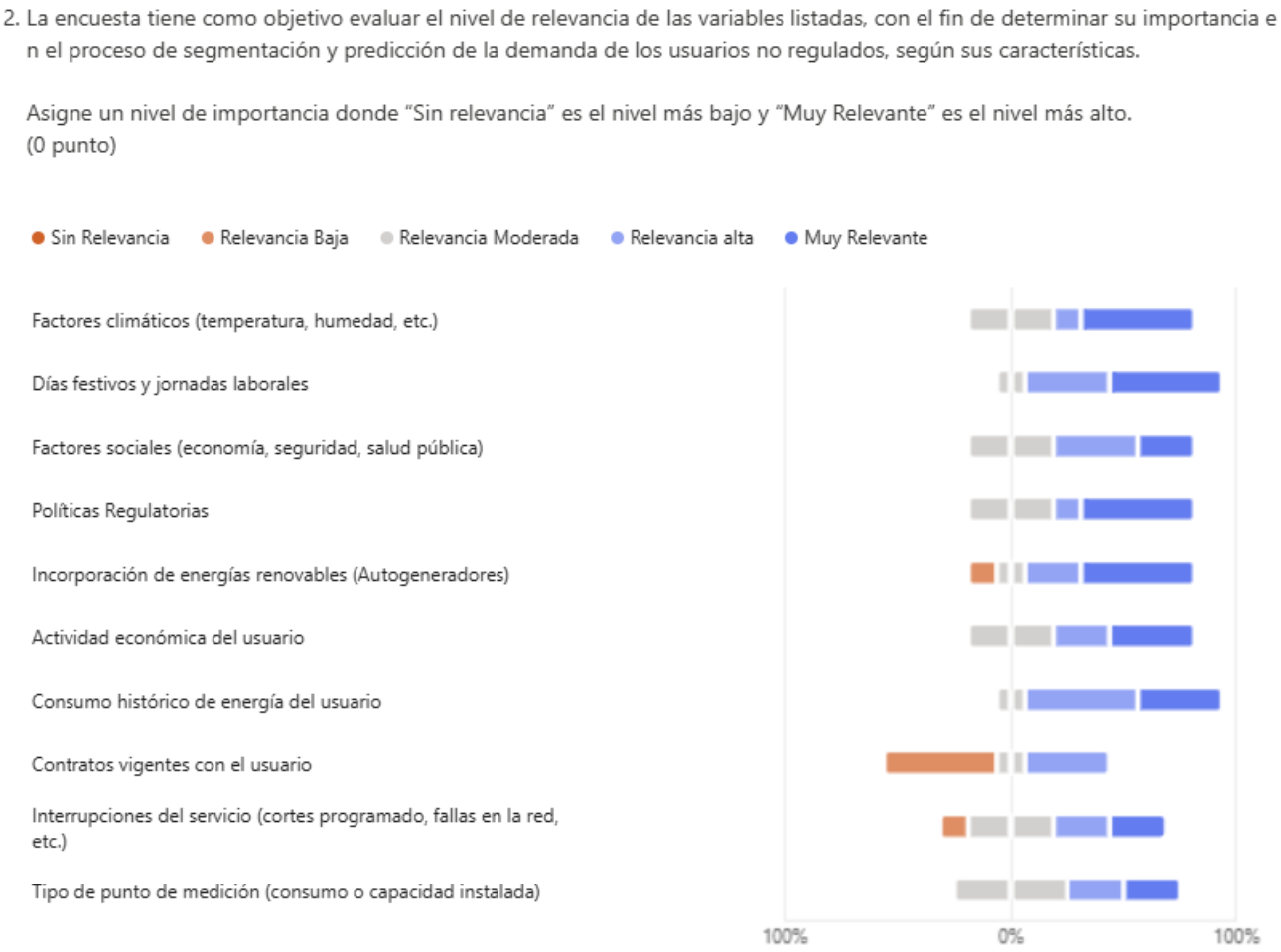
3. ¿Qué otra variable considera relevante para la segmentación y predicción de la demanda de los usuarios no regulados en el contexto del mercado eléctrico colombiano?

Escriba su respuesta

El total de encuestados fue de 8 participantes. En la Figura 2 se presenta un resumen de las respuestas recopiladas mediante Microsoft Forms. Se observa que la mayoría de los participantes asignaron un nivel de importancia muy bajo a los contratos vigentes, mientras que otorgaron el nivel más alto de relevancia a los días festivos y las jornadas laborales.

Figura 2

Resumen de Respuestas de la Encuesta

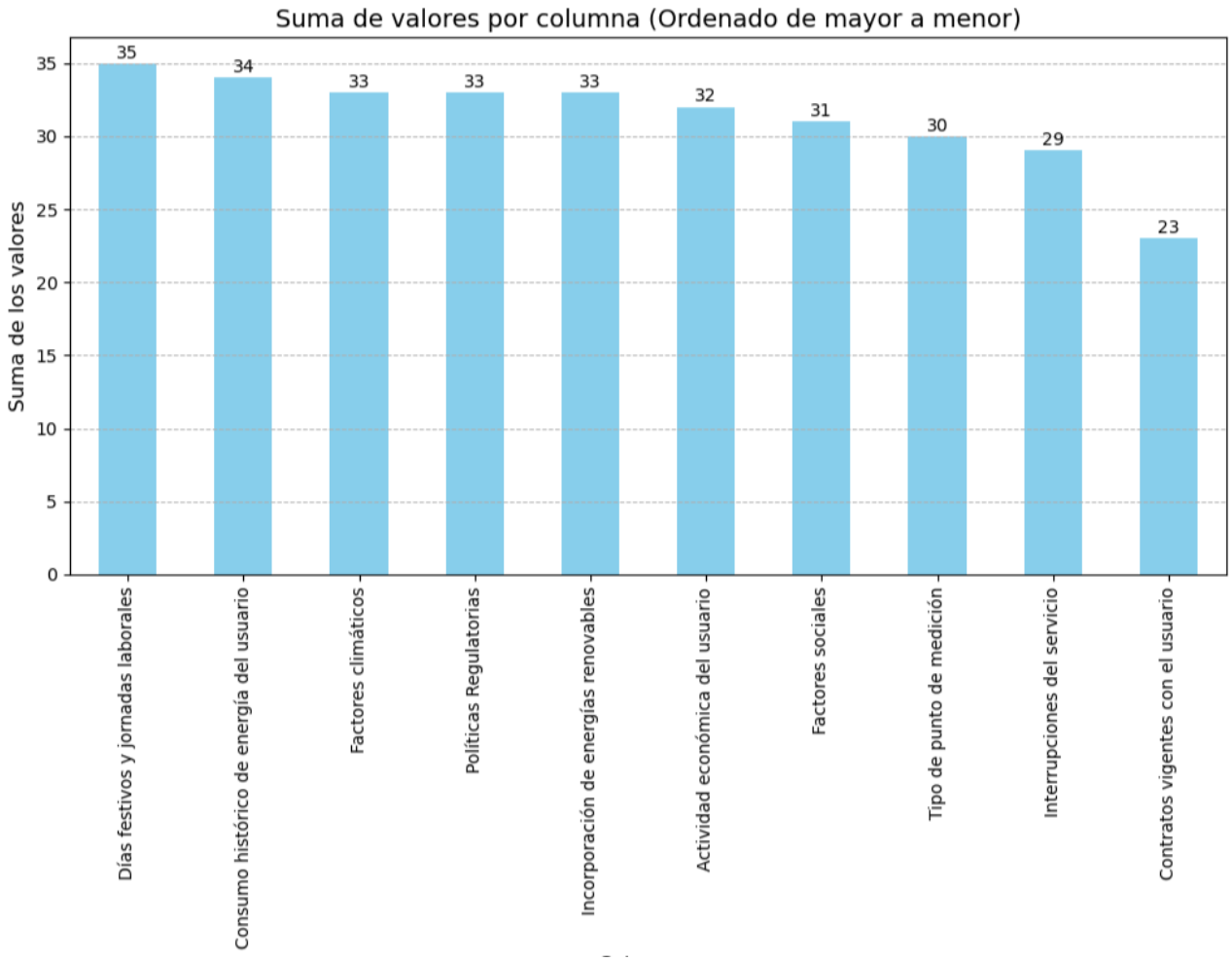


En la Figura 3 se presenta un gráfico de barras que muestra la suma de los niveles de relevancia asignados por los encuestados a cada variable. Las categorías de relevancia están calificadas en una escala de 1 a 5, donde: 1 corresponde a "Sin Relevancia", 2 a "Relevancia Baja", 3 a "Relevancia Moderada", 4 a "Relevancia Alta", y 5 a "Muy Relevante". Los resultados confirman que las variables con los niveles de importancia más altos fueron los días festivos y

las jornadas laborales, seguidas por los consumos históricos. Por otro lado, las interrupciones del servicio y los contratos vigentes obtuvieron los niveles más bajos de relevancia.

Figura 3

Gráfico de Importancia por Variable.

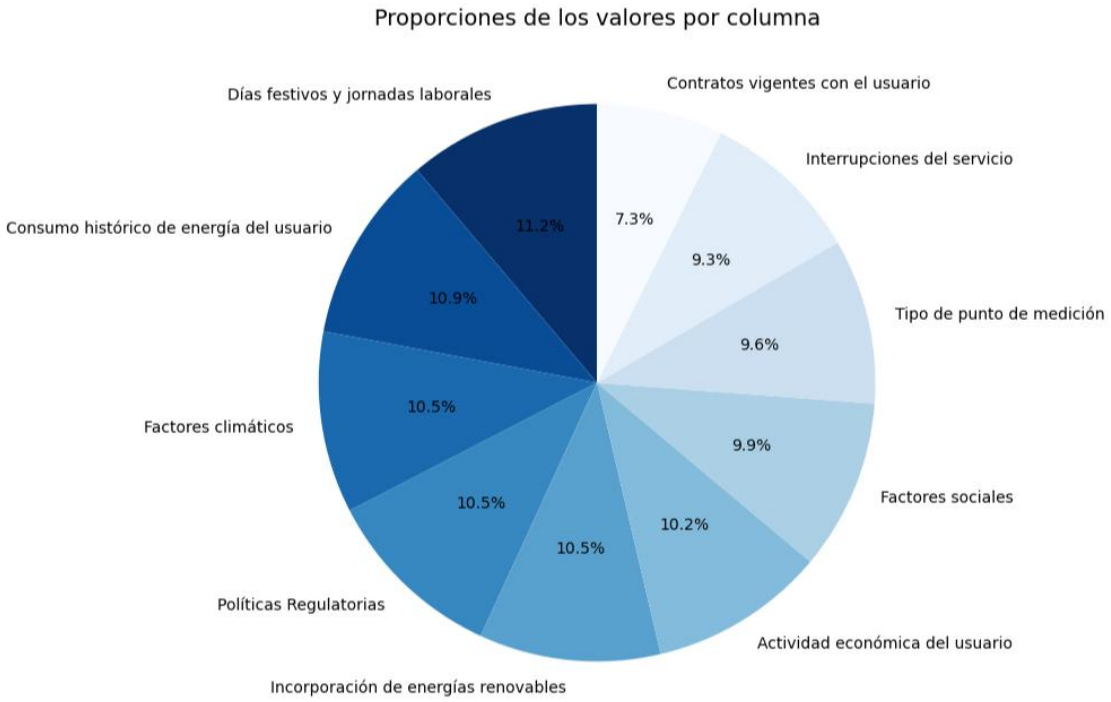


La Figura 4 ilustra de manera proporcional el grado de importancia asignado a cada variable según la suma de las calificaciones proporcionadas por los encuestados. Se destaca que los días festivos obtuvieron la mayor relevancia con un 11.2%. Además, se identificaron cuatro variables con niveles de importancia muy similares, que oscilan entre el 10.5% y el 10.2%, lo

que refleja un grado de relevancia comparable según las evaluaciones realizadas. Estas variables son fundamentales para la segmentación y predicción de la demanda, considerando que presentan una diferencia de solo 3.2% en relación con la variable menos relevante.

Figura 4

Porcentajes de Importancia de las Variables



La encuesta confirmó que las variables seleccionadas son relevantes para la segmentación y predicción de la demanda de usuarios no regulados en el sector eléctrico colombiano. Esto se evidencia en que, desde la segunda variable más importante hasta la penúltima menos importante, la diferencia es de apenas 1.6%, lo que sugiere un grado de importancia relativamente homogéneo entre estas variables. Sin embargo, se destaca que la variable relacionada con los contratos vigentes con los usuarios no regulados fue considerada

mayoritariamente como no relevante, lo que la convierte en una buena candidata para ser excluida en procesos de reducción de dimensionalidad.

En la última pregunta de la encuesta, se solicitó a los participantes que identificaran otras variables relevantes para la segmentación y predicción de la demanda. De las respuestas obtenidas, dos encuestados resaltaron la importancia de considerar aspectos demográficos. Además, se mencionaron, aunque de forma aislada, variables como la presencia de medidores de respaldo en puntos de frontera, las jornadas de mantenimiento en las plantas, las pérdidas técnicas de la red, y la consideración de diferentes puntos de conexión a la red para un solo cliente. Estas observaciones podrían ser valiosas para enriquecer el análisis en futuros estudios.

Análisis Exploratorio de Datos

El análisis exploratorio de datos es una de las primeras tareas que debe llevar a cabo un científico de datos para comprender y adaptar un conjunto de datos a los objetivos de un proyecto. En la Guía práctica de introducción al análisis exploratorio de datos (Gobierno de España, 2024) publicada en el portal de datos abiertos del gobierno español, se presenta una metodología que establece pasos clave para procesar los datos. Entre ellos, se destaca la identificación de datos ausentes y atípicos. Estos dos aspectos son particularmente relevantes, ya que pueden tener un impacto significativo en las predicciones. Un dato ausente podría ser interpretado incorrectamente como un valor cero, mientras que un dato atípico podría distorsionar el modelo al referirse a un valor no representativo del comportamiento habitual de la curva de consumo. Debido a su importancia, este paso es fundamental en el proceso de análisis de series de tiempo, ya que, sin datos limpios, la segmentación y las predicciones podrían no arrojar los resultados esperados. Además, ciertos algoritmos requieren estructuras de datos específicas, lo que hace aún más crucial este proceso de preparación.

Por otro lado, en el artículo sobre la revisión sistemática del pronóstico de la demanda eléctrica y el uso de algoritmos de aprendizaje automático, (Román et al., 2021) destacan la relevancia y la necesidad del análisis exploratorio durante la preparación de los datos antes de su uso en modelos de aprendizaje automático. Esto subraya la importancia de este proceso en el modelado de datos dentro de la ciencia de datos y resalta el papel crucial que esta fase en el ciclo de vida de un proyecto de este tipo.

Métodos de Segmentación

Algoritmos de Aprendizaje Automático para Segmentación

Existen varios modelos en aprendizaje automático que pueden aplicarse a la segmentación de clientes no regulados en el mercado eléctrico. Estos modelos varían en complejidad, interpretabilidad, y rendimiento. Su elección depende de las características del conjunto de datos y de los objetivos específicos del análisis. (Bishop, 2006).

GMM (Gaussian Mixture Model)

El Gaussian Mixture Model (GMM) es un algoritmo de aprendizaje no supervisado que se utiliza para segmentación y agrupamiento de datos. Se basa en el supuesto de que los datos provienen de una mezcla de varias distribuciones normales (o gaussianas). Cada una de estas distribuciones representa un clúster en el conjunto de datos. En lugar de asignar cada punto de datos a un único clúster (como en K-Means), GMM asigna probabilidades de pertenencia a cada clúster, es decir, estima la probabilidad de que un punto pertenezca a una distribución gaussiana específica. (Bishop, 2006).

K-Means (Algoritmo de Clustering)

Aunque K-Means es un algoritmo de agrupamiento no supervisado, es relevante mencionarlo, ya que se utiliza en el análisis de datos para clasificar instancias en grupos (clusters).

K-Means asigna cada instancia a uno de los K clústeres predeterminados, basándose en la distancia a los centroides de los clústeres. El algoritmo intenta minimizar la suma de las distancias cuadradas entre las instancias y los centroides de sus clústeres asignados (Bishop, 2006).

Clustering Jerárquico

El agrupamiento jerárquico es un algoritmo de agrupamiento no supervisado que busca construir una jerarquía de grupos (clústeres). A diferencia de los métodos como K-Means, que requieren que se especifique previamente el número de clústeres, el agrupamiento jerárquico no lo hace. Este algoritmo construye un dendrograma, donde cada nodo representa un grupo de instancias que se van fusionando o dividiendo según una métrica de similitud o distancia (Iglesias et al., 2018).

Comparación de Algoritmos de Segmentación

Tabla 2

*Comparación de Características de Modelos de Segment*ación

Característica	GMM	K-Means	Clustering
Tipo de modelo	No Supervisado	No Supervisado	No Supervisado
Complejidad de implementación	Media, requiere EM y parametrización	Baja, relativamente sencillo, solo se necesitan el número de clusters (k) y un criterio de convergencia.	Alta, más complejo debido a la construcción de la jerarquía de clusters. Requiere elección de método de enlace.
Volumen de datos requerido	Funciona mejor con más datos	Bajo rendimiento en series temporales cortas	Requiere grandes volúmenes de datos
Predicción en diferentes intervalos de tiempo	No diseñado para predicción temporal	Tiene buen rendimiento en predicciones a mediano y largo plazo	Tiene un buen rendimiento en corto, mediano y largo plazo.
Recursos computacionales	Iterativo, cálculos complejos	Requerimientos moderados en la ejecución	Requiere alto poder de cómputo
Aplicabilidad al sector eléctrico	Bueno para segmentaciones complejas	Útil para trabajar con series estacionales	Óptimo para patrones complejos y múltiples entradas

Descripción de Algoritmos Seleccionados

K-Means es uno de los algoritmos más utilizados en el sector eléctrico, especialmente en tareas de segmentación de consumidores, análisis de patrones de consumo energético y agrupación de comportamientos similares. Sus características lo hacen particularmente adecuado para este tipo de aplicaciones. Gracias a su escalabilidad K-Means puede manejar grandes volúmenes de datos, lo cual es esencial en el sector eléctrico, donde se suelen analizar datos de consumo de miles o millones de usuarios. Además, es muy útil para clasificar a los usuarios en diferentes grupos, por ejemplo, en función de su consumo energético (bajo, medio, alto) o en función de su comportamiento en períodos de alta demanda.

A continuación, se mostrará la matriz de criterios seleccionados y porque el K-Means es el algoritmo seleccionado en la investigación**.**

Criterios de Evaluación

1. Tipo de modelo: Supervisado o no supervisado.
2. Complejidad de implementación: Nivel de dificultad de implementación del modelo.
3. Volumen de datos requerido: Cantidad de datos necesarios para que el algoritmo funcione de manera eficiente.
4. Predicción a largo plazo: Capacidad para realizar predicciones a largo plazo.
5. Recursos computacionales: Requerimientos de recursos (poder de cómputo) para ejecutar el modelo.
6. Aplicabilidad al sector eléctrico: ¿Qué tan adecuado es el algoritmo para trabajar con datos del sector eléctrico?

Asignación de Pesos

Cada uno de los criterios que se presentan tiene un peso que refleja su importancia en la selección del algoritmo de segmentación. A continuación, la distribución de los pesos se basa en las características particulares de cada algoritmo y su aplicabilidad al sector eléctrico colombiano.

Todos los algoritmos son no supervisados, por lo que este criterio tiene un peso uniforme, ya que no hay diferencia significativamente entre ellos. La complejidad es variable para cada uno de los algoritmos seleccionados, por lo que tiene un peso distinto para cada uno. En cuanto al volumen, los tres algoritmos se aplican para trabajar con una gran cantidad de datos, pero no todos son muy eficientes al momento de procesarlos. La precisión no es un punto fuerte de los algoritmos estudiados, pero en el caso de K-Means se puede adaptar rápidamente a las predicciones de corto plazo. Respecto al uso de recursos computacionales, los tres algoritmos son exigentes debido a que trabajan mejor con grandes volúmenes de datos. Finalmente, la aplicabilidad al sector eléctrico es el eje principal de la monografía, por lo que su peso es relevante en la investigación.

- Tipo de modelo: 0.1
- Complejidad de implementación: 0.15
- Volumen de datos requerido: 0.2
- Predicción en diferentes intervalos: 0.1
- Recursos computacionales: 0.2
- Aplicabilidad al sector eléctrico: 0.25

Tabla 3

Evaluación de Algoritmos de Segmentación

Criterios	GMM	K-Means	Clustering Jerárquico
Tipo de modelo (Supervisado=1, No Supervisado=5)	5	5	5
Complejidad de implementación (Baja=5, Alta=1)	2	4	2
Volumen de datos requerido (No requiere gran cantidad=5, Requiere grandes volúmenes=1)	5	4	2
Predicción en diferentes intervalos de tiempo (Bueno para corto/largo=5, Solo corto=1)	1	4	5
Recursos computacionales (Bajos=5, Altos=1)	2	4	2
Aplicabilidad al sector eléctrico (Óptimo=5, Menos adecuado=1)	3	4	5

Cálculo de la Puntuación Total para Cada Algoritmo

Ahora multiplicamos las puntuaciones por los pesos y sumamos los resultados para obtener la puntuación final de cada algoritmo.

GMM

$$(5 \cdot 0.1) + (2 \cdot 0.15) + (5 \cdot 0.2) + (1 \cdot 0.1) + (2 \cdot 0.2) + (3 \cdot 0.25) = 3.05 \quad (1)$$

K-Means

$$(5 \cdot 0.25) + (4 \cdot 0.15) + (4 \cdot 0.1) + (4 \cdot 0.2) + (4 \cdot 0.2) + (4 \cdot 0.1) = 4.25 \quad (2)$$

Clustering Jerárquico

$$(5 \cdot 0.25) + (2 \cdot 0.15) + (2 \cdot 0.1) + (5 \cdot 0.2) + (2 \cdot 0.2) + (5 \cdot 0.1) = 3.65 \quad (3)$$

Tabla 4

Puntuación Final de Algoritmos de Segmentación

Algoritmo	Puntuación Total
GMM	3.05
K-Means	4.25
Clustering Jerárquico	3.65

Interpretación de los Resultados

K-Means tiene la puntuación más alta (4.25), lo que lo hace el algoritmo más adecuado bajo los criterios seleccionados.

Clustering Jerárquico tiene la segunda puntuación más alta (3.65), pero debido a su mayor complejidad de implementación y mayor requerimiento de recursos, no es tan adecuado como K-Means en este caso.

GMM, al ser un algoritmo supervisado, tiene la puntuación más baja (3.05), lo que refleja que no se ajusta a los objetivos establecidos en la investigación.

K-Means es el algoritmo más adecuado para la segmentación en el sector eléctrico colombiano, debido a su escalabilidad, facilidad de implementación y eficiencia en el manejo de grandes volúmenes de datos. Su aplicabilidad en la segmentación de clientes de acuerdo con su consumo energético lo hace ideal para este contexto, mientras que GMM y Clustering Jerárquico pueden ser útiles en situaciones más específicas, pero presentan desventajas en cuanto a complejidad computacional y escalabilidad. (Jain, 2010)

Resultados y Discusión

La discusión entre los tres algoritmos de segmentación (GMM, K-Means y Clustering Jerárquico) en el contexto del sector eléctrico se basa en varios criterios clave: escalabilidad, eficiencia computacional, aplicabilidad a grandes volúmenes de datos y objetivos de segmentación específicos.

Tras evaluar los tres algoritmos en función de los criterios mencionados, K-Means emerge como la opción más adecuada para la segmentación en el sector eléctrico, puesto que es capaz de manejar grandes volúmenes de datos, lo cual es esencial en el sector eléctrico, donde los datos de consumo de energía se generan en grandes cantidades, especialmente cuando se gestionan millones de registros de usuarios. En comparación con GMM y el clustering jerárquico, K-Means ofrece una buena relación entre eficiencia y precisión. Aunque no es tan preciso como métodos más avanzados, su rapidez y bajo costo computacional lo hacen adecuado para tareas de segmentación en tiempo real y análisis de grandes bases de datos. Asimismo, K-Means permite segmentar a los usuarios o clientes en diferentes grupos de consumo energético, lo cual es útil para la optimización de redes, la tarificación dinámica y la gestión de demanda. Estas tareas son esenciales para una correcta administración y predicción de la demanda energética.

Este trabajo es fundamental para entender las aplicaciones de algoritmos de clustering como K-Means. En él, se discuten las ventajas y limitaciones de varios algoritmos de segmentación y el impacto de elegir un algoritmo adecuado para distintos contextos, como el de grandes volúmenes de datos, lo que es típico en el sector eléctrico. (Jain, 2010)

Métodos de Predicción

Algoritmos Para la Predicción de Series Temporales

Prophet

De (Taylor & Letham, 2017) Prophet fue presentado en el artículo científico "Forecasting at Scale" por Sean J. Taylor y Ben Letham, documento en el que se describe a Prophet como un modelo estadístico para la predicción de series temporales, en el cual es posible manejar tendencias no lineales, estacionalidad y adicionar eventos especiales o atípicos que puedan afectar la precisión del modelo, este modelo resulta particularmente eficiente si se habla de la predicción del consumo eléctrico de los clientes no regulados en los cuales el consumo eléctrico depende mucho de la estacionalidad, como por ejemplo una universidad cambiaria en gran proporción su consumo si se tiene un día festivo donde no se tendría clase, si este día se compara con un día habitual la predicción podría estar segada por tomar como referencia días con consumos diferentes, así mismo puede ocurrir con las estacionalidades como por ejemplo, en una compañía productora de licor donde es de esperase que los meses previos a las festividades aumente su producción, por lo que resultaría equivocado comparar los meses previos a las festividades con los meses posteriores a las festividades donde baja el consumo de este tipo de productos.

Modelo ARIMA

En el artículo científico (Contreras et al., 2003) se lleva a cabo un análisis del modelo ARIMA implementando un modelo que realice la predicción del precio de la energía eléctrica en el estado de California en Estados Unidos y de la España peninsular, se resaltan las bondades del modelo ARIMA que ha sido utilizado en múltiples aplicaciones para la predicción del consumo eléctrico y de otras industrias como la de hidrocarburos y en la predicción del costo de las

materias primas. En el artículo científico se describe al modelo de medias móviles integradas autorregresivas más conocido por sus siglas en inglés ARIMA (Autoregressive Integrated Moving Average) como un modelo estadístico utilizado para la predicción de series de tiempo, el cual principalmente está basado en tres componentes los cuales son, el *Autorregresivo* (AR) que describe la relación entre una observación y valores pasados en la serie de tiempo, *Integrado* (I) el cual describe una de las características importantes y es que se requiere que las series sean estacionarias por lo que este componente indica la diferencia requerida para que la serie sea estacionaria, finalmente el componente de la *Media Móvil* (MA) se refiere a la captura de la relación entre los rezagos de la serie temporal y una observación.

Redes Neuronales LSTM

El artículo científico (Huang et al., 2023) se establece que el modelo LSTM (Long Short-Term Memory) está diseñado para superar las limitaciones de las dependencias a largo plazo en las redes neuronales recurrentes (RNN). A diferencia de las RNN básicas, el modelo LSTM incorpora una entrada y una salida adicionales, formando una celda de estado que integra una compuerta con la memoria de la red. Estas características adicionales del modelo LSTM permiten integrar variables ambientales y capturar patrones temporales complejos.

Comparación de Algoritmos de Predicción

La **Tabla 5** busca realizar una comparación con base en el análisis de la herramienta en línea Neptuno.ai (Kutzkov, 2022) que realiza seguimiento a experimentos de inteligencia artificial, dentro de sus características claves se incluye la comparación entre diferentes modelos, la comparación busca establecer el mejor modelo enfocado a la predicción en entornos del sector eléctrico.

Tabla 5

Comparación de Características de Modelos de Predicción

Característica	ARIMA	Prophet	LSTM
Tipo de modelo	Estadístico	Aditivo	RNN
Complejidad de implementación	Baja, es un modelo simple y fácil de explicar	Baja, incluso para analistas sin con conocimientos básicos	Alta, requiere amplios conocimientos por la complejidad de las
Volumen de datos requerido	No requiere una gran cantidad de datos	Bajo rendimiento en series temporales cortas	Requiere grandes volúmenes de datos
Predicción en diferentes intervalos de tiempo	Adecuado para predicciones a corto plazo	Tiene buen rendimiento en predicciones a mediano y largo	Tiene buen rendimiento en corto, mediano y largo plazo.
Recursos computacionales	Tiene bajos requerimientos de ejecución	Requerimientos moderados en la ejecución	Requiere alto poder de cómputo
Aplicabilidad al sector eléctrico	Óptimos para análisis de consumos individuales con series estacionarias	Útil para trabajar con series estacionales	Óptimo para patrones complejos y múltiples entradas

Nota. Tomado de (Kutzkov, 2022)

Algoritmos Seleccionados

Por la versatilidad tanto para predecir series a corto y largo plazo, como la capacidad de introducir variables exógenas y su amplia utilización en sector eléctrico se considera que el modelo LSTM es una buena selección para implementaciones que busquen la predicción del consumo eléctrico de los clientes no regulados en el contexto del mercado eléctrico colombiano, este argumento también está fundamentado en el artículo científico (Caicedo & Alfonso, 2023) en el cual presenta un análisis de la implementación de un modelo LSTM en el contexto del sector eléctrico colombiano y se citan múltiples investigaciones donde además de otros modelos de predicción se presenta frecuentemente el uso del modelo LSTM, lo que sugiere que es ampliamente estudiado y utilizado en el sector eléctrico.

A continuación, se presenta la matriz de criterios seleccionados y por qué el LSTM es el algoritmo seleccionado en la investigación.

Criterios de Evaluación

1. Tipo de modelo: Se refiere a si el modelo es estadístico, aditivo o una red neuronal.
2. Complejidad de implementación: Nivel de dificultad de implementación del modelo.
3. Volumen de datos requerido: Cantidad de datos necesarios para que el algoritmo funcione de manera eficiente.
4. Predicción a largo, mediano y corto plazo: Capacidad para realizar predicciones en diferentes intervalos de tiempo.

5. Recursos computacionales: Requerimientos de recursos (poder de cómputo) para ejecutar el modelo.

6. Aplicabilidad al sector eléctrico: ¿Qué tan adecuado es el algoritmo para trabajar con datos del sector eléctrico?

Asignación de Pesos

- Tipo de modelo: 0.1
- Complejidad de implementación: 0.1
- Volumen de datos requerido: 0.1
- Predicción en diferentes intervalos de tiempo: 0.3
- Recursos computacionales: 0.1
- Aplicabilidad al sector eléctrico: 0.3

Evaluación de Cada Modelo

Se asignan puntuaciones de 1 a 5 a cada algoritmo para cada criterio, siendo 5 la mejor puntuación (es decir, la más adecuada para ese criterio).

Tabla 6

Evaluación de Algoritmos de Predicción

Criterios	ARIMA	Prophet	LSTM
Tipo de modelo (Estadístico=1, Aditivo=3, RNN=5)	1	3	5
Complejidad de implementación (Baja=5, Alta=1)	5	5	2
Volumen de datos requerido (No requiere gran cantidad=5, Requiere grandes volúmenes=1)	5	5	2
Predicción en diferentes intervalos (Un intervalo = 1, Dos intervalos = 2, Tres intervalos = 5)	1	3	5
Recursos computacionales (Bajos=5, Altos=1)	5	5	3
Aplicabilidad al sector eléctrico (Óptimo=5, Menos adecuado=1)	5	1	5

Cálculo de las Puntuaciones Totales

Multiplicamos las puntuaciones por los pesos asignados y sumamos los resultados para obtener la puntuación total de cada algoritmo.

Cálculo de la puntuación para ARIMA.

$$(1 \cdot 0.1) + (5 \cdot 0.1) + (5 \cdot 0.1) + (1 \cdot 0.3) + (5 \cdot 0.1) + (5 \cdot 0.3) = 3.4 \quad (4)$$

Cálculo de la puntuación para Prophet.

$$(3 \cdot 0.1) + (5 \cdot 0.1) + (5 \cdot 0.1) + (3 \cdot 0.3) + (5 \cdot 0.1) + (5 \cdot 0.3) = 3.6 \quad (5)$$

Cálculo de la puntuación para LSTM.

$$(5 \cdot 0.1) + (2 \cdot 0.1) + (2 \cdot 0.1) + (5 \cdot 0.3) + (3 \cdot 0.1) + (5 \cdot 0.3) = 4.2 \quad (6)$$

Tabla 7

Puntuación Final de Algoritmos de Predicción

Algoritmo	Puntuación Total
ARIMA	3.4
Prophet	3.6
LSTM	4.2

Interpretación

LSTM tiene el mayor puntaje impulsado por su amplia aplicabilidad en el sector eléctrico y que es muy versátil en los intervalos de tiempo en los cuales es efectivo, ya que se comporta bien a corto, mediano y largo plazo. Aunque tenga menor puntaje en los requerimientos de cómputo, su amplia utilización demostrada en las múltiples investigaciones demuestra que su grado de precisión es alto.

Prophet obtiene la segunda puntuación, aunque funciona bien con series estacionales y tiene un moderado consumo de recursos de cómputo, su baja referenciación en los estudios investigados demuestra que se debe explorar llevando a cabo investigaciones aplicadas para demostrar si se ajusta a las necesidades del mercado eléctrico.

ARIMA se ve afectado en su puntuación, ya que no es óptimo para diferentes rangos de tiempo y su implementación está basada en la estadística clásica, característica que le resta punto por los buenos resultados que han demostrado en los últimos años los modelos basados en técnicas de *machine learning*.

Conclusión del Algoritmo Seleccionado

LSTM es una opción fuerte, especialmente por su capacidad para manejar series complejas y predicciones a corto, mediano y largo plazo. requiere altos volúmenes de datos, pero estos son abundantes por los requerimientos de almacenamiento de las resoluciones vigentes en el sector eléctrico. Su consumo de cómputo es alto, pero las diferentes tecnologías en la nube o equipos con mejor rendimiento hacen que su uso se ajuste a los desarrollos tecnológicos del momento.

Resultados y Discusión del Modelo LSTM

La comparación de los modelos ARIMA, Prophet y LSTM propuestos permitió establecer las ventajas y desventajas de la aplicación de cada uno en la predicción de la carga eléctrica y puntualmente en la predicción de la demanda de clientes del mercado eléctrico no regulado. La siguiente argumentación está basada en el análisis ya realizado para cada modelo y de los argumentos planteados en el artículo de revista académica (Prasad & Kashappa, 2021). Aunque cada uno de los algoritmos tiene puntos a favor y en contra, las características del modelo LSTM se destacan para las series de tiempo asociadas al sector eléctrico, por otro lado, ARIMA es simple y requiere poco poder de cómputo, pero presenta una limitación importante y es que requiere series estacionarias por lo que la conversión de una serie estacional a estacionaria supone pérdida de información, este argumento es importante porque en general los comportamientos de los consumos eléctricos tiene componentes estacionales y de tendencia que se marcan en función de periodos cortos como la hora o incluso largos como un año. Por otro lado, el modelo Prophet aunque presenta características destacadas como un moderado uso de recursos de máquina y permite incluir eventos atípicos para ajustar la predicción, no permite la inclusión de múltiples variables exógenas que influyen de forma directa en la curva de demanda

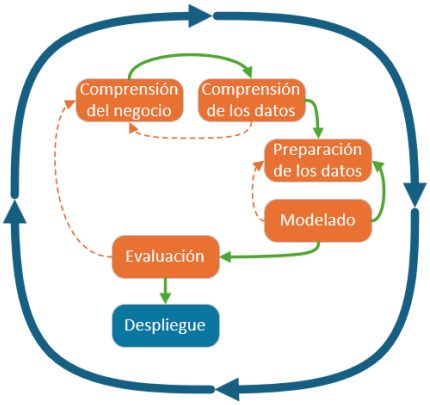
eléctrica, variables como condiciones climáticas, costos del servicio eléctrico o condiciones propias de la actividad comercial de los usuarios, por lo que en el análisis que realiza el autor identifica que las métricas MSE (Mean Squared Error) y RMSE (Root Mean Squared Error) en la implementación del modelo LSTM tienen un mejor rendimiento. El modelo LSTM (Long Short-Term Memory) aunque requiere mayores recursos de cómputo compensa esta desventaja con su versatilidad por la posibilidad de manejar series a corto, mediano y largo plazo, así como la capacidad de incorporar variables exógenas, y el alto volumen da datos requerido para el modelo no presenta una limitación, ya que se cuentan con grandes cantidades de información por la implementación del código de medida (CREG, 2014) que exige medir de forma remota y almacenar los consumos horarios de los clientes no regulados.

Metodología Propuesta

Se plantea una metodología con base en el estándar CRISP-DM que son las siglas en inglés de (Cross-Industry Standard Process for Data Mining). Según (Chapman et al., 2000) CRISP-DM es un modelo ampliamente utilizado en la minería de datos que busca proporcionar una estructura clara y sistemática para los proyectos en este ámbito. Este estándar consta de seis pasos, los cuales se describirán en función de las necesidades planteadas en la investigación y que se ilustran en la **Figura 5**.

Figura 5

Ciclo de Vida Metodología CRISP-DM



Comprensión del Negocio

En esta fase, es crucial entender las necesidades del negocio, enfocándose en el problema de investigación: las posibles inconsistencias en los reportes regulatorios de clientes no regulados. Estas inconsistencias pueden surgir debido a factores que afectan las lecturas de los medidores. Por ello, resulta necesario desarrollar un modelo de predicción del consumo de los clientes para comparar los valores reales con los valores estimados y determinar si las diferencias se encuentran dentro de rangos tolerables.

Comprensión de los Datos

En esta etapa, es esencial iniciar con la recopilación de datos y, posteriormente, llevar a cabo un análisis detallado para familiarizarse con ellos. Las series de tiempo relacionadas con el consumo eléctrico pueden variar en intervalos de tiempo o unidades de medida, por lo que resulta crucial comprender su estructura básica. Este entendimiento permitirá identificar anomalías y definir el enfoque más adecuado para su tratamiento.

Es importante destacar que esta fase puede retroalimentar a la etapa de comprensión del negocio. Esto sucede si se detecta que la fase inicial no planteó completamente las necesidades del negocio o si, por el contrario, dichas necesidades no son alcanzables con los datos disponibles. Este proceso iterativo asegura una mejor alineación entre los datos, los objetivos del negocio y el análisis a realizar.

Preparación de los Datos

Esta fase debe ser cuidadosamente planificada y ejecutada, ya que los diversos modelos de segmentación y predicción requieren estructuras de datos específicas. Dos aspectos que pueden impactar significativamente la modelización son, los datos faltantes, los cuales serán descritos a continuación: en primer lugar, los datos faltantes, si no se manejan adecuadamente, pueden interpretarse como ceros, lo que podría introducir sesgos en los modelos, por esta razón, es fundamental aplicar estrategias para rellenar los valores ausentes. Por otro lado, los datos atípicos también representan un desafío, ya que pueden afectar el rendimiento de los modelos. Por ejemplo, un pico de consumo histórico puede distorsionar las predicciones futuras, por lo que deben tratarse con precaución.

Además, en esta etapa es esencial seleccionar cuidadosamente las tablas, registros y atributos que estén directamente relacionados con el siguiente paso de la modelización, asegurando así la calidad y relevancia de los datos utilizados.

Modelización

En esta etapa, es fundamental seleccionar y aplicar tanto el modelo de segmentación como el modelo de predicción, basándose en las etapas previas de comprensión del negocio, comprensión y preparación de los datos. Esto implica identificar y utilizar algoritmos que aprovechen las características específicas de los datos, previamente analizados y preparados. La preparación incluye la limpieza y selección de datos según los requisitos y estructuras de los modelos.

Es posible que se requiera retroalimentar la fase de preparación si ciertos datos no cumplen con las estructuras necesarias o si alguna característica resulta poco significativa para la aplicación de los modelos. Por ello, esta fase se caracteriza por un enfoque iterativo, que busca garantizar la máxima calidad y relevancia de los datos antes de integrarlos en los modelos.

Evaluación

En esta fase, es fundamental evaluar si los modelos cumplen con los objetivos establecidos desde la etapa de comprensión del negocio.

Si los resultados no alcanzan las metas propuestas, será necesario realizar una revisión exhaustiva de todo el ciclo, comenzando desde la fase de comprensión del negocio, con el objetivo de identificar y corregir aspectos que requieran mejoras o ajustes en las diferentes etapas del proceso.

Dado que el ciclo de vida es iterativo, pueden surgir nuevos requerimientos en cualquier momento. Por ello, el proceso puede ser replanteado parcial o totalmente, adaptándose a los

nuevos objetivos establecidos desde la comprensión del negocio, lo que garantiza su alineación continua con las necesidades actuales.

Despliegue

El despliegue consiste en la presentación de la información generada, que puede variar desde una visualización básica en consola hasta su integración en una interfaz gráfica más sofisticada. Esta fase puede ser ejecutada tanto por el analista como por el usuario final, dependiendo de las necesidades específicas del negocio y del propósito del análisis.

Implicaciones para el Mercado Eléctrico Colombiano

La implementación de modelos que permitan predecir el consumo eléctrico de los clientes no regulados tiene implicaciones directas en el mercado eléctrico colombiano, ya que mejora la calidad de la información suministrada a las entidades regulatorias y permite garantizar procesos de facturación de alta calidad.

La implementación de los algoritmos de segmentación buscan reducir los errores en la predicción del consumo de los usuarios, con el fin de garantizar que las predicciones se ajusten lo mejor posible a los consumos reales, esta minimización del error puede permitir como ya se ha resaltado en otros apartados de la investigación anticipar factores que puedan afectar la calidad de los datos o generar alertas tempranas a los grupos de mantenimiento con el fin de solucionar aspectos técnicos que estén afectando el reporte regulatorio.

La implementación de los modelos también tiene un impacto positivo en las entidades regulatorias encargadas del pronóstico de la demanda del sistema eléctrico colombiano, ya que la identificación de los posibles errores en los reportes regulatorios por parte de los comercializadores garantiza que la información suministrada por estos sea de mayor calidad.

Los usuarios no regulados también pueden beneficiarse positivamente de la predicción de sus consumos, ya que la identificación de anomalías mediante la comparación entre los consumos registrados y los pronósticos permite garantizar que la energía facturada corresponda a la realmente consumida.

Conclusiones

La investigación analizó, mediante una revisión bibliográfica, diversas variables relevantes para la segmentación y predicción de la demanda eléctrica. Se realizó una revisión de modelos de segmentación, con el objetivo de identificar el que mejor se adapte a las necesidades de segmentar a los usuarios no regulados. Asimismo, se evaluaron modelos de predicción de series de tiempo, con el propósito de determinar el más adecuado para prever la demanda de usuarios no regulados y facilitar la identificación de anomalías en los reportes regulatorios.

En la investigación se identificaron las variables más citadas en las publicaciones revisadas como clave para mejorar el rendimiento de los modelos de predicción. Entre las más relevantes se destacaron los factores climáticos, las condiciones económicas, los consumos históricos y los días feriados, considerados elementos esenciales para optimizar la predicción de la demanda eléctrica.

Es importante señalar que, la mayoría de los estudios se enfocan en la predicción de la demanda para generación, transmisión y distribución, el número de investigaciones específicas sobre la predicción de la demanda en usuarios no regulados, particularmente en el contexto del mercado eléctrico colombiano, es limitado. Por ello, además del análisis bibliográfico, se realizó una identificación de las variables que influyen en la predicción de la demanda dentro del contexto eléctrico colombiano. Este análisis se complementó con una encuesta dirigida a profesionales del sector eléctrico en áreas relacionadas con predicción, clientes no regulados y reportes regulatorios, para fortalecer los hallazgos y contextualizarlos mejor. Dentro de las variables identificadas como relevantes y asociadas al sector eléctrico colombiano se puede resaltar los días feriados y jornadas laborales, los consumos históricos, factores climáticos, entre otras. Se destaca que se realizó una encuesta a profesionales del sector eléctrico asociados a roles

de predicción de la demanda y reportes regulatorios, permitiendo confirmar la relevancia de las variables propuestas. Sin embargo, se observó una diferencia notable en la evaluación de los contratos vigentes, que fueron considerados como una de las variables menos relevantes en comparación con otras. En contraste, los días festivos y las jornadas laborales se posicionaron como las variables con mayor importancia. Además, en la pregunta abierta, la variable demográfica emergió como una posible área de interés para investigaciones futuras, resaltando su potencial relevancia en el análisis de la segmentación y predicción de la demanda.

En la investigación se destacó la importancia del análisis exploratorio de datos como paso fundamental tanto para la segmentación de los usuarios como para la predicción de sus consumos. Este proceso permite identificar y tratar datos faltantes o atípicos que, de no gestionarse adecuadamente, pueden afectar negativamente el rendimiento de los algoritmos empleados.

En cuanto a la segmentación de usuarios no regulados, se llevó a cabo un proceso de selección del modelo más adecuado mediante una matriz multicriterio. En esta matriz se establecieron criterios y pesos específicos, ajustados al perfil particular de los clientes no regulados dentro del sector eléctrico colombiano. Los resultados del análisis determinaron que el algoritmo K-Means es el más apropiado para las necesidades de segmentación, dado su desempeño frente a las características distintivas de estos usuarios.

En la comparación de los modelos de predicción de series de tiempo en el entorno del sector eléctrico, se elaboró una tabla comparativa entre tres modelos de aprendizaje automático: ARIMA, Prophet y LSTM. Posteriormente, se utilizó una matriz multicriterio para determinar el modelo que mejor se ajusta a las necesidades de predicción de la demanda de los usuarios no regulados.

Los resultados de este análisis indicaron que el algoritmo LSTM ofreció el mejor desempeño, basándose en los pesos y criterios establecidos. Además, LSTM fue el modelo más referenciado en las investigaciones consultadas para la predicción de la demanda eléctrica, gracias a características destacadas como su adaptabilidad a diferentes rangos de tiempo y su capacidad para integrar múltiples variables, lo que lo convierte en una opción ideal para el contexto analizado.

Finalmente, se presentó la metodología CRISP-DM como guía estructurada para desarrollar un proyecto que implemente los componentes investigados. Esta metodología proporciona un esquema que permite implementar la segmentación de los usuarios no regulados y la predicción de sus consumos, permitiendo realizar comparaciones con los consumos reales, con el objetivo de identificar anomalías en los reportes regulatorios y garantizar la calidad de la información.

Recomendaciones

La implementación de modelos para la segmentación de usuarios es altamente recomendada, ya que permite realizar predicciones de demanda basadas en las características específicas de cada cliente. En el caso de los usuarios no regulados, las actividades comerciales presentan grandes diferencias en sus niveles y patrones de consumo, lo que hace fundamental adaptar los modelos a estas particularidades.

En lo que respecta a la identificación de variables, se han considerado aquellas comúnmente referenciadas en la literatura y el sector eléctrico. No obstante, esto no descarta la posibilidad de incorporar otras características relevantes de los usuarios que puedan enriquecer los modelos de segmentación y predicción.

La encuesta realizada evidenció que la mayoría de las variables propuestas fueron calificadas como relevantes por los participantes, con excepción de la variable correspondiente a los contratos vigentes, la cual podría ser considerada para su exclusión en estudios futuros o en procesos de reducción de dimensionalidad. Asimismo, en la pregunta abierta, donde se invitó a los encuestados a sugerir variables adicionales, se destacó la variable demográfica como una opción importante para ser incluida en futuros análisis, con el objetivo de evaluar su relevancia en el contexto estudiado.

Para fortalecer la investigación en el contexto colombiano, futuras exploraciones podrían incluir casos de implementación de modelos de predicción de demanda en diferentes empresas del sector eléctrico nacional. Además, sería interesante desarrollar estudios que aborden la problemática desde la perspectiva de los usuarios residenciales, quienes representan la mayor proporción de clientes en el sistema eléctrico colombiano.

Se recomienda contar con bases de datos bien estructuradas y robustas para la implementación de los modelos, garantizando así que los datos sean coherentes, de fácil acceso y cumplan con los estándares necesarios para su análisis.

Es igualmente importante realizar revisiones bibliográficas periódicas para mantenerse actualizado sobre las nuevas técnicas y avances en la predicción de series temporales. Cabe destacar que estas exploraciones no deben limitarse exclusivamente al sector eléctrico, ya que investigaciones de otros sectores pueden aportar perspectivas innovadoras y útiles para el desarrollo de nuevos enfoques en la predicción de la demanda eléctrica.

La investigación destacó una amplia gama de estudios que evidencian el creciente interés en la predicción de la demanda en el sector eléctrico. En este contexto, futuras investigaciones deben ser rigurosas y selectivas tanto en sus argumentaciones como en la elección de referencias relevantes, con el fin de mantener el interés del lector y evitar que la extensión del contenido resulte excesiva.

Referencias Bibliográficas

Andrade, N., & Castellanos, M. (2022). Modelo de pronóstico para la demanda de electricidad con un horizonte de tiempo de cinco años en el mercado regulado y no regulado de energía en Cali. [ICESI]. http://repository.icesi.edu.co/biblioteca_digital/handle/10906/95227

Armstrong, J. S. (2001). Principles of Forecasting: A Handbook for Researchers and Practitioners. Springer Science & Business Media.

Aziz, A., Mahmood, D., Qureshi, M. S., Qureshi, M. B., & Kim, K. (2024). AI-based peak power demand forecasting model focusing on economic and climate features. Frontiers in Energy Research, 12. https://doi.org/10.3389/fenrg.2024.1328891

Bishop, C. M. (2006). Pattern recognition and machine learning. Springer.

Caicedo, J. S., & Alfonso, W. (2023). Short-Term Load Forecasting Using an LSTM Neural Network for a Grid Operator. Energies, 16(23), Article 23. https://doi.org/10.3390/en16237878

Chapman, P., Clinton, J., Kerdber, R., & Khabaza, T. (2000). CRISP-DM 1.0: Step-by-step data mining guide. https://www.semanticscholar.org/paper/CRISP-DM-1.0%3A-Step-by-step-data-mining-guide-Chapman/54bad20bbc7938991bf34f86dde0babfbd2d5a72

Chung, J., & Jang, B. (2022). Accurate prediction of electricity consumption using a hybrid CNN-LSTM model based on multivariable data. PLOS ONE, 17(11), e0278071. https://doi.org/10.1371/journal.pone.0278071

Contreras, J., Espinola, R., Nogales, F. J., & Conejo, A. J. (2003). ARIMA models to predict next-day electricity prices. IEEE Transactions on Power Systems, 18(3), 1014-1020. IEEE Transactions on Power Systems. https://doi.org/10.1109/TPWRS.2002.804943

CREG. (1997). Resolución 108 de 1997 CREG. https://gestornormativo.creg.gov.co/gestor/entorno/docs/resolucion_creg_0108_1997.htm

CREG. (1998a). Resolución 70 de 1998 CREG. https://gestornormativo.creg.gov.co/gestor/entorno/docs/resolucion_creg_0070_1998.htm

CREG. (1998b). Resolución 131 de 1998 CREG. https://gestornormativo.creg.gov.co/gestor/entorno/docs/resolucion_creg_0131_1998.htm

CREG. (2002). Resolución 82 de 2002 CREG. https://gestornormativo.creg.gov.co/gestor/entorno/docs/resolucion_creg_0082_2002.htm

CREG. (2013a). Resolución 24 de 2013 CREG. https://gestornormativo.creg.gov.co/gestor/entorno/docs/resolucion_creg_0024_2013.htm

CREG. (2014, marzo 20). Resolución 38 de 2014 CREG [Estatal]. https://gestornormativo.creg.gov.co/gestor/entorno/docs/resolucion_creg_0038_2014.htm

CREG, P. (2013b). Estructura del Sector. Portal CREG. https://creg.gov.co/publicaciones/7819/estructura-del-sector/

CREG, P. (2022, diciembre 9). Energía Eléctrica. Portal CREG. https://creg.gov.co/publicaciones/15004/energia-electrica/

Dangeti, P. (2017). Statistics for Machine Learning (1.ª ed.). Packt Publishing Limited.

García, J. J., Gutierrez, A., Vargas Tobón, L., & Velasquez, H. (2017). Análisis económico del mecanismo de respuesta de la demanda en el sector eléctrico colombiano. http://hdl.handle.net/10784/11751

Gil, R., & Ignacio, A. (2024). Hourly electricity consumption forecasting for Antioquia-Colombia using statistical-machine learning models [Trabajo de grado - Maestría, Universidad Nacional de Colombia]. https://repositorio.unal.edu.co/handle/unal/86415

Gobierno de España. (2024). Guía Práctica de Introducción al Análisis Exploratorio de Datos en R | datos.gob.es. https://datos.gob.es/es/documentacion/guia-practica-de-introduccion-al-analisis-exploratorio-de-datos

Herrera, P., & Natán, G. (2021). Reconocimiento de patrones y pronóstico de consumo eléctrico. http://opac.pucv.cl/pucv_txt/txt-8500/UCD8531_01.pdf. http://repositorio.ucv.cl/handle/10.4151/92840

Huang, X., Lin, Y., Ruan, X., Li, J., & Cheng, N. (2023). Smart grid energy scheduling based on improved dynamic programming algorithm and LSTM. PeerJ Computer Science, 9, e1482. https://doi.org/10.7717/peerj-cs.1482

Iglesias, R. B., Barraqué, A. C., Bakx, G. E., & Izquierdo, S. K. (2018). Inteligencia artificial avanzada.

Jaimes, D. (2019). Desarrollo de un sistema de información para la verificación de conformidad de los sistemas de medición de las fronteras comerciales acorde al nuevo código de medida. https://repository.unab.edu.co/handle/20.500.12749/1507

Jain, A. K. (2010). Data clustering: 50 years beyond K-means. Pattern Recognition Letters, 31(8), 651-666. https://doi.org/10.1016/j.patrec.2009.09.011

Kuhn, M., & Johnson, K. (2013). Applied predictive modeling. Springer.

Kutzkov, K. (2022, septiembre 13). ARIMA vs Prophet vs LSTM for Time Series Prediction. Neptune.Ai. https://neptune.ai/blog/arima-vs-prophet-vs-lstm

Ley 142, de 11 de julio, régimen de los servicios públicos domiciliarios. (1994). *Congreso de la República de Colombia Diario Oficial 41.433 del 11 de julio de 1994*. https://www.funcionpublica.gov.co/eva/gestornormativo/norma.php?i=2752

Mitchell, T. (1997). Machine learning. McGraw-Hill.

Moreno, J., & Gutiérrez, J. (2019). Demanda de la energía eléctrica en Colombia, un modelo de pronostico [Trabajo de grado, Escuela Colombiana de Ingeniería Julio Garavito]. https://repositorio.escuelaing.edu.co/handle/001/1125

Mystakidis, A., Koukaras, P., Tsalikidis, N., Ioannidis, D., & Tjortjis, C. (2024). Energy Forecasting: A Comprehensive Review of Techniques and Technologies. Energies, 17(7), Article 7. https://doi.org/10.3390/en17071662

Prasad, A., & Kashappa, N. (2021). Electrical Load Forecasting using ARIMA, Prophet and LSTM Networks. ResearchGate, 9. https://doi.org/10.37391/IJEER.090404

Provost, F., & Fawcett, T. (2013). Data science for business: What you need to know about data mining and data-analytic thinking (First edition). O'Reilly.

Rojas, J. (2024). Pronóstico de demanda de energía eléctrica en un mercado de comercialización en Colombia. https://bibliotecadigital.udea.edu.co/handle/10495/40395

Román, A., López, M., & Pazos-Arias, J. J. (2021). Systematic Review of Electricity Demand Forecast Using ANN-Based Machine Learning Algorithms. Sensors, 21(13), Article 13. https://doi.org/10.3390/s21134544

Salazar, G., & Panchi, B. (2014). Análisis de la Evolución de la Demanda Eléctrica en el Ecuador Considerando el Ingreso de Proyectos de Eficiencia Energética. Revista Politécnica, 33(1), Article 1. https://revistapolitecnica.epn.edu.ec/ojs2/index.php/revista_politecnica2/article/view/218

Taylor, S. J., & Letham, B. (2017). Forecasting at scale (No. e3190v2). PeerJ Inc. https://doi.org/10.7287/peerj.preprints.3190v2

Torres, A. S. (2023). Predicción de la demanda de energía eléctrica usando modelos de inteligencia artificial para series temporales. https://bibliotecadigital.udea.edu.co/handle/10495/37564

Tukey, J. W. (1977). Exploratory data analysis. Addison-Wesley Pub. Co.

XM. (2024, octubre 2). En agosto, la demanda de energía en Colombia disminuyó -0.41 % en comparación con el mismo mes del año anterior | Portal XM. XM. https://www.xm.com.co/noticias/7179-en-agosto-la-demanda-de-energia-en-colombia-disminuyo-041-en-comparacion-con-el-mismo