

**Integración de series de tiempo y Random Forest para la predicción de disponibilidad de
activos eléctricos**

Juan Camilo Velasquez Molina

Cristhian Camilo Correa Ordoñez

Asesor

José Laureano Cruz Cardozo

Universidad Nacional Abierta y a Distancia UNAD

Escuela de Ciencias Básicas, Tecnología e Ingeniería ECBTI

Especialización en Ciencia de Datos y Analítica

2024

Agradecimientos

Agradezco a mi compañera de vida Laura Marcela por apoyarme en este proceso formativo y poder aportar con conocimientos aplicados a la sociedad

Camilo Correa.

Expreso mi más sincero agradecimiento a mi familia, cuyo apoyo incondicional ha sido fundamental en este proceso. A mi pareja, Valentina Martínez, por su paciencia, motivación y compañía en cada etapa de este camino. A mis amigos, por su respaldo y palabras de aliento. Finalmente, agradezco a Dios, por darme la fortaleza y sabiduría necesarias para alcanzar este logro. A todos ustedes, gracias por ser parte de este importante capítulo de mi vida.

Juan Camilo Velasquez.

Resumen

La planificación y gestión de los activos en los sistemas de distribución de energía eléctrica son esenciales para los operadores de red (OR), especialmente en un contexto de creciente demanda energética. El objetivo principal del proyecto es utilizar técnicas de aprendizaje automático, como Random Forest y series de tiempo, que permitan una mejor previsión del crecimiento de la demanda y de la vida útil de los activos eléctricos. El incremento demográfico y económico ha llevado a las empresas a enfrentar desafíos relacionados con la capacidad del sistema y la continuidad del servicio. Métodos convencionales de predicción, basados en datos históricos, no son siempre efectivos en este entorno dinámico, lo que puede resultar en decisiones no óptimas. Este trabajo enfatiza la necesidad de utilizar datos de demanda histórica (2021-2023) y registros de operaciones de activos (2019-2024) para generar pronósticos más precisos, a través del análisis de datos y la implementación de modelos predictivos, se busca proporcionar proyecciones fiables de la demanda para el periodo 2024-2026 y un mapa georreferenciado que muestre la vida útil de los activos eléctricos. Los objetivos específicos incluyen el preprocesamiento de datos para alimentar los modelos, la predicción de la demanda máxima en circuitos, y la estimación del ciclo de vida de los activos. Esta investigación se alinea con las normativas regulatorias, enfatizando la importancia de garantizar una buena calidad en el servicio de energía. Los resultados obtenidos podrían ser decisivos para que los OR optimicen la gestión de sus activos, reduzcan costos operativos, y mejoren el mantenimiento preventivo, en pro de un sistema eléctrico más eficiente y confiable, además de realizar cambios en activos que cumplieron su vida útil o repotenciándolos para obtener una mejor capacidad nominal.

Palabras claves: Vida útil, predicción de la demanda energética, Random Forest, LSTM

Abstract

Asset planning and management in electric power distribution systems are essential for network operators (ROs), especially in a context of increasing energy demand. The main objective of the project is to use machine learning techniques, such as Random Forest and time series, to enable better forecasting of demand growth and lifetime of electrical assets. Demographic and economic growth has led companies to face challenges related to system capacity and service continuity. Conventional forecasting methods, based on historical data, are not always effective in this dynamic environment, which can result in suboptimal decisions. This work emphasizes the need to use historical demand data (2021-2023) and asset operation records (2019-2024) to generate more accurate forecasts. Through data analysis and the implementation of predictive models, it seeks to provide reliable demand projections for the period 2024-2026 and a georeferenced map showing the useful life of electrical assets. Specific objectives include preprocessing data to feed the models, predicting peak demand on circuits, and estimating the life cycle of the assets. This research is aligned with regulatory standards, emphasizing the importance of ensuring a good quality of energy service. The results obtained could be decisive for ROs to optimize the management of their assets, reduce operating costs, and improve preventive maintenance, in favor of a more efficient and reliable electrical system, in addition to making changes in assets that have reached their useful life or repowering them to obtain a better rated capacity.

Keywords: Useful life, energy demand prediction, Random Forest, LSTM.

Tabla de Contenido

Introducción	9
Justificación	11
Objetivos.....	13
Objetivo General	13
Objetivos Específicos.....	13
Marco Conceptual y Teórico	14
Marco Referencial.....	14
Gestión de la Demanda.....	14
Utilización de Información de la Base de Datos	17
Aplicación de Métodos Estadísticos para Predecir el Crecimiento de la Demanda.....	17
Disponibilidad de Componentes Activos del Sistema Eléctrico de Distribución	18
Crecimiento de la Demanda por Subestaciones	20
Predicción de Vida Útil en Activos Eléctricos	21
Modelo LSTM.....	25
Estructura de una Celda LSTM.....	25
Modelo ARIMA	26
Pruebas ADF (Augmented Dickey-Fuller).....	27
Interpretación de Resultados	28
Modelo SARIMA (Seasonal ARIMA).....	28
Modelo SARIMAX	28
Adaptabilidad y Escalabilidad.....	30
Eficiencia Operacional y Planificación	30

Redes Neuronales Recurrentes	31
Árbol de Decisión.....	32
Modelo Random Forest.....	33
Modelo Holtwinters.....	34
Metodología	38
Resultados	41
Demanda Energética Vista desde Subestación Eléctrica	41
Distribución de Datos	41
Modelos de Predicción Utilizados	44
Modelo LSTM.....	44
Análisis de las Métricas de Evaluación.....	46
Observaciones Clave del Análisis Comparativo	48
Descripción del Gráfico de Cargabilidad.....	48
Activos de Red Eléctrica (Estimación de Vida Útil)	51
Distribución de Datos	51
Modelos de Predicción Utilizados Vida útil	56
Conclusiones	63
Referencias Bibliográficas	65

Lista de Tablas

Tabla 1 <i>Métricas de Evaluación Modelo LSTM.</i>	46
Tabla 2 <i>Resumen Estadístico Variables Categóricas Activos Eléctricos</i>	52
Tabla 3 <i>Resumen Estadístico Variables Enteras o Flotantes Activos Eléctricos</i>	52
Tabla 4 <i>Modelos Predictivos. Desempeño de Métodos para Encontrar la vida Útil del Activo.</i> 57	

Lista de Figuras

Figura 1 <i>Distribución de Valores de Potencia</i>	42
Figura 2 <i>Potencia Promedio a lo Largo de los Años</i>	43
Figura 3 <i>Predicción de la Demanda con el Modelo LSTM</i>	45
Figura 4 <i>Comparación de Modelos SARIMAX y LSTM</i>	47
Figura 5 <i>Cargabilidad del Circuito con los Datos Históricos</i>	49
Figura 6 <i>Cargabilidad Predicha</i>	50
Figura 7 <i>Matriz de Correlación Variables Operativas Mensuales Acumuladas</i>	54
Figura 8 <i>Gráfico de Barras, Comportamiento Histórico de Cantidades de Fallas y Mantenimientos</i>	55
Figura 9 <i>Gráfico de Torta, Relación de Cargabilidad de los Activos Instalados en el SDL</i>	55
Figura 10 <i>Activos de Red Reemplazados Mensualmente en Periodo 2019-2024</i>	56
Figura 11 <i>Equipos de Red Georreferenciado con una Estimación de Vida Baja</i>	58
Figura 12 <i>Gráfica Estimada de Vida de Útil Baja Activo 177</i>	59
Figura 13 <i>Equipos de Red Georreferenciados con un Estimado de Vida de Útil Media</i>	60
Figura 14 <i>Gráfica Estimada de Vida de Útil Media Activo 172</i>	60
Figura 15 <i>Equipos de Red Georreferenciados con un Estimado de Vida de Útil Alta</i>	61
Figura 16 <i>Gráfica Estimada de Vida de Útil Baja Activo 204</i>	61

Introducción

La planificación y gestión eficiente de los activos en el sistema de distribución de energía eléctrica representa un desafío constante para las empresas del sector. Para el operador de red (OR), el crecimiento demográfico y económico ha incrementado la demanda energética, lo que exige prever y preparar adecuadamente el sistema de distribución para evitar problemas de capacidad y garantizar un suministro continuo. Las técnicas convencionales de predicción basadas en datos históricos no siempre proporcionan la precisión necesaria para anticipar los cambios en la demanda de energía, especialmente en un entorno tan dinámico y complejo. Esto plantea el riesgo de subestimar o sobrestimar la demanda futura, lo que podría resultar en sobrecargas del sistema, interrupciones en el suministro y costos adicionales por inversiones urgentes e imprevistas. Por otro lado, los sistemas de distribución local (SDL) deben cumplir con ciertos límites en los indicadores de calidad del servicio que brindan a los usuarios. En Colombia, el objetivo es reducir gradualmente los valores de SAIDI (Indicador que mide la duración promedio de las interrupciones percibidas por un usuario) y SAIFI (Indicador que mide la cantidad promedio de interrupciones experimentadas por un usuario). Para alcanzar este objetivo, cada operador de red debe garantizar, por un lado, la inversión necesaria en la diversificación del sistema, de manera que su confiabilidad aumente mediante la incorporación de activos nuevos e inteligentes. Además, es crucial gestionar adecuadamente los equipos instalados para mantener los activos y asegurar la confiabilidad del sistema. Esto implica recopilar información detallada de cada activo eléctrico (como postes, aislamiento, crucetas y herrajes, transformadores, reconectores, seccionadores, seccionalizadores, cortacircuitos, interruptores, entre otros equipos que aseguran la continuidad del servicio), así como gestionar la asignación de recursos mediante alertas tempranas de intervención. Estas alertas deben basarse

en criterios de sensibilidad relacionados con la vida útil de los equipos, sus condiciones eléctricas, afectaciones, costo de intervención, impacto en los indicadores según la ubicación jerárquica y, finalmente, la estimación de su criticidad. En este contexto, el aprendizaje automático se presenta como una solución prometedora para mejorar la precisión de las predicciones sobre el crecimiento de la demanda y la vida útil de los activos. Sin embargo, la implementación de estas técnicas conlleva varios desafíos. En primer lugar, es esencial integrar y procesar adecuadamente grandes volúmenes de datos históricos, tales como solicitudes de conexión, demanda acumulada, registros de elementos de la red y novedades en el sistema de distribución eléctrica. En segundo lugar, es necesario seleccionar y aplicar métodos estadísticos y algoritmos de aprendizaje automático capaces de manejar esta complejidad, para proporcionar predicciones precisas y confiables.

Justificación

La correcta planificación y gestión del sistema de distribución de energía eléctrica es esencial para cualquier sistema eléctrico en crecimiento, como se da para el OR. El aumento constante en la demanda de energía debido al crecimiento demográfico y económico donde a su vez aumentan los elementos de la infraestructura, requieren de una capacidad de previsión robusta para evitar problemas de capacidad o límite energético o fallas en el sistema, con la finalidad de asegurar un suministro ininterrumpido. Las técnicas tradicionales de predicción basadas en datos históricos no siempre son adecuadas para capturar la complejidad y la dinámica de los datos de consumo energético y registros de eventos en el sistema, así como las características de cada activo, lo que puede llevar a estimaciones imprecisas y, en consecuencia, a decisiones no óptimas. La necesidad de una herramienta predictiva precisa es evidente ante la problemática actual de gestión y aprovechamiento de datos: sin una estimación adecuada del crecimiento de la demanda y aprovechamiento de los registros de los activos eléctricos, el operador de red (OR) corre el riesgo de enfrentar sobrecargas en el sistema, interrupciones en el suministro y costos elevados debido a inversiones urgentes e imprevistas. En este contexto, el aprendizaje automático ofrece una alternativa prometedora, ya que es capaz de manejar grandes volúmenes de datos y captar patrones complejos de los reportes de los activos y la dinámica de los consumos energéticos. Implementar técnicas de aprendizaje automático para predecir si un elemento puede fallar o generar una alerta, así como predecir el crecimiento del sistema de distribución local no solo promete una mayor precisión en las predicciones, sino que también mejora la capacidad de la empresa para planificar estratégicamente y tomar decisiones informadas a corto y mediano plazo. Esto incluye la identificación de áreas específicas que alimenta el prestador del servicio de energía, en el cual pueden requerir más atención y recursos,

basándose en un análisis detallado por subestaciones, circuitos, seccionadores y transformadores. Para el caso de la demanda proporcionar proyecciones fiables para el periodo 2024-2026, por otro lado, la predicción del ciclo de vida de los activos eléctricos genera el insumo para la intervención de los activos que están cerca de culminar su ciclo, por lo que generarían indisponibilidad del servicio, capturar la dinámica del ciclo de vida de cada activo y obtener el dato preciso de su punto probable de pérdida, es de vital utilidad para los grupos de mantenimiento y prevenir alteraciones en la disponibilidad del fluido eléctrico.

La utilización de los datos históricos de la demanda desde el año 2021 hasta el año 2023 como insumo de una aplicación de un sistema de aprendizaje automático que permita desarrollar modelos predictivos robustos, donde se compare el crecimiento de la demanda del circuito año tras año para proporcionar una visión detallada y precisa del futuro del sistema de distribución. Así mismo, para los activos eléctricos, se ejecutará un análisis sobre los activos de seccionamiento de red, debido a que cada uno de estos contiene un conjunto de transformadores y a su vez un grupo de usuarios, por lo que su afectación generaliza sectores y comunas de la zona de influencia del operador, se recopilan registros desde el 01 enero del año 2019 hasta el 01 octubre del 2024 para un total de 296 elementos de red

Objetivos

Objetivo General

Gestionar la disponibilidad técnica de los activos eléctricos asociados a un circuito de distribución (visto desde una subestación eléctrica), utilizando Random Forest y time series.

Objetivos Específicos

Realizar el preprocesamiento de la base de datos que alimentará los modelos de predicción.

Predecir el crecimiento de la demanda máxima presentada en un circuito para los próximos 3 años mediante técnicas de time series.

Pronosticar la vida útil de los activos eléctricos del circuito según sus características técnicas dadas sus condiciones operativas mediante Random Forest

Publicar un mapa georreferenciado con las predicciones gráficas de vida útil de los activos eléctricos para los próximos 2 años y de la cargabilidad anual para los próximos 3 años.

Marco Conceptual y Teórico

Marco Referencial

El desarrollo de modelos predictivos en la gestión de la demanda energética y el mantenimiento de activos eléctricos exige un análisis profundo de los datos históricos, tanto de la demanda de energía como del comportamiento de los equipos y sistemas que integran la infraestructura eléctrica. En este marco, se emplearán diversas herramientas de aprendizaje automático para abordar dos áreas clave: la predicción de la demanda eléctrica del operador de red para los próximos tres años y el análisis de la vida útil de los activos eléctricos de seccionamiento. Con el objetivo de crear un modelo predictivo que describa su vida útil en función de su condición operativa y comportamiento a lo largo del tiempo, en esta sección se abordarán los antecedentes y marco legal.

Gestión de la Demanda

Consiste en verificar las cargas para operar de manera óptima el sistema y obtener el máximo provecho de este; para ello se deberá tener en cuenta varios criterios de optimización relacionados con la curva de demanda, el mayor aprovechamiento de las redes de distribución, transformadores, y capacidades de las subestaciones. Uno de los principales objetivos de la gestión de la demanda es predecir la demanda, para lograr un perfil deseado. Además de anticipar la potencia máxima mensual y anual para poder establecer planes de contingencia en la prestación del servicio eléctrico. Con la implementación de metodologías para la gestión de demanda, existe canales directos e indirectos que cuyo objetivo principal es gestionarla, utilizando medidores inteligentes, incentivos tarifarios, consumo en periodos en la cual exista menor demanda; esto para los controles directos. Por otro lado, existen los controles indirectos,

como lo son la gestión de los usuarios de los consumos de sus aparatos eléctricos, usar algoritmos de optimización para predecir la demanda a corto y mediano plazo.

Pronosticar el consumo de energía, implica un ahorro en costos operacionales a los centros de generación y los operadores de red en la dinámica económica de compra y venta de energía; además permite visualizar el crecimiento energético, para verificar si el sistema de distribución requiere mayor robustez para atender este nuevo punto operativo. El pronóstico de demanda de energía eléctrica se define básicamente como la ciencia o el arte de predecir la carga futura en un sistema dado, para un periodo de tiempo futuro específico. Esta predicción puede ser desde unos minutos con fines de operación, hasta años con fines de planeación. (Instituto de Ingeniería Eléctrica, 2022)

Las redes neurales han sido utilizadas en los sistemas eléctricos de potencia (SEP) para que estos sean diseñados y operados de manera óptima, para ello se realiza el pronóstico de la demanda, la cual permite garantizar el flujo de potencia sea confiable; el uso de este método de aprendizaje, considera, el histórico de la demanda, de los días típicos, no típicos, y festivos; realizando un aprendizaje de estos y prediciendo el comportamiento del mismo; lo que se pretende con el actual proyecto de grado, es realizar dicho aprendizaje, adicionando, las solicitudes de factibilidad de servicio de energía; las cuales son los puntos de carga nuevos que brinda el operador de red, para que se conecten al sistema de distribución; con esta información adicional, no solo predecirá la demanda año a año; si no que se verificará hacia que zonas está creciendo la ciudad de Pereira; a través de estas solicitudes realizadas por los usuarios. (Valencia & Martínez, 2020)

El operador de red (OR), está regulado por la comisión de energía y gas (CREG), la cual vela que esta cumpla con las necesidades de los usuarios que este alimenta; por lo anterior, ha

dispuesto una cantidad considerable de resoluciones las cuales deben cumplir a cabalidad los OR; entre ellas, está la resolución CREG 015 de 2018; la cual habla de las inversiones que deben realizar los OR para asegurar una calidad optima e ininterrumpida del servicio del fluido eléctrico, además, de que el sistema de distribución local (SDL), tenga la capacidad de albergar nueva demanda; lo anterior, lográndolo a través de reposición de elementos obsoletos y que ya cumplieron su vida útil y además, de la instalación de activos eléctricos nuevos, que puedan atender la nueva demanda que entra a operar al SDL; estas inversiones se realizan al nivel de subestaciones, y también al nivel de las redes que se despliegan a lo largo de la zona urbana y rural la cual se encarga de alimentar a los nuevos usuarios. (Comisión Regulatoria de Energía y Gas [CREG], 2018)

Para el caso de los activos eléctricos, las actividades ejecutadas en años previos al 2019 no contienen data, no existía un plan de mantenimiento para las etapas de verificación del ciclo de vida del activo, por lo que para este trabajo se tiene como referencia los datos recopilados a partir de la fecha. Actualmente en algunas compañías utilizan hojas de datos mediante herramientas ofimáticas, en donde bajo una estructuración aprobada por el sistema de gestión de la compañía se registran los datos de las características de los activos, mantenimientos ejecutados entre observaciones del personal de campo. En la industria se tienen software para mantenimiento de subestaciones, sistemas industriales, redes de transmisión y distribución. Por ejemplo, se cuenta con Infraspark que es una de las soluciones más avanzadas de gestión del mantenimiento que ofrece una amplia gama de funciones, incluida la gestión centralizada de activos e inventarios, utiliza la inteligencia artificial para optimizar tus operaciones. De igual forma, existen en el mercado Fractal, Rosmiman entre otros softwares que guardan la esencia de las metas establecidas en el trabajo de grado.

Utilización de Información de la Base de Datos

El registro histórico de demanda de energía disponible abarca datos hora a hora desde el año 2021 hasta el año 2023. Este conjunto de datos se utilizará para implementar un modelo que interprete y analice el comportamiento registrado, con el objetivo de predecir la demanda futura en el circuito bajo análisis. Esto permitirá evaluar si dicho circuito tendrá la capacidad necesaria para soportar nuevas cargas que se deseen conectar al Sistema de Distribución Local (SDL) que lo alimenta. La demanda acumulada de los circuitos muestra un comportamiento característico a lo largo del tiempo. Este patrón puede ser aprovechado por el modelo a implementar, proporcionando suficiente información para aprender y predecir la demanda máxima mensual, así como su crecimiento año tras año.

Dado el gran volumen de datos históricos disponibles, el análisis eficiente de esta información resulta crucial para el entrenamiento de modelos de aprendizaje automático. Las redes neuronales artificiales (RNA) destacan como una herramienta efectiva en la predicción de la demanda eléctrica, gracias a su capacidad para identificar patrones complejos en los datos. Sin embargo, también existen otros modelos estadísticos que pueden ser empleados para realizar predicciones en conjuntos de datos de gran magnitud. (Arévalo & Hernández, 2023).

Aplicación de Métodos Estadísticos para Predecir el Crecimiento de la Demanda

El uso de métodos estadísticos para la predicción de la demanda eléctrica es crucial para la planificación y operación del sistema de distribución, además, para la planeación del crecimiento de la demanda, y el fortalecimiento del mismo sistema de distribución para que se prevea que este tenga la capacidad de soportar la nueva demanda, además de la verificación del estado de activos eléctrico que no cumplan con las condiciones técnicas operativas para abastecer dicha demanda; esto pensado en la expansión de nuevos circuitos y subestaciones que

alimentarán dicha carga. Para la implementación de la predicción se evaluarán los métodos más comunes que incluyen la regresión lineal, los modelos ARIMA (AutoRegressive Integrated Moving Average) y las redes Neuronales Artificiales (RNA). (Carrión, 2017)

Disponibilidad de Componentes Activos del Sistema Eléctrico de Distribución

El objetivo principal de la gestión del activo es asignar costes, tiempo y recursos de forma eficiente, garantizando la eficacia y adecuación de las actividades de mantenimiento. De este modo, se minimiza el sobrecosto de recursos y se evitan fallos inesperados de los equipos que podrían haberse evitado. En esencia, la gestión del mantenimiento adopta un enfoque proactivo, especialmente a través del mantenimiento preventivo, por lo que se apunta a reducir el número de actividades por ejecutar en mantenimientos correctivos. Este tipo de enfoque es común en muchos sectores, incluido el energético, donde el uso de inteligencia artificial y aprendizaje automático puede mejorar la asignación de recursos y la planificación de mantenimiento, como se ha demostrado en estudios previos (Munir, Vogel, & Jacobsson, 2022).

Para nuestro nicho de estudio en las redes de energía eléctrica, se cuenta con un gran número de elementos que se pueden categorizar bajo un esquema para no hacer dispendiosa la clasificación. Es preciso definir que los equipos instalados deben ser clasificados por su diferencia de potencial, ya que el sistema eléctrico colombiano se divide en niveles de tensión:

Nivel 4: Sistemas con tensión mayor o igual a 57.5 kV y menor a 220 kV.

Nivel 3: Sistemas con tensión mayor o igual a 30 kV y menor a 57.5 kV.

Nivel 2: Sistemas con tensión mayor o igual a 1 kV y menor a 30 kV.

Nivel 1: Sistemas con tensión menor a 1 kV.

El circuito evaluado en este proyecto únicamente referencia un nivel de tensión 2, por lo que el conjunto de elementos se reduce a:

Aislamiento. Comprende el elemento que aísla el potencial de tierra con la línea energizada, entre ellos se catalogan elementos de suspensión (linepost, pin) y retención (cadenas de aisladores, aisladores elastoméricos).

Herrajería y Crucetería. Comprende los elementos de fijación a estructura, los soportes de peso y forma de las diferentes configuraciones de la red, estas disposiciones se amplían en la norma de cada compañía, para nuestro caso en la norma del operador de red (OR).

Equipos de Protección. Elementos que sirven para aislar una falla definitiva en un elemento de red, entre las causadas por daño en elemento, rayos, aterrizajes, pérdida de aislamiento entre otras.

Interruptores. Elemento inteligente que contiene a través de un relé la información de variable del sistema eléctrico, actúa ante fallas transitorias y eventuales de mayor envergadura. Se comporta como la protección principal en una red de distribución.

Apoys de Media Tensión. Soporte físico de la red eléctrica con la finalidad de proporcionar distancias de seguridad a los seres vivos, naturales y seres vivos.

Con la clasificación de los elementos de red se proporciona una visión de los elementos asociados a los equipos de seccionamiento evaluados, por lo que su gestión dependerá del estado de cada uno de estos elementos en cuanto aplique, es decir que, un seccionamiento contiene un grupo de herrajes, crucetas aislamientos, protecciones, apoyos de media tensión distribuidos aguas abajo del equipo, al fallar uno de estos genera una condición operativa.

El sistema de gestión de activos, alineado con la normativa ISO 55000, juega un papel fundamental en la evaluación y la predicción de la vida útil de los activos eléctricos (Quiñones, Moncada, & Montenegro, 2022). En este contexto, el proyecto contempla la recopilación de datos históricos de operaciones, fallas y mantenimientos de activos eléctricos entre 2019 y 2024,

lo que permite obtener una visión más completa de su comportamiento a lo largo del tiempo. Este enfoque está relacionado con el trabajo de Karthick, Ravivarman y Priyanka (2024), que muestran cómo los enfoques basados en ML son esenciales para optimizar procesos predictivos, como la estimación de la vida útil de los componentes eléctricos.

Crecimiento de la Demanda por Subestaciones

La demanda eléctrica puede variar significativamente entre diferentes subestaciones debido a factores locales como la densidad de población, el tipo de consumidor, nuevas demandas (comerciales o industriales), y las condiciones socioeconómicas y climáticas; además, de las condiciones en las que se encuentran las redes de distribución que prestan el servicio, sus condiciones técnicas, y vida útil de la misma; por otro lado, las zonas de no interconexión que aún no se encuentran conectadas al sistema de distribución ya sea por una mala práctica en cuanto a la planeación energética a través de los diferentes circuitos alimentados por una subestación y la cual no podría suplir la necesidad energética de estas nuevas demanda. (Instituto de Ingeniería Eléctrica, 2022)

El análisis del crecimiento de la demanda por subestaciones permite identificar áreas de alta demanda y planificar mejoras en la infraestructura de distribución. Este enfoque segmentado ayuda a mejorar la precisión de las predicciones y a optimizar los recursos del sistema en el documento (Implementación de una Red Neuronal Artificial para la predicción de la Demanda Eléctrica a corto plazo) este tipo de análisis facilita la identificación de patrones de consumo específicos de cada subestación, lo cual resulta vital para la planificación y gestión eficiente del sistema de distribución. Por lo cual, para el análisis de este proyecto, es crucial tener el valor histórico, desde el año 2021 hasta el año 2023, de la demanda de la ciudad de Pereira; para revisar el comportamiento de esta. (Ramírez & Torres, 2021)

Predicción de Vida Útil en Activos Eléctricos

Este proyecto contempla una base de datos plana, con información desde el corte del año 2019 y 2024, cada operador de red (OR) contempla un registro de información diaria de operaciones, en esta bitácora se lleva el registro de novedades en los activos que están bajo su propiedad, por lo que esta información es valiosa para identificar las fallas, novedades y operaciones que se generan para los diferentes activos, con la finalidad de obtener una visión global, se generan reportes mensuales para cada elemento de red, por lo que se obtiene una dinámica a nivel de mes para los consolidados de operaciones, mantenimientos o novedades de modificaciones de red. Este conjunto de datos históricos indica la dinámica a lo largo de los años del comportamiento de cada elemento, lo que percibe y la evolución de su entorno, debido a que un aumento de carga a lo largo de los años indica un crecimiento en el número de usuarios residenciales, comerciales o industriales que alimenta, de igual forma se tienen registros del porcentaje de modernización en la red, así como la vegetación asociada, con este insumo pueden modelarse los activos que han fallado en el periodo de tiempo y de esta forma obtener el insumo necesario para estimar la vida útil de los activos según las condiciones operativas establecidas con mayor precisión. Tal como sugieren Karthick, Ravivarman y Priyanka (2024), el uso de enfoques basados en aprendizaje automático es fundamental para la optimización de procesos como la estimación de la vida útil de las baterías de vehículos eléctricos, donde se aprovechan datos históricos para predecir comportamientos futuros. De forma similar, en este contexto, las técnicas de aprendizaje automático aplicadas a los datos históricos de los activos de la red eléctrica pueden ofrecer predicciones precisas sobre las fallas y la vida útil de los activos, permitiendo una gestión más eficiente y sostenible de la infraestructura de distribución.

Antigüedad: El tiempo de uso del activo eléctrico se refiere al tiempo transcurrido en meses desde que fue puesto en servicio o instalado. Esta variable influye en la evaluación de su condición, rendimiento y necesidad de mantenimiento o reemplazo

Número de operaciones mensual: Indica el número de veces que el activo eléctrico (seccionamiento: cortacircuitos) es utilizado o activado durante un mes. Las operaciones de cada equipo están asociadas a fallos de tipo transitorio o permanente, para el caso de las interferencias de corta duración se asocian a caídas parciales de árboles u objetos que hacen contacto con la red momentáneamente, así mismo con arranques de fábricas o industrias que tengan fallo en aplicar correcciones de tipo inductiva o capacitiva para controlar las corrientes de arranque, afectando así la continuidad del servicio según las condiciones para cada caso, en contraste, las fallas permanentes están asociados a los eventos que generen una pérdida de aislamiento de la red eléctrica, reducción de distancias de seguridad o contacto respecto a una edificación o vegetación, pérdida mecánica en el tendido eléctrico entre otras novedades que vulneran la continuidad del servicio y que pueden ser extinguidos desde el equipo estudiado. Esta variable vista desde el enfoque de mantenimiento se define como la cantidad de mantenimientos correctivos.

Número de Mantenimientos Mensuales Ejecutados: Esta variable se refiere al número de intervenciones de mantenimiento realizadas en un activo eléctrico durante el mes. Los mantenimientos pueden ser preventivos o predictivos (se excluyen los correctivos debido a que se asocian en número de operaciones mensuales), estas actividades engloban el despeje de la red energizada respecto a la vegetación si tiene interferencia o con proximidad, cambio de aislamiento defectuoso, inspección visual de los componentes y verificación subjetiva de

condiciones físicas y mecánicas, adicionalmente, comprende la predicción de fallos mediante estudios termográficos, estudios de cortocircuito y ajustes en las protecciones.

Cargabilidad Máxima: Mide el porcentaje de la capacidad máxima de carga que un activo eléctrico puede soportar de manera segura, sin comprometer su integridad o generar riesgos operativos.

Número de Reemplazos: Cantidad de veces que un activo eléctrico ha sido reemplazado a lo largo de su vida útil, este contador está asociado a fallos, obsolescencia, o actualización de tecnología.

Porcentaje de Vegetación: El porcentaje de vegetación hace referencia a la proporción de áreas cercanas a las líneas de distribución que están en cercanía a la vegetación. Esta variable es un insumo para la planificación de tareas de poda y reducir riesgos de interrupciones en el servicio eléctrico debido a contacto con árboles o arbustos. Normalmente este valor es más alto en zonas rurales y es una causal que más frecuente en los fallos del sistema.

Relación Cable Ecológico/Desnudo: Proporción porcentual de cableado ecológico (con recubrimiento de aislamiento XLPE) frente al cable desnudo (sin aislamiento). Entre mayor sea este valor, se obtendrán ventajas en términos de seguridad y durabilidad, ya que reduce riesgos de cortocircuitos y mejora la eficiencia energética al tener mayor cantidad de conductor ecológico instalado en la red.

Latitud: Coordenada geográfica que indica la posición norte-sur de un punto en la Tierra

Longitud: Coordenada geográfica que señala la posición este-oeste de un punto en la superficie terrestre

Los modelos de predicción de la demanda energética desempeñan un papel fundamental en la gestión eficiente de los recursos en redes eléctricas, especialmente cuando se integran

fuentes de energía renovables. El despacho óptimo de energía, en el que se considera la variabilidad y la incertidumbre de la generación renovable, es un desafío clave en este ámbito. Fuentes como la solar y la eólica tienen una producción impredecible, lo que complica la tarea de planificar el suministro energético en períodos múltiples. En este contexto, los modelos de predicción basados en series de tiempo se utilizan para estimar la demanda futura de energía. Estas series permiten analizar patrones históricos de consumo y generar pronósticos que ayudan a optimizar la distribución y programación de la generación eléctrica. Al integrar estos modelos con la variabilidad de la generación renovable, se pueden minimizar costos, mejorar la eficiencia del sistema y garantizar una operación estable de la red eléctrica. (Torres-Riveros & Rivera-Rodríguez, 2018)

Redes Neuronales Artificiales: Las RNA son conocidas por su capacidad para modelar relaciones no lineales y complejas en los datos. Dado que la demanda eléctrica puede verse afectada por múltiples factores (como el clima, el crecimiento económico, y los patrones de consumo) que no siguen una relación lineal simple, las RNA son particularmente efectivas. Según la literatura revisada, las RNA pueden aprender y generalizar patrones complejos en los datos históricos de demanda eléctrica. (Valencia & Martínez, 2020).

Redes Neuronales Recurrentes: (RNN, por sus siglas en inglés) son un tipo de arquitectura de red neuronal diseñada específicamente para trabajar con datos secuenciales o temporales. A diferencia de las redes neuronales tradicionales, las RNN tienen la capacidad de recordar información de entradas anteriores, lo que las hace especialmente útiles para tareas donde el contexto y la secuencia son importantes. Están diseñadas para trabajar con datos secuenciales. A diferencia de las redes neuronales tradicionales, tienen "memoria", lo que

significa que tienen conexiones hacia atrás que les permiten tener en cuenta entradas previas en el procesamiento de la entrada actual. (Unite.AI, n.d.)

Modelo LSTM

El modelo LSTM (Long Short-Term Memory) son una mejora de las Redes Neuronales Recurrentes (RNN) diseñadas para superar sus principales limitaciones, como el problema del desvanecimiento del gradiente y la incapacidad de manejar dependencias a largo plazo en los datos. LSTM Introduce una estructura especial llamada célula de memoria que puede retener, actualizar o descartar información de manera controlada durante múltiples pasos de tiempo. La celda de memoria de una LSTM es controlada por tres "puertas": la puerta de olvido, la puerta de entrada y la puerta de salida. (Unite.AI, n.d.)

Estructura de una Celda LSTM

Puerta de Olvido (Forget Gate). Está puerta decide qué información de la celda de memoria anterior debe ser eliminada o "olvidada". Se utiliza una función sigmoide (activación), donde actúa para regular el flujo de la información en las puertas y toma como entrada el estado oculto previo y la entrada actual. La salida es un valor entre 0 y 1, donde 0 significa "olvidar todo" y 1 significa "recordar todo".

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (1)$$

Donde f_t es la salida de la puerta de olvido, σ es la función sigmoide $\frac{1}{1+e^{-x}}$, W_f son los pesos asociados a la puerta de olvido, h_{t-1}, x_t es la concatenación del estado oculto anterior y la entrada actual y b_f es el sesgo de la puerta de olvido

Puerta de Entrada (Input Gate): Decide qué información nueva debe ser almacenada en la celda de memoria. Hay dos pasos: una función sigmoide decide qué valores actualizar y una función tangente hiperbólica (tanh) crea nuevos valores candidatos que podrían agregarse.

$$i_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (2)$$

$$\widehat{C}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (3)$$

Donde: i_t es la salía de la puerta de entrada y la ecuación (3) es el nuevo candidato y función tangente hiperbólica, que normaliza el valor del candidato.

Actualización del Estado de la Celda: El estado de la celda se actualiza combinando la información retenida (puerta de olvido) y la nueva información (puerta de entrada).

$$C_t = (f_t \cdot [C_{t-1}] + i_t \cdot \widehat{C}_t) \quad (4)$$

Donde C_t es el nuevo estado,

Puerta de Salida (Output Gate): Decide qué parte del estado de la celda debe ser enviada como salida y usada como el nuevo estado oculto. Utiliza una función sigmoide para determinar qué partes del estado de la celda serán salidas y una función tanh para escalar la salida.

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (5)$$

$$h_t = o_t \times \tanh(C_t) \quad (6)$$

Donde o_t es la salida de la puerta de salida, h_t es el nuevo estado oculto y $\tanh(C_t)$ es la activación del estado de la celda.

Modelo ARIMA

ARIMA es un modelo estadístico utilizado para analizar y predecir series de tiempo no estacionarias. Donde una serie de tiempo se dice que es no estacionaria si sus propiedades estadísticas, como la media, varianza y autocorrelación, cambian a lo largo del tiempo; estas series suelen tener tendencias crecientes, decrecientes o cambios estructurales. El propósito de este modelo es hacer que los datos se vuelvan estacionarios mediante diferenciación para luego modelar la relación entre valores pasados y los errores residuales. (Hyndman & Athanasopoulos, 2018, p. X).

Donde ARIMA combina tres componentes principales para modelar series temporales.

AR(Autoregressive). El componente autoregresivo utiliza la relación entre una observación y su(s) observación(es) anterior(es). El parámetro p representa el número de términos de retraso (lags) incluidos. I (Integrated), este componente representa la diferencia entre observaciones para hacer que la serie sea estacionaria (sin tendencia). El parámetro d indica el número de diferencias aplicadas y MA (Moving Average); es el componente de media móvil utiliza la relación entre una observación y el error de predicción de las observaciones anteriores. El parámetro q controla cuántos términos de error retrasados se considera, se expresa como ARIMA (p, d, q), donde p es el orden autoregresivo, d es el grado de diferenciación, y q es el orden de la media móvil. (Ciencia de Datos, n.d.)

Una serie es estacionaria si la media no cambia con el tiempo, la varianza es constante a lo largo del tiempo y la autocorrelación no depende del tiempo. Para verificar su estacionariedad se debe realizar la prueba estadística ADF.

Pruebas ADF (Augmented Dickey-Fuller)

La prueba ADF (Augmented Dickey-Fuller) es una prueba estadística utilizada para determinar si una serie temporal es estacionaria. Es ampliamente utilizada en el análisis de series temporales para validar si los datos cumplen con los supuestos necesarios para modelos como ARIMA.

Hipótesis de la Prueba ADF

Donde la Hipótesis Nula H_0 , es la serie temporal no es estacionaria (tiene raíz unitaria) y la Hipótesis Alternativa, H_1 muestra que la serie temporal es estacionaria.

Interpretación de Resultados

Si el valor p es menor que 0.05, se rechaza la hipótesis nula y se concluye que la serie es estacionaria. Si el valor p es mayor que 0.05, no se puede rechazar la hipótesis nula, lo que indica que la serie no es estacionaria. (Hamilton, 1994).

Modelo SARIMA (Seasonal ARIMA)

El SARIMA extiende ARIMA al manejar series que presentan patrones estacionales. Donde los patrones estacionales ocurren cuando los datos presentan patrones cíclicos regulares que se repiten en intervalos fijos de tiempo (meses, trimestres, años, etc.). Esta extensión del modelo ARIMA introduce términos adicionales para capturar las fluctuaciones cíclicas observadas en intervalos regulares (mensuales, trimestrales, anuales, etc.). SARIMA agrega una capa adicional para capturar la estacionalidad en las series temporales. Esto se logra introduciendo términos estacionales en el modelo, que son análogos a los componentes ARIMA, pero aplicados a intervalos estacionales específicos. (Brockwell & Davis, 2016, p. X).

Modelo SARIMAX

El SARIMAX incorpora, además de los términos estacionales, variables exógenas que afectan a la serie de tiempo, debido a que relacionan factores externos a la serie de tiempo principal, pero que influyen significativamente en su comportamiento. Esto hace que el modelo sea robusto al integrar información externa, como publicidad o factores climáticos. Por lo que este modelo es una versión más avanzada de SARIMA que permite incluir regresores exógenos. (Chatfield, 2004, p. X).

El funcionamiento de SARIMAX sigue estos pasos:

Transformación de la Serie Temporal: Se aplica la diferenciación, si es necesario, para hacer que la serie sea estacionaria (es decir, que su media y varianza no cambien con el tiempo). Se incluyen términos estacionales, si se detecta un patrón estacional.

Construcción del Modelo: Se construye un modelo con los parámetros ARIMA (p,d,q) para capturar las relaciones entre los valores pasados y los errores de predicción.

Se incluyen componentes estacionales (P,D,Q,s) para capturar patrones de repetición estacional.

Si hay regresores exógenos disponibles, se añaden al modelo para considerar el impacto de factores externos

Ajuste del Modelo

El modelo se ajusta a los datos observados para determinar los valores óptimos de los parámetros que minimizan el error de predicción.

Predicción

Una vez ajustado, el modelo puede hacer predicciones sobre valores futuros de la serie temporal. Las predicciones toman en cuenta la influencia de valores pasados, patrones estacionales, y cualquier variable exógena incluida.

Ventajas del Modelo SARIMAX

- Modela Series Temporales Complejas: Puede capturar tanto patrones estacionales como tendencias a largo plazo.
- Flexibilidad: Permite incluir factores externos que influyen en la serie temporal, mejorando las predicciones.
- Interpretabilidad: Los parámetros del modelo son interpretables, lo que ayuda a comprender las relaciones subyacentes en los datos. (Ciencia de Datos, n.d.)

La metodología RNN aplicada con LSTM ofrece una mayor precisión en las predicciones a lo largo del tiempo y pueden capturar patrones complejos y SARIMAX puede proporcionar una base de comparación y validación, asegurando que las predicciones sean robustas y fiables.

Adaptabilidad y Escalabilidad

Los LSTM son adaptables y escalables, lo que significa que pueden ajustarse fácilmente a cambios en los datos de entrada y a nuevas variables que puedan surgir. Esto es crucial para la predicción de la demanda eléctrica en un entorno dinámico y cambiante como el de la ciudad de Pereira, donde factores externos pueden influir significativamente. Los modelos ARIMA es un método bien establecido que pueden ser fácilmente implementados y adaptados a diferentes conjuntos de datos. Su simplicidad y eficiencia computacional permiten una implementación rápida y efectiva.

Eficiencia Operacional y Planificación

Al utilizar estas metodologías, se puede lograr una predicción precisa de la demanda eléctrica, lo que resulta en un ahorro en costos operacionales y una mejor planificación de la infraestructura de distribución, además, de planificar los planes de inversión que se llevarán a cabo a lo largo de 3 años por el operador de red. La capacidad de predecir la demanda permite a los operadores de red optimizar el uso de los recursos existentes y planificar mejoras en la infraestructura de manera eficiente.

Debido a lo anterior, las Redes Neuronales Recurrentes y los modelos ARIMA se ajustan perfectamente al marco referencial del proyecto debido a su capacidad para capturar complejidades en los datos que se presenta a lo largo de un tiempo determinado, proporcionar precisión y validación cruzada, ser adaptables y escalables, tener un historial de éxito en

aplicaciones similares, y contribuir a la eficiencia operacional y planificación del sistema de distribución eléctrica. (Arévalo & Hernández, 2023)

Redes Neuronales Recurrentes

Las RNN se utilizan para predecir la demanda eléctrica a corto y mediano plazo. Estas redes requieren una adecuada selección y clasificación de las variables de entrada, como las del valor máximo mensual, desde el año 2021 hasta el año 2023, de la demanda de la ciudad de Pereira, de un circuito específico del sistema de distribución; para revisar el comportamiento de esta. El documento "Red Neuronal Artificial para la Predicción de la Demanda" detalla los pasos necesarios para la implementación de RNA en la predicción de la demanda, incluyendo el preprocesamiento de datos, la selección de la arquitectura de la red, y el entrenamiento y evaluación del modelo. Lo que se quiere llevar a cabo con RNN es aplicar un modelo más eficiente con series de tiempo, que pueda aprender de patrones que se presenten en los datos históricos y que se presenten valores escalables a lo largo del tiempo.

La aplicación de modelos estadísticos como SARIMAX es esencial para comparar y validar los resultados obtenidos por RNN. La combinación de estos enfoques proporciona una mayor precisión en las predicciones y una mejor comprensión de los factores que influyen en la demanda eléctrica. Los estudios presentados en la "Guía para la Predicción de la Demanda por Subestaciones" y otros documentos revisados sugieren que la integración de métodos estadísticos y de aprendizaje automático puede mejorar significativamente la exactitud de las predicciones de demanda.

En los registros asociados para cada activo de seccionamiento, se pretende predecir la vida útil de cada elemento según sus registros históricos de 6 años desde el año 2019, teniendo en cuenta de que se sabe cuáles son los activos que fueron cambiados a lo largo de este periodo,

se pueden extraer para verificar que tanta afectación aporta cada variable y su evolución a lo largo del tiempo, por lo que se propone un enfoque de cuenta regresiva de vida útil para cada elemento, es decir, que se descuenta la fecha de cambio o reemplazo respecto a la fecha registro, por lo que una vez llegue a la fecha de reemplazo, su valor de vida útil es cero y en caso contrario para su fecha de instalación corresponde al valor máximo que este elemento obtuvo.

Árbol de Decisión

Este modelo de aprendizaje divide los datos de entrada de manera recursiva, se estructura con nodos internos (entrada), en donde se inicia con un nodo raíz la toma de decisiones hasta las hojas (clasificación/predicción) de los ramales siguientes (límite en profundidad), para cada división se busca maximizar la homogeneidad dentro de cada subconjunto. El criterio de la división tiene como fundamento la condición de impureza de cada subconjunto, por lo que se utilizan las siguientes funciones para encontrar el índice de impureza $I(D)$, una vez este valor sea el mínimo, se tiene el resultado predictivo. (Munir et al., 2022)

Impureza de Gini: Se utiliza generalmente para clasificación.

Entropía: Otra métrica para la clasificación, basada en la teoría de la información.

Varianza: Se usa en problemas de regresión. :

Clasificación: El árbol toma decisiones dividiendo el espacio en regiones de manera recursiva utilizando un criterio impureza de Gini. Por lo que se define que, para un nuevo dato, un árbol que predice una clase. Finalmente se predice la clase que recibe la mayoría de los votos de los árboles, donde $I(\gamma_i = c)$ es una función indicadora que toma el valor 1 si el árbol T_i predice la clase C , y 0 en caso contrario.

$$\gamma = \arg \max(\sum_{i=1}^{N_{trees}} I(\gamma_i = C)) \quad (7)$$

Donde N_{trees} = Número de árboles y C = Clase

Predicción. El árbol toma decisiones dividiendo el espacio en regiones de manera recursiva utilizando un criterio de varianza, el modelo suma el promedio de las predicciones de cada uno de los árboles.

$$\gamma = \frac{1}{M} (\sum_{m=1}^M (\gamma_i)) \quad (8)$$

Siendo M el número total de árboles en el bosque

Support Vector Machine: Intenta encontrar un hiperplano óptimo que separe los datos de diferentes clases con el mayor margen posible, es flexible para trabajar en situaciones lineales y no lineales gracias al uso de kernels. Se han implementado en estudios para determinar si un activo de transformador requiere o no de cambio. Quiñones, L. I. A., Moncada, C. A. L., & Montenegro, D. A. B. (2022)

Modelo Random Forest

Esta modelo contiene la esencia de toma de árbol de decisión (descrito en concepto previo) o votación de un cantidad de individuos (múltiples árboles de decisión) que generan una asociación por las diferentes clases para el caso de clasificación o una predicción para una regresión, la metodología del algoritmo selecciona aleatoriamente un subconjunto de características del conjunto total de características disponibles, en donde cada árbol se entrena con una muestra aleatoria (con reemplazo) de los datos original (Bootstrap Aggregating) con la finalidad de cada uno de estos prediga una clase y la última se determine por la mayoría de votos entre todos los árboles para el caso de clasificación, en este documento se implementó la regresión y contiene la misma dinámica, lo que traduce que cada árbol predice un valor numérico y el valor final es la media o la mediana de todas las predicciones de los árboles, por lo que se puede describir mediante la combinación de varios árboles de decisión (ecuación 8) y la agregación de sus predicciones.

$$f(x) = \frac{1}{M} (\sum_{m=1}^M (f_m(x))) \quad (9)$$

donde $f_m(x)$ es la predicción del m-ésimo árbol y M es el número total de árboles en el bosque.

Modelo XGBoost. (Extreme Gradient Boosting): Se fundamenta en el concepto de árbol de decisión en donde estos se construyen secuencialmente (uno a la vez), corrigiendo los errores cometidos por los árboles anteriores mediante la métrica de residuo, esto se logra mediante la optimización con Gradiente descendente, en cada iteración, se ajusta el árbol para minimizar el error de predicción (\hat{y}_i) con base en la función de pérdida (error cuadrático medio) el modelo realiza una combinación ponderada de todas las predicciones previas para dar un resultado. (Karthick et al., 2024)

$$\hat{y}_i = \sum_{m=1}^M f_m(x_i) \quad (10)$$

Donde $f_m(x_i)$ es la predicción del m-ésimo árbol

Modelo Holtwinters

El modelo de holtwinters utiliza el método de suavizamiento exponencial para pronosticar series temporales, tiene tres componentes principales: nivel (nivel de la serie), tendencia (cambió en el nivel de la serie) y estacionalidad (patrón repetitivo dentro de la serie temporal). Este modelo cuenta con dos variantes para la predicción de la serie:

Holtwinters aditivo: La amplitud de la estacionalidad es constante a lo largo del tiempo. Es decir, las fluctuaciones estacionales no dependen del nivel de la serie temporal. Las propiedades del modelo se describen las expresiones matemáticas.

Nivel:

$$I_t = \alpha(y_t - s_{t-m}) + (1 - \alpha)(l_{t-1} + b_{t-1}) \quad (11)$$

Tendencia:

$$b_t = \beta(l_t - l_{t-1}) + (1 - \beta)Ib_{t-1} \quad (12)$$

Estacionalidad:

$$S_t = \gamma(y_t - l_t) + (1 - \beta)s_{t-m} \quad (13)$$

Pronóstico:

$$\gamma_{t+h} = l_t + hb_t + s_{t+m} \quad (14)$$

Donde l_t nivel en el tiempo t , y_t es el valor observado en el tiempo t . S_t Estacionalidad en el tiempo, b_t tendencia en el tiempo t , por otro lado, α , β , γ son los parámetros de suavizado entre (0 y 1) y m es la longitud del ciclo estacional

Holtwinters multiplicativo. La amplitud de la estacionalidad depende del nivel de la serie temporal, es decir, cuando los efectos estacionales son más pronunciados cuando los valores son más altos (Mor et al., 2018). EL modelo describe las siguientes características.

Nivel:

$$I_t = \alpha \left(\frac{y_t}{s_{t-m}} \right) + (1 - \alpha)(l_{t-1} + b_{t-1}) \quad (15)$$

Tendencia:

$$b_t = \beta(l_t - l_{t-1}) + (1 - \beta)Ib_{t-1} \quad (16)$$

Estacionalidad:

$$S_t = \gamma \frac{y_t}{l_t} + (1 - \gamma)s_{t-m} \quad (17)$$

Pronóstico:

$$\gamma_{t+h} = l_t + hb_t + s_{t+m} \quad (18)$$

Modelo de depreciación exponencial. Se utiliza para calcular la depreciación de un activo a lo largo del tiempo, el valor del activo disminuye de manera exponencial, es decir, más rápido en los primeros años y más lentamente en los años posteriores (Tarasov, 2022). Se describe mediante la expresión:

$$v(t) = \gamma(t) \cdot e^{-\lambda t} \quad (19)$$

Donde $v(t)$ define la vida útil, $\gamma(t)$ la función que contiene los pesos para cada instante de tiempo y $e^{-\lambda t}$ la tasa de depreciación exponencial según el tiempo transcurrido desde que se adquirió el activo.

Modelo Holtwinters - depreciación exponencial. Este enfoque permite capturar patrones históricos de tendencias y estacionalidades mediante series temporales: Holtwinters, sin embargo, es necesario complementar esta dinámica con una restricción de tiempo para evitar un comportamiento irreal del ciclo de vida útil del activo eléctrico, por lo que se propone argumentar la variabilidad del ciclo de vida con una depreciación exponencial decreciente, modelar directamente el decaimiento inherente bajo el comportamiento de las variables que alimentan el modelo (operaciones, mantenimientos, porcentaje de vegetación entre otras), por lo que se parametriza según su condición operativa actual e histórica con una depreciación natural de los activos eléctricos.

En la ecuación 19, se agrega una componente adictiva con el modelo Holtwinters debido a que las fluctuaciones estacionales no dependen del nivel de la serie temporal para nuestro caso se genera:

$$v(t) = (l_t + hb_t + S_{t+m}) \cdot e^{-\lambda t} \quad (20)$$

Para el caso de la tasa de depreciación λ , varía en cada cambio de estado según la variación de predicción del modelo Holtwinters, adicionalmente, se contempla que los activos siempre llegarán a una vida útil cercana a cero (0.01), se propone encontrar cada tasa respecto a esta referencia, su condición predicha (ecuación 14) y la antigüedad (T) de cada activo, según se desarrolla en ecuación 20 y se generaliza en 21.

$$\text{Sea: } 0.01 = \gamma_t \cdot e^{-\lambda(T)t} \quad (21)$$

$$\text{se despeja } \lambda = - \frac{\ln\left(\frac{0.01}{\gamma_t}\right)}{T} \quad (22)$$

γ_t Es el resultado del modelo holtwinters, λ Tasa de depreciación exponencial y T es la diferencia entre la edad y el tiempo transcurrido.

Metodología

Preparación de Datos

Se realizará la selección de las potencias máximas mensuales como variable objetivo para el modelo. Se dividen los datos en conjuntos de entrenamiento y prueba y se realizará la normalización de las variables para optimizar el desempeño del modelo que se aplicará

Implementación del Modelo LSTM

Para la configuración del modelo Long Short-Term Memory (LSTM) se procesará datos históricos y realizar predicciones a futuro. Se entrenará el modelo con datos históricos de potencias máximas desde 2021 al 2023 para estimar las potencias máximas mensuales proyectadas para los próximos 36 meses, desde el 1 de enero de 2024, lo anterior realizando lo siguiente:

Olvido de Información: La celda decide qué información del estado anterior debe olvidarse mediante la puerta de olvido. (ecuación 1)

Actualización de la Información: La puerta de entrada determina qué nueva información debe almacenarse en la celda. (ecuación 4)

Actualización del Estado de la Celda: La celda actualiza su estado combinando la información retenida y la nueva información. (ecuación 5)

Generación de Salida: La puerta de salida decide qué información debe emitirse como salida para el paso actual y se convierte en el estado oculto para el próximo paso. (Unite.AI, n.d.) (Ecuación 6)

Donde se verificará una de las ventajas relevantes del modelo LSTM el cual es capaz de recordar información durante periodos prolongados, lo que lo hace útil para tareas de series temporales y secuencias.

Evaluación del Modelo

Se evaluará el modelo donde se calculará las métricas de desempeño para verificar si lo aplicado funcionará para la predicción de la demanda y si representa de manera efectiva el comportamiento de la demanda y representa está a través de una escalabilidad en el tiempo. Las métricas que evaluará el modelo son:

RMSE (Root Mean Squared Error): Para evaluar la precisión general del modelo.

R² Score (Coeficiente de Determinación): Para medir qué tan bien las predicciones se ajustan a los datos reales.

Se comparará los modelos explicados en el marco teórico para verificar como es su representación del histórico de la demanda y su escalabilidad a través del tiempo que se desea estudiar para realizar la predicción de la demanda.

Cargabilidad Presente: Se determinará el porcentaje de utilización de la capacidad máxima del sistema en función de las potencias máximas históricas. Comparar estos valores con la capacidad nominal del sistema (por ejemplo, 6.53 MVA).

Cargabilidad Futura: Se utilizará las predicciones del modelo LSTM para calcular el porcentaje de utilización de la capacidad máxima del sistema, para los nuevos periodos de tiempo, generar un gráfico que proyecte la evolución del uso de la capacidad en función de las tendencias de potencia máxima estimadas e identificar puntos críticos, con la finalidad de evitar que exceda la capacidad actual del sistema, formulando estrategias para ampliar la capacidad del sistema o redistribuir cargas y proponer líneas de investigación adicionales para refinar el modelo predictivo e incluir variables complementarias.

La vida útil de los activos eléctricos se determinará por los activos que tengan registros de falla que genere un reemplazo a lo largo de su existencia operativa, para caracterizar este

comportamiento se evaluarán los modelos Random Forest (ecuación 7), XGBoost (ecuación 8) y SVM. El insumo para el entrenamiento de estos modelos se ejecutará con el 70% de los registros históricos depurados (preprocesamiento de datos, para este estudio no se utiliza reducción de variables por PCA para no perder interpretabilidad en el modelo a ejecutar de la totalidad de registros acumulados que contenga un activo), hasta su reemplazo, en contraste de validación del modelo con el complemento del 30% de los datos. La elección del modelo que mejor se ajuste estará dado por el desempeño del R2 (el valor más cercano a 1) y el menor valor de la desviación estándar del tiempo en meses de vida útil del activo. Una vez se obtiene un modelo que describa la caracterización de los datos, se entabla una extrapolación del modelo a los activos que no han tenido falla para el periodo evaluado, por lo que se ingresan los datos al modelo de predicción con los registros de las condiciones operativas de los elementos que no tienen una fecha de reemplazo. Esta metodología permite conocer la vida útil restante de los activos que han fallado o no, con base en el comportamiento histórico de los activos que presentaron una falla definitiva. Adicionalmente, se implementa un modelo a través de series de tiempo que indique el comportamiento de la vida útil de cada activo en función de la vida útil predicha por el modelo previo, por lo que el nuevo modelo holtwinters - depreciación exponencial descrito en la ecuación (18) generará la dinámica hasta el 2026 de la variación de la vida útil del activo.

Resultados

Demanda Energética Vista desde Subestación Eléctrica

Los valores de referencia para el sistema de potencia evaluado tienen como centro de distribución la subestación eléctrica, por lo que la dinámica de los resultados del comportamiento de la demanda energética se encuentra acotada desde este nodo eléctrico e indica las restricciones principales del sistema en su nivel global, los valores de potencia representan la variable de flujo que caracteriza el sistema para los diferentes instantes del día, para el estudio realizado la estampa de tiempo se tomaron valores máximos mensuales desde el año 2020 al 2023 con unidades de potencia aparente Mega Voltios Reactivos.

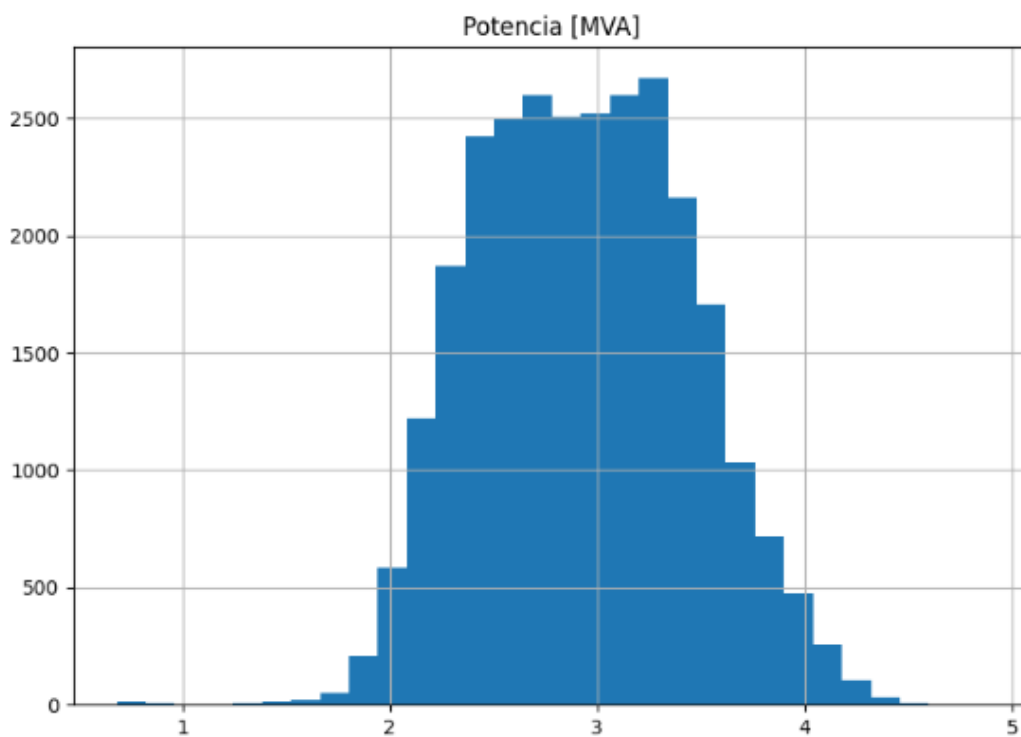
Distribución de Datos

En el marco del proyecto, se llevó a cabo un análisis exploratorio de la base de datos seleccionada, con el objetivo de comprender las características de las variables y su comportamiento, así como identificar posibles patrones y relaciones que puedan influir en el desarrollo del modelo propuesto.

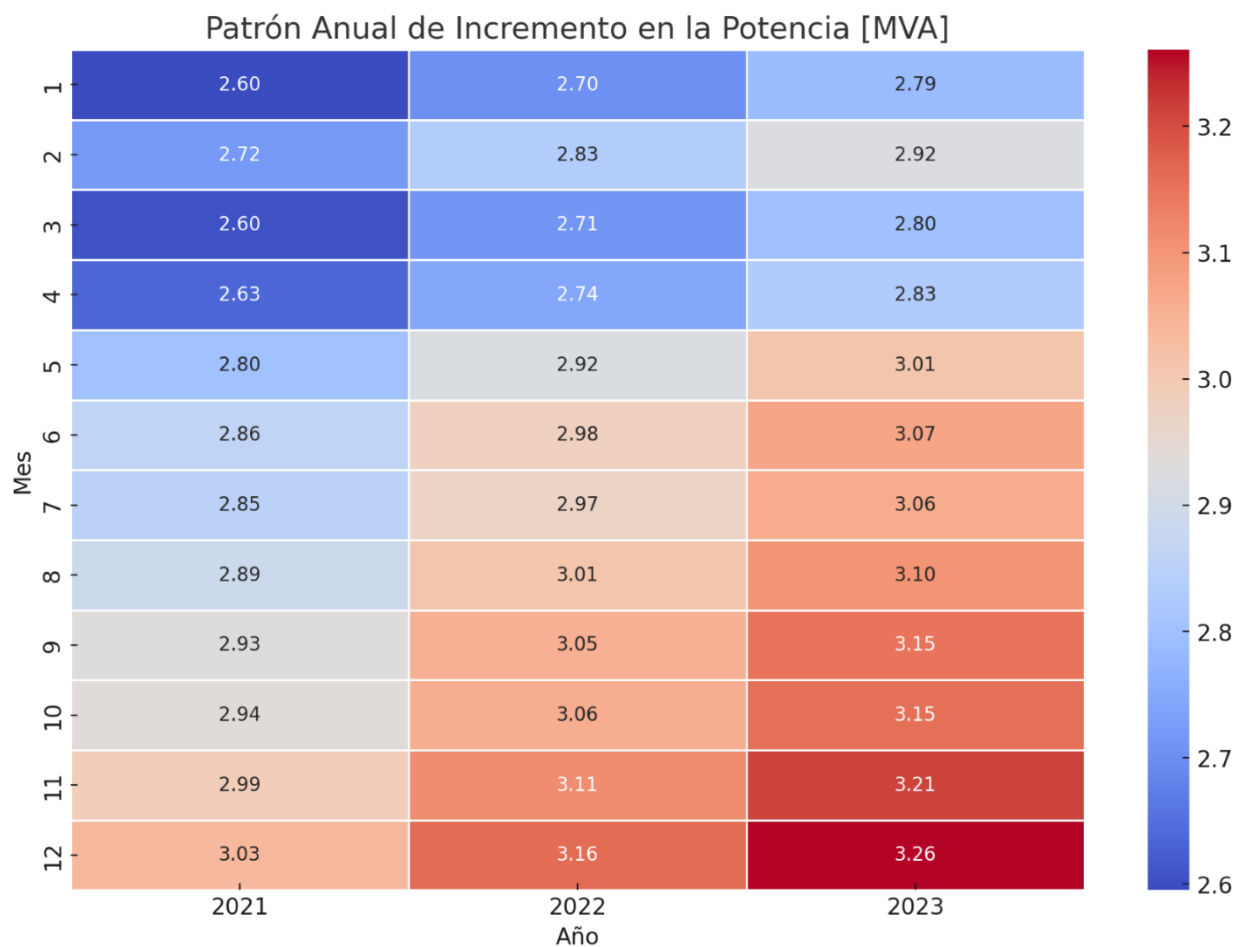
Se examinaron tanto las variables numéricas como las categóricas, asegurando la integridad de los datos y evaluando su distribución. Este análisis inicial permitió detectar posibles valores atípicos, datos faltantes y errores en el registro de las variables.

Figura 1

Distribución de Valores de Potencia



En la figura 1 se muestra un histograma para las variables numéricas, lo que permitió evaluar su distribución normal, proporciona una vista clara de la distribución de frecuencias de la variable Potencia [MVA]. Donde la distribución completamente normal (gaussiana), se observan agrupamientos en ciertos rangos de valores, lo que indica que la mayoría de los datos se concentran en un intervalo específico.

Figura 2*Potencia Promedio a lo Largo de los Años*

En la figura 2 se puede observar la variación mensual del promedio de Potencia [MVA] a lo largo de los años. Los colores más cálidos indican un incremento en los valores de potencia. Se evidencia patrones observados con un incremento recurrente en la potencia promedio, donde este patrón se evidencia año a año, siendo más contundente para el año 2023. Esto se evidencia una clara estacionalidad en los datos presentados en la base de datos de potencia.

Modelos de Predicción Utilizados

Se evalúan modelos predictivos para la descripción futura de la carga energética mediante modelos supervisados con la finalidad de ajustar con precisión el comportamiento para los próximos años bajo el criterio de estacionalidad y error cuadrático medio

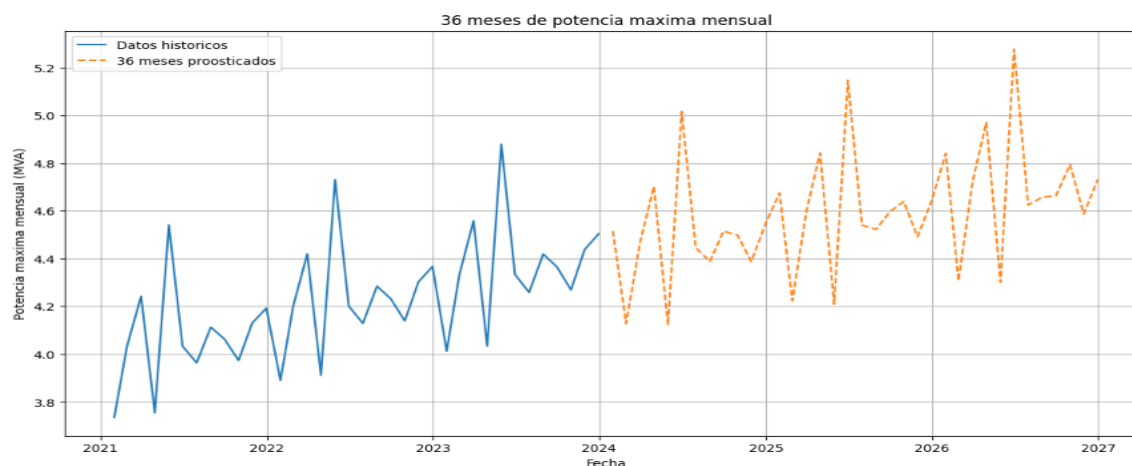
Modelo LSTM

En este proyecto, se implementó un modelo de redes neuronales tipo LSTM, donde se la base de datos presentada se ajusta a los supuestos requeridos para emplear este modelo, donde la información presenta datos organizados de manera secuencial, con marcación temporal, cantidad de datos suficientes para que el modelo aprenda dependencias temporales y evitar que tenga datos nulos. Se ajustaron los parámetros del modelo, incluyendo el número de neuronas en las capas LSTM, la tasa de aprendizaje y el tamaño de lote, para maximizar su capacidad predictiva.

El modelo se entrenó iterativamente, empleando un conjunto de datos preprocesados que incluye normalización de las variables de entrada para mejorar la estabilidad del entrenamiento; este conjunto de datos históricos sirvió para que el modelo obtuviera las predicciones mensuales desde enero de 2024 hasta diciembre de 2026. Donde los valores obtenidos representan las potencias máximas estimadas, destacando fluctuaciones esperadas debido a factores estacionales o tendencias subyacentes observadas en el conjunto de entrenamiento.

Figura 3

Predicción de la Demanda con el Modelo LSTM



En la figura 3 se evidencia una línea sólida que representa los datos históricos de la potencia máxima mensual. Esta línea muestra cómo ha evolucionado la potencia máxima del circuito a lo largo del tiempo. Mientras que, la línea discontinua representa las predicciones generadas por el modelo LSTM para un período de 36 meses en el futuro.

Para este grafico se puede evidenciar que el modelo LSTM intenta capturar y prever patrones a partir de los datos pasados, incluyendo tendencias, estacionalidad y posibles fluctuaciones. Por lo tanto, la predicción busca proyectar cómo evolucionará la potencia máxima mensual en base a estos patrones aprendidos. También, se puede observar que si la línea de predicción LSTM sigue de cerca la tendencia general de los datos históricos, indica que el modelo ha capturado correctamente los patrones subyacentes en los datos; por último, En caso de que la predicción LSTM muestra variaciones y picos que coinciden con patrones cíclicos o estacionales presentes en los datos históricos, esto es una señal de que el modelo ha logrado modelar con precisión la naturaleza no lineal o cíclica de los datos históricos de las potencias máximas.

Análisis de las Métricas de Evaluación

Tabla 1

Métricas de Evaluación Modelo LSTM

Root Mean Squared Error (RMSE):	R ² Score
: 0,01339869	0,97099558

En la tabla 1 se evidencia un RMSE con un valor de 0.0134, esto sugiere que, en promedio, las predicciones del modelo LSTM tienen un error muy pequeño en comparación con los valores observados. Esto indica que el modelo tiene un alto nivel de precisión en sus predicciones. El R² Score indica qué tan bien las predicciones del modelo se ajustan a los datos reales. Un valor de 0.9710 implica que el modelo puede explicar aproximadamente el 97.1% de la variabilidad de los datos. Este es un indicador de que el modelo LSTM ha capturado de manera muy precisa la relación entre las variables en la serie temporal.

Las métricas muestran que el modelo LSTM tiene un desempeño sobresaliente en la predicción de la potencia máxima mensual. El bajo RMSE indica que las predicciones están muy cerca de los valores reales, y el alto valor de R² confirma que el modelo ha capturado de manera precisa la relación entre las variables.

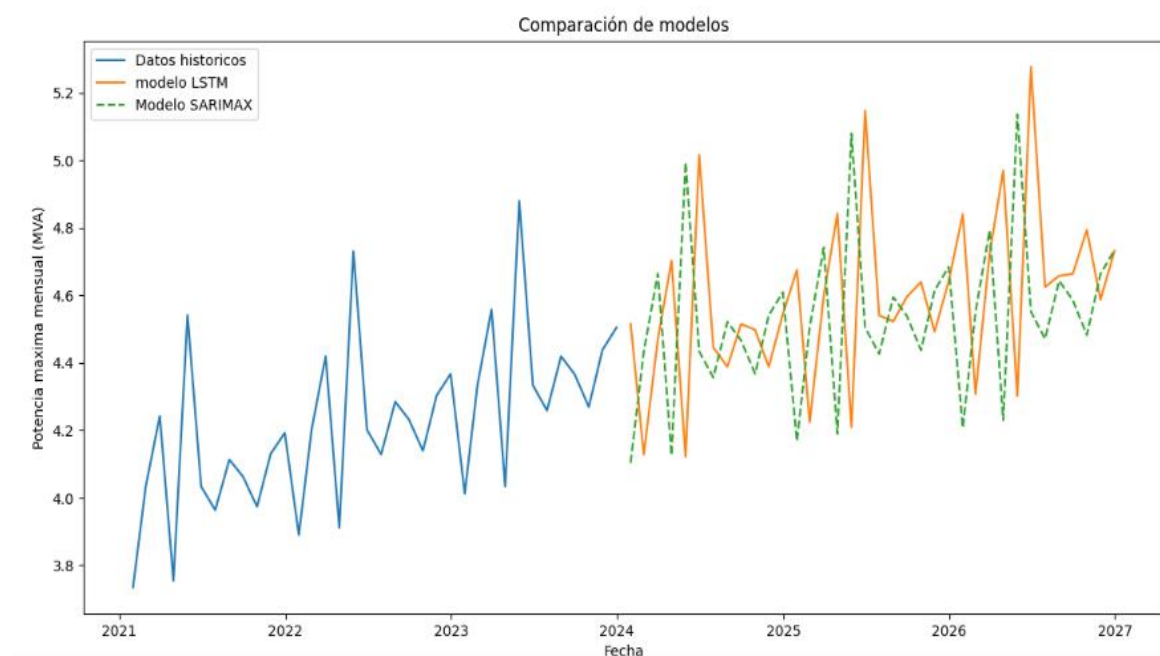
Se realiza la comparativa entre la predicción del modelo LSTM y SARIMAX, para verificar quien representa a lo largo del tiempo de mejor manera la predicción de la demanda tomando las fluctuaciones pronunciadas que se muestran en los datos históricos.

Se verifica que la base de datos utilizada cumpla con las exigencias para aplicar el modelo SARIMAX, donde esta está organizada de forma temporal, proporciona datos de potencia que tienen estacionalidad, donde se presenta un patrón recurrente desde el año 2021 al año 2023 con el aumento promedio de la potencia y para verificar si tiene datos estacionales se

realiza una prueba ADF donde el valor de P dio como resultado $2,6 \times 10^{-25}$; donde este es muy cercano a cero, lo que indica que los datos son estacionales.

Figura 4

Comparación de Modelos SARIMAX y LSTM



En la figura 4, se analiza el desempeño de los modelos LSTM y SARIMAX en la predicción de la potencia máxima mensual, comparando sus predicciones con los datos históricos. Donde para el modelo LSTM representado en la línea sólida, capturan tendencias y fluctuaciones complejas en los datos, reflejando la capacidad del modelo para identificar patrones no lineales en la serie temporal; mientras que, para el modelo SARIMAX, representado con la línea discontinua, se enfocan en modelar patrones estacionales, ofreciendo una perspectiva basada en la periodicidad observada en los datos históricos, aunque logra captar patrones estacionales.

Observaciones Clave del Análisis Comparativo

El modelo LSTM supera claramente a SARIMAX en términos de precisión, capturando mejor los patrones complejos de la serie temporal. Esto se refleja en las métricas de evaluación implementadas para el modelo SARIMAX, donde se mide la métrica RMSE con un resultado de 0.3191, lo que refleja que, en promedio, hay un error significativo en las predicciones.

Aunque el valor puede parecer relativamente bajo, en el contexto del rango de datos, podría ser considerable. El valor del R^2 Score dio como resultado -0.6708, esto muestra que el modelo no ha logrado captar adecuadamente los patrones presentes en los datos y que las predicciones realizadas por el modelo son considerablemente peores que una simple media.

Los resultados muestran que el modelo *SARIMAX* tiene un desempeño deficiente en la predicción de los valores de Potencia. El valor negativo del R^2 indica que el modelo no captura correctamente la tendencia ni la estacionalidad de los datos, y su desempeño es inferior a un modelo base que simplemente predice el valor promedio. Mientras que el alto valor de R^2 Score para el modelo LSTM muestra que este capta la mayoría de la variabilidad en los datos, lo que lo convierte en un modelo más confiable para predicciones futuras en series temporales con patrones no triviales.

Aunque SARIMAX puede ser efectivo para capturar patrones estacionales lineales, su desempeño se ve limitado por su incapacidad para manejar relaciones no lineales o cambios abruptos. En casos donde los datos tienen variabilidad más compleja, SARIMAX puede no ser suficiente sin modificaciones adicionales o modelos complementarios.

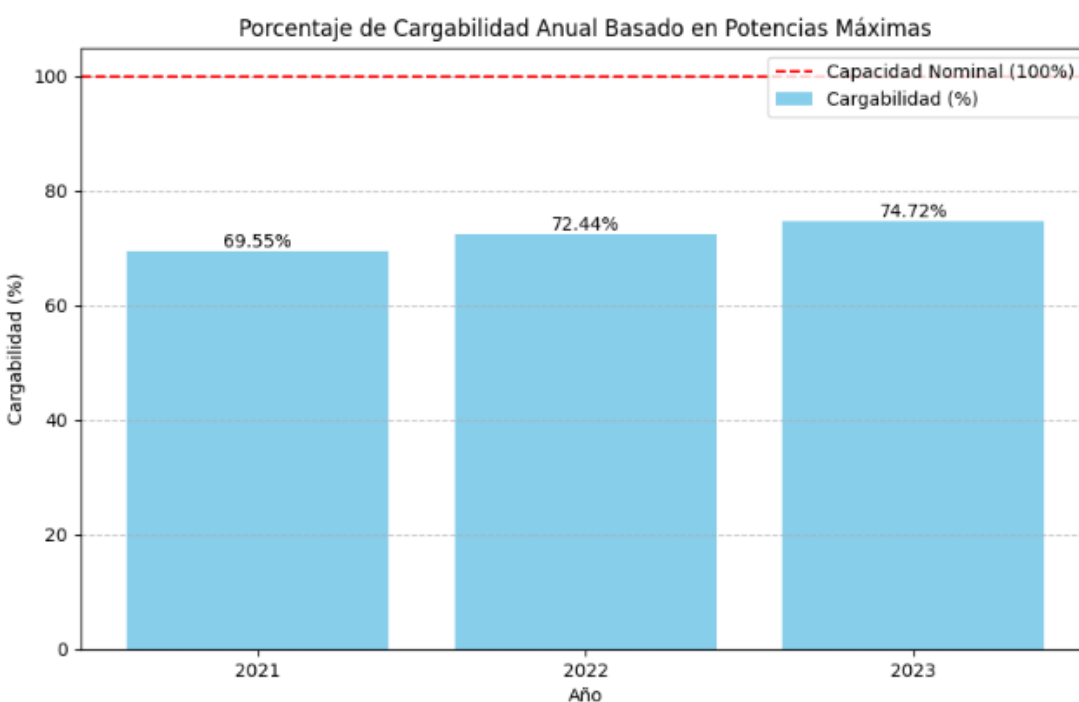
Descripción del Gráfico de Cargabilidad

A través de los datos históricos se realiza el cálculo de las potencias máximas mensual y anuales de los 3 años que se presenta en la base de datos histórica, esto para realizar el cálculo de

la limitante máxima que se obtuvo históricamente anualmente, desde el 2021 hasta el 2023, en cuanto a la cargabilidad del circuito esto debido por la potencia máxima que se presentaba anualmente; esto implica mostrar una de las limitantes operativas que tuvo el circuito para operar durante el año si se requería en su momento aumentar su capacidad de soportar potencias más altas.

Figura 5

Cargabilidad del Circuito con los Datos Históricos



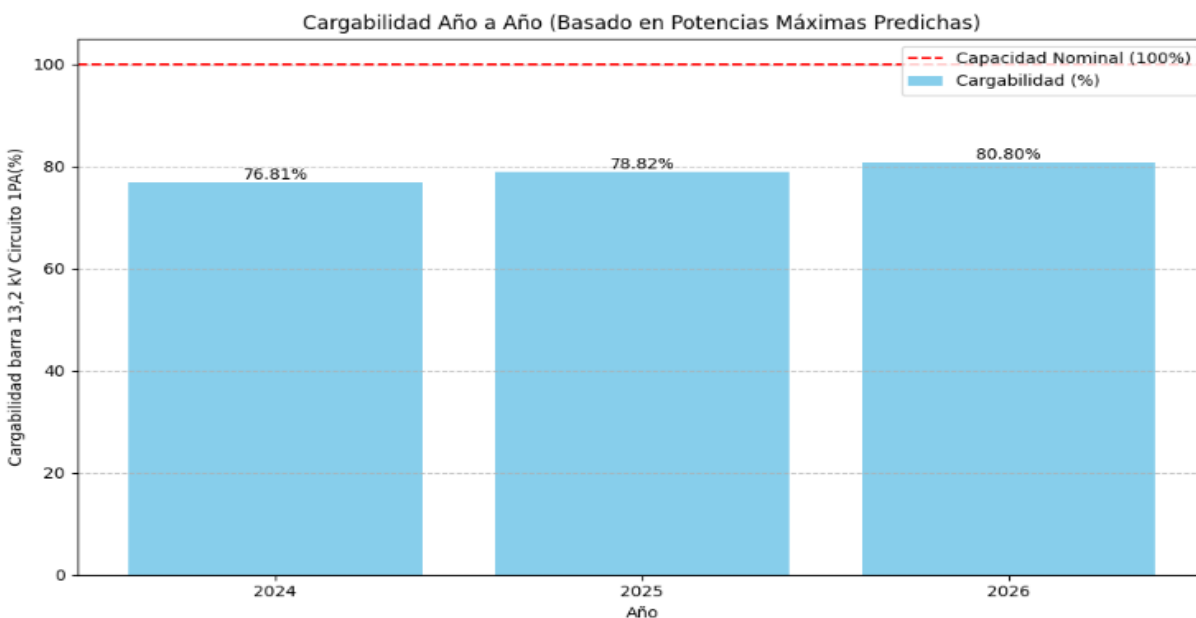
En la ilustración 5, El análisis de los datos históricos de cargabilidad en comparación con la situación actual permite identificar tendencias críticas en la utilización del sistema. Un incremento sostenido en la cargabilidad histórica podría indicar un riesgo potencial de saturación o la necesidad de optimizar la infraestructura para asegurar un desempeño eficiente del sistema crecimiento acelerado en la demanda de energía, lo que a su vez plantea retos significativos para

la infraestructura existente. Este escenario requeriría la implementación de acciones preventivas, tales como la inversión en infraestructura adicional para aumentar la capacidad instalada o la aplicación de estrategias de gestión de la demanda que optimicen el uso de los recursos actuales.

Para conocer el estado del sistema de distribución con la proyección de la demanda se hace necesario conocer las potencias máximas anuales calculadas por los próximos 3 años, para verificar y conocer el estado de los componentes que conforman el circuito y si estos tendrán la capacidad de asumir el crecimiento de la demanda.

Figura 6

Cargabilidad Predicha



Cargabilidad Futura

Para la figura 6, el análisis de las cargabilidades presentadas con los datos históricos en relación con la cargabilidad actual permite identificar tendencias críticas en la utilización del sistema. Un incremento sostenido en la cargabilidad predicha podría señalar un crecimiento acelerado en la demanda de energía, lo que a su vez plantea retos significativos para la

infraestructura existente, esto para realizar inversiones en la infraestructura eléctrica aumentando su capacidad de albergar y transportar mayor cantidad de potencia, además, de mejorar en sus componentes eléctricos para mejorar la calidad del servicio sin afectar a los usuarios existentes y tener mejor capacidad para alimentar a los nuevos usuarios.

Activos de Red Eléctrica (Estimación de Vida Útil)

La composición de la red eléctrica contiene elementos de seccionamiento y/o protección que cumplen el rol de mitigar fallas eléctricas o aislarlas, por lo que en su funcionamiento deben de garantizar cortes de energía bajo ciertos parámetros, para nuestro caso de estudio únicamente se contemplaron elementos “mecánicos”, es decir que no contienen telecontrol ni se encuentran encapsulados con un componente que minimice el impacto de arco eléctrico, por lo que naturalmente se deben operar sin carga. Los cortacircuitos y cuchillas monopolares son los elementos de red estudiados, por lo que su vida útil se evalúa conforme el tiempo, la cantidad de operaciones, entre otras variables.

Distribución de Datos

Los datos obtenidos constan de un total de 21.312 registros tomados desde el año 2019/01/01 al 2024/10/01, estos contienen un resumen mensual de las características más relevantes para los elementos de seccionamiento de red tipo cortacircuitos, teniendo en cuenta que una avería o daño del mismo representa un coste económico y reputacional para las compañías del sector energético, así como una indisponibilidad del servicio de energía eléctrica, es de vital importancia conocer el estado del equipo para dar confiabilidad de disponibilidad energética de ese elemento no sólo en términos de cargabilidad sino, también de su vida útil.

En la tabla 2 y 3, se identifican límites mínimos, máximos y las cotas de los percentiles respectivamente, para el caso de las variables categóricas se identifican 296 activos, 2

condiciones de reemplazo (si/no), fechas de reemplazo mayoritariamente en el 2019 con una mayor concentración en la ubicación de zaragoza.

Tabla 2

Resumen Estadístico Variables Categóricas Activos Eléctricos

Variable	Cuenta	Únicos	Moda	Frecuencia
ID	21312	296	1	72
Reemplazado	21312	2	No	21210
Fecha de reemplazo	102	57	2/03/2019	5
ubicación	21024	188	VIA ZARAGOZA OBANDO	576
zona	21312	1	urbano	21312

Para el caso de las variables en tabla 2, acumulativas de operación y mantenimientos, conglomeran el recuento de intervenciones planeadas o no, en el activo en específico, esta información es relevante para obtener un registro de vida del activo.

Tabla 3

Resumen Estadístico Variables Enteras o Flotantes Activos Eléctricos

Variable	media	Desv mín	máx	cuenta	0.25	0.5
Antigüedad	131,77	39,83	0 332	21.312	53	117
Número de operaciones mensual	2,68	5,08	0 70	21.312	0	0
Número de mantenimientos mensuales ejecutados	0,35	0,48	0 1	21.312	0	0
Cargabilidad max	0,26	0,08	0,15 0,39	21.312	0,2	0,25
No de operaciones acumuladas	79,66	56,5	0 297	21.312	33	72
No de mantenimientos acumulados	11,45	11,77	0 47	21.312	0	8
No de reemplazos	0,19	0,39	0 1	21.312	0	0

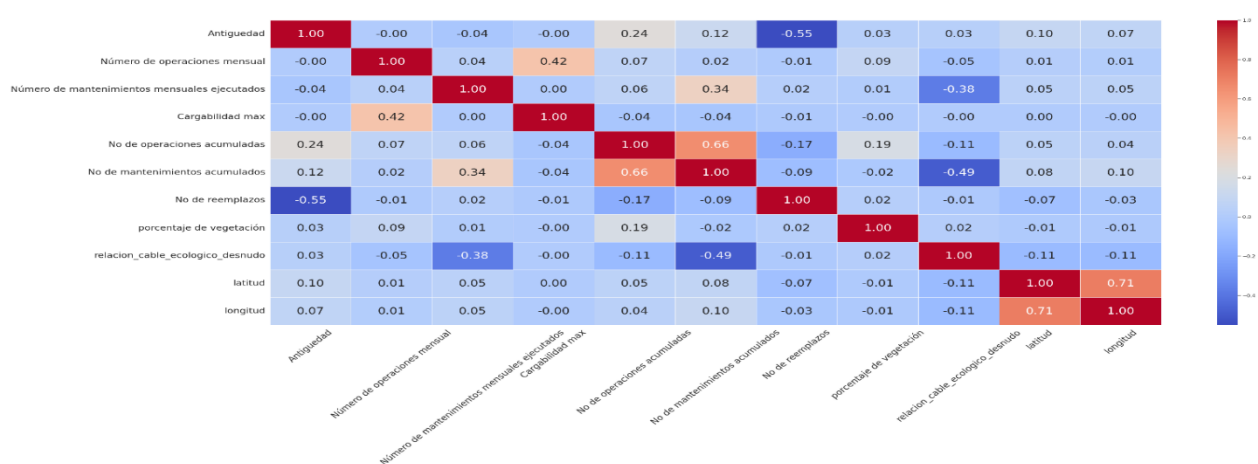
Relación cable ecológico desnudo	14,81	8,41	0	29	21.312	7	15
Latitud	4,69	0,04	4,6	4,74	21.312	4,66	4,7
Longitud	-75,93	0,0275,98	-	75,91	21.312	-75,93	75,92

Teniendo en cuenta de que cada variable aporta una información operativa relevante para cada elemento, se constata una correlación significativa con la finalidad de excluir variables dependientes o redundantes en el modelo. En la tabla 3, se obtienen correlaciones menores a 0.71, para el caso de la georeferenciación de latitud y longitud están relacionadas como complemento para su debida localización, por lo que obtienen una correlación fuerte y positiva, sin embargo, estas obedecen a un sistema de coordenadas, por lo que no hace parte de nuestro interés el evaluarlas individualmente, por otro lado, para las variables acumulativas de operaciones y mantenimientos se obtiene una correlación positiva moderada con un valor de 0.66, lo que implica que un aumento de fallas se correlaciona significativamente con un incremento de mantenimientos efectuados, para el caso de la variable de cargabilidad máxima y número de operaciones se obtiene una correlación moderada positiva con un valor de 0.42, por lo que podemos verificar que existe una correlación significativa pero no fuerte entre las variables, por lo que un aumento de la demanda de la red tiene alguna inferencia en proporción en el elemento de seccionamiento, lo que implicaría que para una mayor carga se tiene más probabilidad de fallo permanente o transitorio. De igual forma, se obtiene una correlación positiva débil de 0.19 entre la cantidad de operaciones acumuladas y el porcentaje de vegetación, para este caso es importante mencionar que la relación existente es posible que no sea lineal, sin embargo, naturalmente la cantidad de vegetación afecta el número de operaciones por lo general. En la variable de antigüedad, se obtiene uno de los indicios con mayor relevancia para el

problema abordado, debido a que está asociada a la variable de interés, vida útil, por lo que la relación débil positiva existente respecto al número de operaciones, indica que, ante mayor edad, mayor número de operaciones, así mismo aplica para el caso de los mantenimientos y contraste con el reemplazo que es en donde el equipo se renueva por completo, por lo que la edad tiende a cero.

Figura 7

Matriz de Correlación Variables Operativas Mensuales Acumuladas

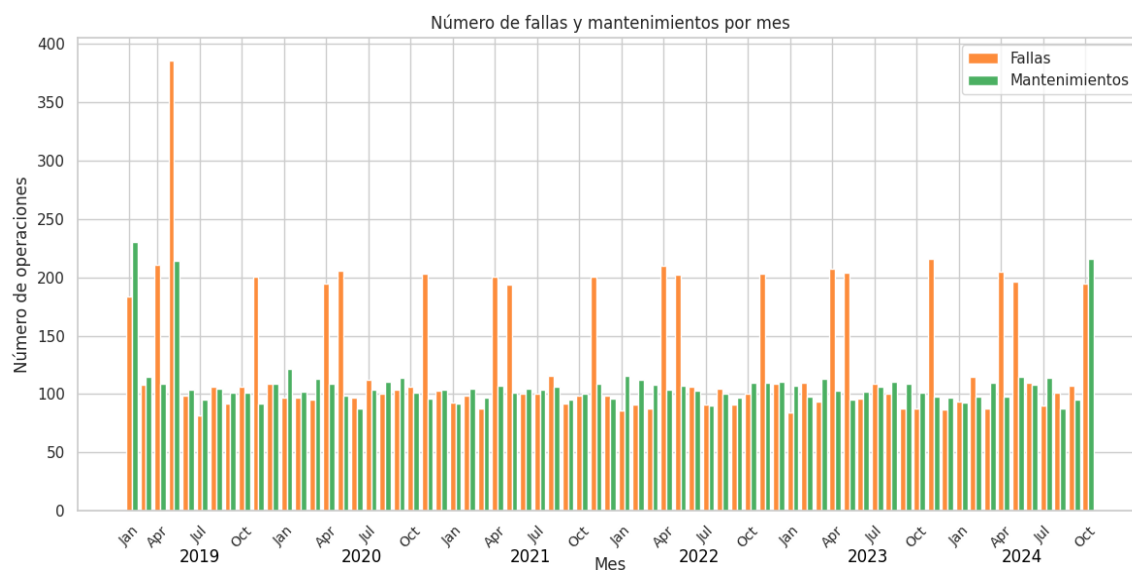


Según los resultados en la figura 7, las variables con mayor impacto en la vida útil del equipo son el número de operaciones y la cantidad de mantenimientos, por lo que su dinámica global para el conjunto de elementos contiene el siguiente comportamiento a través del tiempo.

La cantidad de los mantenimientos correctivos contiene una mayoría respecto a los mantenimientos efectuados a lo largo del periodo evaluado, se obtiene una cantidad de fallas con un volumen significativo y recurrente año a año, para los meses de abril, mayo y noviembre como se indica en la figura 8.

Figura 8

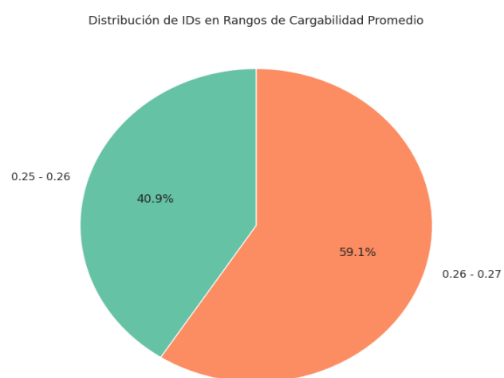
Gráfico de Barras, Comportamiento Histórico de Cantidades de Fallas y Mantenimientos



En la figura 9, se obtiene una distribución cargabilidad, la proporción de 40.9% (25-26) a 59.1% (26-27), lo que implica que los equipos cuentan con una disponibilidad local permisible según sus especificaciones técnicas y restricciones de red vistas desde subestación.

Figura 9

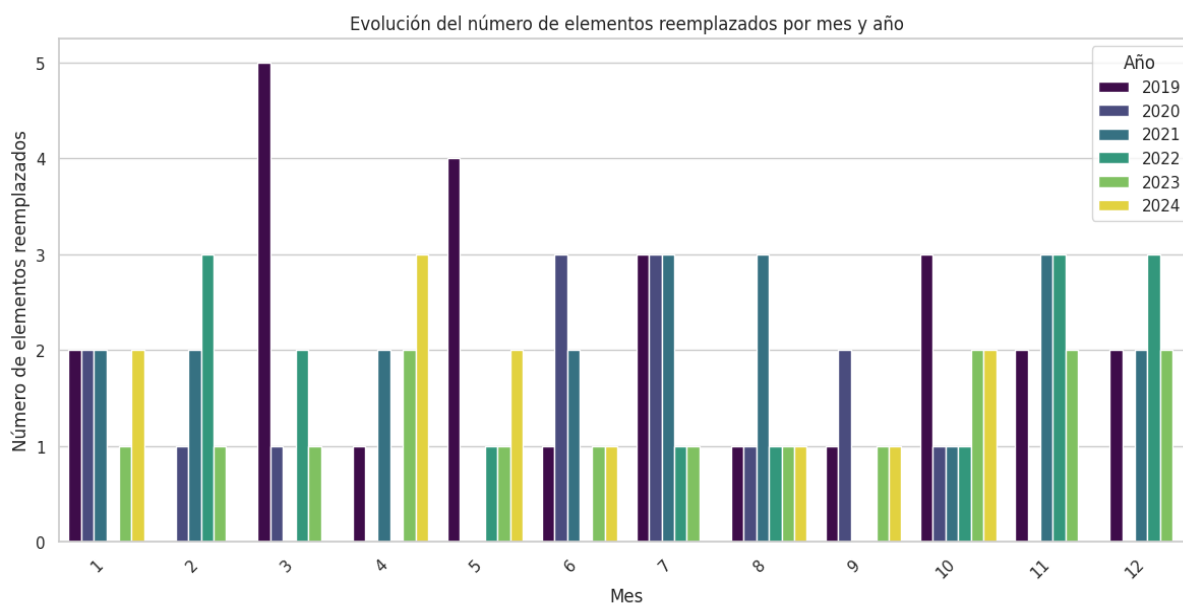
Gráfico de Torta, Relación de Cargabilidad de los Activos Instalados en el SDL



La figura 10, indica el comportamiento de los activos reemplazados a lo largo de los registros en el periodo de los 6 años, se presenta una mayor probabilidad de falla para los años 2019 y 2020 en los meses de marzo y abril. Se observa que en los años recientes la cantidad de equipos reemplazados se ha reducido respecto a los datos iniciales.

Figura 10

Activos de Red Reemplazados Mensualmente en Periodo 2019-2024



Modelos de Predicción Utilizados Vida útil

Se ingresan los IDs que tuvieron que ser reemplazados con sus registros previos teniendo en cuenta de que, al conocerse su fecha de instalación y su fecha de reemplazo, se genera el insumo necesario para verificar las condiciones operativas previas a su falla y la repercusión de estas, en total se realizaron 102 reemplazos en el sistema de distribución durante el periodo evaluado. Estos datos se entrenan los modelos con un 70% y un 30% restante para la validación del modelo (Tabla 4)

Tabla 4

Modelos Predictivos. Desempeño de Métodos para Encontrar la Vida Útil del Activo

Modelo	r2	mse	rmse	Tiempo
XGBoost	0,96	10,28	3,21	0,79
Random Forest	0,97	9,37	3,06	2,95
Support Vector Machine	0,70	88,53	9,41	2,89
MLP	0,96	12,48	3,53	7,50

El modelo Random Forest es el modelo con mayor capacidad explicativa para los datos de los elementos reemplazados en el sistema con un 97%, los modelos propuestos adicionales satisfacen las necesidades según la base de datos para la predicción de la vida útil de los activos eléctricos del circuito, debido a que están por encima del 90% salvo por SVM que, aunque su R2 es el más bajo, un 70% de capacidad explicativa es aceptable.

Teniendo en cuenta que el modelo Random Forest describe la vida útil de los activos reemplazados con una puntuación del 97%, se extrapolan las predicciones a los elementos que no han fallado definitivamente desde su instalación (ya sea que fuese renovado en el periodo evaluado o no), por lo que se genera el estimado de vida útil para cada registro hasta el mes 10 del 2024, con estos valores se produce una tendencia por elemento que interpreta cada una de las variables para su estimación. Una vez se obtienen estas estimaciones se procede con la evaluación del modelo holtwinters-depreciación exponencial decreciente para los dos años siguientes 2025-2026, en donde será posible identificar los elementos que presenten indisponibilidad a lo largo de su vida útil.

Se obtienen 171 elementos con un estimado de vida útil menor al 10% del máx promedio ($x < 13,1$) meses en color rojo. En la ilustración 11, se relacionan los activos que requieren de una intervención rápida debido a su condición de vida útil baja para el sistema eléctrico evaluado.

Figura 11

Equipos de Red Georreferenciado con una Estimación de Vida Baja



La ilustración 12, se selecciona una muestra de datos ubicada en Verona con fecha de instalación 1999-07-31 00:00:00 y la variable predictora con holtwinters de vida útil (meses): 9.615, se obtiene una dinámica de la reducción de la vida útil conforme avanza el tiempo, se logra apreciar una tendencia decreciente exponencial con mayor penalidad respecto al tiempo, esto se asocia al parámetro T de la ecuación 20 y 21, por lo que el modelo refleja una envolvente exponencial decreciente con un parámetro de tendencia a cero al final de su ciclo de vida.

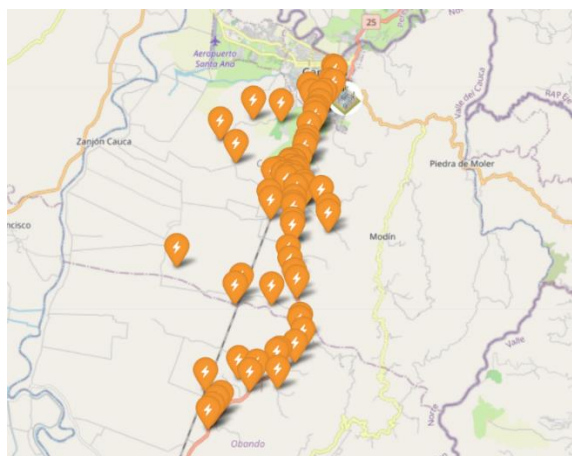
Figura 12

Gráfica Estimada de Vida de Útil Baja Activo 177



Figura 13

Equipos de Red Georreferenciados con un Estimado de Vida de Útil Media

**Figura 14**

Gráfica Estimada de Vida de Útil Media Activo 172



En la figura 14, se selecciona una muestra del activo 172 con dirección en hacienda la miranda con fecha de instalación 2011-03-31 00:00:00 y un estimado de vida útil de 19.21 meses, en este se observa que para el mes 10 del 2026, el valor de vida útil está por debajo de los 6 meses, por lo que representa un activo que debe de inspeccionarse o cambiarse en el periodo evaluado.

Finalmente, para el caso de los activos que contienen una vida útil prolongada mayor al 40% del valor máximo, se obtiene un total de 53 seccionamientos como indica la figura 15. Con una visión de renovación al largo plazo.

Figura 15

Equipos de Red Georreferenciados con un Estimado de Vida de Útil Alta

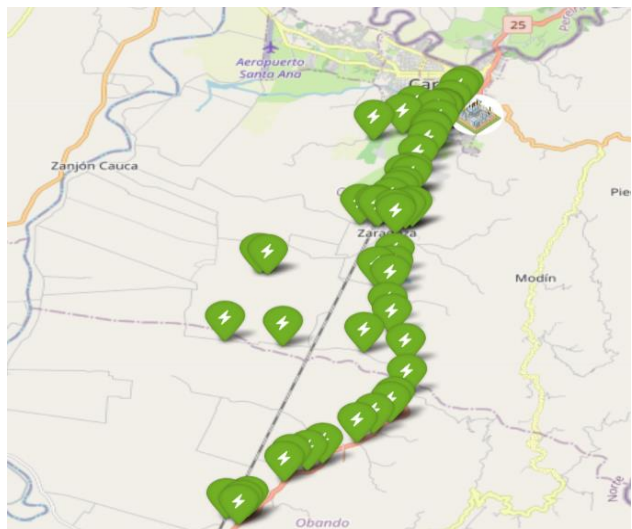
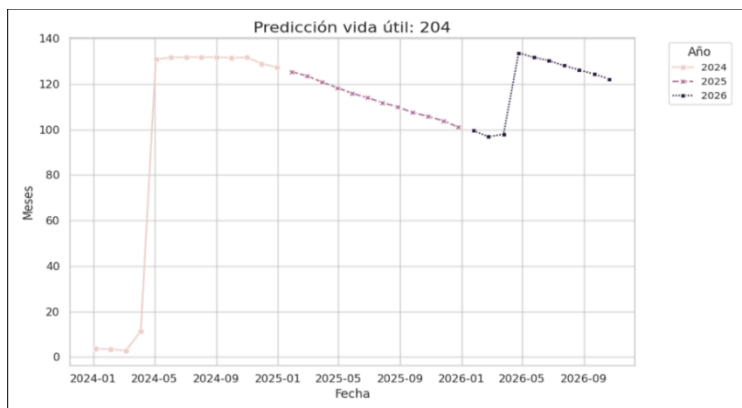


Figura 16

Gráfica Estimada de Vida de Útil Baja Activo 204



En la figura 16, se identifica un activo con una vida útil prolongada de 131.68 meses ubicada en hacienda ucrania - potrero - poste3 con una fecha de instalación 2024-04-04 00:00:00, naturalmente el equipo contiene valores de esperanza de vida útil altos al ser instalado recientemente

Conclusiones

El análisis exploratorio inicial, mediante histogramas, gráficos de correlación y series temporales, proporcionó una base sólida para entender la naturaleza de los datos y seleccionar los modelos adecuados. La visualización de datos históricos y predicciones a futuro permitió identificar patrones, tendencias y fluctuaciones en la demanda, proporcionando una base sólida para la toma de decisiones informadas.

El análisis de los gráficos mostró que el modelo LSTM logra seguir de cerca los datos históricos, capturando no solo las tendencias generales, sino también variaciones y picos específicos en la serie temporal de la demanda eléctrica, esto subraya la capacidad del modelo LSTM para aprender patrones complejos y no lineales en series temporales, siendo especialmente efectivo en escenarios de alta variabilidad. Mientras que, el modelo SARIMAX se le dificulta en seguir las variaciones y proyectarlas de manera más apropiada; esto podría considerarse para futuros casos estudio, en cuanto la parametrización y entrenamiento del modelo, para que pueda representar de manera óptima las fluctuaciones presentadas a lo largo del tiempo.

Las metodologías desarrolladas pueden ser adaptadas y replicadas a otros operadores de red que requieran realizar la predicción de demanda y gestión de activos eléctricos. Esto amplía la aplicabilidad del proyecto más allá del ámbito local.

La predicción de vida útil de los activos que aún no se encuentran en estado de reemplazado, se estima mediante los resultados del modelo random forest hasta el registro del mes 10 del 2024 con una precisión del 97% (obtenidos con dataset de los 6 años evaluados), con la finalidad de proyectar la dinámica del ciclo de vida útil del activo teniendo en cuenta el comportamiento de sus variables (sistema eléctrico) se estima una tendencia de vida útil

mediante holtwinters-depreciación exponencial, en donde se logra capturar el comportamiento de vida útil de cada activo según su histórico y a lo largo del tiempo tiende a decrecer exponencialmente para los 2 años próximos. Se obtiene como resultado 171 elementos con un estimado de vida útil menor al 10% del máx promedio ($x < 13,1$), predijo 72 activos eléctricos (24%) con la condición de un rango de vida útil entre el 10% y el 40% del valor máximo promedio ($13,1 < x < 52,4$). Para el caso de los activos que contienen una vida útil prolongada mayor al 40% del valor máximo, se obtuvo un total de 53 seccionamientos. Con este análisis se plantea como trabajo futuro la implementación de un esquema optimizado de recursos a asignar para las diferentes actividades de remodelación o mantenimientos en los activos identificados en este proyecto.

La publicación del mapa georreferenciado con las predicciones gráficas de la vida útil de los activos eléctricos para los próximos 2 años y de la cargabilidad anual para los próximos 3 años representa un avance significativo en la gestión proactiva de los activos. Al haber generado este mapa interactivo en formato HTML, se facilita el acceso visual y detallado a la información crítica sobre el comportamiento del ciclo de vida restante de cada activo. Este enfoque permite a los gestores y técnicos tomar decisiones informadas para optimizar el mantenimiento, mejorar la eficiencia operativa y prevenir fallos imprevistos, contribuyendo a la sostenibilidad y fiabilidad de la infraestructura eléctrica. Además, la visualización geoespacial de las predicciones de vida útil y cargabilidad anual proporciona una herramienta dinámica que mejora la planificación estratégica y la gestión de recursos en el corto y mediano plazo. El resultado gráfico del proyecto permite visualizar interactivamente los activos más críticos para disponibilidad del servicio, por lo que puede ser el insumo para la elaboración de un plan de mantenimiento o un estudio de remodelación completa según se requiera en el sistema.

Referencias Bibliográficas

- Arévalo, G. J., & Hernández, L. E. (2023). *Red neuronal artificial para la predicción de la demanda*. Universidad de Tecnología de Pereira.
- Atehortua, M. O. C., & Rodríguez, D. M. M. (2019). *Desarrollo de una metodología para el análisis y gestión de activos eléctricos de los operadores de red (OR) del STR y SDL*. Universidad Autónoma de Occidente.
- Brockwell, P. J., & Davis, R. A. (2016). *Introduction to time series and forecasting* (3rd ed.). Springer.
- Carrión, J. (2017). Implementación de una red neuronal artificial para la predicción de la demanda eléctrica a corto plazo. *Revista Técnica Energía*, 13, 36–42. <https://research-ebSCO-com.bibliotecavirtual.unad.edu.co/linkprocessor/plink?id=091f9740-e0f6-38bc-9e65-bf8055791b98>
- Chatfield, C. (2004). *The analysis of time series: An introduction* (6th ed.). Chapman & Hall/CRC.
- Ciencia de Datos. (n.d.). *Modelos ARIMA y SARIMAX en Python*. Ciencia de Datos. [https://cienciadedatos.net/documentos/py51-modelos-arima-sarimax-python#:~:text=ARIMA%20\(AutoRegressive%20Integrated%20Moving%20Average,de%20series%20temporales%20\(forecasting\)](https://cienciadedatos.net/documentos/py51-modelos-arima-sarimax-python#:~:text=ARIMA%20(AutoRegressive%20Integrated%20Moving%20Average,de%20series%20temporales%20(forecasting))
- Comisión Regulatoria de Energía y Gas. (2018). *Resolución CREG 015 de 2018: Planes de inversión aplicados a los operadores de red*. <https://www.creg.gov.co>
- Gómez, R., Cabrera, D., & Robles, P. (2023). Study for localization of fault in the electrical distribution systems. *INGENIUS*, 30, 64-78. <https://doi.org/10.1234/ingenius.2023.30>
- Hamilton, J. D. (1994). *Time series analysis*. Princeton University Press.

- Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: Principles and practice* (2nd ed.). OTexts.
- Kaledio, P. (2024). Machine learning applications in electric power systems: Enhancing efficiency, reliability, and sustainability. *Social Science Research Network*, 14-35. <https://doi.org/10.1234/ssrn.2024>
- Karthick, K., Ravivarman, S., & Priyanka, R. (2024). Optimizing Electric Vehicle Battery Life: A Machine Learning Approach for Sustainable Transportation. *World Electric Vehicle Journal*, 15(2). <https://doi.org/10.3390/wevj15020060>
- Munir, H., Vogel, B., & Jacobsson, A. (2022). Enfoques de inteligencia artificial y aprendizaje automático en la educación digital: Una revisión sistemática. *Information*, 13(4). <https://doi.org/10.3390/info13040203>
- Núñez, V. B., Velandia, R., Hernández, F., Meléndez, J., & Vargas, H. (2013). Atributos relevantes para el diagnóstico automático de eventos de tensión en redes de distribución de energía eléctrica. *RIAI - Revista Iberoamericana de Automática e Informática Industrial*, 10(1), 73-84. <https://doi.org/10.1234/riai.2013.010>
- Quiñones, L. I. A., Moncada, C. A. L., & Montenegro, D. A. B. (2022). Metodología para el mantenimiento predictivo de transformadores de distribución basada en aprendizaje automático. *Ingeniería*, 27(3), e17742. <https://doi.org/10.1234/ingenieria.2022.17742>
- Ramírez, P. M., & Torres, A. F. (2021). *Implementación de una red neuronal artificial para la predicción de la demanda eléctrica a corto plazo*. Universidad Nacional de Ingeniería.
- Tarasov, V. E. (2022). Fractional Dynamics with Depreciation and Obsolescence: Equations with Prabhakar Fractional Derivatives. *Mathematics*, 10(9). 1540. <https://doi.org/10.3390/math10091540>

Torres-Riveros, J., & Rivera-Rodríguez, S. (2018). Optimal energy dispatch in multiple periods of time considering the variability and uncertainty of generation from renewable sources.

Prospectiva, 16(2), 75-81. <https://doi.org/10.1234/prospectiva.2018.16>

Unite.AI. (n.d.). *¿Qué son las RNN y las LSTM en el aprendizaje profundo?* Unite.AI.

<https://www.unite.ai/es/what-are-rnns-and-lstms-in-deep-learning/>

Valencia, M. J., & Martínez, R. C. (2020). *Modelo predictivo de recomendación para el despacho energético del complejo hidroeléctrico Paute*. Universidad de Energía y Minas.