Modelo de pronóstico para la demanda de electricidad con un horizonte de tiempo de cinco años en el mercado regulado y no regulado de energía en Cali

PROYECTO DE GRADO

Nelson Andrés Andrade Bonilla
Mario José Castellanos Valencia

Asesor
Hernán Darío Benítez Restrepo

FACULTAD DE INGENIERÍA Y DISEÑO

MAESTRÍA EN CIENCIA DE DATOS

SANTIAGO DE CALI

2022

Modelo de pronóstico para la demanda de electricidad con un horizonte de tiempo de cinco años en el mercado regulado y no regulado de energía en Cali

PROYECTO DE GRADO

Nelson Andrés Andrade Bonilla
Mario José Castellanos Valencia

Asesor
Hernán Darío Benítez Restrepo

FACULTAD DE INGENIERÍA Y DISEÑO

MAESTRÍA EN CIENCIA DE DATOS

SANTIAGO DE CALI

2022

CONTENIDO

RESUMEN **5**

1. INTRODUCCIÓN **6**
1.1. Antecedentes 6
1.2. Justificación 7
1.3. Contexto 7
1.4. Planteamiento del Problema 7
1.5. Objetivo General 9
1.6. Objetivos Específicos 9
1.7. Organización del Documento 9

2. ANTECEDENTES **10**
2.1. Marco Teórico 10
2.1.1. Dominio del problema 10
2.1.2. Dominio de la solución 11
2.2. Estado del Arte 21
2.2.1. Trabajos seleccionados 22
2.2.2. Resumen de criterios 24

3. METODOLOGÍA **28**
3.1. Esquema de trabajo 28
3.2. Fases del Proyecto 28
3.2.1. Fase 1: Comprensión del negocio 28
3.2.2. Fase 2: Comprensión de los datos 29
3.2.3. Fase 3: Preparación de los datos 29
3.2.4. Fase 4: Modelado 29
3.2.5. Fase 5: Evaluación 29
3.2.6. Fase 6: Despliegue 29

4. COMPARACIÓN DE MODELOS DE APRENDIZAJE AUTOMÁTICO **30**
4.1. Recolección de los datos 30
4.2. Preparación de los datos 32
4.3. Descripción de los datos 33
4.4. Modelamiento 37
4.5. Evaluación 38
4.5.1. Estrategia de Evaluación 38
4.5.2. Resultados de los Modelos 39

4.6. Despliegue 42

5. CONCLUSIONES Y TRABAJO FUTURO **43**

Referencias **44**

Índice de figuras

1. Componentes de la inteligencia artificial como ciencia. 12
2. Capas de una red neuronal. 16
3. Flujo del algoritmo GSA. 19
4. Flujo del algoritmo PSO. 20
5. Metodología CRISP-DM. 28
6. Localización de los sensores de temperatura del CIAT. Fuente: Google Maps. 30
7. Distribución no normalizada de los consumos de energía por tipo de mercado. 32
8. Distribución normalizada de los consumos de energía por tipo de mercado. . 33
9. Distribución de los consumos totales por año en GWh. 34
10. Distribución serializada del consumo por tipo de mercado GW. 34
11. Composición de los consumos por categoría del servicio. 35
12. Temperaturas área CIAT cercanas al alba, cenit y ocaso, promedios mensuales durante los 15 años de recolección. 35
13. Crecimiento del PIB del Valle del Cauca y comportamiento del IPC en el periodo de recolección de los datos. 36
14. Distribución de los suscriptores en el tiempo. 36
15. Correlación entre las variables del estudio. 37
16. Baseline con el modelo ARIMA (0,1,1)(0,0,0), para ventana de tres años (arriba) y ventana de cinco años (abajo). 38
17. Predicciones de test 3 primeros puestos, en ventana de 3 años. 39
18. Predicciones de test 3 primeros puestos, en ventana de 5 años. 41

Índice de tablas

1. Proyecciones de demanda en el mercado regulado en kWh/mes realizadas en EMCALI EICE ESP para compra de energía, en los años 2017 a 2020 usando modelos de media móvil, realizados con tres meses de anticipación para 12 meses posteriores. 8
2. Resumen de los criterios de comparación entre los artículos seleccionados. . . 26
2. Resumen de los criterios de comparación entre los artículos seleccionados. . . 27
3. Configuración de datos para los modelos evaluados. 38

4. Resultados modelo baseline ARIMA del comercializador. 39
5. Resultados de los modelos en ventana de 3 años. 40
6. Resultados de los modelos en ventana de 5 años. 42

RESUMEN

En este trabajo de grado, se formuló una propuesta para abordar el problema de cómo mejorar la predicción de largo plazo en la demanda de energía eléctrica en Cali, utilizando técnicas de aprendizaje automático y modelos de inteligencia artificial. Para hacerlo se propuso una metodología basada en CRISP-DM, la cual propone en primera fase entender el negocio, seguidamente entender y preparar los datos, para lo cual se realizó análisis univariado y multivariado para conocer las posibles influencias y correlaciones entre los datos, posteriormente se realizó el modelamiento y evaluación de los modelos aplicados. Este fue un proceso iterativo dado que en algunas etapas los resultados condujeron a realizar nuevas pruebas y repetir partes del proceso.

El proceso de preparación de los datos exigió un esfuerzo adicional por la dificultad que se encontró en la extracción de los datos del comercializador y de otras entidades. Los modelos de la ciencia de datos usados para abordar la solución al problema fueron algunos tradicionales como ARIMA, Support Vector Regression (SVR), Ridge, Lasso, Random Forest y otros representativos del estado del arte pertenecientes al aprendizaje profundo como Artificial Neural Networks - Particle Swarm Optimization (ANN-PSO) , Extreme Gradient Boost (XGBoost), Recurrent Neural Networks-Long Short-Term Memory (RNN-LSTM).

Los resultados obtenidos mostraron que los modelos SVR y Ridge con optimización PSO y Gravitational Search Algorithm (GSA), muestran un mejor rendimiento cuando los datos no presentan mayores perturbaciones como es el caso del problema que aborda este estudio, mientras que los modelos profundos demostraron menor rendimiento como RNN-LSTM en las métricas seleccionadas. La validación a la que fue sometida la propuesta, consistió en aplicar métricas como RMSE, MAE y MAPE, utilizando validación cruzada y out-of-bag (OOB - muestra para test) con selección de conjuntos de entrenamiento y validación de diferentes horizontes.

Se validaron los supuestos para poder verificar la aplicación de los modelos, los cuales se configuraron con diferentes valores de hiperparámetros y se debió utilizar una estrategia de creación de línea base de predicción a largo plazo (cinco años) basada en modelos ARIMA aplicados por el comercializador. Finalmente, después de todo el trabajo desarrollado y la validación realizada, se puede afirmar que el enfoque de solución propuesto y la metodología empleada para obtenerla resultan apropiados.

1. INTRODUCCIÓN

1.1. Antecedentes

El Mercado Eléctrico en Colombia, se organiza y desarrolla por mandato de la constitución de 1991 al decretar la obligatoriedad de la prestación de los servicios públicos domiciliarios por parte del Estado (Aguilar & Diaz, 2004). Posteriormente, en 1994, se promulgan las leyes 142 de Servicios Públicos Domiciliarios (, CongresodeColombia) y la 143 o ley Eléctrica (, CongresodeColombia), que se fundamentan en la construcción de condiciones de competencia en las actividades de generación, comercialización y monopolio regulado que terminan en la transmisión, distribución y entrega del servicio al usuario final. Al mismo tiempo, se crean las entidades regulatorias como la Comisión de Regulación de Energía y Gas (CREG) y las entidades de control como la Superintendencia de Servicios Públicos Domiciliarios (SSPD) reglamentada por la ley 142.

En Diciembre de 1992, la Comisión Nacional de Energía se transformó en la Unidad de Planeación Minero-Energética (UPME), la cual se encarga de desarrollar el plan energético nacional y las proyecciones de demanda de energía eléctrica, así como también la creación del plan de expansión y transmisión de energía para todo el territorio nacional. Estos planes son necesarios y se llevan a cabo para asegurar la disponibilidad energética permitiendo controlar tanto el flujo como la cantidad de energía eléctrica que debe ser generada o adquirida con anticipación y distribuida conforme a las proyecciones de demanda hacia las regiones. Por otro lado, se entra a regular también la participación de los generadores en el mercado, estableciendo los costos de comercialización y la eficiencia en la contratación de largo plazo, así como la participación en el mercado de energía de corto plazo, además se definió la formulación tarifaria para los usuarios regulados con vigencia a cinco años. Se permite el mercado no regulado de energía y se limita a partir del año 2000 a los usuarios con consumos superiores a los 55 MWh/mes que rige actualmente.

Como consecuencia de estos aspectos, se crea el Estatuto del Usuario de Servicios Públicos de Energía y Gas (, CongresodeColombia), con el fin de proteger al usuario ante prácticas de abuso de posición dominante en los contratos de prestación de estos servicios y también se les da a las empresas la posibilidad de realizar acciones contra el uso no autorizado o fraudulento del servicio y crea normas para la recuperación de pérdidas no técnicas relacionadas.

Finalmente, con respecto al usuario de última milla se ajustan las condiciones relacionadas con los procedimientos del registro de medidores, las fronteras, los contratos y la entrega de las medidas, el manejo de los medidores defectuosos instalados y aspectos relacionados con la medición y facturación. En este contexto los comercializadores deben pronosticar la demanda

para no verse perjudicados dentro del mercado de energía y poder cumplir con el mandato de la constitución y la ley en la prestación de este servicio.

1.2. Justificación

Estas necesidades se suplen actualmente con modelos que se utilizan desde hace varios años pero que deben ser actualizados utilizando enfoques y tecnología de vanguardia, para que permitan ajustar de forma más precisa las variables y proyecciones del consumo, considerando incluso el ingreso de autogeneradores menores de energía a través de paneles solares que modificarán la demanda en un futuro cercano si llegan a tener un crecimiento no calculado. En Colombia por mandato de la constitución, el estado debe proporcionar el servicio de energía eléctrica entre otros servicios públicos, a todos los usuarios y lo hace a través de las empresas generadoras y comercializadoras pertenecientes al mercado de energía creado para tal fin. Es por esto que se hace necesario tener un panorama lo más completo posible del mercado, donde el consumo es el factor principal para poder cumplir con esta obligación y al mismo tiempo tener sostenibilidad económica.

1.3. Contexto

El mercado de energía eléctrica en Colombia contempla dos modalidades de adquisición como son: a) la contratación directa de venta y suministro entre generadores y/o comercializadores, y b) la compra de energía en bolsa. Dado que la capacidad instalada no es suficiente para generar la energía que consume el país, existe un alto grado de incertidumbre en esta actividad primordial de adquisición, que debe ser controlada óptimamente y en la cual se hace necesario sostener un equilibrio en la cantidad de energía por comprar en cada una de las modalidades, la ventana de tiempo anticipado para realizar la compra y factores como el flujo de caja, la demanda de los suscriptores y la fluctuación de precios por la volatilidad del mercado que depende de diversas condiciones económicas, legales y de medio ambiente, principalmente relacionadas con el clima, dado que en Colombia el 68 % de la capacidad instalada proviene de fuentes renovables (ACOLGEN, 2020).

1.4. Planteamiento del Problema

La generación y distribución de energía eléctrica tienen tres horizontes de tiempo en los cuales debe anticiparse el consumo, estos son de corto plazo (Short Term Load Forecasting - STLF) determinado en horas, días o semanas, mediano plazo (Medium Term Load Forecasting - MTLF) en meses o años y largo plazo (Long Term Load Forecasting - LTLF) en años o décadas (Mir et al., 2020).

Para los comercializadores en Colombia es más difícil predecir la demanda de largo plazo para saber cuánta energía comprar porque se introducen variables impredecibles en el tiempo como cambios económicos de orden mundial, nuevas normas, entre otros. En la ciudad de Cali donde el mayor comercializador es EMCALI EICE ESP el cual opera la prestación del servicio al 92 % de los predios de la ciudad, utiliza modelos ARIMA para hacer predicciones en horizontes de corto plazo (ver tabla 1). De igual manera, se hace necesario incorporar los suficientes aspectos determinantes del consumo como la variabilidad de factores climáticos, así como económicos que afectan la capacidad de pago de los usuarios o indicadores de productividad general como el PIB. En ocasiones se incrementan las pérdidas no técnicas o se desborda la demanda por auge en el sector de la construcción de vivienda lo cual también afecta el consumo.

Tabla 1: Proyecciones de demanda en el mercado regulado en kWh/mes realizadas en EMCALI EICE ESP para compra de energía, en los años 2017 a 2020 usando modelos de media móvil, realizados con tres meses de anticipación para 12 meses posteriores.

Año	**Proyección**	**Real**	**Diferencia**	**%**
2017	2.041.430.073	1.918.197.806	123.232.267	-6.4 %
2018	1.935.045.114	1.946.546.712	7.501.598	0,4 %
2019	1.941.746.035	1.979.881.276	38.135.241	1,9 %
2020	2.003.928.335	1.911.229.124	92.699.212	-4,9 %

Este problema genera diversos inconvenientes a estos actores del mercado, que a la postre se trasladan a los usuarios, tales como desabastecimiento energético a la industria, aumento en el precio final de la energía, generación de multas y pérdidas económicas para los comercializadores, produce crisis financiera en las regiones, reducción de la productividad general e incremento del descontento en los usuarios.

Teniendo en cuenta el impacto de las proyecciones del consumo en el proceso de adquisición y comercialización de la energía eléctrica y la necesidad de adquirirla de forma anticipada para abastecer a las regiones, en este proyecto se busca resolver la pregunta: ¿Cómo mejorar la predicción de largo plazo en la demanda de energía eléctrica en Cali, utilizando técnicas de aprendizaje automático y modelos de inteligencia artificial?

1.5. Objetivo General

Desarrollar y validar un modelo de aprendizaje automático para pronosticar el consumo de energía eléctrica en Cali, con un horizonte de tiempo a 5 años para el mercado regulado y no regulado de energía.

1.6. Objetivos Específicos

1. Diseñar diferentes modelos de aprendizaje automático para realizar la predicción de la demanda de energía eléctrica de largo plazo.

2. Seleccionar el mejor modelo de aprendizaje automático para realizar la predicción de la demanda de energía eléctrica de largo plazo a partir de los modelos diseñados.

3. Validar el desempeño del mejor modelo de aprendizaje encontrado, con una estrategia de comparación que pueda constrastar pronósticos contra la realidad.

1.7. Organización del Documento

El presente documento está compuesto por 4 capítulos adicionales a este. El segundo capítulo, presenta el marco teórico y el estado del arte, los cuales brindan el soporte teórico en el que se basa el desarrollo de este estudio. En el tercer capítulo, se discute la metodología que se seguirá para cumplir el objetivo general del proyecto. Por su parte, el cuarto capítulo, presenta el desarrollo de la metodología y los resultados del estudio, mientras que el quinto y último capítulo, ofrece conclusiones y comentarios finales.

2. ANTECEDENTES

2.1. Marco Teórico

2.1.1. Dominio del problema

La energía eléctrica es la principal fuente de progreso para la humanidad, la forma de obtenerla ha ido evolucionando con el tiempo desde la madera, pasando por el carbón, petróleo y el gas natural hasta las fuentes limpias como la hidráulica, eólica y solar. Está asociada al crecimiento económico de las naciones así que todos los países tienen estrategias para desarrollarla de manera segura, eficiente, constante y limpia. Siendo la energía eléctrica el fundamento base del desarrollo socioeconómico es de suma urgencia identificar y entender los factores involucrados en el conocimiento de la cantidad de esta energía que un país necesita producir y distribuir para abastecer su producción, atender el crecimiento de la población y su calidad de vida. Por esta razón se describen a continuación los principales elementos que conforman el dominio de este problema.

- **Variables determinantes del consumo.**

Aunque es cierto que las condiciones de cada país son diferentes, con el conocimiento alcanzado se han logrado identificar muchas de las variables que determinan la demanda de la energía eléctrica. En algunos estudios realizados en Europa se han analizado las posibles relaciones entre el producto interno bruto (GDP) por sus siglas en ingles o PIB en español, el crecimiento poblacional, la formación bruta de capital fijo (GF), el consumo doméstico de electricidad, la generación de energía eléctrica y la energía asociada a las emisiones de CO2 (Ma et al., 2021).

En otros estudios (Mir et al., 2020), usando modelos econométricos se analizaron las relaciones entre las variables para países en vías de desarrollo como Colombia, Venezuela o Pakistán tales como el PIB, el crecimiento poblacional, el ingreso per cápita, la elasticidad de precios, el precio de la electricidad, el número de usuarios y las condiciones climáticas.

En Colombia existe una clasificación socioeconómica de los inmuebles denominada estratificación que fue adoptada de forma estándar en la Ley 142 de 1994 con el fin de unificar los criterios y metodologías que usaban las empresas de energía individualmente hasta entonces para cobrar los servicios públicos prestados. El principal objetivo es identificar los sectores geográficos de una ciudad por sus características socioeconómicas para orientar las políticas y programas de expansión y mejoramiento de infraestructura y servicios en general para la población. Esto facilita la aplicación de los principios constitucionales de solidaridad y redis-

tribución del ingreso en la realización de los cobros por tarifas diferenciales en impuestos y servicios (DANE, 2017).

Esta estratificación es importante para modelar el consumo de electricidad en Colombia (Peña-Guzmán & Rey, 2020) porque cuando se involucra el precio de la energía, que está basado en una tarifa diferencial por estrato debido a que las características socio económicas varían en cada uno, también está relacionado con la distribución de la población y el ingreso per cápita. El sector residencial está dividido en 6 estratos (1. Bajo-bajo, 2. Bajo, 3. Medio-bajo, 4. Medio, 5. Medio-alto y 6. Alto).

- **Pronósticos de la demanda de energía eléctrica.**

El proceso de pronosticar la demanda de electricidad por parte de los actores del mercado regulado de energía se basa en los horizontes de tiempo de pronóstico y en los determinantes del consumo o demanda de electricidad. Rueda et al. (Rueda et al., 2011) describen porqué es importante el pronóstico de la demanda de electricidad por horizontes según el tipo de actor en el mercado.

Por ejemplo, en el caso de los generadores el pronóstico de corto plazo es importante porque les permite definir la cantidad de energía que deben generar al día siguiente, programar las unidades de generación y establecer los precios de bolsa. Con el de mediano plazo definen el mercadeo y venta en contratos y les permite seguir la evolución del mercado y con el de largo plazo pueden detectar excesos y faltantes de capacidad de generación, con el fin de planear las inversiones de expansión del sistema de generación. Para los comercializadores en el mediano y largo plazo, quienes son el objetivo de este proyecto, el pronóstico de demanda es un insumo fundamental para el análisis del comportamiento del mercado y la definición de los planes estratégicos y operativos para la comercialización y adquisición de la electricidad.

Se han desarrollado numerosos métodos para la previsión de la demanda de electricidad, en su mayoría de corto plazo lo cual representa la debilidad para encontrar modelos que proporcionen el pronóstico de largo plazo.

2.1.2. Dominio de la solución

La solución al problema del pronóstico de la demanda de electricidad de largo plazo se encuentra en el ámbito de la inteligencia artificial (ver figura 1)[1] y el aspecto de los determinantes del consumo de energía en el ámbito de la econometría. El aprendizaje de máquina o Machine Learning (ML) es un campo de inteligencia artificial que aprende de los datos más que de la

[1]Síntesis explicativa y gráfica tomadas de https://www.ibm.com/co-es/analytics/machine-learning

programación explícita, los datos son el insumo de un algoritmo que se entrena con ellos y como resultado entrega un modelo que en este caso hace pronóstico a partir de nuevos datos de entrada.

A continuación se describen las categorías que conforman el aprendizaje automático. Algunas de estas categorías serán la base para pronosticar la demanda de electricidad en Colombia en este proyecto.

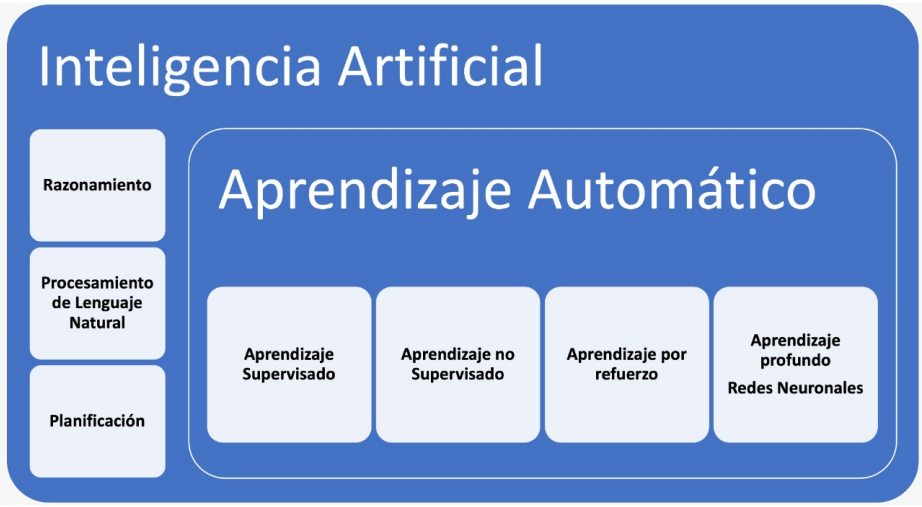

Figura 1: Componentes de la inteligencia artificial como ciencia.

Aprendizaje Automático

La Inteligencia Artificial se enfoca en el estudio de cómo hacer que los computadores hagan cosas que hasta el momento las personas hacen mejor (Ertel, 2017), lograr este objetivo implica investigar sobre los mecanismos de aprendizaje y el desarrollo de algoritmos de aprendizaje. La tarea del aprendizaje automático es el estudio de algoritmos para computador que mejoran automáticamente a través de la experiencia.

En forma general el computador observa unos datos, construye un modelo basado en esos datos luego utiliza el modelo junto con una hipótesis del mundo y un programa para resolver problemas. Estos algoritmos se encuentran en cuatro grandes grupos de aprendizaje automático. Un concepto muy importante con respecto a todos estos modelos es la validación y el compromiso entre sesgo y varianza que debe ser optimizado para llegar a resultados correctos.

- **Aprendizaje supervisado**

El objetivo del aprendizaje supervisado consiste en aprender un modelo a partir de unos datos de entrenamiento etiquetados que se pasan por un agente, en este caso un algoritmo de aprendizaje automático, el agente aprende una función que, cuando se le da una nueva observación predice la etiqueta apropiada de salida. Más formalmente, dado un conjunto de entrenamiento con N muestras de pares de entrada-salida $(x_1, y_1), (x_2, y_2), \ldots (x_N, y_N)$, donde cada par que es generado por una función $y = f(x)$, descubre una función h que se aproxima a la función verdadera f. La función h se llama hipótesis del mundo, extraída del espacio $\mathcal{H}$ de posibles funciones y se asimila a un modelo de datos extraído de una clase modelo $\mathcal{H}$. La salida y_i es la respuesta correcta o predicción que se le pide al modelo y se denomina ground-truth (verdad fundamental) (Russell & Norvig, 2021).

Se espera poder obtener una hipótesis consistente h donde cada x_i del conjunto de entrenamiento tiene $h(x_i) = y_i$. Con salidas de valores continuos esto no es posible, en lugar de eso se busca la función que mejor ajuste los datos para la cual cada $h(x_i)$ está cerca a y_i. La verdadera medida de una hipótesis no es cómo le va con el conjunto de entrenamiento, sino qué tan bien maneja las entradas que aún no ha visto. Se puede evaluar esto con una segunda muestra de pares (x_i, y_i) llamada conjunto de prueba. Se dice que h generaliza bien si predice con precisión las salidas del conjunto de prueba.

A continuación se describen algunos modelos que se utilizan en aprendizaje supervisado:

Regresión lineal: Este es el enfoque más simple del aprendizaje supervisado, pero en realidad también es el punto de partida para otros modelos que son extensiones o generalizaciones de este. Se basa en estimar el valor de la variable objetivo dados unos valores predictores relacionados con ella de manera aproximadamente lineal (James et al., 2021). Este método se llama simple cuando solo hay una variable predictora o múltiple cuando hay mas de una, los parámetros corresponden a los coeficientes de la variables.

K vecinos mas cercanos K-NN: El objetivo de este algoritmo es asignar la clase o valor agregado de las instancias conocidas que se encuentran más cerca de la nueva observación a predecir y se basa en instancias de aprendizaje más que en un modelo subyacente probabilístico (James et al., 2021). Depende de una función de distancia que se escoge de acuerdo con la cantidad y características de las variables independientes y se usa para encontrar los vecinos más cercanos a cada nueva observación, y de un parámetro k que corresponde al número de vecinos más cercanos que se quieren considerar para establecer la clase o valor de una nueva instancia. La selección de k es muy importante porque un valor pequeño produce alta varianza y bajo sesgo lo que conduce a un sobre ajuste y un valor alto produce lo contrario alto sesgo y baja varianza.

Arboles de decisión: Un árbol de decisión es una representación de una función que asigna un vector de valores de atributos a un único valor de salida que se llama “decisión" (Russell & Norvig, 2021). Un árbol alcanza su decisión realizando una secuencia de pruebas, comenzando en la raíz y siguiendo la rama apropiada hasta que se alcanza una hoja. Cada nodo interno en el árbol corresponde a una prueba del valor de uno de los atributos de entrada, las ramas del nodo se etiquetan con los posibles valores del atributo, y los nodos hoja especifican qué valor debe devolver la función con la determinación de una clase.

Regresión logística: Este es un algoritmo de clasificación, se basa en la regresión lineal pero ha sido modificado para distinguir de forma binaria entre dos clases usando la función sigmoide o logística de forma que los valores de esta función se puedan interpretar como probabilidades de que una instancia pertenezca a una clase específica (Géron, 2019). Cuando el valor predicho es mayor que un umbral predeterminado la respuesta es positiva de lo contrario es negativa.

Máquinas de vectores de soporte (SVM): Este algoritmo es de clasificación y está basado en el clasificador de margen máximo, el cual espera que dos clases sean separables por un hiperplano de una dimensión inferior a los datos, si este plano existe entonces los datos que caen por encima de él son de una clase y los que caen por debajo son de otra clase. Tiene algunas restricciones como el margen, que es la distancia mínima desde las observaciones al hiperplano y sirve para determinar los vectores de soporte de los cuales depende el hiperplano de separación los cuales se encuentran justo en las líneas del margen, cuando las clases no son separables entonces se convierte en un clasificador de vectores de soporte, que además flexibiliza la condición del margen permitiendo que dentro de él haya vectores de la otra clase del lado incorrecto del hiperplano de separación (James et al., 2021).

Las observaciones que caen directamente sobre el margen o en el lado equivocado del margen de su clase se llaman vectores de soporte. El clasificador de vectores de soporte se extiende a máquinas de vectores de soporte usando kernels para agrandar el espacio de características, con el fin de acomodar límites no lineales entre las clases y está basado en el producto interno de las observaciones mas que en las observaciones mismas.

- **Aprendizaje no supervisado**

A diferencia del aprendizaje supervisado, el aprendizaje no supervisado se enfoca en un conjunto de características de entrada de un número determinado de observaciones para descubrir aspectos o relaciones interesantes, no le interesa predecir, no tiene una variable de respuesta que guíe el proceso sino únicamente las observaciones y solo de ellas puede extraer información. El aprendizaje no supervisado se refiere a un conjunto de técnicas que pueden descubrir subgrupos en las observaciones o en las variables, descubrir si hay una forma de

visualizar los datos o encontrar una estructura (Géron, 2019), reglas de aplicación o detectar anomalías. Muchas veces se realiza como parte de un proceso de análisis exploratorio de datos.

Estos son algunos análisis que se usan en este tipo de aprendizaje:

Análisis de componentes principales (PCA): Es un procedimiento que busca reducir la dimensionalidad de los datos conservando la mayor cantidad de información posible del conjunto original (James et al., 2021), no requiere supuestos y tampoco requiere conocer la distribución de probabilidad de los datos. Le interesa descubrir la dependencia o interdependencia entre las variables para poder resumir o reducir una gran cantidad de ellas en algunas pocas que contienen la información de las demás, facilitando su análisis.

Agrupamiento: es un gran conjunto de técnicas que busca encontrar subgrupos en los datos. Estos grupos tienen la particularidad de que sus correspondientes observaciones son muy similares entre si pero muy diferentes a otros grupos, el agrupamiento se hace por similitud, proximidad o densidad (James et al., 2021). Los dos mejores métodos representativos de este tipo son el K-means, donde se busca particionar las observaciones en un número determinado de grupos y el agrupamiento jerárquico que busca mediante un árbol llamado dendograma descubrir los grupos que pueden existir en los datos usando distancias Euclidianas o de Manhattan.

- **Aprendizaje por refuerzo**

El aprendizaje por refuerzo (Reinforcement Learning, RL) pertenece el área del aprendizaje automático que se ocupa de la toma de decisiones secuenciales. El RL trata de maximizar una señal de recompensa en lugar de tratar de encontrar una estructura oculta, para la cual debe escoger entre dos acciones: explorar o explotar. Para la fase de explotación el agente ejecuta la acción que considere óptima en el momento de encontrarla, ideal para ir obteniendo la recompensa deseada. Para la fase de exploración el agente explora las distintas posibilidades de ejecución de acciones en situaciones que se presentan en el entorno (Sutton & Barto, 2014). Decidir entre las dos fases es crucial para el rendimiento del agente.

- **Aprendizaje profundo**

El aprendizaje profundo busca aprender representaciones de datos (Goodfellow et al., 2016), con múltiples niveles de abstracción mediante una serie de transformaciones lineales y no lineales, que generen una salida que difiera muy poco con la esperada.

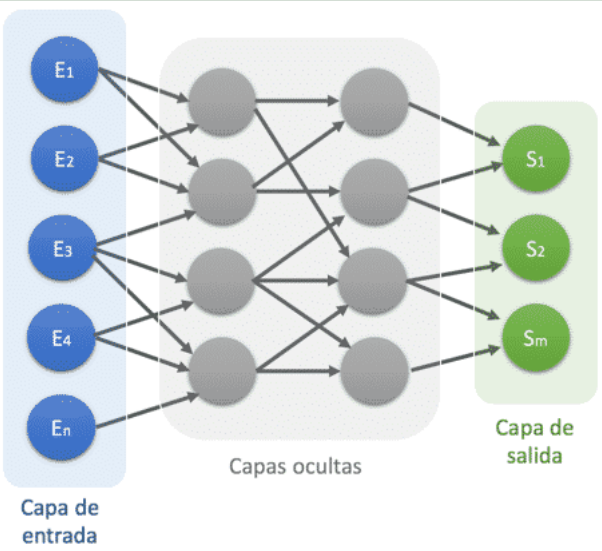

Figura 2: Capas de una red neuronal.

Una aproximación gráfica de una red MLP se muestra en la figura 2[2], representa una red neuronal de tres capas: una de entrada, la cual recibe los datos, una de salida que devuelve la predicción realizada y dos capas ocultas en medio. En la figura se observan varias interconexiones entre capas y neuronas ocultas.

- **Optimización**

La optimización se refiere a encontrar los parámetros o argumentos de una función que produce un máximo o mínimo valor de esa función. Existen numerosos algoritmos de optimización, sin embargo en el ámbito de este proyecto los de mayor interés corresponden al tipo de optimización de funciones continuas y se describen a continuación.

El primer algoritmo se denomina Gravitational Search Algorithm GSA(Rashedi et al., 2009) y se basa en las leyes de la gravitación universal de Newton, consta de objetos llamados agentes cuyo rendimiento se mide por sus masas y se atraen entre sí por la fuerza de la gravedad. Esta fuerza provoca un movimiento global de todos ellos hacia los objetos con masas más pesadas, de tal manera que las masas cooperan entre sí a través de la fuerza gravitacional. Las masas pesadas, que corresponden a buenas soluciones, se mueven más lentamente que las más ligeras, lo que garantiza el paso de explotación del algoritmo. En un lapso de tiempo, se espera que las masas sean atraídas por la masa más pesada. Esta masa presentará una solución óptima en el espacio de búsqueda.

[2]Figura extraída de https://www.diegocalvo.es/wp-content/uploads/2017/07/neural-network.png

Definición de la posición del i-ésimo agente:

$$X_i = \left(x_i^1, ..., x_i^d, ..., x_i^n\right) \; para \; i = (1, 2, ...N) \tag{1}$$

En el tiempo específico t, se define la fuerza que actúa sobre la masa i desde la masa j así:

$$F_{ij}^d(t) = G(t) \frac{M_{pi}(t) \times M_{aj}(t)}{R_{ij}(t) + \varepsilon} \left(x_j^d(t) - x_i^d(t)\right) \tag{2}$$

donde M_{aj}, es la masa gravitacional activa relacionada con el agente j que corresponde a la masa generadora del campo gravitacional, M_{pi} es la masa gravitacional pasiva relacionada con el agente i que responde al campo gravitacional generado por M_{aj}, $G(t)$ es la constante gravitacional en el instante t, ε es una pequeña constante y $R_{ij}(t)$ es la distancia Euclidiana entre los agentes i y j:

$$R_{ij}(t) = \left|\left| X_i(t), X_j(t) \right|\right|_2 \tag{3}$$

Para dar una característica estocástica al algoritmo, se supone que la fuerza total que actúa sobre el agente i en una dimensión d sea una suma ponderada al azar de los d-ésimos componentes de las fuerzas ejercidas por otros agentes:

$$F_i^d(t) = \sum_{j=1, j \neq i}^{N} rand_j F_{ij}^d(t) \tag{4}$$

donde $rand_j$ es un número aleatorio entre $[0, 1]$. Ahora bien, por la ley del movimiento, la aceleración del agente i en el tiempo t, y en la dirección d-ésima, $a_i^d(t)$, se da de la siguiente manera:

$$a_i^d(t) = \frac{F_i^d(t)}{M_{ii}^t} \tag{5}$$

Además, la siguiente velocidad de un agente se considera como una fracción de su velocidad actual sumada a su aceleración. Por lo tanto, su posición y su velocidad podrían calcularse de la siguiente manera:

$$v_i^d(t+1) = rand_i \times v_i^d(t) + a_i^d(t), \tag{6}$$

$$x_i^d(t+1) = x_i^d(t) + v_i^d(t+1) \tag{7}$$

donde $rand_j$ es una variable aleatoria uniforme en el intervalo $[0, 1]$. Se usa este número aleatorio para dar un carácter aleatorio a la búsqueda.

La constante gravitacional G, se inicializa al principio y se reducirá con el tiempo para controlar la precisión de la búsqueda. En otras palabras, G es una función con valor inicial G_0 y el tiempo (t):

$$G\left(t\right) = G\left(G_0, t\right), \tag{8}$$

Las masas gravitacionales y de inercia se calculan simplemente mediante la evaluación de la aptitud física. Una masa más pesada significa un agente más eficaz. Esto significa que los mejores agentes tienen mayores atracciones y caminan más lentamente. Suponiendo la igualdad de la masa gravitacional e inercial, los valores de las masas se calculan utilizando el mapa de aptitud. Se actualizan las masas gravitacionales e inerciales mediante las siguientes ecuaciones:

$$M_{ai} = M_{pi} = M_{ii} = M_i,\ i = 1, 2, ..., N, \tag{9}$$

$$m_i\left(t\right) = \frac{fit_i\left(t\right) - peor\left(t\right)}{mejor\left(t\right) - peor\left(t\right)}, \tag{10}$$

$$M_i\left(t\right) = \frac{m_i\left(t\right)}{\sum_{j=1}^{N} m_j\left(t\right)}, \tag{11}$$

donde $fit_i\left(t\right)$ representa el valor de aptitud del agente i en el tiempo t, y, $peor\left(t\right)$ y $mejor\left(t\right)$ se definen de la siguiente manera (para problemas de minimización):

$$mejor\left(t\right) = \min_{j\in\{1,...,N\}} fit_j\left(t\right), \tag{12}$$

$$peor\left(t\right) = \max_{j\in\{1,...,N\}} fit_j\left(t\right) \tag{13}$$

Para un problema de maximización se invierten las ecuaciones (12) y (13) así:

$$peor\left(t\right) = \min_{j\in\{1,...,N\}} fit_j\left(t\right), \tag{14}$$

$$mejor\left(t\right) = \max_{j\in\{1,...,N\}} fit_j\left(t\right) \tag{15}$$

Una forma de realizar un buen compromiso entre la exploración y la explotación es reducir el número de agentes con lapso de tiempo, en la ecuación (4). Por lo tanto, se propone que solo un conjunto de agentes con mayor masa aplique su fuerza al otro. Sin embargo, se debe tener tener cuidado al usar esta política porque puede reducir el poder de exploración y aumentar la capacidad de explotación.

Para evitar el atrapamiento en un óptimo local, el algoritmo debe usar la exploración al principio. Después de algunas iteraciones, la exploración debe desaparecer y la explotación debe aparecer. Para mejorar el rendimiento de GSA controlando la exploración y explotación, solo los mejores agentes atraerán a los demás. $Kmejor$ es una función del tiempo, con valor inicial K_0 al principio y disminuyendo con el tiempo. De tal manera que, al principio, todos los agentes aplican la fuerza, a medida que pasa el tiempo, $Kmejor$ disminuye linealmente y al final solo habrá un agente que aplique fuerza a los demás. Por lo tanto la ecuación (4) se puede modificar así:

$$F_i^d(t) = \sum_{j \in Kbest, j \neq i}^{N} rand_j F_{ij}^d(t) \tag{16}$$

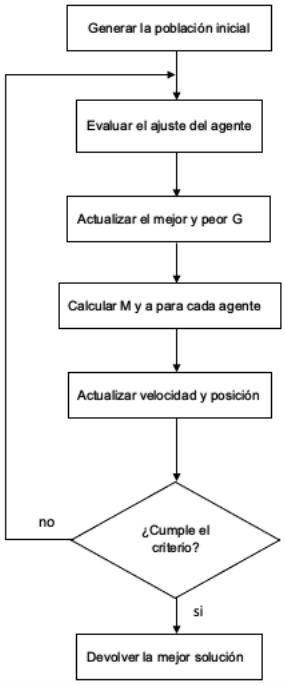

Figura 3: Flujo del algoritmo GSA.

donde $Kbest$ es el conjunto de los primeros agentes de K con el mejor valor de ajuste y la mayor masa. Los pasos del algoritmo se pueden observar en la figura 3.

El segundo algoritmo se denomina Particle Swarm Optimization PSO (Kennedy et al., 2001) y está basado en el movimiento de bandadas de pájaros o peces. En general trata de optimizar un solución a partir de un grupo de soluciones candidatas llamada partículas, moviéndolas por todo el espacio de búsqueda utilizando su posición y velocidad. El movimiento de cada partícula se ve afectado por su mejor posición local y la mejor posición global de todo el grupo y el objetivo es hacer que toda la nube converja de una manera rápida hacia las

mejores soluciones. La tasa de cambio de posición se calcula con la siguiente ecuación:

$$V_{id} = W * V_{id} + c_1 * rand\,[0,1] * (P_{id} - X_{id}) + c_2 * rand\,[0,1] * (G_{id} - X_{id}) \tag{17}$$

donde V_{id} es la velocidad de la partícula, W es la inercia, c_1, c_2 son constantes, $rand\,[1,0]$ son números aleatorios entre 0 y 1, X_{id} es la solución actual de cada individuo, P_{id} es el mejor propio, la mejor solución de cada individuo y G_{id} es el mejor global, la mejor solución de toda la población. La nueva posición de la partícula sería:

$$X_{id} = X_{id} + V_{id} \tag{18}$$

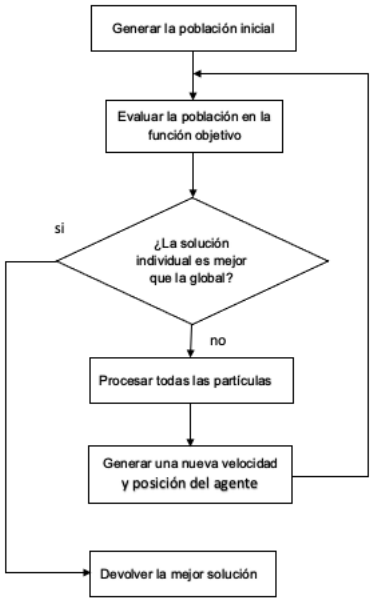

Figura 4: Flujo del algoritmo PSO.

Los pasos del algoritmo se pueden observar en la figura 4.

- **Métricas de evaluación**

Según Koutsandreas (Koutsandreas et al., 2021) las medidas de precisión de pronóstico puntual catalogadas por el tipo de escala (media de la serie, primeras diferencias de la serie o rendimiento relativo frente a un método de referencia) son las usadas generalmente para clasificar las medidas de precisión de pronóstico, pueden ser dependientes de la escala, basadas en errores porcentuales, basadas en errores y medidas relativas y basadas en errores escalados. En este estudio se utilizan dos medidas dependientes de la escala (RMSE y MAE) y otra basada en errores porcentuales (MAPE).

Root mean squared error (RMSE) es la raíz cuadrada del error cuadrado medio y se define así:

$$RMSE = \sqrt{\frac{1}{h}\sum_{i=1}^{h} e^2} \tag{19}$$

donde e^2 es el cuadrado de la diferencia entre el valor estimado y el valor real y h es el horizonte de tiempo del pronóstico.

Mean absolute error (MAE) es el error medio absoluto y se define como:

$$MAE = \frac{1}{h}\sum_{i=1}^{h} \mid e_i \mid \tag{20}$$

donde e_i es la distancia entre el valor estimado y el valor real y h es el horizonte del pronóstico.

Con el fin de hacer que los errores sean independientes de la escala una práctica es expresarlos como porcentajes, ese es el caso de *mean absolute percentage error* (MAPE), el error absoluto porcentual medio que se define como:

$$MAPE = \frac{1}{h}\sum_{i=1}^{h} \mid p_i \mid \tag{21}$$

donde p_i corresponde al valor del error e_t dividido por el valor real y_t en porcentaje:

$$p_t = \frac{e_t}{y_t} \times 100\,\% \tag{22}$$

Estas medidas se utilizarán en el proceso de evaluación de los modelos aplicados en el pronóstico de electricidad de largo plazo en este estudio.

2.2. Estado del Arte

A continuación, se presentan estudios realizados en diferentes países y diferentes años principalmente en temas de análisis de series de tiempo y técnicas de aprendizaje automático, puesto que estos temas resultan fundamentales para el planteamiento del modelo que permita predecir la demanda de energía eléctrica en un horizonte de mediano-largo plazo. Las palabras claves utilizadas en las búsquedas realizadas incluyeron: inteligencia artificial, aprendizaje automático, modelos econométricos, predicción en series temporales, pronóstico de demanda de electricidad.

2.2.1. Trabajos seleccionados

Se encontró mucha literatura relacionada con el pronóstico de demanda de electricidad de corto plazo, a diferencia de la encontrada para largo plazo, sin embargo los enfoques en este aspecto son muy novedosos y variados. Se seleccionaron los siguientes:

1. Técnicas de predicción basadas en Inteligencia Artificial

   a) *Forecasting of Electricity Demand by Hybrid ANN-PSO Models* (Anand & Suganthi, 2017).

   Esta propuesta fue aplicada en la India, con el fin de llenar un vacío en el pronóstico de la demanda de energía eléctrica de largo plazo. En la primera etapa se utilizaron técnicas de optimización usando datos desde 1991 hasta 2000, como optimización por nube de partículas (PSO) y algoritmos genéticos (GA) en formas lineales y cuadráticas que posteriormente fueron utilizados para entrenar redes neuronales artificiales (ANN) y proyectar con estas la demanda entre 2001 y 2015. Se compararon con técnicas de series de tiempo ARIMA y combinaciones ANN-BP (redes neuronales artificiales con backpropagation) donde ANN-PSO en sus dos formas lineal y cuadrática dieron los mejores resultados, basados en los indicadores RMSE y MAPE, haciendo pronósticos de consumo de electricidad más acertados.

   En la segunda etapa se usó este modelo seleccionado como el mejor para hacer pronóstico de demanda de energía hasta 2020. Básicamente cada red neuronal (NN) define los atributos de posición y velocidad. La posición está relacionada con el peso de la red neuronal. La velocidad se refiere a la actualización de los pesos de la red. La función de PSO es obtener el mejor conjunto de pesos (posición de partícula) donde varias partículas intentan moverse para obtener la mejor solución. Para la implementación de la red neuronal, el valor de aptitud corresponde a una propagación hacia adelante a través de la red y al vector de posición de la red. El mejor vecino de la partícula y la mejor partícula global se utilizan para guiar las nuevas soluciones. Al final, la posición de la mejor partícula global es el modelo deseado.

   b) *Predicting long-term monthly electricity demand under future climatic and socio-economic changes using data-driven methods: A case study of Hong Kong* (Liu et al., 2021).

   Este estudio se aplicó en Hong Kong con el fin de cuantificar el impacto de cambios socioeconómicos y climáticos en la demanda de energía eléctrica mensual a largo plazo. Se compararon diferentes técnicas orientadas por datos como máquinas de vectores de soporte (SVM), arboles de decisión (DT), redes neuronales artificiales

(ANN), árboles de decisión de impulso de gradiente (GBDT), procesos de regresión gaussianos (GPR) y regresión lineal múltiple (MLR).

Se utilizaron datos históricos de 40 años para entrenar y validar los diferentes modelos, posteriormente para comparar los rendimientos entre los diferentes algoritmos, se seleccionaron el error cuadrático medio (RMSE) y el error de sesgo medio normalizado (NMBE) para cuantificar las desviaciones entre los valores predichos y los valores reales. Se obtuvo que el modelo GBDT resultó ser el más adecuado para el pronóstico de la demanda de energía eléctrica de largo plazo en Hong Kong, determinando un crecimiento de la demanda en un 89,4 % para el año 2090 en comparación con el 2018.

2. Modelos econométricos de predicción

   a) *Forecasting residential electric power consumption for Bogotá Colombia using regression models* (Peña-Guzmán & Rey, 2020).

   Este estudio fue aplicado en Bogotá, con el fin de identificar las variables que tienen mayor impacto en la predicción de la demanda de energía. Se utilizaron modelos de regresión múltiple econométrica y regresión lineal de doble logaritmo, teniendo en cuenta la segmentación por estratos socio económicos que se da en Colombia para los servicios públicos.

   Se utilizó como data el consumo histórico plurianual de electricidad entre los años 2005 y 2016. Se encontró como resultado que las variables que más impacto tienen en la predicción de la demanda de energía para el primer proceso son los usuarios finales, la temperatura de superficie y el precio de la energía eléctrica y para el segundo (econométrico) fue la elasticidad de precios, el PIB y los usuarios.

   b) *Demand and supply-side determinants of electric power consumption and representative roadmaps to 100 % renewable systems* (Ma et al., 2021).

   Este estudio se aplicó en Suecia, con el fin de involucrar los determinantes de la demanda y la oferta para investigar las relaciones causales a corto y largo plazo entre el sistema de energía eléctrica, el desempeño macroeconómico, la demografía, la calidad ambiental y la formación de capital. Se implementaron modelos econométricos y de aprendizaje automático de dos etapas basados en la atención de la volatilidad-consistencia para la predicción de energía eléctrica. Se usó un modelo de memoria larga de corto plazo (LSTM), un tipo de red neuronal recurrente (RNN).

   La precisión del modelo de previsión de heteroscedasticidad y autocorrelación coherentes (HAC), basado en datos, se estimó dividiendo los conjuntos de datos

relacionados con la energía eléctrica en dos; con datos que cubren desde 1990 a 2013 para el entrenamiento, y del 2014 a 2018 para la validación. Se utilizó un conjunto de validación de cinco años para ajustarse a la definición generalmente aceptada de largo plazo en el sector de la energía eléctrica. El error porcentual medio absoluto (MAPE) se utilizó como métrica para medir el error de previsión registrado en el conjunto de validación y generar una proyección de la demanda hasta 2050.

3. Técnicas de predicción en series temporales

   a) *Electricity Demand Time Series Forecasting Based on Empirical Mode Decomposition and Long Short-Term Memory* (Taheri et al., 2021).

   Este estudio fue realizado para la predicción de la demanda de energía para el Sistema Operador Independiente de California, que surte los estados de California y Nevada para 2.5 años entre 2018 y 2021. Se utilizaron técnicas de aprendizaje automático híbridas como memoria de corto plazo unida a técnicas de procesamiento de señales conocida como descomposición empírica (LSTM + EMD), regresión lineal (LR) y uno de los algoritmos mas usados en la actualidad como es el modelo llamado impulso extremo de gradiente XGBoost.

   Para evaluar la calidad de los modelos se adopta sobre todo la precisión, que incluye el error medio absoluto (MAE), el error cuadrático medio (RMSE), el error porcentual medio absoluto (MAPE) y el coeficiente de determinación (R^2). El modelo LSTM + EMD Híbrido superó con mejores resultados al XGBoost y LR en predicciones de demanda de electricidad de corto y largo plazo.

2.2.2. Resumen de criterios

Para realizar esta selección, fue necesario establecer algunos criterios con el fin de encontrar las publicaciones que mejor se aproximaran a los objetivos del proyecto, tales como:

- Dimensión temporal: antigüedad inferior a cinco años.
- Geografía: realizados en diferentes países y de distintos niveles socioeconómicos.
- Técnicas y herramientas: utilización de modelos o combinaciones de modelos de inteligencia artificial actualizados y que tuvieron resultados exitosos.
- Proceso: clasificación del estudio realizado con base en búsqueda de determinantes o en predicción.
- RMSE: el error cuadrado medio de la desviación entre el valor observado y el valor predicho mas bajo obtenido en el estudio.

- Horizonte: las ventanas de tiempo en las que se hace el pronóstico.
- Contexto: relacionados con el consumo o demanda de electricidad y su pronóstico a largo plazo.

Tabla 2: Resumen de los criterios de comparación entre los artículos seleccionados.

Título	**Dimensión Temporal**	**Geografía**	**Técnicas**	**Proceso**	**RMSE**	**Horizonte**	**Contexto**
1. Demand and supply-side determinants of electric power consumption and representative roadmaps to 100 % renewable systems. (Ma et al., 2021)	2021	Suecia	LSTM-RNN	Econométrico Predictivo	N/A	2020 - 2050	Transformación de sistemas eléctricos tradicionales a fuentes limpias.
2. Predicting long-term monthly electricity demand under future climatic and socioeconomic changes using data-driven methods: A case study of Hong Kong. (Liu et al., 2021)	2021	Hong Kong	ANN GBDT GPR MLR	Basado en datos Predictivo	SVR 369.9767	2026 - 2045 2056 - 2075 2080 - 2099	Impacto de cambios socioeconómicos en la demanda mensual de electricidad a largo plazo. Residencial
3. Forecasting residential electric power consumption for Bogotá-Colombia using regression models. (Peña-Guzmán & Rey, 2020).	2020	Bogotá	MLR2Log	Econométrico Predictivo	N/A	2005 - 2016	Variables de mayor impacto en el consumo de la demanda de electricidad.
4. Forecasting residential electricity consumption using a hybridmachine learning model with online search data. (Gao et al., 2021)	2021	China	SVR SARIMA(X) BPNN ELM	Predictivo	Jaya-ELM 3.78	2011 - 2019	Necesidad de precisión en el pronóstico del consumo de electricidad domiciliaria.
5. Forecasting of Electricity Demand by Hybrid ANN-PSO Models. (Anand & Suganthi, 2017)	2017	India	ARIMA ANN-PSO GA ANNBP	Predictivo	ANN-PSO 1.94	2001 - 2020	Necesidad de proyección de consumo de electricidad por gestación de proyectos energéticos para cerrar la brecha de abastecimiento. Residencial

Tabla 2: Resumen de los criterios de comparación entre los artículos seleccionados.

Título	**Dimensión Temporal**	**Geografía**	**Técnicas**	**Proceso**	**RMSE**	**Horizonte**	**Contexto**
6. The estimation of the electricity energy demand using PSO algorithm: case study of Turkey. (Gulcu & Kodaz, 2017)	2017	Turquía	ANN-PSO	Predictivo	6486.59	2004 - 2013 2014 - 2030	Cerrar brecha de abastecimiento de electricidad por crecimiento económico.
7. Electricity Demand Time Series Forecasting Based on Empirical Mode Decomposition and Long Short-Term Memory. Taheri et al. 2021	2021	USA	LSTM- EMD LR XGBoost	Predictivo	278.76	48 h 24 h 1 week 1 month	Necesidad de precisión en el pronóstico del consumo de electricidad por crecimiento poblacional.

3. METODOLOGÍA

3.1. Esquema de trabajo

La metodología elegida para abordar el proyecto es CRISP-DM (Cross-Industry Standard Process for Data Mining), ver figura 5[3]. Esta metodología fue desarrollada por el consorcio formado por las empresas NCR Systems Engineering Copenhagen (USA y Dinamarca), DaimlerChrysler AG (Alemania), SPSS Inc. (USA) y OHRA Verzekeringen en Bank Groep B.V (Holanda) y publicada en el año 2000 (Chapman et al., 2000).

Las fases vienen dadas en un orden secuencial dentro del proceso genérico, sin embargo, su orden y ejecución pueden variar dependiendo de las características propias del proyecto y de sus actividades (Wirth, 2000).

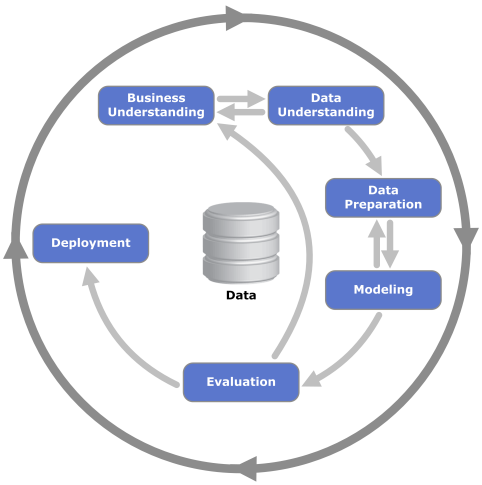

Figura 5: Metodología CRISP-DM.

3.2. Fases del Proyecto

A continuación, se presenta una breve explicación de lo que se realiza en cada una de las fases del proyecto.

3.2.1. Fase 1: Comprensión del negocio

Es la fase inicial del ciclo de vida de un proyecto de analítica. En esta fase se busca conocer los requerimientos y objetivos desde el punto de vista del negocio, con que recursos se cuenta y cual será su alcance en función de la información recopilada por la organización.

[3]Imagen tomada de https://es.wikipedia.org/wiki/Cross_Industry_Standard_Process_for_Data_Mining

3.2.2. Fase 2: Comprensión de los datos

Esta fase busca la familiarización con los datos que van a ser objeto de análisis. Comprende varias actividades, inicialmente se debe recolectar la data, después se realiza su descripción, para posteriormente hacer la exploración y verificación de los datos. En el caso de la predicción del consumo de energía que se pretende obtener, es importante familiarizarse con los datos, teniendo en cuenta que para una posterior fase es necesario verificar su calidad, debido a que los datos son de fuentes públicas sin un proceso de aseguramiento de la calidad óptimo.

3.2.3. Fase 3: Preparación de los datos

Durante esta fase se busca trabajar con los datos crudos hasta convertirlos en datos que puedan ser leídos por los modelos. Incluye todas las actividades de selección, limpieza y transformación de datos, para darles el formato requerido, necesario para la aplicación en herramientas del modelado.

3.2.4. Fase 4: Modelado

Para esta etapa los datos ya se encuentran listos para ser incorporados a una técnica de modelado de ML, pasan a una serie de pruebas validando con indicadores de desempeño (KPI, Key Performance Indicator) previamente definidos. A continuación, se realiza el ajuste de parámetros necesario para la construcción del modelo, validando su efectividad para que pueda generar los insights aprovechables.

3.2.5. Fase 5: Evaluación

En esta fase, se busca evaluar los resultados de los modelos aplicados con el fin de validar si cumplen con los objetivos del negocio. Se revisa tanto el modelo como los pasos que llevaron a su construcción, con el fin de detectar cualquier faltante en los objetivos enunciados en la fase de "entendimiento del negocio".

3.2.6. Fase 6: Despliegue

Como etapa final, se busca poder compartir y difundir los resultados del proyecto, para que puedan ser usados por los usuarios del negocio. El despliegue de un proyecto de analítica/minería depende de los resultados y del objetivo inicial del mismo y abarca desde la presentación de un informe detallado, hasta la implementación de un proceso de analítica, inmerso en el proceso de toma de decisiones de una organización. Para este proyecto, el alcance propuesto comprende la producción de un reporte final, donde se encuentren los resultados obtenidos y las experiencias alcanzadas.

4. COMPARACIÓN DE MODELOS DE APRENDIZAJE AUTOMÁTICO

4.1. Recolección de los datos

Para resolver la pregunta de investigación se utilizan diferentes tipos de datos obtenidos de fuentes como la empresa de servicio de energía de Cali EMCALI EICE, del DANE y del CIAT, en un periodo de tiempo de 15 años desde 2007. Es de anotar que para la variable dependiente se debió asegurar a EMCALI que no se usarían datos sensibles de personas o de la misma empresa, de tal manera que solo se extrajeron datos de dominio público del mercado de energía.

Figura 6: Localización de los sensores de temperatura del CIAT. Fuente: Google Maps.

Por otro lado se decide incluir datos de tres temperaturas en el día, pertenecientes al CIAT, dado que en ellas se puede observar la variación de la temperatura ambiente en momentos claves muy cercanos al alba, cenit y ocaso, esta base de datos estaba prácticamente completa a diferencia de las que se obtuvieron del IDEAM y del DAGMA donde faltan muchos años y muchas lecturas en las estaciones, en la figura 6 se muestra la localización de los sensores. A continuación se describen brevemente las variables recolectadas y sus fuentes:

1. Consumo de Energía: Esta es la variable dependiente del problema de regresión de esta investigación, se obtuvo del sistema de facturación de EMCALI EICE, desde enero de 2007 hasta diciembre de 2021, de forma mensual en kilovatios hora (KWh).

2. Producto Interno Bruto del departamento del Valle del Cauca: esta variable por ciudad es determinante de la demanda de energía en varios trabajos previos (Peña-Guzmán & Rey, 2020), sin embargo para la ciudad de Cali no existe esta discriminación aunque hasta el año de 2009, la universidad ICESI estuvo identificando el cálculo porcentual de

la participación de Cali en el total nacional, por tanto se tomó como referente el PIB del departamento del Valle, el cual se obtuvo de los archivos históricos del DANE.

3. Temperatura 07H: esta variable corresponde a la temperatura recolectada por el CIAT a las 7 de la mañana durante el periodo de 15 años, localizado el sensor en el área del km 17 de la recta a Palmira (3.502689856641904, -76.35488515631718).

4. Temperatura 13H: esta variable corresponde a la temperatura recolectada por el CIAT a la 1 de la tarde durante el periodo de 15 años, localizado el sensor en el area del km 17 de la recta a Palmira (3.502689856641904, -76.35488515631718).

5. Temperatura 19H: esta variable corresponde a la temperatura recolectada por el CIAT a las 7 de la noche durante el periodo de 15 años, localizado el sensor en el area del km 17 de la recta a Palmira (3.502689856641904, -76.35488515631718).

6. Categoría del servicio: esta variable corresponde a la identificación del tipo de servicio que presenta el consumo de energía eléctrica, corresponde a uso residencial que representa el consumo en los hogares, uso comercial que representa el consumo de locales comerciales, industrial para empresas industriales, oficial para instituciones oficiales, especial para instituciones educativas, de salud o asistenciales.

7. Tipo de mercado de energía: esta variable indica si el consumo corresponde a usuarios del mercado regulado o no regulado. Esta diferenciación se puede observar en el capítulo de antecedentes de este escrito.

8. Estrato: por último se utiliza el estrato para aportar la clasificación socioeconómica de los consumos de energía. Esta característica se aplica en Colombia desde 1980 pero entró a formar parte jurídicamente de la ley 142 de servicios públicos de 1994, con el propósito de hacer que las clases más favorecidas subsidien a las menos favorecidas en el cobro la factura de servicios públicos.

9. Periodo: Corresponde al año/mes de las lecturas registradas.

4.2. Preparación de los datos

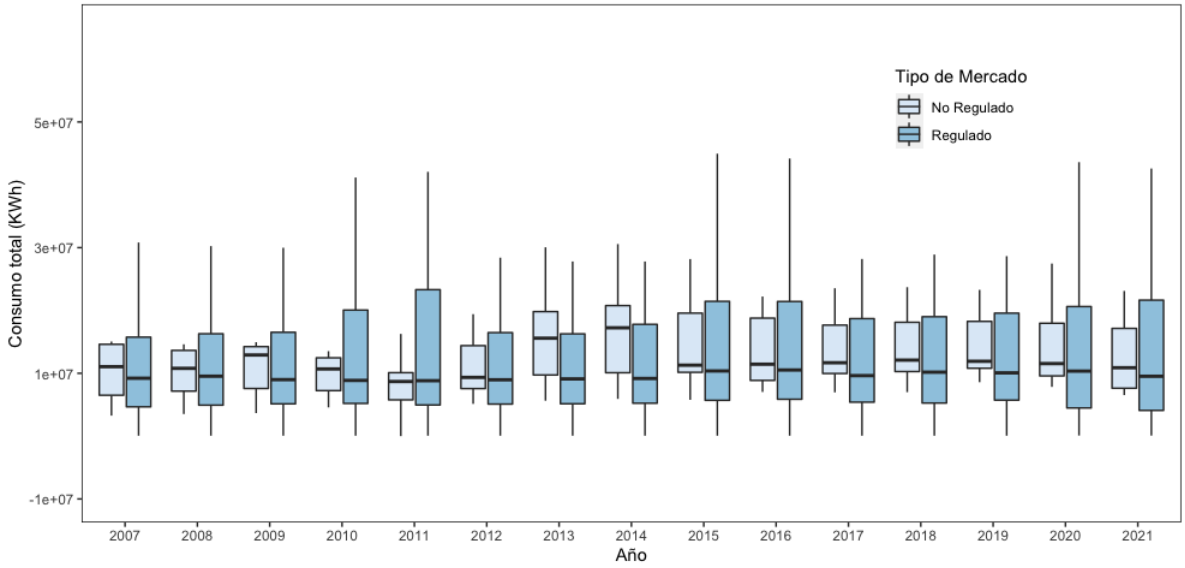

Figura 7: Distribución no normalizada de los consumos de energía por tipo de mercado.

Una vez que se consiguen los datos, se procede a realizar su limpieza y preparación, con el fin de tener la base definitiva con que se estimarán los modelos. Es de anotar que los consumos fueron entregados por producto, lo cual dio como resultado archivos de entre ochocientos mil y ochocientos cincuenta mil registros por mes, que debido a su tamaño, los recursos de computo necesarios para tratarlos resultaron insuficientes, por esta razón se agruparon por categoría, estrato y tipo de mercado, para adicionar información relevante y al mismo tiempo disminuir la granularidad sin pérdida de información, ver figura 7.

Como resultado de lo anteriormente expuesto fue necesario estandarizar los consumos para reducir su magnitud con centro en $\mu = 0$ y $\sigma = 1$, en la figura 8 se observa esta nueva distribución incluyendo las agrupaciones mencionadas excepto el estrato. Este mismo tratamiento se le realiza a las variables continuas con el fin de entrenar los modelos con datos estandarizados.

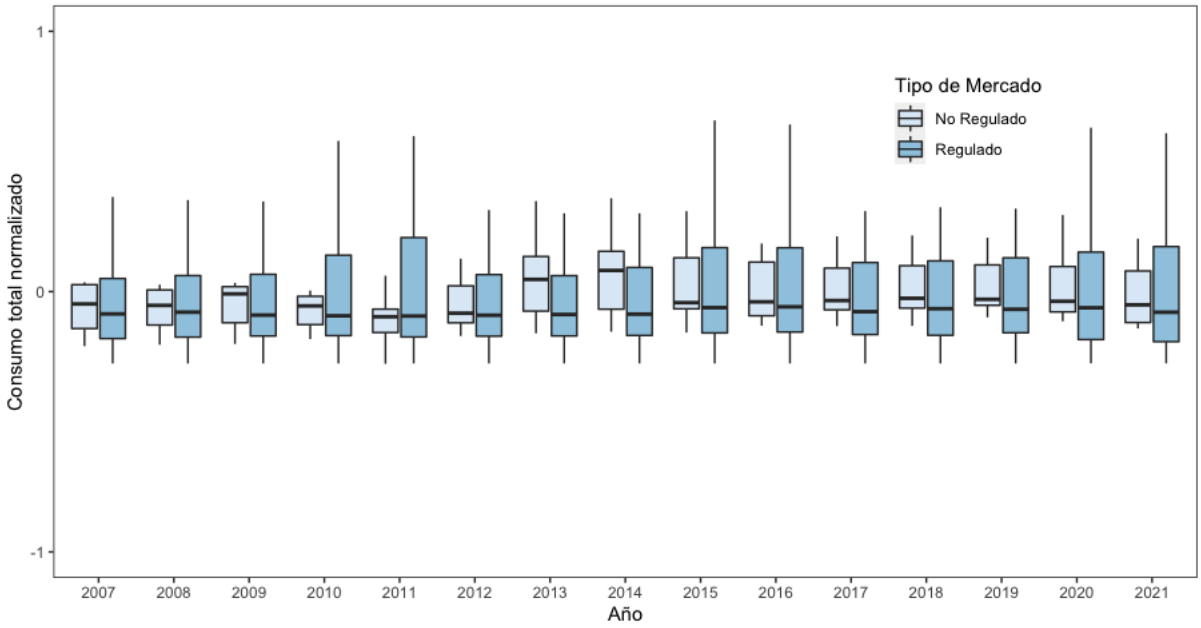

Figura 8: Distribución normalizada de los consumos de energía por tipo de mercado.

Por otro lado la variable período se convierte en un objeto de series de tiempo, con periodicidad de un mes con el fin de que indexe los datos. Esto redujo los datos de 3051 registros a 180.

A las variables de temperatura del CIAT, fue necesario realizarles un tratamiento que consistió en reemplazar alrededor de 20 ceros en 15 años de registros diarios, por el promedio del mes de cada una de las variables horarias donde se encontró el cero, con el fin de no alterar el promedio mensual de la temperatura. Estos datos faltantes se originaron en fallos de los registradores físicos.

Con respecto al PIB del DANE, este aparece de forma anual, por tal razón en la base de datos se coloca el mismo valor para todos los registros mensuales del mismo año.

4.3. Descripción de los datos

La variable dependiente de este estudio es el consumo de electricidad de los usuarios de Emcali que ocupan el área urbana de la ciudad de Cali, Yumbo y Puerto Tejada. Como se observa en la figura 9, el consumo total anual fluctúa entre los 2.500 y 2.900 GWh en el periodo de este estudio que inicia en enero de 2007 y termina en diciembre de 2021.

Otro aspecto importante que se introdujo en la figura 10, se refiere a la distribución del consumo por tipo de mercado de energía, en el cual se observa que el mayor componente es

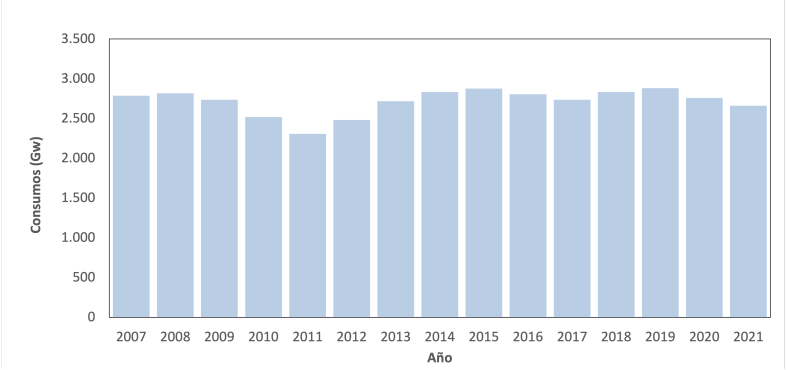

Figura 9: Distribución de los consumos totales por año en GWh.

el mercado regulado. En esta gráfica se observa que la reducción del consumo mencionada anteriormente, se puede explicar parcialmente por la disminución en la demanda en el mercado no regulado. Vale la pena mencionar que este mercado solamente tiene 364 suscriptores a marzo de 2022, pero la proporción de su consumo es muy alta comparada con el consumo del mercado regulado el cual se produce por más de seiscientos mil suscriptores.

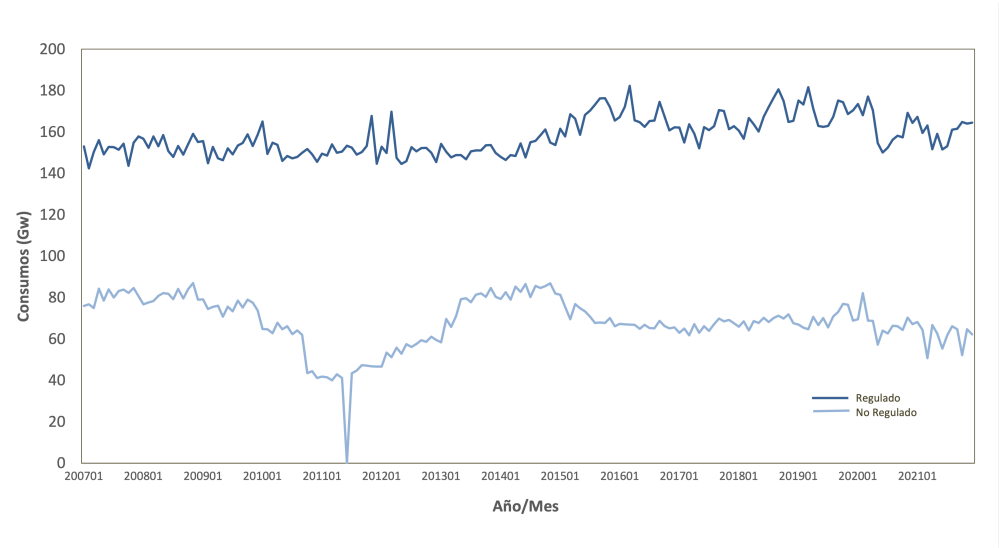

Figura 10: Distribución serializada del consumo por tipo de mercado GW.

Se observa una reducción a mínimos consumos entre 2010 y 2012 debido a la migración de clientes del mercado no regulado hacia la competencia por causas impositivas aplicadas a Emcali, adicionalmente por un mal cálculo en compra de energía, el comercializador se tuvo que exponer en bolsa afectando los precios de compra global lo cual condujo a incrementos de tarifa que hicieron perder la confianza en la marca. Después entre 2012 y 2013 el mercado deja de ser interesante para la competencia y estos clientes regresan a Emcali y por eso se incrementa nuevamente el consumo, para estabilizarse hasta 2019 y por efectos de la pandemia sobre el sector productivo decae nuevamente.

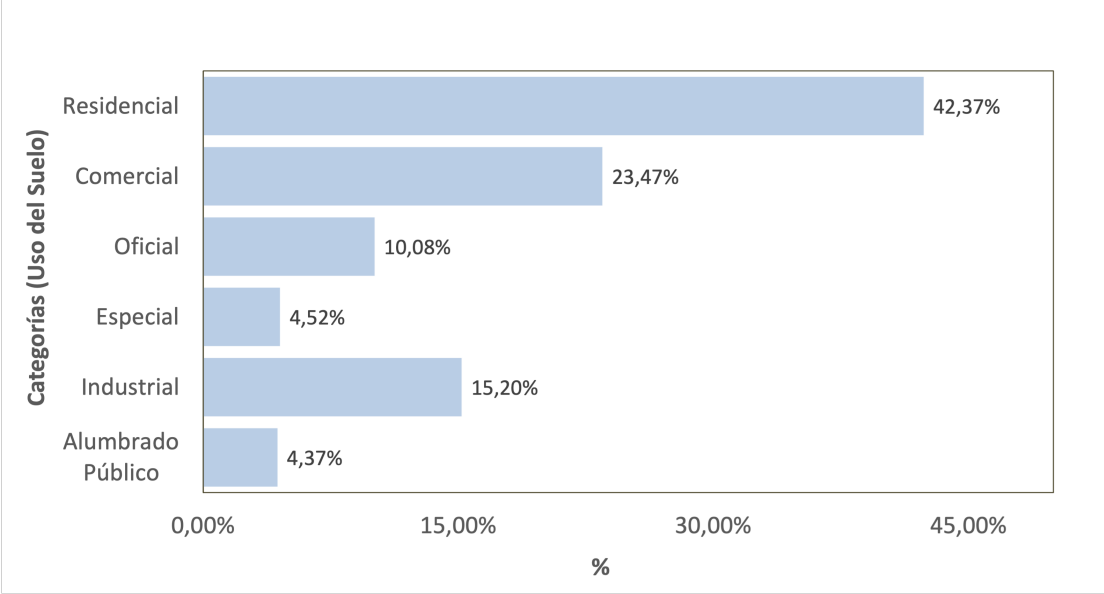

Figura 11: Composición de los consumos por categoría del servicio.

Es importante observar la composición del consumo por el uso del suelo o categoría del servicio. Como se observa en la figura 11, el componente residencial y el comercial consumen el 66 % de la electricidad. El alumbrado público y la categoría de uso especial que corresponde a los predios de hospitales, centros educativos, de salud y asistenciales, presentan un consumo considerable.

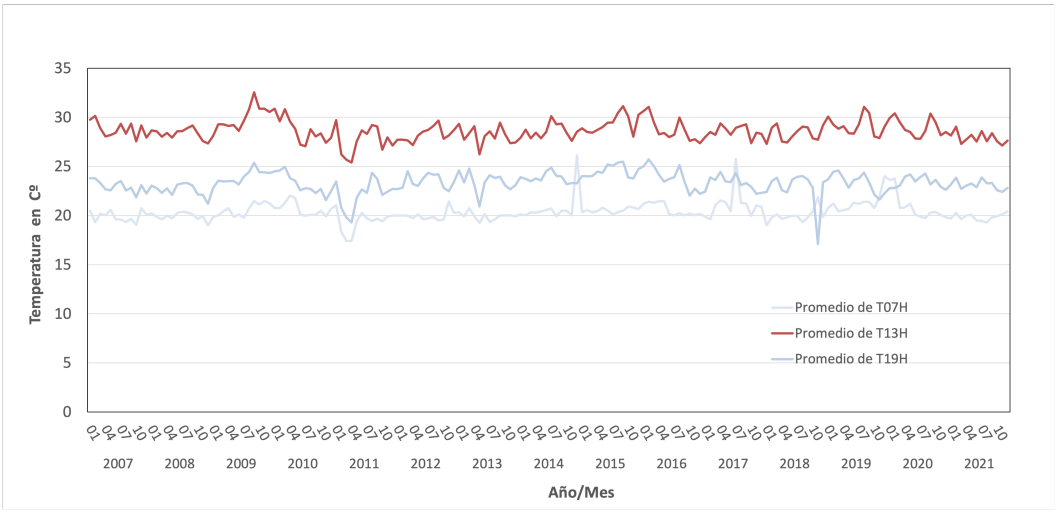

Figura 12: Temperaturas área CIAT cercanas al alba, cenit y ocaso, promedios mensuales durante los 15 años de recolección.

Un tercer aspecto de observación relevante, corresponde a la influencia de la temperatura ambiente en el consumo eléctrico especialmente en las categorías residencial y comercial, lo cual induce a hacer uso de ventiladores o dispositivos acondicionadores de aire en épocas de

mayor temperatura, en la gráfica de la figura 12 se observa incluso pequeño incremento en 2009, 2015 y 2019 por encima de la media.

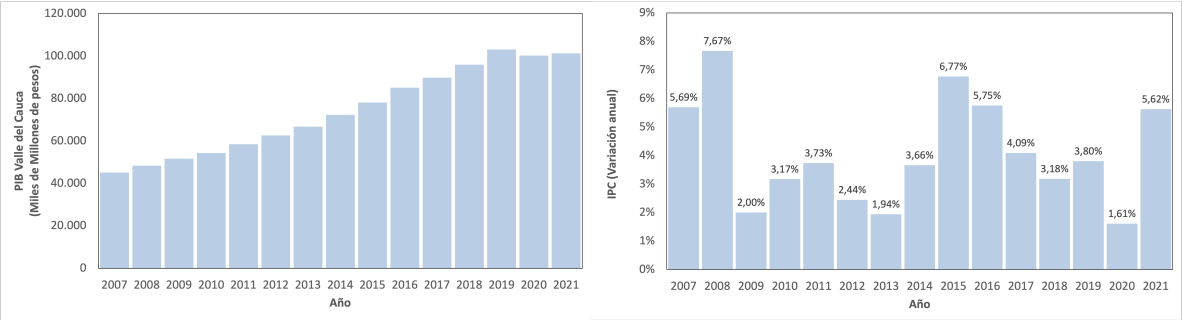

Figura 13: Crecimiento del PIB del Valle del Cauca y comportamiento del IPC en el periodo de recolección de los datos.

En cuarto lugar, se introduce el Producto Interno Bruto del departamento del Valle del Cauca, gráfica de la figura 13, en la que se observa un crecimiento sostenido lo cual podría dar indicios de afectación positiva en el incremento de la demanda de electricidad en el tiempo. Sin embargo se detuvo en 2020, claramente afectado por la pandemia de SARSCOV-2. Adicionalmente, se introduce la variación mensual del IPC que en la figura se integra de forma anual.

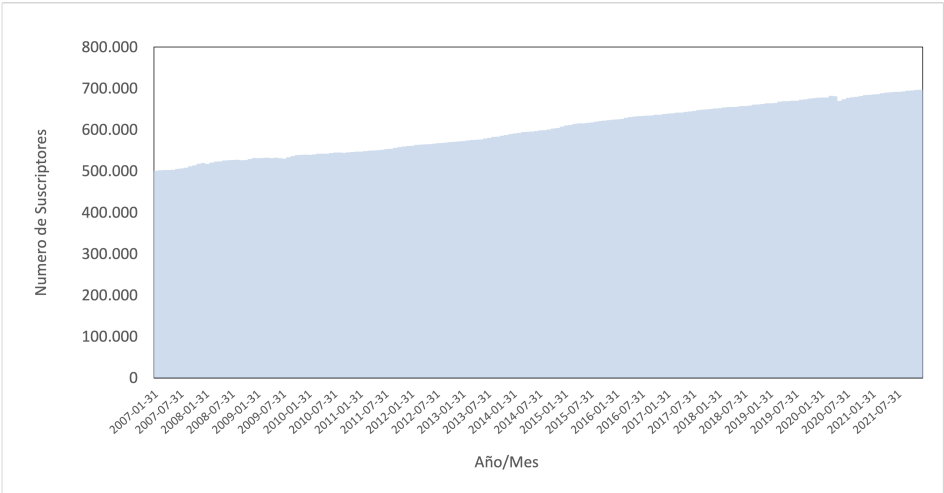

Figura 14: Distribución de los suscriptores en el tiempo.

En cuanto al número de suscriptores se observa en la figura 14 un crecimiento sostenido durante el periodo de estudio de 15 años. Finalmente en la matriz de correlaciones de la figura 15 se puede observar que las variables temperatura, cantidad de suscriptores y el producto interno bruto están correlacionadas de forma débilmente positiva con el consumo de electricidad (coeficiente de correlación de Pearson).

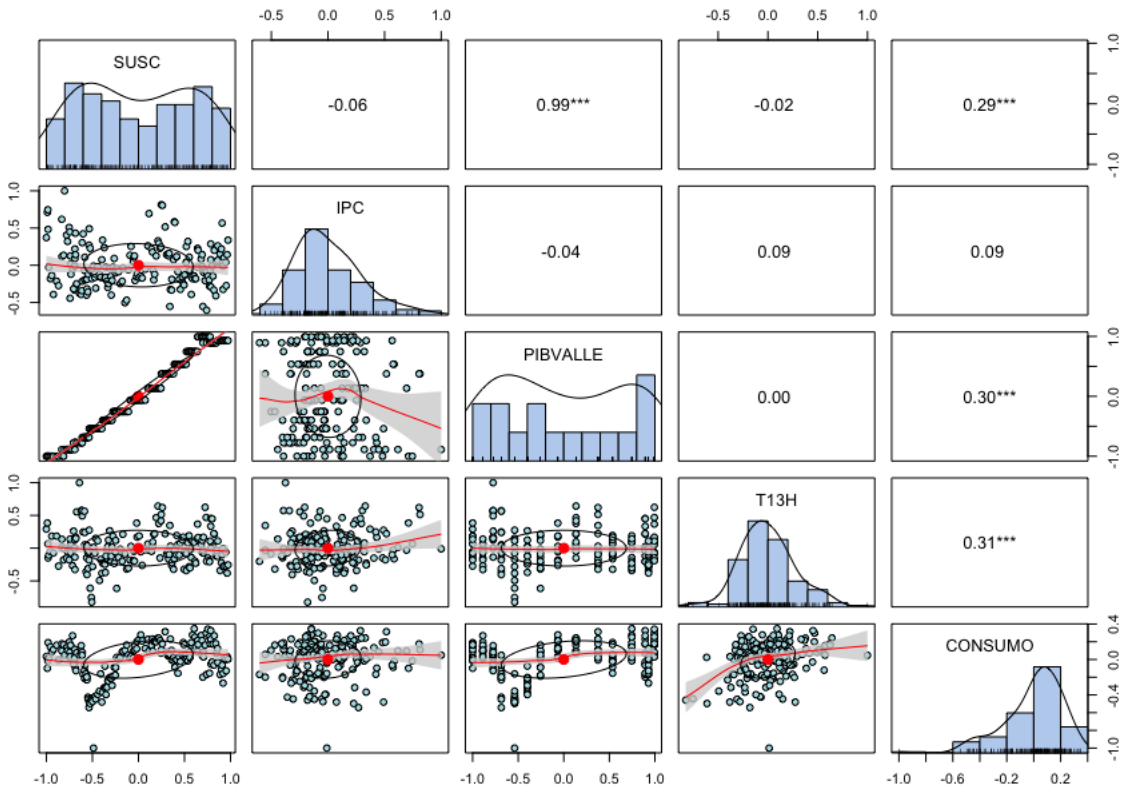

Figura 15: Correlación entre las variables del estudio.

Por otro lado la cantidad de suscriptores y el producto interno bruto del Valle están altamente correlacionados. Cabe mencionar que se normalizaron todas las variables para poder compararlas en la misma escala por un lado, pero también para mejorar el tiempo de ejecución de los modelos y evitar inconsistencias en los datos.

4.4. Modelamiento

Con el fin de alcanzar el objetivo de este estudio, se estimaron modelos de aprendizaje automático basados en las técnicas y configuraciones de conjuntos de datos univariados y multivariados que se presentan en la tabla 3.

En cuanto a los modelos de aprendizaje automático se utilizaron vectores autoregresivos (VAR), Random Forest, máquinas de vectores de soporte con kernel Radial y Lineal (SVM) (James et al., 2021), a los cuales se les realizó optimización de hiperparámetros con algoritmos de nube de partículas (PSO) (Kennedy et al., 2001) y búsqueda gravitacional (GSA)(Rashedi et al., 2009) sobre los modelos SVM, Ridge y Lasso ver sección 2.1.2. Adicionalmente se usó XGBoost (Chen & Guestrin, 2016) y LSTM(Greff et al., 2017).

Modelo	Periodo de los Datos	Nombre Conjunto de Datos	Variables del Conjunto	Horizonte	Datos Train	Datos Test	Datos Pred
ARIMA - Baseline EMCALI	2019/01/01 - 2021/06/17	baseline	FECHA-CONSUMODIA-TIPODIA-DIASEMANA	6 meses	80%	20%	180 dias
Lasso Reg	Ene 2007 - Dic 2021	multivar01	PERI-SUSC-T13H-IPC-CONSUMO				
Lasso Reg GSA	Ene 2007 - Dic 2021	multivar01					
Lasso Reg PSO	Ene 2007 - Dic 2021	multivar01					
LSTM	Ene 2007 - Dic 2021	univar01	PERI-CONSUMO				
Random Forest	Ene 2007 - Dic 2021	multivar02	AÑO-MES-SUSC-T13H-IPC-PIBVALLE-CONSUMO				
Ridge Reg	Ene 2007 - Dic 2021	multivar01					
Ridge Reg GSA	Ene 2007 - Dic 2021	multivar01		3 Años (2019 - 2021)	3 años: 144 meses	3 años: 36 meses	3 años: 36 meses
Ridge Reg PSO	Ene 2007 - Dic 2021	multivar01		5 Años (2017 - 2021)	5 años: 120 meses	5 años: 60 meses	5 años: 60 meses
SVR Kernel Radial - GSA	Ene 2007 - Dic 2021	multivar01					
SVR Kernel Radial - PSO	Ene 2007 - Dic 2021	multivar01					
SVR Kernel Radial	Ene 2007 - Dic 2021	multivar01	PERI-SUSC-T13H-IPC-CONSUMO				
SVR Kernel Lineal - PSO	Ene 2007 - Dic 2021	multivar01					
SVR Kernel Lineal - GSA	Ene 2007 - Dic 2021	multivar01					
SVR Kernel Lineal	Ene 2007 - Dic 2021	multivar01					
VAR	Ene 2007 - Dic 2021	multivar01					
XGBoost	Ene 2007 - Dic 2021	multivar01					

Tabla 3: Configuración de datos para los modelos evaluados.

4.5. Evaluación

4.5.1. Estrategia de Evaluación

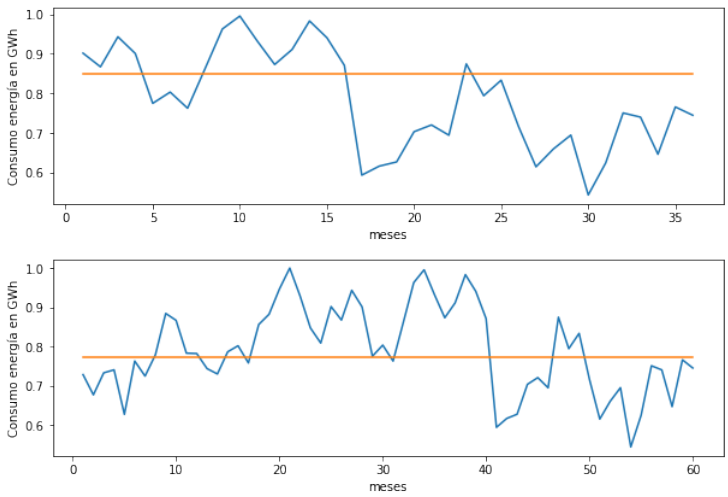

Figura 16: Baseline con el modelo ARIMA (0,1,1)(0,0,0), para ventana de tres años (arriba) y ventana de cinco años (abajo).

Dado que la predicción que se utiliza por el proveedor del servicio de energía en Cali es de corto y mediano plazo, fue necesario construir una línea base para poder comparar el desempeño a largo plazo del modelo ganador usando el modelo ARIMA proporcionado por EMCALI para hacer el pronóstico base a tres y cinco años, ver figura 16. Por otro lado se estimaron en este estudio 29 modelos que se dividieron en horizontes de tres años y cinco años. Se usaron diferentes técnicas de validación como cross_validation y out-of-bag como en Random Forest.

Tabla 4: Resultados modelo baseline ARIMA del comercializador.

Modelo	RMSE	MAE	MAPE
ARIMA H = 36	0.1391	0.1164	0.1370
ARIMA H = 60	0.1113	0.0907	0.1174

Las métricas obtenidas para este modelo en las dos ventanas de tres y cinco años se presentan en la tabla 4, obsérvese que el desempeño de este modelo aplicado en horizontes de largo plazo es mucho menor que la mitad de los ejecutados en este estudio, ver tablas 5 y 6.

4.5.2. Resultados de los Modelos

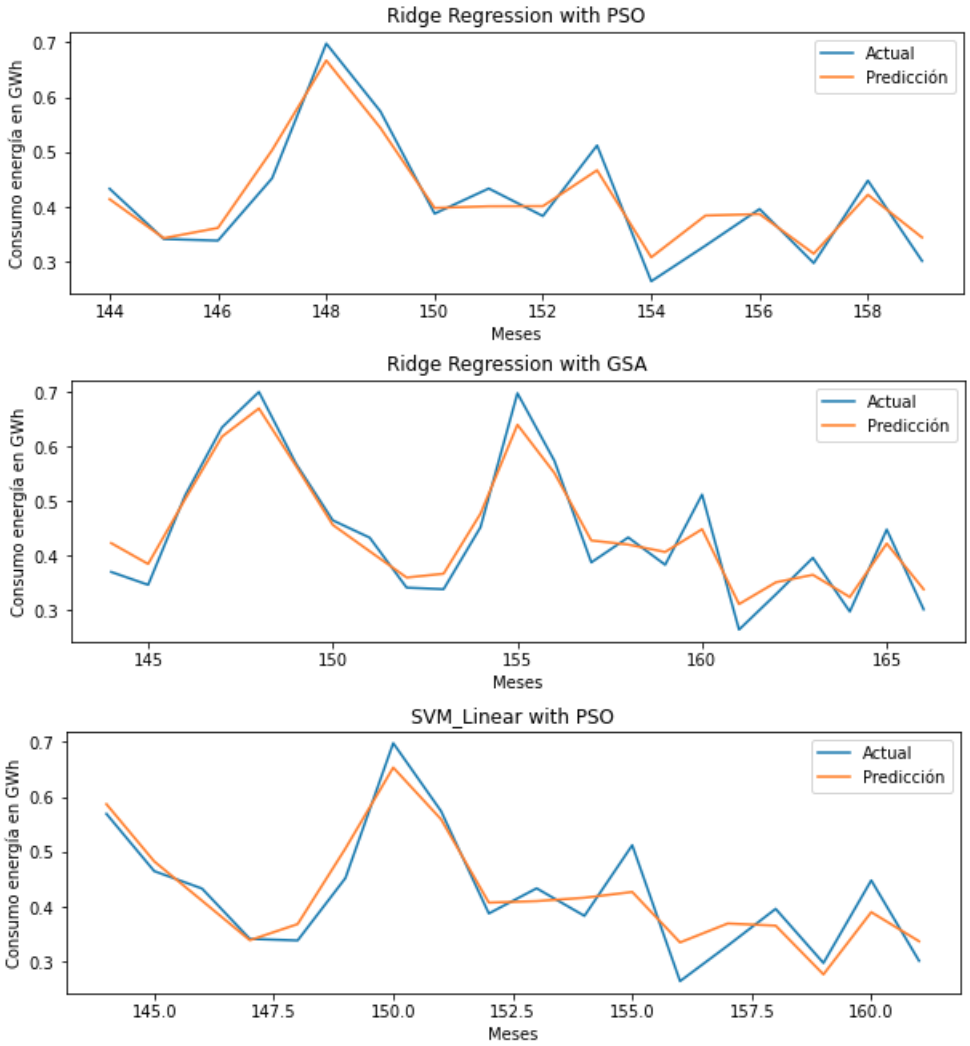

Figura 17: Predicciones de test 3 primeros puestos, en ventana de 3 años.

Tabla 5: Resultados de los modelos en ventana de 3 años.

Modelo	RMSE	MAE	MAPE
1. Ridge Reg PSO	0,0314	0,0304	0,0337
2. Ridge Reg GSA	0,0342	0,0287	0,0672
3. SVM Linear PSO	0,0401	0,0344	0,0847
4. Ridge Regression	0,0433	0,0312	0,1354
5. SVM Linear	0,0551	0,0410	0,1631
6. SVM Linear GSA	0,0557	0,0453	0,1135
7. SVM RBF PSO	0,0694	0,0595	0,1350
8. SVM RBF	0,1231	0,0871	0,2349
9. Lasso Reg PSO	0,1237	0,1000	0,2194
10. Lasso Reg GSA	0,1245	0,1011	0,2228
11. SVM RBF GSA	0,1498	0,1152	0,3230
12. VAR	0,1516	0,1280	0,1819
13. Lasso Regression	0,1574	0,1226	0,3241
14 XGBoost	4.8530	0.0001	0.0003
15. Random Forest	7,7072	6,2633	0.0269

El mejor modelo para la ventana de tres años resultó Ridge Reg PSO ver tabla 5 y para la ventana de cinco años las mejores métricas también fueron para el mismo modelo, ver la tabla 6, los mejores parámetros del optimizador PSO para este modelo son los siguientes: $m = 20$, $maxI = 10$, $npop = 5$, $W_1 = 1$, $wdamp = 0.99$, $c_1 = 2$, $c_2 = 2$, y los de configuración del modelo Ridge son : $Iteration$: 10, $Bestposition$= [5 1 20 5 14 1 1 20], $Bestcost = 0.0013659156$. Hay que mencionar que LSTM fue uno de los modelos de mas bajo rendimiento posiblemente por la reducida cantidad de datos.

Para evaluar los modelos se usaron las métricas RMSE, MAE y MAPE pero con el RMSE como métrica dominante ya que hace hincapié en los errores mas significativos reduciendo el sesgo ya que su objetivo es la media mientras que el MAE le da igual importancia a cada error con objetivo en la mediana (Vandeput, 2021), al cambiar la ventana de tres a cinco años se observa un incremento del RMSE del orden de 1.6 % y en el MAE una disminución del orden de 0.006 % para el modelo Ridge Reg PSO, del grupo con ventana de cinco años, es de anotar que al cambiar el horizonte de tres años a cinco años, se reducen los datos de entrenamiento y test por esta razón incrementa el RMSE mostrando la reducción del desempeño del modelo en el indicador. Se utiliza la siguiente relación:

$$1 - \frac{RMSE.ventana3}{RMSE.ventana5}$$

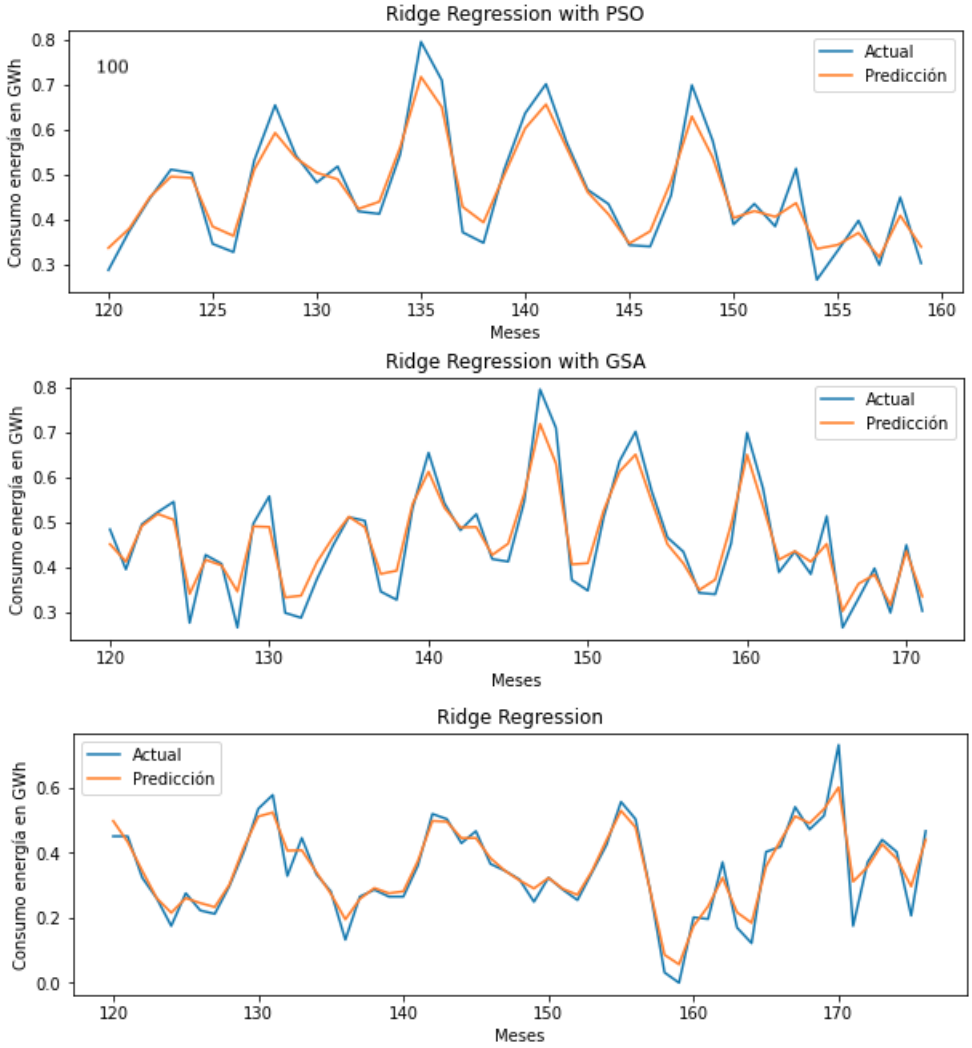

Figura 18: Predicciones de test 3 primeros puestos, en ventana de 5 años.

Tabla 6: Resultados de los modelos en ventana de 5 años.

Modelo	RMSE	MAE	MAPE
1. Ridge Reg PSO	0,0372	0,0306	0,0674
2. Ridge Reg GSA	0,0373	0,0306	0,0694
3. Ridge Regression	0,0403	0,0288	0,1109
4. SVM Linear PSO	0,0463	0,0386	0,0855
5. SVM Linear	0,0476	0,0353	0,1305
6. SVM Linear GSA	0,0666	0,0539	0,1281
7. SVM RBF PSO	0,0709	0,0589	0,1250
8. SVM RBF GSA	0,0959	0,0763	0,1682
9. SVM RBF	0,1009	0,0688	0,1887
10. Lasso Reg PSO	0,1230	0,0992	0,2203
11. VAR	0,1254	0,0978	0,1342
12. Lasso Reg GSA	0,1278	0,1029	0,2283
13. Lasso Regression	0,1461	0,1188	0,3079
14. LSTM	0.2895	0.2895	0.3886
15. XGBoost	5.8656	0.4119	0.1734
16. Random Forest	9,4924	6,6916	0.0305

4.6. Despliegue

El alcance de este estudio no incluye desarrollar algún tipo de interfaz de usuario final para gestionar modelos predictivos, por lo tanto el artefacto entregable es el conjunto de pronósticos mensuales para los próximos 5 años entre 2022 y 2027, con el respectivo código que los genera.

5. CONCLUSIONES Y TRABAJO FUTURO

Predecir la demanda de electricidad de largo plazo no es una tarea fácil, en un horizonte de cinco años como fue el propuesto en este proyecto pueden ocurrir eventos que afecten el pronóstico, como de hecho ocurrió con los datos de este estudio en el año 2011, aunque no hay perturbaciones profundas en este tipo de problema sobre los datos del consumo residencial y comercial regulado, si se presentaron para el mercado no regulado que afectaron la capacidad de pronosticar.

La primera conclusión importante demuestra que la optimización de hiperparámetros sobre modelos de aprendizaje automático juega un papel primordial al momento de pronosticar ya que los modelos que ocuparon los mejores puestos fueron realizados con técnicas como PSO y GSA que los refuerzan.

En segundo lugar para resolver la pregunta ¿Cómo mejorar la predicción de largo plazo en la demanda de energía eléctrica en Cali, utilizando técnicas de aprendizaje automático y modelos de inteligencia artificial? y dada la característica del los datos de el problema de este estudio, modelos superficiales o autoregresivos como SVM y Ridge en este proyecto entregaron mejores resultados que los modelos profundos, sin embargo es interesante revisar en un trabajo futuro si al incrementar la granularidad y en consecuencia cantidad de datos, los modelos profundos como las redes neuronales, ajustan de mejor manera la predicción.

Entre al año 2010 y 2012 hubo una reducción muy notable en el consumo global, esto como ya se explicó correspondió a la pérdida de clientes del mercado no regulado, lo cual produce una perturbación grande en la historia de los datos, por lo tanto como tercera conclusión debe realizarse el pronóstico separando el mercado regulado del no regulado si existe este tipo de problema, con el fin de tener una predicción más precisa en el componente regulado que corresponde a 2/3 partes de toda la demanda de electricidad de Cali y es mucho más estable.

En cuarto lugar la predicción generada con el mejor modelo representa una guía para soportar la gestión de planeación de compra de energía con la que no contaba hasta el momento el comercializador, por que no hacían pronóstico en horizonte de largo plazo, lo que permitirá hacer inversión adecuada con horizontes mas lejanos.

Referencias

ACOLGEN (2020). La energía que impulsa a colombia (capacidad instalada en colombia). https://www.acolgen.org.co/.

Aguilar, A. & Diaz, J. (2004). *Una visión del mercado eléctrico Colombiano*. UPME/Excelsior Impresores, primera edition.

Anand, A. & Suganthi, L. (2017). Forecasting of electricity demand by hybrid ann-pso models. *International Journal of Energy Optimization and Engineering*, 6.

Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Daimlerchrysler, T. R., Shearer, C., & Daimlerchrysler, R. W. (2000). Step-by-step data mining guide. *SPSS inc*, 78.

Chen, T. & Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In ACM (Ed.), *In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794).

(CongresodeColombia) (1994a). Ley 142 de 1994 (julio 11). http://www.alcaldiabogota.gov.co/sisjur/normas/Norma1.jsp?i=2752.

(CongresodeColombia) (1994b). Ley 143 de 1994. *Diario Oficial*, 1994, 347.

DANE (2017). Estratificación socioeconómica para servicios públicos domiciliarios. https://web.archive.org/web/20170329055800/http://www.dane.gov.co/index.php/servicios-al-ciudadano/servicios-de-informacion/estratificacion-socioeconomica.

Ertel, W. (2017). *Introduction to Artificial Intelligence (Undergraduate Topics in Computer Science)*. Cham,Switzerland: Springer, 2nd edition.

Gao, F., Chi, H., & Shao, X. (2021). Forecasting residential electricity consumption using a hybrid machine learning model with online search data. *Applied Energy*, 300.

Géron, A. (2019). *Hands-on machine learning with Scikit-Learn, Keras and TensorFlow: concepts, tools, and techniques to build intelligent systems*. Sebastopol,CA: O'Reilly Media, 2nd edition.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning An MIT Press Book*, volume 29.

Greff, K., Srivastava, R. K., Koutnik, J., Steunebrink, B. R., & Schmidhuber, J. (2017). Lstm: A search space odyssey. *IEEE Transactions on Neural Networks and Learning Systems*, 28.

Gulcu, S. & Kodaz, H. (2017). The estimation of the electricity energy demand using particle swarm optimization algorithm: A case study of turkey. volume 111.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). *An Introduction to Statistical Learning*. Springer Texts, 2nd edition.

Kennedy, J., Eberhart, R. C., Shi, Y., Kennedy, J., Eberhart, R. C., & Shi, Y. (2001). chapter seven - the particle swarm.

Koutsandreas, D., Spiliotis, E., Petropoulos, F., & Assimakopoulos, V. (2021). On the selection of forecasting accuracy measures. *Journal of the Operational Research Society*.

Liu, S., Zeng, A., Lau, K., Ren, C., wai Chan, P., & Ng, E. (2021). Predicting long-term monthly electricity demand under future climatic and socioeconomic changes using data-driven methods: A case study of hong kong. *Sustainable Cities and Society*, 70.

Ma, J., Oppong, A., Adjei, G., Adjei, H., Atta-Osei, E., Agyei-Sakyi, M., & Adu-Poku, D. (2021). Demand and supply-side determinants of electric power consumption and representative roadmaps to 100 *Journal of Cleaner Production*, 299.

Mir, A. A., Alghassab, M., Ullah, K., Khan, Z. A., Lu, Y., & Imran, M. (2020). A review of electricity demand forecasting in low and middle income countries: The demand determinants and horizons. *Sustainability (Switzerland)*, 12.

Peña-Guzmán, C. & Rey, J. (2020). Forecasting residential electric power consumption for bogotá colombia using regression models. *Energy Reports*, 6, 561–566. The 6th International Conference on Energy and Environment Research - Energy and environment: challenges towards circular economy.

Rashedi, E., Nezamabadi-pour, H., & Saryazdi, S. (2009). Gsa: A gravitational search algorithm. *Information Sciences*, 179.

Rueda, V. M., Henao, J. D. V., & Cardona, C. J. F. (2011). Avances recientes en la predicción de la demanda de electricidad usando modelos no lineales. *DYNA (Colombia)*, 78.

Russell, S. & Norvig, P. (2021). *Artificial Intelligence A Modern Approach*, volume 53. Pearson Education Limited 2022, 4th edition.

Sutton, R. S. & Barto, A. G. (2014). Reinforcement learning: An introduction(second edition, in progress). *Decision Theory Models for Applications in Artificial Intelligence: Concepts and Solutions*.

Taheri, S., Talebjedi, B., & Laukkanen, T. (2021). Electricity demand time series forecasting based on empirical mode decomposition and long short-term memory. *Energy Engineering: Journal of the Association of Energy Engineering*, 118.

Vandeput, N. (2021). *Data Science for Supply Chain Forecasting*. Berlin/Boston: De Gruyter, 2nd edition.

Wirth, R. (2000). Crisp-dm : Towards a standard process model for data mining. *Proceedings of the Fourth International Conference on the Practical Application of Knowledge Discovery and Data Mining*.