****

**FACULTAD DE INGENIERÍA**

**-Maestría en Ciencia de Datos-**

**1. Integrantes y Director del Trabajo.**

Integrante(s):

Nelson Andrés Andrade Bonilla

Mario José Castellanos Valencia

Director:

Hernán Darío Benítez Restrepo.

Ingeniero Electrónico

PhD. en Ingeniería Eléctrica

**2. Título del Proyecto o Trabajo de Grado.**

MODELO DE PRONÒSTICO PARA EL CONSUMO DE ENERGÍA ELÉCTRICA CON UN HORIZONTE DE TIEMPO A 1 Y 2 AÑOS PARA EL MERCADO REGULADO DE ENERGÍA EN CALI.

**3. Contexto, Antecedentes y Justificación.**

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 D. & Diaz V., 2004). Posteriormente, en 1994, se promulgan las leyes 142 de Servicios Públicos Domiciliarios ((Congreso de Colombia), 1994) y la 143 o ley Eléctrica (Congreso & Colombia, 1994), 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 ((Congreso de Colombia), 1994), 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.

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).

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 de la demanda, 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.

**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 los modelos actualmente utilizados como ARIMA, que son realizados para horizontes de corto plazo, con los cuales algunos comercializadores como EMCALI EICE ESP no hacen predicción de mediano y largo plazo (ver tabla 1), porque no les ofrecen la precisión requerida o no incorporan 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 la demanda.

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 y modelos de inteligencia artificial?

**5. Objetivos del Proyecto.**

**5.1. Objetivo General.**

Desarrollar y validar un modelo de Machine Learning (ML) para pronosticar el consumo de energía eléctrica en Cali, con un horizonte de tiempo a 1 y 2 años para el mercado regulado de energía.

**5.2. Objetivos Específicos.**

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

5.2.2 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 de Machine Learning diseñados.

5.2.3 Validar el modelo seleccionado para determinar la mejor configuración de parámetros para la cual, el modelo de aprendizaje automático propuesto provee los mejores resultados.

**6. Marco Teórico.**

**6.1 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 el fundamento base del desarrollo socioeconómico es de suma urgencia identificar y entender los factores involucrados en el conocimiento de la cantidad de 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.

### 6.1.1 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 (Ma et al., 2021) 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, como en Suecia.

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, 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 redistribución 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).

6.1.2 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.

**6.2 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) 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 una forma de inteligencia artificial que aprende de los datos mas que de la 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 o machine learning. Alguna selección de ellos será la base para pronosticar la demanda de electricidad en Colombia en este proyecto.

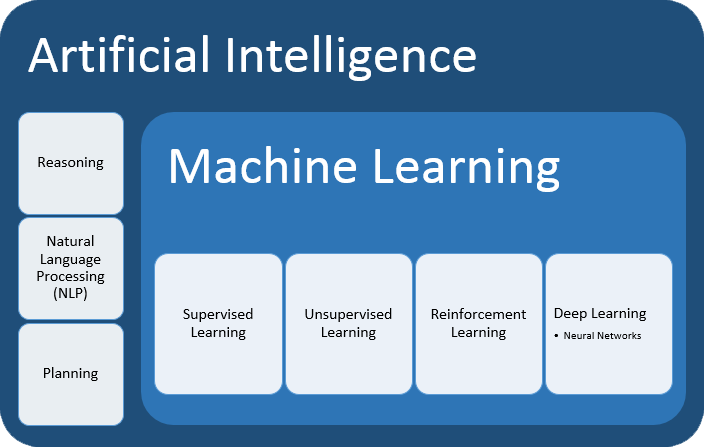


Figura 1: Componentes de la inteligencia artificial como ciencia[[1]](#footnote-1).

**Machine Learning**

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 automático, temas que vienen a ser las ramas mas importantes de la IA, a la que le interesa realmente la adquisición de habilidades matemáticas que no son del plano de la memorización. La tarea del aprendizaje automático (Machine Learning) 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 trade-off sesgo-varianza que debe ser optimizado para llegar a resultados correctos.

* **Aprendizaje Supervisado**

El objetivo del aprendizaje supervisado consiste en estimar un modelo que relacione una respuesta con unos elementos predictores, para explicar el modelo o predecir una respuesta a partir de unas observaciones (A.Géron, 2019). Se trata de predecir un valor a partir de un conjunto de datos de entrenamiento sobre un algoritmo de aprendizaje estimando unos parámetros, bajo una hipótesis con el fin de hallar una respuesta cuantitativa. Otro enfoque del aprendizaje no supervisado es el de encontrar modelos que describan clases para realizar futuras predicciones, en este caso la variable objetivo es de carácter discreto, adicionalmente también estiman la probabilidad de cada clase. Un enfoque mas corresponde a poder inferir o explicar la respuesta encontrando la relación con sus predictores asociados.

A continuación se describen los métodos mas comunes de este tipo de aprendizaje:

Regresión lineal: Este es el enfoque mas 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 se llaman estimadores y corresponden a los coeficientes de la variables que en realidad son constantes.

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 mas cerca de la nueva observación a predecir y se basa en instancias de aprendizaje mas 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 mas cercanos a cada nueva observación , y de un parámetro k que corresponde al numero de vecinos mas 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 overfitting 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" (Norvig, 2019). 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 sigmoidea o logística de forma que los valores de esta función se puedan interpretar como probabilidades de que una instancia pertenezca a una clase especifica (A.Géron, 2019). Se coloca un umbral limite para determinar esa clase si el valor es mayor que ese umbral la respuesta es positiva de lo contrario es negativa.

Maquinas de vectores de soporte (SVM): Este algoritmo es de clasificación y está basado en el clasificador de margen máxima, 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. Esto permite que las separaciones sean de forma polinómica mejorando el clasificador.

* **Aprendizaje no supervisado**

A diferencia del aprendizaje supervisado, el aprendizaje no supervisado se enfoca en un conjunto de características sobre un numero determinado de observaciones para descubrir cosas interesantes acerca de esas observaciones, no le interesa predecir, no tiene una variable de respuesta que guie el proceso sino únicamente las observaciones y solo de ellas se 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 (A.Géron, 2019), reglas de aplicación o detectar anomalías. Se realiza como parte de un proceso de análisis exploratorio de datos.

Estos son los métodos de este tipo de aprendizaje:

Análisis de componentes principales (PCA): Es un procedimiento que busca reducir la dimensionalidad de los datos pero que contenga tanta información como el 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. Su aplicación es muy importante porque se ha usado entre otros, en investigación de cáncer para poder determinar distintos grupos sobre las muestras o sobre la genética para entender mejor la enfermedad.

Clustering: 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, no se tiene una variable objetivo por eso es no supervisado, 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 clusters y el Clustering Jerárquico que busca mediante un árbol llamado dendograma descubrir los grupos que pueden existir en los datos usando distancias Ecludianas 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 explotar 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 debe aprender a reaccionar de manera óptima debido al explorar las distintas posibilidades de ejecución de acciones en situaciones que se presentan por el entorno (Sutton & Barto, 2014). Decidir entre las dos fases es crucial para el rendimiento del agente, por lo tanto la aplicación de algoritmos, según el problema a resolver, resulta una ventaja considerable, entre los más relevantes a considerar: e-Greedy (establecer una probabilidad) y Upper Confidence Bound (Límite de Cota Superior).

* **Aprendizaje profundo**

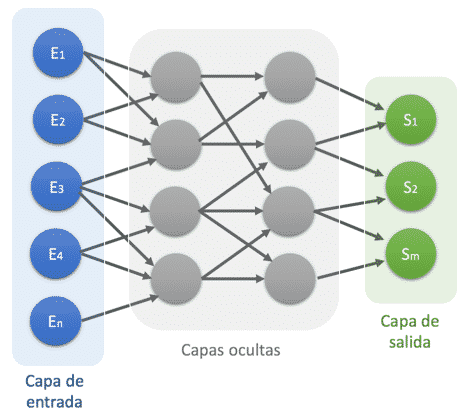


Figura 2. Capas de una red neuronal[[2]](#footnote-2).

El aprendizaje profundo ha tenido un larga historia, con aplicaciones satisfactorias y otras no tanto, pero los avances obtenidos proporcionan un marco muy potente para el aprendizaje supervisado. Se ha vuelto más útil a medida que los datos se incrementan para su entrenamiento y la capacidad de computo esta mas asequible por el auge de tecnologías disponibles en la nube, resolviendo problemas complicados con mayor precisión.

Para (Goodfellow et al., 2016) un ejemplo claro es la red profunda de avance, o perceptrón multicapa (Multilayer Perceptron, MLP), utilizada para representar el concepto de una imagen combinando componentes más simples, como bordes y esquinas, que a su vez conforman aristas. Su comportamiento es aprender representaciones de datos, con múltiples niveles de abstracción que realizan una serie de transformaciones lineales y no lineales, generando una salida que difiera muy poco con la esperada.

Una aproximación gráfica de una red MLP se muestra en la figura 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, se observa varias interconexiones entre capas y neuronas ocultas.

**7. 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 Machine Learning, 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.

**7.1 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 seleccionaros los siguientes:

**7.1.1 Técnicas de predicción basadas en Inteligencia Artificial**

7.1.1.1 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. Se utilizaron técnicas de optimización 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). 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 haciendo pronósticos de consumo de electricidad mas acertados.

7.1.1.2 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).

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 maquinas de vectores de soporte (SVM), arboles de decisión (DT), redes neuronales artificiales (ANN), arboles de decisión de impulso de gradiente (GBDT), procesos de regresión gaussianos (GPR) y regresión lineal múltiple (MLR). Se adoptó el modelo GBDT que resultó ser el mas adecuado para el pronóstico de la demanda de energía eléctrica de largo plazo en Hong Kong.

**7.1.2 Modelos econométricos de predicción**

7.1.2.1 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 encontró como resultado que las variables que mas 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.

7.1.2.2 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 de corto plazo (LSTM) con redes neuronales recurrentes (RNN), con datos de 1990 a 2018 y proyecciones de demanda hasta 2050.

**7.1.3 Técnicas de predicción en series temporales**

7.1.3.1 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 hibridas 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. El modelo LSTM + EMD Híbrido superó con mejores resultados al XGBoost y LR en predicciones de demanda de electricidad de corto y largo plazo.

**7.2 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.
* Contexto: relacionados con el consumo o demanda de electricidad y su pronóstico a largo plazo.

En la tabla 2 se encuentra el comparativo de la literatura estudiada.

Tabla 2 Resumen de los criterios de comparación entre los artículos seleccionados y el proyecto de grado.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Titulo | Dimensión temporal | Geografía | Técnicas | Proceso | Contexto |
| 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 | Transformación de sistemas eléctricos tradicionales a fuentes limpias |
|  |  |  |  |  |  |
| 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 | Data-Driven  Predictivo | Impacto de cambios socioeconómicos en la demanda mensual de electricidad a largo plazo. |
|  |  |  |  |  |  |
| Forecasting residential electric power consumption for Bogotá-Colombia using regression models.(Peña-Guzmán & Rey, 2020) | 2020 | Bogotá | MLR2Log | Econométrico  Predictivo | Variables de mayor impacto en el consumo de la demanda de electricidad |
|  |  |  |  |  |  |
| Forecasting residential electricity consumption using a hybrid machine learning model with online search data.(Gao et al., 2021) | 2021 | China | SVR  SARIMA(X)  BPNN  ELM | Predictivo | Necesidad de precisión en el pronóstico del consumo de electricidad domiciliaria. |
|  |  |  |  |  |  |
| Forecasting of Electricity Demand by Hybrid ANN-PSO Models.(Anand & Suganthi, 2017) | 2017 | India | ARIMA  ANN-PSO  GA  ANNBP | Predictivo | Necesidad de proyección de consumo de electricidad por gestación de proyectos energéticos para cerrar la brecha de abastecimiento. |
|  |  |  |  |  |  |
| The estimation of the electricity energy demand using PSO algorithm: case study of Turkey.(Gulcu & Kodaz, 2017) | 2017 | Turquía | ANN-PSO | Predictivo | Cerrar brecha de abastecimiento de electricidad por crecimiento económico |
|  |  |  |  |  |  |
| 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 | Necesidad de precisión en el pronóstico del consumo de electricidad por crecimiento poblacional. |
|  |  |  |  |  |  |
|  |  |  |  |  |  |

**8. Metodología.**

La metodología elegida para abordar el proyecto es CRISP-DM (Cross-Industry Standard Process for Data Mining). Esta metodología fue desarrollada por el consorcio formado por las empresas NCR Systems Engineering Copenhagen (USA and Denmark), DaimlerChrysler AG (Germany), SPSS Inc. (USA) y OHRA Verzekeringen en Bank Groep B.V (The Netherlands), 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).

**8.1 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.

Actividades de CRISP-DM.

* Determinar objetivos del Negocio:

Teniendo en cuenta el alcance se define el objetivo general del negocio, al cual el proyecto de analítica de impactar con respecto a la demanda de energía eléctrica.

* Evaluar la situación:

Implica una búsqueda de datos más detallada, incluyendo variables endógenas y exógenas, así como otros factores que deben considerarse al determinar el objetivo del análisis de datos y el plan del proyecto.

* Realizar el plan del proyecto:

Se define un listado ordenado y priorizado de los requisitos necesarios para desarrollo del modelo y su posterior validación, basándose en los objetivos específicos del proyecto.

**8.2 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 optimo.

**8.3 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.

**8.4 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.

**8.5 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”.

**8.6 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.

Como etapa final, se busca poder compartir y difundir los resultados del proyecto, para que puedan ser usados por los usuarios del negocio. Para este proyecto, el alcance propuesto comprende el procedimiento general para la construcción del modelo resultante y la producción de un reporte final, donde se encuentren los resultados obtenidos y las experiencias alcanzadas.

**9. Cronograma del Proyecto.**

Tabla 3 Cronograma del proyecto.

|  |  |  |  |
| --- | --- | --- | --- |
| **Objetivos** | **Actividades** | **Fecha Inicio** | **Fecha fin** |
| **Comprensión del negocio** | * Determinar objetivos del Negocio * Evaluar la situación * Realizar el plan del proyecto | Septiembre 2021 | Noviembre 2021 |
| **Comprensión de los datos.** | * Recolectar los datos * Descripción de los datos * Exploración y verificación de los datos | Diciembre  2021 | Enero 2022 |
| **Preparación de los datos** | * Selección de los datos * Limpieza de los datos * Transformación de los datos | Febrero 2022 | Marzo 2022 |
| **Modelado** | * Selección de técnicas de modelado * Generar plan de pruebas * Construir y validar modelo | Marzo 2022 | Abril 2022 |
| **Evaluación** | * Evaluar resultados * Revisión de los pasos | Abril 2022 | Mayo 2022 |
| **Despliegue** | * Compartir y difundir resultados * Producción del informe final | Mayo 2022 | Mayo 2022 |

**10. Referencias bibliográficas.**

(Congreso de Colombia). (1994). LEY 142 DE 1994 (Julio 11). *Secretaría General de La Alcaldía Mayor de Bogotá D.C.*, *142*. http://www.alcaldiabogota.gov.co/sisjur/normas/Norma1.jsp?i=2752

A.Géron. (2019). Hands-on machine learning with Scikit-Learn, Keras and TensorFlow: concepts, tools, and techniques to build intelligent systems. In *O’Reilly Media*.

ACOLGEN. (2020). *La energía que impulsa a Colombia (Capacidad instalada en Colombia)*. https://www.acolgen.org.co/

Aguilar D., A., & Diaz V., J. A. (2004). *Una visión del mercado eléctrico Colombiano*.

Anand, A., & Suganthi, L. (2017). Forecasting of Electricity Demand by Hybrid ANN-PSO Models. *International Journal of Energy Optimization and Engineering*, *6*(4). https://doi.org/10.4018/ijeoe.2017100105

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*.

Congreso, & Colombia, de la R. de. (1994). Ley 143 de 1994. *Diario Oficial*, *1994*(41), 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). In *Book*.

Gao, F., Chi, H., & Shao, X. (2021). Forecasting residential electricity consumption using a hybrid machine learning model with online search data. *Applied Energy*, *300*. https://doi.org/10.1016/j.apenergy.2021.117393

Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning An MIT Press Book. In *Nature* (Vol. 29, Issue 7553).

Gulcu, S., & Kodaz, H. (2017). The estimation of the electricity energy demand using particle swarm optimization algorithm: A case study of Turkey. *Procedia Computer Science*, *111*. https://doi.org/10.1016/j.procs.2017.06.011

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

Liu, S., Zeng, A., Lau, K., Ren, C., Chan, P. wai, & 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*. https://doi.org/10.1016/j.scs.2021.102936

Ma, J., Oppong, A., Adjei, G. K. B., 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% renewable systems. *Journal of Cleaner Production*, *299*. https://doi.org/10.1016/j.jclepro.2021.126832

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. In *Sustainability (Switzerland)* (Vol. 12, Issue 15). https://doi.org/10.3390/SU12155931

Norvig, S. R. and P. (2019). Artificial Intelligence A Modern Approach 4th Ed. In *Journal of Chemical Information and Modeling* (Vol. 53, Issue 9).

Peña-Guzmán, C., & Rey, J. (2020). Forecasting residential electric power consumption for Bogotá Colombia using regression models. *Energy Reports*, *6*. https://doi.org/10.1016/j.egyr.2019.09.026

Rueda, V. M., Velásquez Henao, J. D., & Franco Cardona, C. J. (2011). Avances recientes en la predicción de la demanda de electricidad usando modelos no lineales. *DYNA (Colombia)*, *78*(167).

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*(6). https://doi.org/10.32604/EE.2021.017795

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*, *24959*.

1. Síntesis explicativa y grafica tomadas de <https://www.ibm.com/co-es/analytics/machine-learning> [↑](#footnote-ref-1)
2. Figura extraída de https://www.diegocalvo.es/wp-content/uploads/2017/07/neural-network.png [↑](#footnote-ref-2)