

---

# INSTITUTO TECNOLÓGICO SUPERIOR DE TEZIUTLÁN

---

## Tesis



“Modelo computacional de aprendizaje automático para  
la predicción de la demanda eléctrica”

PRESENTA:

**ADRIÁN AGUILAR TIRADO**

CON NÚMERO DE CONTROL

**21TE0010P**

PARA OBTENER EL GRADO ACADÉMICO DE:

**MAESTRA EN SISTEMAS COMPUTACIONALES**

CLAVE DEL PROGRAMA ACADÉMICO

**MPSCO-0127**

DIRECTOR (A) DE TESIS:

**DR. JACOBO ROBLES CALDERÓN**

CO- DIRECTOR DE TESIS:

“La Juventud de hoy, Tecnología del Mañana”

TEZIUTLÁN, PUEBLA, JUNIO 2023

FOLIO: **SPeI.091.2023**  
ASUNTO: Autorización de  
Impresión de Tesis

Teziutlán, Puebla, 26/junio/2023

**ADRÍAN AGUILAR TIRADO**  
**CANDIDATO AL GRADO DE MAESTRO EN SISTEMAS COMPUTACIONALES**  
**NO. DE CONTROL 21TE0010P**  
**PRESENTE**

Por este medio me permito informar que el Consejo de Posgrado a evaluado el documento de trabajo de Tesis de Grado para el programa de Maestría en **SISTEMAS COMPUTACIONALES**, con el tema: **"MODELO COMPUTACIONAL DE APRENDIZAJE AUTOMÁTICO PARA LA PREDICCIÓN DE LA DEMANDA ELÉCTRICA"**, y al analizar la estructura, el contenido, así como, el impacto que emanan de dicha investigación, y conforme a lo establecido en los Lineamientos para la Operación de los Estudios de Posgrado en el Tecnológico Nacional de México en su apartado 2.15.5. Examen de Grado, esta área a mi cargo emite el presente dictamen.

## AUTORIZACIÓN DE IMPRESIÓN DE TESIS DE GRADO

Debiendo entregar una copia de la Tesis en formato PDF en memoria USB, para agendar su Examen de Grado.

Sin otro particular, le envío un cordial saludo.

### ATENTAMENTE

*Excelencia en Educación Tecnológica*  
*La Juventud de hoy, Tecnología del mañana*



**JULIO VICTOR GALINDO ROJAS**  
**SUBDIRECTOR DE POSGRADO E INVESTIGACIÓN**

ccp. Archivo  
JVGR/lacpj



## **AGRADECIMIENTOS**

Agradezco al Consejo Nacional de Humanidades, Ciencia y Tecnología (CONAHCYT) por la beca otorgada para realizar estudios de posgrado a nivel maestría a través de la convocatoria "BECAS NACIONALES PARA ESTUDIOS DE POSGRADO 2022".

## **DEDICATORIA PERSONAL**

A mis hijos Liam y Santiago por ser ellos mi mayor impulso y motivación en mi vida para el logro de mis metas.

A mis padres por ser esa fuerza necesaria en todas las decisiones que he tomado a lo largo de la vida para cumplir el alcance de mis sueños.

Adriana por hacer que yo creyera en mí en todo momento, sin ti no hubiera podido lograr este posgrado siempre estuviste a mi lado apoyándome y alentándome para no dejarme vencer.

A mis docentes doy gracias por ser esa iluminación y guía para abrirme camino a un nuevo mundo de conocimientos excepcionales.

Y sobre todo a Dios por permitirme llegar hasta aquí y darme los conocimientos necesarios para lograrlo.

A todos ustedes gracias...

# ÍNDICE GENERAL

|   |     |
|---|-----|
| RESUMEN .....   | I   |
| ABSTRACT .....  | II  |
| INTRODUCCIÓN .....                                    | III |
| CAPÍTULO I GENERALIDADES DEL PROYECTO .....           | 1   |
| 1.1 Marco teórico.....                                | 1   |
| 1.1.1 Demanda de energía .....                        | 1   |
| 1.1.2 Modelo de predicción.....                       | 5   |
| 1.1.3 Inteligencia computacional.....                 | 6   |
| 1.1.4 Ciencia de datos.....                           | 7   |
| 1.1.5 Dataset.....                                    | 9   |
| 1.1.6 Minería de datos.....                           | 9   |
| 1.1.7 Metodología CRISP-DM .....                      | 10  |
| 1.1.8 Aprendizaje automático (Machine Learning) ..... | 12  |
| 1.1.9 Aprendizaje supervisado .....                   | 13  |
| 1.1.10 Regresión lineal.....                          | 14  |
| 1.1.11 Algoritmo computacional.....                   | 15  |
| 1.1.12 Aprendizaje profundo (Deep Learning) .....     | 15  |
| 1.1.13 Redes Neuronales Artificiales.....             | 15  |
| 1.1.14 Modelo computacional .....                     | 16  |
| 1.1.15 Lenguaje de programación.....                  | 17  |
| 1.1.16 Mapa de calor .....                            | 17  |
| 1.1.17 Gráfica de dispersión .....                    | 18  |
| 1.1.18 Gráfico de magnitud y perdida .....            | 19  |

|   |  |    |
|---|--|----|
| 1.1.19                                      | Series de tiempo .....                                 | 21 |
| 1.1.20                                      | Histograma de dispersión de valores.....               | 21 |
| 1.2   | Planteamiento del Problema .....                       | 22 |
| 1.3   | Justificación.....                                     | 24 |
| 1.4   | Hipótesis .....  | 24 |
| 1.5   | Objetivo general .....                                 | 25 |
| 1.6   | Objetivos específicos .....                            | 25 |
| 1.7   | Alcances y limitaciones .....                          | 25 |
| 1.7.1                                       | Alcances .....   | 25 |
| 1.7.2                                       | Limitaciones.....                                      | 26 |
| CAPÍTULO II ESTADO DEL ARTE.....            |  | 27 |
| 2.1   | Trabajos relacionados .....                            | 27 |
| 2.2   | Análisis comparativo de los trabajos relacionados..... | 32 |
| 2.3   | Propuesta de solución .....                            | 38 |
| 2.3.1                                       | Lenguaje de programación Python .....                  | 38 |
| 2.3.2                                       | Keras.....   | 39 |
| 2.3.3                                       | Tensorflow.....  | 40 |
| 2.3.4                                       | Jupyter notebook .....                                 | 40 |
| 2.3.5                                       | Anaconda.....  | 41 |
| 2.3.6                                       | Excel .....  | 41 |
| 2.3.7                                       | Algoritmo basado en Redes Neuronales Recurrentes.....  | 41 |
| CAPÍTULO III METODOLOGÍA Y DESARROLLO ..... |  | 43 |
| 3.1   | Metodología.....                                       | 43 |

|   |                                 |    |
|---|---------------------------------|----|
| 3.1.1   | Comprensión de datos: .....     | 45 |
| 3.1.2   | Procesar y preparar datos. .... | 46 |
| 3.1.3   | Evaluación de Algoritmo.....    | 48 |
| 3.1.4   | Mejora de resultados. ....      | 50 |
| CAPÍTULO IV RESULTADOS.....                     |                                 | 51 |
| 4.1   | CASO DE ESTUDIO. ....           | 52 |
| 4.1.1   | Análisis de los datos. ....     | 54 |
| 4.1.2   | Algoritmo de aprendizaje .....  | 68 |
| CAPÍTULO V CONCLUSIONES Y RECOMENDACIONES ..... |                                 | 86 |
| PRODUCTOS ACADÉMICOS.....                       |                                 | 89 |
| REFERENCIAS.....                                |                                 | 90 |

## ÍNDICE DE FIGURAS

|  |    |
|--|----|
| Figura 1 Valores históricos del consumo neto de energía .....        | 3  |
| Figura 2 Consumo de energía por segmento .....                       | 4  |
| Figura 3 Circuito eléctrico de distribución .....                    | 5  |
| Figura 4. Parámetros de predicción.....                              | 6  |
| Figura 5 Ciencia de datos como modelo BI.....                        | 8  |
| Figura 6 Dataset estructura de datos .....                           | 9  |
| Figura 7 Relación de la minería de datos con otras disciplinas ..... | 10 |
| Figura 8 Metodología CRISP-DM.....                                   | 11 |
| Figura 9 Aprendizaje automático .....                                | 13 |
| Figura 10 Esquema de algoritmos de aprendizaje .....                 | 14 |
| Figura 11 Representación de una Red Neuronal .....                   | 16 |
| Figura 12 Representación de un mapa de calor .....                   | 18 |
| Figura 13 Gráfica de dispersión .....                                | 19 |
| Figura 14 Gráficos de validación de magnitud y perdida.....          | 20 |
| Figura 15 Gráfica serie de tiempo .....                              | 21 |
| Figura 16 Histograma de dispersión de valores.....                   | 22 |
| Figura 17 Arquitectura de una RNN .....                              | 42 |
| Figura 18 Metodología Red Neuronal .....                             | 44 |
| Figura 19 Diagrama de flujo de minería de datos.....                 | 47 |
| Figura 20 Perceptrón red neuronal .....                              | 49 |
| Figura 21 Arquitectura del Modelo Neuronal Recurrente RNN .....      | 50 |
| Figura 22 Variables de energía eléctrica .....                       | 53 |
| Figura 23 Variables climatológicas .....                             | 53 |

|   |    |
|---|----|
| Figura 24 Variable totalidad de usuarios por región .....         | 54 |
| Figura 25 Integración del dataset .....                           | 58 |
| Figura 26 Temperaturas por región .....                           | 60 |
| Figura 27 Integración de Área y Región a dataset .....            | 61 |
| Figura 28 Dataset para entrenamiento y prueba .....               | 62 |
| Figura 29 Mapa de calor de correlación de variables .....         | 71 |
| Figura 30 Gráfica de dispersión de correlación de variables ..... | 72 |
| Figura 31 Entrenamiento y validación de perdida .....             | 77 |
| Figura 32 Magnitud de perdida.....                                | 79 |
| Figura 33 Magnitud de perdida de entrenamiento y validación ..... | 80 |
| Figura 34 Valor real frente a valor previsto.....                 | 80 |
| Figura 35 Diagrama de puntos de dispersión.....                   | 81 |
| Figura 36 Histograma valor residual.....                          | 82 |

## **ÍNDICE DE TABLAS**

|   |    |
|---|----|
| Tabla 1. Análisis comparativo .....                                 | 32 |
| Tabla 2 Variables de predicción .....                               | 45 |
| Tabla 3 Tabla representativa de las variables relevantes .....      | 55 |
| Tabla 4 Instancias por municipio .....                              | 56 |
| Tabla 5 Estadística del conjunto de datos.....                      | 57 |
| Tabla 6 Tabla estaciones meteorológicas .....                       | 58 |
| Tabla 7 Transformación de datos. ....                               | 60 |
| Tabla 8 Dataset de entrenamiento y prueba.....                      | 62 |
| Tabla 9 Tabla comparativa de valor real contra valor previsto ..... | 77 |

## **RESUMEN**

El objetivo de esta investigación fue desarrollar un modelo computacional basado en una red neuronal recurrente de regresión para predecir la demanda de energía eléctrica. Se tomaron en cuenta múltiples variables, como el clima y la demanda consumida por los usuarios, entre otros. Para lograrlo, se recopilaron conjuntos de datos históricos de la demanda de energía en la zona oriente de los estados de Puebla y Veracruz, los cuales se utilizaron para entrenar y evaluar el modelo de red neuronal.

El modelo de red neuronal fue capaz de capturar patrones complejos en los datos y generar predicciones cercanas a los valores reales de la demanda de energía eléctrica. Sin embargo, se observó que el rendimiento del modelo estaba influenciado por la calidad de los datos utilizados en el entrenamiento. Además, se exploró la arquitectura de la red neuronal recurrente (RNN) de regresión durante el proceso de prueba del modelo de predicción, demostrando ser una herramienta efectiva para pronosticar la demanda de energía eléctrica.

## **ABSTRACT**

The objective of the research study was to develop a regression recurrent neural network computational model to predict the demand for electrical energy based on multiple variables, such as the weather, the demand consumed per user, among others. Historical data sets of the energy demand of the eastern zone of the states of Puebla and Veracruz were collected and used to train and evaluate the neural network model. The neural network model was able to capture complex patterns from the data and generate close predictions of the actual values of electrical power demand. It was observed that the performance of the model was influenced by the quality of the data used for training. The regression recurrent neural network (RNN) architecture was explored for the prediction model testing process, being this a successful tool for electric power demand prediction forecasts.

## **INTRODUCCIÓN**

La predicción de la demanda de energía eléctrica es crucial para una planificación y gestión eficiente de la generación y distribución de energía. Las redes neuronales recurrentes (RNN) de regresión son una herramienta avanzada de aprendizaje profundo utilizada para modelar y predecir patrones complejos y dinámicos en la demanda de energía eléctrica. Estas redes se implementan a través del lenguaje de programación Python.

Las RNN son especialmente adecuadas para el análisis de secuencias de datos, como las series temporales, lo que les permite utilizar información previa en la predicción futura. Esta característica las hace especialmente efectivas para capturar dependencias temporales y patrones a largo plazo en la distribución de los datos.

En el contexto de la predicción de la demanda de energía eléctrica, las RNN pueden aprender de datos históricos, como la demanda de energía en intervalos anteriores, y utilizar esa información para predecir la demanda futura. Los estudios han demostrado que la predicción de la demanda de energía eléctrica utilizando redes neuronales recurrentes es altamente precisa y efectiva en comparación con modelos de regresión tradicionales.

La capacidad de las RNN para modelar patrones complejos y adaptarse a cambios en los datos las convierte en una herramienta tecnológica sofisticada para abordar los desafíos asociados a la demanda de energía eléctrica.

# **CAPÍTULO I GENERALIDADES DEL PROYECTO**

## **1.1 Marco teórico**

En este capítulo, se presentan una serie de conceptos fundamentales que son esenciales para el análisis de este proyecto de tesis. Se abordan temas como la problemática que se pretende resolver, el objetivo general del estudio, los objetivos específicos y la justificación de esta investigación.

En primer lugar, se expone la problemática que se busca abordar en este proyecto. Se describen los desafíos o dificultades existentes en el área de estudio y se plantea la necesidad de encontrar una solución efectiva. Esto permite establecer el contexto y la relevancia del trabajo de investigación.

A continuación, se establece el objetivo general del estudio. Este objetivo representa la meta principal que se desea alcanzar con la investigación. Se describe de manera clara y concisa qué es lo que se pretende lograr, brindando una orientación clara sobre el propósito del proyecto.

Además del objetivo general, se formulan los objetivos específicos. Estos objetivos detallados descomponen el objetivo general en componentes más pequeños y alcanzables. Cada objetivo específico representa una etapa o tarea que contribuye al logro del objetivo general. Estos objetivos específicos brindan una estructura clara para el desarrollo del trabajo y permiten evaluar el progreso a lo largo del proceso de investigación.

Finalmente, se expone la justificación de este trabajo de investigación. Aquí se exponen las razones y argumentos que respaldan la importancia y relevancia de llevar a cabo esta investigación. Se presentan los beneficios potenciales, tanto teóricos como prácticos, que podrían derivarse de los resultados y conclusiones obtenidos. Asimismo, se pueden mencionar brevemente las limitaciones existentes en investigaciones previas y cómo este estudio busca superarlas.

En resumen, este capítulo proporciona una visión general de los aspectos clave del proyecto de tesis, incluyendo la problemática a resolver, los objetivos generales y específicos, así como la justificación de la investigación.

### **1.1.1 Demanda de energía**

La energía eléctrica la usamos dependiendo de las diversas actividades donde la necesitamos, por ejemplo, el hogar, la industria, el comercio o el entretenimiento; cada una de estas actividades tienen su propio patrón de uso y comportamiento. A la suma de todo esto le llamamos demanda. La demanda de la energía eléctrica es decisiva en el desarrollo económico, social y ambiental y, por consiguiente, en el mejoramiento de las condiciones económicas del país.

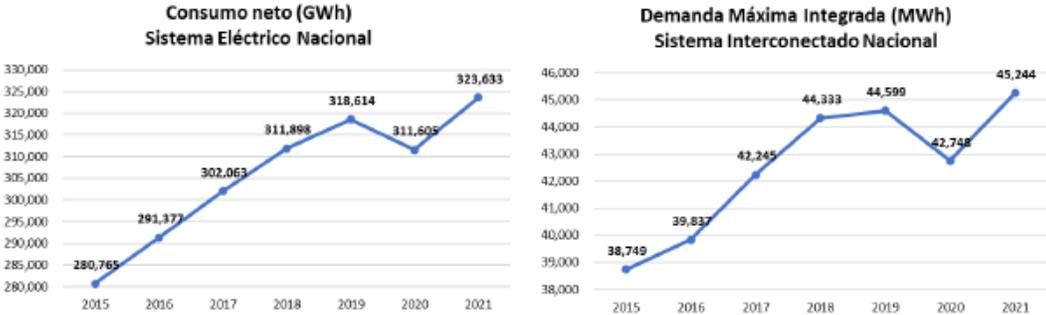
La planeación de la expansión del sistema eléctrico nacional (SEN) toma como base los pronósticos de incremento de demanda y consumo de energía eléctrica. Para ello se estiman tres variables importantes para calificar la demanda: el consumo nacional que se refiere al total de la energía utilizada para los usuarios finales, el consumo de la red que es la energía generada en las centrales y la demanda máxima que corresponde a la potencia eléctrica que debe ser importada para satisfacer las necesidades de los usuarios.

El consumo es la cantidad de energía eléctrica consumida en un periodo de tiempo, presentada en Terawatt-horas (TWH) al año. El pronóstico del consumo nacional estima las necesidades energéticas de los distintos usuarios finales. El consumo de la red permite planear la generación que debe estar disponible en el sistema, incluyendo las distintas tecnologías de generación y considerando las indisponibilidades y degradaciones de las centrales eléctricas.

Por otro lado, la demanda máxima se refiere a la cantidad máxima de energía consumida en un instante de tiempo dado, la cual es medida en Giga watts (GW). Es una medida fundamental para la planeación, ya que indica la generación que

requiere el sistema a cada instante, sin incluir la energía de reserva para mantener la seguridad y confiabilidad del sistema.

En la siguiente gráfica se indican los valores históricos del consumo neto de energía del SEN y la demanda máxima. En el periodo de los años 2015 a 2018 se observa un crecimiento sostenido; en 2019 se observa un ligero decremento; para 2020 se presenta una caída importante por la pandemia de COVID-19; y para 2021 una recuperación importante como se muestra en la Figura 1.



*Figura 1 Valores históricos del consumo neto de energía*

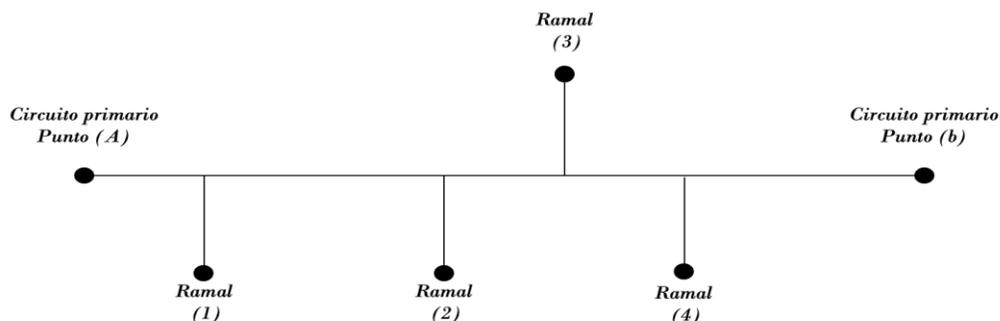
Se espera que los usuarios continúen incrementado su consumo y que la demanda máxima crezca al 2.9 % para el periodo comprendido de 2023-2027.

El servicio público de distribución de energía eléctrica se distribuye a 47 millones de usuarios aproximadamente. En la siguiente Figura 2 se muestra el análisis de consumo por segmento en GWH a junio de 2022.

| Segmento   | Consumo (GWh)  | Ventas MDP       | Clientes          |
|--|----------------|------------------|-------------------|
|  DOMÉSTICO ALTO CONSUMO   | 412            | \$2,626          | 137,834           |
|  ALUMBRADO                | 1,962          | \$7,616          | 179,409           |
|  RIEGO AGRÍCOLA           | 8,252          | \$5,461          | 134,931           |
|  COMERCIAL                | 7,037          | \$26,076         | 4,402,648         |
|  INDUSTRIAL ALTA TENSIÓN  | 14,264         | \$25,360         | 974               |
|  DOMÉSTICO                | 30,296         | \$39,985         | 41,795,882        |
|  INDUSTRIAL MEDIA TENSIÓN | 39,316         | \$93,332         | 424,320           |
|  | <b>101,539</b> | <b>\$200,456</b> | <b>47,075,998</b> |
| <b>ÚLTIMO RECURSO</b>  | <b>308</b>     | <b>\$ 676</b>    | <b>0</b>          |
|  | <b>101,846</b> | <b>\$201,132</b> | <b>47,075,998</b> |

*Figura 2 Consumo de energía por segmento*

La energía eléctrica se distribuye a través de los sistemas de distribución Figura 3 en circuitos primarios (media tensión) y circuitos secundarios (baja tensión) que se transporta de la subestación eléctrica hasta el usuario final, un circuito primario se compone de determinados componentes como son conductores, bancos de transformación, equipo de protección y desconexión, equipos de regulación y capacitores entre otros. De ellos se derivan ramales que es una porción del circuito comprendido entre dos nodos consecutivos, es decir, es una línea de tensión eléctrica que se deriva de un circuito principal que distribuye la energía eléctrica a los puntos de conexión del usuario final. Los valores de la tensión nominal entre fases de los circuitos primarios de media tensión de los sistemas de distribución son de 13.2 Kilo volts (KV), 22.86 KV y 33 KV.



### **1.1.2 Modelo de predicción**

Un modelo de predicción se utiliza para predecir eventos o resultados futuros basados en el análisis de datos. Se define como un proceso en el cual se recopilan, se analizan y se utilizan datos con el objetivo de predecir una o más variables de interés.

En este proceso, se recopilan datos relevantes que están relacionados con la variable o variables que se pretenden predecir. Estos datos pueden provenir de diversas fuentes, como registros históricos, mediciones, encuestas u otras fuentes de información disponibles.

Una vez recopilados los datos, se realiza un análisis detallado para identificar patrones, tendencias o relaciones significativas que existan entre los datos recopilados y la variable objetivo. Esto implica la aplicación de técnicas estadísticas, algoritmos de aprendizaje automático u otras metodologías apropiadas para extraer información relevante y construir un modelo de predicción.

El modelo de predicción se basa en los datos analizados y busca capturar la relación entre las variables de entrada (características o predictores) y la variable objetivo que se desea predecir. Utilizando este modelo, se pueden hacer estimaciones o proyecciones de los valores futuros de la variable objetivo en función de los valores de las variables de entrada.

Es importante tener en cuenta que los modelos de predicción están sujetos a incertidumbre, ya que no pueden predecir con certeza el futuro. Sin embargo, mediante el uso adecuado de datos y técnicas de modelado, se puede obtener una estimación razonable y útil de los eventos o resultados futuros.

En resumen, un modelo de predicción es una herramienta que utiliza datos para predecir variables futuras. Es un proceso que implica la recopilación de datos, el análisis de estos y la construcción de un modelo que permite hacer estimaciones o

proyecciones de eventos o resultados futuros. En la Figura 4 se presentan varios de los factores que están presentes en todo proceso de predicción.

Se representan en 3 vectores: el vector A representa aquello que se va a predecir, el B representa la información utilizada para la predicción como el conjunto de datos de y el C representa los diversos métodos aplicados a la predicción.

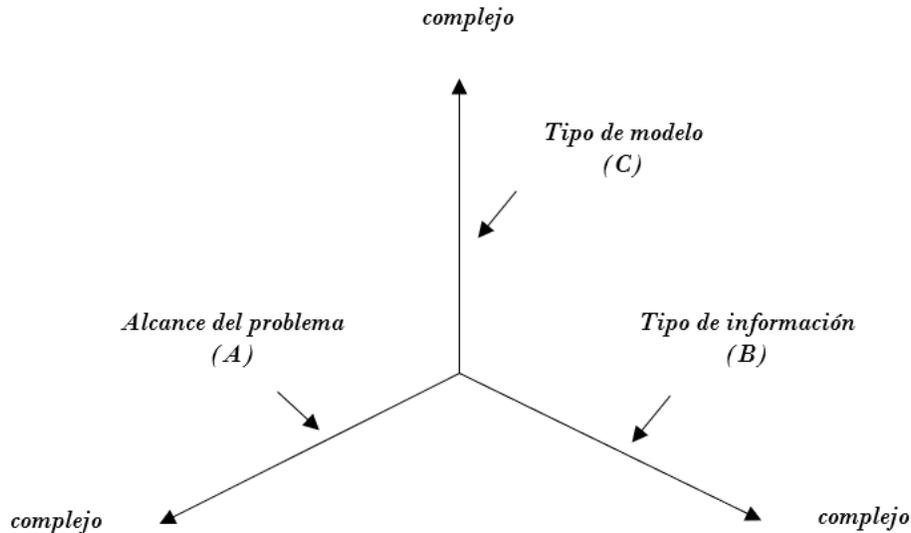


Figura 4. Parámetros de predicción

Los vectores están interrelacionados y parten de un punto central, el cual representa un origen que representa una posición de la cual provienen los estados futuros de complejidad creciente de la problemática a resolver. Cerca del origen los datos objetivos y los conocimientos constituyen una fuente de la cual se pueden formar constructos razonablemente confiables y que se pueden aplicar para el análisis de predicción de nuestro modelo.(Salvendy, 2001)

### 1.1.3 Inteligencia computacional

La inteligencia artificial es un campo de estudio dentro de la informática que se centra en el desarrollo de sistemas y algoritmos capaces de realizar tareas que requieren inteligencia humana. Esta disciplina abarca una amplia gama de áreas,

como el aprendizaje automático, la visión por computadora, el procesamiento del lenguaje natural y la robótica.

Dentro de la inteligencia artificial, existe una rama específica conocida como "aprendizaje automático" o "machine learning" en inglés. Esta rama se ocupa del desarrollo de algoritmos y modelos que permiten a las máquinas aprender de los datos y mejorar su rendimiento con el tiempo, sin ser programadas explícitamente para cada tarea (Rutkowski, 2010).

En el contexto del aprendizaje automático, los mecanismos adaptativos son fundamentales. Estos mecanismos permiten a los sistemas inteligentes ajustar y actualizar sus modelos internos en función de nuevos datos y situaciones cambiantes. Esto les permite adaptarse y mejorar su rendimiento a medida que se enfrentan a nuevos desafíos.

La representación simbólica del conocimiento es otro aspecto importante dentro de la inteligencia artificial. Consiste en capturar y representar el conocimiento humano en forma de símbolos y reglas lógicas, que luego pueden ser utilizados por los sistemas inteligentes para inferir y tomar decisiones.

#### **1.1.4 Ciencia de datos**

La ciencia de datos, como disciplina, se ha desarrollado y ha ganado relevancia en las últimas décadas, con importantes avances y contribuciones a partir de la década de 1990. En este periodo, se establecieron las bases fundamentales para la clasificación y el análisis de los datos, que son aspectos clave en el campo de la ciencia de datos (Liu, 2015).

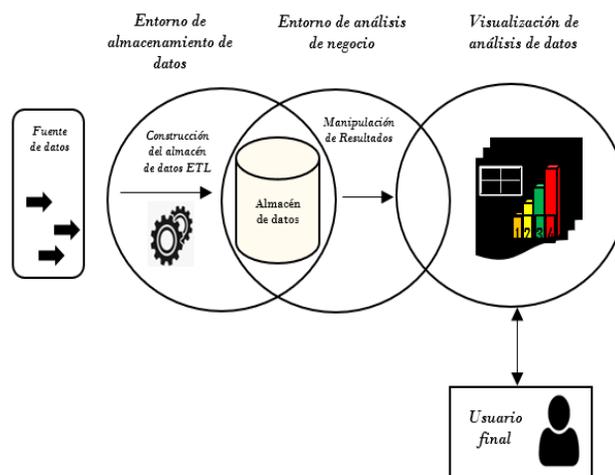
La clasificación de los datos se refiere al proceso de organizar y categorizar los datos en diferentes grupos o categorías, con el objetivo de comprender y extraer información significativa de ellos. Esto se logra a través de técnicas y algoritmos que identifican patrones y características distintivas en los datos, permitiendo asignarlos a clases o categorías específicas. Por otro lado, el análisis de datos se centra en

examinar, interpretar y extraer conocimiento útil de los conjuntos de datos. Esto implica la aplicación de diversas técnicas estadísticas, algoritmos de aprendizaje automático y herramientas de visualización para descubrir patrones, tendencias y relaciones ocultas en los datos.

A medida que la tecnología avanzaba y se hacía más accesible, y con la creciente disponibilidad de grandes volúmenes de datos (Big Data), la ciencia de datos experimentó un crecimiento significativo. Se comenzaron a utilizar técnicas más sofisticadas, como el aprendizaje automático y la minería de datos, para analizar y extraer conocimiento de los datos de manera más precisa y eficiente.

La ciencia de datos ha influido en diversos campos, como el marketing, la medicina, la industria y la investigación científica, entre otros. Ha demostrado ser una disciplina poderosa para tomar decisiones basadas en datos y aprovechar el potencial de los conjuntos de datos existentes.

La ciencia de datos tiene su base en la aplicación de modelo de negocios ver la Figura 5.



*Figura 5 Ciencia de datos como modelo BI*

### 1.1.5 Dataset

Los datos son la parte fundamental para cualquier iniciativa de análisis de negocio como lo son el modelo de negocios (BI), la ciencia de datos y el análisis empresarial. Los datos son la parte principal de lo que producen estas tecnologías de decisión como son la información, el ingenio y el conocimiento. Sin datos ninguna de estas tecnologías podría existir. Por lo que, una vez comprendido el desafío para recopilar, almacenar y administrar los datos, hoy en día se consideran entre los activos más importantes de una organización, con la capacidad de crear información valiosa para comprender mejor a los clientes, los competidores y los procesos comerciales ver Figura 6 (Sharda et al., n.d.).

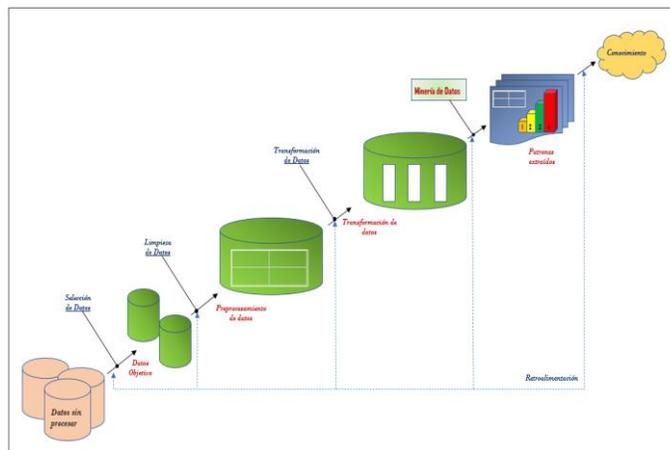


Figura 6 Dataset estructura de datos

### 1.1.6 Minería de datos

El análisis predictivo de datos es el arte de poder construir y usar modelos que hacen predicciones basados en patrones extraídos de datos históricos. Las organizaciones actuales recopilan cantidades masivas de datos y estos deben analizarse para extraer información que pueda usarse para tomar mejores decisiones y ser de gran valor para una organización

La minería de datos es un término que explica el descubrimiento o extracción de conocimientos a partir de grandes cantidades de datos. La minería de datos es un proceso que utiliza técnicas estadísticas, matemáticas y de inteligencia artificial para extraer e identificar información útil y de conocimientos (patrones) posteriores a partir del análisis de grandes cantidades de datos. Estos patrones pueden tener la forma de correlaciones, tendencias o modelos de predicción. (Sharda, Delen, Turban, et al., n.d.)

La minería de datos está estrechamente relacionada con estas disciplinas ver Figura 7.

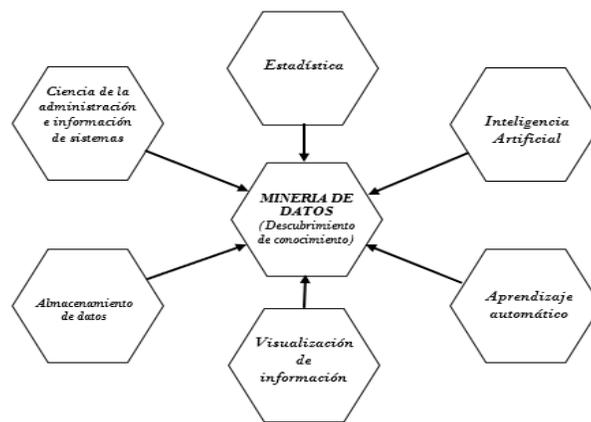
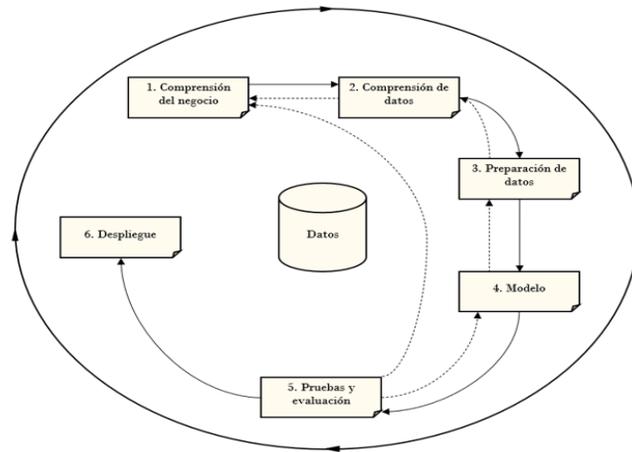


Figura 7 Relación de la minería de datos con otras disciplinas

### 1.1.7 Metodología CRISP-DM

CRISP-DM (Cross Industry Standard Process for Data Mining) es una metodología que utiliza la minería de datos y fue desarrollado originalmente por un grupo de organizaciones, compuesto por líderes de ciencia de datos, usuarios finales, empresas de consultoría e investigadores.

El ciclo de vida de CRISP-DM consta de seis etapas las cuales son: comprensión del negocio, comprensión de los datos, preparación de los datos, modelado, evaluación e implementación ver Figura 8.



*Figura 8 Metodología CRISP-DM*

En las dos primeras etapas, comprensión del negocio y comprensión de datos, el analista de datos intenta definir los objetivos del proyecto al entender las necesidades del negocio y de los datos que la empresa tiene disponibles.

El análisis de datos y selección de características de un problema de negocios son las actividades necesarias para construir un conjunto de datos apropiados que están disponibles para a la siguiente fase del CRISP-DM que es la preparación de datos. El enfoque de la etapa de preparación de datos es la creación de un conjunto de datos que se puede utilizar para el análisis de datos, que incluyen la selección de tablas, registros y atributos, así como la transformación y limpieza de los datos.

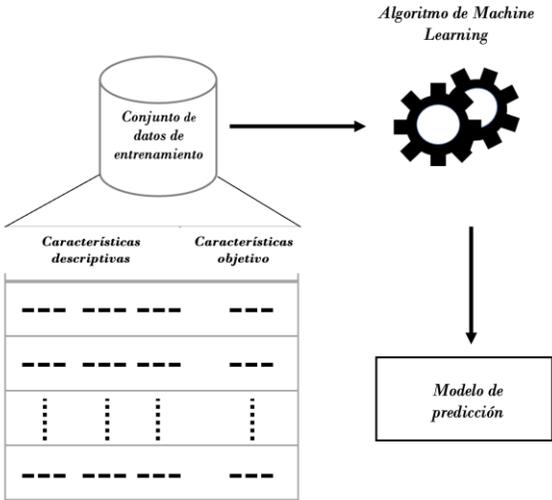
En la etapa de modelado de CRISP-DM se crea un modelo que se entrena en un conjunto de datos mediante la ejecución de un algoritmo para identificar patrones útiles en los datos y devolver un modelo que codifica estos patrones. Generalmente no se sabe qué patrones son los mejores para buscar en los datos, por lo que tiene sentido experimentar con una serie de algoritmos diferentes y ver qué algoritmo devuelve modelos más precisos cuando se ejecutan en el conjunto de datos, el aprendizaje automático es el campo de la informática que se encarga del diseño de estos algoritmos.

En las últimas dos etapas del proceso CRISP-DM, evaluación e implementación, se enfocan principalmente en como los modelos analizados se ajustan al negocio y sus procesos.

Si el proceso de evaluación aprueba un modelo o varios modelos, el proyecto pasa a la etapa final del proceso: implementación. La fase de implementación consiste en examinar cómo implementar los modelos seleccionados en el entorno empresarial y en la toma de decisiones de negocio.(Kelleher, 2018) .

### 1.1.8 Aprendizaje automático (Machine Learning)

El aprendizaje automático es una forma de inteligencia artificial (AI) que permite que un sistema aprenda de un determinado conjunto de datos. Utiliza una variedad de algoritmos que aprenden iterativamente de los datos para describir y predecir resultados. A medida que se incorporan datos de entrenamiento, es posible predecir modelos más precisos basados en el análisis de esos datos. Un modelo de aprendizaje automático es el resultado que se genera cuando se entrena un algoritmo de aprendizaje automático junto con el conjunto de datos. En el entrenamiento, cuando se proporciona un modelo con una entrada ejecutará por lo general una determinada salida. Cuando se proporcionen datos al modelo recibirá una predicción basada en los datos que entrenaron el modelo (Hurwitz, 2018).



### **1.1.9 Aprendizaje supervisado**

Las técnicas de aprendizaje automático se utilizan para mejorar la precisión de los modelos predictivos. Dependiendo de la naturaleza del problema abordado, como en nuestro caso de predicción de energía, existen diferentes enfoques según el tipo y volumen de datos. Uno de estos enfoques es el aprendizaje supervisado, el cual comienza generalmente con un conjunto establecido de datos y cierto conocimiento sobre cómo se clasifican dichos datos.

El aprendizaje supervisado tiene como objetivo encontrar patrones en los datos que se puedan aplicar en un proceso de análisis. Estos datos están etiquetados con características que definen el significado de los mismos. Esto significa que se cuenta con ejemplos previos en los que se conoce la respuesta o la clasificación correcta de los datos, lo que permite entrenar el modelo utilizando estos datos etiquetados.

Durante el proceso de entrenamiento, el modelo aprende a reconocer y generalizar los patrones presentes en los datos etiquetados. Una vez entrenado, el modelo puede aplicarse a nuevos datos para realizar predicciones o clasificaciones basadas en los patrones aprendidos.

El aprendizaje supervisado es especialmente útil cuando se cuenta con datos etiquetados de calidad, ya que permite crear modelos más precisos y confiables. Sin embargo, es importante tener en cuenta que la calidad y representatividad de los datos de entrenamiento son cruciales para obtener resultados precisos y evitar sesgos o errores en las predicciones. En la Figura 10 se da el esquema de los diferentes algoritmos de aprendizaje.

| Data Mining Tasks & Methods | Data Mining Algorithms   | Learning Type |
|-----------------------------|--|---------------|
| <b>Prediction</b>           |  |               |
| Classification              | Decision Trees, Neural Networks, Support Vector Machines, kNN, Naive Bayes, GA | Supervised    |
| Regression                  | Linear/Nonlinear Regression, ANN, Regression Trees, SVM, kNN, GA               | Supervised    |
| Time series                 | Autoregressive Methods, Averaging Methods, Exponential Smoothing, ARIMA        | Supervised    |
| <b>Association</b>          |  |               |
| Market-basket               | Apriori, OneR, ZeroR, Eclat, GA  | Unsupervised  |
| Link analysis               | Expectation Maximization, Apriori Algorithm, Graph-Based Matching              | Unsupervised  |
| Sequence analysis           | Apriori Algorithm, FP-Growth, Graph-Based Matching                             | Unsupervised  |
| <b>Segmentation</b>         |  |               |
| Clustering                  | k-means, Expectation Maximization (EM)   | Unsupervised  |
| Outlier analysis            | k-means, Expectation Maximization (EM)   | Unsupervised  |

Figura 10 Esquema de algoritmos de aprendizaje

### 1.1.10 Regresión lineal

El análisis de regresión es una técnica estadística que determina la relación entre una variable dependiente (por ejemplo, la demanda de un producto o servicio) y la variable independiente (por ejemplo, el tiempo). La fórmula general de regresión entre la variable dependiente  $y$  (o de respuesta) y la variable independiente  $x$  (predictores) está dada como la Formula 1:

$$y = b_0 + b_1x + b_2x^2 + \dots + b_nx^n + \varepsilon \quad (1)$$

donde  $b_0, b_1, \dots, b_n$  son parámetros desconocidos. El error aleatorio  $\varepsilon$  tiene una media cero y una desviación estándar constante ó en su forma simplificada de sumatoria como la Formula 2.

$$Y_i = \sum_{j=0}^P b_j X_{ij} \quad (2)$$

### **1.1.11 Algoritmo computacional**

Un algoritmo se define o se describe como un proceso (programa) que una computadora puede seguir. En el contexto del aprendizaje automático se define como un proceso de búsqueda para analizar un conjunto de datos y elegir la mejor función de un conjunto de funciones posibles con el fin de identificar patrones recurrentes y poder explicar las relaciones entre las características de un conjunto de datos.

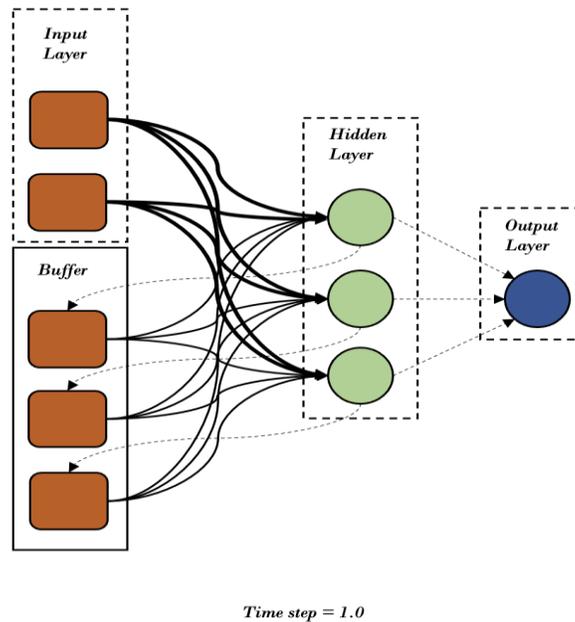
### **1.1.12 Aprendizaje profundo (Deep Learning)**

El aprendizaje profundo permite tomar decisiones basadas en datos al identificar y extraer patrones de grandes conjuntos de datos que se mapean con precisión desde conjuntos de entradas complejas hasta una decisión de salida de resultados. Es un subcampo de la inteligencia artificial que se enfoca en crear grandes modelos de redes neuronales, los patrones que los algoritmos de aprendizaje profundo extraen de los conjuntos de datos son funciones que se representan como redes neuronales profundas. (Kelleher,2019)

### **1.1.13 Redes Neuronales Artificiales**

En su forma más general una red neuronal artificial ver Figura 11 es un proceso informático diseñado para modelar la forma en que el cerebro realiza una tarea o una determinada función de interés particular. Consisten principalmente en una red de unidades simples de procesamiento de información, llamadas neuronas. Una red neuronal es un procesador distribuido masivamente en forma de paralelo, el cual

está compuesto por un procesamiento simple de unidades que tienen como objetivo una tendencia natural de almacenar conocimiento y tenerlo a disposición para la ejecución de un análisis. Por lo que destacan aspectos similares al cerebro humano, como son por una parte la adquisición de conocimiento por la red de su entorno a través de un proceso de aprendizaje y por otra las fuerzas de conexión de las interneuronas como pesos sinápticos para almacenar la información.(Haykin, 2009)



*Figura 11 Representación de una Red Neuronal*

### **1.1.14 Modelo computacional**

Un modelo computacional se refiere a un modelo matemático utilizado en ciencias de la computación para estudiar el comportamiento de sistemas complejos mediante la simulación por computadora. Este modelo requiere de recursos computacionales para realizar dicha simulación y analizar el sistema en cuestión.

Por otro lado, el aprendizaje automático implica un proceso que consta de dos etapas: entrenamiento e inferencia. Durante el entrenamiento, se procesa un conjunto de datos conocidos como datos de entrenamiento. En esta etapa, el modelo

busca encontrar una función que mejor se ajuste a los patrones y características presentes en los datos. Esto implica seleccionar y ajustar parámetros o pesos del modelo para minimizar el error entre las predicciones del modelo y los valores reales de los datos de entrenamiento.

Una vez que el modelo ha sido entrenado, se puede utilizar en la etapa de inferencia. En esta etapa, el modelo se aplica a nuevos ejemplos o datos sin etiquetar y genera estimaciones o predicciones del valor de una función basándose en los patrones aprendidos durante el entrenamiento. La inferencia implica utilizar el modelo para realizar cálculos y obtener resultados precisos y útiles a partir de los nuevos datos.

### **1.1.15 Lenguaje de programación**

Un programa de computadora es un conjunto de instrucciones paso a paso que le indican a la computadora como realizar una tarea específica, y en cada momento se puede elegir ejecutar un programa de acuerdo con la tarea que se requiere realizar. Las instrucciones se deben escribir en un lenguaje en la que la computadora pueda entender. Los lenguajes de programación son lenguajes diseñados especialmente para dar órdenes a una computadora de manera exacta y no interpretarse de diferentes formas.

La solución de un problema por computadora consiste elaborar algoritmos y su respectiva implementación computacional, en algún lenguaje de programación susceptible de ser ejecutado en una computadora para resolver una clase de problemas.(Carrillo Ledesma & Rosas, n.d.)

### **1.1.16 Mapa de calor**

Un mapa de calor es una formación de rectángulos que esta defina por un eje  $x$  que representa por lo general una medida de tiempo, representada por cualquier variable formada por agrupaciones. El eje  $y$  es una variable que define por ciertas categorías de los datos. Todos los rectángulos tienen la misma estructura de tamaño y están identificados por cierto color para mostrar la magnitud de una tercera variable de

magnitud. Los mapas de calor en un principio se usaban para temperaturas y ahora son utilizadas para representar una gran cantidad de tipos de datos.

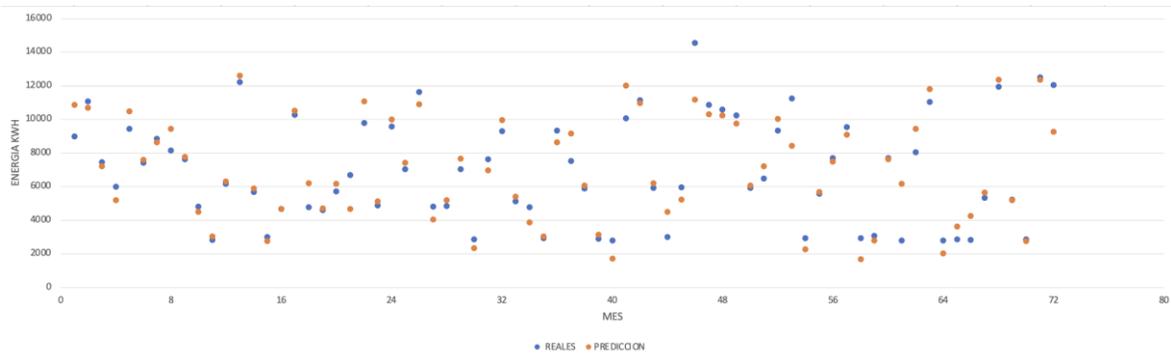
Los mapas de calor son muy útiles para grandes conjuntos de datos. Un mapa de calor con un eje temporal se utiliza para visualizar patrones y cambios en el tiempo. Los rectángulos del mapa de calor se etiquetan con valores de la variable de color, lo cual solo es útil en casos en los que haya muy pocas categorías en el eje como se muestra en la Figura 12.



Figura 12 Representación de un mapa de calor

### 1.1.17 Gráfica de dispersión

Los gráficos de dispersión ver Figura 13 se utilizan para mostrar relaciones. Para la correlación, los gráficos de dispersión ayudan a mostrar la fuerza de la relación lineal entre dos variables. Para la regresión, los gráficos de dispersión suelen incorporar una línea de ajuste. En control de calidad, los gráficos de dispersión pueden con frecuencia incluir límites de especificación o líneas de referencia.



*Figura 13 Gráfica de dispersión*

### **1.1.18 Gráfico de magnitud y pérdida**

El gráfico de magnitud y pérdida ver Figura 14 es una herramienta útil para analizar y comprender el rendimiento de una red neuronal durante el entrenamiento. Proporciona una representación visual de la relación entre la magnitud de los pesos y la función de pérdida, lo que puede ayudar a identificar problemas como la estabilidad, el sobreajuste y la convergencia de la red neuronal.

El eje horizontal del gráfico representa la magnitud de los pesos o parámetros de la red neuronal. Los pesos son los valores numéricos que determinan la fuerza de conexión entre las neuronas en la red. Cada neurona en una red neuronal tiene asociados pesos que se ajustan durante el proceso de entrenamiento.

El eje vertical del gráfico representa la función de pérdida o error de la red neuronal. La función de pérdida mide la discrepancia entre la salida predicha por la red y el valor verdadero o esperado. El objetivo del entrenamiento de una red neuronal es minimizar esta función de pérdida, lo que implica ajustar los pesos de la red para obtener predicciones más precisas.

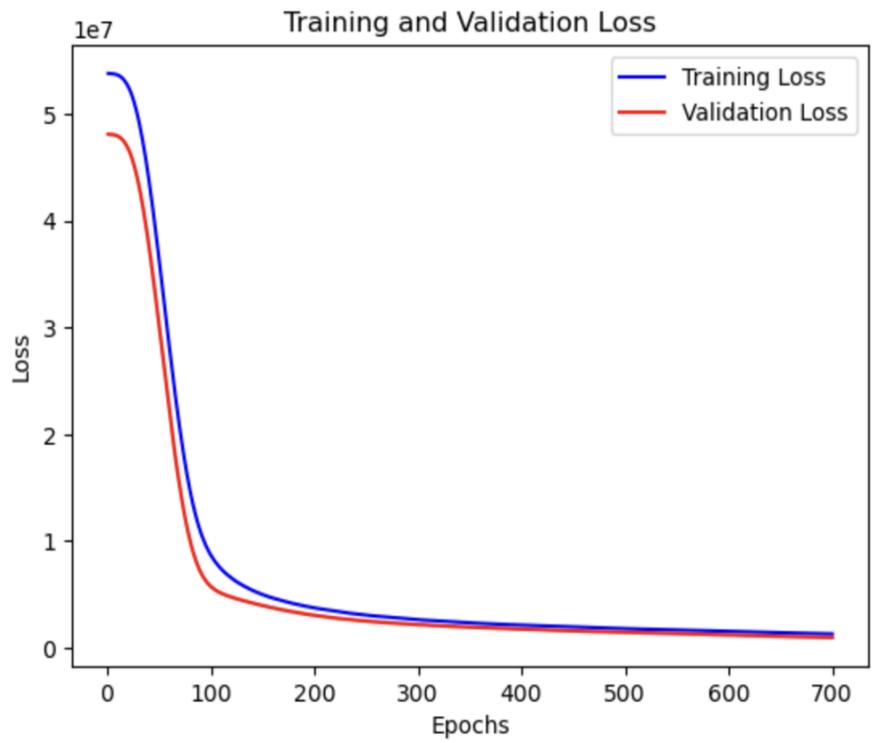
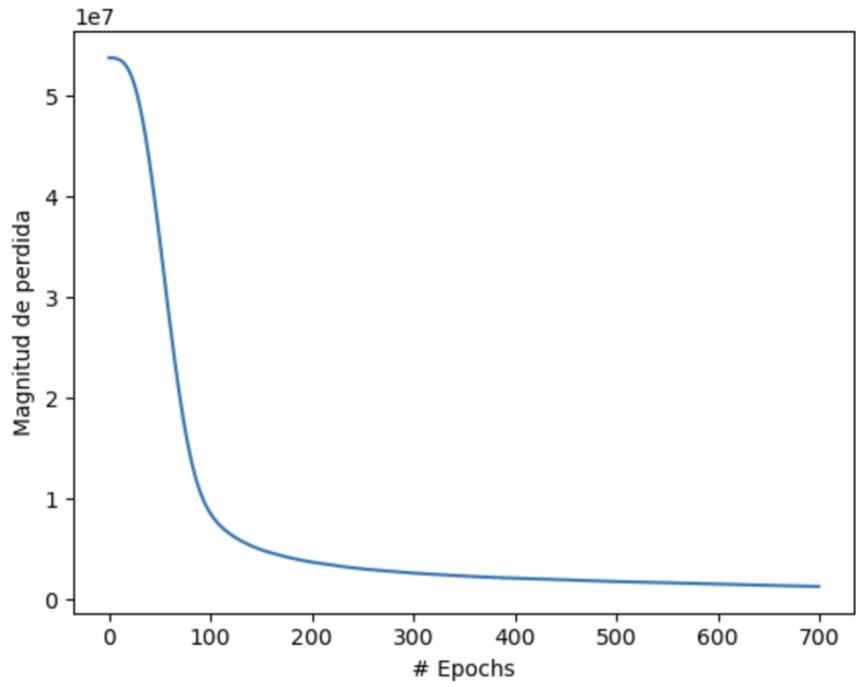


Figura 14 Gráficos de validación de magnitud y pérdida

### 1.1.19 Series de tiempo

Los métodos de series de tiempo se utilizan para predicciones a corto plazo. Una serie de tiempo se define como una lista cronológica de datos históricos en donde la suposición esencial es que la historia predice el futuro de manera razonable. Existen varios modelos y métodos entre los cuales elegir dependiendo del contexto de los datos analizados se dividen en modelo constante, de tendencia y estacional ver Figura 15.

En una serie de tiempo, el tiempo suele ser la variable independiente, mientras que la variable que se está observando y registrando se considera la variable dependiente. Por ejemplo, se pueden recopilar datos diarios de temperatura durante un año, donde el tiempo representa los diferentes días y la variable dependiente es la temperatura registrada en cada día.

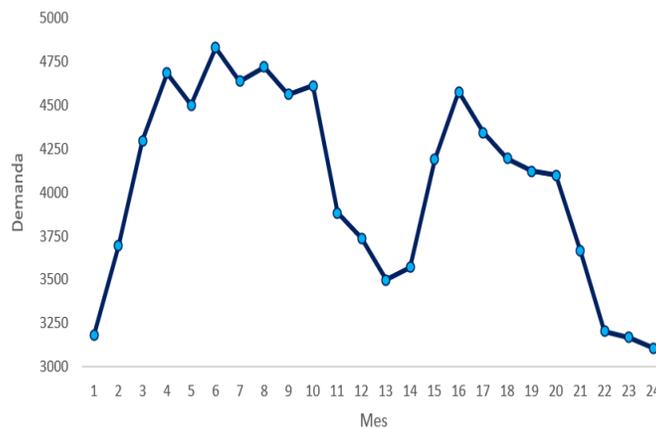


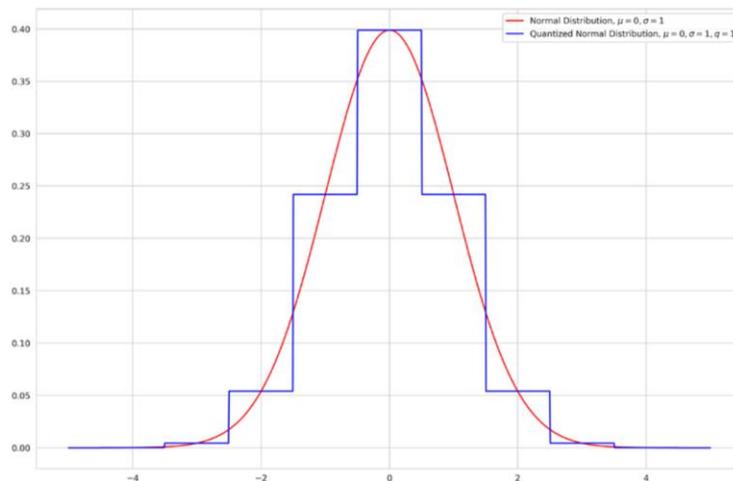
Figura 15 Gráfica serie de tiempo

### 1.1.20 Histograma de dispersión de valores

En estadística, un histograma es una representación gráfica de una distribución de frecuencia. Se utiliza para mostrar la distribución de los datos en diferentes intervalos o categorías. En un histograma, el eje horizontal representa las diferentes

categorías o intervalos, mientras que el eje vertical muestra la frecuencia con la que ocurren los valores en cada categoría.

El histograma proporciona una visualización visual de la distribución de los datos, permitiendo identificar la concentración o dispersión de los valores. También puede revelar información sobre la simetría, sesgo y curtosis de la distribución. Es una herramienta fundamental en el análisis estadístico para comprender y visualizar la estructura de los datos ver Figura 16 fuente (Géron, 2019).



*Figura 16 Histograma de dispersión de valores*

## **1.2 Planteamiento del Problema**

La demanda de energía eléctrica varía debido a diversos factores, como el encendido y apagado de dispositivos eléctricos, períodos vacacionales, días festivos y el clima de la región. En México, la demanda máxima de energía eléctrica se produce principalmente durante la temporada de verano, que generalmente abarca los meses de mayo a septiembre. Durante este período, el calor es más intenso en las zonas norte y sureste del país. Como respuesta a estas condiciones, se incrementa el uso de dispositivos de ventilación para mantener una temperatura confortable y garantizar el correcto funcionamiento de equipos eléctricos y electrónicos.

En este contexto, la idea central es proponer la creación de un modelo computacional basado en el aprendizaje profundo, que es un subcampo de la inteligencia artificial que se enfoca en crear modelos de redes neuronales. Este modelo nos permitirá analizar y tomar decisiones al identificar y extraer patrones precisos de un extenso conjunto de datos históricos complejos sobre la demanda energética. El objetivo es obtener resultados confiables y tomar decisiones óptimas en cuanto a los recursos disponibles para la generación de energía eléctrica como medida preventiva.

El modelo propuesto seguirá un proceso metodológico que involucra la creación del conjunto de datos (dataset) y la implementación de algoritmos de aprendizaje profundo para extraer información mediante funciones de mapeo de entrada, las cuales se representan a través de una red neuronal para obtener valores de salida.

La implementación de una red neuronal artificial utiliza el método de partición conocido como "divide y vencerás", ya que cada neurona de la red aprende una función simple o contribuye a una función general más compleja definida por la propia red.

El uso de redes neuronales es utilizado a medida que se aumenta el número de entradas a la función para una gran cantidad de entradas como es el caso de las variables utilizadas de la medición de la demanda de energía eléctrica y en estos casos se debe aprovechar el poder del aprendizaje profundo para buscar la mejor función para hacer coincidir o ajustar los patrones en el conjunto de datos.

Entre los beneficios esperados se encuentran los siguientes:

- Desarrollar el dataset del conjunto de datos históricos de las variables contenidas para el análisis de las funciones de predicción que el algoritmo buscará para encontrar la mejor coincidencia de los datos.

- Desarrollar un modelo computacional de forma metodología que permita realizar la predicción de la demanda de energía eléctrica en el ámbito del aprendizaje profundo de redes neuronales.

### **1.3 Justificación**

En las revisiones de la literatura encontrada no se identificaron muchos trabajos o artículos relacionados donde se propone la creación del modelo computacional como un proceso metodológico en específico que abarque el análisis de la formación del dataset hasta el algoritmo para la mejor predicción de la demanda de energía eléctrica en el ámbito de la inteligencia artificial y en específico del subcampo de aprendizaje profundo. Si bien el enfoque principal de este proyecto es el desarrollo del modelo computacional de forma metodológica de la eficiencia de la predicción de la demanda eléctrica y por otra la automatización de funciones de las redes neuronales del conjunto de datos históricos a través de los algoritmos de aprendizaje automático que permitan mediante la simulación de la computadora, establecer como solución resultados precisos en la toma de decisiones.

Por lo tanto, es fundamental entrenar el modelo de aprendizaje profundo con datos de entrenamiento adecuados y aumentados. Dentro de cada conjunto de datos la precisión de inferencia de aprendizaje profundo lograda con un modelo entrenado puede cumplir con los objetivos de precisión. No hay metodologías estructuradas conocidas disponibles para selección óptima del modelo y datos de entrenamiento para los niveles de precisión objetivo.

### **1.4 Hipótesis**

H: "El desarrollo del modelo computacional de aprendizaje automático permitirá obtener una predicción precisa y óptima de la demanda de energía eléctrica".

## **1.5 Objetivo general**

Desarrollar un modelo computacional para la predicción de demanda de energía eléctrica en la región oriente basado en técnicas de aprendizaje automático.

## **1.6 Objetivos específicos**

1. Realizar un exhaustivo análisis de la información histórica del conjunto de datos de la demanda energética con el fin de crear un dataset robusto y representativo que permita el análisis preciso de los algoritmos y la función de mapeo de aprendizaje profundo.
2. Desarrollo del modelo computacional óptimo que defina un método para analizar el comportamiento de mapeo de las variables del sistema complejo de la demanda para su implementación en el aprendizaje profundo como método de solución.
3. Seleccionar en el ámbito del cómputo científico, el lenguaje de programación adecuado como solución tecnológica para el análisis de la información de los datos de la demanda de energía eléctrica que mejor se precise en el aprendizaje profundo.
4. Establecer un caso de estudio donde se pruebe y valide el modelo computacional analizado para la predicción de la demanda de energía eléctrica por aprendizaje profundo.

## **1.7 Alcances y limitaciones**

### **1.7.1 Alcances**

- Realizar una investigación descriptiva para estudiar exhaustivamente las características fundamentales de la demanda de energía eléctrica. Este análisis permitirá el desarrollo de un modelo computacional basado en el aprendizaje profundo.

- El modelo computacional propuesto utilizará redes neuronales de regresión como una técnica óptima de aprendizaje profundo para la predicción de la demanda de energía eléctrica.
- El modelo computacional será capaz de utilizar variables de cualquier tipo de interés basadas en datos temporales, ampliando así su flexibilidad y aplicabilidad en el análisis de la demanda energética.
- Se realizará una comparación exhaustiva del modelo computacional propuesto con otros métodos del estado del arte, permitiendo evaluar su rendimiento y efectividad en la predicción de la demanda energética.

### **1.7.2 Limitaciones**

- El enfoque de este proyecto de investigación se centrará exclusivamente en el uso del modelo de predicción de aprendizaje profundo basado en redes neuronales recurrentes (RNN) de regresión, sin considerar otras tecnologías de predicción más complejas como Transformer.
- El alcance de este proyecto se limitará a la toma de decisiones en la predicción de la demanda de energía eléctrica, sin abordar la búsqueda de una solución óptima en términos de generación y distribución de energía.

## **CAPÍTULO II ESTADO DEL ARTE**

El objetivo de este capítulo es realizar una comparación entre diferentes artículos que abordan el tema del aprendizaje automático, centrándose específicamente en las metodologías utilizadas para el análisis de datos y la implementación del aprendizaje profundo en modelos de predicción de la demanda de energía. A continuación, se presenta un resumen de los trabajos seleccionados, los cuales están relacionados directa o indirectamente, total o parcialmente, con el anteproyecto propuesto.

A continuación, se presenta el resumen de trabajos seleccionados por su relación directa o indirecta y/o parcial o total con el anteproyecto propuesto.

### **2.1 Trabajos relacionados**

(Márquez et al., 2019) Proponen un enfoque para el análisis predictivo emergente de la pérdida eficiente de la energía en el consumo de energía mediante la combinación de una red neuronal artificial (ANN) con herramientas de minería de datos (DM por sus siglas en inglés) mediante la regla de asociación de minería (AR). La utilización de estas técnicas puede conducir a nuevos modelos de mantenimiento predictivo usando software que puede manejar un gran volumen de datos. Se utilizó el método de modelo funcional IDEF $\Theta$  para una solución técnica propuesta que considera la aplicación complementaria de la ANN y técnicas DM de la detección de anomalías en el comportamiento de la eficiencia energética.

(Neu et al., 2022) Centra su estudio en una revisión sistemática de la literatura para capturar los métodos de algoritmos de aprendizaje profundo de última generación para la predicción de procesos utilizando diferentes técnicas de procesamiento de datos implementado una variedad de topologías de red centrándose en varios objetivos. Se realiza una revisión de la literatura donde se analiza una comparación cualitativa de las implementaciones donde las clasifica a lo largo de diferentes aspectos como el tipo de red neuronal, tipo de predicción, características de entrada

y métodos de codificación permitiendo una evaluación cuantitativa de comparación de enfoques de predicción de procesos.

(Alzubaidi et al., 2021) En otros trabajos de la literatura revisada se abordaron en ellos solo un aspecto de aprendizaje profundo, lo que lleva una falta general de conocimiento sobre el análisis y aplicación de este enfoque. Realizan su contribución donde se propone presentar un enfoque más holístico con el fin de dar un punto de referencia al desarrollo de una comprensión completa de aprendizaje profundo. Intenta proporcionar un análisis detallado de los aspectos más importantes de Deep Learning (DL) y las mejoras agregadas recientemente en el campo. En esta investigación se destacan la importancia del aprendizaje profundo, las diferentes técnicas y redes neuronales convolucionales (CNN), así como la arquitectura de ellas junto con sus características principales. Donde finalmente presentan a fondo más desafíos y soluciones sugeridas a la comprensión de investigación de aprendizaje profundo.

(Sarker, 2021) En el aprendizaje profundo una rama del aprendizaje automático y de la inteligencia artificial, se considera hoy en día como núcleo de tecnología de la cuarta revolución industrial actual debido a sus capacidades de aprendizaje a través de datos. Sin embargo, construir un modelo es una tarea desafiante y por esa razón su trabajo de investigación presenta una visión estructurada y completa de las técnicas de aprendizaje profundo incluida una taxonomía de su clasificación ordenada y jerárquica considerando varios tipos de tareas del mundo real como supervisadas o no supervisadas. En esta taxonomía aplicada se tienen en cuenta profundas redes para el aprendizaje supervisado e híbrido. Resumen áreas de aplicación donde se pueden aplicar técnicas de aprendizaje profundo. En general pretende diseñar un panorama sobre el modelado basado en datos de aprendizaje profundo que se puede utilizar como guía de referencia en investigaciones para la toma de decisiones inteligentes.

(Ahsan et al., 2021) Se realizó un estudio de 11 algoritmos de predicción de Machine Learning (ML) juntos con 6 distintos escalados de datos que tienen un efecto en las

tecnologías de predicción de ML. Se analiza la variación del rendimiento de los diferentes modelos escalados de datos y muestran resultados del rendimiento de los algoritmos como punto de desarrollo de un modelo de diagnóstico con valores atípicos, características y relación de datos desigual. Se tienen limitaciones abordadas al escalar algoritmos cambiando los parámetros de diferentes técnicas de escalado de datos en Big data.

(Basheer et al., 2021) Se utilizan técnicas de aprendizaje profundo con la intención de desarrollar un algoritmo para proporcionar cambios en el modelo de un diseño de red de cápsulas para obtener mejores resultados de predicción y hacer un modelo computacionalmente eficiente. Se centra en el estudio de la investigación de la profundidad de la identificación de la importancia de las características, la correlación de estudio entre los factores y densidad de datos mediante el análisis exploratorio. Se lleva a cabo un análisis de varias funciones de optimización para la selección de variables y características para que el modelo sea más rápido y preciso. Este modelo se compara con clasificadores de aprendizaje profundo de última generación tomados como punto de referencia utilizando diferentes métricas de rendimiento. El modelo propuesto es computacionalmente más eficiente como factor jerárquico y con mejor precisión con otros modelos que utilizan redes neuronales profundas, CNN y red de cápsula original.

(Shirzadi et al., 2021) El objetivo principal de su estudio es desarrollar y comparar modelos para predecir la demanda de carga eléctrica en técnicas de aprendizaje automático como máquinas de vectores de soporte (SVM,) Random Forest (RF) y aprendizaje profundo como la red neuronal exógena autorregresiva no lineal (NARX) y la red neuronal recurrente (RNN-LSTM). Sobre un conjunto de datos históricos de demanda fusionada con la información climática como la temperatura. Mostrando resultados donde el aprendizaje profundo es empleado como mejor predictor de la demanda de carga eléctrica de forma precisa. El estudio demostró como novedad en el desarrollo de neural recurrente sintonizado de red (memoria a largo plazo) para la predicción de carga, el desarrollo de modelo exógeno autorregresivo no lineal

para evaluar el impacto de la red dinámica recurrente para predecir el comportamiento de la demanda de carga y la evaluación comparativa de varios modelos basados en precisión y costo computacional para la toma de decisiones.

(Jogunola et al., 2022a) Presentan un modelo híbrido de aprendizaje profundo para la predicción de la energía en diferentes edificios comerciales y domésticos de diferentes países. Utilizan una arquitectura compuesta de una capa de red neuronal convolucional (CNN) para la extracción de los datos y para el proceso de predicción en las capas compuestas de un codificador (AE) y una memoria bidireccional a largo plazo (BLSTM) y LSTM. Se recopilaron los datos de consumo real a través de medidores inteligentes de hogares, pymes y los datos climáticos compuestos de variables de temperatura, humedad, velocidad del viento y punto de rocío, así como, el índice de energía por días de la semana, fin de semana y feriados para la limpieza y estructura del conjunto de datos. Para entrenar y probar el modelo híbrido de red neuronal utilizaron un optimizador Adam, una tasa de aprendizaje de 0.001, 70 épocas, 160 tamaños de lote, una división de validación de 0.33 y función de activación RELU. Se evaluó el rendimiento a través de las métricas de error absoluto medio (MAE) y error cuadrático medio (RMSE) obteniendo valores entre el 80 % y 98.7 % de precisión, demostrando su capacidad predictiva del modelo híbrido de la energía sobre el conjunto de datos variables.

(Torres et al., 2022) Analizan una red neuronal profunda para el pronóstico del consumo eléctrico a corto plazo a través de una red de memoria a corto plazo (LSTM) debido a su capacidad de tratar con datos secuenciales, como datos de series de tiempo. La red LSTM propuesta ha obtenido menos errores en comparación con una regresión lineal, un árbol de decisión, dos conjuntos de árboles y dos redes neuronales profundas Feed-forward optimizada mediante una búsqueda aleatoria y un algoritmo de muestreo. Los resultados reportan predicciones muy precisas alcanzando errores inferiores al 1.5 %.

(Hewage et al., 2021) Se propone un modelo de pronóstico del tiempo basado en datos mediante la exploración del enfoque de modelo temporal de memoria a corto

plazo LSTM y redes convolucionales temporales (TCN). En comparación con los modelos regresión de vector soporte (SVR) y Random Forest (RF) como enfoques clásicos de aprendizaje automático. Se tiene una información meteorológica capturada por series de tiempo que se exploran con los modelos LSTM y TCN de última generación, como una forma especializada de red neuronal para la predicción meteorológica. La propuesta consta de una serie de capas que utilizan parámetros meteorológicos superficiales durante un periodo de tiempo determinado para el clima del pronóstico que se evalúan en dos regresiones diferentes. El modelo propuesto produce mejores resultados en comparación con el conocido y complejo modelo pronóstico del tiempo WRF.

(Bacanin et al., 2023) Abordan la predicción de la carga de energía a través de varios factores, como la temperatura, la carga conectada entre otras. Mediante los enfoques de aprendizaje automático y aprendizaje profundo se pronostica a través series de tiempo. El estudio actual de aprendizaje profundo no enfatiza en el ajuste de parámetros para mejorar resultados. Por lo que en su investigación realizada desarrollan un modelo de aprendizaje profundo canónico de memoria a largo plazo LSTM para la carga de electricidad y se ajusta al pronóstico de series de tiempo multivariado. Su investigación destaca que el ajuste de parámetros mejora resultados cuando se utilizan metaheurísticas para este propósito en todos los casos. Se revisaron algunas arquitecturas de DL más recientes para el modelado energético con énfasis en la predicción de carga de energía con el objetivo de proponer un modelo apropiado para demostrar que el ajuste de hiperparámetros a través de metaheurísticas conduce a resultados considerablemente mejores. Enfoques de DL más complejos como BiLSTM o unidades concurrentes mejorarían los resultados.

## 2.2 Análisis comparativo de los trabajos relacionados.

La Tabla 1 contiene información de comparación de cada uno de los trabajos relacionados con el proyecto de tesis con el fin de examinar las semejanzas y diferencias más significativas.

*Tabla 1. Análisis comparativo*

| Autor                  | Título del artículo   | Problema  | Resultado   | Estado     |
|------------------------|---|---|---|------------|
| (Márquez et al., 2019) | A process to implement an artificial neural network and association rules techniques to improve asset performance and energy efficiency | Análisis predictivo emergente de la pérdida eficiente de la energía en el consumo de energía. | Se utilizó el método de modelo funcional IDEF0 para una solución técnica que considera la aplicación complementaria de la ANN y técnicas DM de la detección de anomalías en el comportamiento de la eficiencia energética | Finalizado |
| (Neu et al., 2022)     | A systematic literature review on state-of-the-   | Revisión sistemática de la literatura   | Se analiza una comparación cualitativa  | Finalizado |

|                          |   |   |  |            |
|--------------------------|---|---|--|------------|
|                          | art deep learning methods for process prediction  | para capturar los métodos de algoritmos de aprendizaje profundo de última generación para la predicción de procesos utilizando diferentes técnicas de procesamiento de datos. | donde se clasifica el tipo de red neuronal y tipo de predicción permitiendo una evaluación cuantitativa de comparación de enfoques de predicción.        |            |
| (Alzubaidi et al., 2021) | Review of deep learning: concepts, CNN architectures, challenges, applications, future directions | Revisión de la literatura donde existe una falta general de conocimiento sobre el análisis y aplicación del aprendizaje profundo  | Se propone presentar un enfoque más holístico con el fin de dar un punto de referencia al desarrollo de una comprensión completa de aprendizaje profundo | Finalizado |

|                      |   |  |  |            |
|----------------------|---|--|--|------------|
| (Sarker, 2021)       | Deep Learning: A Comprehensive Overview on Techniques, Taxonomy, Applications and Research Directions | Construcción de un modelo de investigación donde se presenta una visión estructurada y completa de las técnicas de aprendizaje profundo.           | Taxonomía aplicada para modelado basado en datos de DL que se puede utilizar como guía de referencia en investigaciones para la toma de decisiones inteligentes.                                   | Finalizado |
| (Ahsan et al., 2021) | Effect of Data Scaling Methods on Machine Learning Algorithms and Model Performance                   | Estudio de 11 algoritmos de predicción de ML juntos con 6 distintos escalados de datos que tienen un efecto en las tecnologías de predicción de ML | Se analiza la variación del rendimiento de los diferentes modelos escalados de datos y muestran resultados del rendimiento de los algoritmos como punto de desarrollo de un modelo de diagnóstico. | Finalizado |

|                         |  |   |  |            |
|-------------------------|--|---|--|------------|
| (Basheer et al., 2021)  | Computational Modeling of Dementia Prediction Using Deep Neural Network: Analysis on OASIS Dataset | Se centra en el estudio de la investigación de la importancia de las características, de la correlación de estudio entre los factores y densidad de datos mediante el análisis exploratorio | El modelo se compara con clasificadores de aprendizaje profundo CNN de última generación utilizando diferentes métricas de rendimiento y con mejor precisión | Finalizado |
| (Shirzadi et al., 2021) | Medium-term regional electricity load forecasting through machine learning and deep learning       | El objetivo principal de su estudio es desarrollar y comparar modelos para predecir la demanda de carga eléctrica en técnicas de aprendizaje automático                                     | Mostrando resultados donde el aprendizaje profundo LSTM es empleado como mejor predictor de la demanda de carga eléctrica de forma precisa                   | Finalizado |

|                                 |  |   |  |                   |
|---------------------------------|--|---|--|-------------------|
| <p>(Jogunola et al., 2022b)</p> | <p>CBLSTM-AE: A Hybrid Deep Learning Framework for Predicting Energy Consumption</p> | <p>Predicción del consumo de energía de diferentes tipos de edificios</p>   | <p>Se utilizó una capa CNN para extracción de características importantes del conjunto de datos y las capas AE-BLSTM y LSTM se utilizaron para la predicción y se obtuvieron métricas de rendimiento precisas.</p> | <p>Finalizado</p> |
| <p>(Torres et al., 2022)</p>    | <p>A deep LSTM network for the Spanish electricity consumption forecasting</p>       | <p>Analizan una red neuronal profunda para el pronóstico del consumo de eléctrico a corto plazo a través de una red de memoria LSTM</p> | <p>La red LSTM propuesta ha obtenido menos errores en comparación con una regresión lineal, un árbol de decisión, dos conjuntos de árboles y dos redes neuronales</p>  | <p>Finalizado</p> |

|                        |  |   |  |            |
|------------------------|--|---|--|------------|
|                        |  |   | profundas Feed-forward.  |            |
| (Hewage et al., 2021)  | Deep learning-based effective fine-grained weather forecasting model   | Propone un modelo pronóstico basado en series de tiempo   | Su propuesta consta de una serie de capas que utilizan parámetros meteorológicos superficiales durante un periodo de tiempo determinado para el clima del pronóstico que se evalúan en dos regresiones diferentes LSTM y TCN | Finalizado |
| (Bacanin et al., 2023) | On the Benefits of Using Metaheuristics in the Hyperparameter Tuning of Deep Learning Models for Energy Load Forecasting | Abordan la predicción de la carga de energía a través de varios factores, como la temperatura, la carga | Desarrollan un modelo de aprendizaje profundo canónico de memoria a largo plazo LSTM para la carga de electricidad y se  |            |

|  |  |                           |  |  |
|--|--|---------------------------|--|--|
|  |  | conectada<br>entre otras. | ajusta al<br>pronóstico de<br>series de tiempo<br>multivariado |  |
|--|--|---------------------------|--|--|

Después de haber analizado los trabajos relacionados con el proyecto de tesis, se concluye que el enfoque principal de estudio son los pronósticos del consumo eléctrico a través del tiempo con técnicas de aprendizaje profundo mediante holísticas, taxonomía y modelos computacionales. En este proyecto se pretende desarrollar una metodología óptima basada en un modelo computacional para la predicción de la demanda de energía eléctrica de forma regional con el objetivo de analizar un caso de uso donde exista la posibilidad de aplicar algoritmos de autoaprendizaje profundo de redes neuronales midiendo varias características o variables de predicción de la demanda de energía eléctrica.

## **2.3 Propuesta de solución**

Se propone desarrollar un modelo computacional basado en el aprendizaje profundo de redes neuronales como solución principal de este trabajo de tesis. El objetivo es utilizar algoritmos de búsqueda para analizar un conjunto de datos y seleccionar la función más adecuada de entre varias opciones, con el fin de predecir de manera precisa una o más variables relacionadas con la demanda de energía eléctrica. Para garantizar un entorno de trabajo adecuado a la temática de la tesis, se llevó a cabo un análisis y selección de tecnologías y metodologías apropiadas para implementar una solución de predicción efectiva y oportuna.

### **2.3.1 Lenguaje de programación Python**

Python es uno de los lenguajes de programación más populares para la ciencia de datos y de código abierto. Tiene bibliotecas útiles para la computación científica y de aprendizaje automático como son Numpy y SciPy que se construyen sobre

implementaciones de Fortran y C de capa inferior para operaciones factorizadas rápidas en matrices multidimensionales. Para las tareas de aprendizaje automático se utiliza la biblioteca Scikit-learn que es una de las más populares y accesibles de código abierto. (Raschka & Mirjalili, n.d.)

Scikit-learn: Es una biblioteca científica básica que permite entrenar y evaluar diferentes tipos de modelos y dar una solución. Ya que permite la abstracción de datos en menor tiempo a través de comandos muy legibles y específicos. Pues es uno de los kits científicos basados en SciPy para sus operaciones matemáticas.

Pandas: es la librería más utilizada y perfecta para el Data Wrangling. Permite realizar tareas de manipulación, agregación y visualización de datos de forma más sencilla.

NumPy: significa Python Numérico y es el paquete más fundamental sobre el que se construyen todas las bibliotecas científicas.

SciPy: es la biblioteca de software para operaciones científicas como Algebra Lineal, Estadística, Optimización, entre otras y está construida sobre la biblioteca NumPy.

Matplotlib: es la biblioteca más utilizada para crear visualizaciones de datos simples pero potentes en ejecución científica.

Plotly: es una biblioteca de visualización de datos interactiva basada en la web.

### **2.3.2 Keras**

Es un marco de aprendizaje profundo para Python que se utiliza para definir y entrenar en su mayoría cualquier tipo de modelo de aprendizaje profundo. Su desarrollo nació principalmente para investigadores que permitiera una experimentación más eficiente. Keras se distribuye bajo licencia permisiva MIT, lo que permite que pueda ser usado libremente en proyectos. Es compatible con cualquier versión de Python. Dentro de las principales características clave de Keras son:

Permite que se ejecute el mismo código en CPU o GPU, tiene una Api fácil de usar que permite ejecutar cualquier modelo de aprendizaje profundo. Permite soporte para redes convolucionales y recurrentes o combinación de ambas. Admite arquitecturas de red arbitrarias para modelos de múltiples entradas o múltiples salidas, uso compartido de capas y de modelos. Es una biblioteca a nivel de modelo que proporciona bloques de construcción de alto nivel para la construcción de modelos de aprendizaje profundo. Se basa en un tensor especializado y bien optimizado que permite manejar el problema de forma modular, por lo tanto, se puede conectar varios motores Back-end (software subyacente que utiliza algoritmos de aprendizaje profundo) como lo son TensorFlow, de Theano y Microsoft Cognitive.(Francois Chollet, 2017)

### **2.3.3 Tensorflow**

Es una plataforma principal para el aprendizaje profundo desarrollado por Google cualquier código de keras se puede ejecutar sin realizar varios cambios en él y se especializa en entrenar los modelos red neuronal profunda de manera muy eficiente. Es uno de las principales bibliotecas predeterminadas y más utilizadas para las necesidades de aprendizaje profundo, porque tiene características de adaptación, es escalable y eficiente para la ejecución de modelos. A través de TensorFlow, keras se puede ejecutar sin problemas en ambos modos de CPU o GPU, ya que envuelve un nivel más bajo de biblioteca para las operaciones de tensor llamada Eigen por lo que incluye una biblioteca de operaciones de aprendizaje profundo bien optimizadas denominadas NVIDIA CUDA Deep Neural Network.

### **2.3.4 Jupyter notebook**

Es una biblioteca de software de código abierto para poder ejecutar experimentos de aprendizaje profundo. son ampliamente utilizados en la ciencia de datos y modelos de aprendizaje. Un notebook de Jupyter es un archivo generado que tiene la capacidad de combinar y ejecutar código Python con capacidades de edición de texto enriquecidas que permiten realizar una división de experimentos extensos en

unos más pequeños para que se ejecuten de forma independiente, lo que hace que sea de forma interactiva el desarrollo bajo la secuencia de comandos de Python de forma independiente o desde la ejecución de un código desde dentro de un IDE.

### **2.3.5 Anaconda**

Es una distribución alternativa, libre y abierta de Python utilizada en ciencia de datos y aprendizaje automático recomendada para computación científica. Anaconda es un programa gratuito para el uso comercial. Agrupa todo lo esencial de Python para paquetes de datos, matemáticas e ingeniería en una multiplataforma fácil de usar.

Es una plataforma fácil para trabajar con el aprendizaje automático y la inteligencia artificial, pues permite con un conjunto de herramientas y paquetes el análisis de datos en el ámbito de la ciencia. El cual tiene los siguientes componentes principales:

Python: el lenguaje más común y utilizado para la ciencia de datos.

Conda: un administrador de paquetes y una herramienta de entorno virtual.

Navegador: una herramienta GUI que brinda la funcionalidad principal de Conda.

### **2.3.6 Excel**

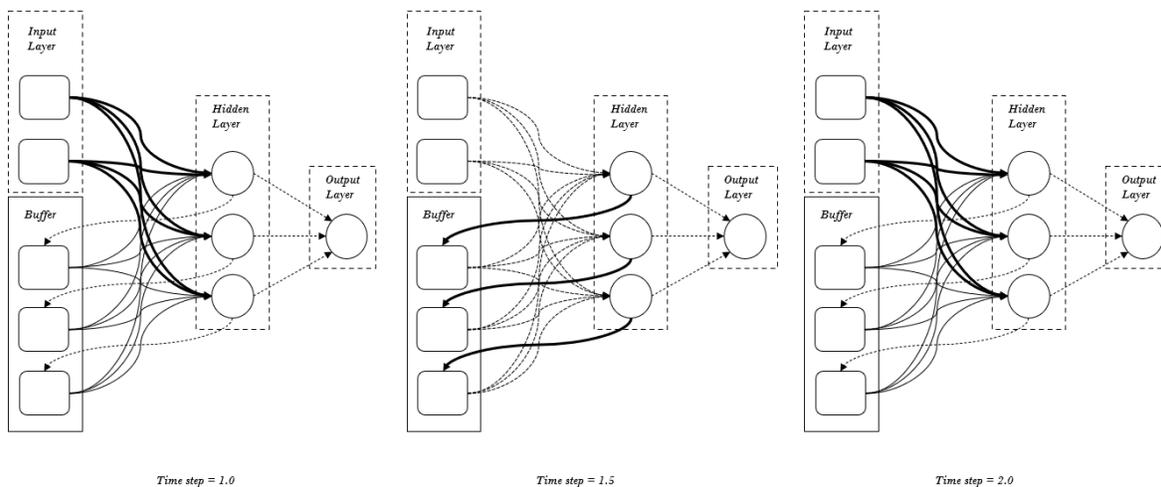
Es una herramienta muy eficaz para obtener información a partir de grandes cantidades de datos. Funciona para cálculos sencillos y para realizar el seguimiento de cualquier tipo de información. Su mayor potencial es la cuadrícula de celdas la cuales pueden contener números, textos o formulas. Los datos se describen en celdas y se agrupan en filas y columnas. Esto permite una manipulación y manejo sencillo para sumar datos, ordenarlos y filtrarlos para ponerlos en tablas.

### **2.3.7 Algoritmo basado en Redes Neuronales Recurrentes**

Una red neuronal recurrente RNN tiene una sola capa de neuronas ocultas, cuya salida se retroalimenta a esta etapa con la siguiente entrada. Esta retroalimentación o recurrencia dentro de la red le da una memoria que le permite procesar cada

entrada dentro del contexto de lo que ha procesado previamente. Las redes neuronales recurrentes son ideales para procesar datos secuenciales o de series temporales.

La Figura 17 ilustra la arquitectura de una RNN y muestra cómo influye la información a través de la red a medida que procesa una secuencia. La profundidad de una RNN nace del hecho de que el vector de memoria se propaga y evoluciona a través de cada entrada de la secuencia del proceso; como resultado, una red RNN se considera tan profunda como larga es la secuencia del procesamiento de los pesos.



*Figura 17 Arquitectura de una RNN*

## **CAPÍTULO III      METODOLOGÍA Y DESARROLLO**

### **3.1 Metodología**

En este capítulo, se presenta la metodología de desarrollo utilizada para construir el modelo computacional de predicción de la demanda de energía eléctrica basado en aprendizaje automático, con un enfoque específico en el aprendizaje profundo. El objetivo principal es garantizar la fiabilidad y precisión del modelo, por lo que se seguirán una serie de pasos lógicos y sistemáticos para generar predicciones confiables.

El proceso de desarrollo del modelo se divide en varias etapas, comenzando con la recopilación y preparación de los datos históricos de la demanda de energía eléctrica. Se realizará un exhaustivo análisis de la calidad de los datos, identificando y tratando posibles anomalías, errores o valores atípicos que puedan afectar la precisión del modelo.

A continuación, se procederá a la selección y diseño de la arquitectura de la red neuronal profunda. Esto implicará la elección de la estructura de capas y neuronas, así como la definición de las funciones de activación y los parámetros del modelo. Se explorarán diferentes configuraciones y se realizarán ajustes para encontrar la arquitectura óptima que se ajuste mejor a los datos y al problema de predicción de la demanda de energía eléctrica.

Una vez que la arquitectura de la red neuronal ha sido establecida, se procederá al entrenamiento del modelo utilizando los datos históricos. Se emplearán algoritmos de optimización y técnicas de retro propagación para ajustar los pesos y los sesgos de la red neuronal, de modo que pueda aprender los patrones y las relaciones en los datos de manera precisa.

Después del entrenamiento, se evaluará el rendimiento del modelo utilizando métricas adecuadas. Se realizarán pruebas adicionales utilizando conjuntos de datos de validación y se realizarán ajustes adicionales si es necesario para mejorar la precisión del modelo.

Una vez que el modelo ha sido validado y se ha obtenido un rendimiento satisfactorio, se procederá a la fase de implementación, donde el modelo se aplicará a nuevos datos de demanda de energía eléctrica para generar predicciones en tiempo real. Se desarrollará una interfaz de usuario intuitiva y amigable que permita la visualización y el seguimiento de las predicciones generadas por el modelo. En la Figura 18 se propone una visión general de la metodología de investigación del flujo de trabajo del sistema o modelo propuesto, así como el papel del algoritmo de aprendizaje como herramienta tecnológica computacional de predicción.

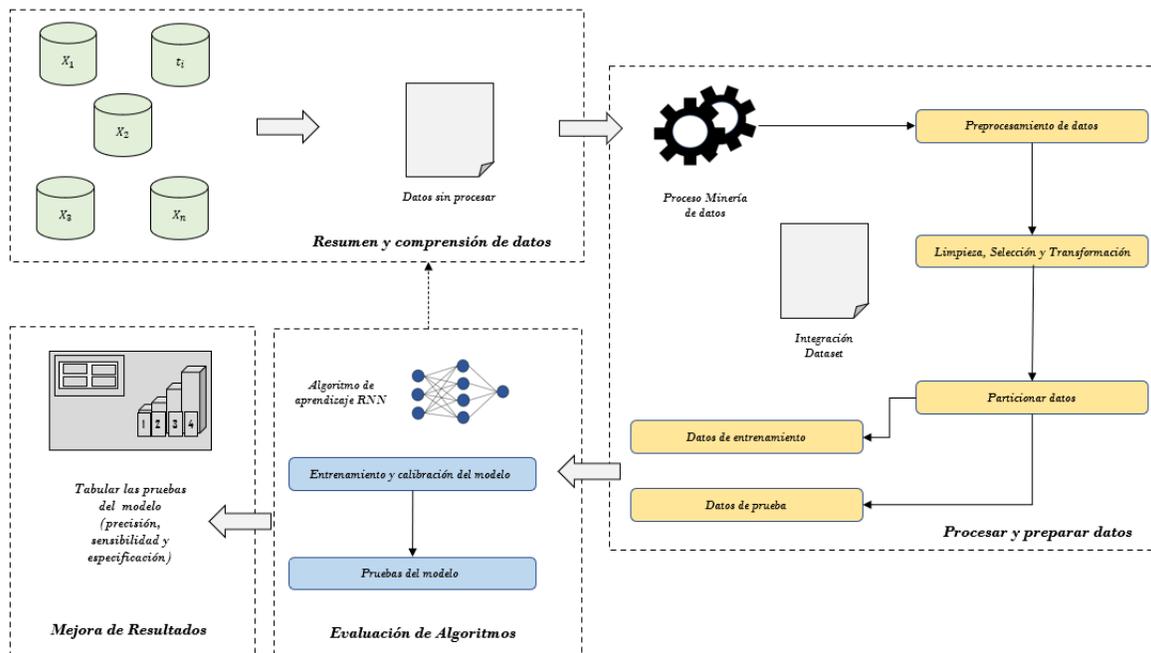


Figura 18 Metodología Red Neuronal

Se analizan las fases importantes para el desarrollo del modelo computacional las cuales se describen mediante la metodología CRISP-DM:

- Comprensión de datos.
- Procesar y preparar datos.
- Evaluación de algoritmos.
- Mejora de resultados.

### 3.1.1 Comprensión de datos:

Esta etapa se centra en la recopilación de información de diversas fuentes de datos que contienen las características o variables relevantes para ser procesadas en el modelo de predicción. Estos datos son fundamentales para alimentar el algoritmo de aprendizaje de nuestro estudio, por lo tanto, es crucial clasificarlos adecuadamente.

Con el fin de garantizar la confiabilidad de nuestro modelo de predicción, se analizaron los datos considerando su peso o importancia relativa. Se buscó identificar y seleccionar las fuentes de información que contengan datos relevantes para obtener un dataset que se ajuste a los objetivos del estudio. Estas fuentes de datos nos permiten clasificar las variables necesarias para la siguiente etapa, que es la construcción de nuestro conjunto de datos. Este conjunto de datos es el fundamento de nuestro estudio, ya que proporciona la información necesaria para la predicción en nuestro modelo.

Las fuentes de datos que se resumieron y comprendieron como análisis de la predicción de la demanda de energía eléctrica dentro de la región de la zona oriente se muestran en la Tabla 2 que se tomaron de un periodo de tiempo mensual comprendido entre los años de 2018 a 2022 para la recolección de datos de las variables de la demanda de energía.

*Tabla 2 Variables de predicción*

| Variables   |
|-------------|
| Municipio   |
| Tiempo      |
| Temperatura |
| Usuarios    |

---

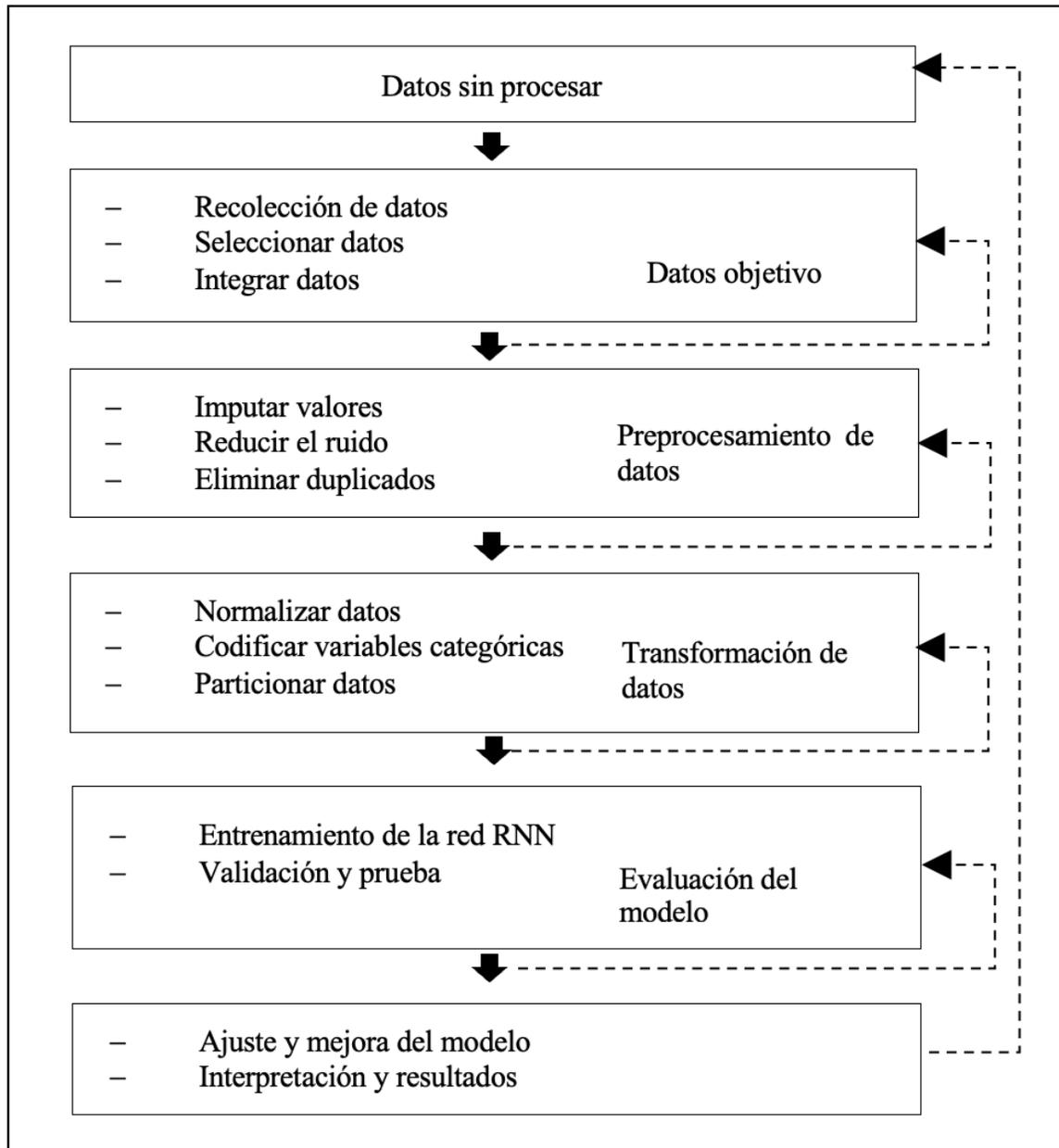
|                    |
|--------------------|
| Demanda            |
| Energía            |
| Factor de potencia |
| Voltaje            |

---

### **3.1.2 Procesar y preparar datos.**

Una vez clasificados nuestros datos para la implementación del modelo de aprendizaje necesitamos darles un tratamiento de extracción basado en un proceso de minería de datos adecuado, de tal forma que podamos crear nuestro dataset y colocar de forma matricial una estructura de  $m \times n$  donde  $m$  son las filas y  $n$  las columnas de nuestras clases e instancias en un contenedor de datos, el cual será ejecutado en el proceso de partición para el entrenamiento y prueba de nuestro modelo de aprendizaje.

La Figura 19 muestra la canalización del proceso de minería de datos mediante el diagrama de flujo propuesto en la investigación.



*Figura 19 Diagrama de flujo de minería de datos*

Respecto al proceso de minería de datos, el total de los datos deben ser preprocesados para ser transformados en datos que puedan ser utilizados en un algoritmo de aprendizaje automático por lo que esto implica, que el conjunto de datos debes ser recolectado, seleccionado e integrado con el objetivo de extraer conocimiento(Colombo-Mendoza et al., 2022). Este procesamiento de datos implica también la limpieza de datos como una manera de corregir o quitar anomalías en

los datos, visto de una forma simple esta limpieza implica imputar valores, estimar valores perdidos y eliminar valores duplicados. Esta parte es trascendental para nuestro modelo computacional pues integrara el conjunto de datos de manera estructurada y ordenada.

La transformación involucra la normalización de datos ajustándolos en valores medidos a diferentes escalas para que se distribuyan de manera normal de nuestro conjunto de datos, lo que permite que se puedan discretizar los datos de nuestras variables a contrapartes discretas de tal forma que puedan ser adecuados para la evaluación numérica en la implementación de los algoritmos de aprendizaje a los cuales estarán evaluados.

La creación de los atributos de los datos que están identificados en nuestras variables de estudio permite estructurar los datos en forma matricial a partir de columnas y filas, lo que genera que se organicen fácilmente en archivos planos como un CSV y sea más fácil seguir su patrón al implementar en el algoritmo del modelo de aprendizaje automático.(El-Amir & Hamdy, 2020)

### **3.1.3 Evaluación de Algoritmo.**

A partir de nuestro dataset y dividiendo nuestros datos en conjuntos de entrenamiento y prueba, podemos implementar el modelo de algoritmo que mejor se ajuste y se adapte a la estructura de nuestros datos. En nuestro estudio, contamos con una serie de secuencias de tiempo que representan datos históricos del consumo de demanda de energía eléctrica. En este contexto, el aprendizaje profundo, especialmente a través de las redes neuronales recurrentes (RNN) de regresión, se presenta como la herramienta más adecuada para el proceso de predicción.

Las RNN de regresión son capaces de adaptarse de manera efectiva al comportamiento de los datos secuenciales, lo que les permite ajustar los pesos de los datos de entrada de forma precisa. Esto se logra a través de la iteración continua y el aprendizaje de patrones a lo largo de la secuencia temporal. Al utilizar estas

redes neuronales en nuestro modelo, permitimos que la función de pérdida se ajuste correctamente a medida que se realizan las predicciones.

Al aplicar el aprendizaje profundo, específicamente las RNN de regresión, a nuestro dataset de secuencias de tiempo de consumo de energía eléctrica, logramos una mejor adaptación a los patrones y comportamientos de los datos, lo cual nos proporciona una mayor precisión en las predicciones.

Una forma general de visualizar una red neuronal como una máquina de procesamiento matemático definida por el recibimiento de varias entradas de algunos nodos o de una fuente externa como las variables ( $x$ ) de nuestro modelo y también con el cálculo de una salida ( $y$ ). Cada entrada tiene un peso asociado ( $w$ ) que se asigna en función de su importancia relativa con respecto a otras entradas. El nodo aplica una función  $f$  a la suma ponderada de sus entradas ver Formula 4.

Fórmula  $y = f ( w_1 \cdot x_1 + w_2 \cdot x_2 + \dots + w_n \cdot x_n + b )$  (4)

La Figura 20 muestra el perceptrón general utilizado y que dará el proceso de matemático de nuestro algoritmo de red neuronal.

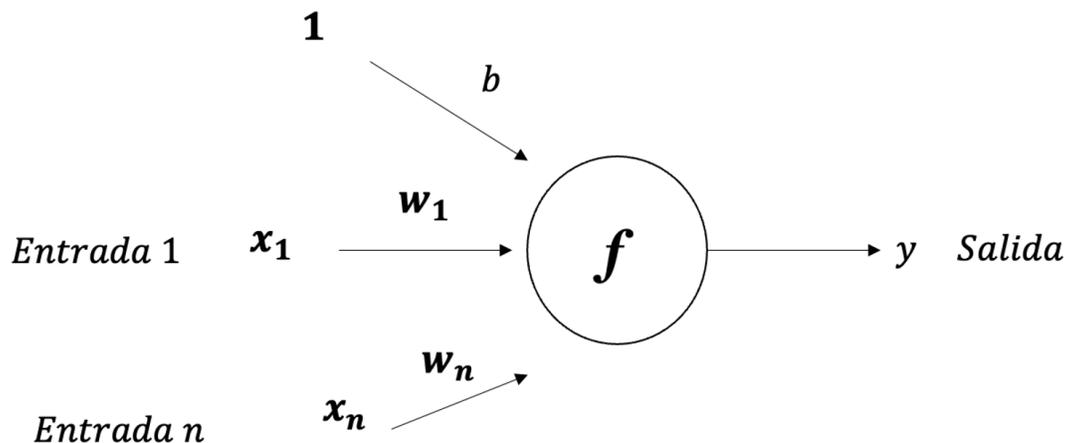
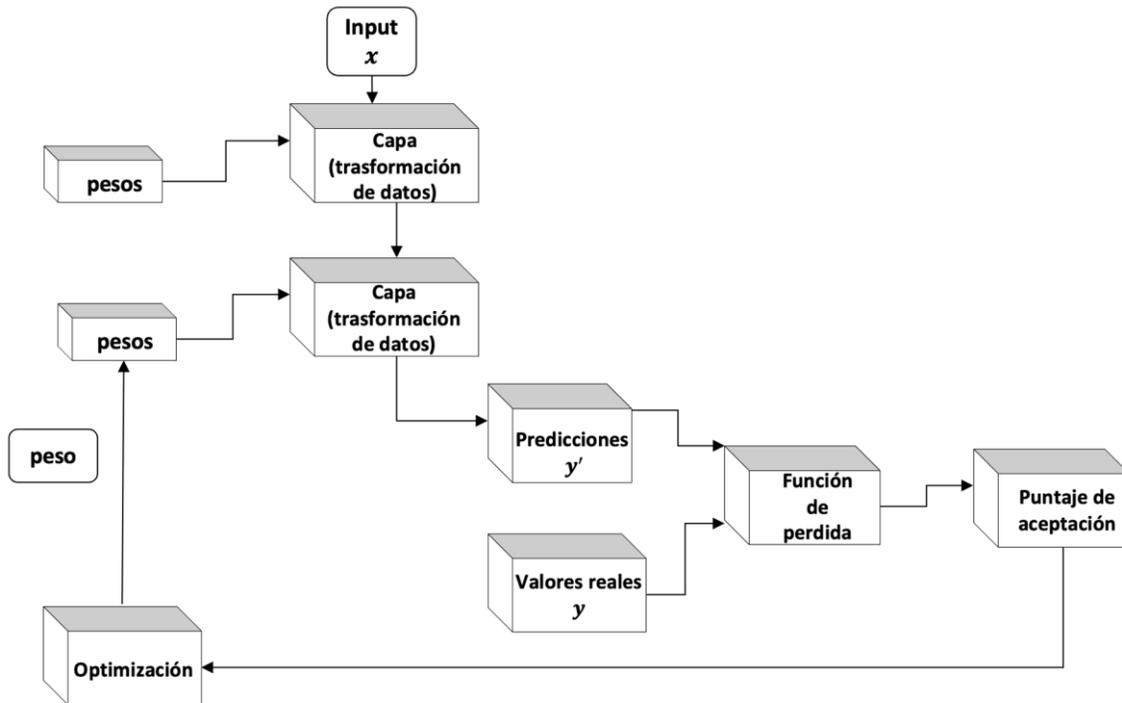


Figura 20 Perceptrón red neuronal

La Figura 21 muestra la Arquitectura de flujo de datos de entrada y salida como el proceso de tubería para el modelo de Red Neuronal Recurrente.



*Figura 21 Arquitectura del Modelo Neuronal Recurrente RNN*

La profundidad de un RNN surge del hecho de que el vector de memoria se propaga y evoluciona a través de cada entrada en la secuencia; como resultado, una red RNN se considera una secuencia de iteraciones para las estimaciones de los valores de entrada..

### **3.1.4 Mejora de resultados.**

En esta etapa, una vez implementado el algoritmo de aprendizaje utilizando los datos de las variables de la demanda eléctrica, se procede a evaluar el proceso de predicción del modelo. Para ello, se utiliza la función de pérdida, la cual proporciona una medida de evaluación del modelo y nos permite optimizar los pesos con el fin de obtener un mejor ajuste de los datos.

El análisis de métricas de las redes neuronales de regresión nos brinda información sobre qué tan bien nuestro modelo realiza las predicciones en relación a los datos reales. Estas métricas nos permiten evaluar la precisión y exactitud del modelo en la tarea de predicción. Algunas de las métricas comunes utilizadas incluyen el error cuadrático medio (MSE), el coeficiente de determinación ( $R^2$ ), y el error absoluto medio (MAE), entre otras.

A través del análisis de estas métricas, podemos obtener una medida objetiva de qué tan bien nuestro modelo se ajusta a los datos y cómo de precisa es su capacidad de predicción. Esto nos brinda información valiosa para realizar ajustes y mejoras en el modelo, en caso de ser necesario, y nos permite evaluar la calidad y confiabilidad de nuestras predicciones.

## **CAPÍTULO IV RESULTADOS**

En este capítulo, se describe el caso de estudio realizado para validar la solución propuesta del modelo computacional de predicción de la demanda de energía eléctrica. Se llevan a cabo pruebas de funcionamiento del modelo y se evalúa su desempeño.

Para realizar la validación, se selecciona un conjunto de datos representativo y se utiliza como entrada para el modelo de predicción. Se ejecutan las predicciones correspondientes y se comparan con los valores reales de la demanda de energía eléctrica.

Se analiza el desempeño del modelo mediante el uso de métricas de evaluación, como el error cuadrático medio (MSE), el coeficiente de determinación ( $R^2$ ) y el error absoluto medio (MAE). Estas métricas nos permiten medir la precisión y la calidad de las predicciones realizadas por el modelo.

Además, se realizan pruebas de funcionamiento del modelo en diferentes escenarios y se evalúa su capacidad para adaptarse a cambios en las condiciones de la demanda energética. Esto nos brinda información sobre la robustez y la efectividad del modelo en diferentes situaciones.

El análisis de los resultados obtenidos en estas pruebas y evaluaciones nos proporciona una visión clara del desempeño del modelo de predicción de la demanda de energía eléctrica, permitiendo validar la solución propuesta y determinar su utilidad y aplicabilidad en el contexto del estudio.

#### **4.1 CASO DE ESTUDIO.**

Dentro de la fase de recopilación de datos, se utilizaron tres fuentes de información para la integración del dataset. La primera fuente consistió en el análisis del sistema de información de la empresa suministradora de energía eléctrica del país, específicamente de la región oriente que abarca los municipios del estado de Puebla y Veracruz.

Se obtuvieron datos de la serie de tiempo que abarca desde el año 2018 hasta el año 2022, correspondientes a las mediciones de las características o variables eléctricas para el análisis. Estos datos fueron recopilados durante un periodo de 60 meses proporcionando un conjunto de datos históricos para el modelo predictivo.

La Figura 22 muestra la representación gráfica de los perfiles característicos de carga obtenidos a partir de los datos proporcionados por CFE Distribución del Sistema de Análisis de Perfiles. Estos perfiles representan las variaciones en la demanda de energía eléctrica a lo largo del tiempo, brindando información relevante para el análisis y la predicción.

La utilización de estas fuentes de datos permite contar con información precisa y detallada para la construcción del dataset, lo cual es fundamental para el desarrollo y entrenamiento del modelo de predicción de la demanda de energía eléctrica. (Comisión Federal de Electricidad [CFE],s.f.)

| Municipio            | Año  | Mes | Demanda (kW) | Energía (kWh) | Factor de Potencia | Voltaje    |
|----------------------|------|-----|--------------|---------------|--------------------|------------|
| MARTINEZ DE LA TORRE | 2019 | 1   | 6748.3103    | 5020742.8639  | 0.9926             | 13937.5245 |
| MARTINEZ DE LA TORRE | 2019 | 2   | 7442.8651    | 5001605.3417  | 0.9902             | 13799.6447 |
| MARTINEZ DE LA TORRE | 2019 | 3   | 7434.1315    | 5530993.8212  | 0.9882             | 13878.4223 |
| MARTINEZ DE LA TORRE | 2019 | 4   | 8828.8937    | 6356803.4874  | 0.9902             | 13769.7784 |
| MARTINEZ DE LA TORRE | 2019 | 5   | 11334.8775   | 8433148.8747  | 0.9688             | 13613.5517 |
| MARTINEZ DE LA TORRE | 2019 | 6   | 11012.2955   | 7928852.7657  | 0.9902             | 13660.5065 |
| MARTINEZ DE LA TORRE | 2019 | 7   | 10721.2571   | 7976615.2540  | 0.9920             | 13694.9407 |
| MARTINEZ DE LA TORRE | 2019 | 8   | 11356.9099   | 8449540.9918  | 0.9736             | 13642.8650 |
| MARTINEZ DE LA TORRE | 2019 | 9   | 10903.6188   | 7850605.5440  | 0.9902             | 13686.9169 |
| MARTINEZ DE LA TORRE | 2019 | 10  | 9220.8207    | 6860290.5666  | 0.9822             | 13808.8356 |
| MARTINEZ DE LA TORRE | 2019 | 11  | 7513.3352    | 5409601.3164  | 0.9902             | 13953.6118 |
| MARTINEZ DE LA TORRE | 2019 | 12  | 6989.9152    | 5200496.9052  | 0.9949             | 14015.6502 |



Figura 22 Variables de energía eléctrica

La segunda fuente de información está relacionada con el clima de la región oriente del país y en específico de la medición de la temperatura en grados Celsius (C°) que se recabo del sistema del centro meteorológico nacional (Comisión Nacional del Agua [CONAGUA], s.f) del periodo de tiempo del año 2018 -2022 ver Figura 23.



Figura 23 Variables climatológicas

Y la tercera fuente de información referente a la totalidad de usuarios por región como consumidores de energía eléctrica, que fue extraído del sistema web de la

empresa suministradora de energía eléctrica en el país (CFE, s.f.) del periodo de tiempo anual de 2018-2022 en una totalidad de análisis de 60 meses ver Figura 24.



Figura 24 Variable totalidad de usuarios por región

#### 4.1.1 Análisis de los datos.

En esta fase de preparación de datos, es fundamental garantizar la calidad y la utilidad de la información que se utilizará para el aprendizaje del algoritmo de aprendizaje automático. Por lo tanto, es importante realizar un examen exhaustivo y un preprocesamiento adecuado del conjunto de datos antes de alimentarlo al algoritmo de aprendizaje profundo.

En cuanto a la selección de las características informativas para la construcción del modelo, se analizaron cuidadosamente las variables más representativas obtenidas de nuestras fuentes de información. Estas variables son de vital importancia para el proceso de predicción de la demanda de energía eléctrica, ya que se basan en la serie de tiempo de los datos recopilados.

En la Tabla 3 se presentan las características seleccionadas, las cuales han demostrado ser relevantes y significativas para el modelo de predicción. Estas características proporcionan la base necesaria para capturar los patrones y las tendencias en la demanda de energía eléctrica, permitiendo así realizar predicciones más precisas y efectivas.

El análisis y la selección adecuada de estas características informativas son fundamentales para el éxito del modelo de predicción, ya que influyen directamente en la capacidad del algoritmo de aprendizaje profundo para aprender y generalizar a partir de los datos proporcionados.:

*Tabla 3 Tabla representativa de las variables relevantes*

| Variables   | Descripción   | Unidad de medida |
|-------------|---|------------------|
| Municipio   | Región del Área geográfica del consumo de demanda y energía eléctrica             | Unidad           |
| Tiempo      | Año y mes de la demanda consumida en la región                                    | T                |
| Temperatura | Grado o nivel térmico de la atmósfera.  | °C               |
| Usuarios    | Persona que hace uso de la energía eléctrica.                                     | Unidad           |
| Demanda     | Medida puntual del consumo eléctrico en un tiempo dado                            | KW               |
| Energía     | Trabajo diseñado por dispositivos eléctricos en un periodo de tiempo determinado. | KWH              |

|                    |   |    |
|--------------------|---|----|
| Factor de potencia | Indicador correcto del aprovechamiento de la energía.         | FP |
| Voltaje            | Cuantificación de la diferencia de potencial entre dos puntos | V  |

Se integra un total de 3240 instancias en nuestro dataset dividido en el análisis de por región de suministro de energía de 2 municipios del estado de Puebla y 4 municipios del estado de Veracruz ver Tabla 4.

*Tabla 4 Instancias por municipio*

| Municipio            | Estado   | Instancias |
|----------------------|----------|------------|
| Martínez de la Torre | Veracruz | 540        |
| Tlapacoyan           | Veracruz | 540        |
| Misantla             | Veracruz | 540        |
| Zacapoaxtla          | Puebla   | 540        |
| Teziutlán            | Puebla   | 540        |
| Perote               | Veracruz | 540        |

Dentro de las variables analizadas tenemos las de factor causal, las cuales son las que influyen en los datos de una manera conocida y pueden ayudar a la predicción. Dentro de los datos analizados la temperatura y el número de usuarios son factores causales del proceso del aumento o disminución de la demanda y el consumo de

energía eléctrica. Porque la demanda y la energía dependen o están sumamente relacionadas con las variables causales para su predicción.

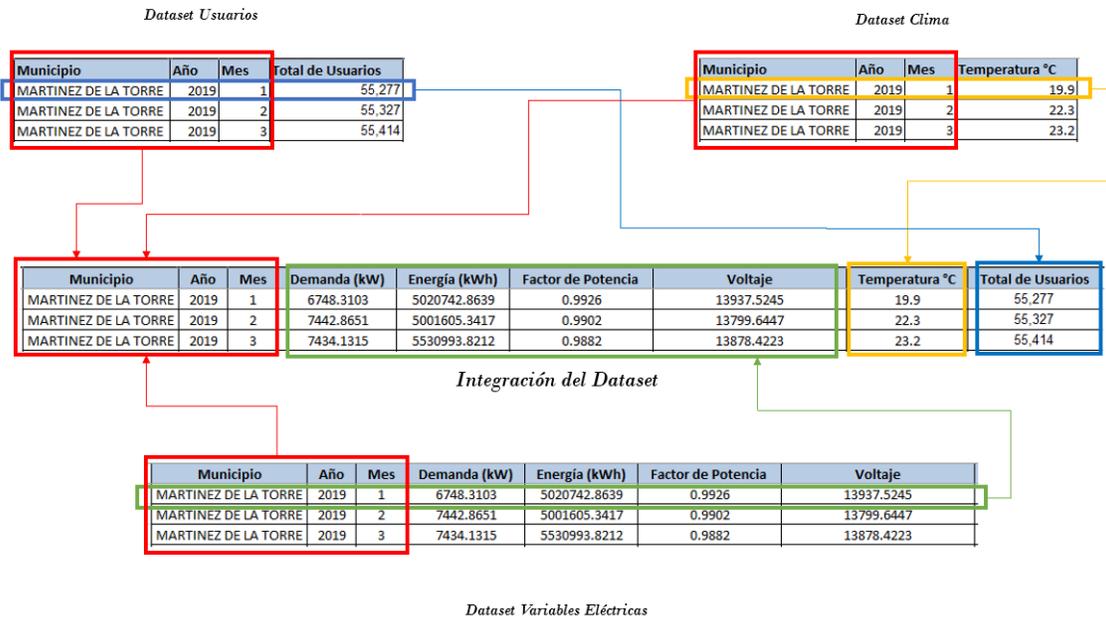
Dentro de la imputación de los valores faltantes del conjunto de datos y de la eliminación de columnas de nuestras fuentes de información al ser integradas en nuestro dataset, no es factible eliminarlas en su totalidad al tener valores nulos o faltantes por lo que podemos perder demasiados datos valiosos. Por lo que se realizó una técnica de interpolación para estimar valores faltantes, con la técnica de la imputación media, se reemplazó el valor faltante por el valor medio de la columna de la variable de datos y que para nuestro análisis de información constaron de 186 valores nulos. Los valores faltantes se integraron para en un total de 3240 datos de nuestro dataset.

La estadística en la Tabla 5 se describe las instancias y valores nulos por demanda de energía eléctrica, temperatura y usuarios.

*Tabla 5 Estadística del conjunto de datos*

| Dataset                   | Instancias | Valores nulos |
|---------------------------|------------|---------------|
| Perfil de carga eléctrica | 1080       | 42            |
| Temperatura               | 360        | 144           |
| Usuarios                  | 360        | 0             |

Es importante mencionar que no se tuvieron valores inconsistentes dentro de las fuentes de información principalmente del dataset de la medición de las variables eléctricas, por lo que la integración de nuestro dataset queda como se muestra en la Figura 25.



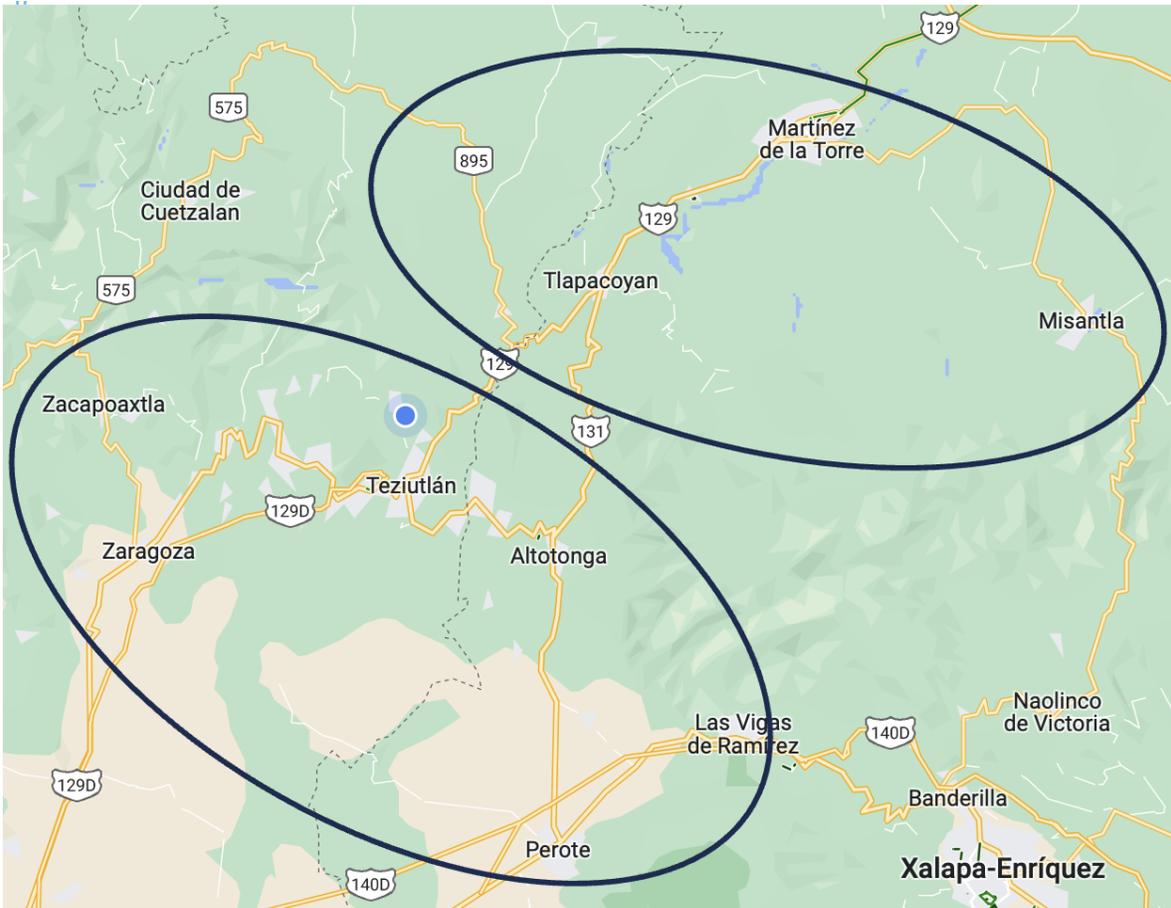
*Figura 25 Integración del dataset*

Para el mejor funcionamiento del modelo de aprendizaje automático se debe transformar los datos alfabéticos en numéricos, de tal forma, que puedan estar normalizados en nuestro dataset de nuestro conjunto de datos. Es por ello por lo que basado en las regiones de temperatura de las estaciones meteorológicas integradas en la Tabla 4 y de los bancos de suministro de energía por región, se procedió a clasificar los datos alfabéticos de la variable municipio a áreas de distribución de energía de forma numérica constituidas en cada una de las subestaciones que integran los bancos para el suministro de energía eléctrica. Y respecto a la región se clasificó de acuerdo con las bases de las estaciones meteorológicas de la medición de la temperatura por altura a nivel del mar dando valores numéricos a cada región Figura 26.

*Tabla 6 Tabla estaciones meteorológicas*

| Municipio            | Estación | Localización   |
|----------------------|----------|--|
| Martínez de la Torre | 30102    | Latitud: 20°04'45" Longitud:097°03'50" Alt: 89.00 MSNM |

|             |                     |  |
|-------------|---------------------|--|
| Tlapacoyan  | 21201<br>margaritas | Latitud: 19°59'14" Longitud:097°17'14" Alt: 500.00<br>MSNM |
| Misantla    | 30108               | Latitud: 19°55'45" Longitud:096°51'20" Alt: 310.0<br>MSNM  |
| Perote      | 30128               | Latitud: 19°34'51" Longitud:097°14'52" Alt: 2392<br>MSNM   |
| Teziutlán   | 21091               | Latitud: 19°49'49" Longitud:097°21'00" Alt: 1818<br>MSNM   |
| Zacapoaxtla | 21103               | Latitud: 19°52'18" Longitud:097°35'18" Alt: 1828<br>MSNM   |



*Figura 26 Temperaturas por región*

Quedando la integración de 2 columnas como se muestra en la Tabla 5, donde se sustituyen a la variable cualitativa de municipio por la variable cuantitativa e integrándola en nuestro dataset y la región como variable cuantitativa definida por la altitud de estaciones meteorológicas para el análisis del modelo computacional.

*Tabla 7 Transformación de datos.*

| Variable categórica | Variable cuantitativa | Región |
|---------------------|-----------------------|--------|
| Martínez            | 1                     | 1      |
| Tlapacoyan          | 5                     | 1      |
| Misantla            | 2                     | 1      |

|             |   |   |
|-------------|---|---|
| Zacapoaxtla | 6 | 2 |
| Perote      | 3 | 2 |
| Teziutlán   | 4 | 2 |

Las variables cuantitativas de la predicción de la demanda de energía eléctrica quedan integradas en nuestro dataset Figura 27.

| Municipio            | Area | Region | Año  | Mes | Demanda (kW) | Energía (kWh) | Factor de Potencia | Voltaje    | Temperatura °C | Total de Usuarios |
|----------------------|------|--------|------|-----|--------------|---------------|--------------------|------------|----------------|-------------------|
| MARTINEZ DE LA TORRE | 1    | 1      | 2019 | 1   | 6748.3103    | 5020742.8639  | 0.9926             | 13937.5245 | 19.9           | 55,277            |
| TEZIUTLAN            | 4    | 3      | 2019 | 1   | 2827.2716    | 2068451.7165  | 0.9744             | 13768.6355 | 13.9           | 54,395            |
| ZACAPOAXTLA          | 6    | 2      | 2019 | 1   | 4928.0050    | 3666435.7120  | 0.9906             | 14237.3338 | 13.3           | 26,025            |
| PEROTE               | 3    | 3      | 2019 | 1   | 10140.8709   | 7544807.9688  | 0.9987             | 14136.2019 | 13.9           | 37,238            |
| MISANTLA             | 2    | 4      | 2019 | 1   | 5755.0380    | 4281748.2403  | 0.9635             | 13892.1672 | 17.8           | 27,882            |
| TLAPACOYAN           | 5    | 1      | 2019 | 1   | 4550.0387    | 3385228.7756  | 0.9799             | 13947.0178 | 19.9           | 28,607            |

  
*Integración de Área y Región a  
Dataset del modelo de predicción.*

*Figura 27 Integración de Área y Región a dataset*

El dataset queda estructurado con el conjunto de características o variables necesarias para la predicción, se omite la variable del Factor de Potencia porque es un valor consecuente de la relación de la demanda con el voltaje lo que permitiría sobre ajustar el modelo. La partición de los datos para el entrenamiento y prueba quedan equilibrados para la implementación del algoritmo que permita predecir correctamente el modelo computacional de la demanda de energía eléctrica.

La integración de los datos se efectúa como se muestra en la siguiente Figura 28.

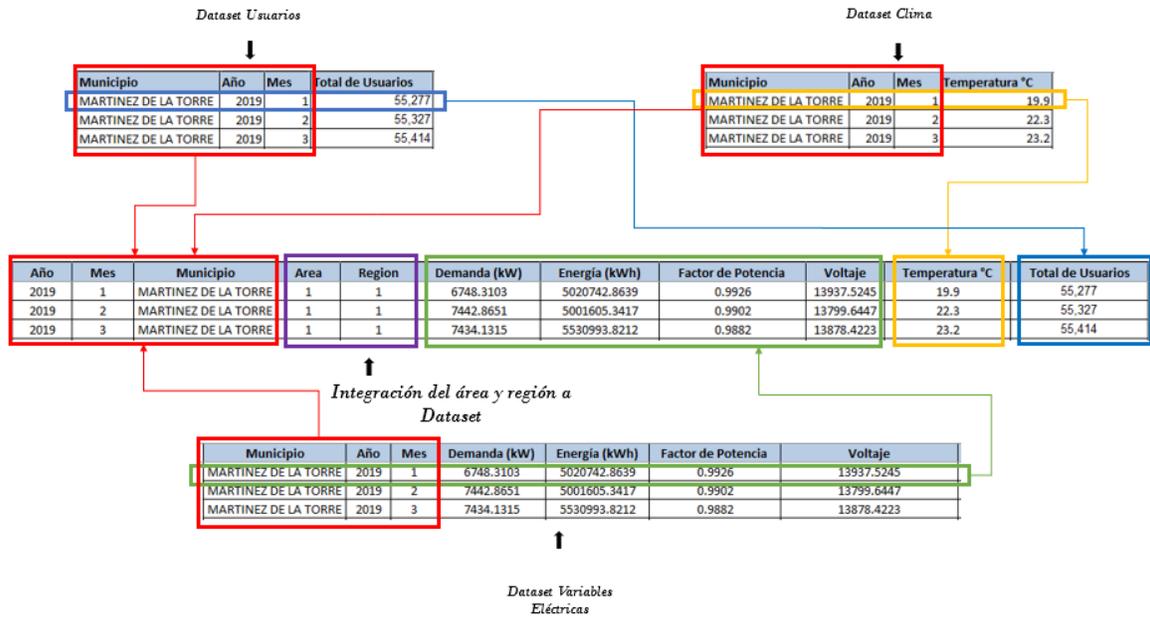


Figura 28 Dataset para entrenamiento y prueba

El dataset para el entrenamiento de calibración y prueba queda constituido por un total de 3240 instancias de valores normalizados quedando como se muestra en la Tabla 8.

Tabla 8 Dataset de entrenamiento y prueba

| Año  | Mes | Municipio | Región | Energía (kWh) | Voltaje    | Temperatura °C | Total de Usuarios | Demanda (kW) |
|------|-----|-----------|--------|---------------|------------|----------------|-------------------|--------------|
| 2018 | 1   | 1         | 1      | 8,745,887     | 13,998.18  | 15.94          | 53,978            | 6786.12      |
| 2018 | 2   | 1         | 1      | 8,439,408     | 131,134.02 | 21.84          | 54,002            | 7148.40      |
| 2018 | 3   | 1         | 1      | 9,324,468     | 131,127.07 | 22.71          | 54,097            | 8057.07      |
| 2018 | 4   | 1         | 1      | 10,911,986    | 127,657.47 | 23.68          | 54,248            | 9523.01      |
| 2018 | 5   | 1         | 1      | 11,758,135    | 127,627.13 | 27.79          | 54,370            | 10861.14     |
| 2018 | 6   | 1         | 1      | 12,544,207    | 13,573.39  | 27.43          | 54,533            | 10556.57     |
| 2018 | 7   | 1         | 1      | 13,836,837    | 13,695.66  | 26.87          | 54,699            | 10922.79     |
| 2018 | 8   | 1         | 1      | 13,346,195    | 13,727.40  | 26.68          | 54,835            | 10840.84     |
| 2018 | 9   | 1         | 1      | 14,504,320    | 13,756.50  | 26.7           | 54,921            | 10216.88     |
| 2018 | 10  | 1         | 1      | 13,622,652    | 13,822.22  | 23.63          | 54,988            | 9277.01      |
| 2018 | 11  | 1         | 1      | 13,008,496    | 13,929.57  | 20.67          | 55,058            | 7913.86      |
| 2018 | 12  | 1         | 1      | 9,188,171     | 13,986.73  | 21.55          | 55,169            | 7363.48      |
| 2019 | 1   | 1         | 1      | 9,304,816     | 13,937.52  | 20.06          | 55,277            | 6748.31      |
| 2019 | 2   | 1         | 1      | 9,088,012     | 13,799.64  | 21.63          | 55,327            | 7442.87      |
| 2019 | 3   | 1         | 1      | 9,742,171     | 13,878.42  | 21.81          | 55,414            | 7434.13      |
| 2019 | 4   | 1         | 1      | 9,524,726     | 13,769.78  | 26.05          | 55,533            | 8828.89      |
| 2019 | 5   | 1         | 1      | 13,292,643    | 13,613.55  | 29.73          | 55,664            | 11334.88     |
| 2019 | 6   | 1         | 1      | 13,462,823    | 13,660.51  | 29.47          | 55,755            | 11012.30     |
| 2019 | 7   | 1         | 1      | 14,446,239    | 13,694.94  | 29.08          | 55,801            | 10721.26     |
| 2019 | 8   | 1         | 1      | 13,641,177    | 13,642.87  | 27.22          | 55,947            | 11356.91     |
| 2019 | 9   | 1         | 1      | 14,500,586    | 13,686.92  | 26.07          | 56,075            | 10903.62     |
| 2019 | 10  | 1         | 1      | 13,369,787    | 13,808.84  | 24.53          | 56,189            | 9220.82      |
| 2019 | 11  | 1         | 1      | 12,668,223    | 13,953.61  | 21.62          | 56,209            | 7513.34      |
| 2019 | 12  | 1         | 1      | 9,743,102     | 14,015.65  | 19.77          | 56,217            | 6989.92      |
| 2020 | 1   | 1         | 1      | 9,994,373     | 14,002.84  | 18.86          | 56,251            | 6841.01      |
| 2020 | 2   | 1         | 1      | 9,320,639     | 13,960.40  | 20.56          | 56,271            | 7026.02      |
| 2020 | 3   | 1         | 1      | 9,578,168     | 13,820.20  | 22.64          | 56,338            | 8545.18      |
| 2020 | 4   | 1         | 1      | 10,212,559    | 13,713.15  | 25.78          | 56,472            | 9818.12      |

|      |    |   |   |            |           |       |        |          |
|------|----|---|---|------------|-----------|-------|--------|----------|
| 2020 | 5  | 1 | 1 | 12,599,603 | 13,741.53 | 27.94 | 56,608 | 9733.89  |
| 2020 | 6  | 1 | 1 | 12,262,705 | 13,720.02 | 28.08 | 56,796 | 10273.43 |
| 2020 | 7  | 1 | 1 | 14,517,849 | 13,643.47 | 27.31 | 56,967 | 11137.09 |
| 2020 | 8  | 1 | 1 | 13,122,370 | 13,726.55 | 27.27 | 57,102 | 10106.48 |
| 2020 | 9  | 1 | 1 | 14,080,047 | 13,768.15 | 26.02 | 57,278 | 9919.03  |
| 2020 | 10 | 1 | 1 | 13,438,679 | 13,782.18 | 24.61 | 57,330 | 9557.54  |
| 2020 | 11 | 1 | 1 | 13,474,780 | 13,882.39 | 21.71 | 57,407 | 8226.90  |
| 2020 | 12 | 1 | 1 | 10,126,032 | 14,021.62 | 19.6  | 57,460 | 6688.98  |
| 2021 | 1  | 1 | 1 | 9,481,756  | 14,020.27 | 18.74 | 57,569 | 6628.69  |
| 2021 | 2  | 1 | 1 | 9,105,851  | 13,911.36 | 20.46 | 57,635 | 7146.18  |
| 2021 | 3  | 1 | 1 | 9,968,332  | 13,851.24 | 22.72 | 57,801 | 8326.30  |
| 2021 | 4  | 1 | 1 | 9,879,424  | 13,788.22 | 25.75 | 57,985 | 9314.35  |
| 2021 | 5  | 1 | 1 | 12,083,741 | 13,732.75 | 27.77 | 58,043 | 10620.44 |
| 2021 | 6  | 1 | 1 | 12,343,957 | 13,137.99 | 27.95 | 58,166 | 10458.48 |
| 2021 | 7  | 1 | 1 | 14,062,255 | 13,678.01 | 27.15 | 58,272 | 11063.00 |
| 2021 | 8  | 1 | 1 | 13,359,329 | 13,800.00 | 27.27 | 58,329 | 11074.40 |
| 2021 | 9  | 1 | 1 | 14,011,951 | 13,720.60 | 26.01 | 58,431 | 10612.22 |
| 2021 | 10 | 1 | 1 | 12,901,240 | 13,745.49 | 24.62 | 58,485 | 10202.30 |
| 2021 | 11 | 1 | 1 | 13,458,590 | 13,929.87 | 21.72 | 58,635 | 8008.30  |
| 2021 | 12 | 1 | 1 | 9,798,861  | 13,957.26 | 19.58 | 58,712 | 7892.41  |
| 2022 | 1  | 1 | 1 | 9,962,471  | 14,032.09 | 18.73 | 58,928 | 6926.45  |
| 2022 | 2  | 1 | 1 | 8,680,164  | 14,025.39 | 20.45 | 59,071 | 6978.53  |
| 2022 | 3  | 1 | 1 | 8,964,053  | 13,898.34 | 22.72 | 59,224 | 7922.68  |
| 2022 | 4  | 1 | 1 | 9,643,493  | 13,771.83 | 25.75 | 59,264 | 10130.70 |
| 2022 | 5  | 1 | 1 | 13,038,789 | 13,648.44 | 27.75 | 59,389 | 11755.36 |
| 2022 | 6  | 1 | 1 | 13,938,119 | 13,775.05 | 27.94 | 59,515 | 10482.09 |
| 2022 | 7  | 1 | 1 | 15,229,627 | 13,766.25 | 27.13 | 59,627 | 10769.82 |
| 2022 | 8  | 1 | 1 | 13,408,243 | 13,740.22 | 27.27 | 59,769 | 10825.58 |
| 2022 | 9  | 1 | 1 | 14,358,671 | 13,850.35 | 26.01 | 59,867 | 9432.64  |
| 2022 | 10 | 1 | 1 | 12,504,159 | 13,952.40 | 24.62 | 59,872 | 8127.40  |
| 2022 | 11 | 1 | 1 | 13,047,175 | 13,952.42 | 21.72 | 59,967 | 7906.90  |
| 2022 | 12 | 1 | 1 | 9,802,198  | 13,952.42 | 19.58 | 59,986 | 7882.60  |
| 2018 | 1  | 4 | 3 | 14,115,984 | 13,812.65 | 8.87  | 52,425 | 2884.76  |
| 2018 | 2  | 4 | 3 | 14,730,011 | 13,788.07 | 13.46 | 52,538 | 2868.25  |
| 2018 | 3  | 4 | 3 | 18,546,694 | 13,758.24 | 14.97 | 52,695 | 2938.97  |
| 2018 | 4  | 4 | 3 | 25,468,743 | 13,623.50 | 15.58 | 52,880 | 2768.38  |
| 2018 | 5  | 4 | 3 | 16,992,834 | 13,581.06 | 15.92 | 53,045 | 2874.11  |
| 2018 | 6  | 4 | 3 | 16,997,453 | 13,717.42 | 16.28 | 53,177 | 2810.35  |
| 2018 | 7  | 4 | 3 | 15,788,012 | 13,751.28 | 14.98 | 53,258 | 2850.38  |
| 2018 | 8  | 4 | 3 | 11,258,080 | 13,728.79 | 14.82 | 53,467 | 2782.36  |
| 2018 | 9  | 4 | 3 | 9,172,355  | 13,786.37 | 15.62 | 53,692 | 2852.24  |
| 2018 | 10 | 4 | 3 | 9,671,035  | 13,835.72 | 14.11 | 53,851 | 2900.00  |
| 2018 | 11 | 4 | 3 | 8,875,092  | 13,853.40 | 13.13 | 54,023 | 2915.03  |
| 2018 | 12 | 4 | 3 | 10,473,056 | 13,826.87 | 12.58 | 54,218 | 2897.43  |
| 2019 | 1  | 4 | 3 | 14,322,641 | 13,768.64 | 12.26 | 54,395 | 2827.27  |
| 2019 | 2  | 4 | 3 | 16,350,726 | 13,757.25 | 15.7  | 54,495 | 2835.33  |
| 2019 | 3  | 4 | 3 | 19,339,080 | 13,788.22 | 14.18 | 54,604 | 3054.59  |
| 2019 | 4  | 4 | 3 | 20,199,207 | 13,732.75 | 17.42 | 54,726 | 2903.04  |
| 2019 | 5  | 4 | 3 | 20,888,665 | 13,137.99 | 18.98 | 54,863 | 3054.10  |
| 2019 | 6  | 4 | 3 | 17,263,708 | 13,678.01 | 16.48 | 55,092 | 2766.25  |
| 2019 | 7  | 4 | 3 | 16,013,634 | 13,800.00 | 15.31 | 55,163 | 2921.21  |
| 2019 | 8  | 4 | 3 | 15,177,370 | 13,720.60 | 13.45 | 55,257 | 2637.24  |
| 2019 | 9  | 4 | 3 | 15,155,280 | 13,745.49 | 13.69 | 55,378 | 2749.55  |
| 2019 | 10 | 4 | 3 | 12,450,653 | 13,929.87 | 12.19 | 55,479 | 2814.56  |
| 2019 | 11 | 4 | 3 | 11,576,026 | 13,957.26 | 11.28 | 55,546 | 2814.58  |
| 2019 | 12 | 4 | 3 | 15,531,433 | 14,032.09 | 11.34 | 55,683 | 2753.52  |
| 2020 | 1  | 4 | 3 | 14,880,760 | 14,025.39 | 10.39 | 55,798 | 2809.43  |
| 2020 | 2  | 4 | 3 | 14,784,944 | 13,898.34 | 12.5  | 55,924 | 2799.84  |
| 2020 | 3  | 4 | 3 | 21,271,740 | 13,771.83 | 13.38 | 56,069 | 2780.36  |
| 2020 | 4  | 4 | 3 | 21,702,754 | 13,648.44 | 15.62 | 56,119 | 2694.80  |
| 2020 | 5  | 4 | 3 | 21,848,582 | 13,775.05 | 16.16 | 56,229 | 2682.32  |
| 2020 | 6  | 4 | 3 | 13,020,431 | 13,766.25 | 15.02 | 56,430 | 2715.37  |
| 2020 | 7  | 4 | 3 | 13,340,124 | 13,740.22 | 13.89 | 56,589 | 2766.43  |
| 2020 | 8  | 4 | 3 | 17,346,696 | 13,850.35 | 13.32 | 56,697 | 2773.18  |
| 2020 | 9  | 4 | 3 | 11,489,595 | 13,952.40 | 13.51 | 56,835 | 2862.11  |
| 2020 | 10 | 4 | 3 | 13,864,490 | 13,952.42 | 12.01 | 56,929 | 2845.79  |
| 2020 | 11 | 4 | 3 | 13,206,439 | 13,952.42 | 11.1  | 57,024 | 2895.17  |
| 2020 | 12 | 4 | 3 | 15,650,977 | 13,752.56 | 11.22 | 57,143 | 2938.89  |
| 2021 | 1  | 4 | 3 | 17,440,336 | 13,753.00 | 10.21 | 57,227 | 2924.14  |
| 2021 | 2  | 4 | 3 | 17,627,758 | 13,756.08 | 12.2  | 57,390 | 2826.91  |
| 2021 | 3  | 4 | 3 | 18,913,086 | 13,753.00 | 13.31 | 57,418 | 2968.70  |
| 2021 | 4  | 4 | 3 | 25,829,710 | 13,391.42 | 15.45 | 57,506 | 2567.49  |

|      |    |   |   |            |           |       |        |         |
|------|----|---|---|------------|-----------|-------|--------|---------|
| 2021 | 5  | 4 | 3 | 24,285,022 | 13,719.02 | 15.89 | 57,650 | 2803.40 |
| 2021 | 6  | 4 | 3 | 16,232,084 | 13,703.52 | 14.88 | 57,772 | 2842.31 |
| 2021 | 7  | 4 | 3 | 16,112,179 | 13,719.26 | 13.76 | 57,899 | 2822.89 |
| 2021 | 8  | 4 | 3 | 17,007,377 | 13,732.62 | 13.31 | 58,058 | 2691.62 |
| 2021 | 9  | 4 | 3 | 11,730,473 | 13,710.90 | 13.49 | 58,231 | 2883.21 |
| 2021 | 10 | 4 | 3 | 13,394,167 | 13,712.61 | 11.99 | 58,347 | 2923.46 |
| 2021 | 11 | 4 | 3 | 12,028,783 | 13,745.82 | 11.09 | 58,546 | 2957.42 |
| 2021 | 12 | 4 | 3 | 13,935,571 | 13,729.39 | 11.21 | 58,632 | 2928.36 |
| 2022 | 1  | 4 | 3 | 16,319,126 | 13,703.58 | 10.19 | 58,774 | 2978.18 |
| 2022 | 2  | 4 | 3 | 18,375,472 | 13,740.62 | 12.17 | 58,944 | 3010.93 |
| 2022 | 3  | 4 | 3 | 23,055,203 | 13,719.93 | 13.3  | 59,087 | 2952.24 |
| 2022 | 4  | 4 | 3 | 21,834,188 | 13,721.40 | 15.44 | 59,231 | 2908.18 |
| 2022 | 5  | 4 | 3 | 27,412,665 | 13,692.20 | 15.87 | 59,380 | 2956.61 |
| 2022 | 6  | 4 | 3 | 22,016,166 | 13,721.91 | 14.86 | 59,509 | 2917.45 |
| 2022 | 7  | 4 | 3 | 19,845,715 | 13,745.66 | 13.75 | 59,641 | 2890.99 |
| 2022 | 8  | 4 | 3 | 20,530,718 | 13,611.60 | 13.31 | 59,825 | 3027.40 |
| 2022 | 9  | 4 | 3 | 12,975,820 | 13,736.69 | 13.49 | 59,978 | 2914.10 |
| 2022 | 10 | 4 | 3 | 12,361,167 | 13,747.97 | 11.99 | 60,185 | 3016.21 |
| 2022 | 11 | 4 | 3 | 14,299,359 | 13,758.10 | 11.08 | 60,303 | 2992.96 |
| 2022 | 12 | 4 | 3 | 13,036,720 | 13,793.44 | 11.21 | 60,427 | 2968.94 |
| 2018 | 1  | 6 | 2 | 2,164,327  | 14,260.79 | 8.87  | 25,533 | 5346.20 |
| 2018 | 2  | 6 | 2 | 2,138,334  | 14,259.05 | 13.46 | 25,533 | 5291.07 |
| 2018 | 3  | 6 | 2 | 2,119,389  | 14,220.30 | 14.97 | 25,581 | 5146.08 |
| 2018 | 4  | 6 | 2 | 2,104,068  | 14,195.94 | 15.58 | 25,623 | 4859.50 |
| 2018 | 5  | 6 | 2 | 2,202,435  | 14,211.03 | 15.92 | 25,700 | 4800.66 |
| 2018 | 6  | 6 | 2 | 2,050,949  | 14,248.84 | 16.28 | 25,693 | 4658.77 |
| 2018 | 7  | 6 | 2 | 2,074,169  | 14,260.24 | 14.98 | 25,749 | 4637.46 |
| 2018 | 8  | 6 | 2 | 2,092,069  | 14,104.78 | 14.82 | 25,770 | 4738.19 |
| 2018 | 9  | 6 | 2 | 2,132,479  | 14,180.33 | 15.62 | 25,820 | 5211.61 |
| 2018 | 10 | 6 | 2 | 2,242,664  | 14,245.74 | 14.11 | 25,852 | 4735.51 |
| 2018 | 11 | 6 | 2 | 2,259,670  | 14,269.29 | 13.13 | 25,886 | 4954.51 |
| 2018 | 12 | 6 | 2 | 2,122,033  | 14,262.61 | 12.58 | 25,907 | 5339.91 |
| 2019 | 1  | 6 | 2 | 2,228,696  | 14,237.33 | 12.26 | 26,025 | 4928.01 |
| 2019 | 2  | 6 | 2 | 2,508,742  | 14,229.93 | 15.7  | 26,073 | 4840.10 |
| 2019 | 3  | 6 | 2 | 2,328,270  | 14,174.53 | 14.18 | 26,135 | 4847.06 |
| 2019 | 4  | 6 | 2 | 2,246,131  | 14,157.93 | 17.42 | 26,353 | 4690.51 |
| 2019 | 5  | 6 | 2 | 2,593,008  | 14,158.01 | 18.98 | 26,396 | 4734.62 |
| 2019 | 6  | 6 | 2 | 2,665,496  | 14,200.26 | 16.48 | 26,441 | 4609.58 |
| 2019 | 7  | 6 | 2 | 2,233,531  | 14,218.38 | 15.31 | 26,545 | 4578.56 |
| 2019 | 8  | 6 | 2 | 2,154,675  | 14,221.75 | 13.45 | 26,623 | 4570.10 |
| 2019 | 9  | 6 | 2 | 2,309,460  | 13,936.24 | 13.69 | 26,679 | 6660.93 |
| 2019 | 10 | 6 | 2 | 2,255,409  | 14,191.23 | 12.19 | 26,725 | 4791.26 |
| 2019 | 11 | 6 | 2 | 2,379,394  | 14,218.96 | 11.28 | 26,758 | 4982.28 |
| 2019 | 12 | 6 | 2 | 2,323,973  | 14,237.62 | 11.34 | 26,804 | 5042.52 |
| 2020 | 1  | 6 | 2 | 2,329,603  | 14,236.55 | 10.39 | 26,810 | 5025.98 |
| 2020 | 2  | 6 | 2 | 2,617,038  | 14,216.92 | 12.5  | 26,859 | 4933.89 |
| 2020 | 3  | 6 | 2 | 2,220,804  | 14,171.39 | 13.38 | 26,929 | 4890.54 |
| 2020 | 4  | 6 | 2 | 2,246,175  | 14,139.26 | 15.62 | 26,947 | 4794.93 |
| 2020 | 5  | 6 | 2 | 2,442,283  | 14,185.21 | 16.16 | 27,035 | 4739.23 |
| 2020 | 6  | 6 | 2 | 2,237,580  | 14,232.91 | 15.02 | 27,054 | 4468.34 |
| 2020 | 7  | 6 | 2 | 2,349,162  | 14,276.33 | 13.89 | 27,097 | 4397.33 |
| 2020 | 8  | 6 | 2 | 2,199,303  | 14,315.49 | 13.32 | 27,122 | 4705.07 |
| 2020 | 9  | 6 | 2 | 2,238,786  | 14,310.67 | 13.51 | 27,191 | 4714.85 |
| 2020 | 10 | 6 | 2 | 2,213,329  | 14,263.90 | 12.01 | 27,221 | 4635.95 |
| 2020 | 11 | 6 | 2 | 2,277,075  | 14,281.36 | 11.1  | 27,282 | 4879.84 |
| 2020 | 12 | 6 | 2 | 2,246,277  | 14,284.87 | 11.22 | 27,317 | 5785.18 |
| 2021 | 1  | 6 | 2 | 2,395,353  | 14,288.97 | 10.21 | 27,367 | 5983.62 |
| 2021 | 2  | 6 | 2 | 2,265,473  | 14,309.03 | 12.2  | 27,437 | 5923.00 |
| 2021 | 3  | 6 | 2 | 2,253,859  | 14,253.36 | 13.31 | 27,525 | 5843.66 |
| 2021 | 4  | 6 | 2 | 2,268,785  | 14,229.16 | 15.45 | 27,596 | 5147.74 |
| 2021 | 5  | 6 | 2 | 2,343,425  | 14,262.91 | 15.89 | 27,657 | 4807.70 |
| 2021 | 6  | 6 | 2 | 2,169,555  | 14,272.51 | 14.88 | 27,719 | 4676.73 |
| 2021 | 7  | 6 | 2 | 2,241,475  | 14,286.27 | 13.76 | 27,762 | 4714.24 |
| 2021 | 8  | 6 | 2 | 2,417,811  | 13,800.00 | 13.31 | 27,813 | 4893.80 |
| 2021 | 9  | 6 | 2 | 2,186,488  | 14,275.67 | 13.49 | 27,871 | 4547.86 |
| 2021 | 10 | 6 | 2 | 2,171,097  | 14,252.41 | 11.99 | 27,924 | 4710.87 |
| 2021 | 11 | 6 | 2 | 2,202,403  | 14,282.51 | 11.09 | 27,971 | 5121.70 |
| 2021 | 12 | 6 | 2 | 2,139,228  | 14,273.21 | 11.21 | 28,045 | 5410.17 |
| 2022 | 1  | 6 | 2 | 2,301,462  | 14,280.30 | 10.19 | 28,124 | 5447.19 |
| 2022 | 2  | 6 | 2 | 2,198,451  | 14,280.30 | 12.17 | 28,212 | 5467.31 |
| 2022 | 3  | 6 | 2 | 2,246,939  | 14,281.91 | 13.3  | 28,333 | 5003.04 |
| 2022 | 4  | 6 | 2 | 2,200,689  | 14,257.40 | 15.44 | 28,397 | 4804.83 |

|      |    |   |   |            |           |       |        |          |
|------|----|---|---|------------|-----------|-------|--------|----------|
| 2022 | 5  | 6 | 2 | 2,302,067  | 14,237.97 | 15.87 | 28,484 | 4921.10  |
| 2022 | 6  | 6 | 2 | 2,366,845  | 14,289.67 | 14.86 | 28,541 | 4880.45  |
| 2022 | 7  | 6 | 2 | 2,348,498  | 14,260.00 | 13.75 | 28,595 | 4859.70  |
| 2022 | 8  | 6 | 2 | 2,217,021  | 14,081.90 | 13.31 | 28,681 | 4783.80  |
| 2022 | 9  | 6 | 2 | 2,272,699  | 14,198.75 | 13.49 | 28,768 | 4922.80  |
| 2022 | 10 | 6 | 2 | 2,334,512  | 14,275.43 | 11.99 | 28,828 | 4803.98  |
| 2022 | 11 | 6 | 2 | 2,334,742  | 14,294.35 | 11.08 | 28,874 | 4834.21  |
| 2022 | 12 | 6 | 2 | 2,210,474  | 14,254.76 | 11.21 | 28,953 | 5121.77  |
| 2018 | 1  | 3 | 3 | 8,836,273  | 14,145.15 | 9.32  | 36,555 | 10060.01 |
| 2018 | 2  | 3 | 3 | 8,168,623  | 14,012.87 | 12.45 | 36,631 | 10614.79 |
| 2018 | 3  | 3 | 3 | 9,321,664  | 13,848.08 | 14.24 | 36,728 | 11803.00 |
| 2018 | 4  | 3 | 3 | 9,691,020  | 13,788.17 | 14.58 | 36,808 | 13256.29 |
| 2018 | 5  | 3 | 3 | 10,002,942 | 13,744.20 | 15.42 | 36,870 | 13329.45 |
| 2018 | 6  | 3 | 3 | 10,843,492 | 13,939.12 | 14.58 | 36,922 | 11206.34 |
| 2018 | 7  | 3 | 3 | 9,548,354  | 13,920.14 | 14.34 | 36,950 | 11618.45 |
| 2018 | 8  | 3 | 3 | 10,033,867 | 13,934.54 | 13.6  | 36,972 | 11222.99 |
| 2018 | 9  | 3 | 3 | 9,759,681  | 14,051.24 | 14.22 | 37,032 | 9620.12  |
| 2018 | 10 | 3 | 3 | 8,608,574  | 14,065.82 | 13.08 | 37,073 | 9571.34  |
| 2018 | 11 | 3 | 3 | 7,908,697  | 14,100.82 | 12.35 | 37,156 | 9629.98  |
| 2018 | 12 | 3 | 3 | 8,059,431  | 14,088.62 | 12.62 | 37,205 | 9853.01  |
| 2019 | 1  | 3 | 3 | 8,346,214  | 14,136.20 | 8.93  | 37,238 | 10140.87 |
| 2019 | 2  | 3 | 3 | 8,357,460  | 14,045.14 | 9.92  | 37,266 | 10800.92 |
| 2019 | 3  | 3 | 3 | 9,489,173  | 13,804.32 | 10.68 | 37,325 | 11252.19 |
| 2019 | 4  | 3 | 3 | 9,606,319  | 13,892.34 | 10.81 | 37,375 | 12851.99 |
| 2019 | 5  | 3 | 3 | 11,195,973 | 13,739.15 | 10.64 | 37,516 | 14630.68 |
| 2019 | 6  | 3 | 3 | 10,595,698 | 14,047.30 | 14.59 | 37,609 | 12037.11 |
| 2019 | 7  | 3 | 3 | 10,163,162 | 14,034.59 | 8.85  | 37,774 | 12460.07 |
| 2019 | 8  | 3 | 3 | 10,830,733 | 14,051.02 | 9.23  | 37,846 | 11914.82 |
| 2019 | 9  | 3 | 3 | 9,934,960  | 14,108.89 | 8.97  | 38,003 | 11061.77 |
| 2019 | 10 | 3 | 3 | 8,523,204  | 14,204.58 | 9.74  | 38,367 | 9803.81  |
| 2019 | 11 | 3 | 3 | 7,903,909  | 14,235.56 | 9.71  | 38,580 | 10004.90 |
| 2019 | 12 | 3 | 3 | 8,329,482  | 14,202.78 | 10.27 | 38,712 | 10245.98 |
| 2020 | 1  | 3 | 3 | 8,412,356  | 14,165.74 | 8.89  | 38,865 | 10048.48 |
| 2020 | 2  | 3 | 3 | 8,395,618  | 14,087.02 | 9.69  | 39,055 | 10667.98 |
| 2020 | 3  | 3 | 3 | 9,179,433  | 13,962.19 | 10.36 | 39,155 | 11947.43 |
| 2020 | 4  | 3 | 3 | 10,573,575 | 13,843.49 | 10.46 | 39,210 | 13128.68 |
| 2020 | 5  | 3 | 3 | 10,664,035 | 13,788.42 | 10.2  | 39,328 | 12558.58 |
| 2020 | 6  | 3 | 3 | 10,018,298 | 13,957.84 | 14.59 | 39,400 | 11229.42 |
| 2020 | 7  | 3 | 3 | 11,648,650 | 13,858.59 | 8.35  | 39,587 | 12202.57 |
| 2020 | 8  | 3 | 3 | 10,480,586 | 13,962.39 | 8.84  | 39,758 | 11026.15 |
| 2020 | 9  | 3 | 3 | 9,368,975  | 14,070.39 | 8.49  | 39,887 | 9649.90  |
| 2020 | 10 | 3 | 3 | 8,926,897  | 14,037.97 | 9.44  | 40,041 | 10145.37 |
| 2020 | 11 | 3 | 3 | 8,849,513  | 14,119.33 | 9.47  | 40,116 | 9730.15  |
| 2020 | 12 | 3 | 3 | 8,319,050  | 14,081.49 | 10.05 | 40,233 | 10379.48 |
| 2021 | 1  | 3 | 3 | 8,686,500  | 14,081.62 | 8.89  | 40,403 | 10958.17 |
| 2021 | 2  | 3 | 3 | 8,194,286  | 13,989.10 | 9.66  | 40,596 | 11330.80 |
| 2021 | 3  | 3 | 3 | 9,570,957  | 13,884.13 | 10.32 | 40,787 | 12906.52 |
| 2021 | 4  | 3 | 3 | 10,513,900 | 13,781.42 | 10.43 | 41,004 | 14103.26 |
| 2021 | 5  | 3 | 3 | 11,369,571 | 13,883.71 | 10.16 | 41,102 | 11610.36 |
| 2021 | 6  | 3 | 3 | 10,152,854 | 14,008.60 | 14.59 | 41,174 | 9466.55  |
| 2021 | 7  | 3 | 3 | 9,108,435  | 14,037.55 | 8.3   | 41,271 | 9747.60  |
| 2021 | 8  | 3 | 3 | 10,540,760 | 13,909.49 | 8.8   | 41,398 | 11139.13 |
| 2021 | 9  | 3 | 3 | 8,548,638  | 14,118.60 | 8.45  | 41,025 | 8974.30  |
| 2021 | 10 | 3 | 3 | 8,385,706  | 14,112.54 | 9.41  | 41,397 | 9028.76  |
| 2021 | 11 | 3 | 3 | 9,208,669  | 14,163.24 | 9.45  | 41,724 | 9012.21  |
| 2021 | 12 | 3 | 3 | 7,977,440  | 14,185.08 | 10.03 | 41,743 | 9094.07  |
| 2022 | 1  | 3 | 3 | 8,604,777  | 14,197.05 | 8.89  | 41,877 | 9092.52  |
| 2022 | 2  | 3 | 3 | 8,364,820  | 13,930.22 | 9.66  | 41,981 | 9659.48  |
| 2022 | 3  | 3 | 3 | 10,636,988 | 13,741.70 | 10.32 | 42,083 | 11105.88 |
| 2022 | 4  | 3 | 3 | 11,666,035 | 13,635.42 | 10.43 | 42,125 | 12941.22 |
| 2022 | 5  | 3 | 3 | 12,976,760 | 13,565.52 | 10.16 | 42,217 | 14518.19 |
| 2022 | 6  | 3 | 3 | 12,424,561 | 13,742.76 | 14.59 | 42,257 | 12092.28 |
| 2022 | 7  | 3 | 3 | 10,590,489 | 13,749.84 | 8.3   | 42,316 | 12063.58 |
| 2022 | 8  | 3 | 3 | 11,463,026 | 13,815.28 | 8.8   | 42,413 | 10811.85 |
| 2022 | 9  | 3 | 3 | 8,798,328  | 13,907.08 | 8.44  | 42,415 | 8794.51  |
| 2022 | 10 | 3 | 3 | 8,714,382  | 13,908.19 | 9.41  | 42,492 | 9307.42  |
| 2022 | 11 | 3 | 3 | 9,306,870  | 13,885.12 | 9.45  | 42,586 | 9772.66  |
| 2022 | 12 | 3 | 3 | 8,892,112  | 13,885.12 | 10.03 | 42,682 | 9692.50  |
| 2018 | 1  | 2 | 4 | 2,767,082  | 13,887.78 | 17.81 | 27,103 | 5986.83  |
| 2018 | 2  | 2 | 4 | 2,255,346  | 13,849.60 | 23.83 | 27,171 | 5990.74  |
| 2018 | 3  | 2 | 4 | 2,782,138  | 13,759.05 | 24.75 | 27,262 | 6326.70  |
| 2018 | 4  | 2 | 4 | 2,477,540  | 13,536.05 | 26.24 | 27,318 | 7198.09  |

|      |    |   |   |           |           |       |        |         |
|------|----|---|---|-----------|-----------|-------|--------|---------|
| 2018 | 5  | 2 | 4 | 3,343,222 | 13,557.01 | 28.77 | 27,389 | 7987.14 |
| 2018 | 6  | 2 | 4 | 2,817,163 | 13,607.47 | 29.76 | 27,455 | 7697.20 |
| 2018 | 7  | 2 | 4 | 3,714,028 | 13,588.01 | 29.45 | 27,491 | 7838.06 |
| 2018 | 8  | 2 | 4 | 2,807,487 | 13,585.59 | 28.65 | 27,589 | 7805.84 |
| 2018 | 9  | 2 | 4 | 3,777,927 | 13,643.52 | 29.16 | 27,682 | 7591.97 |
| 2018 | 10 | 2 | 4 | 2,920,686 | 13,711.11 | 26.44 | 27,733 | 7089.37 |
| 2018 | 11 | 2 | 4 | 3,399,565 | 13,811.03 | 22.4  | 27,781 | 6496.07 |
| 2018 | 12 | 2 | 4 | 2,390,324 | 13,864.08 | 22.16 | 27,837 | 6342.64 |
| 2019 | 1  | 2 | 4 | 2,799,827 | 13,892.17 | 19.51 | 27,882 | 5755.04 |
| 2019 | 2  | 2 | 4 | 2,344,747 | 13,821.58 | 21.49 | 27,920 | 5898.85 |
| 2019 | 3  | 2 | 4 | 2,764,226 | 13,824.38 | 23.77 | 27,991 | 5984.93 |
| 2019 | 4  | 2 | 4 | 2,336,745 | 13,277.66 | 25.84 | 28,040 | 6297.11 |
| 2019 | 5  | 2 | 4 | 3,408,478 | 13,575.65 | 27.96 | 28,079 | 7781.56 |
| 2019 | 6  | 2 | 4 | 3,090,456 | 13,612.90 | 28.69 | 28,154 | 7601.68 |
| 2019 | 7  | 2 | 4 | 3,997,063 | 13,636.35 | 28    | 28,203 | 7422.00 |
| 2019 | 8  | 2 | 4 | 3,045,245 | 13,588.46 | 27.84 | 28,295 | 7848.10 |
| 2019 | 9  | 2 | 4 | 4,037,661 | 13,638.49 | 27.51 | 28,375 | 7675.39 |
| 2019 | 10 | 2 | 4 | 3,496,781 | 13,737.61 | 25.39 | 28,438 | 7046.29 |
| 2019 | 11 | 2 | 4 | 3,711,454 | 13,874.44 | 22.82 | 28,480 | 6281.28 |
| 2019 | 12 | 2 | 4 | 2,596,442 | 13,940.69 | 20.88 | 28,536 | 6226.16 |
| 2020 | 1  | 2 | 4 | 3,110,300 | 13,910.00 | 19.68 | 28,594 | 6130.33 |
| 2020 | 2  | 2 | 4 | 2,624,061 | 13,855.32 | 21.27 | 28,640 | 6105.22 |
| 2020 | 3  | 2 | 4 | 2,889,323 | 13,700.64 | 23.68 | 28,689 | 6746.05 |
| 2020 | 4  | 2 | 4 | 2,716,423 | 13,588.13 | 25.8  | 28,713 | 7601.12 |
| 2020 | 5  | 2 | 4 | 3,906,306 | 13,601.07 | 27.88 | 28,746 | 7613.66 |
| 2020 | 6  | 2 | 4 | 3,090,457 | 13,600.22 | 28.59 | 28,850 | 7576.94 |
| 2020 | 7  | 2 | 4 | 4,098,529 | 13,540.50 | 27.86 | 28,819 | 7976.54 |
| 2020 | 8  | 2 | 4 | 2,824,006 | 13,608.13 | 27.76 | 28,750 | 7446.18 |
| 2020 | 9  | 2 | 4 | 3,919,396 | 13,655.63 | 27.35 | 28,783 | 7401.46 |
| 2020 | 10 | 2 | 4 | 2,881,523 | 13,687.56 | 25.29 | 28,787 | 6875.10 |
| 2020 | 11 | 2 | 4 | 3,508,427 | 13,771.94 | 22.86 | 28,783 | 6485.33 |
| 2020 | 12 | 2 | 4 | 2,480,049 | 13,884.93 | 20.76 | 28,804 | 5964.18 |
| 2021 | 1  | 2 | 4 | 2,937,565 | 13,880.28 | 19.69 | 28,895 | 5866.54 |
| 2021 | 2  | 2 | 4 | 2,418,172 | 13,850.01 | 21.25 | 28,919 | 5818.34 |
| 2021 | 3  | 2 | 4 | 2,889,099 | 13,756.58 | 23.67 | 28,998 | 6172.59 |
| 2021 | 4  | 2 | 4 | 2,686,933 | 13,666.63 | 25.8  | 29,061 | 7232.63 |
| 2021 | 5  | 2 | 4 | 3,683,143 | 13,571.10 | 27.88 | 29,106 | 7857.85 |
| 2021 | 6  | 2 | 4 | 3,365,217 | 13,614.04 | 28.58 | 29,166 | 7674.77 |
| 2021 | 7  | 2 | 4 | 3,971,558 | 13,580.34 | 27.85 | 29,231 | 7864.53 |
| 2021 | 8  | 2 | 4 | 3,162,539 | 13,585.73 | 27.75 | 29,241 | 7670.67 |
| 2021 | 9  | 2 | 4 | 3,959,110 | 13,615.27 | 27.34 | 29,324 | 7692.91 |
| 2021 | 10 | 2 | 4 | 3,090,322 | 13,657.05 | 25.28 | 29,352 | 7430.13 |
| 2021 | 11 | 2 | 4 | 3,617,882 | 13,807.81 | 22.87 | 29,417 | 6476.48 |
| 2021 | 12 | 2 | 4 | 2,510,427 | 13,786.13 | 20.75 | 29,479 | 6647.49 |
| 2022 | 1  | 2 | 4 | 3,122,134 | 13,868.67 | 19.69 | 29,557 | 6195.42 |
| 2022 | 2  | 2 | 4 | 2,414,057 | 13,871.50 | 21.24 | 29,592 | 6140.56 |
| 2022 | 3  | 2 | 4 | 2,829,711 | 13,754.60 | 23.66 | 29,650 | 6403.24 |
| 2022 | 4  | 2 | 4 | 2,636,878 | 13,611.77 | 25.8  | 29,704 | 7661.50 |
| 2022 | 5  | 2 | 4 | 3,745,777 | 13,480.22 | 27.87 | 29,738 | 8695.50 |
| 2022 | 6  | 2 | 4 | 3,431,661 | 13,602.73 | 28.58 | 29,757 | 7935.42 |
| 2022 | 7  | 2 | 4 | 4,331,236 | 13,594.86 | 27.85 | 29,835 | 8089.20 |
| 2022 | 8  | 2 | 4 | 3,273,622 | 13,560.02 | 27.75 | 29,882 | 8258.41 |
| 2022 | 9  | 2 | 4 | 4,197,274 | 13,664.68 | 27.34 | 29,900 | 7598.12 |
| 2022 | 10 | 2 | 4 | 3,154,858 | 13,762.23 | 25.28 | 29,960 | 7005.94 |
| 2022 | 11 | 2 | 4 | 3,509,211 | 13,789.92 | 22.87 | 30,033 | 6741.19 |
| 2022 | 12 | 2 | 4 | 2,598,873 | 13,844.57 | 20.75 | 30,054 | 6532.72 |
| 2018 | 1  | 5 | 1 | 2,782,537 | 13,933.19 | 15.19 | 28,185 | 4767.18 |
| 2018 | 2  | 5 | 1 | 2,937,187 | 13,896.18 | 18.22 | 28,159 | 4788.87 |
| 2018 | 3  | 5 | 1 | 2,815,593 | 13,848.96 | 21.37 | 28,232 | 5959.28 |
| 2018 | 4  | 5 | 1 | 3,261,943 | 13,757.84 | 22.63 | 28,291 | 5672.40 |
| 2018 | 5  | 5 | 1 | 3,186,404 | 13,666.68 | 26.6  | 28,311 | 6159.94 |
| 2018 | 6  | 5 | 1 | 3,591,901 | 13,721.60 | 25.27 | 28,379 | 5918.48 |
| 2018 | 7  | 5 | 1 | 3,355,020 | 13,712.21 | 24.53 | 28,387 | 5987.80 |
| 2018 | 8  | 5 | 1 | 3,575,078 | 13,706.52 | 24.72 | 28,454 | 5961.08 |
| 2018 | 9  | 5 | 1 | 3,362,117 | 13,747.41 | 23.97 | 28,465 | 5763.75 |
| 2018 | 10 | 5 | 1 | 3,760,480 | 13,793.94 | 22.27 | 28,502 | 5523.02 |
| 2018 | 11 | 5 | 1 | 3,177,292 | 13,872.56 | 19.79 | 28,544 | 5101.19 |
| 2018 | 12 | 5 | 1 | 3,317,704 | 13,908.85 | 18.71 | 28,582 | 5020.19 |
| 2019 | 1  | 5 | 1 | 2,935,364 | 13,947.02 | 18    | 28,607 | 4550.04 |
| 2019 | 2  | 5 | 1 | 3,573,479 | 13,883.41 | 18    | 28,670 | 4706.02 |
| 2019 | 3  | 5 | 1 | 2,928,660 | 13,883.99 | 19.87 | 28,766 | 4758.60 |
| 2019 | 4  | 5 | 1 | 3,353,960 | 13,789.87 | 22.7  | 28,788 | 5191.34 |

|      |    |   |   |           |           |       |        |         |
|------|----|---|---|-----------|-----------|-------|--------|---------|
| 2019 | 5  | 5 | 1 | 3,337,780 | 13,634.28 | 24.94 | 28,813 | 6071.97 |
| 2019 | 6  | 5 | 1 | 4,099,757 | 13,714.18 | 25.23 | 28,825 | 5878.45 |
| 2019 | 7  | 5 | 1 | 3,693,157 | 13,743.79 | 24.52 | 28,856 | 5688.18 |
| 2019 | 8  | 5 | 1 | 4,075,148 | 13,699.89 | 24.64 | 28,909 | 5982.09 |
| 2019 | 9  | 5 | 1 | 3,681,924 | 13,739.34 | 23.95 | 29,036 | 5877.66 |
| 2019 | 10 | 5 | 1 | 4,092,499 | 13,807.76 | 22.27 | 29,082 | 5565.51 |
| 2019 | 11 | 5 | 1 | 3,456,595 | 13,924.66 | 19.68 | 29,172 | 4905.28 |
| 2019 | 12 | 5 | 1 | 3,512,343 | 13,952.92 | 17.56 | 29,234 | 4875.74 |
| 2020 | 1  | 5 | 1 | 3,106,536 | 13,944.99 | 16.74 | 29,283 | 4846.04 |
| 2020 | 2  | 5 | 1 | 3,533,569 | 13,917.95 | 17.98 | 29,304 | 4826.97 |
| 2020 | 3  | 5 | 1 | 3,063,659 | 13,804.38 | 20.25 | 29,335 | 5384.06 |
| 2020 | 4  | 5 | 1 | 3,668,416 | 13,732.45 | 22.85 | 29,385 | 5984.73 |
| 2020 | 5  | 5 | 1 | 3,632,926 | 13,683.45 | 24.78 | 29,430 | 5988.30 |
| 2020 | 6  | 5 | 1 | 4,191,393 | 13,724.61 | 25.22 | 29,491 | 6022.56 |
| 2020 | 7  | 5 | 1 | 3,947,575 | 13,678.47 | 24.52 | 29,534 | 6216.58 |
| 2020 | 8  | 5 | 1 | 4,348,522 | 13,729.00 | 24.63 | 29,545 | 5881.87 |
| 2020 | 9  | 5 | 1 | 3,743,615 | 13,775.37 | 23.95 | 29,592 | 5776.93 |
| 2020 | 10 | 5 | 1 | 4,039,279 | 13,799.61 | 22.27 | 29,648 | 5551.32 |
| 2020 | 11 | 5 | 1 | 3,456,031 | 13,871.08 | 19.67 | 29,716 | 5191.74 |
| 2020 | 12 | 5 | 1 | 3,615,221 | 13,957.46 | 17.45 | 29,747 | 4885.71 |
| 2021 | 1  | 5 | 1 | 3,175,943 | 13,953.73 | 16.62 | 29,807 | 4905.53 |
| 2021 | 2  | 5 | 1 | 3,421,474 | 13,882.61 | 17.98 | 29,905 | 4918.73 |
| 2021 | 3  | 5 | 1 | 3,099,531 | 13,847.92 | 20.29 | 29,995 | 5213.79 |
| 2021 | 4  | 5 | 1 | 3,696,831 | 13,777.07 | 22.87 | 30,031 | 5647.25 |
| 2021 | 5  | 5 | 1 | 3,542,881 | 13,727.82 | 24.76 | 30,077 | 6016.02 |
| 2021 | 6  | 5 | 1 | 4,044,456 | 13,739.78 | 25.22 | 30,187 | 5817.09 |
| 2021 | 7  | 5 | 1 | 3,676,786 | 13,727.11 | 24.52 | 30,219 | 5912.50 |
| 2021 | 8  | 5 | 1 | 4,287,795 | 13,719.88 | 24.63 | 30,256 | 5835.44 |
| 2021 | 9  | 5 | 1 | 3,674,585 | 13,729.77 | 23.95 | 30,267 | 5719.66 |
| 2021 | 10 | 5 | 1 | 4,208,822 | 13,748.55 | 22.27 | 30,322 | 5657.25 |
| 2021 | 11 | 5 | 1 | 3,571,014 | 13,858.96 | 19.67 | 30,404 | 4986.42 |
| 2021 | 12 | 5 | 1 | 3,720,830 | 13,826.12 | 17.44 | 30,451 | 5115.83 |
| 2022 | 1  | 5 | 1 | 3,316,850 | 13,887.03 | 16.61 | 30,536 | 4767.13 |
| 2022 | 2  | 5 | 1 | 3,508,606 | 13,900.77 | 17.98 | 30,594 | 4703.76 |
| 2022 | 3  | 5 | 1 | 3,041,759 | 13,859.54 | 20.29 | 30,641 | 4880.68 |
| 2022 | 4  | 5 | 1 | 3,718,574 | 13,731.96 | 22.87 | 30,722 | 5866.36 |
| 2022 | 5  | 5 | 1 | 3,640,501 | 13,621.19 | 24.76 | 30,809 | 6563.46 |
| 2022 | 6  | 5 | 1 | 4,669,633 | 13,707.82 | 25.22 | 30,874 | 5955.82 |
| 2022 | 7  | 5 | 1 | 4,110,533 | 13,699.46 | 24.52 | 30,953 | 6133.94 |
| 2022 | 8  | 5 | 1 | 4,360,984 | 13,677.30 | 24.63 | 30,966 | 6144.94 |
| 2022 | 9  | 5 | 1 | 3,876,564 | 13,745.17 | 23.95 | 31,053 | 5680.75 |
| 2022 | 10 | 5 | 1 | 4,276,994 | 13,819.85 | 22.27 | 31,123 | 5318.00 |
| 2022 | 11 | 5 | 1 | 3,494,844 | 13,835.52 | 19.67 | 31,198 | 5321.22 |
| 2022 | 12 | 5 | 1 | 3,733,958 | 13,898.88 | 17.44 | 31,280 | 5203.49 |

### **4.1.2 Algoritmo de aprendizaje**

Una vez conformado nuestro dataset con valores de datos históricos definidos como series de tiempo de cada una de las variables más representativas de la demanda de energía y para poder generar nuestra predicción necesitamos establecer el algoritmo cuyo sesgo de aprendizaje sea el que mejor se adapte a nuestro conjunto de datos.

Al hablar de series de tiempo se habla de secuencias y las redes neuronales del aprendizaje profundo han tenido un alto impacto en este tipo de pronósticos, por lo tanto, probaremos nuestro dataset en este algoritmo de red neuronal recurrente RNN que basado en el análisis de la información obtenida de las diferentes fuentes de investigación se define que es el más eficaz para este tipo de análisis de datos predictivos de demanda.

Para realizar una evaluación de la información de la predicción de la demanda de energía eléctrica, el algoritmo de aprendizaje toma dos fuentes de información para generar la mejor función que mejor se adapte a nuestros datos, como lo es el dataset (Conjunto de datos) y el sesgo inductivo.

El sesgo es aquel que mejor evalúa las características de la mejor función que se adapte a los datos. A partir de esto las redes neuronales profundas permiten un sesgo débil el cual es el que mejor se ajusta en demasía al conjunto de datos.

Dataset\_Demanda.csv es el conjunto de datos que contiene la demanda y sus características en la región oriente de los estados de Puebla y Veracruz.

Es turno de evaluar nuestro modelo a través del código de Python que se muestran para la evaluación del caso de estudio de las variables eléctricas obtenidas de nuestro dataset de datos.

```
import pandas as pd
import numpy as np
import tensorflow as tf
import matplotlib.pyplot as plt
import seaborn as sns
```

```
# introducir el dataset
datos = pd.read_csv ('Dataset_Demanda.csv')
```

datos

|     | Año  | Mes | Area | Region | Energía (kWh) | Voltaje   | Temperatura °C | Total de Usuarios | Demanda (kW) |
|-----|------|-----|------|--------|---------------|-----------|----------------|-------------------|--------------|
| 0   | 2018 | 1   | 1    | 1      | 8745887       | 13998.18  | 15.94          | 53978             | 6786.1174    |
| 1   | 2018 | 2   | 1    | 1      | 8439408       | 131134.02 | 21.84          | 54002             | 7148.3966    |
| 2   | 2018 | 3   | 1    | 1      | 9324468       | 131127.07 | 22.71          | 54097             | 8057.0711    |
| 3   | 2018 | 4   | 1    | 1      | 10911986      | 127657.47 | 23.68          | 54248             | 9523.0138    |
| 4   | 2018 | 5   | 1    | 1      | 11758135      | 127627.13 | 27.79          | 54370             | 10861.1407   |
| ... | ...  | ... | ...  | ...    | ...           | ...       | ...            | ...               | ...          |
| 355 | 2022 | 8   | 5    | 1      | 4360984       | 13677.30  | 24.63          | 30966             | 6144.9427    |
| 356 | 2022 | 9   | 5    | 1      | 3876564       | 13745.17  | 23.95          | 31053             | 5680.7455    |
| 357 | 2022 | 10  | 5    | 1      | 4276994       | 13819.85  | 22.27          | 31123             | 5318.0013    |
| 358 | 2022 | 11  | 5    | 1      | 3494844       | 13835.52  | 19.67          | 31198             | 5321.2207    |
| 359 | 2022 | 12  | 5    | 1      | 3733958       | 13898.88  | 17.44          | 31280             | 5203.4904    |

360 rows x 9 columns

Se pueden apreciar las variables que mejor componen al conjunto de datos son:

- 1.- Año
- 2.- Mes
- 3.- Municipio
- 4.- Región
- 5.- Energía (KWH)
- 6.- Voltaje
- 7.- Temperatura °C
- 8.- Total de Usuarios
- 9.- Demanda (KW)

Por lo que el conjunto de datos está compuesto por 360 entidades y 9 columnas o variables.

Nuestro conjunto de datos no contiene valores nulos debido a la correcta preparación de datos bajo las técnicas mencionadas en el presente trabajo de investigación.

```
datos.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 360 entries, 0 to 359
Data columns (total 9 columns):
#   Column                Non-Null Count  Dtype
---  ---                ---
0   Año                   360 non-null    int64
1   Mes                   360 non-null    int64
2   Area                  360 non-null    int64
3   Region                360 non-null    int64
4   Energía (kWh)         360 non-null    int64
5   Voltaje               360 non-null    float64
6   Temperatura °C       360 non-null    float64
7   Total de Usuarios     360 non-null    int64
8   Demanda (kW)         360 non-null    float64
dtypes: float64(3), int64(6)
memory usage: 25.4 KB
```

Obtenemos de nuestro conjunto de datos la cantidad de elementos por columna, la media, desviación estándar, los valores percentiles, valores mínimos y máximos como referencia de análisis de datos.

```
datos.describe()
```

|              | Año         | Mes        | Area       | Region     | Energía (kWh) | Voltaje       | Temperatura °C | Total de Usuarios | Demanda (kW) |
|--------------|-------------|------------|------------|------------|---------------|---------------|----------------|-------------------|--------------|
| <b>count</b> | 360.000000  | 360.000000 | 360.000000 | 360.000000 | 3.600000e+02  | 360.000000    | 360.000000     | 360.000000        | 360.000000   |
| <b>mean</b>  | 2020.000000 | 6.500000   | 3.500000   | 2.333333   | 7.798382e+06  | 15160.708528  | 18.139611      | 39723.686111      | 6744.007475  |
| <b>std</b>   | 1.416182    | 3.456857   | 1.710202   | 1.107080   | 5.593940e+06  | 12128.100709  | 6.315284       | 12780.037207      | 2864.493716  |
| <b>min</b>   | 2018.000000 | 1.000000   | 1.000000   | 1.000000   | 2.050949e+06  | 13137.990000  | 8.300000       | 25533.000000      | 2567.494100  |
| <b>25%</b>   | 2019.000000 | 3.750000   | 2.000000   | 1.000000   | 2.933688e+06  | 13729.292500  | 12.965000      | 28591.000000      | 4803.146475  |
| <b>50%</b>   | 2020.000000 | 6.500000   | 3.500000   | 2.500000   | 6.286771e+06  | 13839.605000  | 16.680000      | 33917.500000      | 6019.290350  |
| <b>75%</b>   | 2021.000000 | 9.250000   | 5.000000   | 3.000000   | 1.168214e+07  | 14032.090000  | 23.955000      | 55164.500000      | 9016.346675  |
| <b>max</b>   | 2022.000000 | 12.000000  | 6.000000   | 4.000000   | 2.741266e+07  | 131134.020000 | 29.760000      | 60427.000000      | 14630.882100 |

Se obtiene la desviación con un valor de 2864.49 KW de dispersión de demanda que es la variable a predecir respecto a nuestra media con valor de 6744.00 KW. Los percentiles que indican el valor de la variable de la demanda por debajo del cual se encuentra un porcentaje dado de observaciones en un grupo obtenemos que 25 % están por debajo de 2864.49 KW, el 50% es inferior a 6019.29 KW y el 75 % inferior a 9016.34 KW.

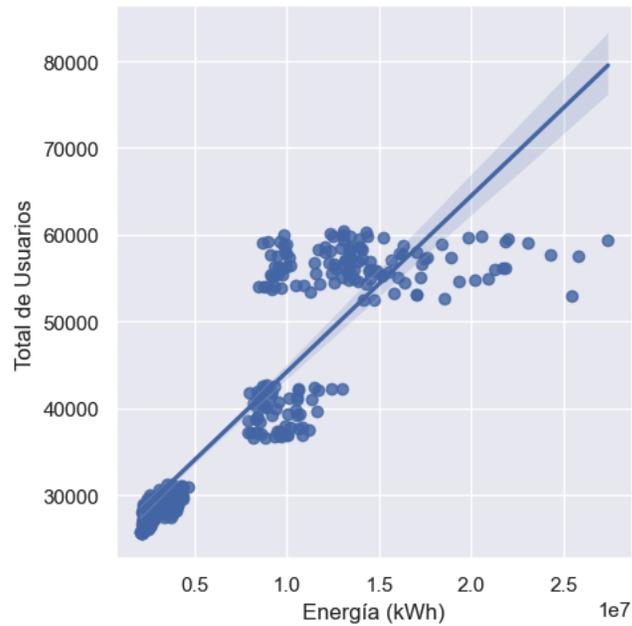
Se genera un análisis de correlación del conjunto de datos utilizado y su relación entre variables. Se genera un mapa de calor que se muestra en la Figura 30 que permite representar los datos en forma bidimensional y observar hasta qué punto las variables están relacionadas linealmente.



Figura 29 Mapa de calor de correlación de variables

Se aprecia la correlación existente entre las 9 variables a través de la intensidad de color. En la gráfica anterior se aprecia que existe una gran relación entre las variables de Energía (KWH) y el Total de Usuarios con un valor de 0.89. por lo que mostramos en un gráfico de dispersión Figura 31 una relación alta donde el valor de  $y$  (total de Usuarios) se incrementa nítidamente a medida que el valor de  $x$  (Demanda) aumenta.

```
sns.lmplot(x='Energía (kWh)',y='Total de Usuarios', data=datos)
<seaborn.axisgrid.FacetGrid at 0x1444ee410>
```



*Figura 30 Gráfica de dispersión de correlación de variables*

Una vez analizada la información de la demanda se procede a entrenar el modelo computacional de red neuronal por lo cual es necesario dividir el conjunto de datos en el *Conjunto de datos de Entrenamiento* y el *Conjunto de datos de Prueba*.

Debido a que se utiliza el algoritmo de RNN se divide el conjunto de datos para obtener el eje  $x$  y el eje de las  $y$ , donde  $x$  son todas las características numéricas de las variables eléctricas predictoras.

```
X = datos[['Año', 'Mes', 'Area', 'Region', 'Energía (kWh)', 'Voltaje',
          'Temperatura °C', 'Total de Usuarios', 'Demanda (kW)']]
```

```
X
```

|     | Año  | Mes | Area | Region | Energía (kWh) | Voltaje   | Temperatura °C | Total de Usuarios | Demanda (kW) |
|-----|------|-----|------|--------|---------------|-----------|----------------|-------------------|--------------|
| 0   | 2018 | 1   | 1    | 1      | 8745887       | 13998.18  | 15.94          | 53978             | 6786.1174    |
| 1   | 2018 | 2   | 1    | 1      | 8439408       | 131134.02 | 21.84          | 54002             | 7148.3966    |
| 2   | 2018 | 3   | 1    | 1      | 9324468       | 131127.07 | 22.71          | 54097             | 8057.0711    |
| 3   | 2018 | 4   | 1    | 1      | 10911986      | 127657.47 | 23.68          | 54248             | 9523.0138    |
| 4   | 2018 | 5   | 1    | 1      | 11758135      | 127627.13 | 27.79          | 54370             | 10861.1407   |
| ... | ...  | ... | ...  | ...    | ...           | ...       | ...            | ...               | ...          |
| 355 | 2022 | 8   | 5    | 1      | 4360984       | 13677.30  | 24.63          | 30966             | 6144.9427    |
| 356 | 2022 | 9   | 5    | 1      | 3876564       | 13745.17  | 23.95          | 31053             | 5680.7455    |
| 357 | 2022 | 10  | 5    | 1      | 4276994       | 13819.85  | 22.27          | 31123             | 5318.0013    |
| 358 | 2022 | 11  | 5    | 1      | 3494844       | 13835.52  | 19.67          | 31198             | 5321.2207    |
| 359 | 2022 | 12  | 5    | 1      | 3733958       | 13898.88  | 17.44          | 31280             | 5203.4904    |

360 rows x 9 columns

La variable objetivo de predicción y que para nuestro caso es la Demanda (KW).

```
y = datos['Demanda (kW)']
```

```
y
```

```
0      6786.1174
1      7148.3966
2      8057.0711
3      9523.0138
4     10861.1407
...
355     6144.9427
356     5680.7455
357     5318.0013
358     5321.2207
359     5203.4904
```

Name: Demanda (kW), Length: 360, dtype: float64

Se evalúa el modelo extrayendo los datos de entrada y salida, así como la división del conjunto de datos de entrenamiento y prueba en el orden 80/20 donde el conjunto de pruebas sea del 20 % y el conjunto de entrenamiento del 80 % utilizando el parámetro de random\_state de 42 que es la semilla de dicho conjunto de datos.

```
# extraer datos de entrada y de salida
X_train = datos.iloc[:, :-1].values
y_train = datos.iloc[:, -1]
```

```
# dividir los datos en conjunto de entrenamiento y prueba
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X_train, y_train, test_size = 0.20, random_state=42)
```

Mediante la ayuda de keras y Tensor Flow procedemos a estandarizar los datos cuya técnica es utilizada eficientemente en el entrenamiento de redes neuronales de regresión RNN por que consiste en transformar los datos de origen con una media cero y una desviación de estándar de uno. Al estandarizar los datos se eliminan las diferencias de escala entre las diferentes características del conjunto de datos logrando se acelere el proceso de entrenamiento teniendo una región más lineal y estable, lo que facilita la propagación hacia atrás de los gradientes durante el entrenamiento y mejora la velocidad de convergencia de la función de activación. La estandarización permite evitar el desequilibrio de los sesgos en los pesos del modelo durante el entrenamiento y por lo tanto mejorar la estabilidad del modelo.

```
# Estandarización del conjunto de datos
# Import `StandardScaler` from `sklearn.preprocessing`
from sklearn.preprocessing import StandardScaler

# Define the scaler
scaler = StandardScaler().fit(X_train)

# Scale the train set
X_train = scaler.transform(X_train)

# Scale the test set
X_test = scaler.transform(X_test)
```

Establecemos la función de activación que es el constructor del modelo de red y que realizará el proceso de las iteraciones de la red neuronal donde indicamos la cantidad de capas de entrada con 9, 18 capas ocultas y 1 capa de salida.

```
# Import `Sequential` from `keras.models`
from keras.models import Sequential

# Import `Dense` from `keras.layers`
from keras.layers import Dense

# Función de activación
# Initialize the constructor
model = Sequential()

# Add an input layer
model.add(Dense(9, activation='relu', input_shape=(8,)))
# Add one hidden layer
model.add(Dense(18, activation='relu'))

# Add an output layer
model.add(Dense(1, activation='linear'))
```

Compilamos nuestro modelo con las métricas de evaluación de la red neuronal de regresión con el optimizador, los errores cuadráticos y absolutos. El modelo de entrenamiento de nuestra red neuronal fue medido con ciertas pruebas y teniendo un valor de optimización correcto de 700 epochs y un número de 8 muestras de

entrenamiento que se propaga a través de la red neuronal haciendo el cálculo eficiente de los pesos.

```
# compilar el modelo
model.compile(optimizer='adam', loss='mean_squared_error', metrics=['mean_absolute_error'])
```

```
# entrenar el modelo
history = model.fit(X_train, y_train, epochs=700, batch_size=8, validation_split=0.2)
```

Generamos las predicciones con la creación de una variable que almacena las predicciones realizadas por el modelo, es decir, el modelo predice el valor de la demanda contenida en el dataset de prueba X\_test a partir de los valores de las columnas *Año, Mes, Municipio, Región, Energía (KWH), Voltaje, Temperatura °C y Total de Usuarios* contenidas en él.

```
loss = model.evaluate(X_test, y_test)
predictions = model.predict(X_test)
3/3 [=====] - 0s 1ms/step - loss: 811152.2500 - mean_absolute_error: 600.2721
3/3 [=====] - 0s 531us/step
```

Se imprime el array de predicciones generadas de la variable predictions:

```
array([[10694.11 ],
       [10636.888 ],
       [ 7281.3965],
       [ 5128.131 ],
       [10222.951 ],
       [ 7528.41  ],
       [ 9414.055 ],
       [ 9344.227 ],
       [ 7564.6646],
       [ 4678.999 ],
       [ 2968.7805],
       [ 6230.5933],
       [13130.173 ],
       [ 5790.567 ],
       [ 2733.628 ],
       [ 4811.3535],
       [10172.616 ],
       [ 5612.5737],
       [ 4782.317 ],
       [ 6166.0444],
       [ 4835.769 ],
       [11240.19  ],
       [ 4607.488 ],
       [ 9662.2    ],
       [ 7372.3433],
       [11753.725 ],
       [ 4467.5503],
```

```
[ 4870.604 ],
[ 7290.923 ],
[ 2531.8672],
[ 7025.016 ],
[ 9462.29  ],
[ 5164.101 ],
[ 4400.627 ],
[ 2810.006 ],
[ 9256.058 ],
[ 8475.474 ],
[ 5881.4707],
[ 3071.527 ],
[ 2187.3088],
[12583.113 ],
[10849.624 ],
[ 5885.1294],
[ 3495.4253],
[ 5020.5996],
[12066.242 ],
[10489.662 ],
[10610.545 ],
[ 9498.595 ],
[ 6112.177 ],
[ 6968.29  ],
[ 9768.278 ],
[ 9198.013 ],
[ 2745.4    ],
[ 5796.671 ],
[ 7623.1426],
[ 8898.817 ],
[ 2014.2316],
[ 3798.0413],
[ 7651.1016],
[ 4646.4697],
[ 8549.484 ],
[12112.313 ],
[ 2788.3506],
[ 2936.0103],
[ 3682.354 ],
[ 5511.7236],
[12634.068 ],
[ 5359.8403],
[ 2889.8376],
[12579.037 ],
[10250.883 ]], dtype=float32)
```

Comparamos con los valores reales de `y_test` que para nuestro modelo fue el valor objetivo de predicción de la Demanda (KW).

```

y_test
224    8974.3000
42     11062.9960
285    7430.1253
302    5959.2800
56     9432.6422
...
199    11914.8210
334    5191.7395
93     2845.7897
198    12460.0658
197    12037.1084
Name: Demanda (kW), Length: 72, dtype: float64

```

*Figura 31 Entrenamiento y validación de pérdida*

Se puede apreciar las predicciones de la demanda con el modelo y las demandas reales respectivamente. Para la demanda de la fila (285) cuyo valor real es 7430.1253 KW, el modelo predijo que su valor sería de 7302.328 KW, para el valor de la fila (334) cuyo valor de demanda real es de 5191.7395 KW, el modelo predijo una demanda de 5331.984 KW. Como podemos apreciar, para estos dos casos los resultados obtenidos se acercan y la diferencia entre esas cantidades es el error del modelo.

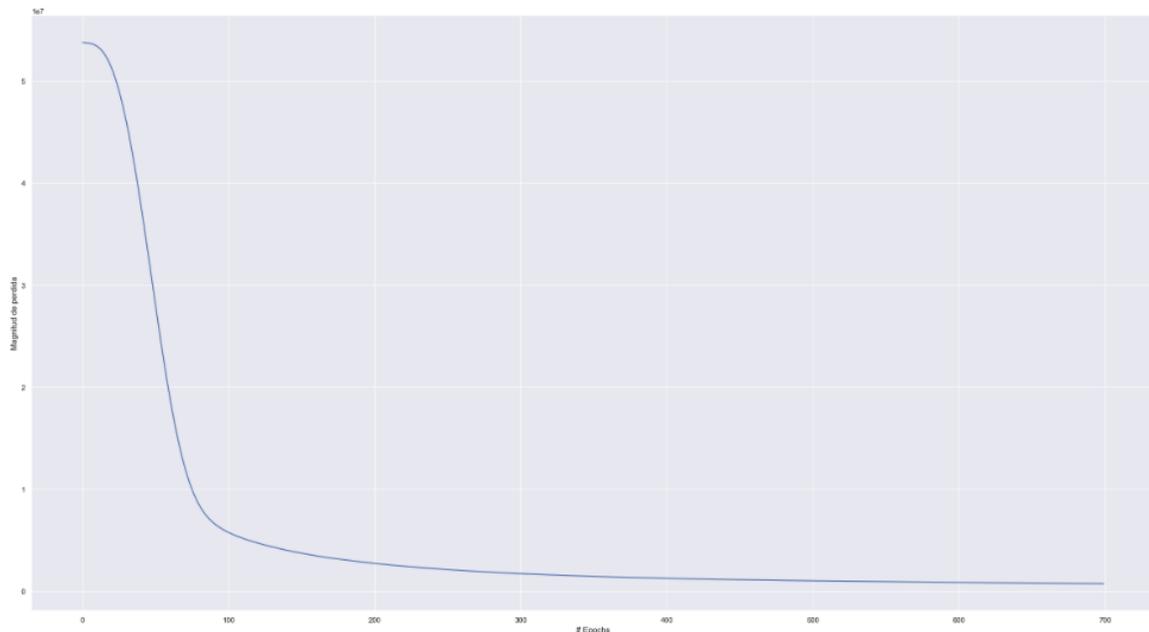
Se comparan los resultados de los valores reales con los valores de predicción obtenidos por el modelo propuesto de red neuronal RNN ver Tabla 8.

Tabla 9 Tabla comparativa de valor real contra valor previsto

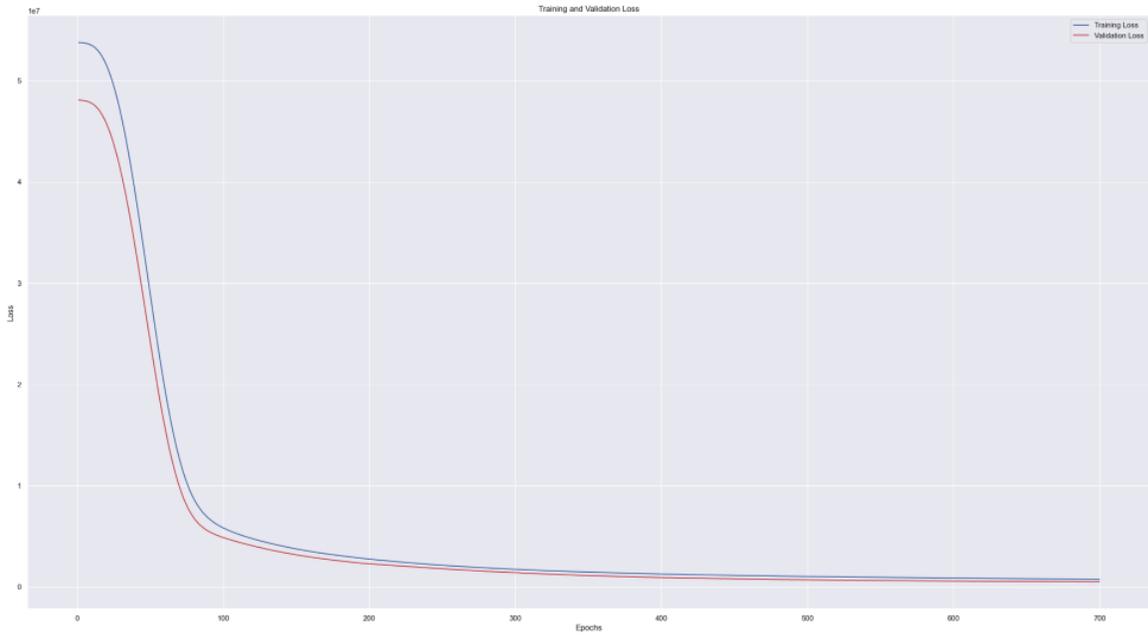
| MES | VALORES REALES | VALORES DE PREDICCIÓN |
|-----|----------------|-----------------------|
| 1   | 8974.3         | 10825.485             |
| 2   | 11062.996      | 10661.036             |
| 3   | 7430.1253      | 7191.62               |
| 4   | 5959.28        | 5187.6426             |
| 5   | 9432.6422      | 10453.458             |
| 6   | 7401.4609      | 7559.4717             |
| 7   | 8828.8937      | 8620.04               |
| 8   | 8127.4         | 9412.714              |
| 9   | 7591.97        | 7732.4536             |
| 10  | 4800.66        | 4467.6484             |
| 11  | 2814.5783      | 3030.0667             |
| 12  | 6133.9353      | 6297.559              |
| 13  | 12202.5696     | 12577.639             |
| 14  | 5647.2472      | 5862.1763             |
| 15  | 2968.9352      | 2749.6147             |
| 16  | 4637.46        | 4647.589              |
| 17  | 10245.978      | 10496.946             |
| 18  | 4767.18        | 6171.419              |
| 19  | 4570.0996      | 4674.2266             |
| 20  | 5680.7455      | 6130.376              |
| 21  | 6660.9276      | 4663.8013             |

|    |            |           |
|----|------------|-----------|
| 22 | 9747.5952  | 11041.842 |
| 23 | 4846.0445  | 5118.6997 |
| 24 | 9557.5379  | 9968.615  |
| 25 | 7005.94    | 7401.545  |
| 26 | 11610.3639 | 10872.132 |
| 27 | 4794.9252  | 4032.0627 |
| 28 | 4826.9704  | 5179.2686 |
| 29 | 7026.0167  | 7644.198  |
| 30 | 2835.3316  | 2307.7537 |
| 31 | 7601.1248  | 6962.9473 |
| 32 | 9277.0135  | 9946.764  |
| 33 | 5115.8347  | 5392.5293 |
| 34 | 4739.2331  | 3865.9524 |
| 35 | 2917.4512  | 3005.2002 |
| 36 | 9314.3498  | 8625.547  |
| 37 | 7513.3352  | 9144.285  |
| 38 | 5866.3595  | 6051.0254 |
| 39 | 2895.1651  | 3104.598  |
| 40 | 2766.4277  | 1712.339  |
| 41 | 10060.01   | 11993.72  |
| 42 | 11137.0904 | 10960.388 |
| 43 | 5918.48    | 6180.784  |
| 44 | 2978.1806  | 4470.506  |
| 45 | 5922.9953  | 5197.0747 |
| 46 | 14518.1948 | 11143.595 |
| 47 | 10825.5838 | 10280.128 |
| 48 | 10556.5735 | 10212.868 |
| 49 | 10202.2954 | 9714.431  |
| 50 | 5912.4961  | 6028.851  |
| 51 | 6476.4839  | 7207.1094 |
| 52 | 9307.417   | 10020.698 |
| 53 | 11229.4159 | 8393.121  |
| 54 | 2921.2149  | 2250.7368 |
| 55 | 5551.3202  | 5644.2964 |
| 56 | 7670.667   | 7463.1533 |
| 57 | 9523.0138  | 9070.989  |
| 58 | 2914.1042  | 1651.248  |
| 59 | 3054.1026  | 2775.429  |
| 60 | 7692.9073  | 7614.0713 |
| 61 | 2768.38    | 6142.218  |
| 62 | 8008.301   | 9419.635  |
| 63 | 11026.1494 | 11785.247 |
| 64 | 2766.2515  | 2019.3687 |
| 65 | 2827.2716  | 3596.8503 |
| 66 | 2809.4308  | 4244.6147 |
| 67 | 5321.2207  | 5622.5527 |
| 68 | 11914.821  | 12344.755 |
| 69 | 5191.7395  | 5179.8574 |
| 70 | 2845.7897  | 2753.432  |
| 71 | 12460.0658 | 12354.925 |
| 72 | 12037.1084 | 9230.553  |

Se mide la magnitud de perdida Figura 32 donde se observa que el modelo se ajusta y converge correctamente en la cantidad de 700 epochs para predecir correctamente en nuestro conjunto de datos, es decir, el número de veces que nuestro algoritmo de entrenamiento recorre el conjunto completo de datos durante el proceso de optimización de entrenamiento. Se utilizo el criterio de parada temprana (Early Stopping) para el rendimiento optimo de nuestro modelo. Se observa que los pesos de la red se ajustaron correctamente por lo que las predicciones se acercaron los mas posible a los valores reales. Pues a medida que nuestra red neuronal se entreno la magnitud y validación de perdida disminuyo Figura 33.

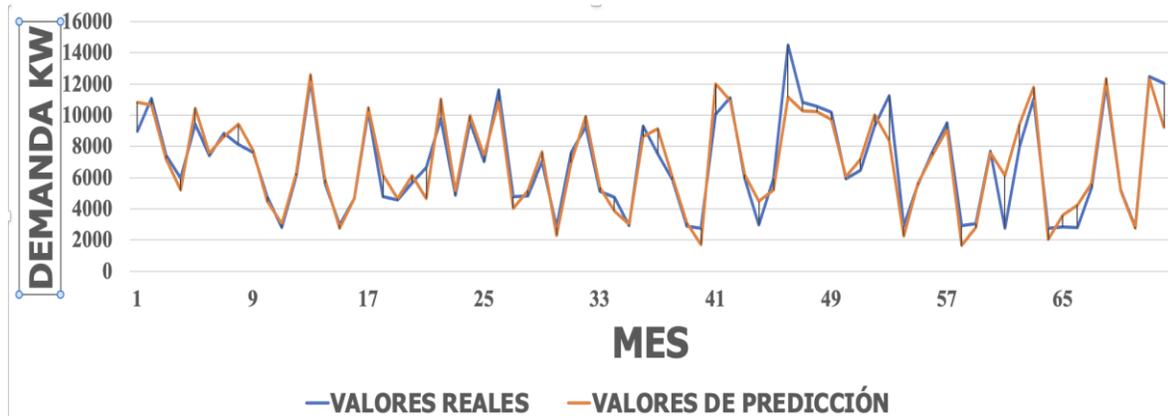


*Figura 32 Magnitud de perdida*



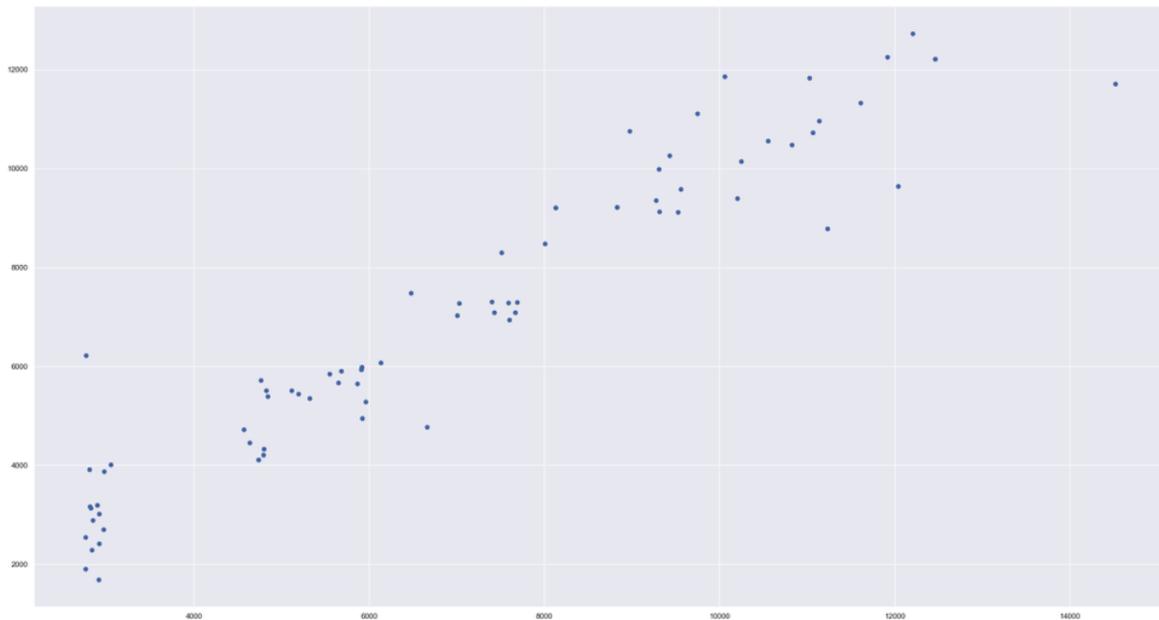
*Figura 33 Magnitud de perdida de entrenamiento y validación*

Se muestra la gráfica para los datos mensuales de la demanda (KWH), donde observa que la diferencia entré él valor real y el valor pronosticado ver Figura 34 de la predicción de un periodo de 72 meses es muy estrecha entre ambas indicando un buen ajuste entré los valores analizados.



*Figura 34 Valor real frente a valor previsto*

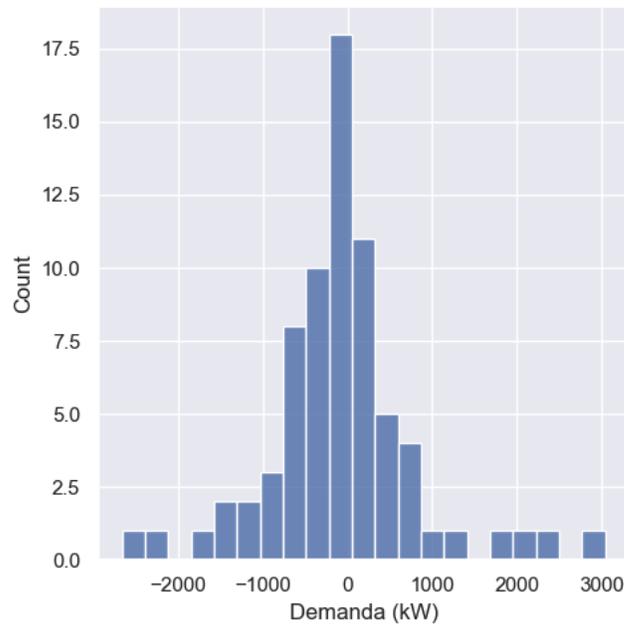
Para obtener una visión general de la comparación entre los valores reales y los obtenidos por las predicciones se genera un gráfico de dispersión de los datos con diagrama de puntos con los valores de  $y_{\text{test}}$  (reales) en el eje  $x$  y los valores de predicción en el eje  $y$ , ver Figura 35.



*Figura 35 Diagrama de puntos de dispersión*

Se observa que los puntos se acercan mucho a la diagonal lo cual indica que se trata del mismo valor para la demanda real y la predicción. Lo antes mencionado significa que el modelo predice de manera eficaz la demanda de energía eléctrica.

Se observa el histograma del error residual donde se observa la diferencia entre la demanda real ( $y_{\text{test}}$ ) y lo que predijo el modelo Figura 36.



*Figura 36 Histograma valor residual.*

Se puede apreciar que los valores del gráfico siguen una distribución normal (una distribución en forma de campana) lo que indica que el modelo de red neuronal es eficiente para capturar las relaciones entre las variables de entrada y la variable objetivo de la demanda de energía eléctrica de nuestro conjunto de datos.

El sistema de predicción de la demanda de energía necesita asegurar los mejores resultados para determinar si el pronóstico se desvía de los resultados reales debido a la aleatoriedad o a un cambio esencial en los datos. Por lo que se evalúan las métricas más significativas para evaluar la eficiencia del modelo de red neuronal recurrente de regresión RNN.

### **Coefficiente de correlación $R$**

El coeficiente de correlación  $R$  mide la fuerza y la dirección de la relación lineal entre dos variables. Es una medida estadística que indica cómo se relacionan dos conjuntos de datos numéricos. El valor de  $R$  varía entre -1 y 1.

El valor absoluto de  $R$  indica la fuerza de la correlación, siendo 1 el valor máximo. Cuanto más cercano esté el valor de  $R$  a 1 o -1, más fuerte será la correlación. Por otro lado, cuanto más cercano esté a 0, más débil será la correlación.

La métrica de coeficiente de correlación  $R$  para medir la eficiencia del ajuste lineal de nuestro modelo de la red neuronal recurrente de regresión, entre los valores obtenidos de la predicción comparados contra los valores reales, tuvo un porcentaje de evaluación de eficiencia del 95.77 %,  $R = 0.9577$ , el cual indica un ajuste lineal demasiado aceptable ya su valor es muy cercano al valor  $|r|$  a 1.

```
correlation = np.corrcoef(predictions.flatten(), y_test)[0, 1]
print("Coeficiente de correlación R:", correlation)
```

Coeficiente de correlación R: 0.9577479573791222

### Coeficiente de determinación ( $R^2$ )

El coeficiente de determinación  $R^2$ , es una medida estadística que proporciona información sobre la proporción de variabilidad de los datos de salida que se explica por el modelo de regresión. Es decir, mide qué tan bien se ajustan los valores predichos por el modelo a los valores reales. El  $R^2$  varía entre un intervalo de 0 a 1, donde:

$R^2 = 0$  indica que el modelo no es capaz de explicar ninguna variabilidad en los datos y su rendimiento es equivalente a predecir la media de los datos de salida.

$R^2 = 1$  indica que el modelo es capaz de explicar toda la variabilidad en los datos, logrando una predicción acertada y perfecta..

Después de evaluar el modelo en el conjunto de datos de entrenamiento se obtiene un valor de rendimiento del 90.91 %,  $R^2 = 0.9091$ . Esto significa que aproximadamente el 91 % de la variabilidad de los datos de la predicción de la demanda de energía es explicada correctamente por el modelo de red neuronal recurrente de regresión.

```
from sklearn.metrics import r2_score
r2 = r2_score(y_test, predictions)
print("R2:", r2)
```

R2: 0.9091973905188271

## Porcentaje absoluto del error (MAPE)

El porcentaje absoluto de error (MAPE, por sus siglas en inglés, Mean Absolute Percentage Error) es una métrica utilizada para evaluar el rendimiento de un modelo de regresión al comparar las predicciones con los valores reales. Mide el promedio de los errores absolutos como un porcentaje del valor real.

El MAPE proporciona una medida de la magnitud promedio de los errores como un porcentaje del valor real. Cuanto más bajo sea el MAPE, mejor será el rendimiento del modelo, ya que indica una menor discrepancia entre las predicciones y los valores reales en términos porcentuales.

El MAPE es una métrica comúnmente utilizada para evaluar la precisión de un modelo de regresión en términos de porcentaje absoluto de error en relación con los valores reales.

Se evaluó el rendimiento del modelo de la red neuronal de regresión y se obtuvo el porcentaje absoluto medio del error (MAPE) con un valor del 10.04 % de diferencia o desviación entre los dos valores medidos, lo que indica una menor discrepancia entre las predicciones y los valores reales en términos porcentual obteniendo un mejor rendimiento del modelo.

```
errors = np.abs((predictions.flatten() - y_test) / y_test)
mape = np.mean(errors) * 100
print("MAPE:", mape)
```

```
MAPE: 10.04309503281514
```

## Error absoluto medio (MAE)

El Error Absoluto Medio (MAE, por sus siglas en inglés, Mean Absolute Error) es una métrica utilizada para evaluar el rendimiento de un modelo de regresión al medir la magnitud promedio de los errores entre las predicciones y los valores reales.

El MAE obtenido de nuestro modelo de red neuronal como métrica de evaluación se obtuvo con un valor promedio de 600.27 KW, que es la diferencia promedio absoluta

de las predicciones del modelo y los valores reales de la variable objetivo de la demanda.

```
# Error absoluto medio
predictions = model.predict(X_test)
mae = np.mean(np.abs(predictions.flatten() - y_test))
print('Error absoluto medio:', mae)

3/3 [=====] - 0s 918us/step
Error absoluto medio: 600.2719796061198
```

## **CAPÍTULO V CONCLUSIONES Y RECOMENDACIONES**

En este trabajo, se ha propuesto un modelo computacional de aprendizaje automático, específicamente en el campo del aprendizaje profundo, basado en redes neuronales recurrentes de regresión para la predicción de la demanda de energía eléctrica. Este modelo presenta diversas ventajas significativas en su aplicación.

Durante el desarrollo de este trabajo, se ha observado que los datos recopilados de diferentes fuentes de información relacionadas con las variables del análisis de la energía eléctrica son de vital importancia. Mediante un adecuado proceso de limpieza y preprocesamiento de datos, se ha logrado estructurar un dataset apropiado que permite la extracción de patrones y dependencias secuenciales utilizando el algoritmo implementado en la red neuronal RNN. Esto ha demostrado ser especialmente efectivo para obtener resultados aceptables en términos de predicción.

La utilización de las técnicas de preprocesamiento de datos ha contribuido en gran medida a mejorar la calidad y la confiabilidad de los datos utilizados en el modelo. Estas técnicas permiten eliminar ruido, eliminar valores atípicos y realizar transformaciones necesarias para asegurar que los datos sean adecuados para el entrenamiento de la red neuronal.

En conjunto, el modelo propuesto y el adecuado preprocesamiento de datos han permitido extraer información valiosa y establecer patrones claros en la demanda de energía eléctrica. Esto ha brindado resultados satisfactorios en términos de predicción, lo que demuestra la eficacia y la aplicabilidad del modelo de aprendizaje profundo propuesto en este trabajo. Al utilizar una RNN para predecir la demanda de energía eléctrica se deben tener variables de medición de ciertos factores, como las estaciones del año respecto a la temperatura y los patrones de consumo de los usuarios, entre otras. Las redes neuronales aprenden de esos patrones pasados y pueden utilizarlos para realizar predicciones futuras con una buena precisión.

Python proporciona bibliotecas y frameworks de aprendizaje profundo, como TensorFlow y Keras, que facilita la implementación y entrenamiento de redes neuronales recurrentes. Estas herramientas ofrecen una gran flexibilidad y permiten ajustar los hiperparámetros de la red, como la arquitectura y la función de pérdida, para adaptarse mejor a los datos y mejorar el rendimiento de la predicción.

Al lograr obtener predicciones precisas de la demanda de energía eléctrica, las empresas pueden aprovechar varios beneficios, como la optimización de la generación y distribución de energía, lo que conlleva a una mayor eficiencia energética, una reducción de costos y una mejor planificación de recursos.

No obstante, para futuros trabajos es importante tener en cuenta que la precisión de las predicciones de la red neuronal está estrechamente ligada a la calidad de los datos de entrada obtenidos de las fuentes de información de la demanda de energía eléctrica, así como al diseño adecuado de la red neuronal en sí. Por lo tanto, es fundamental contar con un proceso continuo de entrenamiento y ajuste a medida que se disponga de nuevos datos, con el fin de mantener la precisión de las predicciones a lo largo del tiempo.

Además, es importante destacar que la demanda de energía eléctrica a menudo exhibe patrones estacionales y cambios a largo plazo, como fluctuaciones diarias, estacionales o anuales. En este sentido, los trabajos futuros podrían centrarse en el desarrollo de modelos de redes neuronales recurrentes capaces de capturar y modelar de manera metaheurística estos patrones estacionales y cambios a largo plazo. Esto permitiría una mejora adicional en la precisión de las predicciones y una mejor comprensión de las tendencias y variaciones en la demanda de energía eléctrica.

A largo plazo el trabajo de investigación debe estar enfocado en mejorar la calidad de los datos, realizar ajustes continuos en el modelo de red neuronal y desarrollar enfoques más sofisticados para capturar y modelar los patrones estacionales y cambios a largo plazo en la demanda de energía eléctrica. Estas acciones contribuirán a un mejor aprovechamiento de los beneficios derivados de la

predicción precisa de la demanda eléctrica. A medida que avanza la investigación en el campo de las redes neuronales, se pueden explorar arquitecturas más complejas y avanzadas para mejorar la predicción de la demanda de energía eléctrica. Por ejemplo, las redes neuronales convolucionales recurrentes (CRNN) o las redes neuronales Transformers podrían ser consideradas para aprovechar la capacidad de capturar patrones espaciales y temporales en los datos.

Por lo que se concluye que una red neuronal recurrente de regresión es una herramienta tecnológica sofisticada para la predicción de la demanda de energía eléctrica como medio para optimizar las decisiones futuras de la generación y distribución de energía eléctrica en la región.

Deep Learning para la predicción de la Demanda de Energía Eléctrica

*Deep Learning for the prediction of Electricity Energy Demand*

Adrián Aguilar Tirado<sup>1</sup>, Jacobo Robles Calderón<sup>2</sup>

<sup>1</sup> Instituto Tecnológico Superior de Teziatlán

<sup>2</sup> Instituto Tecnológico Superior de Teziatlán

\*adrian.aguilari@outlook.com

**Resumen:** La variación del comportamiento de la demanda de energía eléctrica en función del aumento de múltiples variables, como los perfiles de carga, el clima, la demanda consumida por usuarios, entre otros, ocasiona que se incremente significativamente el consumo de energía. Por lo que se necesita predecir los valores óptimos de demanda de energía para periodos futuros y para ello se propone una arquitectura computacional compuesta por una red neuronal recurrente de regresión RNN. Se utilizó la minería de datos basado en KDD (Knowledge Discovery in Databases) para estructurar el conjunto de datos de las variables causales de la demanda de energía en la Región Oriente constituida por 6 municipios de los estados de Puebla y Veracruz en un periodo de 5 años recolectados a través de la Información Estadística de Consumos de la Empresa Comisión Federal de Electricidad y la Comisión Nacional del Agua. Se observó que el rendimiento del modelo fue influenciado por la calidad de los datos utilizados para el entrenamiento por lo que fue capaz de capturar patrones complejos de los datos y generar predicciones cercanas de los valores reales, siendo esta una herramienta acertada para la predicción de la demanda de energía eléctrica.

**Palabras clave:** Demanda, Minería de datos, Red Neuronal, Pronósticos.

**Abstract:** The variation in the behavior of the demand for electrical energy as a function of the increase in multiple variables, such as load profiles, weather, demand consumed per user, among others, causes energy consumption to increase significantly. Therefore, it is necessary to predict the optimal values of energy demand for future periods and for this, a computational architecture composed of a recurrent regression neural network RNN is proposed. Data mining based on KDD (Knowledge Discovery in Databases) was used to structure the data set of the causal variables of energy demand in the Eastern Region made up of 6 municipalities in the states of Puebla and Veracruz in a period of 5 years collected through the Statistical Information on Consumption of the Federal Electricity Commission and the National Water Commission. It was observed that the performance of the model was influenced by the quality of the data used for training which is why it was able to capture complex patterns in the data and generate close predictions of the actual values, making it a successful tool for forecasting the demand for electric power.

**Keywords:** Demand, Data mining, Neural Network, Forecasts.

#### 1 Introducción

La demanda máxima de energía eléctrica ocurre principalmente durante la temporada de verano, entre el periodo comprendido de los meses de mayo a septiembre, cuando el calor es más severo en la zona norte y sur de México. Respecto a ello, se incrementa el uso de dispositivos electrónicos de ventilación para mantener la temperatura en niveles más confortables para el ser humano. Lo que

ocasiona una variación del comportamiento de la demanda de energía eléctrica ocasionando que los valores de carga eléctrica se vean en un aumento significativo para el suministro de energía. Por lo que es importante predecir de manera óptima los valores necesarios de energía en un horizonte de tiempo a corto y largo plazo para tener un equilibrio de los perfiles de carga del suministro de energía eléctrica de manera eficiente para los periodos futuros de consumo.

## PRODUCTOS ACADÉMICOS

Como parte importante de esta tesis se realizó la escritura y publicación de un artículo académico cuyos detalles se muestran a continuación.

### *Deep Learning for the prediction of Electricity Energy Demand*

**Autores:** Adrián Aguilar Tirado, Jacobo Robles Calderón.

**Palabras clave:** Demanda, Red Neuronal, Phytón, Pronósticos.

**Presentado en:** Revista ICYTA Ingeniería, Ciencia y Tecnología Aplicada Edición 2023. Tecnológico Nacional de México.

**Abstract:** The variation in the behavior of the demand for electrical energy as a function of the increase in multiple variables, such as load profiles, weather, demand consumed per user, among others, causes energy consumption to increase significantly. Therefore, it is necessary to predict the optimal values of energy demand for future periods and for this, a computational architecture composed of a recurrent regression neural network RNN is proposed. Data mining based on KDD (Knowledge Discovery in Databases) was used to structure the data set of the causal variables of energy demand in the Eastern Region made up of 6 municipalities in the states of Puebla and Veracruz in a period of 5 years collected through the Statistical Information on Consumption of the Federal Electricity Commission and the National Water Commission. It was observed that the performance of the model was influenced by the quality of the data used for training which is why it was able to capture complex patterns in the data and generate close predictions of the actual values, making it a successful tool for forecasting the demand for electric power.

## REFERENCIAS

- Ahsan, M., Mahmud, M., Saha, P., Gupta, K., & Siddique, Z. (2021). Effect of Data Scaling Methods on Machine Learning Algorithms and Model Performance. *Technologies*, 9(3), 52. <https://doi.org/10.3390/technologies9030052>
- Alzubaidi, L., Zhang, J., Humaidi, A. J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., Santamaría, J., Fadhel, M. A., Al-Amidie, M., & Farhan, L. (2021). Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. *Journal of Big Data*, 8(1). <https://doi.org/10.1186/s40537-021-00444-8>
- Bacanin, N., Stoean, C., Zivkovic, M., Rakic, M., Strulak-Wójcikiewicz, R., & Stoean, R. (2023). On the Benefits of Using Metaheuristics in the Hyperparameter Tuning of Deep Learning Models for Energy Load Forecasting. In *Energies* (Vol. 16, Issue 3). MDPI. <https://doi.org/10.3390/en16031434>
- Basheer, S., Bhatia, S., & Sakri, S. B. (2021). Computational Modeling of Dementia Prediction Using Deep Neural Network: Analysis on OASIS Dataset. *IEEE Access*, 9, 42449–42462. <https://doi.org/10.1109/ACCESS.2021.3066213>
- Carrillo Ledesma, A., & Rosas, K. I. G. (n.d.). *Introducción a la Programación*. <http://academicos.fciencias.unam.mx/antoniocarrilloUnacopiadeestetrabajosepuededescargardelapágina:https://sites.google.com/ciencias.unam.mx/acl/en-desarrollo>
- Colombo-Mendoza, L. O., Paredes-Valverde, M. A., Salas-Zárate, M. D. P., & Valencia-García, R. (2022). Internet of Things-Driven Data Mining for Smart Crop Production Prediction in the Peasant Farming Domain. *Applied Sciences (Switzerland)*, 12(4). <https://doi.org/10.3390/app12041940>

- El-Amir, H., & Hamdy, M. (2020). Deep Learning Pipeline. In *Deep Learning Pipeline*. Apress. <https://doi.org/10.1007/978-1-4842-5349-6>
- Francois Chollet. (2017). *Deep Learning with Python* (Manning Publications, Ed.; Second Edition).
- Géron, A. (2019). *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow SECOND EDITION Concepts, Tools, and Techniques to Build Intelligent Systems*. <http://oreilly.com>
- Haykin, S. S., & Haykin, S. S. (2009). *Neural networks and learning machines*. Prentice Hall/Pearson.
- Hewage, P., Trovati, M., Pereira, E., & Behera, A. (2021). Deep learning-based effective fine-grained weather forecasting model. *Pattern Analysis and Applications*, 24(1), 343–366. <https://doi.org/10.1007/s10044-020-00898-1>
- Hurwitz, J. & K. D. (2018). *Machine Learning for Dummies* (Wiley., Ed.; IBM Limited Edition).
- Jogunola, O., Adebisi, B., Van Hoang, K., Tsado, Y., Popoola, S. I., Hammoudeh, M., & Nawaz, R. (2022a). CBLSTM-AE: A Hybrid Deep Learning Framework for Predicting Energy Consumption. *Energies*, 15(3). <https://doi.org/10.3390/en15030810>
- Jogunola, O., Adebisi, B., Van Hoang, K., Tsado, Y., Popoola, S. I., Hammoudeh, M., & Nawaz, R. (2022b). CBLSTM-AE: A Hybrid Deep Learning Framework for Predicting Energy Consumption. *Energies*, 15(3). <https://doi.org/10.3390/en15030810>
- Kelleher, J. D. \_Tierney, B. (2018). *Data science* (The MIT Press).
- Leszek Rutkowski - *Computational Intelligence - Methods and Techniques-Springer (2010)*. (n.d.).
- Liu, A. (2015). *Data Science and Data Scientist*.

- Márquez, A. C., de la Fuente Carmona, A., & Antomarioni, S. (2019). A process to implement an artificial neural network and association rules techniques to improve asset performance and energy efficiency. *Energies*, *12*(18). <https://doi.org/10.3390/en12183454>
- Neu, D. A., Lahann, J., & Fettke, P. (2022). A systematic literature review on state-of-the-art deep learning methods for process prediction. *Artificial Intelligence Review*, *55*(2), 801–827. <https://doi.org/10.1007/s10462-021-09960-8>
- Raschka, S., & Mirjalili, V. (n.d.). *Python machine learning: machine learning and deep learning with python, scikit-learn, and tensorflow 2*.
- Salvendy, G. (2001). *Handbook of industrial engineering: technology and operations management*. Wiley.
- Sarker, I. H. (2021). Deep Learning: A Comprehensive Overview on Techniques, Taxonomy, Applications and Research Directions. In *SN Computer Science* (Vol. 2, Issue 6). Springer. <https://doi.org/10.1007/s42979-021-00815-1>
- Sharda, R., Delen, D., & Turban, E. (n.d.). *Business Intelligence, Analytics, and Data Science: A Managerial Perspective*.
- Shirzadi, N., Nizami, A., Khazen, M., & Nik-Bakht, M. (2021). Medium-term regional electricity load forecasting through machine learning and deep learning. *Designs*, *5*(2). <https://doi.org/10.3390/designs5020027>
- Torres, J. F., Martínez-Álvarez, F., & Troncoso, A. (2022). A deep LSTM network for the Spanish electricity consumption forecasting. *Neural Computing and Applications*, *34*(13), 10533–10545. <https://doi.org/10.1007/s00521-021-06773-2>
- Comisión Federal de Electricidad. (s.f). *Informe Estadístico de Consumos*. Gestión Comercial. <http://cssnal.cfemex.com/gesCom/>
- Comisión Nacional del Agua. (s.f). *Climatología*

Información Climatológica <https://smn.conagua.gob.mx/es/>

Comisión Federal de Electricidad. (s.f). *Sistemas de Análisis de Perfiles Característicos de Carga*. Distribución <http://10.4.22.4:93>