

	GESTIÓN DE SERVICIOS ACADÉMICOS Y BIBLIOTECARIOS		CÓDIGO	FO-GS-15	
			VERSIÓN	02	
	ESQUEMA HOJA DE RESUMEN			FECHA	03/04/2017
				PÁGINA	1 de 1
ELABORÓ		REVISÓ	APROBÓ		
Jefe División de Biblioteca		Equipo Operativo de Calidad	Líder de Calidad		

RESUMEN TRABAJO DE GRADO

AUTOR(ES):

NOMBRE(S): Paola Andrea APELLIDOS: Castro Correa

NOMBRE(S): Sebastian APELLIDOS: Rojas Ortega

FACULTAD: Ingeniería

PLAN DE ESTUDIOS: Ingeniería Electrónica

DIRECTOR:

NOMBRE(S): Sergio Basilio APELLIDOS: Sepúlveda Mora

CODIRECTOR:

NOMBRE(S): Jhon Alejandro APELLIDOS: Castro Correa

TÍTULO DEL TRABAJO DE GRADO: Desarrollo de un modelo híbrido basado en inteligencia artificial para el despacho de energía en una microrred.

Resumen:

El presente trabajo de grado tuvo como fin desarrollar un algoritmo de inteligencia artificial para controlar de manera automática el despacho de energía dentro de una microrred de energías renovables. La investigación planteó una metodología estructurada para dar cumplimiento a los objetivos propuestos en el documento, mediante este trabajo se obtuvo un algoritmo ensamblado que combina árboles de decisión (DT), refuerzo de gradientes extremo (XGBoost) y un regresor lineal múltiple (MLR) para realizar predicciones de la producción de energía solar, producción de energía eólica y demanda eléctrica dentro de una microrred teniendo en cuenta los datos históricos de la serie de tiempo, obteniendo un error porcentual absoluto medio de 5.17%, 5.64% y 0.52% respectivamente. Estas predicciones fueron utilizadas para el desarrollo de un algoritmo que administra la energía dentro de la microrred basándose en una serie de reglas como la profundidad de descarga de las baterías o el recorte de picos en la demanda energética, se analizó el desempeño del sistema en materia económica y obtuvo un error absoluto medio de 0.004USD/kWh con respecto al comportamiento real del sistema y el predicho por el algoritmo.

PALABRAS CLAVES: Microrredes, despacho de energía, inteligencia artificial, redes neuronales.

CARACTERÍSTICAS:

PÁGINAS: 126 PLANOS: 0 ILUSTRACIONES: 73 CD ROOM: NO

ACTA DE SUSTENTACIÓN DE UN TRABAJO DE GRADO

Fecha: CÚCUTA, 24 DE ABRIL DE 2023

Hora: 16:00

Lugar: SALON AG104

Plan de Estudios: INGENIERÍA ELECTRÓNICA

Título del trabajo de grado: "DESARROLLO DE UN MODELO HÍBRIDO BASADO EN INTELIGENCIA ARTIFICIAL PARA EL DESPACHO DE ENERGÍA EN UNA MICRORRED".

Jurados: IE MSc. ORIANA LOPEZ BUSTAMANTE
IE MSc. MARIO JOAQUIN ILLERA BUSTOS

Director: IE. PhD SERGIO SEPÚLVEDA MORA

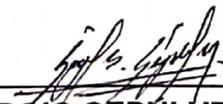
Codirector: IE. MSc JHON CASTRO CORREA.

Nombre del Estudiante:	Código:	Calificación:	
		Número	Letra
PAOLA ANDREA CASTRO CORREA	1161627	5,0	Cinco, cero

LAUREADA



ORIANA LOPEZ BUSTAMANTE

MARIO JOAQUIN ILLERA BUSTOS

SERGIO SEPÚLVEDA MORA
Coordinador Comité Curricular
Ingeniería Electrónica

ACTA DE SUSTENTACIÓN DE UN TRABAJO DE GRADO

Fecha: CÚCUTA, 24 DE ABRIL DE 2023

Hora: 16:00

Lugar: SALON AG104

Plan de Estudios: INGENIERÍA ELECTRÓNICA

Título del trabajo de grado: "DESARROLLO DE UN MODELO HÍBRIDO BASADO EN INTELIGENCIA ARTIFICIAL PARA EL DESPACHO DE ENERGÍA EN UNA MICRORRED".

Jurados: IE MSc. ORIANA LOPEZ BUSTAMANTE
IE MSc. MARIO JOAQUIN ILLERA BUSTOS

Director: IE. PhD SERGIO SEPÚLVEDA MORA

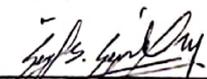
Codirector: IE. MSc JHON CASTRO CORREA.

Nombre del Estudiante:	Código:	Calificación:	
		Número	Letra
SEBASTIAN ROJAS ORTEGA	1161624	5,0	Cinco, cero

LAUREADA


ORIANA LOPEZ BUSTAMANTE


MARIO JOAQUIN ILLERA BUSTOS


SERGIO SEPÚLVEDA MORA
Coordinador Comité Curricular
Ingeniería Electrónica

DESARROLLO DE UN MODELO HÍBRIDO BASADO EN INTELIGENCIA ARTIFICIAL
PARA EL DESPACHO DE ENERGÍA EN UNA MICRORRED

Presentado por:

PAOLA ANDREA CASTRO CORREA

SEBASTIAN ROJAS ORTEGA

UNIVERSIDAD FRANCISCO DE PAULA SANTANDER

FACULTAD DE INGENIERÍA

INGENIERÍA ELECTRÓNICA

SAN JOSÉ DE CÚCUTA

2023

DESARROLLO DE UN MODELO HÍBRIDO BASADO EN INTELIGENCIA ARTIFICIAL
PARA EL DESPACHO DE ENERGÍA EN UNA MICRORRED

Trabajo de grado presentado como requisito para optar por el título de:

INGENIERO ELECTRÓNICO

Presentado por:

PAOLA ANDREA CASTRO CORREA

SEBASTIAN ROJAS ORTEGA

Director:

PhD. SERGIO BASILIO SEPÚLVEDA MORA

Codirector:

MSc. JHON ALEJANDRO CASTRO CORREA

UNIVERSIDAD FRANCISCO DE PAULA SANTANDER

FACULTAD DE INGENIERÍA

INGENIERÍA ELECTRÓNICA

SAN JOSÉ DE CÚCUTA

2023

Dedicatoria

Dedico este logro a mis padres, que confiaron en mí y me han brindado las herramientas necesarias para lograr mis metas.

A mis hermanos, por su apoyo, su compañía y sus valiosos consejos. Mis mejores recuerdos han sido con ustedes.

A todas las personas que me acompañaron en este arduo camino, me orientaron y sobre todo me apoyaron, este logro es para ustedes.

Paola Andrea Castro Correa

Quiero dedicar este trabajo a todas las personas que han influido en mi vida, en primera instancia a mis padres Irit Ortega y Juan Rojas quienes se esforzaron para que yo llegara a este punto; a mis hermanos Angie y Edward que nunca dudaron de mí y me apoyaron todo el tiempo; y también a todos aquellos compañeros, maestros y familiares que me aportaron su conocimiento para llegar a ser una persona profesional. Este no es un logro individual, sino un triunfo en conjunto de todas aquellas personas que se unieron para hacer que esto fuera posible.

Sebastian Rojas Ortega

Agradecimientos

Agradezco primeramente a Dios por haberme dado fortaleza, salud y paciencia para llevar a cabo esta investigación.

Al ingeniero Jhon Castro y al ingeniero Sergio Sepúlveda por su paciencia, dedicación y orientación en el desarrollo de este trabajo de investigación. Su experiencia y sabiduría han sido esenciales para la realización de este proyecto. Son personas admirables.

A mis padres y hermanos por su apoyo incondicional y su constante compañía en este camino académico. Su presencia en mi vida ha sido fundamental para mantenerme motivada.

A Sebastian, que ha sido un gran compañero y es la persona con la que más he colaborado durante el transcurso de mi carrera universitaria.

A mis compañeros con los que he compartido experiencias que han hecho que este proceso sea aún más enriquecedor.

A los ingenieros, quienes han sido parte de mi formación académica y han compartido su conocimiento para guiarme en mi proceso de aprendizaje.

Paola Andrea Castro Correa

De corazón quiero agradecerle en primera instancia a Dios; a mis directores Sergio Sepúlveda y Jhon Castro, quienes nos guiaron durante el proceso; a Paola Castro quien fue mi compañera de proyecto de grado y que juntos logramos todo lo que nos propusimos durante la carrera a nivel profesional; a mis compañeros que fueron una parte clave en este proceso de formación y finalmente a todas aquellas personas que creyeron en mi durante todos estos años, siempre recordaré su apoyo y su esfuerzo incondicional.

Sebastian Rojas Ortega

Tabla de Contenido

	Pág.
Lista de Tablas	IX
Lista de Figuras	XI
Lista de Anexos	XV
Introducción	16
1. Descripción del problema	17
1.1 Planteamiento del problema	17
1.2 Justificación	19
1.2.1 Beneficios tecnológicos.	20
1.2.2 Beneficios económicos.	21
1.2.3 Beneficios sociales.	21
1.2.4 Beneficios institucionales.	22
1.3 Delimitaciones	22
1.3.1 Conceptual.	22
1.3.2 Espacial.	22
1.3.3 Temporal.	23
2. Objetivos	24
2.1 Objetivo general	24
2.2 Objetivos específicos	24
3. Marco referencial	25
3.1 Antecedentes	25
3.1.1 Referentes regionales.	25
3.1.2 Referentes nacionales.	26

3.1.3 Referentes internacionales.	27
3.2 Marco teórico	28
3.2.1 Energías renovables.	28
3.2.1.1 Energía solar fotovoltaica.	29
3.2.1.2 Energía eólica	30
3.2.2 Demanda eléctrica.	31
3.2.3 Microrredes.	32
3.2.3.1 Microrredes on-grid.	32
3.2.4 Series temporales.	33
3.2.5 California Independent Operator (CAISO).	33
3.2.6 National Renewable Energy Laboratory (NREL).	34
3.2.7 Software para simulación de microrredes.	34
3.2.7.1 HOMER Grid.	34
3.2.8 Python para el pronóstico de series temporales.	34
3.2.9 Algoritmos estadísticos.	35
3.2.9.1 Algoritmo SARIMA.	35
3.2.10 Algoritmos predictivos.	36
3.2.10.1 Algoritmo Regresión lineal múltiple (MLR).	37
3.2.10.2 Algoritmo de Árboles de decisión (DT).	38
3.2.10.3 Algoritmo de K-vecinos más cercanos (KNN).	39
3.2.10.4 Algoritmo Extreme Gradient Boosting (XGBoost).	42
3.2.11 Redes neuronales.	43
3.2.11.1 Redes neuronales recurrentes.	45
3.3 Marco legal	46

3.3.1 Ley 603 de 2000: Distribución de licencias y derechos de autor.	47
3.3.2 Ley 1273 de 2009: Manipulación de bases de datos en sistemas informáticos.	47
3.3.3 Ley 23 de 1982: Ley de protección de los derechos de autor.	47
4. Diseño metodológico	48
4.1 Selección y depuración de bases de datos	48
4.1.1 Búsqueda de bases de datos más utilizadas.	48
4.1.2 Comparación y selección de las bases de datos.	50
4.1.3 Adquisición de los datos.	53
4.1.4 Limpieza del conjunto de datos seleccionado.	56
4.2 Exploración y caracterización preliminar de los datos	59
4.2.1 Tendencia de la producción y la demanda eléctrica.	59
4.2.2 Influencia de las estaciones del año.	61
4.2.3 Estacionalidad en la serie temporal.	63
4.2.4 Transformación de los datos.	66
4.3 Identificación de variables de entrada y salida	67
4.3.1 Revisión de variables significativas en antecedentes.	67
4.3.2 Autocorrelación en la serie temporal.	68
4.3.3 Correlación entre variables.	72
4.4 Selección de las herramientas de desarrollo y simulación	73
4.4.1 Lenguaje de programación utilizado en las predicciones.	73
4.4.2 Librerías de aprendizaje automático.	75
4.4.3 Software de simulación de la microrred.	76
4.5 Definición e implementación de los modelos de aprendizaje automático	77
4.5.1 División de los datos.	78

4.5.2 Definición de la lógica de los algoritmos.	80
4.5.3 Métricas utilizadas	81
4.6 Evaluación del desempeño de los modelos en la microrred	83
4.6.1 Funcionamiento de Homer Grid.	83
4.6.2 Energía eólica en Homer Grid.	84
4.6.3 Algoritmo de despacho.	84
5. Resultados	85
5.1 Base de datos para algoritmos de aprendizaje automático	85
5.2 Comparación de los algoritmos utilizados	88
5.3 Integración de los modelos de predicción	96
5.4 Predicción usando redes neuronales	99
5.5 Algoritmo de despacho de energía eléctrica	103
5.6 Evaluación del desempeño del algoritmo de despacho	105
6. Conclusiones	108
7. Recomendaciones	111
Referencias	112
Anexos	120

Lista de Tablas

	Pág.
Tabla 1. Normatividad sobre la creación de algoritmos en Colombia	46
Tabla 2. Características de las bases de datos tenidas en cuenta para la selección final	50
Tabla 3. Matriz de selección de las bases de datos usadas en la realización del proyecto	51
Tabla 4. Pseudo código de la lógica de adquisición de datos	55
Tabla 5. Cantidad de información en la base de datos	58
Tabla 6. Porcentaje total de cobertura de la producción de energía	67
Tabla 7. Variables de entrada para predicción de energía solar	68
Tabla 8. Variables de entrada para predicción de energía eólica	68
Tabla 9. Variables de entrada para predicciones de demanda eléctrica	68
Tabla 10. Definición de variables de entrada	73
Tabla 11. Características de los lenguajes de programación tenidos en cuenta	74
Tabla 12. Matriz de selección de los lenguajes de programación tenidos en cuenta	74
Tabla 13. Algoritmos de aprendizaje automático utilizados en el proyecto	76
Tabla 14. Características de los software de simulación considerados en el proyecto	77
Tabla 15. Matriz de selección de los programas de simulación tenidos en cuenta	77
Tabla 16. Pseudo código de la lógica inicial de programación	80
Tabla 17. Pseudo código de la lógica utilizada en los regresores	81
Tabla 18. Descripción de la base de datos	86
Tabla 19. Propiedades de la base de datos	86
Tabla 20. Análisis inicial de los algoritmos de FV	88
Tabla 21. Análisis de los algoritmos de FV sin variables exógenas	89
Tabla 22. Análisis de los algoritmos de energía eólica sin variables exógenas	89

Tabla 23. Análisis de los algoritmos de demanda eléctrica sin variables exógenas	89
Tabla 24. Análisis de los algoritmos de FV con variables exógenas	90
Tabla 25. Análisis de los algoritmos de energía eólica con variables exógenas	90
Tabla 26. Análisis de los algoritmos de demanda eléctrica con variables exógenas	90
Tabla 27. Hiperparámetros de los modelos predictivos de FV	91
Tabla 28. Hiperparámetros de los modelos predictivos de energía eólica	91
Tabla 29. Hiperparámetros de los modelos predictivos de demanda eléctrica	91
Tabla 30. Métrica MAPE de los modelos AdaBoost combinados	98
Tabla 31. Métricas de error en datos de evaluación para cada red neuronal	102

Lista de Figuras

	Pág.
Figura 1. Tarifas eléctricas finales en Latinoamérica en el año 2020	18
Figura 2. Cobertura eléctrica en Colombia en el año 2020	19
Figura 3. Fuentes de generación de energía desde 1998 hasta 2014	20
Figura 4. Promedio de radiación solar en Colombia	29
Figura 5. Promedio de radiación solar en Cúcuta	30
Figura 6. Países con mayor capacidad instalada	31
Figura 7. Esquema simplificado de una microrred	32
Figura 8. Estructura de una serie temporal	33
Figura 9. Serie temporal con componente estacional	36
Figura 10. Ejemplo de regresión lineal simple	37
Figura 11. Ejemplo de regresión lineal múltiple	38
Figura 12. Estructura de un árbol de decisión	39
Figura 13. Ejemplo KNN: Clasificación	40
Figura 14. Ejemplo KNN: Pesos uniformes	41
Figura 15. Ejemplo KNN: Pesos según la distancia	41
Figura 16. Estructura de XGBoost	42
Figura 17. Esquema de una red neuronal	43
Figura 18. Ecuación de salida de una red neuronal	44
Figura 19. Estructura de red neuronal recurrente	45
Figura 20. Estructura red LSTM	45
Figura 21. Ruta de actividades	48
Figura 22. Localización de las ciudades de los datos de NREL	52

Figura 23. Interfaz disponible en el sitio web de CAISO para producción	53
Figura 24. Interfaz disponible en el sitio web de CAISO para demanda	54
Figura 25. Interfaz disponible en el sitio web de NREL	54
Figura 26. Gráfica de los datos generada con Seaborn sin depurar	56
Figura 27. Gráfica de los datos generada con Seaborn depurada	57
Figura 28. Ejemplo de datos con mediciones erróneas	58
Figura 29. Producción fotovoltaica durante el mes de enero de 2019	59
Figura 30. Producción eólica durante el mes de enero del año 2019	60
Figura 31. Demanda eléctrica durante el mes de enero del año 2019	60
Figura 32. Producción fotovoltaica en 2019 por estaciones	61
Figura 33. Producción eólica en 2019 por estaciones	62
Figura 34. Demanda eléctrica en 2019 por estaciones	63
Figura 35. Descomposición estacional de producción solar	64
Figura 36. Descomposición estacional de producción eólica	64
Figura 37. Descomposición estacional de demanda eléctrica	65
Figura 38. Autocorrelación de producción de energía solar	69
Figura 39. Autocorrelación de producción de energía eólica	69
Figura 40. Autocorrelación de demanda eléctrica	69
Figura 41. Autocorrelación parcial de producción de energía solar	70
Figura 42. Autocorrelación parcial de producción de energía eólica	71
Figura 43. Autocorrelación parcial de demanda eléctrica	71
Figura 44. Mapa de calor de la base de datos	72
Figura 45. División de los datos de producción de energía solar.	78
Figura 46. División de los datos de producción de energía eólica	79

Figura 47. División de los datos de demanda eléctrica	79
Figura 48. Interfaz del software Homer Grid	83
Figura 49. Ruta de resultados obtenidos	85
Figura 50. Cantidad de energía solar producida en MW en 2021	87
Figura 51. Cantidad de energía eólica producida en MW en 2021	87
Figura 52. Cantidad de demanda eléctrica en MW en 2021	87
Figura 53. Predicción de energía solar con SARIMA	92
Figura 54. Predicción de energía solar con DT	92
Figura 55. Predicción de energía solar con XGB	93
Figura 56. Predicción de energía solar con MLR	93
Figura 57. Predicción de energía eólica con SARIMA	94
Figura 58. Predicción de energía eólica con DT	94
Figura 59. Predicción de energía eólica con XGB	94
Figura 60. Predicción de energía eólica con MLR	95
Figura 61. Predicción de demanda eléctrica con SARIMA	95
Figura 62. Predicción de demanda eléctrica con DT	95
Figura 63. Predicción de demanda eléctrica con XGB	96
Figura 64. Predicción de demanda eléctrica con MLR	96
Figura 65. Estructura de integración de los algoritmos	97
Figura 66. Predicción de energía solar con AdaBoost y MLR	98
Figura 67. Predicción de energía eólica con AdaBoost y MLR	98
Figura 68. Predicción de demanda eléctrica con AdaBoost y XGBoost	99
Figura 69. Comparación del MAE en datos de prueba para cada tipo de red neuronal	101
Figura 70. Comparación del MAE en datos de prueba para cada tipo de red neuronal	101

Figura 71. Comparación del RMSE en datos de prueba para cada tipo de red neuronal	102
Figura 72. Estructura de la microrred simulada	103
Figura 73. Despacho de energía en la microrred	104

Lista de Anexos

	Pág.
Anexo 1. Base de datos generada (solo cien datos)	120
Anexo 2. Predicción de energía solar, energía eólica y demanda eléctrica del modelo lineal	122
Anexo 3. Predicción de energía solar, energía eólica y demanda eléctrica del modelo denso	123
Anexo 4. Predicción de energía solar, energía eólica y demanda eléctrica del modelo CNN	124
Anexo 5. Predicción de energía solar, energía eólica y demanda eléctrica del modelo LSTM	125
Anexo 6. Base de datos	126
Anexo 7. Código fuente	126

Introducción

El concepto de microrred para la distribución de energía es relativamente nuevo en comparación a otras tecnologías existentes, y cuyo despacho de energía se basa en modelos implementados con algoritmos de optimización e inteligencia artificial. Por medio de técnicas de aprendizaje automático (machine learning en inglés), es posible conseguir la máxima eficiencia energética dentro de un sistema de despacho de energía o una red de distribución de baja escala con múltiples cargas eléctricas que supla comunidades pequeñas, lo que en esencia, supone el concepto de una microrred [1].

Un sistema híbrido de energía se encarga de generar electricidad a partir de dos o más fuentes de energía de origen renovable o no renovable. Actualmente existen una gran variedad de algoritmos que permiten visualizar el despacho de energía de este tipo de sistemas, con distintas aproximaciones al concepto y cumpliendo con el objetivo de abarcar un aspecto específico del sistema. En este proyecto se desarrolló un modelo híbrido para el despacho de energía en una microrred utilizando múltiples algoritmos de aprendizaje automático [2].

Hoy en día los sistemas de distribución de energía cuentan con distintos tipos de fuentes de alimentación, tales como la energía solar, la energía eólica, generadores de gas natural, entre otras, y que son utilizadas de manera conjunta para suplir la demanda de energía eléctrica en una zona específica. En esta investigación se utilizaron fuentes de energía solar y eólica para entrenar los modelos aprendizaje del sistema. Es por lo anterior, que el algoritmo del modelo híbrido comprendió los patrones en la producción de esas energías teniendo en cuenta variables temporales y estacionales. Adicionalmente, se abordó la predicción de la demanda energética, consiguiendo aprovechar de manera más eficiente la distribución de energía en la microrred [3].

1. Descripción del problema

Desde el año 1974, el consumo eléctrico dentro de Colombia ha tenido un crecimiento promedio anual del 3.4% [4]; esto ha conllevado a la creación de nuevas tecnologías que mejoren la distribución energética y minimicen el impacto en la población que no posee conexión a la red y requiere de ella. Ejemplo de ello son las microrredes, que mediante las energías renovables tiene como uno de sus objetivos generar un suministro eléctrico para las comunidades que tienen acceso limitado a la red eléctrica del país, esta situación se describe de manera más detallada y profunda en las siguientes secciones.

1.1 Planteamiento del problema

En el mundo existe una problemática constante y es la contaminación, que provoca cambios en el entorno generando un desequilibrio medio ambiental que no solo afecta seres vivos sino también ecosistemas, pues el hombre ha venido haciendo uso de energías no renovables que ocasionan, en muchos casos, deterioros irreversibles al medio ambiente [5].

La calidad del aire en algunas regiones de Colombia es deplorable, dejando 10.527 muertos y más de 6.7 millones de personas con efectos negativos en su salud desde el año 2015 por culpa de los efectos de la polución. Lo anterior, no solo ocasionó pérdida de vidas sino costos ambientales por la contaminación atmosférica dados por el uso de energías fósiles [6]. Por tales motivos, se han tomado medidas preventivas para mitigar el impacto ambiental, que, sin embargo, no han sido suficientes para contrarrestar los daños ocasionados al medio ambiente. Lo anterior conllevó a desarrollar nuevas alternativas para reducir los efectos en el ambiente y contrarrestar fallos energéticos derivados cuando las fuentes de energías fósiles se agoten.

Actualmente, Colombia cuenta con fuentes de energía con precios asequibles como hidroeléctricas y termoeléctricas que funcionan a base de agua y carbón, respectivamente. Sin embargo, en ciertos meses del año, estas alternativas no son suficientes para cubrir la demanda

energética, haciendo que el país deba recurrir a los precios de la bolsa de energía cuyo resultado es el incremento del precio de la electricidad [7]. En la figura 1, se presenta el precio de megavatio-hora de los países en Latinoamérica. Realizando un análisis de la gráfica, se evidencia que Colombia es uno de los países latinoamericanos que tiene el valor de la electricidad más elevado, estando por encima de países de mayor envergadura territorial como México, Brasil o Argentina.

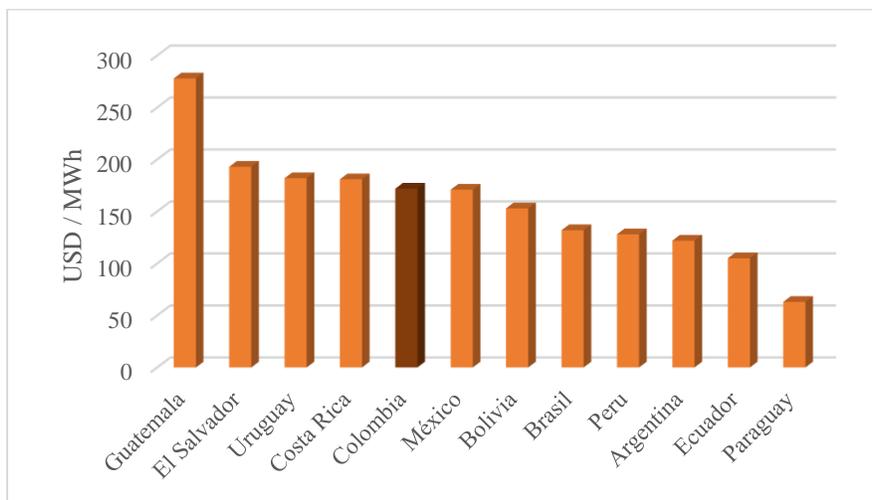


Figura 1. Tarifas eléctricas finales en Latinoamérica en el año 2020. Fuente: Adaptado de [7].

La UPME (Unidad de Planeación Minero Energética) propuso distintas consideraciones para dar solución a la problemática de abastecimiento de energía, dentro de las cuales se encuentra la aplicación de energías renovables como la fotovoltaica, lo que implica indirectamente a una mayor cobertura para toda la región [8]. En la figura 2, se aprecian algunos de los departamentos del territorio nacional y el porcentaje de la población que cuenta con el servicio de electricidad. El análisis de la gráfica conllevó a preocupantes hallazgos, pues en departamentos como Amazonas, Vaupés y Vichada cerca del 30% de los habitantes no cuentan con el servicio debido a que hay cortes de electricidad diarios que provocan que las personas dispongan solo de 8 horas de energía, lo que los afecta cotidianamente.

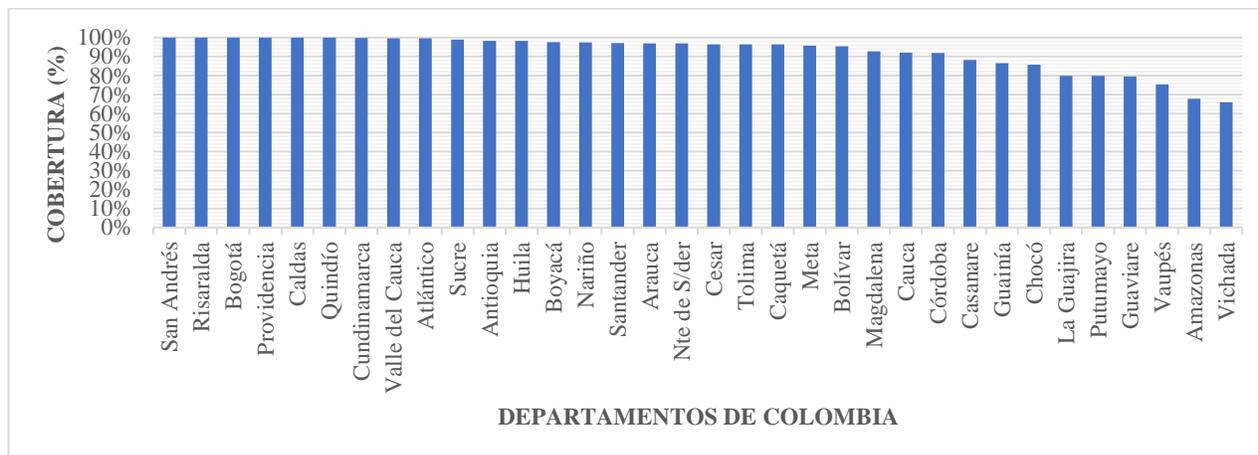


Figura 2. Cobertura eléctrica en Colombia en el año 2020. Fuente: Adaptado de [9].

Debido a su naturaleza, las microrredes no son afectadas de forma directa por cortes de electricidad, dado a que son sistemas descentralizados y son capaces de seguir funcionando debido a las fuentes de energía renovable que poseen, ya que distribuyen el recurso en el momento indicado cuando exista una alta demanda de energía. Es por tal motivo que se propuso la siguiente pregunta: ¿Cómo diseñar un modelo híbrido para predecir el despacho de energía en una microrred que permita aprovechar de forma eficaz los recursos disponibles?

1.2 Justificación

Colombia se ha encaminado al futuro de las tecnologías de la información (TIC) y a la inversión de capital en materia de microrredes basadas en energías renovables, debido al impacto medioambiental producido por las fuentes de energía convencionales. Un estudio realizado en el año 2016 demostró que Colombia, desde el año 1998 hasta el año 2014, produce energía mayormente de fuentes hidroeléctricas, que causan daños considerables al medio ambiente por la desestabilización de los ecosistemas de flora y fauna. Asimismo, varias de las fuentes de energía descritas en el estudio también son contaminantes, siendo las energías renovables las menos utilizadas, como se observa en la figura 3 [10].

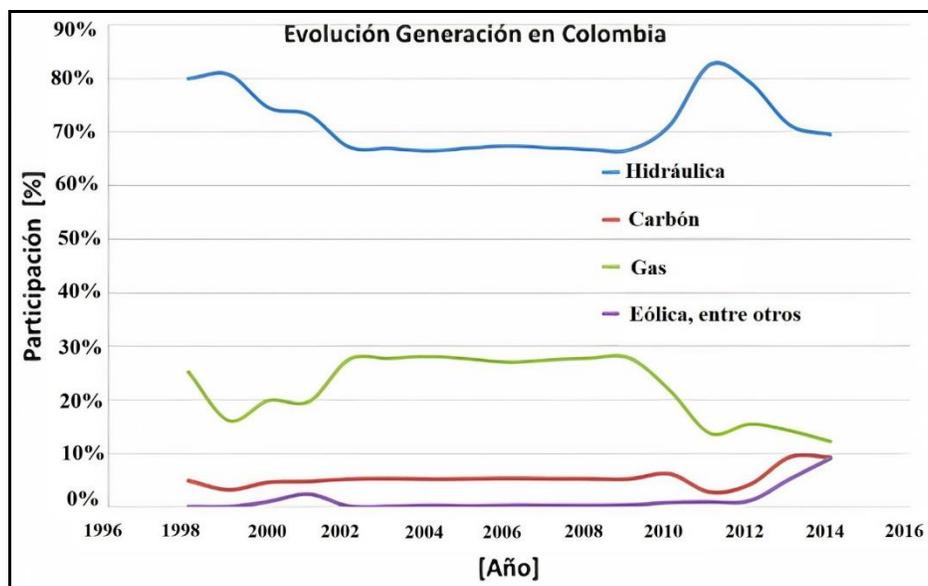


Figura 3. Fuentes de generación de energía desde 1998 hasta 2014. Fuente: Tomado de [10].

Como se observa en la figura 3, las energías renovables hace unos años representaban menos del 5% del total de energía generada en todo el país. En Colombia a la energía solar y eólica se le denomina como FNCER (Fuentes no Convencionales de Energía Renovable). A partir de 2012, el número de proyectos relacionados con las energías limpias comenzó a aumentar. La implementación de microrredes en la actualidad podría aportar al aumento de este índice, puesto que como se observa en la figura 3, la tendencia en el uso de energías renovables ha tenido un incremento lineal los últimos años.

1.2.1 Beneficios tecnológicos. Los resultados de esta investigación fomentan la creación de sistemas de despacho energético con microrredes en Colombia, puesto que el modelo híbrido de algoritmos de inteligencia artificial desarrollado podrá ser aplicado a la distribución de energía en redes eléctricas dadas las necesidades del usuario, esto sentará las bases para la innovación de nuevas tecnologías en esta área.

Se obtuvo un modelo híbrido de inteligencia artificial que combina distintos algoritmos y permite disminuir las pérdidas de energía de una microrred, esto podría tener un impacto a futuro en la sostenibilidad y rentabilidad de las microrredes en Colombia.

1.2.2 Beneficios económicos. Las microrredes eléctricas pueden ser implementadas a gran escala debido a su naturaleza modular. Por lo cual, el sistema híbrido para el despacho energético desarrollado puede reducir el precio de la energía eléctrica debido a una distribución eficiente y el uso de energías alternativas, lo que beneficia directamente a las empresas [11].

Según la nueva normativa publicada por la CREG (Comisión de regulación de energía y gas), una persona que posee una microrred de energía renovable puede disminuir su costo de facturación mensual si esta proporciona a la red más energía de la que consume. Por lo que un algoritmo que estime la producción de energía le permitirá al usuario generar un mayor retorno de inversión [12].

1.2.3 Beneficios sociales. A nivel social, la implementación del proyecto conllevará a una mejora en la calidad de vida de la población que hace uso de las energías renovables, puesto que en Colombia el uso de los combustibles fósiles ha repercutido en la salud de las personas, debido a que el aire contaminado es dispersado por los vientos a sectores habitados por personas mayores o niños [13].

El proyecto impulsa el desarrollo de más tecnologías allegadas a las energías renovables, que buscan mejorar la calidad de los sistemas de distribución de energía en la región. En última instancia, el objetivo es mejorar la calidad de los sistemas de distribución de energía en la región, reducir la dependencia de los combustibles fósiles y promover un futuro más sostenible para Colombia.

1.2.4 Beneficios institucionales. La Universidad Francisco de Paula Santander se verá beneficiada gracias a que la investigación que se realizó servirá de base para proyectos posteriores que se ejecuten en materia de las microrredes y las energías renovables en Colombia, puesto que concuerda con el marco del programa de ingeniería electrónica dentro de la universidad.

El algoritmo híbrido para una mejor distribución energética podría implementarse en futuros proyectos que se lleven a cabo para implementar microrredes dentro de la universidad, de modo que haya una manera automatizar el sistema para un despacho de energía más eficiente.

1.3 Delimitaciones

Aunque el proyecto estuvo enfocado en el despacho de energía en una microrred, el fin de este no fue diseñar ni implementar una, sino que estuvo orientado al diseño de un modelo híbrido que utiliza técnicas de aprendizaje automático para generar predicciones para la toma de decisiones en el despacho de energía teniendo en cuenta datos históricos de libre acceso.

1.3.1 Conceptual. El propósito de esta investigación fue diseñar un modelo híbrido capaz de predecir la producción y la demanda energética dentro de una microrred, para ello, se tomaron datos de acceso público para el desarrollo y entrenamiento del algoritmo; con lo cual, no fue necesaria la implementación en una microrred real para las pruebas del modelo.

1.3.2 Espacial. El modelo híbrido podrá ser utilizado con datos de cualquier parte del mundo. Sin embargo, para el modelamiento del algoritmo se tomaron datos que son de acceso público de otros países, ya que en Colombia no se encuentran disponibles este tipo de datos, debido a que es información privada que las entidades no comparten abiertamente.

1.3.3 Temporal. Este proyecto de investigación tuvo una duración estimada de 12 meses para la culminación total de todos los objetivos, lo que incluye la revisión de bibliografía, desarrollo de los algoritmos y evaluaciones de desempeño, para lo cual se destinaron 20 horas de trabajo semanales por parte de cada uno de los autores.

2. Objetivos

A continuación, se presentan los respectivos objetivos que se plantearon para solucionar sistemáticamente la problemática planteada anteriormente.

2.1 Objetivo general

Proponer y desarrollar un modelo híbrido basado en aprendizaje automático para la estimación de producción, consumo y despacho de energía eléctrica en una microrred de energía renovable.

2.2 Objetivos específicos

- Seleccionar y depurar el conjunto de datos para elegir el modelo para el funcionamiento de la microrred.
- Realizar una exploración y caracterización preliminar de los datos que se seleccionaron para el modelo.
- Identificar las variables de entrada y salida que deberán tener los modelos que se escojan para el funcionamiento de la microrred.
- Escoger las herramientas de desarrollo y simulación necesarias para estimar el despacho de energía eléctrica.
- Definir e implementar los modelos para la estimación de producción, consumo y despacho de energía eléctrica en una microrred.
- Evaluar el desempeño de los modelos en materia de despacho de energía dentro del software de simulación.

3. Marco referencial

3.1 Antecedentes

En la actualidad, las microrredes se han vuelto una tecnología recurrente en la industria, así como también los algoritmos que se utilizan para administrar el despacho de energía de la red. Tales algoritmos han ido mejorando con el paso del tiempo, por lo que fue importante estudiar los avances a nivel nacional e internacional. Para este estudio, se analizaron artículos relevantes en el tema, entre ellos se encuentran dos a nivel regional, dos a nivel nacional y cuatro a nivel internacional que permitieron tener un panorama más amplio de cómo funcionan las microrredes y su aplicación para el despacho de energía.

3.1.1 Referentes regionales.

Análisis de imágenes satelitales de observación de la Tierra y datos geoespaciales a través de machine learning. Utilizando machine learning se desarrolló un modelo que permitió reconocer y clasificar las cubiertas vegetales en distintas ciudades de la región, tomando como uno de los ejemplos a San José de Cúcuta, en donde se evaluó mediante el modelo el porcentaje de vegetación que existía en los parques de la ciudad. Este referente aportó ideas sobre cómo manipular las bases de datos para obtener los resultados esperados en la investigación, así mismo sobre cómo comparar los datos de entrenamiento para establecer los patrones significativos de los conjuntos de datos [14].

Análisis del recurso eólico disponible en la ciudad de Cúcuta usando aprendizaje automático. La investigación evaluó cuáles eran las variables más influyentes del estudio, se determinó que en la ciudad de Cúcuta los datos más importantes del ambiente son el viento y la temperatura, por lo que mediante la aplicación de un algoritmo de aprendizaje automático fue posible crear un modelo para predecir a corto plazo la cantidad de recurso eólico con el que

contaría cada zona de la ciudad, utilizando el método de retro propagación junto con un modelo de validación cruzada para evaluar los resultados del algoritmo [15].

3.1.2 Referentes nacionales.

Microrredes aisladas en la Guajira: Diseño e implementación. Se llevó a cabo un estudio de una comunidad Wayuu que carecía de un suministro de energía estable. Mediante la creación de microrredes fue posible hacer uso de la energía solar y eólica, permitiéndole a la comunidad contar con energía renovable para realizar sus necesidades básicas. Este es un claro ejemplo de que esta tecnología permite evolucionar el estilo de vida, pero a su vez evidencia que se requiere el apoyo de todos los involucrados que utilizan la microrred para que el sistema se mantenga sostenible para los usuarios dentro de él, puesto que sin la colaboración de los individuos no hubiera sido posible la instalación de todos los instrumentos y maquinaria necesaria para la creación del sistema [16].

Una metodología de diseño de micro redes para zonas interconectadas de Colombia. Siguiendo la normativa de instalaciones eléctricas de Colombia se diseñó una microrred para una región ubicada en una zona no interconectada, para esto, se presentó el caso de estudio que contenía los reportes meteorológicos respectivos al Taroa. Con esta información se diseñó la topología del esquema eléctrico para utilizar energías renovables en el suministro eléctrico de la región. A partir del software computacional HOMER fue posible estimar las tensiones y los flujos de potencia en la microrred, la cual tenía una variación del 3% de la tensión nominal, lo cual se encuentra dentro del rango estipulado por el reglamento técnico de instalaciones eléctricas del país. Esto es de vital importancia dentro del territorio Colombiano puesto que así se puede llevar una regulación de la construcción de cualquier esquema eléctrico que se ejecute en el país [17].

3.1.3 Referentes internacionales.

Optimal scheduling of energy storage under forecast uncertainties. En este artículo no se estudiaron las baterías que componen el sistema, sino que se desarrolló un algoritmo para predecir con cierta exactitud la manera en la que la energía fotovoltaica fluye a través de la microrred, a la vez de que se implementó un sistema de seguimiento que permite minimizar los errores del algoritmo. Además de esto, hace un corte de picos de voltaje más allá de los soportados por el sistema ya que esto reduce las desviaciones de voltaje a través de las cargas, permitiendo al sistema ser más estable [18].

Microgrids: A review of technologies, key drivers, and outstanding issues. Los autores demuestran cómo las microrredes serán el futuro de la distribución tecnológica, donde las personas pasarán a un plano donde la energía que generan en sus casas ya sea fotovoltaica o eólica, se vuelva de dominio público, y pueda exportarse e importarse para suplir la demanda de la microrred. Lo anterior, mencionado en el año 2018, y ahora 5 años después se ha visto un gran incremento en la cantidad de investigación que ha habido en este campo, puesto que aún es una tecnología muy nueva y hay muchos ámbitos donde aún hay que innovar y buscar las mejores maneras para su implementación. Dicho tipo de trabajos dan paso a que nuevas investigaciones continúen surgiendo y formando las bases de esta tecnología [19].

Design of a resilient and eco-friendly microgrid for a commercial building. La problemática que se describe en el artículo hace referencia a los desastres creados por el hombre y los que son causados por la naturaleza, como son los huracanes. Todo esto, ya que este tipo de fenómenos naturales provocan disturbios en las redes eléctricas ubicadas en las ciudades. Por lo cual, se propuso una microrred capaz de obtener energía solar y almacenarla en baterías de litio. Se observó que en un principio es un costo elevado, pero a largo plazo es una buena opción sostenible para los hogares además de ser una solución ecológica. En el proyecto se hizo uso del software

HOMER Grid para la simulación de una microrred con cargas y generadores de energía renovable para llevar a cabo el estudio del funcionamiento del algoritmo de distribución [20].

Making the case for time-of-use electric rates to boost the value of battery storage in commercial buildings with grid connected PV systems. Se realizó un riguroso estudio con 25 años de datos, que demostró la relación entre el costo y beneficio que tiene dejar degradar los paquetes de baterías más allá de su límite tradicional (establecido como el 30%). En el estudio se detalla que cuando la degradación es del 30% de su capacidad, las baterías deben reemplazarse dos veces cada 25 años. Asimismo, se realizó un estudio cuando la degradación es del 50% de su capacidad y se concluyó que para este caso las baterías debían cambiarse solo una vez en 25 años. Por último, se investigó un caso extremo en el que la degradación era del 90% de su capacidad y las baterías no debían cambiarse en 25 años, en el cual el sistema comenzaba a presentar fallos debido a que se afectaba a la funcionalidad. Los resultados del estudio permitieron establecer que el punto óptimo en cuanto a funcionamiento y economía es dejar degradar las baterías hasta el 50%, puesto que para este caso el costo de inversión se redujo en un 30% [21].

3.2 Marco teórico

Dentro de esta sección se evidencian todos los fundamentos necesarios para sustentar el desarrollo de este proyecto y que poseen relación con el área de las microrredes y de algoritmos de inteligencia artificial.

3.2.1 Energías renovables. El concepto de energías renovables radica en el aprovechamiento de ciertos recursos de la naturaleza que no contribuyen a la contaminación de estos, como, por ejemplo, la energía hidráulica, la eólica o la fotovoltaica. La implementación de estas tecnologías en la actualidad es lo que le ha permitido al país obtener distintos beneficios en el área del crecimiento económico o industrial, tanto a nivel regional como nacional [22].

El impacto ambiental que generan estas tecnologías es sumamente leve en comparación a los métodos convencionales que se han venido utilizando durante décadas, es por esto que el país ha implementado distintas normativas que regulan a las empresas que trabajan con estas tecnologías, ya que la mayoría de los recursos naturales del país se encuentran en zonas consideradas como reservas naturales o de propiedad indígena [23].

3.2.1.1 Energía solar fotovoltaica. Este tipo de energía es generada mediante el aprovechamiento de la radiación solar sobre el territorio nacional. Por su parte, un sistema fotovoltaico conectado a la red produce energía durante el día; de noche, cuando no hay luz solar, la carga se alimenta de la red eléctrica. En la figura 4 se evidencian los lugares que más reciben radiación solar en Colombia, de modo que mientras más al norte se encuentre, más energía eléctrica (medida en kilovatio-hora, kWh) generan los paneles solares que se instalen; esto también depende del tamaño de la instalación y de otros factores que puedan influir en la recolección y almacenamiento de la energía [24].

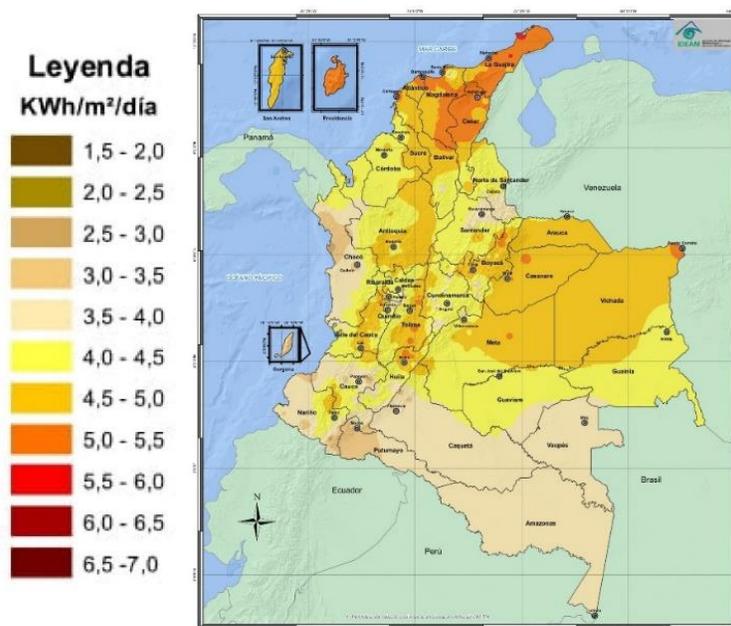


Figura 4. Promedio de radiación solar en Colombia. Fuente: Tomado de [24].

De manera más específica, dentro del municipio de Cúcuta existe un promedio de radiación lo suficientemente alto como para ser aprovechado por sistemas fotovoltaicos. En la figura 5 se muestran los Wh (vatios-hora) que pueden aprovecharse al implementar tecnologías de recolección de energía fotovoltaica en la capital de Norte de Santander.

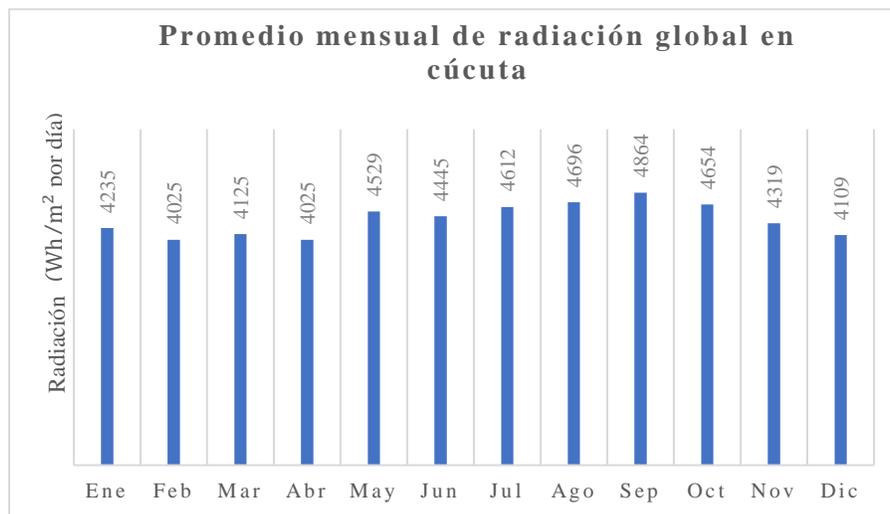


Figura 5. Promedio de radiación solar en Cúcuta. Fuente: Adaptado de [24].

3.2.1.2 Energía eólica. La energía eólica es una de las energías renovables más conocidas, dado a que es producida por las corrientes de aire y es posible convertirla en energía eléctrica por medio de un generador. La energía eólica es una fuente de energía limpia y renovable, que no emite gases de efecto invernadero ni otros contaminantes, lo que la hace importante para combatir el cambio climático y reducir la dependencia de combustibles fósiles.

La energía eólica se ha convertido en una fuente de energía cada vez más importante en todo el mundo. El reporte de 2021 del Consejo Global de la Energía Eólica (GWEC) indica que los países que más producen energía eólica en el mundo son China, Italia, Alemania, India y España, debido a que cuentan con una mayor capacidad instalada, tal y como muestra el gráfico en la figura 6. Esto permite a dichos países reducir su dependencia de combustibles fósiles y reducir sus emisiones de gases de efecto invernadero [25].

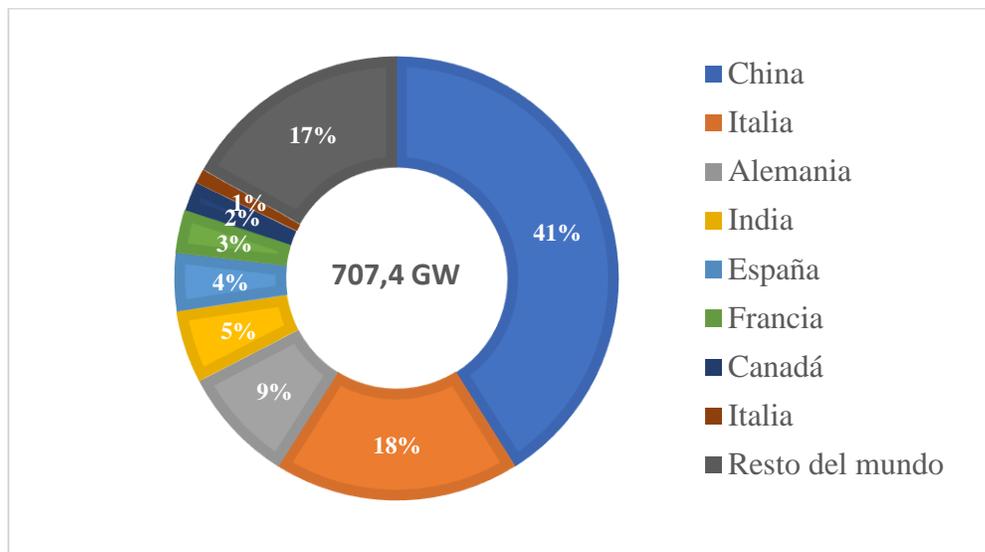


Figura 6. Países con mayor capacidad instalada. Fuente: Adaptado de [25]

3.2.2 Demanda eléctrica. La demanda se refiere a toda la cantidad de energía que los consumidores de la red solicitan en un momento en específico, este valor cambia a lo largo del día y según la época del año, su tendencia depende de variables como el clima, la estación o la hora del día. Este valor de demanda es el que satisface directamente a toda la red o en el caso de una investigación, a una microrred.

Su comportamiento habitual consta de un gráfico bimodal, que presenta un primer pico en las horas de la mañana y otro pico en horas de la noche, esto está relacionado a las costumbres de los consumidores de la red y a su rutina, este patrón siempre es el mismo a lo largo del año, lo que varía en gran medida cada día es la magnitud de ambos picos de la gráfica [26].

3.2.3 Microrredes. Una microrred se considera como una red de energía local que cuenta con un control autónomo, y pueden contener hogares, edificios empresariales, entre otros. Las microrredes pueden estar o no conectadas a la red eléctrica, ya que tienen la posibilidad de trabajar de manera independiente puesto que poseen generadores de energía locales, mayormente de energías renovables como la fotovoltaica o la eólica [27].

3.2.3.1 Microrredes on-grid. Este tipo de microrredes se definen como aquellas que están conectadas a la red eléctrica, en este caso puede poseer fuentes de energía renovable y baterías, por lo que es posible aplicar técnicas que permitan una distribución más eficiente de la energía en la red. En la figura 7 se presentan los componentes de una microrred sencilla teniendo en cuenta lo mencionado anteriormente.

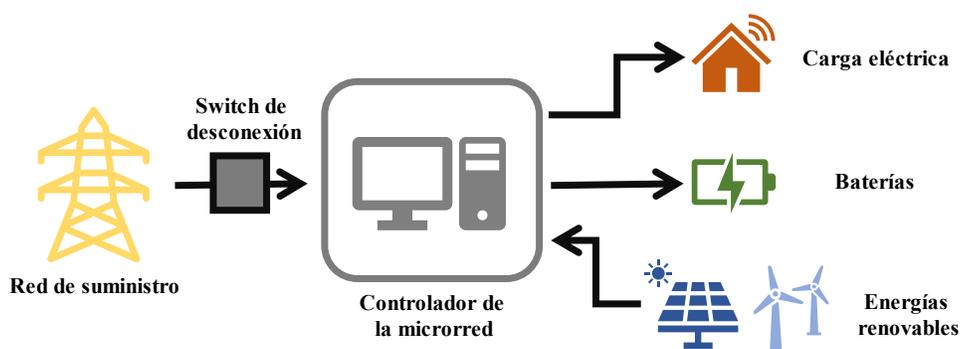


Figura 7. Esquema simplificado de una microrred. Fuente: Elaboración propia

Como se observa en la figura 7, en términos sencillos una microrred tiene la capacidad de acoplarse y desacoplarse de la red de suministro nacional según lo requiera, como por ejemplo en un corte de energía sería capaz de funcionar por sí misma sin afectar los demás componentes que conforman la microrred. Existen variedad de formas de crear una microrred, sin embargo, todas comparten el principio de control de la energía dentro de una localidad en específico [28].

3.2.4 Series temporales. Se define como un conjunto de distintas observaciones ordenadas en intervalos de tiempo que pueden ser regulares o no. Las series temporales son utilizadas en áreas de estadística, ingeniería e inteligencia artificial para analizar el comportamiento de una variable y así poder realizar predicciones basándose en estos patrones de comportamiento [29]. En la figura 8 se evidencia la estructura de una serie temporal, la cual consiste en datos ordenados de manera secuencial según las fechas de la base de datos que se esté utilizando.

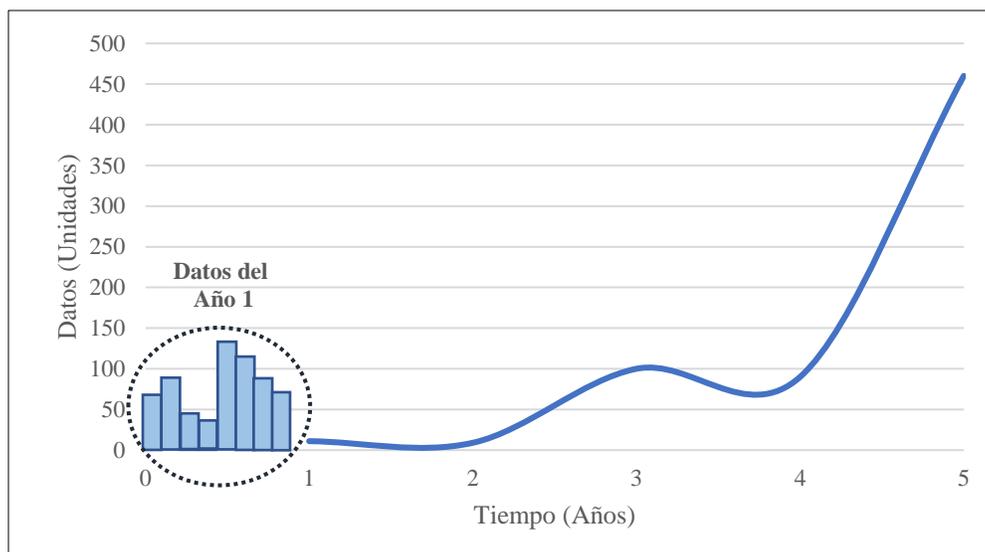


Figura 8. Estructura de una serie temporal. Fuente: Elaboración propia

3.2.5 California Independent Operator (CAISO). Es la organización encargada de mantener la confiabilidad de una de las redes más grandes del mundo, esta entidad recolecta y publica datos temporales en intervalos de cinco minutos y permite su libre acceso, por lo cual es posible realizar análisis y predicciones sobre estos datos. Cabe resaltar que la información que CAISO provee está vinculada a todo el estado de California, por lo que no es posible determinar fácilmente el comportamiento de consumo y producción de energía de una ciudad en específico [30].

3.2.6 National Renewable Energy Laboratory (NREL). Esta entidad es un laboratorio de investigación y desarrollo de energías renovables de Estados Unidos, cuyo objetivo es desarrollar proyectos destinados a mejorar la eficiencia energética y a disminuir la dependencia de los combustibles fósiles. Este laboratorio recopila datos en diferentes intervalos de tiempo utilizando sensores satelitales para recolectar variables como la radiación y la temperatura en distintas partes del mundo [31].

3.2.7 Software para simulación de microrredes. Existen diversos software que permiten la modelación y simulación de microrredes, por lo tanto, fue necesario utilizar uno que cumpliera las necesidades del proyecto de investigación, de modo que se pudiera analizar la distribución de la energía de forma eficaz.

3.2.7.1 HOMER Grid. HOMER Energy es una empresa dedicada al desarrollo y comercialización de software multipropósito que analiza y modela microrredes de energías renovables. En específico, para esta investigación se abordó HOMER Grid, el cual permite simular un sistema eléctrico descentralizado que combina diferentes fuentes de energía renovable y su respectivo almacenamiento. Este programa cuenta con un controlador que posee un algoritmo de despacho que implementa técnicas de optimización energética, la cual determina cuál es la manera más eficiente de aprovechar cada uno de los recursos del sistema [32]. HOMER Grid permitió realizar un análisis y optimización de los diseños propuestos durante el desarrollo del proyecto.

3.2.8 Python para el pronóstico de series temporales. Python permite implementar aplicaciones con redes neuronales y/o algoritmos genéticos gracias a las múltiples librerías especializadas que la comunidad ha desarrollado, por lo que es utilizado principalmente en el área de inteligencia artificial y optimización [33].

Python posee una gran cantidad de librerías orientadas al análisis y pronóstico de series temporales, por ejemplo: Pandas, Numpy, Scikit-learn, TensorFlow, entre otras. Estas librerías

ofrecen herramientas para realizar un análisis exploratorio de los datos, visualizar modelos y realizar predicciones en el ámbito de ciencia de datos [34].

3.2.9 Algoritmos estadísticos. Un algoritmo de este tipo ejecuta un procedimiento computacional que utiliza técnicas de estadística para analizar datos de series temporales y obtener información que se considera útil para su pronóstico, estos algoritmos son utilizados en áreas de inteligencia artificial ya que buscan relaciones entre cada una de las variables para predecir el resultado futuro de una de ellas, existen distintos tipos de algoritmo de este tipo, tales como Mean-Shift, Fuzzy C-means o SARIMA, los cuales poseen distintos enfoques dentro de las técnicas de estadística que utilizan cada uno de ellos [35].

3.2.9.1 Algoritmo SARIMA. El algoritmo “Seasonal Autoregressive Integrated Moving Average), mejor conocido como SARIMA se define como un modelo que evalúa series temporales. Es la extensión de un modelo ARIMA, ya que en este caso considera la estacionalidad que ocurre en la serie de tiempo, este algoritmo es utilizado para modelar y predecir series que poseen un patrón debido a la estación.

El modelo SARIMA posee tres parámetros principales en total, los cuales definen el comportamiento del modelo, estos son:

- El orden de la parte autorregresiva (p)
- El orden de la parte integrada (d)
- El orden de la parte de media móvil (q)

Asimismo, SARIMA incorpora tres parámetros adicionales que se relacionan con la parte estacional del sistema, los cuales son:

- El orden de la parte autorregresiva estacional (P)
- El orden de la parte integrada estacional (D)

- El orden de la parte de media móvil estacional (Q)

De esta forma, utilizando los datos estacionales es posible determinar los seis parámetros requeridos por el modelo para realizar predicciones de datos futuros en la serie, según los antecedentes, SARIMA supera a ARIMA en precisión cuando los datos poseen estas características estacionales [36]. Un ejemplo de un conjunto de datos compatible con este modelo es la serie temporal presentada en la figura 9, donde se evidencia un patrón claramente estacional.

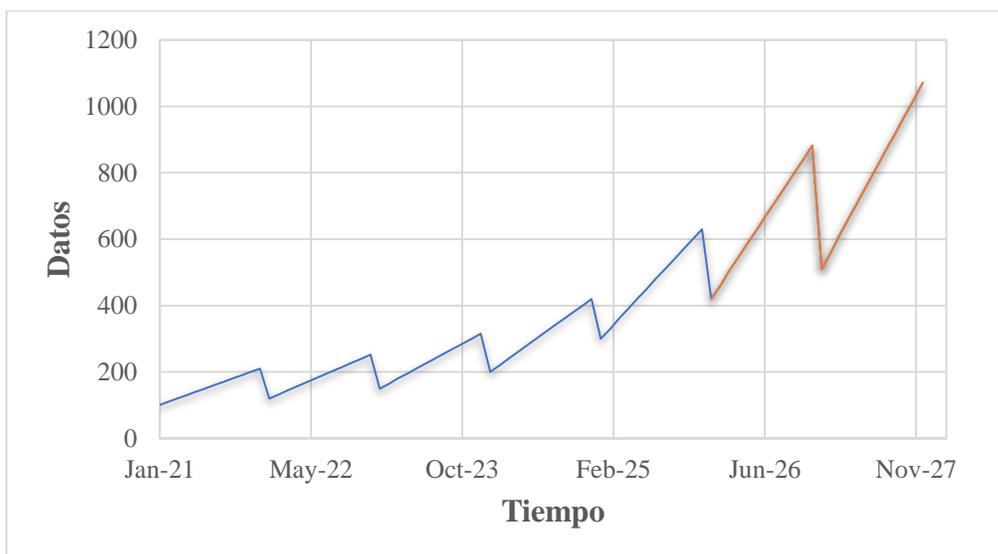


Figura 9. Serie temporal con componente estacional. Fuente: Elaboración propia

3.2.10 Algoritmos predictivos. Son algoritmos diseñados especialmente para analizar datos y realizar predicciones sobre eventos futuros dentro de la misma serie temporal, consisten en la implementación de técnicas de aprendizaje automático para “aprender” y reconocer patrones de los datos históricos de la serie.

El ejemplo más común son los algoritmos de regresión, los cuales utilizan las variables independientes para predecir el valor de una variable dependiente, otros ejemplos incluyen a los árboles de decisión o las redes neuronales artificiales [37].

3.2.10.1 Algoritmo Regresión lineal múltiple (MLR). En primera instancia se define la regresión lineal, la cual es una técnica que analiza la relación entre dos variables que se consideran continuas, una dependiente y una independiente. Un algoritmo de regresión se encarga de estimar la línea que posee el mejor ajuste entre los puntos de los datos y utiliza esa línea para realizar predicciones sobre la variable dependiente. En otras palabras, el algoritmo aprende la tendencia que define los valores futuros que tendrá la variable dependiente. En la figura 10 se muestra un ejemplo de datos donde se aplica una línea de tendencia que posee la forma $y = b_0 + b_1 * x_1$ [38].

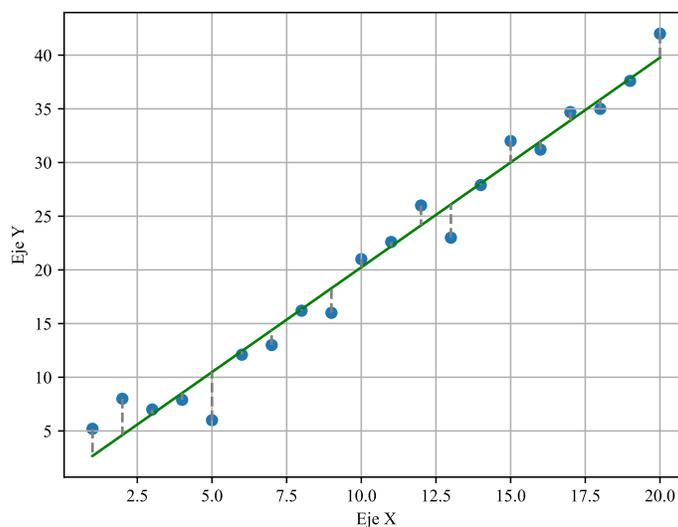


Figura 10. Ejemplo de regresión lineal simple. Fuente: Elaboración propia

Seguidamente, la definición de regresión lineal múltiple (MLR) engloba a todas aquellas técnicas utilizadas para predecir aquella relación que existe entre una variable dependiente y múltiples variables independientes que se relacionan entre sí. En principio se asume que esta relación es lineal y el objetivo es encontrar la mejor combinación de dichos valores que minimizan la distancia entre las predicciones de un modelo y sus valores reales; es decir, que mientras más variables independientes se añadan al modelo, más dimensiones existirán y más complejo será el

cálculo de dicha tendencia [39]. En la figura 11 se muestra cómo se vería el plano de tendencia de los datos de la figura 10 de la forma $y = b_0 + b_1 * x_1 + b_2 * x_2 + \dots + b_n * x_n$.

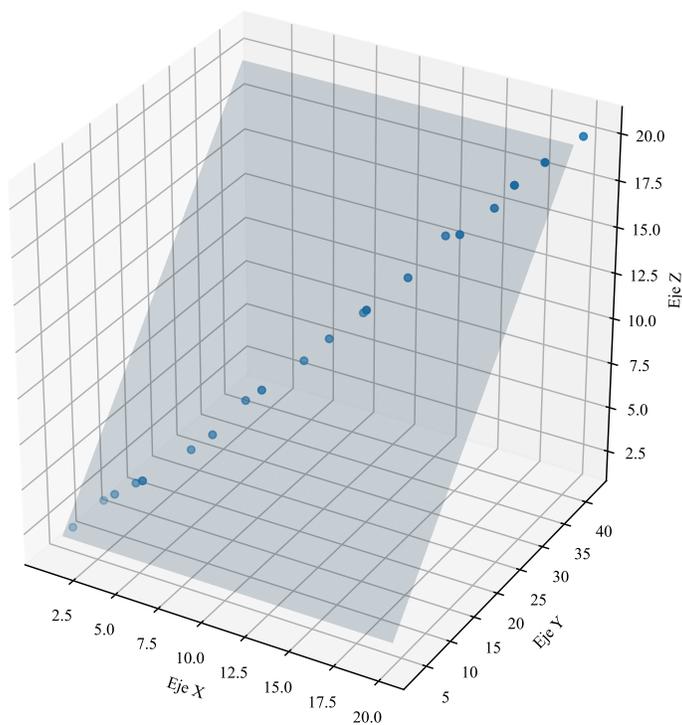


Figura 11. Ejemplo de regresión lineal múltiple. Fuente: Elaboración propia

3.2.10.2 Algoritmo de Árboles de decisión (DT). Un algoritmo de árboles de decisión, o DT (Decision Tree) por sus siglas en inglés, es una técnica de aprendizaje automático que se utiliza para modelar relaciones complejas entre las variables de entrada y de salida, consta de la construcción de una estructura con forma de árbol a partir de un conjunto de datos de entrenamiento, cada nodo dentro de este árbol representa una decisión que toma el modelo y cada rama que se deriva es una posible respuesta a esa pregunta.

Para construir un árbol de decisión se comienza con una pregunta inicial sobre la variable de entrada, esta pregunta debe separar el conjunto de entrenamiento en dos o más grupos, este proceso se repite subdividiendo todos los grupos en subgrupos más pequeños utilizando preguntas adicionales hasta que se llega al criterio de parada definido en el modelo [40]. Un ejemplo de esta

estructura se encuentra en la figura 12, donde se evidencian las hojas y las ramas según cada pregunta dentro del árbol.

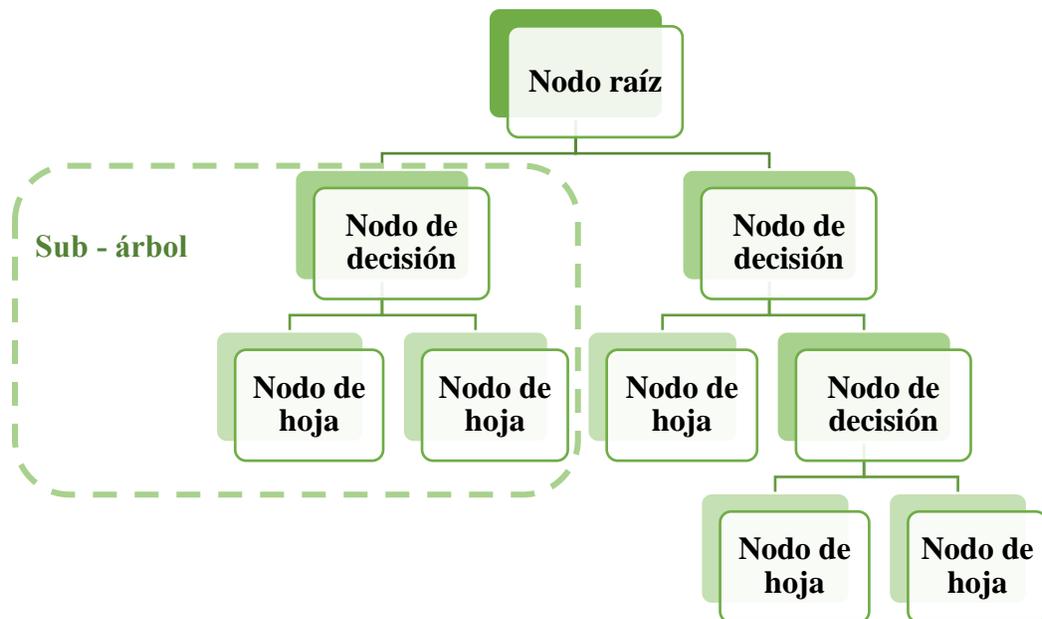


Figura 12. Estructura de un árbol de decisión. Fuente: Elaboración propia

3.2.10.3 Algoritmo de K-vecinos más cercanos (KNN). Un algoritmo de K-vecinos más cercanos, o KNN (K-Nearest Neighbors por sus siglas en inglés), es aquel que realiza tareas de clasificación o regresión basándose en que instancias similares se agrupan en las mismas clases. Lo anterior quiere decir que dado un conjunto de datos con valores previamente conocidos, el algoritmo KNN es capaz de clasificar nuevos datos basado en los K vecinos más cercanos definidos dentro del conjunto de entrenamiento, y que son calculados usando una métrica de distancia definida, como por ejemplo, la Euclidiana o Manhattan [41].

Entrando en detalle sobre el funcionamiento del algoritmo, en la figura 13 se puede apreciar un conjunto de datos agrupados en dos clases, A y B. El recuadro amarillo en la figura corresponde a un nuevo dato por clasificar, y por lo tanto, es necesario escoger la cantidad de vecinos a tener en cuenta para comenzar la tarea de clasificación. En este caso, se realiza el cálculo

de la distancia de los vecinos más cercanos utilizando la fórmula de distancia Euclidiana.

Finalmente, según la K cantidad de vecinos que se hayan definido, el modelo escoge los K puntos ya clasificados más cercanos al punto (amarillo) que se requiere clasificar.

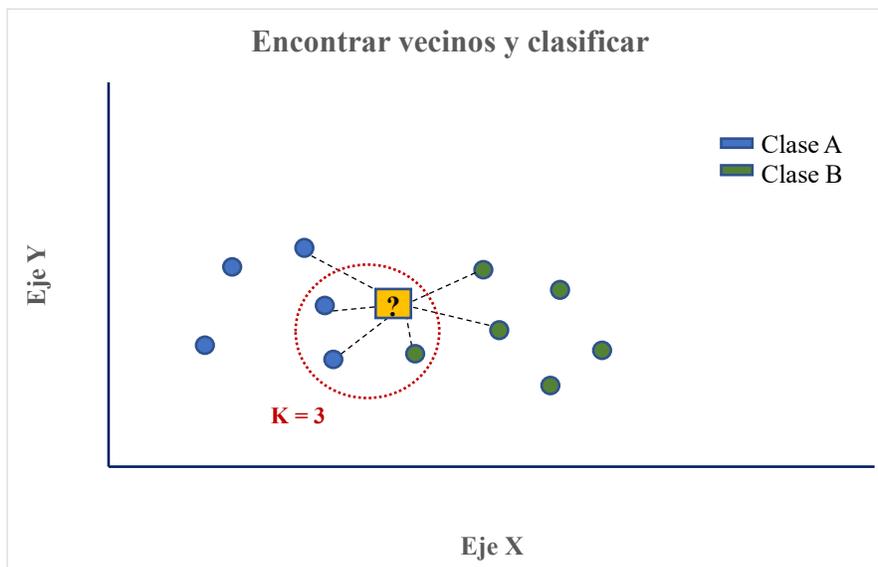


Figura 13. Ejemplo KNN: Clasificación. Fuente: Elaboración propia

En la figura se aprecia que, según la cantidad de vecinos, el nuevo dato corresponde a la “Clase A”, ya que la mayoría de los vecinos más cercanos que posee son de esta misma clase. En el caso de regresión, se utiliza el mismo concepto de vecinos más cercanos, con la salvedad de que para esta aplicación el algoritmo calcula la distancia de las variables de entrada con respecto a los datos conocidos de la serie temporal y según el valor de distancia se selecciona K vecinos.

Seguidamente, los pesos de cada dato se establecen de dos maneras, uniformemente o basado en las distancias. En el caso uniforme, se calcula la media de los k -vecinos, tal y como se evidencia en la figura 14.

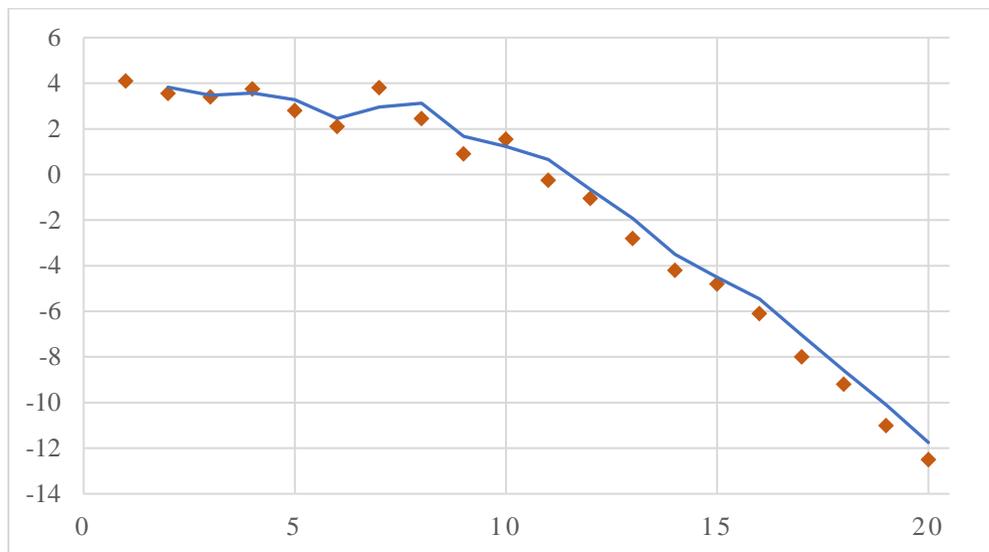


Figura 14. Ejemplo KNN: Pesos uniformes. Fuente: Elaboración propia

La otra forma de configurar los pesos del modelo es según la distancia, que analiza cuál de todos los k-vecinos está más cerca y toma ese dato sin tener en cuenta a los demás, esto se demuestra en la figura 15.

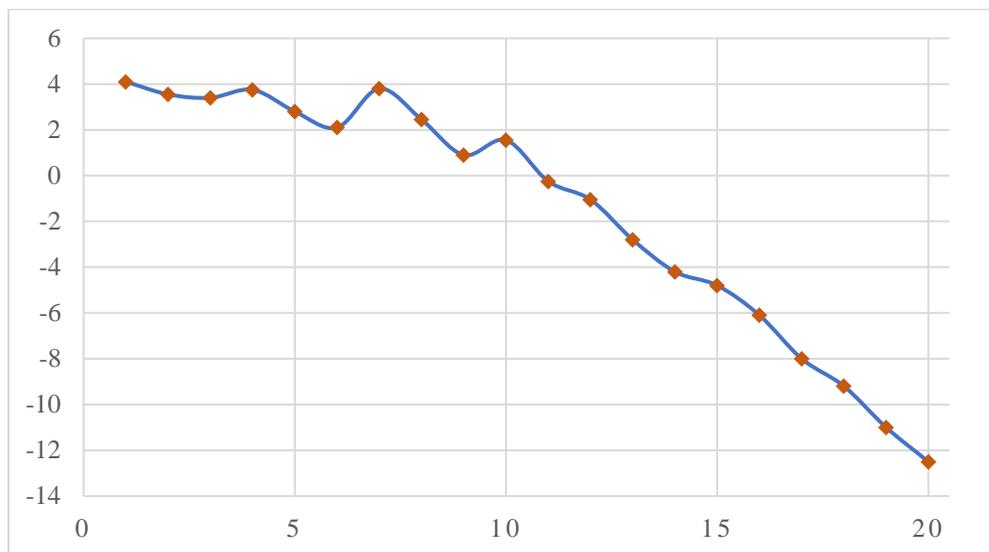


Figura 15. Ejemplo KNN: Pesos según la distancia. Fuente: Elaboración propia

En el caso de esta investigación se utilizó la configuración uniforme, la cual permite tener en cuenta todos los K-vecinos de la serie temporal, de esta forma se evita el sobreajuste de los datos del modelo.

3.2.10.4 Algoritmo Extreme Gradient Boosting (XGBoost). El algoritmo de aumento de gradiente extremo, o XGBoost (eXtreme Gradient Boosting), es una técnica de aprendizaje automático que se utiliza en implementaciones de regresión y clasificación para problemas que requieran realizar predicciones, es una variación de las máquinas de aumento de gradiente ya que a la vez utiliza árboles de decisión para aumentar la precisión del modelo generado [42].

De manera técnica, el modelo genera un conjunto de árboles de decisión en secuencia, lo que significa que en cada iteración se ajusta el árbol a los errores residuales del modelo anterior, además, existe un mecanismo de regularización para controlar la complejidad del modelo y prevenir el sobreajuste de los datos. El proceso de optimización de XGBoost se realiza mediante el algoritmo de gradiente descendente. Este modelo es conocido por su capacidad de manejar conjuntos de datos grandes y de alta dimensionalidad con una gran precisión en comparación a otros algoritmos. La manera en la que funciona se evidencia en la figura 16.

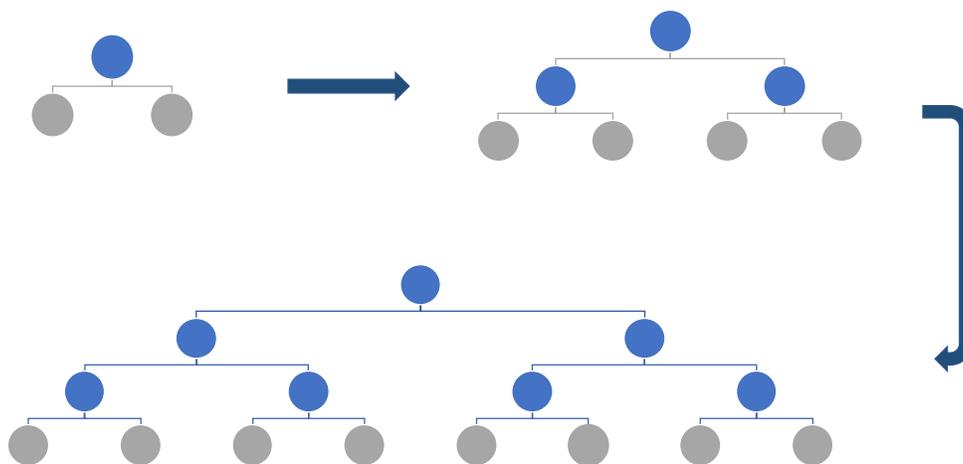


Figura 16. Estructura de XGBoost. Fuente: Elaboración propia

En esta figura se muestra que en cada iteración el modelo realiza una técnica denominada “crecimiento de árboles nivelados”, lo que significa que en cada iteración se calculan todas las ramas al mismo tiempo, al contrario de otros algoritmos que se escoge una rama y se descartan las otras de forma instantánea, esto incrementa el nivel de procesamiento requerido a cambio de una mejor precisión del modelo final.

3.2.11 Redes neuronales. Es considerada una de las aplicaciones más poderosas de la inteligencia artificial, así mismo ha sido la de mayor crecimiento a través de los años debido a sus aplicaciones. Se le considera como un campo del aprendizaje automático que es capaz de enfrentarse a problemas complejos que generalmente impliquen una gran cantidad de datos. Estas redes neuronales se organizan en capas para establecer relaciones entre todos los datos disponibles, por lo que requiere de una gran capacidad de procesamiento.

Su función es intentar asimilar el trabajo que ejercen las neuronas biológicas para reconocer patrones complejos dentro de un conjunto de datos, el sistema responde a las entradas externas para dar la salida respectiva. Un ejemplo sencillo de una red neuronal artificial se presenta en la figura 17, en la cual se presentan las entradas, salidas y capas ocultas con las que cuenta la red.

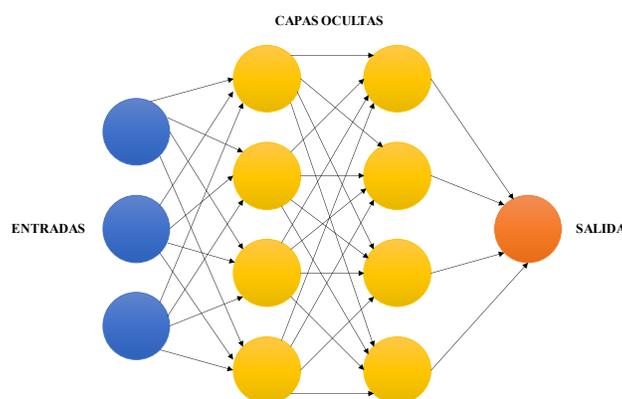


Figura 17. Esquema de una red neuronal. Fuente: Elaboración propia

En el caso de la figura 17 se evidencian varias entradas junto con una única salida, sin embargo, las redes neuronales pueden contar con más de una salida según sea necesario en cada caso. Las capas ocultas representan los procesos que se realizan para manipular las variables de entrada, la cantidad de capas ocultas depende directamente de la complejidad de la red neuronal, a mayor número de capas ocultas, más capacidad de procesamiento es necesaria [43]. En la figura 18 se evidencia la ecuación de salida de una red, se demuestra que la salida de la red se ve influenciada por un sesgo y la sumatoria de todas las entradas multiplicadas con sus respectivos pesos asignados por la red.

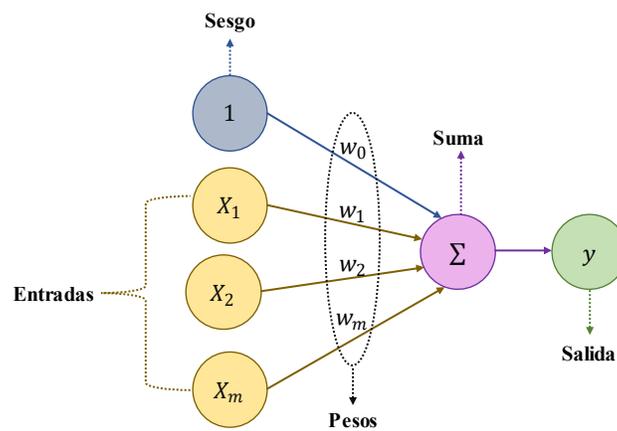


Figura 18. Ecuación de salida de una red neuronal. Fuente: Elaboración propia

Según esta figura, es posible identificar la ecuación de salida “y” de una red neuronal como la que se presenta a continuación, donde “ w_0 ” es el sesgo, “ w_i ” es el peso, y “ x_i ” es la entrada en la posición i .

$$y = w_0 + \sum_{i=1}^m x_i * w_i$$

3.2.11.1 Redes neuronales recurrentes. Una red neuronal recurrente es aquella que está diseñada para procesar datos de series temporales, y que a diferencia de las redes neuronales convencionales, poseen una memoria interna que les permite tomar en cuenta información previa de la base de datos y utilizarla en el procesamiento de los datos actuales, la estructura de estas redes incluye conexiones de retroalimentación, lo que indica que la salida de una capa se utiliza como entrada para la siguiente capa [44]. Un esquema de una red neuronal recurrente puede observarse en la figura 19, donde la salida corresponde al procesamiento de la retroalimentación.

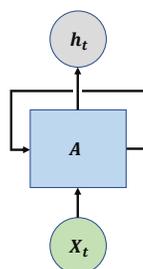


Figura 19. Estructura de red neuronal recurrente. Fuente: Elaboración propia

Un caso específico de este tipo de redes son las Long Short-Term Memory (LSTM), las cuales son un tipo de red neuronal recurrente utilizada para problemas de aprendizaje automático, este tipo de redes son capaces de retener información relevante durante periodos de tiempo más largo ya que estas poseen un componente de celdas de memoria, internamente se realiza un mecanismo de decisión para actualizar la información que está dentro de esas celdas [45]. Ahora, la estructura de este tipo de redes se evidencia en la figura 20.

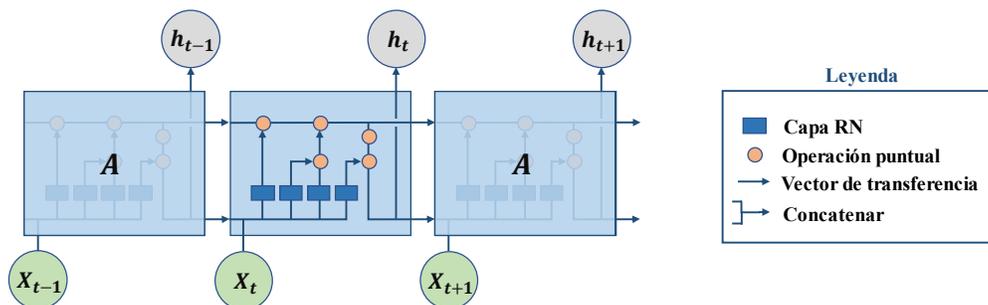


Figura 20. Estructura red LSTM. Fuente: Elaboración propia

La estructura interna de la LSTM posee varias capas de red neuronal que realizan operaciones puntuales y se transfieren al siguiente bloque en forma de memoria. Cada bloque A recibe la información a través de los vectores de transferencia del bloque anterior, así mismo, dentro de cada bloque se realizan operaciones de concatenación para procesar los datos temporales.

3.3 Marco legal

Esta sección especifica todas las leyes y normativas que se tuvieron en cuenta a la hora de llevar a cabo el proyecto relacionado a la programación, esto con el fin de cumplir el marco de la legalidad del país en el cual se llevó a cabo el desarrollo de este proyecto. En la tabla 1 se presenta la normatividad para la creación de algoritmos dentro de la República de Colombia.

Tabla 1. Normatividad sobre la creación de algoritmos en Colombia

Normas que aplican al desarrollador y al proceso de programación			
	Norma	Licencias	Bases de Datos
Ley 603 de 2000	Distribución de Licencias y Derechos de Autor	X	
Ley 1273 de 2009	Manipulación de Bases de Datos en Sistemas Informáticos		X
Normas que aplican al modelo híbrido			
Ley 23 de 1982	Ley de Protección de los Derechos de Autor	X	X

3.3.1 Ley 603 de 2000: Distribución de licencias y derechos de autor. Esta ley está destinada a exigirle a las empresas la presentación de informes de gestión que contengan el estado de cumplimiento de las normas referentes a la propiedad intelectual y los derechos de autor que debe cumplir la sociedad para usar el software, esto con el fin de proteger las licencias generadas por los desarrolladores de software dentro del territorio nacional de Colombia [46].

3.3.2 Ley 1273 de 2009: Manipulación de bases de datos en sistemas informáticos. Ley que modifica el código penal, creando un nuevo bien jurídico al cual denominan como “la protección de la información y los datos que se preservan integralmente en los sistemas de las tecnologías de la información”. Esta ley hace referencia a la normativa que se debe tener en cuenta a la hora de utilizar bases de datos de acceso público o privado, puesto que de lo contrario se estarían rompiendo las normas de derechos de autor [47].

3.3.3 Ley 23 de 1982: Ley de protección de los derechos de autor. La Ley 23 de 1982 (enero 28) decreta todas las normativas sobre derecho de autor que se aplica bajo obras literarias, científicas y artísticas con el fin de estipular la protección de los datos y el uso de dichos datos según lo decida el autor, como lo estipula en el artículo 3. Por lo cual, los autores pueden decidir si su creación de dicha obra es de forma gratuita o remunerada, refiriéndose a la forma en como la persona pueda tener acceso a los datos e información [48].

4. Diseño metodológico

Dentro de esta sección se especifican las actividades que se realizaron para llevar a cabo el desarrollo de los objetivos específicos planteados en la investigación. La ruta de actividades completadas se presenta en la figura 21.

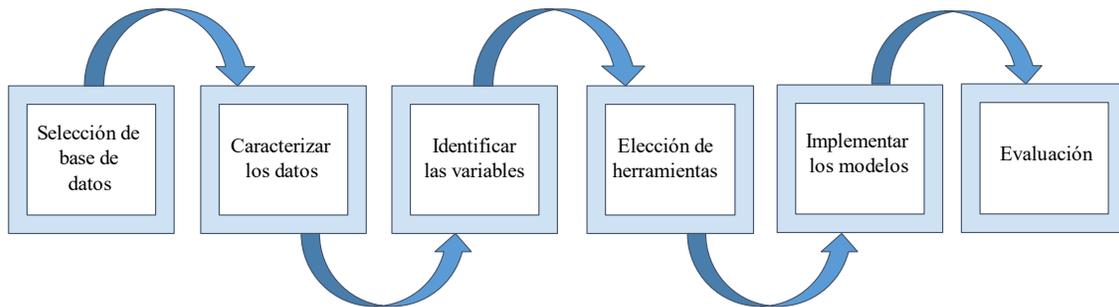


Figura 21. Ruta de actividades. Fuente: Elaboración propia

4.1 Selección y depuración de bases de datos

En esta sección se abordaron todos los conjuntos de datos relacionados a las energías renovables y a la demanda eléctrica. En esta investigación, todos son de organismos y/o empresas de los Estados Unidos ya que a nivel nacional no es posible acceder a este tipo de bases de datos, tal y como se mencionó en las delimitaciones del proyecto.

4.1.1 Búsqueda de bases de datos más utilizadas. Se buscaron distintas bases de datos para llevar a cabo este proyecto, los criterios que se tuvieron en cuenta para realizar dicho proceso fue que los conjuntos de datos tuvieran valores de producción solar fotovoltaica, producción eólica, demanda eléctrica, temperatura, radiación, u otras variables meteorológicas, y que a su vez pertenecieran a una misma localización geográfica; siguiendo estos lineamientos se encontraron en total seis entidades que recopilaban una parte de la información dentro del territorio de los Estados Unidos.

- **California Independent System Operator (CAISO):** Esta es una entidad que regula la información de la red eléctrica de todo el estado de California, y posee información desde

el año 2018 hasta el día de hoy recopilando información de demanda eléctrica, producción fotovoltaica y producción eólica. Existen antecedentes que han utilizado esta entidad para analizar el impacto de la generación de energía solar en la red eléctrica de todo el estado de California gracias a su gran cantidad de datos [49].

- **National Renewable Energy Laboratory (NREL):** Es una entidad que posee un amplio catálogo de bases de datos en Estados Unidos debido a que funciona como un repositorio de datos donde los investigadores pueden subir las mediciones recolectadas durante sus proyectos. La organización también posee sus propios conjuntos de datos recolectados mediante satélites, por lo cual, poseen información de diversas partes del mundo. Algunas de las variables que miden son la temperatura, la radiación solar, la velocidad del viento, entre otras. Los antecedentes han desarrollado estudios donde analizan el impacto de la producción eólica en la red eléctrica de Estados Unidos [50].
- **European Network of Transmission System Operators (ENTSOE):** Esta empresa se encarga de recopilar los datos de generación y demanda eléctrica de toda la red europea, donde algunas de sus bases de datos son abiertas para el público. Investigadores han utilizado este conjunto de datos para analizar el impacto que posee la integración de energías renovables a gran escala en todo el mercado eléctrico europeo, permitiendo hacer un análisis profundo y detallado de la situación [51].
- **International Energy Agency (IEA):** Esta entidad es de escala internacional y recopila datos de las redes eléctricas de distintos países a nivel mundial, algunas variables que mide son producción global, consumo, emisiones de CO₂, inversiones en energías limpias, entre otros. Los conjuntos de información que utilizan en su mayoría son privados, sin embargo, anualmente publican un informe que otorga una visión general de todas las tendencias globales en materia de energía [52].

- **Energy Information Administration (EIA):** Esta es una empresa estadounidense, de la misma forma que otras entidades, se dedica a la recopilación de los datos de la red eléctrica del país, por ejemplo, almacenan información de producción, consumo y emisiones de CO₂. La empresa colabora con otras para generar reportes anuales en temas de energía, como por ejemplo IEA e IRENA.
- **International Renewable Energy Agency (IRENA):** Esta entidad se encarga de recopilar datos a nivel mundial de energías renovables, almacena información de producción de energía eólica y energía solar fotovoltaica, estos datos son publicados en informes anuales donde se observa la tendencia a nivel internacional de la aplicación de estas energías, en uno de sus informes más recientes indicaron cuáles son los costos detallados de generación utilizando este tipo de energías limpias [53].

4.1.2 Comparación y selección de las bases de datos. En la subsección anterior se presentaron las seis entidades que de acuerdo con el estado del arte se utilizan para realizar análisis de producción y consumo de energía eléctrica a través de diferentes países. Para escoger los conjuntos de datos a utilizar en el proyecto, se procedió a realizar una comparación detallada de las características de las organizaciones tal y como se presenta en la tabla 2.

Tabla 2. Características de las bases de datos tenidas en cuenta para la selección final

Características	CAISO	NREL	ENTSOE	IEA	EIA	IRENA
Datos de libre acceso	Si	Si	Si	No	No	No
Intervalo de tiempo (min)	5	5	15	5 – 15	5 – 15	5 – 15
Formato de datos	CSV	CSV	XML	CSV	JSON	API
Datos que contiene	Generación y consumo	Variables meteorológicas	Generación y consumo	Generación y polución	Generación y polución	Costos y generación
Decisión final	Seleccionado	Seleccionado				

Considerando los datos en la tabla 2 se escogieron las bases de datos utilizadas para generar las predicciones dentro de este proyecto de investigación. A continuación, se presenta la tabla 3, donde se evidencia la matriz de selección para escoger las bases de datos.

Tabla 3. Matriz de selección de las bases de datos usadas en la realización del proyecto

Características	CAISO	NREL	ENTSOE	IEA	EIA	IRENA
Datos de libre acceso	5	5	5	1	1	1
Intervalo de tiempo (min)	5	5	2	3	3	3
Formato de datos	5	5	3	5	4	2
Datos que contiene	5	4	5	2	2	3
Puntaje total	625	500	150	30	24	18
Base de datos elegida	Seleccionado	Seleccionado				

El proceso para definir el valor de “Puntaje total” se obtiene multiplicando el puntaje de cada característica, esto se repite para todas las matrices de selección en el documento. Basados en la tabla 3, se decidió utilizar las bases de datos disponibles en el sitio web de CAISO y NREL debido a que poseen un intervalo de tiempo entre cada medición muy bajo; por lo tanto, se delimitó que el modelo híbrido se entrenara con datos del estado de California dado que CAISO solo recopila información de esta zona geográfica. Cada conjunto de datos posee información diferente, a continuación, se enlistan las variables que se utilizaron de cada una de estas entidades:

- CAISO posee datos de producción de energía solar fotovoltaica, energía eólica y demanda de energía eléctrica.
- NREL posee datos de DHI (Irradiancia horizontal difusa), DNI (Irradiancia normal directa), velocidad del viento, humedad, temperatura y GHI (Irradiancia global horizontal). Estas variables fueron consideradas como variables exógenas, es decir, variables independientes que pueden complementar a las variables medidas por CAISO.

Para el caso de los datos de NREL, fue necesario tomar datos de cinco ciudades distintas alrededor de todo el estado de California, ya que esta entidad posee datos satelitales para una latitud y longitud en específico. Por lo tanto, se realizó un promedio de los datos que fuera posible utilizar en conjunto con los datos de CAISO. Para efectos del proyecto, se optó por tomar información de las siguientes ciudades: Fortuna, Sacramento, Bishop, Baker y San Diego, de manera gráfica, las ciudades que se tomaron se muestran en la figura 22.

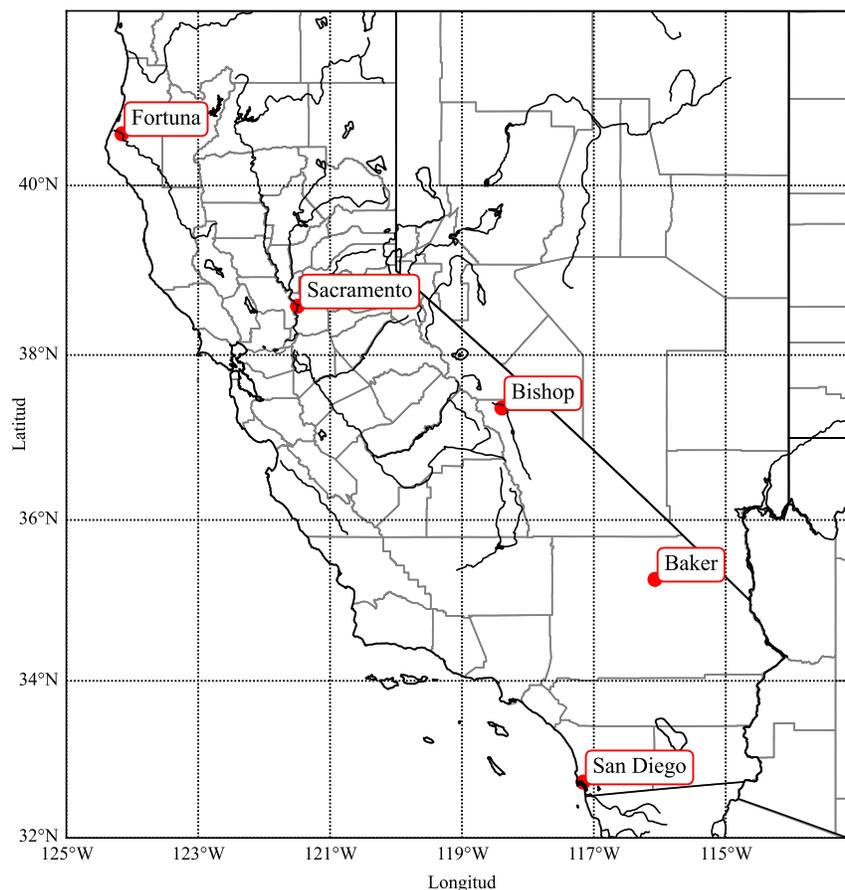


Figura 22. Localización de las ciudades de los datos de NREL. Fuente: Elaboración propia

El criterio que se tuvo en cuenta para escoger las ciudades de la Figura 22 fue de que las ciudades tuvieran una alta cantidad de radiación solar según los datos del buscador de NREL [31], asimismo, que fueran localizaciones esparcidas de norte a sur del estado de California.

4.1.3 Adquisición de los datos. Fue importante establecer el intervalo de datos a utilizar en el proyecto de investigación, para ello se revisaron las fechas de las bases de datos de ambas entidades y se encontró que:

- CAISO posee datos desde el 10 de abril del año 2018, hasta el día de hoy ya que continuamente recopilan información.
- NREL posee datos públicos desde el 1 de enero del año 2019, hasta el 31 de diciembre del año 2021.

Debido a esto, se tomó la decisión de utilizar datos desde el 1 de enero del año 2019 hasta el 31 de diciembre del año 2021 de ambas bases de datos, para que existiera la misma cantidad de información. El proceso de obtención de los datos se llevó a cabo en el sitio web de cada entidad, en el caso de CAISO, para obtener los datos de producción de energía renovable se usó la interfaz de la figura 23.

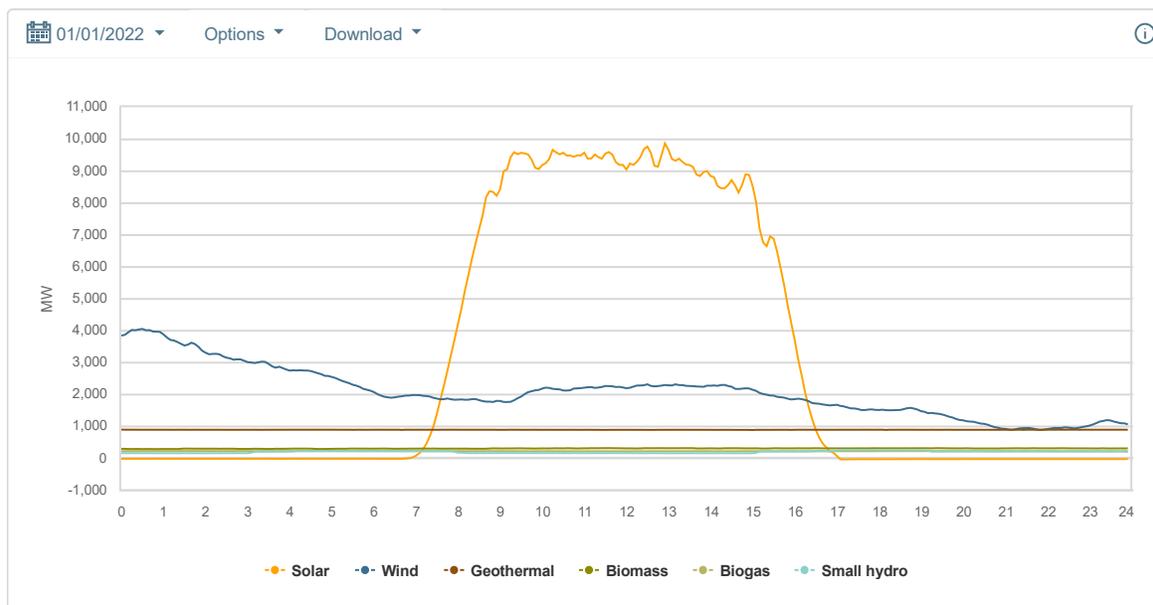


Figura 23. Interfaz disponible en el sitio web de CAISO para producción. Fuente: Tomado de [30]

Asimismo, para el caso de la demanda de energía eléctrica se utilizó la interfaz de la figura 24 disponible en el sitio web de CAISO.

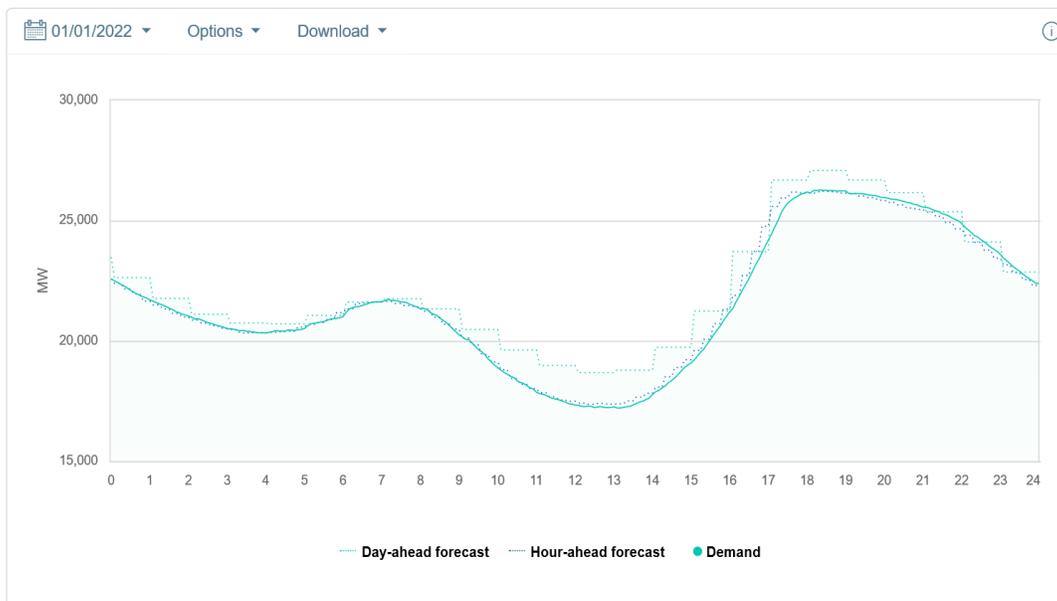


Figura 24. Interfaz disponible en el sitio web de CAISO para demanda. Fuente: Tomado de: [30]

En el caso de NREL se utilizó específicamente la base de datos llamada “National Solar Radiation Database (NSRDB)”, cuya interfaz se muestra en la figura 25.

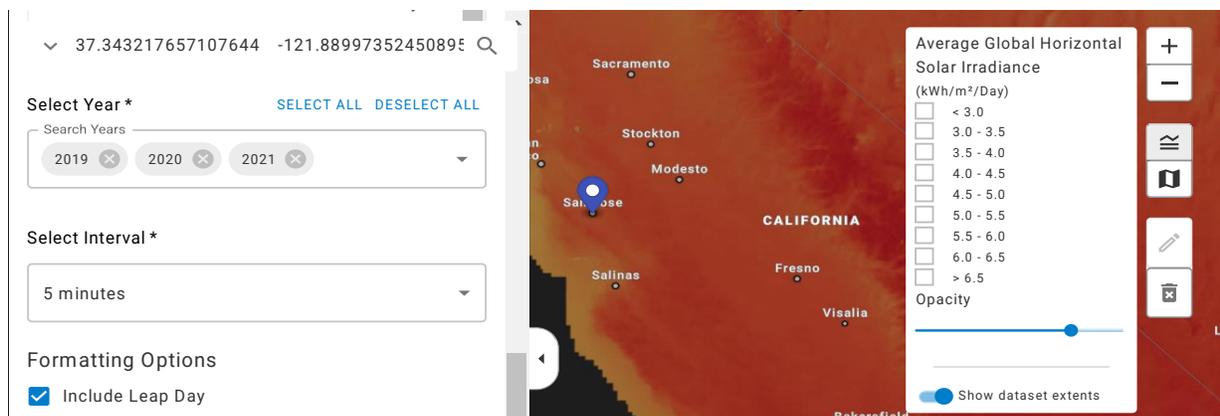


Figura 25. Interfaz disponible en el sitio web de NREL. Fuente: Tomado de [31]

Finalmente, se descargaron todos los datos dentro del intervalo de fechas estipulado (01/01/2019 – 31/12/2021) en pasos de cinco minutos entre cada medición. Se utilizó el lenguaje

de programación Python junto con la librería Pandas, la cual está enfocada en el manejo de grandes bases de datos. La lógica de programación utilizada se presenta en la tabla 4.

Tabla 4. Pseudo código de la lógica de adquisición de datos

Algoritmo para la adquisición de datos
BEGIN
FOR each year between 2019 and 2021
FOR each month of the year
FOR each day of the month
Open the document from CAISO in that date
FOR each hour of the day
FOR each 5 minutes
Read the solar and wind production
Read the demand
Save the values into a list including the date
FOR each city in Baker, Bishop, Fortuna, Sacramento and San Diego
FOR each year between 2019 and 2021
Open the document from NREL in that date
Read the value of DHI, DNI, Wind speed, Humidity, Temperature and GHI
Save the values into a list including the date
FOR each position in NREL lists
Calculate and save the average value into a list
Create a pandas DataFrame with all the lists from CAISO and NREL
Export the DataFrame to a CSV file
END

De esta forma, se realizó la adquisición de los datos de CAISO y NREL desde el año 2019, hasta el año 2021, teniendo en cuenta todas las variables mencionadas anteriormente.

4.1.4 Limpieza del conjunto de datos seleccionado. Es común que un conjunto de datos posea errores en las mediciones debido a situaciones externas a la entidad que recolecta la información, por lo tanto, es necesario realizar una depuración de los datos a nivel general siguiendo un proceso de ingeniería de datos.

En el lenguaje de programación Python existe una librería denominada Seaborn que permite graficar información perteneciente a bases de datos, de esta forma la depuración se puede simplificar en gran medida. Al ingresar el archivo de formato csv generado en la sección anterior se obtiene la gráfica de la figura 26 que relaciona la hora del día, con respecto a la producción eólica, producción solar y demanda eléctrica.

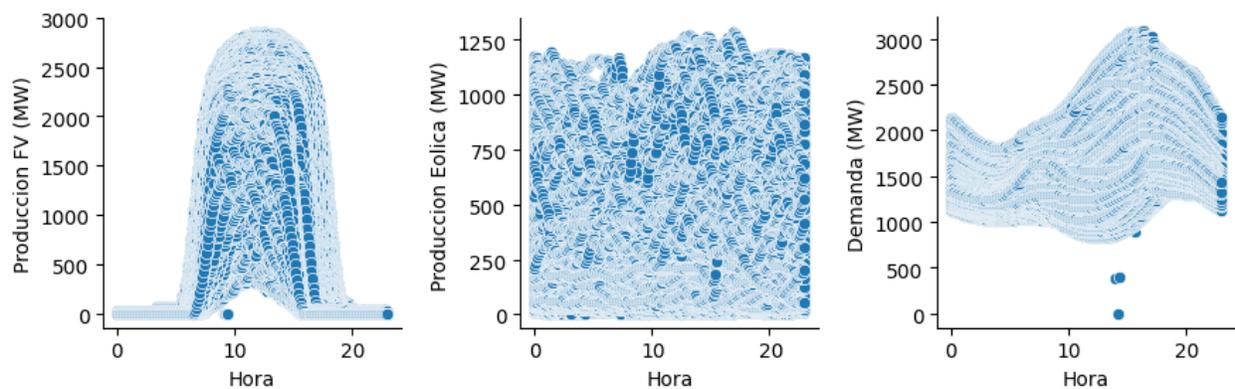


Figura 26. Gráfica de los datos generada con Seaborn sin depurar. Fuente: Elaboración propia

De la figura 26 se pueden realizar distintos análisis con respecto a la base de datos generada, teniendo en cuenta que las gráficas contienen toda la información del conjunto, se llegó al siguiente análisis:

- La energía solar fotovoltaica posee algunos datos erróneos.
- La energía eólica no evidencia a simple vista errores en las mediciones, aunque sí los tiene.
- La demanda eléctrica contiene datos que se encuentran desfasados.

Utilizando técnicas de ingeniería de datos para encontrar aquellos fragmentos de la información que no siguen el mismo patrón que los demás es posible generar una depuración eficaz de esta base de datos, aproximadamente se encontraron 20 fechas alrededor de los tres años que contenían errores esparcidos en las tres series temporales, por lo tanto, en la figura 27 se presenta nuevamente la gráfica generada por Seaborn de todos los datos del conjunto una vez interpolados los datos que se encontraban fuera del patrón.

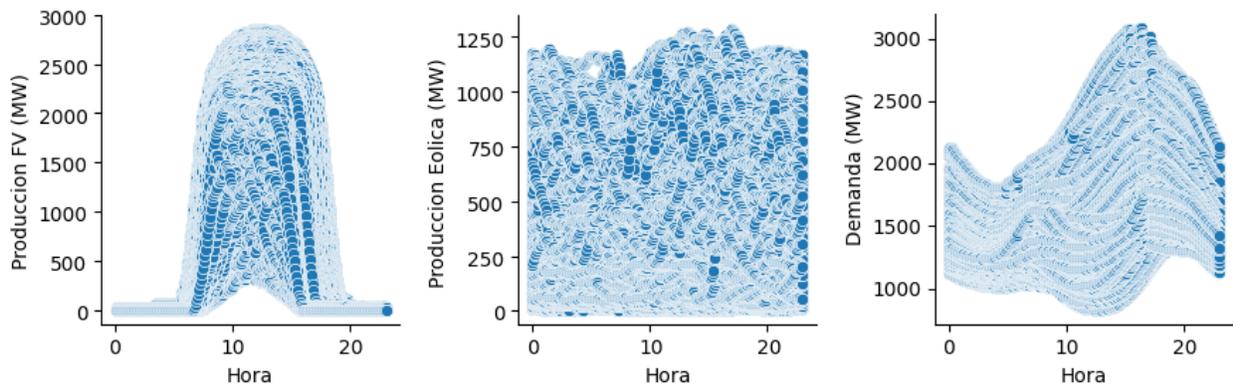


Figura 27. Gráfica de los datos generada con Seaborn depurada. Fuente: Elaboración propia

Los datos fueron depurados, como se evidenció en la figura 27. El proceso que se realizó para depurar los datos es el siguiente, ya que hubo momentos del día que poseían picos negativos, el proceso que se realizó consistió en utilizar el dato anterior y el dato siguiente a la sección que contiene el error para identificar el valor más preciso para ese instante de tiempo, la fórmula de este proceso se describe en la ecuación 1.

$$y_n = y_{n-1} + \frac{x_n - x_{n-1}}{x_{n+1} - x_{n-1}} (y_{n+1} - y_{n-1}) \quad (1)$$

Donde (x_{n-1}, y_{n-1}) son las magnitudes de la variable para el instante anterior y (x_{n+1}, y_{n+1}) del instante de tiempo siguiente del cual se quiere saber el valor interpolado [54]. El código se automatizó para ser aplicado a todos los datos donde ocurrían estos errores en las mediciones.

Asimismo, se identificó otro error en la base de datos, el cual fue que, en el caso de producción solar en las horas que no se estaba generando energía se tomaban mediciones negativas, sin embargo, estos valores deberían ser igual a cero ya que no se está produciendo energía, así que mediante código se estipuló que durante las horas que el Sol no estuviera (entre la noche y la madrugada), los valores negativos se convirtieran en cero. Como ejemplo para lo mencionado anteriormente se muestra una gráfica del día 7 de noviembre del 2020, cuyos datos estaban corruptos con picos y producción negativa tal y como se observa en la figura 28.

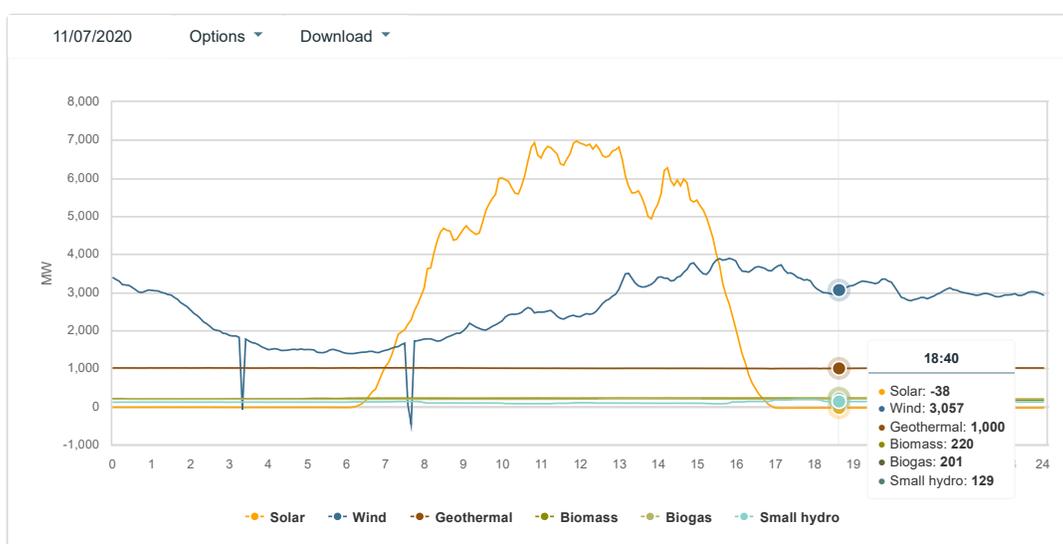


Figura 28. Ejemplo de datos con mediciones erróneas. Fuente: Tomado de [30]

De esta forma, finalmente se obtuvo una base de datos depurada lista para ser utilizada en algoritmos de aprendizaje automático, la cantidad de datos por año dentro del conjunto de datos se presenta en la tabla 5.

Tabla 5. Cantidad de información en la base de datos

Año	Cantidad de datos
2019	105120
2020	105408
2021	105120
Total	315648

La cantidad de datos se obtuvo mediante la siguiente operación, se debió tener en cuenta que los pasos eran cada cinco minutos, y que para el año 2020 son 366 días ya que fue año bisiesto.

$$\text{Cantidad de datos} = \frac{1 \text{ dato}}{5 \text{ min}} * \frac{60 \text{ min}}{1 \text{ h}} * \frac{24 \text{ h}}{1 \text{ día}} * \frac{365 \text{ días}}{1 \text{ año}} = 105120 \frac{\text{datos}}{\text{año}}$$

4.2 Exploración y caracterización preliminar de los datos

Una vez se depuraron las bases de datos seleccionadas, fue necesario realizar una revisión más detallada de la información para conocer la tendencia, estacionalidad y las técnicas que se tuvieron que emplear para realizar una transformación a los datos de la producción fotovoltaica, la eólica y la demanda eléctrica.

4.2.1 Tendencia de la producción y la demanda eléctrica. Dada la exploración de las gráficas de producción fotovoltaica diaria de todo un mes que se muestra en la figura 29 se observó que entre las 8 de la mañana y las 4 de la tarde existe una mayor producción debido a la cantidad de radiación solar que llega a la superficie. Cabe destacar que, aunque todas las gráficas no contienen la misma magnitud en su producción, los datos siguen la misma tendencia de presentar un pico de generación a medio día.

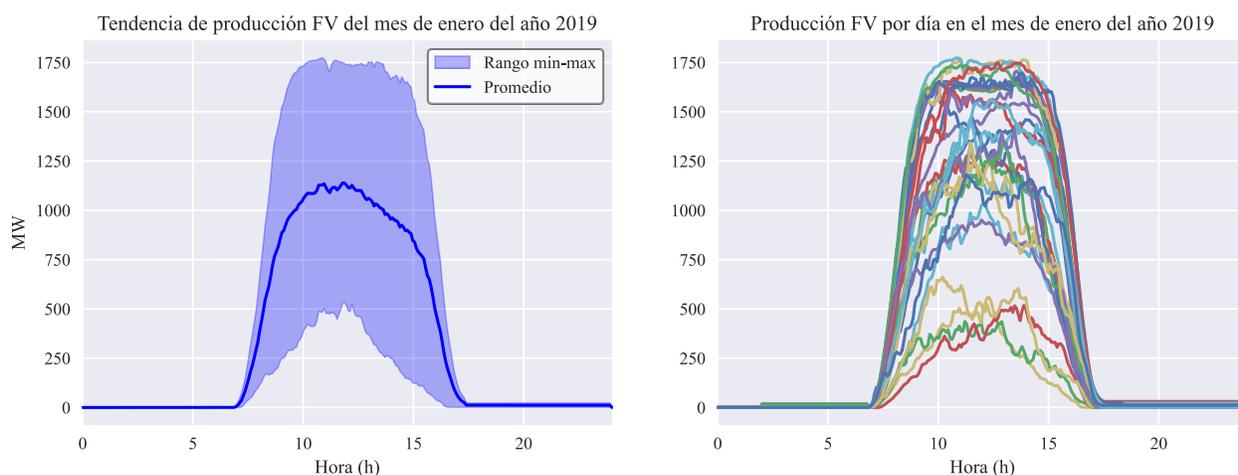


Figura 29. Producción fotovoltaica durante el mes de enero de 2019. Fuente: Elaboración propia

De la misma manera, se analizaron los datos de la producción de energía eólica y se pudo concluir que el comportamiento del viento a lo largo del día varía más, por lo que no es posible detectar una tendencia clara en las mediciones de las gráficas que se encuentran en la figura 30.

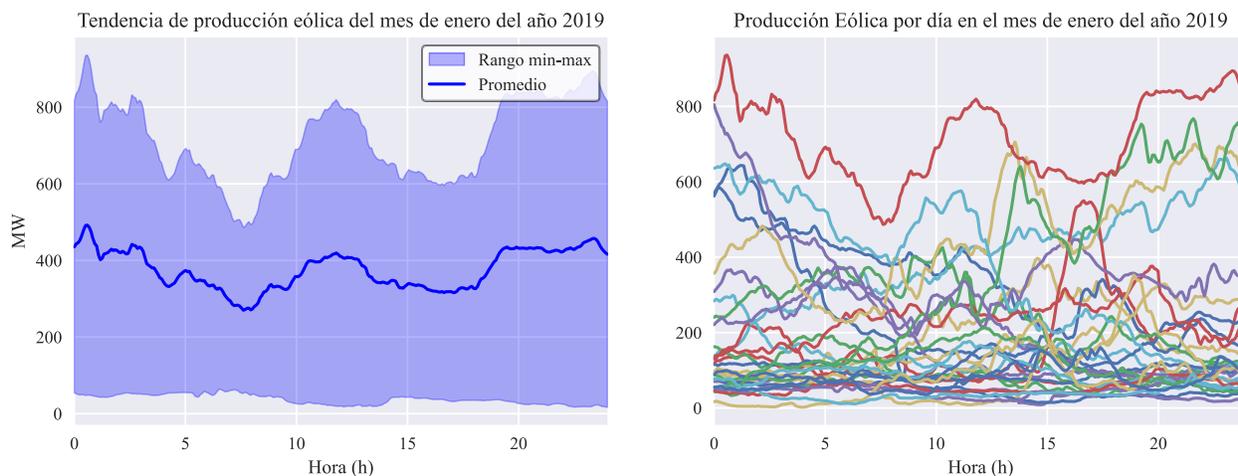


Figura 30. Producción eólica durante el mes de enero del año 2019. Fuente: Elaboración propia

Por último, se analizó la información proporcionada por CAISO correspondiente a la demanda eléctrica en la microrred localizada en California y esta variable sí tuvo una tendencia marcada, caso contrario a la producción eólica y se ilustra en la figura 31, el comportamiento de esta gráfica se debe a que se está experimentando una mayor penetración de energía renovable variable en la red, como la solar y la eólica [55].

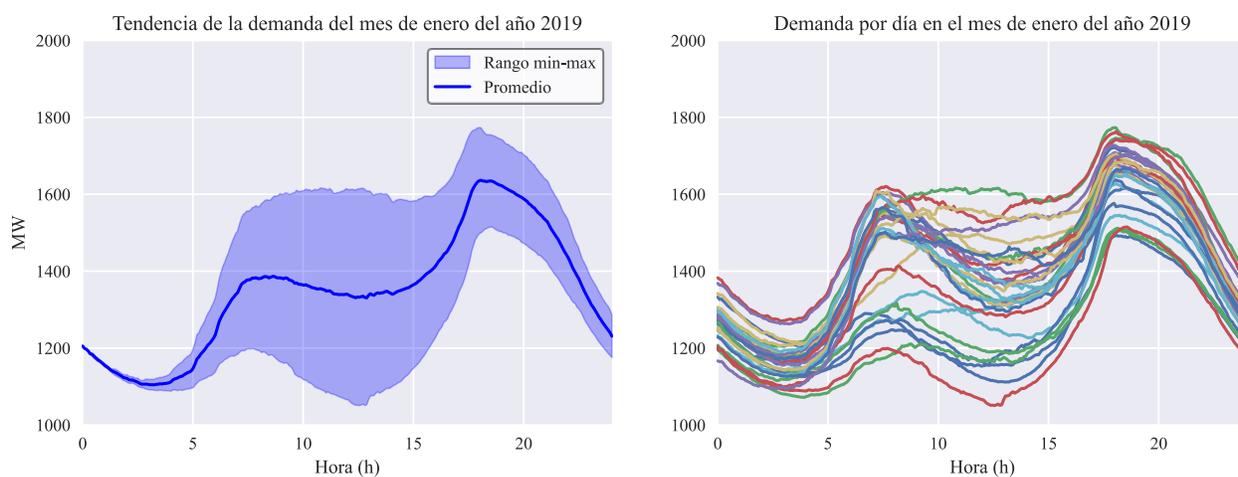


Figura 31. Demanda eléctrica durante el mes de enero del año 2019. Fuente: Elaboración propia

Cabe resaltar que existe un patrón denominado como la curva del pato, el cual se obtiene al restar la demanda eléctrica de un día con respecto a la producción de energía renovable, esto obtiene una gráfica la cual presenta dos picos durante el transcurso del día, los cuales son los momentos antes y después de mediodía.

4.2.2 Influencia de las estaciones del año. Para identificar el comportamiento de las variables temporales con las que se trabajaron en la implementación de los algoritmos de predicción se hizo un análisis por temporada (primavera, verano, otoño e invierno) para identificar qué tanto se ve afectada las producciones y la demanda energética en diferentes épocas del año.

En la figura 32 se muestran gráficamente los datos separados por las estaciones, de los gráficos se concluyó que en primavera y en verano existe mayor estabilidad en la producción fotovoltaica, es decir, la toma de muestras para esos lapsos de tiempo es similar.

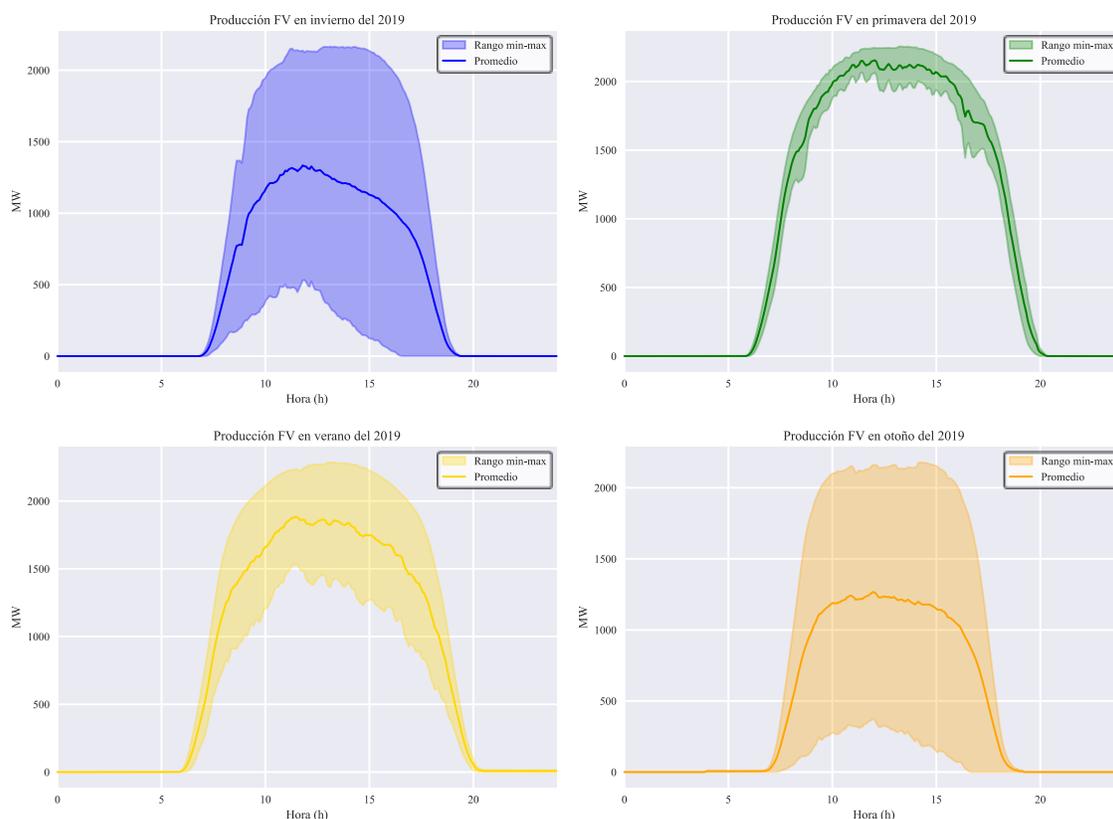


Figura 32. Producción fotovoltaica en 2019 por estaciones. Fuente: Elaboración propia

Nuevamente, en la figura 33 se evidencia que la producción eólica varía mucho con respecto al tiempo, sin importar la estación del año en la que se haga la toma de datos. En la figura se observa el promedio de los datos y las franjas difuminadas indican la distribución de los datos, es decir, se marca el límite superior e inferior.

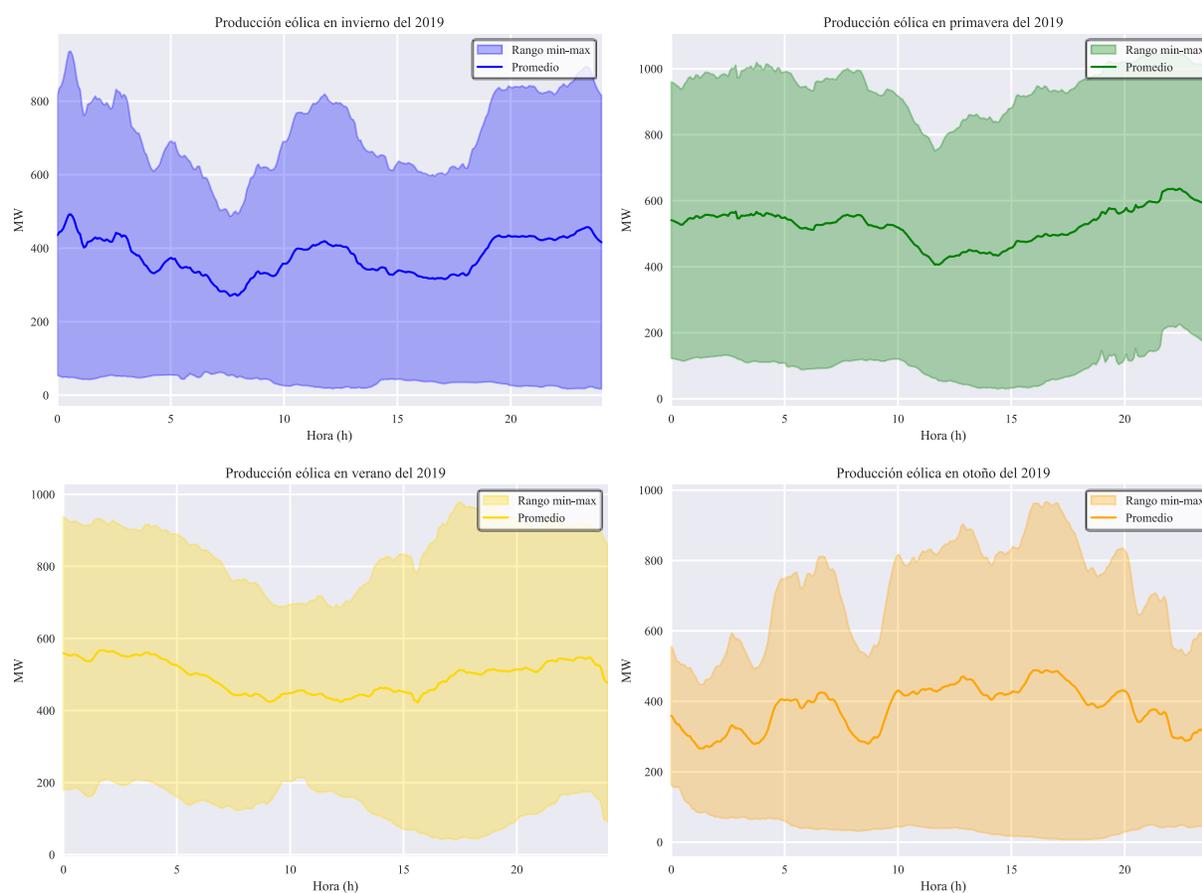


Figura 33. Producción eólica en 2019 por estaciones. Fuente: Elaboración propia

En el caso de la demanda eléctrica se observó en la figura 34 que la tendencia de la gráfica bimodal solo se presenta de manera acentuada en la temporada de invierno, mientras que en las otras gráficas uno de los picos desaparece.

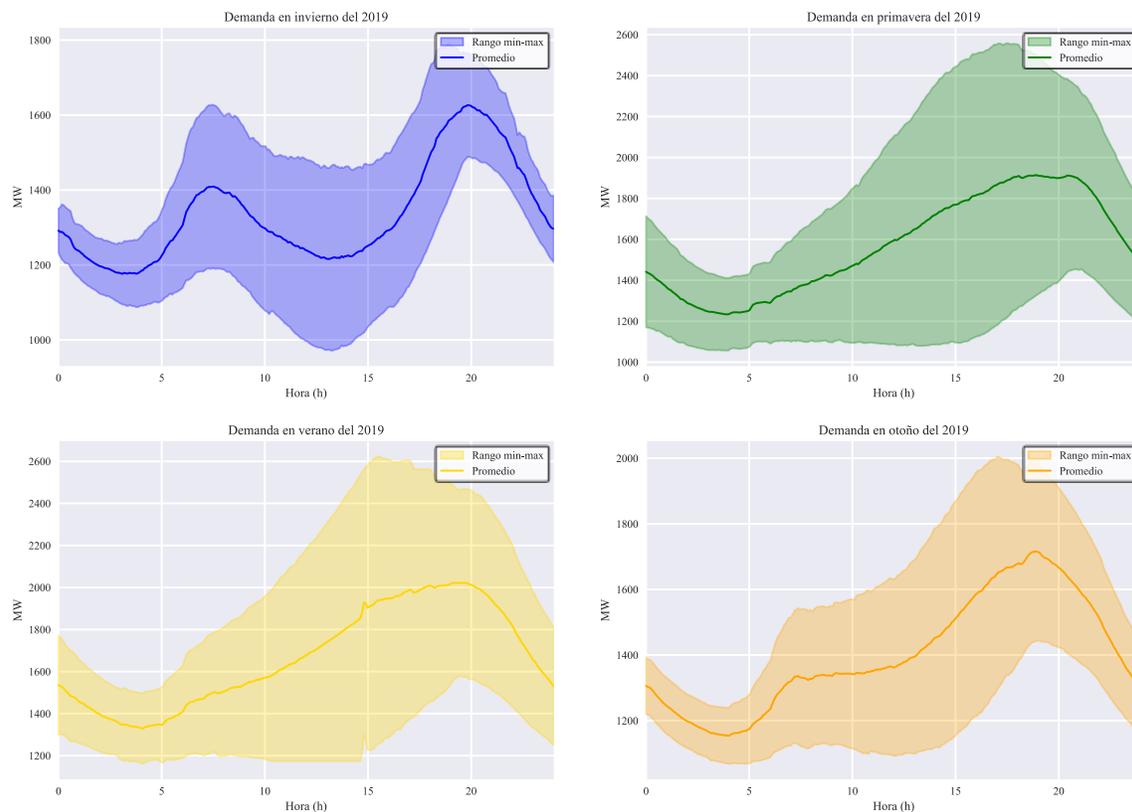


Figura 34. Demanda eléctrica en 2019 por estaciones. Fuente: Elaboración propia

4.2.3 Estacionalidad en la serie temporal. Dado que no fue posible detectar una tendencia en la producción de energía eólica, se utilizó una técnica de descomposición para analizar la tendencia de cada uno de los conjuntos, como se evidencia en la figura 35, figura 36 y figura 37. La información de las gráficas se divide en cuatro secciones, las cuales son:

- **Datos:** Contiene información de toda la serie temporal e identifica el patrón que posee.
- **Tendencia:** Se muestra la dirección que posee la serie a través del tiempo.
- **Estacionalidad:** Se evidencia si existe un ciclo estacional en los datos.
- **Residuos:** Demuestra el “ruido” que existe dentro de la serie.

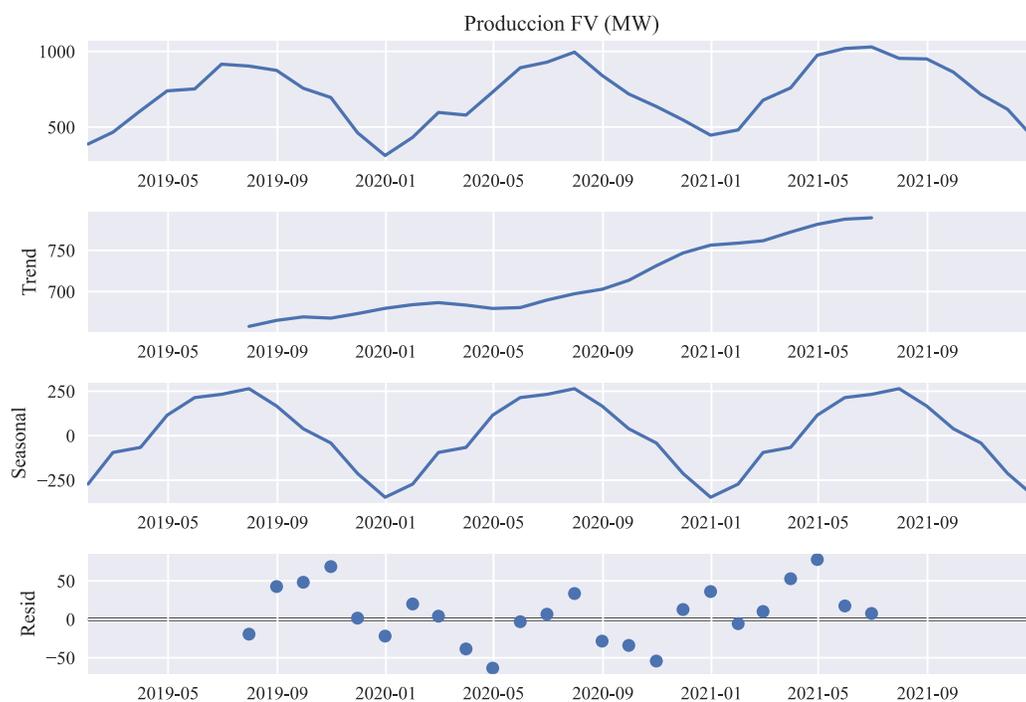


Figura 35. Descomposición estacional de producción solar. Fuente: Elaboración propia

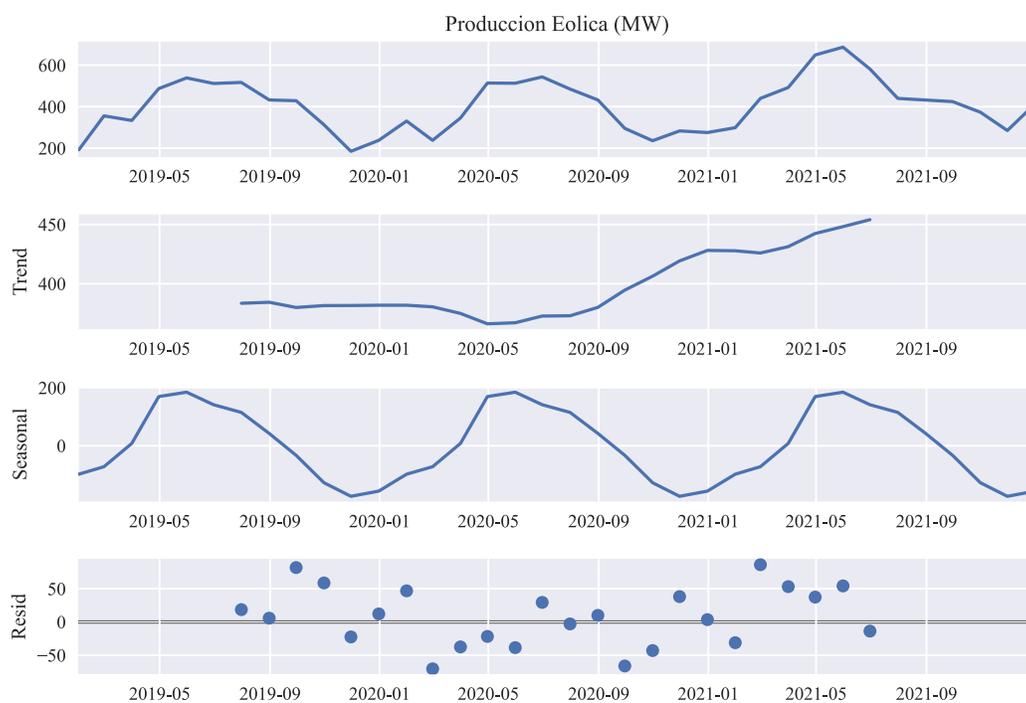


Figura 36. Descomposición estacional de producción eólica. Fuente: Elaboración propia

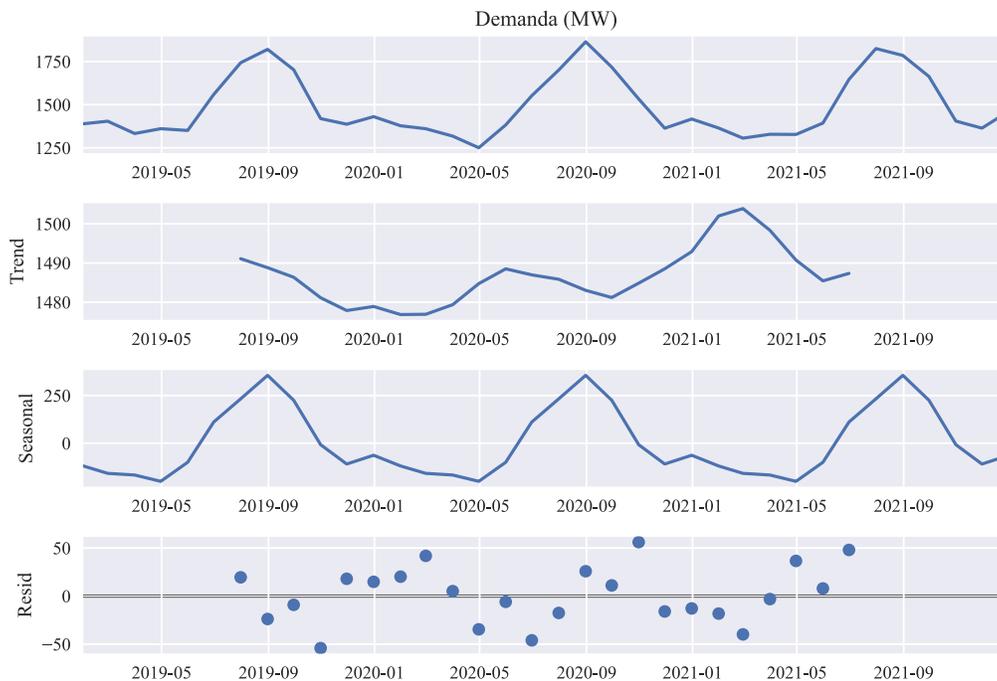


Figura 37. Descomposición estacional de demanda eléctrica. Fuente: Elaboración propia

De las gráficas se obtuvo información relevante con respecto a las características de las series temporales que se utilizaron, por ejemplo:

- Las tres series temporales presentan una componente estacional que evidencia que en cierto momento del año poseen una mayor magnitud.
- La tendencia de la energía solar y eólica va en aumento, lo que indica que cada año se añade más capacidad a las energías renovables.
- Los valores de los residuos son pequeños, esto sugiere que el modelo puede capturar de manera correcta la estructura de la serie.

Asimismo, se utilizó una técnica de ingeniería de datos denominada como prueba de Dickie Fuller, esta técnica analiza los parámetros de una serie temporal y evidencia si posee o no una componente estacional. Para esto se utilizó una librería llamada “statsmodels.tsa.seasonal”, la cual incluye una función para aplicar esta prueba a un DataFrame de pandas, y según esto se

obtuvo que las tres series temporales de producción de energía solar fotovoltaica, energía eólica y demanda eléctrica son estacionales.

4.2.4 Transformación de los datos. Finalmente, fue necesario estandarizar los datos en bruto antes de ingresarlos a modelos de aprendizaje automático, en este proyecto se estandarizaron cada una de las variables utilizadas (columnas del Dataframe) para tener media igual a cero y una desviación estándar igual a uno. El proceso de estandarización se llevó a cabo utilizando la ecuación 2:

$$z = \frac{x - \mu}{\sigma} \quad (2)$$

Donde μ es el promedio de todos los datos de esa columna dentro del DataFrame y se define mediante la ecuación 3:

$$\mu = \frac{1}{N} \sum_{i=1}^N x_i \quad (3)$$

Y donde σ es la desviación estándar, la cual, por definición, es la medida de dispersión o variabilidad utilizada para describir la diferencia en el conjunto, y se define como la ecuación 4:

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2} \quad (4)$$

Esta es una técnica de preprocesamiento, se utilizó la clase “StandardScaler” de la librería scikit-learn de Python para automatizar el proceso de estandarización. Como se mencionó antes, este proceso transformó los datos para que obtuvieran una media de cero y un valor de desviación estándar de uno. La estandarización de datos es vital para cualquier entrenamiento de modelos de inteligencia artificial, pues de este modo el modelo no centra las predicciones en la variable cuya escala sea mayor. Esto mejoró el rendimiento de la serie temporal dentro de los modelos de aprendizaje automático. Al interpretar los resultados fue necesario realizar una transformación inversa para devolver los valores a su magnitud original [56].

Se aplicó un factor de reducción a todas las series temporales, de modo que en promedio la suma de la producción solar fotovoltaica y eólica estuvieran entre el 70% y 80% de la demanda eléctrica anual, ya que en Estados Unidos y Colombia existen restricciones y limitaciones sobre la cantidad de energía renovable que un consumidor puede generar y entregar a la red, así que estableciendo este límite se aseguró el cumplimiento de la normativa [57], [58].

De esta forma, en la tabla 6 se determinó el porcentaje total que abarca la producción solar y eólica con respecto a la demanda de energía anual debido a un factor de reducción en la magnitud de los datos, para cada uno de los años de la base de datos.

Tabla 6. Porcentaje total de cobertura de la producción de energía

Recurso	Año 2019	Año 2020	Año 2021
Producción solar	44,02%	46,77%	52,86%
Producción eólica	25,31%	25,18%	30,85%
Total	69,33%	71,95%	83,71%

4.3 Identificación de variables de entrada y salida

Dentro de esa sección se abordó la revisión de cuáles serían las variables de entrada en cada uno de los modelos de aprendizaje automático, se revisaron técnicas de autocorrelación que indican la relación que tiene la serie temporal con respecto a datos pasados del mismo conjunto.

4.3.1 Revisión de variables significativas en antecedentes. En esta subsección se abordó una revisión de las variables que utilizaron los antecedentes para realizar predicciones sobre series temporales. Para el primer caso se presenta la tabla 7 donde se compararon las variables de energía fotovoltaica.

Tabla 7. Variables de entrada para predicción de energía solar

Referencia	Variables que usa
[59]	Temperatura, irradiancia solar, presión y humedad
[60]	Radiación solar, temperatura, y nubosidad
[61]	Velocidad del viento, humedad y temperatura

Para el siguiente caso se presenta la tabla 8 donde se compararon las variables de entrada para modelos de predicción de energía eólica.

Tabla 8. Variables de entrada para predicción de energía eólica

Referencia	Variables que usa
[62]	Velocidad del viento, dirección del viento, temperatura y humedad
[63]	Presión y temperatura
[64]	Velocidad del viento

Finalmente, en la tabla 9 se realizó la comparación de los parámetros de entrada que se utilizaron en los antecedentes para predicción de demanda eléctrica.

Tabla 9. Variables de entrada para predicciones de demanda eléctrica

Referencia	Variables que usa
[65]	Temperatura, velocidad del viento, humedad y día de la semana
[66]	Temperatura, humedad y día de la semana
[67]	Temperatura, humedad, día de la semana y sociedad

4.3.2 Autocorrelación en la serie temporal. Se utilizó otra técnica de ingeniería de datos denominada “autocorrelación”, la cual es una medida estadística que describe la relación que existe entre una variable de la serie temporal y su versión retrasada en el tiempo, es decir, si una variable está correlacionada con su pasado o futuro, se nombra como “retrasos”, o “lags” por su traducción al inglés.

Por lo tanto, en la figura 38, figura 39 y figura 40 se evidencia la autocorrelación que existe en la serie temporal de producción de energía solar, energía eólica y demanda eléctrica, respectivamente. A la izquierda de la figura se muestra la correlación para un día completo (288 retrasos), mientras que en la parte derecha de la figura se graficó para cincuenta retrasos, de modo que se pudiera apreciar en mejor medida el patrón que sigue la autocorrelación. Este análisis permitió tener en cuenta cuántos retrasos se debían utilizar.

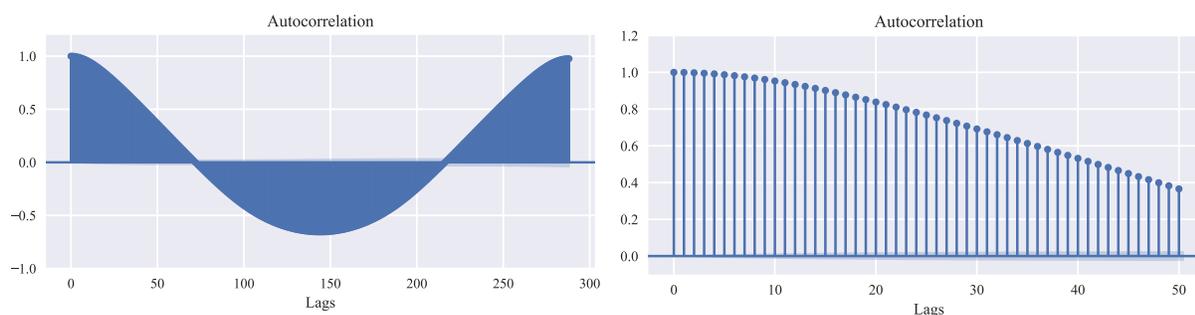


Figura 38. Autocorrelación de producción de energía solar. Fuente: Elaboración propia

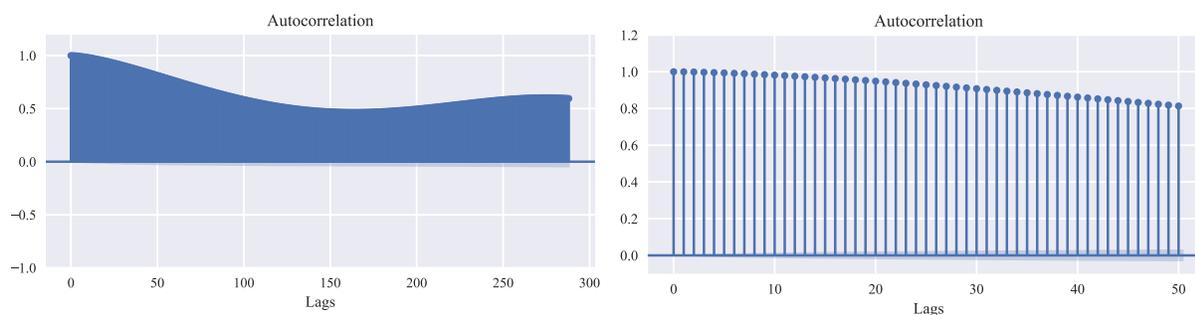


Figura 39. Autocorrelación de producción de energía eólica. Fuente: Elaboración propia

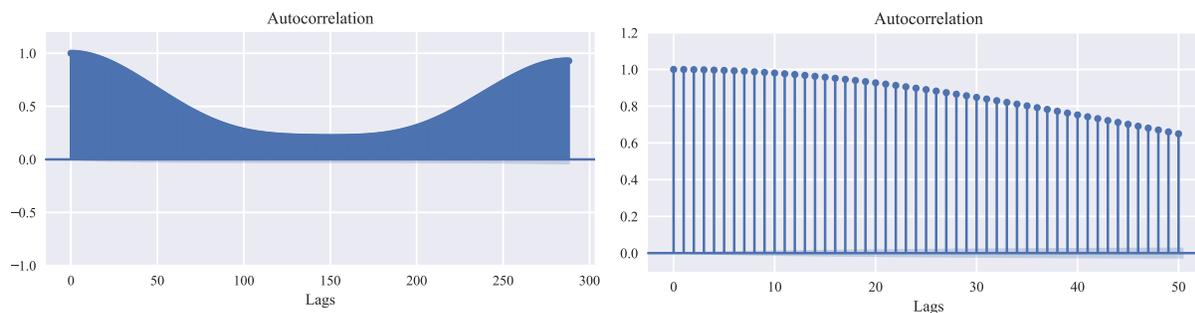


Figura 40. Autocorrelación de demanda eléctrica. Fuente: Elaboración propia

En estas figuras se encontraban 288 retrasos para cada una de las series temporales, y se evidenció que todos influyen en gran medida a los datos futuros del conjunto. Adicionalmente, otro método utilizado fue la autocorrelación parcial, la cual es una modificación que elimina la influencia de los valores intermedios de la serie temporal; es decir, mide la autocorrelación que existe entre la variable y su versión exacta retrasada en el tiempo. Mediante este método se generaron las figuras 41-43, que indican la autocorrelación parcial para la serie temporal de producción de energía solar, energía eólica y demanda eléctrica.

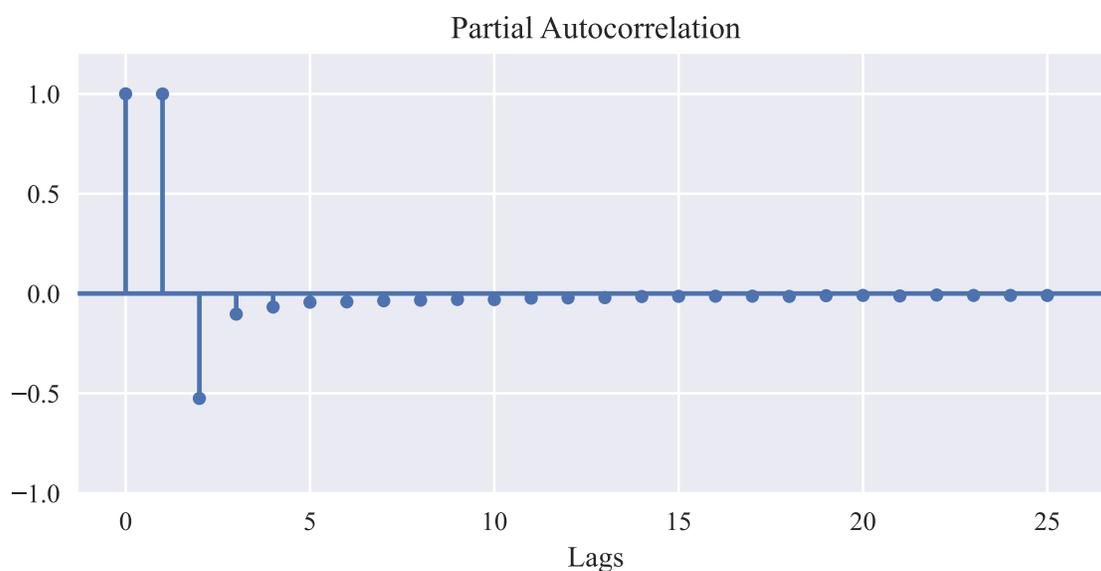


Figura 41. Autocorrelación parcial de producción de energía solar. Fuente: Elaboración propia

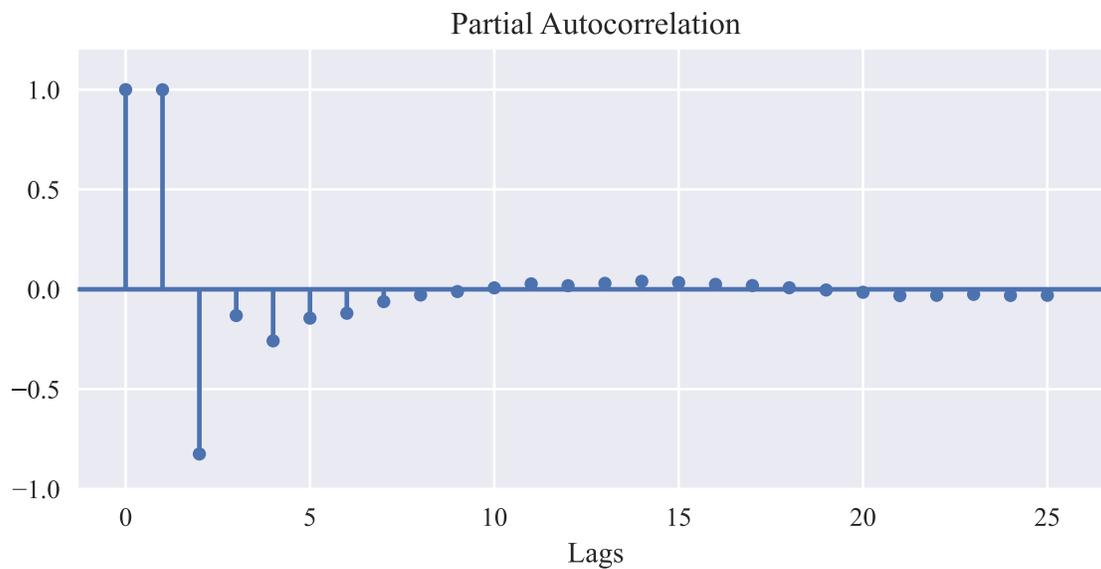


Figura 42. Autocorrelación parcial de producción de energía eólica. Fuente: Elaboración propia

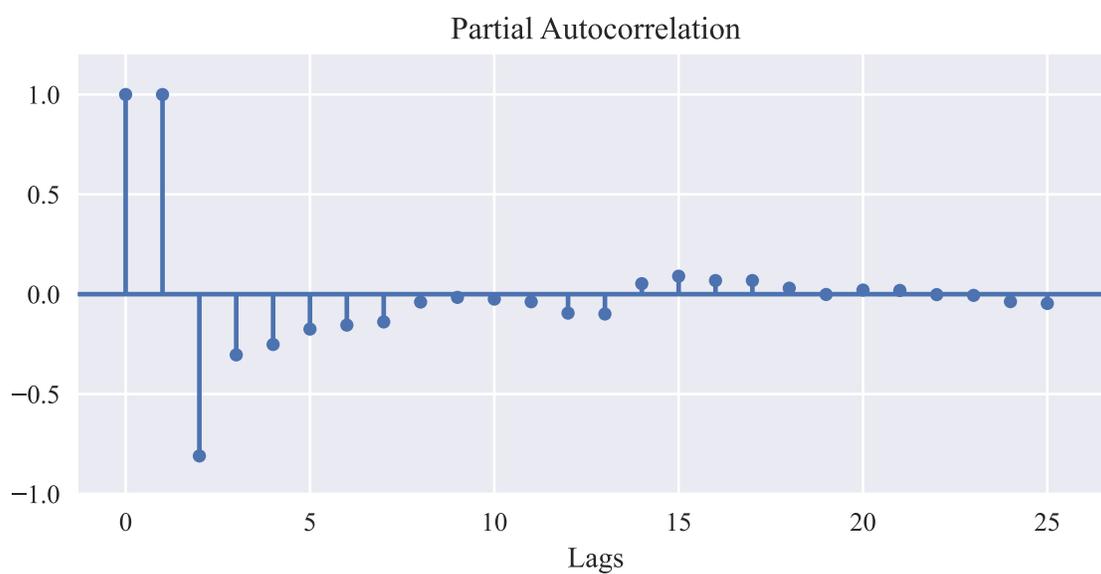


Figura 43. Autocorrelación parcial de demanda eléctrica. Fuente: Elaboración propia

De la misma forma, se evidenció que, en el caso de la autocorrelación parcial, los retrasos que más afectan a los valores futuros son los valores más recientes, entre el primer y el quinto retraso.

4.3.3 Correlación entre variables. Para la implementación del algoritmo de despacho se eligieron como variables de salida la producción fotovoltaica, la producción eólica y la demanda eléctrica, debido a que estos fueron los parámetros más relevantes para el despacho de energía en la microrred que se obtuvieron mediante los estudios presentados en secciones anteriores.

Para determinar cuáles son las variables de entrada necesarias para predecir la producción y demanda de energía se calculó el coeficiente de correlación de Pearson entre todas las variables el cual está definido mediante la ecuación 5:

$$r = \frac{\sum_i^N (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i^N (x_i - \bar{x})^2 \sum_i^N (y_i - \bar{y})^2}} \quad (5)$$

Donde N es el número de mediciones temporales, x_i y y_i corresponden a las variables a comparar, mientras que \bar{x} y \bar{y} son las medias de cada variable, respectivamente. Luego se graficó la matriz de correlación resultante utilizando el método heatmap de la librería Seaborn en Python. De esta forma, se graficó el mapa de calor que se muestra en la figura 44, donde se relacionaron todas las variables que se obtuvieron de las bases de datos de CAISO y NREL.

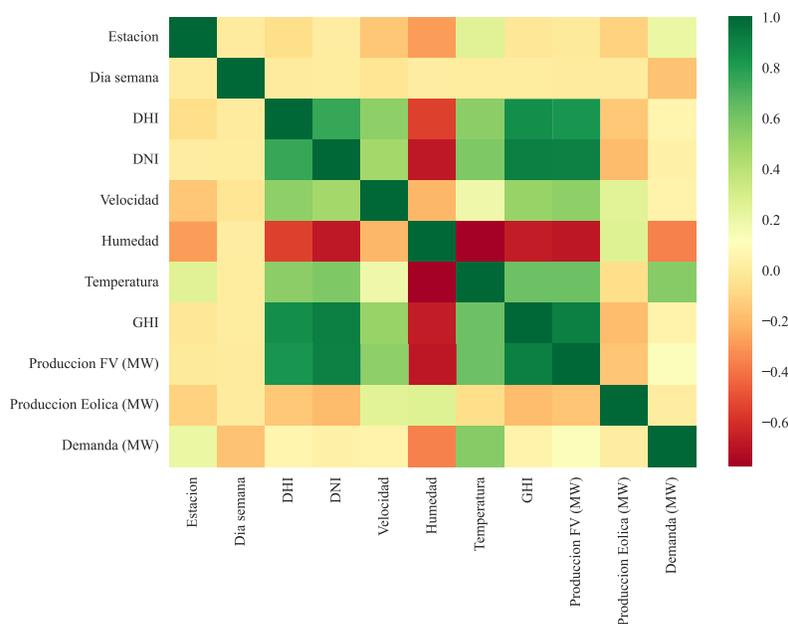


Figura 44. Mapa de calor de la base de datos. Fuente: Elaboración propia

Para determinar cuáles son las variables que se correlacionan es necesario comprender cómo funciona el mapa de calor. En el lateral derecho de la figura se observa una serie de colores que representan los datos dependiendo de la correlación con otra variable, los colores más fuertes significan que existe una mayor correlación, independientemente si es positiva o negativa. Con todo lo abordado en esta sección, se concluyó lo mostrado en la tabla 10.

Tabla 10. Definición de variables de entrada

Parámetros de salida	Parámetros de entrada
Producción fotovoltaica	DHI, DNI, GHI, humedad y temperatura
Producción eólica	Estación y velocidad del viento
Demanda eléctrica	Día de la semana, humedad y temperatura

4.4 Selección de las herramientas de desarrollo y simulación

En esta sección se describen las herramientas utilizadas para llevar a cabo el proyecto de investigación, las cuales incluyen lenguajes de programación y software especializado en la simulación de microrredes. A continuación, se describen las subsecciones necesarias desarrolladas para cumplir este objetivo.

4.4.1 Lenguaje de programación utilizado en las predicciones. Luego de seleccionar y explorar detalladamente las bases de datos se procedió a seleccionar el mejor lenguaje de programación para realizar las predicciones de la producción de energía y la demanda eléctrica. Para ello se consideraron tres lenguajes de programación que son conocidos para el análisis de series temporales: Python, R y MATLAB; sus características se muestran en la tabla 11.

Tabla 11. Características de los lenguajes de programación tenidos en cuenta

Características	Python	R	MATLAB
Puntos fuertes	Ciencia de datos	Análisis estadístico y visualización de datos	Matemáticas aplicadas
Licencia libre	Sí	Sí	No
Complejidad de uso	Baja	Baja	Media
Gama de bibliotecas para aprendizaje automático	Alta: Tensorflow, Scikit-Learn, Keras, Pandas, etc	Media-alta: Caret, MLT, H2O, etc	Media: Neural Network Toolkit, Statistics, etc
Decisión final	Seleccionado		

De acuerdo con la información de la tabla fue posible tomar la decisión de cuál sería el lenguaje para programar los algoritmos para predecir los parámetros de salida nombrados en la metodología del objetivo anterior. A continuación, se presenta la tabla 12, donde se evidencia la matriz de selección.

Tabla 12. Matriz de selección de los lenguajes de programación tenidos en cuenta

Características	Python	R	MATLAB
Puntos fuertes	5	4	3
Licencia libre	5	5	1
Complejidad de uso	5	5	3
Gama de bibliotecas para aprendizaje automático	5	4	3
Puntaje total	625	400	27
Lenguaje elegido	Seleccionado		

La decisión final fue tomada con respecto a la aplicación que se dio en este proyecto de investigación. Aunque Python y R son utilizados comúnmente para trabajos respectivos a aprendizaje automático, se concluyó que Python sería mejor para esta labor, dado a que cuenta con numerosas librerías y documentación.

4.4.2 Librerías de aprendizaje automático. Ya que se utilizó el lenguaje de programación Python fue posible utilizar las librerías disponibles para la manipulación de datos y modelamiento de algoritmos de inteligencia artificial, las cuales son:

- **Scikit Learn:** Esta librería permitió utilizar distintas técnicas de regresión y procesamiento de datos; fue diseñada para ser eficiente manejando grandes conjuntos de información y es ampliamente utilizada en el campo de ciencia de datos. De esta librería se utilizaron los siguientes algoritmos: DT, KNN y MLR, presentados en la sección de marco teórico.
- **XGBoost:** Su función es proveer métodos que permitan aplicar el algoritmo XGBoost a las funciones de la librería de Scikit Learn, gracias a esta librería fue posible combinar los métodos convencionales con métodos avanzados de decisión.
- **Skforecast:** Está especializada en tareas de pronóstico y predicción de series temporales, se utilizó como complemento a la librería Scikit Learn y XGBoost ya que como tal es una expansión de ellas. Permitió combinar diferentes técnicas y métodos de preprocesamiento, así como aplicar los retrasos a la serie temporal de manera efectiva. Se implementó el algoritmo SARIMA utilizando esta librería.
- **Tensorflow:** Es una librería orientada al aprendizaje automático y al procesamiento de datos, la cual se utilizó para construir y entrenar una red neuronal recurrente LSTM, que se caracteriza por poseer un parámetro considerado como memoria y permitió analizar de manera diferente la serie temporal.

De esta forma se escogieron los algoritmos de aprendizaje automático utilizados en este proyecto, en la tabla 13 se evidencia de manera detallada los algoritmos utilizados y a qué librería de aprendizaje automático le corresponden.

Tabla 13. Algoritmos de aprendizaje automático utilizados en el proyecto

Librería	Algoritmo utilizado
Scikit Learn	Decision Tree (DT), K-Nearest Neighbors (KNN) y Multi Linear Regression (MLR)
XGBoost	Extreme Gradient Boosting (XGBoost)
Skforecast	Seasonal Autoregressive Integrated Moving Average (SARIMA)
Tensorflow	Long Short-Term Memory (LSTM)

4.4.3 Software de simulación de la microrred. Se realizó el algoritmo de despacho de energía utilizando el lenguaje de programación Python, sin embargo, se buscó un software para comparar el rendimiento de este modelo, dentro de los cuales se encontraron:

- **HOMER Pro:** La cual es una herramienta para el diseño y análisis de sistemas de energías renovables, orientado a la evaluación económica y técnica de sistemas, así como de análisis de sensibilidad para los diferentes parámetros de un diseño.
- **HOMER Grid:** Es un software especializado en el diseño y análisis de sistemas que utilizan las energías limpias, orientado a implementarse en microrredes on-grid y off-grid, su función es la de optimizar la configuración de una red utilizando sistemas de almacenamiento de energía y controladores.
- **SAM:** Es un software de simulación y modelación de sistemas que se utiliza en la industria para analizar de manera técnica redes de energía, incluye módulos para simular la generación de energía solar, eólica y sistemas de concentración solar.
- **REopt:** Es una herramienta diseñada por NREL que se utiliza para identificar el tamaño y la configuración óptima de una red eléctrica en un edificio o una instalación, se enfoca en el ahorro de costos y minimización de emisiones de gases.

Las características de esos programas se compararon en la tabla 14, donde se evidencian los aspectos de cada software.

Tabla 14. Características de los software de simulación considerados en el proyecto

Características	HOMER Pro	HOMER Grid	SAM	REpot
Especialidad	Sistemas híbridos	Microrredes	Sistemas híbridos	Sistemas de edificios
Simulación	Sí	Sí	No	No
Orientado a microrredes	No	Sí	No	No
Modelado	No	No	Sí	Sí
Decisión final	Seleccionado			

Teniendo en cuenta estas características se desarrolló la tabla 15 que contiene la matriz de selección final.

Tabla 15. Matriz de selección de los programas de simulación tenidos en cuenta

Características	HOMER Pro	HOMER Grid	SAM	REpot
Especialidad	4	5	4	3
Simulación	5	5	3	3
Orientado a microrredes	1	5	1	1
Modelado	3	3	5	5
Puntaje total	60	375	60	45
Programa elegido	Seleccionado			

4.5 Definición e implementación de los modelos de aprendizaje automático

En esta sección se abordó la lógica que se utilizó en los modelos de inteligencia artificial, incluye la separación de los datos, la forma en la que se implementó y la manera en la que se evaluó el desempeño de los modelos.

4.5.1 División de los datos. Inicialmente, la división de la información se abordó de manera aleatoria por medio del método “train_test_split” de scikit-learn, el cual separa el conjunto de datos en dos partes: una que se utiliza para el entrenamiento del algoritmo y la otra que es la sección en la que se realiza la evaluación de las pruebas de funcionamiento. Sin embargo, los resultados obtenidos no fueron exitosos, ya que se estaban dividiendo los datos de manera aleatoria y para el caso de series temporales es necesario una separación cronológica de la información. Es por esta razón, que se utilizó una división respectiva para las series temporales de energía solar, energía eólica y demanda eléctrica, tal y como se muestra en las figuras 45-47, respectivamente.

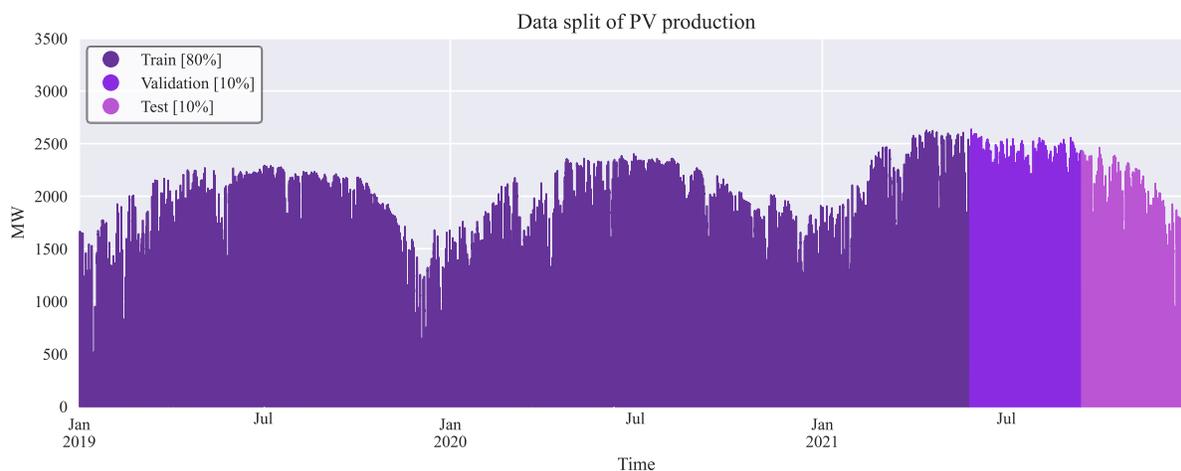


Figura 45. División de los datos de producción de energía solar. Fuente: Elaboración propia

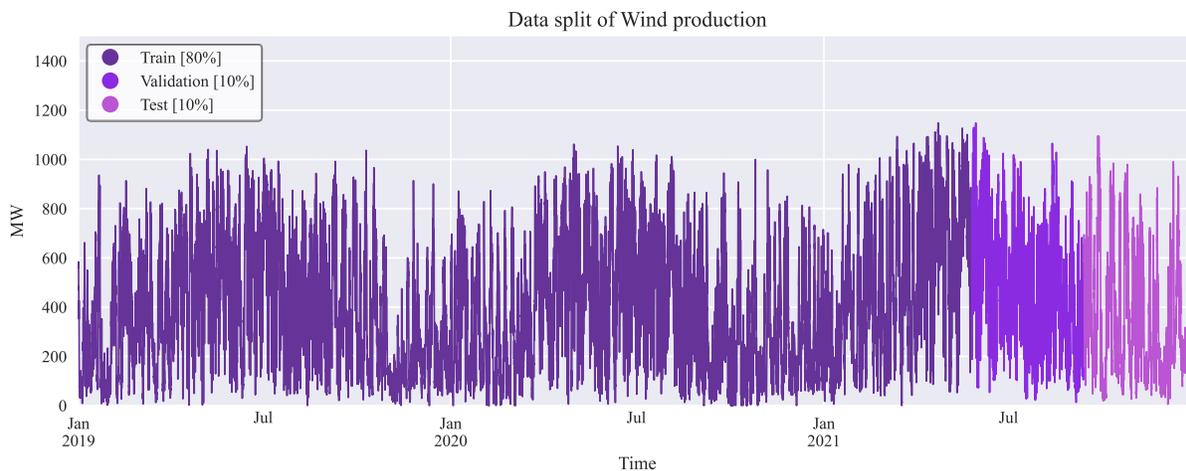


Figura 46. División de los datos de producción de energía eólica. Fuente: Elaboración propia

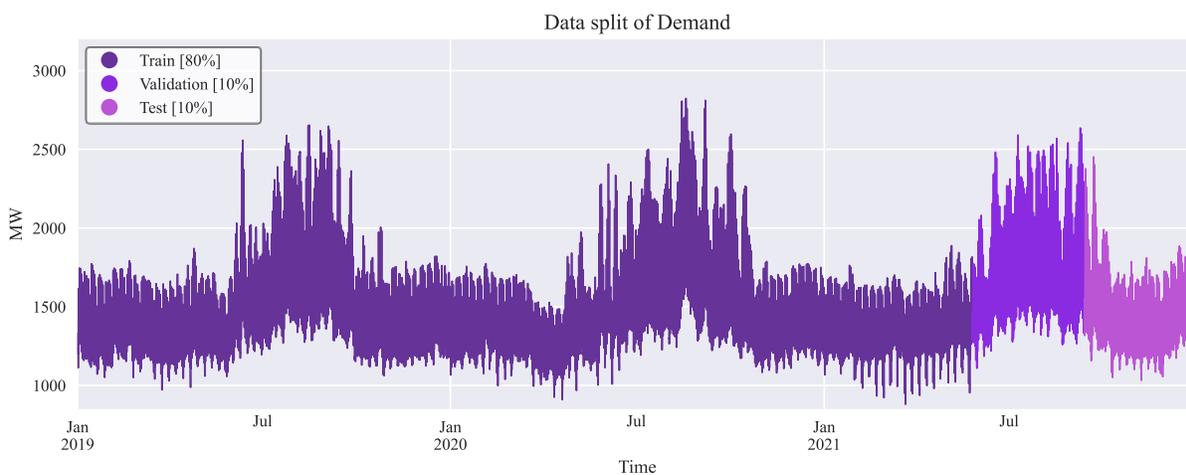


Figura 47. División de los datos de demanda eléctrica. Fuente: Elaboración propia

En las tres figuras anteriores se evidencia la manera correcta de dividir una base de datos para realizar predicciones futuras de series temporales. Las tres secciones en las que se separaron los datos se definen como:

- **Primer subconjunto:** Hace referencia a los datos que se utilizaron en cada modelo para su respectivo entrenamiento, que es igual al 80% del total de la información planteada en la base de datos inicial y se utilizó este tamaño para que se pudiera tener en cuenta la variabilidad de los datos.

- **Segundo subconjunto:** Esta sección fue muy importante para que el algoritmo realizara pruebas internas durante el entrenamiento (el 10% de los datos siguientes) y se ajustaran los hiperparámetros para encontrar el mejor modelo. De esta forma se evitó el sobreajuste en los datos de entrenamiento, más conocido como overfitting por su traducción al inglés.
- **Tercer subconjunto:** En este último subconjunto se tomó el 10% restante de los datos, con esta información se evaluó el desempeño final del modelo comparando el resultado con las métricas de error propuestas.

4.5.2 Definición de la lógica de los algoritmos. La lógica de los algoritmos se abordó de dos maneras distintas a lo largo del desarrollo del proyecto, inicialmente se utilizó la lógica planteada en la tabla 16.

Tabla 16. Pseudo código de la lógica inicial de programación

BEGIN
Import the CSV and save it into a DataFrame
Import the regressor's libraries
Split the dataset into train and test
WHILE true
 Fit the model using the train values
 Calculate the score of the model
 IF *score is higher than 90%*
 Break the infinite loop
 Make the predictions using the test values
 Calculate the error of the models
END

Este código no retuvo información de la serie temporal, por lo que no se estaba aprovechando todas las características que poseía la base de datos que se construyó, así que se

desarrolló la lógica de programación presentada en la tabla 17, la cual utiliza la librería Skforecast para emplear los parámetros de la serie temporal aplicando los retrasos y las variables exógenas.

Tabla 17. Pseudo código de la lógica utilizada en los regresores

BEGIN

Import the CSV and save it into a DataFrame

Import the regressor's libraries

Split the dataset into train, validation and test

Create the forecaster object

FOR *each parameter in the grid*

Tunning using cross validation

Train the forecaster object with the best parameters

Backtesting to make the predictions

Calculate the error of the models

END

A diferencia del anterior algoritmo, los resultados entregados por esta lógica fueron mejores, esto fue debido a que se realizó la búsqueda de los mejores hiperparámetros para cada uno de los modelos mediante validación cruzada utilizando el 90% de los datos, la cual es una técnica utilizada para evaluar resultados dividiendo el conjunto de datos, y posteriormente se realizaron las predicciones utilizando el modelo con la combinación de hiperparámetros cuyo error era el más bajo con el 10% restante de los datos.

4.5.3 Métricas utilizadas. Se implementaron tres diferentes métricas que analizan el error del modelo comparando las predicciones realizadas con los valores reales de esos mismos intervalos de tiempo; a continuación, se describen cuáles se utilizaron y cuáles son sus respectivas ecuaciones matemáticas:

- **Error absoluto medio (MAE):** Se define como la media de las diferencias absolutas de los valores predichos y los valores reales, cuando menor sea este error, mejor es la capacidad del modelo para realizar predicciones consistentes y precisas. Se define mediante la ecuación 6:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - x_i| \quad (6)$$

- **Error absoluto medio porcentual (MAPE):** Es una medida del error relativo promedio que comete el modelo en sus predicciones, se expresa en porcentaje y cuanto menor sea su valor, mejor es la capacidad del modelo para predecir valores precisos. El MAPE se encuentra dado por la ecuación 7:

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - x_i}{y_i} \right| * 100\% \quad (7)$$

- **Raíz del error cuadrático medio (RMSE):** Es una medida de la desviación estándar de los errores en las predicciones, cuando menor sea su valor, mejor es la capacidad del modelo para realizar predicciones correctas. La ecuación 8 que define el RMSE se presenta a continuación:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - x_i)^2} \quad (8)$$

Para las tres ecuaciones anteriores se cumple lo siguiente: n es el número de observaciones en los datos de prueba, y_i es el valor real de la observación en la posición i , y x_i es el valor predicho por el modelo en la posición i .

4.6 Evaluación del desempeño de los modelos en la microrred

En esta sección se aborda la manera en la que se evaluó el desempeño de la red en materia del algoritmo de despacho de energía comparando el rendimiento con el software seleccionado en secciones anteriores.

4.6.1 Funcionamiento de Homer Grid. El software seleccionado (HOMER Grid) posee una interfaz que permitió seleccionar la ciudad de la cual se deseaban tomar las tarifas, en este caso no fue posible seleccionar todo el estado de California ya que el software requiere tomar como referencia un sitio en específico. Para efectos de esta investigación, se tomó como referencia la ciudad de Sacramento en el estado de California (EEUU). La interfaz de Homer Grid se presenta en la figura 48.

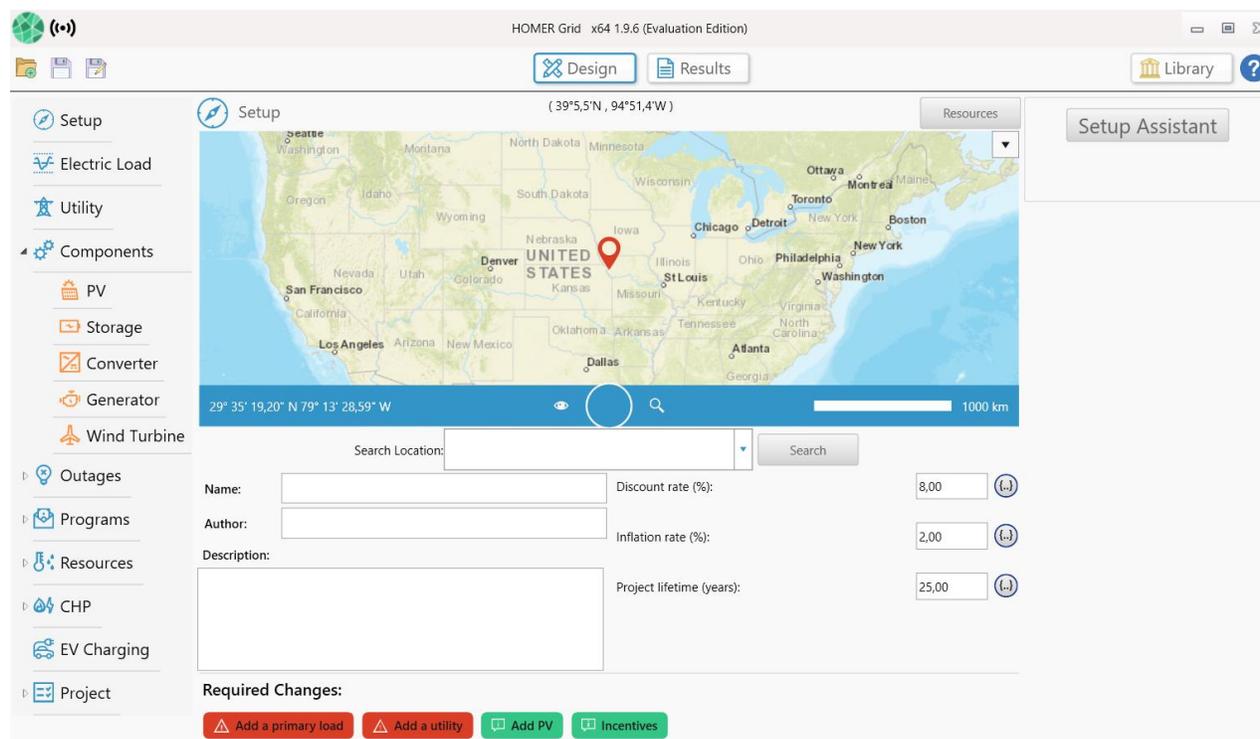


Figura 48. Interfaz del software Homer Grid. Fuente: Tomado de [32]

Se evidenció que el software permite definir el sistema de almacenamiento de energía, la cantidad de producción de energía solar, la demanda eléctrica, la tarifa de energía, entre otros parámetros útiles de la microrred.

4.6.2 Energía eólica en Homer Grid. Una de las limitaciones que presentó el software, es que no fue posible ingresar los datos de predicción de energía eólica generados por los modelos de aprendizaje automático, por lo tanto, fue necesario calcular la cantidad de potencia generadas durante las predicciones e ingresar turbinas eólicas que igualaran dichos valores. Lo anterior, a diferencia de la energía solar y demanda eléctrica que sí fue posible importar los datos predichos de la serie temporal en intervalos de cinco minutos durante un año.

4.6.3 Algoritmo de despacho. El algoritmo de despacho de energía fue elaborado utilizando el lenguaje de programación Python, dentro del cual se definieron las reglas que debía seguir en cada instante de tiempo para determinar si se tomaba energía de la red o directamente de las baterías, de la misma forma, se utilizó la librería MicroGrids para realizar un análisis económico de este modelo y comparar su rendimiento con el software comercial Homer Grid.

5. Resultados

En esta sección se presentan los resultados obtenidos en el proyecto de investigación. La ruta de resultados que se obtuvieron a lo largo del desarrollo se muestra en la figura 49.

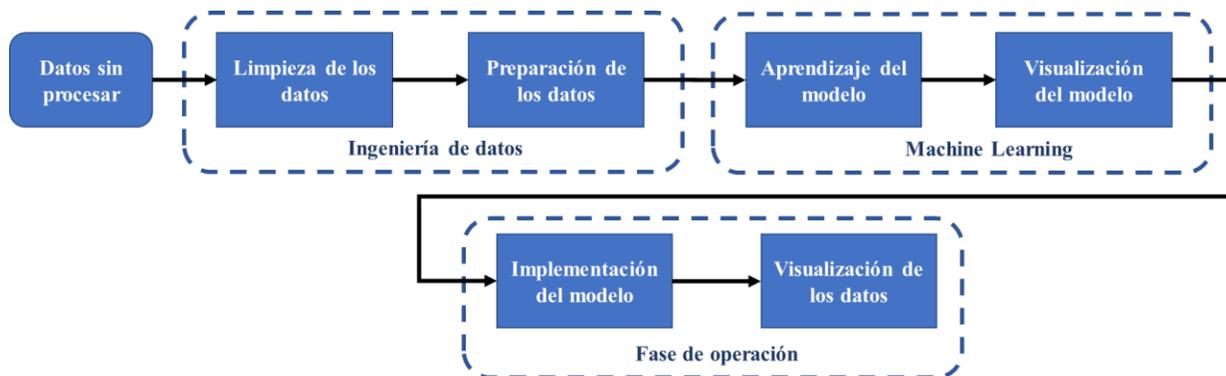


Figura 49. Ruta de resultados obtenidos. Fuente: Elaboración propia

5.1 Base de datos para algoritmos de aprendizaje automático

Durante el desarrollo del proyecto se descargaron los datos sin procesar de las entidades CAISO y NREL que contenían múltiples variables temporales desde el año 2018 a 2021, esta información pasó por un proceso de limpieza y depuración para corregir datos erróneos presentes en las mediciones, conduciendo a una base de datos unificada. Seguidamente, los datos procesados fueron preparados para ser aplicados en algoritmos de aprendizaje automático en formato de serie temporal con pasos de cinco minutos entre cada muestra.

En la tabla 18 se presentan todas las variables dentro de esta base de datos procesada, con su formato y descripción detallada para cada variable. Asimismo, en la tabla 19 se muestra una descripción técnica de algunas propiedades del conjunto de datos generados utilizando una funcionalidad de la librería Pandas que permite representar un conjunto de datos como DataFrame.

Tabla 18. Descripción de la base de datos

Variable	Descripción
Estación	Estación del año, 1: Invierno, 2: Primavera, 3: Verano, 4: Otoño
DHI	Irradiancia horizontal difusa
DNI	Irradiancia normal directa
Velocidad del viento	Velocidad del viento a través de cada instante de tiempo
Humedad	Valor de humedad relativa en porcentaje
Temperatura	Valor de temperatura en grados centígrados
GHI	Irradiancia global horizontal
Producción FV	Producción de energía solar cada cinco minutos
Producción Eólica	Producción de energía eólica cada cinco minutos
Demanda Eléctrica	Demanda de energía eléctrica cada cinco minutos

Tabla 19. Propiedades de la base de datos

Variable	Promedio	Desviación estándar	Valor mínimo	Percentil 25%	Percentil 50%	Percentil 75%	Valor máximo
Estación	NA	NA	1.00	NA	NA	NA	4.00
DHI	53.77	68.83	0.00	0.00	7.60	98.80	431.00
DNI	288.05	336.06	0.00	0.00	38.00	601.60	999.80
Velocidad del viento	2.54	1.04	0.66	1.76	2.24	3.16	8.54
Humedad	51.25	16.31	11.57	38.38	51.66	63.64	88.68
Temperatura	17.47	7.70	-0.54	11.20	16.70	22.80	39.02
GHI	221.78	296.63	0.00	0.00	10.20	427.20	1058.20
Producción FV	713.49	872.29	0.00	0.00	22.00	1590.40	2638.20
Producción Eólica	403.95	260.00	0.00	173.20	365.80	610.60	1148.60
Demanda Eléctrica	1490.02	291.57	879.72	1282.08	1432.74	1608.84	2824.02

De esta forma, en el anexo 1 se encuentran cien filas de la base de datos, a modo de demostración, y en el anexo 6 se encuentra el link de la base de datos obtenida disponible en línea. Estos datos fueron analizados utilizando distintas técnicas de visualización a lo largo de la

investigación, una de ellas fue mediante mapas de calor, los cuales permiten relacionar la producción y consumo de energía en unidades de potencia (MW) por cada hora del día y para cada día del año, tal y como se observa en las figuras 50-52.

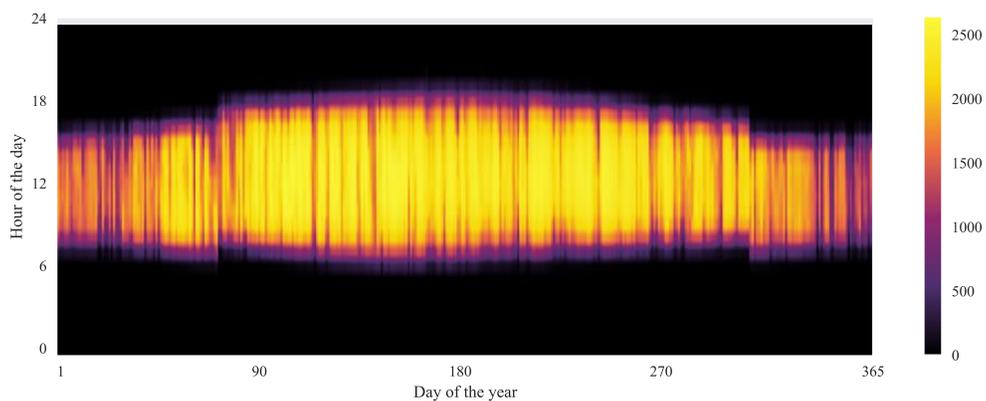


Figura 50. Cantidad de energía solar producida en MW en 2021. Fuente: Elaboración propia

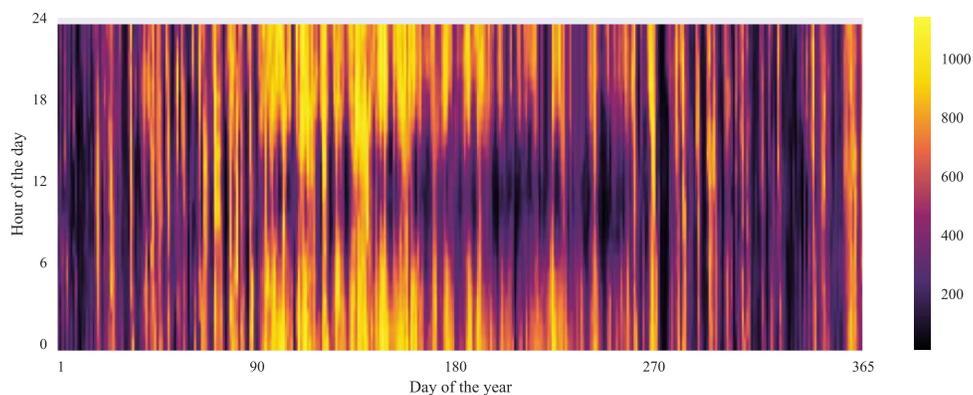


Figura 51. Cantidad de energía eólica producida en MW en 2021. Fuente: Elaboración propia

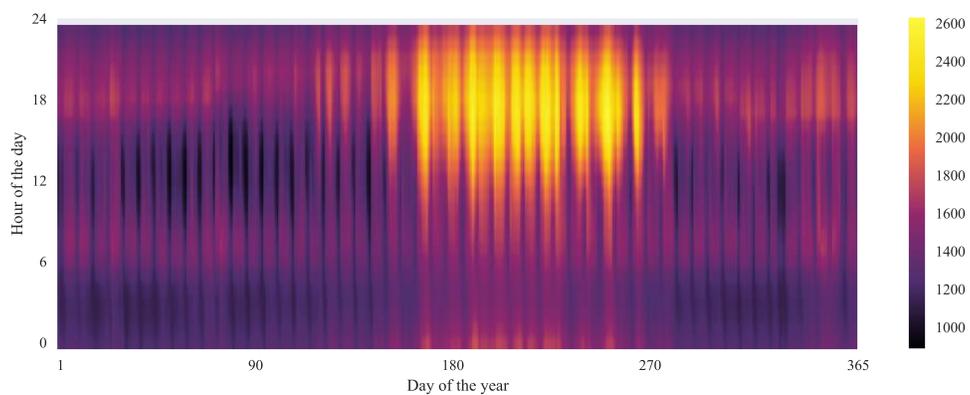


Figura 52. Cantidad de demanda eléctrica en MW en 2021. Fuente: Elaboración propia

5.2 Comparación de los algoritmos utilizados

Utilizando las métricas mencionadas en secciones anteriores, las cuales son MAE, MAPE y RMSE, se evidenciaron diferentes resultados para cada uno de los algoritmos presentados en la tabla 13 dependiendo si fueron utilizados para la predicción de demanda eléctrica, producción de energía solar o eólica. En primera instancia se presentan los resultados obtenidos para los algoritmos sin utilizar variables exógenas.

Se utilizaron los algoritmos de XGBoost, K-Nearest Neighbors y Decision Tree de la tabla 13, cuya implementación se realizó mediante las librerías Scikit Learn y XGBoost de Python. Se evaluó el desempeño de estos algoritmos para la predicción de energía solar, en donde se evidenció que Decision Tree presentó un mejor desempeño en comparación a los demás algoritmos, como se muestra en la tabla 20.

Tabla 20. Análisis inicial de los algoritmos de FV

Algoritmo utilizado	Métrica de error		
	MAE (MW)	MAPE (%)	RMSE
Decision Tree	136.57	22.10	245.47
K-Nearest Neighbors	156.75	49.36	288.10
XGBoost	142.46	54.52	236.72

Como se evidencia en la tabla 20, el MAPE fue alto para todos los modelos, a pesar de estandarizar los datos previamente utilizando la función StandardScaler. Los mismos experimentos se repitieron para la producción de energía eólica y demanda eléctrica, obteniendo resultados similares. Por lo tanto, se optó por utilizar la librería Skforecast, la cual provee funcionalidades para series de tiempo y sirvió como complemento a Scikit Learn y XGBoost para desarrollar algoritmos. Debido al elevado costo computacional se descartó el uso del algoritmo K-Nearest Neighbors (KNN) y en su lugar se utilizaron los algoritmos SARIMA y regresión lineal múltiple

(MLR) Ya que la librería Skforecast utiliza ventanas temporales que permiten optimizar el entrenamiento de modelos de aprendizaje, se obtuvieron mejores resultados en las predicciones. En las tablas 21-23 se presentan los errores resultantes para cada modelo implementado con ventanas temporales y sin utilizar variables exógenas para la predicción de energía solar, energía eólica y demanda eléctrica, respectivamente.

Tabla 21. Análisis de los algoritmos de FV sin variables exógenas

Algoritmo utilizado	Métrica de error		
	MAE (MW)	MAPE (%)	RMSE
SARIMA	107.01	52.08	332.71
Decision Tree	59.22	21.76	123.99
XGBoost	24.70	6.94	53.40
Multi Linear Regression	32.13	14.13	64.18

Tabla 22. Análisis de los algoritmos de energía eólica sin variables exógenas

Algoritmo utilizado	Métrica de error		
	MAE (MW)	MAPE (%)	RMSE
SARIMA	84.15	33.58	105.50
Decision Tree	61.58	24.98	88.42
XGBoost	19.94	6.59	31.39
Multi Linear Regression	19.54	6.73	30.83

Tabla 23. Análisis de los algoritmos de demanda eléctrica sin variables exógenas

Algoritmo utilizado	Métrica de error		
	MAE (MW)	MAPE (%)	RMSE
SARIMA	41.87	2.90	80.94
Decision Tree	45.83	3.15	63.23
XGBoost	9.20	0.62	13.08
Multi Linear Regression	7.13	0.49	10.44

En las tablas se presenta un buen desempeño en las métricas para los algoritmos de XGBoost y regresión lineal múltiple (MLR). Del mismo modo, se repitieron los experimentos con

ventanas temporales utilizando variables exógenas para la predicción de energía solar, energía eólica y demanda eléctrica y cuyos resultados se presentan en las tablas 24-26, respectivamente.

Tabla 24. Análisis de los algoritmos de FV con variables exógenas

Algoritmo utilizado	Métrica de error		
	MAE (MW)	MAPE (%)	RMSE
SARIMA	103.58	60.14	324.53
Decision Tree	59.82	18.23	126.26
XGBoost	24.33	6.70	52.84
Multi Linear Regression	32.01	13.29	63.21

Tabla 25. Análisis de los algoritmos de energía eólica con variables exógenas

Algoritmo utilizado	Métrica de error		
	MAE (MW)	MAPE (%)	RMSE
SARIMA	82.53	32.99	103.46
Decision Tree	61.01	24.64	85.57
XGBoost	19.82	6.56	31.34
Multi Linear Regression	19.35	6.62	30.55

Tabla 26. Análisis de los algoritmos de demanda eléctrica con variables exógenas

Algoritmo utilizado	Métrica de error		
	MAE (MW)	MAPE (%)	RMSE
SARIMA	41.59	2.89	80.57
Decision Tree	31.60	2.19	43.80
XGBoost	9.09	0.61	12.83
Multi Linear Regression	7.07	0.48	10.27

Para obtener estas métricas se entrenó el mejor modelo posible para cada uno de los casos, en decir, se utilizó la combinación óptima de hiperparámetros para cada algoritmo dentro de un rango de búsqueda especificado. El proceso para optimizar el conjunto de hiperparámetros se le denomina afinación o “tuning”, el cual se mediante validación cruzada revisa todas las

combinaciones de hiperparámetros posibles dentro de un rango establecido, de modo que se adapte a los datos que se están utilizando. En este caso, se utilizaron datos para entrenamiento y validación comprendidos desde el 1 de enero de 2019 al 12 de septiembre de 2021 para realizar la optimización de hiperparámetros mediante validación cruzada. Es así como, en las tablas 27-29 se presentan los mejores hiperparámetros encontrados para entrenar los modelos predictivos de energía solar, energía eólica y demanda eléctrica, respectivamente.

Tabla 27. Hiperparámetros de los modelos predictivos de FV

Algoritmo	Valores
SARIMA	<i>order = (0,1,3) seasonal order = (1,1,1,24)</i>
Decision Tree	<i>max_depth = 15 , max_features = 5 min_samples_leaf = 40 , min_samples_split = 40</i>
XGBoost	<i>learning_rate = 0.1 , max_depth = 5 n_estimators = 200</i>
Multi Linear Regression	<i>fit_intercept = True , n_jobs = None</i>

Tabla 28. Hiperparámetros de los modelos predictivos de energía eólica

Algoritmo	Valores
SARIMA	<i>order = (0,1,3) seasonal order = (1,1,1,24)</i>
Decision Tree	<i>max_depth = 13 , max_features = 5 min_samples_leaf = 50 , min_samples_split = 10</i>
XGBoost	<i>learning_rate = 0.1 , max_depth = 5 n_estimators = 200</i>
Multi Linear Regression	<i>fit_intercept = True , n_jobs = None</i>

Tabla 29. Hiperparámetros de los modelos predictivos de demanda eléctrica

Algoritmo	Valores
SARIMA	<i>order = (0,1,3) seasonal order = (1,1,1,24)</i>
Decision Tree	<i>max_depth = 10 , max_features = 10 min_samples_leaf = 30 , min_samples_split = 30</i>
XGBoost	<i>learning_rate = 0.1 , max_depth = 5 n_estimators = 200</i>
Multi Linear Regression	<i>fit_intercept = False , n_jobs = None</i>

De esta manera, se entrenaron los modelos con los hiperparámetros óptimos presentados en las tablas utilizando datos desde el día 1 de enero de 2019 al 25 de mayo de 2021 y cuya validación fue realizada con datos correspondientes al periodo desde el día 26 de mayo al 12 de septiembre del 2021. Finalizado el entrenamiento, se realizaron las predicciones desde el día 13 de septiembre hasta el 19 de septiembre del año 2021, que corresponde a la primera semana después de que terminan los datos de entrenamiento y validación, y que, por ende, son datos que no conocen los modelos. De este modo, en las figuras 53-56 se presentan las predicciones para los cuatro algoritmos implementados para la predicción de energía solar.

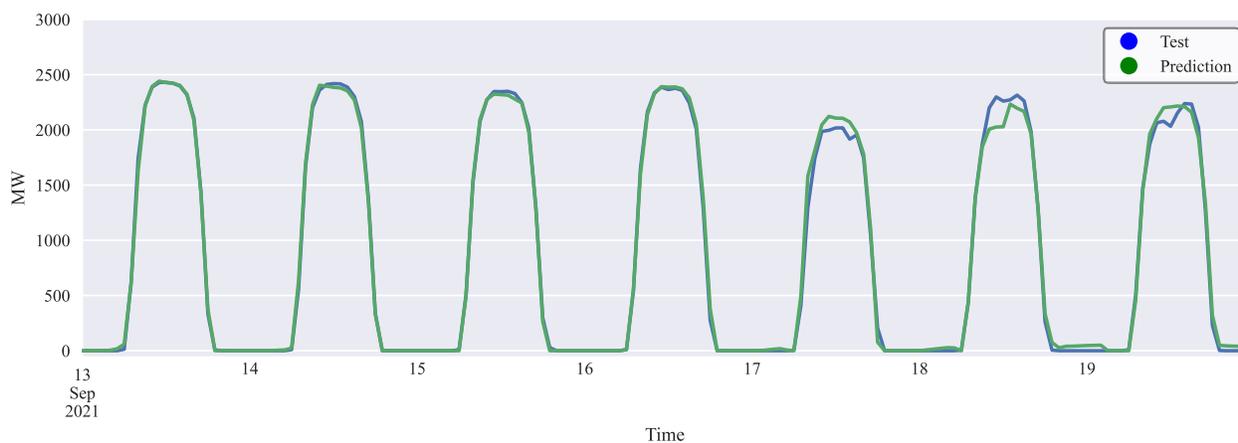


Figura 53. Predicción de energía solar con SARIMA. Fuente: Elaboración propia

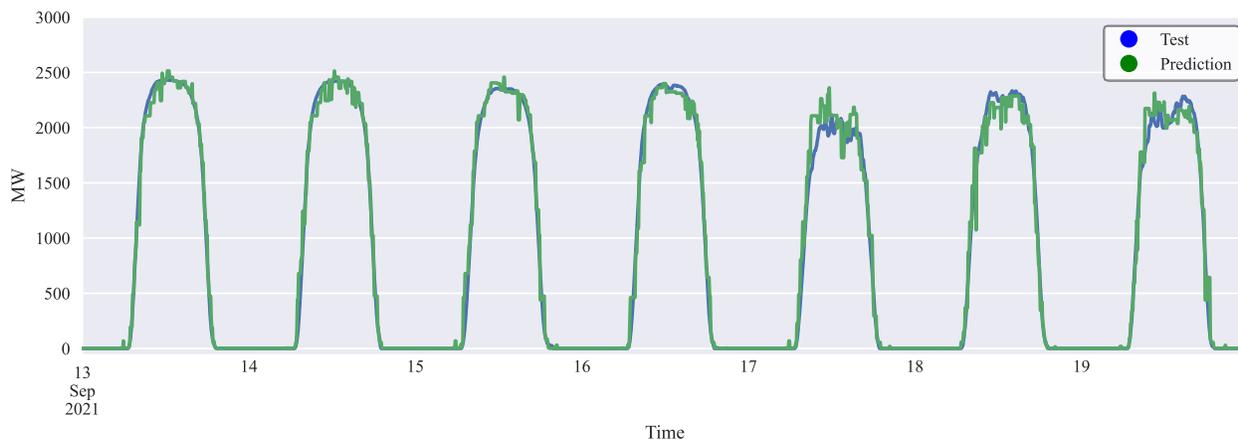


Figura 54. Predicción de energía solar con DT. Fuente: Elaboración propia

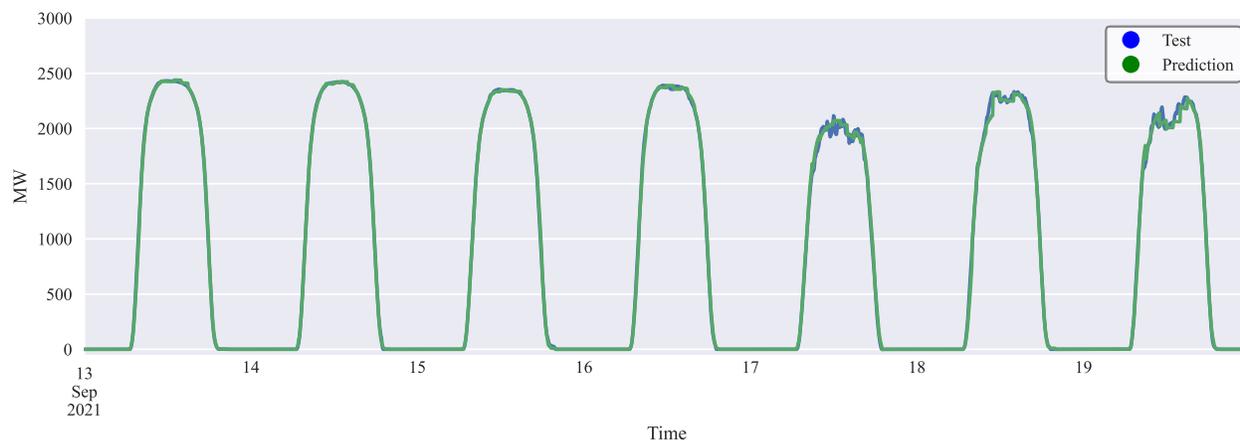


Figura 55. Predicción de energía solar con XGB. Fuente: Elaboración propia

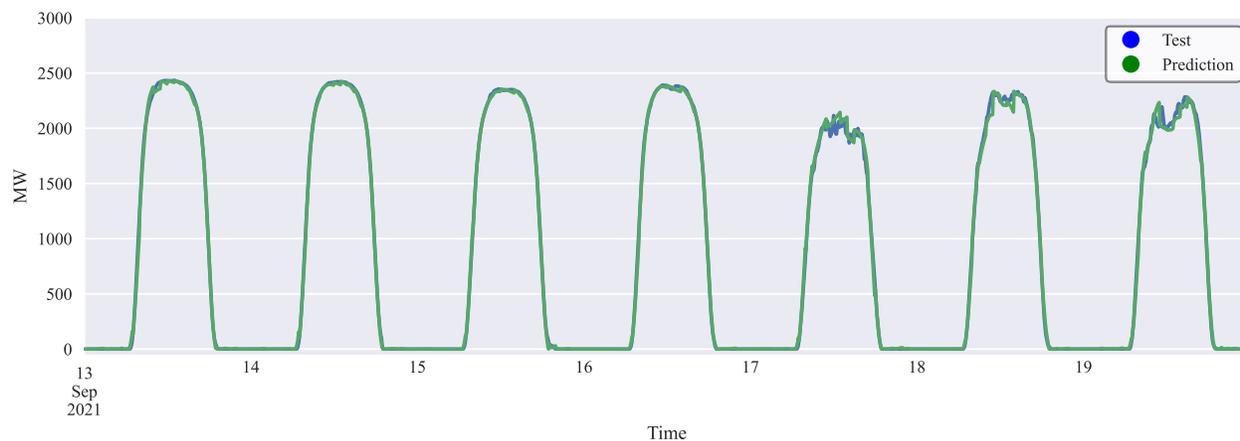


Figura 56. Predicción de energía solar con MLR. Fuente: Elaboración propia

Similarmente, en el mismo rango de fechas mencionadas anteriormente, se realizaron las predicciones para los de producción de energía eólica, como se muestra en las figuras 57-60.

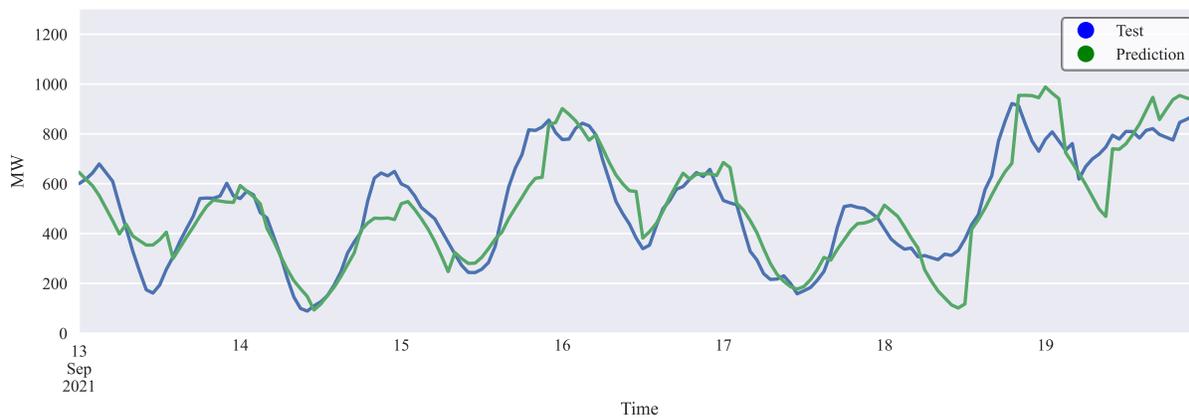


Figura 57. Predicción de energía eólica con SARIMA. Fuente: Elaboración propia

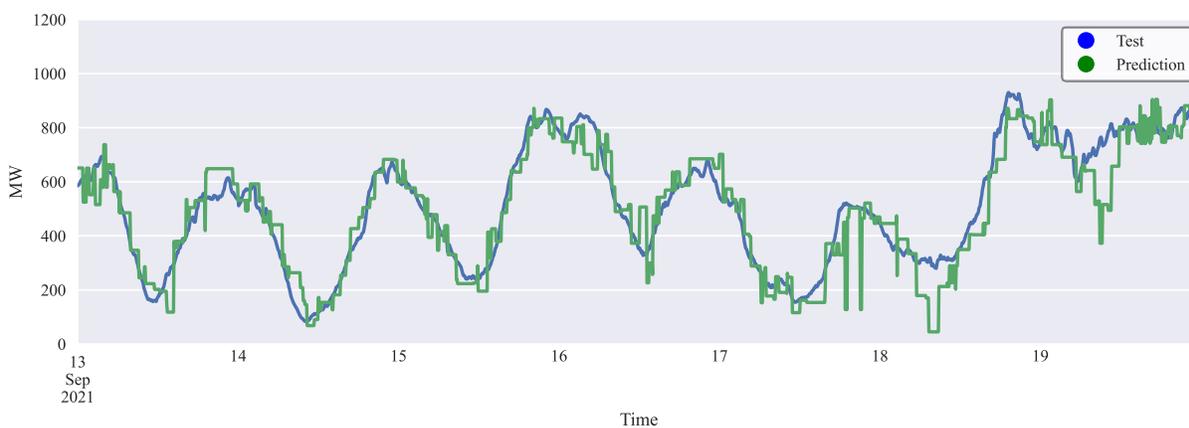


Figura 58. Predicción de energía eólica con DT. Fuente: Elaboración propia

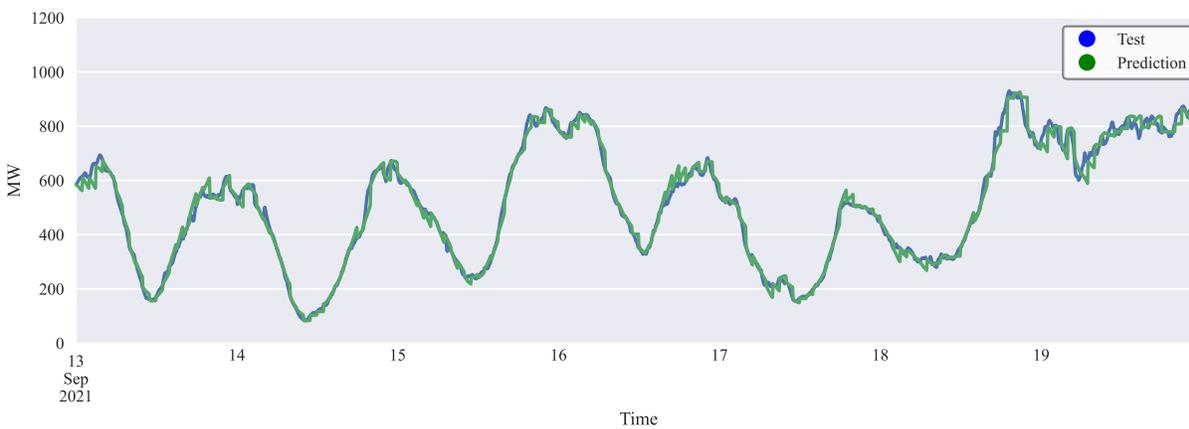


Figura 59. Predicción de energía eólica con XGB. Fuente: Elaboración propia

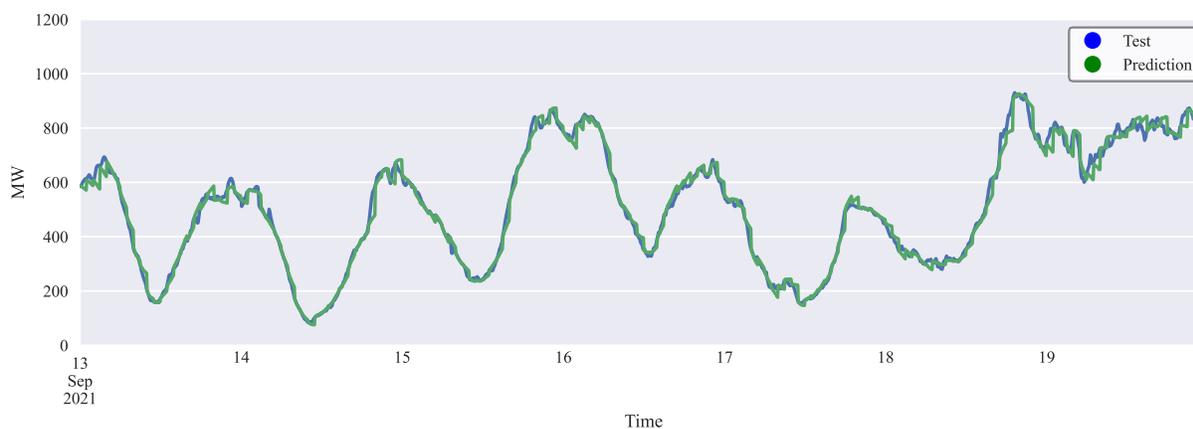


Figura 60. Predicción de energía eólica con MLR. Fuente: Elaboración propia

Finalmente, se realizaron las predicciones para el caso de demanda eléctrica con cada modelo dentro del mismo rango de fechas, las cuales se evidencian en las figuras 61-64.

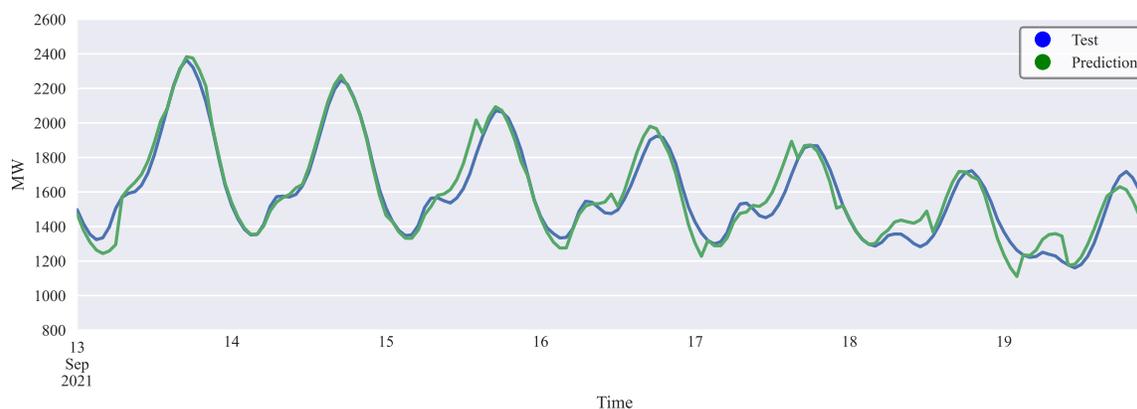


Figura 61. Predicción de demanda eléctrica con SARIMA. Fuente: Elaboración propia

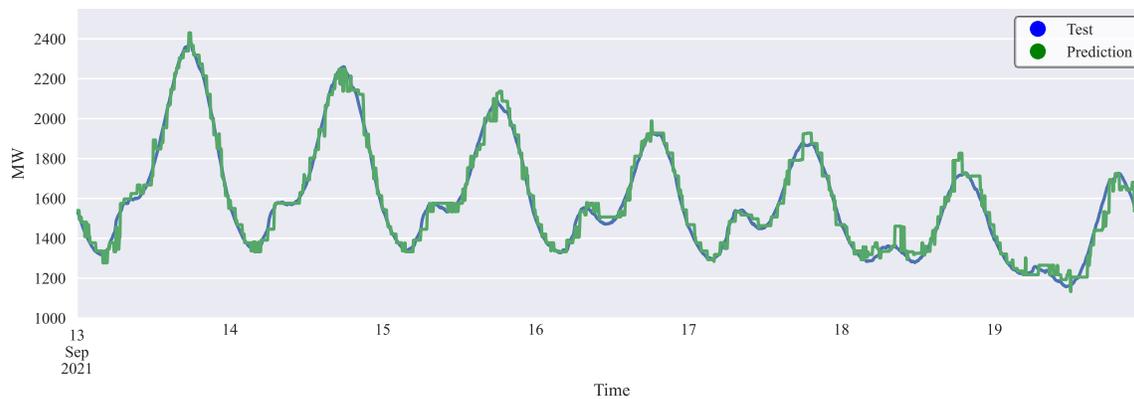


Figura 62. Predicción de demanda eléctrica con DT. Fuente: Elaboración propia

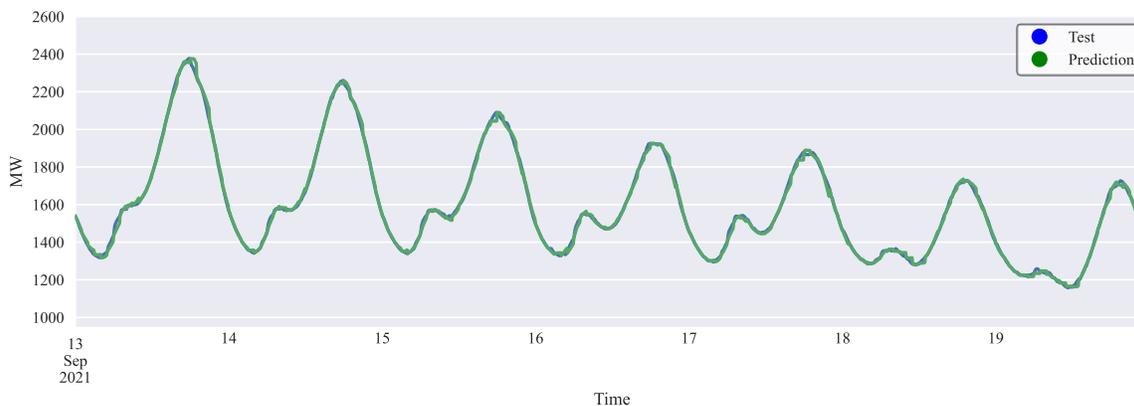


Figura 63. Predicción de demanda eléctrica con XGB. Fuente: Elaboración propia

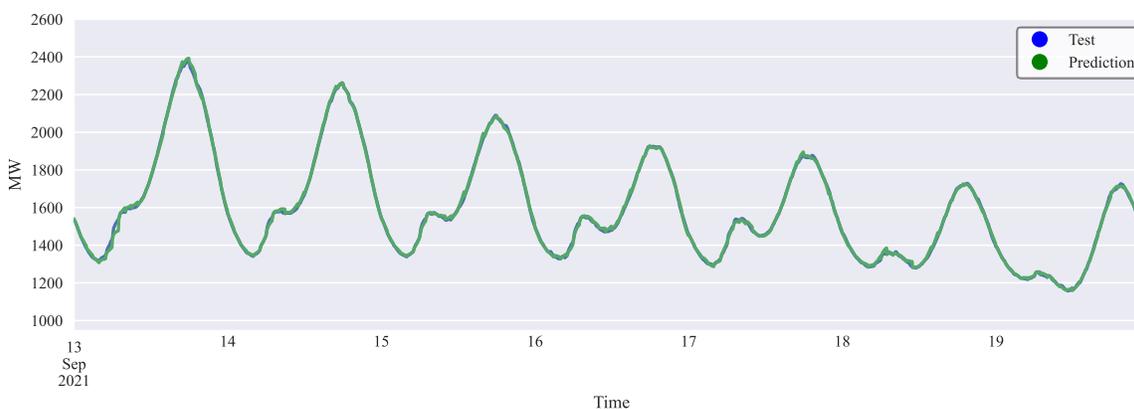


Figura 64. Predicción de demanda eléctrica con MLR. Fuente: Elaboración propia

De esta manera, se logró evaluar visualmente el desempeño de cada algoritmo de aprendizaje automático para la predicción de energía solar, energía eólica y demanda eléctrica y cuyos resultados numéricos se muestran en las tablas 24-26. El proceso de predicción se extendió para las fechas desde el 13 de septiembre al 31 de diciembre del 2021 (alrededor de tres meses y medio). Estas predicciones se almacenaron en un archivo para ser utilizados en las siguientes secciones donde se combinan y evalúan otros aspectos de los modelos de aprendizaje.

5.3 Integración de los modelos de predicción

Se utilizó una técnica de combinación que consiste en entrenar un nuevo modelo con los resultados de los algoritmos previamente entrenados para generar nuevas predicciones. Por lo

tanto, se tomaron como datos de entrenamiento series de tiempo desde el 13 de septiembre del 2021 hasta el 31 de noviembre de 2021, de modo que se realizaran predicciones del mes de diciembre de ese mismo año. Se implementó un algoritmo de regresión combinado AdaBoost utilizando la clase AdaBoostRegressor de Scikit-learn, el cual se utilizó para mejorar la precisión de los modelos de regresión, ya que otorga una ponderación a cada regresor (Decision tree, XGBoost, etc.) en cada iteración y según cuál tenga un mayor “peso” o mayor importancia será el modelo que decida cuál es el valor futuro en ese instante de tiempo. La estructura general del modelo de regresión combinado se evidencia en la figura 65.

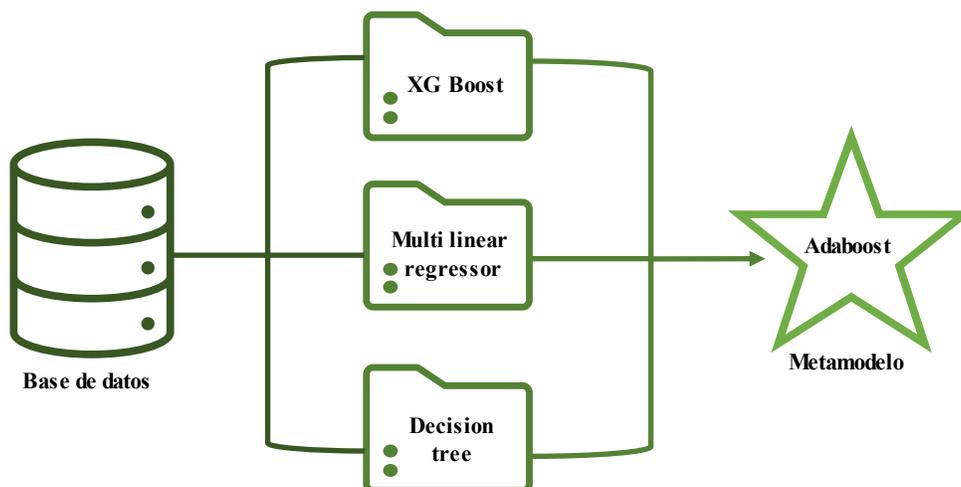


Figura 65. Estructura de integración de los algoritmos. Fuente: Elaboración propia

De esta forma, se evaluó el algoritmo AdaBoost para la predicción de energía solar, energía eólica y demanda eléctrica para todo el mes de diciembre del 2021 (1 al 31 de diciembre). El modelo AdaBoost puede combinarse con estos estimadores para realizar las predicciones de los tres algoritmos, de esta forma, en la tabla 30 se presentan los porcentajes de error MAPE de cada combinación posible, los cuales permitieron analizar estos modelos de una manera técnica.

Tabla 30. Métrica MAPE de los modelos AdaBoost combinados

Serie temporal	MAPE (%)		
	Decision Tree	Multi Linear Regressor	XGBoost
Energía solar	5.619	5.174	5.308
Energía eólica	7.040	5.641	6.325
Demanda eléctrica	0.550	0.524	0.520

De esta forma, la tabla 30 indica que en el caso de energía solar y eólica es óptimo combinar AdaBoost con regresión lineal múltiple, mientras que en el caso de demanda eléctrica es más eficiente utilizar XGBoost. Aplicando esta técnica y prediciendo para el día 1 de diciembre del 2021 se obtienen las gráficas mostradas en las figuras 66-68 para los casos de predicción de energía solar, energía eólica y demanda eléctrica, respectivamente.

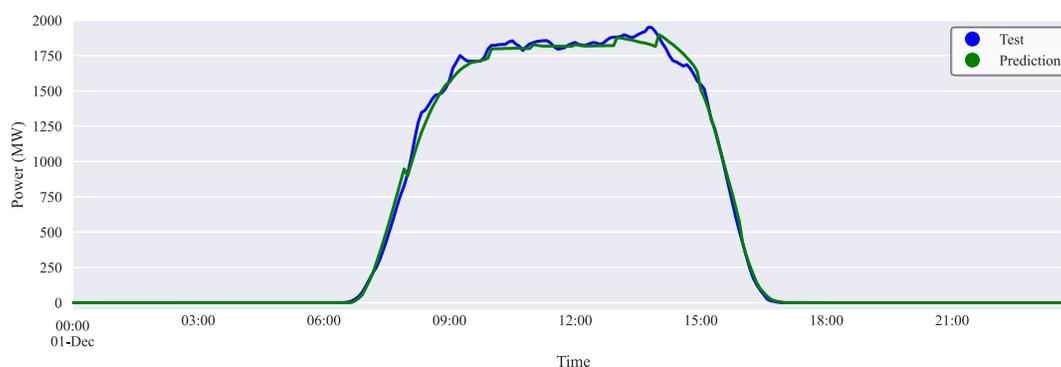


Figura 66. Predicción de energía solar con AdaBoost y MLR. Fuente: Elaboración propia

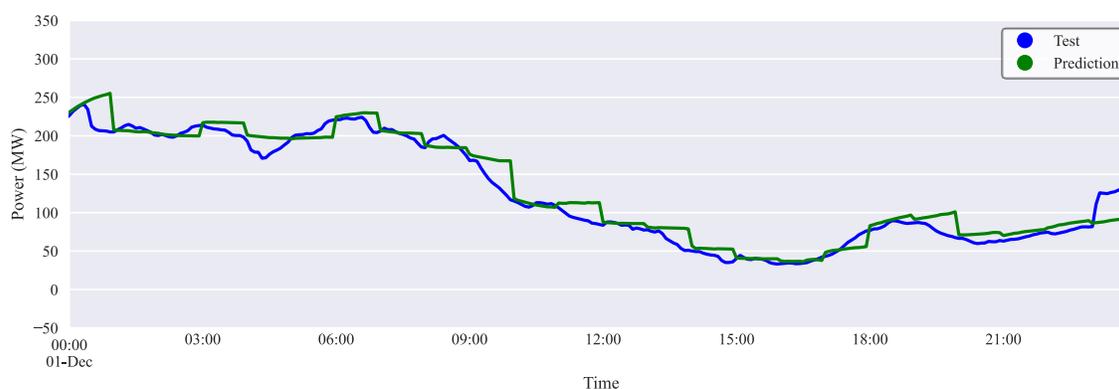


Figura 67. Predicción de energía eólica con AdaBoost y MLR. Fuente: Elaboración propia

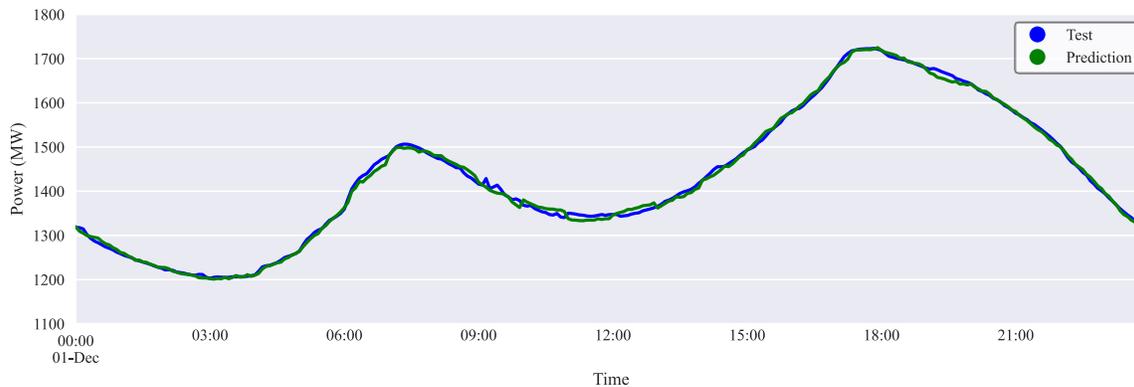


Figura 68. Predicción de demanda eléctrica con AdaBoost y XGBoost. Fuente: Elaboración propia

Cabe resaltar que la combinación de un algoritmo específico con AdaBoost no descarta las predicciones realizadas por el resto de los algoritmos, al contrario, todos se tienen en cuenta y posteriormente AdaBoost en conjunto con regresión lineal múltiple o XGBoost mejora la predicción.

5.4 Predicción usando redes neuronales

A modo de comparación y exploración, en esta sección se entrenaron diversos tipos de redes neuronales para realizar la predicción de energía solar, energía eólica y despacho de energía eléctrica. Específicamente, se entrenaron redes neuronales multicapa (lineales y densas), redes neuronales de convolución (CNN) y redes neuronales recurrentes (LSTM). A diferencia del modelo ensamblado presentado anteriormente, las redes neuronales presentadas en esta sección tienen la habilidad de los valores de energía solar, energía eólica y demanda energética de manera simultánea.

A continuación, se detalla brevemente la metodología utilizada para entrenar estas redes neuronales. Similar a lo presentado secciones anteriores, para cada uno de los modelos, el conjunto de datos procesado de CAISO y NREL fue particionado en subconjuntos de entrenamiento (80%), validación (10%) y evaluación (10%). Los datos de entrenamiento comprenden las fechas desde

enero 1 del 2019 hasta el 25 de mayo de 2021, mientras que el conjunto de validación desde el 26 de mayo al 12 de septiembre de 2021, y el de prueba desde 13 de septiembre al 31 de diciembre de 2021. Asimismo, todos los datos fueron estandarizados del mismo modo que para el modelo ensamblado; es decir, los datos entrenamiento, validación y evaluación fueron restados por el promedio y divididos por la desviación estándar calculada de los datos de entrenamiento. Todos los modelos en esta sección fueron entrenados utilizando GPU con un máximo de 50 épocas y parada temprana del entrenamiento (early stopping), activada por un incremento en el error de validación. Cabe destacar que las redes neuronales tardan más tiempo en ser entrenadas que los modelos presentados en secciones anteriores (Decision Tree, XGBoost, etc.); por lo cual, para estos casos no fue posible realizar la optimización de hiperparámetros con validación cruzada debido al elevado costo computacional.

Para realizar el entrenamiento de las redes neuronales para la predicción de series temporales fue necesario utilizar ventanas temporales de n pasos de ancho para predecir m valores en el futuro. Como se mencionó en secciones anteriores, las series de tiempo de los conjuntos de datos poseen puntos temporales espaciados cada 5 minutos, por lo que, para el entrenamiento, la validación y la evaluación se utilizaron ventanas temporales de 2 días ($n=576$ pasos) para predecir datos de 1 hora en el futuro ($m=12$ puntos). A los modelos entrenados bajo este régimen se les conoce como modelos predictivos de múltiples pasos (en inglés, multi-step models). Los modelos que realizan este tipo de tareas requieren aprender a predecir un rango de valores futuros, así que, en esencia, son capaces de predecir una secuencia de valores futuros. Dentro de esta categoría existen dos tipos de modelos, los que predicen todos los valores a la vez (single-shot) y los modelos autorregresivos, que predicen un único valor que a su vez es retroalimentación a la entrada para realizar nuevas predicciones (similar a los modelos utilizados en secciones anteriores). En este apartado se implementaron modelos single-shot, debido a que el tiempo de cómputo es menor

con respecto a los modelos autorregresivos. Algunos resultados de predicciones sobre una ventana de evaluación de 2 días para cada modelo pueden observarse en los anexos 2-5.

Del mismo modo, las gráficas presentadas en las figuras 69-71 muestran la comparación de errores normalizados de validación y evaluación para las redes neuronales multicapa, la red neuronal convolucional (CNN) y la red neuronal recurrente (LSTM).

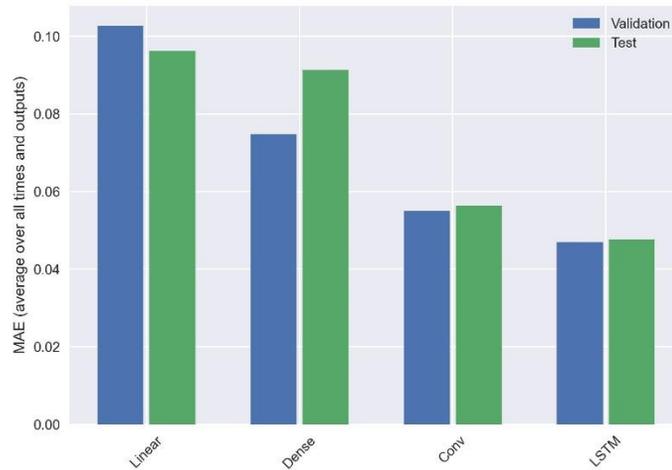


Figura 69. Comparación del MAE en datos de prueba para cada tipo de red neuronal implementado. Fuente: Elaboración propia

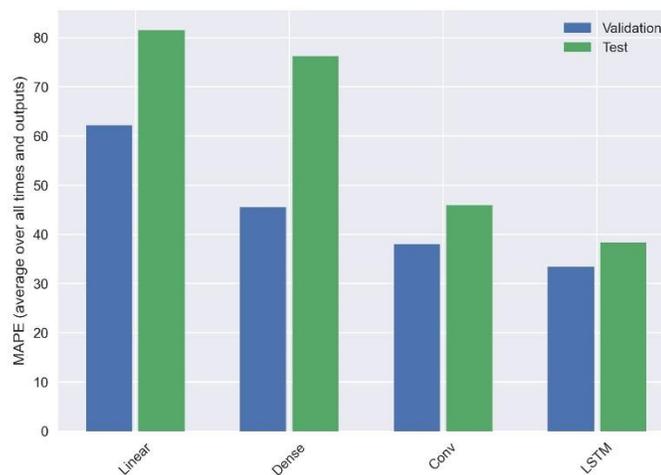


Figura 70. Comparación del MAPE en datos de prueba para cada tipo de red neuronal implementado. Fuente: Elaboración propia

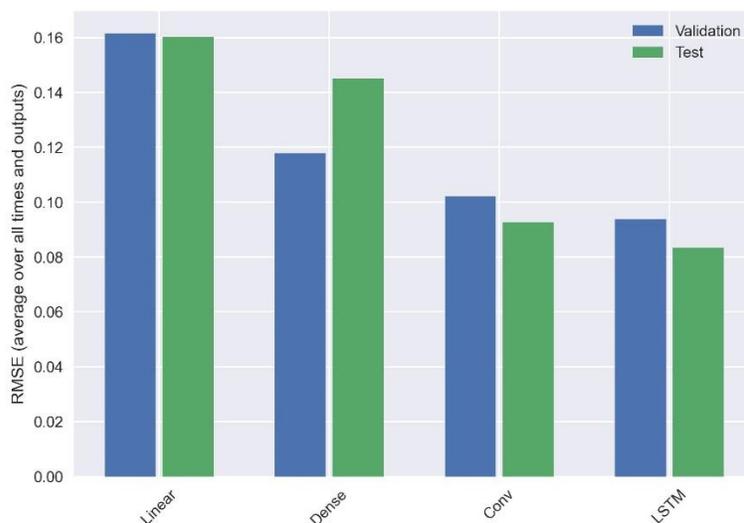


Figura 71. Comparación del RMSE en datos de prueba para cada tipo de red neuronal implementado. Fuente: Elaboración propia

Como se observa en las figuras, la red neuronal recurrente se desempeñó mejor que el resto de las redes implementadas, esto debido a su naturaleza de aprender de secuencias de datos temporales. Del mismo modo, se en la tabla 31 se presentan los errores ponderados para cada modelo obtenidos tras predecir las tres variables (energía solar, energía eólica y demanda eléctrica) de manera simultánea. Al igual que los resultados presentados en las figuras 69-71, se observa que la red LSTM son mejores para este tipo de tareas.

Tabla 31. Métricas de error en datos de evaluación para cada red neuronal

Modelo	Errores (normalizados)		
	MAE	MAPE (%)	RMSE
Red neuronal lineal	0.096	81.507	0.160
Red neuronal densa	0.091	76.248	0.145
Red neuronal convolucional	0.056	45.949	0.093
Red neuronal recurrente (LSTM)	0.048	38.273	0.083

Similar que, a lo ocurrido en el modelo ensamblado, la predicción de energía eólica representa un mayor reto para todas las redes neuronales debido a su naturaleza aleatoria dado que no posee curvas tan representativas como en el caso de la energía solar o demanda eléctrica. Los resultados mostrados en esta sección demuestran que las redes neuronales son una alternativa viable para implementar sistemas de despacho de energía.

5.5 Algoritmo de despacho de energía eléctrica

En esta sección, las predicciones realizadas por el algoritmo combinado fueron utilizadas para modelar el sistema de despacho de energía dentro de la microrred. Para este trabajo de investigación, se modeló el funcionamiento de la microrred por medio de simulación. La estructura de la microrred utilizada se estableció de la siguiente manera: se definieron una línea de energía AC y otra DC en donde los componentes de la microrred se conectan dependiendo de su funcionamiento, esta definición se aprecia en la figura 72. Adicionalmente, las unidades de potencia para la producción y consumo de energía fueron convertidas de MW a kW, de modo que se pudiera usar una batería de tamaño real.

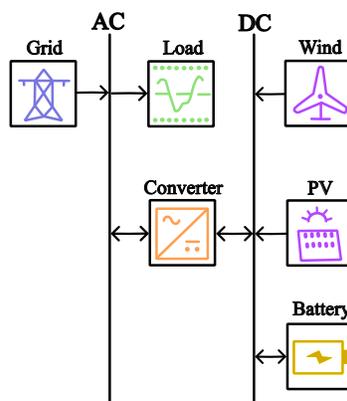


Figura 72. Estructura de la microrred simulada. Fuente: Adaptado de [32].

De la figura 72 se evidencia que la microrred se encuentra conectada a la red, de forma que puede también actuar como fuente de energía en caso de que las baterías y la producción de

energía renovable no cumpla con la demanda energética. Para la transformación de energía DC a AC dentro de la microrred, se contó con un convertidor genérico (inversor / rectificador) cuya eficiencia se estimó en 98.6%. Asimismo, se utilizaron dispositivos de almacenamiento de energía para cumplir con el despacho de manera más eficiente cuando la demanda energética es mayor, la capacidad total de las baterías usadas fue de 20 kWh. Los paneles solares y turbinas eólicas conectadas a la microrred generan energía que es convertida mediante el inversor para satisfacer la carga. El algoritmo fue diseñado teniendo en cuenta que la energía que proviene de los paneles solares y de las turbinas eólicas fuera aprovechada siempre, mientras que la energía de las baterías se utilizara cuando en la demanda existieran picos de alto consumo, de este modo se disminuye el costo a largo plazo de la energía que es comprada a la red eléctrica.

El algoritmo de despacho fue realizado en el lenguaje de programación Python, el cual toma decisiones en cada instante de tiempo para determinar si se distribuye energía de la red eléctrica, si se utilizan las baterías o si se emplea energía de fuentes renovables. Asimismo, se estipuló que la energía almacenada en la batería no cayera por debajo del 20% de su capacidad total, ya que según los antecedentes esto acorta su vida útil en gran medida. En la figura 73 se evidencia la forma en la que se distribuye energía a través de la microrred simulada.

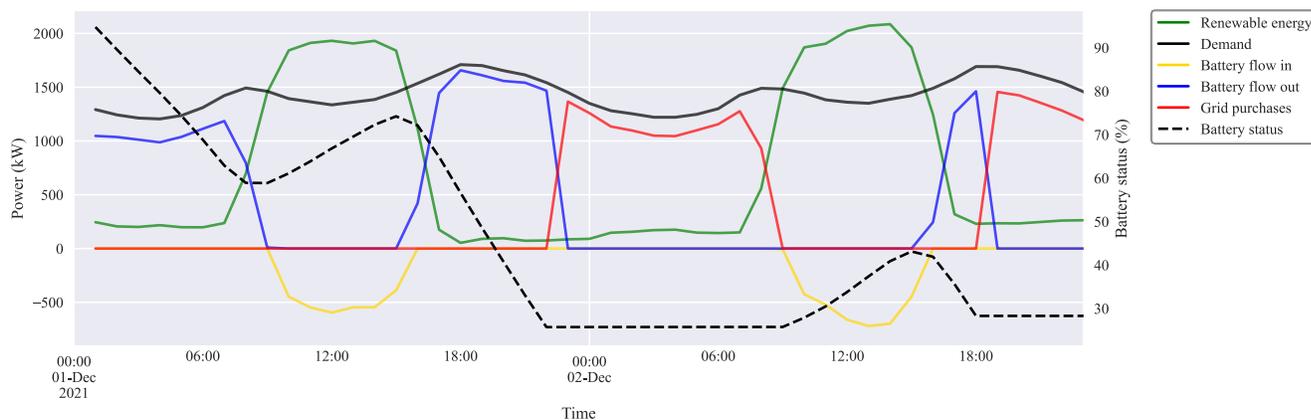


Figura 73. Despacho de energía en la microrred. Fuente: Elaboración propia.

El objetivo de este algoritmo es de proveer energía en los picos más altos de demanda para disminuir el costo generado en esos instantes de tiempo. De esta forma, fue posible aprovechar la cantidad de energía renovable suministrada a la microrred de manera eficiente utilizando la red eléctrica solo en los momentos requeridos, optimizando así el retorno de inversión de la microrred. Los puntos de operación se pueden describir de la siguiente manera: El sistema utilizará la energía de la batería en los puntos donde el pico de la demanda sea mayor en cada día (después de las 6:00pm), no permitirá que la energía de las baterías disminuya por debajo del 30%, y solo se comprará energía en los puntos donde la tarifa sea menor.

5.6 Evaluación del desempeño del algoritmo de despacho

Para la evaluación del desempeño del algoritmo de despacho se utilizó la librería MicroGrid, la cual cuenta con una función que para estimar la métrica del costo nivelado de la energía (Levelized Cost Of Energy, LCOE por sus siglas en inglés), considerada como una medida importante para evaluar la viabilidad económica de los proyectos que utilicen generación de energías renovables. Para calcular el costo nivelado de la energía en la microrred se requiere de la cantidad de años que se desean tener en cuenta para el análisis, así que se optó hacer el cálculo asumiendo 25 años de uso de la microrred. Esta métrica tiene en cuenta todos los gastos económicos durante ese periodo de tiempo y lo divide entre el total de energía producida durante ese mismo intervalo de fechas.

Utilizando los datos predichos para realizar el algoritmo de despacho se encontró que el costo nivelado de la energía fue de 0.017 USD/kWh, el cual tiene en cuenta la demanda eléctrica, la energía producida y la que se adquiere de la red eléctrica. Para corroborar que el valor obtenido fuese correcto, se utilizaron los datos reales para estimar la misma variable, donde se estimó que el costo nivelado de la energía fue de 0.013 USD/kWh. De modo que, aplicando las métricas de error se obtuvo:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - x_i| = |0.013 - 0.017| = 0.004 \text{ USD/kWh}$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - x_i)^2} = \sqrt{(0.013 - 0.017)^2} = 0.004 \text{ USD/kWh}$$

Dentro de la librería MicroGrids se asumieron valores de USD/kWh para el cálculo del costo nivelado de la energía. Los valores utilizados son los siguientes:

- **Costo de las baterías:** 0.4 USD/Wh
- **Costo de las energías renovables:** 1.67 USD/W
- **Tarifa de la red:** 0.14 USD/kWh
- **Tasa de descuento:** 12%
- **Tasa de inflación:** 2%
- **Años de duración del sistema:** 25 años
- **Eficiencia de carga de la batería:** 95%
- **Eficiencia de descarga de la batería:** 95%
- **Profundidad de descarga de las baterías:** 80%
- **Eficiencia del inversor:** 98.6%

Estos fueron los parámetros utilizados con la librería MicroGrids para realizar el cálculo del costo nivelado de la energía en el sistema de despacho diseñado, los costos de los componentes son valores genéricos otorgados por la librería.

A modo de referencia se empleó el algoritmo de despacho que dispone Homer Grid, el cual fue ejecutado utilizando los datos del año 2021 de la base de datos que se creó de CAISO y NREL, teniendo en cuenta que la manera de calcular el valor del costo nivelado de la energía es diferente a como se realiza con la librería MicroGrid, ya que Homer Grid tiene en cuenta el tiempo de degradación de las baterías y los costos de reemplazo. Para este caso, se obtuvo que el costo

nivelado de la energía en una duración de igualmente 25 años fue de 0.127 USD/kWh. Todo el código fuente del proyecto se presenta en el anexo 7.

Este algoritmo puede ser implementado en Colombia teniendo en cuenta que en este país se posee una tarifa plana, de modo que el punto de operación seguiría teniendo en cuenta los otros parámetros y solo compraría energía en los puntos de menor demanda para disminuir los costos. En referencia a los datos, para implementar este algoritmo en Colombia se requiere de una base de datos que contenga las tres series temporales de energía solar, energía eólica y demanda eléctrica, el código fue desarrollado para adaptarse a la base de datos que se le ingrese, por lo tanto, podría modificarse para que aprenda los patrones de las nuevas series temporales.

6. Conclusiones

Se desarrolló un modelo híbrido para el despacho de energía eléctrica utilizando un algoritmo de regresión compuesto por diferentes modelos de aprendizaje automático. Asimismo, se implementó un algoritmo de despacho de energía eléctrica para una microrred simulada utilizando Python y Homer Grid. Los resultados demuestran que tanto el modelo híbrido como el algoritmo de despacho presentan desempeños similares e incluso mejor a los mostrados en el estado del arte.

En Colombia no existen bases de datos públicas disponibles, de modo que se recurrió a utilizar bases de datos extranjeras, en este caso fue de las entidades CAISO y NREL, las cuales están orientadas a recopilar información cada cinco minutos. Sin embargo, se limitaron las predicciones al estado de California en Estados Unidos ya que CAISO solo provee datos de este lugar geográfico; asimismo, se realizó una reducción de los datos para que se asemejara al comportamiento de una microrred. Se verificó que el cambio de la magnitud de la producción y del consumo no influye en el aprendizaje del modelo ya que se aplica una normalización a los datos, y por lo tanto, no existen cambios en las predicciones realizadas.

Se realizó un análisis exploratorio de ambas bases de datos para estudiar la correlación que existía entre las variables y de su estacionalidad a través de toda la serie temporal utilizando técnicas de visualización y matemáticas como la de Dickie Fuller. Se identificaron varias técnicas de depuración de bases de datos dentro de las cuales se utilizaron funciones que encontraban valores vacíos, nulos o negativos, los cuales influyen de manera negativa en el aprendizaje de los modelos de inteligencia artificial.

Durante el desarrollo de los algoritmos de aprendizaje automático se dividieron los datos en tres subconjuntos, los cuales se identifican como entrenamiento, validación y prueba. Para la predicción de valores futuros dentro de las series temporales fue necesario determinar qué tipo de

algoritmo se debía utilizar, para ello se utilizaron técnicas para identificar la estacionalidad de los datos y de su relación con las variables exógenas dentro del conjunto de datos procesado.

Cada uno de los modelos cuenta con hiperparámetros, los cuales son conjuntos de parámetros externos que indican a los algoritmos las características que debe tener en cuenta al modelarse. Para hallar los mejores valores se utilizó una técnica de validación cruzada, la cual itera estos parámetros según una lista ingresada y escoge el de menor error. Este proceso es muy importante para realizar predicciones precisas en cualquier tipo de tarea de aprendizaje automático, ya que redujo el error de los modelos entre un 77.77% hasta un 43.75%.

Las métricas de error más bajas se obtuvieron en las predicciones de demanda eléctrica, ya que este valor siempre se mantiene constante en un rango determinado, nunca llega a ser exactamente cero ya que siempre se está consumiendo energía y siempre presenta un patrón bimodal en su comportamiento, lo que le permite al modelo adaptarse y realizar predicciones de manera precisa, esto se evidencia obteniendo un MAPE de 0.52% hasta un 5.64%, siendo los valores más bajos en comparación a los otros modelos.

El lenguaje de programación Python es de código abierto, por lo tanto, no requiere de una licencia para su uso o distribución de software generado con él, lo que disminuye los costos del proyecto. Asimismo, los otros programas como Homer Grid fueron utilizados bajo licencias académicas gratuitas que permiten utilizar todas las herramientas durante un tiempo limitado.

La implementación de los algoritmos de predicción para el despacho de energía le permite al usuario final observar el comportamiento que tendrá su microrred a lo largo de los días posteriores, esto abre la posibilidad de prepararse para situaciones determinadas en ciertas épocas del año.

En la última instancia del desarrollo de proyecto no se consideró ARIMA dentro del modelo ensamblado por el costo computacional que este implicaba y debido a que el error MAPE

más alto fue de 60.14% en energía solar. A través de un análisis en la temporalidad en los datos, se identificó que la única manera posible de implementarlo es con pasos de tiempo de una hora por cada muestra, lo cual lo hace inviable al ser incompatible con las series de tiempo utilizadas en el resto de los modelos.

Para evaluar el desempeño del algoritmo de despacho de energía se utilizó la métrica de costo nivelado de energía (Levelized Cost of Energy en inglés), la cual es una variable que se utiliza para comparar el costo de utilizar diferentes tecnologías de generación en igualdad de condiciones, donde se obtuvo un error MAE de 0.004 USD/kWh, en este caso se tomó como referente el algoritmo de Homer Grid con el desarrollado en Python.

Se evidenció que utilizar variables exógenas no mejora la capacidad predictiva de los modelos de aprendizaje automático implementados en esta investigación, esto es debido a que su correlación no es tan fuerte con respecto a las variables predichas por los algoritmos de aprendizaje automático utilizados, sin embargo, mejoran las predicciones del modelo entre un 1.64% hasta un 2.05% según la serie temporal.

7. Recomendaciones

Para realizar predicciones más precisas es mejor utilizar una base de datos que contenga todas las variables que se deseen estudiar, ya que, en este caso las entidades CAISO y NREL tomaban sus datos en diferentes localizaciones. Asimismo, es importante que la entidad que genere esos datos los mida en tiempo real utilizando sensores para mejorar la calidad de las predicciones.

Se recomienda adquirir la licencia del software de optimización de IBM CPLEX para realizar análisis de despacho de energía por periodos de tiempo más largos, puesto que, en este caso se limitó la investigación a dos días de despacho, debido a las restricciones de la capa gratuita de IBM CPLEX de solo mil datos.

En caso de utilizar las mismas bases de datos de esta investigación para futuros proyectos, se podría realizar un estudio de cuáles serían los métodos de interpolación disponibles para completar los datos del año 2018 que están incompletos en la base de datos de CAISO. De esta forma sería posible añadir 105120 datos más para el entrenamiento de los modelos, ya que NREL sí posee información completa de ese año.

Se identificó que la implementación de modelos que tuvieran en cuenta una mayor cantidad de atrasos en las predicciones tendría una mayor precisión a cambio de un mayor costo computacional; por lo tanto, un computador con mayor capacidad de procesamiento podría disminuir los tiempos de entrenamiento y solucionar los problemas previamente mencionados.

A pesar de que las variables exógenas representaron una mejora no tan significativa, es recomendable añadirlas a los modelos para tener en cuenta parámetros externos que están correlacionados con la base de datos.

Referencias

- [1] D. F. Ariza Melo, “Microrredes: Una alternativa sostenible para la universalización de la energía eléctrica.” Universidad Santiago de Cali, Cali, 2020.
- [2] CENER, “Introducción a las Microrredes,” *CENER*. <https://www.cener.com/introduccion-a-las-microrredes/> (accessed Jan. 10, 2022).
- [3] J. J. Chacón Guadalix, “Las microrredes como herramienta de gestión energética,” *BBVA*, Feb. 17, 2015. <https://www.bbvaopenmind.com/ciencia/medioambiente/las-microrredes-como-herramienta-de-gestion-energetica/> (accessed Jan. 10, 2022).
- [4] L. Rodríguez Urrego, M. Valderrama Mendoza, H. García León, and P. C. Ocampo, “La gestión para cadena de suministro de sistemas de energía solar fotovoltaica en Colombia y su situación actual,” *Av. Investig. en Ing.*, vol. 15, no. 1, pp. 112–130, Dec. 2018, doi: 10.18041/1794-4953/avances.1.1368.
- [5] M. H. Badii, A. Guillen, and J.L. Abreu, “Renewable Energies and Energy Conservation,” *Daena Int. J. Good Conscienc.*, vol. 11, no. 1, pp. 141–155, Apr. 2016.
- [6] IDEAM.GOV, “Calidad del aire,” *IDEAM*, 2015. <http://www.ideam.gov.co/web/contaminacion-y-calidad-ambiental/calidad-del-aire> (accessed Jan. 10, 2022).
- [7] CIER, “Tarifas eléctricas para grandes usuarios en Latinoamérica,” *Energía Estratégica*, Jan. 29, 2020. <https://www.energiaestrategica.com/comparan-tarifas-electricas-para-grandes-usuarios-en-latinoamerica-paraguay-tienen-las-mas-economicas-guatemala-las-mas-caras/> (accessed Jan. 10, 2022).
- [8] T. Gonzáles and J. Valencia, *Integración de las energías renovables no convencionales en Colombia*, 1st ed., vol. 1. Bogotá: Ministerio de Minas y Energía, 2015.
- [9] La República, “Interconexiones en Colombia,” *LR*, Feb. 07, 2015.

- <https://www.larepublica.co/economia/cuatro-cabeceras-en-colombia-siguen-sin-suministro-de-energia-electrica-2217626> (accessed Jan. 10, 2022).
- [10] W. Ñustes and S. Rivera, “Colombia: Territorio de inversión en fuentes no convencionales de energía renovable para la generación eléctrica,” *Rev. Ing. Investig. y Desarro.*, vol. 17, no. 1, pp. 37–48, 2017.
- [11] CREG, “Generación distribuida y autogeneración a pequeña y gran escala en las zonas no interconectadas,” Bogota, 2018.
- [12] CREG, “Actualización de las reglas de autogeneración a pequeña escala y generación distribuida,” *Comisión de Regulación de Energía y Gas*, Nov. 22, 2021. .
- [13] J. E. Pérez Cardenas, “La calidad del aire en Colombia: Un problema de salud para todos,” *Rev. Biosalud*, vol. 16, no. 2, pp. 5–6, 2017, doi: 10.17151/biosa.2017.16.2.1.
- [14] S. F. Posada Valcárcel, “Análisis de imágenes satelitales de observación de la Tierra y datos geoespaciales a través de Machine Learning,” Universidad de los Andes, 2020.
- [15] G. F. Di Candia Jaimés, B. O. Calderón, S. A. Castro, B. Medina, and D. Guevara, “Análisis del Recurso Eólico disponible en la ciudad de Cúcuta usando Aprendizaje Automático,” in *Actas Del Vi Congreso Investigación, Desarrollo E Innovación De La Universidad Internacional De Ciencia Y Tecnología Idi - Unicyt 2021*, Jan. 2022, pp. 724–738, doi: 10.47300/978-9962-738-04-6-41.
- [16] N. Quijano *et al.*, “Microrredes Aisladas en la Guajira: Diseño e Implementación,” *Rev. Ing.*, no. 48, Jan. 2019, doi: 10.16924/revinge.48.7.
- [17] J. D. Garzón-Hidalgo and A. J. Saavedra-Montes, “Una metodología de diseño de micro redes para zonas no interconectadas de Colombia,” *TecnoLógicas*, vol. 20, no. 39, May 2017, doi: 10.22430/22565337.687.
- [18] Z. Wang, A. Negash, and D. S. Kirschen, “Optimal scheduling of energy storage under

- forecast uncertainties,” *IET Gener. Transm. Distrib.*, vol. 11, no. 17, pp. 4220–4226, 2017, doi: 10.1049/iet-gtd.2017.0037.
- [19] A. Hirsch, Y. Parag, and J. Guerrero, “Microgrids: A review of technologies, key drivers, and outstanding issues,” *Renew. Sustain. Energy Rev.*, vol. 90, pp. 402–411, 2018, doi: 10.1016/j.rser.2018.03.040.
- [20] S. B. Sepúlveda Mora and S. Hegedus, “Design of a Resilient and Eco-friendly Microgrid for a Commercial Building,” *Aibi Rev. Investig. Adm. e Ing.*, vol. 9, no. 1, pp. 8–18, 2021, doi: 10.15649/2346030x.919.
- [21] S. B. Sepúlveda-Mora and S. Hegedus, “Making the case for time-of-use electric rates to boost the value of battery storage in commercial buildings with grid connected PV systems,” *Energy*, vol. 218, p. 119447, 2021, doi: 10.1016/j.energy.2020.119447.
- [22] S. Medina Rincón and A. K. Venegas Camargo, “Energías renovables un futuro optimo para Colombia,” *Punto de Vista*, vol. 9, no. 13, pp. 1–16, 2018.
- [23] H. García, A. Corredor, L. Calderón, and M. Gómez, “Análisis costo beneficio de energías renovables no convencionales en Colombia,” Bogotá, 2013.
- [24] J. Gómez Ramírez, “La energía solar fotovoltaica en Colombia: potenciales, antecedentes y perspectivas,” Universidad Santo Tomás, 2018.
- [25] Global Wind Energy Council (GWEC), “Global Wind Report 2021,” 2021. <https://gwec.net/wp-content/uploads/2021/03/GWEC-Global-Wind-Report-2021.pdf> (accessed Jan. 31, 2022).
- [26] A. Anand and L. Suganthi, “Forecasting of Electricity Demand by Hybrid ANN-PSO Models,” *Int. J. Energy Optim. Eng.*, vol. 6, no. 4, pp. 66–83, Oct. 2017, doi: 10.4018/IJEOE.2017100105.
- [27] ENERGY.GOV, “How Microgrids Work,” *Department of Energy*, 2014.

- <https://www.energy.gov/articles/how-microgrids-work> (accessed Jan. 10, 2022).
- [28] J. Zhu, S. Huang, Y. Liu, H. Lei, and B. Sang, “Optimal energy management for grid-connected microgrids via expected-scenario-oriented robust optimization,” *Energy*, vol. 216, p. 119224, Feb. 2021, doi: 10.1016/j.energy.2020.119224.
- [29] M. Ghofrani and M. Alolayan, “Time Series and Renewable Energy Forecasting,” in *Time Series Analysis and Applications*, InTech, 2018.
- [30] “California Independent System Operator (CAISO),” 2023. <http://www.caiso.com/about/Pages/default.aspx> (accessed Jan. 30, 2022).
- [31] “National Renewable Energy Laboratory (NREL),” 2023. <https://www.nrel.gov/> (accessed Jan. 30, 2022).
- [32] HOMER Energy, “HOMER Grid,” 2023. <https://www.homerenergy.com/index.html> (accessed Jan. 31, 2022).
- [33] Python, “Python Documentation,” 2022. <https://docs.python.org/3/> (accessed Jan. 15, 2022).
- [34] S. Athiyarath, M. Paul, and S. Krishnaswamy, “A Comparative Study and Analysis of Time Series Forecasting Techniques,” *SN Comput. Sci.*, vol. 1, no. 3, p. 175, May 2020, doi: 10.1007/s42979-020-00180-5.
- [35] H. Mishra and S. Tripathi, “A Comparative Study of Data Clustering Techniques,” *Int. Res. J. Eng. Technol.*, vol. 04, no. 05, pp. 1392–1398, May 2017.
- [36] H. Chhabra, “A Comparative Study of Arima and Sarima Models to Forecast Lockdowns Due to Covid-19,” *Res. Sq.*, Sep. 2022, doi: <https://doi.org/10.21203/rs.3.rs-2069129/v1>.
- [37] E. Junqué de Fortuny, D. Martens, and F. Provost, “Predictive Modeling With Big Data: Is Bigger Really Better?,” *Big Data*, vol. 1, no. 4, pp. 215–226, Dec. 2013, doi: 10.1089/big.2013.0037.

- [38] I. Peláez, “Modelos de regresión: lineal simple y regresión logística,” *Rev. Seden*, vol. 14, pp. 195–214, 2016.
- [39] M. F. Akay, O. Bozkurt, E. Cetin, and I. Yarim, “Multiple linear regression-based physical fitness prediction models for Turkish secondary school students,” *New Trends Issues Proc. Humanit. Soc. Sci.*, vol. 5, no. 4, pp. 58–64, Sep. 2018, doi: 10.18844/prosoc.v5i4.3704.
- [40] P. R. Pandurangan, “A Comparative Study of Data Mining Algorithms for Decision Tree Approaches using WEKA Tool,” *Adv. Nat. Appl. Sci.*, vol. 11, pp. 230–241, Jul. 2017.
- [41] Y. Muliono and F. Tanzil, “A Comparison of Text Classification Methods k-NN, Naïve Bayes, and Support Vector Machine for News Classification,” *J. Inform. J. Pengemb. IT*, vol. 3, no. 2, pp. 157–160, May 2018, doi: 10.30591/jpit.v3i2.828.
- [42] T. Chen and C. Guestrin, “XGBoost: A Scalable Tree Boosting System,” in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Aug. 2016, pp. 785–794, doi: 10.1145/2939672.2939785.
- [43] R. Wilmer and M. Bertha, “Generalidades de las redes neuronales artificiales,” in *Redes neuronales artificiales aplicadas al reconocimiento de patrones*, Editorial Utmach, 2018.
- [44] Z. Lipton, J. Berkowitz, and C. Elkan, “A Critical Review of Recurrent Neural Networks for Sequence Learning,” May 2015, doi: <https://doi.org/10.48550/arXiv.1506.00019>.
- [45] R. Staudemeyer and E. Morris, “Understanding LSTM - a tutorial into Long Short-Term Memory Recurrent Neural Networks,” Sep. 2019, doi: <https://doi.org/10.48550/arXiv.1909.09586>.
- [46] El Congreso de Colombia, “Ley 603 De 2000,” 2000. <https://www.funcionpublica.gov.co/eva/gestornormativo/norma.php?i=13960>.
- [47] El Congreso de Colombia, “Ley 1273 De 2009,” 2009. <https://www.funcionpublica.gov.co/eva/gestornormativo/norma.php?i=34492>.

- [48] El Congreso de Colombia, “Ley 23 De 1982,” 1982.
<http://derechodeautor.gov.co:8080/documents/10181/182597/23.pdf/a97b8750-8451-4529-ab87-bb82160dd226>.
- [49] M. A. Cohen and D. S. Callaway, “Effects of distributed PV generation on California’s distribution system, Part 1: Engineering simulations,” *Sol. Energy*, vol. 128, pp. 126–138, Apr. 2016, doi: 10.1016/j.solener.2016.01.002.
- [50] Devashish and A. Thakur, “A Comprehensive Review on Wind Energy Systems for Electric Power Generation: Current Situation and Improved Technologies to Realize Future Development,” *Int. J. Renew. Energy Res.*, vol. 7, no. 4, pp. 1786–1805, 2017.
- [51] L. Corona, A. Mochon, and Y. Saez, “Electricity market integration and impact of renewable energy sources in the Central Western Europe region: Evolution since the implementation of the Flow-Based Market Coupling mechanism,” *Energy Reports*, vol. 8, pp. 1768–1788, Nov. 2022, doi: 10.1016/j.egyr.2021.12.077.
- [52] International Energy Agency, “World Energy Outlook 2022,” 2022. Accessed: Apr. 01, 2022. [Online]. Available: <https://www.iea.org/reports/world-energy-outlook-2022>.
- [53] International Renewable Energy Agency, “Renewable power generation costs in 2021,” 2021. Accessed: Apr. 01, 2022. [Online]. Available: <https://www.irena.org/publications/2022/Jul/Renewable-Power-Generation-Costs-in-2021>.
- [54] F. Vadillo, “Una introducción a la interpolación polinomial,” Valencia, España, Jan. 2021.
- [55] B. J. Albertus, “Confronting the Duck Curve: How to Address Over-Generation of Solar Energy,” *Office of energy efficiency and renewable energy*, Oct. 12, 2017.
<https://www.energy.gov/eere/articles/confronting-duck-curve-how-address-over-generation-solar-energy> (accessed Apr. 02, 2022).
- [56] T. D.K., P. B.G, and F. Xiong, “Auto-detection of epileptic seizure events using deep neural

- network with different feature scaling techniques,” *Pattern Recognit. Lett.*, vol. 128, pp. 544–550, Dec. 2019, doi: 10.1016/j.patrec.2019.10.029.
- [57] California Public Utilities Commission, “Net Energy Metering Rulemaking,” Jan. 2016. Accessed: Apr. 02, 2022. [Online]. Available: <https://docs.cpuc.ca.gov/PublishedDocs/Published/G000/M158/K181/158181678.pdf>.
- [58] Comisión de Regulación de Energía y Gas, “Resolución CREG 030 de 2018,” Bogotá, Feb. 2018. Accessed: Apr. 02, 2022. [Online]. Available: [http://apolo.creg.gov.co/Publicac.nsf/1c09d18d2d5ffb5b05256eee00709c02/83b41035c2c4474f05258243005a1191/\\$FILE/Creg030-2018.pdf](http://apolo.creg.gov.co/Publicac.nsf/1c09d18d2d5ffb5b05256eee00709c02/83b41035c2c4474f05258243005a1191/$FILE/Creg030-2018.pdf).
- [59] K. J. Iheanetu, “Solar Photovoltaic Power Forecasting: A Review,” *Sustainability*, vol. 14, no. 24, p. 17005, Dec. 2022, doi: 10.3390/su142417005.
- [60] K. Vaibhav and V. Singh, “Solar Power forecasting: The State-Of-The-Art,” *J. CPRI*, vol. 10, Jun. 2014.
- [61] S. Amely Jumaat, F. Crocker, M. Helmy Abd Wahab, N. Hanis Mohammad Radzi, and M. Fakri Othman, “Prediction of Photovoltaic (PV) Output Using Artificial Neural Network (ANN) Based on Ambient Factors,” *J. Phys. Conf. Ser.*, vol. 1049, p. 012088, Jul. 2018, doi: 10.1088/1742-6596/1049/1/012088.
- [62] F. Kosanoglu, “Wind Speed Forecasting with a Clustering-Based Deep Learning Model,” *Appl. Sci.*, vol. 12, no. 24, p. 13031, Dec. 2022, doi: 10.3390/app122413031.
- [63] A. Augustyn and J. Kamiński, “A review of methods applied for wind power generation forecasting,” *Energy Policy J.*, vol. 2, no. 2, pp. 139–150, 2018, doi: 10.24425/122767.
- [64] M. Lei, L. Shiyan, J. Chuanwen, L. Hongling, and Z. Yan, “A review on the forecasting of wind speed and generated power,” *Renew. Sustain. Energy Rev.*, vol. 13, no. 4, pp. 915–920, May 2009, doi: 10.1016/j.rser.2008.02.002.

- [65] A. Singh and D. Chaturvedi, “Load Forecasting Techniques and Methodologies: A Review,” 2012.
- [66] R. Weron, “Electricity price forecasting: A review of the state-of-the-art with a look into the future,” *Int. J. Forecast.*, vol. 30, no. 4, pp. 1030–1081, Oct. 2014, doi: 10.1016/j.ijforecast.2014.08.008.
- [67] S. Pawar, P. Mole, S. Phadtare, D. Aghor, and P. Vadtile, “Electricity Forecasting Using Machine Learning: A Review,” 2021, pp. 127–134.

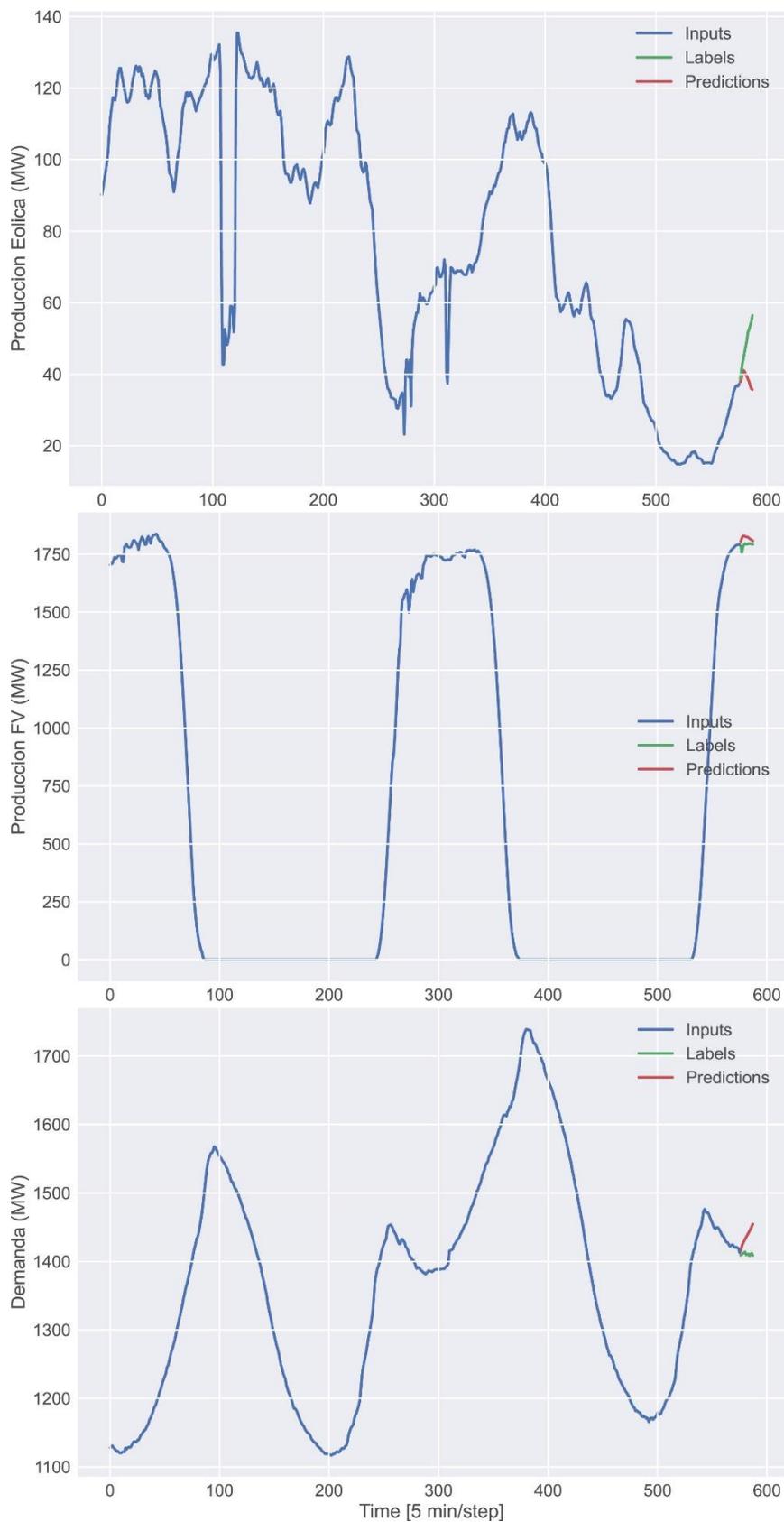
Anexos

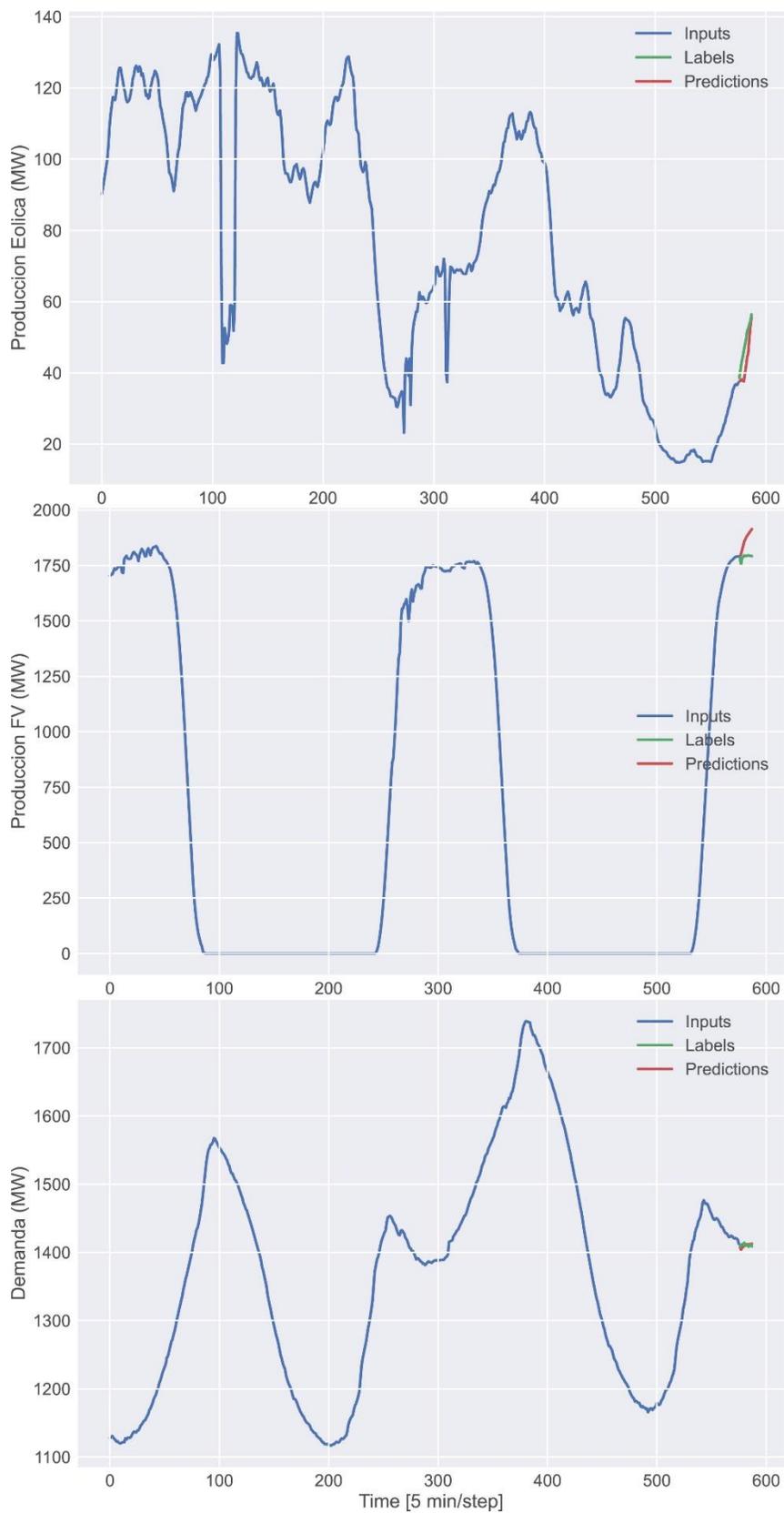
Anexo I. Base de datos generada (solo cien datos)

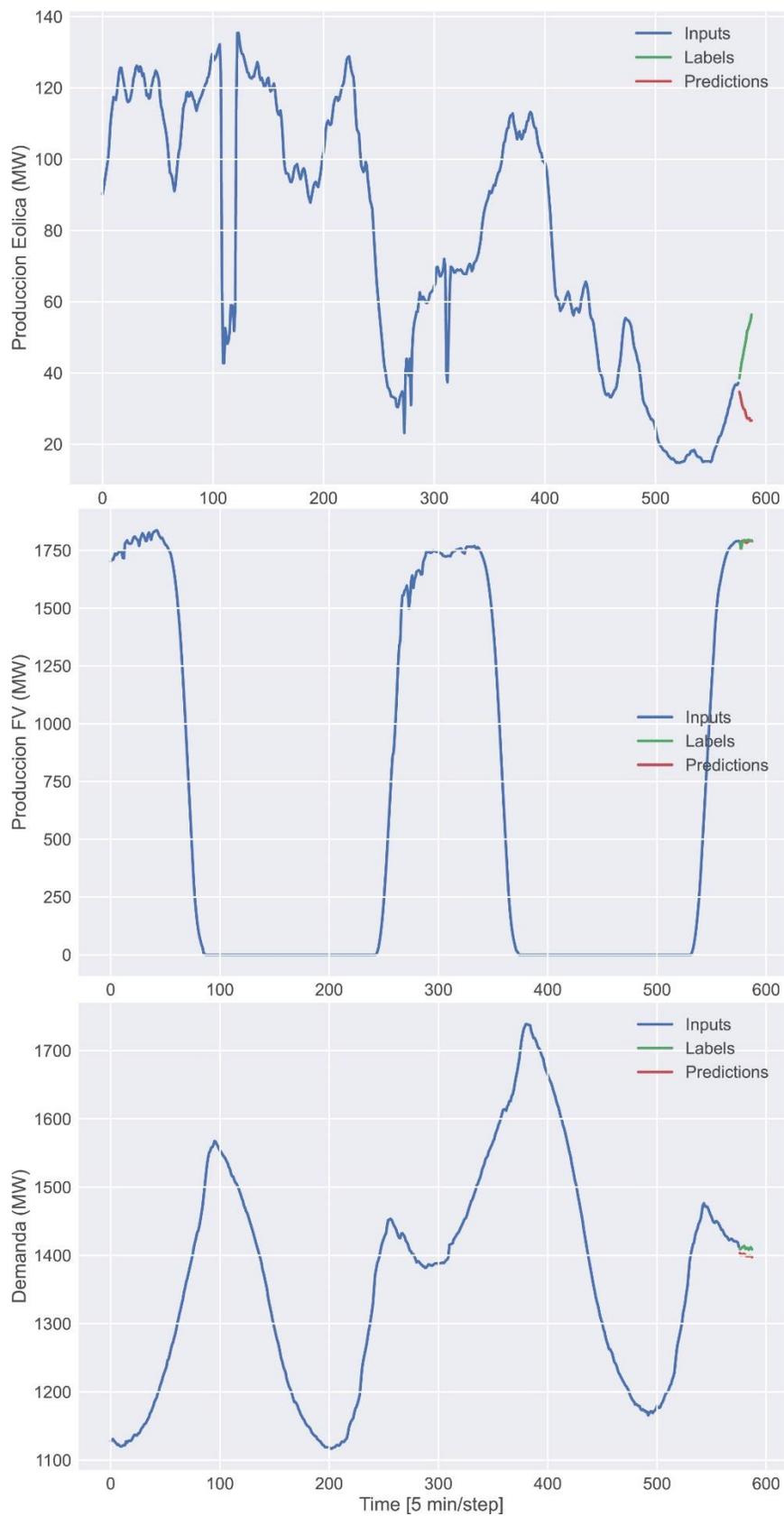
Fecha y hora	Estación	Día de la semana	DHI	DNI	Velocidad del viento	Humedad	Temperatura	GHI	FV	Eólica	Demanda
2019-01-01-T08:20	1	1	37,4	699,6	3,46	40,724	4,54	173,6	4500	1929	20956
2019-01-01-T08:25	1	1	39,2	719,4	3,5	40,098	4,76	187,8	4863	1940	20886
2019-01-01-T08:30	1	1	40,4	738	3,54	39,554	4,96	201,4	5222	1966	20806
2019-01-01-T08:35	1	1	41,8	753,8	3,58	37,278	5,18	216	5579	1965	20728
2019-01-01-T08:40	1	1	43	769,4	3,64	36,854	5,34	229,6	5921	2017	20617
2019-01-01-T08:45	1	1	44,2	784,2	3,7	36,358	5,54	243,6	6223	2060	20553
2019-01-01-T08:50	1	1	45,4	797,6	3,72	35,878	5,74	257,4	6454	2052	20505
2019-01-01-T08:55	1	1	46,4	810,2	3,78	35,34	5,96	271	6580	2007	20426
2019-01-01-T09:00	1	1	47,2	819,8	3,82	34,918	6,14	283,8	6691	2012	20365
2019-01-01-T09:05	1	1	48	830,8	3,84	34,53	6,3	297	7088	2006	20318
2019-01-01-T09:10	1	1	49	841,2	3,82	34,108	6,48	310	7288	1971	20368
2019-01-01-T09:15	1	1	49,8	850,6	3,84	33,77	6,62	322,4	7447	1964	20335
2019-01-01-T09:20	1	1	59,4	772,6	3,84	33,366	6,8	323	7589	1950	20267
2019-01-01-T09:25	1	1	60,2	785	3,86	32,99	6,96	335,4	7719	1940	20216
2019-01-01-T09:30	1	1	52,4	875,8	3,88	32,596	7,14	359,2	7823	1902	20153
2019-01-01-T09:35	1	1	56	872	3,86	32,012	7,3	370,6	7929	1896	20126
2019-01-01-T09:40	1	1	66,6	795,8	3,86	31,668	7,46	368	8014	1869	20051
2019-01-01-T09:45	1	1	66,8	809,4	3,86	31,32	7,62	380,2	8073	1811	19957
2019-01-01-T09:50	1	1	69,2	805,8	3,9	30,964	7,78	388,4	8130	1766	19882
2019-01-01-T09:55	1	1	67,6	830	3,88	30,608	7,96	401,8	8214	1760	19824
2019-01-01-T10:00	1	1	68,6	836,8	3,92	30,236	8,14	412	8232	1766	19767
2019-01-01-T10:05	1	1	67,8	851,2	3,88	30,046	8,22	423,2	8247	1764	19763
2019-01-01-T10:10	1	1	71,2	842,4	3,9	29,728	8,38	429,8	8256	1777	19748
2019-01-01-T10:15	1	1	72,2	843,4	3,86	29,486	8,5	437,8	8262	1808	19704
2019-01-01-T10:20	1	1	70,6	863,2	3,86	29,262	8,62	449	8268	1831	19633
2019-01-01-T10:25	1	1	72,6	859,2	3,82	28,96	8,76	456	8257	1863	19585
2019-01-01-T10:30	1	1	73	865,2	3,82	28,74	8,88	463,8	8253	1885	19541
2019-01-01-T10:35	1	1	74,8	863,2	3,82	28,584	9	470,2	8250	1905	19496
2019-01-01-T10:40	1	1	63	935,6	3,8	28,33	9,14	490,8	8272	1939	19435
2019-01-01-T10:45	1	1	63,2	938,2	3,8	28,144	9,24	497,8	8258	1979	19384
2019-01-01-T10:50	1	1	64	941	3,8	27,894	9,38	504	8245	2017	19316
2019-01-01-T10:55	1	1	64	943,8	3,78	27,668	9,48	509,8	8236	2019	19228
2019-01-01-T11:00	1	1	64,2	946	3,78	27,446	9,6	515,4	8230	2095	19230
2019-01-01-T11:05	1	1	64,4	948,2	3,76	27,292	9,68	520,4	8214	2135	19156
2019-01-01-T11:10	1	1	65,8	931,4	3,78	27,132	9,78	517,2	8197	2089	19108
2019-01-01-T11:15	1	1	64,8	952	3,8	27,006	9,84	529,2	8179	2064	19034
2019-01-01-T11:20	1	1	65,2	953,6	3,76	26,828	9,94	533	8155	2054	18985
2019-01-01-T11:25	1	1	65,4	955	3,76	26,744	10	536,2	8154	2002	18979
2019-01-01-T11:30	1	1	72,4	915,8	3,78	26,62	10,06	529,2	8154	1952	18929
2019-01-01-T11:35	1	1	78,8	893	3,76	26,84	10,16	527	8155	1935	18871
2019-01-01-T11:40	1	1	77,8	898	3,78	26,748	10,22	529,6	8146	1933	18828
2019-01-01-T11:45	1	1	78	898,4	3,78	26,574	10,32	531,2	8136	1915	18819
2019-01-01-T11:50	1	1	81,6	884	3,76	26,43	10,4	529,4	8127	1893	18760
2019-01-01-T11:55	1	1	84,4	872,4	3,78	26,376	10,44	527,2	8131	1866	18724
2019-01-01-T12:00	1	1	83,8	874	3,78	26,226	10,52	527,6	8135	1847	18705
2019-01-01-T12:05	1	1	81,4	885,8	3,76	26,17	10,56	529,8	8135	1845	18688
2019-01-01-T12:10	1	1	77,4	901,4	3,78	26,086	10,6	532,6	8140	1835	18647
2019-01-01-T12:15	1	1	81,6	884	3,78	26,03	10,64	528,2	8137	1830	18633
2019-01-01-T12:20	1	1	83	878	3,8	26,03	10,64	525,8	8136	1803	18594
2019-01-01-T12:25	1	1	87,4	854,4	3,78	25,976	10,68	518,2	8146	1775	18602
2019-01-01-T12:30	1	1	83,8	870,6	3,8	25,86	10,74	520	8148	1713	18574
2019-01-01-T12:35	1	1	88	855	3,78	26,612	10,74	514,4	8093	1657	18569
2019-01-01-T12:40	1	1	89,6	844	3,8	26,534	10,8	509,2	8102	1595	18534
2019-01-01-T12:45	1	1	82,4	873,6	3,78	26,464	10,82	511,8	8092	1598	18545
2019-01-01-T12:50	1	1	85	858,2	3,82	26,426	10,84	505	8136	1617	18546
2019-01-01-T12:55	1	1	78,6	881	3,82	26,384	10,88	505,6	8230	1599	18539

2019-01-01-T13:00	1	1	66,8	937	3,82	26,32	10,92	514	8250	1617	18523
2019-01-01-T13:05	1	1	66,6	934,8	3,8	26,352	10,9	509	8210	1599	18536
2019-01-01-T13:10	1	1	66,8	914,8	3,8	26,384	10,88	495,8	8267	1587	18528
2019-01-01-T13:15	1	1	66	915,6	3,82	26,384	10,88	490,4	8213	1571	18573
2019-01-01-T13:20	1	1	96,8	768,4	3,82	26,406	10,86	453	8225	1514	18574
2019-01-01-T13:25	1	1	65,2	922,4	3,8	26,406	10,86	483	8167	1410	18598
2019-01-01-T13:30	1	1	64,8	919,6	3,76	26,44	10,84	475,6	8237	1338	18602
2019-01-01-T13:35	1	1	64,4	915	3,78	27,566	10,82	467,4	8317	1289	18632
2019-01-01-T13:40	1	1	64,2	911,2	3,76	27,566	10,82	459,6	8328	1279	18663
2019-01-01-T13:45	1	1	63,6	907,2	3,76	27,59	10,8	451	8324	1232	18715
2019-01-01-T13:50	1	1	70	860,2	3,78	27,59	10,8	432,6	8283	1185	18760
2019-01-01-T13:55	1	1	62,6	898	3,78	27,624	10,78	432,8	8229	1181	18784
2019-01-01-T14:00	1	1	76,2	808	3,76	27,624	10,78	404	8187	1202	18847
2019-01-01-T14:05	1	1	61,4	887,8	3,72	27,702	10,72	413,6	8145	1230	18874
2019-01-01-T14:10	1	1	60,8	882,2	3,68	27,922	10,6	403,2	8043	1202	18929
2019-01-01-T14:15	1	1	60	876,2	3,62	28,036	10,52	392,4	7958	1188	18984
2019-01-01-T14:20	1	1	59,4	867	3,58	28,224	10,42	380,8	7910	1180	19045
2019-01-01-T14:25	1	1	58,6	859,2	3,54	28,314	10,36	369,2	7854	1164	19127
2019-01-01-T14:30	1	1	58	852	3,46	28,502	10,26	357,4	7793	1132	19193
2019-01-01-T14:35	1	1	57,6	843	3,42	32,692	10,2	345,4	7695	1088	19277
2019-01-01-T14:40	1	1	57,4	815,8	3,38	32,908	10,1	328	7442	1070	19396
2019-01-01-T14:45	1	1	63,8	760	3,32	33,012	10,04	308,2	7253	1059	19486
2019-01-01-T14:50	1	1	54,8	817,2	3,3	33,286	9,92	308,6	7108	1021	19598
2019-01-01-T14:55	1	1	53,8	807,2	3,28	33,428	9,84	295,8	6928	986	19673
2019-01-01-T15:00	1	1	57,6	748	3,22	33,612	9,76	273,4	6785	942	19728
2019-01-01-T15:05	1	1	52	784,6	3,22	33,612	9,76	269,4	6612	879	19805
2019-01-01-T15:10	1	1	50,6	772,6	3,22	33,612	9,76	256	6410	811	19866
2019-01-01-T15:15	1	1	49,6	759,2	3,22	33,612	9,76	242	6126	790	19939
2019-01-01-T15:20	1	1	48	744,8	3,22	33,612	9,76	228,2	5725	795	20086
2019-01-01-T15:25	1	1	47	729,6	3,22	33,612	9,76	214,2	5291	789	20227
2019-01-01-T15:30	1	1	45,6	713	3,22	33,612	9,76	200	4944	756	20362
2019-01-01-T15:35	1	1	44	695	3,22	33,612	9,76	186	4581	714	20485
2019-01-01-T15:40	1	1	42,4	675,6	3,22	33,612	9,76	171,8	4208	696	20659
2019-01-01-T15:45	1	1	40,8	654,6	3,22	33,612	9,76	157,6	3720	662	20832
2019-01-01-T15:50	1	1	39	631,6	3,22	33,612	9,76	143,2	3377	636	20999
2019-01-01-T15:55	1	1	37,2	606,8	3,22	33,612	9,76	129,2	2951	616	21207
2019-01-01-T16:00	1	1	38,8	494	2,38	43,592	6,82	106,4	2603	582	21381
2019-01-01-T16:05	1	1	36,6	458,8	2,38	43,88	6,72	92,4	2221	574	21551
2019-01-01-T16:10	1	1	33,2	440	2,4	44,31	6,58	81,2	1807	554	21729
2019-01-01-T16:15	1	1	30,2	404,4	2,4	44,658	6,46	68,4	1396	544	21952
2019-01-01-T16:20	1	1	25	434,2	2,42	45,098	6,32	61,6	1151	530	22161
2019-01-01-T16:25	1	1	22,2	388,6	2,4	45,398	6,22	49,6	864	522	22383
2019-01-01-T16:30	1	1	19,4	338,2	2,4	45,71	6,12	38,6	692	518	22605
2019-01-01-T16:35	1	1	14,2	239,8	2,4	44,22	5,98	25,6	546	518	22875
2019-01-01-T16:40	1	1	9,6	162,2	2,42	44,518	5,88	16	276	538	23169

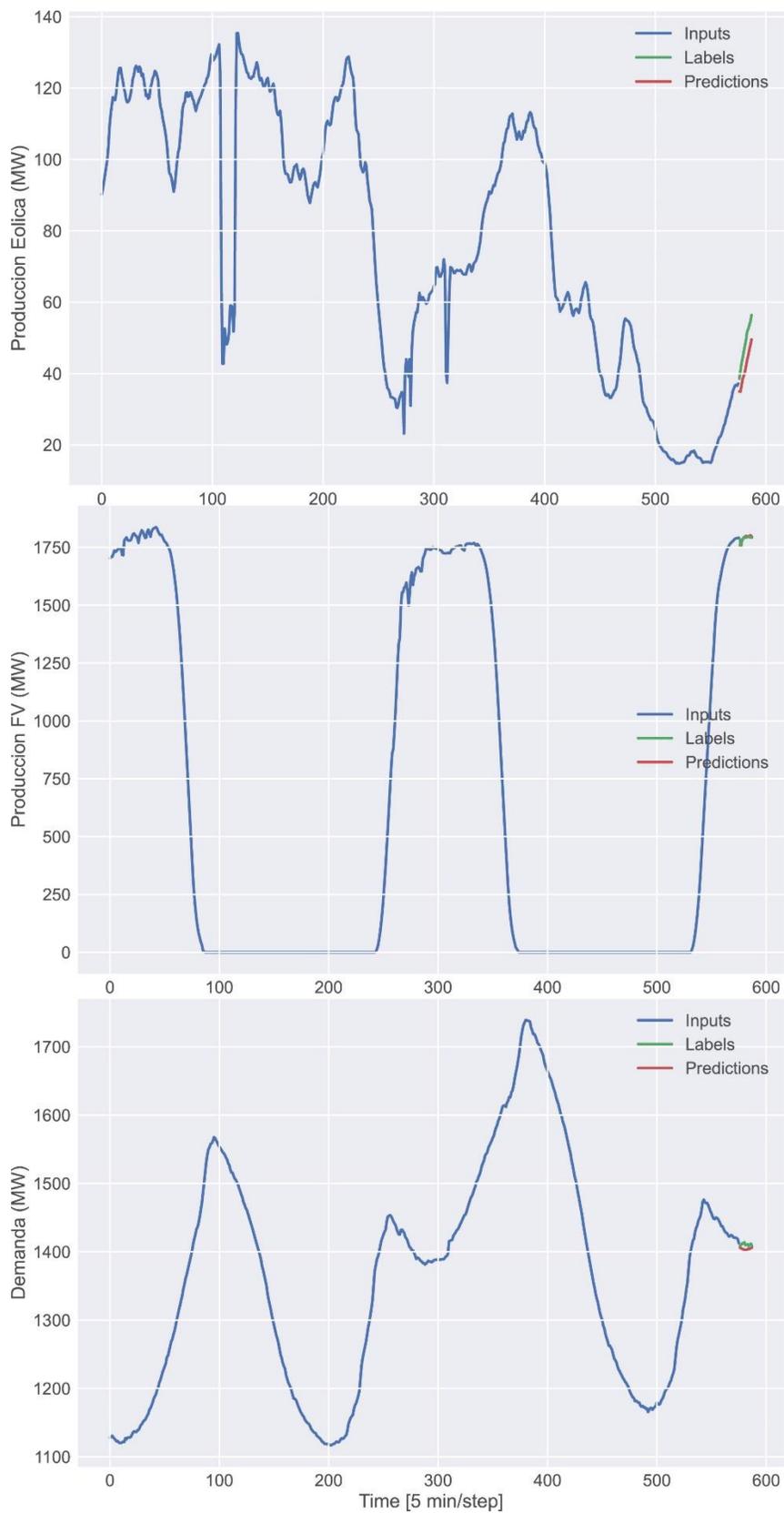
Anexo 2. Predicción de energía solar, energía eólica y demanda eléctrica del modelo lineal



Anexo 3. Predicción de energía solar, energía eólica y demanda eléctrica del modelo denso

Anexo 4. Predicción de energía solar, energía eólica y demanda eléctrica del modelo convolucional

Anexo 5. Predicción de energía solar, energía eólica y demanda eléctrica del modelo LSTM



Anexo 6. Base de datos disponible en: <https://data.mendeley.com/datasets/fdffr3tc2/1>


[Sign In / Register](#)

Renewable Energy and Electricity Demand Time Series Dataset with Exogenous Variables at 5-minute Interval

Published: 26 April 2023 | Version 1 | DOI: 10.17632/fdffr3tc2.1

Contributors: [Sebastian Rojas Ortega](#), [Paola Castro-Correa](#), [Sergio Sepúlveda-Mora](#), [Jhon Castro-Correa](#)

Description

The described database was created using data obtained from the California Independent System Operator (CAISO) and the National Renewable Energy Laboratory (NREL). All data was collected at five-minute intervals, and subsequently cleaned and modified to create a database comprising three time series: solar energy production, wind energy production, and electricity demand. The database contains 12 columns, including date, station (1: Winter, 2: Spring, 3: Summer, 4: Autumn), day of the week (0: Monday, ..., 6: Sunday), DHI (W/m²), DNI (W/m²), GHI (W/m²), wind speed (m/s), humidity (%), temperature (degrees), solar energy production (MW), wind energy production (MW), and electricity demand (MW).

[Download All 7 MB](#) ⓘ

Files



Database.csv

28 MB

Institutions

Universidad Francisco de Paula Santander, University of Delaware

Categories

Forecasting, Time Series, Renewable Energy, Demand Forecasting

Dataset metrics

Usage

Views: 16
Downloads: 4



[View details >](#)

Latest version

Version 1
Published: 26 Apr 2023
DOI: 10.17632/fdffr3tc2.1

Cite this dataset

Rojas Ortega, Sebastian; Castro-Correa, Paola; Sepúlveda-Mora, Sergio; Castro-Correa, Jhon (2023), "Renewable Energy and Electricity Demand Time Series Dataset with Exogenous Variables at 5-minute Interval", Mendeley Data, V1, doi: 10.17632/fdffr3tc2.1

[Copy to clipboard](#)

Anexo 7. Código fuente disponible en: <https://github.com/SRojas28/Hybrid-model-algorithms>

The screenshot shows a GitHub repository page for 'Hybrid-model-algorithm' by user SRojas28. The repository is on the 'master' branch and has 1 branch and 0 tags. It contains 5 commits, with the most recent being 'Update README.md' 2 weeks ago. The repository size is 4.19 MB, and it has 3 stars, 0 forks, and 1 watching. The license is Apache-2.0, and it was last committed in April from Colombia. The repository description states: 'This project was created to develop a hybrid model algorithm based on machine learning to predict production, demand and dispatch energy microgrid or renewable energies, using CAISO database and NREL database. The database we used is available on Mendeley Data.' The right sidebar shows 'About' (Development of a hybrid model based on artificial intelligence for energy dispatch in a microgrid), 'Releases' (No releases published), and 'Packages' (No packages published).