Jessica Castillo

TITULACION_LOJANO_NAVAS (1).pdf

Tesis

Tesis 2024

Universidad Tecnica De Ambato- Direccion de Investigacion y Desarrollo , DIDE

Detalles del documento

Identificador de la entrega
trn:oid:::1:2987630394

Fecha de entrega
21 ago 2024, 4:52 p.m. GMT-5

Fecha de descarga
21 ago 2024, 4:54 p.m. GMT-5

Nombre de archivo
TITULACION_LOJANO_NAVAS_1_.pdf

Tamaño de archivo
1.6 MB

46 Páginas

10,031 Palabras

53,541 Caracteres

4% Similitud general

El total combinado de todas las coincidencias, incluidas las fuentes superpuestas, para ca...

Filtrado desde el informe

- Bibliografía
- Texto citado

Fuentes principales

3%	Fuentes de Internet
0%	Publicaciones
2%	Trabajos entregados (trabajos del estudiante)

Fuentes principales

3% Fuentes de Internet
0% Publicaciones
2% Trabajos entregados (trabajos del estudiante)

Fuentes principales

Las fuentes con el mayor número de coincidencias dentro de la entrega. Las fuentes superpuestas no se mostrarán.

1	Internet	
repositorio.utc.edu.ec		2%
2	Trabajos del estudiante	
Universidad Francisco de Vitoria		0%
3	Trabajos del estudiante	
Universidad Tecnica De Ambato- Direccion de Investigacion y Desarrollo , DIDE		0%
4	Trabajos del estudiante	
Corporación Universitaria Minuto de Dios,UNIMINUTO		0%
5	Publicación	
Jesús Águila León. "Modelo y desarrollo de un sistema de gestión óptima para un...		0%
6	Internet	
www.informatica-juridica.com		0%
7	Internet	
repositorio.upct.es		0%
8	Trabajos del estudiante	
Universidad Internacional de la Rioja		0%
9	Trabajos del estudiante	
Universidad Autónoma de Nuevo León		0%
10	Trabajos del estudiante	
University of the Andes		0%
11	Trabajos del estudiante	
ipn		0%

12 Internet

www.sciencegate.app 0%

RESUMEN

TÍTULO: “DISEÑO DE UNA HERRAMIENTA DE PREDICCIÓN MEDIANTE MACHINE LEARNING PARA LA GENERACIÓN FOTOVOLTAICA EN USUARIOS URBANOS Y RURALES”

Autor: Lojano Navas Cristofer Sergio

El trabajo de investigación responde a la necesidad de la generación eléctrica frente a los problemas de estiaje causados por sequías prolongadas y fenómenos climáticos con El Niño, afectando negativamente la generación hidroeléctrica, es crucial explorar nuevas fuentes de energías renovables. Los sistemas fotovoltaicos emergen como una alternativa, dado que el sol es una fuente accesible y su aprovechamiento para la generación eléctrica representa una alternativa accesible.

Para le proyecto de investigación se recopilaron datos de radiación solar del 2017 al 2013 en Tabacundo, los cuales fueron apropiadamente procesados y depurados asegurando su calidad, además, realizando una indexación en la base de datos, posteriormente son debidamente divididos en conjuntos de entrenamiento, utilizando los años 2017-2022 y parte del 2023 para entrenamiento (80%) y el resto del 2023 para validación (20%). La librería DecisionTreeRegressor permite entrenar y predecir el algoritmo en el software Python.

El modelo resulta confiable para predicciones a corto plazo con una precisión de 0.976, sensibilidad de 0.998, exactitud de 0.974 y R^2 de 0.935, sin embargo, su rendimiento a largo plazo es deficiente obteniendo resultados de MAE de 8,694 en 1 mes y 139,6 para 6 meses, se recomienda considerar otros modelos como LSTM, además, considerar mas variables para mejorar la precisión. Para le dimensionamiento fotovoltaico se requiere 15 paneles para un consumo de 160 kWh en sistemas conectados a la red y 17 en sistemas aislados, resultando más costosos debido a la necesidad de más componentes y un suministro continuo, siendo menos recomendable.

Palabras clave: predicción, radiación solar, sistema fotovoltaico, modelo de predicción.

1. INTRODUCCIÓN

En la actualidad, el precio de los combustibles fósiles y sus derivados del petróleo ha aumentado de manera significativa, al igual que su impacto ambiental. Esto hace imprescindible buscar alternativas para la producción de energía eléctrica, incrementado así el uso de fuentes de energías renovables [1]. Entre las fuentes renovables capaces de generar energía eléctrica, la solar destaca por su accesibilidad, es decir el sol se encuentra presente en todas partes, siendo una fuente de energía sostenible y eficiente.

La investigación tiene como objetivo diseñar una herramienta de predicción de radiación solar mediante Machine Learning y posteriormente dimensionar un sistema fotovoltaico conectado a la red y aislado que permita producir electricidad en función del tipo de usuario rural o urbano para la ciudad de Tabacundo, ubicada en el sector Pedro Moncayo, provincia de Pichincha. El estudio utiliza de datos de radiación, proporcionados por Nasa Power [2], de los años 2017 a 2023, con mediciones realizadas cada hora (7: 00 am – 19:00 pm). La radiación solar es una variable que cambia en función del tiempo, siendo indispensable para la generación fotovoltaica, por lo que una predicción confiable y precisa de la radiación es fundamental para el dimensionamiento adecuado de los sistemas fotovoltaicos [3].

Existen diversidad de técnicas de Machine Learning que permiten desarrollar algoritmos con la capacidad de predecir radiación para la generación fotovoltaica [4]. Esto permite desarrollar análisis y tomar decisiones estratégicas para la generación fotovoltaica. La capacidad de estas técnicas de predicción permite mejorar los sistemas fotovoltaicos, contribuyendo con recursos energéticos, reduciendo la dependencia de combustibles fósiles e incentivando la utilización de fuentes renovables [5].

1.1 SITUACIÓN PROBLÉMICA

El avance del país en el sector eléctrico ha experimentado grandes aumentos, debido al uso cada vez mayor de fuentes renovables. Estas estrategias generan no solo ahorros económicos, además, desencadenan efectos positivos para el medio ambiente al preservar recursos energéticos, así como también la reducción de gases contaminantes, contribuyendo a la matriz energética permitiendo la independencia de recursos no renovables [6].

En muchas áreas la carencia de disponibilidad de la red eléctrica es un problema significativo, afectando negativamente la calidad de vida al no contar con los servicios necesarios, limitando el acceso a una mejor educación, atención médica, la falta de iluminación incrementa el riesgo de accidentes y robos, impidiendo su crecimiento y desarrollo.

Además, el desconocimiento de la potencia requerida para energizar un hogar impide el dimensionamiento adecuado de sistemas fotovoltaicos, como resultado puede recurrir a dependencias de generadores de diésel o gasolina, siendo poco accesibles por su elevado costo y el impacto ambiental negativo que provocan.

1.2 Formulación del problema

La ausencia de conocimiento de herramientas de predicción de radiación solar dificulta la obtención de estimaciones de la potencia necesaria para el dimensionamiento de sistemas fotovoltaicos, impactando negativamente a su desarrollo.

1.3 OBJETO Y CAMPO DE ACCIÓ

1.3.1 Objeto de Investigación:

Herramienta de predicción mediante Machine Learning para la generación fotovoltaica.

1.3.2 Campo de Acción:

07 Ingeniería, Industria y Construcción / 071 Ingeniería y Profesiones Afines / 0713 Electricidad y Energía.

1.4 BENEFICIARIOS

1.4.1 Directo

Sectores rurales y urbanos

1.4.2 Indirecto

Comunidad universitaria

1.5 JUSTIFICACIÓN

El presente proyecto de investigación se realiza previo a la obtención del título de Ingeniero Eléctrico de la Universidad Técnica de Cotopaxi, el mismo que se alinea con

los objetivos de la carrera de Electricidad “Energías alternativas y renovables, eficiencia energética y protección ambiental, asociado a la Sublínea Inteligencia artificial y modelación de sistemas.

En años recientes las fuentes renovables han emergido como alternativas importantes en áreas rurales y urbanas muy alejadas consideradas marginadas, ya que han enfrentado complicaciones en la llegada de la red eléctrica debido a diversos factores, como las características de terreno, condiciones ambientales, entre otros aspectos que los dificultan.

El propósito de estudio es diseñar una herramienta de predicción para la generación fotovoltaica que proporcione información sobre la potencia necesaria para la implementación de paneles fotovoltaicos. Esta herramienta utiliza técnicas de aprendizaje automático con el objetivo de prever la radiación solar, dado que esta variable es crucial para la generación eléctrica, permitiendo una planificación segura y eficiente para la generación fotovoltaica.

El aprendizaje automático (ML) permite realizar predicciones precisas al utilizar múltiples técnicas avanzadas como el reconocimiento de patrones a largo plazo, capacidad de analizar grandes volúmenes de información, clasificar tipos de datos, lo que permite tomar decisiones para prevenir riesgos, implementar mejoras en sus procedimientos con la finalidad de optimizar un trabajo asignado.

1.6 OBJETIVOS

1.6.1 General:

Diseñar una herramienta de predicción mediante Machine Learning para la generación fotovoltaica en usuarios urbanos y rurales.

1.6.2 Específicos:

- Recopilar datos para el desarrollo de la investigación.
- Tratar la base de datos.
- Identificar los parámetros para la implementación del modelo matemático de predicción de radiación solar.
- Validar el programa.

1.6.3 Sistema de Tareas

Objetivos específicos	Actividades (tareas)	Resultados esperados	Técnicas, Medios e Instrumentos
Recopilar datos para el desarrollo de la investigación.	• Investigar información de la investigación • Crear una estructura de base de datos	• Base de datos de la radiación. • Registro de fuente de datos sobre la radiación.	• Nasa Power • Tesis • Artículos científicos
Tratar la base de datos	• Aplicar técnicas de limpieza de información asegurando su calidad. • Utilizar técnicas para el manejo de datos faltantes garantizando infroamción eficiente	• Base de datos limpia y estructurada • Mejor calidad de datos para el algoritmo de predicción	• Software Python • Tesis
Identificar los parámetros para la implementación del modelo matemático de	• Indagar parámetros utilizados en estudios previos. • Investigar algoritmos	• Selección de los parámetros de la herramienta de predicción.	• Tesis. • Bibliografías relacionas a herramientas de predicción.

predicción de radiación solar.	de predicción.		
Validar el modelo	• Evaluar mediante métricas de validación. • Comparar los datos de la predicción con datos reales.	• Evaluación de precisión del modelo • Análisis comparativo	• Software Python • Analizador de radiación

1.7 REVISÓN BIBLIOGRÁFICA

La evolución tecnológica ha ido creciendo, dando lugar al surgimiento de innumerables herramientas facilitando su calidad de vida. Dado que la electricidad es crucial para el desarrollo, existen herramientas capaces de aproximar una cantidad requerida de potencia para la implementación de paneles fotovoltaicos con el propósito de proveer electricidad. Estas tecnologías se han convertido en alternativas viables en zonas donde la red eléctrica a resultado inaccesible, así también se han empleado para reducir los costos de planillas eléctricas, además de ellos resultando positivo para el medio ambiente.

En [7], se presenta un modelo de aprendizaje automático que permite predecir la radiación, lo que permite estimarla de manera diaria o semanales, corto tiempo. Esta estimación se basa en la instalación de una estación meteorológica, la cual registra datos climáticos permitiendo crear un historial de parámetros climáticos. La investigación tiene como objetivo utilizar el aprendizaje automático para planificar futuras generaciones de electricidad mediante paneles fotovoltaicos.

En [8], presenta un modelo de predicción de generación fotovoltaica mediante técnicas denominadas minería de datos. Esta metodología implica un análisis de variables de decisión, univariante como multivariante, con el objetivo de comprender su

comportamiento y su relación con la generación eléctrica. Para llegar a entrenar las variables, se emplea la técnica de árbol de decisión mediante bosques aleatorios, priorizando aquellas con mayor influencia en la estimación de la producción de energía. Este modelo se ha implementado en la comunidad de Paragachi (Imbabura, Ecuador) que cuenta con una central fotovoltaica con 14400 paneles fotovoltaicos, teniendo una potencia nominal de 3,6 MW.

En [9], se emplea un algoritmo de RN que permite predecir la radiación media horaria de las próximas 24 horas. Este modelo considera el cálculo de parámetros de tercer derivado de orden (TOD) permitiendo así detectar las variaciones en la radiación, así también como la diferencia discreta normalizada (NDD) para clasificar los días como nublados o claros. Además, la estructura del modelo se termina mediante validación y se utiliza el método Levenberg-Marquardt para resolver problemas de mínimos cuadrados no lineales.

En [10], expone un modelo de pronóstico de potencia eléctrica a partir de paneles solares, aplicando técnicas de aprendizaje automáticos clásicas y profundas en ubicaciones de la India. Para el desarrollo del modelo, se ha utilizado información de un conjunto de datos de “Solar Power Generation Data” que registra la producción de energía durante 34 días en dos plantas generadoras, con intervalos de 15 minutos. Estos datos incluyen variables de campo de generación y mediciones de sensores. Este estudio evidencia la viabilidad del desarrollo de modelos que pueden llegar a ser instrumentos valiosos para decidir la gestión de la generación eléctrica. Específicamente, estos modelos pueden predecir la potencia que se va a producir para los próximos días, lo que permite mejorar la gestión de redes y sistemas de potencia.

En [11], desarrolla un modelo dinámico basado en dos submodelos que permiten predecir la generación eléctrica en una planta fotovoltaica. El primer submodelo se enfoca en la relación de la irradiación solar, por otra parte, el segundo submodelo describe la relación entre la irradiación y la potencia generada. Los parámetros que considerados para estos modelos son la irradiación y la temperatura. El objeto de investigación es una planta fotovoltaica ubicada en la Universidad Central “Marta Abreu”, Cuba. Cuenta con una potencia nominal de 1,1 MW, además fue puesta en funcionamiento en 2019 constando con 200 bandejas de módulos de paneles fotovoltaicos, produciendo un voltaje de 700 V. El modelo a utilizado datos históricos de la predicción de la planta generadora. Las predicciones del modelo han resultado acertadas, demostrando su efectividad.

En [12], ha desarrollado de predicción para la generación fotovoltaica a corto plazo, basada en técnica de minería, mediante datos aplicadas a una base de datos históricos de producción de una planta de generación fotovoltaica, este modelo se ha creado a partir de información de varios modelos que utilizan distintas técnicas de predicción y aprendizaje. El procedimiento realizado incluyó varias pruebas para obtener predicciones a corto plazo, con el objetivo de instalar una planta fotovoltaica conectada a la red eléctrica. Además, se compararon los resultados con otros modelos, pudiendo determinar que los resultados de predicción obtenidos son superiores a los demás, con mejor precisión y menor porcentaje de error. Para el modelo se utilizaron herramientas como árboles de decisión, algoritmos evolutivos y un sistema de reglas. Además, se proporcionó un punto de predicción de potencia e incertidumbre, mejorando la selección de decisiones.

En [13], ha diseñado una herramienta de predicción de la generación fotovoltaica mediante Deep Learning a corto plazo permitiendo el pronóstico para el siguiente día o la siguiente hora, a partir de datos históricos de SCADA y variables meteorológicas locales, para ello se ha utilizado LSTM, BiLSTM y GRU, se ha logrado este modelo mediante software Matlab desarrollando varios modelos de redes neuronales , constando de variables meteorológicas, su objetivo fue validarlo para un parque fotovoltaico de 5,5 M, mediante los modelos realizados en redes neuronales se ha podido decir que los más eficientes que tiene una mejor precisión y menos porcentaje de error son los que emplean horizontes de tiempo más corto y redes BiLSTM, es decir que tiene una capa RNN que permiten aprender dependencias bidireccionales para las unidades y series de tiempos y datos secuenciales.

En [14], plantea un modelo mediante PowerViewer y una base de datos proporcionados por IDEAM sobre las estaciones meteorológicas para la asignación de radiación. Este modelo emplea técnicas de predicción como la regresión lineal simple y múltiple y se desarrolla en cuatro etapas: Apertura y procesamiento de datos mediante IDEAM; utiliza información de PowerViewer; pronóstico información; análisis de datos y reduciendo errores en la medición de la radiación solar.

En [15], presenta un modelo de producción de generación fotovoltaica, que ha sido analizada mediante varios algoritmos de aprendizaje automático (Machine Learning) así como varios modelos determinando el más preciso. Este análisis considera diversos parámetros, incluyendo escenarios climatológicos, con el objetivo de determinar qué tipo de modelo se aproxima con mayor precisión a la curva real de generación.

En [16], postulan un modelo de predicción mediante Machine Learning utilizando herramientas como Bosques Aleatorios para la generación fotovoltaica en la Haciendo "Campiña" ubicada en Mulaló, Latacunga. Este modelo considera variables como la irradiación solar y temperatura. El modelo integra datos de las variables que se ingresan para entrenar el modelo y brindar predicciones precisas. Utilizando árboles de decisión mediante Python, este modelo proporciona resultados de eficiencia diaria de 91,41% y mensual de 94.47% siendo más efectiva por la mayor cantidad de datos.

1.8 HIPÓTESIS

¿La predicción de radiación solar logrará proporcionar una estimación de la potencia para la generación fotovoltaica?

2. MARCO TEÓRICO:

2.1 ENERGIAS ALTERNATIVAS

Existen diversas maneras de producir energía mediante fuentes alternativas renovables, caracterizadas por ser limpias e inagotables. La producción total de energía del (S.N.I) representa un 82,23% [17]. En la Figura 3.1 se ilustran porcentajes, las fuentes hidráulicas representan un 81,08% equivalente a 20.661,59 GWh. La energía eólica tiene un 0.29% generando 73,7 GWh. Biomasa contribuye un 1,05 % lo que equivale a 382,44 GWh, mientras que el biogás contribuye con un 0,185% generado un 45,52 GWh. Por último, la energía fotovoltaica representa un 0,14% con una producción de 73,3 GWh del total de energía renovable del SIN [17].

Las energías renovables ayudan a ser más independiente de otros recursos, así también ayudan a mitigar el cambio climático, permitiendo reducir emisiones de CO2. La implementación de tecnologías renovables es esencial para el desarrollo y futuro energético.

Producción Total de Energía e Importaciones S.N.I.		GWh	%
Energía Renovable	Hidráulica	20.661,59	81,08%
	Eólica	73,7	0,29%
	Fotovoltaica	34,77	0,14%
	Biomasa	382,44	1,50%
	Biogás	45,52	0,18%
Total Energía Renovable S.N.I.		**21.198,02**	**82,83%**

Figura 2.1 Total Energía Renovable S.N.I [17] .

2.2 ENERGÍA FOTOVOLTAICA.

Esta energía se obtiene mediante una transformación directa del sol, permitiendo generar energía, utilizando componentes tales como placas formadas por módulos, denominadas células fotovoltaicas. Estas células están compuestas por materiales semiconductores que y permiten minimizar las pérdidas de calor. La conversión solar en electricidad ocurre a nivel atómico. Se presenta un efecto conocido como fotoeléctrico, el cual permite captar fotones de luz y liberar electrones. Cuando este efecto se produce, los electrones disponibles son atrapados, esto genera un flujo de corriente que puede ser aprovechado y utilizado como electricidad [18]. Edmundo Bequerel, es un físico francés, fue el primero es observar y descubrir el efecto fotoeléctrico que describe como determinados

materiales, cuando son expuestos a la luz, pueden generar pequeños corrientes de electricidad. [18]. Las aplicaciones fotovoltaicas han evolucionado significativamente desde simples maquetas de aviones, automóviles y radios pequeñas, hasta la actualidad que son capaces de alimentar viviendas enteras, así también como establecer plantas de generación fotovoltaicas que contribuyen a la producción energética de un país.

La habilidad para transformar la energía solar en electricidad ha generado muchos beneficios a lo largo de los años. En lugares donde la red eléctrica no puede llegar, la generación de electricidad a partir de paneles fotovoltaicos se ha convertido en una alternativa sostenible, permitiendo su desarrollo.

2.3 RADIACIÓN SOLAR

Energía transmitida por el Sol llega a la atmósfera en forma de radiación electromagnética, alcanzando una intensidad de 1000 W/m^2 (vatios por metro cuadrado) en la superficie terrestre [19]. Existen tres tipos de radiación solar según la incidencia del os rayos solares:

- Directa
- Difusa
- Reflejada

2.4 MACHINE LEARNING

Representa un subcampo de IA basado en el campo de aprendizaje automático, cuyo objetivo es desarrollar métodos que permitan a los algoritmos descubrir patrones repetitivos en conjuntos de datos, como números, palabras y estadística. Estos algoritmos permiten aprender y mejorar eficiencia [20].

El aprendizaje automático está compuesto por algoritmos que han beneficiado al ser humado. Además, han contribuido significativamente en muchos campos de estudio, desde la ingeniería, salud, predicciones financieras, mercado de inversiones, entre otros. El Machine Learning se clasifica en supervisado, no supervisado y reforzado [20].

2.5 MODELOS DE PREDICCIÓN

También conocidos como modelos predictivos, son un conjunto de herramientas y técnicas estadísticas que nos sirven y ayudan a pronosticar el comportamiento de nuevos eventos a futuro, ya que nos permite tomar decisiones informadas, optimizar recursos e

identificar oportunidades y reducir riesgos. Para los modelos predictivos es necesario recopilar datos históricos, los cuales son utilizados para reconocer patrones que facilitan la predicción [21]. Sus aplicaciones son múltiples, ya que pueden emplearse a cualquier tipo de datos que se deseen predecir, como campos de marketing, electricidad y diferentes áreas de estudio.

2.6 TIPOS DE MACHINE LEARNING

2.6.1 Aprendizaje supervisado

El aprendizaje automático es un sistema que se entrena utilizando datos previamente seleccionados y etiquetados, permitiéndole reconocer y encontrar patrones aplicables en un análisis. Esto resultado en una salida predefinida y conocida [22].

Además, el aprendizaje supervisado puede ser de dos tipos: clasificación o regresión. La clasificación se enfoca en relacionar características con diferentes etiquetas para obtener un resultado categórico. Por otro lado, la regresión tiene como objetivo la obtención de un valor numérico específico, centrando su análisis en las etiquetas numéricas.

2.6.2 Aprendizaje no supervisado.

El aprendizaje no supervisado es un sistema que se enfoca en buscar similitudes entre los datos sin utilizar etiquetas, ya que los datos no están previamente etiquetados, lo que impide que el aprendizaje automático detecte tipos específicos de datos sin una etiqueta previa. Se alimenta de datos que permiten agrupar tareas y se define como un modelo predictivo similar al aprendizaje supervisado, pero que trabaja con datos no clasificados, describiendo patrones de ejemplos similares entre grupos de datos [22]. Este tipo de aprendizaje se clasifica en reducción y clustering. La reducción se enfoca en disminuir el número de variables a considerar para facilitar la toma de decisiones. El clustering, permite clasificar o agripar los datos según sus características similares entre sí.

2.6.3 Aprendizaje reforzado

Este tipo de aprendizaje no pertenece al supervisado y no supervisado. Es un sistema en el que no existe información previa sobre una posible salida y se basa en acciones y resultados obtenidos. Se caracteriza por ser un proceso de ensayo y ajuste hasta encontrar una forma ideal de completar una tarea [22].

2.6.4 Procedimiento para un construir un modelo de Machine Learning

La creación de un modelo de Machine Learning sigue seis pasos fundamentales que deben ser seguidos por la figura 3.2.

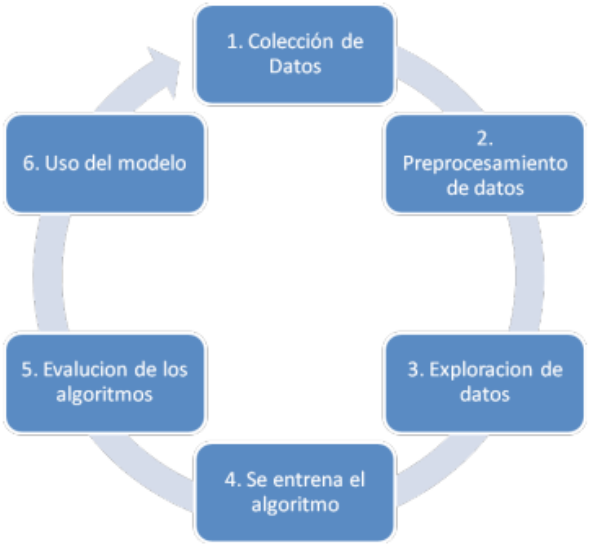


Figura 2.2 Pasos para construir un modelo de ML [22].

Procedimiento:

1. Colección de datos: los datos se extraen de diversas fuentes como bibliografías, sitios web, programaciones, estadísticas, entre otras. El objetivo es obtener información completa
2. Preprocesamiento de información: utilizando la información recopiladas, se realizan múltiples tareas de preprocesamiento. Esto incluye depurar y selecciona los datos asegurando que todos tengan el formato correcto para alimentar el algoritmo de aprendizaje.
3. Exploración de datos: se realiza un análisis preliminar para corregir valores faltantes y descubrir patrones que faciliten la construcción del modelo. Se identifican valores atípicos y se determinan las características más influyentes para realizar predicciones.
4. Entrena el algoritmo: con los pasos anteriores completados, se procede al entrenamiento del algoritmo. El objetivo es que el algoritmo extraiga información útil que permita realizar predicciones precisas.
5. Evaluación de los algoritmos: se evalúa la precisión del algoritmo en sus predicciones comparando los resultados con el conocimiento previo obteniendo

durante el entrenamiento. Si las predicciones no alcanzan un rendimiento aceptable, se regresa a la etapa anterior para ajustar parámetros y mejorar el rendimiento hasta obtener resultados aceptables.

6. Uso del modelo: una vez el modelo ha alcanzo un rendimiento aceptable, se puede aplicar para cualquier tarea deseada.

2.7 TÉCNICAS DE MACHINE LEARNING

2.7.1 Redes Neuronales Artificiales

Son múltiples redes interconectadas de manera paralela de elementos simples [23]. El concepto de red neuronal surgió de la intención de crear un sistema artificial que sea capaz de ejecutar tareas de manera similar a un cerebro humano. Este sistema adquiere información y conocimientos a través del aprendizaje que posteriormente se almacenan [20]. La red neuronal más se ilustra en la Figura 3.3.

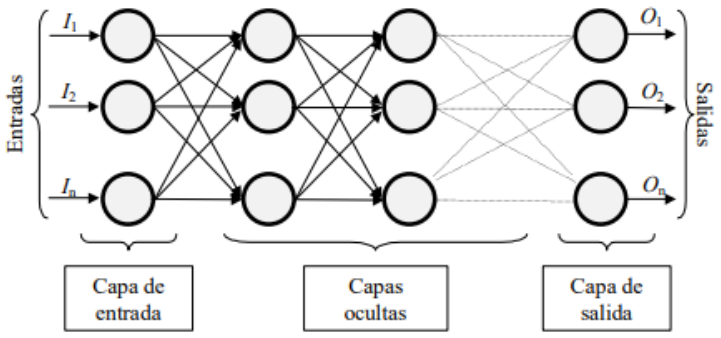


Figura 2.3 Red Neuronal simple de una capa [23].

2.7.2 Árbol de decisión

El sistema de aprendizaje con estructura de árbol, se observa en la Figura 3.4. Un nodo central denota una característica, una rama refleja una norma de decisión y cada nodo terminal representa un desenlace, y toda la información se encuentra en la raíz. [24]. Esta técnica, incluye un conjunto de normas de clasificación vinculadas a una etiqueta de clase particular al término de cada ramificación [24]. Esta técnica aprende a segmentar los datos en función de los valores que contienen los atributos, dividiéndolos de manera recursiva con el objetivo de encontrar un mínimo local utilizando estrategias como la entropía.

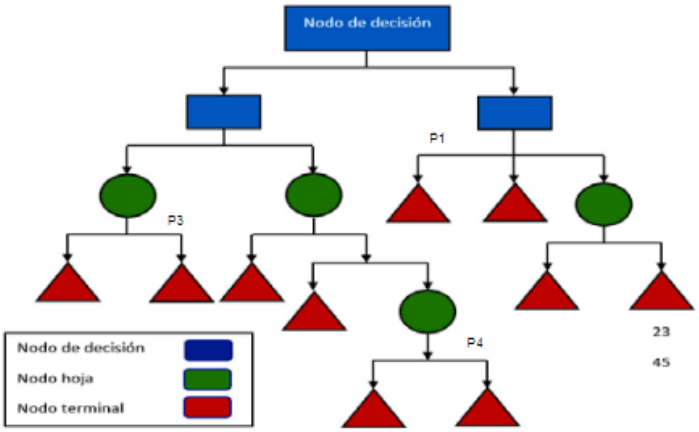


Figura 2.4 Estructura de un árbol de clasificación [24]

2.7.3 Memoria a Corto Plazo y Larga Duración (LSTM)

La red neuronal LSTM es una red neuronal capaz de captar dependencias en mucho tiempo. Fue diseñada para superar el problema de dependencia a largo plazo, permitiendo retener y retener información durante períodos prolongados, lo cual es su característica inherente, facilitando así el aprendizaje [20]. El LSTM se utiliza para predecir secuencia de datos. En la Figura 3.5 se ilustra una celda LSTM de 4 puertas.

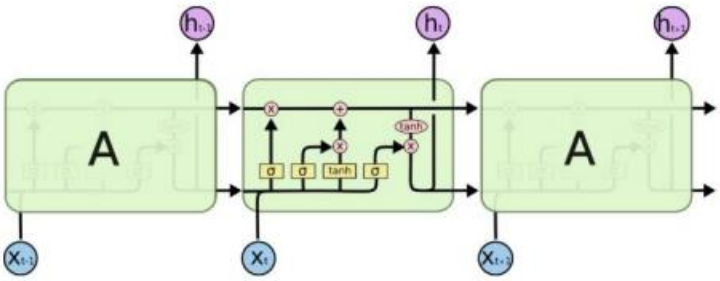


Figura 2.5 Celda LSTM 4 puertas [20].

2.7.4 Red Neuronal Recurrente (RNN)

Una RNN (Red Recurrent Network) es una red compuesta por nodos, semejantes a neuronas, dispuestos en capas secuenciales, donde cada nodo en una capa se conecta de forma unidireccional con cada nodo en la capa siguiente. Tienen una activación de valor real que varía en el tiempo. Los nodos de entrada capturan información externos a la red, los nodos ocultos procesan estos datos para producir salidas, y los nodos de salida generan

los resultados finales. de manera cíclica en una cadena repetitiva [20]. En la figura 3.6 se representa una RNN.

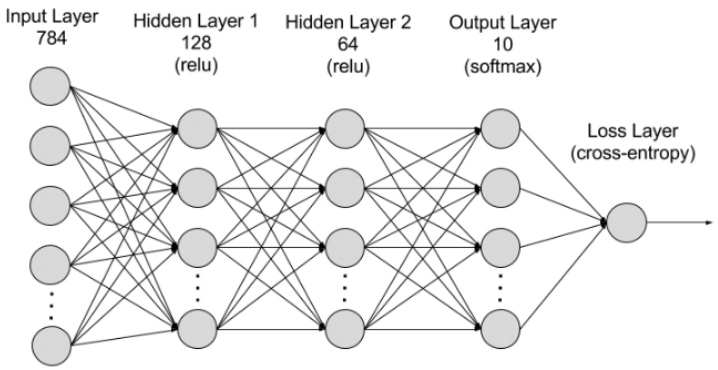


Figura 2.6 Diagrama de una RNN simple [20].

2.8 SISTEMA FOTOLTAICO

2.8.1 TIPOS DE SISTEMAS FOTOVOLTAICO

Los sistemas eléctricos fotovoltaicos son capaces de capturar energía solar y transformarla en energía eléctrica gracias a sus componente mecánicos, eléctricos y electrónicos. En la figura 3.7 se ilustra la clasificación de los sistemas fotovoltaicos [27].

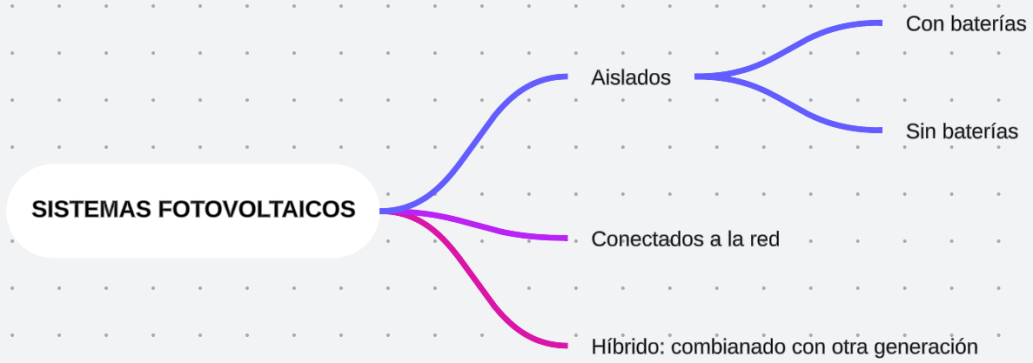


Figura 2.7 Clasificación de sistemas fotovoltaicos

2.8.2 SISTEMAS AISLADOS

Los sistemas fotovoltaicos asilados tienen como objetivo principal cubrir una parte significativa, o idealmente toda la demanda eléctrica en áreas donde la red de distribución eléctrica no ha podido llegar debido a diversas dificultades, como la geografía del terreno, la baja densidad de población en la zona, o los alto costos asociados con la infraestructura necesaria para suministrar electricidad [27].

2.8.3 SISTEMAS HIBRIDOS

Estos sistemas tienen la capacidad de integrarse o combinarse con otras fuentes de generación para mejorar la disponibilidad y estabilidad del suministro energético. Cuando una instalación fotovoltaica se combina con otro tipo de generador, el sistema resultante se denomina sistema híbrido. Este enfoque híbrido busca garantizar una mayor fiabilidad en el suministro energético al aprovechar los beneficios de complementarse a diferentes fuentes de generación eléctrica [27]. La Figura 3.8 ilustra la facilidad del sistema al integrarse con otras fuentes de energía.

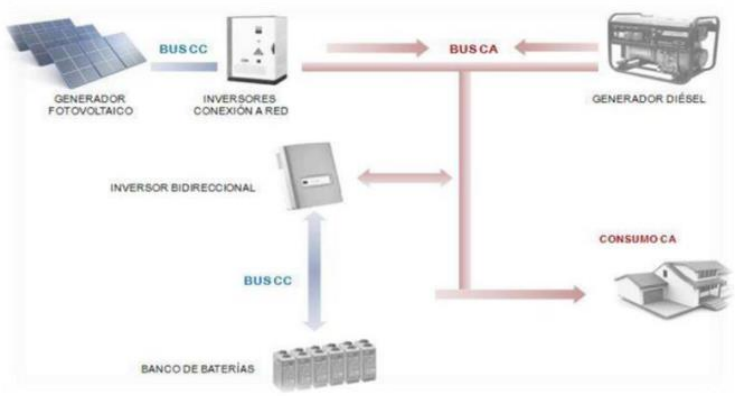


Figura 2.8 Sistema fotovoltaico híbrido [27]

2.8.4 SISTEMAS ON GRID

Se utilizan en áreas donde existe una red pública de electricidad. Además de proporcionar un suministro confiable de electricidad, estos sistemas buscan reducir los costos de la tarifa eléctrica al aprovechar tanto la conexión al Sistema Interconectado como la radiación solar. Durante el día, cuando la radiación solar es fuerte, y su fuente es empleada para abastecer la demanda de electricidad. Por la noche, o cuando la energía solar no es suficiente, se recurre a la red pública para satisfacer la demanda [28].

2.9 COMPONENTES DEL SISTEMA FOTOVOLTAICO

2.9.1 Generador fotovoltaico

Este componente es indispensable ya que su función es generar energía mediante la absorción de radiación solar, convirtiéndola en una fuente de generación [28].

2.9.2 Inversor

Es un dispositivo electrónico con la capacidad para convertir la corriente directa (CD) generada por el sistema en corriente alterna (CA), requerida para suministrar energía eléctrica a los componentes conectados al sistema [28].

2.9.3 Acumulador

Tiene la responsabilidad de almacenar energía excedente con el objetivo de que pueda ser utilizado cuando en el sistema ya no haya recursos solares disponibles para la generación eléctrica [28].

2.9.4 Regulador de carga

Es el dispositivo responsable de controlar el proceso de cargar y descargar la batería garantizando su protección. [28].

2.9.5 Paneles solares

Son un conjunto de celulares fotovoltaicas que tienen una configuración tanto en serie como en paralelo. El objetivo de combinar estas células es aumentar la producción de energía eléctrica, permitiendo que el sistema alcance una escala comercial y sea capaz de alimentar dispositivos [29]. Es importante destacar que la potencia de un panel solar está relacionado y determinado por el rendimiento de sus materiales.

2.9.6 Panel monocristalino

Estos paneles están fabricados con cristal de silicio altamente puros. Los paneles tienen una forma cilíndrica y presentan un color azul oscuro con un brillo metálico, alcanzando un rendimiento de hasta el 17% [29]. En la Figura 3.9 se observa el panel monocristalino.

Figura 2.9 Panel monocristalino [29].

2.9.7 Panel policristalino

Se distinguen por el color azul de sus celdas y generalmente tiene una capacidad de generación que varía entre 5 W y 250 W [29]. Son ampliamente utilizados en zonas residenciales y comerciales debido a su costo más accesible. En la Figura 3.10 se ilustra un panel policristalino.

Figura 2.10 Panel monocristalino solar [29].

2.9.8 Selección de paneles fotovoltaicos

Su clasificación se terminar por el tipo de tecnología empleada, ya que los paneles solares varían en eficiencia según las celdas que los componen. Un mayor rendimiento y eficiencia en las celdas implican la necesidad de utilizar más o menos celdas para alcanzar la capacidad deseada [30]. Esto implica costos variables para las instalaciones solares, los cuales pueden ser alto o bajos en función a la demanda energética a cubrir. La la figura 3.11 ilustran los tipos sus tipos.

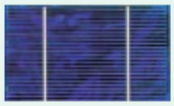
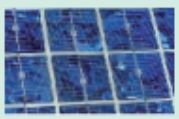
Células	Silicio	Rendimiento laboratorio	Rendimiento directo
	Monocristalino	24 %	15 - 18 %
	Policristalino	19 - 20 %	12 - 14 %
	Amorfo	16 %	< 10 %

Figura 2.11 Características de los paneles solares según su tecnología [30].

2.10 DIMENSIONAMIENTO DE LOS PANELES

Se debe considerar varios factores críticos como la ubicación de los paneles, al no estar debidamente colocado afecta su habilidad para captar la radiación solar. La eficiencia de los paneles fotovoltaicos es convertir la radiación solar en energía eléctrica. Además, las pérdidas de energía debido a componentes del sistema, como inversores, baterías y conexiones. Estas pérdidas se han considerado al calcular la energía suministrada aplicando un factor de ajuste del 90% para reflejar las ineficiencias [31]. La ecuación 3.10 se utiliza para determinar la cantidad de paneles necesarias para una instalación.

$$N_T = \frac{P_{GFV}}{P_{MPP}} \tag{3.10}$$

Dónde:

N_T= cantidad de paneles

P_{GFV}= potencia generada por el sistema fotovoltaico

P_{MPP}= potencia del panel fotovoltaico.

2.10.1.1.1 Inversor

Componente capaz de recibir y convertir la energía de corriente continua a corriente alterna [32]. La ecuación 3.11 me permite calcular el inversor.

$$Pinversor = P_{pv} * FS \tag{3.11}$$

Dónde:

P_{pv} = Potencia del sistema

FS= factor de seguridad

2.10.1.1.2 Capacidad de las baterías

La capacidad de las baterías permite determinar la cantidad de energía que puede almacenarse para consumo [32]. La ecuación 3.12 y 3.13 permite calcular las baterías.

$$Cbatería = \frac{E_{batería}}{V_{batería} * D_{DOD} * \eta_{bateria}} \quad (3.12)$$

$$E_{batería} = E_{diaria} * D_a \quad (3.13)$$

Dónde:

$E_{batería}$= energía almacenada en las baterías.

$V_{batería}$= voltaje.

D_{DOD}= profundidad de descarga de la batería.

$\eta_{bateria}$= eficiencia de la batería.

E_{diaria}= energía diaria.

D_a= días de autonomía.

2.11 SOTFWARE DE SIMULACIÓN

2.11.1 Python

Python es un programa de alto nivel. Su diseño permite a los programadores escribir códigos de manera más eficiente y con menores errores, facilitando la creación de una diversidad de aplicaciones [33]

Este lenguaje se basa en una serie de elementos y estructuras que confirman su sintaxis. Entre estos elementos se encuentran las variables, que permiten guardar información que puede modificarse mientras el programa está en ejecución. En Python, cada variable se identifica por un nombre y contiene un valor asociado, lo que facilita la manipulación de información. Estas características, están diseñadas para optimizar el desarrollo de software, haciendo que los procesos sean más rápidos y efectivos. Además, permite realizar tareas de manera ágil y precisa, garantizando resultados seguros y exactos [33].

3. MÉTODOS Y PROCEDIMIENTOS

3.1 DESCRIPCIÓN DEL PROYECTO

El presente proyecto tiene como objetivo diseñar una herramienta de predicción de la radiación solar para la generación fotovoltaica utilizando técnicas de Machine Learning "Arboles de decisión" mediante el software Python. Se ha recopilado datos históricos de radiación solar del año 2017 a 2023 de la provincia de Pichincha, desde las 7:00 am hasta las 19:00 pm, ya que en horas de la noche no se registran valores debido a la ausencia de luz solar. Con estos datos de radiación, se procede a realizar el dimensionamiento de un sistema fotovoltaico on-grid y off-grid, en función de las características de las zonas rurales y urbanas.

El proyecto de investigación incluye la validación del modelo de Machine Learning mediante métricas como la exactitud, matriz de confusión, precisión y sensibilidad. Además, se realiza una comparación del comportamiento de la radiación predicha con una base de datos obtenida con un analizador de radiación de la misma zona, lo que permite evaluar el comportamiento de la predicción frente a mediciones reales realizadas en campo.

El cambio hacia fuentes de energía renovables y la generación fotovoltaica es crucial, ya que es una de las fuentes de energía más implementadas. La predicción de la radiación es esencial para maximizar la eficiencia de los sistemas fotovoltaicos y asegurar un suministro energético sostenible. El proyecto se desarrolla en un contexto donde la demanda de energías renovables está en aumento, y estas herramientas de predicción ofrecen ventajas significativas.

3.2 ÁREA DE ESTUDIO

La figura 4.1 se ilustra el área de estudio en la provincia de Pichincha, específicamente en Tabacundo, en el sector Pedro Moncayo. La zona de estudio se encuentra a una altitud de 2872 msnm, una latitud de 0° 2' 45.29" N y longitud de 78° 12' 49.01" O.

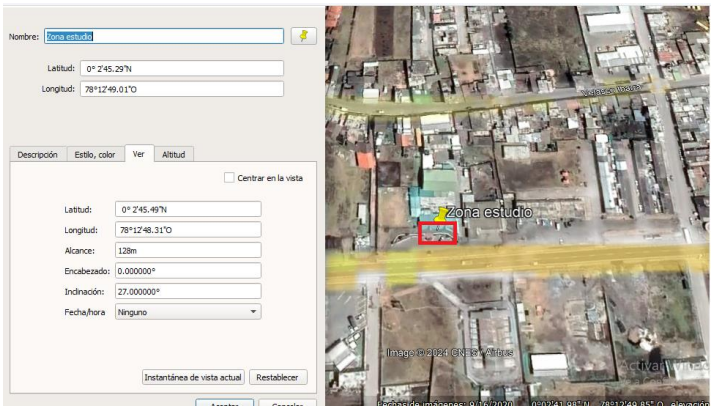


Figura 3.1 Zona para la propuesta.

3.3 SELECCIÓN DE TÉCNICAS DE MACHINE LEARNING

Existen diversos métodos y técnicas de Machine Learning para la predicción. En la Tabla 4.1, se ilustra varias técnicas realizando una comparación los modelos.

Tabla 3.1 Cuadro comparativo de técnica de Machine Learning.

Técnica	Descripción	Ventajas	Desventajas	Aplicación en la predicción de la radiación
Árbol de decisión	Técnicas basadas en reglas de decisiones	• Fácil de entender e interpretar • No requiere normalización de datos	• Pueden sobre ajustarse • Sensibles a pequeñas variaciones	• Eficaz para detectar relaciones sencillas en datos de radiación
LSTM	Tipo de red neuronal recurrente especializada para entrenamiento a largo plazo	• Capacidad para aprender dependencias a largo plazo. • Evita problemas del desvanecimiento del gradiente.	• Tiempo de entrenamiento prolongado • Necesita grandes volúmenes de datos	• Ideal para analizar series temporales y patrones complejos de radiación.
RNN	Red neuronal recurrente que procesar secuencia de datos,	• Eficaz con datos secuenciales • Menos compleja que el LSTM	• Problemas con el desvanecimiento del gradiente • No tiene la capacidad de	• Adecuada para los patrones de radiación de

	recordando información de estados previos		aprender dependencias a largo plazo	corto a medio plazo.
SVM	Máquina de vectores de soporte, utilizada comúnmente para clasificación y regresión	• Alta precisión • Eficaz en espacios de alta dimensionalidad	• Dificultada para interpretar parámetros • Requiere ajustes de parámetros	Útil para clasificar patrones de radiación

Considerando la comparación de modelos en la tabla 4.1, se puede concluir que el árbol de decisión es la técnica de aprendizaje más recomendable por varias razones. En primer lugar, su sencilla interpretación lo cual facilita la comprensión de los resultados, ofrece una estructura clara y comprensible de los datos. En comparación de otras técnicas de Machine Learning, los árboles de decisión no requieren un extenso entrenamiento y un alto esfuerzo computacional. Por estas características, se destacan como una opción eficiente para la predicción de la radiación.

3.4 METODOLOGÍA

La propuesta del proyecto sigue la metodología detallada en el diagrama de flujo ilustrada en la Figura 4.2. El proceso comienza con la lectura y carga de datos en el programa, para lo cual se ha creado previamente una base de datos sobre radiación solar. Esta base de datos contiene información de radiación para un período de 12 horas a lo largo del día. A continuación, se realiza el preprocesamiento o tratamiento de datos, basado en la preparación y transformación de datos antes de su uso en el modelamiento. El preprocesamiento incluye la limpieza de datos, lo que permite verificar y asegurar la calidad de los datos para optimizar su rendimiento. También, se procede con la indexación de datos, lo que facilita una mejor gestión del almacenamiento y la utilización eficiente de la información sobre la radiación.

La siguiente etapa es la división de grupo de entrenamiento de entramiento y grupo de prueba. Para ello, se han utilizado datos del período 2017-2022 y una parte del año 2023

como datos de entrenamiento. El resto del año 2023 se emplea como datos de validación. Es importante destacar que, para esta separación, se utiliza una proporción de 80-20. Posteriormente, se lleva a cabo la ejecución del programa para obtener la información de la predicción de la radiación mediante el entrenamiento del modelo usando árboles de decisión. Además, se calculan las métricas de validación para evaluar la precisión del modelo. Con estos datos de predicción, nos permiten realizar los cálculos necesarios para el sistema fotovoltaico

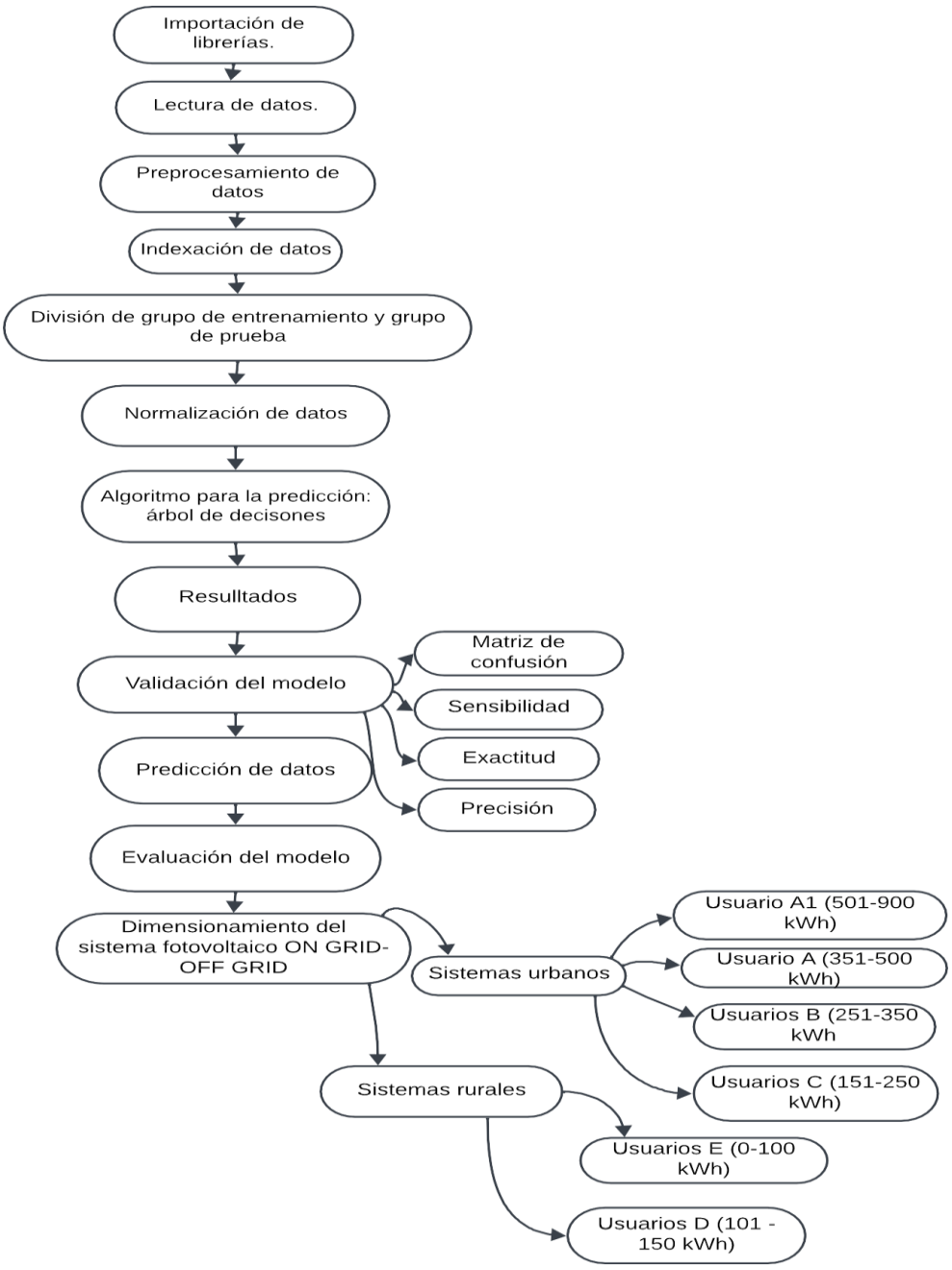


Figura 3.2 Metodología para la predicción de la radiación

3.5 IMPLEMENTACIÓN EN PYTHON

3.5.1 Utilización de librerías

Para el desarrollo de modelos predictivos y la realización de análisis de datos en Python, se utilizan diversas librerías, como se muestra en la Figura 4.3 que incluyen pandas, numpy, matplotlib, que facilitan la manipulación de datos y visualización de gráficos- Por otro lado, existen librerías para el entrenamiento de modelos de Machine Learning tales como "sklearn.three.DecisionThreeRegressor" que permite desarrollar el algoritmo. Además, "sklearn.metrics" proporciona funciones para el cálculo de métricas de validación del modelo, como la matriz de confusión, error cuadrático medio entre otras.

```
import os
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from sklearn.preprocessing import MinMaxScaler
from sklearn.tree import DecisionTreeRegressor
from sklearn.metrics import mean_squared_error, r2_score, confusion_matrix
import seaborn as sns
import time
```

Figura 3.3 Librerías de Python

3.5.2 Lectura de datos

Para asegurar una correcta lectura de los datos, es importante definir adecuadamente la ubicación del archivo. En la figura 4.4 se ilustra tanto la ubicación del archivo como el proceso de lectura de la base de datos en formato Excel.

```
# Definir la ubicación del archivo de forma dinámica
base_dir = r'C:\Users\Usuario\Desktop\TESIS\PROGRAMACION'
file_name = 'RADIACIÓN_HORA_HORA.xlsx'
file_path = os.path.join(base_dir, file_name)

# Leer el archivo Excel
data = pd.read_excel(file_path, header=0)
```

Figura 3.4 Ubicación y lectura de datos

3.5.3 Preprocesamiento de datos

Se trabajará con datos de radiación solar, los cuales incluyen información sobre la hora, día, mes y año, abarcando el período de 2017 a 2023. Estos datos deben ser cuidadosamente depurados, aplicando criterios de eliminación de datos nulos ya que si en alguna fila o columna hay un valor faltante se elimina del conjunto de datos, además de la eliminación de datos atípicos, si un valor se encuentra fuera de un rango esperado se lo considera atípico y lo elimina, mejorando el entrenamiento del modelo, este procedimiento asegura que el modelo se base en información confiable. Este proceso garantiza que se disponga de datos seguros y limpios presentando un formato más comprensible para el desarrollo del modelo, en la tabla 4.2 se ilustra una base de datos más confiable y debidamente estructurada.

Tabla 3.2 Base de datos preprocesados

MES	DIA	HORA	RADIACIÓN 2020	RADIACION 2021	RADIACION 2022
1	1	7	0,000	0,479	0,000
1	1	8	0,000	20,071	0,000
1	1	9	0,000	28,516	0,000
1	1	10	21,359	50,264	21,617
1	1	11	155,288	80,697	157,167
1	1	12	180,461	80,606	182,645
1	1	13	182,480	64,564	184,688
1	1	14	155,545	105,354	157,427
1	1	15	147,245	86,522	149,027

3.5.4 Indexación

El proceso de indexación de datos se basa en una estructura que mejora la gestión y el acceso a los registros de manera más eficiente. Al asignar un nuevo índice a cada registro del DataFrame, relacionado con la hora y fecha, como se observa en la Figura 4.5, facilita una reestructuración de los datos. La reorganización mejora la capacidad de manejar y acceder a los datos, permitiendo una mejor visualización de información.

```
# Combinar las columnas MES, DIA y HORA en una única columna de fecha y hora
data['FECHA_HORA'] = pd.to_datetime(data[['MES', 'DIA', 'HORA']].astype(str).agg('-'.join, axis=1), format='%m-%d-%H')

# Derretir las columnas de radiación en una sola columna para unificar los datos de diferentes años
data_melted = data.melt(id_vars=['FECHA_HORA'],
                        value_vars=['RADIACION 2017', 'RADIACION 2018', 'RADIACION 2019',
                                    'RADIACION 2020', 'RADIACIÓN 2021', 'RADIACIÓN 2022', 'RADIACIÓN 2023'],
                        var_name='AÑO',
                        value_name='RADIACION')


# Crear una nueva columna 'FECHA' que combine 'FECHA_HORA' y 'AÑO'
data_melted['FECHA'] = data_melted.apply(lambda row: row['FECHA_HORA'].replace(year=row['AÑO']), axis=1)

# Configurar la columna 'FECHA' como índice
data_melted.set_index('FECHA', drop=True, inplace=True)
```

Figura 3.5 Indexación de datos.

3.5.5 División de grupo de entrenamiento y grupo de prueba.

En la Figura 4.6 se observa una separación de conjuntos de entrenamiento y prueba, en [34], recomienda una división de datos del 80% para entrenamiento y 20% para validación en aprendizaje automático. Para el modelo, se han utilizado datos del período 2017-2022 y una parte del año 2023 como datos de entrenamiento y el resto del año 2023 se emplea como datos de prueba.

```
# Dividir los datos en conjuntos de entrenamiento y prueba (80% y 20%)
split_index = int(len(X) * 0.8)
X_train, X_test = X[:split_index], X[split_index:]
y_train, y_test = y[:split_index], y[split_index:]
```

Figura 3.6 División del grupo de entrenamiento y prueba.

3.5.6 Normalización de datos

La normalización de datos permite ajustar la escala de las características numéricas para que sean compatible entre sí [35]. En la Figura 4.7, se observa que los datos se transforman en rangos de 0 a 1, asegurando que todas contribuyan de manera equitativa al modelo, evitando que aquellas con mayores magnitudes dominen el aprendizaje. Además, facilita la convergencia de algoritmos mejorando el rendimiento del modelo de Machine Learning.

```
# Normalización de datos
scaler = MinMaxScaler(feature_range=(0, 1))
scaled_data = scaler.fit_transform(data_melted[['RADIACION']])

# Definir ventana de tiempo para las secuencias de entrada
ventana_tiempo = 24
```

Figura 3.7 Normalización de datos.

3.5.7 Elección del algoritmo

Para el proyecto de investigación, se ha seleccionado el modelo de árboles de decisión debido a su facilidad de interpretación y la capacidad de realizar predicciones con bajos volúmenes de datos. En la Figura 4.9 se ilustra la librería de árboles de decisión para el entrenamiento del modelo.

3.5.8 Modelo matemático

La función objetivo emplea regresión lineal, además, consisten en utilizar respuestas verdaderas o falsas a determinadas preguntas y poder clasificarlas los datos, en la ecuación 4.1 se observa su formulación matemática, la cual permite minimizar el error del modelo, ajustándose para que sus predicciones se aproximen lo más posible a los valores reales, [36].

Para los árboles de decisión existen restricciones como la entropía y la impureza de Gini expresadas matemáticamente en las ecuaciones 4.2 y 4.3, que permiten medir el grado de impureza en un nodo y de esta manera clasificar los datos, una menor impureza significa que los datos están correctamente clasificados [37]. La ecuación 4.2 se calcula ordenando los datos de manera ascendente y calculando los promedios de los valores adyacentes, posteriormente se calcula su impureza para cada uno de los promedios, organizando los datos según si los valores de la entidad son menores o mayores que el promedio seleccionado y verificando si dicha clasificación agrupa adecuadamente los datos [38], la ecuación 4.3 permite la cantidad de impureza en los nodos.

Función objetivo

$$\text{minimizar} \frac{1}{\text{n}} \sum_{\text{i=1}}^{\text{n}} (\text{yi} - \hat{\text{y}})^2 \tag{4.1}$$

Restricciones

$$\text{Impureza de Gini}(D) = 1 - \sum_{i=1}^{k} p_i^2 \tag{4.2}$$

$$\text{Entropía} = -\sum_{i=1}^{k} \text{pilog}_2(\text{pi}) \tag{4.3}$$

Dónde

pi es la fracción de observaciones que pertenecen a la clase i

k número de clases

n = total de muestras en el conjunto de datos

y valor real de la variable objetivo para cada una de las n muestras

$\hat{y}$= valor predicho por el modelo para cada una de las n muestras

3.6 DECRIPCIÓN DEL MODELO

3.6.1 Variables de entrada y salida

Las variables de entrada suministran los datos requeridos para realizar predicciones sobre la radiación solar, las variables de entrada consisten en secuencias de 24 horas de radiación solar normalizada, con base en estas secuencias el modelo realiza una predicción de la radiación solar para la siguiente hora. Las variables de salida del modelo son las predicciones de la radiación solar para la hora siguiente a cada secuencia de entrada.

- Variables de entrada: radiación (W/m^2).
- Variables de salida: radiación (W/m^2).

3.6.2 Entrenamiento del modelo

El modelo de árbol de decisión se basa en un conjunto de particiones que permiten clasificar los datos [39]. Este modelo está compuesto por tres elementos principales: la raíz, los ramales de decisión y los nodos.

- Raíz: Representa el conjunto completo de datos.
- Ramales de Decisión: Estos ramales representan las divisiones sucesivas de los datos.
- Nodos: clasificación de los datos.

3.6.3 Selección del mejor atributo para dividir

En cada nodo del árbol de decisión, se selecciona el atributo que mejor divide el conjunto de datos. Esta selección se basa en la ecuación, que mide la impureza del nodo. Además, se calcula la entropía utilizando la ecuación 4.3. Un valor bajo de entropía indica que el nodo es más puro, lo cual sugiere una buena elección del atributo para la división. Este proceso implica tomar los datos de radiación y dividirlos progresivamente hasta lograr valores de impureza y entropía bajos.

3.6.4 División del nodo

En función de la selección del mejor atributo para clasificar los datos de radiación, se evalúan todas las posibles divisiones utilizando las ecuaciones 4.2 y 4.3. Este proceso implica analizar los datos de radiación y determinar cómo se pueden dividir en grupos más puros para cada atributo, se calculan la impureza de Gini y la entropía con el objetivo de identificar la división que minimice estos valores, asegurando así que los datos en cada nodo resultante sean los más puros posibles. La división seleccionada es aquella que maximiza la pureza del nodo, lo que implica que los datos de radiación están mejor organizados en categorías distintivas, si un nodo no alcanza el nivel de pureza deseado, el proceso de división continúa aplicando nuevamente las ecuaciones 4.2 y 4.3, hasta que se logra la menor impureza posible en cada nodo del árbol.

3.6.5 Criterio de parada

El árbol de decisión puede seguir extendiéndose a través de múltiples divisiones de los datos de radiación para mejorar su clasificación, este proceso de división se detiene cuando los valores de la impureza de Gini y la entropía alcanzan niveles suficientemente bajos. En términos de radiación solar, esto significa que los datos en cada nodo del árbol están tan bien clasificados que agregar más divisiones no mejoraría significativamente la pureza del nodo, al alcanzar un nivel adecuado de pureza, el árbol de decisión considera que ha encontrado una estructura adecuada para clasificar los datos de radiación, este criterio de detención evita el sobreajuste del modelo.

En la Figura 4.8 se observa el árbol de decisión y las particiones del espacio, las cuáles se realizan de manera repetitiva hasta encontrar los valores deseados que permitan minimizar la función objetivo.

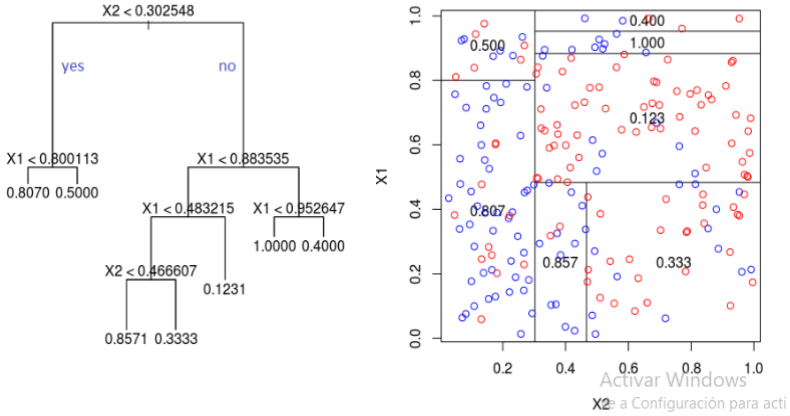


Figura 3.8 Estructura del modelo de árboles de decisión [39].

3.7 MÉTRICAS DE VALIDACÍON PARA MODEOS DE MACHINE LEARNING

3.7.1 Matriz de confusión:

Permite evaluar una precisión del modelo la tabla 4.3 muestra la matriz clasificados como los valores verdaderos positivos (TP), verdaderos negativos (TN), falsos positivos (FP) y falsos negativos (FN) [25]. Los elementos ubicados en la diagonal principal de la matriz representan las predicciones correctas, mientras que los elementos fuera de la diagonal indican errores de clasificación [26].

Tabla 3.3 Matriz de confusión

	Predicho Negativo	Predicho Positivo
Real Negativo	TN	FP
Real Positivo	FN	TP

3.7.2 Precisión

Mide los verdaderos positivos entre el total de las predicciones positivas [25]. En la ecuación 4.1 se observa su expresión matemática. Es importante destacar que una mayor precisión puede indicar que el modelo se encuentre sobre ajustado.

$$\text{Precisión} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{4.1}$$

Dónde:

TP: representa verdadero positivos.

FP: falsos positivos

3.7.3 Sensibilidad

Mide el porcentaje de verdaderos positivos respecto al total de casos positivos reales, permitiendo evaluar la capacidad del algoritmo para identificar correctamente todas las instancias positivas [25]. Se puede observar su formulación matemática en la ecuación 4.2.

$$\text{Sensibilidad} = \frac{TP}{TP + FN} \tag{4.2}$$

Dónde

TP: representa verdadero positivos.

FN: los falsos negativos.

3.7.4 Exactitud

Evalúa la fracción de predicciones acertadas, incluyendo tanto los verdaderos positivos como los verdaderos negativos, respecto al total de casos. evaluados. En la ecuación 4.3 se puede observar su formulación matemática [25].

$$\text{Exactitud} = \frac{TP + TN}{TP + TN + FP + FN} \tag{4.3}$$

Donde:

TN: verdaderos negativos.

TP: verdadero positivos.

FP: falsos positivos.

FN: falsos negativos.

3.7.5 Error cuadrático medio (MSE)

La ecuación 4.4 permite calcular la media de los cuadrados de las diferencias entre los valores predichos y los valores reales [26].

$$\text{MSE} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y})^2 \tag{4.4}$$

3.7.6 Coeficiente de determinación (R^2)

La ecuación 4.5 evalúa la fracción de la variabilidad en la variable dependiente que puede ser explicada por las variables independientes [26].

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y})^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2} \tag{4.5}$$

3.7.7 Error Absoluto Medio (MAE)

La ecuación 4.6 es la diferencia promedio absoluto entre la información predicha e información real [26].

$$MAE = \frac{1}{n}\sum_{j=1}^{n}|y_j - \hat{y}_j| \tag{4.6}$$

3.7.8 Predicción de la radiación

Es importante destacar que, a medida que aumenta el horizonte temporal de la predicción, la precisión disminuye debido a la necesidad de un gran volumen de datos para generar predicciones confiables. Para mejorar la precisión de las predicciones a largo plazo se debe considerar otras técnicas como Redes Neuronales de Memoria a Largo Plazo (LSTM) por su capacidad para manejar grandes cantidades datos, y su mayor tiempo de

entrenamiento que ayuda a retener información y aprender a largo plazo, además, de otros parámetros como humedad, viento, presión atmosférica, etc.

3.7.9 Predicciones de datos

La figura 4.9 se ilustra el proceso de predicción sobre los datos de prueba, seguido del ajuste de la escala de las predicciones y los valores reales para su interpretación. Se utiliza el modelo de árbol de decisión previamente entrenado para realizar predicciones sobre el conjunto de datos de prueba 'x_test'. Estas predicciones se almacenan en 'y_test_pred'. Se recomienda que las predicciones no sean a largo plazo ya que los valores de predicción pueden alejarse más a los valores reales.

```
# Entrenar el modelo
decision_tree_model = DecisionTreeRegressor(random_state=42)
decision_tree_model.fit(X_train, y_train)

# Hacer predicciones
y_test_pred = decision_tree_model.predict(X_test)
y_test_pred_rescaled = scaler.inverse_transform(y_test_pred.reshape(-1, 1))
y_test_rescaled = scaler.inverse_transform(y_test.reshape(-1, 1))
```

Figura 3.9 Predicciones de datos de prueba

3.8 DIMENSIONAMINETO DEL SISTEMA FOTOVOLTAICO

Para el dimensionamiento del sistema fotovoltaico, se ha considerado sistemas conectados a la red y sistemas aislados para usuarios urbanos y rurales. En la Figura 4.10 se observan los estratos de consumo de la Empresa Eléctrica Quito (EEQ) los cuales se clasifican en cinco categorías A1, A, B, C, D y E [40].

Categoría de Estrato de Consumo (Nota 1)	Escalas (kWh/mes/cliente)
E	0 – 100
D	101 – 150
C	151 – 250
B	251 – 350
A	351 – 500
A1	501 – 900

Nota:

1. En los estratos A, B, C, D y E, los rangos están definidos considerando el valor de consumo que registran los equipos eléctricos para uso general y calentamiento de agua; mientras que para el estrato A1 el rango está definido considerando el valor de consumo que registran los equipos eléctricos para uso general, cocción y calentamiento de agua.

Figura 3.10 Categorización de estrato de consumo

El anexo 1 permite identificar a qué categoría pertenece el usuario al grupo rural o urbano, se observan en la Figura 5.9 la categorización, se consideran urbanos a los usuarios dentro de la zona visualizada en color rojo, los usuarios fuera de esa zona se los considera rurales. Entonces, los usuarios de la categoría A1-A – B y C se los considera urbanos y la categoría D y E rurales.

3.8.1 Variables

En la tabla 4.4 se muestran las variables de requeridas por el software. Cada una de estas variables permite calcular los componentes necesarios para el sistema fotovoltaico.

Tabla 3.4 Variables para el dimensionamiento del sistema fotovoltaico.

Tipo de sistema (ON GRID/OFF GRID)
Categoría de consumo del usuario (E, D, C, B, A, A1)
Tiempo para la predicción
Consumo mensual
Potencia del panel
Factor de seguridad

3.8.2 Formulación matemática

Para el dimensionamiento fotovoltaico, se utilizaron diferentes fórmulas que varían en función del tipo de sistema fotovoltaico deseado. Los sistemas conectados a la red (on-grid) se puede dimensionar su operación mediante las siguientes ecuaciones en función del tipo de consumo que se requiera.

3.9 SISTEMAS FOTOVOLTAICOS CONECTADOS A LA RED

3.9.1 Cálculo de la energía diaria promedio

La ecuación 4.7 permite calcular la energía diaria promedio, siendo la división entre el consumo mensual y los días del mes.

$$\text{Ediaria} = \frac{\text{Cmensual}}{30} \tag{4.7}$$

Dónde:

Cmesual_ consumo mensual en kWh

3.9.2 Cálculo de la potencia del sistema

La ecuación 4.8 permite calcular la potencia del sistema siendo la división entre la energía diaria promedio y las horas picos que se pueden encontrar mediate la página de Nasa Power.

$$Ppv = \frac{Ediaria}{Hpsh} \tag{4.8}$$

Donde

Hsph: horas pico del sol

3.9.3 Número de panel

La ecuación 4.9 permite calcular la cantidad de paneles necesarios el cual es necesario la potencia del sistema calculado mediante la ecuación 4.12.

$$Npáneles = \frac{Ppv}{Ppanel} \tag{4.9}$$

Dónde:

Ppanel: potencia del panel.

3.9.4 Dimensionamiento del inversor

La ecuación 4.10 permite calcular el inversor en el cual es importante el factor de seguridad generalmente es un valor de 1.2 [41].

$$Pinversor = Ppv * Fs \tag{4.10}$$

Dónde_

Fs = factor de seguridad

3.10 SISTEMAS AISLADOS (OFF -GRID)

Para los sistemas fotovoltaicos aislados (off-grid) al no estar conectados a la red, dependen de baterías para almacenar la energía que se generar y poder suministrarla. Se utilizan las ecuaciones 4,7 y 4.8 para calcularlas, adicionalmente se utilizan las ecuaciones 4.11 a

3.10.1 Cálculo de la energía almacenada en las baterías

La ecuación 4.11 permite calcular la energía almacenada en las baterías en la cual se considera también los días de autonomía para poder suministrar la demanda de energía eléctrica.

$$\text{Ebatería} = \text{Ediaria} * \text{D.autonomía} \tag{4.11}$$

Dónde:

D.autonomía: días de autonomía.

3.10.2 Capacidad de las baterías

La ecuación 4.12 permite dimensionar la capacidad de las baterías para almacenar la energía eléctrica.

$$C.batería = \frac{E.batería}{Vbatería * D_{DOD} * \eta_{batería}} \tag{4.12}$$

Dónde:

V. batería: voltaje de batería.

D_{DOD}: profundidad de descarga.

$\eta_{batería}$: eficiencia de la batería

3.10.3 Potencia del sistema fotovoltaico

La ecuación 4.13 permite calcular la potencia del sistema que se considera la eficiencia de la batería y horas pico de radiación solar.

$$Ppv = \frac{Ediaria}{Hpsh * \eta batería} \quad (4.13)$$

3.10.4 Dimensionamiento del controlador

La ecuación 4.14 permite calcular la corriente del inversor para la cual se considera el voltaje de la batería y la potencia del sistema fotovoltaico que se lo calcula en la ecuación 5.7

$$I.inversor = \frac{Ppv}{Vbatería} \quad (4.14$$

4. ANÁLISIS Y DISCUSIÓN DE LOS RESULTADOS

Los árboles de decisión es una técnica de Machine Learning que me permite minimizar el error de predicción con el objetivo de ajustarse a los valores observados, mediante restricciones de impureza y entropía, la cual se realiza mediante divisiones del árbol de decisión clasificando los datos obteniendo la menor impureza posible, asegurando la precisión del modelo.

A continuación, se presentan los resultados del proyecto de investigación, que incluyen la predicción, validación y los resultados obtenidos del sistema fotovoltaico. En la Figura 5.1 se observan como primera instancia los valores reales de la radiación solar del sector de Tabacundo.

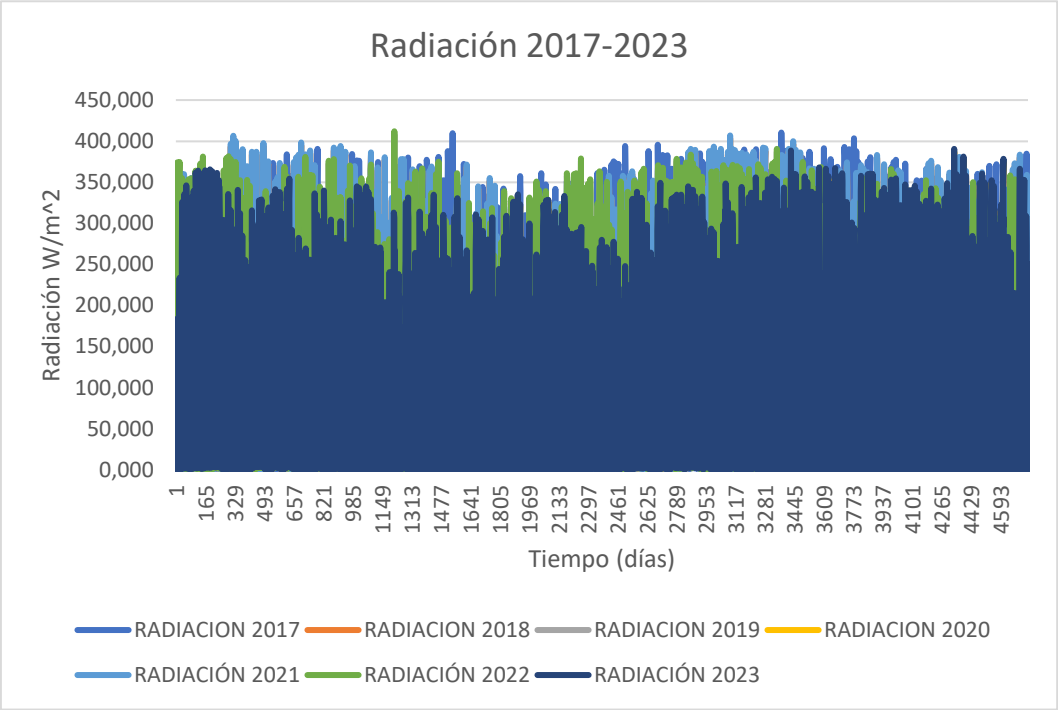


Figura 4.1 Datos reales de la radiación solar de 2017 a 2023.

En la Figura 5.2 se observa la gráfica de los valores que se utilizaron para el entrenamiento, que se han utilizado datos del período 2017-2021 y una parte del año 2022 como datos de entrenamiento, los demás datos se los toma como validación.

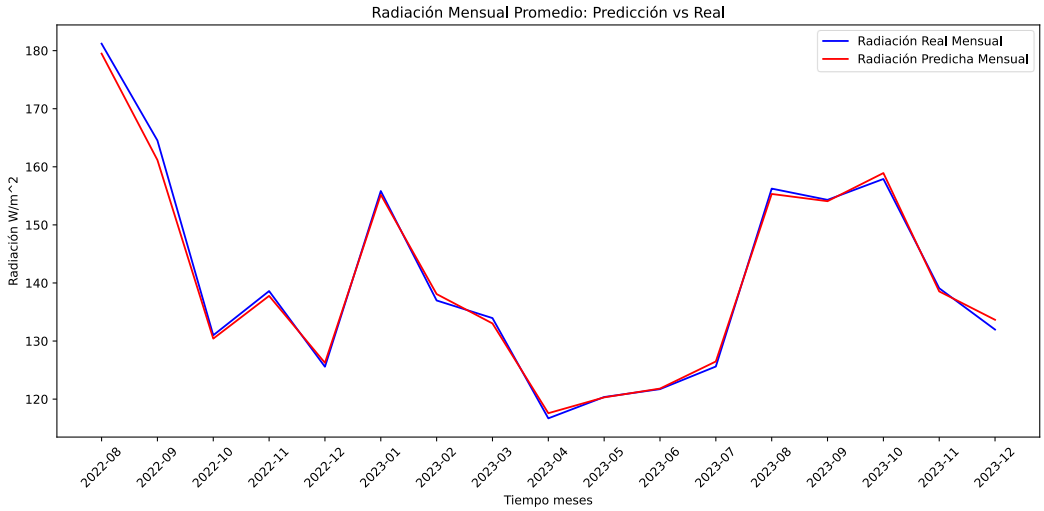


Figura 4.2 Comportamiento de entrenamiento y prueba 2022-2023

4.1 MÉTRICAS DE VALIDACIÓN DEL MODELO DE ENTRENAMIENTO Y PRUEBA.

4.1.1 Matriz de confusión

Al calcular la matriz de confusión, se obtiene la Figura 5.10 que ilustra 6358 verdaderos positivos (TP), 160 falsos positivos (FP), 112 verdaderos negativos (TN) y 9 falsos negativos. Estos valores, indican que dado el número de verdaderos positivos es mayor que falsos negativos, el modelo demuestra una alta capacidad para identificar correctamente los casos positivos, lo que sugiere un buen desempeño en términos de sensibilidad., afirmando que el modelo genera confianza para realizar predicciones.

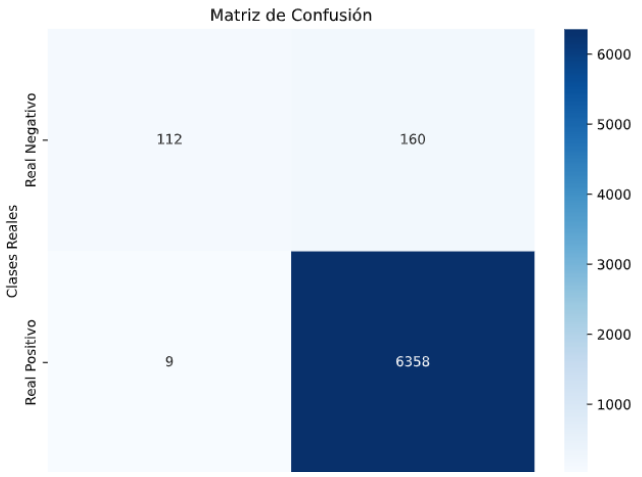


Figura 4.3 Matriz de confusión par la validación del modelo de ML.

Además, se consideran otras métricas de validación del modelo, como la precisión, exactitud y sensibilidad. En la Tabla 5.4 se presentan los valores de estas métricas, en el caso de la sensibilidad y la exactitud cuanto más se acerquen a 1, mejor será el rendimiento del modelo. Otra métrica importante el coeficiente de determinación R^2 con un valor de 0.72 según [26], este valor índico un ajuste moderado, ya que un coeficiente R^2 muy cercano a 1 se consideran un indicativo de un modelo de predicción confiable.

Al analizar las métricas de validación en la Tabla 5.4 y utilizando los criterios en [26], [25], se puede concluir que el modelo de predicción es altamente confiable con una precisión de 0,975, una sensibilidad de 0.998 indicando que es capaz de identificar correctamente los verdaderos positivos, su exactitud de 0,974 lo cual indica que es un modelo eficiente, además de un coeficiente determinación de 0.93 lo que significa en [26]que al ser muy un valor muy cercano a 1 es completamente fiable.

Tabla 4.1 Resultados de las métricas de validación

Precisión	0.975
Sensibilidad	0.998
Exactitud	0.974
Error Cuadrático Medio (MSE)	761.602
Error Cuadrático Medio Normalizado (NME)	5.434
Coeficiente de determinación (R^2)	0.935

Adicional, se ha realizado una comparación de las predicciones para evaluar el error absoluto medio y determinar la precisión del modelo.

CASO 1

Para el primer caso se predice los datos de radiación para un periodo de un mes, en la Figura 5.3 se ilustra su comportamiento, para este caso se ha entrenado el modelo con datos históricos desde el 2017 hasta el 2022 incluyendo una parte del 2023. Al considerar la métrica de error absoluto medio MAE para evaluar el rendimiento del modelo se obtiene un valor de 8.694.

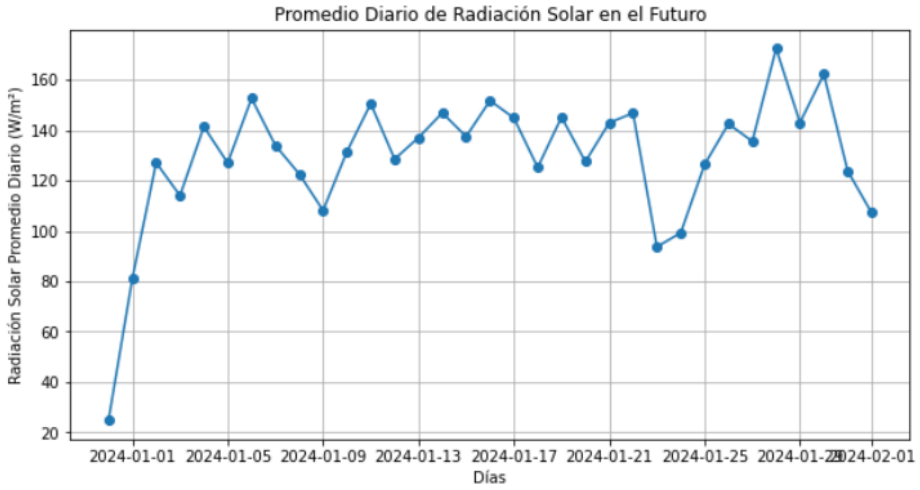


Figura 4.4 Comportamiento de la predicción de la radiación en 1 mes.

CASO 2

Para el segundo caso se predice los datos de radiación para un periodo de seis meses, en la Figura 5.5 se ilustra su comportamiento, el entrenamiento es el mismo que el caso 1. Se obtiene como resultado de MAE de 139.611.

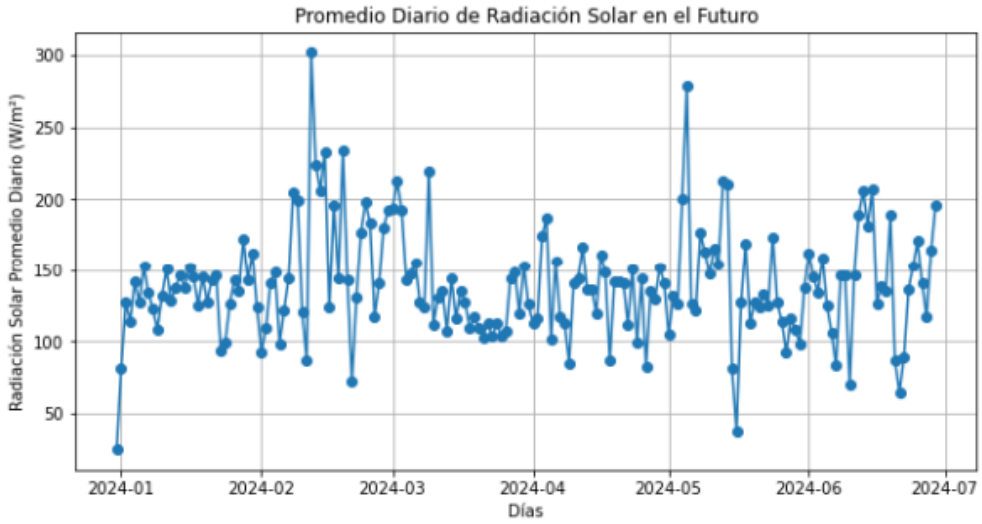


Figura 4.5 Comportamiento de la predicción de la radiación para 6 meses.

En la tabla 5.1 se ilustran los resultados de los casos de predicción, se observa que el error absoluto medio para el caso 1 en un mes es de 8.694 en contraste para el caso 2 para un periodo de 6 meses el error absoluto aumenta gradualmente a 139.611, esta diferencia indica que a medida que se extiende el periodo de predicción estas son menos precisas.

Tabla 4.2 Comparación de MAE para los casos de estudio

Resultados de los casos	Error absoluto medio (MAE)
Caso 1 (1 mes)	8.694.
Caso 2 (6 meses)	139.611

Dimensionamiento fotovoltaico

Caso de estudio

Para el caso de estudio se ha dimensionado un sistema conectado a la red y no conectados a la red. En la tabla 5.2 y 5.3 se pueden observar los resultados de los sistemas fotovoltaicos.

4.1.2 Conectado a la red

Para ello se ha considerado un consumo de 160 kWh, potencia del panel de 100 W un voltaje de la batería de 12 V.

Tabla 4.3 Dimensionamiento de sistema fotovoltaico conectado a la red.

Energía diaria promedio	5,33 kWh/día
Potencia del sistema	1,46 kW
Número de paneles	15.
Dimensionamiento del inversor	275 W

4.1.3 Sistemas aislados

Tabla 4.4 Dimensionamiento de sistema fotovoltaico aislado.

Energía diaria promedio	5,33 kWh/día
Energía almacenada en las baterías	1,200 Wh
Número de paneles necesario	17

Los sistemas fotovoltaicos están dimensionados para satisfacer la misma demanda de energía. Se observa que el sistema aislado requiere un mayor número de paneles ya que al no estar conectado a la red, debe generar y almacenar suficiente energía para satisfacer

su demanda, por otra parte, los sistemas conectados a la red pueden generar y suministrar energía durante el día, sin embargo, durante horas de la noche al no existir radiación el sistema se vuelve dependiente de la red eléctrica para proporciona energía. Se puede concluir que los sistemas on-grid son ideales para zonas urbanas ya que tienen acceso a la red volviéndose confiables y seguros, por otra parte, los sistemas aislados son mayormente utilizados en zonas rurales en áreas de difícil acceso a la red eléctrica.

5. CONCLUSIONES Y RECOMENDACIONES

- El algoritmo de árboles de decisión es un modelo confiable para predicciones de corto plazo, presentando una precisión de 0.975, sensibilidad de 0.998 y una exactitud de 0.974, además de un error cuadrático de 5.43 que es un rango aceptable, y el coeficiente de determinación de 0.935, estos indicadores sugieren que el modelo es confiable en sus predicciones, sin embargo presenta limitaciones para predicciones de largo plazo ya que sus valores tienden a ser menos precisos, es recomendable considerar otros modelos como LSTM que tiene una mejor capacidad para capturar patrones más completos en los datos, además se debe considerar más variables como la humedad, precisión, temperatura para una mayor precisión en la predicciones a largo plazo.

- Al realizar una comparación del modelo de predicción se pudo determinar que el algoritmo de decisión es eficiente a corto plazo obteniendo un resultado de 8.694 de error absoluto para 1 mes, mientras que para 6 meses se obtuvo un resultado de 139.611 concluyendo que a medida que se extiendan las predicciones el modelo no es confiable para largo plazo.
- En el análisis de dimensionamiento fotovoltaico realizado en Tabacundo, en la zona de Pedro Moncayo, para junio del 2024, se consideró el mismo consumo energético de 160 kWh y se identificó que los sistemas conectados a la red requieren 15 paneles en comparación de un sistema aislado que necesita 17 paneles fotovoltaicos, esto se debe a que los sistemas aislados al necesitar un mayor suministro de electricidad durante el día como la noche, además de necesitar más componentes como baterías para su funcionamiento, lo que los hace más costoso. Por otro lado, los sistemas conectados a la red son más accesibles y

recomendables, ya que pueden suministrar energía eléctrica a través de la red pública cuando el sistema lo requiera.