



UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE INGENIERÍA ELÉCTRICA

**MEJORA Y DESARROLLO DE GEMELOS DIGITALES DE PLANTAS
FOTOVOLTAICAS DE ESCALA COMERCIAL E INDUSTRIAL**

MEMORIA PARA OPTAR AL TÍTULO DE INGENIERO CIVIL ELÉCTRICO

NICOLÁS ANDRÉS FERNÁNDEZ VALDÉS

PROFESOR GUÍA:
NICOLÁS ENCINA FLORES

MIEMBROS DE LA COMISIÓN:
MARTÍN CÁDIZ POBLETE
ANDRÉS CABA RUTTE

SANTIAGO DE CHILE
2024

RESUMEN DE LA MEMORIA PARA OPTAR
AL TÍTULO DE INGENIERO CIVIL ELÉCTRICO
POR: NICOLÁS ANDRÉS FERNÁNDEZ VALDÉS
FECHA: 2024
PROF. GUÍA: NICOLÁS ENCINA FLORES

MEJORA Y DESARROLLO DE GEMELOS DIGITALES DE PLANTAS FOTOVOLTAICAS DE ESCALA COMERCIAL E INDUSTRIAL

Este proyecto tiene como objetivo mejorar el rendimiento y la gestión de plantas fotovoltaicas mediante el desarrollo de gemelos digitales, o Digital Twins, que son representaciones virtuales de las plantas reales. Estos modelos permiten simular y predecir el comportamiento de las plantas bajo diversas condiciones, facilitando la toma de decisiones y optimizando la operación de las plantas solares que administra *Solarity*, una empresa con un amplio portafolio de plantas fotovoltaicas.

El enfoque principal del proyecto fue crear y validar estos gemelos digitales utilizando dos metodologías. Por un lado, se empleó la herramienta de simulación *PVLIB*, diseñada para estimar el rendimiento de sistemas fotovoltaicos con base en las configuraciones específicas de cada planta y las condiciones climáticas locales. Por otro lado, se utilizaron redes Transformer, un tipo avanzado de inteligencia artificial, para analizar grandes volúmenes de datos y capturar relaciones complejas, lo que permitió desarrollar un modelo capaz de predecir la energía generada por el sistema a partir de datos meteorológicos y características de las plantas.

El objetivo es que estos modelos digitales puedan predecir con alta precisión el comportamiento de las plantas, comparando sus estimaciones con las obtenidas mediante *HelioScope*, un software reconocido en la industria por su capacidad de simular el rendimiento de plantas fotovoltaicas. Utilizando *HelioScope* como referencia, el proyecto se enfocó en reducir las discrepancias entre las predicciones del gemelo digital y los resultados reales, mejorando así la precisión del modelo basado en *PVLIB*.

El análisis se centró en plantas fotovoltaicas comerciales e industriales que comenzaron a operar hasta antes del año 2024. Para las simulaciones y el entrenamiento de los modelos, se utilizaron datos meteorológicos típicos del año (TMY). Aunque inicialmente el proyecto se enfocó en predecir la energía generada por el sistema, los resultados muestran que estos modelos podrían expandirse en el futuro para predecir otros parámetros eléctricos, proporcionando así una visión más completa del rendimiento de las plantas.

Entre los resultados más relevantes, se logró una precisión del 97.22 % en la predicción de la energía generada comparada con los resultados de *HelioScope*, lo que subraya el potencial de estos enfoques para mejorar la operación y gestión de plantas fotovoltaicas. Estos avances sientan las bases para futuras mejoras en los modelos, permitiendo no solo una gestión más eficiente, sino también una optimización continua del rendimiento de las plantas solares, lo que contribuirá a una operación más eficaz y sostenible a largo plazo.

Agradecimientos

Quiero expresar mi más profundo agradecimiento a todas las personas que me acompañaron y apoyaron durante este proyecto y a lo largo de mi formación académica.

En primer lugar, a mis padres, quienes han sido mi pilar fundamental. Gracias por su amor incondicional, su paciencia, y por haberme brindado siempre el apoyo necesario para alcanzar mis metas. Su confianza en mí y en mis capacidades ha sido una fuente constante de motivación.

A mis amigos, por estar siempre ahí, brindándome su compañía, su ánimo y risas en los momentos buenos y malos. Gracias por su amistad y por compartir conmigo esta etapa tan importante de mi vida.

Quiero agradecer de manera especial a los profesores de mi comisión, quienes con su conocimiento, experiencia y compromiso me han guiado en la realización de esta tesis. Gracias por su tiempo, sus valiosos comentarios y por ayudarme a llevar este trabajo a buen término. Su orientación ha sido clave en mi desarrollo académico.

Finalmente, quiero agradecer a la empresa *Solarity* por la oportunidad de realizar este proyecto en un entorno profesional desafiante y enriquecedor. Su apoyo y confianza me permitieron aplicar y ampliar mis conocimientos, contribuyendo a mi crecimiento como profesional en el área de la energía solar.

Tabla de Contenido

1. Introducción	1
1.1. Motivación y Descripción del Problema	1
1.2. Objetivos	2
1.2.1. Objetivo General	2
1.2.2. Objetivos Específicos	2
1.3. Alcances	2
2. Marco Teórico	4
2.1. Principios de Energía Solar	4
2.1.1. Geometría Solar	4
2.1.2. Interacción de la Radiación Solar con la Atmósfera	5
2.1.3. Tipos de Irradiancia	7
2.1.4. Recurso Solar	8
2.2. Sistemas Fotovoltaicos	8
2.2.1. Componentes de un Sistema Fotovoltaico	10
2.2.1.1. Módulos Fotovoltaicos	10
2.2.1.1.1. Conexión en Strings	11
2.2.1.2. Inversor	12
2.2.1.3. Estructuras de Montaje	12
2.2.1.4. Sistemas de Almacenamiento	13
2.2.1.5. Cableado y Componentes Eléctricos	14
2.2.2. Tipos de Sistemas Fotovoltaicos	14
2.2.2.1. Sistemas Fotovoltaicos Directos (PV Direct)	14
2.2.2.2. Sistemas Conectados a la Red (On-Grid)	14
2.2.2.3. Sistemas Aislados (Off-Grid)	15
2.2.2.4. Sistemas Conectados a la Red con Respaldo de Baterías	15
2.2.3. Rendimiento y Eficiencia de los Sistemas Fotovoltaicos	16
2.2.4. Modelos de Simulación de Sistemas FV	16
2.2.5. Pasos de Simulación de Sistemas Fotovoltaicos	17
2.3. Digital Twins	19
2.3.1. Digital Twins en Energía	19
2.3.2. Digital Twins y ML en Sistemas Fotovoltaicos	20
2.4. <i>PVLIB</i>	21
2.4.1. Componentes Clave de <i>PVLIB</i>	22
2.4.1.1. Modelos de Irradiancia	22
2.4.1.2. Modelos de Pérdidas	22
2.4.1.3. Modelos de Temperatura	23

2.4.1.4.	Conversión de DC a AC	24
2.4.2.	Aplicaciones de <i>PVLIB</i>	24
2.5.	Machine Learning	24
2.5.1.	Tipos de Machine Learning	25
2.5.2.	Machine Learning en Energía y Fotovoltaica	25
2.5.3.	Regresión Lineal y Técnicas de Machine Learning	25
2.6.	Métricas de Evaluación de Rendimiento y Precisión de Modelos	27
2.6.1.	Error Cuadrático Medio	28
2.6.2.	Raíz del Error Cuadrático Medio	28
2.6.3.	Error Absoluto Medio	28
2.6.4.	Coefficiente de Determinación	28
3.	Estado del Arte	30
3.1.	Herramientas de Simulación de Sistemas Fotovoltaicos	30
3.1.1.	<i>HelioScope</i>	30
3.1.1.1.	Principales Usos	30
3.1.1.2.	Características Principales	31
3.1.2.	<i>Solcast Rooftop PV Model</i>	31
3.1.2.1.	Principales Usos	31
3.1.2.2.	Características Principales	32
3.1.3.	<i>PVSyst</i>	32
3.1.3.1.	Principales Usos	32
3.1.3.2.	Características Principales	33
3.1.4.	Diferencias y Similitudes	33
3.2.	Machine Learning	34
3.2.1.	Componentes Principales del Modelo Transformer	34
3.2.1.1.	Mecanismo de Atención	34
3.2.1.2.	Capas de Codificador y Decodificador	34
3.2.1.3.	Positional Encoding	34
3.2.2.	Avances y Aplicaciones	34
3.2.2.1.	Modelos de Lenguaje Avanzados	34
3.2.2.2.	Transferencia de Aprendizaje	35
3.2.2.3.	Visión por Computadora y Otras Áreas	35
3.2.3.	Desafíos y Futuro	35
4.	Metodología	36
4.1.	Obtención de Datos y Configuración Inicial	36
4.1.1.	Datos Meteorológicos y del Sistema Fotovoltaico	37
4.1.2.	Configuración de las Plantas Fotovoltaicas	37
4.2.	Simulación Inicial en <i>PVLIB</i>	37
4.3.	Ajustes Iterativos del Modelo en <i>PVLIB</i>	38
4.4.	Entrenamiento del Modelo Transformer	38
4.4.1.	Preparación del Dataset	38
4.4.2.	Entrenamiento del Modelo	38
4.5.	Evaluación y Comparación de Resultados	38
5.	Desarrollo	39
5.1.	Modelo basado en <i>PVLIB</i>	39

5.2. Modelo Transformer	42
6. Resultados	45
6.1. PVLIB	45
6.1.1. Resultados de la versión final del modelo <i>PVLIB</i>	47
6.2. Modelos Transformers	49
6.2.1. Modelos “Pequeño”, “Mediano” y “Grande”	50
6.2.2. Modelos “Híbrido” y “Híbrido considerando horas de Sol”	55
7. Conclusiones	61
7.1. Conclusiones	61
7.2. Trabajo Futuro	62
Bibliografía	64
Anexos	68
A. Pseudo Códigos de PVLIB	68
A.1. Clase PlantModel	68
A.2. Clase InverterModel	69
A.3. Clase WeatherModel	69
A.4. Clase SimulationRunner	70

Índice de Tablas

5.1.	Hiper-parámetros usados en los primeros 3 modelos Transformers entrenados. . .	43
5.2.	Hiper-parámetros usados en el modelo Transformer “Híbrido”.	44
6.1.	Porcentaje de error promedio calculado por el modelo <i>PVLIB</i> original entregado por la empresa.	45
6.2.	Porcentaje de error promedio calculado por <i>PVLIB</i> respecto a <i>HelioScope</i> . . .	47
6.3.	Porcentaje de error promedio calculado por <i>PVLIB</i> tras la optimización de parámetros.	48
6.4.	Métricas <i>RMSE</i> y <i>MAE</i> para el modelo <i>PVLIB</i> en sus distintas versiones. . .	49
6.5.	Métricas obtenidas al evaluar los modelos “Pequeño”, “Mediano” y “Grande” en el conjunto de prueba.	52
6.6.	Métricas obtenidas al evaluar los modelos “Híbrido” y “Híbrido” considerando horas de Sol en el conjunto de prueba.	56
6.7.	Métricas desnormalizadas obtenidas al evaluar los modelos “Híbrido” y “Híbrido” considerando horas de Sol en el conjunto de prueba.	58

Índice de Ilustraciones

2.1.	Ángulos Zenital, Azimutal y Elevación Solar [4].	5
2.2.	Porcentaje de luz entrante a la superficie de la Tierra [5].	6
2.3.	Tipos de Radiación [9].	7
2.4.	Esquema de un piranómetro (a) y un pirheliómetro (b) [11].	8
2.5.	Funcionamiento del efecto fotoeléctrico dentro de una celda FV [12].	9
2.6.	Diagrama de un Sistema Fotovoltaico en una casa [13].	10
2.7.	Tipos de módulos fotovoltaicos [14].	11
2.8.	Conexión de strings [15].	11
2.9.	Ejemplo de inversores en un sistema fotovoltaico [16].	12
2.10.	Montajes fijos. A la izquierda montaje en suelo y a la derecha montaje sobre techo [17].	13
2.11.	Montaje con sistema tracker [18].	13
2.12.	Sistema Fotovoltaico On-Grid [19].	14
2.13.	Sistema Fotovoltaico Off-Grid [19].	15
2.14.	Sistema Fotovoltaico On-Grid con respaldo de baterías [19].	16
2.15.	Etapas clave en el modelado del desempeño de sistemas fotovoltaicos [21].	18
2.16.	Diagrama de monitoreo y control de una planta FV usando digital twin [28].	20
2.17.	Sistema de monitoreo usando un modelo de ML entrenado para el cálculo de potencia FV [31].	21
2.18.	Técnicas de Machine Learning.	25
2.19.	Arquitectura del modelo Transformer presentada en el paper “Attention is All You Need” [42].	27
4.1.	Metodología seguida para el desarrollo del proyecto.	36
5.1.	Diagrama de funcionamiento de simulaciones usando <i>PVLIB</i>	40
6.1.	Comparación de la generación anual calculada por <i>HelioScope</i> y <i>PVLIB</i> original.	46
6.2.	Histograma de distribución de error en el cálculo de generación anual del modelo <i>PVLIB</i> original respecto a <i>HelioScope</i>	46
6.3.	Comparación de la generación anual calculada por <i>HelioScope</i> y <i>PVLIB</i>	47
6.4.	Histograma de distribución de error en el cálculo de generación anual del modelo <i>PVLIB</i> respecto a <i>HelioScope</i>	48
6.5.	Comparación de la generación anual calculada por <i>HelioScope</i> y <i>PVLIB</i> tras la optimización de parámetros.	49
6.6.	Curvas de pérdidas de entrenamiento y validación de modelo Transformer “Pe- queño”.	50
6.7.	Curvas de pérdidas de entrenamiento y validación de modelo Transformer “Me- diano”.	51
6.8.	Curvas de pérdidas de entrenamiento y validación de modelo Transformer “Gran- de”.	51

6.9.	Comparación de resultados reales y predichos por el modelo “Pequeño”.	53
6.10.	Comparación de resultados reales y predichos por el modelo “Mediano”.	53
6.11.	Comparación de resultados reales y predichos por el modelo “Grande”.	54
6.12.	Curvas de pérdidas de entrenamiento y validación de modelo Transformer “Híbrido”.	55
6.13.	Curvas de pérdidas de entrenamiento y validación de modelo Transformer “Híbrido” considerando horas de Sol.	56
6.14.	Comparación de resultados reales y predichos por el modelo “Híbrido”.	57
6.15.	Comparación de resultados reales y predichos por el modelo “Híbrido” considerando horas de Sol.	57
6.16.	Comparación de resultados desnormalizados reales y predichos por el modelo “Híbrido”.	59
6.17.	Comparación de resultados desnormalizados reales y predichos por el modelo “Híbrido” considerando horas de Sol.	59

Capítulo 1

Introducción

1.1. Motivación y Descripción del Problema

El crecimiento de la energía solar como fuente de energía renovable ha impulsado la instalación masiva de plantas fotovoltaicas a nivel mundial. Sin embargo, la gestión eficiente y la optimización del rendimiento de estas plantas plantean desafíos significativos. Para empresas como *Solarity*, que administran un amplio portafolio de plantas fotovoltaicas, es esencial contar con herramientas que permitan monitorear, analizar y optimizar el rendimiento de cada instalación de manera precisa y eficiente. En este contexto, la creación de gemelos digitales de plantas fotovoltaicas se presenta como una solución innovadora. Estos gemelos digitales permiten simular y predecir el comportamiento de las plantas bajo diversas condiciones operativas y meteorológicas, lo que facilita la toma de decisiones y mejora la eficiencia operativa.

Uno de los principales desafíos que enfrenta *Solarity* en la gestión de su portafolio es la precisión de los modelos de simulación utilizados para estimar la generación de energía y otros parámetros eléctricos. Herramientas como *HelioScope*, ampliamente utilizadas en la industria para modelar plantas fotovoltaicas, ofrecen un enfoque detallado, pero presentan limitaciones cuando se trata de utilizar los modelos automáticamente con datos actuales de radiación, dado que sólo se puede interactuar a través de su interfaz gráfica.

Además, los datos meteorológicos utilizados en las simulaciones provienen de fuentes satelitales. Aunque estos datos son útiles, no siempre reflejan con precisión las condiciones específicas de cada planta a nivel de suelo. La precisión de los valores de irradiancia solar, las condiciones climáticas locales y las características particulares de cada instalación son factores que pueden introducir diferencias significativas en los resultados de las simulaciones. Este problema es especialmente relevante en plantas ubicadas en el sur de Chile, donde la nubosidad puede variar considerablemente y afectar la generación de energía de manera significativa.

Otro desafío crítico es la integración de datos en tiempo real con los modelos de simulación. La calidad y precisión de los datos medidos en tiempo real en las plantas pueden verse comprometidas por fallos en los equipos de medición, variaciones en la eficiencia de los inversores y otros factores operativos. Estas inconsistencias dificultan la comparación precisa entre los valores simulados y los valores reales, afectando la confiabilidad de las simulaciones.

La motivación de este proyecto es abordar estos desafíos desarrollando modelos digitales

que sean precisos y adaptables, utilizando metodologías avanzadas como *PVLIB*, una librería especializada en la simulación de sistemas fotovoltaicos, y redes Transformer, una técnica de *Machine Learning* capaz de analizar grandes volúmenes de datos y capturar relaciones complejas entre variables. Inicialmente, el proyecto se enfocará en la estimación de la potencia AC para validar la capacidad de estos modelos de proporcionar resultados precisos. A largo plazo, se espera expandir estos modelos para incluir otros parámetros eléctricos, lo que permitirá una evaluación integral del rendimiento de las plantas y contribuirá a mejorar la eficiencia operativa y la toma de decisiones estratégicas en *Solarity*.

1.2. Objetivos

1.2.1. Objetivo General

Desarrollar y validar gemelos digitales de plantas fotovoltaicas utilizando metodologías basadas en *PVLIB* y redes Transformer para estimar con precisión diversos parámetros eléctricos comparables a los obtenidos con *HelioScope*.

1.2.2. Objetivos Específicos

1. Implementar y ajustar el modelo basado en *PVLIB* para cada planta fotovoltaica del portafolio de *Solarity*, asegurando que las simulaciones reflejen con precisión las condiciones reales mediante la integración de datos específicos de configuración y condiciones meteorológicas.
2. Desarrollar y entrenar un modelo Transformer que utilice datos meteorológicos y características de las plantas para estimar la potencia de salida del sistema.
3. Realizar comparaciones y ajustes iterativos entre los resultados obtenidos de los modelos *PVLIB* y Transformer con los datos de referencia de *HelioScope*, minimizando las diferencias y mejorando la precisión de los modelos mediante técnicas de validación y optimización de parámetros.

1.3. Alcances

Para el desarrollo de los modelos que generen los gemelos digitales en este proyecto, se utilizaron dos enfoques metodológicos: uno basado en la librería *PVLIB* de *Python* y otro utilizando un modelo Transformer. La característica principal seleccionada para realizar las comparaciones de rendimiento y precisión entre los diferentes modelos es la potencia AC. Esta elección se debe a la importancia de la potencia AC como indicador del rendimiento general de las plantas fotovoltaicas y su relevancia para la toma de decisiones operativas y estratégicas.

Los análisis y desarrollos de los modelos se centraron en el portafolio de plantas fotovoltaicas de *Solarity* que entraron en operación hasta antes del año 2024. La razón detrás de esta selección es que la información de estas plantas se encuentra más consolidada y disponible en la base de datos de la empresa.

Además, el proyecto se enfocó en utilizar datos meteorológicos TMY (Typical Meteorological Year) obtenidos desde el *Explorador Solar* [1], que incluye información sobre irradiancia solar, temperatura ambiente, velocidad del viento y otros factores críticos para la generación de energía. Estos datos fueron utilizados tanto para las simulaciones en *PVLIB* como para el entrenamiento del modelo Transformer.

Finalmente, aunque el enfoque inicial del modelo Transformer fue estimar la potencia AC, existe un gran potencial para expandir su uso a la predicción de otros parámetros eléctricos críticos como voltaje, corriente y temperatura de los paneles. Esta expansión permitirá una evaluación más integral del rendimiento de las plantas y contribuirá a la optimización y gestión del portafolio fotovoltaico de la empresa.

Capítulo 2

Marco Teórico

2.1. Principios de Energía Solar

La energía solar se origina en el núcleo del Sol, donde ocurre el proceso de fusión nuclear. Durante este proceso, los átomos de hidrógeno se combinan para formar helio, liberando enormes cantidades de energía en forma de radiación electromagnética [2]. Esta radiación se emite en todas las direcciones, y una pequeña fracción de ella viaja hacia la Tierra. La radiación solar tarda aproximadamente 8 minutos y 20 segundos en alcanzar nuestro planeta, transportada en línea recta a través del espacio en forma de luz visible, radiación ultravioleta e infrarroja [3].

Al llegar a la Tierra, la radiación se distribuye de diversas maneras, afectada por factores como la atmósfera, la posición relativa del Sol y las condiciones meteorológicas locales. Comprender cómo interactúa esta radiación con la Tierra y los factores que influyen en su captación es esencial para optimizar el rendimiento de los sistemas fotovoltaicos.

2.1.1. Geometría Solar

La geometría solar se refiere a los movimientos relativos del Sol y la Tierra que determinan la posición del Sol en el cielo en diferentes momentos del día y del año. Estos movimientos influyen directamente en la cantidad de radiación solar que un sistema fotovoltaico puede captar [3]. Los principales conceptos que afectan la captación solar son:

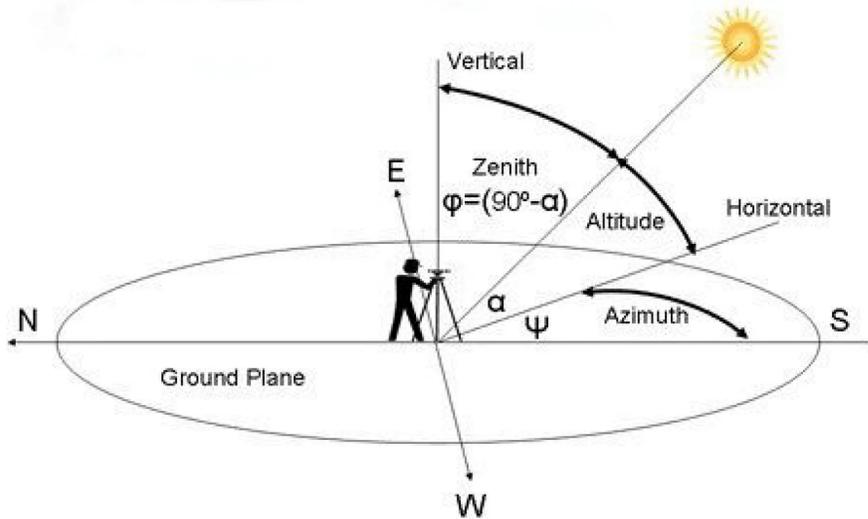


Figura 2.1: Ángulos Zenital, Azimutal y Elevación Solar [4].

- **Ángulo Zenital:** Es el ángulo entre la dirección del Sol y una línea perpendicular a la superficie terrestre. Cuando el Sol está en el punto más alto del cielo (al mediodía solar), el ángulo zenital es mínimo.
- **Ángulo de Elevación Solar:** Es el ángulo entre el horizonte y la posición del Sol en el cielo. Un ángulo de elevación alto significa que el Sol está más alto en el cielo y se recibe más radiación directa.
- **Ángulo Azimutal:** Es el ángulo entre la dirección norte y la proyección horizontal de la dirección del Sol. Este ángulo es crucial para determinar la orientación óptima de los paneles fotovoltaicos.

Como se ilustra en la figura 2.1, los ángulos zenital, azimutal y de elevación solar son fundamentales para entender cómo varía la posición del Sol a lo largo del día y del año. Estos parámetros son esenciales para optimizar la inclinación y orientación de los sistemas fotovoltaicos, con el fin de maximizar la captación de energía solar en cualquier ubicación geográfica. La inclinación de la Tierra y su órbita elíptica alrededor del Sol también causan variaciones estacionales en la cantidad de radiación solar recibida en diferentes latitudes, lo que influye directamente en la producción de energía de los sistemas fotovoltaicos.

2.1.2. Interacción de la Radiación Solar con la Atmósfera

Al atravesar la atmósfera terrestre, la radiación solar se ve afectada por varios procesos de absorción, dispersión y reflexión. Estos procesos pueden reducir significativamente la cantidad de radiación que llega a la superficie terrestre, afectando la eficiencia de los sistemas fotovoltaicos [5].

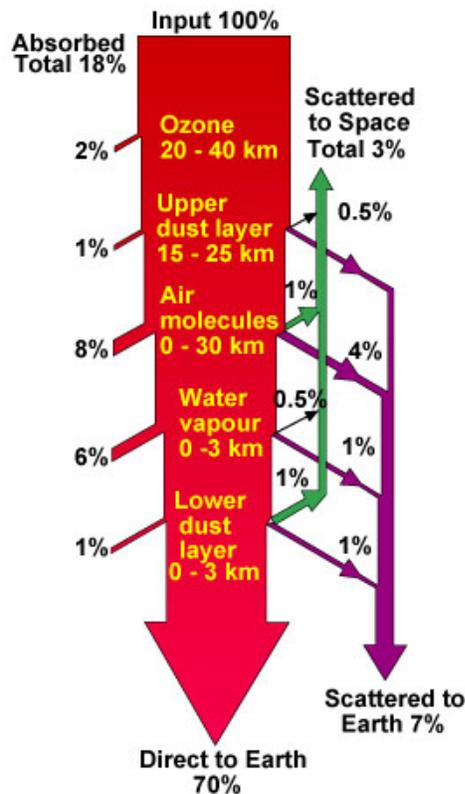


Figura 2.2: Porcentaje de luz entrante a la superficie de la Tierra [5].

La imagen presentada muestra cómo se distribuye la radiación solar a medida que atraviesa la atmósfera. De la radiación solar que llega a la Tierra, solo el 70 % alcanza directamente la superficie terrestre. El resto de la energía es absorbido, dispersado o reflejado por diferentes componentes de la atmósfera. A continuación, se describen los procesos que afectan a la radiación solar y sus impactos en los sistemas fotovoltaicos:

- **Absorción:** Algunos gases en la atmósfera, como el vapor de agua, el dióxido de carbono y el ozono, absorben ciertas longitudes de onda de la radiación solar, convirtiéndolas en calor. Este proceso reduce la cantidad de energía disponible para los sistemas fotovoltaicos [6].
- **Dispersión:** La radiación solar también puede ser dispersada por las moléculas de aire y partículas en suspensión, como el polvo y las gotas de agua. Esta dispersión es responsable de la luz difusa que ilumina el cielo y es más pronunciada en días nublados [7].
- **Reflexión:** Parte de la radiación solar es reflejada de vuelta al espacio por las nubes, el suelo y otros objetos. La fracción reflejada depende del albedo de las superficies y puede variar con el ángulo de incidencia [8].

El ángulo de incidencia es el ángulo en el que los rayos solares inciden sobre una superficie. Un ángulo de incidencia más bajo (cuando el Sol está más cerca del horizonte) resulta en una mayor dispersión y absorción, mientras que un ángulo más alto (cuando el Sol está

directamente sobre la superficie) maximiza la cantidad de radiación directa recibida [6].

2.1.3. Tipos de Irradiancia

Cuando la radiación solar llega a la Tierra, se distribuye en diferentes formas, como se muestra en la Figura 2.3. La cantidad de energía que recibe una superficie terrestre se mide en términos de irradiancia, expresada en W/m^2 , y varía dependiendo de factores como la inclinación del panel y las condiciones atmosféricas [6]. Es fundamental diferenciar los tipos de radiación solar que afectan el rendimiento de los sistemas fotovoltaicos:

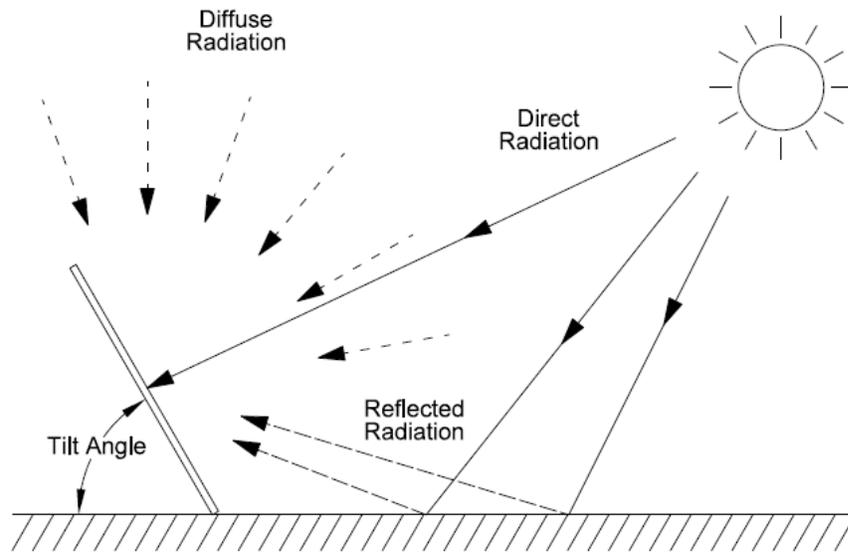


Figura 2.3: Tipos de Radiación [9].

- **Radiación Directa (DNI):** Es la radiación solar que llega directamente desde el Sol sin haber sido dispersada o absorbida por la atmósfera. Esta radiación impacta directamente en la superficie inclinada de un panel fotovoltaico.
- **Radiación Difusa (DHI):** Esta radiación ha sido dispersada en todas direcciones por moléculas y partículas en la atmósfera. A diferencia de la radiación directa, no proviene en línea recta del Sol, sino que llega desde cualquier parte del cielo [10].
- **Radiación reflejada:** Es la radiación solar que se refleja en el suelo o en superficies cercanas al panel solar. La cantidad de radiación reflejada está influenciada por el albedo de la superficie, que es el porcentaje de radiación reflejada en relación con la radiación total incidente.

En conjunto, estos tres tipos de radiación contribuyen a la cantidad total de energía que un sistema fotovoltaico puede convertir en electricidad utilizable. Como se puede ver en la Figura 2.3, la inclinación del panel es clave para optimizar la captación de radiación solar, maximizando el rendimiento del sistema fotovoltaico [6].

2.1.4. Recurso Solar

El recurso solar se refiere a la cantidad total de radiación solar disponible en un lugar determinado y es un factor clave para el diseño y la operación de los sistemas fotovoltaicos. La disponibilidad de este recurso varía considerablemente según la ubicación geográfica, la estación del año y las condiciones meteorológicas locales.

- Caracterización del Recurso Solar a Nivel Global y Local: La cantidad de radiación solar disponible varía según la latitud y la longitud. Las regiones cercanas al ecuador reciben más radiación solar que las zonas polares. A nivel local, la presencia de obstáculos como montañas, edificios y vegetación puede reducir la radiación recibida.
- Métodos de Medición del Recurso Solar: Existen diferentes formas de medir y caracterizar el recurso solar, incluyendo:
 - Estaciones Meteorológicas: Utilizan instrumentos como piranómetros y pirheliómetros para medir la radiación solar directa y difusa, respectivamente.

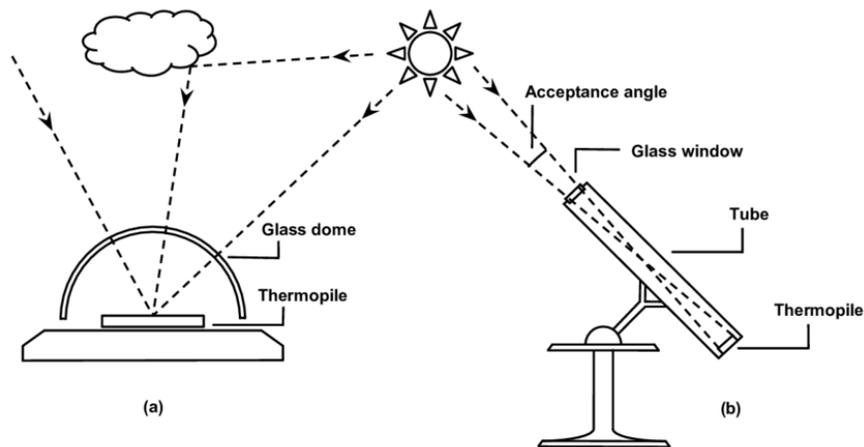


Figura 2.4: Esquema de un piranómetro (a) y un pirheliómetro (b) [11].

- Satélites Meteorológicos: Proveen datos globales sobre la radiación solar mediante sensores remotos. Estos datos son útiles para estimar el potencial solar en áreas donde no existen estaciones meteorológicas locales.
- Factores que Afectan la Disponibilidad de Radiación Solar: La radiación solar recibida en la superficie terrestre puede variar en función de la nubosidad, el polvo atmosférico y otros factores meteorológicos. Las estaciones del año también juegan un papel importante, ya que en invierno los días son más cortos y el ángulo de incidencia es mayor, lo que reduce la cantidad de radiación directa.

2.2. Sistemas Fotovoltaicos

Los sistemas fotovoltaicos (FV) se basan en la conversión directa de la radiación solar en electricidad, utilizando el efecto fotovoltaico. Este fenómeno se produce en materiales semiconductores, como el silicio, que tienen la capacidad de generar corriente eléctrica cuando se

exponen a la luz. Cuando un fotón de la radiación solar incide sobre un átomo en el material semiconductor, si la energía del fotón es suficiente, excita un electrón, haciéndolo saltar de la banda de valencia a la banda de conducción, generando un flujo de electrones o corriente eléctrica continua (DC). Este principio es la base del funcionamiento de las celdas solares que conforman los módulos fotovoltaicos y que se muestra en la siguiente figura 2.5, detallando el funcionamiento de una celda FV.

Inside a photovoltaic cell

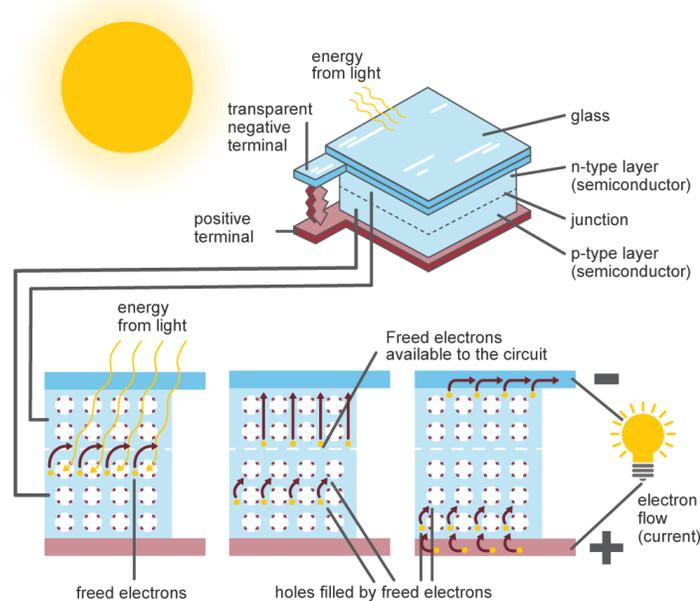


Figura 2.5: Funcionamiento del efecto fotoeléctrico dentro de una celda FV [12].

La energía generada por las celdas solares depende de varios factores, como la intensidad de la radiación solar, el ángulo de incidencia y las características del material semiconductor. Además, la eficiencia con la que los sistemas FV convierten la radiación solar en electricidad varía según la tecnología utilizada, el diseño del sistema y las condiciones climáticas locales. Los avances tecnológicos en el campo de la energía solar han permitido aumentar la eficiencia y reducir el costo de los módulos fotovoltaicos, haciendo que estos sistemas sean una opción viable para la generación de electricidad limpia y renovable en diversas aplicaciones.

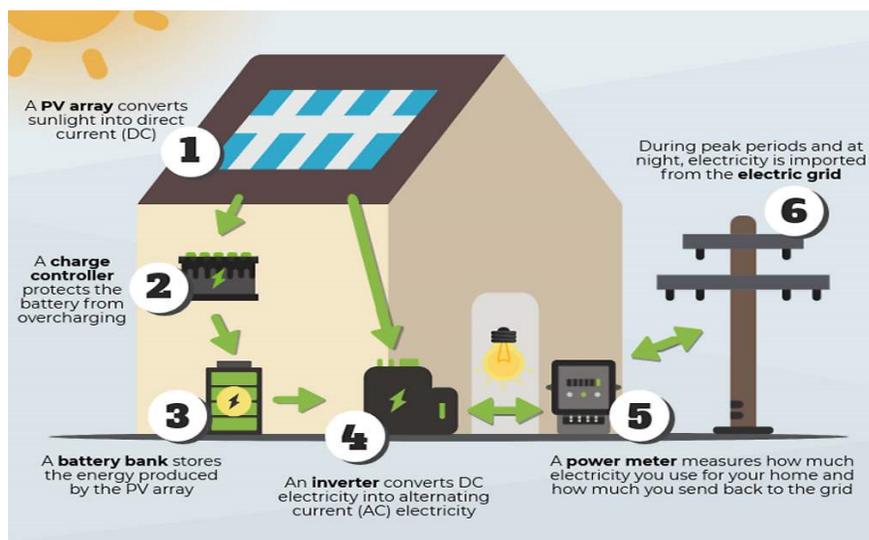


Figura 2.6: Diagrama de un Sistema Fotovoltaico en una casa [13].

La figura 2.6 de un sistema FV muestra cómo la electricidad generada por el arreglo solar se distribuye a la red eléctrica de la vivienda, permitiendo su uso eficiente y el almacenamiento de excedentes para momentos de baja producción o mayor demanda al contar con un sistema de banco de baterías.

2.2.1. Componentes de un Sistema Fotovoltaico

Un sistema fotovoltaico típico consta de varios componentes que trabajan en conjunto para captar la energía solar, convertirla en electricidad utilizable y, en algunos casos, almacenarla para su uso posterior. Los principales componentes son:

2.2.1.1. Módulos Fotovoltaicos

Los módulos fotovoltaicos son el componente principal de un sistema FV, encargados de convertir la radiación solar en electricidad mediante celdas solares. Existen diferentes tipos de módulos, según el material utilizado en las celdas, cada uno con sus ventajas y desventajas en términos de eficiencia, costo y aplicaciones. Los tres tipos principales de módulos son:

- **Monocristalino:** Fabricados con una sola estructura cristalina de silicio, estos módulos se caracterizan por su alta eficiencia y rendimiento en condiciones de baja irradiancia, aunque su costo de producción es generalmente más elevado.
- **Policristalino:** Compuestos de múltiples fragmentos de silicio, estos paneles son menos costosos de producir que los monocristalinos, pero su eficiencia es ligeramente inferior. Son comunes en instalaciones comerciales y residenciales debido a su balance entre costo y rendimiento.
- **Capa Fina (Thin-Film):** Estos paneles son flexibles y ligeros, lo que los hace adecuados para aplicaciones en superficies no tradicionales o en edificios. Sin embargo, su eficiencia es menor comparada con los módulos de silicio cristalino, aunque tienen la ventaja de mantener un rendimiento estable en temperaturas elevadas.

En la figura 2.7, se muestran los tres tipos principales de módulos fotovoltaicos: monocristalino, policristalino y de capa fina.

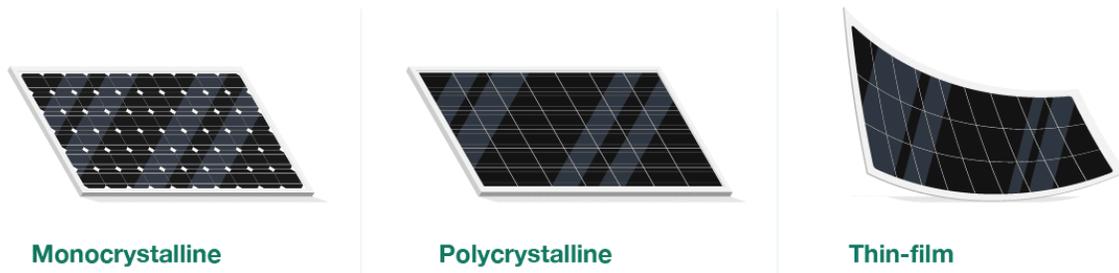


Figura 2.7: Tipos de módulos fotovoltaicos [14].

Cada uno de estos tipos de paneles tiene aplicaciones específicas según las necesidades del sistema y las condiciones del entorno, lo que permite a los diseñadores de sistemas FV optimizar el rendimiento y la viabilidad económica de cada instalación.

2.2.1.1.1. Conexión en Strings

Los módulos fotovoltaicos se agrupan en strings, que son cadenas de módulos conectados en serie para alcanzar el voltaje adecuado para el sistema. Un string está compuesto por varios módulos, dependiendo del diseño del sistema y del tipo de inversor utilizado. Al conectar los módulos en serie, el voltaje de cada módulo se suma, mientras que la corriente permanece constante.

El diseño de los strings es crucial para maximizar la eficiencia del sistema fotovoltaico, ya que permite que la energía generada por los módulos se adapte a los requerimientos del inversor. En la Figura 2.8, se muestra un diagrama que ilustra la conexión de los módulos en strings y su interacción con los inversores.

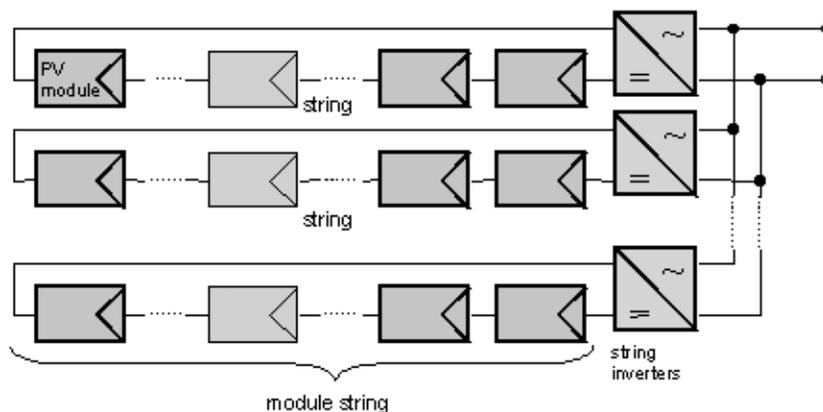


Figura 2.8: Conexión de strings [15].

2.2.1.2. Inversor

Los módulos fotovoltaicos generan electricidad en forma de corriente continua. Sin embargo, la mayoría de los dispositivos eléctricos, tanto en hogares como en instalaciones comerciales, funcionan con corriente alterna. Para adaptar la energía generada por los módulos solares a estos sistemas eléctricos, se utiliza un componente esencial conocido como inversor. Este dispositivo convierte la corriente continua en corriente alterna, asegurando la compatibilidad con la red eléctrica y los aparatos de consumo.

Además de su función básica de conversión de energía, los inversores modernos incluyen características avanzadas como sistemas de monitoreo. Estos permiten a los operadores visualizar el rendimiento del sistema en tiempo real, optimizar su operación y detectar posibles problemas de manera proactiva. Gracias a estas funcionalidades, el inversor no solo desempeña un papel clave en la conversión de energía, sino que también contribuye a la eficiencia general y al mantenimiento de los sistemas fotovoltaicos.



Figura 2.9: Ejemplo de inversores en un sistema fotovoltaico [16].

Como se muestra en la Figura 2.9, los inversores son fundamentales para integrar la energía solar generada en las redes eléctricas tradicionales, lo que permite que los sistemas fotovoltaicos sean viables tanto para el autoconsumo como para la venta de electricidad a la red.

2.2.1.3. Estructuras de Montaje

Los módulos fotovoltaicos requieren una orientación adecuada para captar la mayor cantidad de radiación solar posible. Las estructuras de montaje se encargan de posicionar los paneles de forma óptima, y pueden ser de dos tipos principales: fijas y móviles.

Las estructuras fijas son aquellas donde los paneles permanecen en una posición constante, optimizada para recibir la máxima radiación solar en función de la ubicación geográfica. Estas estructuras se utilizan tanto en instalaciones sobre techos como en sistemas montados

en suelo. En la figura 2.10, se observa un ejemplo de ambas configuraciones: a la izquierda, un sistema montado en suelo y, a la derecha, un sistema instalado en un techo. La simplicidad y el bajo mantenimiento son sus principales ventajas, pero presentan una limitación en la captación de radiación durante todo el día y en distintas estaciones del año.

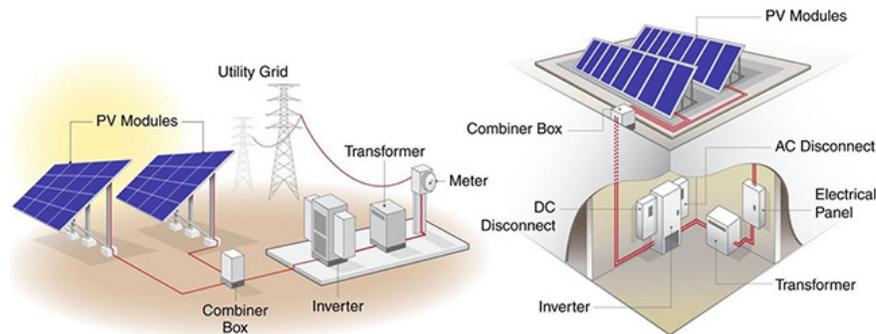


Figura 2.10: Montajes fijos. A la izquierda montaje en suelo y a la derecha montaje sobre techo [17].

Por otro lado, las estructuras móviles o con sistema tracker están diseñadas para seguir el movimiento del Sol a lo largo del día. Estas estructuras ajustan la orientación de los paneles en uno o dos ejes, lo que maximiza la cantidad de radiación solar recibida. Como se puede ver en la figura 2.11, los sistemas tracker pueden moverse en un solo eje, ya sea horizontal o vertical, dependiendo del diseño. Estos sistemas son más complejos y requieren mantenimiento adicional, pero mejoran significativamente la eficiencia del sistema fotovoltaico, especialmente en grandes instalaciones.

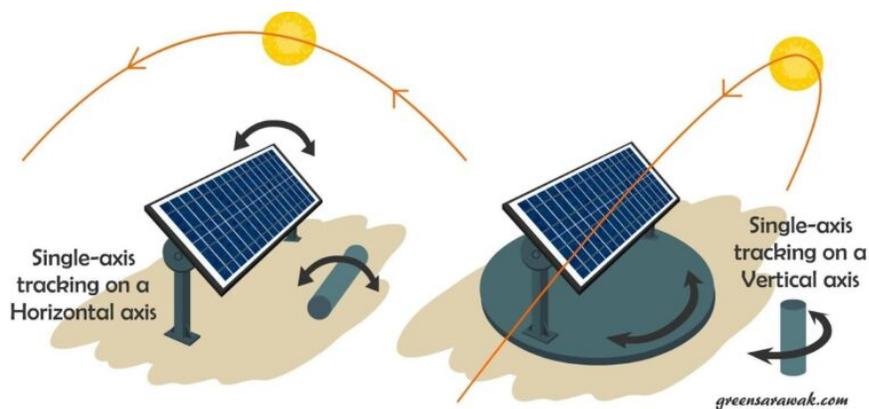


Figura 2.11: Montaje con sistema tracker [18].

2.2.1.4. Sistemas de Almacenamiento

En algunos sistemas, especialmente en instalaciones fuera de la red o donde se busca una mayor independencia energética, se utilizan baterías para almacenar el exceso de electricidad generada durante el día. Esto permite utilizar la energía almacenada durante la noche o en días nublados.

2.2.1.5. Cableado y Componentes Eléctricos

Conectan los módulos, el inversor y, si es necesario, el sistema de almacenamiento. El diseño adecuado del sistema eléctrico es crucial para minimizar pérdidas de energía y asegurar un funcionamiento eficiente.

2.2.2. Tipos de Sistemas Fotovoltaicos

Los sistemas fotovoltaicos pueden clasificarse en cuatro tipos principales: sistemas fotovoltaicos directos, sistemas conectados a la red (On-Grid), sistemas aislados (Off-Grid) y sistemas conectados a la red con respaldo de baterías. Cada uno de estos sistemas tiene aplicaciones específicas en función de las necesidades energéticas y el acceso a la red eléctrica.

2.2.2.1. Sistemas Fotovoltaicos Directos (PV Direct)

Los sistemas fotovoltaicos directos son los más simples, ya que no están conectados ni a la red eléctrica ni a baterías. Solo generan electricidad cuando hay radiación solar directa. Son ideales para aplicaciones como bombas de agua, ventiladores o sistemas de ventilación que solo necesitan funcionar cuando el sol está presente. Son sistemas de bajo costo y mantenimiento, pero no ofrecen energía de respaldo en ausencia de radiación solar.

2.2.2.2. Sistemas Conectados a la Red (On-Grid)

Los sistemas On-Grid, también conocidos como Grid-tied (conceptos equivalentes), son los más comunes y están directamente enlazados con la red eléctrica pública. Estos sistemas permiten utilizar la energía solar durante el día, mientras que el excedente se inyecta en la red, generando beneficios económicos. En momentos de baja producción solar, como por la noche, la energía se toma de la red, garantizando un suministro continuo. Sin embargo, cuando la red eléctrica falla, estos sistemas no pueden operar. El diagrama en la Figura 2.12 ilustra el flujo de energía en este tipo de sistemas.

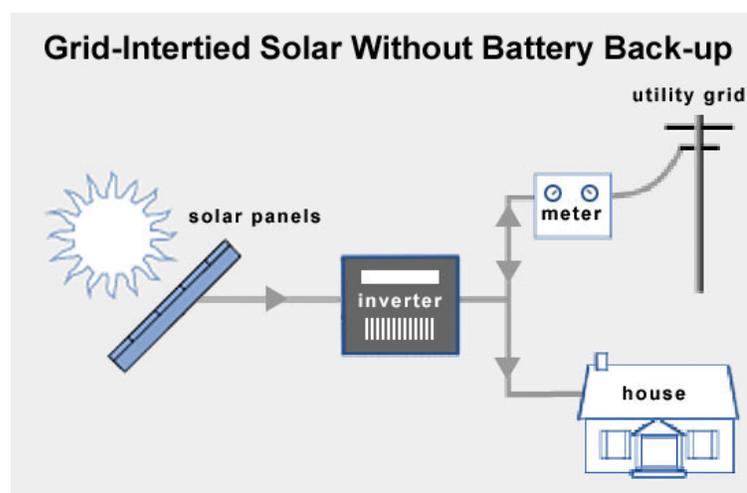


Figura 2.12: Sistema Fotovoltaico On-Grid [19].

2.2.2.3. Sistemas Aislados (Off-Grid)

Los sistemas Off-Grid funcionan de manera independiente a la red eléctrica y requieren baterías para almacenar la energía generada durante el día. Son ideales para ubicaciones remotas o donde no hay acceso a la red pública. Estos sistemas aseguran una fuente de energía continua, pero deben estar diseñados para cubrir completamente las necesidades del usuario, ya que no tienen respaldo de la red. El diagrama en la Figura 2.13 muestra un ejemplo de cómo funciona un sistema Off-Grid.

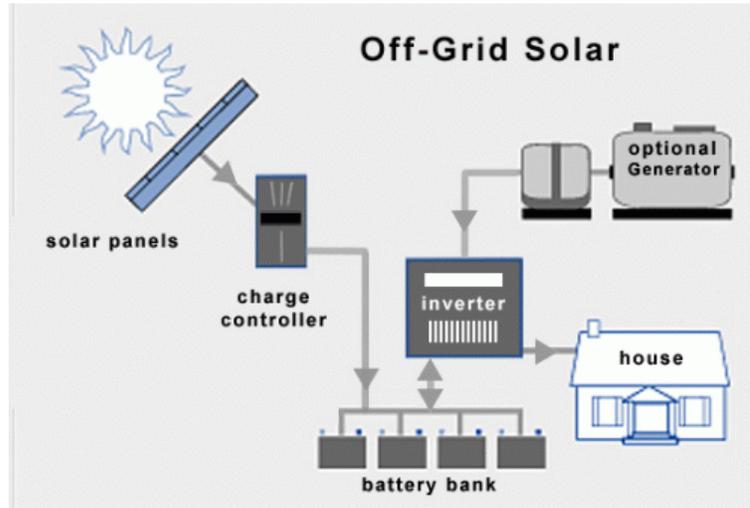


Figura 2.13: Sistema Fotovoltaico Off-Grid [19].

2.2.2.4. Sistemas Conectados a la Red con Respaldo de Baterías

Los sistemas conectados a la red con respaldo de baterías son similares a los sistemas On-Grid, pero añaden almacenamiento en baterías. Esto permite que los usuarios almacenen energía durante el día y la utilicen cuando la red eléctrica falla, proporcionando un nivel adicional de seguridad energética. Aunque estos sistemas son más costosos y complejos, ofrecen la ventaja de seguir funcionando en caso de cortes de energía en la red. El diagrama en la Figura 2.14 ilustra cómo operan estos sistemas.

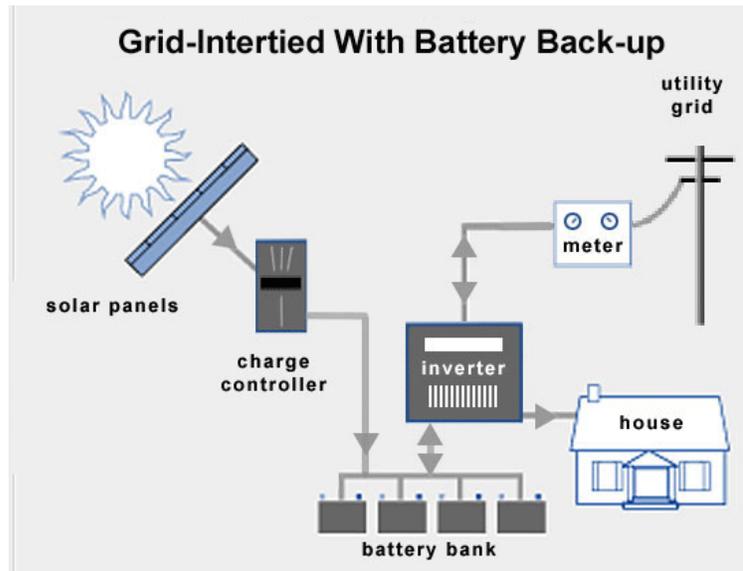


Figura 2.14: Sistema Fotovoltaico On-Grid con respaldo de baterías [19].

2.2.3. Rendimiento y Eficiencia de los Sistemas Fotovoltaicos

El rendimiento de un sistema fotovoltaico depende de varios factores, como la radiación solar disponible, la calidad de los componentes y las condiciones ambientales. Algunos de los principales factores que afectan la eficiencia de un sistema son:

- **Temperatura:** A medida que la temperatura de los módulos aumenta, su eficiencia tiende a disminuir. Los módulos fotovoltaicos funcionan mejor a temperaturas más bajas.
- **Ángulo de Incidencia:** La inclinación y orientación de los módulos respecto al Sol son determinantes para maximizar la captación de radiación.
- **Sombreamiento:** Cualquier obstrucción, como árboles, edificios o nubes, que bloquee la luz solar directa, reduce significativamente la producción de energía.
- **Pérdidas en el Sistema:** Incluyen las pérdidas por cableado, inversores, y la calidad de los módulos.

Un parámetro comúnmente utilizado para evaluar el rendimiento de un sistema FV es el Performance Ratio (PR), que mide la eficiencia general del sistema, comparando la energía producida con la energía potencialmente disponible según la radiación solar recibida. Un PR del 100 % indicaría un sistema sin pérdidas, lo cual es prácticamente imposible en la realidad debido a las pérdidas por temperatura, sombras, suciedad y eficiencia de los inversores y cables. Sin embargo, los sistemas fotovoltaicos tienden a tener un PR realista entre el 70 % y 90 % debido a pérdidas operativas y degradación de los módulos con el tiempo [20].

2.2.4. Modelos de Simulación de Sistemas FV

El diseño, análisis y optimización de sistemas fotovoltaicos se ha facilitado enormemente gracias a la utilización de modelos de simulación. Estas herramientas permiten prever el rendimiento de un sistema bajo diferentes condiciones, optimizando su diseño y asegurando una

mejor integración en el entorno de instalación.

Existen varios programas y bibliotecas que simulan el comportamiento de los sistemas fotovoltaicos. Entre los más utilizados están:

- *PVLIB*: Una biblioteca de Python ampliamente usada para modelar y simular sistemas fotovoltaicos a partir de datos meteorológicos y parámetros del sistema.
- *HelioScope*: Un software web que permite el diseño y simulación de sistemas fotovoltaicos, integrando modelado 3D para optimizar el diseño y cálculo de sombras.
- *PVSyst*: Un programa avanzado de simulación de sistemas FV que incluye herramientas para el diseño, simulación y análisis de sistemas a gran escala.

Estos modelos consideran factores como la radiación solar, el ángulo de los módulos, las pérdidas por temperatura y sombreadamiento, y la eficiencia del inversor, entre otros, proporcionando estimaciones detalladas sobre la generación de energía esperada.

2.2.5. Pasos de Simulación de Sistemas Fotovoltaicos

El modelado de un sistema fotovoltaico sigue una serie de pasos que permiten estimar la potencia generada, desde la medición de la irradiancia hasta la conversión de la energía a corriente alterna. Cada paso puede afectar el resultado final, y es fundamental tener en cuenta las posibles pérdidas y condiciones operativas del sistema. El diagrama de la Figura 2.15 presenta los principales pasos en este proceso de simulación, destacando las etapas clave para evaluar el rendimiento del sistema.

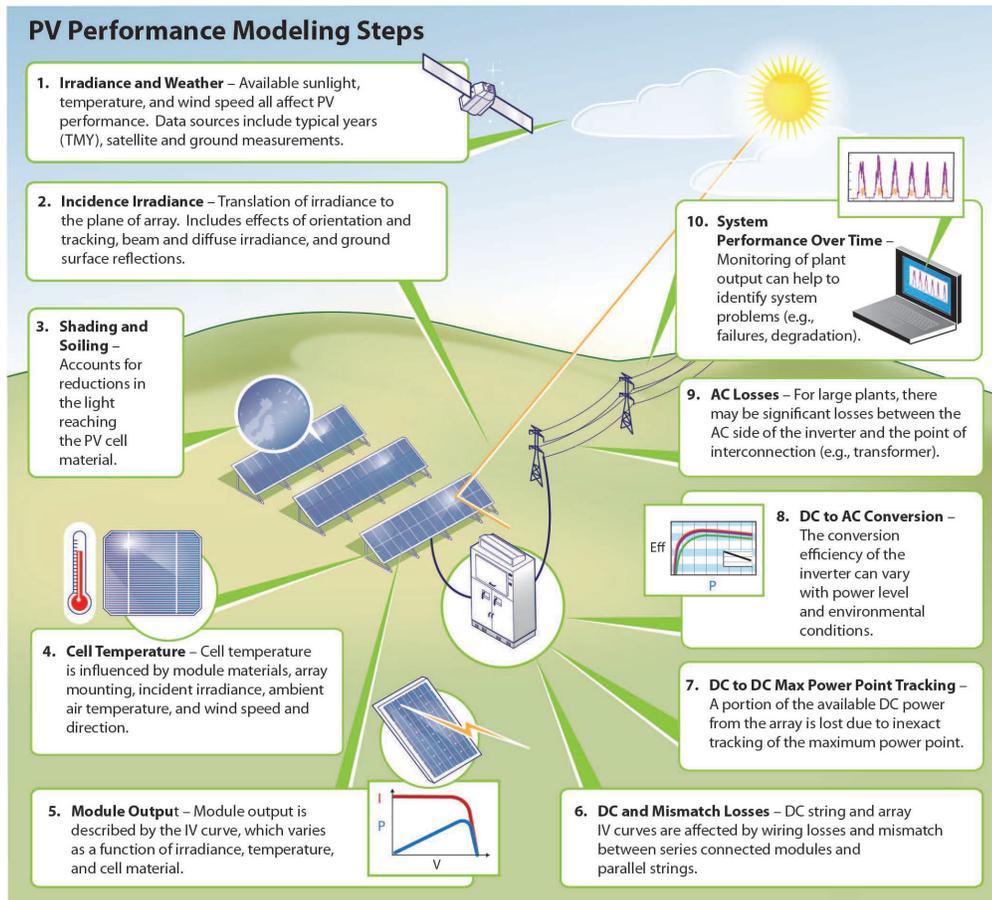


Figura 2.15: Etapas clave en el modelado del desempeño de sistemas fotovoltaicos [21].

El diagrama proporciona una visión general del proceso de simulación de un sistema fotovoltaico, abarcando desde la captura de irradiancia y las condiciones meteorológicas, hasta las pérdidas por temperatura, sombreado y las eficiencias involucradas en la conversión de energía mediante los inversores. A continuación, se ofrece una breve descripción de cada una de las etapas clave del proceso:

- 1. Irradiancia y Clima:** Uso de datos meteorológicos de irradiancia, temperatura y velocidad del viento, todos factores que afectan el rendimiento del sistema fotovoltaico.
- 2. Irradiancia Incidente:** Se calcula la irradiancia que llega al plano de los paneles solares, considerando su inclinación, orientación y las condiciones de radiación directa y difusa.
- 3. Sombreado y Suciedad:** Se modelan las pérdidas asociadas al sombreado de los módulos y la suciedad acumulada sobre ellos.
- 4. Temperatura del Módulo:** La temperatura del módulo se calcula en función de la irradiancia, la temperatura ambiente y el viento, ya que estos factores afectan la eficiencia de los módulos solares.
- 5. Potencia del Módulo:** La potencia generada por el módulo depende de la curva IV (corriente vs. voltaje) del módulo, la cual varía con la irradiancia y la temperatura.

6. **Pérdidas por Mismatch:** Se consideran las pérdidas asociadas a la interconexión de módulos en serie y paralelo, y las diferencias de rendimiento entre ellos.
7. **Seguimiento del Punto de Máxima Potencia (MPPT):** Se simulan las pérdidas debidas a ineficiencias en el seguimiento del punto de máxima potencia de los inversores.
8. **Conversión de DC a AC:** Se modelan las pérdidas en la conversión de la energía generada en corriente continua (DC) a corriente alterna (AC) mediante los inversores.
9. **Pérdidas AC:** En plantas más grandes, se simulan las pérdidas en el lado AC, desde el inversor hasta el punto de interconexión con la red eléctrica.
10. **Desempeño del Sistema en el Tiempo:** Se evalúa el desempeño global del sistema a lo largo del tiempo, monitoreando la salida de energía y detectando posibles problemas o degradaciones.

Con esta referencia visual, pasamos a detallar los pasos específicos del proceso de obtención de datos y simulaciones en *PVLIB*.

2.3. Digital Twins

Los Digital Twins, o gemelos digitales, son representaciones virtuales precisas de objetos, sistemas o procesos físicos que se utilizan para simular, analizar y optimizar su rendimiento en tiempo real [22]. Esta tecnología permite una integración fluida entre el mundo físico y digital, facilitando la monitorización continua y la toma de decisiones informadas. Los gemelos digitales reflejan no solo el estado actual del objeto físico, sino que también pueden predecir su comportamiento futuro mediante modelos analíticos, simulaciones y algoritmos de aprendizaje automático [23].

El concepto de Digital Twin se originó en la industria aeroespacial y de manufactura, donde la precisión y la previsibilidad son cruciales. En estos sectores, los Digital Twins se utilizan para monitorear el rendimiento de equipos complejos, predecir fallas y planificar el mantenimiento preventivo, mejorando la eficiencia operativa y reduciendo costos [24]. La integración de tecnologías de Internet de las Cosas (IoT), inteligencia artificial (IA) y big data ha permitido que los Digital Twins evolucionen y se apliquen en una amplia variedad de industrias, incluyendo la energía, la construcción y la gestión de infraestructuras [25].

2.3.1. Digital Twins en Energía

En el sector energético, los Digital Twins se utilizan para optimizar la generación, distribución y consumo de energía. Estos gemelos digitales permiten a las empresas de energía monitorear en tiempo real el rendimiento de sus activos, como plantas de generación, redes de distribución y sistemas de almacenamiento de energía [26]. Además, pueden simular diferentes escenarios operativos y condiciones ambientales para prever problemas potenciales y optimizar la planificación y operación del sistema energético [27].

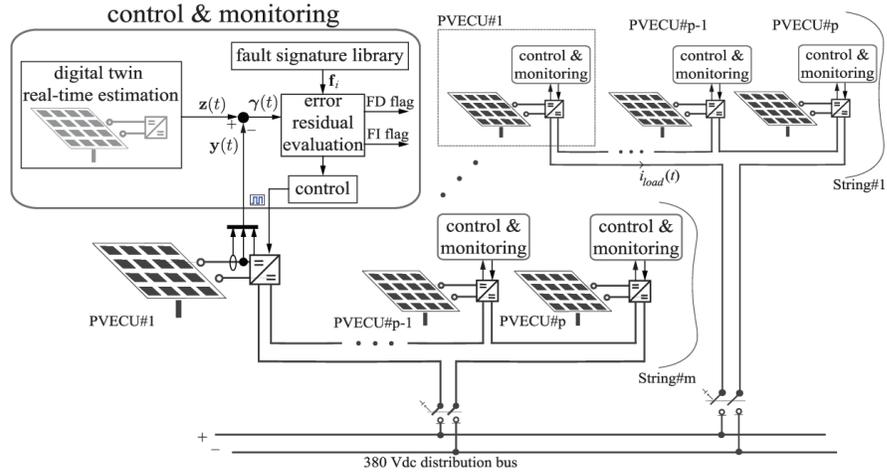


Figura 2.16: Diagrama de monitoreo y control de una planta FV usando digital twin [28].

El diagrama en la Figura 2.16 muestra un sistema de control y monitoreo en una planta FV que utiliza un Digital Twin para la estimación en tiempo real de las variables del sistema. En este esquema, el Digital Twin se integra con un sistema de evaluación de errores que compara las lecturas reales de los sensores con las predicciones del gemelo digital. Este análisis ayuda a detectar fallas o desviaciones en el rendimiento del sistema fotovoltaico activando avisos de fallos, lo que permite una rápida respuesta operativa.

2.3.2. Digital Twins y ML en Sistemas Fotovoltaicos

La integración de Digital Twins y Machine Learning (ML) en sistemas fotovoltaicos permite optimizar su rendimiento mediante análisis predictivos. Los gemelos digitales, en conjunto con algoritmos de ML, pueden procesar grandes volúmenes de datos generados por los sistemas FV para predecir la producción de energía y detectar posibles fallas antes de que ocurran, mejorando la eficiencia operativa [29]. Como se muestra en el diagrama 2.17, los datos de irradiación y producción de energía se monitorean en tiempo real y se comparan con los resultados de un modelo de ML entrenado, lo que permite alertar al equipo de operaciones y mantenimiento (O&M) cuando se detectan anomalías [30].

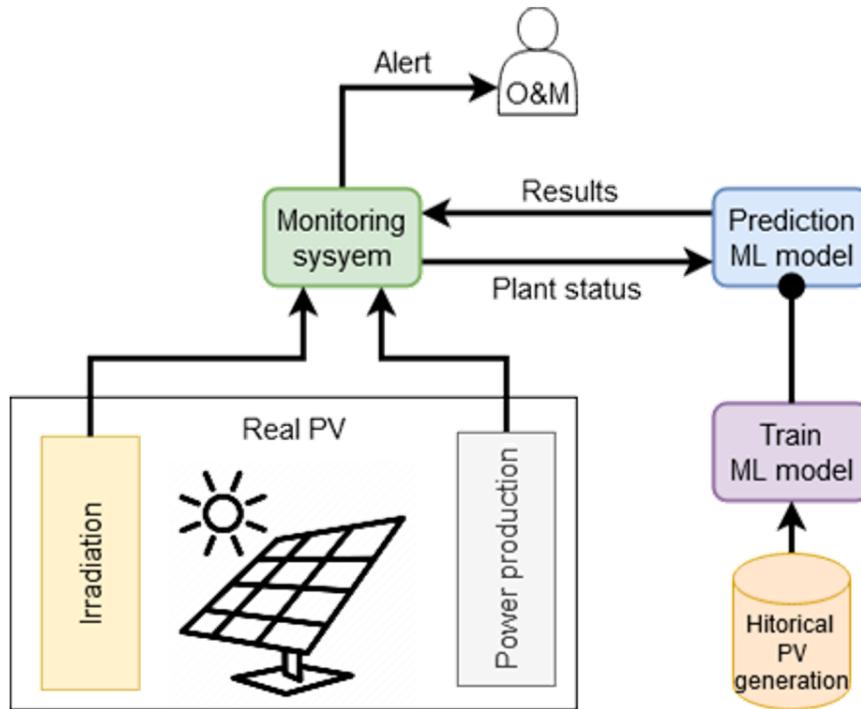


Figura 2.17: Sistema de monitoreo usando un modelo de ML entrenado para el cálculo de potencia FV [31].

Este enfoque también facilita el mantenimiento predictivo, permitiendo a las empresas realizar intervenciones basadas en datos en lugar de seguir calendarios fijos. Con ML, los Digital Twins pueden mejorar continuamente sus predicciones, optimizando la gestión de la planta y reduciendo costos operativos al minimizar la necesidad de mantenimiento rutinario [30].

2.4. *PVLIB*

PVLIB es una biblioteca de Python ampliamente utilizada para la simulación y modelado de sistemas fotovoltaicos (FV). Fue desarrollada inicialmente por Sandia National Laboratories y ha sido adoptada por una comunidad global que continúa expandiendo su funcionalidad. Esta herramienta es de código abierto y proporciona una amplia gama de métodos y modelos que permiten a ingenieros e investigadores simular con precisión la generación de energía de sistemas fotovoltaicos bajo diferentes condiciones meteorológicas y características técnicas del sistema [32].

PVLIB permite modelar diversos componentes de un sistema FV, desde la irradiancia solar que incide en los módulos hasta la conversión de energía de corriente continua (DC) a corriente alterna (AC). Su flexibilidad y precisión han convertido a *PVLIB* en una herramienta de referencia en estudios de rendimiento de sistemas solares. A lo largo de esta sección, se detallarán los componentes clave de *PVLIB*, sus modelos más importantes y su relevancia en el modelado de sistemas FV.

2.4.1. Componentes Clave de *PVLIB*

PVLIB incluye varios módulos que permiten realizar simulaciones detalladas de sistemas fotovoltaicos. A continuación, se presentan algunos de los componentes más importantes para la simulación.

2.4.1.1. Modelos de Irradiancia

Uno de los aspectos más críticos para la simulación de sistemas fotovoltaicos es el cálculo de la irradiancia solar que llega a los módulos. En *PVLIB*, existen varios modelos que permiten realizar este cálculo, ajustándose a diferentes condiciones atmosféricas y características del panel. Los modelos de irradiancia proporcionan estimaciones sobre la cantidad de radiación solar directa, difusa y reflejada que incide sobre una superficie inclinada.

Uno de los modelos más utilizados en *PVLIB* es el modelo de Perez para la transposición de irradiancia, que convierte la irradiancia global horizontal (GHI) en irradiancia en el plano inclinado del panel, teniendo en cuenta la posición del sol y el ángulo de inclinación del panel. Esta es una de las ecuaciones clave utilizadas para calcular la irradiancia en el plano del arreglo (POA):

$$POA = G_{beam} \cdot \cos(AOI) + G_{diffuse} + G_{reflected} \quad (2.1)$$

Donde:

- G_{beam} es la irradiancia directa normal.
- AOI es el ángulo de incidencia de la radiación solar sobre el panel.
- $G_{diffuse}$ es la irradiancia difusa.
- $G_{reflected}$ es la irradiancia reflejada desde el suelo.

El modelo de Perez es particularmente útil en ubicaciones donde la irradiancia difusa es significativa, como en zonas urbanas o con alta nubosidad. Otros modelos que también se incluyen en *PVLIB* son el modelo HayDavies y el modelo de cielo despejado, que permiten estimar la irradiancia en situaciones específicas, como días sin nubes [8].

2.4.1.2. Modelos de Pérdidas

Los sistemas fotovoltaicos experimentan diversas pérdidas que afectan su eficiencia y rendimiento global. *PVLIB* ofrece un conjunto de modelos para estimar las pérdidas asociadas a distintos factores, como la suciedad en los paneles, el sombreado, el desajuste eléctrico entre módulos (mismatch) y las pérdidas por cableado. Uno de los modelos de pérdidas más utilizados es el de PVWatts, desarrollado por NREL, el cual integra una serie de porcentajes que describen las pérdidas asociadas a cada factor.

Las pérdidas se expresan como un porcentaje del total de la potencia generada en corriente continua (DC), y se pueden calcular utilizando la siguiente ecuación:

$$L_{total}(\%) = 100 \left[1 - \Pi_i \left(1 - \frac{L_i}{100} \right) \right] \quad (2.2)$$

Donde cada L_i representa un tipo específico de pérdida (por ejemplo, pérdidas por sombreado o suciedad). *PVLIB* permite ajustar estos valores de acuerdo con las condiciones específicas del sistema fotovoltaico y su entorno.

A continuación, se describen los principales tipos de pérdidas considerados:

- **Pérdidas por ensuciamiento:** La acumulación de polvo y suciedad en los paneles reduce la cantidad de luz solar disponible para la conversión en energía eléctrica.
- **Pérdidas por sombreado:** La sombra proyectada por obstáculos cercanos, como edificios o árboles, disminuye la cantidad de irradiancia solar recibida por los módulos.
- **Pérdidas por cobertura de nieve:** En regiones con nevadas, la acumulación de nieve sobre los paneles bloquea la luz solar, reduciendo la producción de energía.
- **Pérdidas por desajustes:** Las diferencias en las características eléctricas de los módulos dentro de un mismo sistema causan una disminución en la eficiencia global del sistema.
- **Pérdidas por cableado y conexiones:** La resistencia en los cables y las conexiones eléctricas entre los módulos y el inversor provoca una pérdida de energía durante la transmisión.
- **Degradación inducida por luz (LID):** La potencia de salida de los módulos disminuye durante las primeras horas de exposición a la luz solar, estabilizándose posteriormente.
- **Calificación de nombre:** La diferencia entre la potencia nominal de los módulos y su rendimiento real en condiciones operativas contribuye a las pérdidas del sistema.
- **Pérdidas por envejecimiento:** Con el tiempo, la eficiencia de los módulos disminuye debido a la degradación natural de los materiales, generalmente a una tasa de 0.5% anual.

2.4.1.3. Modelos de Temperatura

La temperatura del módulo fotovoltaico tiene un impacto directo en su eficiencia. A medida que los módulos se calientan, su capacidad para convertir la luz solar en electricidad disminuye. Por ello, *PVLIB* ofrece modelos para calcular la temperatura del módulo en función de la irradiancia incidente y la temperatura ambiente.

Uno de los modelos más utilizados es el Sandia Array Performance Model (SAPM), que predice la temperatura del módulo utilizando la siguiente ecuación:

$$T_m = T_a + a \cdot G_{POA} + b \cdot (G_{POA})^2 \quad (2.3)$$

Donde:

- T_m es la temperatura del módulo.
- T_a es la temperatura ambiente.
- G_{POA} es la irradiancia en el plano del arreglo.

- a y b son coeficientes específicos del módulo.

El ajuste preciso de estos coeficientes es crucial para simular correctamente el impacto de la temperatura en el rendimiento del sistema fotovoltaico [33].

2.4.1.4. Conversión de DC a AC

Uno de los últimos pasos en la simulación de sistemas fotovoltaicos es la conversión de la potencia en corriente continua (DC) generada por los módulos en corriente alterna (AC) mediante un inversor. *PVLIB* modela este proceso utilizando ecuaciones que consideran la eficiencia del inversor (η_{inv}) y las pérdidas de conversión.

La ecuación utilizada para calcular la potencia en AC es:

$$P_{AC} = P_{DC} \cdot \eta_{inv} \quad (2.4)$$

Donde:

- P_{AC} es la potencia de salida en corriente alterna.
- P_{DC} es la potencia generada en corriente continua.
- η_{inv} es la eficiencia del inversor [34].

La eficiencia del inversor varía en función de la carga y las condiciones operativas, por lo que este factor es importante para obtener simulaciones precisas.

2.4.2. Aplicaciones de *PVLIB*

PVLIB es utilizado en una amplia variedad de aplicaciones relacionadas con el análisis y la simulación de sistemas fotovoltaicos. Algunas de sus aplicaciones más destacadas incluyen:

- Simulaciones horarias y anuales: *PVLIB* permite realizar simulaciones en diferentes escalas temporales, desde predicciones horarias hasta proyecciones anuales. Esto es esencial para estimar la producción de energía a lo largo de un período y evaluar la viabilidad económica de un proyecto solar.
- Optimización del diseño de sistemas fotovoltaicos: Con *PVLIB*, es posible optimizar el diseño de un sistema solar ajustando la inclinación y orientación de los paneles, además de calcular los efectos de diferentes configuraciones de montaje.
- Estudios de viabilidad: Utilizando datos meteorológicos típicos de un lugar específico (TMY), *PVLIB* permite realizar estudios de viabilidad que evalúan el rendimiento esperado de un sistema FV en un sitio determinado [32].

2.5. Machine Learning

Machine Learning (ML), o aprendizaje automático, es una rama de la inteligencia artificial que se centra en el desarrollo de algoritmos y modelos que permiten a las computadoras

aprender y hacer predicciones o decisiones basadas en datos. A diferencia de los sistemas tradicionales programados explícitamente para realizar una tarea específica, los modelos de ML mejoran su rendimiento a través de la experiencia, es decir, a medida que son expuestos a más datos [35]. Esta capacidad de aprender y adaptarse sin intervención humana directa hace que ML sea una herramienta poderosa para resolver problemas complejos en una amplia variedad de dominios.

2.5.1. Tipos de Machine Learning

El aprendizaje automático se clasifica generalmente en tres tipos: supervisado, no supervisado y de refuerzo. En el aprendizaje supervisado, los modelos son entrenados con un conjunto de datos etiquetados, donde el algoritmo aprende a mapear las entradas a las salidas deseadas. Ejemplos de aplicaciones incluyen la clasificación de imágenes y la predicción de precios de viviendas [36]. En el aprendizaje no supervisado, el modelo trabaja con datos no etiquetados y busca estructuras o patrones ocultos en los datos. Ejemplos de esto incluyen la agrupación (clustering) y la reducción de dimensionalidad. Por último, en el aprendizaje por refuerzo, el modelo aprende a tomar decisiones secuenciales mediante la interacción con un entorno, recibiendo recompensas o penalizaciones según las acciones que realiza [37].

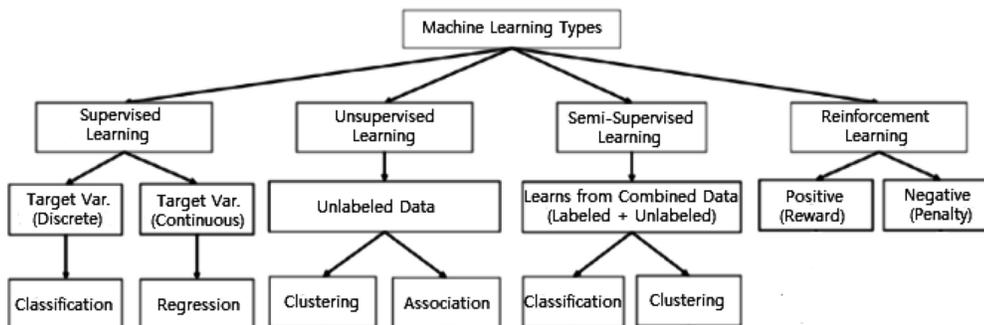


Figura 2.18: Técnicas de Machine Learning.

2.5.2. Machine Learning en Energía y Fotovoltaica

En el sector energético, y específicamente en los sistemas fotovoltaicos, ML juega un papel crucial en la optimización y gestión de la energía. Los algoritmos de ML pueden analizar datos históricos y en tiempo real para predecir la producción de energía solar, identificar patrones de consumo y optimizar la operación de los sistemas. Esto permite mejorar la eficiencia y reducir los costos operativos. Además, ML se utiliza en el mantenimiento predictivo de los sistemas fotovoltaicos, donde los modelos pueden prever fallos y sugerir intervenciones antes de que ocurran, minimizando el tiempo de inactividad y extendiendo la vida útil de los equipos [38].

2.5.3. Regresión Lineal y Técnicas de Machine Learning

La regresión lineal es una técnica de modelado estadístico utilizada para predecir el valor de una variable dependiente (respuesta) basada en el valor de una o más variables independientes

(predictoras). El objetivo principal de la regresión lineal es establecer una relación lineal entre las variables, modelando la relación con una ecuación de la forma

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n + \epsilon \quad (2.5)$$

donde y es la variable dependiente, β_0 es la intersección, $\beta_1, \beta_2, \dots, \beta_n$ son los coeficientes de las variables independientes, y ϵ es el término de error [39].

La regresión lineal se divide en dos tipos principales: regresión lineal simple y regresión lineal múltiple. La regresión lineal simple utiliza una sola variable independiente para predecir el valor de la variable dependiente, mientras que la regresión lineal múltiple utiliza dos o más variables independientes. Esta técnica es ampliamente utilizada en diversas disciplinas, incluyendo la economía, la biología, la ingeniería y las ciencias sociales, debido a su simplicidad y capacidad para interpretar fácilmente las relaciones entre variables [40].

Si bien la regresión lineal es un modelo fácil de interpretar, tiene limitaciones cuando los datos no se ajustan bien a una relación lineal o cuando existe multicolinealidad, es decir, cuando las variables independientes están altamente correlacionadas [41]. Para estos casos, es más apropiado utilizar técnicas de machine learning más complejas, que pueden capturar patrones no lineales y dependencias más profundas en los datos.

Una técnica más avanzada empleada en este proyecto es la arquitectura Transformer, la cual ha revolucionado las tareas que involucran secuencias, como la predicción de series temporales y el procesamiento del lenguaje natural. A diferencia de los modelos tradicionales de redes neuronales, los Transformers se basan en mecanismos de atención, lo que les permite enfocarse en las dependencias más relevantes dentro de grandes secuencias de datos sin la necesidad de procesarlas en orden. Esto les permite procesar secuencias largas con mayor eficacia y precisión.

La arquitectura Transformer se estructura en bloques de codificadores y decodificadores que emplean atención multi-cabezal para capturar relaciones en los datos, lo que es particularmente útil para el análisis de series temporales en sistemas fotovoltaicos. En la figura 2.19, se muestra la arquitectura del modelo Transformer, que destaca cómo el mecanismo de atención y las capas de normalización permiten que los datos fluyan a través del modelo con mayor eficiencia [42].

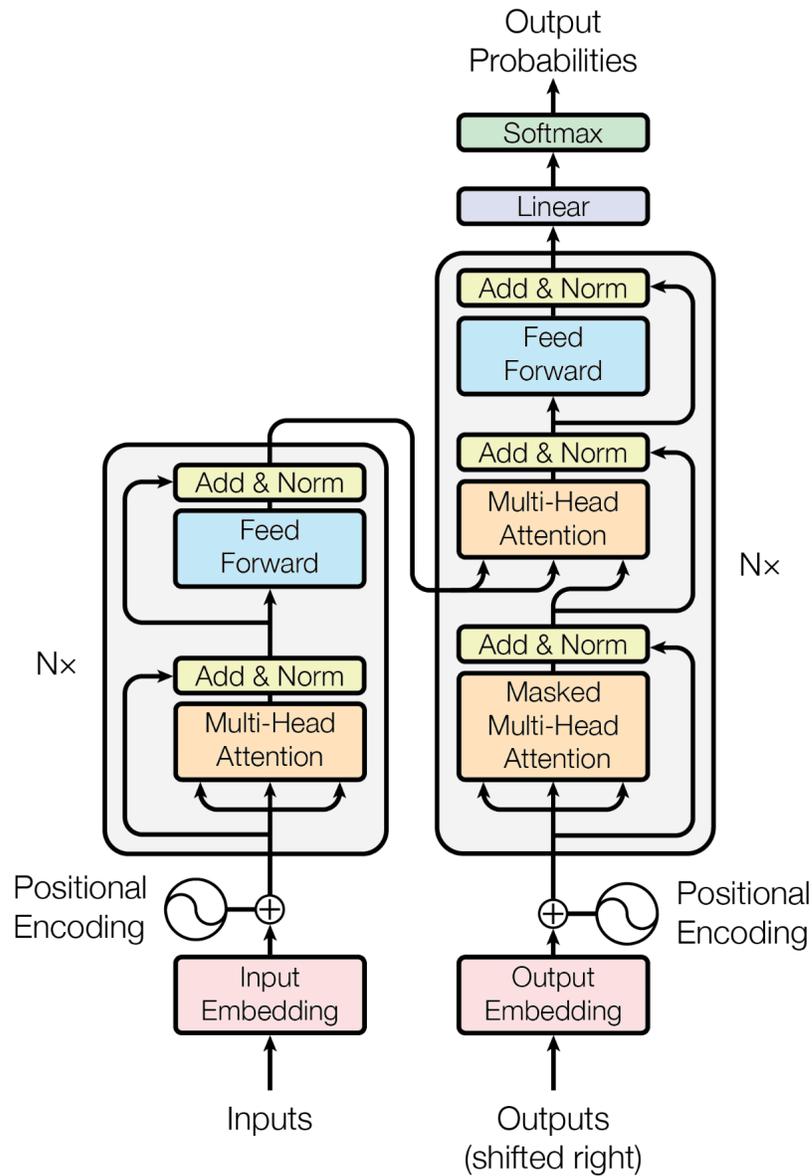


Figura 2.19: Arquitectura del modelo Transformer presentada en el paper “Attention is All You Need” [42].

2.6. Métricas de Evaluación de Rendimiento y Precisión de Modelos

Las métricas de evaluación de rendimiento y precisión son fundamentales en el desarrollo y análisis de modelos de Machine Learning. Estas métricas permiten evaluar la calidad de las predicciones realizadas por los modelos y compararlas con los valores reales. A continuación, se detallan algunas de las métricas más utilizadas: MSE , $RMSE$, MAE y R^2 .

2.6.1. Error Cuadrático Medio

El Error Cuadrático Medio (Mean Squared Error, MSE) es una métrica que mide la media de los cuadrados de los errores o desviaciones, es decir, la diferencia entre los valores predichos por el modelo y los valores reales. Matemáticamente, se define como:

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (2.6)$$

Donde:

- y_i son los valores reales.
- \hat{y}_i son los valores predichos.
- n es el número de observaciones [43].

El MSE es ampliamente utilizado debido a su simplicidad y su capacidad para penalizar los errores grandes, lo que lo hace sensible a valores atípicos. Sin embargo, su principal desventaja es que al estar en unidades cuadradas, puede ser difícil de interpretar directamente en el contexto del problema [44].

2.6.2. Raíz del Error Cuadrático Medio

La Raíz del Error Cuadrático Medio (Root Mean Squared Error, $RMSE$) es la raíz cuadrada del MSE . Proporciona una medida de error en las mismas unidades que las observaciones originales, lo que facilita su interpretación. Matemáticamente, se define como:

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (2.7)$$

El $RMSE$ es útil para comparar la calidad de las predicciones en términos del mismo rango de valores que los datos originales. Al igual que el MSE , es sensible a valores atípicos [45].

2.6.3. Error Absoluto Medio

El Error Absoluto Medio (Mean Absolute Error, MAE) mide la media de las diferencias absolutas entre los valores predichos y los valores reales. A diferencia del MSE y $RMSE$, el MAE no penaliza los errores grandes de manera cuadrática, lo que lo hace más robusto frente a valores atípicos. Matemáticamente, se define como:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (2.8)$$

El MAE es intuitivo y fácil de interpretar, ya que representa el error medio en las mismas unidades que los datos originales. Sin embargo, puede no ser tan útil como el MSE o $RMSE$ para problemas donde los errores grandes son particularmente indeseables [46].

2.6.4. Coeficiente de Determinación

El Coeficiente de Determinación (R-squared, R^2) mide la proporción de la varianza en la variable dependiente que es predecible a partir de las variables independientes. Representa

la calidad del ajuste del modelo. Matemáticamente, se define como:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (2.9)$$

Donde \bar{y} es la media de los valores reales. Un valor de R^2 cercano a 1 indica que el modelo explica bien la variabilidad de los datos, mientras que un valor cercano a 0 indica que el modelo no explica la variabilidad de los datos mejor que una simple media [47].

Capítulo 3

Estado del Arte

3.1. Herramientas de Simulación de Sistemas Fotovoltaicos

Los avances en la simulación de sistemas fotovoltaicos han permitido a los ingenieros y diseñadores optimizar el rendimiento de las instalaciones solares de manera más precisa y eficiente. Entre las herramientas más utilizadas en la industria se encuentran HelioScope, Solcast Rooftop PV Model y PVSyst, cada una con sus características específicas para diferentes tipos de proyectos. A continuación, se presentan sus principales usos, características y una comparación entre ellas.

3.1.1. *HelioScope*

HelioScope es una herramienta de software avanzada diseñada para la planificación, diseño y optimización de sistemas solares fotovoltaicos (fotovoltaicos). Desarrollada inicialmente por Folsom Labs y ahora parte de Aurora Solar, *HelioScope* integra capacidades avanzadas de diseño y simulación solar, convirtiéndose en una herramienta esencial para ingenieros solares, diseñadores y equipos de ventas. Su capacidad para agilizar el proceso de diseño solar, permitiendo iteraciones rápidas y análisis detallados, la hace muy valorada en la industria [48].

3.1.1.1. Principales Usos

HelioScope se utiliza principalmente para diseñar y simular sistemas fotovoltaicos, permitiendo a los usuarios crear modelos precisos de instalaciones solares. Entre sus principales usos se incluyen:

- **Diseño del sistema:** Permite a los usuarios crear diseños detallados de sistemas fotovoltaicos, incluyendo la disposición de los paneles, la orientación y el ángulo de inclinación.
- **Simulación de rendimiento:** Ofrece simulaciones precisas del rendimiento del sistema bajo diferentes condiciones ambientales y de operación.
- **Análisis de sombras:** Proporciona análisis detallados de sombras, permitiendo a los diseñadores evaluar el impacto de los obstáculos y optimizar la disposición de los paneles [49].

- Optimización del sistema: Ayuda a los usuarios a maximizar la producción de energía y reducir costos optimizando el diseño del sistema [50].

3.1.1.2. Características Principales

HelioScope se destaca por una serie de características que lo hacen una herramienta poderosa para el diseño de sistemas fotovoltaicos:

- Diseño basado en web: *HelioScope* es una plataforma basada en la web, lo que significa que no requiere instalación de software y puede ser accesible desde cualquier dispositivo con conexión a internet [48].
- Análisis de sombras en 3D: Proporciona análisis de sombras en 3D para evaluar el impacto de los obstáculos y optimizar la disposición de los paneles.
- Integración con bases de datos meteorológicas: Utiliza datos meteorológicos de alta resolución para mejorar la precisión de las simulaciones de rendimiento.
- Herramientas de optimización: Ofrece herramientas de optimización que permiten a los diseñadores ajustar la configuración del sistema para maximizar la producción de energía y reducir costos [51].
- Reportes detallados: Genera reportes detallados que incluyen análisis de rendimiento, análisis económico y evaluaciones de sombra, que pueden ser utilizados para la planificación y presentación de proyectos [52].

3.1.2. *Solcast Rooftop PV Model*

El modelo *Rooftop PV Model* de *Solcast* es una herramienta empírica diseñada para modelar la producción de sistemas fotovoltaicos instalados en techos. Este modelo es particularmente útil para estimar la producción de flotas de sistemas fotovoltaicos donde las especificaciones del sistema son limitadas o desconocidas. *Solcast* utiliza este modelo en su producto de agregaciones de red detrás del medidor, que proporciona estimaciones de producción y datos de pronóstico para activos de red o regiones de cualquier tamaño, desde vecindarios hasta escalas continentales [53].

3.1.2.1. Principales Usos

El *Rooftop PV Model* se utiliza para proporcionar datos de salida de energía fotovoltaicos para sistemas instalados en techos. Está diseñado para manejar situaciones donde las especificaciones del sistema, como el azimut y la inclinación, la capacidad en corriente alterna (AC) y un factor de pérdida global, son limitadas. Esto permite a los usuarios realizar solicitudes a la API simplemente refiriéndose a la latitud, longitud y capacidad del sistema [53]. Los principales usos incluyen:

- Predicción de producción: Proporciona pronósticos de la producción de energía solar a corto y largo plazo, basados en datos meteorológicos y de irradiancia solar.
- Monitoreo en tiempo real: Permite a los usuarios monitorear el rendimiento de sus sistemas solares en tiempo real, identificando rápidamente cualquier problema o desviación del rendimiento esperado [54].

- Optimización del diseño: Ayuda a los diseñadores de sistemas fotovoltaicos a optimizar la disposición y configuración de los paneles solares para maximizar la producción de energía [55].

3.1.2.2. Características Principales

Algunas de las características más destacadas del *Rooftop PV Model* incluyen:

- Uso de datos de irradiancia y clima de *Solcast*: Los parámetros de irradiancia incluyen la irradiancia directa normal (DNI) y la irradiancia difusa horizontal (DHI), así como la temperatura del aire y una pérdida específica del sitio.
- API accesible: Permite realizar solicitudes a la API sin necesidad de crear sitios específicos, facilitando su uso.
- Validación global: El modelo ha sido validado en colaboración con operadores y empresas en múltiples países, garantizando su fiabilidad y precisión.
- Monitoreo y alertas en tiempo real: Proporciona herramientas de monitoreo en tiempo real y alertas para notificar a los usuarios sobre cualquier problema o desviación en el rendimiento del sistema.
- Informes detallados: Genera informes detallados sobre el rendimiento del sistema, incluyendo análisis económicos y de eficiencia, que pueden ser utilizados para la toma de decisiones y la presentación de resultados [53].

3.1.3. *PVSyst*

PVSyst es un software integral de simulación y diseño para sistemas fotovoltaicos. Desarrollado en Suiza por el Dr. André Mermoud en la Universidad de Ginebra, *PVSyst* se ha convertido en una herramienta esencial para ingenieros, arquitectos, investigadores y profesionales en la industria solar. El software permite realizar estudios detallados, dimensionamiento de sistemas y simulaciones horarias, presentando los resultados en informes completos y gráficos específicos [56].

3.1.3.1. Principales Usos

PVSyst se utiliza principalmente para evaluar y optimizar el diseño de sistemas fotovoltaicos de todo tipo, desde pequeñas instalaciones residenciales hasta grandes plantas solares comerciales e industriales. Entre sus principales usos se incluyen:

- Diseño del sistema: Permite el diseño detallado de sistemas fotovoltaicos, incluyendo la selección de componentes, la configuración del sistema y el análisis de sombreado [57].
- Simulación de rendimiento: Ofrece simulaciones precisas del rendimiento energético del sistema bajo diferentes condiciones meteorológicas y operativas [58].
- Análisis financiero: Proporciona herramientas para evaluar la viabilidad económica de los proyectos fotovoltaicos, incluyendo cálculos de retorno de inversión y análisis de costos [59].
- Investigación y desarrollo: Utilizado por investigadores y desarrolladores para modelar nuevos diseños de sistemas y tecnologías fotovoltaicas [60].

3.1.3.2. Características Principales

PVSyst se destaca por una serie de características que lo hacen una herramienta poderosa para el diseño y simulación de sistemas fotovoltaicos:

- Diseño de sistemas: Permite la creación de diseños detallados de sistemas fotovoltaicos, considerando múltiples orientaciones y tipos de seguimiento solar.
- Simulaciones detalladas: Realiza simulaciones horarias que incluyen el cálculo de pérdidas por sombra, eficiencia del inversor y otros factores que afectan la producción de energía.
- Base de datos de componentes: Incluye una amplia base de datos de módulos fotovoltaicos, inversores y otros componentes del sistema, facilitando la selección y comparación de componentes.
- Análisis económico: Proporciona herramientas para el análisis financiero de los proyectos, incluyendo cálculos de retorno de inversión, análisis de costos y evaluaciones de viabilidad económica.
- Reportes completos: Genera reportes detallados que incluyen análisis de rendimiento, análisis económico y evaluaciones de sombra, que pueden ser utilizados para la planificación y presentación de proyectos.

3.1.4. Diferencias y Similitudes

HelioScope es conocido por su capacidad de realizar simulaciones rápidas y detalladas, utilizando una interfaz basada en la web que facilita el diseño y modelado de sistemas fotovoltaicos complejos, especialmente en techos. Su enfoque en la facilidad de uso y la optimización del diseño lo hace especialmente efectivo en el análisis de sombras y la configuración de sistemas en entornos complicados [49].

Por otro lado, *PVSyst* es reconocido por su precisión en la simulación del rendimiento de sistemas fotovoltaicos grandes y complejos. Ofrece una gama más amplia de opciones de modelado y permite un análisis detallado de las pérdidas del sistema, incluyendo la degradación de los módulos, las pérdidas en los cables y las pérdidas de conversión DC-AC. Su capacidad para integrar datos meteorológicos detallados y realizar simulaciones a largo plazo lo hace ideal para proyectos de gran escala y estudios de viabilidad [57].

Solcast Rooftop PV Model se enfoca en la predicción y monitoreo del rendimiento de sistemas fotovoltaicos en tejados. Utiliza datos satelitales y modelos meteorológicos avanzados para proporcionar estimaciones precisas de la irradiancia solar y la producción de energía. Es particularmente útil para la predicción de la producción de energía a corto y largo plazo y el monitoreo en tiempo real [54].

La elección entre *HelioScope*, *PVSyst* y *Solcast Rooftop PV Model* dependerá en gran medida del tamaño y la complejidad del proyecto, así como de las necesidades específicas del usuario. *HelioScope* es ideal para proyectos más pequeños y medianos que requieren rapidez y facilidad de uso, siendo particularmente efectivo en el análisis de sombras y optimización del diseño en techos complejos. *PVSyst* es más adecuado para proyectos grandes y complejos

que requieren simulaciones detalladas y precisas, ofreciendo una amplia gama de opciones de modelado y análisis detallado de pérdidas. *Solcast Rooftop PV Model* es más apropiado para proyectos que requieren una precisión extrema en la predicción de la producción de energía y el monitoreo en tiempo real, utilizando datos satelitales de alta resolución y modelos meteorológicos avanzados.

3.2. Machine Learning

El modelo Transformer es una arquitectura de red neuronal que ha revolucionado el campo del procesamiento del lenguaje natural (NLP) y otras áreas de la inteligencia artificial. Los Transformers se basan en un mecanismo de atención que permite al modelo enfocarse en diferentes partes de la secuencia de entrada con diferentes niveles de intensidad, mejorando significativamente la capacidad del modelo para capturar dependencias a largo plazo en los datos secuenciales [42].

3.2.1. Componentes Principales del Modelo Transformer

3.2.1.1. Mecanismo de Atención

El mecanismo de atención es el núcleo del modelo Transformer. Este permite al modelo asignar diferentes pesos a diferentes partes de la secuencia de entrada, facilitando la captura de dependencias a largo plazo y mejorando el rendimiento en tareas de traducción y generación de texto. La atención puede ser autoadaptativa, lo que significa que el modelo puede aprender qué partes de la entrada son más relevantes para la tarea en cuestión [61].

3.2.1.2. Capas de Codificador y Decodificador

El Transformer consta de una serie de capas de codificador y decodificador. Cada capa de codificador toma una secuencia de entrada y la transforma en una representación interna, mientras que cada capa de decodificador toma esta representación interna y genera una secuencia de salida. El uso de múltiples capas permite al modelo capturar características complejas y abstracciones de alto nivel [62].

3.2.1.3. Positional Encoding

Dado que el Transformer no utiliza recurrencias ni convoluciones, emplea un esquema de codificación posicional para representar la información de la posición relativa de las palabras en la secuencia. Esto permite al modelo mantener el orden de las palabras en la entrada y generar salidas coherentes [63].

3.2.2. Avances y Aplicaciones

3.2.2.1. Modelos de Lenguaje Avanzados

Los Transformers han dado lugar a varios modelos de lenguaje avanzados, como BERT (Bidirectional Encoder Representations from Transformers) y GPT (Generative Pre-trained Transformer). BERT ha demostrado ser extremadamente efectivo en tareas de comprensión del lenguaje, mientras que GPT se ha destacado en la generación de texto coherente y creativo [64].

3.2.2.2. Transferencia de Aprendizaje

Una de las innovaciones más significativas introducidas por los Transformers es la capacidad de transferencia de aprendizaje. Los modelos preentrenados en grandes conjuntos de datos pueden ser finamente ajustados para tareas específicas con una cantidad relativamente pequeña de datos adicionales. Esto ha permitido avances rápidos en diversas aplicaciones de NLP [65].

3.2.2.3. Visión por Computadora y Otras Áreas

Aunque los Transformers fueron inicialmente diseñados para tareas de NLP, su arquitectura ha sido adaptada para la visión por computadora y otras áreas. Por ejemplo, el modelo Vision Transformer (ViT) ha demostrado un rendimiento competitivo en tareas de clasificación de imágenes. Además, los Transformers se están explorando en biología computacional y en la simulación de sistemas físicos [63].

3.2.3. Desafíos y Futuro

A pesar de sus éxitos, los Transformers enfrentan varios desafíos. Uno de los principales es el alto costo computacional asociado con su entrenamiento, especialmente para modelos muy grandes. Además, la interpretabilidad de los modelos Transformer sigue siendo un área activa de investigación.

Capítulo 4

Metodología

El objetivo principal de esta metodología es obtener valores eléctricos de sistemas fotovoltaicos que sean comparables a los reportes obtenidos por herramientas como *HelioScope*, pero utilizando una aproximación programática basada en la librería *PVLIB* de Python y un modelo Transformer. Este enfoque permite la automatización del análisis, facilitando el trabajo con grandes volúmenes de datos de diferentes sistemas fotovoltaicos sin la necesidad de interactuar directamente con interfaces gráficas.

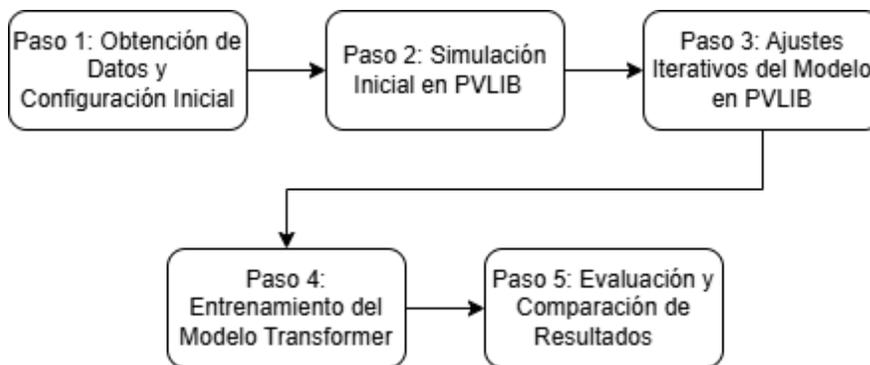


Figura 4.1: Metodología seguida para el desarrollo del proyecto.

A continuación, se describe en detalle cada uno de los pasos llevados a cabo para configurar, simular, y ajustar los modelos utilizados, partiendo desde la obtención de datos hasta la evaluación final de los resultados.

4.1. Obtención de Datos y Configuración Inicial

La primera fase de la metodología consiste en la obtención de los datos necesarios para realizar las simulaciones de los modelos basados en *PVLIB* como de los modelos Transformers. Los datos recopilados incluyen información meteorológica y las características técnicas de los sistemas fotovoltaicos almacenadas en la base de datos de la empresa.

4.1.1. Datos Meteorológicos y del Sistema Fotovoltaico

El primer paso es la recopilación de datos meteorológicos de Typical Meteorological Year (TMY), los cuales incluyen variables como irradiancia solar, temperatura ambiente y velocidad del viento. Estos datos son esenciales para calcular la irradiancia incidente en los módulos solares y modelar correctamente el comportamiento del sistema. La información meteorológica se obtuvo desde el Explorador Solar [1] y se utilizará tanto para simular las plantas en *PVLIB* como para entrenar el modelo Transformer.

4.1.2. Configuración de las Plantas Fotovoltaicas

Cada planta en el portafolio de la empresa se configuró con sus características específicas, incluyendo la latitud, longitud, inclinación, orientación de los paneles solares y la configuración de los inversores. Estas configuraciones son fundamentales para calcular con precisión:

- La geometría solar, que determina la posición del Sol en relación con los paneles.
- El ángulo de incidencia (AOI), que afecta la cantidad de irradiancia que recibe el panel.

Estos datos, junto con las características del sistema (potencia de los módulos, pérdidas por sombreado, etc.), proporcionan la base para las simulaciones en *PVLIB*.

4.2. Simulación Inicial en *PVLIB*

La simulación inicial en *PVLIB* se llevó a cabo utilizando los datos meteorológicos y las configuraciones de las plantas descritas anteriormente. En este punto, se siguen los pasos del diagrama de modelado de rendimiento mostrado en la figura 2.15 de la sección 2.2.5 presentado en el Marco Teórico:

1. Cálculo de la Irradiancia Incidente: Se utilizó el modelo de transposición de irradiancia para convertir la irradiancia global horizontal en irradiancia en el plano de los paneles, teniendo en cuenta su inclinación y orientación.
2. Cálculo de Pérdidas por Suciedad y Otros Factores: Se aplicó la fórmula de $p_{watts\ losses}$ para estimar las pérdidas totales asociadas a diferentes factores, como suciedad, sombreado y desajustes. Este modelo proporciona un porcentaje de pérdidas global que afecta la cantidad de energía utilizable.
3. Cálculo de la Temperatura del Módulo: La temperatura del módulo se estimó utilizando los modelos de *PVLIB*, considerando la irradiancia y la temperatura ambiente.
4. Cálculo de Potencia DC y Conversión a AC: Se calculó la potencia en corriente continua (DC) generada por los módulos y su conversión a corriente alterna (AC), teniendo en cuenta la eficiencia del inversor y las pérdidas en la conversión. Estos cálculos proporcionaron una primera estimación de la energía generada por el sistema.

4.3. Ajustes Iterativos del Modelo en *PVLIB*

Durante las simulaciones iniciales, se observó una diferencia significativa entre los resultados de *PVLIB* y los obtenidos de *HelioScope*, particularmente en la potencia AC. A partir de estas observaciones, se realizaron ajustes iterativos en el modelo, incluyendo:

- Variaciones en los porcentajes de pérdidas de las categorías del modelo PVWatts
- Ajustes en los valores de los coeficientes de temperatura

Cada iteración fue comparada con los resultados obtenidos en *HelioScope* para minimizar la diferencia de potencia AC.

4.4. Entrenamiento del Modelo Transformer

4.4.1. Preparación del Dataset

El siguiente paso fue el desarrollo de un modelo Transformer para la predicción de la potencia AC de las plantas fotovoltaicas. Para ello, se preparó un dataset que incluyó las siguientes variables:

- Datos meteorológicos horarios (irradiancia, temperatura, viento).
- Ángulo de incidencia (AOI) calculado.
- Inclinación y orientación de los paneles solares.
- Latitud y longitud de las plantas fotovoltaicas.

Estos datos fueron normalizados y formateados para que el modelo Transformer pudiera procesarlos eficientemente.

4.4.2. Entrenamiento del Modelo

El modelo Transformer fue configurado con una arquitectura adecuada para la tarea de predicción de potencia AC, y entrenado utilizando el dataset descrito anteriormente. Se empleó un enfoque de early stopping para evitar el sobreajuste, evaluando continuamente el rendimiento del modelo en un conjunto de validación.

4.5. Evaluación y Comparación de Resultados

Finalmente, se evaluaron los resultados obtenidos de las simulaciones en *PVLIB* y las predicciones del modelo Transformer, comparándolos con los datos de referencia generados por *HelioScope*.

Capítulo 5

Desarrollo

5.1. Modelo basado en *PVLIB*

El desarrollo del modelo de simulación basado en *PVLIB* fue un proceso en el cual se diseñó un sistema capaz de replicar el comportamiento de plantas fotovoltaicas, ajustándose a las características específicas de cada planta dentro del portafolio de la empresa. El objetivo fue generar resultados comparables a los obtenidos por *HelioScope*, con un enfoque particular en la simulación de potencia AC y otras métricas clave relacionadas con la eficiencia y el rendimiento de las plantas. A lo largo del desarrollo, se integraron mejoras para optimizar el modelo, mejorar la precisión de los resultados y facilitar su uso a gran escala.

El primer paso consistió en definir el modelo utilizando la información almacenada en la base de datos de la empresa. Cada planta fotovoltaica está identificada por un ID único, y a partir de este se obtienen parámetros clave como la ubicación geográfica (latitud y longitud), la potencia AC total, la distribución de strings, la inclinación de los paneles, el ángulo acimutal, el tipo de montaje, y las características específicas de los módulos e inversores. Estos datos fueron esenciales para la creación del modelo de simulación y garantizaron que este reflejara fielmente las condiciones de operación de cada planta.

Para facilitar el proceso de simulación y su gestión, el código se estructuró en varias clases especializadas, las principales correspondiendo a **PlantModel**, **InverterModel**, **WeatherModel** y **SimulationRunner**. Cada una de estas clases fue diseñada para gestionar diferentes aspectos del proceso de simulación, desde la obtención de datos meteorológicos hasta la simulación final del sistema fotovoltaico. Esta modularidad permite un enfoque más flexible y escalable, lo que facilita la simulación tanto a nivel de inversor como de planta completa.

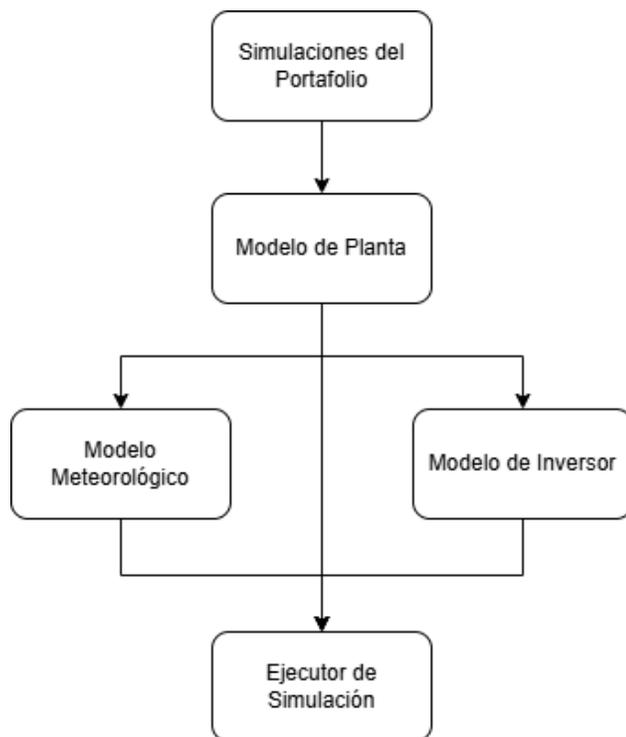


Figura 5.1: Diagrama de funcionamiento de simulaciones usando *PVLIB*.

El diagrama anterior ilustra las interacciones entre las clases principales del modelo. Como se puede observar, el flujo de simulación comienza con la clase **PlantModel**, que se encarga de inicializar las configuraciones de la planta fotovoltaica, tales como su ubicación, configuración de inversores y pérdidas. Esta clase recupera la información necesaria desde la base de datos y utiliza el método `create_model()` para agrupar los inversores y preparar la simulación (ver Anexo A.1). El proceso de simulación propiamente dicho se realiza mediante el método `simulate()`, el cual inicializa la creación de los modelos de inversores y datos meteorológicos, y luego ejecuta la simulación completa con el soporte de las otras clases claves.

Los inversores se modelan utilizando la clase **InverterModel** (ver Anexo A.2). Esta clase se encarga de inicializar cada inversor con los datos técnicos relevantes y asociar los strings correspondientes. El método `create_model()` genera el modelo del inversor mediante la creación de arrays fotovoltaicos y la aplicación de pérdidas. Uno de los elementos más cruciales es la función `calculate_pvwatts_losses()`, que calcula las pérdidas totales utilizando el modelo de *PVWatts*, considerando factores como ensuciamiento, sombreado y pérdidas por cableado, entre otros. Estas pérdidas se aplican para ajustar los resultados de la simulación de manera precisa.

La clase **WeatherModel** es la encargada de obtener los datos meteorológicos que se necesitan para la simulación, como la irradiancia, la temperatura y el viento. Dependiendo de la fuente de datos especificada (por ejemplo, *Explorador Solar* u otras fuentes meteorológicas satelitales), el método `get_weather_df()` se encarga de obtener los datos meteorológicos necesarios y organizarlos en un `DataFrame` adecuado para la simulación (ver Anexo A.3). Estos datos son cruciales para ajustar la producción de energía de la planta en función de las condiciones ambientales. La correcta integración de los datos meteorológicos permite una

simulación precisa del rendimiento de la planta fotovoltaica a lo largo del tiempo.

Finalmente, la simulación se ejecuta completamente a través de la clase `SimulationRunner` (ver Anexo A.4), que recibe los modelos de inversores y el modelo meteorológico para calcular la producción de energía de la planta. El método `run_model()` es responsable de ejecutar los cálculos de potencia para cada inversor, utilizando los datos meteorológicos proporcionados por `WeatherModel`. Los resultados de la simulación se consolidan en un `DataFrame` mediante el método `inverters_results()`, que combina los datos de los inversores con las variables meteorológicas para generar los resultados finales de la simulación. Este enfoque modular permite escalar la simulación a diferentes plantas y configuraciones, proporcionando un análisis detallado y ajustado a cada planta en particular.

Con el fin de facilitar las simulaciones a gran escala, se añadió la capacidad de simular automáticamente las plantas de todo el portafolio de la empresa. Se creó una función que permitía ejecutar la simulación tanto para una planta específica como para todas las plantas, almacenando automáticamente los resultados en la base de datos para su posterior análisis. Esta capacidad automatizada fue crucial para la gestión de un portafolio extenso y permitió obtener resultados detallados de cada planta sin intervención manual.

Una vez implementada la automatización de las simulaciones, el siguiente paso fue realizar múltiples pruebas para optimizar el modelo y ajustar los parámetros de simulación. Se utilizaron técnicas como el análisis de regresión lineal y la minimización del error cuadrático medio (MSE) entre los resultados de *PVLIB* y *HelioScope*. Aunque la regresión lineal no mostró mejoras significativas, la minimización del MSE permitió optimizar los valores de simulación y alcanzar una precisión mayor en términos de generación anual y potencia AC.

La división modular del código permitió una mayor flexibilidad para implementar nuevas funciones y mejorar el modelo sin afectar su funcionamiento global. `PlantModel` actúa como la clase principal que desencadena todas las funciones necesarias para la simulación, llamando a las otras clases según sea necesario. Esto permitió un control más preciso del proceso de simulación y facilitó la incorporación de nuevas características o ajustes en el futuro.

La estructura modular también permitió simplificar el proceso de simulación a través de una única función que permitía ajustar la resolución temporal (horaria, diaria, mensual o anual) de las simulaciones, evitando la necesidad de ejecutar simulaciones separadas para cada escala temporal. Esto mejoró la eficiencia del código y permitió obtener resultados a nivel de inversor o planta completa, según fuera necesario.

Tras corregir los errores iniciales y ajustar el modelo para obtener resultados más precisos, el modelo basado en *PVLIB* fue capaz de replicar el comportamiento de las plantas fotovoltaicas bajo diversas condiciones operativas. Su capacidad para manejar grandes volúmenes de datos y adaptarse a las características específicas de cada planta en el portafolio de la empresa lo convierten en una herramienta poderosa y flexible para la simulación de sistemas fotovoltaicos.

5.2. Modelo Transformer

Para este método de solución del problema, se procedió a entrenar una red Transformer “Vanilla” (se podría llamar de esta forma ya que existen distintas versiones de modelos Transformers, pero para el proyecto se utilizó el modelo original desarrollado en el paper [42]) que fuera capaz de estimar el valor de potencia AC de una planta fotovoltaica teniendo como valores de entrada los datos meteorológicos medidos y algunas características de la planta. Notar que en este caso, el modelo no tiene ningún “conocimiento” ni ecuación para calcular potencia AC de una planta, si no que la idea es que aprenda de forma autónoma cómo se realizan estos cálculos para obtener un resultado lo más similar al esperado.

Para entrenar el modelo se comenzó armando el dataset a utilizar. Este cuenta con los datos meteorológicos almacenados en la base de datos (y que fueron utilizados para realizar las simulaciones en *HelioScope*), ángulo de incidencia del sol calculados por *HelioScope*, características generales de la planta tal como promedio de inclinación de los paneles, promedio de acimutal de la planta, potencia de los paneles de la planta, tipo de montaje, orientación y porcentajes de pérdidas. Finalmente, el valor objetivo a estimar corresponde a la potencia AC calculada por *HelioScope* para los mismos datos meteorológicos.

- GHI
- DNI
- DHI
- Temperatura [°C]
- Velocidad del viento
- Altitud del Sol [°]
- Acimutal del Sol [°]
- Incidencia del Sol [°]
- Latitud
- Longitud
- Inclinación de paneles [°]
- Acimutal de paneles [°]
- Tipo de montaje
- Soiling [%]
- Sombra [%]
- Mismatch [%]
- Wiring [%]
- Connections [%]
- LID [%]
- Disponibilidad [%]
- Nameplate Rating [%]
- Potencia peak de planta [W]
- Potencia peak de panel [W]
- Potencia AC [W]
- Tipo de orientación

Dado que el objetivo de este modelo es estimar potencia AC según los datos meteorológicos de entrada, y no pronosticar la potencia AC a futuro según los datos meteorológicos actuales, no se incluyó la columna de “fecha y hora” pues principalmente no se buscaba que el modelo aprendiera alguna relación con la fecha, si no que se enfocara principalmente en los datos meteorológicos.

El entrenamiento del modelo se realizó en *Google Colab* con una tarjeta gráfica *NVIDIA T4* de 16GB y el dataset se dividió en los conjuntos de entrenamiento, validación y prueba en las proporciones 80 %, 10 % y 10 % respectivamente. Para todas las pruebas se utilizaron los mismos datos en cada conjunto, en donde el conjunto de prueba fue aleatorizado para que el modelo no aprendiera patrones agrupados por planta fotovoltaica.

Dado que los resultados obtenidos dependen directamente de las características del modelo (hiper-parámetros), se inició entrenando 3 modelos con distintos sets de hiper-parámetros:

Tabla 5.1: Hiper-parámetros usados en los primeros 3 modelos Transformers entrenados.

	Modelo Pequeño	Modelo Mediano	Modelo Grande
Dimensión	128	256	512
Cabezas	4	8	16
Capas	2	4	6
Retroalimentación	512	1024	2048
dropout	0.1	0.2	0.3
Tasa de aprendizaje	1e-3	1e-4	5e-4
Batch size	128	64	32
Función de Pérdida	MSELoss	MSELoss	MSELoss
Optimizador	Adam	Adam	Adam

Notar que las características ‘Tasa de aprendizaje’ y ‘Batch size’ no corresponden al modelo precisamente, si no que al proceso de entrenamiento y dataloader respectivamente, sin embargo, estos valores afectan de igual forma los resultados obtenidos, y se probaron distintas combinaciones de estos para obtener los mejores resultados.

Dado que el portafolio de plantas de la empresa presenta una amplia variación en la potencia peak, que oscila desde unas pocas decenas de kWp hasta 2 MWp, la potencia AC objetivo fue normalizada en función de la capacidad de cada planta. El resto de las características del dataset se escalaron de manera estándar, restando la media y dividiendo por la desviación estándar, utilizando la función `StandardScaler` de la librería `sklearn` [66].

Tras entrenar estos 3 modelos, y analizar los resultados obtenidos con cada uno de estos, se notó que los mejores resultados se obtuvieron con los modelos “Pequeño” y “Mediano”, por lo que se decidió entrenar un nuevo modelo “Híbrido” que tuviera una mezcla de los hiper-parámetros de estos 2 modelos.

Tabla 5.2: Hiper-parámetros usados en el modelo Transformer “Híbrido”.

	Modelo Híbrido
Dimensión	128
Cabezas	8
Capas	4
Retroalimentación	512
dropout	0.1
Tasa de aprendizaje	1e-4
Batch size	128
Función de Pérdida	MSELoss
Optimizador	Adam

Con este modelo “Híbrido” se observó una leve mejora en los resultados en comparación con los modelos “Pequeño” y “Mediano”, por lo que se consideró que este set de hiper-parámetros era óptimo.

Luego, se volvió a entrenar un nuevo modelo “Híbrido” pero con una diferencia en el dataset entregado, esta vez se agregó la hora del día. El objetivo de esto era que el modelo aprendiera a relacionar las horas de “poco sol” (amanecer y atardecer) con una baja potencia AC.

Capítulo 6

Resultados

6.1. PVLIB

Para las pruebas realizadas en este proyecto, se consideró un total de 142 plantas del portafolio de Solarity, que a la fecha de este informe cuenta con más de 170 plantas en operación. Las plantas evaluadas corresponden a aquellas que entraron en operación hasta antes del año 2024. Los resultados obtenidos mediante *PVLIB* se centraron principalmente en la estimación de la generación anual de las plantas, dado que este valor es crucial para las métricas de desempeño que utiliza la empresa. Sin embargo, el modelo también entrega un conjunto completo de resultados que pueden desglosarse por inversor, o consolidarse a nivel planta para generar reportes a modo de “Digital Twin” de cada una de las plantas fotovoltaicas en el portafolio de la empresa.

En los primeros experimentos realizados con el modelo original proporcionado por la empresa, se calculó el porcentaje de error simple en relación con los resultados de *HelioScope*, obteniendo los siguientes valores:

Tabla 6.1: Porcentaje de error promedio calculado por el modelo *PVLIB* original entregado por la empresa.

Promedio de error [%]	Promedio de error absoluto [%]
7.13	7.52

Como se observa, el modelo subestimaba la generación anual con un promedio de error simple del 7.13 %, es decir, el modelo calculaba aproximadamente el 93 % de la generación teórica estimada por *HelioScope*. Este error resultaba significativo, sobre todo al aumentar la capacidad de las plantas, lo que se reflejaba en la discrepancia entre la potencia real y la simulada.

Para visualizar esta discrepancia, se graficaron los resultados de ambos modelos, como se muestra a continuación:

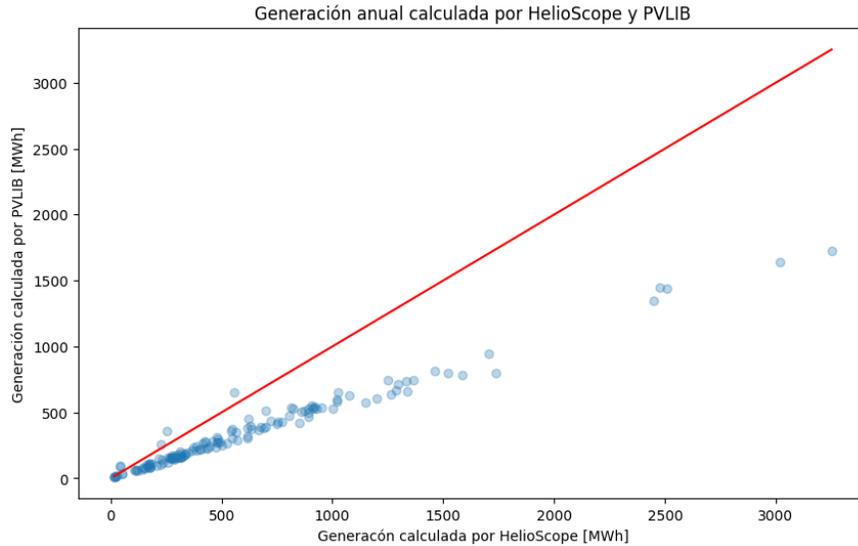


Figura 6.1: Comparación de la generación anual calculada por *HelioScope* y *PVLIB* original.

En este gráfico, se puede observar que el modelo *PVLIB* subestimaba la generación anual de las plantas, particularmente a medida que la capacidad instalada aumentaba, lo que provocaba una mayor discrepancia con respecto a los valores calculados por *HelioScope*. Esto pone de manifiesto la necesidad de ajustar el modelo para mejorar su precisión en la predicción de generación anual.

Además, se generó un histograma para visualizar la distribución del error de cálculo de la generación anual respecto al modelo de referencia:

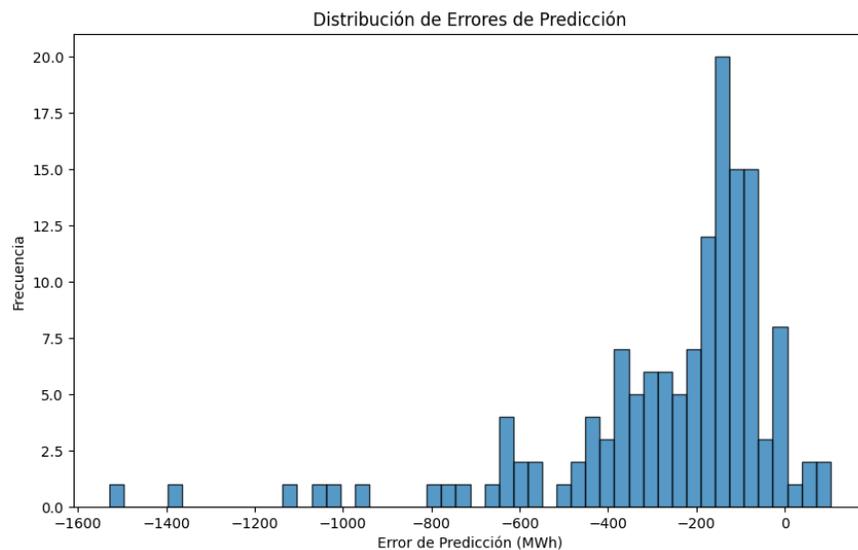


Figura 6.2: Histograma de distribución de error en el cálculo de generación anual del modelo *PVLIB* original respecto a *HelioScope*.

La distribución muestra que los errores se concentran mayoritariamente en torno al 7%

de subestimación, con algunos valores dispersos que reflejan un desempeño inconsistente del modelo original para distintas plantas.

6.1.1. Resultados de la versión final del modelo *PVLIB*

Tras varias iteraciones y ajustes en el modelo, los resultados obtenidos con la versión final del modelo *PVLIB* mejoraron significativamente en comparación con los del modelo original:

Tabla 6.2: Porcentaje de error promedio calculado por *PVLIB* respecto a *HelioScope*.

Promedio de error [%]	Promedio de error absoluto [%]
-0.7872	3.0920

El porcentaje de error promedio de -0.78% indica que el modelo *PVLIB* subestima la generación anual de manera leve respecto a *HelioScope*. Sin embargo, el valor absoluto de error promedio del 3.09% señala una mejora considerable en la precisión, alcanzando un 96.91% de precisión en comparación con el modelo de referencia. Esto marca un avance significativo respecto al error inicial del 7.13% .

En el gráfico siguiente se comparan los resultados de generación anual obtenidos por ambos modelos:

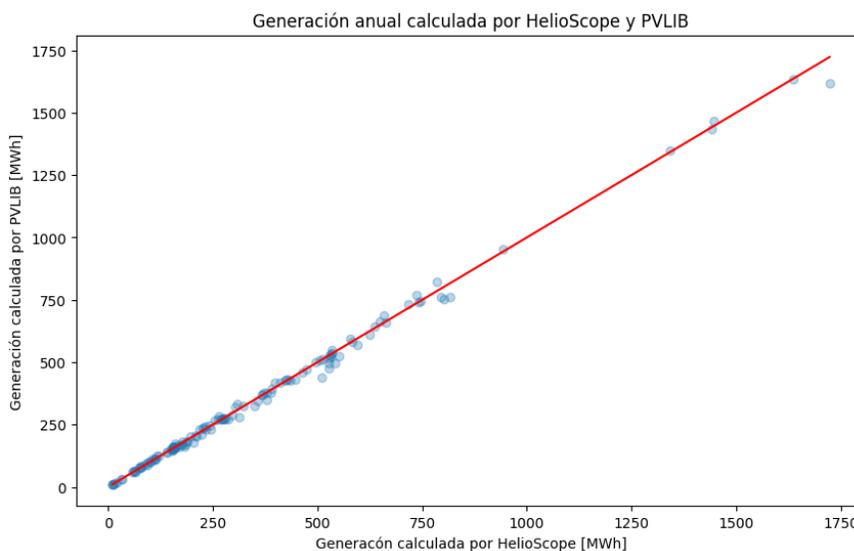


Figura 6.3: Comparación de la generación anual calculada por *HelioScope* y *PVLIB*.

En este gráfico, los puntos azules representan los valores calculados por *PVLIB*, mientras que la línea roja actúa como referencia, mostrando los valores calculados por *HelioScope*. La cercanía de los puntos a la línea indica una reducción significativa del error en la estimación de la generación anual, a excepción de un outlier en torno a los $1750[MWh]$.

La mejora en la precisión del modelo también se refleja en la distribución del error, la cual se muestra a continuación:

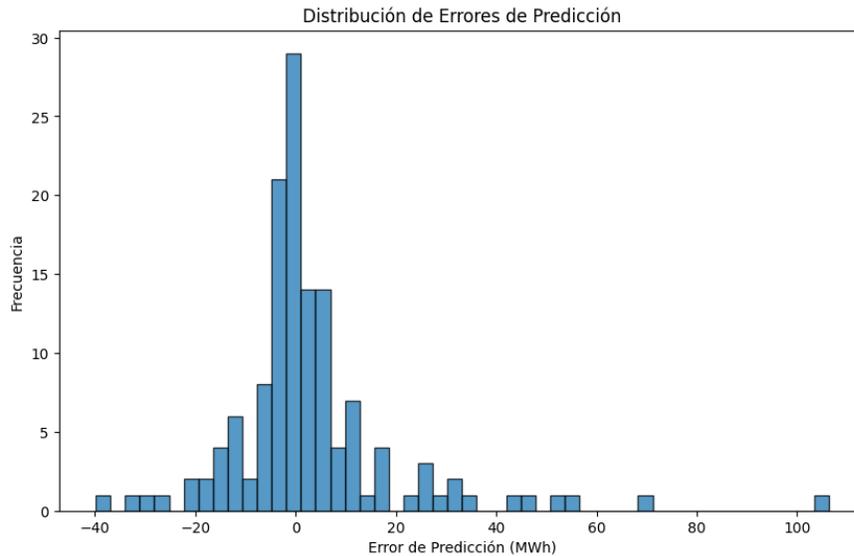


Figura 6.4: Histograma de distribución de error en el cálculo de generación anual del modelo *PVLIB* respecto a *HelioScope*.

En el histograma se observa que la mayoría de los errores de cálculo se concentran entre $-10[MWh]$ y $10[MWh]$, lo que indica una clara mejora en la precisión del modelo, con errores mucho menores que los observados en el modelo original.

Finalmente, se realizaron simulaciones utilizando los mejores parámetros de optimización para el portafolio de la empresa, obteniendo los siguientes resultados:

Tabla 6.3: Porcentaje de error promedio calculado por *PVLIB* tras la optimización de parámetros.

Promedio de error [%]	Promedio de error absoluto [%]
1.4141	2.7805

A pesar de que el promedio de error se invirtió a una sobreestimación de 1.41 %, la mejora en el error absoluto a 2.78 % indica una mayor precisión en el modelo. Sin embargo, la diferencia respecto a la versión sin parámetros optimizados es mínima, con solo una mejora del 0.3 %, lo que cuestiona la viabilidad de esta optimización debido al alto costo computacional asociado.

El siguiente gráfico muestra los resultados tras la optimización:

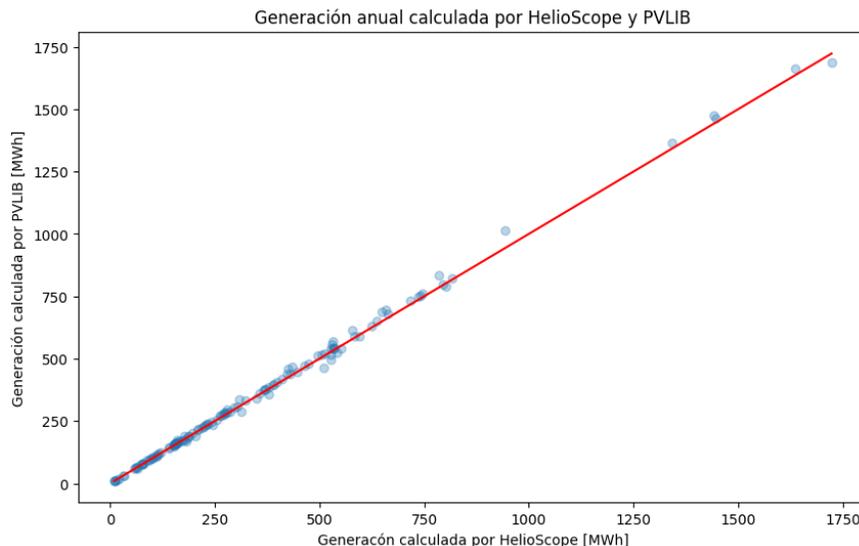


Figura 6.5: Comparación de la generación anual calculada por *HelioScope* y *PVLIB* tras la optimización de parámetros.

Finalmente, las métricas estadísticas $RMSE$ y MAE también se observaron en las tres versiones del modelo:

Tabla 6.4: Métricas $RMSE$ y MAE para el modelo *PVLIB* en sus distintas versiones.

	<i>PVLIB</i> original	<i>PVLIB</i> final	<i>PVLIB</i> optimizado
$RMSE$	61.3154[MWh]	17.9542[MWh]	15.1127[MWh]
MAE	52.0286[MWh]	10.1750[MWh]	9.7382[MWh]

En resumen, las mejoras en el modelo *PVLIB* desde su versión inicial a su versión optimizada son notables, especialmente en términos de precisión, lo que permite reducir el error absoluto a solo unas pocas decenas de MWh al año, mejorando drásticamente el rendimiento del modelo en comparación con su estado original.

6.2. Modelos Transformers

En esta sección, se presenta un análisis detallado de los modelos Transformers entrenados para predecir la potencia AC de plantas fotovoltaicas. Los modelos se comparan en función de su rendimiento en distintas métricas de error, como $RMSE$, MAE y R^2 , así como su comportamiento durante el proceso de entrenamiento. Se incluyen cinco modelos principales: “Pequeño”, “Mediano”, “Grande”, “Híbrido” y “Híbrido considerando horas de Sol”. A lo largo de esta sección, se busca resaltar las fortalezas y debilidades de cada modelo, con un enfoque en cómo cada ajuste y modificación afecta el desempeño general.

6.2.1. Modelos “Pequeño”, “Mediano” y “Grande”

Se comenzó realizando pruebas con estos 3 tamaños de modelos Transformers con el objetivo de identificar los mejores hiperparámetros para el problema a resolver. El modelo “Pequeño” se entrenó durante 101 épocas por un tiempo de 2 horas y 10 minutos, el modelo “Mediano” se entrenó durante 56 épocas por un tiempo de 3 horas y 20 minutos, y finalmente, el modelo “Grande” se entrenó durante 32 épocas por un tiempo de 6 horas y 10 minutos.

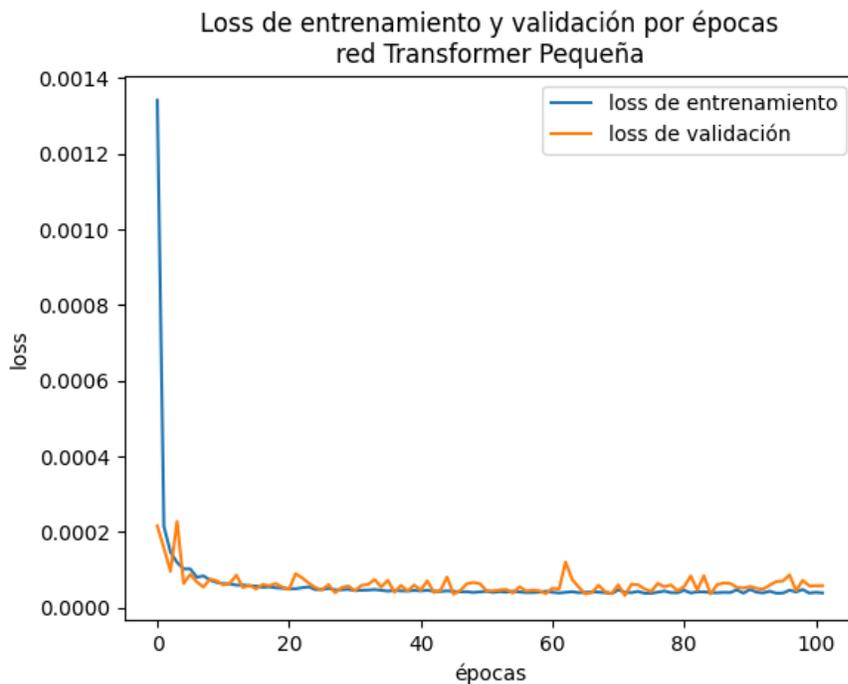


Figura 6.6: Curvas de pérdidas de entrenamiento y validación de modelo Transformer “Pequeño”.

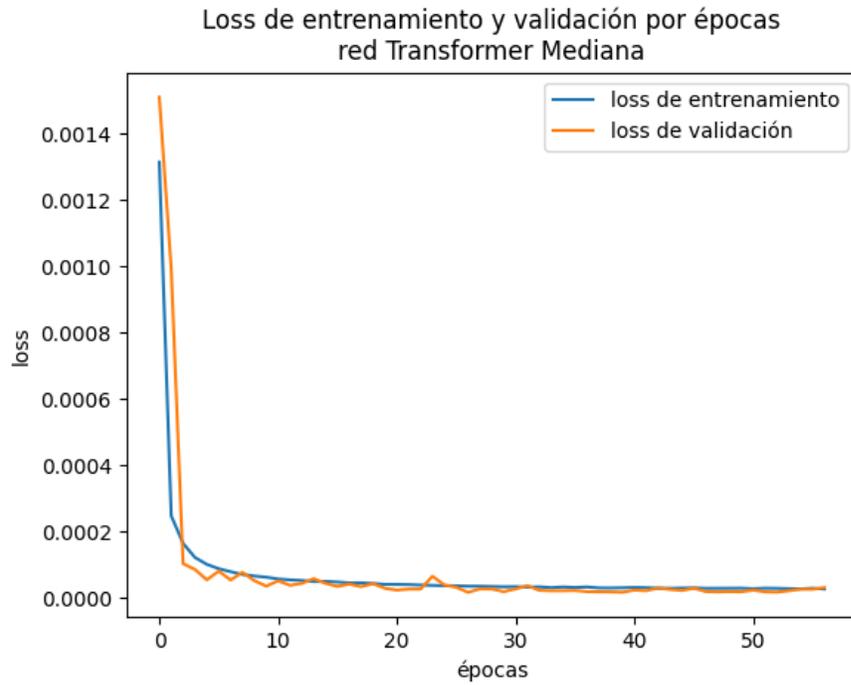


Figura 6.7: Curvas de pérdidas de entrenamiento y validación de modelo Transformer “Mediano”.

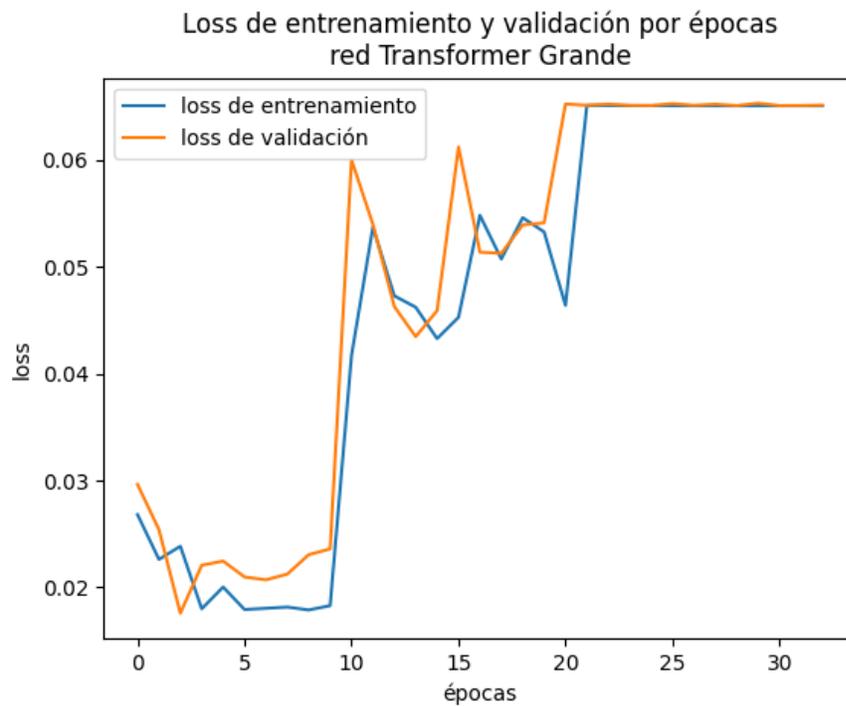


Figura 6.8: Curvas de pérdidas de entrenamiento y validación de modelo Transformer “Grande”.

El análisis de las curvas de entrenamiento revela comportamientos distintos entre los modelos. El modelo “Pequeño” alcanza un plateau poco antes de la época 20, pero la presencia

de ‘peaks’ en la curva de pérdidas del conjunto de validación sugiere un sobreajuste, donde el modelo ajusta excesivamente los datos de entrenamiento, aumentando la variabilidad en las pérdidas de validación. Por otro lado, el modelo “Mediano” también alcanza un plateau alrededor de la época 20, pero con menos ‘peaks’ en la curva de validación, lo que podría indicar una mejor capacidad de generalización debido a su mayor complejidad, resultando en una menor variabilidad en las pérdidas. En contraste, el modelo “Grande” presenta una curva de entrenamiento errática, con valores de pérdida inestables tanto en entrenamiento como en validación, sin alcanzar un plateau claro, lo que sugiere que el modelo no está aprendiendo de manera estable y podría estar experimentando sobreajuste, posiblemente debido a un ajuste inadecuado de la tasa de aprendizaje.

Al momento de evaluar los modelos en el conjunto de prueba, se obtuvieron las siguientes métricas:

Tabla 6.5: Métricas obtenidas al evaluar los modelos “Pequeño”, “Mediano” y “Grande” en el conjunto de prueba.

Métrica	Modelo “Pequeño”	Modelo “Mediano”	Modelo “Grande”
MSE	0	0	0.0175
RMSE	0.0058	0.0045	0.1325
MAE	0.0027	0.0018	0.0847
R^2	0.9995	0.9997	0.7308

Los valores mostrados en la tabla 6.5 permiten analizar el rendimiento de los tres modelos evaluados: “Pequeño”, “Mediano” y “Grande”. En el caso de los modelos “Pequeño” y “Mediano”, las métricas MSE , $RMSE$ y MAE son bajas, con el MSE igual a 0 y $RMSE$ y MAE muy cercanos a 0, lo que indica que ambos modelos realizan predicciones con alta precisión y pequeños errores. Además, el coeficiente de determinación R^2 es prácticamente 1 para ambos modelos (0.9995 y 0.9997, respectivamente), lo que significa que estos modelos explican casi toda la variabilidad en los datos, sugiriendo un ajuste adecuado a los datos y una capacidad sólida para generalizar. En contraste, el modelo “Grande” muestra métricas menos favorables; un MSE más elevado sugiere mayores errores en las predicciones, mientras que los valores más altos de $RMSE$ y MAE reflejan una menor precisión, con desviaciones más significativas respecto a los valores reales. El R^2 de 0.7308 indica que el modelo “Grande” explica una menor proporción de la variabilidad en los datos, lo que sugiere problemas de ajuste y generalización en comparación con los otros dos modelos.

Es importante destacar que los modelos fueron entrenados con valores de potencia AC normalizados por la potencia peak de la planta. Esta normalización implica que los valores de potencia AC se encuentran entre 0 y 1, sin unidad de medida, ya que resultan de la división de la potencia en $[W]$ por la potencia peak de la planta (también en $[W]$). En otras palabras, la potencia AC se expresa como una fracción de la capacidad máxima de la planta.

Debido a esta normalización, los valores de las métricas calculadas también reflejan esta escala de 0 a 1. Por lo tanto, aunque estas métricas cuentan con unidades físicas, proporcionan una medida relativa del error del modelo en términos de la capacidad máxima de la

planta. Esto permite evaluar el modelo en una escala uniforme, independientemente de la capacidad de cada planta en el conjunto de datos.

Al graficar los resultados predichos por el modelo respecto a los valores reales de potencia del conjunto de prueba, se puede observar que el modelo en general estima de buena forma los valores de potencia AC.

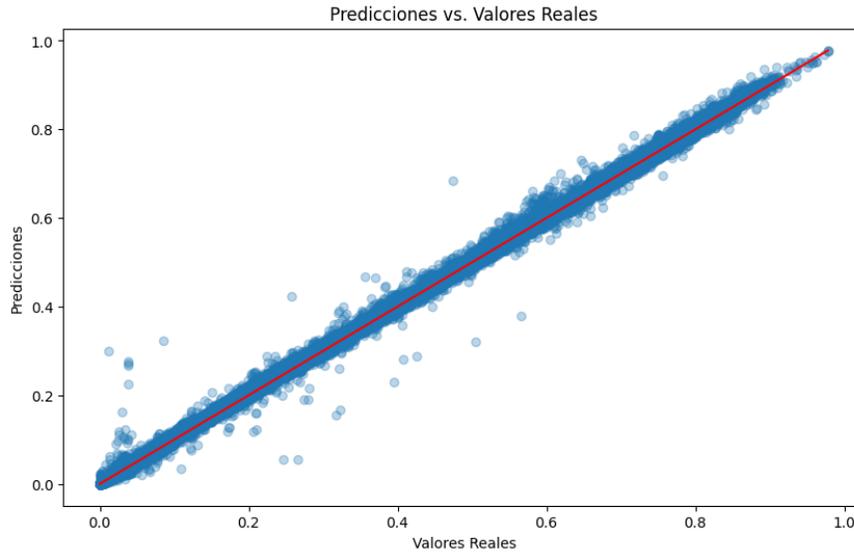


Figura 6.9: Comparación de resultados reales y predichos por el modelo “Pequeño”.

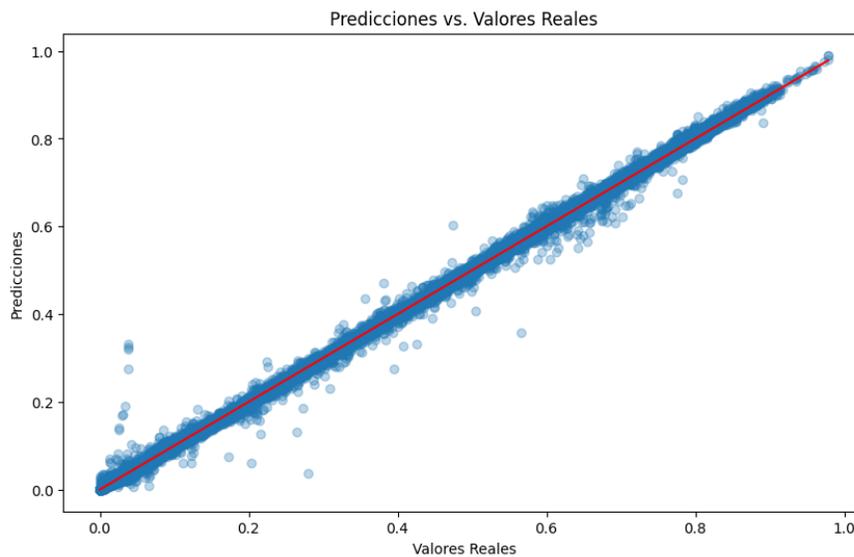


Figura 6.10: Comparación de resultados reales y predichos por el modelo “Mediano”.

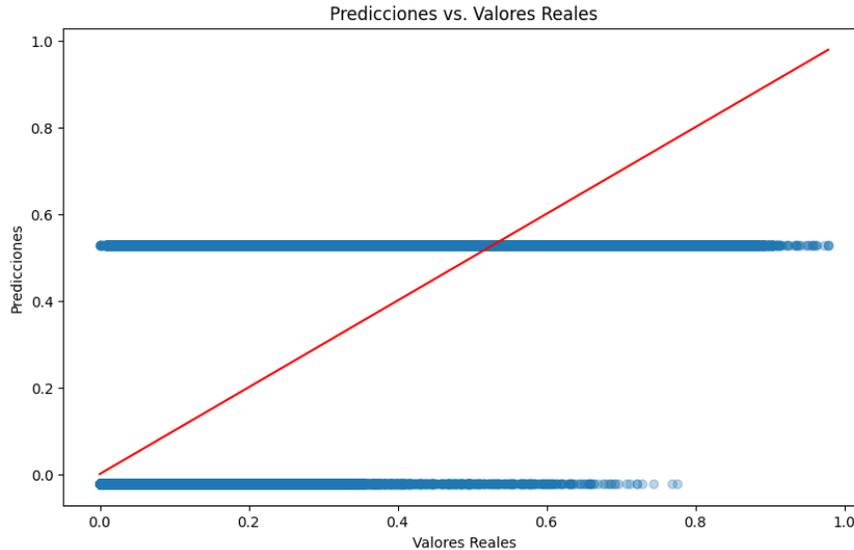


Figura 6.11: Comparación de resultados reales y predichos por el modelo “Grande”.

Se puede observar en la figura 6.9, la línea roja corresponde a los valores reales de potencia AC normalizada por la potencia total de cada planta, lo cual se representa como una diagonal en el gráfico, mientras que los puntos azules corresponden a los valores de potencia AC normalizada que el modelo estima para cada conjunto de datos de entrada..

Los valores predichos por los modelos “Pequeño” y “Mediano” muestran una alta precisión, con predicciones muy cercanas a los valores reales de potencia AC, lo que indica un buen rendimiento general en ambos casos. Sin embargo, se observan algunos outliers donde ambos modelos tienden a sobreestimar la potencia en condiciones de baja irradiación, con valores cercanos a 0. Este fenómeno podría ocurrir principalmente durante el amanecer o atardecer, o en situaciones de invierno o nubosidad intensa. Aunque tanto el modelo “Pequeño” como el “Mediano” presentan esta sobreestimación, la cantidad de outliers es menor en el modelo “Mediano”, lo que sugiere una mejora en su capacidad de generalización y precisión. Además, se detecta una subestimación de la potencia en el rango de 0.2 a 0.6, posiblemente relacionada con las características específicas de los datos de entrada, aunque esta subestimación es menos pronunciada en el modelo “Mediano”, reflejando un rendimiento superior comparado con el modelo “Pequeño”.

Por otro lado, el modelo “Grande” muestra un rendimiento significativamente inferior. El análisis gráfico revela que este modelo no ha logrado aprender la relación entre las características de entrada y la potencia AC, manifestándose en dos líneas de predicciones relativamente continuas: una cercana a 0 y otra alrededor de 0.6. Este patrón sugiere que el modelo ha memorizado ciertos valores del conjunto de entrenamiento, en lugar de generalizar adecuadamente para nuevos datos. La falta de variabilidad en las predicciones y la presencia de errores constantes para un amplio rango de entradas explican los altos valores de $RMSE$ y MAE , así como el bajo R^2 , lo que indica que el modelo “Grande” no logra una predicción precisa y sufre de un sobreajuste significativo.

6.2.2. Modelos “Híbrido” y “Híbrido considerando horas de Sol”

Tras las pruebas realizadas con los tres modelos entrenados utilizando distintos sets de hiperparámetros, se identificó que los mejores resultados se obtenían con los hiperparámetros utilizados en los modelos “Pequeño” y “Mediano”. Con base en esta observación, se procedió a entrenar nuevos modelos que combinan estos hiperparámetros. El segundo modelo entrenado surgió debido a la observación una concentración significativa de valores de potencia AC menores a 200 [kW] en el modelo “Híbrido”, junto con una mayor dispersión de los puntos en la figura 6.16, por lo que se decidió realizar un nuevo entrenamiento incorporando la hora del día al dataset, con la esperanza de que el modelo aprendiera a relacionar las horas de baja irradiación, como el amanecer y el atardecer, con los correspondientes valores bajos de potencia AC.

El modelo “Híbrido” se entrenó durante 106 épocas por un tiempo de 3 horas y 30 minutos mientras que el modelo “Híbrido considerando horas de Sol” se entrenó durante 181 épocas por un tiempo de 3 horas.

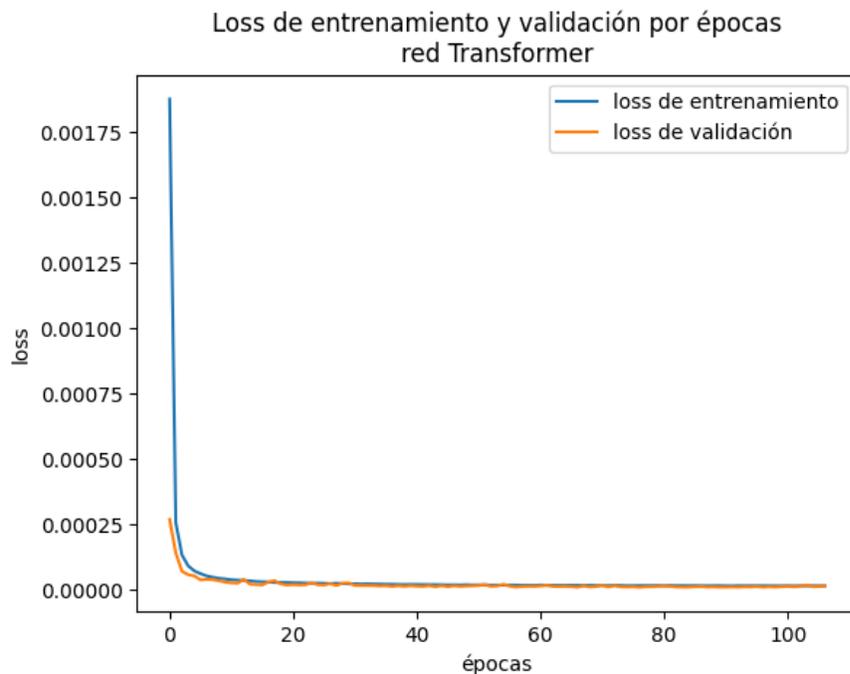


Figura 6.12: Curvas de pérdidas de entrenamiento y validación de modelo Transformer “Híbrido”.

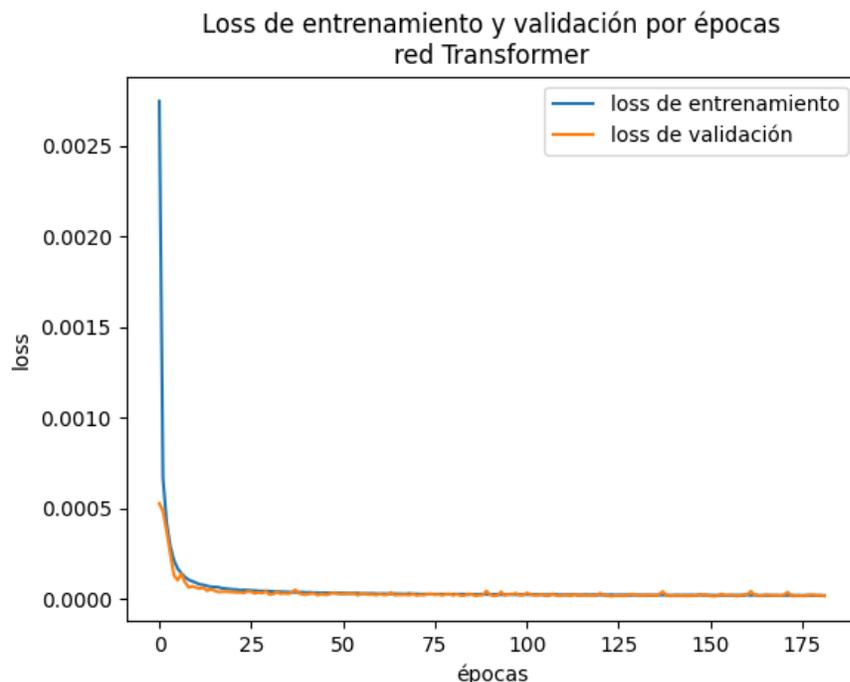


Figura 6.13: Curvas de pérdidas de entrenamiento y validación de modelo Transformer “Híbrido” considerando horas de Sol.

El análisis de las curvas de entrenamiento muestra que el modelo híbrido alcanza el plateau aproximadamente en la época 10, manteniendo una curva de pérdida de validación constante a lo largo del proceso, lo que indica una buena capacidad de generalización sin signos de sobreajuste. De manera similar, el modelo híbrido modificado para considerar las horas del Sol también alcanza su plateau alrededor de la época 25, con una curva de pérdida de validación igualmente estable, lo que sugiere que este modelo también generaliza correctamente, sin presentar indicios de sobreajuste a pesar de la inclusión de la nueva variable temporal.

Al momento de evaluar los modelos en el conjunto de prueba, se obtuvieron las siguientes métricas:

Tabla 6.6: Métricas obtenidas al evaluar los modelos “Híbrido” y “Híbrido” considerando horas de Sol en el conjunto de prueba.

Métrica	Modelo “Híbrido”	Modelo “Híbrido” con horas de Sol
MSE	0	0
RMSE	0.0036	0.0042
MAE	0.0012	0.0020
R^2	0.9998	0.9998

Los valores mostrados en la tabla 6.6 indican que tanto el modelo “Híbrido” como su variante que considera las horas de Sol presentan métricas que reflejan un rendimiento muy bueno. En ambos casos, el MSE es igual a 0, y tanto el $RMSE$ como el MAE son extremadamente bajos, lo que sugiere predicciones muy precisas. Además, el coeficiente de determinación R^2 se aproxima a 1, lo que implica un ajuste casi perfecto a los datos y una

excelente capacidad de generalización. Sin embargo, aunque el modelo que incorpora las horas de Sol mantiene un alto nivel de precisión, sus métricas son ligeramente inferiores a las del modelo “Híbrido” original, lo que indica un desempeño marginalmente menor en comparación con la versión sin la consideración de esta variable temporal.

Al graficar los resultados predichos por el modelo respecto a los valores reales de potencia del conjunto de prueba, se puede observar que el modelo en general estima de buena forma los valores de potencia AC.

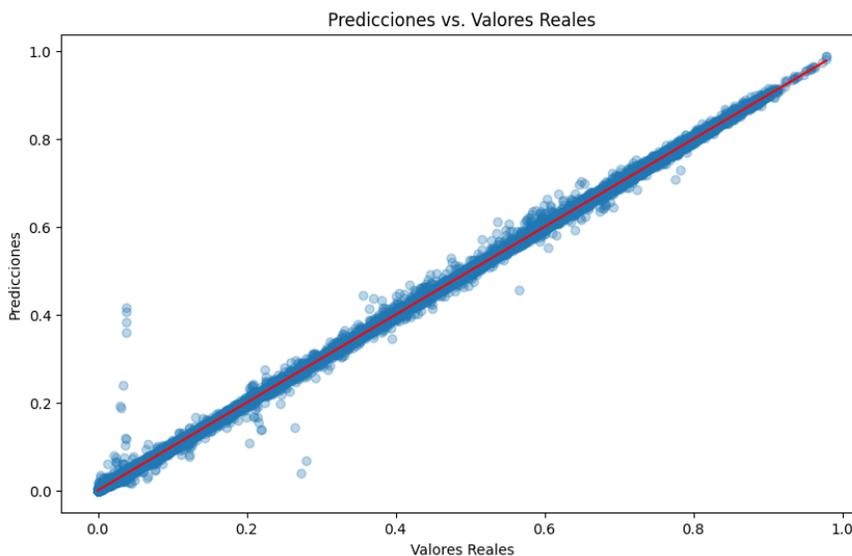


Figura 6.14: Comparación de resultados reales y predichos por el modelo “Híbrido”.

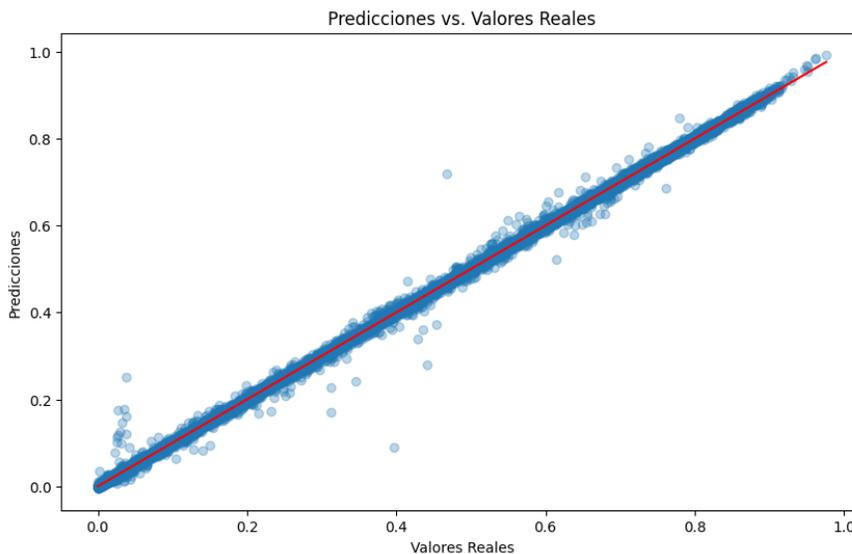


Figura 6.15: Comparación de resultados reales y predichos por el modelo “Híbrido” considerando horas de Sol.

Como se observa en las figuras 6.14 y 6.15, los valores predichos por ambos modelos

“Híbridos” son en general muy cercanos a los valores reales de potencia AC, mostrando un buen rendimiento en la predicción. Al igual que en los modelos “Pequeño” y “Mediano”, se identifican algunos outliers, principalmente en la sobreestimación de la potencia para valores cercanos a 0 y la subestimación en rangos específicos, aunque con diferencias entre las versiones. En el modelo “Híbrido” original, la subestimación se concentra entre los valores de 0.2 y 0.3, con una menor cantidad de outliers en comparación con el modelo “Mediano”, lo que se refleja positivamente en sus métricas. En contraste, la versión del modelo que incorpora las horas de Sol muestra un incremento en los outliers, tanto en la sobreestimación para valores cercanos a 0 como en la subestimación para el rango de 0.4 a 0.8, lo que indica un leve deterioro en la precisión respecto al modelo “Híbrido” original.

Dado que estos modelos “Híbridos” mostraron una mejora en rendimiento en comparación a los 3 modelos iniciales, se decidió desnormalizar los valores de potencia AC para observar cuán diferentes eran los valores de potencia predichos por el modelo en comparación con los valores reales de potencia AC, ambos en $[W]$.

Las métricas desnormalizadas obtenidas por los modelos “Híbridos” corresponden a las siguientes:

Tabla 6.7: Métricas desnormalizadas obtenidas al evaluar los modelos “Híbrido” y “Híbrido” considerando horas de Sol en el conjunto de prueba.

Métrica	Modelo “Híbrido”	Modelo “Híbrido” con horas de Sol
MSE	678814.6875	962173.7500
RMSE	823.9021	980.9045
MAE	259.3701	418.0567
R^2	0.9999	0.9999

Notar que el conjunto de prueba sigue siendo el mismo, la única diferencia es que las métricas fueron calculadas con los valores de potencia en $[W]$ en vez de una relación que toma valores entre 0 y 1.

Las métricas desnormalizadas mostradas en la tabla 6.7 destacan el alto rendimiento de ambos modelos “Híbrido”. Para el modelo “Híbrido” original, el $RMSE$ de 823.9 W y el MAE de 259.3 W indican un error de predicción promedio relativamente bajo, lo que refleja la alta precisión del modelo en la predicción de la potencia AC, con errores absolutos pequeños. El R^2 de 0.9999 refuerza esta precisión, sugiriendo que el modelo explica el 99.99% de la variabilidad en los datos, lo que demuestra un excelente ajuste. Por otro lado, el modelo “Híbrido” que incorpora las horas de Sol muestra un $RMSE$ de 980.9 W y un MAE de 418.0 W, lo que, aunque sigue indicando una predicción precisa, representa un ligero incremento en los errores promedio en comparación con el modelo original. A pesar de este aumento en los errores, el R^2 se mantiene en 0.9999, indicando que este modelo también explica el 99.99% de la variabilidad en los datos de potencia AC, aunque con una leve disminución en la precisión general.

Al graficar los resultados desnormalizados predichos por ambos modelos respecto a los valores reales de potencia del conjunto de prueba, se puede observar que los modelos, en general, estiman de buena forma los valores de potencia AC.

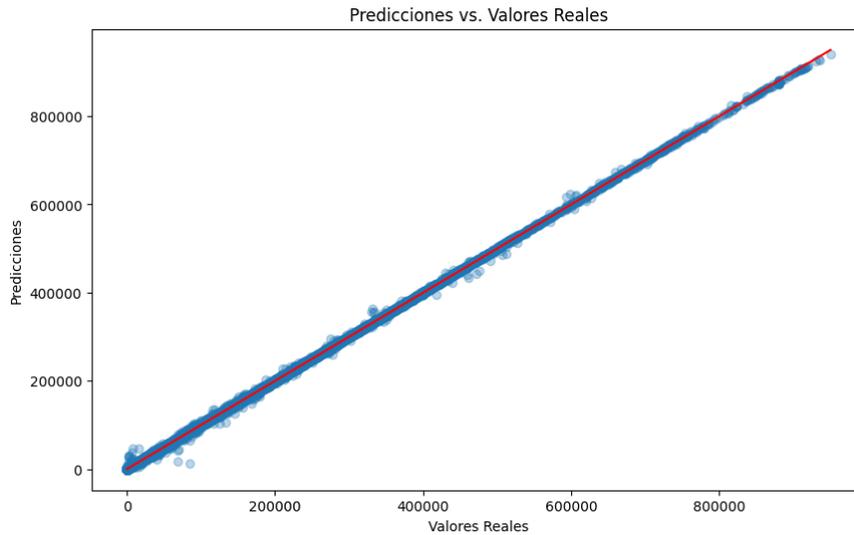


Figura 6.16: Comparación de resultados desnormalizados reales y predichos por el modelo “Híbrido”.

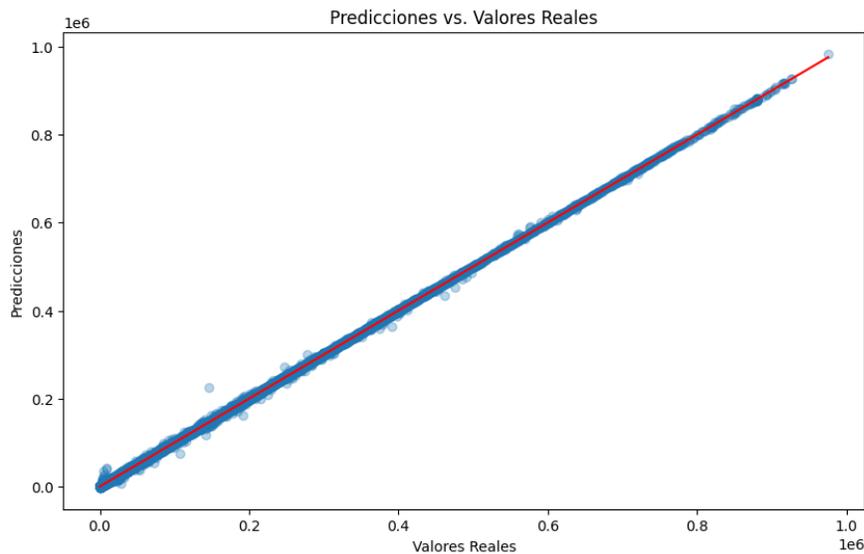


Figura 6.17: Comparación de resultados desnormalizados reales y predichos por el modelo “Híbrido” considerando horas de Sol.

Las figuras 6.16 y 6.17 muestran que los valores predichos por ambos modelos “Híbridos” son en su mayoría muy cercanos a los valores reales de potencia en $[W]$, lo que indica una alta precisión en las predicciones. La línea roja diagonal en ambos gráficos representa los valores reales de potencia AC, mientras que los puntos azules corresponden a las predicciones desnormalizadas de los modelos. En el caso del modelo “Híbrido” original, la mayoría de los puntos se agrupan muy cerca de la línea roja, aunque se identifican dos outliers donde el modelo subestima la potencia AC en aproximadamente 100 kW. Este alto grado de alineación entre predicciones y valores reales confirma el excelente rendimiento del modelo después de la desnormalización, manteniendo la consistencia observada en las métricas desnormalizadas. Por otro lado, el modelo “Híbrido” que incorpora las horas de Sol también muestra predic-

ciones precisas, pero presenta un outlier donde se sobreestima la potencia AC en un caso cercano a 200 kW. Aunque las métricas de este modelo siguen siendo buenas, son ligeramente inferiores a las del modelo original, lo que sugiere que la inclusión de la variable de las horas de Sol no mejoró significativamente la capacidad del modelo para relacionar periodos de baja irradiación con menores valores de potencia AC.

Capítulo 7

Conclusiones

7.1. Conclusiones

El proyecto desarrollado para la creación de gemelos digitales de plantas fotovoltaicas de escala comercial e industrial mediante el uso de la librería *PVLIB* y modelos basados en redes Transformer ha demostrado ser exitoso en varios aspectos clave. Los resultados obtenidos muestran mejoras significativas en la precisión y la capacidad de simulación de los modelos, proporcionando una base sólida para la optimización y gestión del portafolio de plantas de Solarity.

Comenzando por el modelo *PVLIB*, en su versión inicial proporcionada por la empresa subestimaba considerablemente la generación anual de las plantas fotovoltaicas, con un promedio de error de 7.13 %. Sin embargo, tras una serie de iteraciones y mejoras, se logró reducir este error de manera significativa. La versión final del modelo *PVLIB* mostró un promedio de error de -0.78% , indicando que subestima ligeramente la generación anual en comparación con *HelioScope*, pero con un error promedio absoluto de 3.09 %, alcanzando una precisión del 96.91 %.

El proceso de ajuste incluyó la mejora de los modelos de cálculo, la incorporación de nuevas variables de entrada y la optimización de los parámetros existentes. Estas mejoras permitieron una simulación más precisa y detallada de los sistemas fotovoltaicos, proporcionando resultados comparables a los de *HelioScope* y facilitando la generación de reportes completos a nivel de planta y de inversores individuales. La búsqueda de los mejores parámetros de simulación resultó en una precisión del 97.22 %. Aunque el proceso fue computacionalmente intensivo, su importancia radica en establecer una base de datos precisa que puede ser utilizada en simulaciones futuras, asegurando que las decisiones operativas y de mantenimiento se basen en datos confiables durante toda la vida útil de la planta.

Los modelos Transformer entrenados para estimar la potencia AC de plantas fotovoltaicas presentaron resultados diversos, destacándose especialmente el modelo “Híbrido”, que combinó los mejores hiperparámetros de los modelos “Pequeño” y “Mediano”. Este modelo “Híbrido” logró un *RMSE* de 0.0036, un *MAE* de 0.0012 y un R^2 de 0.9998, mostrando un rendimiento superior tanto en predicciones normalizadas como desnormalizadas, con un *RMSE* de 823.9 W y un *MAE* de 259.3 W.

En base a esto, se puede considerar que los objetivos del proyecto se cumplieron, ya que el modelo original de *PVLIB* fue modificado para obtener simulaciones más precisas (al ser comparadas con resultados obtenidos por *HelioScope*), a lo que además se modificó para que fuera modular, permitiendo así tener mayor capacidad de escalabilidad y mantención en el tiempo. Por parte de los modelos Transformers, las pruebas realizadas para estimar la potencia AC de una planta fotovoltaica teniendo como datos de entrada los valores meteorológicos y características de la planta, mostraron resultados satisfactorios con alta precisión lo que a su vez invita a realizar más pruebas en un futuro para conseguir un modelo más completo considerando más características eléctricas de las plantas fotovoltaicas. Finalmente, estos modelos fueron probados en diversas iteraciones variando diferentes parámetros o configuraciones iniciales de estos, siempre con el objetivo de obtener resultados lo más cercanos posibles al modelo de referencia *HelioScope*, precisión que logró ser obtenida y es reflejada en las buenas métricas obtenidas por los modelos *PVLIB* y Transformer.

7.2. Trabajo Futuro

Los resultados obtenidos indican que tanto el modelo *PVLIB* como los modelos Transformer tienen un gran potencial para ser utilizados en la creación de gemelos digitales de plantas fotovoltaicas. Sin embargo, existen varias áreas en las que se pueden mejorar estos modelos para aumentar su precisión y utilidad. Una dirección futura importante será seguir ajustando y mejorando el modelo de *PVLIB* para reducir las discrepancias observadas cuando se compara con el modelo de referencia basado en *HelioScope*. Esto asegurará una mayor precisión en las predicciones y reportes generados, proporcionando así una herramienta más fiable para la toma de decisiones.

En cuanto a los modelos basados en redes Transformer, que actualmente se enfocan en predecir la potencia AC de salida a partir de datos meteorológicos y características de la planta, se presentan oportunidades significativas para su expansión. Un área clave de trabajo futuro será ampliar el alcance de estos modelos para incluir la predicción de otros parámetros críticos como voltaje, corriente y temperatura de los paneles. Esta expansión podría lograrse ya sea entrenando un único modelo de Transformer capaz de generar un reporte integral, o desarrollando una serie de modelos especializados, cada uno enfocado en un aspecto específico de la planta fotovoltaica.

Otra línea de trabajo futura consiste en mejorar la precisión de las simulaciones mediante la implementación de piranómetros en plantas específicas, lo que permitiría medir con mayor exactitud la radiación real recibida por las plantas. Dado el alto costo de estos dispositivos y el amplio portafolio de más de 170 plantas de *Solarity*, se sugiere realizar pruebas en zonas estratégicas a lo largo de Chile para identificar clústers de plantas donde las diferencias en la radiación medida por sistemas satelitales y la real sean más significativas. Esto podría ayudar a refinar los modelos de simulación, mejorando la correlación entre los datos teóricos y los medidos en las plantas.

Además, una vez que se logre reducir la incertidumbre en los datos reales medidos en las plantas fotovoltaicas, se podría aprovechar la capacidad del modelo *PVLIB* para realizar análisis a nivel de inversor, e incluso a nivel de string. Aunque actualmente este nivel de

detalle no es necesario para *Solarity*, en el futuro, ajustar la precisión del modelo *PVLIB* a los datos medidos por cada inversor del portafolio podría ofrecer *insight* valiosos para la operación y mantenimiento de las plantas.

Bibliografía

- [1] Ministerio de Energía de Chile, “Energía Solar,” 2024, <https://solar.minenergia.cl/inicio>. [En línea; accedido: 20-08-2024].
- [2] Picó, S., “Desarrollo e implantación de energías renovables: situación actual y tendencias,” 2002, <https://upcommons.upc.edu/handle/2099.1/3428>.
- [3] Messenger, R. A. y Ventre, J., Photovoltaic Systems Engineering. CRC Press, 2010.
- [4] “Solar geometry,” n.d., <https://www.of-convalence.com/entries/solar-geometry>. Accessed: 2024-08-30.
- [5] Hu, C. y White, R. M., Solar Cells: From Basic to Advanced Systems. New York: McGraw-Hill, 1983.
- [6] Pearsall, N., The Performance of Photovoltaic (PV) Systems. Woodhead Publishing, 2017.
- [7] Masters, G. M., Renewable and Efficient Electric Power Systems. Wiley, 2004.
- [8] Duffie, J. A. y Beckman, W. A., Solar Engineering of Thermal Processes. Wiley, 2013.
- [9] Mallon, K. R., Assadian, F., y Fu, B., “Analysis of on-board photovoltaics for a battery electric bus and their impact on battery lifespan,” Energies, vol. 10, 2017, [doi:10.3390/en10070879](https://doi.org/10.3390/en10070879).
- [10] Green, M. A., Third Generation Photovoltaics: Advanced Solar Energy Conversion. Springer, 2001.
- [11] Widén, J. y Munkhammar, J., Solar Radiation Theory. Uppsala University, 2019, [doi:10.33063/diva-381852](https://doi.org/10.33063/diva-381852).
- [12] Roseline, P., Okonkwo, O., Mbeledogu, N., y Chinwe, G., “The benefits and challenges of intelligent self- diagnostic model for fault detection in photovoltaic system,” 2023.
- [13] Akshay, “What are the basic components of the photovoltaic system?,” Republic Of Solar, 2022, <https://arka360.com/ros/basic-components-of-photovoltaic-system/>. Accessed: 2024-08-26.
- [14] David, L., “Types of solar panels,” EcoWatch, 2022, <https://www.ecowatch.com/solar/types-of-solar-panels>. Accessed: 2024-08-26.
- [15] Luque, A. y Hegedus, S., Hegedus, S.: Handbook of Photovoltaic Science and Engineering. Wiley, New York. John Wiley&Sons, 2003, [doi:10.1002/0470014008](https://doi.org/10.1002/0470014008).
- [16] Deltavolt, “Inversores Fotovoltaicos,” <https://deltavolt.pe/pv-systems/inversores-red/>. [En línea; accedido: 20-08-2024].
- [17] Woodhouse, M., Feldman, D., Ramasamy, V., Smith, B., Silverman, T., Barnes, T.,

- Zuboy, J., y Margolis, R., “Research and development priorities to advance solar photovoltaic lifecycle costs and performance,” Rep. Tec. NREL/TP-7A40-80505, National Renewable Energy Laboratory, 2021.
- [18] Green Sarawak, “Going Solar Chapter 11 : Know Your Solar PV Panels Mounts and Supports.”, https://greensarawak.com/things-to-know-before-going-solar/going-solar-chapter-11-know-your-solar-pv-panels-mounts-and-supports/#google_vignette. [En línea; accedido: 20-08-2024].
- [19] EnerzyTech, “Types of Solar PV System.”, <https://www.enerzytech.com/resources/articles/13-types-of-solar-pv-systems>. [En línea; accedido: 20-08-2024].
- [20] Dierauf, T. *et al.*, “Weather-corrected performance ratio,” National Renewable Energy Laboratory (NREL), 2012, <https://www.nrel.gov/publications>.
- [21] Hansen, C. y Martin, C., “Photovoltaic system modeling: Uncertainty and sensitivity analyses,” 2015, [doi:10.13140/RG.2.1.4444.2081](https://doi.org/10.13140/RG.2.1.4444.2081).
- [22] Grieves, M. y Vickers, J., *Digital Twin: Mitigating Unpredictable, Undesirable Emergent Behavior in Complex Systems*. Springer, 2017.
- [23] Tao, F., Zhang, H., Liu, A., y Nee, A. Y. C., “Digital twin in industry: State-of-the-art,” *IEEE Transactions on Industrial Informatics*, vol. 15, no. 4, pp. 2405–2415, 2018.
- [24] Glaessgen, E. y Stargel, D., “The digital twin paradigm for future nasa and u.s. air force vehicles,” en *53rd AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conference 20th*, 2012.
- [25] Boschert, S. y Rosen, R., *Digital Twin—The Simulation Aspect*. Springer, 2016.
- [26] Negri, E., Fumagalli, L., y Macchi, M., “A review of the roles of digital twin in cps-based production systems,” *Procedia Manufacturing*, vol. 11, pp. 939–948, 2017.
- [27] Fuller, A., Fan, Z., Day, C., y Barlow, C., “Digital twin: Enabling technologies, challenges and open research,” *IEEE Access*, vol. 8, pp. 108952–108971, 2020.
- [28] Jain, P., Poon, J., Singh, J. P., Spanos, C., Sanders, S. R., y Panda, S. K., “A digital twin approach for fault diagnosis in distributed photovoltaic systems,” *IEEE Transactions on Power Electronics*, vol. 35, no. 1, pp. 940–956, 2020, [doi:10.1109/TPEL.2019.2911594](https://doi.org/10.1109/TPEL.2019.2911594).
- [29] Lu, Y., Liu, C., Wang, K., Huang, H., y Xu, X., “Digital twin-driven smart manufacturing: Connotation, reference model, applications and research issues,” *Robotics and Computer-Integrated Manufacturing*, vol. 61, p. 101837, 2019.
- [30] Yaqoob, I., Hashem, I. A. T., y Gani, A., “Internet of things architecture: Recent advances, taxonomy, requirements, and open challenges,” *IEEE Wireless Communications*, vol. 24, no. 3, pp. 10–16, 2020.
- [31] Yalçın, T., Solà, P., Stefanidou-Voziki, P., Dominguez-Garcia, J. L., y Demirdelen, T., “Exploiting digitalization of solar pv plants using machine learning: Digital twin concept for operation,” *Energies*, vol. 16, p. 5044, 2023, [doi:10.3390/en16135044](https://doi.org/10.3390/en16135044).
- [32] Will Holmgren, C. H. y Mikofski, M., “pvlib python: a python package for modeling solar energy systems,” *Journal of Open Source Software*, vol. 3, no. 29, p. 884, 2018, <http://dx.doi.org/10.21105/joss.00884>.
- [33] King, D. L., Boyson, W. E., y Kratochvil, J. A., “Photovoltaic array performance model,” 2004. Sandia National Laboratories.

- [34] Laboratories, S. N., “Pvlib python documentation,” 2015. Retrieved from PVLIB Documentation.
- [35] Bishop, C. M., Pattern Recognition and Machine Learning. Springer, 2006.
- [36] Goodfellow, I., Bengio, Y., y Courville, A., Deep Learning. MIT Press, 2016.
- [37] Sutton, R. S. y Barto, A. G., Reinforcement Learning: An Introduction. MIT Press, 2018.
- [38] Abu-Mahfouz, A. M. y Hamam, Y., “The role of machine learning in optimizing solar energy production: A comprehensive review,” Renewable Energy, vol. 135, pp. 143–152, 2018.
- [39] Montgomery, D. C., Peck, E. A., y Vining, G. G., Introduction to Linear Regression Analysis. Wiley, 2012.
- [40] Kutner, M. H., Nachtsheim, C. J., Neter, J., y Li, W., Applied Linear Statistical Models. McGraw-Hill/Irwin, 2005.
- [41] James, G., Witten, D., Hastie, T., y Tibshirani, R., An Introduction to Statistical Learning: with Applications in R. Springer, 2013.
- [42] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., y Polosukhin, I., “Attention is all you need,” en Advances in Neural Information Processing Systems, vol. 30, pp. 5998–6008, 2017.
- [43] Yi, J., Chen, Y., Li, J., Sett, S., y Yan, T. W., “Predictive model performance: Offline and online evaluations,” 2013, <https://dl.acm.org/doi/10.1145/2487575.2488215>.
- [44] Werra, L. V. *et al.*, “Evaluate & evaluation on the hub: Better best practices for data and model measurements,” 2022, <https://arxiv.org/abs/2210.01970>.
- [45] Hutchinson, B., Rostamzadeh, N., Greer, C., Heller, K., y Prabhakaran, V., “Evaluation gaps in machine learning practice,” 2022, <https://arxiv.org/pdf/2205.05256v1.pdf>.
- [46] Shankar, S., Herman, B., y Parameswaran, A. G., “Rethinking streaming machine learning evaluation,” 2022, <https://arxiv.org/pdf/2205.11473.pdf>.
- [47] Ding, Y., Liu, J., Xiong, J., y Shi, Y., “Revisiting the evaluation of uncertainty estimation and its application to explore model complexity-uncertainty trade-off,” 2020, <https://ieeexplore.ieee.org/document/9150782/>.
- [48] Labs, F., “Helioscope documentation,” 2015. Retrieved from HelioScope.
- [49] Sinha, A. y Chandel, S. S., “Review of software tools for hybrid renewable energy systems,” Renewable and Sustainable Energy Reviews, vol. 32, pp. 192–205, 2014.
- [50] Jain, P. y Sudhakar, K., “Solar pv performance evaluation using helioscope and pvsyst,” Renewable Energy, vol. 45, pp. 409–417, 2017.
- [51] Hill, R. y Mayer, T., “Integration of high-resolution meteorological data in solar pv simulation tools,” Journal of Solar Energy Engineering, vol. 140, no. 1, p. 012004, 2018.
- [52] Drury, E., Denholm, P., y Margolis, R., “The impact of different economic performance metrics on the perceived value of solar photovoltaics,” rep. tec., National Renewable Energy Laboratory (NREL), 2011.
- [53] Solcast, “Solcast documentation,” 2021. Retrieved from Solcast.

- [54] Ren, Z. y Foliente, G., “A methodology for predicting the performance of rooftop photovoltaic systems,” *Renewable Energy*, vol. 76, pp. 742–750, 2015.
- [55] O’Shaughnessy, E. y Cutler, D., “Solar pv system optimization using solcast data,” 2018.
- [56] Mermoud, A., *PVSyst 6.0: Software for the Study and Simulation of Photovoltaic Systems*. University of Geneva, 2012.
- [57] Freitas, S. y Collares-Pereira, M., “Evaluation of pv system performance using pvsyst software: A case study,” *Renewable Energy*, vol. 80, pp. 278–285, 2015.
- [58] Chianese, D. y Antonelli, G., “Pvsyst software analysis for the simulation of a photovoltaic system,” *Energy Procedia*, vol. 126, pp. 220–226, 2017.
- [59] Ayompe, L. M. y Duffy, A., “Analysis of the performance of a grid-tied pv system in ireland,” *Energy Conversion and Management*, vol. 87, pp. 654–662, 2014.
- [60] Nemet, G. F., “Beyond the learning curve: Factors influencing cost reductions in photovoltaics,” *Energy Policy*, vol. 34, no. 17, pp. 3218–3232, 2006.
- [61] Devlin, J., Chang, M. W., Lee, K., y Toutanova, K., “Bert: Pre-training of deep bidirectional transformers for language understanding,” *arXiv preprint arXiv:1810.04805*, 2019.
- [62] Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., y Sutskever, I., “Language models are unsupervised multitask learners,” *OpenAI*, 2019.
- [63] Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., y Houlsby, N., “An image is worth 16x16 words: Transformers for image recognition at scale,” *arXiv preprint arXiv:2010.11929*, 2021.
- [64] Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., y Amodei, D., “Language models are few-shot learners,” *arXiv preprint arXiv:2005.14165*, 2020.
- [65] Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., y Stoyanov, V., “Roberta: A robustly optimized bert pretraining approach,” *arXiv preprint arXiv:1907.11692*, 2019.
- [66] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., y Duchesnay, E., “Scikit-learn: Machine learning in python,” *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011, <http://jmlr.org/papers/v12/pedregosa11a.html>.

Anexos

Anexo A. Pseudo Códigos de PVLIB

A.1. Clase PlantModel

```
1 class PlantModel:
2
3     def __init__(id):
4         - Inicializa una planta utilizando el ID proporcionado
5         - Crea una conexión a la base de datos
6         - Establece el nombre de la planta y sus características
7
8     def get_plant_losses(id_planta):
9         - Consulta las pérdidas de la planta desde la base de datos
10        - Devuelve un diccionario con las pérdidas de la planta, si existen
11
12    def create_model(with_degradation):
13        - Obtiene los strings de la planta (los inversores)
14        - Obtiene la ubicación de la planta (latitud y longitud)
15        - Aplica las pérdidas de la planta según la base de datos o los parámetros
16        ↪ proporcionados
17        - Crea el modelo de inversores para cada inversor de la planta y lo almacena en
18        ↪ una lista
19
20    def simulate(start_str, end_str, with_degradation, meteo_data_source):
21        - Llama al método create_model para crear los inversores
22        - Crea un modelo meteorológico con la fuente de datos especificada
23        - Ejecuta el simulador (SimulationRunner) usando los inversores y el modelo
24        ↪ meteorológico
25        - Devuelve los resultados de la simulación
26
27    def simulate_plant(start_str, end_str, with_degradation, meteo_data_source,
28        ↪ resolution):
29        - Simula la planta con diferentes resoluciones (horaria, diaria, mensual, anual)
30        ↪ llamando al método simulate
```

Código A.1: Pseudo código de la clase PlantModel.

A.2. Clase InverterModel

```
1 class InverterModel:
2
3     def __init__(id_equipo, device_data, strings):
4         - Inicializa el inversor con los datos del dispositivo y los strings asociados
5
6     def create_model(site_location, plant_losses, with_degradation):
7         - Crea un array con los datos del string llamando al método create_array
8         - Obtiene los parámetros del inversor y calcula las pérdidas con pvwatts
9         - Crea un ModelChain para el inversor utilizando los datos del array y las
10            ↪ pérdidas
11         - Devuelve el ModelChain del inversor
12
13     def create_array(array_data):
14         - Crea un array de módulos fotovoltaicos para un inversor
15         - Asigna los parámetros de temperatura según el tipo de montaje
16         - Devuelve el array del sistema fotovoltaico
17
18     def calculate_pvwatts_losses(plant_losses, installation_date, with_degradation):
19         - Calcula las pérdidas totales del inversor utilizando el modelo de pvwatts
20         - Ajusta las pérdidas según la edad del sistema si 'with_degradation' es
21            ↪ verdadero
22         - Devuelve un diccionario con los parámetros de pérdidas
```

Código A.2: Pseudo código de la clase InverterModel.

A.3. Clase WeatherModel

```
1 class WeatherModel:
2
3     def __init__(plant, start_str, end_str, source):
4         - Inicializa el modelo meteorológico para la planta
5         - Obtiene los datos meteorológicos para la planta desde la fuente especificada
6
7     def set_weather(weather_df):
8         - Establece el DataFrame con los datos meteorológicos
9
10    def get_weather_df(start_str, end_str, source):
11        - Dependiendo de la fuente, obtiene los datos meteorológicos
12        - Devuelve un DataFrame con los datos meteorológicos
```

Código A.3: Pseudo código de la clase WeatherModel.

A.4. Clase SimulationRunner

```
1 class SimulationRunner:
2
3     def __init__(inverters, plant_id, strings, weather_model):
4         - Inicializa el simulador con los inversores, el ID de la planta, los strings y el
5           ↪ modelo meteorológico
6         - Almacena los datos meteorológicos para su uso en la simulación
7
8     def run_model():
9         - Ejecuta el modelo de cada inversor usando los datos meteorológicos
10        - Guarda los resultados de cada inversor en una lista
11
12    def inverters_results():
13        - Crea un DataFrame con los resultados de la simulación para cada inversor
14        - Combina los datos de los inversores y agrega las variables meteorológicas
15        - Devuelve un DataFrame con los resultados finales
```

Código A.4: Pseudo código de la clase SimulationRunner.